

NMF-PY Workflow

The steps in this notebook are intended to replicate the preprocessing, base model building, and base model post-processing steps of PMF5.

The error estimation functionality has not yet been implemented in the new code base.

```
In [3]: # Notebook imports
import os
import sys
import json

module_path = os.path.abspath(os.path.join('..'))
if module_path not in sys.path:
    sys.path.append(module_path)
```

Sample Dataset

The three sample datasets from PMF5 are available for use, but a new dataset can be used in their place.

```
In [2]: # Baton Rouge Dataset
br_input_file = os.path.join("data", "Dataset-BatonRouge-con.csv")
br_uncertainty_file = os.path.join("data", "Dataset-BatonRouge-unc.csv")
br_output_path = os.path.join("data", "output", "BatonRouge")
# Baltimore Dataset
b_input_file = os.path.join("data", "Dataset-Baltimore_con.txt")
b_uncertainty_file = os.path.join("data", "Dataset-Baltimore_unc.txt")
b_output_path = os.path.join("data", "output", "Baltimore")
# Saint Louis Dataset
sl_input_file = os.path.join("data", "Dataset-StLouis-con.csv")
sl_uncertainty_file = os.path.join("data", "Dataset-StLouis-unc.csv")
sl_output_path = os.path.join("data", "output", "StLouis")
```

Code Imports

```
In [3]: from src.data.datahandler import DataHandler
from src.model.nmf import NMF
from src.model.batch_nmf import BatchNMF
from src.data.analysis import ModelAnalysis
```

Input Parameters

```
In [4]: index_col = "Date" # the index of the input/uncertainty datasets
factors = 6 # the number of factors
method = "ls-nmf" # "Ls-nmf", "ws-nmf"
models = 20 # the number of models to train
init_method = "col_means" # default is column means "col_means", "kmeans", "cmeans"
init_norm = True # if init_method=kmeans or cmeans, normalize the data prior to clustering.
seed = 42 # random seed for initialization
max_iterations = 20000 # the maximum number of iterations for fitting a model
converge_delta = 0.1 # convergence criteria for the change in loss, Q
converge_n = 10 # convergence criteria for the number of steps where the loss changes by less than converge_delta
verbose = True # adds more verbosity to the algorithm workflow on execution.
optimized = True # use the Rust code if possible
parallel = True # execute the model training in parallel, multiple models at the same time
```

Dataset Selection

One of the three sample datasets can be selected or a new cleaned dataset can be used. Datasets should be cleaned, containing no missing data (either dropping missing/NaNs, or interpolating the missing values).

```
In [5]: # Loading the Baton Rouge dataset  
input_file = br_input_file  
uncertainty_file = br_uncertainty_file  
output_path = br_output_path
```

Load Data

Assign the processed data and uncertainty datasets to the variables V and U. These steps will be simplified/streamlined in a future version of the code.

```
In [6]: data_handler = DataHandler(  
    input_path=input_file,  
    uncertainty_path=uncertainty_file,  
    index_col=index_col  
)  
V, U = data_handler.get_data()
```

18-Jan-24 12:24:15 - Input and output configured successfully

Input/Uncertainty Data Metrics and Visualizations

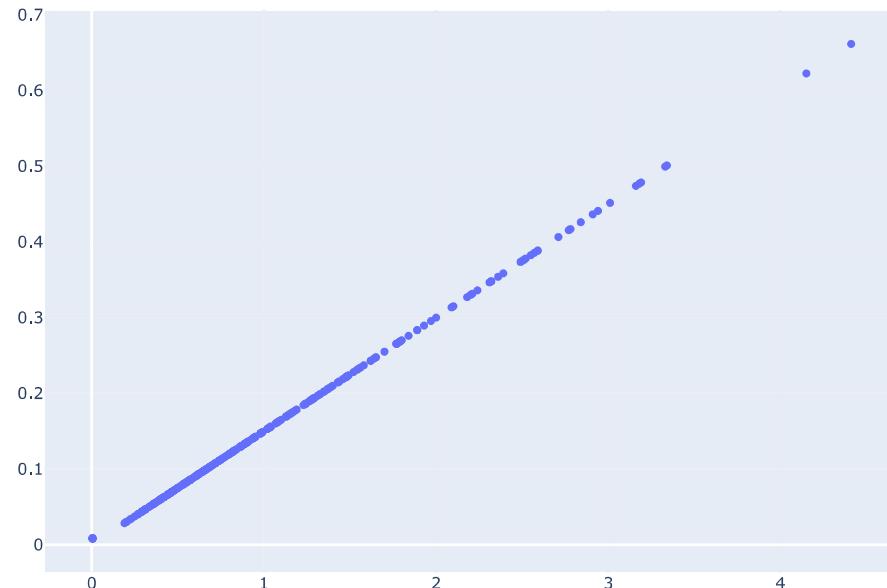
```
In [7]: # Show the input data metrics, including signal to noise ratio of the data and uncertainty  
data_handler.metrics
```

Out[7]:		Category	S/N	Min	25th	50th	75th	Max
	124-Trimethylbenzene	Strong	5.445168	0.005000	0.820001	1.290001	1.865001	5.470003
	224-Trimethylpentane	Strong	5.666667	0.410000	1.580001	2.490002	3.865002	13.560008
	234-Trimethylpentane	Strong	5.537459	0.005000	0.530000	0.820001	1.300001	4.410003
	23-Dimethylbutane	Strong	5.500543	0.005000	0.640000	1.110001	2.285001	10.500007
	23-Dimethylpentane	Strong	5.463626	0.005000	0.340000	0.490000	0.780000	3.310002
	2-Methylheptane	Strong	5.039088	0.005000	0.215000	0.330000	0.535000	2.480002
	3-Methylhexane	Strong	5.648208	0.005000	0.655000	1.050001	1.510001	7.780005
	3-Methylpentane	Strong	5.611292	0.540000	1.720001	2.990002	5.945004	29.100018
	Acetylene	Strong	5.666667	0.380000	1.410001	1.990001	2.835002	8.070005
	Benzene	Strong	5.666667	0.590000	1.960001	2.770002	4.440003	9.330006
	Cis-2-Butene	Strong	3.100977	0.005000	0.005000	0.120000	0.310000	21.510013
	Cis-2-Pentene	Strong	5.057546	0.005000	0.220000	0.370000	0.720000	19.130012
	Ethane	Strong	5.666667	2.570002	8.285005	12.070008	17.890011	49.970031
	Ethylbenzene	Strong	5.666667	0.200000	0.620000	0.940001	1.450001	3.760002
	Ethylene	Strong	5.666667	0.890001	3.055002	5.260003	10.370006	49.490031
	Isobutane	Strong	5.666667	0.690000	2.705002	4.680003	7.700005	36.330023
	Isopentane	Strong	5.666667	2.500002	6.530004	10.260006	16.080010	125.500079
	Isoprene	Strong	5.555917	0.005000	0.890001	2.520002	4.840003	9.420006
	Isopropylbenzene	Strong	2.067318	0.005000	0.005000	0.005000	0.220000	6.640004
	M_P Xylene	Strong	5.666667	0.470000	1.845001	2.950002	4.510003	13.050008
	M-Diethylbenzene	Strong	2.436482	0.005000	0.005000	0.005000	0.230000	1.740001
	M-Ethyltoluene	Strong	5.519001	0.005000	0.530000	0.890001	1.395001	4.670003
	N-Butane	Strong	5.666667	1.650001	4.790003	7.360005	11.445007	56.230035
	N-Decane	Strong	5.168295	0.005000	0.280000	0.430000	0.690000	1.760001
	N-Heptane	Strong	5.666667	0.210000	0.730000	1.060001	1.515001	5.350003
	N-Hexane	Strong	5.611292	0.660000	2.170001	4.530003	9.820006	92.560058
	N-Nonane	Strong	5.408252	0.005000	0.310000	0.460000	0.670000	1.950001
	N-Octane	Strong	5.574376	0.005000	0.405000	0.630000	0.890001	3.100002
	N-Pentane	Strong	5.666667	0.710000	3.770002	5.450003	9.270006	38.580024
	N-Propylbenzene	Strong	3.617807	0.005000	0.005000	0.220000	0.330000	0.940001
	N-Undecane	Strong	4.983713	0.005000	0.250000	0.370000	0.550000	2.860002
	O-Ethyltoluene	Strong	4.946797	0.005000	0.240000	0.390000	0.585000	1.620001
	O-Xylene	Strong	5.666667	0.220000	0.800001	1.160001	1.735001	5.290003
	Propane	Strong	5.666667	2.260001	8.250005	12.270008	18.440012	70.960044

	Category	S/N	Min	25th	50th	75th	Max
Propylene	Strong	5.666667	0.380000	1.285001	2.450002	4.705003	24.900016
Styrene	Strong	4.891422	0.005000	0.270000	0.500000	0.885001	7.650005
Toluene	Strong	5.666667	1.310001	3.890002	5.950004	8.690005	26.870017
Trans-2-Butene	Strong	2.971770	0.005000	0.005000	0.090000	0.300000	3.210002
Trans-2-Pentene	Strong	5.408252	0.005000	0.405000	0.700000	1.300001	17.280011
Unidentified	Strong	1.000000	1.020135	28.923549	50.394956	76.587424	385.184875
TNMOC	Strong	0.000000	44.420000	113.616462	158.820000	249.795150	708.467700

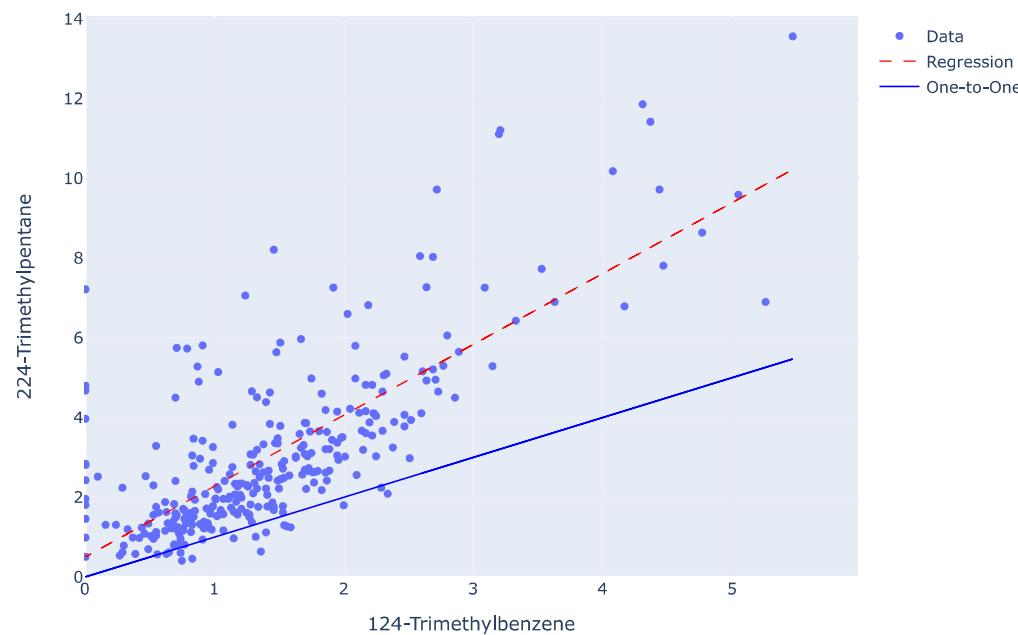
```
In [8]: # Concentration / Uncertainty Scatter plot for specific feature, feature/column specified by index
data_handler.data_uncertainty_plot(feature_idx=2)
```

Concentration/Uncertainty Scatter Plot - 234-Trimethylpentane



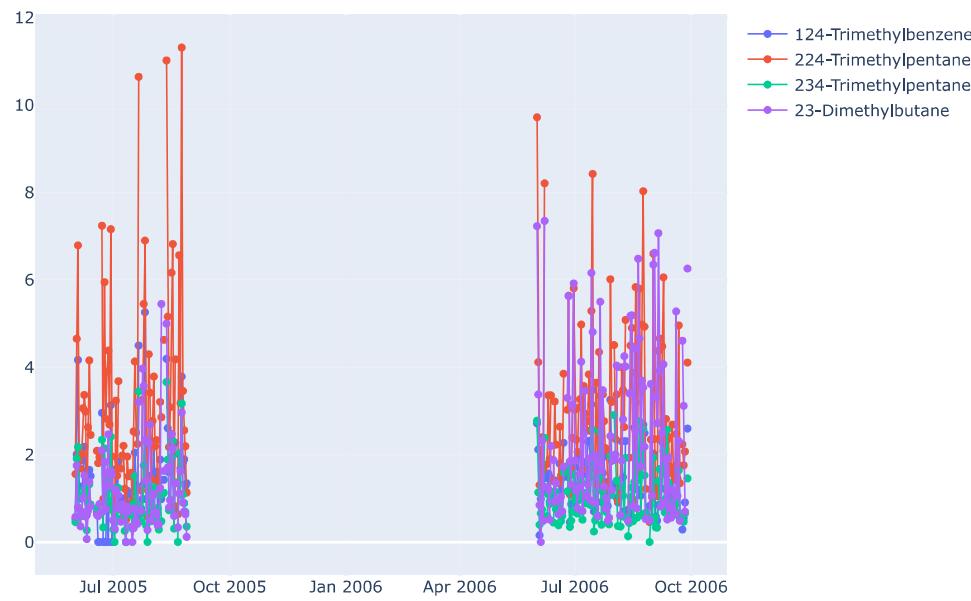
```
In [9]: # Species Concentration plot comparing features, features/columns specified by index
data_handler.feature_data_plot(x_idx=0, y_idx=1)
```

Feature vs Feature Plot: 224-Trimethylpentane/124-Trimethylbenzene



```
In [10]: # Species Timeseries, a single or list of features/columns specified by index  
data_handler.feature_timeseries_plot(feature_selection=[0, 1, 2, 3])
```

Concentration Timeseries



```
In [11]: %time
# Training multiple models, optional parameters are commented out.
nmf_models = BatchNMF(V=V, U=U, factors=factors, models=models, method=method, seed=seed, max_iter=max_iterations,
                      init_method=init_method, init_norm=init_norm,
                      converge_delta=converge_delta, converge_n=converge_n,
                      parallel=parallel, optimized=optimized,
                      verbose=verbose
                     )
_= nmf_models.train()
```

```

18-Jan-24 12:24:24 - Model: 1, Q(true): 65069.5744, Q(robust): 55644.9141, Seed: 8925, Converged: True, Steps: 2300/20000
18-Jan-24 12:24:24 - Model: 2, Q(true): 65541.8666, Q(robust): 55585.2486, Seed: 77395, Converged: True, Steps: 1291/20000
18-Jan-24 12:24:24 - Model: 3, Q(true): 65049.2052, Q(robust): 55589.223, Seed: 65457, Converged: True, Steps: 1348/20000
18-Jan-24 12:24:24 - Model: 4, Q(true): 66051.8651, Q(robust): 56544.1801, Seed: 43887, Converged: True, Steps: 1134/20000
18-Jan-24 12:24:24 - Model: 5, Q(true): 63928.0347, Q(robust): 54497.1779, Seed: 43301, Converged: True, Steps: 1495/20000
18-Jan-24 12:24:24 - Model: 6, Q(true): 63840.4524, Q(robust): 54423.8403, Seed: 85859, Converged: True, Steps: 2343/20000
18-Jan-24 12:24:24 - Model: 7, Q(true): 66533.3494, Q(robust): 56896.0526, Seed: 8594, Converged: True, Steps: 1011/20000
18-Jan-24 12:24:24 - Model: 8, Q(true): 65055.9772, Q(robust): 55644.0452, Seed: 69736, Converged: True, Steps: 2167/20000
18-Jan-24 12:24:24 - Model: 9, Q(true): 63894.2365, Q(robust): 54459.2191, Seed: 20146, Converged: True, Steps: 2286/20000
18-Jan-24 12:24:24 - Model: 10, Q(true): 66038.0929, Q(robust): 56519.5267, Seed: 9417, Converged: True, Steps: 1677/20000
18-Jan-24 12:24:24 - Model: 11, Q(true): 66365.8347, Q(robust): 56776.6322, Seed: 52647, Converged: True, Steps: 1088/20000
18-Jan-24 12:24:25 - Model: 12, Q(true): 66462.6334, Q(robust): 56901.4507, Seed: 97562, Converged: True, Steps: 1069/20000
18-Jan-24 12:24:25 - Model: 13, Q(true): 66053.1798, Q(robust): 56880.8998, Seed: 73575, Converged: True, Steps: 2570/20000
18-Jan-24 12:24:25 - Model: 14, Q(true): 68234.1053, Q(robust): 57304.2993, Seed: 76113, Converged: True, Steps: 1182/20000
18-Jan-24 12:24:25 - Model: 15, Q(true): 63913.2814, Q(robust): 54463.6814, Seed: 71747, Converged: True, Steps: 1513/20000
18-Jan-24 12:24:25 - Model: 16, Q(true): 68121.432, Q(robust): 58292.0844, Seed: 78606, Converged: True, Steps: 997/20000
18-Jan-24 12:24:25 - Model: 17, Q(true): 64827.6197, Q(robust): 55184.941, Seed: 51322, Converged: True, Steps: 921/20000
18-Jan-24 12:24:25 - Model: 18, Q(true): 65340.5877, Q(robust): 55586.3584, Seed: 12811, Converged: True, Steps: 1972/20000
18-Jan-24 12:24:25 - Model: 19, Q(true): 66006.3531, Q(robust): 56131.7228, Seed: 83974, Converged: True, Steps: 2296/20000
18-Jan-24 12:24:25 - Model: 20, Q(true): 65945.8613, Q(robust): 56316.9819, Seed: 45038, Converged: True, Steps: 2113/20000
18-Jan-24 12:24:25 - Results - Best Model: 6, Q(true): 63840.4524, Q(robust): 54423.8403, Converged: True
18-Jan-24 12:24:25 - Runtime: 0.14 min(s)
CPU times: total: 422 ms
Wall time: 8.37 s

```

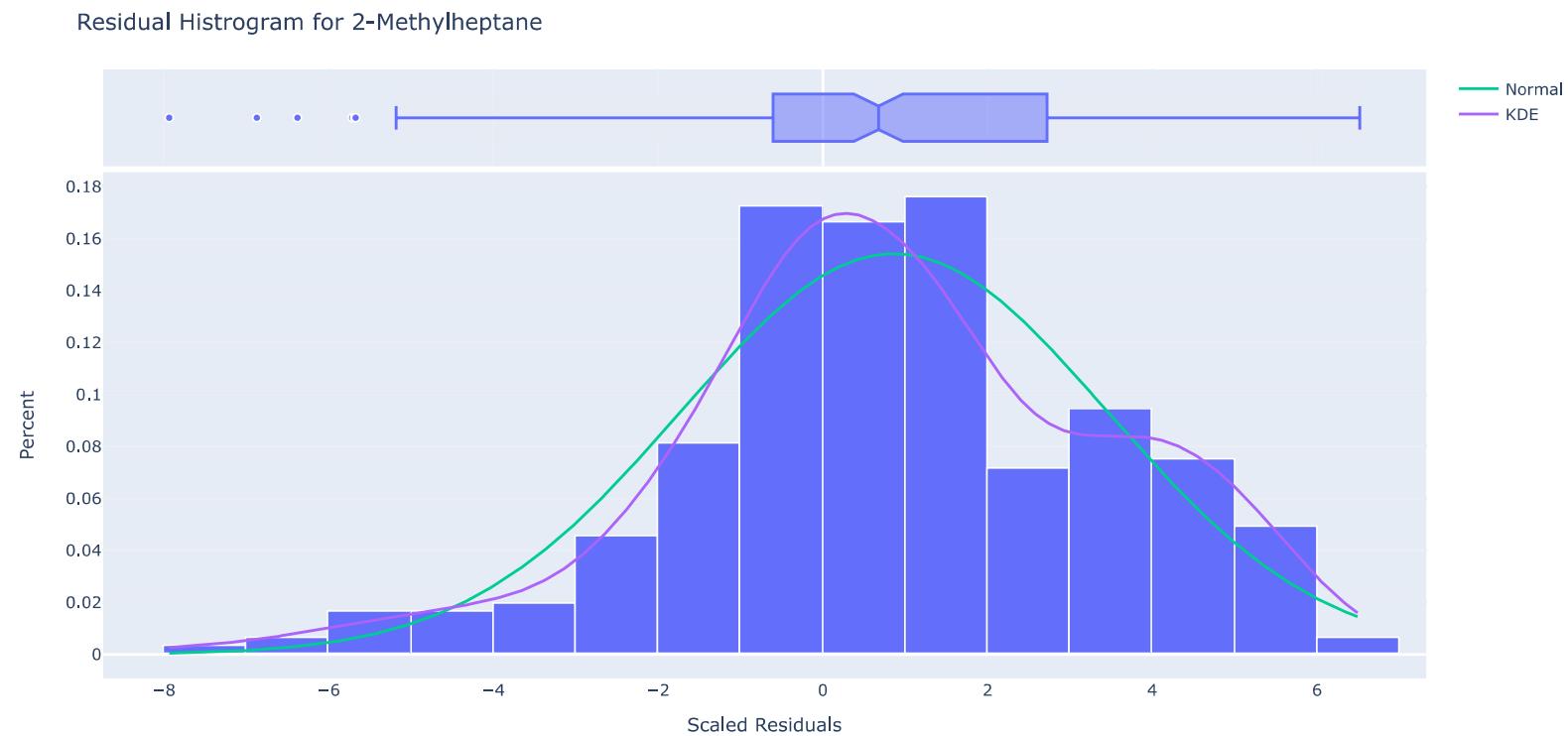
Train Model

```
In [12]: # Select the best performing model to review
best_model = nmf_models.best_model
nmf_model = nmf_models.results[best_model]
best_model
```

