Sandbars in the Colorado River: an Environmental Consulting Project

Jennifer A. Hoeting
Colorado State University

Brian Cluer
National Park Service

Kristina Varga*
Colorado State University

October 8, 1996

Abstract

The National Park Service funded a study to determine the impact of water released from the Glen Canyon Dam on sandbars downriver through Grand Canyon National Park. The project involves messy data with small sample sizes and considerable missing data. A weighted mixed regressive auto-regressive model was developed to predict average net change in sandbar size by the previous two observations as well as mean daily discharge from the dam and sediment input from a tributary. Model limitations and sampling issues are discussed.

KEYWORDS: Weighted Mixed Regressive Auto-Regressive model, sampling interval, environmental monitoring.

1 Sandbars?

From September 1990 to July 1991, 17 helicopter flights were made above 230 miles of the Colorado river below the Glen Canyon Dam (Figure 1). On each flight approximately 58 out of the total population of about 600 sandbars along the river were photographed. These photographs were digitized in order to determine the size of the sandbar at each flight (Cluer 1995b).

Measuring sandbar sizes may sound like another government boondoggle, but sandbars play a key role in the ecosystem of the Colorado River. For birds and insects, the sandbars offer a small strip of riparian habitat in a harsh desert environment. The sandbars also create eddies where endangered fish and other fauna feed. Finally, rafters camp on the sandbars during their trip down the Colorado River.

When Glen Canyon Dam was opened in 1966, the annual flood cycle was eliminated as the dam controlled all water flow. Floods scour the river bottom, bringing up sediment deposited there. When flood waters recede, the sediment is left on the shore of the river in the form of sandbars. Surveys of the river show that sandbars have decreased in size and number since the dam opened in 1966 (Kearsley *et al.* 1994). Not only do fewer sandbars mean reduced habitat for fish and other wildlife, but reduced numbers of sandbars force all campers to use the same sandbars, thereby increasing the user impact on a fragile environment. For these reasons, the National Park Service would like to determine how changing water release rates or patterns from Glen Canyon Dam influences sandbar size.

2 The Data

Net change in sandbar size (the difference between the sandbar size for the current flight and

---

*Jennifer Hoeting (jah@stat.colostate.edu) is Assistant Professor of Statistics and Kristina Varga is Graduate Research Assistant at the Department of Statistics, Colorado State University, Fort Collins, CO 80523. Brian Cluer is Hydrologist at the National Park Service. This research was supported by the National Park Service.

Table 1: Selected Summary Statistics

Variable	Mean	s.d.	Range
Days Between Flights	20 (Days)	15	12 – 70
Mean Daily Discharge	336 (cms)	83	194 – 491
Sediment supply	86,268 (tons)	215,315	0 – 780,000
Net Change	14 (m^2)	414	−2,368 – 2,513

the sandbar size for the previous flight) was selected as the response (Figure 2).

In addition to the sandbar size measurements recorded via photographs, hydrological data and sandbar characteristics were recorded. The helicopter flights were made during periods when the water release was at a constant level. In between flights, water was released from the dam in different patterns of discharge. The mean daily discharge, transport capacity (potential for sediment movement), and several other discharge characteristics were recorded. Sediment supply from the Little Colorado River was also measured, as sediment from this tributary of the Colorado river could impact sandbar size (Figure1). Sediment supply is the average amount of sediment per day entering from the Little Colorado River during each inter-flight period. Sandbar characteristics included location in terms of miles from the dam, left or right bank, and type of sandbar. Selected summary statistics are given in Table 1.

The dam release measurements are summarized on a "per flight" basis while the sandbar characteristics are collected on a "per sandbar" basis. While this difference in measurement scale makes the analyses somewhat complicated, the real challenge is the large amount of missing data. Out of 986 possible observations for sandbar size (58 sandbars × 17 flights), only 692 were available. The data was missing for various reasons, but primarily due to blurry photographs. This leads to a very incomplete database with which to model a complicated space/time relationship.

