WALAILAK JOURNAL
http://wjst.wu.ac.th Engineering and Physical Sciences

An Integrated Event Detection and Decision Support System for Managing the Health of Ocean and Climatic Sensor

Premrudee NOONSANG, Krisanadej JAROENSUTASINEE*, Mullica JAROENSUTASINEE and Peeravit KOAD

Center of Excellence for Ecoinformatics, School of Science, Walailak University, Thasala, Nakhon Si Thammarat 80161, Thailand

(*Corresponding author's e-mail: krisanadej@gmail.com)

Received: 16 July 2015, Revised: 25 August 2015, Accepted: 28 September 2015

Abstract

Real-time online coral sensor and climatic data are vital for long term ecological observatory research. It is difficult to maintain a continuous real-time online data stream using various underwater sensors feeding to a server. To solve this difficulty, we integrated an event detection and decision support system (DSS) to help researchers to make decisions about sensor maintenance and to proactively prevent system malfunctions. This paper describes the components of pre-processing data in-situ, using behavior learning from historical data, creating anomaly detection, and verifying sensor data. This online system is flexible, for both health monitoring and decision-making. When data loss or sensor malfunctions occur, then an automatic email will be sent to researchers, to show both sensor health and anomaly reports. Researchers will then make a decision on who will fix the problems, and plan for maintenance on site. The integrated decision support system and event detection uses various tools and techniques, such as *Mathematica* for analysis and open source software (PHP, MySQL, and Highcharts) for visualization. This is the first practical system that uses heterogeneous sensors integrated with flexible cloud storage, which shows both alerts regarding system health and anomaly reports through social media, and helps researchers in decision-making, in order to reduce maintenance time and costs, and to enable long term use of sensor data in order to expand understanding on climate change.

Keywords: Decision Support System, event detection, system health, coral sensor data, ecoinformatics

Introduction

The advanced technology of environmental sensors allows ecoinformatics researchers to get real-time weather and water quality data [1]. Maintenance is the most expensive task in a product's life cycle [2], and it is difficult to develop technologies and technical solutions to reduce maintenance costs. There are various reasons to monitor the health of a system, including (1) verifying performing tasks [3], (2) detecting anomalous data [4,5], (3) detecting system faults [3], (4) investigating battery health [6], (5) predicting future faults [3,7], and (6) warning of any upcoming events before causing any severe damage [8]. To reduce the maintenance costs for monitoring real-time systems that send data to a server, we need to develop a surveillance system [9]. System health monitoring is a form of system diagnosis that aims to detect system failure or malfunction and relaying the identifying results to the administrators or technical users who are in charge of the task [9]. The functions of system health monitoring consist of diagnosing malfunctioned sensors and sending alerts to researchers.

Event detection plays a precious role in mitigating the effects of natural disasters [10], and in preventive maintenance. The main goals of such systems are similar to reduce economic losses, to mitigate the number of deaths from disaster, and to prepare for emerging disaster [10]. The systems rely on gathering data from sensors, storage, infrastructure, and applications for performing various analyses of data such as simulation-based prediction [10].

Decision support system (DSS) is an important tool that retrieves all relevant information and visualizes the integrated information in order to support decision-making. DSS can be automatic computerized, semi human-powered, or a combination of both. For environmental researches, DSS have been implemented for flood risk management [11], and have supported the activities of emergency agencies and government departments.

The development of environmental sensor networks requires a unique combination of technological and environmental understanding [12-14]. Different types of data are collected by the sensor nodes, and specific applications for monitoring, visualizing, and alerting concerning important events and scenarios are used [14]. Before a system are designed and deployed, a detailed understanding of the physical environment and deployment is required. An effective system has to meet goals including sensor integration, data collection and quality, data size, cost, robustness, battery use, and data loss management. Low cost maintenance is a goal of a system, due to the harsh nature of the environment, distance of the site, Internet outage, and the high cost of transportation. Finally, applications are developed to monitor the health of each sensor, and to alert administrators and researchers of bad health, to enable them to take action to fix the malfunction.

System health monitoring success factors depend on sensor deployment, Internet, and cloud storage for real-time sensor data, administrator actions, and weather conditions. Furthermore, the system must process sensor data within a certain time frame [10]. The alerting information should employ a dynamic and powerful infrastructure that can be adapted to current requirements. These features are commonly supported by cloud computing systems [15-17], forecasting the maintenance and estimating the time for replacing batteries or sensors.

This paper describes a methodology and supporting tools that integrates event detection and decision support system techniques that can be used for the development of system health monitoring, data analysis, and visualization. There were 4 steps taken: (1) computational data pre-processing, (2) behavior learning, (3) anomaly detection, and (4) sensor quality assessment. Online system health monitoring and visualization tools (OSHM) are very useful for administrators and technical users to take initiatives for preventive maintenance, and can be applied for climate change, environmental, and disaster monitoring systems.

