FORECASTING HYPOXIA IN CORPUS CHRISTI BAY, TEXAS BY MODEL FUSION

Authors: Indu Chinta[1], Barbara S. Minsker[1]
[1]Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

Abstract
Hypoxic or dead zones, which occur when dissolved oxygen levels in water drop below 2 mg/L, are prevalent worldwide. An example of such an hypoxic zone forms intermittently in Corpus Christi Bay (CCBay), Texas, a USEPA-recognized estuary of national significance. Hypoxia in CCBay occurs primarily due to inflow of hypersaline waters that enter from adjacent bays and estuaries, fluctuations in oxygen levels due to the oxygen production-consumption cycle of the aquatic flora and fauna, seasonal fluctuations, and lastly, discharges from several wastewater treatment plants. The aim of this paper is to create more accurate and efficient near-real-time forecasts of hypoxia that will give researchers advance notice for manual sampling during hypoxic events. The approach involves developing and testing a suite of data-driven model fusion approaches with the help of cyberinfrastructure. The models that will be used as a part of this study are: a data-driven k-nearest neighbor model (Coopersmith, 2008), and a physics-based valve model (Sin Chit To, 2009). The k-nearest neighbor predicts the probability of occurrence of hypoxia 24 hours later, and the levels of dissolved oxygen. On the other hand, the valve model determines the distance traveled by a gravity-current upon entering Corpus Christi Bay from adjacent estuaries and bays and the period of time it persists. A comparison will be made between the results of the fused model and those of the individual models to test the effectiveness of model fusion in predicting the estuarine condition in the model.

Keywords: Hypoxia, Corpus Christi Bay, Model fusion, Cyberinfrastructure

1. Introduction

Hypoxia can be defined as the depletion of dissolved oxygen levels in aquatic environments to less than 2 mg/l (Dauer et al., 1992). However, benthic effects have been found to occur at dissolved oxygen levels of 3 mg/l (Ritter and Montagna, 1999). Hypoxia in Corpus Christi Bay was first observed and documented in 1988 (Montagna and Kalke, 1992), and its reoccurrence has been confirmed every following summer (Martin and Montagna, 1995; Ritter and Montagna, 1999; Applebaum et al., 2005).

Hypoxia is highly detrimental and stressful to benthic life, reducing biodiversity, relocating benthic organisms to the surface (Montagna and Ritter, 2006), and sometimes resulting in ‘inactivity and death’ (Tyson and Pearson, 1991; Diaz et al., 1992). Further, hypoxia in Corpus Christi Bay has implications for the agricultural economy of the Gulf of Mexico’s watershed (Donnelly and Scavia, 2007).

This phenomenon is most often the result of water pollution and eutrophication. In Corpus Christi Bay, it is a consequence of the following causes. First, seasonal periodicity of dissolved oxygen (DO) levels (Ritter and Montagna, 1999). DO (dissolved oxygen) levels have been found to be lowest in summers as the evaporation rates and temperatures are highest and precipitation is lowest in summer. Second, fluctuations due to the oxygen production-consumption cycle by the aquatic flora and fauna also play an important role (Goldshmid et al, 2004). The oxygen that is produced by the plants throughout the day is consumed by the fauna in the nights. This causes a minimum in DO levels immediately after sunrise and a maximum just after sunset. Third and most important cause is the inflow of hypersaline waters from adjacent bays such as the Oso Bay and Laguna Madre (Hodges and Furnans, 2007a). Hypersalinity in Laguna Madre is because inflow of freshwater is less than evaporation and also because Laguna Madre is separated from the Gulf of Mexico by a barrier island. The occurrence of this condition in Oso Bay is attributed to the Barney Davis Power Plant which draws a 400 MGD flow of cooling water from upper Laguna Madre and discharges it into Oso Bay (To, 2009). These hypersaline waters induce density stratification in the shallow water bay of Corpus Christi. The benthic layer is most influenced by this process and has a higher tendency of becoming hypoxic (Hodges and Furnans, 2007b). Lastly, several wastewater treatment plants discharge directly into Oso Bay or its tributaries. When the Oso bay overflows, the water in it flows into the Corpus Christi Bay and the latter is contaminated as well. The excessive nutrients from the wastewater treatment plant discharges have also been found to add to hypoxia (Coopersmith, 2008).

