

Article

River Basin Cyberinfrastructure in the Big Data Era: An Integrated Observational Data Control System in the Heihe River Basin

Jianwen Guo ^{1,2,*} , Minghu Zhang ³, Qingsheng Shang ⁴, Feng Liu ¹, Adan Wu ^{1,2} and Xin Li ^{5,6} 

- ¹ Key Laboratory of Remote Sensing of Gansu Province, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; liufeng@lzb.ac.cn (F.L.); wuadan@lzb.ac.cn (A.W.)
 - ² Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
 - ³ School of Computer and Communication Technology, Lanzhou University of Technology, Lanzhou 730050, China; zhangmh@lut.cn
 - ⁴ School of Information Engineering, Lanzhou University of Finance and Economics, Lanzhou 730020, China; shangqsh@lzufe.edu.cn
 - ⁵ National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China; xinli@itpcas.ac.cn
 - ⁶ CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China
- * Correspondence: guojw@lzb.ac.cn

Abstract: River basin cyberinfrastructure with the Internet of Things (IoT) as the core has brought watershed data science into the big data era, greatly improving data acquisition and sharing efficiency. However, challenges in analyzing, processing, and applying very large quantities of observational data remain. Given the observational needs in watershed research, we studied the construction of river basin cyberinfrastructure and developed an integrated observational data control system (IODCS). The IODCS is an important platform for processing large quantities of observational data, including automated collection, storage, analysis, processing, and release. This paper presents various aspects of the IODCS in detail, including the system's overall design, function realization, big data analysis methods, and integrated models. We took the middle reaches of the Heihe River Basin (HRB) as the application research area to show the performance of the developed system. Since the system began operation, it has automatically received, analyzed, and stored more than 1.4 billion observational data records, with an average of more than 14 million observational data records processed per month and up to 21,011 active users. The demonstrated results show that the IODCS can effectively leverage the processing capability of massive observational data and provide a new perspective for facilitating ecological and hydrological scientific research on the HRB.

Keywords: Heihe River; river basin cyberinfrastructure; wireless sensor networks; automated observational data integration



Citation: Guo, J.; Zhang, M.; Shang, Q.; Liu, F.; Wu, A.; Li, X. River Basin Cyberinfrastructure in the Big Data Era: An Integrated Observational Data Control System in the Heihe River Basin. *Sensors* **2021**, *21*, 5429. <https://doi.org/10.3390/s21165429>

Academic Editor: Paolo Bellavista

Received: 29 June 2021

Accepted: 8 August 2021

Published: 11 August 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/>).

1. Introduction

Big data, first proposed by Alvin Toffler in *The Third Wave*, is characterized by broad sources, large volumes, multiple modes, and high values, and it is changing how people understand the world [1–3]. Big data has started to support the rapid development of Earth sciences [4–6]. Specifically, it has begun to play an increasingly important role in the research of Earth surface system sciences, which mainly focus on the overall characteristics of the interaction among the elements in the hydrosphere, pedosphere, atmosphere, and biosphere [7–9]. Therefore, big data can well support the integrated research of regional ecological economy [10]. After years of the constant construction of Earth surface observational systems, massive quantities of observational data have been accumulated

into big Earth data at a considerable scale. In particular, with the rapid development of technologies related to the Internet of Things (IoT) [11,12], observational systems on the Earth's surface are becoming increasingly mature, with the quantity of observational data reaching the petabyte scale and the number of modes of observational data significantly increasing [11,13,14]. The rise of the IoT has made it possible to acquire Earth surface observational and monitoring data in real time. We reviewed, compared, and analyzed several typical instances of data processing systems for big data, as shown in Table 1. Most of these systems were constructed for data delivery, storage, and visualization. However, unified standards and integrated systems of descriptions, organization, transmissions, interfaces, management, and applications of massive quantities of observational data are missing. This issue has become one of the most significant challenges for managing and sharing big Earth data [15–17]. In particular, regarding observational data from the IoT, which are usually termed streaming data and feature high speeds, large volumes, and uncertainties, traditional data reception, management, and visualization are confronted with challenges. Therefore, we must develop fully automated data receiving, intelligent data quality control, and more efficient and convenient data management and visualization methods.

Table 1. Typical instances of big data processing systems.

No.	System Name	Year	Main Functions	References
1	BigDancing	2015	A big data cleaning system to tackle efficiency, scalability, and ease-of-use issues in data cleaning; it can be run on most general-purpose data processing platforms, ranging from DBMSs to MapReduce-like frameworks.	[18]
2	SWITCH	2019	It offers a flexible co-programming architecture that provides an abstraction layer and an underlying infrastructure environment, which can help to both specify and support the life cycle of time-critical cloud native applications.	[19]
3	DRIP	2019	It was developed for the dynamic optimization of data services in research support environments and might be used for a number of similar applications involving distributed services and large, dynamic datasets with further investigation and development.	[20]
4	SPS-IUTO	2020	To achieve significant improvements in terms of energy and redundant data, a matrix completion-based sampling point selection joint intelligent unmanned trajectory optimization (SPS-IUTO) scheme for unmanned aerial vehicles (UAVs) was proposed to plan sampling points for UAVs in both time and space.	[21]
5	BD-VTE	2020	A novel baseline data based verifiable trust evaluation (BD-VTE) scheme was proposed to guarantee security at a low cost for massive data. The BD-VTE scheme includes a verifiable trust evaluation (VTE) mechanism, an effectiveness-based incentive (EI) mechanism, and a secondary path planning (SPP) strategy, which are used for reliable trust evaluation, reasonable reward, and efficient path adjustment, respectively.	[22]
6	DRMCS	2020	DRMCS, a data collection scheme for mobile crowdsensing vehicular networks, was proposed to enhance the data collection rate in vehicular networks for opportunistic communication.	[23]
7	SDAC	2021	A novel secure and dynamic access control (SDAC) model was developed for IoT networks (smart traffic control and roadside parking management). It allows IoT devices to securely communicate and share information through using wired and wireless networks (cellular networks or Wi-Fi).	[24]
8	aiRe	2021	This open-access tool simplifies air quality data analysis and visualization, with the desirable effects of removing ownership costs, fostering appropriation by nonexpert users, and ultimately promoting informed decision making for the general public and local government authorities.	[25]

Cyberinfrastructure (CI) is the bridge between information engineering and scientific research [26]. With the help of information technology and network technology, CI integrates observational instruments, storage devices, computing devices, and other resources to provide more efficient, faster, and more flexible collaboration space for scientific research. Therefore, as an application instance of CI—an integrated observational data control system (IODCS) that combines the IoT, wireless sensor networks (WSNs), cloud computing, and data visualization—is developed in this paper. The IODCS is considered a standardized system that makes it possible to achieve the automated reception and storage, distributed storage, quality control, security sharing, and visualized analysis of observational data according to descriptive data specifications. For instance, if the observational data obtained by each observational system could be automatically received and stored according to the database index type and data format, data storage efficiency would be improved. The introduction of distributed storage into the system serves as important technical support for the scientific and efficient management of massive quantities of observational data. Developing and integrating automated data quality control can achieve automated anomaly detection and observational data processing. Moreover, data sharing is an important driving force for promoting Earth surface research, and online data visualization allows users to more intuitively and conveniently analyze and understand data.

In this article, the authors present a new CI application framework for automatic observational networks and demonstrate it in the Heihe River Basin (HRB), a typical endorheic river basin in Northwest China. The new CI application framework facilitates scientific observational and data service related to the river basin and brings new insights into applications in other areas.

2. Description of the Proposed System

2.1. System Overview

The IODCS was designed to provide a universal data sharing platform for basin-scale scientific studies. The system is composed of the following functions: automated data reception and storage, automated data quality control, distributed data storage, data sharing services, model integration, and visualization. Figure 1 shows the framework of the IODCS, and the details are presented as follows.