Materials and methods

Study site

This study presents the application of system health monitoring and diagnosis of system failure or malfunction at Racha Island (Latitude 7.60488 °N, Longitude 98.37660 °E), situated in the Andaman Sea (**Figure 1**). The coral reefs in this area are typically shallow (1 - 15 m depth) fringing [17,18]. The monthly temperature ranges between 25 and 30 °C. It was not easy for us to maintain the sensors and diagnose malfunctions of the system, due to a lack of both a stable electricity supply and a stable Internet connection, as well as due to heavy rainfall during the monsoon season [17,18].

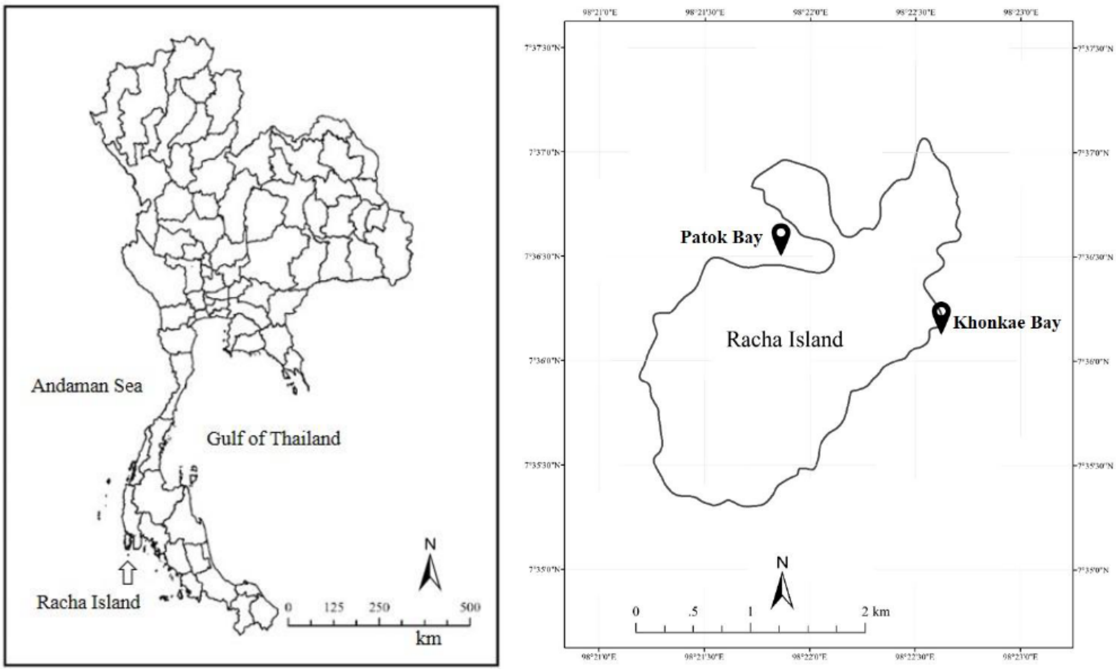


Figure 1 Map of study site; Racha Island, Thailand, is located at Latitude 7.60488 °N, and Longitude 98.37660 °E.

System architecture

Coral reef ecosystems are sensitive to climate changes in the physical environment. To expand the understanding of coral reefs, we deployed 5 real-time online sensors in both Khonkae and Patok Bays: (1) two CTD (conductivity, temperature, and depth) sensors (Sea-bird SBE 37 IM), with a 5 min sampling frequency at depths of 12 and 16 m, (2) two TD (temperature and depth) sensors (Sea-bird SBE 39 IM), with a 5 min sampling frequency at depths of 4 and 8 m, and (3) one automatic weather station (AWS) (Davis Vantage Pro II Plus, measuring 36 weather parameters with a 1 min sampling frequency), at a height of 40 m. The OSHM was programmed using the computer language Hypertext Markup Language (HTML), with PHP technology, while Apache Web Server (Version 2.2.8) and MySQL (Version 5.0.51b) were used as the web server and database management system. Highchart JS and Amcharts JS were used for visualization. Three servers were used for archiving, computation, and visualization (cloud storage, data center, and social media) (**Figure 2**).

