# Electronic supplementary material

This electronic supplementary material is attached to the paper ‘*Inversion of resistivity properties through whole-rock geochemistry and IR spectroscopy using Machine-Learning’* and provides additional information regarding the specific implementations of the different ML models used in this paper.

The study aimed at predicting resistivity values (as log values) using whole-rock ICP-MS geochemistry, IR spectroscopy (350-2500nm) and a combination of both dataset (data fusion). Datasets used in this study are exploration datasets collected for uranium exploration in the Athabasca Basin (Saskatchewan, Canada) and its underlying basement. As the Athabasca Basin and its basement have very different physico-chemistry, textures and petrophysical properties, they were divided between two subsets, i.e. a basin and a basement dataset on which all ML steps were run separately.

This supplement provides for each scenario:

* Total training time of the 100 iterations
* Results of R² and RMSE on the validation and training datasets.
* The combination of hyperparameters found by optuna.
* Scatterplots of predicted values against actual values, along with .95 prediction confidence intervals determined using bootstrap resampling (with 100 iterations).

Testing and training datasets were divided using a blind-well split strategy. Blind well splitting is a splitting method associated to ML using drillhole data in which samples are split based on drillhole names instead of random sampling of individuals.

The electronic supplementary material is organized as follows:

[**Search spaces** 2](#_Toc171523483)

[**Results** 2](#_Toc171523484)

[**Basin** 2](#_Toc171523485)

[Geochemistry – Centered log-ratio (CLR) 2](#_Toc171523486)

[Geochemistry – Pairwise Log-ratio (PWLR) 4](#_Toc171523487)

[Geochemistry – Isometric log-ratio (ILR) 6](#_Toc171523488)

[IR spectroscopy – Simple Smoothing (SS) 8](#_Toc171523489)

[IR spectroscopy – First-order derivative (FOD) 10](#_Toc171523490)

[IR spectroscopy – Second-order derivative (SOD) 12](#_Toc171523491)

[IR spectroscopy – Continuum removal (CR) 14](#_Toc171523492)

[Data fusion – PWLR + FOD 16](#_Toc171523493)

[**Basement** 18](#_Toc171523494)

[Geochemistry – Centered log-ratio (CLR) 18](#_Toc171523495)

[Geochemistry – Isometric log-ratio (ILR) 19](#_Toc171523496)

[Geochemistry – Pairwise log-ratio (PWLR) 20](#_Toc171523497)

[IR spectroscopy – Simple Smoothing (SS) 22](#_Toc171523498)

[IR spectroscopy – First-order derivative (FOD) 23](#_Toc171523499)

[IR spectroscopy – Second-order derivative (SOD) 24](#_Toc171523500)

[IR spectroscopy – Continuum removal (CR) 26](#_Toc171523501)

[Data fusion – PWLR + FOD 28](#_Toc171523502)

# Search spaces

Table 1 provides details regarding the hyperparameters search spaces for each model. Given the fact that the studied dataset was rather small (1301 samples for the basement and 2350 for the basin), the choice was made to reduce model complexity (i.e., low values of maximum depth of trees, number of estimators…) in order to avoid overfitting and limit computing time.

|  |  |
| --- | --- |
| **Model** | **Search space** |
| **RandomForestRegressor** | N\_estimators [2 to 75]  Max\_depth [3 to 10]  Max\_leaf\_nodes [10 to 50]  Min\_samples\_split [0.1 to 1.0]  Min\_samples\_leaf [0.1 to 1.0] |
| **XGBoostRegressor** | Max\_depth : [3 to 9]  Gamma [0 to 2]  Learning\_rate [0.1 to 0.03]  Subsample [0.6 to 0.9]  Min\_child\_weight [1 to 6]  Lambda [1e-3 to 1000]  Alpha [1e-1 to 1000]  Colsample\_bytree [0.6 to 0.9] |
| **LGBRegressor** | N\_estimators [2 to 75]  Reg\_alpha [1e-3 to 1000]  Reg\_lambda [1e-3 to 1000]  Colsample\_bytree [0.4 to 0.6]  Subsample [0.4 to 1.0]  Learning\_rate [0.01 to 0.05]  Max\_depth [3 to 8]  Num\_leaves [2 to 10]  Min\_child\_samples [1 to 50] |
| **SupportVectorRegressor** | Kernel [‘linear’ or ‘rbf’]  C [0.1 to 2]  Epsilon [1e-3 to 0.1]  Gamma [1e-4 to 0.1] *(for rbf kernel only)* |
| **PLSRegressor** | N\_components : [1 to 100]  Tol : [1e-8 to 1e-2]  Scale : [True, False] |

**Table 1. Search spaces for hyperparameters**

# Results

# Basin

## Geochem – Centered log-ratio (CLR)

**LGBR**

LGBR took 24.72 seconds

Best trial parameters for LGBR: {'n\_estimators': 73, 'reg\_alpha': 0.32336004246760947, 'reg\_lambda': 0.0025006190048306974, 'colsample\_bytree': 0.6, 'subsample': 0.8, 'learning\_rate': 0.04911724195633387, 'max\_depth': 7, 'num\_leaves': 10, 'min\_child\_samples': 17}

Best trial CV R² for LGBR: 0.7450

**XGBR**

XGBR took 31.55 seconds

Best trial parameters for XGBR: {'max\_depth': 6, 'gamma': 0.6028414105608607, 'eta': 0.02404278237291413, 'subsample': 0.8715396547703093, 'min\_child\_weight': 6, 'lambda': 190.58592551467606, 'alpha': 4.6303717835297284, 'colsample\_bytree': 0.6126573878456539}

Best trial CV R² for XGBR: 0.5949

**RFR**

RFR took 28.86 seconds

Best trial parameters for RFR: {'n\_estimators': 48, 'max\_depth': 4, 'max\_leaf\_nodes': 40, 'min\_samples\_split': 0.13574894012771357, 'min\_samples\_leaf': 0.10056800271491852}

Best trial CV R² for RFR: 0.6243

**SVR**

SVR took 45.22 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.8626696485737283, 'epsilon': 0.015453160675024787, 'gamma': 0.013672234129079568}

Best trial CV R² for SVR: 0.7879

**PLSR**

PLSR took 7.55 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 7.168380449301725e-05, 'scale': True}

Best trial CV R² for PLSR: 0.7275

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

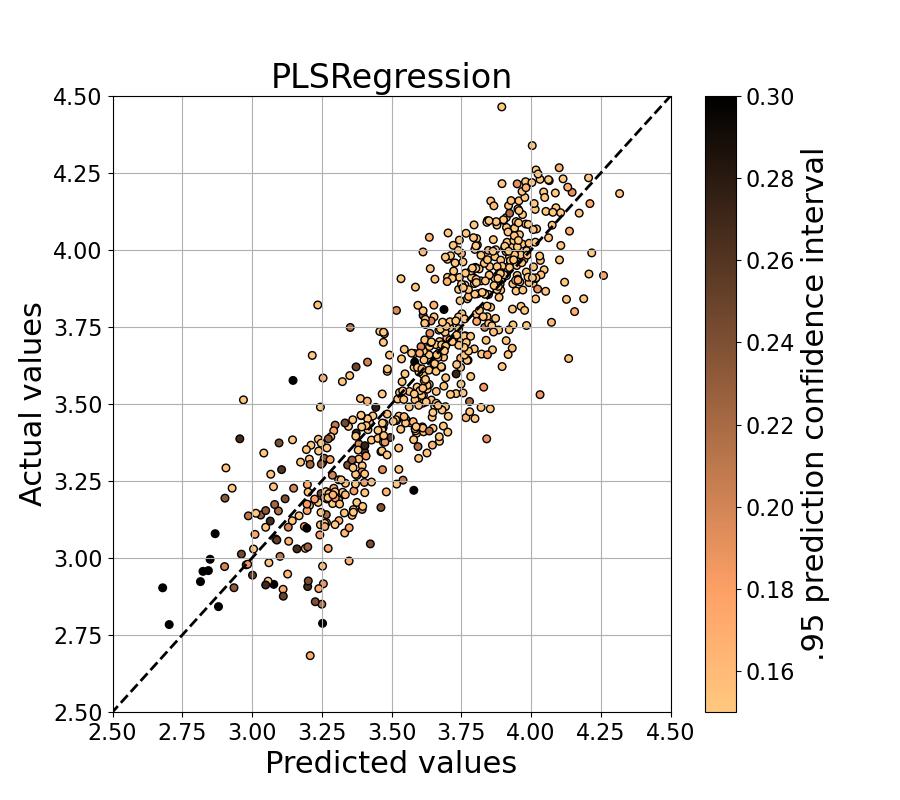
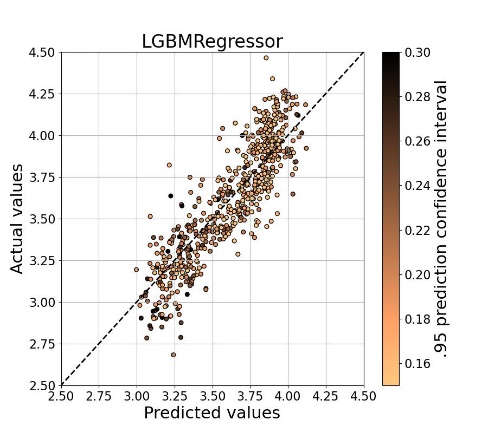
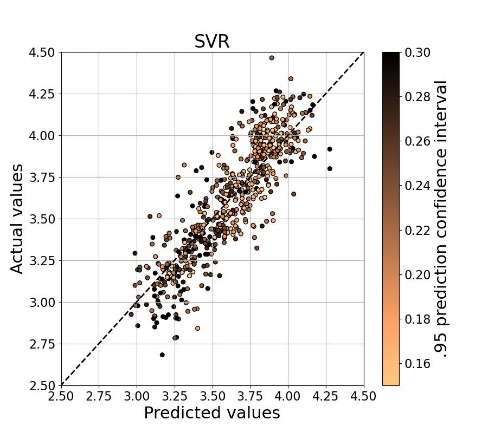
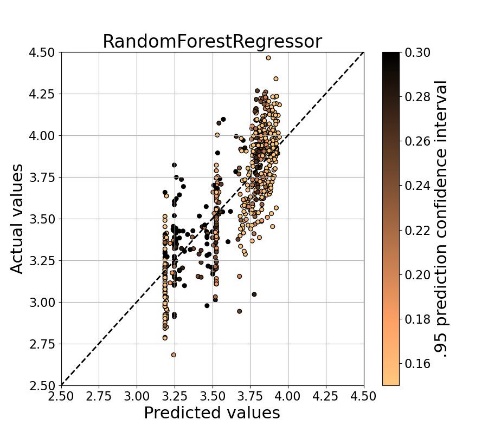
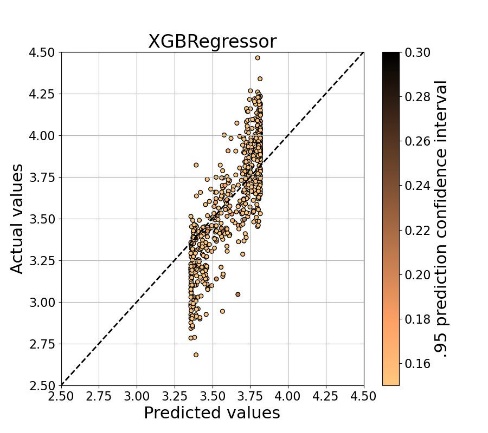
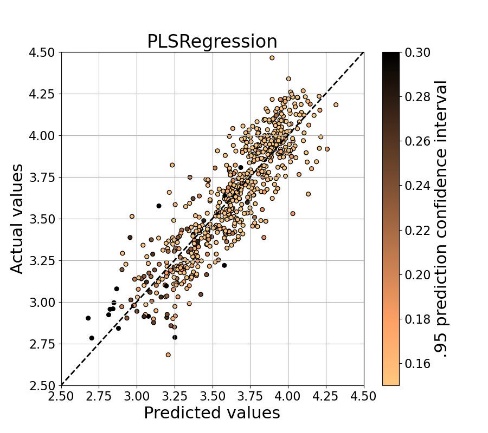
LGBR 0.8246 0.7744 0.1423 0.1719

