Research on Data Quality Control Method of Eucalyptus Plantation Sensor Network

Deng Guang¹, *, Yu Xinwen¹, Fan Dongpu¹, Zhu Chen² and Zhang xu¹

¹Research Institute of Forest Resource Information Techniques, China Academy of Forestry Sciences, 100091 Beijing, China
²Suzhou Yingchuang Kyushu Information Technology Co., Ltd., Suzhou, China

Abstract. The Internet of Things technology can provide data support for precise management measures and reflect the effect of management activities in time through long-term monitoring of plantation forest stands and environmental factors. However, the networking and deployment of the Internet of Things based on sensor network in the plantation forest has just started, and the stability of the sensor network itself is not very reliable, so the data quality will be greatly affected. In this paper, based on the experimental data and the experience of continuous data post-processing such as meteorology, the data quality of eucalyptus plantation forest sensor network is preliminarily analyzed. Then, on this basis, some algorithms of data interpolation and data integration are given according to different observation indexes, which lays a foundation for the further application of sensor network data of plantation forest in the future.

1. Introduction
At present, general problems such as poor plantation quality and low quality of plantation forest are widespread in China. Accurate improvement of forest quality, that is, based on specific stand characteristics, expected functions and objectives, the implementation of precise management plans and measures, and the process of comprehensively improving forest ecological, social and economic benefits. Sensor networks in the planting forest can provide help for long-term accurate monitoring of forest quality.[1,2] Eucalyptus plantation is a forest under intensive management and operations, so its potential impacts or contribution on biodiversity may vary with respect to management. To commit sound environmental responsibility, we develops a set of plantation sensor network to measure the environmental and growth changes of eucalyptus plantation. Through long-term observation data, it will lay a foundation for scientific and sustainable management of eucalyptus plantation in the future. Firstly, this paper describes the deployed sensor network of eucalyptus plantation, and then analyzes the obtained data. Then, according to meteorological data and DBH measurement data, some methods of judging abnormal values and quality control are given. Finally, through some published studies on monitoring of eucalyptus plantation, the practical application value of the data obtained by the sensor network of eucalyptus plantation is illustrated, and the future research prospect is given.

2. Study area and data
The study area is located in Yinling branch of Guangxi Gaofeng Forest Farm, which belongs to Xixiangtang District of Nanning City, Guangxi. Please refer to the paper[3] for the introduction of study area and data. What needs to be added is that we have deployed monitoring sites in the field twice. First three node sensor devices were deployed in the eucalyptus plantations which initial planted in 2015 on July 3, 2019. At the same time, two microclimate observation devices were deployed, one in the...
eucalyptus sample field and the other outside the forest. Then four node sensor devices were deployed in the eucalyptus plantations which initial planted in 2019 in January, 2020. The three nodes in the first deployment are numbered 1, 2 and 3, and the four nodes in the second deployment are numbered 4, 5, 6 and 7. The power supply system in the forest uses solar panels, but it does not perform well in the forest. Therefore, except node 5, the nodes deployed for the second time have a long continuous working time, and the other three nodes have no data acquisition soon. Therefore, in August 2020, we went to the experimental site to replace the solar panels, and all the nodes have obtained data up to now. The ground deployment and platform of sensor network for eucalyptus plantation have been described in detail in published paper[3], so they are not described here.

3. Method

The theme of this article is to deal with the problems often encountered in the research of forestry Internet of Things, that is, what to do when faced with the data whose quality cannot meet the expectations. There are many restrictions on the normal operation of the sensor in the field, including the power supply condition and the environment where the sensor is placed. For example, the first generation products of soil temperature sensor and soil moisture sensor in eucalyptus plantation can not adapt to the local rainy and backward irrigation, and the sensors are damaged by rain.

The characteristics of the field sensor determine that long-term stable and high-quality data acquisition is unrealistic. In order to obtain high-quality and long-term stable IoT data, sensors, power supply systems, layout methods, networking methods and transmission networks that are suitable for the field environment are indispensable, and long-term professional personnel are required to maintain the monitoring network on site and on the platform. These involve the cost of equipment and the cost of personnel input. We should consider the significance of online data acquisition for plantation monitoring.