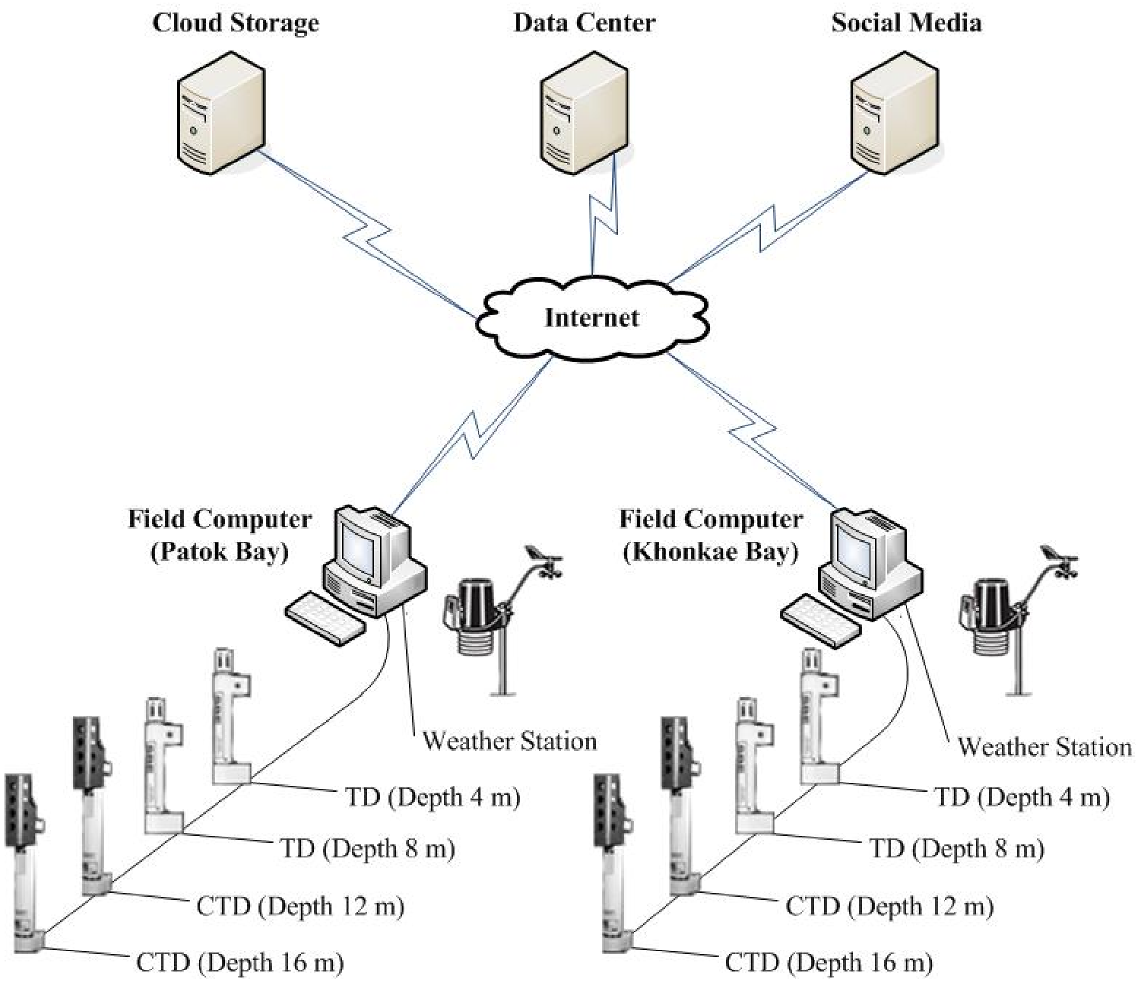


Figure 2 System Architecture of field computers at Patok and Khonkae Bays at Racha Island, and all servers (cloud storage, data center and social media).

Components of the OSHM

The OSHM was composed 4 steps: (1) computational data pre-processing, (2) behavior learning, (3) anomaly detection, and (4) sensor quality assessment. Data from the field computers at Patok and Khonkae Bays were from underwater sensors (CTD and TD) and AWS. The context diagram of this OSHM is shown in **Figure 3**. In step one, underwater sensors collected conductivity, temperature, and depth from CTD, and temperature and depth from TD. AWS collected 36 parameters, such as temperature, UV, rainfall, rain rate, humidity, wind speed, wind direction, and etc. Data from CTD, TD, and AWS were exported, cleaned-up, and converted to .xls files and database tables. These database files were uploaded to the server for utilization in data analysis and visualization at cloud storage and data center. In step two, the behavior learning process, historical data with anomalistic data removed were statistically processed offline and a Poisson distribution was observed. So, to create adaptive behavior rules, which are different from the Gaussian or normal distributions where the idea of +/- SD can be applied, the 98 % Quartile (or any standard rejection value to be selected, e.g. 5 or 10 %) are computed. This record can provide the value needed in behavior rules for anomaly detection. For example, at Khonkae Bay, at the 98 % Quartile lag time was about 14 days. On the other hand, at Patok Bay, the 98 %

Quartile lag time was 27 days. We found 2 behaviors from Patok and Khonkae Bays (**Figure 4**). In step three, the behavior rules were used for detecting anomalies, and anomaly reports were sent to the administrator. In step four, for near real-time, conductivity, temperature, and depth data from 4 CTDs and 4 TDs were uploaded to the cloud storage every one hour. The process checked the state of all of the sensors processed on the server side, and sent sensor health reports to the administrator and fed them to social media automatically.

