PROBABILISTIC DEPENDABLE HYDRO CAPACITY: THE BENEFITS OF SYNTHETIC HYDROLOGY

Konstantin Staschus[1] (non-member, ASCE)
Jerson Kelman[2] (member, ASCE)

ABSTRACT

Dependable hydro capacity has historically been defined deterministically as the capacity available in the worst drought on record. Recently, a probabilistic approach has been suggested which takes into account all historical water years, be they dry, average or wet. Taking the probabilistic approach a step further, a Monte Carlo simulation model can be used to generate synthetic streamflows that preserve certain parameters of the historical flows, but may include more extreme droughts and floods.

This paper focuses on the benefits of using synthetic streamflows (the "synthetic hydrology" approach) for determining a probabilistically defined dependable hydro capacity, as opposed to using historical flows only. A three-stage sampling experiment is used to determine which of the two approaches provides a more precise estimate of probabilistic dependable capacity in a case study. The case study analyzes the Central Valley Project in California. The results indicate that there is no significant advantage to using the more elaborate synthetic hydrology approach for most hydro reliability levels in this case. Possible explanations for this result are also offered.

1. INTRODUCTION

Synthetic hydrology, the generation of synthetic streamflows that preserve certain statistical properties of a set of observed streamflows, has been the subject of much work in the hydrological research literature (see, e.g., Loucks et al., 1981, Salas et al., 1980), and has occasionally been applied to help solve water resource-related decision problems (Frevert et al., 1986, Pereira et al., 1984, INTASA, 1981). Literature extolling the benefits of synthetic hydrology has been available for decades (e.g. Maass et al., 1962). Empirical studies comparing the benefits of different synthetic hydrology models are also available (e.g. Klemes & Bulu, 1979, Burges & Lettenmaier, 1981, Klemes et al., 1981). However, quantification of the benefits of synthetic hydrology vs. the use of historical data has been attempted very rarely (Lenton, 1978, Vogel & Stedinger, 1986).

Lenton (1978) showed that estimation of the mean flow is better

1 PG&E, 77 Beale, Rm. 2475, San Francisco, CA 94106
2 CEPEL - Electric Energy Research Center, Rio de Janeiro, Brazil

WATER RESOURCES OPERATION AND MANAGEMENT WORKSHOP, 3.
COLORADO, Julho 1988. 16p.

done without synthetic hydrology. He thus proved wrong the assumption that synthetic hydrology is always preferable to the use of historical data only. Vogel & Stedinger (1986) studied the required storage capacity to meet a given firm water demand over n consecutive years. The required storage capacity was defined as the median (over infinitely many n-year sequences) maximum cumulative deficit in n consecutive years (Gomide, 1975). Vogel & Stedinger considered two estimators for the required storage capacity:

a) The required storage capacity in n historical years, following Rippl (1883). The resulting single value provides an estimate of the median maximum cumulative deficit.

b) The median of M maximum cumulative deficits obtained from M n-year synthetic flow sequences generated with a synthetic hydrology model fitted to the historical n-year sequence.

Vogel & Stedinger concluded that (b) has a smaller root-mean-square error than (a), and that there were therefore significant benefits to the use of synthetic hydrology in this case.

Given Lenton's (1978) finding of no benefits for the determination of mean flows (a measure which is of little interest for reservoir design or reliability problems) and Vogel & Stedinger's (1986) finding of significant benefits for the determination of required storage capacity, the question of which problems justify the application of synthetic hydrology remains unanswered. This paper helps to answer that question. It employs the same test design suggested by Lenton (1978) and used by Vogel & Stedinger (1986). However, it applies the test to a different problem, that of evaluating reservoir reliability for electric power production. This problem requires an additional layer of modeling to translate water releases into energy generation, and consequently the implementation of the test design becomes more complex than in the previous two papers. Furthermore, while the previous papers addressed academic test data, the results of this study are of interest in the determination of the dependable hydroelectric capacity of California's Central Valley Project.

After a brief summary of synthetic hydrology and its uses in Section 2, Section 3 introduces the paper's case study. Section 4 outlines the test employed to evaluate the benefits of synthetic hydrology. Section 5 presents the results of the test; Section 6 discusses the results and possible implications; and Section 7 presents conclusions.

2. SYNTHETIC HYDROLOGY AND ITS USES

The modeling of a streamflow time series by a stochastic model to produce a large number of synthetic traces is usually done in four stages: model identification, parameter estimation, streamflow generation, and validation. The four steps are employed in the given order, but with an important feedback loop from validation back to model identification. In applications of synthetic hydrology the generated flows are usually input to a simulation model of the water resource system being considered. The output of this simulation model is then used in the evaluation of the system.

Of the four steps of the synthetic hydrology approach, parameter estimation and streamflow generation are fairly straightforward, although they may require significant computer resources. Model identification, however, requires significant user input and judgment, since it aims to select the member of a family of models which is most likely to best represent the underlying process from which the observed streamflow record was drawn. Selecting the most appropriate set of models for the different parts of a complex river basin can require many iterations between model identification and validation. In some cases, two quite different models may appear to fit equally well. The choice of models often has a significant impact on the results of the simulation model. Therefore, to the extent that the analyst is unsure whether the model choice is correct, water resource system evaluations based on synthetic hydrology results could be flawed.

