Chapter Plan: Information Content

Ewan Pinnington

February 2, 2016

1 Introductory material

- Discuss observability of system. Consider observability of the linearised problem by computing $\hat{\mathbf{H}}$ and calculating the rank for a given number of observations, then comparing this to the length of the augmented state vector. Comment that for state estimation with DALEC1 the system is observable by only assimilating Net Ecosystem Exchange (NEE) observations. This is not the case for joint parameter and state estimation with DALEC2 (check this!).

- Definition of information content measures:

    - Shannon information content [Rodgers et al., 2000]
    - Degrees of freedom for signal [Fowler and van Leeuwen, 2011]
    - The influence matrix [Cardinali et al., 2004]
    - Adjoint sensitivity [Langland and Baker, 2004]

- How measures are implemented with data assimilation system introduced in earlier chapters for a time series of observations. We can consider the information content in a time series of observations by using the $\hat{\mathbf{R}}$ and $\hat{\mathbf{H}}$ matrices in the normal definitions of information content.

2 DALEC1 information content analytic expressions

- Derivation of analytic equations for Shannon information content and degrees of freedom for signal using the DALEC1 data assimilation system, set up for state estimation. Discuss what information content is a function of for NEE (temperature, gross primary production, observation and background standard deviations, etc.). Extend the analytic expressions for information content for multiple observations of NEE in time. This will allow us to see the impact on information content of including a correlation in $\hat{\mathbf{R}}$ by specifying the off diagonal term.

- Show table displaying values of information content calculated from analytic expressions against values calculated numerically in Python. This will give us confidence that our numerical implementation of the measures is correct.

3 Information content with the DALEC2 model

- Repeat work with DALEC1 from previous information content report using the DALEC2 data assimilation system for joint state and parameter estimation. Investigate if we have

similar results with DALEC2 when using the measures from the report (Shannon information content, degrees of freedom for signal). Previous results had shown information content varying seasonally for NEE (as expected from 2) and that information content tends to a limit as successive observations of the same type are added to the assimilation in time.

- Investigate the effect on information content when using a parameter set describing an evergreen forest vs. a parameter set describing a deciduous forest for DALEC2. Does the added phenology in DALEC2 (especially for a deciduous parameter set) impact on temporal nature of information content found for NEE with DALEC1?

- Extend work to include new measures. The Cardinali et al. [2004] influence matrix measures the sensitivity of the analysis in observation space to the observations. The adjoint technique proposed by Langland and Baker [2004], approximates the sensitivity of a scalar forecast error norm to the observations.

  The Influence matrix will allow us to see the spreading of information in NEE observations when using the correlated observation error covariance matrix defined in the previous chapters.

- Investigate the effect of using correlated error covariance matrices (introduced in earlier chapters) on the information content in observations. Effect: Spreading of information from observations.

  See the effect of using different representations of $\hat{\mathbf{R}}$ with different correlation functions in time and between observation types.

4 Twin experiments with DALEC2

- Determine which parameter and state variables are recoverable when assimilating different synthetic observation types (NEE, soil respiration, phenological parameters, etc.) in a set of twin experiments. Using background standard deviations and analysis augmented state from previous chapter as the truth to generate synthetic observations for assimilation. Perturbing the truth twice to find a background vector.

- Can we accurately predict gross primary productivity and total ecosystem respiration by only assimilating NEE. If not what extra observations/constraints (does using $\mathbf{B}_{corr}$ help) are required?

- Use twin experiment setup to further investigate results from information content measures.

5 Future work

- Investigate effect of miss-specification of error, quantity and time of sampling of observations using twin experiments.

- Investigate information content in NEE and total ecosystem respiration (night time NEE) observations when observations are treated as half hourly, averaged twice daily or averaged daily by using a variable time step in DALEC2. This will involve modifying the DALEC2 model to run at a variable timestep (requires updating the mechanism for autotrophic respiration in the model). Modifying the model in this way will also help when parameterising DALEC2 for thinned/unthinned flux data as it will make more data available to us (we wont have to have a full day of observations in order to calculate a daily NEE value for assimilation).

References

C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 130(603):2767–2786, 2004. ISSN 1477-870X. doi: 10.1256/qj.03.205. URL `http://dx.doi.org/10.1256/qj.03.205`.

A. Fowler and P. J. van Leeuwen. Measures of observation impact in gaussian data assimilation. *Reading University*, 2011.

R. H. Langland and N. L. Baker. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 56(3):189–201, 2004. ISSN 02806495. doi: 10.1111/j.1600-0870.2004.00056.x.

C. D. Rodgers et al. *Inverse methods for atmospheric sounding: Theory and practice*, volume 2. World scientific Singapore, 2000.