Mapping Chemical Performance Space

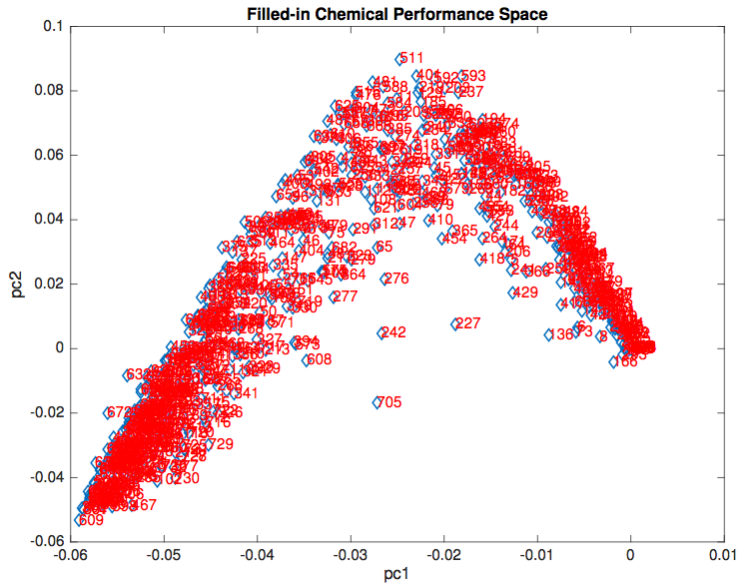
Maleniel

University of Bath

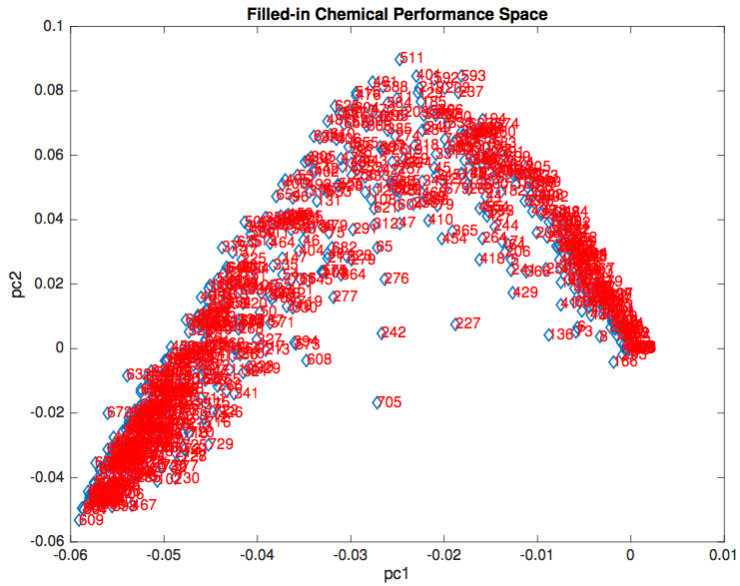
February 3, 2017

Mapping Chemical Performance Space

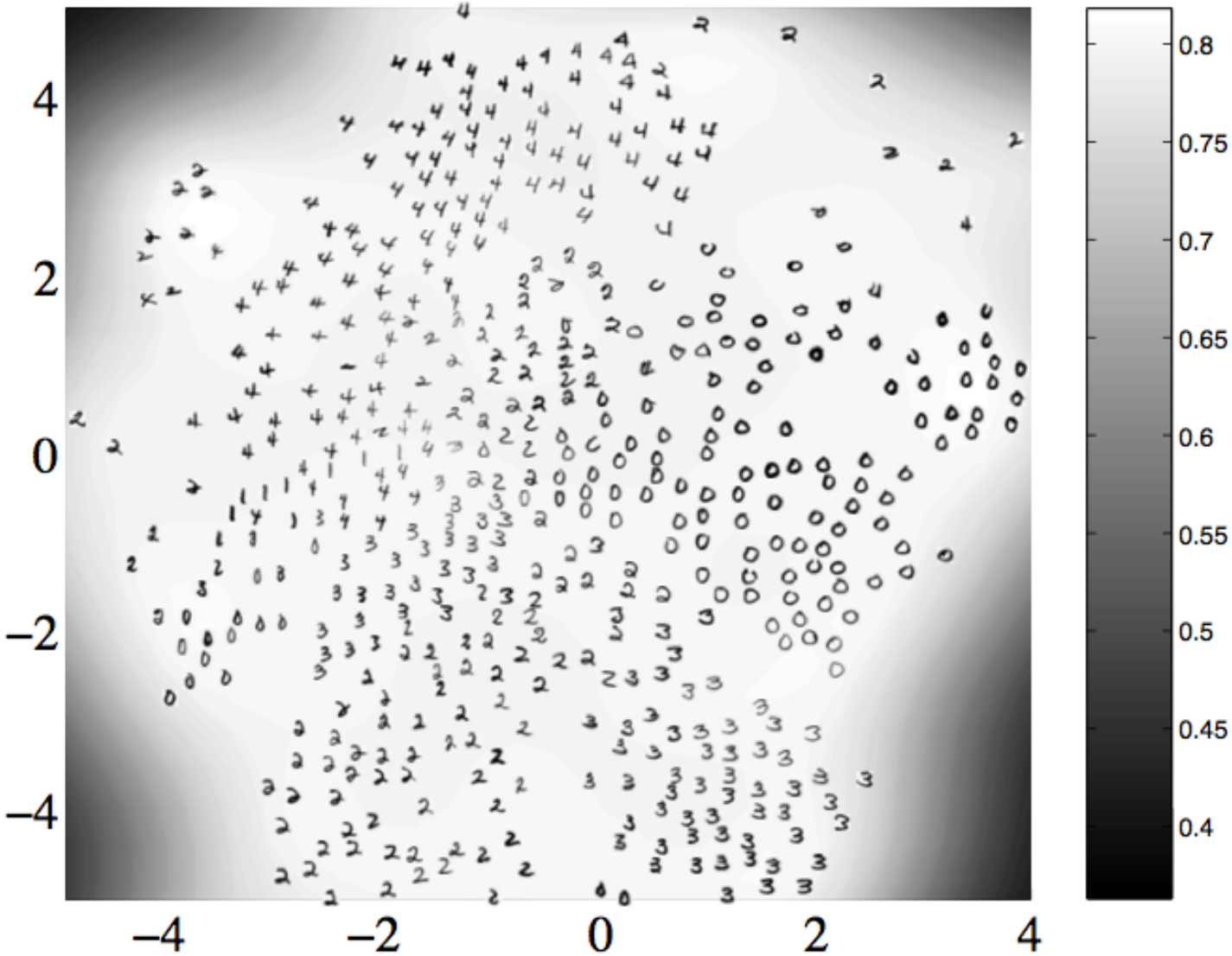
Chemical	Screen P1	Screen Q1	Screen R1	Screen P2
Standard				
Best				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				



