

Statistical Data Assimilation for Parameter Estimation in the Advanced Circulation Model

Talea L. Mayo

Postdoctoral Research Associate
Civil and Environmental Engineering
Princeton University

April 4, 2014



Outline

1 Coastal Ocean Modeling

2 Data Assimilation

3 Numerical Results

Outline

1 Coastal Ocean Modeling

2 Data Assimilation

3 Numerical Results

Coastal Ocean Modeling

- model coastal ocean under moderate conditions:
 - prediction of tides
 - time history of elevation, velocity, temperature, salinity
 - ecological forecasts
 - coastal navigation
 - transport of contaminants



Coastal Ocean Modeling

- model extreme events such as hurricane storm surge and coastal inundation



Hurricane Ike, 2008



Hurricane Katrina, 2005

Coastal Ocean Modeling

- accuracy dependent on many factors:
 - approximate physics
 - numerical discretization
 - uncertain boundary/initial conditions
 - inaccurate model parameters**

$$\frac{\partial H}{\partial t} + \frac{\partial}{\partial x}(Q_x) + \frac{\partial}{\partial y}(Q_y) = 0$$

$$\frac{\partial Q_x}{\partial t} + \frac{\partial UQ_x}{\partial x} + \frac{\partial VQ_x}{\partial y} - fQ_y = -gH \frac{\partial [\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial x} + \frac{\tau_{sx}}{\rho_0} - \frac{\tau_{bx}}{\rho_0} + M_x - D_x - B_x$$

$$\frac{\partial Q_y}{\partial t} + \frac{\partial UQ_y}{\partial x} + \frac{\partial VQ_y}{\partial y} - fQ_x = -gH \frac{\partial [\zeta + P_s/g\rho_0 - \alpha\eta]}{\partial y} + \frac{\tau_{sy}}{\rho_0} - \frac{\tau_{by}}{\rho_0} + M_y - D_y - B_y$$

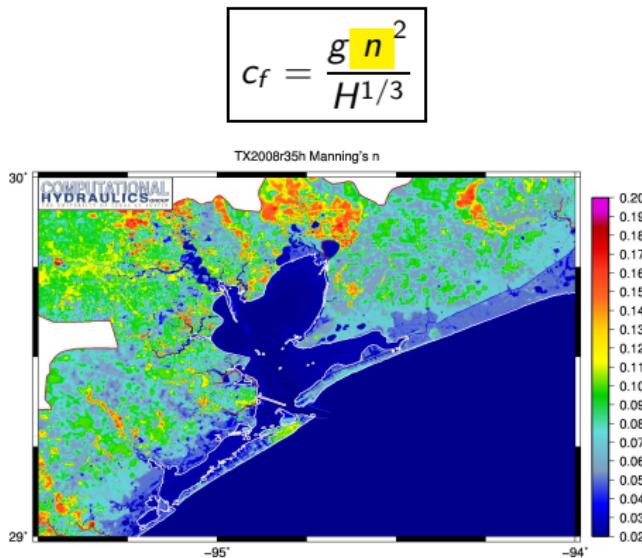
Bottom Stress

$$\frac{\tau_{bx_i}}{\rho_0} = c_f |\mathbf{u}| \frac{Q_{x_i}}{H}$$

- bottom stress terms in momentum equations
 - significant source of uncertainty
 - critical to accurately modeling free surface circulation
 - can be defined using linear or quadratic drag law (shown)
- c_f standard friction coefficient
 - often given uniform value due to lack of information about seabed
 - can be used as a tuning parameter (Signell et al., 2000)

Bottom Friction

- Manning's n coefficient of roughness
- spatially dependent
- highly variable
 - surface roughness
 - vegetation
 - silting and scouring
- cannot be measured directly, no exact method for selection
- values typically obtained from land cover databases



Bottom Friction



- use measured data to reduce uncertainties in model
- invert water elevation data using parameter estimation
 - traditional methods can be computationally intensive
- assimilate water elevation data using state estimation
 - does not provide information about uncertain model parameters

Outline

1 Coastal Ocean Modeling

2 Data Assimilation

3 Numerical Results

Kalman Filtering

- statistical data assimilation method
- determine the best estimate of state by combining model output with observed data
- quantify uncertainty in the state estimate
- can be performed sequentially, on-line
- nonintrusive and can be implemented in two steps

Kalman Filtering: Forecast

- given initial state, \mathbf{x}_{k-1}^a , and associated error covariance, \mathbf{P}_{k-1}^a
- can use the numerical forecast model to estimate the state and error covariance at a later time t_k :

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a$$

$$\begin{aligned}\mathbf{P}_k^f &= \overline{(\mathbf{x}_k^t - \mathbf{x}_k^f)(\mathbf{x}_k^t - \mathbf{x}_k^f)^T} \\ &= \mathbf{M}_k \overline{(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a)(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a)^T} \mathbf{M}_k^T + \overline{\eta_k^2} \\ &\quad + 2\mathbf{M}_k \overline{(\mathbf{x}_{k-1}^t - \mathbf{x}_{k-1}^a)\eta_k} \\ &= \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k\end{aligned}$$

Kalman Filtering: Analysis

- use observation operator to determine residual between forecasted state and data:

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f$$

- update model forecast by adding residual weighted by Kalman gain matrix:

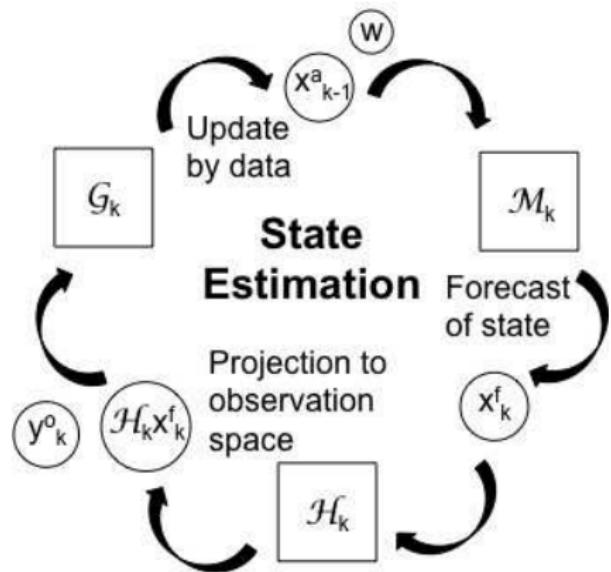
$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \mathbf{d}_k$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$$

- Kalman gain matrix minimizes analysis error covariance:

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H} \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}.$$

State Estimation



Parameter Estimation

State estimation:

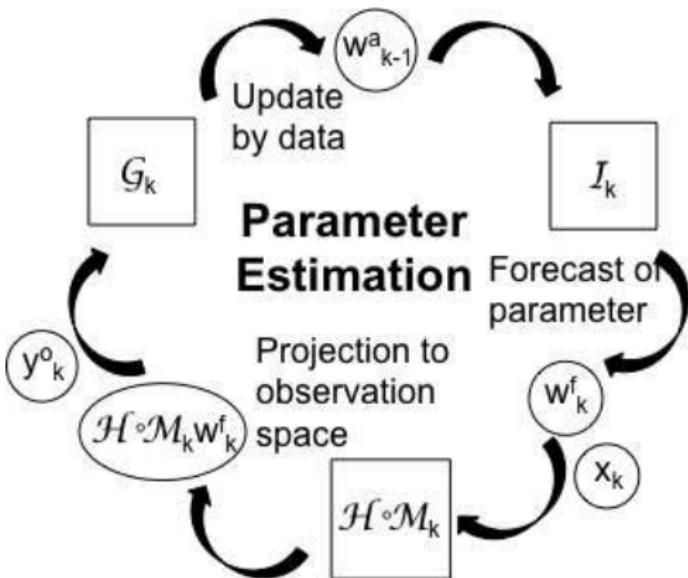
$$\begin{aligned}\mathbf{x}_k^f &= \mathcal{M}_k(\mathbf{x}_{k-1}^a, \mathbf{w}) \\ \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k^f \\ \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k \mathbf{d}_k\end{aligned}$$

Parameter estimation:

$$\begin{aligned}\mathbf{w}_k^f &= \mathcal{I} \mathbf{w}_{k-1}^a + \epsilon_w \\ \mathbf{y}_k^w &= \mathbf{H}_k \circ \mathcal{M}_k(\mathbf{x}_{k-1}^a, \mathbf{w}_k^f) \\ \mathbf{w}_k^a &= \mathbf{w}_k^f + \mathbf{K}_k^w \mathbf{d}_k^w\end{aligned}$$

- traditionally statistical data assimilation methods are used for state estimation (Butler et al., 2012; Altaf et al., 2013)
- methods can be used for parameter estimation by reformulating the operators \mathcal{M}_k and \mathcal{H}_k
 - the evolution of the model parameters considered a static process, $\mathcal{M}_k^w = \mathcal{I}$
 - composition of \mathcal{M}_k and \mathcal{H}_k projects parameter into observation space,
 $\mathcal{H}_k^w = \mathbf{H}_k \circ \mathcal{M}_k$

Parameter Estimation



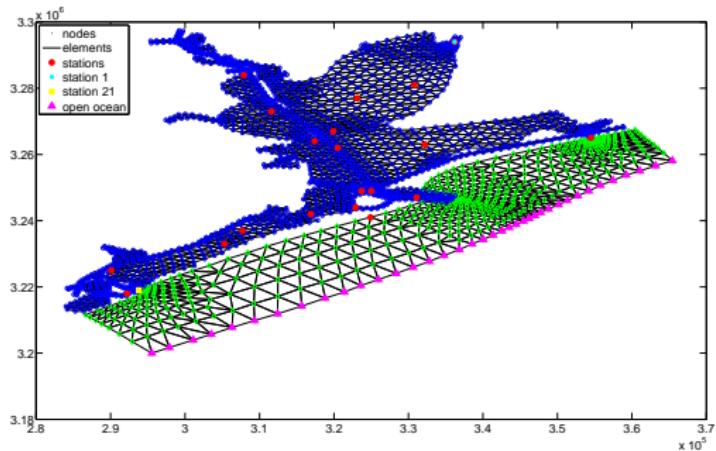
Outline

1 Coastal Ocean Modeling

2 Data Assimilation

3 Numerical Results

Galveston Bay

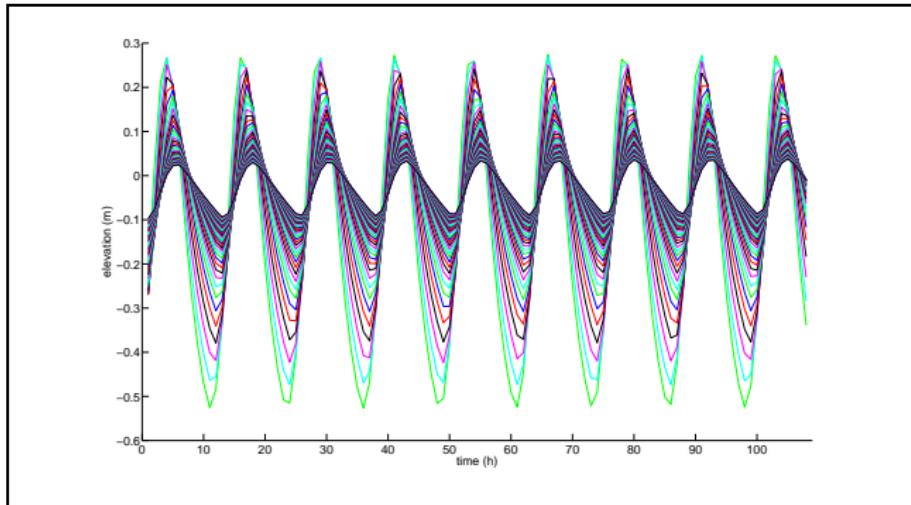


- 2,113 nodes, 3,397 elements
- bathymetry varies from 0.354 - 17.244 m
- 17 small islands within the bay
- 21 observation stations

Parameter Estimation Methodology: OSSEs

- use three low dimensional parameterizations to define various “true” fields of Manning’s n coefficients
 - constant
 - piecewise constant
 - realization of a stochastic process
- systematically choose true parameters and generate synthetic water elevation data
- using synthetic data, estimate the true parameters from various incorrect initial guesses
- determine accuracy of the parameter estimation based on estimated water elevations

Field of Constant Manning's n Coefficients

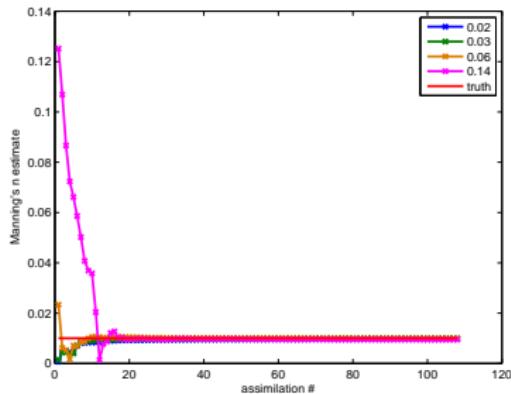


- model tides by forcing ADCIRC with M_2 tidal constituent
- let field of constant Manning's n coefficients range from 0.005-0.2
- divide coefficients into five classes based on mean tidal amplitude at station 1

Field of Constant Manning's n Coefficients

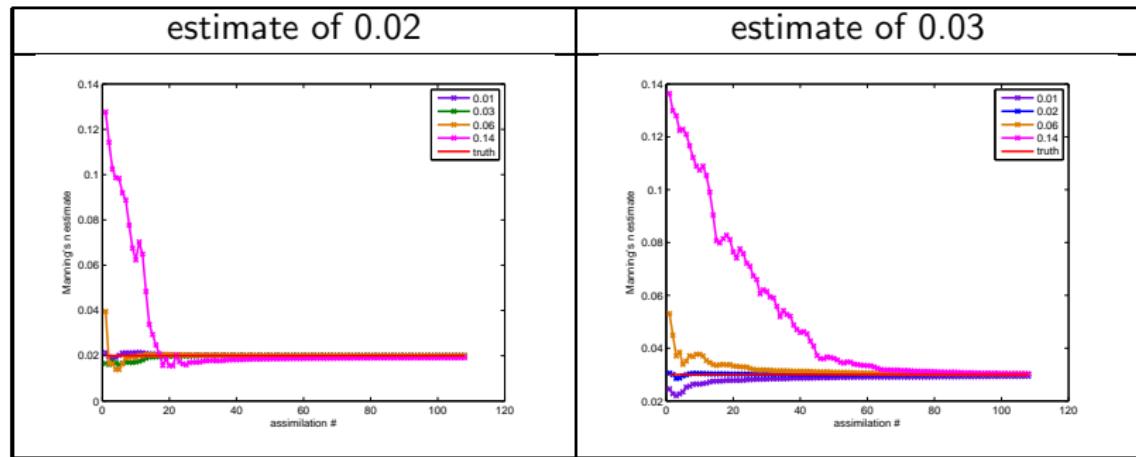
- for each experiment, define true parameter as value from each class
- let various initial guesses of parameter be values from the four remaining classes
- assimilate synthetic water elevation data every hour over a 5 day simulation period
- consider the parameter estimation successful if the final estimate lies within the correct class

