

Contents

1	Introduction	1
2	The matrix completion problem	5
2.1	Fundamentals	5
2.1.1	Definition	5
2.1.2	Fields of application	6
2.2	Matrix completion methods (MCMs)	7
2.2.1	Low-rank matrix completion.....	7
2.2.2	MCM based on iterative principal component analysis (PCA).....	9
2.2.3	Bayesian MCM	12
3	Illustration of characteristics of a low-rank MCM using synthetic test matrices.....	15
3.1	Introduction	15
3.2	Generation of test matrices.....	16
3.2.1	Complete matrices	16
3.2.2	Incomplete training matrices	18
3.3	Results and discussion.....	18
3.3.1	Low-rank assumption.....	19
3.3.2	Noise suppression.....	20
3.3.3	Influence of the number of observed entries	23
3.4	Conclusions	24
4	Prediction of limiting activity coefficients at 298.15 K	27
4.1	Introduction	27
4.2	Experimental data basis.....	28
4.3	Predictions obtained by matrix completion.....	30
4.3.1	Results from the Bayesian MCM	30
4.3.2	Results from the MCM based on iterative PCA.....	32
4.3.3	Comparison of both MCMs.....	33

4.4	Comparison with modified UNIFAC (Dortmund).....	35
4.5	Conclusions	39
5	Prediction of limiting activity coefficients for varying temperature	41
5.1	Introduction	41
5.2	Experimental data basis.....	42
5.3	Matrix completion for isothermal data at varying temperature.....	43
5.3.1	Results for the largest dataset at 303.15 K	45
5.3.2	Comparison of the results for different temperatures.....	47
5.4	Introduction of temperature dependency using physical knowledge	50
5.4.1	Description of the approach	51
5.4.2	Training dataset	53
5.4.3	Evaluation procedure.....	54
5.4.4	Results of the matrix completion approach.....	55
5.4.5	Comparison with modified UNIFAC (Dortmund).....	56
5.5	Conclusions	58
6	Conclusions	61
	Literature	65
	Appendix	73
A	Implementation of the Stan model	74
B	Additional results	75
B.1	Isothermal datasets	75
B.2	Temperature dependent dataset	89
C	Alternative approach for the temperature dependency.....	90