REAL TIME FORECASTING AND AUTOMATIC SPECIES CLASSIFICATION OF HARMFUL ALGAL BLOOMS (HAB) FOR FISHERIES MANAGEMENT

BY JOSEPH H. W. LEE, J. H. GUO, TREE S. N. CHAN, DAVID K. W. CHOI, W. P. WANG AND KENNETH M. Y. LEUNG

Fish is an important source of animal protein in the diet of the Asian population and 60 percent of this is from aquaculture. Asia contributes about 90 per cent of the global aquaculture production and has become the most important supplier to the global seafood trade [7]. It is expected that population growth and economic development will lead to increasing fish consumption and global demand for food fish. In Hong Kong, marine fish culture (mariculture) has been a major supplier of high value fish including groupers, snappers and sea breams. Local mariculture is carried out in cages suspended by floating fish farm rafts in designated fish culture zones (FCZ) which are typically weakly-flushed tidal inlets.

In subtropical eutrophic coastal waters around Hong Kong and the region, the explosive growth of phytoplankton (algal blooms) is often observed. These blooms can lead to water discoloration (e.g. red tides), severe dissolved oxygen depletion, and shellfish poisoning – resulting in beach closures and massive fish kills [1]. For example, in April 1998, a devastating red tide resulted in the worst fish kill in Hong Kong's history - over 80% (3,400 tonnes) of fish stocks in Hong Kong were wiped out, with an estimated loss of over USD 40 million. Despite significant upgrades of the water pollution control infrastructure over the past two decades, massive harmful algal blooms (HAB) still recur and present formidable challenges to fisheries management (Figure 1). Worldwide, HAB is an important problem related to the global challenges of water and food security. The onset of a HAB is also notoriously difficult to predict.

Figure 1. Typical marine fish culture zone located in a coastal tidal inlet and examples of coastal algal blooms and fish kills.
(a) Examples of coastal algal blooms.

(b) Typical marine fish culture zone and massive fish kill in April 1998.

Traditional approaches of red tide monitoring and fisheries management rely on field sampling and laboratory analysis of chlorophyll-a concentration (Chl-a) - an indicator of algal biomass - and manual cell counting and species identification, which are resources intensive and time consuming. With the increasing availability of real time water quality sensors, the development of HAB early warning systems has become a practical possibility. In this article, an overview of recent research on the use of remotely sensed data in a HAB early warning system is described. Two aspects of the system are presented: (i) daily forecast of algal bloom risk based on prediction of vertical density gradients using *in-situ* real time (10 min sampling period) water quality data; and (ii) use of machine learning to automatically detect target HAB species from images (30,000 numbers/hour) monitored by a submerged Imaging Flow Cytometer at a marine fish farm. Further details can be found in the cited references.

Real time forecasting of algal blooms using real time water quality data

The occurrence of HABs in eutrophic coastal waters depends on the complex interaction of physical and biological factors that include: nutrient supply (e.g. inorganic nitrogen and phosphorus), algal growth rate, hydro-meteorological conditions (e.g. solar radiation,

