A multivariate model for flood forecasting of lake levels

M. Mohssen
Department of Environmental Management,
Lincoln University, New Zealand

Abstract

A new multivariate flood forecast model for lake levels has been developed. The model is based on the concept of the projection theorem, to obtain the projection of lakes flood levels on total rainfall forecasted or observed at nearby rainfall gauges. A comprehensive analysis of 20 historical and significant rainfall events on Lake Wakatipu and Lake Wanaka in the South Island of New Zealand has been carried out to detect the best time series and functions to be used for this projection.

The study showed that linear relationships exist between total lake rise and total rainfalls observed at available rain gauges. In addition, during flood events, linear relationship can be utilised to represent the function between observed cumulative rainfall and cumulative lake rise at 10 hours (Lake Wakatipu) and 7 hours (Lake Wanaka) of lead time. Analysis of relationships between lake rise and alternative options for available rainfall sites resulted in the choice of rainfall sites for the best forecast model. Parameter estimation of a multivariate model which includes both Lake Wakatipu and Lake Wanaka flood levels and the selected rainfall sites was carried out by using the projection equations of orthonormal sets. The derived model was applied to two observed events which were not included in the calibration process. The new model has performed well in forecasting flood levels of both lakes.
Keywords: flood forecast, flood modelling, rainfall-runoff, lake level, projection theorem, Hilbert space.

www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/FRIAR120101

1 Introduction

Floods are the most common natural disasters, and they cause devastating damage to communities. In the absence of flood protection infrastructures, flood warning is one of the most effective schemes to mitigate the impact of these natural hazards. The Environment Agency of UK and the strategic plan for the US National Weather Service indicated the urgent need for major investment to develop new forecast models for flood warning [1, 2].

A reliable flood forecast model is the most important component of an efficient flood warning system. If reliable estimates of flood levels are forecasted before the event happens, authorities/communities can have sufficient time to mitigate the impact of the coming flood event by getting prepared, evacuating, moving stock away from flood prone areas, and relocating precious items. Despite the fact that literature is rich with research on flood forecasting, many of these models/techniques fail to accurately forecast flood events on the shorter scale (hourly) due to the high variability of the associated hydrological and meteorological processes [1]. High variability of precipitation, the main driving factor for floods, along with complex hydrological characteristic of the catchment area, makes it hard to obtain the right variables representing these processes at the right resolution to be capable of producing a reliable flood estimate of an occurring or incoming event.

Despite the fact that the hydrologic process of lake flooding is quite different from river floods, available modelling approaches in the literature are still similar. Watershed modelling is used to simulate the inflows to the lake, and then a hydrologic budget model is applied to translate the difference between inflows and outflows to a rise (or drop) in lake level [3, 4].

Time series modelling, by using ARIMA models have been applied in the literature, but due to the stationary requirement of the modelling process, usually they apply to longer time periods (days or more), while during flood events we are usually concerned in hourly flows [5]. ANN has been recently applied to forecast lake levels, and several techniques have been suggested for their applications to hourly time steps [6–8].

Floods are the most damaging and costly natural hazard in New Zealand, with about 935 damaging flood events occurred during the period 1920 to 1983 [9]. Queenstown, a major town by Lake Wakatipu which is full of attractions for tourism, has been badly flooded during the events of 1878 and 1999. Flooding of Lake Wakatipu takes a long time (could be weeks) to recede, and this will have persistent impact on the economic activity of Queenstown township and its community. Lake Wanaka, about 60 km north east of Queenstown has also flooded the township of Wanaka which is on the lake front.

2 Multivariate forecast model for lake levels

Lake Wakatipu has an area of about 293 km^2, with a total catchment area of about 3059 km^2, and lies in the western ranges of Otago district, in the South Island of New Zealand. Lake Wanaka has an area of about 196 km^2 and its catchment is also in the western ranges of Otago with an area of 2564 km^2. Lake Wanaka discharges its outflows to the head of the Clutha River, which is the largest river in New Zealand with average flows of 534 m^3/s, while Lake Wakatipu outflows to the Kawarau River, which joins the Clutha River further downstream. Figure 1 shows the two lakes, their outflows and the location of rainfall sites of this study.