The application of synthetic hydrology can thus require not only quite a bit of effort, but also a significant amount of statistical expertise and familiarity with the river basins to be modeled. Furthermore, it is imperative that the simulation models which use streamflows as an input are designed so that they can cope with synthetic streamflows that may contain droughts and floods more severe than those observed in the historical record. Many simulation models are designed around historical data, often with year-specific rule curves for reservoir operations, and may therefore be impossible to run with synthetic streamflows without major modifications. To the extent that year-specific rule curves are derived using perfect foresight, such models are also likely to yield biased, overly optimistic results. Thus, the use of synthetic hydrology may require analysts to redevelop such models, making them more realistic. The benefits of this improved modeling notwithstanding, such a redevelopment may constitute a major effort.

A legitimate question is which classes of problems justify the additional effort associated with the use of synthetic hydrology. For certain applications synthetic hydrology is a very helpful tool: these typically occur when the problem to be addressed requires a longer streamflow record than is available. Four examples follow.

Example 1: It may be desired to design a system of cascaded reservoirs for a 100-year planning horizon, but the available streamflow record may contain only 25 years of data. In such a case, the designer needs synthetic hydrology unless a rule of thumb design criterion is accepted.

Example 2: Dependable hydroelectric energy production is traditionally determined through driest period modeling. Under such an approach, if a hydro plant under study is to be integrated within a larger hydroelectric system, the driest historical period observed for the system is used to evaluate the benefits associated with the new plant. But during that period, there may have been above-average flows in the proposed plant's river, if that river is geographically remote from the system. Since above-average flows would lead to above-average energy production, a dependable energy larger than the mean energy would result. This is a counter-intuitive result. With synthetic hydrology, the river's dependable energy could alternatively be defined as the mean energy produced by the new plant during many driest periods of the system within synthetic streamflow traces that are as long as the historical record (Kelman et al., 1979).

Example 3: Section 1 described the work of Vogel & Stedinger (1986), who compared estimators of the median of the maximum cumulative deficit in an n-year sequence. There would not be a simple non-synthetic hydrology-based estimator if the decision criterion were any quantile of the maximum cumulative deficit other than the median.

Example 4: There is often a need to test operating policies against a wider range and combination of circumstances than may be captured by a limited historical record. The use of synthetic hydrology can help to prevent tuning an operating policy to the peculiarities of circumstances captured in that limited record.

The above examples describe cases in which synthetic hydrology is needed unless the evaluation criterion itself is changed. In contrast, this paper deals with a case in which synthetic hydrology could be used but is not necessary, because a long historical streamflow record is available. Given the substantial costs associated with the application of synthetic hydrology, a quantification of its benefits is appropriate.

Synthetic hydrology does not create any additional observed streamflow data. Therefore its use can only be beneficial because of the information imbedded in the mathematical model of streamflows used in the synthetic hydrology approach. The model structures typically employed can be interpreted as systematic and concise descriptions of the streamflow properties of many river basins. However, in the application of synthetic hydrology to any particular basin, the magnitude of this benefit is unknown.

3. THE CENTRAL VALLEY PROJECT CASE STUDY

California's Central Valley Project (CVP) is an assembly of federally owned and operated reservoirs, hydro power plants and irrigation canals. The CVP is contractually integrated with the Pacific Gas and Electric Company (PG&E), which provides gas and electric service to most of Northern California, for the purposes of coordinated electric power generation. The CVP's installed hydroelectric capacity of over 1800 megawatts (MW) constitutes an important part of Northern California's power resources.

"Capacity payments" between the federal government and PG&E are determined on the basis of the dependable hydroelectric capacity of the CVP, a possible definition of which is developed below. Since it generally takes at least five years to build new capacity, the CVP's dependable capacity (known as its Project Dependable Capacity or PDC) is determined with a five year leadtime. Therefore, the random variables of interest are the monthly capacities of the CVP in the fifth year in the future, given today's reservoir storage volumes. Let these random variables be denoted C_m, for the months m=1,..,12. To derive the cumulative distribution functions $F_m(c_m)$ of these random variables, all possible five-year sequences of CVP reservoir inflows would need to be run through the model that simulates the CVP operations (called the CVPower model). Assuming that this could be done, the monthly PDCs could then be defined as a quantile $c_m(p)$ - corresponding to probability p - of these distributions:

$$p=F_m(c_m(p)) \text{ or } c_m(p)=F^{-1}{}_m(p)$$

(The current integration contract prescribes deterministic, driest period-based, annual PDCs. The PDC definition used in this study is being discussed but is currently not implemented.)

The monthly PDCs would provide a measure of the extent to which the CVP's integration with PG&E's electric system would increase the reliability of the integrated CVP-PG&E system, and could become the basis for the capacity payments between the two parties, which could amount to tens of millions of dollars per year. Therefore, the PDCs should be determined as accurately as possible.

However, the true distributions $F_m(c_m)$ are not known since only a limited number of five-year sequences of reservoir inflows have been observed: one 83-year sequence of inflows (1895-1977) is available.

The $F_m(c_m)$ distributions, and with them the PDCs, can be estimated from the available historical data either using, or not using, synthetic hydrology. Without synthetic hydrology, the PDCs can be estimated in two different ways. First, one could run the CVPower program using the 16 non-overlapping five-year sequences that can be extracted from the historical record. This will result in 16 observations of C_m for m=1,..,12. These observations can be used to construct the empirical probability distributions $\hat{F}_m(c_m)$ (estimates of the monthly distributions $F_m(c_m)$), from which estimates $\hat{c}_m(p)$ of the monthly PDCs can be constructed.