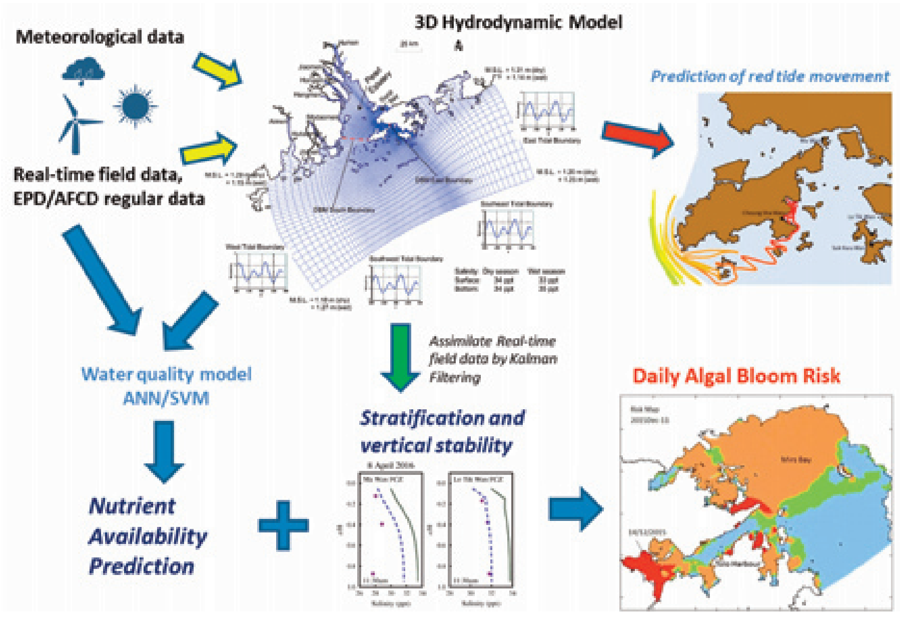

Figure 2. Conceptual framework for a harmful algal bloom (HAB) early warning system for prognostic forecast of algal bloom.

rainfall, air and water temperature, wind), tidal currents, water column transparency (light extinction) and turbulent mixing which is strongly affected by density stratification. The impacts of HAB on water quality also depend on algal and dissolved oxygen dynamics, and nutrient recycling. An early warning system of HAB occurrence (even with a lead time of 1-2 days) can benefit fisheries management and emergency response greatly.

Building on field observations of algal blooms, the use of data-driven methods such as Artificial Neural Networks (ANN) to predict coastal algal blooms has been attempted [2], [9]. However, the measurement frequency (typically monthly or biweekly) of most water quality monitoring protocols was insufficient to capture the highly dynamic variation of hydrodynamics and water quality, and in particular algal biomass. In recent years, HAB early warning systems have increasingly been reported [5], [6], [10]. Nevertheless, the development of field validated HAB forecast systems remains a formidable challenge.

Recently we have developed a daily algal bloom risk forecast system based on: (i) a vertical stability theory; and (ii) a data-driven artificial neural network (ANN) model that assimilates high frequency data to predict sea surface temperature (SST) and vertical density stratification on a daily basis. The model does not rely on past chlorophyll measurements and has been validated against extensive field data.

Field observations show that a stable water column is necessary for an algal bloom to form. In weakly flushed tidal inlets, it can be shown that the vertical turbulent diffusivity, E, must be less than a turbulence threshold defined by the net algal growth rate and the euphotic depth – with $E<E_c=4\mu l^2/\pi^2$, where $\mu=$ net algal growth rate and $l=$euphotic depth (proportional to Secchi depth) [13], [14]. If the vertical mixing exceeds the critical turbulence threshold, too much algae will be mixed out of the photic zone into the non-productive lower layer, and a bloom cannot be formed. The vertical stability criterion has been verified against 191 algal blooms over the past three decades [8].

In addition to the water column stability condition, a nutrient threshold, i.e. total inorganic nitrogen > 120 μg/L and orthophosphate > 18 μg/L should be met. If both the stability and nutrient criteria are fulfilled, there is no restriction for the algal population to grow in either physical or biological aspects

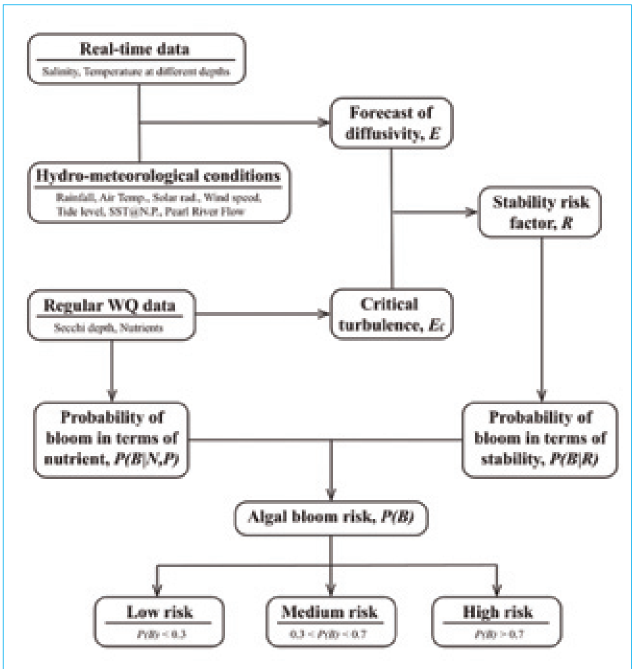

Figure 3. Daily algal bloom risk forecast framework as a function of hydro-meteorological and water quality data expressed in terms of a hydrodynamic stability risk factor and nutrient availability for the Yim Tin Tsai (YTT) Fish Culture Zone, Tolo Harbour, Hong Kong.

