Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh

22nd Annual Progress Report
August 2001

Chapter 11:
Improving Salinity Estimates at the Martinez Boundary

Author: Eli Ateljevich

11 Improving Salinity Estimates at the Martinez Boundary

11.1 Introduction

Empirical modeling plays an important role both in planning studies and in real-time modeling. Both of these modeling applications involve simulations over periods for which there is no observed data at the ocean boundary, as there is with historical runs. Without a stream of recorded data, Martinez is not a "boundary" for salinity in a simple physical sense. Rather, the location marks a compromise between the appropriateness of a 1-dimensional approximation (which is better upstream) and independence from the mechanics of the Delta (which is better downstream). At Martinez, salinity can be modeled empirically using aggregate boundary quantities such as Net Delta Outflow (NDO), without considering hydrodynamics within the model domain[1]. Since the quantities required for the estimate are available prior to the DSM2 run, the estimates can be computed off-line ahead of time and used in DSM2 just like any other boundary condition.

Under current practice, the boundary empirical estimate is carried out as a preprocessing step to a planning run. Artificial neural networks (ANNs) estimate tidally averaged salinity as a function of the time history of Net Delta Outflow[2,3]. A tidal component is then added to the daily average estimate, constructed using *Kristof coefficients*. These are a series of 25 hourly coefficients arranged in a tidal pattern that are multiplied by daily averaged salinity to produce a scaled tidal fluctuation. The tidal fluctuation is synchronized with the Martinez design ("19-year mean") tide used for stage, so that the hydrodynamic (DSM2-HYDRO) and transport (DSM2-QUAL) parts of the simulation are realistically phased.

Several developments in Delta modeling practice have motivated the development of a better boundary salinity estimator. One is the real-time modeling program. Real-time modeling combines recent historical information with short-term projections and hypothetical operational decisions (see Chapter 12 for more information about real-time modeling). The empirical boundary estimate used for the hypothetical part of the run must agree with the earlier historical data. This is accomplished with a combination of better absolute accuracy and data assimilation. Another impetus for improving the boundary salinity model comes from planning studies. Planning studies traditionally have been conducted using a simplified 25-hour synthetic stage and salinity boundary. The computational limitations requiring this simplification have ceased to be relevant, and the Section is endeavoring to produce a more sophisticated planning simulation based on realistic tides for both stage and salinity.

[1] A reported exception are some activities in Suisun Marsh.
[2] See http://modeling.water.ca.gov/delta/studies/CalFed/models/mtz_ec_boundary/boundary.html.
[3] In practice, a salinity surrogate such as EC is used, and the choice of surrogate is currently under review.

Because of the contrived nature of a planning scenario, the requirement for boundary estimates is not absolute accuracy, but rather realism. To be realistic, the salinity estimate must be consistent with Net Delta Outflow (NDO). In addition, the tidal fluctuation in salinity must be an appropriate counterpart to the tidal flows and water levels at the boundary – as much as possible, we want to avoid patterned aberrations. Repetitive, patterned error is more damaging than white noise to a water quality simulation – the system is damped and effectively smoothes out any mistake that does not repeat itself. Similarly, stage and empirical salinity must remain correctly in phase with one another over an entire 19-year lunar nodal cycle.

This report introduces an improved method of salinity estimation at the Martinez boundary. The model was designed to accommodate both real-time and planning applications. However, the real-time data assimilation component will not be discussed. The basis of the method is G-model and the work of Denton (1993). G-model was originally designed to predict daily-averaged salinity (EC). It is modified here with one major additional assumption: to derive tidal (15-minute) salinity estimates. The proposed model is not, however, a disaggregation of a daily-averaged model. In fact, heuristic arguments are presented here to suggest that the current brand of daily-averaged salinity estimate is a poor basis for any tidally disaggregated model.

11.2 G-model Basics

G-model is a conceptual-empirical model of salinity transport along the main stem of the Sacramento River. Details of the model may be found in Denton (1993) and in documents supplied to DWR and Delta modeling community by Dr. Gregory Gartrell and Dr. Richard Denton of Contra Costa Water District. The parametric form of the G-model is based on a successive steady-state argument. The development begins with the steady-state solution to a simple, 1-dimensional advective-diffusion problem in a steady current in an infinitely long channel with a downstream ocean and an upstream river boundary condition. The solution may be written as an exponential longitudinal salinity profile as follows (Denton, 1993):

$$s(x) = s_b + (s_o - s_b)\exp(-\tilde{\alpha} q_{\text{steady}} x) \qquad \text{[Eqn. 0-1]}$$

where
$s(x)$ is salinity or a surrogate over longitudinal distance (x),
s_b is the ambient salinity that exists in upstream fresh water,
s_o is the ocean salinity at the downstream boundary,
q_{steady} is the steady, uniform flow rate, and
$\tilde{\alpha}$ is a dispersion parameter.

The most rudimentary successive steady state approximation would be to substitute time-varying $q_{\text{out}}(t)$ for q_{steady} in Equation 0-1. This substitution would be justified if the differential system was fast compared to the rate of change of $q_{\text{out}}(t)$. This is not the case in the Delta during dynamic periods, so the authors substituted *antecedent outflow* $g(t)$ instead, where:

$$\dot{g}(t) = \frac{g(q_{\text{out}} - g)}{\beta} \qquad \text{[Eqn. 0-2]}$$

where
$g(t)$ is antecedent outflow,
$q_{\text{out}}(t)$ is net delta outflow, and
β is a parameter that determines how slowly the system reacts to changes in delta outflow.

Antecedent outflow converges to q_{steady} under steady conditions, and changes much more slowly than $q_{\text{out}}(t)$ does under dynamic circumstances. Empirically, $g(t)$ produces good transient results, although the author is not aware of any formal connection between the solution of the original time-varying transport equations and the solution of the steady-state form with the antecedent flow substitution.

After antecedent outflow is substituted in Equation 0-1, and considering a fixed point x, the standard form of G-model is as follows:

$$s(t) = s_b + (s_o - s_b)\exp\left(-\alpha g^n(t)\right) \qquad \text{[Eqn. 0-3]}$$

where
$s(t)$ is salinity (or a surrogate such as EC),
s_o and s_b respectively represent ocean and river salinity (or upper and lower bounds),
α is the result of consolidating upstream distance into the dispersion parameter, and
n is an additional empirical shape parameter close to unity, used to compensate for imperfections in the antecedent outflow and exponential profile assumptions.