Alternatively, 79 overlapping five-year flow sequences (without wrap-around) can be extracted from the historical record and processed by the CVPower program. This will yield 79 observations of C_m for m=1,..,12, but due to the overlapping, these observations will not constitute an independent random sample - they will exhibit high autocorrelations. These observations can again be used to construct estimates $\hat{F}_m(c_m)$ of the distributions $F_m(c_m)$, and thus to construct estimates $\hat{c}_m(p)$ of the monthly PDCs. This method, which has been suggested by Anderson et al., 1982 (also see House & Ungvari, 1983, Labadie et al., 1987), has been used within this paper to represent the non-synthetic hydrology-based approach. (This method was tested against the method using non-overlapping periods, using the same test that is discussed later in this paper, and was found to be superior.)

The synthetic hydrology-based approach used here is as follows. The 83 years of historical streamflow data are used with a synthetic hydrology program to derive a large number (in this case 200) of five-year sequences that, run through CVPower, will result in 200 observations of C_m, in estimates of the monthly distributions $\tilde{F}_m(c_m)$ and in estimates $\tilde{c}_m(p)=\tilde{F}^{-1}_m(p)$ of the monthly PDCs.

Throughout this paper, $c_m(p)$ and $F_m(c_m)$ are used to represent true PDCs and distributions, $\hat{c}_m(p)$ and $\hat{F}_m(c_m)$ represent PDCs and distributions derived using the non-synthetic hydrology-based approach, and $\tilde{c}_m(p)$ and $\tilde{F}_m(c_m)$ represent PDCs and distributions derived using the synthetic hydrology-based approach.

4. TESTING THE BENEFITS OF SYNTHETIC HYDROLOGY

For the PDC application, the benefits of synthetic hydrology can be defined in terms of the root-mean-square errors of the PDC estimates obtained using the synthetic hydrology-based approach ($RMSE[\tilde{c}_m(p)]$, $m=1,..,12$) and compared to the root-mean-square errors obtained using the non-synthetic hydrology-based approach ($RMSE[\hat{c}_m(p)]$, $m=1,..,12$). If $RMSE[\tilde{c}_m(p)]<RMSE[\hat{c}_m(p)]$ for most m, the use of synthetic hydrology would be beneficial since the resulting PDCs would be expected to be closer to the true PDCs.

However, the true PDCs and the root-mean-square errors of $\hat{c}_m(p)$ and $\tilde{c}_m(p)$ are unknown, since under either approach there is only one estimate per method for each month. This testing problem can be addressed through the application of synthetic hydrology on three different levels, as described below and shown in Figure 1. The first level is used to define 'true' flow distributions and PDCs. The second level generates 20 sequences of 83 years of flows. By interpreting each 83-year sequence as a possible 'historical' sequence, one can obtain 20 PDC estimates based on 'historical' data only, i.e. using the non-synthetic hydrology-based approach. At the third level, synthetic hydrology is applied to each of the 20 second-level 'historical' sequences, yielding 20 synthetic hydrology-based PDC estimates. This scheme is described in detail below.

4.1 Parent Model Level (Calculate 'True' PDCs)

The testing scheme starts with a synthetic hydrology model of CVP reservoir inflows: The SPIGOT model (Grygier & Stedinger, 1987, Stedinger & Taylor, 1982, Stedinger et al., 1985), a disaggregation-based synthetic hydrology model, generates annual flows at an 'aggregate' site and then disaggregates them to monthly aggregate flows, which can then be further disaggregated to monthly flows at major 'key sites' and finally at minor 'control points' (Grygier & Stedinger, 1987). In order to make test conditions as realistic as possible, the

