Supplementary Material

Salinity intrusion in a modified river-estuary system: an integrated modeling framework for source-to-sea management

S1: Required River Discharges

The focus here was on low discharge conditions during which salinity extends farther upstream so the model was run with fixed values of river discharges Q_r = 75m^{3}s^{-1}, 150m^{3}s^{-1}, and 300m^{3}s^{-1} for 50 days after reaching an equilibrium to incorporate a representative range of tidal forcing (Fig. S1). Solutions for salinity intrusions L_x were obtained for each discharge case, with the range of L_x for a given discharge due to tidal fluctuations. A relationship between intrusions and discharges was approximated by a power law, $L_x = \beta Q_r{}^{\gamma}$, where β and γ were parameters that governed the fit (Abood 1974; Monismith *et al*. 2002). The fitted relationship was used to calculate a river discharge $\hat{Q}_r$ that satisfied Pr($L_x \geq 120$km) = δ. (Equivalently, a river discharge equating to an exceedance of the salinity threshold at Poughkeepsie, δ% of the time was calculated.) The estimator was:

$$\hat{Q}_r = \left(\frac{120 - \Phi(1-\delta)\hat{\sigma}}{\hat{\beta}}\right)^{\frac{1}{\hat{\gamma}}}$$

where $\hat{\beta}$ and $\hat{\gamma}$ are least squares estimates, $\hat{\sigma}$ is an estimated residual variance, and $\Phi(1-\delta)$ is the value of the standard normal distribution for which δ% of the model solutions for L_x have been exceeded at Poughkeepsie.

A $(1-\alpha)$ confidence interval for Q_r can be found through a non-parametric bootstrap (Davison and Hinkley 1997). n_i, i = 1, …, m, values were sampled with replacement from the salinity intrusions L_x associated with each of the m flows from the model runs. The power law was fitted to the bootstrapped dataset, and $\widehat{Q^*}$ was estimated. This procedure was repeated a large number of times, and the estimated values of $\widehat{Q^*}$ were ordered. An approximate confidence interval for Q was found by selecting the values of $\widehat{Q^*}$ that satisfied $\alpha/2$ and $(1-\alpha/2)$ percentiles.

S2: Extreme-Value Theory Methodology

Consider the following representation. Let $Z_n = max\{X_1, \dots, X_n\}$, where $X_1, \dots, X_n$ are a series of independent random variables with a common distribution function, F. The $X_i's$ represent the daily river flow data, and Z_n corresponds to the maximum daily flow over n time units. If n equals 365, then Z_n represents the annual maximum. Further, if the data are blocked into m sequences of length n, generating a sequence of block maxima, $Z_{n,1}, \dots, Z_{n,m}$, then a Generalized Extreme Value (GEV) distribution can be fitted.

The GEV distribution has the form:

$$G(z) = exp\{-[1+\xi\left(\frac{z-\mu}{\sigma}\right)]^{-1/\xi}\}$$

defined on the set $\left\{z: 1+\frac{\xi(z-\mu)}{\sigma} > 0\right\}$ where $-\infty < \mu < \infty$, $-\infty < \xi < \infty$, and $\sigma > 0$. The model comprises a location parameter, μ, a scale parameter, σ, and a shape parameter, ξ. The GEV distribution corresponds to the Frèchet family for $\xi < 0,$ the Gumbel family for $\xi = 0$, and the Weibull family for $\xi > 0$. The flexibility of the GEV distribution means that the user need not select one of these three distributions *a priori*, instead allowing the data to determine the most appropriate form.

From the fitted distribution, a return level

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1-\{-log(1-p)\}^{-\xi}\right], & for\ \xi \neq 0 \\ \mu - \sigma log\{-log(1-p)\}, & for\ \xi = 0 \end{cases}$$

is found with an associated return period. The level z_p is exceeded by the annual maximum in any particular year with probability p, implying a return period of 1/p years.

