ATTACHMENT E

Bootstrap Analysis of Turbidity Errors in RMA Adult Delta Smelt Entrainment Prediction

Bootstrap Analysis of Turbidity Errors in RMA Adult Delta Smelt Entrainment Prediction

Resource Management Associates

DRAFT –October 9, 2009

Contents

1 Introduction

This document describes the initial work to consider the impact of errors in the turbidity field used to drive the adult delta smelt behavior model utilizing a bootstrap analysis technique. The behavior algorithm used within the RMA particle tracking model to represent movement of adult delta smelt is strongly dependent on the time dependent turbidity field produced by the RMA Bay-Delta model (RMA, 2008). The turbidity modeling capability is in an early phase of development. While the model provides reasonable results overall, there remain uncertainties related to incomplete information for turbidity sources from boundary flows, Delta island return flows, and re-suspension related to wind and tidal flows. To determine the potential impact of errors in the turbidity prediction on the prediction of adult delta smelt distributions and entrainment, the RMA particle tracking model has been enhanced to simulated multiple particle model realizations driven by a single run of the RMA Bay-Delta model with bootstrap errors applied to the predicted time dependent turbidity field. The output from the set of realizations can be used to estimate confidence intervals about the mean predicted result. This technique has been developed in cooperation with Dr. Bryan Manly of Western EcoSystems Technology, Inc.

The bootstrap technique assumes that the best estimate for error distribution in future predictions of a quantity comes from computed previous errors. Future bootstrap predictions can be made by adding a randomly selected historic error to a raw prediction.

In the study case of interest here, the quantity analyzed is turbidity and errors are calculated from differences between RMA model computed predictions and observed values for 11 locations in the Delta through the period from December 1st, 2007 to March 31st, 2008. The errors are therefore going to depend on space (station) and time (day number).

Applying the bootstrap idea to the RmaSim program, the turbidity is computed at a different year (not necessarily in the future) and it is modified by adding an error selected at random from the list of errors described in the previous paragraph. Repeating this process produces a set of sample turbidity time series that conform to the error distribution assumption.

The data produced is then used to simulate the smelt behavior in the Delta using RMA particle tracking code. The entrainment of fish in the CVP, SWP pumping stations is of special interest. Entrainment time series are produced for each bootstrap experiment. The result is a set of samples of Entrainment data corresponding to different bootstrap turbidity errors. The data can be statistically analyzed to estimate the smelt behavior within a selected degree of confidence.

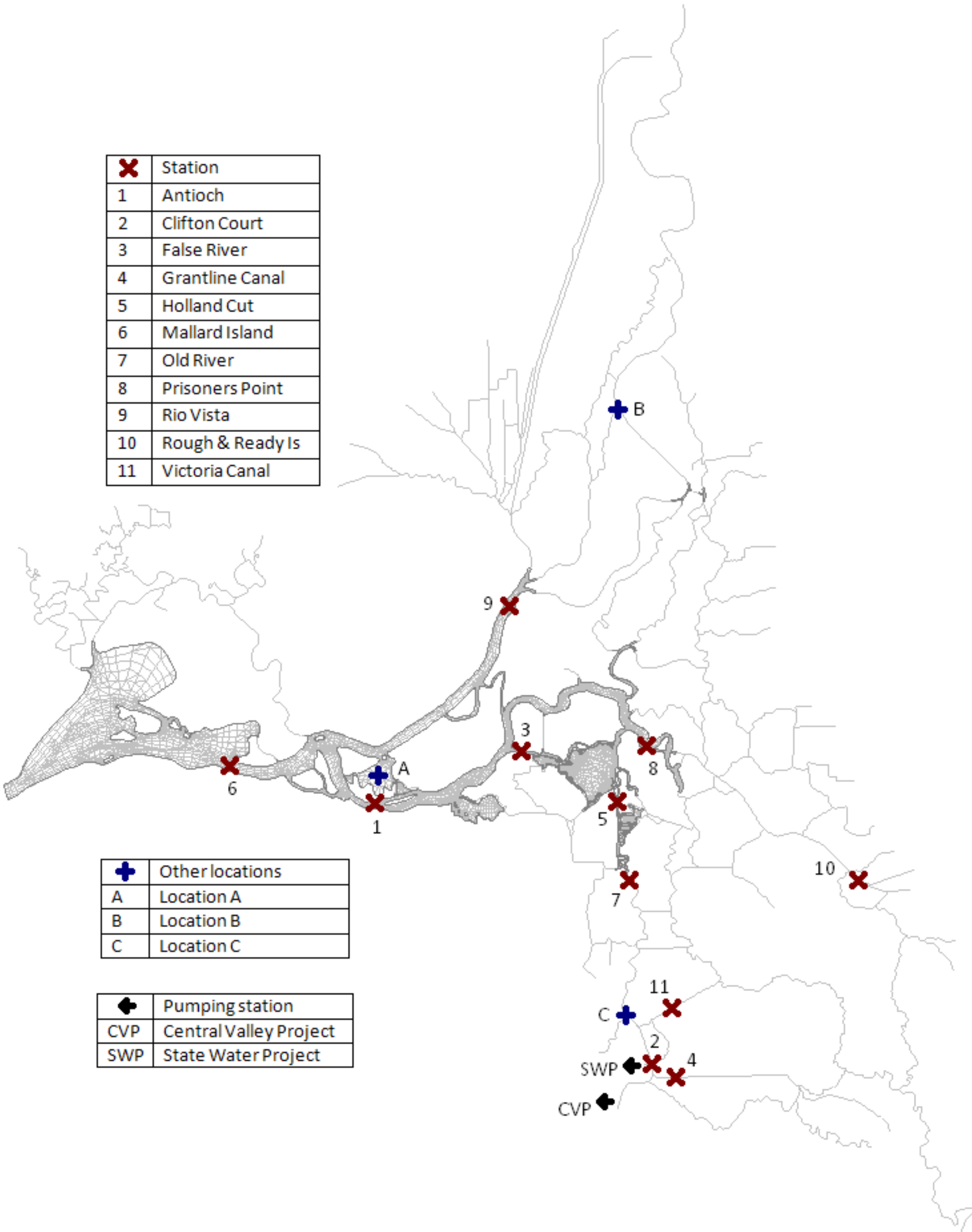


Figure 1. Map of the delta, showing stations where turbidity data was collected (red X), pumping stations (black ←) and other locations (blue +).

