Statistical Validation Methods: Application to Unit Hydrographs

By Bing Zhao,[1] Associate Member, ASCE, Yeou-Koung Tung,[2] Member, ASCE, Keh-Chia Yeh,[3] Associate Member, ASCE, and Jinn-Chuang Yang,[4] Member, ASCE

Abstract: In hydraulic and hydrologic studies, engineers develop new models or calibrate existing models by various techniques. One is often concerned with the model validity regarding its ability to predict future events. Five potentially useful statistical validation methods are presented. For illustration, they were applied to examine the predictability of unit hydrographs derived by various methods in the framework of the least squares and its variations. It was found that storm-stacking (conventional multistorm analysis) together with storm-scaling yields the most desirable UH. The general framework of these validation methods can also be applied to a validation study of other hydrologic and hydraulic models.

INTRODUCTION

Developing a model consists of derivation of a model, validation of the derived model, and maintenance of the validated model. Validation is an important task in the process of developing a model. By validation, one evaluates how well the derived model can predict future events. If the derived model performs well in a validation test, then the model will be useful in prediction. Otherwise, the model is of limited use and its predictions are unreliable.

Model validation can be performed in several ways. One easy approach is to collect new data. However, this approach implies two things: (1) additional costs; and (2) possible postponement of the decision making affected by the model results. Hence, collection of new data for model-validation purposes could be impractical for some real-life problems.

Alternatively, data splitting could be a viable approach. The idea is to divide the observed data set into two subsets, namely, the estimation subset and the validation subset. The estimation subset is used to estimate the model parameters or coefficients. Then, the validation set is used to imitate future random observations to validate the estimated model coefficients. Validation by data splitting is sometimes found in hydrologic studies. The most commonly used practice is to split data into two halves.

In data splitting, how to split the data is an important issue. In the context of the regression model, Snee (1977) gave an excellent description of the DUPLEX algorithm, developed by R. W. Kennard, which finds the estimation subset and the validation subset on the basis of the Euclidean distance between the standardized and orthonormalized observed data points. Bruen and Dooge (1992) applied the DUPLEX algorithm to evaluate a derived UH. Another approach to data splitting regarding the linear regression model was discussed by several authors (Picard and Cook 1984; Picard and Berk 1990). Since these two data splitting methods are only applicable in the context of a regression model, it is necessary to have a general method for data splitting. The cross-validation technique could be considered as a general method (Allen 1971; Stone 1980; Geisser 1975; McCarthy 1976).

A more general method for model validation uses the bootstrap technique (Efron 1977, 1982). The validation by bootstrap has significant appeal when theories like the ordinary least squares is intractable and when it is difficult or impractical to collect additional data for validation. Halfon (1989) used the bootstrap technique in model validation without performing more experiments on fish toxicity. Efron (1983) presented an excellent discussion on the basis of the prediction and on estimating the prediction error using the observed data points. In his paper, several model validation methods including the leave-one-out cross-validation, the bootstrap validation methods (ordinary, randomized, and double bootstrap), the jackknife validation, and the 0.632-estimator validation were discussed.

The unit hydrograph (UH) proposed by Sherman (1932) is defined as a direct runoff hydrograph (DRH) resulting from one unit of effective rainfall (ER) distributed uniformly over a watershed for a specified duration. In UH theory, the watershed is considered as a system with effective rainfall hyetograph (ERH) being the input, DRH being the output, and UH being the kernel function. If the watershed is assumed to be a linear and time-invariant system, one can derive a convolution relationship between the input and output with UH being the transformation function.

Since the derived UH is most likely to be used to predict DRH when effective rainfall data are available, it is important to investigate how well the derived UH can predict the DRH. After the derived UH is evaluated through proper validation procedures and found to have a good predictability of DRH, engineers will have more confidence in using the derived UH to predict the design runoff discharge for designing and evaluating hydrosystems.

In this paper, five statistical validation methods are presented and applied to examine the predictive capability of UHs derived from various methods. The validation methods considered herein are (1) the leave-one-out cross-validation (1CV); (2) the leave-half-out cross-validation with replications (HCV); (3) the ordinary bootstrap validation (BV); and (4) the 0.632-estimator validation (0.632) methods. In addition, a new validation method, called the bootstrap-leave-half-out validation method (HBV), is proposed herein. All five validation methods were applied to examine the predictability of UHs derived from 15 methods that are the combinations of the unconstrained ordinary least squares (OLS) and two types of unconstrained ridge least squares (RLS), together with the multistorm UH and the averaged single-storm UHs. A brief description of the OLS and RLS methods applied to multistorm UH determination is given in Appendix I. The RLS method essentially adds a positive number called

[1]Res. Assoc., Civ. Engrg. Dept., Arizona State Univ., Tempe, AZ 85287.
[2]Prof., Wyoming Water Resour. Ctr. and Stat. Dept., Univ. of Wyoming, Laramie, WY 82071.
[3]Assoc. Prof., Civ. Engrg. Dept., Nat. Chiao-Tung Univ., Hsinchu, Taiwan, R.O.C.
[4]Prof., Civ. Engrg. Dept., Nat. Chiao-Tung Univ., Hsinchu, Taiwan, R.O.C.

Note. Discussion open until January 1, 1996. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on July 28, 1994. This paper is part of the *Journal of Hydraulic Engineering*, Vol. 121, No. 8, August, 1995. ©ASCE, ISSN 0733-9429/95/0008-0618–0624/$2.00 + $.25 per page. Paper No. 8975.

J. Hydraul. Eng. 1995.121:618-624.

ridge parameter k to the diagonal of the design matrix in (13). The first type of unconstrained RLS method is to find the ridge parameter by minimizing the mean-squared-error (MSE) of the UH, and the second type is to find the ridge parameter by minimizing the MSE of the DRH (Zhao 1992).

