A blind test of monthly homogenisation algorithms

http://www.clim-past-discuss.net/7/2655/2011/cpd-7-2655-2011-discussion.html

Content

- Benchmark dataset
- Results homogenisation benchmark
  - RMSE
  - Errors trends
  - Contingency scores
- Discussion
  - Homogenisation
  - Benchmarking
  - Our common future
Intercomparison study

- Compare full homogenisation algorithms
  - Detection, correction
  - If applicable: reference, iterations, remove outliers, etc.

- Benchmark dataset
  - Monthly temperature and precipitation networks
  - Homogenized blind
  - Random small inhomogeneities (Gaussian distribution)

- Sections
  - Real (inhomogeneous) climate records
  - Synthetic data
  - Surrogate data
Surrogate temperature section

- Generated homogeneous surrogate temperature networks
  - Based on statistical properties of homogenized data
  - 15 networks
  - Length: 100 years
  - Number of stations: 5, 9, 15
### WWII

**Missing data**

#### True deviations & true and detected inhomogeneities for network: inho temp sur1 000004

<table>
<thead>
<tr>
<th>Location</th>
<th>Y-axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NANTES-BOUGUENAISS</td>
<td>-2</td>
</tr>
<tr>
<td>SAINT-NAZARE</td>
<td>-1</td>
</tr>
<tr>
<td>SAINT-GEORGES-DES-GARDES</td>
<td>0</td>
</tr>
<tr>
<td>LAVAL-ENTRAMMES</td>
<td>1</td>
</tr>
<tr>
<td>EVRON</td>
<td>2</td>
</tr>
<tr>
<td>PRE-EN-PAIL</td>
<td>3</td>
</tr>
<tr>
<td>GROIX</td>
<td>4</td>
</tr>
<tr>
<td>CHARTRES</td>
<td>5</td>
</tr>
<tr>
<td>RENNES</td>
<td>6</td>
</tr>
<tr>
<td>DIEVRES</td>
<td>7</td>
</tr>
<tr>
<td>ST-CORNIER-DES-LANDES</td>
<td>8</td>
</tr>
<tr>
<td>LE MANS</td>
<td>9</td>
</tr>
<tr>
<td>ILE-YEU</td>
<td>10</td>
</tr>
<tr>
<td>LA-MOTHE-AHARD</td>
<td>11</td>
</tr>
<tr>
<td>BIARD</td>
<td>12</td>
</tr>
</tbody>
</table>

Red cross: detected outlier (y=0); red plus: true outlier

Detected inhomogeneities: drawn vertical line; true inhomogeneities: lighter stripped line

Black line: break; Green line: begin trend; Red line: end trend
Global trend

- Stochastic signal, Fourier filtered noise
- $E(k) \sim k^{-4}$
- Standard deviation (max-min) = 1 °C
Participant returned the data

- 25 blind contributions
- Some algorithms multiple contributions
  - Test versions
  - Test influence operator (manual methods)
- Algorithms/software
  - USHCN
  - PRODIGE
  - MASH
  - Craddock
  - AnClim
  - RhTestV2
  - SNHT
  - Climatol
  - ACMANT
Scatterplots monthly CRMSE
Monthly CRMSE complete contributions

![Graph showing monthly CRMSE contributions for temperature and precipitation.](image_url)
Decadal CRMSE complete contributions

Temperature

Inhom. data
MASH main
PRODIGE main
PRODIGE monthly
PRODIGE trendy
USHCN main
USHCN 52x
USHCN cx8
AnClim main
PMTred rel
PMFred abs
C3SNHT
SNHT DWD
ACMANT

CRMSE [°C]

Precipitation

Inhom. data
MASH main
PRODIGE main
PRODIGE monthly
PRODIGE trendy
AnClim main
PMTred rel
PMFred abs
C3SNHT
Climatol

CRMSE [mm]
Temporal behaviour CMRSE

- Monthly
- Yearly
- Decadal

CRMSE temperature [°C]

CRMSE precipitation [mm]

Time [a]

1900 1920 1940 1960 1980 2000
Errors in trends

![Graph showing errors in trends for different datasets including ACMANT, iCraddock Vertacnik, AnClim main, MASH main, PRODIGE main, and SNHT DWD. Each dataset displays trends in temperature and precipitation over time.]
Errors in trends

Trend difference [°C/100a]

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
ACMANT
SNHT DWD
C3SNHT
PMFred abs
PMTred rel
AnClim main
USHCN cx8
USHCN 52x
USHCN main
MASH main
Inhom. data
PRODIGE trendy
PRODIGE monthly
PRODIGE main

Trend difference [mm/100a]

-50 -40 -30 -20 -10 0 10 20 30 40 50
Climatol
C3SNHT
SNHT DWD
ACMANT
Scatterplot POD vs POFD

Probability of Detection vs Probability of False Detection
<table>
<thead>
<tr>
<th>Contribution</th>
<th>No stations</th>
<th>POD</th>
<th>POFD Skill</th>
<th>Pierce Skill</th>
<th>Heidke Skill</th>
<th>Heidke Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASH main</td>
<td>111</td>
<td>0.63</td>
<td>0.09</td>
<td>0.53</td>
<td>0.31</td>
<td>-0.20</td>
</tr>
<tr>
<td>PRODIGE main</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.33</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>PRODIGE monthly</td>
<td>111</td>
<td>0.39</td>
<td>0.02</td>
<td>0.37</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>PRODIGE trendy</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.32</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>USHCN main</td>
<td>111</td>
<td>0.34</td>
<td>0.00</td>
<td>0.33</td>
<td>0.46</td>
<td>0.61</td>
</tr>
<tr>
<td>USHCN 52x</td>
<td>111</td>
<td>0.40</td>
<td>0.01</td>
<td>0.39</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>USHCN cx8</td>
<td>111</td>
<td>0.35</td>
<td>0.01</td>
<td>0.35</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>AnClim main</td>
<td>111</td>
<td>0.18</td>
<td>0.03</td>
<td>0.15</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>iCraddock Vertacnik</td>
<td>55</td>
<td>0.60</td>
<td>0.03</td>
<td>0.57</td>
<td>0.54</td>
<td>0.49</td>
</tr>
<tr>
<td>PMTred rel</td>
<td>111</td>
<td>0.41</td>
<td>0.04</td>
<td>0.37</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>PMFred abs</td>
<td>111</td>
<td>0.21</td>
<td>0.01</td>
<td>0.20</td>
<td>0.27</td>
<td>0.46</td>
</tr>
<tr>
<td>C3SNHT</td>
<td>111</td>
<td>0.23</td>
<td>0.05</td>
<td>0.18</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>SNHT DWD</td>
<td>111</td>
<td>0.12</td>
<td>0.01</td>
<td>0.11</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>Climatol</td>
<td>111</td>
<td>0.38</td>
<td>0.01</td>
<td>0.37</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>ACMANT</td>
<td>111</td>
<td>0.50</td>
<td>0.03</td>
<td>0.47</td>
<td>0.44</td>
<td>0.41</td>
</tr>
</tbody>
</table>
## Contingency scores

<table>
<thead>
<tr>
<th>Contribution</th>
<th>No stations</th>
<th>POD</th>
<th>POFD Skill Score</th>
<th>Pierce Skill Score</th>
<th>Heidke Skill Score</th>
<th>Heidke Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASH main</td>
<td>111</td>
<td>0.63</td>
<td>0.09</td>
<td>0.53</td>
<td>0.31</td>
<td>-0.20</td>
</tr>
<tr>
<td>PRODIGE main</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.33</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>PRODIGE monthly</td>
<td>111</td>
<td>0.39</td>
<td>0.02</td>
<td>0.37</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>PRODIGE trendy</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.32</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>USHCN main</td>
<td>111</td>
<td>0.34</td>
<td>0.00</td>
<td>0.33</td>
<td>0.46</td>
<td>0.61</td>
</tr>
<tr>
<td>USHCN 52x</td>
<td>111</td>
<td>0.40</td>
<td>0.01</td>
<td>0.39</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>USHCN cx8</td>
<td>111</td>
<td>0.35</td>
<td>0.01</td>
<td>0.35</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>AnClim main</td>
<td>111</td>
<td>0.18</td>
<td>0.03</td>
<td>0.15</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>iCraddock Vertacnik</td>
<td>55</td>
<td>0.60</td>
<td>0.03</td>
<td>0.57</td>
<td>0.54</td>
<td>0.49</td>
</tr>
<tr>
<td>PMTred rel</td>
<td>111</td>
<td>0.41</td>
<td>0.04</td>
<td>0.37</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>PMFred abs</td>
<td>111</td>
<td>0.21</td>
<td>0.01</td>
<td>0.20</td>
<td>0.27</td>
<td>0.46</td>
</tr>
<tr>
<td>C3SNHT</td>
<td>111</td>
<td>0.23</td>
<td>0.05</td>
<td>0.18</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>SNHT DWD</td>
<td>111</td>
<td>0.12</td>
<td>0.01</td>
<td>0.11</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>Climatol</td>
<td>111</td>
<td>0.38</td>
<td>0.01</td>
<td>0.37</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>ACMANT</td>
<td>111</td>
<td>0.50</td>
<td>0.03</td>
<td>0.47</td>
<td>0.44</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Contingency scores

<table>
<thead>
<tr>
<th>Contribution</th>
<th>No stations</th>
<th>POD</th>
<th>POFD Skill Score</th>
<th>Pierce Skill Score</th>
<th>Heidke Skill Score</th>
<th>Heidke Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASH main</td>
<td>111</td>
<td>0.63</td>
<td>0.09</td>
<td>0.53</td>
<td>0.31</td>
<td>-0.20</td>
</tr>
<tr>
<td>PRODIGE main</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.33</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>PRODIGE monthly</td>
<td>111</td>
<td>0.39</td>
<td>0.02</td>
<td>0.37</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>PRODIGE trendy</td>
<td>111</td>
<td>0.35</td>
<td>0.02</td>
<td>0.32</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>USHCN main</td>
<td>111</td>
<td>0.34</td>
<td>0.00</td>
<td>0.33</td>
<td>0.46</td>
<td>0.61</td>
</tr>
<tr>
<td>USHCN 52x</td>
<td>111</td>
<td>0.40</td>
<td>0.01</td>
<td>0.39</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>USHCN cx8</td>
<td>111</td>
<td>0.35</td>
<td>0.01</td>
<td>0.35</td>
<td>0.47</td>
<td>0.61</td>
</tr>
<tr>
<td>AnClim main</td>
<td>111</td>
<td>0.18</td>
<td>0.03</td>
<td>0.15</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>iCraddock Vertacnik</td>
<td>55</td>
<td>0.60</td>
<td>0.03</td>
<td>0.57</td>
<td>0.54</td>
<td>0.49</td>
</tr>
<tr>
<td>PMTred rel</td>
<td>111</td>
<td>0.41</td>
<td>0.04</td>
<td>0.37</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>PMFred abs</td>
<td>111</td>
<td>0.21</td>
<td>0.01</td>
<td>0.20</td>
<td>0.27</td>
<td>0.46</td>
</tr>
<tr>
<td>C3SNHT</td>
<td>111</td>
<td>0.23</td>
<td>0.05</td>
<td>0.18</td>
<td>0.16</td>
<td>0.04</td>
</tr>
<tr>
<td>SNHT DWD</td>
<td>111</td>
<td>0.12</td>
<td>0.01</td>
<td>0.11</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>Climatol</td>
<td>111</td>
<td>0.38</td>
<td>0.01</td>
<td>0.37</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>ACMANT</td>
<td>111</td>
<td>0.50</td>
<td>0.03</td>
<td>0.47</td>
<td>0.44</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Heidke skill score – normal vs special
Temporal behaviour contingency scores
Temporal behaviour contingency scores - PMTred
Fuzzy (5yr) POD vs POD
Conclusions - Homogenisation

- **Best algorithms**
  - Function with an inhomogeneous reference (multiple breaks)
  - (ACMANT), Craddock, MASH, PRODIGE, USHCN
  - Automatic algorithms among the best

- **Relative homogenisation improves temperature records**
  - Absolute homogenisation can make data more inhomogeneous

- **The coder is the best operator**
  - Training operators is important

- **Work on handling missing data**

- **Correction methods: statistical model selection**
  - Year, season, month
  - Correlations with reference, length HSP
Conclusions - Benchmarking

- Moderate correlation between error metrics
  - Contingency scores not good predictor of skill
- Benchmark with multiple networks is needed to obtain reliable results
- Surrogate data is more difficult as Gaussian white noise
- Non-blind Validation studies also valuable
  - Recompute in case of errors (optimal result)
  - Sensitivity studies
    - Change settings, parameters homogenisation algorithms
    - Use unrealistic data (too many breaks, (almost) no breaks, etc.)
- Understanding problem, not just numerics
  - Mathematical basis
Technical recommendations benchmarking

- Realism is important for benchmarking
  - Competitive character
  - Accurate number for remaining errors in real networks
- Outliers
  - No clear influence
  - No outliers or realistic ones
- Correlations in perturbations within network
- Random missing data
- Study seasonal cycle
- Study frequency, size of local trends
- Remove decadal variability in difference series explicitly
- Model trend and natural variability separately, realistically
  - Validate absolute homogenisation
- Networks in real data section also in surrogate section
  - Real data section is important to validate the benchmark
  - Same cross-correlations
  - (Missing) data pattern
Future benchmarking exercise I

- Please analyse the current benchmark
  - Still lots of interesting questions remaining
  - Many more interesting validation metrics
  - Mixing detection and correction
  - Results for local trends (UHI, VGE)

- Surface temperature initiative (STI)
  - Temperature globally

- Similar benchmark or project proposal
  - Volunteer?
Future benchmarking exercise II

- New project: write proposal
  - Innovative

- Elements
  - Precipitation (room for improvement)
  - Humidity, pressure

- Other regions
  - Need to study properties inhomogenities globally (STI)

- Daily benchmark
  - Can we do this already?
  - Temperature, precipitation (?)

- Provide metadata
  - Partial and including wrong entries
  - Work on automatic use of metadata

- Participants
  - Survey?
Email distribution list

- Do we want to keep contact after the Action?
- homogenisation@listserv.uni-bonn.de
- Propose
  - All HOME people
  - Please inform your colleagues