Supplementary Information

for

Accelerating Multicomponent Phase-Coexistence Calculations with Physics-informed Neural Networks

Satyen Dhamankar,[†,‡] Shengli Jiang,[†,‡] and Michael A. Webb[*,†]

†*Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA*

‡*These authors contributed equally to this work.*

E-mail: mawebb@princeton.edu

Contents

S1 Additional optimized phase diagrams

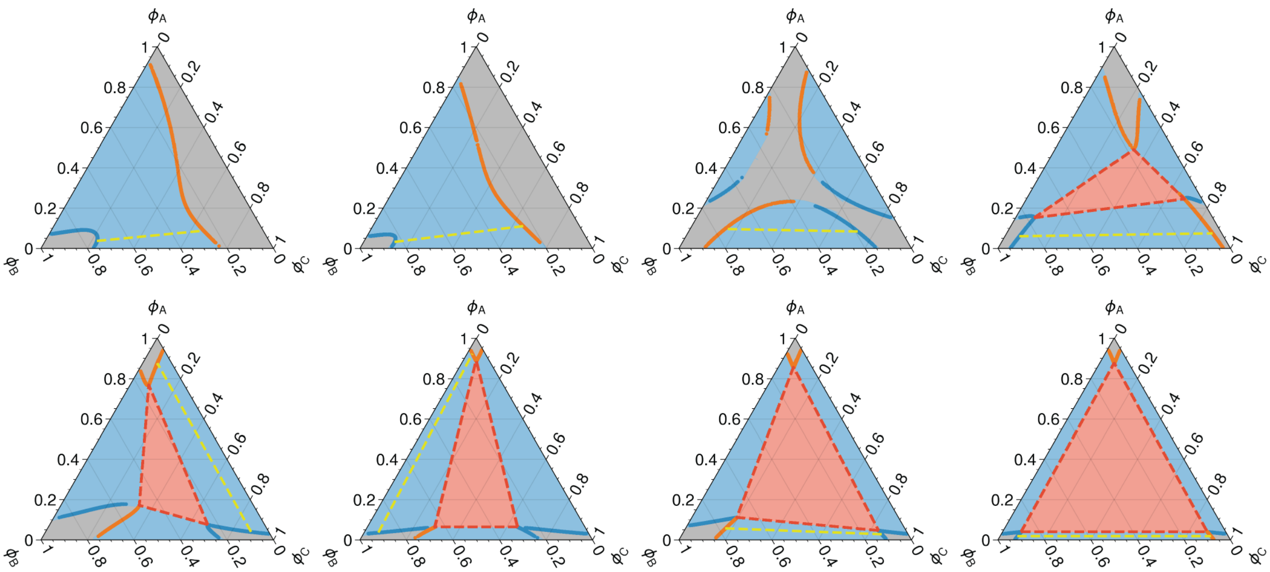


Figure S1: **Additional coexistence curves produced with the post-ML optimization strategy.** The results are obtained using the PI model inference trained with full training data to warm-start Newton-CG optimization. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. The system parameters $[\chi_{\mathrm{AB}}, \chi_{\mathrm{BC}}, \chi_{\mathrm{AC}}, v_{\mathrm{A}}, v_{\mathrm{B}}, v_{\mathrm{C}}]$, for the top row, from left to right are $[1.0, 2.0, 1.5, 1.5, 1.5, 1.5]$, $[1.0, 2.0, 2.0, 2.0, 1.5, 1.0]$, $[1.5, 1.5, 1.5, 1.5, 2.0, 1.5]$, and $[1.5, 1.5, 2.0, 1.5, 1.5, 2.0]$. For the bottom row, the parameters from left to right are $[2.0, 1.5, 1.5, 2.0, 1.5, 1.5]$, $[2.0, 2.0, 1.5, 2.0, 1.5, 1.5]$, $[2.0, 1.5, 1.5, 2.0, 1.5, 2.0]$, and $[2.0, 2.0, 2.0, 2.0, 1.5, 1.5]$.

