National Physical Laboratory

Centre for Carbon Measurement

Testing and validating an uncertainty budget

Teresa Goodman

4 April 2017

fiducial reference measurements for satellite ocean colour

Metrology for Earth Observation and Climate

http://www.emceoc.org

EMRP

European Metrology Research Programme

Programme of EURAMET

The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union

EURAMET

European Association of National Metrology Institutes

Recap: Calculating the measurement uncertainty

$$u_c^2(y) = \boxed{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)} + \boxed{2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i, x_j)}$$

Need to know / work out:

- What affects the measurement result?
- How big is the uncertainty associated with each of these effects?
- How sensitive is the result to each of these effects?
- Are effects correlated?

We have the answer!

- “Plug everything into the equation”
- And we have our uncertainty!

$$u_c^2(y) = \boxed{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)} + \boxed{2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i, x_j)}$$

Or do we?

How do we know we’ve got the ‘right’ uncertainty?

Testing and validating an uncertainty budget

- Make a new experiment to achieve the same end point
- Change as much as possible between the two experiments
- Work out new uncertainty budget
- Do they agree within uncertainties?

Testing and validating an uncertainty budget

NPL National Physical Laboratory | Centre for Carbon Measurement

- Repeat the measurement on another occasion; vary as much as possible
  - Use different reference(s)
  - Use different instrument(s)
  - Get someone else to do the measurement
  - Change the order of the measurements

- Use a different measurement method e.g.
  - Calibration directly against another radiance gauge rather than lamp-tile combination
  - Array spectrometer or scanning instrument

- Compare with someone else
  - Compare results, uncertainties and uncertainty budgets
  - Helps quantify ‘known unknowns’
  - Helps identify ‘unknown unknowns’

Testing and validating an uncertainty budget

- Repeat the measurement on another occ… possible
  - Use different reference(s)
  - Use different instrument(s…
  - Get someone else to…
  - Change the ord…
- Use a diff…
  - C… …nce gauge rather than lamp-
  - … …g instrument
- Compa… …e else
  - Comp… …ults, uncertainties and uncertainty budgets
  - Helps quantify ‘known unknowns’
  - Helps identify ‘unknown unknowns’

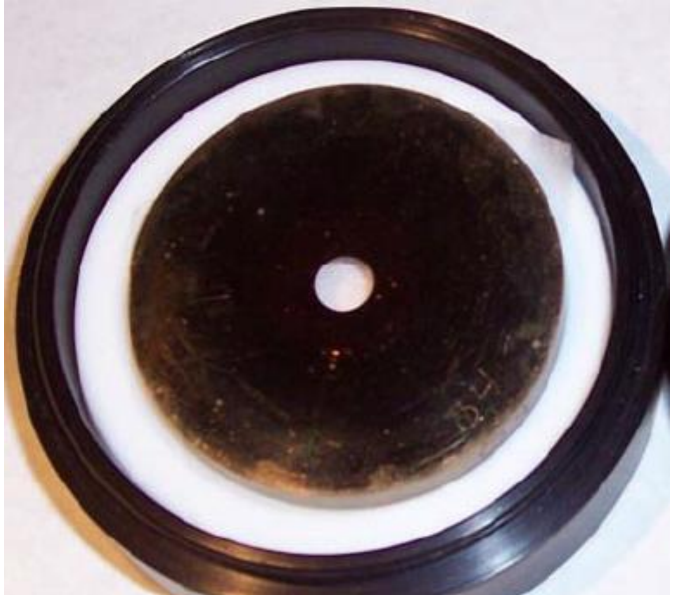
Use a range of artefacts

Example 1: Measuring aperture area

Example 1: Measuring aperture area

Comparing methods

Optical method – scan laser beam over surface



Source of Uncertainty	Value (μm)	Value/√3 (μm)
Interferometer drift	±0.02	±0.012
Inteferometer alignment	±0.03 (5 mm ϕ)	±0.017
	±0.12 (20 mm ϕ)	±0.069
	±0.15 (25 mm ϕ)	±0.087
Laser stability	±0.04	±0.023
Environmental conditions	**	**
Resolution	±0.06	±0.035
Combined uncertainty		
Nominal 5 mm ϕ	±0.11	±0.05
Nominal 20 mm ϕ	±0.16	±0.08
Nominal 25 mm ϕ	±0.19	±0.10

Example 1: Measuring aperture area

Comparing methods

Contact method – CMM: mechanically scan ball-ended stylus over surface

Source	µm	ppm
Machine scale uncertainty	0.04	
Temperature difference in beam paths during calibration		0.01
Laser frequency difference		0.02
Measurement Reproducibility	0.04	0.04
Edlen Equation		0.03
Index of Refraction-Air Temperature		0.01
Index of Refraction –Air Pressure		0.04
Index of Refraction – Humidity		0.03
Thermal Expansion		0.05
Coefficient of Thermal Expansion		0.05
Contact Deformation	0.002	
Gage Surface Geometry	0.004	
Gage Form Estimation Technique	0.005-0.100	
U (um) = 0.11 µm +0.20x10^{-6} L (k=2) best		

Do the results agree?

$$E_N = \frac{|E_1 - E_2|}{\sqrt{{U_1}^2 + {U_2}^2}}$$

E is measurement result

U is associated expanded uncertainty

$E_N < 1.0$ indicates results agree with each other within the limits expected based on their associated uncertainties

Do the results agree?

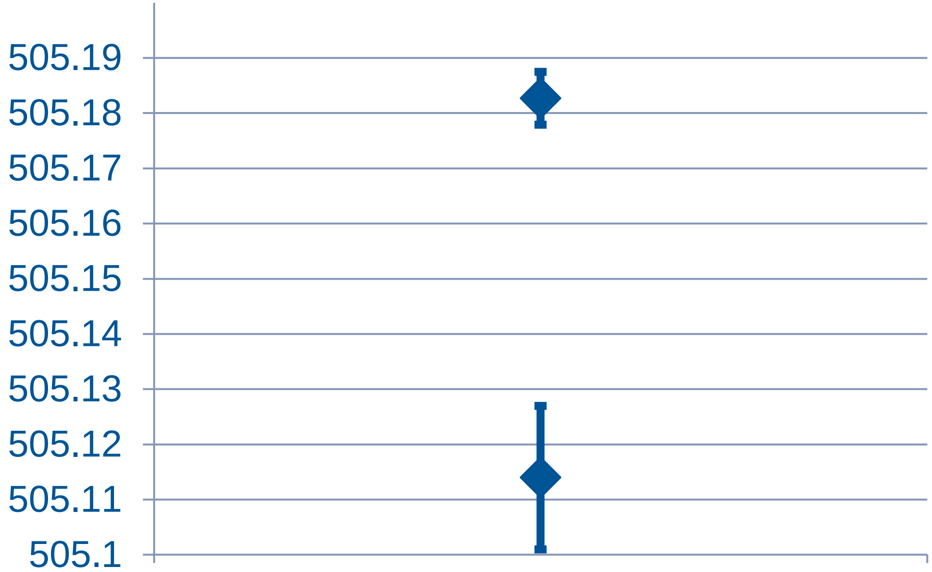
Error bars $k = 1$

$$E_N = 0.4$$

Results agree to well within uncertainties

⇨ Uncertainties (probably) correctly evaluated

Do the results agree?



