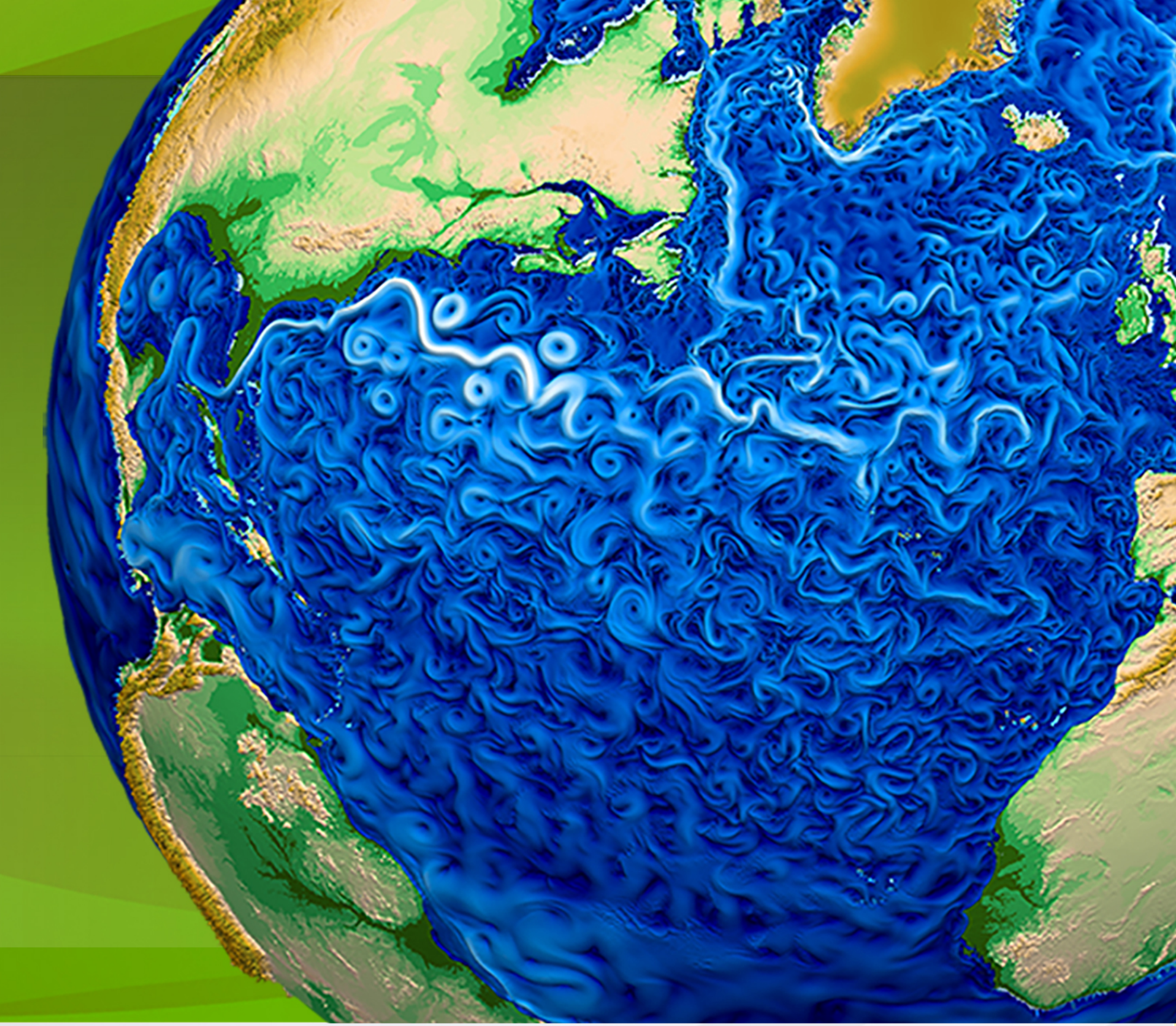


R: Bit Groom ACME Data

Charles S. Zender (UCI) and Jeremy D. Silver (U. Melbourne)



Objective

Why compress?

