**Extracting the iron concentration in silicon solar cells using photovoltaic parameters and machine learning**

Oleg Olikh, Oleksii Zavhorodnii

*Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Street, Kyiv, 01601, Ukraine*

olegolikh@knu.ua

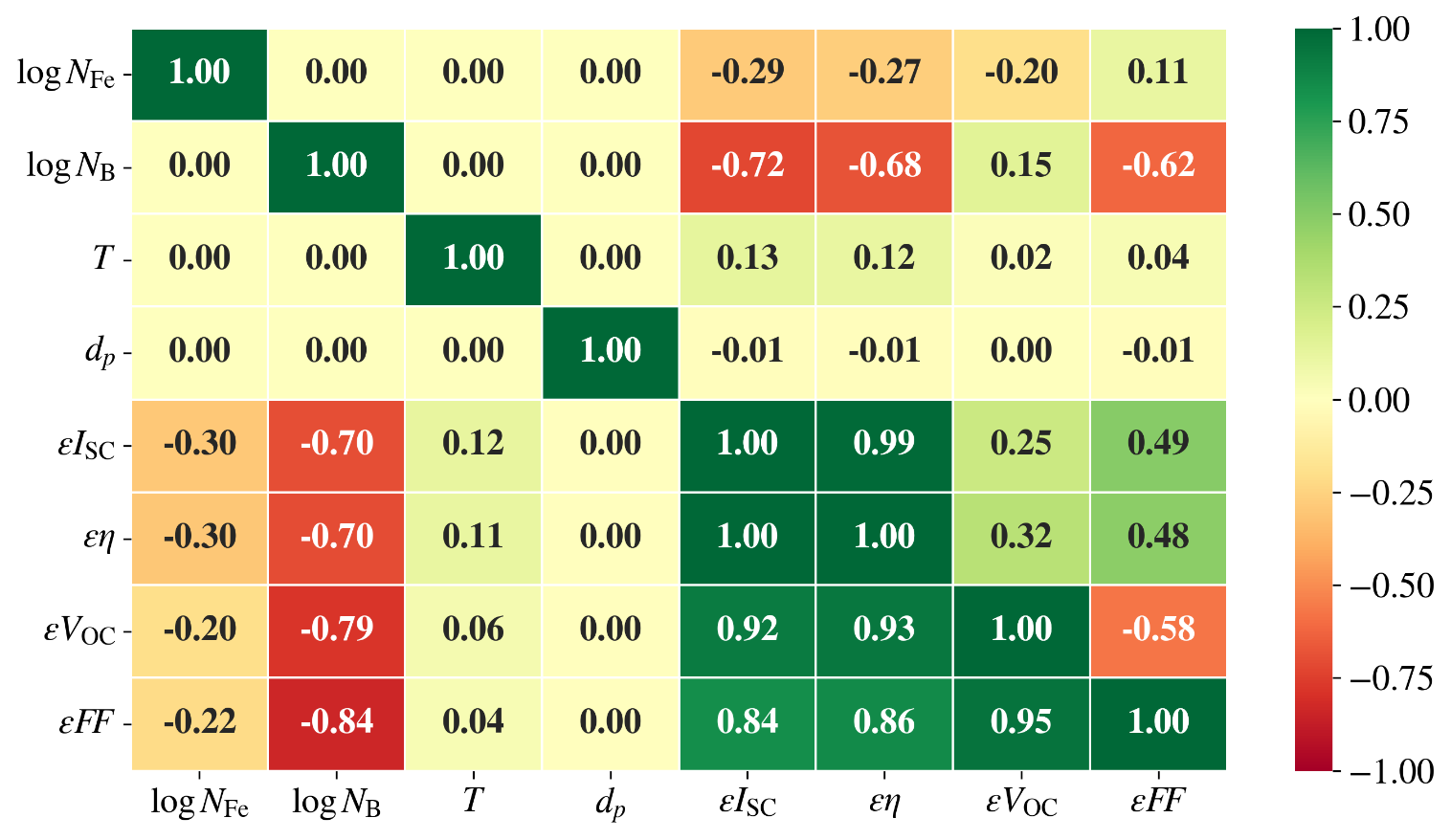


Fig.S1. Correlation plot of features in training set. Data above and below the main diagonal correspond to AM1.5 and 940 nm illumination, respectively.

Table S1. Hyperparameter space for RF

|  |  |
| --- | --- |
| Hyperparameter | Values |
| # estimators | 100, 200, 250, 300, 350, 400, 450,500,550, 600, 650, 700 |
| max depth | 10, 15, 20, 25, 30, 35, 40, 45 |
| min samples leaf | 1, 2, 3, 4, 5, 6, 7 |
| min samples split | 2, 3, 4, 5, 6, 7 |
| bootstrap | True, False |
| max features | 'log2', 'sqrt', 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 |

Table S2. Hyperparameter space for GB

|  |  |
| --- | --- |
| Hyperparameter | Values |
| # estimators | 100, 200, 250, 300, 350, 400, 450,500,550, 600, 650 |
| max depth | 15, 20, 25, 30, 35, 40, 45 |
| min samples leaf | 1, 2, 3, 4, 5, 6, 7 |
| min samples split | 2, 3, 4, 5, 6, 7 |
| learning rate | [10-3, 10-1] |
| max features | 'log2', 'sqrt', 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 |

Table S3. Hyperparameter space for XGB

|  |  |
| --- | --- |
| Hyperparameter | Values |
| booster | gbtree, gblinear, dart |
| max depth\* | 3, 4, 5, 6, 7, 10, 15, 20 |
| min split loss\* | [10-6; 5] |
| min child weight\* | [0; 15] |
| subsample\* | 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 |
| сolsample by tree\* | 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 |
| # estimators | 200, 300, 400, 500, 600, 700, 800, 900 |
| learning rate | [10-5; 1] |
| L1 | [10-8; 1] |
| L2 | [10-8; 10] |
| \* for all boosters except gblinear | |

Table S4. Hyperparameter space for SVR

|  |  |
| --- | --- |
| Hyperparameter | Values |
| kernel | linear, poly, rbf, sigmoid |
| degree\* | 2, 3, 4, 5, 6 |
| C0 | [0; 5] |
| Tolerance | [10-5; 10-2] |
| C | [10-2; 15] |
| Epsilon | [10-3; 1] |
| \* for poly kernel only | |

Table S5. Hyperparameter space for DNN

|  |  |
| --- | --- |
| Hyperparameter | Values |
| hidden layers configuration\* | Pipe, Trapezium, Triangle, Butterfly, Fir |
| # nodes for first hidden layer | 5, 10, 20, 30, 50, 75, 100, 120, 150, 200, 250 |
| # hidden layers\*\* | 5, 6, 8, 10, 12, 15 |
| batch size | 8, 16, 32, 64, 128 |
| activation function | ReLu, sigmoid, tanh, SELU, ELU |
| optimizer | SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, Ftrl |
| learning rate | [10-5; 10-2] |
| # epochs | 100, 300, 400, 500, 600, 700, 1000, 1500 |
| weight initializer | Xavier Normal, Xavier Uniform, He Normal, He Uniform, Random Normal, Random Uniform |
| \* The configurations are shown in Fig.S2. | |
| \*\* For Pipe configuration only | |

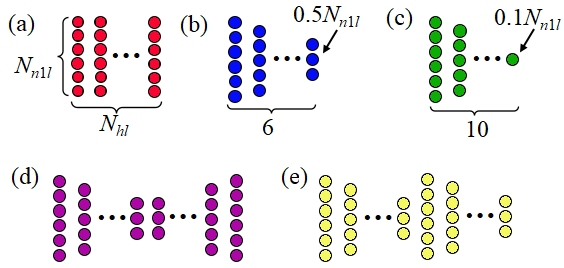
 Fig.S2. The considered configuration of the hidden layers for DNN models: a) pipe; b) trapezium; c) triangle; d) butterfly (two serial reflected trapezium); e) fir (two serial trapezium). *Nn*1*l* is the number nodes for first hidden layer, *Nhl* is the number of hidden layers.

Table S6. Chosen hyperparameter combinations for RF models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Hyperparameter | | | | | |
| # estimators | max depth | min sample split | min sample leaf | max features | bootstrap |
|  | 200 | 15 | 3 | 1 | 1.0 | True |
|  | 650 | 15 | 2 | 1 | 0.9 | True |
|  | 500 | 30 | 2 | 1 | 0.9 | True |
|  | 200 | 20 | 2 | 1 | 0.9 | True |
|  | 600 | 20 | 2 | 1 | 0.6 | False |
|  | 300 | 15 | 4 | 1 | 0.6 | False |
|  | 500 | 35 | 2 | 1 | 1.0 | True |
|  | 700 | 45 | 2 | 1 | 0.6 | True |
|  | 500 | 30 | 2 | 1 | 0.5 | False |
|  | 600 | 30 | 2 | 1 | 0.6 | False |
|  | 400 | 30 | 2 | 1 | 0.5 | False |
|  | 500 | 25 | 2 | 1 | 0.7 | False |
|  | 600 | 45 | 2 | 1 | 0.7 | False |
|  | 550 | 40 | 2 | 1 | 0.6 | False |
|  | 450 | 40 | 2 | 1 | 0.7 | False |
|  | 450 | 30 | 2 | 1 | 0.6 | False |

Table S7. Chosen hyperparameter combinations for GB models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Hyperparameter | | | | | |
| # estimators | max depth | min sample split | min sample leaf | max features | learning rate |
|  | 500 | 15 | 4 | 7 | 0.8 | 8.3e-03 |
|  | 550 | 15 | 5 | 7 | 0.8 | 8.1e-03 |
|  | 650 | 30 | 6 | 6 | 0.8 | 1.1e-02 |
|  | 650 | 45 | 4 | 7 | 0.7 | 1.3e-02 |
|  | 650 | 45 | 2 | 4 | 0.8 | 4.3e-02 |
|  | 550 | 15 | 4 | 3 | 0.6 | 7.2e-03 |
|  | 600 | 40 | 4 | 6 | 0.7 | 1.9e-02 |
|  | 600 | 40 | 6 | 6 | 0.5 | 3.8e-02 |
|  | 600 | 45 | 7 | 5 | 0.7 | 3.4e-02 |
|  | 550 | 30 | 7 | 7 | 0.6 | 1.9e-02 |
|  | 650 | 45 | 7 | 7 | 0.7 | 2.7e-02 |
|  | 400 | 35 | 3 | 7 | 0.9 | 3.5e-02 |
|  | 450 | 15 | 4 | 6 | 0.7 | 2.3e-02 |
|  | 550 | 40 | 2 | 6 | 0.6 | 2.1e-02 |
|  | 650 | 15 | 5 | 7 | 0.7 | 2.8e-02 |
|  | 600 | 40 | 3 | 5 | 0.6 | 2.7e-02 |

Table S8. Chosen hyperparameter combinations for XGB models

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Hyperparameter | | | | | | | | | |
| booster | max depth | min split loss | min child weight | sub  sample | сolsample bytree | # estimators | learning rate | L1 | L2 |
|  | dart | 10 | 5.3e-4 | 10.8 | 0.6 | 1 | 400 | 1.7e-2 | 2.8e-2 | 3.1e-2 |
|  | gbtree | 15 | 1.9e-3 | 11.1 | 0.7 | 1 | 500 | 9.5e-3 | 1.8e-4 | 2.6e-3 |
|  | dart | 15 | 2.3e-5 | 5.2 | 0.7 | 1 | 700 | 1.1e-2 | 6.6e-3 | 5.3e-2 |
|  | gbtree | 15 | 1.8e-6 | 4 | 0.7 | 1 | 800 | 8.2e-3 | 1.1e-4 | 1 |
|  | gbtree | 15 | 1.1e-5 | 1.9 | 0.7 | 1 | 800 | 7.3e-2 | 9.9e-4 | 9.8 |
|  | dart | 15 | 1.6e-5 | 10.4 | 0.3 | 1 | 800 | 6.7e-3 | 7.0e-3 | 0.5 |
|  | dart | 20 | 1.3e-5 | 1.5 | 0.7 | 1 | 900 | 1.1e-2 | 1.4e-4 | 2.4 |
|  | dart | 20 | 1.3e-4 | 6.9 | 0.6 | 1 | 900 | 1.1e-2 | 2.5e-3 | 1.4e-3 |
|  | dart | 20 | 5.5e-5 | 0.7 | 0.5 | 1 | 500 | 1.7e-2 | 1.7e-3 | 0.3 |
|  | dart | 15 | 6.8e-6 | 5.7 | 0.8 | 1 | 400 | 6.3e-2 | 8.1e-2 | 6.9e-2 |
|  | gbtree | 20 | 1.0e-5 | 3 | 0.5 | 1 | 900 | 1.0e-2 | 7.2e-4 | 1.3e-3 |
|  | dart | 15 | 4.3e-6 | 9.9 | 0.7 | 1 | 500 | 4.7e-2 | 4.9e-4 | 3.3e-2 |
|  | dart | 15 | 1.4e-5 | 3.9 | 0.3 | 1 | 700 | 5.6e-2 | 2.0e-3 | 5.9 |
|  | gbtree | 20 | 4.7e-6 | 12.6 | 0.6 | 0.9 | 900 | 4.4e-2 | 7.9e-2 | 0.3 |
|  | dart | 20 | 2.5e-6 | 8.3 | 0.5 | 1 | 900 | 8.4e-2 | 1.1e-3 | 7.4 |
|  | gbtree | 15 | 1.3e-4 | 1.5 | 0.4 | 1 | 600 | 2.4e-2 | 1.2e-4 | 4.9e-2 |

Table S9. Chosen hyperparameter combinations for SVR models

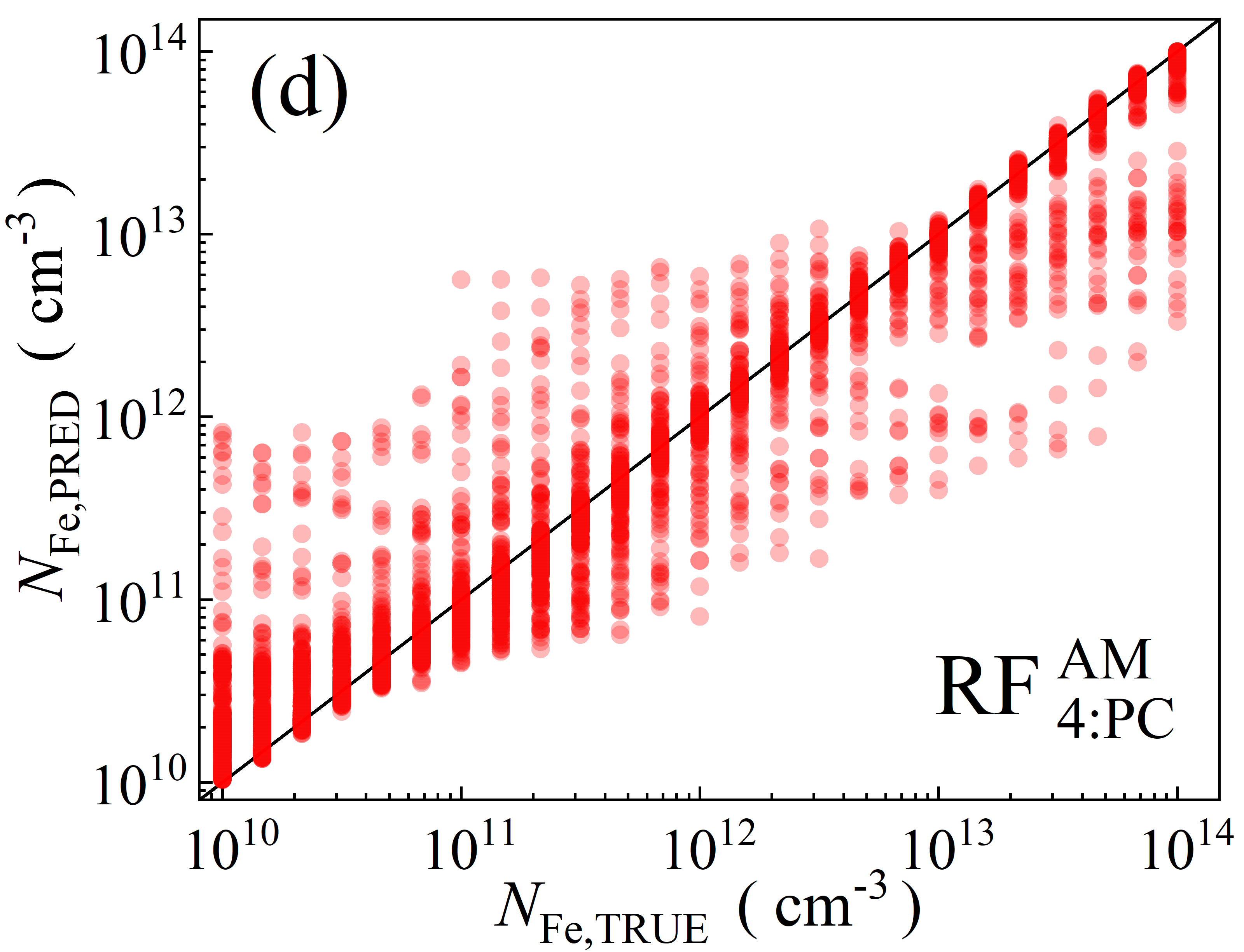
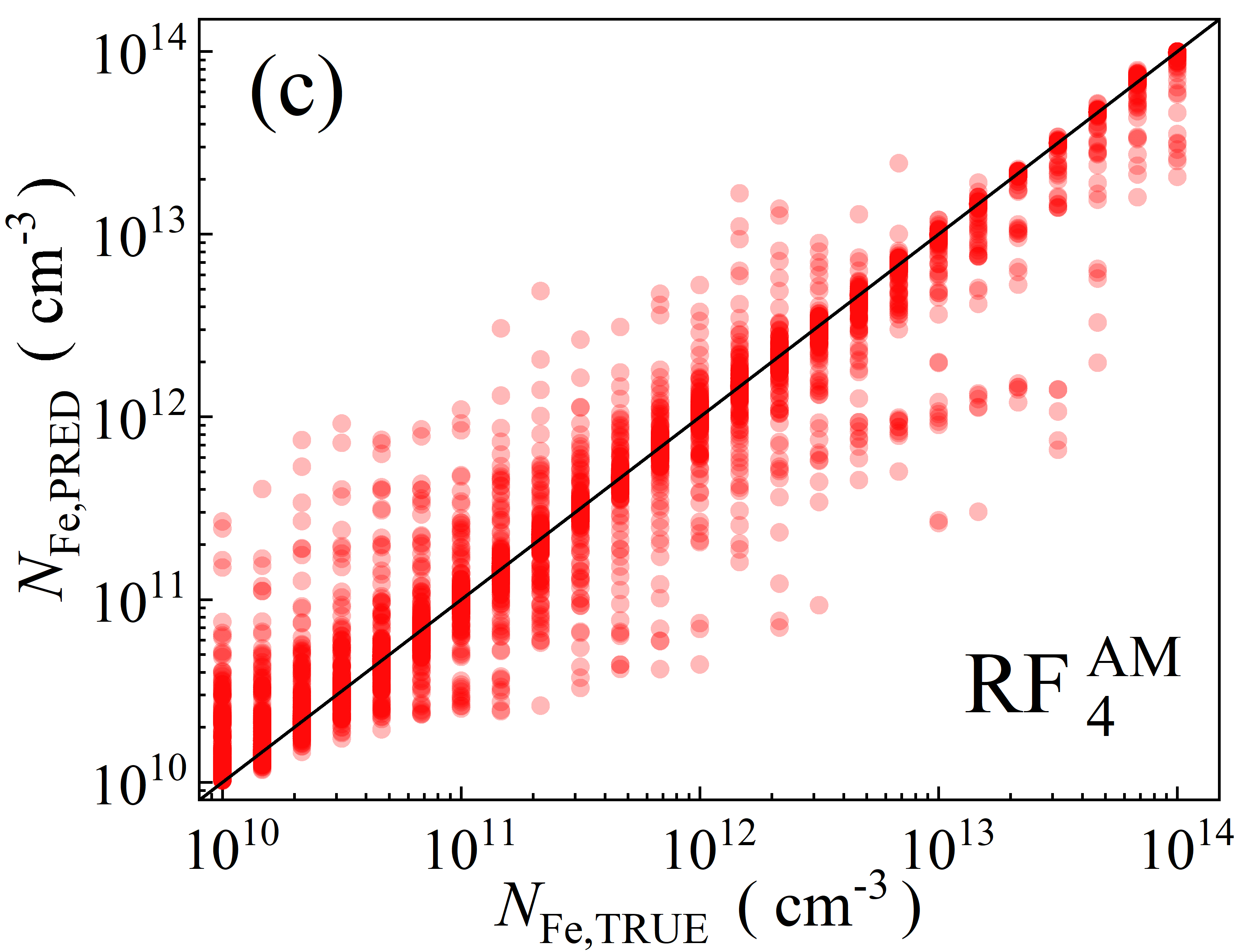
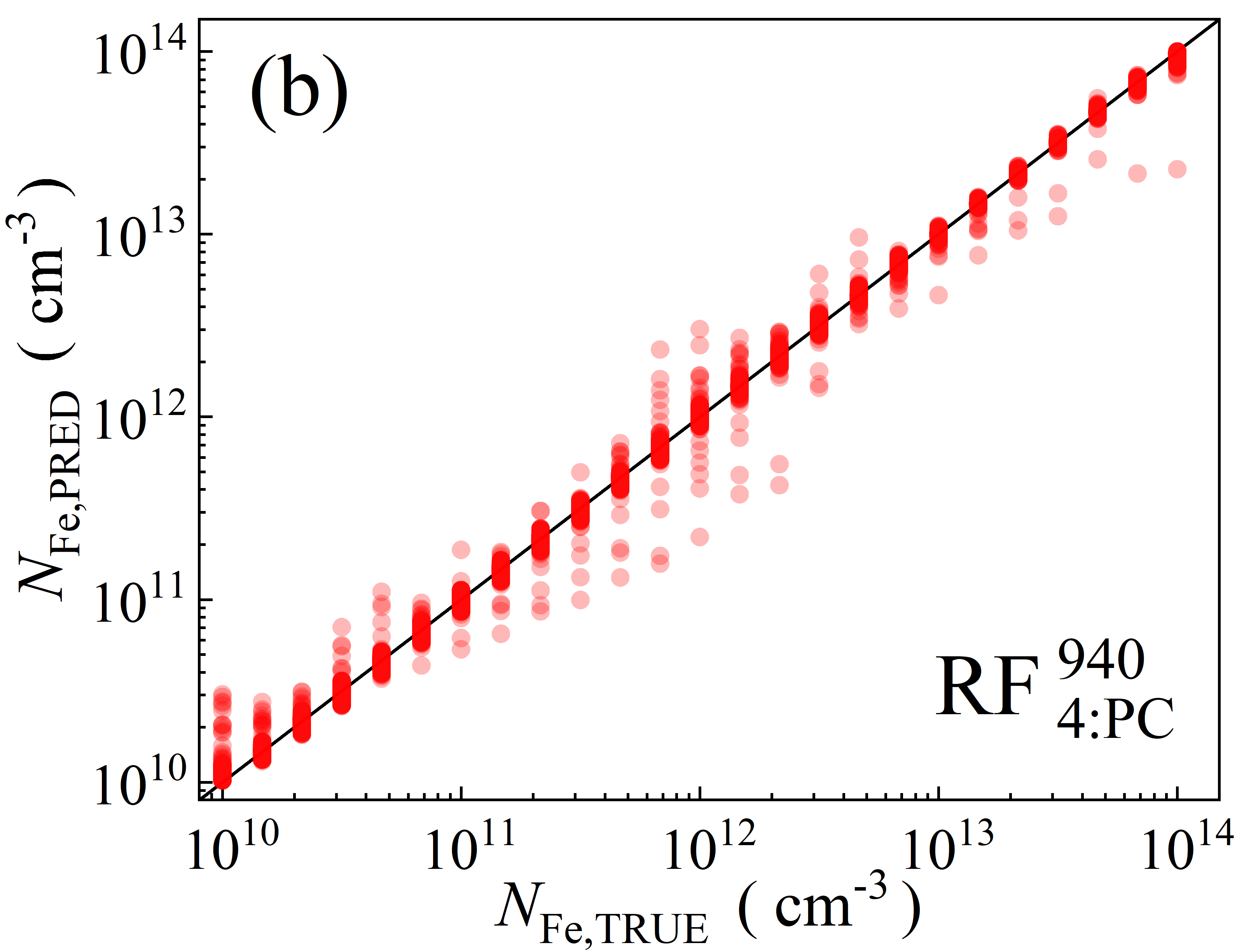
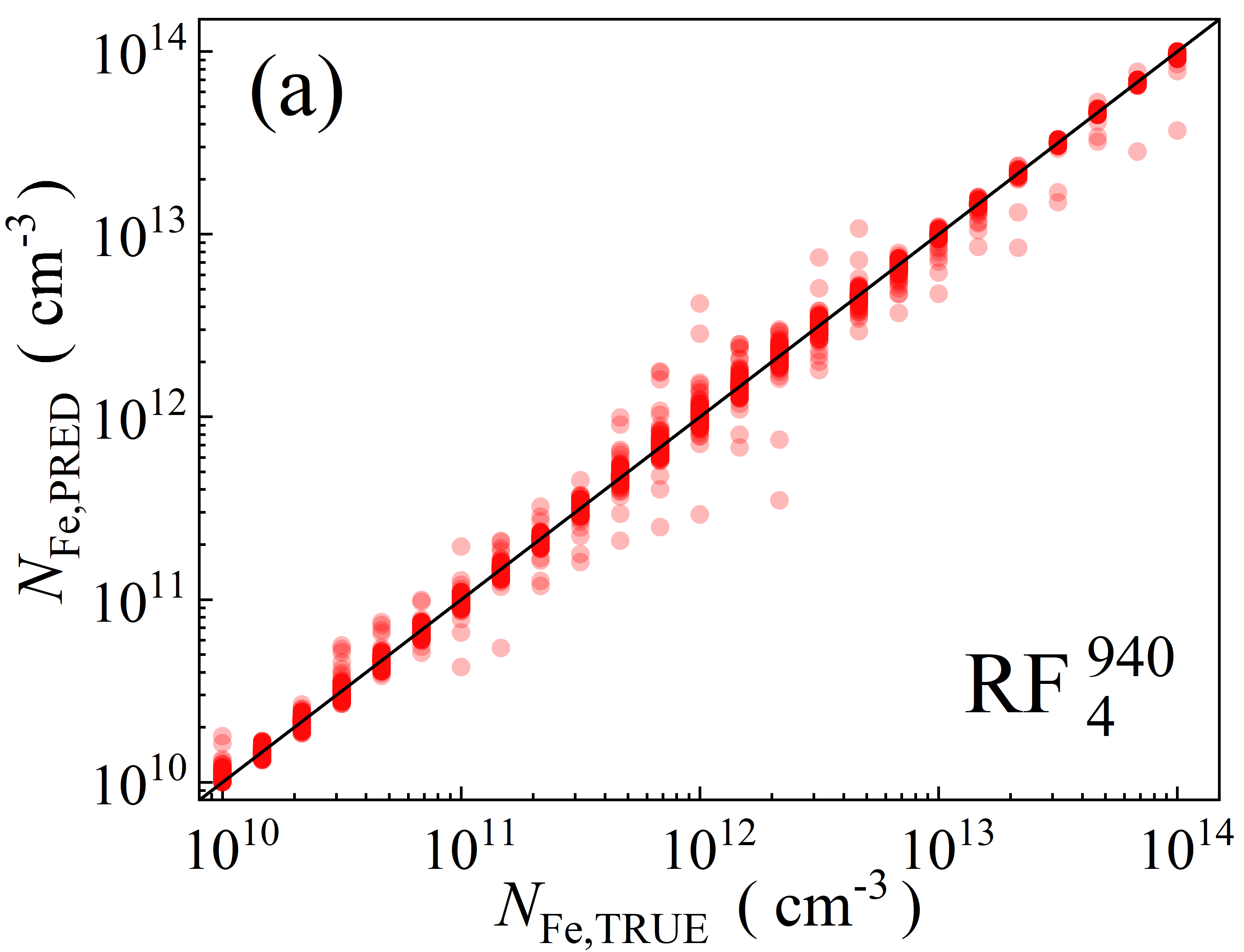
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Hyperparameter | | | | |
| kernel | C0 | Tolerance | C | Epsilon |
|  | rbf | 0.78 | 3.3e-05 | 15 | 0.15 |
|  | rbf | 0.81 | 1.1e-03 | 15 | 0.15 |
|  | rbf | 0.31 | 2.3e-03 | 15 | 0.11 |
|  | rbf | 0.81 | 1.4e-03 | 15 | 0.09 |
|  | rbf | 0.56 | 2.7e-04 | 15 | 0.16 |
|  | rbf | 0.88 | 4.5e-03 | 15 | 0.19 |
|  | rbf | 0.67 | 1.8e-04 | 15 | 0.15 |
|  | rbf | 0.69 | 3.8e-04 | 15 | 0.17 |
|  | rbf | 0.35 | 6.3e-03 | 15 | 0.19 |
|  | rbf | 0.11 | 1.9e-04 | 15 | 0.24 |
|  | rbf | 0.94 | 2.9e-04 | 15 | 0.16 |
|  | rbf | 0.42 | 1.7e-04 | 15 | 0.19 |
|  | rbf | 0.42 | 4.8e-04 | 15 | 0.22 |
|  | rbf | 0.02 | 9.5e-04 | 15 | 0.19 |
|  | rbf | 0.82 | 7.2e-04 | 15 | 0.19 |
|  | rbf | 0.95 | 1.2e-04 | 15 | 0.17 |

