Extracting heavy metal stress indicators from remote sensing imagery using WOFOST model and wavelet packet decomposition algorithm

Zhao Xu1,* Shuang Zhao2, Xu Qian3
1 State Grid Energy Research Institute Co., Ltd., Changping District, Beijing 102209, China
2 School of Geology and Geometrics, Tianjin Chengjian University, Tianjin 300384, China
3 96669 Troops, Changping District, Beijing 102208, China
* Correspondence: xuzhao1@sgeri.sgcc.com.cn; Tel: +86-010-6660-3529

Abstract: Heavy metal pollution of crops seriously endangers food security and indirectly threatens human health. Direct measures in the fields and laboratories through on-site sample collection, testing, and analysis are time-consuming and labor intensive, thereby prohibiting their applications in large-scale monitoring. Remote sensing techniques provide an alternative means through examining above-ground vegetation status, e.g. leaf area index (LAI). Heavy metals, however, are typically accumulated in the root of crops, which may also be affected by a large number of external environmental factors besides heavy metals. The objective of this paper, therefore, is to identify heavy metal stress indicators of crops through integrating LAI extraction from remote sensing imagery, weight of rice roots (WRT) estimation by the World Food Study (WOFOST) model, and heavy metal stress indicator identification with the wavelet packet decomposition (WPD) method. First, LAI was retrieved from the HJ CCD data over three continuous years through constructing a relationship between LAI and normalized difference vegetation index (NDVI). Next, dry weight of rice roots (WRT) curves over these three continuous years were estimated using the WOFOST model with multi-temporal LAIs as inputs. Finally, a component (e.g. cfs 14) was identified to represent the heavy metal pollution status with the Wavelet Packet Decomposition (WPD) of the WRT curves of these three years. Validation results suggest that the identified component can successfully represent different levels of heavy metal stress.

1. Introduction
Heavy metal pollution in soil is an essential environmental pollution problem in China. It not only affects crop growth but also indirectly threatens human health [6][7]. Therefore, it is important to monitor heavy metal pollution in crops accurately [5][8]. Traditional methods for monitoring heavy metal pollution can be roughly divided into laboratory monitoring and field monitoring; both utilize on-site sample collection, testing, and analysis, which require much more time, manpower, and material resources, thereby prohibiting their applications for large-scale monitoring [1][9]. In recent years, remote sensing techniques have been employed to detect heavy metal pollution in soils, through examining physiological changes in crops caused by heavy metal pollution.
To achieve the goals of extracting heavy metal stress indicators from remotely sensed imagery, we developed a procedure that includes 1) deriving time-series LAI of rice from multi-temporal remote sensing imagery, 2) estimating time-series WRT through employing the WOFOST model with LAIs as inputs, and 3) extracting heavy metal stress indicators from estimated WRT images using the Wavelet Packet Decomposition (WPD). This study aims at finding an effective indicator of heavy metal stress through decomposing the WRT curves in crop growth period of continuous years and select the most stable components.

2. Materials and Methods

2.1. Study Area and Data

2.1.1. Study Area. The study area is located in Zhuzhou city (112°17′-114°07′ E, 26°03′-28°01′ N), The main pollution sources in Zhuzhou are industrial pollution, such as mining, non-ferrous metal smelting, machinery parts processing. In this research, the study area includes two parts: a large rice-growing area (labeled A) as the experimental area, and two small rice-growing areas (labeled B and C) as the validation area (Figure 1).

2.1.2. Data collecting and processing. The dataset employed in this study includes remote sensing imagery, meteorological data, and field-measured data. For the remote sensing data, we employed the Charge-coupled Device (CCD) data of the Environment and Disaster Reduction Small Satellites acquired from the China Center for Resources Satellite Data and Application. For this study, the CCD data is advantageous due to its high time resolution (2 days) and mid-high spatial resolution (30 m). Based on the consideration of data quality and crop growth period, 14 images over three years (2013, 2014 and 2015) were employed for experiments, and 5 images of 2016 were used for validation.

2.2. Methods

Prior knowledge indicated that the roots absorbed and enriched heavy metals much more than other parts of the rice plant, so the WRT was used to obtain the most stable characteristic as an index reflecting the heavy metal stress. There are two parts of the research method: (1) An improved WOFOST model with two stress factors (the f in the b part of Figure 2) was used to simulate the WRT, and (2) the wavelet
packet decomposition was used to decompose the WRT to get the most stable characteristic. The procedure is generalized in Figure 2.

![Figure 2](image)

**Figure 2.** A general flow chart of the improved WOFOST model and the wavelet packet decomposition for obtaining stable components as a representation of heavy metal stress.

