TOWARDS TRACEABILITY AND TRANSPARENCY

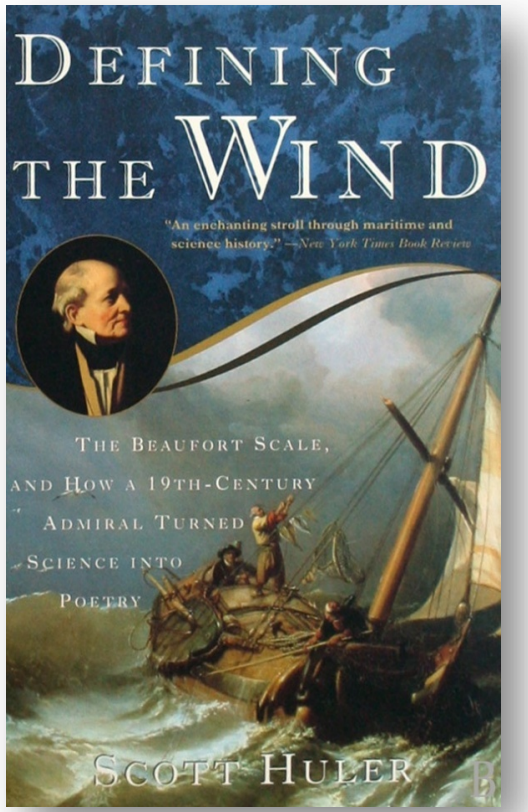
EUMETSAT

Rob Roebeling and Joerg Schulz

Thanks to CORE-CLIMAX Team

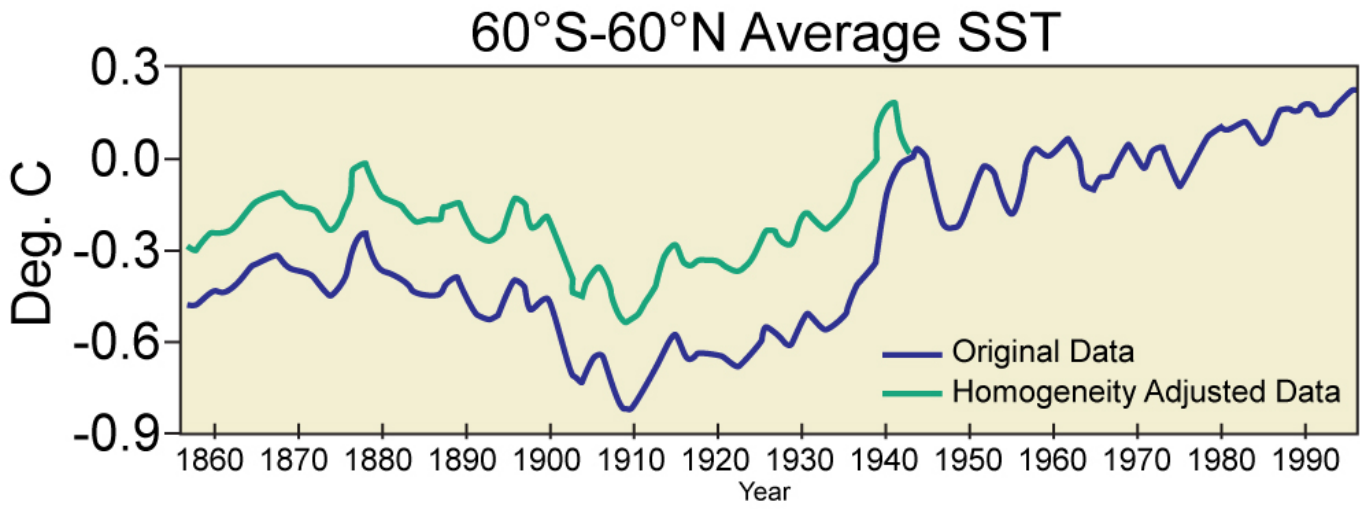
Motivation

- **What is at stake?**
- History shows that weather observations became useful for society after a lexicon was agreed to
  - ✓ The Beaufort scale did this for wind climatology and maritime commerce in the 19th century
- For the Climate Service to benefit society, it must adopt a lexicon that sets expectations, accessible to the public, for openness, process and transparency
  - ✓ How might we define a climate record lexicon useful to both scientists and general public in the 21st century?



Common Climate Observations
Building upon Best Practices

- **Steps to long-term monitoring**
  - Over the last 20-30 years many investigators have developed methods for seaming together observations with evolving coverage and accuracies
  - From these experiences, common elements are emerging on how climate scientists do business
- **How do we capture and make available these best practices?**



Examples of steps that can be taken

- **To evaluate if the Production System follows Best Practices;**
- **To provide infrastructure to compare the specifications of Data Records** *(e.g. ECV inventory: http://ecv-inventory.com);*
- **To work towards using standards cross different Data Records** *(e.g. data format, doi's, aggregation methods, multiple algorithm ensembles..);*
- **To improve the consistency between Data Records of different ECVs** *(e.g. clouds, precipitation, temperature and water vapor profiles, ..)*
- **To identify and address requirements brought up by operational and scientific users** *(e.g. nowcasting, numerical weather prediction, climate and weather model analysis, climate monitoring)*
- **…**

Evaluation of the

Data Record Production Process

- EU Core Climax -

- NOAA-NCDC -

- ESA-CCI -

Common Climate Observations
Approach proposed in FP7 CORE-CLIMAX

Three step approach to classify the maturity of ECV CDRs:

- **System Maturity Matrix (SMM)**

Evaluates if the production of the ECV CDR follows best practices for science, engineering and utilization;

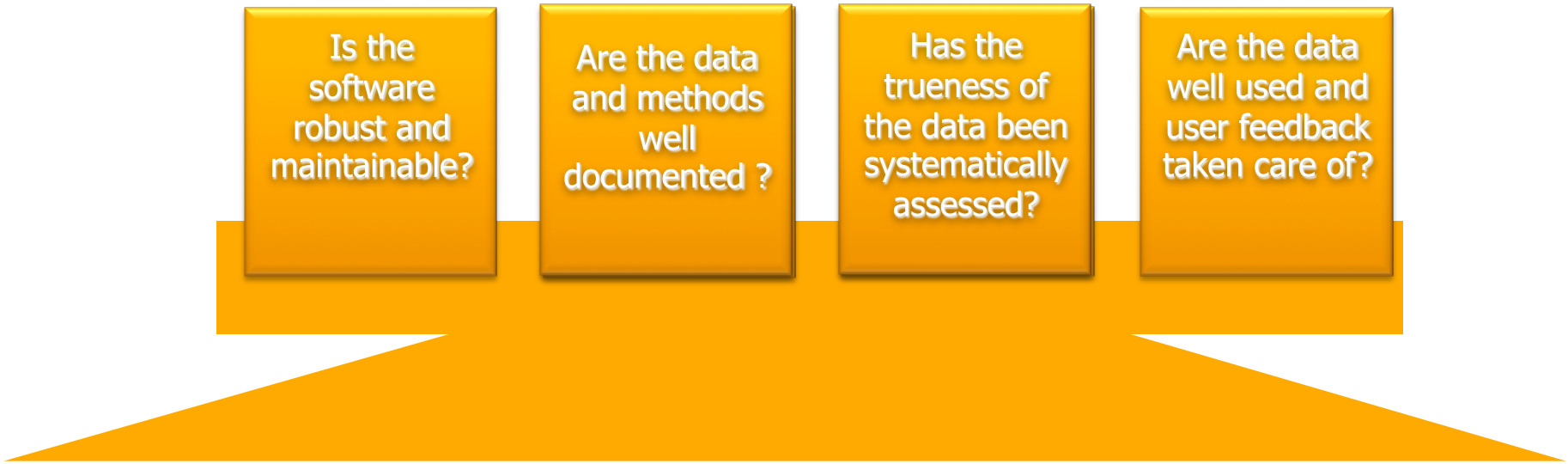
- **Data Record Inventories (DRI)**

Contains the Product Specification Tables and links to documented information on quality, calibration and inter-calibration *(e.g. ecv-inventory.com);*

- **Application Performance Metric (APM)**

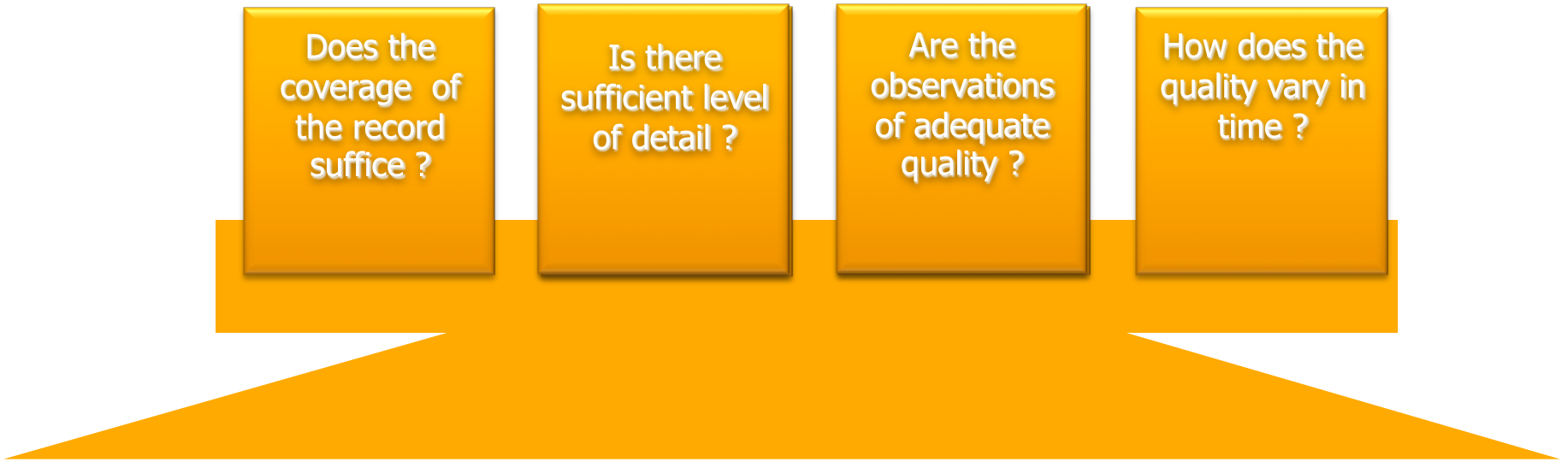
Evaluates the performance of an ECV CDR with respect to a specific application.

Concept - System Maturity Matrix



Software readiness	Metadata	User documentation	Uncertainty Characterization	Public Access, Feedback, Update	Usage
Are the codes compliant with standards, stable, portable and reproducible?	Do the metadata meet international standards, and allow provenance tracking?	Are the formal documents and peer-reviewed papers up-to-date and public?	Are the uncertainties assessed systematically in a standard manner?	Are the data, source code, and documents publicly available and regularly updated?	Are the data wildly used in the scientific, and decision and policy making communities?

Concept – Application Performance Matrix



Coverage	Sampling	Uncertainty	Stability
Are the record length and spatial coverage meeting the application's requirements?	Do the spatial and temporal sampling meet the applications requirements?	Do the random and systematic uncertainties meet the specifications?	Do the temporal and spatial stability meet the specifications?

Core-Climax **S**ystem **M**aturity **M**atrix Concept

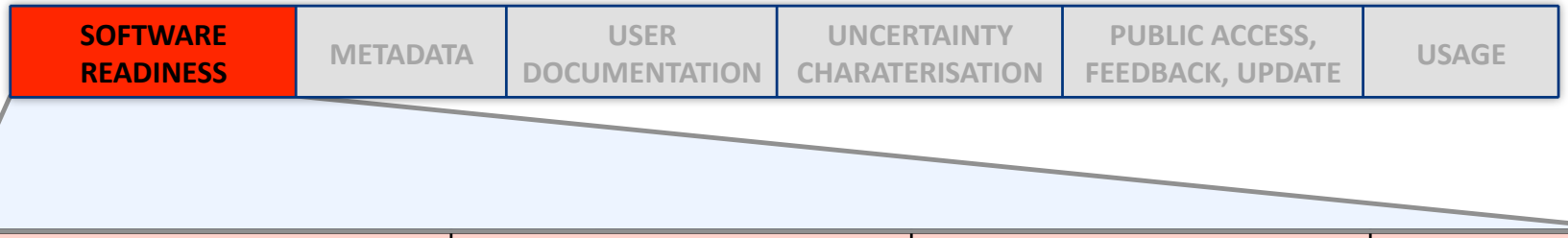
Acknowledgement

We thank ESA CCI, DWD, and CMSAF for testing earlier versions of the maturity matrices and Chris Merchant, University of Reading and John Bates, NOAA/NCDC for useful suggestions.

Core-Climax: System Maturity Matrix

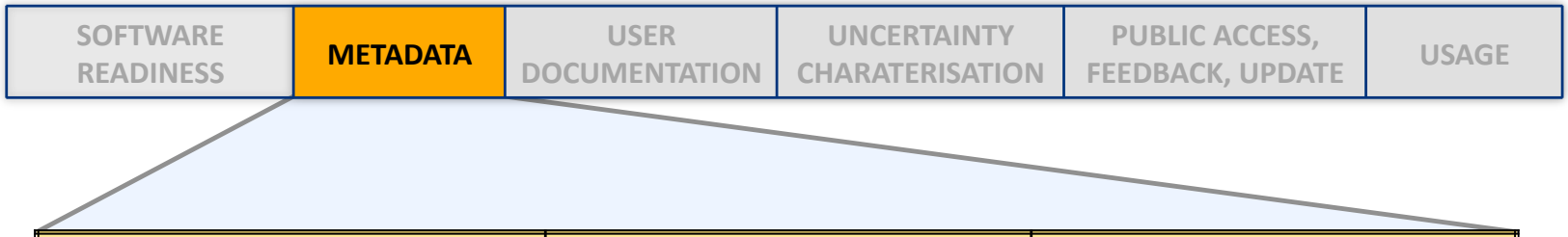
Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is sumitted for peer-review	Standard uncertainty nomenclature is idenitified or defined; limited validation done; limited information on uncertainty available	Data avaliable from PI, feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability, numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	**Score 2** + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations concept available from PI	**Score 2** + standard nomenclature applied; validation extended to full product data coverage, comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI, feedback through scientifc exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	**Score 3** + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	**Score 3** + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	**Score 3** + comprehensive scientific description available from data provider; report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	**Score 3** + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	**Score 3** + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	**Score 4** + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party installs the code operationally	**Score 4**+ fully compliant with standards; complete discovery metadata; complete location level metadata	**Score 4** + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	**Score 4** + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	**Score 4** + source code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	**Score 4+** Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	**Score 5** + fully compliant with standards; Turnkey System	**Score 5** + regularly updated	**Score 5** + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	**Score 5** + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	**Score 5** + source code available to the public and capability for continuous data provisions established (ICDR)	**Score 5** + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated

Core-Climax: Main Matrix and Sub Matrices



	Coding standards	Software Documentation	Numerical Reproducibility and Portability	Security
1	No coding standard or guidance identified or defined	No documentation	Not evaluated	Not evaluated
2	Coding standard or guidance is identified or defined, but not applied	Minimal documentation	PI affirms reproducibility under identical conditions	PI affirms no security problems
3	Score 2 + standards are partially applied and some compliance results are available	Header and process description (comments) in the code, README complete	PI affirms reproducibility and portability	Submitted for data provider's security review
4	Score 3 + compliance is systematically checked in all code, but not yet compliant to the standards.	Score 3 + a draft Software Installation/User Manual	3rd party affirms reproducibility and portability	Passes data provider's security review
5	Score 4 + standards are systematically applied in all code and compliance is systematically checked in all code. Code is not fully compliant to the standards. Improvement actions to achieve full compliance are defined.	Score 4 + enhanced process descriptions throughout the code; software installation/user manual complete	Score 4 + 3rd party can install the code operationally	Continues to pass the data provider's review
6	Score 5 + code is fully compliant with standards.	As in score 5	Score 5 + Turnkey system	As in score 5

Main Matrix and Sub Matrices



	Standards	Collection level	File level
1	No standard considered	None	None
2	No standard considered	Limited	Limited
3	Metadata standards identified and/or defined but not systematically applied	Sufficient to use and understand the data independent of external assistance; Sufficient for data provider to extract discovery metadata from meta data repositories	Sufficient to use and understand the data independent of external assistance
4	Score 3 + standards systematically applied at file level and collection level by data provider. Meets international standards for the dataset	Score 3 + Enhanced discovery metadata	Score 3 + Limited location (pixel, station, grid-point, etc.) level metadata
5	Score 4 + meta data standard compliance systematically checked by the data provider	Score 4 + Complete discovery metadata meets international standards	Score 4 + Complete location (pixel, station, grid-point, etc.) level metadata
6	Score 5	Score 5 + Regularly updated	Score 5

Main Matrix and Sub Matrices

	Formal description of scientific methodology	Formal Validation Report	Formal Product User Guide	Formal description of operations concept
1	Limited scientific description of methodology available from PI	None	None	None
2	Comprehensive scientific description available from PI and Journal paper on methodology submitted	Report on limited validation available from PI	Limited product user guide available from PI	None
3	Score 2 + Journal paper on methodology published	Report on comprehensive validation available from PI; Paper on product validation submitted	Comprehensive User Guide available from PI	Limited description of operations concept available
4	Score 3 + Comprehensive scientific description available from Data Provider	Report on inter-comparison to other CDRs, etc. Available from PI and data Provider; Journal paper on product validation published	Score 3 + available from data provider	Comprehensive description of operations concept available
5	Score 4 + Comprehensive scientific description maintained by data provider	Score 4 + Report on data assessment results exists	Score 4 + regularly updated by data provider with product updates and/or new validation results	Operations concept and description of practical implementation available
6	Score 5 + Journal papers on product updates published	Score 5+ Journal papers more comprehensive validation, e.g., error covariance, validation of qualitative uncertainty estimates published	Score 5	Score 5 + Operations concept regularly updated

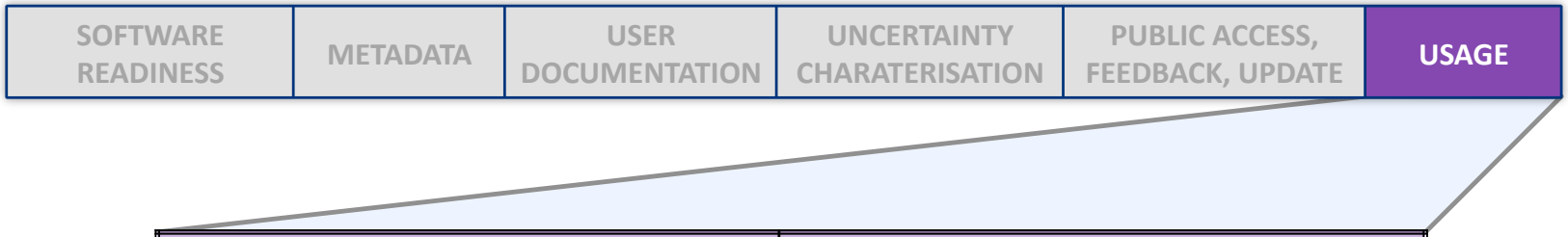
Main Matrix and Sub Matrices

	Standards	Validation	Uncertainty quantification	Automated Quality Monitoring
1	None	None	None	None
2	Standard uncertainty nomenclature is identified or defined	Validation using external reference data done for limited locations and times	Limited information on uncertainty arising from systematic and random effects in the measurement	None
3	Score 2 + Standard uncertainty nomenclature is applied	Validation using external reference data done for global and temporal representative locations and times	Comprehensive information on uncertainty arising from systematic and random effects in the measurement	Methods for automated quality monitoring defined
4	Score 3 + Procedures to establish SI traceability are defined	Score 3 + (Inter)comparison against corresponding CDRs (other methods, models, etc)	Score 3 + quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points	Score 3 + automated monitoring partially implemented
5	Score 4 + SI traceability partly established	Score 4 + data provider participated in one inter-national data assessment	Score 4 + temporal and spatial error covariance quantified	Score 3 + monitoring fully implemented (all production levels)
6	Score 5 + SI traceability established	Score 4 + data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle	Score 5 + comprehensive validation of the quantitative uncertainty estimates and error covariance	Score 5 + automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation

Main Matrix and Sub Matrices

	Public Access/Archive	Version	User Feedback Mechanism	Updates to Record
1	Data may be available through request to PI	None	None	None
2	Data available through PI	Preliminary versioning by PI	PI collects and evaluates feedback from scientific community	Irregularly by PI following scientific exchange and progress
3	Data and documentation archived and available to the public from PI	Versioning by PI	PI and Data provider collect and evaluate feedback and from scientific community	Irregularly by PI following scientific exchange and progress
4	Data and documentation archived and available to the public from Data Provider	Version control institutionalised	Data provider establishes feedback mechanism such as regular workshops, advisory groups, user help desk, etc. and utilises feedback jointly with PI	Regularly by PI utilising input from established feedback mechanism
5	Score 4 + source code archived by Data Provider	Fully established version control considering all aspects	Established feedback mechanism and international data quality assessment results are considered in periodic data record updates	Regularly operationally by data provider as dictated by availability of new input data or new methodology following user feedback
6	Score 5 + source code available to the public from Data Provider	Not used	Score 5 + Established feedback mechanism and international data quality assessment results are considered in continuous data provisions (Interim Climate Data Records)	Score 5 + capability for fast improvements in continuous data provisions established (Interim Climate Data Records)

Main Matrix and Sub Matrices



	Research	Decision Support System
1	None	None
2	Benefits for research applications identified	Potential benefits identified
3	Benefits for research applications demonstrated by publication	Use occurring and benefits emerging
4	Score 3 + Citations on product usage occurring	Score 3 + societal and economical benefits discussed
5	Score 4 + product becomes reference for certain applications	Score 4 + societal and economical benefits demonstrated
6	Score 5 + Product and its applications becomes references in multiple research field	Score 5 + influence on decision (including policy) making demonstrated

Examples of Typical Maturity Matrices

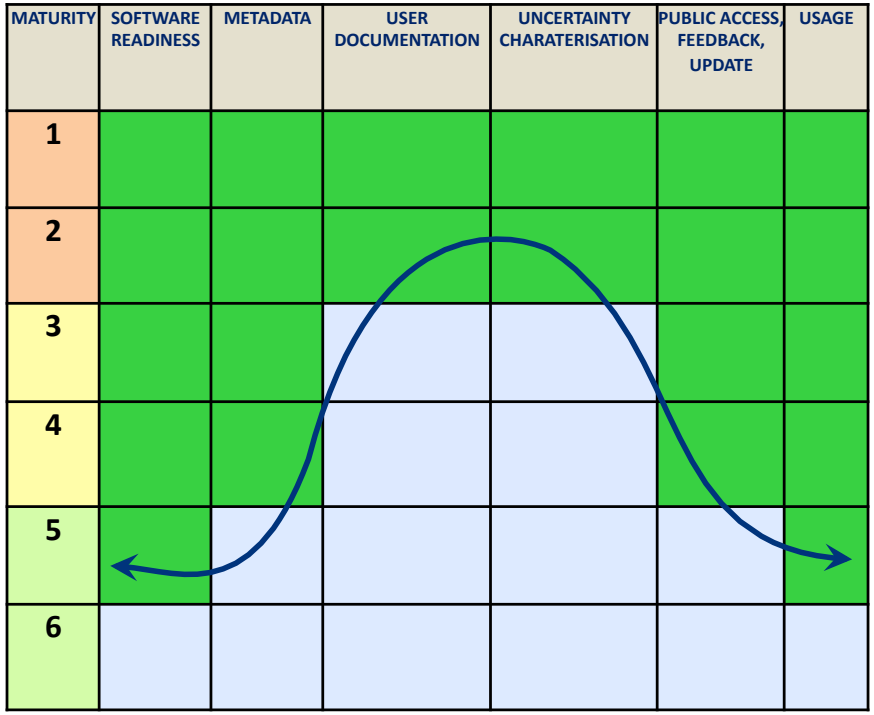


Fig: Typical SMM for a dataset from an operational provider (e.g. FCDR)

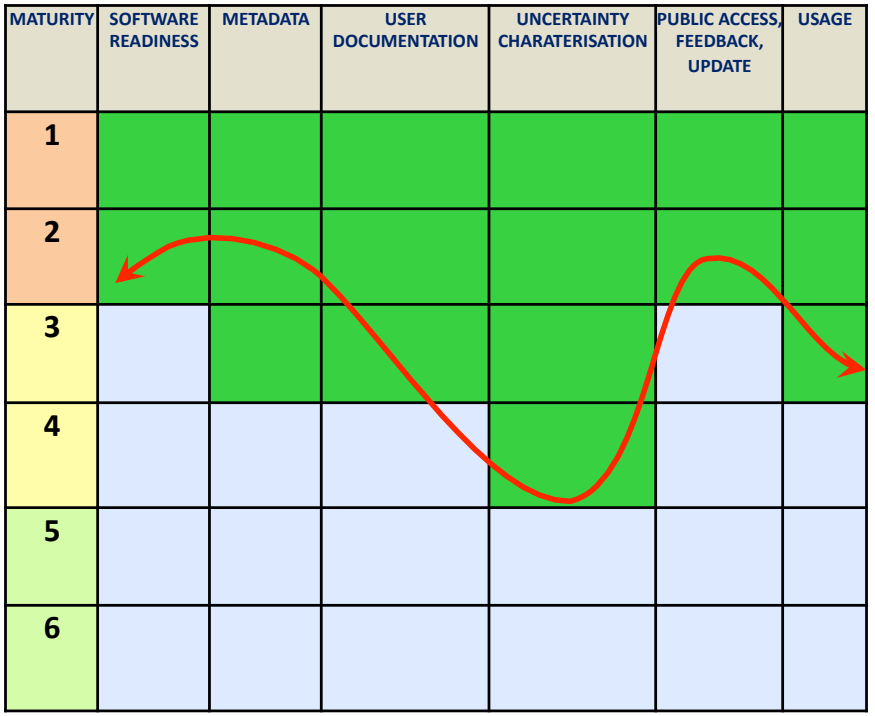


Fig: Typical SMM for a dataset from an scientific provider(e.g. TCDR)

Maturing Takes Time!

MATURITY	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARATERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	←					→
2						
3						
4						
5						
6						

Fig: Evolution of the SMM levels with time

TCDR ESA-CCI SST (~5 years)

ESA SST CCI AVHRR L2P long-term product version 1.0

maturity level as of 2013/2014

CORE-CLIMAX System Maturity Matrix

Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is asserted for peer-review	Standard uncertainty nomenclature is identified or defined; limited validation done; limited information on uncertainty available	Data available from PI; feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability; numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations concept available from PI	Score 2 + standard nomenclature applied; validation extended to full product data coverage; comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI; feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	Score 3 + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	Score 3 + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party assaults the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	Score 4 + soure code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	Score 4+ Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	Score 5 + fully compliant with standards; Turnkey System	Score 5 + regularly updated	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	Score 5 + source code available to the public and capability for continuous data provisions established (ICDR)	Score 5 + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated

TCDR CM-SAF Clouds (~12 years)

CLARA-A1 Cloud Properties

maturity level as of 19/12/2013

CORE-CLIMAX System Maturity Matrix

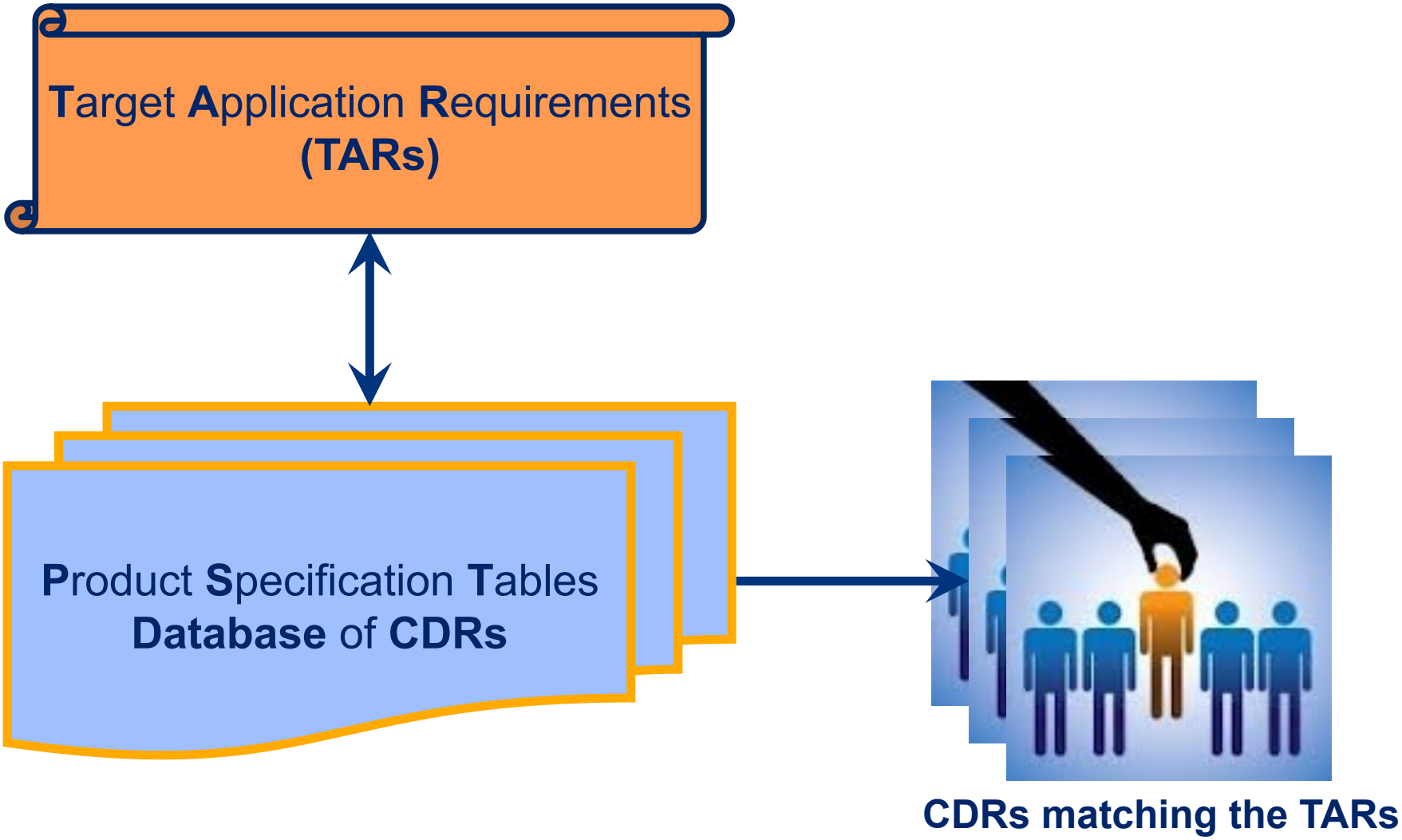
Maturity	SOFTWARE READINESS	METADATA	USER DOCUMENTATION	UNCERTAINTY CHARACTERISATION	PUBLIC ACCESS, FEEDBACK, UPDATE	USAGE
1	Conceptual development	None	Limited scientific description of the methodology available from PI	None	Restricted availability from PI	None
2	Research grade code	Research grade	Comprehensive scientific description of the methodology, report on limited validation, and limited product user guide available from PI; paper on methodology is asserted for peer-review	Standard uncertainty nomenclature is identified or defined; limited validation done; limited information on uncertainty available	Data available from PI; feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications identified DSS: Potential benefits identified
3	Research code with partially applied standards; code contains header and comments, and a README file; PI affirms portability; numerical reproducibility and no security problems	Standards defined or identified; sufficient to use and understand the data and extract discovery metadata	Score 2 + paper on methodology published; comprehensive validation report available from PI and a paper on validation is submitted; comprehensive user guide is available from PI; Limited description of operations concept available from PI	Score 2 + standard nomenclature applied; validation extended to full product data coverage; comprehensive information on uncertainty available; methods for automated monitoring defined	Data and documentation publically available from PI; feedback through scientific exchange, irregular updates by PI	Research: Benefits for applications demonstrated. DSS: Use occuring and benefits emerging
4	Score 3 + draft software installation/user manual available; 3rd party affirms portability and numerical reproducibility; passes data providers security review	Score 3 + standards systematically applied; meets international standards for the data set; enhanced discovery metadata; limited location level metadata	Score 3 + comprehensive scientific description available from data provider; report on inter comparison available from PI; paper on validation published; user guide available from data provider; comprehensive description of operations concept available from PI	Score 3 + procedures to establish SI traceability are defined; (inter)comparison against corresponding CDRs (other methods, models, etc); quantitative estimates of uncertainty provided within the product characterising more or less uncertain data points; automated monitoring partially implemented	Data record and documentation available from data provider and under data provider's version control; Data provider establishes feedback mechanism; regular updates by PI	Score 3 + Research: Citations on product usage in occurring DSS: societal and economical benefits discussed
5	Score 4 + operational code following standards, actions to achieve full compliance are defined; software installation/user manual complete; 3rd party assaults the code operationally	Score 4+ fully compliant with standards; complete discovery metadata; complete location level metadata	Score 4 + comprehensive scientific description maintained by data provider; report on data assessment results exists; user guide is regularly updated with updates on product and validation; description on practical implementation is available from data provider	Score 4 + SI traceability partly established; data provider participated in one inter-national data assessment; comprehensive validation of the quantitative uncertainty estimates; automated quality monitoring fully implemented (all production levels)	Score 4 + soure code archived by Data Provider; feedback mechanism and international data quality assessment are considered in periodic data record updates by Data Provider	Score 4+ Research: product becomes reference for certain applications DSS: Societal and economic benefits are demonstrated
6	Score 5 + fully compliant with standards; Turnkey System	Score 5 + regularly updated	Score 5 + journal papers on product updates are and more comprehensive validation and validation of quantitative uncertainty estimates are published; operations concept regularly updated	Score 5 + SI traceability established; data provider participated in multiple inter-national data assessment and incorporating feedbacks into the product development cycle; temporal and spatial error covariance quantified; Automated monitoring in place with results fed back to other accessible information, e.g. meta data or documentation	Score 5 + source code available to the public and capability for continuous data provisions established (ICDR)	Score 5 + Research: Product and its applications becomes references in multiple research field DSS: Influence on decision and policy making demonstrated

Transparency of the available Data Records **A**pplication **P**erformance **M**etric Concept

Acknowledgement

We thank ESA CCI, DWD, and CMSAF for testing earlier versions of the maturity matrices and Chris Merchant, University of Reading and John Bates, NOAA/NCDC for useful suggestions.

Concept of APM



Target Application Requirements (TAR)

General Query Parameters	Input
Essential Climate Variable (ECV)	<ECV name>
Temporal Sampling	<Threshold><Break through><Optimum>
Horizontal Sampling	<Threshold><Break through><Optimum>
Vertical Sampling	<Threshold><Break through><Optimum>
Temporal Coverage	<Threshold><Break through><Optimum>
Spatial Coverage	region of interest (e.g. Global, Europe, Africa, etc...)

Specific Query Parameters	Input
Uncertainties	<Threshold><Break through><Optimum>
Stability	<Threshold><Break through><Optimum>
Statistics	quantities (e.g. mean, error estimate, histograms, etc..)
Sensitivity to auxiliary data (Purity)	%
Consistency with other ECVs	<>
Continuity	<>
<Other Suggestions>	

Example: Product Specification Table (PST)

ORGANISATION INFORMATION	
Indicator	**Example Input**
Respondent name	Rainer Hollmann (DWD)
Respondent e-mail	rainer.hollmann@dwd.de
…etc...	…etc...

CLIMATE DATA RECORD SPECIFICATIONS	
Indicator	**Example Input**
Essential Climate Variable (ECV)	cloud water path
Systematic uncertainty (bias)	15%
Random uncertainty (rms)	30%
Temporal Stability	15 g/m2
…etc...	…etc...

DOCUMENTATION & DATA ACCESS	
Indicator	**Example Input**
Algorithm Theoretical Baseline Document (ATBD) (link)	http://wui.cmsaf.eu/safira/action/viewProduktDetails?id=20288
…etc...	…etc...

Discussion

Common Data Format

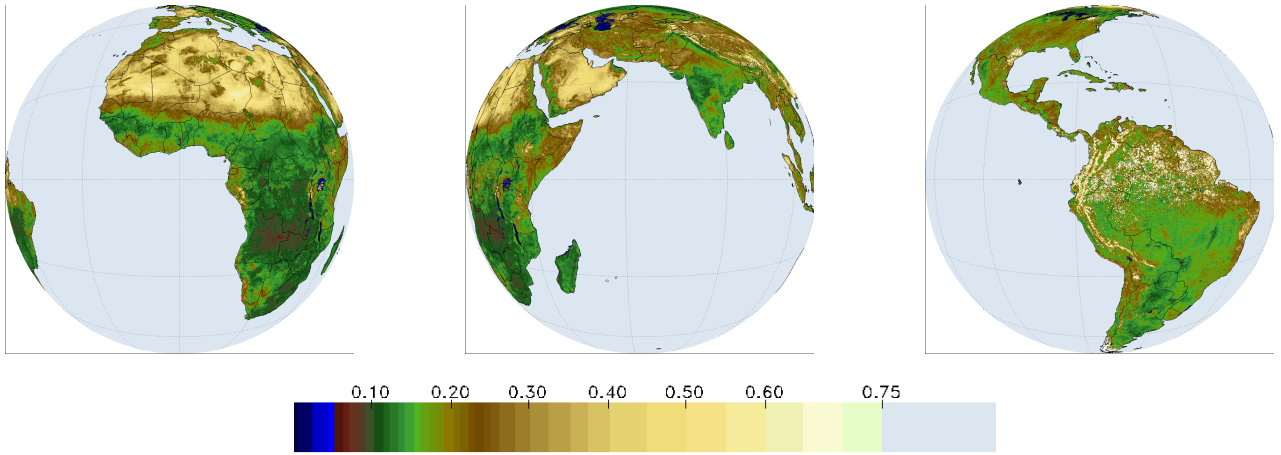
Common Coding Standards

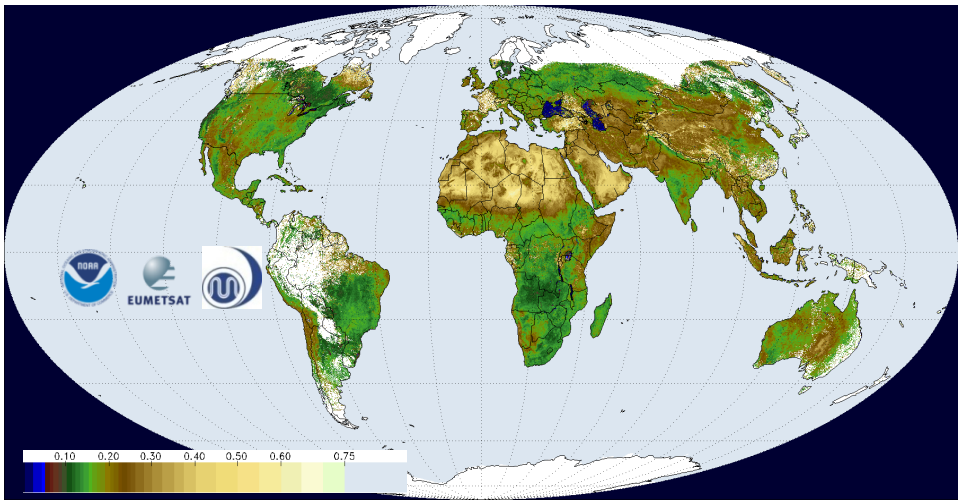
Common Data Formats and Common Metadata

Prime Mission centred over Africa

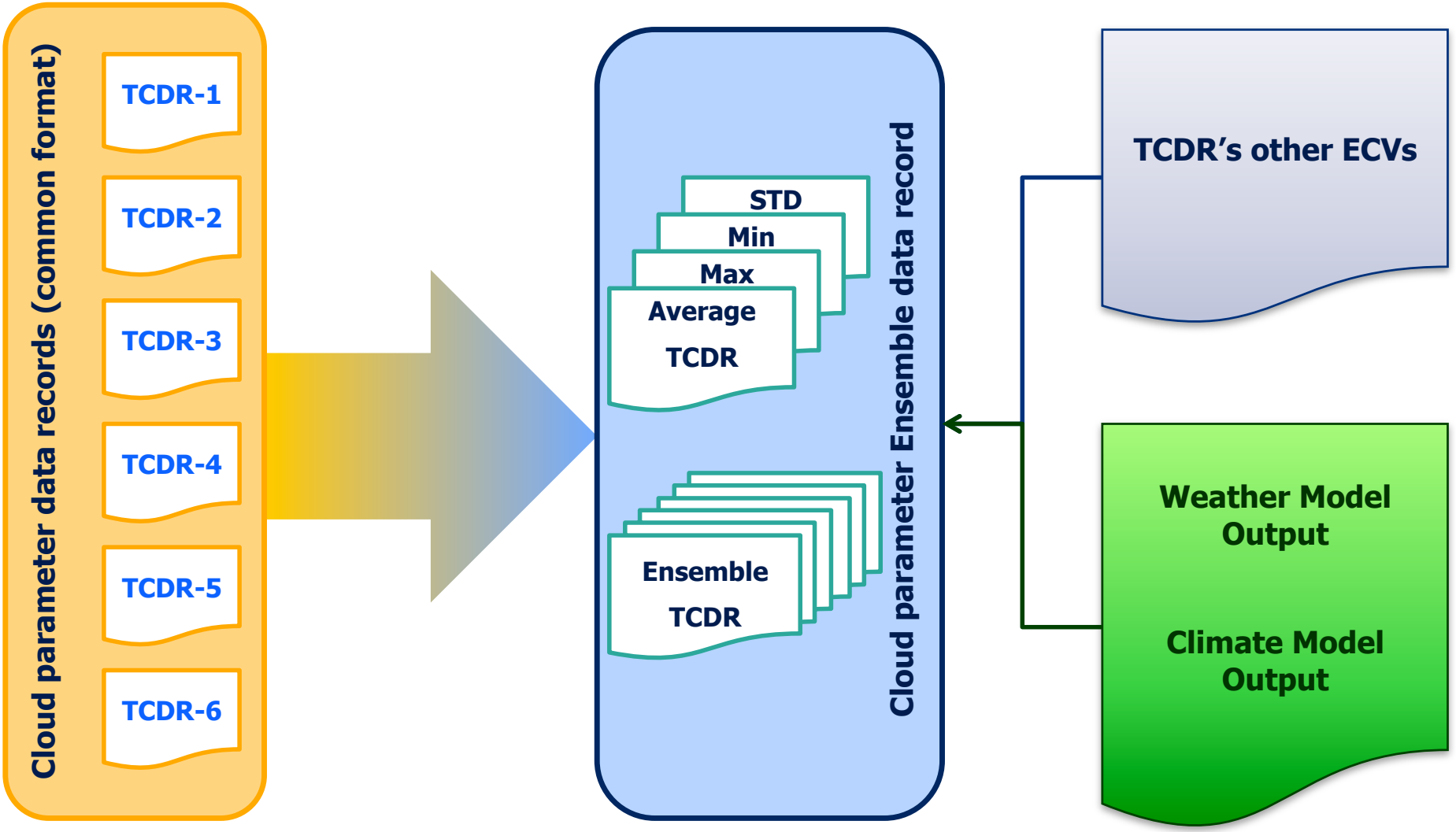
Indian Ocean Data Coverage

Atlantic Data Coverage





Common Data Formats and Common Grids

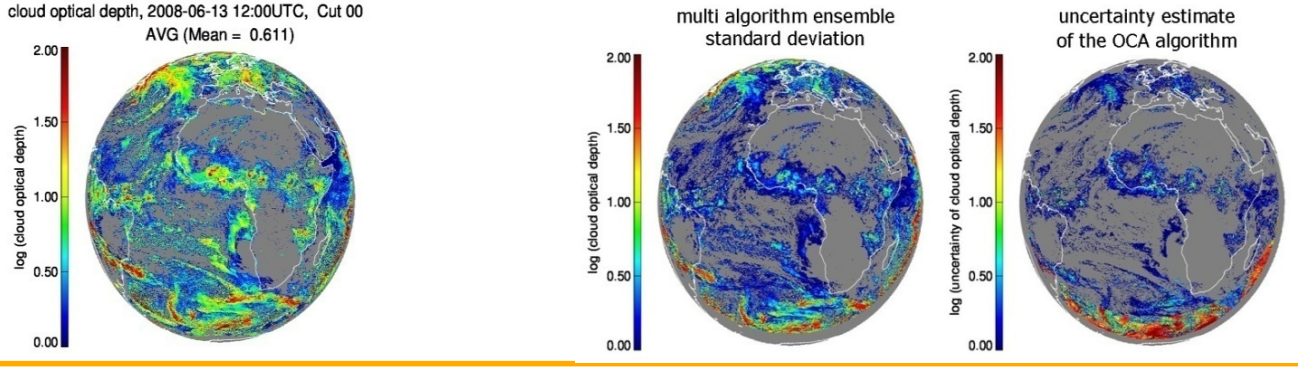


** TCDR: **T**hematic **C**limate **D**ata **R**ecord*
** ECV: **E**ssential **C**limate **V**ariable*

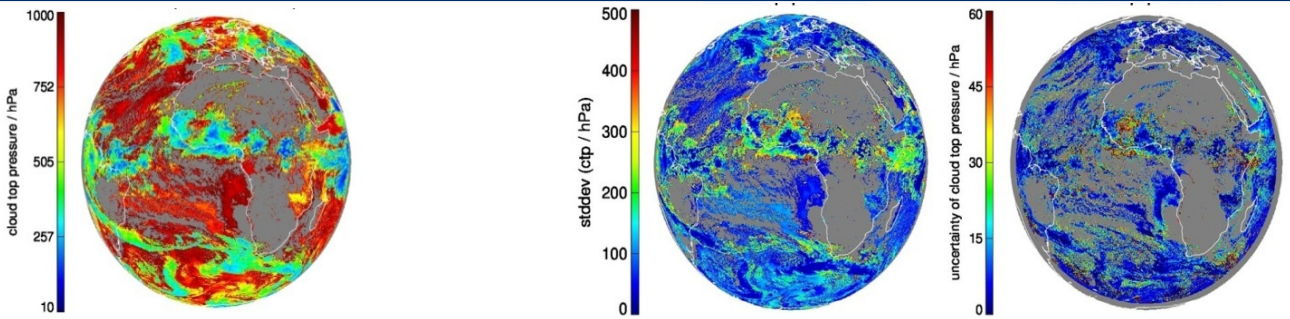
Common grid

Courtesy of Ulrich Hamann , KNMI

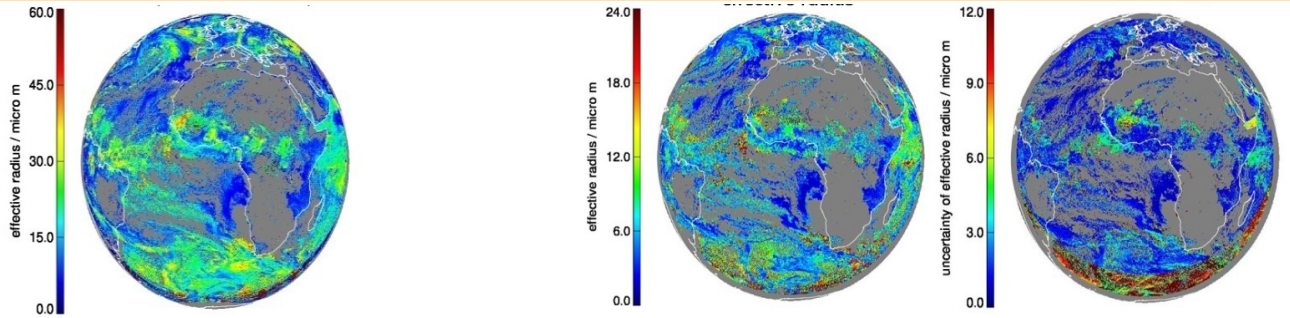
COT



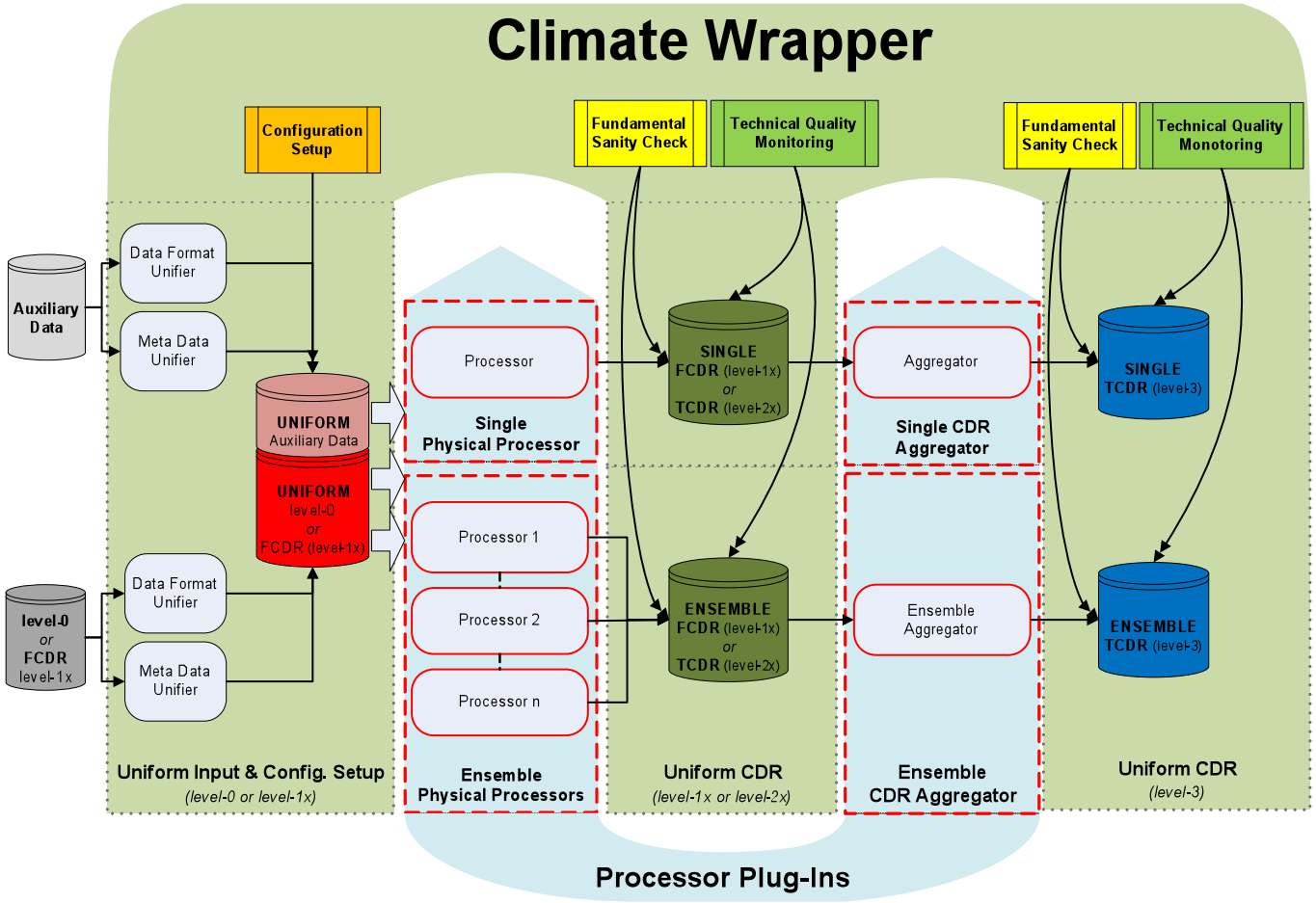
CTP in hPa



Re in micron



Common Coding Standards and I/O concepts



Summary

Summary

- International programs ask transparency and assessments of product production facilities and specifications of product data records (ESA, EU, NCDC, EUMETSAT)
- There is consensus on manner of assessing the maturity of production system
- A pilot on manners of assessing the applicability of data records is starting
- Users ask for more coordination on sharing data formats and metadata

Thank You

Any Questions?

EUMETSAT