



Climate Change

Coordination across C3S EQC contracts QA4Seas review

DECM project meeting, Oslo, 2017-10-12
nicolau.manubens@bsc.es





Climate
Change

OUTLINE

- QA4Seas overview
- Surveys and interviews
- EQC framework
 - Data standards and Provenance
 - KPIs for CDS, EQC and products
 - ...
- EQC prototype
 - Design and status
 - Interaction with the CDS
- Conclusions



Climate
Change

QA4SEAS OVERVIEW

- QA4Seas aims at developing a strategy for the evaluation and quality control (EQC) of the multi-model seasonal forecasts provided by the C3S to respond to the needs identified among a wide range of stakeholders.
- To achieve the objective, the consortium will:
 - Be user-driven with a **two stage consultation** process (coordinated with other lots). **ADAPTATION**
 - Perform a **gap analysis** of the current information available to the users + elaborate a **verification framework** considering recent methods developed in research and the evaluation of multi-faceted quality aspects. **INNOVATION**



Climate
Change

Q A 4 S E A S O V E R V I E W

- Elaborate **data storage recommendations** for the CDS and forecast providers and develop a **provenance framework**. This will ensure a high degree of coherence and make products open to scrutiny and comparison, thus building trust on the service.
- Develop a **computing performance framework** with an analysis on existing verification software, recommendations for its efficient use and recommendations for the CDS infrastructure.
- Develop a **proof-of-concept prototype**, making the most of existing open-source tools.

CONSISTENCY

EFFICIENCY



- First consultation stage – carried out
 - Survey
 - A survey has been conducted, on the following themes:
 - Organization and role of the participant
 - Use of ECVs and CIIIs
 - Using and accessing climate forecasts
 - Post-processing
 - Uncertainty
 - Metadata and software tools
 - The survey focuses on what the respondents currently do.
 - Interviews
 - An interview protocol has been established. Same themes as the survey.
 - Feedback requested on a drafted list of functional requirements for the EQC prototype.



Climate
Change

SURVEYS AND INTERVIEWS - ADAPTATION

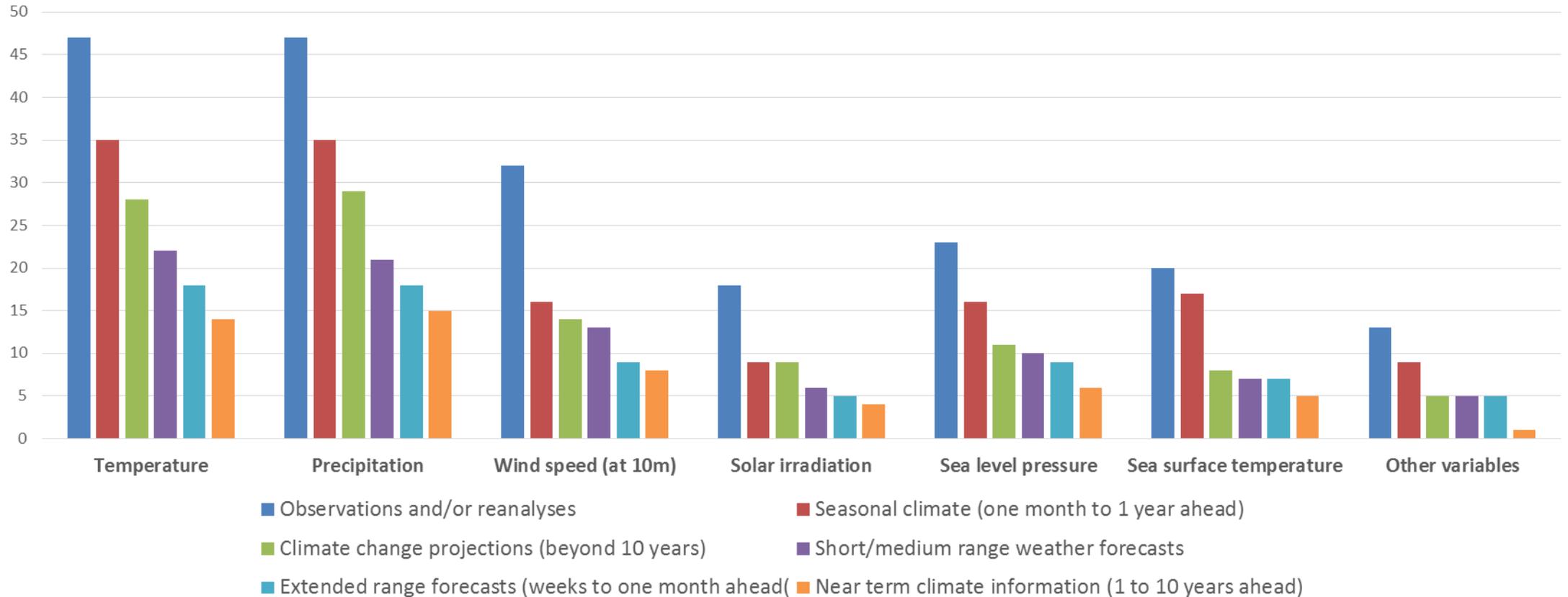
- First consultation stage – carried out
 - Respondent summary