Chemical	Screen P1	Screen Q1	Screen R1	Screen P2
Standard				
Best				
1				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
20				
21				
22				

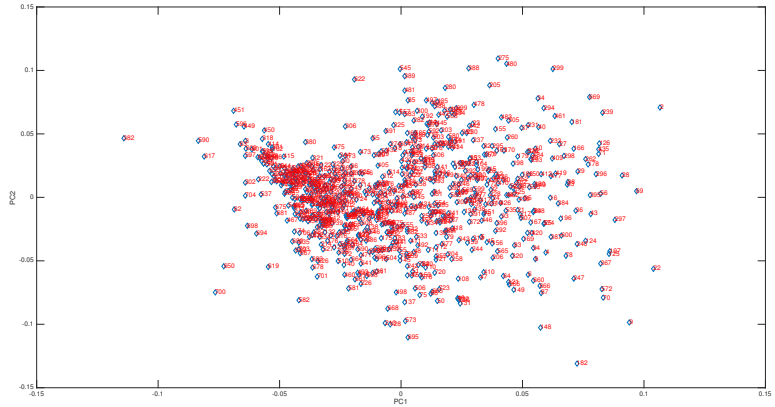


Latent variable model

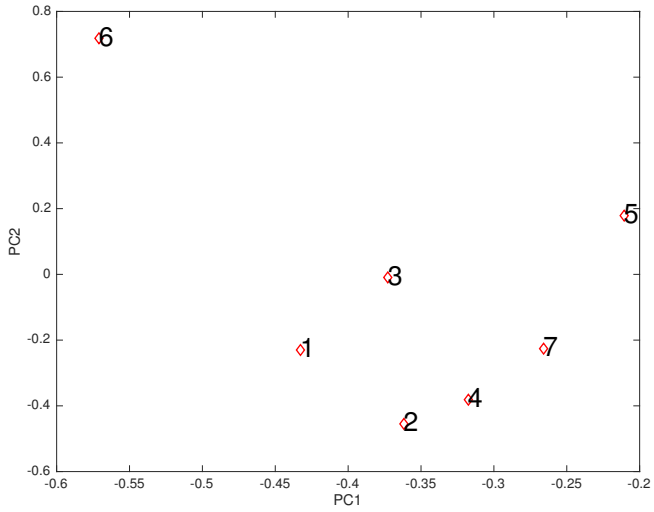


N.D. Lawrence. *Gaussian Process Latent Variable Models for Visulaisation of High Dimensional Data*.

