THE PLANETARY SCIENCE INFERENCE ENGINE: APPLICATIONS TO LUNAR SCIENCE.  A. H. Parker¹, R. Ghent², ¹SETI Institute (aparker@seti.org), ²Planetary Science Institute.

Introduction: Conducting rigorous Bayesian inference – including parameter estimation and model selection – in planetary science settings can be a monumental challenge. In many settings, even defining an appropriate likelihood integral may prove difficult or impossible, let alone actually evaluating it. However, many of the science questions we would like to answer at the interface of our observations and the models we build to explain them can only be appropriately tackled with such an inferential step. Our field is not alone in this challenge. The human genomics community pioneered an approach (originally called “Approximate Bayesian Computation,” Pritchard et al. 1999; see Marin et al. 2011 for a review) that has evolved into the modern family of “likelihood-free inference” methods. This family of approaches have since been adapted for application in other fields, including astronomy (Ishida et al. 2015, Hahn et al. 2017, Hsu et al. 2018, Witzel et al. 2018) and planetary science (Parker 2015, Mazrouei et al. 2019, Parker 2021).

To make the adoption of these methods more straightforward for planetary science applications, we have developed the Planetary Science Inference Engine (PSIE; https://github.com/alex-parker/psie). The repository includes worked examples and a standardized interface to core utilities that simplify the implementation of likelihood-free inference workflows in existing planetary science data analysis and model selection problems. At its most basic level, PSIE permits the following:

1) Rigorous estimation of parameter posterior probability functions for proposed models given an observed dataset.
2) Model selection by Bayes factor estimation for any given model in a suite of n models that may best describe the processes that produced a given dataset.

… if the following can be provided:

a) A generative simulation that produces discrete samples of synthetic data given a proposed model and parameter set.

b) An appropriate “distance metric” that describes the similarity of an observed dataset and a discrete sample of synthetic simulated data.

Much of the challenge of applying this method is in appropriate selection of this distance metric. In planetary science, this method has been applied to determining the intrinsic properties of orbit distributions given unknown observational biases (Parker 2015), determining the age distribution of lunar craters (Mazrouei et al. 2019), the parameters describing the size-frequency distribution of impactors that cratered Pluto and Charon and that a model containing binary impactors is preferred over a model only containing solitary impactors (Parker 2021).

Lunar Science Applications: A huge range of lunar science could benefit from the application of likelihood-free inference. The most direct examples are problems relating to the distribution of crater sizes, locations, depths, diameters, and superposition. PSIE provides a coherent framework to understand the odds of rare events through time and the interrelationship of events across different bodies and populations. Further, it provides a clear path to combining observations of very different provenance. The properties make it ideal for tackling complex inference challenges such as those presented by understanding the late heavy bombardment. We will present several worked examples of lunar science applications and which will be archived at https://github.com/alex-parker/psie/tree/main/examples

Future Work: PSIE currently has well-developed distance metrics for planetary science problems relating to n-dimensional distances on manifolds in orbital space, distances on spheres or planes, and distances between arbitrary one-dimensional distributions (and can be invariant under cyclic transformations). Future work plans include development of modules for working on comparisons between models and data in single images (permitting, for example, direct inference on terrain evolution models given imaging datasets). We are also pursuing algorithmic efficiency improvement to permit handling of much larger datasets and a greater diversity of complex models.

Acknowledgements: Contents of this presentation include work supported by the NASA Solar System Exploration Research Virtual Institute nodes Project EXPRESSO and GEODES, under the cooperative agreement numbers 80ARCOM0008 and 80NSSC19M0216 respectively.