Out[12]: 5

```
In [13]: # Initialize the Model Analysis module
model_analysis = ModelAnalysis(datahandler=data_handler, model=nmf_model, selected_model=best_model)
```

```
In [14]: # Residual Analysis shows the scaled residual histogram, along with metrics and distribution curves. The abs_threshold parameter specifies the condition for the returned values of the
abs_threshold = 3.0
threshold_residuals = model_analysis.plot_residual_histogram(feature_idx=5, abs_threshold=abs_threshold)
```



```
In [15]: print(f"List of Absolute Scaled Residual Greather than: {abs_threshold}. Count: {threshold_residuals.shape[0]}")  
threshold_residuals
```

List of Absolute Scaled Residual Greather than: 3.0. Count: 88

Out[15]:	2-Methylheptane	datetime
0	5.127830	6/1/2005 6:00
5	3.498680	6/4/2005 6:00
9	4.784237	6/6/2005 6:00
19	4.723600	6/12/2005 6:00
24	6.348824	6/19/2005 3:00
...
294	4.395149	9/18/2006 6:00
295	5.399104	9/19/2006 3:00
296	4.342209	9/19/2006 6:00
302	-4.799767	9/24/2006 6:00
305	4.124265	9/26/2006 6:00

88 rows × 2 columns

```
In [16]: # The model output statistics for the estimated V, including SE: Standard Error metrics, and 3 normal distribution tests of the residuals (KS Normal is used in PMF5)
model_analysis.calculate_statistics()
model_analysis.statistics
```

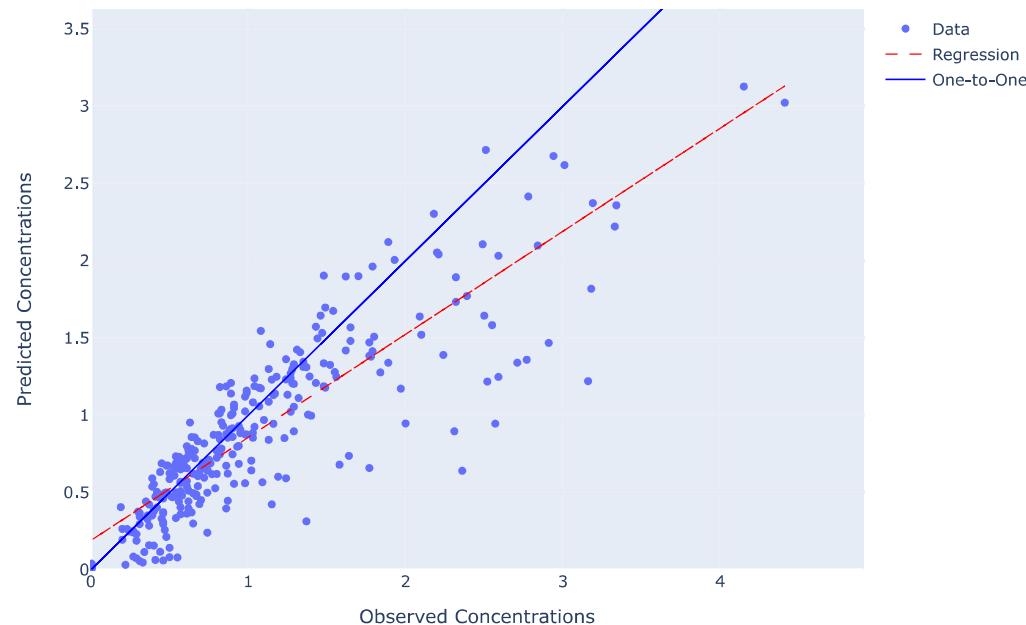
Out[16]:

	Features	Category	r2	Intercept	Intercept SE	Slope	Slope SE	SE	SE Regression	Anderson Normal Residual	Anderson Statistic	Shapiro Normal Residuals	Shapiro PValue	KS Normal Residuals	KS PValue	KS Statistic
0	124-Trimethylbenzene	Strong	0.588973	0.197586	0.054143	0.657848	0.031468	0.034396	0.511612	No	2.847252	No	5.039691e-06	Yes	1.540227e-01	0.064066
1	224-Trimethylpentane	Strong	0.827492	0.695924	0.064181	0.651170	0.017024	0.057165	0.649723	15.0	0.214079	Yes	7.244727e-01	Yes	9.710335e-01	0.027341
2	234-Trimethylpentane	Strong	0.752289	0.189695	0.027837	0.666840	0.021910	0.021412	0.282952	No	1.469420	No	5.827030e-04	Yes	2.716517e-01	0.056446
3	23-Dimethylbutane	Strong	0.552614	0.207565	0.083120	0.639939	0.032970	0.067606	1.004347	No	1.306493	No	9.724029e-04	Yes	3.156685e-01	0.054238
4	23-Dimethylpentane	Strong	0.784656	0.099089	0.016666	0.739934	0.022196	0.011492	0.167215	No	3.303195	No	1.367640e-08	No	3.784841e-02	0.079814
5	2-Methylheptane	Strong	0.573763	0.056348	0.016459	0.674406	0.033284	0.011334	0.173263	No	1.438563	No	6.605691e-04	Yes	3.804811e-01	0.051341
6	3-Methylhexane	Strong	0.707751	0.183255	0.036302	0.676358	0.024886	0.025434	0.357434	No	4.307489	No	2.793057e-10	No	2.475226e-03	0.103789
7	3-Methylpentane	Strong	0.541216	1.198529	0.161995	0.454739	0.023973	0.183461	1.957702	No	1.571499	No	1.082421e-03	Yes	2.083368e-01	0.060135
8	Acetylene	Strong	0.794266	0.326204	0.059588	0.761776	0.022200	0.034662	0.517446	15.0	0.199257	Yes	8.460033e-01	Yes	9.064345e-01	0.031752
9	Benzene	Strong	0.775260	0.663911	0.083857	0.707843	0.021822	0.051291	0.713236	5.0	0.700786	Yes	8.472921e-02	Yes	6.284243e-01	0.042233
10	Cis-2-Butene	Strong	0.023800	0.151174	0.014345	0.020631	0.007566	0.102968	0.241215	No	18.915672	No	2.767349e-17	No	2.331389e-12	0.210025
11	Cis-2-Pentene	Strong	0.537683	0.307519	0.027610	0.326511	0.017336	0.059417	0.426863	No	1.928500	No	1.374298e-04	Yes	1.425911e-01	0.065031
12	Ethane	Strong	0.544614	4.078593	0.479100	0.581442	0.030444	0.285990	3.937299	15.0	0.414018	Yes	1.631195e-01	Yes	6.544188e-01	0.041353
13	Ethylbenzene	Strong	0.817406	0.138257	0.028056	0.794919	0.021513	0.016394	0.252133	15.0	0.525375	Yes	3.090723e-01	Yes	3.678631e-01	0.051878
14	Ethylene	Strong	0.471828	3.061338	0.169182	0.254368	0.015410	0.337223	2.005987	2.5	0.793250	No	9.883852e-03	Yes	3.272898e-01	0.053691
15	Isobutane	Strong	0.620601	2.372488	0.160504	0.399963	0.017907	0.236354	1.913976	1.0	0.958293	Yes	7.874400e-02	Yes	3.837266e-01	0.051205
16	Isopentane	Strong	0.838153	4.432128	0.336064	0.543320	0.013671	0.538535	4.371680	No	2.232366	No	1.968678e-04	Yes	5.991538e-02	0.075025
17	Isoprene	Strong	0.607921	0.402327	0.110799	0.621228	0.028567	0.085110	1.187726	No	2.534531	No	4.233360e-04	No	1.759649e-02	0.087215
18	Isopropylbenzene	Strong	0.065906	0.014652	0.000680	0.005685	0.001226	0.029486	0.011118	No	27.576149	No	8.292839e-19	No	6.916830e-23	0.287047
19	M_P Xylene	Strong	0.879736	0.303012	0.070059	0.807864	0.017103	0.044009	0.648513	No	4.595439	No	2.066715e-14	No	2.037189e-02	0.085849
20	M-Diethylbenzene	Strong	0.424919	0.017390	0.001070	0.059285	0.003949	0.012386	0.015868	No	19.039798	No	5.743467e-16	No	1.319692e-12	0.212172

		Features	Category	r2	Intercept	Intercept SE	Slope	Slope SE	SE	SE Regression	Anderson Normal Residual	Anderson Statistic	Shapiro Normal Residuals	Shapiro PValue	KS Normal Residuals	KS PValue	KS Statistic
21	M-Ethyltoluene	Strong	0.737561	0.128704	0.033906	0.715840	0.024450	0.024291	0.354324	No	2.089867	No	2.437133e-05	No	3.588567e-02	0.080350	
22	N-Butane	Strong	0.861195	2.032279	0.222625	0.680979	0.015654	0.227036	2.588542	1.0	0.913657	Yes	1.972165e-01	Yes	3.270502e-01	0.053702	
23	N-Decane	Strong	0.646456	0.065161	0.018069	0.697876	0.029552	0.011562	0.174827	No	2.696522	No	4.029847e-05	Yes	8.269459e-02	0.071474	
24	N-Heptane	Strong	0.786343	0.199844	0.032514	0.770747	0.023005	0.018760	0.285489	1.0	0.925536	No	1.855063e-02	Yes	3.030518e-01	0.054847	
25	N-Hexane	Strong	0.374079	2.722746	0.209605	0.198064	0.014670	0.550488	2.935321	No	2.976483	No	3.395951e-06	No	1.201662e-02	0.090676	
26	N-Nonane	Strong	0.5711549	0.112225	0.019595	0.636112	0.031536	0.012337	0.180355	No	1.615565	No	2.123029e-03	Yes	7.281159e-02	0.072898	
27	N-Octane	Strong	0.572930	0.161886	0.026229	0.631725	0.031230	0.016429	0.238561	1.0	1.028387	No	3.798392e-02	Yes	1.316307e-01	0.066016	
28	N-Pentane	Strong	0.808734	1.554608	0.188007	0.662499	0.018448	0.175513	2.123449	No	2.247843	No	3.335715e-07	Yes	1.460849e-01	0.064730	
29	N-Propylbenzene	Strong	0.854405	-0.009551	0.006470	0.932711	0.022046	0.004473	0.077207	No	5.478733	No	1.501627e-08	No	2.184853e-03	0.104754	
30	N-Undecane	Strong	0.568061	0.108332	0.014957	0.553423	0.027633	0.012198	0.156862	No	2.234528	No	7.059643e-04	No	2.232979e-02	0.084982	
31	O-Ethyltoluene	Strong	0.794382	-0.004099	0.013312	0.861548	0.025098	0.008042	0.134368	10.0	0.647315	Yes	1.082856e-01	Yes	7.666526e-01	0.037491	
32	O-Xylene	Strong	0.843374	0.167732	0.031338	0.787187	0.019425	0.019157	0.284334	15.0	0.358267	Yes	5.803114e-01	Yes	8.006758e-01	0.036251	
33	Propane	Strong	0.627356	5.263036	0.451150	0.520448	0.022968	0.417401	4.692228	No	1.244810	No	3.112880e-03	Yes	3.036319e-01	0.054819	
34	Propylene	Strong	0.424983	1.496252	0.098400	0.249397	0.016611	0.200413	1.265902	15.0	0.328430	Yes	3.714180e-01	Yes	9.011820e-01	0.032018	
35	Styrene	Strong	0.392001	0.150323	0.025410	0.382166	0.027253	0.029928	0.320007	No	5.353592	No	4.322498e-09	No	2.171340e-03	0.104802	
36	Toluene	Strong	0.840805	1.149875	0.152740	0.749139	0.018665	0.100553	1.396239	15.0	0.261386	Yes	6.658835e-01	Yes	9.661129e-01	0.027821	
37	Trans-2-Butene	Strong	0.426870	0.073160	0.011585	0.312128	0.020709	0.021939	0.178890	No	22.085591	No	1.347885e-18	No	5.186542e-17	0.247213	
38	Trans-2-Pentene	Strong	0.761284	0.349360	0.036996	0.499011	0.016000	0.063126	0.538775	No	3.212229	No	6.090579e-06	No	1.737571e-02	0.087332	
39	Unidentified	Strong	0.030235	15.270432	1.088075	0.040005	0.012973	3.103851	12.491844	No	22.988677	No	3.701235e-22	No	1.586847e-09	0.183561	
40	TNMOC	Strong	0.822566	24.371812	5.008934	0.805524	0.021422	2.918889	45.378317	15.0	0.557870	Yes	6.784214e-02	Yes	7.495149e-01	0.038097	

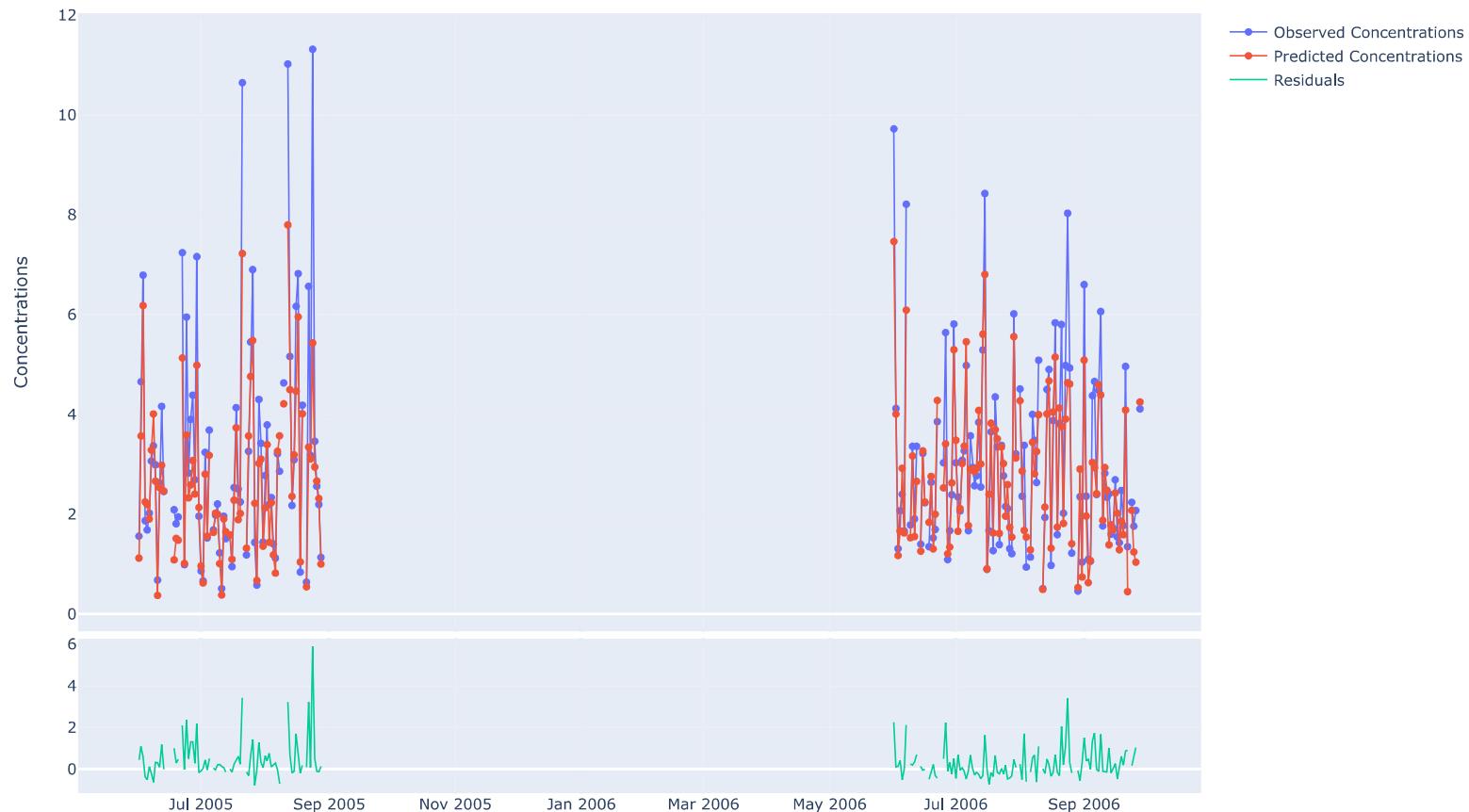
```
In [17]: # Model feature observed vs predicted plot with regression and one-to-one lines. Feature/Column specified by index.  
model_analysis.plot_estimated_observed(feature_idx=2)
```

