# Estimation of Water Quality of Fish Farms using Multivariate Statistical Analysis

Hee-Taek Ceong, Hae-Ran Kim, *Member, KIMICS*

***Abstract*— In this research, we have attempted to estimate the water quality of fish farms in terms of parameters such as water temperature, dissolved oxygen, pH, and salinity by employing observational data obtained from a coastal ocean observatory of a national institution located close to the fish farm. We requested and received marine data comprising nine factors including water temperature from Korea Hydrographic and Oceanographic Administration. For verifying our results, we also established an experimental fish farm in which we directly placed the sensor module of an optical mode, YSI-6920V2, used for self-cleaning inside fish tanks and used the data measured and recorded by a environment monitoring system that was communicating serially with the sensor module. We investigated the differences in water temperature and salinity among three areas—Goheung Balpo, Yeosu Odongdo, and the experimental fish farm, Keumho. Water temperature did not exhibit significant differences but there was a difference in salinity (significance <5%). Further, multiple regression analysis was performed to estimate the water quality of the fish farm at Keumho based on the data of Goheung Balpo. The water temperature and dissolved-oxygen estimations had multiple regression linear relationships with coefficients of determination of 98% and 89%, respectively. However, in the case of the pH and salinity estimated using the oceanic environment with nine factors, the adjusted coefficient of determination was very low at less than 10%, and it was therefore difficult to predict the values. We plotted the predicted and measured values by employing the estimated regression equation and found them to fit very well; the values were close to the regression line. We have demonstrated that if statistical model equations that fit well are used, the expense of fish-farm sensor and system installations, maintenances, and repairs, which is a major issue with existing environmental information monitoring systems of marine farming areas, can be reduced, thereby making it easier for fish farmers to monitor aquaculture and mariculture environments.**

***Index Terms*— water quality, fish farms, multiple regression analysis, mariculture, multivariate statistical analysis.**

Manuscript received June 7, 2011; revised June 23, 2011; accepted July 8, 2011.
Hee-Taek Ceong and Hae-Ran Kim are with the Department of Digital Convergence, Chonnam National University, Yeosu, 550-749, Korea (Email: htceong@jnu.ac.kr)

## I. INTRODUCTION

**THE** farming of aquatic organisms involving the cultivation of freshwater and seawater populations under controlled conditions is called aquaculture. For the last 40 years, the practice of aquaculture, including mariculture, i.e., the cultivation of marine organisms, has been rapidly increasing worldwide. Today, aquatic seafood represents important and substantial portions of the seafood consumed by humans, and fisheries are decreasing tendency. Moreover, the demand for seafood and fish products is expected to increase continuously. In addition, aquaculture, which has an annual growth rate of 10%, contributes to food security in the form of a provisional source of future food supply.

The importance of aquaculture has been recognized and acknowledged. The futurologist Alvin Toffler had predicted that in an information-oriented age, the seafood industry, including aquaculture, will become one of the four major industries, while the social ecologist Peter Drucker has written that fish farming would be more important than the Internet in the 21st century.

In order to improve our living standards and ensure reliable food supply in the near future, information and communication technologies must be applied to fishery farming treating it as a high-value industry. In particular, it is important to achieve improvements in the productivity and quality of produced seafood, assure reliability, and ensure immediate disposal in the case of abnormal phenomena such as uncommon temperatures or winds. Moreover, it is important to ensure that the industry is environment friendly.

Efforts to perform real-time monitoring of the water quality of fish farm areas in order to produce seafood that is both environment friendly as well as of high quality have been previously attempted in researches such as "the fish tank information monitoring service business utilizing USN" of "ubiquitous fish farm" in the Jeju special self-governing province and the "environmental information monitoring system" demonstration project of the Eco Aquafarm Research Center, Chonnam National University [1][2][3].

By using these monitoring systems, fishery producers monitor water-quality information such as inner-tank water temperature and salinity at the production site. Moreover, they are also able to conduct rapid treatments and make correspondence possible. In this manner, the systems provide fish farmers with usability and convenience[1][3]. However, in these systems, regular corrections are required, and moreover, sensors must be replaced regularly, causing operational and maintenance issues. In particular, the first stage involves huge purchase costs for sensors and monitoring system equipment and places a large burden on small-scale fish farmers [3].

Therefore, we have attempted to solve these problems by providing water-quality information using oceanic environment data from a trusted national institution, Korea Hydrographic and Oceanographic Administration that is located closer to the fish farm areas; note that this is different from the current method in which the monitoring data is collected using sensors installed in fish tanks. This research is an expansion of a paper "the statistical analysis of water quality in a land-based fish farm[3] whose analysis scientifically summarized the characteristics of water-quality factors and presented a dissolved-oxygen predictive model. In this model, in the cost reduction dimension about the dissolved-oxygen sensor in which the cyclical replacement is needed.

