Applying Information-Theoretic Approaches for Objective Model Selection and Quantification of a Model Selection Uncertainty

Odele Coddington *[Odele.Coddington@lasp.colorado.edu], Erik Richard, Peter Pilewskie, and Andrew Kren, Laboratory for Atmospheric and Space Physics (LASP), University of Colorado, Boulder.*

Space optics exposed to harsh solar radiation degrade as a function of wavelength and exposure amount. For time-series data, correction of on-orbit degradation must be applied during post-processing to achieve record stability. Correction methods require the assumption of a degradation model and the suitability of this model is typically evaluated by the minimum of the sum square errors between the modeled and the measured data. However, a superior fit can be achieved for reasons other than model fidelity, suggesting that a good fit, while necessary, isn't sufficient to quantify how well a model captures the underlying processes.

Various information-theoretic techniques allow for the objective selection of a model, where the "best" model is the one that makes the observed data *most likely*. For example, An Information Criterion (AIC) [*Akaike*, 1973] applies a metric of model fit called the Kullback-Leibler distance [*Kullback and Leibler*, 1951] to define the information lost when a model is used to approximate the data. Here, we apply AIC to compare several candidate degradation models against the SORCE SIM (v19) data record (without degradation corrections applied). The candidate models have common features where degradation is a function of wavelength and exposure amount, but relax constraints where the degradation rate is assumed constant with SIM channel, and that the Sun has unchanged during the simultaneous measurements (the SIM A/B channels are mirror images; while multi-channel degradation monitoring measurements are taken on the same day, the time lag between measurements at a common wavelength can approach a day). We use AIC to rank and scale the candidate degradation models and demonstrate how multi-model inference can be used to incorporate a model selection uncertainty into the stability estimates of a long-term data record.