BUG or FEATURE?
The importance of data quality in science

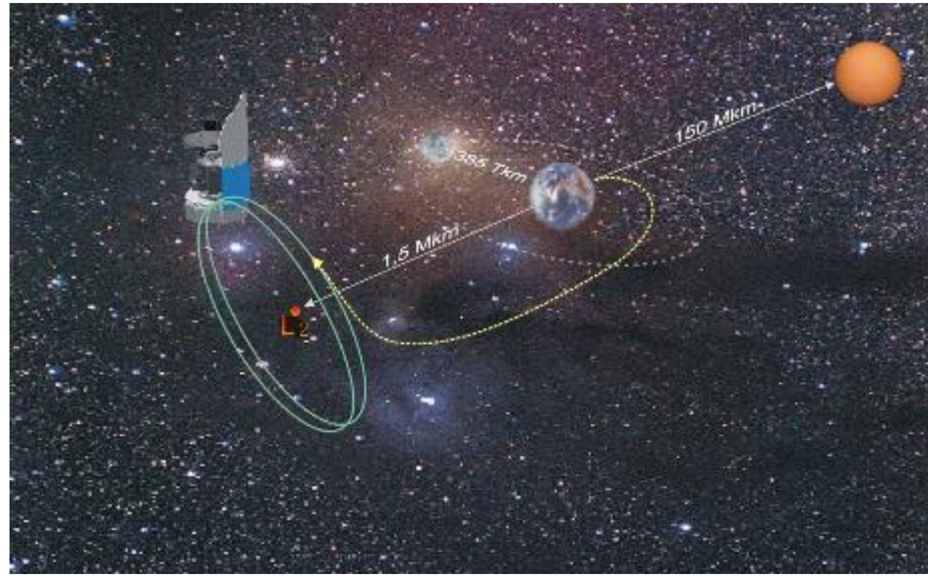
Martin Kunz
Physique Théorique, Université de Genève

The scientific results that we present today are a product of the Planck Collaboration, including individuals from more than 100 scientific institutes in Europe, the USA and Canada

Planck is a project of the European Space Agency, with instruments provided by two scientific Consortia funded by ESA member states (in particular the lead countries: France and Italy) with contributions from NASA (USA), and telescope reflectors provided in a collaboration between ESA and a scientific Consortium led and funded by Denmark.

Planck: a microwave telescope in space

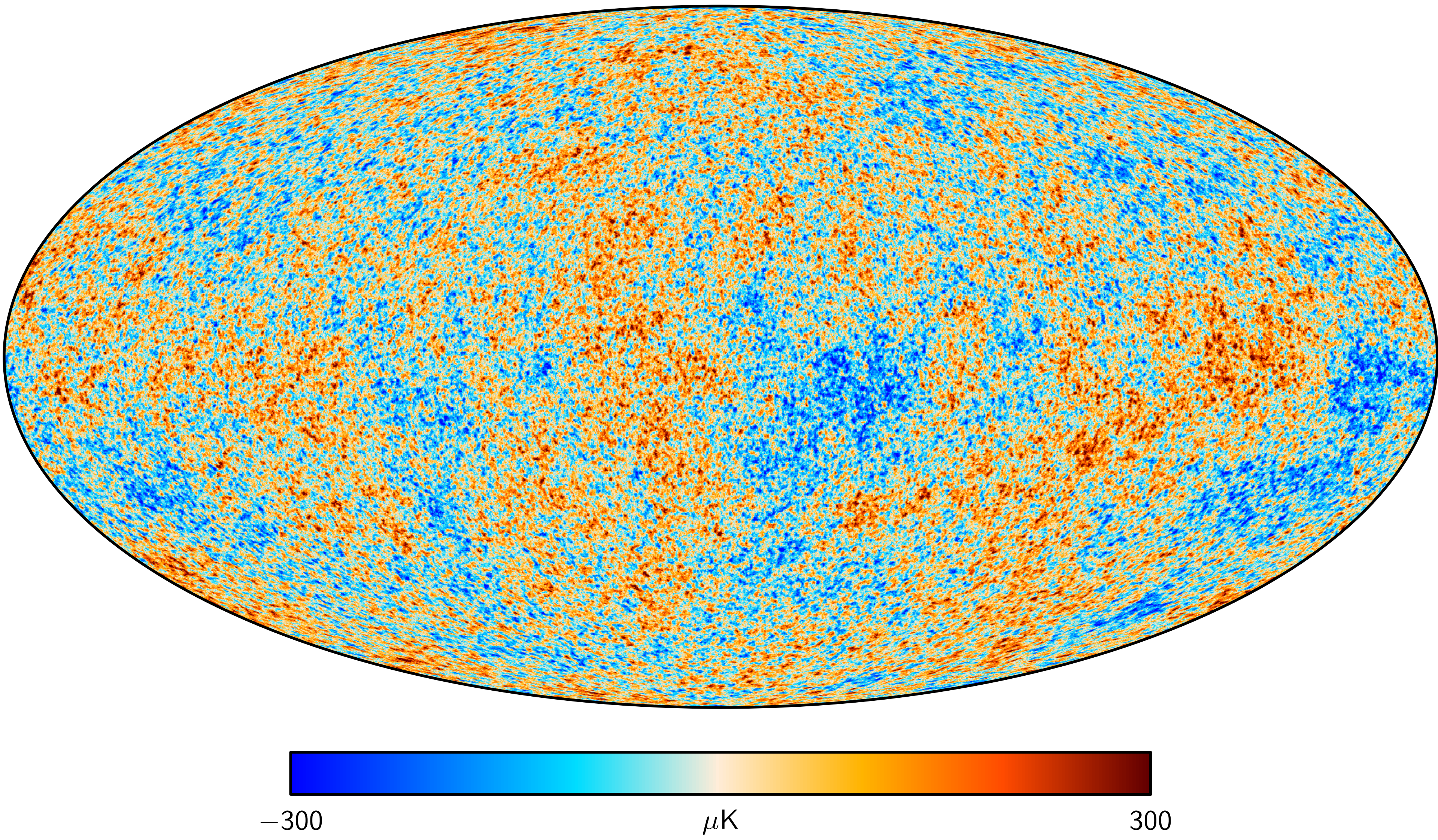
planck



esa

Agency

the cosmic microwave background: a photo of the adolescent universe



How to get good data?