Four rainfall sites (the Hillocks, Makarora, Peats Hut and West Wanaka) are included in this study, as they have a suitable long record of data, while other sites are shorter and they don't include the most significant flood event of November 1999. Lake Wakatipu outflows are recorded at Willow Place, while the site for Lake Wanaka outflows is at Roys Bay, as shown in Fig. 1. The third lake "Lake Hawea" which lies just to the east of Lake Wanaka was not included in this study as it is completely controlled, and there is no major settlement beside the lake.

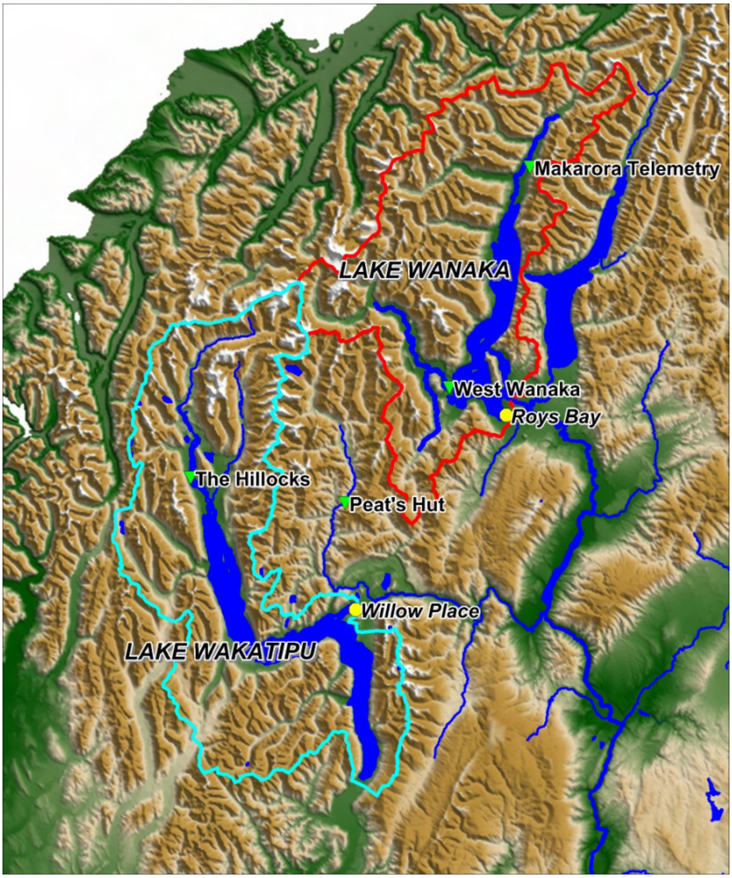


Figure 1: Rainfall and flow sites of Lake Wakatipu and Lake Wanaka.

Lake flooding is different from river flooding as the lake has a huge storage and its outflows are a function of its level rather than inflows to the lake. Thus lake rise is a function of cumulative rainfall over a period of time rather than its intensity during a short period of the rainfall event [10]. The new concept for modelling which has been applied in this research utilises the projection theorem [11] to obtain the projection of the total rise of lake levels on the cumulative rainfall of the selected rain gauges. The projection theorem guarantees that the obtained model will produce the best forecasted flood levels, if the correct function which relates the time series was selected. The work carried out here is an extension of the univariate case presented by [10], where Lake Wakatipu levels were estimated by rainfall at only one rainfall gauging site (the Hillocks). In this research, multiple rainfall sites will be used to forecast lake levels at two lakes (Wakatipu and Wanaka).

2.1 Model formulation

For large lakes, such as Lake Wakatipu and Lake Wanaka, flooding will be slow and flood levels are more dependent on cumulative rain rather than rainfall intensity, and usually peak flood levels recede slowly as the lake outlet takes long time to dispose the huge amounts of water stored in the lake during the flood event. An hourly hydrologic balance for a lake, in the univariate case, can be expressed as follows [10]:

$$L_{t+t_f} = L_{t-t_1} - \sum_{j=t-t_1}^{j=t+t_f-1} Q_j + f(R_t) \tag{1}$$

Assuming that at the start of the rainfall event t=0, then L_t is the lake level (mm) at present time t, t_f is the forecast time in hours after t, t_1 is a lag time (hrs) before t, Q_j is the lake outflow (mm) at time j, R_t is the cumulative rainfall (mm) over a period of hours and $f(R_t)$ is a function of this cumulative rainfall. Of course $f(R_t)$ will have units of level (mm).

Mohssen and Goldsmith [10], by analysing the 1999 flood event, showed that the best value for t1 is the present time t, thus $t-t_1=0$, which is the start of the rainfall event. Level L_0 at the start of the event is usually used to get the rise of the lake level. Thus, Eq. (1) can be written as follows:

$$L_{t+t_f} - L_0 + \sum_{j=0}^{j=t+t_f-1} Q_j = f(R_t) \tag{2}$$

Let us denote the forecasted total lake rise since the start of the event $L_{t+t_f} - L_0$ as ΔL, and the cumulative outflows since the start of the event $\sum_{j=0}^{j=t+t_f-1} Q_j$ as ΔQ, then Eq. (2) in the multivariate form is:

$$\begin{bmatrix} \Delta \boldsymbol{L_1} \\ \Delta \boldsymbol{L_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{R})_{\boldsymbol{1}} \\ \boldsymbol{f}(\boldsymbol{R})_{\boldsymbol{2}} \end{bmatrix} - \begin{bmatrix} \Delta \boldsymbol{Q_1} \\ \Delta \boldsymbol{Q_2} \end{bmatrix} \tag{3}$$

where the subscripts 1 and 2 refer to Lake Wakatipu and Lake Wanaka, respectively. Total lake inflows during a flood event equal the lake rise plus cumulative outflows, i.e. $\Delta\mathbf{I} = \Delta\mathbf{L} + \Delta\mathbf{Q}$, thus:

$$\Delta\mathbf{I} = \mathbf{f}(\mathbf{R}_t) \tag{4}$$

where $\Delta\mathbf{I}$ is a column vector of total lake inflows during the flood event at the two sites, $\mathbf{f}(\mathbf{R}_t)$ is a column vector of functions of rainfall at multi rainfall sites. Mohssen and Goldsmith [10] showed that using total lake inflows does not result in improving the reliability of the forecast model for the univariate case, and to the contrary, model testing showed that it resulted in less reliable models. The same has been shown in the results of this study for multi rainfall sites, and thus we will show only the results for the projection of total lake rise since the start of the event on cumulative rainfalls at selected sites.

For the best projection of $\Delta\mathbf{L}$ on $\mathbf{f}(\mathbf{R}_t)$, $\Delta\mathbf{L} - \mathbf{f}(\mathbf{R}_t)$ should be orthogonal to $\mathbf{f}(\mathbf{R}_t)$. In Hilbert space, this is written as:

$$<\Delta\mathbf{L} - \mathbf{f}(\mathbf{R}_t)\, ,\, \mathbf{f}(\mathbf{R}_t)> = \mathbf{0} \tag{5}$$

Assuming that:

$$\mathbf{f}(\mathbf{R}_t) = \propto \boldsymbol{X} = \sum_{\mathbf{1}}^{\boldsymbol{n}} \propto_{\boldsymbol{i}} \boldsymbol{X}_{\boldsymbol{i}} \tag{6}$$

where α is a row vector of parameters, and $\mathbf{X}$ is a column vector of power functions of rainfall at each site R_i, Eq. 5 can be re-written as:

$$<\Delta\mathbf{L} - \propto \boldsymbol{X}\, ,\, \boldsymbol{X}> = \mathbf{0} \tag{7}$$

Application of Eq. (7) would produce n equations in the parameters ∝, which can be solved simultaneously to obtain the values of the parameters ∝ as a function of statistical properties of the time series ΔL and X.

The solution of (7) can be written, in matrix form, as follows:

$$\propto = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\, \boldsymbol{X}^{\boldsymbol{T}}\, \Delta\boldsymbol{L} \tag{8}$$

In case $\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-\mathbf{1}}$ is singular, there will be infinite solutions of (8). However, by the uniqueness of the projection theorem, all of them will produce the same forecast [11].

The projection theorem guarantees that the model provided by (8) will produce coefficients of $\mathbf{f}(\mathbf{R}_t)$ for the best forecasts of $\Delta\mathbf{L}$. Of course, the type of the functions $\mathbf{f}(\mathbf{R}_t)$ is not known, and part of the analysis carried out here is to obtain the best estimate of this function. In the case of a linear relationship, the power of R is one, and the power can be zero for a constant parameter (intercept). There are two different cases to be considered herein. The first case is the total lake rise (inflow) due to the total rainfall of an event. This is required in the case of long term forecast, when an event is expected to hit the region with an estimated total rainfall at a selected site, for instance on the second day or two days later. In this case, real rainfall of this event is not observed yet, but it is important to forecast the impact of this event on the potential flooding of Queenstown or Wanaka. The other case is flood forecasting during the event itself, and in this case one forecasts the cumulative lake rise at both lakes after several hours (forecast time) due to observed cumulative rainfall at the selected rainfall sites.

2.2 Case 1: Total lake rise due to total rainfall

Figures 2 and 3 present the relation between total rainfall of 20 events, and the corresponding total lake rise. The figures show that a linear relationship is quite suitable to represent this relationship. It is quite clear from Fig. 2 that total rise of Lake Wakatipu is represented well by a linear relation of total rainfalls at the

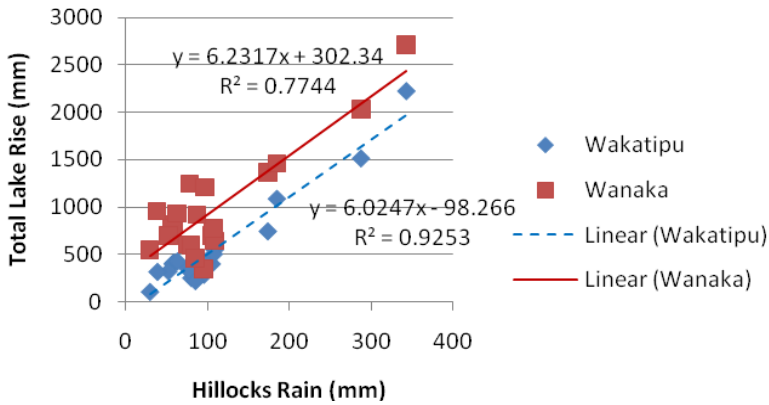


Figure 2: Total rainfall at Hillocks and Lake Wakatipu Rise.

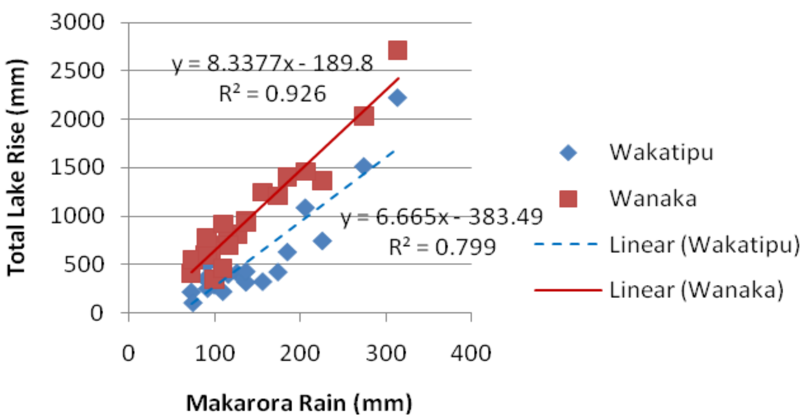


Figure 3: Total rainfall at Makarora and Lake Wanaka Rise.

Hillocks (R^2 = 0.93), while this is not the case for Lake Wanaka. Makarora rainfall site offers a significantly improved linear relationship to the total rise of Lake Wanaka (R^2 = 0.93) compared to the Hillocks rainfall (R^2 = 0.8). These results indicate the importance of carrying out the analysis to choose the right rainfall sites to select for the modelling process. Figures 4 and 5 show the linear relationship for the multivariate case. Both lakes have slightly improved their relationship with rainfall at the selected sites. However, the figures again confirm the fact that the wrong choice of rainfall sites would result in an unreliable forecast model. It should be mentioned that, in this study, other alternatives for rainfall sites were included, but only cases which improved the linearity of the relationship are presented.

Note that the variable x of the trend line in Figs 4 and 5 is the weighted average of total rain at several rainfall sites according to the weights of the projection (multiple regression) of lake rise onto total rainfall of these sites. Moreover, the projected weighted coefficients of the selected rainfall sites are not the same values for both lakes. For instance, in Fig. 4, the shown graphs are for the fitted parameters for Lake Wakatipu, not Lake Wanaka, and that is why it shows much higher determination coefficient for Lake Wakatipu than for Lake Wanaka, and the opposite is true for Fig. 5. Cases where multiple regression produced negative coefficients for any of the rainfall sites were excluded, as one would consider these coefficient to represent the contribution of this rainfall site to the lake under study. The models derived from this study are as follows:

$$\begin{bmatrix} \Delta L\ (Wakatipu) \\ \Delta L\ (Wanaka) \end{bmatrix} = \begin{bmatrix} 2.54 & 1.42 & 3.72 & -165 \\ 0.00 & 7.13 & 1.69 & -134 \end{bmatrix} \begin{bmatrix} H \\ M \\ W \\ 1 \end{bmatrix} \tag{9}$$

where total rainfall of the event at the Hillocks is H, at Makarora is M and at West Wanaka is W.

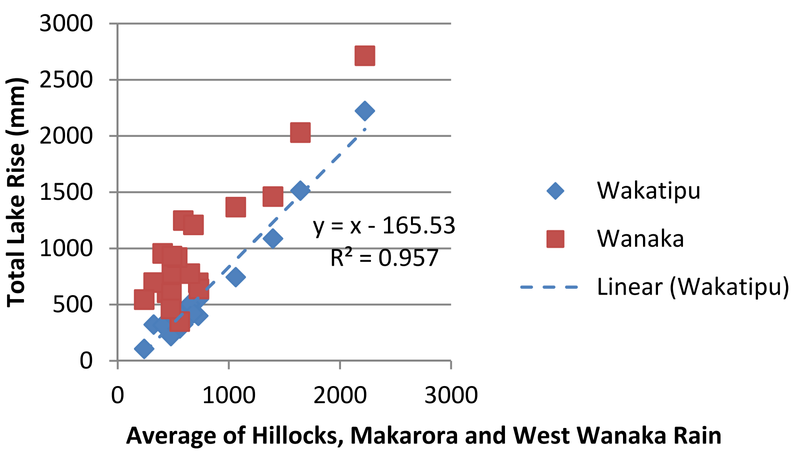


Figure 4: Average weighted rainfalls and Wakatipu level rise.

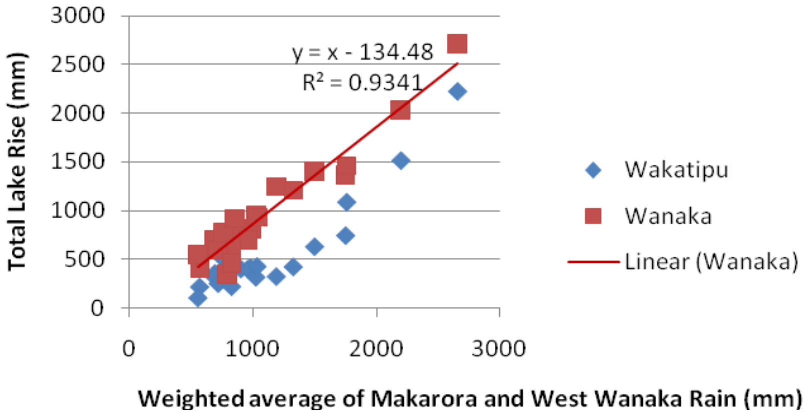
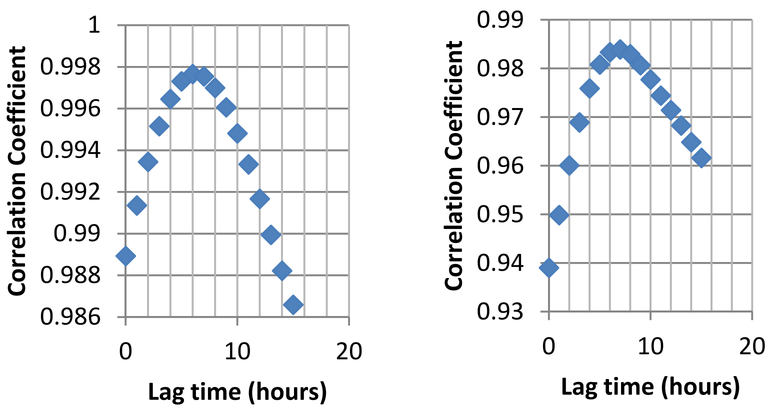


Figure 5: Average weighted rainfalls and Wanaka level rise.

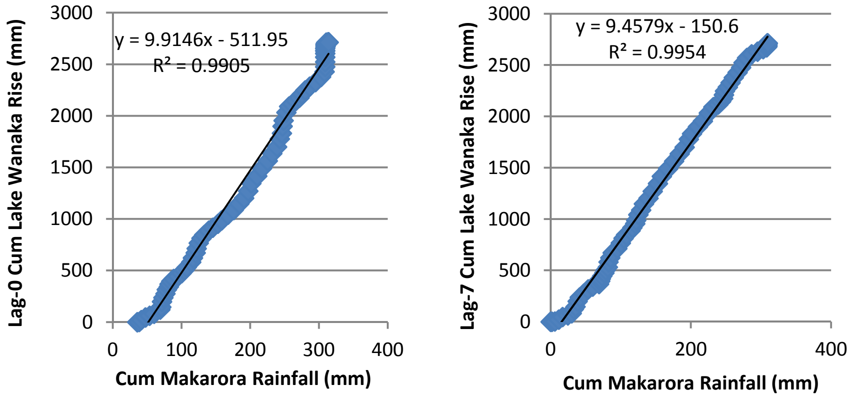
2.3 Case 2: forecast during the rainfall event

Mohssen and Goldsmith [10] showed that a lag of 10 hours produces a good linear relationship between cumulative Lake Wakatipu rise and cumulative rainfall at the Hillocks since the start of the rainfall events. As shown in Fig. 6, the optimum correlation between cumulative increase in Lake Wanaka rise and cumulative rainfall since the start of the rainfall event is 7 hours. Figure 7 confirms that linear relationship exists between cumulative rainfall and cumulative Lake Wanaka rise, and that the use of lag 7 hours significantly improved this linear relation, which will definitely improve any modelling process. The results of Fig 7 show the effectiveness of this new technique in improving the derived model.



(a) November 1999 Event (b) April 2010 Event

Figure 6: Lagged correlations between cumulative rain and Lake Wanaka rise.



(a) Lag-0 (b) Lag-7 hrs

Figure 7: Lagged correlations between cumulative rain and Lake Wanaka rise.

Knowing that the function $\mathbf{f(R_t)}$ of Eq. (6) is linear, the projection Eq. (8) was applied to the most significant two events, November 1999 and April 2010, to estimate the matrix of parameters $\propto$. Thus, the multivariate model for hourly forecast of Lake Wakatipu and Lake Wanaka levels is:

$$\begin{bmatrix} \Delta L\ (Wakatipu) \\ \Delta L\ (Wanaka) \end{bmatrix} = \begin{bmatrix} 4.28 & 0.00 & 3.3 & -101 \\ 0.00 & 5.07 & 5.2 & -19.4 \end{bmatrix} \begin{bmatrix} H \\ M \\ W \\ 1 \end{bmatrix} \qquad (10)$$

The determination coefficient for the fitted model is 0.985 for both Lake Wakatipu and Lake Wanaka.

3 Model testing

Validation of the fitted models was carried out by applying the model to rainfall events which were not included in its calibration. Table 1 shows the forecasted total lake rise versus the observed one for the two events, while Figures 8 and 9 show the simulation of the fitted model to forecast hourly cumulative lake rise with a lead time of 10 hours for lake Wakatipu and 7 hours for Lake Wanaka. The figures show that the model is capable of forecasting the lake levels with determination coefficients of about 0.98 for both cases. The Filliben correlation coefficient (FC) is a little bit higher than the determination coefficient. The percentage of error ranges from 1.5% to 22%.

Table 1: Observed vs Forecasted total Lake Rise (mm).

Event Date	Lake	Observed	Forecast	R^2	FC	% Error
Nov 2001	1	357	278	0.978	0.989	22.1
	2	605	561	0.977	0.988	7.3
Dec 2010	1	424	514	0.972	0.986	-21.2
	2	1210	1192	0.968	0.984	1.5

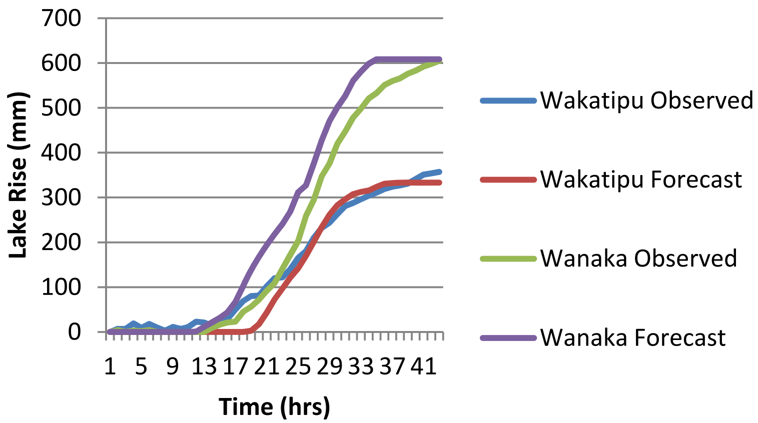


Figure 8: Observed vs forecasted cumulative lake's rise for the rainfall event November 2001.

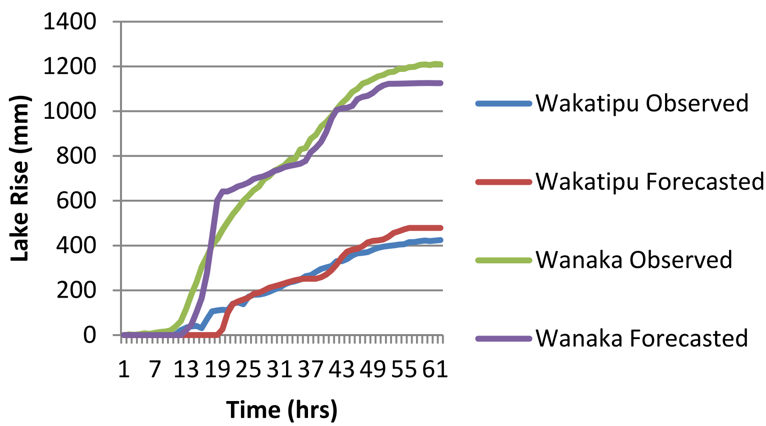


Figure 9: Observed vs. forecasted cumulative lake's rise for the rainfall event December 2010.

4 Conclusions

A multivariate model has been developed to forecast flood levels for Lake Wakatipu and Lake Wanaka in the South Island of New Zealand. The model is based on the projection theorem to obtain the optimum projection of the cumulative/total rise of lake levels onto cumulative/total rainfall at selected rainfall sites in the region. Analysis of lagged correlations between cumulative/total lake rise and cumulative/total rainfalls at the selected proved to be vital to estimate the optimum lag and to identify the function of the rainfall which is used in the projection equations. The projection equations guarantee that the produced model would produce the best forecasts. The derived model performed well in simulating observed events which were not included in the calibration process. The developed models can be used either for short term forecast during the flood event, and they give a lead time of 10 and 7 hours for Lake Wakatipu and Lake Wanaka, respectively, or they can be used for longer term forecast based on forecasted total rains at the selected sites. This modelling technique is unique and adds a new approach to the literature of flood modelling of lakes' levels.

References

[1] Bye, P. and M. Horner, Easter 1998 Floods Report by the Independent Review Team to the Board of the Environmental Agency, vol.1, Environmental Agency, Bristol, 1998.

[2] Demeritt D., H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes and Maria-Helena Ramoset, Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, Environmental Hazards, vol 7, pp.115-127, 2007.

[3] Jowitt, P.W. A Conceptual System Model of Rainfall-Runoff on the Haast River. Journal of Hydrology, New Zealand, 38(1), pp.121-144, 1999.

[4] Moore, R. *et al.*, Forecasting for flood warning, C.R. Geoscience vol 337, pp.203-217, 2005.

[5] Vaziri, M., Predicting Caspian Sea Surface water level by ANN and ARIMA models, JWtrwy, Port, Coast, Ocean Engineering, 123(4), 158-162, 1997.

[6] Tiwari, M. K., Chatterjee, C., Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach, J. of Hydrology 394 (2010) 458–470.

[7] Chen-ShenHsien; Lin-YongHuang; Chang-LiChiu; Chang-FiJohn, The strategy of building a flood forecast model by neuro-fuzzy network. Journal of Hydrological Processes, 20(7), pp.1525-1540, 2006.

[8] Kisi, O., *et al.*, Forecasting daily lake levels using artificial intelligence approaches. Computers and Geosciences (2011), doi:10.1016/j.cageo.2011.08.027 [7].

[9] McSaveney, E., Floods - New Zealand's number one hazard, Te Ara - the Encyclopedia of New Zealand, updated 2-Mar-09.

[10] Mohssen, M. and Goldsmith, M., Flood Forecasting of Lake Levels: A New Concept. Int. J. of Safety and Security Eng., 1(4), pp.363–375, 2011.
[11] Brockwell, P. J. and R. A. Davis., Time Series: Theory and Methods, Springer-Verlag New York Inc., pp.46-51, 1991.