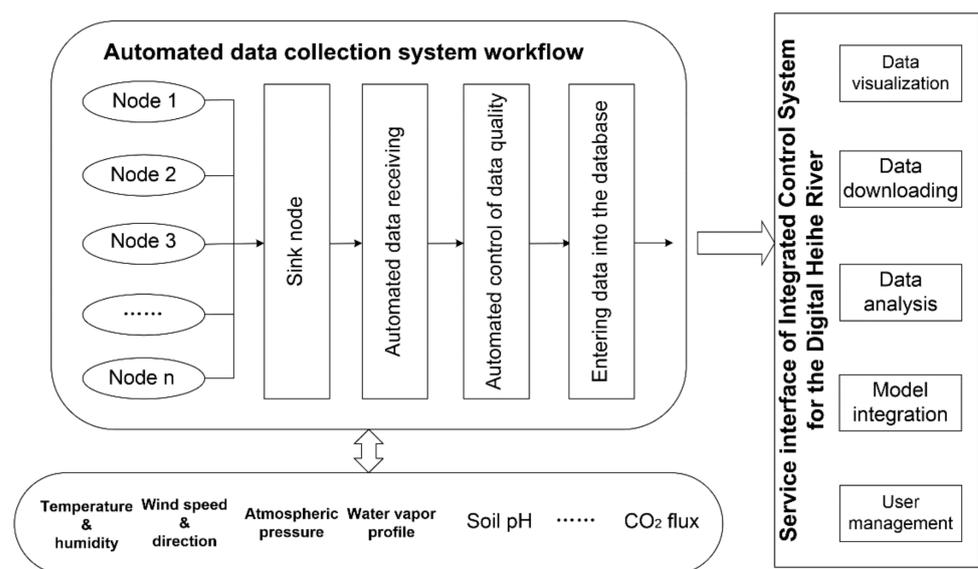


Figure 1. Framework of the IODCS.

2.2. Automated Data Reception and Storage

The automated data reception module aims to normalize different data organization modes of observational devices from different sources. Through the programming of the

server-side interface of the specific WSN observational device with the software development kit (SDK) [27], the data string transmitted to the server is directly intercepted, and the data are then reorganized as a defined standard data string. The normalized data string simplifies the back-end automated data preprocessing and warehousing procedure.

During the automated data reception process, the data status of the WSN on the nodes is automatically monitored, collected, and normalized in an unattended manner. The observational equipment regularly sends the observational data in the local cache to the data receiving interface program on the remote server. The server can also send instructions to request data uploading from the observational equipment. Figure 2 shows the process.

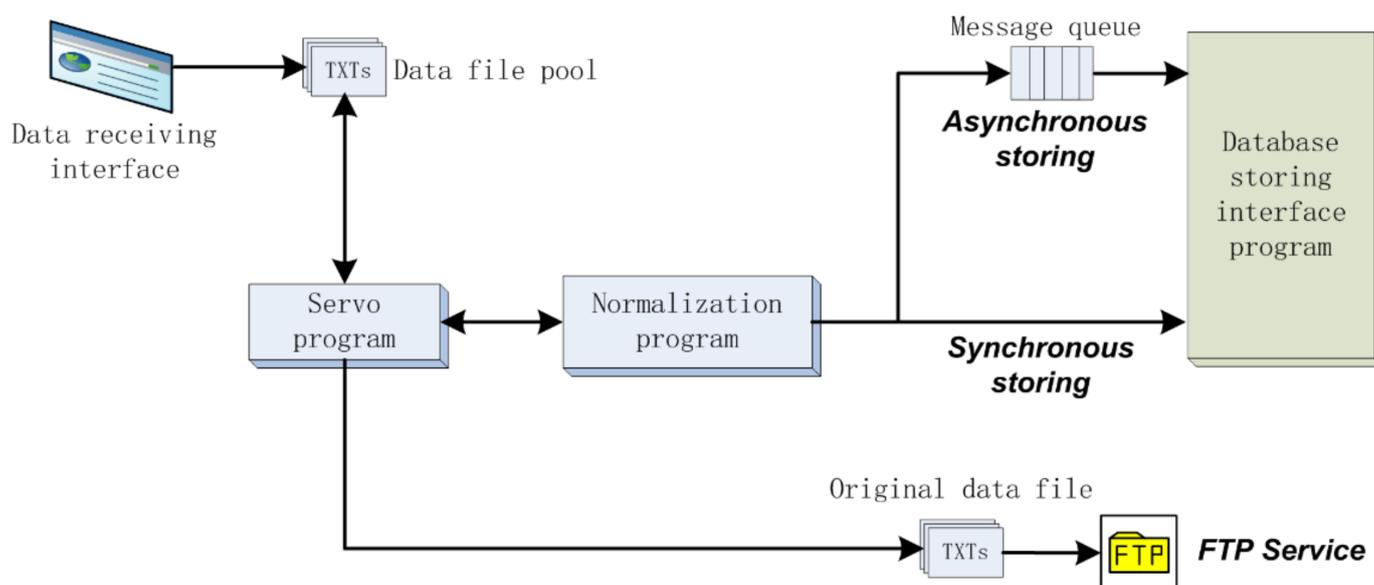


Figure 2. Process of the automatic receiving, normalizing, and storing of observational data in the IODCS.

The servo program automatically polls whether the data files in the data file pool have changed. When there are new data, the data reception process automatically starts, and the obtained data are pushed to a normalization program. After normalization, the disordered multisource data string becomes a simple and standard data string that is then sent to the database storage interface program. The data storage process has no human intervention, and it provides a unified database access interface for various data reception programs.

The automated storage process of observational data can be accomplished synchronously or asynchronously based on the actual needs of the system. In the synchronous mode, upon receiving the data, the normalization program invokes the storage program to immediately store the observational data in the database. In the asynchronous mode, the normalization program puts data into a message queue, and then the storage program obtains data from the message queue and stores them in the database when the system is not busy.

2.3. Automated Data Quality Control

For automated observational data, errors, such as source errors, calibration errors, and transmission errors, occur. Source errors are caused by the observational device itself, calibration errors occur during the calibration and correction of observational data, and transmission errors consist of the operational errors and interference errors incurred during data transmission [28–32]. The data quality control model is deployed in the IODCS to evaluate and control the quality of the observational data using unified standards and a consistent quality control system during the generation, processing, and transmission of data. Moreover, the transformation model deployed in the IODCS describes the conversion type factors during automated data processing, which contributes to the application of a

consistent conversion system in the automated process. Additionally, automated observational devices arranged in the field are produced by different manufacturers and use different parameters and data storage formats. Thus, these differences should be converted before data storage to guarantee that the data are stored with unified standards, namely semantics and format. The automated conversion algorithm and the quality evaluation algorithm are sequentially executed in the logical flow of data processing [33]. Figure 3 shows a flow chart of automated data quality control.

