**2017/07/20**

In Remy’s version of MyLake model

* Uncommented lines 143-241 in MyL-application.m
* Imported modelinputs\_v12.m into v12\_1 folder

Raoul suggested I take the working v1 model for the Vansjo example and modify it for Lake 227. So, I referred to the user manual for MyLake v1.2 and set up a folder on my desktop that mimics the MyLake\_public\_master repository in GitHub (KRS ELA Model). This includes a folder for v12 containing the MyLake script files, a folder containing the air\_sea toolbox, and a folder for the specific L227 application that contains init, input, and param files for L227. I have taken these from the IO folder in the ELA\_MyLake repository and have ensured that they match the Vansjo example (column headings, formatting, etc.). I duplicated the working code for running MyLake for Lake Vansjo and put it into the L227\_application folder as well. I replaced path names and files to match the corresponding files for the L227 model and commented out the “observed” data from Vansjo (including script within figures that calls out observed data). Once the model is up and running, I plan to add in observed data from L227. The reason I haven’t done this yet is because the observed data for Vansjo and L227 are not in the same format (i.e., I will need to either re-format L227 spreadsheets or amend script files to fit the L227 spreadsheets).

I got the model to run (!!!), but the outputs in the graphs will require comparisons to observed data. I will start on this next time.

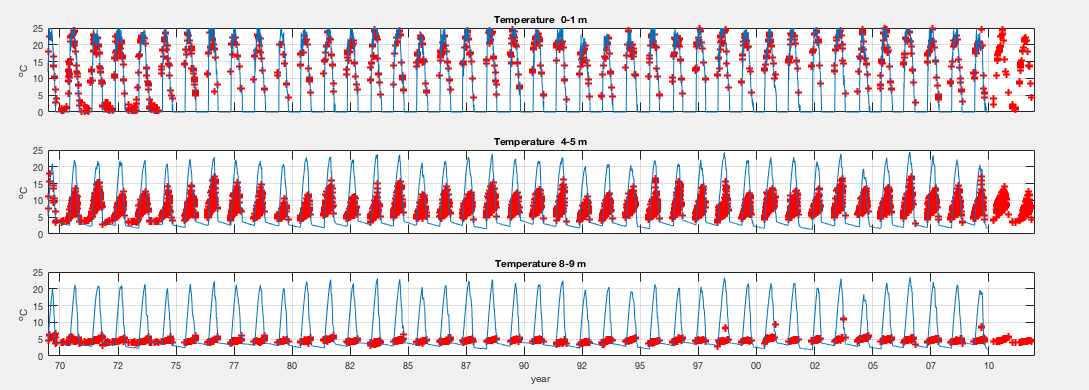
**2017/07/22**

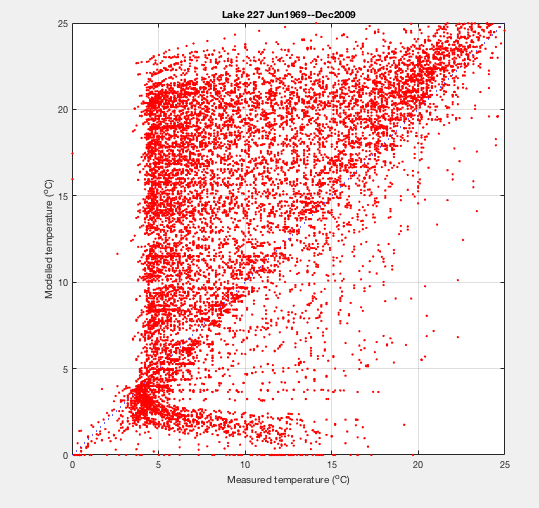
The first priorities for the model are to get the physical parameters correct: lake hydraulic residence time, temperature, ice cover, and wind. It will thus be key to get the input file correct, as this is the file from which the outputs are generated (along with the previous day’s outputs). I see that some of the variables written into the input matrix (e.g., inflow volume) are made by multiplying another variable (e.g., precip) by a scaling factor. This might be a good place to target changes to get the water level and temperature right.

Today, my goal is to generate observation matrices for physical parameters that can be compared to modeled outputs. I will start with temperature profiles.

I started by looking at the vansjotemp file, which has four columns: the first is a date column formatted as yyyymmdd, the second I think is time of day (the model doesn’t use this column), the third is the depth in the water column, and the fourth is temperature. I took the L227 temperature profile file from the historical data and made a new file in .xls and .txt called “L227temp” and put it into an Observations folder. I un-commented the lines of code that use observed temperature and updated the source file code. I also changed the code for figure 22 to display temperature at 4-5 m depth and 8-9 m depth rather than 10-11 and 30-31 m depth (Vansjo example comes from a deep lake). The model now runs, and I am able to compare modeled vs observed temperatures, visualized in figures 2, 3, 4, and 22. The model now runs between m\_start =[1969,6,27] and m\_stop=[2009,12,31]. It ran into an error (“undefined function or variable ‘z0’”) when I put m\_stop=2011,12,31].

By a visual comparison, the model predicts surface temperature pretty well, but has issues with the metalimnion and hypolimnion. Perhaps this has something to do with the parameters that control diffusion?





I’m not sure how I will model the water level or hydraulic residence time, as this is not listed as an output for the model. I know getting the water volume mass balance right will enable getting the nutrients right, so this is a question I will need to ask Raoul or Jason.

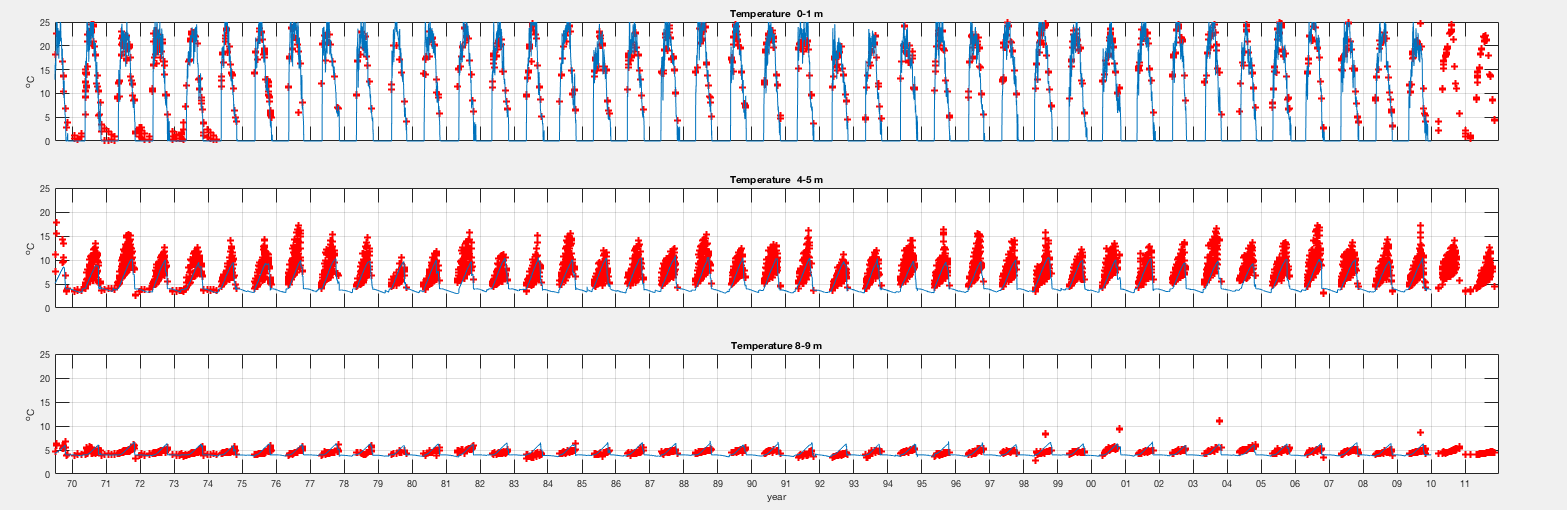
I noticed that the input file for the L227 does not contain any chemistry/biology for the inflows (passive and sedimenting tracer, total P, dissolved organic P, chl a). This will need to be taken from the Lake 239 inflow data (NW and NE inflows, perhaps averaged) and written into the matrix for the inputs (using the ImportInputs script in the Input\_generator folder from Remy and written to the IO folder). This may be tricky, since the rest of the inputs are taken from the climate dataset, which provides daily measurements of each variable. Scaling the inputs based on when they were measured will require some interpolating across the dataset (transferring weekly or biweekly measurements into daily estimates). This can wait for now.

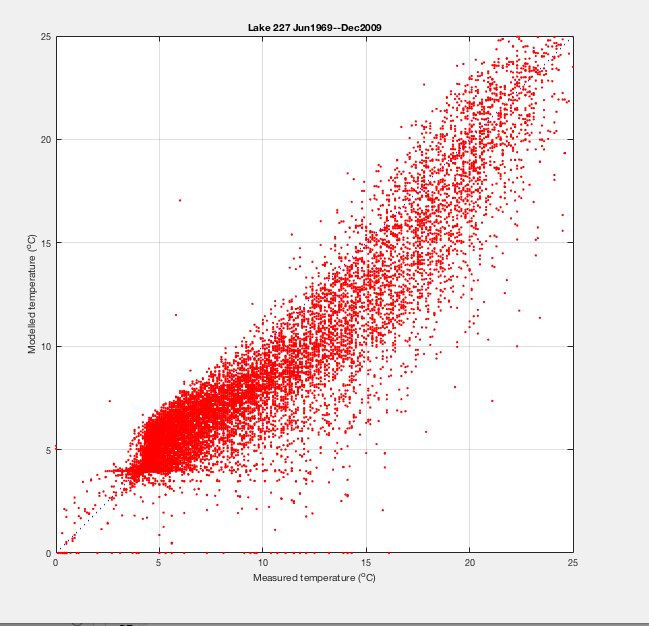
**2017/07/24**

Today, I will try to get the temperature for the MyLake model for L227 to match the observed temperatures. Some key items that could contribute to temperature profiles:

In parameter file:

* I highlighted parameters that differ from the Vansjo parameter file.
* I\_scT (scaling coefficient for inflow temperature)
  + Initially set at 0, can be calibrated.
  + I tried switching this parameter to 1, which visually didn’t seem to improve temperature prediction
  + I tried switching this parameter to 5, which visually didn’t seem to improve temperature prediction
  + I set this parameter back to 0
* I\_scV (fraction scalar for volume inflows)
  + Raoul informed me that this is initially set at 1. I discovered that it was set at 0.1. This may be a mistake, or it might be due to the fact that we are using L239 inflow data for L227 (perhaps the scaling factor of 0.1 is based on the watershed area?).
  + I set the parameter back to 1 and ran the model. As the model was running, the command window displayed “Large inflow!!” several times each year. This seemed to make modeled temperatures more erratic, but the high modeled temperatures in the hypolimnion haven’t gone away.
* Kz\_ak (diffusion parameter for open water periods)
  + I noticed that this was set so 0.0322, which is exactly double what is set at for the Vansjo example.
  + Using equation 18 in the MyLake v1.2 user manual, I calculated that for L227, with an area of 0.05 km2, this number should be 0.0807.
  + I tried switching this parameter to 0.0807, which didn’t seem to improve temperature prediction.
  + I then saw that the instructions indicate that if NaN is entered for this parameter value, ak will be calculated from lake surface area. I tried doing this, and the predictions for temperature in the lower water column improved.