### 2.2.1. Simulation of time-series WRT from the improved WOFOST model

WOFOST model is a descriptive dynamic model for simulating plant growth processes by considering the plant growth climate and other parameters. The original WOFOST model is conducted without heavy metal pollution, thus an improved WOFOST model was established to solve this problem.

WOFOST model adjusts the parameters by measuring and simulating values. The measured values of LAI can be obtained by CCD image data. The formulas for inversion of LAI using NDVI is shown as follows:

\[ y = 0.3629e^{3.075x} \]  

(1)

Where, \( x \) is the value of NDVI and \( y \) is the value of LAI [2].

The correlation coefficient (\( R^2 \)) between measured data and simulated data is greater than 0.8, which proves that it is feasible to use this inversion formula to calculate LAI(Figure 3).
The improved WOFOST model embeds two stress factors ($f_{DTGA}$, $f_{CVF}$). The improved WOFOST model can reduce the difference between the measured and simulated values effectively by optimizing the value of the two stress factors that are embedded[2].

$$f_{DTGA}(D) = f_{DTGA} \times DTGA(D), f_{DTGA} \in (0.65,1)$$  \hspace{1cm} (2)

$$f_{CVF}(D) = f_{CVF} \times CVF(D), f_{CVF} \in (0.65,1)$$  \hspace{1cm} (3)

where $D$ is the day number of the year, $DTGA(D)$ and $CVF(D)$ are, respectively, the daily total gross assimilation and carbohydrate-to-dry matter conversion coefficients under the potential growth level of day D; $f_{DTGA}(D)$ and $f_{CVF}(D)$ correspond to the daily total gross assimilation of CO$_2$ and the carbohydrate-to-dry matter conversion coefficient, respectively.

The stress factors were tuned by obtaining the optimal value of the difference between simulated LAI value and the retrieved LAI value. The particle swarm optimization algorithm (PSO) was used to achieve the optimal value:

$$Q = \frac{1}{N} \sum_{i=1}^{N} (LAI_{R(ti)} - LAI_{S(ti)})^2$$  \hspace{1cm} (4)

Where $Q$ is the value of the merit function, $LAI_{R(ti)}$ is the LAI derived from remotely sensed data on day $ti$, $LAI_{S(ti)}$ is the LAI simulated by the WOFOST model on day $ti$, and $N$ is the number of images.

2.2.2. WRT time-series decomposition using WPD. The WPD method, as well as the wavelet decomposition method for comparison purpose, was developed, and the decomposition structures are shown in Figure 4.

$$\text{Figure 4. Wavelet decomposition and wavelet packet decomposition (WPD) flow chart with a three-layer structure.}$$

Where the L represents the signal through low-pass filtering, and correspondingly, the H represents the signal through high-pass filtering. Compared to wavelet decomposition, which decomposes the low
frequency signal continuously, WPD decomposes the high and low frequency bands at the same time, and the frequency bands are divided into homogeneous bands.

The WPD formula can be expressed as follows:

\[ P_j^{2i-1} = LP_{j-1}^{2i-1}(t) \]  
\[ P_j^{2i} = HP_{j-1}^{2i}(t) \]

Where \( t=1,2,\ldots,N/2^j \); \( i=1,2,\ldots,2^j \); \( j=1,2,\ldots,j \); \( j \) is the maximum number of decomposition layers; \( N \) is the number of the sampling points; the spectral original signal is decomposed into \( 2^i \) signals by the \( j^{\text{th}} \) layer of the wavelet packet; and \( L \) and \( H \) are the low-pass filter and the high-pass filter, respectively.

In this study, a three-layer WPD was used to decompose the simulated WRT curves. The WPD analysis yielded 15 components, named cfs1-15. Component cfs1 is the simulated WRT curve (the original signal). In the first decomposition layer, the high frequency component is cfs2 and the low frequency component is cfs3. In the second decomposition layer, cfs2 was decomposed into the high frequency component cfs4 and the low frequency component cfs5, cfs3 was decomposed into the high frequency component cfs6 and the low frequency component cfs7. In turn, the components of the third layer decomposition were obtained from the second layer.

In order to obtain the most stable component of all decomposition components, the correlation of decomposition curves is needed to analyse. The related tables and figures can only reflect the relationships between the components, but cannot exactly show the details of the relationship between the components. The correlation coefficient designed by Pearson can be used to reflect the degree of closeness of the relationship between variables[3].

The formula is as follows:

\[ R = \frac{\sum_{i} X_i Y_j - \frac{\sum_{i} X_i \sum_{j} Y_j}{N}}{\sqrt{\left( \sum_{i} X_i^2 - \frac{\left(\sum_{i} X_i\right)^2}{N} \right) \left( \sum_{j} Y_j^2 - \frac{\left(\sum_{j} Y_j\right)^2}{N} \right)}} \]

\( (i = 2013,2014,2015; j = 2014,2015,2013) \)

Where \( R \) is the correlation coefficient, \( N \) is the length of crop growth period, \( X_i \) and \( Y_j \) the component value of the corresponding year, respectively.

