CHAPTER 5

MODEL TESTING ON STREAMFLOWS: II. ANNUAL MAXIMUM FLOWS

5.1 RESULTS AND DISCUSSION

5.1.0 Preamble

The methodology was again tested on the three streamflow gauges, i.e. D1H003, D1H006 and D1H009 as depicted in Figure A.1 of Appendix A and section 3.4 of Chapter 3. The regime of the time-series considered was the annual maximum flows of these gauges (1960-1989). This type of data is used for design-oriented studies such as reservoirs, bridges, etc. Thus, the accuracy of the estimated values is of great importance as it may have an impact on the design of these hydraulic structures.

The theory of the EM algorithms and the entropy computations are much easier for normal distributions than for other distributions. Thus, the first assumption of the serial independence of the annual maximum flows at each gauge was tested by computing the value of the first order serial correlation of the flow series at each gauge. It was found that the values of the first order serial correlation were 0.127, 0.16, and 0.173 for gauges D1H003, D1H006, and D1H009 respectively. It was concluded that the independence assumption could not be rejected statistically at 95 % confidence interval. As mentioned by Elshorbagy et al. (2000b), by just assuming a threshold value of 0.5 for the correlation coefficient (below which the data series is considered to be insignificantly correlated), one can say that the annual maximum flow series at these gauges are assumed to be independent. The second assumption, normality, is not always met, as real data is rarely normal, especially for a set of natural streamflow. One way of meeting this assumption was to transform strategically the data to follow approximately the normal distribution. It was found in this study the Box-Cox family of transformations to be suitable. The annual maximum flows series for the three gauges were approximated normally distributed with a square root transformation.

5.1.1 Selection of base/subject station

The entropy computations for the respective gauges are shown in Tables 5.1 (a-c). The cross-correlation values of lag zero for the station pairs, i.e. D1H003- D1H006, D1H003-D1H009 and D1H006- D1H009 were found to be 0.7679, 0.8963 and 0.74041 respectively.

Table 5.1a Marginal entropy of annual maximum flows for different gauges

	D1H003	D1H006	D1H009
Marginal entropy (napiers)	3.6845	2.5787	3.6871

Table 5.1b Informational matrix (e.g. T) of annual maximum flows for station-pairs.

	D1H003	D1H006	D1H009
D1H003	1	0.4453	0.8131
D1H006	0.4453	1	0.3973
D1H009	0.8131	0.3973	1

Table 5.1c Informational matrix (e.g. DIT) of annual maximum flows for different station-pairs.

	D1H003	D1H006	D1H009
D1H003	1	0.1209	0.2206
D1H006	0.1727	1	0.15055
D1H009	0.2205	0.1078	1

By assuming a threshold value of 0.20 (e.g. Threshold 1) for the directional information transfer index (above which, in a given station-pair, the first station is potentially considered as the one transferring physically information to the other station), the following station pairs were selected: D1H003 (base)-D1H009 (subject), and D1H009 (base)-D1H003 (subject). From the selected gauge pairs, it could be said that the potential predictor (base) gauge for the predicted (subject) gauge D1H003 is D1H009. Similarly, D1H003 is potential predictor (base) gauge for D1H009. Thus, the two gauging stations, i.e. D1H003 and D1H009 can mutually infer information about one another.

5. 1.2 Training and assessment of streamflow data infilling techniques

Since the streamflow data exhibited no gaps, different gap sizes (duration) of missing values were created at the target (subject) station, starting for example from 1965. The different values of gap duration of 6.7 %, 13.4 %, 20% and 30 % were considered and were then interpolated (infilled) by the selected potential predictor (base) station. For the station pairs D1H003 (subject) - D1H009 (base) and D1H009 (subject)-D1H003 (base), the only results discussed here are for gap duration of 6.7 %. This was done so, as the entropy criterion for the selection of flow data infilling techniques could be satisfied. (Refer to the next section). The results for the rest of gap duration are presented in Appendix B. Nonetheless, the rest of gap duration (e.g. 13.4 %, 20 % and 30 %) could be also considered to investigate the relationship between gap duration (size) and accuracy of the estimated annual maximum flows.

A three-layered ANN was used as in Zealand et al. (1999). The learning rate was set to the range between 0.2 and 0.4 for quite reasonable results, unless stated otherwise. The number of nodes in the hidden layer was set to the range between 2 to 4 for reasonable results. The ANN techniques were then trained on the concurrent parts of the observed values (e.g. annual maximum flows) and the parameters obtained (weights, etc.) were subsequently used to estimate the missing values. Training comprises of the annual maximum flow series pertaining to input and output to the network and obtaining the inter-connection weights for the back propagation network. Initially, the transfer function (e.g. sigmoid, hyperbolic tangent, etc.) is defined and the network is assigned values to the interconnected weights. A sigmoid function was first tried unless stated otherwise. The input data and corresponding output data were scaled within the interval (0.1, 0.9).

On the other hand the EM algorithms were applied, thus the observed data could remain intact throughout and only the missing values were repeatedly estimated until convergence.

The different techniques were first assessed among themselves and in turn compared to the other stream of techniques. Based on the results of the application of the different techniques, the proposed and existing techniques were assessed predominantly through entropy concept. The graphical and statistical criteria were also done to verify the results from entropy concept. The graphical representation here is done as in Bennis et al. (1997), where the observed parts remain intact and only the missing values are estimated. This can enhance essentially one of the steps of the EM techniques.

5.1.3 Infilling annual maximum flows at gauge D1H003 with base gauge D1H009

5.1.3.1 Using 7.6 % of missing annual maximum flows at gauge D1H003

5.1.3.1.1 Selection of ANN and EM techniques for flow infilling at D1H003

The results of performance for the different streamflow data interpolation (infilling) techniques at the subject (predicted) gauge D1H003 are summarized in Table 5.2 and Figures 5.1 (a-j). D1H009 is the base (predictor) gauge for filling in the missing annual maximum flows at D1H003. Table 5.2 summarizes entropy calculations and statistics at D1H003 using ANN and EM techniques.

The techniques that are presented in Table 5.2 are thought to have performed well (e.g. the threshold value-Threshold 2 for DIT was set to 30 %). At least 30 % of uncertainty in the annual maxim flows at D1H003 can be physically removed by the application of these techniques. The values for DIT (T) range from 0.3865 (2.686 napiers) to 0.5366 (3.729 napiers). These two limits fall under the entropy criterion. In terms of percentage, these two limits are 38.65 % and 53.66 % respectively. It could be said that the standard BP performed as well as its variants. The same can be said for the standard EM and its variants.

The statistical criteria (Table 5.2) and the graphical plot (Figures 5.1 (a-g)) for each respective technique were then made just to crosscheck result found from entropy calculations. The statistical indicators viz. RMSEp, RMEp and EV range respectively from 96.69 Mm^3 to 265.59 Mm^3 of flows, from 0.0337 to 0.0816 and from -0.0387 to -0.0987. The values for RMEp and EV, which are dimensionless, are relatively small. This can eventually confirm the results obtained from entropy criterion. The graphical plot enabled a visual comparison between the observed and the estimated hydrographs for each selected technique. The two hydrographs were found to be close for each respective technique (Figure 5.1 (a-j)). This also confirms the results from entropy calculations.

Hence, the selected techniques are thought to be good estimators of the missing annual maximum flows at D1H003 for gap duration of 6.7 %.

Table 5.2 Entropy calculations and statistics at the subject gauge D1H003 (6.7% missing annual maximum flows) using the base gauge D1H009

	Description	T (napiers)	DIT	RMSEp (Mm^3)	RMEp	EV	Ratio Variance	Ratio Mean
ANN Techniques	BP (sigmoid hidden layer)	2.686	0.3865	265.59	0.0816	-0.0987	1.047	1.009
	McL1BP (hyperbolic tangent hidden layer)	3.729	0.5366	96.69	0.0337	-0.0387	1.018	1.003
	GenerBP (sigmoid hidden layer), s=5	2.834	0.4078	230.95	0.0746	-0.089	1.041	1.008
	QBP (sigmoid hidden layer), acc = 0.015, lr = 0.85, Weight Cond	3.009	0.43302	190.131	0.0712	-0.0528	1.033	1.004
	GoldSBP (sigmoid hidden layer), lr = 0.25	2.693	0.3875	263.71	0.0805	-0.0972	1.046	1.008
EM Techniques	Standard EM	3.262	0.4694	149.080	0.0545	-0.0414	1.027	1.004
	MEM1-3	3.262	0.4694	149.080	0.0545	-0.0414	1.027	1.004
	MEM2, u = 0.89, lr = 0.001	3.264	0.4697	148.00	0.0502	-0.0407	1.025	1.0032
	ECM1-2	3.262	0.4694	149.080	0.0545	-0.0414	1.027	1.004
	ECME1-2-3	3.262	0.4694	149.080	0.0545	-0.0414	1.027	1.004

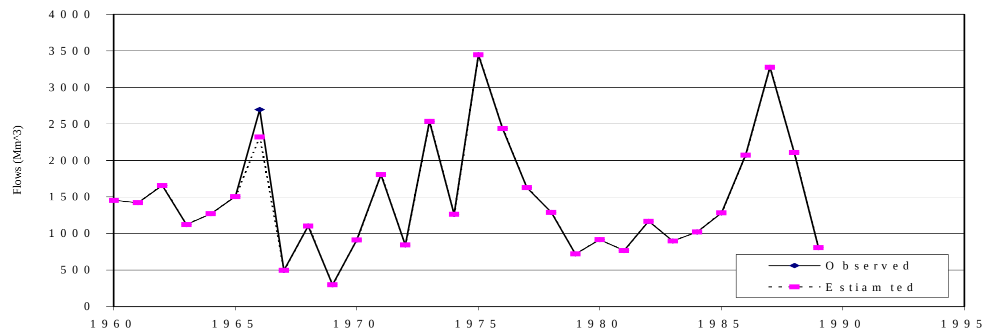


Figure 5.1a Annual maximum flows at D1H003 (6.7 % missing from 1965) using standard BP with base gauge D1H009

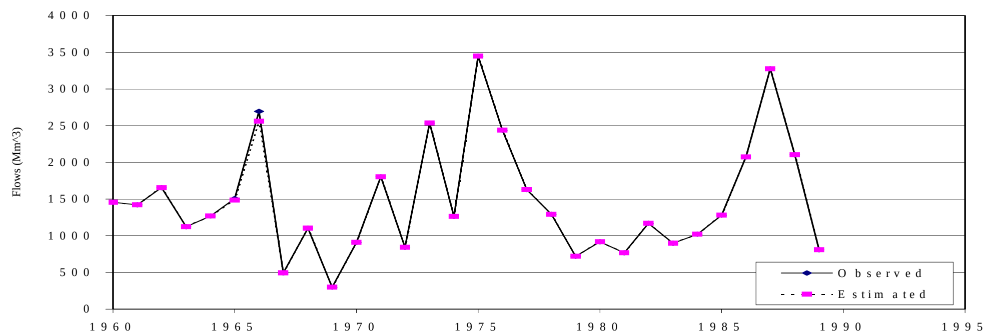


Figure 5.1b Annual maximum flows at D1H003 (6.7 % missing from 1965) using McL1BP with base gauge D1H009

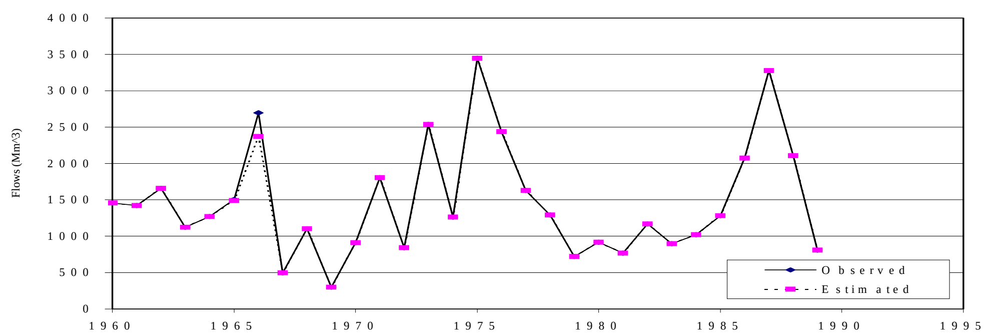


Figure 5.1c Annual maximum flows at D1H003 (6.7 % missing from 1965) using GenerBP with base gauge D1H009

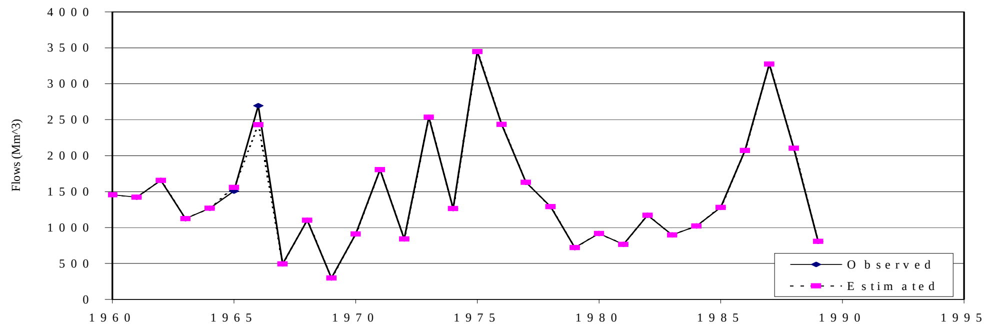


Figure 5.1d Annual maximum flows at D1H003 (6.7 % missing from 1965) using QBP with base gauge D1H009

