Verification & Validation

Rasmus E. Benestad

Winter School in eScience

Geilo January 20-25, 2013

3 double lectures

Rasmus.benestad@met.no

Objective

“reproducible science and modern techniques for scientific software development.”

There is a curse in validation - the ability to see others' faults – others can find it really annoying

My background

Physics training.

Experience with computer programming

Worked on replication of published results

Working with statistics

Practical approach – climate sciences

Contribution to debating climate sciences in terms of critical reviews – [RealClimate.org]

Lecture Questions

What is meant by **verification & validation**?

What is 'agnotology'?

Science & replication...?

What are models?

What information do we have?

Types of predictions?

How to assess skill?

Lecture Structure

3 x 2 lectures:

1a. General introduction (36)

1b. Scientific replication & Models (43).

2a. Physics-based evaluation (38)

2c. Statistics-based evaluation (45)

3a. Exercises (6).

3b. Re-cap: validation & verification (37)

Central themes in V&V

Proof
Tests
Evidence
Data
Replication
Hypotheses
Documentation

1a. General introduction

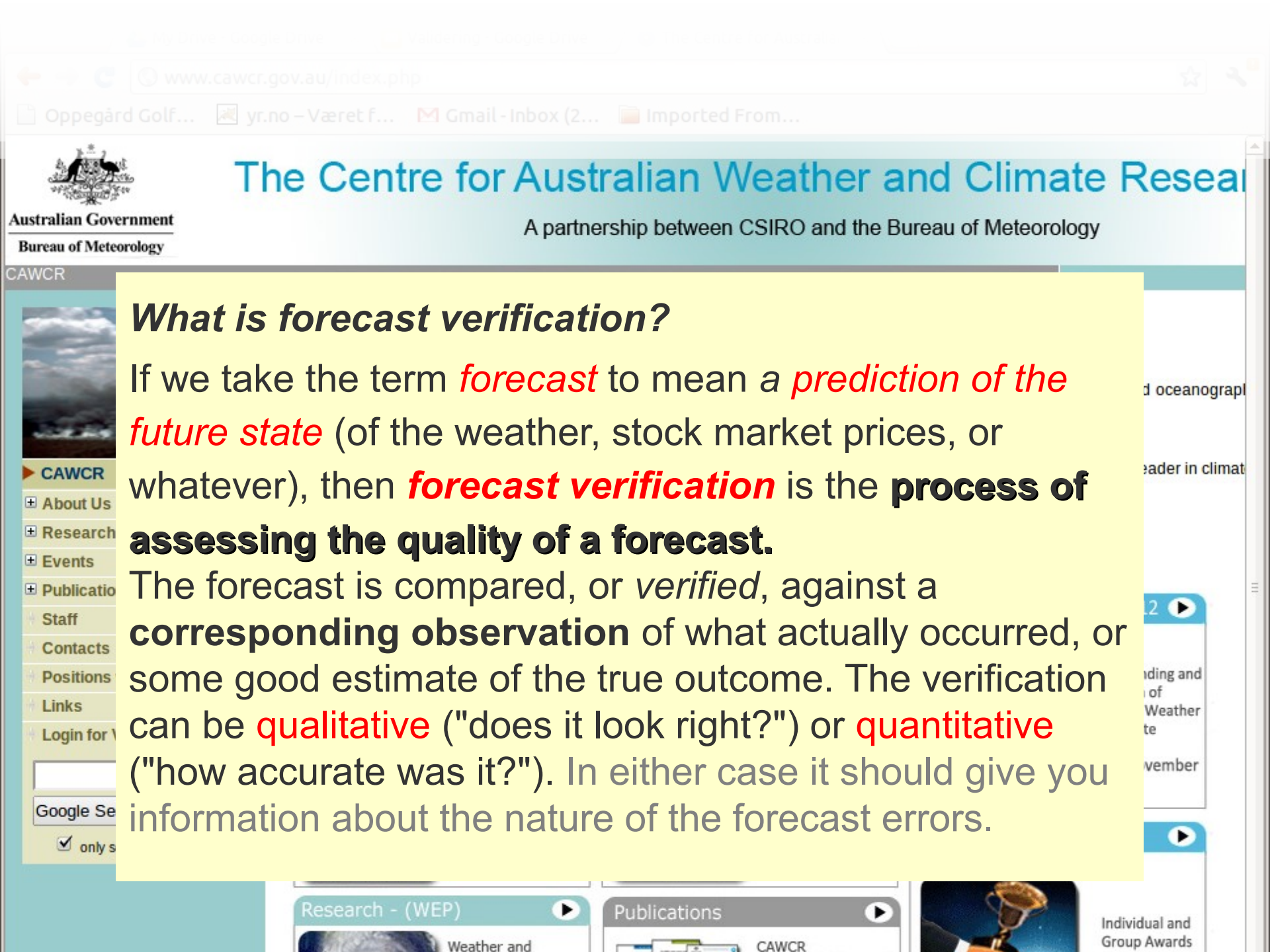
Verification & validation
What are they?

