Partitioning the variance between space and time

Fubao Sun,1 Michael L. Roderick,1,2 Graham D. Farquhar,1 Wee Ho Lim,1 Yongqiang Zhang,3 Neil Bennett,4 and Stephen H. Roxburgh5

Received 22 March 2010; revised 17 May 2010; accepted 20 May 2010; published 22 June 2010.

Here we decompose the space–time variance of near-surface air temperature using monthly observations for the global land surface (excluding Antarctica) from 1901–2000. To do that, we developed a new method for partitioning the total space–time variance, here called the grand variance, into separate spatial and temporal components. The temporal component is, in turn, further partitioned into the variance relating to different time periods and we use monthly data to decompose intra- and inter-annual components of the variance. The results show that the spatial and temporal components of the variance of near-surface air temperature have both, on average, decreased over time primarily because of reductions in the equator-to-pole (northern) temperature gradient, and because in cold regions, winter is generally warming faster than summer. We also found that in most regions, the inter-annual variance in near-surface air temperature has increased.

1. Introduction

Regional climates are usually defined using long term averages. Currently, there is much interest in changes in the mean, but equally, many hydrologic [Arora and Boer, 2006; Milly et al., 2008] and ecologic [Zimmermann et al., 2009] processes are known to be sensitive to changes in the extremes of climate variables. For that reason, the variance of near-surface air temperature has been studied over daily [Karl et al., 1995], monthly [Parker et al., 1994; Michaels et al., 1998] and annual [Seneviratne et al., 2006; Stouffer and Weathersald, 2007; Boer, 2009] time scales. Sometimes, spatial variance is used [e.g., Hernández-Deckers and von Storch, 2010], as opposed to the more familiar temporal variance.

The space and time components of the above-noted variances must be inter-related in some way and the aim of this paper is to establish those relationships. In section 2, we present a new theoretical approach that partitions the total space–time variance into separate spatial and temporal components. The temporal variance is further decomposed into intra- and inter-annual components. The new method builds on the traditional Analysis of Variance (ANOVA) technique [von Storch and Zwiers, 1999; Wilks, 2006] and extends that by deriving analytical relationships linking the various time and space scales. In section 3, we describe the temperature data. In section 4, we use the new approach to examine the climatology of, and changes over time in, the variance of the global near-surface air temperature over land from 1901 to 2000.

2. Mathematical Derivation

Historical databases are typically available at regular time intervals (e.g., monthly) and on regular spatial grids. Here we consider the more general case of irregularly distributed regions. As shown in Figure S1 (in the auxiliary material), a two-dimensional array is formed as

\[ Z = \begin{bmatrix} z_{ij} \end{bmatrix}_{m \times n} \] (1)

where \( i = 1, m \) months (time) and \( j = 1, n \) regions (space).

2.1. Partitioning Between Space and Time

The grand variance across space and time \( \sigma_g^2 \) and total sum of squares \( SS_g \) for the differences from the grand mean \( \mu_g = \sum_{j=1}^{m} \sum_{i=1}^{n} w_j z_{ij} / (m \times n) \) are,

\[ \sigma_g^2 = \frac{SS_g}{(m \times n) - 1} \] (2)

\[ SS_g = \sum_{j=1}^{m} \sum_{i=1}^{n} w_j (z_{ij} - \mu_g)^2 \] (3)

where for \( n \) regions having area \( A_i \), we define weights, \( w_j = A_j / (\sum_{j=1}^{n} A_j) \), as the proportional area of the \( j \)th region scaled as a proportion of the mean regional area. For the special case where each region has the same area, we have \( w_j = 1 \).