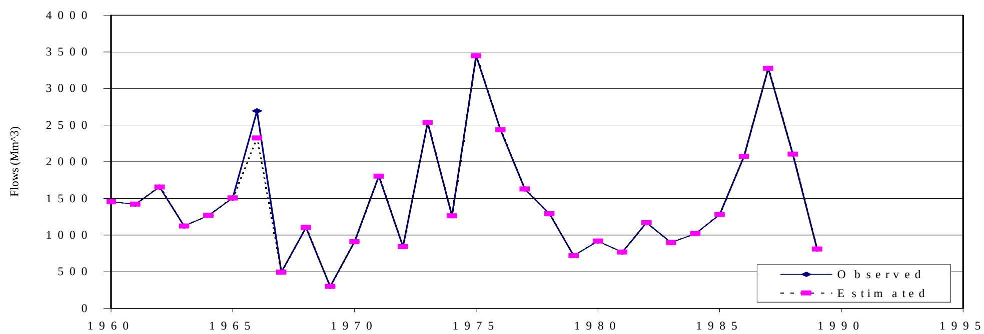


Figure 5.1.e Annual maximum flows at D1H003 (6.7 % missing from 1965) using GoldSBP with base gauge D1H009

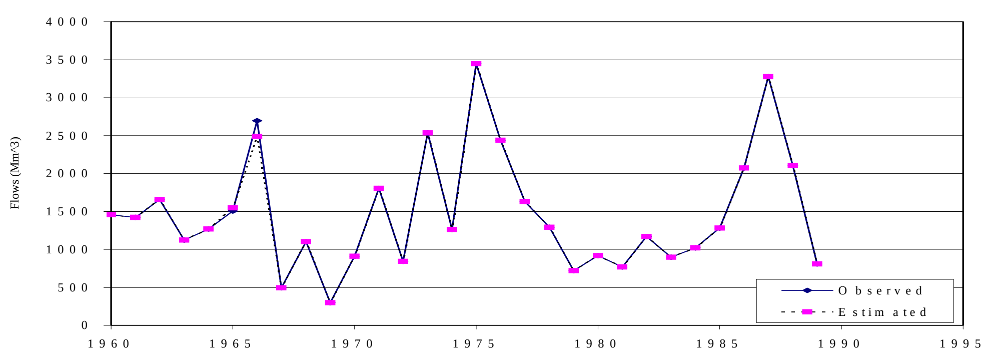


Figure 5.1f Annual maximum flows at D1H003 (6.7 % missing from 1965) using standard EM with base gauge D1H009

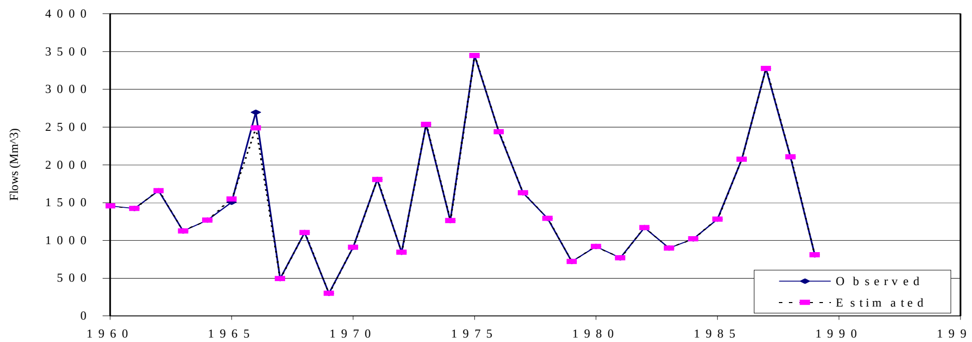


Figure 5.1g Annual maximum flows at D1H003 (6.7 % missing from 1965) using MEM1-3 with base gauge D1H009

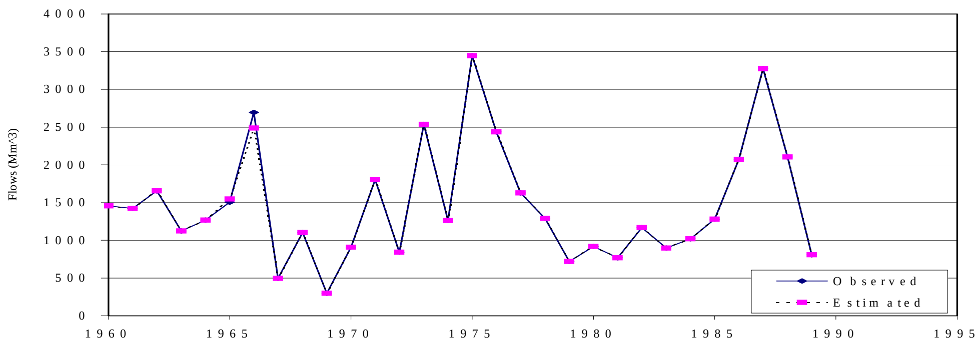


Figure 5.1.h Annual maximum flows at D1H003 (6.7 % missing from 1965) using MEM2 with base gauge D1H009

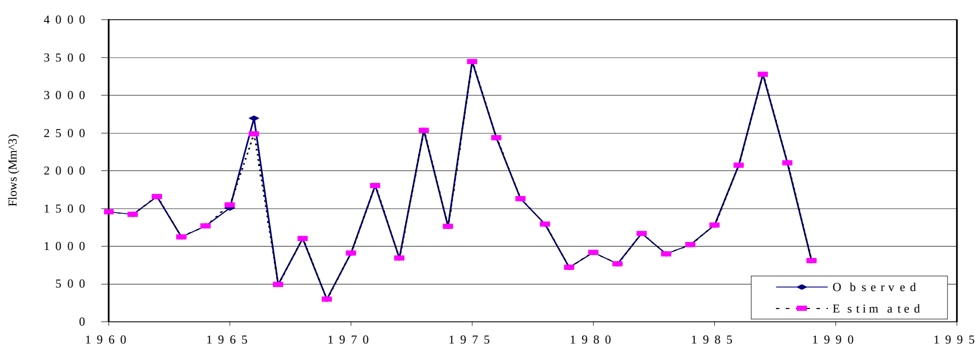


Figure 5.1i Annual maximum flows at D1H003 (6.7 % missing from 1965) using ECM1-2 with base gauge D1H009

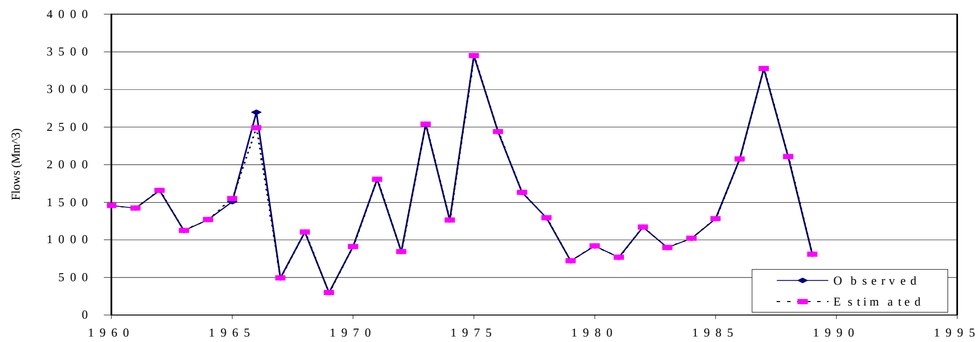


Figure 5.1j Annual maximum flows at D1H003 (6.7 % missing from 1965) using ECME1-2-3 with base gauge D1H009

5.1.3.1.2 Comparison of performance among ANN techniques at gauge D1H003 using base gauge D1H009

Based on the entropic values for the selected techniques, model performance assessment was then made for the estimation of missing annual maximum values at gauge D1H003. Considering D1H009 as base gauge and for gap duration of 6.7 % at D1H003, the performance (in a decreasing order) of the different ANN techniques in terms of DIT (T) is as follows:

-McL1 BP

-QBP, acc = 0.015, lr = 0.85 and Weight cond.

-GenerBP (lr = 0.25, s = 3)

-GoldSBP, lr = 0.25

-Standard BP

The VLR technique was not selected, as it couldn't satisfy the entropy criterion. From Table 5.2 and Figures 5.2 (a-b), the values for DIT (T) range from 0.3865 (2.686 napiers) to 0.5366 (3.729 napiers). Thus, the first best technique (i.e. McL1BP) among the selected ANN techniques has the highest entropic value of 0.5366 (3.729 napiers) for DIT (T). In other words, 53.66 % of uncertainty can be removed from the annual maximum flow series at gauge D1H003 via McL1BP. The Mac Laurin order 1 power series is believed to be a better approximation for the sigmoid function within the scaled input data interval (0.1, 0.9) compared to the standard BP for this specific station pair and for a gap duration of 6.7 % annual maximum flows at D1H003.

The QBP technique is also thought to improve slightly the results in some cases when compared to the standard technique as noticed by Patnaik (1996) and Alexander (1994) in their respective applications. The GenerBP makes it slightly better than the standard technique through its generalization parameter (e.g. s), which is chosen between 1 and 50 as proposed by Ng et al. (1996) in their applications. It was noticed that a value of this parameter greater than 5 decreases the accuracy of the estimated values. Despite that, the last best technique, i.e. standard BP enabled to remove 38.65 % of uncertainty in the annual maximum flow series at gauge D1H003. Thus, the standard BP technique is also thought to have relatively a good capability of estimating the missing annual maximum flows at D1H003.

It is believed that the standard BP leads to solutions in most cases as stated by Minns and Hall (1996). Figures 5.2 (a-b) summarize also the comparison of the performance of the different ANN techniques in terms of entropic values. These figures show that the results are very close; except for McL1BP.

On the other hand, the values for RMSEp range from 96.69 Mm^3 to 265.59 Mm^3 of flows. These two values correspond to the Mac Laurin order 1 BP and standard BP. The values of RMEp and EV ranged from 0.0337 to 0.0816 and from 0.087 to 0.0987. The two limits are related to the first best and the last best techniques respectively. The McL1BP technique has relatively low values of statistical indicators while the standard BP has relatively high values for the same statistical indicators. These results can just confirm the conclusions drawn from the entropy calculations.

Figure 5.3 did not show generally any substantial difference between hydrographs for different techniques. However, a difference could be noticed between the first best technique (i.e. McL1BP) and the last best one (i.e. BP).

In general, all these techniques presented here appear to be good estimators of the missing annual maximum flow series (6.7 %) at gauging station D1H003 (with base gauge D1H009) as long as the values for DIT comply with the entropy criterion. For 6.7 % of missing annual maximum flows, the statistical parameters for the data series at D1H003 such as mean and variance did not vary considerably as shown (in Table 5.2) by the ratio of observed statistic to the estimated statistic. In other words, the selected ANNs preserved the mean and the variance of the estimated series compared to the observed one. Nonetheless, these techniques appear to underestimate slightly the mean and variance of annual maximum flow series at D1H003. These parameters are of great importance for design-oriented studies.

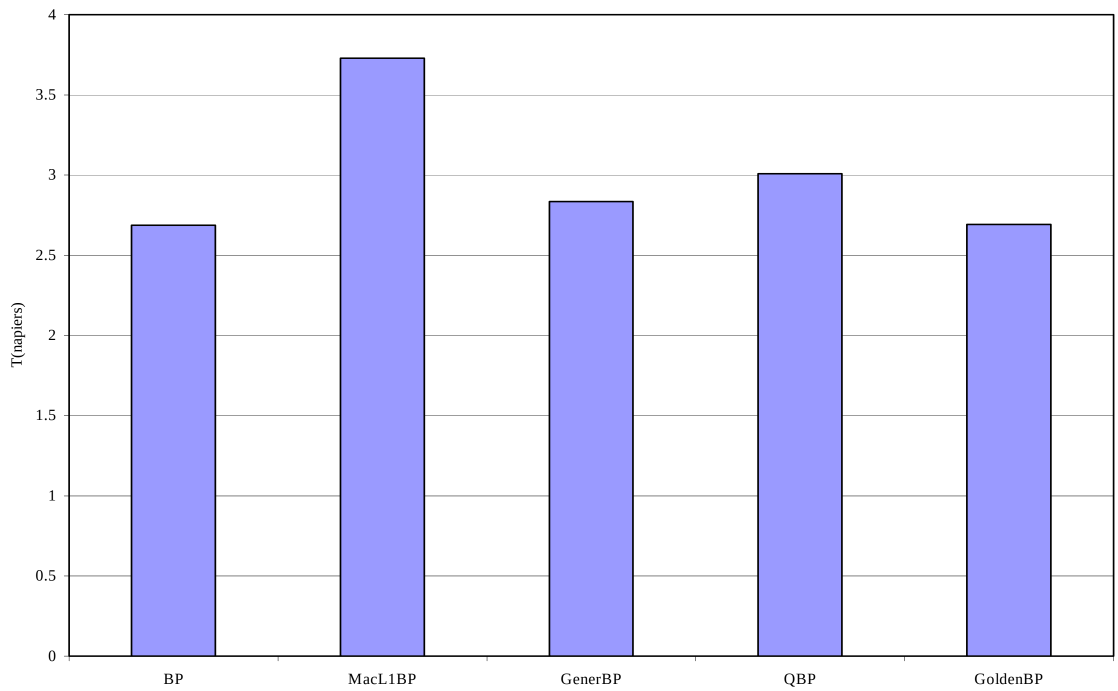


Figure 5.2a Comparison of ANNs performance in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

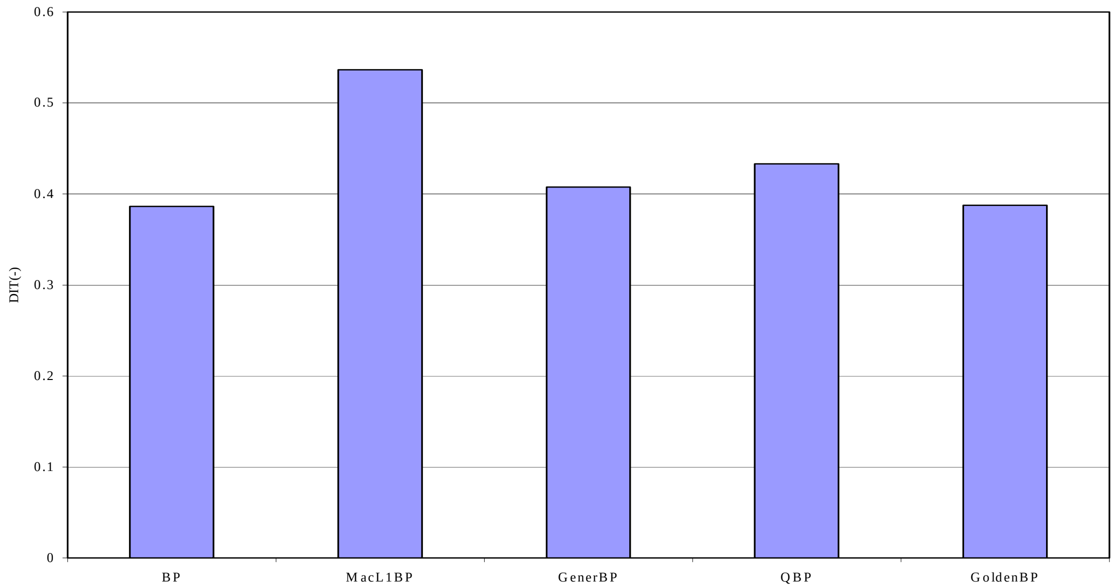


Figure 5.2b Comparison of ANNs performance in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

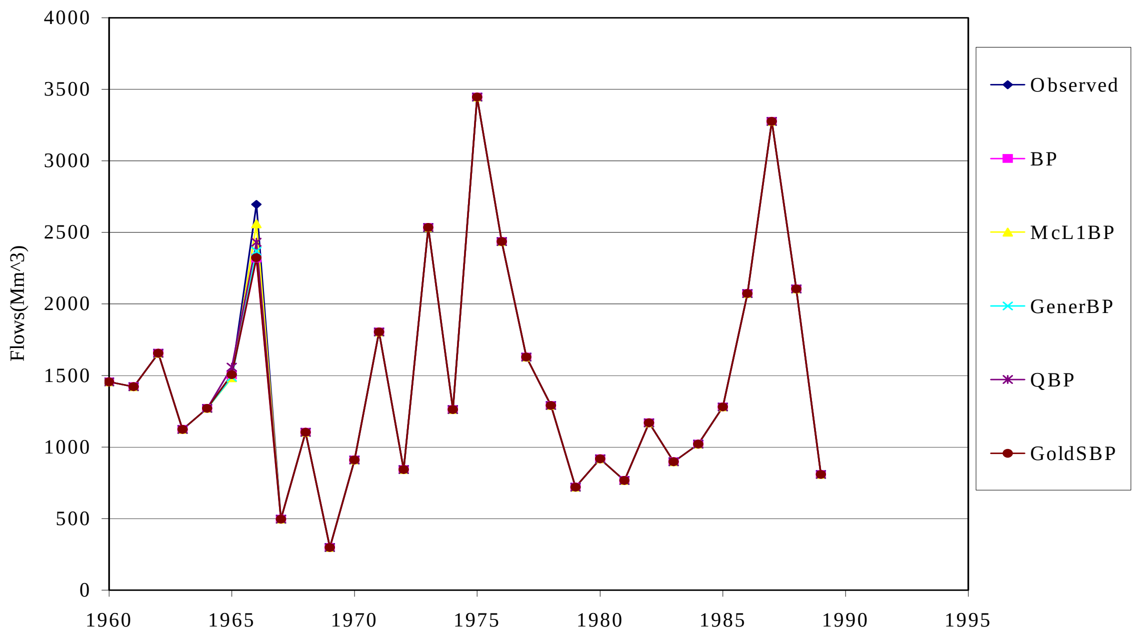


Figure 5.3 Comparison of ANNs in terms of hydrographs (6.7 % annual maximum flows from 1965 at D1H003) using base gauge D1H009

5.1.3.1.3 Comparison of performance of EM techniques at gauge D1H003 using base gauge D1H009

These selected techniques are thought to be promising estimators of the missing values at D1H003 for gap duration of 6.7 % as long as they satisfy the entropy criterion.

From the results presented in Table 5.2 and Figures 5.4 (a-b), it can be seen that the standard EM technique and its variants have almost identical entropic values although a slight (negligible) difference in entropic values was noticed for MEM2. In other words, the proportion of the uncertainty in the annual maximum flow data, which has been physically removed by the application of the standard EM and its variants, is almost the same. This shows that the standard EM leads to the same solution as its variants in some cases. It is believed in this thesis that these results could be due most probably to the form of the maximum likelihood equations (refer to equations of section 3.3.6.2 of Chapter 3). These equations are, in fact, not complicated as taken in this study. However, for complicated cases, Meng and Xu (1993) recommended using other variants.

The purpose here was to determine the accuracy of the streamflow data interpolation (infilling) techniques mainly through entropy approach, not their computational efficiency. Nonetheless, it was noticed some differences in terms of iterations between the standard method (e.g EM) and its variants. It can be concluded from entropy calculations, that both the standard EM and its variants perform likely equally in estimating the missing annual maximum flow values (e.g. 6.7 %) at gauge D1H003. The use of momentum or learning rate in the different versions, as used in this study, does not have any substantial impact on the accuracy of the estimated values. This was also noticed in the case of annual mean flow series.

The statistical indicators were almost identical for all the EM techniques. These results enhance the conclusions drawn from entropy calculations. The EM techniques preserved the mean and the variance of the data series at gauge D1H003 as shown in Table 5.2. However, these techniques seem to underestimate slightly the mean and the variance of the annual maximum flows at D1H003.

Figure 5.5 summarizes the comparison of the performance of the different EM techniques in terms of hydrographs. From a visual observation, all the estimated hydrographs were almost identical. Again, this is a confirmation of the results obtained from the entropy calculations.

In an overall assessment, the EM techniques lead almost to the same solution and are thought to be promising estimators of the missing annual maximum flows at D1H003 for gap duration of 6.7 % using the base gauge D1H009.

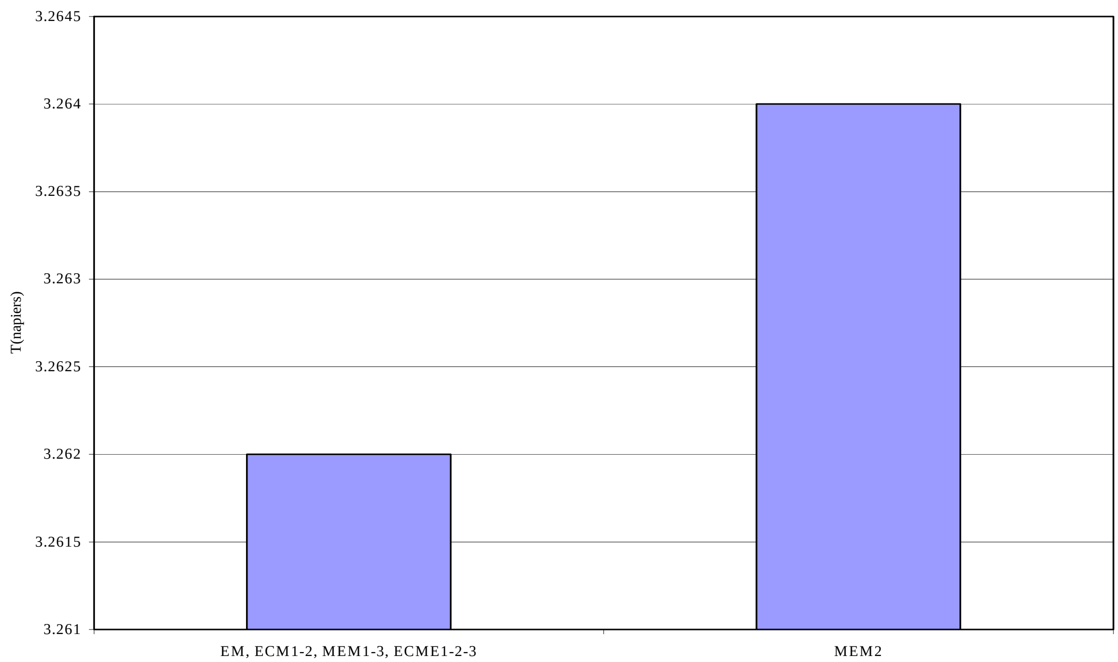


Figure 5.4a Comparison of EM techniques in terms of T (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

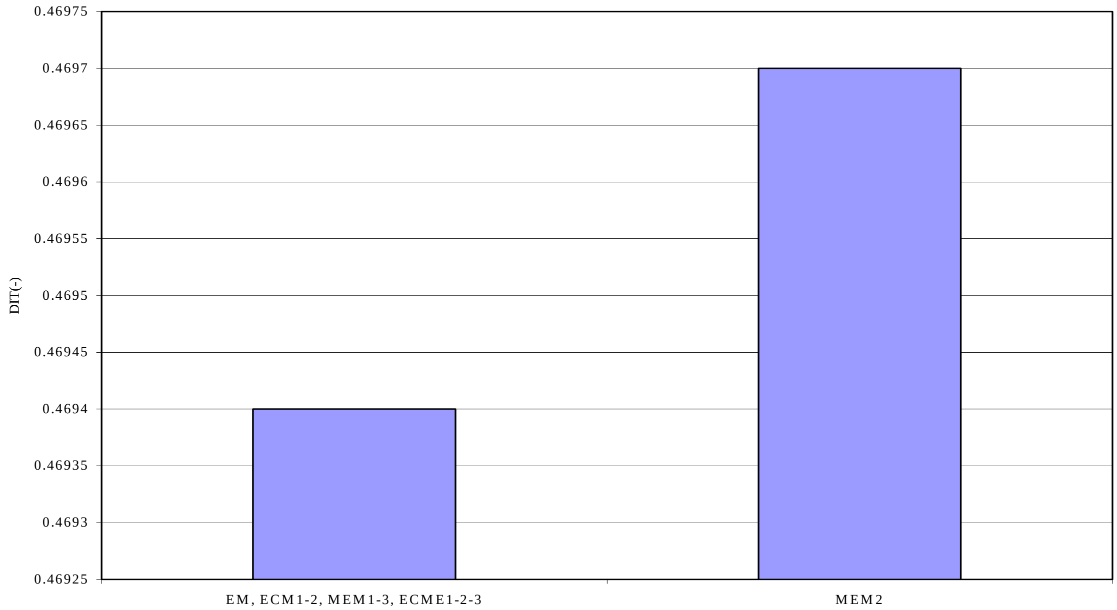


Figure 5.4b Comparison of EM techniques in terms of DIT (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

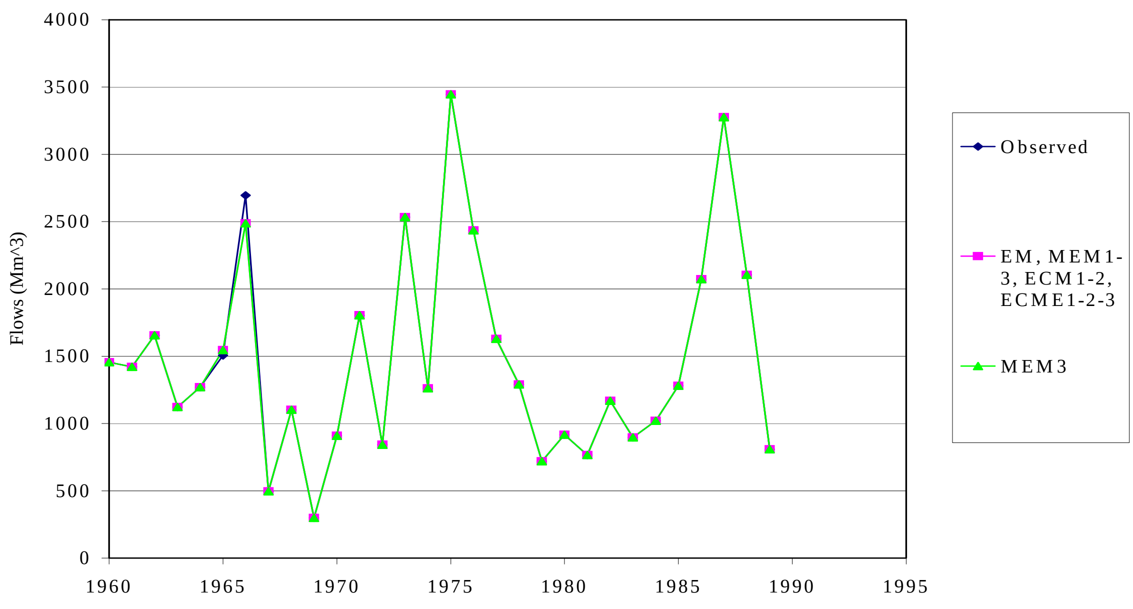


Figure 5.5 Comparison of EM in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

5.1.3.1.4 Comparison of performance of ANN and EM techniques at gauge D1H003 using base gauge D1H009

For visual convenience, only the best first two ANN (i.e. McL1BP and QBP) and EM techniques are presented in Figures 5.6 (a-b) and Figure 5.7. The EM techniques come in the second position in terms of their performance and show, however, good estimation capabilities for the missing annual maximum flows at D1H003. This fact may be due to the presence of a linear relationship in the data series between gauges D1H003 and D1H009 since the EM techniques used were developed within the linear regression context. Khalil et al. (2001) made also a similar observation, when using linear regression methods.

The best first ANN technique is thought to have relatively a higher capability (than the EM techniques) of mapping the non-linear characteristics in the missing annual maximum flow values through the use of a hyperbolic tangent hidden layer. On the other hand, the EM techniques were shown to perform slightly better than the rest of ANN techniques.

The comparison of hydrographs between the EM techniques and the best first two ANNs for gauge D1H003 is exhibited in Figure 5.7. The visual observation revealed that the hydrographs for QBP and EM techniques were very close.