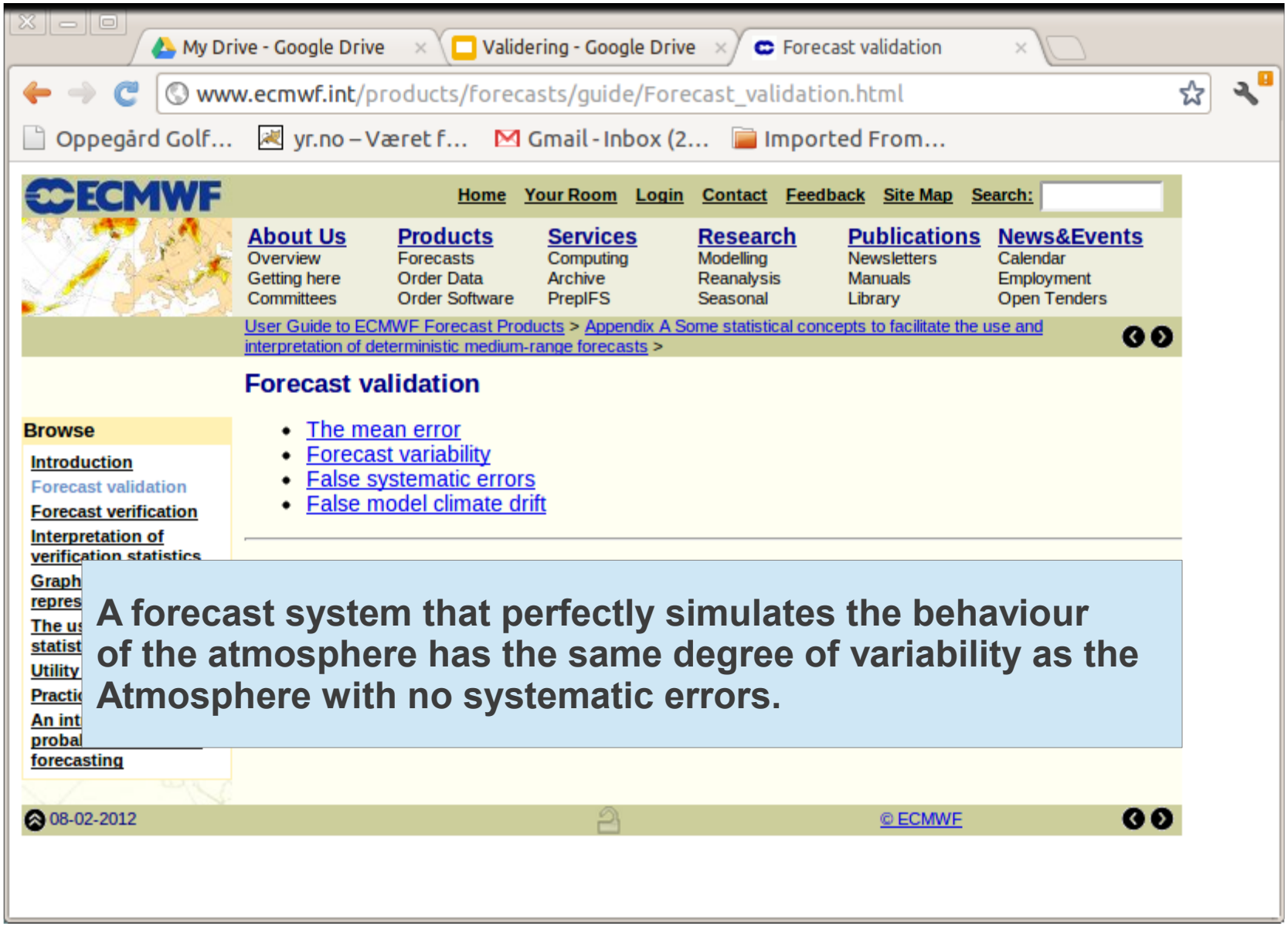
Con troll?