This study aims at creating near real-time forecasts of hypoxia by fusing existing models using data mining techniques. The paper is structured as follows. Section two is a review of the literature on model fusion strategies applied to environmental hydrology. Section three is a description of Corpus Christi Bay, which is the focus of this research. Section four discusses the methods that will be applied to fuse the existing models. Results will be presented at the conference.

2.Literature review

Model fusion started as early as 1969 with Bates and Granger combining forecasts of airline passenger data to “form a composite set of forecasts” that would lower the mean-square error compared to the individual forecasts. Some of the other earlier contributions in this area include Dickinson (1971) and Newbold and Granger (1974). Later, statistical techniques such as Bayesian model averaging were introduced by Bunn (1975, 1977) and Bordley (1982), which were pursued by Duan et al. (2007), Raftery et al. (1997), Vrugt and Robinson (2007), and more recently by Hsu et al. (2009) to predict “daily watershed streamflow”.

Data fusion strategies such as simple averaging, neural networks, fuzzy logic, M5 model trees, and instance based learning were illustrated for fusing flow forecasting models built over the River Ouse catchment in the United Kingdom (See, 2008). Prior to this, See and Abrahart, in 2001, tested four data fusion “experiments”- combining mean and median of four individual forecasts and two others using neural networks- and observed that the two neural networks fusion methods produced better results. Ajami et al. (2006) shed light on the use of simple averaging method, multimodel superensemble method, modified multimodel superensemble method, and the weighted averaging method for studying hydrological simulations. Simple averaging method, weighted averaging method, and neural network method were also employed for integrating rainfall-runoff models (Shamseldin and O'Connor, 1999, and, Shamseldin et al., 1997).

More closely related to this study, is the work of Coulibaly et al. (2005) wherein the authors combined "three dynamically different" hydrological models – a nearest neighbor model, conceptual model, and artificial neural network model- using the improved weighted average method. This type of combination provided an improved "4-day ahead prediction".

Apart from hydrology, model fusion has also been extensively applied in geophysics and meteorology (Chakraborthy et al., 2007) for combining climatic models. Dietrich et al. (2008) transformed meteorological ensemble forecasts into "discharge ensemble forecasts of rainfall-runoff models" to effectively predict flood situations. Furthermore, "multiple climate and hydrological models" were integrated for predicting uncertainty in streamflow using techniques like pooling and linear regression weighting (Block et al., 2009).

3.Area of study – Corpus Christi Bay

The United States Environmental Protection Agency (USEPA) has designated the Corpus Christi bay system as an estuary of national significance. Aransas, Corpus Christi and upper Laguna Madre in Texas comprise three of the seven estuaries of the system. The Gulf Intracoastal Waterway connects Corpus Christi Bay to Aransas Bay (north) and Laguna Madre (south) and the Mustang and North Padre Islands separate the entire system from Gulf of Mexico on the west. The Aransas pass allows exchange of water between the Corpus Christi Bay system and the Gulf of Mexico (Encyclopedia Britannica, 2010).

Corpus Christi Bay is located along the coast of Southern Texas stretching 25 miles (40 km) long and 3–10 miles (5–16 km) wide (Encyclopedia Britannica, 2010). It is a shallow water body with limited inflow freshwater with distinct climatic conditions (Montagna and Kalke, 1995). Its only sources of freshwater drainage are the Nueces River and Oso River.