Observed/Predicted Scatter Plot - 234-Trimethylpentane

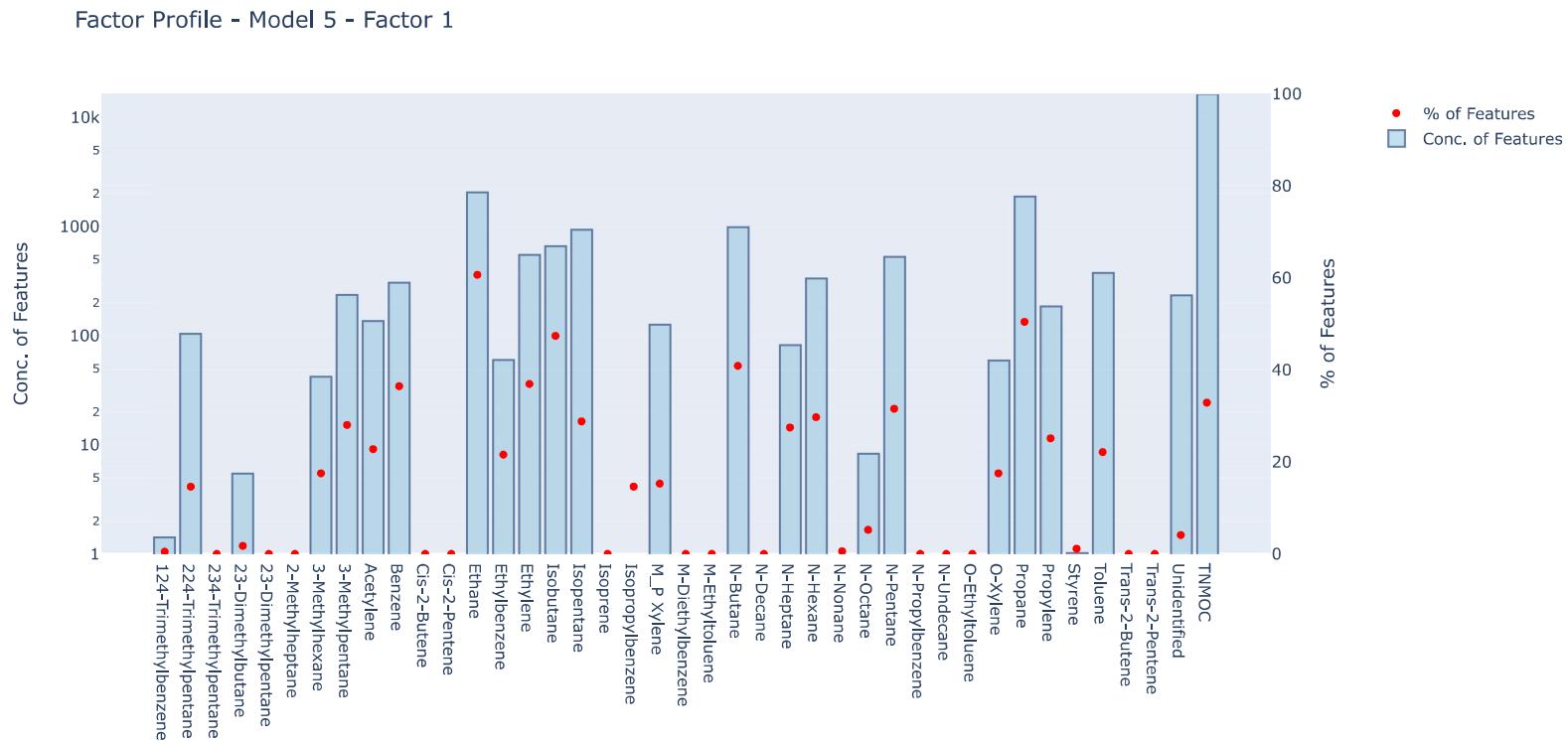


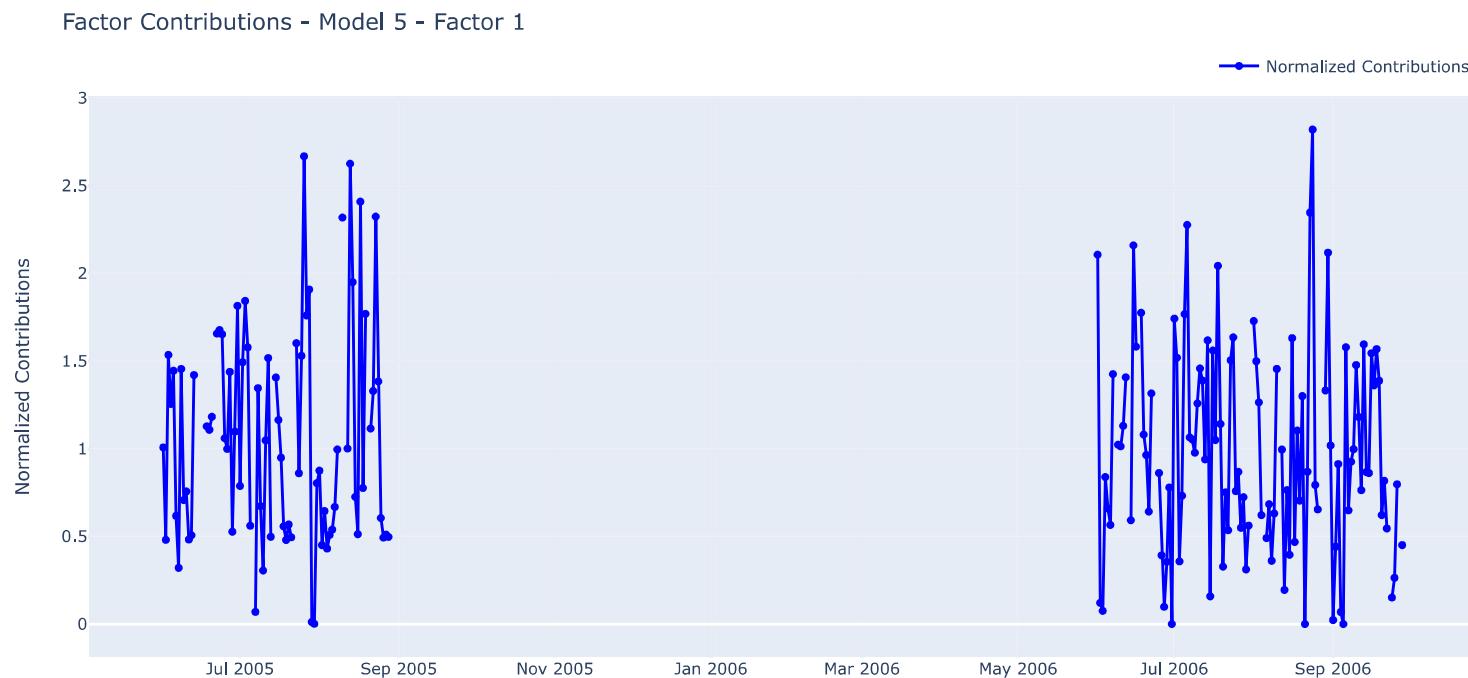
```
In [18]: # Model feature timeseries analysis plot showing the observed vs predicted values of the feature, along with the residuals shown below. Feature/column specified by index.  
model_analysis.plot_estimated_timeseries(feature_idx=1)
```

Estimated Time-series for 224-Trimethylpentane - Model 5



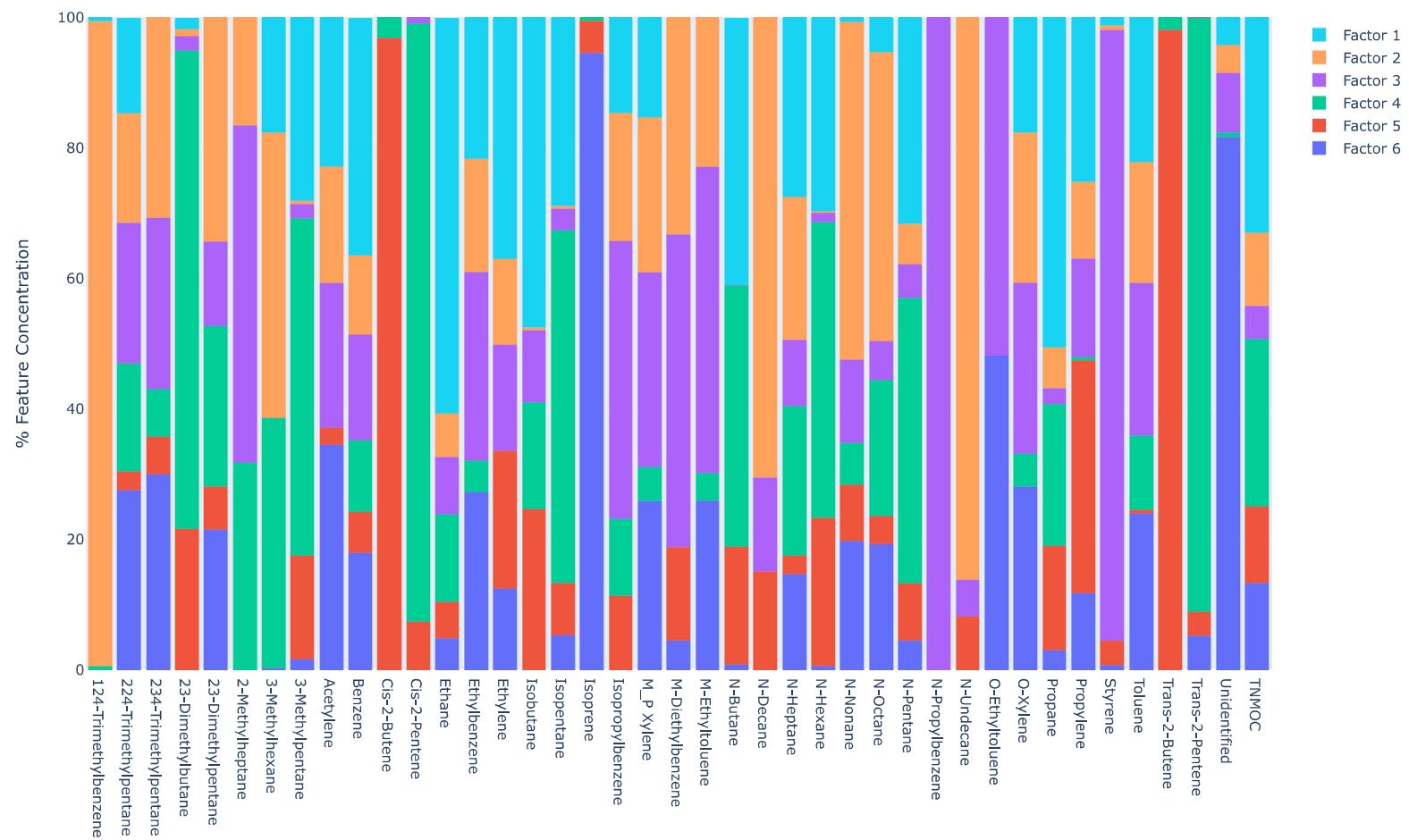
```
In [19]: # Factor profile plot showing the factor sum of concentrations by feature (blue bars), the percentage of the feature as the red dot, and in the bottom plot the normalized contributions  
# Factor specified by index.  
model_analysis.plot_factor_profile(factor_idx=1)
```





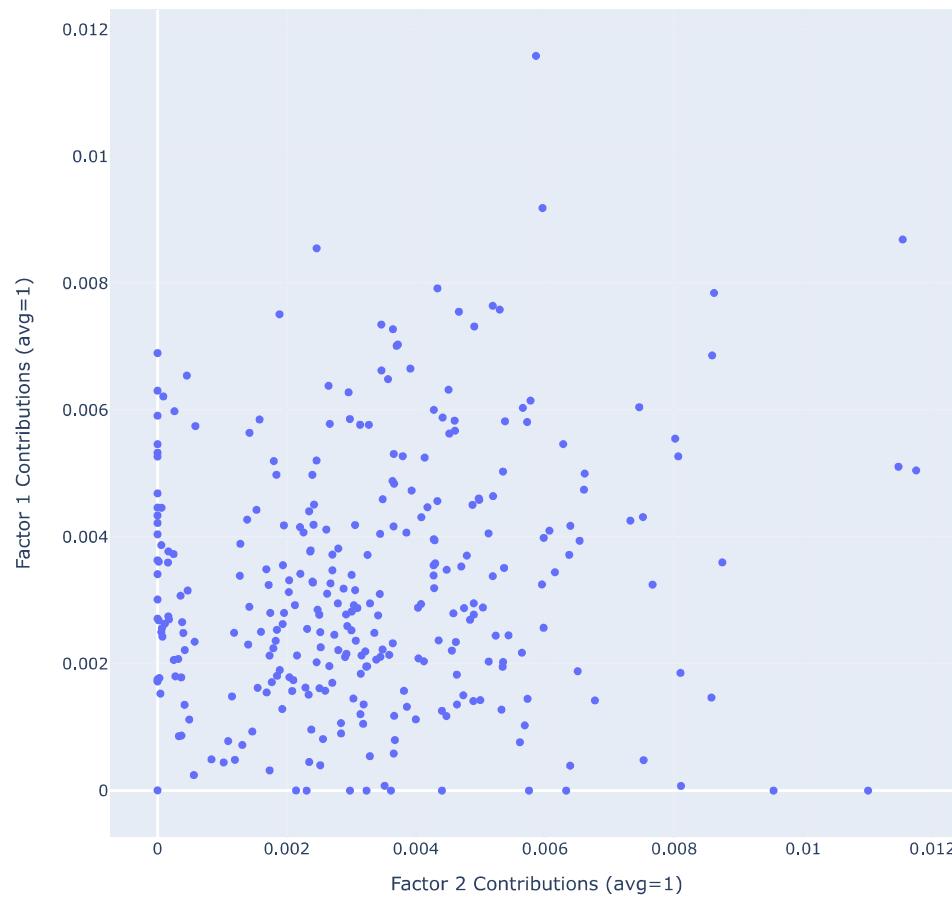
```
In [20]: # Model factor fingerprint specifies the feature percentage of each factor.  
model_analysis.plot_factor_fingerprints()
```

Factor Fingerprints - Model 5



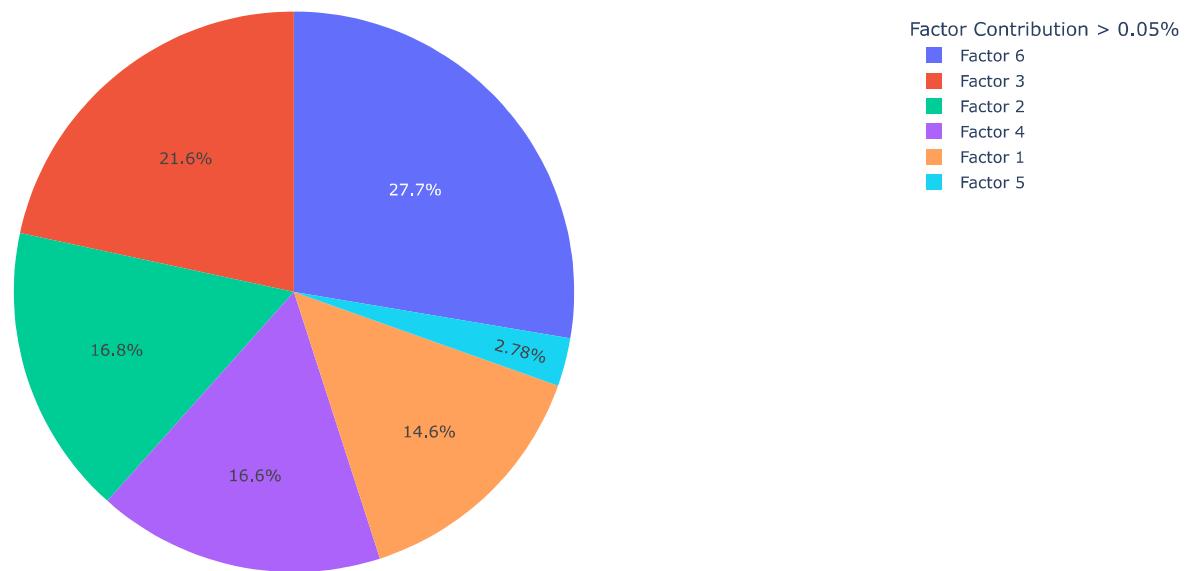
```
In [21]: # Factor G-Space plot shows the normalized contributions of one factor vs another factor. Factor specified by index.
model_analysis.plot_g_space(factor_1=2, factor_2=1)
```

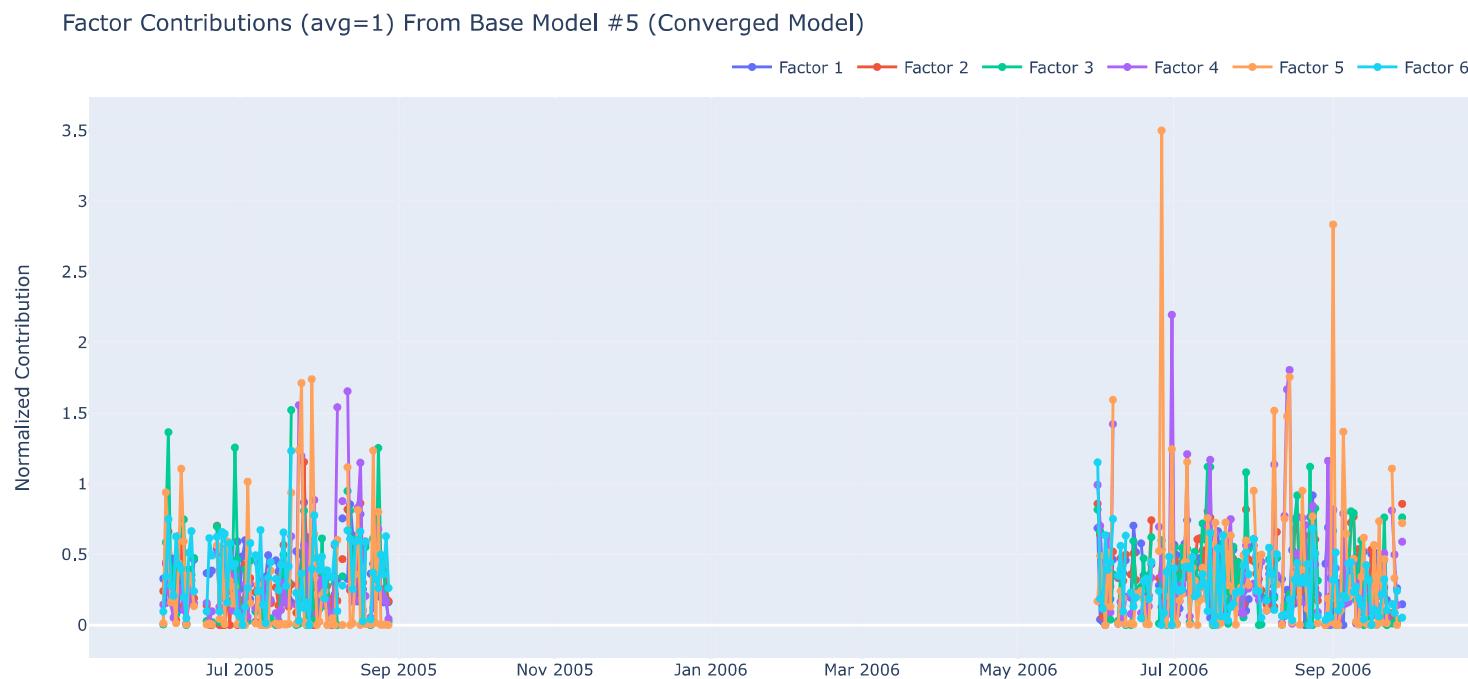
G-Space Plot - Model 5



```
In [22]: # Factor contribution pie chart shows the percentage of factor contributions for the specified feature, and the corresponding normalized contribution of each factor for that feature (b  
model_analysis.plot_factor_contributions(feature_idx=1)
```

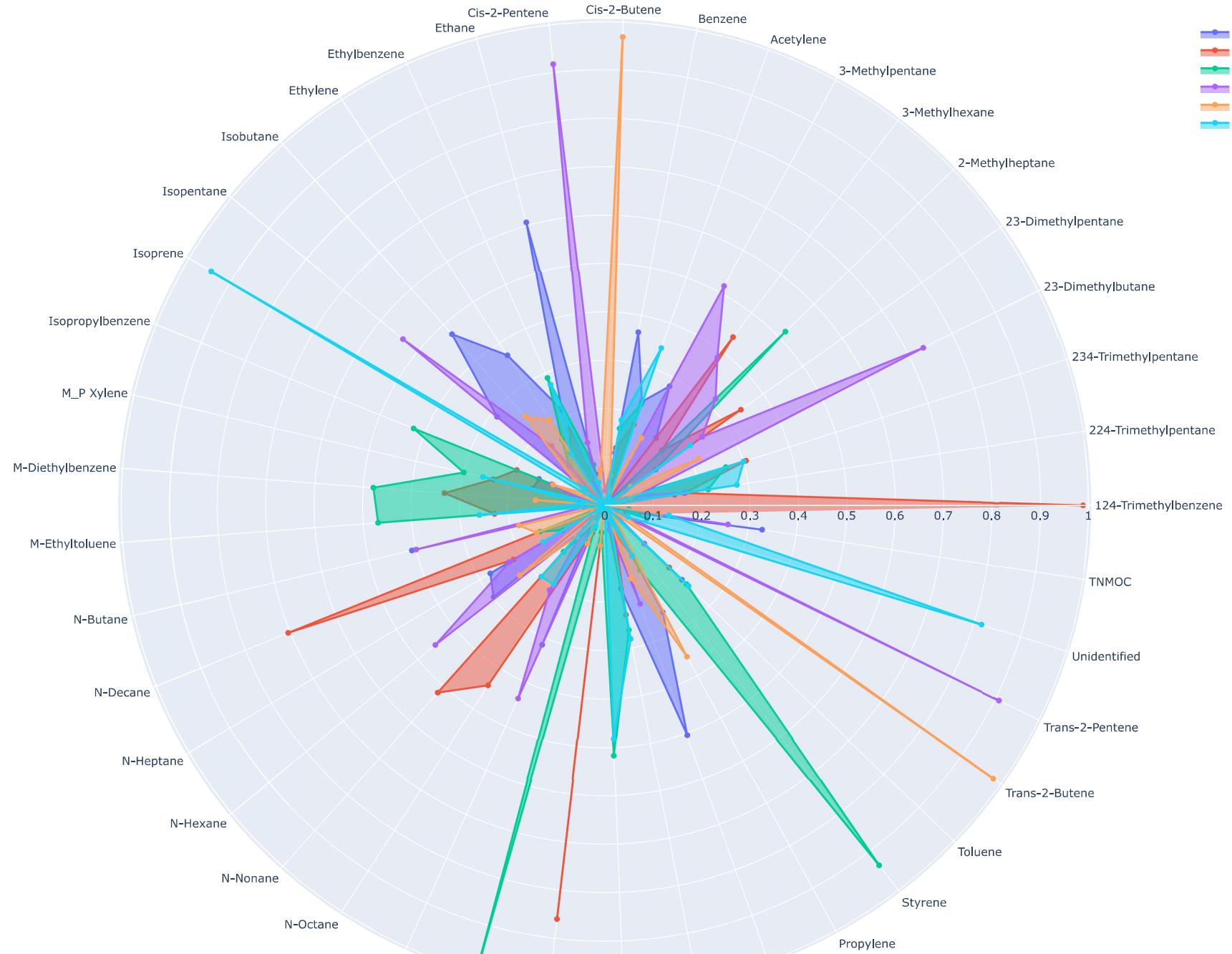
Factor Contributions to Feature: 224-Trimethylpentane - Model 5





```
In [23]: # New Graphic: Factor Profile Composition Radar Graph  
model_analysis.plot_factor_composition()
```