For the multistorm analysis, storm-scaling and no-storm-scaling were used along with storm-stacking and storm-combining with time alignment at the origin (Zhao 1992). Storm-stacking is the conventional multistorm analysis (Mays and Coles 1980). Storm-combining is to directly add the DRH ordinates and ERH ordinates of all selected storms, respectively. Storm-scaling is to scale all selected storms so that all storms have the same amount of total effective rainfall depth of one unit with the intent to eliminate the dominant effect of larger storms over smaller storms in UH determination (see Appendix I). Even when an engineer is interested in "large storms" for the design, the scaling may still be useful because the storm magnitude, small or large, is only relative and one may have several "large" storms whose magnitudes are not identical.

With 15 methods for UH determination, comparisons were made for the predictability of the resulting UHs from (1) multistorm analysis versus the averaged UHs from single-storm analysis; (2) OLS versus RLS with minimization of the MSE of UH; (3) OLS versus RLS with minimization of the MSE of DRH; (4) two types of RLS methods; (5) storm-scaling versus no-storm-scaling; and (6) storm-stacking versus storm-combining.

The premise for the five validation methods presented in the next section is that the observed storms are considered to be statistically independent random samples. It is generally difficult to theoretically verify statistical independence among different storm events. However, storm independence can be intuitively justified if the selected rainfall-runoff events are not produced by the same storm system. For practical purposes, as long as the time occurrence of the select storm events is sufficiently large, the storm dependence assumption is generally justifiable.

METHODS OF VALIDATION

The presented statistical validation methods (excluding the leave-one-out cross-validation) are based on resampling, by which the population is approximated by resampling the observed storm sample for a large number of times. In doing so, the problem that the number of selected storms is small can be alleviated.

Leave-One-Out Cross-Validation (1CV)

Suppose there are R observed storms. The 1CV method is to leave out one storm from R storms at a time, then derive a UH based on the remaining $R - 1$ storms. Further, use the derived UH to predict the excluded storm. The value of the RMSE for predicting the DRH of the excluded storm is then calculated. After the leave-one-out is carried for each of the R storms, the R values of the RMSE are averaged. Then, the averaged RMSE is used for comparing the predictive performance of the UHs from different methods. The algorithm of the 1CV method can be outlined as follows:

1. Consider R storm events, $\mathbf{S} = (S_1, S_2, \ldots, S_R)$.
2. Let $r = 1$.
3. Leave out storm event S_r from the original data set $\mathbf{S}$.
4. Use the selected method to determine a UH called $UH_{(r)}$ based on the remaining $R - 1$ storms, that is, $(S_1, S_2, S_{r-1}, S_{r+1}, \ldots, S_R)$.
5. Use the derived $UH_{(r)}$ along with the effective rainfall data of the excluded storm event S_r to predict the DRH of that storm event. Then, the corresponding $RMSE_{(r)}$ value is calculated as

$$RMSE_{(r)} = \sqrt{\frac{\sum_{n=1}^{N_r} (q_{r,n} - \hat{q}_{r,n})^2}{N_r}} \quad (1)$$

in which $q_{r,n}$ = the nth observed DRH ordinate of the removed storm event S_r; and $\hat{q}_{r,n}$ = the corresponding predicted DRH ordinate.
6. Repeat steps (3)–(5) for $r = r + 1$ till $r = R$.
7. Compute the average of R RMSEs as the prediction error

$$RMSE_{1CV} = \frac{\sum_{r=1}^{R} RMSE_{(r)}}{R} \quad (2)$$

Leave-Half-Out Cross-Validation (HCV)

Consider that the number of storms R is even. The HCV is to randomly split the storm sample into halves. The UH derived from storms in the first half, using a selected method, is used to predict the DRHs in the second half of the storms. For each storm in the second half, one value of the RMSE can be computed. Then, the mean of $R/2$ values of the RMSE over the second half of the storms is calculated. Let B denote a large integer number. This procedure is repeated B times producing B values of the mean RMSE. Finally, the average value of the B mean RMSEs is calculated and serves as the prediction error. The algorithm of the HCV method can be outlined as follows:

1. Consider R storm events, $\mathbf{S} = (S_1, S_2, \ldots, S_R)$, where R is an even number.
2. Let $b = 1$.
3. Randomly select $R/2$ storms *without* replacement from $\mathbf{S}$ to obtain the first half of the storms, $\mathbf{S}_1 = (S_{1,1}, S_{1,2}, \ldots, S_{1,R/2})$. The remaining storms are the second half of the storms, $\mathbf{S}_2 = (S_{2,1}, S_{2,2}, \ldots, S_{2,R/2})$.
4. Use the selected method under consideration to determine a UH called $UH_{(b)}$, based on storms in set $\mathbf{S}_1$.
5. Use the derived $UH_{(b)}$ to predict the DRH of each storm event in set $\mathbf{S}_2$. Then, calculate the corresponding $RMSE_{(r)}$, $r = 1, 2, \ldots, R/2$, for each storm in set $\mathbf{S}_2$ according to (1).
6. Average $R/2$ values of RMSE over the storms in set $\mathbf{S}_2$ by

$$RMSE_b = \frac{\sum_{r=1}^{R/2} RMSE_{(r)}}{R/2} \quad (3)$$

7. Repeat steps (3)–(6) with $b = b + 1$ till $b = B$.
8. Compute the mean of the B RMSEs as the prediction error

$$RMSE_{HCV} = \frac{\sum_{b=1}^{B} RMSE_{(b)}}{B} \quad (4)$$

Bootstrap Half Validation (HBV)

Suppose that $\mathbf{S} = (S_1, S_2, \ldots, S_R)$ is the observed random sample from an unknown distribution. Let $\hat{F}$ be the empirical probability distribution putting mass $1/R$ on each storm

$$\hat{F}: \text{mass } 1/R \text{ on } S_1, S_2, \ldots, S_R$$

in which $\hat{F}$ implies that each storm is equally sampled. Then,

J. Hydraul. Eng. 1995.121:618-624.

$R/2$ storms are randomly selected, with replacement, from **S** to constitute the first storm set $\mathbf{S}_1$ for the purpose of deriving a UH. Then, a second storm set $\mathbf{S}_2$ containing $R/2$ storms are randomly selected, with replacement, from **S** for prediction purpose. As can be seen, the establishment of the second data set for prediction by this method is different from that of the HCV method. The difference is "without" or "with" replacement sampling.

The UH determined from $\mathbf{S}_1$ is used to predict the DRH of each storm in $\mathbf{S}_2$. For each storm in $\mathbf{S}_2$, the value of RMSE can be found. Then, the mean of $R/2$ RMSE values over the storms in $\mathbf{S}_2$ is calculated. This procedure is repeated B times. Then, the average of the B RMSE means is calculated and used as the prediction error. The advantage of the HBV over HCV is that one can do the validation analysis when the number of storms is small. The algorithm of the HBV method can be outlined as follows:

1. Consider R storm events, $\mathbf{S} = (S_1, S_2, \ldots, S_R)$, where R is an even number.
2. Let $b = 1$.
3. Randomly select $R/2$ storms *with* replacement from **S** to obtain the first set of the storms, $\mathbf{S}_1 = (S_{1,1}, S_{1,2}, \ldots, S_{1,R/2})$.
4. Randomly select $R/2$ storms *with* replacement from **S** to obtain the second set of the storms, $\mathbf{S}_2 = (S_{2,1}, S_{2,2}, \ldots, S_{2,R/2})$.
5. Use the selected method to determine a UH, $\mathrm{UH}_{(b)}$, based on $\mathbf{S}_1$.
6. Use the derived $\mathrm{UH}_{(b)}$ to predict the DRH of each storm event in $\mathbf{S}_2$, then calculate the corresponding $\mathrm{RMSE}_{(r)}$ value for each storm in $\mathbf{S}_2$ according to (1).
7. Average $R/2$ values of the RMSE over the storms in set $\mathbf{S}_2$ according to (3).
8. Repeat steps (3)–(7) with $b = b + 1$ till $b = B$.
9. Compute the mean of the B RMSEs as the prediction error

$$\mathrm{RMSE}_{\mathrm{HBV}} = \frac{\sum_{b=1}^{B} \mathrm{RMSE}_{(b)}}{B} \tag{5}$$

Bootstrap Validation (BV)

Suppose that $\mathbf{S} = (S_1, S_2, \ldots, S_R)$ are the observed random samples from an unknown distribution. Let $\hat{F}$ be the empirical probability distribution placing probability mass $1/R$ on each storm.

The BV procedure is composed of two steps. The first step is to use all original R storms in **S** to derive a UH. Then, use the derived UH to predict the DRHs of these R storms. For each storm in **S**, the value of the RMSE is calculated. Thus, the mean of the R values of the RMSE can be calculated, yielding RMSE_0.

The second step is to randomly select R storms from **S**, with replacement, to construct a random storm sample, called the bootstrap random sample, denoted by $\mathbf{S}^* = (S_1^*, S_2^*, \ldots, S_R^*)$. This second step is repeated B times, yielding B bootstrapped random storm samples $\mathbf{S}_b^* = (S_{b,1}^*, S_{b,2}^*, \ldots, S_{b,R}^*)$, $b = 1, 2, \ldots, B$.

The UH, denoted by UH_b^*, is derived based on $\mathbf{S}_b^*$, $b = 1, 2, \ldots, B$. Then, the derived UH_b^* is used to predict the DRHs of the storms in the original sample **S** from which the value of $\mathrm{RMSE}_{b,r}$ can be calculated, $b = 1, 2, \ldots, B$ and $r = 1, 2, \ldots, R$. For the bth bootstrapped random storm sample, the mean value of RMSE denoted by RMSE_b over R storms is calculated. Then, the mean of B RMSE_b is computed. Also, for each bootstrap random sample $\mathbf{S}_b^*$, one can find the proportion

$$P_r^{*b} = \frac{\text{counts for } \{S_j^* = S_r\}}{R} \tag{6}$$

where $j = 1, 2, \ldots, R$; $r = 1, 2, \ldots, R$; and $b = 1, 2, \ldots, B$. The final prediction error of bootstrap validation is

$$\mathrm{RMSE}_{\mathrm{HBV}} = \mathrm{RMSE}_0 + \sum_{b=1}^{B} \frac{op^{*b}}{B} \tag{7}$$

where

$$op^{*b} = \sum_{r=1}^{R} \left(\frac{1}{R} - P_r^{*b}\right) \cdot \mathrm{RMSE}_b \tag{8}$$

The algorithm of the BV method can be outlined as follows:

1. Consider R storm events, $\mathbf{S} = (S_1, S_2, \ldots, S_R)$.
2. Derive a UH based on **S** using a selected method. Then, use the derived UH to regenerate the DRHs of the storms in **S** to compute the RMSE for each storm in **S**. Average the R values of RMSE in **S** to yield RMSE_0.
3. Let $b = 1$.
4. Randomly select R storms *with* replacement from **S** to obtain the bootstrap random sample $S_b^* = (S_{b,1}^*, S_{b,2}^*, \ldots, S_{b,R}^*)$.
5. For the bootstrap sample $\mathbf{S}_b^*$, compute the proportion based on (5).
6. Use the selected method to determine a UH, UH_b^*, based on the storms in bootstrap sample $\mathbf{S}_b^*$.
7. Use the derived UH_b^* to predict the DRH of each storm event in the original random sample **S**, then calculate the corresponding $\mathrm{RMSE}_{b,r}$ according to (1).
8. Average the R values of $\mathrm{RMSE}_{b,r}$ for each bth bootstrap sample, $r = 1, 2, \ldots, R$

$$\mathrm{RMSE}_b = \frac{\sum_{r=1}^{R} \mathrm{RMSE}_{b,r}}{R} \tag{9}$$

9. Repeat steps (3)–(7) with $b = b + 1$ till $b = B$.
10. The final prediction error of the bootstrap validation can be computed by (8)–(9).

0.632-Estimation Validation (0.632BV)

The 0.632BV method was proposed by Efron (1983). As in the bootstrap validation method, B bootstrap random samples are obtained. Also, the quantity P_r^{*b} is calculated. The derived UH_b^* is used to predict the storms in **S** for which $P_r^{*b} = 0$, $b = 1, 2, \ldots, B$ and $r = 1, 2, \ldots, R$. Then, average the values of the RMSE to get RMSE_1. Therefore, the final prediction error by the 0.632BV method is

$$\mathrm{RMSE}_{0.632\mathrm{BV}} = 0.368\mathrm{RMSE}_0 + 0.632\mathrm{RMSE}_1 \tag{10}$$

The algorithm of the 0.632BV method can be outlined as follows:

1. Consider R storm events, $\mathbf{S} = (S_1, S_2, \ldots, S_R)$.
2. Derive a UH using a selected method based on the storm in **S**. Then use the derived UH to regenerate the DRHs of the storms in **S** to compute the RMSE for each storm in **S**. Average the R values of RMSE in **S** to yield RMSE_0.
3. Let $b = 1$.
4. Randomly select R storms, with replacement, from **S** to obtain the bootstrap random sample, $\mathbf{S}_b^* = (S_{b,1}^*, S_{b,2}^*, \ldots, S_{b,R}^*)$.
5. For each bootstrap random sample $\mathbf{S}_b^*$, compute the proportion based on (6).

J. Hydraul. Eng. 1995.121:618-624.

6. Use the selected method to determine a UH, UH_b^*, based on the storms in bootstrap sample S_b^*.
7. Repeat steps (3)–(6) with $b = b + 1$ till $b = B$.
8. Use the derived UH_b^* to predict the DRH of each storm event in the original storm sample **S** for $P_r^{*b} = 0$, $b = 1, 2, \ldots, B$ and $r = 1, 2, \ldots, R$, then calculate the corresponding RMSEs according to (1). Average these RMSEs to yield $RMSE_1$.
9. The final prediction error for the 0.632BV method is computed according to (10).

APPLICATION AND DISCUSSIONS

Storms from three watersheds were used: one watershed is from Bree's paper (1978) with 20 storms and two others (Lan-Yang watershed and Tong-Tou watershed) are in Taiwan, with 10 and nine storms, respectively. The 20 storms in Bree's paper (1978) occurred in 1958 (five storms in winter), 1959 (nine storms in winter) and 1960 (three storms in winter and three storms in summer) in the Nenagh River at the catchment upstream of Clarianna with an area of 295 km^2. The time between observations of rainfall and discharge was chosen as large as possible to minimize the amount of data to be handled, and also to reduce the degree of collinearity in the rainfall series (Bree 1978). The peak discharges of DRH for these 20 storms vary from 9.44 m^3/s to 41.6 m^3/s.

The drainage areas for Lan-Yang watershed and Tong-Tou watershed in Taiwan are 820.69 km^2 and 259.2 km^2, respectively. The 19 storms selected in the analysis are from typhoon events occurred in summer season. The 10 storms for Lan-Yang watershed occurred from 1980 to 1987, and the nine storms for Tong-Tou watershed occurred from 1970 to 1981. The peak discharges of DRH for Lan-Yang and Tong-Tou vary from 1,163 to 2,929 m^3/s and 955 to 3,149 m^3/s, respectively.

For the purpose of reducing the amount of computation, the adopted UH duration was 3 h. All five validation methods were applied to evaluate the UHs derived from 15 methods that are unconstrained OLS and two types of unconstrained RLS, together with multistorm analysis and single-storm analysis. The first type of RLS, called RLS/UH herein, is to find the ridge parameter that minimizes the MSE of the UH. The second type, called RLS/DR, is to find the ridge parameter that minimizes the MSE of the DRH. For multistorm analysis, storm-scaling and no-storm-scaling were used along with storm-stacking and storm-combining with time alignment at the origins of the DRH.

The comparison is based on the prediction error criterion, which should be consistent with the objective function or the prediction rule. Because the least-squares method and its variations were used to derive the UH, the mean of the RMSE is adopted as the prediction error for comparison. It should be noted that if the determination of model performance was based on another prediction rule, the validation methods could also be applied by comparing the prediction error related to the prediction rule. Sometimes, hydrologists are interested in comparing how the UHs by different methods can predict the peak discharge of the DRH. This type of comparison can only serve as a reference because the error of peak discharge is not the objective function to be minimized by least-squares methods. Although assigning higher weight to the error in peak discharge yields a UH that can regenerate a more accurate peak of DRH, there is no guarantee that this UH will predict a more accurate peak of DRH for other storms, because regeneration and prediction are different (Zhao and Tung 1994). If the hydrologists are more interested in predicting the peak of DRH, a different method, one that directly minimizes peak discharge error, should be used. Herein, the prediction rule used in UH validation evaluation is the RMSE, which is consistent with the various LS methods used in this study.

Efron (1983) discussed the validation problems in a general context and suggested that $B = 200$ would give accurate estimation of the prediction error. Results of validation study are presented in Tables 1–3, each corresponding to a watershed. A general look of Tables 1–3 indicates that the prediction errors for Taiwan's two watersheds are much larger than those for Bree's watershed. This may be because the runoff magnitudes for storms in the two Taiwan watersheds are much larger than those from Bree's.

Multistorm UH versus Averaged Single-Storm UH

If the OLS is used, one can observe that, from Tables 1–3, the averaged RMSE for DRHs based on the averaged single-storm UHs, in most cases, is larger than that for the UH from the storm-stacking. However, in most cases, it is smaller than that for the UH from the storm-combining with time-origin alignment.

The same observation can be made for RLS/UH and RLS/DR as that for the OLS. Accordingly, the predictability of the averaged single-storm UHs is less desirable than that of the storm-stacking, but more desirable than that of the storm-combining with time alignment at the origin. Storm-combining is inferior mainly because it is essentially a single-storm procedure.

OLS versus RLS/UH and OLS versus RLS/DR

For UHs from the storm-stacking and from the single-storm averaging, RLS/UH is not better than the OLS. For the storm-stacking with scaling, the predictability of RLS/UH is even worse. It can also be observed that, when using the storm-combining procedure, the averaged RMSE for DRHs for the RLS/UH, in most cases, is smaller than that for the OLS. It appears that RLS/UH is effective only when the two-norm condition number of the ERH matrix is large. This is the case for the storm-combining. If the two-norm condition number is small (i.e., the storm-stacking with the storm-scaling), the RLS/UH procedure does not have advantage over the OLS method to derive a UH.

For a given watershed, by using the storm-stacking with scaling, there is no significant difference in the averaged RMSE values between the OLS and RLS/DR procedures. For the storm-stacking without storm-scaling, the predictability of the RLS/DR was found better than the OLS in most cases. Also, the RLS/DR is better than the OLS using the storm-combining with time-origin alignment in most cases. By using single-storm averaging, there is no significant difference between the OLS and RLS/DR. Observations from Tables 1–3 indicate that the RLS/DR procedure is useful only when the two-norm condition number is large when the storm-combining is used. Otherwise, it is not necessary to use the RLS/DR.

RLS/UH versus RLS/DR

Using the storm-stacking with scaling, the predictability of the resulting UH by the RLS/DR is better than that of the RLS/UH in most cases. For the storm-stacking without scaling, in more than half of the cases considered, the RLS/DR is better than the RLS/UH. For the storm-combining with storm-scaling, the RLS/DR is better than the RLS/UH for 11 cases out of 15. Using the storm-combining without storm-scaling one cannot draw a clear conclusion with regard to the superiority of predictability of the RLS/UH and RLS/DR. For the single-storm averaging, the RLS/DR performs better than the RLS/UH in most cases.

J. Hydraul. Eng. 1995.121:618-624.

TABLE 1. Averaged RMSE Associated with UHs from Different Methods by Five Validation Techniques for Watershed in Bree (1978)

		Validation Method				
		1CV	HCV	HBV	BV	0.632BV
(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLS	S/S	2.1731e + 00	2.2340e + 00	2.1891e + 00	2.0647e + 00	2.1720e + 00
OLS	S/N	2.2643e + 00	2.3457e + 00	2.2469e + 00	2.1041e + 00	2.2608e + 00
OLS	C/S	2.1963e + 00	2.3554e + 00	2.2651e + 00	2.0852e + 00	2.1878e + 00
OLS	C/N	2.2982e + 00	2.4137e + 00	2.4139e + 00	2.1616e + 00	2.3098e + 00
OLS	AVG.	2.2580e + 00	2.3834e + 00	2.3661e + 00	2.1489e + 00	2.2915e + 00
RLS/UH	S/S	2.2027e + 00	2.2079e + 00	2.2405e + 00	2.1013e + 00	2.2142e + 00
RLS/UH	S/N	2.2695e + 00	2.3175e + 00	2.2911e + 00	2.1255e + 00	2.2759e + 00
RLS/UH	C/S	2.1881e + 00	2.3004e + 00	2.2263e + 00	2.0788e + 00	2.2170e + 00
RLS/UH	C/N	2.2850e + 00	2.3914e + 00	2.3357e + 00	2.1512e + 00	2.2632e + 00
RLS/UH	AVG.	2.2254e + 00	2.2900e + 00	2.3355e + 00	2.1237e + 00	2.2157e + 00
RLS/DR	S/S	2.1770e + 00	2.2557e + 00	2.1299e + 00	2.0709e + 00	2.1605e + 00
RLS/DR	S/N	2.2578e + 00	2.3225e + 00	2.2407e + 00	2.1033e + 00	2.2003e + 00
RLS/DR	C/S	2.1893e + 00	2.2853e + 00	2.2866e + 00	2.0795e + 00	2.1773e + 00
RLS/DR	C/N	2.2873e + 00	2.3641e + 00	2.3812e + 00	2.1523e + 00	2.2931e + 00
RLS/DR	AVG.	2.2175e + 00	2.2996e + 00	2.1987e + 00	2.1143e + 00	2.2328e + 00

Note: S/S = storm-stacking with storm-scaling; S/N = storm-stacking without storm-scaling (or no-storm-scaling); C/S = storm-combining with storm-scaling; C/N = storm-combining without storm-scaling (or no-storm-scaling); and AVG. = averaged single-storm UH.

TABLE 2. Averaged RMSE Associated with UHs from Different Methods by Five Validation Techniques for Lan-Yang Watershed in Taiwan

		Validation Method				
		1CV	HCV	HBV	BV	0.632BV
(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLS	S/S	1.7504e + 02	1.8417e + 02	1.7598e + 02	1.6132e + 02	1.7377e + 02
OLS	S/N	1.7823e + 02	1.8643e + 02	1.7982e + 02	1.6226e + 02	1.7993e + 02
OLS	C/S	2.0315e + 02	2.4003e + 02	2.3399e + 02	1.9463e + 02	2.1836e + 02
OLS	C/N	2.4085e + 02	2.5452e + 02	2.3922e + 02	2.1242e + 02	2.3875e + 02
OLS	AVG.	1.78094 + 02	1.8826e + 02	1.8138e + 02	1.6275e + 02	1.7568e + 02
RLS/UH	S/S	1.7710e + 02	1.8571e + 02	1.7945e + 02	1.6379e + 02	1.8067e + 02
RLS/UH	S/N	1.7805e + 02	1.8683e + 02	1.7890e + 02	1.6231e + 02	1.7851e + 02
RLS/UH	C/S	1.9144e + 02	2.1751e + 02	2.0017e + 02	1.7375e + 02	1.9777e + 02
RLS/UH	C/N	2.0501e + 02	2.2645e + 02	2.1225e + 02	1.8163e + 02	2.1192e + 02
RLS/UH	AVG.	1.7977e + 02	1.8691e + 02	1.8292e + 02	1.6630e + 02	1.7998e + 02
RLS/DR	S/S	1.7585e + 02	1.8064e + 02	1.7083e + 02	1.6253e + 02	1.7223e + 02
RLS/DR	S/N	1.7738e + 02	1.8702e + 02	1.7777e + 02	1.6200e + 02	1.7409e + 02
RLS/DR	C/S	1.9129e + 02	2.1115e + 02	2.0779e + 02	1.7346e + 02	1.9653e + 02
RLS/DR	C/N	2.0234e + 02	2.2021e + 02	2.1311e + 02	1.8011e + 02	2.0991e + 02
RLS/DR	AVG.	1.7834e + 02	1.8877e + 02	1.8097e + 02	1.6486e + 02	1.7840e + 02

Note: S/S = storm-stacking with storm-scaling; S/N = storm-stacking without storm-scaling (or no-storm-scaling); C/S = storm-combining with storm-scaling; C/N = storm-combining without storm-scaling (or no-storm-scaling); and AVG. = averaged single-storm UH.

TABLE 3. Averaged RMSE Associated with UHs from Different Methods by Five Validation Techniques for Tong-Tou Watershed in Taiwan

		Validation Method				
		1CV	HCV	HBV	BV	0.632BV
(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLS	S/S	1.2326e + 02	1.3013e + 02	1.2619e + 02	1.1440e + 02	1.2133e + 02
OLS	S/N	1.3234e + 02	1.3829e + 02	1.2774e + 02	1.1514e + 02	1.2925e + 02
OLS	C/S	1.6790e + 02	1.7864e + 02	1.7033e + 02	1.4054e + 02	1.6598e + 02
OLS	C/N	1.6253e + 02	1.8701e + 02	1.7060e + 02	1.3447e + 02	1.6198e + 02
OLS	AVG.	1.3273e + 02	1.4215e + 02	1.3689e + 02	1.2150e + 02	1.3364e + 02
RLS/UH	S/S	1.2475e + 02	1.2887e + 02	1.2432e + 02	1.1681e + 02	1.2488e + 02
RLS/UH	S/N	1.2899e + 02	1.3579e + 02	1.2963e + 02	1.1564e + 02	1.2566 + 02
RLS/UH	C/S	1.7154e + 02	1.8286e + 02	1.6448e + 02	1.5428e + 02	1.7498e + 02
RLS/UH	C/N	1.5663e + 02	1.7333e + 02	1.6818e + 02	1.3205e + 02	1.5398e + 02
RLS/UH	AVG.	1.3617e + 02	1.4296e + 02	1.3959e + 02	1.2687e + 02	1.3629e + 02
RLS/DR	S/S	1.2312e + 02	1.2832e + 02	1.2355e + 02	1.1480e + 02	1.2434e + 02
RLS/DR	S/N	1.3008e + 02	1.3472e + 02	1.2305e + 02	1.1467e + 02	1.2590e + 02
RLS/DR	C/S	1.6527e + 02	1.7232e + 02	1.5893e + 02	1.4587e + 02	1.5936e + 02
RLS/DR	C/N	1.5731e + 02	1.7474e + 02	1.6633e + 02	1.3205e + 02	1.5523e + 02
RLS/DR	AVG.	1.3301e + 02	1.3676e + 02	1.3783e + 02	1.2334e + 02	1.3316e + 02

Note: S/S = storm-stacking with storm-scaling; S/N = storm-stacking without storm-scaling (or no-storm-scaling); C/S = storm-combining with storm-scaling; C/N = storm-combining without storm-scaling (or no-storm-scaling); and AVG. = averaged single-storm UH.

J. Hydraul. Eng. 1995.121:618-624.

Storm-Scaling versus No-Storm-Scaling

For the OLS, RLS/UH, and RLS/DR, the predictability of the storm-scaling is better than that of no-storm-scaling in most storm-stacking cases. In most cases for the Bree's watershed and Lan-Yang watershed, the storm-scaling is better than no-storm-scaling for the storm-combining as well. However, in the Tong-Tou watershed, there is no clear evidence indicating the superiority of the storm-scaling over no-storm-scaling for the storm-combining.

Storm-Stacking versus Storm-Combining

In most cases the storm-stacking results in better predictability than does the storm-combining under both the storm-scaling and no-storm-scaling conditions when the UH is derived by the OLS, RLS/UH, and RLS/DR.

SUMMARY AND CONCLUSIONS

Five statistical validation methods were presented and applied to evaluate the predictability of UHs derived by 15 methods. A total of 39 storm events from three watersheds were used in the analysis. The evaluation was based on the prediction error represented by the averaged RMSE that is consistent with the objective function of the adopted least-squares methods. From the numerical results, one can conclude that, among the 15 methods, a UH obtained from the storm-stacking with scaling in the framework of OLS and its variations is the most desirable in predicting the future storm events.

The estimation of an optimal UH, in essence, is an exercise of parameter estimation or model calibration that is often encountered in hydrologic and hydraulic engineering practices. The validation methods presented here can be applied equally well to the calibration of other models in hydrologic and hydraulic analyses. For example, if one adopted a particular method for model calibration, the validation methods presented can be applied to assess the predictive error associated with the method. On the other hand, if one had several methods for model calibration, the validation methods can be applied to evaluate the predictability of the model calibrated by the different methods.

When the available data are limited, methods such as leave-half-out validation, bootstrap-half validation, bootstrap validation, and 0.632 validation have significant appeal because they are based on resampling procedure. By resampling, one can "generate" a simulated population. Some words of caution should be said regarding the use of the validation methods in hydrologic and hydraulic applications. First, the resampling procedures used in cross-validation are based on the condition that data are statistically independent. This should be checked or justified before the methods can be used. When data are correlated, resampling techniques cannot be directly applied. In such circumstances, appropriate transformations can be made to break the correlation in the original data and apply resampling techniques to the transformed uncorrelated data. Second, one should check the consistency and homogeneity of hydrologic events selected for model calibration. Consider the context of UH estimation as an example. Although it is desirable to use as many independent storms as possible, one should make sure that the selected storms occur in a time period during which the hydrologic conditions in the watershed are considered stationary.

ACKNOWLEDGMENTS

This study was supported by the Agricultural Council of the Executive Yuan, Republic of China. The writers are thankful to Wen-Zhang Hu of the Agricultural Council for his encouragement and to Yu-Chuang Huang of Taiwan Provincial Bureau of Water Conservancy for her supply of data. The writers are grateful to the anonymous reviewers for their in-depth comments and helpful suggestions.

APPENDIX I. DETERMINATION OF UNIT HYDROGRAPH BY LEAST SQUARE PROCEDURE AND ITS VARIATIONS

In deriving a UH using multistorm analysis, one considers solving the following equation involving R storms:

$$\mathbf{Pu} = \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_R \end{pmatrix} \mathbf{bu} = \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \vdots \\ \mathbf{q}_R \end{pmatrix} = \mathbf{q} \quad (11)$$

where $\mathbf{P}_r$ and $\mathbf{q}_r$ = the ERH matrix and DRH vector of the rth storm, respectively; and $\mathbf{u}$ = the vector of UH ordinates to be determined. This equation is for stacking individual storms. Note that the number of ordinates of the unknown UH $\mathbf{u}$ is suggested to be $J = \max\{J_r\}$ where $J_r = N_r - M_r + 1$ (Diskin and Boneh 1975; Bree 1978; Singh 1988) with M_r, N_r, and J_r being the number of ordinates for ERH, DRH, and UH of the rth storm. Thus, before stacking the individual equations, the DRHs for individual storms should be adjusted.

The multistorm UH by the unconstrained OLS method for a watershed is

$$\mathbf{u}_{ols/u} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{q} = \left(\sum_{r=1}^{R} \mathbf{P}_r'\mathbf{P}_r\right)^{-1} \left(\sum_{r=1}^{R} \mathbf{P}_r'\mathbf{q}_r\right) \quad (12)$$

Using the ridge regression procedure, the multistorm UH can be obtained as

$$\mathbf{u}_{rls/u} = (\mathbf{P}'\mathbf{P} + k\mathbf{I})^{-1}\mathbf{P}'\mathbf{q} = \left(\sum_{r=1}^{R} \mathbf{P}_r'\mathbf{P}_r + k\mathbf{I}\right)^{-1} \left(\sum_{r=1}^{R} \mathbf{P}_r'\mathbf{q}_r\right) \quad (13)$$

where k = a positive-valued number, $k > 0$, called the ridge parameter. The ridge parameter k in (13) can be determined based on two criteria: (1) minimization of the MSE of the derived UH, MSE($\hat{\mathbf{u}}$); and (2) minimization of the MSE of predicted DRH, MSE($\hat{\mathbf{q}}$). The MSE($\hat{\mathbf{u}}$) measures the expected Euclidean distance between the estimated UH($\hat{\mathbf{u}}$) and the true but unknown UH($\mathbf{u}$), whereas the MSE($\hat{\mathbf{q}}$) measures the expected Euclidean distance between the estimated DRH($\hat{\mathbf{q}}$) and the true but unknown DRH($\mathbf{q}$). Mathematically, the two MSEs can be expressed as

$$\text{MSE}(\hat{\mathbf{u}}) = E[(\hat{\mathbf{u}} - \mathbf{u})'(\hat{\mathbf{u}} - \mathbf{u})] \quad (14)$$

$$\text{MSE}(\hat{\mathbf{q}}) = E[(\hat{\mathbf{q}} - \mathbf{q})'(\hat{\mathbf{q}} - \mathbf{q})] = E[(\hat{\mathbf{u}} - \mathbf{u})'\mathbf{P}'\mathbf{P}(\hat{\mathbf{u}} - \mathbf{u})] \quad (15)$$

Based on Hoerl and Kennard (1970, 1976) and Lee and Campbell (1985), the MSE($\hat{\mathbf{u}}$) and MSE($\hat{\mathbf{q}}$) can further be expressed, respectively, as

$$\text{MSE}(\hat{\mathbf{u}}) = \sum_{j=1}^{J} \frac{\lambda_j\sigma^2}{(\lambda_j + k)^2} = \sum_{j=1}^{J} \frac{k^2\alpha_j^2}{(\lambda_j + k)^2} \quad (16)$$

$$\text{MSE}(\hat{\mathbf{q}}) = \sum_{j=1}^{J} \frac{\lambda_j^2\sigma^2}{(\lambda_j + k)^2} + \sum_{j=1}^{J} \frac{\lambda_j k^2\alpha_j^2}{(\lambda_j + k)^2} \quad (17)$$

where $\lambda_j (j = 1, 2, \ldots, J)$ are the eigenvalues of $\mathbf{P}'\mathbf{P}$; σ^2 can be estimated by $\|\mathbf{q} - \mathbf{Pu}_{ols/u}\|^2/(N - J)$; $\boldsymbol{\alpha} = \mathbf{H}'\mathbf{u}_{ols/u}$ with $\mathbf{H}$ being an eigenvector matrix of $\mathbf{P}'\mathbf{P}$. Appropriate numerical optimization techniques can be applied to determine the values of ridge parameter that minimize MSE($\hat{\mathbf{u}}$) and MSE($\hat{\mathbf{q}}$),

J. Hydraul. Eng. 1995.121:618-624.

respectively. The optimal ridge parameter then is used in (13) to compute the corresponding UH.

Consider that in (11) some storms have DRH ordinates that are significantly higher than those of other storms. In this situation, the value of objective function (whichever is adopted) containing deviations between the observed and computed DRH ordinates is more sensitive to large storms than smaller ones. Hence the resulting UH from minimizing an error criterion is dominated by larger storms. This implies that the inclusion of smaller storms is redundant and has little or no contribution to the determination of the representative UH in the watershed. Neglecting the system information embodied in smaller stones could lead to a biased estimation of the UH.