- **Avoid contamination: go to space**
  - Clean environment (except when it isn’t)
  - But: in space no-one can hear you scream (hard to fix problems)

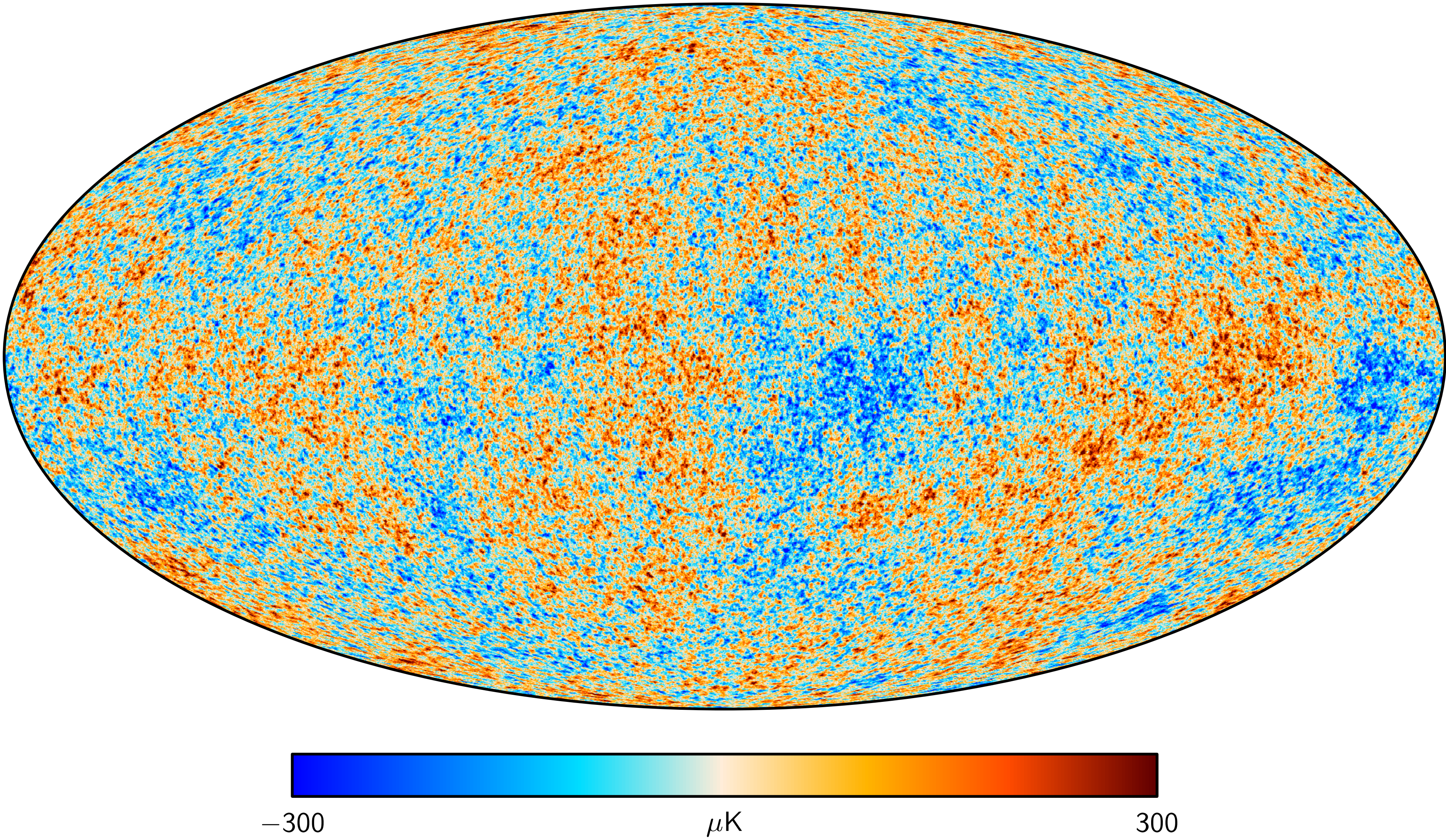
- **High sensitivity**
  - Reduce ‘statistical’ error bars as far as possible
    - Makes any detection much more convincing
  - But needs control of ‘systematic’ problems

- **Build in redundancy**
  - Measure (and analyze) same things in different ways
    - Multiple detectors, redundant scanning, independent analyses
    - Allows for cross-checks
  - Prefer cross-correlations over auto-correlations

