An integrated database for marine environment monitoring and management system at the Tongyoung bay in Korea

Sung-Dae Kim, Sang-Hwa Choi, Hyuk-Min Park*
Korea Institute of Ocean Science and Technology, *sdkim@kiost.ac.kr

Abstract

KIOST (Korea Institute of Ocean Science and Technology) started to develop a marine environmental monitoring and management system to support efficient operation of the aqua-farm at Tongyoung bay in 2017. The system consists of an intensive observation system, a data management system, a hybrid environment prediction system, and a facility management system based on artificial intelligence technology. During the 1st phase of the project (2017-2018), we are setting up a database system to manage marine data of the intensive observation and to archive all available environmental data of Tongyoung bay area. Several physical parameters and chemical parameters are measured at the surface buoy and automatic vertical profiler, and the collected data are provided to the researchers in real-time mode. Observed data of biological parameters and biogeochemical parameters are being submitted to the database system by the related researchers in delayed mode. We also collected all oceanographic data produced by several organizations and satellites. We manipulated all collected data in accordance with the metadata standard and quality control procedures of KIOST which had been prepared for research data management. An internet web site was established to provide real-time data to researchers. To develop marine environment prediction system, a data driven model and a numerical dynamic model are under development. We are trying to build a data driven model based on LSTM (Long Short tem Memory) network of RNN (Recurrnet Neural Network) to simulate biological parameters using physical parameters and chemical parameters. Tensorflow librarys was used to set up machine learning programs. An ensemble model will be set up to combine the results of the data driven model and the numerical model and provide prediction information to the aquafarm operation system.

1. Integrated DB system

A DB system was developed to manage collected marine data at Tongyoung bay and provide them to data driven model and numerical model. DB structure was designed after analyzing characteristics of the collected data and DB system was established using Oracle DB software. All ocean data obtained during physical, chemical, and biological observations was converted to the ASCII data files which meet DB structure. Table 1 shows data items of the integrated DB system and Figure 1 shows the DB structure. We can handle DB tables and the contents of DB system using DB tools (Figure 2). Figure 3 shows all observation stations of Tongyoung bay operated by other organizations.

Table 1. Data items collected and archived at Tongyoung bay

Observation	Data Item
In-situ observation	Temperature, Salinity, Current, Chlorophyll, NO4, NO2, NO3, PO4, SIOH4, SOD, Micro Phytoplankton, Primary Production
Real-time observation (Buoy, Vertical Profiler)	Temperature, Salinity, Air Temperature, Air Pressure, Humidity, Wind, pH, Air PCO2, Sea PCO2
Archived data from other institutes	SST, Temperature, Salinity, Air Temperature, Air Pressure, Wind, DO, pH, COD, NH4, NO2, NO3, DIN,TN, DIP, TP, SIO2, SS, Chlorophyll

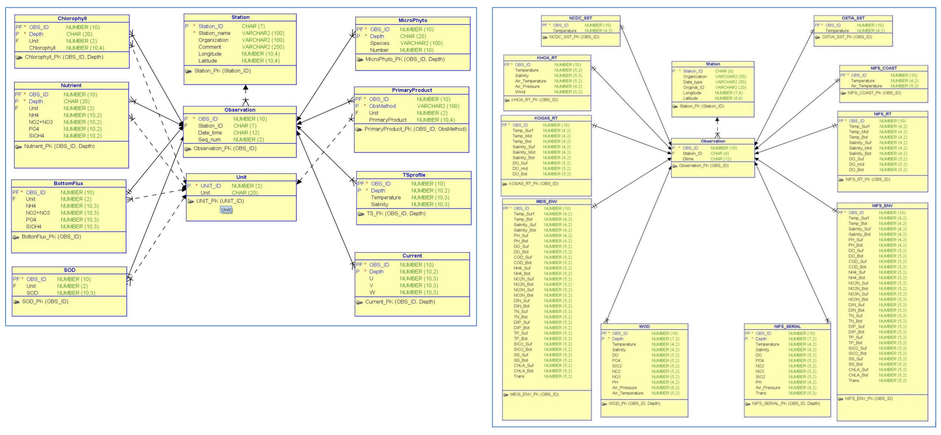

Figure 1. Entity relational diagram of the integrated DB system.

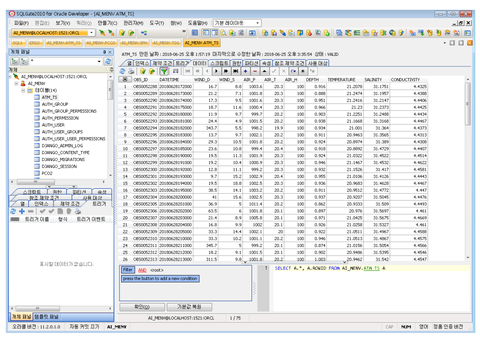

Figure 2. DB tables and contents of the DB system.

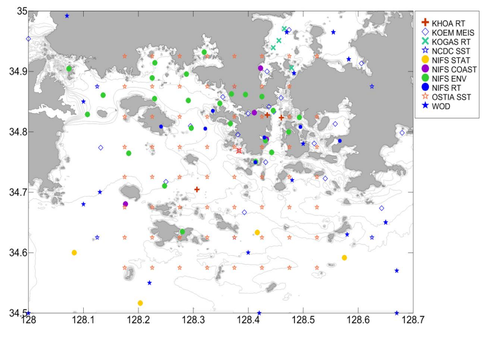

Figure 3. Observation stations operated by other organizations.

2. Real-time service of the automatic observation data

Some physical data items and chemical data items are measured at automatic observation platforms. All measured data at the ocean buoy and the automatic profiling system are transmitted to KIOST data server through the telephone communication network in real-time mode. An intranet web site was set up to provide them to the researchers in near real-time mode. It was developed by Python language and several open source packages like Django, CX-Oracle, Fusion Chart. Figure 4 shows the screen of real-time data service set up by this work.

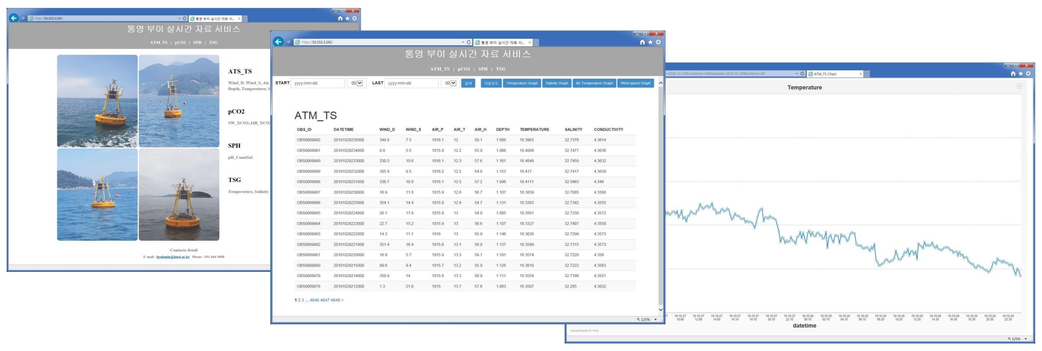

Figure 4. Screens of the real-time data service

3. Development of the data driven model

Before developing data driven model which can predict marine ecosystem, machine learning techniques were tested using open source programs and several ocean data. To check possibility of chlorophyll prediction, 5 kind of tests were accomplished using LSTM network of RNN in Google Tensorflow package (Table 2). The collected observation data and satellite data were retreated to make up for deficient data. Test 5 uses ocean ecosystem model output as training data sets (Figure 5). It's result shows LSTM network can predict chlorophyll when sufficient data was provided. Maximum correlations was around 0.5 when we use satellite data and meteorological model output. The low correlations are caused by mainly insufficiency of chlorophyll data.

Table 2. List and results of machine learning tests to predict Chlorophyll

Test No.	Training Data Sets		Period	Data Interval	Number of Data	Result (R)
1	-Chl-a (MODIS) -SST (REMSS)	- Day Light (KMA) - Humidity (KMA)	2002.07.17. - 2017.12.31	10 days	564	~ 0.25
2	-Chl-a (GOCI) -SST (REMSS)	-Solar radiation (ECMWF)	2011.04.01. - 2017.12.31	3 days	822	~ 0.54
3	-Chl-a (GOCI) -SST (REMSS)	-Solar radiation (ECMWF)	2011.04.01. - 2017.12.31	5 days	493	~ 0.53
4	*Observation Data -Temperature -Solar radiation (ECMWF)	-Nutrient (NO2, NO2+ NO3, NH4, PO4,SIO2) -Chl-a	2014.01. - 2015.08.	15 days	40	< 0.1
5	*Dynamic Model Data -Solar radiation -Temperature -Chl- a	-PO4(t) -NO2+NO3(t) -Chla(t-1)	1980.1.1. – 2016.12.31	1 day	13,514	~ 0.995

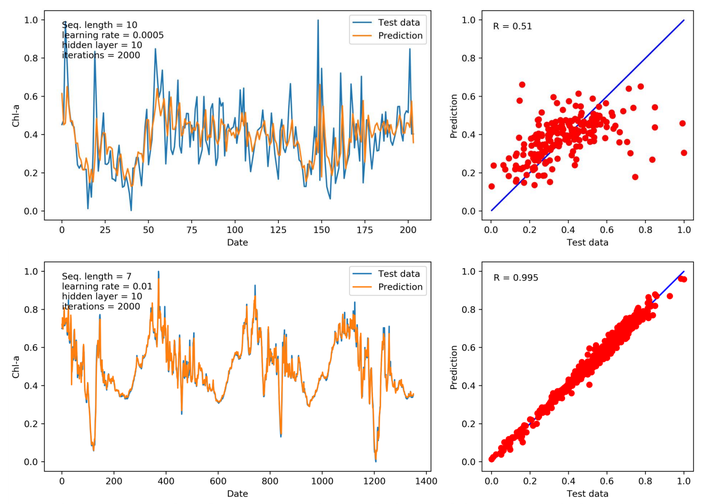

Fig 4. Result of test running of data driven model
Test 3 (upper), Test 5 (lower)

Acknowledgments

This study was supported by the project titled "Development of marine environment monitoring and management system in aquafarm based on artificial intelligence (PE99611)".