First Steps Towards a Homogeneous Solar Spectral Irradiance Dataset

Selection, Merging and Quality Assessment

Micha Schöll, Matthieu Kretzschmar, Thierry Dudok de Wit & Margit Haberreiter
SOLID Project, LPC2E/CNRS Orléans and Université d'Orléans, France & pmodwrc Davos, Switzerland

Data available now!
micha.scholl@cnrs-orleans.fr

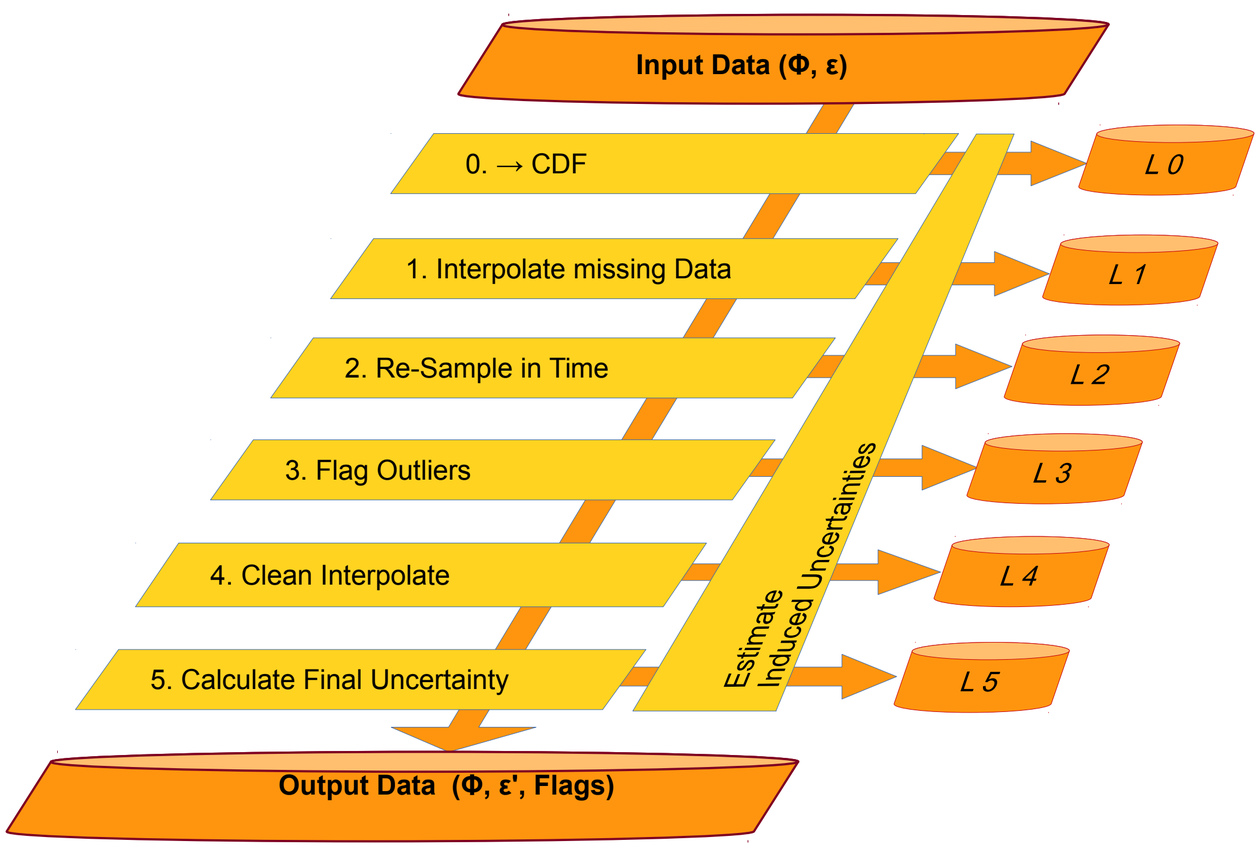
Introduction: SOLID Project

SOLID (First European comprehensive **Sol**ar **I**rradiance **D**ata Exploitation) project is part of the 7th European framework programme. It aims to bring together a large variety of different solar spectral irradiance datasets, to combine them into one dataset and to reconstruct the spectral solar variability further back in time. The overall goal is to deliver one single homogeneous dataset that can be used in astrophysics and climate studies alike to account for non-constant solar forcing.