2 Data preparation

Analysis of the data at every station also shows that the errors are skewed. That means that the natural scale provided by the unit system (NTU) is not appropriate for the analysis. Instead, the data is transformed to logarithms of turbidity providing a more normally distributed distribution of errors. The rest of the analysis will be made in this new unit scale.

$$E_i = \ln(\mathbf{Comp.Turb}) - \ln(\mathbf{Obs.Turb}) = \ln\left(\frac{\mathbf{Comp.Turb}}{\mathbf{Obs.Turb}}\right)$$

As seen in Table 1, there are missing observed turbidity values for some days in six of the 11 stations. This problem is solved by interpolating the values from neighboring stations where data is present. This step is justified from the strong spatial correlations between station values.

Location	First Day	Last Day	Days	Mean Obs	Mean RMA11	Mean Error	SD Error	Serial Corr
Antioch	18-Jan-08	31-Mar-08	74	35.6	31.5	4.1	8.2	0.9
Clifton Court	1-Dec-07	31-Mar-08	120	21.8	14.7	7	9.2	0.72
False River	3-Jan-08	31-Mar-08	89	30.6	27.2	3.5	9.8	0.86
Grantline Canal	1-Dec-07	31-Mar-08	122	23.5	24.1	-0.6	9.8	0.52
Holland Cut	1-Dec-07	31-Mar-08	122	21.2	16.6	4.6	13.9	0.55
Mallard Island	1-Dec-07	3-Mar-08	94	40.3	32.1	8.2	8.7	0.88
Old River	28-Jan-08	31-Mar-08	64	24.5	21.9	2.5	7.8	0.74
Prisoners Point	1-Dec-07	31-Mar-08	122	13	24.3	-11.3	10.1	0.97
Rio Vista	1-Feb-08	31-Mar-08	60	37	40.8	-3.8	14.8	0.74
Rough & Ready Island	6-Dec-07	31-Mar-08	117	25.6	22.9	2.7	19.2	0.56
Victoria Canal	1-Dec-07	31-Mar-08	122	10	11.9	-1.8	5.3	0.96

Table 1. Locations where turbidity information was measured and their statistical indicators.

The bootstrap technique assumes that the errors in the data are driven by random processes and not by systematic inaccuracies. Two different approaches can be followed in order to remove any systematic bias.

The first correction approach is of order zero. It consists in centering the data so that the new set of errors has mean zero. In other words the formula

$$E_i^{\mathbf{0}} = E_i - \overline{E}_i,$$

is applied to the data at every station where $\overline{E}_i$ is the average of all errors at station i.

The second correction approach is of order one. It arises from the fact that in all stations there is a clear upward trend between days 1 and 40, followed by a downward trend between days 41 and 50. The corrections consist in subtracting linear estimates of these trends from the data.

$$E_i^{\mathbf{1}} = E_i - (a_i t + b_i)\ \mathbf{for}\ t = \mathbf{1}, \dots, \mathbf{40},$$

$$E_i^{\mathbf{1}} = E_i - (c_i t + d_i)\ \mathbf{for}\ t = \mathbf{41}, \dots, \mathbf{50},$$

$$E_i^{\mathbf{1}} = E_i - k_i\ \mathbf{for}\ t = \mathbf{51}, \dots, \mathbf{122},$$

where t is the time in days and for every station i , $a_i t + b_i$ is the least squares approximation of the data for days $1, \ldots, 40$; $c_i t + d_i$ is the least squares approximation of the data for days $51, \ldots, 50$; and k_i is the average of the data for days $51, \ldots, 122$.

3 Correlation in time

The bootstrap is complicated by the fact that the data is correlated over time. Analysis of the errors through all stations shows a strong serial correlation between the errors computed in two consecutive days. Table 2 shows that, as the analysis is carried over data taken more days apart, the correlation reduces to a weak correlation for 15 days. It is therefore suitable to apply the technique of block bootstrapping (Chernick, 1999) and a block size of 2 weeks is selected.

Days Apart	Antioch	Clifton Court	False River	Grantline Canal	Holland Cut	Mallard Island	Old River	Prisoners Point	Rio Vista	R & R River	Victoria Canal
1	0.93	0.79	0.91	0.87	0.83	0.94	0.8	0.89	0.82	0.85	0.97
5	0.62	0.51	0.49	0.41	0.3	0.54	0.37	0.51	0.56	0.46	0.74
10	0.33	0.27	0.22	0.28	0.27	0.29	0.14	0.25	0.25	0.22	0.45
15	-0.01	0.16	0.07	0.11	0.05	0.11	-0.03	0.22	-0.07	0.1	0.17

Table 2. Correlation between observed and estimated errors in Ln(turbidity) values 1, 5, 10 and 15 days apart.