3 Predicting Sandbar Sizes

The initial goal of the project was to produce a space/time model to predict sandbar sizes based on sandbar characteristics and dam release measurements. The model was hoped to provide scientists with some guidelines on how different patterns of dam water releases impacted different types of sandbars.

Statisticians became involved only after the flights had been made and all the data collected. Preliminary data analysis along with a survey of available methodology indicated that the large amount of missing data made this goal unattainable. This was disappointing because this is the best data set ever obtained for a sample of Grand Canyon sandbars; indeed, a large sample of sandbars was monitored over a long period of time. By averaging over the net change in sandbar size, we reduced the space/time problem to a time series model. This prevents us from making inferences about different *types* of sandbars, but still allows us to investigate the relationship between water release patterns and sandbar size.

We adopted a Weighted Mixed Regressive Auto-Regressive (WMRAR) model. This is an extension of the Mixed Regressive Auto-Regressive model of Ord (1975). The WMRAR model has the following form:

$$Y = X\beta + \rho WY + \epsilon,$$

where ϵ is an n-vector of random errors assumed to follow a multivariate normal distribution with mean 0 and variance $\sigma^2\Sigma$ where Σ is known, and $W = [w_{i,j}]$, where w_{ij} is a nonnegative weight which represents the 'degree of possible interaction' of observation i and j and $w_{ii} = 0$.

Direct extension of Ord's approach to maximum likelihood estimation for the parameters of the model yields the estimates

$$\hat{\beta} = (X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}(I - \rho W)Y, \text{ and}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left[Y^t(I - \rho W)^t A\Sigma^{-1}A(I - \rho W)Y\right]$$

where $A = I - X(X^t\Sigma^{-1}X)^{-1}X^t\Sigma^{-1}$, and $\hat{\rho}$ is the value of ρ that minimizes

$$-2n^{-1}ln|(I - \rho W)| + ln\hat{\sigma}^2.$$

The WMRAR model allowed us to account for the non-symmetric sampling intervals and unequal sample sizes. To account for the non-symmetric flight intervals, we used $w_{ij} = 1/(\# \text{ of days between flights})$. Since the ith response is an average of n_i equally weighted observations, we assumed $\text{var}(y_i) = \sigma^2/n_i$ and thus Σ is a diagonal matrix with $1/n_i$ on the diagonal and 0's elsewhere.

Model selection, driven by both residual plots and model fit as measured by minimizing $\hat{\sigma}^2$, indicated that mean daily discharge and the natural log of the average sediment supply from the Little Colorado River were the best predictors. Since sediment input from the Litte Colorado can only influence those sandbars down stream of the confluence of the Little Colorado and the Colorado, only those 37 sandbars downstream of the confluence were included in the analysis. Including the observations from the previous 2 flights in the auto-regressive portion of the model produced the best fit.

The estimates for the coefficients of this model are given in Table 2. A limited amount of variability in the response was explained by the model as reflected in an R^2 of 50%. The model suggests that as mean daily discharge increases, net change increases and as sediment supply increases, net change decreases. Finally, $\rho = -0.81$ indicates that as the net change for the previous 2 flights increases, net change decreases. The large negative value of ρ is rather surprising for these highly variable data, but there is considerable uncertainty about this value as reflected by the large standard deviation $\hat{\sigma}_\rho = 0.23$.

Table 2: Coefficients for Model

Parameter	Estimate	s.d.
Intercept	-135.33	14.73
Mean Daily Discharge	0.55	0.04
ln(Sediment supply)	-9.64	0.87
ρ	-0.81	0.23
σ^2	321.53	

Attempts to simplify this model, including alternative lags, alternative predictor sets, accounting for initial sandbar size in the model, and simple weighted least squares, all yielded less satisfactory models.

4 Limitations of the Analysis

There are several limitations with the model described above. A primary constraint is the small sample size. Maximum likelihood estimation for a WMRAR model based on 15 observations is questionable at best. We have indicated to our client that the main use of the model should be for selection of potentially important predictors, but the coefficient estimates are somewhat questionable.

