Community Whitepaper on Uncertainty Quantification

David Moroni[1], Hampapuram "Rama" Ramapriyan[2], Ge Peng[3]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

[2]Science Systems and Applications, Inc. & NASA Goddard Space Flight Center

[3]North Carolina State University, Cooperative Institute for Satellite Earth System Studies - (CISESS) at NOAA's National Centers for Environmental Information (NCEI)

POS Panel Breakout Session on Uncertainty Quantification

US-CLIVAR Summit – August 6, 2019

Long Beach, CA

This work was a result of the authors' participation in the ESIP IQC. Ramapriyan's work was supported by NASA under a contract with Science Systems and Applications, Inc. Peng's work was supported by NOAA under a grant with NCSU. Moroni's work was supported by NASA under a contract with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. U.S. government sponsorship is acknowledged.

ESIP Information Quality Cluster

- Vision: "Become **internationally recognized** as an **authoritative and responsive information resource** for guiding the implementation of **data quality standards and best practices** of the science data systems, datasets, and data/metadata dissemination services."
- Information Quality = {Science Quality, Product Quality, Stewardship Quality, Service Quality}
- What do we do?
  - Share experiences; collaborate internationally; invited speakers at monthly telecons; sessions and/or presentations at AGU, AMS, ESIP, E2SIP, and OGC meetings
  - Maintain wiki site with many useful references http://wiki.esipfed.org/index.php/Information_Quality
- Publications
  - Peng, G. *et al.,* 2016: Scientific stewardship in the Open Data and Big Data era - Roles and responsibilities of stewards and other major product stakeholders. *D.-Lib Magazine,* **22** (5/6), doi: https://doi.org/10.1045/may2016-peng.
  - Ramapriyan, H K, Peng G, Moroni D, Shie C-L, Ensuring and Improving Information Quality for Earth Science Data and Products. *D-Lib Magazine*, **23** (7/8), July/August 2017, DOI: https://doi.org/10.1045/july2017-ramapriyan
  - Moroni, et al.(22 authors), "Understanding and Communicating Uncertainty in Earth Science Data Informatics", White Paper (in preparation)

Information Quality - Definition

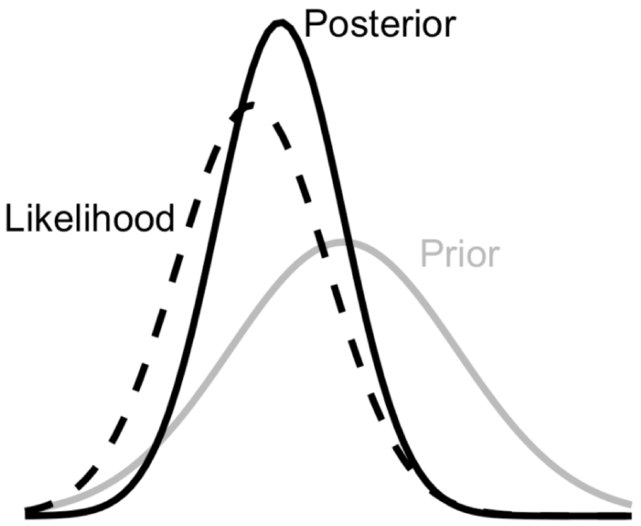
- Scientific quality
  - Accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose) in various applications
- Product quality
  - How well the scientific quality is assessed and documented
  - Completeness of metadata and documentation, provenance and context, etc.
- Stewardship quality
  - How well data are being managed, preserved, and cared for by an archive or repository
- Service Quality
  - How easy it is for users to find, get, understand, trust, and use data
  - Whether archive has people who understand the data available to help users.