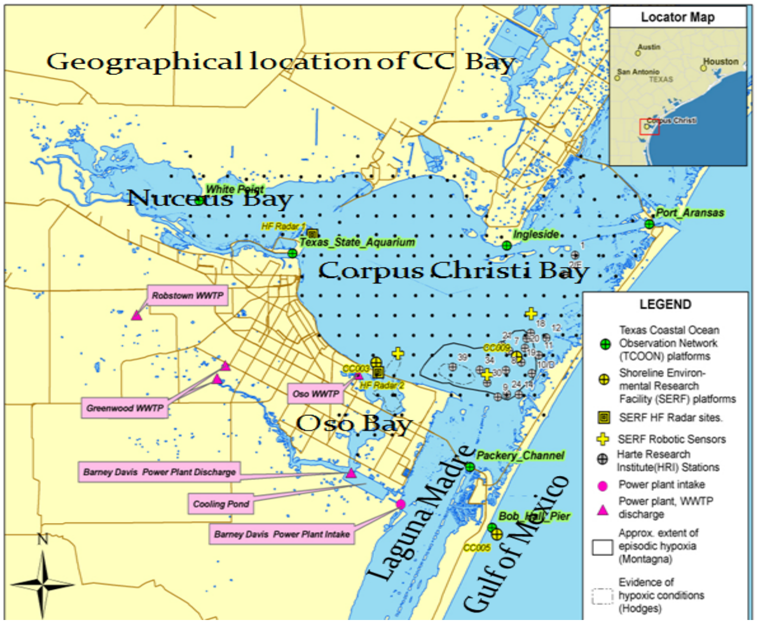

Figure 1. Study area – Corpus Christi Bay (Source: Sin Chit To, 2009)

4. Methodology

This study aims at creating a single robust model by fusing two existing models that predict hypoxia in Corpus Christi Bay using significantly different approaches – a data-driven k-nearest neighbor (knn) model (Coopersmith, 2008), and a physics-based hydrodynamic model or valve model (To, 2009). Data-driven models such as the knn model analyze input and output, and derive statistical relationships between the two that best fit the available data. The knn model derives a relationship between dissolved oxygen level in the bay and parameters such as salinity, temperature, wind, and time of day.

Physics-based models are those built on equations and fundamentals of physics, and describe the system behavior. The physical model in this case is a 2-D valve model. The valve model is based on the gravity current theory proposed by Dr. Ben Hodges and Paula Kulis of the University of Texas at Austin. According to this theory, gravity currents, i.e., the hypersaline waters from Oso Bay and Laguna Madre flowing into Corpus Christi Bay, flow towards the bottom of the bay, and prevent transfer of dissolved oxygen through the current owing to their higher density and greater energy requirement. Data for both models comes from sensor networks such as Texas Coastal Ocean Observation Network (TCOON), Harte Research Institute (HRI), Shoreline Environmental Research Facility (SERF), Texas Parks and Wildlife Department (TPWD).

The fusion methodology comprises the following steps. First, the training dataset for the data-driven model is split into several subsets for the purpose of quantifying uncertainty in the data, and the k-nearest neighbor model is trained on each of the subsets to form models known as local models or experts (Haykin, 1999). Splitting of training set is categorized into soft splitting of the training set and hard splitting of the training set (Solomatine and Ostfeld, 2008). Types of soft splitting include mixture of experts (Jordan and Jacobs, 1995), bagging or bootstrap aggregating (Breiman, 1996), and boosting (Freund and Schapire, 1997). Types of hard splitting include decision trees (Quinlan, 1986), regression trees (Breiman et al., 1984), M5 model trees (Quinlan, 1992). Here, both the knn and valve model are bootstrapped.

In the next step, both models predict the probability of hypoxia at the same grid of locations (with interpolation where common grids are not possible) to enable comparisons between the two models and fusion of the results. The fusion is then performed using simple averaging, regression trees, and neural networks to obtain a single output. A testing data set is then used to assess which model or fusion approach is most accurate.

5. Conclusion

Hypoxia in Corpus Christi Bay, which is partly naturally-occurring (stratification) and partly human-induced (discharges from wastewater treatment plants) is a severe threat to the living aquatic organisms in the bay.

We expect that model fusion will enable the strengths of multiple types of models, both statistical and physics-based, to be merged to produce forecasts of hypoxia that are more reliable and accurate. Results testing this hypothesis will be presented at the conference.