Overview

We use a pipeline approach and process all instrument data with the same procedures. At each step the induced uncertainty is estimated and stored and all changes to data are flagged. Currently, we have 6 levels, starting with the converted raw data (level 0), with its gaps filled (level 1) and re-gridded in time (level 2). Possible outliers are removed and flagged as such with physical outliers accounted for (level 3). Missing data is interpolated via models (level 4).
Finally, we translate all adoptions into uncertainties, resulting into original data ready for further processing (level 5).

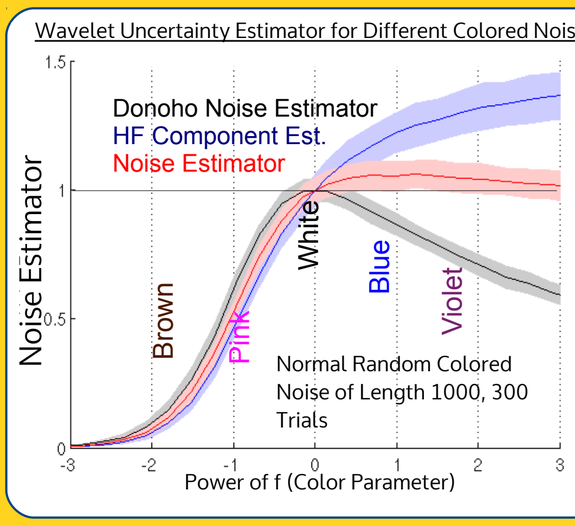
Pipeline Approach



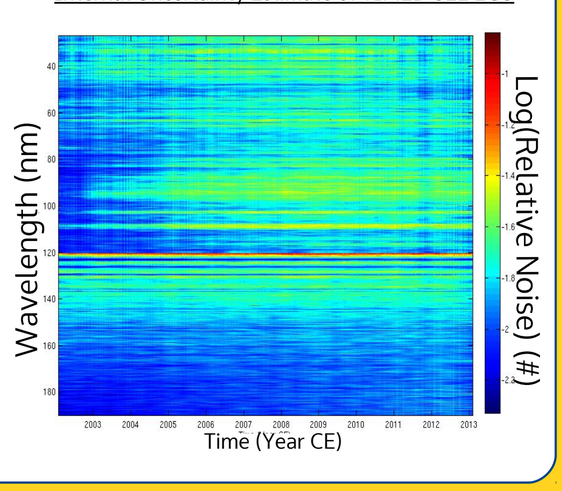
All Levels: Internal Uncertainty Estimation

For e.g. re-gridding we adapt the Donoho Wavelet estimator to accurately assess HF colored noise, i.e. blue and violet noise, by combining a weighted averaging of the original Donoho estimator together with an high frequency component noise estimator.