	Survey	Interviews
NMHS	35	0
Research Institutes	8	4 (JRC, ETH, CNR, IRI)
Private companies	6	1 (Swiss Re)
International organisations	1	2 (WMO, WFP)
Gov. agency	2	1 (Alvac)
Other	6	0
	58	8



SURVEYS AND INTERVIEWS - ADAPTATION

Use of climate variables

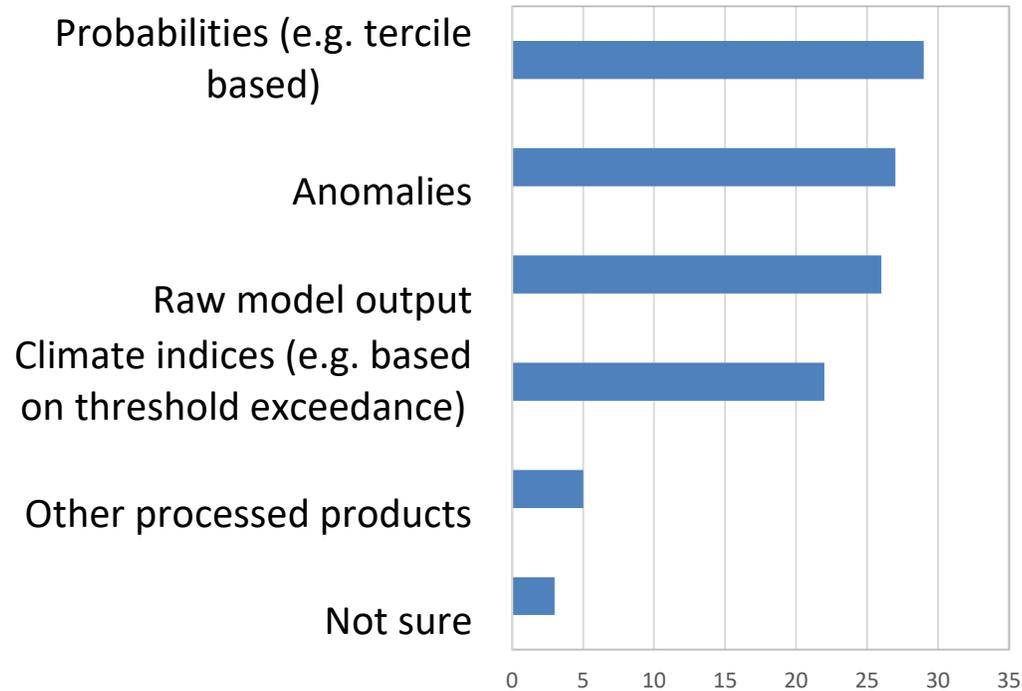


Users tend not to employ derived quantities or indices (n=27); those who do tend to use HDD and heavy precipitation/rainfall index (n=17 each), consecutive dry days (n=14) and cooling degree days (n=13).