Error bars k = 1

$$E_N = 2.5$$

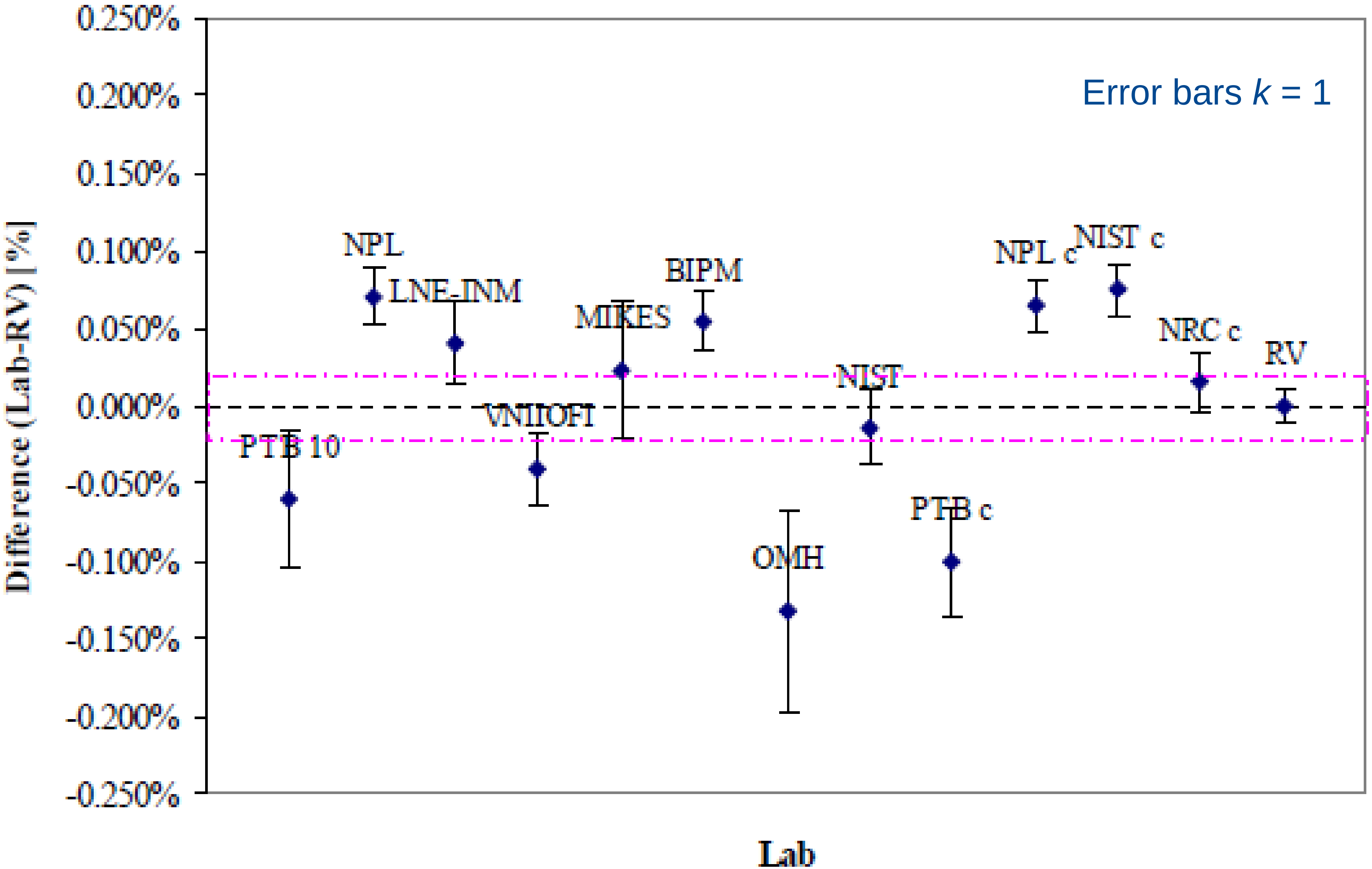
Results differ by much more than uncertainties