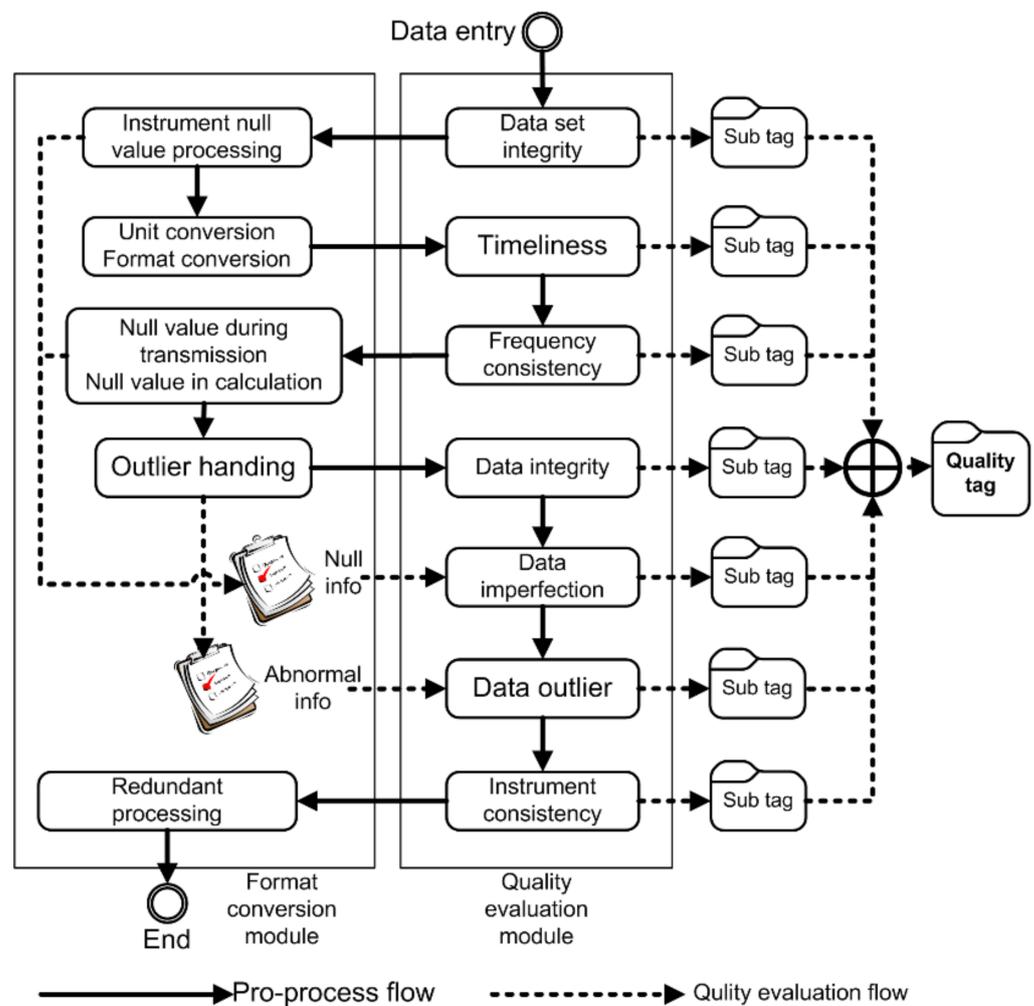


Figure 3. Data quality control items and flow in the IODCS.

There are two kinds of control logic processes in the data quality control flow: format conversion and quality evaluation. The format conversion process mainly deals with possible format problems in observational data. Quality evaluation evaluates the quality status of data and attaches the results as a label. These two processes include a variety of different methods or algorithms to process observational data. The details are shown in Table 2.

Table 2. Format conversion and quality evaluation methods used in the IODCS system.

Method	Type	Function	Impact on Data
Instrument null value	Format conversion	Detect null values caused by the instrument	Depending on the strategy, the data may be modified
Unit and format conversion	Format conversion	Detect and handle unit and format inconsistencies in observational data	The data will be modified
Null value during transmission and calculation	Format conversion	Detect and handle null values caused by other reasons	Depending on the strategy, the data may be modified
Outlier	Format conversion	Detect and handle data that do not adhere to data trends	Depending on the strategy, the data may be modified
Redundant processing	Format conversion	Detect and delete duplicate data based on timestamps	The duplicate data will be deleted
Dataset integrity	Quality evaluation	Check whether all the variables of integrated observations have observational values. For example, is there a missing value in a temperature profile?	The data will not be modified but will be tagged
Timeliness	Quality evaluation	Check the timeliness of warehousing data	No data will be modified, but the system will tag the variable
Frequency consistency	Quality evaluation	Check whether the data are collected according to the acquisition frequency	No data will be modified, but the system will tag the variable
Data integrity	Quality evaluation	Tag the data according to the result of outlier detection	No data will be modified, but the system will tag the variable
Data imperfection	Quality evaluation	Tag the data according to the result of null value detection	No data will be modified, but the system will tag the variable
Data outlier	Quality evaluation	Tag the abnormal data according to the result of outlier detection	No data will be modified, but the system will tag the variable
Instrument consistency	Quality evaluation	Detect abnormal data caused by instrument	No data will be modified, but the system will tag the variable

2.4. Distributed Storage System

In the past, observational data were mainly stored in the form of text or spreadsheets. Since the data files from different sources contained different contents and were in different formats, the efficiency of data management, analysis, and use was rather low. As the automated observational data stored and managed in an observational database are strongly structured, the relational database was selected as the basis of storage design in this study, and the Greenplum distributed database [34,35] was used for data storage. After fully investigating and analyzing the observational projects and observational elements that had been carried out or were planned, we designed and constructed the IODCS with an object-oriented relational database design method. The new data storage and management mode significantly improved efficiency. Figure 4 shows the structure of the relational database designed in this study.

As shown in Figure 5, we built a high-performance computer cluster with one master node, multiple segment nodes, and a Greenplum distributed observational database. The number of segment nodes can be dynamically expanded according to actual needs. The master node is responsible for organizing and dispatching the cluster operation and connecting with the external network. All data are stored on segment nodes. Each segment

node can contain multiple segments. A segment is the basic unit for performing database-concurrent operations. The number of segments per segment node may be different according to user requirements and server hardware performance. This kind of cluster application mode has excellent advantages in data query efficiency, especially when the quantity of query data is large [36].

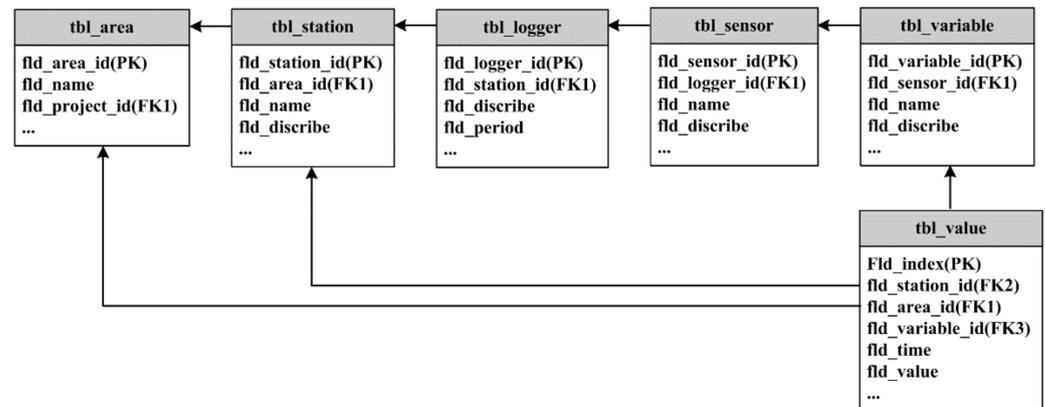


Figure 4. Conceptual structure of the relational observational database in the IODCS.

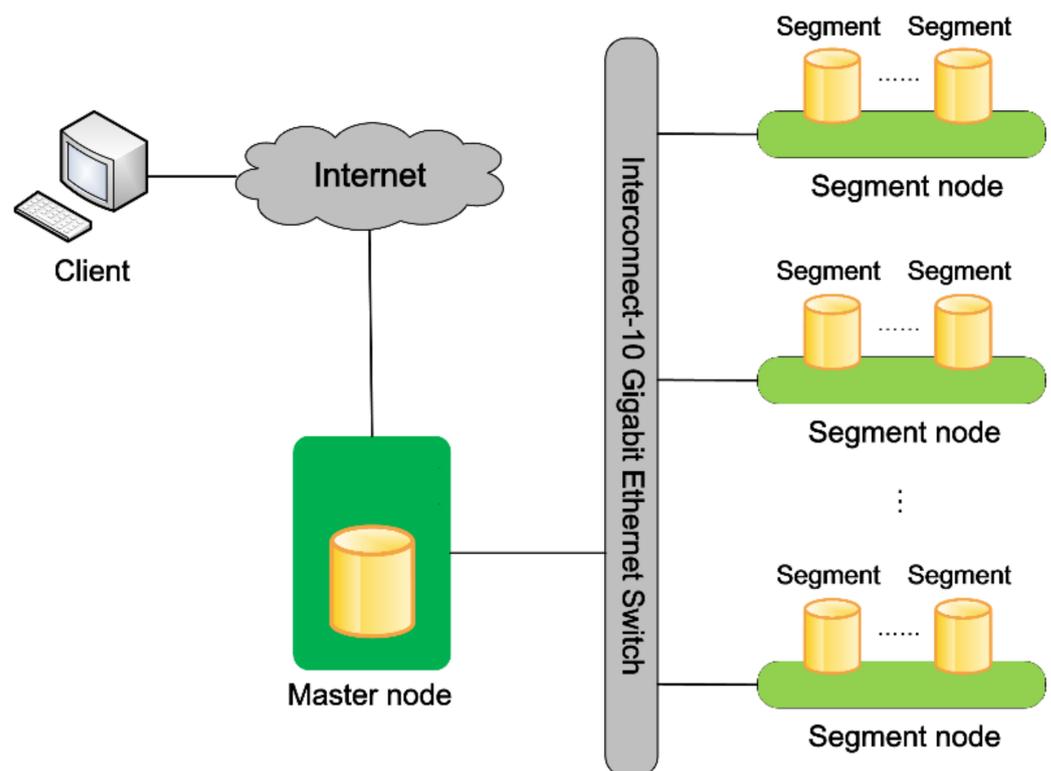


Figure 5. A high-performance computer cluster for the distributed observational database.