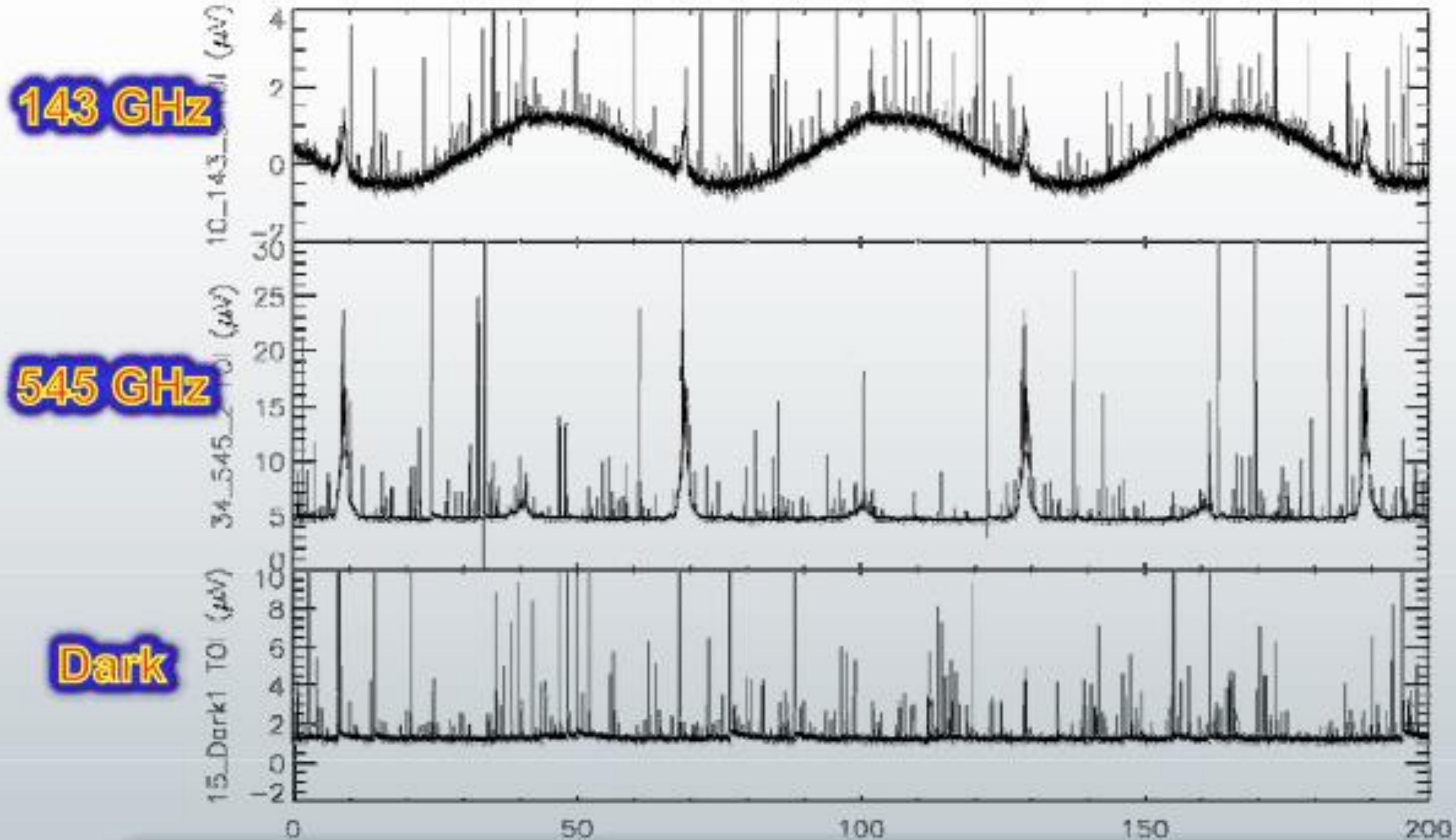
- **Keep looking for problems**

Ceci n'est pas une mesure …

planck



Ceci est une mesure!



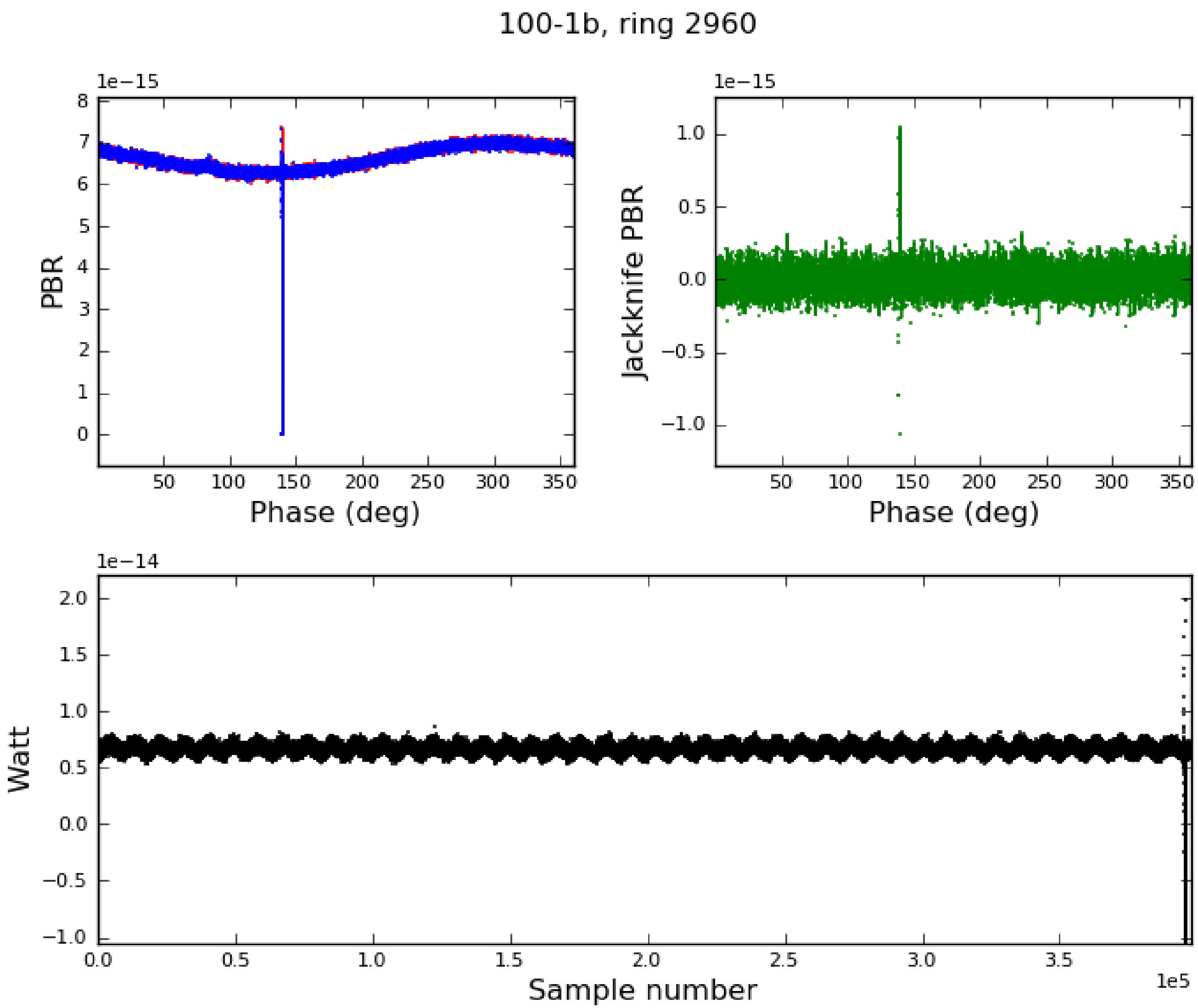
3 minutes of quasi 'raw" data (i.e. only demodulated).The Solar (cosmological) dipole is clearly visible at 145GHz with a 60 seconds period (the satellite rotates at 1 rpm), while the Galactic plane crossings (2 per rotation) are more visible at 545 GHz than at 143 GHz. The Dark bolometer sees no sky signal, but displays a similar population of glitches from cosmic rays.

