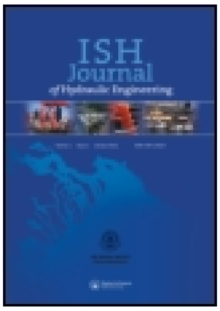

ISH Journal of Hydraulic Engineering
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tish20

Reply to the "Discussion by Haddad et al. on 'Hydroclimatic stream flow prediction using least square-support vector regression' by Bhagwat and Maity (2013)"

Parag P. Bhagwat[a] & Rajib Maity[b]

[a] School of Civil Engineering, Lovely Professional University, Punjab, India

[b] Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India

Published online: 05 Feb 2014.

ISH Journal of Hydraulic Engineering, 2014
http://dx.doi.org/10.1080/09715010.2014.881075

Reply to the "Discussion by Haddad et al. on 'Hydroclimatic stream flow prediction using least square-support vector regression' by Bhagwat and Maity (2013)"

Parag P. Bhagwat[a] and Rajib Maity[b]*

[a]*School of Civil Engineering, Lovely Professional University, Punjab, India;* [b]*Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India*

(*Received 20 October 2013; accepted 30 November 2013*)

We appreciate the discussion by Haddad et al. (2013) (hereinafter HAN2013) that reflects the interest of the hydrologic community to the subject area presented in Bhagwat and Maity (2013) (hereinafter BM2013). This reply helps the authors to provide some more explanation that was not explicitly stated in BM2013. In this article, our replies may be read in the same order of the discussion points in HAN2013.

Regarding the comment on missing values, we wish to reiterate that (quoting from BM2013) "For some (noncontiguous) days (approximately 150 days out of 23 years), rainfall and temperature data are missing. These values are replaced by the average of other surrounding stations (mentioned before) for that date". Thus, the missing values are less than 2% of the total record available. This is for the Narmada basin. For the Mahanadi basin only "Temperature data is missing for some (noncontiguous) days (approximately 250 days out of 31 years)". We agree that even for such a small number of missing data (~2%) also any sophisticated methods, including those suggested by H2013 ("geostatistical or data mining methods"), can be used. However, LSSVM approach is lumped in nature and Thiessen polygon method is adopted in BM2013 to estimate the average values for the entire basin. Thus, the missing values are replaced by the average of the surrounding stations before using the Thiessen weights. Alternatively, revised Thiessen polygons (ignoring the station having missing data) may be constructed for those particular dates when the data from one station is missing.

Regarding "normalization", we feel "scaling" would have been a better word. As the structure suggests, only rescaling is done through Equation (11). It does not transform the data to follow Gaussian (or Normal) distribution. Thus, test for "normalization" is not applicable. Secondly, it is true that Bray and Han (2004) used support vector machine (SVM) in predicting the runoff. However, both SVM and LS-SVM are based on the statistical learning theory. LS-SVM, proposed by Suykens and Vandewalle (1999), is a modified version of the standard SVM, which uses equality constraints instead of inequality constraints used in SVM whose solution results in a set of linear equations. There are many studies where scaled data are used in the case of LS-SVM as well (Lin et al. 2006; Qin et al. 2005). Hence, rescaled data was used in BM2013. We also like to mention that transformation through Equation (11) is also termed as "normalization" in many studies (e.g. Ceryan et al. 2013; Ismail et al. 2012; Kisi 2012; Samsudin et al. 2011). Thus, the word "normalization" was used in BM2013.

A reference to the bounds of the parameters can be found in Samsudin et al. (2011). They used 0.01–1 for σ^2 and 10–1000 for γ. In BM2013, the same range for σ^2 was selected. The selected range for γ was even wider (1–1000). Grid search method can always provide the global solution since all possible combinations are checked. As a consequence, the best possible values that result in minimum prediction error are obtained (Kalra et al. 2013). Naturally, it is computationally huge. This is the reason to invent different algorithms to arrive at the global solution quickly. However, with the availability of fast computing facilities, grid search method is manageable and adopted in this study.

Regarding the validation data-set, we agree with HAN2013 that a separate validation data-set is always recommended in order to avoid the chance of overfitting. Since the available data-set was small, a separate data-set for validation was not used. However, referring to Table 1 in BM2013, comparable performance during training and testing periods indicates that the developed models are not overfitted. The training period data-set was not systematically selected. The data-set is approximately divided into 2/3rd and 1/3rd parts. The first 2/3rd (continuous) data was selected for training and the rest 1/3rd (continuous) data was used for testing.