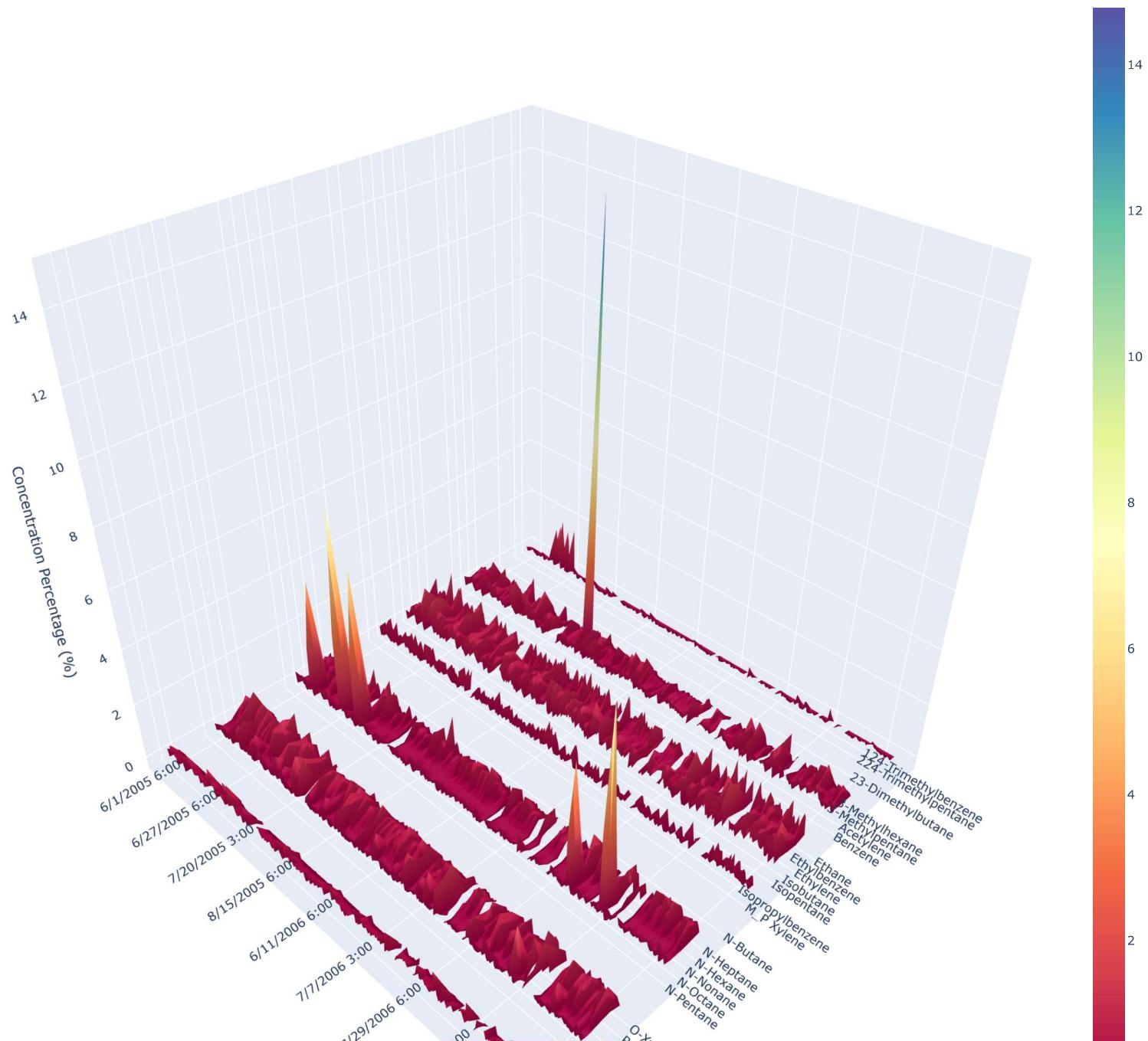
Factor Profile Composition





```
In [24]: # New Graphic: Factor Contribution Surface Plot
factor_idx = 1
feature_idx = None
model_analysis.plot_factor_surface(factor_idx=factor_idx, feature_idx=feature_idx, percentage=True)
```

Feature Concentrations for Factor 1





Error Estimation - Displacement

The displacement method for error estimation works by make slight adjustments to the factor profile values, individually, until a specific change in the loss value (dQ) is reached. There are 4 dQ values that are targetted, $dQ = 4, 8, 16, 32$.

The target dQ value is run for both an increase and decrease in the factor profile value, a single value in the factor profile matrix at a time. The change in factor profile is found by running a modified binary search to identify the value change within a small threshold, 0.1 of the target dQ . The value search is stopped if the change in the factor profile value is less than $1e-8$, in the instance that decreasing a factor profile value already near zero.

Once the change in the factor profile value is found that produces the target dQ , both increasing and decreasing, the modification to the H matrix is used as an initial guess for retraining a NMF model. The W matrix is reinitialized, using the original base model seed, and the model is trained to convergence.

The resulting model factor profile is checked to see if any factors swapped base upon the highest factor correlation with the base model factors. The output shows the swap %, based upon the number of retrained models where that factor was modified and the number of times a swap was detected.

The factor profile plot shows the variability in the factor profile feature values that correspond to the dQ target values, default shown in the plots is for a $dQ=4$. The last plot show the factor feature contribution variability based upon the same changes to the H matrix.

```
In [25]: # Import Error Estimation Displacement Method
from src.error.displacement import Displacement
```

```
In [26]: # Initialize the Displacement method, passing in the results of the batch nmf run and the features Labels from the data handler.
disp = Displacement(nmf=nmf_model, feature_labels=data_handler.features, model_selected=best_model, features=None)
```

```
In [27]: %%time
# Execute the displacement model, which will test both increasing and decreasing changes to the individual values of H for all dQ targets. Results are then compiled and prepared.
disp.run()
```

```
18-Jan-24 07:48:10 - Starting DISP for batch: -1
+ : Batch -1, Factor 1 - Features: 100% | 41/41 [00:28<00:00, 1.44it/s]
+ : Batch -1, Factor 2 - Features: 100% | 41/41 [00:29<00:00, 1.41it/s]
+ : Batch -1, Factor 3 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
+ : Batch -1, Factor 4 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
+ : Batch -1, Factor 5 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
+ : Batch -1, Factor 6 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
Increasing value for factors: 100% | 6/6 [02:50<00:00, 28.45s/it]
- : Batch -1, Factor 1 - Features: 100% | 41/41 [00:28<00:00, 1.44it/s]
- : Batch -1, Factor 2 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
- : Batch -1, Factor 3 - Features: 100% | 41/41 [00:28<00:00, 1.45it/s]
- : Batch -1, Factor 4 - Features: 100% | 41/41 [00:28<00:00, 1.44it/s]
- : Batch -1, Factor 5 - Features: 100% | 41/41 [00:28<00:00, 1.43it/s]
- : Batch -1, Factor 6 - Features: 100% | 41/41 [00:27<00:00, 1.47it/s]
Decreasing value for factors: 100% | 6/6 [02:50<00:00, 28.35s/it]
18-Jan-24 07:53:51 - Completed DISP for batch: -1
CPU times: total: 45min 19s
Wall time: 5min 40s
```

```
In [28]: # The swap table shows the percentage of times a factor was found to be more highly correlated as a result of the change in the factor feature value. This percentage is the number of t
# The Largest change in the dQ value is the largest difference (both from increasing and decreasing changes to the factor feature values) between a retrained model and the base model L
```

```
disp.summary()
```

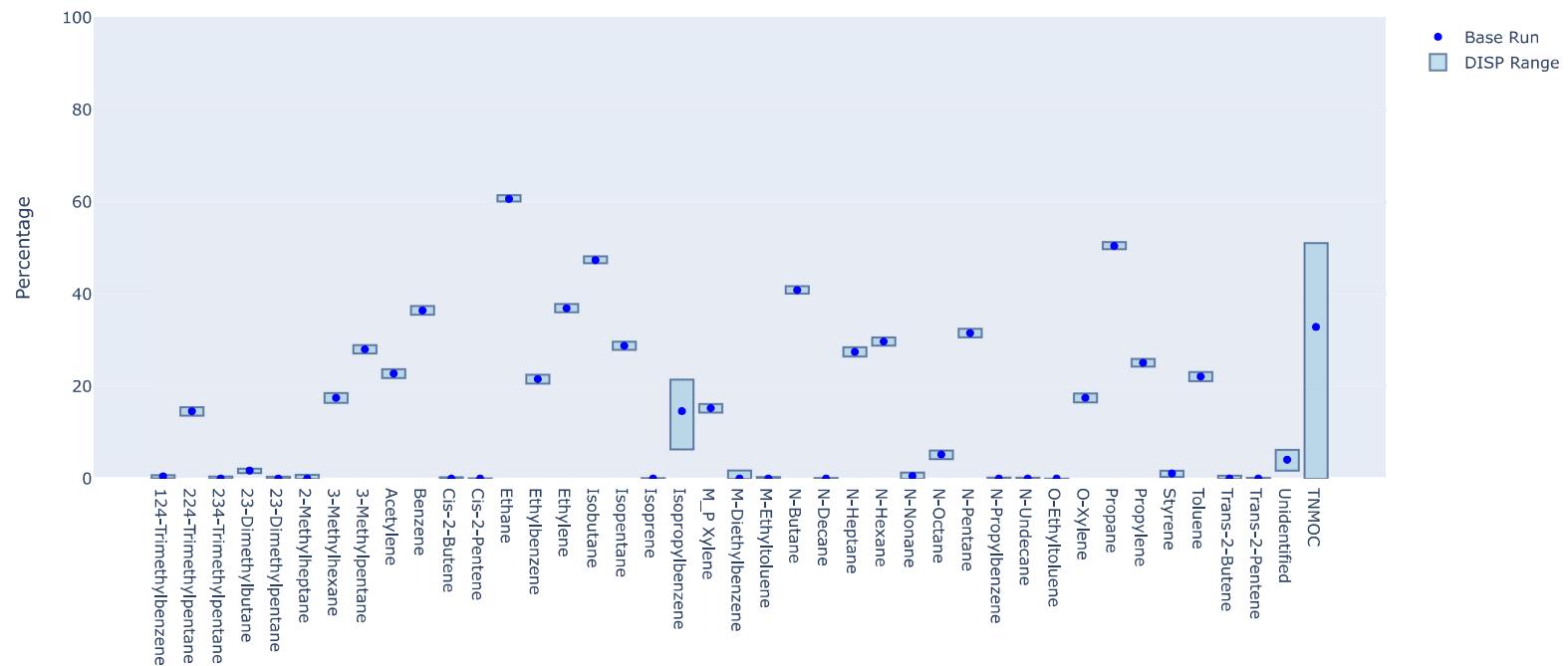
18-Jan-24 07:53:51 - Largest dQ Decrease: 0.0

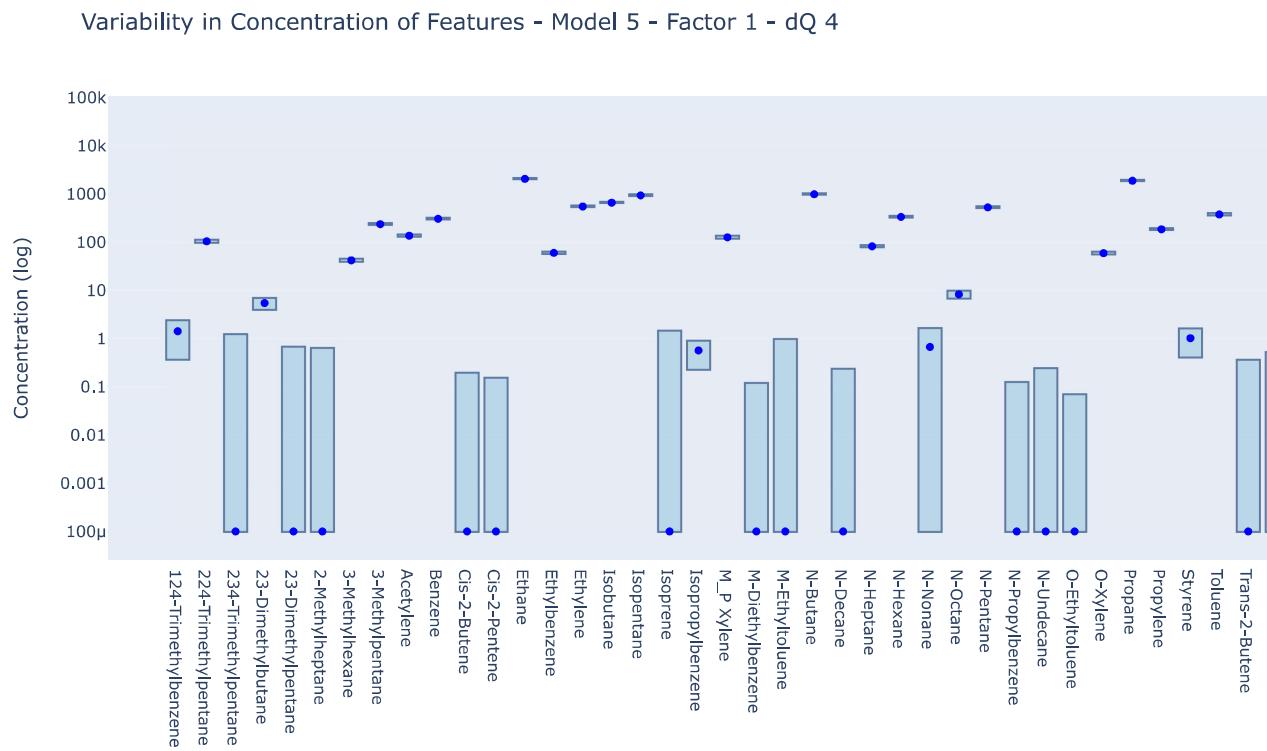
Swap %

dQ Max	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
4	0	1.22	1.22	0	0	0
8	0	1.22	1.22	0	0	0
16	0	1.22	1.22	0	0	0
32	0	1.22	1.22	0	0	0

In [29]: *# The results for a specific factor can be plotted, showing both the variability in the profile(%) and contribution for a given dQ value.*
 factor_i = 1
 disp.plot_results(factor=factor_i)

Variability in Percentage of Features - Model 5 - Factor 1 - dQ 4





Error Estimation - Bootstrap

The bootstrap method used is the block bootstrap method for time-series data. Here the initial dataset is broken into chunks of a specified size, containing sequential samples, and randomly added to a bootstrap dataset until the bootstrap dataset is the same size as the initial dataset.

The recommended block size calculation feature has not yet been implemented.

The aim of the bootstrap method is to quantify the variability in the factor profiles and contributions when the order of the datasets have been shuffled, resampled. In this case the block bootstrap method is the default method, while the full bootstrap method can be used by setting the block parameter in the run function to false, i.e. `bs.run(block=False)`. The blocks are randomly selected with replacement, allowing for the same block to be added to the bs dataset more than once. The final selected block is reduced in size until the bs dataset is exactly the same as the initial dataset.

The resampling is completed a specified number of times, set by `bootstrap_n`. During each bootstrap run, the initial datasets are resampled (both the data and uncertainty datasets are resampled using the same indeces) and used to retrain a NMF model. The NMF model uses the resampled data and uncertainty data, the base model H matrix and the base model random seed. W is reinitialized and the model is trained to convergence. The resulting factor contributions, V'_{bs} are mapped to the base model V'_{base} where the correlation between all combinations of the factor contributions are checked. The mapping between the highest correlation is then noted, if the correlation is above the user specified threshold (default=0.6), and the results are shown in the summary statistics. The summary statistics also include metrics for the distribution of the bootstrap model results to compare against the base model results, these are shown as tables for each factor in the summary statistics.

The results of all the bootstrap runs are compiled to provide a distribution for each factor/feature percentage and contribution, these are shown in the plotted results.

```
In [30]: # Import Bootstrap Error Estimation Method
from src.error.bootstrap import Bootstrap
```

```
In [31]: # Bootstrap input parameters
model_selected = nmf_models.best_model
nmf_model = nmf_models.results[model_selected]
feature_labels = data_handler.features

bootstrap_n = 20
block_size = data_handler.optimal_block
threshold = 0.6
seed = seed
print(f"Optimal BS block size: {data_handler.optimal_block}")

Optimal BS block size: 4
```

```
In [32]: # Initialize the bootstrap object
bs = Bootstrap(nmf=nmf_model, feature_labels=feature_labels, model_selected=model_selected, bootstrap_n=bootstrap_n, block_size=block_size, threshold=threshold, seed=seed)
```

```
In [33]: %%time
# Execute the bootstrap runs with default parameters
bs.run()
```

Bootstrap resampling, training and mapping: 100%|██████████| 20/20 [00:10<00:00, 1.82it/s]

CPU times: total: 37.8 s
Wall time: 11 s

```
In [34]: # Print the output summary of all the bootstrap runs.
bs.summary()
```

NMF Bootstrap Error Estimation Summary
----- Input Parameters -----
Base model run number: 5
Number of bootstrap runs: 20
Min. Correlation R-Value: 0.6
Number of Factors: 6

Mapping of bootstrap factors to base factors

Boot Factors	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Base Factor 5	Base Factor 6	Unmapped
Boot Factor 1	20	0	0	0	0	0	0
Boot Factor 2	0	17	3	0	0	0	0
Boot Factor 3	0	4	12	0	0	4	0
Boot Factor 4	0	0	0	20	0	0	0
Boot Factor 5	0	0	0	2	18	0	0
Boot Factor 6	0	0	0	0	0	20	0

Q(Robust) Percentile Report

Base	Min	25th	Median	75th	Max
54424	49590	51789	52613	53426	54697

Bootstrap run uncertainty statistics - Factor 1

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.002892	true	0.035323	0.076582	0	0.000798	0.000798
224-Trimethylpentane	0.211926	true	0.209144	0.042722	0.149392	0.171307	0.211926
234-Trimethylpentane	0	true	0.002726	0.005743	0	0	0.211926
23-Dimethylbutane	0.011113	true	0.016089	0.012091	0	0.006275	0.006275
23-Dimethylpentane	0	true	0.001087	0.004863	0	0	0.001087
2-Methylheptane	0	true	0.000209	0.000934	0	0	0
3-Methylhexane	0.085673	true	0.073987	0.050369	0	0.028089	0.028089
3-Methylpentane	0.481379	false	0.510838	0.057761	0.418721	0.484277	0.500000
Acetylene	0.27739	true	0.273052	0.048256	0.186193	0.257576	0.280000
Benzene	0.62234	true	0.636866	0.048327	0.562957	0.590917	0.610000
Cis-2-Butene	0	true	0	0	0	0	0
Cis-2-Pentene	0	true	0	0	0	0	0
Ethane	4.18482	false	4.578196	0.482773	4.042894	4.374417	4.490000
Ethylbenzene	0.122034	true	0.122544	0.021367	0.096194	0.107488	0.122034
Ethylene	1.119339	true	1.10757	0.195657	0.900167	0.951777	1.060000
Isobutane	1.344812	false	1.543332	0.193974	1.269784	1.440817	1.540000
Isopentane	1.903126	false	2.013977	0.134983	1.786202	1.950074	2.010000
Isoprene	0	true	0	0	0	0	0
Isopropylbenzene	0.001158	true	0.001329	0.00131	0	0.00007	0.00007
M_P Xylene	0.256893	true	0.26845	0.06721	0.193409	0.214663	0.256893
M-Diethylbenzene	0	true	0	0	0	0	0
M-Ethyltoluene	0	true	0.002382	0.005055	0	0	0
N-Butane	2.008926	false	2.224243	0.20496	1.910714	2.109026	2.160000
N-Decane	0	true	0	0	0	0	0
N-Heptane	0.166933	true	0.173121	0.01413	0.153088	0.161714	0.170000
N-Hexane	0.680714	true	0.708213	0.066782	0.628859	0.647569	0.710000
N-Nonane	0.001367	true	0.001965	0.002651	0	0	0.001367
N-Octane	0.016889	true	0.018057	0.012056	0.006069	0.009772	0.016889
N-Pentane	1.073978	false	1.171829	0.089053	1.010532	1.126365	1.110000
N-Propylbenzene	0	true	0	0	0	0	0
N-Undecane	0	true	0	0	0	0	0
O-Ethyltoluene	0	true	0	0	0	0	0
O-Xylene	0.120623	true	0.122873	0.027894	0.086711	0.108508	0.120623
Propane	3.829896	false	4.216249	0.473323	3.465101	3.913911	4.300000
Propylene	0.377601	true	0.36796	0.040432	0.307048	0.337399	0.370000
Styrene	0.00208	true	0.00197	0.002589	0	0	0.00208
Toluene	0.765138	true	0.787184	0.124347	0.627878	0.692937	0.790000
Trans-2-Butene	0	true	0	0	0	0	0
Trans-2-Pentene	0	true	0	0	0	0	0
Unidentified	0.47688	true	0.684477	0.722517	0	0.331313	0.390000
TNMOC	33.375656	true	33.838976	2.312745	28.931145	33.006883	34.100000

Bootstrap run uncertainty statistics - Factor 2

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.589181	false	0.495343	0.09338	0.359745	0.485128	0.52
224-Trimethylpentane	0.243338	true	0.263469	0.046082	0.197703	0.231956	0.24
234-Trimethylpentane	0.145374	true	0.130598	0.025565	0.080363	0.119953	0.13
23-Dimethylbutane	0.006475	true	0.021738	0.049518	0	0	0.00
23-Dimethylpentane	0.098541	true	0.092034	0.014068	0.066628	0.085201	0.09
2-Methylheptane	0.024015	true	0.026033	0.021961	0.005862	0.010284	0.01
3-Methylhexane	0.213784	true	0.177295	0.063573	0.071528	0.143284	0.11
3-Methylpentane	0.008076	true	0.037378	0.035561	0	0.00036	0.00
Acetylene	0.216568	true	0.231027	0.044463	0.165468	0.192337	0.21
Benzene	0.206193	true	0.242737	0.056369	0.15736	0.203643	0.24
Cis-2-Butene	0	true	0	0	0	0	0
Cis-2-Pentene	0	true	0	0	0	0	0
Ethane	0.459201	true	0.568885	0.224851	0.163443	0.456629	0.56
Ethylbenzene	0.098626	true	0.105436	0.018531	0.075923	0.096204	0.10
Ethylene	0.397574	true	0.512447	0.16579	0.323916	0.391887	0.51
Isobutane	0.013263	true	0.033847	0.033403	0.000034	0.003884	0.00
Isopentane	0.033387	true	0.106467	0.103489	0	0.006011	0.08
Isoprene	0	true	0	0	0	0	0
Isopropylbenzene	0.00155	true	0.001192	0.001223	0	0	0.00
M_P Xylene	0.397847	true	0.388724	0.06594	0.286004	0.348401	0.31
M-Diethylbenzene	0.004377	true	0.003501	0.00269	0	0.001895	0.00
M-Ethyltoluene	0.108736	true	0.112182	0.034811	0.066409	0.087991	0.10
N-Butane	0	true	0	0	0	0	0
N-Decane	0.156316	false	0.136416	0.030691	0.108915	0.126589	0.14
N-Heptane	0.132947	true	0.131589	0.019851	0.107377	0.113853	0.11
N-Hexane	0.005896	true	0.041993	0.049371	0	0.000104	0.00
N-Nonane	0.125645	false	0.115565	0.009943	0.100387	0.111488	0.11
N-Octane	0.143128	false	0.125137	0.015208	0.104806	0.117143	0.11
N-Pentane	0.211229	true	0.234568	0.075723	0.15378	0.179945	0.21
N-Propylbenzene	0	true	0	0	0	0	0
N-Undecane	0.157211	false	0.129805	0.014588	0.109509	0.118889	0.11
O-Ethyltoluene	0	true	0	0	0	0	0
O-Xylene	0.15849	true	0.161296	0.028737	0.119615	0.140435	0.16
Propane	0.482815	true	0.492057	0.173913	0.201126	0.424988	0.46
Propylene	0.176567	true	0.201845	0.064484	0.122321	0.163236	0.19
Styrene	0.001318	false	0.001129	0.002617	0	0	0
Toluene	0.638893	true	0.665554	0.106708	0.531842	0.601678	0.67
Trans-2-Butene	0	true	0	0	0	0	0
Trans-2-Pentene	0	true	0.000631	0.001954	0	0	0
Unidentified	0.502719	true	0.978541	1.306864	0	0	0.50
TNMOC	11.428929	true	12.851594	2,402053	9,42191	10.824211	12.4

Bootstrap run uncertainty statistics - Factor 3

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0	true	0	0	0	0	0
224-Trimethylpentane	0.313731	true	0.288845	0.048706	0.214313	0.277108	0.29
234-Trimethylpentane	0.124578	true	0.117016	0.019888	0.091777	0.10701	0.11
23-Dimethylbutane	0.013978	true	0.024506	0.024547	0	0	0.02
23-Dimethylpentane	0.036894	true	0.03448	0.016318	0.009397	0.024882	0.03
2-Methylheptane	0.075709	true	0.066606	0.015583	0.044687	0.064493	0.06
3-Methylhexane	0	true	0	0	0	0	0
3-Methylpentane	0.039057	true	0.042146	0.046824	0	0.000679	0.01
Acetylene	0.270467	true	0.252254	0.03708	0.215488	0.229461	0.24
Benzene	0.276213	true	0.254283	0.040934	0.207722	0.227875	0.24
Cis-2-Butene	0	true	0	0	0	0	0
Cis-2-Pentene	0.001897	true	0.006082	0.009337	0	0.000014	0.00
Ethane	0.602721	false	0.483166	0.165322	0.275412	0.314016	0.52
Ethylbenzene	0.163238	false	0.153016	0.018276	0.130999	0.140405	0.14
Ethylene	0.493328	false	0.368236	0.121265	0.190574	0.286133	0.31
Isobutane	0.312853	true	0.262008	0.070891	0.16807	0.212281	0.26
Isopentane	0.216567	true	0.239331	0.137415	0.032332	0.183594	0.21
Isoprene	0	true	0	0	0	0	0
Isopropylbenzene	0.003369	true	0.003896	0.002398	0.000818	0.002631	0.00
M_P Xylene	0.502572	true	0.506232	0.052165	0.434635	0.475123	0.49
M-Diethylbenzene	0.006317	true	0.00569	0.004422	0.001757	0.003103	0.01
M-Ethyltoluene	0.223713	true	0.213547	0.024209	0.177053	0.199243	0.20
N-Butane	0.005738	true	0.054502	0.086834	0	0.000034	0.00
N-Decane	0.031986	true	0.030816	0.024379	0.001327	0.01907	0.02
N-Heptane	0.061271	true	0.054122	0.015555	0.026218	0.04784	0.01
N-Hexane	0.029721	true	0.050729	0.066349	0	0.000286	0.01
N-Nonane	0.030978	true	0.027441	0.010798	0.01153	0.020193	0.02
N-Octane	0.019337	true	0.016406	0.011089	0	0.007089	0.01
N-Pentane	0.175807	true	0.167014	0.070712	0.060746	0.117439	0.16
N-Propylbenzene	0.086975	true	0.081958	0.00735	0.071059	0.077913	0.0
N-Undecane	0.010243	true	0.015678	0.016314	0.000594	0.002673	0.00
O-Ethyltoluene	0.104785	true	0.105194	0.014671	0.08496	0.092531	0.10
O-Xylene	0.180475	true	0.168579	0.020577	0.14249	0.156224	0.11
Propane	0.184809	true	0.115055	0.147566	0.000066	0.002907	0.01
Propylene	0.227466	true	0.204888	0.032222	0.148849	0.188541	0.20
Styrene	0.17325	true	0.155645	0.02449	0.111784	0.142795	0.11
Toluene	0.810481	true	0.769692	0.119157	0.607605	0.686876	0.78
Trans-2-Butene	0	true	0	0	0	0	0
Trans-2-Pentene	0.000117	true	0.00527	0.012353	0	0	0
Unidentified	1.059308	true	0.962474	1.08545	0	0	0.54
TNMOC	5.121967	true	4.705033	1.934483	2.063911	3.475338	4.71

Bootstrap run uncertainty statistics - Factor 4

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.003589	true	0.003865	0.004929	0	0	0.004929
224-Trimethylpentane	0.240273	true	0.22674	0.059395	0.166371	0.186813	0.240273
234-Trimethylpentane	0.034866	true	0.0359	0.01967	0.014919	0.025544	0.034866
23-Dimethylbutane	0.467402	true	0.431824	0.055055	0.340655	0.391915	0.467402
23-Dimethylpentane	0.070421	true	0.071495	0.01136	0.057478	0.062007	0.070421
2-Methylheptane	0.046492	true	0.045974	0.01353	0.031003	0.035813	0.046492
3-Methylhexane	0.187685	true	0.180497	0.032157	0.137505	0.16232	0.187685
3-Methylpentane	0.885938	true	0.863411	0.135843	0.705847	0.755595	0.885938
Acetylene	0	true	0	0	0	0	0
Benzene	0.188794	true	0.166317	0.049699	0.095523	0.132499	0.188794
Cis-2-Butene	0.003325	true	0.003237	0.002539	0.00012	0.001017	0.003325
Cis-2-Pentene	0.218562	true	0.214567	0.03195	0.162582	0.189639	0.218562
Ethane	0.928439	true	0.881518	0.266756	0.443727	0.765786	0.928439
Ethylbenzene	0.026985	false	0.022702	0.013776	0.00324	0.013973	0.026985
Ethylene	0	true	0	0	0	0	0
Isobutane	0.466149	false	0.505661	0.149274	0.182147	0.467271	0.505661
Isopentane	3.567409	true	3.572216	0.476505	2.728486	3.297761	3.567409
Isoprene	0.007936	true	0.022348	0.029622	0	0.00139	0.007936
Isopropylbenzene	0.000932	true	0.000847	0.000795	0	0.000128	0.000932
M_P_Xylene	0.088699	true	0.082986	0.057183	0.026921	0.050998	0.088699
M-Diethylbenzene	0	true	0	0	0	0	0
M-Ethyltoluene	0.020277	true	0.020808	0.016576	0.001649	0.01283	0.020277
N-Butane	1.964461	true	1.932866	0.228721	1.549673	1.843049	1.964461
N-Decane	0	true	0.000449	0.001467	0	0	0
N-Heptane	0.139152	true	0.134246	0.027558	0.101412	0.112717	0.139152
N-Hexane	1.037974	true	1.028915	0.190817	0.77724	0.869648	1.037974
N-Nonane	0.015491	true	0.022525	0.014456	0.005752	0.013601	0.015491
N-Octane	0.067465	true	0.072648	0.025191	0.039507	0.056957	0.067465
N-Pentane	1.487912	true	1.487932	0.152595	1.215784	1.380602	1.487912
N-Propylbenzene	0	true	0.00011	0.000286	0	0	0
N-Undecane	0	true	0	0	0	0	0
O-Ethyltoluene	0	true	0	0	0	0	0
O-Xylene	0.035161	true	0.031562	0.017406	0.005837	0.020481	0.035161
Propane	1.639279	true	1.666738	0.316788	1.214462	1.523692	1.639279
Propylene	0.007901	true	0.035159	0.035558	0	0.002755	0.007901
Styrene	0	true	0.000203	0.000756	0	0	0
Toluene	0.392567	false	0.340518	0.069113	0.245562	0.294553	0.392567
Trans-2-Butene	0.001861	true	0.002512	0.001423	0.000916	0.001191	0.001861
Trans-2-Pentene	0.391407	true	0.379846	0.045425	0.304327	0.350456	0.391407
Unidentified	0.086532	true	0.38795	0.695709	0	0	0.086532
TNMOC	26.12796	true	25.308418	2.861794	21.137777	23.880904	24.9

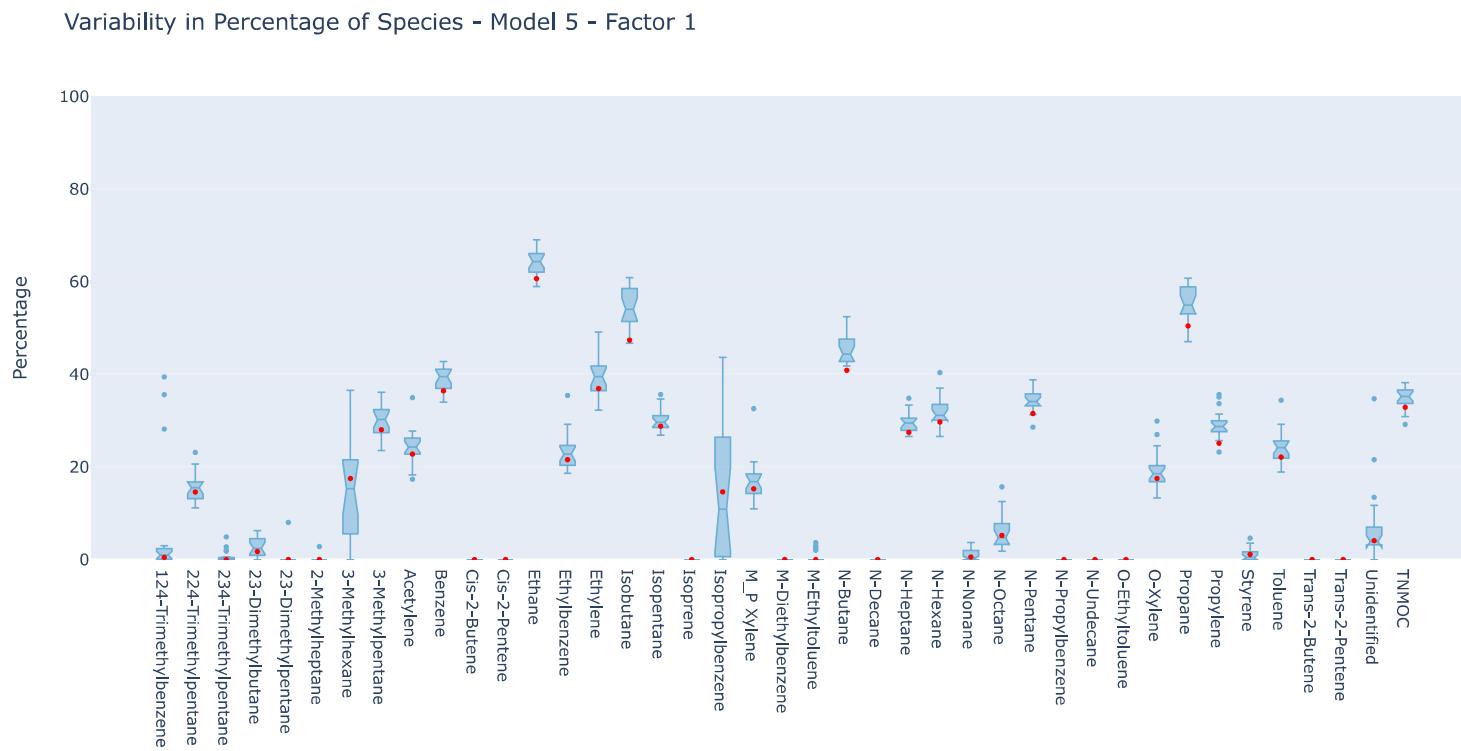
Bootstrap run uncertainty statistics - Factor 5

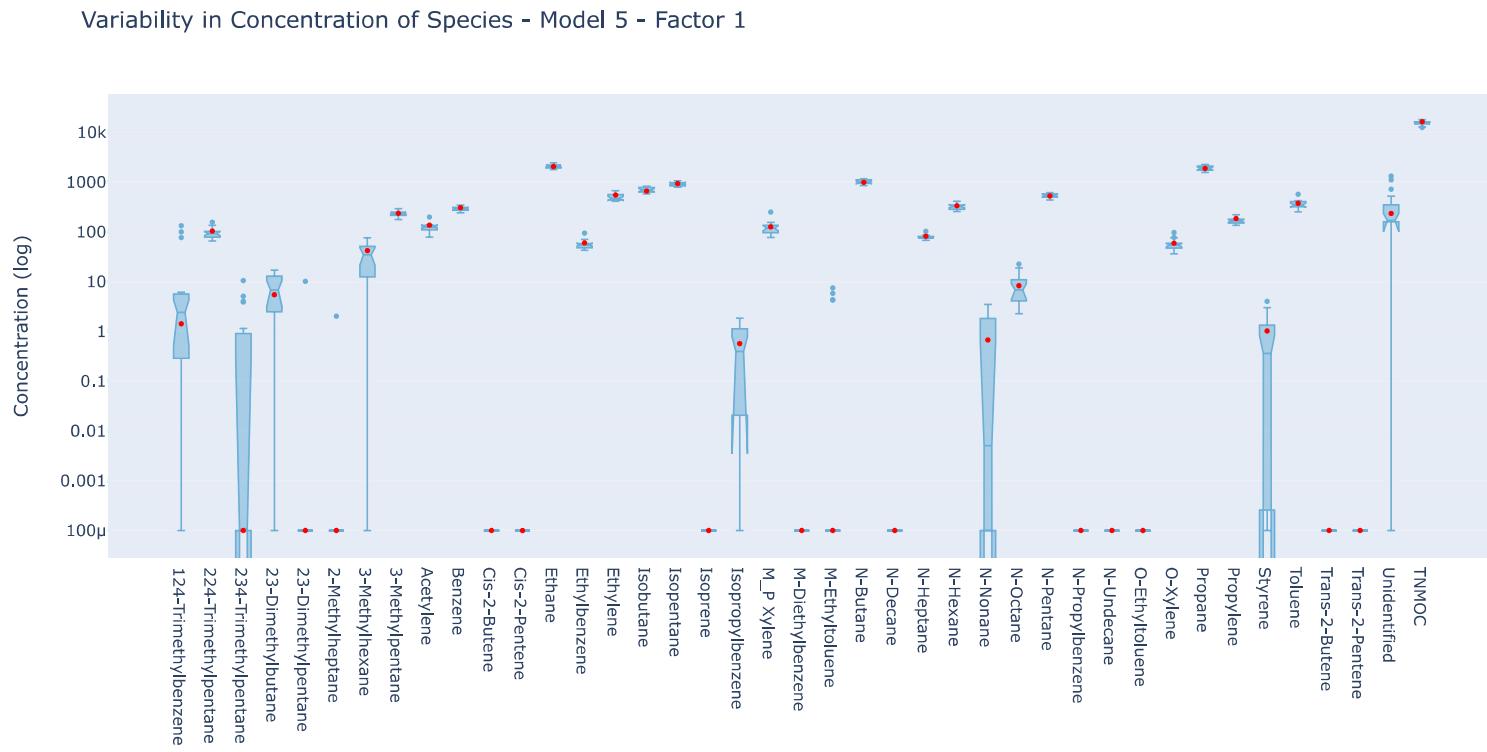
features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0	true	0.004338	0.015682	0	0	0
224-Trimethylpentane	0.040363	true	0.054384	0.039923	0.010182	0.0187	0.01
234-Trimethylpentane	0.026936	true	0.027196	0.019328	0.000467	0.009562	0.00
23-Dimethylbutane	0.13778	false	0.112486	0.038644	0.060888	0.088168	0.11
23-Dimethylpentane	0.018895	true	0.01437	0.008329	0.001417	0.009415	0.00
2-Methylheptane	0	true	0.002453	0.006291	0	0	0
3-Methylhexane	0	true	0	0	0	0	0
3-Methylpentane	0.272057	true	0.226789	0.104313	0.064788	0.164048	0.21
Acetylene	0.032152	false	0.018433	0.014508	0	0.005404	0.00
Benzene	0.10668	false	0.073503	0.046077	0.026368	0.038681	0.06
Cis-2-Butene	0.105556	false	0.07169	0.018475	0.043542	0.058713	0.06
Cis-2-Pentene	0.017702	true	0.020087	0.023966	0	0.000701	0.00
Ethane	0.380834	false	0.23446	0.252096	0.001975	0.053081	0.16
Ethylbenzene	0	true	0.00114	0.003058	0	0	0
Ethylene	0.637753	false	0.46901	0.283117	0.178079	0.281044	0.31
Isobutane	0.700365	false	0.447101	0.149008	0.20924	0.352915	0.41
Isopentane	0.531097	true	0.471563	0.406165	0.007951	0.190759	0.40
Isoprene	0.07779	false	0.033211	0.034131	0.000553	0.001743	0.00
Isopropylbenzene	0.000911	true	0.001054	0.001036	0	0.000018	0.00
M_P_Xylene	0	true	0	0	0	0	0
M-Diethylbenzene	0.001888	true	0.007898	0.014465	0	0	0.00
M-Ethyltoluene	0	true	0.001609	0.005059	0	0	0
N-Butane	0.893476	false	0.623585	0.28879	0.287237	0.443138	0.51
N-Decane	0.033585	false	0.023278	0.012426	0.00199	0.016126	0.02
N-Heptane	0.017067	true	0.017603	0.01216	0.002682	0.009057	0.01
N-Hexane	0.519971	false	0.360181	0.13753	0.175467	0.264875	0.31
N-Nonane	0.021147	true	0.014545	0.010657	0	0.00284	0.00
N-Octane	0.01364	true	0.014962	0.013332	0	0.002783	0.00
N-Pentane	0.295467	true	0.244553	0.154707	0.038238	0.154798	0.21
N-Propylbenzene	0	true	0.000206	0.000707	0	0	0
N-Undecane	0.015264	true	0.011958	0.008706	0.001008	0.005756	0.00
O-Ethyltoluene	0	true	0	0	0	0	0
O-Xylene	0	true	0.002271	0.007291	0	0	0
Propane	1.217287	false	0.824581	0.332583	0.440137	0.633701	0.79
Propylene	0.534636	false	0.298685	0.097819	0.162806	0.234147	0.2
Styrene	0.007095	true	0.014083	0.024187	0	0.000958	0.00
Toluene	0.01875	true	0.044825	0.035149	0	0.017841	0.04
Trans-2-Butene	0.103144	false	0.068001	0.014874	0.045421	0.057964	0.01
Trans-2-Pentene	0.015818	true	0.031268	0.040295	0	0.001103	0.00
Unidentified	0	true	0	0	0	0	0
TNMOC	11.816523	false	8.351712	2.362828	4.705613	7.22744	8.60