Not everything can be anticipated

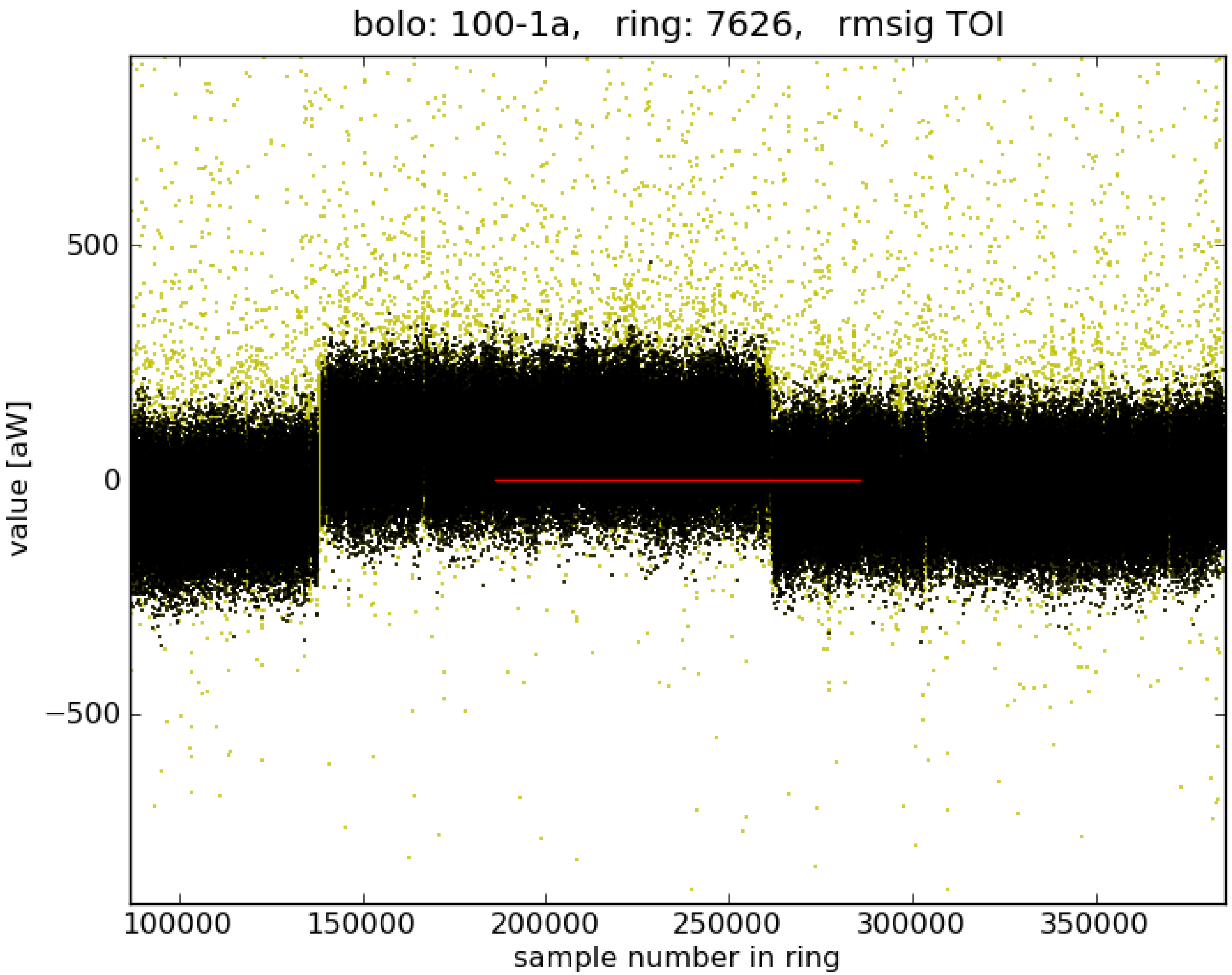
planck

- **There are always surprises**
  - Checking everything is crucial
  - Simpler in the past: just look at all the data, see if something appears ‘weird’
- **No-one will be able to look at all the data ever again … only machines!**
  - Planck data set: ca 10^{12} samples, a few terabytes
  - SKA (large radio telescope) expected data rate: several GB/s!
  - But how do you tell a computer what is ‘weird’?
- **With Planck we found some surprises ‘the hard way’**
  - Active solar period: solar rays much worse than expected
    - Needed to build a detailed model of the satellite to understand signatures of cosmic ray hits to subtract them out
  - Space-qualified analog-digital converter was badly suited
    - In principle known, but no-one realized what it meant
    - Unexpected gain variations observed in data
    - Needed to characterize ADC on actual data (in space no-one…)
  - The ADC was an unknown unknown … there were others …

