A validation suite for downscaled climate model data

Elisabeth Orskaug[1,2]; Ola Haug[1,2]; Ida Scheel[3,2] and Arnoldo Frigessi[4,2,1]

1) BACKGROUND

MOTIVATION:

- Intensified climate research produces an increasing number of data sets combining different global circulation models, CO_2 emission scenarios and downscaling techniques.
- For impact studies, but also as an issue of separate interest, the quality of these data need to be verified.
- Hence, there is an apparent demand for validation of past and present climate projections against real observations at different spatial scales.

AIM:

Identify distributional discrepancies in ERA40 re-analysis data as compared to interpolated observation data on a 25x25km² grid nationwide.

- Point out global and local differences in the distributions
- Assess their seriousness by appropriate local measures
- Formally: Test H_0: "$f_X = f_Y$" against H_{alt}: "$f_X \neq f_Y$" (literally or properties thereof)
- Concern is equally much on a modest FNR (leaving in true discrepancies) as on keeping the FDR low (leaving out true similarities)

DATA:

40 years of daily precipitation data (1961-2000) organized into 777 25x25km² grid cells covering mainland Norway.

<u>**ERA40 re-analysis data (dynamically downscaled, ENSEMBLES):**</u>

- Day to day correlation exhibited by GCM ERA40 data with observations partly lost in the downscaling process
- Reliant on the downscaling, still supposed to possess properties similar to real weather locally over longer time periods

<u>**Observation data:**</u>

- Interpolations (1x1 km²) from a triangulation of the official measurement stations operated by the Norwegian Meteorological Institute
- Aggregated to 25 x 25 km² scale by collecting 1x1 km² grid cells with centre points within the ERA40 cell, taking their mean as a representation of the precipitation inside that grid cell.

PROSPECTS:

From the identification of distributional discrepancies in the model data, develop locally supported transfer functions that can bring downscaled climate model data closer to observational truth.

PARTNERS:

- Norwegian Meteorological Institute (Ole Einar Tveito, Jan Erik Haugen, Eirik Førland)
- Peter Guttorp (University of Washington and Norwegian Computing Center)

CONTACT PERSON:

- Elisabeth Orskaug, elisabeth.orskaug@nr.no

2) COMPARISONS (SEASONAL)

GLOBAL MEASURE:

Kolmogorov Smirnov test

LOCAL MEASURES:

Measure	Test	Data
Mean	t-test	All data
Standard deviation	t-test	*
Quantiles (q05,q10,q25,q50,q75,q90,q95)	Fisher exact test	All data
Wet day frequency	t-test	*
Largest 5 day precipitation total	t-test	*
Maximum number of consecutive dry days	t-test	*
GPD (via tail parameter and return period)	t-test	**

* The original data set is divided into years (and seasons). Each measure is calculated on annual data, producing a total of 40 values for each season. T-tests are then performed on those 40 values using CLT.

** The standard deviation of the return period, x_T, is calculated from 1000 simulations. Since x_T is a function of the GPD parameters estimated by ML, the return level is approximately normal.

Return period of GPD: *Brabson, B.B. and Palutikof, J.P. (2000): Tests of the generalized Pareto distribution for predicting extreme wind speeds. Journal of Appl. Meteorol. vol 39, no 9, 1627-1640.*

COMBINED TEST:

Sectionwise pdf test, see box 5

3) SIGNIFICANCE LEVEL α OF TESTS

Type II errors is most dangerous. The table shows FNR and FDR values , and number of rejected grid cells for different α for the Kolmogorov Smirnov test in season autumn.

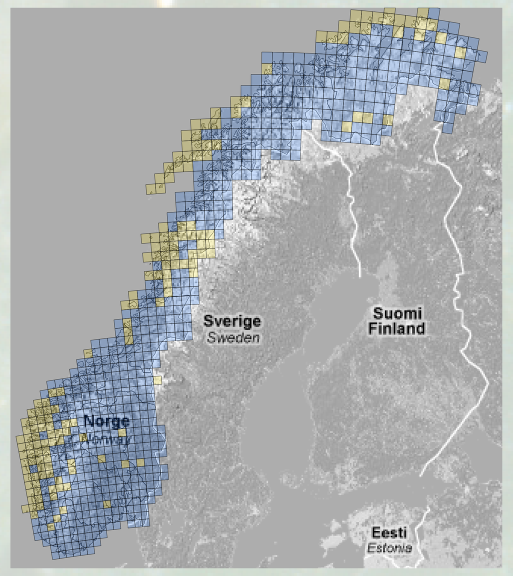
Critical value (α)	FNR	FDR	Rejections
0.0001	0.41	4.37e-6	449
0.001	0.33	4.38e-5	515
0.01	0.19	0.00044	609
0.05	0.10	0.0023	673
0.10	0.062	0.0045	701
0.15	0.046	0.0069	717
0.20	0.034	0.0090	730
0.30	0.012	0.013	743
0.50	0.000	0.023	759

The critical value is chosen such that FNR is less than 5%. Still, we want FDR to be as low as possible.

Storey estimator is considered: *Zehetmayer, S. and Posch, M. (2010): Post hoc power estimation in large-scale multiple testing problems. Journal of Bioinformatics vol. 26, no. 8, 1050-1056.*

4) GRAPHICS FOR SELECTED TESTS (SEASON: AUTUMN)

Kolmogorov Smirnov test



Hypothesis:

H_0: $F_{ERA40} = F_{OBS}$

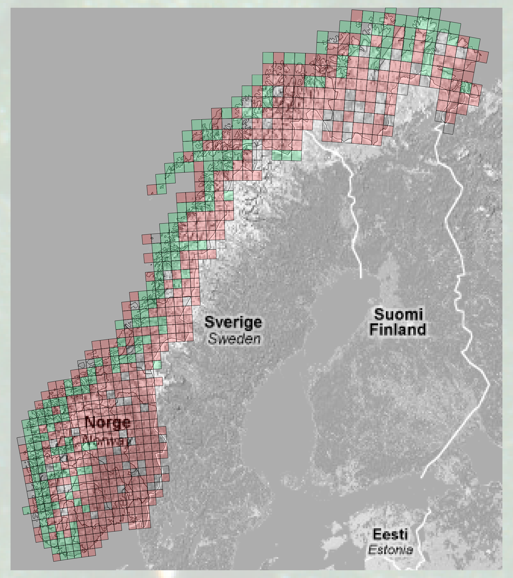
H_{alt}: $F_{ERA40} \neq F_{OBS}$

Colours on map:

<u>Yellow:</u> Keep H_0

<u>Blue:</u> Reject H_0

Test of median



Hypothesis:

H_0: $q50_{ERA40} = q50_{OBS}$

H_{alt}: $q50_{ERA40} \neq q50_{OBS}$

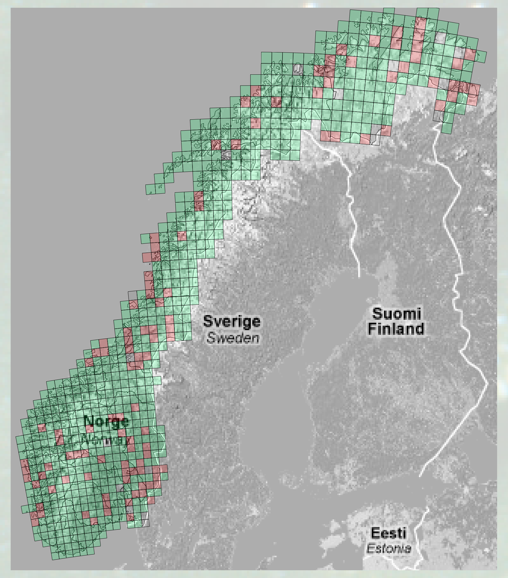
Colours on map:

<u>Clear:</u> Keep H_0

<u>Green:</u> Reject H_0, and $q50_{ERA40} < q50_{OBS}$

<u>Red:</u> Reject H_0, and $q50_{ERA40} > q50_{OBS}$

Test of return level in GPD (u=q95, T=1 year)



Hypothesis:

H_0: $x_{T,\,ERA40} = x_{T,\,OBS}$

H_{alt}: $x_{T,\,ERA40} \neq x_{T,\,OBS}$

Colours on map:

<u>Clear:</u> Keep H_0

<u>Green:</u> Reject H_0, and $x_{T,\,ERA40} < x_{T,\,OBS}$

<u>Red:</u> Reject H_0, and $x_{T,\,ERA40} > x_{T,\,OBS}$

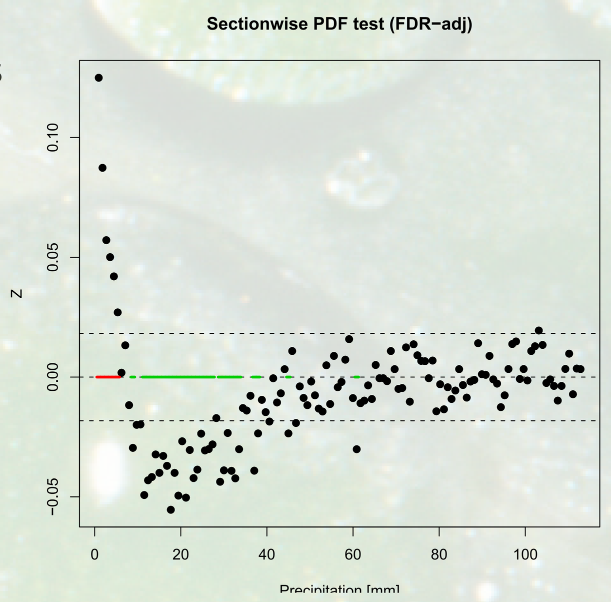
5) SECTIONWISE PDF TEST:

<u>Shortcomings of global tests:</u>

- Do not tell which part(s) of the distribution that differs

<u>Shortcomings of local tests considered so far:</u>

- Focus one single pdf property at the time



Alternative idea that performs section-wise testing on the probability density function (ongoing work by Glad and Mohammed, University of Oslo):

i) Test statistics, including variable transformation:

- Data sets: Let X = ERA40 and Y = OBS
- Divide the range of the data set $X \cup Y$ into a suitable number of bins, T (Friedman & Diaconis)
- Count the number of observations in each bin, n_{Xi} and n_{Yi}
- After adding random Poisson(10) counts to account for a heavy tail, root-transform the modified counts and compute differences $Z_i = \tilde{n}_{Xi} - \tilde{n}_{Yi}$
- Now, $Z_i \overset{appr}{\sim} N(\mu_{zi}, \sigma^2_{zi})$ with σ^2_{zi} fixed and known. Z_i can be modeled as $Z_i = \mu_{zi} + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2_{\varepsilon i})$ and i = 1, …, T

ii) Testing H_0: $f_X = f_y$ vs H_{alt}: $f_X \neq f_Y$ amounts to identifying non-zero elements of the vector $\boldsymbol{\mu_z}$:

- Test separately each bin (two-sided test of normality for normalized Z_i), and FDR-correct by a Benjamini-Hochberg approach

NR Norsk Regnesentral NORWEGIAN COMPUTING CENTER · UNIVERSITY OF OSLO · Meteorologisk institutt · (sfi)² Statistics for Innovation

1: Norwegian Computing Center, Oslo, Norway; 2: Statistics for Innovation, (sfi)²; 3: Department of Mathematics, University of Oslo, Oslo, Norway; 4: Department of Biostatistics, University of Oslo, Oslo, Norway