White Paper Scope

- Began as a pilot plenary/breakout session at the ESIP 2017 meeting, featuring invited researchers: Carol Anne Clayson (WHOI), Isla Simpson (NCAR), and Amy Braverman (JPL).
- Primary focus on “discovery” of the breadth of approaches with regard to Earth science data UQ, UC, and the dissemination/utilization of UQ/UC information by data providers and end users.
- Considers 4 perspectives: Mathematical, Programmatic, Observational, User.
- Will identify both commonalities and differences between perspectives.
- Authors and co-authors represent various aspects of Earth science data informatics, metrology, data science/statistics, remote sensing, in situ, and disciplinary fundamental research.
- Numerical modeling was considered for the sake of use case discussion, but was decided to be left out for the sake of focusing on approaches using observational data.

Mathematical

- Championed by Jonathan Hobbs - JPL
- Considered to be the foundational section of the paper, establishing the key mathematically-based definitions of uncertainty and related constructs such as UQ, UC, mean square error, PDFs, quantiles, confidence intervals, confidence levels, etc…
- Presents directly applicable use cases by which these mathematical definitions are applicable to observational Earth science data, primarily from a remote sensing perspective, but much of which utilizes consistent metrology for a variety of measurement types, including in situ and sub-orbital.



Schematic implementation of Bayes' theorem for a univariate QOI. The prior distribution is combined with information from an observation (via the likelihood) to produce a posterior distribution.

Programmatic

- Championed by Rama – SSAI/NASA GSFC.
- Captures the governmental and intergovernmental approaches, starting with specific US-based agencies and moving into the international arena.
- Considers US law that drives policy at key agencies, including but not limited to NASA and NOAA.
- Considers international agreements, such as by the U.N, IPCC, WMO, and CEOS.
- Considers multi-lateral agreements, statements and policies by EU-sponsored agencies/organizations, such as by: ESA, FIDUCEO, UncertWeb, and MetEOC.

Observational

- Championed by Justin Goldstein – NOAA.
- Discusses the foundational approaches to UQ and UC from an Earth observation perspective, including perspectives from both point-based studies, invariant in space but not in time (e.g., Eulerian Specifications), and those that conduct observations varying in *both* space and time (e.g., Lagrangian Specifications).
- Cal/Val: looks at UQ and UC approaches from a calibration and validation perspective and the role played by "ground truth" data.
- Product Development: examines a variety of approaches and considerations toward making uncertainty information available for common types of observational data products, with a focus on making this information available at the production stage of data.

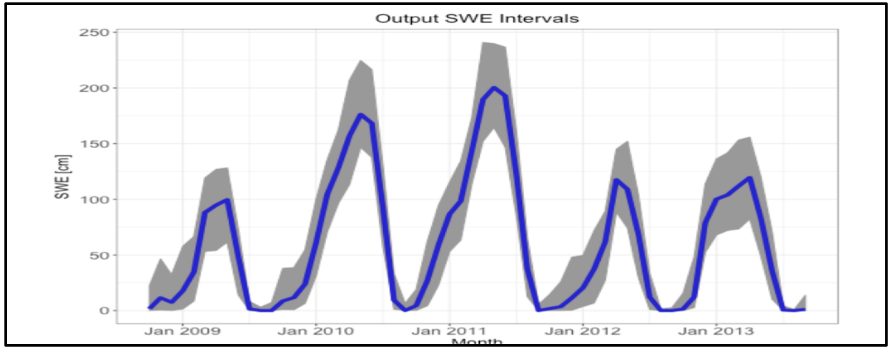


Figure: Snow water equivalent with uncertainty shaded with 95% confidence intervals. Blue line represents the point estimate. Hobbs et al. 2017, unpublished.

User

- Championed by Bob Downs – Columbia University.
- Focuses on the ways in which uncertainty information can be effectively or ineffectively consumed, interpreted and ultimately leveraged by the typical data user.
- Provides insights in to methods of communication, dissemination, visualization tools/services, and multi-variate analysis.
- Examples considered include: ISO-19157, UncertML, CO2SYS, and OGC's Testbed-12 innovation program (OGC, 2017).

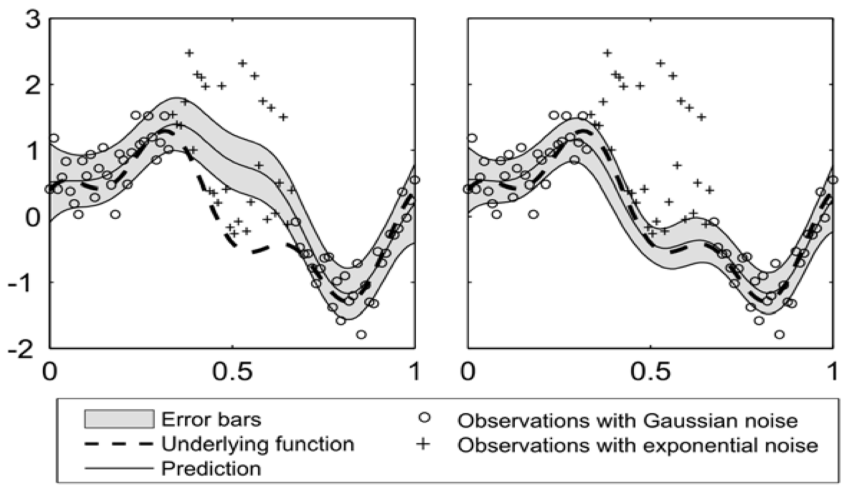


Figure: Applying UncertML to automated Bayesian interpolation algorithm. Left plot assumes Gaussian random error while right plot incorporates obs-specific error characteristics via UncertML encoding, improving the performance of interpolation. Williams et al. 2009.

Next Steps

- Complete by August:
  - Commonalities, differences, conclusions.
  - Re-write the introduction to better align with main sections.
  - Include more graphics/figures.
- Complete by September
  - Prep for white paper publication; consult with ESIP student fellow to apply improved styling and consistent references/citation styling adhering to AGU standard.
  - Publish on ESIP Figshare site.

Ideas beyond this publication…

- Draft and publish a shortened “executive summary” paper in a more prominent journal, such as Data Science or EOS.
- Draft a part-2 paper, focusing on recommendations and actionable solutions.

Authorship Acknowledgements (19)

Jonathan Hobbs[1], Robert Wolfe[4], Chung-Lin Shie[5], Christopher J. Merchant[6], Janae Csavina[7], Mark Bourassa[8], *Isla Simpson*[9], Jessica L. Matthews[3], *Matthew Plumlee*[10], Peter Cornillon[11], Justin C. Goldstein[12], Lucy Bastin[13], Kenneth Kehoe[14], Benjamin Smith[15], Jeffrey L. Privette[15], Robert R. Downs[16], Aneesh C. Subramanian[17], Otis Brown[3], Ivana Ivánová[18]

Co-Author Affiliations

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
2. Science Systems and Applications, Inc. & NASA Goddard Space Flight Center
3. North Carolina State University (NCSU), Cooperative Institute for Climate and Satellites-North Carolina (CICS-NC) at NOAA’s National Centers for Environmental Information (NCEI), Asheville, NC
4. NASA Goddard Space Flight Center
5. NASA Goddard Space Flight Center and University of Maryland, Baltimore County
6. University of Reading and National Centre for Earth Observation, Reading, UK
7. National Ecological Observatory Network
8. Center for Ocean-Atmospheric Prediction Studies, Florida State University
9. National Center for Atmospheric Research
10. Northwestern University
11. University of Rhode Island
12. Riverside Technology, Inc. supporting the NOAA NESDIS Technology, Planning, and Integration for Observation division
13. Joint Research Centre, European Commission / Aston University, UK
14. University of Oklahoma- Cooperative Institute for Mesoscale Meteorological Studies
15. NOAA’s National Centers for Environmental Information
16. Center for International Earth Science Information Network, Columbia University
17. University of Colorado Boulder
18. Curtin University, Australia