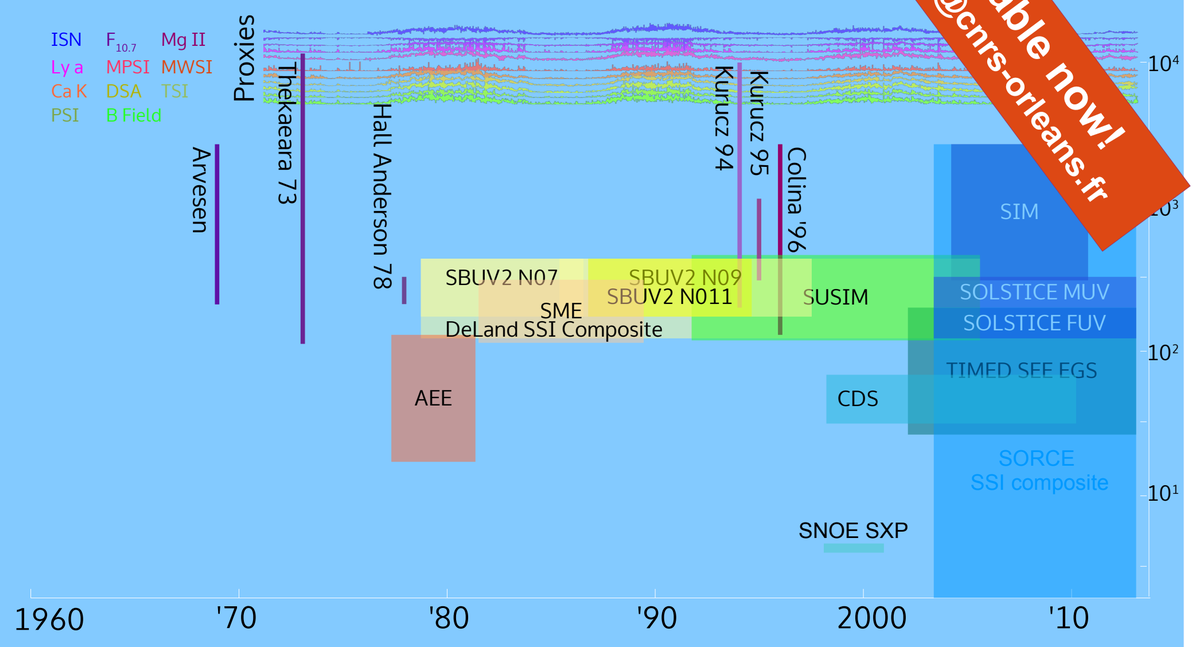
Wavelet Uncertainty Estimator for Different Colored Noise



Internal Uncertainty Estimate of TIMED SEE EGS



L. 0) Instrument Data



Data are collected from different instruments, solar proxies and reference spectra. All data with accompanying uncertainty, error estimates and meta data are stored in NCDF CF-1.6 (Climate Forecasting) files. These data will be available as level 0 data. Currently, the database consists of 18 instruments, 11 reference spectra and 13 proxies, i.e. 42 data sources.