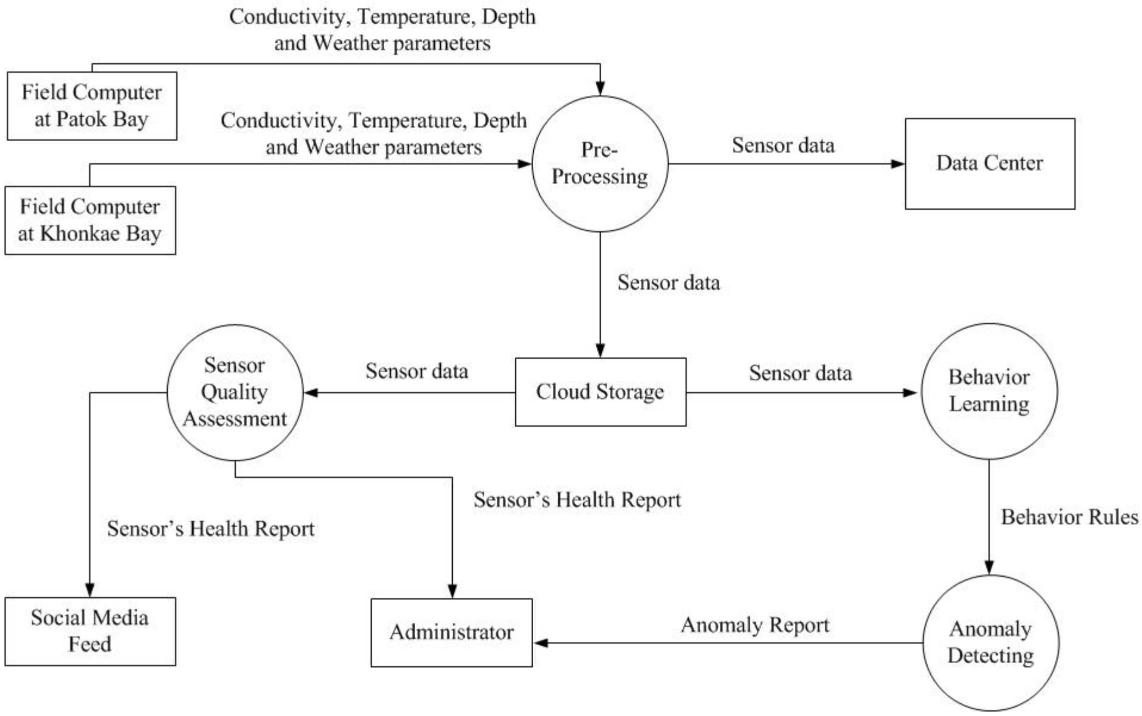


Figure 3 Context Diagram of pre-processing, behavior learning, anomaly detecting, and sensor quality assessment of sensor data from field computers at Patok and Khonkae Bays.

Results and discussion

Lag time behavior

There were 2 patterns of lag time from Patok and Khonkae Bays (**Figures 4a** and **4b**). **Figure 4a** represents the percentage of probability of lag time from Patok Bay. This site has very low Internet bandwidth and a lack of staff that are able to check the computers, equipment and Internet. **Figure 4b** represents the normal lag time from Khonkae Bay. The highest amount of data from this site was sent to the data center to the date of the day. The maximum lag time was 17 days at Khonkae Bay. Internet connection at Khonkae Bay has higher bandwidths than at Patok Bay, and also at this site, there is a technical staff on site. The results show that Khonkae Bay during April to June and September has less lag time than Patok Bay. The least lag time was at Khonkae Bay in June 2014. It was only a half day lag time. This could be because there were fewer tourists that shared Internet bandwidth during June due to low tourist season in the area.