3. Results and Discussion

3.1. Decomposition of time-series WRT and extraction of stable component

The WRT curves of the growth period show similar trends but obviously different values over three years (Figure 5). As no significant changes occurred in the exterior environment, the degree of heavy metal stress in soil did not change drastically. In a former study, it was found that WRT only was not sufficient as the monitoring indicator of the heavy metal stress, and a number of other factors gave rise to the difference of WRT values among years. Light, precipitation, fertilization, irrigation, pests and diseases, and other stress factors are different year to year, while the heavy metal stress is relatively stable. The high frequency decomposition component in each layer shows the same trend as the original signal. While the low frequency decomposition components are complicated, there are some differences in positive or negative directions.
In order to quantitatively compare the curves, the correlation coefficients of the curves for three years were calculated and are summarized in Table 1. The range of the average correlation for the three curves is from -0.058 to 0.97. The correlation coefficient is regarded as the evaluation index of stability between components. A previous study indicated that almost all the information of the original signal is concentrated in the low frequency component [4]. As such, the high frequency components of each layer, included cfs2, cfs4, and cfs8 (Figure 6 (c),(d), and (h)), whose correlation coefficients are 0.963, 0.966, 0.970, respectively, are not counted in the stability analysis. In addition to the high frequency decomposition components, the 13th component, cfs14 (Figure 6 (n)), has the optimal average value of 0.693, even though other components have higher correlation coefficients among years. For example, the correlation coefficient for cfs9 is 0.740 (Figure 6 (i)), but its performance is not stable; and the difference between 2013 and 2015 are significantly lower than the average. As various factors were considered, the cfs14 component was considered to be the most stable component for representing heavy metal stress monitoring.

**Table 1. Correlation coefficients of the curves for three years.**

| Component | $R_{2013−2014}$ | $R_{2014−2015}$ | $R_{2013−2015}$ | Average  |
|-----------|-----------------|-----------------|-----------------|----------|
| CFS2      | 1               | 0.943           | 0.946           | 0.963    |
| CFS3      | 0.386           | 0.139           | 0.311           | 0.279    |
| CFS4      | 1               | 0.947           | 0.95            | 0.966    |
| CFS5      | 0.887           | 0.123           | 0.182           | 0.397    |
| CFS6      | 0.515           | -0.304          | -0.247          | -0.012   |
| CFS7      | 0.562           | 0.473           | 0.596           | 0.544    |
| CFS8      | 1               | 0.954           | 0.957           | 0.97     |
| CFS9      | 0.928           | 0.954           | 0.339           | 0.74     |
| CFS10     | 0.884           | -0.044          | 0.012           | 0.284    |
| CFS11     | 0.931           | 0.424           | 0.407           | 0.587    |
| CFS12     | 0.203           | -0.088          | -0.039          | 0.025    |
| CFS13     | 0.868           | -0.551          | -0.492          | -0.058   |
| CFS14     | 0.767           | 0.658           | 0.653           | 0.693    |
| CFS15     | 0.379           | -0.153          | 0.113           | 0.113    |

Note: $R_{2013−2014}$ is the correlation coefficient of decomposition results between 2013 and 2014; $R_{2014−2015}$ is
the correlation coefficient of decomposition results between 2014 and 2015; \( R_{2013-2015} \) is the correlation coefficient of decomposition results between 2013 and 2015; the average is the average correlation coefficient value of the decomposition results.

Figure 6. The results of wavelet packet decomposition.

The cfs14 is a low frequency component, and the frequency range of cfs14 is from -0.91 to 0.75. Before the 230th day, the curve has small fluctuations, from -0.4 to 0.3, while the curve for 2015 shows small differences from the other curves. After the 230th day, the rice enters the ripening phase, the roots stop growing and gradually lose weight, and the WRT is not affected by heavy metals anymore. The difference between 2013 and 2014 is minimal, and the two curves are approximately coincident before the middle jointing-booting phase (Figure 7).