This paper is organized as follows. In the section Materials and Methods, we describe the research areas, data, and statistical analysis method. The section Results and Discussion describes the water-quality characteristics of the fish farm, the difference analysis of water temperatures and salinity of three inter-regional areas, water-quality characteristics of nine factors obtained from the abovementioned national institution, multiple regression model estimation for the water quality of fish farms, the fit of the estimated result by the regression equation, and the actually measured values; we end this section by discussing an improved environmental information monitoring system. The paper concludes with the section Conclusions, with a summary of estimation of water quality for fish farms as a whole.

## II. MATERIALS AND METHODS

### *A. Research Area*

The following three regions were used for this research. The first was a self-testing fish farm, the Keumho of a land-based fish farm located at Geogeum-do, Goheung-gun, Jeollanam-do. The other two regions were oceanic environment areas observed by the National Oceanographic Research Institute[4], the Odongdo Island of Yeosu and Balpo areas of Goheung.

In November, 2008, the abovementioned demonstration project of the environmental information monitoring system established a sensor and monitoring system in the fish farm at Geogeum-do. Furthermore, environmental information (per minute) such as water temperature, salinity, dissolved oxygen, chlorophyll, and pH were obtained from a smart sensor system installed inside the tank [1][3].

We also obtained measured data (per hour) such as tide level, wind speed, wind gust, wind direction, temperature, air pressure, conductivity, water temperature, and salinity of Goheung Balpo and Yeosu Odongdo areas from the Korea Hydrographic and Oceanographic Administration [5] and used them in this study.

Fig. 1. Map showing the location of fish farm and monitoring stations.

### *B. Dataset*

The dataset was analyzed and classified into five sets for the three areas, as shown in Table 1.

TABLE 1
DATASET FOR STATISTICAL ANALYSIS

| Set | Items | Period | Areas |
|---|---|---|---|
| set1 | water temperature, salinity, dissolved oxygen, chlorophyll, pH, oxygen saturation | Jan. 2009 ~ Feb. 2009 | A |
| set2 | tide level, wind speed, wind gust, temperature, conductivity, wind direction, air pressure, salinity, water temperature | Jan. 2009 ~ Feb. 2009 | B,C |
| set3 | water temperature, salinity | Jan. 2009 ~ Feb. 2009 | A,B,C |
| set4 | A: water temperature, salinity, pH, dissolved oxygen<br>B: water temperature, salinity, conductivity, air pressure, wind speed | Jan. 2009 ~ Feb. 2009 | A,B |
| set5 | A: water temperature, dissolved oxygen<br>B : water temperature, salinity | Feb. 4, 2010 ~ July 30, 2010 | A,B |

Note : A = Keumho fish farm; B = Balpo area; C = Yeosu area.

The first dataset was constructed to investigate the characteristics of the water quality of the Keumho fish farm; the second dataset was used to examine the characteristics of the oceanic environment data; and the third one analyzes the differences in the water temperature and salinity, i.e., the co-factors between the three areas. The fourth dataset is for estimating the environmental information of the Keumho fish farm by using the nine factors observed in the Goheung Balpo area, i.e., the near field. The fifth dataset is for plotting the predicted values by the regression estimation equation and actually measured values.

The first dataset was constructed to investigate the characteristics of the water quality of the Keumho fish farm; the second dataset was used to examine the characteristics of the oceanic environment data; and the third one analyzes the differences in the water temperature and salinity, i.e., the co-factors between the three areas. The fourth dataset is for estimating the environmental information of the Keumho fish farm by using the nine factors observed in the Goheung Balpo area, i.e., the near field. The fifth dataset is for plotting the predicted values by the regression estimation equation and actually measured values.

### *C. Statistical analysis*

The statistical processing for the statistical analysis was performed by using SAS® 9.2, a representative statistical package. In our analysis, we used the daily means of the water-quality data that are measured every second in the experimental fish farm.

Analysis of variance (ANOVA) was performed for the difference verification of the water temperatures and salinities among the three areas. In addition, the daily means of the oceanic environment (nine factors) were used for the multiple regression analysis in order to estimate the water quality of the fish farm and to plot the estimated results and actually measured values. Correlation analysis and factor analysis were applied to the measured raw data (per hour) of the oceanic environment (nine factors) that were obtained from the Korea Hydrographic and Oceanographic Administration.

Factor analysis was performed on correlation matrix of rearranges data, so that it explains the structure of the underlying data set. The correlation coefficient matrix measures how well the variance of each constituent can be explained by relationship with each of others[6][7]. In general, a factor with an eigenvalue greater than 1 is considered significant [8]. We decided the total number of factors to be number of factors with eigenvalue greater than 1.