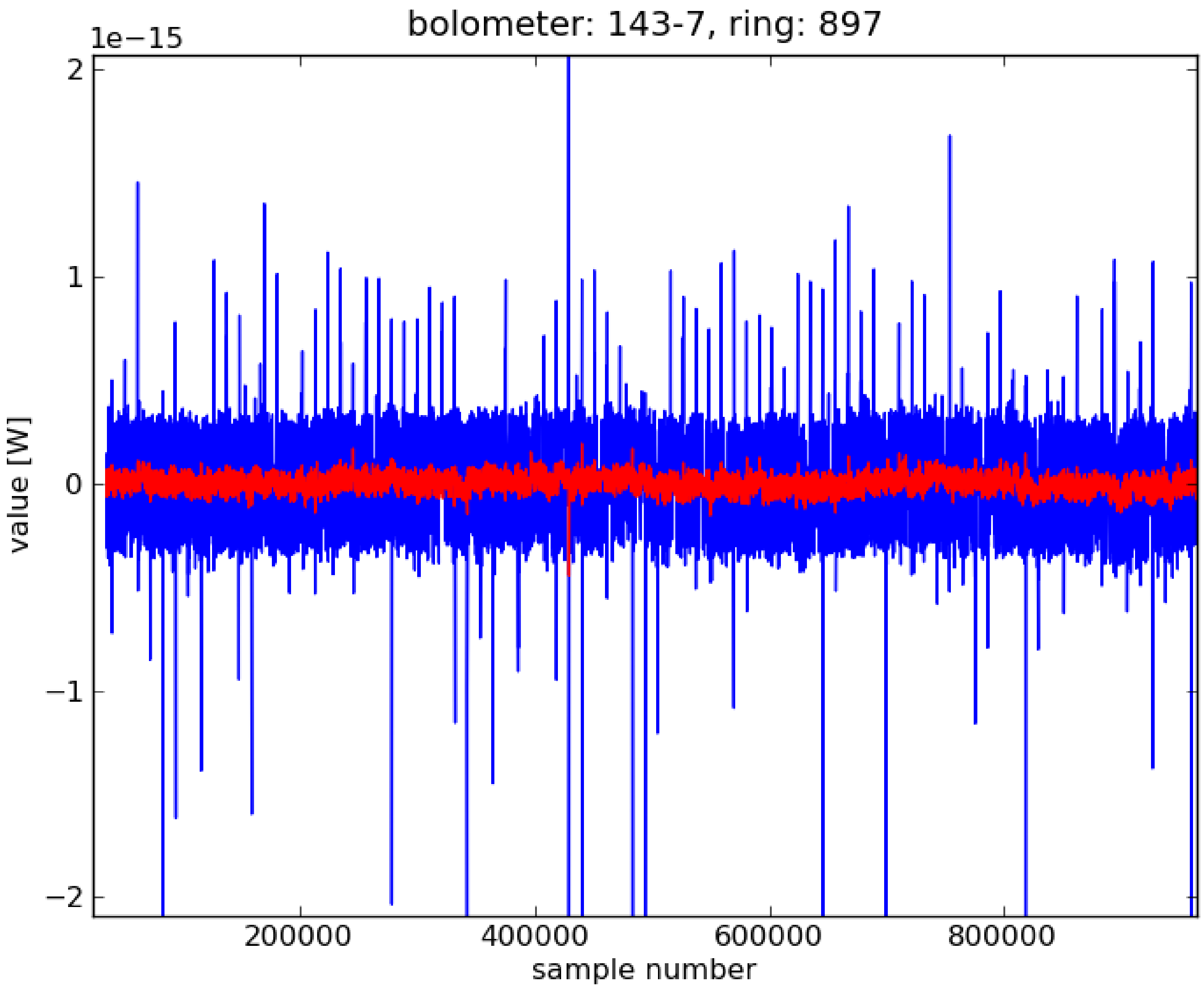
This should be white noise …



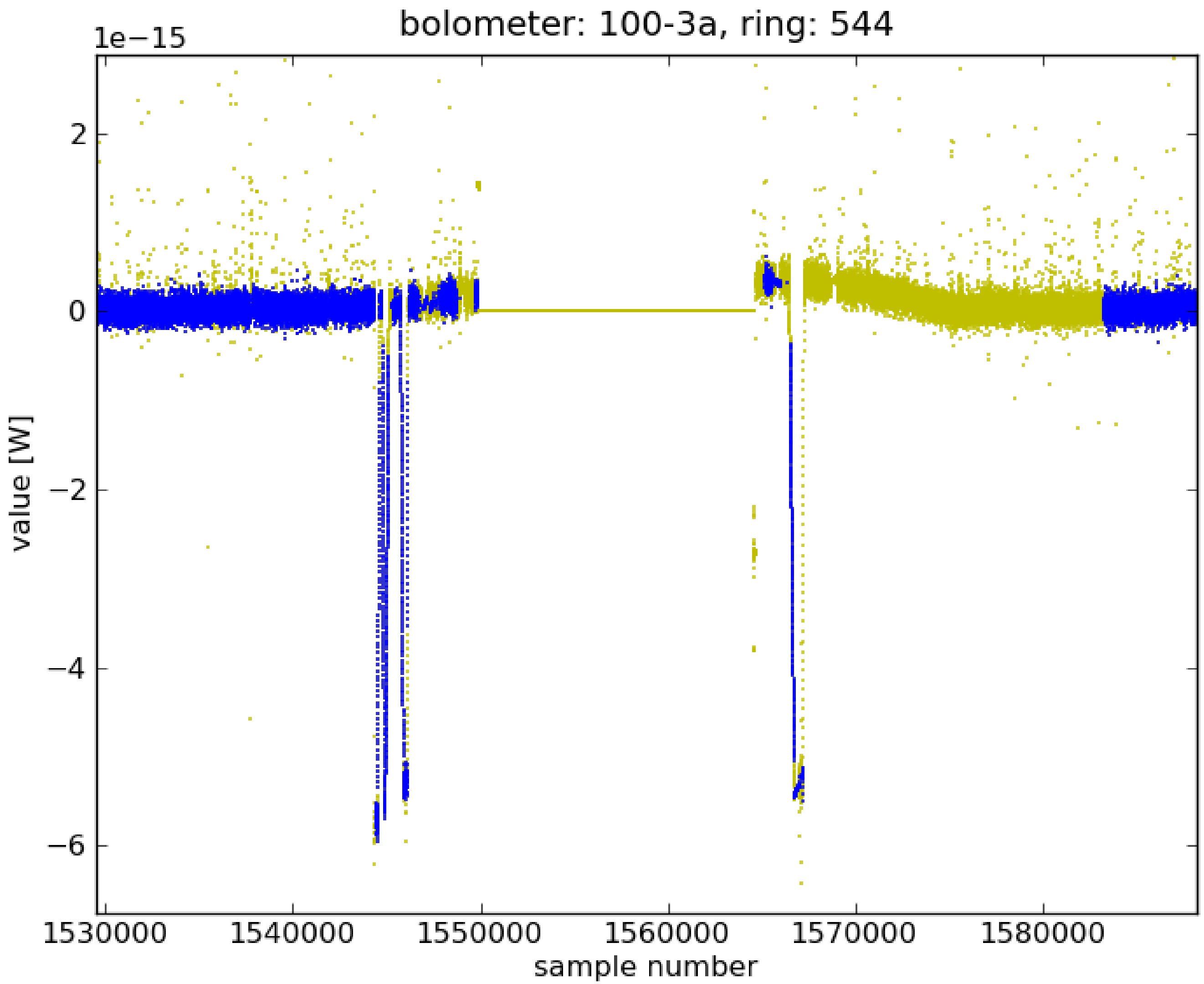
This should be white noise ...



