**Principal component analysis module of combined EBSD/EDS datasets – AstroEBSD v2**

10.1016/j.ultramic.2020.112944

Tom McAuliffe 21/02/20 – [t.mcauliffe17@imperial.ac.uk](mailto:t.mcauliffe17@imperial.ac.uk) or [tmcauliffe@icloud.com](mailto:tmcauliffe@icloud.com)

We have provided an example deck for combined analysis of EBSD and EDS datasets.

Short version:

* Data should be in ‘.h5’ format, exported from a Bruker BCF file (with patterns and spectra) using the BCF2HDF5 converter.
* You can cluster on EBSD information, EDS or both.
* Fill out the ‘Directories’ and ‘Critical’ settings specific to your problem.
* RTM module needs ‘.bin’ files in *\phases\masterpatterns* and AstroEBSD needs associated ‘.pha’ and ‘.cif’ files in *\phases\phasefiles* and *\phases\cifs* respectively. More information on this is provided in the other tutorial documents. After generating a ‘.bin’ file with a ‘.cif’ file and DynamicS, you can copy and paste a ‘.pha’ in the phases folder, changing the name and bin file location and the ‘ishex’ toggle.
* PostAnalysis and PatternComparison decks are provided that are fairly self-explanatory.

Settings guide

* All user settings are in the deck scripts.

**Directories**: - fairly self explanatory.

* Set up **InputUser.HDF5\_folder** as where your H5 files are stored (make sure it’s a string).
* Files should be a cell array of strings of files in that folder to analyse, eg. **files**={‘sample1.h5’,’sample2.h5’}; The code will run for all of these consecutively, creating a results folder in in **InputUser.SavingFolder**.
* **InputUser.SavingFolder** should be a string of this output location.
* **InputUser.Phases** should be a cell string array of the candidate phases (by name of the .bin file), eg. {‘Ni’, ‘M6C’, ‘M23C6’…};
* Specify where your mtex and AstroEBSD directories are.

**Critical settings**: - what analysis are you going to do?

* EBSD, EDS, and the associated weighting.
* **PCA\_Setup.EBSD\_weighting**: The extent to which you want to prioritise the EBSD data in the PC calculation. 1 is standard, and is a quite heavy EBSD weighting by nature of having far more EBSD measurements than EDS. If you want EDS weighting, put it to 0.1 or less. …
* **PCA\_Setup**.**RTM**: Do you want to do the RTM analysis on the rotated PCs?
* **PCA\_Setup.SpatialKernel**: Employ a spatial kernel to introduce some localisation to the analysis, after Guo *et al* (doi.org/10.1080/10618600.2014.912135).
* **PCA\_Setup.KernelRadius**: What should the kernel square dimension be?
* **Refine.run**: Run a pattern centre refinement step using RTM on selected patterns? This can improve RTM peak heights if they were indexed poorly by Bruker Esprit etc.

**Medium level** **settings**: -

* **Printing**: display and save figures…
* **PCA\_Setup.crop\_factor**: What should the tiling of the total area of interest be? Eg. 3 gives 3-by-3 tiling. This has significant implications on RAM usage.
* **PCA\_Setup.variance\_tolerance**: in %, the ‘best’ way of choosing number of copmonents to retain. 0.1 is standard, if you want better spatial resolution at the cost of possible oversampling this can be reduced. If you want to group more stuff together (ie. think you’re already oversampling), raise it to 0.2 or 0.3. If you want to force a specific number of components (applied to every tile) set this in **PCA\_Setup.components**. … Read the paper for help with choosing this.
* RTM settings as discussed in the RTM paper. **RTM\_setup.Screensize** is (square) resolution of the EBSPs for cross-correlation, reduce this to save memory (esp. wrt to the template library).

**Low level** **settings**: - these only have to be adjusted fairly infrequently. Of particular note are…

* **RTM\_setup.Sampling\_Freq** is the SO3 angular spacing of template EBSPs. Increase this to save memory and processing time, but 7 or 8˚ is standard.
* AstroEBSD background corrections as required.

Action of the code (series loading):

For each file in the files cell of strings…

* Divide the map into an x\*x tiling grid.
* Load EBSPs and spectra for each tile.
* Calculate PCs for each tile, and get loadings of each scan point for each PC. Number of PCs calculated from variance tolerance. Varimax rotate these to get RCCs and RC loadings of each scan point. Nb. There is a ‘weightloop’ functionality in there, where you can cycle different weightings and save results for the same area of interest.
* Reassemble results from all tiles in the right shape / place / etc
* Generate a set of RC-EBSPs as candidates for template matching and indexing
* Perform the RTM
* Save results
* Generate ‘.spx’ Bruker spectra files for each dataset (both label average and RC-spectra).

After this is all done, to do the post-**processing you need to load and quantify the ‘.spx’ output files using Bruker Esprit**. Save these as a results table in ‘.xls’ format, which is read by the post-analysis deck. An example of this format is provided in */tutorial\_docs/Example\_EDS\_Quant*.

Post processing

* The PostAnalysis deck provides a lot of functionality for plotting phase maps, IPF maps, and phase-based chemistry evaluation. What you want to analyse should be described by the **post** structure. NB: *post.comb* provides element combinations for EDS maps (eg RGB for C, Ni, Ta respectively). If you put numbers > number of elements in your results.xls there will be errors.
* The PatternComparison script is fairly self-explanatory. It compares the RC-EBSP for a given component number to the corresponding simulations for visual evaluation.

Generating template libraries with Bruker DynamicS

* Need a ‘.pha’, ‘.cif’, and ‘.bin’ file for each phase in the correct locations.
* .bin files can be generated using DynamicS.
* Navigate to (currently) F drive, where there is a folder called ‘Esprit DynamicS CIF’.
* In ‘bin’ alter the DynamicS\_CIF.txt config file with the settings you need (nb. resolution choice 3 is 1024). ‘CIF filename’ and ‘XML filename’ should point to the CIF file you want to generate a library from, eg:
  + CIF\_FILENAME='C:\Users\tpm416\Documents\Borides\_050719\Borides\_050719\W3CoB3\_Cmcm.cif'
  + XML\_FILENAME='C:\Users\tpm416\Documents\Borides\_050719\Borides\_050719\W3CoB3\_Cmcm.xml'
* Open command prompt, then navigate to the F drive:
  + ‘F:’ then enter
  + ‘cd Esprit DynamicS’ CIF
  + ‘cd bin’ enter
  + ‘DynamicS\_CIF’ enter
* There will be a load of OpenGL warnings you need to keep clicking through.
* It will start running the simulation. I have experienced a bug where it will not choose the correct resolution unless you cancel the auto-run, then re-click ‘simulation’ to restart it.
* Results will be saved in the directory you specified.