If the correlation between independent variables is strong in the multiple regression analysis wherein the independent variables are included in the model, the independent variables have an unexpected effect on the dependent variable; this causes the accuracy of the estimated coefficient of regression to reduce, and the results may be distorted [9]. Therefore, a process to remove the exposition variable after diagnosing the multicollinearity problem was performed; We also conducted the process included input order decisions of the exposition variable, outlier, and influential observation diagnosis besides the multicollinearity diagnosis to obtain final regression model conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

## III. RESULTS AND DISCUSSION

### *A. Water Quality of Fish Farm*

Table 2 lists the water-quality data of the Keumho fish farm in 2009, and N indicates the observational days. In particular, in May and June, the observational days are few due to the cleaning of water tanks and system non-operation, and representativeness is insufficient. Nevertheless, it is possible to analyze the annual data transition of the fish farm. The water temperature was lowest in January and highest in August due to seasonal factors. The concentration of the dissolved oxygen required for the fishes to remain alive is 4 ppm or greater. The dissolved oxygen in the water temperature and inverse proportion relations were observed at pH greater 6.2. The pH was mostly greater than 7.0, implying that the water was alkaline. No large-scale fluctuations in the chlorophyll density could be observed; further, the concentration was high when great quantities of nutritive salts were present. Moreover, the values for July, October, November, and December were slightly higher than those in the other months.

TABLE II
WATER-QUALITY STATISTICS OF KEUMHO FISH FARM, 2009

| Var. / Mon. | Temp. | Sal. | DO | Chl. | | pH | | N |
|---|---|---|---|---|---|---|---|---|
| | X±SD | X±SD | X±SD | X±SD | | X±SD | | (days) |
| 1 | 7.37±0.52 | 30.75±0.56 | 10.05±0.24 | 1.31 | 0.52 | 8.14 | 0.06 | 30 |
| 2 | 8.82±0.42 | 30.08±0.93 | 9.78±0.16 | 1.22 | 0.33 | 8.18 | 0.03 | 26 |
| 3 | 10.26±0.94 | 29.14±0.85 | 9.40±0.34 | 1.64 | 0.50 | 8.19 | 0.04 | 30 |
| 4 | 12.92±0.94 | 30.02±1.48 | 8.33±0.40 | 1.51 | 0.55 | 7.99 | 0.20 | 26 |
| 5 | 14.88±0.86 | 27.50±0.18 | 8.21±0.15 | 1.28 | 0.32 | 8.18 | 0.07 | 5 |
| 6 | 22.07±1.53 | 29.54±0.36 | 6.26±1.26 | 1.48 | 0.66 | 8.18 | 0.08 | 13 |
| 7 | 21.55±0.50 | 27.99±0.78 | 6.86±0.27 | 2.91 | 1.37 | 7.98 | 0.13 | 27 |
| 8 | 24.69±1.87 | 28.44±0.81 | 6.31±1.15 | 1.06 | 0.22 | 8.27 | 0.17 | 28 |
| 9 | 24.10±0.70 | 30.57±0.62 | 6.41±0.20 | 1.68 | 0.79 | 8.29 | 0.05 | 30 |
| 10 | 20.32±1.65 | 30.17±0.48 | 7.24±0.33 | 2.98 | 0.47 | 8.17 | 0.06 | 29 |
| 11 | 14.80±1.97 | 28.60±0.87 | 8.09±0.38 | 3.30 | 0.68 | 7.97 | 0.13 | 26 |
| 12 | 10.62±1.29 | 28.50±0.77 | 9.26±0.66 | 2.85 | 0.39 | 8.19 | 0.08 | 15 |

Note : Mon. = Month; Var. = Variables; Temp. = Temperature; Sal. = Salinity; DO = Dissolved Oxygen; Chl. = Chlorophyll

TABLE III
ESTIMATED MULTIPLE LINEAR REGRESSION models (n = 285)

| Multiple linear regression Model | Adj. $R^2$ | Sig. |
|---|---|---|
| Temp. = -9.902 - 4.15 × DO + 7.29 × pH | 0.94 | * |
| DO = -1.909 - 0.227 × Temp. + 1.669 × pH | 0.94 | * |
| pH = 8.169 + 0.004 × Temp. - 0.051 × Chl. | 0.13 | * |
| Sal. = 30.512 - 0.027 × Temp. - 0.317 × Chl. | 0.09 | * |
| Chl. = 23.522 + 0.02 × Temp. - 2.082 × pH - 0.168 × Sal. | 0.16 | * |

Note : Temp. = Temperature; Sal. = Salinity; DO = Dissolved Oxygen; Chl. = Chlorophyll;
* = Significance at the 95% confidence level

The optimum regression model is shown in Table 3; here, among the measured five qualities of water factors, one quality was not measured. The number of observational days in Table 3 was 285 days, and the daily mean for five factors was used for the estimation.