2.5. Model Integration

An essential function of IODCS is to integrate the online scientific model. Establishing a scientific model for the HRB and its surrounding areas is an indispensable approach to basin-scale scientific research. Researchers have developed different ecological and hydrological models for the Heihe River Basin. To efficiently combine these models with real-time observational data and to more quickly serve basin research, we developed an online model integration module in the IODCS. The models are integrated with the model integration module through a web service interface. As shown in Figure 6, the relevant

models are called by the IODCS through the online integration module, and the input parameters, output results, real-time observational data, and basic spatial data support of the model are controlled by the IODCS.

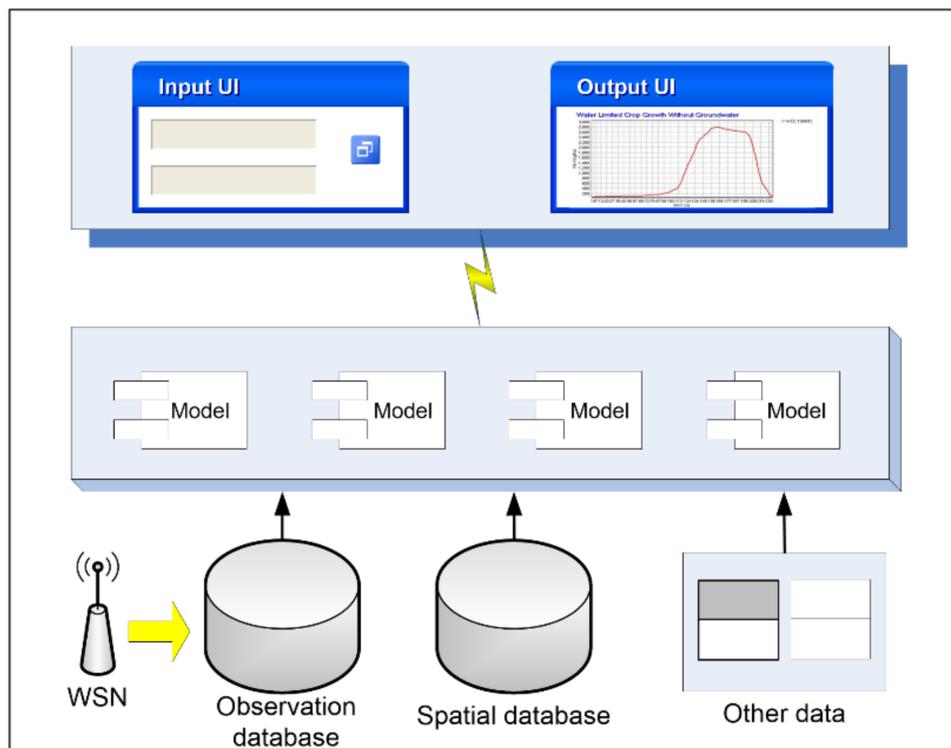


Figure 6. The logic architecture of model integration in the IODCS.

By inputting the needed parameters of the integrated model via the visual interface of the IODCS, users can obtain the real-time predicted results online.

2.6. Visualizations

The IODCS has many functions for observational data, including automated reception and storage, automated format conversion and quality evaluation, statistical analysis, and classification according to standard data description and relevant rules. All these functions are intended to provide convenience for users, and every bit of the system's information is vital. Thus, we developed a visualization module to intuitively show this information.

The data visualization module was developed using the OpenLayers plug-in and Apache ECharts control [37–39]. Figure 7 shows the functional structure of the visualization system. The map visualization function provides visualization support for the underlying surface, regional boundary, and distribution of the observational stations. The data visualization function provides visualization support for data trends, data comparison, data analysis, and data sharing paths. With the help of the visualization module, we can easily check the location and configuration of the observational stations and the type and data trends of the observational variables online. Moreover, the observational data can be intuitively analyzed online in real time.

The IODCS optimizes the technical process of data visualization to avoid the performance loss caused by excessively frequent connection and visualization requests between the client and server. Based on WebSocket [40], the IODCS establishes a continuous long-term connection between the client browser and the server, which not only reduces the server load but also constantly provides the client with the latest data and their trends through data visualization. In other words, once the server receives the new field data, the client immediately updates the data.

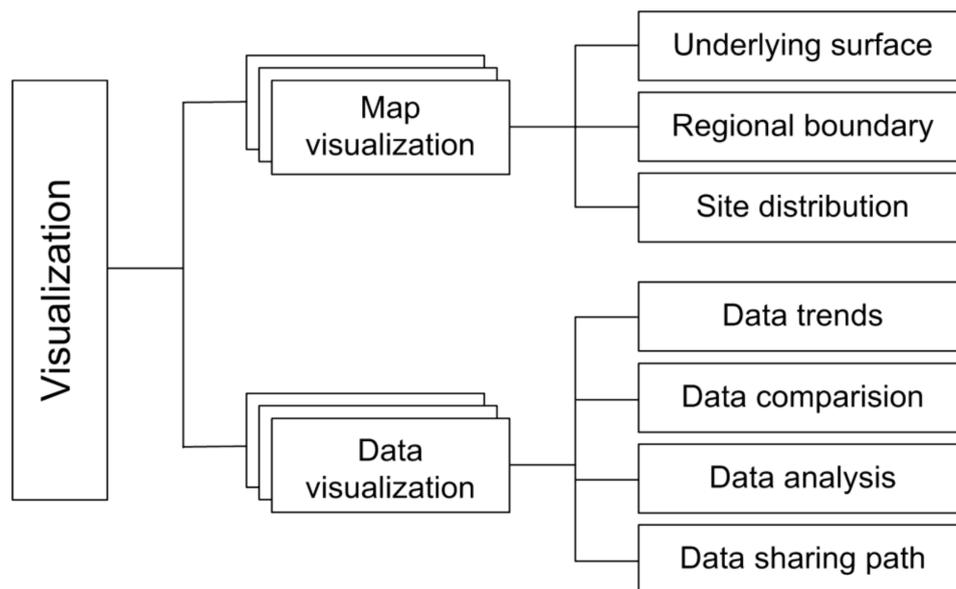


Figure 7. Data visualization function system in the IODCS.

3. Application in the Heihe River Basin

3.1. Case Study Area and the Overall Implementation

The Heihe Watershed Allied Telemetry Experimental Research (HiWATER) is a large-scale comprehensive observational experiment [41–44] in which midstream observations were launched in 2012. To effectively capture the spatial heterogeneity of the surface elements in the river basin and to verify the authenticity of remote sensing data, HiWATER was carried out in a 5.5×5.5 km core observational area located in the Yingke/Daman Irrigation District in the middle reaches of the Heihe River. The spatial optimization algorithm selected a total of 198 observational stations. With the WSN as the bridge, a great variety of meteorological, hydrological, and ecological observational projects with multisource sensors densely distributed on the scale of the river basin/irrigation district were integrated to establish a fully automated observational system with spatiotemporally coordinated ecohydrological sensors. Figure 8 shows the layout of the WSN observational nodes in the middle reaches of the HRB.

The core observational area was equipped with four types of surface variable observational devices with remote wireless data transmission, namely SoilNET for observing the soil moisture/temperature, WATERNET for observing the soil moisture/temperature and surface temperature, LAINET for observing the leaf area index, and AWS (automatic weather stations), as shown in Table 3. The nodes of BNUNET (designed by Beijing Normal University for observing the soil moisture/temperature), LAS (large aperture scintillometer), and EC (eddy covariance system) have no remote wireless data transmission functionality.

The IODCS automatically collected more than 300 observational variables (observational variables at different heights or depths were regarded as different observational variables), with a 10 min sampling period for each type of observational variable. During the observational period of the synergetic enhancement between the Earth and the satellite, the sampling period was 1 min, and a small number of other observational variables were sampled every 30 min [45,46]. In this study, the designed fully automated observations and IODCS comprehensively improved the overall observational ability, information level, and observational data sharing service for the ecohydrological processes in the river basin.