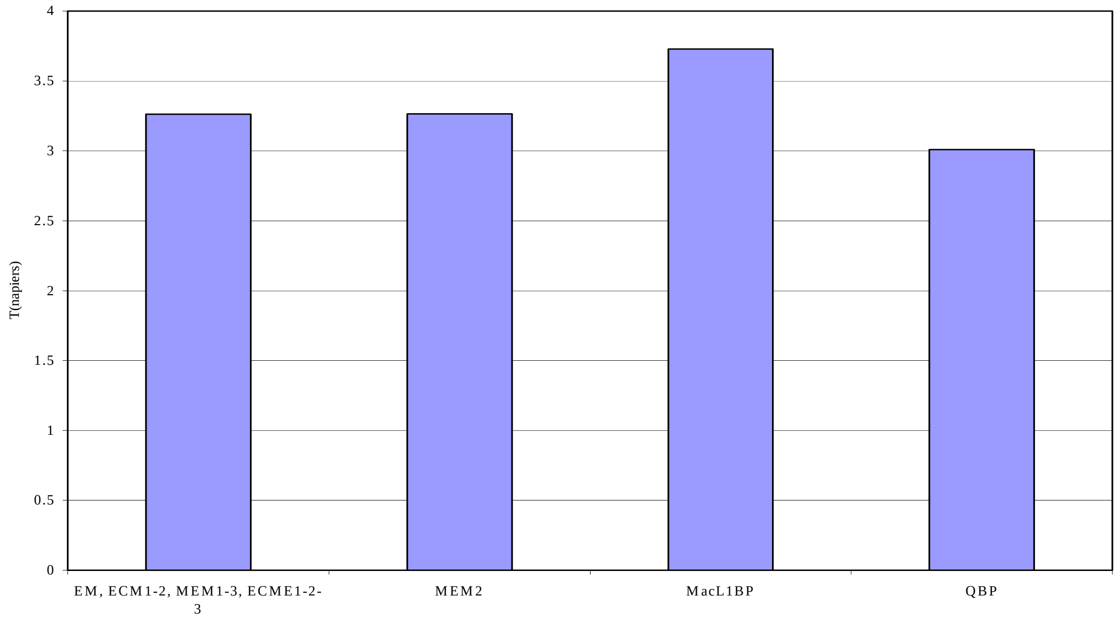


Figure 5.6a Comparison of ANN and EM techniques in terms of T (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

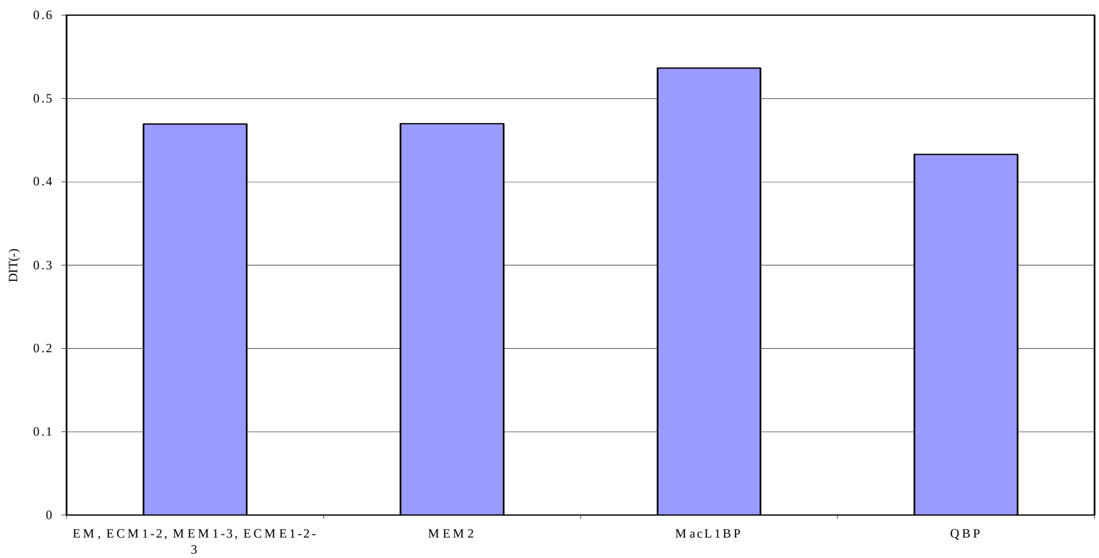


Figure 5.6b Comparison of ANN and EM techniques in terms of DIT (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

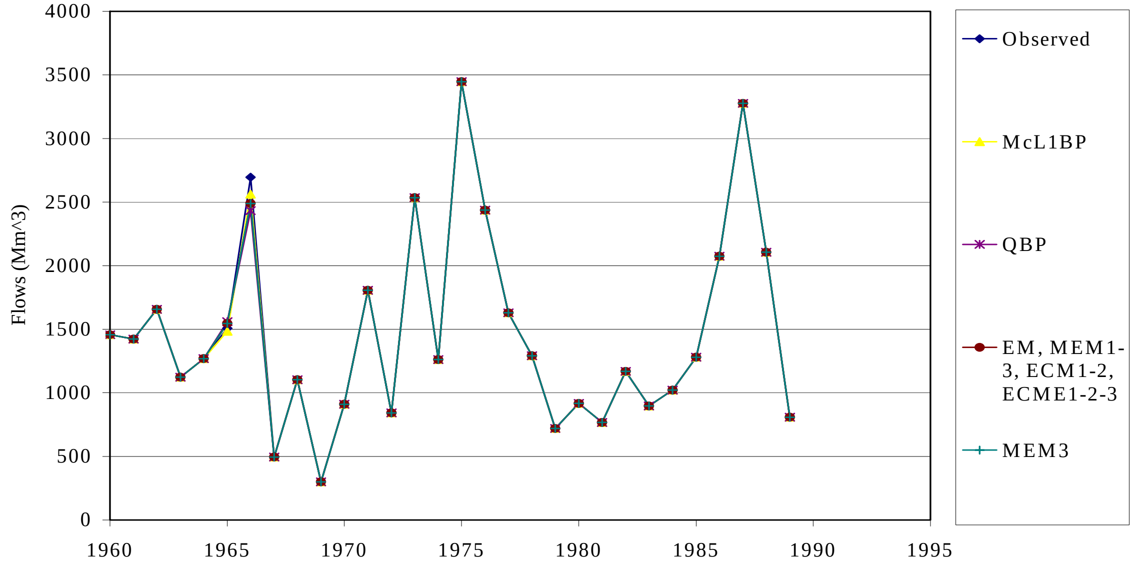


Figure 5.7 Comparison of ANN and EM techniques in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H003) using base gauge D1H009

5.1.4 Infilling annual maximum flows at gauge D1H009 with base gauge D1H003

5.1.4.1 Using 7.6 % of missing annual maximum flows at gauge D1H009

5.1.4.1.1 Selection of ANN and EM techniques for flows infilling at D1H009

The results of performance for the different streamflow data infilling techniques (e.g. ANN and EM) at the subject (predicted) gauge D1H009 are summarized in Table 5.3 and Figures 5.8 (a-l). D1H003 is the base (predictor) station for interpolating (filling in) the missing annual maximum flows at D1H009. Table 5.3 summarizes entropy calculations and statistics at D1H009 using ANN and EM techniques.

The values of DIT (T) for the selected techniques range from 0.4904 (3.408 napiers) to 0.7060 (4.906 napiers). These two limits fall under the entropy criterion, e.g. at least 30 % of uncertainty in the annual maximum flow data at D1H003 can be removed by the application of these techniques. It could be said that the standard BP performed as well as its variants. The same is noticed for the EM technique and its variants.

The statistical criteria (Table 5.3) and the graphical plot (Figures 5.8 (a-l)) for each respective technique were just made to crosscheck the results found from entropy calculations. The statistical indicators viz. RMSEp RMEp and EV range respectively from 20.043 Mm^3 to 116.83 Mm^3 of flows, from 0.0114 to 0.0612 and from 0.00842 to –0.065. The values for RMEp and EV, which are dimensionless, are relatively small. The graphical plot enabled a visual comparison between the observed and the estimated hydrographs for each selected technique. The two hydrographs were found to be close for each respective technique (Figures 5.8 (a-l)). It was noticed that the match between the two hydrographs was very strong for both McL1BP and GenerBP.

The results from the graphical comparison and the statistical indicators confirm just the results from entropy calculations. Hence, the selected techniques are thought to be good estimators of the missing annual maximum values at D1H009 for gap duration of 6.7 %.

Table 5.3 Entropy calculations and statistics at the target gauge D1H009 (6.7% missing annual maximum flows) using the base gauge D1H003

	Description	T (napiers)	DIT	RMSEp	RMEp	EV	Ratio Variance	Ratio Mean
ANN Techniques	BP (sigmoid hidden layer)	3.971	0.5714	51.74	0.0294	-0.0227	0.997	0.0998
	McL1BP (hyperbolic tangent hidden layer)	4.663	0.6710	26.497	0.0176	-0.0383	0.997	1.000
	GenerBP (sigmoid hidden layer), s = 5	4.906	0.7060	20.043	0.0114	0.00842	0.9994	0.9992
	QBP (sigmoid hidden layer), acc = 0.015, lr = 0.85, Weight Cond	3.414	0.4913	116.317	0.0609	-0.0647	1.025	1.006
	GoldSBP (sigmoid hidden layer), lr = 0.25	4.073	0.5861	46.34	0.0276	-0.0152	1.001	0.998
	VLR (sigmoid hidden layer)	3.408	0.4904	116.83	0.0612	-0.065	1.025	1.006
EM Techniques	Standard EM	3.448	0.4962	106.72	0.0642	-0.0634	1.022	1.006
	MEM1	3.448	0.4962	106.72	0.0642	-0.0634	1.022	1.006
	MEM2, u = 0.89, lr = 0.001	3.446	0.4959	106.750	0.06510	0.0640	1.020	0.9900
	MEM3	3.449	0.4963	106.00	0.0641	0.0635	1.019	0.9930
	ECM1-2	3.448	0.4913	106.72	0.0642	-0.0634	1.022	1.006
	ECME1-2-3	3.448	0.4913	106.72	0.0642	-0.0634	1.022	1.006

Figure 5.8a Annual maximum flows at D1H009 (6.7 % missing from 1965) using standard BP with base gauge D1H003

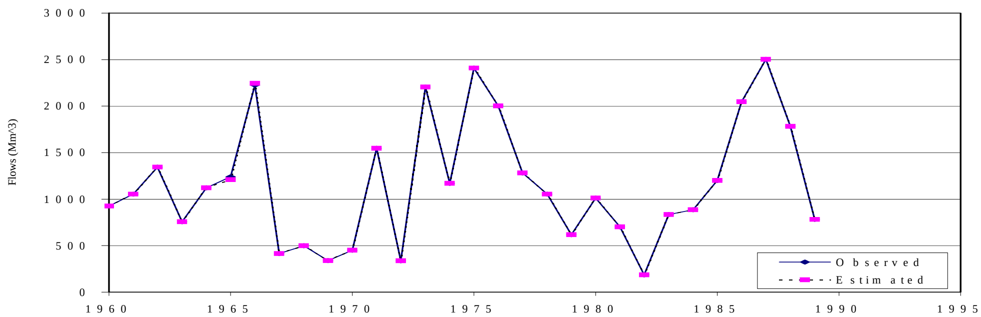


Figure 5.8b Annual maximum flows at D1H009 (6.7 % missing from 1965) using McL1BP (hyperbolic tangent hidden layer) with base gauge D1H003

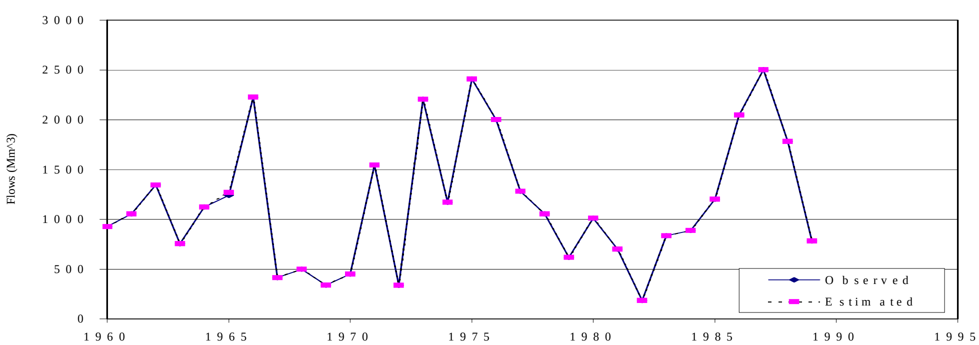


Figure 5.8c Annual maximum flows at D1H009 (6.7 % missing from 1965) using GenerBP with base gauge D1H003

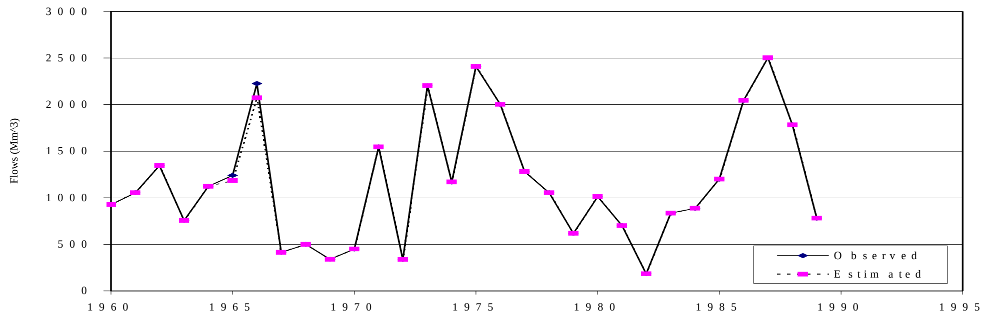


Figure 5.8d Annual maximum flows at D1H009 (6.7 % missing from 1965) using QBP with base gauge D1H003

