1   **Title:** The accuracy of climate models' simulated season lengths and the effectiveness of
2   grid scale correction factors

3   **Author:** Wade E. Winterhalter

4   **Institution:** University of Central Florida

5   **Contact:** 4000 Central Florida Blvd.

6   Orlando, FL, 32816-2368

7   Phone: 407-230-0134

8   Fax: 407-823-5769

9   Email: wwinterh@mail.ucf.edu

10   **Article Type:** Article

11 ABSTRACT

12   Global climate change is expected to impact biological populations through a
13 variety of mechanisms including increases in the length of their growing season. Climate
14 models are useful tools for predicting how season length might change in the future.
15 However, the accuracy of these models tends to be rather low at regional geographic
16 scales. Here, I determined the ability of several Atmosphere and Ocean General
17 Circulating Models (AOGCMs) to accurately simulate historical season lengths for a
18 temperate ectotherm across the continental United States. I also evaluated the
19 effectiveness of regional scale correction factors to improve the accuracy of these
20 models. I found that both the accuracy of simulated season lengths and the effectiveness
21 of the correction factors to improve the model's accuracy varied geographically and
22 across models. These results suggest that regional specific correction factors do not
23 always adequately remove potential discrepancies between simulated and historically
24 observed environmental parameters. As such, an explicit evaluation of the correction
25 factors' effectiveness should be included in future studies of global climate change's
26 impact on biological populations.

27
28 KEY WORDS:

29 climate change, season length, AOGCM, *Allonemobius socius*, model accuracy

INTRODUCTION

How global climate change will impact both wild and domesticated biological populations is of interest to evolutionary ecologists (Walther et al. 2002, Parmesan and Yohe 2003), agriculturalists (Fuhrer 2003, Nardone et al. 2010), conservationists (McCarty 2001, Thomas et al. 2004), economists (Nordhaus 2001, Stern 2006), policy makers (Jaffe et al. 2009, Ekholm et al. 2010), and the general public (Reiner et al. 2006, Hamilton and Keim 2009). Much of this work involves comparing the environmental tolerances of a species to its predicted future environmental conditions (i.e. projecting climate envelopes; Pearson and Dawson 2003, Thomas et al. 2004). In these studies, local extinctions are expected whenever future environmental conditions within the focal organism's current range fall outside its environmental tolerances. Similarly, range expansions are expected whenever future environmental conditions outside the focal organism's current range fall within its environmental tolerances (assuming unlimited dispersal capability). Together, these local extinctions and range expansions allow researchers to predict how species distributions might shift over time as a result of global climate change.

Although environmental tolerances are undoubtedly an important mechanism through which global climate change will impact biological populations, other mechanisms are likely to be important as well. One of these 'other mechanisms' involves changes in the length of a population's growing season (Linderholm 2006, Christidis et al. 2007). The general warming trend predicted by most climate change scenarios (Nakicenovic and Swart 2000), should increase the amount of time many populations have to complete growth and reproduction. Such increases could lead to shifts in both

53 evolutionary (Zani 2008, Kivela et al. 2009) and ecologically (White et al. 1999, Hudson
54 and Henry 2009) relevant parameters. These shifts are most likely to occur in species
55 with temperature dependent physiologies (i.e. where growth and reproduction rates are a
56 function of external temperatures); and in populations located in temperate regions (i.e.
57 where season lengths are relatively short).

58 For example, the striped ground crickets of the *Allonemobius socius* species group
59 (Traylor et al. 2008) are small bodied terrestrial ectotherms found throughout the
60 continental United States and southern Canada. Differences in the length of these
61 crickets' natal growing seasons are closely associated with variation in several crucial
62 life-history traits including: nymphal development time (Mousseau 1988), body size
63 (Mousseau and Roff 1989aa), wing dimorphism (Mousseau and Roff 1989bb), and
64 diapause incidence (i.e. the proportion of individuals that enter an arrested physiological
65 state; Mousseau and Roff 1989aa, Winterhalter and Mousseau 2007). Determining how
66 these life-history characters will respond to increases in the length of the growing season
67 is a critical step towards our understanding how global climate change will impact this
68 particular group, as well as other temperate ectothermic species.

69 Climate models are an important tool for anticipating the potential environmental
70 changes that will occur over the next century and beyond. One class of models that are
71 frequently used in these projections is called the Atmosphere and Ocean General
72 Circulating Models or AOGCMs. AOGCMs divide the planet into a series of grids, and
73 then use previously established mathematical descriptions of both the internal dynamics
74 (e.g. ocean currents and carbon cycles) and the external forcings (e.g. solar radiance, CO₂
75 emissions) that contribute to climate in order to simulate historical and future

76 environmental conditions. These models are considered one of the most sophisticated
77 physical climate models currently available, and are the primary models used in
78 international climate policy assessments (IPCC 2007). One of the strengths of these
79 models is that they are capable of simulating historical environmental conditions at the
80 continental scale (Christensen et al. 2007). These results suggest that the models'
81 mathematical dynamics adequately approximate climatic processes, and that they will be
82 able to accurately simulate future conditions at the continental scale as well.

83   However, the ability of AOGCMs to simulate historical conditions at the scale of
84 individual grids is considerably lower (Giorgi 2005). This deficiency is particularly
85 significant to those interested in predicting the response of biological populations to
86 global climate change for two reasons. First, the entire range of many species is
87 substantially less than a continent (Brown et al. 1996). As such, applying broad scale
88 environmental predictions to narrowly distributed species may not be appropriate. And
89 second, species that have relatively broad ranges often exhibit geographic variation for a
90 number of phenotypes within their distributions (Endler 1977, Morrison and Hero 2003).
91 How these regional level differences might be affected by global climate change cannot
92 be adequately evaluated using continental scale projections.

93   To compensate for these inaccuracies (i.e. model biases), regional scale correction
94 factors are often applied to a model's raw output (Giorgi and Francisco 2000, Kearney
95 and Porter 2004, Lawler et al. 2009). These correction factors effectively compare the
96 model's simulated historical conditions to the historically observed conditions (usually
97 obtained from weather stations), and then apply the differences to the model's future
98 predictions. In this manner, the magnitude of the change predicted by the model is

maintained, while the discrepancies between the simulated and observed environmental conditions (i.e. the model's inaccuracies) are reduced or eliminated. Although these correction factors have been applied to a variety of environmental parameters (e.g. average temperature, precipitation percent sunlight; Walther et al. 2002, Parmesan 2006), they have not been made within the context of predicting future season lengths of temperate ectotherms.

The purpose of this study was three-fold: 1) to determine the ability of AOGCMs to simulate the historical season lengths of a temperate ectotherm at the regional geographic scale; 2) to generate regional level correction factors for the environmental parameters that contribute to season length; and 3) to evaluate the ability of these correction factors to improve the accuracy of the models. I chose the striped ground crickets from the *A. socius* species group as my focal organism because of their wide distribution (Howard 1983, Traylor et al. 2008), and the intimate relationship between the length of their growing season and their life-history characters (Mousseau and Roff 1989aa, Winterhalter and Mousseau 2007).

METHODS

Season length calculations

Because of their temperature dependent physiology, the length of the growing season for temperate ectotherms is typically measured in accumulated thermal units or degree-days (Bonhomme 2000). Degree-days are dependent on both the external environmental conditions and the range of temperatures that permit the focal organism to metabolize and grow. For example, the upper and lower thermal thresholds for growth in

the *A. socius* species group are 35°C and 13°C, respectively (Mousseau 1988). If these crickets were reared at a constant temperature of 23°C for twenty-four hours, then they would accumulate 10 degree-days greater than 13°C (i.e. 23°C – 13°C = 10 deg-days > 13°C). Likewise, if they were reared at 33°C for twenty-four hours, they would accumulate 20 deg-days.

Unlike most laboratory conditions, temperatures in the wild fluctuate both daily and seasonally. To approximate the number of degree-days accumulated by the *A. socius* species group in a given year (i.e. the season length), I used Allen's dual sine wave methodology (Allen 1976). Briefly, this method fits a series of sine waves to the daily maximum and minimum temperatures experienced in a given habitat over an entire year. The number of degree-days accumulated during this period is set equal to the area underneath these sine wave curves that also fall within the organism's thermal limits for growth.