We have demonstrated that our system developed for HiWATER and its online observational data support platform can provide observational managers and researchers with online data services, including the visualization of two-dimensional, three-dimensional, or

3.2. Data Management and Service

Figure 9 shows the data service interface for data management and release. To date, the IODCS of the HRB has automatically received, processed, and stored more than 1.4 billion observational data records, with an average of over 14 million observational data records per month. In August 2012, the month in which the most intensive observations were carried out, the number of stored observational data records reached 340 million. On the premise of efficiently completing the automated preprocessing and distributed storage of observational data, the IODCS provides observers and researchers with flexible and convenient practical functions such as online data visualization, online data querying, on-demand data downloading, alarms for anomalies in observational devices, and the automated generation of daily equipment inspection reports. The function of the daily equipment inspection reports is to automatically sort the observational data generated by all the observational equipment in the HRB observational network and generate reports every day, which can help users intuitively understand the operational status of the observational network and the quality of the observational data.



Figure 9. Data management and service interface of the IODCS of the HRB (translated from the user interface in Chinese).

3.3. Real-Time Online Data Browsing and Analysis

As shown in Figure 10, during the observational period in the HRB, the observational data are collected every 10 min (1 min in intensive mode), and the related data curves are also automatically updated every 10 min (1 min in intensive mode). Users can arbitrarily call and display the data of any observational elements at any observational station during any observational period and browse the data in various online visualization formats, e.g., curve charts, area charts, scatter charts, or column charts. Moreover, the system supports the multivariable comparative analysis of observational data. Users can combine multiple correlating variables together for the browsing of visual contrast.

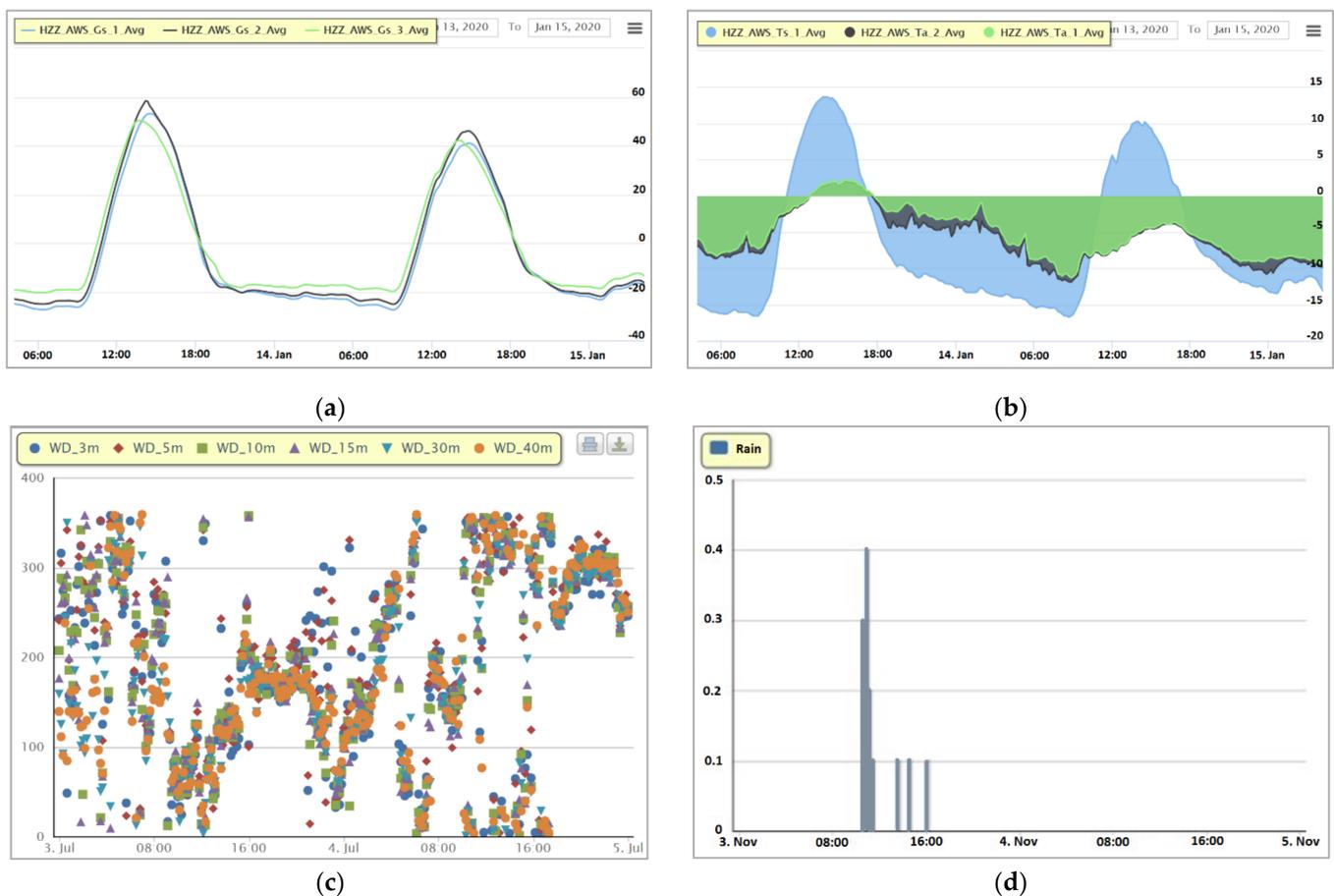


Figure 10. Different online visualization forms of the observational data in the HRB. (a) A curve chart example of soil heat flux (in this example, in a corn field near Daman station); (b) an area chart example of soil temperature (in this example, 2, 4, and 6 cm underground near Daman station); (c) a scatter chart example of wind direction (in this example, near Dashedalong station on 4 November 2019); (d) a column chart example of rainfall (in this example, near Dashedalong station on 4 November 2019).

3.4. Data Downloading on Demand

The application of relational databases facilitates the on-demand downloading of observational data. In the past, observational data in the HRB were shared through data files. Each file stored some observational elements obtained in specific areas within a given period in a certain format. After gaining the shared data files, researchers extract and sort the data that they need from many data files based on their research demands. This approach is very inefficient and time consuming. To solve this problem, the authors of this study developed the IODCS in the HRB so that researchers may directly obtain the observational elements collected from a specific area within the required time interval from the system according to their research needs while keeping data acquisition process simple and efficient. Figure 11 shows the data downloading interface.

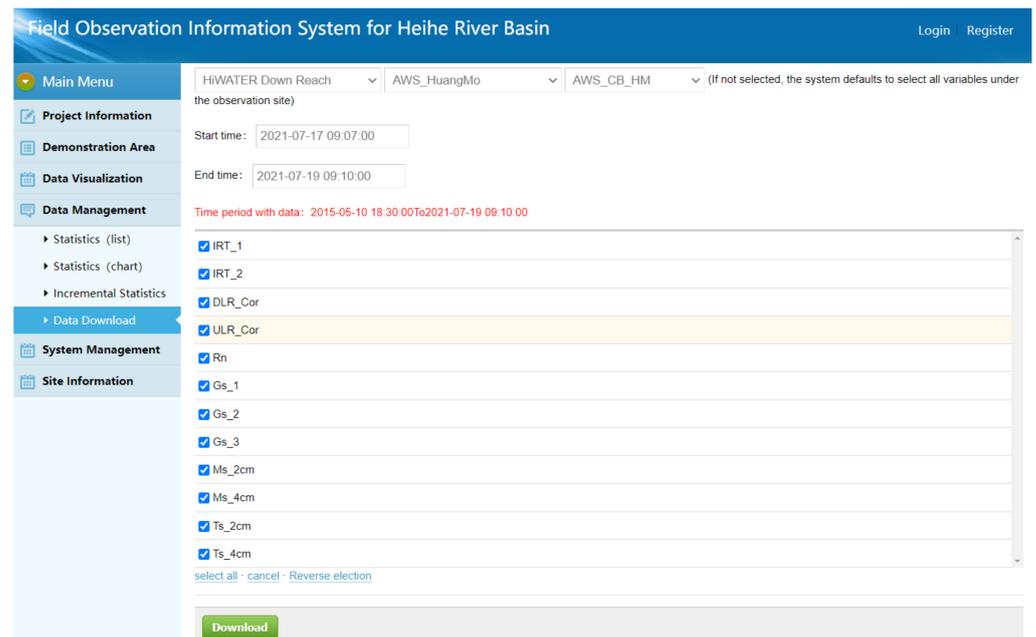


Figure 11. Data downloading interface (translated from the user interface in Chinese).

