Procedia Computer Science

Volume 29, 2014, Pages 1236–1245

ICCS 2014. 14th International Conference on Computational Science

A Dynamic Data-driven Decision Support for Aquaculture Farm Closure

Md. Sumon Shahriar and John McCulluch

Computational intelligence, Autonomous systems,
CSIRO Computational Informatics (CCI), CSIRO, Castray Esplanade, Hobart, 7001, Australia
mdsumon.shahriar@csiro.au,john.mcculloch@csiro.au

Abstract

We present a dynamic data-driven decision support for aquaculture farm closure. In decision support, we use machine learning techniques in predicting closures of a shellfish farm. As environmental time series are used in closure, we propose two approaches using time series and machine learning for closure prediction. In one approach, we consider time series prediction and then using expert rules to predict closure. In other approach, we use time series classification for closure prediction. Both approaches exploit a dynamic data-driven technique where prediction models are updated with the update of new data to predict closure decisions. Experimental results at a case study shellfish farm validate the applicability of the proposed method in aquaculture decision support.

Keywords: Aquaculture decision support, Machine learning, Dynamic data-driven decision support

1 Introduction

Data-driven approaches are widely used in many decision support systems [1]. Most of the decision support systems consider historical data for analysis and decision making. Decisions are usually made in a *static* way meaning that there is not much update in data and hence in the data-driven models used in the decision support systems. On the contrary, in *dynamic* data-driven approaches for decision support systems [2], data and associated models are updated to provide decisions.

Although, there exist decision support systems in aquaculture domain [3], the use of *dynamic* approach where models are updated and decisions are predicted based on the updated data fitted to the models is limited to the best of our knowledge. The current decision support system for Tasmanian Shellfish Quality Assurance Program (TSQAP) provides status of the closure based on the expert rules applied to the observed environmental sensor data such as rainfall, river flow and salinity. These environmental sensor data provide proxy for contaminants in the water. A snapshot of the current system is shown in Figure 1. In the system, the current closure status or decisions of the shellfish farms which are made manually by relevant authorities are shown.

