Statistical Analysis of Compressed Climate Data

Joseph Nardi
Noah Feldman
Andrew Poppick
Allison Baker
Dorit Hammerling
ERRATA

After publishing this technote, we discovered that we made a mistake in our python API when applying the two compressors. In particular, we inadvertently transposed the x and y dimensions of the climate data (which we blame on our frequent use of Fortran). This mistake caused misalignment in ZFP as well as compression artifacts in SZ, as we essentially introduced artificial discontinuities by transposing the data, which also negatively affected the compression ratios. We note that this mistake does not affect the statistical methods proposed, nor does it invalidate most of the conclusions of the paper.
The Technical Notes series provides an outlet for a variety of NCAR Manuscripts that contribute in specialized ways to the body of scientific knowledge but that are not yet at a point of a formal journal, monograph or book publication. Reports in this series are issued by the NCAR scientific divisions, serviced by OpenSky and operated through the NCAR Library. Designation symbols for the series include:

**EDD – Engineering, Design, or Development Reports**
Equipment descriptions, test results, instrumentation, and operating and maintenance manuals.

**IA – Instructional Aids**
Instruction manuals, bibliographies, film supplements, and other research or instructional aids.

**PPR – Program Progress Reports**
Field program reports, interim and working reports, survey reports, and plans for experiments.

**PROC – Proceedings**
Documentation or symposia, colloquia, conferences, workshops, and lectures. (Distribution maybe limited to attendees).

**STR – Scientific and Technical Reports**
Data compilations, theoretical and numerical investigations, and experimental results.

The National Center for Atmospheric Research (NCAR) is operated by the nonprofit University Corporation for Atmospheric Research (UCAR) under the sponsorship of the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

National Center for Atmospheric Research
P. O. Box 3000
Boulder, Colorado 80307-3000
Statistical Analysis of Compressed Climate Data

Joseph Nardi, Noah Feldman, Andrew Poppick
Department of Mathematics and Statistics,
Carleton College, Northfield, MN
Allison Baker
Computational and Information Systems Laboratory,
I/O Workflow Application Group,
National Center for Atmospheric Research, Boulder, CO
Dorit Hammerling
Computational and Information Systems Laboratory,
Analytics and Integrative Machine Learning,
National Center for Atmospheric Research, Boulder, CO
Statistical Analysis of Compressed Climate Data

Joseph Nardi¹, Noah Feldman¹, Andrew Poppick*¹, Allison Baker², and Dorit Hammerling²

¹Carleton College, Northfield, MN
²National Center for Atmospheric Research, Boulder, CO

August 22, 2018

Abstract

The data storage burden resulting from large climate model experiments only continues to grow. Lossy data compression methods are required to alleviate this burden, but lossy methods introduce the possibility that key climate variable fields could be altered to the point of affecting scientific conclusions. It is therefore important to develop a detailed understanding of how compressed climate model output differs from the original for different compression algorithms and compression rates. In this work, we evaluate the effects of two leading compression algorithms, SZ and ZFP, on daily average and monthly maximum temperature data, and daily average precipitation rate data, from a historical run of CESM1 CAM5.2. While both algorithms show promising fidelity with the original model output, detectable artifacts are introduced even at relatively low error tolerances. Examples for temperature data include biases in temperature gradient fields, temporal autocorrelation, and seasonal cycles; precipitation data show, for example, biases in the number of rainy days. We highlight the need for evaluation methods that are sensitive to errors at different spatiotemporal scales and specific to the particular climate variable of interest.

Keywords: data compression, earth system models, climate models, CESM, climate variability

*apoppick@carleton.edu
Contents

1 Introduction 6

2 Experiment Details 6
  2.1 CESM Data 6
  2.2 Compression Methods 7
  2.3 Data for Analysis 8

3 Exploratory Analysis 8
  3.1 Daily Temperature (TS) 9
  3.2 Monthly Maximum Temperature (TSMX) 11
  3.3 Daily Average Rainfall Rate (PRECT) 13

4 Gridcell-level Analysis 16
  4.1 Daily Temperature (TS) 17
    4.1.1 Mean Errors 19
    4.1.2 Contrast Variances 25
    4.1.3 Standard Deviations 28
    4.1.4 Biases in Seasonality 30
    4.1.5 Temporal Correlations 33
    4.1.6 Time of Maximum and Minimum Errors 36
  4.2 Monthly Maximum Temperature (TSMX) 39
    4.2.1 Mean Errors 41
    4.2.2 Contrast Variances 44
    4.2.3 Standard Deviations 46
    4.2.4 Biases in Seasonality 48
    4.2.5 Temporal Correlations 50
    4.2.6 Time of Maximum and Minimum Errors 53
  4.3 Daily Average Rainfall Rate (PRECT) 56

5 Discussion 60
1 Introduction

Earth system models, such as the popular Community Earth System Model (CESM) (Hurrell et al., 2013), have been generating increasingly larger (and often unwieldy) volumes of data in recent years as the models take advantage of impressive advances in supercomputing resources. While data compression as a tool to reduce the storage burden for climate data has been the subject of a number of studies in recent years (e.g., see Woodring et al. (2011); Hübbe et al. (2013); Baker et al. (2014); Kuhn et al. (2016); Zender (2016); Baker et al. (2016, 2017)), more analysis needs to be done to satisfy concerns in the climate community regarding data compression-induced artifacts (Baker et al., 2016).

Climate data, such as that from CESM (our focus in this work), is typically stored by time slice, meaning that variables are stored by spatial fields. Even for the so-called CESM “time-series” files, the multidimensional data arrays are laid out such that time is the outer array dimension, meaning that the most efficient access is by spatial slices. For this reason, compression algorithms applied to CESM data thus far (e.g., Baker et al. (2014, 2016, 2017)) have been applied to spatial fields, and any dependence on time has been ignored. It is therefore natural to investigate whether compression applied to spatial fields independently introduces any artifacts that change either the temporal or spatial characteristics of the model output. These questions are worth investigating in detail: examining climate characteristics over time is a critical component to most climate data analyses, as is examining coherent spatial features of climate variables.

Most compression algorithm development teams evaluate their methods with simple metrics, such as root mean squared error or maximum pointwise error, which are quite appropriate for many visualization applications, but less so for floating-point data from climate simulations. Because of our familiarity with climate data, our aim is to thoroughly explore the effects of data compression on climate model data so that we can work together with the compression algorithm development teams to address specific issues that arise. (Note that the work in Lindstrom (2017) contains an interesting empirical analysis of error bounds from multiple compressors, including two that we study in detail in this work.) Ultimately, to mitigate the burden of the increasing climate data storage demands, we need suitable compression methods that reduce data volume while preserving information such that scientific conclusions are not (substantively) affected.

2 Experiment Details

In this section, we describe the CESM data used in our study, the two different compression algorithms chosen, and how the compressors were applied to the data.

2.1 CESM Data

We use CESM data from the publicly available CESM Large Ensemble (CESM-LE) project (Kay et al., 2015). The project includes a set of 40 ensemble runs for the period 1920-2100. All simulations use the fully coupled one degree latitude-longitude (lat-lon) version of CESM-CAM5 (Community Atmosphere Model version 5). In particular, we experiment with the first 86 years (1920-2005), which is the historical forcing period, for ensemble member 30. Note that when the floating-point data are written to file, they are truncated from double precision (64 bits) to single precision (32 bits).

As with the study in Baker et al. (2016), we focus only on the atmospheric model output in CESM-LE. Each CAM variable (159 total) is stored in its own time-series file, and variable output is at temporal frequencies of monthly, daily, or 6-hourly (though 6-hourly data is limited to three short time periods). Daily time-series data consists of 31,390 time slices, and monthly data consists of 1,032 slices. The approximate one degree lat-lon grid corresponds to 192 × 288 grid points per vertical level, and 3D variables consist of 30 vertical levels. The 192 × 288 lat-lon grid layout is such that the 192 rows indicate the latitude range from 90° to -90° and the 288 columns indicate the longitude from 0° to 360°. (Note
that most rectangular maps of the Earth’s surface - including those in this study - are depicted with 0° longitude at the center of the grid, which does not reflect how the CESM grid is stored.) We study the following 2D variables (time-series file size is listed in parentheses):

- TS: Daily average surface temperature, in °K (3.8G).
- TSMX: Monthly maximum surface temperature, in °K (126M).
- PRECT: Daily average precipitation rate, in m/s (5.4G).

These variables were chosen as temperature is considered “easy” to compress because of its smoothness and relatively small data range. Precipitation variables, on the other hand, change more abruptly and have vastly larger ranges. (Note that we will look at maximum precipitation in future work, as it is not available from CESM-LENS.)