The parameters of the GEV are estimated by maximum likelihood, the log-likelihood for the GEV when $\xi \neq 0$ is

$$l(\mu,\xi,\sigma) = -mlog\sigma - (1+1/\xi)\sum_{i=1}^{m} log\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-1/\xi}$$

For $1+\xi\left(\frac{z_i-\mu}{\sigma}\right) > 0, for\ i = 1, \dots, m$

And when $\xi = 0$ is

$$l(\mu,\sigma) = -mlog\sigma - \sum_{i=1}^{m}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{m} exp\left\{-\left(\frac{z_i-\mu}{\sigma}\right)\right\}$$

Maximization is obtained using standard numerical optimization algorithms.

A $(1-\alpha)$ confidence interval for the return period can be found through a parametric bootstrap approach (Davison & Hinkley 1997). Simulate a sample of n_i, i= 1, …, m values from the fitted GEV distribution with MLE parameters, $\hat{\mu}, \hat{\sigma}, \hat{\xi}$. Refit the GEV model to this boot-strapped sample and estimate the return period based on the original 1995 level, z_p. Repeat the procedure a large number of times. An approximate confidence interval for the return period is found by selecting the bootstrapped values of return period that satisfy the $\alpha/2$ and $(1-\alpha/2)$ percentiles.

S2: Supplementary Tables and Figures

Table S1: Parameter estimates, return period estimates, and 95% confidence interval for return period under a 30d drought.

Duration (d)	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	Return period (y)	95% CI for return period (y)
30	140.43	38.30	-0.451	20.28	[12.4,97.0]

Table S2: Expected, minimum, and maximum volumes of water (in millions of cubic meters) for δ = 1% and 5% risk of salinity intrusion

	Water Volume (million m^3)						
δ	Expected volume for baseline	95% CI	Expected volume for counterfactual	95% CI	Expected increase due to dredging	Min	Max
1%	1061	[1028,1098]	1437	[1369,1505]	376	271	477
5%	883	[855,907]	1179	[1133,1225]	296	226	370

