**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.

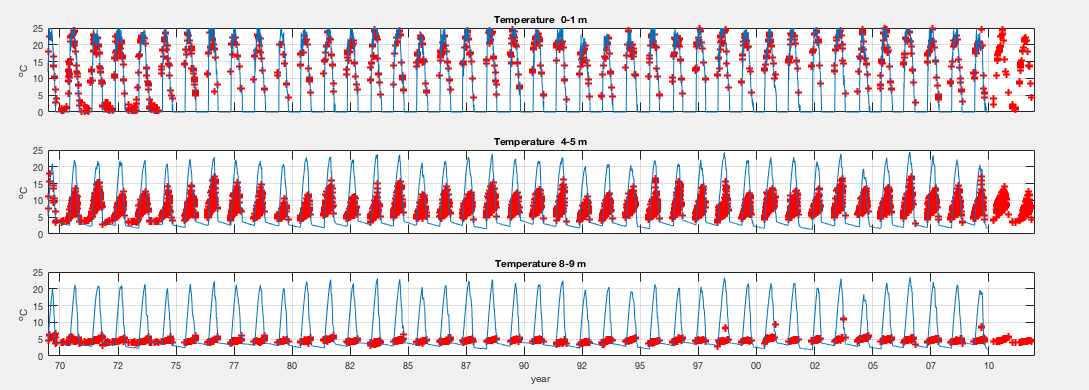
**2017/07/22**

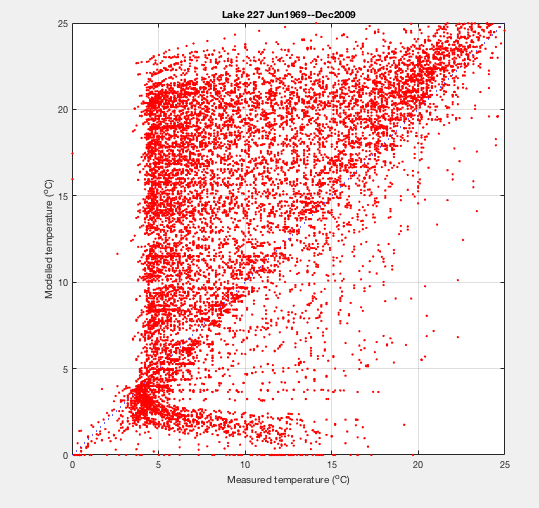
The first priorities for the model are to get the physical parameters correct: lake hydraulic residence time, temperature, ice cover, and wind. It will thus be key to get the input file correct, as this is the file from which the outputs are generated (along with the previous day’s outputs). I see that 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.

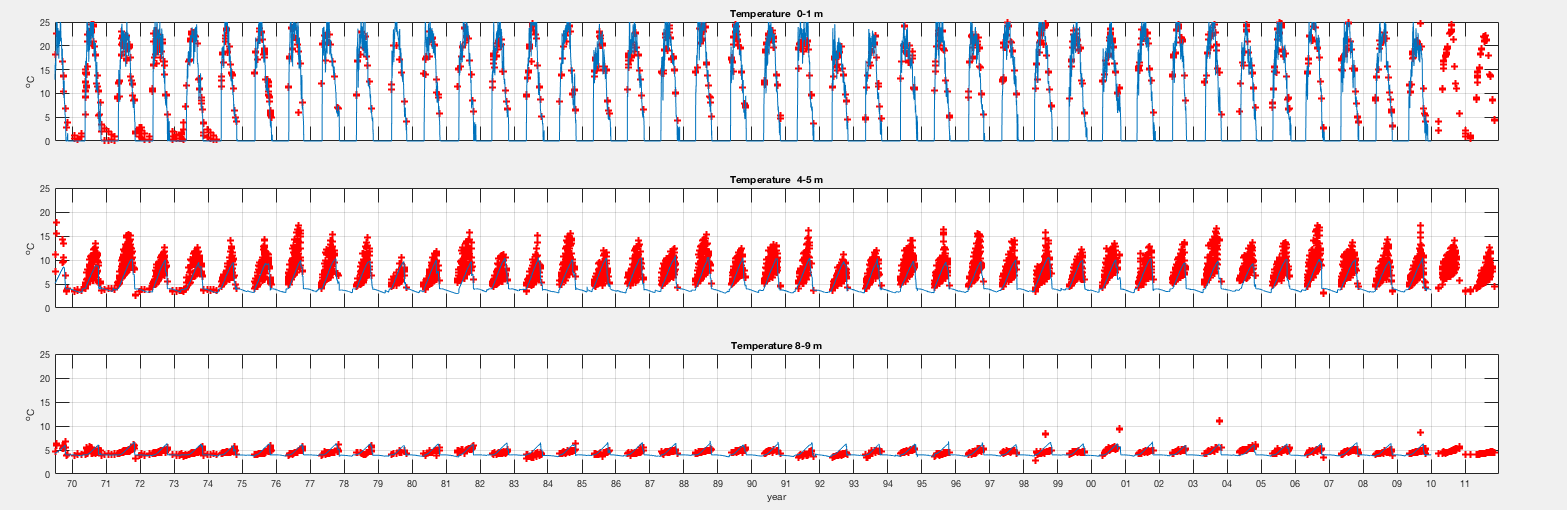
I noticed that the input file for the L227 does not contain any chemistry/biology for the inflows (passive and sedimenting tracer, total P, dissolved organic P, chl a). This will need to be taken from the Lake 239 inflow data (NW and NE inflows, perhaps averaged) and written into the matrix for the inputs (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.