### 2.2 Compression Methods

Compression algorithms are either lossless or lossy. The former type preserves the original data exactly when the compressed data is reconstructed (i.e., decompressed). Lossy methods, on the other hand, only approximate the original data upon reconstruction, where the quality of the approximation is typically controlled by algorithm-dependent parameters. Our interest is in lossy methods, as they offer the most meaningful data reduction for floating-point data. While many lossy methods have been advocated for use on floating-point data in recent years, we experiment with two of the more popular methods: ZFP (Lindstrom, 2014) and sz (Di and Cappello, 2016; Tao et al., 2017).

The ZFP compressor is a lossy approach that was designed to facilitate random data access, but can also be used for error-bounded sequential compression, depending upon the specified input parameters. The compressor partitions \( d \)-dimensional arrays into blocks of \( 4^d \) values and compresses each block independently via a floating-point representation with a single common exponent per block, an orthogonal block transform, and embedded encoding. In our experiments, we use version 0.5.3 of ZFP in fixed-accuracy mode. We note that specifying an absolute error tolerance of 0 indicates that the compressor should achieve lossless (if possible) or near lossless compression. While we limit ourselves to fixed accuracy mode in this study, the ZFP compressor can also be used in a fixed rate mode (required for random access) or fixed precision mode. See Lindstrom (2014) for more details on ZFP.

The sz compressor is a predictive lossy method that uses adaptive error-controlled quantization and variable-length encoding to optimize compression. The most recent version in Tao et al. (2017) uses a multilayer prediction model, which we use in fixed-accuracy mode. Given an absolute error tolerance, \( \epsilon \), sz defines a size \( 2\epsilon \) interval that is centered on the predicted value, as well as \( 2m - 2 \) additional adjacent intervals. The interval that the actual value falls into determines the identifying index (size \( m \) bits). If the actual value does not lie in any of the intervals, it is given an index that flags it as being unpredictable and uses an alternative coding scheme (with longer codes). All the predicted values are then subjected to Huffman encoding and compressed further with gzip. In our experiments, we use version 1.4.13 of sz in fixed-accuracy mode (errorBoundMode = ABS), with single layer prediction (layers = 1), with optimized auto-selection of quantization intervals (quantization_intervals = 0), with the maximum number of intervals specified (max_quant_intervals = 65536, which corresponds to \( m = 16 \)), with the mode set to default (szMode = SZ_DEFAULT_COMPRESSION), and to no offset (offset = 0). Note that sz offers several error bound modes in addition to fixed accuracy: fixed relative error (normalized by the global data range), fixed PSNR (peak signal-to-noise ratio), and fixed point-wise relative error bound. See Di and Cappello (2016) and Tao et al. (2017) for more details on sz.

To simplify the comparison of sz and ZFP, we chose to use both compressors in their fixed absolute error modes as those modes are equivalent. ZFP’s fixed precision mode is similar to sz’s fixed relative error bound, but they differ by a scaling factor. Note that sz does not have a fixed rate option, nor does ZFP have a fixed PSNR or point-wise error mode.
<table>
<thead>
<tr>
<th>Absolute error</th>
<th>Effective Absolute Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1e-1</td>
<td>6.25e-2</td>
</tr>
<tr>
<td>1e-2</td>
<td>7.82e-3</td>
</tr>
<tr>
<td>1e-3</td>
<td>9.77e-4</td>
</tr>
<tr>
<td>1e-4</td>
<td>6.10e-5</td>
</tr>
<tr>
<td>1e-5</td>
<td>7.63e-6</td>
</tr>
<tr>
<td>1e-6</td>
<td>9.54e-7</td>
</tr>
<tr>
<td>1e-7</td>
<td>5.96e-8</td>
</tr>
<tr>
<td>1e-8</td>
<td>7.45e-9</td>
</tr>
<tr>
<td>1e-9</td>
<td>9.31e-10</td>
</tr>
<tr>
<td>1e-10</td>
<td>5.82e-11</td>
</tr>
<tr>
<td>1e-11</td>
<td>7.28e-12</td>
</tr>
</tbody>
</table>

Table 1: Effective error tolerances used by ZFP for a given specified absolute error tolerance in our study.

2.3 Data for Analysis

To create the compressed data for analysis, we applied both ZFP and SZ to the previously described time-series data for TS, TSMX, and PRECT. Note that for a single time-series file, we call the compressor on each time slice sequentially in sequence over time. For the 2D variables in this study, this approach means that each lat-lon slice is compressed independently. For a 3D variable, each lat-lon-lev slice is compressed independently. Note that both ZFP and SZ can typically achieve more data reduction (i.e., a smaller compression ratio) when applied to larger data fields. For SZ, the overhead associated with storing the Huffman tree is better mitigated for larger data sets, and for ZFP, higher dimensional data is also better suited to hiding the overhead due to the compression (e.g., using $4^3$ block size for 3D data instead of $4^2$ blocks for 2D data). In the case of the CESM time-series files, for 2D variables in particular, we could consider compressing multiple time slices at once so as to operate on 3D data. However, this choice introduces further complexity due to memory constraints. While we could certainly do 10 or even 100 time slices together on this one degree grid, it is unlikely that we could compress an entire time-series file worth of data with one call to the compressor for high resolutions, long durations, or frequent temporal output. Therefore some dividing between time slices would be necessary, and to simplify our analysis here, we compress each time slice independently.

Both compressors were applied with multiple absolute error tolerances, including large tolerances such as 1.0 and 0.5. As far as the small tolerances, the characteristics of the PRECT data led us to use much smaller tolerances (e.g., $10^{-11}$) than needed for TS and TSMX ($10^{-5}$). We also note that absolute error tolerances for ZFP are converted to the nearest power of 2 (that is smaller than the specified value), resulting in an effective tolerance that differs from the user-specified tolerance (see Table 1 for conversions).

The data that we analyze resulted from applying compression, followed by reconstruction, to each variable file for multiple absolute error tolerances for each compressor (via a Python script that called the compressors’ APIs). We then created a new NetCDF file containing the reconstructed data. We additionally created a corresponding second NetCDF file containing the difference at each gridcell between the original and reconstructed data for each compressor and error tolerance.

3 Exploratory Analysis

In this section, we give a brief exploration of the original data fields that have been subjected to compression, along with some global summaries of the compression errors by algorithm and error tolerance.
3.1 Daily Temperature (TS)

Figure 1 shows minimum and maximum TS by gridcell, along with the gridcell mean values and their standard deviations. As might be expected, minimum TS values tend to be colder towards the poles and over land; maximum TS values tend to be warmer over land locations and towards the equator. Overall, TS values tend to be more variable in time at high latitudes and over land, as can be seen in the gridcell standard deviations. (Note that this measure of temporal variability includes both regular seasonal variation as well as other sources of atmospheric variability.)

Table 2 summarizes global characteristics of the quality of compression for both sz and ZFP at all error tolerances considered. We show mean errors (to capture any systematic biases in the sign of the compression error), along with mean absolute errors (MAEs) and root mean square errors (RMSEs) (where the latter is more sensitive to extreme errors than the former). We also include the achieved compression ratio (CR), which we define as the ratio of the size of the compressed variable data to that of the original variable data.

Overall, the quality of compression scales with the error tolerance. The sz output becomes lossless at error tolerance $10^{-4}$ and the ZFP at error tolerance $10^{-5}$. (While ZFP is not lossless at error tolerance $10^{-4}$, errors are very frequently zero and only take five unique values; see Table 3.) For a fixed error tolerance, ZFP shows substantially smaller MAEs and RMSEs than sz, but larger compression ratios; however, ZFP shows larger magnitude mean errors themselves (see column 2 of Table 2), indicating that ZFP errors have a more consistent sign than sz errors and therefore result in a stronger systematic mean
Table 2: Summary statistics of TS errors. The compression algorithm and error tolerance are given in the first column. The metrics are then, from left to right, the mean error, mean absolute error, root mean square error, and compression ratio.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Error</th>
<th>MAE</th>
<th>RMSE</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>sz1.0</td>
<td>-1.45e-03</td>
<td>1.35e-01</td>
<td>3.68e-01</td>
<td>.06</td>
</tr>
<tr>
<td>sz0.5</td>
<td>-7.18e-04</td>
<td>6.77e-02</td>
<td>1.84e-01</td>
<td>.09</td>
</tr>
<tr>
<td>sz1e-1</td>
<td>-1.48e-06</td>
<td>1.69e-02</td>
<td>4.60e-02</td>
<td>.16</td>
</tr>
<tr>
<td>sz1e-2</td>
<td>4.47e-06</td>
<td>2.11e-03</td>
<td>4.06e-03</td>
<td>.29</td>
</tr>
<tr>
<td>sz1e-3</td>
<td>2.72e-06</td>
<td>1.38e-04</td>
<td>3.66e-04</td>
<td>.60</td>
</tr>
<tr>
<td>sz1e-4</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.78</td>
</tr>
<tr>
<td>sz1e-5</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.78</td>
</tr>
<tr>
<td>sz1e-6</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.78</td>
</tr>
<tr>
<td>zfp1.0</td>
<td>7.71e-03</td>
<td>3.42e-02</td>
<td>6.75e-02</td>
<td>.23</td>
</tr>
<tr>
<td>zfp0.5</td>
<td>-3.86e-03</td>
<td>1.71e-02</td>
<td>3.43e-02</td>
<td>.26</td>
</tr>
<tr>
<td>zfp1e-1</td>
<td>4.82e-04</td>
<td>2.16e-03</td>
<td>4.38e-03</td>
<td>.36</td>
</tr>
<tr>
<td>zfp1e-2</td>
<td>-6.01e-05</td>
<td>2.69e-04</td>
<td>5.48e-04</td>
<td>.45</td>
</tr>
<tr>
<td>zfp1e-3</td>
<td>7.76e-06</td>
<td>3.35e-05</td>
<td>6.93e-05</td>
<td>.54</td>
</tr>
<tr>
<td>zfp1e-4</td>
<td>9.17e-08</td>
<td>5.49e-07</td>
<td>3.20e-06</td>
<td>.67</td>
</tr>
<tr>
<td>zfp1e-5</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.76</td>
</tr>
<tr>
<td>zfp1e-6</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.77</td>
</tr>
<tr>
<td>zfp0</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.76</td>
</tr>
</tbody>
</table>