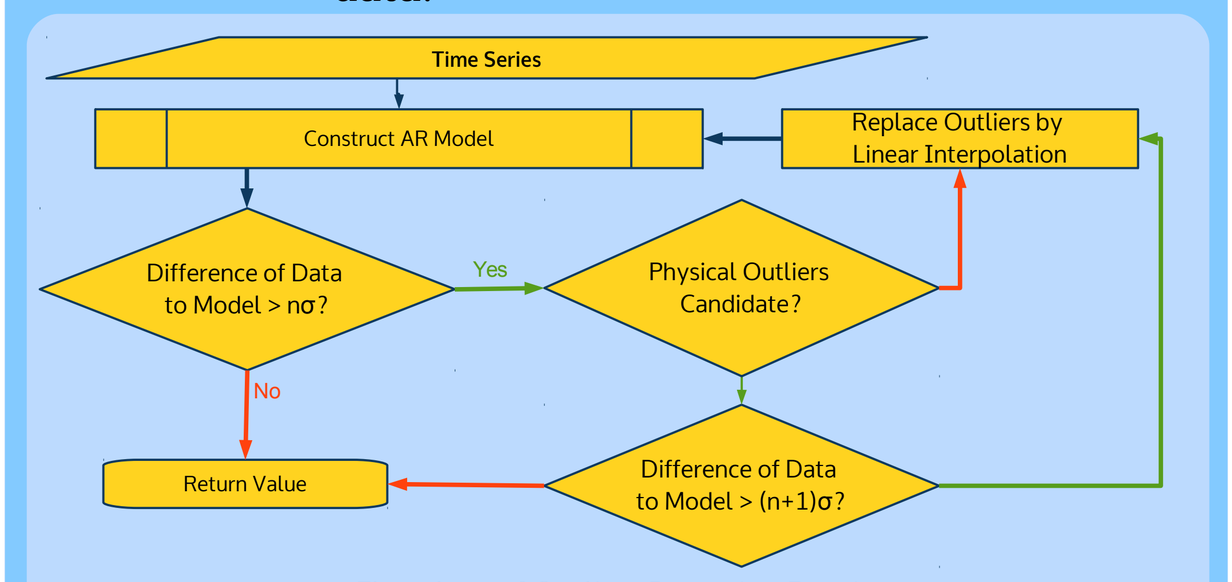
L. 1-2) Gap Filling and Re-Gridding

The instrumental data gaps range from < 1 % to ~13%. Some further processing requires the data on a uniform grid without missing data. Missing data are flagged and interpolated by an iterative SVD interpolation up to 10th order, starting with 1st order interpolation, increasing the order once convergence is reached, resulting in a homogeneous spectrum over both wavelength and time, available as level 1 and 2 data.

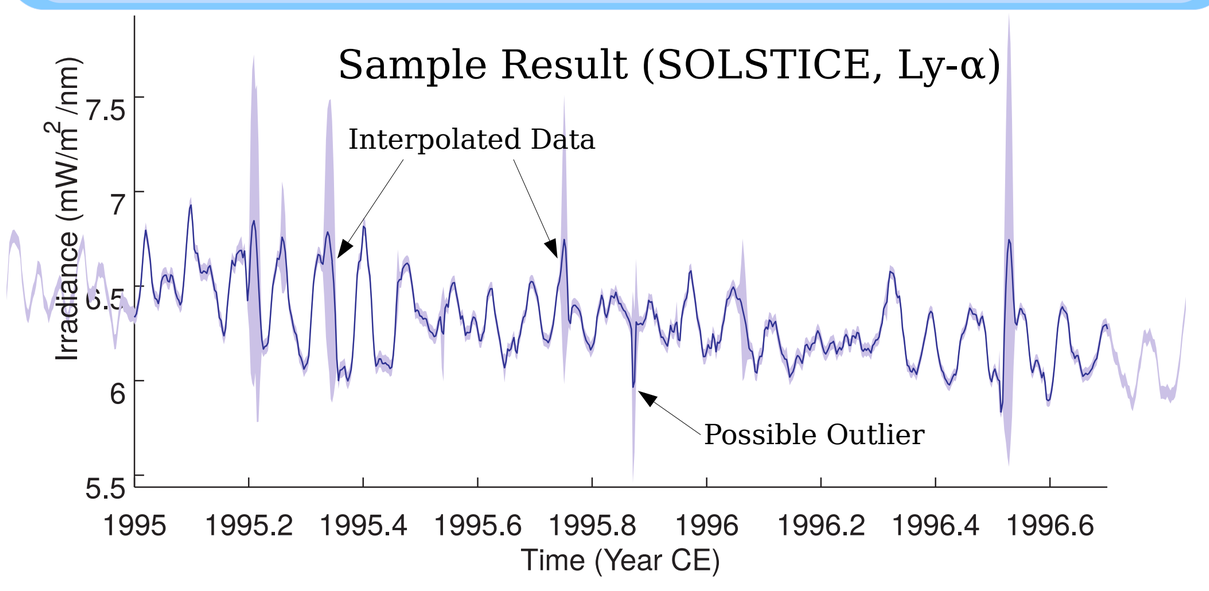
L. 3) Outliers

Outliers are possible non-physical data. We apply an iterative autoregressive distance model to detect possible outliers. The flagging of corresponding values and their replacement by interpolation results in level 3 data.

Physical Outliers
The same outliers detection is used on proxy data. For detected outliers in the proxy data, the outliers criterion is relaxed.



Flowchart of Outliers Detection Routine



Acknowledgement This project is founded by European union FP7-SPACE program # 313188