For this study, I assumed that the upper and lower thermal limits for growth of the *A. socius* species group were constant across their entire range (upper limit = 35°C, lower limit = 13°C). The daily maximum and minimum temperatures were obtained from either the output files of an AOGCM in the case of the simulated season lengths, or from weather stations in the case of historical observations.

Simulated data

The primary climate model examined in this study was the AOGCM produced by the Geophysical Fluid Dynamics Laboratory designated cm2.0 (Delworth et al. 2006). I chose this particular model because the arrangement of its grids happened to correspond

to regional differences in the life-history strategies of the *A. socius* species group (Mousseau and Roff 1989aa). Simulated daily maximum and minimum temperatures (i.e. tasmax and tasmin) were obtained for all grids within the continental United States. Only grids that were composed of at least 50% land area, based on visual inspection (number of grids = 150). All available records from January 1, 1961 to December 31, 2000 were included in the analysis.

In addition to the GFDL model cm2.0, I also examined 10 other AOGCMs: GFDL's model cm2.1, CCCMA's model cgm3.1, CNRM's model cm3, CSIRO's model mk3.0, GISS's model e20, IAP's model cm4v1, MIRCO's model v3.2, MPI's model echam5, and MRI's model cgcm2.3.2a. For each of these models, I randomly selected 20 grids from within the continental United States that were comprised of at least 50% land area and obtained all of the simulated daily maximum and minimum temperatures that were available from January 1, 1961 to December 31, 2000. The simulated temperatures from all of the models examined in this study represented environmental conditions two meters from the surface (Meehl et al. 2007).

For each model, grid, and year; the length of the growing season was estimated in terms of accumulated degree-days > 13°C using Allen's dual sine wave method (Allen 1976). Because I focused on the geographic scale of individual grids, downscaling methods were not employed. All of the simulated data were obtained with permission from the World Climate Research Programme's Coupled Model Intercomparison Project 3 (i.e. WCRP-CMIP3; Meehl et al. 2007). In some cases, the data files contained internal errors and could not be accessed. These data were excluded from the analysis.

Historical data

The historically observed daily maximum and minimum temperature records from January 1, 1961 to December 31, 2000 were obtained from multiple weather stations located within 1° latitude and 1° longitude of the center of each grid at standard height (1.25m). Within the timeframe, an average of 6.2 ± 0.5 ($\pm$ stdev) stations were available for each grid examined per year. Missing data from the historical records were replaced by that year's monthly average minimum or maximum temperature. Months in which records were missing for 15 or more days were excluded from the analysis. Years in which less than 12 months of records were available were also removed. These decisions led to an average of 36.9 ± 1.7 ($\pm$ stdev) years being examined for each grid.

For each weather station and year, the length of the growing season was estimated in terms of accumulated degree-days $> 13^{\circ}\text{C}$ using Allen's dual sine wave method (Allen 1976). Within each grid and year, these season lengths were averaged ($\pm$ stdev) across weather stations in order to obtain estimates of both the mean season length and its variation within a given grid. All of the historical data were obtained with permission from the National Climate Data Center (NCDC 2009).

Season length analysis

For each grid and year, the simulated season length produced by a given model was compared to the mean ($\pm$ stdev) of the historically observed season lengths using a one sample t-test. A total of 5,089 of these one sample t-tests were performed for the primary analysis (GFDL model cm2.0) and 7,364 one sample t-tests were performed for the other models. The proportion of years in which a significant difference between the

191 simulated and historical season lengths occurred was noted. Based on an $\alpha = 0.05$, I
192 expected 5% of the years examined would be significant due to chance alone. Finally,
193 the percent difference between the simulated and historical season lengths were estimated
194 for each grid, year and model.

196 *Correction factors and adjusted season lengths*

197 The correction factors were estimated independently for each model and grid.
198 First, the simulated daily maximum and minimum temperatures were subtracted from the
199 average historically observed daily maximum and minimum temperatures respectively
200 for each day in which data were available. These differences were then averaged within
201 each year, and the results were finally averaged across years. This procedure led to a
202 single correction factor for each of the temperature extremes that represented the average
203 difference between the simulated and historical data across all years. Standard deviations
204 for these estimates were also calculated and represent the across year variation in the
205 correction factor.

206 These model and grid specific correction factors were then added to each of the
207 simulated maximum and minimum temperatures. The simulated season lengths were
208 recalculated, now using the adjusted maximum and minimum temperatures, and analyzed
209 in the same manner as the unadjusted model estimations.

210 All statistics and calculations were performed using R (R 2009).

212 RESULTS

213 *Primary model*

The AOGCM that was the primary focus of this study (i.e. GFDL's model cm2.0) consistently underestimated the length of the growing season across the continental United States. Across all grids and years, 0.91 ± 0.13 ($\pm$ stdev) of the simulated season lengths were significantly different from the historical records (Table 1). On average, the simulated season lengths were $41.4 \pm 18.2\%$ ($\pm$ stdev) less than the historical records (Table 1). The largest discrepancies between the simulated and historical season lengths were found in western Montana, Idaho and Wyoming (Figure 1A). In this area, the simulated season lengths were as much as 80% less than the historical observations (see 45°N , 108.75°W ; Table 1). The model performed best in the deep south (Figure 1A). But even in this region, the simulated season lengths were at least 11% lower than the historical observations (see 33°N , 91.25°W ; Table 1).

Not surprisingly, the GFDL model cm2.0 also under-estimated both the daily maximum and minimum temperatures. However, the magnitude of this under-estimation was greater for the daily maximum temperatures. When averaged across all grids, the maximum temperatures were under-estimated by $5.9 \pm 1.9^\circ\text{C}$ ($\pm$ stdev), while the daily minimum temperatures were under-estimated by only $2.6 \pm 1.6^\circ\text{C}$ ($\pm$ stdev; Table 2). The largest differences in daily maximum temperature estimates were generally found in the north-western quarter of the country with the exception of a few regions along the east coast (Figure 2A). In contrast, the largest differences for the minimum temperatures were found primarily in the south-western portion of the country (Figure 2B).

After using the grid specific correction factors (Table 1) to adjust the simulated daily maximum and minimum temperatures (see Methods above), the accuracy of the model's simulated season lengths improved. The proportion of season lengths across all

years and grids in which significant differences were detected was cut nearly in half from 0.91 before the adjustments to 0.47 after the adjustments were applied (Table 1). In addition, the average difference in season length estimates went from an under-estimation of $41.4 \pm 18.2\%$ to an over-estimation of $2.1 \pm 8.2\%$ (Table 1). The adjusted season lengths were closest to the historical season lengths in the central portion of the United States and along the west coast (Figure 1B). In addition to improving the accuracy of the simulated season lengths, the correction factors also significantly reduced the variation in accuracy across grids (Figure 3). The standard deviation of the model's accuracy was 18.2% before the adjustment, but only 8.2% after the adjustment was applied ($F_{149,149} = 10.0$, $P < 0.0001$).

Cross model comparisons:

Substantial variation in the ability to accurately simulate season lengths was observed across all of the AOGCMs examined in this study (Figure 4). The GISS model was the least accurate, and on average under-estimated the season length by $44.3 \pm 26.8\%$ ($\pm$ stdev). The IAP and MIRCO models were the most accurate, over-estimating the season length by an average of $1.2 \pm 32.6\%$ and $3.5 \pm 29.4\%$ respectively ($\pm$ stdev). These were the only two models in which the simulated season lengths based on the models' raw data were not significantly different from the historical observations (IAP: $t_{19} = 0.16$, $P = 0.8710$; MIRCO: $t_{19} = 0.53$, $P = 0.6006$). All of the other models examined significantly under-estimated the historical season lengths across grids ($\alpha = 0.05$).

Generally after the correction factors were added to the models' output, the accuracy of the simulated season lengths was improved. For most models, the average difference between the simulated and historical season lengths decreased after the correction factors were applied (Figure 4). Across all models, the difference between the simulated and historical season lengths went from an under-estimation of $18.6 \pm 4.5\%$ ($\pm$ stdev) prior to the correction factors being applied to an over-estimation of $1.2 \pm 6.7\%$ ($\pm$ stdev) afterwards. In addition, the proportion of years in which a significant difference was detected between the simulated and historical season lengths decreased for all models (Table 2). Before the correction factors were applied, 0.71 ± 0.11 ($\pm$ stdev) of the simulated season lengths were significantly different from the historical records (averaged across all models), but only 0.41 ± 0.08 were significantly different afterwards (Table 2). This difference represented a significant improvement in the models' ability to simulate historical season lengths ($t_{20} = 7.4$, $P < 0.0001$).