One difficulty with the use of *antecedent outflow* is finding an initial condition for Equation 0-2, the differential relationship between $g(t)$ and $q_{\text{out}}(t)$. Because $g(t)$ is a contrived quantity, it does not have a natural initial condition. To get around this problem, the original G-model tools use $g(0) = q_{\text{out}}(0)$. This analogy between antecedent outflow and actual outflow is a good approximation near the end of long periods of steady Net Delta Outflow. It is particularly useful in planning models, where the circumstances are synthetic, there is no field salinity to be matched, and the duration of the simulation is so long that salinity initial conditions influence only a small fraction of the model period.

In historical studies where an initial salinity is available and short-term accuracy is desired, we can use Equation 0-3 to convert $s(0)$ directly to $g(0)$ instead of using a flow. Equation 0-3 represents a functional, rather than differential, relationship between salinity and $g(t)$, so we can translate $s(0)$ into an initial condition without any assumptions of the equilibrium of the system. This choice of initial conditions dramatically improves the short-term performance of

G-model. In real-time applications, recent data are assimilated into the empirical model by means of a recursive filter, which more or less trivializes the initial condition.

In the conversion to a tidal model, a G-like model is integrated over steps smaller than one day, and there are a few special steps required to do this. First, Net Delta Outflow is interpolated to smaller time steps using a spline, although its interpretation is still as a daily average (just with a moving window over smaller time steps). The differential model for antecedent outflow is integrated with a standard second-order trapezoidal method. Note that for standard applications of G-model to estimate daily-averaged salinity, there is no reason to interpolate NDO or integrate the standard G-model at time steps smaller than a day.

Finally, the original authors of G-model have proposed corrections to Net Delta Outflow to account for both Delta filling and draining and better island consumptive use estimates. Only the Net Delta Outflow correction was used in the work presented here. This correction is of the form:

$$q_{\text{corrected}}(t) = q_{\text{ndo}}(t) - A\Delta z(t) \qquad \text{[Eqn. 0-4]}$$

where A is a coefficient that represents the storage area filled by an incremental change in (tidally filtered) water surface height $\Delta z(t)$ or an estimate of $\Delta z(t)$ (see Section 11.6). The value of A currently being used is equivalent to $A = 40,000\ ft^2$ for daily increments of water surface height.

11.3 Tidal Model

The tidal extension of G-model is based on the following simplification of tidal dynamics:

> *Tidally-varying salinity is the result of a uniform, harmonic advection acting on the exponential salinity profile from G-model.*

This advection is illustrated in Figure 11-1. Between time t_0 and t_1, the salinity profile is shifted upstream (to the right) by a flood tide. Later, between t_1 and t_2, the salinity profile moves back downstream with the ebb tide. To emphasize the role of advection, the profile has been drawn the same each time step. Figure 11-2 shows the full model, with independent changes in the G-model and advection components: the profile changes shape as it oscillates before the observer.

The interaction of the tide and the concentration profile is significant from the point of view of an observer. Imagine that t_0 represents a "centered" (or median) period during the tidal day and that the profiles t_1 and t_2 represent the tidal excursion in either direction. Due to the convex shape of the concentration profile, salinity at the point of the observer will achieve a greater extreme during flood than at ebb (dots have been placed on the plot to show the peaks in both directions). This lopsided tidal effect favoring the flood is regularly observed at Martinez and nearby stations. It is shown in Figure 11-3 during a period in 1998 with low-moderate salinity. In periods of high salinity, the salinity gradient is more linear over the length of the tidal excursion, and the phenomenon is less important.

It was asserted earlier that the temporal mean salinity is not an appropriate basis for a tidal model. The reason for this is tied to the lopsided tidal fluctuation just described. The tide does not attain its mean salinity when the displacement of the profile is at its tidally "centered" position. Instead, the mean salinity is biased above the centered value, and is attained sometime when the profile is upstream of the centered position. Just how far upstream depends on the curvature of the concentration profile, which, in turn, is dependent on flow regime and difficult to correct. Since the mean salinity and mean displacement do not coincide, schemes that begin with a mean estimate of salinity and then add a tidal fluctuation centered on this estimate tend to show common shortcomings. They fail to realistically capture the lopsided shape of the tidal fluctuation, and they grow increasingly inaccurate as concentrations decrease.

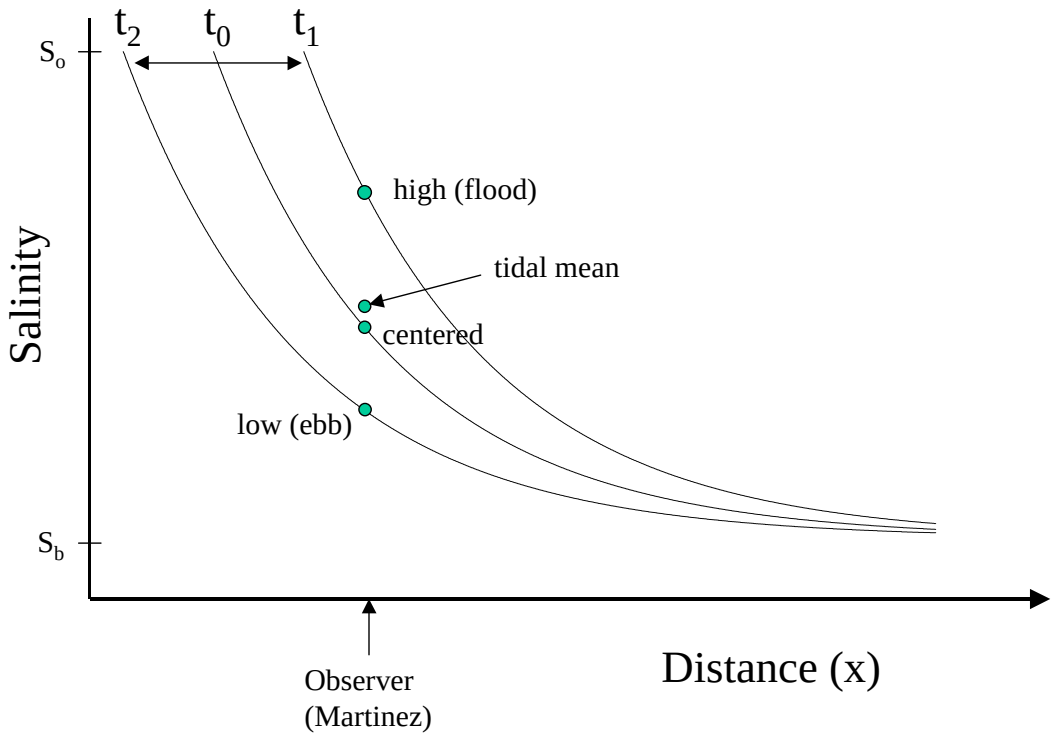

Figure 11-1: Advection of a Concentration Profile Back and Forth of an Observer.

