InfoVeg - Bug #3611

Newell R module = supersample?

11/07/2008 09:54 PM - Michael Lee

Status:	In Progress	**Start date:**	11/07/2008
Priority:	Normal	**Due date:**	
Assignee:	Michael Lee	**% Done:**	0%
Category:	DataFix	**Estimated time:**	0.00 hour
Target version:	2009-June	**Spent time:**	0.00 hour
Bugzilla-Id:	3611		

Description

I have discovered a rather surprising anomaly in our database. I am working on making sure that the calculation of R module size is correct for tree stems, which is necessary for getting basal area correct, at least at that modular level. This is relatively straightforward, because the tree size of the plot is known, and so are the number of intensive tree modules (generally- sometimes if no trees are in a module, it doesn't get recorded and so it's not clear if the empty module is really empty or part of R. This email is not addressing that minor issue).

There are 30 plots in the database that have the same number of intensive modules as tree plot size (e.g. plot size of 2 and intensive modules are 1 and 2) that ALSO have tree module R. So it begs the question, what is the R module representing. Claire Newell's plots (project 10 and 11) are the only ones this way, and all of them are supersampled. It seems clear to me that the R module is where she parked the extra stems from supersampling, which is to say that the intensive module numbers are NOT supersampled. The R module, when added to the 1 or more intensive modules, would accurately represent the super sample.

So I'm not exactly sure what to do about these. Her data do make sense if only analyzed at the plot level, ignoring the modules. That would be option A for these plots: remove all module information and call all rows module R. I briefly thought about creating R modules and increasing the plot size to match them, but the difficulty there is that not all species are supersampled always, thus module R is not really a complete module. That would be option B, which I no longer consider hopeful. Option C would be to split up the R modules's stems and distribute them amongst the various intensively sampled modules, of which there are 4 times only 1 (that's easy enough), 19 times 2 modules, 3 times 3 modules, and 4 times 4 modules.

I like option A: to remove all intensive module information from these 30 plots and then the supersampling is accurate. Please let me know (soon) what you would like me to do.

To see the raw data, see text files:
\\Bioark\peetlab\CVS\CVS_Projects\10_Linville\10_trees.arc
\\Bioark\peetlab\CVS\CVS_Projects\11_Shining\11_trees.arc
plots:
010-0C-0013
010-0C-0020
010-0C-0025
010-0C-0027
010-0C-0029
010-0C-0053
010-0C-0056
010-0C-0057

010-0C-0059
010-0C-0062
010-0C-0082
010-0C-0085
010-0C-0087
010-0C-0091
010-0C-0096
010-0C-0097
010-0C-0101
010-0C-0112
010-0C-0176
011-0C-0302
011-0C-0309
011-0C-0314
011-0C-0350
011-0C-0396
011-0C-0400
011-0C-0405
011-0C-0410
011-0C-0416
011-0C-0434
011-0C-0441

Related issues:

Blocked by InfoVeg - Bug #3609: subsampling typo? 11-C-309	**Resolved**	**11/07/2008**

History

#1 - 03/20/2009 10:39 AM - Michael Lee

4 of these plots are 1 module, so those can just have the R modules switched to 1:
authorObsCode
010-0C-0029
011-0C-0396
011-0C-0350
010-0C-0062

This is done for these 4 plots.

#2 - 03/20/2009 10:40 AM - Michael Lee

For the remaining 25 plots, the only thing to be done that I can see is to extend the bounds of these plots for trees so that they really do have an R module. Then we decrease the supersamples to 100% and the subsamples accordingly (i.e. a 20% subsample of the plot would become a 10% subsample if the plot was 200% supersampled).

I will create a list of all species on these plots with # stems and basal area as we think it is now, then recreate this after I am done to be sure it went correctly.

#3 - 03/20/2009 10:45 AM - Michael Lee

OK, the bugzilla thread is getting tired. After reading the email thread about this issue once more, I realize that we can't do what I suggested (creating real R modules) because the basal area and density would be skewed within each module or for the full plot for species not sampled (that is, ignored) in the R module.

The other possibility would be to divvy up the stems from R into the modules in as even a way as we can, keeping the subsamples as they are. I will see how feasible this is.

#4 - 03/20/2009 11:43 AM - Michael Lee

Updates have been queued for switching the stems over to randomly picked intensive modules (then stratified by me somewhat to prevent stacking too many stems in one module). I'd like to check with Forbes or Bob before committing the update to the database (revision project 33).

#5 - 03/20/2009 11:46 AM - Michael Lee

More OLD email thread:

From: Robert K. Peet
Date: Mon, Nov 10, 2008 at 5:14 PM
To: Michael Lee
Cc: Forbes Boyle

If Michael is correct, it does seem to beg the question of what she did when she supersampled and also had Rs. Were teh extra trees from the intensives placed in the intensives or in the Rs? We should look into this. Is the ratio of trees in the Rs relative to intensives what it should be in such supersampled plots. Michael, might you check this out and let us know whether we need to investigate further?

Bob

From: Michael Lee
Date: Mon, Nov 10, 2008 at 6:25 PM
To: "Robert K. Peet"
Cc: Forbes Boyle

Sure, that's a good point, Bob. I have migrated the data into the new archive now (still doing QA on it, though, so far, so good), and this task should be much easier in the new archive than in the old one. I'll let you know what I find out.