3.5. Intelligent Analysis of the Status of the Observational Network

Because there are many observational nodes and elements in the observational area, it is very important to know the status of the observational network in real time. The IODCS can automatically collect and organize the observational data of all nodes in the whole observational network every day and generate a graphical status report of the observational network. This report can help maintenance personnel monitor the operational status of the observational system and analyze whether there are errors in the data. Figure 12 shows part of the equipment status inspection reports.

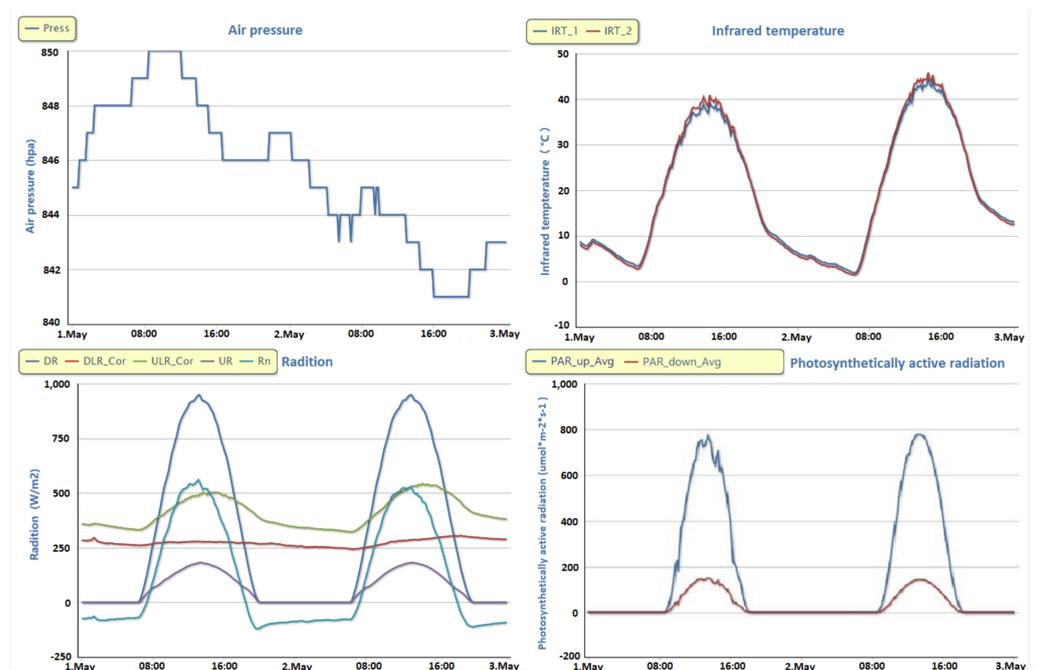


Figure 12. Part of the equipment status inspection report of the IODCS.

3.6. Anomaly Detection

The IODCS applies some data analysis methods to assist in the analysis and detection of observational data. For example, a fitting method is used to detect abrupt anomalies in the data. An artificial neural network analyzes the abnormal fluctuation of the data, and the integrity of the dataset is analyzed using a statistical method.

Figure 13 shows the detection of data exceptions using the polynomial curve fitting method (using the same strategy as MATLAB; see <https://www.mathworks.com/help/matlab/ref/polyfit.html> (accessed on 28 June 2021)). Considering the WSN data for soil moisture within a certain period as an example [47], according to the time-series scatter diagram and the fitted results of the sample data, the observed soil moisture value sharply increased at approximately 01:20:00 on 26 June 2012. On that day, due to the change (rain) in the external environment, the fitted data revealed the fluctuation well.

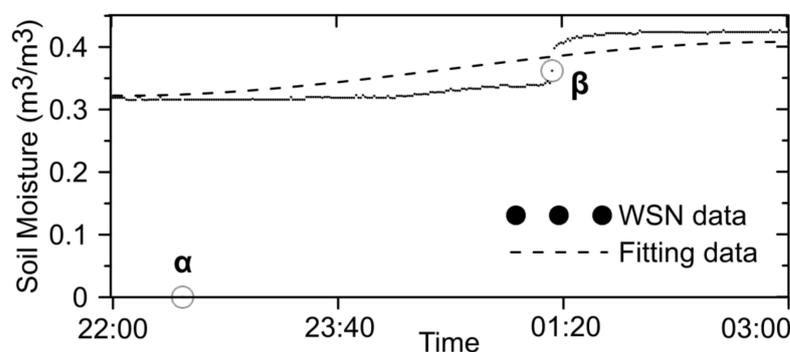


Figure 13. Detection of data exceptions using the fitting method.

The data at point β might be misjudged as abnormal data if there were accidental errors but can be correctly identified among the quality control elements with abnormal fluctuation. However, with a residual error of 0.3135, which was much greater than three standard deviations (standard deviation = 0.0278), the data at point α were abnormal.

Figure 14 shows the analysis of data exceptions using the backpropagation (BP) neural network method. Taking the measured WSN data for soil moisture on a specific day as an example, we analyzed the fluctuation in the observational data. Figure 14 illustrates that as the number of training iterations increased, the model's accuracy was greatly improved. Then, we used the true data range from 18:36 to 20:24 as the samples by which the BP algorithm was applied to train the network, and the corresponding weight matrixes and their eigenvectors were calculated. By comparing these eigenvectors with the standard eigenvector, e.g., the eigenvector of the data collected during a rain spell, researchers could determine whether data are abnormal.

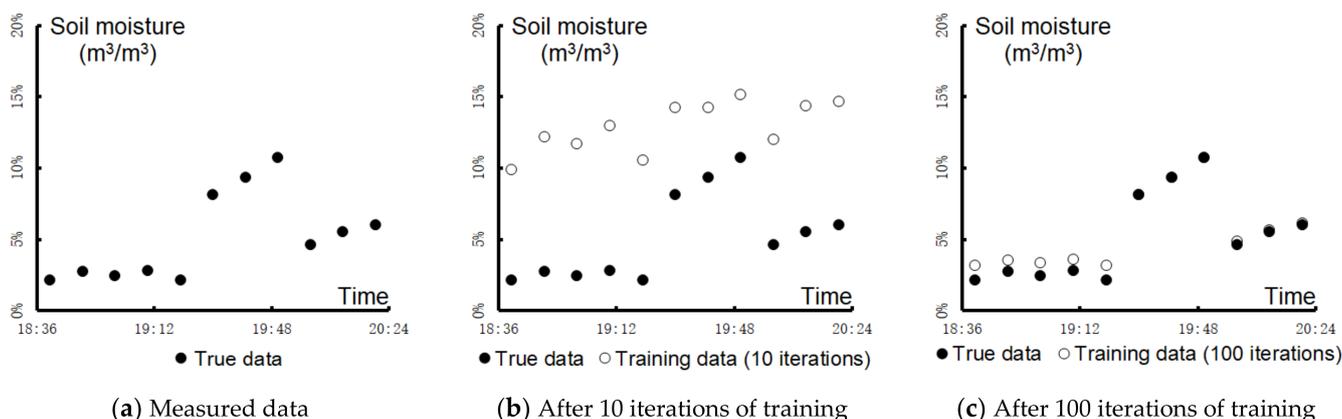


Figure 14. Analysis of data exceptions using the BP neural network method.

3.7. Online Computing with Integrated Models

We used online calculations for growth monitoring and yield prediction [48,49] and soil moisture spatial interpolation of crops as examples of these integrated models [50,51]. Table 4 shows the details of the integrated models.

Table 4. Models integrated into the IODCS.

No.	Model	Function	Development Language
1	WOFOST crop growth model	With the physiological and ecological processes of crops, e.g., assimilation, respiration, transpiration, and dry matter partitioning, as the simulation basis, the WOFOST crop growth model simulates the growth of crops under the circumstances of potential growth, restricted water, and limited nutrients.	Fortran [52]
2	Spatial kriging interpolation model	After examining soil moisture in the typical irrigated farmland in the upper reaches of the Heihe River as the object of study, relevant extension packs of the Python language are applied to analyze the spatial variability in the observational data and build the spatial kriging interpolation model to estimate the soil moisture in the study area.	Python

The predicted results of crop growth, yield, and biomass are shown in Figure 15. The predicted results of the soil moisture spatial interpolation model are shown in Figure 16.

