Betsy Cowdery

December 12, 2014

Abstract

The goal of this project is to develop a multivariate Bayesian meta-analytical model that synthesizes plant trait data from multiple studies while accounting for various sources of uncertainty. Using observed sample mean, sample size, and a sample error statistics for multiple plant traits, we aim to produced well constrained estimates of mean and precision for a single trait. This has be done in PEcAn (the Predictive Ecosystem Analyzer) using a univariate model [1], however, a multivariate model can leverage the fact that many plant traits are highly correlated [2] to constrain our estimates even further. This may be especially useful improving predictions for studies in which observations are missing.

1 Introduction

- what is meta-analysis
- why is it necessary in this case?
 - what is PEcAn?
- what does PEcAn already do
 - what did they do in the nature paper?
- how do we plan on using their findings

The goal of this project is to develop a multivariate Bayesian meta-analytical model that synthesizes plant trait data from multiple studies while accounting for various sources of uncertainty. Using observed sample mean, sample size, and a sample error statistics for multiple plant traits, we aim

1 2 3 4 <u>5</u> 6 7 8 9 10 11 12 <u>13</u> 2 $\underline{14}$ <u>15</u> $\underline{16}$ <u>17</u> <u>18</u> <u>19</u> <u>20</u> 21 <u>22</u>

<u>23</u>

 $\frac{24}{25}$

 $\frac{26}{27}$

28 29

 $\frac{30}{31}$

 $\frac{32}{33}$

<u>35</u>

37

<u>39</u>

 $\frac{40}{41}$ $\frac{42}{43}$ $\frac{44}{44}$

to produced well constrained estimates of mean and precision for a single trait. This has be done in PEcAn (the Predictive Ecosystem Analyzer) using a univariate model, however, a multivariate model can leverage the fact that many plant traits are highly correlated to constrain our estimates even further. This may be especially useful improving predictions for studies in which observations are missing.

2 Materials and Methods

2.1 Experimental Data

For this project I am using the data set compiled from the global plant trait network (Glopnet), a database created to quantify leaf economics across the worlds plant species (Wright et al., 2004).

These data are described and are available as supplemental material in (Wright et al., 2004).

From this data set I will be focusing on six plant-traits:

- (1) Leaf mass per area (LMA)
- (2) Photosynthetic capacity (Amass) photosynthetic assimilation rates measured under high light, ample soil moisture and ambient CO2
- 34 (3) Leaf nitrogen (N)
- $\underline{36}$ (4) Leaf phosphorus (P)
- 38 (5) Dark respiration rate (Rmass)
 - (6) Leaf lifespan (LL)

 $\frac{2}{3}$

<u>5</u>

1

<u>3</u> 1 _

2.2

Data Model:

 $\frac{6}{7} \quad Y_{i,j} \sim N(\mu_{i,j}, \sigma^2)$

9

 $\frac{10}{11}$ Process Model:

 $\frac{12}{13} \quad \sigma^2 \sim IG(s_1, s_2)$

 $\mu_{i,j} \sim N(\mu_0, V_\mu)$

 $\frac{16}{17}$

<u>18</u>

19

2.3 Multivariate Model

Univariate Model

 $\frac{20}{21}$

Data Model:

 $\frac{4}{5}$ $Y_i \sim N_p(\vec{\mu}_i, \Sigma)$

 $\frac{26}{27} Y_{i,j}^{(0)} \sim N_p(Y_{i,j}, \sigma^2)$

28 29

<u>31</u>

<u>33</u>

<u>35</u>

Process Model:

 $\sigma^2 \sim IG(s_1, s_2)$

 $\mu_{i,j} \sim N(\mu_0, V_\mu)$

 $\frac{36}{87}$ $\Sigma \sim IW(V, df)$

39

 $\frac{40}{41}$

4

43 44

 $\frac{44}{45}$ $\underline{46}$

 $\underline{47}$

3 Results and Discussion

Each of the two plot matrices in figure 1 consist of three parts. The lower left half of the plot matrix

shows scatterplots of each combination of variables with with a regression line in red if the regression

4

1

8

 $\frac{18}{19}$

13

14

 $\frac{20}{21}$ $\frac{22}{23}$

 $\frac{24}{25}$

26

 $\frac{32}{33}$ $\frac{34}{34}$

<u>35</u>

36

31

37 38 39

 $\frac{40}{41}$

 $\frac{42}{43}$

 $\frac{44}{45}$

 $\frac{46}{47}$

is statistically significant (p_i.01). The top right half of the plot matrix shows pairwise correlation. The plots on the diagonal show histograms of the collected data.

The lefthand matrix uses all the available data and the righthand matrix uses only studies in which every plant trait was measured. This significantly reduces the number of data points.

Using the full dataset, every plant trait has a statistically significant correlation with all the others, which suggests that a multivariate analysis will in fact be informative. Using the limited dataset, correlation values are lower and one pair of traits no longer have a statistically significant correlation.

When using data that excludes all studies with missing observations, there is practically no difference between the two models estimated means for each of the variables. However, for both models, including studies with NAs produces estimated means that are significantly different from those produced with data excluding NAs. For the variables Log.LMA and Log.Amass, the estimated means from the univariate and multivariate models were very close, but for the remaining variables, they were noticeably different, with the estimated mean from the univariate model always closer to the data mean than the estimated mean from the multivariate model.

Contrary to what I expected, the variation in the multivariate model showed little to no improvement over the univariate. The multivariate model produced posterior distributions with larger variance around the mean (except for Log.Nmass where the SE for the univariate model was larger by 3.3e-06, a small amount relative to the size of the mean.)

When studies with missing observations are included, the standard errors begin to behave more like one might expect. The posterior distributions from the multivariate model have smaller variances for all the variables except Log.Pmass. However, it is difficult to see the difference since the expected means of the variables are no longer similar.

2

1

3 4 <u>5</u>

<u>11</u> <u>13</u> $\underline{14}$

10

<u>16</u> $\underline{17}$ <u>18</u>

<u>19</u>

<u>20</u> <u>21</u> <u>22</u> <u>23</u>

 $\underline{24}$

 $\underline{25}$ <u>26</u> <u>27</u> <u>28</u> <u>29</u> <u>30</u>

<u>32</u> <u>33</u> 34 <u>35</u> <u>36</u>

<u>37</u>

<u>31</u>

<u>39</u> <u>40</u> $\underline{41}$

<u>42</u> <u>43</u> <u>44</u> $\underline{45}$

<u>46</u> <u>47</u>

Univariate Model 3.1

Excluding data with missing values 3.1.1

Including data with missing values 3.1.2

Multivariate Model 3.2

Excluding data with missing values 3.2.1

3.2.2 Including data with missing values

Conclusion 4

Applying Bayesian models to the GLOPNET data has proven to be useful in two very important ways: First of all a Bayesian models (both univariate and multivariate) can handle data with missing values. This changed the sample size from 42 to 2,.... Such a considerable increase in sample size helped to constrain parameter estimates. Secondly, introducing a multivariate model, changed the shape and sizes of the joint probability confidence regions. The changes at times were subtle and could go undetected when projected back into single probability functions. However, changes such as decreased overall area and increases in elongation of the confidence ellipse will improve our estimation of plant traits.

How to Count Lines 6

 $\frac{1}{2}$ $\frac{3}{4}$ $\frac{5}{6}$

? Wright et al. (2004)

 $\frac{6}{7}$

8

9

<u>14</u>

32

 $\frac{33}{34}$

 $\frac{35}{36}$

 $\frac{37}{38}$

 $\frac{39}{40}$

 $\frac{41}{42}$

 $\frac{43}{44}$

 $\frac{45}{46}$

47

References

 $\frac{10}{11}$ Clark, J. S. (2007), Models for Ecological Data: An Introduction, 617 pp., Princeton University Press.

Clark (2007) Dietze et al. (2013) Dietze et al. (2014) Gelman et al. (2013) LEBAUER et al. (2013)

- Dietze, M. C., D. S. Lebauer, and R. Kooper (2013), On improving the communication between models and data., *Plant, cell & environment*, 36(9), 1575–85, doi:10.1111/pce.12043.
- Dietze, M. C., S. P. Serbin, C. Davidson, A. R. Desai, X. Feng, R. Kelly, R. Kooper, D. Lebauer,
 J. Mantooth, K. McHenry, and D. Wang (2014), A quantitative assessment of a terrestiral biosphere
 models data needs across North American biomes, Journal of Geophysical Research: Biogeosciences,
 119(3), 286–300, doi:10.1002/2013JG002392.Received.
- 19 Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013), Bayesian Data Analysis, Third Edition, vol. 1, 675 pp., CRC Press.
- LEBAUER, D. S., D. WANG, K. T. RICHTER, C. C. DAVIDSON, and M. C. DIETZE (2013), Facilitating feedbacks between field measurements and ecosystem models, *Ecological monographs*, 83(2), 133–154.
- Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-bares,
 T. Chapin, J. H. C. Cornelissen, M. Diemer, J. Flexas, E. Garnier, P. K. Groom, J. Gulias,
 K. Hikosaka, B. B. Lamont, T. Lee, W. Lee, C. Lusk, J. J. Midgley, M.-L. Navas, U. Niinemets,
 J. Oleksyn, N. Osada, H. Poorter, P. Poot, L. Prior, V. I. Pyankov, C. Roumet, S. C. Thomas,
 M. G. Tjoelker, E. J. Veneklaas, and R. Villar (2004), The worldwide leaf economics spectrum.,
 Nature, 428 (6985), 821–7, doi:10.1038/nature02403.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis

ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu,

5 6 7

<u>13</u> 14

15

<u>16</u>

12

17 18 <u>19</u>

20

<u>25</u> 26 <u>27</u> 28

29

31

32 33

<u>35</u>

36 37 38

> <u>40</u> 41 42

43

39

<u>44</u>

46 47 Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo.

accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan

bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit

mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus

et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper

vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada portitior diam. Donec felis erat, conque non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus conque quam, in hendrerit risus eros eget felis. Maecenas eget erat in

 $\frac{43}{44}$ $\frac{45}{46}$ $\frac{47}{47}$

sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer. Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.