

Manuscript Number: JQSR-D-07-00145R1

Title: Obtaining Accurate and Precise Environmental Reconstructions from the Modern Analog Technique and North American Surface Pollen Dataset

Article Type: Research and Review Paper

Corresponding Author: Dr. John W Williams, PhD

Corresponding Author's Institution: University of Wisconsin

First Author: John W Williams, PhD

Order of Authors: John W Williams, PhD; Bryan Shuman, PhD

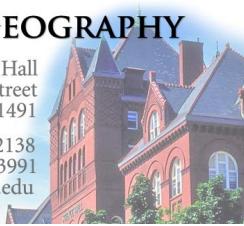
Abstract: Using a recently compiled surface-sample pollen dataset for North America, we test methods to improve the skillfulness (i.e. accuracy and precision) of the modern analog technique (MAT) for continental- to sub-continental-scale paleoenvironmental reconstructions. Interregional floristic differences challenge such reconstructions, because with widening spatial extent each pollen type represents an increasing number of species, each with its unique niche, thus blurring pollen-climate relationships. We conduct a series of cross-validation experiments in which we first explore 1) whether increasing the spatial extent of the pool of potential analogs improves or worsens MAT skill, 2) whether MAT skill is improved by increasing the number of pollen taxa and splitting pollen taxa into regional groups, and 3) the differences in MAT skill among environmental variables. Regional splits are guided by the range boundaries of parent species. Results are analyzed for multiple environmental variables. We then systematically explore options for a) the value of the no-analog/analog threshold ( $T$ ), b) the maximum number of modern analogs ( $N$ ) allowed for a target fossil sample, and c) whether the environmental average constructed from the modern analogs should be weighted by their compositional dissimilarity ( $D$ ) to the target sample. We demonstrate that substantial interregional differences in pollen-climate relationships, particularly between eastern and western North America, degrade MAT precision and accuracy, but these adverse effects can be overcome by expanding the list of taxa used and/or splitting pollen types into regional groups. MAT precision was best when pollen types were regionally split and better when more taxa were used, although increasing the taxon list from 64 to 135 types did not substantially increase performance. Temperature-related variables were reconstructed more precisely than hydrological variables, and there was little difference in MAT skill between climatic and bioclimatic variables (e.g. mean July temperatures versus growing degree days).  $T$  scales with the number of taxa analyzed, and there is a tradeoff between skill (best when  $T$  is low) and utility (if  $T$  is low, few samples can receive environmental inferences). For  $N$ , there is a tradeoff between precision and accuracy, such that retaining just the single best analog had the worst precision and best accuracy. Strongly weighting by taxonomic dissimilarity ( $1/D^2$ ) consistently reduced precision, but MAT precision was similar for unweighted and inverse-distance weighted averages. Given the above tradeoffs, we recommend using the 64-taxon list with taxa split by region,

$0.20 \leq T \leq 0.30$ ,  $3 \leq N \leq 7$ , and either no weighting or an inverse-distance weighting, for North American applications of the MAT.



## DEPARTMENT OF GEOGRAPHY

160 Science Hall  
550 North Park Street  
Madison, WI 53706-1491  
Telephone: 608-262-2138  
Fax: 608-265-3991  
Web: [www.geography.wisc.edu](http://www.geography.wisc.edu)



December 31, 2007

Dr. Jim Rose  
Editor in Chief, Quaternary Science Reviews  
Department of Geography  
University of London

Re: Ms. Ref. No.: JQSR-D-07-00145 ("Obtaining Accurate and Precise Environmental Reconstructions from the Modern Analog Technique and North American Surface Pollen Dataset")

Dear Dr. Rose:

Thank you for the opportunity to make the minor revisions requested by Reviewer 1 (Tom Webb). We have made the requested changes, and enclose the revised ms. for your consideration.

Detailed responses to Tom's comments are provided below. We followed almost all recommendations.

The revised ms. has 11,996 words (including references, tables, figure legends, etc.), 403 words in the abstract, 76 references, 8 figures, and 4 tables.

Thanks again and best wishes for 2008.

Sincerely,

John (Jack) Williams  
Department of Geography  
University of Wisconsin – Madison  
[jww@geography.wisc.edu](mailto:jww@geography.wisc.edu)

Bryan Shuman  
Department of Geography  
University of Minnesota, Twin Cities  
[bshuman@umn.edu](mailto:bshuman@umn.edu)

## REVIEWER 1 (TOM WEBB'S) COMMENTS & RESPONSES

### Reviewer #1: General Comments

1. This paper is well organized and addresses some important issues in the use of the widely used MAT in Quaternary palynology. It presents many different interesting results that will help palynologists better know how to interpret their data with the MAT in light of having access to extensive modern data sets.
2. I found a few places where the authors need to explain more about their results to make them understandable. But mostly the paper presents the results in a clear fashion and his telling figures to help illustrate the points.
3. Informative abstract. Much useful detail.
4. Clear, purposeful Introduction. Well written.
5. Clear set of useful conclusions.

**Thanks.**

### Specific Comments

1. P. 4, line 3: Is it statistical or just a numerical technique? **Well, it could be argued that the MAT is a non-parametric statistical technique, but ‘numerical’ works too. We’ve made the change.**
2. P. 7, line 7: delete the comma. **Done**
3. P. 12, lines 6-7: Not sure what you are driving at here. **Clarified concept of ‘analytical utility’ by changing paragraph to:**

**Very restrictive thresholds can improve analytical skill but reduce the analytical utility of analog approaches, because if restrictive thresholds disallow all possible analogs for most target samples, then the MAT will fail to make any environmental inferences for those samples (although its inferences for the remaining samples might be highly precise). Thus, in our leave-one-out analyses, we also note the number of samples with at least one modern analog, to indicate when high precision is gained at the cost of low analytical utility.**

4. P. 15, line 8: Good to have the 95% confidence intervals and non-overlap for some of the variables. You could note the statistical significance for those results. **Done. Added asterixes to table 3 to indicate significance and added new text:**

**The ENA-NA reconstructions also tend to be less accurate (have a slope further from 1) than ENA-ENA reconstructions, with significantly lower slopes ( $p<0.05$ ) for the temperature-related variables (TSum, TWin, MTCO, MTWA, GDD5) and summer precipitation (PSum).**

5. P. 15, Table 3: Why are the standard climatic variables left out here and only results for bioclimatic variables shown? Alright with me if you don't use them, but might also be helpful if they showed significant differences between ENA-NA and ENA-ENA. **Variables have been added to Table 3.**

6. P. 15, lines 22-23: With all the symbols on Figure 4, and the changing scales for the y-axis on Figure 4, I am finding it hard to compare the numerical values on Table 3 with the right open or black triangle, square or circle on Figure 4. Can't you add a figure to show just this comparison? **We opted not to add a new figure because we felt it would be too redundant with Table 3. Instead, we addressed this comment by editing this text to remove joint references to Table 3 and Fig. 4. Each is now reviewed separately, removing the need for cross-comparisons.**

7. P. 16, lines 1-2: Ok, here you give the numbers. Good, but are these differences significant? At least say something about the reduction in RSME in temperature terms because that is 3.95 deg C vs 2.86 deg. C, which is a reduction by over 1 deg C or by 2 deg C in a confidence interval of plus or minus 2.86 vs 3.95. Just citing the numbers without units does not give the reader much to hold on to. And what about 2.86 deg C. Do you really know temperature to hundredths of a degree. You could list these as 4.0 and 2.9 deg. C and perhaps make a stronger point. The concern of precision is a concern with significant figures and confidence interval widths. Don't let the computer's ability to produce many figures after the decimal point fool you into using them. **Good points. We've now noted that the units are °C and modified the sentence as follows:** "For example, in the NA-NA experiment and no distance weighting, RMSE= 3.31°C for TWin for the 36-taxon list, 2.94°C for the 80-taxon list, 2.84°C for the 135-taxon list, and 2.66°C for the 64S list (Fig. 3), resulting in a net 0.65°C (20%) improvement in precision." We're not exactly sure how to count the number of significant figures when assessing RMSE, but given that we are working with >4500 samples, two significant figures seems reasonable, especially when our interpretations are limited to comparisons among the modern validation experiments (and not, e.g. actual paleoclimatic reconstructions).

8. P. 16, line 8: You could note white triangles and squares in parentheses as showing this. "The ENA-NA and ENA-ENA results for the 64S-taxon list are particularly close (white triangles and squares in Fig. 4)." **Done**

9. P. 16, line 8: The suggesting phrase dangles and needs rephrasing. Try: This close match allows you to infer that splitting pollen types into eastern and western groups effectively excludes floristically false analogs from western North America from being assigned in ENA.

#### **Done**

10. P. 17, lines 1 and 2: Good point. **Thanks**

11. P. 17, line 7: worse? Why not be precise and say larger. **Done**

12. P. 17, line 13: reference Figure 5 on this line. **Done**

13. P. 18: lines 1&2: you could mention that you may explain more variance in a relative sense while the RSME is larger because the larger gradient can tolerate a larger RSME and still explain more variance. It is a case of the denominator (gradient magnitude or total variance) increasing more than the numerator (RSME). Both increase but one goes up far more than the other. You can explain more variance but still have larger error bars in terms of mm or deg C.

#### **Done. Modified text on p. 18 to insert some of this language.**

14. P. 18, line 13: Fig. not Figs. Also I think that RSME is a direct measure of precision because it relates to error bars or confidence intervals, which are what precision is all about. R-square measures percent variance explained and is not a direct measure of precision. I would rephrase the your statements in terms of what Fig. 4 shows for RSME. **Done – removed references to precision and instead mention percent variance explained.**

15. P. 18, lines 15-21: so why is this? What is D and why might weighting reduce precision? I just had to go back to p. 6, lines 10 & 11 to find out about what D is. It is the dissimilarity. What gets divided by D or D-square and who introduced this idea? I need to know more of what you did to figure out why you got the results that you got. **Clarified this concept by 1) adding sentence to intro, and 2) adding sentence to beginning of paragraph on p. 18:**

**1): When calculating the environmental averages, a simple unweighted average can be used, or the environmental variables associated with the most compositionally similar calibration samples can be upweighted using a 1/D or 1/D<sup>2</sup> weighting (Jackson and Williams 2004).**

**2): Because the MAT is based on the assumption that compositionally similar samples represent similar environments, using an inverse distance weighting ( $1/D$  or  $1/D^2$ ) to upweight the most similar samples should improve the skillfulness of MAT reconstructions (e.g. Calcote, 2003; Nakagawa et al., 2002). However, in these analyses, MAT skill is insensitive to the choice of dissimilarity weighting (Fig. 3). There is no clear advantage to using no weighting versus the  $1/D$  weighting (Fig. 3). The  $1/D$  weighting has slightly higher precision for PSum (Fig. 3d), whereas no weighting is slightly more precise for PWin (Fig. 3c), and there is no difference for temperature-related variables and fractional woody cover (Fig. 3a,b,e). There is, however, a small but consistent loss of precision for all variables when using the  $1/D^2$  distance weighting.**

16. P. 19, lines 1-9: here you explain how T works a bit and give an example. Do something of this nature for the section on taxonomic weighting. **OK, see above.**

17. **P. 19, line 20: How is it that Figure 6 shows regression statistics? Clarified sentence as follows: The high precision observed at  $T < 0.05$  occurs because most samples (70%) have no analog (Fig. 6) and hence are not included in the regression statistics (Fig. 6).**

18. P. 20, lines 17-18: How do I see this? Nothing much happens on the A-A' or B-B' transects. You imply that the values plotted in the upper right hand of the graphs in Figures 7 and 8 are lowest or highest. I guess that they do but not substantially at the  $N=10$  and  $T=0.5$ . **We've modified the last paragraph in this sentence to better highlight the key trend:**

**The T and N parameters strongly interact (Figs. 7, 8). When one is set to a fairly restrictive value (e.g.  $N=3$ ,  $T=0.10$ ), MAT skill is largely insensitive to changes in the other parameter (Figs. 7, 8). Conversely, both precision and accuracy substantially worsen when N and T are relaxed simultaneously, as shown by the decreases in precision and accuracy from the lower left to upper right of each plot in Figs. 7, 8.**

19. P. 25, lines 7-13: Say more. When I looked in the methods section, I missed finding where it tells just what D is and how it is measured. Is it the SCD? And then what does it divide as  $1/D$ ? I think that I may be getting it that it allows the sites with the smallest values of D (all less than T) to have more weight in determining the climate value. If so, then it makes some of N analogs be more important than the others. This could be somewhat equivalent to reducing N because only fewer samples determine the final value. But if T is set correctly, then dividing by D or D-square should not have an impact. But if T is large, then maybe it would compensate and make a difference. I would say a little more in the results about why you wanted to try this weighting to set up the discussion here. **We think these comments have been addressed by the changes made in response to comment#15 – i.e., by clarifying what D is earlier in the introduction and results.**

20. P. 26, lines 1-5: good point. **Thanks.**

21. Clear set of useful conclusions. **Thanks.**

[Note: This section duplicates information included in cover letter.]

## REVIEWER 1 (TOM WEBB'S) COMMENTS & RESPONSES

### Reviewer #1: General Comments

1. This paper is well organized and addresses some important issues in the use of the widely used MAT in Quaternary palynology. It presents many different interesting results that will help palynologists better know how to interpret their data with the MAT in light of having access to extensive modern data sets.
2. I found a few places where the authors need to explain more about their results to make them understandable. But mostly the paper presents the results in a clear fashion and his telling figures to help illustrate the points.
3. Informative abstract. Much useful detail.
4. Clear, purposeful Introduction. Well written.
5. Clear set of useful conclusions.

### Thanks.

### Specific Comments

1. P. 4, line 3: Is it statistical or just a numerical technique? **Well, it could be argued that the MAT is a non-parametric statistical technique, but ‘numerical’ works too. We’ve made the change.**
2. P. 7, line 7: delete the comma. **Done**
3. P. 12, lines 6-7: Not sure what you are driving at here. **Clarified concept of ‘analytical utility’ by changing paragraph to:**

**Very restrictive thresholds can improve analytical skill but reduce the analytical utility of analog approaches, because if restrictive thresholds disallow all possible analogs for most target samples, then the MAT will fail to make any environmental inferences for those samples (although its inferences for the remaining samples might be highly precise). Thus, in our leave-one-out analyses, we also note the number of samples with at least one modern analog, to indicate when high precision is gained at the cost of low analytical utility.**

4. P. 15, line 8: Good to have the 95% confidence intervals and non-overlap for some of the variables. You could note the statistical significance for those results. **Done. Added asterixes to table 3 to indicate significance and added new text:**

**The ENA-NA reconstructions also tend to be less accurate (have a slope further from 1) than ENA-ENA reconstructions, with significantly lower slopes ( $p<0.05$ ) for the temperature-related variables (TSUM, TWIN, MTCO, MTWA, GDD5) and summer precipitation (PSUM).**

5. P. 15, Table 3: Why are the standard climatic variables left out here and only results for bioclimatic variables shown? Alright with me if you don't use them, but might also be helpful if they showed significant differences between ENA-NA and ENA-ENA. **Variables have been added to Table 3.**

6. P. 15, lines 22-23: With all the symbols on Figure 4, and the changing scales for the y-axis on Figure 4, I am finding it hard to compare the numerical values on Table 3 with the right open or black triangle, square or circle on Figure 4. Can't you add a figure to show just this comparison? **We opted not to add a new figure because we felt it would be too redundant with Table 3. Instead, we addressed this comment by editing this text to remove joint**

**references to Table 3 and Fig. 4. Each is now reviewed separately, removing the need for cross-comparisons.**

7. P. 16, lines 1-2: Ok, here you give the numbers. Good, but are these differences significant? At least say something about the reduction in RSME in temperature terms because that is 3.95 deg C vs 2.86 deg. C, which is a reduction by over 1 deg C or by 2 deg C in a confidence interval of plus or minus 2.86 vs 3.95. Just citing the numbers without units does not give the reader much to hold on to. And what about 2.86 deg C. Do you really know temperature to hundredths of a degree. You could list these as 4.0 and 2.9 deg. C and perhaps make a stronger point. The concern of precision is a concern with significant figures and confidence interval widths. Don't let the computer's ability to produce many figures after the decimal point fool you into using them. **Good points. We've now noted that the units are °C and modified the sentence as follows:** “**For example, in the NA-NA experiment and no distance weighting, RMSE= 3.31°C for TWin for the 36-taxon list, 2.94°C for the 80-taxon list, 2.84°C for the 135-taxon list, and 2.66°C for the 64S list (Fig. 3), resulting in a net 0.65°C (20%) improvement in precision.**” We're not exactly sure how to count the number of significant figures when assessing RMSE, but given that we are working with >4500 samples, two significant figures seems reasonable, especially when our interpretations are limited to comparisons among the modern validation experiments (and not, e.g. actual paleoclimatic reconstructions).

8. P. 16, line 8: You could note white triangles and squares in parentheses as showing this. "The ENA-NA and ENA-ENA results for the 64S-taxon list are particularly close (white triangles and squares in Fig. 4)." **Done**

9. P. 16, line 8: The suggesting phrase dangles and needs rephrasing. Try: This close match allows you to infer that splitting pollen types into eastern and western groups effectively excludes floristically false analogs from western North America from being assigned in ENA.

**Done**

10. P. 17, lines 1 and 2: Good point. **Thanks**

11. P. 17, line 7: worse? Why not be precise and say larger. **Done**

12. P. 17, line 13: reference Figure 5 on this line. **Done**

13. P. 18: lines 1&2: you could mention that you may explain more variance in a relative sense while the RSME is larger because the larger gradient can tolerate a larger RSME and still explain more variance. It is a case of the denominator (gradient magnitude or total variance) increasing more than the numerator (RSME). Both increase but one goes up far more than the other. You can explain more variance but still have larger error bars in terms of mm or deg C.

**Done. Modified text on p. 18 to insert some of this language.**

14. P. 18, line 13: Fig. not Figs. Also I think that RSME is a direct measure of precision because it relates to error bars or confidence intervals, which are what precision is all about. R-square measures percent variance explained and is not a direct measure of precision. I would rephrase the your statements in terms of what Fig. 4 shows for RSME. **Done – removed references to precision and instead mention percent variance explained.**

15. P. 18, lines 15-21: so why is this? What is D and why might weighting reduce precision? I just had to go back to p. 6, lines 10 & 11 to find out about what D is. It is the dissimilarity. What gets divided by D or D-square and who introduced this idea? I need to know more of what you did to figure out why you got the results that you got. **Clarified this concept by 1) adding sentence to intro, and 2) adding sentence to beginning of paragraph on p. 18:**

**1): When calculating the environmental averages, a simple unweighted average can be used, or the environmental variables associated with the most compositionally similar**

calibration samples can be upweighted using a  $1/D$  or  $1/D^2$  weighting (Jackson and Williams 2004).