The technique consists in that once every 14 days; a block of consecutive 14 days of error data will be chosen at random from the 122 days of data. This ensures that the bootstrap errors show correlation in time just like the real errors do, but there is little correlation between the data at the beginning and end of every block of data. In case the end of the 122 days is reached, the block is completed with data from the beginning of the period, guaranteeing that all errors have the same chance of being selected.

4 Spatial Interpolation

Observed data for turbidity is available only in $m = 11$ stations through the delta. However, the particle tracking simulation requires turbidity information in every node of the grid. Fortunately, the continuity of the turbidity implies that there is a strong correlation of the turbidity values at nearby locations. This allows for the implementation of some interpolation algorithm to estimate turbidity values everywhere on the grid.

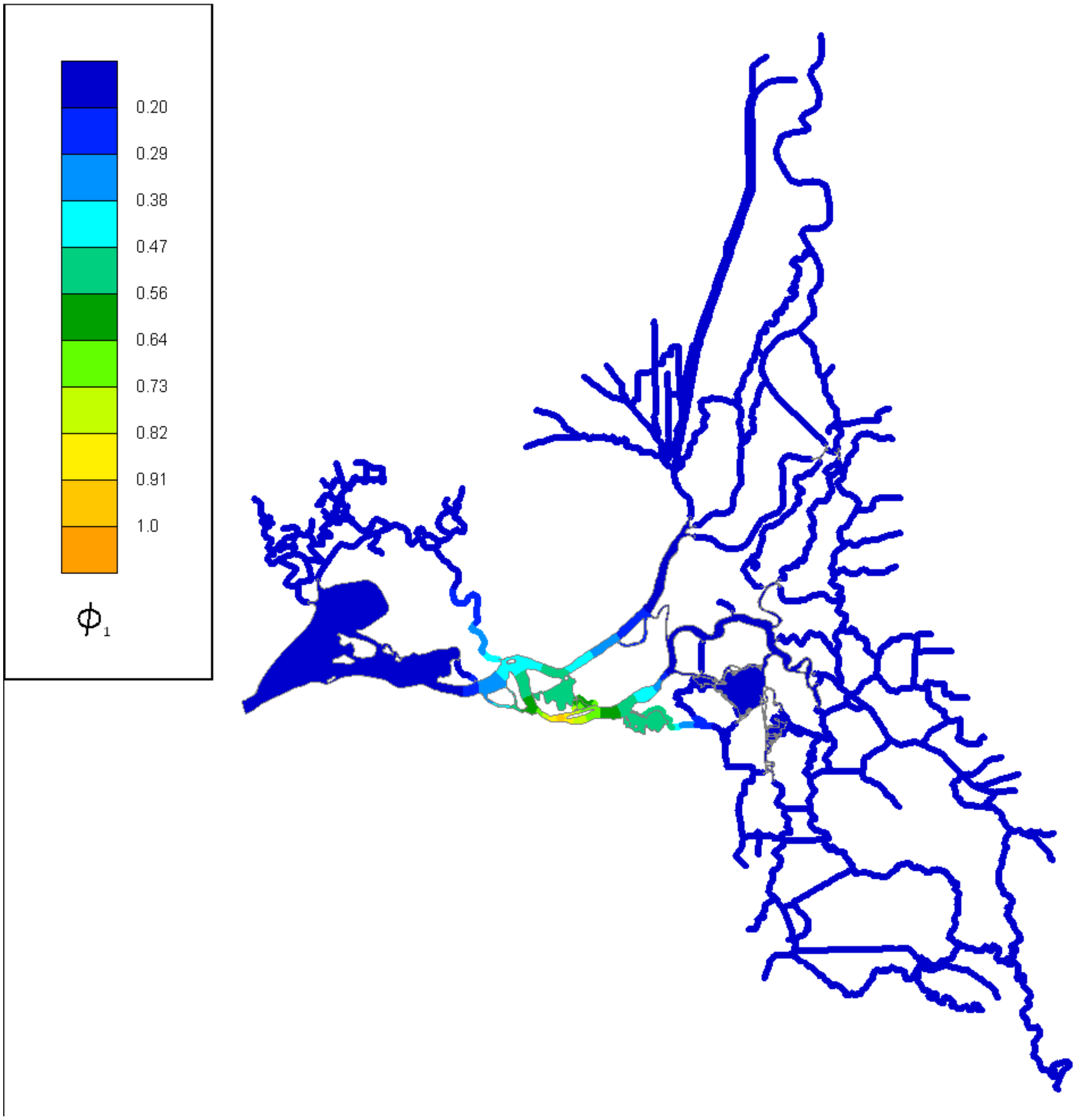


Figure 2. Weight function for station 1 (Antioch).

A sequence of weight functions ϕ_i, $i = 1, \ldots, m$ is constructed such that

- $0 \le \phi_i(x) \le \mathbf{1}$ for all nodes x on the grid.
- $\phi_i = \mathbf{1}$ at station i .
- $\phi_i = \mathbf{0}$ at stations different from i .

- $\sum_{i=1}^{m} \phi_i(x) = 1$ at any node x .

Given the complicated geometry of the delta, the best approach to produce the weight functions is to solve the Laplace equation with Neumann boundary conditions

$$\frac{d^2\phi_i}{dx^2} + \frac{d^2\phi_i}{dy^2} = 0$$

$$\phi_i = 1 \text{ at station } i$$

$$\phi_i = 0 \text{ at every station } \neq i$$

$$\frac{d\phi_i}{dn} = 0 \text{ at boundaries}$$

Once the weight functions have been computed, the error value at a given node of the grid is given by

$$E(x) = \sum_{i=1}^{m} E_i \phi_i(x).$$

5 Algorithm

The particle tracking area of the program RmaSim has a module to simulate the behavior of smelt in a watershed. It requires information about the fish environment such as water depth, velocity and quality (turbidity is one of them). It also reads some parameters that determine how the fish react to its environment, like preferred turbidity range and swimming velocity upstream and downstream.

This program is extended to allow the possibility of iteration with different input turbidity fields and to be able to construct those turbidity fields in a bootstrap fashion. A catalog of errors is prepared in advance. At every time step, a new turbidity field is prepared for the particle tracking code, by adding a selected error field to the original turbidity field. This procedure is done in a very careful way because in general the error data and the particle tracking code use different time step lengths. Moreover, some bookkeeping has to be done to be able to reset the time window every two weeks.

The user provides number of bootstrap iterations, time block size (14 days), error data at stations and their corresponding weight function information. User also provides typical particle tracking data such as water depth, velocity and quality fields.

```
For k = 1,..,<number of bootstrap iterations>
    For j = 1,..,<number of time steps>
        Prepare bootstrap error for time step j
        At every grid node make: <turbidity> = <turbidity> * <interpolated error>
        Compute one step of the particle tracking code
    End
End
```

The algorithm to implement bootstrapping is described above. Since all the calculation are made in terms of the natural logarithm of the turbidity, the turbidity obtained from adding the bootstrap error is given by the formula

$$\ln(\text{bootstrap turbidity}) = \ln(\text{turbidity}) + \ln(\text{interpolated error}).$$

An algebraic manipulation using the laws of logarithms yields the formula used in the psudo-code,

$$\text{bootstrap turbidity} = \text{turbidity} \times \text{interpolated error}.$$

The preparation of the bootstrap error entails the construction of an error field. The bootstrap error data for the stations is given in time steps of a day, but the particle tracking code works with a lot smaller time step. As a consequence, the code must figure out the field of the logarithm of the error at the beginning and the end of the day (**LnError0, LnError1**) and interpolate to find the (logarithm) error field at the precise time of the day required by the particle tracking code. The interpolation constant **Lambda** balances the importance of the error at the beginning and end of the day. It can take a value between 0 and 1.

Compute time block length
If ((<time step> == 1) or (<time block length> == 2 weeks))
    Record new beginning of the time block
    Select a random day to start new time block
End
If (simulation time corresponds to a new day)
    Recompute the error field at the beginning (*t0*) and end (*t1*) of the day

$$\text{LnError0}(x) = \sum_{i=1}^{m} \ln E_i(t0)\,\phi_i(x), \qquad \text{LnError1}(x) = \sum_{i=1}^{m} \ln E_i(t1)\,\phi_i(x)$$

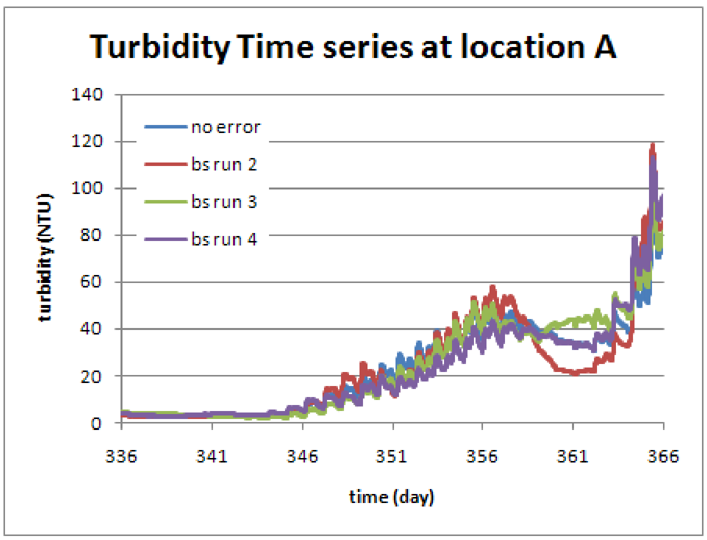
End
    Recompute interpolation constant
        Lambda = fraction of the day corresponding to simulation time
Compute error field: $\text{LnError}(x) = (1 - \text{Lambda})\text{LnError0}(x) + (\text{Lambda})\text{LnError1}(x)$

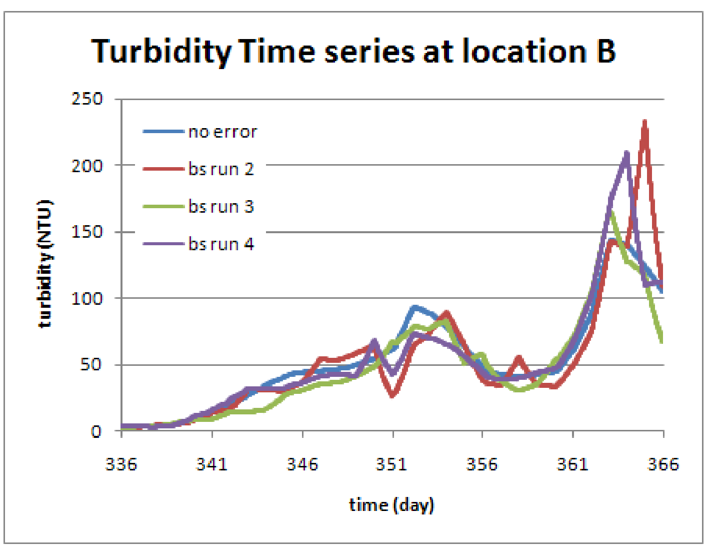
Since the errors are sequentially computed for consecutive days inside of any time block, **LnError0** does not have to be computed from scratch using the formula given. In fact **LnErrro0** is identical to the **LnError1** field of the previous day. So when the simulation time indicated that a new day of error fields must be computed, it suffices to make **LnError0 = LnError1**, and recomputing **LnError1**,saving about half the computing time.