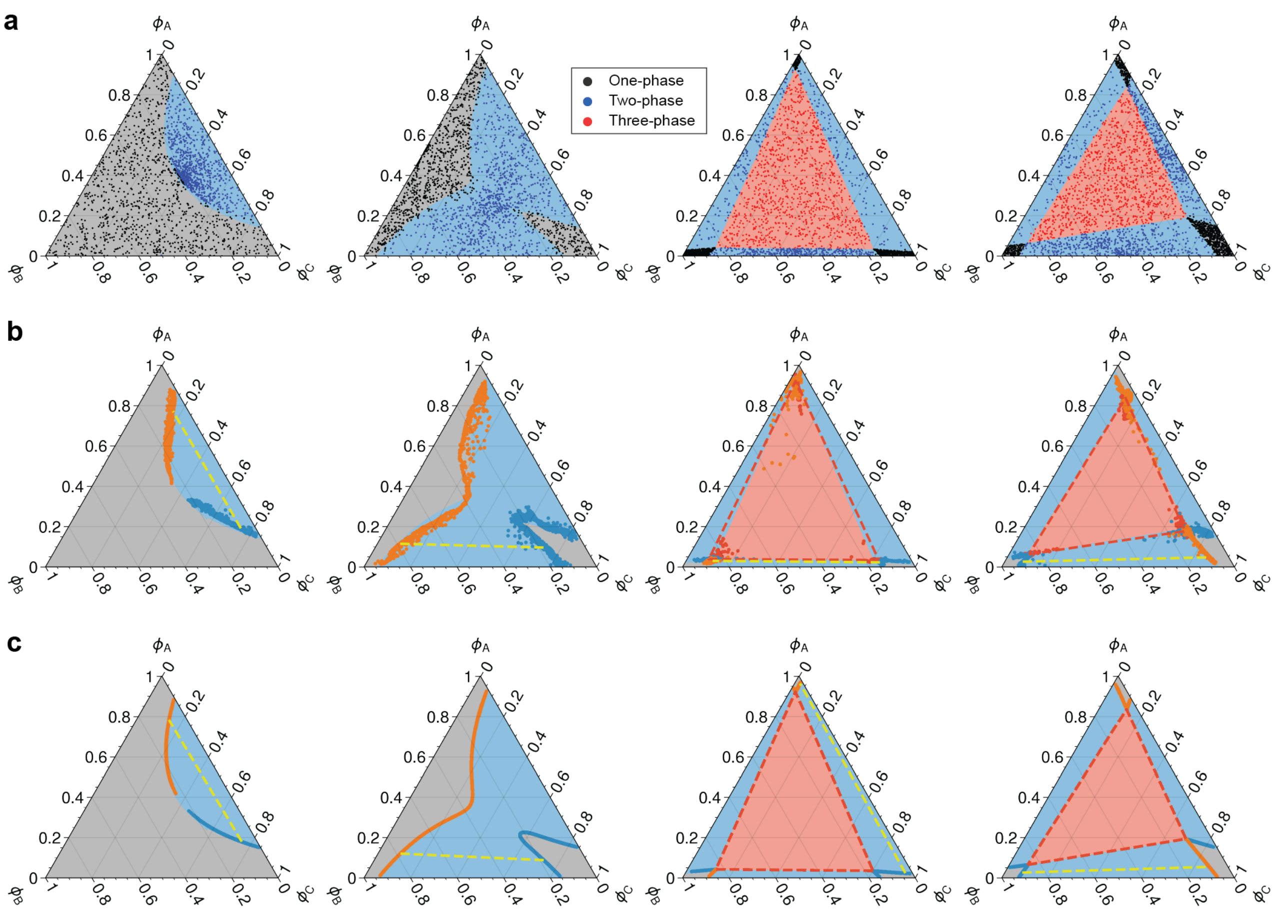


Figure S2: **Baseline model performance in phase-coexistence prediction with 10% of training data**. **a)** Classification of the number of coexisting phases. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Colors in the legend denote the types of predicted splits. **b)** Predicted coexistence curves. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. **c)** Coexistence curves produced with the post-ML optimization strategy. The results are obtained using ML inference to warm-start Newton-CG optimization.

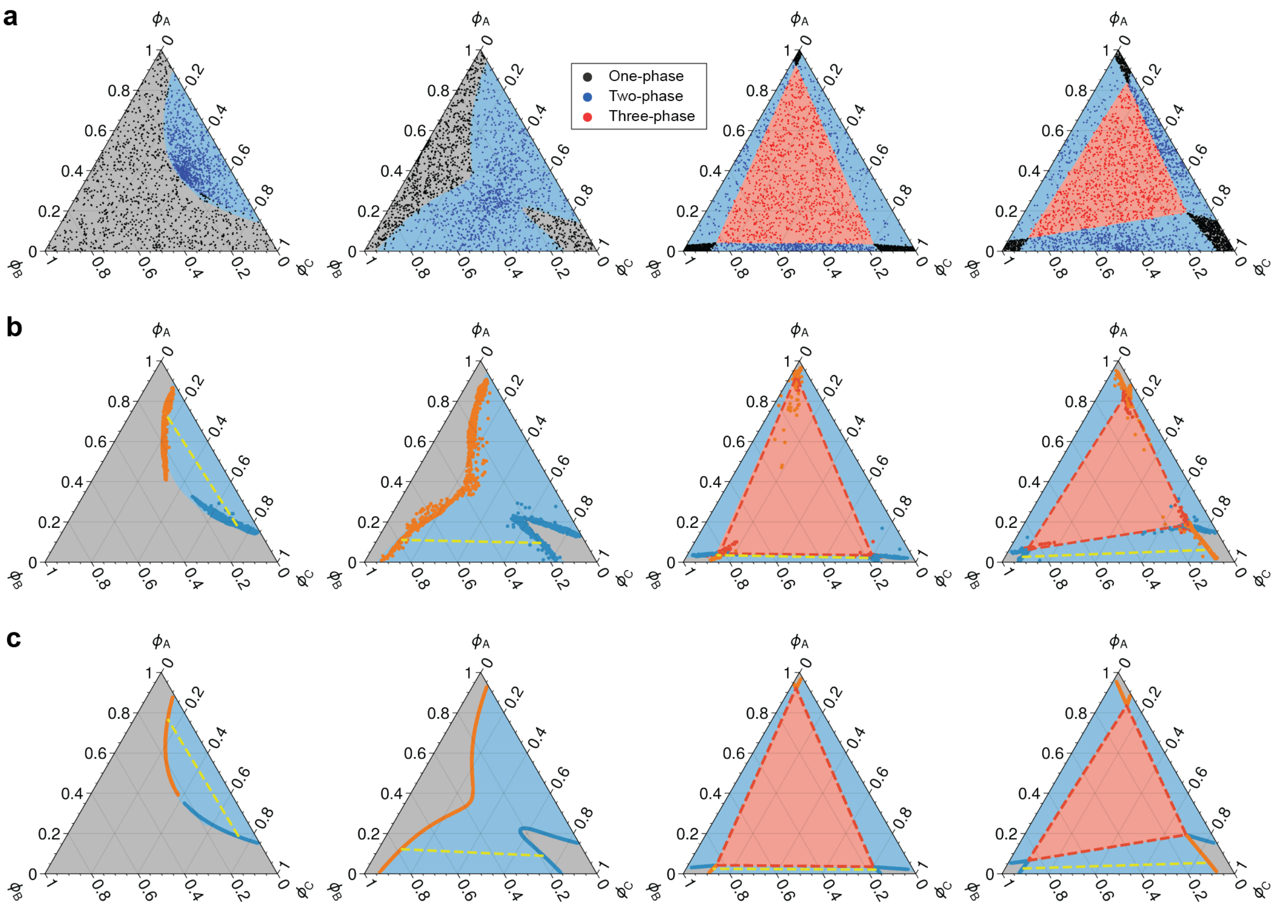


Figure S3: **Baseline model performance in phase-coexistence prediction with full training data**. **a)** Classification of the number of coexisting phases. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Colors in the legend denote the types of predicted splits. **b)** Predicted coexistence curves. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. **c)** Coexistence curves produced with the post-ML optimization strategy. The results are obtained using ML inference to warm-start Newton-CG optimization.

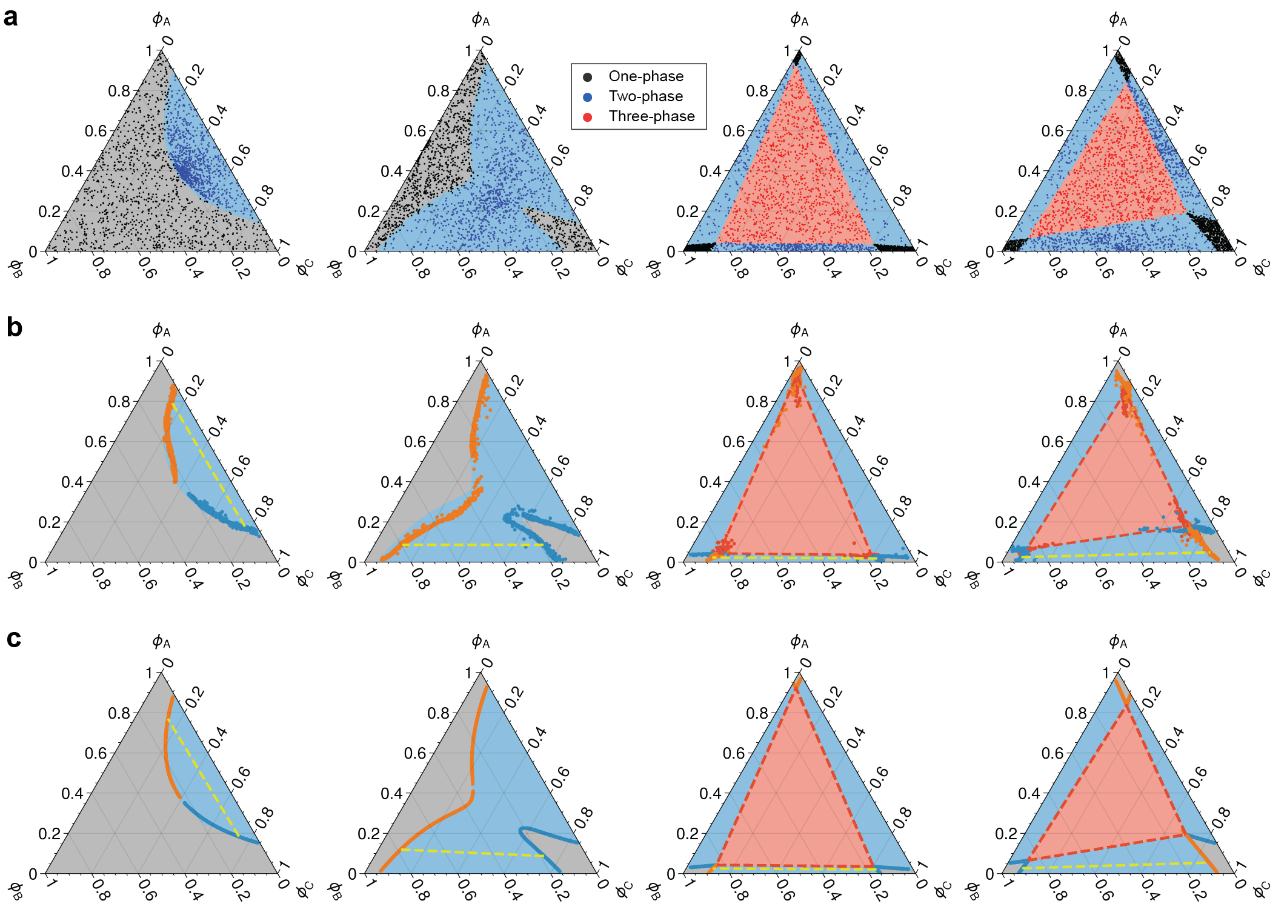


Figure S4: **PI model performance in phase-coexistence prediction with 10% of training data**. **a)** Classification of the number of coexisting phases. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Colors in the legend denote the types of predicted splits. **b)** Predicted coexistence curves. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. **c)** Coexistence curves produced with the post-ML optimization strategy. The results are obtained using ML inference to warm-start Newton-CG optimization.

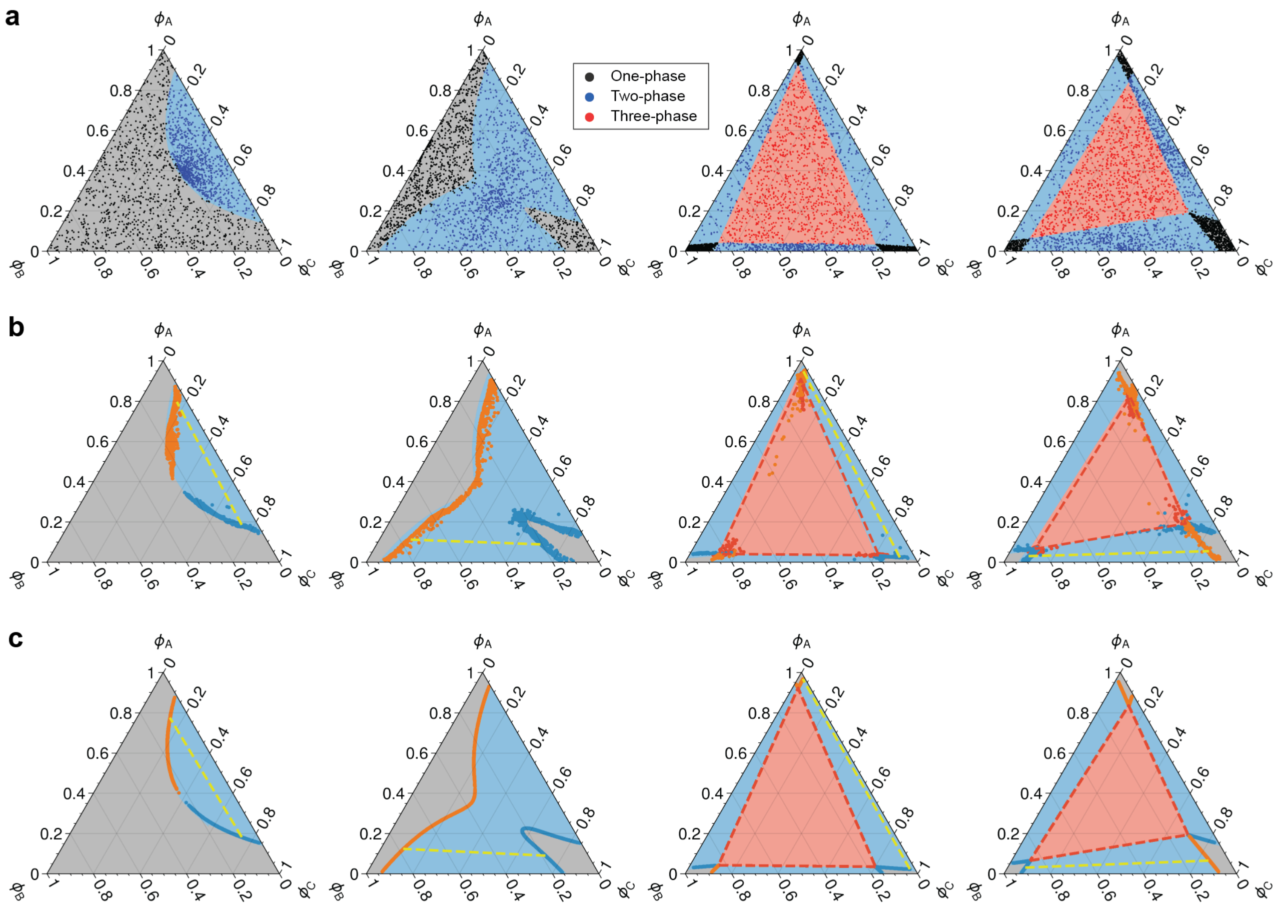


Figure S5: **PI+ model performance in phase-coexistence prediction with 10% of training data**. **a)** Classification of the number of coexisting phases. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Colors in the legend denote the types of predicted splits. **b)** Predicted coexistence curves. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. **c)** Coexistence curves produced with the post-ML optimization strategy. The results are obtained using ML inference to warm-start Newton-CG optimization.

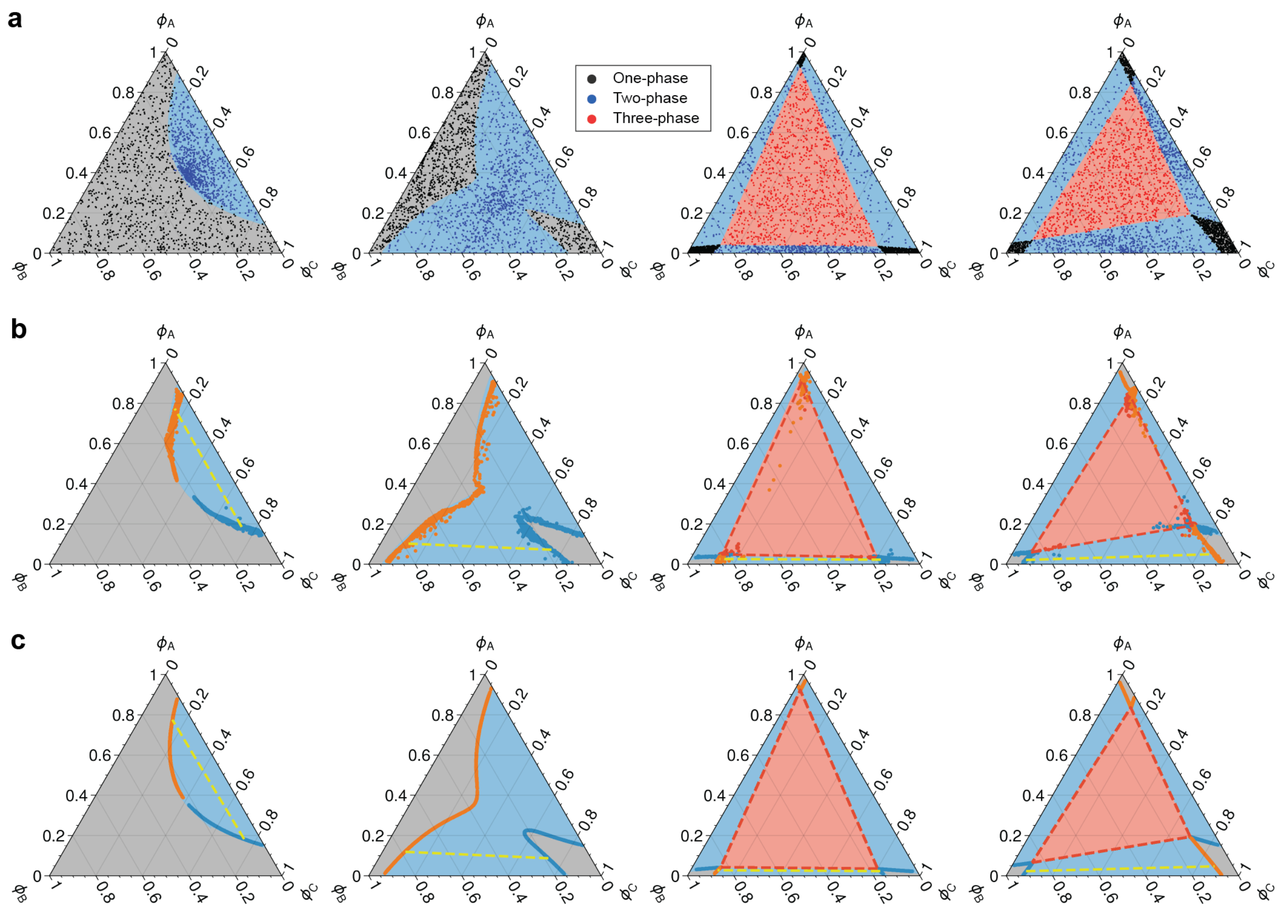


Figure S6: **PI+ model performance in phase-coexistence prediction with full training data**. **a)** Classification of the number of coexisting phases. The background color in all phase diagrams denotes the true phase: gray (one-phase), blue (two-phase), and red (three-phase). The scatter points indicate the predicted phase splits for a given initial composition. Colors in the legend denote the types of predicted splits. **b)** Predicted coexistence curves. Blue and orange scatter points indicate two-phase coexistence curves, with the yellow dashed line denoting an example tie line. The vertices of the red triangle indicate three-phase coexistence points. **c)** Coexistence curves produced with the post-ML optimization strategy. The results are obtained using ML inference to warm-start Newton-CG optimization.

Table S1: Parameters for representative systems depicted in main text figures.

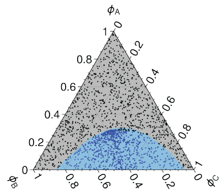
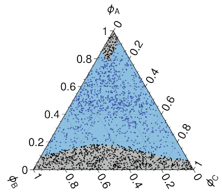
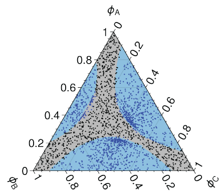
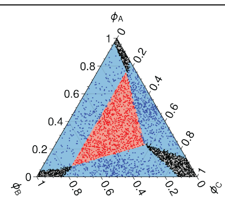
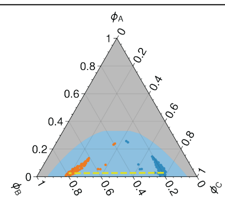
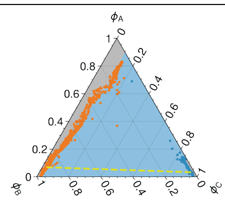
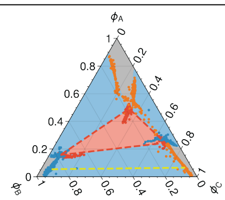
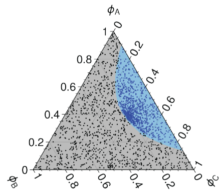
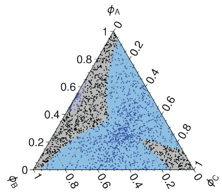
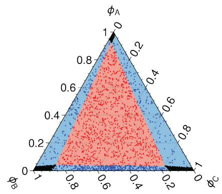
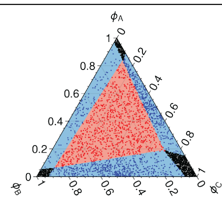
Location	Image	χ_{AB}	χ_{BC}	χ_{AC}	v_A	v_B	v_C
Fig. 3(a) leftmost		1.0	1.0	2.0	1.5	1.0	2.0
Fig. 3(a) center-left		1.5	1.5	1.0	2.0	2.0	1.5
Fig. 3(a) center-right		1.5	2.0	2.0	1.5	1.5	1.0
Fig. 3(a) rightmost		1.531	1.869	1.477	1.739	1.497	1.975
Fig. 5(a) top-left		1.0	1.0	2.0	1.0	1.5	1.5
Fig. 5(a) center-left		1.5	1.0	2.0	1.5	2.0	2.0
Fig. 5(a) bottom-left		1.5	1.5	2.0	1.5	1.5	2.0

Table S2: Parameters for representative systems depicted in main text figures (continued).

Location	Image	χ_{AB}	χ_{BC}	χ_{AC}	v_A	v_B	v_C
Fig. 6(a) leftmost		1.5	1.0	1.0	1.5	2.0	2.0
Fig. 6(a) center-left		2.0	2.0	2.0	1.0	2.0	1.0
Fig. 6(a) center-right		2.0	2.0	1.5	2.0	2.0	1.5
Fig. 6(a) rightmost		1.5	2.0	1.5	1.5	2.0	2.0

S2 Phase classification confusion matrices

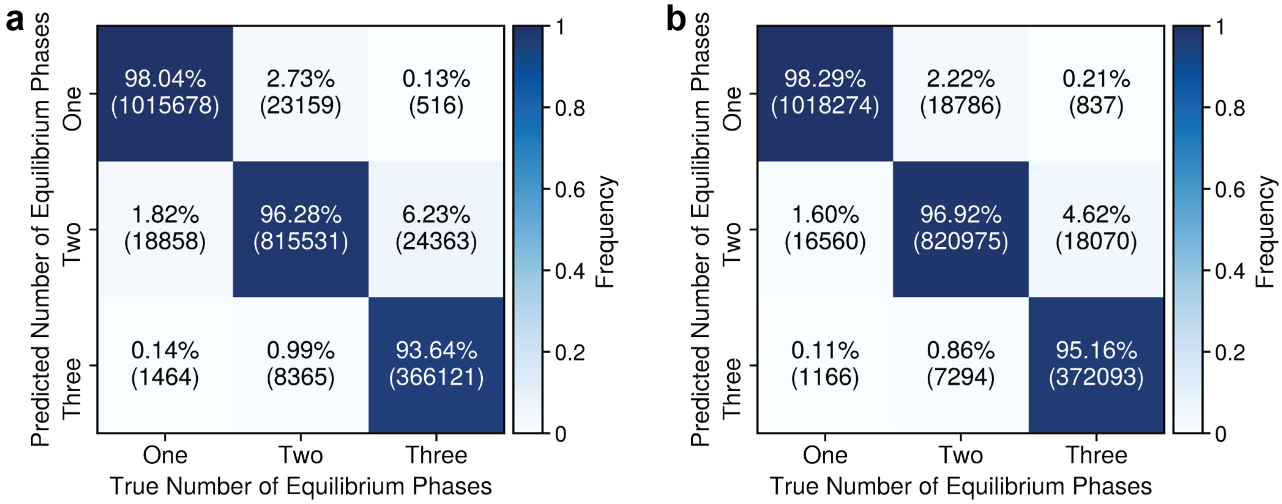


Figure S7: **Confusion matrices for the predicted number of equilibrium phases using the baseline model with a)** 10% of training data, and **b)** with full training data. Diagonal entries represent correctly classified instances, while off-diagonal entries represent misclassifications.

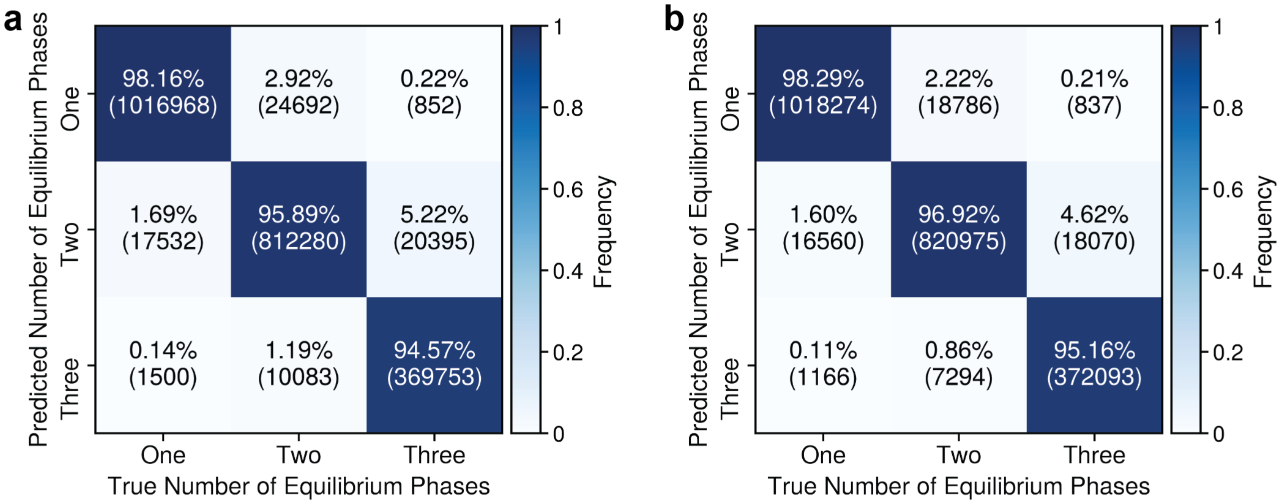


Figure S8: **Confusion matrices for the predicted number of equilibrium phases using the PI model with a)** 10% of training data, and **b)** with full training data. Diagonal entries represent correctly classified instances, while off-diagonal entries represent misclassifications.

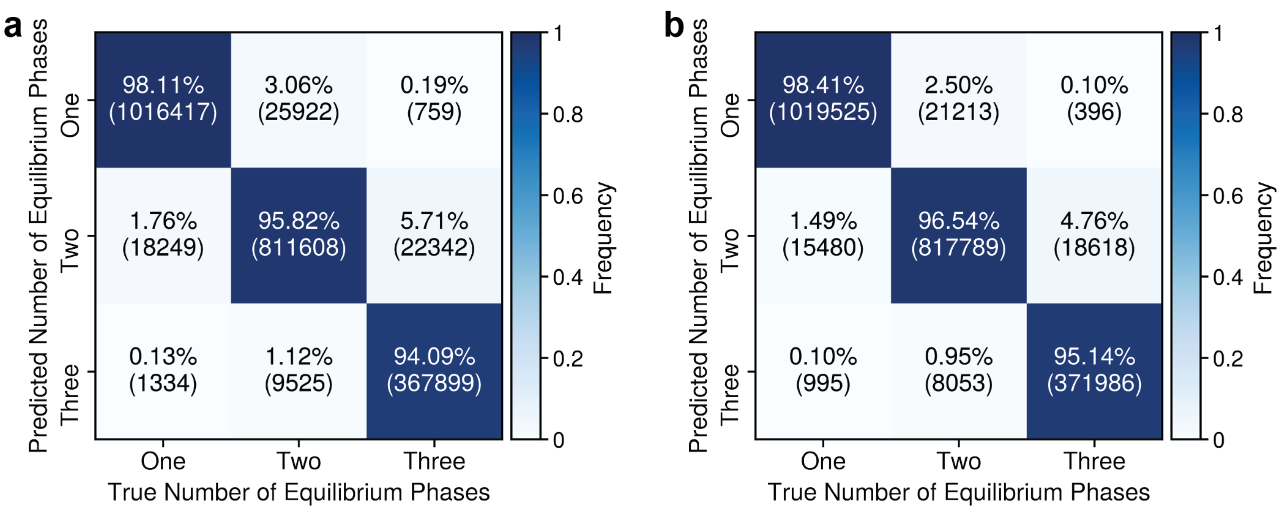


Figure S9: **Confusion matrices for the predicted number of equilibrium phases using the PI+ model with a)** 10% of training data, and **b)** with full training data. Diagonal entries represent correctly classified instances, while off-diagonal entries represent misclassifications.

S3 Equilibrium composition prediction parity plots

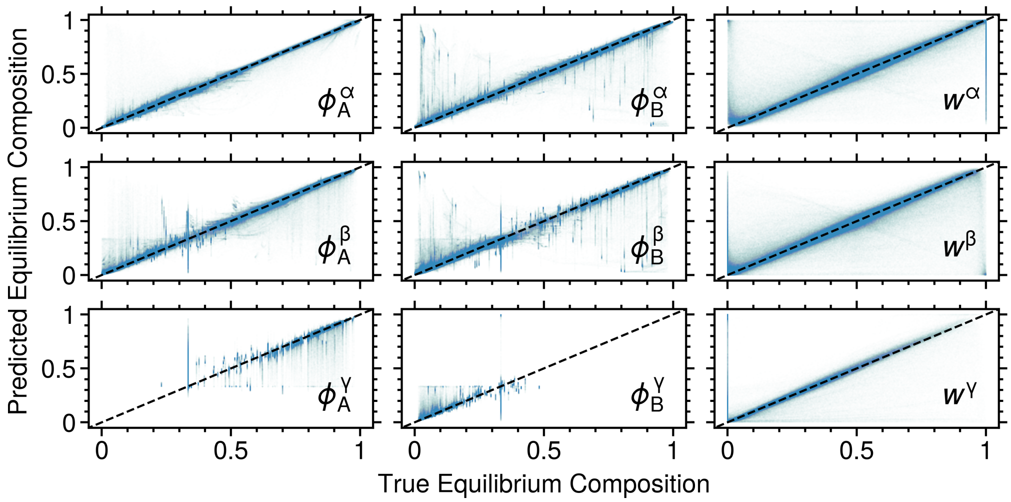


Figure S10: **Parity plot for predicted equilibrium composition using the baseline model with 10% of training data.** The diagonal dashed line represents perfect performance.

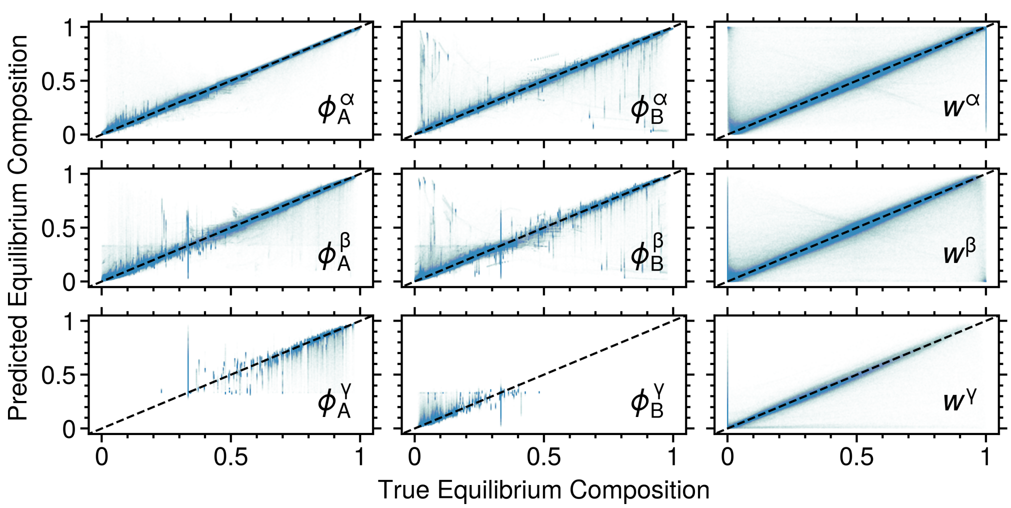


Figure S11: **Parity plot for predicted equilibrium composition using the PI model with 10% of training data.** The diagonal dashed line represents perfect performance.

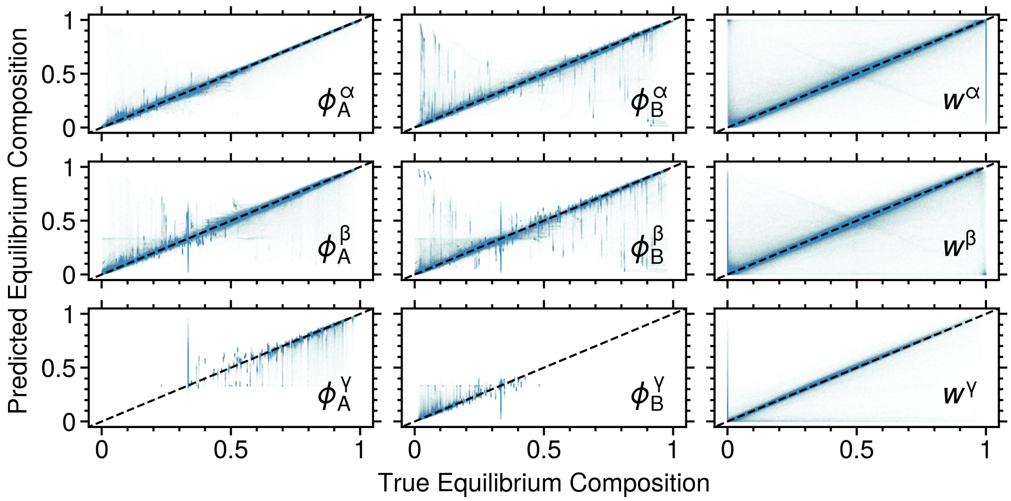


Figure S12: **Parity plot for predicted equilibrium composition using the PI model with full training data.** The diagonal dashed line represents perfect performance.

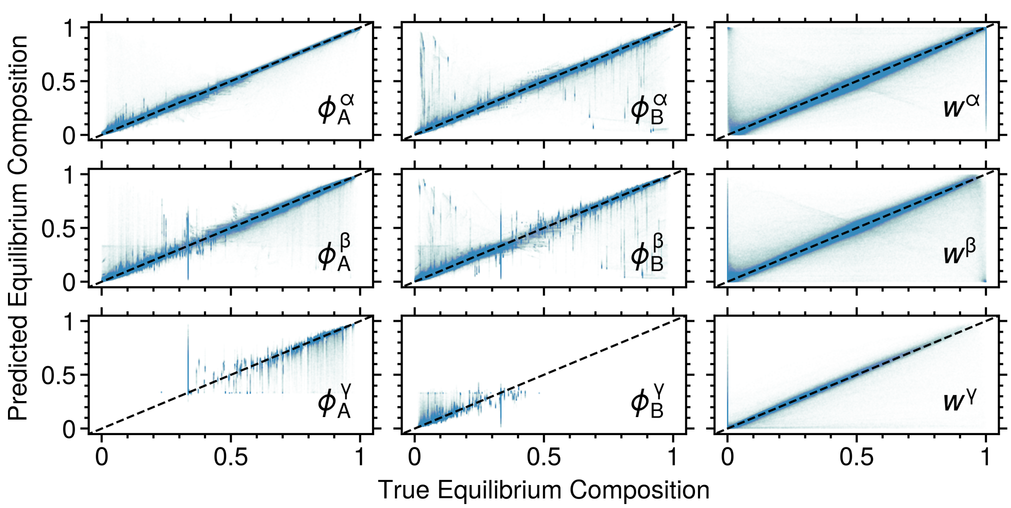


Figure S13: **Parity plot for predicted equilibrium composition using the PI+ model with 10% of training data.** The diagonal dashed line represents perfect performance.

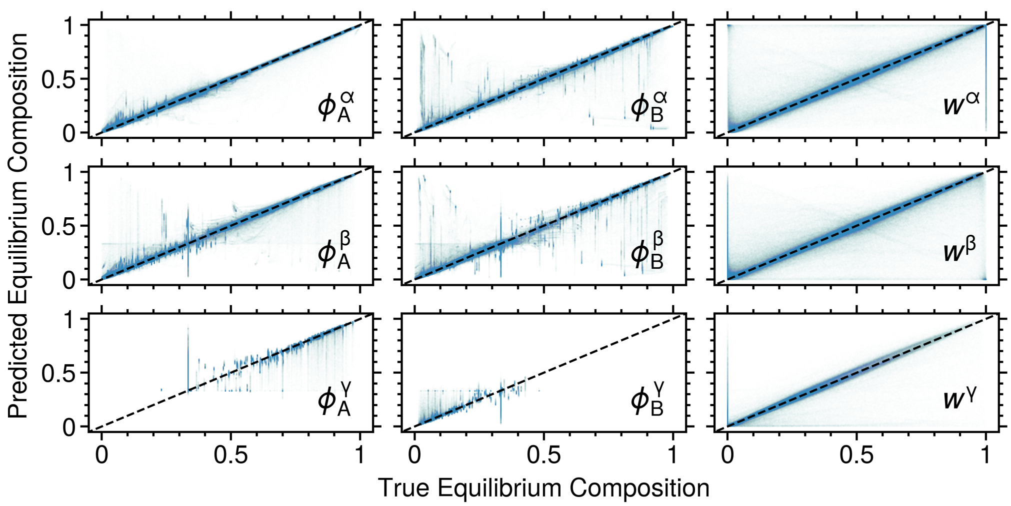


Figure S14: **Parity plot for predicted equilibrium composition using the PI+ model with full training data.** The diagonal dashed line represents perfect performance.

S4 Post-ML optimization performance

Table S3: **Convergence time and success rate of equilibrium composition prediction with post-ML Newton-CG optimization.** Mean values are reported with standard errors in parentheses. The best result is highlighted in bold and underlined.

	Data size	Convergence time per point (sec)	Two-phase success rate	Three-phase success rate
Base	100%	**0.415 (0.024)**	0.993 (0.000)	**0.960 (0.001)**
PI	100%	0.431 (0.021)	0.990 (0.000)	0.959 (0.001)
PI+	100%	0.542 (0.028)	0.990 (0.000)	0.923 (0.001)
Base	10%	0.520 (0.030)	0.988 (0.000)	0.935 (0.001)
PI	10%	0.537 (0.021)	0.987 (0.000)	0.922 (0.001)
PI+	10%	0.489 (0.029)	**0.994 (0.000)**	0.944 (0.001)

S5 Impact of Weighting Parameters on the PI+ Loss Function

Table S4: **Impact of weighting parameters in the loss function (Eq. 15) on the performance of the PI+ model.** The rows are ranked by the mean of the sum of equilibrium composition regression R^2 and phase classification F_1. The parameters used in this study is highlighted in bold and demonstrates statistically equivalent performance to other top-performing parameters. The PI model without physics-informed loss is indicated with an underscore. Performance deteriorates as λ_{f} increases, while the impact of other parameters is minor.

λ_{split}	$\lambda_{\Delta\mu}$	λ_{f}	Mean R^2	Std. R^2	Mean F_1	Std. F_1	Mean $R^2+F_1\downarrow$	Std. R^2+F_1
0.01	0.1	0.001	0.95	0.008	0.973	0.002	1.923	0.008
1	0.1	0.001	0.948	0.011	0.973	0.004	1.922	0.012
0.01	0.001	0.001	0.948	0.009	0.972	0.002	1.92	0.009
0.001	0.001	0.001	0.949	0.011	0.97	0.002	1.919	0.012
0.1	0.1	0.001	0.947	0.009	0.972	0.003	1.919	0.009
1	0.01	0.001	0.947	0.012	0.971	0.002	1.918	0.012
0.01	**0.01**	**0.001**	**0.946**	**0.012**	**0.972**	**0.002**	**1.917**	**0.012**
0.001	0.1	0.001	0.946	0.008	0.972	0.002	1.917	0.009
0.1	0.001	0.001	0.945	0.01	0.971	0.002	1.916	0.01
0.1	0.01	0.001	0.945	0.009	0.972	0.003	1.916	0.01
0.1	1	0.001	0.943	0.009	0.973	0.004	1.916	0.01
0.001	0.01	0.001	0.945	0.01	0.97	0.002	1.915	0.01
1	0.001	0.001	0.945	0.01	0.97	0.002	1.915	0.01
1	0.1	0.01	0.944	0.007	0.972	0.003	1.915	0.007
<u>0</u>	<u>0</u>	<u>0</u>	<u>0.945</u>	<u>0.007</u>	<u>0.97</u>	<u>0.003</u>	<u>1.915</u>	<u>0.007</u>
0.001	1	0.001	0.942	0.01	0.973	0.003	1.914	0.011
0.01	0.1	0.01	0.941	0.009	0.973	0.003	1.914	0.01
0.01	1	0.001	0.94	0.009	0.974	0.003	1.914	0.01
0.01	1	0.01	0.941	0.008	0.973	0.003	1.914	0.008
1	0.01	0.01	0.943	0.009	0.971	0.002	1.914	0.009
1	1	0.001	0.942	0.006	0.972	0.002	1.914	0.006
0.1	0.1	0.01	0.94	0.008	0.973	0.003	1.913	0.008
0.1	1	0.01	0.94	0.006	0.973	0.002	1.913	0.007
1	1	0.01	0.94	0.008	0.973	0.003	1.913	0.008
0.001	1	0.01	0.938	0.011	0.973	0.003	1.911	0.011
0.001	0.1	0.01	0.936	0.008	0.973	0.002	1.909	0.009
0.1	0.01	0.01	0.934	0.008	0.973	0.003	1.906	0.008
1	0.001	0.01	0.933	0.012	0.973	0.003	1.906	0.012
0.01	0.01	0.01	0.933	0.01	0.973	0.002	1.905	0.01
0.001	0.01	0.01	0.931	0.008	0.971	0.003	1.902	0.008
0.1	0.001	0.01	0.92	0.015	0.973	0.002	1.893	0.015
0.01	0.001	0.01	0.914	0.01	0.973	0.002	1.888	0.01
0.001	0.001	0.01	0.909	0.01	0.973	0.002	1.883	0.01

Table S5: **Impact of weighting parameters in the loss function (Eq. 15) on the performance of the PI+ model (continued).** The rows are ranked by the mean of the sum of equilibrium composition regression R^2 and phase classification F_1. The parameters used in this study is highlighted in bold and demonstrates statistically equivalent performance to other top-performing parameters. The PI model without physics-informed loss is indicated with an underscore. Performance deteriorates as λ_f increases, while the impact of other parameters is minor.

λ_split	$\lambda_{\Delta\mu}$	λ_f	Mean R^2	Std. R^2	Mean F_1	Std. F_1	Mean $R^2+F_1\downarrow$	Std. R^2+F_1
1	1	0.1	0.165	0.061	0.959	0.004	1.124	0.061
0.01	0.01	0.1	0.123	0.019	0.965	0.003	1.088	0.019
0.1	0.01	0.1	0.124	0.03	0.964	0.002	1.088	0.03
0.1	0.001	0.1	0.119	0.013	0.963	0.005	1.082	0.014
0.001	0.01	0.1	0.113	0.008	0.963	0.002	1.076	0.008
0.01	0.001	0.1	0.104	0.031	0.962	0.007	1.066	0.032
0.001	0.001	0.1	0.086	0.035	0.964	0.005	1.05	0.036
0.1	1	0.1	0.076	0.045	0.959	0.007	1.035	0.045
1	0.01	0.1	0.045	0.126	0.964	0.002	1.009	0.126
0.01	1	0.1	0.045	0.083	0.954	0.006	0.999	0.084
0.001	1	0.1	-0.033	0.064	0.956	0.005	0.922	0.064
1	0.1	0.1	-0.042	0.028	0.965	0.003	0.922	0.028
1	0.001	0.1	-0.08	0.045	0.964	0.004	0.884	0.046
0.01	0.1	0.1	-0.316	0.124	0.958	0.01	0.642	0.125
0.001	0.1	0.1	-0.331	0.152	0.959	0.007	0.628	0.152
0.1	0.1	0.1	-0.402	0.087	0.959	0.007	0.556	0.088
0.1	0.01	1	-1.38	0.14	0.935	0.012	-0.445	0.14
0.01	0.001	1	-1.39	0.248	0.928	0.009	-0.463	0.248
1	0.01	1	-1.417	0.091	0.923	0.014	-0.495	0.092
1	0.1	1	-1.456	0.164	0.919	0.014	-0.538	0.165
0.001	0.1	1	-1.477	0.221	0.911	0.017	-0.566	0.222
0.01	0.01	1	-1.508	0.082	0.93	0.003	-0.578	0.082
1	0.001	1	-1.517	0.216	0.934	0.011	-0.583	0.216
0.01	0.1	1	-1.508	0.139	0.924	0.01	-0.584	0.14
0.1	0.001	1	-1.51	0.199	0.924	0.015	-0.587	0.2
1	1	1	-1.521	0.381	0.907	0.022	-0.614	0.381
0.001	0.001	1	-1.574	0.14	0.922	0.008	-0.652	0.14
0.001	0.01	1	-1.622	0.213	0.928	0.011	-0.694	0.213
0.1	0.1	1	-1.678	0.297	0.918	0.021	-0.76	0.297
0.1	1	1	-1.999	0.39	0.9	0.016	-1.1	0.39
0.001	1	1	-2.057	0.354	0.893	0.014	-1.164	0.355
0.01	1	1	-2.247	0.303	0.898	0.02	-1.349	0.304