Bootstrap run uncertainty statistics - Factor 6

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0	true	0.009312	0.040557	0	0	0
224-Trimethylpentane	0.401161	false	0.2976	0.047403	0.240892	0.264623	0.29
234-Trimethylpentane	0.142911	false	0.106349	0.019939	0.086161	0.092552	0.10
23-Dimethylbutane	0	true	0	0	0	0	0
23-Dimethylpentane	0.06191	false	0.044281	0.012873	0.020311	0.041151	0.04
2-Methylheptane	0	true	0	0	0	0	0
3-Methylhexane	0.001823	false	0.055321	0.042635	0	0.013511	0.01
3-Methylpentane	0.030292	true	0.02586	0.022388	0	0.002779	0.02
Acetylene	0.421066	false	0.340981	0.03337	0.294318	0.30891	0.34
Benzene	0.306758	false	0.258388	0.029691	0.222761	0.23381	0.21
Cis-2-Butene	0	true	0	0	0	0	0
Cis-2-Pentene	0	true	0	0	0	0	0
Ethane	0.341451	true	0.365223	0.140825	0.182038	0.271909	0.36
Ethylbenzene	0.154643	false	0.12122	0.015739	0.097478	0.113046	0.13
Ethylene	0.381845	false	0.344681	0.06871	0.267517	0.318257	0.33
Isobutane	0	true	0.011183	0.028651	0	0	0
Isopentane	0.355507	true	0.306953	0.085228	0.164337	0.274477	0.30
Isoprene	1.524843	false	1.407022	0.228394	1.137496	1.203519	1.39
Isopropylbenzene	0	true	0	0	0	0	0
M_P Xylene	0.435352	false	0.310263	0.090872	0.179237	0.238309	0.32
M-Diethylbenzene	0.000605	false	0.000354	0.000534	0	0	0.00
M-Ethyltoluene	0.124044	false	0.073649	0.046073	0.008322	0.031362	0.01
N-Butane	0.041146	true	0.063167	0.045501	0.000152	0.015132	0.06
N-Decane	0	true	0.002189	0.004212	0	0	0
N-Heptane	0.089983	false	0.07352	0.015268	0.053743	0.061812	0.07
N-Hexane	0.016022	true	0.020924	0.022974	0	0.000251	0.01
N-Nonane	0.047872	false	0.02835	0.010727	0.01059	0.021926	0.03
N-Octane	0.062615	false	0.049087	0.009641	0.037907	0.042124	0.04
N-Pentane	0.158715	true	0.111301	0.067948	0	0.070254	0.11
N-Propylbenzene	0	true	0	0	0	0	0
N-Undecane	0	true	0	0	0	0	0
O-Ethyltoluene	0.097974	true	0.07617	0.035122	0.01988	0.043634	0.08
O-Xylene	0.193504	false	0.149946	0.027332	0.11695	0.129843	0.14
Propane	0.237432	true	0.293898	0.128764	0.141042	0.188811	0.28
Propylene	0.178763	false	0.152699	0.0282	0.116622	0.131919	0.14
Styrene	0.001362	true	0.001771	0.001894	0	0.000262	0.00
Toluene	0.83019	false	0.633774	0.096148	0.522664	0.582219	0.6
Trans-2-Butene	0	true	0	0	0	0	0
Trans-2-Pentene	0.022976	true	0.021732	0.018036	0.00489	0.007518	0.01
Unidentified	9.5062	false	7.517555	1.786855	4.522776	6.786293	7.8
TNMOC	13.615918	false	11.68424	1.614244	8.898436	10.466381	12.

```
In [35]: # Plot the results of the bootstrap runs for a specific factor, showing the variability in percentage and concentration for each feature of the specified factor.
factor_i = 1
bs.plot_results(factor=factor_i)
```





Error Estimation - Bootstrap-Displacement

The bootstrap-displacement (BS-DISP) method is a combination of the bootstrap and displacement methods. An existing BS instance can be used, or a new BS instance will be created.

The BS-DISP method runs a BS instance and for each model in the BS run, DISP is run for each of the features specified or all by default. The results are similar to the DISP results by are aggregated across all bootstrap runs.

```
In [36]: from src.error.bs_disp import BSDISP
```

```
In [37]: # Bootstrap input parameters (Shared with BS-DISP)
model_selected = nmf_models.best_model
nmf_model = nmf_models.results[model_selected]
feature_labels = data_handler.features

bootstrap_n = 10
block_size = data_handler.optimal_block
threshold = 0.6
seed = seed

# Displacement input parameters (Shared with BS-DISP)
threshold_dQ = 0.1
max_search = 50
features = None
```

```
In [38]: # Initialize BS-DISP with the BS and DISP parameters
bsdisp = BSDISP(nmf=nmf_model, feature_labels=feature_labels, model_selected=model_selected, bootstrap_n=bootstrap_n, block_size=block_size, threshold=threshold, max_search=max_search,
```

```
In [39]: %%time
# Execute the BS-DISP instance, which will first run BS (if required) then will run DISP for each BS model.
# parallel = False (CPU times: total: 1h 21min 25s, Wall time: 10min 13s)
bsdisp.run(parallel=True)
```

18-Jan-24 07:54:03 - Running new Bootstrap instance with 10 runs and block size 4
 Bootstrap resampling, training and mapping: 100%|██████████| 10/10 [00:04<00:00, 2.05it/s]
 18-Jan-24 07:54:08 - Bootstrap model created for BS-DISP instance

18-Jan-24 10:12:46 - Completed all BS-DISP calculations, BS runs: 10, Features: 41, Factors: 6, Runtime: 2:18:38.402242 hr:min:sec
 CPU times: total: 19.9 s
 Wall time: 2h 18min 43s

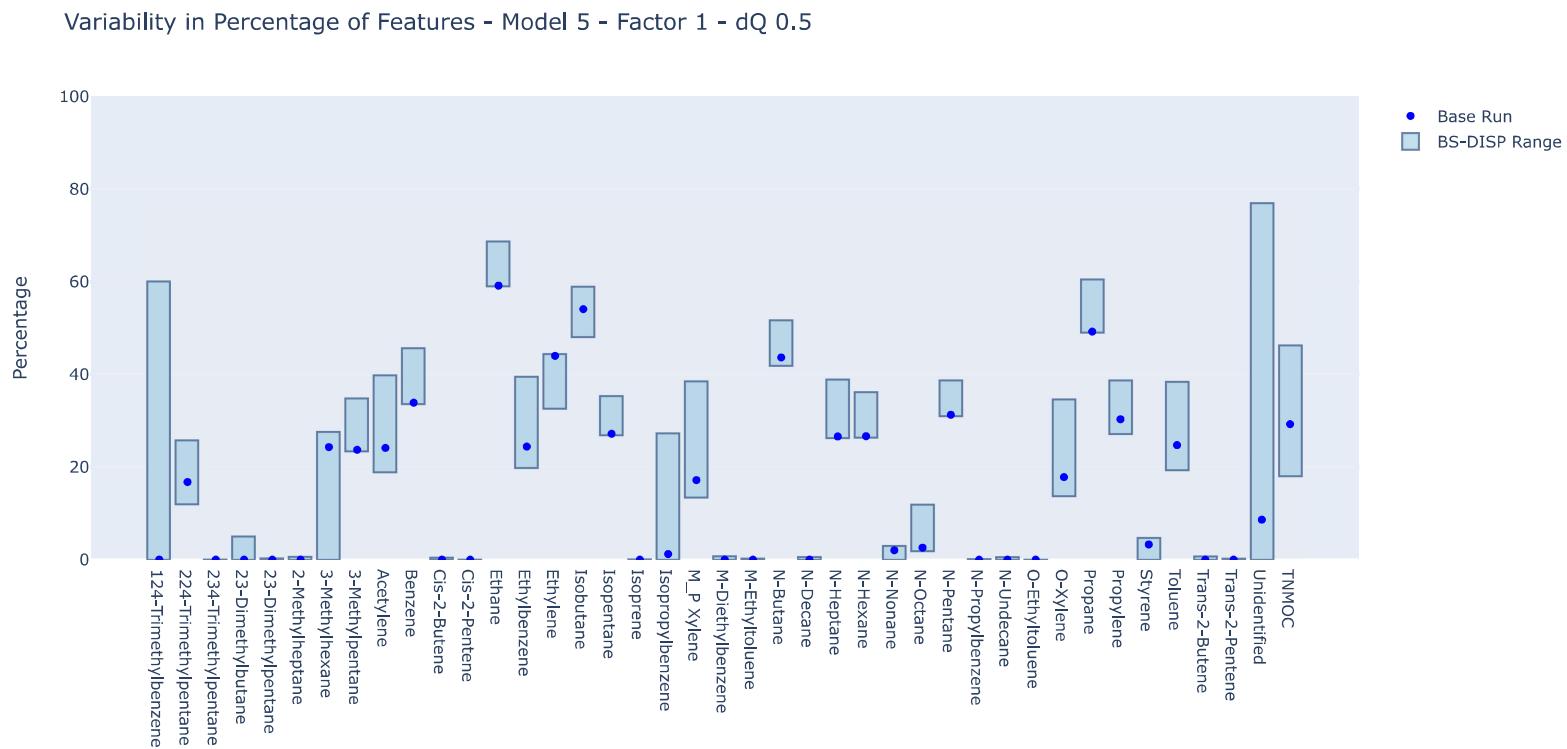
```
In [40]: # Print the summary table and general metrics
bsdisp.summary()
```

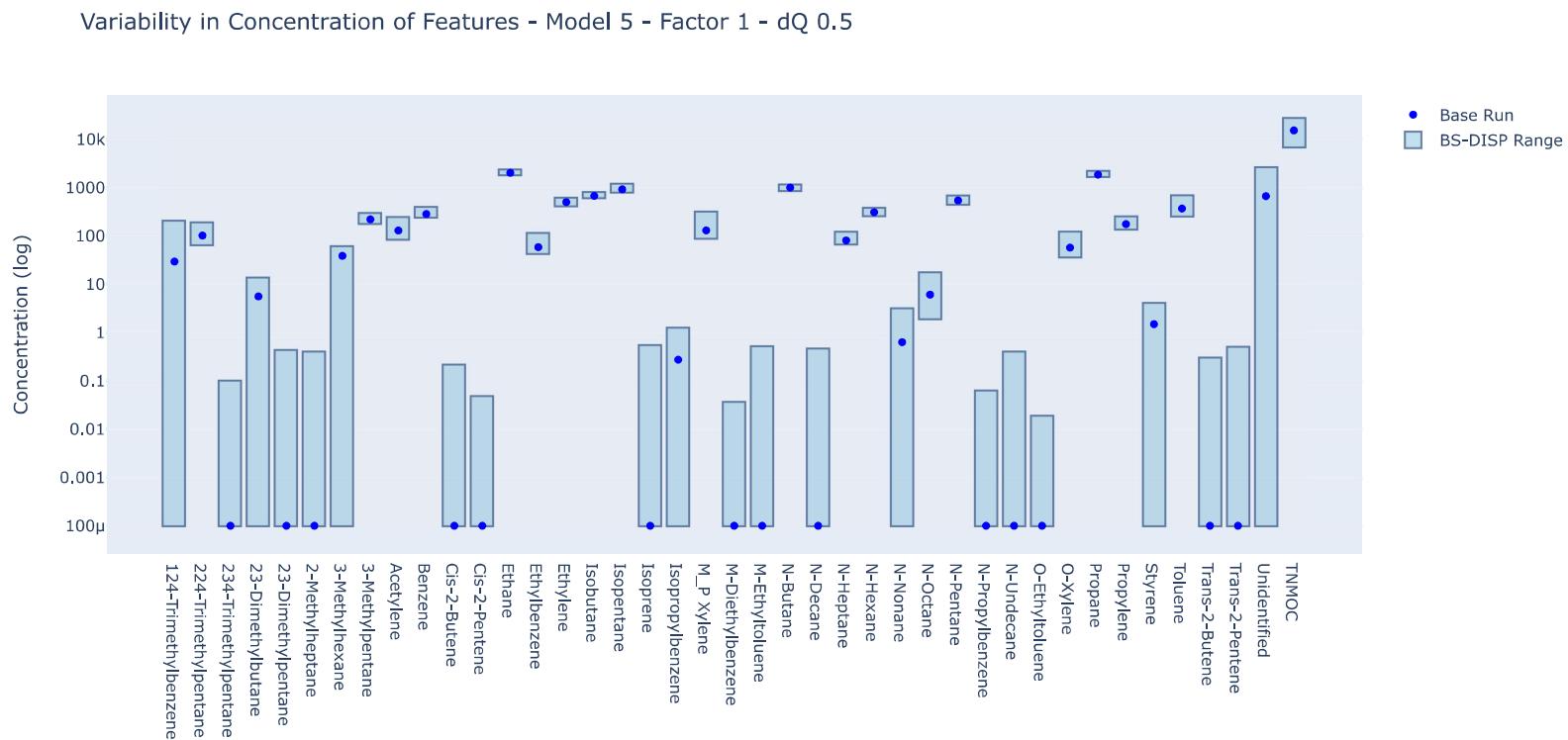
18-Jan-24 10:12:46 - Largest dQ decrease: -0.51, # of Q drops: 1, # of best fit swaps: 4, # of DISP swaps: 0

Swap %

dQ Max	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
0.5	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
4	0	0	0	0	0	0

```
In [41]: # Plot the BS-DISP results, the profile and contribution boxplots for a specific factor.
factor_i = 1
bsdisp.plot_results(factor=factor_i)
```

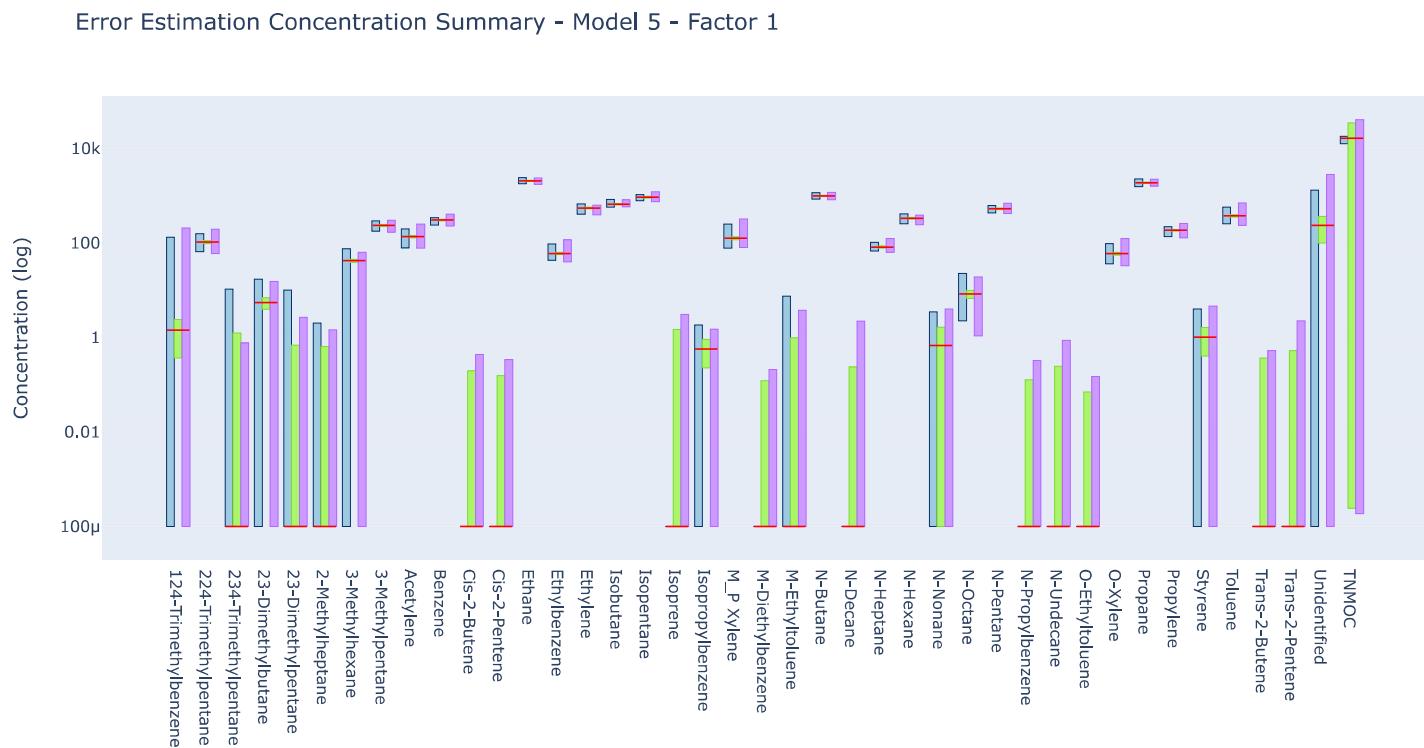




```
In [42]: # The overall error summary can be shown through the follow method that will take in an existing bootstrap and displacement object and plot the error estimation for a given factor.
from src.error.error import Error
```

```
In [43]: # Pass in the previously completed bootstrap and displacement instances
error = Error(bs=bs, disp=disp, bsdisp=bsdisp)
```

```
In [44]: # Plot the error estimation of the concentration for a specified factor.
factor_i = 1
error.plot_summary(factor=factor_i)
```