Galveston Bay: 0.01

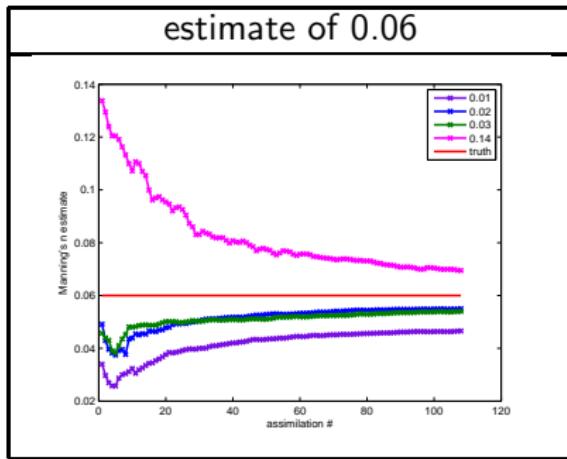


initial guess	0.02	0.03	0.06	0.14
Manning's n estimate of 0.01	0.009776	0.009903	0.009921	0.009479
Manning's n relative error	0.022383	0.009720	0.007877	0.052067
Elevation RMSE (SEIK filter)	0.003502	0.004191	0.007290	0.015166
Elevation RMSE (baseline)	0.016351	0.024598	0.033514	0.038380

Galveston Bay: 0.02 and 0.03

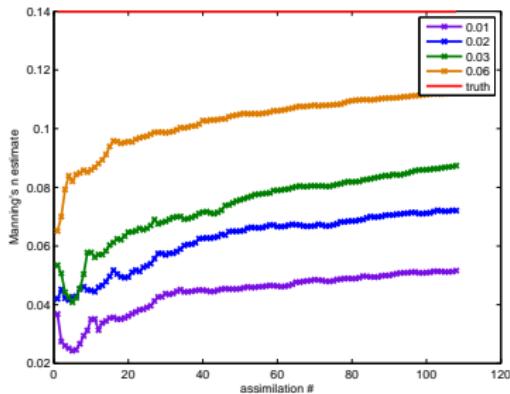


Galveston Bay: 0.06



Truth	Initial Guess	Parameter Recovered
0.02	0.01	y
0.02	0.03	y
0.02	0.06	y
0.02	0.14	y
0.03	0.01	y
0.03	0.02	y
0.03	0.06	y
0.03	0.14	y
0.06	0.01	y
0.06	0.02	y
0.06	0.03	y
0.06	0.14	y

Galveston Bay: 0.14



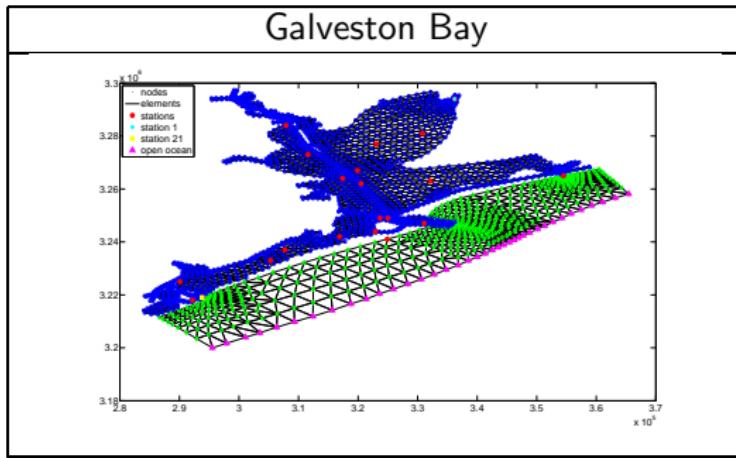
initial guess	0.01	0.02	0.03	0.06
Manning's n estimate of 0.14	0.051604	0.072078	0.087391	0.112387
Manning's n relative error	0.631399	0.485155	0.375777	0.197233
Elevation RMSE (SEIK filter)	0.013696	0.008704	0.006591	0.002847
Elevation RMSE (baseline)	0.038380	0.026017	0.018304	0.007729

Discussion: Constant Manning's n Coefficients

- able to recover parameter from all classes except class E
 - unable to recover class E from smallest initial guesses
- relative error in estimate increases with the true value
 - ADCIRC not as sensitive to variations in larger coefficients
 - smaller updates to parameters by SEIK filter
 - more updates and/or better initial guess required to obtain estimates in correct class
- RMSEs of estimated water elevations always reduced with respect to baseline case

Field of Piecewise Constant Manning's n Coefficients

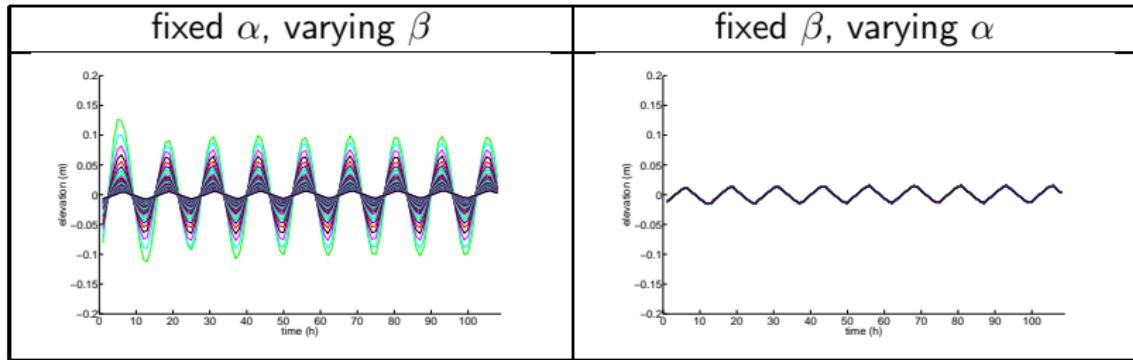
- let true Manning's n coefficient equal 0.005 in the deep water (green) and 0.1 in the shallow bay area (blue)



Field of Piecewise Constant Manning's n Coefficients

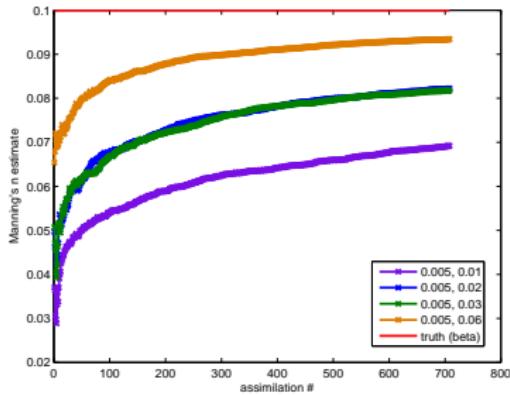
- choose various initial guesses
- assimilate synthetic water elevation data every hour over a 30 day simulation period
- consider the parameter estimation successful if the estimated mean tidal amplitude at station 21 is within 20% of the truth

Galveston Bay: $n_{0.005,0.1}$



- assume $\alpha = 0.005$ is known
- initial guesses:
 - $n_{0.005,0.01}$
 - $n_{0.005,0.02}$
 - $n_{0.005,0.03}$
 - $n_{0.005,0.06}$

Galveston Bay: $n_{0.005, 0.1}$

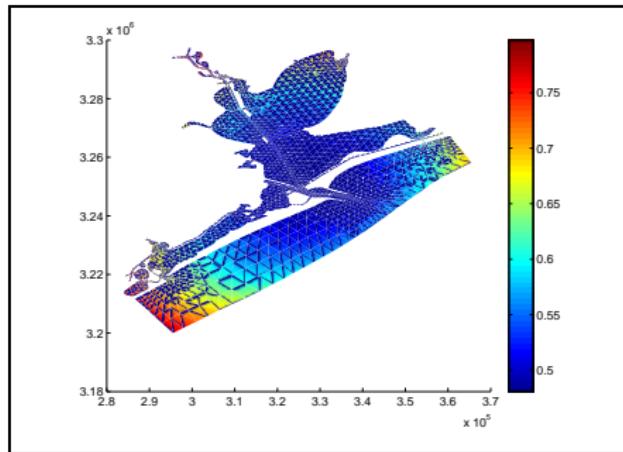


initial guess	$n_{0.005, 0.01}$	$n_{0.005, 0.02}$	$n_{0.005, 0.03}$	$n_{0.005, 0.06}$
estimate of $\beta = 0.1$	0.069189	0.082208	0.081744	0.093426
Relative error (β)	0.308112	0.177918	0.182555	0.065742
RMSE (SEIK)	0.005968	0.003506	0.003380	0.001313
RMSE (baseline)	0.036703	0.024333	0.016423	0.005416
Mean tidal amplitude (sta. 21)	0.100471	0.100696	0.100684	0.100858
Relative error	0.004738	0.002519	0.002635	0.000905

Discussion: Piecewise Constant Manning's n Coefficients

- able to accurately estimate parameters from all initial guesses
- find that the model is more sensitive to β , value of the Mannings n coefficient on the shallow side of the domain
 - less initial error in $\beta = 0.1$ produces more accurate final estimate
 - for Galveston Bay domain, variation in α has minimal effect on water elevations
- RMSEs of water elevations reduced by at least 75%
- estimated mean tidal amplitudes within 20% of truth (Mayo et al., 2014)

Realization of a Stochastic Process

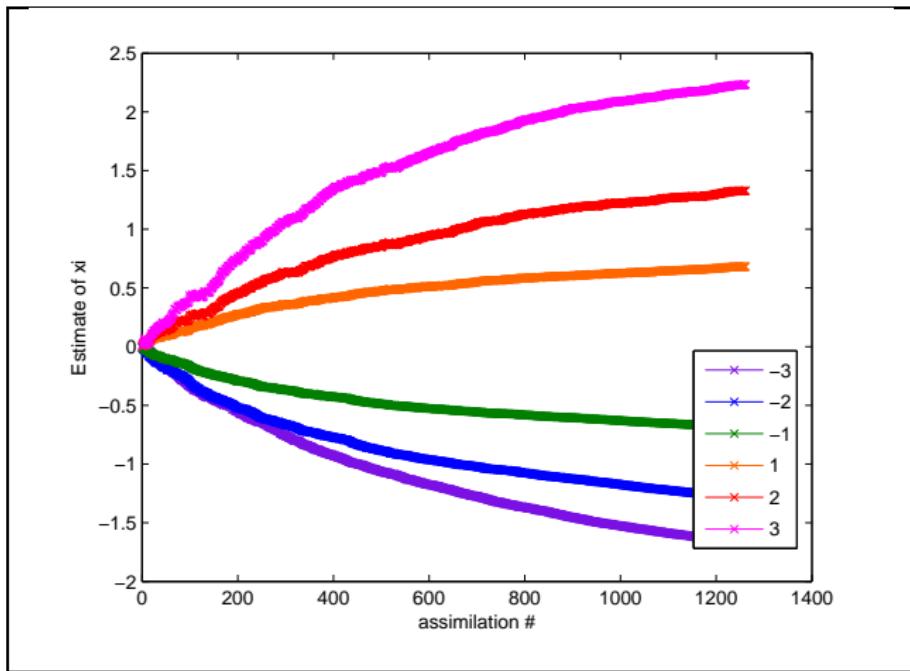


- choose true parameters to define fields of Manning's n coefficients that are physical, $\xi = -3, -2, -1, 1, 2, 3$
 - estimate natural log of process, $n(\mathbf{x}) = \exp [\bar{\omega}(\mathbf{x}) + \xi_1(\theta)\sqrt{\lambda_1}f_1(\mathbf{x})]$
 - let expectation equal value from class C
 - choose true parameters ξ_1 so that corresponding fields span the Manning's n coefficients in the remaining classes

Realization of a Stochastic Process

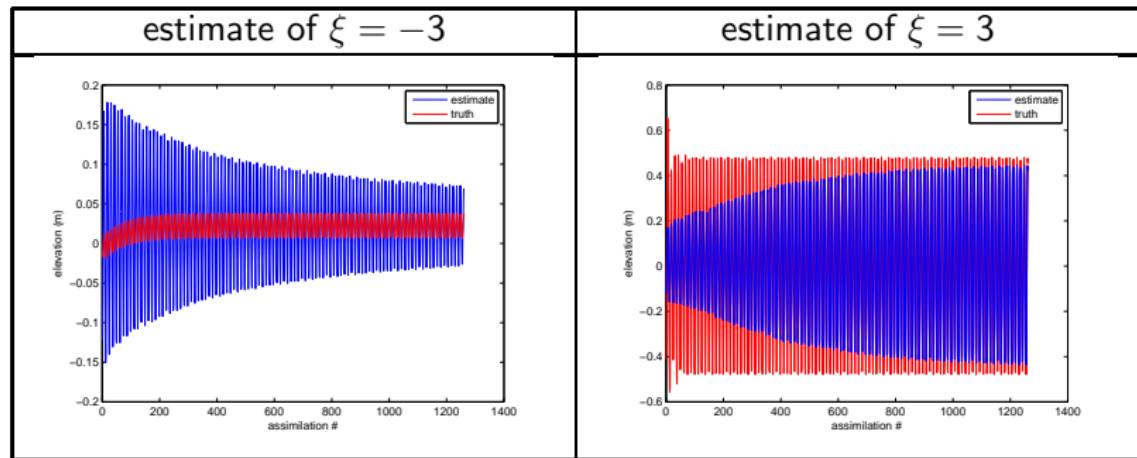
- let initial guess equal the expectation of the process, i.e. $\xi_1 = 0$
- assimilate synthetic water elevation data every hour over a 53 day simulation period
- consider the parameter estimation successful if the estimated mean tidal amplitude at station 21 is within 20% of the truth

Galveston Bay: $\xi = -3, -2, -1, 1, 2, 3$



Galveston Bay: $\xi = -3, 3$

- estimates of model states:



Galveston Bay: $\xi = -3, -2, -1, 1, 2, 3$

true value of ξ	-3	-2	-1	1	2	3
Final estimate	-1.685	-1.292	-0.689	0.682	1.330	2.233
Absolute error	1.315	0.708	0.311	0.318	0.670	0.767
RMSE of field (SEIK)	0.065	0.022	0.006	0.002	0.003	0.002
RMSE of field (baseline)	0.097	0.043	0.015	0.008	0.013	0.016
RMSE (SEIK)	0.042	0.031	0.019	0.033	0.077	0.096
RMSE (baseline)	0.077	0.062	0.038	0.056	0.118	0.165
True mean tidal amplitude (sta. 21)	0.455	0.458	0.460	0.471	0.488	0.504
Est. mean tidal amplitude (sta. 21)	0.460	0.460	0.461	0.466	0.470	0.481
Relative error	0.011*	0.006*	0.003*	0.011*	0.037*	0.047*

Discussion: Realization of a Stochastic Process

- able to accurately estimate parameters from all initial guesses
 - more time required for estimation of spatially varying field
 - numerical instability before true parameters were attained
- estimation of positive values more accurate than estimation of negative values
 - for true ξ negative, initial guess $\xi = 0$ is overestimate of state
 - for true ξ positive, initial guess $\xi = 0$ is underestimate of state
 - for consistent forcing, more difficult to remove fluid in system by increasing magnitude of bottom stress than the converse
- RMSEs of corresponding fields reduced between 33% and 97%
- maximum global RMSE in water elevations is 0.096 m
- estimated tidal amplitudes within 20% of truth

Thank you

- Altaf, M., Butler, T., Luo, X., Dawson, C., Mayo, T., and Hoteit, I. (2013). Improving short range ensemble kalman storm surge forecasting using robust adaptive inflation. *Monthly Weather Review*, 141(2013):2705–2720.
- Butler, T., Altaf, M. U., Dawson, C., Hoteit, I., Luo, X., and Mayo, T. (2012). Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Monthly Weather Review*, 140:22152231.
- Mayo, T., Butler, T., Dawson, C., and Hoteit, I. (2014). Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning's friction coefficient. *Ocean Modelling*. accepted for publication.
- Signell, R., List, J., and Farris, A. (2000). Bottom currents and sediment transport in long island sound: A modeling study. *Journal of coastal research*, pages 551–566.