Table 3: Summary of unique values of ZFP errors at error tolerance 10^{-4}. Left, error value. Right, the proportion taking that value (aggregating over both days and gridcells).

<table>
<thead>
<tr>
<th>Value</th>
<th>Prop. Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.31 × 10^{-4}</td>
<td>6.9 × 10^{-5}</td>
</tr>
<tr>
<td>-0.15 × 10^{-4}</td>
<td>0.015</td>
</tr>
<tr>
<td>0</td>
<td>0.97</td>
</tr>
<tr>
<td>0.15 × 10^{-4}</td>
<td>0.013</td>
</tr>
<tr>
<td>0.31 × 10^{-4}</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

bias. The average ZFP errors are positive at error tolerances of 1.0, 0.1, 10^{-3}, and 10^{-4}; they are negative for tolerances 0.5 and 0.01. The overall error tolerances are much smaller than the scale of variation in the temperatures themselves. For example, the globally pooled TS standard deviation is about 8.6°K, whereas the maximum global RMSE achieved is about 0.37°K (for sz at error tolerance 1.0). This implies that pointwise comparisons of the original and compressed output will appear very similar relative to the scale of variability in the data (i.e., the correlation between the original output and compressed output is nearly one when averaged globally).

Since it is possible that mean errors may vary seasonally, Figure 2 shows the mean absolute errors by day of the year for each tolerance level. Each value represents the mean error across the entire globe, averaged over all instances of the given day in the 86 years of the data set. When averaged globally, there do not appear to be any strong biases in seasonality introduced by compression.
Figure 2: Mean absolute error by day of the year for both zfp and sz for a range of tolerances. The legend contains the compression algorithm and the tolerance that the line represents. There do not appear to be any strong biases in seasonality introduced by compression.

3.2 Monthly Maximum Temperature (TSMX)

We now repeat the analyses shown in Section 3.1 but for monthly maximum temperatures (TSMX).

Figure 3 shows the same summary statistics for TSMX as were shown for TS in Figure 1. The characteristics of TSMX are qualitatively similar to those of TS, with some important distinctions. For example, gridcell maximum TSMX values show a stronger contrast between land and ocean compared to TS. Additionally, areas where there is typically sea ice at some times of the year show comparatively less variability in monthly TSMX than in daily TS.

Likewise, Table 4 repeats the global error summaries shown in Table 2 except for TSMX. They are overall fairly similar to the error summaries discussed for TS, never deviating by more than a factor of two. As with TS, sz is lossless at error tolerance $10^{-4}$ and zfp at error tolerance $10^{-5}$; zfp errors at error tolerance $10^{-4}$ take the same five unique values shown for TS errors in Table 3, with similar percent occurrences (not shown). When not lossless, TSMX errors are generally somewhat larger than those for TS, especially as measured by MAEs; the RMSEs are more similar. The fact that the RMSEs are more similar than the MAEs when comparing TSMX and TS errors would seem to imply that extreme errors are similar between the two fields whereas smaller errors tend to be larger for TSMX than for TS. As with TS errors, there does not appear to be substantial seasonality in the global TSMX errors (Figure 4).
Figure 3: Exploratory analysis of the original TSMX data. Top left, minimum TSMX; top right, maximum TSMX; bottom left, mean TSMX; bottom right, log_{10}(standard deviation). Compare to Figure 1.

<table>
<thead>
<tr>
<th></th>
<th>Mean Error</th>
<th>MAE</th>
<th>RMSE</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>sz1.0</td>
<td>-1.53e-03</td>
<td>4.99e-01</td>
<td>5.76e-01</td>
<td>.06</td>
</tr>
<tr>
<td>sz0.5</td>
<td>-6.36e-04</td>
<td>2.49e-01</td>
<td>2.88e-01</td>
<td>.09</td>
</tr>
<tr>
<td>sz1e-1</td>
<td>-1.83e-05</td>
<td>4.99e-02</td>
<td>5.76e-02</td>
<td>.15</td>
</tr>
<tr>
<td>sz1e-2</td>
<td>4.51e-06</td>
<td>4.99e-03</td>
<td>5.76e-03</td>
<td>.30</td>
</tr>
<tr>
<td>sz1e-3</td>
<td>1.05e-06</td>
<td>3.24e-04</td>
<td>4.63e-04</td>
<td>.59</td>
</tr>
<tr>
<td>sz1e-4</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.77</td>
</tr>
<tr>
<td>sz1e-5</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.77</td>
</tr>
<tr>
<td>zfp1.0</td>
<td>9.24e-03</td>
<td>6.41e-02</td>
<td>8.16e-02</td>
<td>.23</td>
</tr>
<tr>
<td>zfp0.5</td>
<td>-4.63e-03</td>
<td>3.29e-02</td>
<td>4.16e-02</td>
<td>.26</td>
</tr>
<tr>
<td>zfp1e-1</td>
<td>5.80e-04</td>
<td>4.23e-03</td>
<td>5.35e-03</td>
<td>.35</td>
</tr>
<tr>
<td>zfp1e-2</td>
<td>-7.21e-05</td>
<td>5.30e-04</td>
<td>6.71e-04</td>
<td>.45</td>
</tr>
<tr>
<td>zfp1e-3</td>
<td>9.29e-06</td>
<td>6.62e-05</td>
<td>8.44e-05</td>
<td>.54</td>
</tr>
<tr>
<td>zfp1e-4</td>
<td>1.34e-07</td>
<td>3.75e-07</td>
<td>2.79e-06</td>
<td>.66</td>
</tr>
<tr>
<td>zfp1e-5</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.76</td>
</tr>
<tr>
<td>zfp0</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>0.00e+00</td>
<td>.76</td>
</tr>
</tbody>
</table>

Table 4: Summary statistics of TSMX errors. The compression algorithm and error tolerance are given in the first column. The metrics are then, from left to right, the mean error, mean absolute error, root mean square error, and compression ratio. Compare to Table 2.
Figure 4: Mean absolute error by month for both ZFP and SZ for a range of tolerances. The legend contains the compression algorithm and the tolerance that the line represents. Compare to Figure 2.

3.3 Daily Average Rainfall Rate (PRECT)

We now turn to daily average rainfall (PRECT). Figure 5 shows minimum and maximum PRECT values by gridcell, along with the gridcell mean values and the probability that PRECT is positive across days and gridcells. Minimum rainfall values are typically very small, but are positive in about 34% of gridcell locations. (Note that negative PRECT values appear at about 30% of gridcell locations, but the minimum PRECT value observed in the original dataset is \(-1.1 \times 10^{-19}\).) Maximum rainfall tends to be larger in equatorial ocean locations. Except in desert regions, the probability of positive rainfall is very close to one.

Table 5 gives a global summary of PRECT errors. We show the mean rainfall produced in the compressed output, the global MAEs and RMSEs, and the global probability of positive and of negative rainfall (i.e., the proportion of days with positive or negative rainfall across all days and gridcells). Even at the smallest error tolerances, MAEs and RMSEs are of the same order of magnitude as the mean PRECT values themselves. (This is at least in part because PRECT is recorded in units m/s so values tend to be very small.) Globally, for an error tolerance of \(10^{-4}\) or larger, ZFP sets PRECT to zero on every day at every location of the globe; for smaller error tolerances, the compressed output does preserve some rainy days but with a negative bias in the number of rainy days. By contrast, the SZ algorithm produces an abundance of rainy days for error tolerances larger than or equal to \(10^{-6}\). (For these error tolerances, the probability of rain is near one at every location on the globe, and the few days where rainfall is zero are the same days at every location on the globe.) For smaller error tolerances, both produce substantially more negative PRECT values than are present in the original output.

Figure 6 shows the percentage of gridcells with positive rainfall by day of the year, aggregating globally and across years, as well as the log-odds ratios comparing these percentages in the compressed output to the original output. The odds of rain are defined as

\[
\omega = \frac{p}{1-p},
\]

(1)

where \(p\) is the probability of rain on that day (again aggregating globally and across years). Denoting the odds of rain under the compressed output as \(\tilde{\omega}\), the odds ratio, comparing compressed to original, is then \(OR = \tilde{\omega}/\omega\). Note that for the purposes of calculating odds and odds ratios (but not for displaying the probabilities themselves), the probability \(p\) is calculated adding one rainy day and one dry day to the
Figure 5: Exploratory analysis of the original PRECT data. Top left, log$_{10}$(minimum PRECT + 1e-18) (we add 10$^{-18}$ because the smallest PRECT value is about $-1.1 \times 10^{-19}$; note that values near -18 should therefore be interpreted as near zero PRECT); top right, log$_{10}$(maximum PRECT); bottom left, log$_{10}$(mean PRECT); bottom right, probability of positive PRECT. Note that the color scales are different for the minimum, maximum, and mean values because the range of these values vary dramatically.
<table>
<thead>
<tr>
<th></th>
<th>Mean PRECT</th>
<th>MAE</th>
<th>RMSE</th>
<th>Prob &gt;0</th>
<th>Prob &lt;0</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>2.82e-08</td>
<td>–</td>
<td>–</td>
<td>0.985</td>
<td>0.000209</td>
<td>–</td>
</tr>
<tr>
<td>sz1e-1</td>
<td>1.95e-09</td>
<td>2.76e-08</td>
<td>6.51e-08</td>
<td>0.992</td>
<td>6.37e-05</td>
<td>.0002</td>
</tr>
<tr>
<td>sz1e-2</td>
<td>1.95e-09</td>
<td>2.76e-08</td>
<td>6.51e-08</td>
<td>0.992</td>
<td>6.37e-05</td>
<td>.0002</td>
</tr>
<tr>
<td>sz1e-3</td>
<td>1.95e-09</td>
<td>2.76e-08</td>
<td>6.51e-08</td>
<td>0.992</td>
<td>6.37e-05</td>
<td>.0002</td>
</tr>
<tr>
<td>sz1e-4</td>
<td>1.95e-09</td>
<td>2.76e-08</td>
<td>6.51e-08</td>
<td>0.992</td>
<td>6.37e-05</td>
<td>.0002</td>
</tr>
<tr>
<td>sz1e-5</td>
<td>2.06e-07</td>
<td>1.91e-07</td>
<td>2.50e-07</td>
<td>0.999</td>
<td>0</td>
<td>.0002</td>
</tr>
<tr>
<td>sz1e-6</td>
<td>4.36e-08</td>
<td>4.62e-08</td>
<td>8.31e-08</td>
<td>1</td>
<td>5.65e-05</td>
<td>.0009</td>
</tr>
<tr>
<td>sz1e-7</td>
<td>2.87e-08</td>
<td>3.95e-08</td>
<td>8.07e-08</td>
<td>0.875</td>
<td>0.125</td>
<td>.01</td>
</tr>
<tr>
<td>sz1e-8</td>
<td>2.82e-08</td>
<td>3.87e-08</td>
<td>8.06e-08</td>
<td>0.92</td>
<td>0.0801</td>
<td>.06</td>
</tr>
<tr>
<td>sz1e-9</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.951</td>
<td>0.0488</td>
<td>.15</td>
</tr>
<tr>
<td>sz1e-10</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.964</td>
<td>0.0358</td>
<td>.29</td>
</tr>
<tr>
<td>sz1e-11</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.964</td>
<td>0.0358</td>
<td>.61</td>
</tr>
<tr>
<td>zfpl1e-1</td>
<td>0</td>
<td>2.82e-08</td>
<td>6.58e-08</td>
<td>0</td>
<td>0</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-2</td>
<td>0</td>
<td>2.82e-08</td>
<td>6.58e-08</td>
<td>0</td>
<td>0</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-3</td>
<td>0</td>
<td>2.82e-08</td>
<td>6.58e-08</td>
<td>0</td>
<td>0</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-4</td>
<td>0</td>
<td>2.82e-08</td>
<td>6.58e-08</td>
<td>2.30e-08</td>
<td>2.30e-08</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-5</td>
<td>2.50e-09</td>
<td>3.15e-08</td>
<td>7.98e-08</td>
<td>0.00647</td>
<td>0.00181</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-6</td>
<td>2.40e-08</td>
<td>4.35e-08</td>
<td>8.55e-08</td>
<td>0.277</td>
<td>0.069</td>
<td>.002</td>
</tr>
<tr>
<td>zfpl1e-7</td>
<td>2.81e-08</td>
<td>3.91e-08</td>
<td>8.07e-08</td>
<td>0.76</td>
<td>0.118</td>
<td>.11</td>
</tr>
<tr>
<td>zfpl1e-8</td>
<td>2.82e-08</td>
<td>3.87e-08</td>
<td>8.06e-08</td>
<td>0.892</td>
<td>0.0797</td>
<td>.19</td>
</tr>
<tr>
<td>zfpl1e-9</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.938</td>
<td>0.0538</td>
<td>.28</td>
</tr>
<tr>
<td>zfpl1e-10</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.96</td>
<td>0.0365</td>
<td>.41</td>
</tr>
<tr>
<td>zfpl1e-11</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.969</td>
<td>0.0294</td>
<td>.50</td>
</tr>
<tr>
<td>zfpl0</td>
<td>2.82e-08</td>
<td>3.86e-08</td>
<td>8.06e-08</td>
<td>0.976</td>
<td>0.011</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 5: Summary statistics for PRECT and PRECT errors. The compression algorithm and error tolerance are given in the first column. The metrics are then, from left to right, the mean PRECT value, mean absolute error, root mean square error, probability of positive rainfall, probability of negative rainfall, and compression ratio. Note that in the second column we are showing mean PRECT and not mean PRECT errors, to emphasize that zfp sets PRECT to zero at error tolerances greater than or equal to $10^{-4}$. 

Figure 6: Top, percentage of gridcells with positive rainfall by day of the year, aggregated globally and across years. Left, $sz$; right, $zfp$ (error tolerances indicated in figure legend). The images are then repeated zoomed-in, excluding the results for $zfp$ at error tolerances $10^{-5}$ and $10^{-2}$, where the percentage of rainy days is near zero or exactly zero, respectively. Bottom, the log-odds ratios comparing the compressed output percentages to those in the original output. Odds ratios are compressed/original, so positive log-odds ratios mean that it rains more often in the compressed output, and the opposite for negative log-odds ratios.

There is a discernible seasonal pattern in global probabilities of rainfall, as well as in errors in these probabilities (as measured by the odds ratios). In general, rainfall is more likely in the first half of the year and this is also the period when the odds ratios (comparing the compressed and original data) are largest in magnitude.

4 Gridcell-level Analysis

In this section, we explore compression errors at finer spatial and temporal scales than are presented in the preceding section. While globally, errors appear small and without many noticeable temporal effects, gridcell-level errors can show detectable artifacts that are missed by global summaries and that can themselves produce artifacts in quantities of interest derived from the compressed output.
4.1 Daily Temperature (TS)

First, we investigate artifacts in the compression errors (and corresponding compression output) for daily average temperatures (TS) at the gridcell level.

Before examining artifacts in the compression output at each gridcell, we briefly discuss the nature of some of the temporal artifacts created by SZ and ZFP in two example locations. Figures 7 and 8 show the error time-series at two locations where SZ and ZFP, respectively, produce artifacts (both are shown for an error tolerance of 0.1), along with the histogram of daily errors and their periodogram\(^1\). We also show the mean error, the error standard deviation, and Z-statistic. Note that the Z-statistic at a location is defined as

\[
Z = \frac{\bar{e}}{s_e / \sqrt{N}},
\]

where \(\bar{e}\) is the mean error across time at that location, \(s_e\) is the standard deviation of the errors across time at that location, and \(N = 31,390\) is the number of days. If the errors are mean zero and independent and identically distributed across time, these Z-statistics will be approximately standard normal, so Z-statistics that are very large in magnitude are an indication of locations where the errors are not mean zero and/or are not independent and identically distributed across time.