Nevertheless, it is natural to want to interpret the signs of the coefficients. The negative sign of the coefficient for sediment supply is somewhat counterintuitive as scientific evidence indicates that increased sediment should increase sandbar size. Exploratory plots suggest that increased sediment may take longer than one flight period to impact sandbars (which is not surprising since sediment moves slower than water); however, the unsymmetric flight intervals make inclusion of a lagged sediment supply predictor somewhat difficult. This issue will be explored in future analyses.

Another limitation of the model is that it is based upon *average* net change. While computing the mean net change for each flight did reduce variability, there is always the risk of ecological correlation, the inflation of correlations based on average values. This concern sheds some light upon why our model estimates a highly negative auto-correlation (ρ) for very variable data.

The primary concern, however, is the long time interval between observations. Recent data has shown that it is common for large-scale rapid erosion events to occur in a matter of days or even over several hours. An ad-hoc analysis of previous sandbar surveys indicates that as the time span between measurements increases, estimates of net erosion decrease (Cluer 1995a). Since data analyzed in this paper were collected at intervals between 12 and 70 days, we have a rather incomplete picture of what actually happened to the sandbars.

The issue of the importance of a smaller sampling interval is included in our final report to the NPS. We argue that not only will more frequent sampling result in better understanding of the underlying natural processes, but more frequent sampling of fewer sandbars can save money. Traditional sampling techniques use either aerial photography or land-based surveying. Aircraft flying at low altitude deep in the Grand Canyon is expensive, dangerous and ecologically unsound. Land surveying is similarly expensive and time consuming so it is best to budget for few flights or few surveys where many sandbars are measured. An alternative design would be to set up automatic cameras at a few sandbars to take photographs at specified intervals. While the second set-up would reduce the number of sandbars included in a study, we show that more information about the problem of interest is gained through this design.

5 Conclusions

What did the National Park Service gain from our work on this project? The limitations of our analyses underline the importance of considering an appropriate analysis before the data are collected. In this case, the data were difficult to model well because of several faults in the design that could be avoided in future studies. The NPS principal investigator (Cluer) came away from the project with additional ammunition to convince other scientists and resource managers of the importance of collecting more frequent sandbar size measurements for fewer sandbars as opposed to completing a few measurements at many sandbars.

What did the statisticians learn from analyzing these data? Hoeting and Varga were again reminded of the limitations of available methods for modeling data with missing values. For Hoeting, who headed up the project, the credo that a consultant should learn as much as possible about the scientific background of the project was reinforced. In this case, the statisticians were unaware of the main limitations of the data until well into the project: the data were collected on what seems to be an inappropriate scale. However, the scientists were also unaware of the time scale limitation when the sampling design was established. While our results will definitely be useful, similar analysis of *daily* data would greatly improve the scope of our results.

Another goal was to use the project as a teaching tool. Varga learned a great deal about the challenges of modeling real data and her efforts on the project resulted in the project paper required for her master's degree.

We did provide the NPS with evidence that mean daily discharge and sediment supply may be useful in predicting changes in sandbar size. In the past, statistical analyses of sandbar size data has been limited. Previous analyses include simple models that ignore spatial and time correlations (Beus and Avery 1992; Cluer 1995b). While our models are not completely satisfactory, the results presented here are a step in the right direction towards the development of a useful model to predict sandbar sizes.

Discharge of water through Glen Canyon dam has been controlled to optimize peak load hydropower production. The U.S. Bureau of Reclamation (USBOR), which runs the dam, faces the continual challenge of balancing the needs of the ecosystem with the needs of the power companies. A large controlled flood was carried out in the spring of 1996. One of the goals of the controlled flood was to reinvigorate the sandbars on the Colorado River. Preliminary observations show that this goal was at least partially achieved (Dave Wegner, USBOR, personal communication).

Our analyses, along with analyses of data from the controlled flood will help scientists learn more about the relationship between sandbar size and dam water releases. Statisticians can play a key role in this process by helping to guide good data collection as well as by continuing to develop methodology to assist in the analyses of these and similar data. Involvement of statisticians can not occur soon enough as the need to understand the impact of dams on ecosystems is continually increasing: it is predicted that by the year 2000 over 60% of the world's rivers will be regulated (Gore and Petts 1989).