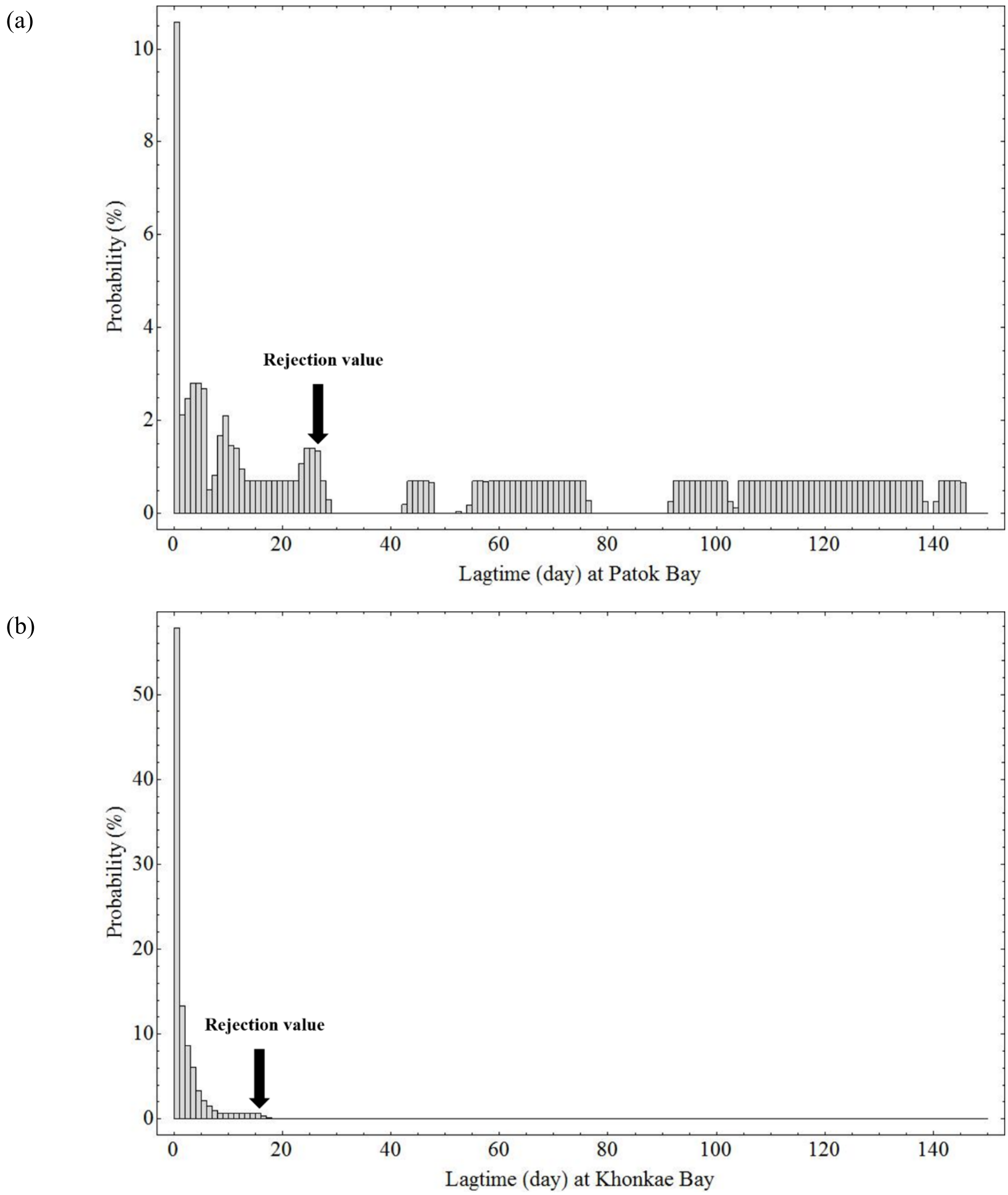


Figure 4 Probability (%) of 2 patterns of lag time from Patok and Khonkae Bays.

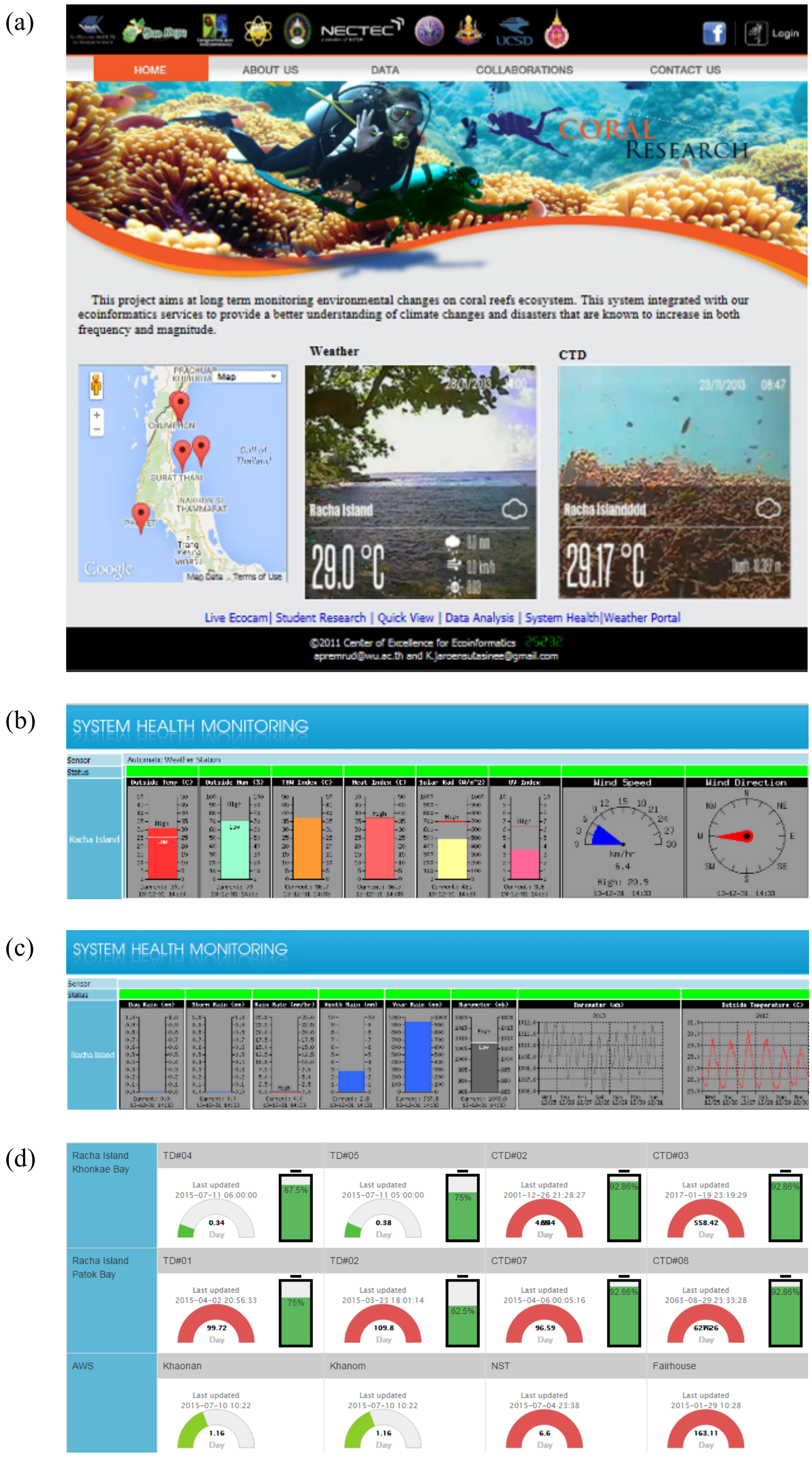


Figure 5 Online system health monitoring for coral database; (a) homepage of www.twibl.org/coraldbn, (b) temperature, (c) rainfall, (d) last updated data, and battery's status (http://164.115.22.116:81/systemhealth/).

Contents in system health monitoring

It is essential that all illustrations are as clear and as legible as possible. Online system health monitoring is an automatic alerting system. This process is on the server side (**Figures 5b - 5d**). For reporting systems, each row shows real-time online sensors status. For example, at Racha Island, **Figures 5b** and **5c** show weather data, namely outside temperature and humidity, total heat index, heat index, solar radiation, UV index, wind speed, wind direction, daily rainfall, storm rain, rain rate, monthly rain, yearly rain, barometer, etc. Users can hover the curser over each parameter to see more detail. **Figure 5d** shows the underwater parameters, namely water temperature, conductivity, and pressure. Due to the advancement in sensor technology, we could store data in sensor memory for more than 6 months (CTD and TD). However, our behavior rules required that most data were sent to the server within one day. So we defined 3 sensor status indicators: green referred to data sent within one day, yellow referred to data sent between one and 2 days, and red referred to data sent more than 2 days prior. The alert criteria could be changed by the administrator if the behavior rules changed.

Contents in data analysis and visualization tools

The data analysis and visualization tools provide data accessible in 6 web pages: Home, About Us, Data, Collaboration, and Contact Us. The home page shows the study site, using Google Maps, on-beach environment data, with real-time temperature, wind speed, and UV index, from AWS, and water temperature and depth at the depth of 10 m from CTD. The About Us page provides details of our research aims. The Data page provides details on the sensors that we deployed, and users can select study sites and parameters for graph plotting. The Collaboration page lists organizations that support funding and the sharing of knowledge for our research. The Contact Us page shows telephone, email, and Facebook page details. These services are available online at www.twibl.org/coraldbn (**Figure 5a**). The bottom of this page shows a quick link to student research, quick view, data analysis, system health, and a weather portal.

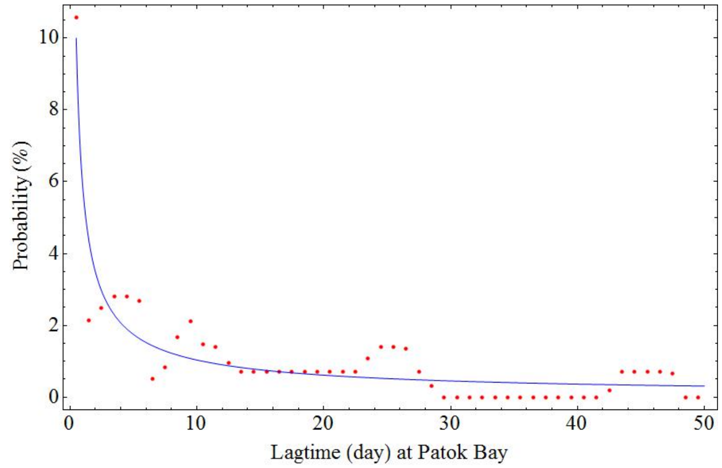
Discussion

The lag time from Khonkae Bay each day was monitored, it was noticed that the least lag time was of one day, and the most was 17 days obeying a Poisson distribution. This is because, when the sensor's health report was sent to the administrator, the staff on site could check the Internet and equipment during 3 to 5 PM only and sometimes had to perform other duties on land. The staff at Khonkae Bay has knowledge of information technology. Khonkae Bay has a stable Internet connection and electricity supply, and this site also has an electricity converter.

