WW-CO2-models

EW

February 18, 2016

## Methods (abbreviations likely used in paper already have been used without full defintion)

Seasonal CO2 fluxes of Wascana were calculated based on physicochemical measurements and constants following methods outlined in Finlay et al. (2009). Fluxes were therefore directly related to measures such as pH, conductivity, salinity, and DIC, with pH likely to be the predominant control seasonally given its high correlation with annual CO2 flux (Finlay et al. 2015).

Generalised additive models (GAMs) were used, firstly, to enumerate the variation in flux accounted for by pH alone, and secondly, to evaluate the relative contribution of lake metabolism correlates (O2, DOC, GPP, TDN, chl *a*) and climate indices (Pacific Decadal Oscillation, PDO; Southern Oscillation Index, SOI) to seasonal variation in pH. All numerical analyses were performed using R (Version 3.2.3; R Core Team 2015) with the mgcv package (Version 1.8-10; Wood 2011).

Predictor variables were chosen based on availability (minimal missing observations), previously identified relationships with CO2 flux and pH (Finlay et al. 2009; 2015), and independence (minimal covariation). Variables that were right-skewed (TDN, chl *a*, DOC) were log-transformed (base 10). Owing to the potential for an interactive effect of SOI and PDO (Shabbar & Yu 2012), the two terms were modelled as a bivariate smooth using a tensor product spline. In order to explicitly model seasonal, not annual, patterns in the data, the year of each measurement was added to the model as a random effect. The double penalty method of Marra & Wood (2011) was used for model selection. In the double penalty method, as well as the usual penalty to promote smoothness, an additional penalty term is applied to smooths which penalises the linear part of the spline. The effect of this is that redundant model terms can be effectively removed entirely from the model. The double penalty method of model selection has good performance relative to other model selection methods in tests with simulated data (Marra & Wood, 2011).

## Results

* Numbers for a table: NS = not significant
  + CO2 ~ pH (p < 0.001): DEVIANCE EXPLAINED 95%
  + pH ~ DOC (NS), TDN (NS), chl *a* (p < 0.001), O2 (NS), PDO-SOI (p < 0.001), GPP (NS): DEVIANCE EXPLAINED 39%

As expected, pH accounted for almost all of the variation in CO2 flux (95%) in Wascana with a strong negative relationship (Fig. [co2model]). Only two (chl *a*, PDO-SOI) out of six predictors of pH were significant (p < 0.001). TDN, O2, and GPP were shrunk from the model, and DOC had a small negative effect on pH (not shown). Chl *a* had a strong positive effect on pH (> 1 unit-change; Fig. [phmodel]), though above 50 that rate of change in pH with increasing Chl *a* diminished. Climate (PDO & SOI) had a slightly lesser predictive effect on pH (< 1 unit-change; not shown).

## Full references to stuff Peter may not have in the paper yet

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Amir Shabbar & Bin Yu (2012) Intraseasonal Canadian Winter Temperature Responses to Interannual and Interdecadal Pacific SST Modulations, Atmosphere-Ocean, 50:1, 109-121, DOI: 10.1080/07055900.2012.657154

Marra, G. & Wood, S. N. (2011) Practical variable selection for generalized additive models. *Computational Statatistics and Data Analysis* 55:2372–-2387

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1):3-36