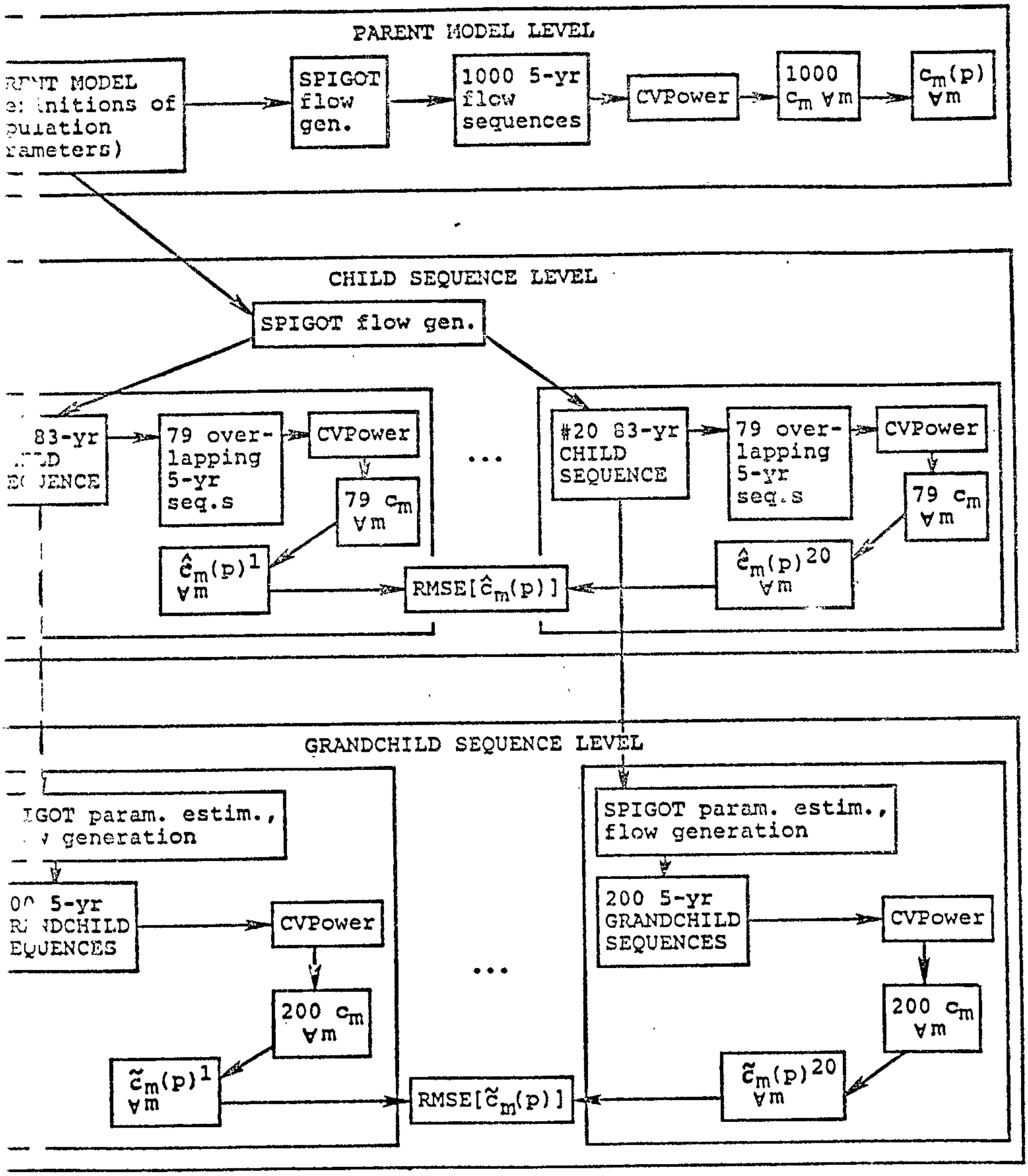

FIGURE 1: Benefits Assessment Testing Scheme

best-fitting model choices were used. For example, for the transformation of annual aggregate flows, which overshadows all other transformations in importance since all other flows are generated through disaggregation of these flows, a three-parameter lognormal model was assumed. The mean, standard deviation and skewness of annual aggregate flows are $13.4 \cdot 10^9 m^3$, $4.8 \cdot 10^9 m^3$, and 0.46, respectively, making this choice more appropriate than assuming normal or two-parameter lognormal flows. This model of CVP reservoir inflows will be called the 'parent' model.

From this parent model, 1000 five-year streamflow sequences were generated. CVPower was run with each sequence, resulting in 1000 observations of C_m for each month m, 12 monthly distributions $F_m(c_m)$, and 12 monthly PDCs $c_m(p)=F^{-1}{}_m(p)$. For the purposes of this test, these PDCs were defined to be the true monthly PDCs. $F_m(c_m)$ (and later $\hat{F}_m(c_m)$ and $\tilde{F}_m(c_m)$) were defined using piecewise linear functions, with straight lines connecting the points $(c_m, f(c_m))$, where $f(c_m)$ is the Weibull plotting position associated with observation c_m.

4.2 Child Sequence Level (Non-Synthetic Hydrology-Based Approach)

From the same parent model, 20 83-year CVP reservoir inflow sequences were generated. Let these 20 sequences be called the 'children'. Each of the 20 83-year child sequences can be regarded as a set of possible 'historical' data. The 83-year sequence actually observed could have been one of the 20 child sequences, and should be similar to many of them. From each of these 20 children, a set of monthly estimates $\hat{c}_m(p)$ was obtained using the non-synthetic hydrology-based approach: 79 overlapping five-year sequences were extracted from each 83-year child sequence, CVPower was run with each of these five-year sequences, and the 79 sets of output from CVPower were used to generate a set of monthly $\hat{F}_m(c_m)$ estimates of the fifth year CVP capacity distributions. The 20 sets of monthly estimates $\hat{c}_m(p)$ were then determined from $\hat{c}_m(p)=\hat{F}^{-1}{}_m(p)$. From these estimates, the PDC estimate variances $Var[\hat{c}_m(p)]$, biases $Bias[\hat{c}_m(p)]$ and root-mean-square errors $RMSE[\hat{c}_m(p)]$, m=1,..,12, were computed.

4.3 Grandchild Sequence Level (Synthetic Hydrology-Based Approach)

Just as one has the option to apply synthetic hydrology to the actually observed 83-year sequence, one can also interpret each of the 20 child sequences as a possible 'historical' sequence and apply synthetic hydrology. From each child sequence, 200 separate five-year sequences called the 'grandchildren' were generated. The CVPower program was then run with these 20 sets of 200 five-year inflow sequences, resulting in 20 sets of monthly $\tilde{F}_m(c_m)$ estimates of the fifth year CVP capacity distributions. From these distributions, 20 sets of monthly estimates $\tilde{c}_m(p)=\tilde{F}^{-1}{}_m(p)$, and the corresponding PDC variances $Var[\tilde{c}_m(p)]$, biases $Bias[\tilde{c}_m(p)]$ and root-mean-square errors $RMSE[\tilde{c}_m(p)]$, m=1,..,12, were obtained.