Wikipedia:

"Verify or **verification** may refer to:

- Verification and validation, in engineering or quality management systems, it is the act of reviewing, inspecting or testing, in order to establish and document that a product, service or system meets regulatory or technical standards.
- Verification (spaceflight), in the space systems engineering area, covers the processes of qualification and acceptance
- Verification theory, philosophical theory relating the meaning of a statement to how it is verified
- Third-party verification, use of an independent organization to verify the identity of a customer
- Authentication"


Australian Government
Bureau of Meteorology
The Centre for Australian Weather and Climate Resear
A partnership between CSIRO and the Bureau of Meteorology
CAWCR
CAWCR
About Us
Research
Events
Publicatio
Staff
Contacts
Positions
Links
Login for
Google Se
only s
What is forecast verification?
If we take the term forecast to mean a prediction of the future state (of the weather, stock market prices, or whatever), then forecast verification is the process of assessing the quality of a forecast.
The forecast is compared, or verified, against a corresponding observation of what actually occurred, or some good estimate of the true outcome. The verification can be qualitative ("does it look right?") or quantitative ("how accurate was it?"). In either case it should give you information about the nature of the forecast errors.
Research - (WEP)
Weather and
Publications
CAWCR
Individual and
Group Awards


My Drive - Google Drive
Validering - Google Drive
Forecast validation
www.ecmwf.int/products/forecasts/guide/Forecast_validation.html
Oppegård Golf…
yr.no – Været f…
Gmail - Inbox (2…
Imported From…
ECMWF
Home Your Room Login Contact Feedback Site Map Search:
About Us Overview Getting here Committees
Products Forecasts Order Data Order Software
Services Computing Archive PrepIFS
Research Modelling Reanalysis Seasonal
Publications Newsletters Manuals Library
News&Events Calendar Employment Open Tenders
User Guide to ECMWF Forecast Products > Appendix A Some statistical concepts to facilitate the use and interpretation of deterministic medium-range forecasts >
Forecast validation
The mean error
Forecast variability
False systematic errors
False model climate drift
Browse
Introduction
Forecast validation
Forecast verification
Interpretation of verification statistics
Graph repres
The us statist
Utility
Practi
An int probal forecasting
A forecast system that perfectly simulates the behaviour of the atmosphere has the same degree of variability as the Atmosphere with no systematic errors.
08-02-2012
© ECMWF

Wikipedia:

"Validation may refer to:

- Verification and validation, in engineering,` confirming that a product or service **meets the needs of its users**
- Verification and validation (software), checking that a software system **meets specifications and fulfills its intended purpose**
- Validation of foreign studies and degrees, processes for transferring educational credentials between countries
- Validation (drug manufacture), documenting that a process or system meets its pre-determined specifications and quality attributes
- Data validation, in computer science, **ensuring that data inserted into an application satisfies defined formats** and other input criteria
- Regression model validation, in statistics, determining whether a **model fits the data well**
- XML validation, the process of checking a document written in XML to confirm that it both is "well-formed" and follows a defined structure
- Social validation, compliance in a social activity to fit in and be part of the majority
- the validation of a analytical test method, to show that it is **suitable for the purpose it is used**"

For all intents and purposes, 'verification' and 'validation' are used like synonyms.

Henceforth, these concepts will be abbreviated as '**V&V**'.

V&V is fairly fuzzy and can embrace a wide range of considerations.

V&V is the backbone of modern science and engineering.

