ANOVA correction model in relative homogenization: Optional or obligatory?

Peter Domonkos
dpeterfree@gmail.com
https://acmant.eu

DEVELOPMENTS IN WEATHER AND CLIMATE SCIENCE 3

SERIES EDITOR: PAUL D. WILLIAMS

CLIMATE OBSERVATIONS

Data Quality Control and Time Series Homogenization

PETER DOMONKOS, RÓBERT TÓTH AND LÁSZLÓ NYITRAI

ELSEVIER

RMetS
Royal Meteorological Society

Chapter 1. Land surface observations

Chapter 2. Upper air observation and remote sensing

Chapter 3. Data quality control and dataset development

Chapter 4. Homogenization task and its principal approaches

Chapter 5. Relative homogenization: The basis

Chapter 6. Relative homogenization: Optional tools

Chapter 7. Relative homogenization: Special problems

Chapter 8. A selection of statistical homogenization methods

Chapter 9. Accuracy of homogenization results

Chapter 10. Use of quality controlled and homogenized data

Aknowledgement

- Catalan Meteorological Service partially funded my works in the last few years.

Kinds of homogenization problems

(i) **Sufficiently studied and the clear advantage of one or a few of methods is proven**

(ii) Sufficiently studied, but various methods yield similar homogenization accuracy

(iii) Exact evaluation would need further studies

(iv) Special problems, need of individual decisions

ANOVA correction model: Part of (i), its advantage is significant, applicable in a wide range of homogenization tasks

Simple ANOVA correction model

Simple model: all inhomogeneities are breaks, climate signal is spatially uniform; *x* – observed value, *u* – climate signal, *v* – station effect, *i* – time point, *s* – station, *k* – section without break in series *s*

$$\mathbf{X} = \mathbf{U} + \mathbf{V} + \boldsymbol{\varepsilon} \qquad (\mathbf{X} = x_1, x_2, \ldots x_i \ldots x_n \quad (i = 1,2 \ldots n)) \qquad (1)$$

- For each point of the time series:

$$\widehat{u_i} + \frac{1}{N}\sum_{s=1}^{N} \widehat{v_{s,k(i)}} = \frac{1}{N}\sum_{s=1}^{N} x_{s,i} \qquad (2)$$

- For each homogeneous section of each time series:

$$\frac{1}{j_{s,k+1} - j_{s,k}}\sum_{i=j_{s,k}+1}^{j_{s,k+1}} \widehat{u_i} + \widehat{v_{s,k}} = \overline{\mathbf{x}_{\mathbf{s,k}}} \qquad (3)$$

ANOVA model: optimal estimations (Why?)

- Joint use of all pieces of information in an equation system
- The model conditions are expected in any other correction method (with some exceptions of limited importance)
- In the simple ANOVA model the number of parameter estimations from noisy, inhomogeneous sample is minimized.

- A problem: Results of Eqs. (2)&(3) do not include direct solution to *u* and *v*, but only to the break sizes. --- How do we know then that the break size estimations are really optimal? A full proof is shown by:

Lindau, R. and Venema, V.K.C. (2018) Int. J. Climatol., http://doi.org/10.1002/joc.5728.

History of ANOVA model use in homogenization

- Used first by Olivier Mestre in 1996. Mestre was modest and not very effective in convincing colleagues.
- Conditions for achieving good results with ANOVA model were over-emphasised.
- Metadata role sometimes overestimated
- Individualism and incomplete understanding of colleagues’ results; geographical – political division
- Climatol method: spatial interpolation does not need the accuracy of break dates (except for the last break of time series)
- Around 2010 Mestre, Venema and Domonkos managed most the ANOVA correction model.

Use of ANOVA model: practical advice

- Equations of (2) can be merged for sections without break in all time series, reducing the dimension of the problem matrix; hence the ANOVA correction model can easily be used in any time resolution.
- Internal data gaps are infilled before using ANOVA model. In HOMER, a modified model version is applied, no previous gap filling is needed there, but tests have not proven yet its correctness (I do not know…).
- External missing data (those before the starting and ending dates of a relatively short time series) are left as they are.
- ANOVA correction model can be applied to any kind of variable, not only to section means (e.g. to summer-winter differences).
- Coincidental breaks in all time series is not allowed.

ANOVA model and metadata use

- **Station specific, non-quantified metadata**: their importance is the greatest for small networks.
- Quantitative analysis with efficiency tests: Domonkos (2022). Automatic homogenization of time series: How to use metadata? Atmosphere, https://doi.org/10.3390/atmos13091379.
- Principal conclusion of efficiency tests: The benefit of metadata is the highest when **metadata dates are included in the ANOVA model as break dates without testing statistical significance**. More than 50% of metadata dates are expected to point on true technical changes.

Weighted ANOVA model

- Weighted ANOVA model considers spatial changes of climate. Its idea comes from Szentimrey’s study on weighting neighbour series (2008) and Venema’s oral communications. Eq. (2) is transformed to (5) by weighting *x*, as well as its *u* and *v* components.
- $$\sum_{s=1}^{N} w_s \widehat{u_{s,i}} + \frac{1}{N}\sum_{s=1}^{N} w_s \widehat{v_{s,k(i)}} = \frac{1}{N}\sum_{s=1}^{N} w_s x_{s,i} \qquad (5)$$
- The calculations are performed separately to each series, i.e. each series once takes the role of candidate series for which $w \equiv 1$.
- Theoretical optimal weights can be calculated by ordinary kriging.
- Practical application in ACMANT uses squared spatial correlations.

Use of ANOVA model in ACMANT

- Adjustments for inhomogeneity biases are performed by the ANOVA correction model, except for quality problems of shorter than 5 months periods.
- Intermediate phases: simple ANOVA model, annual time resolution
- Final adjustments: weighted ANOVA model, daily or monthly resolution
- ANOVA model is applied to time series of annual means and summer-winter differences (in sinusoid seasonality of inhomogeneity bias model) in various time resolution
- ANOVA model is applied to annual series of monthly means (in irregular seasonality model) and to annual series of seasonal means (in precipitation homogenization of rainy seasons and snowy seasons)

More about my results/products:

https://acmant.eu

Thank you for your attention!

MULTITEST results:

J Climate, 2021, https://doi.org/10.1175/JCLI-D-20-0611.1

- green: ACMANTv4 was significantly better than any other method;
- yellow: ACMANT was tied to the first place with some other method(s);
- red: some other method was significantly better than ACMANT (Y5: series with synchronous breaks)

	RMSE-m	RMSE-year	Trend bias	NetworkTr	Net-RMSEy
Y1					
Y2					
Y3					
Y4					
Y5					
Y6					
U1					
U2					
U3					
U4					
U5					
U6					