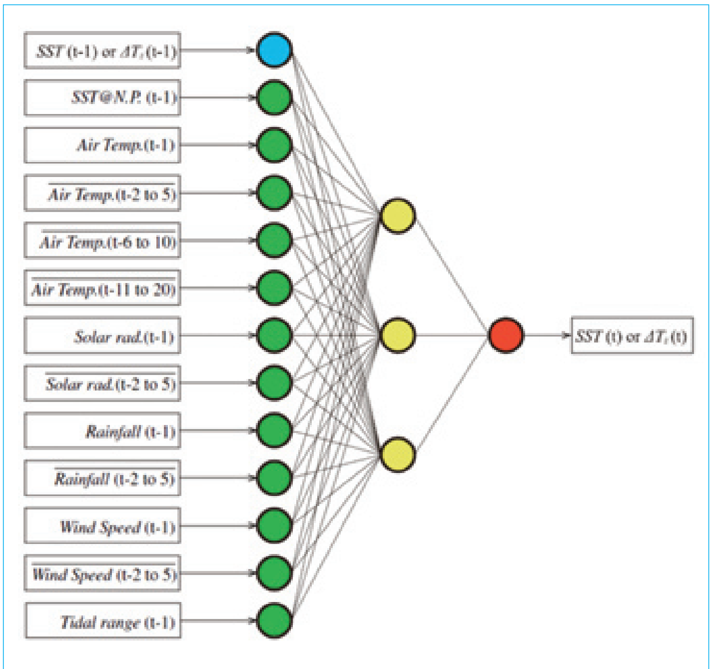

Figure 4. Artificial Neural Network (ANN) for daily prediction of sea surface temperature (SST) and vertical temperature difference (ΔT_z); the uppermost neuron in the input layer shows the most current real time measurement (when data is available). Time averages over several days indicated by over bar.

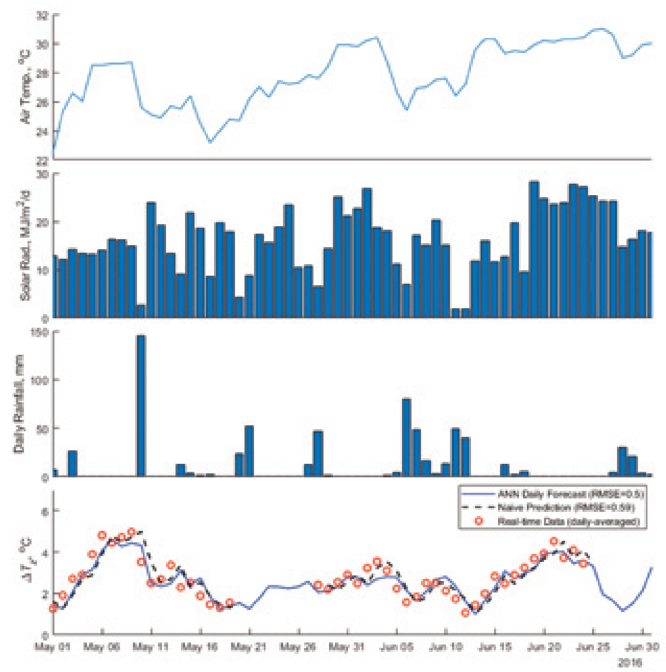

Figure 5. Example daily forecast of vertical temperature differential ΔT_z using hybrid ANN model, compared with daily-averaged real-time data and naive prediction given by data on the previous day. Note that the ANN daily forecast is continuous while naive prediction is limited by gaps of real-time data.

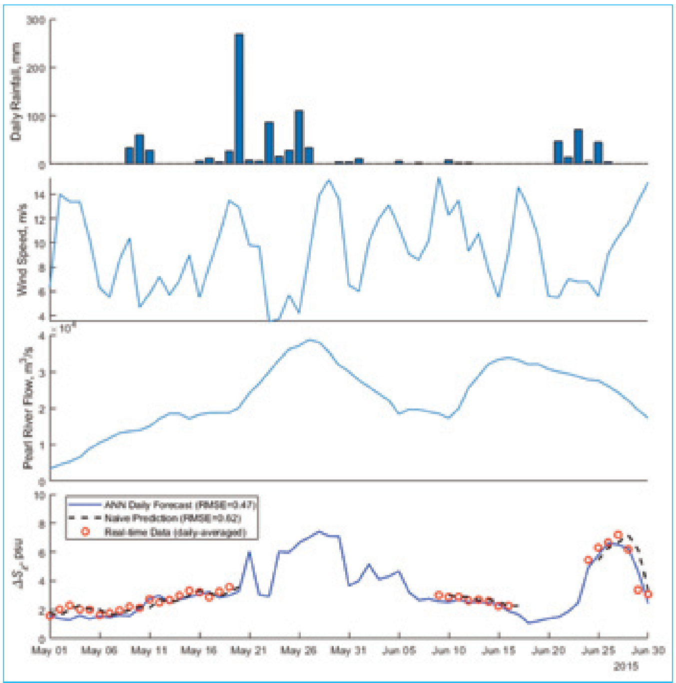

Figure 6. Example daily forecast of vertical salinity differential ΔS_z using hybrid ANN model, compared with daily-averaged real-time data and naive prediction given by data on the previous day. Note that the ANN daily forecast is continuous while naive prediction is limited by gaps of real-time data.

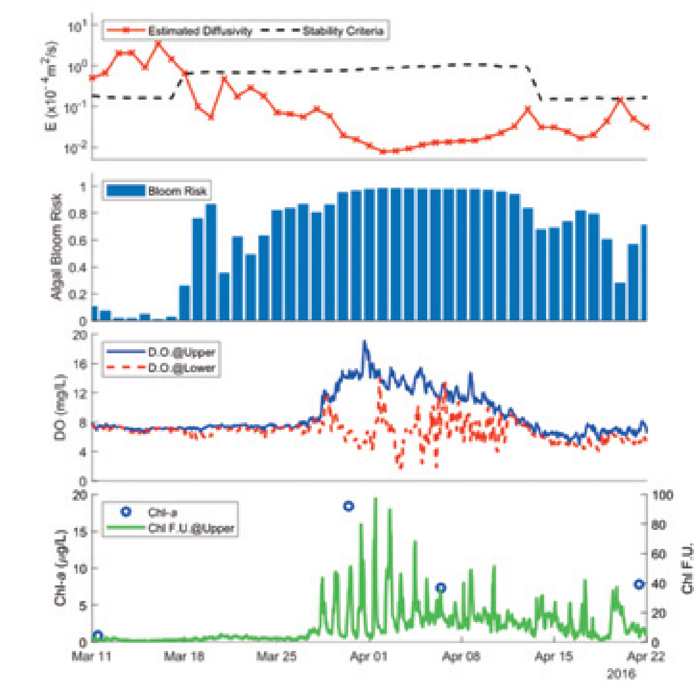

Figure 7. Example daily forecast of vertical turbulent diffusivity and bloom risk compared with measured surface and bottom dissolved oxygen and chlorophyll fluorescence for a dinoflagellate bloom observed at YTT FCZ in Mar-Apr 2016 (causative species: *Akashiwo sanguinea*; cell count:1,000-10,000 cells/mL).

and hence a bloom is likely to occur. Based on long term data, the vertical stability criterion and the nutrient threshold can be cast into probabilistic or risk terms and combined to give a prognostic forecast of algal bloom risk (high, medium, low) levels. Figure 2 shows a conceptual framework of a possible data assimilation system based on the integration of 3D and data-driven models, and field data.