The optimum regression model for each variable of the water quality is the model chosen through the significance test (significance level = 5%) of the regression model, multiple probability problem, significance test (significance level = 5%) of the exposition variable for each case, and the adjusted coefficients of determination.

However, the adjusted coefficients of the determination of pH, salinity, and chlorophyll are 13%, 9%, and 16% of the total variance. In the case of the water-temperature forecasting model, the dissolved oxygen showing the negative correlation of the water temperature has most of the descriptive abilities, followed by the pH. In addition, in the case of the dissolved-oxygen forecasting model, the water temperature exhibiting the negative correlation of the dissolved oxygen has most of the descriptive abilities, followed by the pH.

*B. Water Quality at Two Oceanographic Observation Facilities*

Table 4 presents the averages and standard deviations of the monthly water-quality data at the Goheung Balpo area and Yeosu Odongdo Island. The tide levels and wind directions wherein the averages were meaningless have been excluded. The Goheung Balpo area had lower temperatures than the Yeosu Odongdo Island area. Further, the water temperatures of the Goheung Balpo area were lower than those of Yeosu Odongdo Island in October, November, December, and January, while they were similar during the other months.

The salinity of the Goheung Balpo area was high than that of Yeosu Odongdo Island, presumably because Yeosu Odongdo Island is adjacent to Yeosu downtown, and there was various inflows such as domestic sewage.

TABLE IV
STATISTICAL SUMMARY OF FIVE FACTORS AT TWO SITES (BALPO, YEOSU) FROM JAN 2009 TO DEC 2009

| Month | wtp(°C) | | tp(°C) | | ap(hPa) | | ws(m/s) | | cd(mS/cm) | | sa(psu) | | gu(°) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | C | B | C | B | C | B | C | B | C | B | C | B | C |
| 1 | 4.98±0.64 | 6.28 ±0.82 | 1.73±3.22 | 2.64±3.01 | 1022.1±4.08 | 1022.0±4.03 | 3.28±1.33 | 3.78±1.75 | 31.92±0.51 | 28.42±5.03 | 33.24±0.16 | 28.09±7.51 | 4.82±2.13 | 5.08±2.46 |
| 2 | 7.87±0.62 | 7.63±0.66 | 5.81±3.15 | 7.06±3.16 | 1017.3±5.87 | 1017.4±5.98 | 2.90±1.26 | 3.01±1.67 | 33.42±1.03 | 27.56±8.19 | 32.14±1.13 | 26.49±6.88 | 4.01±1.74 | 3.97±2.37 |
| 3 | 9.48±1.05 | 9.52±0.85 | 7.42±2.89 | 8.79±2.82 | 1016.8±5.11 | 1017.0±5.19 | 3.51±1.49 | 3.42±1.64 | 35.06±1.83 | 29.37±4.06 | 32.36±1.02 | 28.33±2.05 | 4.93±2.23 | 4.65±2.18 |
| 4 | 13.10±1.38 | 13.06±1.29 | 11.71±2.07 | 13.45±2.13 | 1013.7±6.07 | 1013.9±6.05 | 2.91±1.45 | 2.87±1.85 | 38.33±1.66 | 32.28±4.07 | 32.44±1.56 | 27.97±2.60 | 4.13±2.27 | 3.96±2.47 |
| 5 | 16.63±0.97 | 16.93±0.94 | 16.20±1.36 | 18.09±1.53 | 1010.8±4.03 | 1011.2±4.18 | 2.36±0.96 | 2.13±1.17 | 41.14±1.32 | 33.93±4.41 | 32.05±1.27 | 23.72±4.93 | 3.19±1.49 | 3.02±1.43 |
| 6 | 20.36±1.31 | 20.03±1.14 | 19.45±1.82 | 21.22±1.79 | 1004.9±4.44 | 1005.5±4.55 | 2.00±0.79 | 2.08±0.84 | 43.96±2.03 | 41.22±4.93 | 31.50±0.91 | 27.37±4.67 | 2.63±1.04 | 3.09±1.03 |
| 7 | 21.66±0.48 | 21.66±0.54 | 22.18±1.45 | 23.80±1.64 | 1004.9±3.05 | 1005.7±3.06 | 2.71±1.23 | 3.30±1.54 | 45.66±0.95 | 38.00±4.12 | 31.89±0.61 | 24.85±4.09 | 3.65±1.67 | 4.59±2.20 |
| 8 | 25.06±1.09 | 24.36±1.02 | 23.74±1.33 | 24.92±1.23 | 1007.7±3.19 | 1008.3±3.26 | 2.97±0.88 | 2.99±1.38 | 47.03±1.31 | 39.67±3.61 | 30.53±0.62 | 26.06±3.04 | 3.88±1.19 | 4.01±1.65 |
| 9 | 24.46±0.67 | 24.35±0.32 | 21.05±1.10 | 22.00±1.61 | 1012.1±3.58 | 1012.5±3.71 | 2.78±0.50 | 2.97±1.11 | 47.31±0.60 | 40.47±4.26 | 31.14±0.14 | 27.79±2.08 | 3.54±0.73 | 3.86±1.28 |
| 10 | 19.97±1.63 | 21.10±1.30 | 16.64±1.61 | 17.74±1.40 | 1015.8±3.99 | 1015.6±4.39 | 2.97±0.90 | 3.67±1.43 | 43.58±1.03 | 36.04±8.32 | 31.51±0.45 | 28.64±3.47 | 3.93±1.34 | 4.62±1.74 |
| 11 | 13.80±2.52 | 15.72±2.07 | 9.21±4.35 | 10.98±4.27 | 1021.1±5.52 | 1020.9±5.31 | 3.39±1.37 | 3.98±2.13 | 38.67±2.15 | 36.81±5.40 | 32.13±0.27 | 31.60±1.63 | 4.74±2.08 | 5.16±2.58 |
| 12 | 8.45±1.74 | 10.39±1.60 | 3.54±3.90 | 5.11±3.96 | 1020.3±3.35 | 1019.7±3.51 | 3.70±1.23 | 4.31±1.50 | 34.31±1.43 | 34.47±2.68 | 32.54±0.17 | 32.09±0.59 | 5.28±2.01 | 5.56±2.01 |

Note : wtp = water temperature; tp = temperature; ap = atmospheric pressure; ws = wind speed; cd = conductivity; sa = salinity; gu = wind gust; B = Balpo area; C = Yeosu area.

TABLE V
CORRELATION ANALYSIS

| | wtp (°C) | tp (°C) | ap (hPa) | ws (m/s) | cd (mS/cm) | jw (cm) | sa (psu) | gu (°) | Wd (°) |
|---|---|---|---|---|---|---|---|---|---|
| wtp | 1 | | | | | | | | |
| tp | **0.91665** | 1 | | | | | | | |
| ap | **-0.63639** | **-0.73614** | 1 | | | | | | |
| ws | -0.12125 | -0.14116 | 0.05036 | 1 | | | | | |
| cd | **0.98328** | **0.91336** | **-0.63713** | -0.13365 | 1 | | | | |
| jw | 0.10679 | 0.09684 | -0.05920 | -0.03842 | 0.10177 | 1 | | | |
| sa | **-0.57335** | -0.47027 | 0.31902 | 0.00918 | -0.41586 | -0.07576 | 1 | | |
| gu | -0.17269 | -0.19029 | 0.06974 | **0.95125** | -0.18302 | -0.04444 | -0.03941 | 1 | |
| wd | -0.23717 | -0.21864 | 0.06461 | 0.23902 | -0.23054 | -0.05751 | -0.14136 | 0.28925 | 1 |

Note : wtp = water temperature; tp = temperature; ap = atmospheric pressure; ws = wind speed; cd = conductivity; jw = sea surface height; sa = salinity; gu = wind gust; wd = wind direction

In the correlation analysis given in Table 5, the water temperature exhibits positive relationships r = 0.92 and r = 0.98 with the temperature and conductivity and negative relationships r = -0.64 and r = -0.57 with the air pressure and salinity.

The temperature has a positive relationship r = 0.91 with the conductivity and negative relationship r = -0.74 with the air pressure. The air pressure exhibits a negative correlation r = -0.64 with the conductivity, while the wind direction exhibits a positive correlation r = 0.95 with the wind blast.

Table 6 lists the results of the factor analysis based on the data from Goheung Balpo area in 2009. The eigenvalue ≥ 1, while the number of factors is 2. Water temperature, air pressure, and conductivity are included in factor 1, whereas air pressure is inversely proportional to other variables. Factor 2 includes the wind speed and blast.

TABLE VI
FACTOR ANALYSIS FOR GOHEUNG BALPO DATA

| Variables | Factor 1 | Factor 2 |
|---|---|---|
| Water temperature | 0.95852 | 0.15739 |
| Temperature | 0.94598 | 0.13967 |
| Atmospheric pressure | -0.74344 | -0.22233 |
| Wind speed | -0.28201 | 0.92122 |
| Conductivity | 0.93510 | 0.13301 |
| Seasurface height | 0.14548 | -0.02601 |
| Salinity | -0.58391 | -0.19198 |
| Wind gust | -0.33143 | 0.91430 |
| Wind direction | -0.32830 | 0.35987 |
| Eigenvalue | 3.90 | 1.96 |
| Proportion | 0.43 | 0.22 |

In the factor analysis of Yeosu Odongdo Island, conductivity was not included in factor 1, in contrast to the case of Balpo. Therefore, the results of the correlation analysis showed that the conductivity has no correlation with temperature, water temperature, and air pressure. you submit your final version, after your paper has been accepted, print it in two-column format, including figures and tables.

### *C. ANOVA Analysis of Water Temperatures and Salinity of Three Areas*

The local daily means for the water temperature and salinity were obtained; these parameters were observed in all three areas for the gap analysis. ANOVA analysis was then conducted. In the ANOVA analysis given in Table 7, there are no differences between the water temperature and salinity in the case of the three areas, and the significance level was < 5%. The salinity for Balpo was the highest, while that for Yeosu Odongdo Island was the lowest. The low salinity for Yeosu Odongdo Island is because Yeosu Odongdo Island is adjacent to Yeosu downtown, and inflows including domestic sewage are high.

TABLE VII
ANOVA OF THREE SITES

| Dependent variable | Type-III sum of squares | Degrees of freedom | Mean square | F | p-value |
|---|---|---|---|---|---|
| Water temperature | 57.7084 | 2 | 28.8542 | 0.69 | 0.503 |
| Salinity | 3226.1824 | 2 | 1613.0912 | 185.39 | <.0001 |

Figures 2 and 3 show the trends for the changes in the water temperature and salinity (2009) for the three areas, respectively. The separator (line) in each graph indicates the starting date of each month.

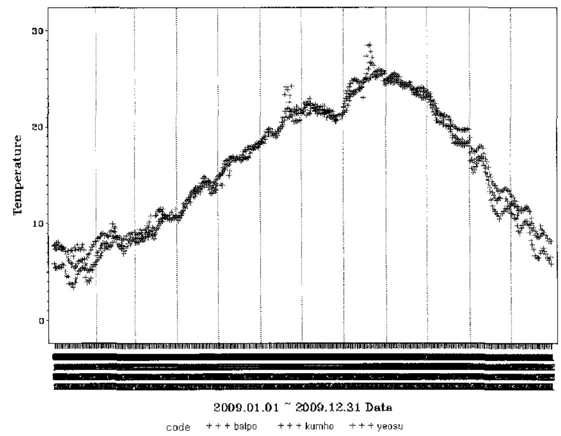



Fig. 2. Temperature variations in the three areas.

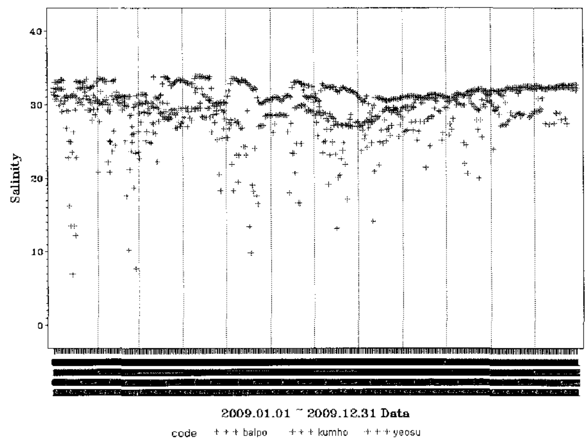



Fig. 3. Salinity variations in the three areas.

There is no difference in the water temperature between the three areas but salinity is different.

### *D. Regression Model Equation for Water Quality of Fish Farm*

The results of the factor analysis and correlation analysis obtained using the Goheung Balpo data, which have been described in section B, were referred to predict the water temperature of the Keumho fish farm located close to Goheung Balpo area. Consequently, the water temperature, wind speed, and salinity were considered to be independent variables, and the multiple regression analysis was performed.

The p-value of the F-test statistic was smaller than the significance level of 0.05; from this, it can be concluded that either water temperature, salinity, or wind speed has an effect on the water temperature of the fish farm. However, in the t-test for the significance test of the coefficient of regression, the exposition variables noted for the significance level of 0.05 were water temperature and salinity. Therefore, the estimation results, excluding the wind speed, are given in Table 8.

The conformance of the regression model was determined in the form of the F-test results and coefficient of determination R2. F = 11,167.8 (p < .0 0 01) was small and the regression equation was found to be statistically significant with a significance level of 5%. Further, the regression model wherein the coefficient of determination is close to 1 illustrates the most of the total variance. The T test result for the coefficient of regression was statistically significant, and there was no multicollinearity effect since the variance inflation factor (VIF) value is close to 1. The diagnosis of the multicollinearity leads to the result that if the VIF value is 10 or greater, the exposition variable is generated with the multicollinearity problem[9][10].

TABLE VIII
MULTIPLE REGRESSION ANALYSIS FOR WATER TEMPERATURE ESTIMATION

| Item | Ustd. coefficient | | Std. coefficient | t | Significance probability | VIF |
|---|---|---|---|---|---|---|
| | β | Std. Error | Beta | | | |
| Constant | -4.402 | 1.667 | | -2.64 | <.0088 | 0 |
| Water temperature at Balpo | 0.906 | 0.007 | 1.014 | 121.75 | <.0001 | 1.5667 |
| Salinity at Balpo | 0.198 | 0.050 | 0.033 | 3.97 | <.0001 | 1.5667 |
| $R^2$=.988 Adjusted $R^2$=.988 F(p)=11167.8(<.0001) | | | | | | |

Note : VIF = Variance Inflation Factor; Ustd. = Unstandardization; Std = Standardization

In addition, the standardized regression coefficient value of the water temperature is 1.014. We know that the water temperature constitutes most of the total regression model. The prediction model (Eq. (1)) used in this study is as follows:

$$A_t = -4.402 + 0.906 * B_t + 0.198 * B_s \quad (1)$$

Here $A_t$ is the water temperature of the fish farm; $B_t$, the water temperature at Balpo; and $B_s$, salinity at Balpo. Table 10 lists the regression model equations for the dissolved oxygen and salinity.

TABLE IX
MULTIPLE REGRESSION MODELS (N = 285)

| Multiple linear regression Model | $R^2$ | Sig |
|---|---|---|
| pH = 5.368-0.024*sa+0.003*ap+0.003*wtp | 0.06 | ** |
| Sal = 30.151-0.044*tp+0.004*sa-0.056*gu | 0.07 | ** |
| DO = 15.833-0.209*wtp-0.142*sa | 0.89 | ** |

Note : sa = salinity; wtp = water temperature; tp = temperature; gu = wind gust;
** significance at the 90% confidence level

The coefficients of determination of the pH and salinity are very low and the regression model equation is not appropriate. For the coefficient of determination of 89% level, the regression equation for the dissolved-oxygen estimation has a linear regression relationship, as given by Eq. (2), where is the dissolved oxygen of the fish farm, the water temperature at Balpo and salinity at Balpo.

$$A_d = 15.833 - 0.209 * B_t - 0.142 * B_s \quad (2)$$

In Eq. (2), most of the dissolved-oxygen forecasting model comprises the water temperature having negative correlation, followed by salinity having negative correlation. However, the estimation of pH and salinity was difficult from the provided oceanic environment nine factors because the adjusted coefficient of determination is very low, less than 10%.

*E. Comparison of Predicted and Measured Values*

We compared the predicted and measured values using the regression equation for the optimum regression model described in section D. On the basis of the dataset 5 (set 5) given in section B of **II**, reasonableness check of the predicted and measured values was performed.

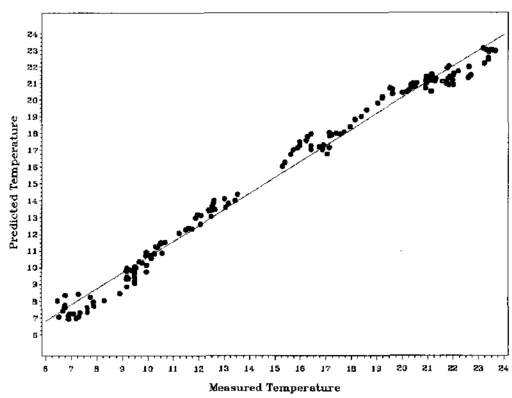



Fig. 4. Predicted vs. measured values for water temperature.

Since the values of the predicted and measured water temperatures fit the regression line shown in Fig. 4 well, the statistical model is appropriate.

In the regression line graph for the predicted and measured values of the dissolved oxygen, shown in Fig. 5, the points are further from the regression line than in the case of the water temperature in Fig. 4. However, this model still yields values that are reasonably close to the regression line and is appropriate for the dissolved oxygen as well.

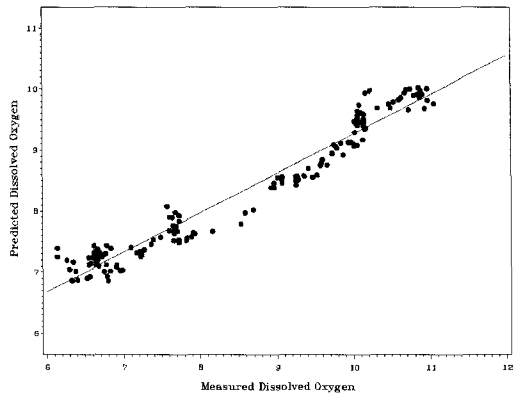



Fig. 5. Predicted vs. measured values for dissolved oxygen.

*F. Environmental Information Monitoring System*

From section E, the predicted values of the water temperature and dissolved-oxygen value were close to the experimental values. Therefore, we have shown that the water-quality information in marine farming areas can be monitored on the basis of oceanic environment data obtained from nearby authorized organizations. Since this technique does not involve and is independent of sensors, sensor costs are saved. Moreover, other equipment costs such as fish farm costs, maintenance costs of the sensor, software costs, and purchases of monitoring equipment that stores and transmits data and also displays it.

We have also shown that in our method, if the statistical model equation is well fitted using trustworthy oceanic environment data from an authorized organization, it can be considered to be different and better than the existing monitoring methods. Moreover, it becomes possible to easily achieve environmental information monitoring of marine farming areas using modern computers with internet and communication devices such as cell phones.

## IV. CONCLUSIONS

There exists an enormous burden on small fish farms for setting up sensors in water tanks directly and to manage it. In order to resolve this problem, we have attempted to establish a sequence of tasks for estimating the water-quality environment information of fish farms using information obtained from oceanic environment observatories that is operated and managed in national institutions located close to the fish farm.

In the estimated regression model, the adjusted coefficients of determination of the pH, salinity, and chlorophyll were very low and they did not fit well with the measured values. ANOVA analysis of the water temperature and salinity for the three areas of Goheung Balpo, Yeosu Odongdo, and experimental fish farm, Keumho did not indicate difference in the water temperature; however, the salinity exhibited differences under 5% significance level. In particular, Balpo had highest salinity, while Yeosu Odongdo Island, which is adjacent to Yeosu downtown and has tremendous inflows including domestic sewage, had lowest.

Multiple regression analysis was performed with nine water quality factors of the Goheung Balpo area located close to the Keumho fish farm. The estimation of water temperature and dissolved oxygen of the fish farm had multiple regression linear relationship with the coefficient of determination of 89% and 95% level, respectively. However, the adjusted coefficient of determination for pH and salinity was very low (<10%) and therefore not appropriate. Moreover, the estimated models of water temperature and dissolved oxygen were also shown to be useful when the predicted and measured values were close to the regression line.

In this manner, if we use a well fitted statistical model equation, we can manage sensor and data with low costs and burden on small-scale fish farms, in contrast to the current scenario with the existing environmental information monitoring systems.

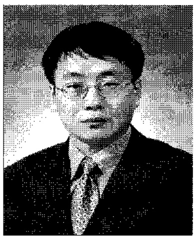
## ACKNOWLEDGMENT

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation.

## REFERENCES

[1] National Fisheries Research and Development Institute, 2009, A Study on the standardization of Aquaculture Industry
[2] Kim, Ceong, Han, Kang and Jang, 2007. Aqua Data Sensing System for Eco-friendly Aquaculture Products. 2nd International Conference on Ubiquitous Information
[3] Ceong, Park and Han, 2007. IT Convergence Application System for Eco Aquafarm. 07 Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies
[4] Korea Hydrographic and Oceanographic, 2009. Measurement data for Balpo and Odongdo Coastal http://http://www.khoa.go.kr/
[5] Kim and Ceong, 2010. Statistical Analysis of Water Quality in a Land-based Fish Farm. The Korea Institute of Electronic Communication Sciences 5(6)
[6] Liu, C.W., Lin, K.H.,Kuo,Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Tot. Environ. 313, 77—89
[7] Singh, Malik and Mohan, 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River(India) - a case study, Water Research 38 : 3980-3992
[8] Hair et al., 1998. Multivariate data analysis. Prentice Hall, New Jersey, USA.

[9] Kim, Koh, Ko and Yeo, 2008. Prediction of Nitrate Contamination of Groundwater in the Northern Nonsan area Using Multiple Regression Analysis. Korea. Geoscience and Mineral Res.13(5)
[10] S. Shrestha, F. Kazama, 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environment Modelling & Software 22, 464-475



**Hee-Taek Ceong** received the B.S., M.S., and Ph.D. degrees in Computer Science from Chonnam National University, Kwangju, Korea in 1992, 1995, and 1999, respectively. Since 1999, he has been an associate professor in Department of Digital Convergence, Chonnam University, Yeosu, Korea. His research interests include the data analysis, RFID/USN system, and Multimedia.

**Hae-Ran Kim** received the B.S., M.S degrees in Computer Science and Statistics from Chonnam National University, Kwangju, Korea in 1993 and 1996. She was a researcher in Hyundai motor company's Namyang R&D center, Korea, from 1996 to 2000. She is a Ph. D. candidate in Dept. of digital convergence at Chonnam National University, Yeosu, Korea. Her research interests include web application, data analysis and RFID/USN system.