XGBR 0.6358 0.6474 0.2051 0.2149

RFR 0.6417 0.6529 0.2034 0.2133

SVR 0.8999 0.7773 0.1075 0.1708

PLSR 0.7446 0.7707 0.1718 0.1733

## Geochemistry – Pairwise Log-ratio (PWLR)

**LGBR**

LGBR took 102.09 seconds

Best trial parameters for LGBR: {'n\_estimators': 72, 'reg\_alpha': 0.14002824939901107, 'reg\_lambda': 0.006438311069718575, 'colsample\_bytree': 0.4, 'subsample': 0.4, 'learning\_rate': 0.04757491442812937, 'max\_depth': 9, 'num\_leaves': 9, 'min\_child\_samples': 13}

Best trial CV R² for LGBR: 0.7705

**XGBR**

XGBR took 332.71 seconds

Best trial parameters for XGBR: {'max\_depth': 9, 'gamma': 1.1974919811841576, 'eta': 0.02924590419754341, 'subsample': 0.7428858092977704, 'min\_child\_weight': 3, 'lambda': 236.42523546456687, 'alpha': 1.6366832219981422, 'colsample\_bytree': 0.8791242497344556}

Best trial CV R² for XGBR: 0.6223

**RFR**

RFR took 163.50 seconds

Best trial parameters for RFR: {'n\_estimators': 30, 'max\_depth': 8, 'max\_leaf\_nodes': 17, 'min\_samples\_split': 0.17262940878952981, 'min\_samples\_leaf': 0.10113749899984904}

Best trial CV R² for RFR: 0.6717

**SVR**

SVR took 583.73 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.8420728738541695, 'epsilon': 0.014209003166813531, 'gamma': 0.0007269575574330742}

Best trial CV R² for SVR: 0.7951

**PLSR**

PLSR took 22.56 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.00541164127601286, 'scale': False}

Best trial CV R² for PLSR: 0.7289

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

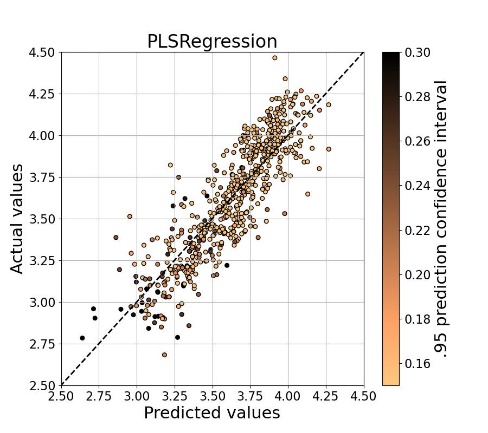
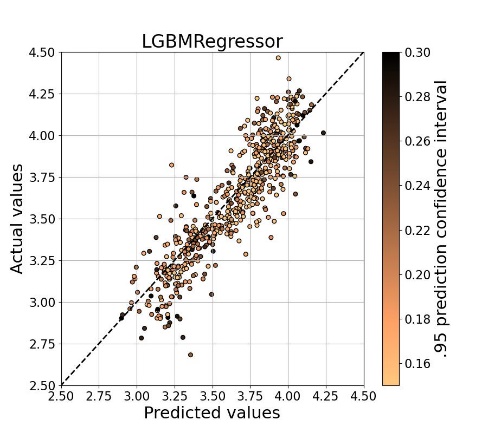
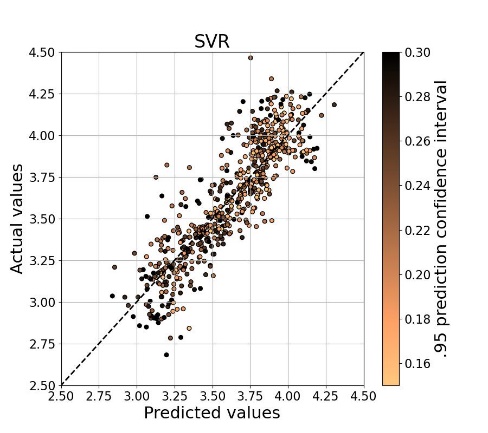
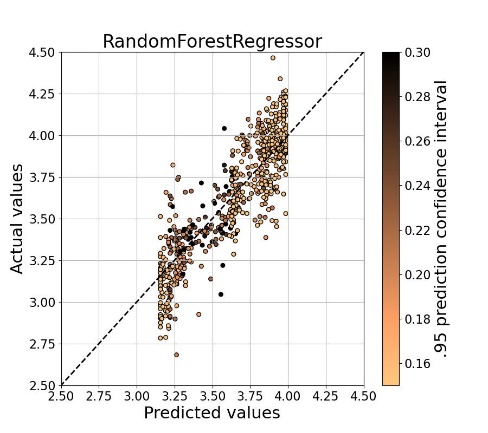
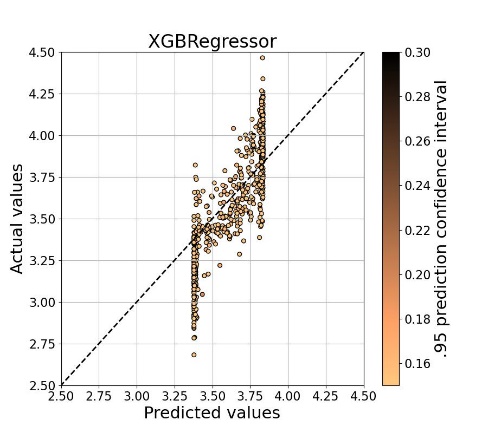
LGBR 0.8542 0.7989 0.1298 0.1623

XGBR 0.6611 0.6977 0.1978 0.1990

RFR 0.6920 0.7448 0.1886 0.1829

SVR 0.9199 0.7697 0.0962 0.1737

PLSR 0.7471 0.7715 0.1709 0.1730



## Geochemistry – Isometric log-ratio (ILR)

**LGBR**

LGBR took 7.68 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 0.39134305467079444, 'reg\_lambda': 0.1315655901135208, 'colsample\_bytree': 0.6, 'subsample': 1.0, 'learning\_rate': 0.04753890399752513, 'max\_depth': 6, 'num\_leaves': 9, 'min\_child\_samples': 10}

Best trial CV R² for LGBR: 0.7383

**XGBR**

XGBR took 14.32 seconds

Best trial parameters for XGBR: {'max\_depth': 9, 'gamma': 0.030061593160582323, 'eta': 0.02362712662215172, 'subsample': 0.8687698345494517, 'min\_child\_weight': 3, 'lambda': 14.197300135711032, 'alpha': 6.496370881416329, 'colsample\_bytree': 0.7819381429564798}

Best trial CV R² for XGBR: 0.6645

**RFR**

RFR took 19.76 seconds

Best trial parameters for RFR: {'n\_estimators': 31, 'max\_depth': 8, 'max\_leaf\_nodes': 22, 'min\_samples\_split': 0.11295686555423667, 'min\_samples\_leaf': 0.10612695146502205}

Best trial CV R² for RFR: 0.6074

**SVR**

SVR took 19.45 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.9401203310002514, 'epsilon': 0.06688793611682943, 'gamma': 0.007648074878319484}

Best trial CV R² for SVR: 0.7833

**PLSR**

PLSR took 4.05 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.008911496469487088, 'scale': False}

Best trial CV R² for PLSR: 0.7284

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

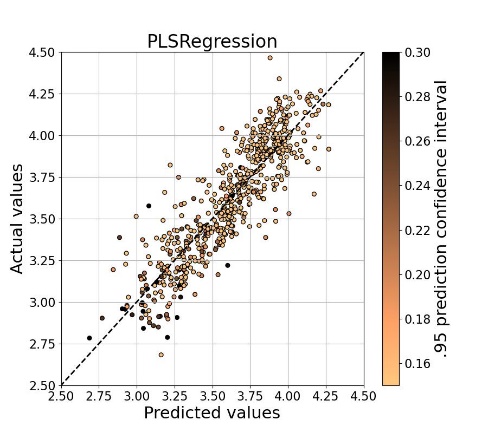
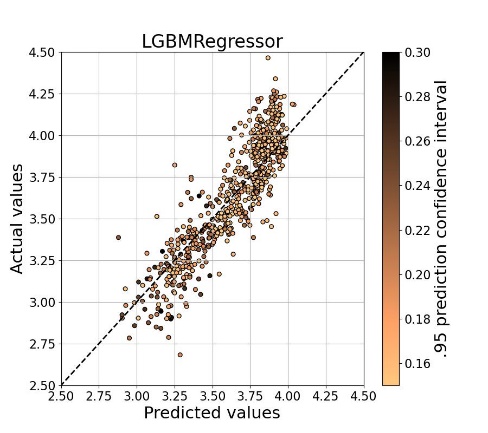
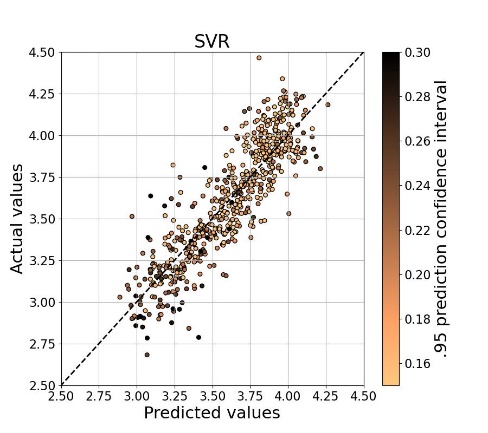
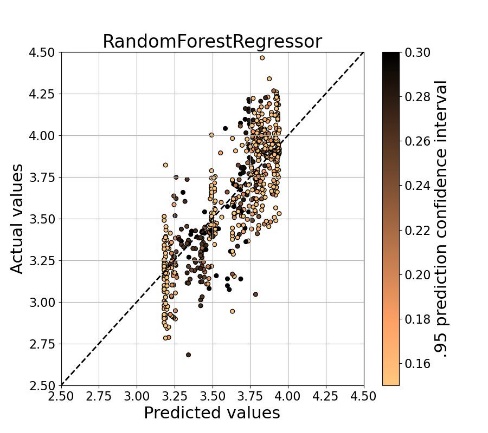
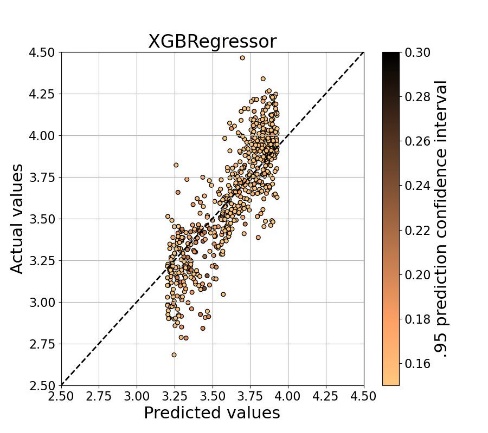
LGBR 0.8128 0.7756 0.1470 0.1715

XGBR 0.7409 0.6983 0.1730 0.1988

RFR 0.6248 0.6228 0.2082 0.2223

SVR 0.8565 0.7890 0.1287 0.1663

PLSR 0.7440 0.7717 0.1720 0.1730



## IR spectroscopy – Simple Smoothing (SS)

**LGBR**

LGBR took 218.31 seconds

Best trial parameters for LGBR: {'n\_estimators': 73, 'reg\_alpha': 0.12831565199121436, 'reg\_lambda': 0.00780842114572355, 'colsample\_bytree': 0.4, 'subsample': 0.7, 'learning\_rate': 0.048830307801204495, 'max\_depth': 6, 'num\_leaves': 10, 'min\_child\_samples': 31}

Best trial CV R² for LGBR: 0.6353

**XGBR**

XGBR took 550.17 seconds

Best trial parameters for XGBR: {'max\_depth': 6, 'gamma': 0.6625392980894005, 'eta': 0.025500940374067336, 'subsample': 0.7146788189997815, 'min\_child\_weight': 9, 'lambda': 181.36207707946645, 'alpha': 0.6921432561947078, 'colsample\_bytree': 0.7170538979003657}

Best trial CV R² for XGBR: 0.4900

**RFR**

RFR took 201.50 seconds

Best trial parameters for RFR: {'n\_estimators': 22, 'max\_depth': 6, 'max\_leaf\_nodes': 35, 'min\_samples\_split': 0.20592311755989195, 'min\_samples\_leaf': 0.10008357611938819}

Best trial CV R² for RFR: 0.4821

**SVR**

SVR took 660.15 seconds

Best trial parameters for SVR: {'kernel': 'linear', 'C': 0.6571031402803712, 'epsilon': 0.06370014455979468}

Best trial CV R² for SVR: 0.7080

**PLSR**

PLSR took 31.74 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.007667867228957925, 'scale': True}

Best trial CV R² for PLSR: 0.5921

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

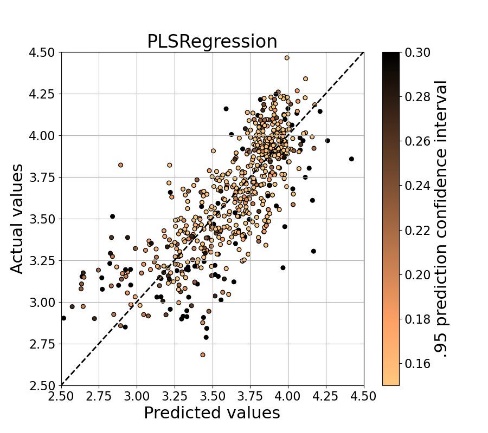
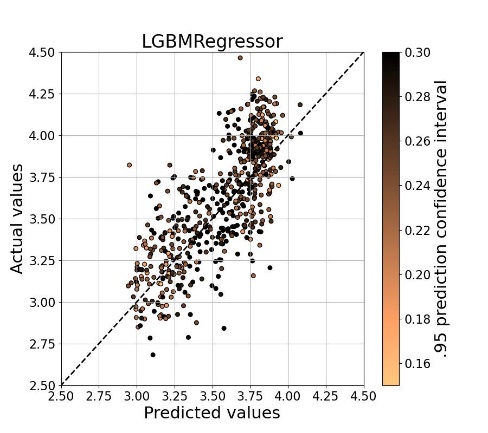
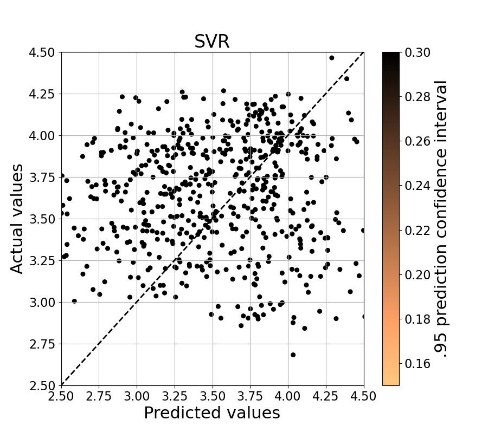
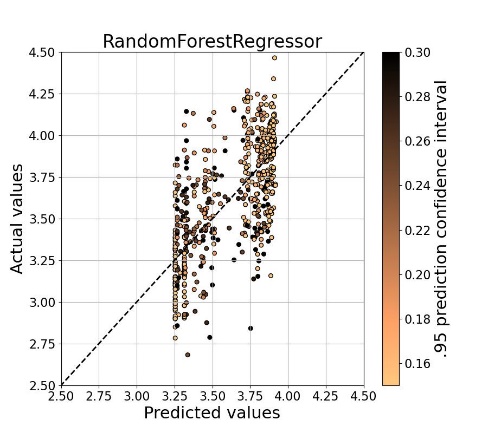
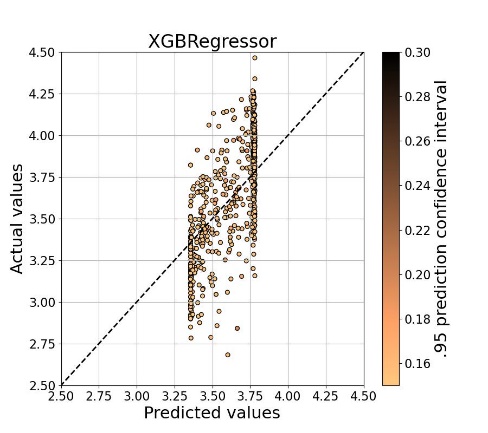
LGBR 0.7632 0.5473 0.1654 0.2435

XGBR 0.5516 0.5160 0.2276 0.2518

RFR 0.5076 0.5123 0.2385 0.2528

SVR 0.7614 -0.5383 0.1660 0.4490

PLSR 0.6076 0.5472 0.2129 0.2436



## IR spectroscopy – First-order derivative (FOD)

**LGBR**

LGBR took 176.64 seconds

Best trial parameters for LGBR: {'n\_estimators': 69, 'reg\_alpha': 0.18443387834697694, 'reg\_lambda': 0.027515640951675625, 'colsample\_bytree': 0.4, 'subsample': 0.5, 'learning\_rate': 0.042628270803302004, 'max\_depth': 6, 'num\_leaves': 10, 'min\_child\_samples': 20}

Best trial CV R² for LGBR: 0.7034

**XGBR**

XGBR took 549.80 seconds

Best trial parameters for XGBR: {'max\_depth': 6, 'gamma': 0.7811835455965672, 'eta': 0.02906100659106366, 'subsample': 0.8655546610346003, 'min\_child\_weight': 7, 'lambda': 81.70711824352989, 'alpha': 1.32763465537745, 'colsample\_bytree': 0.6287152021152509}

Best trial CV R² for XGBR: 0.6448

**RFR**

RFR took 221.18 seconds

Best trial parameters for RFR: {'n\_estimators': 44, 'max\_depth': 4, 'max\_leaf\_nodes': 40, 'min\_samples\_split': 0.1726928594067987, 'min\_samples\_leaf': 0.10001271874239952}

Best trial CV R² for RFR: 0.6170

**SVR**

SVR took 820.64 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.8900627547642105, 'epsilon': 0.0895732977088781, 'gamma': 0.0007680037273826458}

Best trial CV R² for SVR: 0.7379

**PLSR**

PLSR took 32.32 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.0009320004259617912, 'scale': True}

Best trial CV R² for PLSR: 0.6290

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

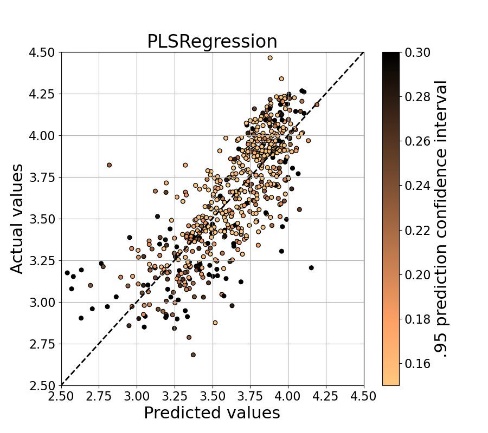
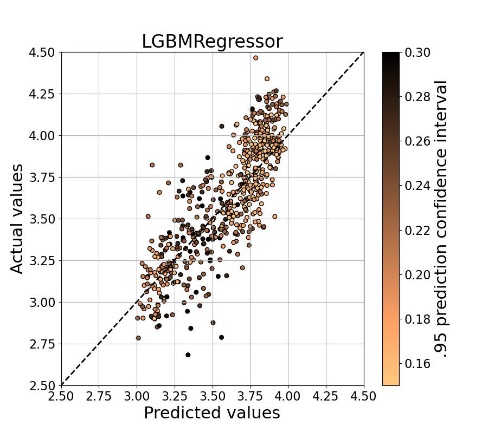
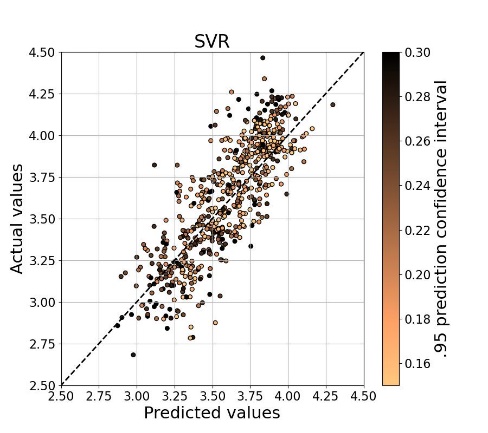
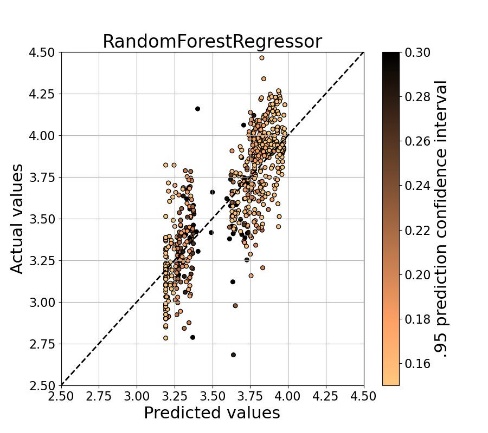
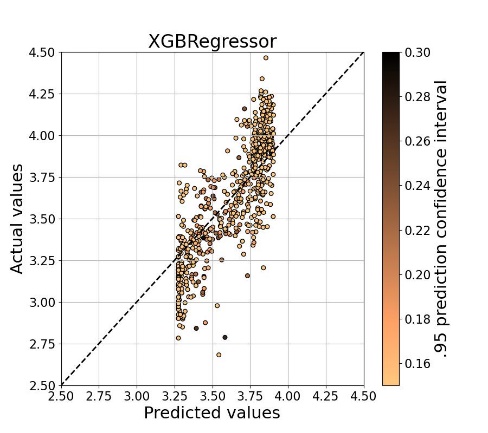
LGBR 0.8031 0.7128 0.1508 0.1940

XGBR 0.6969 0.6554 0.1871 0.2125

RFR 0.6486 0.6479 0.2015 0.2148

SVR 0.8609 0.6956 0.1267 0.1997

PLSR 0.6563 0.6479 0.1992 0.2148



## IR spectroscopy – Second-order derivative (SOD)

**LGBR**

LGBR took 143.62 seconds

Best trial parameters for LGBR: {'n\_estimators': 71, 'reg\_alpha': 0.2183506669226763, 'reg\_lambda': 0.1719947782230465, 'colsample\_bytree': 0.4, 'subsample': 0.5, 'learning\_rate': 0.0486536315332095, 'max\_depth': 4, 'num\_leaves': 10, 'min\_child\_samples': 38}

Best trial CV R² for LGBR: 0.7024

**XGBR**

XGBR took 513.70 seconds

Best trial parameters for XGBR: {'max\_depth': 5, 'gamma': 0.9063573154541369, 'eta': 0.02988748424334134, 'subsample': 0.7413418777474785, 'min\_child\_weight': 5, 'lambda': 250.00213341540695, 'alpha': 1.714955424745635, 'colsample\_bytree': 0.8858933597462454}

Best trial CV R² for XGBR: 0.5541

**RFR**

RFR took 373.41 seconds

Best trial parameters for RFR: {'n\_estimators': 39, 'max\_depth': 6, 'max\_leaf\_nodes': 22, 'min\_samples\_split': 0.19836174635869086, 'min\_samples\_leaf': 0.10230462449130501}

Best trial CV R² for RFR: 0.5626

**SVR**

SVR took 1129.13 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.9882716598304179, 'epsilon': 0.010572109601044478, 'gamma': 0.0005715510713027042}

Best trial CV R² for SVR: 0.7617

**PLSR**

PLSR took 40.12 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.006752701386383963, 'scale': True}

Best trial CV R² for PLSR: 0.6470

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

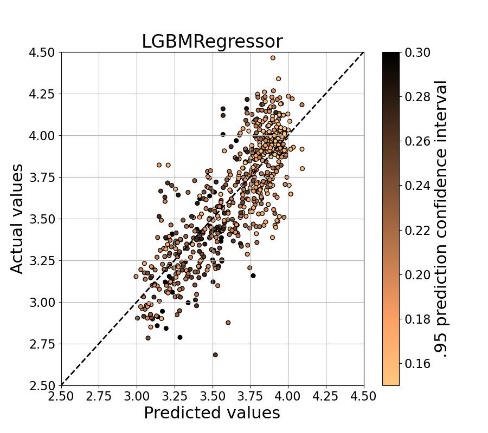
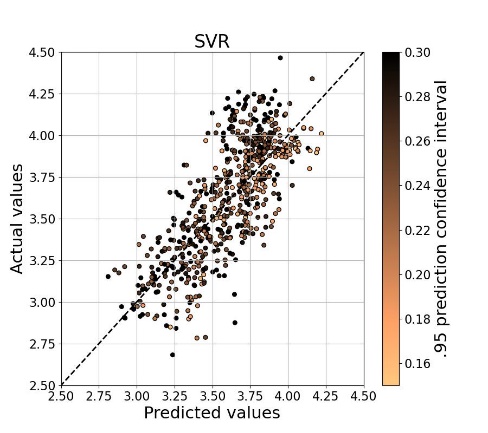
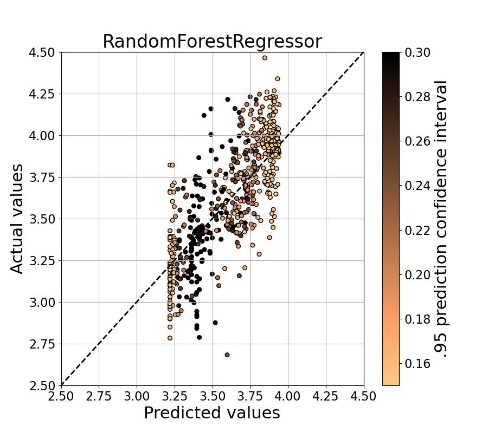
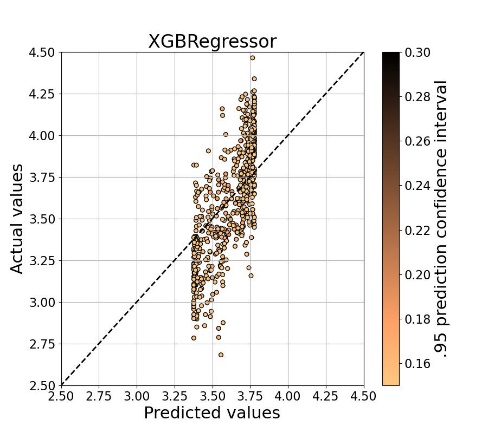
LGBR 0.8135 0.6708 0.1468 0.2077

XGBR 0.6113 0.5888 0.2119 0.2321

RFR 0.5926 0.5812 0.2169 0.2343

SVR 0.9170 0.6582 0.0979 0.2116

PLSR 0.7004 0.5523 0.1860 0.2422



## IR spectroscopy – Continuum removal (CR)

**LGBR**

LGBR took 149.52 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 0.13721974032444745, 'reg\_lambda': 0.014457107401063582, 'colsample\_bytree': 0.4, 'subsample': 1.0, 'learning\_rate': 0.04850082427870976, 'max\_depth': 8, 'num\_leaves': 10, 'min\_child\_samples': 15}

Best trial CV R² for LGBR: 0.6358

**XGBR**

XGBR took 497.57 seconds

Best trial parameters for XGBR: {'max\_depth': 9, 'gamma': 0.12100078322069763, 'eta': 0.01746950534266912, 'subsample': 0.8535637500247882, 'min\_child\_weight': 1, 'lambda': 56.28037422561087, 'alpha': 0.6261558869175665, 'colsample\_bytree': 0.8159657258213877}

Best trial CV R² for XGBR: 0.5255

**RFR**

RFR took 112.51 seconds

Best trial parameters for RFR: {'n\_estimators': 12, 'max\_depth': 3, 'max\_leaf\_nodes': 28, 'min\_samples\_split': 0.20948779971215867, 'min\_samples\_leaf': 0.10142169564170023}

Best trial CV R² for RFR: 0.4800

**SVR**

SVR took 761.22 seconds

Best trial parameters for SVR: {'kernel': 'linear', 'C': 0.9915257735569005, 'epsilon': 0.1284498794698518}

Best trial CV R² for SVR: 0.7225

**PLSR**

PLSR took 35.12 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.005004813289445323, 'scale': True}