This should be white noise …



This should be white noise ...

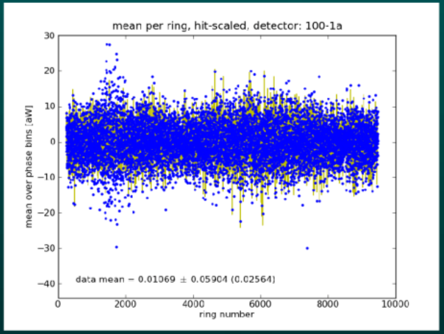


Jackknife PBR calculations

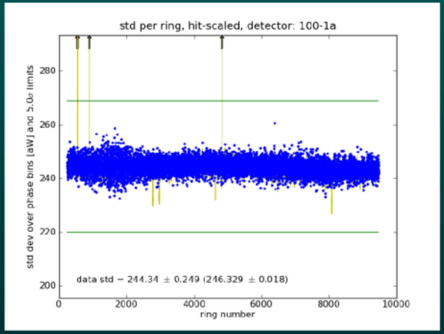
- Calculate various stats per ring and look for outliers
- Currently flagging rings for outliers in std dev, skewness, kurtosis

TEST EVERYTHING!
EXPECT THE UNEXPECTED!

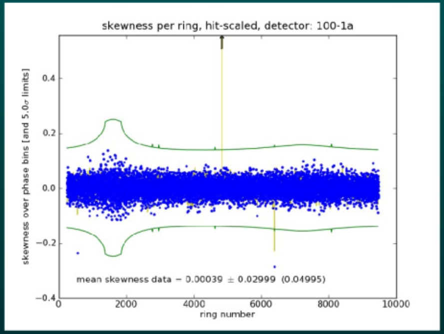
Mean



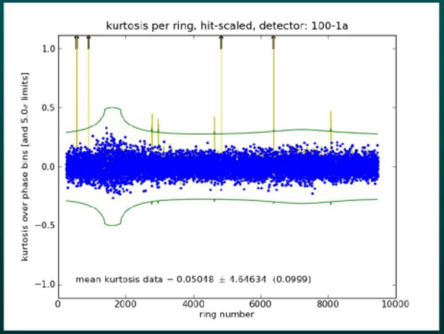
Standard deviation



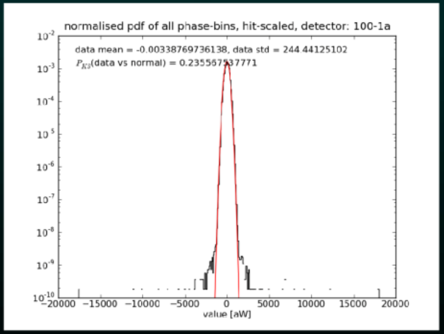
Skewness



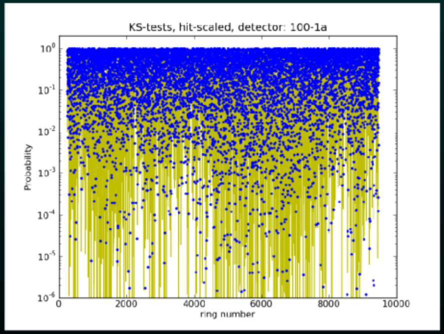
Kurtosis



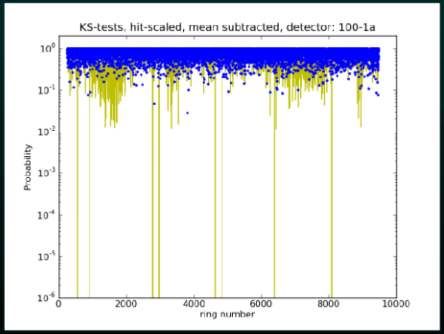
PBR histogram



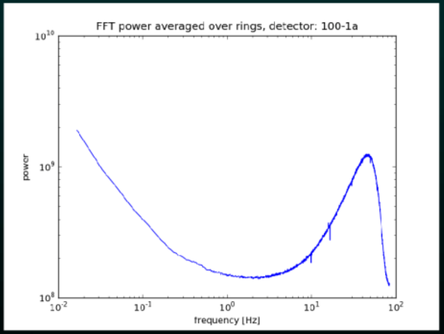
KS test



Mean-subtracted KS test



Average PSD



Is it unexpected 'new physics'?