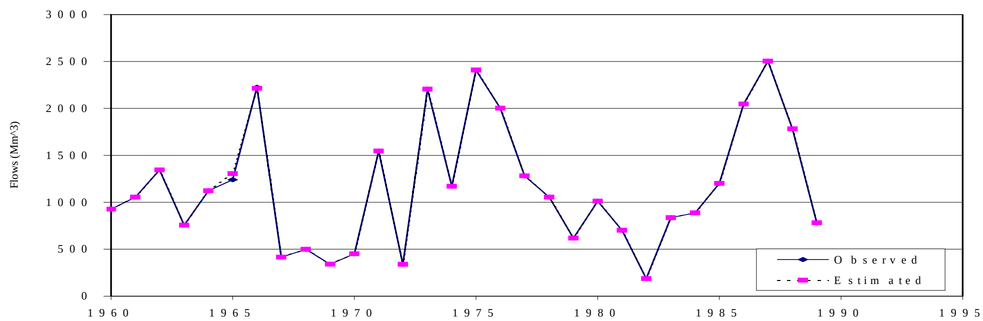


Figure 5.8e Annual maximum flows at D1H009 (6.7 % missing from 1965) using GoldSBP with base gauge D1H003

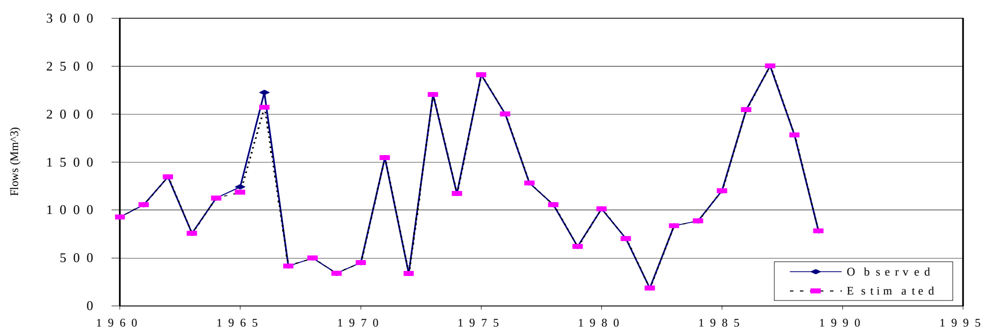


Figure 5.8f Annual maximum flows at D1H009 (6.7 % missing from 1965) using VLRBP with base gauge D1H003

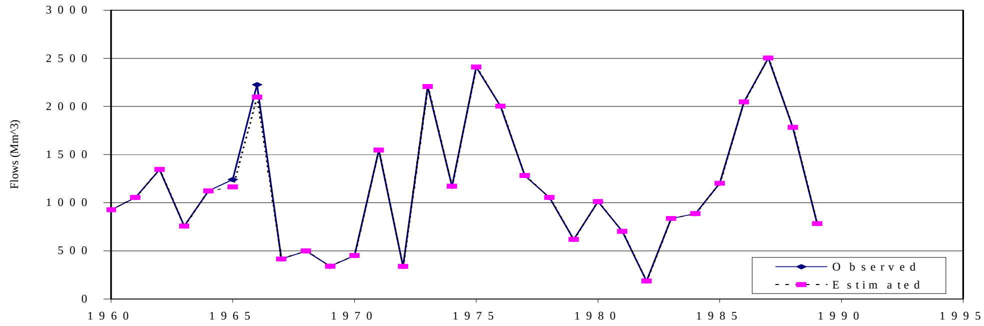


Figure 5.8g Annual maximum flows at D1H009 (6.7 % missing from 1965) using standard EM with base gauge D1H003

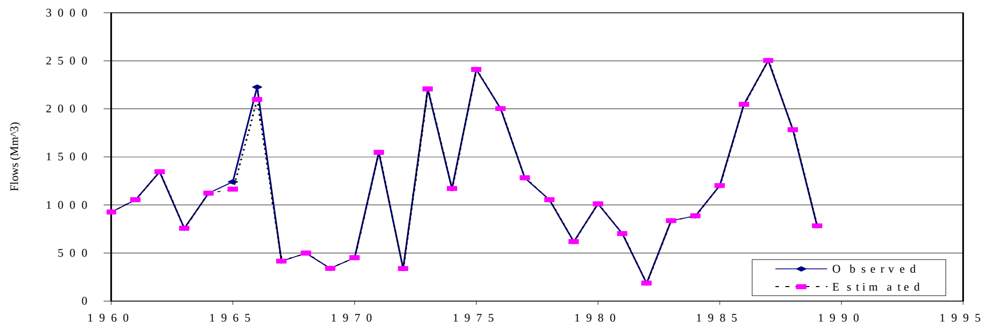


Figure 5.8h Annual maximum flows at D1H009 (6.7 % missing from 1965) using MEM1 with base gauge D1H003

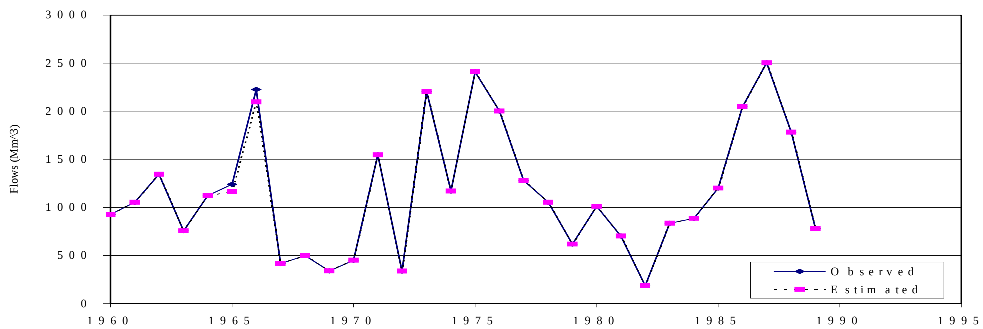


Figure 5.8i Annual maximum flows at D1H009 (6.7 % missing from 1965) using MEM2 with base gauge D1H003

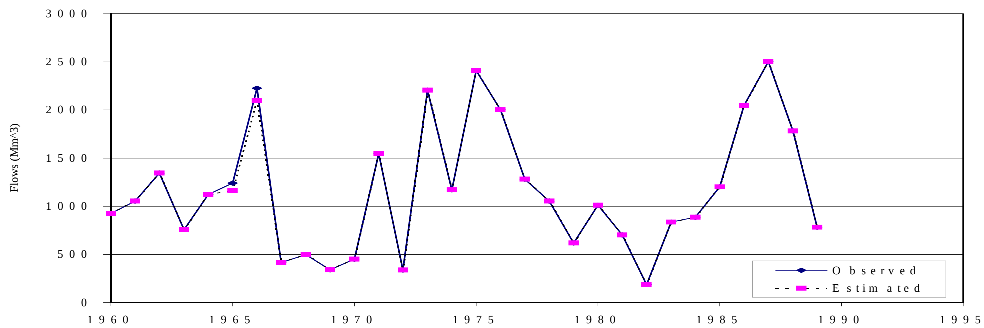


Figure 5.8j Annual maximum flows at D1H009 (6.7 % missing from 1965) using MEM3 with base gauge D1H003

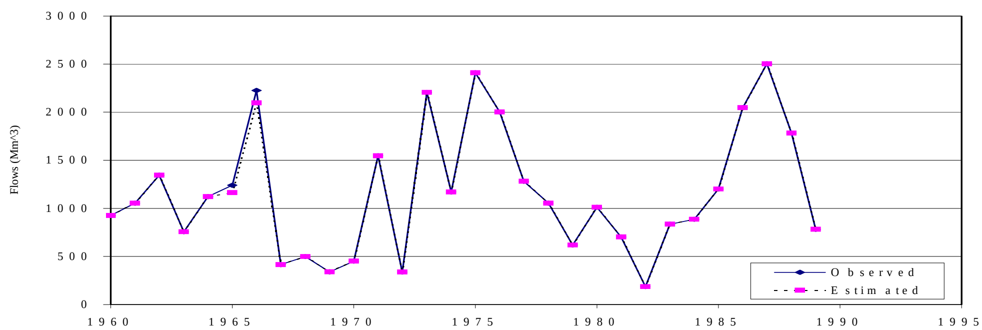


Figure 5.8k Annual maximum flows at D1H009 (6.7 % missing from 1965) using ECM1-2 with base gauge D1H003

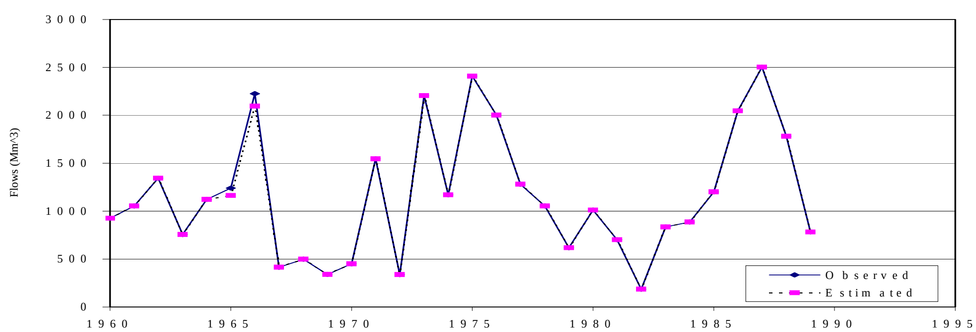


Figure 5.8l Annual maximum flows at D1H009 (6.7 % missing from 1965) using ECME1-2-3 with base gauge D1H003

5.1.4.1.2 Comparison of performance among ANN techniques at gauge D1H009 using base gauge D1H003

Based on the entropic values for the selected techniques, model performance assessment was then made for the estimation of missing annual maximum values at gauge D1H009. Taking gauge D1H003 as base station and for gap duration of 6.7 % at D1H009, the performance of the different ANN techniques in terms of DIT (T) is as follows:

-GenerBP (lr = 0.25, s = 2)

-McL1BP

-GoldSBP, lr = 0.25

-Standard BP

-QBP, acc = 0.015, lr = 0.85 and Weight cond.

-VLR

From Table 5.3 and Figures 5.9 (a-b) and Figure 5.10, the values of DIT (T) range from 0.4904 (3.408 napiers) to 0.7060 (4.906 napiers). Thus, the first best technique (i.e. GenerBP) among the selected ANN techniques has the highest entropic value of 0.7060 (4.906 napiers) for DIT (T). In other words, 70.60 % of uncertainty can be removed from the annual maximum flow data at gauge D1H009, via the GenerBP technique. The last best technique, i.e. VLR enabled to remove 49.04 % of uncertainty at gauge D1H009. The standard BP technique is also thought to have relatively a good capability of estimating the missing annual maximum flows at D1H009. This technique enabled 57.14 % of uncertainty in the to be removed from the annual maximum flows at gauge D1H009.

On the other hand, the values for RMSEp range from 20.043 Mm^3 to 116.83 Mm^3 of flows. These two values correspond to the GenerBP and the VLR respectively. The values of RMEp and EV ranged from 0.0114 to 0.0612 and from 0.00842 to –0.065. The two limits are related to the first best and the last best techniques respectively.

The GenerBP technique has relatively low values of statistical indicators while the standard BP has relatively higher values for the same statistical indicators. The GenerBP, McL1BP and GoldSBP are thought to be slightly better than the standard BP. The explanation of this situation can be found in section 5.1.3.1.2. These results just confirm the conclusions drawn from the entropy calculations.

Figure 5.10 shows that the different hydrographs are very close. This plot did not show any substantial difference between any two successive techniques as ranked above. However, a small difference could be noticed between the first best technique and the last best technique. For 6.7 % missing annual maximum flows, the statistical parameters (for the data series at D1H009) such as mean and variance did not vary considerably as shown (in Table 5.3) by the ratio of the observed statistic to the estimated statistic. In other words, the above-selected ANNs preserved the mean and the variance of the estimated series compared to the observed one. Nonetheless, these techniques appear to underestimate slightly the mean and the variance of annual maximum flows at D1H009. Recall that these parameters are very important for design-oriented studies.

In general, all these techniques presented here, appear to be good estimators of the missing annual maximum flow series (6.7 %) at the target gauge D1H009 using the base gauge D1H003.

