Journal of
open research software

Krakauer, N Y 2016 SeFo: A Package for Generating Probabilistic Forecasts from NMME Predictive Ensembles. *Journal of Open Research Software,* 4: e19, DOI: http://dx.doi.org/10.5334/jors.112

SOFTWARE METAPAPER

SeFo: A Package for Generating Probabilistic Forecasts from NMME Predictive Ensembles

Nir Y Krakauer

Department of Civil Engineering and NOAA-CREST, City College of New York, New York, USA 10031
nkrakauer@ccny.cuny.edu

Long-range weather forecasts based on output from ensembles of computer simulations are attracting increasing interest. A variety of methods have been proposed to convert the ensemble outputs to calibrated probabilistic forecasts. The package presented here (SeFo, for Seasonal Forecasting) implements a number of methods for producing forecasts of monthly surface air temperature anomalies up to 9 months in advance using output from the North American Multi-Model Ensemble (NMME). The package contains modules for downloading and reading past observations and ensemble output; producing forecast probability distributions; and verifying and calibrating a user-determined subset of methods using arbitrary past periods. By changing individual modules, the package could be extended to use other model ensembles, forecast other weather variables, or apply other forecast methods. SeFo is written in the numerical computing language Octave and is available on Bitbucket under the GNU General Public License (Version 3 or later).

Keywords: Probabilistic forecasting; Weather prediction; Model ensembles; Meteorology; Seasonal forecasting; GNU Octave

Funding Statement: The author gratefully acknowledges support from NOAA awards NA11SEC4810004, NA12OAR4310084, NA15OAR4310080; PSC-CUNY Award 68346-00 46; and CUNY CIRG Award 2207. All statements made are the views of the author and not the opinions of any funders or government agencies.

(1) Overview

Introduction

Long-range weather forecasts based on output from ensembles of computer simulations are attracting increasing interest as being useful for various weather-sensitive socioeconomic sectors, including agriculture, energy, and water management [1–5]. A variety of methods have been proposed to convert ensemble outputs to calibrated probabilistic forecasts of future meteorologic variables such as temperature and precipitation [6–19]. The workflow for applying, assessing and comparing the performance of different methods involves many steps: downloading historic observations and ensemble outputs; applying the selected methods to produce probabilistic forecasts; comparing the performance of the selected methods over a common past period using any of several possible metrics; and producing summary graphics. Until now, there has not been any publicly available implementation of this workflow, forcing different investigators and end users to write their own and limiting reproducibility of published results and the ability to isolate causes for differences between studies.

To address these problems, a new SeFo (Seasonal Forecasting) package was written, with functionality as given in the following section. This SeFo package builds on the Logocline package previously published by the author [20].

This paper is intended to serve as a general description of the philosophy and computational architecture of the package. Technical details on specific forecasting methods implemented and skill assessments of seasonal forecasting using this package will be published in appropriate peer-reviewed venues.

Implementation and architecture

SeFo is written as a package for the free numerical computing language and environment GNU Octave [21–22], exploiting built-in Octave functions and capabilities wherever possible. (Although Octave is deliberately designed to be able to run programs written for the proprietary environment MATLAB, no effort has been made to make SeFo usable in MATLAB.) The package has a modular architecture that isolates components of the workflow for probabilistic seasonal forecasts using model ensembles. Each module is an Octave function that writes its output to a file, from which it can be accessed by other modules. An input structure is used to hold options that need to be passed on to different modules, such as which data sets, dates, and forecast methods to use. (The list of possible options is given in a README text file.)

The core SeFo functions (modules) all have names starting with "sefo_", and their interrelationships are diagrammed in **Figure 1**. There are also a number of ancillary functions included in the package, for example for obtaining regression models used for particular forecast methods, regridding to the 1 degree grid used for the forecast ensemble output, and making maps. The capabilities of the core functions are as follows:

sefo_obs_read: Download and regrids observational data (currently surface air temperature, either from Berkeley Earth