6. Works cited

Abrahart, R., See, L. 1999. Fusing multi-model hydrological data. *IEEE*, 2757-2762.

Ajami et al. 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *American Meteorological Society*, 755- 768.

Applebaum, S., Montagna, P.A., and Ritter, C., 2005. Status and trends of dissolved oxygen in Corpus Christi Bay, Texas, U.S.A., *Environmental Monitoring Assessment,* Vol. 107, pp. 297-311.

Bates, J. M., and Granger, C.W.J. 1969. The combination of forecasts. *Operational Research Quarterly*, Vol. 20, No.4, 451-468.

Block et al. 2009. A streamflow forecasting framework using multiple climate and hydrological models. *Journal of the American Water Resources Association*, Vol. 45, No. 4, 828-843.

Bordley, Robert F. The combination of forecasts: a Bayesian approach. Journal of the Operational Research Society, Vol 33, No. 2, 171-174.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984. *Classification and Regression Trees*. Wadsworth International, Belmont.

Breiman, L., 1996. Bagging predators. *Machine Learning*, Vol. 24, No. 2, pp. 123-140.

Chakraborty et al. 2007. Prediction of the diurnal cycle using a multimodel superensemble. Part II: Clouds. *American Meteorological Society*, Vol. 135, 4097-4116.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5 (1989) 559-583.

Coopersmith, E., 2008. Understanding and forecasting hypoxia using machine learning algorithms. MS diss., University of Illinois at Urbana-Champaign.

Coulibaly et al. 2005. Improving daily reservoir inflow forecasts with model combination. *Journal of Hydrologic Engineering*, 91-99.

Dauer, D.M., Rodi Jr, A.J., and Ranasinghe, J.A., 1992. effects of low dissolved oxygen events on the macrobenthos of the lower Chesapeake Bay. *Estuaries*, Vol. 15, pp. 384-391.

Diaz, R.J., Neubauer, R.J., Schaffner, L.C., Pihl, L., and Baden, S.P., 1992. Continuous monitoring of dissolved oxygen in an estuary experiencing periodic hypoxia and the effect of hypoxia on macrobenthos and fish. *Science of the Total Environment*, pp. 1055-1068.

Dickinson, J.P. 1973. Some statistical results in the combination of forecast. *Operational Research Quarterly*, Vol. 24, No. 2, 253-260.

Dietrich et al. 2008. Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event. *Nonlinear Processes in Geophysics*, 15, 275-286.

Donnelly, K., and Scavia, D., 2007. Reassessing hypoxia forecasts for the Gulf of Mexico. *Environmental Science & Technology*, Vol. 41, pp. 8111-8117.

Duan et al. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources* 30 (2007) 1371-1386.

Encyclopedia Britannica, 2010. Corpus Christi Bay. *Encyclopedia Britannica Online*. <http://www.britannica.com/EBchecked/topic/138506/Corpus-Christi-Bay>.

Freund, Y., and Schapire, R., 1997. A decision-theoritic generalization of on-line learning and an application of boosting. *Journal of Computer and System Sciences*, Vol. 55, pp. 119-139.

Georgakakos et al. 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology* 298 (2004) 222-241.

Goldshmid, R., Holzman, R., Weihs, D., and Genin, A., 204. Aeration of corals by sleep-swimming fish. *Limnology & Oceanography*, Vol. 49, No. 5, pp. 1832-1839.

Haykin, S., 1999. Neural networks: a comprehensive foundation. McMillan, New York.

Hodges, B., and Furnans, J., 2007. Thin-layer gravity currents in a shallow estuary. Proceedings of 18th Engineering Mechanics Division Conference (EMD2007) June 3-6, Blacksburg, VA, USA. (a)

Hodges, B., and Furnans, J., 2007b. Linkages between hypoxia and thin-layer stratification in Corpus Christi Bay. Manuscript in revision for Environmental Fluid Mechanics (July, 207). (b)

Jordan, M.I., and Jacobs, R.A., 1995. Modular and hierarchical learning system. In: Arbib, M (Ed.), *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, M.A.