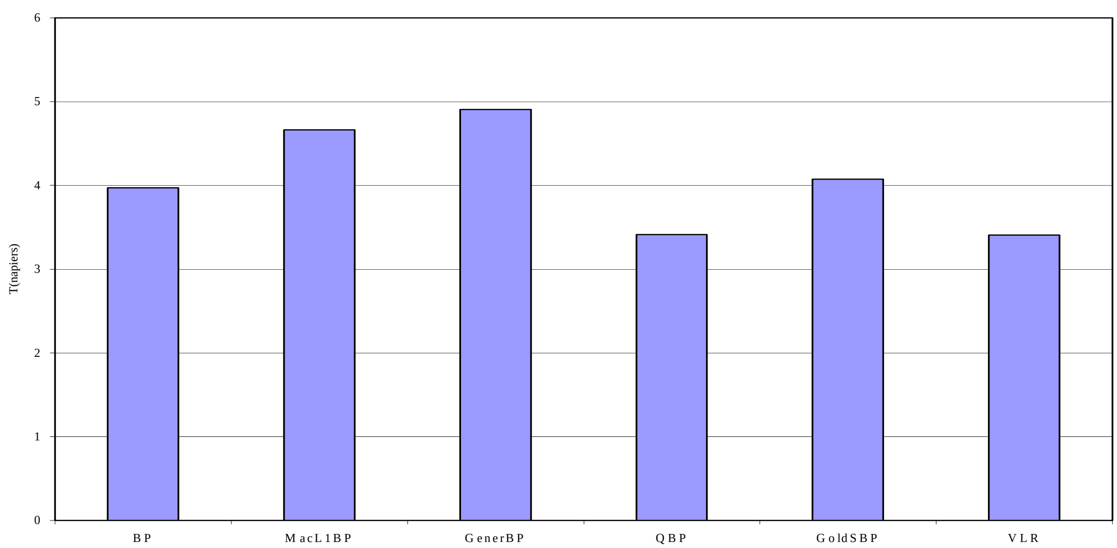


Figure 5.9a Comparison of ANNs in terms of T (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

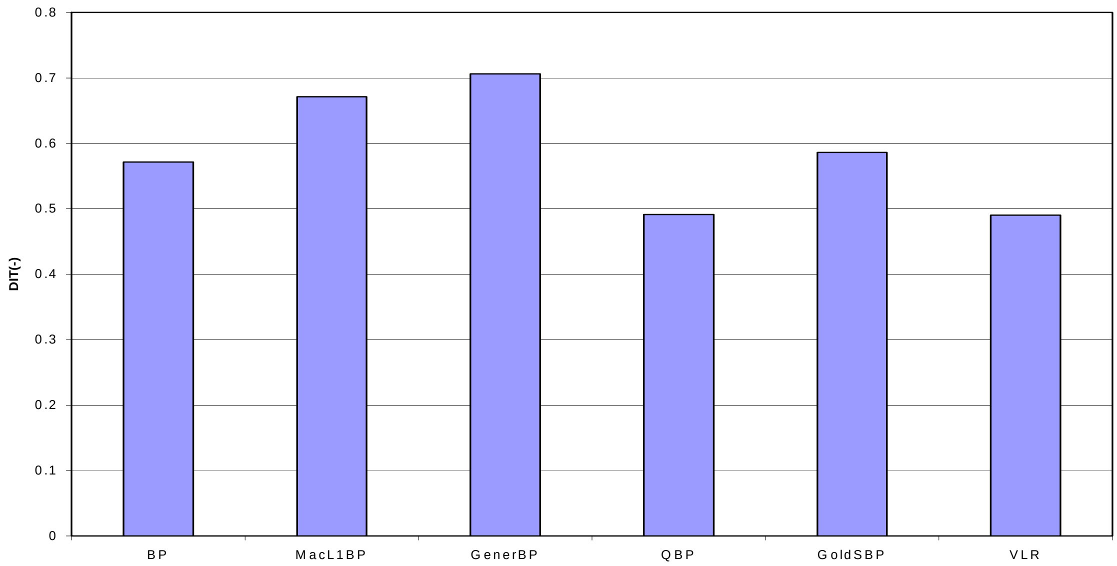


Figure 5.9b Comparison of ANNs in terms of DIT (6.7 % annual maximum flows from 1965 at D1H009) using base gauge D1H003

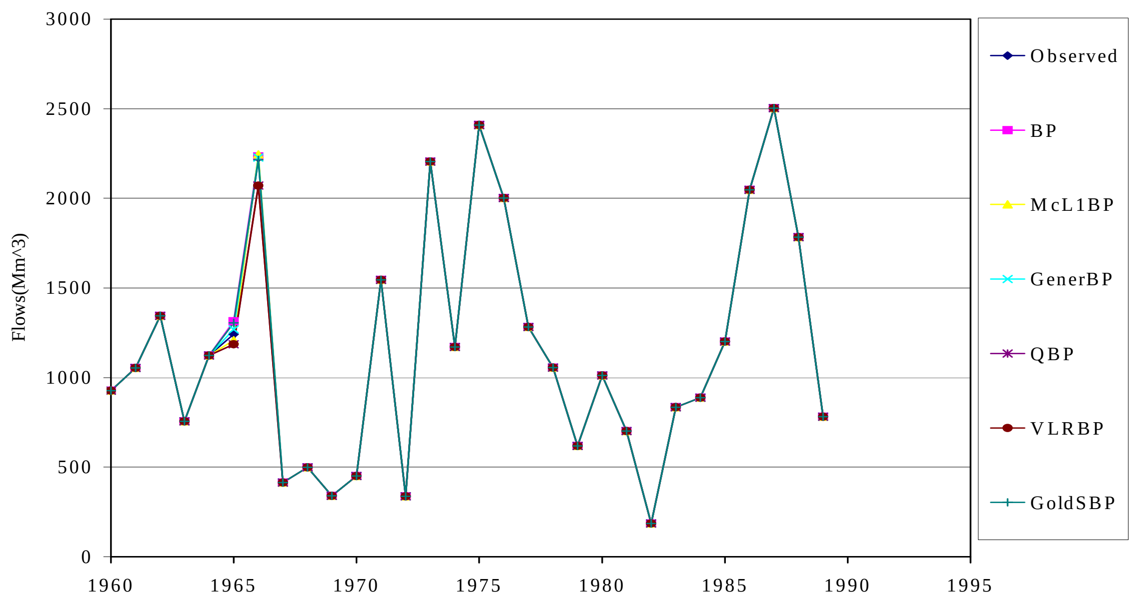


Figure 5.10 Comparison of ANN techniques in terms of hydrographs (6.7 % annual maximum flows from 1965 at D1H009) using base gauge D1H003

5.1.4.1.3 Comparison of performance of EM techniques at gauge D1H009 using base gauge D1H003

The EM techniques are thought to be promising estimators of the missing values at D1H009 for gap duration of 6.7 % since they satisfy the entropy criterion. From the results presented in Table 5.3 and Figure 5.11 (a-b), it can be seen that the standard EM technique and its variants have almost identical entropic values although a slight (negligible) difference in entropic values was noticed for MEM3. In other words, the proportion of the uncertainty in the annual maximum flow data, which has been physically removed by the application of the standard EM and its variants, is almost the same. The standard EM leads to the same solution as its variants in some cases. It is believed that these results could be due most probably to the form of the maximum likelihood function, which is in fact not complicated as taken in this research work. However, for complicated cases Meng and Rubin (1993) recommended using other variants.

The purpose here was to estimate the accuracy of the annual maximum flow infilling techniques pre-dominantly through entropy approach, not their computational burden. Nonetheless, it was noticed some differences in terms of iterations between the standard method and its variants. Thus, from entropy calculations both the standard EM and its variants perform likely equally in estimating the missing annual maximum flow values (e.g. 6.7 %) at the target gauge D1H009, using the base gauge D1H003. The use of momentum or learning rate in the different versions, as used in this study, does not have any substantial impact on the accuracy of the estimated values.

The statistical indicators were also identical for all the EM techniques. These results enhance the conclusions drawn from entropy calculations. The EM techniques preserved the mean and variance as shown in Table 5.3. However, these techniques seem to underestimate slightly the mean and the variance of the annual maximum flows at D1H009.

Figure 5.12 summarizes the comparison of the performance in terms of hydrographs for the different EM techniques. From a visual observation, all the estimated hydrographs were almost identical. This is a confirmation of the results obtained from the entropy calculations.

In an overall assessment, the EM techniques lead almost to the same solution and are thought to be promising estimators of the missing annual maximum flows values at D1H009 for gap duration of 6.7 %, using the base gauge D1H003.

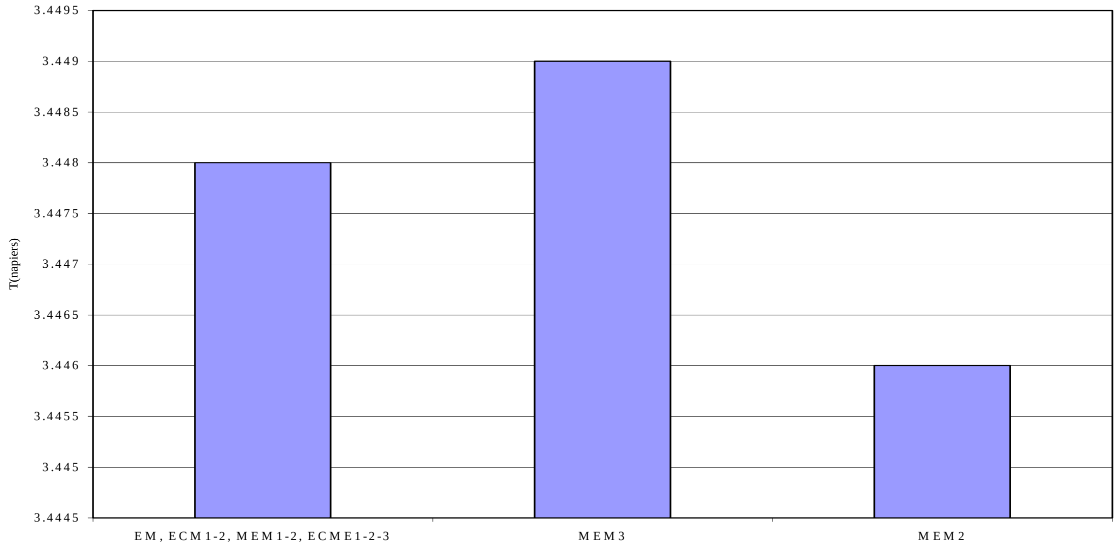


Figure 5.11a Comparison of EM techniques in terms of T (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

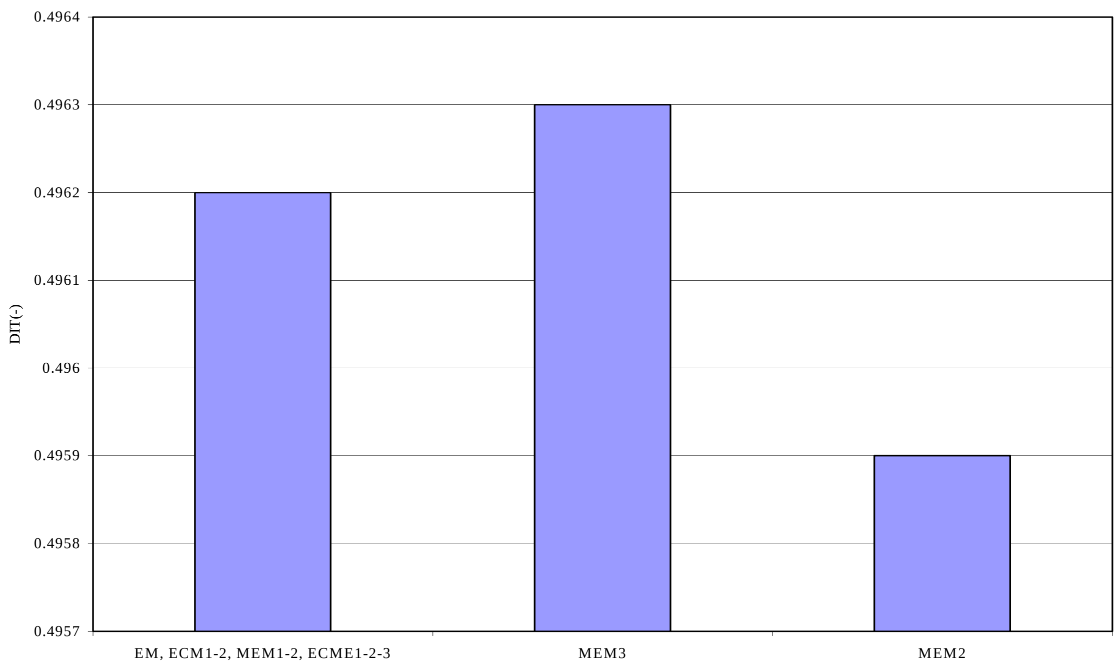


Figure 5.11b Comparison of EM techniques in terms of DIT (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

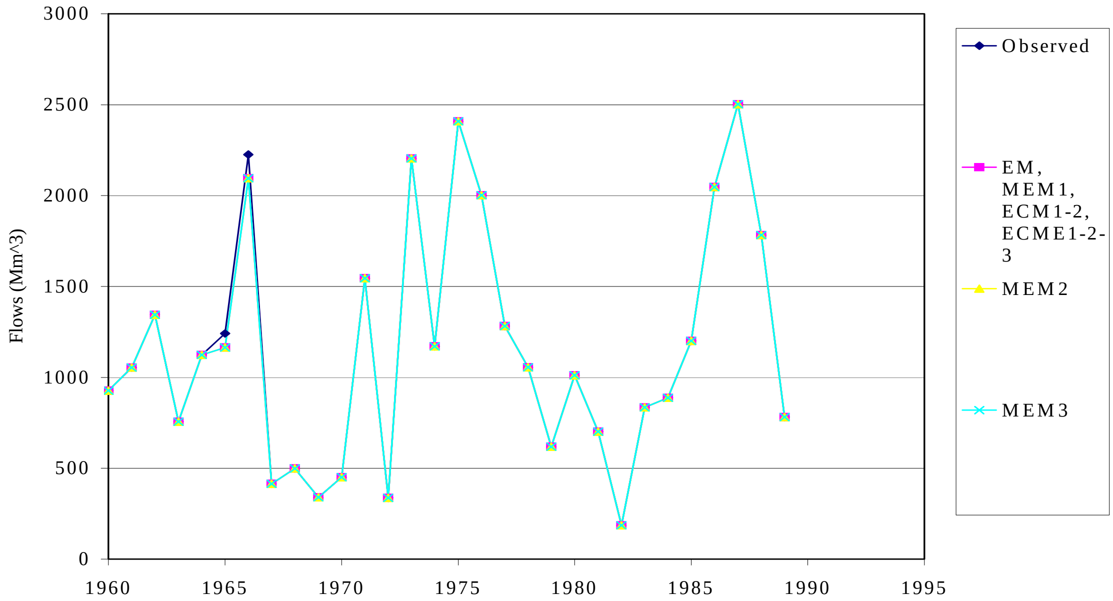


Figure 5.12 Comparison of EM techniques in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

5.1.4.1.4 Comparison of performance of ANN and EM techniques at gauge D1H009 using base gauge D1H003

The results for the comparison between ANNs and EM techniques are depicted in Figures 5.13 (a-b) and Figure 5.14. The EM techniques come in the fourth position in terms of performance (through entropy criterion) and show, however, good estimation capabilities for the missing annual maximum flows at D1H009. This fact may be due to the presence of linear relationship in the flow series between gauges D1H003 and D1H009, since the EM techniques used here were developed within the linear regression context. Khalil et al. (2001) made also a similar observation, when using linear regression methods.

The first best four ANN techniques are thought to have relatively a higher capability (than EM techniques) of mapping the non-linear characteristics in the missing values through the use of a sigmoid/hyperbolic hidden layer. On the other hand, the EM techniques were shown to perform slightly better than the rest of ANNs.