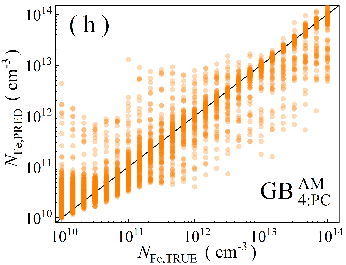
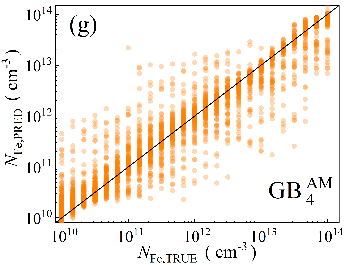
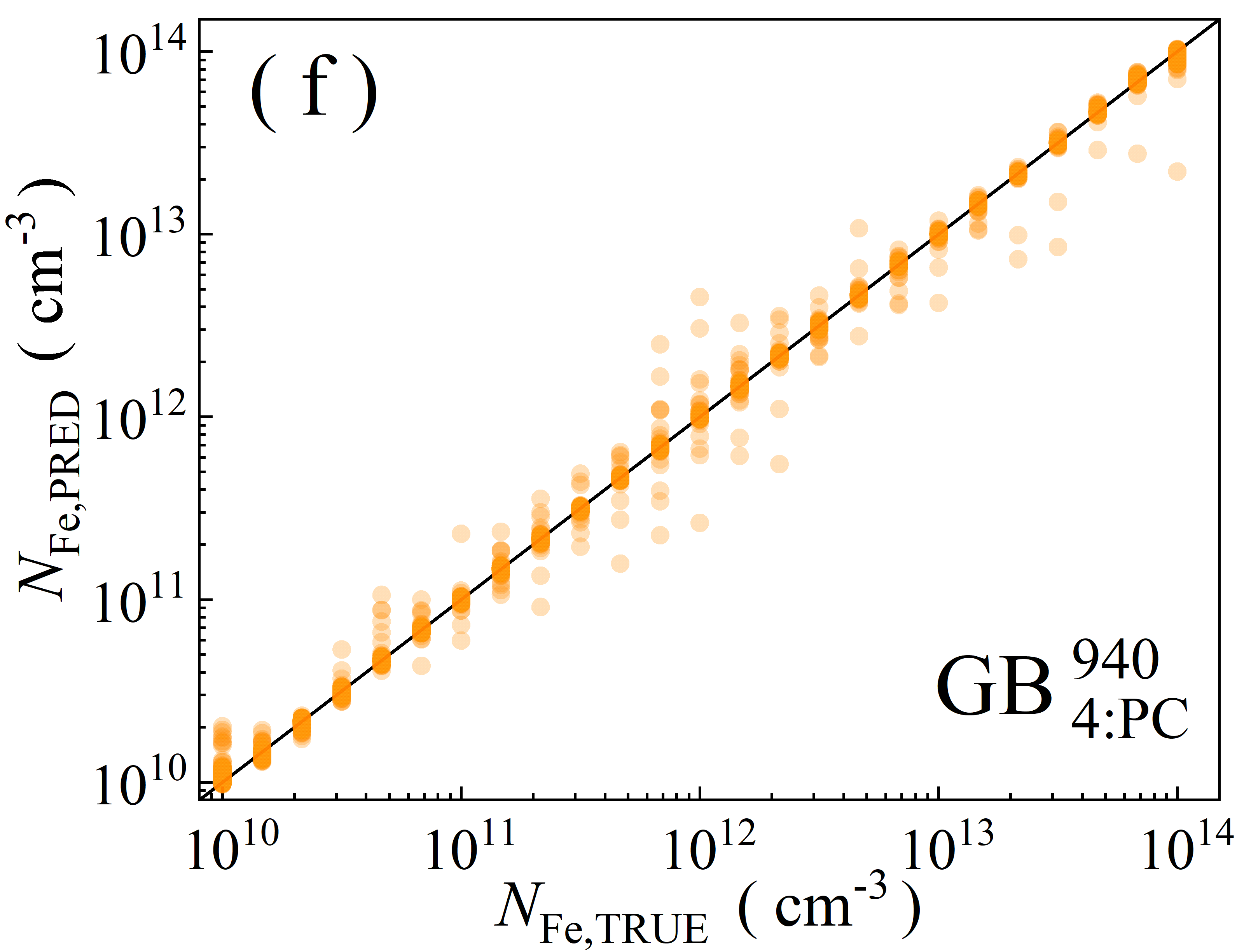
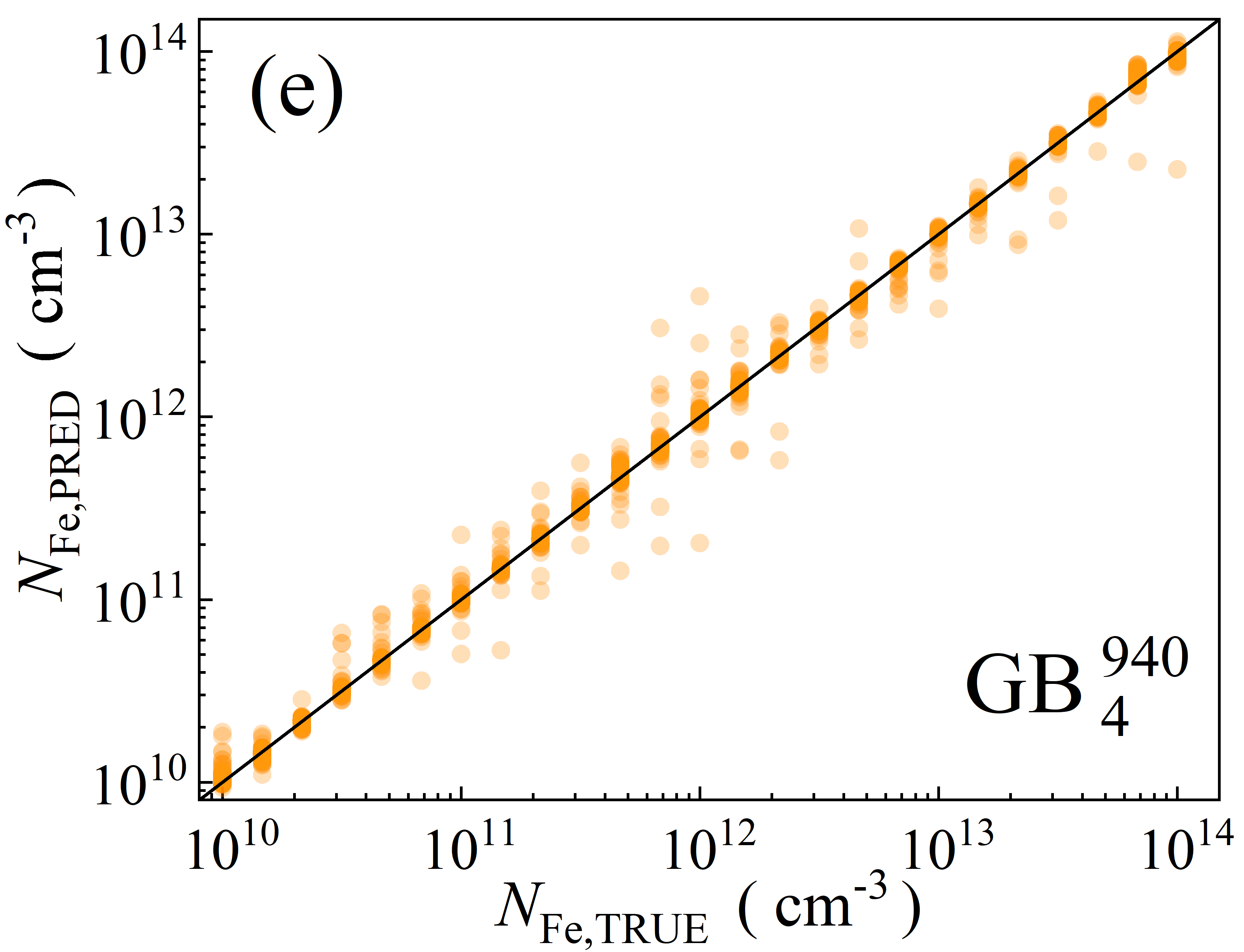
Table S10. Chosen hyperparameter combinations for DNN models

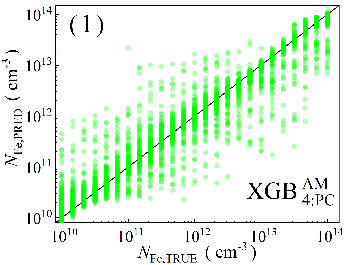
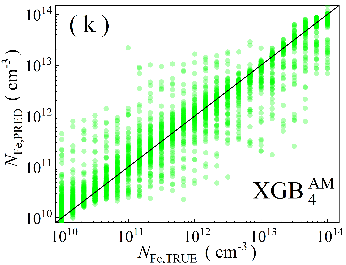
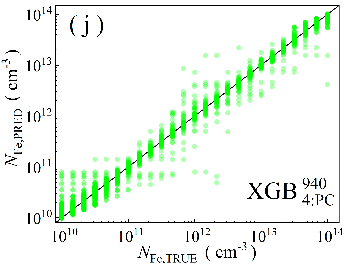
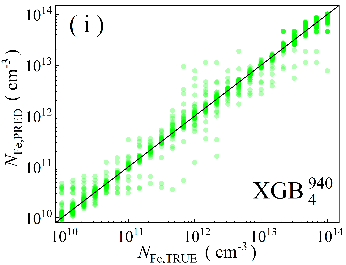
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Мережа | Параметр | | | | | | | | | |
| config | *Nhl* | *Nn*1*l* | BS | Epochs | LR, 10-4 | Optim | Activ | Init |
|  | Pipe | 8 | 50 | 64 | 500 | 12.0 | Adamax | tanh | XN |
|  | Pipe | 8 | 120 | 16 | 500 | 5.4 | Nadam | tanh | XU |
|  | Pipe | 5 | 200 | 128 | 500 | 3.4 | Nadam | tanh | XU |
|  | Trapezium | - | 250 | 16 | 500 | 3.7 | Adamax | tanh | XN |
|  | Trapezium | - | 50 | 16 | 500 | 2.7 | Adam | tanh | XN |
|  | Pipe | 10 | 50 | 16 | 500 | 6.3 | Adamax | relu | XN |
|  | Trapezium | - | 200 | 64 | 500 | 7.8 | Adamax | tanh | XU |
|  | Pipe | 5 | 100 | 16 | 500 | 1.9 | Adamax | relu | XN |
|  | Trapezium | - | 150 | 8 | 500 | 3.5 | Nadam | tanh | XN |
|  | Trapezium | - | 150 | 128 | 500 | 1.5 | Nadam | tanh | XU |
|  | Pipe | 5 | 50 | 128 | 500 | 3.9 | Nadam | elu | XU |
|  | Trapezium | - | 150 | 64 | 500 | 1.0 | Nadam | tanh | XU |
|  | Pipe | 5 | 120 | 16 | 500 | 10.3 | Adamax | elu | XN |
|  | Trapezium | - | 100 | 32 | 500 | 18.3 | Adamax | relu | XN |
|  | Pipe | 6 | 50 | 16 | 500 | 16.4 | Nadam | tanh | XN |
|  | Pipe | 6 | 100 | 8 | 500 | 60.5 | Adamax | elu | XU |

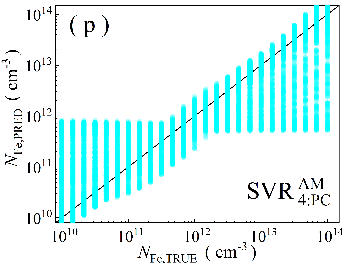
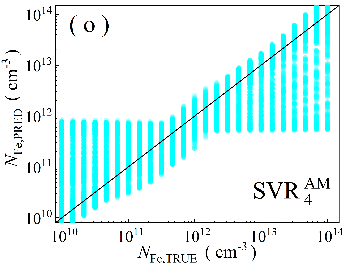
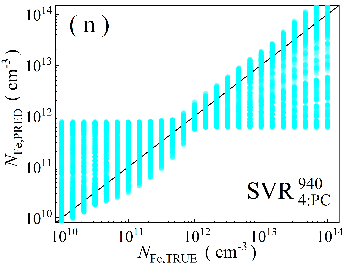
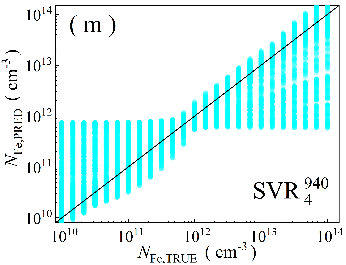
Table S11. Performance metrics of the models using fivefold cross-validation of train dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | MSE, 10-3 | MAPE, % | R2, 10-3 | Model | MSE, 10-3 | MAPE, % | R2, 10-3 |
|  | 5.2 ± 0.1 | 11 ± 2 | 980 ± 5 |  | 32 ± 3 | 41 ± 7 | 970 ± 10 |
|  | 7.0 ± 1.0 | 12 ± 1 | 976 ± 4 |  | 52 ± 5 | 90 ± 30 | 910 ± 10 |
|  | 4.1 ± 0.5 | 9.6 ± 0.5 | 994 ± 1 |  | 12 ± 2 | 17 ± 3 | 993 ± 3 |
|  | 6 ± 1 | 13 ± 1 | 970 ± 10 |  | 76 ± 6 | 100 ± 20 | 895 ± 15 |
|  | 3.0 ± 0.8 | 7.1 ± 0.5 | 997 ± 2 |  | 4.3 ± 0.4 | 7.9 ± 0.5 | 994 ± 4 |
|  | 6 ± 1 | 14 ± 4 | 971 ± 8 |  | 25 ± 5 | 34 ± 10 | 940 ± 15 |
|  | 3.0 ± 0.8 | 6.6 ± 0.3 | **998 ± 1** |  | 3.9 ± 0.4 | 7.0 ± 0.7 | 994 ± 6 |
|  | 3.6 ± 0.9 | 9.0 ± 0.5 | 993 ± 2 |  | 7 ± 1 | 12 ± 1 | 966 ± 5 |
|  | 3.7 ± 0.9 | 7.4 ± 0.7 | 986 ± 3 |  | 34 ± 6 | 60 ± 20 | 940 ± 7 |
|  | 5.0 ± 0.9 | 10 ± 1 | 975 ± 8 |  | 55 ± 8 | 100 ± 30 | 910 ± 10 |
|  | 2.5 ± 0.5 | 6.5 ± 0.2 | 995 ± 1 |  | 10 ± 2 | 14 ± 4 | 993 ± 3 |
|  | 4 ± 1 | 9.4 ± 0.7 | 983 ± 4 |  | 77 ± 7 | 130 ± 10 | 900 ± 15 |
|  | 1.9 ± 0.3 | 5.2 ± 0.3 | 997 ± 1 |  | 4.1 ± 0.6 | 7.7 ± 0.9 | **997 ± 2** |
|  | 5 ± 1 | 11 ± 4 | 976 ± 7 |  | 23 ± 4 | 30 ± 10 | 937 ± 7 |
|  | 1.9 ± 0.2 | 5.3 ± 0.3 | **998 ± 1** |  | 3.3 ± 0.6 | 6.5 ± 0.4 | 992 ± 4 |
|  | 3.2 ± 0.6 | 8.0 ± 0.6 | 992 ± 3 |  | 5.4 ± 0.9 | 9.6 ± 0.5 | 970 ± 10 |
|  | 4.8 ± 0.5 | 9.2 ± 0.8 | 964 ± 4 |  | 36 ± 4 | 50 ± 10 | 925 ± 15 |
|  | 8.7 ± 0.5 | 15 ± 1 | 960 ± 6 |  | 52 ± 5 | 110 ± 50 | 900 ± 10 |
|  | 2.8 ± 0.4 | 6.4 ± 0.4 | 982 ± 4 |  | 10 ± 2 | 19 ± 7 | 985 ± 2 |
|  | 6.5 ± 0.4 | 11.2 ± 0.3 | 966 ± 6 |  | 80 ± 4 | 130 ± 30 | 870 ± 10 |
|  | **1.4 ± 0.3** | **4.3 ± 0.3** | 996 ± 1 |  | 3.3 ± 0.5 | 6.8 ± 0.7 | **997 ± 3** |
|  | 5.8 ± 0.9 | 11 ± 1 | 968 ± 6 |  | 22 ± 2 | 35 ± 10 | 950 ± 10 |
|  | 1.5 ± 0.2 | 5.4 ± 0.1 | 996 ± 1 |  | 2.7 ± 0.3 | 6.5 ± 0.6 | 992 ± 4 |
|  | 4 ± 1 | 8.1 ± 0.5 | 970 ± 20 |  | 5 ± 1 | 10 ± 1 | 961 ± 7 |
|  | 220 ± 20 | 200 ± 20 | 540 ± 30 |  | 230 ± 5 | 220 ± 20 | 500 ± 20 |
|  | 221 ± 7 | 205 ± 15 | 530 ± 40 |  | 230 ± 10 | 215 ± 10 | 510 ± 20 |
|  | 210 ± 10 | 180 ± 15 | 480 ± 40 |  | 200 ± 10 | 180 ± 30 | 520 ± 20 |
|  | 215 ± 9 | 180 ± 15 | 520 ± 20 |  | 243 ± 8 | 220 ± 8 | 500 ± 20 |
|  | 180 ± 7 | 127 ± 7 | 600 ± 10 |  | 180 ± 15 | 59 ± 2 | 420 ± 30 |
|  | 204 ± 8 | 150 ± 10 | 515 ± 20 |  | 210 ± 15 | 76 ± 4 | 300 ± 60 |
|  | 161 ± 9 | 115 ± 10 | 610 ± 30 |  | 140 ± 8 | 55 ± 1 | 390 ± 60 |
|  | 188 ± 9 | 124 ± 8 | 600 ± 20 |  | 140 ± 9 | 50 ± 1 | 450 ± 10 |
|  | 6 ± 1 | 10 ± 2 | 971 ± 3 |  | 38 ± 4 | 57 ± 3 | 940 ± 10 |
|  | 6 ± 2 | 9 ± 2 | 980 ± 10 |  | 80 ± 20 | 125 ± 40 | 860 ± 30 |
|  | 6 ± 1 | 11 ± 2 | 970 ± 7 |  | 9 ± 6 | 25 ± 15 | 985 ± 5 |
|  | 7 ± 2 | 12 ± 3 | 973 ± 6 |  | 80 ± 8 | 230 ± 110 | 930 ± 15 |
|  | 4 ± 1 | 11 ± 3 | 970 ± 9 |  | 10 ± 6 | 14 ± 5 | 980 ± 15 |
|  | 5 ± 2 | 12 ± 7 | 967 ± 6 |  | 2.2 ± 0.7 | 7 ± 2 | 994 ± 4 |
|  | 20 ± 10 | 45 ± 25 | 880 ± 100 |  | **0.8 ± 0.4** | 3.2 ± 0.7 | 997 ± 2 |
|  | 15 ± 10 | 14 ± 2 | 975 ± 7 |  | 0.9 ± 0.5 | **3 ± 1** | 997 ± 2 |









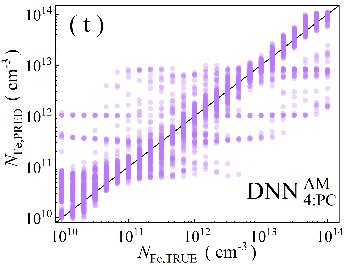
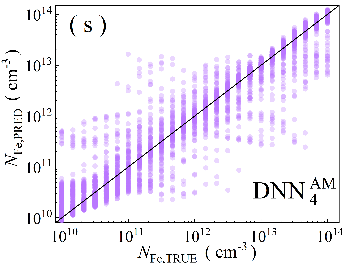
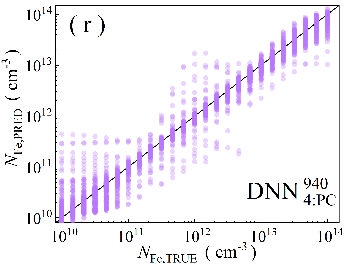
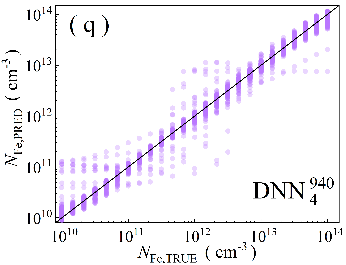
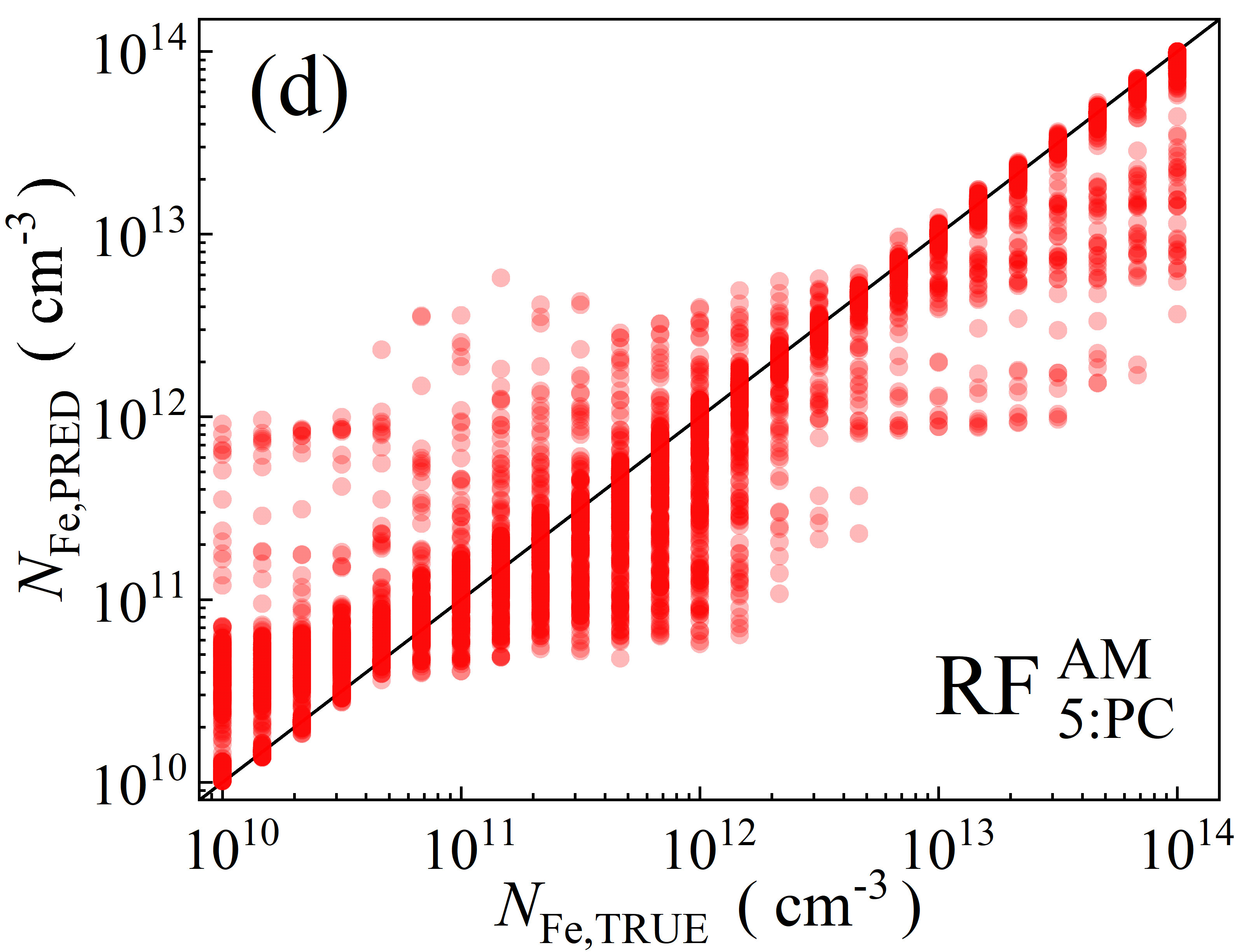
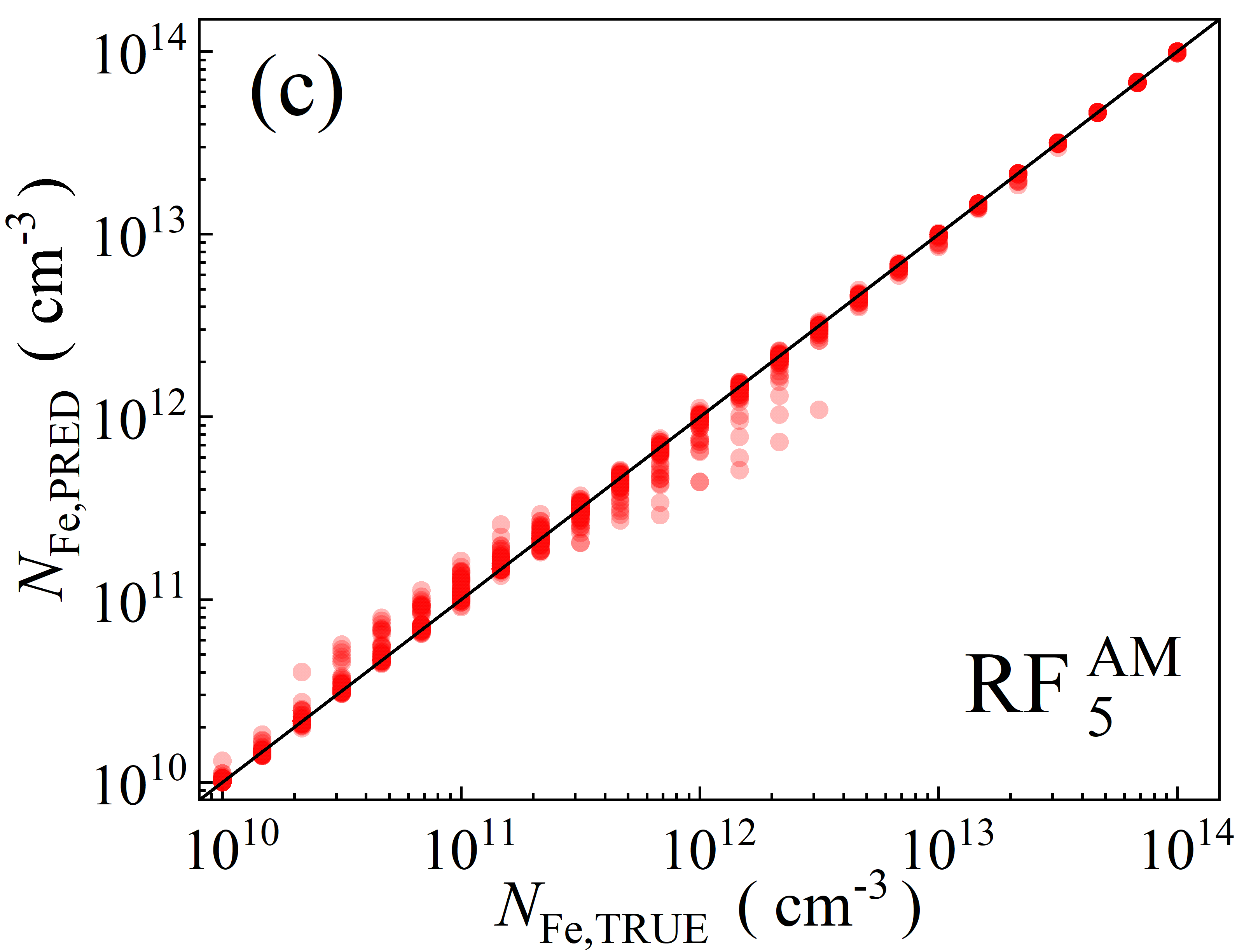
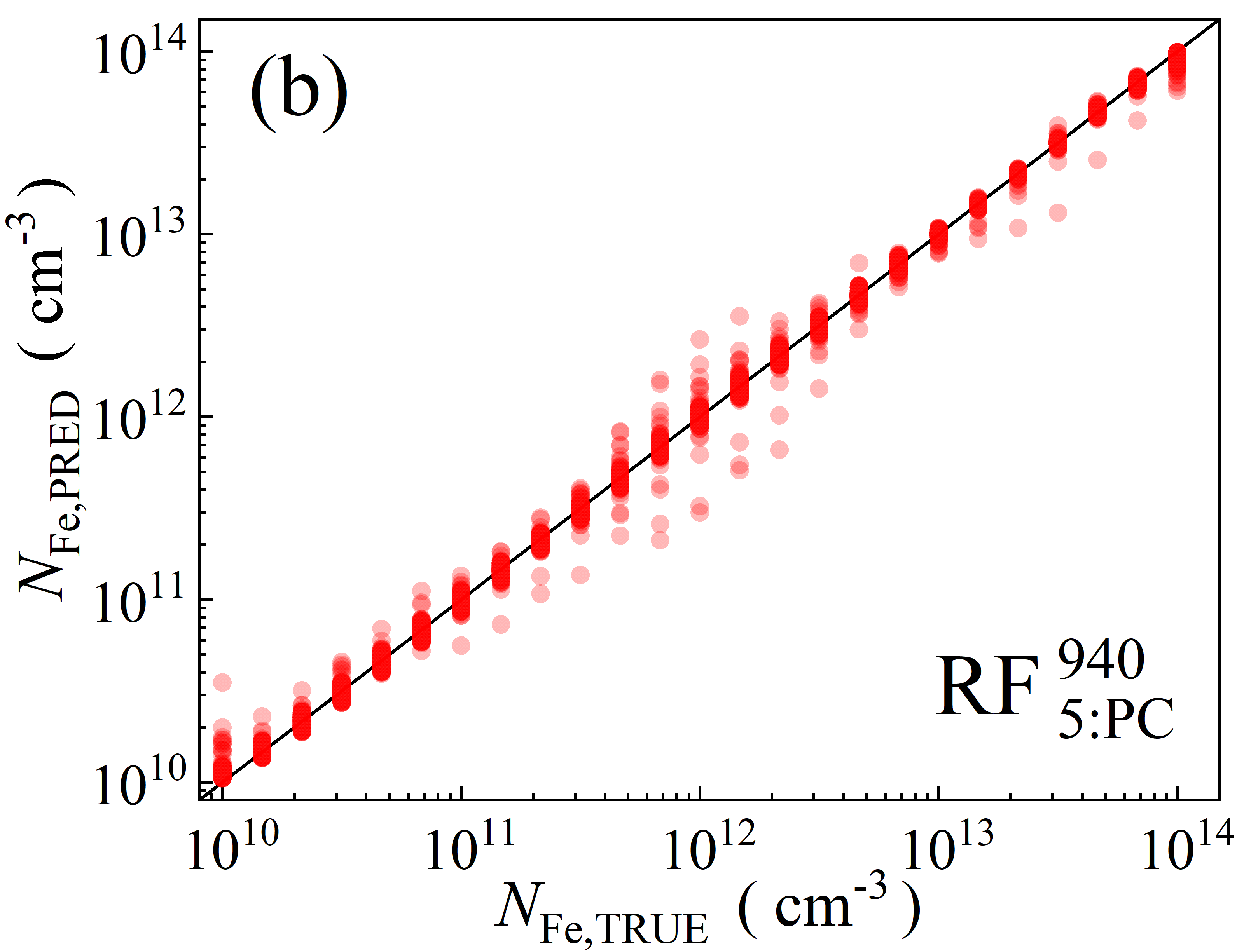
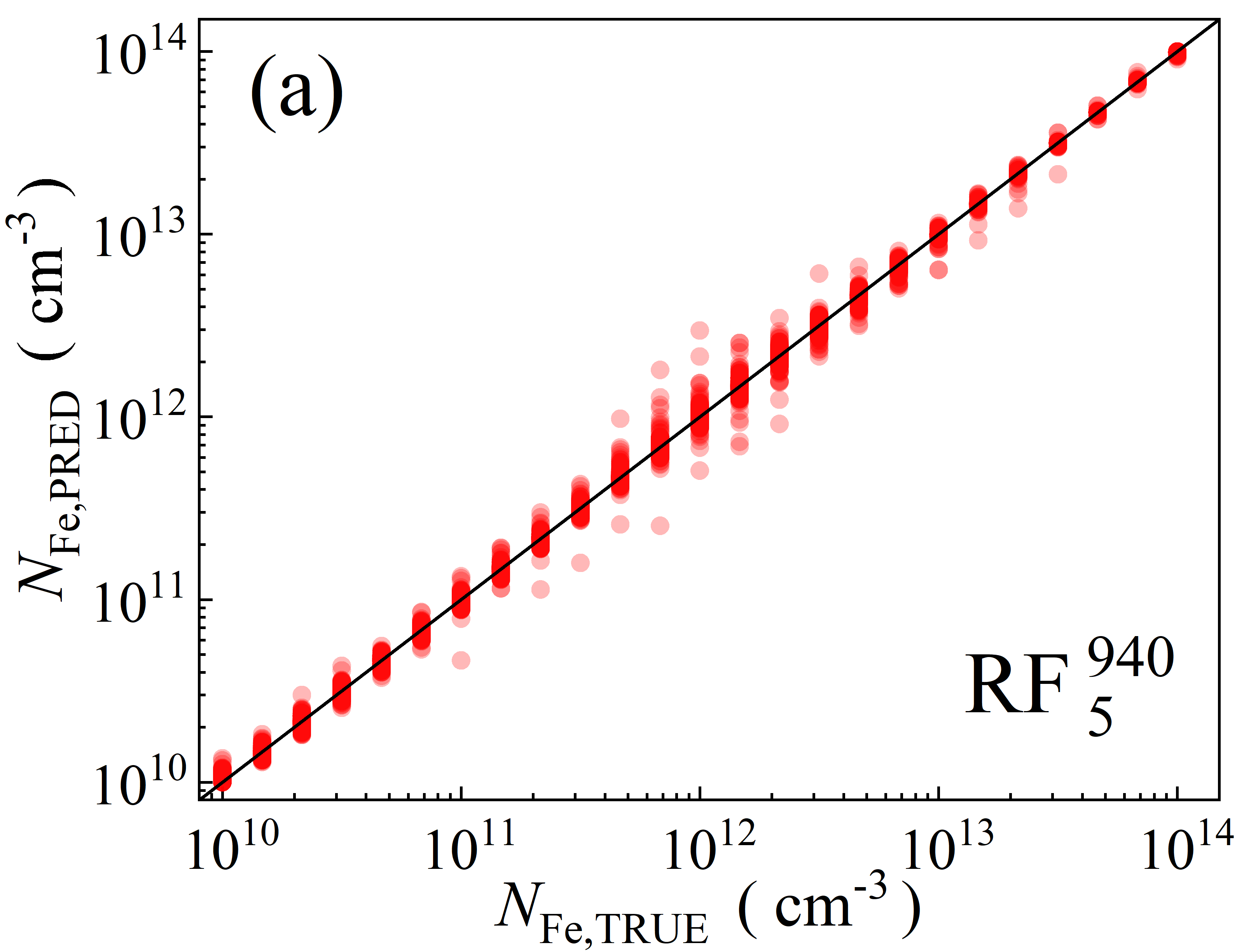
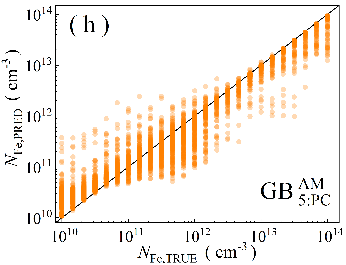
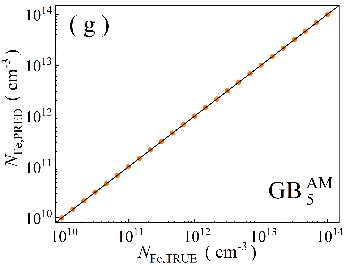
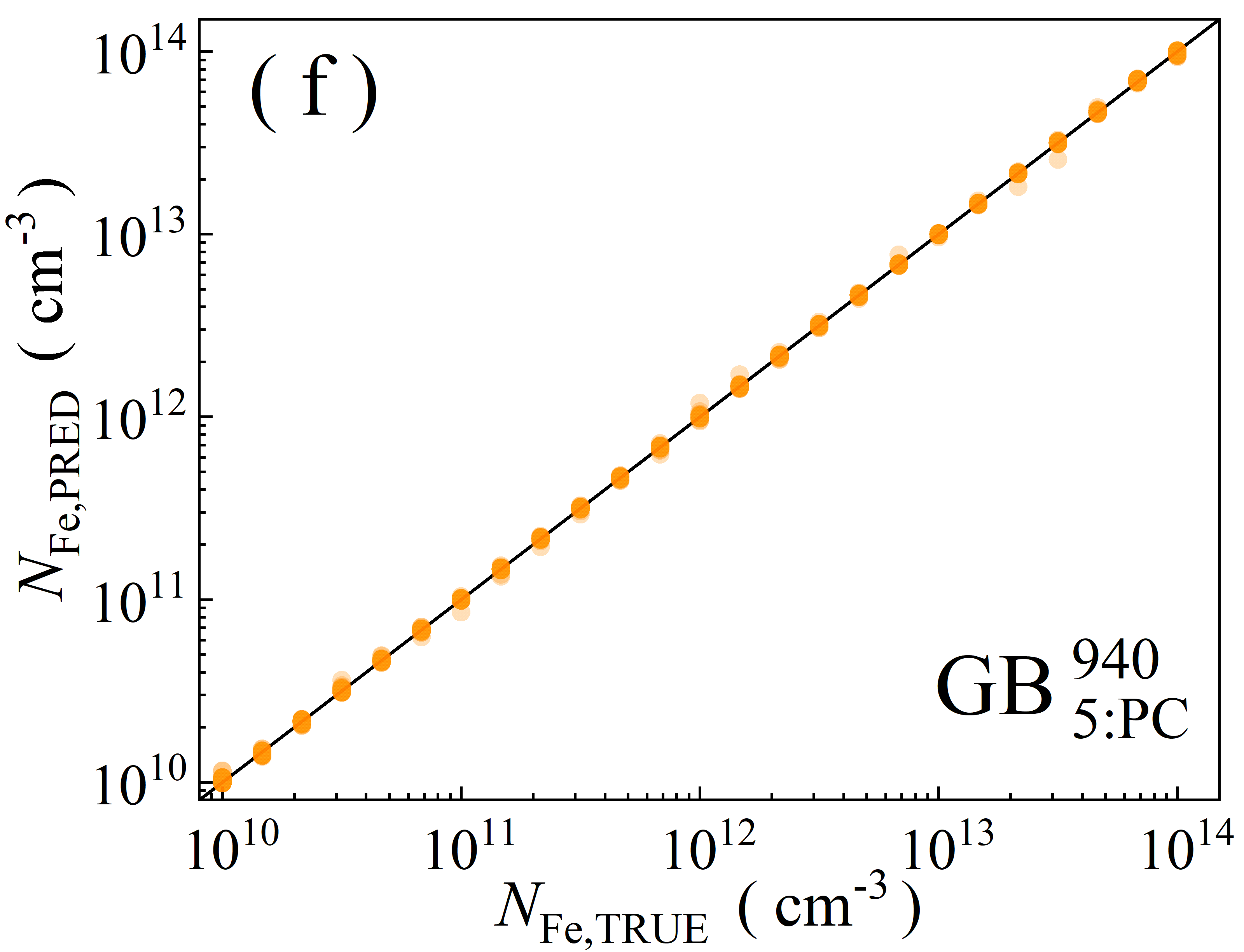
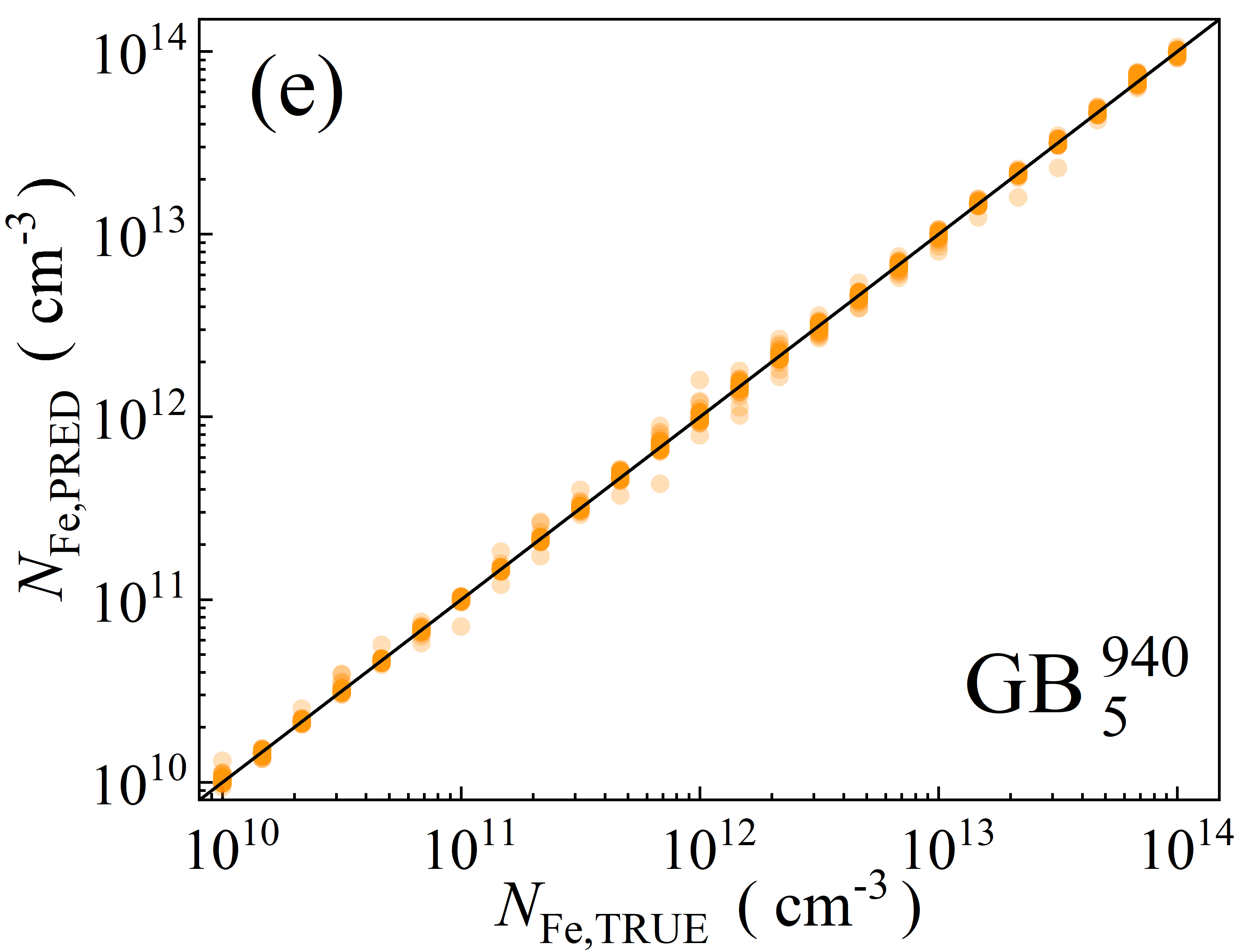
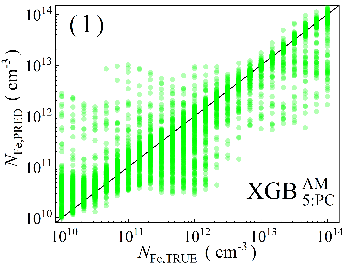
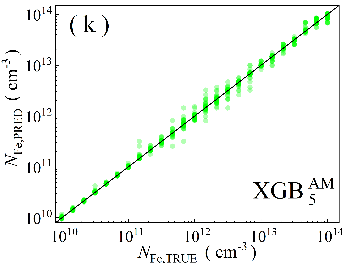
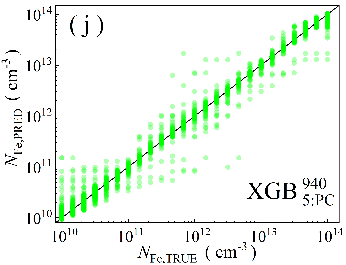
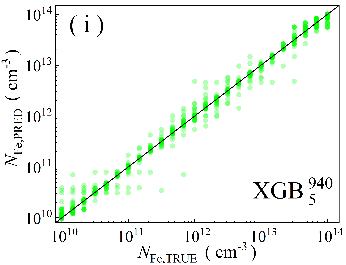
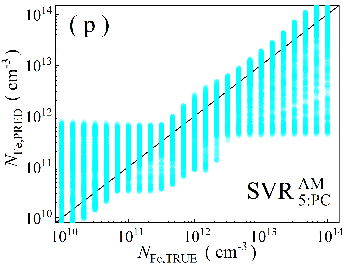
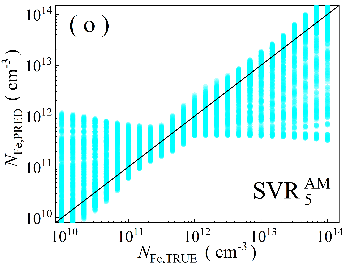
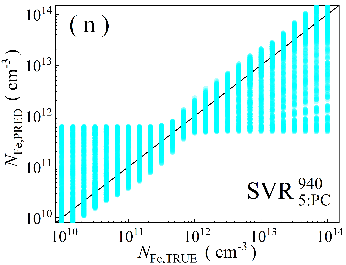
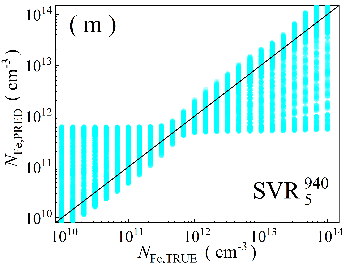


Fig.S3. Scatter plots of the iron concentrations between the reference values and ML predicted values for training phase in the case of 4D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