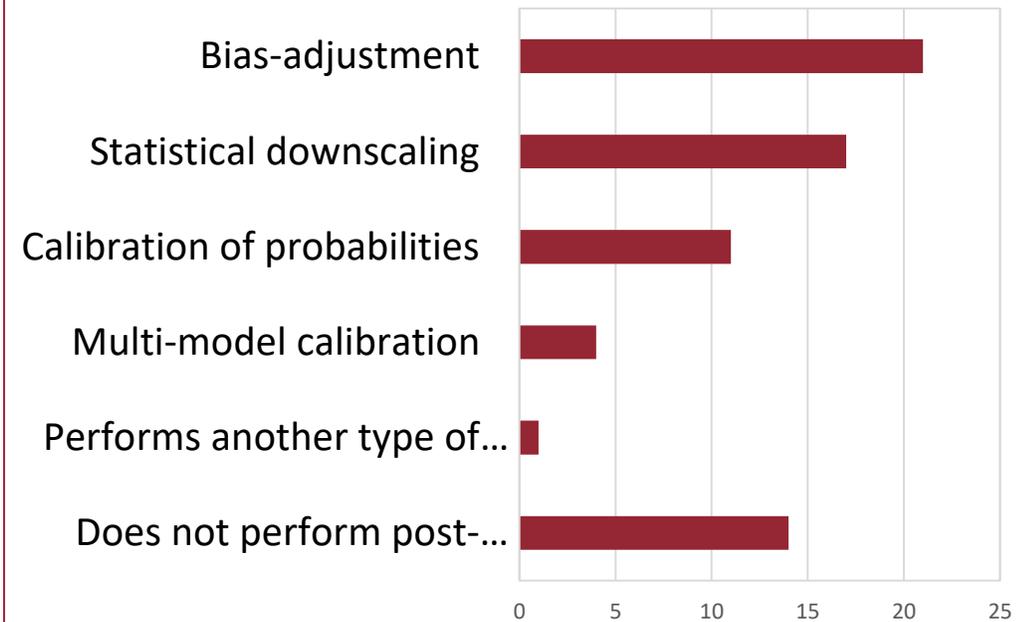


- First consultation stage – carried out

"What kind of data from global SF do you use?"



"What type of adjustment post-processing do you perform on the SF data before using it?"

