Adaptive Sampling in The COlumbia RIvEr Observation Network

Thanh Dang, Nirupama Bulusu, and Wu-chi Feng
Department of Computer Science
Portland State University
dangtx,nbulusu,wuchi@cs.pdx.edu

Sergey Frolov and Antonio Baptista
Department of Environmental and Biomolecular Systems
OGI School of Science and Engineering
Oregon Health and Science University
frolovs, baptista@stccmop.org

Categories and Subject Descriptors

I.6 [**Simulation and Modeling**]: Model Validation and Analysis; J.2 [**Physical sciences and Engineering**]: Earth and atmospheric sciences

General Terms

Sensor networks, adaptive sampling, data assimilation, sigma point Kalman filter

1 Introduction

The Columbia River (CoRie) Observation Network includes an extensive array of fixed stations monitoring the Columbia River estuary and nearby coastal ocean. At each station, variable combinations of in-situ sensors measure one or more physical properties of water or atmosphere. Using a multi-scale data assimilation model, the CORIE modeling system integrates models and field controls to produce a simulation of 3D circulation, in a region centered in the estuary and plume. The CORIE data assimilation framework [1] combines observational data with numerical data models to produce an estimated system state for the physical process. To augment the fixed observational network, additional data is collected during periodical cruises of a mobile sensor station. Because these cruises are expensive and rare, an important goal for scientists is to sample data at points that most reduce the uncertainty of the data assimilation model. This is challenging, since the estuary environment is very dynamic, and therefore the optimal cruise path cannot be determined in advance. The goal of our system is to move the mobile station as the data assimilation proceeds in order to maximally reduce the uncertainty in the data assimilation process.

In this work, we propose an *adaptive sampling* algorithm to guide a mobile cruise to collect data that reduces uncertainty in the estimation. Using the linear observation feature in the circulation model, the algorithm generates in advance observation matrices for all points (grid point) that model the CORIE physical map. During the estimation, the algorithm searches for the point that can reduce the most uncertainty. The algorithm is fast and can reduce the uncertainty by 7% compared to random data collection.

SenSys'07, November 6–9, 2007, Sydney, Australia.
ACM 1-59593-763-6/07/0011

2 CORIE Data Assimilation Framework

The CORIE data assimilation framework is based on the Sigma point Kalman filter [2] on reduced state space. The dimension reduction is illustrated in the following equation.

$$x_f = PO \times x_s + \mu + \xi \tag{1}$$

where

- x_f is the system state in the full space.
- PO is the projection matrix.
- x_s is the system state in the reduced space.
- μ is the ensemble mean of the full system state.
- ξ is the noise due to the dimension reduction.

Let

- t is the time step after taking the measurement.
- t^- is the time step before taking the measurement.
- $x_s(t^-_{t+1})$ is the predicted system state before taking the measurement.
- w is the weight.
- $\widehat{P}_{xx}$ is the predicted state covariance matrix.
- K is the Kalman gain.
- H_s is the observation matrix in the reduced space.
- $\hat{y}, y$ are the estimated and the true measurement.

The sigma point Kalman filter has two main steps, time update and measurement update. In the time update step, the filter predicts the system state and uncertainty using a dynamic model of the physical process.

$$x^a(t_i) = [\hat{x}^a(t_i), \hat{x}^a(t_i) + \varsigma\sqrt{P^a(t_i)}, \hat{x}^a(t_i) - \varsigma\sqrt{P^a(t_i)}] \tag{2}$$

$$x^x(t_i) = f^s(x^a(t_i), u(t_i), v(t_i)) \tag{3}$$

$$x_s(t^-_{i+1}) = \sum_{i=0}^{2L} w_i^m x_i^x(t_i) \tag{4}$$

$$\widehat{P}_{xx}(t^-_{i+1}) = \sum_{i=0}^{2L} w_i^m [x_i^x(t_i) - x_s(t^-_{i+1})][x_i^x(t_i) - x_s(t^-_{i+1})]^T \tag{5}$$

In the measurement update, the predicted system state is adjusted based on the actual measurement.

$$y_{t_i} = H(t_i)(PO \times x_s(t_i^-) + \mu + \xi) \quad (6)$$
$$= H_s(t_i)x_s(t_i^-) + n(t_i) \quad (7)$$

$$y^x(t_i) = Hx^x(t_i) + n(t_i) \quad (8)$$

$$y(t_i^-) = \sum_{i=0}^{2L} w_i^m y_i^x(t_i) \quad (9)$$

$$\hat{y}(t_{i+1}) = \sum_{i=0}^{2L} w_i^m y_i^x(t_i) \quad (10)$$

$$\widehat{P}_{yy}(t_{i+1}^-) = \sum [y_i^x(t_i) - \hat{y}(t_{i+1}^{)}][y_i^x(t_i) - \hat{y}(t_{i+1})]^T \quad (11)$$

$$\widehat{P}_{xy}(t_{i+1}^-) = \sum [x_{i+1}^x(t_i) - x(t_{i+1}^-)][y_i^x(t_i) - \hat{y}(t_{i+1})]^T \quad (12)$$

$$K(t_{i+1}) = P_{xy}(t_{i+1})P_{yy}(t_{i+1})^-1 \quad (13)$$

$$x_s(t_{i+1}) = x_s(t_(i+1)^-) - K(t_{i+1})(y(t_{i+1}) - \hat{y}(t_{i+1})) \quad (14)$$

The predicted covariance matrix is calculated as follows

$$\widehat{P}_{xx}(t_{i+1}) = \widehat{P}_{xx}(t_{i+1}^-) - K(t_{i+1})\widehat{P}_{yy}(t_{i+1})K(t_{i+1})^T \quad (15)$$

The trace of $\widehat{P}_{xx}(t_{i+1})$ tells us how uncertain the estimated system state is. The larger the trace, the more uncertain the estimated system is. We would like to reduce the uncertainty in the estimation by minimizing the trace of $\widehat{P}_{xx}(t_{i+1})$.

3 Problem Formulation

Since the observation is linear time variant, we can substitute

$$y_{t_i} = H(t_i)(PO \times x_s(t_i^-) + \mu + \xi) \quad (16)$$
$$= H_s(t_i)x_s(t_i^-) + n(t_i) \quad (17)$$

in the equation 15 above to achieve

$$P_{xx}(t_i) = \widehat{P}_{xx}(t_i^-) - \widehat{P}_{xx}(t_i^-)H_s(t_i)^T \quad (18)$$
$$[H_s(t_i)\widehat{P}_{xx}(t_i^-)H_s(t_i)^T + R]^{-1}H_s(t_i)\widehat{P}_{xx}(t_i^-) \quad (19)$$

We can formulate the problem of adaptive sampling to reduce estimated uncertainty as

$$\max_{H_s(t_i)} trace(\widehat{P}_{xx}(t_i^-)H_s(t_i)^T \quad (20)$$
$$[H_s(t_i)\widehat{P}_{xx}(t_i^-)H_s(t_i)^T + R]^{-1}H_s(t_i)\widehat{P}_{xx}(t_i^-)) \quad (21)$$

4 Algorithm

The model for the Columbia river consists of a set of grid points. The observation function from each grid point is constant and linear in the system state. Hence, we can search all the grid points and calculate the estimated uncertainty without doing the estimation. The algorithm has two phases

Offline We generate the observation matrices for all grid points. These matrices are not changed. Hence, we can store them in advance.

Online In the data assimilation, before taking a new measurement, we loop through all the grid points, merge the observation matrix H_s and calculate the trace of the predicted covariance matrix. The point that results in the lowest trace will be the sampling point of the next time step.

5 Results

Figure 1 shows the traces of the covariance matrices of our sampling scheme and the random sampling scheme, where the cruise randomly move to a location to take a measurement. The simulation is based on data collected over two days. The error trace of our sampling scheme stays lower than the other most of the time. On average, the uncertainty

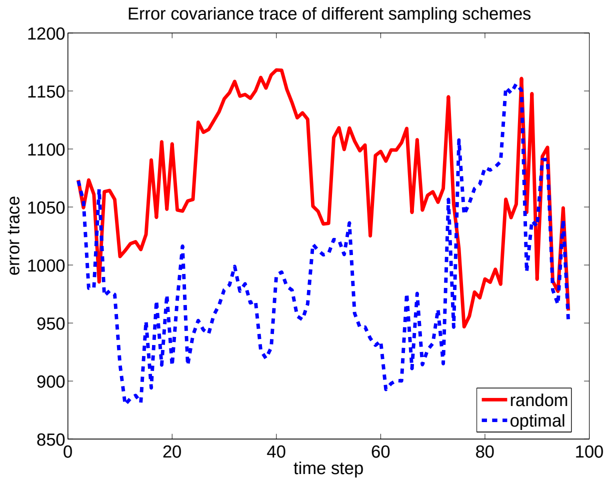

Figure 1. Error traces of different sampling schemes

of our adaptive sampling scheme is reduced by 7% compared to the random sampling scheme. The actual computation time to do the search is only *15* seconds.

6 Future Work

In the future, we would like to consider the correlation between the estimated covariance and the true covariance to develop a sampling scheme that reduce both uncertainty and error. We also would like to extend the algorithm to multiple cruises.

7 Acknowledgements

The research described in this paper was supported by National Science Foundation grants NSF 05-14818 and 01-21475.

8 References

[1] S. Frolov, A. Baptista, Z. Lu, R. V. D. Merwe, and T. Leen. Fast data assimilation using a nonlinear kalman filter and a model surrogate: an application to the columbia river estuary. *Ocean Modelling (Submitted with revision)*, 2007.

[2] R. V. D. Merwe. *Sigma-point kalman filters for probabilistic inference in dynamic state-space models*. PhD thesis, 2004. Supervisor-Eric A. Wan.