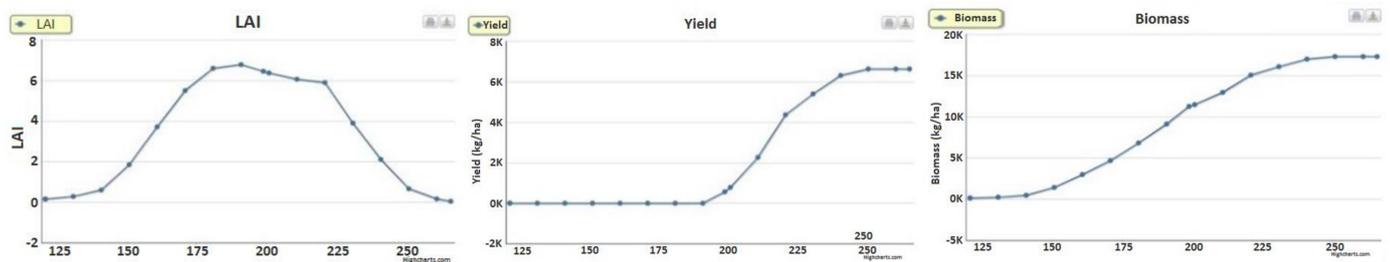


Figure 15. Prediction results of the crop growth model.

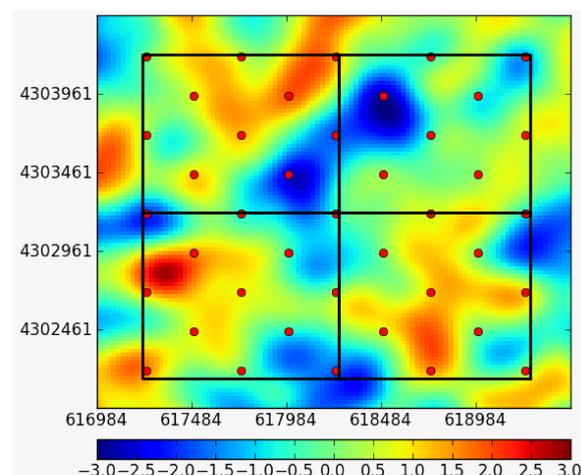


Figure 16. Online calculation results of the soil moisture spatial interpolation model (using discrete soil moisture observational data as the input, the platform calls the online spatial interpolation model to calculate the spatial distribution data for soil moisture).

4. Summary and Outlook

The IODCS developed in this paper is a highly standardized, strongly interactive, secure, and reliable instance of CI application. The HRB area application shows that the IODCS has completely changed the method of storing and managing observational data using data files by applying a relational database to manage strongly structured observational data, which greatly increases data management efficiency. The change in the data management mode also directly influences the subsequent use of observational data, not only making it possible to efficiently query, analyze, and download the observational data but also remarkably improving the method and efficiency of data service and sharing. The IODCS facilitates other application modes for observational data, such as online model integration application and the intelligent analysis of the status of the observational network.

However, due to the current level of automated data processing, only a limited number of automated preprocessing algorithms are used in the IODCS, which, to a certain degree, confines the processing capacity of the CI system. To meet the demands of scientific field observations in the big data era, the IODCS must be further optimized to combine big data with machine learning and deep learning to provide a one-stop platform for data management and model development for massive observational data to support big Earth data sharing services for scientific research in the HRB.

Author Contributions: Writing—original draft, J.G.; methodology, J.G., Q.S., F.L., A.W. and M.Z.; writing, reviewing, and editing, X.L. and M.Z. conceptualization, J.G. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (2016YFC0500105) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060602).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. This data can be found here: <http://210.77.68.221:10066/iframe>.

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

References

1. Toffler, A. *The Third Wave*, 1st ed.; Bantam Books, Inc.: New York, NY, USA, 1980; pp. 199–236.
2. Bertot, J.C.; Choi, H. Big data and e-government: Issues, policies, and recommendations. In Proceedings of the 14th Annual International Conference on Digital Government Research, Quebec, QC, Canada, 17–20 June 2013; pp. 1–10. [\[CrossRef\]](#)
3. Simović, A. A Big Data smart library recommender system for an educational institution. *Libr. Hi Tech* **2018**, *36*, 498–523. [\[CrossRef\]](#)
4. Guo, H.; Wang, L.; Liang, D. Big Earth Data from space: A new engine for Earth science. *Sci. Bull.* **2016**, *61*, 505–513. [\[CrossRef\]](#)
5. Baumann, P.; Mazzetti, P.; Ungar, J.; Barbera, R.; Barboni, D.; Beccati, A.; Bigagli, L.; Boldrini, E.; Bruno, R.; Calanducci, A.; et al. Big data analytics for earth sciences: The EarthServer approach. *Int. J. Digit. Earth* **2016**, *9*, 3–29. [\[CrossRef\]](#)
6. Li, D.R. Towards Geo-spatial Information Science in Big Data Era. *Acta Geod. Cartogr. Sin.* **2016**, *45*, 379–384. (In Chinese)
7. Jones, M.O.; Jones, L.A.; Kimball, J.S.; McDonald, K.C. Satellite passive microwave remote sensing for monitoring global land surface phenology. *Remote Sens. Environ.* **2011**, *115*, 1102–1114. [\[CrossRef\]](#)
8. Li, X. Characterization, controlling and reduction of uncertainties in the modeling and observation of land-surface systems. *China Earth Sci.* **2014**, *57*, 80–87. [\[CrossRef\]](#)
9. El-Zeiny, A.M.; Effat, H.A. Environmental monitoring of spatiotemporal change in land use/land cover and its impact on land surface temperature in El-Fayoum governorate, Egypt. *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 266–277. [\[CrossRef\]](#)
10. Cheng, G.S.; Li, X.; Zhao, W.; Xu, Z.; Feng, Q.; Xiao, S.; Xiao, H. Integrated study of the water-ecosystem-economy in the Heihe River Basin. *Natl. Sci. Rev.* **2014**, *1*, 413–428. [\[CrossRef\]](#)
11. Li, X.; Zhao, N.; Jin, R.; Liu, S.; Sun, X.; Wen, X.; Wu, D.; Zhou, Y.; Guo, J.; Chen, S.; et al. Internet of Things to network smart devices for ecosystem monitoring. *Sci. Bull.* **2019**, *64*, 1234–1245. [\[CrossRef\]](#)
12. Zhang, M.H.; Li, X. Drone-enabled Internet-of-Things relay for environmental monitoring in remote areas without public networks. *IEEE Internet Things J.* **2020**, *7*, 7648–7662. [\[CrossRef\]](#)