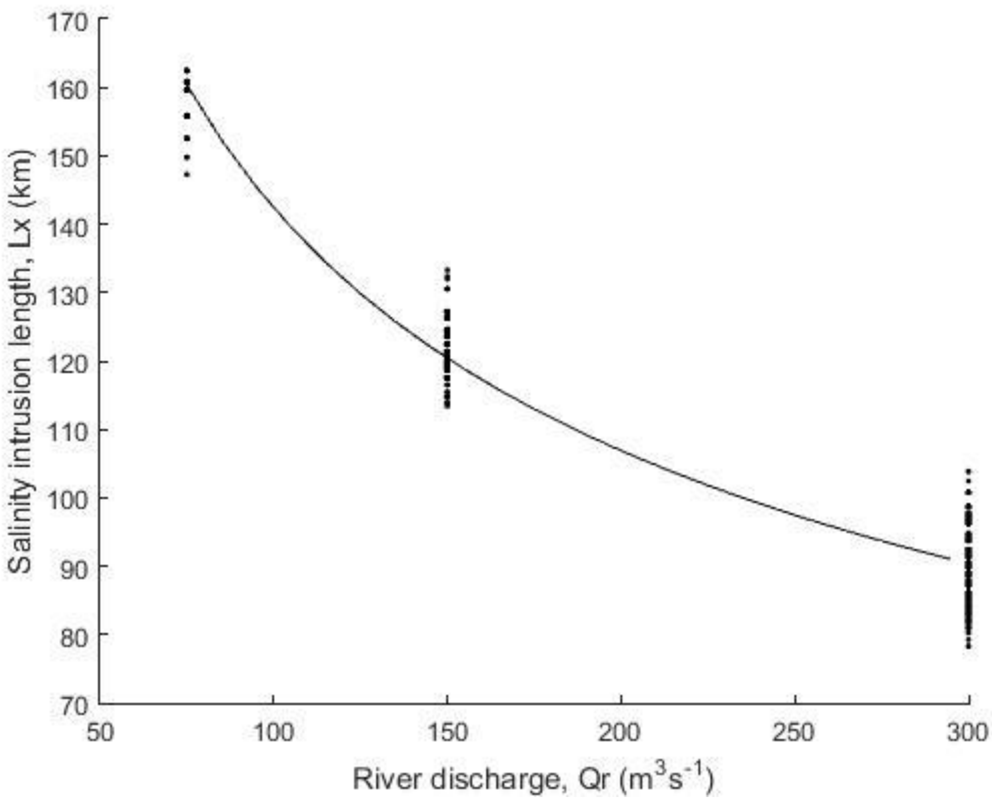
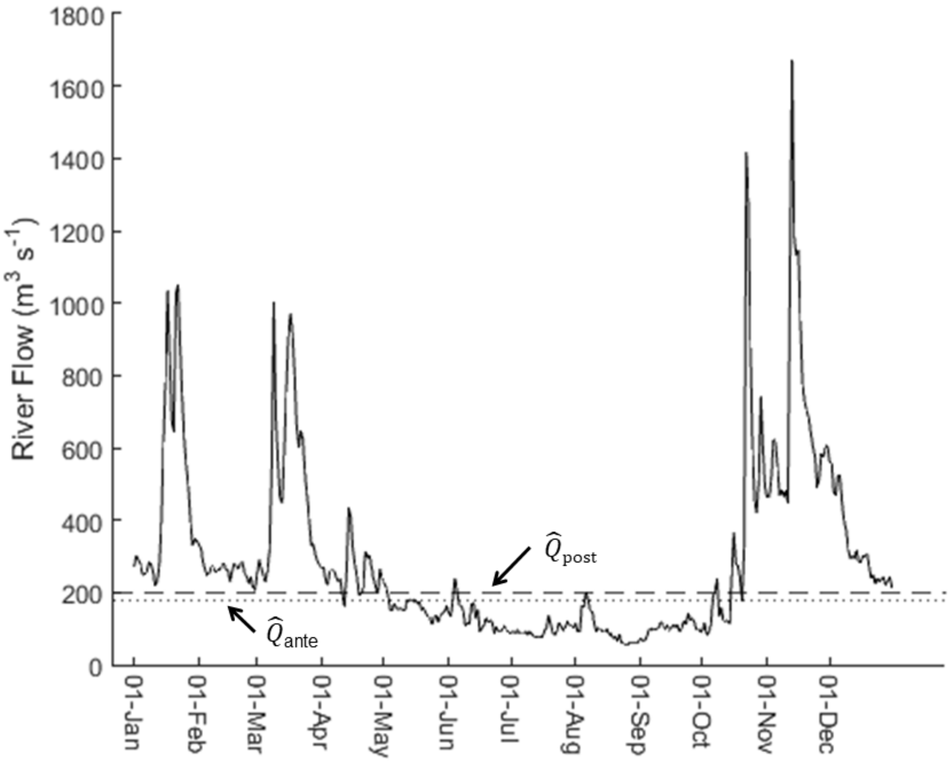


Fig. S1 The simulated position (km), upstream of Upper NY Harbor, of a salinity threshold of 0.324 PSU in relation to Hudson River discharges, shown for the river's 1995 (pre-dredged) bathymetry. The salinity intrusion length, L_x, was fitted using a power law to Hudson River discharges, Q_r, as simulated by a ROMS hydrodynamic model at flows of $75m^3s^{-1}$, $150m^3s^{-1}$, and $300m^3s^{-1}$. Observations at each level of discharge comprised tidal variations in the ROMS model output.



<u>Fig. S2</u>: Observed daily river flow (m^{3}s^{-1}) during 1995 from the Green Island gauge, immediately upstream of the Troy Lock and Dam on the Hudson River (42°45'08"N, 73°41'20"W). The difference between $\hat{Q}_{\text{post}}$ (dotted line) and $\hat{Q}_{\text{ante}}$ (dashed line) is the estimated additional water required to be released from Great Sacandaga Lake under the counterfactual a deepened channel in 1995).