References

Barnett, G.; Kohn, R.; and Sheather, S. (1996) "Bayesian Estimation of an Autoregressive Model Using Markov Chain Monte Carlo", Working Paper, Australian Graduate School of Management, University of New South Wales.

Beus, S.S. and Avery, C.C. (1992) "Final Report: The Influence of Variable Discharge on Colorado River Sand Bars Below Glen Canyon Dam: Final Report", National Park Service technical report.

Cluer, B. (1995a), "Cyclic Fluvial Processes and Bias in Environmental Monitoring, Colorado River in Grand Canyon", *The Journal of Geology*, **103**, 411-421.

Cluer, B. (1995b), "Final Report: Aerial Photography of the GCES-II Test Flows and Interim Flows", National Park Service technical report.

Gore, J.A. and Petts, G.E. (1989) *Alternatives in Regulated River Management*, CRC Press, 344 p.

Kearsley, L.H.; Schmidt J.C.; and Warren K.D. (1994) "Effects of Glen Canyon Dam on Colorado River Sand Deposits Used as Campsites in Grand Canyon National Park, USA", *Regulated Rivers: Research and Management*,**9**, 137-149.

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction", *Journal of the American Statistical Association*, **70:349**, 120-126.

Approximate location of Figure 1, see attached map.

Figure 1: Geography of the Colorado River in the Grand Canyon.

