Homogenization of monthly long-term temperature series of mainland Norway

Seminar for Homogenization and Quality Control in Climatological Databases, Budapest 24-28 October 2011 by Lars Andresen
Software package

- **Petr Štěpánek**
  - **LoadData**
    - Data from KDVH - Norwegian Climate Data Warehouse (dbf-files)
  - **AnClim**
    - Homogeneity analysis (using txt-files)
  - **ProClimDB**
    - Automating the homogenization procedure (using mainly dbf-files)
Homogenization tools (Windows)

ProClimDB

Processing window (profile: Default)

Menu: Adjust
Adjusts series according to reference series

Item: Adjust - from Reference Series (1)
Adjusts series according to existing reference series

Source files:
Ref Info file: Oslo11_all_recd_adj_info_9-8.dbf
Data file: Oslo11_all_recd_adj_series_10-3_S1.dbf
(Refer. Series)
(Inhomogeneities)

Destination files:
(Output Adj. Data)
Adj Info File:
(Adjust - Control)
(Inhom. corr.)

Settings
Vals. Transformation (precip, No. of years around): 10
Min. years from end: 4

Options
- One Decimal place
- Process without halt
- Add Statistics info

Conditions
- Smooth monthly ADJ
- No ADJUST below Cor+
- Cor+ value: 0.05

Input
Output
Batch-job
Conditions
Homogeneity methods

• Different methods available in AnClim (and ProClimDB)

• Chosen method for this study: SNHT
Reconstruction of series

- Series of stations less than 10 km away
- Maximum altitude difference = 150 m
- Reconstitution min length = 30 years
  - Shorter series may contribute
- Max gap = 4 years
- Within the same region

- Totally 95 out of 231 series are reconstructed
The Oslo-series (1837-2011)

- Oslo I and II
- Oslo I
- Oslo II
- Oslo-Blindern

1837 - 1937 - 1933 - 2011
Temperature series running in 2007

Length ≥ 30 years

Length ≥ 100 years
Homogenization tools (ProClimDB)
Homogenization tools (ProClimDB)

**MON 8-2**
- **Settings**
  - Create Info File only
  - Number of Stations: 5
  - Limit - correlation (; dist.): 0.8-0.1
  - Maximum altitude diff.: 300
  - "Clever" nbg. selection
  - Years per one part: 40
  - Overlap - years: 10
  - Common period
  - Allow length +/- overlay
  - Correlations column: K13
  - Use only deleted
  - Weighted average
  - Region selection:
    - the same region
    - different region
    - any region

**MON 9-3**
- **Settings**
  - Years around: 4
  - %Y_possible for no Metadata: [20,10]
  - Compare 1:
    - ID - HomResults2: ID_ORIG
    - ID - Metadata 1: STNR
  - Compare 2:
    - ID - HomResults2: ID_ORIG
    - ID - Metadata 1: STNR
    - Link Metadata 2 with Metadata
    - ID_link - Metadata 2: NGB_ORIG
    - ID_link - Metadata 1: STNR
  - Output:
    - Sort output

**MON 10-2**
- **Settings**
  - Vals. Transformation (precip.):
    - No. of years around: 10
    - Min. years from end: 4
Testing of conditions in ProClimDB

45 realistic break points

- Probable break detection: Y_POSSIBLE (y_p) based on weighting (5(year)/2(seasons)/1(month)). Max possible score is 100% (year, 4 seasons and 12 months). Ex: A break in 5 months, 2 seasons $\rightarrow$ 36%

- Correlation after adjustments, corr + value = 0.000 or 0.005. No adjustments below value. Or none correlation check.
### Number of breaks with different settings_1
40 year periods, different overlap years

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>None</th>
<th>0.000</th>
<th>0.005</th>
<th></th>
<th>None</th>
<th>0.000</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>40-10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>False</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>y_p=20</td>
<td>33</td>
<td>33</td>
<td>29</td>
<td></td>
<td>y_p=20</td>
<td>37</td>
<td>36</td>
<td>16</td>
</tr>
<tr>
<td>y_p=30</td>
<td>25</td>
<td>25</td>
<td>23</td>
<td></td>
<td>y_p=30</td>
<td>12</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>y_p=40</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
<td>y_p=40</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>

| **40-20** |         |      |       |       |       |      |       |       |
| y_p=20 | 35      | 34   | 32    |       | y_p=20 | 35   | 35    | 16    |
| y_p=30 | 29      | 29   | 28    |       | y_p=30 | 15   | 14    | 8     |
| y_p=40 | 24      | 24   | 24    |       | y_p=40 | 10   | 9     | 5     |
## Number of breaks with different settings_2
### Corr.criterion, different values

<table>
<thead>
<tr>
<th>y_p</th>
<th>Correct</th>
<th>0.005</th>
<th>0.010</th>
<th>0.015</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>32</td>
<td>26</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>26</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>32</td>
<td>27</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>30</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>y_p</th>
<th>False</th>
<th>0.005</th>
<th>0.010</th>
<th>0.015</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>21</td>
<td>10</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>18</td>
<td>8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>16</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>9</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Number of breaks with different settings_3

### Perfect metadata

<table>
<thead>
<tr>
<th>y_p</th>
<th>Correct</th>
<th>None</th>
<th>0.000</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>y_p=15;10</td>
<td>37</td>
<td>37</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=20;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=25;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=30;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=35;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=40;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=45;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=50;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

### False

<table>
<thead>
<tr>
<th>y_p</th>
<th>Correct</th>
<th>None</th>
<th>0.000</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>y_p=15;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=20;10</td>
<td>38</td>
<td>38</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>y_p=25;10</td>
<td>19</td>
<td>14</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>y_p=30;10</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>y_p=35;10</td>
<td>11</td>
<td>14</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>y_p=40;10</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>y_p=45;10</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>y_p=50;10</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
1. \( y_p=45;25 \) in the first run and no correlation criterion
2. \( y_p=45;25 \) in the first iteration and no correlation criterion
3. \( y-p=40;20 \) in the second iteration and \( \text{corr} + 0.5\% \)
4. \( y_p=30;15 \) in the third iteration and \( \text{corr} + 0.5\% \)
5. \( y_p=25;10 \) in the fourth iteration and \( \text{corr} + 0.5\% \)
## Chosen procedure_2

<table>
<thead>
<tr>
<th></th>
<th>1.run</th>
<th>1.it.</th>
<th>2.it.</th>
<th>3.it.</th>
<th>4.it</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metadata not available</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detected breaks</td>
<td>23</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>Correct detection</td>
<td>20</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>False detection</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1.run</th>
<th>1.it.</th>
<th>2.it.</th>
<th>3.it.</th>
<th>4.it</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metadata available</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detected breaks</td>
<td>36</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>51</td>
</tr>
<tr>
<td>Correct detection</td>
<td>32</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>43</td>
</tr>
<tr>
<td>False detection</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96 %</td>
</tr>
</tbody>
</table>
Break detection

Series with many false detections

Successful homogenization is dependent on series from at least a few stations in the neighbourhood
Homogenization procedure

**Original Data**

**Quality control**
- Rank of monthly values
  - Dist. / Stand. to alt. / Outliers
- Comparing with neighbours
  - Replacing suspicious values
- Stations within 10 km
- Demands on data coverage
- Merging of different series

**Reconstruction of series**
- Reference series (40 years, 20 years overlap) from correl. / weights
  - Standardization to base station (AVG/STD)
  - Assessment of hom. results
- Reference series (10 years around inhomogeneity) from distances
  - Standardization to base station (AVG/STD)
  - Smoothing monthly adjustments / Demands on corr. after adjustm.

**Homogeneity testing**

**Adjusting Data**

**Iteration process**
Choosing area of analysis

Regionalization of the temperature climate of Norway results in roughly 6 regions.

The borders between the temperature regions coincide very well with the county borders.

From report 25/98 Klima (IH-B & N)
Regions for analysis, defined by counties

Red squares: Analyzed station series
Green squares: Supporting station series

Norwegian Meteorological Institute  met.no
Result of homogenization

83 series considered homogeneous without any adjustments
Result of homogenization

48 series with traces of less extensive breaks, not adjusted
Result of homogenization

20 series considered homogeneous after adjustments, not confirmed by metadata
Result of homogenization

80 series considered homogeneous after adjustments, confirmed by metadata
Result of homogenization

All together 231 series with minimum length of 30 years are analysed

Legend
- **Green circle**: Homogeneous series, no traces of breaks
- **Red triangle**: Traces of breaks, not adjusted
- **Orange square**: Adjusted breaks, confirmed by metadata
- **Blue pentagon**: Adjusted breaks, not confirmed by metadata
The breaks - size and number

```
<table>
<thead>
<tr>
<th>Size of breaks (C)</th>
<th>No. of breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 0.1</td>
<td>0</td>
</tr>
<tr>
<td>0.2-0.3</td>
<td>40</td>
</tr>
<tr>
<td>0.4-0.5</td>
<td>50</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>40</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>10</td>
</tr>
<tr>
<td>≥ 1.0</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Reasons for breaks, from metadata (%)

- Instrumentation
- Relocation
- Screen
- Obs.hour
- Observer
- Other

The chart shows the percentage distribution of reasons for breaks, with Relocation accounting for a significant portion.