xskillscore: an open source python package to verify forecasts

presented by **Ray Bell**

Wednesday, April 1, 2020 | 6:30 PM - 8:30 PM
The Idea Center @ MDC


xskillscore:
an open source python package
to verify forecasts
presented by Ray Bell
Coauthors: Aaron Spring, Andrew Huang and Riley Brady
Thanks to: Ram Narasimhan
Wednesday, April 1, 2020 | 6:30 PM - 8:30 PM
The Idea Center @ MDC

xskillscore in the wild

Matteo De Felice @matteodefelice · Oct 12, 2019

Can we use a #Python only workflow to calculate the skill scores on @CopernicusECMWF seasonal forecasts? Yes, and the whole process is *unbelievably* faster than a few years ago. The @GoogleColab notebook is accessible here: colab.research.google.com/drive/1wWHz_SM... 1/4

Google Colaboratory
colab.research.google.com

4 15 70

Show this thread

Matteo De Felice @matteodefelice · Oct 12, 2019

Replying to @matteodefelice

3. xskillscore (github.com/raybellwaves/x...) to calculate skill scores @RayBell_RC

4. xESMF (xesmf.readthedocs.io/en/latest/) to regrid the data @Jiawei_Zhuang_

5. cdsapi (pypi.org/project/cdsapi/) to access the CDS @BaudouinRaoult @alexamici @StephanSiemen

3/4

https://colab.research.google.com/drive/1wWHz_SMCHNuos5fxWRUJTcB6wqkTJQCR

Motivation for xskillscore

- Applying verification metrics to climate/weather prediction is at the heart of all model development

- Climate/weather data is big:

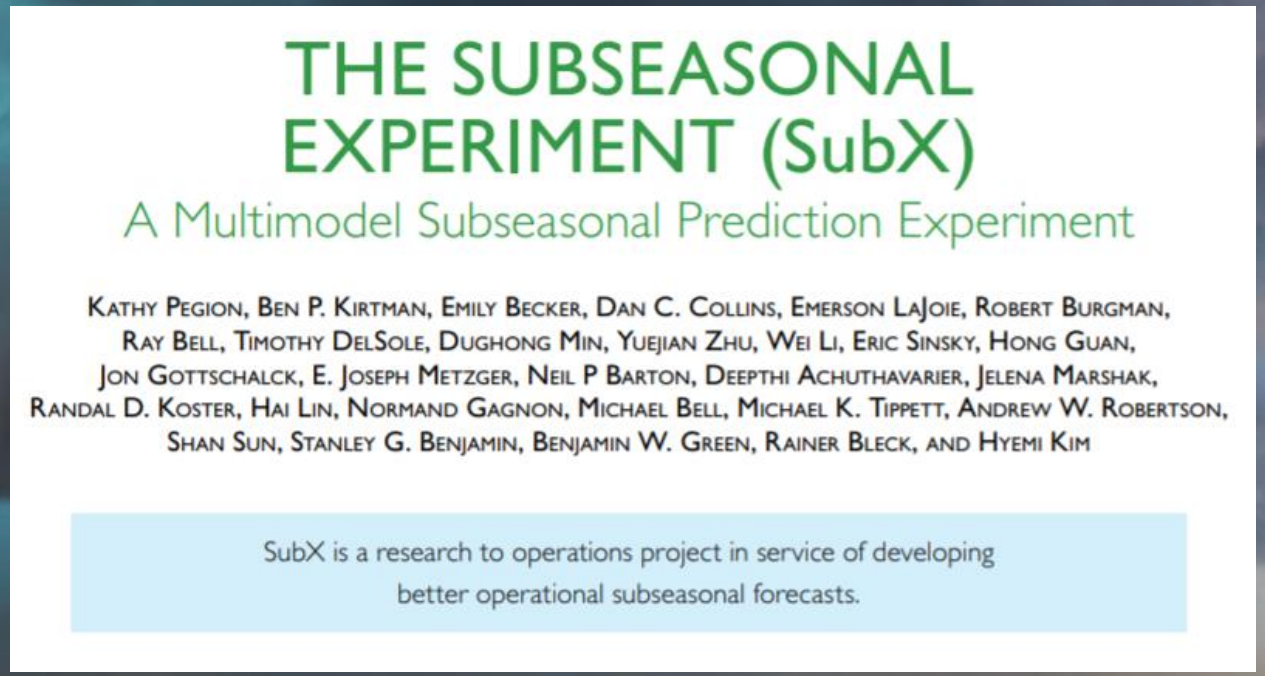
(member x lead time x height x longitude x latitude x time)

e.g. one model in SubX (there are many):

(10 x 45 x 4 x 360 x 181 x 7000) ~= 821,016,000,000

SubX data is around **~20 Tb**

- Apply verification metrics across multiple dimensions and avoid for loops!



THE SUBSEASONAL EXPERIMENT (SubX)

A Multimodel Subseasonal Prediction Experiment

Kathy Pegion, Ben P. Kirtman, Emily Becker, Dan C. Collins, Emerson LaJoie, Robert Burgman, Ray Bell, Timothy DelSole, Dughong Min, Yuejian Zhu, Wei Li, Eric Sinsky, Hong Guan, Jon Gottschalck, E. Joseph Metzger, Neil P Barton, Deepthi Achuthavarier, Jelena Marshak, Randal D. Koster, Hai Lin, Normand Gagnon, Michael Bell, Michael K. Tippett, Andrew W. Robertson, Shan Sun, Stanley G. Benjamin, Benjamin W. Green, Rainer Bleck, and Hyemi Kim

SubX is a research to operations project in service of developing better operational subseasonal forecasts.

https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-18-0270.1

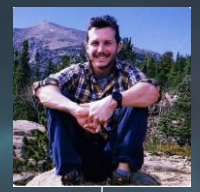
The beauty of open source

mse
1/4/2019

accessor
5/15/2019

probabilistic
8/30/2019

weights; skipna
10/17/2020



First commit
6/14/2018

climpred

Dependency of climpred
January 2019

Data science tutorial
4/1/2020

xskillscore overview

Deterministic metrics (correlation and distance)

- RMSE
- MAPE
- MAE
- R2
- Pearson R

Probabilistic metrics

- CRPS
- Brier score

Scales with dask

Extension of xarray