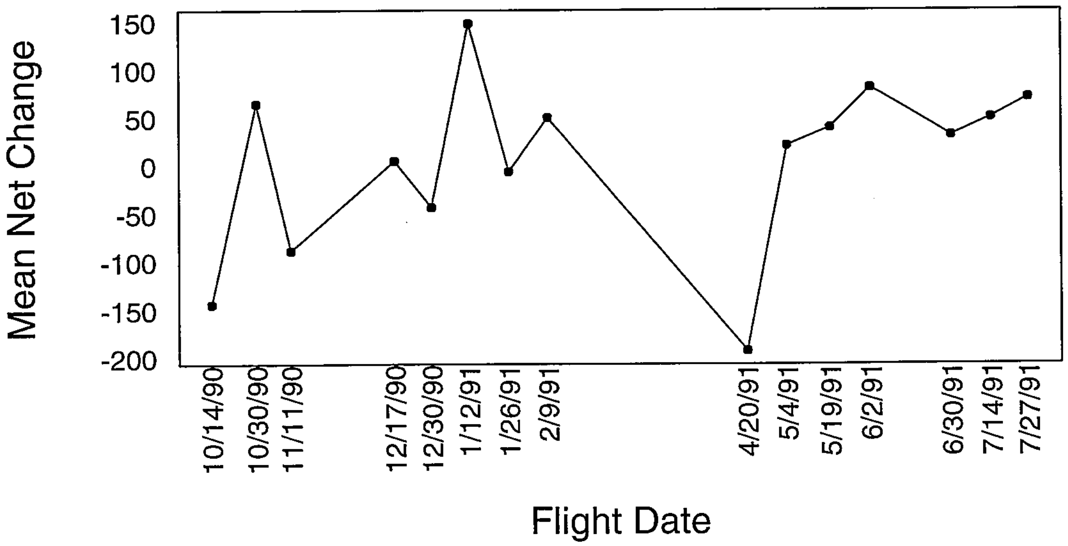
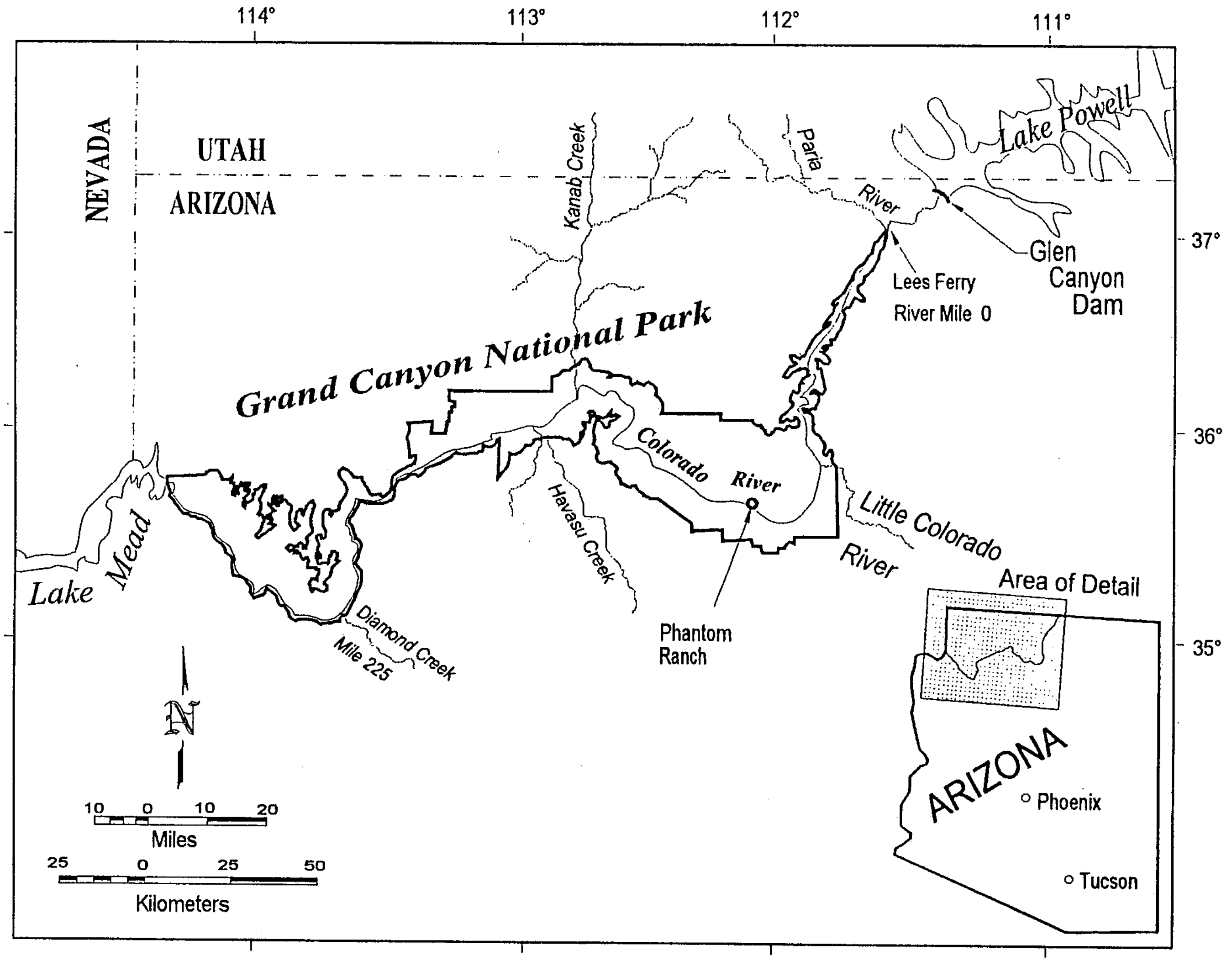


Figure 2: Mean net area change (in square meters) for 37 sandbars downstream of the Little Colorado River versus flight date.

114°
113°
112°
111°
37°
36°
35°
NEVADA
UTAH
ARIZONA
Kanab Creek
Paria
River
Lake Powell
Glen Canyon Dam
Lees Ferry River Mile 0
Grand Canyon National Park
Colorado River
Little Colorado River
Havasu Creek
Phantom Ranch
Lake Mead
Diamond Creek
Mile 225
Area of Detail
ARIZONA
Phoenix
Tucson
N
10 0 10 20
Miles
25 0 25 50
Kilometers