13. Hart, J.K.; Martinez, K. Toward an environmental Internet of Things. *Earth Space Sci.* **2015**, *2*, 194–200. [[CrossRef](#)]
14. Lvovich, I.Y.; Lvovich, Y.E.; Preobrazhenskiy, A.P.; Preobrazhenskiy, Y.P.; Choporov, O. Modeling of information processing in the internet of things at agricultural enterprises. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2019; Volume 315, p. 032029. [[CrossRef](#)]
15. Guo, J.W.; Shang, Q.S.; Chang, H.L.; Liu, F.; Li, J.; Wu, A. Design of Field Observation Data Automatic Assembling System. *Remote Sens. Technol. Appl.* **2013**, *28*, 399–404. (In Chinese)
16. Wu, A.D.; Guo, J.W.; Wang, L.X. Improvement and Application of automatic data in Heihe river basin downloading system. *Remote Sens. Technol. Appl.* **2015**, *30*, 1027–1032. (In Chinese)
17. Wang, H.W.; Zhang, W.G.; Yu, X.W.; Zhang, X.; Deng, G.; Liu, Y.; Wang, J.; Li, F. Design and Operation of Network Management Platform for Forest Ecological Positioning Observation System. *World For. Res.* **2018**, *31*, 28–33. (In Chinese)
18. Khayyat, Z.; Ilyas, I.F.; Jindal, A.; Madden, S.; Ouzzani, M.; Papotti, P.; Quiané-Ruiz, J.-A.; Tang, N.; Yin, S. Bigdancing: A system for big data cleansing. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 31 May 2015–4 June 2015; Association for Computing Machinery: New York, NY, USA, 2015.
19. Štefanič, P.; Cigale, M.; Jones, A.C.; Knight, L.; Taylor, I.; Istrate, C.; Suci, G.; Ulisses, A.; Stankovski, V.; Taherizadeh, S.; et al. SWITCH workbench: A novel approach for the development and deployment of time-critical microservice-based cloud-native applications. *Future Gener. Comput. Syst. Int. J. Escience* **2019**, *99*, 197–212. [[CrossRef](#)]
20. Koulouzis, S.; Martin, P.; Zhou, H.; Hu, Y.; Wang, J.; Carval, T.; Grenier, B.; Heikkinen, J.; De Laat, C.; Zhao, Z. Time-critical data management in clouds: Challenges and a Dynamic Real-Time Infrastructure Planner (DRIP) solution. *Concurr. Comput. Pract. Exp.* **2020**, *32*. [[CrossRef](#)]
21. Liu, X.; Song, H.; Liu, A. Intelligent UAVs Trajectory Optimization from Space-Time for Data Collection in Social Networks. *IEEE Trans. Netw. Sci. Eng.* **2020**, *8*, 853–864. [[CrossRef](#)]
22. Huang, S.; Liu, A.; Zhang, S.; Wang, T.; Xiong, N. BD-VTE: A Novel Baseline Data based Verifiable Trust Evaluation Scheme for Smart Network Systems. *IEEE Trans. Netw. Sci. Eng.* **2020**. [[CrossRef](#)]
23. Ren, Y.; Wang, T.; Zhang, S.; Zhang, J. An intelligent big data collection technology based on micro mobile data centers for crowdsensing vehicular sensor network. *Pers. Ubiquitous Comput.* **2020**. [[CrossRef](#)]
24. Aftab, M.U.; Oluwasanmi, A.; Alharbi, A.; Sohaib, O.; Nie, X.; Qin, Z.; Ngo, S.T. Secure and dynamic access control for the Internet of Things (IoT) based traffic system. *Peerj Comput. Sci.* **2021**. [[CrossRef](#)]
25. Díaz, J.J.; Mura, I.; Franco, J.F.; Akhavan-Tabatabaei, R. aiRe-A web-based R application for simple, accessible and repeatable analysis of urban air quality data. *Environ. Model. Softw.* **2021**, *138*. [[CrossRef](#)]
26. Gorton, I. Cyberinfrastructures: Bridging the Divide between Scientific Research and Software Engineering. *Computer* **2014**, *47*, 48–55. [[CrossRef](#)]
27. Wang, L.; Wang, S.; Ran, Y. Data sharing and data set application of watershed allied telemetry experimental research. *IEEE Geosci. Remote. Sens. Lett.* **2014**, *11*, 2020–2024. [[CrossRef](#)]
28. Shih, Y.T.; Cheng, H.M.; Sung, S.H.; Hu, W.C.; Chen, C.H. Quantification of the calibration error in the transfer function-derived central aortic blood pressures. *Am. J. Hypertens.* **2011**, *24*, 1312–1317. [[CrossRef](#)] [[PubMed](#)]
29. Zhou, B.; Chen, Q.; Xiao, P. The error propagation analysis of the received signal strength-based simultaneous localization and tracking in wireless sensor networks. *IEEE Trans. Inf. Theory* **2017**, *63*, 3983–4007. [[CrossRef](#)]
30. Fei, H.; Xiao, F.; Li, G.H.; Sun, L.J. An Anomaly Detection Method of Wireless Sensor Network Based on Multi-Modals Data Stream. *Chin. J. Comput.* **2017**, *40*, 1829–1842. (In Chinese)
31. Zhang, M.H.; Li, X.; Wang, L.L. An adaptive outlier detection and processing approach towards time series sensor data. *IEEE Access* **2019**, *7*, 175192–175212. [[CrossRef](#)]
32. Zhang, M.H.; Guo, J.W.; Li, X.; Jin, R. Data-driven anomaly detection approach for time-series streaming data. *Sensors* **2020**, *20*, 5646. [[CrossRef](#)]
33. Guo, J.W.; Liu, F. Automatic data quality control of observations in wireless sensor network. *IEEE Geosci. Remote. Sens. Lett.* **2014**, *12*, 716–720.
34. Schwichtenberg, H. Installing Entity Framework Core. In *Modern Data Access with Entity Framework Core*; Apress: Berkeley, CA, USA, 2018; pp. 15–29.
35. Albertini, O.R.; Bhargov, D.; Denissov, A.; Guerrero, F.; Jayaram, N.; Kak, N.; Khanna, E.; Kislal, O.; Kumar, A.; McQuillan, F.; et al. Image classification in Greenplum database using deep learning. In Proceedings of the International Conference on Management of Data, Portland, OR, USA, 14–19 June 2020; ACM SIGMOD Record. pp. 1–4.
36. Wu, A.D.; Guo, J.W.; Yang, P.F. Research on an Application of Shared Architecture for Ecological Monitoring-oriented IoT Streaming Data. *IEEE Access* **2020**, *8*, 195385–195397. [[CrossRef](#)]
37. Wu, A.D.; Che, T. Application Research of 3D Visualization System for Three Poles Scientific Discovery. *J. Glaciol. Geocryol.* **2021**, *43*, 1–11. (In Chinese)
38. OpenLayers API Docs. Available online: <https://openlayers.org/en/latest/apidoc/> (accessed on 26 March 2021).
39. ECharts Docs. Available online: <https://echarts.apache.org/en/api.html#echarts> (accessed on 26 March 2021).
40. WebSocket. Available online: <https://developer.mozilla.org/en-US/docs/Web/API/WebSocket> (accessed on 26 March 2021).

41. Li, X.; Cheng, G.; Liu, S.; Xiao, Q.; Ma, M.; Jin, R.; Che, T.; Liu, Q.; Wang, W.; Qi, Y.; et al. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific objectives and experimental design. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1145–1160. [[CrossRef](#)]
42. Li, X.; Liu, S.; Xiao, Q.; Ma, M.; Jin, R.; Che, T.; Wang, W.; Hu, X.; Xu, Z.; Wen, J.; et al. A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. *Sci. Data* **2017**, 170083. [[CrossRef](#)]
43. Jin, R.; Li, X.; Yan, B.; Luo, W.; Li, X.; Guo, J. Introduction of eco-hydrological wireless sensor network in the Heihe River Basin. *Adv. Earth Sci.* **2012**, *27*, 993–1005.
44. Liu, S.; Li, X.; Xu, Z.; Che, T.; Xiao, Q.; Ma, M.; Liu, Q.; Jin, R.; Guo, J.; Wang, L.; et al. The Heihe integrated observatory network: A basin-scale land surface processes observatory in China. *Vadose Zone J.* **2018**, *17*, 180072. [[CrossRef](#)]
45. Xu, Z.; Liu, S.; Li, X.; Shi, S.; Wang, J.; Zhu, Z.; Xu, T.; Wang, W.; Ma, M. Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. *J. Geophys. Res. Atmos.* **2013**, *118*, 13140–13157. [[CrossRef](#)]
46. Jin, R.; Li, X.; Yan, B.; Li, X.; Luo, W.; Ma, M.; Guo, J.; Kang, J.; Zhu, Z.; Zhao, S. A nested eco-hydrological wireless sensor network for capturing the surface heterogeneity in the midstream area of the Heihe River Basin, China. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2015–2019. [[CrossRef](#)]
47. Liu, F.; Guo, J.W. Study on quality control approach for Heihe wireless sensor network observation data. *Remote Sens. Technol. Appl.* **2013**, *28*, 252–257. (In Chinese)
48. Wang, J.; Li, X.; Lu, L.; Fang, F. Parameter sensitivity analysis of crop growth models based on the Extended Fourier Amplitude Sensitivity Test method. *Environ. Model. Softw.* **2013**, *48*, 171–182. [[CrossRef](#)]
49. Wang, J.; Li, X.; Lu, L.; Fang, F. Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *Eur. J. Agron.* **2013**, *49*, 126–140. [[CrossRef](#)]
50. Jian, K.; Rui, J.; Shaojie, Z.; Linna, C. Spatial Sampling Design of the Sensor Network for Monitoring the Surface Freeze / thaw Cycles over the Heterogeneous Surface in the Heihe River Basin. *Remote Sens. Technol. Appl.* **2014**, *29*, 833–838. (In Chinese)
51. Jin, R.; Li, X.; Liu, S.M. Understanding the heterogeneity of soil moisture and evapotranspiration using multiscale observations from satellites, airborne sensors, and a ground-based observation matrix. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 2132–2136. [[CrossRef](#)]
52. Fortran 90. Available online: <https://www.fortran90.org/> (accessed on 26 March 2021).