Phenylketonuria landscape analysis v 1.9.0

Pipeline:

1. Apply Gaussian model and extract features
2. Perform QC (rmse - var)
3. Detect non-responders
4. Subpopulation detection

Run pipeline

Install packages

```
 numpy
pandas
scipy
matplotlib
openpyxl
tqdm
```

To run the pipeline use the script `gm.py` , which requres experimant files as `*.xlsm` .

Example 1) run the pipeline for one experimant

```
python gm.py --input_dir Data/experiments_all/20240305_EXP13.xlsm
```

Example 2) run the pipeline for all experimants in a directory (*e.g.* `Data/experiments_all`)

```
python gm.py --input_dir Data/experiments_all
```

Promp terminal parameters:

- `--input_dir` : input directiry. Either a `*.xlsm` file or a folder containing them.
- `--param_dir` : directory to parameters json file (see below).
- `no_plot` : no plot will be created if apecified.

A parameter *.json* file is also needed to specify the parameters. By default the json file should be in the same folder as the `gm.py` named `gm_parameters.json` . You can change this by the parameter `--param_dir` .

Parameters in the json file:

- `save_image2d_dir` : Directory to save 2D landscapes; set it as `""` to prevent saving.
- `save_image3d_dir` : Directory to save 3D plots; set it as `""` to prevent saving.
- `save_feature_dir` : Directory to save feature csv file; set it as `""` to prevent saving.
- `tag` : Name tag for the feature *.csv* file.
- `qc_thr_rmse` : QC threshold for RMSE; `default=[0.2, 0.25]`
- `qc_thr_n_peaks` : QC threshold for N Peacks; `default=[5, 8]`
- `qc_thr_variation` : QC threshold for Variations; `default=[0.1, 0.25]`
- `elev` : Elevation of the camera in the 3D plots; `default=30` ,
- `azim` : Angle of the camera in the 3D plots (in degrees); `default=-120` ,
- `nbins` : Number of bins used for the interpolation (mesh grid of size nbins x nbins); `default=1000` ,
- `sm_method` : Method used for the interpolation and smoothing ("regular_grid" or "linear_ndi"); `default='regular_grid'` ,
- `rescale` : Binary value to rescale the data with respect to the *WT*; `default=True` ,
- `max_val_scale` : Maximum value of the non-rescaled data; `default=10000` ,
- `info_box` : Binary value to add an information box to the plots; `default=True` ,
- `max_val` : Binary value to add the maximum value of the data to the information box; `default=True` ,
- `peak_coords` : Binary value to add the coordinates of the peaks to the information box; `default=True` ,
- `fifty_coords` : Binary value to add the coordinates of the 50% of the maximum value to the information ; `default=True` ,
- `plot_replicates` : Binary value to plot and save data for all the replicates; `default=True`