Towards Unified Error Reporting (TUNER): Strategies for survival in the jungle of random, systematic, smoothing and headache errors

Thomas von Clarmann, Doug Degenstein, Nathaniel Livesey and the TUNER Team

TUNER

TOWARDS UNIFIED ERROR REPORTING

- A paper on recommendations on error reporting has been published by the TUNER team.
- In this presentation I will focus on issues relevant for the TOAR project (as far as I can judge)

Atmos. Meas. Tech., 13, 4393–4436, 2020
https://doi.org/10.5194/amt-13-4393-2020

Atmospheric Measurement Techniques

Overview: Estimating and reporting uncertainties in remotely sensed atmospheric composition and temperature

Thomas von Clarmann[1], Douglas A. Degenstein[2], Nathaniel J. Livesey[3], Stefan Bender[4], Amy Braverman[3], André Butz[5], Steven Compernolle[6], Robert Damadeo[7], Seth Dueck[2], Patrick Eriksson[8], Bernd Funke[9], Margaret C. Johnson[3], Yasuko Kasai[10], Arno Keppens[6], Anne Kleinert[1], Natalya A. Kramarova[11], Alexandra Laeng[1], Bavo Langerock[6], Vivienne H. Payne[3], Alexei Rozanov[12], Tomohiro O. Sato[10], Matthias Schneider[1], Patrick Sheese[13], Viktoria Sofieva[14], Gabriele P. Stiller[1], Christian von Savigny[15], and Daniel Zawada[2]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Karlsruhe, Germany
[2]Department of Physics & Engineering Physics, University of Saskatchewan, Saskatoon, Canada
[3]NASA Jet Propulsion Laboratory and California Institute of Technology, Pasadena, CA, USA
[4]Department of Physics, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway
[5]Institut für Umweltphysik, Department of Physics and Astronomy, Universität Heidelberg, Heidelberg, Germany
[6]Department of Atmospheric Composition, Royal Belgian Institute for Space Aeronomy (BIRA-IASB), 1180 Brussels, Belgium
[7]NASA Langley Research Center, Hampton, VA, USA
[8]Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden
[9]Instituto de Astrofísica de Andalucía, CSIC
[10]National Institute of Information and Communications Technology (NICT), 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan
[11]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[12]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany
[13]Department of Physics, University of Toronto, Toronto, Canada
[14]Finnish Meteorological Institute, Helsinki, Finland
[15]Institute of Physics, Greifswald University, Greifswald, Germany

Correspondence: T. von Clarmann (thomas.clarmann@kit.edu)

Received: 20 September 2019 – Discussion started: 14 October 2019
Revised: 26 June 2020 – Accepted: 4 July 2020 – Published: 17 August 2020

Abstract. Remote sensing of atmospheric state variables typically relies on the inverse solution of the radiative transfer equation. An adequately characterized retrieval provides information on the uncertainties of the estimated state variables as well as on how any constraint or a priori assumption affects the estimate. Reported characterization data should be intercomparable between different instruments, empirically validatable, grid-independent, usable without detailed knowledge of the instrument or retrieval technique, traceable and still have reasonable data volume. The latter may force one to work with representative rather than individual characterization data. Many errors derive from approximations and simplifications used in real-world retrieval schemes, which are reviewed in this paper, along with related error estimation schemes. The main sources of uncertainty are measurement noise, calibration errors, simplifications and idealizations in the radiative transfer model and retrieval scheme, auxiliary data errors, and uncertainties in atmospheric or instrumental parameters. Some of these errors affect the result in a random way, while others chiefly cause a bias or are of mixed character. Beyond this, it is of utmost importance to know the influence of any constraint and prior information on the solution. While different instruments or retrieval schemes may require different error estimation schemes, we provide a list of recommendations which should help to unify retrieval error reporting.

Published by Copernicus Publications on behalf of the European Geosciences Union.

A side note:

- GUM claims that “error” and “uncertainty” are two different concepts.
- The difference is not clear; neither within GUM nor in the rest of the literature.
- I use the terms “error” and “uncertainty” interchangeably.
- At worst I will get a ticket/penalty from the terminology police…

What are systematic errors?
What are random errors?

TUNER
TOWARDS UNIFIED ERROR REPORTING

- You find multiple, partly contradicting definitions in the literature.
- We use the following definition:
- Systematic errors are errors which generate a bias between observations of the same airmass by different instruments.
- Random errors are errors that generate a standard deviation of the differences between observations of the same airmass by different instruments.
- Headache errors are errors which generate both.

Advantage: error estimates according to these definitions are empirically testable

Can systematic errors be represented by a pdf?

- Error estimates are often characterized by statistical quantities (variance, standard deviation) etc. Often, error bars represent standard deviations.
- Some people claim that this is nonsense because a systematic error is a single value and not a part of a frequency distribution.
- Since in the context of systematic errors frequentist statistics fails, we can conceive ‘probability’ as the belief of a rational agent (often called `subjective probability’);
- The ‘fair bet ratio’ is a means to get an intuitive handle on this concept.
- With this concept it is possible to characterize systematic errors with a standard deviation.
- This is a precondition to combine systematic and random errors to an estimate of the total error.

Should random and systematic errors be reported as a combined, total error?

TUNER
TOWARDS UNIFIED ERROR REPORTING

- NO!
- For some applications only the random error is relevant (e.g. time series)
- For some application only the systematic error is relevant (when large amounts of data are averaged, the random component averages out)
- Error components should be reported separately.

How to evaluate systematic errors?

TUNER
TOWARDS UNIFIED ERROR REPORTING

- Either you can use linear (Gaussian) error propagation: $\mathbf{S}_x = \mathbf{GKS}_b\mathbf{K}^T\mathbf{G}$

  $\mathbf{S}_x$ error covariance matrix of the target quantities

  G Gain (sensitivity of the target variables wrt the direct measurements)

  K Jacobian (sensitivity of the direct measurement wrt the error-causing parameter)

  $\mathbf{S}_b$ error covariance matrix of the error-causing parameter

- Or you can perform sensitivity studies.

  $\sigma_x = \mathbf{G}(\mathbf{y};\mathbf{b}+\boldsymbol{\sigma}_b) - \mathbf{G}(\mathbf{y};\mathbf{b}_b)$

Headache errors

- Some error sources cause both a systematic and a random error component;
- Example: in remote sensing, spectroscopic data uncertainties cause a systematic error, because the same spectroscopic data are used for each retrieval. But variations of the atmospheric state `modulate' this originally systematic error; i.e., the sensitivity of the target quantity to the uncertain parameter may depend on the atmospheric state. This causes a random component.
- Often one can categorize the error depending on the dominance of either the random or the systematic component.
- With MIPAS we try to do sensitivity studies for larger samples and calculate the bias and the scatter individually

Error correlations in various domains

TUNER
TOWARDS UNIFIED ERROR REPORTING

- Often we say about an error “it is systematic in altitude” because the state value is too high in all altitudes. Strictly speaking, it can still be a random error, but correlated in the altitude domain.
- Correlations in various domains are important; depending on the application of the data, the same error can behave like random or systematic errors
- Domains: altitude, among species, …
- Applications:

  calculation of vertical column amounts from profiles;

  Trace gas budgets

A priori and smoothing issues

- For retrievals where a priori constraints are used, the a priori and the averaging kernels should be reported.
- In order to reduce data volume, often averaged data are reported; if in such cases averaged averaging kernels are reported, also the correlation between the state variable and the averaging kernel is needed (von Clarmann and Glatthor, AMT, 2019)
- Instead of the smoothing error (Rodgers 2000), the averaging kernels should be provided to the user. The reason is that the smoothing error cannot be propagated to finer grids (direct evaluation on the finer grid would render a larger smoothing error!) (von Clarmann, AMT, 2014). Attention: Rodgers' S_x error covariance matrix does include the smoothing error.

Further TUNER recommendations:

TOWARDS UNIFIED ERROR REPORTING

- Define your terminology
- The error budget should be as complete as possible
- Report confidence limits (or whatever the meaning of the error bars is)
- Error diagnostics should be reported for the same discretization which is used for the data as reported.
- If representative error estimates are reported: are the error components additive or multiplicative?

THANK YOU!