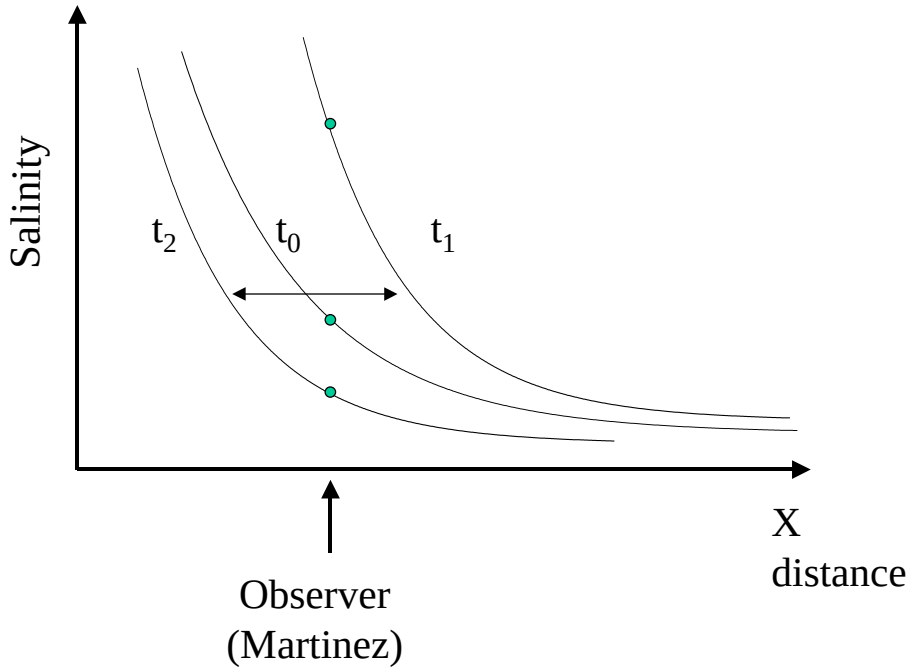

Figure 11-2: Advection of a Changing Concentration Profile.

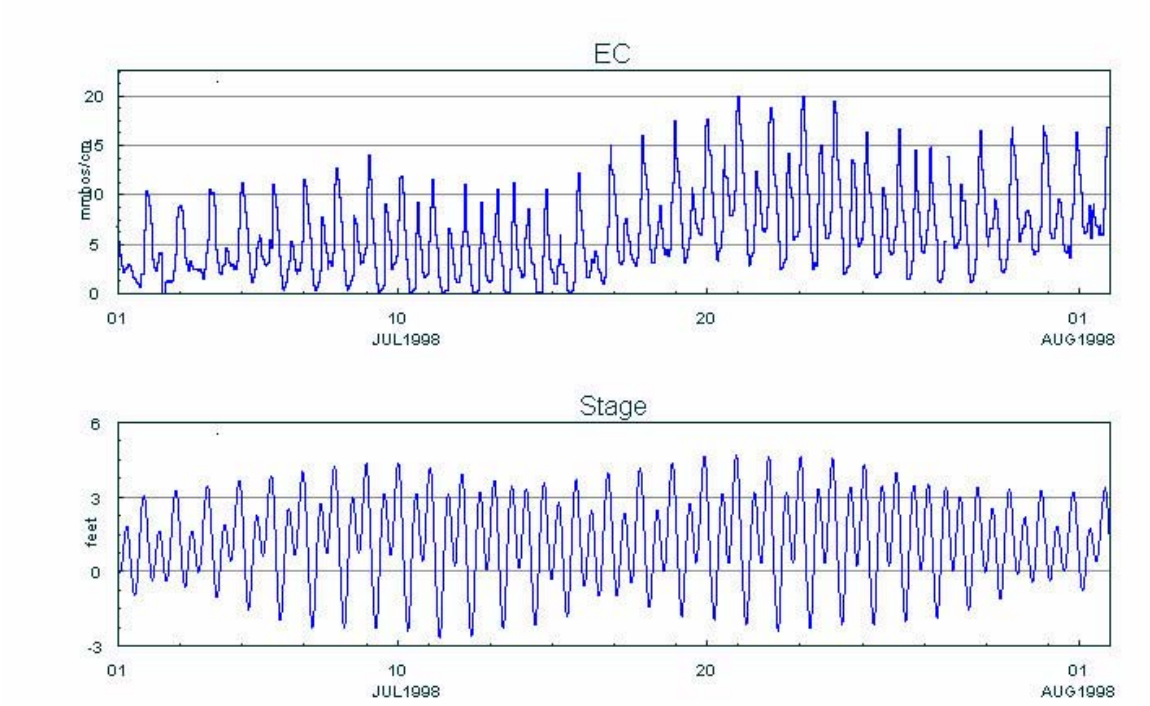

Figure 11-3: Comparing the Shape of the Stage and EC Tidal Fluctuation.

11.4 Mathematical Formulation

The conceptual model thus described can be written symbolically as follows:

$$s(t) = s_b + (s_o - s_b)\exp\left(-\tilde{\alpha} g^n(t) x(t)\right) \qquad \text{[Eqn. 0-5]}$$

where the terms are as before and
$s(t)$ is (tidally varying) salinity,
$\tilde{\alpha}$ is the decay parameter of Equation 0-1 before distance was bundled into it for G-model (the longitudinal salinity profile is not used explicitly in G-model), and
$x(t)$ is a harmonic position, reflecting displacement of the profile.

Rearranging Equation 0-5, we obtain:

$$\ln\left(\frac{s(t) - s_b}{s_o - s_b}\right) = -\tilde{\alpha} g^n(t) x(t) \qquad \text{[Eqn. 0-6]}$$

We can further write $x(t) = x_0 + x'(t)$ as a centered position x_0 and harmonic perturbation $x'(t)$, and combine the scalar product $-\tilde{\alpha} x_0$ into a coefficient β_1:

$$\ln\left(\frac{s(t) - s_b}{s_o - s_b}\right) = \beta_1 g^n(t) + x'(t) g^n(t) \qquad \text{[Eqn. 0-7]}$$

Comparing the model to the original G-model, it is clear that in the centered position $x'(t) = x_0$, the tidal model and G-model share the same parametric form, with β_1 analogous to the G-model parameter $-\alpha$. The only difference is that: β_1 will be chosen to fit the tidal model (i.e., it represents salinity when the concentration profile is tidally centered), whereas α is chosen to estimate daily-averaged values.

Next, we must model $x'(t)$, the component that reflects tidal displacement of the concentration profile. Because the local flow processes are dominated by linear terms, we will assume that $x'(t)$ is composed mostly of the same harmonic constituents as the stage. However, we opted not to fit an independent harmonic tide for the tidal displacement because tidal constituents would be resolved much less precisely for tidal displacement than for stage. The stage signal is directly observable, relatively noise free, and more perfectly harmonic in character. In contrast, tidal displacement is not directly observable – it is an intermediate term in our nonlinear model whose ultimate output (EC or salinity) is particularly noisy.

a.

b.