Despite the fact that the adjusted season lengths were generally more accurate than the unadjusted estimates, significant differences between the adjusted and historical season lengths were still detected in four of the eleven models examined (Figure 4). In two of these four models (GFDL 2.0 and GFDL 2.1) the differences were relatively small, and in the case of the GFDL 2.0 model's the statistical power was quite large (number of grids = 150). The accuracy of the GISS model was improved after the correction factors were applied (Figure 4), although the simulated season lengths were still $19.3 \pm 12.9\%$ lower than the historical observations. In contrast, the accuracy of the IAP model actually decreased after the correction factors were applied (Figure 4). Prior to the adjustment, the season length estimate in the average grid of the IAP model was

1.2% greater than the historical observation, while after the adjustment it was 9.9% greater. However, the variation in accuracy of this model across grids did decrease significantly after the correction factors were applied. The standard deviation before the adjustment was 32.6%, but only 11.4% afterwards ($F_{19,19} = 2.86$, $P = 0.0135$).

DISCUSSION

Although the general inability of AOGCMs to simulate historical season lengths at the regional geographic scale was expected, the amount and levels of variation were surprising. In this study, the accuracy of the season length predictions varied both geographically (Figure 1A), as well as across models (Figure 4).

One potential explanation for the geographic variation observed in this study involves the number of weather stations within each grid that were available for analysis. 46 of the 150 grids examined in the primary model had fewer than 10 weather stations (Table 1). Within this sub-group, sample size had a significant effect on the accuracy of the simulated season length ($F_{1,45} = 8.5$, $P = 0.006$, $r^2 = 0.16$). For each additional weather station, the accuracy of the simulated season lengths increased by an average of $4.2 \pm 1.4\%$ ($\pm$ s.e.). However, this sample size effect was not present for grids that had 10 or more weather stations available ($F_{1,101} < 0.1$, $P = 0.995$, $r^2 < 0.01$). In addition, the grids that exhibited the greatest discrepancies between the simulated and historical season lengths (i.e. Montana, Wyoming, Idaho; Figure 1A) had an average of 10.4 ± 1.5 (s.e.) weather stations. This relatively large sample size in an area of low accuracy suggests that sample size alone was not the sole cause of the inaccuracies found in Montana, Wyoming and Idaho. The most likely explanation for these inaccuracies is the low

spatial resolution of AOGCMs at higher altitudes. This lack of spatial resolution has been found to under-estimate albedo feedback and daily high temperatures (Christensen et al. 2007), which in turn would lead to the large under-estimation of season length found in this study. Thus, geographic variation in the ability of AOGCMs to simulate historical season lengths appears to be due, in part, to their spatial resolutions. As such, I would expect geographic variation to be lower for climate models with higher spatial resolutions. These meso-scale climate models may be better than AOGCMs in predicting how biological populations will respond to climate change, particularly if research interests are confined to a specific area or region. Furthermore, as AOGCMs are improved and refined, I expect the geographic variation in their accuracy to decrease as well.

Unlike the geographic variation observed within the primary model, variation across AOGCMs did not appear to be related to their spatial resolutions. The relationship between grid size and the accuracy of the simulated season length across models was not significant ($F_{1,9} = 0.4$, $P = 0.536$, $r^2 = 0.04$). This result was not surprising given that the AOGCMs used in this study had similar spatial resolutions. On average, grids were $2.6 \pm 0.3^\circ$ in longitude and $2.3 \pm 0.3^\circ$ in latitude ($\pm$ s.e.). The inter-model variation observed in this study is most likely caused by differences in the mathematical composition of the AOGCMs, particularly how cloud feedback is incorporated (Randall et al. 2007). This inter-model variation is likely to persist as different research groups employ alternate strategies to optimize mathematical complexity, simulation run time and parameterization.

The variation observed across both geographic areas and separate AOGCMs emphasize the fact that no universal set of correction factors will consistently improve the predictions of all AOGCMs. As such, the most appropriate set of correction factors for a given study will be dependent on its specific goals, particularly on the geographic areas of interest and the specific AOGCM being used.

Despite the large amounts of variation observed in the raw AOGCMs season length simulations, most models consistently under-estimated historical season lengths. This under-estimation of season length appears to be driven primarily by the model's under-estimation of daily maximum temperatures. The geographic distribution of season length discrepancies (Figure 1A) and daily maximum temperature correction factors (Figure 2A) were not only closely associated, but the correlation between accuracy and the daily maximum correction factor across models (Table 2) was also significantly positive ($r = 0.811$, $t_{10} = 4.2$, $P = 0.0025$). In contrast, the distribution of daily minimum correction factors (Figure 2B) and season lengths discrepancies (Figure 1) did not appear to be strongly associated, while the correlation of accuracy and daily minimum correction factors across models was positive, but not significant (Table 2; $r = 0.458$, $t_{10} = 1.5$, $P = 0.1564$).

The stronger association between daily maximum temperatures and season length estimates should be expected, given that for a large portion of the year, daily minimum temperatures will fall below the lower thermal threshold for development of the focal species (i.e. 13°C). As such, under-estimations of the daily minimum temperatures will not contribute to the accumulation of degree-days during these portions of the year. In contrast, daily maximum temperatures will frequently be within the focal organism's

upper (35°C) and lower (13°C) thermal thresholds for growth. Because of this relationship, any consistent under-estimation in simulated daily high temperatures will lead to an under-estimation of accumulated degree-days (i.e. season length).

Because the season length estimates of temperate ectotherms are dependent on both the daily maximum and minimum temperatures, and the upper and lower thermal thresholds for development of the focal organism; the accuracy of simulated season lengths may vary across species. Organisms with a lower thermal threshold substantially below 13°C should be more sensitive to under-estimations of daily low temperatures than *A. socius* species group. This increased sensitivity would be due to the fact that more of the observed daily minimum temperatures would fall within the organism's usable thermal range. Likewise, organisms with upper thermal threshold substantially higher than 35°C should be more sensitive to under-estimations of daily maximum temperatures than the *A. socius* species group, because a higher proportion of the daily maximum temperatures would fall within their usable thermal range.

Interestingly, regardless of the upper and lower thermal thresholds of a particular organism, the correction factors used to improve the accuracy of the simulated season lengths would be the same. Because the estimation of the correction factors are based only on the differences between the simulated and historical maximum and minimum temperatures, and do not take into consideration the organism's thermal thresholds, the grid specific correction factors estimated here (Table 1) could be used for any ectothermic organism residing within the continental United States. However, because the models simulated daily maximum and minimum temperatures are "near surface", meaning that they approximate air temperatures two meters above the ground (Meehl et

al. 2007), they do not necessarily represent the range of temperature micro-habitats available to organisms within an area (Sinclair et al. 2003; Inouye 2008). Organisms capable of selecting micro-habitats within a particular region may not respond as predicted by any climate simulation. As such, population level predictions based on either raw or corrected model data should be interpreted with caution.

Perhaps the most interesting observation of this study was the large amount of variation in the ability of the correction factors to improve the accuracy of the simulated season lengths. The average difference between the simulated season lengths after the correction factors were applied and the historical observations varied both geographically (Figure 1B) and across models (Figure 4). Although this variation in the effectiveness of region specific correction factors is likely to exist for all environmental parameters; it has not been previously reported, and it is not typically included in studies of the impact of global climate change on biological populations. Usually in these studies, the manner in which region specific correction factors were obtained is outlined in the methods section without any evaluation of their effectiveness. The implicit assumption of these studies is that the correction factors adequately remove potential discrepancies between the simulated and historical environmental parameters. However, the results of this study demonstrated that even after the application of correction factors, average simulated environmental parameters can still be off by as much as 19.3% (GISS, Figure 4). Therefore, the effectiveness of region specific correction factors should be explicitly considered and reported in future studies of global climate change impacts on biological populations.

