

Python Language Based Data Analysis Toolkit

T.H. Osborne

A set of experimental data analysis tools have been created using Python as the scripting language. Python offers the advantages of: 1) open source with strong support in the open source community, 2) simple structured syntax, 3) simple object oriented structure, 4) very large set of existing extensions not limited to data analysis, 5) easy extensibility with well document C API. The tools are strongly integrated to MDSplus data archives of raw and processed experimental results. The basic building block for general purpose and custom tools is a class representing signal type data with many processing methods defined, e.g. FFT, smoothing, splines. Objects of this class (Data) are typically instantiated by reading an MDSplus signal node. Instances of the Data class can be combined algebraically with automatic interpolation and masking of undefined operations (e.g. zero divide at some time point). Python interfaces to pgplot, BLT, and Gnuplot are used for visulization. Python interfaces to Tk and GTK+ (Gnome) widget sets are installed and used and an interface to Qt (KDE) widgets also exists. Higher level routines based on Data class objects have been written for general data display, profile fitting, signal processing, and drivers for equilibrium fitting (EFIT), transport (ONETWO), and stability (BALOO) codes. The tool kit had web based documentation and is available for LINUX in RPM form in a YUM archive as well as source packages and is managed through CVS. The Python package set has been installed at PPPL as well as GA.