The availability of high-frequency real-time temperature, salinity, dissolved oxygen (DO) and chlorophyll fluorescence data (at 10-minute intervals) opens the possibility of forecasting algal bloom risks on daily basis. Real-time telemetry data monitoring stations have now been set up in 12 key fish culture zones in Hong Kong, with spatial distances ranging from 2.5 to 20 km. Figure 3 shows the flow chart of the implementation of the forecasting framework for the Yim Tin Tsai marine fish culture zone in Tolo Harbour, Hong Kong. The vertical temperature and salinity gradients (and hence the density gradient) can be forecast by assimilation of data and/or model predictions using an Artificial Neural Network (ANN). Figure 4 shows an ANN model with three layers (input, hidden and output layers) for daily forecast of SST and vertical temperature difference using inputs of daily averaged real-time data in the previous day together with past hydro-meteorological data. A similar network can be obtained for the vertical salinity difference. The tidally and wind-induced vertical diffusivity E can then be estimated (based on 3D hydrodynamic models and predicted density stratification) and compared with the critical turbulence criterion E_c to give a stability risk factor R. By analysing all historical algal bloom events, the likelihood of a bloom occurrence based on hydrodynamic stability can be cast in terms of a probabilistic risk, $P(B|R)$. Similarly, the likelihood of a bloom based on nutrient availability (i.e. concentration of total inorganic nitrogen and orthophosphate) can be obtained as $P(B|N)$ and $P(B|P)$. The algal bloom risk for the next day can then be obtained using the multiplication rule and the Liebig's Law of the Minimum: $P(B) = P(B|R).min[P(B|N),P(B|N)]$.[8]

Figure 5 and Figure 6 show respectively a daily forecast of vertical temperature and salinity differential (at two levels). Based on

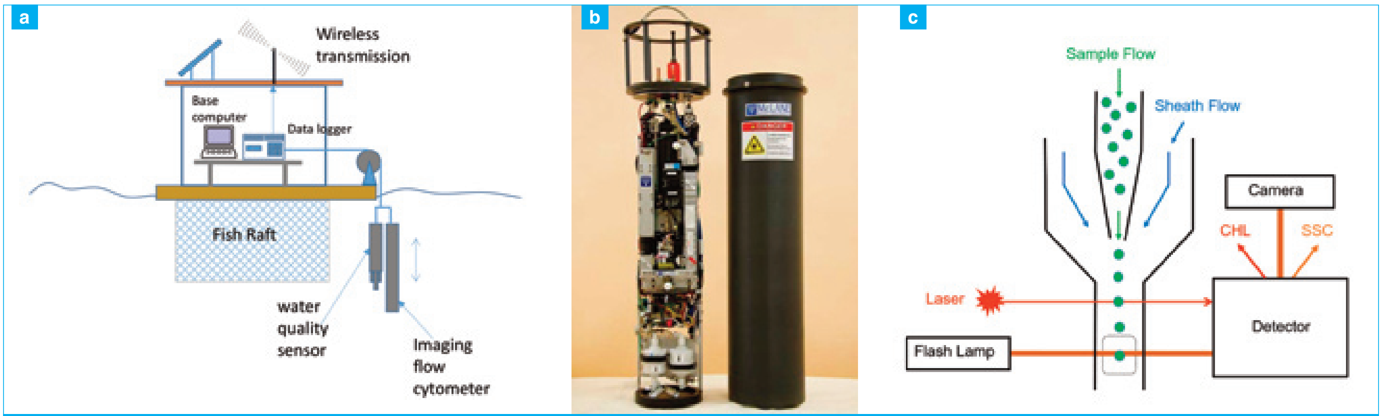

Figure 8. HAB species monitoring at fish culture zone using Imaging Flow CytoBot (IFCB). (a) Field deployment of IFCB at fish raft. (b) IFCB (c) Hydrodynamic focusing.

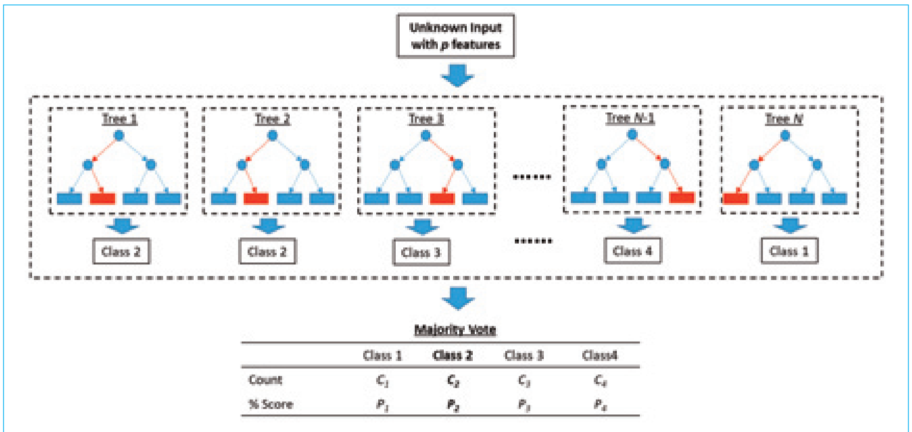

Figure 9. Classification using a random forest classifier (ensemble of decision trees trained with bootstrap sampling and random feature subspace methods). Extracted features of input image are presented to classifier and each tree makes a prediction independently. The number of instances that each class i being predicted are counted (C_i) and a percentage score is obtained (P_i). The final decision is the class with the maximum score (majority vote).

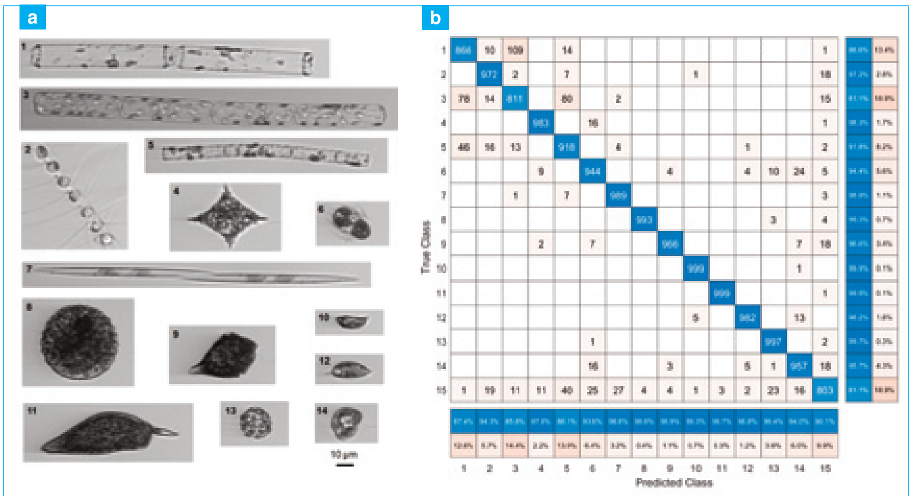

Figure 10. Automated classification of 14 target harmful algal bloom (HAB) species using machine learning. (a) Examples of IFCB images for 14 target HAB species. (b) Confusion matrix of classification result of test images. Numbers in blue boxes along the diagonal line indicates the correctly classified images. Class No. 15 refer to all species other than the 14 targets in (a).

the forecast, the vertical density gradient at the site can then be determined. The ANN model is a hybrid model that is capable of making short term forecasts even in the absence of *in-situ* data (e.g. due to data logger failure, system malfunctioning or equipment maintenance). The system has been validated against four years of field data, with an accuracy comparable to the field performance of commercially available systems (0.51 ˚C and 0.58 psu for the temperature and salinity, respectively). It should be noted that the model is clearly superior to the naïve prediction (prediction of today's conditions being same as yesterday). In practical deployment, the presence of real-time data gaps is the norm rather than the exception and it is essential to have a model that can perform short-term forecasts even in the absence of *in-situ* real-time data.

Figure 7 shows the variation of the estimated vertical diffusivity, algal bloom risk, DO and chlorophyll fluorescence in March-April 2016. It is seen that with the decrease in vertical diffusivity towards the end of March 2016, the bloom risk becomes steadily high ($P(B)>0.8$) around 26 March, and the stable water column resulted in an algal bloom which was sighted on 29 March, 2016. The onset of the dinoflagellate bloom was indicated by the sharp rise in chlorophyll fluorescence and was confirmed by direct onsite measurements which revealed the causative species to be *Akashiwo sanguinea* with cell counts of 1,000-10,000 counts/mL and chlorophyll-*a* > 10 μg/L. The photosynthetic production in the surface layer resulted in DO supersaturation (up to 16 mg/L) and a marked DO differential between surface and bottom of 4-10 mg/L. The bottom DO was depleted to a low level of around 4 mg/L during the bloom which subsided after about two weeks. The algal and DO dynamics is also associated with nitrogen and phosphorus uptake [8].

The vertical turbulence at the site is dominated by wind-induced mixing prior to the bloom which was coincident with a period of low wind (< 2 m/s), neap tide, high water transparency (large Secchi depth), and increasing temperature and vertical temperature (salinity) differentials of 4 ˚C (2 psu) respectively. The bloom occurrence is clearly correlated with the predicted algal bloom risks. As a bloom will occur if nutrients are sufficient, it is found that the bloom risk due to stability risk is often a good indicator of a bloom.

Automated classification of high-frequency microalgae images

High-frequency microalgae image data can be acquired *in-situ* through an imaging FlowCytobot (IFCB) that enables the identification of HAB species and estimation of cell abundance in real time. The IFCB is an automated, submersible equipment that can be continuously deployed underwater for months [11]. Designed using the principle of hydrodynamic focusing and flow cytometry, the IFCB is able to capture up to 30,000 high-resolution images (3.4 μm/pixel) in an hour (three 5 mL samples). The observation range is from 10 μm to 150 μm, which covers most of the common algal bloom species in Hong Kong. Analysis of image data at such a high sampling rate requires automated taxonomic classification using machine learning techniques [12].

Since March 2019 we have been deploying an IFCB at the Yim Tin Tsai (YTT) Fish Culture Zone in Tolo Harbour, Hong Kong, to collect algal image data and monitor algal species. The system is equipped with a 4G cellular network connection to facilitate remote equipment control and data transfer (Figure 8). To collect training samples for development of auto-classifier, we have performed manual annotation of over 330,000 images collected by IFCB during the deployment in YTT. These images cover 40 categories from species to group levels, including diatoms and dinoflagellates. Automated classification approach of IFCB images has been developed using both (i) random forest algorithm with robust image processing and feature selection techniques; and (ii) state-of-the-art transfer learning with a pre-trained Convolution Neural Network (CNN) (i.e. GoogLeNet). The random forest (RF) [3] is an efficient machine learning approach predicting the label of an unknown image based on extracted image features. As illustrated in Figure 9, an ensemble of decision trees trained with bootstrap sampling and feature bagging make predictions independently and the final decision is based on majority votes. Fourteen commonly observed HAB species of particular interest are selected as the training targets (Figure 10(a)). Both RF and CNN approaches reach classification accuracies of over 80% for all target species. Figure 10(b) shows the confusion matrix of classification results (using the RF approach) of 1,000 test images for each species. The columns of the confusion matrix represent the number of predictions in each class while its rows represent the actual observations in each class. Testing against unlabelled IFCB samples shows that our developed classification approach is very efficient with near real-time cell abundance estimation of prevailing species - results can be obtained within 1-7 minutes after a sample is acquired. This opens the possibility of adapting IFCB into a real-time HAB detection and early warning system.

Acknowledgement

The research reported herein has been supported at various times by the Hong Kong Research Grants Council (RGC). The current study is supported by a research project on "Pilot Study of Red Tide Early Warning System" from the Agriculture, Fisheries and Conservation Department of the Hong Kong SAR Government (AFCD/FIS/01/18). ■

References

[1] AFCD, 2016. "History of red tide/HAB in Hong Kong", Agriculture, Fisheries and Conservation Department, HKSAR (https://www.afcd.gov.hk/english/fisheries/hkredtide/redtide/red03.html)

[2] Blauw, A., Anderson, P., Estrada, M., Johansen, M., Laanemets, J., Peperzak, L., Purdie, D., Raine, R., & Vahtera, E. (2006). The use of fuzzy logic for data analysis and modelling of European harmful algal blooms: results of the HABES project. *African Journal of Marine Science*, 28, 365–369.

[3] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

[4] Cheng, K. H., Chan, S. N., & Lee, J. H. W. (2020). Remote sensing of coastal algal blooms using unmanned aerial vehicles (UAVs). *Marine Pollution Bulletin*, 152, 110889.

[5] Coad, P., Cathers, B., Ball, J. E., & Kadluczka, R. (2014). Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. *Environmental Modelling and Software*, 61, 393–409.

[6] Dabrowski, T., Lyons, K., Nolan, G., Berry, A., Cusack, C., & Silke, J. (2016). Harmful algal bloom forecast system for SW Ireland. Part I: Description and validation of an operational forecasting model. *Harmful Algae*, 53, 64–76.

[7] FAO (2016). Sustainable intensification of aquaculture in the Asia-Pacific region. Documentation of successful practices. Miao, W. and Lal, K.K. (Ed.), Bangkok, Thailand. Food and Agriculture Organization, United Nations.

[8] Guo, J. H. , Dong, Y.H. , & Lee, J. H. W. (2020). A real time data driven algal bloom risk forecast system for mariculture management. *Marine Pollution Bulletin*, 161

[9] Lee, J. H. W., Huang, Y., Dickman, M., & Jayawardena, A. W. (2003). Neural network modelling of coastal algal blooms. *Ecological Modelling*, 159, 179–201.

[10] Lee, J. H. W., Wong, K. T. M., & Choi, K. W. (2012). Forecasting and management of coastal water quality. In H. J. S. Fernando (Ed.), *Handbook of Environmental Fluid Dynamics, Volume One*, Chapter 7. (pp. 75–90). CRC Press/Taylor & Francis Group, LLC.

[11] Olson, R. J., & Sosik, H. M. (2007). A submersible imaging in flow instrument to analyze nano and microplankton: Imaging FlowCytobot. *Limnology and Oceanography: Methods*, 5(6), 195-203.

[12] Sosik, H. M., & Olson, R. J. (2007). Automated taxonomic classification of phytoplankton sampled with imaging in flow cytometry. *Limnology and Oceanography: Methods*, 5(6), 204-216.

[13] Wong, K. T. M., Lee, J.H.W., & Hodgkiss, I. J. (2007). A simple model for forecast of coastal algal blooms. Estuarine, *Coastal and Shelf Science*, 74(1-2), 175-196.

[14] Wong, K. T. M., Lee, J.H.W., & Harrison, P. J. (2009). Forecasting of environmental risk maps of coastal algal blooms. *Harmful Algae*, 8(3), 407-420.

Prof. Joseph H. W. Lee is Senior Advisor to the President and Senior Member of the Jockey Club Institute for Advanced Study at the Hong Kong University of Science and Technology (HKUST). He is an internationally recognized expert in water environment engineering and has been working on red tide and fisheries management problems since 1986 – with a passion to apply engineering systems approach to solve the complex water environment problems. He is the master mind of the Hong Kong WATERMAN coastal management and forecast system and the Principal Investigator of the ongoing group research project on HAB monitoring and prediction.

Mr J. H. Guo is a first-class honours graduate of the University of Hong Kong and a Chartered Civil Engineer and a member of the Institution of Civil Engineers (ICE) in the UK. After working in the construction industry for 7 years he is now pursuing a Ph.D. degree in environmental engineering in the Hong Kong University of Science and Technology - on application of real-time data and AI techniques in HAB early warning for coastal fisheries management.

Dr Tree S. N. Chan is a Research Assistant Professor in the Hong Kong University of Science and Technology. He worked as a research fellow in Singapore Nanyang Technological University with the Singapore-MIT Research and Technology Alliance (SMART) in 2014-15. Dr. Chan is an expert on the experimental and numerical modelling of fluid/hydrodynamics and water quality. His has been a key researcher in projects including beach water quality forecast, optimization of chlorine disinfection dosage for wastewater, and air-water interactions in urban drainage.

Dr David K. W. Choi is a Research Associate in the Department of Civil & Environmental Engineering at the Hong Kong University of Science and Technology and is specialist in hydraulic and water quality modelling. He has extensive experience in applied research involving the use of numerical models for a range of environmental problems including mariculture management, coastal and beach water quality, urban drainage hydraulics, and disinfection dosage control for wastewater treatment.

Prof. W. P. Wang is Chair Professor and former Head, Department of Computer Science at the University of Hong Kong (HKU). Professor Wang conducts research in computer graphics, computer vision, robotics, virtual reality, visualization, medical image analysis, and geometric modelling. He has made fundamental research contributions in collision detection, shape modelling and analysis, mesh generation, and architectural geometry. He is a key member of the research team in the development of the VISJET buoyant jet modelling software and the WATERMAN coastal water quality management system.

Prof. Kenneth M.Y. Leung is Chair Professor of Environmental Toxicology and Chemistry, Department of Chemistry and Director of State Key Laboratory of Marine Pollution (SKLMP) at City University of Hong Kong. Research themes of SKLMP include innovative technologies for monitoring and controlling of marine pollution, environment risk assessment of chemical contaminants and algal blooms, ecosystem responses to pollution or mitigation, and ecological restoration using eco-engineering. He is a member of the Red Tide/Harmful Algal Blooms Expert Advisory Group of the Hong Kong SAR Government.