First of all, we need to know that the principle and realistic composition of field deployment of sensor networks determine that its instability is the normal state. In a certain period of time, the technical improvements of sensors, power supply systems and transmission networks adapted to the field environment are limited. The significance of the sensor network data management platform lies in the continuous improvement of sensors, power supply systems, layout methods, networking methods and transmission networks after discovering data problems. Because only when the data obtained by the sensor is correct and the power supply system guarantees certain continuous data, the following data analysis and application can be carried out. Figure 1 shows such a continuous iterative process.

For the monitoring of eucalyptus plantation sensor networks, different application research purposes determine the quality requirements of different continuous measurement data. It should be said that there is no uniform data quality criterion. This paper only discusses data quality control of some microclimate observation factors and DBH sensors.

For the monitoring of eucalyptus plantation in Guangxi, if the data of microclimate weather stations inside and outside the forest are missing for 3 days, the data of two days before and after the missing value should be averaged and then replaced. If the missing value exceeds 3 days, the data quality should indicate that the data quality is inferior reliable, and replace it with the latest standard weather station data. The data that can be replaced by interpolation here are daily average, daily maximum and daily
minimum. See formula 1 for the interpolation formula, where \( \hat{x}_i \) represents the interpolation value of the missing value of the i-day temperature record, and \( \bar{x}_i \) and \( S_j \) represent the long-term average and standard deviation of the i-day temperature in the interpolation station. The missing daily raw data, that is, the meteorological record value of ten minutes, cannot be interpolated using above method.

\[
\hat{x}_i = \bar{x}_i * S_j + \hat{x}_i
\]  

Interpolation of DBH value of trees is relatively simple. No matter how much data is missing in the middle, the data of the day before the missing value is used for interpolation. This method is relatively simple, but it can ensure the physical rationality of DBH value of trees. In practice, we also find that the continuous DBH value jumps, and the abnormal values that become larger or smaller should be removed, and the normal data of the previous day should still be used for interpolation.

4. Results and Discussions

Figure 2 (1) shows the temperatures outside the eucalyptus plantation in August, 2020. Unfortunately, the data missing from August 1 to August 18 is difficult to be interpolated by the monitoring station's own data, and its data quality should be medium. The data missing on August 22 can be interpolated by the method in the previous section. It is calculated that the maximum temperature on August 28th is 30 degrees, the average temperature is 28 degrees, and the minimum temperature is 25 degrees.

Figure 2 (2) shows the temperatures inside and outside the eucalyptus plantation on December 1, 2020. Temperature 1 is the temperature of microclimate observation in the forest, and temperature 2 is the temperature of microclimate observation outside the forest. We can see the same change trend and different details of the temperature inside and outside the forest. Using the present observation technology and the data quality control method in this paper, we can do more research on the relationship between microclimate and environment and growth of eucalyptus forest.

5. Conclusions

In this paper, we discuss the imperfect plantation sensor network data, which is a topic rarely discussed in other papers. Then we give data quality control methods based on the real data, which are simple and practical. Then, through practical examples, compared with previous studies, it gives how to use existing data to get some new understanding of application. Of course, this paper is only the beginning of this
research. With the accumulation of data, there will be more and more problems in the quality control of sensor networks. Many important sensor data interpolation methods will also be discussed in the future.

Acknowledgments
This work was supported by National Key Research and Development Program of China under Grant No. 2017YFD0600901.

References
[1] Campbell, John L; Rustad, Lindsey E.; etc.. Quantity is nothing without quality: automated QA/QC for streaming sensor networks [J]. BioScience. 63(7): 574-585 (2013)
[2] https://smartforests.org
[3] Deng Guang,Yu Xinwen,Fan Dongpu,Zhu Chen. (2019) Preliminary Study on Data Fusion Based on Internet of Things of Eucalyptus Plantation. ICGEC 2019.E3S Web of Conferences. 131(7):01117
[4] Tan Shaoman;Huang Jinlong.Preliminary study on microclimate characteristics of Eucalyptus torii mixed forest .ACTA ECOLOGICA SINICA.5(3):241-248(1985)
[5] Chen Liyan.Study on soil water and water use efficiency of Eucalyptus plantation on Multi-time scale[D].Hainan Normal University(2019)