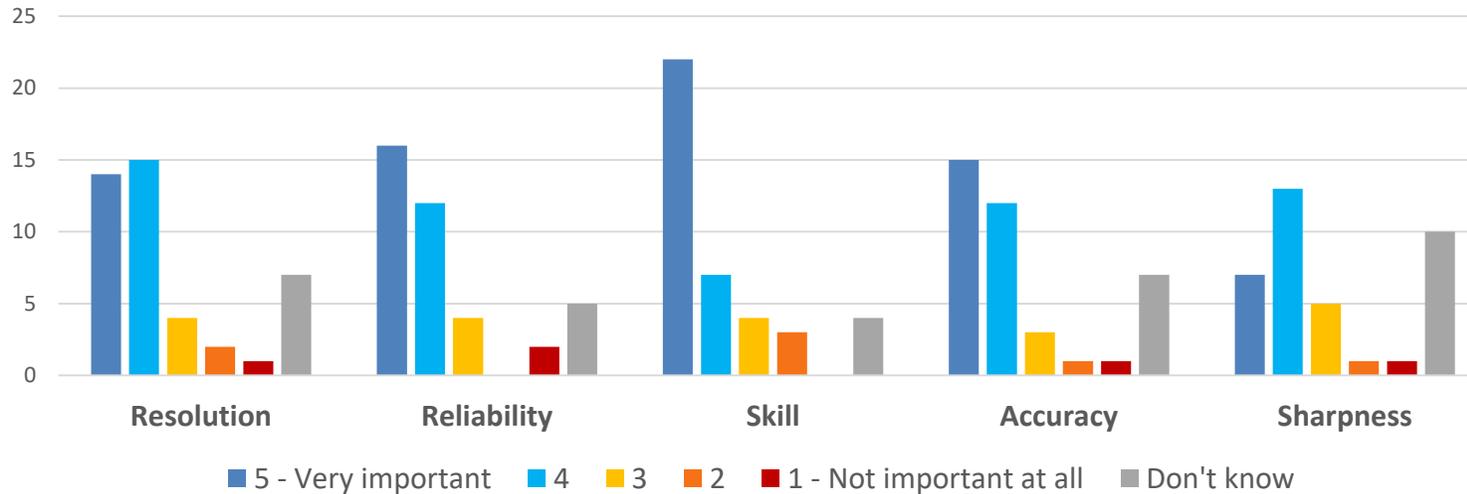




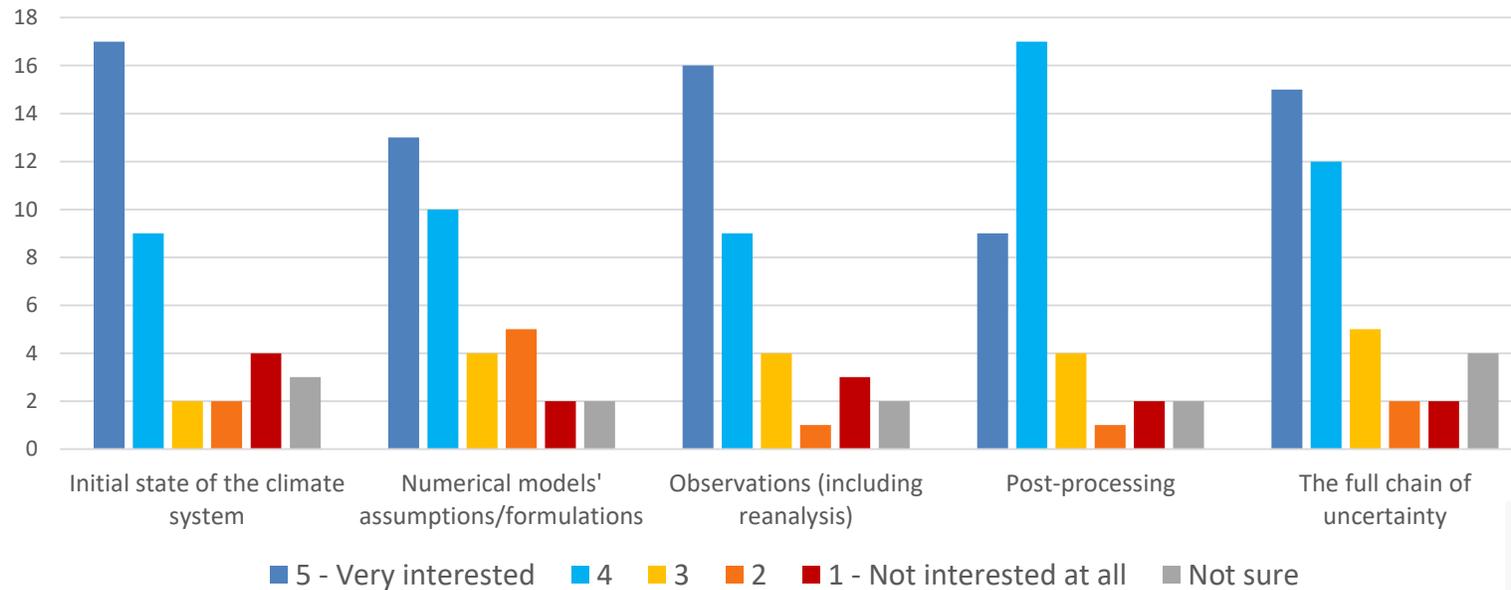
Climate
Change

SURVEYS AND INTERVIEWS - ADAPTATION

Relative importance of forecast quality attributes



Interest in sources of uncertainty in seasonal forecasts



M. Soares, A. Taylor (Univ. Leeds)



- First consultation stage – carried out
 - Survey result highlights
 - Majority of survey respondents were from NMHS (n = 30).
 - Many of the survey respondents use seasonal climate forecasts (n = 35).
 - The climate variables most used were temperature and precipitation (n = 35 each) followed by sea surface temperature (n = 17), wind speed (n = 16) and sea level pressure (n = 16).
 - Indices were less used; some used heating degree days and heavy precipitation/rainfall index (both n = 17) and consecutive dry days and cooling degree days were also fairly used (n = 14 and n = 13, respectively).



- First consultation stage – carried out
 - Interview facts/results (under analysis)
 - Interviewees were mixed from research institutes, international organisations and private companies.
 - Only half of the interviewees (n = 5) use seasonal forecasts – some are forecast providers or intermediary organisations and not users of forecasts per se.
 - Different uses e.g. for research, to inform/support action, or for business purposes as well as by using the information quantitatively and/or qualitatively.
 - User requirements are a moving target that depends on the amount of information they receive.



Climate
Change

SURVEYS AND INTERVIEWS - ADAPTATION

- Second consultation stage
 - Pending, due in February 2018.
 - It will be a QA4Seas-specific survey.
- User requirement collaboration framework
 - Key identified areas for collaboration:
 - Agreeing on methods for assessment of user requirements.
 - Sharing information between EQC contracts via URDB.
 - Summarizing collated information in a common URAD.
 - A presentation is available with details.



Climate
Change

EQC FRAMEWORK - INNOVATION

- Verification in a nutshell
 - A thorough assessment of the existing C3S seasonal forecasts from C3S_433 has been done:
 - More than 30,000 plots made available via an ad-hoc Shiny app.
 - An assessment of the results is being carried out and will be made available as a paper.
 - A verification framework has been drafted with:
 - Suggestions on which forecast quality measures are needed according to C3S and users needs.
 - Suggestions of other quality measures that are strongly recommended.
 - Rhetorical questions for the users to better understand the meaning of the quality measures.



Climate
Change

EQC FRAMEWORK - INNOVATION

- Verification in a nutshell
 - Recommendations on EQC products and visual representation of uncertainty are being drafted.
 - EQC information is not neutral. Precise definitions are necessary, and documentation (context, provenance) is key.



Climate
Change

EQC FRAMEWORK - CONSISTENCY

- Data standards in a nutshell
 - Identified and homogeneously catalogued ECVs relevant to seasonal forecast verification.
 - Defined conventions for the forecast providers and CDS to store climate data and metadata.
 - Developed a collaboration framework to define standards on observational products.

The data standards outcomes and collaborations will be discussed in the afternoon



Climate
Change

EQC FRAMEWORK - CONSISTENCY

- Provenance
 - **Reproducibility** (being able to duplicate a process) and **traceability** (being able to verify the history, location and application of an item by means of a documented record) are key factors for trust building.
 - Sprov-Flow provides a solution emphasizing the technical traceability
 - QA4Seas has developed METACLIP, a solution that makes emphasis on the high-level (scientific) aspect of the traceability. This aspect has been deemed more relevant in QA4Seas for the users to understand and properly select the products.
 - Both solutions are not mutually exclusive. ECMWF and CDS toolbox team pending to analyse them and give indications for SC2.



Climate
Change

EQC FRAMEWORK - CONSISTENCY

- Provenance
 - METACLIP relies on RDF. RDF is an W3C standard that allows to model relationships between concepts.
 - A set of modelled relationships between concepts in a particular field or theme conform a vocabulary. Plenty of RDF vocabularies have contributed and evolved by the community.
 - METACLIP uses some of the existing vocabularies (e.g. PROV for Climate Data provenance), and defines some others from scratch to represent typical concepts in climate forecasting: data source, transformation, (no) adjustment, seasonal forecast verification, climate validation (w.i.p., COST Action Value), and product.



Climate
Change

EQC FRAMEWORK - CONSISTENCY

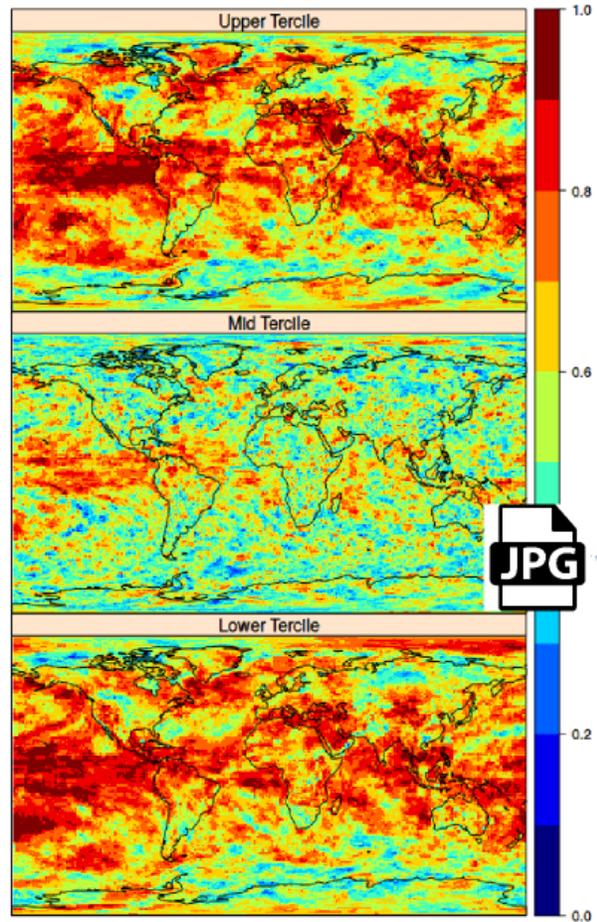
- Provenance
 - METACLIP generates and populates RDF graphs with the following information:
 - A high-level chain of labelled steps that have been followed to produce a product.
 - A record of software versions and calls issued throughout the validation proces.
 - Metadata, which is propagated from its source to the final product.
 - METACLIP is a framework that can be implemented in any programming language.
 - RDF graphs can easily be serialized and attached to figures or other products.