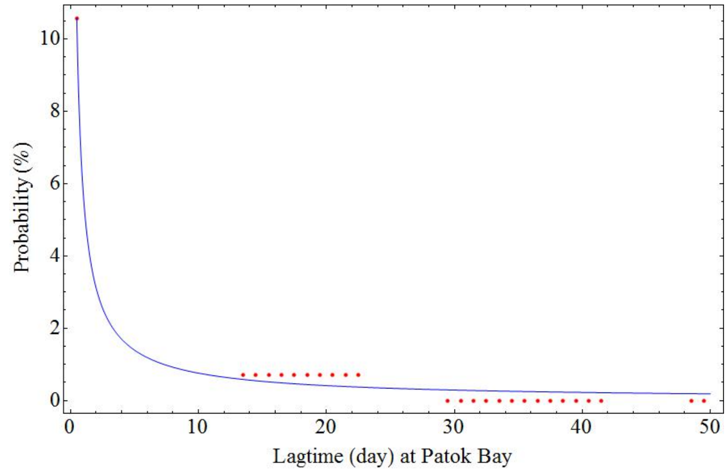
On the other hand, the lag time from Patok Bay did not obey a Poisson distribution. The behavior of the lag time looks like slot. The most lag time was one day, and the others were slot during 0 - 30 days, 42 - 47 days, 55 - 78 days and 92 - 138 days, respectively. When we went to perform both underwater sensors and AWS maintenance on site usually every one to 3 months, there was no lag time lasting for 14 - 28 days. The staff at this Patok site has little technical knowledge, so when the sensor health report was sent to the administrator, it was difficult to solve the problem with the help of the Patok staff. In addition, this site also does not have a stable Internet connection or electricity supply. From the observation data at Patok Bay, **Figure 6a** represented an unusual lag time. There were 4 points indicated the maintenance time for administrator. **Figure 6b** represented the normal lag time that followed the power-law distribution.

For future work, we can apply the power-law distribution to detect unusual lag times which would helping us monitor the Internet strength 24/7. When there is a slow Internet connection, we can train staff to manage it, so that the important sensor data will continue to be sent to the server.

(a)



(b)



(c)

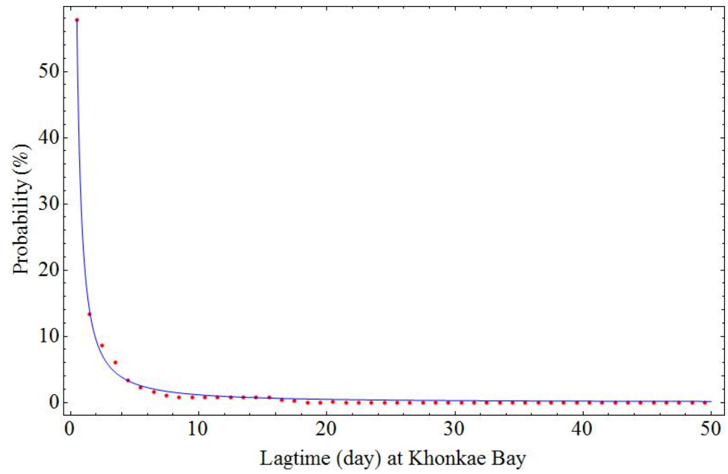


Figure 6 The Power Law Distribution (a,b) at Patok Bay and (c) at Khonkae Bay.

Conclusions

We developed an OSHM tool that provides an essential tool for alerting of messaging, querying, and visualizing coral reef sensor, based at 6 sites, and publicly accessible for researching environmental changes on coral reefs. The OSHM provides 3 distinct advantages. First, system failure or malfunction detection can be relayed to notify the administrators or technical users in charge of the task by social networking. Second, researchers can visualize large data set results online and perform long term monitoring. Third, analyzing lag time is an important tool for administrators to manage archived time of sensor data from each site. This is a new way to sense and understand the environment, which can be applied to many areas of the environmental sciences, and also allows a new style of monitoring [1]. We believe that our proposed methods and solutions are practical for system health monitoring, as well as a key component of environmental analysis [1], and continuous system monitoring and maintenance that can make systems operate for longer and run more safely and robustly.

Acknowledgements

We thank John Barker and 2 anonymous referees for useful comments on previous versions of this manuscript. This work was supported in part by the Office of the Higher Education Commission for supporting by grant fund under the program Strategic Scholarships for Frontier Research Network for the Ph. D. Program Thai Doctoral degree (2/2552), Walailak University Fund (10/2552), Bioresources Research and Technology Program-NSTDA: BRT Program-NSTDA (Grant No. P-12-01254), Walailak University (Grant No. WU55107), Center of Excellence for Ecoinformatics, the Institute of Research and Development, Walailak University and NECTEC, AIMS, UCSD, Ban Raya Resort & Spa and Raya Resort for providing research facilities at Racha Island. We are also thankful to Asst. Prof. Dr. Uthai Kuhapong, Dr. Sirilak Chumkiew, Sunthon Suraban, Kiettisak Chaidej and Yudtachai Suwanarat for invaluable assistance in the field.

References

[1] JK Hart and K Martinez. Environmental sensor networks: A revolution in the earth system science? *Earth Sci. Rev.* 2006; **78**, 177-91.
[2] P Bect, Z Simeu-Abazi and P Maisonneuve. Identification of abnormal events by data monitoring: Application to complex systems. *Comput. Ind*. 2015; **68**, 78-88.
[3] R Wald, TM Khoshgoftaar, P Beaujean and JC Sloan. A review of prognostics and health monitoring techniques for autonomous ocean systems. *In*: Proceeding of the 16th ISSAT International Conference on Reliability and Quality in Design, Washington DC, USA, 2010.
[4] DJ Hill and BS Minsker. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environ. Model. Software* 2010; **25**, 1014-22.
[5] M Mourad and JL Bertrand-Krajewski. A method for automatic validation of long time series of data in urban hydrology. *Water Sci. Tech*. 2002; **45**, 263-70.
[6] K Goebel, B Saha, A Saxena, JR Celaya and J Christophersen. Prognostics in battery health management. *IEEE Instrum. Meas. Mag.* 2008; **11**, 33-40.
[7] J Rabatel, S Bringay and P Poncelet. Anomaly detection in monitoring sensor data for preventive maintenance. *Expert Syst. Appl.* 2011; **38**, 7003-15.
[8] B Balis, M Kasztelnik, M Bubak, T Bartynski, T Gubała, P Nowakowski and J Broekhuijsen. The Urban flood common information space for early warning systems. *Procedia Comput. Sci.* 2011; **4**, 96-105.
[9] R Kothamasu, HS Huang and WH VerDuin. System health monitoring and prognostics: A review of current paradigms and practices. *Int. J. Adv. Manuf. Tech*. 2006; **28**, 1012-24.
[10] B Balis, M Kasztelnik, M Bubak, T Bartynski, T Gubała, P Nowakowski and J Broekhuijsen. The Urban Flood common information space for early warning. *Syst. Procedia Comput. Sci*. 2011; **4**, 96-105.

[11] FEA Horita, JP Albuquerque, LC Degrossi, EM Mendiondo and J Ueyama. Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Comput. Geosci.* 2015; **80**, 84-94.
[12] Computer Science and Telecommunications Board, Embedded. *Everywhere: A Research Agenda for Networked Systems of Embedded Computers*. National Academy Press, Washington, 2001, p. 1-34.
[13] CY Chong and SP Kumar. Sensor networks: Evolution, opportunities, and challenges. *Proc. IEEE* 2003; **91**, 1247-56.
[14] K Martinez, JK Hart and R Ong. Environmental sensor networks. *Computer* 2004; **37**, 50-6.
[15] MN Bennani and D Menasce. Resource allocation for autonomic data centers using analytic performance models. *In:* Proceeding of the 2nd International Conference on Automatic Computing, IEEE Computer Society, Washington DC, USA, 2005, p. 229-40.
[16] X Wang, Z Du, Y Chen, S Li, D Lan, G Wang and Y Chen. An autonomic provisioning framework for outsourcing data center based on virtual appliances. *Cluster Comput.* 2008; **11**, 229-45.
[17] B Sotomayor, RS Montero, IM Llorente and I Foster. Virtual infrastructure management in private and hybrid clouds. *IEEE Internet Comput.* 2009; **13**, 14-22.
[18] M Jaroensutasinee, K Jaroensutasinee, T Fountain, M Nekrasov, S Chumkiew, P Noonsang, U Kuhapong and SJ Bainbridge. Coral sensor network at Racha Island, Thailand. *In*: Proceeding of the Environmental Information Management Conference, Santa Barbara, USA, 2011, p. 82-6.
[19] M Jaroensutasinee, K Jaroensutasinee, SJ Bainbridge, T Fountain, S Chumkiew, P Noonsang, U Kuhapong, S Vannarat, A Poyai and M Nekrasov. Sensor network applications for reefs at Racha Island, Thailand. *In*: Proceeding of the 12th International Coral Reef Symposium, Cairns, Australia, 2012, p. 95.