Rotational Tools - Fpeak

Placeholder for fpeak summary and algorithm.

```
In [13]: # import the module
from src.rotational.fpeak import Fpeak
```



```
In [14]: # initialize the fpeak instance and set the fpeak values
# fpeak_list = [1.0, -1.0, 1.5, -1.5, 2.5, -2.5, 5.0, -5.0]
fpeak_list = [1.0, -1.0, 1.5, -1.5]
s = 0.1      # The softness parameter, setting lowercase s sets all values of the S array to s.
S = None     # An array of size N, that is the softness values corresponding to each nth auxiliary equation (sample in the dataset). Used when all values of S are not the same.

fp = Fpeak(base_model=nmf_model, data_handler=data_handler, fpeaks=fpeak_list, s=s, S=S)
```



```
In [15]: # Run the Fpeak instance
max_iterations = 10000
converge_delta = 1e-4
converge_n = 20

fp.run(max_iter=max_iterations, converge_delta=converge_delta, converge_n=converge_n)
```

```

W Update - Q(robust): 54623.3327, Q(main): 64106.843, Q(aux): 133.8011:  3%|| | 315/10000 [00:00<00:27, 358.50it/s]
H Update - Q(robust): 54544.46, Q(main): 63941.1:  4%|| | 377/10000 [00:00<00:24, 389.84it/s]
W Update - Q(robust): 54403.3389, Q(main): 64124.2616, Q(aux): 203.175:  5%|| | 473/10000 [00:01<00:24, 389.61it/s]
H Update - Q(robust): 54411.11, Q(main): 63909.21:  2%|| | 203/10000 [00:00<00:22, 426.74it/s]
W Update - Q(robust): 54704.7221, Q(main): 64280.9606, Q(aux): 236.4304:  3%|| | 287/10000 [00:00<00:24, 391.80it/s]
H Update - Q(robust): 54599.7, Q(main): 63994.67:  4%|| | 436/10000 [00:01<00:24, 396.94it/s]
W Update - Q(robust): 54413.9916, Q(main): 64333.3745, Q(aux): 364.9943:  6%|| | 593/10000 [00:01<00:25, 361.98it/s]
H Update - Q(robust): 54419.81, Q(main): 63948.57:  4%|| | 398/10000 [00:01<00:26, 369.11it/s]

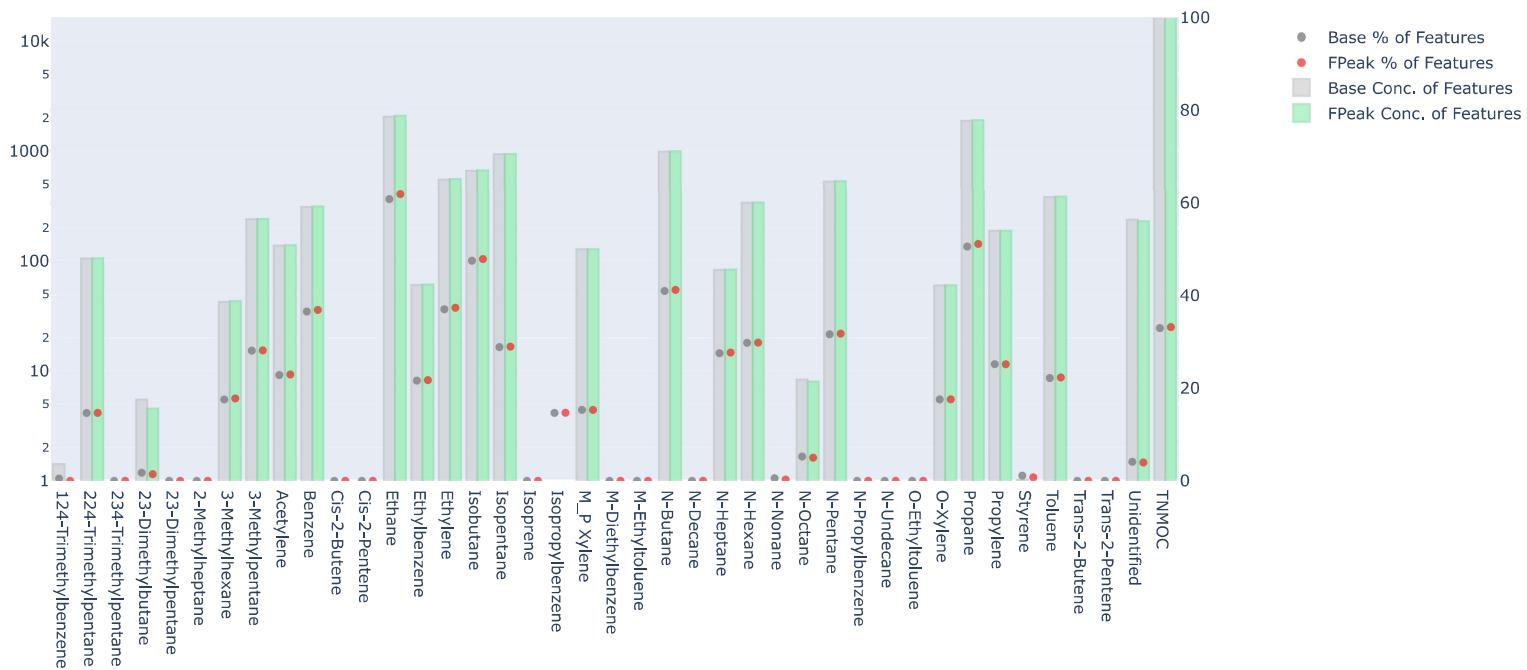
```

In [16]: # View the tabled results of the fpeak runs (as shown in PMF5 - Fpeak Run Summary)
`fp.results_df`

	Strength	dQ(Robust)	Q(Robust)	% dQ(Robust)	Q(Aux)	Q(True)	Converged
0	1.0	120.62	54544.46	0.22	133.80	63941.10	True
1	-1.0	-12.73	54411.11	-0.02	203.18	63909.21	True
2	1.5	175.86	54599.70	0.32	236.43	63994.67	True
3	-1.5	-4.03	54419.81	-0.01	364.99	63948.57	True

In [17]: # Plot the Profiles/Contributions of a specific Fpeak value and factor
`factor_idx = 1`
`fpeak = 1.0`
`fp.plot_profile_contributions(factor_idx=factor_idx, fpeak=fpeak)`

Fpeak Factor Profile - FP=1.0 - Factor 1

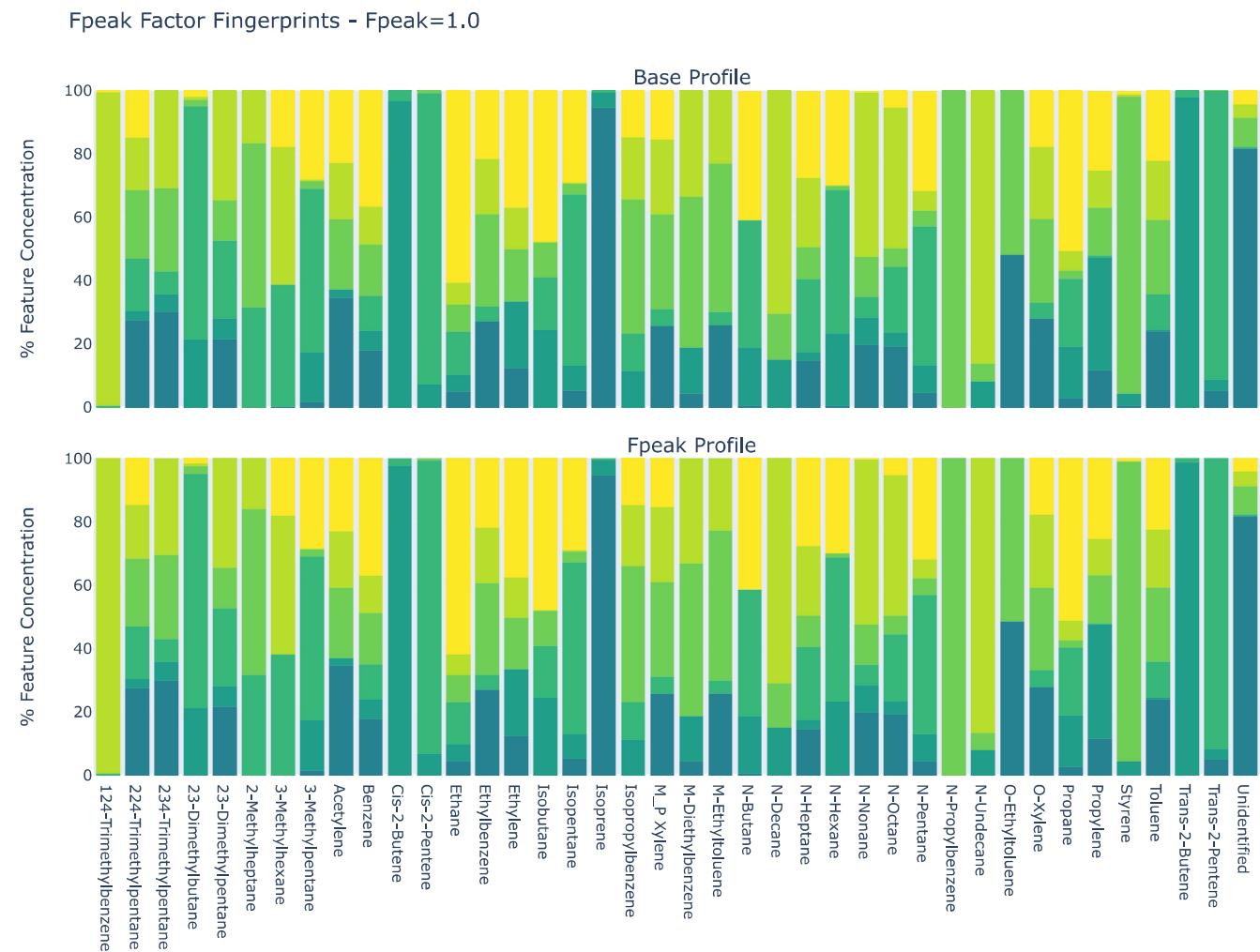


Peak Factor Contributions - Fpeak=1.0 - Factor Factor 1



```
In [18]: # Plot the Factor profiles/fingerprints
fpeak = 1.0

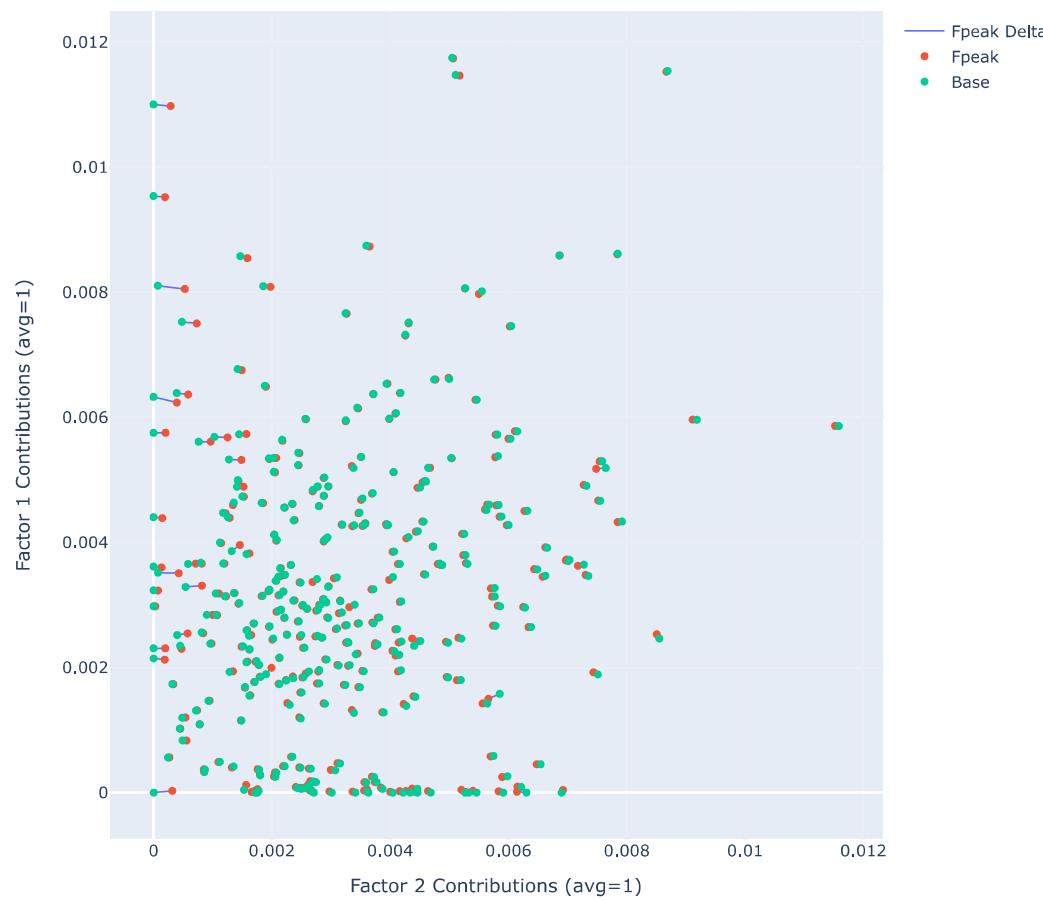
fp.plot_factor_fingerprints(fpeak=fpeak)
```



```
In [19]: # Plot the Factor G-Space graph
fpeak = 1.0
factor_idx1 = 1
factor_idx2 = 2
show_base = True
show_delta = True

fp.plot_g_space(fpeak=fpeak, factor_idx1=factor_idx1, factor_idx2=factor_idx2, show_base=show_base, show_delta=show_delta)
```

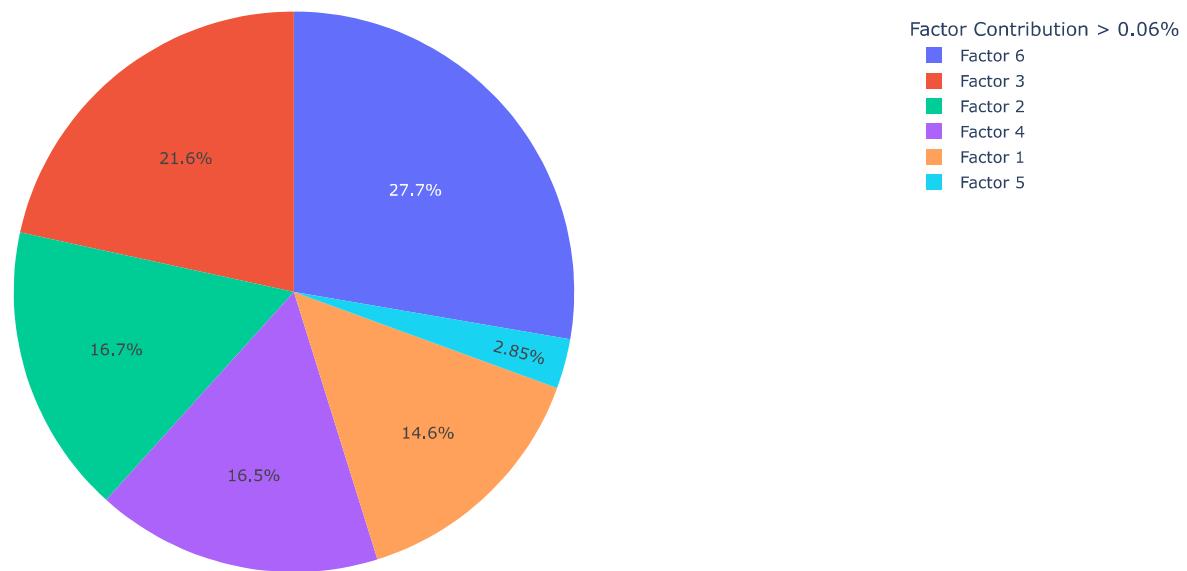
Fpeak G-Space Plot - Fpeak=1.0



```
In [20]: # Plot the factor contributions
fpeak = 1.0
feature_idx = 2
threshold = 0.06

fp.plot_factor_contributions(fpeak=fpeak, feature_idx=feature_idx, threshold=threshold)
```

Factor Contributions to Feature: 224-Trimethylpentane - Fpeak=1.0





```
In [21]: # Run Bootstrap on the Fpeak results
bootstrap_n = 20
block_size = data_handler.optimal_block
threshold = 0.6
bs_seed = nmf_model.seed

fp.run_bs(bootstrap_n=bootstrap_n, block_size=block_size, threshold=threshold, seed=bs_seed)
```

```
Bootstrap resampling, training and mapping: 100%|██████████| 20/20 [04:03<00:00, 12.19s/it]
Bootstrap resampling, training and mapping: 100%|██████████| 20/20 [03:39<00:00, 10.95s/it]
Bootstrap resampling, training and mapping: 100%|██████████| 20/20 [04:06<00:00, 12.33s/it]
Bootstrap resampling, training and mapping: 100%|██████████| 20/20 [03:23<00:00, 10.19s/it]
```

```
In [22]: # Display results of BS of FPeak
fpeak = 1.0
factor_idx = 1
fp.display_bs_results(fpeak=fpeak, factor_idx=factor_idx)
```

```
NMF Bootstrap Error Estimation Summary
----- Input Parameters -----
Base model run number: 1.0
Number of bootstrap runs: 20
Min. Correlation R-Value: 0.6
Number of Factors: 6
```

Mapping of bootstrap factors to base factors

Boot Factors	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Base Factor 5	Base Factor 6	Unmapped
Boot Factor 1	20	0	0	0	0	0	0
Boot Factor 2	0	19	1	0	0	0	0
Boot Factor 3	1	2	13	0	0	4	0
Boot Factor 4	0	0	0	19	0	1	0
Boot Factor 5	0	0	0	1	19	0	0
Boot Factor 6	0	0	5	0	0	15	0

Q(Robust) Percentile Report

Base	Min	25th	Median	75th	Max
54544	49905	51113	52356	53777	55553

Bootstrap run uncertainty statistics - Factor 1

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.000024	true	0.015386	0.035242	0	0	0.0
224-Trimethylpentane	0.212493	true	0.207125	0.038575	0.147389	0.173107	0.2
234-Trimethylpentane	0	true	0.005184	0.010154	0	0	0.00
23-Dimethylbutane	0.009172	true	0.014741	0.021182	0	0	0.00
23-Dimethylpentane	0	true	0.002669	0.005559	0	0	0.00
2-Methylheptane	0	true	0.002243	0.005356	0	0	0.00
3-Methylhexane	0.08676	true	0.096229	0.046067	0.022952	0.061598	0.10
3-Methylpentane	0.482889	true	0.512019	0.078117	0.391297	0.453461	0.51
Acetylene	0.279006	true	0.273767	0.053412	0.195691	0.23495	0.21
Benzene	0.62862	true	0.653538	0.055041	0.595229	0.611709	0.64
Cis-2-Butene	0	true	0	0	0	0	0.0
Cis-2-Pentene	0	true	0	0	0	0	0.0
Ethane	4.25556	false	4.567324	0.457938	3.98578	4.339195	4.6
Ethylbenzene	0.122714	true	0.114745	0.022708	0.08438	0.093803	0.1
Ethylene	1.128775	false	1.281506	0.195569	0.991327	1.166409	1.24
Isobutane	1.355398	false	1.567131	0.199013	1.320319	1.470426	1.51
Isopentane	1.910661	false	2.002286	0.182122	1.713373	1.930883	1.96
Isoprene	0	true	0.001619	0.002946	0	0	0.0
Isopropylbenzene	0.00116	true	0.001589	0.001209	0	0.000665	0.00
M_P_Xylene	0.25679	true	0.229285	0.073175	0.138943	0.169096	0.21
M-Diethylbenzene	0	true	0	0	0	0	0.0
M-Ethyltoluene	0	true	0.001363	0.002429	0	0	0.0
N-Butane	2.021155	false	2.238926	0.279234	1.996275	2.050863	2.18
N-Decane	0	true	0.000496	0.00159	0	0	0.0
N-Heptane	0.16791	true	0.172173	0.016306	0.145881	0.159794	0.17
N-Hexane	0.682958	true	0.722783	0.098262	0.559135	0.661612	0.72
N-Nonane	0.000652	true	0.004862	0.00637	0	0	0.0
N-Octane	0.016145	true	0.021374	0.010795	0.008393	0.013737	0.01
N-Pentane	1.079524	false	1.156234	0.143487	0.926237	1.111777	1.11
N-Propylbenzene	0	true	0	0	0	0	0.0
N-Undecane	0	true	0.000205	0.000633	0	0	0.0
O-Ethyltoluene	0	true	0	0	0	0	0.0
O-Xylene	0.120892	true	0.118067	0.024934	0.083099	0.099506	0.11
Propane	3.877604	false	4.24383	0.411826	3.844425	3.992819	4.2
Propylene	0.378105	true	0.391035	0.05913	0.279064	0.348846	0.41
Styrene	0.001386	true	0.001151	0.001319	0	0	0.00
Toluene	0.770797	true	0.728957	0.118022	0.556888	0.63263	0.71
Trans-2-Butene	0	true	0.000093	0.000416	0	0	0.0
Trans-2-Pentene	0	true	0.00037	0.001247	0	0	0.0
Unidentified	0.460446	false	0.272576	0.252891	0	0	0.31
TNMOC	33.646591	true	34.13208	2.376117	30.327497	32.524118	34.3