396 ACKNOWLEDGEMENTS

397 I would like to thank A. Weltzian, J. Hays, V. Lee, S. Brimner, K. Lindholm, E. Chandler
398 and A. Taraboletti for their assistance in data collection. I would also like to thank two
399 reviewers whose comments improved an earlier draft. This research was supported by
400 the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-
401 07ER64455 to W.E. Winterhalter.

LITERATURE CITED

- Allen, J. C. 1976. Modified sine wave method for calculating degree days. *Environmental Entomology* **5**:388-396.
- Bonhomme, R. 2000. Bases and limits to using 'degree.day' units. *European Journal Of Agronomy* **13**:1-10.
- Brown, J. H., G. C. Stevens, and D. M. Kaufman. 1996. The geographic range: Size, shape, boundaries, and internal structure. *Annual Review Of Ecology And Systematics* **27**:597-623.
- Christensen, J. H., B. Hewitson, A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R. K. Kolli, W. T. Kwon, R. Laprise, V. Magaña Rueda, L. Mearns, C. G. Menéndez, J. Räisänen, A. Rinke, A. Sarr, and P. Whetton. 2007. Regional Climate Projections. *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. *Climate change 2007: the physical science basis. contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, UK, New York, NY, USA.
- Christidis, N., P. A. Stott, S. Brown, D. J. Karoly, and J. Caesar. 2007. Human contribution to the lengthening of the growing season during 1950-99. *Journal Of Climate* **20**:5441-5454.
- Delworth, T. L., A. J. Broccoli, A. Rosati, R. J. Stouffer, V. Balaji, J. A. Beesley, W. F. Cooke, K. W. Dixon, J. Dunne, K. A. Dunne, J. W. Durachta, K. L. Findell, P. Ginoux, A. Gnanadesikan, C. T. Gordon, S. M. Griffies, R. Gudgel, M. J. Harrison, I. M. Held, R. S. Hemler, L. W. Horowitz, S. A. Klein, T. R. Knutson,

425 P. J. Kushner, A. R. Langenhorst, H. C. Lee, S. J. Lin, J. Lu, S. L. Malyshev, P.
 426 C. D. Milly, V. Ramaswamy, J. Russell, M. D. Schwarzkopf, E. Shevliakova, J. J.
 427 Sirutis, M. J. Spelman, W. F. Stern, M. Winton, A. T. Wittenberg, B. Wyman, F.
 428 Zeng, and R. Zhang. 2006. GFDL's CM2 global coupled climate models. Part I:
 429 Formulation and simulation characteristics. *Journal Of Climate* **19**:643-674.
 430 Ekholm, T., S. Soimakallio, S. Moltmann, N. Hohne, S. Syri, and I. Savolainen. 2010.
 431 Effort sharing in ambitious, global climate change mitigation scenarios. *Energy*
 432 *Policy* **38**:1797-1810.
 433 Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton University
 434 Press, Princeton, New Jersey.
 435 Fuhrer, J. 2003. Agroecosystem responses to combinations of elevated CO₂, ozone, and
 436 global climate change. *Agriculture Ecosystems & Environment* **97**:1-20.
 437 Giorgi, F. 2005. Climate change prediction. *Climatic Change* **73**:239-265.
 438 Giorgi, F. and R. Francisco. 2000. Evaluating uncertainties in the prediction of regional
 439 climate change. *Geophysical Research Letters* **27**:1295-1298.
 440 Hamilton, L. C. and B. D. Keim. 2009. Regional variation in perceptions about climate
 441 change. *International Journal of Climatology* **29**:2348-2352.
 442 Howard, D. J. 1983. Electrophoretic Survey Of Eastern North-American *Allonemobius*-
 443 (*Orthoptera*, *Gryllidae*) - Evolutionary Relationships And The Discovery Of 3
 444 New Species. *Annals Of The Entomological Society Of America* **76**:1014-1021.
 445 Hudson, J. M. G. and G. H. R. Henry. 2009. Increased plant biomass in a High Arctic
 446 heath community from 1981 to 2008. *Ecology* **90**:2657-2663.

447 Inouye D.W. 2008. Effects of climate change on phenology, frost damage, and floral
 448 abundance of montane wildflowers. *Ecology* **89**:353-362.

449 IPCC. 2007. Summary for policymakers. *in* S. Solomon, D. Qin, M. Manning, Z. Chen,
 450 M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. Climate change
 451 2007: the physical science basis. contribution of working group I to the fourth
 452 assessment report of the intergovernmental panel on climate change Cambridge
 453 University Press, Cambridge, UK
 454 New York, NY, USA.

455 Jaffe, J., M. Ranson, and R. N. Stavins. 2009. Linking Tradable Permit Systems: A Key
 456 Element of Emerging International Climate Policy Architecture. *Ecology Law*
 457 *Quarterly* **36**:789-808.

458 Kearney, M. and W. P. Porter. 2004. Mapping the fundamental niche: Physiology,
 459 climate, and the distribution of a nocturnal lizard. *Ecology* **85**:3119-3131.

460 Kivela, S. M., P. Valimaki, J. Oksanen, A. Kaitala, and V. Kaitala. 2009. Seasonal Clines
 461 of Evolutionarily Stable Reproductive Effort in Insects. *American Naturalist*
 462 **174**:526-536.

463 Lawler, J. J., S. L. Shafer, D. White, P. Kareiva, E. P. Maurer, A. R. Blaustein, and P. J.
 464 Bartlein. 2009. Projected climate-induced faunal change in the Western
 465 Hemisphere. *Ecology* **90**:588-597.

466 Linderholm, H. W. 2006. Growing season changes in the last century. *Agricultural And*
 467 *Forest Meteorology* **137**:1-14.

468 McCarty, J. P. 2001. Ecological consequences of recent climate change. *Conservation*
 469 *Biology* **15**:320-331.

470 Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J.
471 Stouffer, and K. E. Taylor. 2007. The WCRP CMIP3 multimodel dataset - A new
472 era in climate change research. *Bulletin Of The American Meteorological Society*
473 **88**:1383-1394.

474 Morrison, C. and J. M. Hero. 2003. Geographic variation in life-history characteristics of
475 amphibians: a review. *Journal Of Animal Ecology* **72**:270-279.

476 Mousseau, T. A. 1988. Life history evolution in a seasonal environment: a case study.
477 McGill University.

478 Mousseau, T. A. and D. A. Roff. 1989a. Adaptation to seasonality in a cricket - patterns
479 of phenotypic and genotypic variation in body size and diapause expression along
480 a cline in season length. *Evolution* **43**:1483-1496.

481 Mousseau, T. A. and D. A. Roff. 1989b. Geographic Variability In The Incidence And
482 Heritability Of Wing Dimorphism In The Striped Ground Cricket, *Allonemobius-*
483 *Fasciatus*. *Heredity* **62**:315-318.

484 Nakicenovic, N. and R. Swart, editors. 2000. Special report on emissions scenarios. A
485 special report of working group III of the intergovernmental panel on climate
486 change. Cambridge University Press, Cambridge, UK
487 New York, NY, USA.

488 Nardone, A., B. Ronchi, N. Lacetera, M. S. Ranieri, and U. Bernabucci. 2010. Effects of
489 climate changes on animal production and sustainability of livestock systems.
490 *Livestock Science* **130**:57-69.

491 NCDC. 2009. Climate-radar inventories. US Department of Commerce.

492 Nordhaus, W. D. 2001. Climate change - Global warming economics. *Science* **294**:1283-
 493 1284.
 494 Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change.
 495 *Annual Review Of Ecology Evolution And Systematics* **37**:637-669.
 496 Parmesan, C. and G. Yohe. 2003. A globally coherent fingerprint of climate change
 497 impacts across natural systems. *Nature* **421**:37-42.
 498 Pearson, R. G. and T. P. Dawson. 2003. Predicting the impacts of climate change on the
 499 distribution of species: are bioclimate envelope models useful? *Global Ecology*
 500 *and Biogeography* **12**:361-371.
 501 R. 2009. R: a language and environment for statistical computing. R foundation for
 502 statistical computing, Vienna, Austria.
 503 Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A.
 504 Pitman, J. Shukla, J. Srinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor. 2007.
 505 Climate models and their evaluation. *in* S. Solomon, D. Qin, M. Manning, Z.
 506 Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. Climate
 507 change 2007: the physical science basis. contribution of working group I to the
 508 fourth assessment report of the intergovernmental panel on climate change
 509 Cambridge University Press, Cambridge, UK
 510 New York, NY, USA.
 511 Reiner, D. M., T. E. Curry, M. A. De Figueiredo, H. J. Herzog, S. D. Ansolabehere, K.
 512 Itaoka, F. Johnsson, and M. Odenberger. 2006. American exceptionalism?
 513 Similarities and differences in national attitudes toward energy policy and global
 514 warming. *Environmental Science & Technology* **40**:2093-2098.