2): Because the MAT is based on the assumption that compositionally similar samples represent similar environments, using an inverse distance weighting ( $1/D$  or  $1/D^2$ ) to upweight the most similar samples should improve the skillfulness of MAT reconstructions (e.g. Calcote, 2003; Nakagawa et al., 2002). However, in these analyses, MAT skill is insensitive to the choice of dissimilarity weighting (Fig. 3). There is no clear advantage to using no weighting versus the  $1/D$  weighting (Fig. 3). The  $1/D$  weighting has slightly higher precision for PSum (Fig. 3d), whereas no weighting is slightly more precise for PWin (Fig. 3c), and there is no difference for temperature-related variables and fractional woody cover (Fig. 3a,b,e). There is, however, a small but consistent loss of precision for all variables when using the  $1/D^2$  distance weighting.

16. P. 19, lines 1-9: here you explain how T works a bit and give an example. Do something of this nature for the section on taxonomic weighting. **OK, see above.**

17. **P. 19, line 20: How is it that Figure 6 shows regression statistics? Clarified sentence as follows:** The high precision observed at  $T < 0.05$  occurs because most samples (70%) have no analog (Fig. 6) and hence are not included in the regression statistics (Fig. 6).

18. P. 20, lines 17-18: How do I see this? Nothing much happens on the A-A' or B-B' transects. You imply that the values plotted in the upper right hand of the graphs in Figures 7 and 8 are lowest or highest. I guess that they do but not substantially at the  $N=10$  and  $T=0.5$ . **We've modified the last paragraph in this sentence to better highlight the key trend:**

The T and N parameters strongly interact (Figs. 7, 8). When one is set to a fairly restrictive value (e.g.  $N=3$ ,  $T=0.10$ ), MAT skill is largely insensitive to changes in the other parameter (Figs. 7, 8). Conversely, both precision and accuracy substantially worsen when N and T are relaxed simultaneously, as shown by the decreases in precision and accuracy from the lower left to upper right of each plot in Figs. 7, 8.

19. P. 25, lines 7-13: Say more. When I looked in the methods section, I missed finding where it tells just what D is and how it is measured. Is it the SCD? And then what does it divide as  $1/D$ ? I think that I may be getting it that it allows the sites with the smallest values of D (all less than T) to have more weight in determining the climate value. If so, then it makes some of N analogs be more important than the others. This could be somewhat equivalent to reducing N because only fewer samples determine the final value. But if T is set correctly, then dividing by D or D-square should not have an impact. But if T is large, then maybe it would compensate and make a difference. I would say a little more in the results about why you wanted to try this weighting to set up the discussion here. **We think these comments have been addressed by the changes made in response to comment#15 – i.e., by clarifying what D is earlier in the introduction and results.**

20. P. 26, lines 1-5: good point. **Thanks.**

21. Clear set of useful conclusions. **Thanks.**

1

2      **Obtaining Accurate and Precise Environmental Reconstructions from**  
3      **the Modern Analog Technique and North American Surface Pollen**  
4    **Dataset**

5

6      Williams, J. W.\* and Shuman, B.<sup>†‡</sup>

7

8      \*Corresponding Author:

9      jww@geography.wisc.edu

10     550 North Park St.

11     Department of Geography

12     University of Wisconsin – Madison

13     USA

14     608-265-5537 (office)

15     608-265-3991 (fax)

16

17     <sup>†</sup>Department of Geography, 216 19<sup>th</sup> Ave. South, University of Minnesota, Minneapolis,  
18     MN 55455

19

20     <sup>‡</sup>Present Address: Department of Geology and Geophysics, 1000 University Ave,  
21     University of Wyoming, Laramie, WY 82071

22

1    **Abstract**

2    Using a recently compiled surface-sample pollen dataset for North America, we test  
3    methods to improve the skillfulness (i.e. accuracy and precision) of the modern analog  
4    technique (MAT) for continental- to sub-continental-scale paleoenvironmental  
5    reconstructions. Interregional floristic differences challenge such reconstructions,  
6    because with widening spatial extent each pollen type represents an increasing number of  
7    species, each with its unique niche, thus blurring pollen-climate relationships. We  
8    conduct a series of cross-validation experiments in which we first explore 1) whether  
9    increasing the spatial extent of the pool of potential analogs improves or worsens MAT  
10   skill, 2) whether MAT skill is improved by increasing the number of pollen taxa and  
11   splitting pollen taxa into regional groups, and 3) the differences in MAT skill among  
12   environmental variables. Regional splits are guided by the range boundaries of parent  
13   species. Results are analyzed for multiple environmental variables. We then  
14   systematically explore options for a) the value of the no-analog/analog threshold ( $T$ ), b)  
15   the maximum number of modern analogs ( $N$ ) allowed for a target fossil sample, and c)  
16   whether the environmental average constructed from the modern analogs should be  
17   weighted by their compositional dissimilarity ( $D$ ) to the target sample. We demonstrate  
18   that substantial interregional differences in pollen-climate relationships, particularly  
19   between eastern and western North America, degrade MAT precision and accuracy, but  
20   these adverse effects can be overcome by expanding the list of taxa used and/or splitting  
21   pollen types into regional groups. MAT precision was best when pollen types were  
22   regionally split and better when more taxa were used, although increasing the taxon list  
23   from 64 to 135 types did not substantially increase performance. Temperature-related

1 variables were reconstructed more precisely than hydrological variables, and there was  
2 little difference in MAT skill between climatic and bioclimatic variables (e.g. mean July  
3 temperatures versus growing degree days). T scales with the number of taxa analyzed,  
4 and there is a tradeoff between skill (best when T is low) and utility (if T is low, few  
5 samples can receive environmental inferences). For N, there is a tradeoff between  
6 precision and accuracy, such that retaining just the single best analog had the worst  
7 precision and best accuracy. Strongly weighting by taxonomic dissimilarity ( $1/D^2$ )  
8 consistently reduced precision, but MAT precision was similar for unweighted and  
9 inverse-distance weighted averages. Given the above tradeoffs, we recommend using the  
10 64-taxon list with taxa split by region,  $0.20 \leq T \leq 0.30$ ,  $3 \leq N \leq 7$ , and either no weighting or  
11 an inverse-distance weighting, for North American applications of the MAT.

12

13 **Key Words:** Bioclimatic variables, Cross-Validation, Modern Analog Technique, North  
14 America, Pollen, Surface Samples

1      **1. Introduction**

2            The modern analog technique (MAT, Jackson and Williams, 2004; Overpeck et  
3       al., 1985; Prell, 1985) is one of a handful of widely employed numerical techniques for  
4       constructing quantitative estimates of past environments from micropaleontological  
5       assemblages. Like other statistical techniques that employ a ‘space for time’ substitution  
6       (e.g. Imbrie and Kipp, 1971; ter Braak, 1995; Birks, 1995; ter Braak and Juggins, 1993;  
7       Malmgren et al., 2001; Guiot et al., 1996), the MAT relies on modern calibration  
8       datasets, consisting of spatial networks of samples from sites with known environmental  
9       characteristics (e.g. macroclimatic attributes, vegetation type or structure) to draw  
10      inferences about temporal sequences of fossil samples with unknown environmental  
11      characteristics (e.g. Overpeck et al., 1985; Overpeck et al., 1992; Williams, 2003; Guiot  
12      et al., 1993). In the MAT, the compositional dissimilarity ( $D$ ) between a ‘target’ fossil  
13      sample and each member of the calibration dataset is measured, and the environmental  
14      characteristics of the  $N$  most similar (least dissimilar) modern samples are averaged and  
15      assigned to the target (Overpeck et al., 1985; Jackson and Williams, 2004). When  
16      calculating the environmental averages, a simple unweighted average can be used, the  
17      environmental variables associated with the most compositionally similar calibration  
18      samples can be upweighted using a  $1/D$  or  $1/D^2$  weighting (Jackson and Williams, 2004),  
19      or other distance-based weightings can be applied (Lytle and Wahl, 2005). Modern  
20      samples with dissimilarities to the target sample that are greater than a threshold  $>T$  are  
21      not considered; if all modern samples have dissimilarities greater than a threshold  $>T$ ,  
22      then the target sample is deemed to have no modern analog and no environmental  
23      inference is made. Cross-method comparisons show that MAT skill (i.e. accuracy and

1 precision) is similar to that for other methods (Bartlein and Whitlock, 1993; Malmgren et  
2 al., 2001; Malmgren and Nordlund, 1997; Crosta et al., 1998) although its apparent  
3 performance may be inflated by spatial autocorrelation (Telford and Birks, 2005).

4 The recently published North American surface pollen dataset {Whitmore, 2005  
5 #1625, [www.lpc.uottawa.ca/data/index.html](http://www.lpc.uottawa.ca/data/index.html)} was compiled to aid pollen-based  
6 paleoenvironmental reconstructions, by providing a spatially dense calibration dataset  
7 comprising over 4500 records, 134 pollen types, and climatic and vegetational  
8 information for each site. The availability of this data set raises several interesting issues  
9 concerning how best to obtain accurate and precise estimates of past environmental  
10 variables when using the MAT. Should searches for analogs be made over the entire  
11 continental-scale calibration dataset or will regional datasets yield more accurate results?  
12 In theory, a spatially dense and extensive calibration dataset should improve MAT skill,  
13 due to the enlarged pool of potential analogs from a wide variety of plant communities.  
14 However, in scaling up from regional datasets to North America, the taxonomic  
15 resolution of pollen data effectively decreases, because the number of species  
16 contributing to each pollen morphotype increases. This decrease raises the risk of  
17 floristically inappropriate matches among apparently analogous pollen samples,  
18 particularly between those from eastern and western North America. For example, the  
19 *Picea* (undiff.) pollen type is produced by *P. glauca*, *P. mariana*, and *P. rubens* in the  
20 eastern US, Alaska, and much of Canada, and by *P. engelmannii*, *P. pungens*, and *P.*  
21 *sitchensis* in the west. Improved capabilities for species-level identifications partially  
22 mitigates this issue, e.g. *Acer*, *Fraxinus*, and *Juglans* can often be identified to species  
23 (Finkelstein et al., 2005), as can eastern species of *Picea* (Lindbladh et al., 2003), but few

1 pollen records currently include this level of taxonomic precision. Proposed analytical  
2 solutions include defining regional calibration datasets or otherwise geographically  
3 constraining potential analogs (Williams et al., 2000; Pflaumann et al., 1996; Birks,  
4 1995). Here we experiment with an alternate approach, in which we split some pollen  
5 types into regional groups, using range maps of the parent species to constrain the splits.

6 The analyst is also confronted with several operational decisions and parameters  
7 potentially critical to the skilful functioning of the MAT (Jackson and Williams, 2004;  
8 Sawada et al., 2004; Wahl, 2004; Gavin et al., 2003; Lytle and Wahl, 2005; Overpeck et  
9 al., 1985; Oswald et al., 2003; Bryson, 1985). Analysts must choose 1) the pool of  
10 modern samples, 2) the pollen taxa for dissimilarity calculations, 3) the analog/no-analog  
11 threshold that disallows too-dissimilar modern samples from matching to the target  
12 sample (T), 4) the maximum number of modern analogs allowed per target sample (N), 5)  
13 whether the environmental average constructed from the modern analogs is weighted by  
14 their taxonomic dissimilarity (D) to the target sample, and 6) the environmental variables  
15 to be reconstructed (See Jackson and Williams (2004) for a fuller review of the MAT and  
16 these operational parameters). One analysis indicated that optimizing these parameters  
17 matters more to MAT skill than does increasing the number of pollen grains counted per  
18 sample (Lytle and Wahl, 2005). There is no way to determine *a priori* the optimal  
19 parameter value, and some will vary according to the size and density of the calibration  
20 dataset and the desired ecological resolution (Sawada et al., 2004; Wahl, 2003).

21 Here we attempt to maximize the precision and accuracy of subcontinental- to  
22 continental-scale climatic and vegetational reconstructions with the North American  
23 surface sample dataset and the MAT. Our primary analytical tool is a variant of the

1 leave-one-out cross-validation technique (Malmgren and Nordlund, 1997; Williams,  
2 2003; Sawada et al., 2004; Wahl, 2003), in which we apply the MAT to produce climatic  
3 and vegetational reconstructions for each surface sample (never allowing the sample to  
4 match to itself or other nearby samples), then check the analog-inferred values against the  
5 original observations for each target site. We first show that eastern and western taxa  
6 display substantially different pollen-climate relationships and that these differences can  
7 reduce the accuracy and precision of subcontinental climatic reconstructions. We then  
8 show that expanded taxa lists and splitting pollen types into regional groups improves  
9 MAT precision and accuracy. We compare the precision of pollen-based reconstructions  
10 among many climatic, bioclimatic, and vegetational variables, including comparisons  
11 between temperature and hydrological variables and between bioclimatic and climatic  
12 variables. We explore the effect of various taxon lists and parameter combinations on  
13 MAT skill, in terms of both analytical precision and accuracy. We conduct Receiver-  
14 Operator-Curve (ROC) analysis (Gavin et al., 2003; Wahl, 2004; Oswald et al., 2003) to  
15 help constrain the analog/no-analog threshold. We identify tradeoffs, e.g. between  
16 precision and accuracy, inherent to some parameter choices, and provide a recommended  
17 range of MAT parameter values for use with the North American surface pollen dataset.

18 **2. Materials and Methods**

19 **2.1. North American Surface Pollen Dataset**

20 Whitmore et al. (2005) describe the contents of the North American surface  
21 sample dataset and its methods of assembly , and the dataset continues to be updated.  
22 The version used here consists of 4549 surface pollen samples and counts or relative  
23 abundances for 134 pollen types (Table 1). Climatic properties at each site are from

1 1961-1990 data from the Climate Research Unit (New et al., 2002) and are lapse-rate  
2 corrected and bilinearly interpolated to each location (Whitmore et al., 2005). Fractional  
3 tree cover data are based upon data from the Advanced Very High Resolution  
4 Radiometer (AVHRR) (DeFries et al., 2000; Williams, 2003). The North American  
5 surface sample dataset was the basis for a recent atlas of pollen distributions in  
6 geographic space and arrayed along various climatic and vegetational gradients (Williams  
7 et al., 2006). This dataset represents the collective effort of scores of palynologists and  
8 integrates several preexisting regional calibration databases. All pollen data are  
9 represented here as percentages relative to the sum of pollen types used in the MAT  
10 analysis (Table 1).

11 **2.2. Range-Based Splits of Pollen Types**

12 Many tree genera have distinct suites of species between eastern and western  
13 North America, and these species often differ in their distributions along climate  
14 gradients, particularly with respect to hydrological variables (Thompson et al., 1999b; a;  
15 Thompson et al., 1999c). Here, we use the regional splits of several pollen types made by  
16 Williams et al. (2006): *Abies*, *Acer*, *Alnus*, *Celtis*, Cupressaceae, *Fraxinus*,  
17 *Larix/Pseudotsuga*, *Picea*, *Pinus*, *Quercus*, and *Tsuga* were divided into eastern and  
18 western groups (Table 1), taking advantage of the fact that the eastern and western  
19 species from these taxa minimally overlap at present and are unlikely to have overlapped  
20 further during the late Quaternary (Williams et al., 2006). Assignments of pollen types to  
21 eastern and western groups are based upon the aggregated eastern/western range maps  
22 created by Thompson et al. (1999b; 1999a; 1999c). The position of the east-west division  
23 thus differs for each pollen type, and is tailored to the ca. 1960's distributions of the

1 parent taxa. Because the distribution of a pollen type may extend beyond the range limits  
2 of its parent taxon, we added buffers to the aggregated range maps. Buffers placed  
3 around eastern aggregated ranges were not allowed to extend into western counterparts,  
4 and vice versa. Shapefiles for the polygons used to split sites are included in the  
5 Supplementary Information.

6 A few taxa – *Alnus*, *Picea*, *Pinus*, and *Quercus* – currently have small range  
7 overlaps between eastern and western species, which results in locally ambiguous  
8 attributions of the corresponding pollen types. The number of sites in overlap regions is  
9 small (14% of 4549 sites). In the analyses presented here all ambiguous pollen types are  
10 assigned to eastern species.

11 In addition, we split eastern *Pinus* into northeastern and southeastern groups  
12 (Williams et al., 2006). The northeastern and southeastern *Pinus* species are  
13 indistinguishable palynologically (except for *P. strobus*) but have very different climatic  
14 tolerances (Thompson et al., 1999b), confounding pollen-based climate reconstructions.  
15 Previous pollen-based paleoclimate reconstructions have also split the northeastern and  
16 southeastern pines (Webb et al., 1993b). We also split Chenopodiaceae/Amaranthaceae  
17 into two regions: Florida and central North America. This split is motivated by the  
18 occurrence of two distinct modes in Chenopodiaceae/Amaranthaceae pollen abundances,  
19 one centered over the Great Plains, and the other in the Florida Everglades (Williams et  
20 al., 2006). Of the 14 Chenopodiaceae/Amaranthaceae species native to Florida  
21 (Wunderlin and Hansen, 2004) and the ca. 60 native Midwestern species, only three are  
22 common to both regions (Anonymous, 2006). Many of the Florida  
23 Chenopodiaceae/Amaranthaceae species occupy saline or locally xeric habitats within the

1 Florida Everglades and coastal regions (Willard, pers. comm.). These taxa therefore  
2 occupy similar microclimates but very different macroclimates.

3 **2.3. Taxa Lists**

4 We test four taxon lists (Table 1): 1) a 135-taxon list, comprising all 134 pollen  
5 types in the Whitmore et al dataset and the split of southeastern *Pinus* versus other *Pinus*;  
6 2) a 36-taxon list, comprising 35 selected pollen types and the southeastern split for  
7 *Pinus*; 3) a 64-taxon list, comprising 63 selected pollen types and the southeastern split  
8 for *Pinus*; 4) a '64-split' list (64S), comprising the same 63 types and geographical splits  
9 for many pollen types (Table 1). The number of possible taxon lists is over  $10^{40}$  for 134  
10 taxa), and therefore precludes an exhaustive search for the optimal list. Choice of pollen  
11 types for analysis was based on our judgment and prior usage (Overpeck et al., 1985;  
12 Williams, 2003; Williams et al., 2000; Sawada et al., 2004; Webb et al., 1993b; Bartlein  
13 and Whitlock, 1993; Bartlein et al., 1984). We use the southeastern split for *Pinus* in all  
14 lists because other analyses (not shown) indicate that not splitting northeastern and  
15 southeastern pines strongly and negatively affects the accuracy and precision of the MAT  
16 reconstructions for ENA, regardless of which other taxa are included (see also Webb et  
17 al., 1993b). MAT skill may also be improved in the future by also splitting some western  
18 taxa into subregional types (Davis, 1995), but this option is not explored here.

