

Phenylketonuria landscape analysis v 1.5

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 requires experiment files as `*.xlsm`.

Example 1) run the pipeline for one experiment

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

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

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

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`: irectory to save 2D landscapes; set it as `""` to prevent saving.
- `save_image3d_dir`: irectory 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 Peaks; default=[5, 8]
- `qc_thr_variation`: QC threshold for Variations; default=[0.1, 0.25]