2.1.1. Space-First Formulation

The total sum of squares (equation (3)) can be rewritten by incorporating the spatial mean of each month \( \mu_s(i) = \sum_{j=1}^{n} w_j z_{ij} / n \) as follows,

\[ SS_g = \sum_{i=1}^{m} \sum_{j=1}^{n} w_j (z_{ij} - \mu_s(i) + \mu_g(i) - \mu_g)^2 \] (4)

Auxiliary materials are available in the HTML. doi:10.1029/2010GL043323.
This can be expanded and rearranged as follows,
\[
SS_g = \sum_{i=1}^{m} \sum_{j=1}^{n} w_j (z_{ij} - \mu_s(i))^2
+ 2 \sum_{i=1}^{m} (\mu_s(i) - \mu_g) \left( \sum_{j=1}^{n} w_j (z_{ij} - \mu_s(i)) \right)
+ \sum_{i=1}^{m} (\mu_s(i) - \mu_g)^2 \sum_{j=1}^{n} w_j
= \sum_{i=1}^{m} \sum_{j=1}^{n} w_j (z_{ij} - \mu_s(i))^2 + m \sum_{i=1}^{m} (\mu_s(i) - \mu_g)^2 \sum_{j=1}^{n} w_j
= (n-1) \sum_{i=1}^{m} \sigma_s^2(i) + n(m-1)\sigma_t^2(\mu)
\] (5)

[7] In classical experimental analysis using the ANOVA method, equation (5) with \( w_j = 1 \) is usually termed the total sum of squares and is partitioned into treatment and error components [von Storch and Zwiers, 1999; Wilks, 2006]. Here we express \( SS_g \) in terms of variances as
\[
SS_g = (n-1) \sum_{i=1}^{m} \sigma_s^2(i) + n(m-1)\sigma_t^2(\mu)
\] (6)
where \( \sigma_s^2(i) \equiv \sum_{j=1}^{n} w_j (z_{ij} - \mu_s(i))^2/(n-1) \) is the spatial variance of the \( i \)th month and can be displayed as a time series; and \( \sigma_t^2(\mu) \equiv \sum_{i=1}^{m} (\mu_s(i) - \mu_g)^2/(m-1) \) is the temporal variance of the spatial means, shown in Figure S1.

2.1.2. Time-First Formulation

[8] The above analysis was based on the inclusion of the spatial mean for each month (equation (4)) in the original expression for \( SS_g \). That is, we considered the spatial component first. We can also derive an alternative expression for \( SS_g \) by first including the temporal mean for the \( j \)th region given by \( \mu_t(j) = \sum_{i=1}^{n} z_{ij}/m \) in equation (3) as follows,
\[
SS_g = \sum_{j=1}^{n} w_j \left[ \sum_{i=1}^{m} (z_{ij} - \mu_t(j))^2 \right]
+ 2 \sum_{j=1}^{n} w_j (\mu_t(j) - \mu_g) \left( \sum_{i=1}^{m} (z_{ij} - \mu_t(j)) \right)
+ m \sum_{j=1}^{n} (\mu_t(j) - \mu_g)^2 \sum_{i=1}^{m} w_i
= (m-1) \sum_{j=1}^{n} \sigma_t^2(j) + m(n-1)\sigma_s^2(\mu)
\] (7)

that is again expanded and rearranged as follows,
\[
SS_g = \sum_{j=1}^{n} w_j \sum_{i=1}^{m} (z_{ij} - \mu_t(j))^2/(m-1)\sigma_t^2(j)
+ 2 \sum_{j=1}^{n} w_j (\mu_t(j) - \mu_g) \sum_{i=1}^{m} (z_{ij} - \mu_t(j))/(m-1)\sigma_t^2(j)
+ m \sum_{j=1}^{n} (\mu_t(j) - \mu_g)^2 \sum_{i=1}^{m} w_i/(m-1)\sigma_s^2(\mu)
= (m-1) \sum_{j=1}^{n} w_j\sigma_t^2(j) + m(n-1)\sigma_s^2(\mu)
\] (8)

where \( \sigma_t^2(j) \equiv \sum_{i=1}^{m} (z_{ij} - \mu_t(j))^2/(m-1) \) is the temporal variance of the \( j \)th region and can be displayed as a map; \( \sigma_s^2(\mu) \equiv \sum_{j=1}^{n} w_j (\mu_t(j) - \mu_g)^2/(n-1) \) is the spatial variance of the temporal means, as shown in Figure S1.

2.1.3. Partitioning Schemes

[9] By substituting equations (6) and (8) into equation (2), alternative relationships that partition the grand variance into spatial and temporal components with area weights are derived. For the space-first formulation we have,
\[
\sigma_s^2 = \frac{m(n-1)}{(m \times n) - 1} \sum_{j=1}^{n} \sigma_t^2(j) + \frac{m(n-1)}{(m \times n) - 1} \sigma_t^2(\mu)
\] (9)
and for the time-first formulation,
\[
\sigma_t^2 = \frac{n(m-1)}{(m \times n) - 1} \sum_{j=1}^{n} \sigma_s^2(j) + \frac{n(m-1)}{(m \times n) - 1} \sigma_s^2(\mu)
\] (10)

with definitions of \( \overline{\sigma}_s^2 \equiv \sum_{i=1}^{m} \sigma_s^2(i)/m \), the mean of the spatial variances of the \( m \) months and \( \overline{\sigma}_t^2 \equiv \sum_{j=1}^{n} \sigma_s^2(j)/n \), the mean of the temporal variances of the \( n \) regions. The coefficients appearing in both expressions, \( \frac{m(n-1)}{(m \times n) - 1} \) and \( \frac{n(m-1)}{(m \times n) - 1} \), vary from \( \frac{n-1}{n} \rightarrow 1 \) and \( \frac{m-1}{m} \rightarrow 1 \), respectively, e.g., both are 0.99 when \( m = 100 \) and \( n = 100 \) and larger than 0.9 when \( m \geq 10 \) and \( n \geq 10 \). Note that when \( m \gg 1 \) and \( n \gg 1 \), the sample variances calculated here will approach the respective population variances and the coefficients become unity (see Section S.1 in the supporting online material). In summary, when \( m \) and \( n \) are both larger than 10, the influence of the sample size on the partitioning is small.

[10] The grand variance is the sum of two terms. In the space-first formulation (equation (9)), the first term is an average of the spatial variances at each time step and the second is a single number that represents the temporal variance of the spatial means. In the time-first formulation (equation (10)), the first term is an average of the temporal variances for each region and the second is a single number that represents the spatial variance of the climatology. The choice between equations (9) and (10) depends on the purpose of the study. The space-first approach is useful for calculating globally aggregated statistics. The time-first approach is more useful for regional studies and has the added property, explained below, that the temporal component can be decomposed into the variance contribution from different time periods.

2.2. Partitioning the Temporal Variance in the Time-First Formulation

[11] For a given region, \( \sigma_t^2(j) \) in equation (10) is the temporal variance of the time series. This variance will depend on the level of temporal aggregation, e.g., months, years, etc. Similarly, the numerical values of the spatial variance terms will also depend on the degree of spatial aggregation. Here we assume the spatial sampling units are fixed and focus on formulating the problem in terms of temporal aggregation but note that the same underlying approach could be adopted for spatial aggregation.

[12] Here we aggregate from months to years by again reorganizing the data into a 2-dimensional array (Figure S1). Formally, the \( m \)-months time series of the \( j \)th region is:
\[
[z_{ik}]_{i=1}^{p \times q}, \text{where } k = 1, p \text{ months per year (} p = 12 \text{) and } l = 1, q \text{ year (} q = m/p \text{).}
\] Following the previous mathematical
derivation but without the weights, the temporal variance \( \sigma_t^2(j) \) is further partitioned into separate components as,

\[
\sigma_t^2(j) = \frac{q(p - 1)}{(p \times q) - 1} \sum_{l=1}^{q} \frac{\sigma_t^2(l)}{q} + \frac{p(q - 1)}{(p \times q) - 1} \sigma_t^2(\mu) \quad (11)
\]

\[
\sigma_t^2(j) = \frac{p(q - 1)}{(p \times q) - 1} \sum_{k=1}^{p} \frac{\sigma_t^2(k)}{p} + \frac{q(p - 1)}{(p \times q) - 1} \sigma_t^2(\mu) \quad (12)
\]

with definitions of \( \sigma_t^2 = \sum_{i=1}^{q} \sigma_t^2(i)/q \) the inter-annual mean of the intra-annual variances \( \sigma_t^2(i) \) for the q years and \( \sigma_t^2 = \sum_{p=1}^{p} \sigma_t^2(k)/p \) the intra-annual mean of the inter-annual variances \( \sigma_t^2(k) \) for the p months; \( \sigma_t^2(\mu) \) is the inter-annual variance for the yearly means; \( \sigma_t^2(\mu) \) is the intra-annual variance for the mean annual cycle (Figure S1).

2.3. Relation to Other Approaches

[13] Much previous climate related research has been performed using anomalies and that can be readily incorporated within the theory presented here. For example, Westra and Sharma [2010] calculated anomalies by subtracting the spatial (i.e., global) mean at each month. The variance of those anomalies is the spatial component in equation (9). The same applies where climate data is deseasonalized by subtracting the mean annual cycle [e.g., Parker et al., 1994; Held and Soden, 2006]. The resulting variance equals the inter-annual component of the temporal variance in equation (12). Similar examples can be found throughout the literature.

3. Data and Methods

[14] Here we use the theory to examine the variance in near-surface air temperature. We use the CRU TS2.1 gridded 0.5° × 0.5° monthly air temperature database for the global land surface from 1901 to 2000 [Mitchell and Jones, 2005]. Note that this database excludes Antarctica. We aggregated the database to a spatial resolution of 2.5° × 2.5°. The climatology has been calculated for the 100 year period. To examine changes in variance over time, we divided the 100-yr period into 10 × 10 year blocks and calculated the decadal grand variance and the components for each successive 10 year block. Trends were calculated using linear regression (ordinary least squares).

4. Results and Discussion

4.1. Twentieth Century Climatology, 1901–2000

[15] The 20th century climatology for the grand variance and components is described in Table 1. In the space-first formulation (equation (9)), most of the grand variance (246.44°C² = 0.9997 × 217.53 + 0.9992 × 29.00) is due to the spatial variance for each month (217.53°C²) (see Figure S2 in the auxiliary material for the time series from which this component is computed) with a much smaller contribution from the temporal variance of the spatial means (29.00°C²). The spatial component dominates. That arises because the spatial differences across the global land surface at a given time are generally much larger than the differences over time in a given region.

<table>
<thead>
<tr>
<th>Time-First (equation (10))</th>
<th>0.9992</th>
<th>75.11</th>
<th>0.9997</th>
<th>171.44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-annual Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( q(p - 1)/(p \times q - 1) )</td>
<td>( \sigma_t^2 )</td>
<td>( q(p - 1)/(p \times q - 1) )</td>
<td>( \sigma_t^2(\mu) )</td>
<td></td>
</tr>
<tr>
<td>0.9174</td>
<td>81.36</td>
<td>0.9908</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>Inter-annual Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p(q - 1)/(p \times q - 1) )</td>
<td>( \sigma_t^2 )</td>
<td>( p(q - 1)/(p \times q - 1) )</td>
<td>( \sigma_t^2(\mu) )</td>
<td></td>
</tr>
<tr>
<td>0.9908</td>
<td>2.63</td>
<td>0.9174</td>
<td>79.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Century Climatology (1901–2000) of the Grand Variance (°C²) of Monthly Air Temperature and the Underlying Variance Components⁴

The coefficients are based on the 100-year block: \( m = 1200 \) months (\( q = 100 \) years and \( p = 12 \) months per year) and \( n = 3399 \) grid-boxes. Partitioning the grand variance: \( \sigma_t^2 = 246.44 \). Partitioning the temporal variance: \( \sigma_t^2 = 75.11 \).
between years are much smaller than variations within a year.

4.2. Decadal Trends, 1901–2000

The decadal grand variance generally declined over time (Figures 1a and 2a) and was approximately complementary with the global mean near-surface air temperature (Figures 1b and S4). The grand variance decreased from 250.59°C² in the first decade to 240.56°C² in the last decade of the 20th century (Figure 1b). The average rate of change is −8.57°C² per century and the corresponding standard deviation $\sigma_g$ decreased at 0.27°C per century. (Note: The variance change can be related to the change in the standard deviation by $d\sigma_g^2 = 2\sigma_g^2 dt$.)

The overall trend has been broken down into the respective components (Table 2). (Trends of the corresponding population variances in Table S1.) The spatial components, in either the space– (equation (9) and Figure 1c) or time–first (equation (10) and Figure 1f) approaches generally dominate the overall decline in the grand variance (Figure 1a) while the temporal components also decline (Figures 1d and 1e). As in the climatology, the trends in the grand variances are conserved in the partitioned spatial and temporal components (Table 2). Of interest here is the trend in the temporal variances for individual regions (Figure 1e). This component is calculated as an average over space and the underlying grid-box-level data can be displayed as a map (Figure 2a) that is more readily interpreted. In general, the climatological variance in colder regions (e.g., Northern China, Russia, Greenland) is larger (Figure S3) and the trends in the temporal variance in those regions are also larger (Figure 2a) than in warm tropical regions. The large declines in the temporal variance in the above–noted cold regions are generally the result of winter temperatures increasing faster than summer temperatures [e.g., Stine et al., 2009]. Similarly, the decline in the spatial variance (Figures 1c and 1f) is due to reductions in the equator–to–pole (northern) temperature gradient [e.g., Karl et al., 1995].

The trend in the temporal variance, regardless of the scheme used (equations (11) or (12)), is dominated by the intra–annual component (Figures 1g and 1j) that decreased
Figure 2. Trends in the decadal temporal variance and the intra- and inter-annual components for 1901–2000. (a) Temporal variance, $\sigma^2_t(j)$ and components as follows: (b) mean of the intra-annual variances $\sigma^2_a$ (equation (11)), (c) inter-annual variance $\sigma^2_e(m)$ of the yearly means (equation (11)), (d) mean of the inter-annual variances $\sigma^2_e$ (equation (12)), (e) intra-annual variance $\sigma^2_a(m)$ of the mean annual cycle (equation (12)).

Table 2. Trends of the Decadal Grand Variance ($^\circ$C$^2$ per Century) of Monthly Air Temperature and the Underlying Variance Components for the Global Land Surface (1901–2000)$^a$

<table>
<thead>
<tr>
<th>Space-First (equation (9))</th>
<th>$m(n-1)((m+n)-1)$</th>
<th>$\frac{d\sigma^2_t}{dt}$</th>
<th>$n(m-1)((m+n)-1)$</th>
<th>$\frac{d\sigma^2_e}{dt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9997</td>
<td>−7.12</td>
<td>0.9917</td>
<td>−1.46</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time-First (equation (10))</th>
<th>$m(n-1)((m+n)-1)$</th>
<th>$\frac{d\sigma^2_{a}}{dt}$</th>
<th>$m(n-1)((m+n)-1)$</th>
<th>$\frac{d\sigma^2_{e}}{dt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9917</td>
<td>−3.58</td>
<td>0.9997</td>
<td>−5.01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intra-First (equation (11))</th>
<th>$q(p-1)((p+q)-1)$</th>
<th>$\frac{d\sigma^2_a}{dt}$</th>
<th>$q(p-1)((p+q)-1)$</th>
<th>$\frac{d\sigma^2_e}{dt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9244</td>
<td>−4.08</td>
<td>0.9076</td>
<td>0.21</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inter-First (equation (12))</th>
<th>$p(q-1)((p+q)-1)$</th>
<th>$\frac{d\sigma^2_a}{dt}$</th>
<th>$p(q-1)((p+q)-1)$</th>
<th>$\frac{d\sigma^2_e}{dt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9076</td>
<td>0.82</td>
<td>0.9244</td>
<td>−4.68</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Trends are calculated based on 10-year blocks: $m = 120$ ($q = 10$, $p = 12$) and $n = 3399$. Trend of the grand variance: $\frac{d\sigma^2_g}{dt} = -8.57$. Trend of the temporal variance: $\frac{d\sigma^2_t}{dt} = -3.58$. 
at a similar rate to the overall temporal variance (Figure 1e). This can be confirmed by the maps in Figures 2a, 2b, and 2c. Of interest here is that irrespective of the partitioning scheme adopted, the inter-annual component of the temporal variance is increasing, although the magnitude of the trends are relatively small (Figures 1h and 1l). The map (Figures 2c and 2d) shows that the inter-annual variance is increasing in most regions with decreases prominent in only a few regions, e.g., US, eastern Australia, northern China and the Himalayan plateau. That is consistent with earlier findings of a generally small increase in the inter-annual component of the temporal variance in most land regions [Parker et al., 1994]. The anomalous decrease in the US reported here is also consistent with earlier findings [Karl et al., 1995].

In previous regional studies of changes in the temporal variance of air temperature, the decrease at sub-annual time scales (e.g., daily, monthly) in much of the northern hemisphere [Karl et al., 1995; Michaels et al., 1998] along with an increase over longer periods (e.g., inter-annual) throughout much of Europe [Seneviratne et al., 2006] have both been attributed to increasing greenhouse gas concentrations. That can potentially explain the globally averaged trends (Figures 1g–1j). However, that explanation cannot be universal because the sign (+/−) of the trend in the inter- and intra-annual components of the temporal variance is not always opposite. For example, throughout much of the tropics both the inter- and intra-annual variances have increased (Figure 2).

5. Conclusions

Based on the long established concept of ANOVA, we have developed a new method of partitioning variance between spatial and temporal components. The grand (or total space-time) variance can be split using two different partitioning schemes that we call the space- and time-first approaches. In the space-first scheme, the grand variance is equal to the sum of two components: the average of the spatial variances at each time step and the temporal variance of the spatial means. This scheme is useful for calculating globally aggregated statistics. In the time-first scheme, the grand variance is equal to the sum of two components: the average of the temporal variances from all regions and the spatial variance of the temporal means. This scheme is more useful for regional studies. A further advantage of the time-first scheme is that the resulting variance of the temporal component can be further partitioned into components relating to different time periods. Here we used monthly data to decompose the intra- and inter-annual variance.

Using observations of monthly temperature over the global land surface from 1901 to 2000 we calculated the climatology and examined trends in the variance components over time. The climatology showed the expected result: the grand variance is dominated by the spatial component since the differences in air temperature between equator-to-pole are generally larger than differences between seasons in a given region. In turn, the temporal component is dominated by the intra-annual component, i.e., in a given region, the temperature difference within a year is larger than the difference between years.

The trend analysis showed declines in the decadal grand variance from 1901 to 2000. Using the time-first approach, the total decline was more or less equally partitioned between the spatial and temporal components of the variance (Table 2). The decline in the temporal component is generally due to the northern pole warming faster than the equator. The decline in the temporal component arises from a similar phenomenon: winter temperatures are increasing faster than summer temperatures, particularly in cold regions (e.g., Northern China, Russia, Greenland) that dominate the overall temporal trends (Figure 2a). However, in many parts of the tropics, the temporal component of variance has increased slightly (Figure 2a). The temporal component was further partitioned into inter- and intra-annual components with the latter dominating trends. Interestingly, changes in the inter-annual variance (Figures 2c and 2d), whilst of small magnitude, are generally increasing in most regions and should be a high priority for further investigation.

Acknowledgments

This work was supported by the Australian Research Council (DP0879763). We thank two reviewers for comments that improved the manuscript.

References