Bit-Grooming: Shave Your Bits with Razor-sharp Precision

Zender Charles¹ and Silver Jeremy²

¹Affiliation not available
²University of Melbourne

November 15, 2022

Abstract

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. We evaluated Bit Grooming against competitors Linear Packing, Layer Packing, and GRIB2/JPEG2000.
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. We evaluated Bit Grooming against competitors Linear Packing, Layer Packing, and GRIB2/JPEG2000.

Conclusions

How does Bit Grooming perform?
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 can have better compression ratios, yet Bit Grooming is the most accurate, usable, and portable.

Why don’t we Bit Groom 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. Users could reduce their long term storage costs, and show leadership in the elimination of false precision, by adopting Bit Grooming.

Implementation

netCDF Operators (NCO) produce Bit Groomed datasets with —ppc option:

```
ncks -7 —ppc default=5 in.nc out.nc # 5 sig. digits
ncks -7 —ppc p,w,z=5 —ppc q,RH=4 —ppc T,u,v=3 in.nc out.nc
ncks -7 —ppc default=5|q,RH=4|T,u,v=3 in.nc out.nc # Same
```

Bit-Groomed data is IEEE format, requires no special software to read!

References


Software documentation: https://nco.sf.net/nco.html#ppc

Support