The visual observation revealed that the hydrographs were somehow close (Figure 5.14).

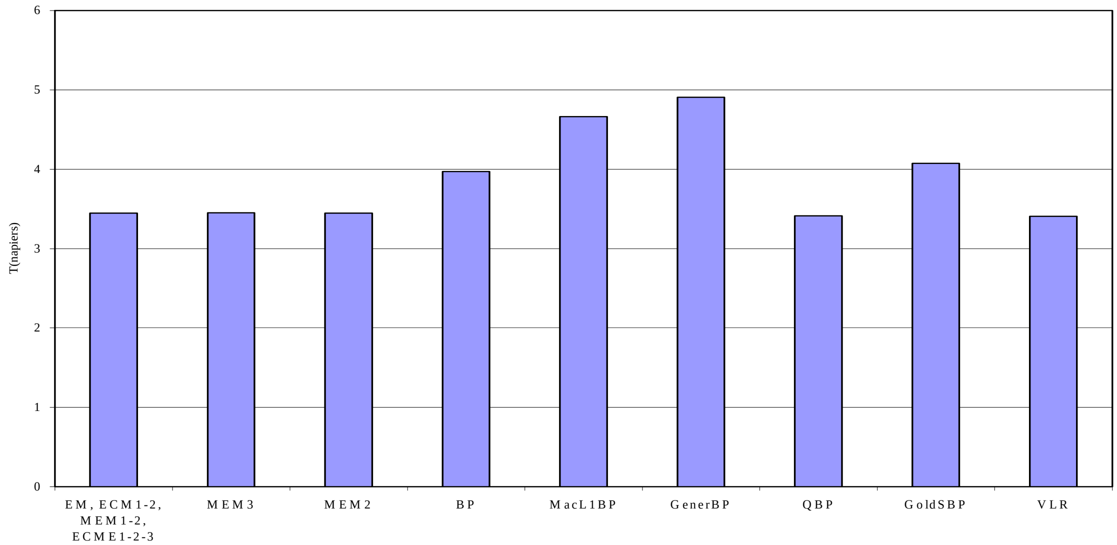


Figure 5.13.a Comparison of ANN and EM techniques in terms of T (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

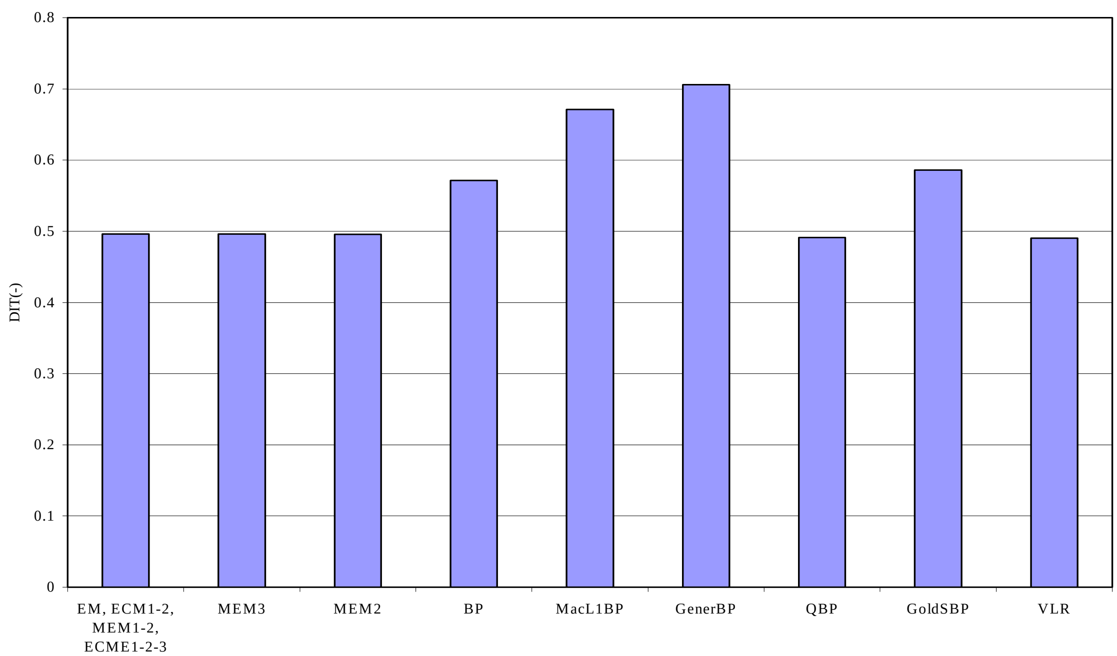


Figure 5.13b Comparison of ANN and EM techniques in terms of DIT (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

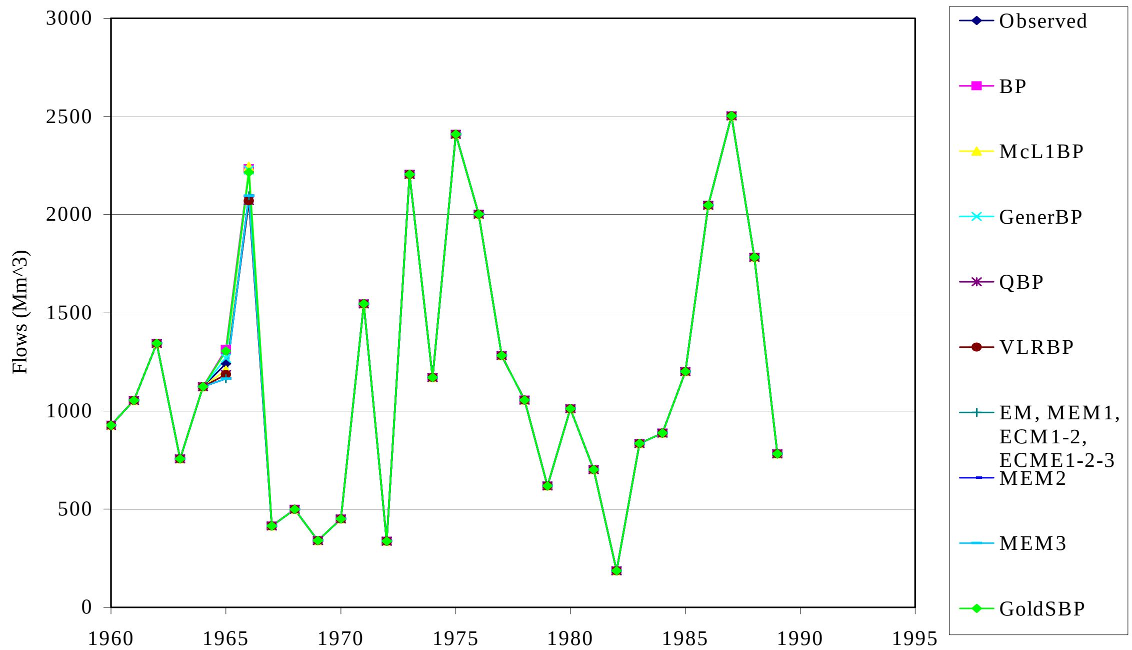


Figure 5.14 Comparison of ANN and EM techniques in terms of hydrographs (6.7 % missing annual maximum flows from 1965 at D1H009) using base gauge D1H003

The values of DIT could enable the comparison of the performance of the different techniques on different catchment areas, e.g. gauges D1H003 and D1H009. In general, for a given gap size, the overall value of DIT for the station pair D1H003-D1H009 was higher than the one for station pair D1H009-D1H003. Thus, D1H003 was revealed to be a better predictor for the interpolation of missing annual maximum flows at D1H009 than when gauge D1H009 is used for interpolating the data series at D1H003.

5.2 SUMMARY

The entropy concept was shown to be a versatile tool (through DIT) in the sense that it firstly enables to know the information content of the streamflow gauges. Secondly, it enables also the selection of predicted/predictor gauge. Laslty, it is also used to assess a technique performance when applied to annual maximum series infilling. Recall that these data are used in design-oriented studies of hydraulic structures such as reservoirs, bridges, flood studies etc.

It was shown that the gauges in the station-pairs D1H003-D1H009 and D1H009-D1H003 could infer mutually information (contained in annual maximum flow series) about one another. In other words, when one gauge could be considered as predicted station, the other one was predictor station and vice-versa.

Only the results for gap duration of 6.7 % in annual maximum flows at either station were discussed as the entropy criterion was satisfied with regard to the performance of the different techniques. The results of the rest of gap duration were not satisfactory and therefore, were not discussed. Nonetheless, the results of the rest of the gap duration were just used to investigate the relationship between gap duration and accuracy of estimated values for the different techniques.

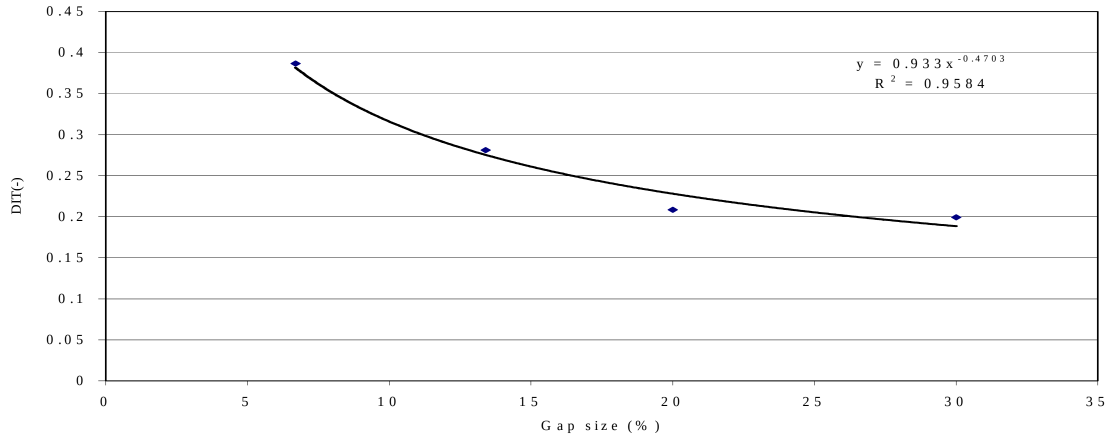
Considering the values for DIT, the different techniques could also be compared for different catchment areas. In general, It was shown that D1H003 was a better predictor for the estimation process of missing annual maximum flows at D1H009 (than when D1H009 was used for estimating the missing values at D1H003).

It was also noticed that the directional information transfer index (DIT) between observed and estimated values at either target gauge generally decreases when the proportion of missing annual maximum flows increases. In other words, the proportion of information physically transferred by the knowledge of the estimated series into the process to make the annual maximum flow series better defined will decrease as the gap duration increases at the subject station.

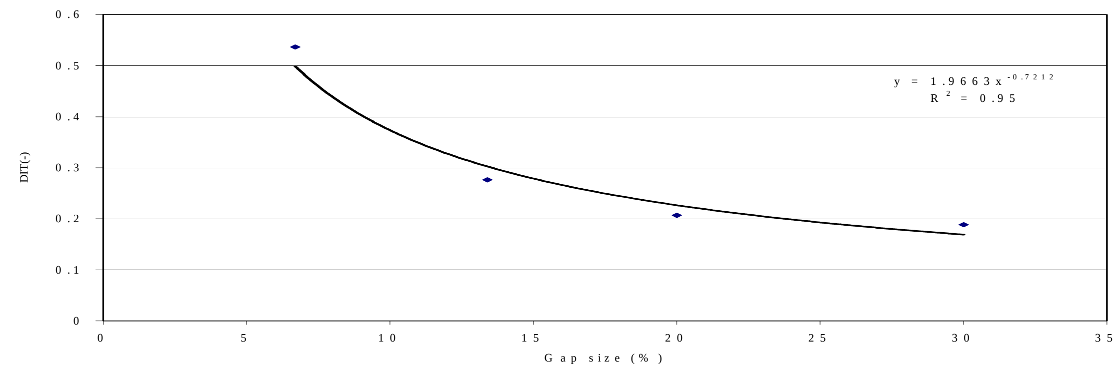
Generally, a decay power function could describe better (than exponential/linear) the relationship between the gap duration (e.g. from 6.7 % to 30%) and the technique accuracy in terms of DIT. However, only the results for gap duration of 6.7 % of annual maximum flows at either target gauge could comply with the entropy criterion. Figures 5.15 (a-b) and Figure 5.16 (a-b) give just an illustration. Thus, for a given technique, it is possible to find approximately the expected accuracy of the estimated annual maximum flow values at the subject station when the gap size (duration) is known. Increasing the number of data points (e.g. up to seven values of gap size at the target gauge) did not sensitively affect the above-mentioned relationship. It was also noticed that an earlier start (e.g. at 1963) or later start (e.g. 1970) for the gaps created on the records of the subject station did not show any substantial impact on the accuracy of the estimated values.

An overall assessment of the techniques used here revealed that the standard BP and the standard EM as well as their respective variants showed good capabilities of estimating the missing annual maximum flows at the target gauge D1H009 using the base gauge D1H003 (and vice-versa). Nonetheless, the results from the EM techniques were almost identical as explained so far.

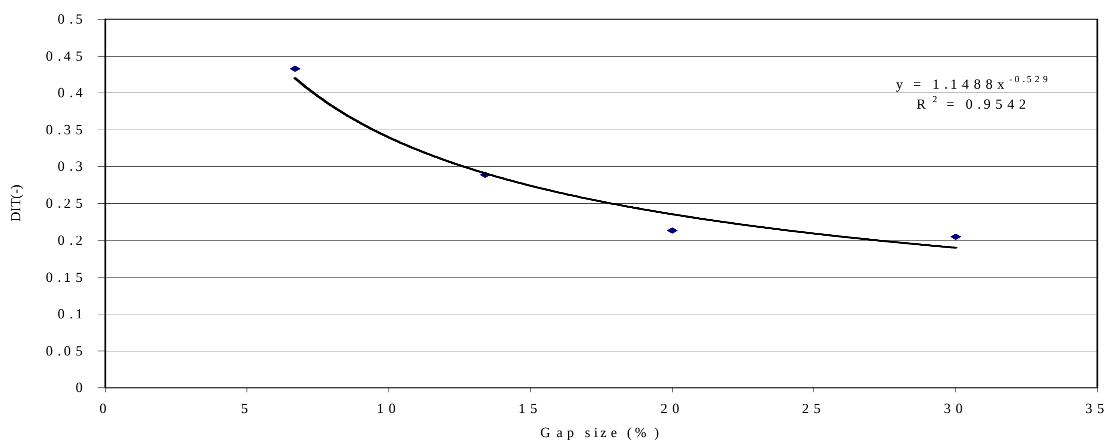
Generally, it was noticed that the performance of the different techniques depends on the gap duration at the target gauge and the station-pair involved in the estimation of missing annual maximum flows.