The WPD method we used has proven to be applicable and advantageous over other methods, largely because it decomposes the signal precisely by decomposing the high frequency and low frequency parts together.
3.2. Validation of stable characteristic indicated the status of heavy metal stress

A validation was carried out using data from areas in different heavy metal stress status of 2016 for evaluating the status of heavy metal stress by using the stable characteristic. The cfs14 curves which were obtained from the WRT curves as the previous work is shown in Figure 8. There are four curves presented in Figure 8, potential productivity level, mild level, moderate level, and severe level. Potential productivity level means the crops grow in the proper moisture and nutrition conditions. Meanwhile, the crops were not subjected to any heavy metal stress. The ranges of cfs14 value in different heavy metal stress level show significant difference. The potential productivity level ranges from -0.879 to 1.245, the mild level ranges from -0.684 to 0.970, the moderate level ranges from -0.572 to 0.812, and the severe level ranges from -0.510 to 0.722. With the decrease of the heavy metal stress level, the ranges of value are gradually reduced. And the differences became more apparent after the middle heading and flowering phase.

The validation results indicate that the cfs14 can distinguish heavy metal stress levels effectively, though the specific mechanism of results needs further explanation. The previous studies about heavy metal pollution had focused on the data processing, analyzing, and comparison of different areas, in order to obtain an indicator that represents the degree of the heavy metal pollution. The external environmental differences of different areas have not been eliminated. This study characterized the effect of heavy metal stress on crops and obtained a stable characteristic that can be seen as the representation of heavy metal stress in rice. The results also confirm the feasibility of long time series analysis for heavy metal crop stress monitoring. This implies that remote sensing is subject to environmental concern. According to the analysis of heavy metal stress at different levels, we determined a heavy metal pollution stress indicator that removes the impact of other stress.
4. Conclusion
In this study, a new representation of heavy metal stress is obtained. The new representation, cfs14 of the WPD components, was extracted from the time-series WRT curves for monitoring heavy metal stress. An improved WOFOST model with two stress factors was applied to obtain the WRT by assimilating the LAI, which was retrieved from the CCD data. WPD, whose kernel function is the Db5, was used to acquire the decomposed components. By comparing the correlation coefficients of the time-series WRT over three years, the average correlation coefficient for cfs14 is 0.693, and it is the optimal value; thus, cfs14 is considered to be the most stable component. In conclusion, cfs14 showed good correlation among three years though the WRT curves of three years varied, and it was proved to be able to distinguish between situations of heavy metal stress in different status by the validation. Therefore, the long time series analysis method shows a great ability to monitor heavy metal stress by eliminating the interference of external environmental factors. The cfs14 can be used as a representation of the heavy metal stress apart from the effects of the external environment. This study indicated that the time-series method demonstrates good applicability for heavy metal stress monitoring in the future.

Acknowledgments
This work was supported by the Young Talents Programme of State Grid Energy Research Institute Co., Ltd. (Research on the new urbanization development indicators of energy and electricity supply and demand analysis based on night lighting remote sensing data for smart grid, No. XM2018020037579) and the Science and Technology Project of State Grid Corporation of China (Research and Development of Comprehensive Analysis and Forecasting Model System for Multi-regional Energy Supply and Demand in China).

References
[1] Clemens S. (2006). Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie, 88(11), 1707-1719.
[2] Jin M., Liu X. N., Wu L., & Liu M. L. (2015). An improved assimilation method with stress factors incorporated in the WOFOST model for the efficient assessment of heavy metal stress levels in rice. International Journal Of Applied Earth Observation And Geoinformation, 41, 118-129. doi:10.1016/j.jag.2015.04.023
[3] Lin L. I. (1989). A CONCORDANCE CORRELATION-COEFFICIENT TO EVALUATE REPRODUCIBILITY. Biometrics, 45(1), 255-268. doi:10.2307/2532051
[4] Liu M., Liu X., Ding W., & Wu L. (2011). Monitoring stress levels on rice with heavy metal pollution from hyperspectral reflectance data using wavelet-fractal analysis. International Journal Of Applied Earth Observation And Geoinformation, 13(2), 246-255.

[5] Liu Y., Wen C., & Liu X. (2013). China's food security soiled by contamination. Science, 339(6126), 1382-1383.

[6] Shi T. Z., Liu H. Z., Wang J. J., Chen Y. Y., Fei T., & Wu G. F. (2014). Monitoring Arsenic Contamination in Agricultural Soils with Reflectance Spectroscopy of Rice Plants. Environmental Science & Technology, 48(11), 6264-6272. doi:10.1021/es405361n

[7] Wang M. L., Lynch J. P., & Sohn H. (2014). Introduction to sensing for structural performance assessment and health monitoring Sensor Technologies for Civil Infrastructures (Vol. 55, pp. 1-22): Woodhead Publishing.

[8] Werbrouck S. P. O., & Zhang G. P. (2015). Plant growth regulators involved in abiotic stress: preface. Plant Growth Regulation, 76(1), 1-2. doi:10.1007/s10725-015-0057-7

[9] Yadav S. K. (2010). Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. South African Journal Of Botany, 76(2), 167-179. doi:10.1016/j.sajb.2009.10.007