6 Results

6.1 Turbidity

The turbidity results are analyzed for the three locations A, B, C (see Figure 1). The RMA computed turbidity data at three locations is presented in Figure 3 together with three more turbidity time series obtained by adding bootstrap errors. The data correspond to the month of December 2003. Locations A and C show oscillations due to the tidal flows. The oscillations persist in the bootstrap turbidity curves. Location B is further upstream and it does not present tidal oscillations. The transition between time windows at days 350 and 364 is noticeable in location B due to the relative smoothness of the time series.





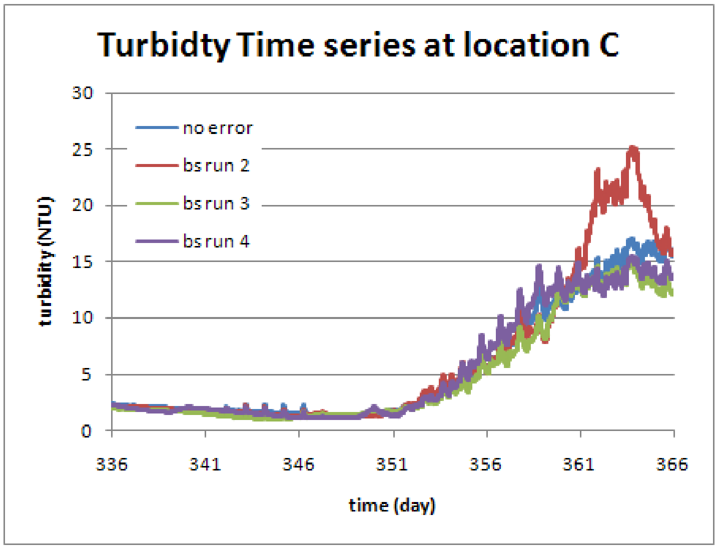


Figure 3. Comparison of original RMA model predicted turbidity (no error) and three examples of turbidity obtained using the bootstrap method (bs run 2-4). The time series correspond to locations A, B, C shown in Figure 1.

The turbidity results for day 362 (December 28th, 2003) at 00:00 hours are presented in Figure 4. The contour levels follow a logarithmic scale for convenience. As expected from the time series in Figure 4, the bootstrap turbidity values around location A are lower than the RMA model values. It can be observed that the effect is produced by negative turbidity errors in the Antioch (1), Mallard Island (6), False River (3) and Rio Vista (9) stations, which are all in the neighborhood of location A.

The opposite behavior is observed around location C. The bootstrap turbidity is higher in that area, due to a positive turbidity error at nearby Victoria Canal (11) and Clifton Court (2). Grantline Canal (4) station has a negative turbidity error, but it does not have an effect on location C because of the geometry of the region: there is no way to travel from location C to Grantline Canal without going through or nearby other stations.

Finally, bootstrap turbidity at location B is slightly lower than the RMA model because of the effect of Rio Vista (9) station negative error.