19 **2.4. Cross-Validation Experiments with the Modern Analog Technique**

20 We perform several series of experiments with the modern analog technique,  
21 using a variant of the leave-one-out cross-validation technique (Malmgren and Nordlund,  
22 1997; Williams, 2003; Sawada et al., 2004). In this technique, each modern pollen  
23 sample is compared to every other modern sample, and self-comparisons are not allowed.

1 We use the squared-chord distance (SCD) metric (Overpeck et al., 1985) to measure  
2 palynological dissimilarity among samples. Our variant of the leave-one-out technique is  
3 more stringent than the norm, because it disallows matches between the target pollen  
4 sample and any other sample within 50 km (Williams, 2003). The 50-km distance is  
5 chosen to ensure that target surface samples and their analogs do not have overlapping  
6 pollen source areas (Sugita, 1994).

7 Skillful reconstructions will be characterized by both a high precision and  
8 accuracy of the analog-predicted environmental variables relative to the original  
9 observations. To assess precision, we rely on the root mean squared error (RMSE) and  
10 the coefficient of determination ( $R^2$ ), both calculated using ordinary least-squares (OLS)  
11 regression. RMSE measures the expected prediction error for paleoenvironmental  
12 reconstructions, but is an absolute measure of model error that is scaled to the units of the  
13 reconstructed environmental variable, which hinders intervariable comparisons. Even  
14 variables with the same units but different variances along a common spatial gradient  
15 (e.g. January vs. July temperatures) will tend to have different RMSE's. Conversely,  $R^2$   
16 is a standardized metric of precision, defined as the fraction of variance explained by the  
17 regression model relative to the total variance contained in the original environmental  
18 observations, facilitating intervariable comparisons. However, because of this  
19 standardization,  $R^2$  values cannot be easily compared between experiments employing  
20 datasets with different total variance (e.g. an all-North American dataset vs. a dataset  
21 from eastern North America; see below). We therefore use RMSE for most assessments  
22 of precision and reserve  $R^2$  for assessing differences in precision among environmental

1 variables within a single experiment. Precise reconstructions will have a low RMSE and  
2 high R<sup>2</sup>.

3 To assess accuracy, we calculate the slope of the linear relationship between  
4 observed and predicted environmental variables. Accurate reconstructions will have a  
5 slope close to 1. Because the slope estimated by ordinary least-squares regression will be  
6 biased towards too-low values when there is uncertainty in the observational data  
7 (Legendre and Legendre, 1998), we use major-axis regression to estimate slope  
8 (Williams et al., 2008). For reasons of space, we primarily focus on analytical precision,  
9 and less on accuracy. We analyze 40 climatic, bioclimatic, and vegetational variables,  
10 with detailed analyses done for growing-degree-days (GDD5), mean temperature of the  
11 coldest month (MTCO), mean temperature of the warmest month (MTWA), the ratio of  
12 actual to potential evapotranspiration (AE/PE), mean DJF temperature (TWin), mean JJA  
13 temperature (TSum), mean DJF precipitation (PWin), mean JJA precipitation (PSum),  
14 and fractional woody cover (Woody).

15 Very restrictive thresholds can improve analytical skill but reduce the analytical  
16 utility of analog approaches, because if restrictive thresholds disallow all possible analogs  
17 for most target samples, then the MAT will fail to make any environmental inferences for  
18 those samples (although its inferences for the remaining samples might be highly  
19 precise). Thus, in our leave-one-out analyses, we also note the number of samples with at  
20 least one modern analog, to indicate when high precision is gained at the cost of low  
21 analytical utility.

22 We first apply the leave-one-out method to test whether expanding the pool of  
23 potential analogs to all of North America (NA), without accounting for regional floristic

1 differences, adversely affects subcontinental paleoclimate reconstructions. Eastern North  
2 America (ENA) was chosen as the case study because it is a classic region for pollen-  
3 based paleoclimatic reconstructions and technique development (Webb et al., 1993a;  
4 Webb et al., 1998; Webb et al., 1993b; Overpeck et al., 1985; Jackson et al., 2000;  
5 Williams, 2003; Bartlein and Whitlock, 1993; Bartlein et al., 1984; Gajewski, 1988). We  
6 conduct three experiments in which 1) the set of target samples is restricted to ENA (east  
7 of 100°W) and the pool of potential analogs encompasses all of North America (**ENA-**  
8 **NA**), 2) the target samples are restricted to ENA and the pool of analogs is limited to  
9 samples east of 110°W (**ENA-ENA**), and 3) both the target samples and potential analogs  
10 span North America (**NA-NA**). Comparing the ENA-NA and ENA-ENA experiments  
11 provide information about the effect of the spatial extent of the pool of potential analogs  
12 upon MAT skill; comparisons between the NA-NA and ENA-NA experiments provide  
13 information about the effect of the spatial extent of the target pool upon MAT skill.  
14 Adding a 10° buffer to the pool of eastern analogs in the second experiment removes  
15 edge-effect artifacts. For these experiments, we vary the choice of taxon list, application  
16 of geographic splitting, and distance-weighting in order to determine which  
17 combination(s) maximize MAT precision for subcontinental- to continental-scale  
18 environmental reconstructions. N and T are set to standard values (N=10, T=0.25).

19 In the second set of analyses, we systematically explore the parameter space for  
20 the N, T, and distance-weighting parameters (D) to assess the sensitivity of the MAT to  
21 particular parameters and to identify combinations that maximize agreement between the  
22 observed and analog-inferred climate and vegetational characteristics (Table 2). In all,  
23 1410 parameter combinations were evaluated. Five variables were chosen for these

1 analyses: mean DJF temperature (TWin), mean JJA temperature (TSum), mean DJF  
2 precipitation (PWin), mean JJA precipitation (PSum), and fractional woody cover  
3 (Woody).

4 **2.5. Receiver-Operator-Curve (ROC) Analysis**

5 For estimating T, we also used ROC analysis. ROC analysis, originally  
6 developed for medical applications (Henderson, 1993; Zweig and Campbell, 1993),  
7 identifies the threshold for a test statistic that optimally discriminates between two states,  
8 e.g. infected or not infected. In applications to the MAT, the dissimilarities between all  
9 possible pairs of pollen samples are calculated and used as the test statistic for ROC  
10 analysis. ROC then identifies an dissimilarity threshold ( $T_{opt}$ ) that optimally discriminates  
11 between pollen samples drawn from the same land cover type versus pollen samples from  
12 different land cover types (Gavin et al., 2003; Oswald et al., 2003; Wahl, 2004). This  
13 approach is conceptually similar to the original approach used to determine analog/no-  
14 analog thresholds (Overpeck et al., 1985), but is more statistically rigorous and allows  
15 explicit consideration of tradeoffs between the identification of false-positive and false-  
16 negative analogs (Wahl, 2004). Here we use continental-scale vegetation units (biomes)  
17 from the DISCover land-cover maps for North America, IGBP vegetation classification  
18 (Loveland et al., 2000) to determine  $T_{opt}$ .

19 **3. Results**

20 **3.1. Interregional Variations in Pollen-Climate Relationships**

21 Eastern and western tree species have distinctly different climatic distributions  
22 (Thompson et al., 1999b; a), so it is unsurprising that these taxa also display substantially  
23 different pollen-climate relationships (Figs. 1, 2). Differences between eastern and

1 western species distributions are particularly pronounced for hydrological variables (e.g.  
2 AE/PE, Fig. 1). Most western taxa occupy drier climates than their eastern counterparts,  
3 although the degree of difference varies among taxa (Fig. 1). For several taxa the eastern  
4 modal AE/PE value tends to lie at or outside the range of AE/PE values for western taxa.  
5 Unlike AE/PE (Fig. 1), eastern and western taxa generally have similar GDD5  
6 distributions (Fig. 2), Eastern and western *Tsuga* differ the most, with eastern species of  
7 *Tsuga* growing at lower GDD5. Southeastern *Pinus* and Florida  
8 Chenopodiaceae/Amaranthaceae have narrow climate ranges relative to their counterparts  
9 elsewhere in North America, and as a consequence of their location grow in relatively  
10 moist and warm climates.

11 When a fairly short taxon list (36 taxa, Table 1) is used, these interregional  
12 differences adversely affect the accuracy and precision of analog-based climate  
13 reconstructions, if samples outside of a region are allowed to be candidates for analogs  
14 (Table 3). For all climatic variables, climate reconstructions for ENA samples were  
15 substantially less precise (i.e. have higher RMSE and lower  $R^2$  values) when the pool of  
16 potential analogs encompassed all of North America (ENA-NA) than when the pool of  
17 analogs was restricted to eastern North America (ENA-ENA) (Table 3). Both  
18 temperature and hydrological reconstructions lose precision, and the loss is somewhat  
19 more severe for hydrological variables: on average, RMSE values are 39.4% higher  
20 (worse) for moisture-related variables (PSum, PWin, AE/PE, AnnP), versus 33.6% higher  
21 for temperature-related variables (TSum, MTWA, TWin, MTCO, GDD5) (Table 3). The  
22 ENA-NA reconstructions also tend to be less accurate (have a slope further from 1) than  
23 ENA-ENA reconstructions, with significantly lower slopes ( $p < 0.05$ ) for the temperature-

1 related variables (TSum, TWin, MTCO, MTWA, GDD5) and summer precipitation  
2 (PSum). This analysis thus demonstrates that the negative effects of interregional  
3 differences in species-climate relationships can overwhelm the positive effects of a larger  
4 and more comprehensive North American pollen calibration dataset when the taxonomic  
5 resolution in the dataset is too low to allow for discrimination among geographically  
6 distant samples.

7 **3.2. Effect of Taxonomic Resolution on MAT Skill**

8 In general, expanding the list of taxa improves MAT precision, and splitting  
9 pollen types into regional groups further improves precision (Fig. 3). In an analysis of  
10 the entire North American dataset (Fig. 3), RMSE's are consistently worse for the 36-  
11 taxon list relative to the 64- and 135-taxon lists, and similar or slightly worse for the 64-  
12 taxon list relative to the 135-taxon list. The 64S-list (with pollen types split into regional  
13 groups, Table 1) consistently equaled or outperformed the other variables. The  
14 importance of taxonomic resolution varies among environmental variables; with the  
15 largest improvements observed for TWin and the smallest improvements for TSum and  
16 PWin. For example, in the NA-NA experiment and no distance weighting, RMSE=  
17 3.31°C for TWin for the 36-taxon list, 2.94°C for the 80-taxon list, 2.84°C for the 135-  
18 taxon list, and 2.66°C for the 64S list (Fig. 3), resulting in a net 0.65°C (20%)  
19 improvement in precision.

20 Moreover, when more extensive and/or geographically precise taxon lists are  
21 used, expanding the pool of potential analogs to include western taxa no longer  
22 substantially degrades MAT precision for ENA samples. This is demonstrated by the  
23 close overlap between RMSE's for the ENA-NA (Fig. 4, triangles) and ENA-ENA (Fig.

1 4, squares) experiments for both the 135- and 64S-taxonomic lists. The ENA-NA and  
2 ENA-ENA results for the 64S-taxon list are particularly close (open triangles and squares  
3 in Fig. 4). This closeness suggests that splitting pollen types into eastern and western  
4 groups effectively excludes the assignment of floristically false analogs from western  
5 North America to ENA samples. On the other hand, expanding the pool of potential  
6 analogs from ENA to NA never improves MAT precision, i.e. NA-ENA RMSE's (Fig. 4,  
7 squares) are never better than ENA-ENA RMSE's (Fig. 4, triangles). Therefore, no skill  
8 is gained for eastern sites by including potential analogs from the western US, even with  
9 a high taxonomic resolution.

10 For all three experiments (ENA-ENA, ENA-NA, NA-NA), MAT precision is  
11 superior for the 64S-taxon list relative to the 135-taxon list (Fig. 4, open vs. black  
12 symbols). The spread between RMSE values for the 135- and 64S-taxon lists tends to be  
13 smaller for growing-season monthly temperatures and larger for winter monthly  
14 temperatures. For monthly precipitation, RMSE values for the 64S-taxon list are  
15 consistently better than those for the 135-taxon list, with little seasonal variation.  
16 Differences between the 135- and 64S-taxon lists are pronounced for annual bioclimatic  
17 variables (GDD0, GDD5, PE, AE, and AE/PE), indicating that MAT precision for these  
18 variables is particularly improved by geographically splitting pollen types. In general, it  
19 appears that geographically splitting 12 common pollen types improves MAT skill more  
20 than increasing taxonomic resolution by adding 71 relatively minor types.

21 3.3. Effect of Spatial Extent on Precision of MAT Reconstructions

22 For all variables, expanding the set of target samples from eastern North America  
23 to North America reduces overall MAT precision. Compare the ENA-NA (Fig. 4,

1 triangles) and NA-NA (Fig. 4, circles) reconstructions: for almost all variables  
2 (broadleaved woody cover is the lone exception), RMSE's are larger for the NA-NA  
3 reconstructions than the ENA-NA reconstructions. The pool of potential analogs is  
4 identical between the ENA-NA and NA-NA experiments, so that loss of analytical  
5 precision is solely due to including reconstructions for western pollen samples. This  
6 indicates that, in this analysis, MAT reconstructions are generally less precise for western  
7 pollen samples.

8 However, when  $R^2$  is used to measure precision (Fig. 5), a different pattern is  
9 observed: NA-NA reconstructions of temperature-related variables are still worse (lower  
10  $R^2$  values) than the ENA-NA and ENA-ENA reconstructions, but the NA-NA  
11 reconstructions of hydrological variables are apparently more precise (Fig. 5, compare  
12 circles to squares and triangles). This is opposite of the pattern observed for RMSE (Fig.  
13 4). This reversal is caused by the sensitivity of the  $R^2$  metric to both the variance  
14 explained by the regression model (which is inversely proportional to RMSE) and the  
15 total variance in the dataset (i.e. the denominator for  $R^2$ ). In North America, moisture  
16 gradients primarily run west-east, so reconstructions limited to only ENA samples (i.e.  
17 the ENA-ENA and ENA-NA experiments) do not capture the full moisture gradient and  
18 linear regression models fitted to this incomplete gradient tend to have  $R^2$  values lower  
19 than those for linear regressions encompassing the full gradient (i.e. the NA-NA  
20 experiment). This effect is particularly pronounced for AE/PE (Fig. 5). RMSE is not  
21 affected by gradient length. Thus, when comparing across the NA-NA, ENA-NA, and  
22 ENA-ENA experiments, the RMSE statistic is a better index of precision.

23 3.4. Intervariable Differences in MAT Precision

1           Reconstructions for temperature-related variables consistently have more variance  
2       explained than reconstructions of hydrological variables (Fig. 5).  $R^2$  values for  
3       temperature-related variables typically range between 0.8 and 0.95, versus 0.4 to 0.8 for  
4       hydrological variables. Pollen-based reconstructions of bioclimatic variables do not seem  
5       to outperform or underperform those for corresponding climatic variables. For example,  
6        $R^2$  values for monthly precipitation and bioclimatic indices of soil moisture and  
7       precipitation seasonality (AE/PE, AE/PEev, AE/PEdc, PJANPANN, PJULPANN) in the  
8       NA-NA experiment for the 64S taxon list are all between 0.68 and 0.76 and average ca.  
9       0.71 (Fig. 5). Reconstructions of growing-season monthly temperatures and precipitation  
10      tend to have somewhat higher  $R^2$  values than those for winter monthly temperatures and  
11      precipitation (Fig. 5).

12      3.5. Sensitivity of MAT Skill to Parameter Choice

13      3.5.1. *Weighting by Taxonomic Dissimilarity (D)*

14       Because use of the MAT assumes that compositionally similar samples represent  
15       similar environments, using an inverse distance weighting to upweight the most similar  
16       samples should in theory improve MAT skill (e.g. Calcote, 2003; Nakagawa et al., 2002;  
17       Lytle and Wahl, 2005). However, in these analyses, MAT skill is insensitive to the choice  
18       of dissimilarity weighting (Fig. 3). There is no clear advantage to using no weighting  
19       versus the 1/D weighting (Fig. 3). The 1/D weighting has slightly higher precision for  
20       PSum (Fig. 3d), whereas no weighting is slightly more precise for PWin (Fig. 3c), and  
21       there is no difference for temperature-related variables and fractional woody cover (Fig.  
22       3a,b,e). There is, however, a small but consistent loss of precision for all variables when  
23       using the 1/ $D^2$  distance weighting.

1      3.5.2. *Threshold (T)*

2            More restrictive thresholds limit the number of samples that can find analogs  
3        elsewhere the surface sample dataset (Fig. 6). Because the opportunities for  
4        compositional discrepancies between pollen samples accumulate with the number of taxa  
5        considered, longer taxonomic lists require larger thresholds (Sawada et al., 2001). For  
6        example, over 50% of surface samples match to another sample in the dataset when  
7         $T=0.07$  for the 36-taxon list, when  $T=0.09$  for the 64 and 64S-taxon lists, and when  
8         $T=0.10$  for the 135-taxon list (Fig. 6). Over 95% of surface samples have a match when  
9         $T=0.20$  for the 36-taxon list,  $T=0.25$  for the 64 and 64S-taxon lists, and  $T=0.30$  for the  
10      135-taxon list.

11            In the ROC analyses, the optimal T for discriminating among IGBP land-cover  
12      types also increases with length of the taxon list (Table 4).  $T_{opt}$  ranges from 0.12 for the  
13      36-taxon list to 0.24 for the 135-taxon list, and approximates 0.19 for the 64 and 64S-  
14      taxon lists. Given these values of T, 88-93% of surface pollen samples have analogs  
15      elsewhere in the dataset (Fig. 6).

16            MAT precision is non-linearly related to T (Fig. 7). RMSE's rise quickly  
17      (worsen), then abruptly level off and are insensitive to further changes in T (Fig. 7). This  
18      pattern is common to the five variables shown here, although details vary. For the 64S-  
19      taxon list, the breakpoint between RMSE sensitivity and insensitivity to T occurs at  
20       $T=0.05$  (Fig. 7). The high precision observed at  $T<0.05$  occurs because most samples  
21      (70%) have no analog (Fig. 6) and hence are not included in the regression statistics.

22            MAT accuracy is also highest for low thresholds (regression slopes are closest to  
23      1) and decreases as T increases (Fig. 8). Slopes are  $<1$ , indicating that MAT

1 reconstructions systematically underpredict observations. MAT accuracy non-linearly  
2 decreases as T increases, but, unlike MAT precision (Fig. 7), does not abruptly level off.  
3 Reconstruction accuracy varies widely among variables, as does the sensitivity of the  
4 accuracy to T. TWin reconstructions are highly accurate (slope>0.95), regardless of T,  
5 whereas Woody and PWin reconstruction accuracy quickly drops as T increases. TSUM  
6 and PSum are intermediate. As with precision, gains in accuracy at very low T's are  
7 balanced against the low utility of the MAT. Thus, when choosing T, there is a tradeoff  
8 between the skill of the MAT and its utility, and the tradeoff is strongest for low T's.

9     3.5.3. *Maximum Number of Analogs (N)*

10    There is also a tradeoff inherent to the choice of N, but here the tradeoff lies  
11    between precision and accuracy. Precision is worst when N=1 but quickly rises,  
12    plateauing for N $\geq$ 3 (Fig. 7). For some variables (TSUM, Woody), RMSE's improve  
13    gradually as N increases above three, whereas for others (TWin, PSum), RMSE is nearly  
14    constant for N>3. Conversely, accuracy is best when N=1 (Fig. 8) but worsens as N  
15    increases, with the largest drop in accuracy between N=1 and N=3.

16    The T and N parameters strongly interact (Figs. 7, 8). When one is set to a fairly  
17    restrictive value (e.g. N=3, T=0.10), MAT skill is largely insensitive to changes in the  
18    other parameter (Figs. 7, 8). Conversely, both precision and accuracy substantially  
19    worsen when N and T are relaxed simultaneously, as shown by the decreases in precision  
20    and accuracy from the lower left to upper right of each plot in Figs. 7, 8.

21    **4. Discussion**

22    4.1. Interregional Differences in Pollen-Climate Relationships

1       The observed interregional differences in pollen-climate relationships are  
2       consistent with the hypothesis that each species has a unique environmental niche  
3       (Hutchinson, 1957; Jackson and Overpeck, 2000; Chase and Leibold, 2003; however, see  
4       Hubbell, 2001). Because species-level identifications of pollen grains are often not  
5       possible with standard light microscopy, the resulting taxonomic uncertainty translates to  
6       climatic uncertainty for any pollen-based climate reconstructions, except where plant  
7       macrofossils help constrain species identifications (Jackson et al., 1997; Birks, 2003).  
8       Our analyses clearly demonstrate strong differences in pollen-climate relationships  
9       between eastern and western congeneric species; these differences are consistent with  
10      those observed for eastern and western tree species (Thompson et al., 1999b; a).

11       Our analyses also demonstrate that these differences significantly diminish the  
12      precision of analog-based climate reconstructions, if the pool of potential analogs  
13      encompasses North America and relatively restricted taxon lists are used (Fig. 3).  
14       However, these interregional differences in pollen-climate relationships can be largely  
15      overcome by extending the list of pollen types considered and/or splitting pollen types  
16      into regional groups (Figs. 3, 4). Both are viable solutions, although our analyses suggest  
17      that splitting pollen types into geographic groups (using the range maps of the parent  
18      species to guide splits) generally improves MAT skill more than simply adding more taxa  
19      (Figs. 3, 4). One key reason is that the SCD metric is most sensitive to more-abundant  
20      pollen types and less sensitive to rare types (Overpeck et al., 1985), so adding rare types  
21      often does not substantially alter dissimilarity scores. (In general, the most abundant  
22      pollen types are already included in the 36- and 64-taxon lists, so the types added to the  
23      135-taxon list are mostly minor components.) Geographic splitting improves MAT skill

1 because it enables more precise identification of the most abundant pollen types (e.g.  
2 western vs. eastern *Quercus*), which contribute more weight to the SCD.

3 For the purposes of paleoenvironmental reconstruction, geographically  
4 segregating the North American surface sample dataset is only useful if fossil pollen  
5 records can also be segregated. Fortunately, the geographic separation between eastern  
6 and western forests has probably been in place since the late Miocene (Axelrod, 1985).  
7 The Great Plains is a natural barrier between eastern and western arboreal populations,  
8 and geographic segregation was likely more pronounced during glacial periods, when the  
9 Laurentide and Cordilleran Ice Sheets acted as an additional barrier between eastern and  
10 western taxa (although the division between eastern and western species is imperfect;  
11 some ‘eastern’ tree species, e.g. *Picea glauca*, likely had small glacial refugia in eastern  
12 Beringia) (Anderson et al., 2006). Therefore, late-Quaternary distributions of eastern and  
13 western species are unlikely to have been more mixed than they are today, and we can  
14 apply the present divisions between eastern and western taxa to most late-Quaternary  
15 pollen records.

16 Splitting northeastern and southeastern *Pinus* and  
17 Chenopodiaceae/Amaranthaceae must be applied more cautiously. The boundary  
18 between northeastern and southeastern *Pinus* is not static during the Quaternary, but at  
19 many sites, can be constrained from plant macrofossil data (Jackson et al., 1997). The  
20 division of Chenopodiaceae/Amaranthaceae into central North American and Floridian  
21 species is unconstrained by paleoecological evidence, so application to fossil pollen  
22 records requires assuming that the present geographic separation persisted during the late  
23 Quaternary, perhaps due to relatively restricted edaphic tolerances for the Floridian taxa.

1 When applying these splits to fossil pollen records, we recommend that analysts only use  
2 the regional splits appropriate to their location, and whenever possible, check their splits  
3 against species-level identifications from macrofossil data (Birks and Birks, 2000). For  
4 example, for a site in northern Georgia at the last glacial maximum, the regional split for  
5 Chenopodiaceae/Amaranthaceae might be best avoided, but the east/west split for  
6 *Quercus*, *Picea*, *Pinus*, etc. would be solid, and, depending on macrofossil evidence,  
7 splitting northeastern and southeastern *Pinus* might be justified.

8 A simple alternative to the regional splitting of individual pollen taxa is to  
9 geographically constrain the pool of potential analogs, either by allowing matches  
10 between samples within a prescribed distance (Williams et al., 2000; Pflaumann et al.,  
11 1996) or by using regional training sets (Birks, 1995). The disadvantage of this simpler  
12 approach is the absence of clear criteria for choosing any particular prescribed distance or  
13 ‘region’. In contrast, using the geographical ranges of the parent plant species as the  
14 basis for splitting pollen types is firmly grounded in niche theory (Hutchinson, 1957).

15 Splitting taxa into regional groups reduces but does not eliminate the risk of  
16 inaccurate paleoclimatic reconstructions caused by floristically inappropriate matches,  
17 because within the broad regions defined here, multiple species still can correspond to a  
18 single pollen morphotype. In eastern North America, for example, the pollen grain  
19 *Fraxinus nigra*-type is produced by both *Fraxinus nigra* and *Fraxinus quadrangulata*,  
20 with distinctly different geographic ranges and bioclimatic envelopes (Thompson et al.,  
21 1999a). Similarly, the presence of *Picea critchfieldii*, a now-extinct taxon, in full-glacial  
22 communities from the southeastern US (Jackson and Weng, 1999), complicates pollen-  
23 based paleoclimate reconstructions for this time and place (Jackson and Williams, 2004).

1 Furthermore, even if pollen morphologies were uniformly identifiable to species,  
2 phenotypic and genotypic variation among individuals and populations (e.g. Etterson and  
3 Shaw, 2001) blur somewhat the plant-climate relationships established at the species  
4 level. These factors weaken (but do not invalidate) the underlying assumption of the  
5 MAT that the comparative similarity between two assemblages corresponds to a causal  
6 similarity (Jackson and Williams, 2004) and reduce the skill of paleoclimatic  
7 reconstructions. Thus, wherever possible, improved taxonomic precision is desirable and  
8 should improve the skill of paleoenvironmental inferences (e.g. Finkelstein et al., 2005;  
9 Lindbladh et al., 2007).

10 **4.2. Choosing Parameters for the Modern Analog Technique**

11 Based on the above cross-validation analyses of the North American surface  
12 sample dataset and on-going work with fossil pollen data, we recommend using a  
13 geographically split taxon list with at least ‘moderate’ palynological diversity (i.e. a list  
14 resembling the 64S-taxon list), using either no distance weighting or a 1/D weighting,  
15 and restricting the number of analogs (N) to between 3 and 7. The choice of analog/no-  
16 analog threshold (T) depends on the taxon list used; for the 64S-list, we recommend  
17 setting T between 0.2 and 0.3.

18 The rationale for splitting taxa into regional groups is discussed above. In the  
19 absence of geographic splits, both the mid-sized taxon list (64 taxa) and full taxon list  
20 (135 taxa) outperformed the short (36 taxa) lists, both by improving precision and by  
21 minimizing the negative effects of using an all-North American calibration dataset. This  
22 result shows that, in some cases, enhanced taxonomic resolution yields additional  
23 vegetational and environmental information (Finkelstein et al., 2005; Whitmore et al.,

1 2005; Sawada et al., 2004). However, precision gains between the 135-taxon and 64-  
2 taxon lists are small to negligible (Fig. 3). Some pollen types are simply too rare,  
3 inconsistently identified, or affected by site-specific factors to be environmentally  
4 informative at continental scales. However, at finer spatial scales, less-abundant pollen  
5 types can be crucial aids to distinguishing among plant communities (Oswald et al.,  
6 2003). It is likely that the optimal taxon list for North America is not considered here.  
7 Furthermore, a taxon list optimal for continental-scale applications may not be optimal  
8 for site-specific applications. We encourage analysts to experiment with their own lists.

9 It is not entirely clear why the  $1/D^2$  weighting performs more poorly than  $1/D$  and  
10 no weightings (Fig. 3), but it must be mainly because dissimilarity in pollen composition  
11 is not a perfect predictor of climatic or vegetational dissimilarity. In these analyses, most  
12 inappropriate matches are already excluded by the N and T parameters. Therefore, once  
13 the N best analogs are identified, subtle differences in palynological dissimilarity  
14 between the N best analogs and target sample (which then are magnified by the  $1/D^2$   
15 weighting) are not environmentally informative.

16 The tradeoff between accuracy and precision observed for N occurs because N  
17 effectively operates as a smoothing parameter, in which high N's result in more smoothed  
18 MAT reconstructions (Viau and Gajewski, 2007). On the one hand, the low precision of  
19 the MAT when only a single best analog is retained (N=1, Fig. 7) suggests that the  
20 identity of the single best analog is as much due to chance as to environmental  
21 determinants, so that allowing multiple matches per sample reduces stochasticity and  
22 improves precision. On the other hand, higher N's smooth the MAT reconstructions,  
23 reducing their accuracy (Fig. 8). The tradeoff between precision and accuracy

1 demonstrated here clarifies recent methodological debates about whether to use multiple  
2 analogs (Jackson and Williams, 2004) or the single best match (Viau and Gajewski,  
3 2007; Viau et al., 2006). Given that a) gains in precision are largely limited to increasing  
4 N from 1 to 3, but b) accuracy continues to decrease as N increases, we recommend using  
5 three and no more than seven analogs. Using 10 analogs, a default choice for many MAT  
6 applications (e.g. Prell, 1985; Allen et al., 2000), reduces accuracy and does not  
7 substantially improve precision.

8 Our recommended range of threshold values for the 64S-taxon list ( $0.2 \leq T \leq 0.3$ )  
9 trades off utility against precision, accuracy, and the need to avoid unsupported  
10 environmental interpretations for fossil pollen samples with no modern analog. We use  
11 the ROC-identified  $T_{opt}=0.2$  (optimal for discriminating among pollen samples from  
12 different biomes) to set the lower-limit recommendation. Utility rapidly decreases if T is  
13 lowered below 0.2 (Fig. 6) with only minor improvements in precision and accuracy (as  
14 long as a reasonably restrictive N is chosen, Figs. 7-8). Interestingly, our finding that  
15 MAT precision is highest at very low thresholds ( $T < 0.05$ ) is *opposite* to the results of  
16 Lytle and Wahl (2005), which we attribute to a) the broader geographic extent of this  
17 study (increasing the potential penalty for false analogs) and b) the limited number of  
18 taxa (11) in that study (which reduces the information content of pollen samples such that  
19 even the most compositionally similar samples may have low predictive value).

20 Setting 0.3 as the upper bound further increases the number of samples that can  
21 receive environmental determinations from the MAT (98% of surface samples have an  
22 analog elsewhere in the dataset.) at a still-minor cost to precision and accuracy. Prior  
23 ROC analyses have shown that reducing T below  $T_{opt}$  severely diminishes the

1 discriminating ability of the MAT (because too many samples from the same vegetation  
2 type are ruled to be from different vegetation types), but that there is a broad plateau of  
3 near-optimal values of T above  $T_{opt}$  (Wahl, 2004; Gavin et al., 2003). Nevertheless,  
4 relaxing T to above  $T_{opt}$  should be done cautiously, because it increases the risk of false  
5 positives, i.e. ruling two samples analogous when in fact they come from substantively  
6 different environmental settings (Wahl, 2004). This is particularly important for  
7 paleoenvironmental applications because many late-glacial and early Holocene fossil  
8 pollen assemblages lack modern analogs (Overpeck et al., 1992; Williams et al., 2001).  
9 Recent work with fossil pollen records from the upper Midwest (Shuman et al. in prep.)  
10 and the 64S-taxon list has shown that minimum SCD's consistently range between 0.15  
11 and 0.25 for middle- to late-Holocene pollen samples (by which time modern vegetation  
12 patterns had been established, Webb, 1987) and exceed 0.35 for early-Holocene and late-  
13 glacial ‘no-analog’ assemblages.

14 4.3. Intervariable Differences in Analytical Skill

15 Our findings do not show a clear superiority in analytical skill in the bioclimatic  
16 indices of temperature and moisture availability versus the monthly mean values of  
17 temperature and precipitation (Fig. 5). These climatic and bioclimatic variables of course  
18 are highly correlated, and pollen abundances along these various climate axes are  
19 arranged similarly enough that there is little difference in MAT skill. However,  
20 temperature-related and moisture-related variables clearly differ, suggesting that  
21 continental-scale plant abundances are more closely tied to temperature variations than to  
22 moisture availability (Guiot et al., 1993). Similar conclusions have been reached for  
23 subregional-scale applications (Wahl, 2003). The Priestly-Taylor coefficient, computed

1 as the ratio of AE to PE, has been advocated as an index of plant-available moisture  
2 (Prentice et al., 1992), but here we found that AE/PE is not well predicted by modern  
3 pollen distributions, and particularly not in eastern North America, where variations in  
4 AE/PE are small (Williams et al., 2006). AE alone has also been advocated as a  
5 biologically meaningful predictor of plant distributions (Stephenson, 1990; 1998), and  
6 previous work has shown a high correlation between AE and the distribution of  
7 vegetation types and tree diversity (Currie, 1991; Currie and Paquin, 1987). AE is  
8 reconstructed relatively skillfully by the MAT (Fig. 5), which also supports the use of AE  
9 as a bioclimatic predictor of plant distributions.

10 **4.4. Spatial Autocorrelation**

11 Spatial autocorrelation in environmental variables helps inflate the apparent MAT  
12 skill for individual variables, e.g. by increasing  $R^2$  values (Telford and Birks, 2005). As  
13 Telford and Birks (2005) discuss, the MAT is a holistic approach, in which suites of  
14 environmental variables are attributed to a target sample. These environmental variables  
15 are cross-correlated and autocorrelated, so that similarity in one environmental dimension  
16 is often a good predictor of similarity in other climatic dimensions. Consequently, the  
17 expected  $R^2$  value is not zero. However, because the spatial autocorrelation structure is  
18 the same for all cross-validation analyses,  $R^2$  and RMSE can still be used as a metric of  
19 relative analytical skill.

20 **5. Conclusions**

21 Our analyses show that the positive aspects of a comprehensive modern pollen  
22 calibration dataset for North America (a spatially dense and extensive network of sites  
23 from a variety of plant communities) can be outweighed by the negative aspects

1 (decreased precision of pollen-climate relationships), unless the appropriate analytical  
2 steps are taken. The key issue is the existence of interregional differences in pollen-  
3 climate relationships, which result from a) species-level niche variation, b) regional  
4 floristic differences, and c) the taxonomic imprecision inherent to pollen data. If these  
5 interregional differences are not corrected for, they can significantly degrade the accuracy  
6 of pollen-based climate reconstructions for North America. We show that both  
7 expanding the list of pollen types used for MAT analysis from 36 to 64 pollen types and  
8 geographically splitting key pollen types (using the range maps of the parent species to  
9 guide the splits) significantly improves the accuracy and precision of continental-scale  
10 climate reconstructions. Increasing from 64 to 135 pollen types resulted in small or no  
11 gains in MAT skill. In general, splitting major pollen types into regional groups  
12 improved MAT skill more than adding minor pollen types.

13 Based upon a comprehensive series of cross-validation analyses with the North  
14 American surface sample dataset, we recommend the following parameter set for  
15 subcontinental- to continental-scale paleoenvironmental reconstructions: a 64-taxon list  
16 with some taxa split regionally  $T=0.20$  to  $0.30$  (this assumes the squared-chord  
17 dissimilarity metric is used),  $N=3$  to  $7$ , and no distance weighting or an inverse-distance  
18 weighting.

19 The MAT reconstructs temperature-related variables more precisely than  
20 hydrological variables. There is little difference in MAT precision between most  
21 ‘climatic’ (i.e. monthly temperature and precipitation) and ‘bioclimatic’ variables (e.g.  
22 growing-degree days, the ratio of actual to potential evapotranspiration, mean

1 temperature of the coldest month). However, integrative bioclimatic variables, such as  
2 actual evapotranspiration, are skillfully reconstructed by the MAT.

3 **6. Acknowledgments**

4 This paper is indebted to all scientists who have helped assemble or contributed data to  
5 the North American Surface Sample Database, with particular thanks to Johanne  
6 Whitmore, Mike Sawada, Andre Viau, and Konrad Gajewski. Patrick Bartlein and Sara  
7 Shafer contributed the climate data used here. This paper was improved by comments  
8 from Patrick Bartlein, Gene Wahl, Tom Webb, and an anonymous reviewer. Support for  
9 this work was provided by the NSF Earth System History Program (ATM-0507999,  
10 ATM-9910638, DEB-0716471).

1      **7. References**

- 2      Allen, J.R.M., Watts, W.A., Huntley, B., 2000. Weichselian palynostratigraphy,  
3            palaeovegetation and palaeoenvironment; the record from Lago Grande di  
4            Monticchio, southern Italy. Quaternary International 73/74, 91-110.
- 5      Anderson, L.L., Hu, F.S., Nelson, D.M., Petit, R.J., Paige, K.N., 2006. Ice-age  
6            endurance: DNA evidence of a white spruce refugium in Alaska. Proceedings of  
7            the National Academy of Sciences 103, 12447-12450.
- 8      Anonymous, 2006. PLANTS Database. US Department of Agriculture,  
9            <http://plants.usda.gov>. Accessed 2006.
- 10     Axelrod, D.I., 1985. Rise of the central grassland biome, central North America.  
11            Botanical Review 51, 163-202.
- 12     Bartlein, P.J., Webb, T., III, Flerl, E., 1984. Holocene climate change in the northern  
13            Midwest: Pollen-derived estimates. Quaternary Research 22, 361-374.
- 14     Bartlein, P.J., Whitlock, C., 1993. Paleoclimatic interpretation of the Elk Lake pollen  
15            record. In: Bradbury, J.P., Dean, W.E. (Eds.), Elk Lake, Minnesota: Evidence for  
16            rapid climate change in the north-central United States. Geological Society of  
17            America, Boulder, CO pp. 275-293
- 18     Birks, H.H., 2003. The importance of plant macrofossils in the reconstruction of  
19            Lateglacial vegetation and climate: examples from Scotland, western Norway,  
20            and Minnesota, USA. Quaternary Science Reviews 22, 453-473.
- 21     Birks, H.H., Birks, H.J.B., 2000. Future uses of pollen analysis must include plant  
22            macrofossils. Journal of Biogeography 27, 31-35.
- 23     Birks, H.J.B., 1995. Quantitative paleoenvironmental reconstruction. In: Maddy, D.,  
24            Brew, J.S. (Eds.), Statistical modelling of Quaternary Science Data. Technical  
25            Guide 5. Quaternary Research Association, Cambridge pp. 116-254
- 26     Bryson, R.A., 1985. On climatic analogs in paleoclimatic reconstruction. Quaternary  
27            Research 23, 275-286.
- 28     Calcote, R., 2003. Mid-Holocene climate and the hemlock decline: the range limit of  
29            *Tsuga canadensis* in the western Great Lakes region, USA. Holocene 13, 215-  
30            224.
- 31     Chase, J.M., Leibold, M.A., 2003. Ecological Niches: Linking Classical and  
32            Contemporary Approaches. University of Chicago Press, Chicago 212pp.
- 33     Cramer, W., Prentice, I.C., 1988. Simulation of regional soil moisture deficits on a  
34            European scale. Norsk Geografisk Tidsskrift 42, 149-151.
- 35     Crosta, X., Pichon, J.-J., Burckle, L.H., 1998. Application of modern analog technique to  
36            marine Antarctic diatoms: Reconstruction of maximum sea-ice extent at the Last  
37            Glacial Maximum. Paleoceanography 13, 284-297.
- 38     Currie, D.J., 1991. Energy and large-scale patterns of animal- and plant-species richness.  
39            American Naturalist 137, 27-49.
- 40     Currie, D.J., Paquin, V., 1987. Large-scale biogeographical patterns of species richness in  
41            trees. Nature 329, 326-327.
- 42     Davis, O.K., 1995. Climate and vegetation patterns in surface samples from arid western  
43            U.S.A.: Application to Holocene climatic reconstructions. Palynology 19, 95-  
44            117.

- 1 DeFries, R.S., Hansen, M.C., Townshend, J.R.G., 2000. Global continuous fields of  
2 vegetation characteristics: a linear mixture model applied to multi-year 8 km  
3 AVHRR data. *International Journal of Remote Sensing* 21, 1389-1414.
- 4 Etterson, J.R., Shaw, R.G., 2001. Constraint to adaptive evolution in response to global  
5 warming. *Science* 294, 151-154.
- 6 Finkelstein, S.A., Gajewski, K., Viau, A.E., 2005. Improved resolution of pollen  
7 taxonomy allows better biogeographical interpretation of post-glacial forest  
8 development: analyses from the North American Pollen Database. *Journal of  
9 Ecology* 94, 415-430.
- 10 Gajewski, K., 1988. Late Holocene climate changes in eastern North-America estimated  
11 from pollen data. *Quaternary Research* 29, 255-262.
- 12 Gavin, D.G., Oswald, W.W., Wahl, E.R., Williams, J.W., 2003. A statistical approach to  
13 evaluating distance metrics and analog assignments for pollen records. *Quaternary  
14 Research* 60, 356-367.
- 15 Guiot, J., Cheddadi, R., Prentice, I.C., Jolly, D., 1996. A method of biome and land  
16 surface mapping from pollen data: application to Europe 6000 years ago.  
17 *Palaeoclimates* 1, 311-324.
- 18 Guiot, J., Harrison, S.P., Prentice, I.C., 1993. Reconstruction of Holocene precipitation  
19 patterns in Europe using pollen and lake-level data. *Quaternary Research* 40, 139-  
20 149.
- 21 Hare, F.K., 1980. Long-term annual surface heat and water balances over Canada and the  
22 United States south of 60°N: reconciliation of precipitation, run-off and  
23 temperature fields. *Atmosphere-Ocean* 18, 127-153.
- 24 Henderson, A.R., 1993. Assessing test accuracy and its clinical consequences: a primer  
25 for receiver operating characteristic curve analysis. *Annals of Clinical  
26 Biochemistry* 30, 521-539.
- 27 Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography.  
28 Princeton University Press, Princeton, NJ 448pp.
- 29 Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbor Symposia on  
30 Quantitative Biology* 22, 425-427.
- 31 Imbrie, J., Kipp, N.G., 1971. A new micropaleontological method for quantitative  
32 paleoclimatology: application to a late Pleistocene Caribbean core. In: Turekian,  
33 K. (Ed.), *The Late Cenozoic Glacial Ages*. Yale University Press, New Haven pp.  
34 71-181
- 35 Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to  
36 environmental changes of the late Quaternary. *Paleobiology* 26 (Supplement),  
37 194-220.
- 38 Jackson, S.T., Overpeck, J.T., Webb, T., III, Keatitch, S.E., Anderson, K.H., 1997.  
39 Mapped plant-macrofossil and pollen records of late Quaternary vegetation  
40 change in eastern North America. *Quaternary Science Reviews* 16, 1-70.
- 41 Jackson, S.T., Webb, R.S., Anderson, K.H., Overpeck, J.T., Webb, T., III, Williams,  
42 J.W., Hansen, B.C.S., 2000. Vegetation and environment in eastern North  
43 America during the last glacial maximum. *Quaternary Science Reviews* 19, 489-  
44 508.

- 1 Jackson, S.T., Weng, C., 1999. Late Quaternary extinction of a tree species in eastern  
2 North America. *Proceedings of the National Academy of Sciences* 96, 13847-  
3 13852.
- 4 Jackson, S.T., Williams, J.W., 2004. Modern analogs in Quaternary paleoecology: Here  
5 today, gone yesterday, gone tomorrow? *Annual Review of Earth and Planetary  
6 Sciences* 32, 495-537.
- 7 Legendre, L., Legendre, P., 1998. *Numerical Ecology*. Elsevier Scientific Pub.,  
8 Amsterdam 853pp.
- 9 Lindbladh, M., Jacobson, G.L., Jr., Schauffler, M., 2003. The postglacial history of three  
10 *Picea* species in New England, USA. *Quaternary Research* 59, 61-69.
- 11 Lindbladh, M., Oswald, W.W., Foster, D.R., Faison, E.K., Hou, J., Huang, Y., 2007. A  
12 late-glacial transition from *Picea glauca* to *Picea mariana* in southern New  
13 England. *Quaternary Research* 67, 502-508.
- 14 Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J., Yang, L., Merchant, J.W.,  
15 2000. Development of a global land cover characteristics database and IGBP  
16 DISCover from 1-km AVHRR data. *International Journal of Remote Sensing* 21,  
17 1303-1330.
- 18 Lytle, D.E., Wahl, E.R., 2005. Palaeoenvironmental reconstructions using the modern  
19 analogue technique: the effects of sample size and decision rules. *The Holocene*  
20 15, 554-566.
- 21 Malmgren, B.A., Kucera, M., Nyberg, J., Waelbroeck, C., 2001. Comparison of statistical  
22 and artificial neural networks techniques for estimating past sea surface  
23 temperatures from planktonic foraminifer census data. *Paleoceanography* 16, 520-  
24 530.
- 25 Malmgren, B.A., Nordlund, U., 1997. Application of artificial neural networks to  
26 paleoceanographic data. *Palaeogeography, Palaeoclimatology, Palaeoecology*  
27 136, 359-373.
- 28 Nakagawa, T., Tarasov, P.E., Nishida, K., Gotanda, K., Yasuda, Y., 2002. Quantitative  
29 pollen-based climate reconstruction in central Japan: application to surface and  
30 Late Quaternary spectra. *Quaternary Science Reviews* 21, 2099-2113.
- 31 New, M.G., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface  
32 climate over global land areas. *Climate Research* 21, 1-25.
- 33 Oswald, W.W., Brubaker, L.B., Hu, F.S., Gavin, D.G., 2003. Pollen-vegetation  
34 calibration for tundra communities in the Arctic Foothills, northern Alaska.  
35 *Journal of Ecology* 91, 1022-1033.
- 36 Overpeck, J.T., Webb, R.S., Webb, T., III, 1992. Mapping eastern North American  
37 vegetation change of the past 18 ka: No-analogs and the future. *Geology* 20,  
38 1071-1074.
- 39 Overpeck, J.T., Webb, T., III, Prentice, I.C., 1985. Quantitative interpretation of fossil  
40 pollen spectra: dissimilarity coefficients and the method of modern analogs.  
41 *Quaternary Research* 23, 87-108.
- 42 Pflaumann, U., Duprat, J., Pujol, C., Labeyrie, L.D., 1996. SIMMAX: A modern analog  
43 technique to deduce Atlantic sea surface temperatures from planktonic  
44 foraminifera in deep-sea sediments. *Paleoceanography* 11, 15-35.

- 1 Prell, W.L., 1985. The stability of low-latitude sea-surface temperatures: An evaluation  
2 of the CLIMAP reconstruction with emphasis on the positive SST anomalies. US  
3 Department of Energy, Washington, D. C.pp.
- 4 Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, R.A.,  
5 1992. A global biome model based on plant physiology and dominance, soil  
6 properties and climate. *Journal of Biogeography* 19, 117-134.
- 7 Sawada, M., Viau, A.E., Gajewski, K., 2001. Critical thresholds of dissimilarity in the  
8 modern analog technique (MAT) for quantitative paleoclimate reconstruction. In:  
9 Chylek, P., Lesins, G. (Eds.) 1st Annual Conference on Global Warming and the  
10 Next Ice Age. Dalhousie University, Halifax, Nova Scotia 149-152pp.
- 11 Sawada, M., Viau, A.E., Vettoretti, G., Peltier, W.R., Gajewski, K., 2004. Comparison of  
12 North-American pollen-based temperature and global lake-status with CCCma  
13 AGCM output at 6ka. *Quaternary Science Reviews* 23, 225-244.
- 14 Stephenson, N.L., 1990. Climatic control of vegetation distribution: The role of the water  
15 balance. *American Naturalist* 135, 649-670.
- 16 Stephenson, N.L., 1998. Actual evapotranspiration and deficit: biologically meaningful  
17 correlates of vegetation distribution across spatial scales. *Journal of Biogeography*  
18 25, 855-870.
- 19 Sugita, S., 1994. Pollen representation of vegetation in Quaternary sediments: theory and  
20 method in patchy vegetation. *Journal of Ecology* 82, 881-897.
- 21 Telford, R.J., Birks, H.J.B., 2005. The secret assumption of transfer functions: problems  
22 with spatial autocorrelation in evaluating model performance. *Quaternary Science  
23 Reviews* 24, 2173-2179.
- 24 ter Braak, C.J.F., 1995. Non-linear methods for multivariate statistical calibration and  
25 their use in paleoecology: a comparison of inverse (k-nearest neighbors, partial  
26 least squares and weighted averaging partial least squares) and classical  
27 approaches. *Chemometrics and Intelligent Laboratory Systems* 28, 165-180.
- 28 ter Braak, C.J.F., Juggins, S., 1993. Weighted averaging partial least squares regression  
29 (WA-PLS): an improved method for reconstructing environmental.  
30 *Hydrobiologia* 269-270, 485-502.
- 31 Thompson, R.S., Anderson, K.H., Bartlein, P.J., 1999a. *Atlas of Relations Between  
32 Climatic Parameters and Distributions of Important Trees and Shrubs in North  
33 America--Hardwoods*. U.S. Department of the Interior, U.S. Geological Survey,  
34 Denver 423pp.
- 35 Thompson, R.S., Anderson, K.H., Bartlein, P.J., 1999b. *Atlas of Relations Between  
36 Climatic Parameters and Distributions of Important Trees and Shrubs in North  
37 America--Introduction and Conifers*. U.S. Department of the Interior, U.S.  
38 Geological Survey, Denver 269pp.
- 39 Thompson, R.S., Anderson, K.H., Bartlein, P.J., Smith, S.A., 1999c. *Atlas of Relations  
40 Between Climatic Parameters and Distributions of Important Trees and Shrubs in  
41 North America - Additional Conifers, Hardwoods, and Monocots*. U.S.  
42 Geological Survey 386pp.
- 43 Viau, A.E., Gajewski, K., 2007. Comments on: "The magnitudes of millennial- and  
44 orbital-scale climatic change in eastern North America during the Late  
45 Quaternary." *Quaternary Science Reviews* 26, 264-267.

- 1 Viau, A.E., Gajewski, K., Sawada, M.C., Fines, P., 2006. Millennial-scale temperature  
2 variations in North America during the Holocene. *Journal of Geophysical*  
3 *Research* 111.
- 4 Wahl, E., 2004. A general framework for determining cutoff values to select pollen  
5 analogs with dissimilarity metrics in the modern analog technique. *Review of*  
6 *Palaeobotany and Palynology* 128, 263-280.
- 7 Wahl, E.R., 2003. Assigning climate values to modern pollen surface sample sites and  
8 validating modern analog climate reconstructions in the southern California  
9 region. *Madrono* 50, 271-285.
- 10 Webb, R.S., Anderson, K.H., Webb, T., III, 1993a. Pollen response-surface estimates of  
11 late-Quaternary changes in the moisture balance of the northeastern United States.  
12 *Quaternary Research* 40, 213-227.
- 13 Webb, T., III, 1987. The appearance and disappearance of major vegetational  
14 assemblages: Long-term vegetational dynamics in eastern North America.  
15 *Vegetatio* 69, 177-187.
- 16 Webb, T., III, Anderson, K.H., Bartlein, P.J., Webb, R.S., 1998. Late Quaternary climate  
17 change in eastern North America: a comparison of pollen-derived estimates with  
18 climate model results. *Quaternary Science Reviews* 17, 587-606.
- 19 Webb, T., III, Bartlein, P.J., Harrison, S.P., Anderson, K.H., 1993b. Vegetation, lake  
20 levels, and climate in eastern North America for the past 18,000 years. In: Wright,  
21 H.E., Jr., Kutzbach, J.E., Webb, T., III, Ruddiman, W.F., Street-Perrott, F.A.,  
22 Bartlein, P.J. (Eds.), *Global Climates Since the Last Glacial Maximum*.  
23 University of Minnesota Press, Minneapolis, MN pp. 415-467
- 24 Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B., Bartlein, P.J.,  
25 Minckley, T., Viau, A.E., Webb, T., III, Anderson, P.M., Brubaker, L.B., 2005.  
26 North American and Greenland modern pollen data for multi-scale  
27 paleoecological and paleoclimatic applications. *Quaternary Science Reviews* 24,  
28 1828-1848.
- 29 Williams, J.W., 2003. Variations in tree cover in North America since the Last Glacial  
30 Maximum. *Global and Planetary Change* 35, 1-23.
- 31 Williams, J.W., Bartlein, P.J., Webb III, T., 2000. Data-model comparisons for eastern  
32 North America - Inferred biomes and climate values from pollen data. In:  
33 Braconnot, P. (Ed.) *Proceedings of the Third Palaeoclimatic Modelling*  
34 *Intercomparison Project Workshop*. World Climate Research Program, Montreal,  
35 Canada 77-86pp.
- 36 Williams, J.W., Shuman, B., Bartlein, P.J., Whitmore, J., Gajewski, K., Sawada, M.,  
37 Minckley, T., Shafer, S., Viau, A.E., Webb, T., III, Anderson, P.M., Brubaker,  
38 L.B., Whitlock, C., Davis, O.K., 2006. An Atlas of Pollen-Vegetation-Climate  
39 Relationships for the United States and Canada. American Association of  
40 Stratigraphic Palynologists Foundation, Dallas, TX 293pp.
- 41 Williams, J.W., Shuman, B.N., Webb, T., III, 2001. Dissimilarity analyses of late-  
42 Quaternary vegetation and climate in eastern North America. *Ecology* 82, 3346-  
43 3362.
- 44 Williams, J.W., Zajac, L., Kaplan, J.O., 2008. Leaf area index for northern and eastern  
45 North America at the Last Glacial Maximum: A data-model comparison. *Global*  
46 *Ecology and Biogeography* 17, 122-134.

- 1 Wunderlin, R.P., Hansen, B.F., 2004. Atlas of Florida Vascular Plants. University of  
2 South Florida, Tampa [www.plantatlas.usf.edu](http://www.plantatlas.usf.edu). Accessed 2006.  
3 Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a  
4 fundamental evaluation tool in clinical medicine. Clinical Chemistry 39, 561-577.

1    **Figure Legends and Tables**

2    **Figure 1:** Pollen abundances versus the Priestley-Taylor index of soil moisture  
3    availability (Cramer and Prentice, 1988; Hare, 1980), calculated as the ratio between  
4    actual evapotranspiration and potential evapotranspiration (AE/PE). AE/PE varies  
5    between 0 (dry) and 1 (wet). Nine taxa are shown: *Abies*, *Alnus* undiff.,  
6    Cupressaceae/Taxaceae, *Fraxinus*, *Picea*, *Quercus*, *Tsuga*, *Pinus*, and  
7    Chenopodiaceae/Amaranthaceae. Pollen abundances for all taxa are split into regional  
8    subsets, which are based on the aggregated range maps for regionally co-occurring  
9    species. Designation of species as eastern or western is based upon Thompson et al.  
10   (1999b; 1999a), as is the split of *Pinus* into western pines, northeastern pines, and  
11   southeastern pines. Note that '*Alnus* undiff.' does not include *A. rugosa* and *A. rubra*, but  
12   does include *A. crispa* (Table 1, ALNUSXC). The split of  
13   Chenopodiaceae/Amaranthaceae between Florida and the Midwest is based upon a  
14   bimodal distribution of pollen abundances between these regions (Williams et al., 2006)  
15   and state-level species distributions from (Anonymous, 2006) and (Wunderlin and  
16   Hansen, 2004). Filled circles indicate abundances >0.5%; open circles indicate  
17   abundances <0.5%.

18  
19   **Figure 2:** As Figure 1, but for pollen abundances versus growing degree-days on a 5°C  
20   base (GDD5).

21  
22   **Figure 3:** Plots of the root mean squared error of prediction (RMSE) for linear  
23   regressions of observed versus analog-reconstructed estimates of modern environmental  
24   variables, for the entire North American dataset (i.e. NA-NA experiment). RMSE's are  
25   compared across 12 experiments using four different taxon lists (36 taxa, 64 taxa, 135  
26   taxa, and the 64-taxon list with geographic splits) and three different distance weightings  
27   (no weighting [open circles], inverse-distance [1/D, black circles], and inverse-distance-  
28   squared [1/D<sup>2</sup>, black triangles]. The five plots are for a) mean JJA temperatures (TSum),  
29   b) mean DJF temperatures (TWin), c) mean JJA precipitation (PSum), d) mean DJF  
30   precipitation (PWin), and e) fractional woody cover (Woody). For this visualization, the  
31   number of analogs and threshold were not held constant; instead, we used whichever  
32   values a) allowed at least 95% of modern samples to find an analog (see Fig. 6) and b)  
33   given a), also produced the maximal RMSE for each combination of distance-weighting,  
34   taxon list, and climatic variable. The RMSE axis is reversed so that better (lower)  
35   RMSE's are plotted above worse (higher) RMSE's.

36  
37   **Figure 4:** A comparison across climatic, bioclimatic, and vegetational variables of the  
38   root mean squared error of prediction (RMSE) for six cross-validation experiments:  
39   NA-NA135 (filled circles), NA-NA64S (open circles), ENA-NA135 (black triangles),  
40   ENA-NA64S (open triangles), ENA-ENA135 (black squares), and ENA-ENA64S (open  
41   squares). Variable abbreviations: [TJAN...TDEC]= mean monthly temperatures (°C) for  
42   1961-1990, [PJAN...PDEC]= monthly precipitation total (mm), TAVE= mean annual  
43   temperature (°C), TMAX= maximum single monthly temperature between 1961 and  
44   1990 (°C), MTWA= mean temperature of the warmest month of the year (°C), TMIN=  
45   minimum single monthly temperature between 1961 and 1990 (°C), MTCO= mean

1 temperature of the coldest month of the year ( $^{\circ}\text{C}$ ),  $\text{GDD}_0$ = growing-degree days,  $0^{\circ}\text{C}$   
2 base (degree-days),  $\text{GDD}_5$ = growing-degree days,  $5^{\circ}\text{C}$  base (degree-days),  $\text{AE}$ = actual  
3 evapotranspiration (mm),  $\text{PE}$ = potential evapotranspiration (mm),  $\text{AE/PE}$ = ratio of actual  
4 to potential evapotranspiration,  $\text{AE/PE}_{\text{EV}}$ = ratio of actual to potential evapotranspiration  
5 for the evergreen assimilation period,  $\text{AE/PE}_{\text{DC}}$ = ratio of actual to potential  
6 evapotranspiration for the deciduous assimilation period,  $\text{PJANPANN}$ = January/Annual  
7 precipitation,  $\text{PJULPANN}$ = July/Annual precipitation,  $\text{NEED}$ = fractional area around the  
8 site covered by woody needleleaved plants,  $\text{BROAD}$ = fractional broadleaved woody  
9 cover,  $\text{WOODY}$ =fractional woody cover,  $\text{OPEN}$ =fractional open/unforested cover. For  
10 all experiments,  $T=0.25$ ,  $N=10$ , and environmental averages were unweighted. RMSE  
11 axes are inverted, with zero (perfect precision) at top. RMSE units and maximum y-axis  
12 value are indicated at bottom left of each plot. The RMSE metric of precision is best  
13 suited for interexperiment comparisons. See Fig. 5 for intervariable comparisons.  
14

15 **Figure 5.** As Figure 4, for the percent variance explained ( $R^2$ ) in the analog-based  
16 estimates by the original observations. Unlike Fig. 4, all variables share a common  $R^2$   
17 scale. The  $R^2$  metric of precision is best suited for intervariable comparisons. See Fig. 4  
18 for interexperiment comparisons.  
19

20 **Figure 6.** Scatter plot showing that the likelihood of finding an analog for any sample is  
21 directly determined by the choice of threshold, and is a saturating function. More  
22 restrictive (i.e. smaller) dissimilarity thresholds result in fewer samples having at least one  
23 match in the rest of the surface sample dataset. These ‘no-analog’ samples are not  
24 assigned any inferred environmental characteristics and are not included in the  $R^2$   
25 statistics.  
26

27 **Figure 7.** The distribution of RMSE values (shown as contours) with respect to two  
28 MAT parameters: the analog/no-analog threshold and the number of analog matches  
29 allowed. For these visualizations, we used the 64-split taxon list (Table 1) and no  
30 distance weighting.  
31

32 **Figure 8.** As Figure 7, showing instead the distribution of regression slopes (as an index  
33 of MAT accuracy).

1 **Table 1:** Taxa used in analog analyses

Scientific Name	Common Name	35	64	64S	135
<i>Abies</i>	Fir	x	x	we	x
<i>Acer negundo</i>	Box Elder	a	a	a	x
<i>Acer pennsylvanicum</i>	Striped Maple	a	a	a	x
<i>Acer rubrum</i>	Red Maple	a	a	a	x
<i>Acer saccharinum</i>	Silver Maple	a	a	a	x
<i>Acer saccharum</i>	Sugar Maple	a	a	a	x
<i>Acer undifferentiated</i>	Maple undif.	a	a	a	x
<i>ACER</i> <sup>1</sup>	Maple	x	x	we	
<i>Alnus crispa</i>	Mountain Alder	a	a	a	x
<i>Alnus rubra</i>	Red Alder	a	x	x	x
<i>Alnus rugosa</i>	Speckled Alder	a	x	x	x
<i>Alnus undifferentiated</i>	Alder	a	a	a	x
<i>ALNUSXC</i> <sup>2</sup>	Alder undif. + Mountain Alder		x	we	
<i>ALNUS</i> <sup>3</sup>	Alnus	x			
<i>Ambrosia</i>	Ragweed				x
<i>Amorpha</i>	Leadplant				x
Anacardiaceae	Sumac Family				x
Apiaceae	Celery Family		a	a	x
Aquifoliaceae	Holly Family		x	x	x
<i>Arceuthobium</i>	Mistletoe		x	x	x
Arecaceae	Palm Family				x
<i>Armeria</i>	Thrift				x
<i>Artemisia</i>	Sagebrush, Wormwood	x	x	x	x
Asteraceae undifferentiated	Daisy tribe	a	a	a	x
Asteraceae subf. Cichorioideae	Dandelion tribe	a	a	a	x
ASTERACEAE <sup>4</sup>	Asteraceae Family	x	x	x	
<i>Betula</i>	Birch	x	x	x	x
Boraginaceae	Borage Family				x
<i>Botrychium</i>	Grape Fern (spore)				x
Brassicaceae	Mustard Family		x	x	x
Cactaceae	Cactus Family	x	x	x	x
Campanulaceae	Bluebells				x
Caprifoliaceae	Honeysuckle Family				x
<i>Carya</i>	Hickory	x	x	x	x
Caryophyllaceae	Chickweed Family		x	x	x
<i>Castanea</i>	Chestnut	x	x	x	x
<i>Ceanothus</i>	New Jersey Tea				x
<i>Celtis</i>	Hackberry		x	we	x
<i>Cephalanthus</i>	Button Bush		x	x	x
<i>Cercocarpus</i>	Mahogany				x
<i>Chamaedaphne/Ledum</i>	Leather Leaf/Labrador Tea		a	a	x
Chenopodiaceae/Amaranthaceae	Goosefoot/Pigweed Families	x	x	fc	x
<i>Chrysolepis/Lithocarpus</i>	Chinquapin		x	x	x
<i>Cornus</i>	Dogwood				x
<i>Corylus</i>	Hazel	x	x	x	x

Cupressaceae	Cedar Family	x	x	we	x
Cyperaceae	Sedge	x	x	x	x
<i>Dodecatheon</i>	Shooting Star				x
<i>Dryas</i>	Mountain Avens		x	x	x
Elaeagnaceae	Oleaster Family				x
<i>Ephedra</i>	Mexican Tea	x	x	x	x
<i>Equisetum</i>	Horsetails (spore)				x
Ericaceae undifferentiated	Heath Family		a	a	x
ERICACEAE <sup>5</sup>	Heath Family		x	x	
Ericales undifferentiated	Ericales Order				x
<i>Eriogonum</i>	Buckwheat		a	a	x
Euphorbiaceae	Spurge Family		a	a	x
Fabaceae	Pea Family				x
<i>Fagus</i>	Beech	x	x	x	x
<i>Fraxinus nigra</i>	Black Ash	a	x	x	x
<i>Fraxinus pennsylvanica/americana</i>	White/Green Ash	a	x	x	x
<i>Fraxinus</i> undifferentiated	Ash	a	x	x	x
FRAXINUS <sup>6</sup>	Ash	x		we	
<i>Iva</i>	Marsh Elder	a	x	x	x
<i>Juglans cinerea</i>	Butternut	a	a	a	x
<i>Juglans nigra</i>	Black Walnut	a	a	a	x
<i>Juglans</i> undifferentiated	Walnut Family	a	a	a	x
JUGLANS <sup>7</sup>	Walnut	x	x	x	
<i>Koenigia islandica</i>	Island Purslane				x
Lamiaceae	Mint Family				x
<i>Larix/Pseudotsuga</i>	Larch/Douglas Fir	x	x	we	x
<i>Larrea</i>	Creosote Family		x	x	x
Liliaceae	Lily Family				x
<i>Liquidambar</i>	Sweet Gum	x	x	x	x
<i>Liriodendron</i>	Tulip Tree				x
<i>Lycopodium annotinum</i>	Club moss		a	a	x
<i>Lycopodium clavatum</i>	Club moss		a	a	x
<i>Lycopodium complanatum</i>	Club moss		a	a	x
<i>Lycopodium selago</i>	Club moss		a	a	x
<i>Lycopodium</i> undifferentiated	Club moss		a	a	x
LYCOPODIUM <sup>8</sup>	Club moss	x	x		
Magnoliaceae	Umbrella Tree Family				x
Malvaceae	Mallow Family	a	a		x
MINOR FORBS <sup>9</sup>	Minor Forbs	x	x		
Moraceae	Mulberry Family				x
Myricaceae	Myrtle Family	x	x		x
<i>Nyssa</i>	Sour Gum	x	x		x
Onagraceae	Evening Primrose Family				x
Osmundaceae	Royal Fern Family				x
<i>Ostrya/Carpinus</i>	Hop/Hornbeam, Ironwood	x	x	x	x
<i>Oxyria</i>	Mountain Sorrel		x	x	x
Papaveraceae	Poppy Family	x	x	x	x

<i>Pedicularis</i>	Louse Wort				x
<i>Picea glauca</i>	White Spruce	a	a	a	x
<i>Picea mariana</i>	Black Spruce	a	a	a	x
<i>Picea undifferentiated</i>	Spruce	a	a	a	x
<i>PICEA</i> <sup>10</sup>	Spruce	x	x	we	
<i>Pinus diploxylon</i>	Hard Pines	a	x	wnese	x
<i>Pinus haploxylon</i>	Soft Pines	a	x	wnese	x
<i>Pinus undifferentiated</i>	Pine undifferentiated	a	x	wnese	x
<i>Pinus Southeast</i>	Southeastern Pines	x	x		x
<i>PINUS</i> <sup>11</sup>	Pine	x			
Plantaginaceae	Plantain Family		x	x	x
<i>Platanus</i>	Sycamore	x	x	x	x
Poaceae	Grass	x	x	x	x
Polemoniaceae	Phlox Family				x
Polygonaceae	Buckwheat undifferentiated				x
<i>Polygonum bistortoides</i>	Knotweed				x
<i>Polygonum undifferentiated</i>	Bistort				x
<i>Polygonum viviparum</i>	Fern Family				x
Polypodiaceae	Poplar, Aspen	x	x	x	x
<i>Populus</i>	Cinquefoil				x
<i>Potentilla</i>	Mesquite	x	x	x	x
<i>Prosopis</i>	Bracken (fern)		x	x	x
<i>Pteridium</i>	Oak	x	x	we	x
<i>Quercus</i>	Buttercup Family				x
Ranunculaceae	Buckthorn/Grape				x
Rhamnaceae/Vitaceae	Rose Family	x	x		x
Rosaceae	Bedstraw Family				x
Rubiaceae	Raspberry, Cloudberry				x
<i>Rubus</i>	Sorrel				x
<i>Rumex</i>	Sorrel/Mountainsorrel				x
<i>Rumex/Oxyria</i>	Willow	x	x	x	x
<i>Salix</i>	American Burnet				x
<i>Sanguisorba</i>	Greasewood	x	x	x	x
<i>Sarcobatus</i>	Nodding Saxifrage	a	a	a	x
<i>Saxifraga cernua</i>		a	a	a	x
<i>Saxifraga hieracifolia</i>	Purple Saxifrage	a	a	a	x
<i>Saxifraga oppositifolia</i>	Prickled Saxifrage	a	a	a	x
<i>Saxifraga tricuspidata</i>	Saxifrage	x	x	x	
<i>SAXIFRAGA</i> <sup>12</sup>	Saxifrage Family				x
Saxifragaceae undifferentiated	Figwort Family				x
Scrophulariaceae	Club moss				x
<i>Selaginella</i>	Soapberry				x
<i>Shepherdia canadensis</i>	Peat moss	x	x	x	x
<i>Sphagnum</i>	Cypress	x	x	x	x
<i>Taxodium</i>	Yew				x
<i>Taxus</i>	Meadow Rue	x			x

<i>Tilia</i>	Basswood	x	x	x	x
<i>Tsuga heterophylla</i>	Western Hemlock	x	x	x	x
<i>Tsuga mertensiana</i>	Mountain Hemlock		x	x	x
<i>Tsuga</i>	Hemlock unifferentiated	x	x	we	x
<i>Ulmus</i>	Elm	x	x	x	x
Urticaceae	Nettle Family				x
<i>Vaccinium</i>	Blueberry		a	a	x
<i>Xanthium</i>	Cocklebur types	a	a	a	x

x -- included in analog analyses as standalone pollen type

a -- included as part of an aggregated pollen type (aggregated types shown in capital letters)

we -- included but split into western and eastern groups

fc -- included but split into central and Floridian groups

wnese -- included but split into western, northeastern, and southeastern groups

<sup>1</sup> Sum of *Acer negundo*, *A. pennsylvanicum*, *A. rubrum*, *A. saccharinum*, *A. saccharum*, *A. undiff.*

<sup>2</sup> Sum of *Alnus crispa*, *A. undiff.*

<sup>3</sup> Sum of *Alnus crispa*, *A. rubra*, *A. rugosa*, *A. undiff.*

<sup>4</sup> Sum of Asteraceae undiff., Asteraceae subf. Cichorioideae, *Iva*, *Xanthium*. Does not include *Artemisia*, *Ambrosia*

<sup>5</sup> Sum of *Chamaedaphne/Ledum*, Ericaceae undif., *Vaccinum*

<sup>6</sup> Sum of *Fraxinus nigra*, *F. pennsylvanica/americana*, *F. undiff.*

<sup>7</sup> Sum of *Juglans cinerea*, *J. nigra*, *J. undiff.*

<sup>8</sup> Sum of *Lycopodium annotinum*, *L. clavatum*, *L. complanatum*, *L. selago*, *L. undiff.*

<sup>9</sup> Sum of Apiaceae, *Eriogonum*, Euphorbiaceae, Malvaceae

<sup>10</sup> Sum of *Picea glauca*, *P. mariana*, *P. undiff.*

<sup>11</sup> Sum of *Pinus diploxyylon*, *P. haploxyylon*, *P. undiff.* Does not include *Pinus* South

<sup>12</sup> Sum of *Saxifraga cernua*, *S. hieracifolia*, *S. oppositifolia*, *S. tricuspidata*

1	<b>Table 2:</b> Experimental values for each parameter
2	Threshold 0.01, 0.03, 0.05, 0.07, 0.09, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50*, 0.60*, 0.70*
3	
4	Number of Analogs 1, 3, 5, 7, 9, 10, 15, 20, 30*, 40*, 50*
5	Weighting None, Inverse-Distance ( $1/D$ ), Inverse-Distance-Squared ( $1/D^2$ )
6	Number of Taxa <sup>†</sup> 36, 64, 135, 64S
7	*Not used in all permutations
8	<sup>†</sup> See Table 1
9	
10	
11	

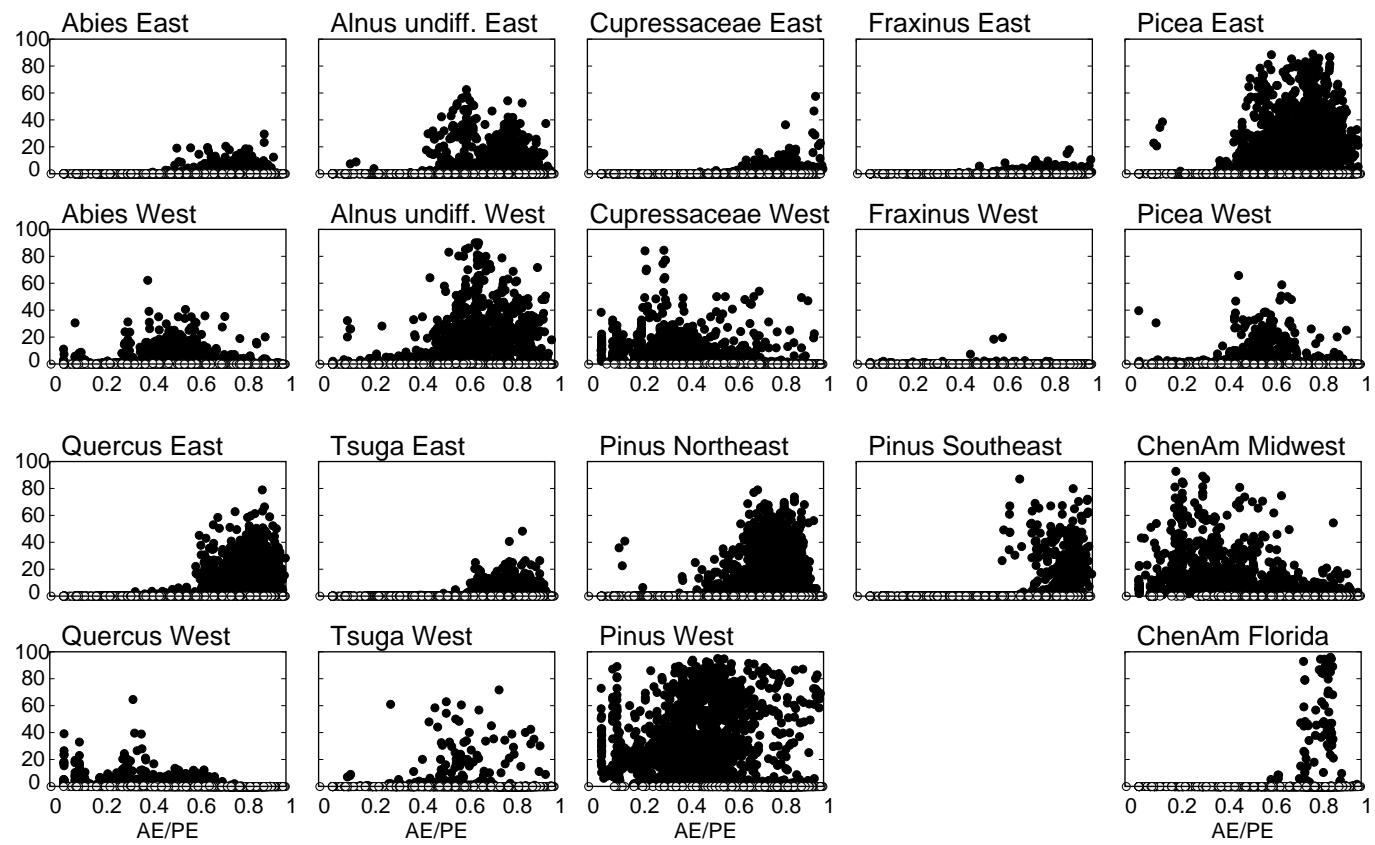
12 **Table 3:** Cross-validation statistics for ENA-NA and ENA-ENA experiments, using 36  
13 taxa, T=0.25, N=10, and an unweighted average. Asterixes in the last column indicate  
14 slopes significantly different (at  $p<0.05$ ) between the ENA-ENA and ENA-NA  
15 experiments.

	RMSE		$R^2$		Slope (with 95% confidence intervals)	
	ENA- ENA	ENA- NA	ENA- ENA	ENA- NA	ENA-ENA	ENA-NA
TSum (°C)	1.24	1.53	0.955	0.927	0.977 (0.968-0.986)	0.952 (0.941-0.964)*
MTWA (°C)	1.19	1.47	0.954	0.926	0.976 (0.967-0.985)	0.951 (0.939-0.962)*
GDD5 (°C-day)	283	418	0.963	0.908	0.987 (0.980-0.997)	0.928 (0.916-0.941)*
TWin (°C)	2.72	3.82	0.928	0.856	0.971 (0.960-0.983)	0.961 (0.944-0.977)*
MTCO (°C)	2.82	3.93	0.927	0.855	0.972 (0.960-0.983)	0.956 (0.940-0.973)*
PSum (mm)	13.1	18.2	0.782	0.573	0.877 (0.858-0.897)	0.827 (0.798-0.858)*
PWin (mm)	15.7	20.5	0.725	0.538	0.817 (0.796-0.838)	0.799 (0.768-0.831)
AnnP (mm)	135	195	0.781	0.564	0.886 (0.866-0.906)	0.896 (0.863-0.929)
AE/PE	0.0585	0.0799	0.502	0.219	0.670 (0.642-0.698)	0.658 (0.607-0.712)

16  
17  
18 **Table 4:** Optimal threshold values based on ROC analysis, in which the number of false-  
19 positive and false-negative analog identifications are jointly minimized.  
20

21	Taxon List	$T_{opt}$
22	36	0.118
23	64	0.188
24	135	0.239
25	64S	0.194

**Figures 1-8**



**Fig. 1**

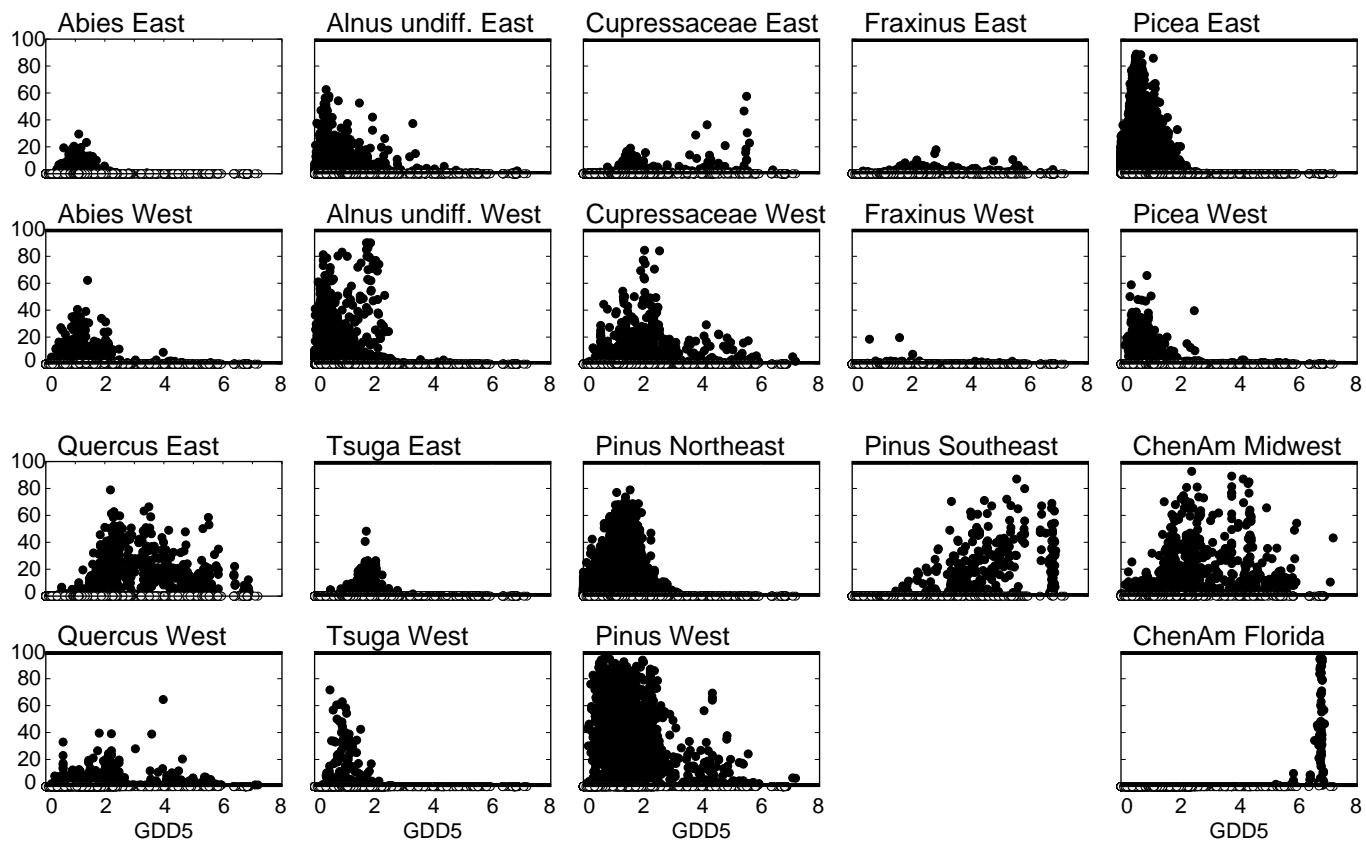


Fig. 2

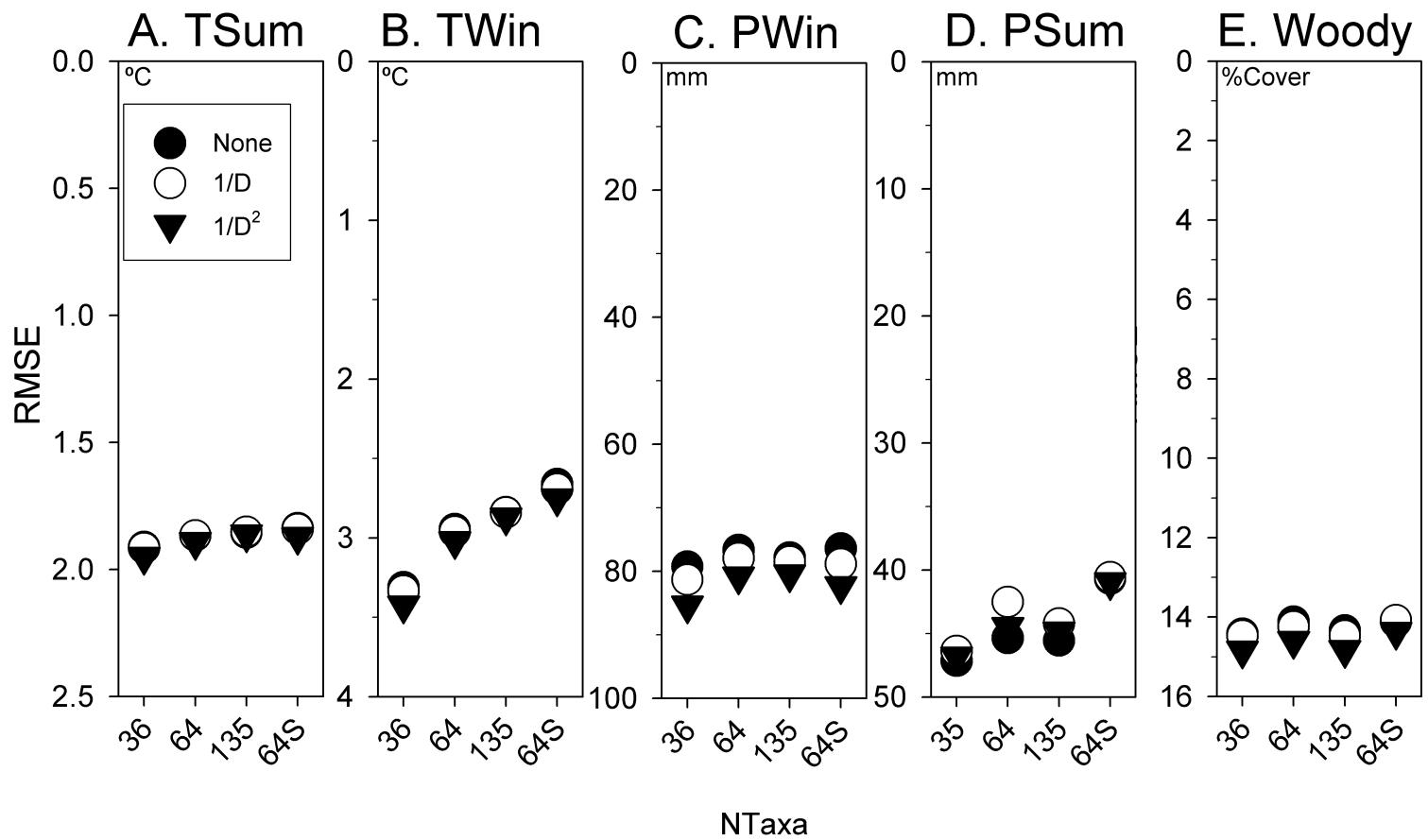


Fig. 3

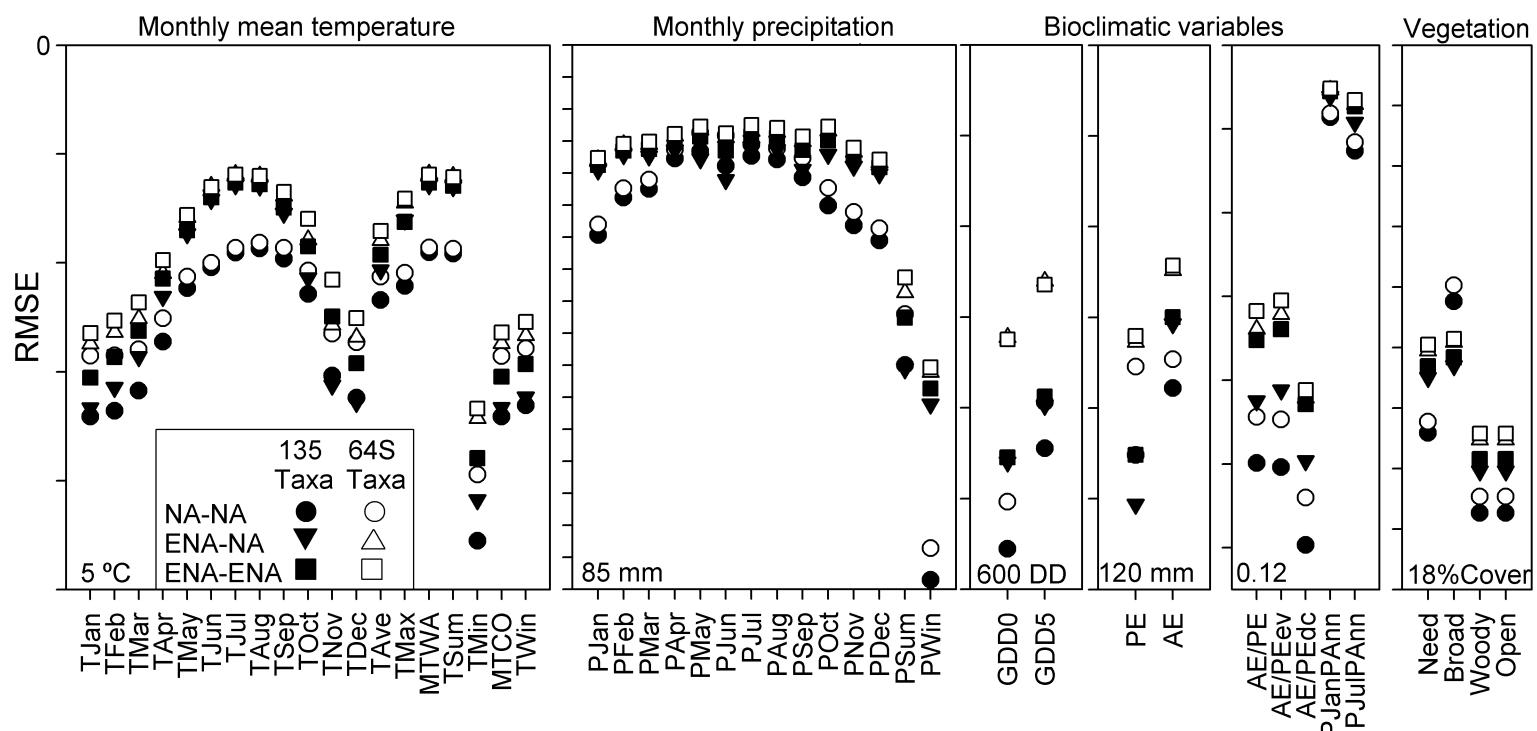


Fig. 4

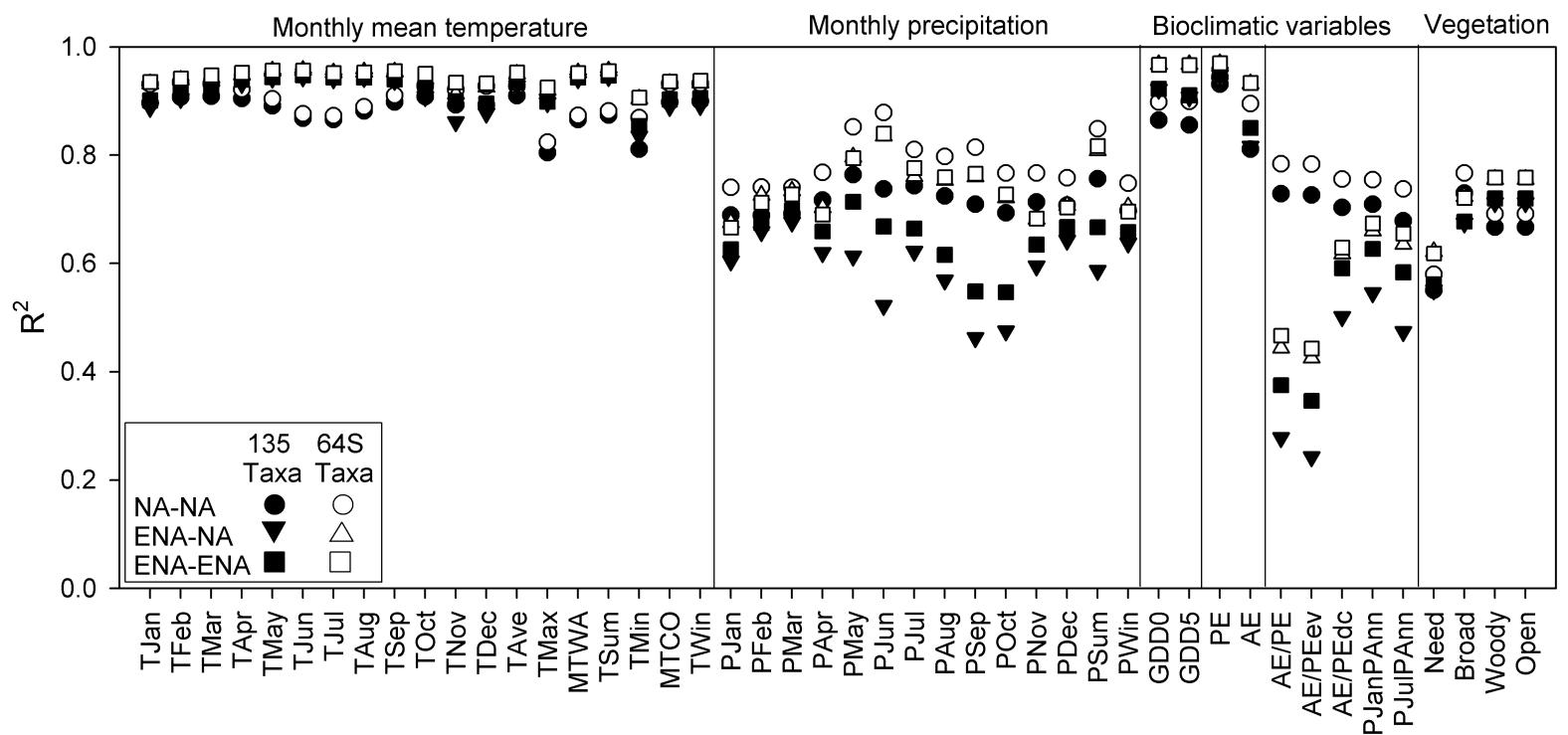


Fig. 5

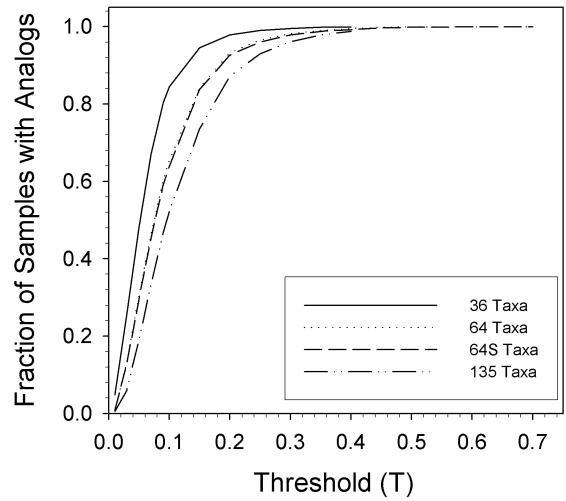


Fig. 6

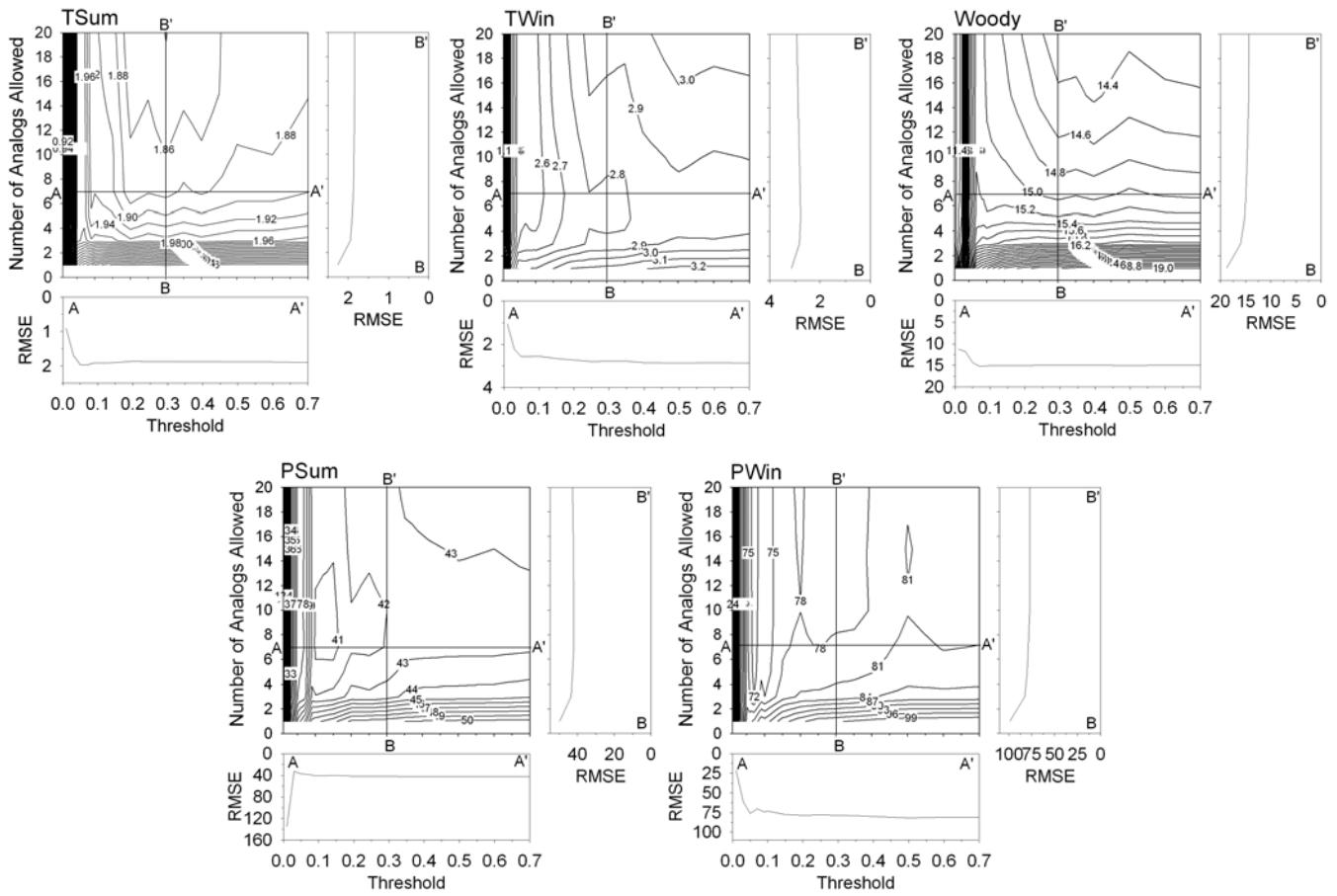


Fig. 7

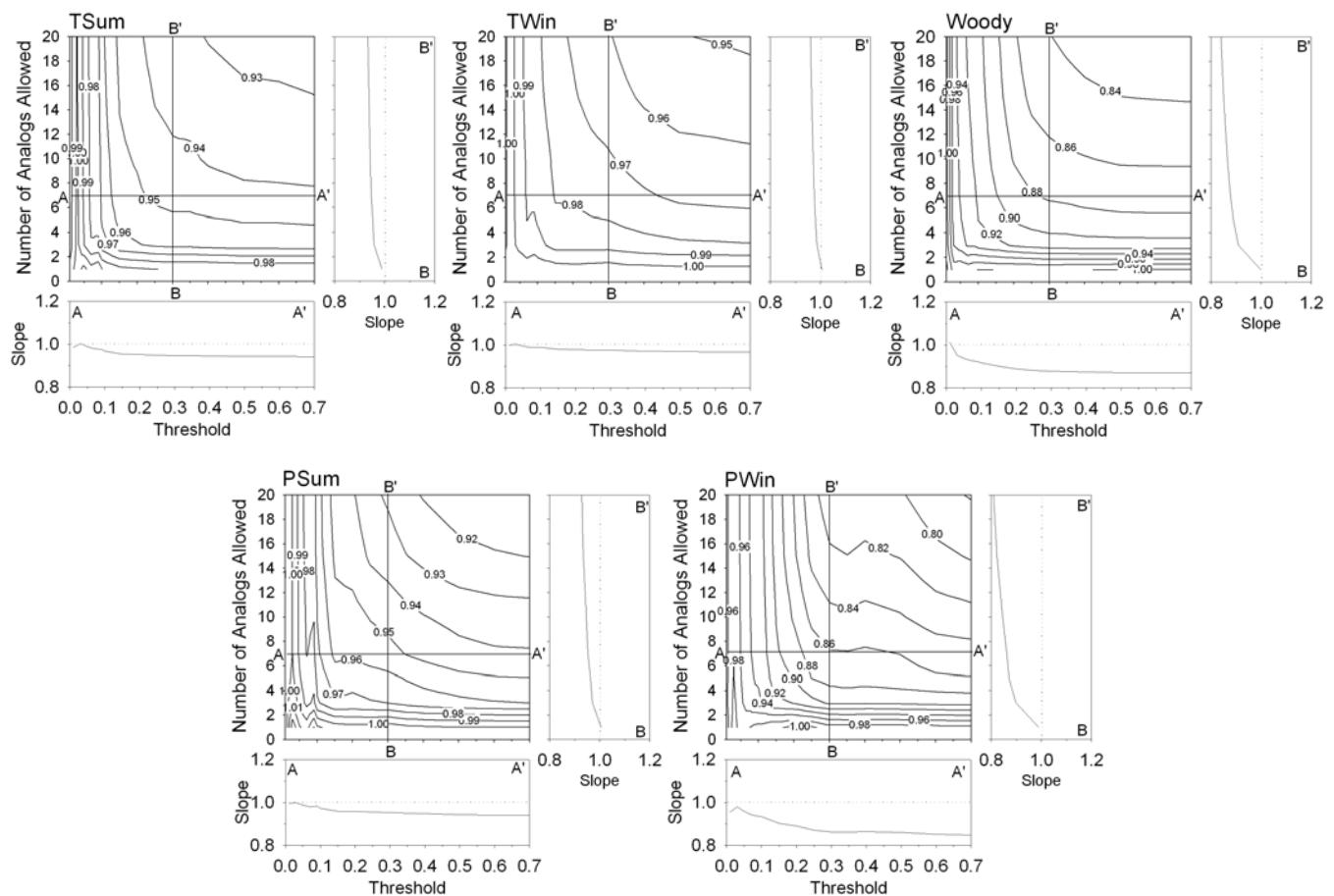


Fig. 8

**Supplementary text**

[\*\*Click here to download Supplementary Data: SupplementaryInfo\\_README.doc\*\*](#)

[Click here to download Supplementary Data: Abies\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Acer\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Alnus\\_EastWest.pdf](#)

[Click here to download Supplementary Data: \*Celtis\*\\_EastWest.pdf](#)

[Click here to download Supplementary Data: ChenAm\\_Florida.pdf](#)

[Click here to download Supplementary Data: Cypress\\_EastWest.pdf](#)

[Click here to download Supplementary Data: \*Fraxinus\\_EastWest.pdf\*](#)

[Click here to download Supplementary Data: LarixPseud\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Picea\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Pinus\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Quercus\\_EastWest.pdf](#)

[Click here to download Supplementary Data: Tsuga\\_EastWest.pdf](#)

[Click here to download Supplementary Data: abieEAST.dbf](#)

[Click here to download Supplementary Data: abieEAST.sbn](#)

[Click here to download Supplementary Data: abieEAST.sbx](#)

[Click here to download Supplementary Data: abieEAST.shp](#)

[Click here to download Supplementary Data: abieEAST.shp.xml](#)

[Click here to download Supplementary Data: abieEAST.shx](#)

[Click here to download Supplementary Data: abieWEST.dbf](#)

[Click here to download Supplementary Data: abieWEST.sbn](#)

[Click here to download Supplementary Data: abieWEST.sbx](#)

[Click here to download Supplementary Data: abieWEST.shp](#)

[Click here to download Supplementary Data: abieWEST.shp.xml](#)

[Click here to download Supplementary Data: abieWEST.shx](#)

[Click here to download Supplementary Data: acerEAST.dbf](#)

[Click here to download Supplementary Data: acerEAST.sbn](#)

[Click here to download Supplementary Data: acerEAST.sbx](#)

[Click here to download Supplementary Data: acerEAST.shp](#)

[Click here to download Supplementary Data: acerEAST.shp.xml](#)

[Click here to download Supplementary Data: acerEAST.shx](#)

[Click here to download Supplementary Data: acerWEST.dbf](#)

[Click here to download Supplementary Data: acerWEST.sbn](#)

[Click here to download Supplementary Data: acerWEST.sbx](#)

[Click here to download Supplementary Data: acerWEST.shp](#)

[Click here to download Supplementary Data: acerWEST.shp.xml](#)

[Click here to download Supplementary Data: acerWEST.shx](#)

[Click here to download Supplementary Data: AlnuEAST.dbf](#)

[Click here to download Supplementary Data: AlnuEAST.sbn](#)

[Click here to download Supplementary Data: AlnuEAST.sbx](#)

[Click here to download Supplementary Data: AlnuEAST.shp](#)

[Click here to download Supplementary Data: AlnuEAST.shp.xml](#)

[Click here to download Supplementary Data: AlnuEAST.shx](#)

[Click here to download Supplementary Data: AlnuWEST.dbf](#)

[Click here to download Supplementary Data: AlnuWEST.sbn](#)

[Click here to download Supplementary Data: AlnuWEST.sbx](#)

[Click here to download Supplementary Data: AlnuWEST.shp](#)

[Click here to download Supplementary Data: AlnuWEST.shp.xml](#)

[Click here to download Supplementary Data: AlnuWEST.shx](#)

[Click here to download Supplementary Data: celtEAST.dbf](#)

[Click here to download Supplementary Data: celtEAST.sbn](#)

[Click here to download Supplementary Data: celtEAST.sbx](#)

[Click here to download Supplementary Data: celtEAST.shp](#)

[Click here to download Supplementary Data: celtEAST.shp.xml](#)

[Click here to download Supplementary Data: celtEAST.shx](#)

[Click here to download Supplementary Data: celtWEST.dbf](#)

[Click here to download Supplementary Data: celtWEST.sbn](#)

[Click here to download Supplementary Data: celtWEST.sbx](#)

[Click here to download Supplementary Data: celtWEST.shp](#)

[Click here to download Supplementary Data: celtWEST.shp.xml](#)

[Click here to download Supplementary Data: celtWEST.shx](#)

[Click here to download Supplementary Data: cheamFLOR.dbf](#)

[Click here to download Supplementary Data: cheamFLOR.sbn](#)

[Click here to download Supplementary Data: cheamFLOR.sbx](#)

[Click here to download Supplementary Data: cheamFLOR.shp](#)

[Click here to download Supplementary Data: cheamFLOR.shp.xml](#)

[Click here to download Supplementary Data: cheamFLOR.shx](#)

[Click here to download Supplementary Data: cheamMID.dbf](#)

[Click here to download Supplementary Data: cheamMID.sbn](#)

[Click here to download Supplementary Data: cheamMID.sbx](#)

[Click here to download Supplementary Data: cheamMID.shp](#)

[Click here to download Supplementary Data: cheamMID.shp.xml](#)

[Click here to download Supplementary Data: cheamMID.shx](#)

[Click here to download Supplementary Data: fraxEAST.dbf](#)

[Click here to download Supplementary Data: fraxEAST.sbn](#)

[Click here to download Supplementary Data: fraxEAST.sbx](#)

[Click here to download Supplementary Data: fraxEAST.shp](#)

[Click here to download Supplementary Data: fraxEAST.shp.xml](#)

[Click here to download Supplementary Data: fraxEAST.shx](#)

[Click here to download Supplementary Data: fraxWEST.dbf](#)

[Click here to download Supplementary Data: fraxWEST.sbn](#)

[Click here to download Supplementary Data: fraxWEST.sbx](#)

[Click here to download Supplementary Data: fraxWEST.shp](#)

[Click here to download Supplementary Data: fraxWEST.shp.xml](#)

[Click here to download Supplementary Data: fraxWEST.shx](#)

[Click here to download Supplementary Data: LARIX.dbf](#)

[Click here to download Supplementary Data: LARIX.sbn](#)

[Click here to download Supplementary Data: LARIX.sbx](#)

[Click here to download Supplementary Data: LARIX.shp](#)

[Click here to download Supplementary Data: LARIX.shp.xml](#)

[Click here to download Supplementary Data: LARIX.shx](#)

[Click here to download Supplementary Data: piceEAST.dbf](#)

[Click here to download Supplementary Data: piceEAST.sbn](#)

[Click here to download Supplementary Data: piceEAST.sbx](#)

[Click here to download Supplementary Data: piceEAST.shp](#)

[Click here to download Supplementary Data: piceEAST.shp.xml](#)

[Click here to download Supplementary Data: piceEAST.shx](#)

[Click here to download Supplementary Data: piceWEST.dbf](#)

[Click here to download Supplementary Data: piceWEST.sbn](#)

[Click here to download Supplementary Data: piceWEST.sbx](#)

[Click here to download Supplementary Data: piceWEST.shp](#)

[Click here to download Supplementary Data: piceWEST.shp.xml](#)

[Click here to download Supplementary Data: piceWEST.shx](#)

[Click here to download Supplementary Data: pinuEAST.dbf](#)

[Click here to download Supplementary Data: pinuEAST.sbn](#)

[Click here to download Supplementary Data: pinuEAST.sbx](#)

[Click here to download Supplementary Data: pinuEAST.shp](#)

[Click here to download Supplementary Data: pinuEAST.shp.xml](#)

[Click here to download Supplementary Data: pinuEAST.shx](#)

[Click here to download Supplementary Data: pinuNE.dbf](#)

[Click here to download Supplementary Data: pinuNE.sbn](#)

[Click here to download Supplementary Data: pinuNE.sbx](#)

[Click here to download Supplementary Data: pinuNE.shp](#)

[Click here to download Supplementary Data: pinuNE.shp.xml](#)

[Click here to download Supplementary Data: pinuNE.shx](#)

[Click here to download Supplementary Data: pinuSE.dbf](#)

[Click here to download Supplementary Data: pinuSE.sbn](#)

[Click here to download Supplementary Data: pinuSE.sbx](#)

[Click here to download Supplementary Data: pinuSE.shp](#)

[Click here to download Supplementary Data: pinuSE.shp.xml](#)

[Click here to download Supplementary Data: pinuSE.shx](#)

[Click here to download Supplementary Data: pinuWEST.dbf](#)

[Click here to download Supplementary Data: pinuWEST.sbn](#)

[Click here to download Supplementary Data: pinuWEST.sbx](#)

[Click here to download Supplementary Data: pinuWEST.shp](#)

[Click here to download Supplementary Data: pinuWEST.shp.xml](#)

[Click here to download Supplementary Data: pinuWEST.shx](#)

[Click here to download Supplementary Data: PSEUDO.dbf](#)

[Click here to download Supplementary Data: PSEUDO.sbn](#)

[Click here to download Supplementary Data: PSEUDO.sbx](#)

[Click here to download Supplementary Data: PSEUDO.shp](#)

[Click here to download Supplementary Data: PSEUDO.shp.xml](#)

[Click here to download Supplementary Data: PSEUDO.shx](#)

[Click here to download Supplementary Data: querEAST.dbf](#)

[Click here to download Supplementary Data: querEAST.sbn](#)

[Click here to download Supplementary Data: querEAST.sbx](#)

[Click here to download Supplementary Data: querEAST.shp](#)

[Click here to download Supplementary Data: querEAST.shp.xml](#)

[Click here to download Supplementary Data: querEAST.shx](#)

[Click here to download Supplementary Data: querWEST.dbf](#)

[Click here to download Supplementary Data: querWEST.sbn](#)

[Click here to download Supplementary Data: querWEST.sbx](#)

[Click here to download Supplementary Data: querWEST.shp](#)

[Click here to download Supplementary Data: querWEST.shp.xml](#)

[Click here to download Supplementary Data: querWEST.shx](#)

[Click here to download Supplementary Data: tctEAST.dbf](#)

[Click here to download Supplementary Data: tctEAST.sbn](#)

[Click here to download Supplementary Data: tctEAST.sbx](#)

[Click here to download Supplementary Data: tctEAST.shp](#)

[Click here to download Supplementary Data: tctEAST.shp.xml](#)

[Click here to download Supplementary Data: tctEAST.shx](#)

[Click here to download Supplementary Data: tctWEST.dbf](#)

[Click here to download Supplementary Data: tctWEST.sbn](#)

[Click here to download Supplementary Data: tctWEST.sbx](#)

[Click here to download Supplementary Data: tctWEST.shp](#)

[Click here to download Supplementary Data: tctWEST.shp.xml](#)

[Click here to download Supplementary Data: tctWEST.shx](#)

[Click here to download Supplementary Data: tsugEAST.dbf](#)

[Click here to download Supplementary Data: tsugEAST.sbn](#)

[Click here to download Supplementary Data: tsugEAST.sbx](#)

[Click here to download Supplementary Data: tsugEAST.shp](#)

[Click here to download Supplementary Data: tsugEAST.shp.xml](#)

[Click here to download Supplementary Data: tsugEAST.shx](#)

[Click here to download Supplementary Data: tsugWEST.dbf](#)

[Click here to download Supplementary Data: tsugWEST.sbn](#)

[Click here to download Supplementary Data: tsugWEST.sbx](#)

[Click here to download Supplementary Data: tsugWEST.shp](#)

[Click here to download Supplementary Data: tsugWEST.shp.xml](#)

[Click here to download Supplementary Data: tsugWEST.shx](#)