<u>Fig. S3</u>: Annual minima for transformed Hudson River flow data for r = 30d, constituting historical occurrences of a 30d drought.

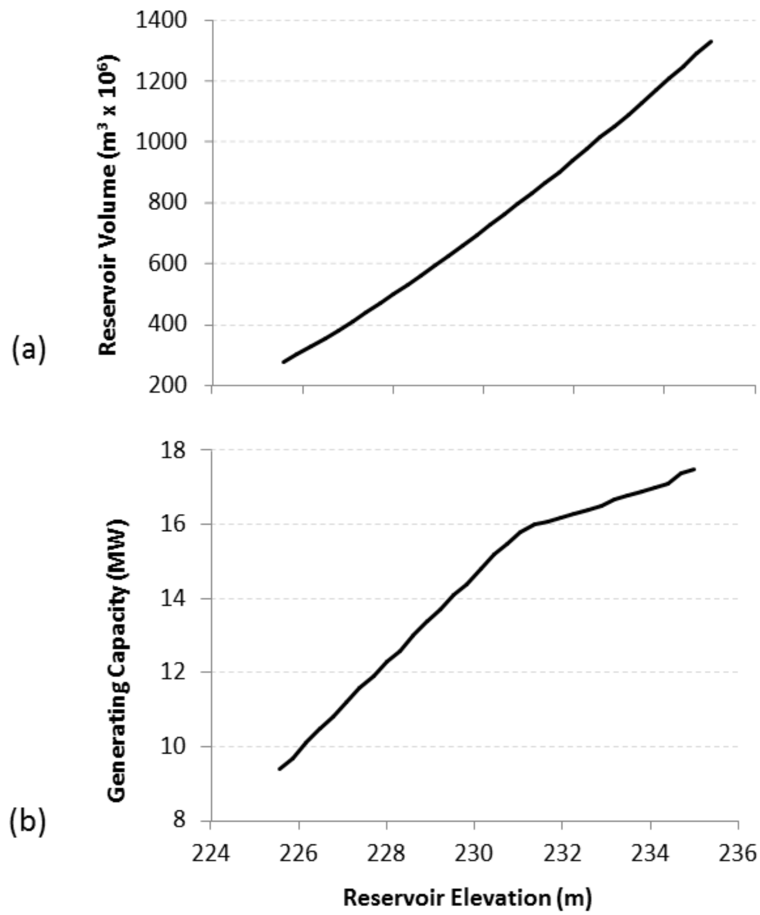
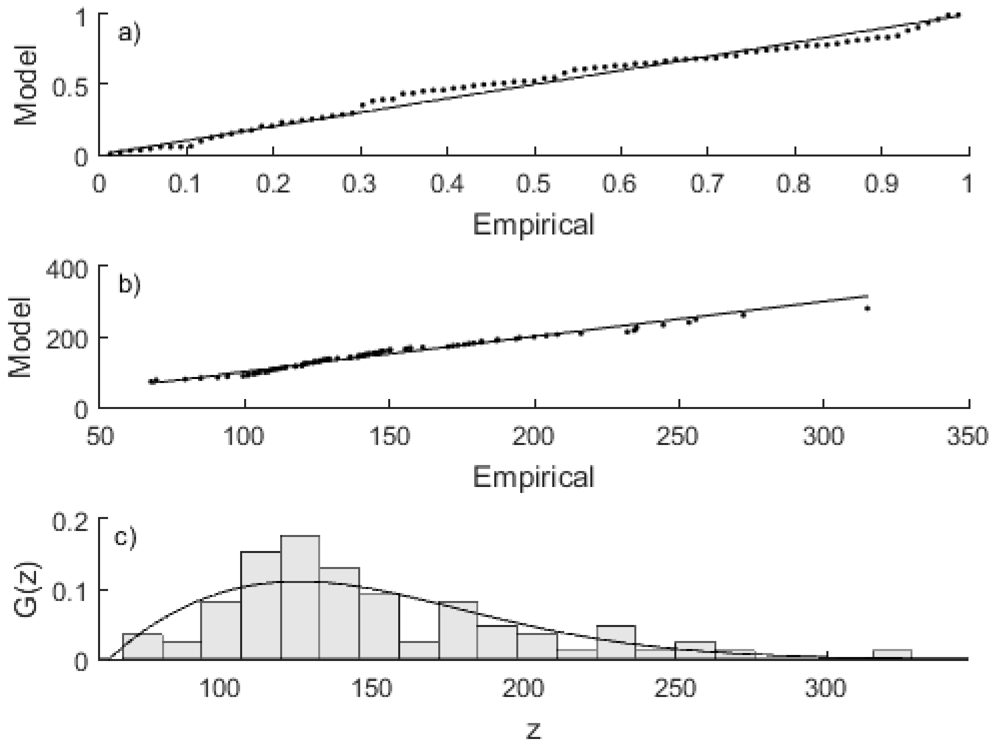