An overall comment was made regarding the relative performance of the two basins. It is found that the results of the Mahanadi river basin, in most of the cases, are better as compared to that of the Narmada river basin. In brief, during training, correlation coefficient (CC) and Nash-Sutcliffe Efficiency (NSE) were found to be better in the case of the Mahanadi river basin (Table 1 in BM2013). Only Root Mean Square Error (RMSE) was better for the Narmada river basin. During testing, both RMSE and NSE were found to be better in the case of the Mahanadi river basin (Table 2 in BM2013). Only CC was found to be marginally better (0.85 vs. 0.86) in the case of the Narmada river basin. So far as the performance of a predictive model is concerned, NSE may be given higher weightage as compared to the CC

*Corresponding author. Email: rajib@civil.iitkgp.ernet.in

(McCuen et al. 2006). Moreover, RMSE was found to be better in the case of the Mahanadi river basin during both training and testing periods. Thus, an overall conclusion was made that the performance for the Mahanadi river basin was found to be better compared to the Narmada river basin. To arrive at this conclusion, HAN2013 suggests the use of Multi Criteria Decision Making. We leave this to the discretion of the readers.

References

Bhagwat, P.P., and Maity, R. (2013). "Hydroclimatic streamflow prediction using least square-support vector regression." *ISH J. Hydraul. Eng.*, 19(3), 320–328. doi:10.1080/09715010.2013.819705.

Bray, M., and Han, D. (2004). "Identification of support vector machines for runoff modelling." *J. Hydroinform.*, 6(4), 265–280.

Ceryan, N., Okkan, U., Samui, P., and Ceryan, S. (2013). "Modeling of tensile strength of rocks materials based on support vector machines approaches." *Int. J. Numer. Anal. Methods Geomech.*, 37(16), 2655–2670. doi:10.1002/nag.2154.

Haddad, O.B., Aboutalebi, M., and Nejad, I.G. (2013). "Discussion of 'Hydroclimatic stream flow prediction using least square-support vector regression' by Parag P Bhagwat and Rajib Maity, communicated."

Ismail, S., Shabri, A., and Samsudin, R. (2012). "A hybrid model of self organizing maps and least square support vector machine for river flow forecasting." *Hydrol. Earth Syst. Sci.*, 16, 4417–4433. doi:10.5194/hess-16-4417-2012.

Kalra, A., Li, L., Li, X., and Ahmad, S. (2013). "Improving streamflow forecast lead time using oceanic-atmospheric oscillations for Kaidu river basin, Xinjiang, China." *J. Hydrol. Eng.*, 18, 1031–1040. doi:10.1061/(ASCE)HE.1943-5584.0000707.

Kisi, O. (2012). "Modeling discharge-suspended sediment relationship using least square support vector machine." *J. Hydrol.*, 456–457, 110–120. doi:10.1016/j.jhydrol.2012.06.019.

Lin, J.Y., Cheng, C.T., and Chau, K.W. (2006). "Using support vector machines for long-term discharge prediction." *Hydrol. Sci. J.*, 51 (4), 599–612, doi:10.1623/hysj.51.4.599.

McCuen, R.H., Knight, Z., and Cutter, A.G. (2006). "Evaluation of the nash–sutcliffe efficiency index." *J. Hydrol. Eng.*, 11(6), 597–602. doi:10.1061/(ASCE)1084-0699(2006) 11:6(597.

Qin, Z., Yu, Q., Li, J., Wu, Z., and Hu, B. (2005). "Application of least squares vector machines in modelling water va-por and carbon dioxide fluxes over a cropland." *J. Zhejiang Univ. Sci.*, B6(6), 491–495. doi:10.1631/jzus.2005.B0491.

Samsudin, R., Saad, P., and Shabri, A. (2011). "River flow time series using least squares support vector machines." *Hydrol. Earth Syst. Sci.*, 15, 1835–1852. doi:10.5194/hess-15-1835-2011.

Suykens, J.A.K., and Vandewalle, J. (1999). "Least squares support vector machine classifiers." *Neural Process. Lett.*, 9(3), 293–300.