(a)

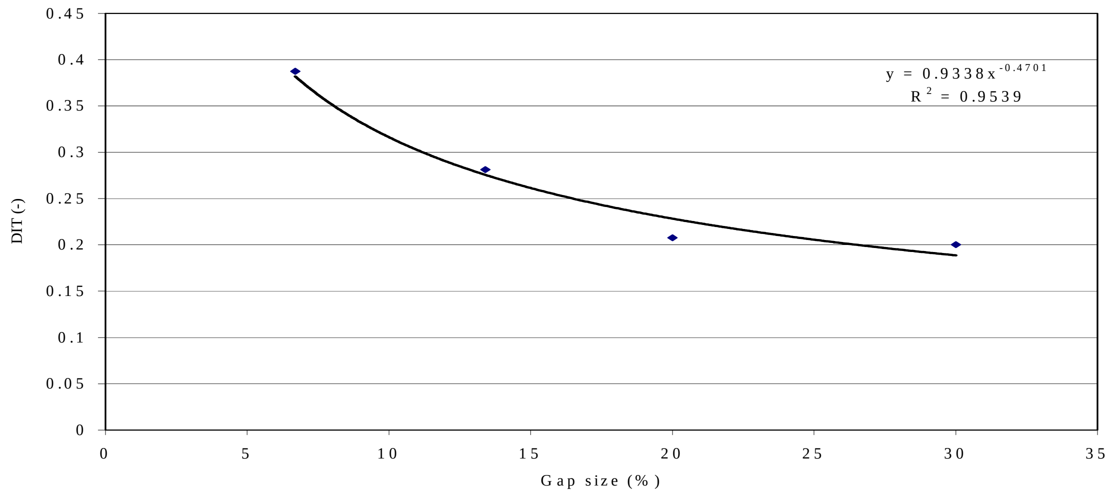


(b)

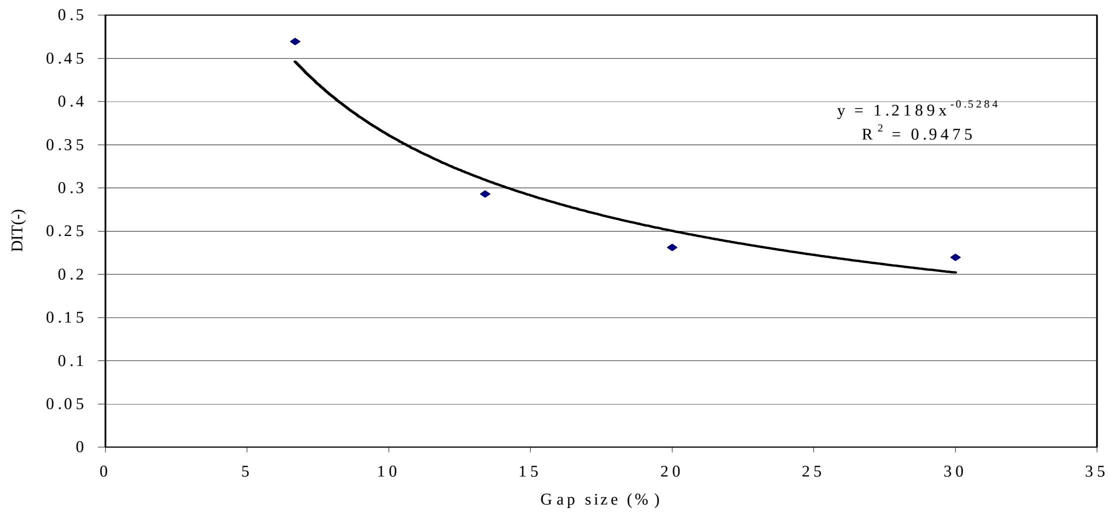


(c)

Figures 5.15 (a-c) DIT versus gap size for annual maximum flows at D1H003 (base gauge D1H009): (a) BP, (b) McL1BP, (c) QBP

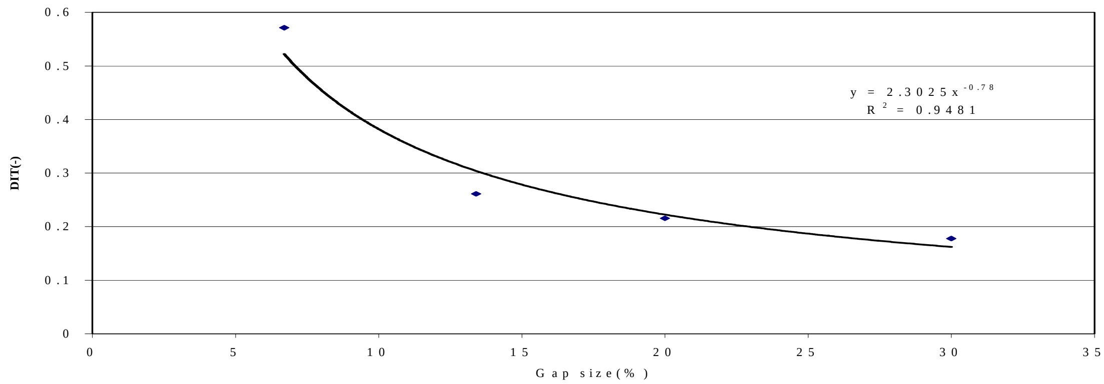


(d)

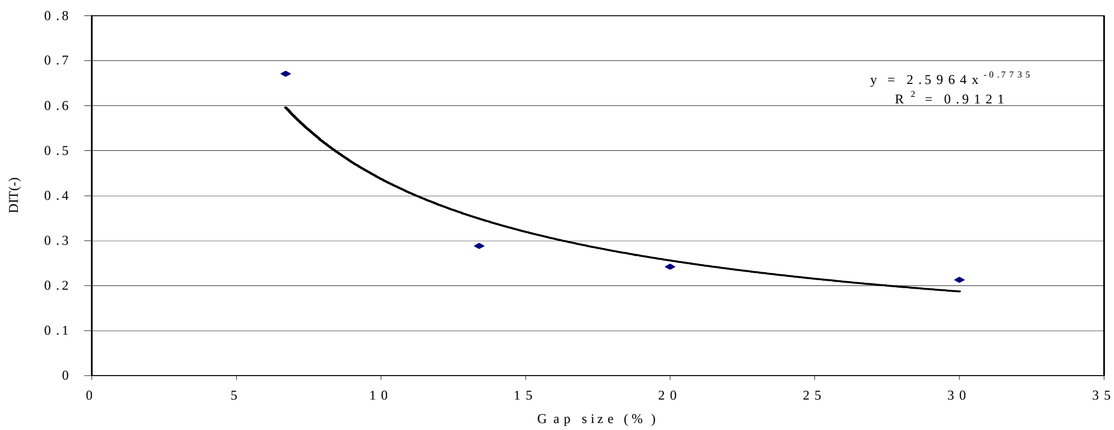


(e)

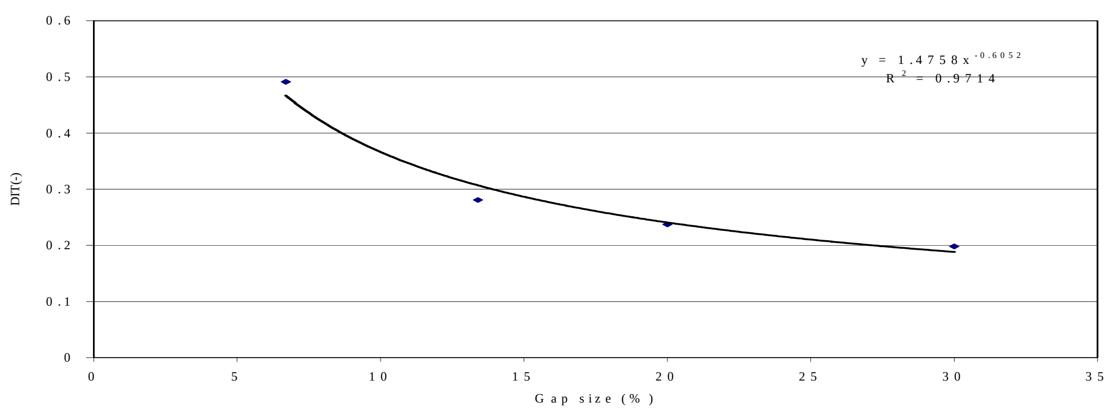
Figures 5.15 (d-e) DIT versus gap size for annual maximum flows at D1H003 (base gauge D1H009): (d) GoldSBP, (e) EM techniques



(a)

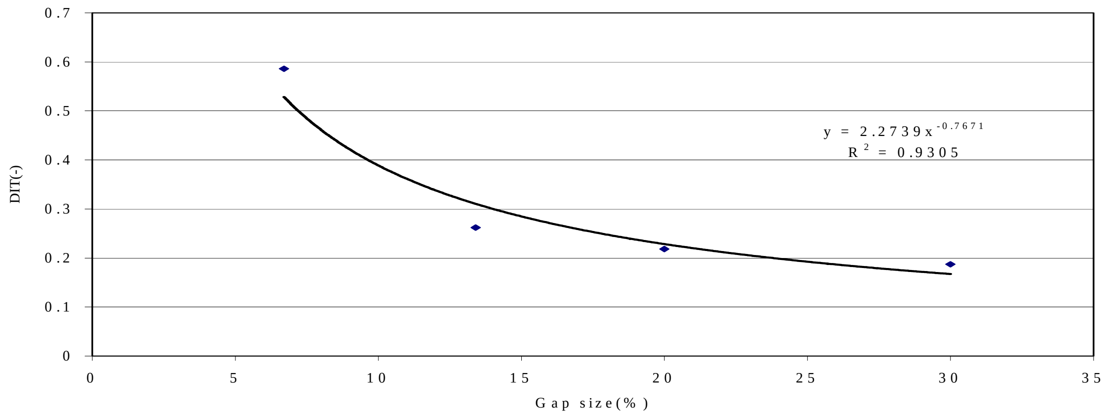


(b)

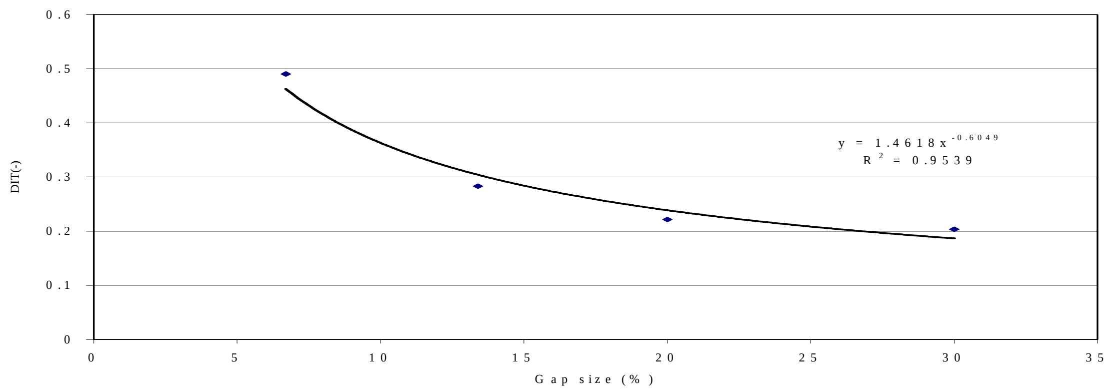


(c)

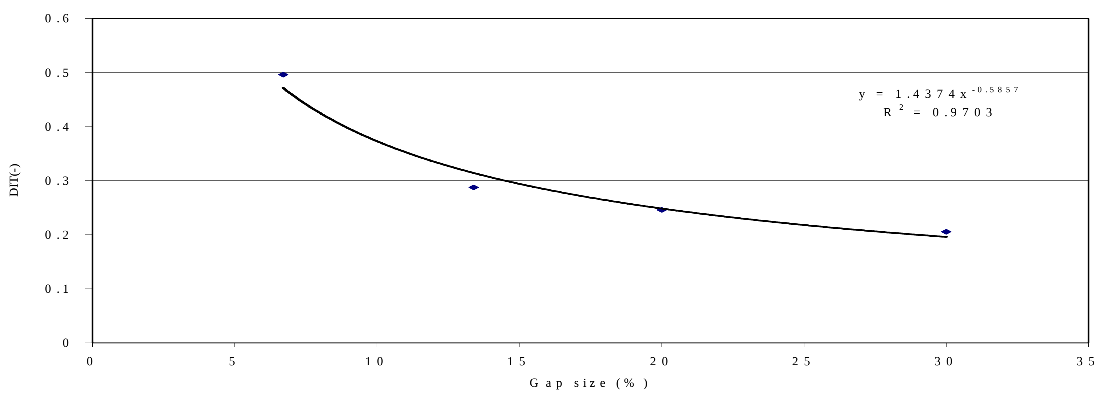
Figures 5.16 (a-c) DIT versus Gap size for annual maximum flows at D1H009 (base gauge D1H003) : (a) BP, (b) MacL1BP, (c) QBP



(d)



(e)



(f)

Figures 5.16 (d-f) DIT versus Gap size for annual maximum flow at D1H009 (base gauge D1H003): (d)GoldSBP, (e)VLR, (f) EM techniques