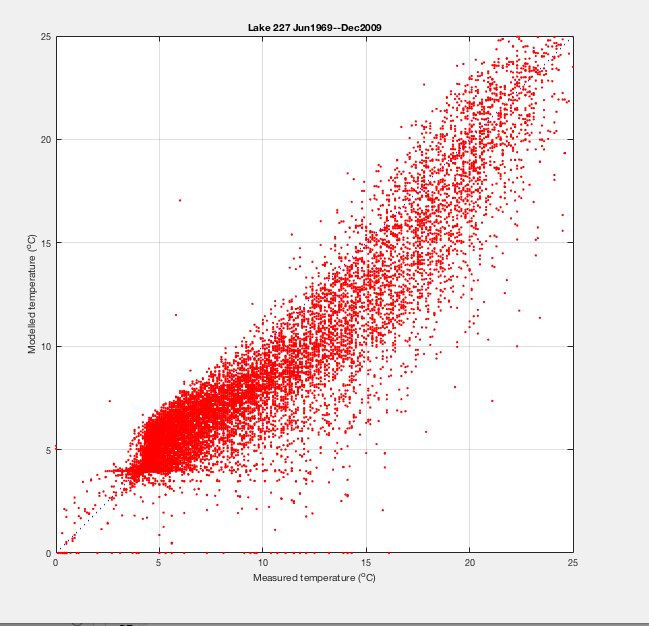
**2017/07/24**

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.

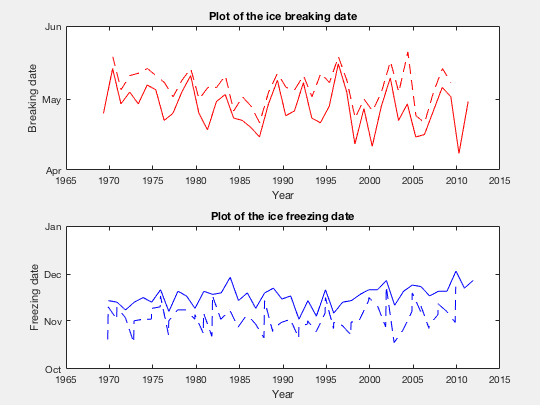
However, not 100% of the precipitation in the catchment reaches the lake, so the I\_scV parameter must be scaled to reflect this. In the “L227 hydrology” document, a regression between precipitation and L227 outflow is presented (R2 = 0.68, N = 28):

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.



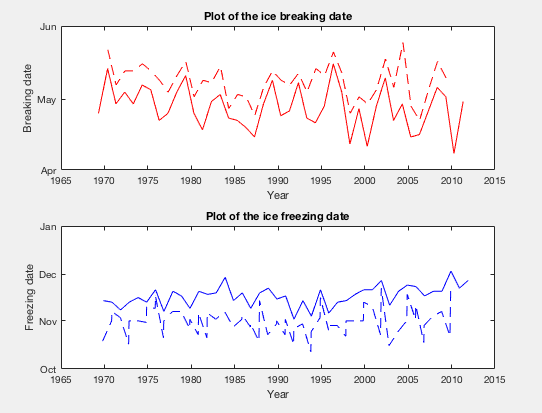
BreakModel-BreakObs = 7.000 ± 4.478

FreezeModel-FreezeObs = -10.525 ± 6.445

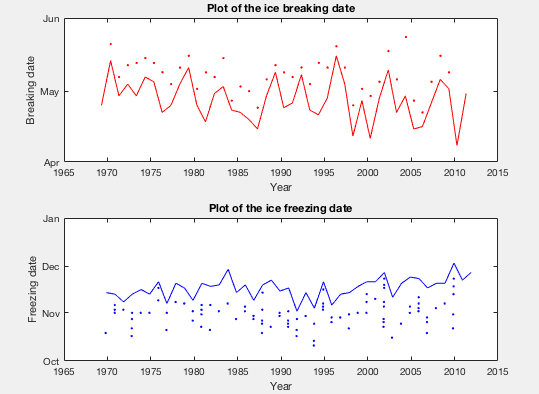
However, I ran the PerformanceIce.m script from Remy, and the rmse for ice break and freeze are 6.9250 and 16.5854, respectively. This is higher than the rmse values in Remy’s thesis (3.78 and 2.08 for breaking and freezing, respectively). This may not be a huge deal – I will ask Raoul and Jason.

**2017/07/30**

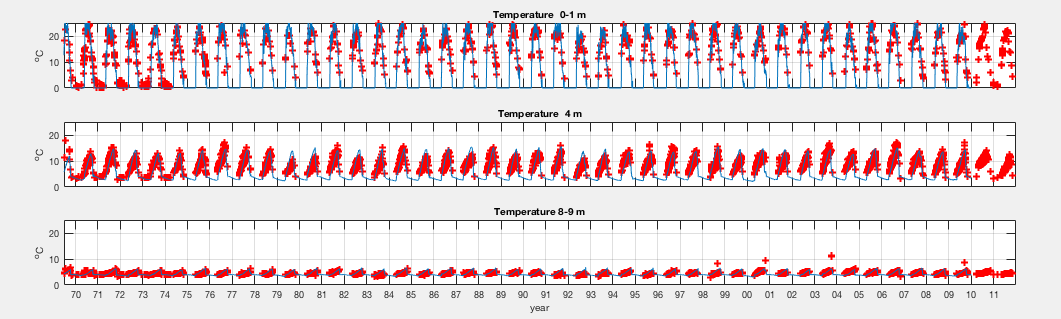
I updated the heatflux\_v12.m script file today. I noticed that on lines 76-80, the script specifies total transmissivity if global radiation data are missing from the input file (they are missing for L227). The file originally had the empirical determination of transmissivity from Beisjo and Vansjo 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**

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?”