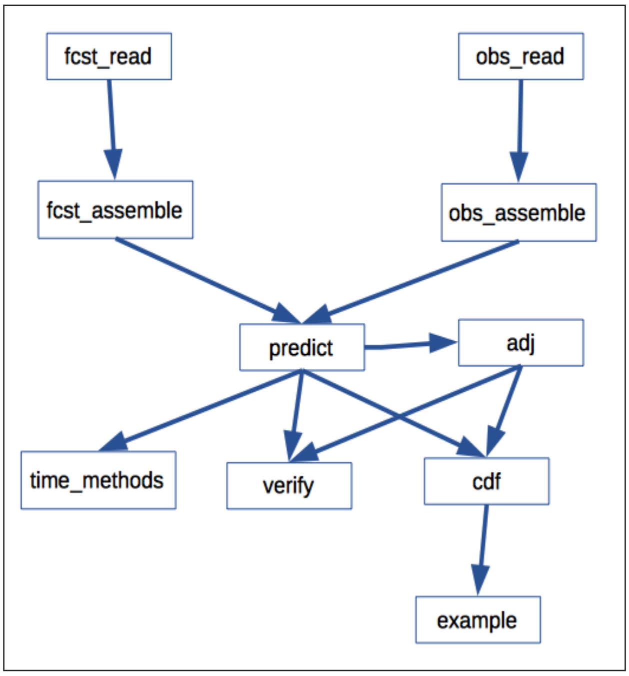

Figure 1: Calling dependencies between the core functions in sefo. Each has "sefo_" prefixed to its name.

Surface Temperature [23]; NCEP/NCAR reanalysis [24]; or Climate Prediction Center [25]) for a specified month.

sefo_obs_assemble: Collect the observations for a sequence of months.

sefo_fcst_read: Download and store ensemble forecasts from a given climate model and month (currently the data source is the North American Multi-Model Ensemble (NMME) Phase 1 [26] accessed via the IRI Data Library [27]).

sefo_fcst_assemble: Collect ensemble forecasts for a sequence of months.

sefo_predict: Apply one of several (currently 23) available prediction methods to estimate a probability distribution values for a given month from current ensemble predictions plus a set of past prediction-observation pairs. Currently the implemented forecast methods all return t distributions as the forecast probability distributions.

sefo_adj: Apply an optional calibrating adjustment to the forecast t distribution to better match the distribution of verifying observations over some specified past period.

sefo_cdf: Calculate, and optionally map, requested quantiles of a forecast probability distribution.

sefo_verify: Compare probabilistic forecasts for a past period against observations using several metrics, including forecast root mean square error, bias, mean negative log likelihood, and Kolmogorov-Smirnov statistic.

sefo_time_methods: Compare the computation times for selected forecast methods.

sefo_example: Exercise the key components of the package by generating a sample forecast for next month **(Figure 2)**.

Basic installation instructions are provided in the README file.

Quality control

Each of the core functions (all the functions with names beginning in sefo, except sefo_example) has a demonstration script that tests and illustrates its basic capabilities. "demo function_name" will run this script in Octave. Some of the ancillary functions have their own unit tests defined ("test function_name"). Development and testing was carried out in a Linux environment, specifically the Debian distribution (versions Jessie (Stable) and Unstable), and in Mac OS X with a Macports Octave installation.

Limitations and potential improvements

Currently, the package routines are not fully generalized. For example, NMME is currently the only supported source of ensemble predictions.

Documentation for the package and unit tests and demos for non-core functions are also not complete.

More information could be provided while the software is running, such as percentage progress of the downloads and data analysis.

Better input checking for the options structure could be provided with analogues of the odeset and odeget functions used for supplying parameters to Octave's differential equation solvers.

While the current data sources for SeFo, referenced above, are, to the author's knowledge, available without restrictions on use, abilities to handle and display different data licenses could potentially be added.

Once the functionality has been extended to more use cases and the documentation is more complete, it is envisioned that SeFo might be added to the Octave Forge repository, from which it might be accessed by a wider user base.

Users are encouraged to submit bugs and patches to the repository issue tracker on Bitbucket.

(2) Availability

Operating system

While in theory the package should run in any operating system for which Octave is available, including Windows, it has only been tested in Unix-like environments (Linux and Mac OS X).

Programming language

The package requires GNU Octave (Version 3.8 or newer) with the linear-algebra [28], nan [29], netcdf [30], and splines [31] packages installed.

Additional system requirements

An Internet connection is required to download observational data and numerical weather prediction model ensemble output. Data and intermediate files are stored locally, which will typically require one to several gigabytes of space, depending on the use case.

Dependencies

There are no dependencies beyond those for Octave with the indicated packages.