Lossless compression can reduce climate data storage by 30-40%. In general, further reductions require lossy compression that also reduces precision. Fortunately, geoscientific models and measurements generate false precision (scientifically meaningless data bits) that can be eliminated without sacrificing scientifically meaningful data. We introduce Bit Grooming, a lossy compression algorithm that removes the bloat due to false-precision, those bits and bytes beyond the meaningful precision of the data.

How good is Bit Grooming?

Bit Grooming is statistically unbiased, applies to all floating point numbers, and is easy to use. Bit-Grooming reduces ACME data storage requirements by 40-80%. We compared Bit Grooming to competitors Linear Packing, Layer Packing, and GRIB2/JPEG2000. The other compression methods have the edge in terms of compression, but Bit Grooming is the most accurate and certainly the most usable and portable.

Why don't we use it already?

We're lazy. Bit Grooming provides flexible and well-balanced solutions to the trade-offs among compression, accuracy, and usability required by lossy compression. ACME and its users could reduce their long term storage costs, and show leadership in the elimination of false precision, by adopting Bit Grooming.

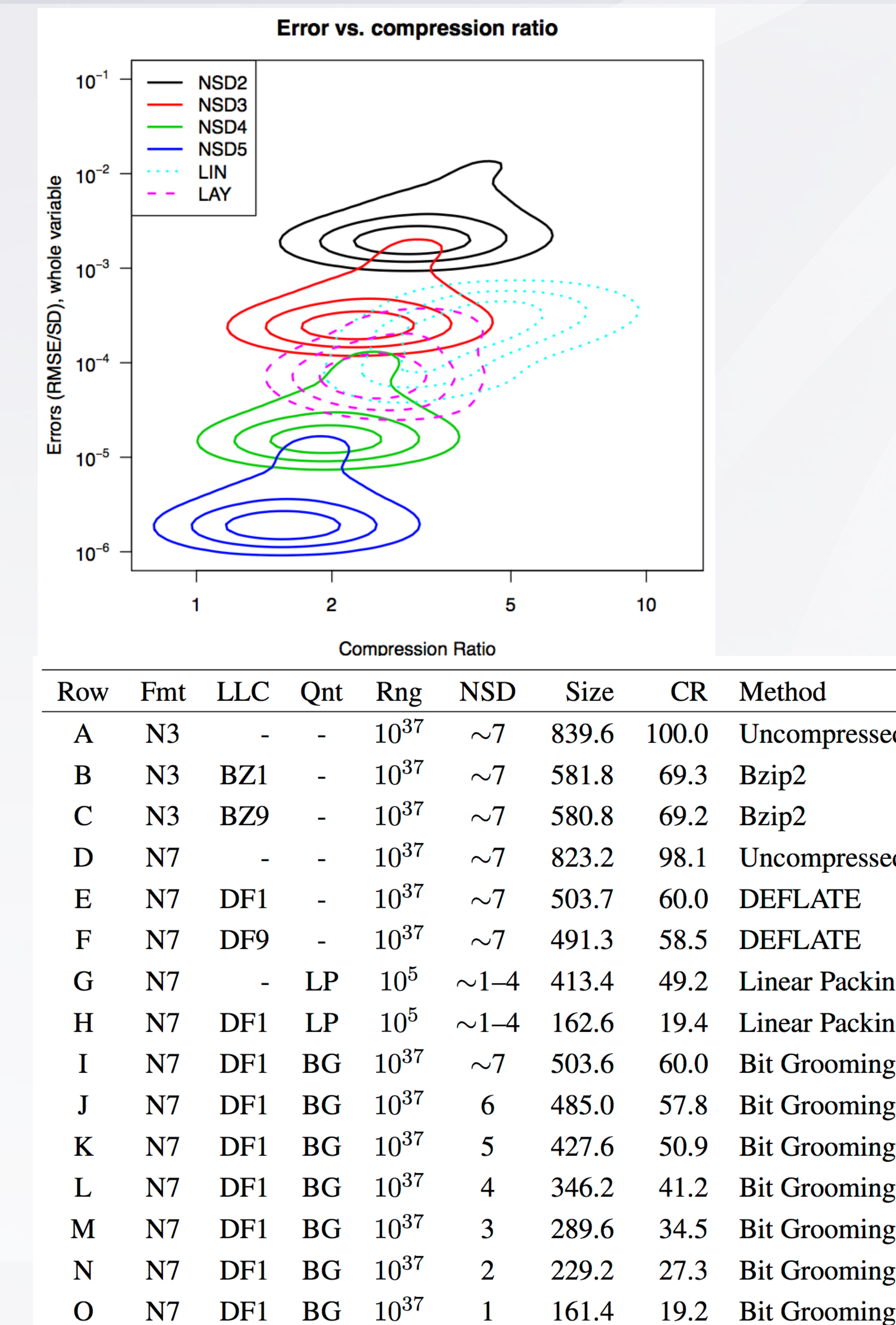
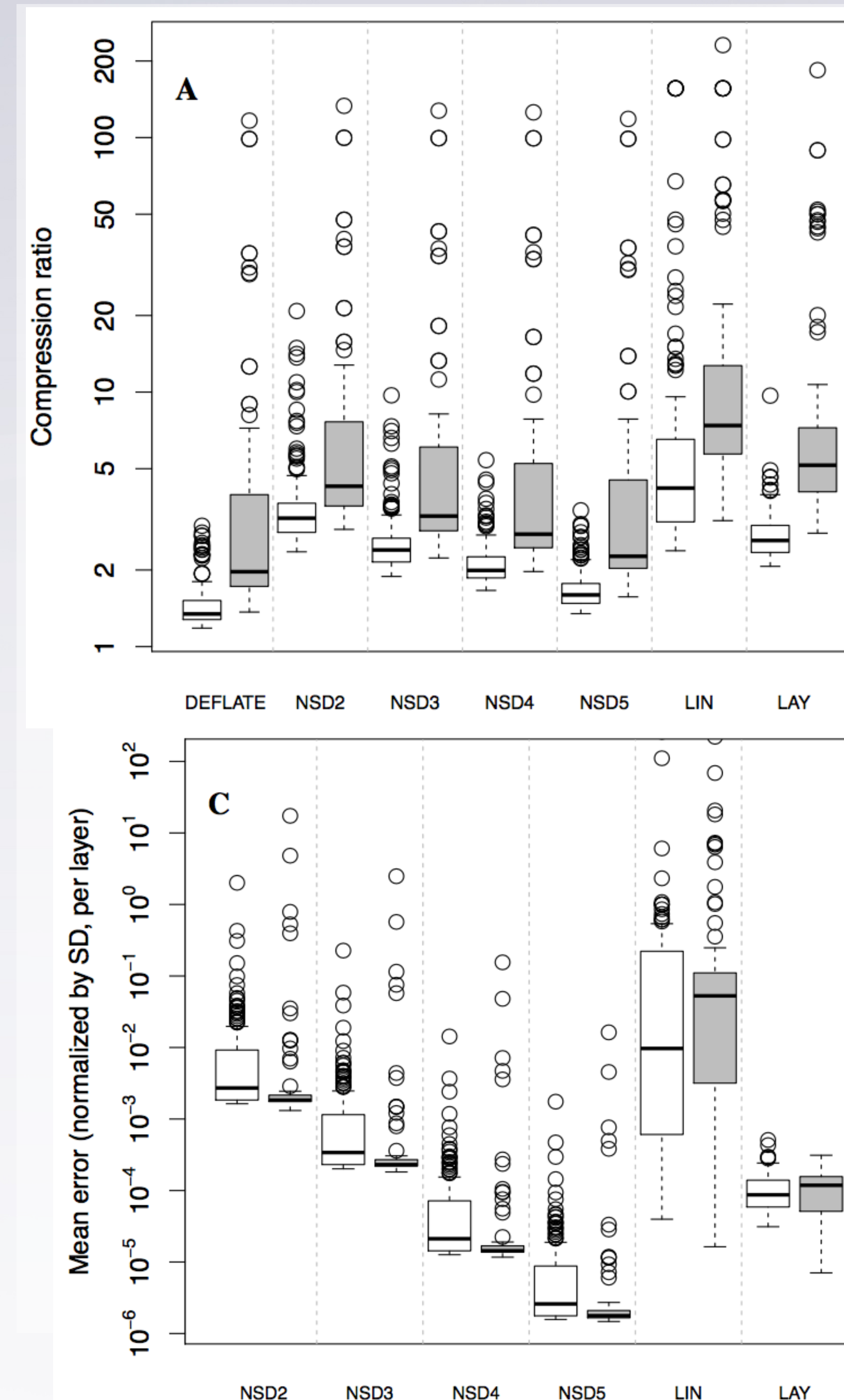
Approach