Building knowledge

V&V involves knowledge and information seeking – examining, analysis, probing and testing.

Building a picture of the system, process, theory, or the situation.

Learning plays a central role.

Certification...

Outside the scope here...

There has been developed guide lines for certification of various products, organisations, etc.

E.g. ISO 9001.

Here, the emphasis will be on the science and the aspects of V&V in science.

Validate or verify what?

Results & forecasts

Theory

Method & algorithm

Data

How consistent is it with...?

Information

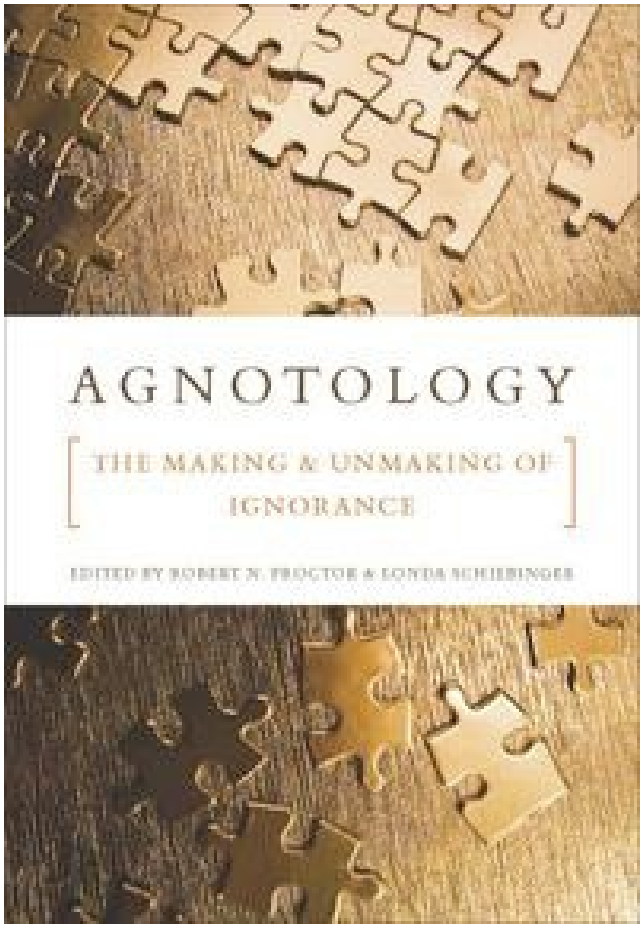
Wikipedia:

"...in its most restricted technical sense, is a sequence of symbols that can be interpreted as a message".

Related to the mathematical concepts of combinations and permutations. In statistics, the information content is related to the degrees of freedom (DOF):

"Estimates of statistical parameters can be based upon different amounts of information or data. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (df)".

Why we don't know things we don't know.



Is sufficient efforts devoted to replication, validation, & validation?

Pebesma et al (2012):
"*Reproducibility is an important aspect of scientific research, because the credibility of science is at stake when research is not reproducible*".

What don't we know, and why don't we know it?

Agnotology—the study of ignorance:

"Why don't we know what we don't know?"
"Ignorance is often more than just an absence of knowledge."

"Ignorance has a history and a political geography"

The connection between V&V and agnotology

- V&V is a process for which the objective it to gain specific knowledge and information.
- Question whether the specific knowledge can be obtained. Is it possible to validate?

“How do I know if something is true?”

Question: **What is relevant and what is not?**

How to proceed:

- Search relevant and independent information – journals are a good start.
- Google – Wikipedia – check other accounts. Not as reliable. Reflects mass perception.

“How do I know if something is true?”

Can you **trust the information**?

Agnotology – research provided by interest groups? (Climate change).

Bring together & make the connections.

Neutral and objective information: **ubiquitously valid**.

Trial & error.

“How do I know if something is good or different?”

Reference & benchmark – similar or not? Skill – better.

Question of expectation.

Depends on situation.

How did we come up with this piece of information or knowledge?

reproducible science

Missing piece of knowledge?

Interpretations

Data analysis
Predictions
Testings

Data control

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1 X_2} \cdot \sqrt{\frac{2}{n}}}$$

Source of information

Source of information

Science - universal truths

Different to 'audit' (specific case)

Universal truths **objective** and **robust**

When are results similar – when are they drawn from the same process?

Degree of complexity due to dependency on many factors

An '*audit*' case: Downhill skiing

Non-universal!

Δt=0.21s; t=110.5s. (**0.19%**). Difference between 8th & 9th rank: Δt=0.01s (**0.009%**). **What is the real precision & Errors?**

Equipment: ciruits, EM induction, transmission, temperature, humidity, random fluctuations?

Inbox - r... | Dealers ... | Laineet.... | 80 til 15... | Infrared... | Alpin...

www.nrk.no/sport/resultat/alpint-verdenscup/?seson

resultater alpin

Most Visited | Latest Headlines | Oppegård Golfklubb -... | yr.no – Været for Nor...

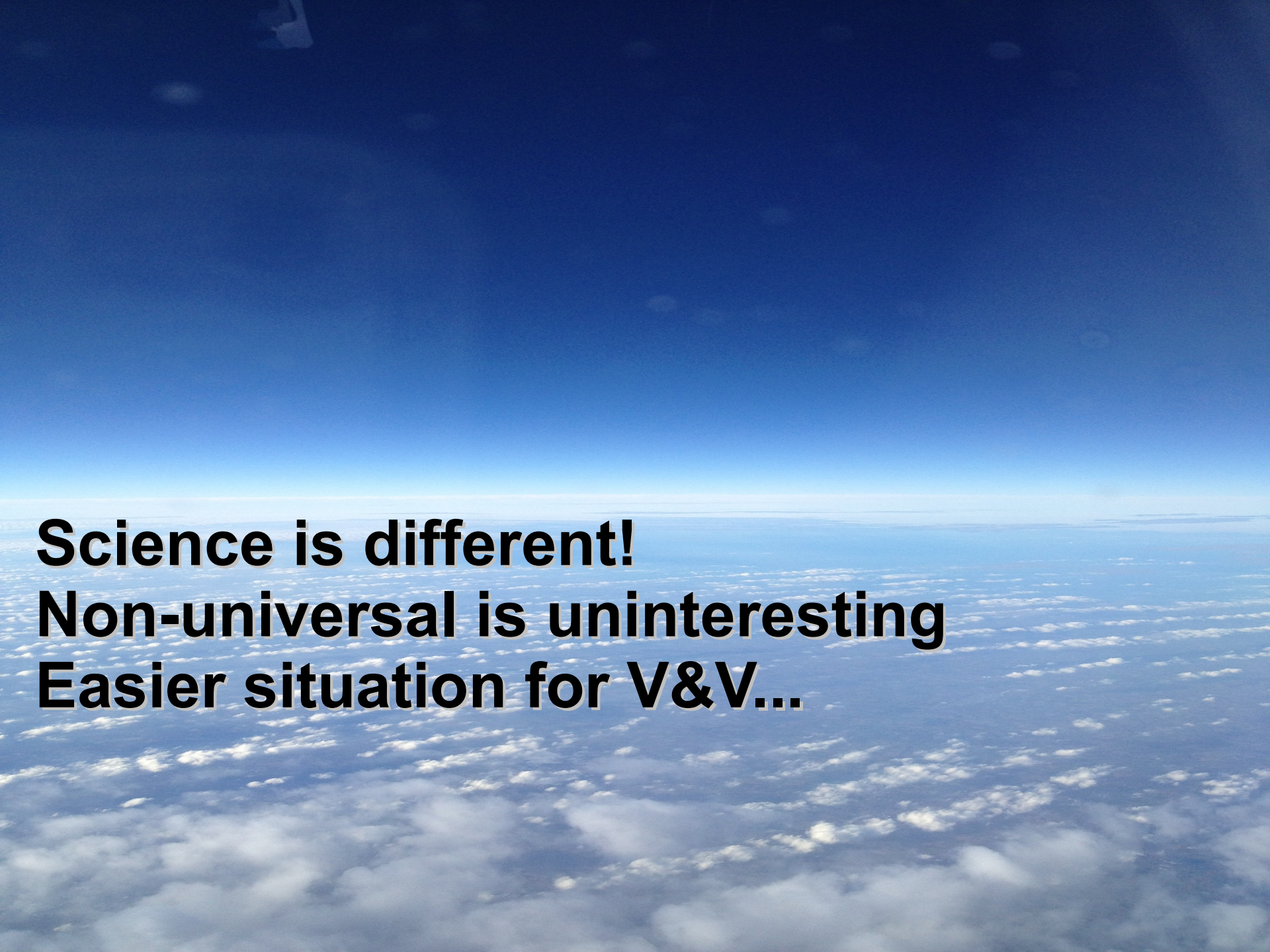
Alpint Verdenscup 2012/2013

Kombinasjon etter Slalåm (menn) | Kombinasjon utfor (menn)
Utfor (menn) | Slalåm 1. omgang (menn)
Slalåm 2. omgang (menn)

Slalåm 2. omgang finale - menn 20.01.13, 13:15

Pl	Navn	Resultat	Merk
1	Felix Neureuther	1:50.53	
2	Marcel Hirscher	+0.21	
3	Ivica Kostelic	+0.25	
4	Manfred Mölgg	+0.46	
5	Stefano Gross	+0.62	
6	Fritz Dopfer	+0.66	
7	Giuliano Razzoli	+1.04	
8	Mattias Hargin	+1.10	
9	Markus Larsson	+1.11	
9	Axel Baeck	+1.11	
11	Steve Missillier	+1.22	
12	Mario Matt	+1.26	
13	Ted Ligety	+1.33	
14	Jean-Baptiste Grange	+1.41	
15	Reinfried Herbst	+1.51	
16	Manfred Pranger	+1.65	
17	Patrick Thaler	+1.79	
18	Benjamin Raich	+1.88	
19	Cristian Deville	+1.97	
20	Manuel Feller	+2.02	
21	Ramon Zenhäusern	+2.13	
22	Wolfgang Hörl	+2.21	
23	Jens Byggmark	+2.24	
24	Akira Sasaki	+2.31	
25	Marc Gini	+2.33	
26	Michael Janyk	+2.81	
27	Markus Vogel	+3.09	
28	Thomas Mermillod Blondin	+3.34	
29	Alexander Khoroshilov	+3.79	
	Philipp Schmid		Fullførte ikke

zotero


Science is different!
Non-universal is uninteresting
Easier situation for V&V...

Science - social truths?

Expectations

Danger of deceiving ourselves

What is objective?

History of science & philosophy can give some guidance...

V&V can provide some insight too.

“How much effort is spent in replication?”

Takes time and resources.

Little direct benefit – a long-term value: underpins science.

Normal expectations: scientific production.

Interest: new discoveries.

Replication is 'a red listed specie' – undervalued.

Identical replication

Bit-by-bit identical (Linux 'diff' command).

Fixed number of states (digital information).

Constructed information/states.

Often the case

The model only captures the essential feature

Measurement errors

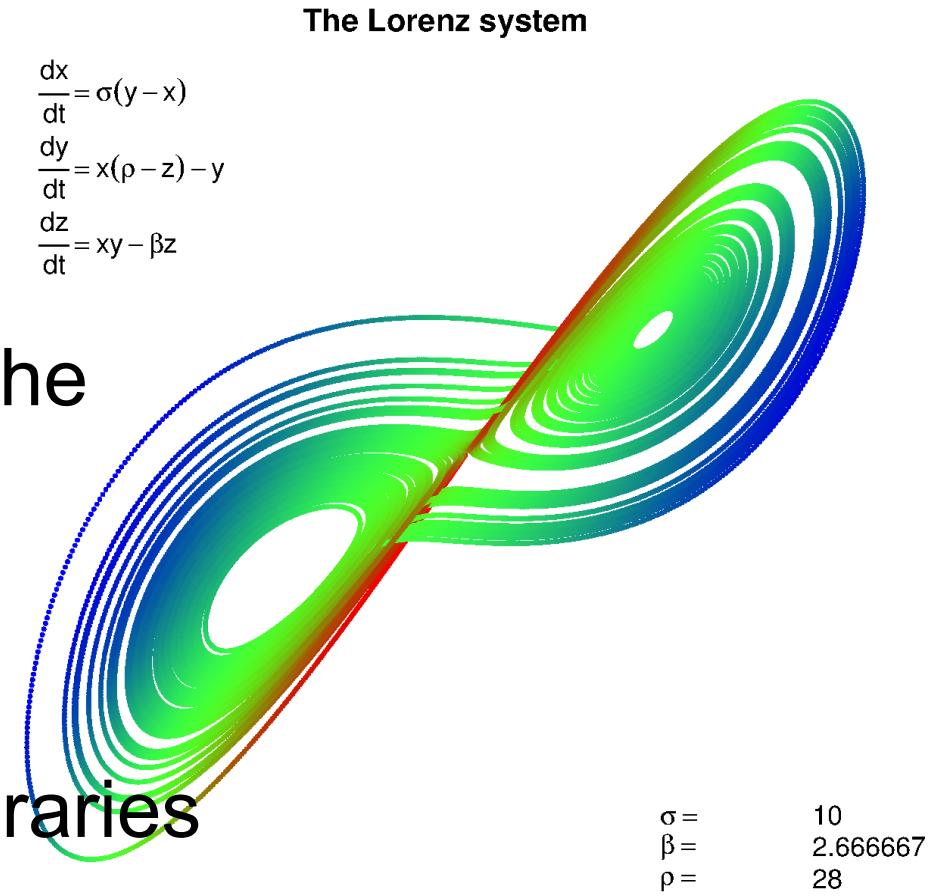
Round-off errors

Different compilers and libraries

Approximations

Non-deterministic processes – chaos.

Unmeasured aspects



What is the probabilities that...

… scientific replication

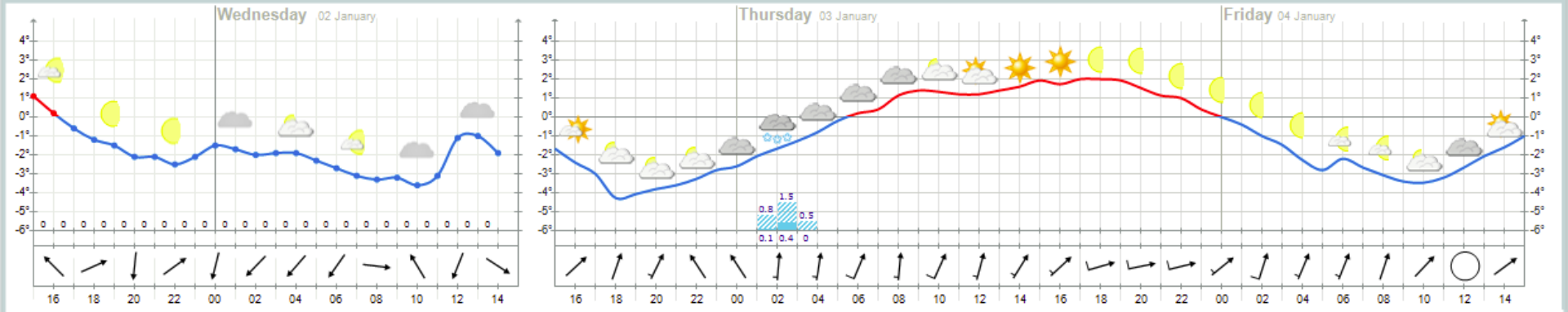
- Complex real world and probabilities are often the context for verification and validation of scientific work.

Examples from climate research

- Weather & climate forecasting
- Rich sample of different cases – varied
- Convenient for V&V – readily available data
- Direct experience – easy to conceptualise
- Some experience

Don't expect identical numbers...

Observations from Oslo (Blindern) observation site Last 24 h | **Meteogram for Oslo** Next 48 hours



The observations are from Oslo (Blindern) observation site, 3.6 km from Oslo. The observations should not be compared directly with the forecasts for Oslo, but give an indication on how accurate the forecasts are right now.

The forecast is issued for Oslo, not for Oslo (Blindern) observation site. The next forecast update is expected around 19:00. The blue bars shows expected max and min values for precipitation per hour.

Next lecture