Software location

Archive (e.g. institutional repository, general repository)

Name: figshare

Persistent identifier: https://dx.doi.org/10.6084/m9.figshare.3114844.v1 (tarball of version 0.0.2)

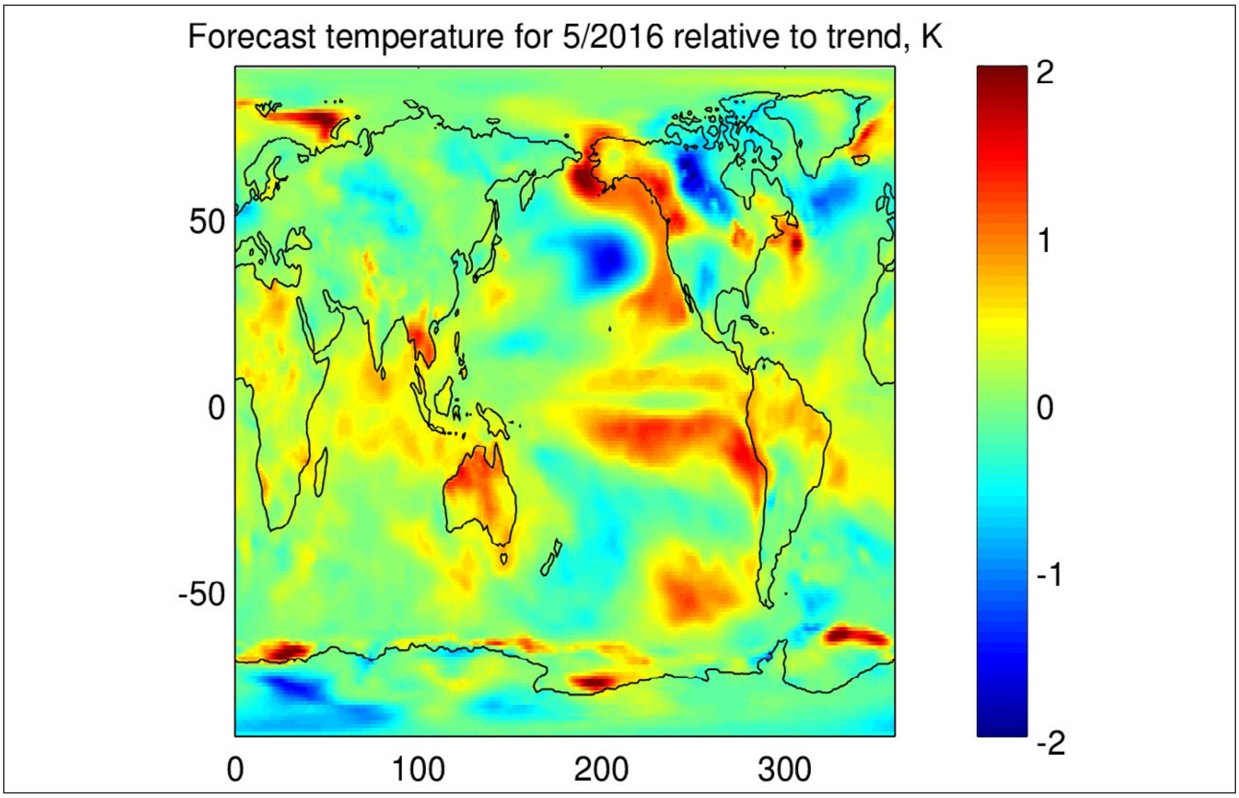

Figure 2: Example graphical output, generated with the sequence "predict_year = 2016; predict_month = 5; lag = 2; sefo_example" using version 0.0.2 of SeFo.

License: CC-BY
Publisher: NY Krakauer
Date published: 15/03/16

Code Repository (e.g. SourceForge, GitHub etc.)
Name: Bitbucket
Identifier: https://bitbucket.org/niryk/sefo
License: GPL V3+
Date published: 15/03/16

Language
Octave

(3) Reuse potential

Given the modular structure of SeFo, it could be extended with comparatively little additional work within the seasonal forecast context to accommodate alternative weather variables (such as precipitation or sunniness, although because these are farther than temperature from a normal distribution, some modification in the forecast methods would be advisable [32-33]), observation data sources, sources of ensemble outputs (besides NMME), and methods for generating forecasts from ensemble outputs and past observations. Many of the components of SeFo, including the specific forecast and verification methods implemented in the functions called by sefo_predict and sefo_verify, could also be reused for forecasting applications in fields outside of weather prediction.

Competing Interests
The author declares that they have no competing interests.

