



# 2018-2019 Mid-Term Credibility Plan Review

PI: Danny Bluestein

| # | Ten Simple Rules               | REVIEWER #1                         |   | REVIEWER #2                         |   |
|---|--------------------------------|-------------------------------------|---|-------------------------------------|---|
|   |                                | Considered in the Credibility Plan? | Comments  | Considered in the Credibility Plan? | Comments  |
| 1 | Define context clearly         | sufficient                          | COU is relatively specific; the one word that I'd be interested in understanding more is "reflect".   | sufficient                          | Investigators clearly state model capabilities and potential domains of application. However, specific application descriptions would provide clearer context for the interested user in evaluating the suitability of the mode |
| 2 | Use appropriate data           | sufficient                          | This seems the key aspect of their credibility plan; getting appropriate data for their model.  | sufficient                          | Investigator provides substantial detail on data source and data uncertainty and how it pertains to the model parameters  |
| 3 | Evaluate within context        | insufficient                        | The main emphasis of their credibility plan hinges on appropriate data; but it seems that "evaluate within context" is misunderstood. They emphasized how they run their simulations, not evaluate credibility. | insufficient                        | UQ is well described. Validation, for model parameter tuning is described, but no subsequent validation to an independent referent. There is insufficient discussion of verification activities.                                |
| 4 | List limitations explicitly    | insufficient                        | they imply that the main limitations are with HPC and software capabilities – not wrt to assumptions or decisions. Not sure if their communicated in publications.  | insufficient                        | Limitations and assumptions are not adequately discussed in the context of the models use   |
| 5 | Use version control            | sufficient                          | Versioning data and files.  | insufficient                        | As presented, the activities do not describe sufficient version control processes   |
| 6 | Document adequately            | sufficient                          | Data and models are stored and shared with team members – not sure how this is shared externally.   | insufficient                        | This description actual fulfills more of the version control activity. Documentation activity for such as those communicating model , development, validity, use, and revision are not described                                |
| 7 | Disseminate broadly            | sufficient                          | Considering broad sharing with Google Cloud and present at conferences.   | sufficient                          | Dissemination to community is discussed. Could be strengthened with planning for feedback and evaluation.   |
| 8 | Get independent reviews        | sufficient                          | Using IMAG colleagues for evaluation; not clear how this is conducted.  | sufficient                          | Independent review is planned with general and specific researchers. Extent of the review should be presented in future reporting.  |
| 9 | Test competing implementations | sufficient                          |   | N/A                                 | Investigators imply model is so unique as to not have contemporary implementations within the context of use.   |



# 2018-2019 Mid-Term Credibility Plan Review

|    |                      |            |   |              |   |
|----|----------------------|------------|---|--------------|---|
| 10 | Conform to standards | sufficient | Emphasis was on standards for data collection | insufficient | Investigators appear to confound experimental standards with modeling/programming/data handling standards |
|----|----------------------|------------|---|--------------|---|

## General Comments

### Reviewer 1:

This group seems to mix-up calibration and validation. They stated that “model parameters are compared with experimental results for validation and iteratively adjusted until differences between model and experimental data is minimized. This is calibration not validation.

### Reviewer 2:

Thank you for providing a very interesting and thought provoking credibility plan update on your modeling application. In general the credibility plan update was informative, but lacked clarity on how specific actions support informing the broader ten simple rules categories. The description of the data acquisition is to be applauded, however more detail on model development history, maintenance and documentation needs to accompany the experimental data acquisition. The weakest aspect of the credibility plan was in the definition of the assumptions and limitations of the modeling approach. With the detail that has gone assessing the modeling scheme and associated parameters, there are likely local and global assumptions and limitations that impact model predictive performance. The plan should discuss how these are identified and documented, and possibly addressed during the development phase.