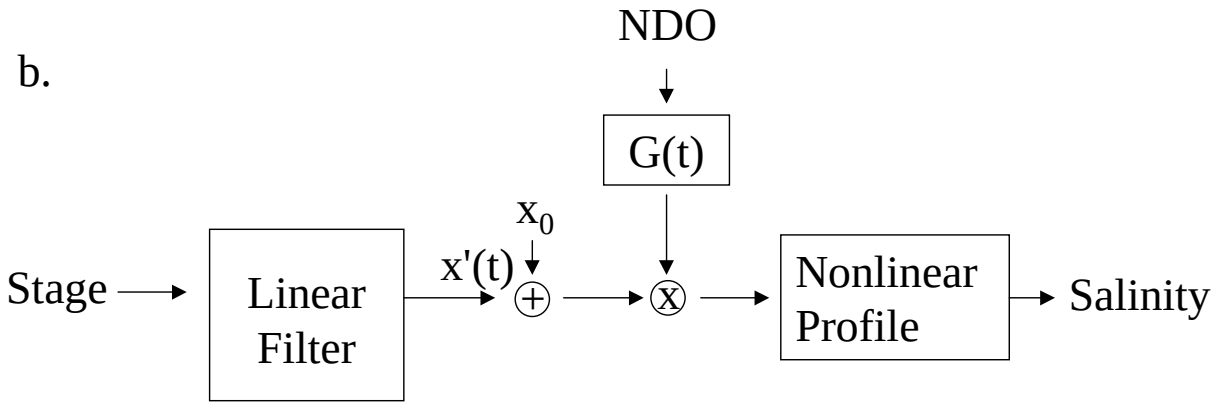

Figure 11-4: Models of Tidal Displacement. a) The Displacement Model as a Linear Filtration of Stage. b) The Displacement Model as Embedded in the Full Salinity Model.

A better way to control the relationship between stage and displacement of the salinity profile is to model $x'(t)$ as a linear filtration of stage. This solution allows stage to be estimated at whatever level of refinement is possible, and for the details to be inherited by the salinity model at whatever level is appropriate. The description here is based on the rudimentary system depicted in Figure 11-4a, a linear filter which takes stage as input and tidal displacement as output. We must bear in mind, however, that this an abstraction – the filter is embedded in a larger nonlinear model (Figure 11-4b), and since our only observations are salinity or EC, we will be grappling with the full non-linear model for estimation purposes.

Linear filters (such as the one represented schematically in Figure 11-4a) are very general and ubiquitous analysis tools for describing input-output systems. A familiar example in the water resources field is the "hydrograph", which relates rainfall to runoff using a convolution filter. In the present case, we will also use a convolution filter modeling displacement on lagged values of stage:

$$x'(t) = \sum_{k=0}^{n_k-1} a_k z\left(t + k_0\Delta t - k\Delta t\right) \qquad \text{[Eqn. 0-8]}$$

where

a_k are the filter coefficients

$z(t)$ is a band-limited estimate of the tide,

n_k is the length of the convolution kernel (number of lagged input values used),

Δt is a spacing between lagged values (for instance we may take every third hourly value), and

k_0 is an offset of the filter allowing a shift in the first hour used.

Although the convolution sum of lagged values is a bonefide and familiar representation of a linear system, it is not particularly useful for discussing the system's effect on tides. Instead, we are interested in the frequency response, or "transfer function," of the system – the gain or phase shift at various frequencies in the spectrum. For instance, in converting stage to tidal displacement, we might want to know whether the diurnal species are damped or shifted more or less compared to the semidiurnal species. These, incidentally, are the only types of effects that a linear filter can have – it cannot give rise to new frequencies that are not present in the input. For more on the spectral characteristics of linear filters, the reader is referred to any standard linear systems texts; Oppenheim and Willsky (1996) is recommended.

In the present case, the physical processes are consistent with a fairly simple transfer function. The linear filter used here is one that varies very slowly with frequency – enough that waves from different species (long waves, diurnal, semidiurnal, etc.) can be treated differently, but not so much to allow great variation in the treatment of waves within a species (e.g. M2 and S2). The simplicity in the transfer function will be achieved by choosing the convolution filter from the same family proposed by Munk and Cartwright (1966).

Munk and Cartwright (1966) was a landmark paper in tidal literature, which treated the oceanic tide as a filtration of gravitational tidal potential[4]. The authors posited a "credo of smoothness" for the transfer function of their model, much like the one suggested in the previous paragraph for the relationship between stage and the position of the salinity profile. They showed that a rudimentary family of smooth transfer functions could be generated using convolution filters with evenly (and fairly widely) spaced weights.

The filter proposed here is based on a "credo of even more smoothness". Munk and Cartwright imposed smoothness within tidal bands, but used separate filters for long wave, diurnal and semidiurnal species. In the present model, smoothness is imposed across the entire important part of the spectrum, from long wave to semidiurnal. Because this broadness of the transfer function is associated with smaller time steps in the time domain, the filter spacing $\Delta t = 3$ hours was selected instead of the $\Delta t = 48$ hours adopted by Munk and Cartwright for their application.[5] The actual "wiggliness" of the frequency response is determined by the number of terms the convolution kernel n_k; longer filters admit more detail. In the present case $n_k = 7$ equally spaced

[4] Gravitational potential is a fairly complicated function of the spherical geometry between (mainly) the sun, moon, and earth. It is a theoretical value used as the input or forcing of the filter. Oceanic tide is the output.

[5] See the original paper for the arguments relating filter spacing to the domain of the of the transfer function.

components was thought to be the most detail that could be justified based on the data and the postulated simplicity of the system.

Finally, the input series $z(t)$ must be band limited. The filter is not designed for high frequencies (terdiurnal and higher species), but does have a frequency response at these frequencies and will exhibit unwanted behavior if the input includes them. This undesirable behavior can be avoided by excluding high frequencies from the stage estimate (z). If an observed stage signal is used as input – which is not particularly recommended – it should be low-pass filtered with a sharp cutoff frequency under 3 cycles/day. If the semidiurnal limitation suggested here proves too strict, the filter domain can be extended to terdiurnal species by taking $\Delta t = 2$ days.

With the filter in place, the full salinity model can now be written:

$$\ln\left(\frac{s(t)-s_b}{s_o-s_b}\right) = \beta_1 g^n(t) + g(t)\sum_{k=0}^{n_k-1} a_k z(t+k_0\Delta t - k\Delta t) \qquad \text{[Eqn. 0-9]}$$

Notice that the shaping exponent n has been dropped from the tidal term. This allows for slight deviation from a perfectly exponential profile, and was found to yield a better fit over a variety of flow regimes.

11.5 Estimation

The model (Equation 0-9) has thus far been described deterministically and mechanically. In order to estimate the parameters of the model from field data, however, we must consider the uncertainty due to model imperfections and generally noisy data. A naïve least squares approach is not appropriate for this fit, and the results it gives are poor.

No standard statistical model applies to the full tidal model. However, if we assume that we know $g(t)$, Equation 0-9 can be estimated conditional on $g(t)$ from a wide family of appropriate models. Therefore, the approach used to estimate the tidal model iterated informally between the following two steps:

- An outer step to fit the G-model parameter β from Equation 0-2 and the Net Delta Outflow correction A from Equation 0-4, and the parameter.

- An inner iteration for all the remaining parameters relating salinity or EC to $g(t)$ using a statistical model and Generalized Estimating Equations. This fitting procedure is statistically consistent conditional on a single set of G-model parameters from the previous step (these guarantees of consistency do not necessarily apply to the iterative process when the outer step is included).

There is little of a technical nature to describe about the outer procedure. The G-model fit in the outer step was mostly qualitative – although some refinement was carried out based on squared

residual error from the inner iteration. The main qualitative criterion for adjusting β was to match the response-time salinity data to long-period (three week or more) variations in outflow observed in the field data. The main criterion for assessing the Delta outflow correction was correctness of the shape of the tidal envelope over two-week cycles. These two goals are somewhat conflicting, a point that will be taken up again during the discussion of results. The parameter s_b was fixed at a value of 200 µmhos/cm based on physical arguments, and the parameter s_o was nominally set at 32,000 µmhos/cm (it is effectively corrected in the next step, anyway). The shaping factor n was taken to be 0.7.

The inner procedure is based on a more formal statistical model. Assume that an adequate guess for the G-model parameters is available and $g(t)$ has been pre-calculated. We will introduce the random component into our formula by means of expectations:

$$\ln\left[E\frac{s(t)-s_b}{s_o-s_b}\right] = \beta_0 + \beta_1 g^n(t) + g(t)\sum_{k=0}^{n_k-1} a_k z(t+k_0\Delta t - k\Delta t) \qquad \text{[Eqn. 0-10]}$$

In Equation 0-10, the expected value is internal to the logarithm. By applying the expectation to a term linear in $s(t)$, we avoid bias due to the log transformation, and can cast the expression conveniently as a Generalized Linear Model (GLM). The difference $s_o - s_b$ can also be removed successively from both the expectation and the log, in which case it just becomes an additive scalar that can be absorbed or corrected by the intercept term β_0. This is the reason that the initial choice of s_o was described before as "nominal".

Given that the products of $g(t)$ with the lags of $z(t)$ are pre-calculated, Equation 0-10 is in the form of a GLM, with log link function, response variable $y = \dfrac{s(t)-s_b}{s_o-s_b}$ or $y = s(t)-s_b$ as convenient and linear parameters $\beta_0, \beta_1, a_k \ (k = 0 \ldots n_k)$. The log function is a natural link of the Gamma family of distributions, which have skew, sign, and mean-variance relationship that are appropriate for the physical circumstances. The choice of a Gamma distribution is more general and robust than, say, a Gaussian distribution. To add further robustness, the Gamma family for the GLM was converted to a robust family using the Splus function for this purpose.

The fit can be improved further if we bear in mind that the model errors are time-correlated. We can account for this with minimal variation in approach by using Generalized Estimating Equations (Liang and Zeger 1986), which are an extension of GLM that allows a correction for patterned, correlated error. Generalized Estimating Equations (GEE) allow a "working covariance" model for the covariance, and one popular choice is the AR1 time-correlated model. GEE gives consistent coefficients for the linear parameters even if the working covariance model is only approximate. The AR1 model is, of course, only approximate, but it is better than the naïve assumption of white noise. In fact, the YAGS (Yet Another GEE Solver) software performing the fit estimated the residual autocorrelation parameter as high as 0.909, which indicates that the errors are far from "white".

Table 11-1 lists the calculated coefficients for the model, fit to data spanning from August 20, 1991 to September 5, 1992. Examples in the tide literature using similar filter design (the "response" method) indicate that the coefficients vary considerably more in the time domain than in the frequency domain, so the convolution weights should not be over-interpreted. The robust standard error estimates are constructed to account for model failures: the "naïve" standard error estimates, which do not take these into account, tend to be just less than half the robust values. The z-score statistic is based on the asymptotic (normal) distribution of the coefficient estimates and can be interpreted in pretty much the same way a t-statistic is in regression. Several of the coefficients (β_0, a_1, and a_2) are not significant according to this statistic, but they were retained in order to maintain the evenly spaced pattern of the filter and interpretation of the model.

Table 11-1: Coefficient Estimates, Standard Error, and Z-Scores (GEE).

Component	*Coefficient*	*Robust std. Error*	*Robust z-score*
β_0	1.37E-02	3.58E-02	0.3821742
β_1	-6.43E-05	6.74E-06	-9.5344983
a_0	1.59E-04	1.91E-05	8.3407955
a_1	-1.28E-05	1.50E-05	-0.8485398
a_2	6.96E-06	2.40E-05	0.2898766
a_3	4.83E-05	1.34E-05	3.6016491
a_4	-7.67E-05	1.63E-05	-4.7145223
a_5	6.93E-05	8.77E-06	7.8985614
a_6	-3.85E-05	1.42E-05	-2.7054738

It is perhaps more relevant (and stable) to view the transfer function in the frequency domain when the filter components $a_k \left(k = 0 \ldots n_k\right)$ are interpreted literally as a linear filter on $z\left(t\right)$. Figure 11-5 shows the transfer function under the GEE/YAGS estimate, including both gain (magnitude amplification) and phase shift.[6] For comparison, the corresponding fit for GLM is shown in Figure 11-6. The GEE/YAGS estimate spans a greater range, but the differences are modest near the important frequencies of 1c/d and 2c/d. The main trend in either case is greater attenuation of higher frequencies than of lower frequencies. There is an apparent phase shift induced by the filter that varies linearly with frequency. This, however, is just the artifact of using a 3-hour initial offset ($k_0 = -1$) for the filter. Three hours comprise a greater proportion of a shorter (say, 12-hour) period than it does of a longer (say, 25-hour) period constituent. So, in terms of degrees or cycles, the offset causes a greater shift in high frequency, short period waves. When the transfer function is adjusted to remove this offset (not pictured), the phase response is otherwise quite flat.

[6] Magnitudes are small because the filter output is normally multiplied $g\left(t\right)$ to produce salinity estimates and this multiplication is not represented here.

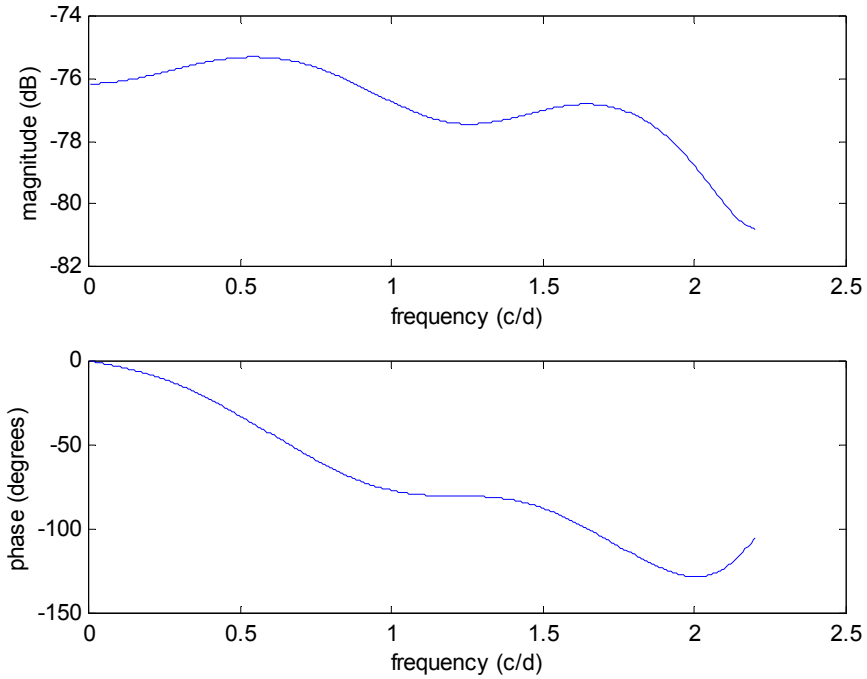

Figure 11-5: Transfer Function Characteristics of the Filter Relating Concentration Profile Displacement x'(t) to Water Surface Height. Estimate Using GEE with Robust Gamma Distribution and Log Link.

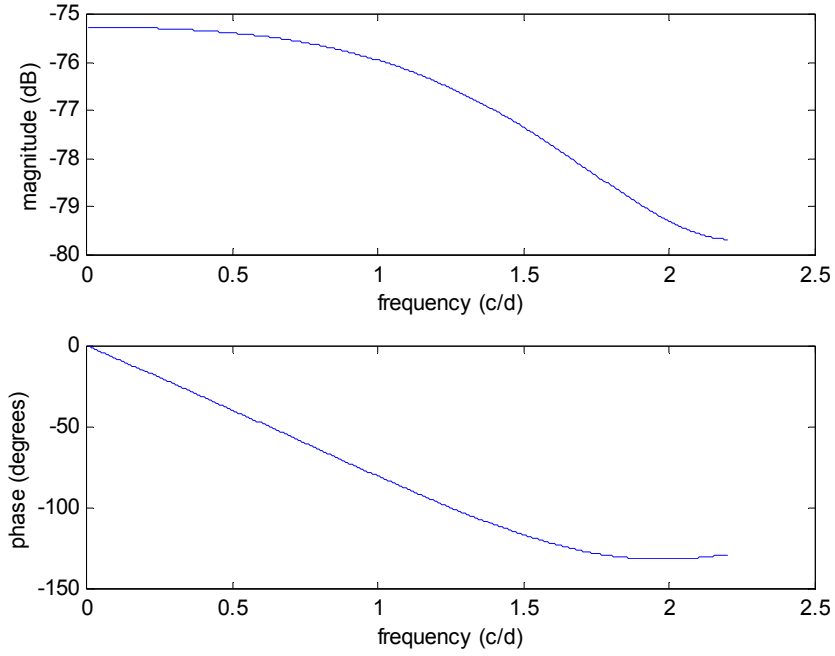

Figure 11-6: Transfer Function Characteristics of the Filter Relating Concentration Profile Displacement x'(t) to Water Surface Height. Estimate Using GLM with Robust Gamma Distribution and Log Link.

11.6 Validation and Discussion

The coefficients were estimated from the period August 20, 1991 to September 5, 1992. The fit model was then applied over a validation period spanning the two calendar years 1993 to 1994. The results and residuals are pictured in Figure 11-7. Little detail is apparent from this plot, but we can see the ability of the model to pick up the main trend and the tidal envelope. Root mean squared error over the validation period is 2,828 µmhos/cm. When rms error is applied to the daily average salinity, the result is 1,863 µmhos/cm. This is lower than cross-validation rms error of the best daily-averaged (i.e., standard) G-model that the author was able to fit (2020 µmhos/cm), illustrating that it is advantageous to account for mechanics at finer time scales even if the output is to be time-averaged.

To illustrate the tidal component of the model, a shorter subset of the test data during a period with fairly constant average salinity is pictured in Figure 11-8. Several features are worth noting. First, under moderate dynamics, the residuals show little diurnal or semidiurnal fluctuation, indicating that the tidal model, even when it fails, does not produce patterned error that would bias a planning run. Around May 1 each year, for instance, there is a period when salinity is overestimated for several days. Despite this error, the tidal fluctuation is reasonable.

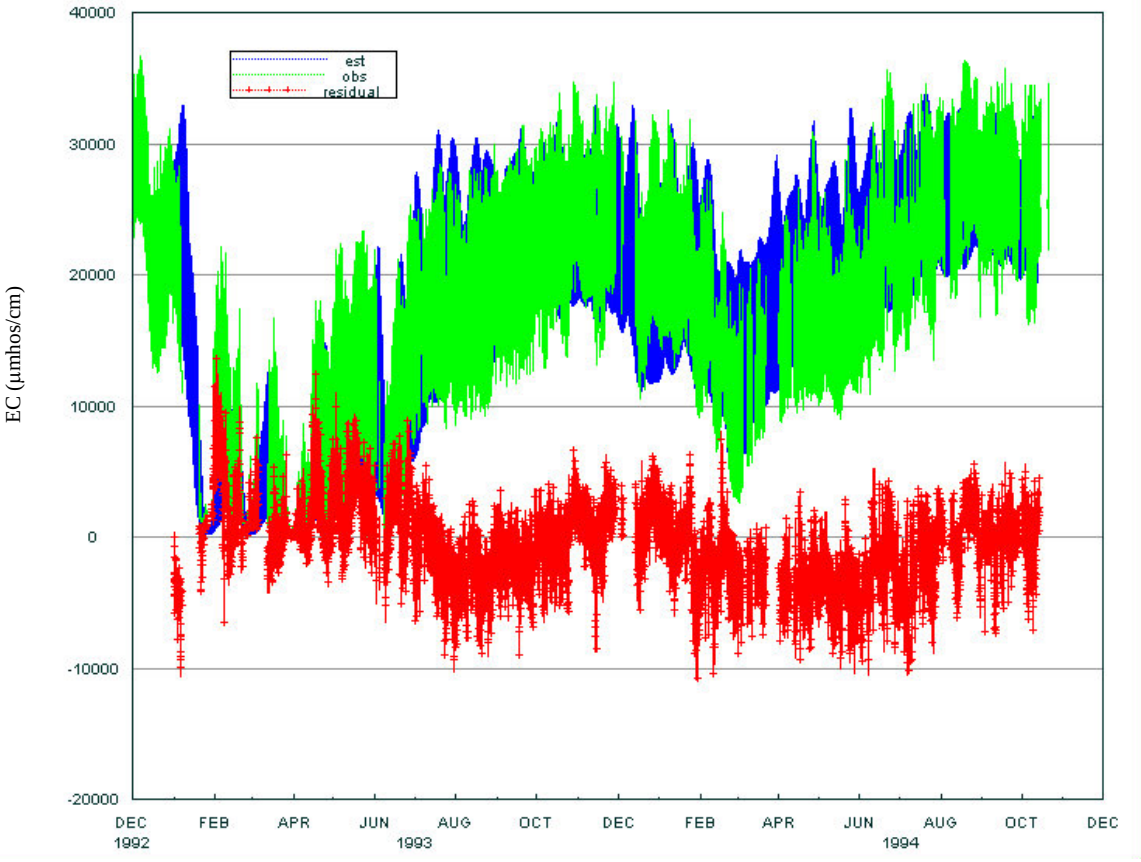

Figure 11-7: Validation Estimates for the Years 1993-1994

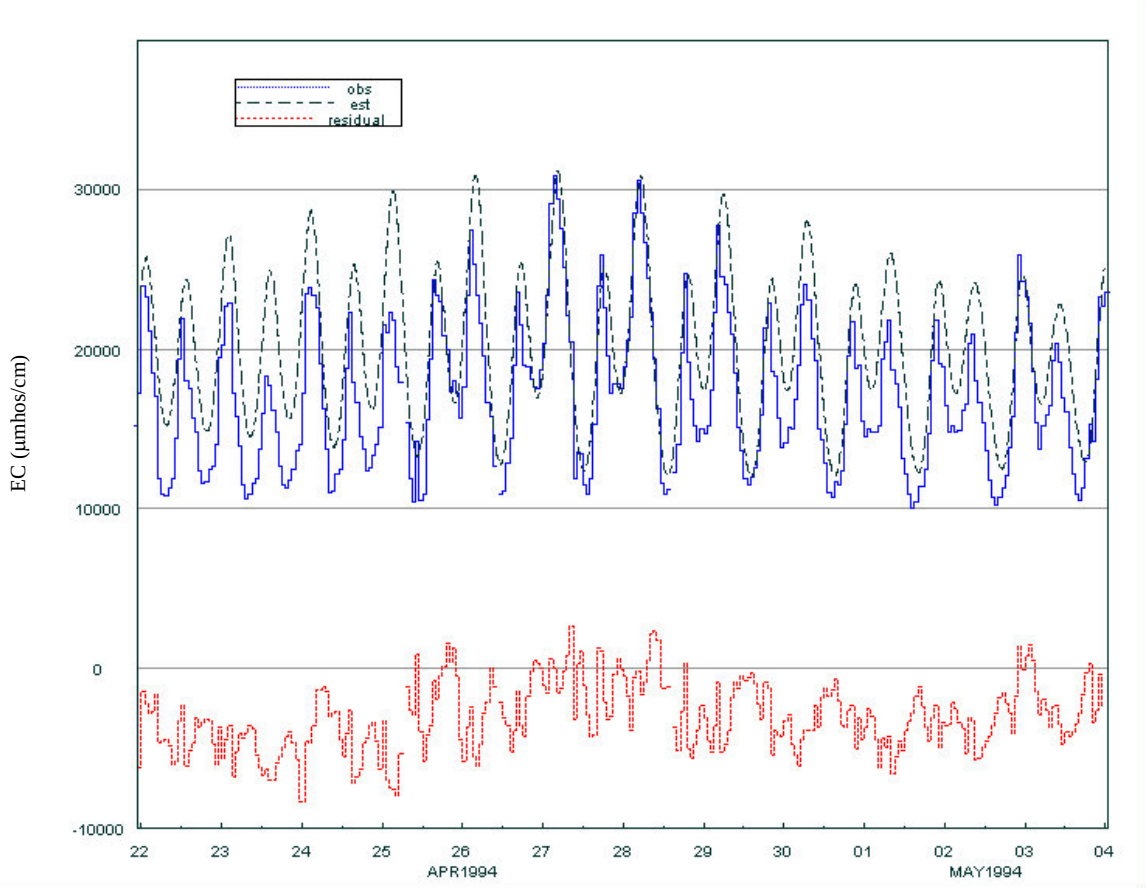

Figure 11-8: Model Performance Over a Two-Week Period.

In contrast, the error on the left hand side of Figure 11-8 (and to some extent the far right) represents the most vexing problem in the analysis. To see the problem, note the shape of the tidal envelope over a 14-day cycle. The estimate, mimicking stage, has a fairly symmetric envelope top and bottom. During the neap tide at the left side of the plot, the envelope is curved toward the center as in the center of an hourglass. Field salinity, on the other hand, does not exhibit a symmetric envelope. The upper envelope is curved as expected, but the lower envelope is flat over the entire spring-neap cycle at about 10,000 µmhos/cm – the middle part of the hourglass has been pressed down. This phenomenon is common whenever salinity is high enough to expect a tidal shape.

Analysis by Denton attributes these anomalies in salinity over the spring-neap cycle to the draining of the Delta. There is a non-linear interaction between tidal constituents that tends to produce high water levels and salinity when tidal energy is high, and low water levels and salinity when tidal energy is low. The resulting 14-day fluctuation in (tidally averaged) water levels is known as the "filling and draining" of the Delta. The storage correction to Net Delta Outflow is intended to compensate for this cycle – evacuated storage is considered to augment Net Delta Outflow. Denton reported G-model results in which this correction dramatically improves the performance of the standard G-model. A similar correction was used here, but the correction has proven to be problematic both practically and conceptually.

Astronomical stage estimates contain a classic spring-neap cycle, but this is just a linear superposition of waves – its tidal average is zero. The planning tides described in Chapter 12 augment astronomical tides with low-frequency components modeled on San Francisco. These low-frequency component estimates also provide good surrogates for use in the filling and draining correction (Equation 0-4). See Chapter 12 for a more complete discussion of the low-frequency Martinez and San Francisco tides.

The limitation of this method is that historical San Francisco stage data must be available. This is true over most of the century (planning runs), but not for the future (real-time runs). When San Francisco estimates are not available, tidal energy can be used to predict filling and draining. Figure 11-9 shows the correspondence between filling and draining (low-passed observed tide) and predicted tide energy (low-passed squared astronomical tide). Periods of very high tidal energy often coincide with high values of stage, but the relationship is mild and often overwhelmed by other factors during less powerful spring-neap cycles. Occasionally, the filling fails to occur even during times of high tidal energy. An example of this is shown in the plot around July 12, 1993. Even during periods where the relationship holds well qualitatively, the fluctuations are not an adequate basis for a numerical prediction. [7]

Barometric pressure is usually assumed to be an important contributing physical factor determining water levels, but it is somewhat collinear with tidal energy and only explains a tiny fraction (about 6%) of the remaining squared error in tidal height when added to a model based on tidal energy. Perhaps this is fortunate, because pressure is not available for prediction anyway.

In addition to these practical difficulties, there are two conceptual problems with the filling and draining correction. The first is that it gives rise to two conflicting time scales. For the draining correction to have any effect at all on salinity, the 14-day fluctuations must be allowed to register rather strongly in $g(t)$. This, in turn, means that the parameter β must be small enough to allow response at this time scale. The problem is that a small value of β is not particularly consistent with changes in Net Delta Outflow at longer time scales. Even though great care was taken to avoid compromising accuracy at these longer time scales in order to accommodate filling and draining, a slight tendency for the estimate to respond too quickly to changes in flow regime is still apparent in Figure 11-7.

[7] Richard Oltmann and Michael Simpson of the USGS have also studied this phenomenon. Their plots and text at http://sfbay.wr.usgs.gov/access/delta/tidalflow/tidalcycle.html include both stage and flow observations at Jersey Point. They concluded that at this location filling and draining coincide well with the spring-neap cycle; however, their plots also include irregularities and exceptions of the types described here.

The second conceptual problem is the question of causality: is it the draining of the Delta that causes salinity to fall or is it reduced tidal energy and dispersion? The discussion so far borrows the G-model premise that outflow determines salinity – filling and draining merely provide a correction to outflow. In contrast, one preliminary DSM2 study[8] suggests that in DSM2, the most important mechanism is tidal energy and dispersion and that the spring-neap phenomena can be reproduced even if low-frequency water levels are held artificially flat. The issue is probably best settled by analysis of observed data, since neither model was calibrated in a way that ensures that one transport mechanism was not traded off in favor of the other.

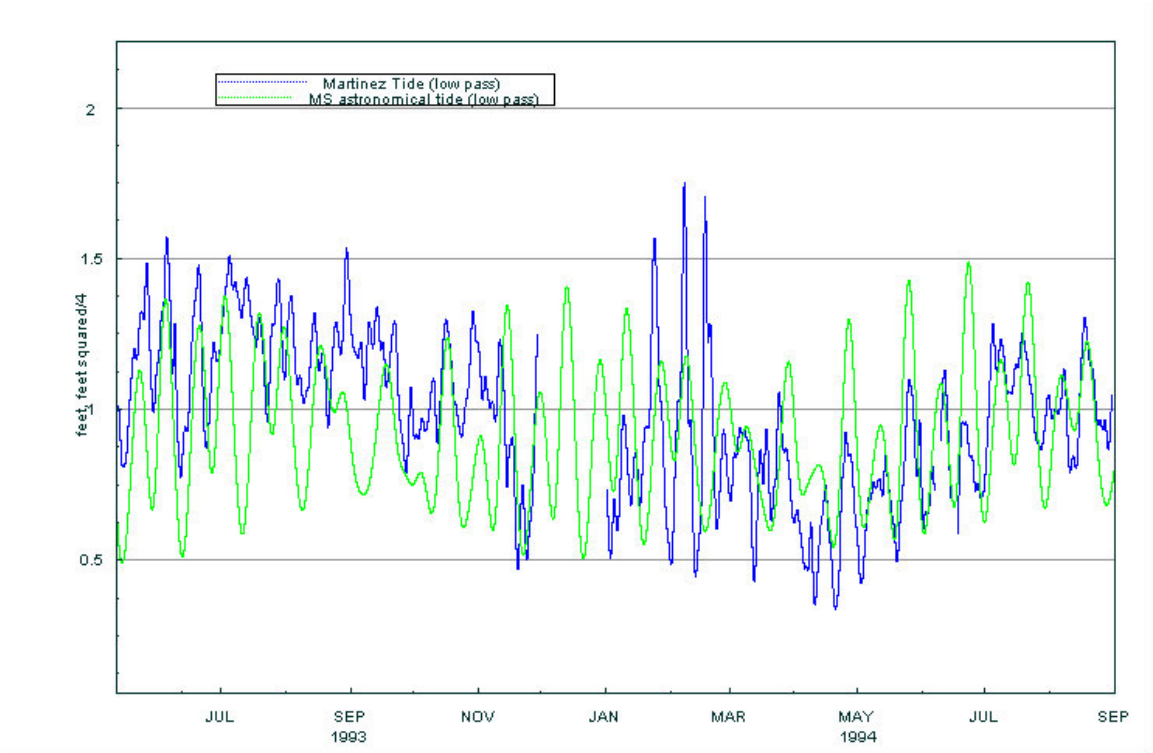

Figure 11-9: Comparing Low-Passed Martinez Stage (Filling and Draining) to Low-Passed Square Tide (a Measure of Tidal Energy).

11.7 References

Denton, R. and G. Sullivan. (1993). *Antecedent Flow-Salinity Relations: Applications to Delta Planning Models.* Contra Costa Water District.

Liang, K.-Y. and S.L. Zeger. (1986). *Longitudinal Data Analysis Using Generalized Linear Models.* Biometrika, **73**. pp.13-22.

[8] The study, over August – October 1994, compared a simulation using historical Martinez EC and stage to a simulation using historical EC but with stage that had been high passed to remove low frequency fluctuations (below 0.5 cycles per day). The two runs gave nearly identical results at interior stations and both exhibited lower salinity during periods of reduced tidal energy.

Munk, W.H. and D.E. Cartwright. (1966). *Tidal Spectroscopy and Prediction*, Phil. Trans. Roy. Soc. (London), Ser. A **259**, pp. 533-81.

Oppenheim, A.V. and A. S. Willsky. (1996). *Signals and Systems*, Prentice Hall.