No-one will believe you

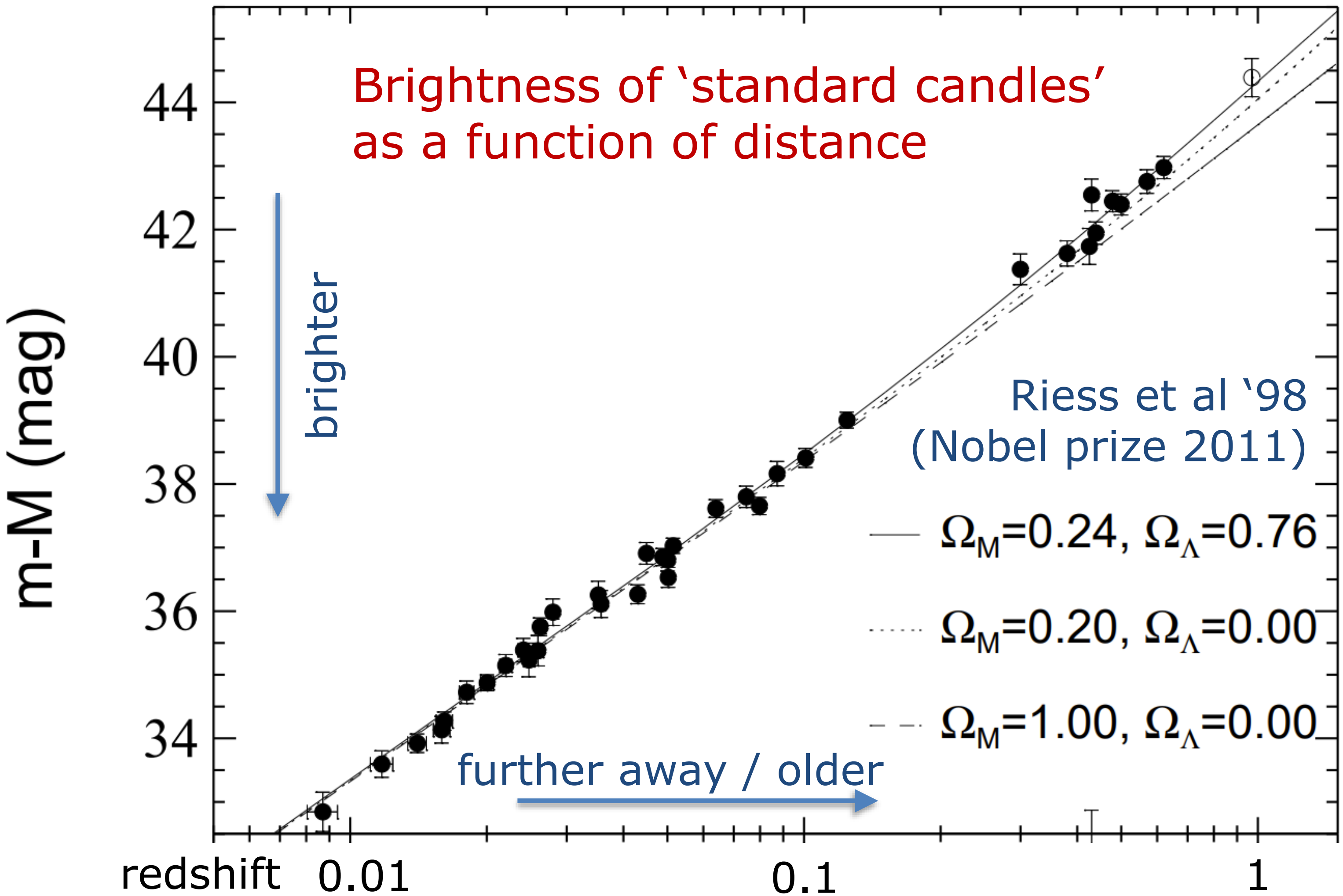
- In the 1990's there were multiple claims that the then-standard model of cosmology was not compatible with data
- For example from the distribution of galaxies in the Universe, or from the observed sizes of radio galaxies – but no-one trusted that data
- Here the opinion of the community in 1995:

(telescoper.wordpress.com)

scientific revolutions

But the scientific method works:

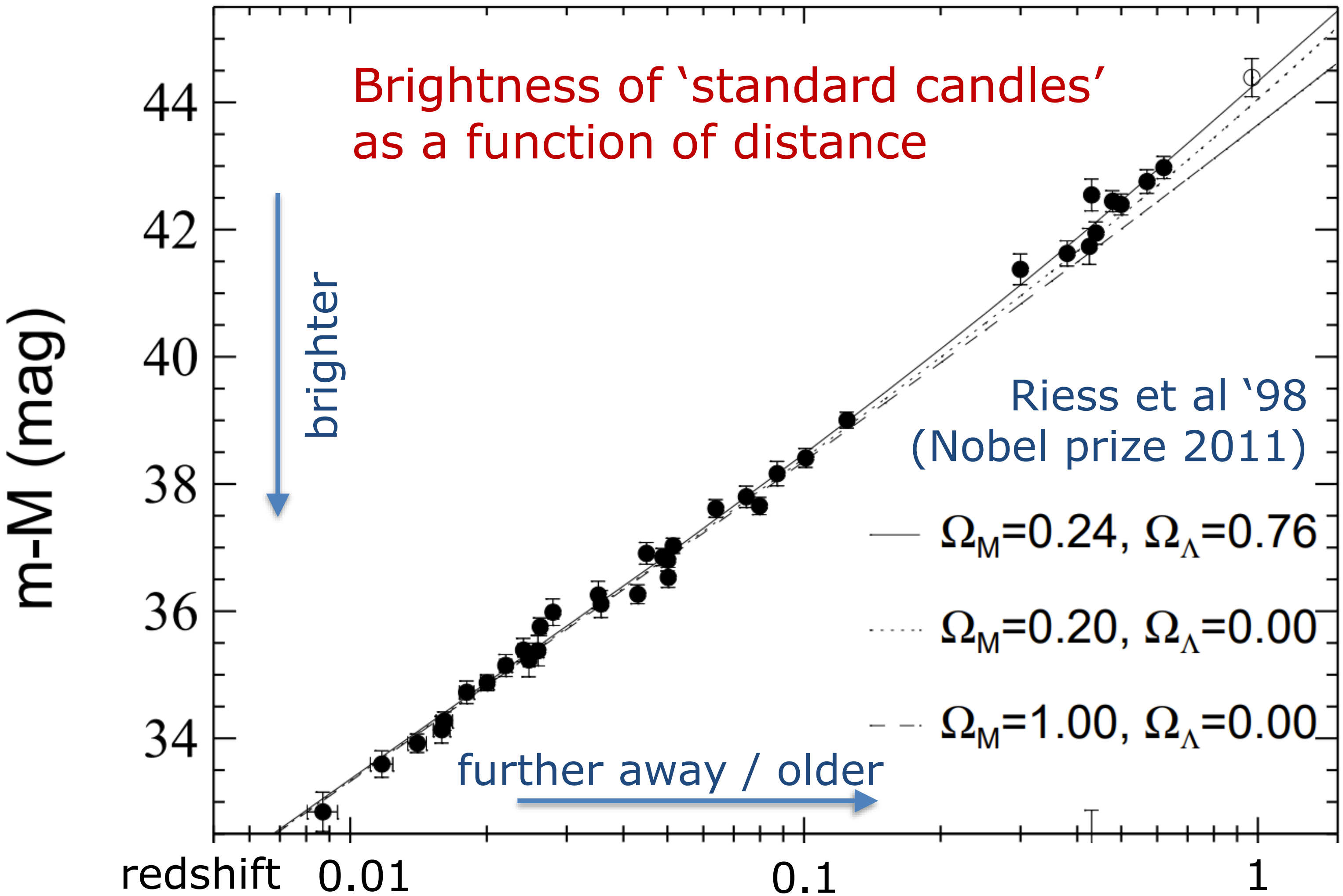
- Evidence accumulates and eventually new data triggers a paradigm 'phase transition' – trust in the data is critical!



scientific revolutions

But the scientific method works:

- Evidence accumulates and eventually new data triggers a paradigm 'phase transition' – trust in the data is critical!



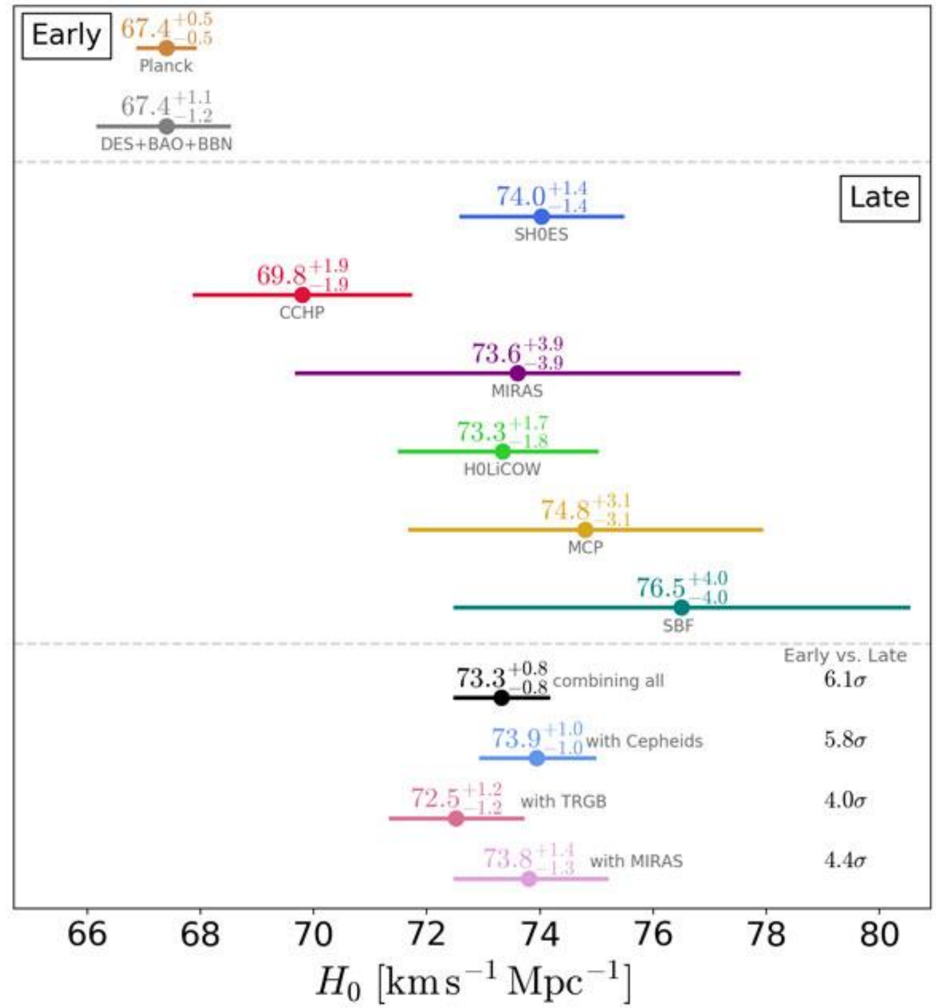
And how about today?

Enigma of the day: Different observations find a different expansion rate of the Universe!

- Either some observations are wrong, or the model that connects them is wrong.
- We don't know yet…

- All of these observations were made taking outmost care to avoid problems.
- But we all know that there are surprises.

So far the community is skeptical as to whether this is 'new physics'



Vivien Bonvin, Sky&Telescope 2019


Thank you
X. Ferrara, Dec 2014