Best trial CV R² for PLSR: 0.5938

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

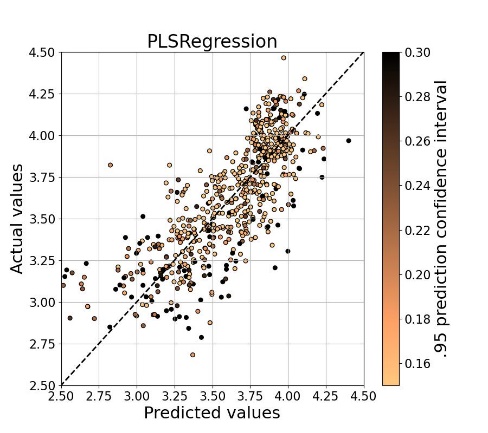
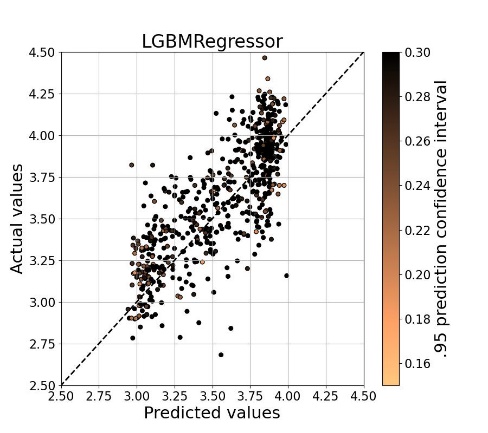
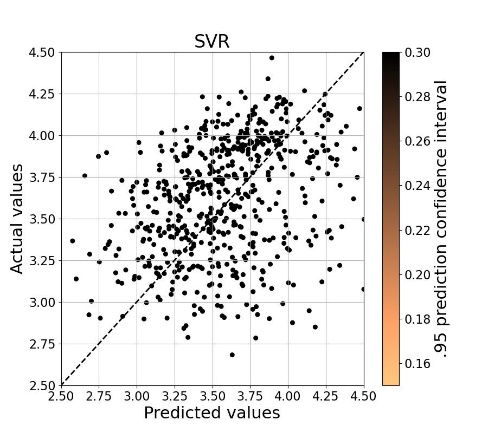
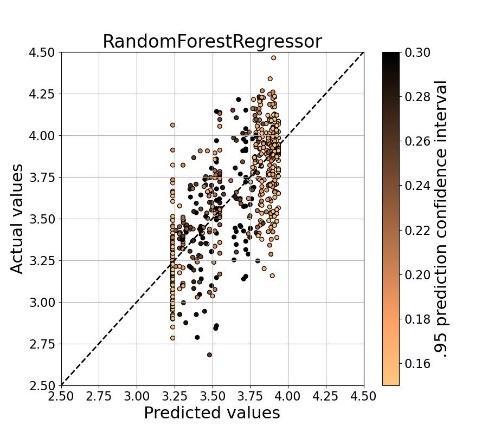
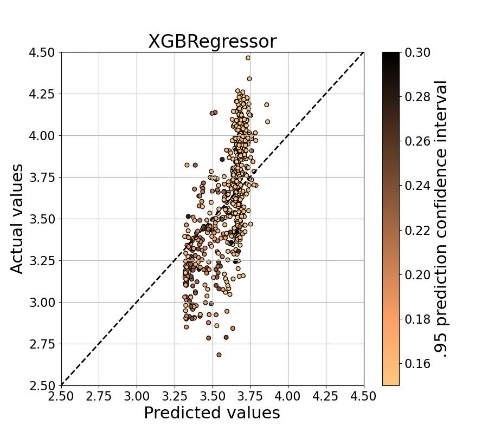
LGBR 0.7758 0.4356 0.1609 0.2719

XGBR 0.6502 0.3913 0.2010 0.2824

RFR 0.5046 0.5147 0.2392 0.2522

SVR 0.8215 -0.3137 0.1436 0.4149

PLSR 0.6095 0.5481 0.2124 0.2434



## Data fusion – PWLR + FOD

**LGBR**

LGBR took 168.62 seconds

Best trial parameters for LGBR: {'n\_estimators': 71, 'reg\_alpha': 0.7536754627480934, 'reg\_lambda': 0.007085425647253488, 'colsample\_bytree': 0.6, 'subsample': 0.5, 'learning\_rate': 0.045793565068882255, 'max\_depth': 3, 'num\_leaves': 10, 'min\_child\_samples': 5}

Best trial CV R² for LGBR: 0.7718

**XGBR**

XGBR took 1049.75 seconds

Best trial parameters for XGBR: {'max\_depth': 6, 'gamma': 0.12815811894291274, 'eta': 0.02875326064513437, 'subsample': 0.8070344419290124, 'min\_child\_weight': 10, 'lambda': 47.281151349305, 'alpha': 2.336618211856589, 'colsample\_bytree': 0.7856244098339019}

Best trial CV R² for XGBR: 0.7346

**RFR**

RFR took 332.03 seconds

Best trial parameters for RFR: {'n\_estimators': 7, 'max\_depth': 3, 'max\_leaf\_nodes': 10, 'min\_samples\_split': 0.14320030407935902, 'min\_samples\_leaf': 0.10254003289490636}

Best trial CV R² for RFR: 0.6797

**SVR**

SVR took 2123.26 seconds

Best trial parameters for SVR: {'kernel': 'linear', 'C': 0.11616729297644866, 'epsilon': 0.08567899565016367}

Best trial CV R² for SVR: 0.7882

**PLSR**

PLSR took 54.60 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.00011248266084848998, 'scale': False}

Best trial CV R² for PLSR: 0.7536

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

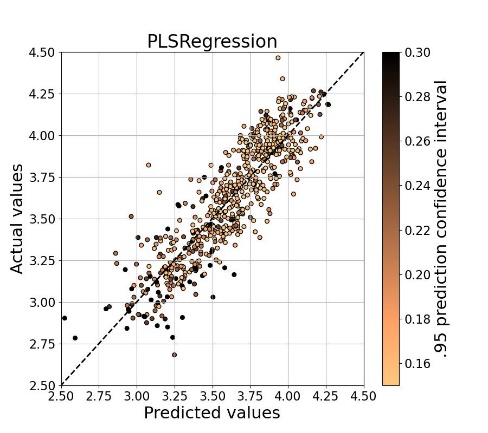
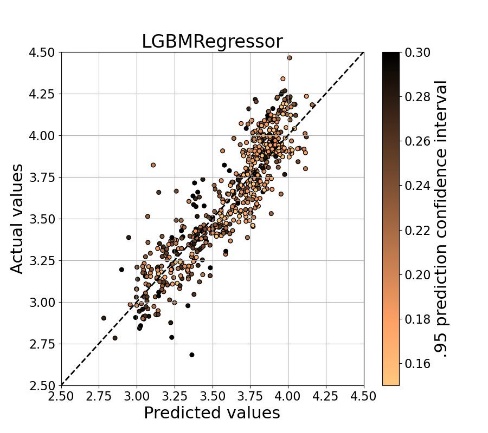
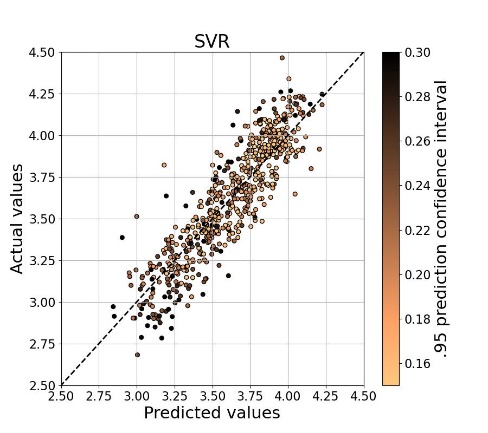
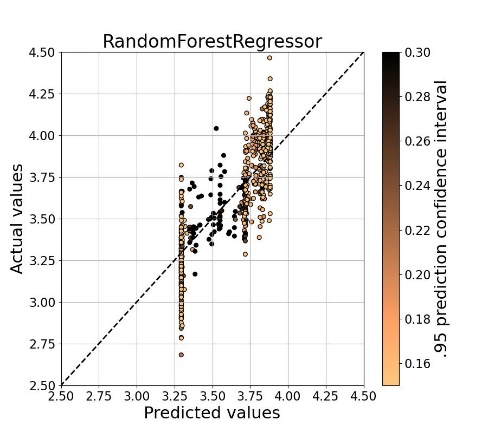
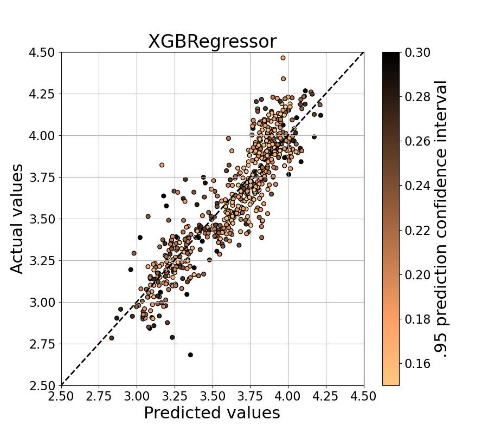
LGBR 0.8419 0.7994 0.1351 0.1621

XGBR 0.8203 0.7486 0.1441 0.1815

RFR 0.6904 0.7209 0.1891 0.1912

SVR 0.8657 0.3074 0.1245 0.3013

PLSR 0.7797 0.7801 0.1595 0.1698



# Basement

## Geochemistry – CLR

**LGBR**

LGBR took 16.54 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 0.12441994254121294, 'reg\_lambda': 0.030904882671394605, 'colsample\_bytree': 0.6, 'subsample': 0.4, 'learning\_rate': 0.04919730754782928, 'max\_depth': 8, 'num\_leaves': 10, 'min\_child\_samples': 16}

Best trial CV R² for LGBR: 0.8080

**XGBR**

XGBR took 31.87 seconds

Best trial parameters for XGBR: {'max\_depth': 3, 'gamma': 0.9535141850497593, 'eta': 0.029217618764923587, 'subsample': 0.8039622394008821, 'min\_child\_weight': 4, 'lambda': 116.41649073799965, 'alpha': 1.461683536553623, 'colsample\_bytree': 0.7046667219205766}

