Karlsruher Institut für Technologie

Towards unified error reporting

Thomas von Clarmann

IMK-ASF-SAT

Why do we need error estimates*?

*) When I say ‘error estimates’ I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

Why do we need error estimates*?

- To judge if the data are adequate to investigate a certain process.

*) When I say ‘error estimates’ I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

Why do we need error estimates*?

- To judge if the data are adequate to investigate a certain process.
- To judge if any difference is significant.

*) When I say 'error estimates' I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

Why do we need error estimates*?

- To judge if the data are adequate to investigate a certain process.
- To judge if any difference is significant.
- To optimally weight multiple data-sets when they are combined. Particularly important for DA!

*) When I say ‘error estimates’ I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

Why do we need error estimates*?

- To judge if the data are adequate to investigate a certain process.
- To judge if any difference is significant.
- To optimally weight multiple data-sets when they are combined. Particularly important for DA!
- To know how errors co-vary. Particularly important for DA!

*) When I say ‘error estimates’ I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

Why do we need error estimates*?

- To judge if the data are adequate to investigate a certain process.
- To judge if any difference is significant.
- To optimally weight multiple data-sets when they are combined. Particularly important for DA!
- To know how errors co-vary. Particularly important for DA!
- For a lot of other reasons I might have forgotten.

*) When I say ‘error estimates’ I always mean the total diagnostic metadata including information and weight of any a priori information used, spatial resolution etc…

What do we have?

What do we have?

- A mess!

What do we have?

- A mess!
- Dozens of different methods are used!

What do we have?

- A mess!
- Dozens of different methods are used!
- How are these comparable?

Various approaches:

- Noise + smoothing + Parameter errors (Rodgers 1987)
- Retrieval error (Rodgers 1976)
- Error estimate from fit residuals
- Ensemble studies (von Clarmann et al. 1992; Povey & Grainger 2015)
- Errors estimated from repeated measurements (Laeng et al 2015)
- Errors from comparison of multiple instruments.
- …

Various approaches:

Karlsruher Institut für Technologie

- Noise + smoothing + Parameter errors (Rodgers 1987)
- Retrieval error (Rodgers 1976)
- Error estimate from fit residuals
- Ensemble studies (von Clarmann et al. 1992; Povey & Grainger 2015)
- Errors estimated from repeated measurements (Laeng et al 2015)
- Errors from comparison of multiple instruments.
- …

The number of approaches is somewhere between 6 and the number of authors, probably closer to the latter☺

Rodgers 1990/2000 approach:

- The total error is composed of noise, parameter and model error, and the smoothing error (JGR 95 5587-5595,1990; World Scientific 2000)

Pro:

- All known error sources are considered.
- Error correlations in the altitude domain are provided.
- For error estimation, quantities available from a retrieval using the Rodgers formalism can be used.

Con:

- Error correlations in other domains are missing.
- The concept of the smoothing error is problematic.
- Many groups do not have the algorithms to provide these error terms, since their retrievals are not based on the Rodgers formalism.

Smoothing error pitfalls

- The smoothing error evaluated on one grid and then propagated onto a finer grid differs from the smoothing error evaluated on the fine grid, even if the propagation is made in full consistence with generalized Gaussian error propagation (T.v.C., AMT 7 3023-3034, 2014).
- This behaviour is not acceptable for an “error”, because data users rely on the applicability on the established error propagation laws.
- The reason of the problem is that the smoothing error evaluated on a discrete grid ignores any atmospheric variability not resolved by this grid.

Remedy: Distribute the averaging kernels instead.

Rodgers 1976 approach:

- $S_x = (K^T S_y^{-1} K + S_a^{-1})^{-1}$ Rev. Geophys. Space Phys, 14 609-624, 1976

Pro:
- Noise and smoothing error considered.
- Error correlations in the altitude domain are provided.
- If a Rodgers-type formalism is used for the retrieval, no additional effort is needed

Con:
- Error correlations in other domains are missing.
- The concept of the smoothing error is problematic.
- The concept is only adequate if x_a and S_a are the true expectation values and their covariances.
- Many groups do not have the algorithms …

Error estimates from fit residuals

- There exist several ways to do this. How these approaches are mathematically justified is not always clear.

Pro:

- Does not require the full matrix algorithm, i.e. can possibly applied to “USA East coast algorithms”
- Does not contain the smoothing error and is usually applied to retrievals which have no formal smoothing error.
  ➔self-consistent

Con:

- Not always clear which errors show up in the residuals.
- Error components whose spectral responses are correlated with the target Jacobian may not be detected.

Ensemble studies:

A.C. Povey and R. G. Grainger: Known and unknown unknowns: the application of ensemble techniques to uncertainty estimation in satellite remote sensing data (AMTD 8, 8509-8562, 2015)

Pro:

- Uncertainty estimates do not depend on one single retrieval algorithm.

Con:

- Effort.
- Common heritage of algorithms.
- There are not always multiple independent systems available.
- Single poor algorithms can lead to a false negative result about our knowledge.
- No (!!!) access to unknown unknowns.

Repeated Measurements

The standard deviation of repeated measurements is an estimator of the uncorrelated component of the error of each measurements.

Pro:

- Not as theory-laden as the other error estimates; does not depend on any particular model.

Con:

- Correlated error components remain undetected.
- Repeated measurements are not always possible (Way out: structure analysis, Sofieva et al., AMT 7, 2147-2158, 2014; Laeng et al., AMTD 8 5565-5590, 2015)

Multiple Instruments

Karlsruher Institut für Technologie

Comparison of data of multiple instruments can uncover specific problems

Pro:

- To a large degree independent of the assumptions made in the context of each single data set.

Con:

- Not always clear which of the intercompared data sets is closer to truth.
- Collocated measurements are typically available only for a small subset of the data.

All in all better suited for validation than for routine error estimation. But it does help to test if the error estimates are realistic.

What do we need: (rigorously speaking)

What do we need: (rigorously speaking)

- Complete error budget including noise, parameter errors, model errors.

What do we need: (rigorously speaking)

- Complete error budget including noise, parameter errors, model errors.
- Complete covariance information of these in all domains

What do we need: (rigorously speaking)

- Complete error budget including noise, parameter errors, model errors.
- Complete covariance information of these in all domains
- Averaging kernels and prior used

What do we need: (rigorously speaking)

- Complete error budget including noise, parameter errors, model errors.
- Complete covariance information of these in all domains
- Averaging kernels and prior used
- All this for all atmospheric situations (because these quantities can be state-dependent).

What do we need: (rigorously speaking)

- Complete error budget including noise, parameter errors, model errors.
- Complete covariance information of these in all domains.
- Averaging kernels and prior used.
- All this for all atmospheric situations (because these quantities can be state-dependent).

But:

- We will never get all this.
- If we had all this: is the user who downloads our data educated enough to use all this in an appropriate manner?
- All these diagnostics need magnitudes more storage on the data server than the atmospheric state data themselves. Is it worthwhile?

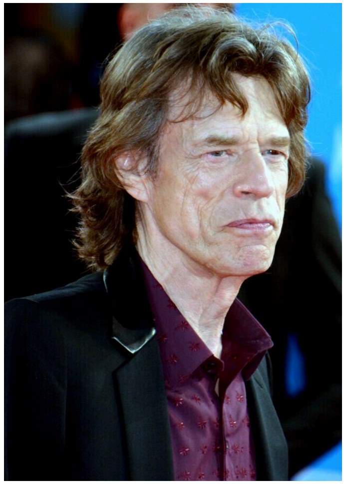
W at do we need: (rigorously speaking)

- ror budget including noise, parameter errors, model errors.
- nce information of these in all domains.
- rior used.
- All th uations (because these quantities can be state-dep

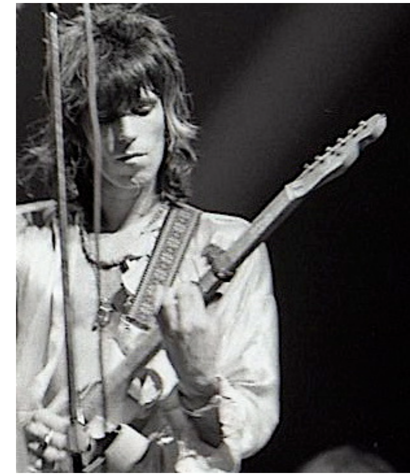
Decades before us some wise men have already discovered that

But:

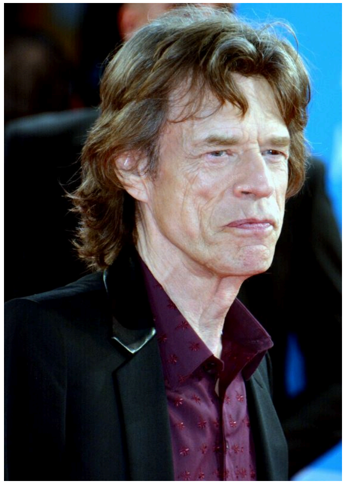
- We will never get all thi
- If we had all this: is the user ata educated enough to use all this in ner?
- All these diagnostics need magnitudes data server than the atmospheric state data it worthwhile?

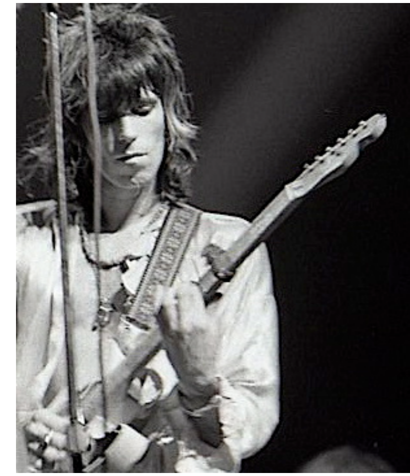


You can't always get what you want
You can't always get what you want
You can't always get what you want
…



M. Jagger & K. Richards 1969





You can't always get what you want
You can't always get what you want
You can't always get what you want

And if you try sometime
You get what you need

M. Jagger & K. Richards 1969
Photo: “Keith Richards and Guitar” by Dina Regine, CC BY SA-2.0
Photo: “Mick Jagger Deauville 2014” by Georges Biard, CC BY SA-3.0

What do we really need?:

Fully quantitative analyses should be made in cooperation with the data provider.

- The data providers (should) know how to deal with the diagnostics provided.
- The data providers (should) know the limitations and weaknesses of their data.

But how can the data users be helped who just want to download data for some more or less basic applications? Should these not be provided with some reasonable uncertainty estimates?

What about the confusion caused by the different ways to report errors?

How shall error reporting look like?

Solution: still unknown!

:

I might prefer this

Clive might prefer that

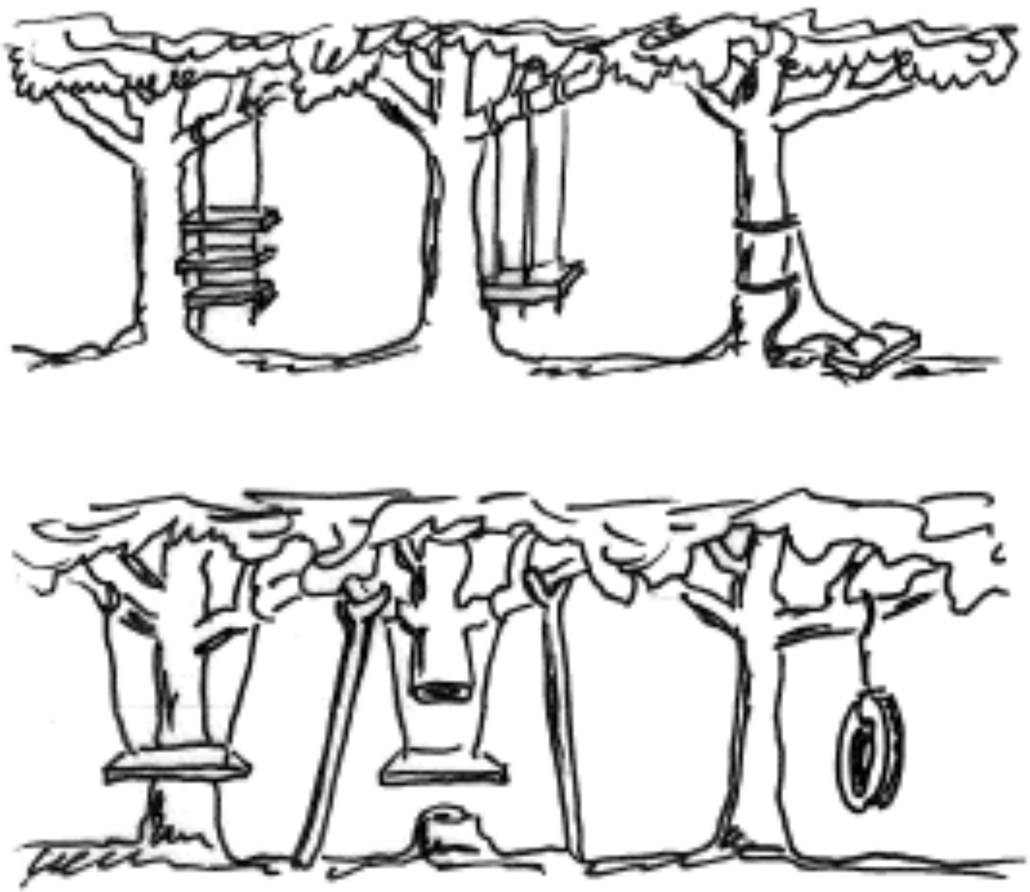
Give the problem to your software engineer and you might get this

Your instrument PI might prefer this

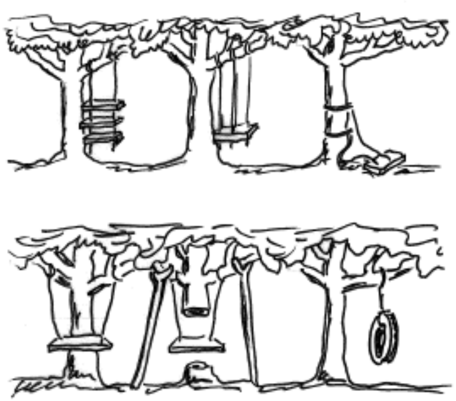
Karlsruher Institut für Technologie

Space agency engineers then might come up with a solution like this

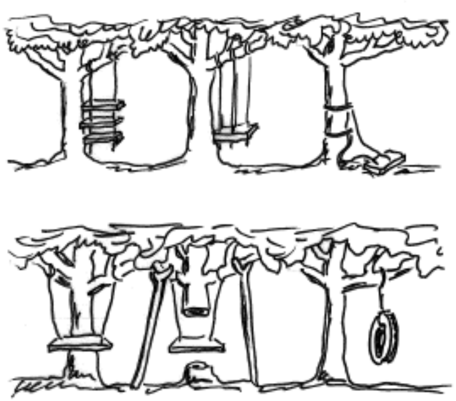
What the data user really needs may simply be this



How can we arrive at unified error reporting?

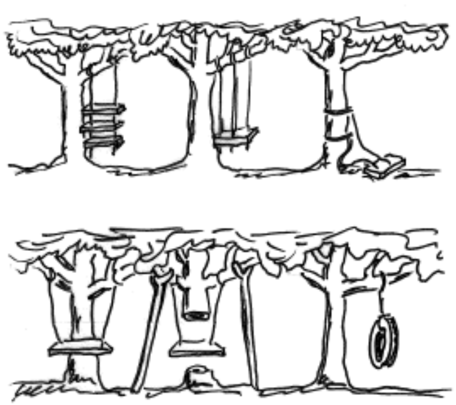


We must sit together and discuss:



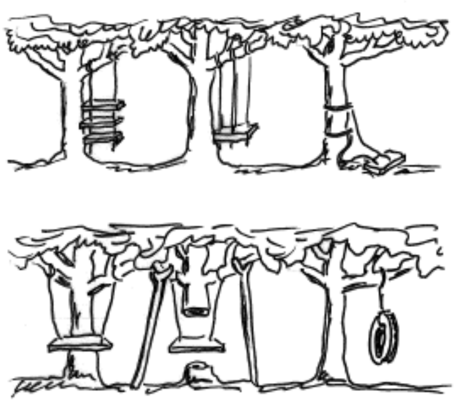
We must sit together and discuss:

- Which diagnostics are really important for the user?



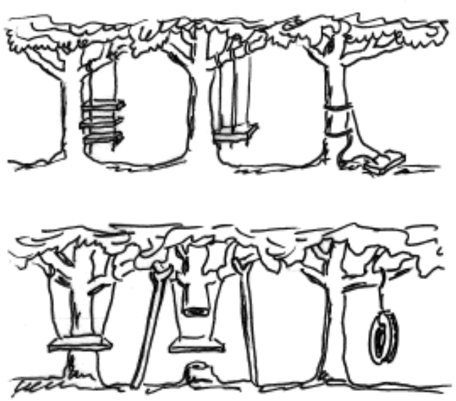
We must sit together and discuss:

- Which diagnostics are really important for the user?
- How can errors from different provider be made comparable?



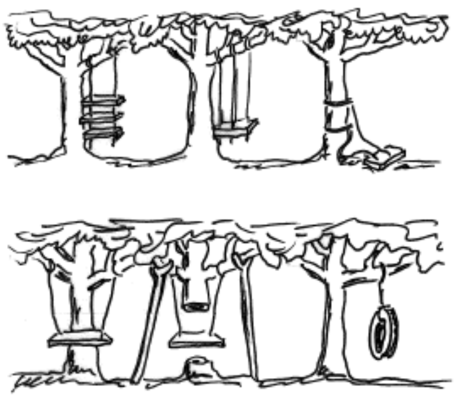
We must sit together and discuss:

- Which diagnostics are really important for the user?
- How can errors from different provider be made comparable?
- How to deal with missing diagnostics?



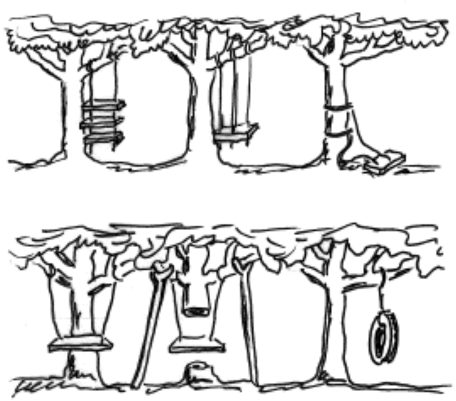
Karlsruher Institut für Technologie

- TUNER has been suggested as a SPARC activity.



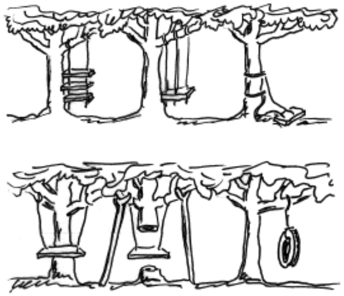
- TUNER has been suggested as a SPARC activity.
- A team has been formed, including PIs and retrieval experts of 12 important space-borne EO missions; further participants welcome!


- TUNER has been suggested as a SPARC activity.
- A team has been formed, including PIs and retrieval experts of 12 important space-borne EO missions; further participants welcome!
- `Liaison scientists' from data using SPARC activities shall be included in the core team to represent the data users.

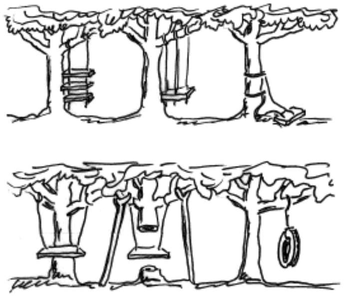


KIT Karlsruher Institut für Technologie

- TUNER has been suggested as a SPARC activity.
- A team has been formed, including PIs and retrieval experts of 12 important space-borne EO missions; further participants welcome!
- `Liaison scientists' from data using SPARC activities shall be included in the core team to represent the data users. SPARC-DA is herewith encouraged to nominate a liaison scientist.

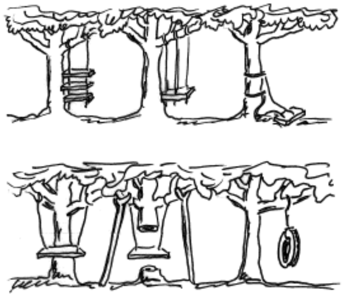


Work plan:



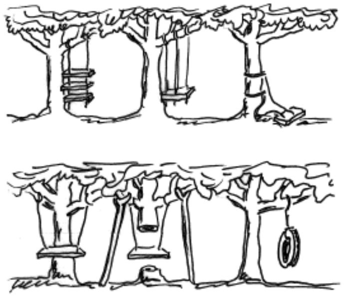
Work plan:

- Inventory of error estimation methods actually used.



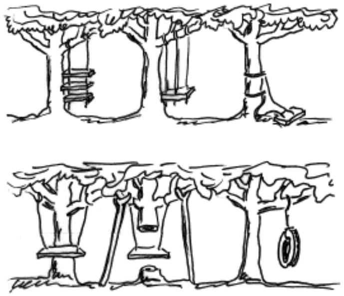
Work plan:

- Inventory of error estimation methods actually used.
- Review of user requirements.



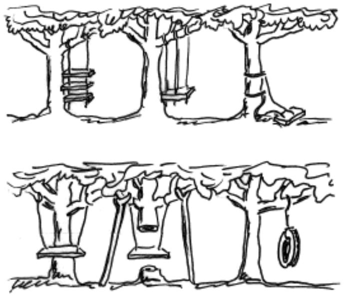
Work plan:

- Inventory of error estimation methods actually used.
- Review of user requirements.
- Develop recommendations how unified error reporting should look like.



Work plan:

- Inventory of error estimation methods actually used.
- Review of user requirements.
- Develop recommendations how unified error reporting should look like.
- Provide diagnostic metadata for selected data sets; this includes emulation of missing diagnostics.



Work plan:

- Inventory of error estimation methods actually used.
- Review of user requirements.
- Develop recommendations how unified error reporting should look like.
- Provide diagnostic metadata for selected data sets; this includes emulation of missing diagnostics.
- Provide recommendations how the diagnostic metadata should be used (e.g. how to use the averaging kernel matrix to construct the H operator)

THANK YOU!

THANK YOU!

And if you think that this activity is useful, please support us in getting TUNER approved as a SPARC activity.