Both locations show an overall negative bias, indicated by a negative mean error and associated very large in magnitude Z-statistic. There is also a detectable mean seasonal cycle in the errors at both locations (indicated by a spike in the periodogram at the frequency 1/365), and positive correlation in the errors (indicated by a decay in the periodogram in frequency). Errors for SZ are bounded by 0.1 and relatively frequently achieve values close to that bound, whereas ZFP errors are smaller and their distribution has tails that decay more rapidly. The SZ distribution is unimodal, centered near zero; the ZFP error distribution is bimodal, with a large negative mode and a smaller positive mode. Both error time series show evidence of temporal nonstationarity, with periods of increased variability (particularly for the location chosen for SZ). Because of this apparent nonstationarity, the histograms and periodograms shown should be interpreted as representing the average behavior of the error time-series across time and not necessarily its behavior in a particular window of time.

In part motivated by the behavior illustrated in these time-series, in the following, we investigate the behavior of the mean errors, seasonal cycles, standard deviations across time, and temporal dependence structure across time, at each gridcell. Since biases and artifacts can vary by spatial location, we also investigate artifacts in the fine-scale spatial variation produced by compression.

In the following sections, we show results for error tolerances 1.0, 0.5, 0.1, and 0.01. We show results for these highest error tolerances because they show the clearest artifacts. Unless otherwise specified, the behavior of the compressed output at lower error tolerances is qualitatively very similar to the behavior at error tolerance 0.01, until the output becomes lossless. Recall that SZ becomes lossless for daily TS at error tolerance \(10^{-4}\) and ZFP at error tolerance \(10^{-5}\) (Table 2), but ZFP errors take only 5 unique values at error tolerance \(10^{-4}\), most frequently the value zero (Table 3).

---

\(^1\)The periodogram of a time-series is the squared magnitude of the Fourier transform of the time-series, typically evaluated at the Fourier frequencies \(\omega_j = 2\pi j / N\) for \(j = 1, \ldots, N/2\), where \(N\) is the length of the time-series (assuming even \(N\), otherwise \((N + 1)/2\)). The periodogram value represents the amplitude of variability associated with oscillations with the corresponding frequency.
Figure 7: Error time-series for Lon -120 Lat 64.6 (a location in northwest Canada) for $SZ$ at a tolerance of 0.1 for TS. Left, time-series for the first 3 years; middle, histogram of entire time-series; right, periodogram of entire time-series. The mean, standard deviation, and associated $Z$-statistic are given in the right margin.

Figure 8: Error time-series for Lon -83.8 Lat -82.5 (a location in Antarctica) for $ZFP$ at a tolerance of 0.1 for TS. Left, time-series for first 3 years; middle, histogram of entire time-series; right, periodogram of entire time-series. The mean, standard deviation, and associated $Z$-statistic are given in the right margin.
4.1.1 Mean Errors

We first show mean compression errors at the gridcell level, indicating locations with consistent positive or negative biases. Figure 9 shows the mean errors produced by the ZFP and SZ algorithms at a range of error tolerances (0.01, 0.1, 0.5, 1.0) for the daily temperature data.

In the SZ output there are longitudinal bands showing large mean errors (at longitudes -120, 0 and 120) at all error tolerances shown, in addition to large mean errors at the South Pole and in parts of the Atlantic Ocean and in Oceania. At error tolerances of 0.5 and 1, mean errors over the ocean appear to be strongly spatially correlated.

To provide more detail for the behavior around the aforementioned longitudes, Figure 10 shows zoomed-in plots of mean errors for SZ at an error tolerance of 0.01 around longitudes -120 degrees (left), 0 degrees (middle), and 120 degrees (right). In the first two plots there are stripes of alternating large positive and then small mean values at the specified longitudes. In the last plot there is a stripe of alternating large negative and then small mean values at the specified longitude.

In comparison to Sz, ZFP mean errors show much stronger patterns than do Sz errors (Figure 9, right). In particular, a 4 × 4 gridding pattern is apparent in the mean errors for all compression levels where ZFP is not lossless, with periodic mean errors that are large in magnitude. The gridding pattern does not align with the positions of the 4 × 4 partition used in the ZFP algorithm across the entire globe: there is misalignment at particular longitudes (again at -120, 0, and 120 degrees). The large mean errors are mostly positive for error tolerances of 0.1 and 1 and mostly negative for error tolerances of 0.01 and 0.5. Whatever the dominant sign for the error tolerance of interest, there is a shift towards the South Pole where the mean errors become close to zero and then reverse in sign.

Figure 11 is a zoomed-in plot of mean errors at an error tolerance of 0.01 for the ZFP algorithm, around the longitudes -120, 0, and 120. The structure of the 4 × 4 gridding, as well as the misalignment noted above, is more apparent here. In particular, there are two dominant patterns of 4 × 4 gridding that alternate moving along a longitude band. Every four longitudes, there is a band containing large negative mean errors every three and then five latitudes. Along this band, from north to south, the pattern is large negative, small negative, small, large negative, small positive, small, small negative, small positive. The next (eastward) longitudinal band has mean errors with small absolute values, the next contains moderately-sized mean errors that are opposite of the first band in sign, and the final band contains moderately-sized mean errors that match the first band in sign. The aforementioned misalignment occurs at the -120, 0, and 120 longitudes, where the whole pattern shifts south by three, two, and three gridcells, respectively.

Because the 4 × 4 gridding pattern in the ZFP mean errors does not align with this partition, artifacts are even more apparent when errors are aggregated to the 4 × 4 partition used by ZFP. Figure 12 shows the ZFP mean errors when the errors are aggregated to the ZFP 4 × 4 partition (again for an error tolerance of 0.01). Distinct regions of striping appear between longitudes -120 to 0, 0 to 120, and 120 to -120.

To emphasize that the artifacts in mean errors discussed above are much larger than would be expected if compression errors were mean zero and independent and identically distributed in time, we show Z-statistics associated with the mean TS errors (Figure 13). We calculate Z-statistics for the errors in the same way as described in Eq. (2), except at each location separately. The patterns for SZ and ZFP at different error tolerances are very similar to the mean errors in Figure 9. The magnitude of the Z-statistics can be quite large, indicating clear evidence at the gridcell level that the compression algorithm produces mean biases in the TS field. This is particularly true for ZFP errors, where large errors are widespread spatially.

In Figure 13, the percentage of gridcells that are considered to be significant and the cut-off Z-statistic value for significance are given in the plot titles. The percentage of significant gridcells in based on p-values calculated for the Z-statistics in comparison to the standard normal distribution, which is the reference distribution for the Z-statistics if the errors have a mean of zero and are independent and identically distributed. Gridcell-level p-values themselves, however, cannot be taken at face value because of the fact
Figure 9: Mean errors for TS. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Overall mean errors are indicated in the plot titles.
Figure 10: Zoomed-in plot of the mean errors of $sz$ at a tolerance of 1e-2 for TS. Note the longitudinal stripes at -120, 0, and 120 degrees longitude.

Figure 11: Zoomed-in plot of the mean errors of $zfp$ at a tolerance of 1e-2 for TS. Note the misalignment at -120, 0, and 120 degrees longitude.
Figure 12: Mean errors in TS aggregated to the ZFP $4 \times 4$ partition for ZFP compression with an error tolerance of .01. Three distinct regions of striping are visible with boundaries at -120, 0, and 120 degrees longitude. These are also the locations of misalignment in the ZFP gridding structure shown in Figure 9, right.
that multiple gridcells are being tested simultaneously (i.e., even if at every gridcell there were no mean bias, 5% of gridcells would be expected to show p-values less than 0.05). To account for this multiple testing problem, we control the false discovery rate at 1%, using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), which has been advocated as a method for making significance statements at the gridcell level for atmospheric variables (e.g., Wilks (2006, 2016)). In this context, the false discovery rate is defined as the expected proportion gridcells declared “significant” (i.e., discoveries) that in fact do not show true mean biases. The Benjamini-Hochberg procedure for controlling false discovery rates was developed for independent data, but is conservative for positively correlated p-values (as would be expected for atmospheric variables like temperatures). Over 90% of gridcells in zfp have a $Z$-statistic that is deemed significant at the false discovery rate of 1%, whereas under 2% of those in sz are deemed significant.

As discussed in the preceding section, the mean error behavior at lower error tolerances is qualitatively similar to that discussed here, until the algorithms become lossless. The one exception is that, since zfp errors at error tolerance $10^{-4}$ take only five unique values, most often the value zero (Table 3), the pattern of mean errors is slightly different (Figure 14). There is still a gridding pattern, but at about 20% of locations the output is exactly lossless (i.e., all errors are zero).
Figure 13: Z-statistics associated with the mean TS errors. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. The thresholds for Z-statistics considered significant at a false discovery rate of 1% are shown in the plot titles, along with the percentage of significant gridcells. Compare to mean errors (Figure 9).
4.1.2 Contrast Variances

The artifacts seen in the mean TS errors (Figure 9) result in similar artifacts in the TS contrast variance fields in the compressed output. Contrast variances are natural quantities to inspect for artifacts of compression, because differences in temperatures between adjacent gridcells are typically small. (In comparison, raw temperatures are typically more variable than the compression error tolerance, which can mask artifacts of compression in the raw temperature fields.) Contrast variances were also used in Guinness and Hammerling (2018) to assess quality of compression.

We calculate the East-West contrast variances for month $m$ at latitude $L$ and longitude $l$ as

$$c_{l,L}(m) = \frac{1}{86D} \sum_{y=1}^{86} \sum_{d=1}^{D} (T_{l,L}(m,d,y) - T_{l+1,L}(m,d,y))^2,$$

(3)

where $T_{l,L}(m,d,y)$ is the temperature value at that location in day $d$, month $m$, and year $y$, and $D$ is the number of days in the month of interest. Artifacts in contrast variances would indicate biases in the fine-scale spatial dependence structure of the compressed output. East-West contrast variances for January are shown in Figure 15. (Comparisons between the original and compressed data in other months are qualitatively very similar, but East-West contrast variances themselves do vary by month. For example, in regions where there is seasonal sea ice, contrast variances are larger in months with sea ice than in months without ice.)

As stated above, some of the artifacts shown in the mean errors (Figure 9) are apparent in the contrast variances, particularly at high error tolerances: e.g., in sz, patterns around the 0 longitude in the Norwegian Sea, and in zfp a gridding structure. Moreover, and perhaps more importantly, contrast variances in the compressed output are typically larger than in the original output. This bias is larger at higher error tolerances and over oceans, and larger for SZ compared to ZFP. (The ratios of average contrast variances in compressed vs. original output are shown in Figure 15, averaging globally and
also separately over land and ocean.) Thus the compressed output, particularly at high error tolerances, appears to *increase* fine-scale spatial variation.
Figure 15: January East-West TS contrast variances, on the log scale base 10. Top, the contrast variances for the original data. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. For high tolerances, sz compression shows an overall positive bias in contrast variances (indicating that algorithm increases fine-scale spatial variation) and ZFP compression shows a gridding pattern. The global mean contrast variance is provided for the original data in the plot title, while for compressed data the ratio of the global mean contrast variance to that of the original is provided; land and ocean averages are also given.
4.1.3 Standard Deviations

In addition to spatial patterns of mean biases, spatial patterns of error variances would indicate locations where the compression is more vs. less successful. Figure 16 shows the log ratio of the error variance at a location divided by the global pooled variance\(^2\). Negative log ratios correspond to locations where the local error variance is smaller than the global pooled error variance, the opposite for positive log ratios. (Note that, given an error tolerance, global pooled variances are always larger in sz compared to zfp, similar to the MAEs and RMSEs shown in Table 2.)

In the sz output, patterns similar to those shown for the means appear in the variances as well: reduced variances along longitudinal bands at longitudes -120, 0, and 120, as well as over the Atlantic Ocean, in Oceania, and in the Himalayas.

The zfp output again shows a gridding pattern (the pattern for error tolerance 10\(^{-4}\), not shown, is also similar to that for the mean shown in Figure 14). There are also areas of smaller variances at both of the poles at all error tolerances shown. Additionally, there are wide bands in the Arctic and Antarctic (but away from the poles) showing enhanced error variability; this is due to a seasonal cycle in the errors (see Section 4.1.4). At larger error tolerances, yet another pattern emerges over the ocean in the Southern Hemisphere: regions of reduced error variability. Perhaps surprisingly, these regions of reduced error variability also show biases in the mean seasonal cycle (again see Section 4.1.4). Embedded within this region of reduced error variability are three repeating patterns in the shape of South America and Australia, each separated by 120 degrees longitude.

---

\(^2\)In this setting, where the time series length is the same at each gridcell, the pooled variance is defined simply as the average of the gridcell-level variances. The pooled standard deviation is then the square root of the pooled variance.
Figure 16: Ratio of the error variance to the global pooled error variance for TS, shown on the log scale (base 10). Negative log ratios correspond to locations where the variance is smaller than the global pooled variance, and the opposite for positive values. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. The global pooled standard deviation is indicated in the title of each graph.
4.1.4 Biases in Seasonality

As demonstrated in Figures 7 and 8, variation in errors over time can in part be due to biases in the mean seasonal cycle. Figure 17 shows the mean seasonal cycles at the two gridcells used in the aforementioned figures, one where ZFP produces a bias and the other where SZ produces a bias, along with the error seasonal cycle produced by the relevant algorithm (standardized to have a standard deviation of 1, so that errors at different compression levels can be more easily compared). At both locations, the absolute bias in the seasonal cycle is small because the error tolerances are much smaller than the scale of the original seasonal cycle; however, the biases are strongly detectable relative to the error variability. In the location where ZFP produces a bias, the sign of the bias is inconsistent across the varying error tolerances (as also seen in the overall mean errors in Figure 9 and Table 2). In the location where SZ produces a bias, the seasonal cycle is slightly enhanced in the SZ output. Note that while there is substantial variability in the estimated error seasonal cycle for the location chosen for SZ, the seasonality is still strongly detectable (i.e., the error seasonal cycle does not look like white noise).

To identify the biases in mean seasonal cycles globally, we examine the amplitude of the first seasonal harmonic of the error at each grid cell (i.e., the value of the periodogram of the errors at a frequency of 1/365 days$^{-1}$). Figure 18 shows the relative strength of the error seasonal cycles, where we define this quantity as the ratio of the amplitude of the first seasonal harmonic of the errors compared to the average value of the periodogram in a neighborhood of 50 frequencies around the annual frequency (excluding the annual frequency). This quantity is a measure of how spiked the periodogram is at the annual frequency. If there is no error seasonal cycle and if the spectral density of the errors is close to constant in the window around the annual frequency, the aforementioned ratio will have approximately an $F_{2,100}$ distribution. Thus ratios much larger than would be expected under that distribution indicate areas with strong error seasonality. In Figure 18, grid cells showing “significant” error seasonal cycles are indicated, where significance is assessed by computing a $p$-value with respect to the $F_{2,100}$ distribution and controlling the false discovery rate at 1%, using the Benjamini-Hochberg procedure.

In SZ, error seasonality is less common than in ZFP (for any given tolerance, the percentage of gridcells with significant correlation is lower), but is apparent in a few longitudinal bands at longitudes -120, 0 and 120 and in some regions of the ocean.

In ZFP, error seasonality is most consistently seen in the Arctic and Antarctic (apparent for all error tolerances shown). For higher error tolerances, additional areas with error seasonality emerge over much of the ocean. At tolerances of 0.5 and 1, about half of all gridcells show significant error seasonality (in particular over most of the ocean).
Figure 17: Top left, seasonal cycles for daily TS in the original output and ZFP compressed output at a location in Antarctica where ZFP errors show a strong mean seasonal cycle (so that ZFP temperatures themselves have a biased mean seasonal cycle). Top right, the error seasonal cycles at these locations (standardized to have a standard deviation of one). Bottom, the same, but for SZ output at a location in western Canada where SZ produces biases in the seasonal cycle.
Figure 18: Amplitudes of the error annual harmonic relative to the average periodogram value in a neighborhood of 50 frequencies around the annual frequency for TS errors. Values are shown on the log scale (base 10). Left, sz; right, zfp. Error tolerances are indicated on the left margin. Locations where the annual harmonic amplitude is “significant” at a 1% false discovery rate are marked with a gray dot; the percentage of significant gridcells for a given error tolerance and algorithm is given in each plot title.
4.1.5 Temporal Correlations

We now investigate temporal correlation in compression errors and compression TS output at the gridcell level. Figure 19 shows the lag-1 correlations of the deseasonalized errors for daily TS for error tolerances of 0.01, 0.1, 0.5, and 1.0 under the sz and zfp algorithms. (We deseasonalized the errors by subtracting their daily averages, so as not to conflate the error seasonality with other sources of autocorrelation.) In Figure 19, gridcells showing “significant” lag-1 correlations are indicated, controlling the false discovery rate at 1% (again using the Benjamini-Hochberg procedure). In assessing significance, \( p \)-values at each gridcell are computed with respect to the normal distribution with a mean of zero and a standard deviation of \( 1/\sqrt{N} \), where \( N = 31,390 \) is the number of days in the simulation; this is the approximate distribution of the sample lag-1 autocorrelation coefficient when the true correlation is zero (e.g., Shumway and Stoffer (2017) Eq. (1.38)).

In the sz output at a tolerance of 0.01, there are significant correlations in some vertical stripes at longitudes of -120, 0, and 120 and there are also significant correlations at both poles. At tolerances above 0.1, there is striping pattern of stronger positive lag-1 correlations that becomes more widespread for larger error tolerances. Most of the ocean shows significant positive lag-1 correlations at an error tolerance of 1.0.

In the zfp output, gridcells towards both poles show significantly positive autocorrelations at all error tolerances shown. At tolerances above 0.1, there is also a repeated pattern of strong, positive lag-1 correlation in the Southern Hemisphere oceans that are marked as significant, similar to that shown in Figure 18. Overall, for any given tolerance level, the number of gridcells with significant correlation is always equal or greater for zfp than for sz.

Typically we will be more concerned about changes to the temporal correlation structure of the TS series themselves rather than temporal correlations in the errors. It is important to recognize that temporal correlations in the raw TS field are affected both by temporal correlations in the errors and also, e.g., by the overall error standard deviation. Figure 20 compares the temporal correlation structure in the original output to that in the compressed output. We show the lag-1 autocorrelations of the first differences of the de-seasonalized TS values. (We first-difference and de-seasonalize the data as a naive way to remove temporal trends and seasonal cycles that would otherwise be confounded with other sources of temporal correlation.)

Both algorithms strongly suppress temporal correlation at high error tolerances, particularly over the ocean (where temporal correlation in the original output is strongest). The suppression is stronger for sz than for zfp; however, zfp also produces gridding artifacts and the patterns of South America and Australia in the Southern Hemisphere oceans, also mentioned above. Even at an error tolerance of 0.1, the sz TS field shows temporal correlations over oceans that are much weaker than in the original. At an error tolerance of 0.01, the temporal correlations appear to match those in the original.
Figure 19: Lag-1 correlations of the deseasonalized errors for TS. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Locations where the lag-1 correlation is “significant” at a 1% false discovery rate are marked with a gray dot. The percentage of locations that are significant is indicated in each plot title.
Figure 20: Lag-1 correlations of the first differences of the de-seasonalized TS values. Top, original output. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. We emphasize that these are correlations of first differences; negative values are a result of over-differencing and do not imply that the original series is negatively correlated in time.
4.1.6 Time of Maximum and Minimum Errors

Finally, we investigate the day of the year that the mean absolute error (averaged across years) is largest and smallest for each gridcell (Figures 21 and 22, respectively). For small error tolerances, there are no clear spatial patterns in the day that errors tend to be smallest or largest. For error tolerances of 0.5 and especially 1, however, spatial correlation is apparent in the days of both the minimum and maximum MAE, indicating that large or small errors are achieved contemporaneously at nearby spatial locations. (If maximum and minimum errors did not tend to occur contemporaneously at nearby spatial locations, the distribution of the day with the maximum or minimum MAE would not show spatial patterns and would be spatially uniform.)
Figure 21: Day with the largest mean absolute error for TS. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Spatial correlation is apparent particularly in the oceans at higher error tolerances.
Figure 22: Day with the smallest mean absolute error TS. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Spatial correlation is apparent particularly in the oceans at higher error tolerances.
4.2 Monthly Maximum Temperature (TSMX)

In the following, we repeat the analyses shown in Section 4.1, but for monthly maximum temperatures (TSMX) at the gridcell level.

Figures 23 and 24 concern the time-series of sz and zfp TSMX errors, respectively, at the same locations used for Figures 7 and 8. As with the daily TS errors, zfp errors show an overall negative bias and a mean seasonal cycle; the sz errors also show a mean seasonal cycle, but do not show strong evidence of an overall bias. Unlike for daily TS, the monthly TSMX errors do not appear to be strongly correlated over time besides the seasonality in the zfp errors (i.e., the periodogram of the errors is otherwise fairly flat).

Similar to our results for TS, in the following sections, we show results for error tolerances 1.0, 0.5, 0.1, and 0.01. As with TS, unless otherwise specified, the behavior of the compressed output at lower tolerances looks qualitatively similar to that for error tolerance 0.01 until the algorithm output becomes lossless. Recall that sz produces lossless output at error tolerance $10^{-4}$ and zfp at error tolerance $10^{-5}$ (Table 4).
Figure 23: Error time-series for Lon -120 Lat 64.6 for $sz$ at a tolerance of 0.1 for TSMX. Left, time-series for first 5 years; middle, histogram of entire time-series; right, periodogram of entire time-series. The mean, standard deviation, and associated $Z$-statistic are given in the right margin. Compare to Figure 7.

Figure 24: Error time-series for Lon -83.8 Lat -82.5 for $zfp$ at a tolerance of 0.1 for TSMX. Left, time-series for first 5 years; middle, histogram of entire time-series; right, periodogram of entire time-series. The mean, standard deviation, and associated $Z$-statistic are given in the right margin. Compare to Figure 8.
4.2.1 Mean Errors

Figure 25 shows spatial patterns of mean errors produced by sz and zfp. (Compare to Figure 9.)

For the sz output, similar to the daily TS errors, there are longitudinal bands showing large mean errors at longitudes of -120, 0, and 120. There are also large, positive mean errors north of New Zealand and in the middle of the Atlantic Ocean. At higher error tolerances, there are large mean errors in the ocean near Norway.

Likewise, in the zfp algorithm, similar to the daily TS errors, a gridding pattern is apparent. (The pattern for error tolerance $10^{-4}$, not shown, is also similar to that shown in Figure 14.) Also, as for daily TS errors, the signs of the zfp errors change based on the error tolerance. Again, at higher tolerances, there are additional patterns that emerge, especially over the Southern Ocean.

Figure 26 shows the $Z$-statistics associated with the mean monthly TSMX errors. (Compare to Figure 13.) $Z$-statistics are calculated in the same way as described in Eq. (2), except at the gridcell level and for monthly data the number of time points is $N = 1,032$. Patterns of $Z$-statistics for with TSMX errors are qualitatively similar to those for TS errors; however, $Z$-statistics tend to be smaller for TSMX errors and somewhat fewer gridcells are deemed significant at the 1% false discovery rate.
Figure 25: Mean errors for TSMX. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Overall mean errors are indicated in the plot titles. Compare to 9.
Figure 26: Z-statistics associated with the mean TSMX errors. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. The threshold for Z-statistics considered significant at a false discovery rate of 1% is shown in each plot title, along with the percentage of significant gridcells. Compare to Figure 13.
4.2.2 Contrast Variances

Figure 27 shows the January contrast variances for monthly TSMX. (Compare to Figure 15.) These quantities are computed in the same way as described in Eq. (3) except that, since this is monthly data, there is no summing over the days of the month.

The January contrast variances for TSMX look very similar to those for January TS. Namely, the artifacts seen in the mean error appear here, and the contrast variances are biased large particularly in SZ at high error tolerances over oceans.
Figure 27: January East-West TSMX contrast variances. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Compare to Figure 15.
4.2.3 Standard Deviations

Figure 28 shows the log ratio of the local error variance divided by the global pooled error variance. (Compare to Figure 16.) The patterns shown are similar to those for TS errors; however, artifacts are somewhat more apparent for SZ at higher error tolerances here compared to for TS errors, and unlike for TS errors, ZFP TSMX errors do not show enhanced variability in the Arctic.
Figure 28: Ratio of the error variance to the global pooled error variance for TSMX. Values are shown on the log scale (base 10). Negative log ratios correspond to locations where the error variance is smaller than the global pooled variance, and the opposite for positive log ratios. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. The overall pooled standard deviation is indicated in each plot title. Compare to Figure 16.
4.2.4 Biases in Seasonality

Figure 29 shows the relative strength of the error seasonal cycles for monthly TSMX. (Compare to Figure 18.) The definition of this quantity, and the assessment of its significance, is identical to that described in Section 4.1.4.

Error seasonality is less commonly detected in monthly TSMX errors compared to daily TS errors. However, as with daily TS errors, in zfp, error seasonality is most consistently seen in the Arctic and Antarctic (apparent for all error tolerances shown). For error tolerances of 0.5 and 1, additional areas with error seasonality emerge over oceans in the Southern Hemisphere and the tropics. Note also that the three regions in the shape of Australia and South America also shown in Figure 28 and elsewhere have little or no error seasonality within this larger region.

Likewise, as with daily TS errors, in sz, error seasonality is less common than in zfp but is apparent at small error tolerances in a few longitudinal bands over the Pacific Ocean and Australia. At larger error tolerances, error seasonality emerges in some regions of the ocean.
Figure 29: Amplitudes of the error annual harmonic relative to the average periodogram value in a neighborhood of 50 frequencies around the annual period for TSMX errors. Values are shown on the log scale (base 10). Left, sz; right, zfp. Error tolerances are indicated on the left margin. Locations where the annual harmonic amplitude is “significant” as defined in the text are marked with a gray dot; the percentage of significant gridcells for a given error tolerance and algorithm are given in the plot title. Compare to Figure 18.
4.2.5 Temporal Correlations

Figure 30 shows lag-1 correlations for de-seasonalized monthly TSMX errors. (Compare to Figure 20.) Here the errors were deseasonalized by subtracting their monthly averages, and gridcells showing “significant” lag-1 correlations are indicated, controlling the false discovery rate at 10% (assessed in the same way as described in Section 4.1.5, except that here \( N = 1,032 \) is the number of months rather than days in the simulation). The reason we use a 10% false discovery rate, whereas in the rest of this document we use a 1% false discovery rate, is that monthly TSMX errors appear nearly uncorrelated over time at almost all locations (unlike for daily TS errors).

