

Breakout_C-3 (can map to issue name later)

Issue: https://github.com/CFDSI/Kickoff_Workshop/issues/#

Related Issues:

Issue Statement: [Prediction and Calibration - standardizing protocols for using data from experiments and observations for robust prediction.](#)

Discussion topic:

Moderator: Jed Brown

Note taker: Assad Oberai

Reporter:

Group Members:

Abani Patra, Steve
Wojtkiewicz
Omar Knio
Bob Moser
Subhayan De
Paul Bauman
Alireza Doostan
Ryan Skinner
Jed Brown
Assad Oberai

Please address these topics in your discussion (moderators please make sure that there is enough time to cover all three before the session ends).

1. Describe the problem:

- **Anything that is said below applies to experimental and computational data!**
- **Model provenance:** All the inputs to a model must be specified: data, specification of geometry, boundary conditions, LES model - prescription, how long was it run, calibration.
- **Calibration is done with modest data.** This should restrict the domain over which predictions should be made. Therefore it is **important to state up-front what data you have calibrated** with. CFDSI could help with formalizing this.
- Recursive thinking: have run into something that I cannot predict - so need to

rethink/re-calibrate our model.

- This is **beyond verification and validation**.
- Define calibration errors, and use error analysis to characterize errors.
- Calibration is a statistical inference problem which is hard to solve/understand.
- **Data semantics** - what does the data actually mean?

2. What are potential solutions?

- **Protocols** for provenance, calibration need to be established. (look at Kaggle - for ML).
- Encouraging internal sanity checks. **Full disclosure**.
- **Common language** to describe and quantify errors/uncertainty among experiments and computation.
- CFDSI should **bring together experimental and computational collaborations**. Then this will become a part of our culture.
- How do we accomplish **more complete reporting** of error? **Elevate folks who do this. Provide software to make this happen.**
- **Reproducibility - incentivize this for experiments and software.**
- Standardized software tools for calibration solved as an inverse problem.
- Is **working with raw data** a solution? If not raw, then provide provenance - **report how the raw data was treated.**
- Deeper sense of provenance - entire data tracking chain. **Recognizing that data** (computational or experimental) **is probabilistic.**
- **Data passport** - data should be stamped each time it is processed.
- **Proprietary data processing** is tough to unravel. Can CFDSI help with a greater bargaining power.
- Tools that have been used to process the data can be stored in a repository in CFDSI.
- Experimental design that accounts for the what will be done with the data.
- **Badging system** for data
- **Data Quality Hierarchy**
 1. Baseline - data is posted on the website. Some metadata.
 2. Next tier - Provide some information about data processing.
 3. Higher tier - Good characterization of uncertainty; open source pipeline all the way to “raw” sources

3. What can CFDSI do to help?

- **Protocols** for provenance, calibration need to be established. (look at Kaggle - for ML).
- CFDSI should **bring together experimental and computational collaborations**. Then this will become a part of our culture.
- How do we accomplish **more complete reporting** of error? **Elevate**

folks who do this. Provide software to make this happen.

- **Data passport** - data should be stamped each time it is processed.
- **Proprietary data processing** is tough to unravel. Can CFDSI help with a greater bargaining power.
- **Reproducibility - incentivize this for experiments and software.**
- **Badging system** for data. **Data Quality Hierarchy**
 - Baseline - data is posted on the website. Some metadata.
 - Next tier - Provide some information about data processing.
 - Higher tier - Good characterization of uncertainty; open source pipeline all the way to “raw” sources
-

4. Misc ideas so they don't get lost (e.g., Did you find new issues? If yes, create the issues on GitHub!):

-

5. Summary for report-back (Alternatively, just bold the key points above):

-