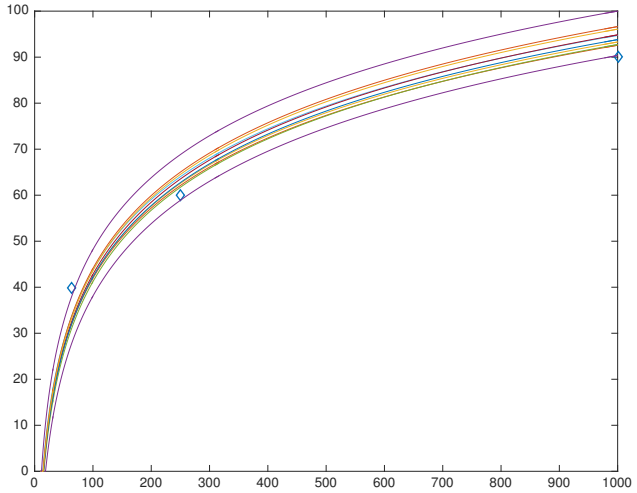
- Chemical performance space (latent space) Y

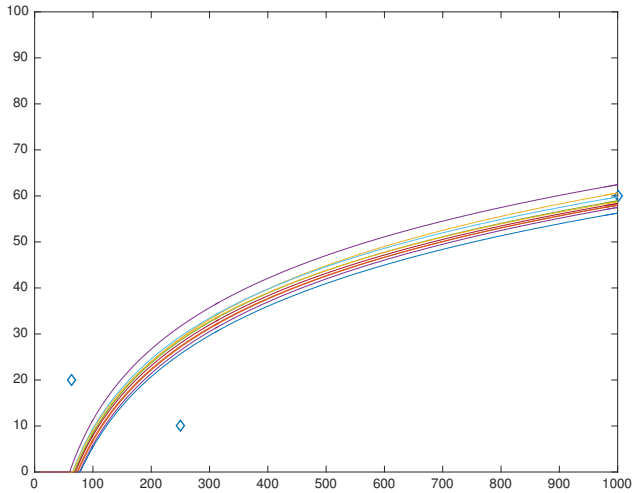


- Chemical performance space (latent space) Y
  - ↓ linear map W



- Chemical performance space (latent space) Y
  - ↓ linear map W
- Manisfest Space X=WY





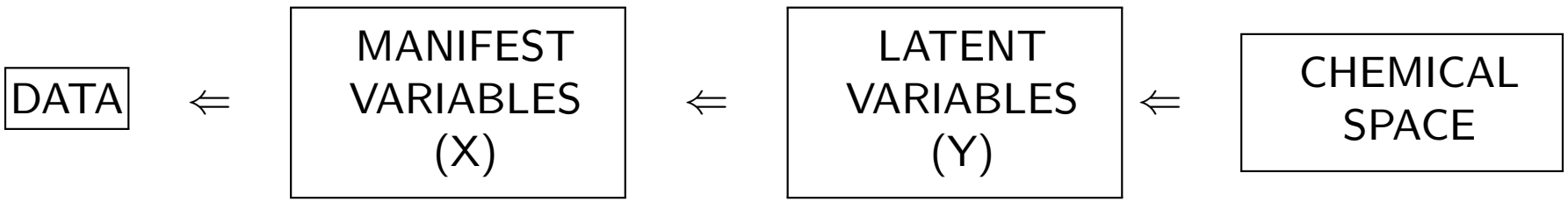
- Chemical performance space (latent space) Y
  - ↓ linear map W
- Manisfest Space X=WY
  - ↓ deterministic function
- Data Space

0	110	21	43	0	9	103
58	90	90	90	100	90	90
60	100	94	92	100	66	99
84	90	90	100	100	90	100
91	97	103	100	100	103	100
101	94	99	100	100	99	100

- Chemical performance space (latent space) Y
  - ↓ linear map W
- Manisfest Space X=WY
  - ↓ deterministic function
- Data Space
  - ↓ noise and missing data
- Real data

NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	90	90	90	100	90	90
NaN	NaN	NaN	NaN	100	NaN	NaN
NaN	90	90	100	100	90	100
NaN	NaN	NaN	100	100	NaN	100
NaN	NaN	NaN	100	100	NaN	100

Future Improvements



- **1)** Use non-linear maps from Y to X, e.g. Gaussian Process Latent Variable Model.
- **2)**Incorporate chemical structure information into the model.
- Let y_i be the latent $1 \times k$ vector representing chemical i (ith row of matrix Y).
- how about saying

$$cov(y_i, y_j) \propto e^{-\|F_i - F_j\|}$$

where F_i is the chemical feature vector.

Future Improvements

- **3)** Missing data not at random.
- Reasons that data is missing.
- If data is missing because scientists think the chemical will perform poorly: