Observing System Development and UQ in a Parallel Bayesian Framework: Applications for Weather, Clouds, Convection, and Precipitation

27 May 2021

Presenter and PI: Derek J. Posselt\textsuperscript{1}

Team Members: Brian Wilson\textsuperscript{1}, Rachel Storer\textsuperscript{2}, Derek Tropf\textsuperscript{1}, Vishal Lall\textsuperscript{1},
George Duffy\textsuperscript{1}, Matt Lebsock\textsuperscript{1}, Simone Tanelli\textsuperscript{1}

Program: AIST-18

\textsuperscript{1}Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
\textsuperscript{2}University of California, Los Angeles, CA

©2021. All rights reserved. A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.
• Technologies in this session provide information highly relevant for weather
• This talk describes information technology for weather formulation
• Rapidly expanding trade space – what is needed? Which instruments? Combinations/constellations? Accuracy?
• We have developed a system that is designed to more thoroughly and efficiently explore the science trade-space for new missions.
• It is flexible, parallelizes over diverse architectures, and includes several robust techniques with which to measure uncertainty.
• When combined with tools (e.g., TAT-C) that assess sampling needs (orbits, swaths, etc) it is now possible to evaluate a much larger number of options.
Scientific Challenge

- Clouds and precipitation are central to weather and climate
- After decades of space-borne measurements, *key processes are still missing*
- Goal: design a new observing system (e.g. ACCP*)
  - Address specific science objectives
  - Consider the vast array of possible measurements
  - Rigorously quantify uncertainties

*Aerosol, Clouds, Convection, and Precipitation [https://science.nasa.gov/earth-science/decadal-accp](https://science.nasa.gov/earth-science/decadal-accp)
Technical Challenge

- The design trade-space is *large* and clouds are *diverse*
- The dimensionality of the design problem is *immense*
  - Multiple different geophysical scenarios (different cloud types)
  - Diversity of measurement types (active, passive, single-point, distributed)
  - Multiple sources of uncertainty (instrument noise, forward models, ambiguity)

For each geophysical variable

For each cloud type

For each combination of measurements

Considering all sources of uncertainty

Determine whether measurements meet mission requirements
Solution: Accelerate OSSEs

Any **observing system simulation experiment** (OSSE) requires at least four components:

1. **Nature run**: Realistically represent the real world
Solution: Accelerate OSSEs

Any **observing system simulation experiment** (OSSE) requires at least four components:

1. **Nature run**: Realistically represent the real world
2. **Instrument simulators**: Synthetic measurements
Solution: Accelerate OSSEs

Any **observing system simulation experiment** (OSSE) requires at least four components:

1. **Nature run**: Realistically represent the real world
2. **Instrument simulators**: Synthetic measurements
3. **Quantify uncertainty**: Sources of noise and error

![Graph showing Measurement Uncertainty and Geophysical Variable Uncertainty]
Any **observing system simulation experiment** (OSSE) requires at least four components:

1. **Nature run**: Realistically represent the real world
2. **Instrument simulators**: Synthetic measurements
3. **Quantify uncertainty**: Sources of noise and error
4. **Assess impact***: Did observations meet science and applications goals and objectives?

*NWP (weather forecast OSSE) is just one example of impact. OSSEs must grow to encompass advances in knowledge and traceability to applications.
Parallel OSSE Toolkit for Mission Design

**OSSE Components**
- **ParMAP**: Flexible Parallelism
- **Clusters and HPC**
- **Cloud Computing**: AWS
- **OSSE Components**
  - **Nature Runs**
    - Large Eddy Simulations
    - Cloud Resolving Models
    - Global Simulations
  - **Instrument Simulation**
    - Radar
    - Passive Microwave (Extensible via pluggable containers)
  - **Bayesian Retrievals**
    - Optimal Estimation
    - Ensemble Kalman Filter
    - Markov chain
    - Monte Carlo

**Measurement Uncertainty**

**Uncertainty Analysis**

**Mission Design Decisions**

2017 EARTH SCIENCE DECADAL SURVEY
DESIGNATED OBSERVABLE

Derek Posselt@jpl.nasa.gov

2021 Earth Science Technology Forum
Example 1: Uncertainty Inherent in Clouds

- Accurate estimates of cloud properties and evolution are important
  - Precipitation
  - Atmospheric dynamics
  - Earth’s radiative balance
  - Chemical reactions
- Many processes of interest are governed by cloud microphysics:
  - Phase change, collisions, etc
  - Particle size, number, and shape
Example 1: Uncertainty Inherent in Clouds

- There is a diversity of available remote sensing measurements
- All are sensitive to some degree to cloud microphysics
- What are the measurement requirements for successfully observing cloud properties and processes?
Example 1: Uncertainty Inherent in Clouds

• Quick example: run a radar simulator using two different sets of ice shapes (spheres vs multiple habits)
Example 1: Uncertainty Inherent in Clouds

• Quick example: run a radar simulator using two different sets of ice shapes (spheres vs multiple habits)
Example 1: Uncertainty Inherent in Clouds

- Quick example: run a radar simulator using two different sets of ice shapes (spheres vs multiple habits)
Example 1: Uncertainty Inherent in Clouds

- Quick example: run a radar simulator using two different sets of ice shapes (spheres vs multiple habits)
Example 1: Uncertainty Inherent in Clouds

Experiment Configuration:
• 2 input model profiles
• 3 radar frequencies (Ku, Ka, W)
• 5 uncertain parameters, 11 possible values each
• $2 \times 3 \times 11^5 = 966,306$ forward model runs

Inputs:
• Nature run profiles
• Range of uncertainty

Outputs:
• Ensemble of possible radar profiles for each input model profile and frequency
• Improved understanding of uncertainty in radar observations of convection

Posselt et al. 2021 (IGARSS)
Example 2: Shallow Cloud Retrieval

- Shallow convection is crucial for climate (hydrologic cycle and cloud-radiation feedbacks)
- Rain retrievals are challenging: sensitive to radar design parameters (sensitivity, footprint, surface clutter)
- Constructed an optimal estimation (Bayesian) retrieval based on the CloudSat algorithm
- Conducted an initial test of retrieval uncertainty using 6000 shallow rain profiles from nature run
Parallel OSSE System (ParOSSE) Performance

• Sensitivity and retrieval experiments are embarrassingly parallel (can be done nearly independently)

• ParMAP library makes ParOSSE deployable on a single machine (Par), cluster (Dask), and AWS Lambdas

• Our initial tests have indicated excellent scaling efficiency*

*Efficiency > 1 is due to I/O limitations with a single CPU
ParOSSE Capability to Date

• Pluggable nature runs and instrument simulators enable a wide range of trade space studies

• Flexible parallelism enables experiments on diverse architectures and more thorough exploration of uncertainty in measurements and retrievals

• Have implemented various sensitivity analysis techniques
  • Method of Morris, Sobol sensitivity, Monte Carlo, grid search

• Retrievals can utilize several Bayesian methodologies
  • Optimal estimation, MCMC, ensemble Kalman filter, Gamma-Inverse Gamma filter

Derek.Posselt@jpl.nasa.gov

User Interface
(Python Notebooks, Web Services API)

- Nature Runs
- Measurement Design Parameters
- Profile Selection
- Sources of Uncertainty

JSON Docs, Input Tagging

- Parallel MAP
- Ensemble of Measurement Simulations and Geophysical Variable Retrievals

Knowledge Base (Elasticsearch)

- Parallel Reduce
- Compute Statistics Evaluate Simulation Output Produce Plots

- Python Wrappers Docker Containers
Future Directions: Span the SATM?

Can we quantify ability to meet science and applications goals and objectives?

**Science:**
- Quantify state of knowledge – sources of uncertainty and relevant variables?
- Models as a laboratory, and ensembles as the tool.
- ParOSSE is flexible - spawn ensembles of *process simulations* and assess reduction in uncertainty (metrics from information theory, ensemble forecasting, etc)

**Applications:**
- Map from GV uncertainty to uncertainty in stakeholder quantities of interest (e.g., rainfall duration and intensity vs. needs of reservoir managers)
- Note: NOAA’s ASPEN system considers a large database of user-defined requirements and then quantifies observing system effectiveness by inputting expected GV uncertainty.
Example: Convection-Environment Interaction

• Which observations are necessary to improve state of knowledge of convective storms?

• First: determine which are the most important control variables

• How? Models as a laboratory

• This is a small number of runs of one case, each with a slightly different environment

• Can we scale up to many types of convection in many different environments?

• ParOSSE’s flexible configuration makes this straightforward

Cross-section through ensemble of 25 simulations of deep convection, showing transport of pollution from the boundary layer upward into the free troposphere.
References


