**AWRA 2008 SPRING SPECIALTY CONFERENCE**
**San Mateo, California**
**March 17-19, 2008** 

# USING SATELLITE DATA TO IMPROVE THE ACCURACY OF RIVER FLOW FORECASTING MODELS A CASE STUDY OF ZAYANDE-ROOD RIVER BASIN, IRAN

Ahmad Abrishamchi, Massoud Tajrishy, Ali Tafarojnoruz,
Behdad Chehrenegar, and Shadab Shadzad *

ABSTRACT: River flow forecasting is important in reservoir operation and water allocation to different uses such as irrigation. In this paper, a river flow forecasting model is developed based on Principal Component Regression (PCR) analysis using some earth data such as precipitation, river discharge, snow water equivalent, climate signals as well as snow cover area extracted from NOAA-AVHRR satellite images. Satellite data were related to period of 1989 to 2004 and were downloaded via Satellite Active Archive (SAR) and about 3-4 images per month were processed to determine SCA. To distinct snow from land, albedo data of channel 2, temperature data of channel 4, and channel 3 data were used, and to separate snow and cloud, channel 3 and channel 4 data were applied. Furthermore, to separate composite pixels of 2 or 3 phenomena, a linear mixing model was developed. The proposed model was applied to one of the Zayande-rood river inflow tributaries to the Zayande-rood River in Iran. In the proposed model, river flow is forecasted for three periods in a year. In the first period at the end of January, river flow volume is forecasted for the period of February -June using observed data. Furthermore, volume of March-July (second period) and April-July (third period) is forecasted at the end of February and March, respectively. The results show that extracted SCA from satellite images is an effective parameter that improves the accuracy of river flow forecasting models.
KEY TERMS: Forecasting, Principal component regression, Snow Cover Area, Zayande-rood

## INTRODUCTION

Seeking maximum accuracy in regression-based forecasts is useful for several reasons. First, it is the basis of the practical seasonal streamflow volume forecasting. Physically-based watershed models, which can produce forecasts of the daily hydrograph as well as the seasonal or monthly runoff volumes, are promising improved forecasting technique (Pearson 1974; Twedt et al. 1977; Kuehl 1979; Speers and Versteeg 1982; Druce 1984; Day 1985), but they are used now only on a limited basis, and their widespread use is still many years away. Second, in some cases, regression may remain the forecast method of choice if adequate water management decisions can be made without more detailed hydrologic information or if more complex methods (such as physically-based models) are not sufficiently accurate. Finally, regression forecasts provide a baseline level of accuracy against which to test physically-based models. Of course, it would be an unfair test to compare physically-based model forecasts with non-optimal regression forecasts.

In recent years, many researchers tried to develop forecasting regression equations using ground-based data in which Snow Water Equivalent (SWE) was the only snow parameter in extracting those equations. In this paper, Snow Cover Area (SCA) data sets were used as a complement in development of equations to improve accuracy of forecasting equations. In other words, a combination of ground-based and Satellite-based data sets in improvement of forecasting equations was considered.

### Study Area

Zayande-rood River basin is located at the centre of Iran. The main river of this basin, the Zayande-rood River, flows from the Karivash Mountain in the west to the Gavkhooni Swamp in the east.

*Respectively, Professor, Associate Professor, Water Resources Specialist, GIS Specialist, Water Resources Specialist, Environment and Water Research Centre (EWRC), Department of Civil Engineering, Sharif University of Technology, Tehran, Iran, Phone: +00982166164238, Fax: +00982166014828 E-Mail: abrisham@sharif.edu.

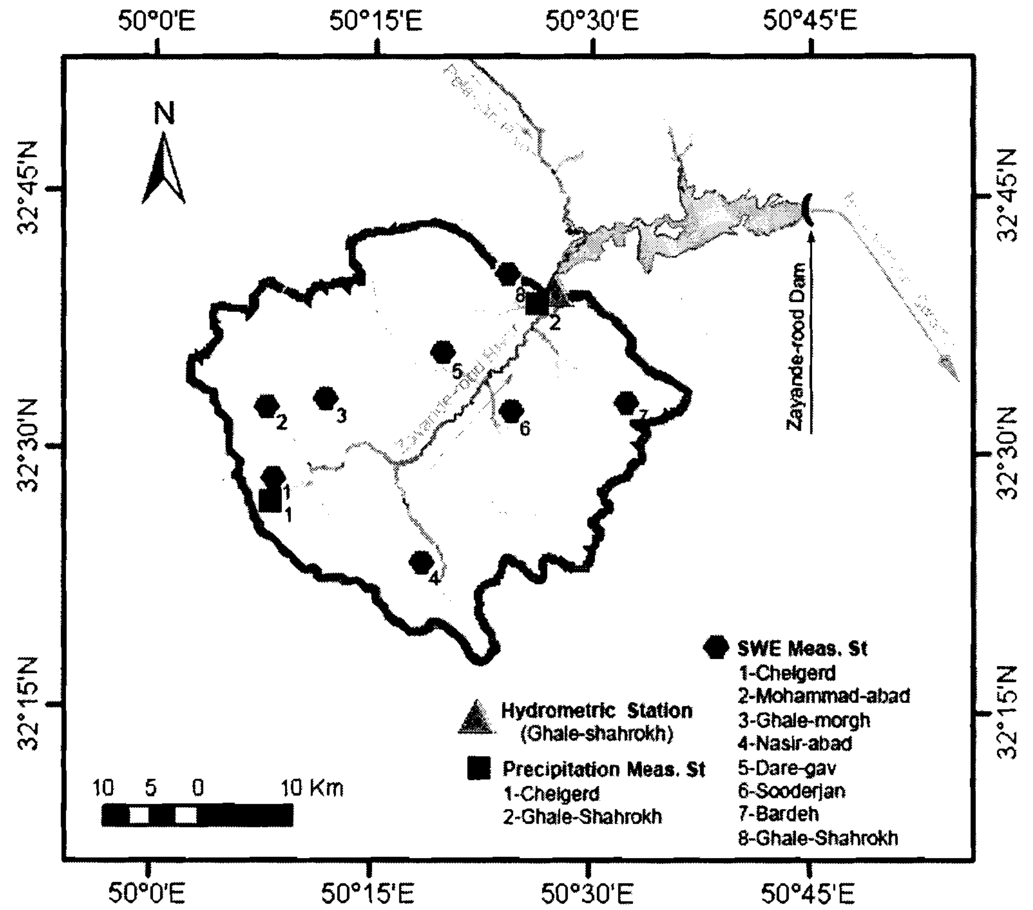


Figure 1. Zayande-rood River Basin and Location of SWE and Precipitation Stations

Zayande-rood reservoir as the main reservoir in the basin is a multi-purpose reservoir operating since 1970. Municipal and agricultural water supplies are the primary purposes of the Zayande-rood reservoir. However, production of hydropower and protection of the Isfahsn city monuments are the other purposes. The Zayande-rood reservoir is fed by two main tributaries; Plasjan and Zayande-rood.

In this paper, the sub-basin which has the most hydrological contribution to the reservoir-the Zayande-rood sub- basin-was selected (see Fig. 1). The area of this sub-basin is about 1432 sq. Kilometers. Its outflow is measured daily at the Ghale-shahrokh hydrometric station. The long term mean annual yield of this basin is about 800 MCM. There are eight snow water equivalent (SWE) and two precipitation gage stations in this basin (Fig. 1.).

## METHOD OF SNOW COVER AREA (SCA) ESTIMATION

Satellite data provide the only practical way to obtain the necessary spatial and temporal coverage of aerial extent of snow cover required for hydro-meteorological applications. Since NOAA-AVHRR data are one of the long term and most accessible satellite data, in this research, NOAA-AVHRR resource data were used to prepare required SCA data.

An improved technique in estimation of the aerial extent of snow covers from AVHRR data was proposed by Simpson, et al, (1998) which is an effective method of separation of snow from cloud and land. This method has two major steps. In the first step labeling is designed to separate clear land pixels from pixels of either snow or cloud. Low reflection and low digital number of snow-free land pixels in channel 2 rather than high reflection of snow and cloud pixels is an appropriate tool to separate these two phenomena. Furthermore, channel 4 temperature of snow free land surface is much higher than snow surface or cloud pixels. In the second step, snow surface lands are separated from cloud pixels. To apply this idea, it is necessary to calculate Delta and Gamma; Delta is defined as the Channel 3 and Channel 4 temperature difference in °C and Gamma is $R4/\sqrt{R3}$, where R4 and R3 are the radiances in Channels 4 and 3, respectively, for a given pixel. Clouds tend to have a large value of Delta and Gamma whereas snow tends to have a lower value. Above all, a mixed pixel class is also identified and pixels in this class can be assigned a pereentage composition (cloud, snow, and land) using a linear mixing model. For instance, channel 2 reflection, ehannel 4 temperature, Delta, Gamma, no snow land, cloud, and snow maps on March 23rd, 1993 are shown in Fig. 2.

In this research, more than 60 NOAA-AVHRR satellite images were processed to determine snow cover area (SCA). The images were taken in months of Decemher, January, February, and March of 1989 to 2004 and about 3-4 images in each month were processed to determine SCA.

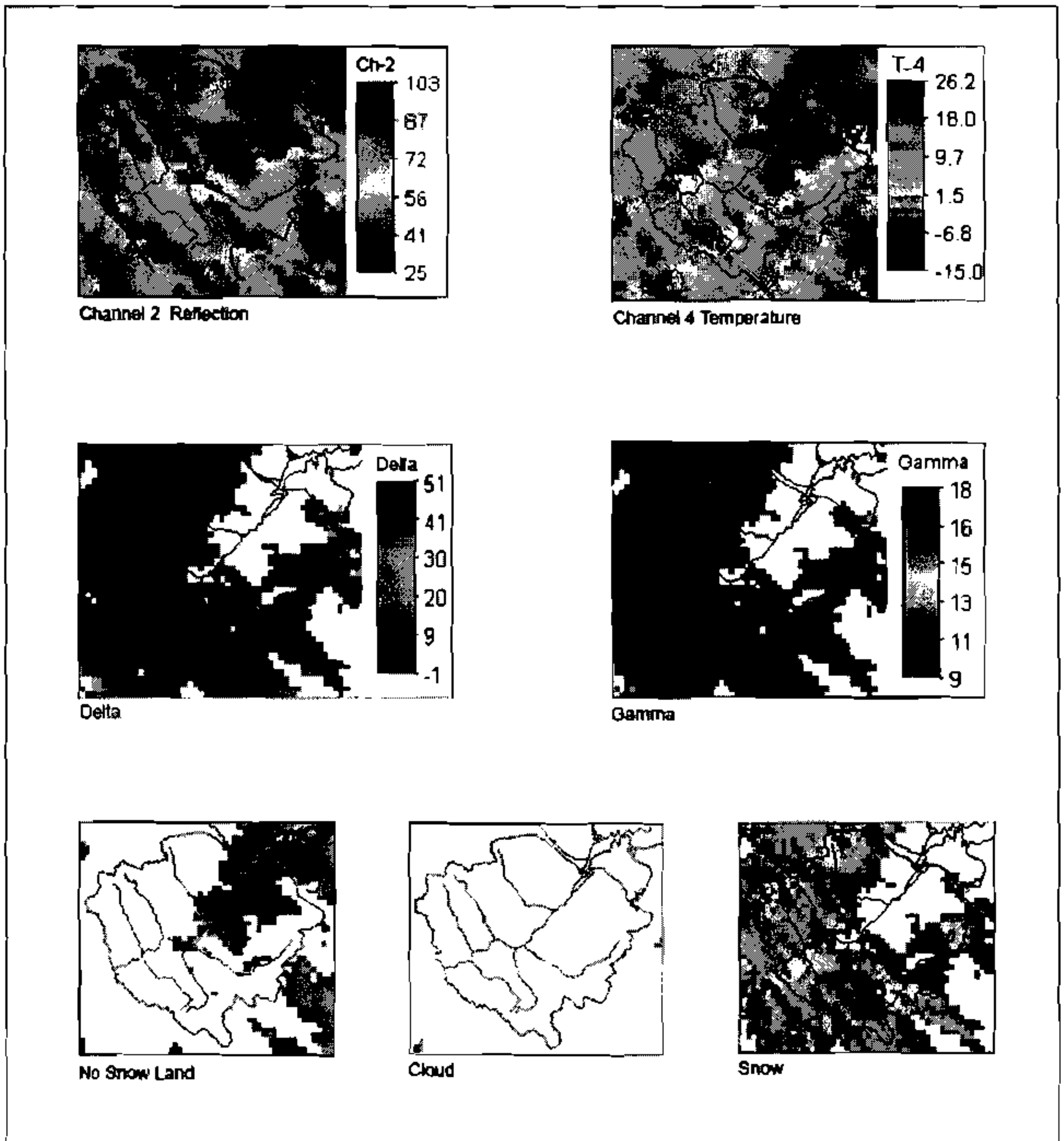


Figure 2. Required Maps in Extracting Snow Cover Map (March 28th, 1993)

METHOD OF FORECASTING

Typical practice in hydrologic forecasting has often included variables in multiple regression equations that describe parameters such as future snow accumulation or precipitation which are unknown at the time the forecast is made. The predictor variables used in hydrologic forecasting are usually correlated with each other, particularly data for different stations of the same data type at the same time (e.g., snow water equivalent on a given date at several snowcourses). Considering the significance of the regression coefficients, standard multiple regression will keep only a very few such variables in the equation. If all of these variables are retained anyway, the coefficients will not be accurately estimated, and they may not make physical sense (e.g., negative coefficients for variables having a positive correlation with streamflow). Such an equation may not give consistently accurate predictions over time and is not conceptually acceptable. If only the few significant variables are retained in the equation, too heavy reliance is placed on a few data sites to represent spatially variable snowpack and precipitation consistently; again one might expect erratic performance over time.

In hydrologic forecasting, the practice of constructing composite indexes that were used as independent variables had the effect of at least partially circumventing the intercorrelation problem. These composite indexes were typically weighted sums of data from stations of the same data type to produce snow, fall precipitation, winter precipitation, and spring precipitation terms. By combining data from several highly correlated data sites into a single variable before entering the regression, the major source of intercorrelation was removed.

The most satisfactory and statistically rigorous way to deal with intercorrelation is the use of principal components regression. Principal components analysis is a statistical technique that restructures a set of intercorrelated variables into an equal number of uncorrected variables.

## Forecasting Regression Models in Zayande-rood sub-basin

Forecasting the Zayande-rood river flow before beginning of April is very important for planning of water allocation to various sectors (i.e., agricultural, industrial, and municipal users). In this research, three separate equations were developed to predict river discharge at three forecast times before beginning of April (beginning of February, March, and April). In each equation, a forecast period (February to July, March to July, and April to July) and a forecast time (beginning of February, March, and April) were defined. For instance, an equation was developed to forecast streamflow volume in February to July (Forecast period) of each year at the beginning of February (forecast time) of the same year using the available data before February. Two other equations were developed to forecast future streamflow volume at the beginning of January and March.

Selecting effective parameters in developing regression models plays an important role in accuracy of forecasted values. In an extensive study conducted by Garen (1992) to improve techniques in regression-based streamflow volume forecasting, Snow Water Equivalent (SWE), precipitation, and antecedent (October-December) streamflow were considered in developing the forecasting equations. In a similar modeling, large-scale climatological parameters (tropical Pacific Sea Surface Temperatures (SST), Pacific-North American (PNA) atmospheric teleconnection and Pacific Decadal Oscillation (PDO) as well as the local precipitation data were used to predict the April–August Columbia River streamflow at Donald, British Columbia, Canada (Hsieh et al, 2003). In both of the above modeling, "Principal Component Regression" method was used.

In this study, hydro-climatological parameters (e.g. SWE, precipitation, and antecedent streamflow), climate signals of SOI (Seasonal Oscillation Index) and NAO (North Atlantic Oscillation), and Snow Cover Area (SCA) data were used in developing forecasting equations.

Three to four NOAA-AVHRR images were processed to calculate SCA during each month (December, January, February, and March). Using these data, maximum, average, and minimum snow cover areas of each four months were estimated. So, 12 data series were developed to resemble effects of SCA in river forecasting equations. The processed snow cover images show that in this sub-basin, snow cover area extends rapidly during December and the maximum snow cover area occurs in February, afterward it starts to shrink and disappears completely in June or July.

Since all SWE measuring stations were established in 1989, a 15 years period (1989 to 2004) of data was used to derive the forecasting equations. The modeling was done in two scenarios for three forecast periods. In the first scenario, all available data series excluding SCA data sets were used to develop forecasting equations, and in the second scenario, SCA data series were also used in modeling procedure. Table 1 shows the derived equations for the two scenarios. Table 2 shows coefficients of determination and standard errors of modeling.

Table 1. Derived Equations With and Without SCA Data Series

| Forecast period | With SCA data sets | Without SCA data sets |
|---|---|---|
| February - July | $Y = 0.49X_1 + 0.76X_2 + 0.84X_3 + 0.32X_4 + 0.17X_5 - 66.89X_6 + 205.39$<br><br>Y= February-July streamflow ($10^6$ $m^3$)<br>$X_1$= SWE in late January at Chelgerd St. (mm)<br>$X_2$= SWE in late January at Daregav St. (mm)<br>$X_3$= SWE in late January at Sooderjan St. (mm)<br>$X_4$= SWE in late January at Qalemorgh St. (mm)<br>$X_5$= Minimum SCA in January (Sq. Km)<br>$X_6$= SOI in October | $Y = 0.54X_1 + 0.78X_2 + 0.87X_3 + 0.35X_4 - 63.23X_5 + 214.56$<br><br>Y= February-July streamflow ($10^6$ $m^3$)<br>$X_1$= SWE in late January at Chelgerd St. (mm)<br>$X_2$= SWE in late January at Daregav St. (mm)<br>$X_3$= SWE in late January at Sooderjan St. (mm)<br>$X_4$= SWE in late January at Qalemorgh St. (mm)<br>$X_5$= SOI in October |
| March - July | $Y = 1.87X_1 + 0.73X_2 + 0.19X_3 - 123.9$<br><br>Y= March-July streamflow ($10^6$ $m^3$)<br>$X_1$= SWE in late February at Chelgerd St. (mm)<br>$X_2$= Precipitation in February at Chelgerd St. (mm)<br>$X_3$= Minimum SCA in January (Sq. Km) | $Y = 2.02X_1 + 0.81X_2 - 143.1$<br><br>Y= March-July streamflow ($10^6$ $m^3$)<br>$X_1$= SWE in late February at Chelgerd St. (mm)<br>$X_2$= Precipitation in February at Chelgerd St. (mm) |
| April - July | $Y = 0.23X_1 + 0.11X_2 + 0.49X_3 + 0.72X_4 + 0.43X_5 + 0.065X_6 + 27.73$<br><br>Y= April-July streamflow ($10^6$ $m^3$)<br>$X_1$= Precipitation of March at Chelgerd St. (mm)<br>$X_2$= Precipitation of October-March at Chelgerd St. (mm)<br>$X_3$= SWE in late January at Chelgerd St. (mm)<br>$X_4$= SWE in late January at Nasirabad St. (mm)<br>$X_5$= SWE in late March at Bardeh St. (mm)<br>$X_6$= Minimum SCA in January (Sq. Km) | $Y = 0.28X_1 + 0.14X_2 + 0.57X_3 + 0.74X_4 + 0.46X_5 + 31.56$<br><br>Y= April-July streamflow ($10^6$ $m^3$)<br>$X_1$= Precipitation of March at Chelgerd St. (mm)<br>$X_2$= Precipitation of October-March at Chelgerd St. (mm)<br>$X_3$= SWE in late January at Chelgerd St. (mm)<br>$X_4$= SWE in late January at Nasirabad St. (mm)<br>$X_5$= SWE in late March at Bardeh St. (mm) |

Table 2. Coefficient of Determination and Standard Error of the Models

| Forecast period | With SCA data sets | | Without SCA data sets | | Improvement, % | |
|---|---|---|---|---|---|---|
| | Coefficient of Determination | Standard Error ($10^6$ $m^3$) | Coefficient of Determination | Standard Error ($10^6$ $m^3$) | Coefficient of Determination | Standard Error |
| February - July | 0.71 | 101 | 0.58 | 115 | 40 | 12 |
| March - July | 0.79 | 78 | 0.69 | 89 | 14 | 9 |
| April - July | 0.90 | 19 | 0.85 | 20 | 6 | 5 |

## CONCLUSION

In this research, principal component regression approach was used to develop forecasting equations with and without SCA data series. Results show that the minimum SCA in December is an effective parameter which can increase coefficient of determination by 6 to 40 percent and decrease standard error by 5 to 12 percent for three forecasting periods. Although SCA parameter is not as informative as some other parameters such as SWE, it can be used as a complementary parameter to improve the accuracy of forecasting equations. It is important to note that in this research, only 15 years of data were available to develop the forecasting equations and longer time series when available may result in different forecast equations which may improve the forecasts.

## ACKNOWLEDGEMENT

The authors would like to thank research staffs of the Environment and Water Research Centre (EWRC) of Sharif university of Technology specially Dr. Hamidreza Taheri for his valuable insights that contributed significantly to the development of the algorithms used in calculating Snow Cover Area. We also benefited from helpful discussions with Mr. Karamalian, Mr. Asadi, and Mr. Kazemi of Isfahan Regional Water Authority. This work was supported by Isfahan Regional Water Authority.

## REFERENCES

Day, G. N. (1985). "Extended streamflow forecasting using NWSRFS." /. Water Resour. Planning and Mgmt., ASCE, 111(2), 157-170.

Druce, D. J. (1984). "Seasonal inflow forecasts by a conceptual hydrologic model for Mica Dam, British Columbia." A critical assessment of forecasting in western water resources management; Proc. of AWRA Symp., J. J. Cassidy and D. P. Lettenmaier, eds., Seattle, Wash., 85-91.

Garen, David C. (1992). "Improved Techniques in Regression-Based Streamflow Volume Forecasting.", Journal of Water Resources Planning and Management, ASCE, Vol. 118 No. 6, 654-670.

Kuehl, D. W. (1979). "Volume forecasts using the SSARR model in a zone mode." Proc, Western Snow Conference, 38-47.

Liu. R.X., Kuang J. , Gong Q. , Hou X.L. , (2002)." Principal Component regression Analysis With SPSS", Computer Methods and programs in Biomedicine , Vol(71), 141-147.

Pearson, T. (1974). "Simulating runoff to the Hungry Horse reservoir of western Montana." Proc, Western Snow Conference, 96-102.

Simpson, J.J., J.R. Stitt, and M.Sienko (1998) "Improved estimates of the areal extent of snow cover from AVHRR data." Journal of Hydrology, No. 204, pp. 1-23.

Speers, D. D., and Versteeg, J. D. (1982). "Runoff forecasting for reservoir operations the past and the future." Proc, Western Snow Conference, 149-156.

Twedt, T. M., Schaake, J. C. Jr., and Peck, E. L. (1977). "National Weather Service extended streamflow prediction." Proc, Western Snow Conference, 52-57.

William W. Hsieh, Yuval, Jingyang Li, Amir Shabbar, and Stephanie Smith (2003), "Seasonal Prediction with Error Estimation of Columbia River Streamflow in British Columbia", Journal of water Resources Planning and Management, Vol. 129, No. 2.