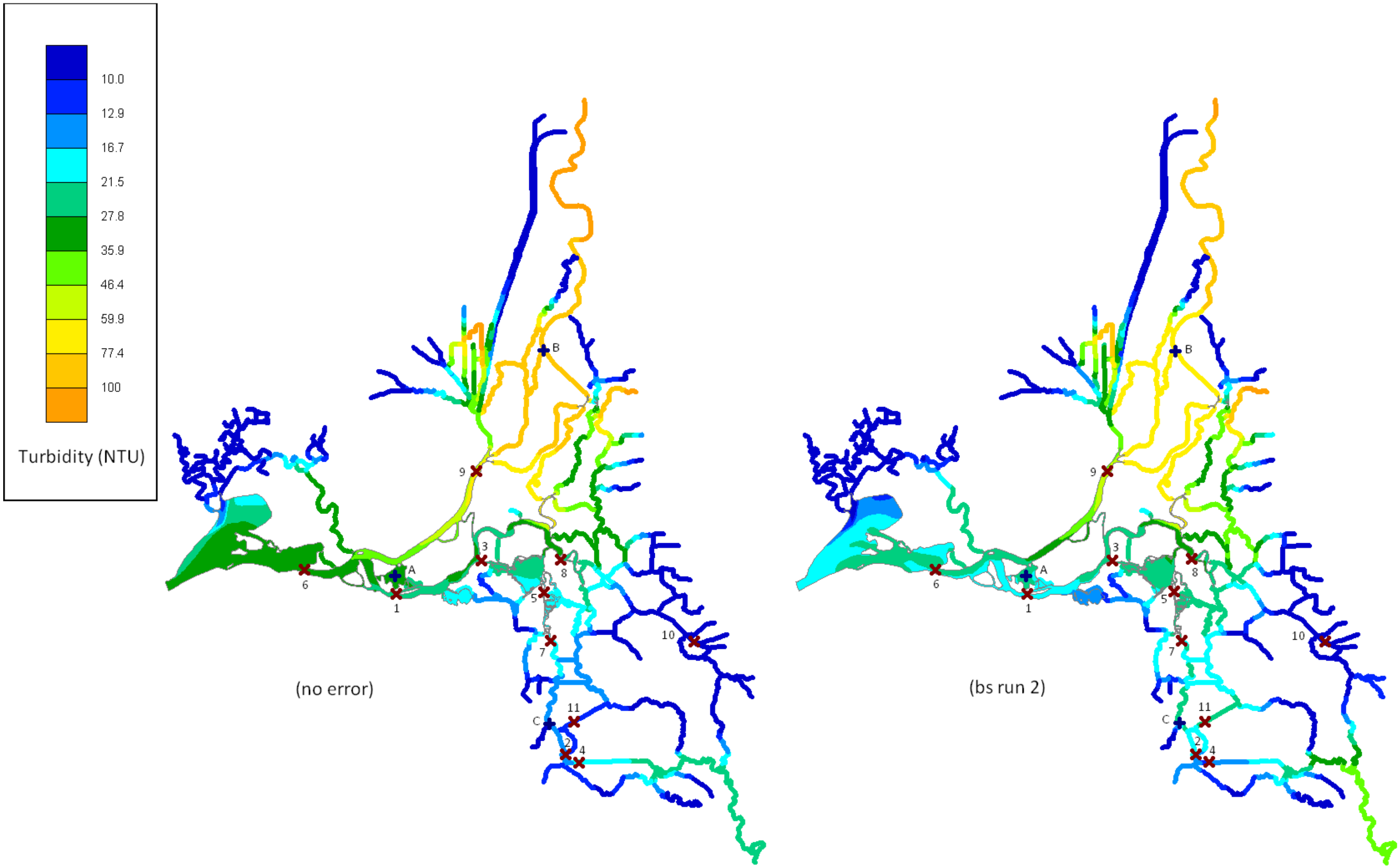


Figure 4. Comparison of the turbidity field at day 263. RMA model results are on the left (no error) and bootstrap modified turbidity on the right (bs run 2).

6.2 Entrainment

The number of fish entrained at pumping stations CVP and SWP (see Figure 1) is of particular interest to this study. The fish count is computed over the period of four months from December 1st, 2003 to March 31st, 2004. Cumulative Entrainment count is presented for the plain particle tracking as well as for a bootstrapped particle tracking with 39 bootstrap iterations.

For a given time, a new, independently calculated entrainment count has 1/40 chance of being above all 39 bootstrapped counts. Similarly, there is a 1/40 chance the count will be below the bootstrap curves. Conversely, there is a 38/40 = 95% chance the data will be between the lower and upper values. The range spanning the values of all 39 curves can therefore be seen as an interval for a 95% confidence level.

To test the bootstrap error analysis methodology, a set of simulation were performed based on the three of the initial 2003-2004 simulations developed for the 2-Gate project. The runs selected included Historic conditions (HIST), OCAP Lower Bound (OCAP-LB), and 2-Gate Adult Operation for the Lower Bound (2-Gate-LB). When reviewing the draft results below, note that export reductions for the OCAP-LB run are based on the three station turbidity average and start later than the export reductions for the 2-Gate-LB simulation which is based on the Jersey Point Trigger. The bootstrap analysis will be re-run when simulation results are available for updated No Project and With Project conditions.

The results for Historic, OCAP-LB and 2-Gate simulations are presented in the figures below. It can be observed that OCAP-LB represents a reduction in entrainment of about 93% with respect to the historic data, and that the 2-Gate model represents an additional reduction of around 85% with respect to OCAP-LB.

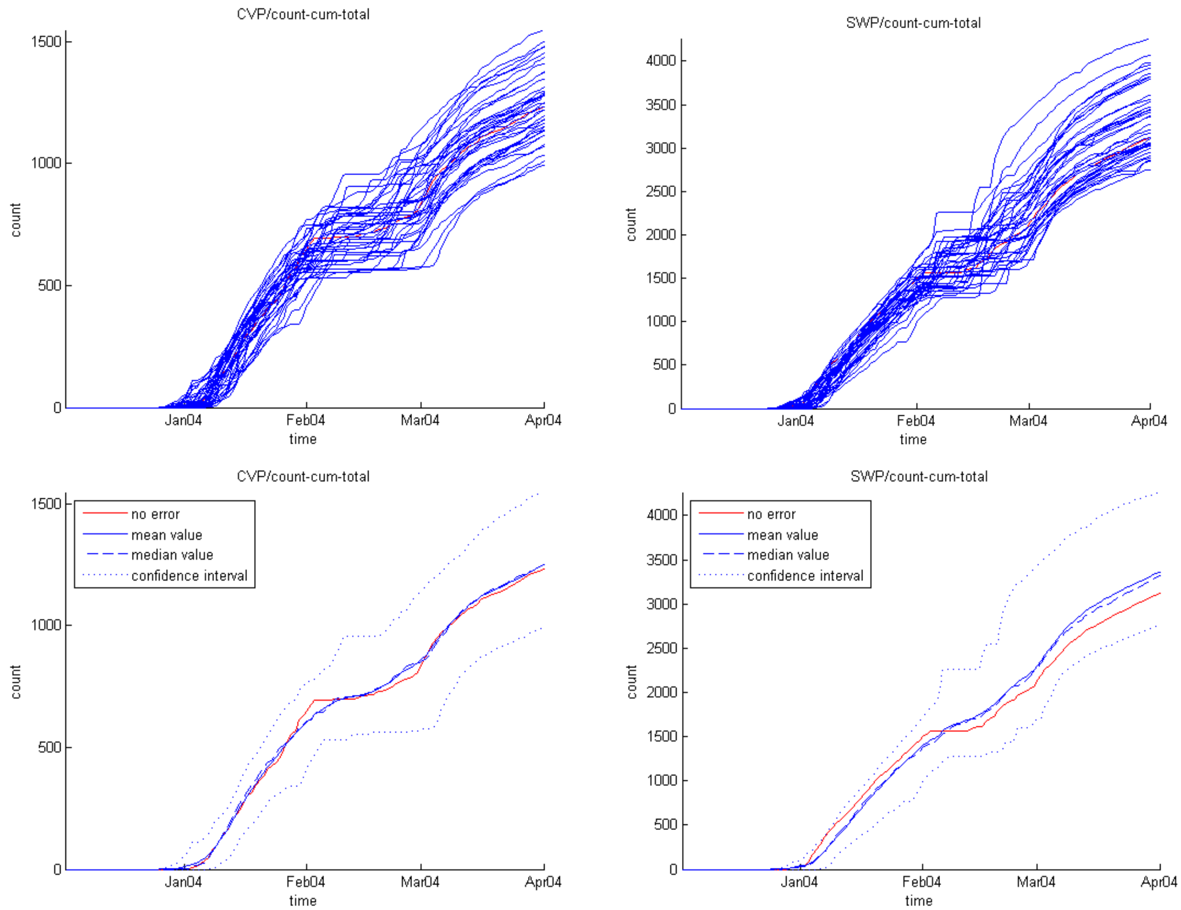


Figure 5. Historic simulation. Smelt bootstrap entrainment at CVP (left) and SWP (right) and simulation with no bootstrap error (red). All 39 bootstrap runs (above) and confidence interval (below) are shown.

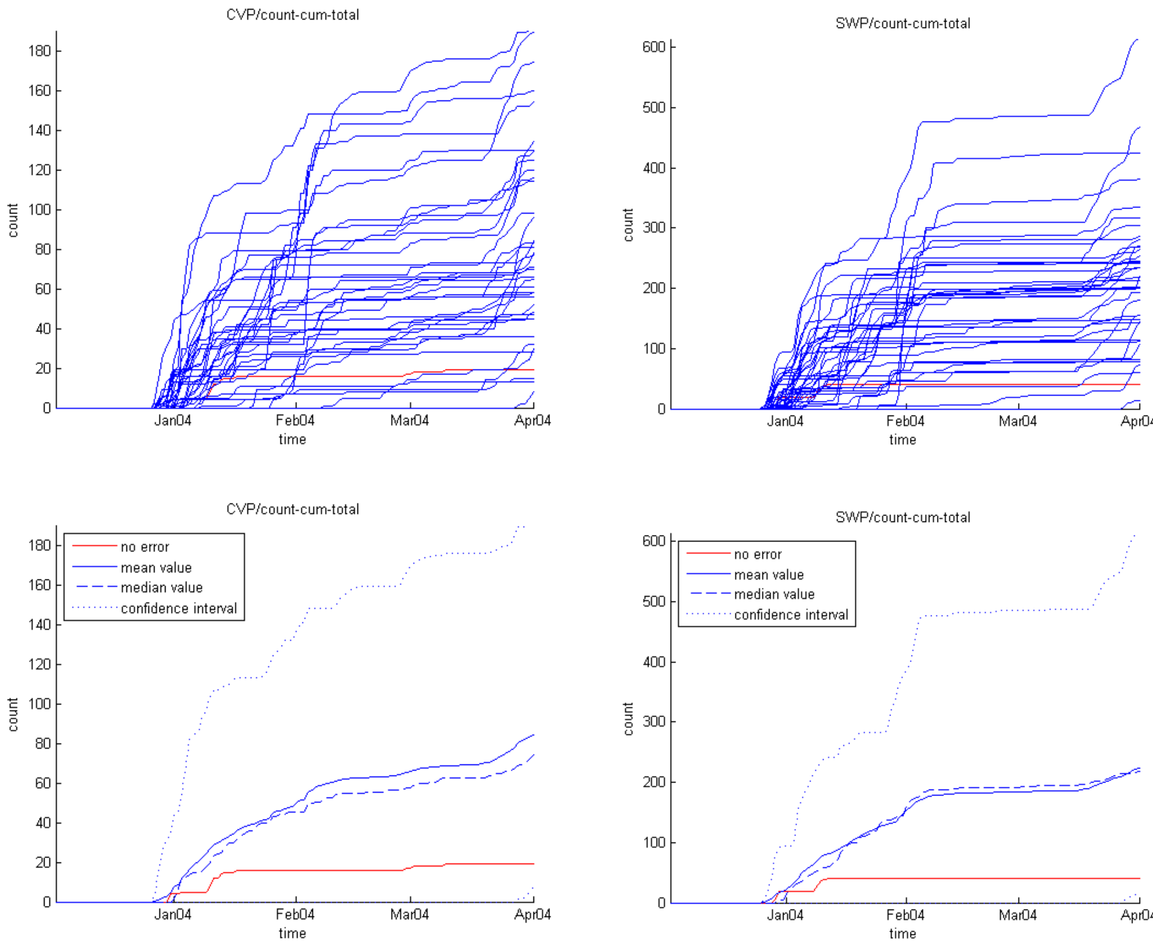


Figure 6. OCAP-LB simulation. Smelt bootstrap entrainment at CVP (left) and SWP (right) and simulation with no bootstrap error (red). All 39 bootstrap runs (above) and confidence interval (below) are shown.

In the OCAP-LB and 2 gate simulations, the bootstrap runs produce more entrainment than the no error simulation. A reason for that could be that the bootstrap errors produce secondary gradients in turbidity that could become significant when the turbidity field is flat. Those secondary gradients could be large enough to guide the fish to surf in an unexpected direction and produce extra diffusion of the particles.

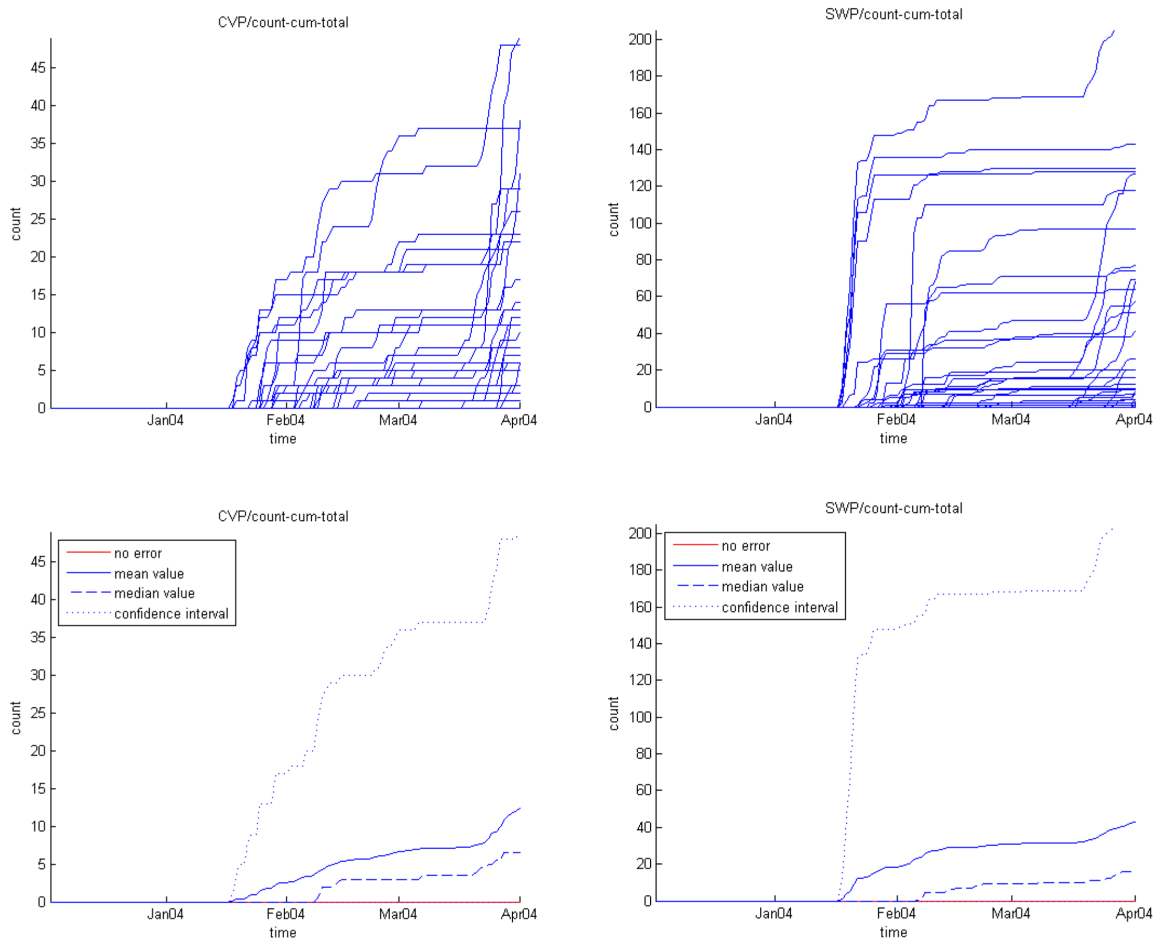


Figure 7. 2Gate-LB simulation. Smelt bootstrap entrainment at CVP (left) and SWP (right) and simulation with no bootstrap error (red). All 39 bootstrap runs (above) and confidence interval (below) are shown.

7 Initial Conclusions

The RMA particle tracking model has been enhanced to simulate multiple particle model realizations driven by a single run of the RMA Bay-Delta model with bootstrap errors applied to the predicted time dependent turbidity field. The output from the set of realizations has been used to estimate confidence intervals about the mean predicted entrainment for several sample simulation conditions. Results from the example simulations suggest that the methodology may provide useful insights for the uncertainty in entrainment related to uncertainty in the turbidity model.

For the historic condition simulation where there were moderate levels of entrainment, the mean entrainment from the bootstrap analysis is very similar to the entrainment from direct simulation without modification of the turbidity field. However for the other cases where entrainment levels were very low, the mean result from the bootstrap analysis is higher than the entrainment from direct simulation without modification of the turbidity field. The approach to applying bootstrap errors to the turbidity field may be leading to subtle discontinuities or local turbidity gradients that increase particle movement along the turbidity front, which is important in simulations where conditions such that entrainment is near zero.

This effort has been a first step in implementation of bootstrap error analysis. The approach will be carried forward as additional modeling is performed for analysis of the 2 Gate project.

References

Chernick, M.R. (1999). Bootstrap Methods: a Practitioner's Guide. Wiley, New York.

Elmore, K.L.; Baldwin, M.E.; Shultz, D.M. (2005). Field Significance Revisited: Spatial Bias Errors in Forecasts as Applied to the Eta Model. Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

Srinivas, V.V.[1]; Srinivasan, K.[2] (2005). Matched-block bootstrap for resampling multiseason hydrologic time series. [1] Department of Civil Engineering, Indian Institute of Science, Bangalore, India; [2] Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India; Wiley InterScience On-line.

Srinivas, V.V.[1]; Srinivasan, K.[2] (2005). Hybrid matched-block bootstrap for stochastic simulation of multiseason streamflows. [1] Department of Civil Engineering, Indian Institute of Science, Bangalore, India; [2] Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India; Wiley InterScience On-line.

Resource Management Associates, RMA (2008). Sacramento-San Joaquin Delta Turbidity Modeling, project report presented to Metropolitan Water District of Southern California.