Martin, C.M., and Montagna, P.A., 1995. Environmental assessment of LaQuinta Channel, Corpus Christi Bay, Texas. *Texas Journal of Science*, Vol. 47, pp. 203-222.

Montagna, P.A., and Kalke, R.D., 1992. The effect of freshwater inflow on meiofaunal and macrofaunal populations in the Guadalupe and Nueces Estuaries, Texas. *Estuaries*, Vol. 15, pp. 307-326.

Montagna, P.A., and Kalke, R.D., 1995. Ecology of infaunal Mollusca in south Texas estuaries. *American Malacological Bulletin*, Vol. 11, No. 2, pp. 163-175.

Montagna, P., and Ritter, C., 2006. Direct and indirect effects of hypoxia on benthos in Corpus Christi Bay, Texas, U.S.A. *Journal of Experimental Marine Biology and Ecology*, Vol. 330, pp. 119-131.

Montagna, P.A., and Froeschke, J., 2009. Long-term biological effects of coastal hypoxia in Corpus Christi Bay, Texas, USA. *Journal of Experimental Marine Biology and Ecology*, Vol. 381, pp. S21-S30.

Parasraman, K., Elshorbagy, A. 2008. Toward improving the reliability of hydrologic prediction: Model structure uncertainty and its quantification using ensemble-based genetic programming framework. *Water Resources Research*. Vol 44, W12406.

Quinlan, J.R., 1986. Induction of decision trees. *Machine Learning*, Vol. 1, pp. 81-106.

Quinlan, J.R., 1992. Learning with continuous classes. In: Adams, A., and Sterling, L. (Eds.). *Proceedings of AI'92, 5th Australian Joint Conference on Artificial Intelligence*. World Scientific, Singapore, pp. 343-348.

Ritter, C., and Montagna, P.A., 1999. Seasonal hypoxia and models of benthic response in a Texas bay. *Estuaries*, Vol. 22, pp. 7-20.

See, L. 2008. Data fusion methods for integrating data-driven hydrological models. *Studies in Computational Intelligence* (SCI) 79, 1-18.

See, L., Abrahart, R. 2001. Multi-model data fusion for hydrological forecasting *Computers & Geosciences* 27 (2001) 987-994.

Shamseldin, Assad Y., and O'Connor, Kieran M. 1999. A real-time combination method for the outputs of different rainfall-runoff models. *Hydrological Sciences-Journal*, Vol. 44, No. 6, 895-912.

Shamseldin, Assad Y., O'Connor, Kieran M., and Liang, G.C. 1997. Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, Vol. 197, 203-229.

Solomatine, D.P., and Ostfeld, A., 2008. Data-driven modeling: some past experiences and new approaches. *Journal of Hydroinformatics*, Vol. 10, No. 1, pp. 3-22.

Srinivasulu, S., Jain, A. 2009. River flow prediction using an integrated approach. *Journal of Hydrologic Engineering*, Vol. 14, No. 1, 75-83.

Thompson, Philip D. 1976. How to improve accuracy by combining independent forecasts. Monthly Weather Review, Vol. 105, 228-229.

To, E.S.C, 2009. Hypoxia modeling in Corpus Christi Bay using a hydrologic information system. PhD diss., University of Texas at Austin.

Tyson, R.V., and Pearson, T.H., 1991. Modern and ancient continental shelf anoxia: an overview. In: Tyson, R.V., and Pearson, T.H. (Eds.), *Modern and Ancient Continental Shelf Anoxia*, Geological Society of London, Vol. 58, pp. 1-24. Special Publication.

USEPA, 1999. ecological condition of estuaries in the Gulf of Mexico. EPA 620-R-98-004. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, Gulf Breeze, Florida, pp. 80.

Vrugt, Jasper A., and Robinson, Bruce A. treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, Vol. 43, W01411, doi:10.1029/2005WR004838.