References

1. **National Research Council** Assessment of Intraseasonal to Interannual Climate Prediction and Predictability. Technical report, 2010. DOI: http://dx.doi.org/10.17226/12878
2. **Troccoli, A** 2010 Seasonal climate forecasting. *Meteorological Applications*, 17: 251–268. DOI: http://dx.doi.org/10.1002/met.184
3. **Alemu, E T, Palmer, R N, Polebitski, A** and **Meaker, B** 2011 Decision support system for optimizing reservoir operations using ensemble streamflow predictions. *J. Water Resour. Plng. and Mgmt.*, 137 (1): 72–82. DOI: http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000088
4. **Smith, D M, Scaife, A A** and **Kirtman, B P** 2012 What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environmental Research Letters*, 7 (1): 015602. DOI: http://dx.doi.org/10.1088/1748-9326/7/1/015602
5. **Doblas-Reyes, F J, García-Serrano, J, Lienert, F, Biescas, A P** and **Rodrigues, L R L** 2013 Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4 (4). DOI: http://dx.doi.org/10.1002/wcc.217
6. **Arribas, A, Glover, M, Maidens, A, Peterson, K, Gordon, M, MacLachlan, C, Graham, R, Fereday, D, Camp, J, Scaife, A A, Xavier, P, McLean, P, Colman, A** and **Cusack, S** 2011 The GloSea4 ensemble prediction system for seasonal forecasting. *Monthly Weather Review*, 139 (6): 1891–1910. DOI: http://dx.doi.org/10.1175/2010MWR3615.1

7. **Wang, Q J, Schepen, A** and **Robertson, D E** 2012 Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. *Journal of Climate*, 25 (12): 5524–5537. DOI: http://dx.doi.org/10.1175/JCLI-D-11-00386.1
8. **Peng, P, Kumar, A, Halpert, M S** and **Barnston, A G** 2012 An analysis of CPC's operational 0.5-month lead seasonal outlooks. *Weather and Forecasting*, 27 (4): 898–917, 2012. DOI: http://dx.doi.org/10.1175/WAF-D-11-00143.1
9. **Pavan, V** and **Doblas-Reyes, F J** 2013 Calibrated multi-model ensemble summer temperature predictions over Italy. *Climate Dynamics*, 41 (11): 2115–2132. DOI: http://dx.doi.org/10.1007/s00382-013-1869-7
10. **Krakauer, N Y, Grossberg, M D, Gladkova, I** and **Aizenman, H** 2013 Information content of seasonal forecasts in a changing climate. *Advances in Meteorology*, 2013: 480210. DOI: http://dx.doi.org/10.1155/2013/480210
11. **Hawthorne, S, Wang, Q J, Schepen, A** and **Robertson, D** 2013 Effective use of general circulation model outputs for forecasting monthly rainfalls to long lead times. *Water Resources Research*. DOI: http://dx.doi.org/10.1002/wrcr.20453
12. **Dutton, J A, James, R P** and **Ross, J D** 2013 Calibration and combination of dynamical seasonal forecasts to enhance the value of predicted probabilities for managing risk. *Climate Dynamics*, 40 (11–12): 3089–3106. DOI: http://dx.doi.org/10.1007/s00382-013-1764-2
13. **DelSole, T, Jia, L** and **Tippett, M K** 2013 Scale-selective ridge regression for multimodel forecasting. *Journal of Climate*, 26 (20): 7957–7965. DOI: http://dx.doi.org/10.1175/JCLI-D-13-00030.1
14. **Yatagai, A, Krishnamurti, T N, Kumar, V, Mishra, A K** and **Simon, A** 2014 Use of APHRODITE rain gauge–based precipitation and TRMM 3B43 products for improving Asian monsoon seasonal precipitation forecasts by the superensemble method. *Journal of Climate*, 27 (3): 1062–1069. DOI: http://dx.doi.org/10.1175/JCLI-D-13-00332.1
15. **Wilks, D S** 2014 Comparison of probabilistic statistical forecast and trend adjustment methods for North American seasonal temperatures. *Journal of Applied Meteorology and Climatology*, 53 (4): 935–949. DOI: http://dx.doi.org/10.1175/JAMC-D-13-0294.1
16. **Mwangi, E, Wetterhall, F, Dutra, E, Di Giuseppe, F** and **Pappenberger, F** 2014 Forecasting droughts in East Africa. *Hydrology and Earth System Sciences,* 18 (2): 611–620, DOI: http://dx.doi.org/10.5194/hess-18-611-2014
17. **Mo, K C** and **Lettenmaier, D P** 2014 Hydrologic prediction over the Conterminous United States using the National Multi-Model Ensemble. *Journal of Hydrometeorology*, 15 (4): 1457–1472. DOI: http://dx.doi.org/10.1175/JHM-D-13-0197.1
18. **Kumar, S, Dirmeyer, P A** and **Kinter, J L III** 2014 Usefulness of ensemble forecasts from NCEP Climate Forecast System in sub-seasonal to intra-annual forecasting. *Geophysical Research Letters,* 41 (10): 3586–3593. DOI: http://dx.doi.org/ 10.1002/2014GL059586
19. **Dutra, E, Wetterhall, F, Di Giuseppe, F, Naumann, G, Barbosa, P, Vogt, J, Pozzi, W** and **Pappenberger, F** 2014 Global meteorological drought – Part 1: Probabilistic monitoring. *Hydrology and Earth System Sciences,* 18 (7): 2657–2667. DOI: http://dx.doi.org/10.5194/hess-18-2657-2014
20. **Krakauer, N Y** and **Devineni, N** 2015 Up-to-date probabilistic temperature climatologies. *Environmental Research Letters*, 10 (2): 024014, DOI: http://dx.doi.org/10.1088/1748-9326/10/2/024014
21. **Eaton, J W, Bateman, D** and **Hauberg, S** 2008 *GNU Octave Manual Version 3.* Network Theory Limited. URL http://www.network-theory.co.uk/octave/manual/.
22. **Eaton, J W** 2012 GNU Octave and reproducible research. *Journal of Process Control,* 22 (8): 1433–1438. DOI: http://dx.doi.org/10.1016/j.jprocont.2012.04.006
23. **Rohde, R, Muller, R, Jacobsen, R, Perlmutter, S, Rosenfeld, A, Wurtele, J, Curry, J, Wickham, C** and **Mosher, S** 2013 Berkeley Earth temperature averaging process. *Geoinformatics and Geostatistics: An Overview,* 1 (2): 1000103. DOI: http://dx.doi.org/10.4172/2327-4581.1000103
24. **Kalnay, E, Kanamitsu, M, Kistler, R, Collins, W, Deaven, D, Gandin, L, Iredell, M, Saha, S, White, G, Woollen, J, Zhu, Y, Leetmaa, A, Reynolds, R, Chelliah, M, Ebisuzaki, W, Higgins, W, Janowiak, J, Mo, K C, Ropelewski, C, Wang, J, Jenne, R** and **Joseph, D** 1996 The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society,* 77 (3): 437–471. DOI: http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2
25. **Fan, Y** and **van den Dool, H** 2008 A global monthly land surface air temperature analysis for 1948-present. *Journal of Geophysical Research.* DOI: http://dx.doi.org/10.1029/2007JD008470
26. **Kirtman, B P, Min, D, Infanti, J, M, Kinter, J L, Paolino, D A, Zhang, Q, van den Dool, H, Saha, S, Mendez, M P, Becker, E, Peng, P, Tripp, P, Huang, J, DeWitt, D G, Tippett, M K, Barnston, A G, Li, S, Rosati, A, Schubert, S D, Rienecker, M, Suarez, M, Li, Z E, Marshak, J, Lim, Y-K, Tribbia, J, Pegion, K, Merryfield, W J, Denis, B** and **Wood, E F** 2014 The North American Multi-Model Ensemble (NMME): Phase-1 Seasonal to Interannual Prediction, Phase-2 Toward Developing Intra-Seasonal Prediction. *Bulletin of the American Meteorological Society,* 95: 585–601. DOI: http://dx.doi.org/10.1175/BAMS-D-12-00050.1
27. **Blumenthal, M, Benno, B, Michael, C, John d, Cousin, R** and **Khomyakov, I** 2014 IRI Data Library: enhancing accessibility of climate knowledge. *Earth Perspectives,* 1(1): 1-12. DOI: http://dx.doi.org/10.1186/2194-6434-1-19
28. **Linear Algebra** 2015 Package for GNU Octave. URL http://octave.sourceforge.net/linear-algebra.
29. **Schloegl, A** 2015 NaN Package for GNU Octave. URL http://octave.sourceforge.net/nan.
30. **Barth, A** 2015 NetCDF Package for GNU Octave. URL http://octave.sourceforge.net/netcdf.
31. **Krakauer, N Y** et al. Splines Package for GNU Octave, 2015. URL http://octave.sourceforge.net/splines.

32. **Krakauer, N Y, Pradhanang, S M, Panthi, J, Lakhankar, T** and **Jha, A K** 2015b Probabilistic precipitation estimation with a satellite product. *Climate*, 3 (2): 329–348. DOI: http://dx.doi.org/10.3390/cli3020329

33. **Aizenman, H, Grossberg, M D, Krakauer, N Y and Gladkova, I** 2016 Ensemble forecasts: probabilistic seasonal forecasts based on a model ensemble Climate. 4(2): 19. DOI: http://dx.doi.org/10.3390/cli4020019

How to cite this article: Krakauer, N Y 2016 SeFo: A Package for Generating Probabilistic Forecasts from NMME Predictive Ensembles. *Journal of Open Research Software,* 4: e19, DOI: http://dx.doi.org/10.5334/jors.112

Submitted: 13 January 2016 **Accepted:** 04 May 2016 **Published:** 20 May 2016

]u[*Journal of Open Research Software* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS