A Comprehensive Pipeline for Modeling and Separating Signal and Systematics from Lunar Far Side Radio
Observations. D. Rapetti1,2,3, N. Bassett3, J. Hibbard4, J. O. Burns3, 1Universities Space Research Association-2NASA
Ames Research Center, Moffett Field, California, 3University of Colorado Boulder, Boulder, Colorado.

Introduction: The Dark Ages is a key, unexplored epoch of the Early Universe that was identified as the
discovery area in cosmology by the 2020 Astro Decadal Survey. The highly redshifted, spin-flip hyperfine line
of primordial neutral hydrogen provides a powerful technique to study the Dark Ages and Cosmic Dawn via
low radio frequency observations. The Lunar Surface Electromagnetics Experiment (LuSEE) is to be de-
ployed on the lunar far side on 2025 via NASA’s Commercial Lunar Payload Services (CLPS) program.
LuSEE will provide pioneering low radio frequency ob-
servations from this pristine, radio quiet environment
through the lunar night for a total mission duration of
up to two years.

To extract the coveted sky-averaged (global) 21-cm
signal from these revolutionary measurements is crucial
to have an analysis pipeline capable of properly model-
ing and separating each of the data components. The sky
is several orders of magnitude brighter than the signal
at the relevant frequencies and current information
about the antenna beams is not available at this level of
precision. To address this crucial challenge, we de-
veloped a pattern recognition technique based on training
sets built from realizations encompassing the modeling
uncertainties to capture the critical modes of variation
in the models. Figure 1 shows an example of the power
of this methodology from [1].

Data analysis pipeline: I will discuss the character-
istics and utilization of our hydrogen cosmology pipe-
line that allows for the rigorous and efficient separation
between signal and systematics modeled using infor-
mation from theory, observations, lab measurements
and simulations. This software is publicly available and
employs pattern recognition and Bayesian inference
techniques. It can analyze multiple data sets of interest,
such as those from which to extract the global 21-cm
signal itself or those with which to describe the beam-
weighted, low-radio-frequency sky that can then be em-
ployed to model the foreground component.

The pipeline can be used for ground and space-based
experiments and was particularly constructed towards
instruments aimed at observing from the pristine lunar
far side. Available examples of ongoing simulations for
instrumental design will be presented.

Figure 1: Simulated signal extractions using horizons seen from locations near the center of the Schrödinger
Basin on the lunar farside [1]. Top: Horizon profiles used for the simulated data realization (red) and the
foreground training set curves (gray). In the case on the left, only one horizon is included in the training set,
whereas on the right, six different horizons are included. Middle: Residuals between the horizon profiles used to
create the foreground training set (gray) and the fiducial horizon used to create the data realization (red). Bottom:
68 and 95% confidence intervals of the extracted signal compared to the input (black). While the signal extrac-
tion is inaccurate when the training set horizon is slightly incorrect (as in the case on the left), including
several realizations of the horizon within the training set produces a significant improvement in both the fit and
the signal extraction. The reduce chi-square for the full fit is 221 when the horizon is constant in the training set
and 1.03 when the horizon varies within the training set.

Reference:
[1] Bassett, N., Rapetti, D., Tauscher, K., Nhan, B., Bor-
denave, D., Hibbard, J., Burns, J. “Lost Horizon: Quan-
tifying the Effect of Local Topography on Global 21 cm