From: Robert K. Peet
Date: Tue, Nov 11, 2008 at 6:18 PM
To: Michael Lee
Cc: Forbes Boyle

Michael,

Forbes and I have looked into this problem a bit more (via phone with me at NESCent).

We looked at 010-0C-0013. In this case Claire sampled just three intensive modules at 150% and there were no Rs, except that for trees there were Rs. Michael and I agree that the modules should be at 100% and not 150%, and the there should be 1.5 modules of R for trees for a total tree area of 4.5 modules rather than the 3 reported. She had also marked the individual tree lines in the modules as 150% but this does not appear in the digital data. There were two species present in the R module not present in the intensives, so we can be confident that the R stems do not appear in both places. I expect this is the pattern for pretty much all the 30 plots you mention.

We looked at 011-0C-0441. In this case Claire sampled two intensive modules at 150% and the trees showed up in Rs. But for the intensives she did a 20% sample for Kalmia. Kalmia does not show up in the Rs. This measn that the Rs are again adding area, her an extra 100m2, but that the percent suybsample for Kalmia in the Rs should be 0%. This is a possible problem in all of the 30 plots you identify, so you will need to screen for individual taxa with their own subsamples ansd adjust like we report here.

We looked at 010-0C-104 because it is a case of Claire doing 10 modules and havind a supersample of 150%. At least in this case the total stems in the Rs relative to the stems in the intensives was consistent with the ratio of area of Rs being 6/10 of the plot, so we think the supersample was done correctly. I don't know how to check for this elsewher, ,except to repeat the analysis we did of looking at stem number rations for misfits.

As an aside, in the process of looking at the date, we used the viewer program 15a. This has an error in that the total area of the Rs is consistently reported as 100 rather than the correct value. For example, in 010-0C-104 referenced above teh area of R should be 600 m2, but is given as 100 m2. In contrast, the total area of for the S records is correctly given as 1000 m2. Perhaps this does matter as we will not be using the viewer much once we are in the new database.

Let me konw if you have any questions about this or wish how to discuss how to continue from here.

Best,

From: Michael Lee
Date: Thu, Nov 27, 2008 at 9:35 PM
To: "Robert K. Peet"
Cc: Forbes Boyle

Hi Bob and Forbes,

Happy Thanksgiving!

I have run an analysis on Claire's plots. There are 202 plots with R modules as part of them in the stems. 30 of these are the type of plot we are already talking about: where R does not logically fit as a stand-alone module. All 30 of these had super-sampling on them.

Of the remaining 172 plots, 136 had R modules with stem BA within 10% of the range of stem densities of the other modules (10% below the least dense up to 10% more than the most dense module). That's 79%.

I thought it would be interesting to see what was going on with this if you split the plots out that have sub- or super-sampling. Considering plots without any super- or sub-sampling, there are 114 out of 141 that have R modules are within the 10% window (81%). Of plots that have sub- or super-sampling, 22 out of 31 plots are within the 10% window (71%). Looking at the plots as a whole, it doesn't

seem that there is a systematic issue here. Some plots have significantly more BA in R or significantly less. Those I looked at were generally sparse and might have a huge stem in the residuals, which skews BA significantly.

Once I had that set up, it was simple to run the routine for all projects. Claire's plots (not counting the 30 odd balls with no area for R and supersampling) come out right in the middle. I attach an Excel spreadsheet of what I did. The first chart is all projects with

1. of plots, # super/sub sampled, # not (as bars). The lines are the percent of plots with R modules within 10% of min and max of intensives (BA), as well as percent of plots that were super/subsampled vs. no super/subsamples. I think, as you do, that Claire's data are OK with the exception of the 30 plots that will need adjusting as you outline above.

The only thing that doesn't make sense to me is your sentence: but this does not appear in the digital data." My copy does show 150% for all stems in 010-0C-0013.

#6 - 12/02/2009 01:48 PM - Michael Lee

One module plots only were fixed a while ago, simply merging R into module 1.

I have updated these and fixed them I think. I looked through all cases where there were fewer than 5 species in the R module. In these cases, it was fairly simple and most accurate to split the stems into the 1,2 or 1,2,3 or 1,2,3,4 modules as the case may be.

For all the other plots (with at least 5 species in the R module), I removed module data for stems so that the stems are accurate for the full-plot with supersampling as originally indicated (<=100% usually for saplings and >100% for trees).

I will compare more explicitly the data before I updated it and after to ensure that stem density is what it should be before resolving this bug.

#7 - 12/03/2009 02:19 PM - Michael Lee

this has been fixed! version 1.1.51 archive database. QA succeeded.

#8 - 01/05/2010 08:02 AM - Michael Lee

This bug has been reopened due to the new issue of non-standard cutoffs for supersampling stems. Bob and I discussed this issue today and think that it would be helpful if we could have an hourly employee go through project 10,11, and 12 and check each plot to see if this sort of issue exists. The issue exists when there are two numbers for stem tallies in the 5 and or 10 cm columns, one which is the actual sampled version and another that increases the number to fit the supersampling for the larger stems.

Once we know how many plots we have to deal with like this, we can determine how to fix it.

#9 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3611