Best trial CV R² for XGBR: 0.6866

**RFR**

RFR took 28.34 seconds

Best trial parameters for RFR: {'n\_estimators': 22, 'max\_depth': 10, 'max\_leaf\_nodes': 22, 'min\_samples\_split': 0.1790631295788889, 'min\_samples\_leaf': 0.10010508949735887}

Best trial CV R² for RFR: 0.6632

**SVR**

SVR took 16.75 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.8715818565204648, 'epsilon': 0.08928484162313116, 'gamma': 0.02149073945361247}

Best trial CV R² for SVR: 0.8241

**PLSR**

PLSR took 7.38 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.005269855418393144, 'scale': True}

Best trial CV R² for PLSR: 0.7670

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

LGBR 0.9006 0.7914 0.2432 0.3701

XGBR 0.7437 0.7058 0.3904 0.4396

RFR 0.6982 0.6781 0.4236 0.4598

SVR 0.9419 0.7846 0.1859 0.3761

PLSR 0.7917 0.7474 0.3519 0.4073

## Geochemistry – Isometric log-ratio (ILR)

**LGBR**

LGBR took 7.96 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 1.090757903211582, 'reg\_lambda': 1.1270781552763924, 'colsample\_bytree': 0.6, 'subsample': 0.4, 'learning\_rate': 0.048629244720344765, 'max\_depth': 6, 'num\_leaves': 10, 'min\_child\_samples': 16}

Best trial CV R² for LGBR: 0.7970

LGBR - Train R²: 0.8885, Test R²: 0.7917

LGBR - Train RMSE: 0.2575, Test RMSE: 0.3699

**XGBR**

XGBR took 13.34 seconds

Best trial parameters for XGBR: {'max\_depth': 7, 'gamma': 1.2969303787794748, 'eta': 0.027141505811964443, 'subsample': 0.8729974738036191, 'min\_child\_weight': 5, 'lambda': 221.44507679156715, 'alpha': 1.5502997657443096, 'colsample\_bytree': 0.7850381263898787}

Best trial CV R² for XGBR: 0.6402

XGBR - Train R²: 0.6925, Test R²: 0.6538

XGBR - Train RMSE: 0.4276, Test RMSE: 0.4769

**RFR**

RFR took 15.36 seconds

Best trial parameters for RFR: {'n\_estimators': 48, 'max\_depth': 3, 'max\_leaf\_nodes': 26, 'min\_samples\_split': 0.1708849531569665, 'min\_samples\_leaf': 0.1021328043079633}

Best trial CV R² for RFR: 0.6724

**SVR**

SVR took 5.68 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.241291949183715, 'epsilon': 0.08535537527092955, 'gamma': 0.01325583468658834}

Best trial CV R² for SVR: 0.7807

**PLSR**

PLSR took 4.12 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.00565545170921812, 'scale': False}

Best trial CV R² for PLSR: 0.7665

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

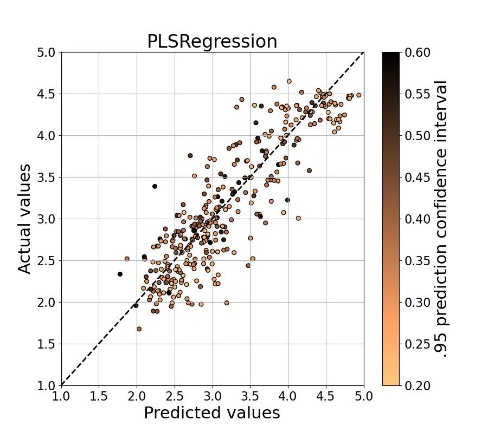
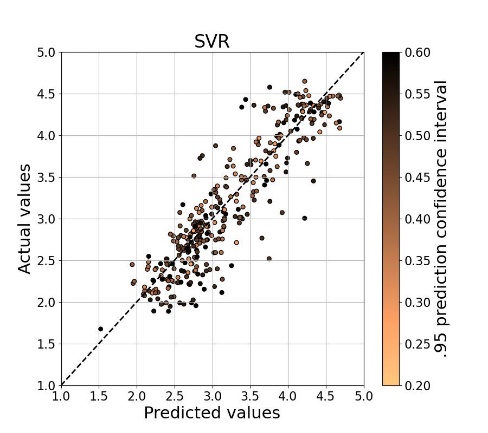
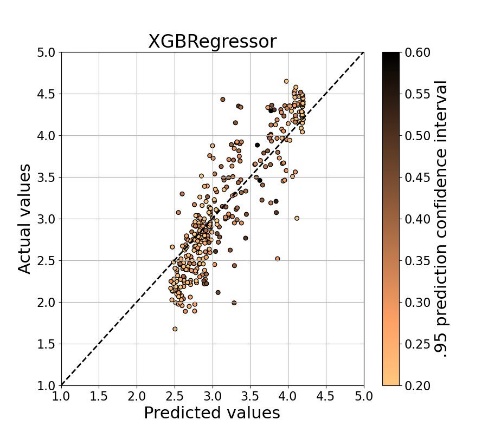
LGBR 0.8885 0.7917 0.2575 0.3699

XGBR 0.6925 0.6538 0.4276 0.4769

RFR 0.7020 0.6774 0.4209 0.4603

SVR 0.8370 0.7898 0.3113 0.3715

PLSR 0.7886 0.7458 0.3546 0.4086



## Geochemistry – Pairwise log-ratio (PWLR)

**LGBR**

LGBR took 140.34 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 0.163451639778182, 'reg\_lambda': 0.25434067224654566, 'colsample\_bytree': 0.5, 'subsample': 1.0, 'learning\_rate': 0.0492347699530963, 'max\_depth': 7, 'num\_leaves': 10, 'min\_child\_samples': 12}

Best trial CV R² for LGBR: 0.8279

**XGBR**

XGBR took 433.23 seconds

Best trial parameters for XGBR: {'max\_depth': 4, 'gamma': 0.9787306606371388, 'eta': 0.025668952050468076, 'subsample': 0.820606424407513, 'min\_child\_weight': 5, 'lambda': 125.35990246702654, 'alpha': 2.4462806903236967, 'colsample\_bytree': 0.6932718606021472}

Best trial CV R² for XGBR: 0.7091

**RFR**

RFR took 104.46 seconds

Best trial parameters for RFR: {'n\_estimators': 13, 'max\_depth': 4, 'max\_leaf\_nodes': 21, 'min\_samples\_split': 0.14643499285901948, 'min\_samples\_leaf': 0.10238335657874206}

Best trial CV R² for RFR: 0.7210

**SVR**

SVR took 281.47 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.9937012899546515, 'epsilon': 0.054847532746431574, 'gamma': 0.0006765365640979172}

Best trial CV R² for SVR: 0.8329

**PLSR**

PLSR took 18.15 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.0007343000969334157, 'scale': False}

Best trial CV R² for PLSR: 0.7705

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

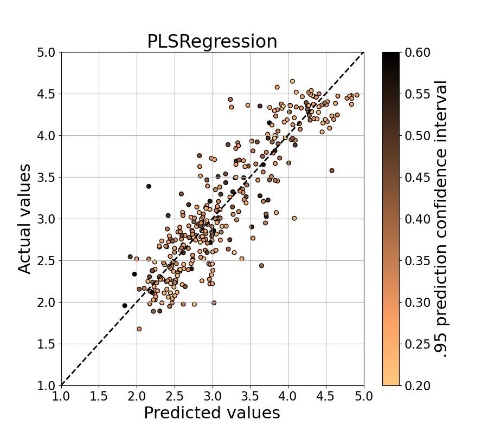
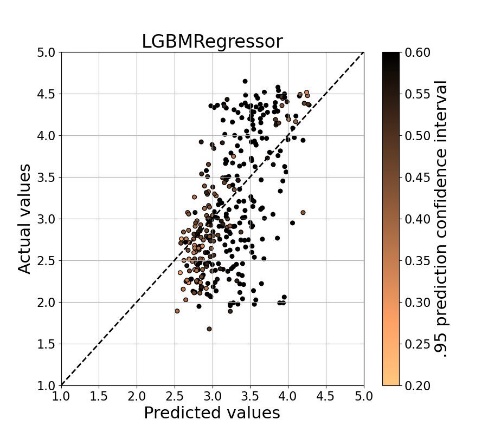
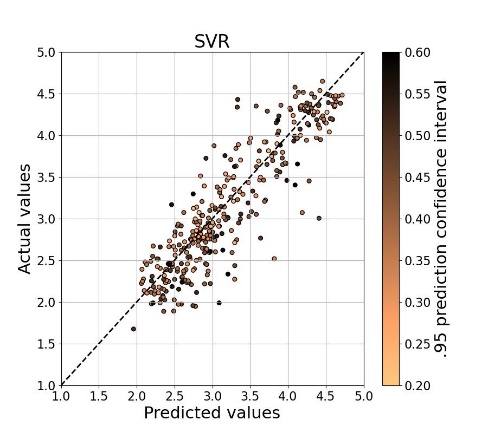
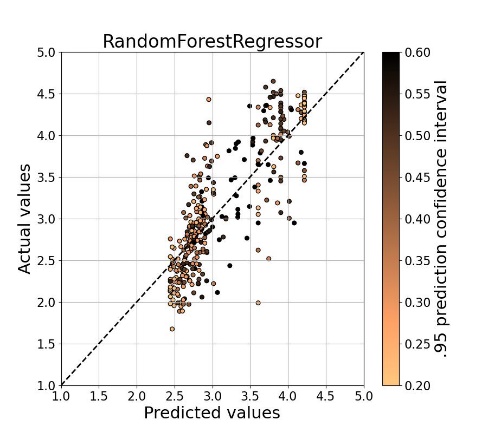
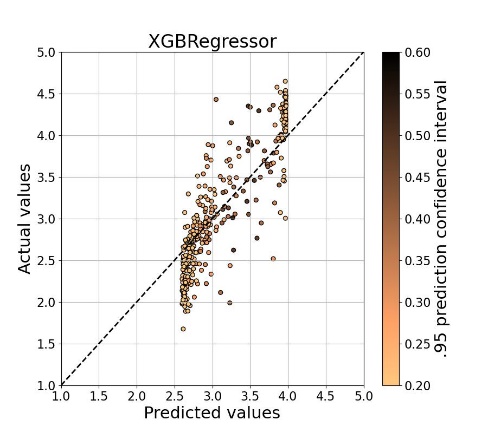
LGBR 0.9317 0.7989 0.2015 0.3634

XGBR 0.7795 0.7235 0.3621 0.4261

RFR 0.7578 0.6973 0.3795 0.4459

SVR 0.9403 0.7997 0.1884 0.3627

PLSR 0.7957 0.7494 0.3486 0.4057



## IR spectroscopy – SS

**LGBR**

LGBR took 66.54 seconds

Best trial parameters for LGBR: {'n\_estimators': 75, 'reg\_alpha': 0.668530663362835, 'reg\_lambda': 2.193421475871177, 'colsample\_bytree': 0.4, 'subsample': 1.0, 'learning\_rate': 0.04257130970580314, 'max\_depth': 4, 'num\_leaves': 6, 'min\_child\_samples': 34}

Best trial CV R² for LGBR: 0.3581

**XGBR**

XGBR took 316.48 seconds

Best trial parameters for XGBR: {'max\_depth': 4, 'gamma': 1.3458189389702122, 'eta': 0.028808363409804565, 'subsample': 0.8941553962991411, 'min\_child\_weight': 4, 'lambda': 521.7646295973493, 'alpha': 1.8987739145881846, 'colsample\_bytree': 0.8815269155044768}

Best trial CV R² for XGBR: 0.1813

**RFR**

RFR took 146.63 seconds

Best trial parameters for RFR: {'n\_estimators': 33, 'max\_depth': 8, 'max\_leaf\_nodes': 47, 'min\_samples\_split': 0.13358118615269682, 'min\_samples\_leaf': 0.1011315630694369}

Best trial CV R² for RFR: 0.2532

**SVR**

SVR took 103.36 seconds

Best trial parameters for SVR: {'kernel': 'linear', 'C': 0.3135622909972069, 'epsilon': 0.47206970562869427}

Best trial CV R² for SVR: 0.5603

**PLSR**

PLSR took 18.93 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.009178348586759394, 'scale': False}

Best trial CV R² for PLSR: 0.3919

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

LGBR 0.5281 0.3491 0.5297 0.6539

XGBR 0.2570 0.2251 0.6647 0.7134

RFR 0.3009 0.2830 0.6448 0.6863

SVR 0.6280 0.4923 0.4703 0.5775

PLSR 0.4239 0.3832 0.5853 0.6365

## IR spectroscopy – FOD

**LGBR**

LGBR took 130.99 seconds

Best trial parameters for LGBR: {'n\_estimators': 73, 'reg\_alpha': 0.9160233563629611, 'reg\_lambda': 0.33366663943809427, 'colsample\_bytree': 0.4, 'subsample': 0.7, 'learning\_rate': 0.04999819507401147, 'max\_depth': 7, 'num\_leaves': 10, 'min\_child\_samples': 31}

Best trial CV R² for LGBR: 0.5536

**XGBR**

XGBR took 942.22 seconds

Best trial parameters for XGBR: {'max\_depth': 8, 'gamma': 0.8873371049729264, 'eta': 0.02790110332625754, 'subsample': 0.877416871997709, 'min\_child\_weight': 3, 'lambda': 0.3024285434437246, 'alpha': 1.3450339611059832, 'colsample\_bytree': 0.6607541334383757}

Best trial CV R² for XGBR: 0.5462

**RFR**

RFR took 165.24 seconds

Best trial parameters for RFR: {'n\_estimators': 50, 'max\_depth': 7, 'max\_leaf\_nodes': 10, 'min\_samples\_split': 0.1747524880999196, 'min\_samples\_leaf': 0.10044693741506303}

Best trial CV R² for RFR: 0.3680

**SVR**

SVR took 244.79 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.4228947729205939, 'epsilon': 0.011008002670199721, 'gamma': 0.0006775979901830446}

Best trial CV R² for SVR: 0.5644

**PLSR**

PLSR took 19.19 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 6.575111027924598e-05, 'scale': False}

Best trial CV R² for PLSR: 0.4905

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

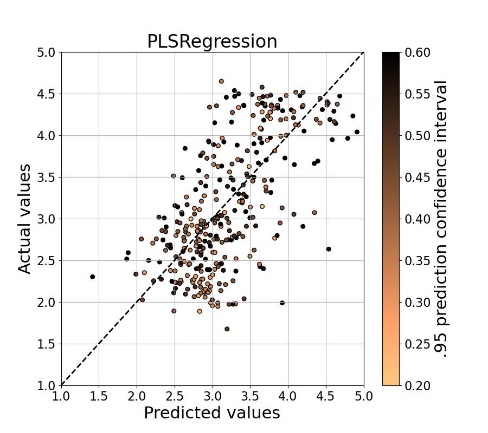
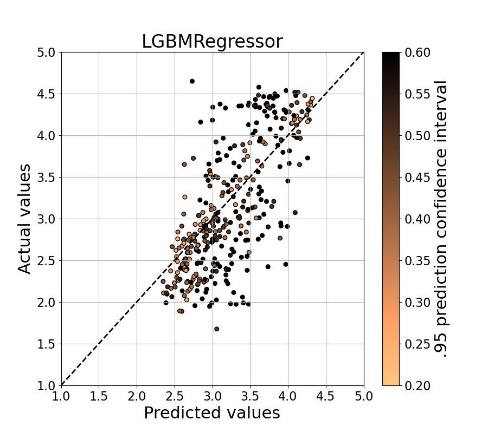
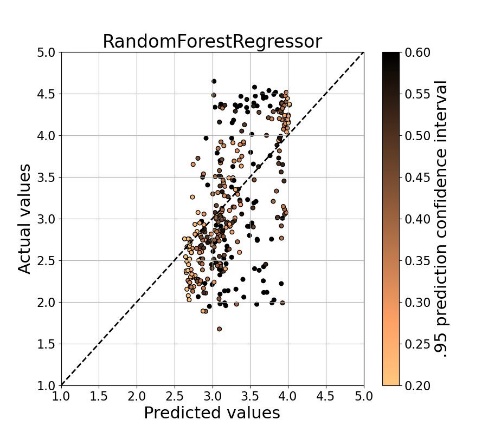
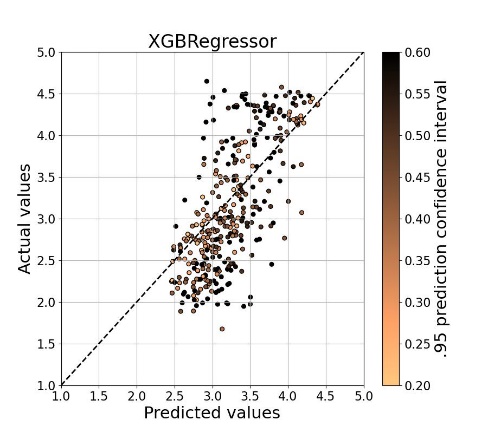
LGBR 0.7986 0.4990 0.3460 0.5737

XGBR 0.8559 0.4789 0.2928 0.5850

RFR 0.4182 0.2699 0.5882 0.6925

SVR 0.6853 0.5488 0.4326 0.5444

PLSR 0.5443 0.3847 0.5206 0.6357



## IR spectroscopy – SOD

**LGBR**

LGBR took 153.02 seconds

Best trial parameters for LGBR: {'n\_estimators': 74, 'reg\_alpha': 0.10350027782838847, 'reg\_lambda': 0.23218042835472583, 'colsample\_bytree': 0.6, 'subsample': 0.7, 'learning\_rate': 0.045515546290046054, 'max\_depth': 6, 'num\_leaves': 9, 'min\_child\_samples': 32}

Best trial CV R² for LGBR: 0.5170

**XGBR**

XGBR took 478.13 seconds

Best trial parameters for XGBR: {'max\_depth': 5, 'gamma': 1.9072535609382792, 'eta': 0.022685720441478176, 'subsample': 0.756362071610848, 'min\_child\_weight': 4, 'lambda': 49.5367199791086, 'alpha': 0.6591676494836101, 'colsample\_bytree': 0.7348761038977595}

Best trial CV R² for XGBR: 0.4425

**RFR**

RFR took 163.49 seconds

Best trial parameters for RFR: {'n\_estimators': 45, 'max\_depth': 10, 'max\_leaf\_nodes': 12, 'min\_samples\_split': 0.2161591688483735, 'min\_samples\_leaf': 0.14284406411017533}

Best trial CV R² for RFR: 0.3258

**SVR**

SVR took 317.52 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.9064380756909942, 'epsilon': 0.016107691852297053, 'gamma': 0.0006403896610549008}

Best trial CV R² for SVR: 0.5828

**PLSR**

PLSR took 17.59 seconds

Best trial parameters for PLSR: {'n\_components': 7, 'tol': 0.00562112405028612, 'scale': False}

Best trial CV R² for PLSR: 0.4989

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

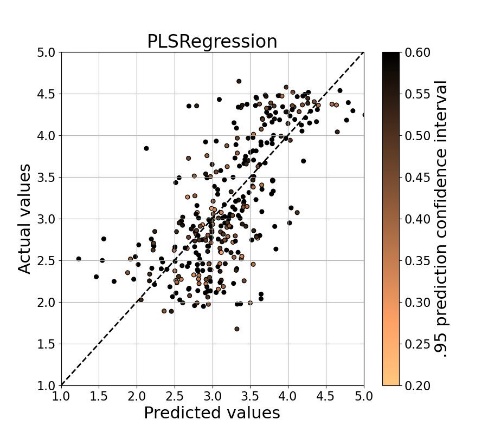
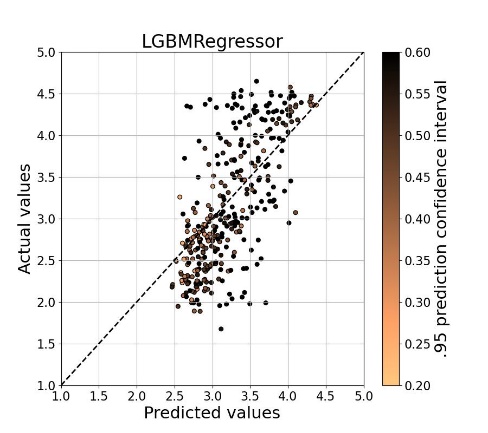
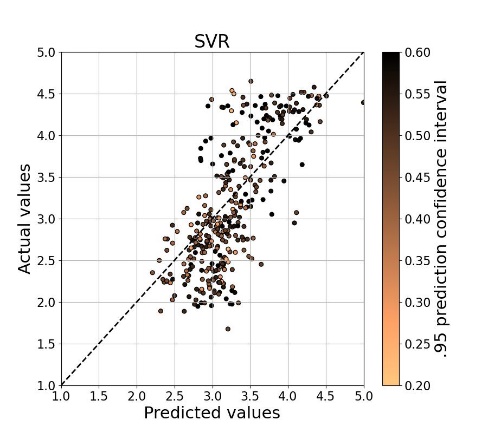
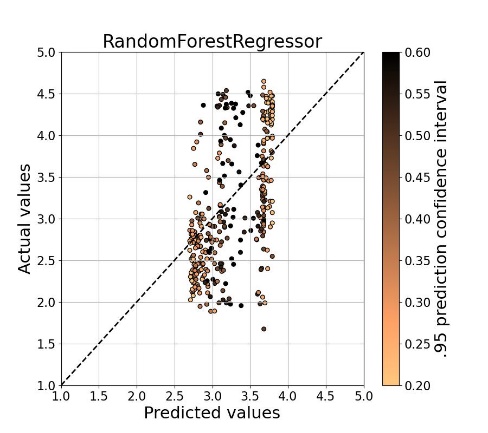
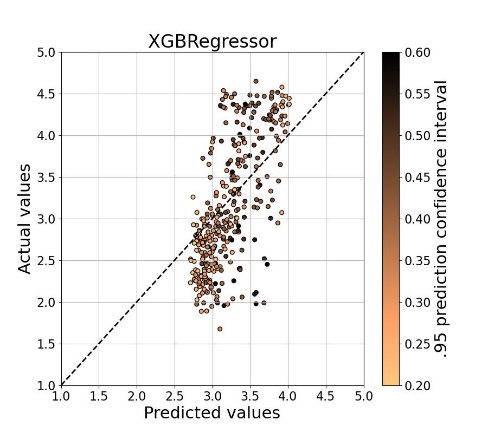
LGBR 0.7929 0.4799 0.3509 0.5845

XGBR 0.6220 0.4123 0.4741 0.6213

RFR 0.3479 0.2709 0.6227 0.6920

SVR 0.8208 0.4990 0.3264 0.5736

PLSR 0.6200 0.4577 0.4754 0.5969



## IR spectroscopy – Continuum removal

**LGBR**

LGBR took 135.03 seconds

Best trial parameters for LGBR: {'n\_estimators': 73, 'reg\_alpha': 0.38851770737944696, 'reg\_lambda': 0.03328098675881456, 'colsample\_bytree': 0.4, 'subsample': 0.4, 'learning\_rate': 0.046313715632654226, 'max\_depth': 9, 'num\_leaves': 10, 'min\_child\_samples': 16}

Best trial CV R² for LGBR: 0.3973

**XGBR**

XGBR took 687.60 seconds

Best trial parameters for XGBR: {'max\_depth': 7, 'gamma': 0.605673834098159, 'eta': 0.023841118587276677, 'subsample': 0.7554245845219756, 'min\_child\_weight': 1, 'lambda': 2.955665297008025, 'alpha': 1.9074580011554794, 'colsample\_bytree': 0.709144989199581}

Best trial CV R² for XGBR: 0.3944

**RFR**

RFR took 212.18 seconds

Best trial parameters for RFR: {'n\_estimators': 50, 'max\_depth': 5, 'max\_leaf\_nodes': 19, 'min\_samples\_split': 0.11720660124594127, 'min\_samples\_leaf': 0.10464642374808049}

Best trial CV R² for RFR: 0.2490

**SVR**

SVR took 233.41 seconds

Best trial parameters for SVR: {'kernel': 'linear', 'C': 0.5947830858781697, 'epsilon': 0.3843483121456853}

Best trial CV R² for SVR: 0.5580

**PLSR**

PLSR took 20.54 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.009568595935091136, 'scale': False}

Best trial CV R² for PLSR: 0.3939

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

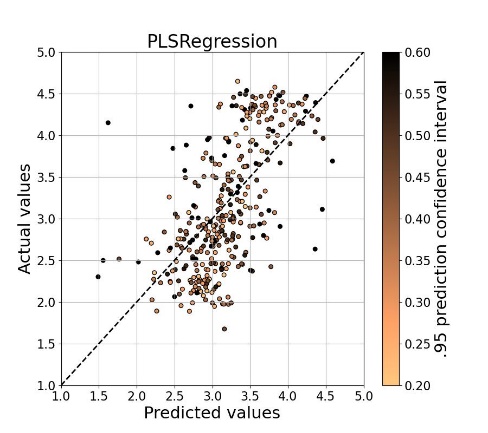
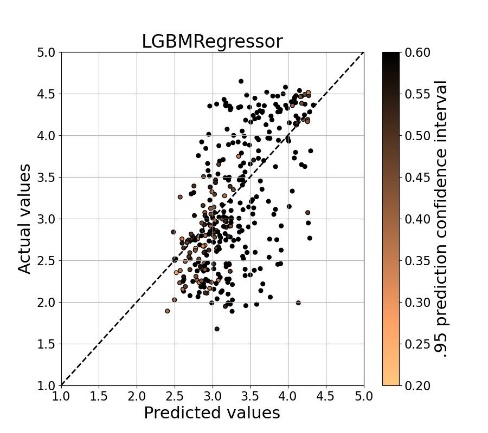
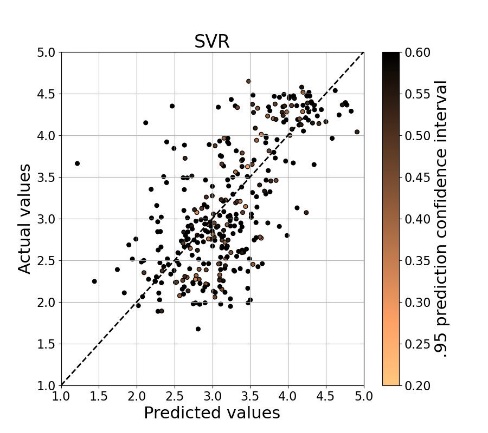
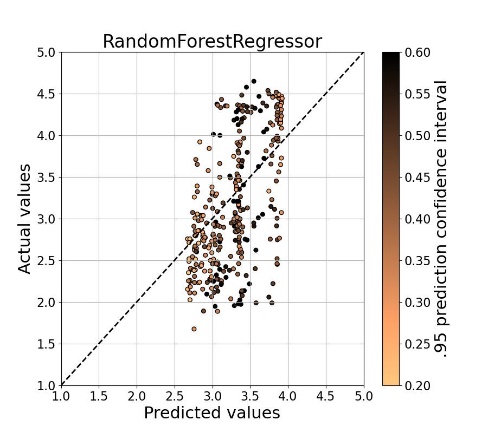
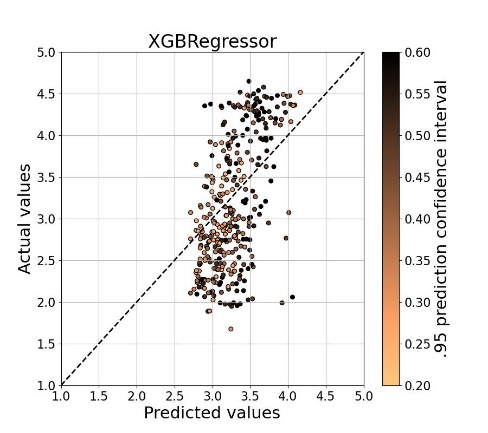
LGBR 0.6824 0.3378 0.4346 0.6595

XGBR 0.7510 0.2840 0.3848 0.6858

RFR 0.2986 0.2804 0.6458 0.6875

SVR 0.6949 0.4152 0.4260 0.6198

PLSR 0.4264 0.3857 0.5840 0.6352



## Data fusion

**LGBR**

LGBR took 232.03 seconds

Best trial parameters for LGBR: {'n\_estimators': 58, 'reg\_alpha': 0.10914357438914202, 'reg\_lambda': 0.057071650151390534, 'colsample\_bytree': 0.5, 'subsample': 1.0, 'learning\_rate': 0.04821223214253982, 'max\_depth': 9, 'num\_leaves': 10, 'min\_child\_samples': 14}

Best trial CV R² for LGBR: 0.8209

**XGBR**

XGBR took 1004.52 seconds

Best trial parameters for XGBR: {'max\_depth': 9, 'gamma': 1.3057004678715194, 'eta': 0.02510327016745332, 'subsample': 0.7582104638930908, 'min\_child\_weight': 1, 'lambda': 224.12314153396875, 'alpha': 6.4730161836979185, 'colsample\_bytree': 0.6478715903598818}

Best trial CV R² for XGBR: 0.6412

**RFR**

RFR took 315.55 seconds

Best trial parameters for RFR: {'n\_estimators': 33, 'max\_depth': 9, 'max\_leaf\_nodes': 14, 'min\_samples\_split': 0.20160567466621027, 'min\_samples\_leaf': 0.11051287001491997}

Best trial CV R² for RFR: 0.7173

**SVR**

SVR took 1343.78 seconds

Best trial parameters for SVR: {'kernel': 'rbf', 'C': 0.9974330494850988, 'epsilon': 0.01619400157133186, 'gamma': 0.00025944403829962986}

Best trial CV R² for SVR: 0.8320

**PLSR**

PLSR took 34.27 seconds

Best trial parameters for PLSR: {'n\_components': 8, 'tol': 0.0033343160679535936, 'scale': False}

Best trial CV R² for PLSR: 0.7857

**Summary of Results:**

Model Train R² Test R² Train RMSE Test RMSE

----------------------------------------------------------------

LGBR 0.9176 0.7934 0.2214 0.3683

XGBR 0.7052 0.6589 0.4187 0.4734

RFR 0.7416 0.6840 0.3920 0.4556

SVR 0.9366 0.8173 0.1941 0.3464

PLSR 0.8199 0.7515 0.3273 0.4040