Each of the 20 83-year child sequences input to the SPIGOT synthetic hydrology model results in 20 sets of parameters that are similar but not equal to the parameters of the parent model, since for instance the sample mean of a generated 83-year sequence is not necessarily equal to the mean in the parent model. Such differences between child sequence parameters and parent model parameters would have been even more pronounced if SPIGOT's option to include parameter uncertainty had been used (see, e.g., Vicens et al., 1975, Davis, 1977, McLeod and Hipel, 1978, Damazio and Kelman, 1981, Stedinger et al., 1985). This was not done for the following reason.

The use of parameter uncertainty in synthetic hydrology attempts to account for the unknown bias introduced through the assumption that the population parameters equal the sample parameters of the observed streamflows. Not knowing whether this bias is positive or negative, synthetic hydrology models must assume that it could be either way, and generate parameters both smaller and larger than those observed. Because the resultant derived distribution of capacity reflects both the natural variability of streamflows and the parameter uncertainty, it yields biased (more conservative) quantile estimates, especially in the tails, of the distribution of capacity which reflects only the natural streamflow variability (Stedinger, 1983).

In many applications of synthetic hydrology, this effect is desirable, since it coincides with traditional risk-averse engineering practice, in which underdesign is considered worse than overdesign. For the test described above, however, the inclusion of parameter uncertainty in the generation of the grandchild sequences would decrease the value of $\tilde{F}^{-1}_m(p)$ for small p for every grandchild sequence, thus decreasing the $\tilde{c}_m(p)$ estimates for small p for every grandchild sequence and introducing a negative bias into these estimates. This would invalidate the use of the RMSE as a comparison criterion between the two approaches (recall that $RMSE=\sqrt{Bias^2+Var}$).

For similar reasons, the uncertainty in the choice of a model to describe flows within SPIGOT was ignored: the same model that was used to generate the child sequences was adopted for the generation of the grandchild sequences. This is equivalent to assuming perfect knowledge that nature's true distribution of flows belongs to a certain parametric family in the generation of each grandchild sequence. Although it is conceivable that a 'wrong model', with fewer parameters, could in some cases do a better job, this perfect knowledge assumption removes one possible source of error in the synthetic hydrology-based approach, and probably makes the resulting PDC estimates more precise.

5. RESULTS

Tables 1a-1e provide the results of the test for five distinct levels of p (0.02, 0.05, 0.1, 0.2, and 0.5). Means, standard deviations (S.D.), biases and root-mean-square errors (RMSE) of the monthly PDC estimates are given. For each level of p, first the 'true' c(p) are shown, followed by the statistics for the non-synthetic hydrology-based $\hat{c}(p)$ and the statistics for the synthetic hydrology-based $\tilde{c}(p)$.

TABLE 1
Monthly PDC Results for Synthetic Hydrology-Based and Non-Synthetic Hydrology-Based Approaches

Table 1a: p=0.02	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
'True' PDC	377	496	410	441	1463	1469	1319	1165	1088	435	388	423	802
Non-SH Mean	367	480	404	463	1233	1360	1317	1210	981	446	378	347	809
Non-SH SD	21	29	16	19	125	168	169	186	176	38	28	38	73
Non-SH Bias	-9	-15	-5	22	-224	-108	-1	45	-106	11	-9	-75	7
Non-SH RMSE	23	33	17	30	258	201	170	192	206	39	30	85	73
SH Mean	376	486	409	470	1302	1443	1400	1302	1078	470	392	368	851
SH SD	22	30	10	12	109	152	167	168	167	33	41	47	69
SH Bias	0	-9	0	29	-160	-25	81	137	-9	35	4	-54	49
SH RMSE	22	32	11	31	195	155	186	217	167	48	41	72	84

Table 1b: p=0.05	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
'True' PDC	411	535	433	462	1508	1560	1490	1392	1231	465	422	445	890
Non-SH Mean	401	527	425	504	1375	1520	1461	1367	1139	487	417	400	884
Non-SH SD	18	27	13	38	100	93	115	150	150	26	25	39	73
Non-SH Bias	-9	-7	-7	42	-132	-39	-28	-24	-91	22	-4	-44	-5
Non-SH RMSE	21	28	15	56	166	102	119	153	176	35	25	60	73
SH Mean	404	538	426	510	1409	1544	1503	1423	1198	496	420	401	907
SH SD	23	34	11	34	97	76	75	84	89	29	22	37	50
SH Bias	-6	3	-6	48	-98	-15	13	31	-32	31	-1	-43	17
SH RMSE	25	35	13	59	138	78	76	90	96	42	22	57	53

Table 1c: p=0.10	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
'True' PDC	429	567	454	476	1560	1596	1552	1489	1274	510	443	482	943
Non-SH Mean	429	573	452	580	1503	1590	1547	1483	1251	516	442	446	945
Non-SH SD	12	34	19	36	51	20	31	46	62	13	22	28	41
Non-SH Bias	0	6	-1	104	-56	-5	-4	-5	-22	6	0	-35	2
Non-SH RMSE	12	34	19	110	77	24	34	47	66	15	22	46	41
SH Mean	431	575	450	603	1503	1588	1545	1486	1278	518	448	446	951
SH SD	18	37	18	28	59	27	40	42	68	21	20	27	43
SH Bias	2	8	-3	127	-56	-7	-6	-2	4	8	5	-35	8
SH RMSE	18	38	18	130	82	28	40	42	68	22	21	45	44