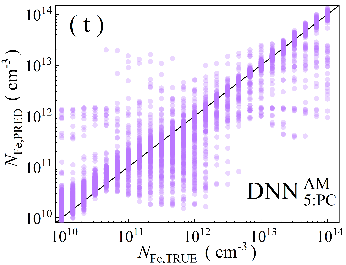
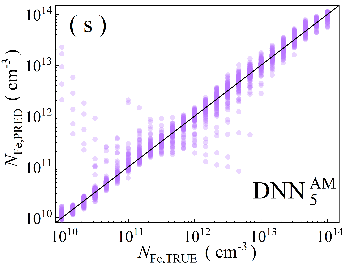
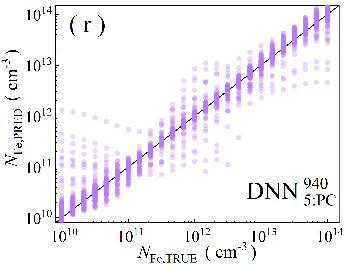
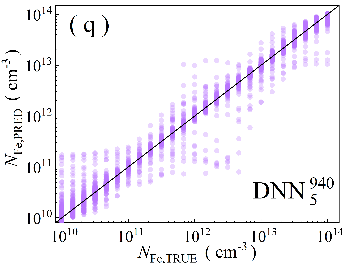
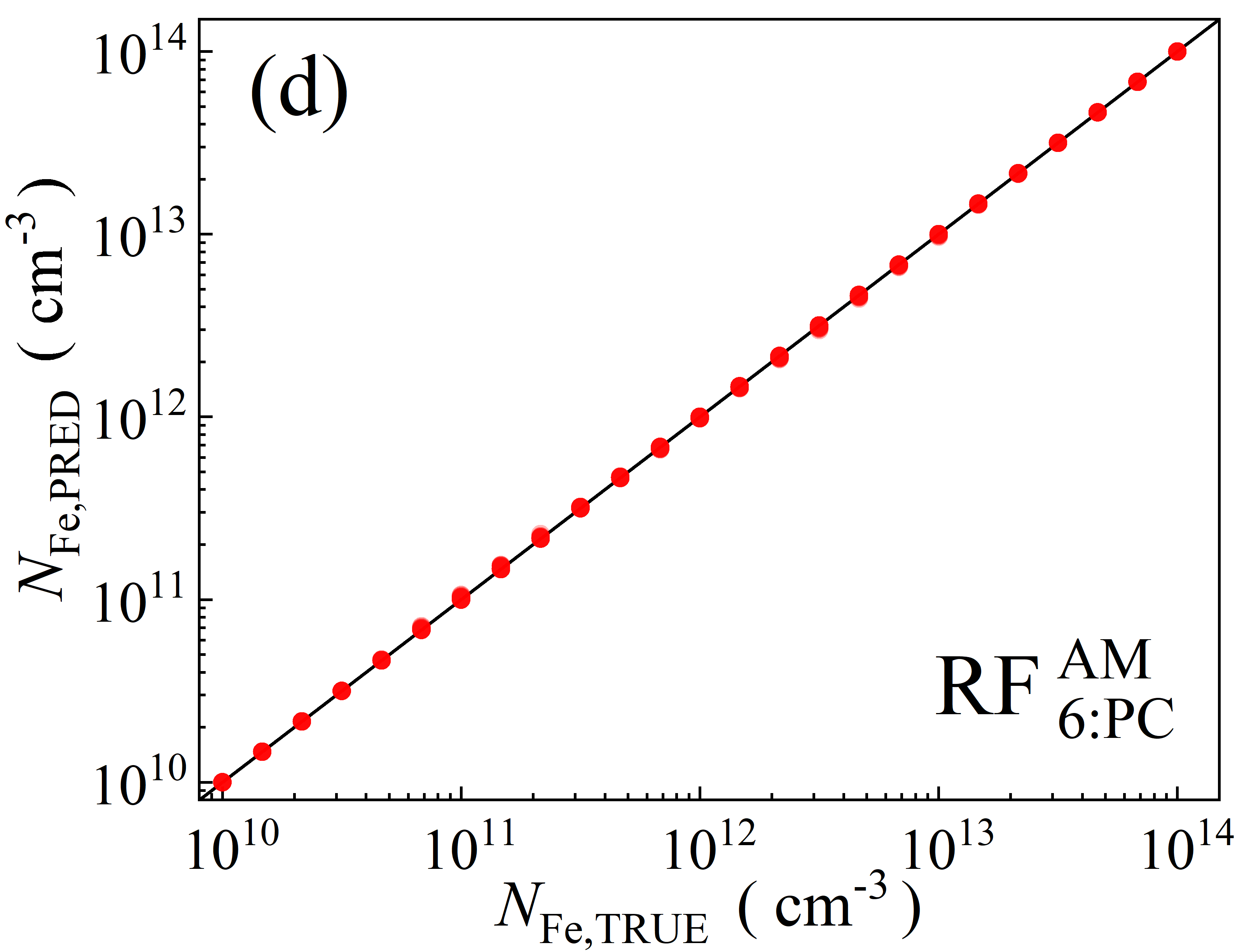
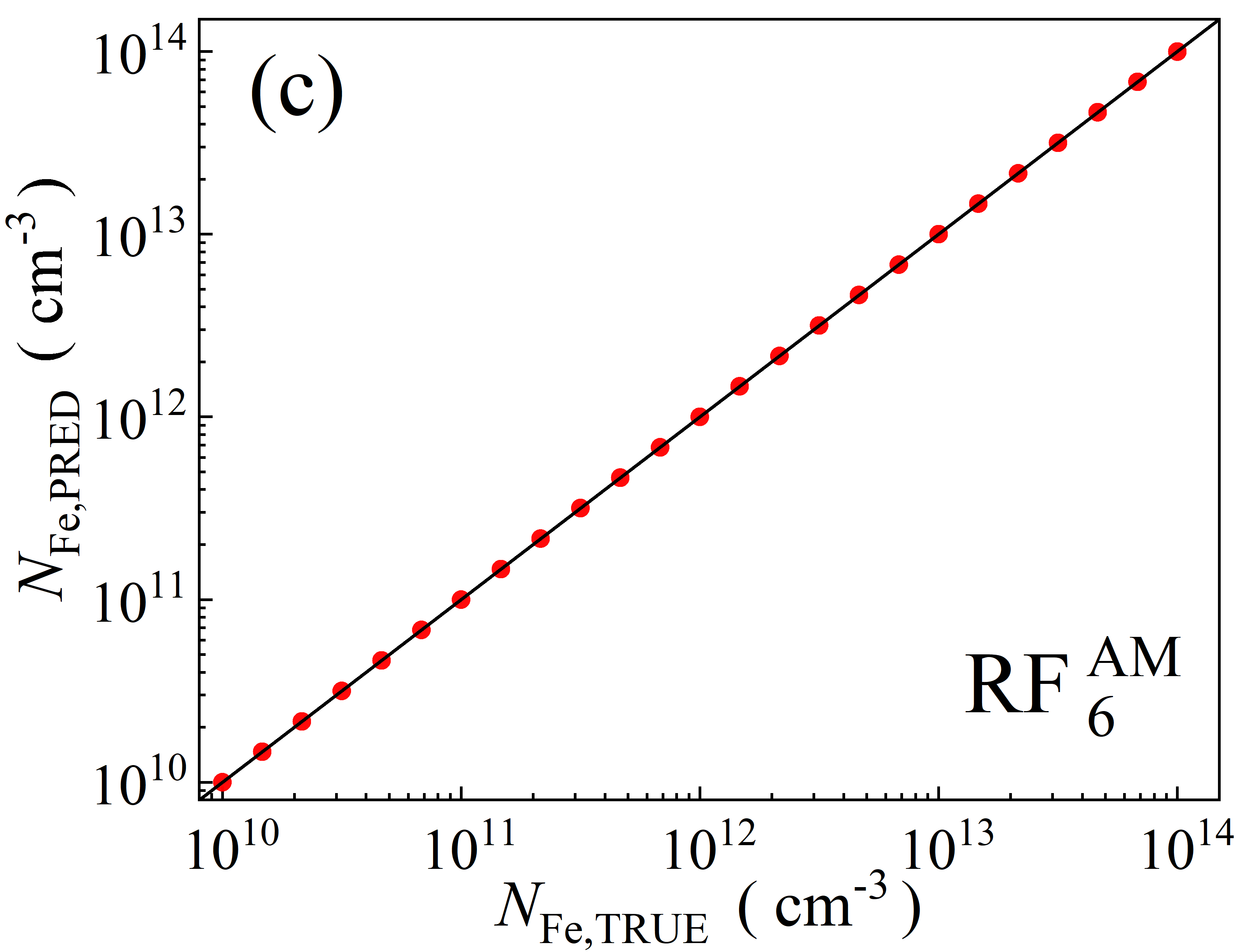
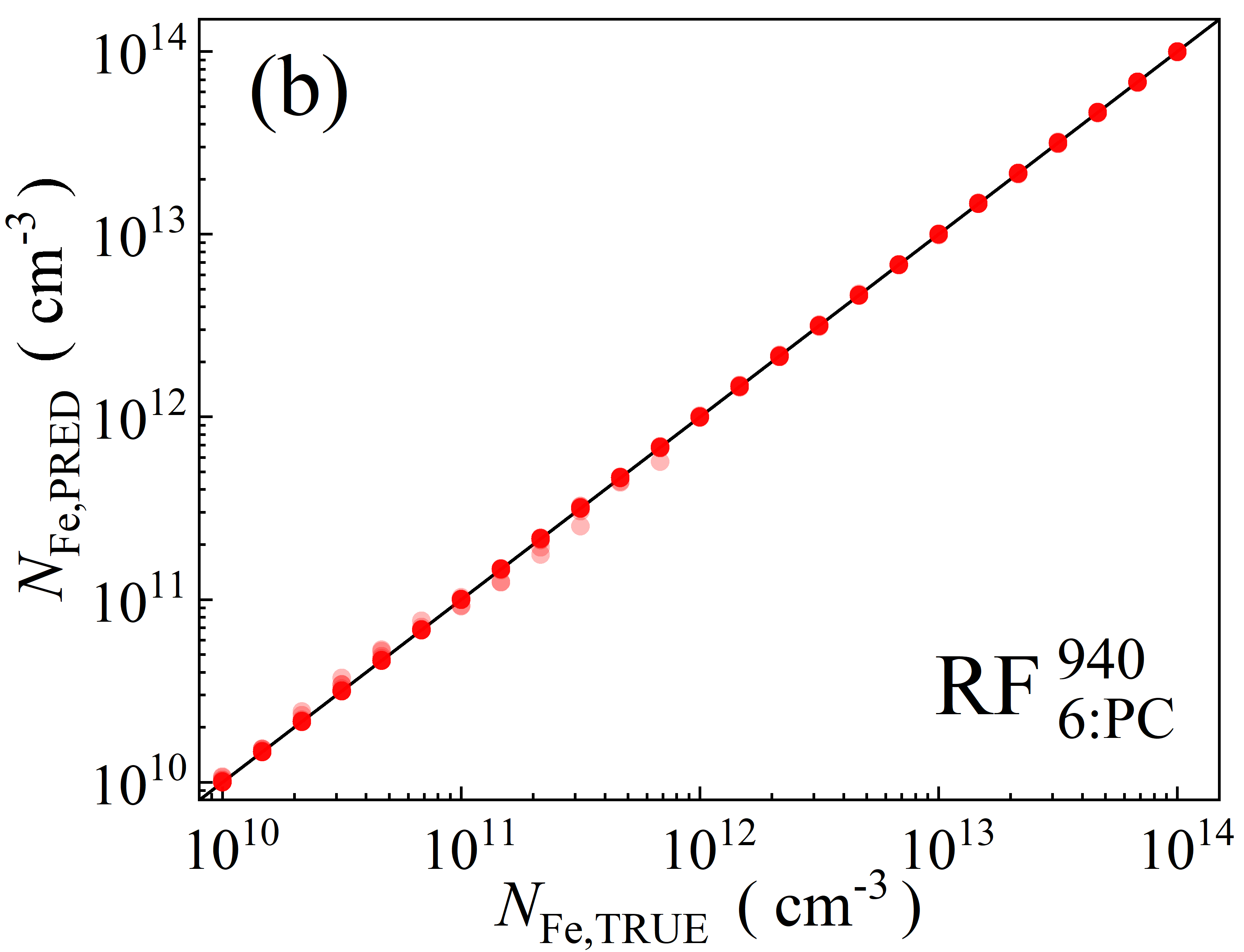
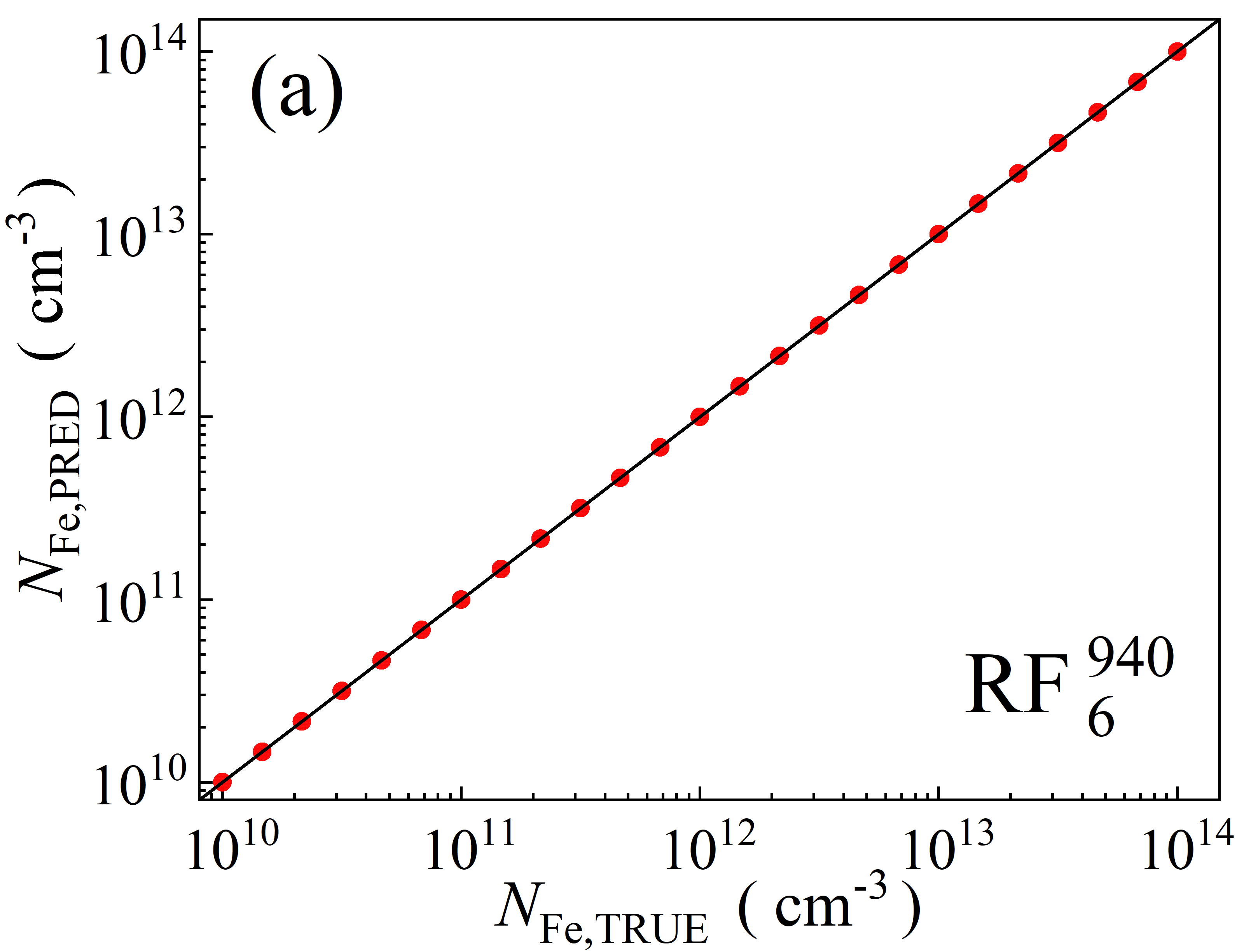
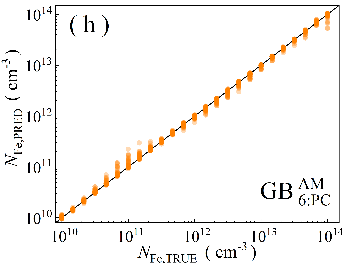
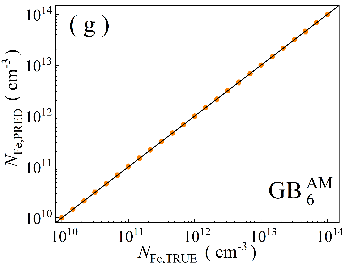
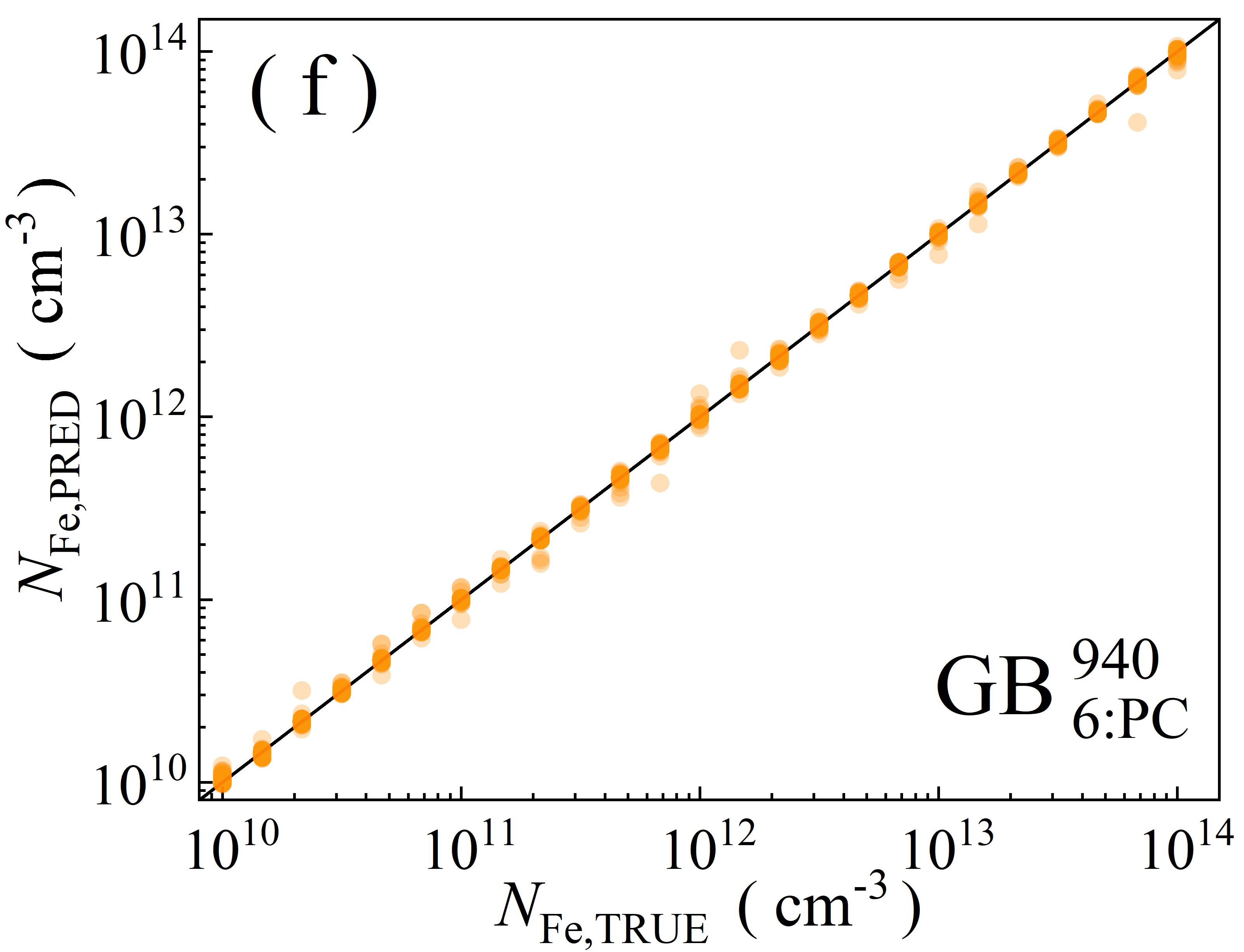
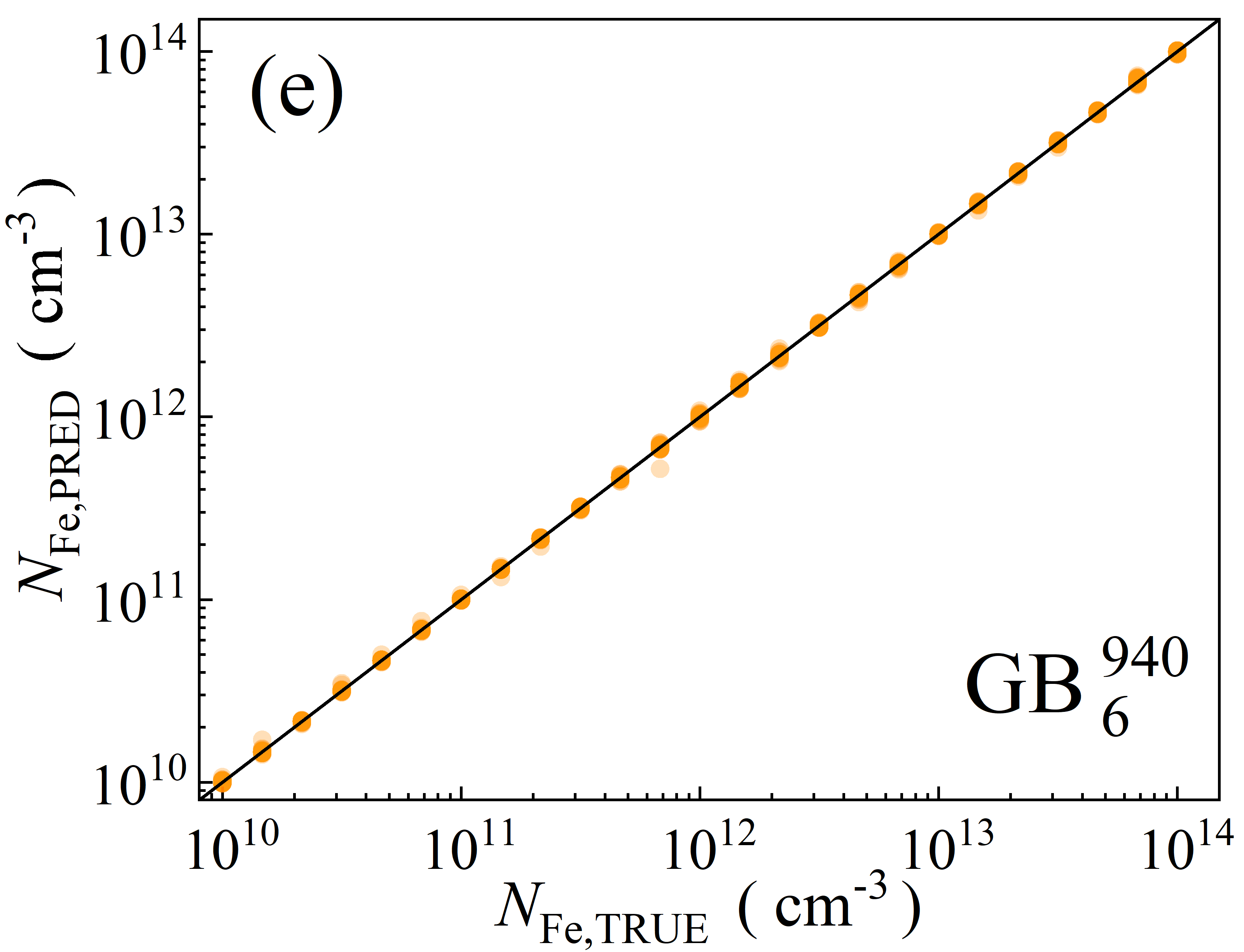
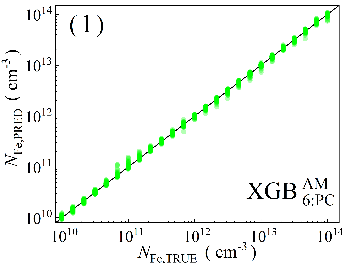
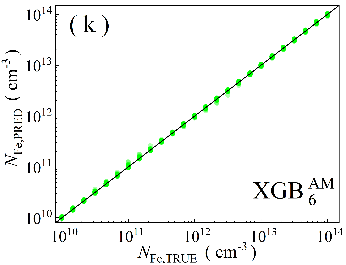
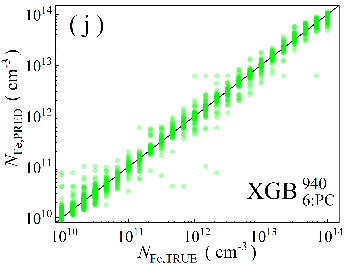
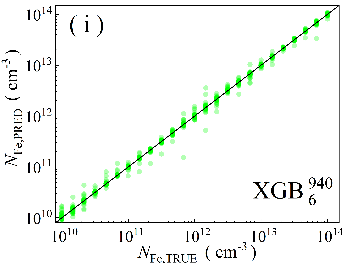
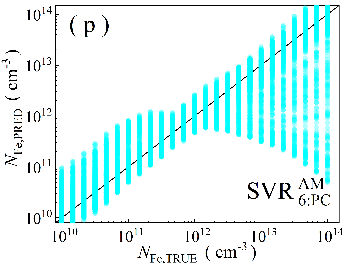
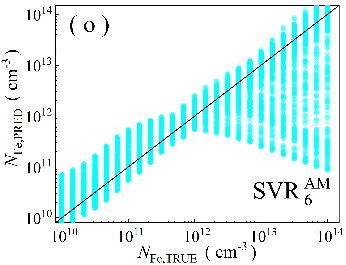
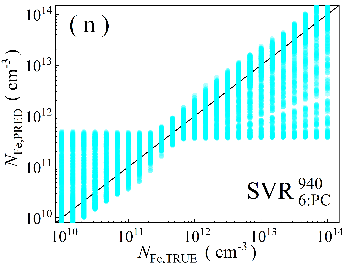


Fig.S4. Scatter plots of the iron concentrations between the reference values and ML predicted values for training phase in the case of 5D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









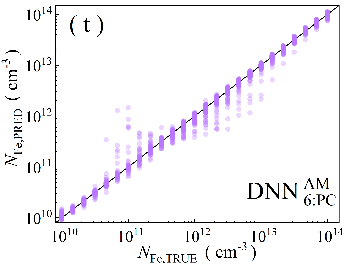
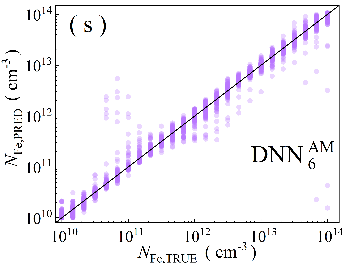
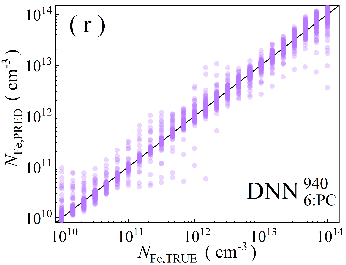
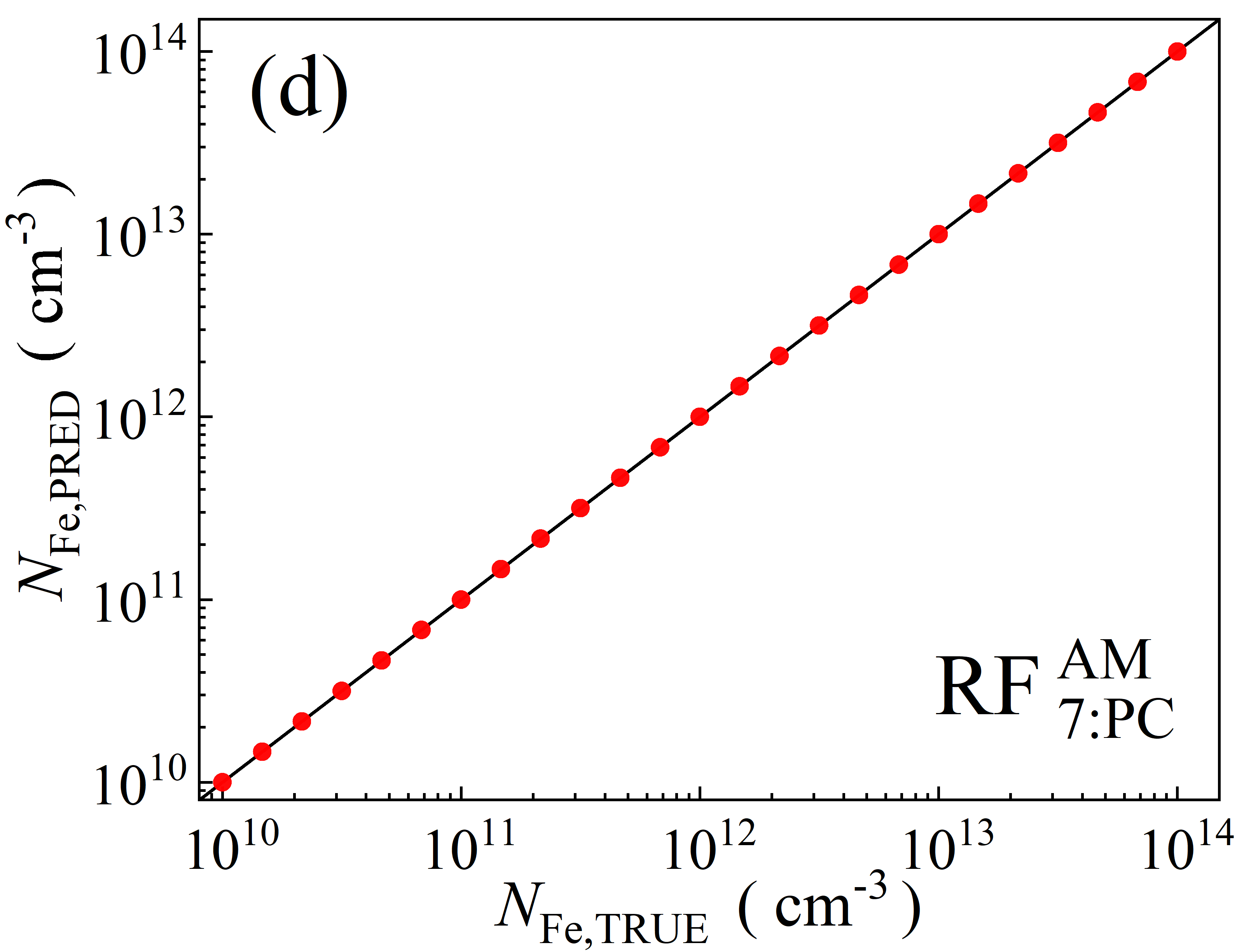
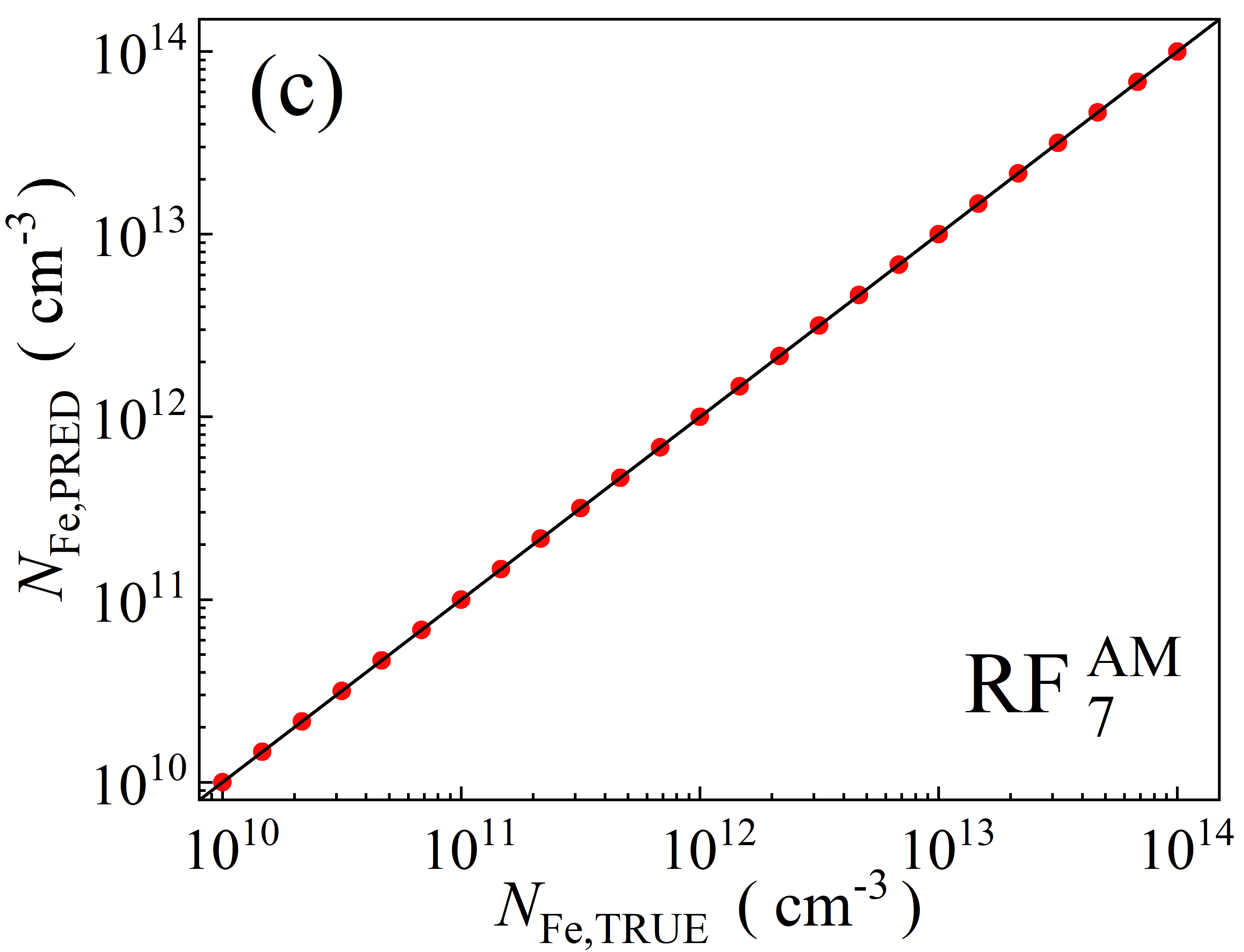
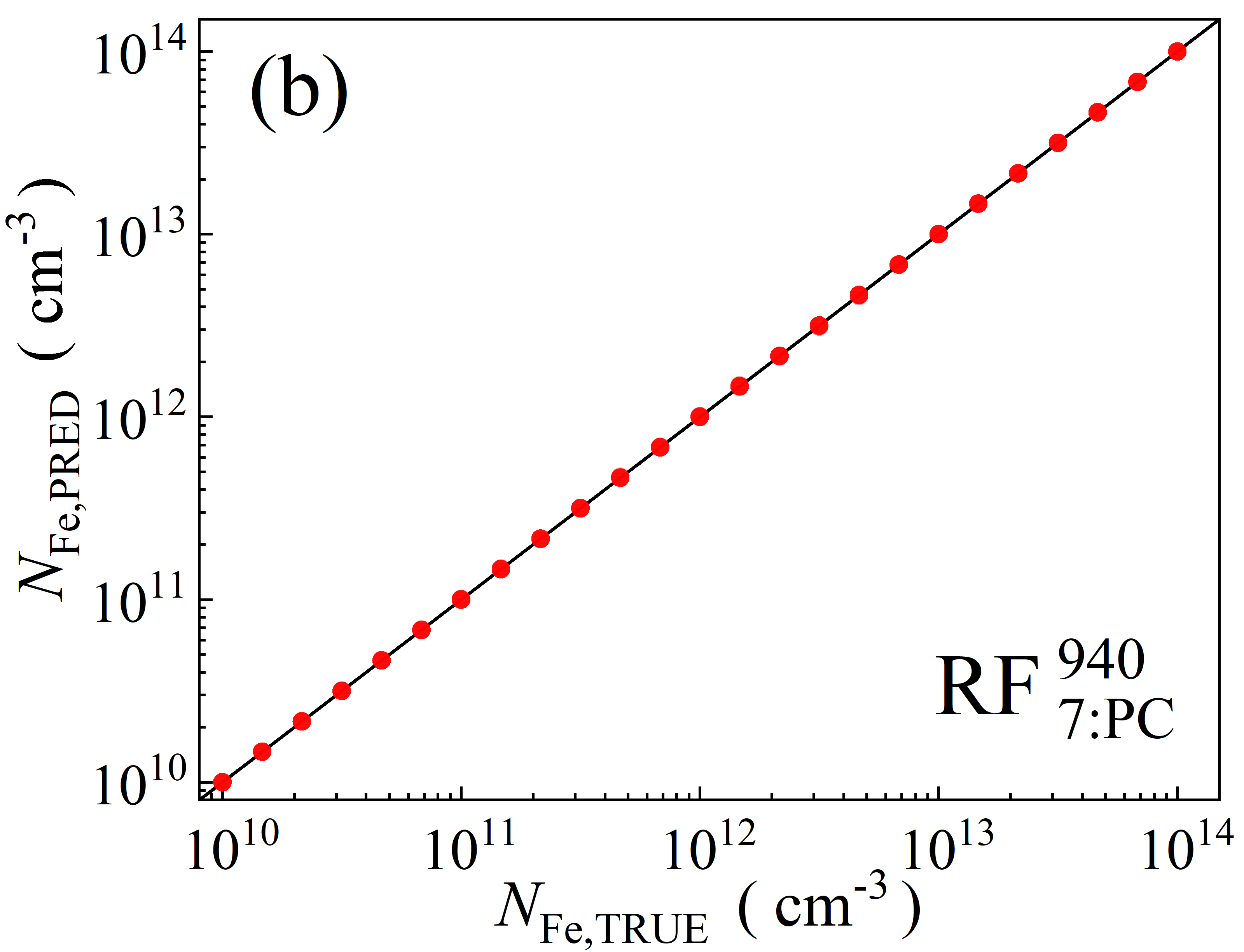
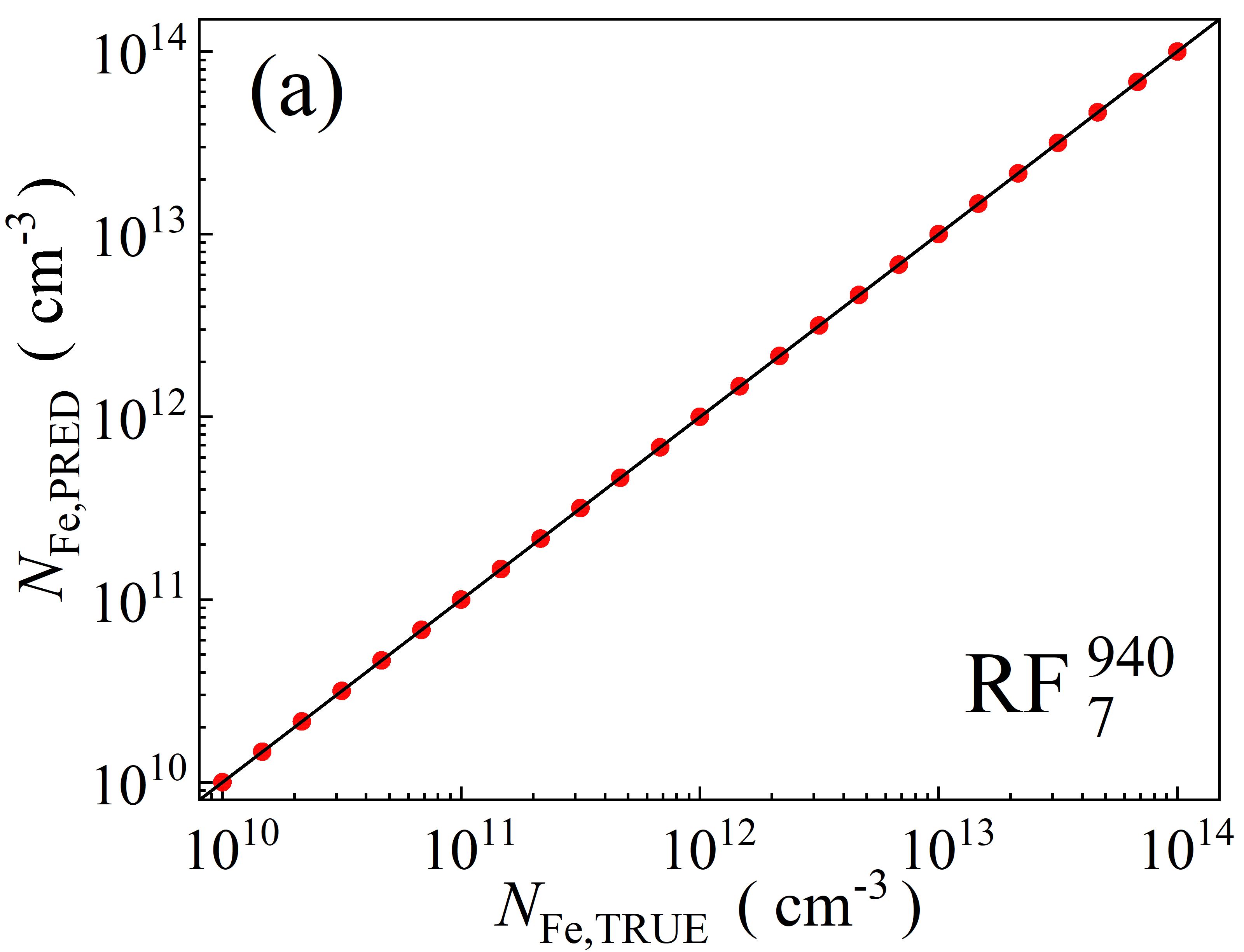
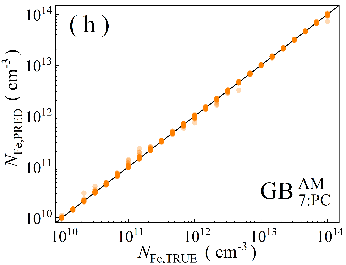
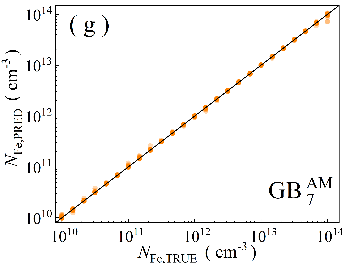
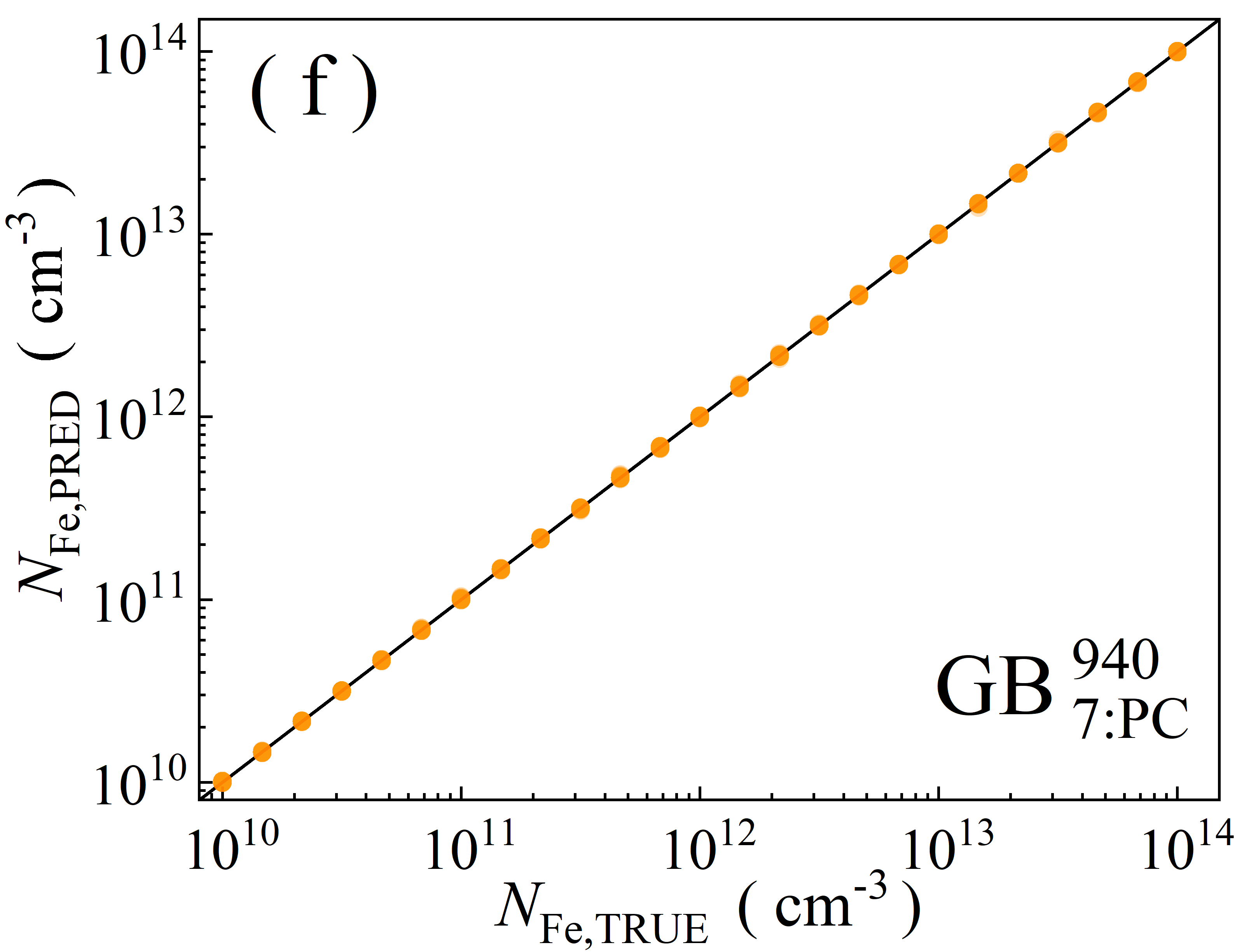
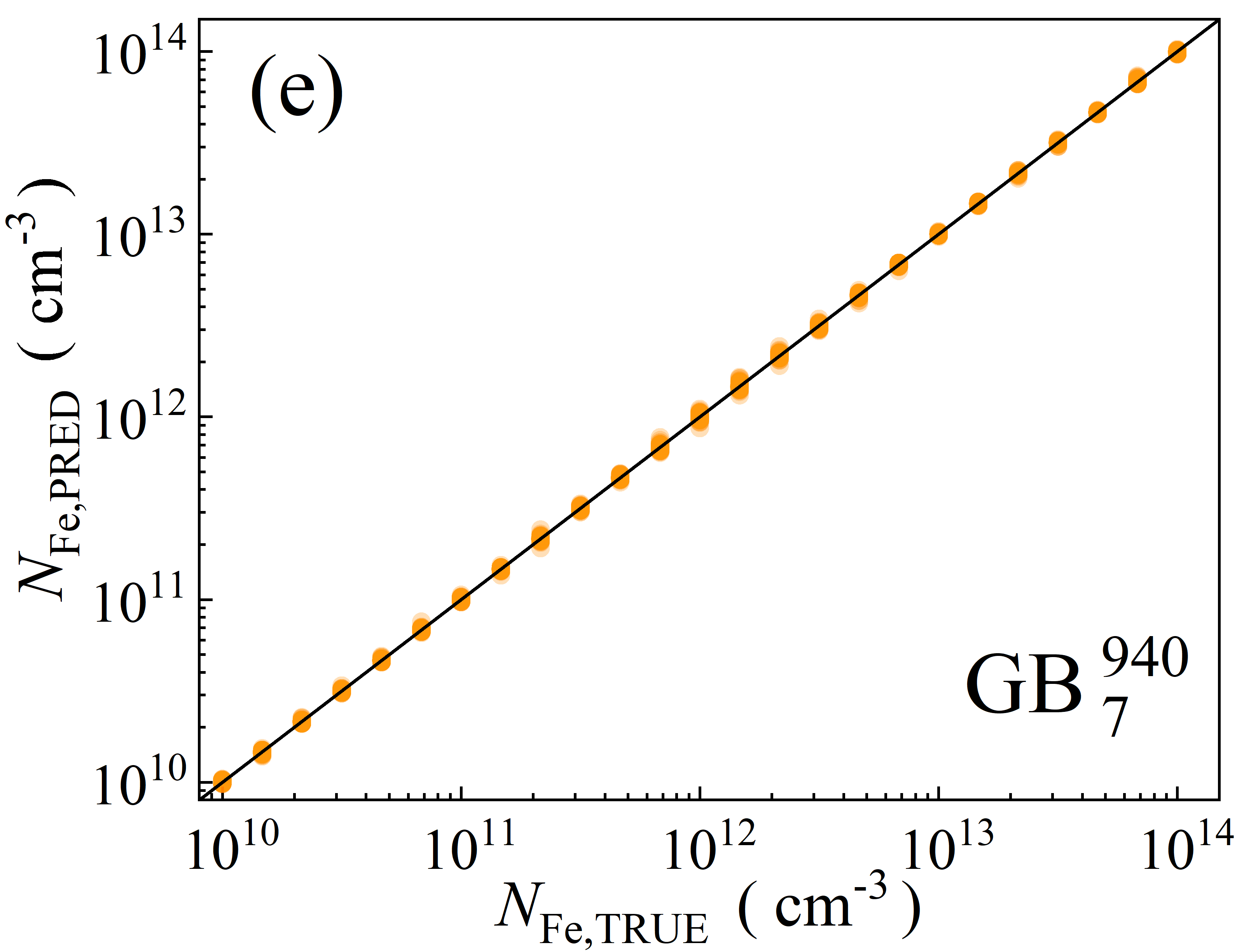
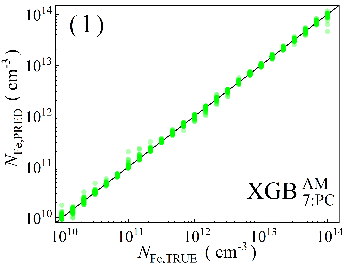
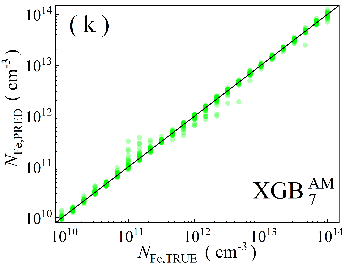
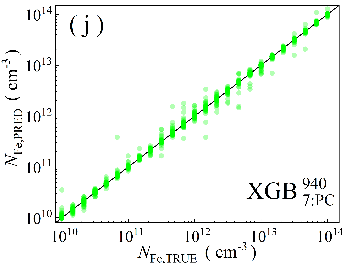
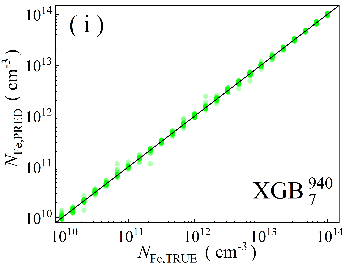
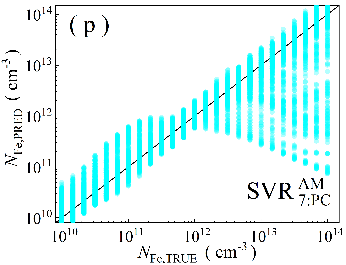
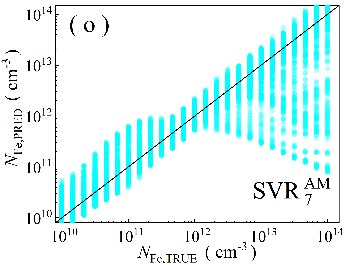
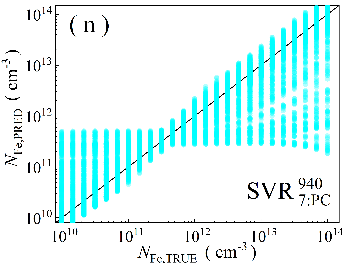
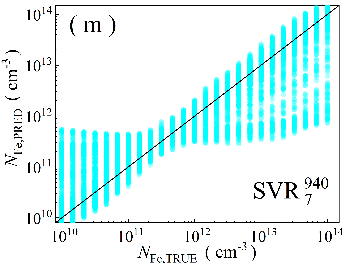


Fig.S5. Scatter plots of the iron concentrations between the reference values and ML predicted values for training phase in the case of 6D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









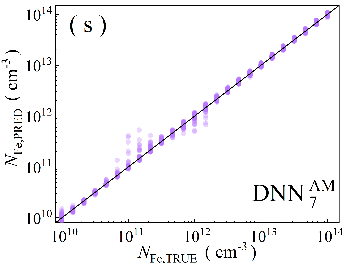
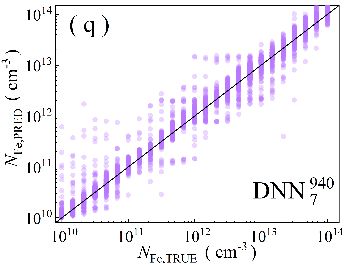


Fig.S6. Scatter plots of the iron concentrations between the reference values and ML predicted values for training phase in the case of 7D features ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.

Table S12. Performance metrics of the models for train dataset. Illumination 940 nm.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | MSE, 10-3 | R2 | MAPE, % | MdAPE, % | *p*01, % | *p*10, % |
|  | 0.60 | 0.9979 | 3.11 | 1.468 | 39.85 | 95.37 |
|  | 1.05 | 0.99639 | 4.20 | 2.134 | 32.19 | 91.22 |
|  | 0.444 | 0.99976 | 2.91 | 1.077 | 48.39 | 94.63 |
|  | 0.662 | 0.9967 | 3.89 | 2.174 | 29.56 | 92.1 |
|  | **0** | **1.000** | 0.002 | **8.2E-13** | **100** | **100** |
|  | 0.0036 | 0.999 | 0.045 | **8.2E-13** | 99.47 | 99.91 |
|  | **0** | **1.000** | **0** | **8.2E-13** | **100** | **100** |
|  | **0** | **1.000** | 0.0026 | **8.2E-13** | 99.96 | **100** |
|  | 0.499 | 0.99725 | 0.160 | 0.5048 | 71.02 | 97.87 |
|  | 0.47 | 0. 9977 | 1.52 | 0.4638 | 72.64 | 97.98 |
|  | 0.0402 | 0.99970 | 0.604 | 0.2315 | 87.26 | 99.47 |
|  | 0.0087 | 0.99993 | 0.36 | 0.1678 | 93.65 | 99.93 |
|  | 0.0058 | 0.999 | 0.26 | 0.1112 | 95.68 | 99.98 |
|  | 0.0370 | 0. 99970 | 0.599 | 0.2469 | 86.68 | 99.56 |
|  | 0.0119 | 0.99992 | 0.50 | 0.2419 | 87.62 | 99.9 |
|  | 0.0007 | 1.000 | 0.107 | 0.05002 | 99.14 | **100** |
|  | 2.282 | 0.98091 | 4.64 | 1.559 | 36 | 92.15 |
|  | 3.929 | 0.9846 | 7.41 | 2.17 | 30.5 | 82.72 |
|  | 0.788 | 0.99490 | 1.94 | 0.617 | 66.3 | 97.31 |
|  | 3.729 | 0.9857 | 6.56 | 1.85 | 32.43 | 88.35 |
|  | 0.174 | 0.99954 | 1.060 | 0.4432 | 76.06 | 98.93 |
|  | 1.97 | 0.9924 | 4.63 | 1.645 | 34.76 | 91.14 |
|  | 0.052 | 0.99983 | 0.84 | 0.4336 | 77.72 | 99.52 |
|  | 0.35 | 0.9992 | 2.01 | 1.156 | 44.52 | 98.1 |
|  | 215 | 0.549 | 201 | 33.24 | 1.358 | 15.48 |
|  | 215 | 0.537 | 200 | 29.55 | 1.673 | 17.46 |
|  | 204 | 0.492 | 171 | 34.94 | 1.552 | 15.05 |
|  | 210 | 0.535 | 177 | 40.48 | 1.139 | 12.77 |
|  | 174 | 0.609 | 124 | 35.86 | 1.608 | 13.71 |
|  | 199 | 0.548 | 142 | 41.63 | 1.277 | 11.81 |
|  | 155 | 0.626 | 112 | 37.34 | 1.188 | 12.42 |
|  | 182 | 0.628 | 122 | 36.34 | 1.349 | 14.22 |
|  | 5 | 0.962 | 12 | 6.787 | 7.717 | 67.28 |
|  | 11 | 0.952 | 17 | 7.833 | 6.804 | 60.34 |
|  | 8 | 0.963 | 12 | 5.977 | 8.727 | 72.91 |
|  | 8 | 0.940 | 17 | 6.983 | 7.572 | 65.73 |
|  | 4 | 0.966 | 9 | 5.205 | 10.26 | 75.68 |
|  | 4 | 0.958 | 9 | 4.611 | 12.86 | 77.29 |
|  | 12 | 0.931 | 24 | 14.4 | 4.186 | 38.54 |
|  | 4 | 0.983 | 8 | 5.109 | 10.03 | 77.92 |

Table S13. Performance metrics of the models for train dataset. Illumination AM1.5.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | MSE, 10-3 | R2 | MAPE, % | MdAPE, % | *p*01, % | *p*10, % |
|  | 15.7 | 0.986 | 16.3 | 2.264 | 37.96 | 78.81 |
|  | 27.6 | 0.9407 | 30.9 | 3.939 | 18.95 | 73.54 |
|  | 0.25 | 0.99993 | 1.0 | 0.1921 | 90.97 | 98.26 |
|  | 43.1 | 0.929 | 44.1 | 3.89 | 24.28 | 68.2 |
|  | 0.0009 | 0.99999 | 0.1 | **8.1E-13** | **100** | **100** |
|  | 0.0008 | **1.000** | 0.0032 | 1.6E-12 | 98.95 | **100** |
|  | **0** | **1.000** | **0** | **1.2E-12** | **100** | **100** |
|  | **0** | **1.000** | **0** | **8.2E-13** | **100** | **100** |
|  | 12.9 | 0.9693 | 16.3 | 4.521 | 10.55 | 81.41 |
|  | 19.4 | 0.9559 | 21.9 | 3.506 | 14.34 | 82.42 |
|  | **0** | 1.000 | 0.008 | 0.002666 | 100 | 100 |
|  | 20.3 | 0.9586 | 23.4 | 6.337 | 7.232 | 70.86 |
|  | 0.0002 | 1.000 | 0.0027 | 0.007811 | 99.77 | 100 |
|  | 0.00142 | 0.99858 | 1.01 | 0.3292 | 80.75 | 98.55 |
|  | 0.026 | 0.99984 | 0.48 | 0.1894 | 89.64 | 99.77 |
|  | 0.0078 | 0. 99971 | 0.197 | 0.08304 | 97.52 | 99.88 |
|  | 24.2 | 0.9465 | 26.9 | 4.362 | 14.63 | 73.26 |
|  | 32.3 | 0.933 | 43.9 | 4.954 | 12.61 | 69.65 |
|  | 1.820 | 0.99777 | 1.158 | 0.541 | 73.43 | 98.59 |
|  | 55.0 | 0.906 | 66.4 | 7.317 | 9.891 | 57.49 |
|  | 0.025 | 0.999 | 0.897 | 0.6342 | 68.15 | 99.85 |
|  | 0.161 | 0.9983 | 2.27 | 1.592 | 33.94 | 98.52 |
|  | 0.17 | 0.9991 | 1.38 | 0.731 | 61.61 | 98.93 |
|  | 0.154 | 0.9979 | 2.04 | 1.389 | 38.34 | 98.63 |
|  | 225 | 0.520 | 209 | 42.62 | 1.131 | 11.71 |
|  | 224 | 0.521 | 209 | 41.86 | 1.366 | 12.08 |
|  | 197 | 0.551 | 181 | 38.96 | 1.244 | 11.14 |
|  | 238 | 0.525 | 213 | 50.62 | 0.897 | 9.939 |
|  | 170 | 0.465 | 56 | 38.36 | 1.261 | 11.85 |
|  | 202 | 0.336 | 74 | 46.59 | 0.9212 | 9.228 |
|  | 133 | 0.444 | 53 | 38.09 | 1.172 | 12.66 |
|  | 133 | 0.504 | 49 | 34.6 | 1.584 | 14.17 |
|  | 31 | 0.936 | 37 | 10.17 | 5.067 | 49.32 |
|  | 47 | 0.870 | 61 | 14.32 | 3.79 | 36.55 |
|  | 8 | 0.987 | 61 | 5.279 | 10.41 | 75.47 |
|  | 48 | 0.941 | 57 | 6.544 | 9.156 | 63.31 |
|  | 6 | 0.979 | 11 | 5.287 | 10.69 | 75.45 |
|  | 2 | 0.995 | 5 | 3.01 | 19.61 | 92.74 |
|  | 0.3 | 0.998 | 3 | 1.747 | 29.81 | 98 |
|  | 0.4 | 0.998 | 3 | 2.146 | 24.64 | 97.9 |

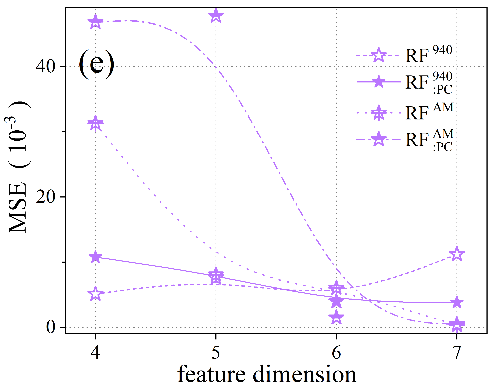
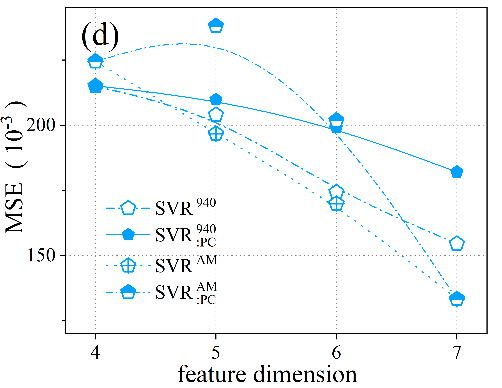
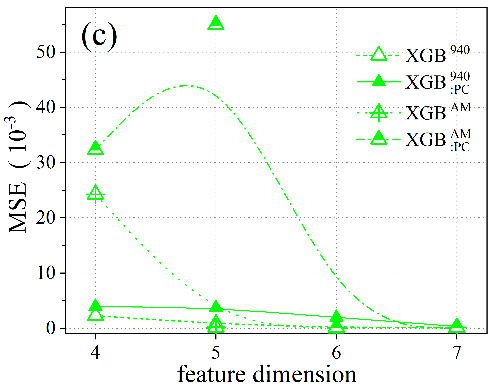
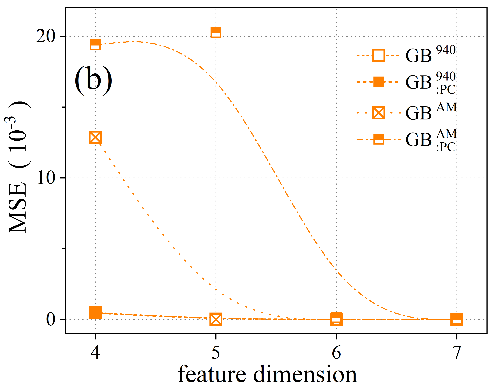
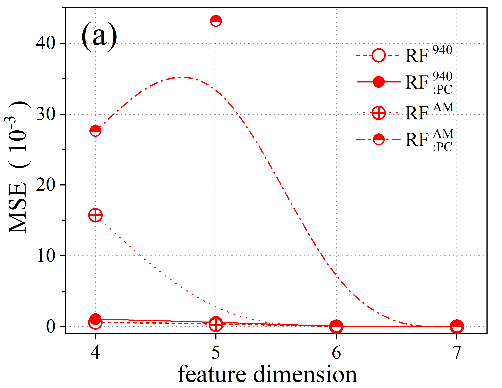
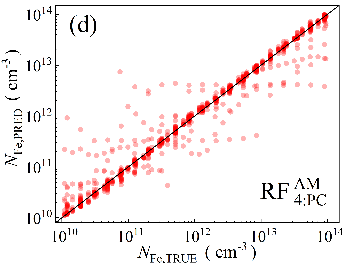
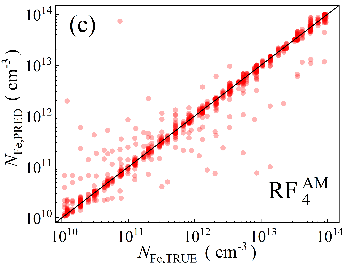
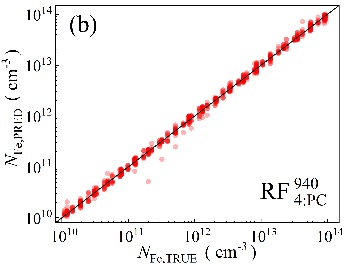
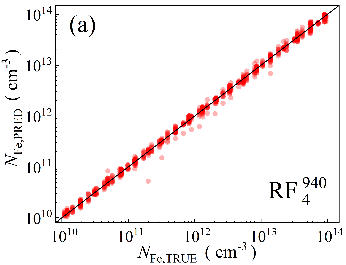
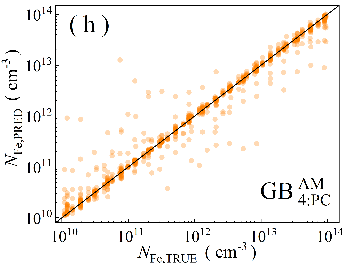
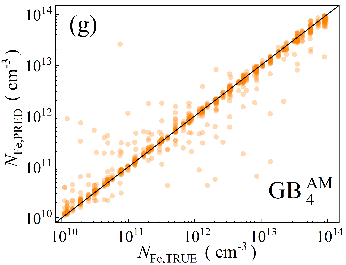
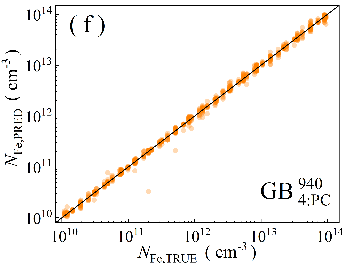
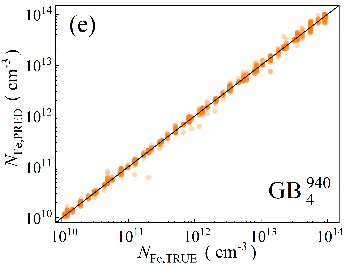
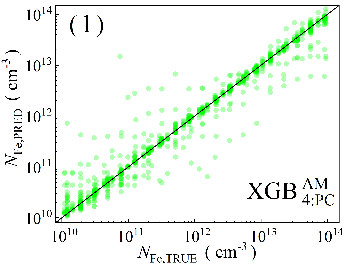
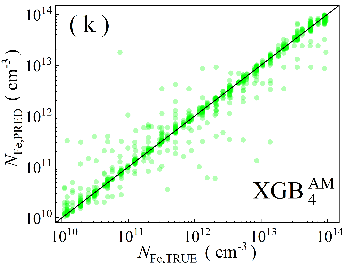
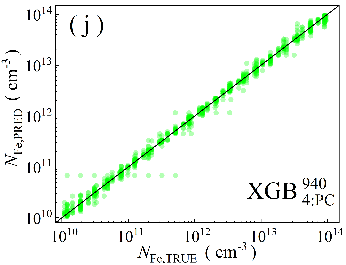
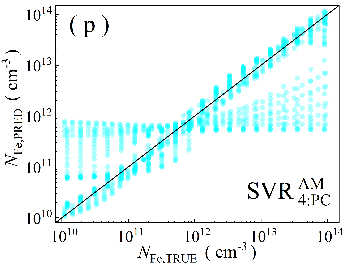
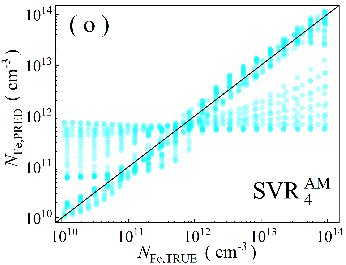
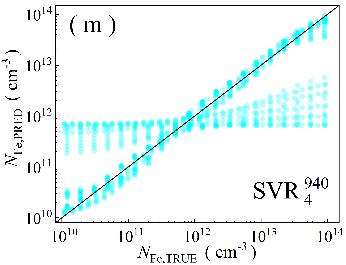


Fig.S7. Dependencies of MSE on the input feature dimension for train data.









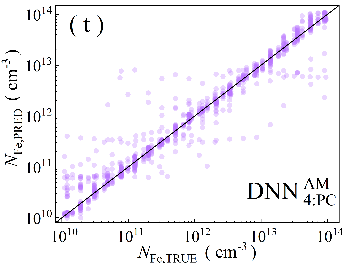
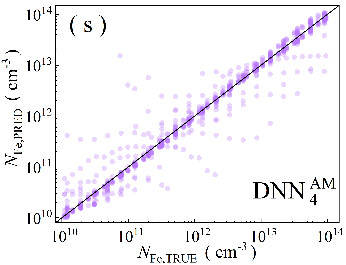
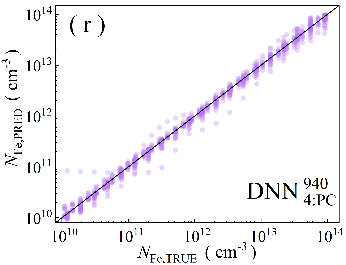
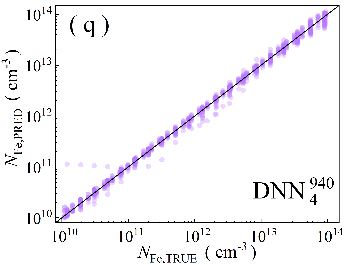
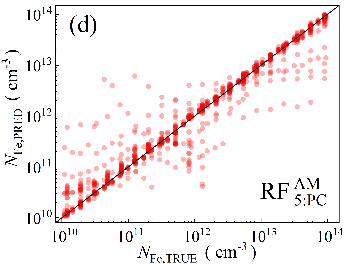
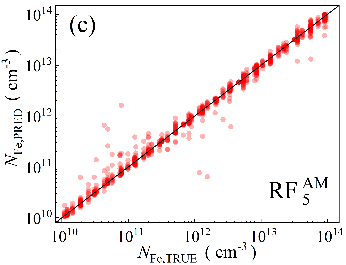
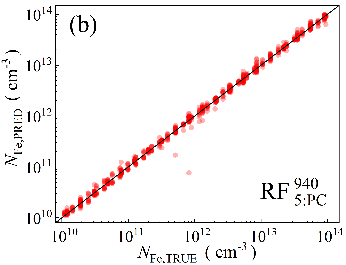
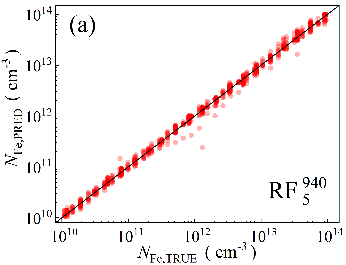
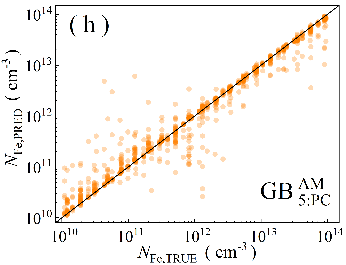
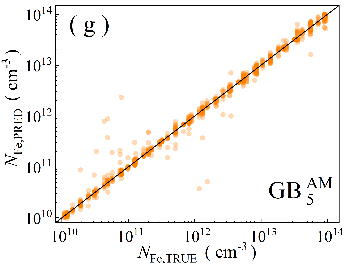
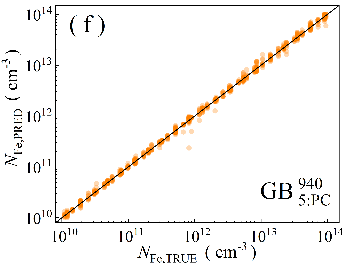
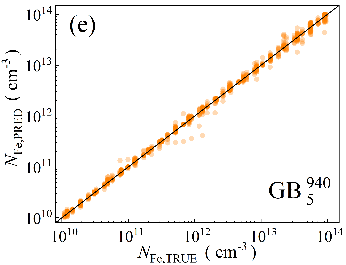
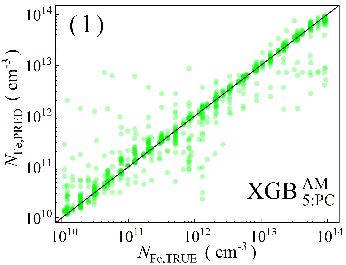
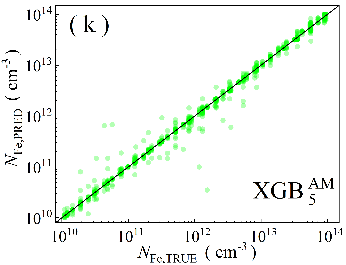
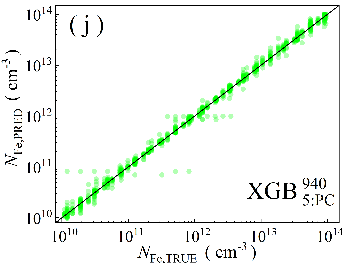
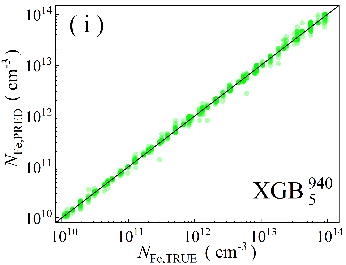
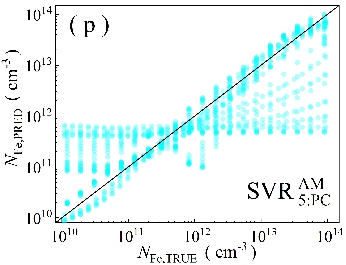
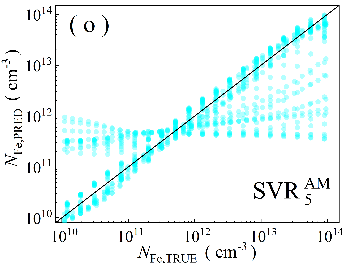
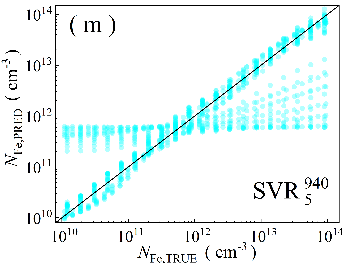


Fig.S8. Scatter plots of the iron concentrations between the reference values and ML predicted values for *N*Fe-altered test phase in the case of 4D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









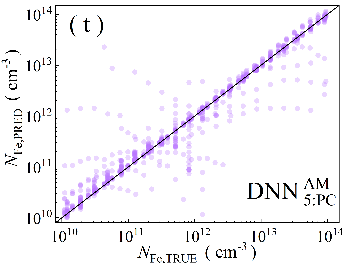
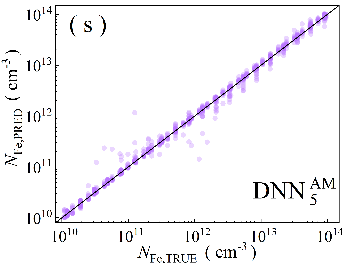
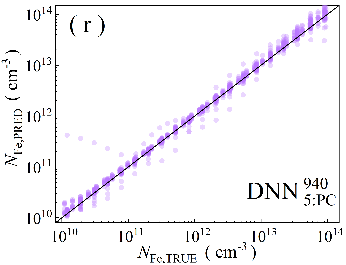
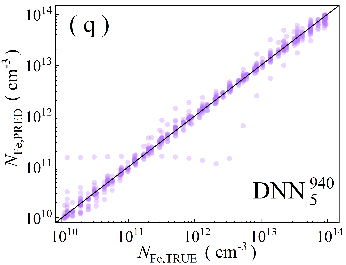
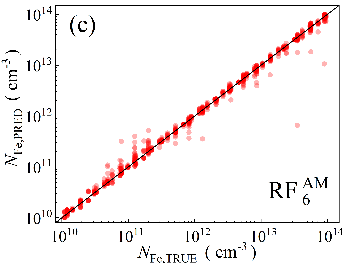
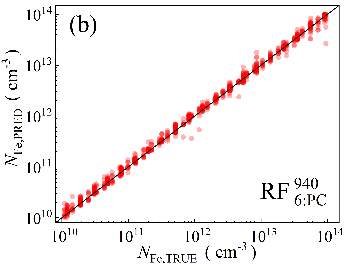
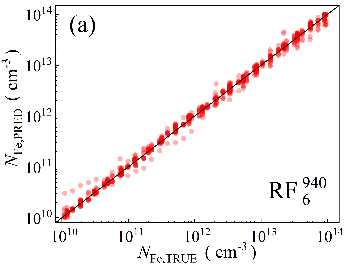
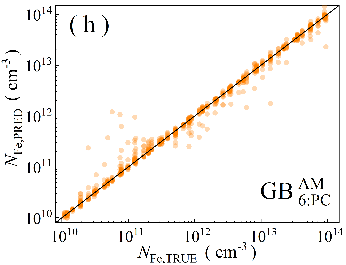
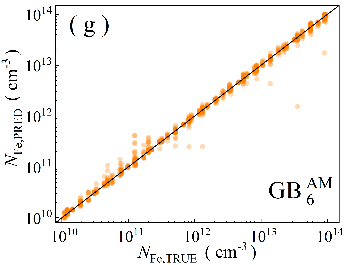
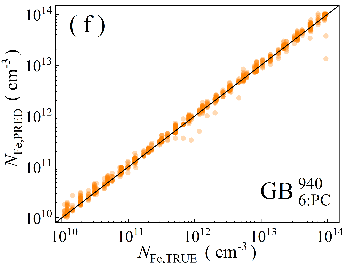
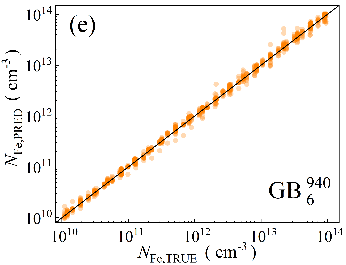
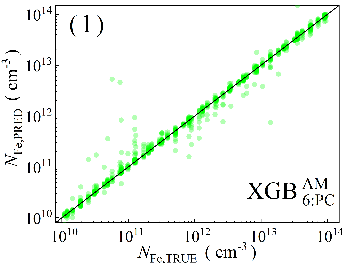
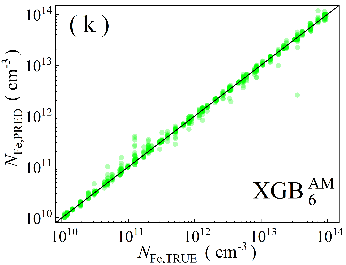
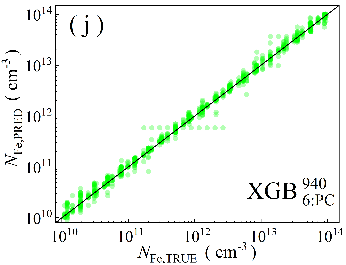
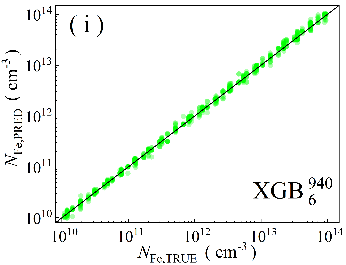
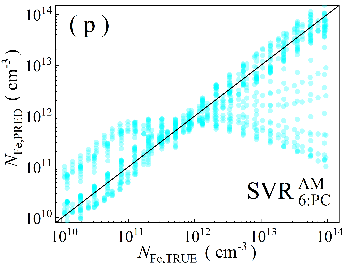
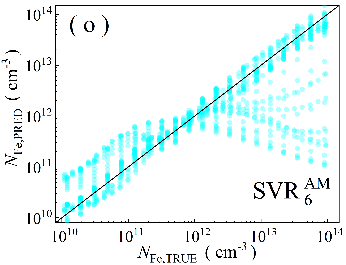
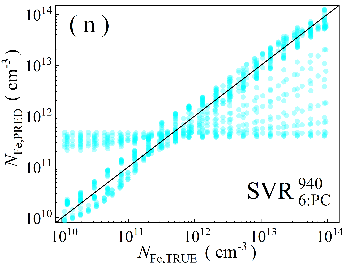
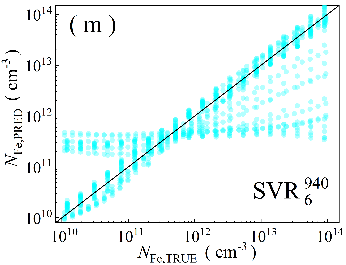


Fig.S9. Scatter plots of the iron concentrations between the reference values and ML predicted values for *N*Fe-altered test phase in the case of 5D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









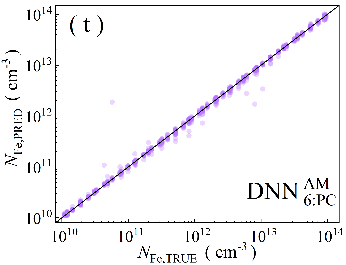
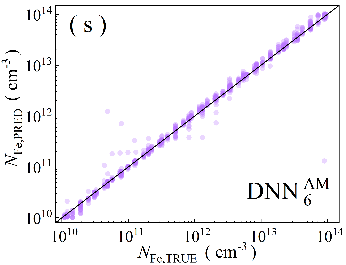
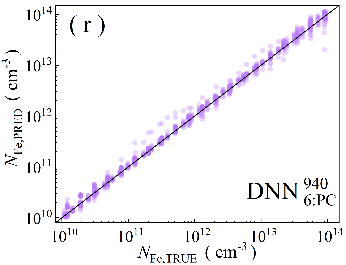
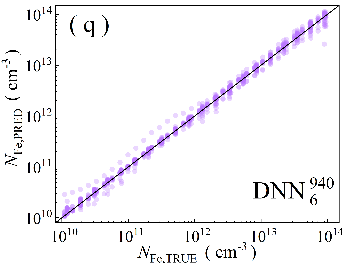
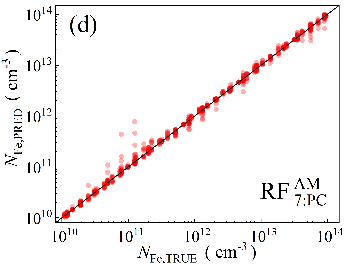
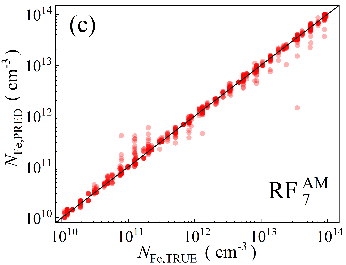
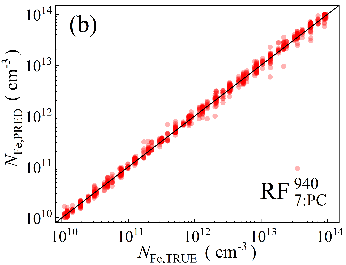
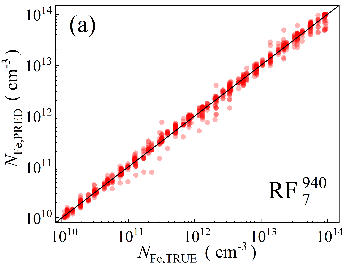
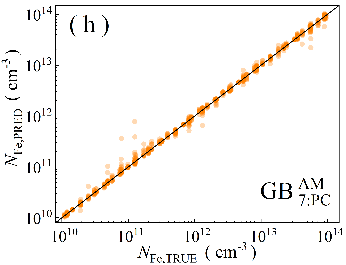
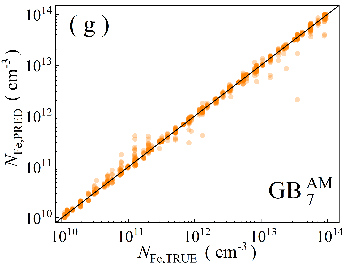
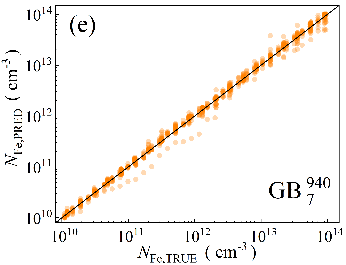
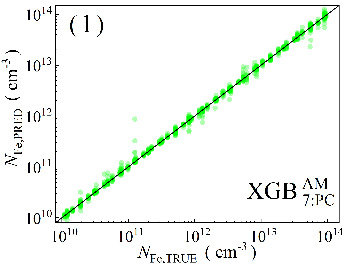
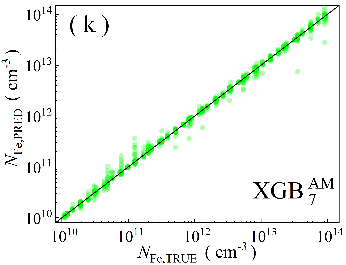
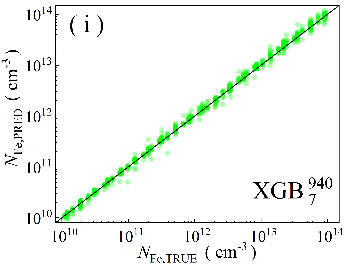
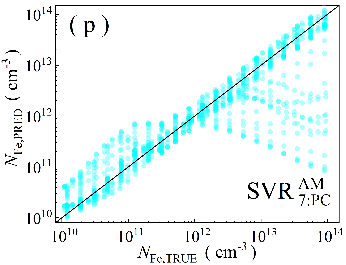
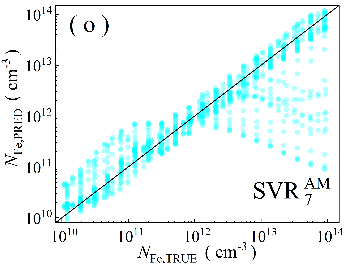
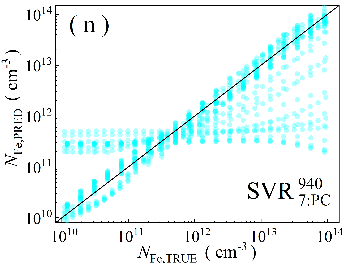
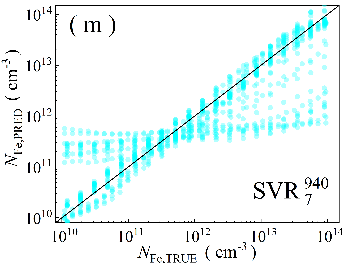


Fig.S10. Scatter plots of the iron concentrations between the reference values and ML predicted values for *N*Fe-altered test phase in the case of 6D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









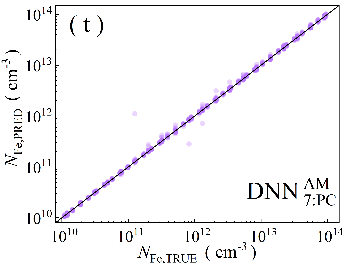
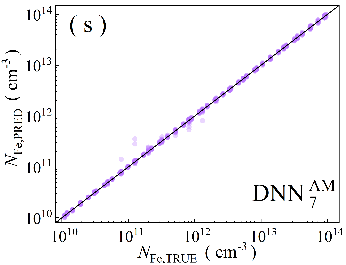
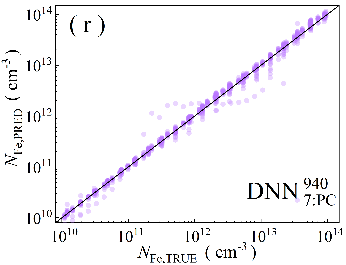
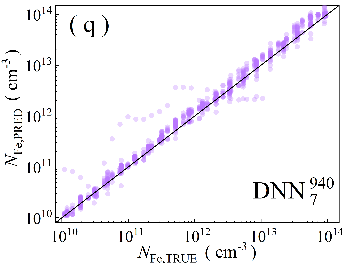
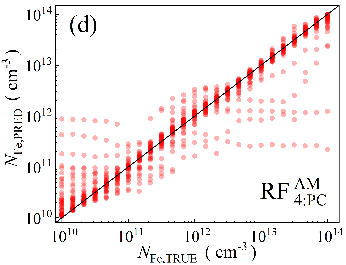
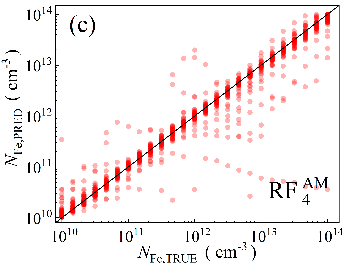
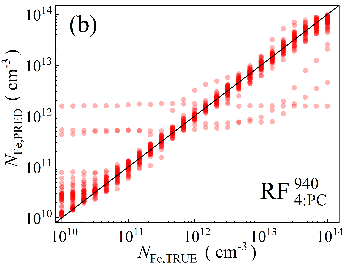
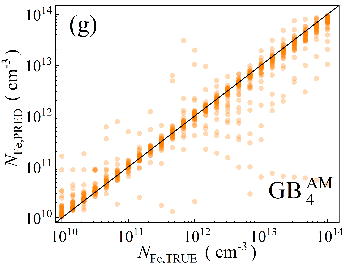
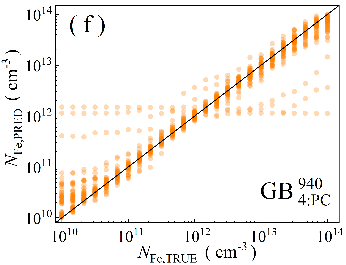
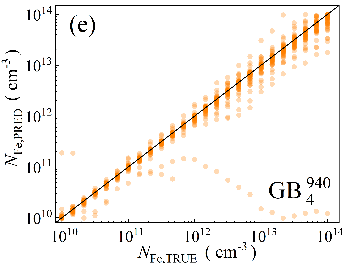
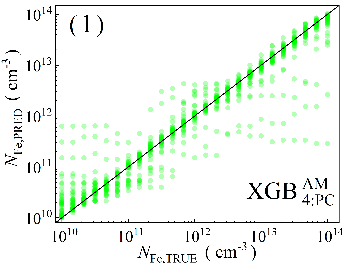
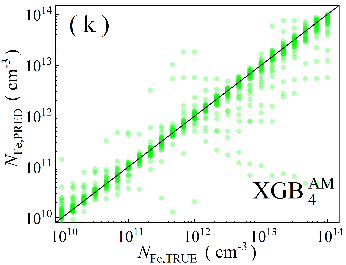
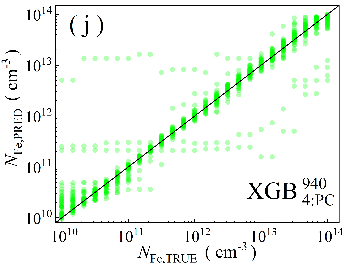
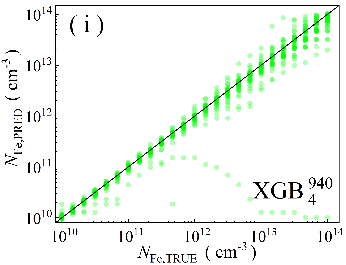
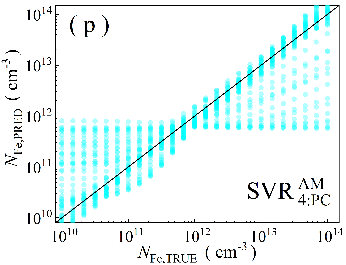
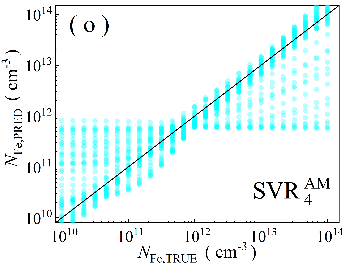
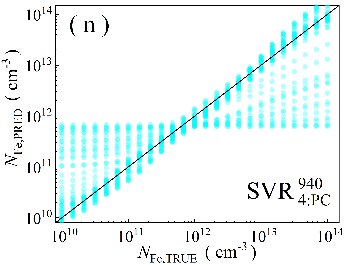
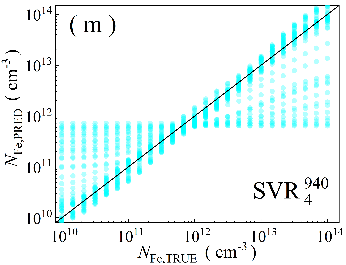


Fig.S11. Scatter plots of the iron concentrations between the reference values and ML predicted values for *N*Fe-altered test phase in the case of 7D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









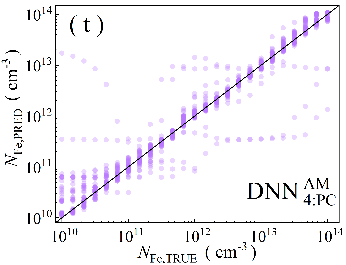
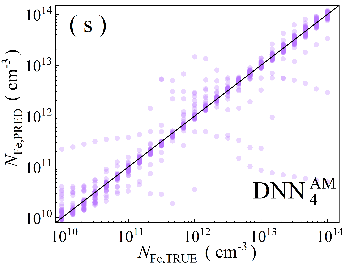
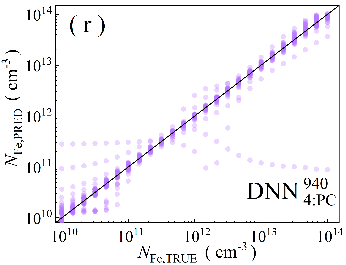
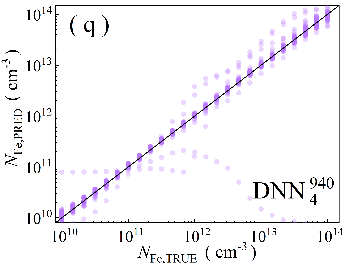
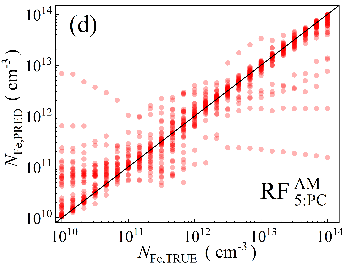
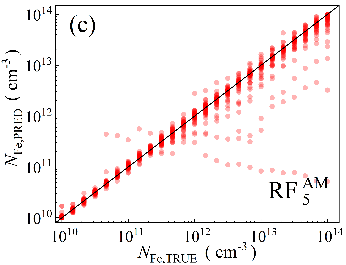
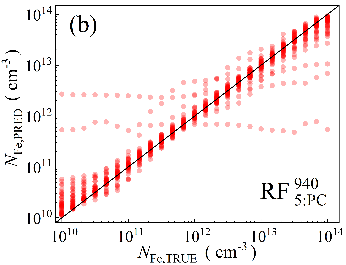
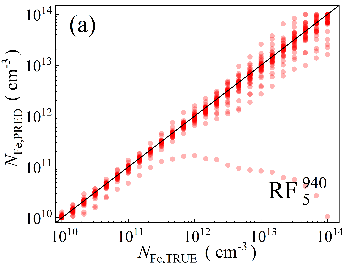
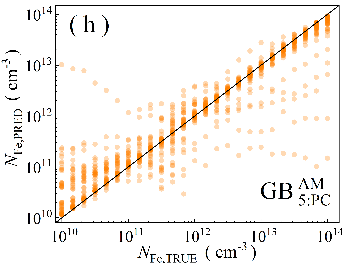
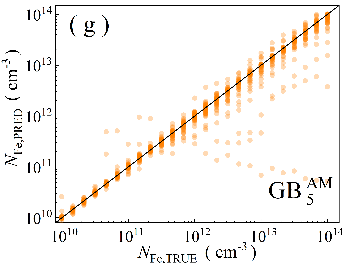
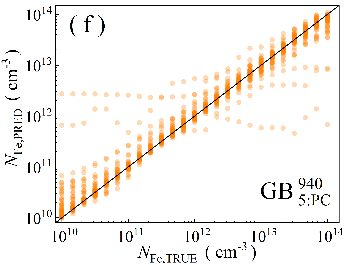
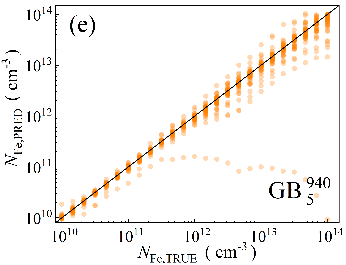
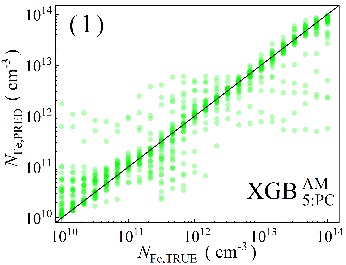
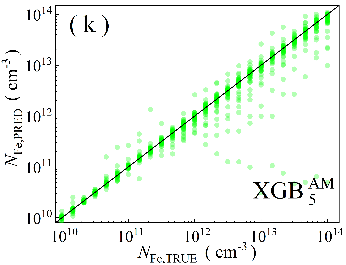
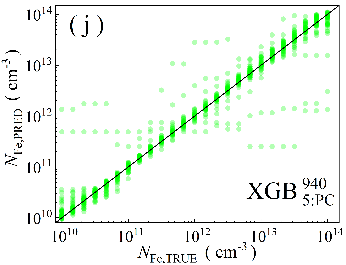
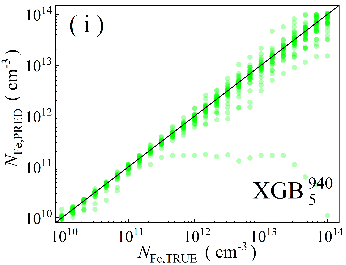
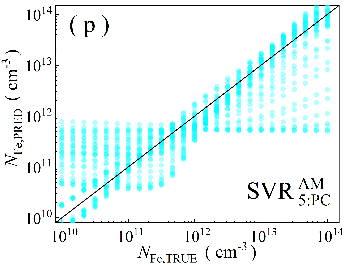
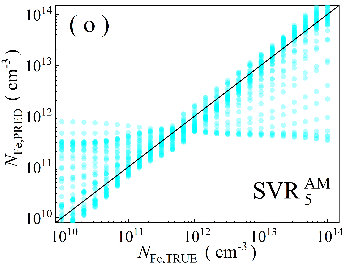
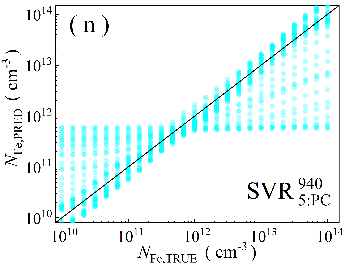
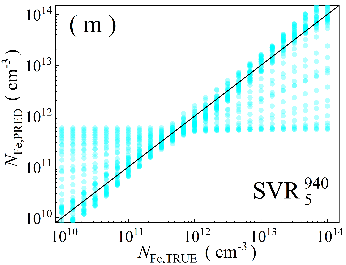


Fig.S12. Scatter plots of the iron concentrations between the reference values and ML predicted values for *T*-altered test phase in the case of 4D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









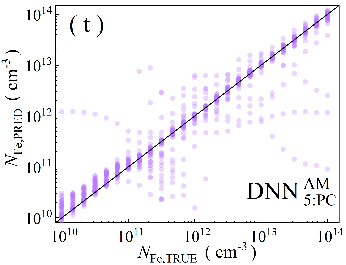
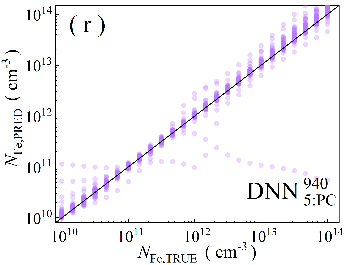
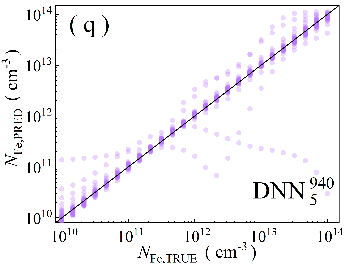
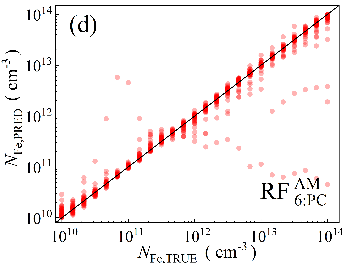
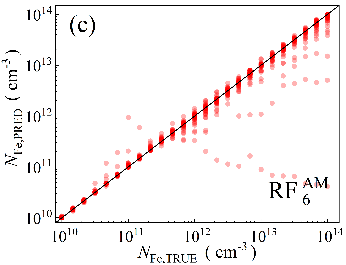
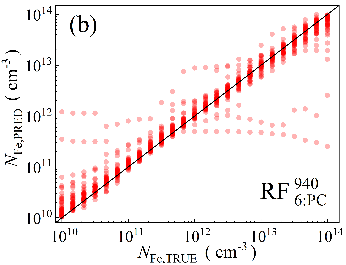
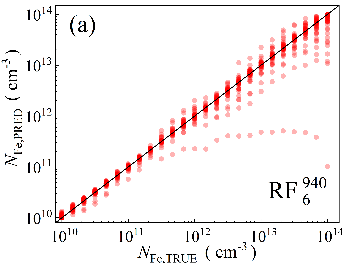
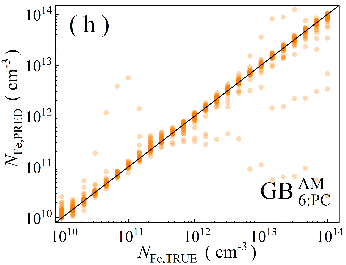
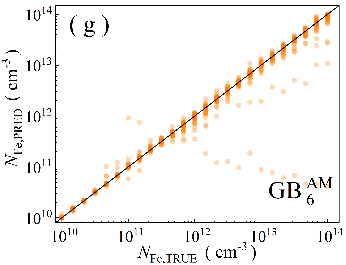
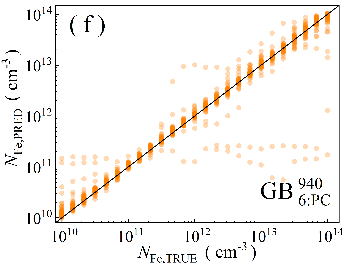
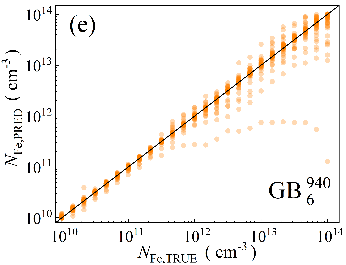
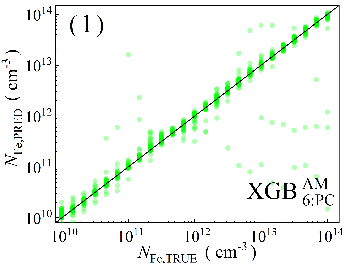
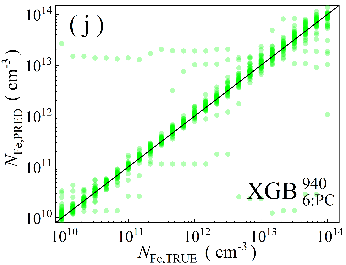
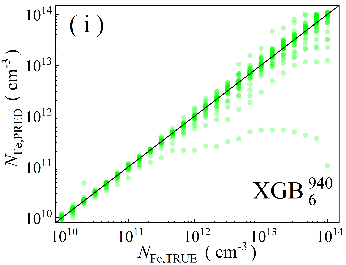
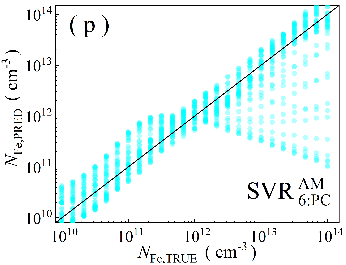
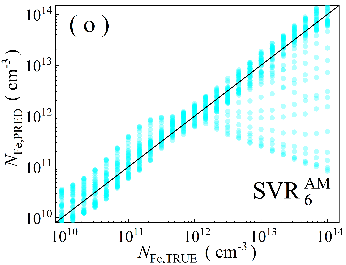
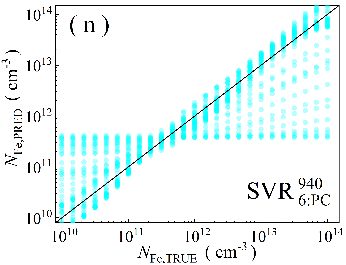
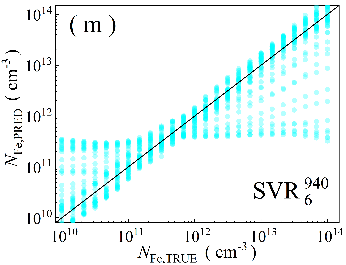


Fig.S13. Scatter plots of the iron concentrations between the reference values and ML predicted values for *T*-altered test phase in the case of 5D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









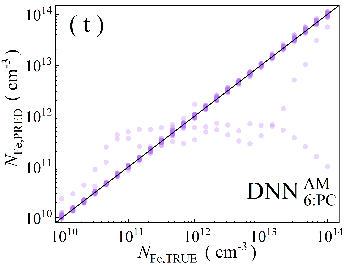
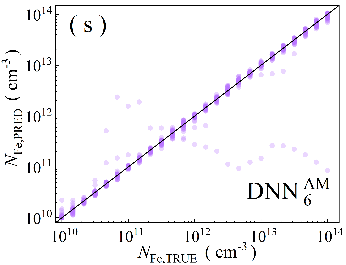
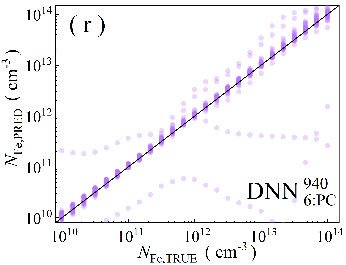
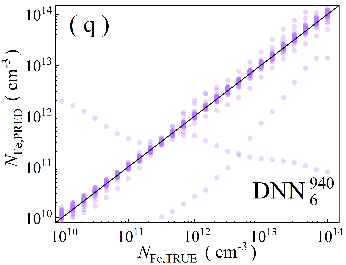
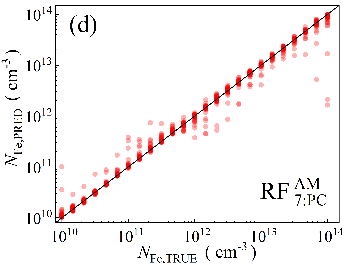
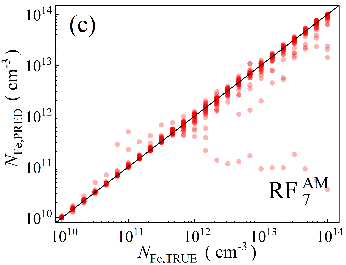
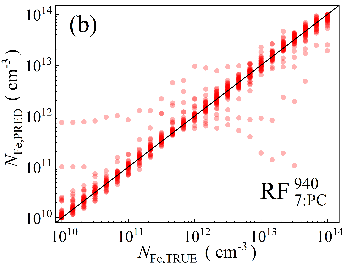
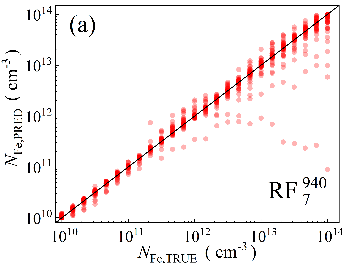
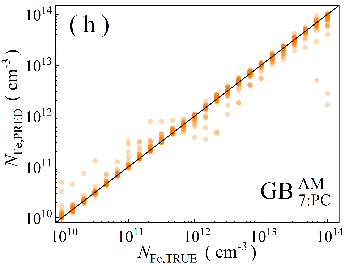
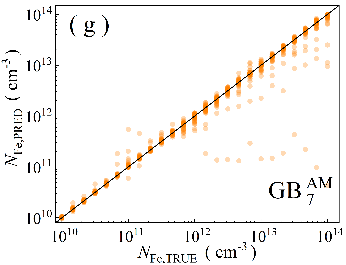
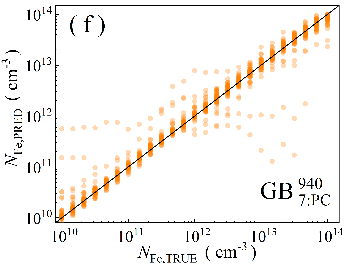
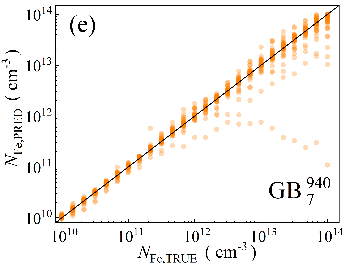
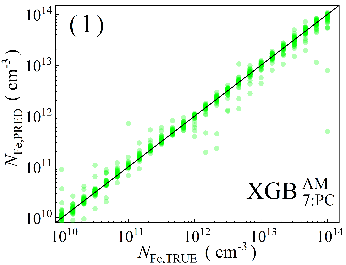
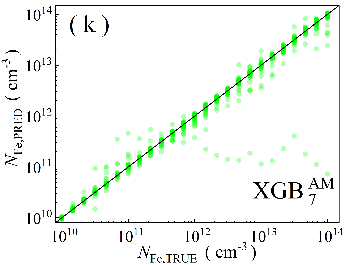
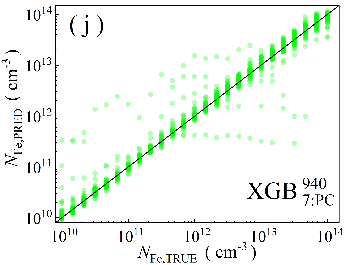
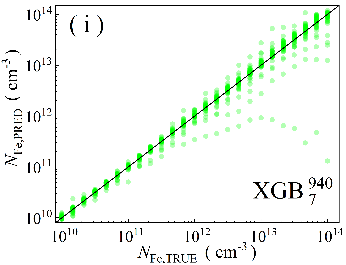
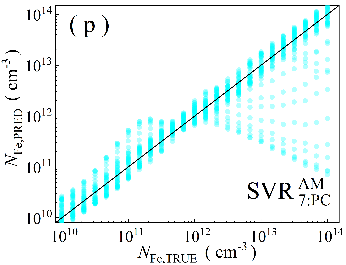
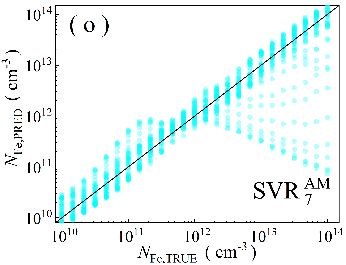
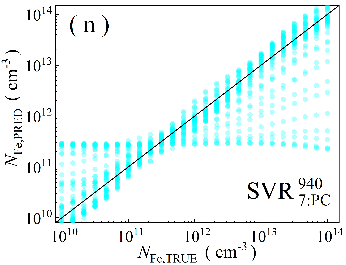
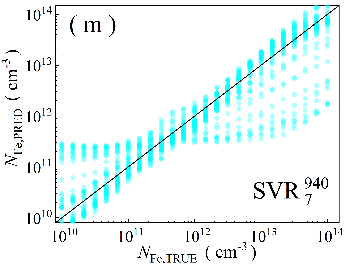


Fig.S14. Scatter plots of the iron concentrations between the reference values and ML predicted values for *T*-altered test phase in the case of 6D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.









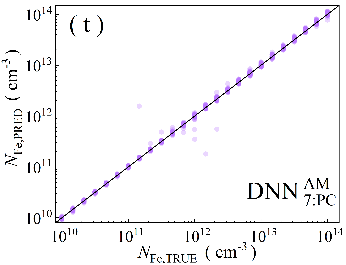
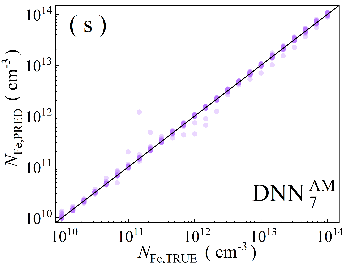
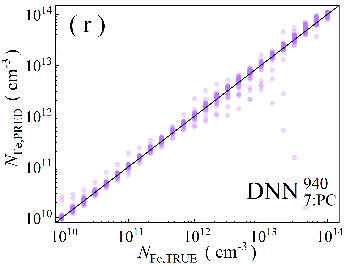
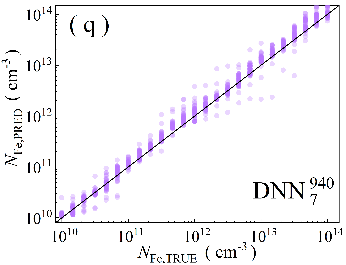


Fig.S15. Scatter plots of the iron concentrations between the reference values and ML predicted values for *T*-altered test phase in the case of 7D features. ML algorithms: RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), DNN (q-t). The data are obtained for monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 (c, d, g, h, k, l, o, p, s, t illuminations. PCA was used for the panels b, d, f, h, j, l, n, p, r, and t. The black lines are the identify lines servings as the references.

Table S16. Performance metrics of the models for *N*B-altered dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Feature dimension | MdAPE, % | | | | *p*01, % | | | | *p*10, % | | | |
| 940 nm | | AM1.5 | | 940 nm | | AM1.5 | | 940 nm | | AM1.5 | |
| Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA |
| RF | 4 | 46.4 | 19.27 | 44.0 | 16.2 | 0.91 | 5.09 | 1.82 | 4.64 | 11.9 | 28.7 | 16.6 | 34.2 |
| 5 | 46.6 | 17.7 | 33.5 | 19.5 | 2.09 | 3.55 | 2.00 | 2.46 | 12.2 | 31.6 | 14.7 | 29.7 |
| 6 | 35.7 | 18.4 | 14.4 | 14.4 | 2.18 | 4.27 | 8.46 | 4.64 | 12.7 | 35.1 | 41.1 | 37.9 |
| 7 | 34.4 | 21.4 | 14.5 | 11.6 | 2.55 | 5.46 | 9.36 | 6.09 | 13.9 | 29.6 | 38.5 | 45.2 |
| GB | 4 | 46.1 | 21.3 | 39.7 | 15.7 | 0.91 | 3.55 | 2.82 | 5.55 | 13.6 | 29.3 | 17.4 | 36.7 |
| 5 | 44.9 | 16.3 | 33.2 | 18.1 | 0.36 | 4.27 | 3.09 | 6.36 | 7.36 | 33.9 | 16.1 | 33.7 |
| 6 | 34.2 | 13.8 | 11.5 | 12.0 | 1.55 | 7.36 | 14.3 | 4.55 | 18.0 | 42.7 | 47.2 | 43.8 |
| 7 | 34.2 | 19.8 | 12.4 | 10.1 | 0.64 | 5.46 | 9.18 | 6.91 | 16.3 | 29.9 | 44.4 | 49.7 |
| XGB | 4 | 53.1 | 17.5 | 52.7 | 15.7 | 2.55 | 6.36 | 1.27 | 5.46 | 16.1 | 34.5 | 13.0 | 37.0 |
| 5 | 53.5 | 15.4 | 40.9 | 14.4 | 3.46 | 5.91 | 2.36 | 5.36 | 15.2 | 38.3 | 14.9 | 40.5 |
| 6 | 46.4 | 15.0 | 11.9 | 12.4 | 1.82 | 5.09 | 12.6 | 6.09 | 19.3 | 38.8 | 46.9 | 43.5 |
| 7 | 42.0 | 18.4 | 12.2 | 11.4 | 1.36 | 5.36 | 6.91 | 6.91 | 15.8 | 34.0 | 46.6 | 45.7 |
| SVR | 4 | 32.3 | 31.5 | 36.9 | 35.9 | 1.73 | 1.64 | 1.82 | 1.73 | 15.6 | 17.4 | 16.3 | 16.5 |
| 5 | 32.3 | 35.9 | 35.7 | 44.8 | 1.46 | 1.73 | 1.00 | 1.73 | 14.8 | 15.5 | 13.1 | 11.9 |
| 6 | 29.6 | 35.3 | 31.6 | 40.7 | 2.64 | 1.09 | 1.73 | 1.46 | 22.5 | 15.9 | 19.4 | 12.4 |
| 7 | 34.0 | 30.5 | 34.4 | 32.3 | 1.91 | 1.36 | 1.64 | 2.46 | 17.0 | 18.2 | 17.0 | 17.7 |
| DNN | 4 | 7.14 | 8.05 | 10.6 | 14.9 | 8.09 | 5.64 | 3.82 | 4.55 | 63.3 | 59.6 | 47.8 | 34.2 |
| 5 | 6.05 | 8.06 | 5.76 | 8.36 | **10.7** | 8.00 | 8.64 | 5.91 | **71.4** | 57.4 | 68.0 | 54.9 |
| 6 | **5.83** | 7.18 | 5.65 | **2.93** | 10.6 | 8.18 | 9.18 | **17.1** | 68.3 | 62.5 | 71.1 | **86.0** |
| 7 | 17.6 | **6.15** | **2.49** | 3.27 | 1.64 | **10.9** | **23.5** | 16.3 | 27.5 | **67.0** | **82.8** | 81.0 |

Table S17. Performance metrics of the models for All-altered dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Feature dimension | MdAPE, % | | | | *p*01, % | | | | *p*10, % | | | |
| 940 nm | | AM1.5 | | 940 nm | | AM1.5 | | 940 nm | | AM1.5 | |
| Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA |
| RF | 4 | 55.8 | 17.2 | 42.8 | 16.0 | 0.92 | 4.37 | 1.93 | 4.03 | 10.7 | 35.5 | 13.2 | 35.8 |
| 5 | 55.2 | 17.4 | 33.4 | 18.2 | 1.26 | 3.53 | 1.35 | 3.70 | 10.3 | 32.4 | 15.8 | 34.6 |
| 6 | 41.6 | 19.3 | 15.6 | 14.5 | 2.02 | 3.95 | 2.52 | 4.37 | 17.8 | 32.7 | 33.5 | 39.3 |
| 7 | 38.4 | 20.4 | 14.3 | 11.4 | 1.01 | 3.03 | 3.61 | 5.21 | 15.5 | 29.2 | 36.2 | 44.5 |
| GB | 4 | 58.4 | 17.5 | 53.5 | 15.1 | 1.01 | 4.45 | 1.18 | 4.45 | 10.8 | 33.3 | 12.4 | 36.1 |
| 5 | 56.1 | 16.5 | 39.4 | 15.6 | 0.76 | 4.54 | 1.60 | 4.29 | 11.7 | 37.0 | 15.3 | 38.1 |
| 6 | 43.1 | 13.4 | 14.5 | 12.7 | 1.85 | 5.13 | 4.03 | 6.05 | 16.7 | 41.5 | 36.0 | 43.3 |
| 7 | 39.9 | 19.9 | 13.3 | 10.4 | 0.92 | 3.70 | 2.69 | 6.47 | 15.0 | 30.8 | 40.6 | 49.1 |
| XGB | 4 | 57.6 | 18.0 | 65.1 | 14.8 | 0.42 | 5.46 | 0.84 | 4.20 | 7.48 | 34.7 | 7.90 | 38.8 |
| 5 | 59.3 | 14.8 | 49.7 | 12.1 | 1.09 | 4.37 | 1.26 | 5.71 | 8.91 | 39.6 | 9.92 | 44.5 |
| 6 | 45.7 | 15.0 | 16.8 | 12.3 | 0.42 | 6.05 | 3.11 | 3.87 | 11.2 | 38.7 | 33.8 | 43.8 |
| 7 | 44.5 | 22.7 | 20.3 | 11.7 | 1.35 | 2.69 | 1.77 | 5.29 | 8.24 | 30.3 | 27.7 | 45.5 |
| SVR | 4 | 30.6 | 28.4 | 34.5 | 33.8 | 1.77 | 1.43 | 1.68 | 2.02 | 17.7 | 19.6 | 15.1 | 16.1 |
| 5 | 32.8 | 37.9 | 35.1 | 41.7 | 1.68 | 1.77 | 1.68 | 1.01 | 15.9 | 14.0 | 15.4 | 13.1 |
| 6 | 29.8 | 37.3 | 31.0 | 41.0 | 2.27 | 1.26 | 1.85 | 1.68 | 20.9 | 13.3 | 18.2 | 12.4 |
| 7 | 34.0 | 30.7 | 32.5 | 29.0 | 2.02 | 2.10 | 1.68 | 2.35 | 17.1 | 17.0 | 15.2 | 18.7 |
| DNN | 4 | **6.29** | 7.34 | 11.6 | 13.4 | 8.07 | 6.98 | 3.36 | 3.53 | 67.1 | 63.0 | 43.4 | 39.9 |
| 5 | 6.95 | 10.2 | 5.51 | 7.74 | 7.65 | 4.62 | 10.3 | 7.56 | 64.5 | 49.4 | 71.7 | 58.0 |
| 6 | 6.52 | 7.60 | 5.13 | **3.06** | **8.24** | **9.33** | 11.0 | **17.7** | **69.4** | 60.5 | 73.4 | **88.0** |
| 7 | 19.1 | **7.16** | **2.05** | 3.46 | 2.69 | 8.99 | **27.8** | 17.1 | 28.6 | **63.3** | **85.4** | 80.6 |

————————————————————————————————--

Table S14. Performance metrics of the models for *N*Fe-altered dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Feature dimension | MdAPE, % | | | | *p*01, % | | | | *p*10, % | | | |
| 940 nm | | AM1.5 | | 940 nm | | AM1.5 | | 940 nm | | AM1.5 | |
| Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA |
| RF | 4 | 6.47 | 6.38 | 8.29 | 8.00 | 10.1 | 9.48 | 7.54 | 8.12 | 68.0 | 70.2 | 57.0 | 57.9 |
| 5 | 7.67 | **6.00** | 6.96 | 9.46 | 7.06 | 8.51 | 10.1 | 6.96 | 60.5 | 72.2 | 64.2 | 52.1 |
| 6 | 8.01 | 7.34 | 5.46 | 5.66 | 6.58 | 7.74 | 11.0 | 10.9 | 59.5 | 60.4 | 72.4 | 72.9 |
| 7 | 8.89 | 8.74 | 6.27 | 4.62 | 5.71 | 5.42 | 10.4 | 13.3 | 54.2 | 56.4 | 68.8 | 82.2 |
| GB | 4 | 4.63 | 4.35 | 6.95 | 6.27 | **13.6** | **14.1** | 9.87 | 9.38 | 76.5 | 80.3 | 62.4 | 67.2 |
| 5 | 4.96 | 4.63 | 6.55 | 8.05 | **13.6** | 12.5 | 11.2 | 7.25 | 72.7 | **82.4** | 66.6 | 57.4 |
| 6 | 6.78 | 6.73 | 5.31 | 4.41 | 8.03 | 9.09 | 11.9 | 15.4 | 65.5 | 64.0 | 74.4 | 79.1 |
| 7 | 7.87 | 7.41 | 4.97 | 3.73 | 7.16 | 7.64 | 12.2 | 17.4 | 60.7 | 61.8 | 75.3 | 84.3 |
| XGB | 4 | 4.82 | 6.91 | 6.43 | 7.75 | 11.0 | 9.77 | 10.1 | 8.32 | 75.6 | 63.5 | 62.7 | 60.2 |
| 5 | **4.46** | 6.20 | 4.26 | 9.04 | 11.7 | 10.2 | 13.8 | 6.58 | **78.6** | 68.6 | 74.7 | 52.3 |
| 6 | 6.49 | 7.57 | 4.13 | 4.79 | 9.87 | 8.51 | 14.8 | 11.8 | 68.7 | 59.9 | 83.2 | 77.2 |
| 7 | 6.16 | 7.96 | 3.71 | 3.94 | 8.80 | 7.54 | 15.2 | 15.6 | 66.0 | 60.9 | 83.3 | 83.0 |
| SVR | 4 | 31.1 | 26.7 | 41.2 | 40.5 | 1.45 | 1.35 | 1.06 | 0.77 | 14.9 | 18.2 | 11.4 | 11.8 |
| 5 | 36.3 | 42.7 | 39.7 | 51.9 | 1.64 | 0.87 | 1.74 | 1.35 | 14.4 | 10.8 | 10.9 | 9.77 |
| 6 | 37.0 | 44.8 | 37.3 | 44.2 | 1.16 | 0.48 | 1.64 | 0.77 | 13.4 | 8.99 | 14.4 | 10.1 |
| 7 | 39.1 | 38.1 | 36.5 | 32.5 | 1.16 | 1.06 | 0.68 | 1.26 | 9.48 | 13.7 | 13.4 | 15.2 |
| DNN | 4 | 8.06 | 7.48 | 11.4 | 13.0 | 7.74 | 7.64 | 3.97 | 4.35 | 58.8 | 60.5 | 44.8 | 41.6 |
| 5 | 7.42 | 9.02 | 5.97 | 6.84 | 6.48 | 6.67 | 10.1 | 6.77 | 66.0 | 54.7 | 72.3 | 61.4 |
| 6 | 6.98 | 8.15 | 6.30 | 3.18 | 7.06 | 5.80 | 8.80 | 17.8 | 64.8 | 59.3 | 69.9 | 90.2 |
| 7 | 17.3 | 6.93 | **1.63** | **2.40** | 4.06 | 8.90 | **32.6** | **23.5** | 33.1 | 64.7 | **98.7** | **96.7** |

Table S15. Performance metrics of the models for *T*-altered dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm | Feature dimension | MdAPE, % | | | | *p*01, % | | | | *p*10, % | | | |
| 940 nm | | AM1.5 | | 940 nm | | AM1.5 | | 940 nm | | AM1.5 | |
| Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA | Init | PCA |
| RF | 4 | 6.50 | 13.1 | **1.88** | 11.6 | 18.2 | 8.92 | **28.1** | 11.8 | 60.4 | 40.8 | **95.6** | 45.2 |
| 5 | 5.59 | 13.2 | 9.58 | 5.90 | 22.4 | 6.75 | 12.1 | 16.6 | 62.9 | 40.7 | 51.3 | 62.7 |
| 6 | 4.61 | 13.3 | 14.2 | 2.13 | 20.8 | 8.58 | 9.50 | 35.3 | 68.8 | 43.5 | 42.1 | 77.3 |
| 7 | 5.07 | 9.82 | 5.05 | **1.94** | 21.5 | 12 | 17.7 | 38.7 | 67.4 | 50.4 | 68.3 | 79.8 |
| GB | 4 | 5.29 | 12.5 | 3.47 | 10.1 | 20.7 | 6.67 | 21.5 | 5.08 | 62.0 | 42.9 | 80.6 | 49.8 |
| 5 | 4.91 | 12.6 | 9.34 | 4.89 | 19.1 | 7.00 | 6.58 | 22.5 | 66.2 | 44.7 | 51.8 | 65.0 |
| 6 | 3.41 | 9.12 | 12.8 | 1.81 | 22.3 | 11.2 | 4.67 | **40.9** | 73.1 | 52.4 | 43.2 | 78.3 |
| 7 | 5.03 | 9.72 | 5.14 | 2.44 | 15.6 | 10.2 | 13.9 | 31.3 | 68.3 | 51.0 | 70.7 | 79.3 |
| XGB | 4 | 5.03 | 10.9 | 3.40 | 8.72 | 22.8 | 8.92 | 20.8 | 8.42 | 64.6 | 46.9 | 80.8 | 53.8 |
| 5 | 3.64 | 9.01 | 9.88 | 5.14 | 28.4 | 8.75 | 7.00 | 23.7 | 67.3 | 53.0 | 50.3 | 66.3 |
| 6 | **2.47** | 10.4 | 10.5 | 2.08 | **33.6** | 6.75 | 6.58 | 34.4 | **73.7** | 49.2 | 48.6 | 78.0 |
| 7 | 9.93 | **5.37** | 5.78 | 3.38 | 9.92 | **14.9** | 12.3 | 19.3 | 50.3 | **67.9** | 69.2 | 78.3 |
| SVR | 4 | 30.9 | 27.8 | 5.11 | 36.4 | 1.50 | 1.58 | 13.6 | 0.67 | 15.3 | 18.3 | 72.9 | 12.8 |
| 5 | 33.7 | 38.0 | 35.7 | 37.5 | 1.75 | 1.00 | 1.17 | 0.92 | 15.3 | 12.9 | 13.5 | 12.8 |
| 6 | 38.3 | 40.9 | 47.6 | 37.1 | 1.50 | 1.50 | 0.67 | 1.25 | 14.3 | 15.0 | 9.83 | 12.8 |
| 7 | 18.6 | 6.85 | 46.2 | 36.3 | 3.83 | 7.50 | 0.75 | 1.83 | 29.8 | 66.3 | 9.58 | 15.4 |
| DNN | 4 | 6.92 | 7.54 | 10.5 | 14.2 | 7.83 | 8.33 | 4.58 | 4.67 | 65.3 | 59.0 | 47.3 | 37.2 |
| 5 | 6.53 | 8.90 | 5.46 | 7.44 | 9.17 | 5.50 | 10.9 | 7.42 | 68.8 | 53.8 | 71.6 | 58.4 |
| 6 | 7.09 | 7.18 | 4.92 | 2.99 | 7.50 | 9.25 | 10.7 | 18.3 | 66.8 | 61.1 | 79.2 | 90.9 |
| 7 | 39.3 | 35.9 | 32.2 | 2.53 | 1.25 | 1.50 | 1.00 | 23.2 | 10.7 | 13.3 | 14.7 | **92.5** |

Performance comparison of considered models in four different seasons (