Indeed, for TSMX errors, correlation is generally much weaker than in TS errors, and far fewer gridcells show significant correlations. In sz, errors appear significantly correlated at the South Pole at an error tolerance of 0.1; at error tolerances of 0.5 and 1, sample autocorrelations appear to be strongly spatially correlated, but are not deemed significant at the gridcell level at the 10% false discovery rate. For zfp, errors appear correlated in the Southern Ocean at error tolerances of 0.5 and 1.

Figure 31 shows the lag-1 autocorrelations of the first differences of the de-seasonalized TSMX values themselves. (Compare to Figure 20.) Overall, temporal correlations are also unsurprisingly much weaker for monthly TSMX values than for daily TS values. As such, the suppression in temporal correlation seen for daily TS values is not as severe here; however, at error tolerances of 0.5 and 1, sz still noticeably suppresses temporal correlations and zfp produces a noticeable gridding pattern in the temporal correlations.
Figure 30: Lag-1 correlations of the deseasonalized errors for TSMX. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Locations where the lag-1 correlation is “significant” at the 10% false discovery rate are marked with a gray dot. The percentage of locations that are deemed significant is indicated in each plot title. Note that we are using a 10% false discovery rate here, whereas other figures use a 1% false discovery rate. (Compare to Figure 19.)
Figure 31: Lag-1 correlations of the first differences of the de-seasonalized TSMX values. Top, original output. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. We emphasize that these are correlations of first differences; negative values are a result of over-differencing and do not imply that the original series is negatively correlated in time. (Compare to Figure 20.)
4.2.6 Time of Maximum and Minimum Errors

Figures 32 and 33 show the month of the year that the mean absolute monthly TSMX error (averaged across years) is largest and smallest, respectively. (Compare to Figures 21 and 22.) The patterns look similar to those shown for daily TS. In particular, spatial correlation is apparent at high error tolerances.
Figure 32: Month with the largest mean absolute error for TSMX. Left, $sz$; right, $zfp$. Compression error tolerances are indicated on the left margin. Compare to Figure 21.
Figure 33: Month with the smallest mean absolute error for TSMX. Left, sz; right, zfp. Compression error tolerances are indicated on the left margin. Compare to Figure 22.
4.3 Daily Average Rainfall Rate (PRECT)

We now turn our attention to daily average rainfall rates (PRECT). In part because rainfall has a positive probability of being zero, the metrics we use to evaluate the quality of compression are necessarily different from those used for evaluating temperatures. As discussed in Section 3.3, the compressed output shows global biases in the percentage of days with positive rainfall and also produces days with negative PRECT values (in excess of those days for which PRECT is negative in the original output).

To examine these biases at the gridcell level, Figure 34 shows the first 50 days of daily average rainfall rates (PRECT) at two locations on the globe (one gridcell in Minnesota and one off the coast of South America) in the original output and at various error tolerances using both algorithms. The biases described above are readily apparent, particularly for the larger error tolerances, where PRECT is set to zero in ZFP and positive but too small in SZ. Both algorithms also produce negative PRECT values at small error tolerances; however, ZFP appears to perform somewhat better at smaller error tolerances at these two locations.

To assess the bias in the number of rainy days produced by the compression algorithm, we compare the odds of daily rain at each gridcell in the original and compressed output. The odds of daily rain, and the odds ratio comparing compressed to original output, are calculated similarly to how described in Section 3.3, except that the probability of rain is calculated at the gridcell level across all days.

Figure 35, shows the odds of daily rain, and odds ratio, comparing original and compressed output at each gridcell for SZ (left) and ZFP (right), under four error tolerances (and for the “lossless” mode of ZFP). There is a negative bias in the odds of rain under the compressed output at all locations at small error tolerances. The odds of rainfall in SZ are qualitatively similar to those in ZFP at these error tolerances, except that the 4 × 4 gridding used in the ZFP algorithm is also apparent in that output. The locations where the bias is strongest are equatorial ocean gridcells and off the western coasts of continents, where the odds of rain in the original output are on the order of 100 times greater than in the compressed output, even at the smallest error tolerances. In the median over the globe, odds of rain in the original output is about 2.7 times greater than in the ZFP “lossless” output. Similarly, the odds of rain in the original output is about 11.5 times greater than in SZ at its lowest error tolerance of 10\(^{-11}\), in the median. At larger error tolerances, ZFP produces no rainfall whereas SZ produces positive rainfall at all locations on nearly every day (and the few nonpositive rain days occur at every location simultaneously).

In addition to biases in the number of days with positive rainfall, both algorithms produce days with negative rainfall (Figure 36, which shows the probability of negative rainfall at each gridcell). The ZFP algorithm produces negative rainfall at some locations for error tolerances smaller than 10\(^{-5}\), and the SZ algorithm produces negative rainfall at some locations for error tolerances smaller than 10\(^{-8}\). Negative PRECT values are most widespread at error tolerances of 10\(^{-6}\) and 10\(^{-7}\). While PRECT can be negative in the original output, negative values are much more common in compressed output with small error tolerances. Negative rainfall appears to be most common in desert regions at the smallest error tolerances, although we note that for other error tolerances negative rainfall can appear fairly widespread globally.
Figure 34: The first 50 days of PRECT at two locations (coordinates indicated in plot titles). Top, a location in Minnesota; bottom, a location in the Pacific Ocean west of South America. Left, $sz$; right, $zfp$ (error tolerances indicated in the plot legends). Both algorithms produce absolute biases in PRECT as well as in number of rainy days, and both can produce days with negative PRECT values. Note in the bottom right panel that the $zfp$ output at error tolerance of $10^{-5}$ produces one very large value that we have cut off in order to more easily see the behavior on the other days.
Figure 35: Odds of daily rainfall, and odds ratio comparing compressed to original output. Values are shown on the log scale (base 10). Odds ratios are computed as compressed / original, so negative log-odds ratios correspond to locations where daily rainfall is less likely in the compressed output compared to the original. The compression algorithm is indicated in the columns and the compression error tolerance in the rows. Note that at the tolerance 0.01, nearly all days in sz show positive rainfall and no days in zfp show positive rainfall.
Figure 36: Probability of negative daily PRECT. Top, in the original output; left, sz; right, zfp. Compression error tolerances are indicated on the left margin.
5 Discussion

In this work we have sought to give a detailed exploration of the effects of SZ and ZFP compression on surface temperature and precipitation fields from a historical run of CESM. The differences between the characteristics of the distributions of surface temperatures vs. precipitation call for different methods of evaluation for these two variables, and result in differing levels of success in compression.

For surface temperatures, it appears that both algorithms can achieve good fidelity with the original output for daily average and monthly maximum temperatures, for modest compression error tolerances. However, for larger error tolerances, both algorithms produce detectable artifacts that can impact important features of the spatiotemporal structure of temperatures. The ZFP algorithm tends to produce more artifacts in the temporal mean structure of temperatures (e.g., mean biases and biases in seasonality), whereas the SZ output tends to produce larger biases in the spatiotemporal correlation structure of temperatures (e.g., enhancement of contrast variances and reduction of temporal correlations).

It is worth emphasizing that, even at the highest error tolerances considered here, errors are substantially smaller than natural temperature variability, which means that pointwise comparisons of the original and compressed output will tend to make the compressed output look successful (i.e., high correlation between compressed and original output); however, small but detectable artifacts in errors can produce detectable and possibly important biases in fine-scale spatial and temporal correlations. These biases in small-scale spatial and temporal variation will typically not be captured by global measures of compression quality, highlighting the importance of inspecting compressed output at multiple spatial and temporal scales.

For daily precipitation rates, both algorithms inflate the number of days without positive rainfall even at very small error tolerances. It is possible that rainfall values are more difficult to compress than are temperatures, because rainfall has a positive probability of being zero on a given day but also a strongly skewed distribution of positive values with a large range. It is also possible that the compression algorithm struggles to preserve very small but positive precipitation values (perhaps in part due to the fact that precipitation values were stored in the units m/s).

Lossy data compression is promising for reducing the storage requirements resulting from large climate model experiments. However, in order to ensure that minimal scientific information is lost due to compression, it is important to evaluate the quality of compression with a sensitivity both to the characteristics of the climate variables of interest and also to climate variability at multiple spatiotemporal scales. These questions are worth investigating in detail so that we may then work together with compression algorithm development teams to address important issues that may arise for climate model data compression.

References