515 Sinclair B.J., Vernon P., Klok C.J. & Chown S.L. 2003. Insects at low temperatures: an
 516 ecological perspective. *Trends In Ecology & Evolution* **18**:257-262.

517 Stern, N. 2006. Stern review on the economics of climate change. *Stern Review on the*
 518 *Economics of Climate Change*, UK Treasury.

519 Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C.
 520 Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L.
 521 Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-
 522 Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk
 523 from climate change. *Nature* **427**:145-148.

524 Traylor, T., A. C. Birand, J. L. Marshall, and D. J. Howard. 2008. A zone of overlap and
 525 hybridization between *Allonemobius socius* and a new *Allonemobius* sp. *Annals*
 526 *Of The Entomological Society Of America* **101**:30-39.

527 Walther, G. R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. M.
 528 Fromentin, O. Hoegh-Guldberg, and F. Bairlein. 2002. Ecological responses to
 529 recent climate change. *Nature* **416**:389-395.

530 White, M. A., S. W. Running, and P. E. Thornton. 1999. The impact of growing-season
 531 length variability on carbon assimilation and evapotranspiration over 88 years in
 532 the eastern US deciduous forest. *International Journal Of Biometeorology* **42**:139-
 533 145.

534 Winterhalter, W. E. and T. A. Mousseau. 2007. Patterns of phenotypic and genetic
 535 variation for the plasticity of diapause incidence. *Evolution*.

536 Zani, P. A. 2008. Climate Change Trade-Offs in the Side-Blotched Lizard (*Uta*
537 *stansburiana*): Effects of Growing-Season Length and Mild Temperatures on
538 Winter Survival. *Physiological And Biochemical Zoology* **81**:797-809.
539

Table 1: Summary statistics for GFDL cm2.0. Grids are organized by longitude (°W) and latitude (°N). Sample sizes are broken down into the number of Years examined and the number of weather stations per year (WS). Pr(years signif) represents the proportion of years in which a significant difference in season length was detected, and Accuracy represents the percent difference between the simulated and historical season lengths. Both of these variables were calculated using the model's raw (Model) and adjusted (Adjust) output. The Correction Factors used to make the adjustments are also presented. Accuracy and Correction factors are given $\pm$ stdev (the across year variation).

	Long	Lat	Years	WS	Pr (years signif.)		Accuracy (%)			Correction Factors (°C)	
					Model	Adjust	Model	Adjust		Maximum	Minimum
544	123.75	43	35	10	0.97	0.20	-46.9 $\pm$ 12.0	-5.6 $\pm$ 15.3		-4.8 $\pm$ 0.9	-0.2 $\pm$ 0.8
545		41	35	14	0.69	0.14	-33.0 $\pm$ 11.6	8.2 $\pm$ 14.4		-4.9 $\pm$ 0.8	0.3 $\pm$ 0.9
546	121.25	47	35	10	0.89	0.34	-51.1 $\pm$ 16.0	9.8 $\pm$ 22.6		-5.4 $\pm$ 1.0	-4.3 $\pm$ 0.9
		45	35	5	1.00	0.23	-62.2 $\pm$ 11.7	-11.4 $\pm$ 16.9		-5.9 $\pm$ 1.0	-3.3 $\pm$ 0.9
547		43	35	14	1.00	0.46	-45.0 $\pm$ 14.0	-1.6 $\pm$ 18.9		-5.3 $\pm$ 1.1	-0.4 $\pm$ 0.9
		41	35	9	0.94	0.26	-39.4 $\pm$ 11.8	8.8 $\pm$ 16.4		-5.8 $\pm$ 1.1	-1.2 $\pm$ 0.9
548		39	35	10	0.89	0.00	-43.2 $\pm$ 7.9	0.6 $\pm$ 11.5		-5.5 $\pm$ 1.1	-4.5 $\pm$ 0.8
		37	35	8	0.66	0.06	-29.7 $\pm$ 10.2	4.7 $\pm$ 13.6		-3.8 $\pm$ 0.9	-2.5 $\pm$ 1.1
549	118.75	47	35	10	1.00	0.60	-68.2 $\pm$ 9.8	-11.9 $\pm$ 13.5		-7.4 $\pm$ 1.0	-4.6 $\pm$ 0.9
		45	28	10	0.96	0.11	-70.0 $\pm$ 9.3	-11.6 $\pm$ 14.7		-7.8 $\pm$ 1.1	-3.6 $\pm$ 0.9
550		43	17	5	0.76	0.00	-61.3 $\pm$ 6.8	-2.1 $\pm$ 10.8		-7.6 $\pm$ 1.0	-2.9 $\pm$ 0.8
		41	20	5	1.00	0.20	-64.8 $\pm$ 6.6	-2.4 $\pm$ 10.5		-9.7 $\pm$ 1.2	-3.9 $\pm$ 1.1
551		39	35	5	0.94	0.03	-62.6 $\pm$ 7.6	-0.9 $\pm$ 15.4		-9.6 $\pm$ 1.3	-4.6 $\pm$ 1.5
		37	35	22	1.00	0.20	-56.9 $\pm$ 7.1	-4.1 $\pm$ 13.0		-7.6 $\pm$ 1.4	-6.7 $\pm$ 1.3
552	116.25	35	35	10	1.00	0.14	-40.6 $\pm$ 7.9	2.3 $\pm$ 11.4		-5.4 $\pm$ 1.2	-5.3 $\pm$ 0.9
		47	31	10	0.84	0.10	-67.1 $\pm$ 11.3	-1.5 $\pm$ 15.8		-7.4 $\pm$ 1.1	-4.4 $\pm$ 1.0
553		45	35	10	1.00	0.17	-71.7 $\pm$ 9.8	-13.2 $\pm$ 15.5		-8.1 $\pm$ 1.2	-2.7 $\pm$ 1.1
		43	35	21	1.00	0.51	-71.0 $\pm$ 7.8	-4.8 $\pm$ 13.1		-9.6 $\pm$ 1.2	-5.8 $\pm$ 0.9
554		41	19	4	1.00	0.11	-61.6 $\pm$ 7.6	2.1 $\pm$ 11.2		-9.2 $\pm$ 1.4	-3.6 $\pm$ 1.1
		39	21	3	1.00	0.00	-57.9 $\pm$ 7.6	8.3 $\pm$ 13.9		-9.0 $\pm$ 1.1	-5.4 $\pm$ 0.8
555		37	34	6	0.97	0.12	-60.5 $\pm$ 6.2	-2.1 $\pm$ 11.8		-10.2 $\pm$ 1.2	-7.7 $\pm$ 1.3
		35	31	5	0.45	0.00	-42.7 $\pm$ 8.2	-0.2 $\pm$ 12.5		-5.9 $\pm$ 1.8	-6.5 $\pm$ 0.9
556	113.75	33	35	16	0.89	0.06	-28.6 $\pm$ 8.5	-0.8 $\pm$ 10.6		-3.8 $\pm$ 1.2	-3.0 $\pm$ 0.8
		47	35	10	1.00	0.43	-76.2 $\pm$ 10.3	-4.4 $\pm$ 18.3		-8.2 $\pm$ 1.2	-4.3 $\pm$ 0.9
557		45	35	8	1.00	0.26	-78.1 $\pm$ 9.0	-8.4 $\pm$ 17.0		-9.1 $\pm$ 1.1	-3.3 $\pm$ 0.9
		43	35	10	1.00	0.43	-66.5 $\pm$ 9.7	2.1 $\pm$ 16.3		-8.8 $\pm$ 1.3	-4.9 $\pm$ 0.9
558		41	21	4	0.90	0.00	-63.8 $\pm$ 8.9	-5.3 $\pm$ 13.9		-9.1 $\pm$ 1.3	-5.1 $\pm$ 1.1
		39	35	10	1.00	0.46	-55.2 $\pm$ 8.1	4.4 $\pm$ 13.5		-8.9 $\pm$ 1.1	-4.6 $\pm$ 1.0
559		37	35	6	0.63	0.00	-45.3 $\pm$ 8.1	1.7 $\pm$ 12.7		-7.2 $\pm$ 1.3	-4.5 $\pm$ 1.0
		35	21	6	0.52	0.00	-48.3 $\pm$ 5.4	-5.4 $\pm$ 6.8		-7.9 $\pm$ 1.0	-6.4 $\pm$ 1.0
560	111.25	33	34	5	1.00	0.26	-40.3 $\pm$ 4.8	-3.9 $\pm$ 5.9		-6.2 $\pm$ 1.0	-6.5 $\pm$ 0.8
		47	35	10	1.00	0.60	-78.5 $\pm$ 8.1	-7.5 $\pm$ 15.1		-8.9 $\pm$ 1.4	-4.3 $\pm$ 1.0
561		45	35	10	1.00	0.06	-78.3 $\pm$ 9.3	-7.9 $\pm$ 20.1		-8.3 $\pm$ 1.6	-5.0 $\pm$ 1.3
		43	35	19	1.00	0.54	-63.9 $\pm$ 11.8	5.1 $\pm$ 20.1		-7.7 $\pm$ 1.2	-2.8 $\pm$ 0.8
562		41	35	10	1.00	0.20	-61.5 $\pm$ 8.6	7.7 $\pm$ 15.6		-9.3 $\pm$ 1.2	-5.7 $\pm$ 1.0
		39	35	7	0.97	0.43	-47.8 $\pm$ 8.8	14.3 $\pm$ 14.7		-8.2 $\pm$ 1.2	-4.5 $\pm$ 1.0
		37	35	13	0.86	0.06	-42.2 $\pm$ 8.5	7.1 $\pm$ 14.4		-6.6 $\pm$ 1.3	-4.8 $\pm$ 1.1
		35	35	10	0.43	0.00	-30.1 $\pm$ 9.6	6.2 $\pm$ 11.7		-5.6 $\pm$ 1.0	-2.2 $\pm$ 0.8
		33	35	10	1.00	0.03	-46.1 $\pm$ 4.9	-3.7 $\pm$ 6.3		-7.8 $\pm$ 0.9	-6.8 $\pm$ 1.0

Table 1 cont.:

					Pr (years sigif.)		Accuracy (%)			Correction Factors (°C)		
	Long	Lat	Years	WS	Model	Adjust	Model	Adjust		Maximum	Minimum	
564	108.75	47	35	6	1.00	0.34	-70.4 ±	9.6	-8.9 ±	15.4	-7.8 ±	1.5
565		45	35	10	1.00	0.51	-80.3 ±	6.2	-9.7 ±	13.4	-9.7 ±	1.4
566		43	35	6	1.00	0.31	-77.7 ±	7.0	-9.4 ±	14.2	-10.2 ±	1.4
		41	34	4	1.00	0.06	-65.5 ±	8.4	2.9 ±	14.8	-10.3 ±	1.4
		39	35	6	0.94	0.03	-58.1 ±	8.2	10.0 ±	15.3	-9.9 ±	1.2
567		37	35	21	1.00	0.49	-41.6 ±	9.4	15.2 ±	14.3	-7.4 ±	0.9
		35	35	15	0.69	0.77	-18.9 ±	11.5	22.8 ±	14.5	-5.4 ±	0.9
568		33	35	10	0.46	0.09	-25.2 ±	10.0	9.8 ±	12.1	-6.1 ±	1.1
	106.25	47	35	8	1.00	0.83	-68.0 ±	9.9	-17.1 ±	13.9	-7.5 ±	1.7
569		45	35	10	1.00	0.83	-72.7 ±	7.1	-11.9 ±	14.2	-8.4 ±	1.5
		43	29	4	1.00	0.28	-73.3 ±	8.2	-5.2 ±	17.0	-9.1 ±	1.5
570		41	35	10	0.97	0.40	-58.3 ±	13.7	14.6 ±	23.4	-7.7 ±	1.3
		39	35	10	0.51	0.40	-36.8 ±	20.8	22.5 ±	29.8	-6.0 ±	1.4
571		37	35	10	0.71	0.43	-33.8 ±	15.7	21.5 ±	21.2	-6.5 ±	1.2
		35	35	10	0.74	0.34	-34.5 ±	11.4	17.0 ±	15.3	-7.2 ±	1.1
572		33	35	10	0.60	0.09	-28.4 ±	10.8	11.2 ±	13.2	-6.3 ±	1.3
	103.75	47	35	10	1.00	0.71	-60.8 ±	11.8	-16.3 ±	16.1	-6.1 ±	1.8
573		45	35	10	1.00	0.54	-63.7 ±	9.1	-11.0 ±	16.3	-6.8 ±	1.6
		43	35	10	1.00	0.49	-64.0 ±	7.9	-9.5 ±	14.3	-7.6 ±	1.5
574		41	35	10	1.00	0.40	-59.5 ±	11.6	-0.8 ±	18.2	-7.9 ±	1.6
		39	35	8	1.00	0.31	-59.3 ±	12.9	0.9 ±	16.5	-9.3 ±	1.7
575		37	35	7	0.46	0.54	-25.1 ±	22.5	20.2 ±	25.4	-5.8 ±	1.7
		35	33	10	0.76	0.48	-30.4 ±	14.8	10.1 ±	17.1	-6.6 ±	1.9
576		33	33	9	0.67	0.33	-23.7 ±	11.9	5.1 ±	13.5	-5.7 ±	1.7
		31	23	8	0.65	0.22	-23.7 ±	10.2	4.9 ±	12.2	-5.5 ±	1.6
577	101.25	47	34	10	0.91	0.62	-51.9 ±	13.7	-18.2 ±	16.8	-4.4 ±	1.7
		45	35	9	0.91	0.54	-53.7 ±	12.0	-16.5 ±	16.4	-5.3 ±	1.8
578		43	35	10	1.00	0.71	-55.3 ±	10.0	-12.0 ±	14.6	-6.6 ±	1.7
		41	35	10	1.00	0.51	-51.2 ±	11.1	-4.8 ±	15.3	-6.9 ±	1.6
579		39	35	10	0.94	0.66	-40.7 ±	15.9	-1.8 ±	18.0	-6.3 ±	1.9
		37	35	10	0.89	0.71	-33.4 ±	17.6	0.6 ±	17.9	-6.0 ±	2.1
580		35	35	10	0.80	0.60	-26.3 ±	15.4	3.5 ±	15.8	-5.2 ±	2.0
		33	35	9	0.66	0.34	-18.2 ±	13.3	4.1 ±	14.2	-4.1 ±	1.9
581		31	35	6	0.80	0.43	-20.9 ±	11.1	2.4 ±	12.4	-4.3 ±	1.8
	98.75	47	35	10	1.00	0.74	-43.8 ±	14.2	-12.7 ±	17.1	-4.1 ±	1.7
582		45	35	10	1.00	0.80	-46.5 ±	13.0	-13.4 ±	16.3	-4.7 ±	1.8
		43	35	10	1.00	0.77	-45.8 ±	12.2	-8.0 ±	16.3	-5.6 ±	1.9
583		41	35	10	0.97	0.77	-40.3 ±	12.9	-4.7 ±	16.2	-5.3 ±	1.8
		39	35	10	0.97	0.71	-36.9 ±	14.4	-2.1 ±	15.5	-5.8 ±	2.0
		37	35	10	0.94	0.71	-29.6 ±	14.9	0.1 ±	14.4	-5.3 ±	2.0
584		35	35	10	0.91	0.66	-20.9 ±	13.4	1.9 ±	13.3	-4.3 ±	1.9
		33	35	10	0.94	0.69	-18.7 ±	11.7	2.9 ±	12.4	-3.9 ±	1.8
585		31	35	9	0.77	0.66	-14.7 ±	10.9	3.7 ±	11.6	-3.5 ±	1.7
		29	33	7	0.88	0.45	-20.1 ±	9.3	2.3 ±	10.3	-3.6 ±	1.7
		27	20	3	1.00	0.40	-25.3 ±	7.1	0.2 ±	7.4	-4.1 ±	1.5