Bootstrap run uncertainty statistics - Factor 2

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.592619	false	0.52396	0.090401	0.393645	0.477925	0.51
224-Trimethylpentane	0.242887	true	0.237218	0.070644	0.106092	0.194711	0.21
234-Trimethylpentane	0.144199	false	0.129456	0.020793	0.086688	0.118838	0.13
23-Dimethylbutane	0.004817	true	0.031355	0.056329	0	0	0.00
23-Dimethylpentane	0.098375	true	0.092857	0.011693	0.078535	0.085155	0.09
2-Methylheptane	0.02313	true	0.02204	0.021025	0	0.011195	0.01
3-Methylhexane	0.21472	false	0.178661	0.051608	0.090184	0.16227	0.11
3-Methylpentane	0.0055	true	0.038353	0.044992	0	0.000591	0.00
Acetylene	0.216593	true	0.210959	0.071469	0.075114	0.180823	0.21
Benzene	0.203443	true	0.204282	0.066194	0.119328	0.166019	0.19
Cis-2-Butene	0	true	0.000051	0.000229	0	0	0
Cis-2-Pentene	0	true	0	0	0	0	0
Ethane	0.431436	true	0.462819	0.291208	0.144609	0.168414	0.41
Ethylbenzene	0.098372	true	0.102258	0.036527	0.042483	0.075583	0.11
Ethylene	0.390873	true	0.363577	0.133376	0.134502	0.280249	0.38
Isobutane	0.005615	true	0.027909	0.056765	0	0	0.00
Isopentane	0.024624	true	0.096763	0.112428	0	0	0.01
Isoprene	0	true	0	0	0	0	0
Isopropylbenzene	0.001524	true	0.000961	0.000827	0	0.000197	0.00
M_P Xylene	0.398763	true	0.43471	0.149245	0.166938	0.323905	0.44
M-Diethylbenzene	0.004349	false	0.002668	0.002007	0	0.001052	0.00
M-Ethyltoluene	0.107508	true	0.124656	0.038338	0.072498	0.101456	0.11
N-Butane	0	true	0.003931	0.017581	0	0	0
N-Decane	0.156642	false	0.132938	0.033097	0.087214	0.122567	0.12
N-Heptane	0.132677	true	0.126338	0.027022	0.081336	0.10623	0.11
N-Hexane	0.003227	true	0.045715	0.060662	0	0	0.00
N-Nonane	0.126473	false	0.112637	0.018465	0.085641	0.101205	0.11
N-Octane	0.144067	false	0.124925	0.014907	0.095794	0.121787	0.11
N-Pentane	0.205492	true	0.198417	0.08223	0.079453	0.149842	0.20
N-Propylbenzene	0	true	0	0	0	0	0
N-Undecane	0.15775	false	0.133347	0.028274	0.092283	0.119621	0.11
O-Ethyltoluene	0	true	0.005895	0.016003	0	0	0
O-Xylene	0.158633	true	0.154102	0.047112	0.082776	0.128953	0.11
Propane	0.46909	true	0.494194	0.180477	0.209311	0.397419	0.52
Propylene	0.175188	true	0.166067	0.046038	0.09276	0.134693	0.16
Styrene	0.000334	true	0.00352	0.004218	0	0	0.00
Toluene	0.637131	true	0.655362	0.183981	0.389257	0.509366	0.70
Trans-2-Butene	0	true	0.000127	0.00039	0	0	0
Trans-2-Pentene	0	true	0.00226	0.002594	0	0	0.00
Unidentified	0.540739	true	0.665516	0.945472	0	0	0.40
TNMOC	11.374198	true	11.714851	3.4189	7.852111	8.527768	11.2

Bootstrap run uncertainty statistics - Factor 3

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0	true	0.002239	0.010012	0	0	
224-Trimethylpentane	0.313728	true	0.297665	0.049087	0.227111	0.27348	0.3
234-Trimethylpentane	0.1256	true	0.114437	0.02117	0.083636	0.10035	0.12
23-Dimethylbutane	0.015158	true	0.01779	0.030701	0	0	0.00
23-Dimethylpentane	0.036692	true	0.035908	0.014452	0.003899	0.031971	0.01
2-Methylheptane	0.076266	true	0.066291	0.015384	0.035673	0.061638	0.06
3-Methylhexane	0	true	0	0	0	0	
3-Methylpentane	0.040427	true	0.045904	0.037404	0	0	0.01
Acetylene	0.269447	true	0.255212	0.04224	0.19986	0.222728	0.24
Benzene	0.275268	true	0.272965	0.057395	0.195432	0.237613	0.26
Cis-2-Butene	0	true	0.000282	0.000616	0	0	
Cis-2-Pentene	0.001291	true	0.004725	0.007036	0	0.000828	0.00
Ethane	0.602229	true	0.637445	0.322417	0.246657	0.339683	0.64
Ethylbenzene	0.163481	true	0.154135	0.020514	0.117317	0.140989	0.11
Ethylene	0.490254	true	0.460972	0.159966	0.265263	0.339326	0.40
Isobutane	0.311133	false	0.26749	0.077206	0.164149	0.223603	0.26
Isopentane	0.215953	true	0.231267	0.130293	0.044793	0.136122	0.21
Isoprene	0	true	0	0	0	0	
Isopropylbenzene	0.003386	true	0.003457	0.001232	0.001825	0.002498	0.00
M_P Xylene	0.50253	true	0.451426	0.12124	0.21107	0.396973	0.46
M-Diethylbenzene	0.006331	true	0.006459	0.004487	0.001935	0.003924	0.00
M-Ethyltoluene	0.225894	false	0.202464	0.031524	0.163839	0.18277	0.19
N-Butane	0.005907	true	0.064074	0.075004	0	0	0.0
N-Decane	0.031406	true	0.034255	0.017477	0.013331	0.02045	0.01
N-Heptane	0.060975	true	0.061501	0.012531	0.042033	0.056207	0.01
N-Hexane	0.030383	true	0.038717	0.043022	0	0	0.01
N-Nonane	0.030411	true	0.030013	0.008583	0.017365	0.022423	0.01
N-Octane	0.019014	true	0.019787	0.010369	0.003026	0.010267	0.01
N-Pentane	0.176295	true	0.199477	0.054193	0.124266	0.164772	0.19
N-Propylbenzene	0.086864	true	0.081624	0.008395	0.069605	0.077058	0.08
N-Undecane	0.009777	true	0.018779	0.018862	0.001769	0.004346	0.00
O-Ethyltoluene	0.104211	true	0.097869	0.012849	0.07553	0.090426	0.0
O-Xylene	0.180625	true	0.171443	0.026924	0.140018	0.149772	0.11
Propane	0.175055	true	0.200784	0.178416	0	0.057528	0.1
Propylene	0.226619	true	0.227721	0.057671	0.167683	0.183507	0.20
Styrene	0.174806	true	0.157659	0.031049	0.103375	0.130953	0.16
Toluene	0.807965	true	0.782367	0.129763	0.593522	0.680823	0.71
Trans-2-Butene	0	true	0.000433	0.000677	0	0	
Trans-2-Pentene	0	true	0.002242	0.005548	0	0	
Unidentified	1.027174	true	1.434295	1.247723	0	0.376349	1.24
TNMOC	5.016948	true	5.692441	1.706188	3.133807	4.317498	5.9

Bootstrap run uncertainty statistics - Factor 4

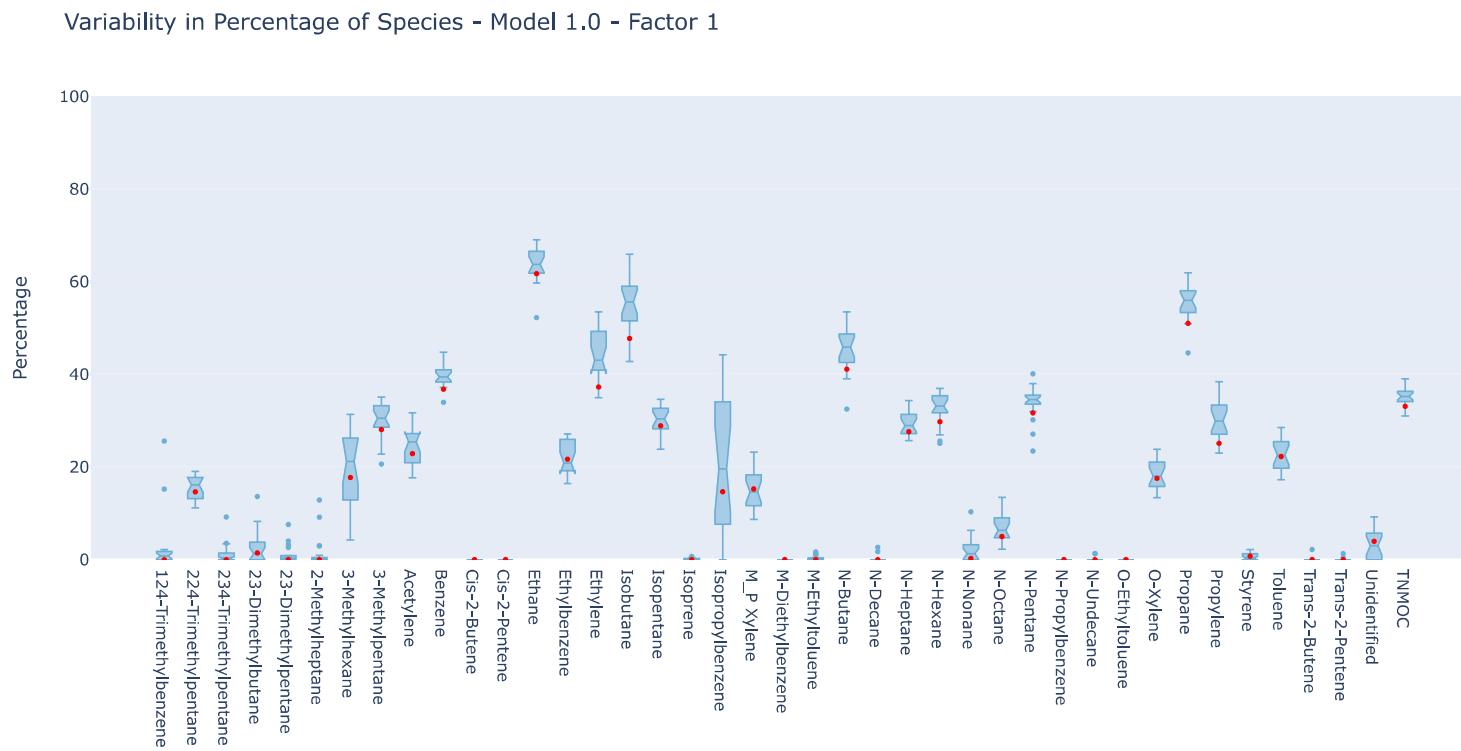
features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0.003973	true	0.004639	0.006398	0	0	0.004639
224-Trimethylpentane	0.239628	false	0.220471	0.062057	0.139481	0.195239	0.200000
234-Trimethylpentane	0.03447	true	0.036414	0.028443	0.011009	0.025023	0.025023
23-Dimethylbutane	0.470238	false	0.407996	0.071263	0.306937	0.355498	0.355498
23-Dimethylpentane	0.070723	true	0.072413	0.009768	0.051888	0.069234	0.070723
2-Methylheptane	0.046605	true	0.044584	0.013504	0.024261	0.039341	0.046605
3-Methylhexane	0.186858	true	0.167643	0.037323	0.10489	0.1376	0.186858
3-Methylpentane	0.886423	true	0.83989	0.141352	0.700602	0.797888	0.886423
Acetylene	0	true	0.014482	0.040531	0	0	0.014482
Benzene	0.187798	true	0.199784	0.0378	0.159093	0.172135	0.199784
Cis-2-Butene	0.002369	true	0.00327	0.002298	0.000014	0.001316	0.002369
Cis-2-Pentene	0.220323	true	0.220618	0.03759	0.169225	0.208138	0.220323
Ethane	0.923016	true	0.902912	0.274699	0.463501	0.778183	0.923016
Ethylbenzene	0.026725	true	0.027815	0.025877	0.009574	0.018563	0.026725
Ethylene	0	true	0	0	0	0	0.000000
Isobutane	0.459032	true	0.549628	0.162692	0.327768	0.438355	0.549628
Isopentane	3.583094	true	3.638076	0.68118	2.884478	3.354442	3.638076
Isoprene	0.007465	true	0.026102	0.031224	0	0.005944	0.007465
Isopropylbenzene	0.000937	true	0.001137	0.00068	0.000049	0.000527	0.000937
M_P_Xylene	0.08804	true	0.106897	0.113351	0.030522	0.070714	0.08804
M-Diethylbenzene	0	true	0	0	0	0	0.000000
M-Ethyltoluene	0.019823	true	0.021613	0.03662	0	0.002916	0.019823
N-Butane	1.956719	true	1.96166	0.353064	1.720275	1.847094	1.956719
N-Decane	0	true	0.001705	0.003601	0	0	0.000000
N-Heptane	0.139127	true	0.13167	0.018142	0.099516	0.126343	0.139127
N-Hexane	1.036852	true	1.004515	0.203212	0.771536	0.88584	1.036852
N-Nonane	0.015476	true	0.018118	0.010635	0.005934	0.010783	0.015476
N-Octane	0.067513	true	0.064025	0.01461	0.049299	0.05266	0.067513
N-Pentane	1.491938	true	1.501038	0.268348	1.201413	1.373684	1.491938
N-Propylbenzene	0	true	0.000489	0.000548	0	0	0.000489
N-Undecane	0	true	0	0	0	0	0.000000
O-Ethyltoluene	0	true	0	0	0	0	0.000000
O-Xylene	0.034873	true	0.038623	0.032861	0.016402	0.023565	0.034873
Propane	1.628622	true	1.6201	0.33759	1.235456	1.469863	1.628622
Propylene	0.005259	false	0.041278	0.029652	0	0.019385	0.005259
Styrene	0	true	0.00122	0.004395	0	0	0.000000
Toluene	0.386207	true	0.379609	0.130256	0.255778	0.312518	0.386207
Trans-2-Butene	0.001122	true	0.001714	0.001724	0	0.000697	0.001122
Trans-2-Pentene	0.393985	true	0.390443	0.060648	0.328784	0.379085	0.393985
Unidentified	0.075829	true	0.848826	1.901053	0	0	0.100000
TNMOC	26.12743	true	25.244376	2.627568	20.366958	24.449945	25.300000

Bootstrap run uncertainty statistics - Factor 5

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS Median	BS 75th	BS Max
124-Trimethylbenzene	0	true	0.010447	0.019925	0	0	0	0	0
224-Trimethylpentane	0.04145	true	0.058617	0.04736	0	0.024797	0.04145	0.06291	0.06291
234-Trimethylpentane	0.027659	true	0.030956	0.015993	0.01039	0.020405	0.020405	0.020405	0.020405
23-Dimethylbutane	0.137434	false	0.119284	0.043324	0.073775	0.095883	0.101086	0.101086	0.101086
23-Dimethylpentane	0.018912	true	0.016915	0.00957	0.002873	0.010869	0.010869	0.010869	0.010869
2-Methylheptane	0	true	0.008247	0.011655	0	0	0	0	0
3-Methylhexane	0	true	0.01179	0.026774	0	0	0	0	0
3-Methylpentane	0.273831	false	0.207286	0.117306	0.064425	0.114802	0.114802	0.114802	0.114802
Acetylene	0.031265	false	0.017245	0.017798	0	0	0	0	0
Benzene	0.105939	false	0.046818	0.037247	0.006822	0.012284	0.012284	0.012284	0.012284
Cis-2-Butene	0.10706	false	0.073444	0.020235	0.050364	0.058098	0.058098	0.058098	0.058098
Cis-2-Pentene	0.016279	false	0.013616	0.019981	0	0	0	0	0
Ethane	0.357463	false	0.116455	0.182227	0	0	0	0	0
Ethylbenzene	0	true	0.002821	0.005951	0	0	0	0	0
Ethylene	0.638139	false	0.435268	0.209944	0.195636	0.239591	0.239591	0.239591	0.239591
Isobutane	0.706885	false	0.391898	0.161444	0.214078	0.297337	0.297337	0.297337	0.297337
Isopentane	0.519438	false	0.321536	0.383792	0	0.02093	0.02093	0.02093	0.02093
Isoprene	0.073959	false	0.031242	0.036099	0.000929	0.001973	0.001973	0.001973	0.001973
Isopropylbenzene	0.000911	true	0.000732	0.001053	0	0.000067	0.000067	0.000067	0.000067
M_P_Xylene	0	true	0.003618	0.009448	0	0	0	0	0
M-Diethylbenzene	0.001905	true	0.006033	0.010444	0	0.000278	0.000278	0.000278	0.000278
M-Ethyltoluene	0	true	0.006554	0.009209	0	0	0	0	0
N-Butane	0.894487	false	0.514224	0.27268	0.126351	0.342797	0.342797	0.342797	0.342797
N-Decane	0.033808	true	0.027136	0.010882	0.002421	0.021945	0.021945	0.021945	0.021945
N-Heptane	0.016893	true	0.016544	0.0145	0	0.005993	0.005993	0.005993	0.005993
N-Hexane	0.525691	false	0.363029	0.149786	0.159882	0.249023	0.249023	0.249023	0.249023
N-Nonane	0.021248	true	0.01877	0.01071	0	0.011519	0.011519	0.011519	0.011519
N-Octane	0.013599	true	0.020498	0.01414	0.003976	0.008173	0.008173	0.008173	0.008173
N-Pentane	0.29226	false	0.173747	0.167729	0	0.033364	0.033364	0.033364	0.033364
N-Propylbenzene	0	true	0.000485	0.00098	0	0	0	0	0
N-Undecane	0.014972	true	0.013416	0.010467	0.000496	0.004932	0.004932	0.004932	0.004932
O-Ethyltoluene	0	true	0.000697	0.001823	0	0	0	0	0
O-Xylene	0	true	0.002457	0.004213	0	0	0	0	0
Propane	1.224312	false	0.716968	0.32529	0.387172	0.503038	0.503038	0.503038	0.503038
Propylene	0.543582	false	0.320909	0.095804	0.185506	0.226921	0.226921	0.226921	0.226921
Styrene	0.007883	true	0.013843	0.025156	0	0.000636	0.000636	0.000636	0.000636
Toluene	0.029688	true	0.029186	0.043743	0	0	0	0	0
Trans-2-Butene	0.104585	false	0.070017	0.017333	0.04628	0.057495	0.057495	0.057495	0.057495
Trans-2-Pentene	0.014416	false	0.015012	0.030547	0	0	0	0	0
Unidentified	0	true	0	0	0	0	0	0	0
TNMOC	11.834067	false	7.615524	2,964927	3.328892	5.84621	5.84621	5.84621	7.29

Bootstrap run uncertainty statistics - Factor 6

features	Base Run Profile	Within IQR	BS Mean	BS Std. Dev.	BS 5th	BS 25th	BS N
124-Trimethylbenzene	0	true	0.012469	0.0496	0	0	
224-Trimethylpentane	0.402652	false	0.315684	0.033632	0.270532	0.297708	0.3
234-Trimethylpentane	0.143406	false	0.109445	0.013641	0.088169	0.098185	0.1
23-Dimethylbutane	0	true	0.007682	0.027779	0	0	
23-Dimethylpentane	0.062418	false	0.040198	0.01372	0.023318	0.03171	0.04
2-Methylheptane	0	true	0.0015	0.002204	0	0	
3-Methylhexane	0.000631	true	0.037915	0.047799	0	0	0.01
3-Methylpentane	0.030058	true	0.055193	0.049447	0.0001	0.021588	0.04
Acetylene	0.4222	false	0.341006	0.033289	0.294206	0.319246	0.32
Benzene	0.307184	false	0.270139	0.033701	0.229453	0.245821	0.21
Cis-2-Butene	0	true	0.000071	0.000235	0	0	
Cis-2-Pentene	0	true	0.001369	0.003167	0	0	
Ethane	0.322156	false	0.457956	0.16769	0.195899	0.336004	0.4
Ethylbenzene	0.154486	false	0.122666	0.013524	0.106374	0.113749	0.12
Ethylene	0.380532	false	0.335804	0.085315	0.248046	0.271099	0.32
Isobutane	0	true	0.029398	0.06571	0	0	
Isopentane	0.358602	true	0.392288	0.112274	0.24189	0.297305	0.39
Isoprene	1.53091	false	1.313916	0.265107	0.901316	1.081781	1.38
Isopropylbenzene	0	true	0.000015	0.000347	0	0	
M_P Xylene	0.436064	false	0.310563	0.112764	0.102163	0.283522	0.34
M-Diethylbenzene	0.000059	true	0.0000555	0.0000592	0	0	0.00
M-Ethyltoluene	0.123823	false	0.074651	0.048352	0.012478	0.019124	0.09
N-Butane	0.037928	false	0.147611	0.095636	0.054799	0.086816	0.12
N-Decane	0	true	0.001878	0.003368	0	0	
N-Heptane	0.090343	false	0.07898	0.017691	0.054623	0.064816	0.07
N-Hexane	0.015096	true	0.040241	0.047983	0	0	0.01
N-Nonane	0.048484	false	0.030987	0.012155	0.008624	0.0265	0.02
N-Octane	0.06304	false	0.048367	0.012803	0.030882	0.041344	0.04
N-Pentane	0.160291	true	0.181809	0.061555	0.094457	0.140297	0.11
N-Propylbenzene	0	true	0	0	0	0	
N-Undecane	0	true	0	0	0	0	
O-Ethyltoluene	0.098679	true	0.074729	0.036189	0.006859	0.074102	0.08
O-Xylene	0.193694	false	0.152715	0.018927	0.131654	0.138135	0.11
Propane	0.225112	false	0.342642	0.171685	0.12802	0.266109	0.32
Propylene	0.17752	false	0.149226	0.034558	0.103795	0.123912	0.11
Styrene	0.000582	true	0.001013	0.001467	0	0	
Toluene	0.831276	false	0.665462	0.073801	0.563878	0.626786	0.66
Trans-2-Butene	0	true	0	0	0	0	
Trans-2-Pentene	0.022223	true	0.024125	0.016377	0.008937	0.011027	0.02
Unidentified	9.540117	false	6.604383	2.125003	3.631221	4.886609	7.01
TNMOC	13.629334	false	12.449045	1.631769	9.960256	11.193229	12.2



Variability in Concentration of Species - Model 1.0 - Factor 1