Fig. S4: (a) Great Sacandaga Lake elevation-volume relationship; (b) Great Sacandaga Lake and Conklingville Dam (E.J. West hydropower facility) generating capacity-elevation relationship. Source: ORNL (2011).



<u>Fig. S5</u>: Diagnostic (a) probability; (b) quantile; and (c) density plots for the GEV fit for a 30d drought.

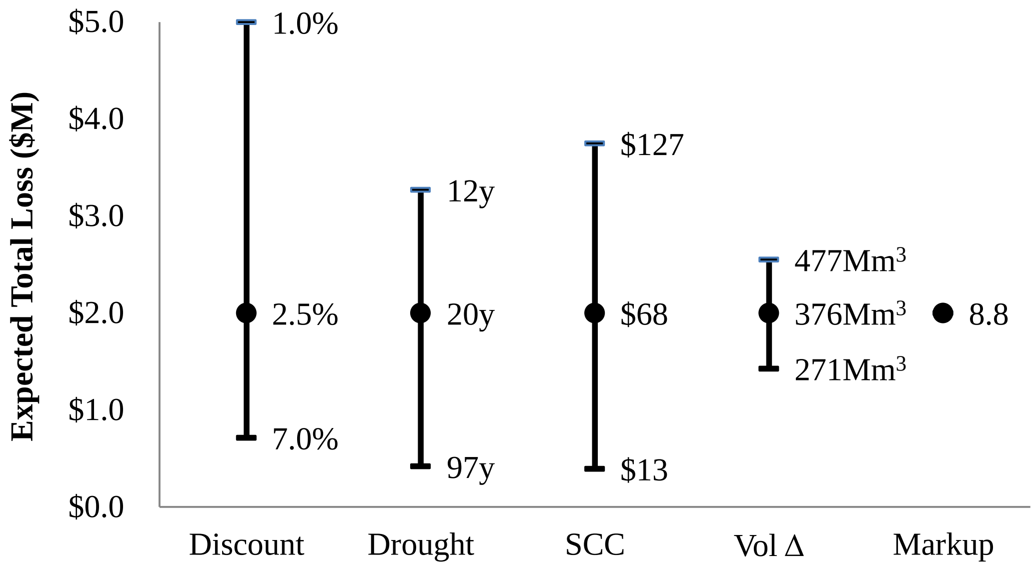


Fig. S6: Ranges of estimates of expected present value of total loss (2018 $M) under the counterfactual. Social discount rate (Discount; %); odds ratio representing expected years between droughts (Drought; y); social cost of carbon (SCC; 2018 $); expected augmented water release (Vol Δ; Mm^3); ratio of load-based marginal price of electricity in NY State Electricity Zone J (a proxy for the marginal cost of replacement electric power) to the average operating and maintenance costs of hydropower (Markup; unitless). Estimates for each parameter range were undertaken with all other parameters at their central values.