Table 1d: p=0.20	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
'True' PDC	462	636	501	631	1586	1617	1581	1529	1413	538	507	546	1021
Non-SH Mean	467	670	506	656	1567	1613	1575	1523	1377	543	505	519	1008
Non-SH SD	20	69	29	24	22	6	9	11	57	19	43	46	35
Non-SH Bias	5	34	5	25	-18	-3	-5	-5	-35	5	-1	-26	-12
Non-SH RMSE	20	76	29	34	30	7	11	12	68	20	41	53	38
SH Mean	470	661	516	680	1569	1612	1574	1523	1382	540	507	521	1011
SH SD	139	96	130	35	12	2	2	2	5	19	54	50	35
SH Bias	8	25	15	49	-16	-4	-6	-5	-30	2	0	-24	-9
SH RMSE	25	63	36	60	30	9	13	15	65	18	51	58	38

Table 1e: p=0.50	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
'True' PDC	754	1112	938	872	1617	1628	1592	1550	1464	618	836	754	1152
Non-SH Mean	711	1059	875	876	1616	1628	1591	1549	1462	616	821	735	1135
Non-SH SD	125	88	100	46	12	2	2	2	4	27	32	56	31
Non-SH Bias	-38	-51	-61	7	0	0	0	0	-1	-1	-14	-17	-16
Non-SH RMSE	129	105	120	46	11	2	2	2	4	26	36	58	35
SH Mean	739	1060	911	881	1616	1628	1591	1549	1461	615	811	754	1143
SH SD	143	99	133	36	12	2	2	2	5	19	57	59	36
SH Bias	-10	-50	-25	12	0	0	0	0	-2	-2	-24	1	-8
SH RMSE	139	109	132	37	12	2	2	2	6	18	61	58	36

Table 1 Notes
- All values given are in MW
- Starting reservoir levels for each five-year simulation were set at flood control levels. In dry years, this assumption can be overly optimistic. Therefore the PDCs reported should not be considered indicative of the CVP's true capacity.
- Annual average PDC data is based on the random variable $c_a=(1/12)\sum_m c_m$ (the annual average fifth year capacity) and its distribution $F_a(c_a)$ before rank-ordering. Because of the independent rank-ordering by month for the $F_m(c_m)$ distributions, the annual PDC, $c_a(p)=F^{-1}{}_a(p)$ is not equal to the average of the monthly PDCs.
- Non-SH data is based on 20 non-synthetic hydrology-based PDC estimates $\hat{c}(p)$.
- SH data is based on 20 synthetic hydrology-based PDC estimates $\tilde{c}(p)$.

It would have been preferable to analyze more than 20 child sequences, in order to increase the precision of the estimates given in Table 1. For example, for p=0.02, a two standard deviation-wide confidence interval for the standard deviation of the synthetic hydrology-based annual PDC is approximately (37 MW, 90 MW), while the point estimate is 69 MW. However, the consideration of more child sequences would have been computationally prohibitive, and was also unnecessary for the conclusions drawn from the results.

Table 2 summarizes the results given in Table 1, noting by measure and value of p how many times each approach was superior, i.e. had a smaller S.D., bias or RMSE. The first number indicates how many times out of 13 (12 monthly PDCs plus annual average PDC) the synthetic hydrology-based approach was superior; the second number indicates how many times the non-synthetic hydrology-based approach was superior. The two numbers may not sum to 13 if neither approach was superior in some months.

TABLE 2
Number of Times SH Superior : Number of Times Non-SH Superior

p	0.02	0.05	0.10	0.20	0.50	All
S.D.	9 : 4	10 : 3	6 : 7	5 : 6	2 : 7	32 : 27
Bias	8 : 5	9 : 4	2 : 9	7 : 5	5 : 4	31 : 28
RMSE	7 : 6	9 : 4	4 : 9	3 : 8	2 : 7	25 : 34
All	24 : 15	28 : 11	12 : 25	15 : 19	9 : 18	88 : 89

Table 3 provides further summary information: it contains the RMSE of the synthetic hydrology-based and the non-synthetic hydrology-based annual PDCs, both scaled by the 'true' PDC for the five values of p.

TABLE 3
Normalized RMSE of Estimated Annual PDC

p	SH RMSE / True PDC	Non-SH RMSE / True PDC
0.02	0.105	0.091
0.05	0.060	0.082
0.10	0.047	0.043
0.20	0.037	0.037
0.50	0.031	0.030

It can be seen from Tables 1 through 3 that neither of the two approaches is clearly superior to the other, although for very low levels of p, the synthetic hydrology-based approach appears to be slightly superior (see Table 2). The Smirnov two-sample test was applied for all months m and all values of p in order to check the null hypothesis that $\hat{C}_m(p)$ and $\tilde{C}_m(p)$ have the same underlying population distribution. In no case could the null hypothesis be rejected at a five percent significance level.

A more accurate comparison would have been possible if more than

20 child sequences had been used. However, the goal of the test was not to obtain a precise measure of any possibly existing difference in RMSE between the two approaches. Rather, it was sufficient to learn that if indeed there is a difference, the difference must be small.

6. DISCUSSION

This section discusses the test results given above. Three issues will be addressed.

(i) The non-synthetic hydrology-based annual average fifth year capacities exhibited significant lag-1 autocorrelations due to the dependence introduced through the overlapping. These autocorrelations varied from child sequence to child sequence, ranging from 0.21 to 0.53, and averaging 0.39. Because of the absence of any such autocorrelation in the grandchildren, one might have expected an advantage to the synthetic hydrology-based approach. However, the test results suggest that even such autocorrelation did not lead to a significant advantage for the synthetic hydrology-based approach.

(ii) Going beyond the data-based conclusion that neither of the two approaches is clearly superior in this case study, it is next possible to speculate under which conditions a synthetic hydrology-based approach might be superior, and thus offer ideas for future research. It can be seen from Tables 1 through 3 that any future search for benefits of the synthetic hydrology-based approach should concentrate on small values of p, for which the test results show some benefits of synthetic hydrology. Such benefits could be due to the tails of $\hat{F}_m(c_m)$ containing very few data points from which to interpolate the desired low quantiles.

The possible superiority of synthetic hydrology at low levels of p concurs with the findings of Vogel & Stedinger (1986), who reported that the use of synthetic hydrology "leads to much more precise estimates" of the required storage capacity of a reservoir given a 50-year flow record. Required reservoir storage capacity is a function of the severity of the worst drought within a flow sequence. It is thus a measure akin to the PDC corresponding to a very low p. In fact, for values of p smaller than 1/80, there is no alternative to the synthetic hydrology-based approach, unless one fits a parametric distribution to $\hat{F}_m(c_m)$.

The possible benefits associated with the use of synthetic hydrology appear to be based on the extent to which the actual application requires an accurate definition of the tail quantiles of the random variable of interest. The value of using synthetic hydrology would therefore depend on two conditions. First, it must be difficult to define the tails of the distribution of the variable of interest directly from the available observations of this variable. This could be the case, for example, when the relationship between streamflows and the variable of interest is highly nonlinear. If this relationship is linear, additional synthetic low-flow values may not provide more information about the distribution of the variable of interest than interpolations between the observed values of that variable. Second, the streamflow model used in the synthetic hydrology approach must produce rare events with correct probabilities. (This condition must in fact be fulfilled in any application of synthetic hydrology.)

(iii) Figure 2 shows the relationship between streamflows (the fifth year aggregate annual CVP reservoir inflows) and the variable of interest (the fifth year annual capacity c_a) for this case study, for 200 synthetic five year sequences generated with the parent model. The correlation between flows and c_a is 0.88. The figure shows that the relationship between the two variables is almost linear, even in the range of very low flows. (In the range of very high flows, the curve flattens out because c_a is bounded by the installed generating capacity.) Given point (ii) above, this linearity could possibly have led one to expect that the synthetic hydrology-based approach would not be superior to the non-synthetic hydrology-based approach for the estimation of $c_a(p)$. Similarly, nonlinearities in the flow-c_m relationships for $m=6,\ldots,9$ at $p=0.05$ correspond to some superiority of the synthetic hydrology approach for the estimation of $c_m(0.05)$ for these months.

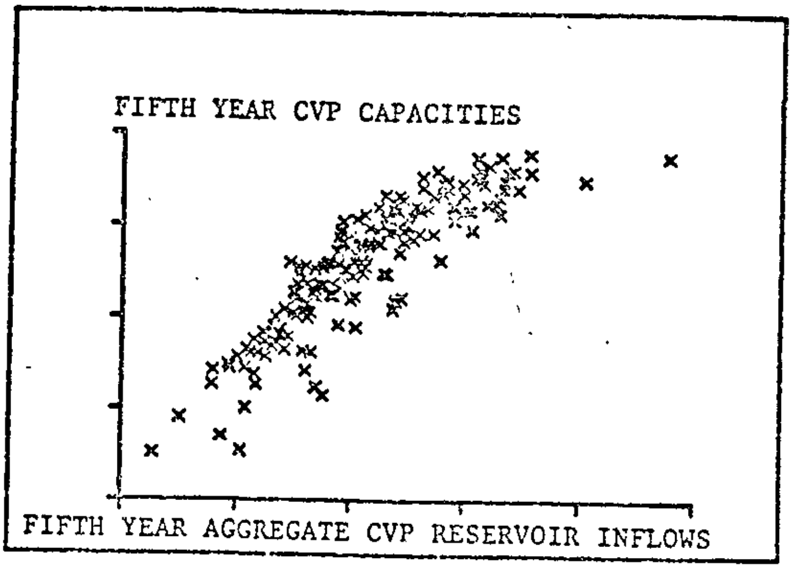

Figure 2: Central Valley Project Flow-Capacity Relationship

7. CONCLUSION

This paper has described an experiment to evaluate the benefits of using synthetic hydrology, as opposed to using historical data only, for determining reliability-based dependable electric generating capacity levels for California's Central Valley Project. The experiment showed that there is no significant advantage to using the synthetic hydrology-based approach for most hydro reliability levels. The test results were sufficient to conlude that if in fact one of the approaches is superior, the difference is of small practical importance for the application studied in this paper. Because the non-synthetic hydrology-based approach is much simpler to implement, it is therefore difficult to justify the use of synthetic hydrology in this particular case. No claim is made that this is a general result. As mentioned in Section 2, there are important problems for whose solution synthetic hydrology is needed. However, our results show that there is no reason to assume that a synthetic hydrology-based approach will always be preferable to simpler solutions.

ACKNOWLEDGEMENTS

We thank C. F. Imparato, J. R. Stedinger and R. M. Vogel for their extensive comments on drafts of this paper. Parts of our discussion section are the offspring of discussions with J. R. Stedinger.

REFERENCES

Anderson, J.D., Freeny, G.B., Ungvari, J.L., House, P.M. 1982: Application of the Indexed Sequential Streamflow Hydrological Modeling Approach to the Central Valley Project, Golden, CO: WAPA

Burges, S.J., Lettenmaier, D.P. 1981: Reliability Measures for Water Supply Reservoirs and the Significance of Long-Term Persistence, International Symposium on Real-Time Operation of Hydrosystems, Waterloo, Ontario: University of Waterloo

Damazio, J., Kelman, J. 1981: Bayesian Approach to Hydroelectric Studies (in Portuguese), Brazilian Journal of Engineering 4

Davis, D.R.: 1977, Comment on "Bayesian Generation of Synthetic Streamflows" by Vicens, G.J., Rodriguez-Iturbe, I., Schaake Jr., J.C., Water Resources Research 13

Frevert, D.K., Cowan, M.S., Lane, W.L.: 1986, Use of Stochastic Hydrology in Reservoir Operation, Water Forum '86, Vol. 2, 1904-1911, New York, NY: ASCE.

Gomide, F.L.S.: 1975, Range and Deficit Analysis Using Markov Chains, Hydrology Papers 4, No. 79, Colorado State Univ.

Grygier, J.C., Stedinger, J.R.: 1987, SPIGOT: A Synthetic Flow Generation Software Package, User's Manual, Ithaca, NY: Cornell University

House, P.M., Ungvari, J.L. 1983: Indexed Sequential Modeling of Hydropower System Operations, Waterpower 83 Conference Proceedings

INTASA, Inc. 1981: A Statistical Method to Determine Dependable Hydrogeneration, INTASA and Linsley, Kraeger Associates

Kelman, J., Gomes, F.B.M., Pinheiro, S.F., Pereira, M.V.F. 1979: Substitution of Firm Energy Through the Use of Synthetic Streamflow (in Portuguese), Fifth Brazilian Seminar on Production and Transmission of Electric Energy, Recife, Brazil

Klemes, V., Bulu, A. 1979: Limited Confidence in Confidence Limits Derived by Operational Stochastic Hydrologic Models, Journal of Hydrology 42, 9-22

Klemes, V., Srikanthan, R., McMahon, T.A. 1981: Long-Memory Flow Models in Reservoir Analysis: What is Their Practical Value?, Water Resources Research 17, 737-751

Labadie, J.W., Fontane, D.G., Tabios III, G.Q., Chou, N.F. 1987: Stochastic Analysis of Dependable Hydropower Capacity, Journal of Water Resources Planning and Management 113, 422-437

Lenton, R.L. 1978: A Review of the Use of Operational Hydrology for River Basin Simulation Models, unpublished manuscript, New Delhi, India: The Ford Foundation

Loucks, D.P., Stedinger, J.R., Haith, D.A. 1981: Water Resource Systems Planning and Analysis, Englewood Cliffs, NJ: Prentice Hall

Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas Jr., H.A., Marglin, S.A., Fair, G.M. 1962: Design of Water Resource Systems, Cambridge, MA: Harvard University Press

McLeod, A.I., Hipel, K.W. 1978: Simulation Porcedures for Box-Jenkins Models, Water Resources Research 14

Pereira, M.V.F., Oliveira, G.C., Costa, C.C.G., Kelman, J. 1984: Stochastic Streamflow Models for Hydroelectric Systems, Water Resources Research 20, 379-390

Rippl, W. 1883: The Capacity of Storage Reservoirs for Water Supply, Proceedings of the Institute of Civil Engineers 71

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L. 1980: Applied Modeling of Hydrologic Series, Littleton, CO: Water Resources Public.s

Stedinger, J.R. 1983, Design Events With Specified Flood Risk, Water Resources Research 19, 511-522

Stedinger, J.R., Pei, D., Cohn, T.A. 1985: A Condensed Disaggregation Model for Incorporating Parameter Uncertainty Into Monthly Reservoir Simulations, Water Resources Research 21, 665-675

Stedinger, J.R., Taylor, M.R. 1982: Synthetic Streamflow Generation, 1: Model Verification and Validation, 2: Effect of Parameter Uncertainty, Water Resources Research 18, 909-924

Vicens, G.J., Rodriguez-Iturbe, I., Schaake Jr., J.C. 1975: Bayesian Generation of Synthetic Streamflows, Water Resources Research 11

Vogel, R.M., Stedinger, J.R. 1986: The Value of Stochastic Streamflow Models in Over-Year Reservoir Design Applications, working paper, Medford, MA: Tufts University

PROBABILISTIC DEPENDABLE HYDRO CAPACITY: THE BENEFITS OF SYNTHETIC HYDROLOGY

List of Table and Figure Captions

Keywords

Synthetic hydrology, stochastic hydrology, Central Valley Project, dependable capacity, hydroelectric capacity reliability, quantile estimation