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.111

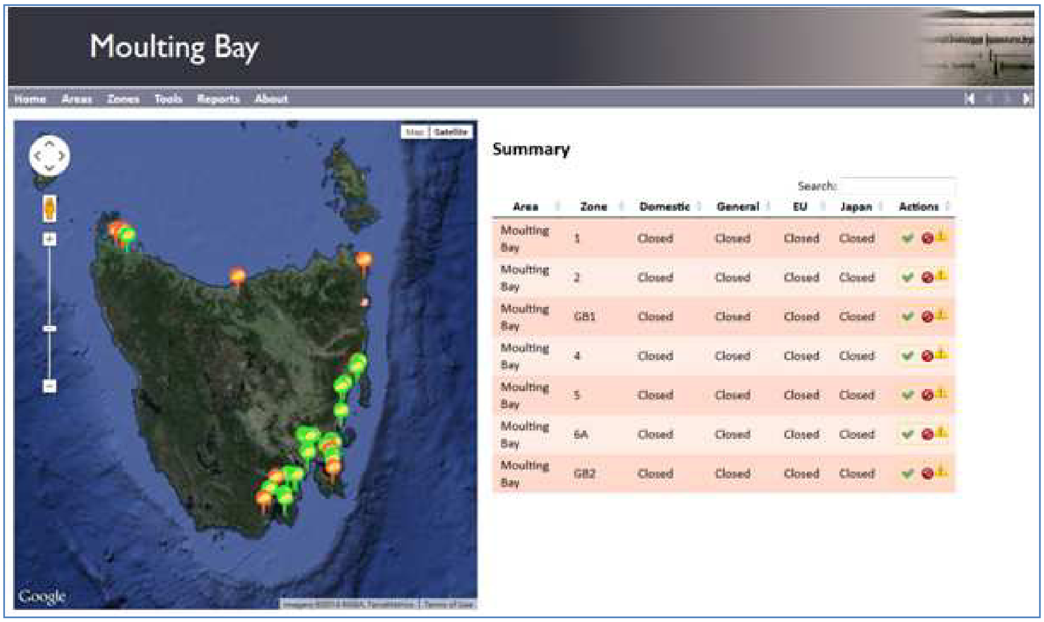

Figure 1: A snapshot of current decision support of TSQAP

In aquaculture decision support, we want to add the capability of predicting closure decision at least one day ahead using dynamic data-driven approaches. This predictive capability will benefit shellfish farms and relevant authorities to take decisions early related to the closure or opening status. We explain our proposed approach in Figures 2 and 3. First, we show the data-driven model learning in Figure 2. Historical data of related environmental variables for closure of shellfish farms are transformed and fused together as time series. Closure information for shellfish farms are also integrated. Then we consider two alternative ways in predicting closure decisions. One way is to predict the variable causing the farm closure. In this paper, we consider salinity. To predict salinity one day ahead, we consider salinity, water temperature, rainfall and river flow as input variables. After predicting salinity, we apply expert rules (thresholds) on the predicted salinity whether farm may be closed one day ahead. The other way is to predict closure using time series classification. In this technique, rather predicting salinity one day ahead and using rules on the predicted salinity, we use time series features of salinity, water temperature, river flow and rainfall to predict the closure.

When the models are learned and recommended from Figure 2, we use the dynamic approach where prediction is made using time series data. In the dynamic approach, we feed the input parameters to the learned model and a decision is recommended. The learned models are also updated in the module. This is shown in Figure 3.

We consider the following contributions in this paper.

- We propose and develop a dynamic data-driven decision support for aquaculture. In dynamic data-driven approach, we use time series prediction and classification models.

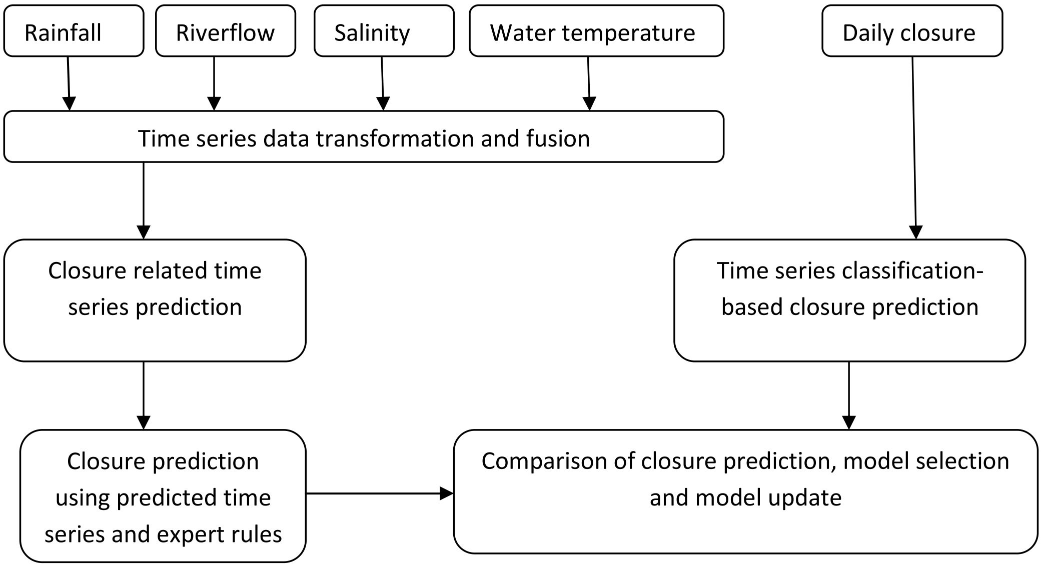

Figure 2: Data-driven predictive model learning (off-line)

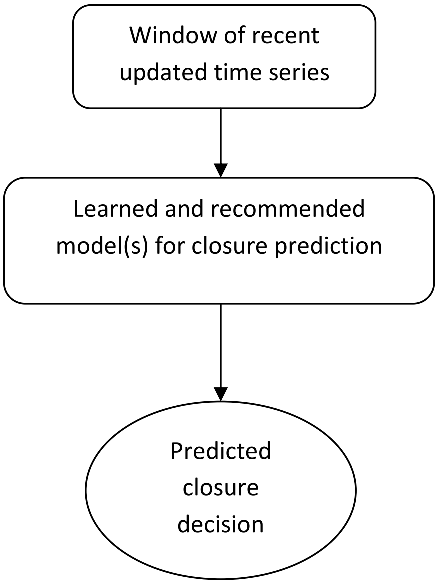

Figure 3: Predicting closure based on learned model and recent data arrival(on-line)

- We exploit two ways in predicting closure decisions. One way is to predict environmental time series that is considered proxy for closure and then to use expert rules on the predicted data. Other way is to use time series classification for predicting closures.

This paper is organized as follows. In Section 2, we briefly present related research. Methods are discussed in Section 3. We present experimental results for a shellfish farm in Section 4. Finally we conclude in Section 5.

2 Related research

Decision support systems can be data-driven where decisions are usually made based on the intelligent data analysis. Most of these data-driven decision support systems are also static meaning that decisions are made based on the models derived from existing historical data. The applications of data-driven decision support systems include many domains such as finance, supply-chain [1], clinical[4] and environment[5].

Dynamic data-driven application systems (DDDAS) are being developed for many areas such as environmental monitoring [6], natural disaster [7], hazard analysis [8], transport systems [9, 10], emergency management [11], healthcare [12] and cyber-physical systems [13]. DDDAS systems can include decision support systems [2, 14]. In [2], real-time data is assimilated with simulation model to update the initial parameters for oil spill incidents management and decision making. For dynamic capabilities of a decision support system, there is a need to include functionalities that can cope with changes in data [15].

In aquaculture, there exist some data-driven decision support systems [16, 17, 18]. The decision support system for aquaculture facility management and planning is studied in [16]. In cage aquaculture [17], a decision support system is developed with functionalities of site classification and selection, capacity of the farms and economic appraisal for a specific site. For aquaculture licensing, a decision support that can provide environmental impacts on aquaculture for a location is studied in [18]. All these systems exhibit the static properties of a data-driven decision support for aquaculture.

Time series analysis and prediction are used in decision support systems for business [19] and agriculture [20]. Different soft computing (i.e. fuzzy) and statistical techniques are used in predicting time series in [19]. In agricultural decision support system for crops, rainfall and evaporation time series are predicted using fuzzy logic.

Time series classification are also used in decision support systems in stock trend prediction [21] and supply-chain management [22] . In [21], classification and clustering techniques are used in producing a multidimensional decision support indicator for predicting stock trends. An intelligent system to classify time series data using support vector machine for supply-chain domain is investigated in [22].

3 Methods

In this section, we discuss the general methods. First, we consider data transformation and fusion. Then we consider two approaches in predicting closure status. First approach is to predict environmental variable related to closure and to apply expert rules as classifier on the predicted value. The other approach is to use time series classification-based prediction where multiple classifiers are used in predicting closure. Finally, prediction models are updated when new data arrive in the system.

3.1 Time series data transformation and fusion

In this method, time series data are extracted, transformed and integrated. As time series data are collected from heterogeneous sources, sampling rates can be different. We transform salinity and water temperature data from minutes to daily resolution. River flow data is transformed from minutes to daily resolution. Rainfall and closure data are already at daily resolution.

3.2 Prediction models

We consider the following two approaches in prediction.

3.2.1 Closure prediction using time series prediction and expert rules

In this approach we first predict the environmental variable that is considered as proxy (reason) for closure. Then we apply the expert rules on the predicted value to predict closure decisions. For example, if salinity is the main variable for closure decision, we consider the following steps:

Step 1: We define the prediction as $sal_{t+1} = f(sal_t, sal_{t-1}, \cdots, sal_{t-p}, rf_t, rf_{t-1}, \cdots, rf_{t-q}, rain_t, rain_{t-1}, \cdots, rain_{t-r}, wt_t, wt_{t-1}, \cdots, wt_{t-s})$ where sal is salinity, rf is river flow, wt is water temperature, $rain$ is rainfall and t is the current day. The lag values p, q, r, s can be different in the model. f is the regression model used in prediction.

Step 2: If sal_{t+1} `op` $threshold$, then $decision_{1_{t+1}}$ else $decision_{2_{t+1}}$ where `op` is the operator such as greater than or less than and $threshold$ is the value decided by the domain expert. $decision$ can be either open or close status of the farm.

We exploit machine learning techniques in predicting time series [23]. We use regression techniques such as M5P [24], support vector regression (SMOReg) [25], linear regression (LR) and artificial neural network (NN) [26]. In NN, we use multilayer perceptron (MLP) with different hidden layers. As input to the regression methods, we provide combination of time lags with a maximum of 5 days and a minimum of 1 day for one day ahead prediction. A model with minimum mean squared error (MSE) is recommended for prediction of a time series.

3.2.2 Closure prediction using time series classification

In this approach, we also exploit machine learning on time series data to classify (i.e. predict) closure [27]. This approach is different from the previous approach where we predict time series to apply expert rules for closure prediction. We use time series features to predict a closure decision.

We define this method in the following.

$decision_{i_{t+1}} = f(sal_t, sal_{t-1}, \cdots, sal_{t-p}, rf_t, rf_{t-1}, \cdots, rf_{t-q}, rain_t, rain_{t-1}, \cdots, rain_{t-r}, wt_t, wt_{t-1}, \cdots, wt_{t-s})$ where f is the classifier. $decision_{i_{t+1}}$ is the closure decision predicted for one day ahead.

In time series classification, lagged window of time series is used as feature in classifier model to predict the decision (output label). We use support vector regression (SMO) and NN as classifier for predicting closure one day ahead.

3.3 On-line prediction based on learned models

When we have learned models, we use those models to predict closure using recent values of the environmental time series. The models can be either time series prediction and expert rule-based approach or time series classification-based approach. The initial models are learned from the existing historical data. As the data arrive in the system, the new data can be added to the historical data to refine the model. In this method, learned models are updated periodically with the arrival of new data in the system. This provides the dynamic and incremental capability of learning models in predicting closure decisions.

4 Experimental results

We use data set for a shellfish farm at Moulting Bay Zone 1 at Tasmania, Australia as a case study. The shellfish farm at this location use salinity as primary trigger for closure of the farm. We consider data from 1/07/2012 to 31/08/2013. Last two month data is considered for testing and the remaining data is considered for training regression and classification models.

We transform environmental data as daily time series data and integrate with closure data from multiple sources. Salinity, water temperature and closure data are collected from Tasmanian Shellfish Quality Assurance Program (TSQAP)[28]. River flow data is collected from Department of Primary Industries Parks, Water and Environment (DPIPWE) [29]. Rainfall data is collected from Bureau of Meteorology (BOM) [30].

Technique	Lag days(River flow-Rainfall-Water temperature-Salinity)	MSE
M5P	1-1-1-3	0.361
SMOReg	1-1-3-1	0.365
LR	1-2-2-1	0.375
NN (Hidden layer=3)	2-1-1-1	0.395

Table 1: Models ordered on MSE

4.1 Time series prediction-based closure prediction

In Table 1, we present different regression models and lag days used in building prediction model for salinity time series data. The models are ordered according to the lowest mean squared error (MSE) on the training data.

We show the actual and predicted time series in Figure 4 using trained models shown in Table 1. We find that the model M5P using river flow, rainfall and water temperature with 1 day lag and salinity with 3 days lag performs best. We now apply the TSQAP rules (if Salinity $\leq$ 30 PSU, then close) on the predicted salinity using M5P model for closure prediction. The accuracy of closure prediction on the testing data is shown in the Table 2. The overall accuracy of prediction is 83%. The accuracy of the opening prediction is 92% and the accuracy of the closure prediction is 66%.

	Closed	Opened
Closed	14	7
Opened	3	36

Table 2: Closure prediction using time series prediction with expert rules

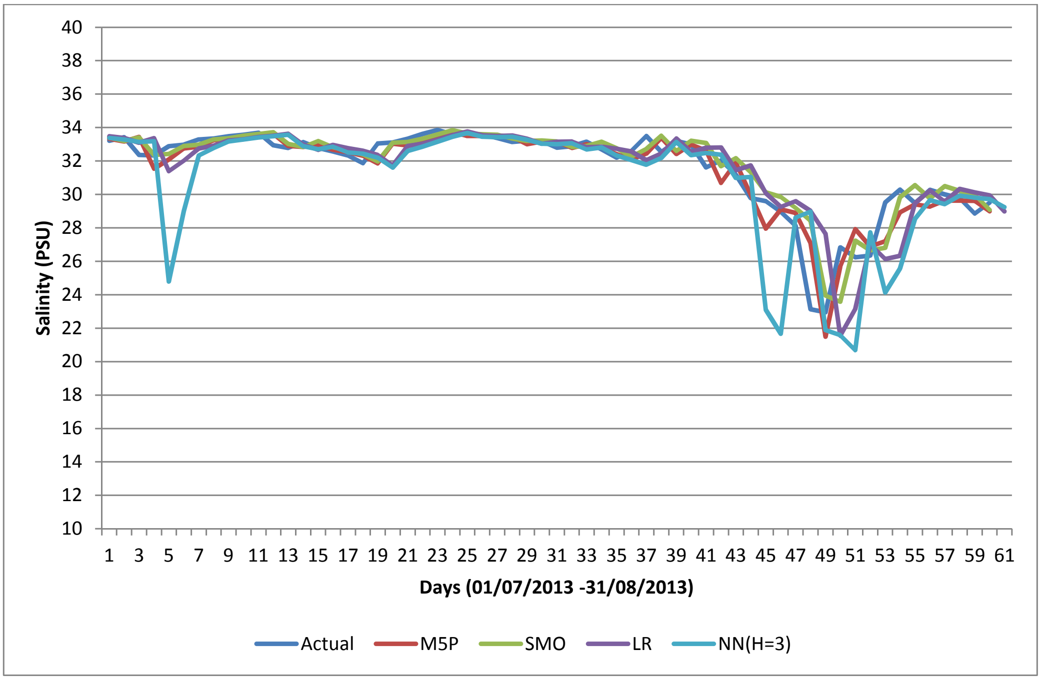

Figure 4: Prediction results of salinity

4.2 Time series classification-based closure prediction

We now present the prediction of closures using time series classification. In this case, we use a window of lagged time series as input and the closure information (open/close) as output (labels) for classifiers.

The prediction accuracy using support vector machine (SMO) and neural network with single hidden layer are shown in Tables 3 and 4 respectively. We find that the accuracy of SMO in Table 3 is similar to the expert rule-based prediction shown in Table 2. However, the accuracy of predicting close using NN (81%) is higher than the accuracy of predicting close using SMO (66%).

	Closed	Opened
Closed	14	7
Opened	3	36

Table 3: Closure prediction using time series classification with SMO

	Closed	Opened
Closed	17	4
Opened	8	31

Table 4: Closure prediction using time series classification with NN(H=1)

5 Conclusions

We presented a dynamic data-driven decision support for shellfish farm closure. Machine learning techniques on time series data are used in predicting closure of a farm. In building prediction models, we used a dynamic data-driven technique where learned models are updated to predict decisions. We plan to include capabilities to adapt missing values in prediction models. We also plan to evaluate the performance of the prediction capabilities of the system.

6 Acknowledgement

CSIRO Computational Informatics CCI, Hobart(formerly known as The Intelligent Sensing and Systems Laboratory) and the Tasmanian node of the Australian Centre for Broadband Innovation is assisted by a grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts. This research is also supported with Aquaculture Decision Support (AquaDS) project under CSIRO Food Futures Flagship. We also acknowledge Claire D'Este, Ashfaqur Rahman, Daniel Hugo, Greg Smith and Chris Sharman for their comments and support.

References

[1] A. Gunasekaran and E.W.T. Ngai. Decision support systems for logistics and supply chain management. *Decision Support Systems*, 52(4), 2012.

[2] Q. Zeng, Q. Zhang, and Z. Yang. A dynamic data driven application simulation model for oil spill emergency decision in port water area. *Transport*, pages 1–5, 2013.

[3] J.G. Ferreira, J. Aguilar-Manjarrez, C. Bacher, K. Black, S.L. Dong, J. Grant, E. Hofmann, J. Kapetsky, P.S. Leung, R. Pastres , . Strand, and C.B. Zhu. Progressing aquaculture through virtual technology and decision-support tools for novel management, farming the waters for people and food. In *Proceedings of the Global Conference on Aquaculture*, pages 643–704, 2010.

[4] T. J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R.R. Coeytaux, G. Samsa, V. Hasselblad, J.W. Williams, M.D. Musty, L. Wing, A. S. Kendrick, G.D. Sanders, and D.Lobach. Effect of clinical decision-support systems a systematic review. *Annals of Internal Medicine*, 157(1):29–43, 2012.

[5] T. Sullivan, J. Zhang, E. O'Connor, B.B. Ciprian Constantin, B. Heery, G. Leonardo, A. F. Smeaton, N. E. O'Connor, and F. Regan. Improving data driven decision making through integration of environmental sensing technologies. In *Ocean 13*, pages 10–13, 2013.

[6] L. Peng, D. Lipinski, and K. Mohseni. Dynamic data driven application system for plume estimation using uavs. *Journal of Intelligent & Robotic Systems*, 72(1):421–436, 2014.

[7] Jan Mandel, LynnS. Bennethum, Mingshi Chen, JaniceL. Coen, CraigC. Douglas, LeopoldoP. Franca, CraigJ. Johns, Minjeong Kim, AndrewV. Knyazev, Robert Kremens, Vaibhav Kulkarni, Guan Qin, Anthony Vodacek, Jianjia Wu, Wei Zhao, and Adam Zornes. Towards a dynamic data driven application system for wildfire simulation. In *International conference on computational science*, volume 3515 of *Lecture Notes in Computer Science*, pages 632–639, 2005.

[8] A.K. Patra, M. Bursik, J. Dehn, M.D. Jones, M. Pavolonis, E.B. Pitman, T. Singh, P. Singla, and O. Webley. A dddas framework for volcanic ash propagation and hazard analysis. In *International Conference on Computational Science*, pages 1090–1099, 2012.

[9] Szu-Yin Lin, Kuo-Ming Chao, and Chi-Chun Lo. Service-oriented dynamic data driven application systems to urban traffic management in resource-bounded environment. *SIGAPP Appl. Comput. Rev.*, 12(1):35–49, 2012.

[10] Szu-Yin Lin, Kuo-Ming Chao, and Chi-Chun Lo. Service-oriented dynamic data driven application systems to traffic signal control. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 3463–3470, 2010.

[11] Mark Gaynor, Margo Seltzer, Steve Moulton, and Jim Freedman. A dynamic, data-driven, decision support system for emergency medical services. In *International conference on computational science*, volume 3515 of *Lecture Notes in Computer Science*, pages 703–711, 2005.

[12] J.T. Oden, D. Fuentes, J. Bass, and Y. Feng. Dynamic-data-driven systems aid patient-specific cancer therapy. *SPIE*, pages 1–3, 2008.

[13] Glynis Dsouza, Salim Hariri, Youssif Al-Nashif, and Gabriel Rodriguez. Resilient dynamic data driven application systems (rdddas). In *International Conference on Computational Science*, volume 18, pages 1929–1938, 2013.

[14] Tahsin Kurc, Xi Zhang, Manish Parashar, Hector Klie, MaryF. Wheeler, Umit Catalyurek, and Joel Saltz. Dynamic data-driven systems approach for simulation based optimizations. In *International Conference on Computational Science*, volume 4487, pages 1213–1221, 2007.

[15] Frederica Darema. Dynamic data driven applications systems: New capabilities for application simulations and measurements. In *International Conference on Computational Science*, volume 3515, pages 610–615, 2005.

[16] Douglas H Ernst, John P Bolte, and Shree S Nath. Aquafarm: simulation and decision support for aquaculture facility design and management planning. *Aquacultural Engineering*, 23(1):121–179, 2000.

[17] Halmar Halide, A. Stigebrandt, M. Rehbein, and A.D. McKinnon. Developing a decision support system for sustainable cage aquaculture. *Environmental Modelling & Software*, 24(6):694–702, 2009.

[18] W. Silvert. Decision support systems for aquaculture licensing. *Journal of Applied Ichthyology*, 10(4):307–311, 1994.

[19] J. D. Bermúdez, J. V. Segura, and E. Vercher. A decision support system methodology for forecasting of time series based on soft computing. *Comput. Stat. Data Anal.*, 51(1):177–191, 2006.

[20] M.I. Animas, B.D. Gerardo, Yung-Cheol Byun, and M.B. Concepcion. Decision support system for agricultural management using prediction algorithm. In *Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on*, pages 191–194, 2013.

[21] Kuralmani Vellaisamy and Jinyan Li. Multidimensional decision support indicator (mdsi) for time series stock trend prediction. In *Advances in Knowledge Discovery and Data Mining*, volume 4426, pages 841–848, 2007.

[22] Fernando Turrado García, Luis Javier García Villalba, and Javier Portela. Intelligent system for time series classification using support vector machines applied to supply-chain. *Expert Syst. Appl.*, 39(12):10590–10599, 2012.

[23] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Al Le Borgne. Machine learning strategies for time series forecasting. In *Business Intelligence*, volume 138 of *Lecture Notes in Business Information Processing*, pages 62–77, 2013.

[24] Y. Wang and I.H. Witten. Induction of model trees for predicting continuous classes. In *9th European Conference on Machine Learning*, 1997.

[25] S.K. Shevade, S.S. Keerthi, C. Bhattacharyya, and K. R K Murthy. Improvements to the smo algorithm for svm regression. *Neural Networks, IEEE Transactions on*, 11(5):1188–1193, Sep 2000.

[26] M. J. Aitkenhead and R. J. Cooper. Neural network time series prediction of environmental variables in a small upland headwater in ne scotland. *Hydrological Processes*, 22(16):3091–3101, 2008.

[27] Marcella Corduas. Mining time series data: A selective survey. In Francesco Palumbo, Carlo Natale Lauro, and Michael J. Greenacre, editors, *Data Analysis and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 355–362. Springer Berlin Heidelberg,

2010.
[28] Tasmanian shellfish quality assurance program (tsqap). http://www.dhhs.tas.gov.au/peh/tsqap.
[29] Department of primary industries, parks, water and environment. http://www.dpiw.tas.gov.au.
[30] Bureau of meteorology. http://www.bom.gov.au.