⇨ Uncertainties underestimated and/or some contributions missing

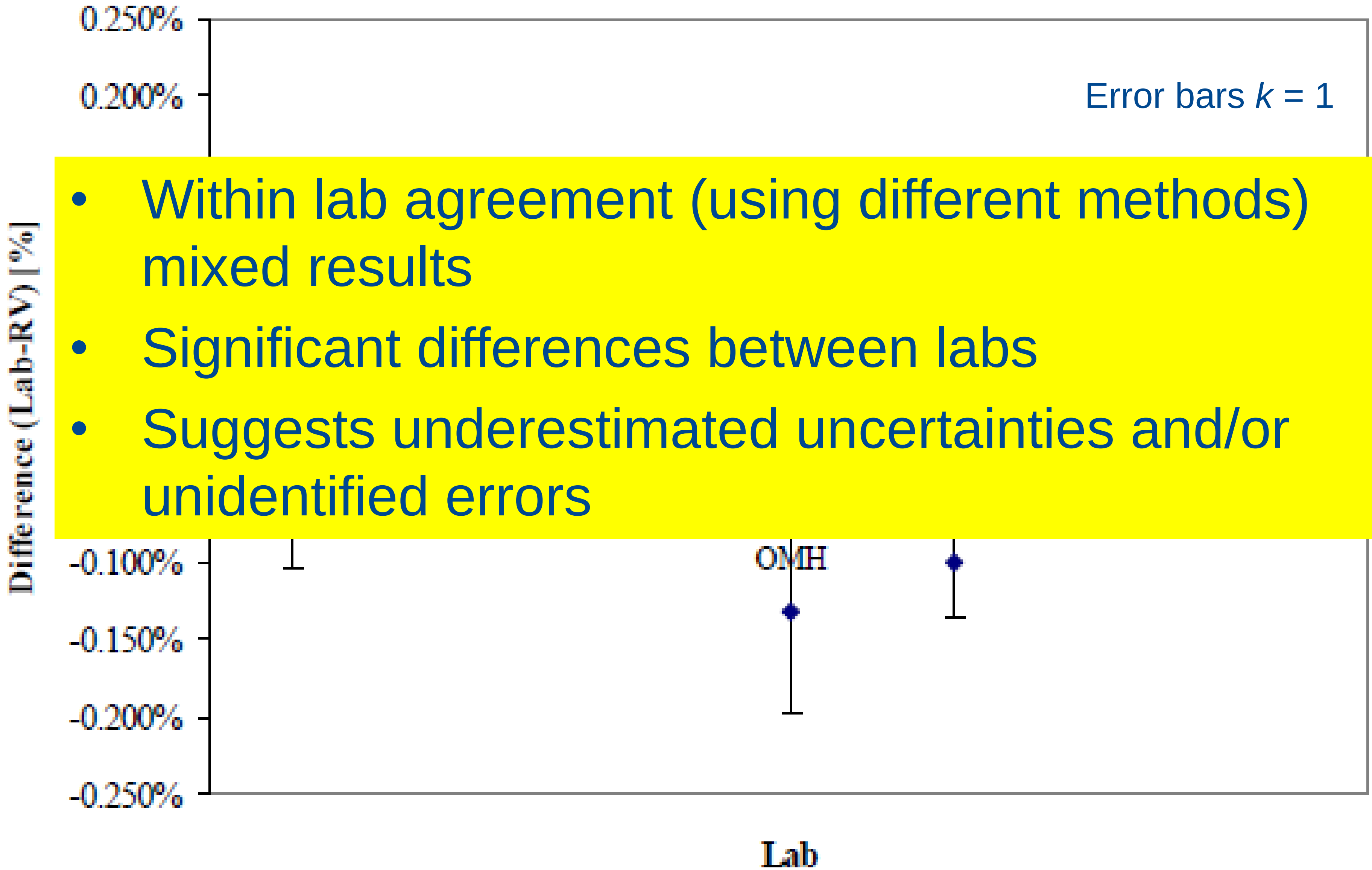
Example 1: Measuring aperture area

Comparing with another lab

Do the results agree?

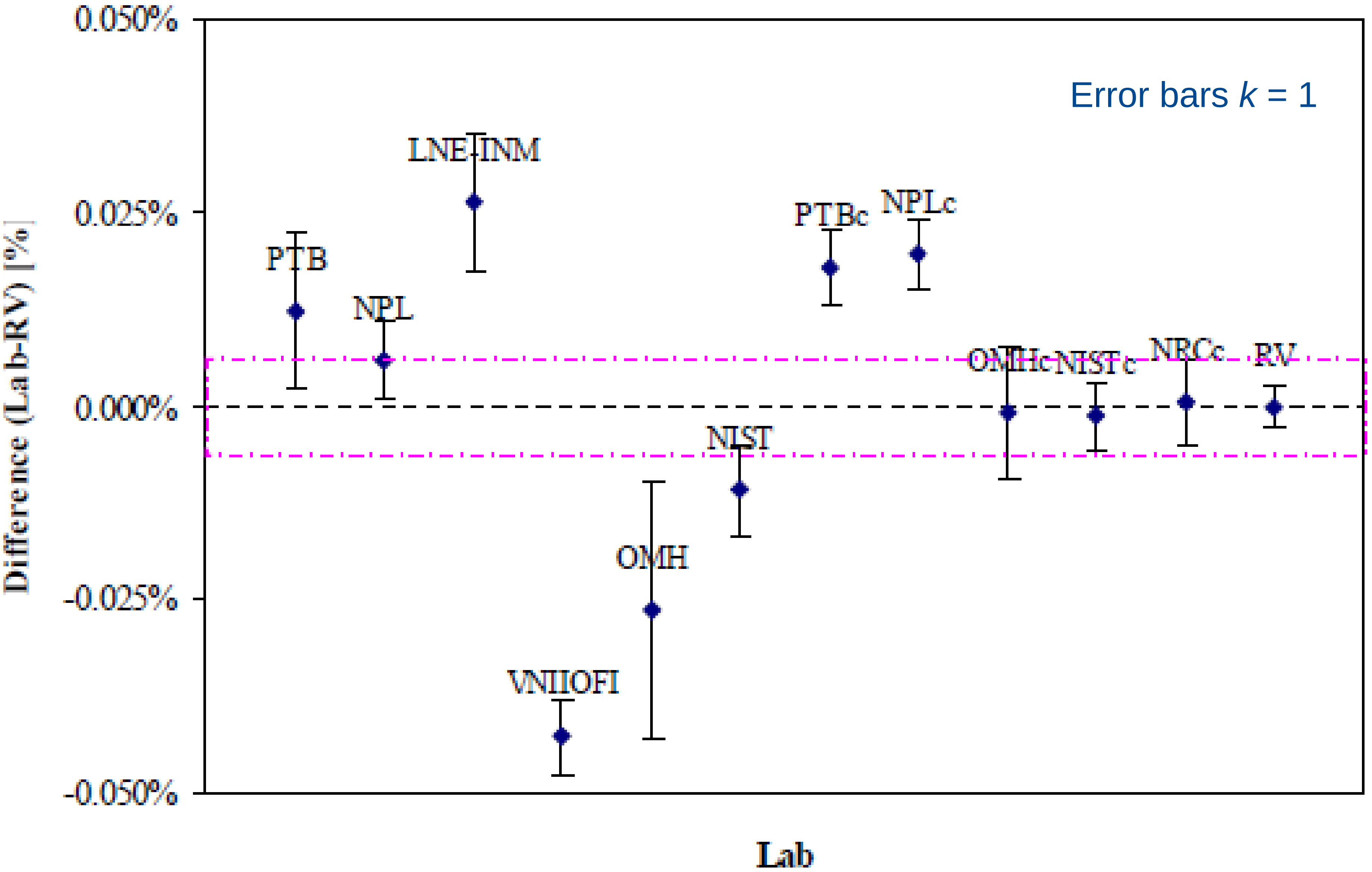


Do the results agree?

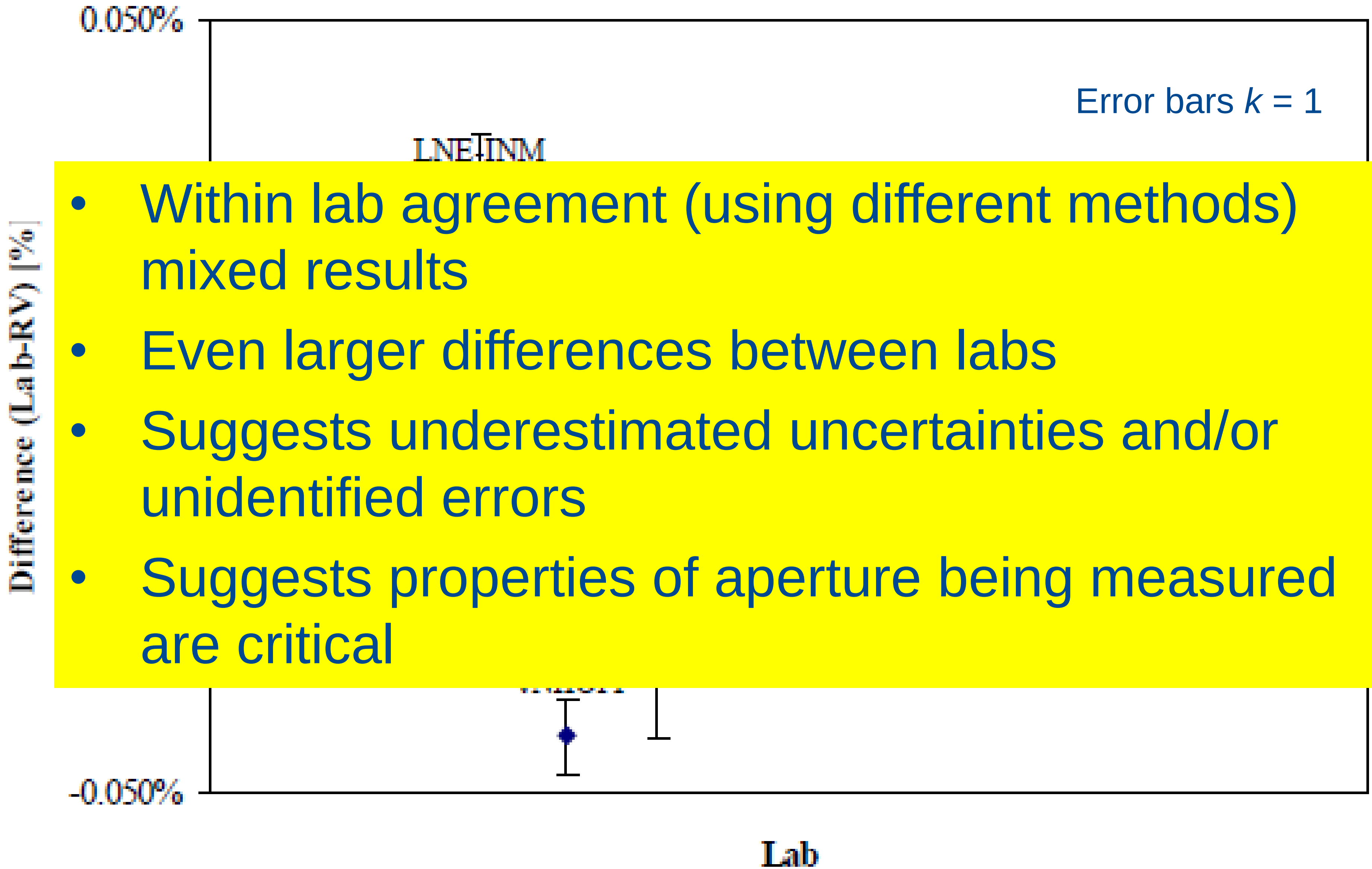


- Within lab agreement (using different methods) mixed results
- Significant differences between labs
- Suggests underestimated uncertainties and/or unidentified errors

Do the results agree?



Do the results agree?

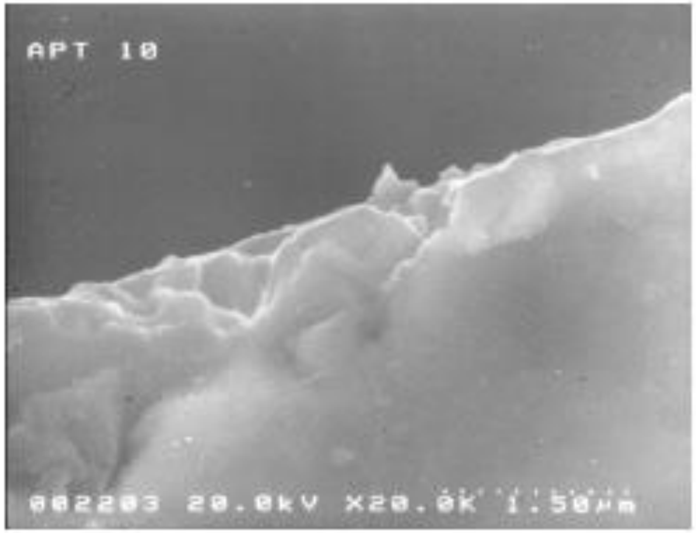


- Within lab agreement (using different methods) mixed results
- Even larger differences between labs
- Suggests underestimated uncertainties and/or unidentified errors
- Suggests properties of aperture being measured are critical

Example 1: Measuring aperture area

NPL National Physical Laboratory | Centre for Carbon Measurement

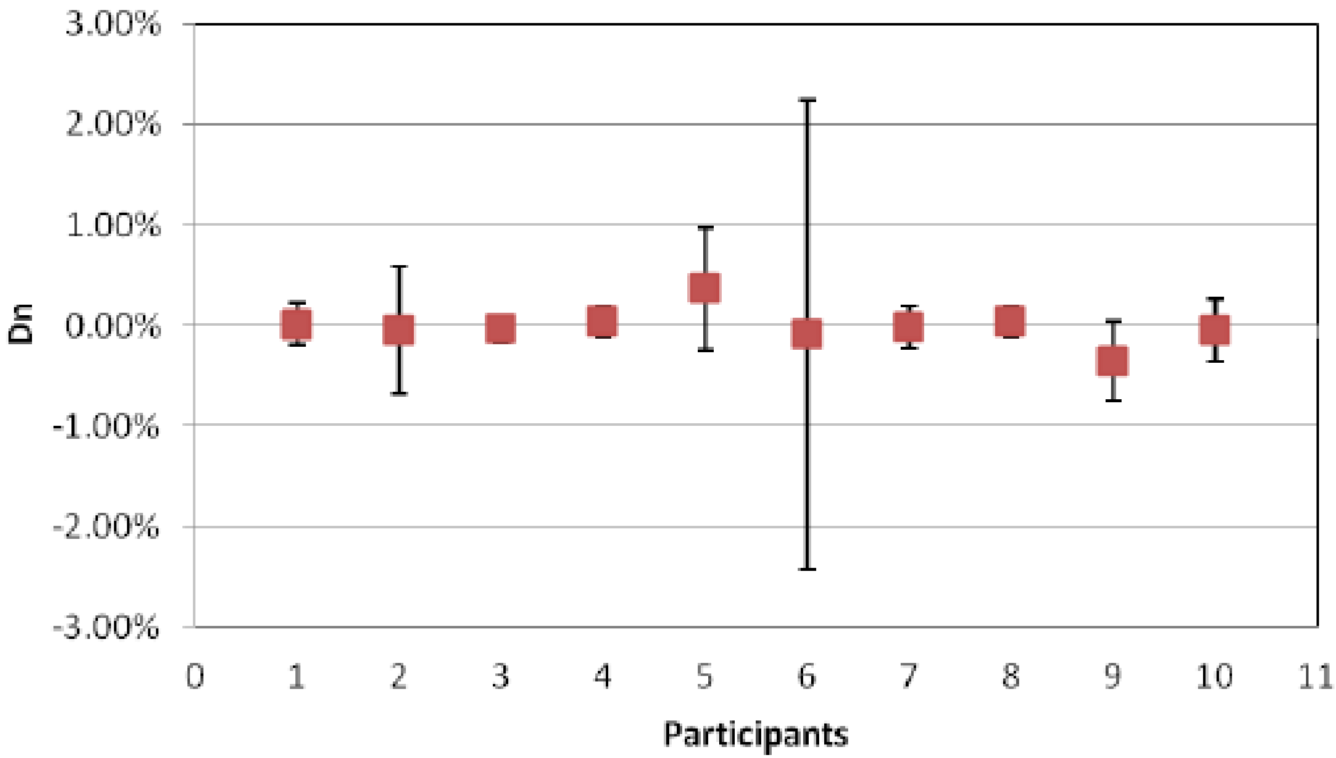
Conclusions



- Even where a lab had good agreement internally between different methods, agreement with other labs was not good
- Uncertainties underestimated by most labs
- Using a range of artefacts revealed additional differences and helped understand measurements better

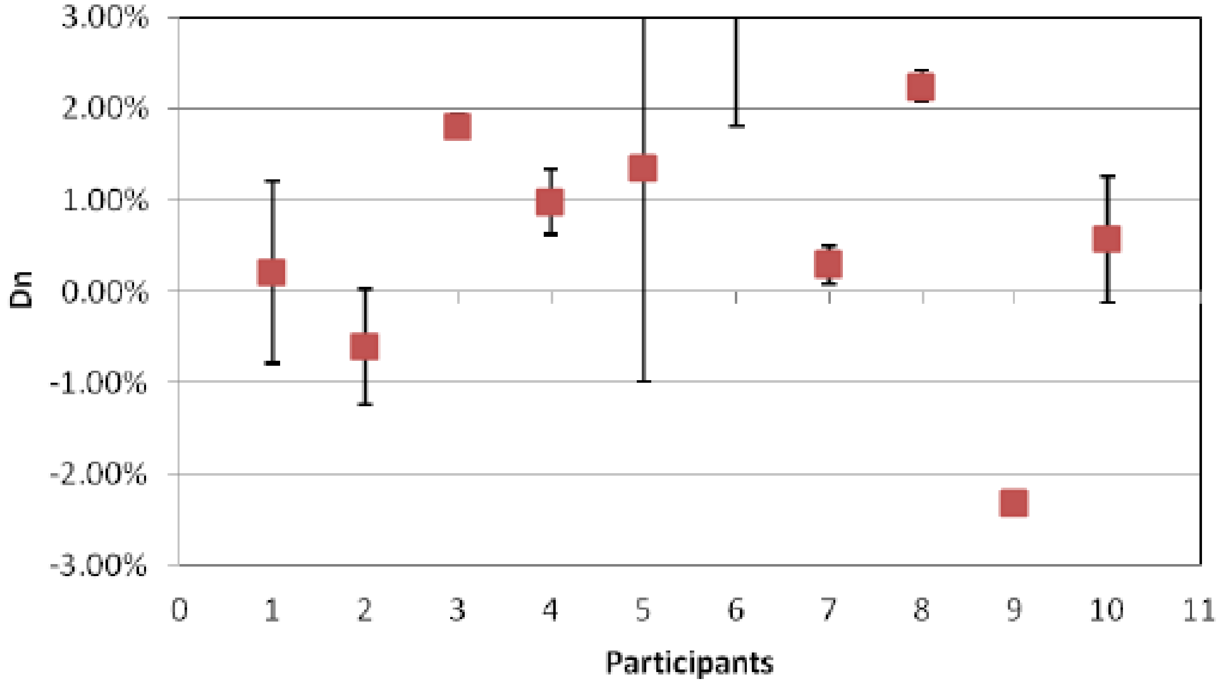
Example 2: Electrical power of LED lighting products

Example 2: Electrical power of LED lighting products



- All results are in agreement
- Large variation in claimed uncertainties

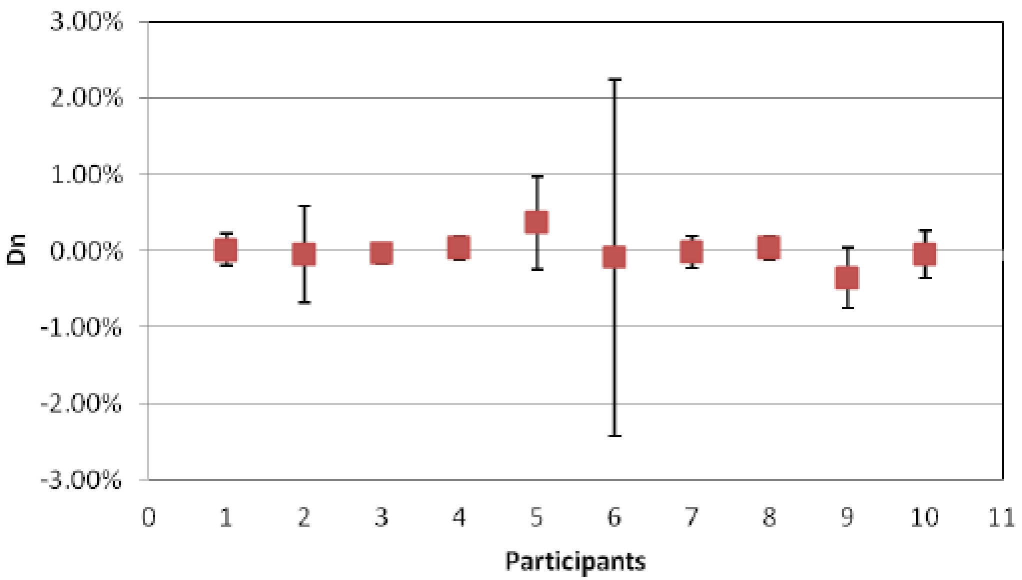
Example 2: Electrical power of LED lighting products

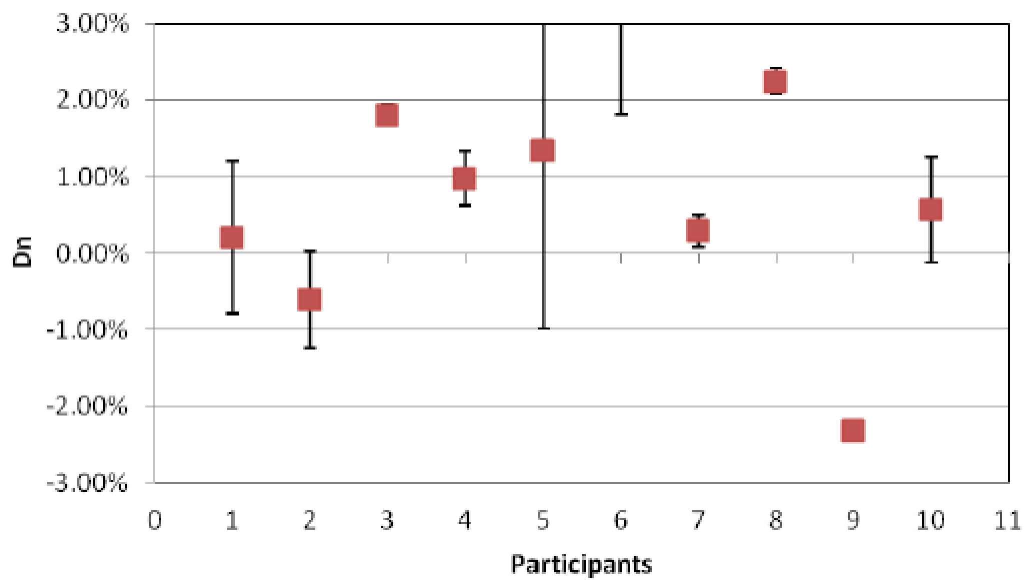


- Only a few results are in agreement
- Large variation in claimed uncertainties
- Many uncertainties appear to be underestimated

Example 2: Electrical power of LED lighting products

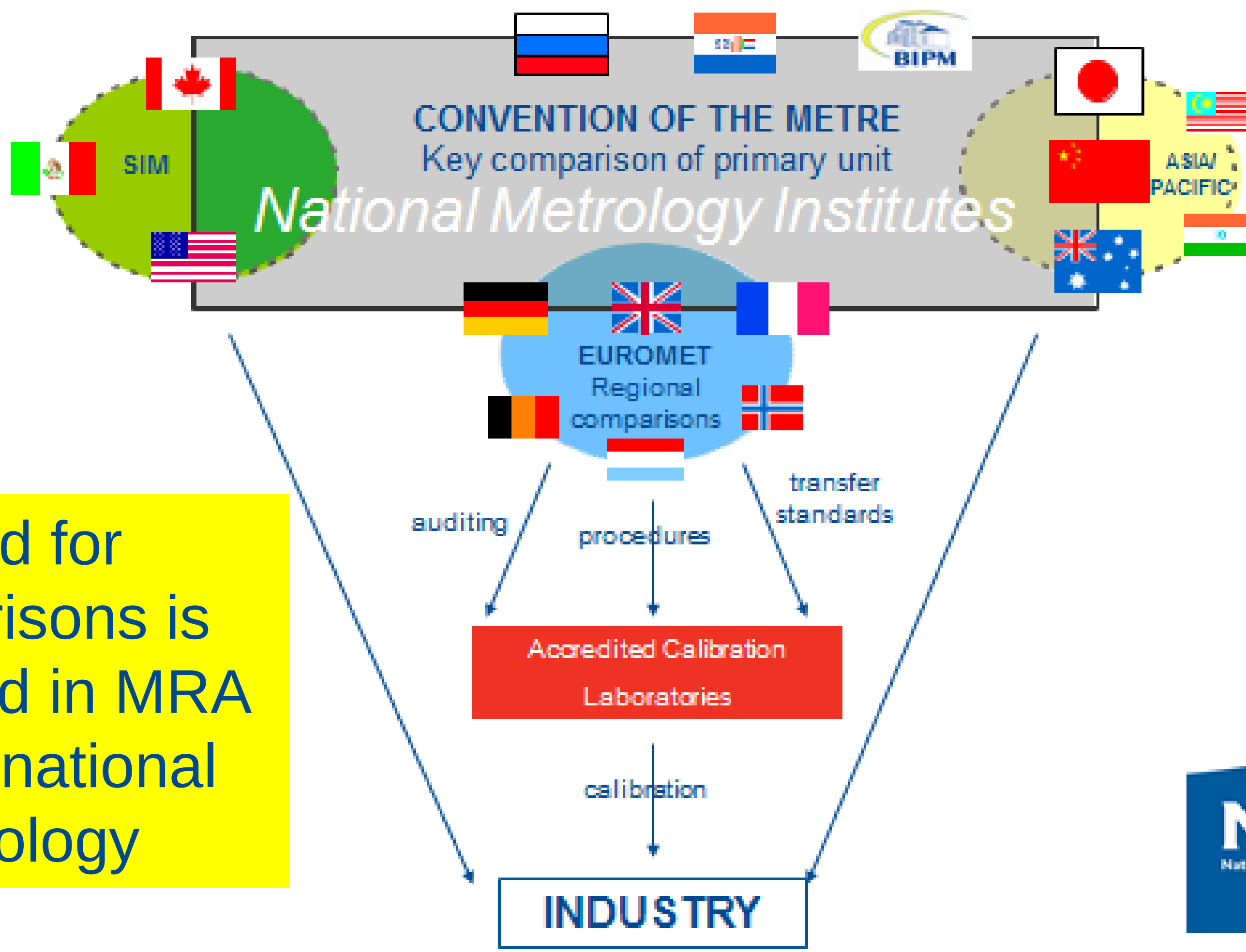
- Uncertainties appear ‘correct’ for one lamp, but generally underestimated for the other
- Using a range of artefacts helps understand measurements better
- Errors would not have been known without comparison





International system for assuring quality of results

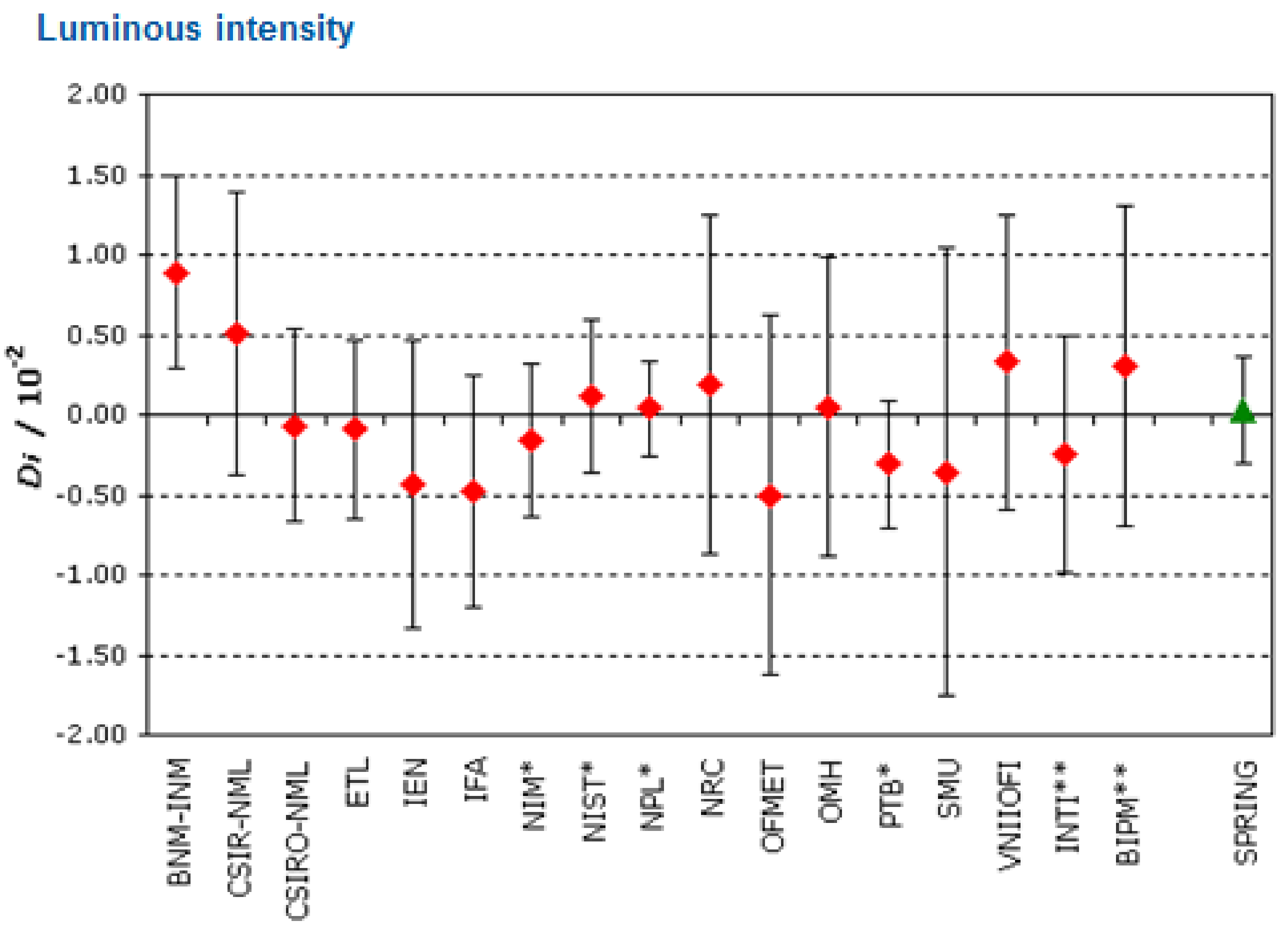
SI traceability



Need for comparisons is enshrined in MRA for international metrology

International system for assuring quality of results

'Rules' have been established by NMIs for conducting and analysing comparisons



Agreed rules for NMI comparisons

- Must define artefacts and set up to be used in advance ('technical protocol')
  - Minimise additional uncertainties due to comparison itself
  - Use more than one artefact type if possible
- Participants use own traceability routes and methods
  - Reveal differences due to these effects
- Participants provide detailed uncertainty budgets
  - Allows sharing of ideas and methods
- Blind comparison
  - Ensures reliability and integrity of comparison results
- Agree in advance how to present results e.g. what is comparison reference value
  - Avoids debates on how to do this!

Framework for assuring quality of EO data

A QUALITY ASSURANCE FRAMEWORK FOR EARTH OBSERVATION

- The Quality Assurance for Earth Observation (QA4EO) framework
- Looks to make the GUM accessible to the EO community

Community-specific guidelines Click to close

Identifier	Description
QA4... CLP...	
QA4... CLP...	
QA4... CLP...	
QA4EO-WGCV-IVO-CLP-004	A best practice guide to land "test-site" characterisation
QA4EO-WGCV-IVO-CLP-005	Guidelines for the identification and set-up of new systems complying with requirements for satellite ocean color applications
QA4EO-WGCV-IVO-CLP-006	Methodologies that should be applied to determine immersion factors for both radiance and irradiance underwater sensors
QA4EO-WGCV-IVO-CLP-007	Absolute Calibration using Rayleigh Scattering
QA4EO-WGCV-IVO-CLP-008	Protocol for the CEOS WGCV pilot Comparison of techniques/instruments used for vicarious calibration of land surface imaging through a ground reference standard test site

QA4EO Principle:
'All data and derived products shall have associated with them a fully traceable indicator of their quality'

NPL National Physical Laboratory Training

Framework for assuring quality of EO data

NPL National Physical Laboratory | Centre for Carbon Measurement

QA4EO A QUALITY ASSURANCE FRAMEWORK FOR EARTH OBSERVATION

The QA4EO PRINCIPLES provides the background to QA4EO and introduces the key guidelines:

QA4EO-QAEO-GEN-DQK-001	A guide to establish a Quality Indicator on a satellite sensor derived data product
QA4EO-QAEO-GEN-DQK-002	A guide to content of a documentary procedure to meet the Quality Assurance requirements of CEOS
QA4EO-QAEO-GEN-DQK-003	A guide to "reference standards" in support of Quality Assurance requirements of QA4EO
QA4EO-QAEO-GEN-DQK-004	A guide to comparisons – organisation, operation and analysis to establish measurement equivalence to underpin the Quality Assurance requirements of QA4EO
QA4EO-QAEO-GEN-DQK-005	A guide to establishing validated models, algorithms and software to underpin the Quality Assurance requirements of QA4EO
QA4EO-QAEO-GEN-DQK-006	A guide to expression of uncertainty of measurements
QA4EO-QAEO-GEN-DQK-007	A guide to establishing quantitative evidence of traceability to underpin the Quality Assurance requirements of QA4EO

Guidelines are based on NMI 'rules' for conducting and analysing comparisons

FRM4SOC LCE-1

- Purpose is to verify the performance of reference irradiance sources used in calibration of Ocean Colour Radiometers
  - All are FEL-type, but several different manufacturers
- Participants use own traceability routes; all traceable to SI
  - Confirm consistency between these routes
- Lamps have been used for varying amounts of time since calibration
  - Allow for up to 50 hours ageing in uncertainty budget for comparison

Analysis of FRM4SOC LCE-1

- Want to confirm consistency of traceability routes, not to establish a single reference scale for all calibrations

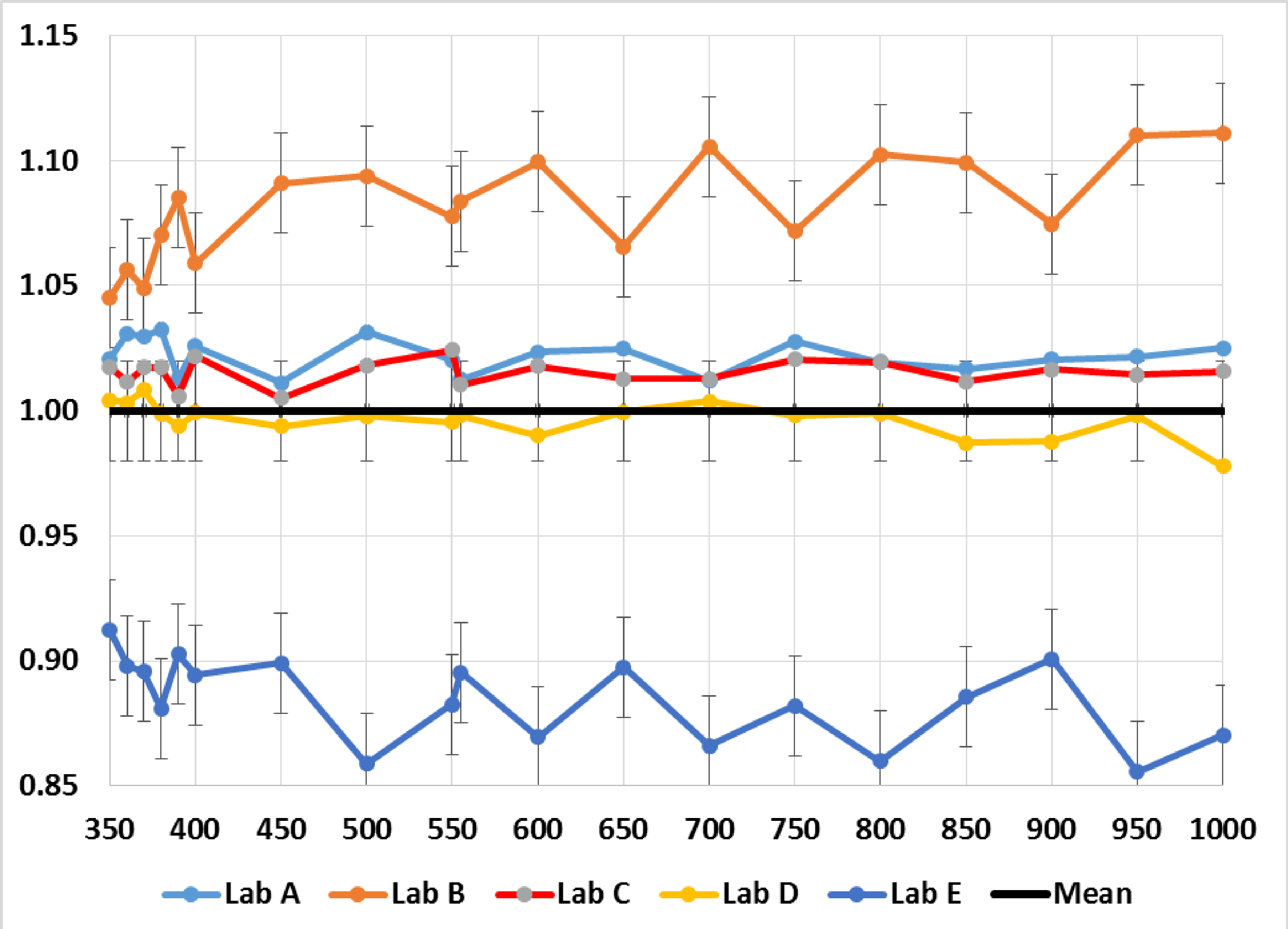
  Will compare each individual lamp to the mean of all lamps
- Uncertainty associated with each lamp will be combination of:
  - Assigned calibration uncertainty (from certificate)
  - Uncertainty allowance for up to 50 hours ageing
  - Measurement uncertainties associated with the comparison at NPL (e.g. system noise and stability, lamp current, lamp alignment, distance setting)
- Results will be consistent if all lamps agree with mean value to within the associated uncertainties

Analysis of FRM4SOC LCE-1

- Want to confirm consistency of traceability routes, not to establish a single reference scale for all calibrations

  Will compare each individual lamp to the mean of all lamps
- Uncertainty associated with each lamp will be combination of:
  - Assigned calibration uncertainty (from certificate)
  - Uncertainty allowance for up to 50 hours ageing
  - Measurement uncertainties associated with the comparison at NPL (e.g. system noise and stability, lamp current, lamp alignment, distance setting)
- Results will be consistent if all lamps agree with mean value to within the associated uncertainties

Analysis of FRM4SOC LCE-1



Things to take away from this module

- It's important to test / validate calculated uncertainties
  - Best done experimentally
  - Repeat the measurement on another occasion; vary as much as possible
  - Use a different measurement method
  - Compare with someone else

- Internal consistency can give a false sense of security
  - Ultimate test is always to compare with other labs
  - Looking at other people's uncertainty budgets can give new ideas

- Using a range of artefacts can reveal unexpected issues
  - Helps to understand measurement system better
  - Can identify 'unknown unknowns'

- Comparisons help develop a 'gut feeling' about uncertainty
  - If calculated uncertainty 'feels too low', it often is!

Things to take away from this module

NPL National Physical Laboratory | Centre for Carbon Measurement

- It's important to test / validate calculated uncert
  - Best done experimentally
  - Repeat the measurement on anothe sible
  - Use a different measurement m
  - Compare with someone e

- Internal consis
  - Ultima
  - I e new ideas

- Us xpected issues
  - F ystem better
  - Ca ns'

- Comparis velop a 'gut feeling' about uncertainty
  - If calcul ncertainty 'feels too low', it often is!

Principles apply equally to other laboratories as to NMIs!