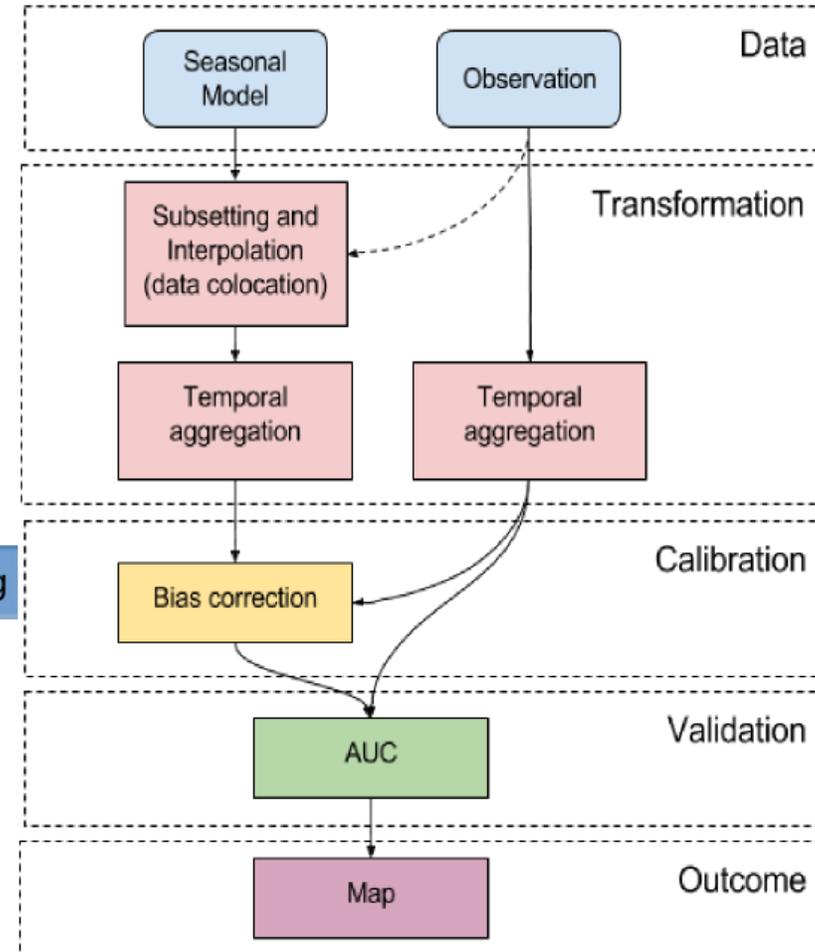
Climate
Change

EQC FRAMEWORK - CONSISTENCY

- Provenance



Metadata encoding



<http://demo.predictia.es/qa4seas/metadata/>



Climate
Change

EQC FRAMEWORK - EFFICIENCY

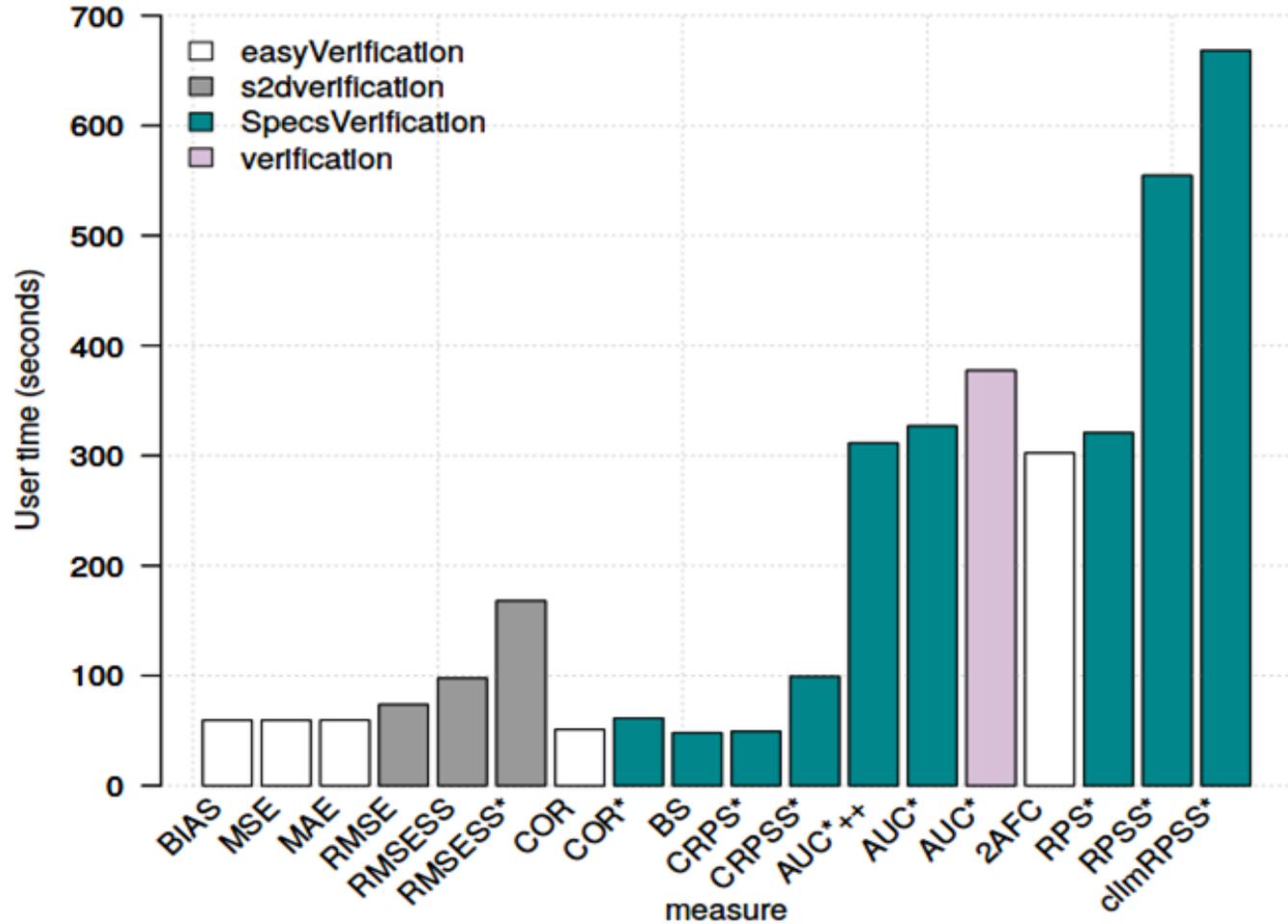
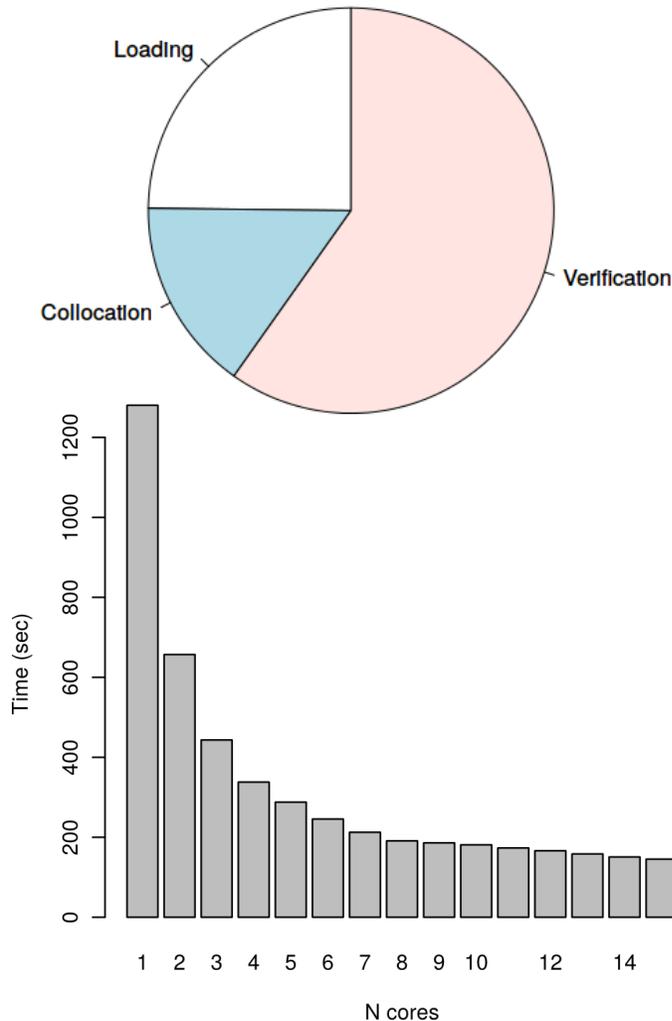
- Performance in a nutshell
 - A performance analysis of existing validation tools has been conducted, evaluating the performance and efficiency KPIs (read further for info).
 - Parallelization over clusters of nodes is being implemented. A workflow manager (ecFlow) is transparently used underneath.
 - The performance analysis and developments strongly depend on the CDS infrastructure. We will tailor recommendations as we receive information. Regardless, we will provide general recommendations on NetCDF chunking, bringing compute to the data, caché-ing products, pre-calculation, CPU usage, ...



Climate Change

EQC FRAMEWORK - EFFICIENCY

- Performance in a nutshell





Climate
Change

EQC FRAMEWORK – KPIS

- Elaborated 3 groups of KPIs:
 - To evaluate the CDS service
 - Availability of ECVs
 - Performance of requests
 - Service availability
 - To evaluate CDS and EQC products
 - Fulfillment of standards
 - Reproducibility
 - Plot dissemination quality
 - To evaluate the EQC service
 - Completeness of EQC functionality provided
 - Performance
 - Wallclock time
 - Parallelization ability
 - Memory footprint
- The KPIs will be evaluated at the end of the contract. However they have been designed to be used as a monitoring tool.



Climate
Change

E Q C P R O T O T Y P E

- Design and status

- The goal is to deliver a set of command line functions. No graphical user interface planned.

- Development plan based on existing open-source R packages, using bridging functions to communicate data from one to the other.

SpecsVerification

- Probabilistic and deterministic scores
- Works on [time x members] arrays

easyVerification

- Applies SpecsVerification scores to arrays of any dimensions, multi-core
- Probabilistic and deterministic scores

downscaleR + loader

- Data retrieval and homogenization
- Bias adjustment, modes, downscaling
- Probabilistic and deterministic scores
- Visualisation of data and results

s2dverification

- Data retrieval and homogenization
- Bias adjustment, filtering, modes
- Probabilistic and deterministic scores
- Visualisation of data and results

S
C
O
R
E
S

UNIVERSITY OF
EXETER

MeteoSwiss

F
R
A
M
E
W
O
R
K
S

UC
UNIVERSIDAD
DE CANTABRIA

CSIC





- Design and status

- Score and processing functions in all involved packages are being adapted to operate arrays with only the essential dimensions. Then, they are applied automatically in parallel (multicore, multinode) over arrays with additional dimensions.
- s2dverification and downscaleR are two coexisting frameworks with similar data structures.

\$Dataset
\$Variable
\$Data
\$Dates
 \$start
 \$end
\$xyCoords

Main components of an R s2dverification / downscaleR object. In green, the components that fully match. In red, the components for which there are differences between the two frameworks, even if the existence and name of the components is common.



Climate
Change

EQC PROTOTYPE

- Design and status
 - Bridging functions have been developed for communication s2dverification <-> downscaleR.
 - Work has been done in SPECS on bridging functions downscaleR <-> esd.
 - Each of the involved packages is being evolved to implement METACLIP.
 - New individual processing functions can optionally have attributes with key METACLIP information to be fully recognized by the provenance engine.
 - Expected result: framework to process (large) multidimensional arrays, in a delayed-operation mode.



Climate
Change

EQC PROTOTYPE

- Design and status
 - Time series plotting:
 - A generic time-series plotting engine is under development.
 - Uses the well known R package ggplot2.
 - Plots different kinds of time series automatically (anticipates user's needs) and plots titles, units and legends according to the available metadata.
 - Map plotting:
 - Various map plotting functions coming from multiple packages will be considered for the prototype.
 - It is not clear yet how to easily homogenize their look.



Climate
Change

EQC PROTOTYPE

- Interaction with the CDS
 - Each EQC framework should define its own strategy for data provenance. If possible, a common framework will be established by a “data standards” working group.
 - Current knowledge on CDS:
 - It will provide pointers to THREDDS servers distributed over Europe.
 - The CDS toolbox will provide a python framework that will allow creating “use cases” that call functionality from EQC prototypes. A loading function will be made available. (confirmation required)
 - We should not focus on the graphical interface, it will be provided by the CDS toolbox in a shiny-like fashion.
 - Workshop taking place soon on an early release of the CDS toolbox.



Climate
Change

EQC PROTOTYPE

- Interaction with the CDS
 - Current assumptions on CDS:
 - It will be possible to call our R packages from the CDS toolbox use-case definition environment via the python-R interface rpy2.
 - The CDS toolbox will run on a cloud computing platform, where a cluster of nodes will be abstracted as a single node with multiple computing cores.



Climate
Change

EQC PROTOTYPE

- Conclusions / potential areas for collaboration
 - Work has been done on collation of user requirements. Potential collaboration framework defined for other EQCs.
 - KPIs ready to be evolved by other EQCs.
 - Provenance solution designed, with plans for climate projections. QA4Seas could provide support for its implementation in DECM.
 - Development of bridging functions seems a first viable working approach to unify EQC software.