Table 1 cont.:

	Long	Lat	Years	WS	Pr (years sigif.)		Accuracy (%)				Correction Factors (°C)	
					Model	Adjust	Model		Adjust		Maximum	Minimum
588	96.25	47	35	8	1.00	0.60	-41.3 ±	14.0	-9.1 ±	17.2	-4.1 ± 1.6	-1.1 ± 1.5
589		45	35	10	1.00	0.83	-39.2 ±	14.1	-7.8 ±	17.4	-3.9 ± 1.7	-1.3 ± 1.4
		43	35	10	0.97	0.66	-39.3 ±	13.7	-5.9 ±	17.5	-4.3 ± 1.9	-1.7 ± 1.4
590		41	35	10	0.94	0.80	-35.0 ±	14.1	-3.0 ±	16.4	-4.5 ± 1.9	-1.6 ± 1.2
		39	35	10	0.94	0.71	-30.0 ±	14.4	0.4 ±	15.2	-4.7 ± 1.8	-1.8 ± 1.2
591		37	35	10	0.94	0.74	-21.0 ±	15.1	3.2 ±	15.1	-4.0 ± 1.8	-1.5 ± 1.2
		35	35	8	0.89	0.69	-18.0 ±	12.6	4.6 ±	12.6	-3.9 ± 1.7	-1.8 ± 1.1
592		33	35	10	0.91	0.57	-17.9 ±	10.1	3.8 ±	10.9	-3.5 ± 1.6	-2.4 ± 0.9
		31	34	10	0.91	0.59	-19.2 ±	8.4	3.9 ±	9.2	-3.6 ± 1.5	-2.8 ± 0.8
593		29	32	10	0.91	0.47	-17.8 ±	7.4	1.1 ±	8.1	-3.7 ± 1.2	-0.5 ± 0.8
	93.75	47	29	9	0.97	0.52	-42.5 ±	15.0	3.6 ±	19.8	-5.3 ± 1.5	-1.5 ± 1.5
594		45	35	10	0.97	0.60	-41.8 ±	13.1	-1.5 ±	17.2	-4.6 ± 1.6	-2.3 ± 1.3
		43	34	10	0.94	0.65	-36.6 ±	14.1	-1.6 ±	17.3	-4.2 ± 1.7	-2.0 ± 1.3
595		41	35	10	0.91	0.57	-28.9 ±	16.1	1.3 ±	18.3	-4.0 ± 1.8	-1.6 ± 1.3
		39	35	7	0.91	0.71	-26.3 ±	15.2	4.4 ±	16.1	-4.3 ± 1.8	-2.1 ± 1.2
596		37	35	10	0.83	0.71	-17.9 ±	16.1	7.1 ±	16.2	-3.9 ± 1.8	-1.4 ± 1.2
		35	35	8	0.71	0.60	-14.5 ±	14.5	7.1 ±	14.3	-3.9 ± 1.8	-1.2 ± 1.1
597		33	35	10	0.83	0.63	-11.0 ±	10.6	6.7 ±	10.8	-3.1 ± 1.5	-1.3 ± 0.8
		31	35	5	0.91	0.63	-13.1 ±	8.7	6.0 ±	9.3	-2.8 ± 1.5	-2.1 ± 0.8
598	91.25	45	35	10	0.97	0.49	-42.0 ±	13.1	6.4 ±	18.3	-5.3 ± 1.4	-2.6 ± 1.3
		43	35	10	0.94	0.54	-35.3 ±	14.6	3.3 ±	18.4	-4.3 ± 1.5	-2.3 ± 1.2
599		41	35	21	1.00	0.80	-32.9 ±	14.3	2.9 ±	16.5	-4.4 ± 1.6	-2.5 ± 1.1
		39	35	10	0.97	0.60	-24.5 ±	15.5	4.5 ±	16.3	-4.0 ± 1.7	-1.7 ± 1.1
600		37	35	10	0.77	0.63	-18.2 ±	16.6	6.4 ±	16.3	-4.1 ± 1.9	-1.0 ± 1.1
		35	35	23	0.89	0.80	-17.8 ±	14.1	5.4 ±	13.8	-3.9 ± 1.7	-1.6 ± 1.1
601		33	35	10	0.89	0.63	-13.6 ±	11.2	5.5 ±	11.1	-3.3 ± 1.5	-1.4 ± 0.9
		31	35	10	0.89	0.71	-12.7 ±	9.3	6.7 ±	9.6	-2.7 ± 1.4	-2.3 ± 0.9
602	88.75	45	35	10	0.94	0.46	-35.0 ±	14.8	12.4 ±	19.8	-4.9 ± 1.3	-2.1 ± 1.3
		43	35	10	1.00	0.66	-35.7 ±	14.5	9.1 ±	18.6	-4.8 ± 1.4	-2.7 ± 1.2
		41	35	10	0.97	0.69	-33.9 ±	13.9	4.5 ±	16.6	-4.6 ± 1.5	-2.5 ± 1.0
603		39	35	10	0.89	0.66	-28.2 ±	15.0	5.4 ±	16.2	-4.5 ± 1.7	-2.2 ± 1.1
		37	35	9	0.89	0.80	-24.7 ±	15.0	6.5 ±	15.1	-4.4 ± 1.7	-2.3 ± 1.0
604		35	35	10	0.89	0.86	-20.8 ±	15.0	6.4 ±	14.6	-4.3 ± 1.7	-1.7 ± 1.0
		33	35	10	0.77	0.69	-14.2 ±	13.8	6.4 ±	13.2	-3.8 ± 1.7	-0.8 ± 1.0
605		31	35	10	0.89	0.71	-16.1 ±	10.2	7.3 ±	10.5	-3.1 ± 1.4	-2.5 ± 0.8
	86.25	41	35	10	0.97	0.71	-35.6 ±	13.3	7.4 ±	16.5	-4.8 ± 1.3	-2.7 ± 1.0
606		39	35	10	0.89	0.57	-31.2 ±	13.9	5.2 ±	15.9	-4.5 ± 1.5	-2.2 ± 1.0
		37	35	10	0.94	0.80	-29.3 ±	14.0	7.7 ±	14.5	-5.1 ± 1.5	-2.3 ± 0.9
607		35	35	10	0.89	0.77	-22.3 ±	16.5	7.3 ±	16.6	-4.5 ± 1.7	-1.2 ± 1.0
		33	35	10	0.89	0.63	-22.9 ±	14.4	6.8 ±	13.7	-4.6 ± 1.7	-1.9 ± 0.9
608		31	35	9	0.69	0.66	-11.5 ±	11.7	6.3 ±	11.4	-3.1 ± 1.5	-0.9 ± 0.9

Table 1 cont.:

					Pr (years sigif.)		Accuracy (%)				Correction Factors (°C)	
	Long	Lat	Years	WS	Model	Adjust	Model		Adjust		Maximum	Minimum
	83.75	45	35	10	0.97	0.71	-34.7 ±	14.0	17.5 ±	18.8	-5.3 ± 1.2	-1.9 ± 1.1
		43	35	10	1.00	0.69	-43.8 ±	12.5	12.9 ±	18.1	-5.7 ± 1.3	-4.0 ± 1.0
		41	35	10	0.97	0.66	-39.4 ±	12.0	6.3 ±	15.8	-5.1 ± 1.2	-2.6 ± 1.0
		39	35	10	1.00	0.60	-38.1 ±	10.7	5.7 ±	13.3	-5.1 ± 1.2	-2.8 ± 0.9
		37	35	10	1.00	0.74	-38.0 ±	11.7	8.5 ±	13.0	-6.2 ± 1.2	-2.2 ± 0.8
		35	35	10	0.77	0.49	-24.4 ±	14.8	13.4 ±	15.1	-5.1 ± 1.4	-1.2 ± 0.8
		33	35	10	0.94	0.57	-29.2 ±	12.6	7.5 ±	12.3	-5.3 ± 1.5	-2.6 ± 0.9
		31	35	10	0.91	0.66	-20.2 ±	11.4	5.5 ±	11.0	-4.3 ± 1.6	-1.7 ± 0.9
	81.25	41	35	9	1.00	0.49	-47.4 ±	9.5	10.1 ±	14.4	-5.9 ± 1.0	-3.7 ± 1.0
		39	35	9	1.00	0.43	-47.9 ±	8.4	6.9 ±	11.8	-6.8 ± 1.0	-2.8 ± 0.9
		37	35	10	0.91	0.37	-33.2 ±	11.7	11.6 ±	14.0	-5.1 ± 1.1	-1.6 ± 0.8
		35	35	9	0.97	0.37	-36.3 ±	10.2	9.9 ±	11.1	-6.3 ± 1.1	-2.4 ± 0.7
		33	35	10	1.00	0.63	-26.0 ±	10.1	6.1 ±	10.0	-5.2 ± 1.2	-1.6 ± 0.9
		27	35	20	0.97	0.74	-15.1 ±	4.8	0.8 ±	5.1	-3.2 ± 0.7	-0.4 ± 0.6
	78.75	41	35	10	1.00	0.37	-48.6 ±	9.5	8.8 ±	14.5	-6.0 ± 1.0	-2.3 ± 0.9
		39	35	10	0.97	0.23	-43.3 ±	9.3	10.1 ±	13.1	-5.9 ± 1.0	-2.5 ± 0.7
		37	35	10	1.00	0.57	-42.9 ±	7.9	8.6 ±	9.9	-6.6 ± 0.9	-3.0 ± 0.7
		35	35	10	1.00	0.54	-36.3 ±	8.0	7.5 ±	8.9	-6.1 ± 1.0	-2.7 ± 0.7
	76.25	41	35	6	0.94	0.17	-51.0 ±	9.4	7.8 ±	15.5	-5.7 ± 1.1	-3.8 ± 1.0
		39	35	10	1.00	0.71	-49.2 ±	7.1	9.6 ±	10.6	-6.5 ± 0.9	-5.0 ± 0.7
		37	34	6	0.79	0.06	-39.7 ±	8.7	3.7 ±	12.0	-8.8 ± 1.2	1.6 ± 1.2
	73.75	43	35	9	1.00	0.23	-61.1 ±	8.8	-3.2 ±	15.2	-6.0 ± 1.1	-3.4 ± 1.0
		41	35	10	1.00	0.37	-52.4 ±	8.8	-3.6 ±	13.4	-5.6 ± 1.1	-2.3 ± 0.8
	71.75	43	35	10	1.00	0.49	-61.3 ±	8.6	-5.7 ±	14.1	-6.4 ± 1.0	-2.1 ± 0.9
	68.75	45	35	5	1.00	0.31	-72.2 ±	7.0	-9.2 ±	14.3	-6.9 ± 1.1	-2.8 ± 1.2
	Ave		33.9	9.6	0.91	0.47	-41.4 ±	18.2	2.1 ±	8.2	-5.9 ± 1.9	-2.6 ± 1.6

Table 2: The summary statistics for all eleven AOGCMs examined in this study. Sample sizes are broken down into Grids (number of grids examined), Yrs/Grid (average number of years available per grid), WS/Yr (average number of weather stations per year within each grid). The proportion of years in which the simulated season length differed significantly from the historical observation (Years Signif.) are present both before the correction factors were included (Model) and after (Adjust). Average Correction factors (Corr. Factors) for both maximum (Tmax) and minimum (Tmin) temperatures are also reported. All values are averaged ($\pm$ stdev) across grids, with the exception of the Overall row which is the average across all models ($\pm$ stdev).

Model [†]	Sample Size:			Years Signif. (Pr):		Corr. Factors (°C):	
	Grids	Yrs/Grid	WS/Yr	Model	Adjust	Tmax	Tmin
GISS	20	38.5 $\pm$ 3.4	6.8 $\pm$ 3.6	0.84 $\pm$ 0.21	0.59 $\pm$ 0.30	-3.6 $\pm$ 3.0	2.2 $\pm$ 1.7
GFDL 2.0	150	33.6 $\pm$ 4.1	6.4 $\pm$ 2.7	0.91 $\pm$ 0.13	0.47 $\pm$ 0.25	-5.9 $\pm$ 1.4	-2.6 $\pm$ 1.0
GFDL 2.1	20	39.3 $\pm$ 3.4	6.1 $\pm$ 1.6	0.83 $\pm$ 0.65	0.49 $\pm$ 0.22	-5.3 $\pm$ 1.8	-1.9 $\pm$ 1.9
CSIRO	20	34.8 $\pm$ 9.0	6.0 $\pm$ 2.8	0.65 $\pm$ 0.33	0.30 $\pm$ 0.28	-4.2 $\pm$ 2.6	0.1 $\pm$ 2.3
IPSL	19	37.0 $\pm$ 3.4	5.5 $\pm$ 1.6	0.71 $\pm$ 0.24	0.41 $\pm$ 0.23	-4.6 $\pm$ 2.3	0.7 $\pm$ 2.0
MPI	20	36.9 $\pm$ 7.8	6.3 $\pm$ 2.7	0.67 $\pm$ 0.28	0.34 $\pm$ 0.27	-2.7 $\pm$ 2.6	2.1 $\pm$ 2.5
CCCMA	20	37.3 $\pm$ 6.8	5.5 $\pm$ 1.5	0.63 $\pm$ 0.26	0.41 $\pm$ 0.27	-1.7 $\pm$ 2.8	-1.4 $\pm$ 2.1
CNRM	20	38.1 $\pm$ 4.0	5.7 $\pm$ 1.6	0.65 $\pm$ 0.34	0.38 $\pm$ 0.30	-3.0 $\pm$ 1.7	0.4 $\pm$ 2.2
MRI	20	37.3 $\pm$ 6.9	6.2 $\pm$ 2.3	0.60 $\pm$ 0.27	0.43 $\pm$ 0.24	-1.5 $\pm$ 1.9	0.1 $\pm$ 1.5
IAP	20	37.1 $\pm$ 8.6	6.5 $\pm$ 1.9	0.59 $\pm$ 0.31	0.33 $\pm$ 0.28	-2.1 $\pm$ 2.4	0.8 $\pm$ 2.1
MIROC	19	35.5 $\pm$ 9.4	7.0 $\pm$ 3.5	0.68 $\pm$ 0.32	0.39 $\pm$ 0.26	-1.9 $\pm$ 3.1	3.6 $\pm$ 1.7
Overall		36.9 $\pm$ 1.7	6.2 $\pm$ 0.5	0.71 $\pm$ 0.11	0.41 $\pm$ 0.08	-3.3 $\pm$ 1.5	0.4 $\pm$ 1.9

[†] GISS- Goddard Institute for Space Studies; GFDL- Geophysical Fluid Dynamics Laboratory; CSIRO- Commonwealth Scientific and Industrial Research Organization; IPSL- Institut Pierre Simon Laplace; MPI- Max Planck Institute for Meteorology; CCCMA- Canadian Centre for Climate Modeling and Analysis; CNRM- Centre National de Recherches Meteorologique; MRI- Meteorological Research Institute; IAP- Institute of Atmospheric Physics; MIRCO- Model for Interdisciplinary Research on Climate.

655 FIGURES

656 Figure 1: The absolute percent difference between the simulated and historical season
657 lengths averaged across all years from 1961 to 2000 for the GFDL model cm2.0 using
658 (A) raw model output, and (B) adjusted model output.

659

660 Figure 2: The absolute difference between the simulated and historical daily maximum
661 (A) and minimum (B) temperatures for the GFDL model cm2.0 averaged across all days
662 from January 1, 1961 to December 31, 2000.

663

664 Figure 3: The percent differences between the simulated and historical season lengths
665 across all grids examined in the GFDL model cm2.0 using raw model output (dark bars)
666 and adjusted output (light bars).

667

668 Figure 4: The percent differences between the simulated and historical season lengths for
669 all 11 AOGCMs examined in this study using raw data (dark bars) and adjusted data
670 (light bars). Each estimate is the average across both grids and years ($\pm$ stderr).

671 Significance tests based on a one sample t-test for each model ($\alpha = 0.05$) are summarized
672 below the zero line for the raw output and above the zero line for the adjusted output.

Figure 1

Figure 2

Figure 3

Figure 4