The scaling technique is used to take into account the effect of storm magnitude in multistorm analysis by dividing the ERH and DRH ordinates by the ER amount of the corresponding storm. That is

$$\tilde{\mathbf{p}}_r = \frac{\mathbf{p}_r}{\mathbf{1}'\mathbf{p}_r}, \quad r = 1, 2, \ldots, R \tag{18}$$

$$\tilde{\mathbf{q}}_r = \frac{\mathbf{q}_r}{\mathbf{1}'\mathbf{p}_r}, \quad r = 1, 2, \ldots, R \tag{19}$$

where $\tilde{\mathbf{p}}_r = \tilde{\mathbf{q}}_r$ = scaled (or standardized) ERH and DRH for the rth storm, respectively. Through the scaling procedure, (11) can be modified as

$$\tilde{\mathbf{P}}\mathbf{u} = \begin{pmatrix} \tilde{\mathbf{P}}_1 \\ \tilde{\mathbf{P}}_2 \\ \vdots \\ \tilde{\mathbf{P}}_R \end{pmatrix} \mathbf{u} = \begin{pmatrix} \tilde{\mathbf{q}}_1 \\ \tilde{\mathbf{q}}_2 \\ \vdots \\ \tilde{\mathbf{q}}_R \end{pmatrix} = \tilde{\mathbf{q}} \tag{20}$$

The scaled $\tilde{\mathbf{q}}_r$ can be viewed as the DRH resulting from a total of one unit of ER. Hence, the scaled DRH has the same dimension as the UH. Note that the original DRH is affected by both the total amount of ER and its temporal distribution. The scaling procedure removes the influence of ER amount on the UH determination; leaving the temporal distribution of ER with a total of one unit as the sole factor affecting the shape of UH. Eqs. (12) and (13) for the OLS and RLS procedures are also applicable to the scaled ERHs and DRHs by simply replacing $\mathbf{P}$ by $\tilde{\mathbf{P}}$ and $\mathbf{q}$ by $\tilde{\mathbf{q}}$.

APPENDIX II. REFERENCES

Allen, D. M. (1971). "The prediction sum of squares as a criterion for selecting variables." *Tech. Rep. No. 23*, Dept. of Stat., University of Kentucky, Lexington, Ky.

Bree, T. (1978). "The stability of parameter estimation in the general linear model." *J. Hydro.*, 37(1/2), 47–66.

Bruen, M., and Dooge, J. C. I. (1992). "Unit hydrograph estimation with multiple events and prior information: II. Evaluation of the method." *Hydro. Sci. J.*, 37(5), 445–462.

Diskin, M. H., and Boneh, A. (1975). "Determinations of an optimal IUH for linear, time invariant systems from multi-storm records." *J. Hydro.*, 24(1/2), 57–76.

Efron, B. (1977). "Bootstrap methods: another look at the jackknife." *Rep. No. 32*, Div. of Biostatistics, Stanford University, Stanford, Calif.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans, CBMS 38*, SIAM, Philadelphia, Pa.

Efron, B. (1983). "Estimating the error rate of a prediction rule: improvement on cross-validation." *J. Am. Stat. Assoc.*, 78(382), 316–331.

Geisser, S. (1975). "The predictive sample reuse method with applications." *J. Am. Stat. Assoc.*, 70(350), 320–328.

Halfon, E. (1989). "Probabilistic validation of computer simulations using the bootstrap." *Ecological Modelling*, 46(3/4), 213–219.

Hoerl, A. E., and Kennard, R. W. (1970). "Ridge regression: biased estimation for nonorthogonal problems." *Technometrics*, 12(1), 55–67.

Hoerl, A. E., and Kennard, R. W. (1976). "Ridge regression iterative estimation of the biasing parameter." *Communications in Stat.—Theory and Meth.*, A5(1), 77–88.

Lee, T. S., and Campbell, D. B. (1985). "Selecting the optimum k in ridge regression." *Com. Stat.—Theor. Meth.*, 14(7), 1589–1604.

Mays, L. W., and Coles, L. (1980). "Optimization of unit hydrograph determination." *J. Hydr. Div.*, ASCE, 106(1), 85–97.

McCarthy, P. J. (1976). "The use of balanced half-sample replication in cross-validation studies." *J. Am. Stat. Assoc.*, 71(355), 596–604.

Picard, R. R., and Cook, R. D. (1984). "Cross-validation of regression models." *J. Am. Stat. Assoc.*, 79(387), 575–583.

Picard, R. R., and Berk, K. N. (1990). "Data splitting." *Am. Stat.*, 44(2), 140–147.

Sherman, L. K. (1932). "Stream flow from rainfall by unit-graph method." *Engrg. News Record*, 108(Apr. 7), 501–505.

Singh, V. P. (1988). *Hydrologic systems*, Vol. I., Prentice Hall, Inc., Englewood Cliffs, N.J.

Snee, R. D. (1977n). "Validation of regression models: methods and examples." *Technometrics*, 19(4), 415–428.

Stone, M. (1980). "Cross-validatory choice and assessment of statistical predictions." *J. Royal Stat. Soc., Series B*, 36, 111–147.

Zhao, B. (1992). "Determination of a unit hydrograph and its uncertainty applied to reliability analysis of hydraulic structures," MSc thesis, University of Wyoming, Laramie, Wyo.

Zhao, B., and Tung, Y-K. (1994). "Determination of optimal unit hydrograph by linear programming." *Water Resour. Mgmt.*, 8, 101–119.

APPENDIX III. NOTATION

The following symbols are used in this paper:

- B = number of resampling;
- $\mathbf{H}$ = eigenvector matrix of $\mathbf{P}'\mathbf{P}$;
- J_r = number of ordinates for UH of rth storm;
- k = ridge parameter;
- M_r = number of ordinates for ERH of rth storm;
- MSE($\hat{\mathbf{q}}$) = mean squared error of DRH;
- MSE($\hat{\mathbf{u}}$) = mean squared error of UH;
- N_r = number of ordinates for DRH of rth storm;
- $\mathbf{P}_r$ = ERH matrix of rth storm;
- $\mathbf{p}_r$ = ERH vector of rth storm;
- $\tilde{\mathbf{p}}_r$ = scaled ERH vector for rth strom;
- $\mathbf{q}_r$ = DRH vector of rth storm;
- $\tilde{\mathbf{q}}_r$ = scaled DRH vector for rth storm;
- $\mathbf{S}$ = random sample consisting of storms;
- $\mathbf{u}_r$ = vector of UH ordinates of rth storm; and
- λ_j = eigenvalue of $\mathbf{P}'\mathbf{P}$.

J. Hydraul. Eng. 1995.121:618-624.