ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

GAIA-CLIM

Task 6.1.1

User Requirements Survey

-Results-

Hanna Mäkelä
Gerrit de Leeuw
Otto Hyvärinen
Antti Mäkelä

Climate Service Centre
Finnish Meteorological Institute

GAIA-CLIM User Requirements Workshop
October 2015

Objective of Task 6.1.1

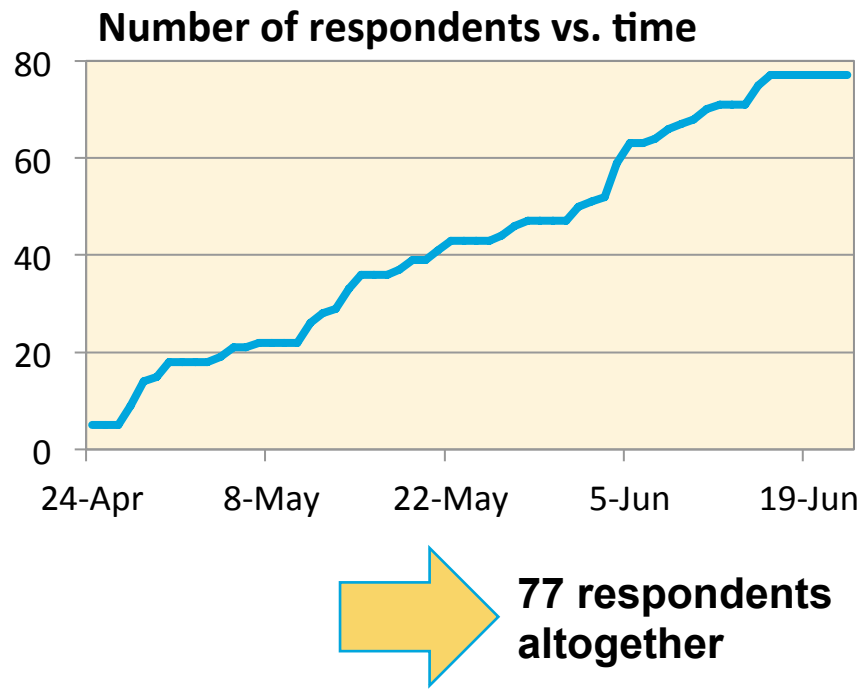
- This task **will collect feedback and information from the potential users of the Virtual Observatory** and other envisaged GAIA-CLIM outcomes and products

→ A user requirements survey was disseminated for different target groups and users; at least 5000 unique expert recipients

- The outcomes of the survey will provide **a baseline set of user expectations** and **help in identifying the group of core users** who may be invited to future workshops

Survey distribution

- WCRP, GCOS
- ECMWF: C3S & CAMS
- EUMETSAT: SAF leads
- ESA: CCI user community CMUG
- International Ozone Commission (IO3C)
- NDACC: SC members & UV/visible WG members
- QA4ECV
- FIDUCEO
- ConnectinGEO
- CLIMLIST
- Everyone in GAIA-CLIM: all participants & advisory panel



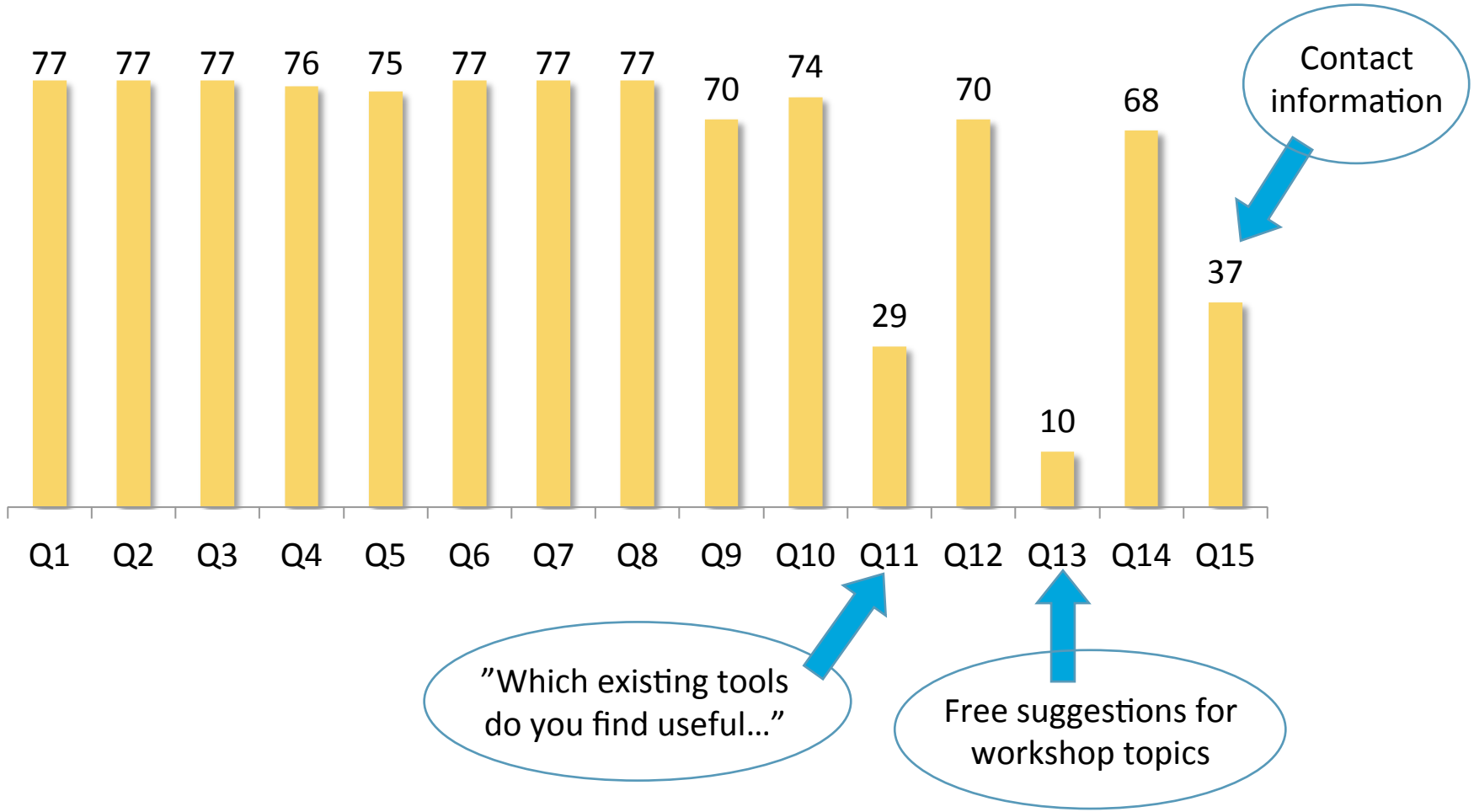
Survey design

- The survey was designed to help in prioritizing in particular how data and information are to be served through Virtual Observatory
- Five survey fields:
  1. Respondent's background (2 questions)
  2. Use of satellite, ground-based remote sensing, balloon-borne observations and aircraft observations (2 questions)
  3. Uncertainties (3 questions)
  4. Data availability, format and tools (4 questions)
  5. GAIA-CLIM Workshops (4 questions)

→ **15 questions altogether**

1. RESPONDENT'S BACKGROUND	2. USE OF SATELLITE, GROUND-BASED REMOTE SENSING, BALLOON-BORNE OBSERVATIONS AND AIRCRAFT OBSERVATIONS	3. UNCERTAINTIES	4. DATA AVAILABILITY, FORMAT AND TOOLS	5. GAIA-CLIM WORKSHOPS WITH EXTERNAL USERS
Q1. What is your sector of work?	Q3. Are you a user of satellite data? If so how do you use it, and if not why don't you?	Q5. What is you level of expertise in using information about observational uncertainties related to satellite, ground-based, balloon-borne or aircraft data?	Q8. What timeliness of the observational data do you need for your intended applications?	Q12. Would you be interested in taking part in GAIA-CLIM workshops?
Q2. What is your principal field of work?	Q4. Are you a user of data arising from ground-based remote sensing, balloon-borne observations or aircraft observations? If so how do you use it, and if not why don't you?	Q6. Do you use uncertainty estimates?	Q9. GAIA-CLIM is planning to distribute all its data in NetCDF4 format containing Climate and Forecast compliant metadata. What is your preference for NetCDF4?	Q13. Do you have any suggestions for topics to be addressed in more detail during the workshops within the GAIA-CLIM project?
		Q7. Which guidance on how to utilize observational uncertainty information related to satellite and sub-orbital data would be valuable to you?	Q10. What kind of tools do you expect to be available on the Virtual Observatory?	Q14. Would you be interested in testing the tools developed in GAIA-CLIM?
			Q11. Which existing tools do you find useful to compare data sets?	Q15. If you wish you can leave your contact info here.

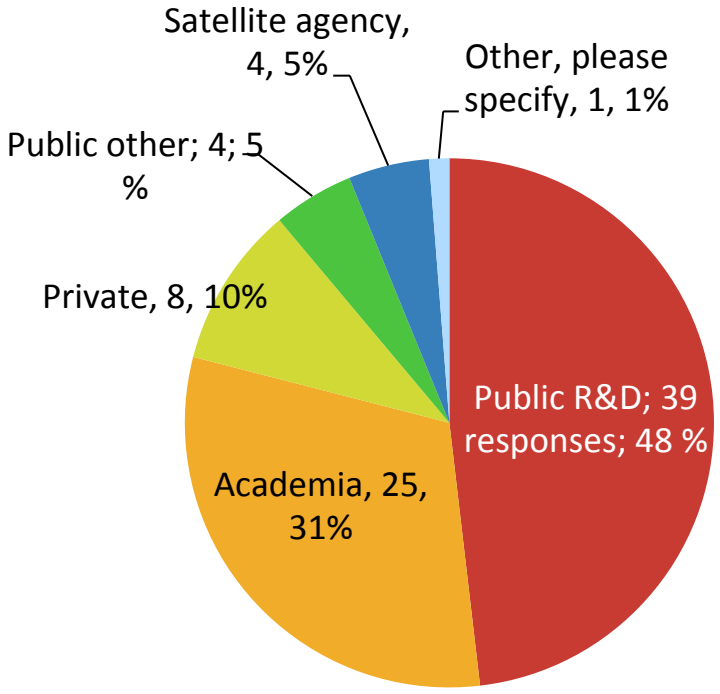
Number of respondents per question



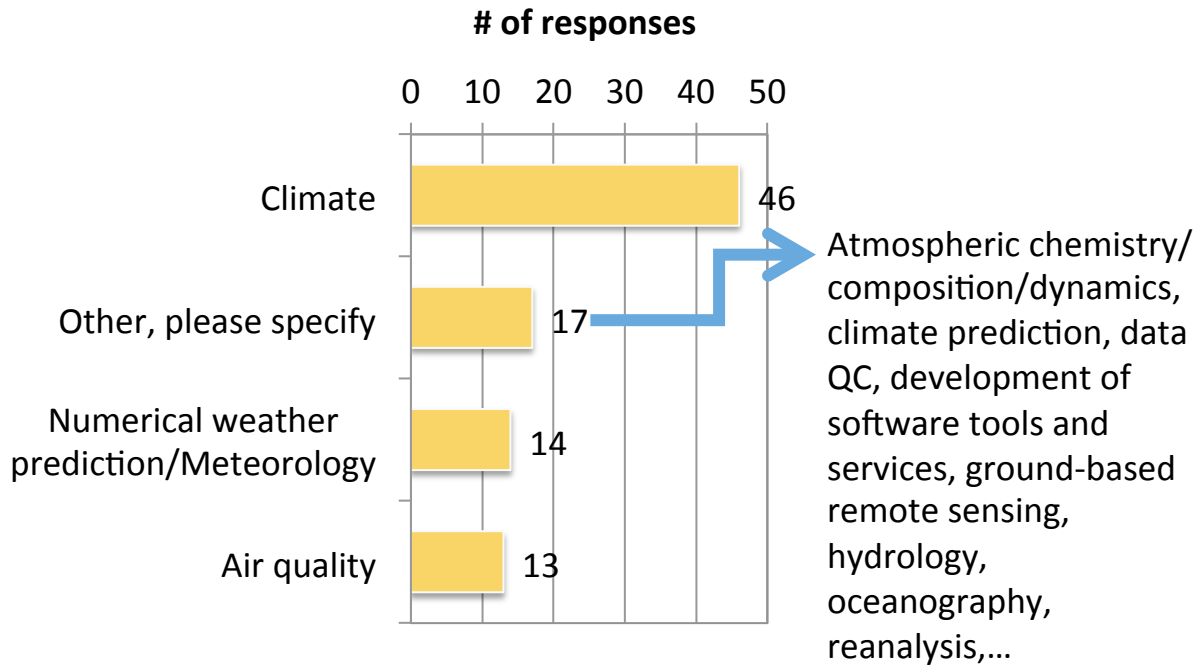
Q1 & Q2: Respondent's background

- 77 respondents
- Over 75% of the respondents are from public sector's R&D or from academia
- Respondents work mostly with "Climate" (60% of the respondents)
- Respondents picked, on average, 1.2 different fields of work

What is your sector of work?



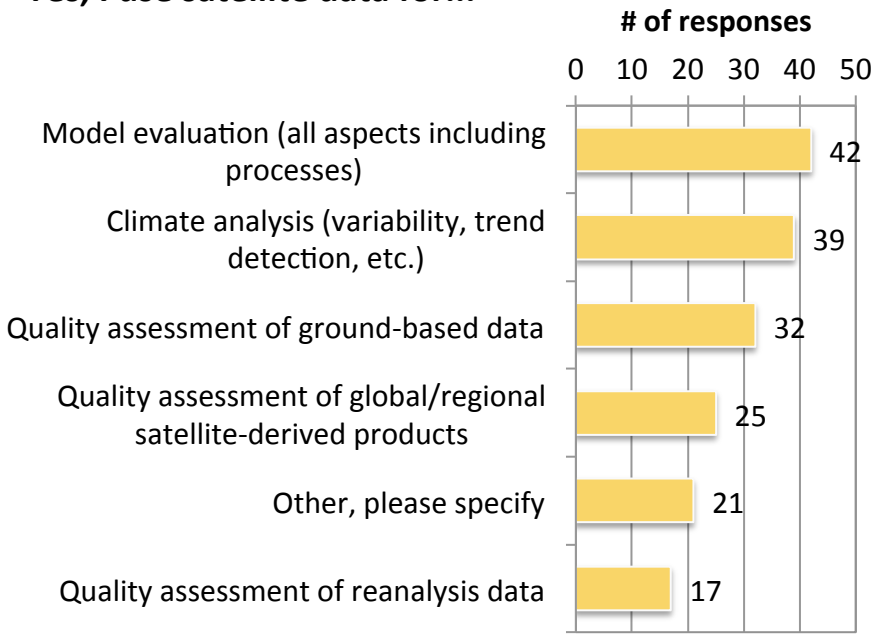
What is your principal field of work?



Q3: Are you a user of satellite data? If so how do you use it, and if not why don't you?

- 77 respondents
- On average, 2.4 choices per respondent

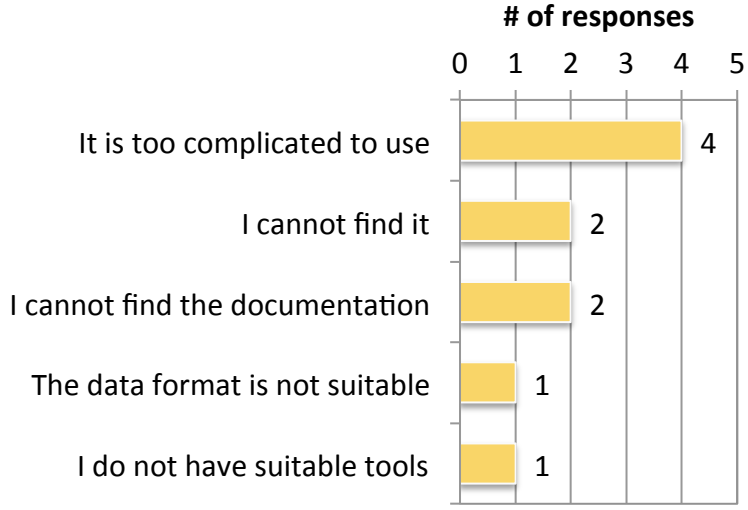
Yes, I use satellite data for...



"Other, please specify":

- Data assimilation, generation of meteorological products, as model input, case studies, educating forecasters,…

No, I do not use satellite data because...



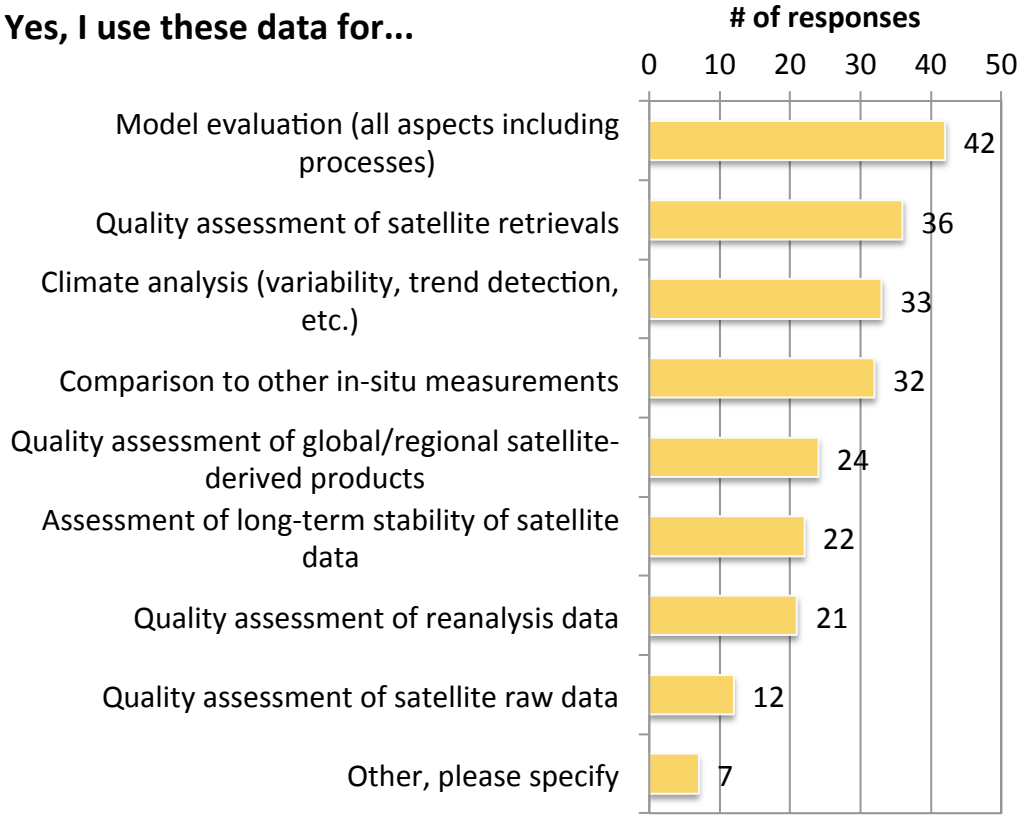
None for:

- I do not need it
- I do not have access to it
- There are no suitable datasets for my applications
- I do not trust it

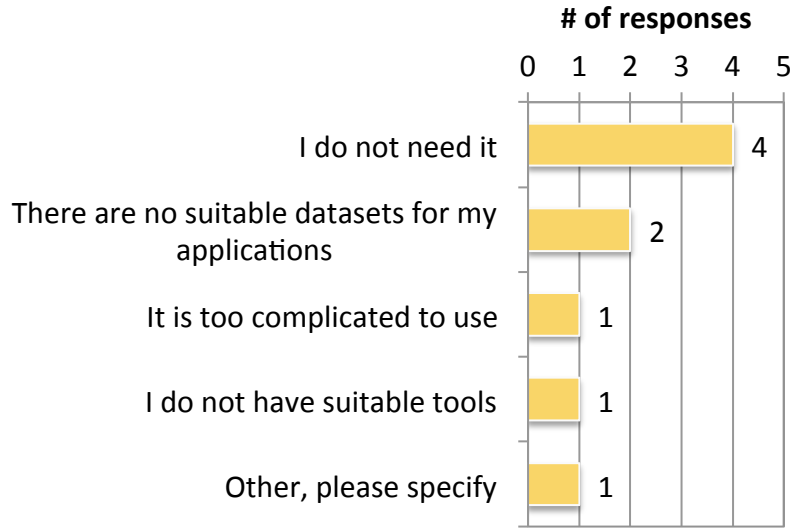
Q4: Are you a user of data arising from ground-based remote sensing, balloon-borne observations or aircraft observations? If so how do you use it, and if not why don't you?

- 76 respondents
- On average,3.1 choices per respondent

Yes, I use these data for...



No, I do not use these data because...

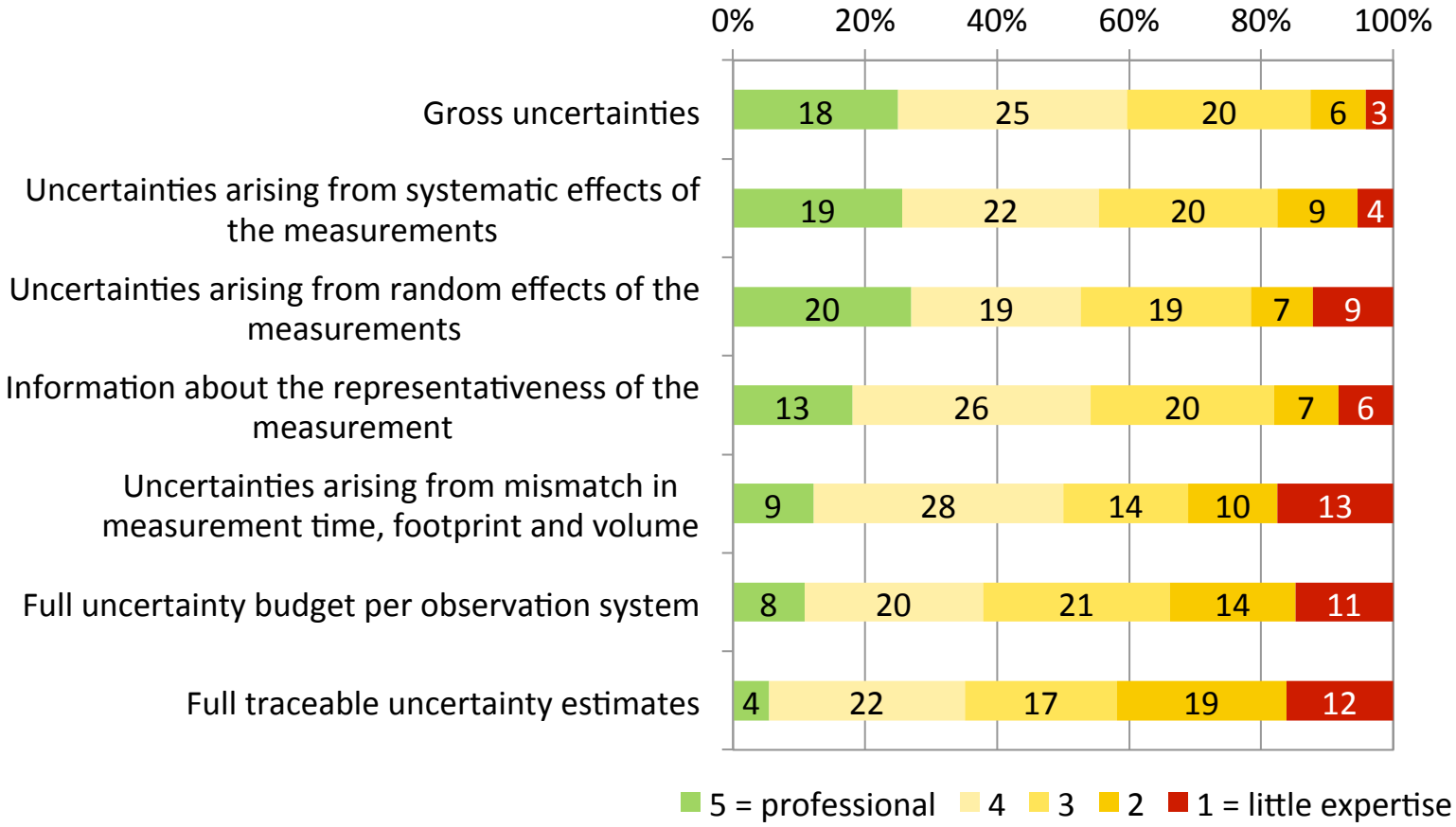


None for:

- I do not have access to it
- I cannot find it
- I cannot find the documentation
- The data format is not suitable
- I do not trust it

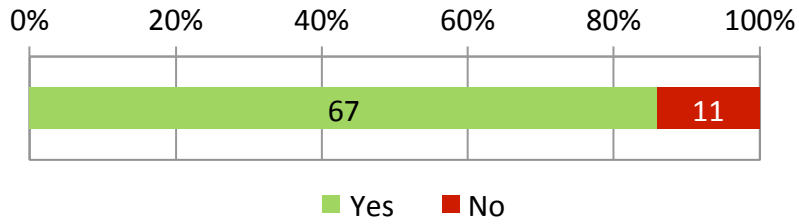
Q5: What is your level of expertise using information about observational uncertainties related to satellite, ground-based, balloon-borne or aircraft data?

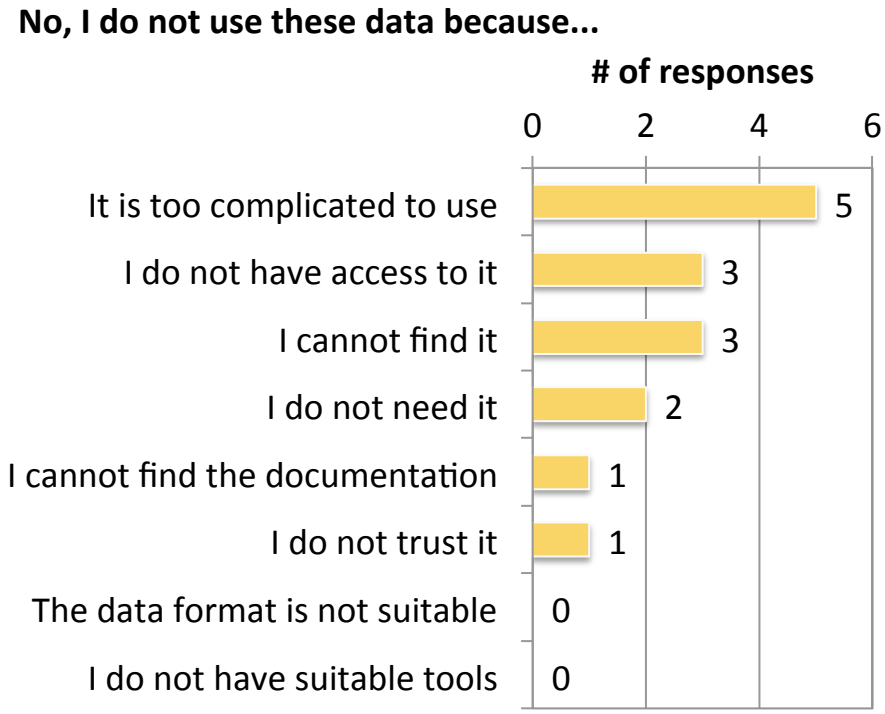
- 75 respondents



Q6: Do you use uncertainty estimates?

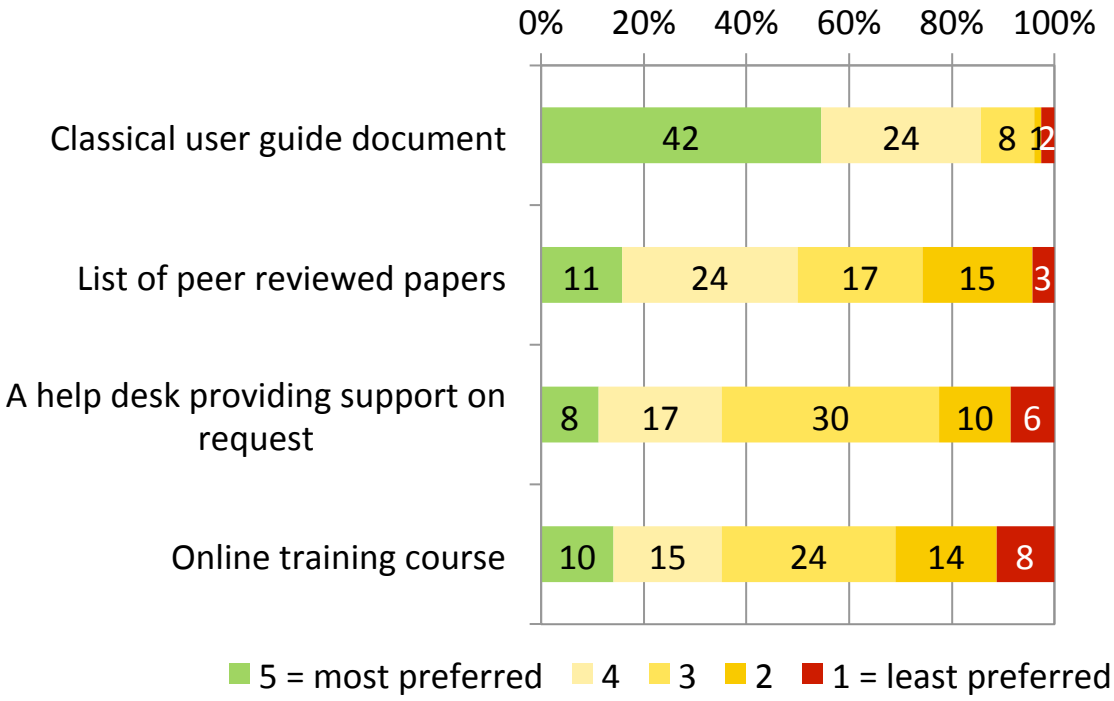
- 77 respondents
- 86% of the respondents uses uncertainty estimates
- Those who don't use, find it too complicated to use or do not have access to it or cannot find it





Q7: Which guidance on how to utilize observational uncertainty information related to satellite and sub-orbital data would be valuable to you?

- 77 respondents

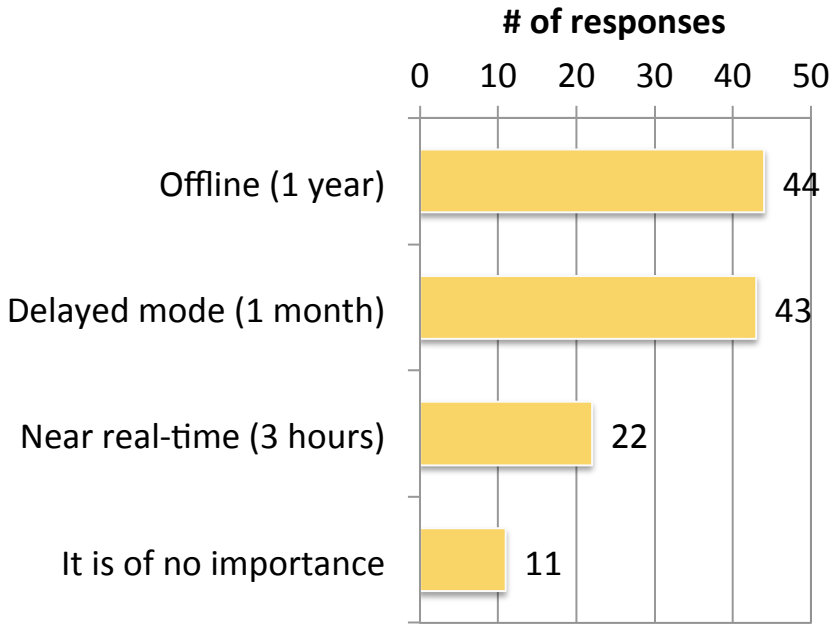


”Other, please specify”:

- A workshop run by instrument experts
- A web page showing examples
- A concise traceable document providing key information
- Ensemble of observations

Q8: What timeliness of the observational data do you need for your intented applications?

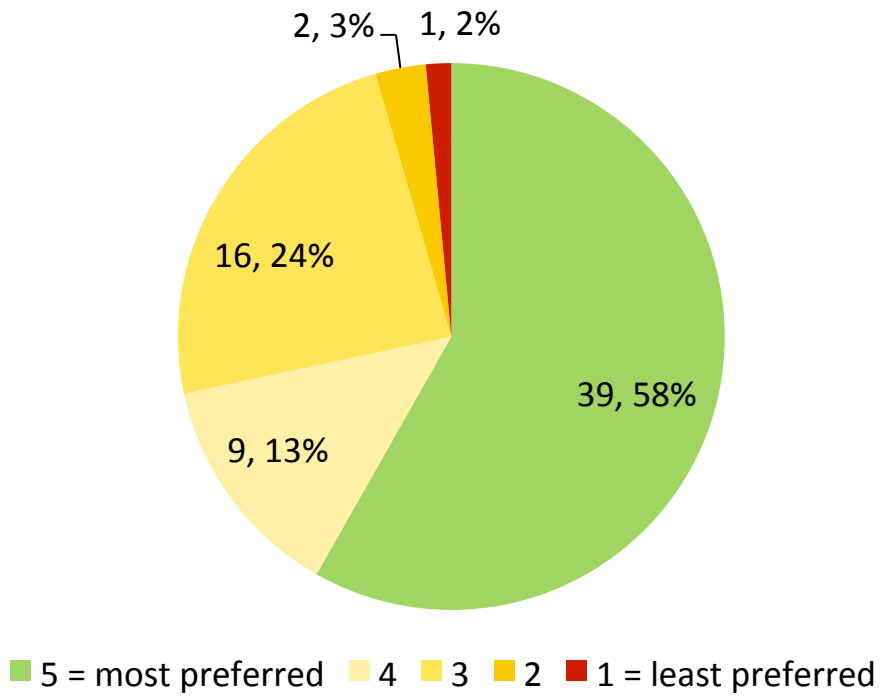
- 77 respondents
- On average, 1.6 choices per respondent
- Over 70% of the respondents favour timeliness either of ”Offline (1 year)” or ”Delayed mode (1 month)” or don’t consider it to be important



Q9: GAIA-CLIM is planning to distribute all its data in NetCDF4 format containing Climate and Forecast compliant metadata.

- 70 respondents
- Only three respondents picked ”1” or ”2” for NetCDF4 preference → their suggestions are HDF5 and ASCII

What is your preference for NetCDF4?



”Other format preferences”:

	# of responses	preference
HDF5	9	4.00
ASCII	4	4.25
HDF	4	4.25
ODB	2	4.00
BUFR	1	5.00
CSV	1	5.00
FITS	1	5.00
GRIB	1	5.00
NetCDF3	1	5.00
HDF4	1	4.00
NetCDF	1	4.00
v4	1	4.00
ARTS-XML	1	2.00

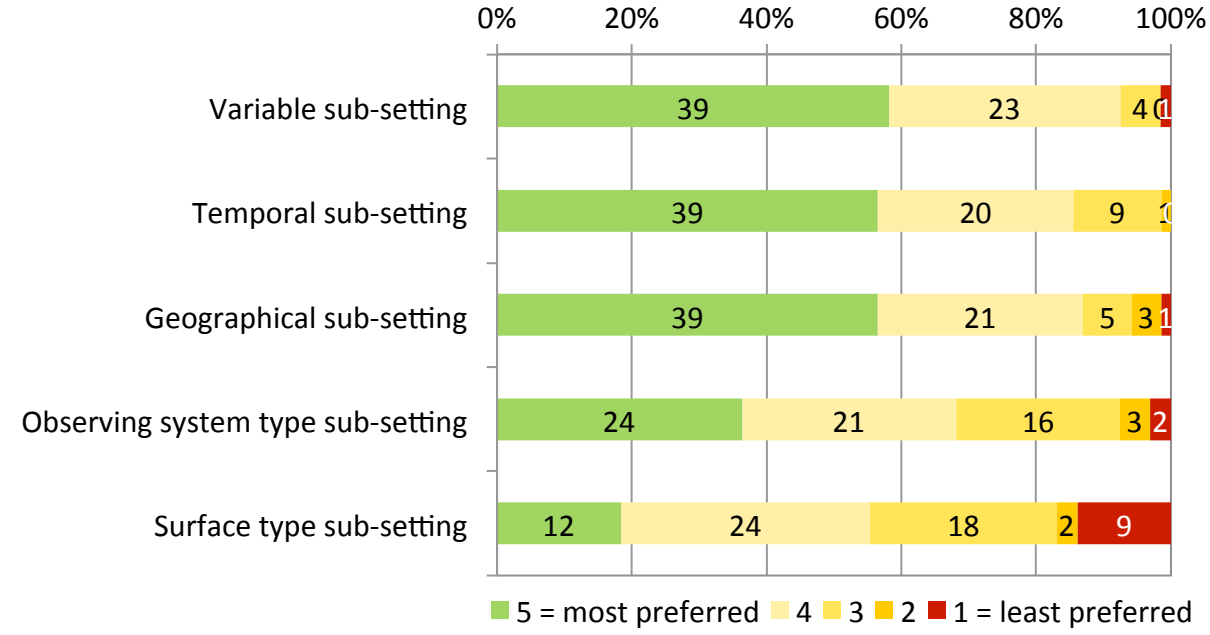
Q10: What kind of tools do you expect to be available on the Virtual Observatory? 1/3

- 74 respondents

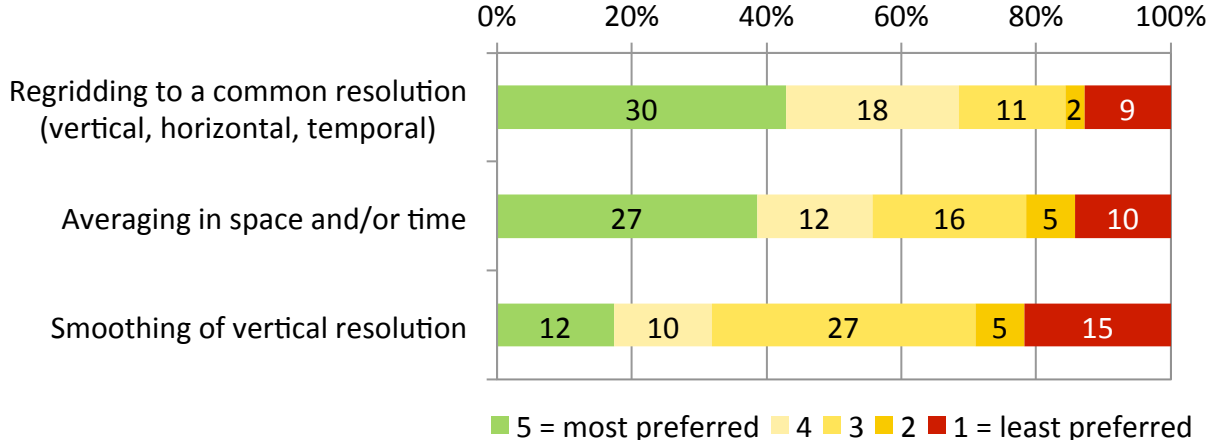
”Other, please specify”:

- A fast SQL interface
- An interface to get metadata
- Possibility for FTP downloads
- Ability to download data directly (i.e., direct download URLs)

Data search and discovery



Online data postprocessing

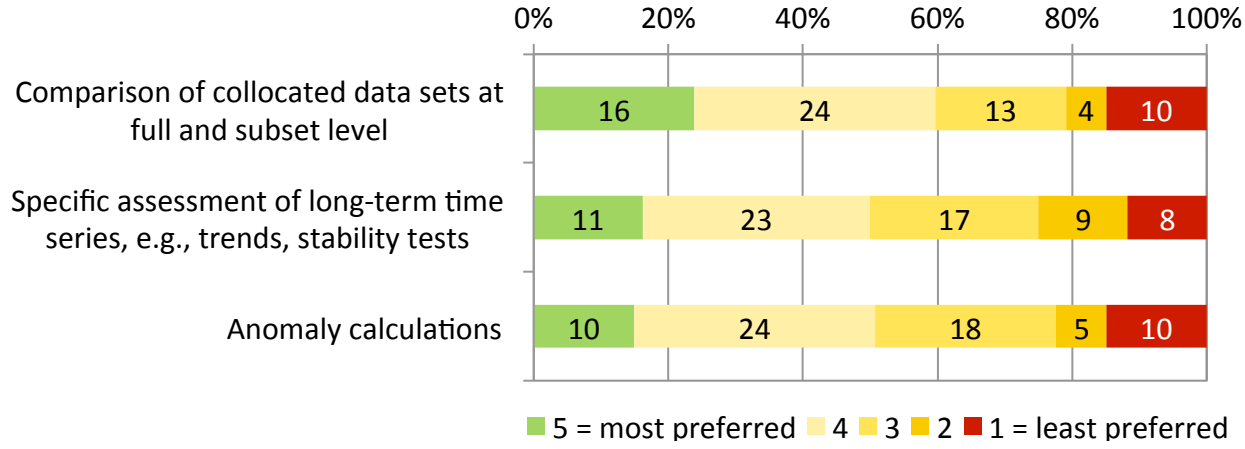


Q10: What kind of tools do you expect to be available on the Virtual Observatory? 2/3

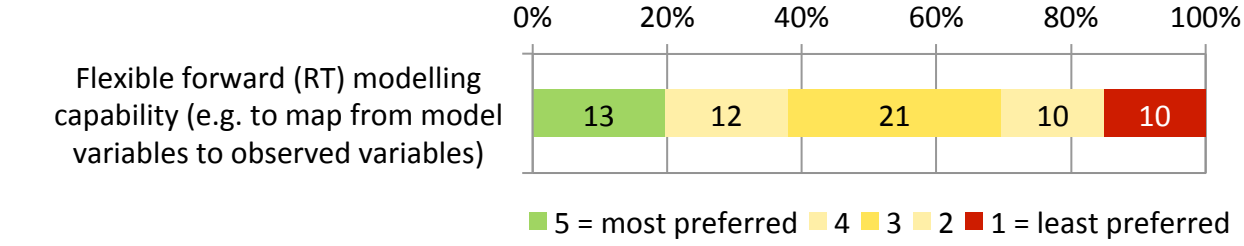
”Other, please specify”:

- “All the above, but not performed by button pressing, but perhaps given a library, performed by a script that I can edit and test locally and/or submit online“

Online data analysis (statistical analysis etc.)



Flexible forward (RT) modelling capability

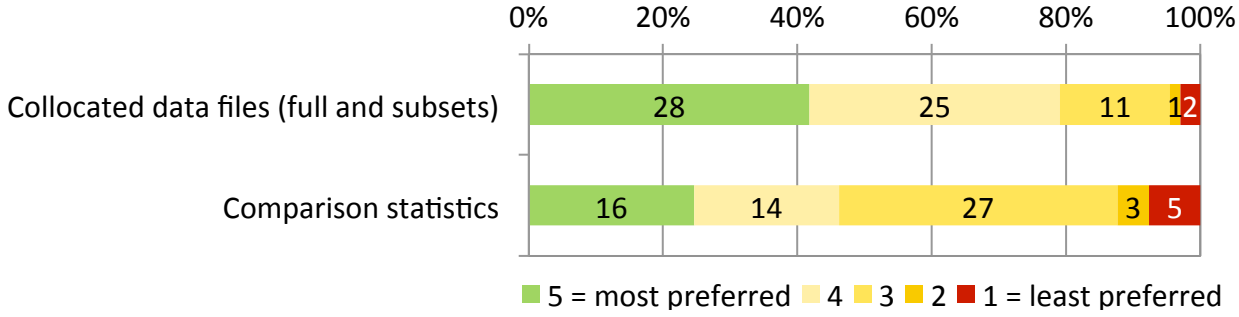


Q10: What kind of tools do you expect to be available on the Virtual Observatory? 3/3

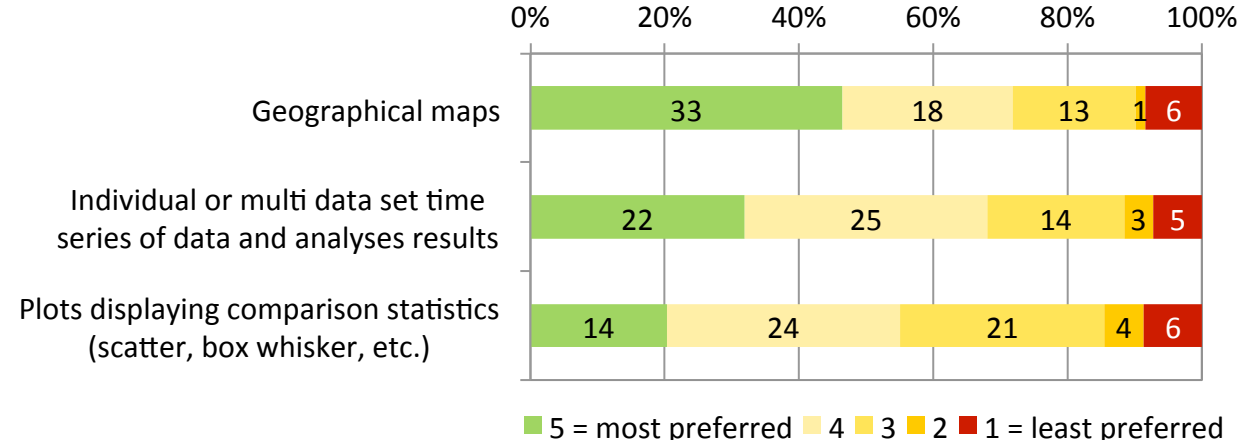
”Other, please specify”:

- Open formats and metadata standards to allow tailoring tools for one’s own applications and studies
- Toolbox of scripts for data handling that can be downloaded and modified for personal use

Data extraction

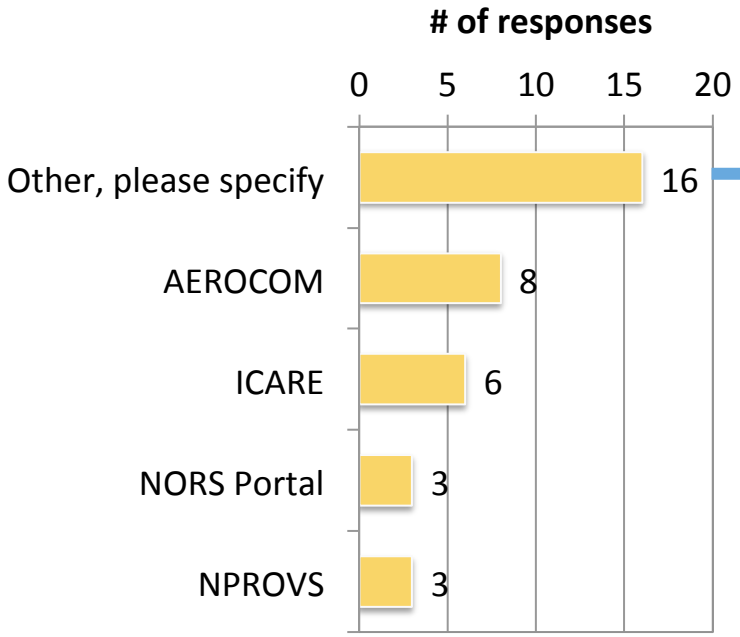


Data visualization



Q11: Which existing tools do you find useful to compare data sets?

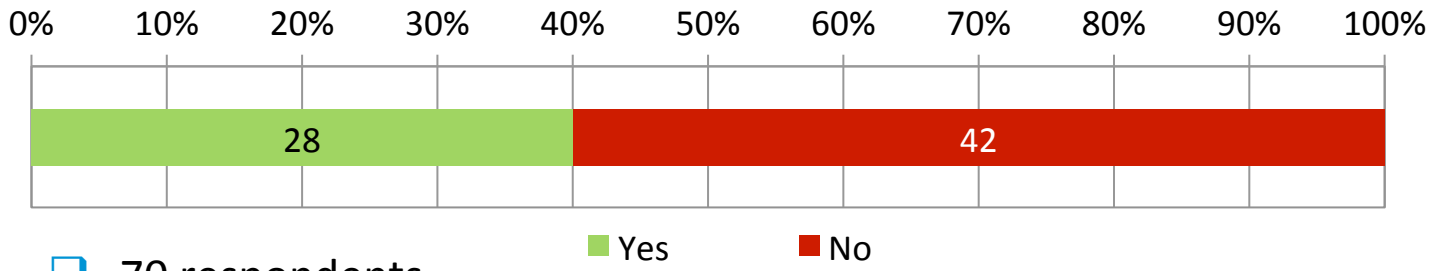
- 29 respondents
- On average, 1.2 choices per respondent
- Most of the respondents picked something else than the predefined options



”Other, please specify”:	# of responses
My/Institute's own	4
GIOVANNI	2
IDL	2
NDACC	2
AERONET, African Climate Atlas, CIS, Igor Pro, programming tools, satellite data websites, TCCON, TEMIS, NASA AVDC	1

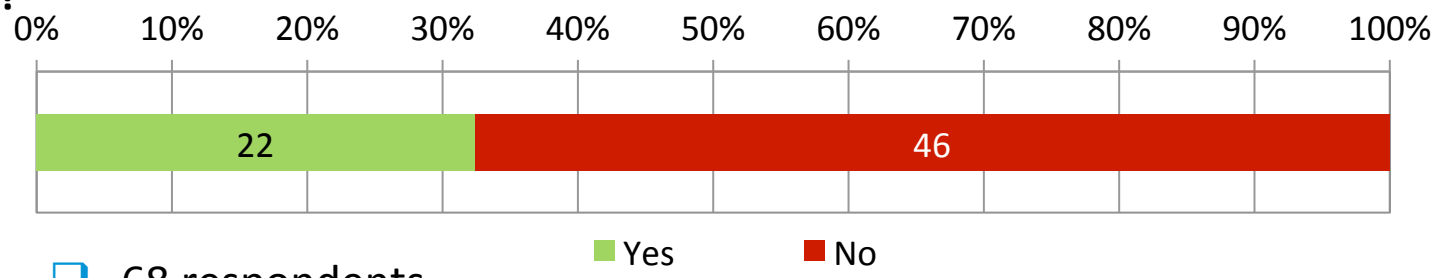
Q12 & Q14

Would you be interested in taking part in GAIA-CLIM workshops?



- 70 respondents
- 21 of the ”Yes” respondents left their contact info

Would you be interested in testing the tools developed in GAIA-CLIM?



- 68 respondents
- 20 of the ”Yes” respondents left their contact info

Q13: Do you have any suggestions for topics to be addressed in more detail during the workshops within the GAIA-CLIM project?

- 10 respondents

- Topical campaigns for specific research themes
- Potential inter-project synergies between FIDUCEO and GAIA-CLIM
- Statistical methods to incorporate measurement uncertainties in the calculation of spatial and temporal means
- Address the issue of disseminating level 1 products and its uncertainty instead of satellite retrievals
- Methods for relating global surface measurement networks to satellite observations
- Forward modelling - and how to do it properly
- Long-term trends; data set stability; averages, e.g., monthly zonal means
- The most relevant model of propagation of uncertainties/errors when performing comparisons between two measurements of the same object with two different methods
- Ways to handle space/time collocation mismatches
- Be clear about definitions of uncertainty (bias, accuracy, stability ...)

Conclusions 1/2

- Based on the survey the priorities of VO should be
  - to present long time series (because majority of the responders were interested in climate)
  - to serving the application areas of
    - model evaluation,
    - climate analysis,
    - quality assessment of ground-based data/satellite retrievals.
- Easily understandable explanations are needed (e.g., uncertainty measures) and desired in the form of a classical user guide document
- No strong timeliness requirement → VO can be successfully working offline in its initial implementation (but we should not forget NWP requirements for NRT)

Conclusions 2/2

- The central gaps revealed by the survey:
  - the lack of response rate → a possible gap in education and knowledge about how sub-orbital and/or satellite data can be used for scientific and practical applications
  - the lack of broad-scale use of existing match-up facilities → a clear need for such a facility, but a difficulty to define a functional one, and a potential reluctance of users to take up new tools
  - the broad range of user familiarity with and use of uncertainties → a need for education and capacity building in this area