Computational Approach:

- Alternately shave (to zero) and set (to one) the least significant bits of consecutive values
- Symmetric, two-sided variant of Bit Shaving algorithm that solely zeroes bits
- Alternation eliminates artificial low-bias produced by always zeroing bits
- Implemented as bit-mask, no floating-point arithmetic (or rounding) required
- Bit Grooming *guarantees* preserving the requested Number of Significant Digits (NSD):

Sign	Exponent	Fraction (significand)	Decimal	Notes
0	10000000	10010010000111111011011	3.14159265	Exact
0	10000000	10010010000111111011011	3.14159265	NSD = 8
0	10000000	10010010000111111011010	3.14159262	NSD = 7
0	10000000	10010010000111111011000	3.14159203	NSD = 6
0	10000000	10010010000111111000000	3.14158630	NSD = 5
0	10000000	10010010000111100000000	3.14154053	NSD = 4
0	10000000	10010010000000000000000	3.14062500	NSD = 3
0	10000000	10010010000000000000000	3.14062500	NSD = 2
0	10000000	10010000000000000000000	3.12500000	NSD = 1

- Compression achieved by losslessly compressing (with DEFLATE) Bit Groomed data
- More accurate, greater range, less compression than linear packing (netCDF default)
- Unlike all viable competitors, Bit Grooming *guarantees* specified precision for all data
- Preserves IEEE floating-point format—no special software required to read it



Impact

Implementation:

netCDF Operators 4.4.8+ produce Bit Groomed datasets with `-ppc` command:
`ncks -7 -ppc default=5 in.nc out.nc # 5 significant digits`
`ncks -7 -ppc p,w,z=5 -ppc q,RH=4 -ppc T,u,v=3 in.nc out.nc # 5, 4, 3`
`ncks -7 -ppc default=5#q,RH=4#T,u,v=3 in.nc out.nc # Same as above`

Try it, You'll Like It:

Bit Grooming can easily reduce ACME data storage volume by 60-70% without sacrificing scientific precision or burdening users. It preserves IEEE floating-point format—no special software or command is necessary to read Bit Groomed data.

Algorithm details and error analysis:

Zender, C. S. (2016), Bit Grooming: Statistically accurate precision-preserving quantization with compression, evaluated in the netCDF Operators (NCO, v4.4.8+), Geosci. Model Dev., 9, 3199-3211, doi:10.5194/gmd-9-3199-2016.

Intercomparison with other lossy compression algorithms:

Silver, J. D. and C. S. Zender (2016), Compression-error trade-off for large gridded datasets, in revision after review, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-177.

Software implementation:

<http://nco.sf.net/nco.html#ppc>