I plan to fiddle around with a few more temperature-related things next time, but this is looking very good. Next on the list:

* Adjust inflow temperature-related parameters 🡪 effect on modeled temperatures
* Adjust snow- and ice-related parameters 🡪 effect on modeled temperatures, ice on and ice off
* Adjust wind-related parameters 🡪 effect on modeled temperatures (mixing)
* Adjust inflow volume and associated parameters 🡪 effect on water level, hydraulic residence time

**2017/07/26**

Today, I wanted to optimize snow and ice. I ran the model with the parameters from 2017/07/24 and compared modeled ice-on and ice-off dates with observed ice-on and ice-off dates from L239. In the documentation for the ELA datasets, it states that L239 is the only lake that has continuous records of these dates (thus we use them to approximate L227 dates). It also says that lakes in the area with smaller surface areas and less depth usually have ice-on and ice-off dates a few days to a week in advance of those observed on L239.

From the model run, I found that with an inflow volume coefficient of 1:

* Modeled ice off occurs 11 ± 6 days later than on L239
* Modeled ice on occurs 9 ± 7 days earlier than on L239

With an inflow volume coefficient of 0.1:

* Modeled ice off occurs 19 ± 6 days later than on L239
* Modeled ice on occurs 12 ± 6 days earlier than on L239

I noticed that the parameters for melting ice and snow albedo were set to 0.6 and 0.9, respectively. This differs from the MyLake documentation, which specifies 0.3 and 0.77 as default values, respectively. I changed these to the default values and ran the model again (with an inflow volume coefficient of 0.1).

* Modeled ice off occurs 8 ± 5 days later than on L239
* Modeled ice on occurs 11 ± 6 days earlier than on L239

Setting the albedo to the default improved the estimate of ice off to about a week later than L239. Given that this lake is sheltered, this makes sense to me. Since we don’t know the actual dates of ice off for L227, this estimate seems fine for now. Note: changing ice and snow albedo didn’t change the estimate for ice on.

**2017/07/27**

Today, I read through Remy Buoyssou’s thesis, with an eye to the way inflow to the catchment was calculated. In there, it specifies that inflow is calculated as the precipitation (mm) x catchment area (ha).

Relevant parameters:

* Precipitation measured at met site (mm)
* Watershed area L227 = 49 ha
* Watershed area L239 (inflow parameters other than volume come from here) = 335 ha

I looked through the L227 input file, and I noticed that the inflow (m3 d-1) is calculated for the L239 watershed area, not the L227 watershed area. This explains why I was seeing “Large inflow” warnings when I had the inflow volume scaling parameter set to 1.

I went through and multiplied the precipitation by the L227 watershed area rather than the L239 watershed area in the input file, and I set the I\_scV parameter back to 1 in the parameter file. This essentially says that all the precipitation hitting the ground in the L227 catchment makes it to the lake. Running the model with these inputs and parameters yielded only 3 “Large inflow” warnings from 1969-2009.

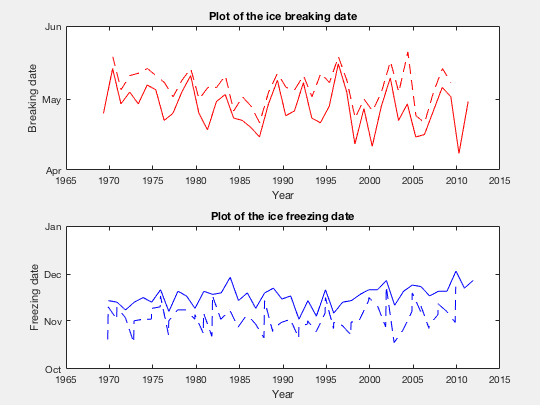
However, not 100% of the precipitation in the catchment reaches the lake, so the I\_scV parameter must be scaled to reflect this. In the “L227 hydrology” document, a regression between precipitation and L227 outflow is presented (R2 = 0.68, N = 28):

L227 runoff (mm) = 0.7359 \* precipitation (mm) – 281.25

Given this relationship, it is expected that 73.6 % of precipitation leaves the lake as outflow. If we assume outflow = inflow - evaporation, we can apply 0.7359 as a scaling parameter for inflow volume (I\_scV). This is not perfect, as it ignores evaporation, but it will be better than using 1 for now.

I found the script file in Remy’s documentation that generates matrices and comparison figures for observed ice break and freeze and modeled ice break and freeze. I edited the code to run with my files, and I stored the script files (Iceplot.m and PerformanceIce.m) in a new folder entitled “Model Output Evaluations.”

With the new changes to parameters (I\_scV = 0.7359) and inputs (inflow scaled for L227 rather than L239), here’s what the comparisons of modeled (dashed) vs. observed (solid) ice dates look like:



BreakModel-BreakObs = 8.350 ± 4.721

FreezeModel-FreezeObs = -11.025 ± 6.479

I added 1.8 degrees to the inflow temperature (I\_scT = 1.8, not 0). This improved freeze and break estimates.



BreakModel-BreakObs = 7.875 ± 4.592

FreezeModel-FreezeObs = -10.825 ± 6.320

I added 1.8 degrees to the inflow temperature (I\_scT = 5, not 0). This improved freeze and break estimates.



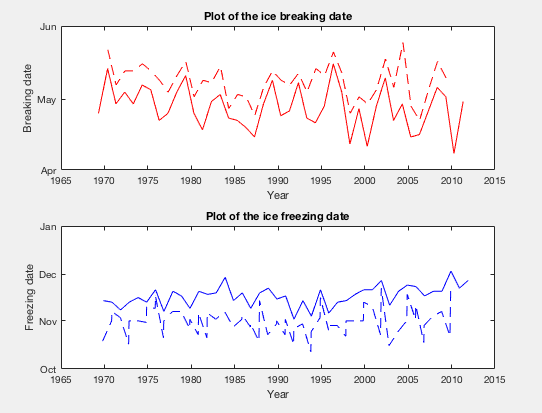
BreakModel-BreakObs = 7.000 ± 4.478

FreezeModel-FreezeObs = -10.525 ± 6.445

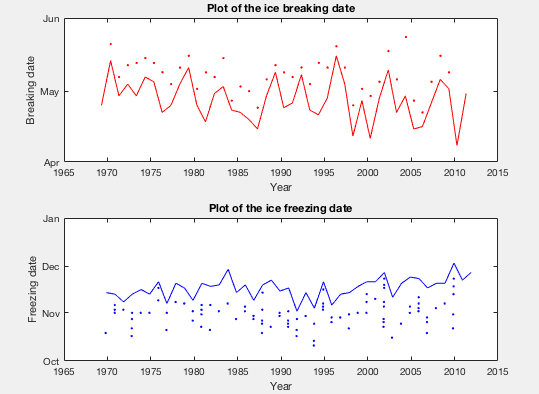
However, I ran the PerformanceIce.m script from Remy, and the rmse for ice break and freeze are 6.9250 and 16.5854, respectively. This is higher than the rmse values in Remy’s thesis (3.78 and 2.08 for breaking and freezing, respectively). This may not be a huge deal – I will ask Raoul and Jason.

**2017/07/30**

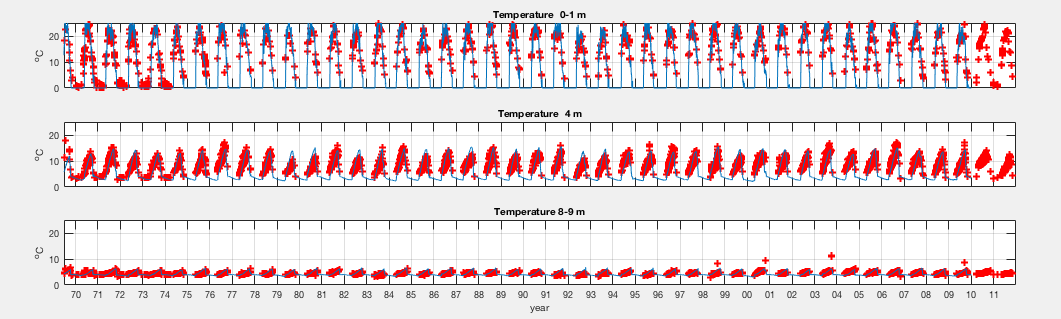
I updated the heatflux\_v12.m script file today. I noticed that on lines 76-80, the script specifies total transmissivity if global radiation data are missing from the input file (they are missing for L227). The file originally had the empirical determination of transmissivity from Beisjo and Vansjo activated (line 79) but had the generic formula commented out (line 80). I switched the two: commented out the Beisjo/Vansjo example (line 79) and activated the generic formula (line 80). However, doing this made the predictions of ice break/freeze worsen. The rmse for breaking and freezing ice is now 10.48 and 19.68, respectively (given an I\_scT of 0). rmse improves slightly (<0.5) when I\_scT is increased up to 5.



However, I noticed that iceplot.m and PerformanceIce.m take ALL modeled dates for ice breaking and freezing in a year, which can happen multiple times in a year (plotted with dots below). Perhaps this comparison would improve if we only took the latest date for ice breaking and freezing in a single year?



I also updated the code for figure 22 in KRS\_model227\_v12.m to display only temperatures measured at 4 m rather than from 4-5 m. This took care of the issue of the metalimnion modeled temperatures plotting cooler than the observed temperatures (modeled temps were displayed from 5 m depth and observed temps were displayed from 4 m depth).



**2017/07/31**

Today, I sent an email to Raoul and Jason regarding the outstanding questions I have about the physical parameters.

In the meantime, I will start to work on the phosphorus part of the model. One issue that arises is that P was actually added to L227 in a continuous fashion, whereas the model needs P inputs to be scaled to inflow volumes. There are two paths possible moving forward:

1. Force P inputs to scale with inflow volumes
2. Change the code so that P inputs don’t need to scale with inflow volumes

Option 1 is obviously easier to do for now. I will go with option 1 and check how this goes compared to the observed P concentrations in the lake.

I made a spreadsheet “Pinputcalcs.xlsx” to calculate the scaled P inputs to the lake. I acquired the amount of P added (kg/y) each year from Schindler et al. 2010 (PNAS) (note: P load through 2009 is the same as 1999-2005). I calculated what the P load would be if the yearly load (May-October) was split proportionally with inflow. From this, I calculated the hypothetical P concentration in the inflows on each day. I then put these values into the Inflow\_TP (mg/m3) column in the “L227\_input\_basin1\_land\_doc\_var\_new\_species\_simplified.xls” file. I also updated I\_scTP as 1/I\_scV in the parameter file, as the actual P inflow is the actual load and should not be scaled.

I set the Min and Max values for I\_scTP and I\_scS to NaN in the parameter file, as I am not sure of what the bounds of these parameters should be.

However, I am getting an error when I run the model that did not appear when Inflow\_TP (mg/m3) was filled with zeroes:

Error using roots (line 27)

Input to ROOTS must not contain NaN or Inf.

Error in solvemodel\_v12>Ppart (line 1086)

Pdiss(w) = max(real(roots([a b c])));

Error in solvemodel\_v12 (line 545)

[Pdz\_store, Psz\_store]=Ppart(VolFrac,TIP\_sed,Psat\_L,Fmax\_L\_sed,rho\_sed,Fstable);

Error in KRS\_modelL227\_v12 (line 42)

= solvemodel\_v12(m\_start,m\_stop,initfile,'lake',inputfile,'timeseries', parafile,'lake');

**2017/08/02-04**

I worked on the inflow P concentrations today. I noticed that when the very large inputs from Remy’s spreadsheet were included, the model ran with no problem. So, I played with the inflow concentrations to see at what values I start to get the error. For Inflow\_TP, values of actual Inflow\_TP \* 3 allowed the model to run (lowest value was 218 mg m-3), but Inflow\_TP \* 2 did not. I will look into the calculation that gives the error and determine why it won’t take values lower than this.

I received some feedback from Raoul about optimizing physical parameters of the model, among other things. Below are his suggestions, with my notes on progress embedded.

Re: progress so far

“Temperatures look very good. I think we should clarify the issues with ice phenology, then continue with MyLake v1.2 to capture total phytoplankton biomass (for which Chl is a proxy).

Re: errors in ice break/freeze predications

“MyLake usually performs very well for ice, with RMSE of a few days. Remi had a good RMSE, but the parameters he used were out of range and unrealistic. Your parameter set is better and with temperature closely modelled it is a good start. 11 days off is quite a large error, so we should definitely try to improve that as it will affect turnover time, light, oxygen, etc.”

“Also, how good is L 239 as a proxy for L227 regarding ice formation?”

* From Ken Beaty’s notes on ice on/off: “It has been observed that lakes in the area having smaller surface area and less depth are usually a few days to a week in advance of Lake 239 in both respects.” Thus, if modeled ice break/freeze are less than about 7 days earlier than observed, we should be good.

“Have you looked at the timing of turnover, modelled vs measured? E.g., the date at which the lake becomes isothermal in the spring and fall?”

“You can try to look at the wind, a key input that will influence temperature and ice formation. In many cases wind has to be corrected from the weather station to the actual site. Based on your observations on site, is the wind measured representative of the wind affecting the lake?  the C\_shelter parameter modulates the wind sheltering by the topography surrounding the lake. Wind time-series scan also be corrected if necessary.”

Re: model recording multiple freeze dates

“I am curious as to why the data set reports many freezing dates per year, but not many break-up dates per year. The lake must break-up before freezing again? In any case it is true that the final freezing and the final break-up are the target we aim for.”

“You could play with the albedos of ice and snow, if there is reason to believe that e.g., snow absorbs more heat than the default parameterization. A colleague of mine once reported that this had an effect on modelled ice break-up.”

Re: choosing I\_scT

“I am not aware of a “default” value as it is very site specific. To have a more solid estimates of inflow temperature and how it follow air temperature we can use the empirical relationship  provided in the attached  “Erickson, T.; Stefan, H. Linear Air/Water Temperature Correlations for Streams during Open Water Periods. Journal of Hydrologic Engineering 2000, 5 (3), 317-321.” I have used it before. If there is groundwater seepages stream T could be warmer”

Re: lake residence time

“We may have discussed this, but it is not a normal output of the model. You’d have to code it. Further, the model assumes constant volume (I hope that this is indeed the case for L227) so given the outflows and constant volume, residence time can be calculated. However we should just do our QC on state variables it is just as robust.”

Re: moving forward

“I think that the goal would be to have MyLake 1.2 calibrated as much as possible (temp, ice, phytoplankton biomass), ported to MyLake 2 so that is works just as well, and the inputs required for MyLake 2 prepared (the inflowing concentrations of elements). We should talk about phytoplankton dynamics and what we hope to learn by adding iron (or if this is indeed our goal) . The issue is that if observed phytoplankton behaves erratically or is chaotic, then how would we know that adding e.g., Fe limitations actually improve the model ? We could also try an offline statistical model to predict cyanos, then see if Fe helps explain the time series. I feel we should discuss this point together with Jason so that we clearly define our modelling hypothesis.”

“Also, MyLake does not do phytoplankton community structures, it has only 2 pools (2 “species”) both P and light limited. MyLake 2 has the same phytoplankton module as MyLake 1.2. Assessing if Mylake is the right model for the job I part of this testing phase … Can we model total biomass given the complexity of the system (widely varying proportion of different species throughout the years?) Can we find over which time period the phytoplankton model performs better, and understand why?”

**2017/08/04**

I looked through the inflow temperature data, which in Remy’s dataset was calculated as:

Inflow temp = 5 + 0.75\*air temp

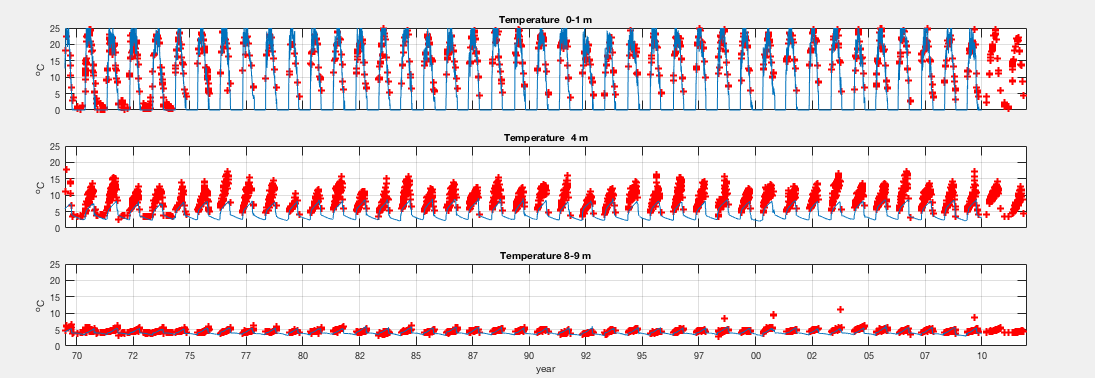
This resulted in (a) inflow temperatures being higher than air temperatures (not likely here) and (b) inflow temps being negative when air temps were negative. Rather than using this relationship, I looked at the Vansjo example and parameterized that relationship instead. When I regressed air temperature by inflow temperature for Vansjo (positive air temps only), the relationship was:

Inflow temp = 0.0201 + 0.9782\*air temp

This brought the RMSE for ice break and ice freeze to 5.52 and 19.92, respectively.

I also updated the code to run with the solvemodel\_v12\_1b function rather than the solvemodel\_v12 function. I started by commenting out the second chlorophyll group and getting the model to run. The issue with phosphorus inputs yielding an error message has gone away. I just now need to add in the P inputs from actual inflows in addition to the experimental additions.

These caused the temperature measurements to worsen in the metalimnion (see images), but the RMSE for ice break and ice freeze improved to 5.67 and 16.80.



However, when I try to add in the parameters for the 2nd group of chlorophyll (lines 145-153), I get an error: “Index exceeds matrix dimensions.” I went into modelinputs\_v12.m to try to debug this issue. In lines 53-55, I put in 26:48 for all Bio\_Par components rather than 26:40. This didn’t make the issue go away though. I will continue to adjust the code to try to get this issue fixed next time.

**2017/08/07**

I tried adjusting the file path for the model, and it worked – no more “index exceeds matrix dimensions” error. I just needed to have the terminus of the path to be KRS\_ELA\_Model rather than L227\_application.

I have a skype meeting with Raoul today. Some items we need to talk about:

* Adding in biological parameters causes temperature profiles to change. I think it will be best to get all of the aspects of the model running and then work from the beginning with the physical parameters.
* I’ve started working with solvemodel\_v12\_1b.m rather than solvemodel\_v12.m. This model adds in a 2nd group of chlorophyll and also frazil ice (plus code rearrangements). Once everything is up and running, this will allow me to put in separate growth equations for N fixers and non-N fixers.
* Ways to adjust ice break/freeze (I am down to RMSE values of 5.67 and 16.80).
  1. Wind sheltering
  2. Wind input series
  3. Inflow temperature
  4. Transmissivity
  5. Ice and snow albedo (this helped!)
* Inflow temperatures: the relationship 5 + 0.75(air temp) did not seem to be accurate (inflow temps warmer than air temp during warm periods, something we would not expect in this system). I have substituted the empirical relationship from the Vansjo example instead. I will also try the empirical relationship presented in Erickson and Stefan of 4.18 + 0.82(air temp) for Minnesota streams (Minnesota is nearby the ELA).
* Questions from Raoul’s email:
  1. “Also, how good is L 239 as a proxy for L227 regarding ice formation?”
     + From Ken Beaty’s notes on ice on/off: “It has been observed that lakes in the area having smaller surface area and less depth are usually a few days to a week in advance of Lake 239 in both respects.” Thus, if modeled ice break/freeze are less than about 7 days earlier than observed, we should be good.
  2. “Have you looked at the timing of turnover, modelled vs measured? E.g., the date at which the lake becomes isothermal in the spring and fall?”
     + Is there code we can insert to report this? I know we get a matrix of temperature by depth each day, but looking through this code would be tedious. I can compare this to the observed data for L227.
     + Note: I can use DoF and DoM from the workspace to look at this rather than Remy’s script
  3. Re: model recording multiple freeze dates: “I am curious as to why the data set reports many freezing dates per year, but not many break-up dates per year. The lake must break-up before freezing again? In any case it is true that the final freezing and the final break-up are the target we aim for.”
     + The model records multiple break ups when there are multiple freezes, but the post-hoc code only allows break up to be observed during the appropriate time of year (spring). Is there any code I can put into the ice script that will allow it to only record the last ice freeze date? This will improve the RMSE a bit.

**Next steps:**

1. Get v2 running
   1. Optimize parameters and inputs for L227
      1. Model should be able to predict temperature, ice on/off, chlorophyll stocks, P concentrations, fluxes at the sediment-water interface
   2. Add complexity only when model fails to predict observed patterns
      1. May need to add in N and/or Fe into the model
      2. Stella model for Fe may help with setting parameters and differential equations
      3. From Raoul: “We should talk about phytoplankton dynamics and what we hope to learn by adding iron (or if this is indeed our goal) . The issue is that if observed phytoplankton behaves erratically or is chaotic, then how would we know that adding e.g., Fe limitations actually improve the model ? We could also try an offline statistical model to predict cyanos, then see if Fe helps explain the time series.”
2. Clarify goals: do we want to demonstrate new processes (e.g., Fe cycling) with the model, or take rates from the literature and test them?
3. Potential option for predicting phytoplankton dynamics: use MyLake to generate physical and chemical variables, then use a post-hoc statistical modeling approach (Bayesian?) to generate
4. Look at observed trends in oxygen, chemical species 🡪 make figures ahead of time in order to look at potential changes to the model

**Potential directions for paper(s)**

1. *Ice record for ELA*: has the timing of ice on/off changed over 5 decades? If so, what are the drivers of this phenomenon, and how does this affect limnology during the ice-free period? Temperature, oxygen, phytoplankton growth, etc.
2. *L&O long-term data special issue*: value of long-term data (present unpublished time series) and the extra layer of knowledge modeling can add. Potential topics:
   1. Internal loads of P and Fe (in hypolimnion and coupling with the sediment) 🡪 role on the blooms
   2. Prediction of complex phytoplankton dynamics
3. *Diagenetic model of sediment-water interactions:* very few studies have demonstrated sediment-water interactions, including reduced chemical species. The model will give fluxes, whereas we have concentrations. Potentially powerful but we need to demonstrate matching between model output and observed trends.

**2017/08/08**

I updated the inflow temperatures in the inflow file to match the empirical relationship reported in Erickson and Stefan (2000) (degrees C):

Inflow temp = 4.18 + 0.82(air temp)

This relationship was derived from Minnesota streams, which will likely make a nice proxy for ELA.

I generated inflow chemistry data from the Lake\_239\_chemistry.csv spreadsheet. I took data collected from NEIFL, NEIFW, and NWIF and averaged the three measurements. Details:

* **Inflow\_TP:** taken from TDP in the chemistry file. This was combined in Pinputcalcs.xlsx to add experimental P inputs to the inflow P inputs. I had to add a factor of 1/0.7359 to the experimental P inputs so that they would scale correctly to the inflows (inflow volume is scaled by 0.7359, but experimental P inputs should not be).
* **Inflow\_DOP, Inflow\_Chla, Inflow\_CH4, Inflow\_Al3, Inflow\_SiO4,** and **Inflow\_diatom** were all either absent or nearly absent from the L239 chemistry dataset. These were all set at 0 throughout the observation period. I’m not sure if it would be better to set these at 0 or as NaN.
* For all other variables, dates that were available are marked with numbers, and dates that were not available are marked with NaN.
* **Inflow\_NO3** is the combination of NO3- and NO2- in the inflows.
* **Inflow\_Fe2**and **Inflow\_Fe3­** were estimated from sparse Fe data. In the L239 chemistry spreadsheet, there are data for total Fe and ferrous Fe. The ferrous Fe dataset was much more sparse than the total Fe. What I did was divide the ferrous concentration by the total Fe concentration for the dates that had both and averaged this proportion (0.819). Inflow\_Fe2 was thus calculated as 0.819\*total Fe, and Inflow\_Fe3 was calculated as (1-0.819)\*total Fe.
* **Inflow\_SiO2** is taken as SRSi in the L239 chemistry sheet (i.e., I presumed that soluble silicate is SiO2 rather than SiO4).

Some of these variables won’t be used until the sediment model is up and running. I added the relevant variables into the input file for MyLake v1.2.1 today, and it ran with no errors.

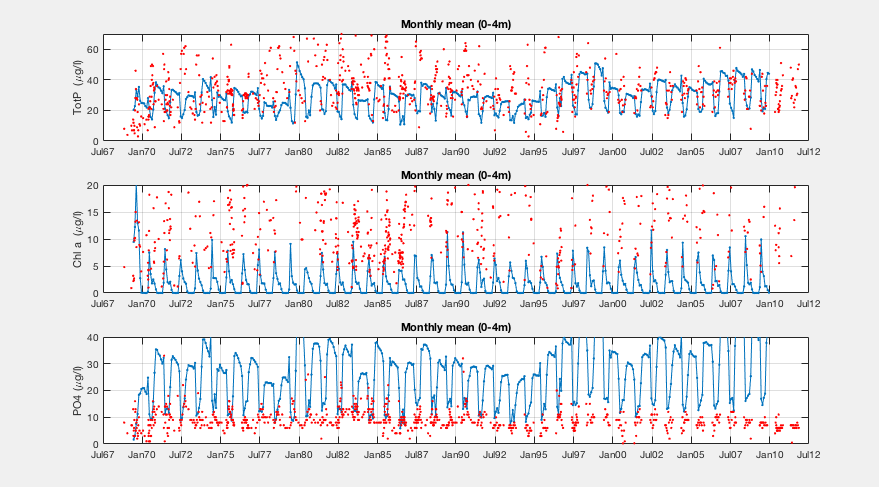
**2017/08/09**

I made observation files for dissolved P, total P, and chl a to match the Vansjo example. This will allow for the comparison between model outputs and observations in the system. I took the data from the Lake 227 chemistry dataset. Concentrations are taken from integrated epilimnion samples.

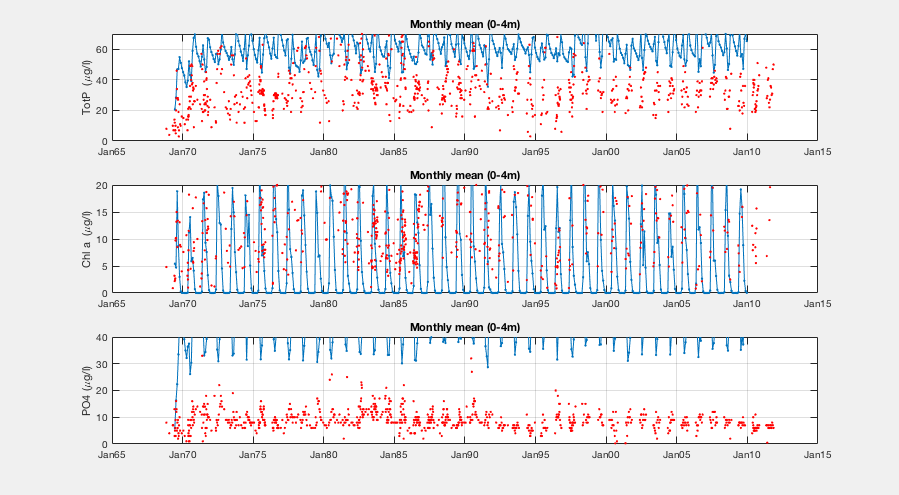
I made the following files:

* **L227\_TP\_chla.xls**: TP was calculated at TDP + suspended P.
* **L227TDP.txt**: PO4 was approximated as TDP.

The model runs without error and I can now compare observations to outputs. Total P is approximated fairly well, but Chl a and dissolved P are not.



I also updated the parameter file for the two pools of phytoplankton. From *Freshwater Ecology* (textbook by Dodds and Whiles 2010), the Ks values (concentration when uptake rate is half of max uptake rate) for phytoplankton and cyanobacteria are 0.0357 mg/L and 0.5 mg/L, respectively. Here’s what the output looks like now. Chl a is now approximated well, but total P and dissolved P are highly overestimated.



Next it might be good to update growth rates at 20 degrees C for the two groups. *Freshwater Ecology* has max growth rates, and I might be able to look this up in the other papers I have gathered for the two phytoplankton pools.

**2017/08/10**

I’m looking up phytoplankton growth parameters today to input into the model for the two phytoplankton groups under P limited growth.

**Growth parameters needed for phytoplankton species in MyLake – phosphorus**

Ks ½ saturation P concentration for growth (mg m-3 = μg L-1)

μmax maximum specific growth rate @ 20° C (d-1)

\*I suspect Dodds and Whiles may be off by a factor of 1000. They report Ks in units of mg L-1, but the conversion to mg m-3 results in abnormally high values compared to the other studies.

**Taxa Species Ks μmax Study**

Phytoplankton 35.7 0.222 Dodds & Whiles 2010

Diatoms 65 0.262

Green algae/Chlorophytes 280 0.317

Cyanobacteria 500 0.585

Chlorophyte *Scendesmus* 0.4830.630 Grover 1991 (12° C)

Chlorophyte *Chlorella* 0.067 0.811

Cyanobacteria (Nfix) *Anabaena* 11 0.792 Nobel et al. 1997

Cyanobacteria (Nfix) *Aphanizomenon flos-aquae* 21 0.600

Cyanobacteria *Prochlorothrix* 0.015 0.600

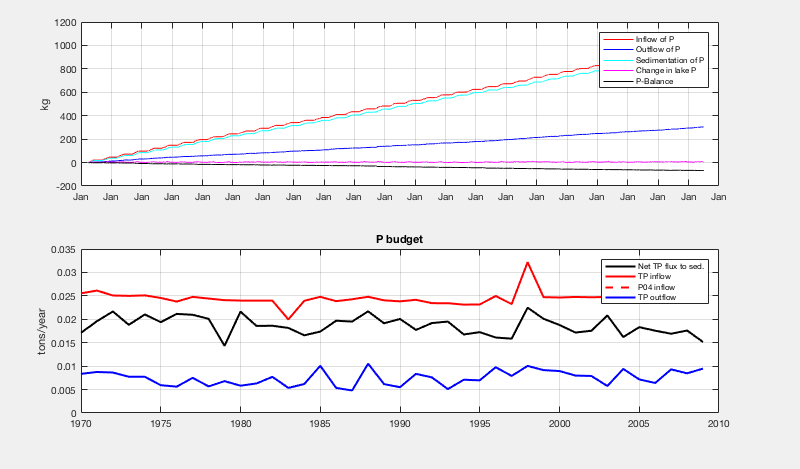
Cyanobacteria (no Nfix)*Planktothrix* 0.17 0.864 Ducobu et al. 1998

One important note: the model looks like it determines phytoplankton (chl a) stocks based on Monod growth relationship. However, Grover (1991) found that the Droop relationship outperformed Monod under non-equilibrium situations. Further, Monod doesn’t account for the possibility of P uptake outpacing P cellular needs (luxury uptake). We may want to consider re-coding the model for Droop rather than Monod if we aren’t able to reproduce phytoplankton stocks.

**2017/08/18**

I checked the input file to make sure the inflow P concentrations are correct. In Pinputcalcs.xlsx, the dissolved P inputs from the experimentally added P is 980.3 kg (1969-2009). This matches with the historical records of how much P was added (985.1 kg; Schindler et al. 2008). The estimated amount of dissolved P from the catchment is 27.6 kg. The relative amounts of these different inputs seems well in line for what we would expect for the lake (i.e., experimental P inputs greatly outweigh catchment inputs). Note: catchment inputs are multiplied by 0.7359 to account for the volume of water that doesn’t reach the lake, whereas the experimental P inputs were not scaled. To do this, I divided experimental P inputs by 0.7359 for the input file so that they could be multiplied by 0.7359 in the parameter file to achieve a final scaling factor of 1.

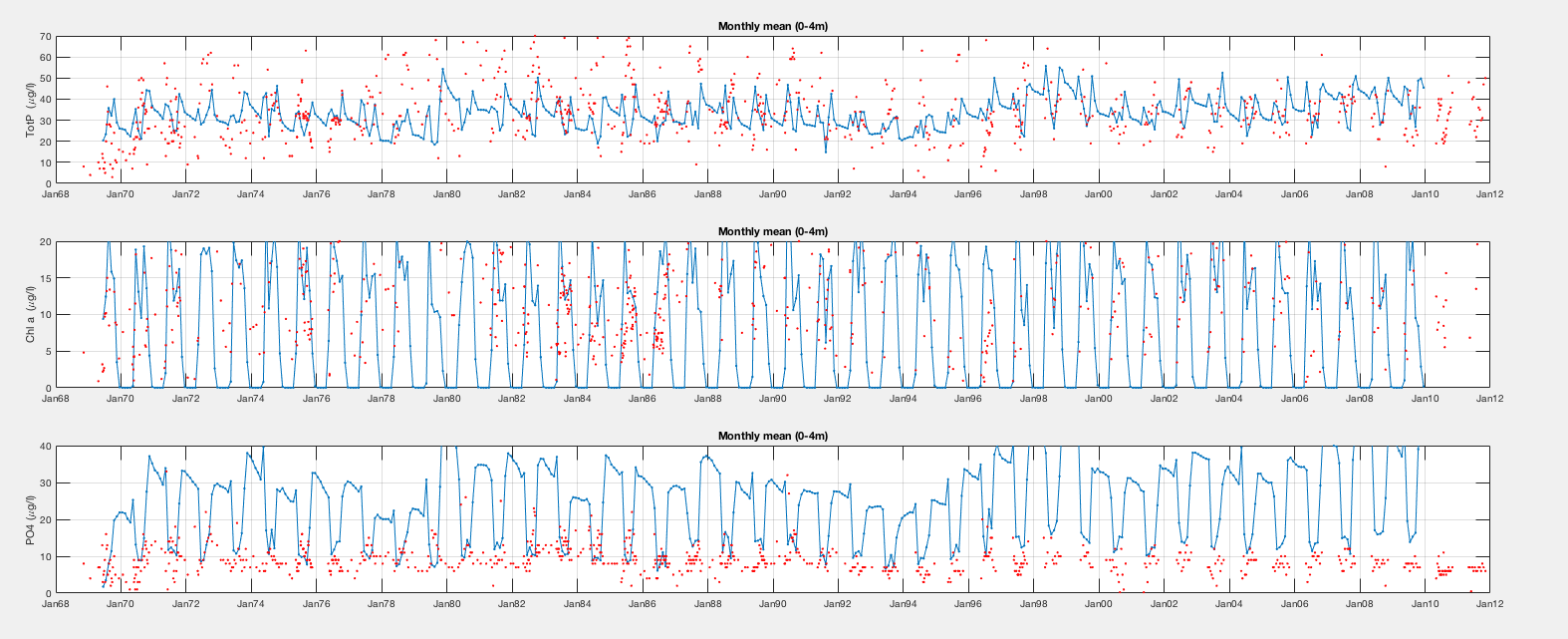
In the model output, the inflow of P (about 1000 kg) lines up with this calculation, which tells me that the input files were put into the model correctly. The model predicts that most of the P is sedimented, and the rest flows out of the lake.



Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 21

All other parameters set to defaults

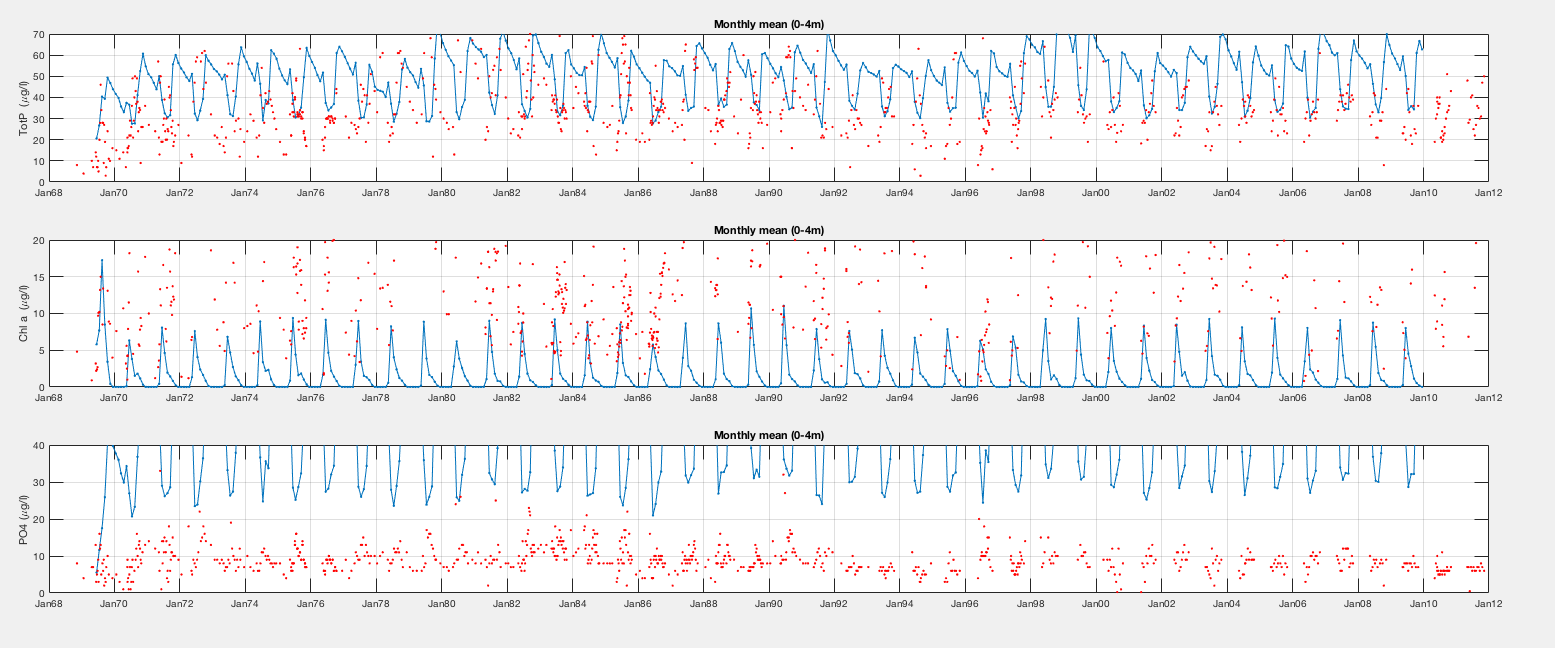
Total P is in the right ballpark (not precise but values straddle the observed values), chlorophyll is well-predicted, PO4 is overestimated.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 21

P\_half2 = 21

All other parameters set to defaults



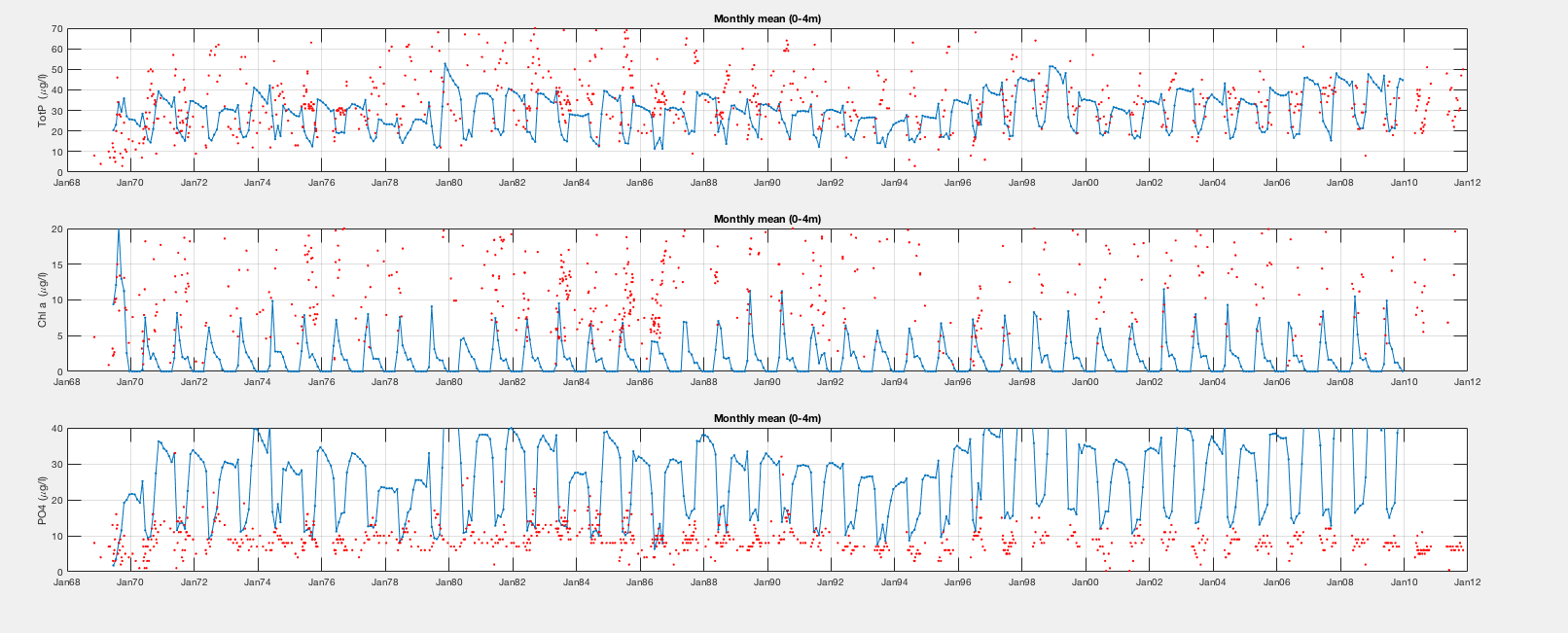
Total P is overestimated, chlorophyll is underestimated, and PO4 is severely overestimated.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 0.483

All other parameters set to defaults



Total P is underestimated, chl a is underestimated, and PO4 is overestimated.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

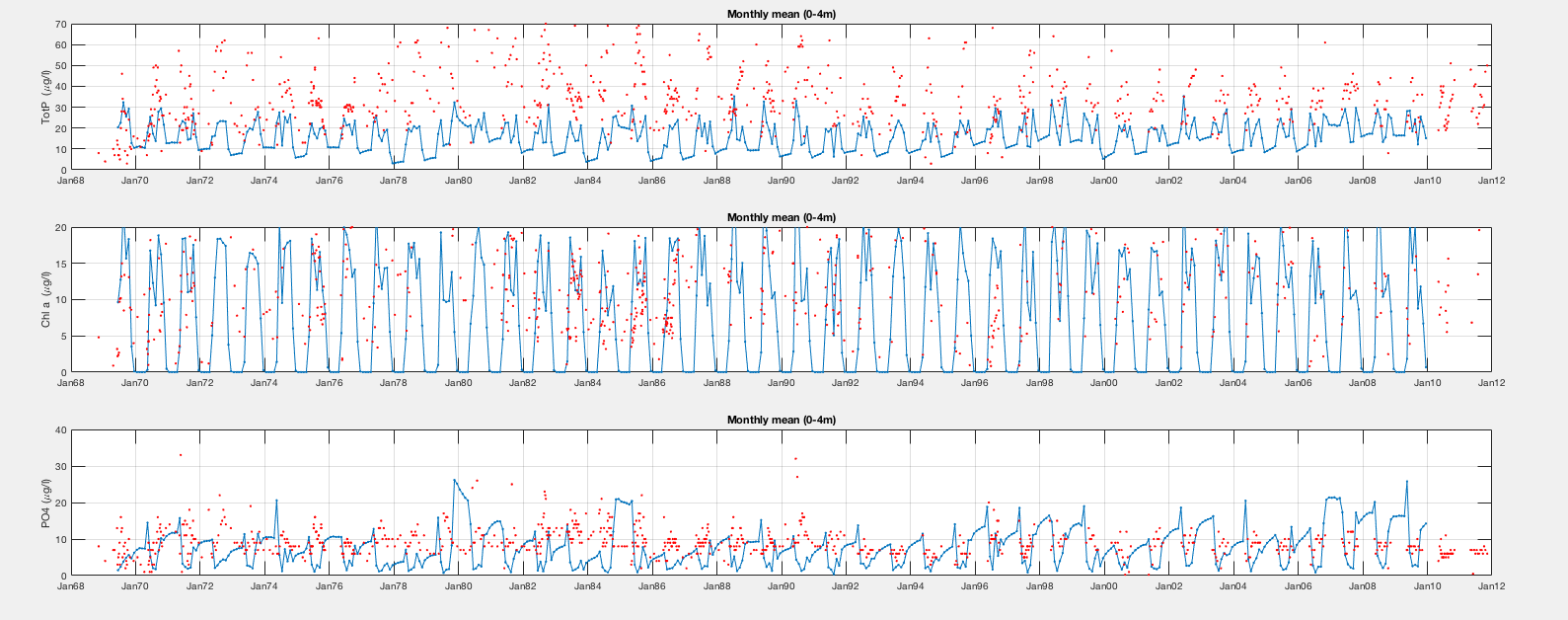
P­\_half = 0.483

P\_half2 = 21

m\_twty = 0.1

m\_twty2 = 0.1

All other parameters set to defaults



Total P is underestimated, chl a is well-predicted, PO4 is (finally) in the ballpark. Conclusion: decreasing the loss rate (m\_twty) decreases the amount of PO4 that goes back into the water column. I imagine this happens proportionally to the growth rate (g\_twty). I think it will be preferable to use g\_twty values from the literature, which tend to be near 0.6 rather than 1.5.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 21

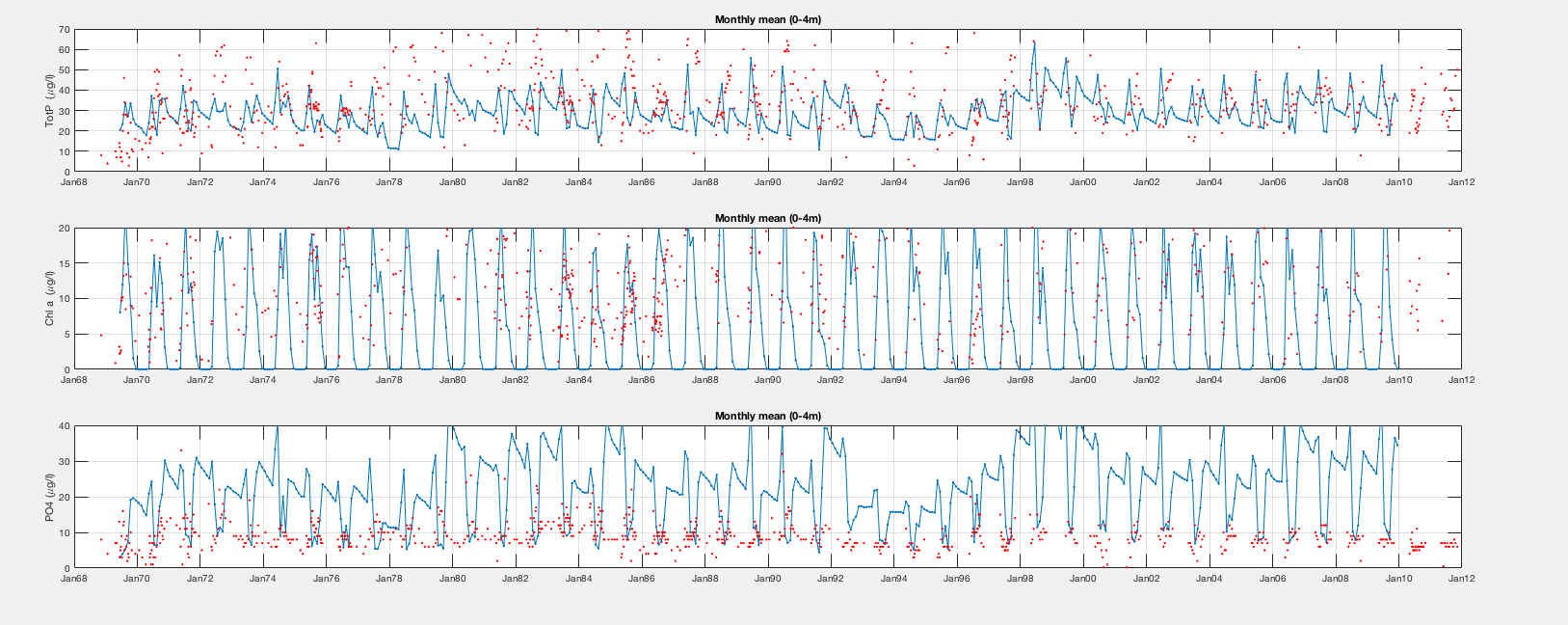
g­­­\_twty = 0.6

g\_twty2 = 0.6

m\_twty = 0.06

m\_twty2 = 0.06

All other parameters set to defaults



Total P is back in the right ballpark, chl a is well-predicted, and PO4 is overestimated.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 21

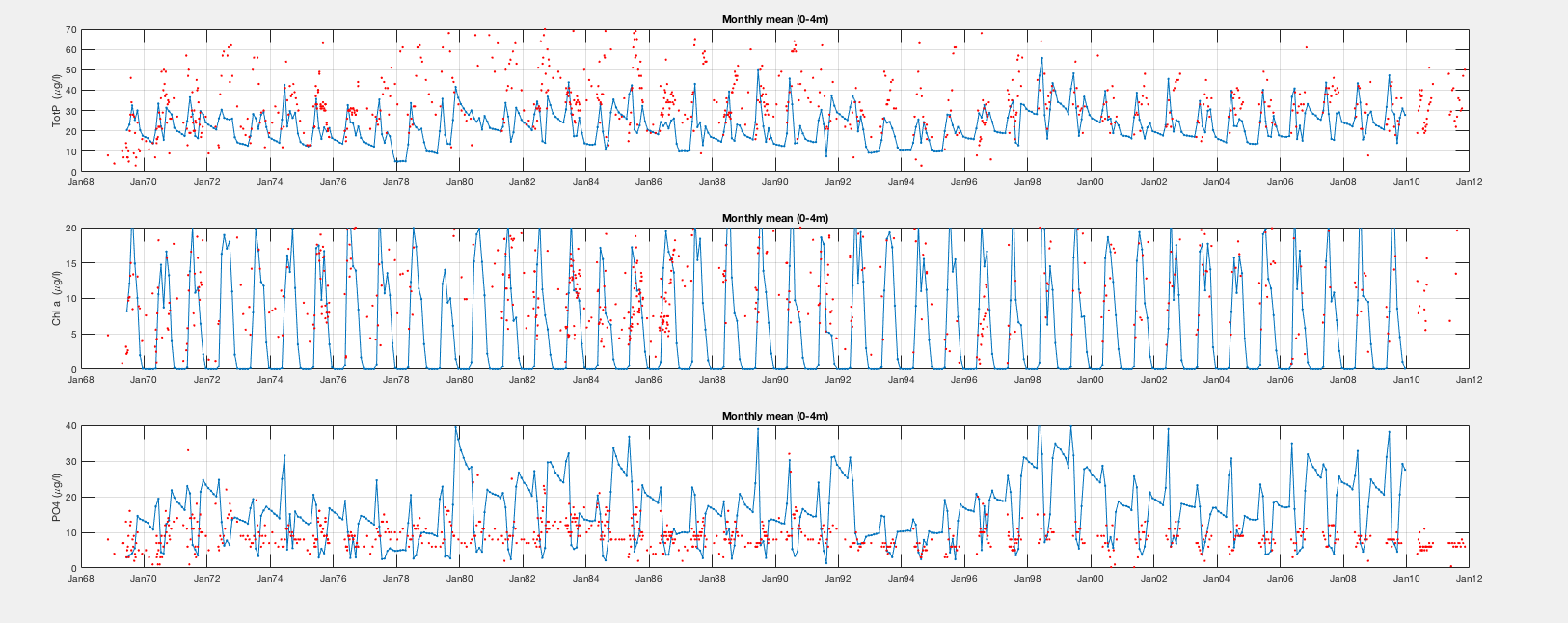
g­­­\_twty = 0.6

g\_twty2 = 0.6

m\_twty = 0.04

m\_twty2 = 0.04

All other parameters set to defaults



Total P is underestimated, chl a is well-predicted, and PO4 is still overestimated.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 21

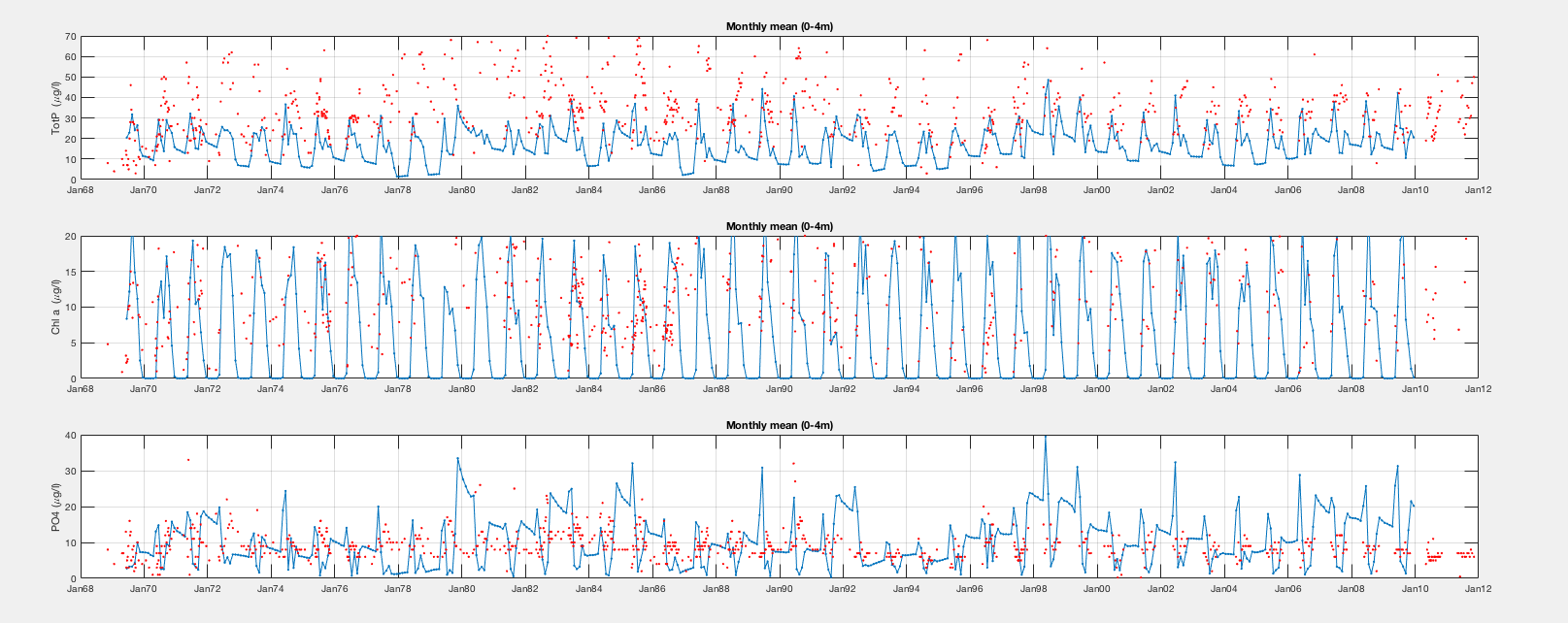
g­­­\_twty = 0.6

g\_twty2 = 0.6

m\_twty = 0.02

m\_twty2 = 0.02

All other parameters set to defaults



Total P is underestimated, chl a is well-predicted, and PO4 is in the right ballpark except for large spikes in the spring that are not borne out on the rare occasion that it was sampled.

Total P, chl a, and PO4 when phytoplankton growth parameters are both set at:

P­\_half = 0.483

P\_half2 = 21

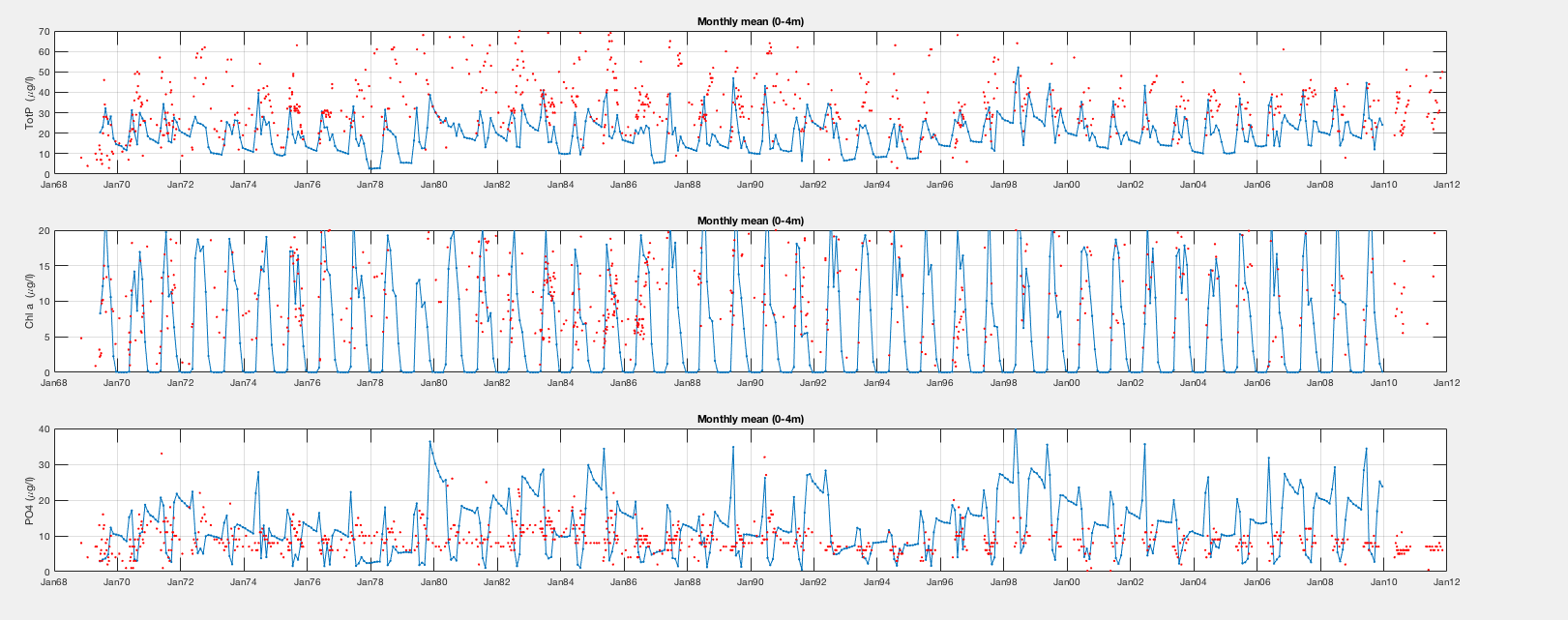
g­­­\_twty = 0.6

g\_twty2 = 0.6

m\_twty = 0.03

m\_twty2 = 0.03

All other parameters set to defaults



Total P is underestimated, chl a is well-predicted, and PO4 is still overestimated.

It seems that lower m\_twty values result in underestimation of total P, whereas higher m\_twty values results in overestimation of PO4.

It’s also interesting to note that visually, post-1990 is better predicted than pre-1990. It will be useful to statistically compare these periods once the parameters are fully optimized.

Next I also want to unpack the differences in phytoplankton abundance between the two pools of chlorophyll. They drastically change over time and it would be good to see if the model predicts these changes well.

**2017/08/21**

Igor Markelov sent me the current version of MyLake v2 that he is working up for Vansjo. My first step today was to look at the solvemodel function and compare v1.2.1 to v2. Here is a summary of the changes and additions.

**Changes**

* The function now generates two datasets: MyLake\_results and sediment\_results. This replaces the list of variables generated in the previous version of the model.
* Reads input data from modelinputs\_v2 rather than from modelinputs\_v12
* Error messages for outputs being out of range commented out
* Code for PDz\_store, Psz\_store, and chlsz\_store removed
* Uses convection\_v2 rather than convection\_v1
* Some calculations removed to make way for new chemistry and sediment module
  + Growth and loss for chl groups 1 and 2 removed
  + Sediment porewater fluxes removed
  + Sedimentation removed
  + Suspended solids correction commented out
* dRdz and TCz commented out in exchange for un-commenting duplicate equations (for different versions of Matlab (lines 1258-1262)
* New formulation for MixStat matrix (lines 1534-1596)
* Changes to computations for tridiagonal matrix

**Additions**

* Initial DIC and O2 profiles
* Output DIC, CO2, O2, O2 saturation (relative and absolute) profiles
* Fokema module
  + CDOM absorption (m-1) (line 77)
  + CDOM, DOC (3 pools), DOCtfrac (3 pools), daily\_BB (3 pools bacteria), daily\_pb (photobleaching) (lines 297-308)
* Switches
  + Deposition switch (set to 0)
  + Matsedlab\_sediment\_module (set to 1)
  + Wc\_chemistry\_module (set to 1)
  + Wc\_int\_method (set to 0)
  + Photobleaching (set to 0)
  + Flocculation (set to 1)
  + Resuspension (set to 0)
  + Rate estimator (set to 0, setting to 1 increases computational time by 20%)
* Scaling parameters
  + POP
  + O2
  + DIC
  + NO3
  + NH4
  + SO4
  + Fe2
  + CA2
  + pH
  + CH4
  + Fe3
  + Al3
  + SiO4
  + SiO2
  + Diatom
* Parameters for oxygen (lines 197-210)
* Code for setup of output data matrices, initial profiles, new profiles by day step, inflows, output matrices
  + DIC
  + CO2
  + O2
  + NO3
  + NH4
  + SO4
  + HS
  + H2S
  + Fe2
  + Ca2
  + pH
  + CH4
  + Fe3
  + Al3
  + SiO4
  + SiO2
  + Diatom
  + POP
  + H\_sw
  + H\_sw 2
  + Par
  + O2\_diff
  + O2\_sat (relative and absolute)
* Fluxes, solubility, profiles, transformations etc. for CO2 and O2 (lines 310-324, 717-746, 1218-1225)
* Sediment initial profiles and parameters (lines 442-450)
* Code for Pz if deposition switch = 1 (lines 600-602)
* Code for calculation of DOC profiles
* New chemistry module and sediment module (lines 1278-1438)
* Fully new set of script from lines 1845-end (calculating and generating outputs)

**2017/08/23**

Today, I had a look through the MyLake v2 Vansjo repository from Igor and have begun to port it into the repository I have made for L227. The model runs for Vansjo on my machine. I decided to port into my repository rather than putting L227 files into a clone of Igor’s repository, as there seem to be extra scripts in Igor’s repository that won’t be necessary for my application to L227.

I will need to change the code in some of the Vansjo functions, as I prefer to store parameters, etc. in txt or xls files that can be called up in the code rather than storing the parameters themselves in the code. When I need to change parameters, the code can then stay the same.

There are several scripts for calibration routines for the v2 Vansjo model. I’m not sure how these work exactly, but they may be useful later. I am making a note here to ask Igor about them once the model is up and running.

Today, I added the following scripts to KRS\_ELA\_Model in a new folder called v2:

* carbondioxideflux.m
* carbonequilibrium.m
* convection\_v2.m
* fokema\_new.m
* modelinputs\_v2.m
* MyLake\_save\_result\_for\_init\_conc.m
* nansum.m
* oxygenflux.m
* relative\_oxygen.m
* solvemodel\_v2.m

In order to get v2 working for L227, I need to complete the following steps:

* Update code in MyLake\_single\_run.m
* Update code in fn\_MyL\_application.m
* Update parameter file for L227: new rows
  + k\_BOD
  + k\_SOD
  + theta\_BOD
  + theta\_BOD\_ice
  + open\_slot
  + open\_slot
  + theta\_T
  + pH
  + I\_scDIC
  + Mass\_Ratio\_C\_Chl
  + 57
  + 58
  + 59
  + I\_scO
* Update initial file for L227: new columns
  + DIC
  + NO3
  + NH4
  + SO4
  + HSz
  + H2Sz
  + Fe2z
  + Ca2z
  + pHz
  + CH4z
  + Fe3z
  + Al3z
  + SiO4z
  + SiO2z
  + Diatomz
  + POCz
* Update input file for L227: new columns (same as init)
* Check on needs for adding sediment module
  + Sediment initial file
* Look into observation data for sediments in L227

**2017/08/24-30**

Updates for v2 scripts and functions:

**MyLake\_single\_run.m**

\*Lines may be off by a few due to additions

* Updated paths
* Added lake, year
* Changed is\_save\_results to false (final concentrations will not be saved as new inits)
* Lines 30-204: commented out
* Lines 208-298: commented out
* Lines 13, 314: commented out (do not want the model to run multiple times)
* Line 318: changed path for MyLakeResults.mat to L227 Application folder
* Lines 351-379: commented out (will add L227 observations back in when the model runs)
* Line 388: Removed “end” – didn’t correspond to anything
* Line 13: added run\_ID = 'L227 run 1'; to avoid an error
* Line 14: added clim\_ID = run\_ID

**Fn\_MyL\_application.m**

* Line 18: path for Vansjo parameters changed to L227 parameters
* Line 38: Lake 227 specified
* Line 39: 1969 specified
* Line 58: path for Vansjo params changed to L227 params
* Line 59: path for Vansjo inits changed to L227 inits
* Line 63: path for Vansjo inputs changed to L227 inputs
* Lines 93-177: commented out second basin

**modelinputs\_v2.m**

* Changed the function from init\_filename, etc. to initsheet, etc. to match modelinputs\_v12.m
* Line 45: uncommented line and changed to match modelinputs\_v12.m
* Lines 50-64: commented out
* Lines 68-71: uncommented, range changed from 3:25 to 3:40
* Lines 76-80: uncommented, range changed from 26:48 to 41:74
* Line 92: uncommented
* Line 94: commented
* Lines 96-127 and 151-153: changed parentheses from (1:end…) to (3:end…)
* Line 100: In\_Sz changed to In\_POCz
* Line 144: uncommented
* Line 146: commented
* Line 153: column range changed to 1:30 rather than 1:33

**Input file**

* Added new variables (new columns)

**Init file**

* Added new variables (new columns)

**Param file**

* Added new parameters from the Vansjo v2 parameter file (same values)

Note: at the end of 2017/08/30, I have now worked out all the errors related to the port from v1.2 to v2. However, I still need sediment data for the model to run (this is the error I’m getting now).

**2017/08/31**

Today I am reading through the Langtjern wiki to make sure I know the ins and outs of MyLake v2. I am updating notes here to document things I need to change in the L227 application:

1. MyLake\_application.m should call the following function:

[zz,Az,Vz,tt,Qst,Kzt,Tzt,Czt,Szt,Pzt,Chlzt,PPzt,DOPzt,DOCzt,DICzt,CO2zt,O2zt,NO3zt,NH4zt,SO4zt,HSzt,H2Szt,Fe2zt,Ca2zt,pHzt,CH4zt,Fe3zt,Al3zt,SiO4zt,SiO2zt,diatomzt,O2\_sat\_relt,O2\_sat\_abst,BODzt,Qzt\_sed,lambdazt,...

P3zt\_sed,P3zt\_sed\_sc,His,DoF,DoM,MixStat,Wt,surfaceflux,O2fluxt,CO2\_eqt,K0t,O2\_eqt,K0\_O2t,...

CO2\_ppmt,dO2Chlt,dO2BODt,testi1t,testi2t,testi3t,...

MyLake\_results\_basin1, sediment\_data\_basin1] = ...

solvemodel\_v2(m\_start,m\_stop,initfile,'lake', inputfile,'timeseries', parafile,'lake',In\_Z,In\_Az,tt,In\_Tz,In\_Cz,In\_Sz,In\_TPz,In\_DOPz,In\_Chlz,In\_DOCz,In\_DICz,In\_O2z,In\_NO3z,In\_NH4z,In\_SO4z,In\_HSz,In\_H2Sz,In\_Fe2z,In\_Ca2z,In\_pHz,In\_CH4z,In\_Fe3z,In\_Al3z,In\_SiO4z,In\_SiO2z,In\_diatomz,In\_TPz\_sed,In\_Chlz\_sed,In\_FIM, ...

Ice0,Wt,Inflw,Phys\_par,Phys\_par\_range,Phys\_par\_names,

Bio\_par,Bio\_par\_range,Bio\_par\_names, Deposition);

Then, the outputs of interest can be added to the MyLake\_application.m function call and header

1. A point source P addition could be added using deposition switch in solvemodel\_v2.m (line 91). I should look into this to decouple experimental P additions from catchment inputs.
2. Switches added to solvemodel\_v2.m that can be switched off if I want to run a simpler version of the model:
   1. Deposition (atmospheric deposition, point sources)
   2. Matsedlab\_sediment\_module (diagenetic model)
   3. Wc\_chemistry\_module (reaction network in water column)
   4. Photobleaching (0 = TSA model, 1 = FOKEMA model)
   5. Flocculation (testing only, should remain off)
   6. Resuspension
   7. Rate\_estimator
3. Note: dissolved and solid species are denoted in input and init files by “z” or “z\_sed.” E.g., dissolved iron is Fez (mg/m3), and solid iron in the sediment is Fez\_sed (mg/m3).
4. pH module: I need to install the optimization toolbox to use this component of the sediment model. Perhaps not needed?
5. Very detailed information describing the sediment module is detailed in section 2 of the Langtjern Wiki. It also describes in detail how to extract rates and fluxes when coupling the sediment and water column portions of the model. I should look back on this as I spin up the L227 sediment model.

Also, I spoke with members of my lab to find out if we have access to sediment profiles in L227. Jason sent several papers along. He also suggested taking a look at Ray’s Stella model to see how that model handles phosphorus.

* Rudd and Hamilton 1975
* Hesslein 1976
* Hesslein 1980
* Schindler et al. 1987

Rich also mentioned that there is a manuscript in preparation which explores phosphorus cycling, including sediments, in L227. This was done by Diane Orihel and Philippe Van Cappellen, and Sherry will speak with them about the possibility of using their data.

https://uwaterloo.ca/science/news/phosphorus-legacy-causes-recurring-blooms-according-ela-lake

Regarding the problem of not having very much water column SRP data from ELA (Jason):

“For the P species, that is a tricky question. I suppose with alkaline phosphatase the DOP fraction (which is pretty much TDP) is available but the half-life of an actual PO4(3-) molecule in that lake must be incredibly short. In some ways DOP is just actual PO4(3-) with a time lag and SRP is going to be hard to measure because of its reactivity. A few days after the weekly P spike any actual PO4(3-) must be produced in the lake and simply recycled through the plankton biomass.

“A simple test would be to scale the TDP using an average of the SRP:TDP ratio where both measurements exist. (Pretending that SRP is something approximating PO4(3-) after sample collection and the boat rides back to the chemlab at camp.) But that still relies on the TDP. How does Ray do it with his Stella-based model?”