Supplementary Information

A Simplified Machine Learning Workflow for Identifying Potential Singlet Fission Candidates: Benzannulated Biphenylenes as a Case Study

Iqra Sarfraz, Sergei F. Vyboishchikov, Miquel Solà* and Albert Artigas*

Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, C/ Maria Aurèlia Capmany, 69, 17003 Girona, Catalonia, Spain.

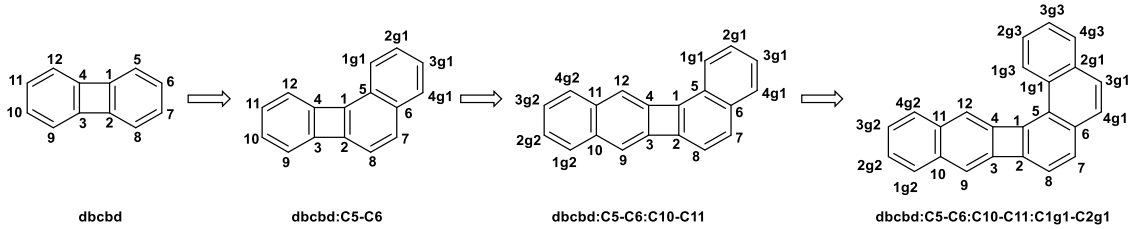
e-mail: albert.artigas@udg.edu; miquel.sola@udg.edu

Computational details

The library of benzannulated biphenylenes was automatically generated using *Sbuilder.py*[1], a purposed made python code, using the following command:

```
>> python Sbuilder.py dbcbd.xyz -gen 6 -out dbcbd_6_gen.xyz -opt mmff
```

The code sequentially attaches benzene rings to an input scaffold and assigns an internal code in the following way:



Scheme S1. *Example of internal code nomenclature used in the structure generation process. Indices of each next-generation structure show the C–C bond to which the new benzene ring is aligned. Numbers such as* ***3g2*** *mean the "3rd atom of the 2nd attachment generation", while numbers without g mean atoms from the original biphenylene scaffold.*

Initial geometries were automatically pre-optimized using the Merck molecular force field (MMFF).[2] The geomatries were gathered in a multi-xyz file provided separately as supplementary information (dbcbd_6_gen.xyz).

Randomly picked members of the library were reoptimized at the (U)M06-2X[3]/def2-SVP[4] level of theory for both the S_0 and T_1 states and single point energy calculations were carried out with the same functional using the def2-TZVPP[4] basis set. The restricted formalism was used for the calculation of the S_0 states and the unrestricted one for the T_1 states. Calculations of vibrational frequencies verified that the optimized geometries are minima in the potential energy surfaces. Vertical S1 (at S_0 geometry) excited states were computed using time-dependent density functional theory (TD-DFT)[5] with the same method. The reported S1/T1 ratios were obtained using electronic energies obtained computed at the (U)M06-2X/def2-TZVPP level of theory.

All calculations were performed using Gaussian16 program.[6] Gaussian input files were automatically generated using the QPREP module included in the AQME program.[7] using the following command line:

```
>> python -m aqme --qprep --program "gaussian" --qm_input "m062x/def2svp
opt freq=noraman" --files "dbcbd_6_gen.xyz"
```

Manipulation of input files to include single point energy calculations and/or TD-DFT calculations was done using standard bash scripting. Examples of input files are provided separately as supplementary information (dbcbd_6_gen_conf_1.com and dbcbd_6_gen_conf_1_t1.com)

The input file used for training set generation (biphenylenes_gen_6_round_1_code_smiles.csv) contained a code name for each member of the training set and their corresponding SMILES, which was automatically obtained from xyz files using the *Open Babel* program.[8]

```
>> obabel -i *.xyz -osmi
```

Training set generation was done using the CSEARCH and QDESCP modules included in the AQME program (v. 1.7.2) using the following commands:

```
>> python -m aqme --csearch --program "rdkit" --input "biphenylenes_gen_6_round_1_code_smiles.csv"

>> python -m aqme --qdescp --program "xtb" --files "CSEARCH/*.sdf"
```

The .csv file resulting from the QDESCP module of AQME (biphenylenes_gen_6_round_1_positives_QDESCP_interpret_descriptors) was used for model training using ROBERT[9] (v 2.0.0) after including a new column that corresponds to the target variable (S1/T1), obtained from the corresponding Gaussian output files. The command line used was:

```
>> python -m robert --y "S1/T1" --names "code_name" --csv_name "QDESCP_interpret_descriptors.csv"
```

To predict the S1/T1 ratio of all members in the library, a new descriptor database (biphenylenes_gen_6_full_QDESCP_interpret_descriptors.csv) was created using the CSEARCH and QDESCP modules included in the AQME program, using the following command:

```
>> python -m robert --predict --csv_test biphenylenes_gen_6_full_QDESCP_interpret_descriptors.csv.csv
```

A new training set containing 89 extra points with predicted S1/T1 values greater than 2.0 (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors) was constructed repeating the steps described above.

A test (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors) set made of randomly picker members of the library was used in this second round of model training to assess model's performance. The command use was the following:

```
>> python -m robert --y "S1/T1" --names "code_name" --csv_name "QDESCP_interpret_descriptors.csv" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv"
```

Predictions were done again using the full database using the same command. Predicted values are gathered in a .csv file provided separately as supplementary information (biphenylenes_gen_6_round_2_predicted_NN_No_PFI).

Additional model training testing custom combinations of descriptors and algorithms was done using ROBERT's graphical user interface *EasyROB* available at: https://robert.readthedocs.io/en/latest/Install/gui.html. The results obtained are gathered in Table S1 of this document.

All input and output files necessary to reproduce the results are included as supplementary information together with the corresponding ROBERT reports.

Table S1. *Metrics obtained in round 2 of model training using custom combinations of descriptors and algorithms*.

Entry	Descriptors						Algorithm	R2		MAE		RMSE		Robert Score
	S_0-T_1 gap	Electronegativity	Second EA	G of H-bonds H_2O	MolLogP	Dipole module		5xCV	TEST	5xCV	TEST	5xCV	TEST	
1	yes	yes	yes	yes	yes	yes	NN	0.95	0.96	0.061	0.057	0.086	0.076	8
2	yes	yes	yes	No	No	No	NN	0.94	0.94	0.067	0.064	0.093	0.094	8
3	yes	yes	yes	No	No	No	RF	0.91	0.9	0.086	0.082	0.12	0.12	8
4	yes	yes	yes	No	No	No	GB	0.93	0.94	0.072	0.063	0.1	0.094	8
5	yes	yes	yes	No	No	No	MVL	0.88	0.87	0.097	0.1	0.13	0.13	8
6	yes	yes	No	No	No	No	NN	0.93	0.94	0.072	0.067	0.098	0.094	9
7	yes	yes	No	No	No	No	RF	0.9	0.89	0.093	0.09	0.12	0.13	8
8	yes	yes	No	No	No	No	GB	0.92	0.92	0.077	0.072	0.1	0.1	8
9	yes	yes	No	No	No	No	MVL	0.88	0.87	0.097	0.1	0.13	0.13	8
10	yes	No	No	No	No	No	NN	0.92	0.9	0.078	0.079	0.11	0.12	8
11	yes	No	No	No	No	No	RF	0.91	0.91	0.081	0.077	0.11	0.11	8
12	yes	No	No	No	No	No	GB	0.92	0.89	0.081	0.081	0.11	0.12	8
13	yes	No	No	No	No	No	MVL	0.88	0.86	0.096	0.1	0.13	0.14	8

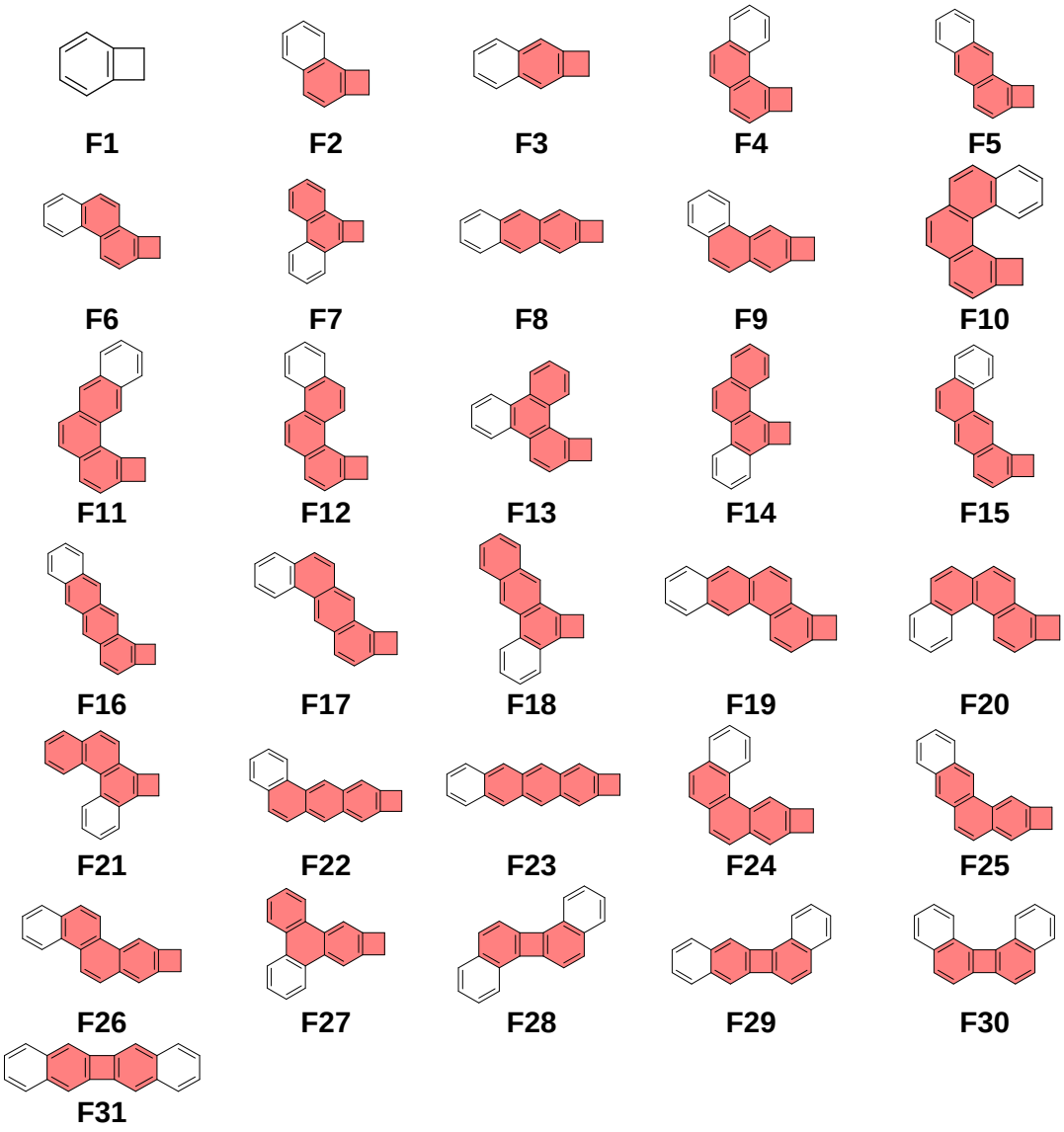


Figure S1. *Chemical structure of the fragments considered in the structural analysis of hit compounds.*

Table S2. *Fragment statistics in the 3835 molecules (with 505 positives). The most indicative fragments (#5 (cyclobuta[a]anthracene), #29 (dibenzo[a,g]biphenylene) and #30 (dibenzo[a,i]biphenylene) are shown in blue.*[a]

F	Fragment	False Negatives	False Positives	True Positives	True Negatives	Accuracy (%)	Precision (%)	Sensitivity (%)	True-Negative Rate (%)	F-score
1	c1(cccc2)c2cc1	0	3330	505	0	13.2	13.2	100.0	0.0	0.23
2	c1(c(cccc2)c2cc3)c3cc1	6	2757	499	573	28.0	15.3	98.8	17.2	0.27
3	c1(cc(cccc2)c2c3)c3cc1	480	1508	25	1822	48.2	1.6	5.0	54.7	0.03
4	c1(c(c(cccc2)c2cc3)c3cc4)c4cc1	333	1338	172	1992	56.4	11.4	34.1	59.8	0.17
5	**c1(c(cc(cccc2)c2c3)c3cc4)c4cc1**	**107**	**661**	**398**	**2669**	**80.0**	**37.6**	**78.8**	**80.2**	**0.51**
6	c1(c(ccc2c3cccc2)c3cc4)c4cc1	372	1377	133	1953	54.4	8.8	26.3	58.6	0.13
7	c1(c(cccc2)c2c3c4cccc3)c4cc1	226	718	279	2612	75.4	28.0	55.2	78.4	0.37
8	c1(cc(cc(cccc2)c2c3)c3c4)c4cc1	493	375	12	2955	77.4	3.1	2.4	88.7	0.03
9	c1(cc(c(cccc2)c2cc3)c3c4)c4cc1	498	906	7	2424	63.4	0.8	1.4	72.8	0.01
10	c1(c(c(c(cccc2)c2cc3)c3cc4)c4cc5)c5cc1	462	366	43	2964	78.4	10.5	8.5	89.0	0.09
11	c1(c(c(cc(cccc2)c2c3)c3cc4)c4cc5)c5cc1	485	295	20	3035	79.7	6.3	4.0	91.1	0.05
12	c1(c(c(ccc2c3cccc2)c3cc4)c4cc5)c5cc1	465	369	40	2961	78.3	9.8	7.9	88.9	0.09
13	c1(c(c(cccc2)c2c3c4cccc3)c4cc5)c5cc1	492	475	13	2855	74.8	2.7	2.6	85.7	0.03
14	c1(c(c(cccc2)c2cc3)c3c4c5cccc4)c5cc1	398	339	107	2991	80.8	24.0	21.2	89.8	0.23
15	c1(c(cc(c(cccc2)c2cc3)c3c4)c4cc5)c5cc1	433	243	72	3087	82.4	22.9	14.3	92.7	0.18
16	c1(c(cc(cc(cccc2)c2c3)c3c4)c4cc5)c5cc1	353	83	152	3247	88.6	64.7	30.1	97.5	0.41
17	c1(c(cc(ccc2c3cccc2)c3c4)c4cc5)c5cc1	443	253	62	3077	81.9	19.7	12.3	92.4	0.15
18	c1(c(cc(cccc2)c2c3)c3c4c5cccc4)c5cc1	363	210	142	3120	85.1	40.3	28.1	93.7	0.33
19	c1(c(ccc2c3cc4c(cccc4)c2)c3cc5)c5cc1	493	303	12	3027	79.2	3.8	2.4	90.9	0.03
20	c1(c(ccc2c3c4c(cccc4)cc2)c3cc5)c5cc1	477	381	28	2949	77.6	6.8	5.5	88.6	0.06
21	c1(c(ccc2c3cccc2)c3c4c5cccc4)c5cc1	426	367	79	2963	79.3	17.7	15.6	89.0	0.17
22	c1(cc(cc(c(cccc2)c2cc3)c3c4)c4c5)c5cc1	503	203	2	3127	81.6	1.0	0.4	93.9	0.01
23	c1(cc(cccc2)c2c3)c3cc(cc(cc4)c4c5)c5c1	497	84	8	3246	84.9	8.7	1.6	97.5	0.03
24	c1(cc(c(c(cccc2)c2cc3)c3cc4)c4c5)c5cc1	505	310	0	3020	78.7	0	0	90.7	nan
25	c1(cc(c(cc(cccc2)c2c3)c3cc4)c4c5)c5cc1	500	226	5	3104	81.1	2.2	1.0	93.2	0.01
26	c1(cc(c(ccc2c3cccc2)c3cc4)c4c5)c5cc1	505	310	0	3020	78.7	0	0	90.7	nan
27	c1(cc(c(cccc2)c2c3c4cccc3)c4c5)c5cc1	505	197	0	3133	81.7	0	0	94.1	nan
28	**c1(ccc2c3cccc2)c3c4c1c5c(cccc5)cc4**	**156**	**450**	**349**	**2880**	**84.2**	**43.7**	**69.1**	**86.5**	**0.54**
29	c1(c2c(cccc2)cc3)c3c4c1cc5c(cccc5)c4	486	936	19	2394	62.9	2.0	3.8	71.9	0.03
30	**c1(c(cccc2)c2cc3)c3c4c1c5c(cccc5)cc4**	**154**	**449**	**351**	**2881**	**84.3**	**43.9**	**69.5**	**86.5**	**0.54**
31	c1(cc(cccc2)c2c3)c3c4c1cc5c(cccc5)c4	502	202	3	3128	81.6	1.5	0.6	93.9	0.01

[a] Definitions: True positives: fragment present, $S_1/T_1>1.9$; True negatives: fragment not present, $S_1/T_1<=1.9$; False positive: fragment present, $S_1/T_1<=1.9$; False negative: fragment not present, $S_1/T_1>1.9$.

References

1. S. F. Vyboishchikov, Sbuilder, GitHub repository, 2024, https://github.com/vyboishchikov/Sbuilder (accessed 18 November 2025).
2. T. A. Halgreen MMFF VI. MMFF94s Option for Energy Minimization Studies, *J. Comput. Chem.* 1999, **20**, 720-729.
3. Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
4. F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
5. M. E. Casida, M. Huix-Rotllant, Progress in Time-Dependent Density-Functional Theory, *Annu. Rev. Phys. Chem.* 2012, **63**, 287.
6. Gaussian 16, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
7. J. V Alegre-Requena, S. S. Sowndarya V, R. Pérez-Soto, T. M. Alturaifi, R. S. Paton, C. V Juan Alegre-Requena, D. de and S. S. Sowndarya V contributed equally, AQME: Automated quantum mechanical environments for researchers and educators, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2023, **13**, e1663.
8. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, Open Babel: An Open chemical toolbox, *J. Cheminform*, 2011, **3**, 33.
9. D. Dalmau and J. V. Alegre-Requena, ROBERT: Bridging the Gap Between Machine Learning and Chemistry, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2024, **14**, e1733.

ROBERT v 2.0.0 2025/07/17 10:44:32

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

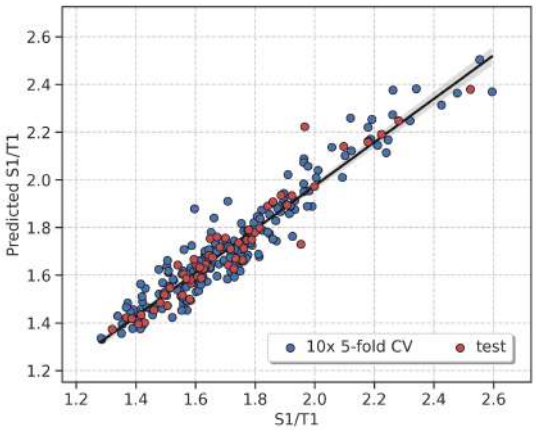
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 224:6



10x 5-fold CV : R^2 = 0.9, MAE = 0.056, RMSE = 0.074

Test : R^2 = 0.92, MAE = 0.046, RMSE = 0.066

Severe warnings

- No severe warnings detected

Moderate warnings

- Uneven y distribution (Section C)
- Highly correlated features (Section D)

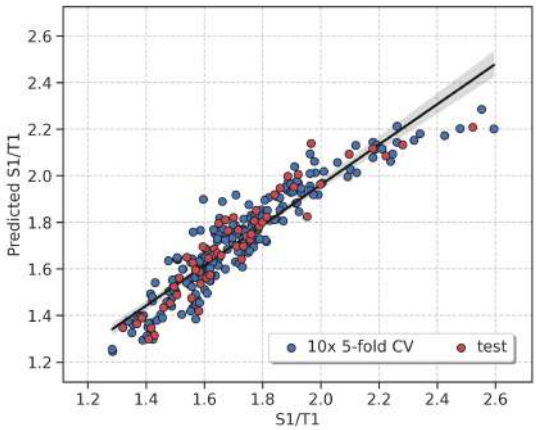
Overall assessment

- Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 224:4



10x 5-fold CV : R^2 = 0.86, MAE = 0.066, RMSE = 0.088

Test : R^2 = 0.87, MAE = 0.065, RMSE = 0.087

Severe warnings

- No severe warnings detected

Moderate warnings

- Uneven y distribution (Section C)
- Highly correlated features (Section D)

Overall assessment

- Decent model, but it has limitations

Section B. Advanced Score Analysis

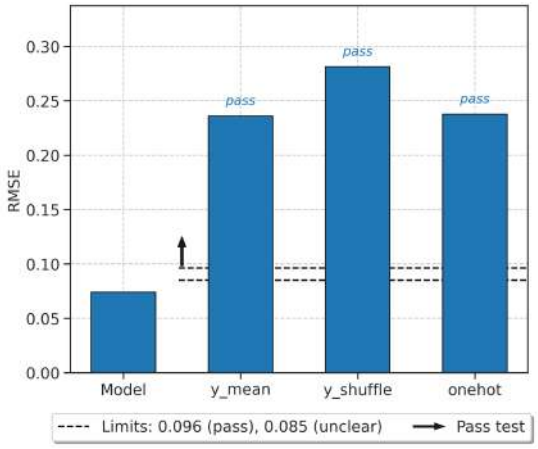
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· *Scoring from -6 to 0* ·

Pass: 0, Unclear: -1, Fail: -2.

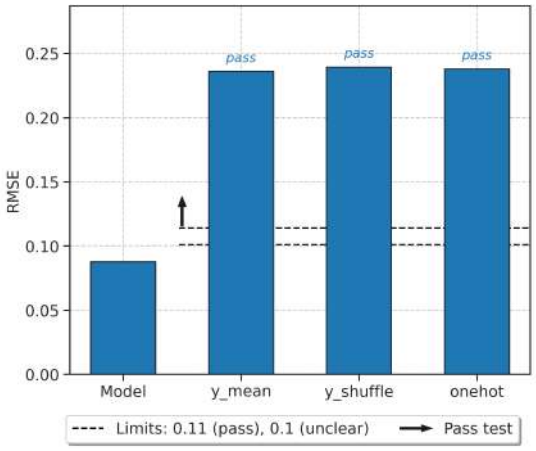


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· *Scoring from -6 to 0* ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 5.69%.

R^2 (10x 5-fold CV) = 0.9.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.77%.

R^2 (10x 5-fold CV) = 0.86.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 5.08%.

R^2 (test set) = 0.92.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.89*scaled RMSE (CV).

· *Scoring from 0 to 2* ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.69%.

R^2 (test set) = 0.87.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.99*scaled RMSE (CV).

· *Scoring from 0 to 2* ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

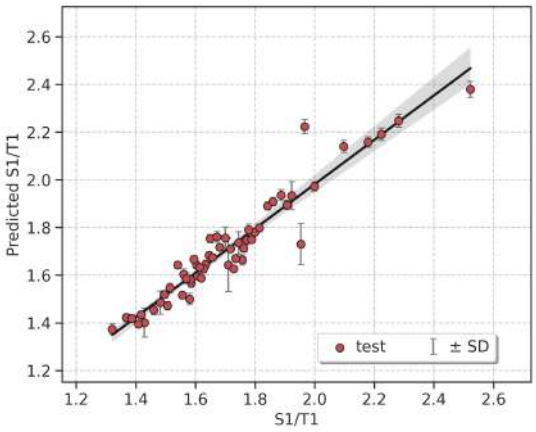
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (7% y-range).

· *Scoring from 0 to 2* ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

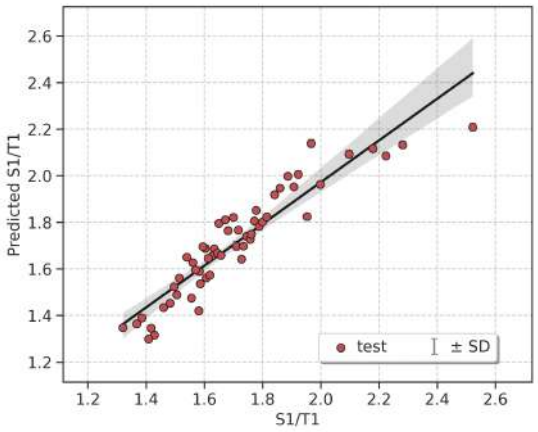


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (3% y-range).

· *Scoring from 0 to 2* ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[12.31%, 6.92%, 5.38%, 7.69%, 16.15%]

· *Scoring from 0 to 2* ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

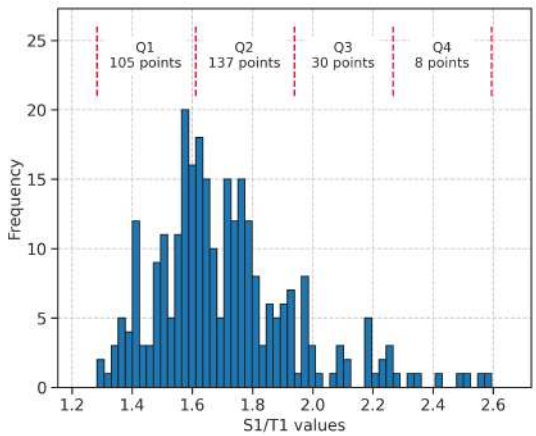
[6.15%, 7.69%, 6.92%, 4.62%, 16.15%]

· *Scoring from 0 to 2* ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

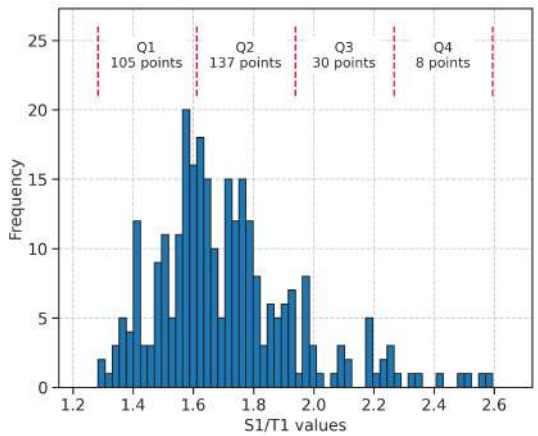
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 8 points while Q2 has 137)

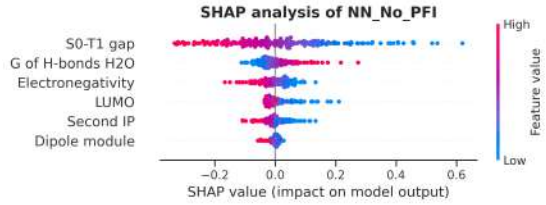


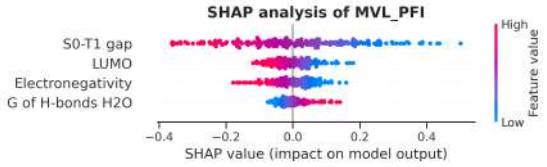
y distribution analysis

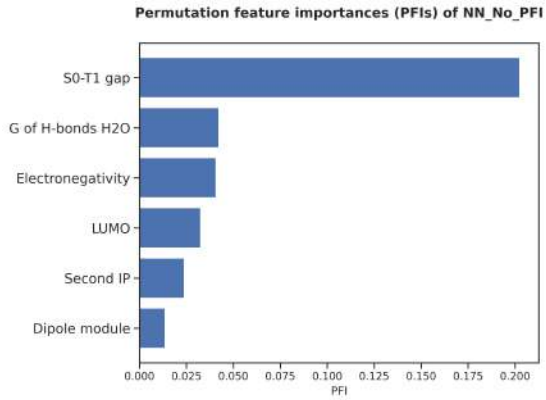
x WARNING! Your data is not uniform (Q4 has 8 points while Q2 has 137)

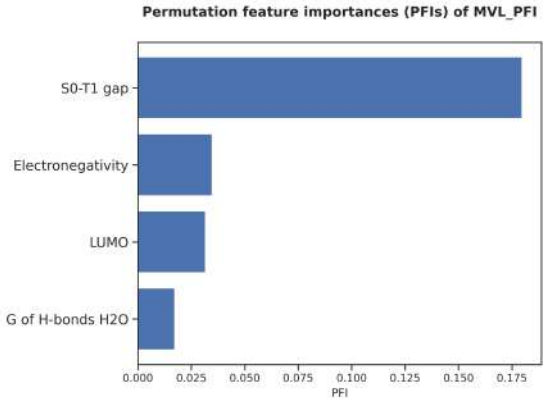
Section D. Feature Importances

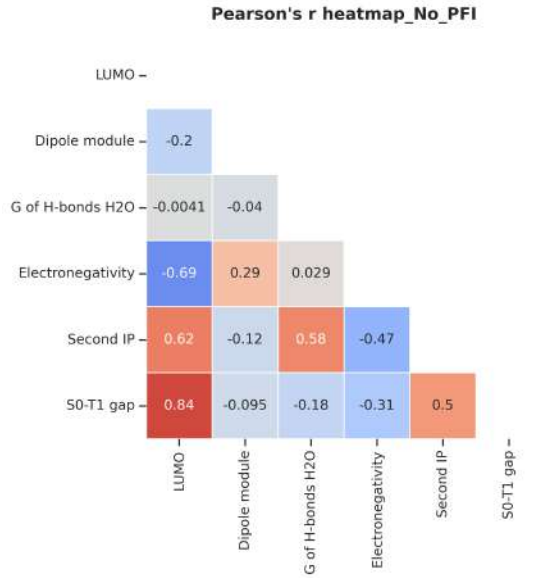
This section presents feature importances measured using the validation set.

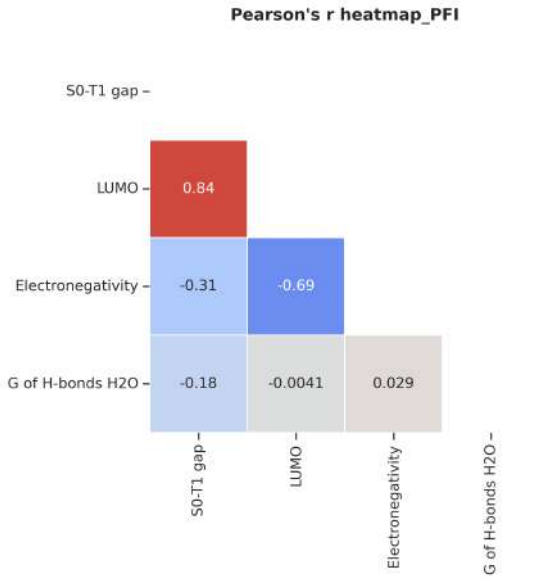












Correlation analysis

x WARNING! High correlations observed (up to r = 0.84 or R^2 = 0.71, for LUMO and S0-T1 gap)

Correlation analysis

x WARNING! High correlations observed (up to r = 0.84 or R^2 = 0.71, for S0-T1 gap and LUMO)

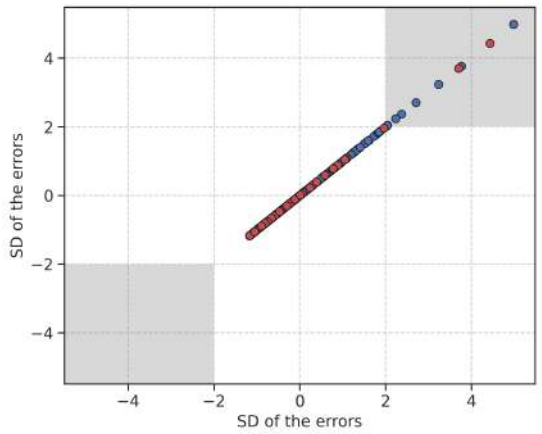
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 9 outliers out of 224 datapoints (4.0%)
- 501 (2.0 SDs)
- 1127 (3.8 SDs)
- 1362 (2.7 SDs)
- 1582 (2.4 SDs)
- 2053 (2.2 SDs)
- 2195 (5.0 SDs)
- 3044 (2.0 SDs)
- 3264 (3.2 SDs)
- 3555 (2.0 SDs)

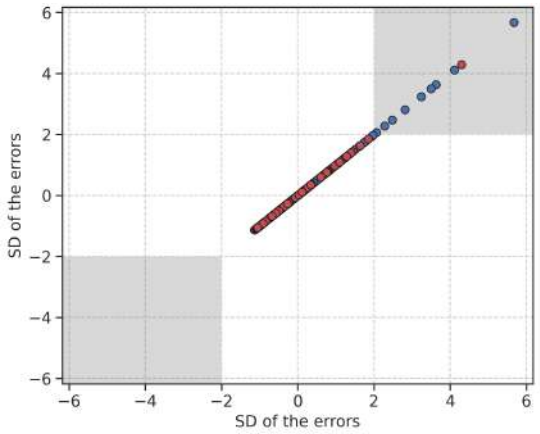
Test: 2 outliers out of 56 datapoints (3.6%)
- 345 (4.4 SDs)
- 2224 (3.7 SDs)



PFI (only important descriptors):

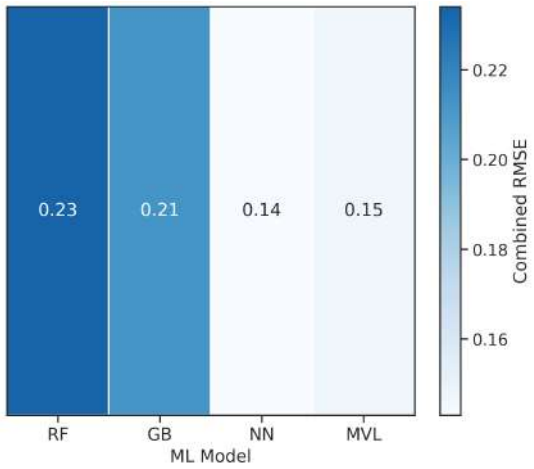
Outliers (max. 10 shown)
Train: 9 outliers out of 224 datapoints (4.0%)
- 245 (3.6 SDs)
- 609 (3.5 SDs)
- 875 (3.2 SDs)
- 1127 (5.7 SDs)
- 1362 (2.8 SDs)
- 2195 (4.1 SDs)
- 2920 (2.1 SDs)
- 3044 (2.3 SDs)
- 3264 (2.5 SDs)

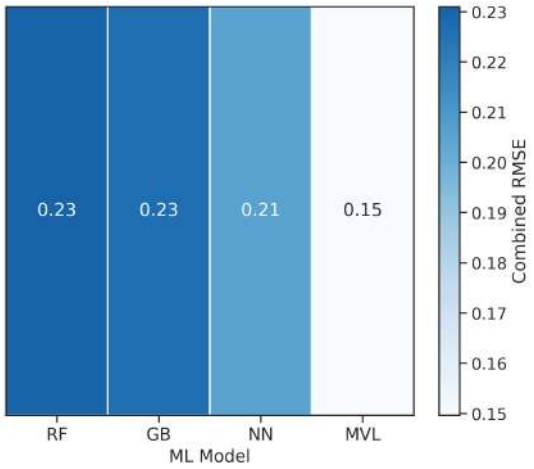
Test: 1 outliers out of 56 datapoints (1.8%)
- 207 (4.3 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.0
- Install scikit-learn-intelex: pip install scikit-learn-intelex==2025.0.1

(if scikit-learn-intelex is not installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV database:

python -m robert --y "S1/T1" --names "code_name" --csv_name "QDESCP_interpret_descriptors.csv"

4. Execution time, Python version and OS:

Originally run in Python 3.10.16 using Linux #1 SMP Thu Dec 7 03:06:13 EST 2023

Total execution time: 185.35 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):

sklearn model: MLPRegressor
hidden_layer_1: 7
hidden_layer_2: 6
max_iter: 385
alpha: 0.09493732706631618
tol: 7.136382691931351e-05
random_state: 0
solver: lbfgs

PFI (only important descriptors):

sklearn model: LinearRegression

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):

type: reg
kfold: 5
repeat_kfolds: 10
seed: 0
error_type: rmse

PFI (only important descriptors):

type: reg
kfold: 5
repeat_kfolds: 10
seed: 0
error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.0 2025/07/17 19:24:41

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

Section A. ROBERT Score

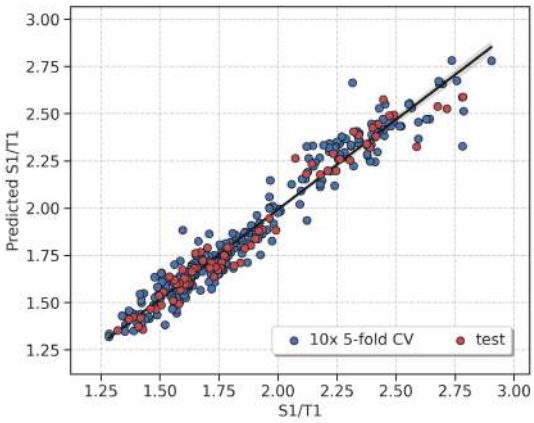
This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:6

MODERATE



10x 5-fold CV : R^2 = 0.95, MAE = 0.061, RMSE = 0.086

Test : R^2 = 0.96, MAE = 0.057, RMSE = 0.076

Severe warnings

- No severe warnings detected

Moderate warnings

- Uneven y distribution (Section C)
- Potential "faulty" outliers (Section E)

Overall assessment

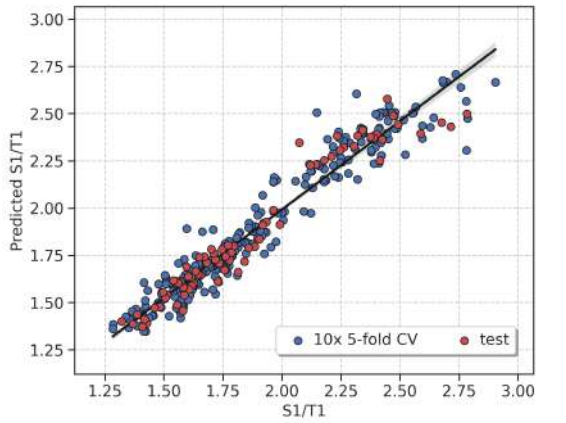
- Decent model, but it has limitations

PFI (only important descriptors) · Score 9

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:3

STRONG



10x 5-fold CV : R^2 = 0.94, MAE = 0.068, RMSE = 0.094

Test : R^2 = 0.94, MAE = 0.066, RMSE = 0.092

Severe warnings

- No severe warnings detected

Moderate warnings

- Uneven y distribution (Section C)

Overall assessment

- The model seems reliable

Section B. Advanced Score Analysis

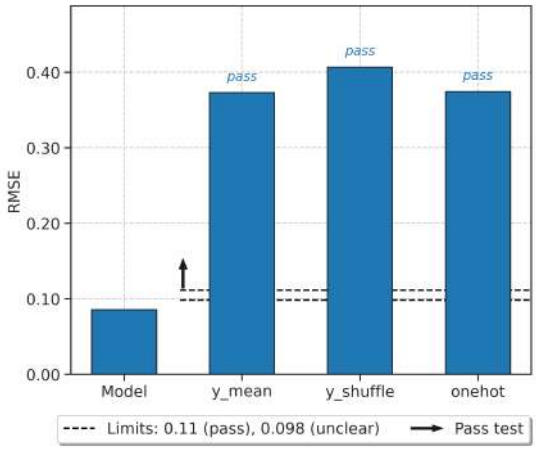
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· *Scoring from -6 to 0* ·

Pass: 0, Unclear: -1, Fail: -2.

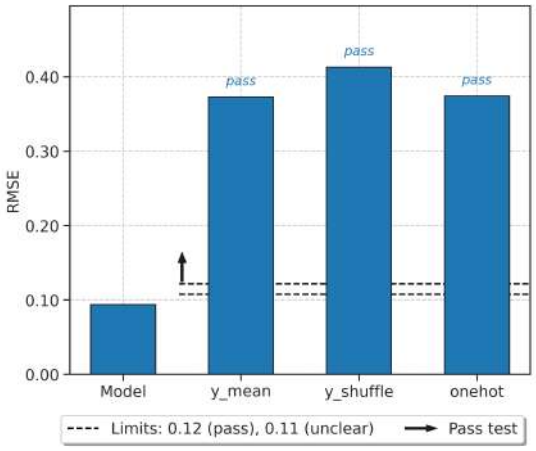


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· *Scoring from -6 to 0* ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 5.37%.

R^2 (10x 5-fold CV) = 0.95.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 5.88%.

R^2 (10x 5-fold CV) = 0.94.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 4.75%.

R^2 (test set) = 0.96.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.88*scaled RMSE (CV).

· *Scoring from 0 to 2* ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 5.75%.

R^2 (test set) = 0.94.

· *Scoring from 0 to 2* ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.98*scaled RMSE (CV).

· *Scoring from 0 to 2* ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

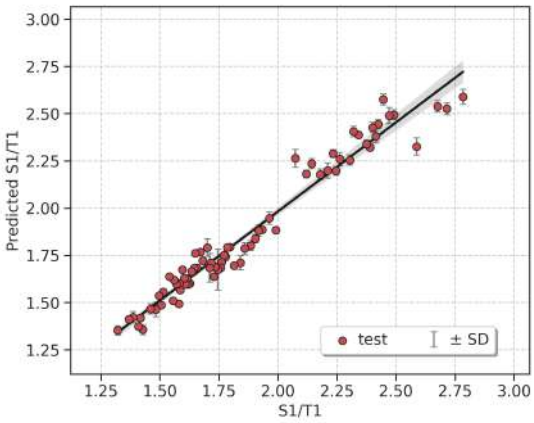
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· *Scoring from 0 to 2* ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[5.62%, 5.0%, 4.38%, 8.12%, 15.0%]

· *Scoring from 0 to 2* ·

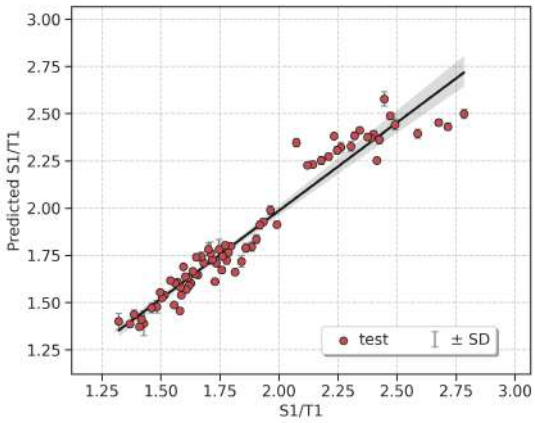
Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· *Scoring from 0 to 2* ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (1 / 2)

Scaled RMSEs across 5-fold CV:

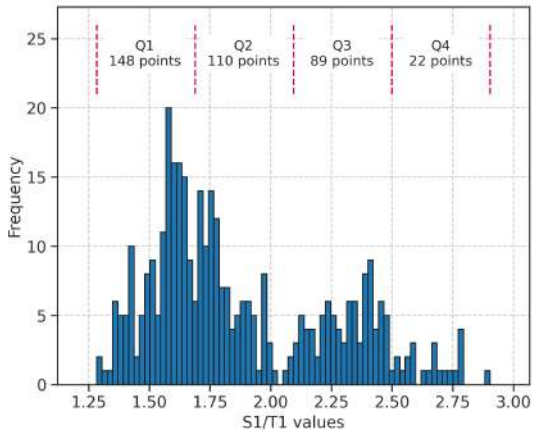
[6.25%, 5.0%, 5.0%, 9.37%, 14.37%]

· *Scoring from 0 to 2* ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

Section C. Distribution of y Values

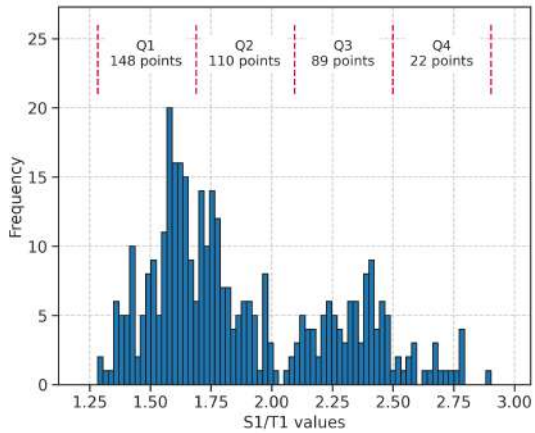
This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

Q1
148 points
Q2
110 points
Q3
89 points
Q4
22 points
Frequency
S1/T1 values

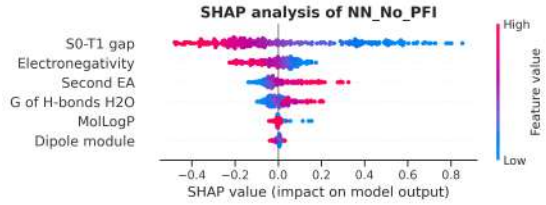


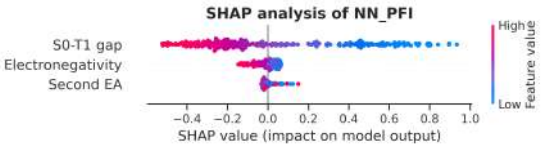
y distribution analysis

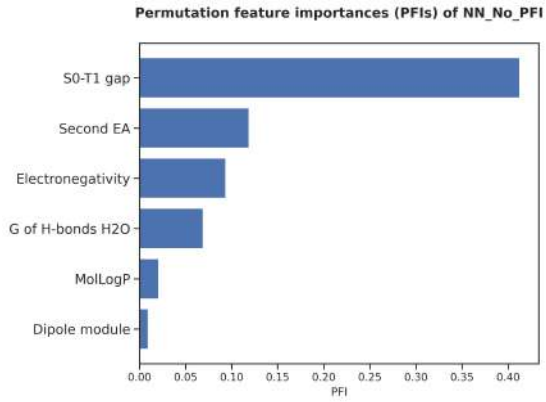
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

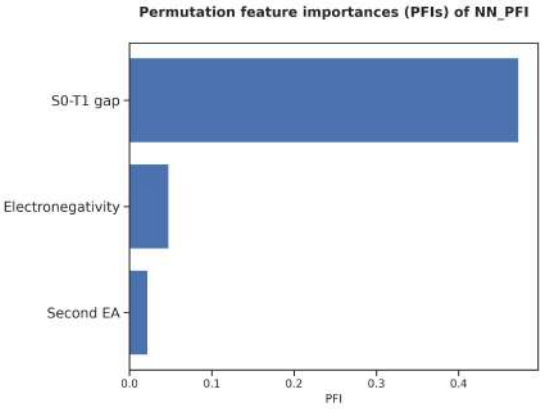
Section D. Feature Importances

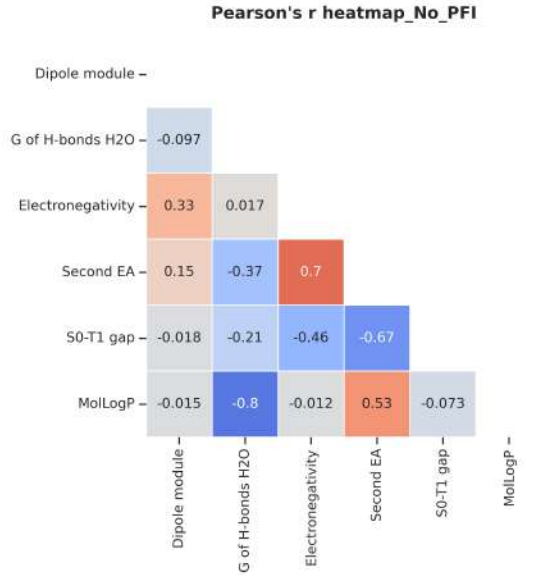
This section presents feature importances measured using the validation set.

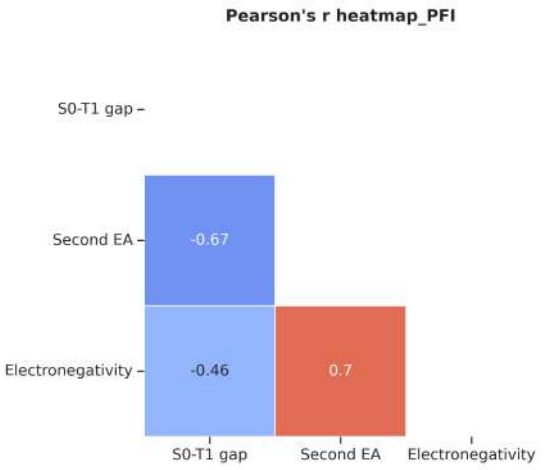












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

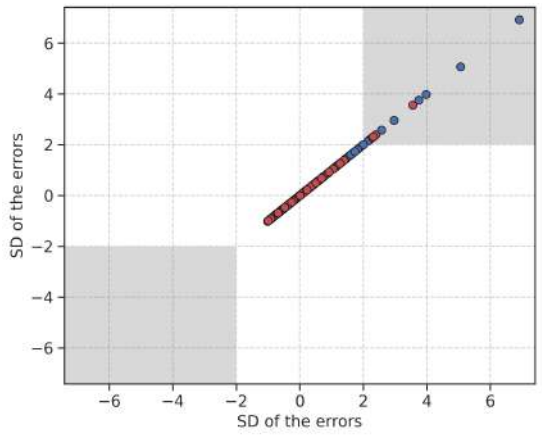
Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 885 (3.8 SDs)
- 1007 (6.9 SDs)
- 1130 (2.3 SDs)
- 1126 (3.0 SDs)
- 1006 (5.1 SDs)
- 2651 (2.2 SDs)
- 3117 (2.3 SDs)
- 879 (2.6 SDs)
- 345 (2.1 SDs)
- 1582 (2.0 SDs)

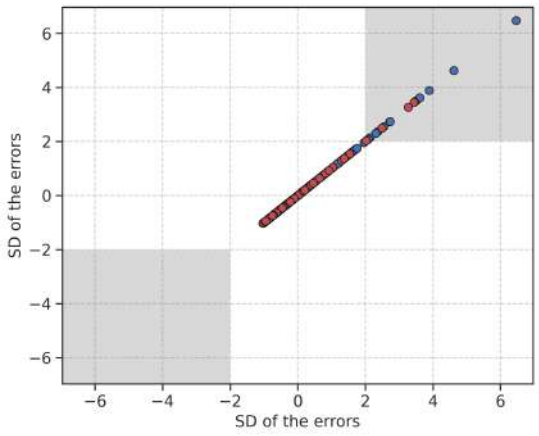
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (2.4 SDs)
- 884 (2.3 SDs)
- 3715 (3.6 SDs)
- 1483 (2.3 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 12 outliers out of 295 datapoints (4.1%)
- 2279 (2.7 SDs)
- 885 (3.9 SDs)
- 1007 (6.5 SDs)
- 1130 (2.4 SDs)
- 1003 (2.1 SDs)
- 1126 (2.5 SDs)
- 207 (2.7 SDs)
- 1006 (3.5 SDs)
- 2651 (4.6 SDs)
- 345 (2.1 SDs)

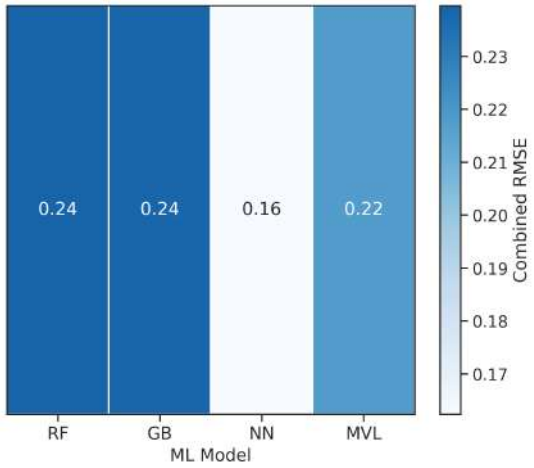
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (3.5 SDs)
- 884 (3.4 SDs)
- 890 (2.5 SDs)
- 3715 (2.0 SDs)
- 1483 (3.3 SDs)

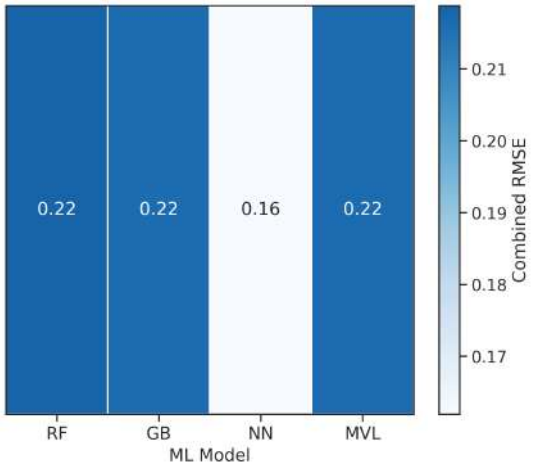




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.0
- Install scikit-learn-intelex: pip install scikit-learn-intelex==2025.0.1

(if scikit-learn-intelex is not installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --y "S1/T1" --names "code_name" --csv_name "QDESCP_interpret_descriptors.csv" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv"

4. Execution time, Python version and OS:

Originally run in Python 3.10.16 using Linux #1 SMP Thu Dec 7 03:06:13 EST 2023

Total execution time: 178.67 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: MLPRegressor	sklearn model: MLPRegressor
hidden_layer_1: 9	hidden_layer_1: 9
hidden_layer_2: 3	hidden_layer_2: 3
max_iter: 465	max_iter: 465
alpha: 0.0959440906711455	alpha: 0.0959440906711455
tol: 3.615185752570025e-05	tol: 3.615185752570025e-05
random_state: 0	random_state: 0
solver: lbfgs	solver: lbfgs

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.91, MAE = 0.048, RMSE = 0.059

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.24 ± 0.03
399	2.14	2.2 ± 0.03
1099	2.24	2.19 ± 0.02
2764	2.13	2.15 ± 0.02
997	2.12	2.11 ± 0.03
1015	2.13	2.09 ± 0.03
1141	2.12	2.09 ± 0.02
1754	2.08	2.09 ± 0.04
1769	2.01	2.07 ± 0.02
1123	2.04	2.06 ± 0.05
...	...	...
3687	1.42	1.39 ± 0.01
855	1.37	1.38 ± 0.01
3789	1.36	1.38 ± 0.01
3468	1.33	1.38 ± 0.04
3765	1.4	1.38 ± 0.01
2007	1.45	1.37 ± 0.01
3657	1.4	1.36 ± 0.01
3503	1.29	1.34 ± 0.02
788	1.3	1.33 ± 0.02
3779	1.26	1.31 ± 0.01

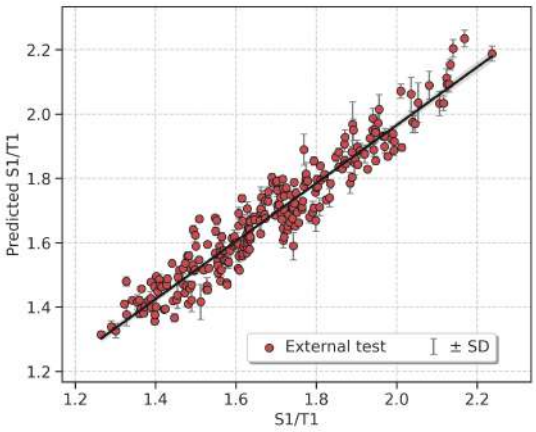
External test metrics

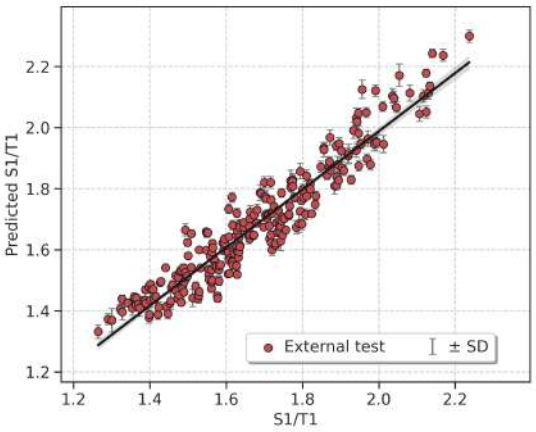
R2 = 0.9, MAE = 0.053, RMSE = 0.063

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1099	2.24	2.3 ± 0.02
399	2.14	2.24 ± 0.01
1853	2.17	2.24 ± 0.02
997	2.12	2.18 ± 0.01
20	2.05	2.17 ± 0.04
2764	2.13	2.14 ± 0.01
1883	1.96	2.13 ± 0.03
2749	1.99	2.12 ± 0.02
1015	2.13	2.11 ± 0.02
1754	2.08	2.11 ± 0.03
...	...	...
3516	1.37	1.41 ± 0.03
3801	1.32	1.41 ± 0.01
3468	1.33	1.4 ± 0.03
2007	1.45	1.39 ± 0.02
3765	1.4	1.39 ± 0.02
3687	1.42	1.39 ± 0.02
3657	1.4	1.38 ± 0.02
3503	1.29	1.37 ± 0.02
788	1.3	1.37 ± 0.04
3779	1.26	1.33 ± 0.02


Predicted S1/T1
S1/T1
External test
± SD


Predicted S1/T1
S1/T1
External test
± SD

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 14:30:46

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

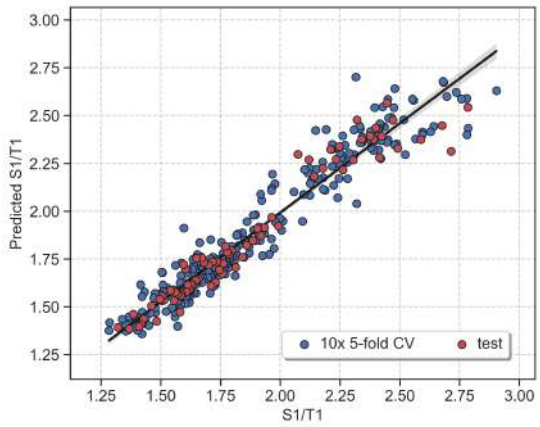
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:3



10x 5-fold CV : R^2 = 0.93, MAE = 0.072, RMSE = 0.1

Test : R^2 = 0.94, MAE = 0.063, RMSE = 0.094

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

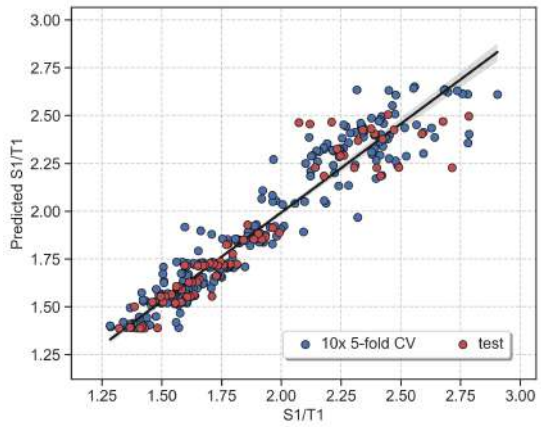
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.89, MAE = 0.081, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

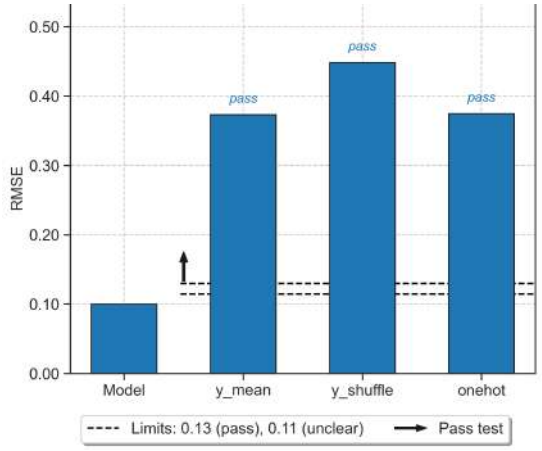
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

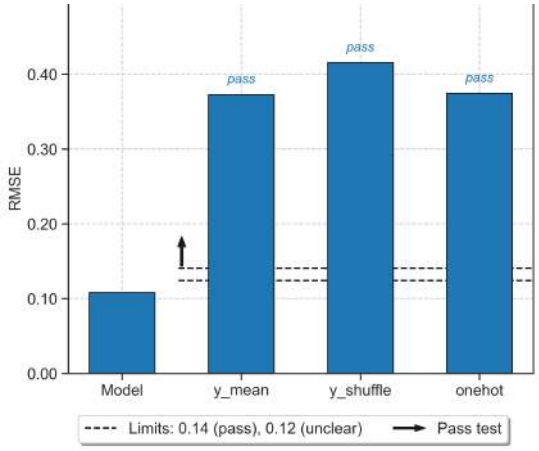


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.25%.

R^2 (10x 5-fold CV) = 0.93.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 5.88%.

R^2 (test set) = 0.94.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.94*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.89.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

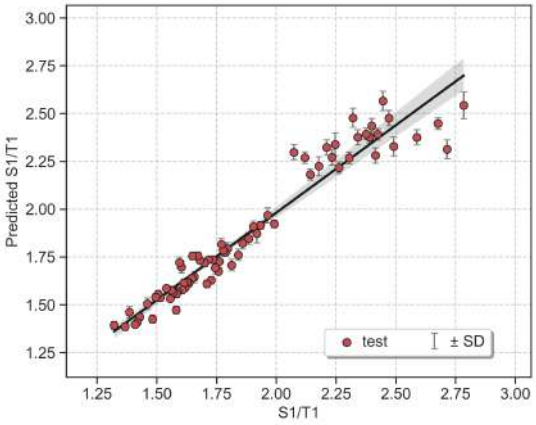
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (7% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[10.62%, 5.62%, 6.87%, 12.5%, 21.87%]

· Scoring from 0 to 2 ·

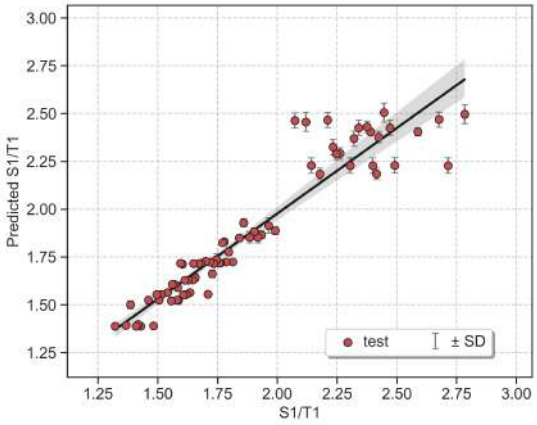
Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

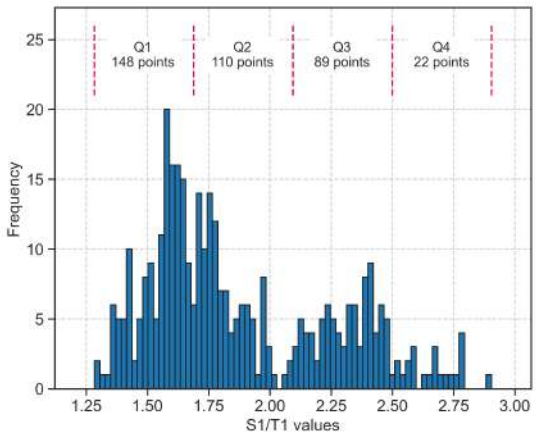
[10.0%, 6.25%, 7.5%, 14.37%, 20.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

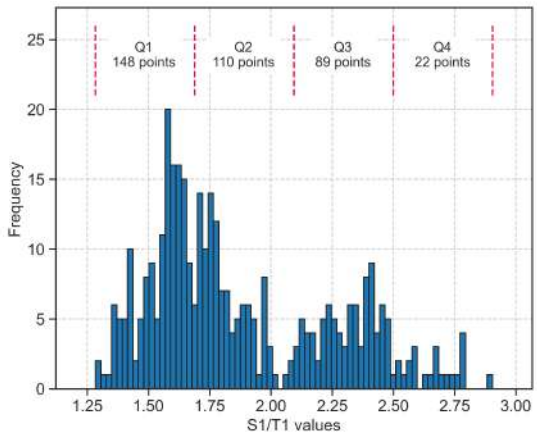
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

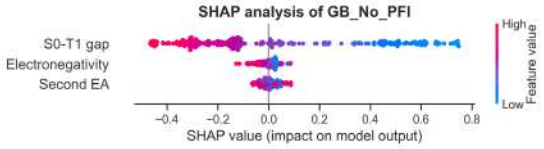


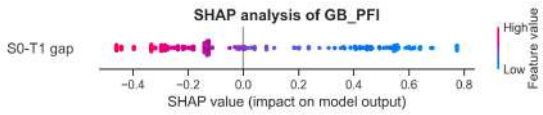
y distribution analysis

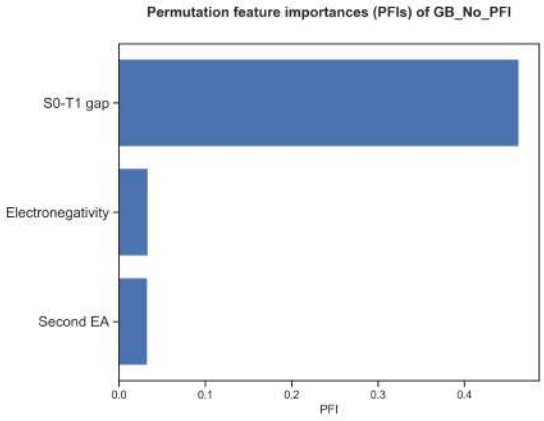
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

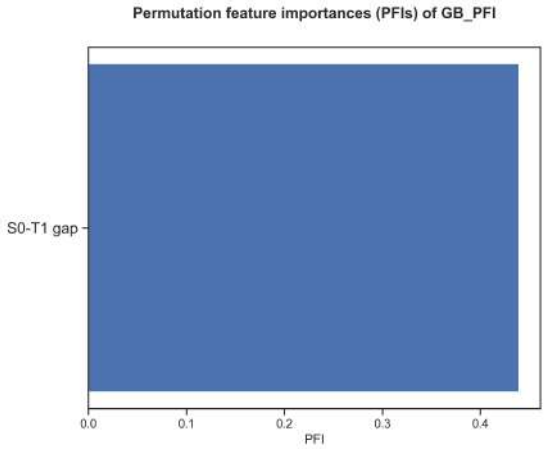
Section D. Feature Importances

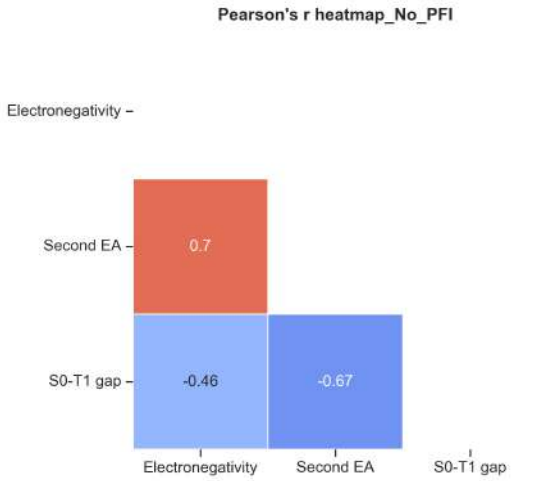
This section presents feature importances measured using the validation set.

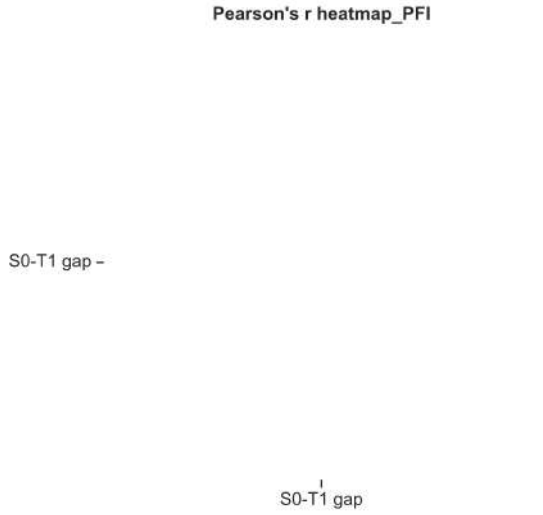












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

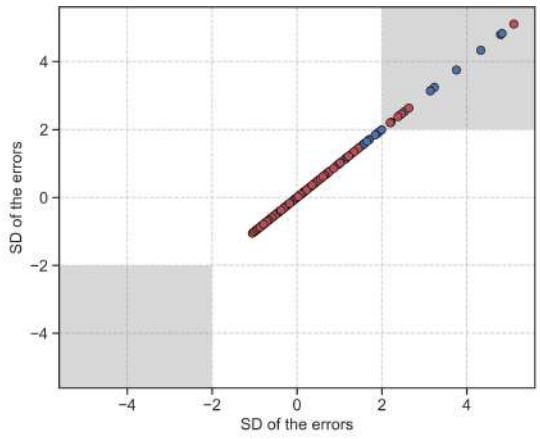
Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (3.1 SDs)
- 885 (4.3 SDs)
- 1007 (4.8 SDs)
- 1003 (2.2 SDs)
- 207 (2.4 SDs)
- 49 (3.2 SDs)
- 1006 (4.8 SDs)
- 2763 (2.5 SDs)
- 2651 (3.1 SDs)
- 345 (2.4 SDs)

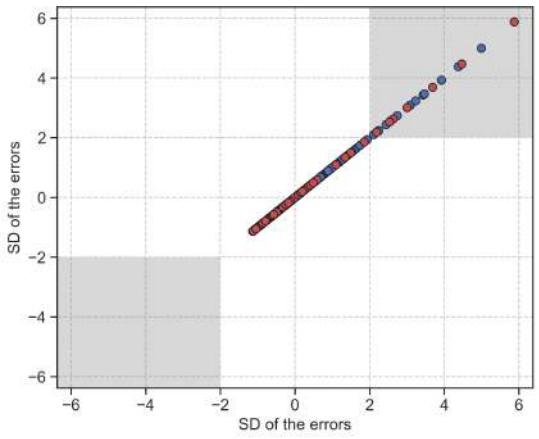
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (2.6 SDs)
- 884 (5.1 SDs)
- 890 (2.5 SDs)
- 3715 (2.2 SDs)
- 1483 (2.4 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 15 outliers out of 295 datapoints (5.1%)
- 2279 (3.1 SDs)
- 885 (4.4 SDs)
- 1007 (5.0 SDs)
- 978 (2.6 SDs)
- 1126 (3.1 SDs)
- 207 (2.2 SDs)
- 223 (2.4 SDs)
- 889 (2.2 SDs)
- 1128 (2.1 SDs)
- 49 (3.9 SDs)

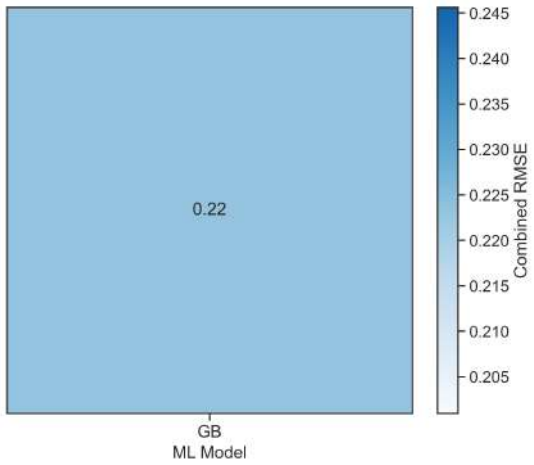
Test: 7 outliers out of 74 datapoints (9.5%)
- 888 (3.0 SDs)
- 884 (5.9 SDs)
- 872 (2.6 SDs)
- 891 (2.2 SDs)
- 3222 (2.5 SDs)
- 412 (3.7 SDs)
- 1483 (4.5 SDs)

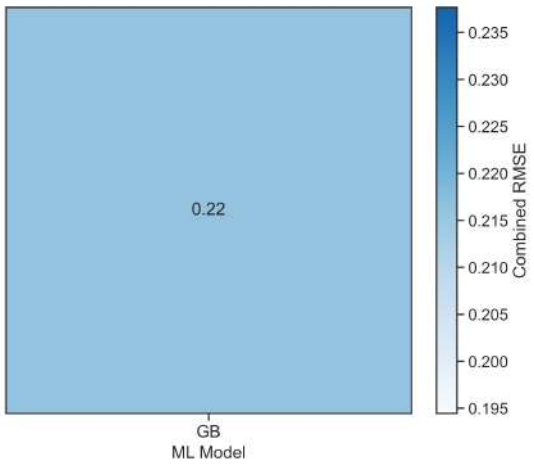




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'IP', 'LUMO', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Second IP', 'Nucleophilicity idx', 'MolLogP']" --model "['GB']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 105.42 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: GradientBoostingRegressor	sklearn model: GradientBoostingRegressor
n_estimators: 39	n_estimators: 39
learning_rate: 0.2035490101894677	learning_rate: 0.2035490101894677
max_depth: 7	max_depth: 7
min_samples_split: 8	min_samples_split: 8
min_samples_leaf: 3	min_samples_leaf: 3
subsample: 0.754957408602135	subsample: 0.754957408602135
max_features: 0.6898847011075624	max_features: 0.6898847011075624
validation_fraction: 0.10402150923749871	validation_fraction: 0.10402150923749871
min_weight_fraction_leaf: 0.04144700146086816	min_weight_fraction_leaf: 0.04144700146086816
ccp_alpha: 4.695476192547066e-05	ccp_alpha: 4.695476192547066e-05
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.9, MAE = 0.053, RMSE = 0.065

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1099	2.24	2.3 ± 0.05
1853	2.17	2.27 ± 0.04
2764	2.13	2.22 ± 0.04
997	2.12	2.21 ± 0.04
399	2.14	2.2 ± 0.04
1754	2.08	2.14 ± 0.04
1015	2.13	2.11 ± 0.06
1769	2.01	2.11 ± 0.03
1123	2.04	2.07 ± 0.06
3211	1.94	2.07 ± 0.03
...	...	...
3765	1.4	1.41 ± 0.01
3779	1.26	1.41 ± 0.02
3657	1.4	1.4 ± 0.01
3801	1.32	1.4 ± 0.03
3468	1.33	1.4 ± 0.03
3687	1.42	1.4 ± 0.02
788	1.3	1.39 ± 0.02
3815	1.36	1.39 ± 0.02
1985	1.39	1.39 ± 0.02
3503	1.29	1.35 ± 0.02

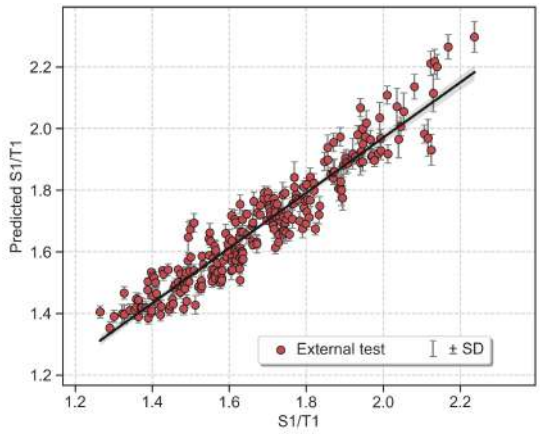
External test metrics

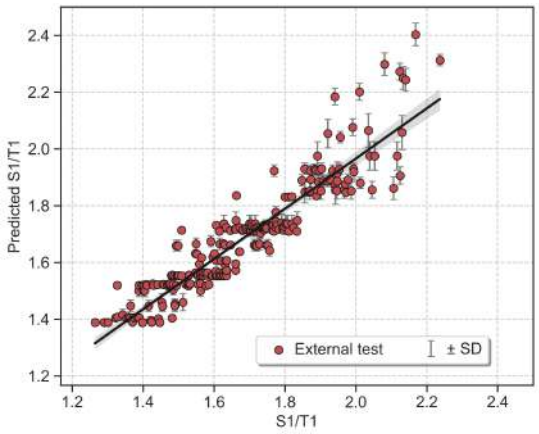
R2 = 0.86, MAE = 0.058, RMSE = 0.075

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.4 ± 0.04
1099	2.24	2.31 ± 0.02
1754	2.08	2.3 ± 0.04
997	2.12	2.27 ± 0.03
2764	2.13	2.25 ± 0.04
399	2.14	2.24 ± 0.04
1769	2.01	2.2 ± 0.03
3211	1.94	2.18 ± 0.03
2749	1.99	2.08 ± 0.03
1123	2.04	2.06 ± 0.06
...	...	...
3626	1.44	1.4 ± 0.01
3687	1.42	1.4 ± 0.01
855	1.37	1.39 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.39 ± 0.01
3765	1.4	1.39 ± 0.01
3657	1.4	1.39 ± 0.01
788	1.3	1.39 ± 0.01
3779	1.26	1.39 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2
2.4

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 14:13:13

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

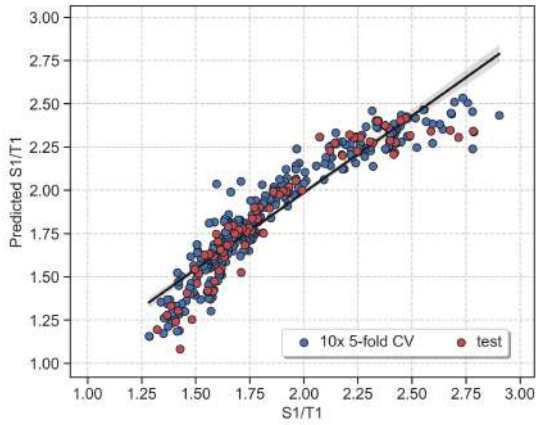
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:3



10x 5-fold CV : R^2 = 0.88, MAE = 0.097, RMSE = 0.13

Test : R^2 = 0.87, MAE = 0.1, RMSE = 0.13

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

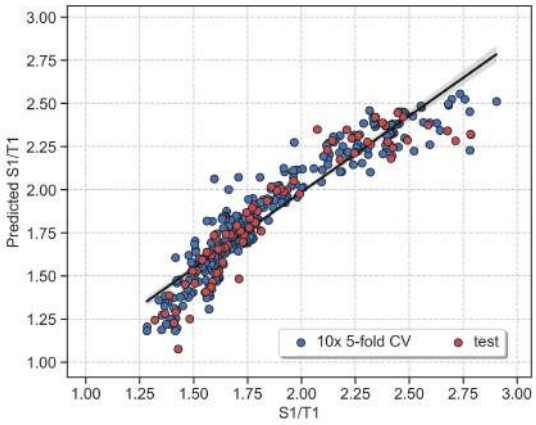
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.88, MAE = 0.096, RMSE = 0.13

Test : R^2 = 0.86, MAE = 0.1, RMSE = 0.14

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

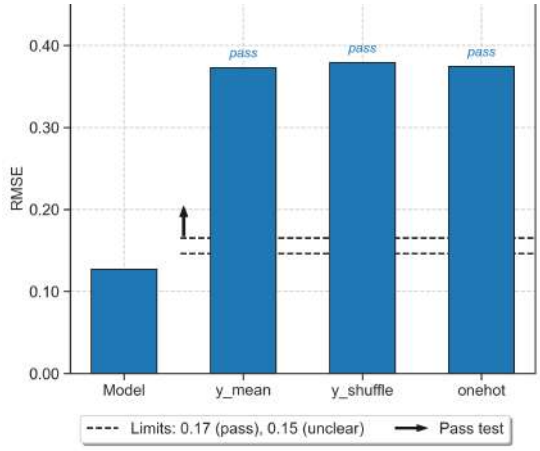
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

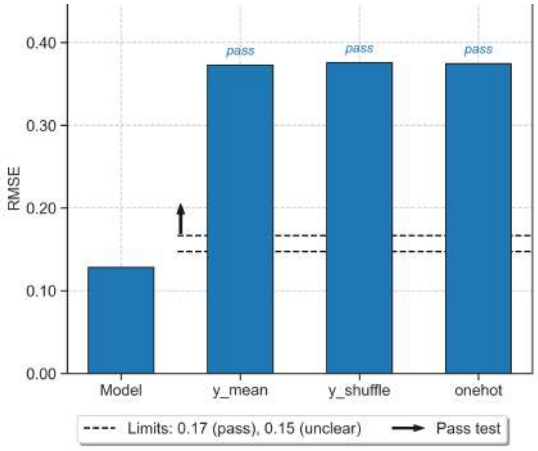


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.12%.

R^2 (test set) = 0.87.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.75%.

R^2 (test set) = 0.86.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.08*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

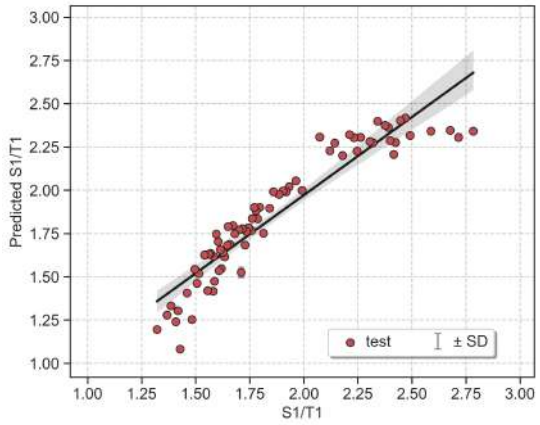
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

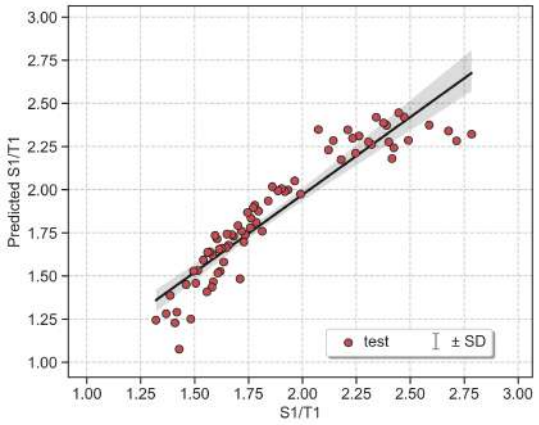


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (1% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[10.62%, 8.12%, 6.25%, 7.5%, 20.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

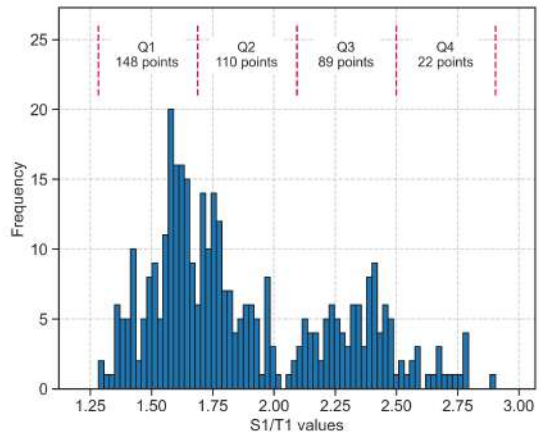
[9.37%, 8.12%, 6.25%, 6.87%, 18.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

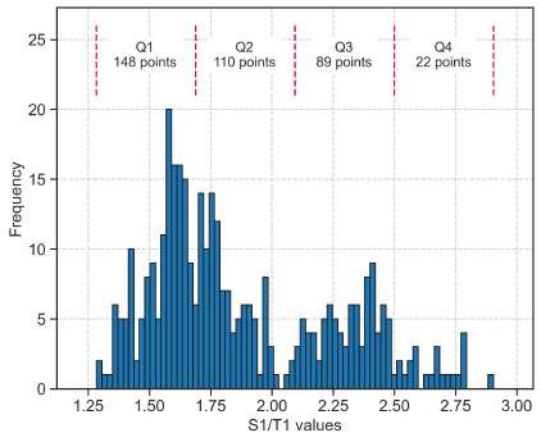
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

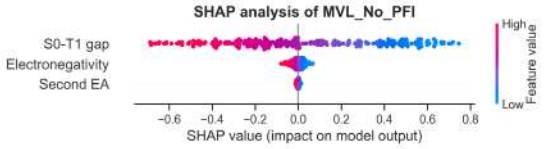


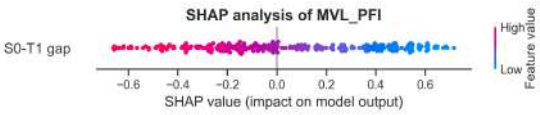
y distribution analysis

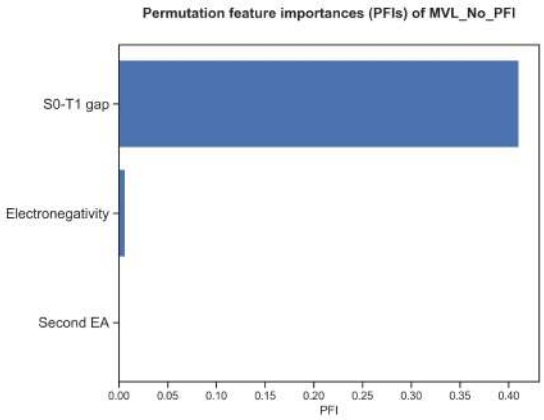
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

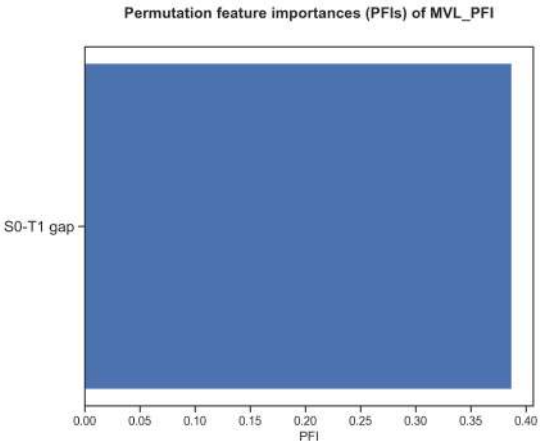
Section D. Feature Importances

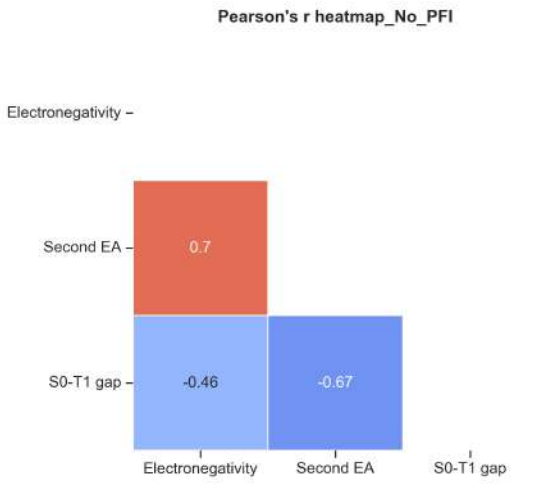
This section presents feature importances measured using the validation set.

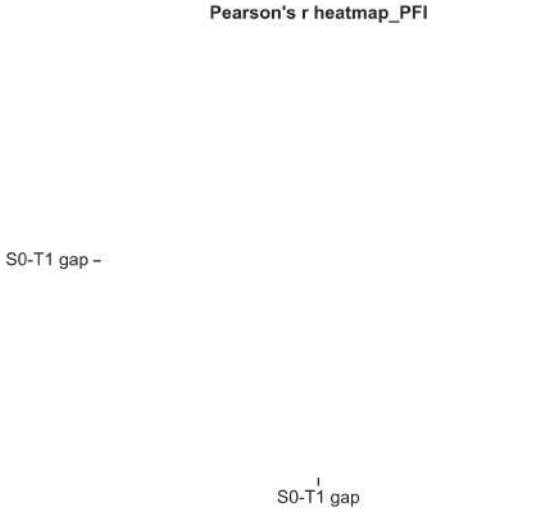












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

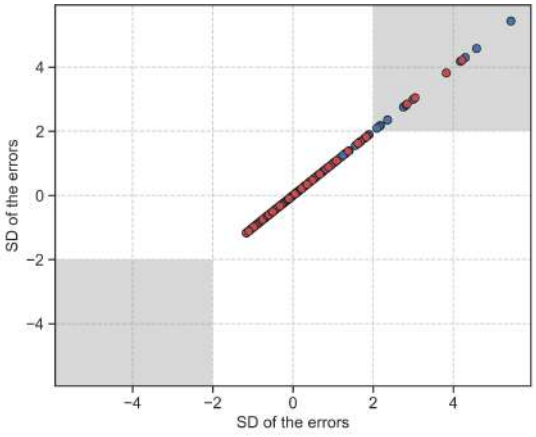
Outliers (max. 10 shown)
Train: 12 outliers out of 295 datapoints (4.1%)
- 2279 (4.6 SDs)
- 885 (4.3 SDs)
- 1007 (5.4 SDs)
- 1130 (2.8 SDs)
- 978 (2.4 SDs)
- 1126 (2.8 SDs)
- 207 (2.2 SDs)
- 345 (2.1 SDs)
- 3264 (3.0 SDs)
- 1362 (2.8 SDs)

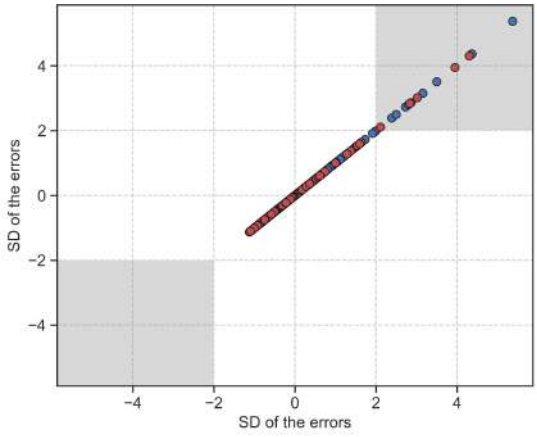
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.2 SDs)
- 884 (3.8 SDs)
- 890 (2.8 SDs)
- 46 (3.0 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (3.5 SDs)
- 885 (4.4 SDs)
- 1007 (5.4 SDs)
- 1130 (2.7 SDs)
- 978 (2.4 SDs)
- 1126 (2.8 SDs)
- 207 (2.4 SDs)
- 345 (2.5 SDs)
- 3264 (3.2 SDs)
- 1362 (2.9 SDs)

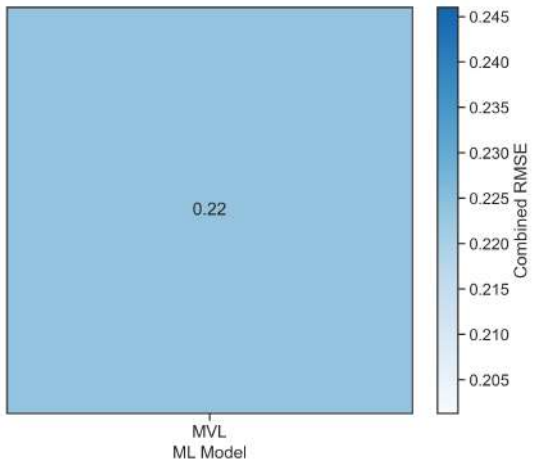
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.3 SDs)
- 884 (4.0 SDs)
- 890 (2.8 SDs)
- 1483 (2.1 SDs)
- 46 (3.0 SDs)

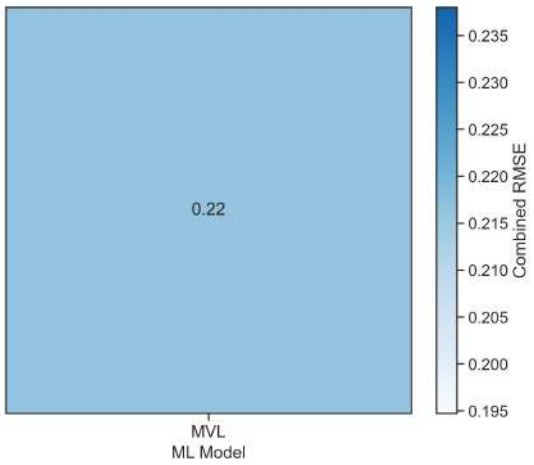




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'MolLogP']" --model "['MVL']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 16.96 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: LinearRegression	sklearn model: LinearRegression

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.9, MAE = 0.077, RMSE = 0.095

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.24 ± 0.01
1754	2.08	2.23 ± 0.01
1099	2.24	2.22 ± 0.01
399	2.14	2.21 ± 0.01
997	2.12	2.2 ± 0.01
2764	2.13	2.19 ± 0.01
1769	2.01	2.18 ± 0.01
1123	2.04	2.15 ± 0.01
3211	1.94	2.14 ± 0.01
1883	1.96	2.13 ± 0.01
...	...	...
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3765	1.4	1.25 ± 0.01
3468	1.33	1.24 ± 0.01
2007	1.45	1.24 ± 0.01
3657	1.4	1.23 ± 0.01
3503	1.29	1.23 ± 0.01
3779	1.26	1.18 ± 0.01
788	1.3	1.14 ± 0.01

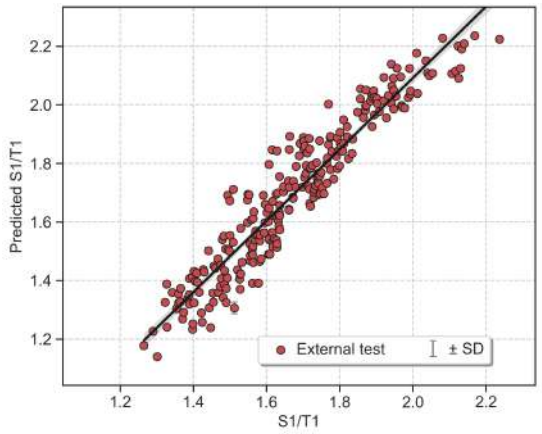
External test metrics

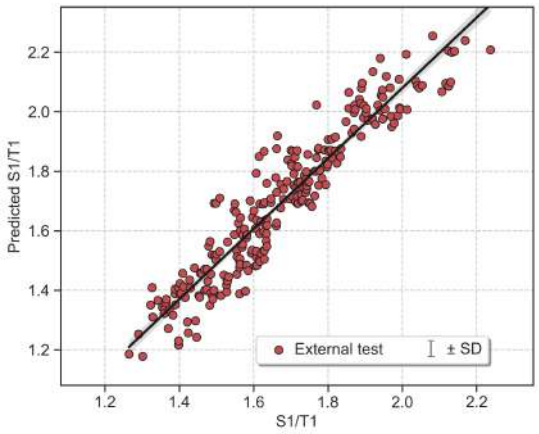
R2 = 0.89, MAE = 0.072, RMSE = 0.093

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.25 ± 0.01
1853	2.17	2.24 ± 0.01
1099	2.24	2.21 ± 0.01
997	2.12	2.21 ± 0.01
399	2.14	2.2 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.18 ± 0.01
2236	1.92	2.13 ± 0.01
1883	1.96	2.12 ± 0.01
...	...	...
3626	1.44	1.3 ± 0.01
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3503	1.29	1.25 ± 0.01
2007	1.45	1.24 ± 0.01
3765	1.4	1.23 ± 0.01
3657	1.4	1.22 ± 0.01
3779	1.26	1.19 ± 0.01
788	1.3	1.18 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 14:24:48

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

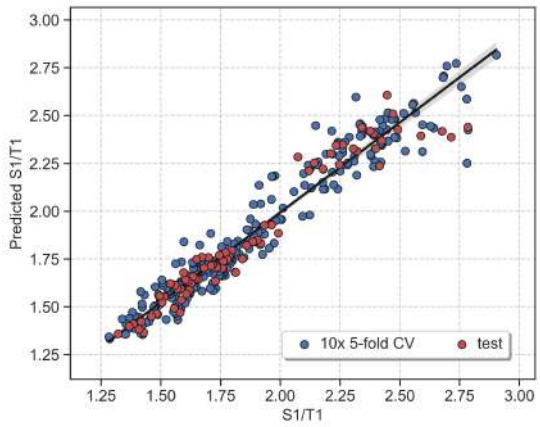
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:3



10x 5-fold CV : R^2 = 0.94, MAE = 0.067, RMSE = 0.093

Test : R^2 = 0.94, MAE = 0.064, RMSE = 0.094

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Potential "faulty" outliers (Section E)

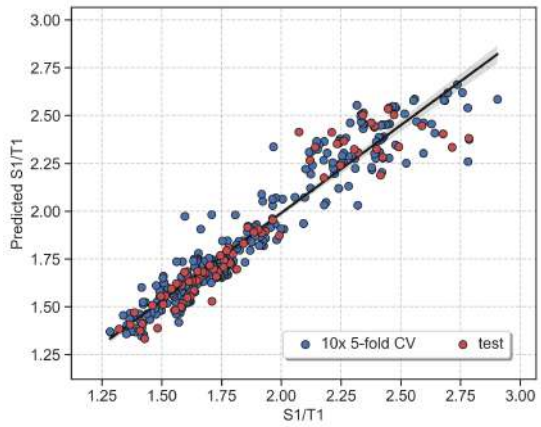
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.079, RMSE = 0.11

Test : R^2 = 0.91, MAE = 0.077, RMSE = 0.11

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

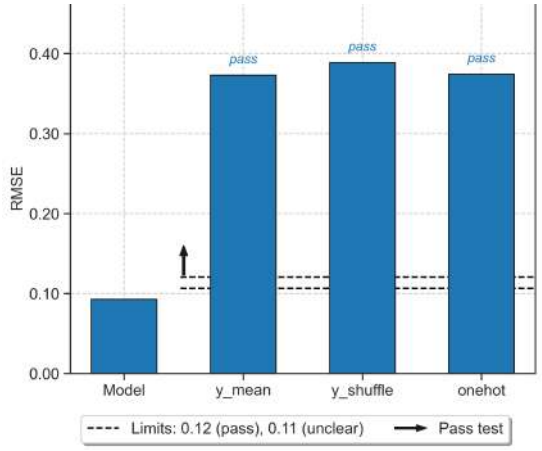
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

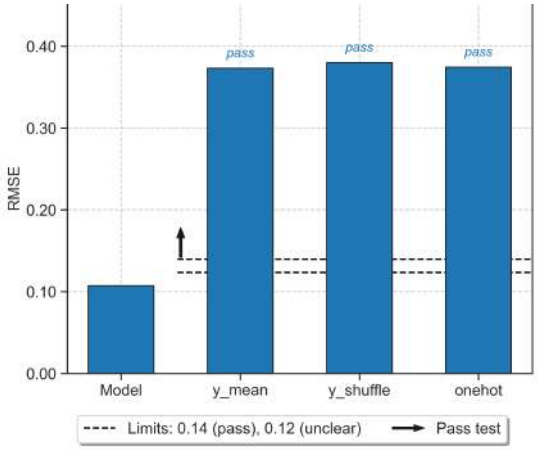


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 5.81%.

R^2 (10x 5-fold CV) = 0.94.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 5.88%.

R^2 (test set) = 0.94.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.01*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.87%.

R^2 (test set) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

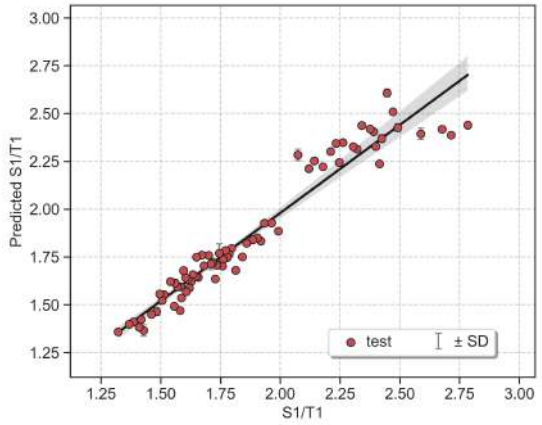
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (3% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[13.75%, 4.38%, 5.62%, 10.0%, 12.5%]

· Scoring from 0 to 2 ·

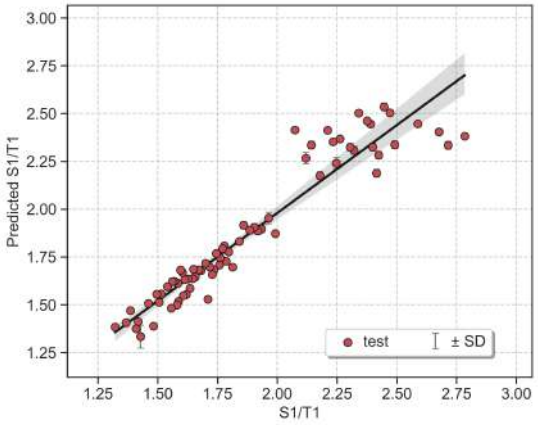
Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

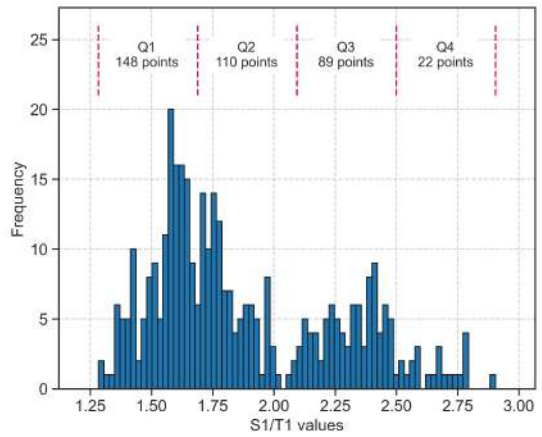
[8.12%, 6.25%, 7.5%, 11.88%, 20.62%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

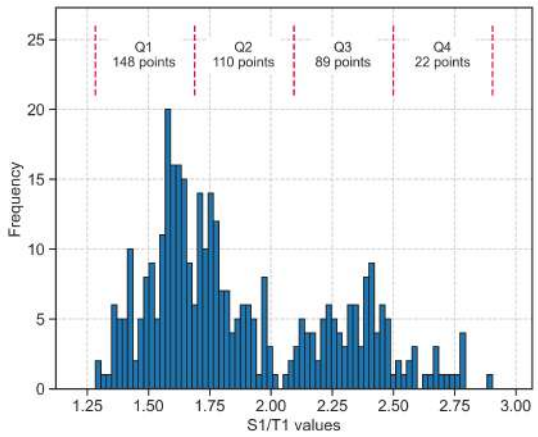
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

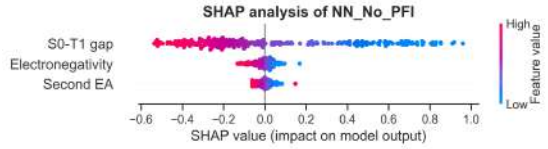


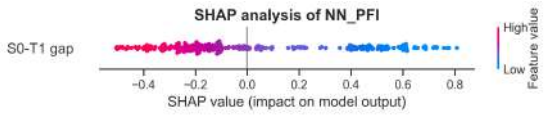
y distribution analysis

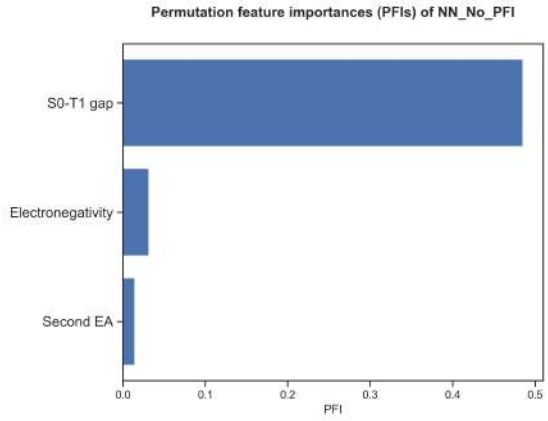
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

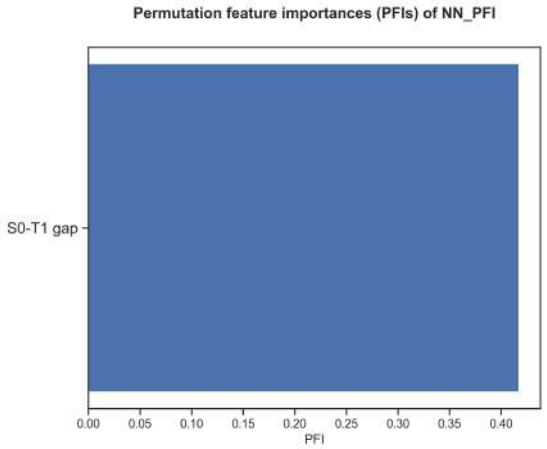
Section D. Feature Importances

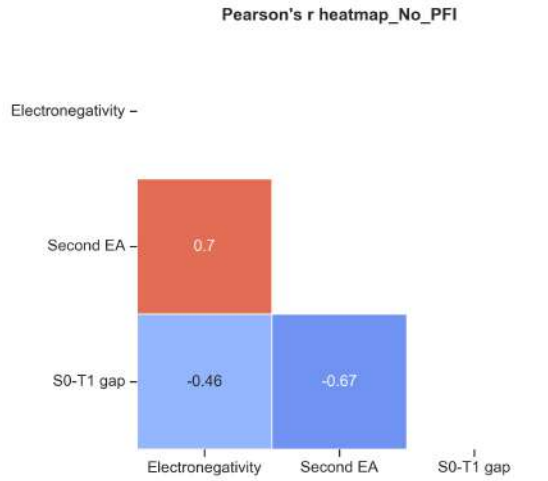
This section presents feature importances measured using the validation set.

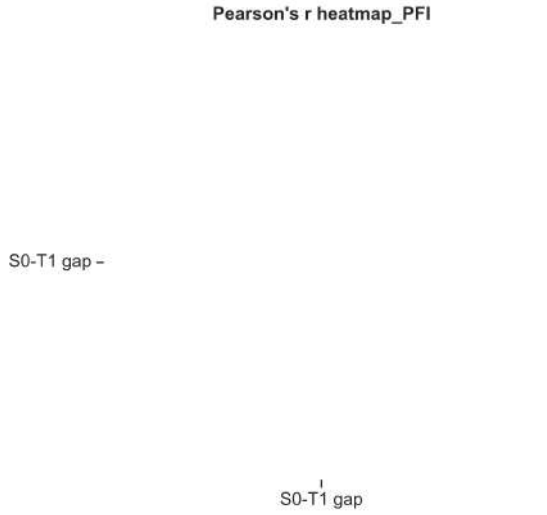












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

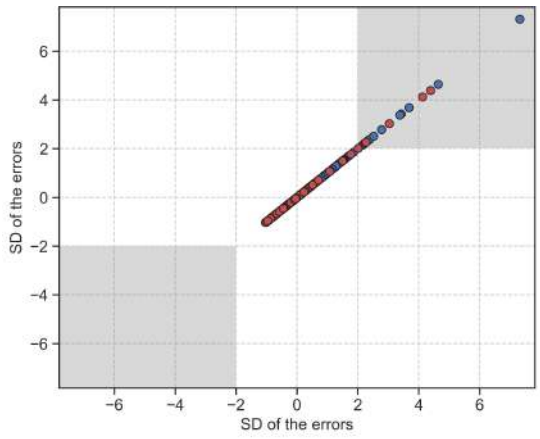
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 14 outliers out of 295 datapoints (4.7%)
- 885 (4.6 SDs)
- 1007 (7.3 SDs)
- 1130 (2.0 SDs)
- 978 (2.2 SDs)
- 1126 (3.4 SDs)
- 207 (2.2 SDs)
- 1006 (3.4 SDs)
- 1157 (2.2 SDs)
- 2651 (3.7 SDs)
- 2649 (2.2 SDs)

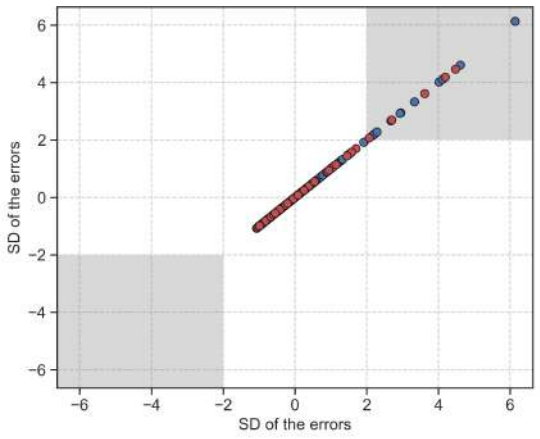
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.4 SDs)
- 884 (4.1 SDs)
- 890 (3.0 SDs)
- 3715 (2.0 SDs)
- 1483 (2.3 SDs)



PFI (only important descriptors):

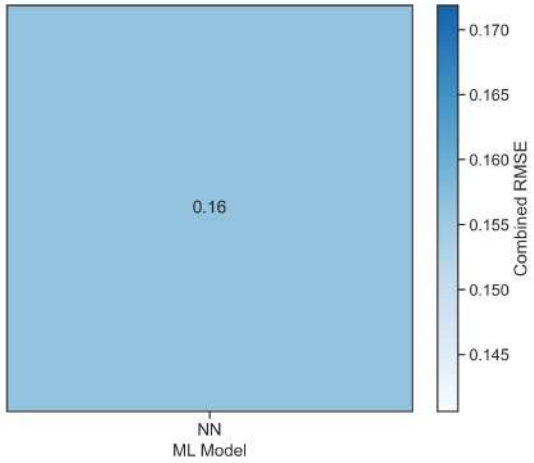
Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (3.3 SDs)
- 885 (4.6 SDs)
- 1007 (6.1 SDs)
- 1130 (2.3 SDs)
- 978 (2.2 SDs)
- 1126 (2.9 SDs)
- 207 (2.7 SDs)
- 49 (2.9 SDs)
- 1006 (2.2 SDs)
- 345 (4.0 SDs)

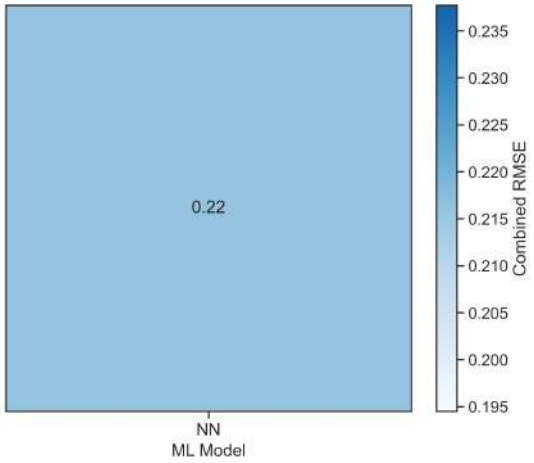
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.5 SDs)
- 884 (4.2 SDs)
- 890 (2.7 SDs)
- 891 (2.1 SDs)
- 1483 (3.6 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'Dipole module', 'Total charge', 'EA', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total FOD', 'Total polariz. alpha', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'MolLogP']" --model "['NN']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 181.21 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: MLPRegressor	sklearn model: MLPRegressor
hidden_layer_1: 4	hidden_layer_1: 4
hidden_layer_2: 4	hidden_layer_2: 4
max_iter: 409	max_iter: 409
alpha: 0.015420292446634285	alpha: 0.015420292446634285
tol: 7.00090043901101e-05	tol: 7.00090043901101e-05
random_state: 0	random_state: 0
solver: lbfgs	solver: lbfgs

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.91, MAE = 0.051, RMSE = 0.061

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1099	2.24	2.24 ± 0.01
1853	2.17	2.21 ± 0.01
1123	2.04	2.2 ± 0.02
399	2.14	2.2 ± 0.01
997	2.12	2.16 ± 0.01
1754	2.08	2.15 ± 0.01
2764	2.13	2.13 ± 0.01
892	2.11	2.11 ± 0.01
1015	2.13	2.1 ± 0.01
20	2.05	2.09 ± 0.02
...	...	...
3687	1.42	1.4 ± 0.01
3765	1.4	1.4 ± 0.01
2007	1.45	1.4 ± 0.01
3474	1.35	1.39 ± 0.01
3516	1.37	1.39 ± 0.01
3657	1.4	1.39 ± 0.01
3503	1.29	1.37 ± 0.01
3468	1.33	1.36 ± 0.02
3779	1.26	1.34 ± 0.01
788	1.3	1.33 ± 0.02

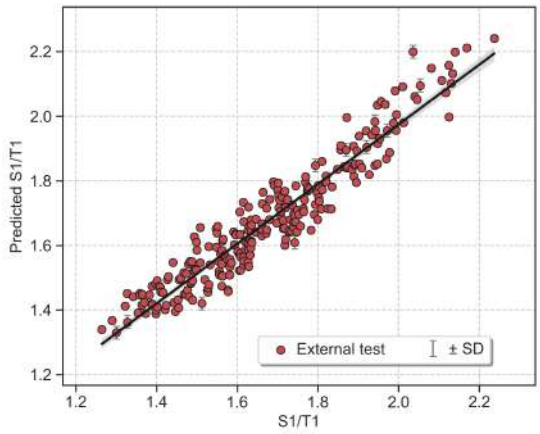
External test metrics

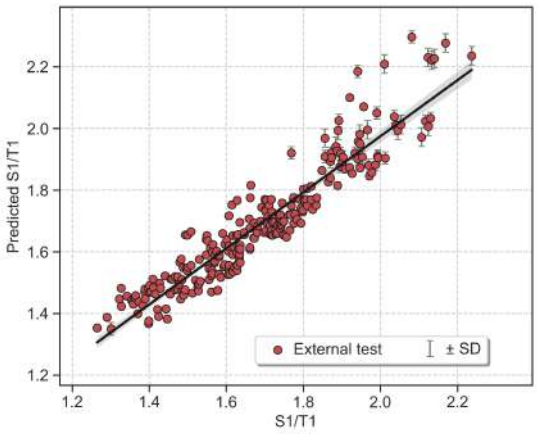
R2 = 0.88, MAE = 0.057, RMSE = 0.07

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.3 ± 0.02
1853	2.17	2.28 ± 0.03
1099	2.24	2.24 ± 0.03
997	2.12	2.23 ± 0.03
399	2.14	2.23 ± 0.03
2764	2.13	2.22 ± 0.03
1769	2.01	2.21 ± 0.03
3211	1.94	2.18 ± 0.02
2236	1.92	2.1 ± 0.01
1883	1.96	2.07 ± 0.01
...	...	...
3626	1.44	1.41 ± 0.01
3687	1.42	1.41 ± 0.01
855	1.37	1.4 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.38 ± 0.01
3765	1.4	1.37 ± 0.01
3657	1.4	1.37 ± 0.01
3779	1.26	1.35 ± 0.01
788	1.3	1.35 ± 0.02


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 14:19:42

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

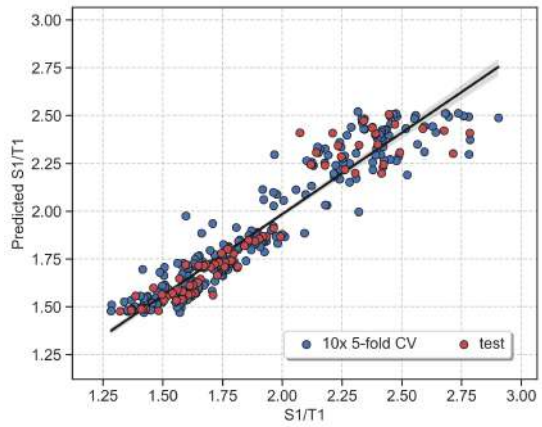
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:3



10x 5-fold CV : R^2 = 0.91, MAE = 0.086, RMSE = 0.12

Test : R^2 = 0.9, MAE = 0.082, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

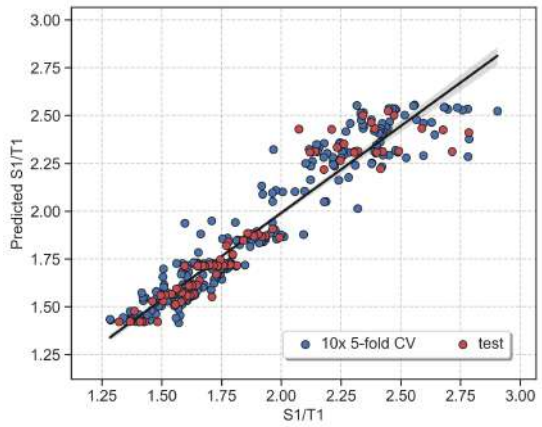
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.91, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.91, MAE = 0.077, RMSE = 0.11

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

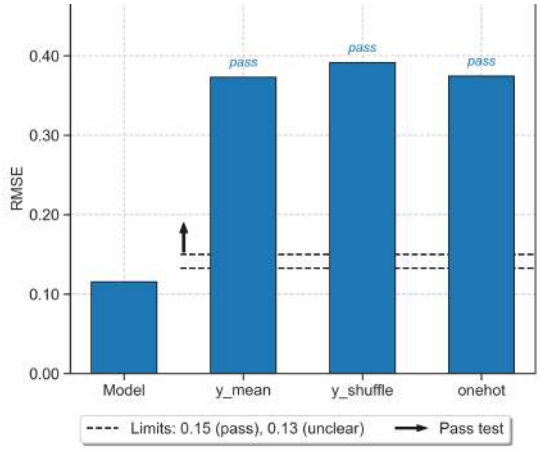
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

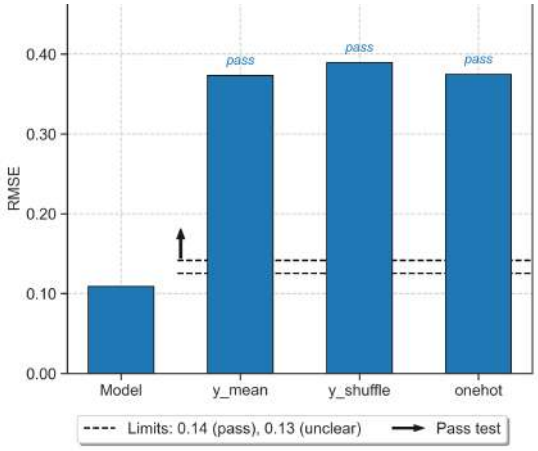


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 7.5%.

R^2 (10x 5-fold CV) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.9.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.87%.

R^2 (test set) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

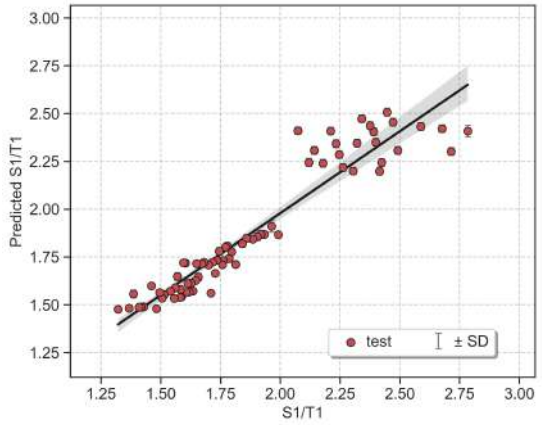
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (4% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

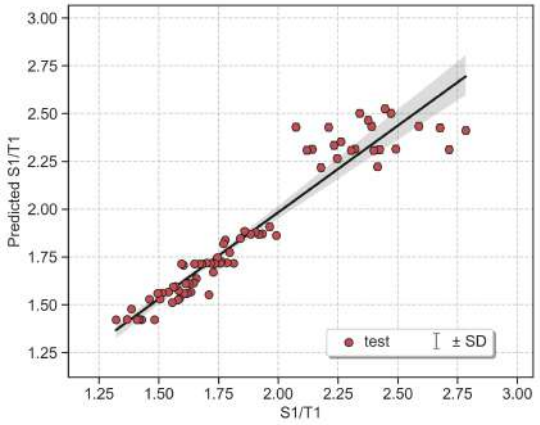


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (3% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[13.12%, 6.87%, 7.5%, 10.0%, 23.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

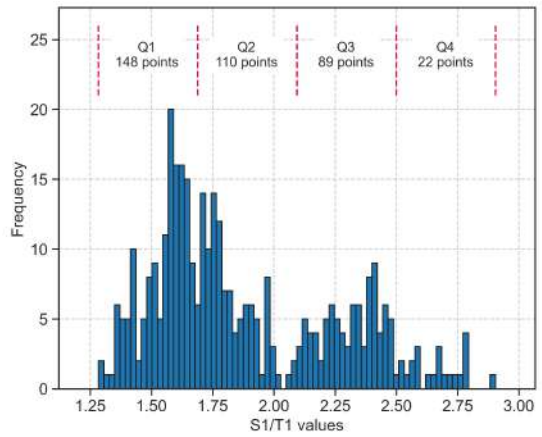
[11.25%, 6.87%, 8.75%, 13.12%, 20.62%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

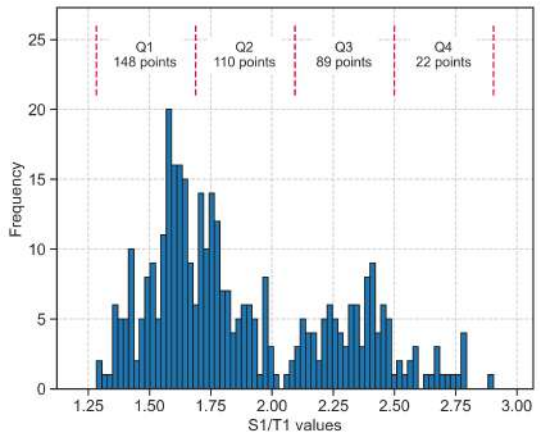
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

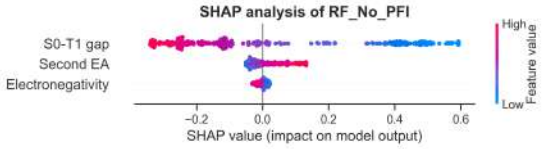


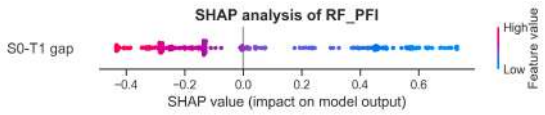
y distribution analysis

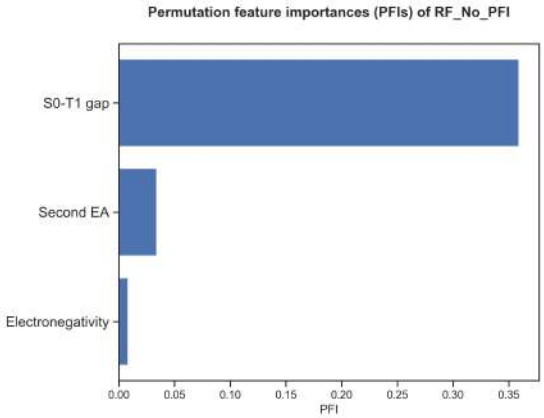
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

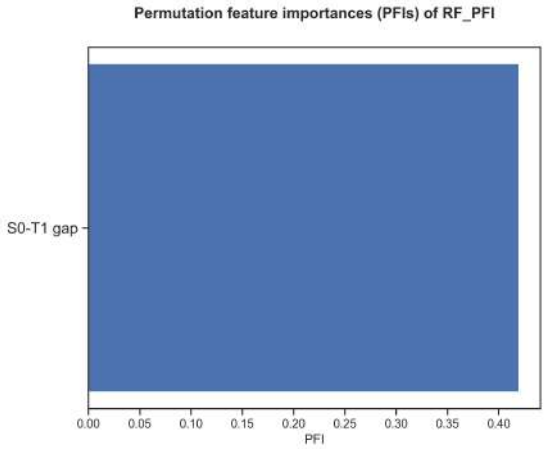
Section D. Feature Importances

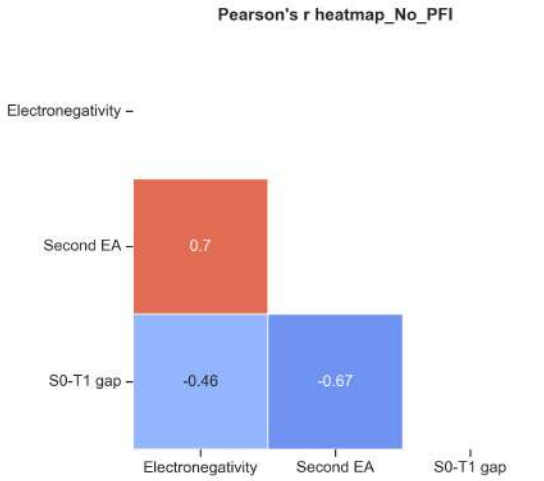
This section presents feature importances measured using the validation set.

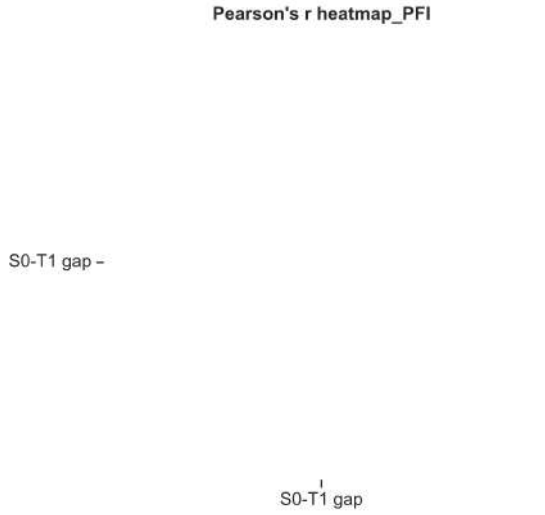












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

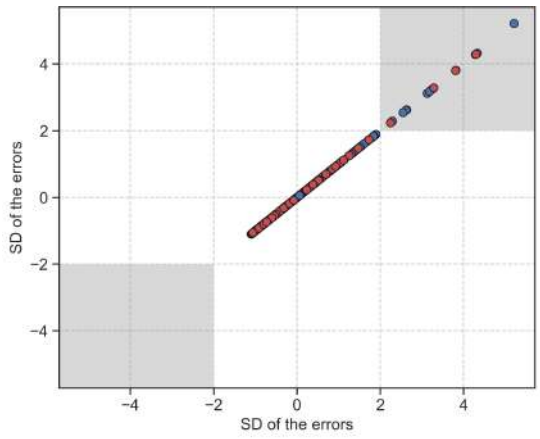
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (4.3 SDs)
- 885 (4.3 SDs)
- 1007 (5.2 SDs)
- 1130 (2.6 SDs)
- 2767 (2.3 SDs)
- 1126 (2.6 SDs)
- 207 (3.3 SDs)
- 49 (3.1 SDs)
- 345 (3.2 SDs)
- 2195 (3.8 SDs)

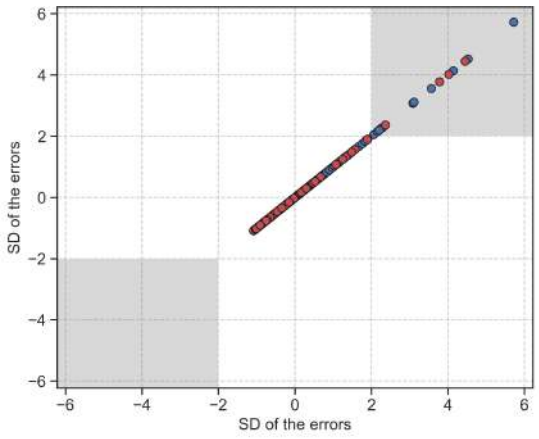
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (3.8 SDs)
- 884 (4.3 SDs)
- 890 (2.2 SDs)
- 1483 (3.3 SDs)



PFI (only important descriptors):

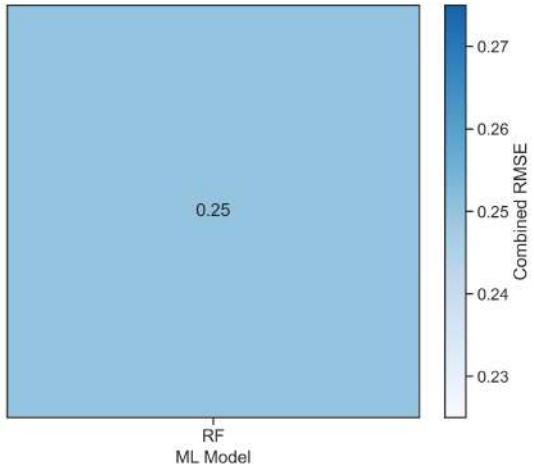
Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (4.1 SDs)
- 885 (4.5 SDs)
- 1007 (5.7 SDs)
- 1130 (2.3 SDs)
- 2767 (2.1 SDs)
- 978 (2.2 SDs)
- 1126 (3.1 SDs)
- 207 (2.3 SDs)
- 49 (3.1 SDs)
- 1006 (2.2 SDs)

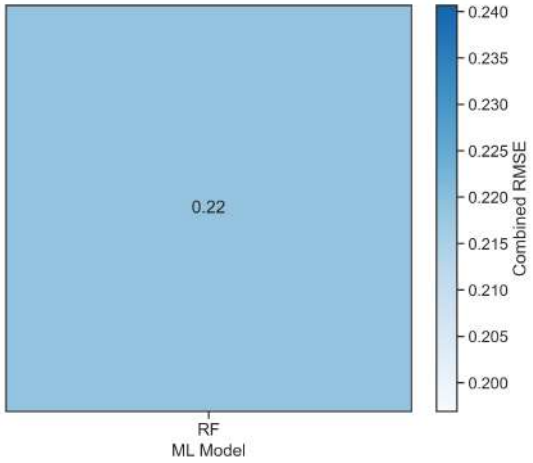
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.0 SDs)
- 884 (4.4 SDs)
- 890 (2.4 SDs)
- 1483 (3.8 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'IP', 'LUMO', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Total FOD', 'Fermi-level', 'Total polariz. alpha', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'MolLogP']" --model "['RF']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 207.26 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: RandomForestRegressor	sklearn model: RandomForestRegressor
n_estimators: 97	n_estimators: 97
max_depth: 11	max_depth: 11
min_samples_split: 8	min_samples_split: 8
min_samples_leaf: 4	min_samples_leaf: 4
min_weight_fraction_leaf: 0.028402228054696617	min_weight_fraction_leaf: 0.028402228054696617
max_features: 0.9441974787194958	max_features: 0.9441974787194958
ccp_alpha: 0.0007103605819788694	ccp_alpha: 0.0007103605819788694
max_samples: 0.31534697477615553	max_samples: 0.31534697477615553
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.84, MAE = 0.065, RMSE = 0.083

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.31 ± 0.02
1754	2.08	2.29 ± 0.02
1099	2.24	2.27 ± 0.02
997	2.12	2.24 ± 0.02
3211	1.94	2.24 ± 0.02
399	2.14	2.24 ± 0.02
2764	2.13	2.23 ± 0.02
1769	2.01	2.21 ± 0.02
1883	1.96	2.1 ± 0.03
2236	1.92	2.1 ± 0.03
...	...	...
3789	1.36	1.48 ± 0.01
3799	1.34	1.48 ± 0.01
3687	1.42	1.48 ± 0.01
1985	1.39	1.48 ± 0.01
3779	1.26	1.48 ± 0.01
2007	1.45	1.48 ± 0.01
2396	1.48	1.48 ± 0.01
855	1.37	1.48 ± 0.01
788	1.3	1.48 ± 0.01
3503	1.29	1.47 ± 0.01

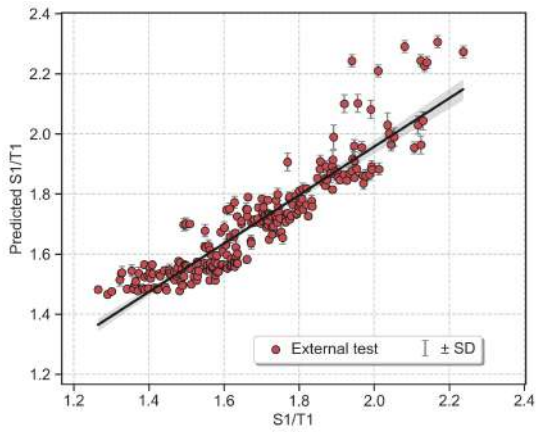
External test metrics

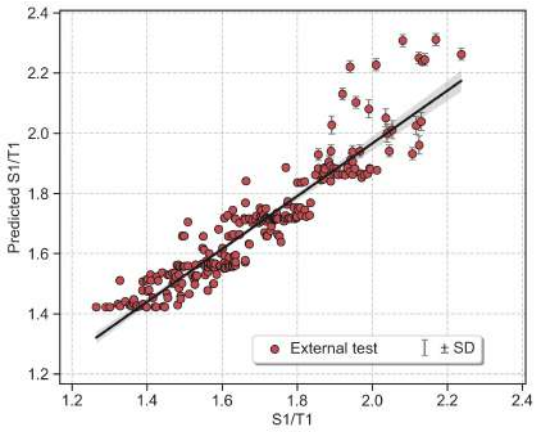
R2 = 0.86, MAE = 0.058, RMSE = 0.074

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.31 ± 0.02
1754	2.08	2.31 ± 0.02
1099	2.24	2.26 ± 0.02
997	2.12	2.25 ± 0.02
399	2.14	2.24 ± 0.02
2764	2.13	2.24 ± 0.01
1769	2.01	2.23 ± 0.02
3211	1.94	2.22 ± 0.02
2236	1.92	2.13 ± 0.02
1883	1.96	2.1 ± 0.02
...	...	...
3626	1.44	1.42 ± 0.01
3687	1.42	1.42 ± 0.01
855	1.37	1.42 ± 0.01
2007	1.45	1.42 ± 0.01
816	1.42	1.42 ± 0.01
3765	1.4	1.42 ± 0.01
3657	1.4	1.42 ± 0.01
788	1.3	1.42 ± 0.01
3503	1.29	1.42 ± 0.01
3779	1.26	1.42 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD


Predicted S1/T1
S1/T1
External test
± SD

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/07 10:59:15

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

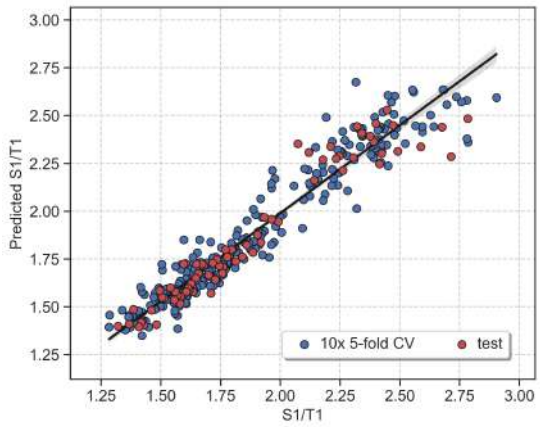
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:2



10x 5-fold CV : R^2 = 0.92, MAE = 0.077, RMSE = 0.1

Test : R^2 = 0.92, MAE = 0.072, RMSE = 0.1

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

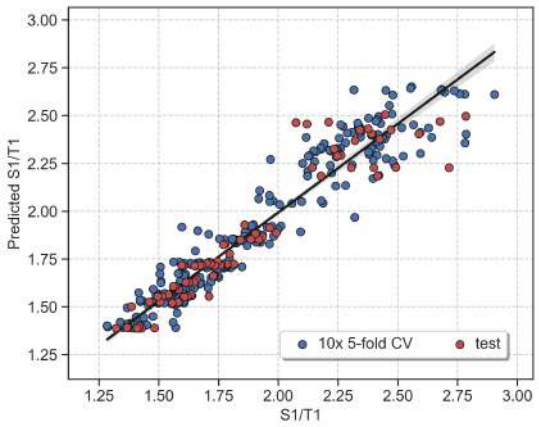
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.89, MAE = 0.081, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

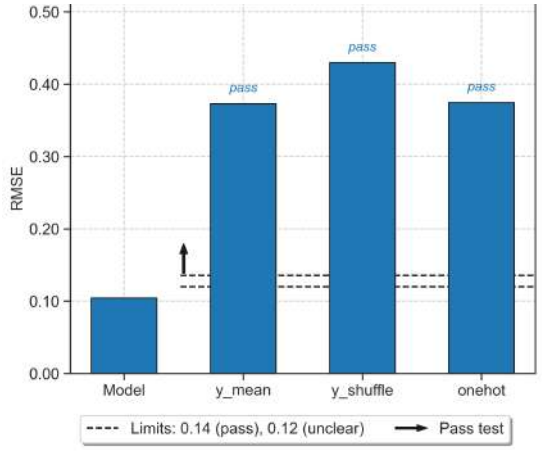
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

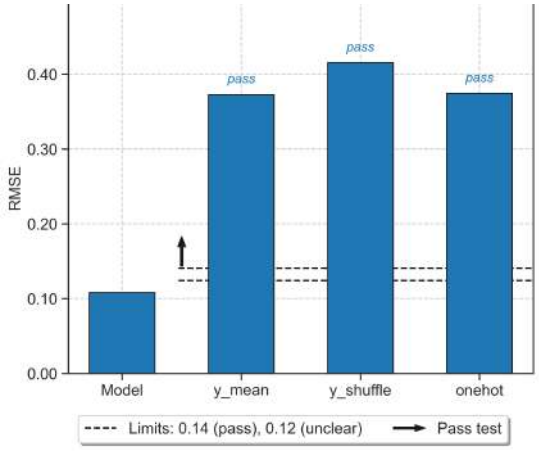


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.25%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.25%.

R^2 (test set) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.89.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

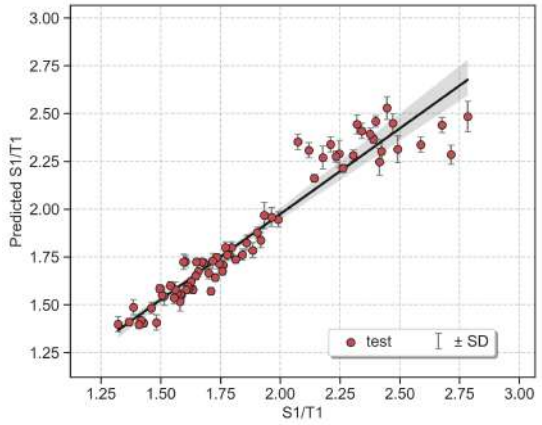
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (8% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

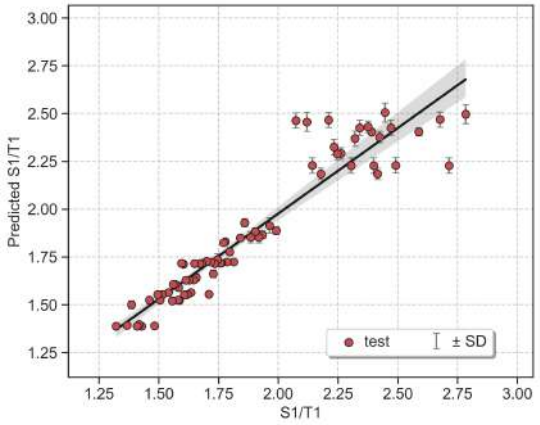


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[11.25%, 5.62%, 6.87%, 13.12%, 21.87%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

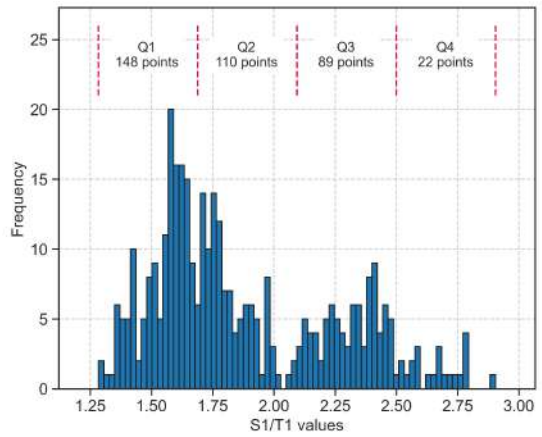
[10.0%, 6.25%, 7.5%, 14.37%, 20.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

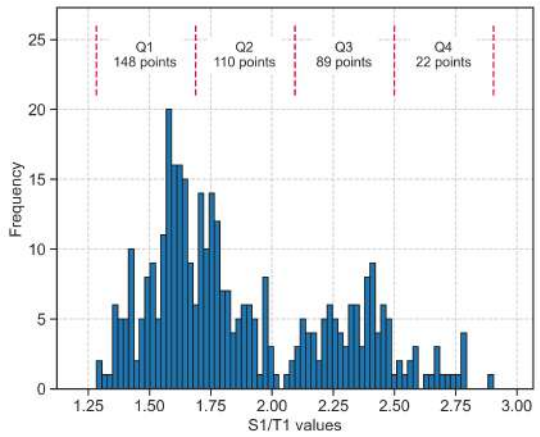
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

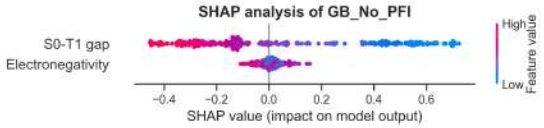


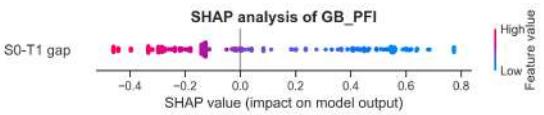
y distribution analysis

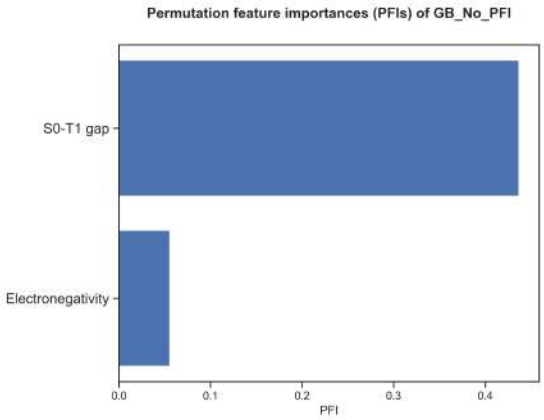
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

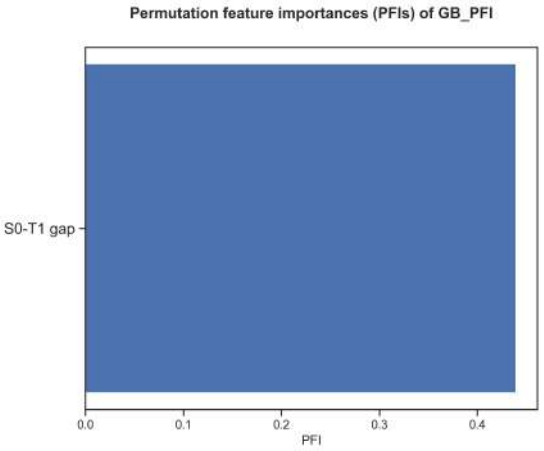
Section D. Feature Importances

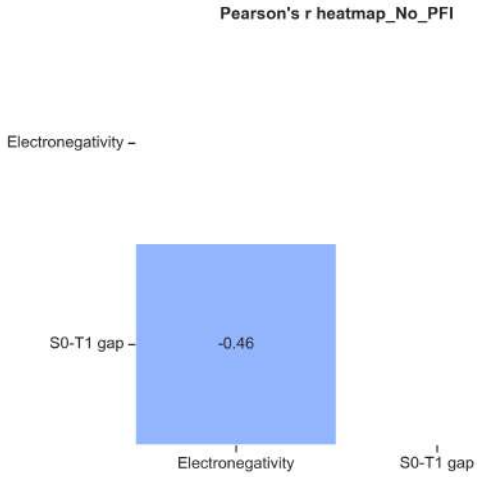
This section presents feature importances measured using the validation set.

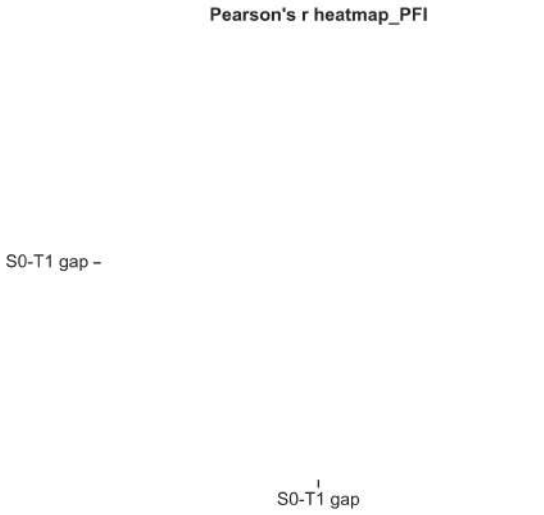












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

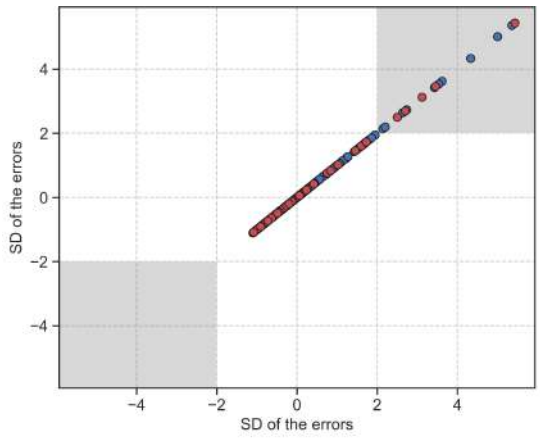
Outliers (max. 10 shown)
Train: 10 outliers out of 295 datapoints (3.4%)
- 2279 (3.6 SDs)
- 885 (5.4 SDs)
- 1007 (5.0 SDs)
- 883 (2.1 SDs)
- 49 (3.5 SDs)
- 1006 (4.3 SDs)
- 2763 (3.4 SDs)
- 345 (2.6 SDs)
- 2195 (2.7 SDs)
- 2041 (2.2 SDs)

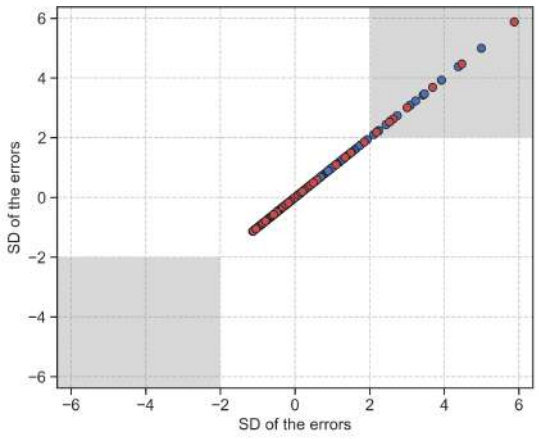
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (3.5 SDs)
- 884 (5.4 SDs)
- 890 (2.5 SDs)
- 3715 (2.7 SDs)
- 1483 (3.1 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 15 outliers out of 295 datapoints (5.1%)
- 2279 (3.1 SDs)
- 885 (4.4 SDs)
- 1007 (5.0 SDs)
- 978 (2.6 SDs)
- 1126 (3.1 SDs)
- 207 (2.2 SDs)
- 223 (2.4 SDs)
- 889 (2.2 SDs)
- 1128 (2.1 SDs)
- 49 (3.9 SDs)

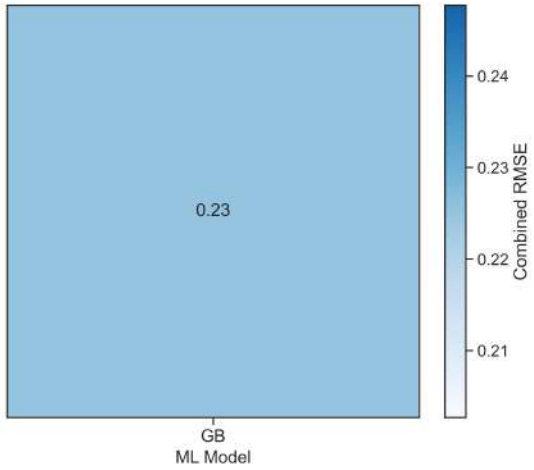
Test: 7 outliers out of 74 datapoints (9.5%)
- 888 (3.0 SDs)
- 884 (5.9 SDs)
- 872 (2.6 SDs)
- 891 (2.2 SDs)
- 3222 (2.5 SDs)
- 412 (3.7 SDs)
- 1483 (4.5 SDs)

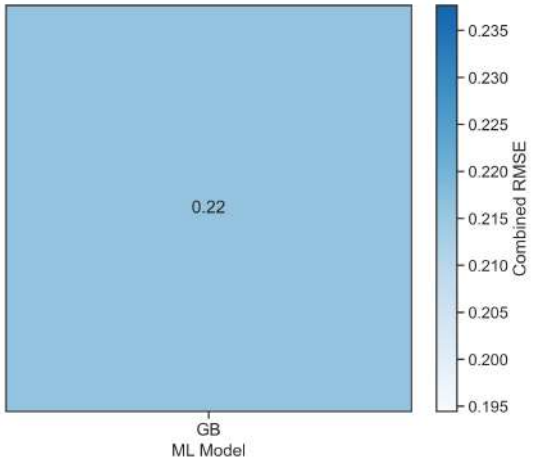




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Total charge', 'Dipole module', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Softness', 'Hardness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP']" --model "['GB']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 92.03 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: GradientBoostingRegressor	sklearn model: GradientBoostingRegressor
n_estimators: 39	n_estimators: 39
learning_rate: 0.2035490101894677	learning_rate: 0.2035490101894677
max_depth: 7	max_depth: 7
min_samples_split: 8	min_samples_split: 8
min_samples_leaf: 3	min_samples_leaf: 3
subsample: 0.754957408602135	subsample: 0.754957408602135
max_features: 0.6898847011075624	max_features: 0.6898847011075624
validation_fraction: 0.10402150923749871	validation_fraction: 0.10402150923749871
min_weight_fraction_leaf: 0.04144700146086816	min_weight_fraction_leaf: 0.04144700146086816
ccp_alpha: 4.695476192547066e-05	ccp_alpha: 4.695476192547066e-05
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.89, MAE = 0.053, RMSE = 0.066

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1099	2.24	2.29 ± 0.07
1853	2.17	2.25 ± 0.03
399	2.14	2.24 ± 0.04
997	2.12	2.21 ± 0.03
2764	2.13	2.2 ± 0.03
1015	2.13	2.14 ± 0.08
1754	2.08	2.14 ± 0.05
2749	1.99	2.13 ± 0.04
1769	2.01	2.09 ± 0.04
3211	1.94	2.08 ± 0.04
...	...	...
3779	1.26	1.42 ± 0.03
2396	1.48	1.42 ± 0.04
833	1.42	1.42 ± 0.04
816	1.42	1.42 ± 0.03
3516	1.37	1.41 ± 0.05
3474	1.35	1.4 ± 0.03
3815	1.36	1.39 ± 0.02
1985	1.39	1.39 ± 0.03
788	1.3	1.39 ± 0.03
3503	1.29	1.38 ± 0.05

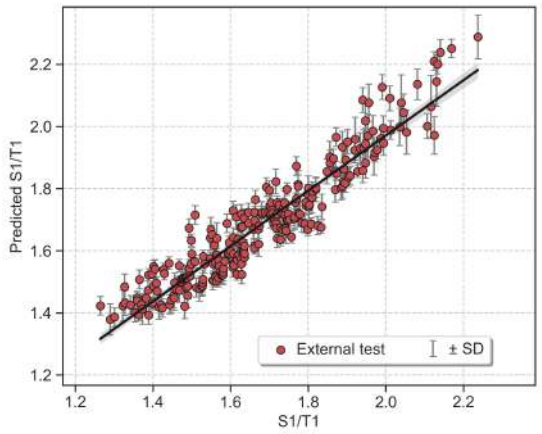
External test metrics

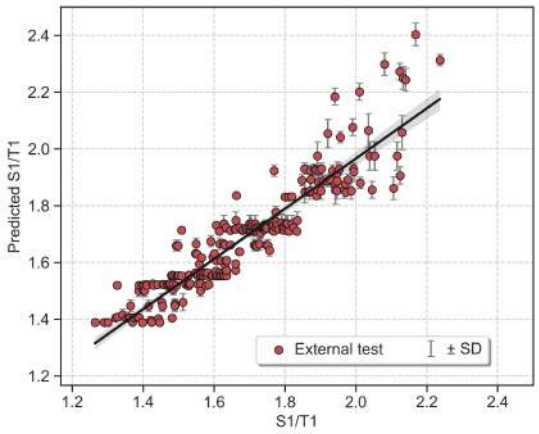
R2 = 0.86, MAE = 0.058, RMSE = 0.075

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.4 ± 0.04
1099	2.24	2.31 ± 0.02
1754	2.08	2.3 ± 0.04
997	2.12	2.27 ± 0.03
2764	2.13	2.25 ± 0.04
399	2.14	2.24 ± 0.04
1769	2.01	2.2 ± 0.03
3211	1.94	2.18 ± 0.03
2749	1.99	2.08 ± 0.03
1123	2.04	2.06 ± 0.06
...	...	...
3626	1.44	1.4 ± 0.01
3687	1.42	1.4 ± 0.01
855	1.37	1.39 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.39 ± 0.01
3765	1.4	1.39 ± 0.01
3657	1.4	1.39 ± 0.01
788	1.3	1.39 ± 0.01
3779	1.26	1.39 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2
2.4

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/07 11:03:39

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

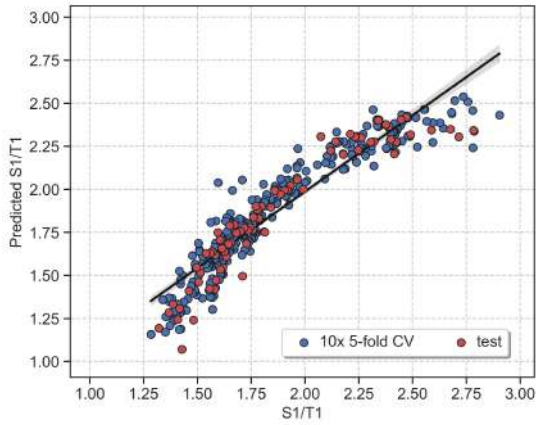
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:2



10x 5-fold CV : R^2 = 0.88, MAE = 0.097, RMSE = 0.13

Test : R^2 = 0.87, MAE = 0.1, RMSE = 0.13

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

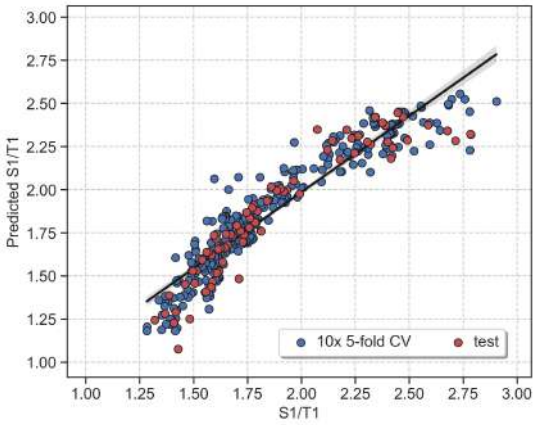
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.88, MAE = 0.096, RMSE = 0.13

Test : R^2 = 0.86, MAE = 0.1, RMSE = 0.14

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

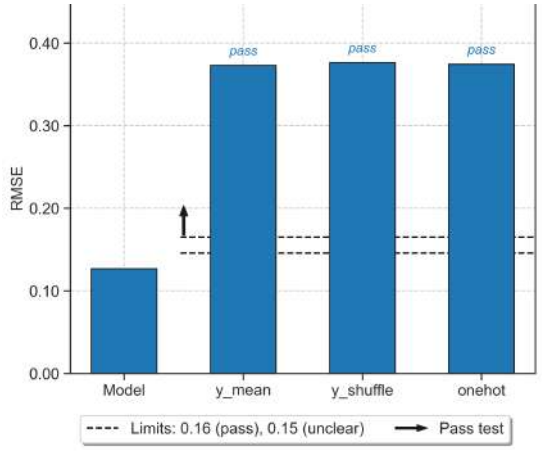
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

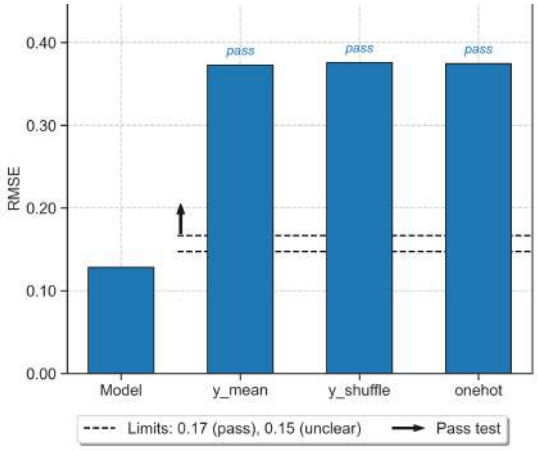


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.12%.

R^2 (test set) = 0.87.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.75%.

R^2 (test set) = 0.86.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.08*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

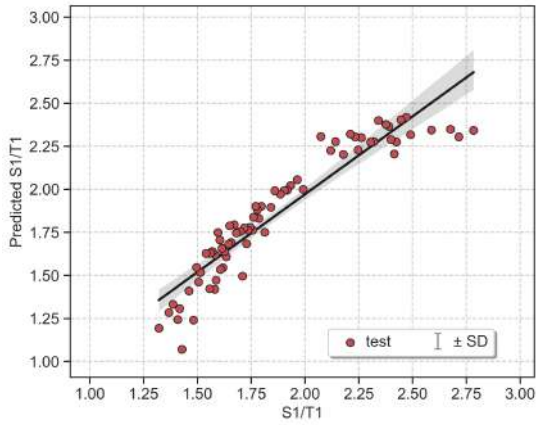
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

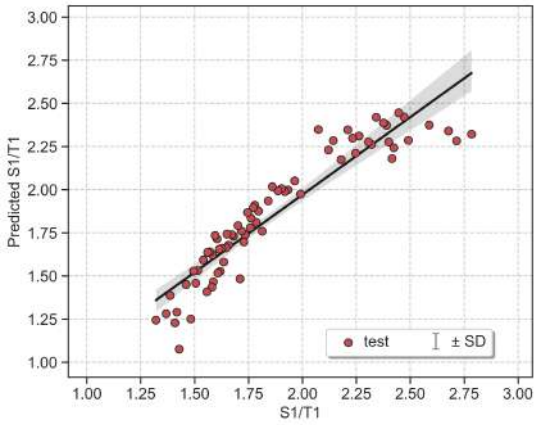


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (1% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[10.62%, 8.12%, 6.25%, 7.5%, 19.37%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

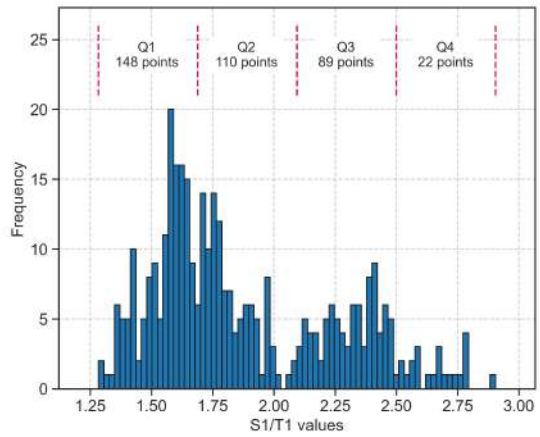
[9.37%, 8.12%, 6.25%, 6.87%, 18.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

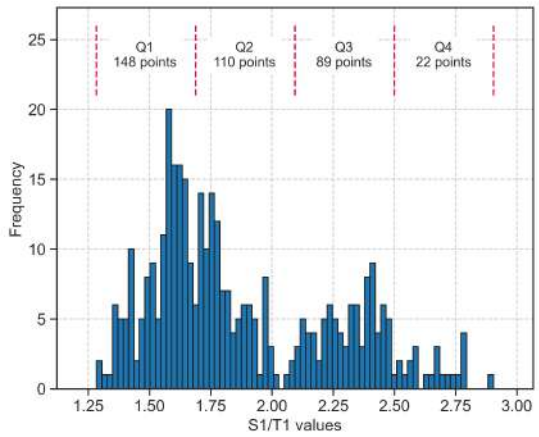
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

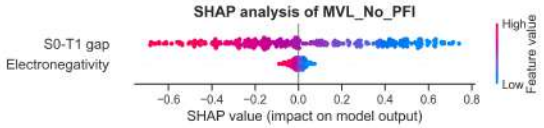


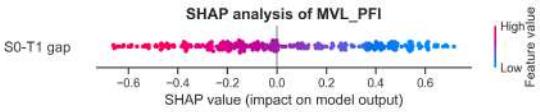
y distribution analysis

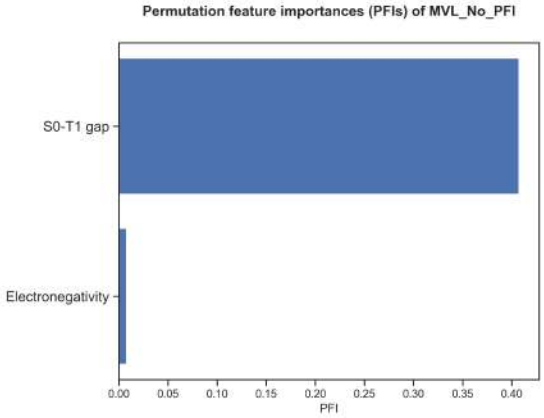
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

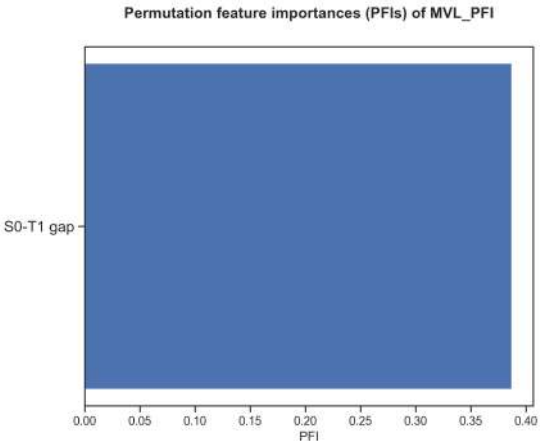
Section D. Feature Importances

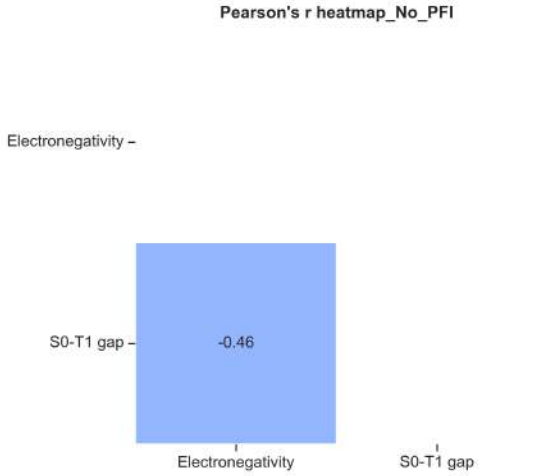
This section presents feature importances measured using the validation set.

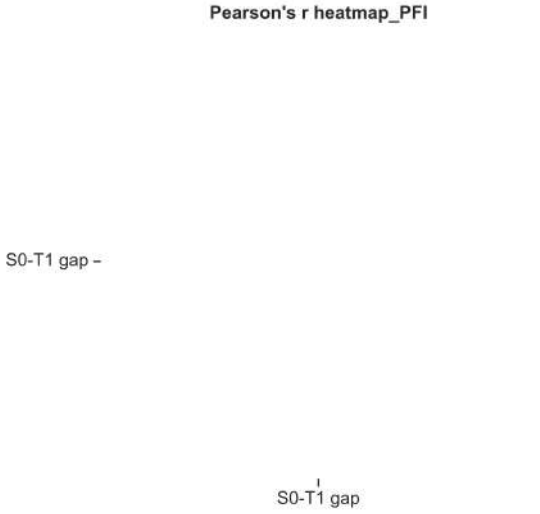












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

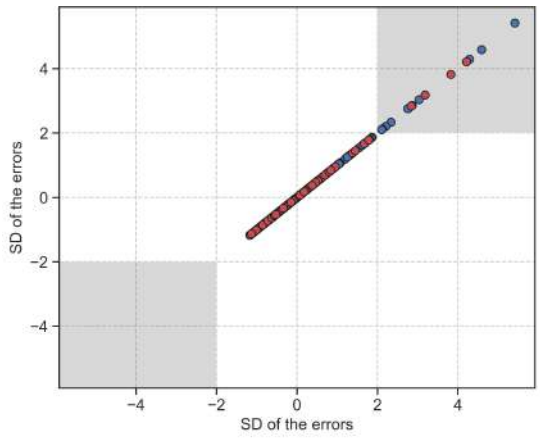
Outliers (max. 10 shown)
Train: 12 outliers out of 295 datapoints (4.1%)
- 2279 (4.6 SDs)
- 885 (4.3 SDs)
- 1007 (5.4 SDs)
- 1130 (2.8 SDs)
- 978 (2.3 SDs)
- 1126 (2.8 SDs)
- 207 (2.2 SDs)
- 345 (2.1 SDs)
- 3264 (3.0 SDs)
- 1362 (2.9 SDs)

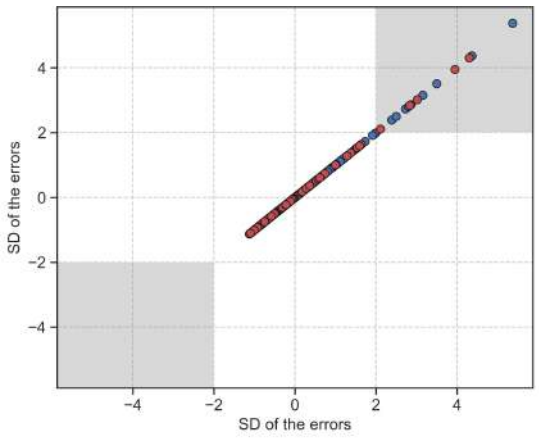
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.2 SDs)
- 884 (3.8 SDs)
- 890 (2.8 SDs)
- 46 (3.2 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (3.5 SDs)
- 885 (4.4 SDs)
- 1007 (5.4 SDs)
- 1130 (2.7 SDs)
- 978 (2.4 SDs)
- 1126 (2.8 SDs)
- 207 (2.4 SDs)
- 345 (2.5 SDs)
- 3264 (3.2 SDs)
- 1362 (2.9 SDs)

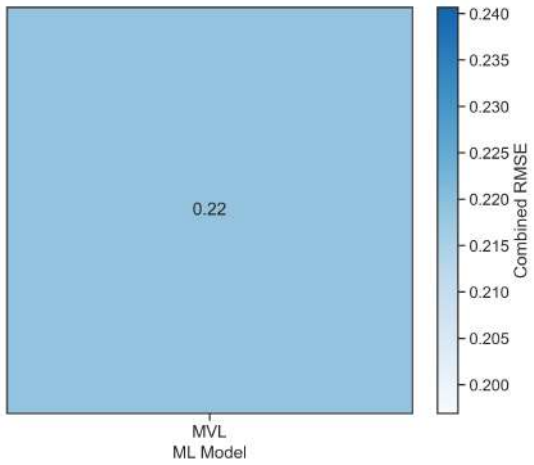
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.3 SDs)
- 884 (4.0 SDs)
- 890 (2.8 SDs)
- 1483 (2.1 SDs)
- 46 (3.0 SDs)

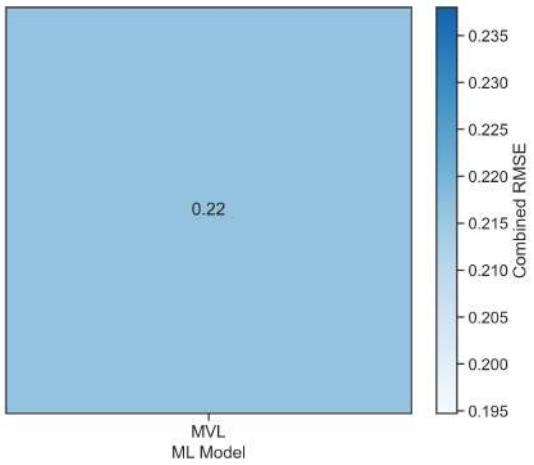




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'LUMO', 'HOMO', 'IP', 'Dipole module', 'EA', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP']" --model "['MVL']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 16.05 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: LinearRegression	sklearn model: LinearRegression

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.9, MAE = 0.077, RMSE = 0.096

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.24 ± 0.01
1754	2.08	2.23 ± 0.01
1099	2.24	2.23 ± 0.01
399	2.14	2.21 ± 0.01
997	2.12	2.2 ± 0.01
2764	2.13	2.19 ± 0.01
1769	2.01	2.18 ± 0.01
1123	2.04	2.15 ± 0.01
3211	1.94	2.14 ± 0.01
1883	1.96	2.13 ± 0.01
...	...	...
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3765	1.4	1.25 ± 0.01
3468	1.33	1.24 ± 0.01
2007	1.45	1.24 ± 0.01
3657	1.4	1.24 ± 0.01
3503	1.29	1.23 ± 0.01
3779	1.26	1.18 ± 0.01
788	1.3	1.14 ± 0.01

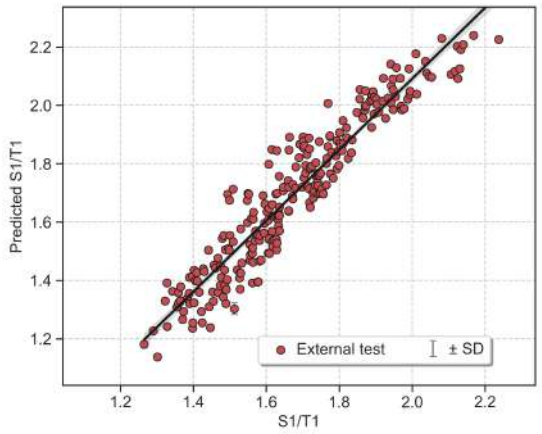
External test metrics

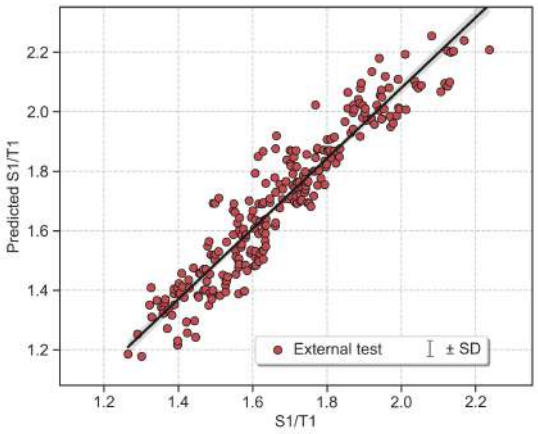
R2 = 0.89, MAE = 0.072, RMSE = 0.093

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.25 ± 0.01
1853	2.17	2.24 ± 0.01
1099	2.24	2.21 ± 0.01
997	2.12	2.21 ± 0.01
399	2.14	2.2 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.18 ± 0.01
2236	1.92	2.13 ± 0.01
1883	1.96	2.12 ± 0.01
...	...	...
3626	1.44	1.3 ± 0.01
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3503	1.29	1.25 ± 0.01
2007	1.45	1.24 ± 0.01
3765	1.4	1.23 ± 0.01
3657	1.4	1.22 ± 0.01
3779	1.26	1.19 ± 0.01
788	1.3	1.18 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/07 11:09:08

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

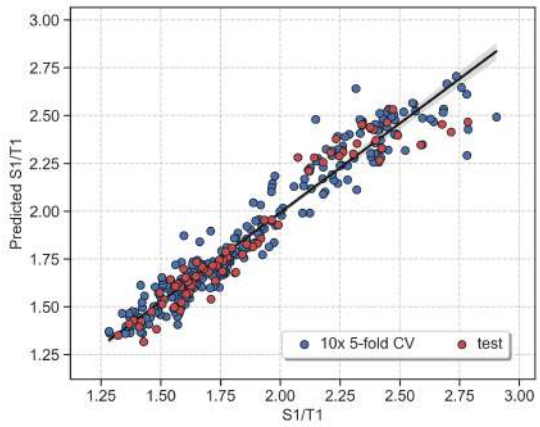
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 9

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:2



10x 5-fold CV : R^2 = 0.93, MAE = 0.072, RMSE = 0.098

Test : R^2 = 0.94, MAE = 0.067, RMSE = 0.094

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

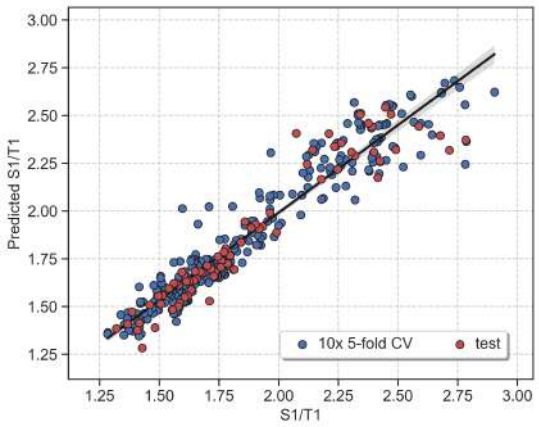
Overall assessment

The model seems reliable

PFI (only important descriptors) · Score 9

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.078, RMSE = 0.11

Test : R^2 = 0.9, MAE = 0.079, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

The model seems reliable

Section B. Advanced Score Analysis

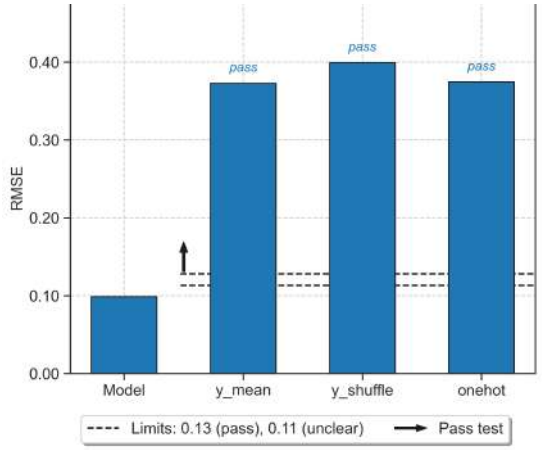
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

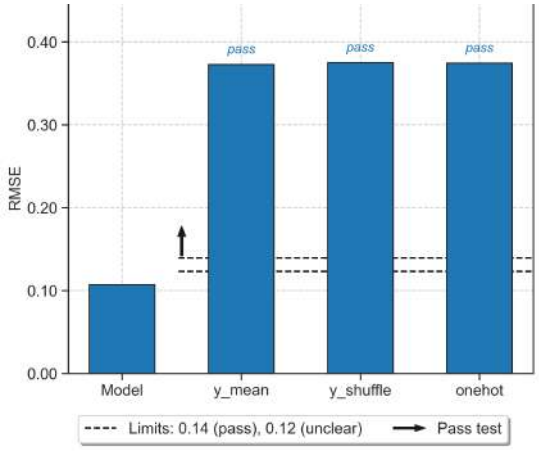


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.12%.

R^2 (10x 5-fold CV) = 0.93.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 5.88%.

R^2 (test set) = 0.94.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.96*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.9.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

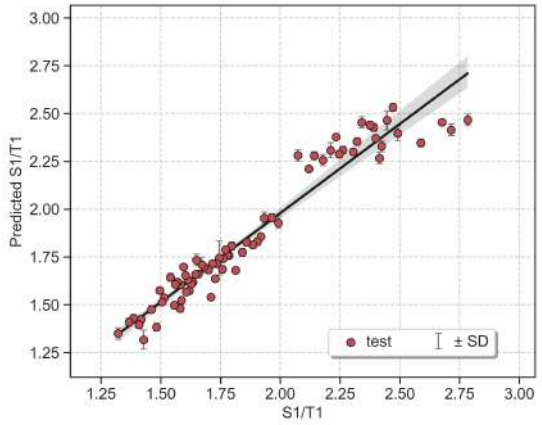
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

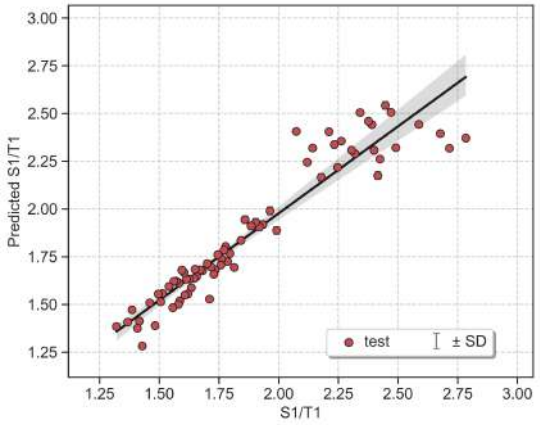


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (1 / 2)

Scaled RMSEs across 5-fold CV:

[6.25%, 5.0%, 5.62%, 7.5%, 13.12%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (1 / 2)

Scaled RMSEs across 5-fold CV:

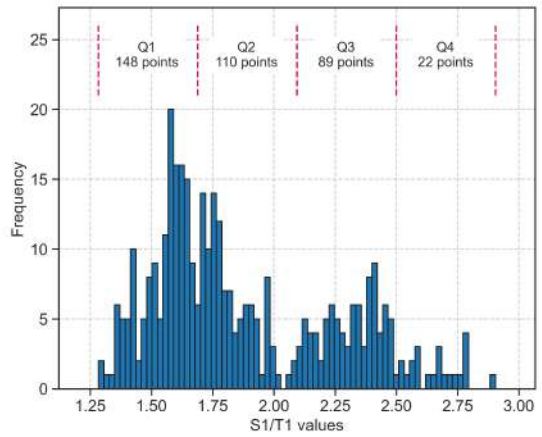
[7.5%, 6.87%, 6.87%, 10.0%, 15.62%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

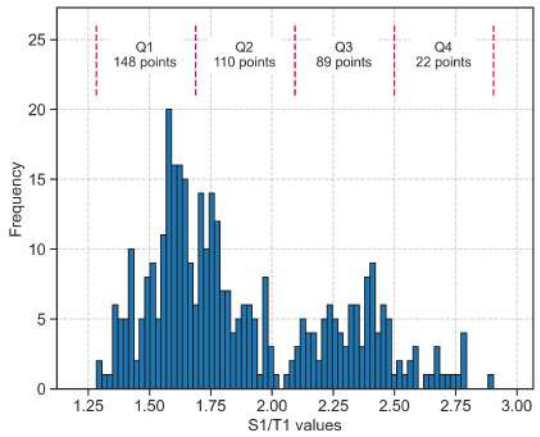
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

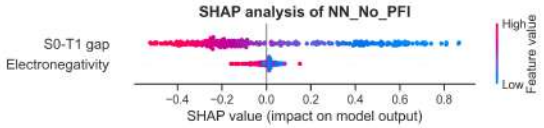


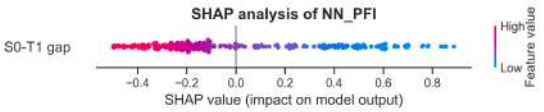
y distribution analysis

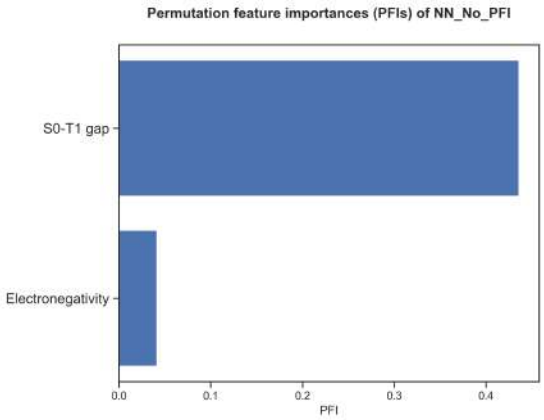
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

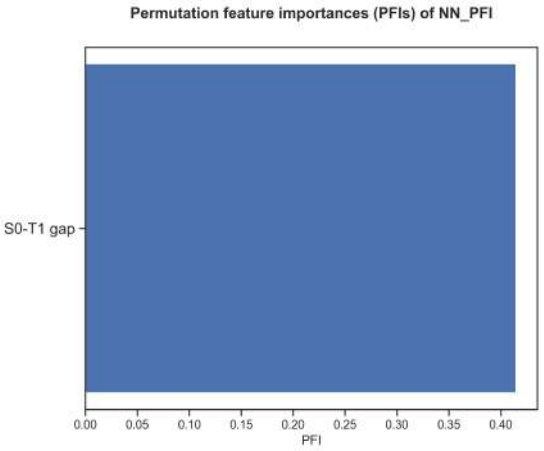
Section D. Feature Importances

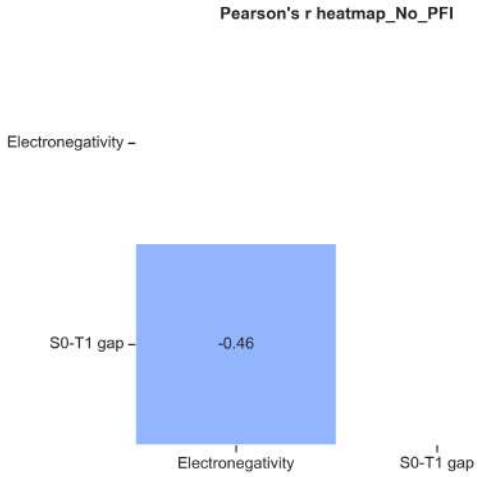
This section presents feature importances measured using the validation set.

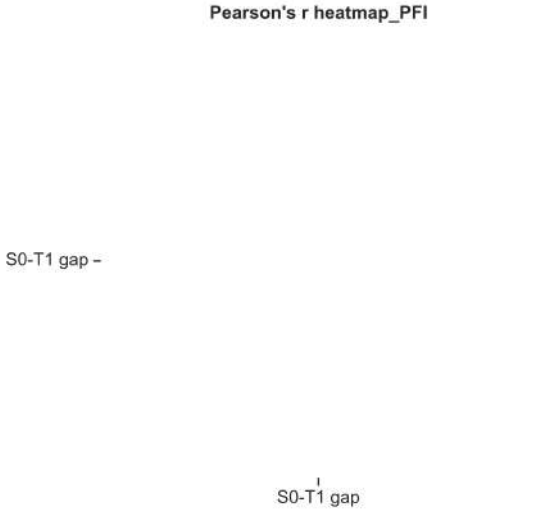












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

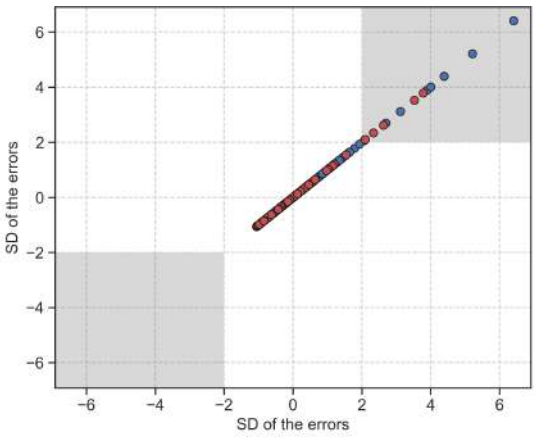
Outliers (max. 10 shown)
Train: 10 outliers out of 295 datapoints (3.4%)
- 2279 (5.2 SDs)
- 885 (4.4 SDs)
- 1007 (6.4 SDs)
- 1126 (2.7 SDs)
- 207 (2.1 SDs)
- 49 (2.1 SDs)
- 1006 (3.9 SDs)
- 2651 (4.0 SDs)
- 2649 (2.1 SDs)
- 2195 (3.1 SDs)

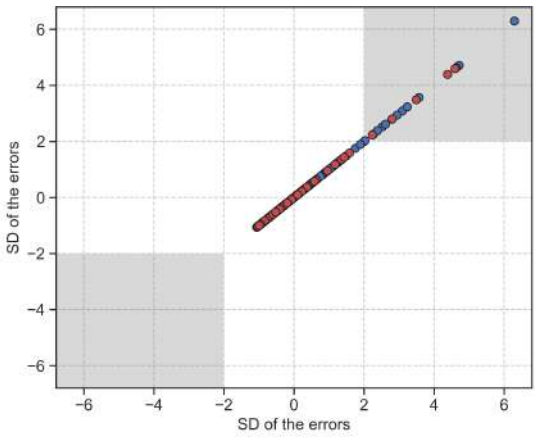
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (3.8 SDs)
- 884 (3.5 SDs)
- 890 (2.3 SDs)
- 3715 (2.6 SDs)
- 1483 (2.1 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 14 outliers out of 295 datapoints (4.7%)
- 2279 (2.8 SDs)
- 885 (4.7 SDs)
- 1007 (6.3 SDs)
- 1130 (2.0 SDs)
- 978 (2.3 SDs)
- 1126 (3.1 SDs)
- 207 (2.9 SDs)
- 889 (2.0 SDs)
- 49 (2.5 SDs)
- 1006 (2.4 SDs)

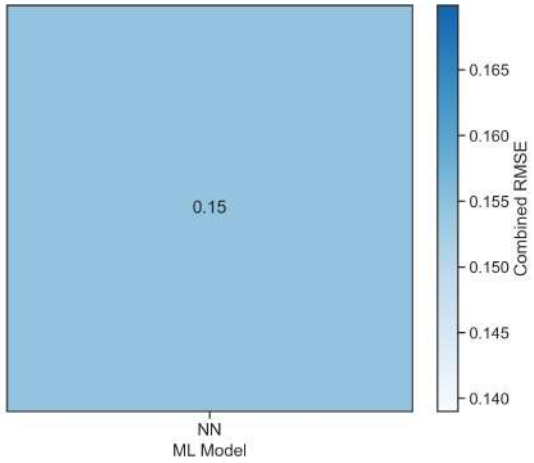
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.6 SDs)
- 884 (4.4 SDs)
- 890 (2.8 SDs)
- 891 (2.2 SDs)
- 1483 (3.5 SDs)

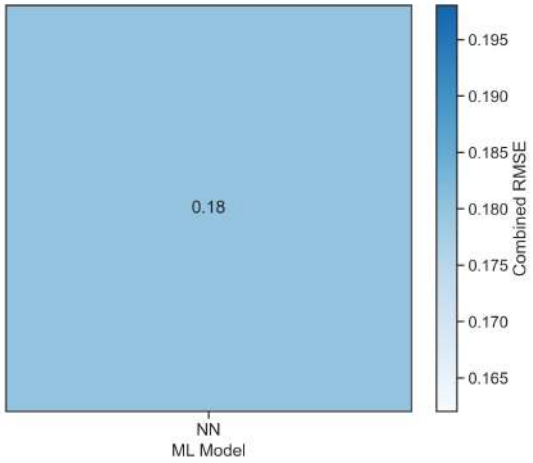




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'Total charge', 'Global SASA', 'G solv. in H2O', 'EA', 'Dipole module', 'Fermi-level', 'G of H-bonds H2O', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP']" --model "['NN']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 158.04 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: MLPRegressor	sklearn model: MLPRegressor
hidden_layer_1: 8	hidden_layer_1: 8
hidden_layer_2: 4	hidden_layer_2: 4
max_iter: 389	max_iter: 389
alpha: 0.01288474041197673	alpha: 0.01288474041197673
tol: 3.177395036752787e-05	tol: 3.177395036752787e-05
random_state: 0	random_state: 0
solver: lbfgs	solver: lbfgs

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.91, MAE = 0.051, RMSE = 0.062

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1099	2.24	2.28 ± 0.02
1853	2.17	2.25 ± 0.01
399	2.14	2.21 ± 0.01
997	2.12	2.18 ± 0.01
1754	2.08	2.15 ± 0.03
2764	2.13	2.14 ± 0.02
1015	2.13	2.14 ± 0.03
2749	1.99	2.12 ± 0.02
1139	2.12	2.11 ± 0.02
1013	2.04	2.11 ± 0.02
...	...	...
3765	1.4	1.4 ± 0.02
3516	1.37	1.4 ± 0.02
855	1.37	1.4 ± 0.02
816	1.42	1.39 ± 0.02
3657	1.4	1.39 ± 0.02
2007	1.45	1.38 ± 0.02
3468	1.33	1.38 ± 0.03
3503	1.29	1.37 ± 0.01
3779	1.26	1.36 ± 0.02
788	1.3	1.33 ± 0.04

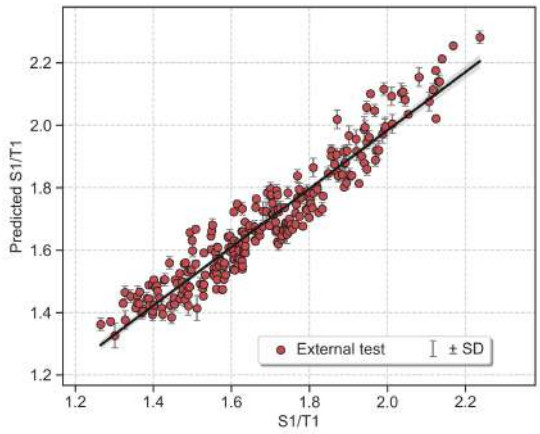
External test metrics

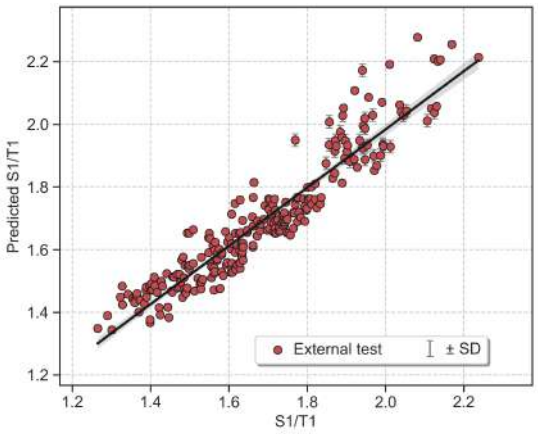
R2 = 0.88, MAE = 0.057, RMSE = 0.069

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.28 ± 0.01
1853	2.17	2.26 ± 0.01
1099	2.24	2.21 ± 0.01
997	2.12	2.21 ± 0.01
399	2.14	2.21 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.17 ± 0.02
2236	1.92	2.11 ± 0.01
1883	1.96	2.09 ± 0.01
...	...	...
3626	1.44	1.42 ± 0.01
3687	1.42	1.41 ± 0.01
855	1.37	1.4 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.38 ± 0.01
3765	1.4	1.38 ± 0.01
3657	1.4	1.37 ± 0.01
3779	1.26	1.35 ± 0.01
788	1.3	1.34 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/07 10:52:48

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

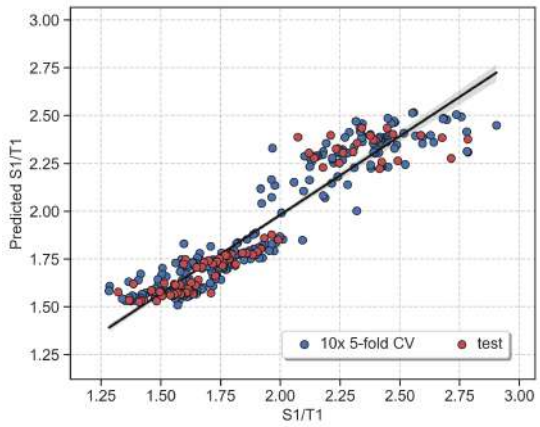
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:2



10x 5-fold CV : R^2 = 0.9, MAE = 0.093, RMSE = 0.12

Test : R^2 = 0.89, MAE = 0.09, RMSE = 0.13

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

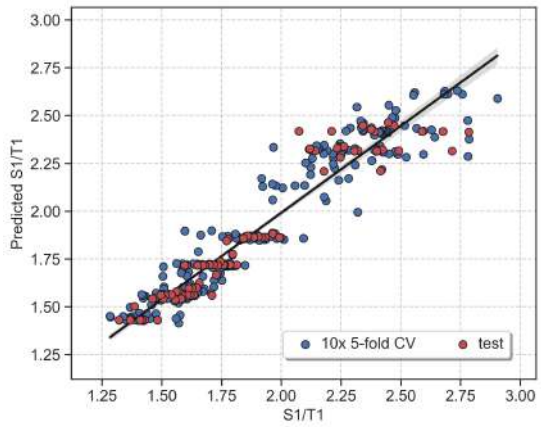
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.91, MAE = 0.082, RMSE = 0.11

Test : R^2 = 0.91, MAE = 0.077, RMSE = 0.11

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

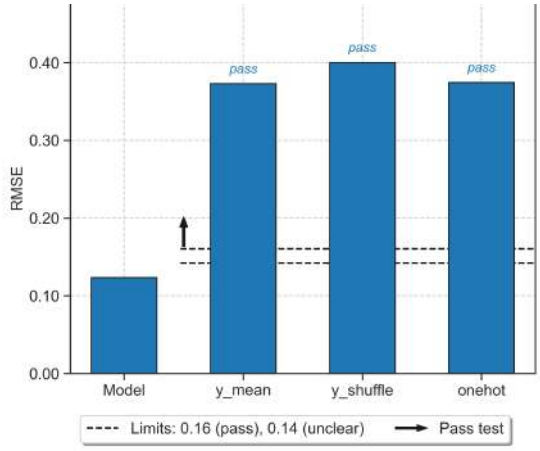
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

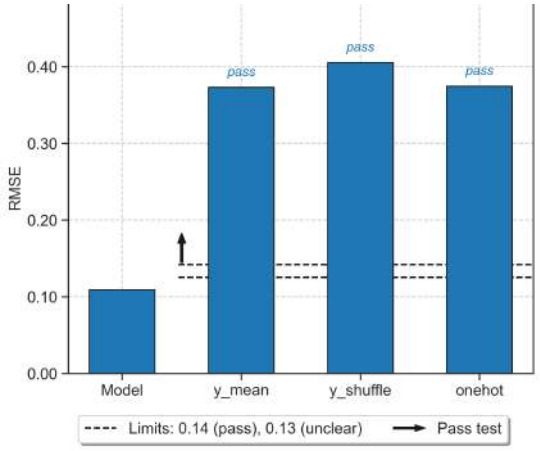


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2 ▬▬)

Scaled RMSE (10x 5-fold CV) = 7.5%.

R^2 (10x 5-fold CV) = 0.9.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2 ▬▬)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2 ▬▬)

Scaled RMSE (test set) = 8.12%.

R^2 (test set) = 0.89.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2 ▬▬)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.08*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2 ▬▬)

Scaled RMSE (test set) = 6.87%.

R^2 (test set) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2 ▬▬)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

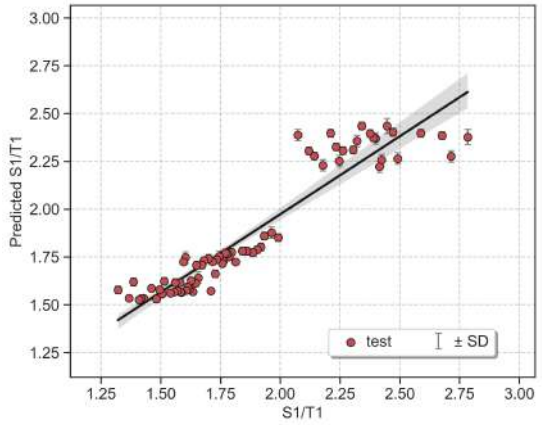
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

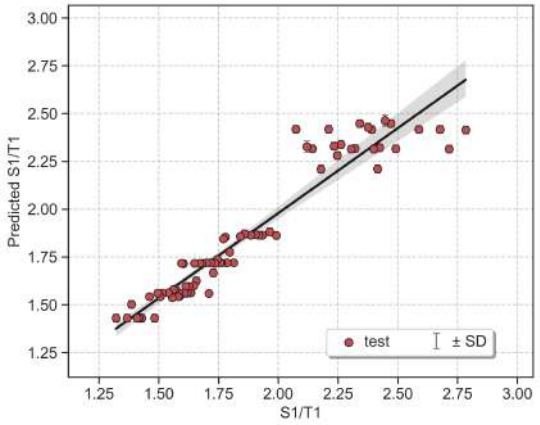


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (4% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[15.62%, 6.87%, 8.75%, 13.12%, 25.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

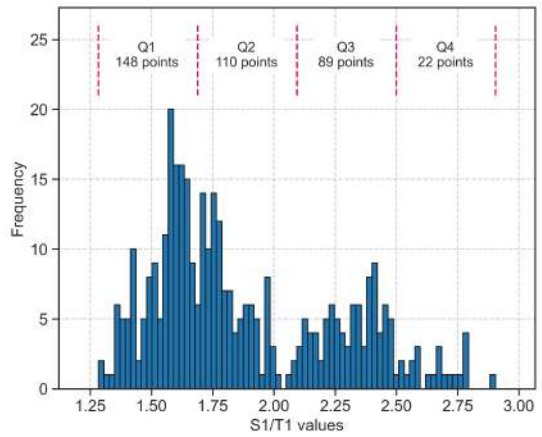
[12.5%, 8.12%, 10.0%, 14.37%, 21.25%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

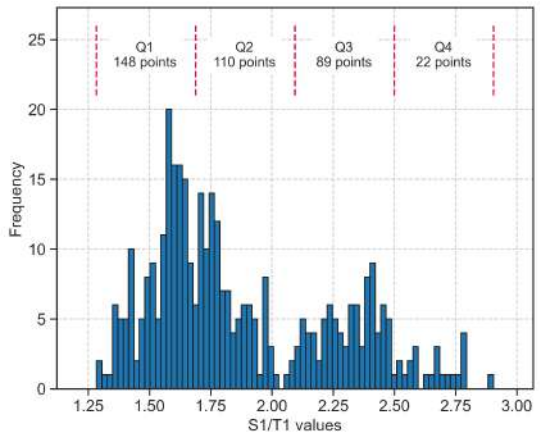
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

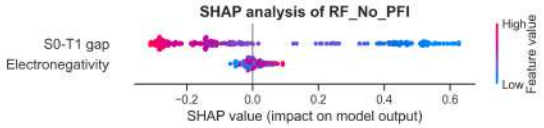


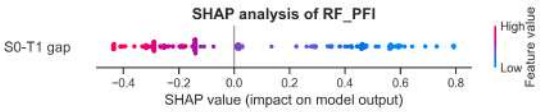
y distribution analysis

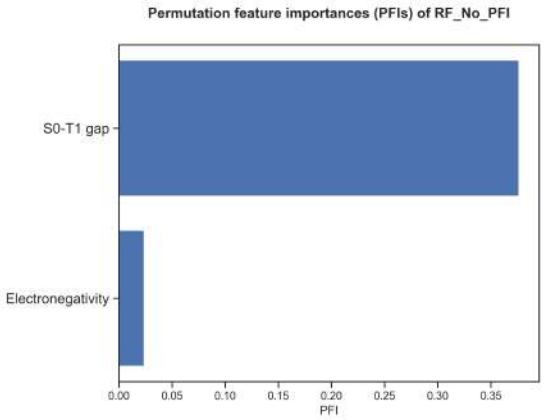
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

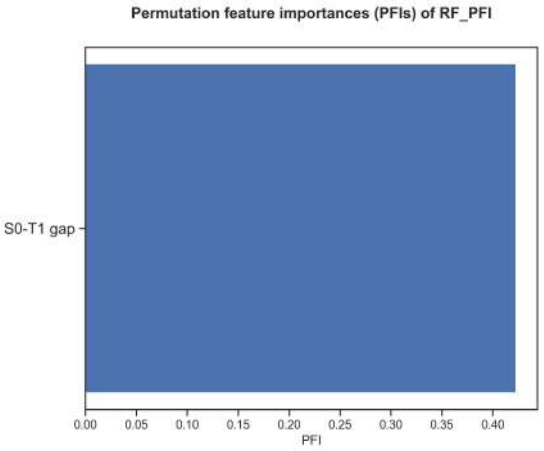
Section D. Feature Importances

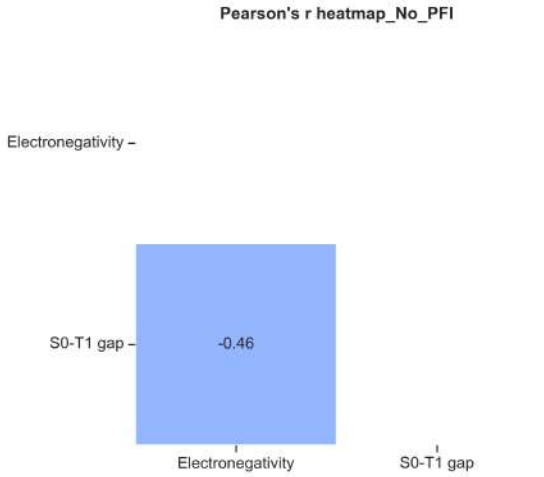
This section presents feature importances measured using the validation set.

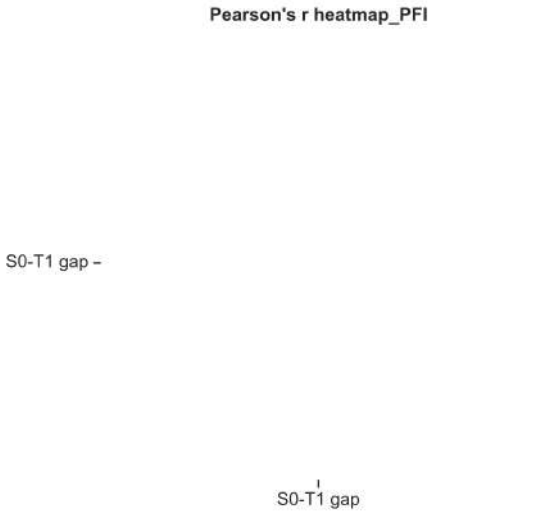












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

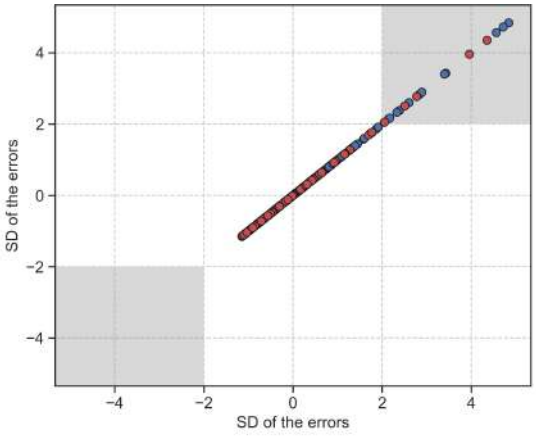
Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (4.6 SDs)
- 885 (4.8 SDs)
- 1007 (4.7 SDs)
- 1130 (3.4 SDs)
- 2767 (2.2 SDs)
- 978 (2.4 SDs)
- 207 (2.3 SDs)
- 49 (2.8 SDs)
- 345 (3.4 SDs)
- 3555 (2.0 SDs)

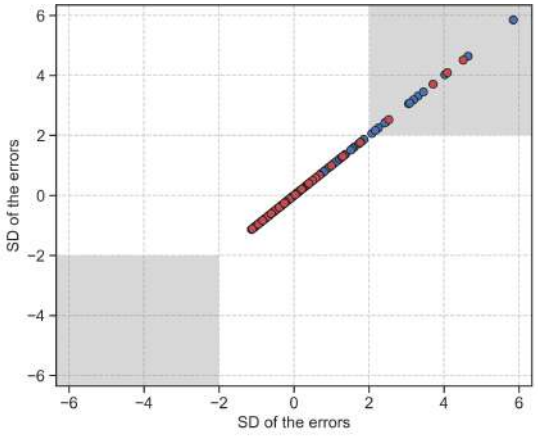
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.0 SDs)
- 884 (4.4 SDs)
- 890 (2.5 SDs)
- 1483 (2.8 SDs)
- 775 (2.1 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (3.3 SDs)
- 885 (4.6 SDs)
- 1007 (5.8 SDs)
- 1130 (3.2 SDs)
- 978 (2.5 SDs)
- 1126 (3.1 SDs)
- 207 (2.3 SDs)
- 49 (3.5 SDs)
- 1006 (2.1 SDs)
- 1014 (2.2 SDs)

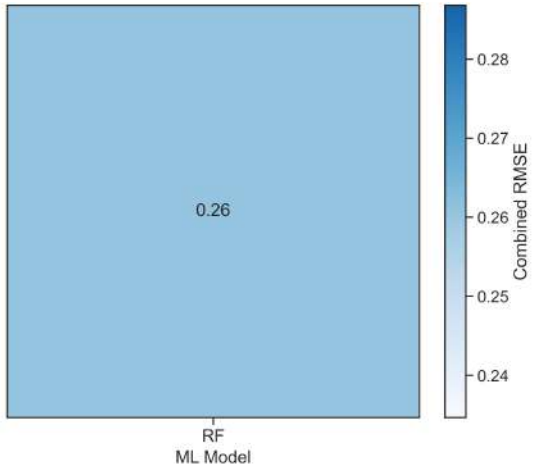
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.1 SDs)
- 884 (4.5 SDs)
- 890 (2.5 SDs)
- 1483 (3.7 SDs)

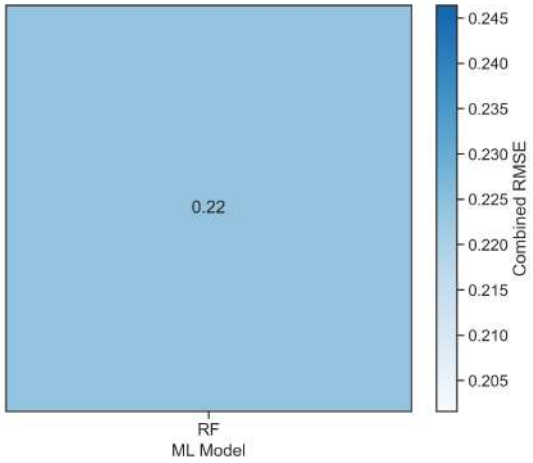




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Global SASA', 'Total charge', 'Dipole module', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP']" --model "['RF']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 221.96 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: RandomForestRegressor	sklearn model: RandomForestRegressor
n_estimators: 99	n_estimators: 99
max_depth: 12	max_depth: 12
min_samples_split: 9	min_samples_split: 9
min_samples_leaf: 3	min_samples_leaf: 3
min_weight_fraction_leaf: 0.014621720493395468	min_weight_fraction_leaf: 0.014621720493395468
max_features: 0.5933073827833075	max_features: 0.5933073827833075
ccp_alpha: 0.0012369559075742199	ccp_alpha: 0.0012369559075742199
max_samples: 0.8102811493262625	max_samples: 0.8102811493262625
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.77, MAE = 0.078, RMSE = 0.1

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.3 ± 0.02
1754	2.08	2.29 ± 0.03
399	2.14	2.28 ± 0.02
997	2.12	2.27 ± 0.02
2764	2.13	2.27 ± 0.02
1099	2.24	2.25 ± 0.03
3211	1.94	2.24 ± 0.03
1769	2.01	2.23 ± 0.02
1883	1.96	2.14 ± 0.04
2749	1.99	2.12 ± 0.04
...	...	...
3687	1.42	1.53 ± 0.01
2396	1.48	1.53 ± 0.02
3779	1.26	1.53 ± 0.01
460	1.49	1.53 ± 0.02
3765	1.4	1.53 ± 0.01
855	1.37	1.53 ± 0.02
3644	1.36	1.53 ± 0.01
3663	1.38	1.53 ± 0.01
3657	1.4	1.52 ± 0.01
3799	1.34	1.52 ± 0.01

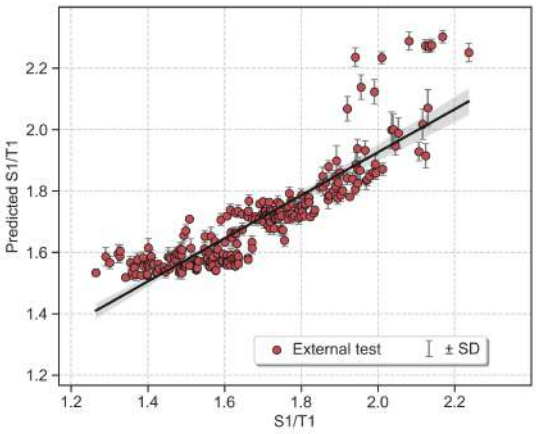
External test metrics

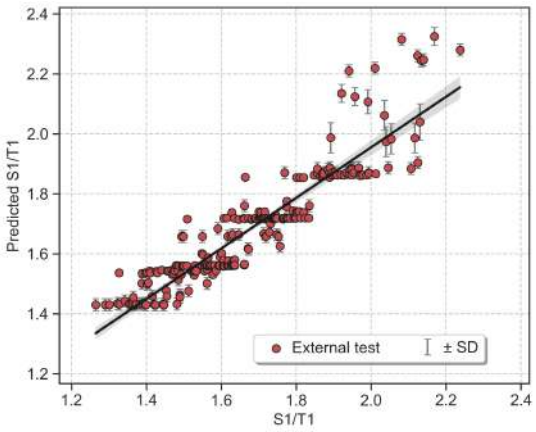
R2 = 0.84, MAE = 0.062, RMSE = 0.079

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.32 ± 0.03
1754	2.08	2.31 ± 0.02
1099	2.24	2.28 ± 0.02
997	2.12	2.26 ± 0.02
399	2.14	2.25 ± 0.02
2764	2.13	2.24 ± 0.02
1769	2.01	2.22 ± 0.02
3211	1.94	2.21 ± 0.02
2236	1.92	2.13 ± 0.03
1883	1.96	2.12 ± 0.03
...	...	...
3626	1.44	1.43 ± 0.02
3687	1.42	1.43 ± 0.02
855	1.37	1.43 ± 0.02
2007	1.45	1.43 ± 0.02
816	1.42	1.43 ± 0.02
3765	1.4	1.43 ± 0.02
3657	1.4	1.43 ± 0.02
788	1.3	1.43 ± 0.02
3503	1.29	1.43 ± 0.02
3779	1.26	1.43 ± 0.02


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2
2.4

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 13:18:37

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

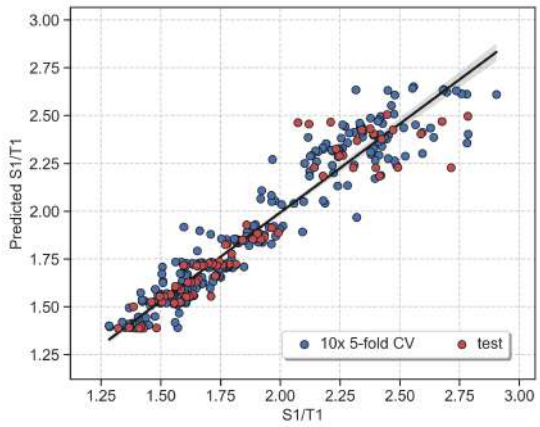
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.89, MAE = 0.081, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

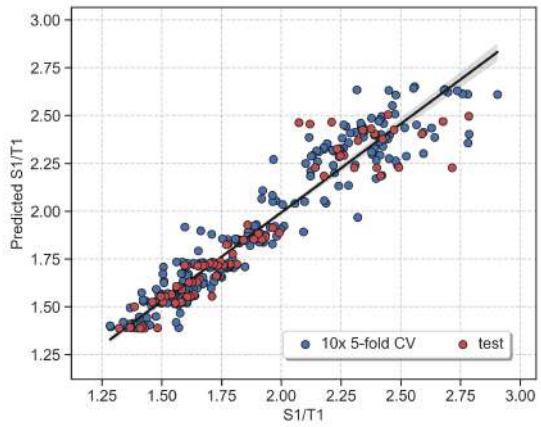
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = GB · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.89, MAE = 0.081, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

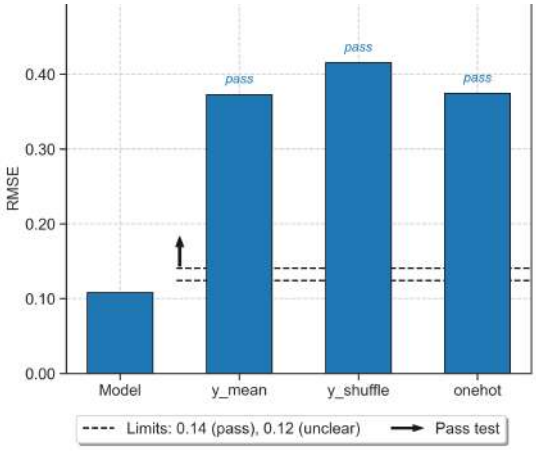
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

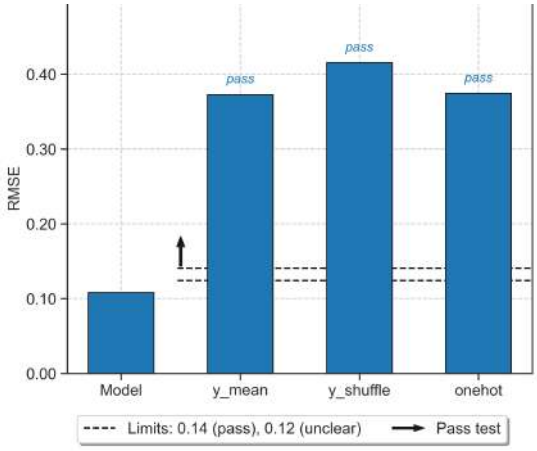


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.89.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.89.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

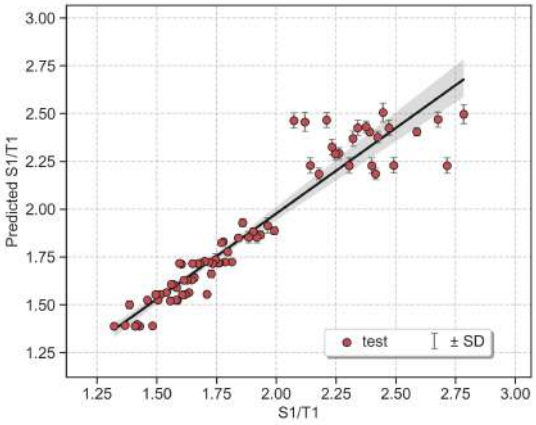
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

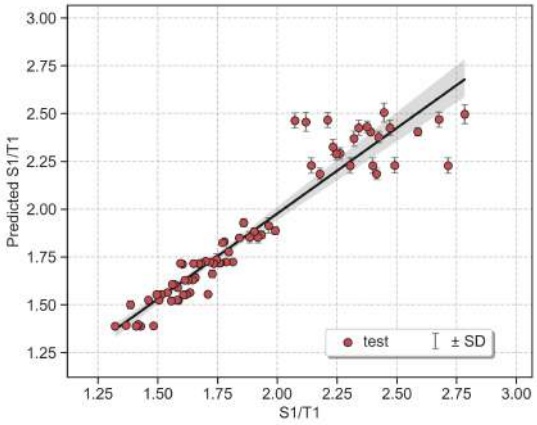


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (5% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[10.0%, 6.25%, 7.5%, 14.37%, 20.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

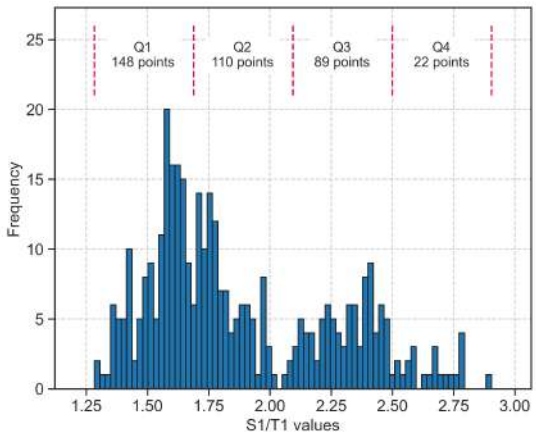
[10.0%, 6.25%, 7.5%, 14.37%, 20.0%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

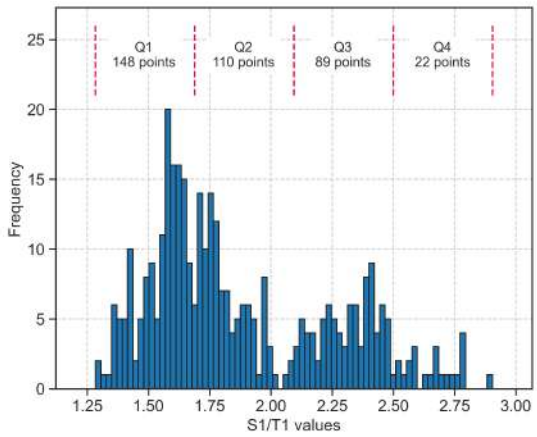
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

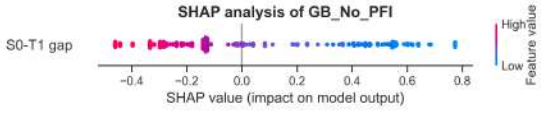


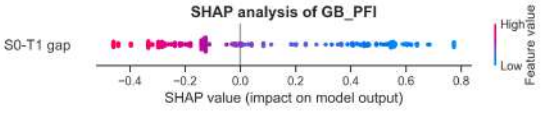
y distribution analysis

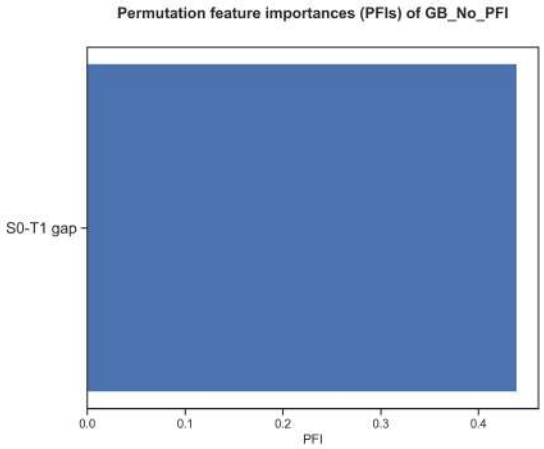
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

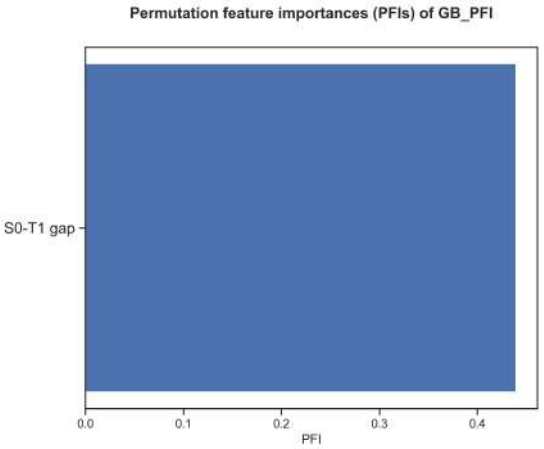
Section D. Feature Importances

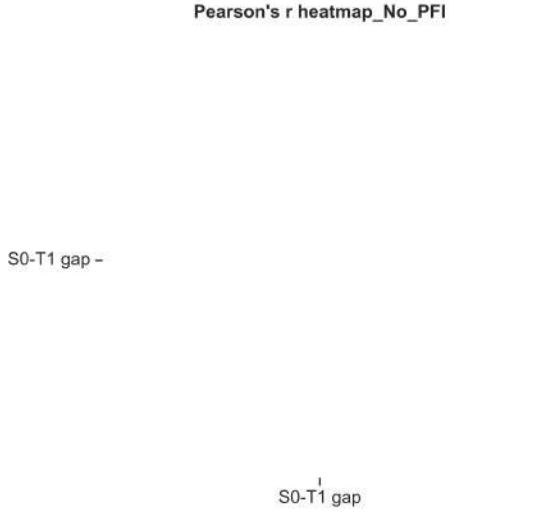
This section presents feature importances measured using the validation set.

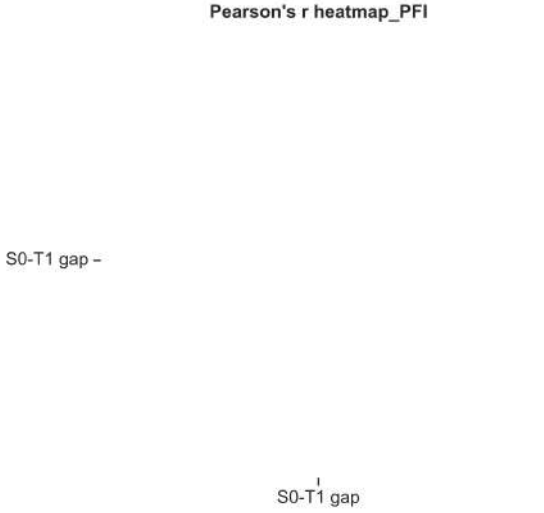












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

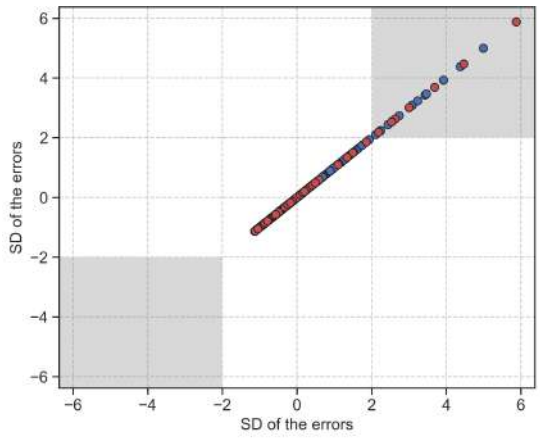
Outliers (max. 10 shown)
Train: 15 outliers out of 295 datapoints (5.1%)
- 2279 (3.1 SDs)
- 885 (4.4 SDs)
- 1007 (5.0 SDs)
- 978 (2.6 SDs)
- 1126 (3.1 SDs)
- 207 (2.2 SDs)
- 223 (2.4 SDs)
- 889 (2.2 SDs)
- 1128 (2.1 SDs)
- 49 (3.9 SDs)

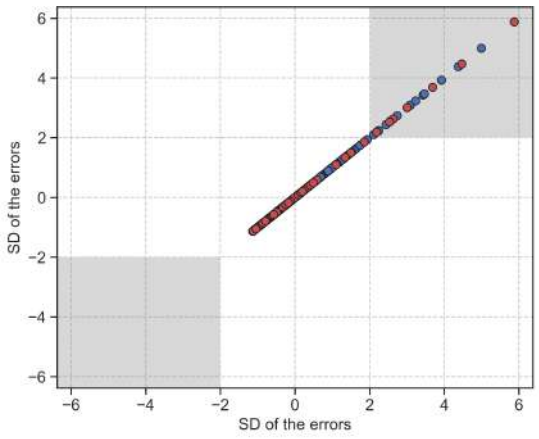
Test: 7 outliers out of 74 datapoints (9.5%)
- 888 (3.0 SDs)
- 884 (5.9 SDs)
- 872 (2.6 SDs)
- 891 (2.2 SDs)
- 3222 (2.5 SDs)
- 412 (3.7 SDs)
- 1483 (4.5 SDs)

PFI (only important descriptors):

Outliers (max. 10 shown)
Train: 15 outliers out of 295 datapoints (5.1%)
- 2279 (3.1 SDs)
- 885 (4.4 SDs)
- 1007 (5.0 SDs)
- 978 (2.6 SDs)
- 1126 (3.1 SDs)
- 207 (2.2 SDs)
- 223 (2.4 SDs)
- 889 (2.2 SDs)
- 1128 (2.1 SDs)
- 49 (3.9 SDs)

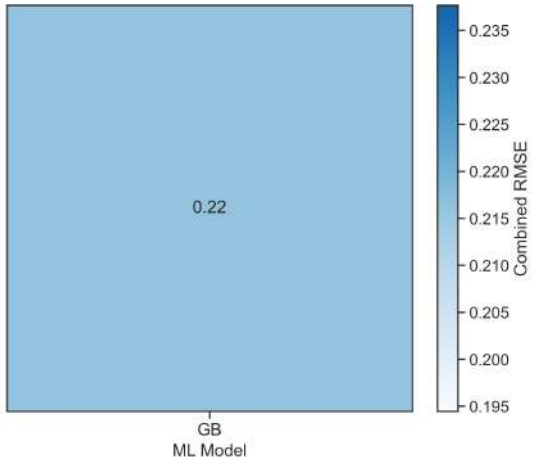
Test: 7 outliers out of 74 datapoints (9.5%)
- 888 (3.0 SDs)
- 884 (5.9 SDs)
- 872 (2.6 SDs)
- 891 (2.2 SDs)
- 3222 (2.5 SDs)
- 412 (3.7 SDs)
- 1483 (4.5 SDs)

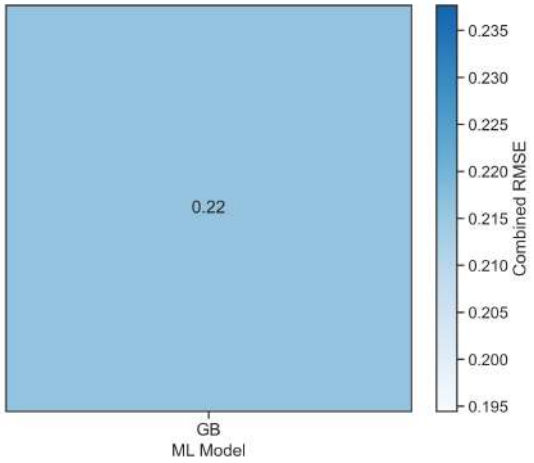




Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP', 'Electronegativity']" --model "['GB']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 96.13 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: GradientBoostingRegressor	sklearn model: GradientBoostingRegressor
n_estimators: 39	n_estimators: 39
learning_rate: 0.2035490101894677	learning_rate: 0.2035490101894677
max_depth: 7	max_depth: 7
min_samples_split: 8	min_samples_split: 8
min_samples_leaf: 3	min_samples_leaf: 3
subsample: 0.754957408602135	subsample: 0.754957408602135
max_features: 0.6898847011075624	max_features: 0.6898847011075624
validation_fraction: 0.10402150923749871	validation_fraction: 0.10402150923749871
min_weight_fraction_leaf: 0.04144700146086816	min_weight_fraction_leaf: 0.04144700146086816
ccp_alpha: 4.695476192547066e-05	ccp_alpha: 4.695476192547066e-05
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.86, MAE = 0.058, RMSE = 0.075

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.4 ± 0.04
1099	2.24	2.31 ± 0.02
1754	2.08	2.3 ± 0.04
997	2.12	2.27 ± 0.03
2764	2.13	2.25 ± 0.04
399	2.14	2.24 ± 0.04
1769	2.01	2.2 ± 0.03
3211	1.94	2.18 ± 0.03
2749	1.99	2.08 ± 0.03
1123	2.04	2.06 ± 0.06
...	...	...
3626	1.44	1.4 ± 0.01
3687	1.42	1.4 ± 0.01
855	1.37	1.39 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.39 ± 0.01
3765	1.4	1.39 ± 0.01
3657	1.4	1.39 ± 0.01
788	1.3	1.39 ± 0.01
3779	1.26	1.39 ± 0.01

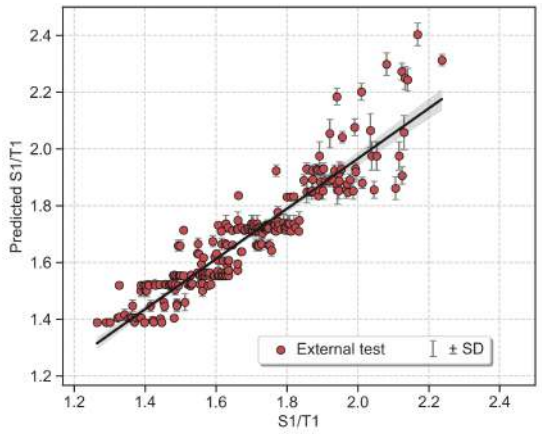
External test metrics

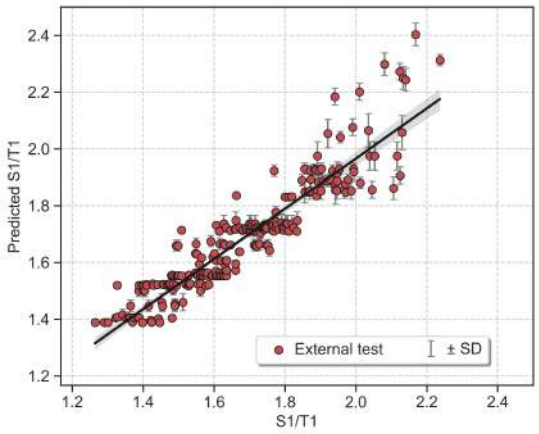
R2 = 0.86, MAE = 0.058, RMSE = 0.075

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.4 ± 0.04
1099	2.24	2.31 ± 0.02
1754	2.08	2.3 ± 0.04
997	2.12	2.27 ± 0.03
2764	2.13	2.25 ± 0.04
399	2.14	2.24 ± 0.04
1769	2.01	2.2 ± 0.03
3211	1.94	2.18 ± 0.03
2749	1.99	2.08 ± 0.03
1123	2.04	2.06 ± 0.06
...	...	...
3626	1.44	1.4 ± 0.01
3687	1.42	1.4 ± 0.01
855	1.37	1.39 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.39 ± 0.01
3765	1.4	1.39 ± 0.01
3657	1.4	1.39 ± 0.01
788	1.3	1.39 ± 0.01
3779	1.26	1.39 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2
2.4


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2
2.4

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 13:21:33

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

Section A. ROBERT Score

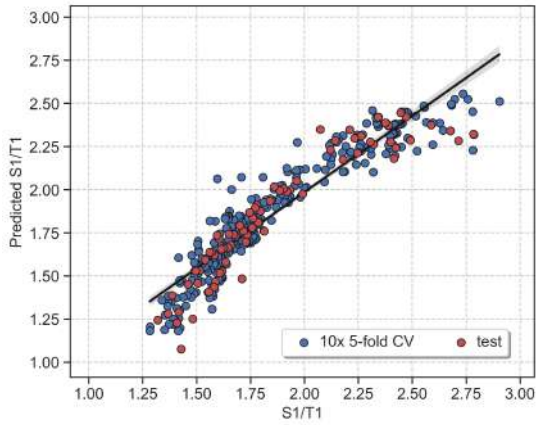
This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1

MODERATE



10x 5-fold CV : R^2 = 0.88, MAE = 0.096, RMSE = 0.13

Test : R^2 = 0.86, MAE = 0.1, RMSE = 0.14

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

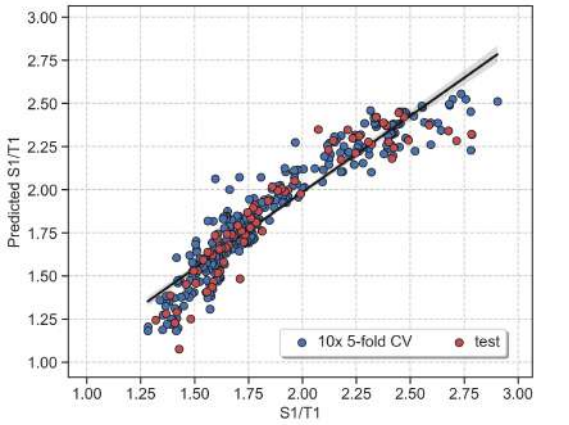
Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1

MODERATE



10x 5-fold CV : R^2 = 0.88, MAE = 0.096, RMSE = 0.13

Test : R^2 = 0.86, MAE = 0.1, RMSE = 0.14

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

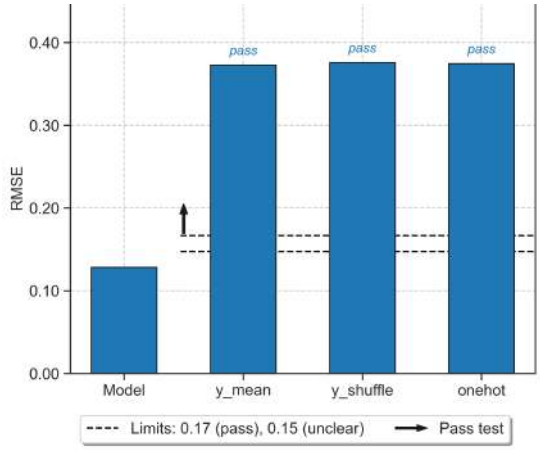
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

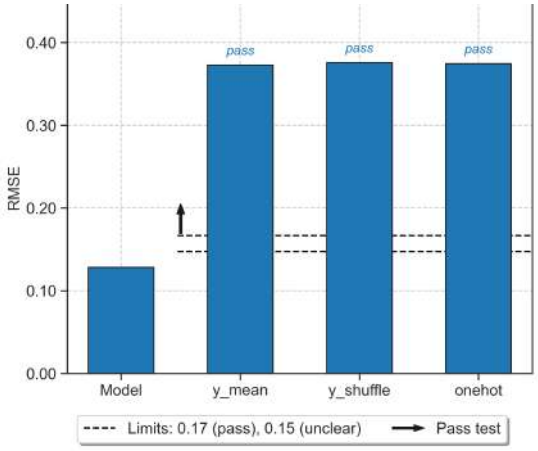


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 8.12%.

R^2 (10x 5-fold CV) = 0.88.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.75%.

R^2 (test set) = 0.86.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.08*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.75%.

R^2 (test set) = 0.86.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.08*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

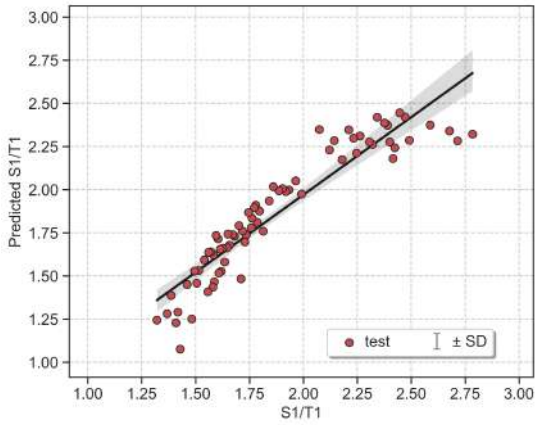
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (1% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

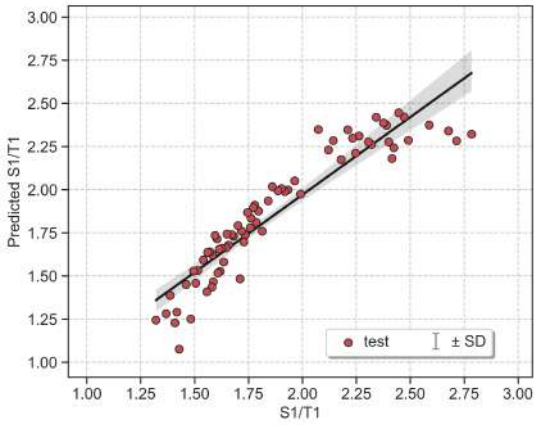


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (1% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[9.37%, 8.12%, 6.25%, 6.87%, 18.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

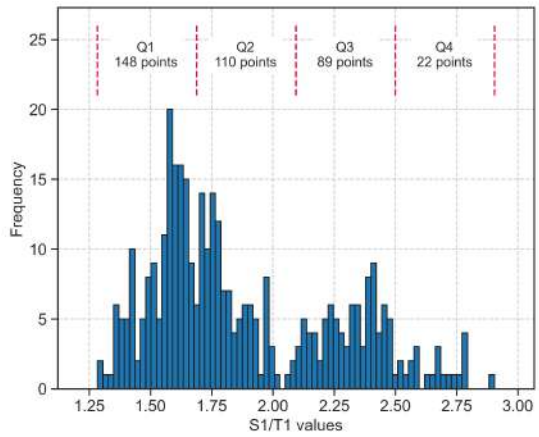
[9.37%, 8.12%, 6.25%, 6.87%, 18.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

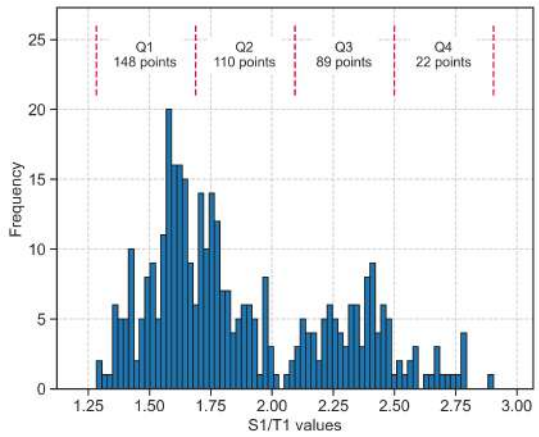
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

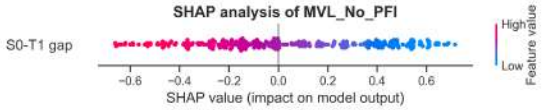


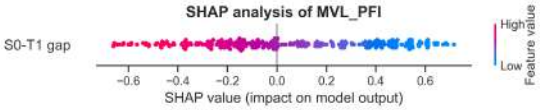
y distribution analysis

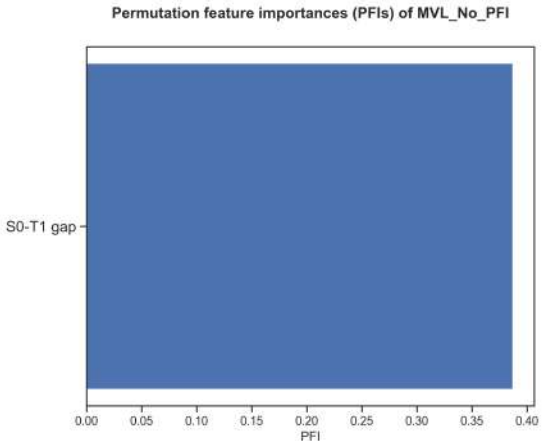
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

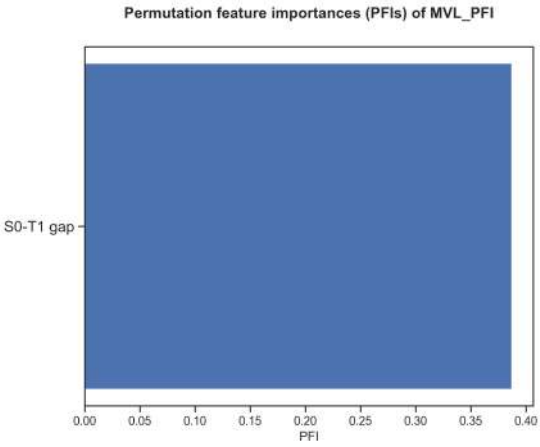
Section D. Feature Importances

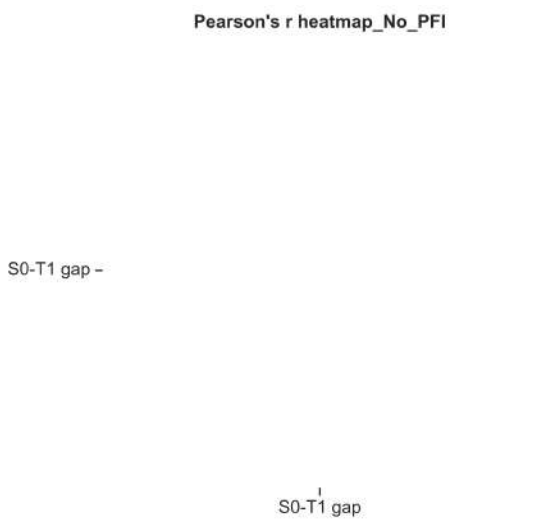
This section presents feature importances measured using the validation set.

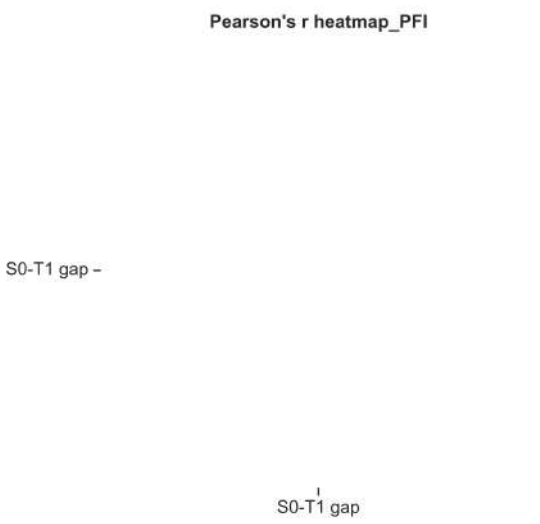












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

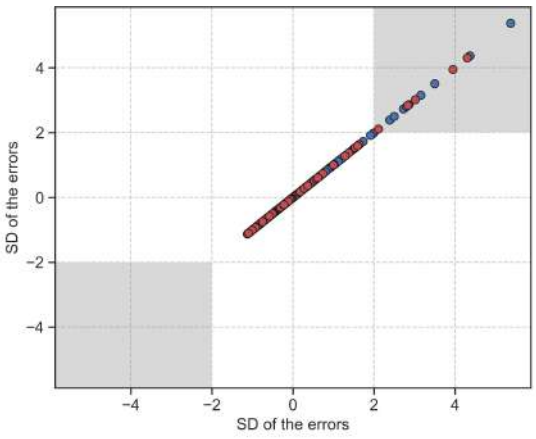
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (3.5 SDs)
- 885 (4.4 SDs)
- 1007 (5.4 SDs)
- 1130 (2.7 SDs)
- 978 (2.4 SDs)
- 1126 (2.8 SDs)
- 207 (2.4 SDs)
- 345 (2.5 SDs)
- 3264 (3.2 SDs)
- 1362 (2.9 SDs)

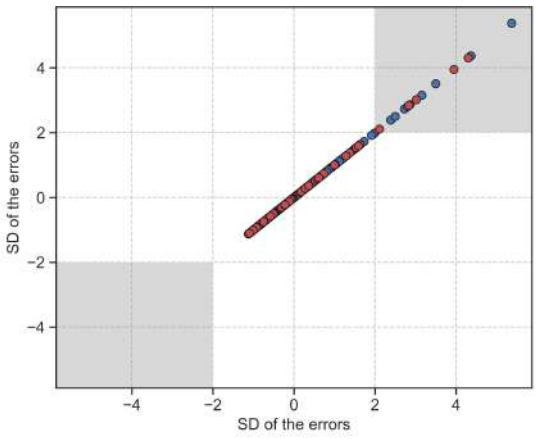
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.3 SDs)
- 884 (4.0 SDs)
- 890 (2.8 SDs)
- 1483 (2.1 SDs)
- 46 (3.0 SDs)



PFI (only important descriptors):

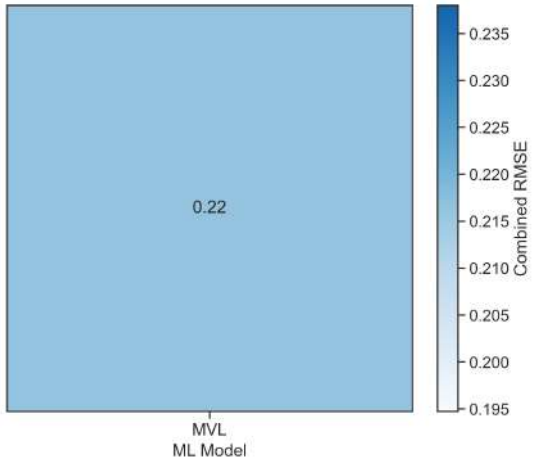
Outliers (max. 10 shown)
Train: 11 outliers out of 295 datapoints (3.7%)
- 2279 (3.5 SDs)
- 885 (4.4 SDs)
- 1007 (5.4 SDs)
- 1130 (2.7 SDs)
- 978 (2.4 SDs)
- 1126 (2.8 SDs)
- 207 (2.4 SDs)
- 345 (2.5 SDs)
- 3264 (3.2 SDs)
- 1362 (2.9 SDs)

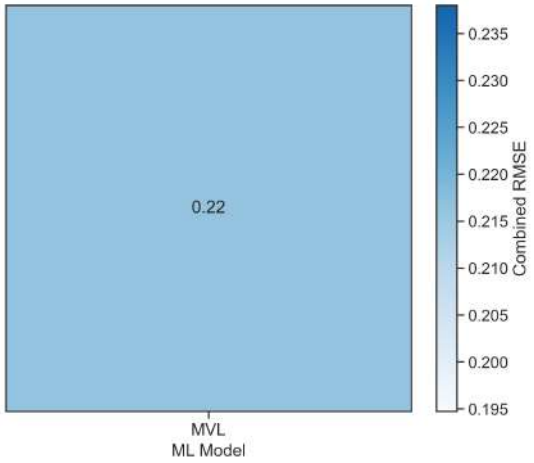
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.3 SDs)
- 884 (4.0 SDs)
- 890 (2.8 SDs)
- 1483 (2.1 SDs)
- 46 (3.0 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electronegativity', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP', 'Electrophil. idx']" --model "['MVL']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 13.76 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: LinearRegression	sklearn model: LinearRegression

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.89, MAE = 0.072, RMSE = 0.093

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.25 ± 0.01
1853	2.17	2.24 ± 0.01
1099	2.24	2.21 ± 0.01
997	2.12	2.21 ± 0.01
399	2.14	2.2 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.18 ± 0.01
2236	1.92	2.13 ± 0.01
1883	1.96	2.12 ± 0.01
...	...	...
3626	1.44	1.3 ± 0.01
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3503	1.29	1.25 ± 0.01
2007	1.45	1.24 ± 0.01
3765	1.4	1.23 ± 0.01
3657	1.4	1.22 ± 0.01
3779	1.26	1.19 ± 0.01
788	1.3	1.18 ± 0.01

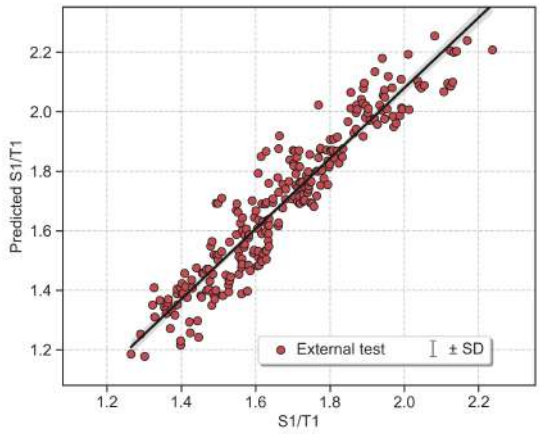
External test metrics

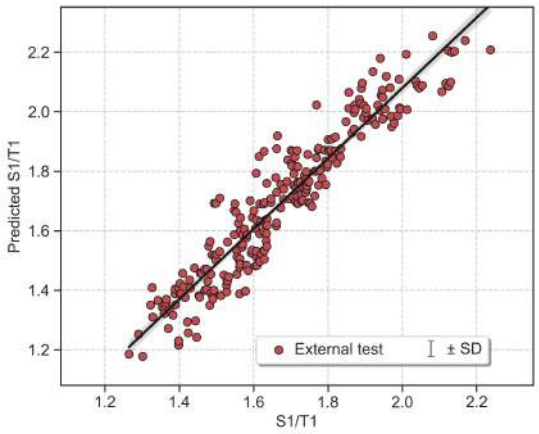
R2 = 0.89, MAE = 0.072, RMSE = 0.093

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.25 ± 0.01
1853	2.17	2.24 ± 0.01
1099	2.24	2.21 ± 0.01
997	2.12	2.21 ± 0.01
399	2.14	2.2 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.18 ± 0.01
2236	1.92	2.13 ± 0.01
1883	1.96	2.12 ± 0.01
...	...	...
3626	1.44	1.3 ± 0.01
3687	1.42	1.29 ± 0.01
855	1.37	1.27 ± 0.01
816	1.42	1.26 ± 0.01
3503	1.29	1.25 ± 0.01
2007	1.45	1.24 ± 0.01
3765	1.4	1.23 ± 0.01
3657	1.4	1.22 ± 0.01
3779	1.26	1.19 ± 0.01
788	1.3	1.18 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/07 10:12:53

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

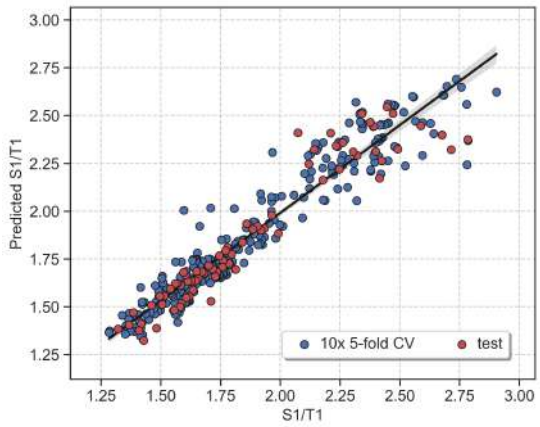
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.078, RMSE = 0.11

Test : R^2 = 0.9, MAE = 0.079, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

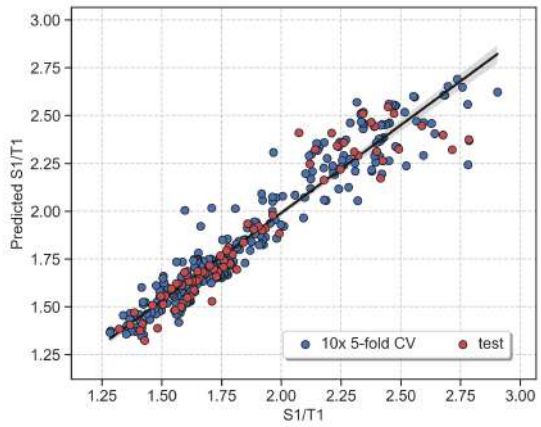
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.92, MAE = 0.078, RMSE = 0.11

Test : R^2 = 0.9, MAE = 0.079, RMSE = 0.12

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

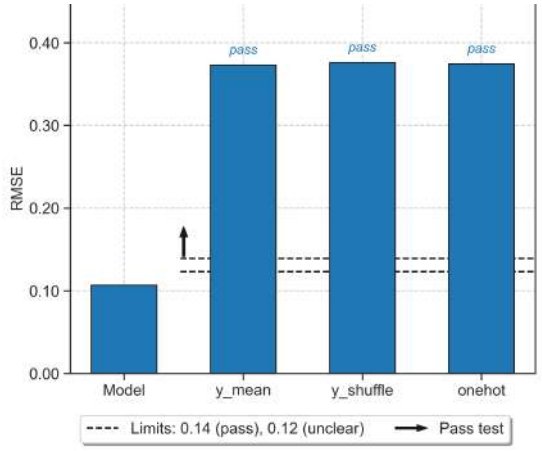
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

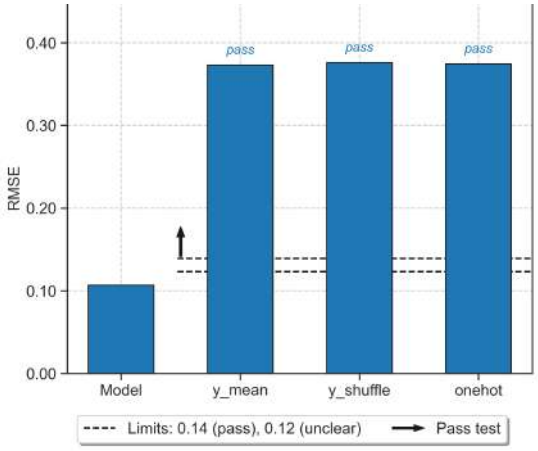


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.92.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.9.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 7.5%.

R^2 (test set) = 0.9.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.09*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

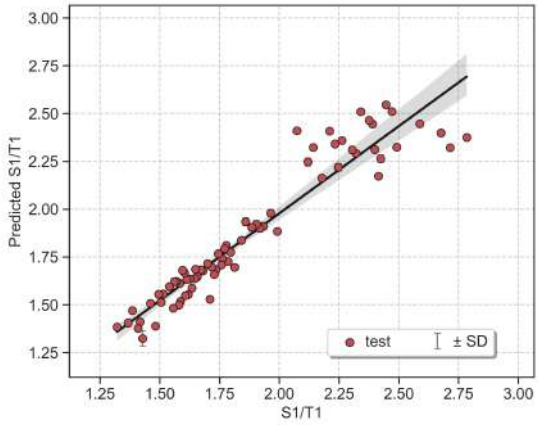
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

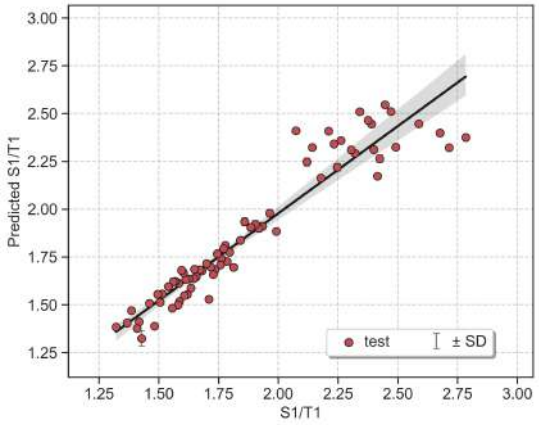


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (2% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[10.0%, 6.25%, 6.25%, 10.0%, 13.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

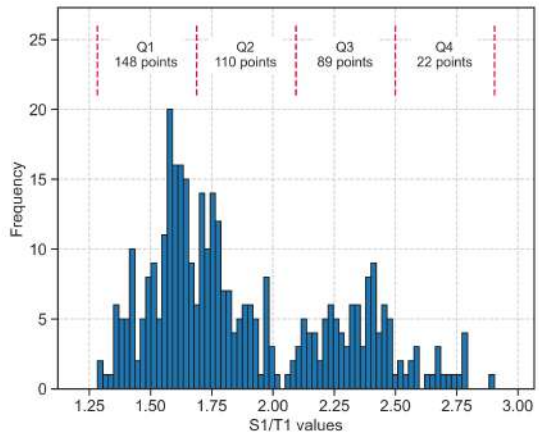
[10.0%, 6.25%, 6.25%, 10.0%, 13.75%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

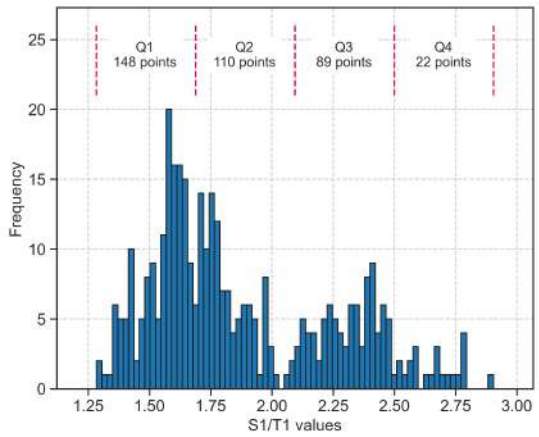
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

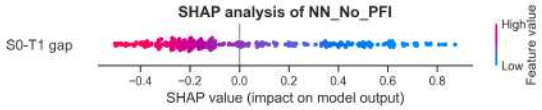


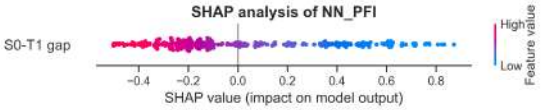
y distribution analysis

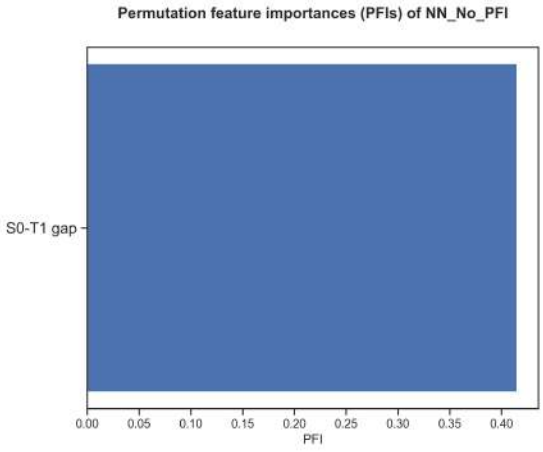
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

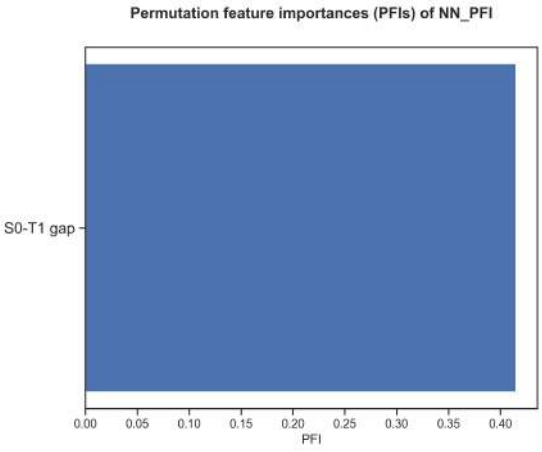
Section D. Feature Importances

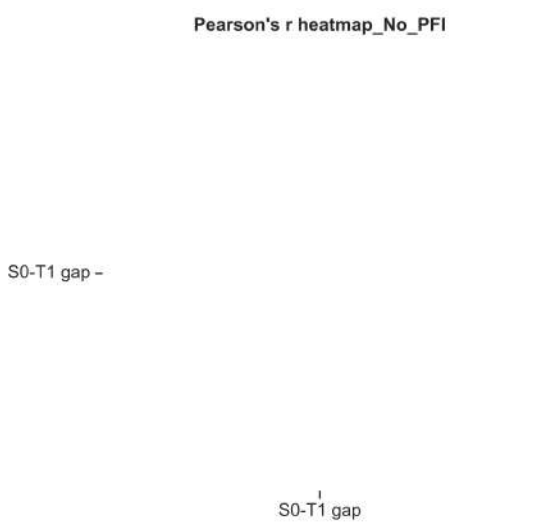
This section presents feature importances measured using the validation set.

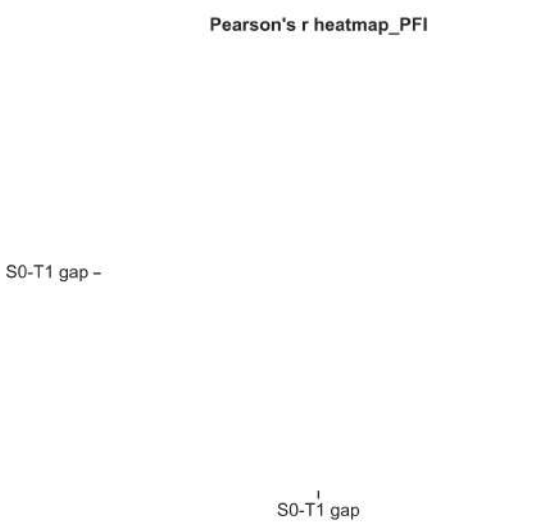












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

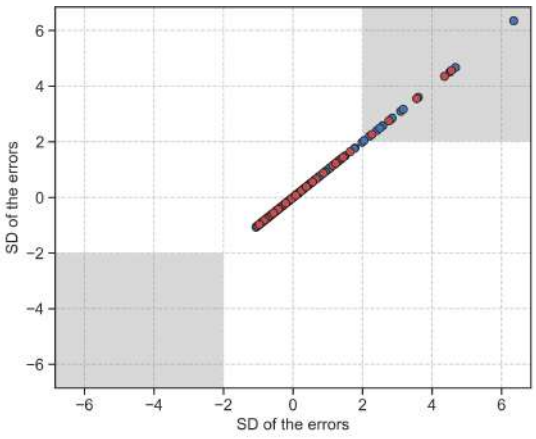
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (2.8 SDs)
- 885 (4.7 SDs)
- 1007 (6.3 SDs)
- 978 (2.2 SDs)
- 1126 (3.1 SDs)
- 207 (2.9 SDs)
- 889 (2.0 SDs)
- 49 (2.6 SDs)
- 1006 (2.4 SDs)
- 345 (3.6 SDs)

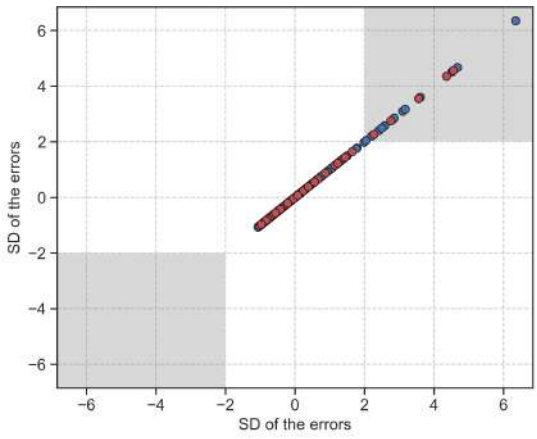
Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.6 SDs)
- 884 (4.4 SDs)
- 890 (2.8 SDs)
- 891 (2.3 SDs)
- 1483 (3.6 SDs)



PFI (only important descriptors):

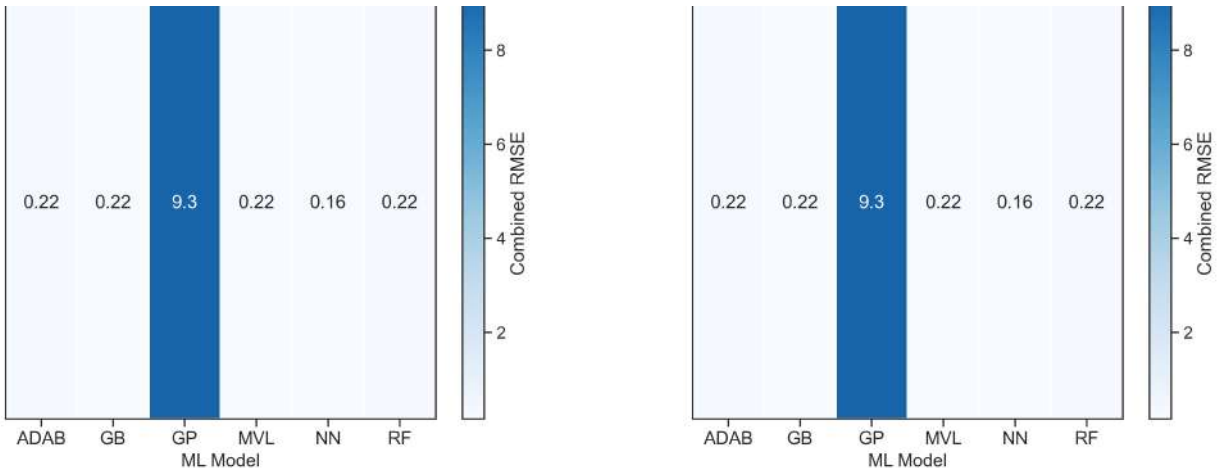
Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (2.8 SDs)
- 885 (4.7 SDs)
- 1007 (6.3 SDs)
- 978 (2.2 SDs)
- 1126 (3.1 SDs)
- 207 (2.9 SDs)
- 889 (2.0 SDs)
- 49 (2.6 SDs)
- 1006 (2.4 SDs)
- 345 (3.6 SDs)

Test: 5 outliers out of 74 datapoints (6.8%)
- 888 (4.6 SDs)
- 884 (4.4 SDs)
- 890 (2.8 SDs)
- 891 (2.3 SDs)
- 1483 (3.6 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.



Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'LUMO', 'IP', 'HOMO', 'EA', 'Global SASA', 'G solv. in H2O', 'Fermi-level', 'G of H-bonds H2O', 'Total polariz. alpha', 'Total FOD', 'Total charge', 'Dipole module', 'Hardness', 'Softness', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP', 'Electronegativity']" --model "['AdaB','GB','GP','MVL','NN','RF']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 355.22 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: MLPRegressor	sklearn model: MLPRegressor
hidden_layer_1: 7	hidden_layer_1: 7
hidden_layer_2: 1	hidden_layer_2: 1
max_iter: 483	max_iter: 483
alpha: 0.056966348957506456	alpha: 0.056966348957506456
tol: 4.7319574599147124e-05	tol: 4.7319574599147124e-05
random_state: 0	random_state: 0
solver: lbfgs	solver: lbfgs

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.88, MAE = 0.056, RMSE = 0.069

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.28 ± 0.01
1853	2.17	2.26 ± 0.02
1099	2.24	2.21 ± 0.02
997	2.12	2.21 ± 0.01
399	2.14	2.21 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.17 ± 0.01
2236	1.92	2.1 ± 0.01
1883	1.96	2.08 ± 0.01
...	...	...
3626	1.44	1.41 ± 0.01
3687	1.42	1.41 ± 0.01
855	1.37	1.4 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.38 ± 0.01
3765	1.4	1.38 ± 0.01
3657	1.4	1.37 ± 0.01
3779	1.26	1.35 ± 0.01
788	1.3	1.35 ± 0.01

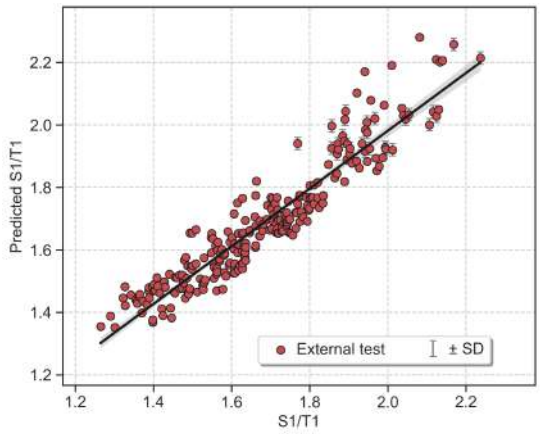
External test metrics

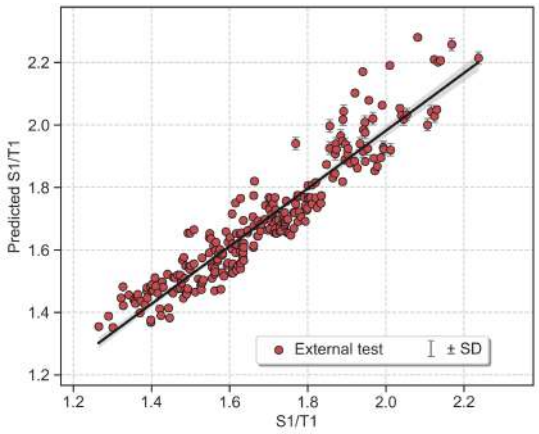
R2 = 0.88, MAE = 0.056, RMSE = 0.069

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1754	2.08	2.28 ± 0.01
1853	2.17	2.26 ± 0.02
1099	2.24	2.21 ± 0.02
997	2.12	2.21 ± 0.01
399	2.14	2.21 ± 0.01
2764	2.13	2.2 ± 0.01
1769	2.01	2.19 ± 0.01
3211	1.94	2.17 ± 0.01
2236	1.92	2.1 ± 0.01
1883	1.96	2.08 ± 0.01
...	...	...
3626	1.44	1.41 ± 0.01
3687	1.42	1.41 ± 0.01
855	1.37	1.4 ± 0.01
816	1.42	1.39 ± 0.01
3503	1.29	1.39 ± 0.01
2007	1.45	1.38 ± 0.01
3765	1.4	1.38 ± 0.01
3657	1.4	1.37 ± 0.01
3779	1.26	1.35 ± 0.01
788	1.3	1.35 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2


Predicted S1/T1
S1/T1
External test
± SD
1.2
1.4
1.6
1.8
2.0
2.2

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2025/11/10 13:14:02

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

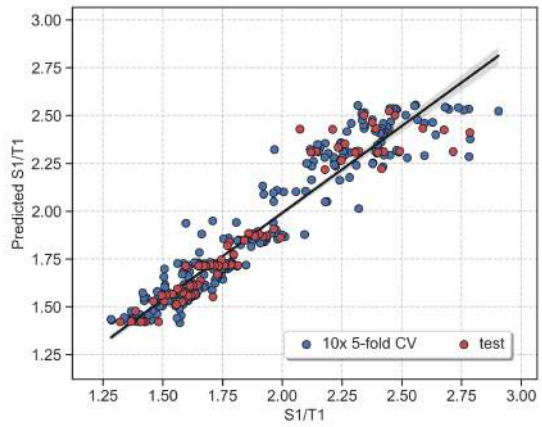
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.91, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.91, MAE = 0.077, RMSE = 0.11

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

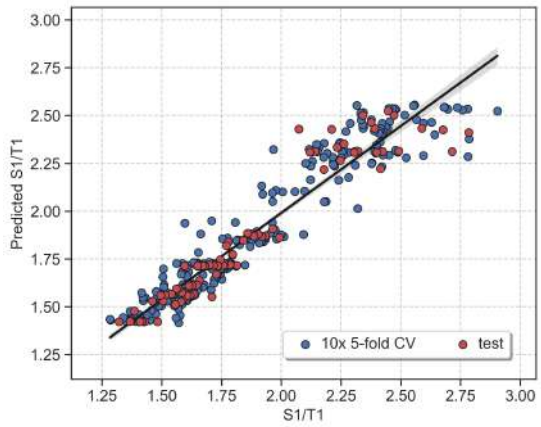
Overall assessment

Decent model, but it has limitations

PFI (only important descriptors) · Score 8

Model = RF · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 295:1



10x 5-fold CV : R^2 = 0.91, MAE = 0.081, RMSE = 0.11

Test : R^2 = 0.91, MAE = 0.077, RMSE = 0.11

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

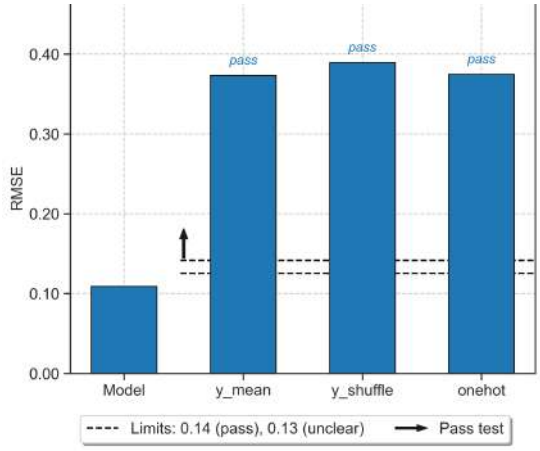
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

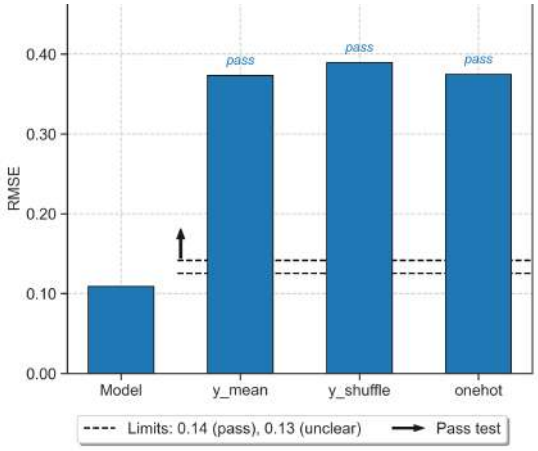


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 6.87%.

R^2 (10x 5-fold CV) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.87%.

R^2 (test set) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 6.87%.

R^2 (test set) = 0.91.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.0*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

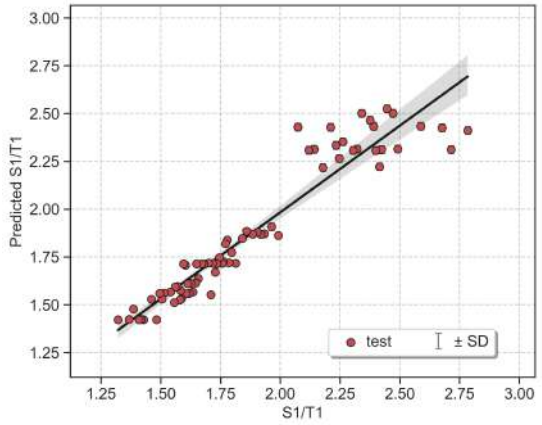
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (3% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

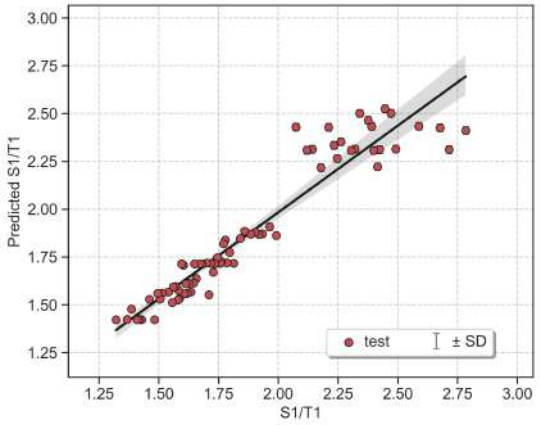


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.1 (3% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

[11.25%, 6.87%, 8.75%, 13.12%, 20.62%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

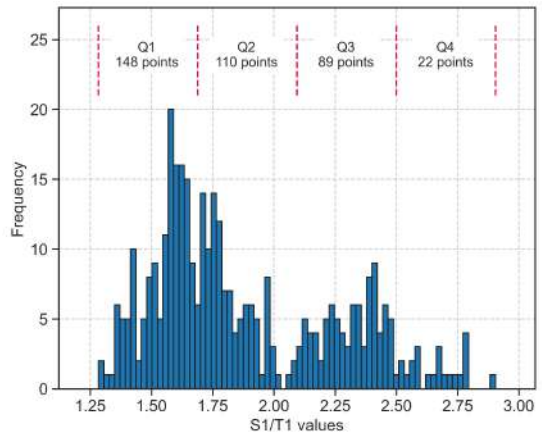
[11.25%, 6.87%, 8.75%, 13.12%, 20.62%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

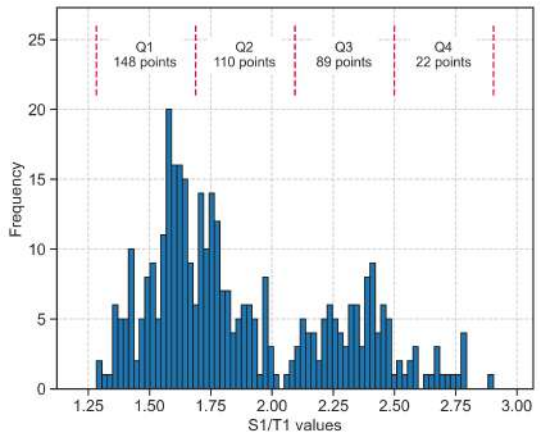
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

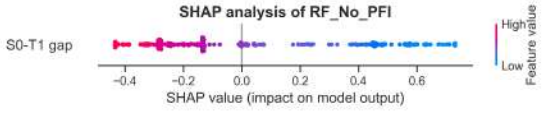


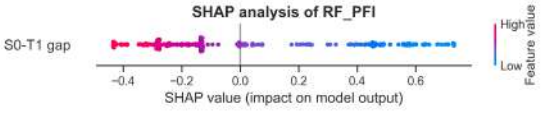
y distribution analysis

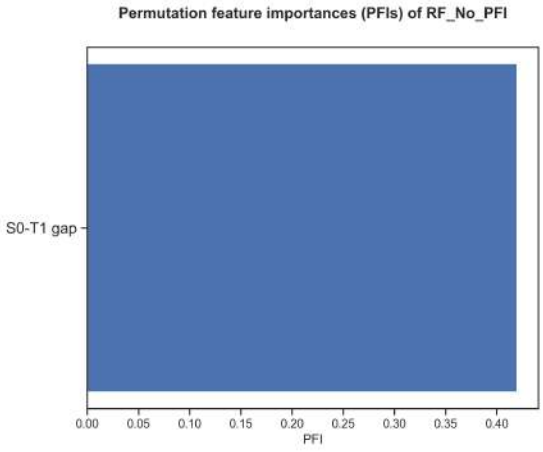
x WARNING! Your data is not uniform (Q4 has 22 points while Q1 has 148)

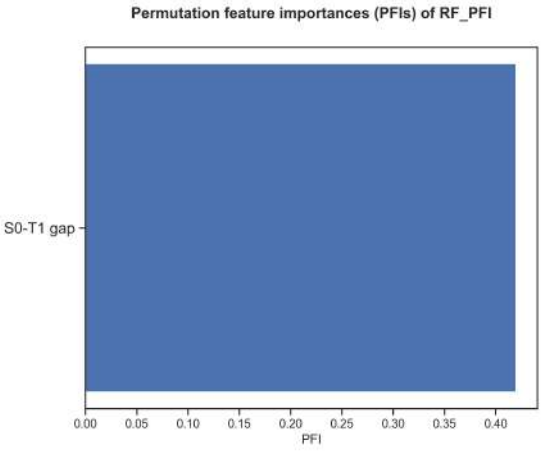
Section D. Feature Importances

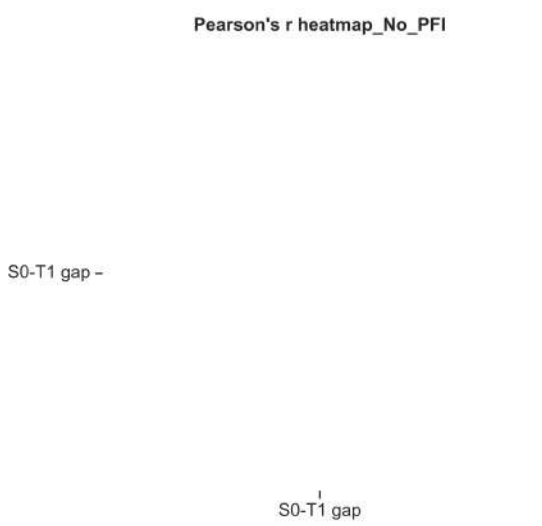
This section presents feature importances measured using the validation set.

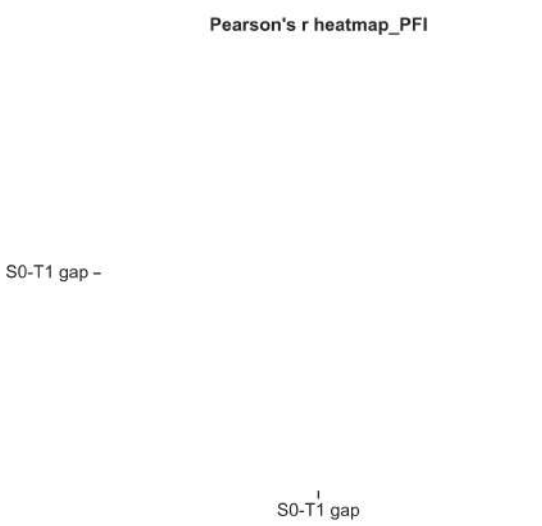












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

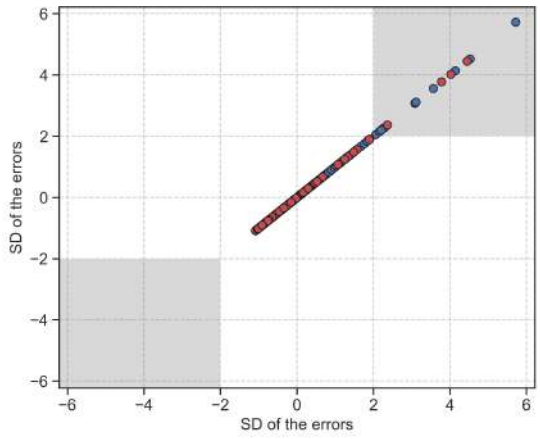
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (4.1 SDs)
- 885 (4.5 SDs)
- 1007 (5.7 SDs)
- 1130 (2.3 SDs)
- 2767 (2.1 SDs)
- 978 (2.2 SDs)
- 1126 (3.1 SDs)
- 207 (2.3 SDs)
- 49 (3.1 SDs)
- 1006 (2.2 SDs)

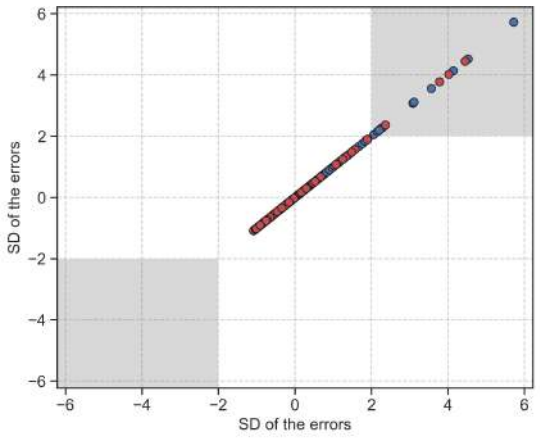
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.0 SDs)
- 884 (4.4 SDs)
- 890 (2.4 SDs)
- 1483 (3.8 SDs)



PFI (only important descriptors):

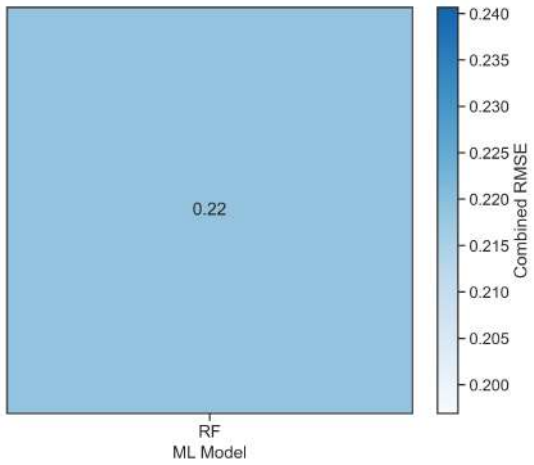
Outliers (max. 10 shown)
Train: 13 outliers out of 295 datapoints (4.4%)
- 2279 (4.1 SDs)
- 885 (4.5 SDs)
- 1007 (5.7 SDs)
- 1130 (2.3 SDs)
- 2767 (2.1 SDs)
- 978 (2.2 SDs)
- 1126 (3.1 SDs)
- 207 (2.3 SDs)
- 49 (3.1 SDs)
- 1006 (2.2 SDs)

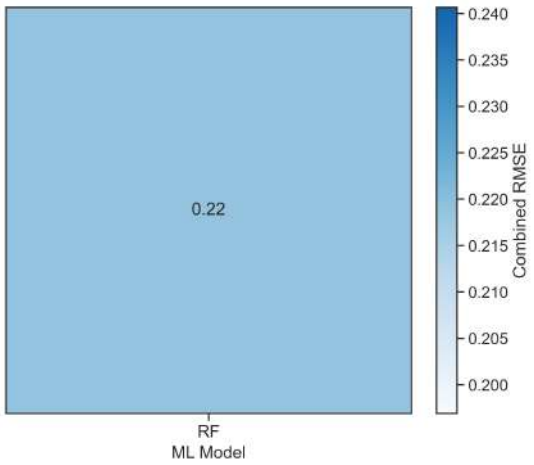
Test: 4 outliers out of 74 datapoints (5.4%)
- 888 (4.0 SDs)
- 884 (4.4 SDs)
- 890 (2.4 SDs)
- 1483 (3.8 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv)
- External test set (biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV databases:

python -m robert --csv_name "biphenylenes_gen_6_round_2_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --csv_test "biphenylenes_gen_6_round_2_test_set_QDESCP_interpret_descriptors.csv" --ignore "['HOMO-LUMO gap', 'HOMO', 'LUMO', 'IP', 'EA', 'Dipole module', 'Total charge', 'Global SASA', 'G solv. in H2O', 'G of H-bonds H2O', 'Fermi-level', 'Total polariz. alpha', 'Total FOD', 'Hardness', 'Softness', 'Electronegativity', 'Electrophil. idx', 'Nucleophilicity idx', 'Second IP', 'Second EA', 'MolLogP']" --model "['RF']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.17 using Windows 10.0.22631

Total execution time: 190.63 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
sklearn model: RandomForestRegressor	sklearn model: RandomForestRegressor
n_estimators: 97	n_estimators: 97
max_depth: 11	max_depth: 11
min_samples_split: 8	min_samples_split: 8
min_samples_leaf: 4	min_samples_leaf: 4
min_weight_fraction_leaf: 0.028402228054696617	min_weight_fraction_leaf: 0.028402228054696617
max_features: 0.9441974787194958	max_features: 0.9441974787194958
ccp_alpha: 0.0007103605819788694	ccp_alpha: 0.0007103605819788694
max_samples: 0.31534697477615553	max_samples: 0.31534697477615553
random_state: 0	random_state: 0

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):	**PFI (only important descriptors):**
type: reg	type: reg
kfold: 5	kfold: 5
repeat_kfolds: 10	repeat_kfolds: 10
seed: 0	seed: 0
error_type: rmse	error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Section J. New Predictions

Predictions of the external test set added with the csv_test option.

External test metrics

R2 = 0.86, MAE = 0.058, RMSE = 0.074

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/...No_PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.31 ± 0.02
1754	2.08	2.31 ± 0.02
1099	2.24	2.26 ± 0.02
997	2.12	2.25 ± 0.02
399	2.14	2.24 ± 0.02
2764	2.13	2.24 ± 0.01
1769	2.01	2.23 ± 0.02
3211	1.94	2.22 ± 0.02
2236	1.92	2.13 ± 0.02
1883	1.96	2.1 ± 0.02
...	...	...
3626	1.44	1.42 ± 0.01
3687	1.42	1.42 ± 0.01
855	1.37	1.42 ± 0.01
2007	1.45	1.42 ± 0.01
816	1.42	1.42 ± 0.01
3765	1.4	1.42 ± 0.01
3657	1.4	1.42 ± 0.01
788	1.3	1.42 ± 0.01
3503	1.29	1.42 ± 0.01
3779	1.26	1.42 ± 0.01

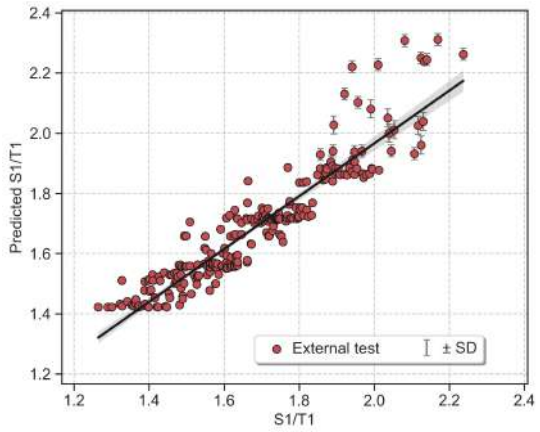
External test metrics

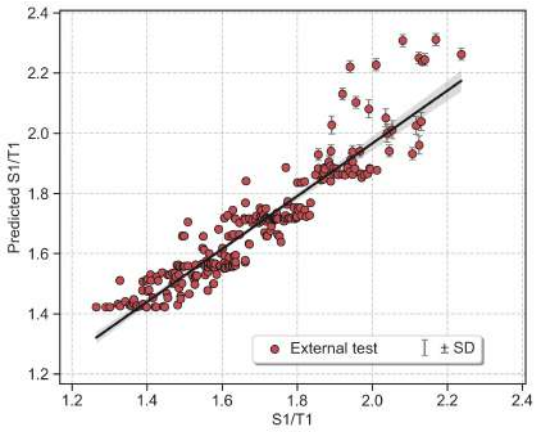
R2 = 0.86, MAE = 0.058, RMSE = 0.074

External test predictions (sorted, max. 20 shown)

From /PREDICT/csv_test/..._PFI.csv

code_name	S1/T1	S1/T1_pred ± sd
1853	2.17	2.31 ± 0.02
1754	2.08	2.31 ± 0.02
1099	2.24	2.26 ± 0.02
997	2.12	2.25 ± 0.02
399	2.14	2.24 ± 0.02
2764	2.13	2.24 ± 0.01
1769	2.01	2.23 ± 0.02
3211	1.94	2.22 ± 0.02
2236	1.92	2.13 ± 0.02
1883	1.96	2.1 ± 0.02
...	...	...
3626	1.44	1.42 ± 0.01
3687	1.42	1.42 ± 0.01
855	1.37	1.42 ± 0.01
2007	1.45	1.42 ± 0.01
816	1.42	1.42 ± 0.01
3765	1.4	1.42 ± 0.01
3657	1.4	1.42 ± 0.01
788	1.3	1.42 ± 0.01
3503	1.29	1.42 ± 0.01
3779	1.26	1.42 ± 0.01


Predicted S1/T1
S1/T1
External test
± SD


Predicted S1/T1
S1/T1
External test
± SD

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.

ROBERT v 2.0.2 2026/01/08 10:23:26

How to cite: Dalmau, D.; Alegre Requena, J. V. WIREs Comput Mol Sci. 2024, 14, e1733.

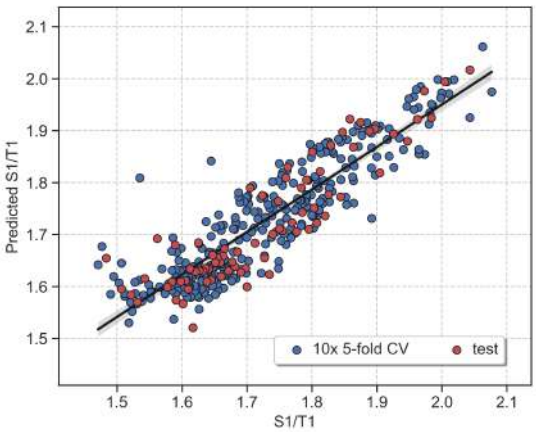
Section A. ROBERT Score

This score is designed to evaluate the models using different metrics.

No PFI (standard descriptor filter) · Score 9

Model = NN · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 312:3



10x 5-fold CV : R^2 = 0.81, MAE = 0.042, RMSE = 0.055

Test : R^2 = 0.82, MAE = 0.043, RMSE = 0.054

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Potential "faulty" outliers (Section E)

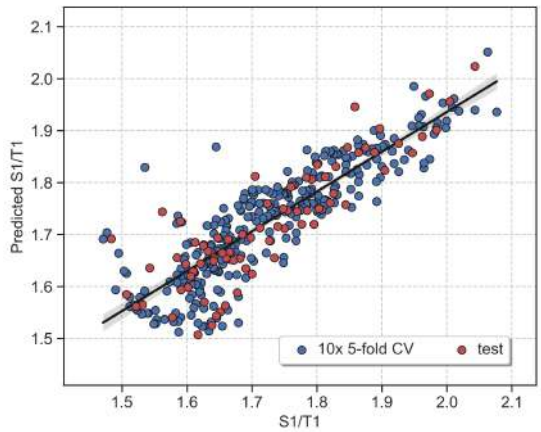
Overall assessment

The model seems reliable

PFI (only important descriptors) · Score 7

Model = MVL · CV (train+valid.):Test = 80:20

Points(train+validation):descriptors = 312:1



10x 5-fold CV : R^2 = 0.76, MAE = 0.046, RMSE = 0.061

Test : R^2 = 0.75, MAE = 0.048, RMSE = 0.063

Severe warnings

No severe warnings detected

Moderate warnings

Uneven y distribution (Section C)

Overall assessment

Decent model, but it has limitations

Section B. Advanced Score Analysis

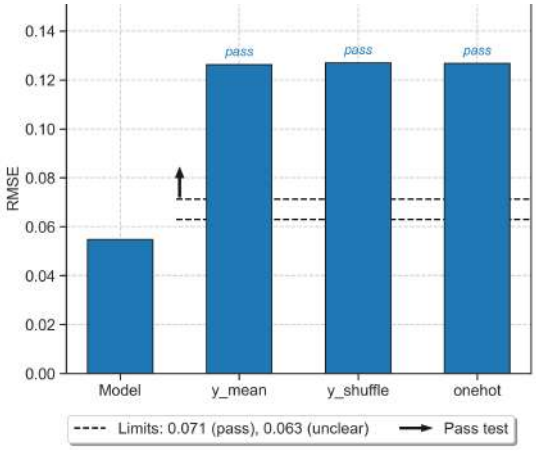
This section explains each component that comprises the ROBERT score. More details here.

1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.

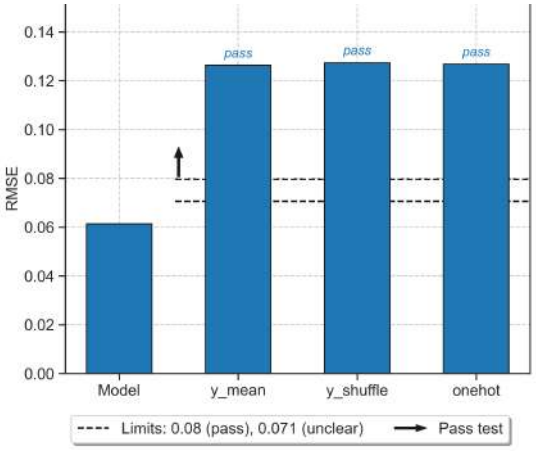


1. Model vs "flawed" models (0 / 0)

The model predicts right for the right reasons.

· Scoring from -6 to 0 ·

Pass: 0, Unclear: -1, Fail: -2.



2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 9.02%.

R^2 (10x 5-fold CV) = 0.81.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

2. CV predictions of the model (2 / 2)

Scaled RMSE (10x 5-fold CV) = 10.0%.

R^2 (10x 5-fold CV) = 0.76.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3. Predictive ability & overfitting

3a. Predictions test set (2 / 2)

Scaled RMSE (test set) = 8.85%.

R^2 (test set) = 0.82.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 0.98*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3. Predictive ability & overfitting

3a. Predictions test set (1 / 2)

Scaled RMSE (test set) = 10.33%.

R^2 (test set) = 0.75.

· Scoring from 0 to 2 ·

Scaled RMSE ≤ 10%: +2, Scaled RMSE ≤ 20%: +1.

R^2 < 0.5: -2, R^2 < 0.7: -1

3b. Prediction accuracy test vs CV (2 / 2)

Relative differences in values from sections 2 and 3a.

RMSE in test is 1.03*scaled RMSE (CV).

· Scoring from 0 to 2 ·

Scaled RMSE (test) ≤ 1.25*scaled RMSE (CV): +2.

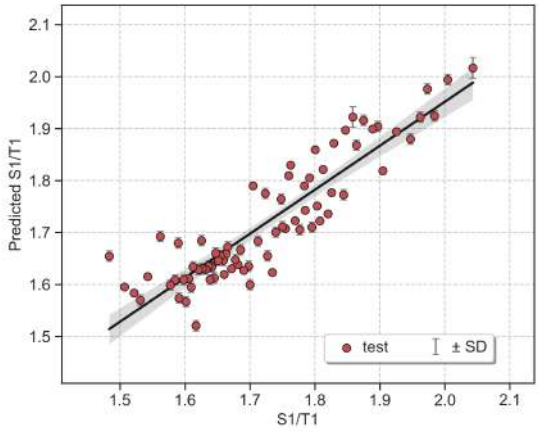
Scaled RMSE (test) ≤ 1.50*scaled RMSE (CV): +1.

3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (4% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.

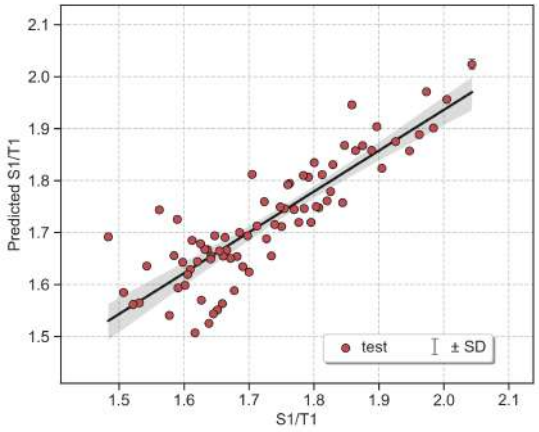


3c. Avg. standard deviation (SD) (2 / 2)

Low variation, 4*SD = 0.0 (0% y-range).

· Scoring from 0 to 2 ·

4*SD ≤ 25% y-range: +2, 4*SD ≤ 50% y-range: +1.



3d. Extrapolation (sorted CV) (1 / 2)

Scaled RMSEs across 5-fold CV:

[14.75%, 8.2%, 9.84%, 9.84%, 16.39%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

3d. Extrapolation (sorted CV) (0 / 2)

Scaled RMSEs across 5-fold CV:

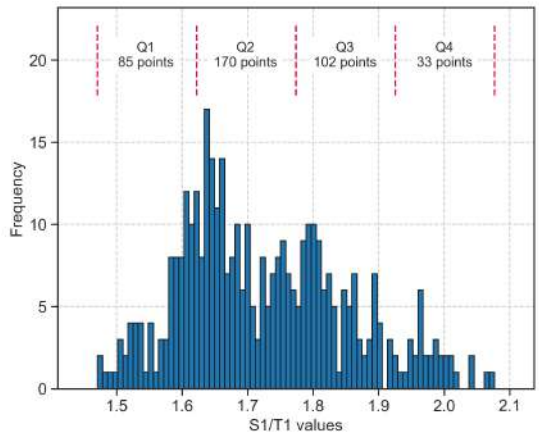
[14.75%, 11.48%, 9.84%, 6.56%, 18.03%]

· Scoring from 0 to 2 ·

Every two folds with RMSEs ≤ 1.25*min RMSE: +1.

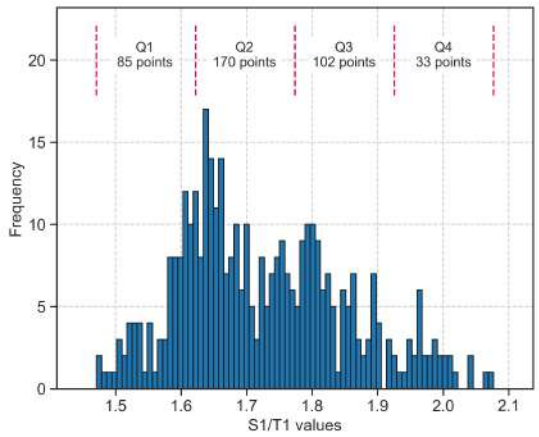
Section C. Distribution of y Values

This section shows the distribution of y values within the training and validation sets.



y distribution analysis

x WARNING! Your data is not uniform (Q4 has 33 points while Q2 has 170)

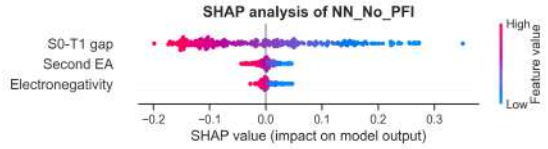


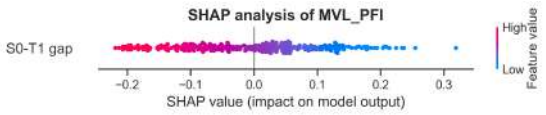
y distribution analysis

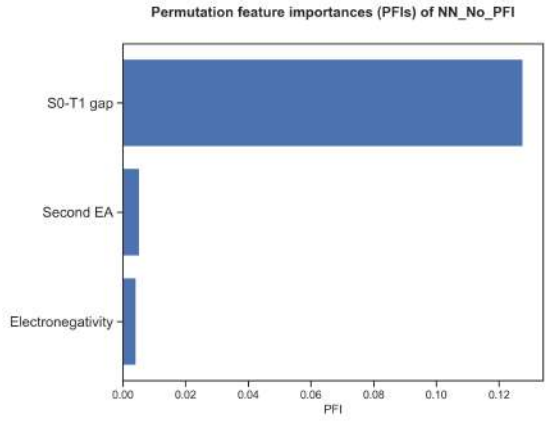
x WARNING! Your data is not uniform (Q4 has 33 points while Q2 has 170)

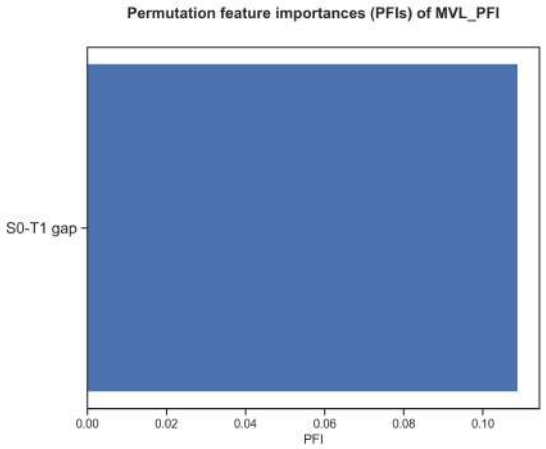
Section D. Feature Importances

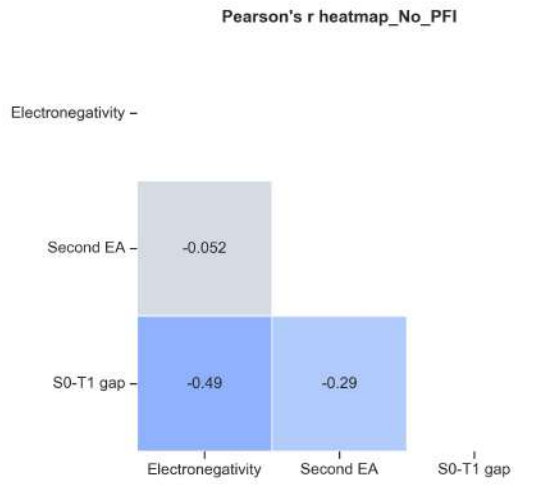
This section presents feature importances measured using the validation set.

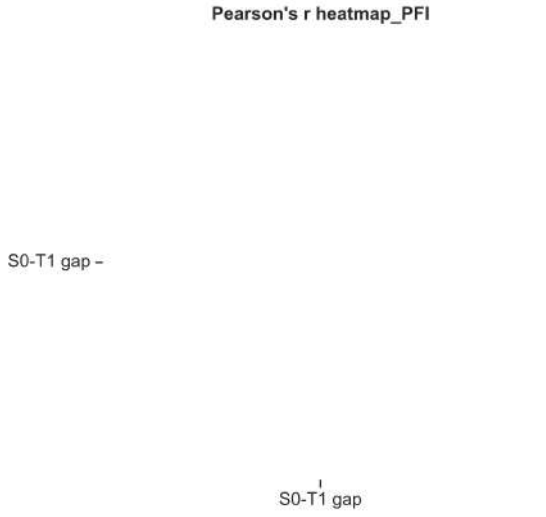












Correlation analysis

o Correlations between variables are acceptable

Correlation analysis

o Correlations between variables are acceptable

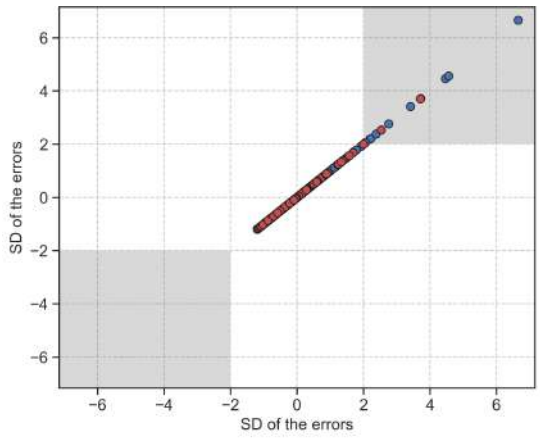
Section E. Outlier Analysis

This section detects outliers using the standard deviation (SD) of errors from the training set.

No PFI (standard descriptor filter):

Outliers (max. 10 shown)
Train: 10 outliers out of 312 datapoints (3.2%)
- 2947 (2.2 SDs)
- 2929 (2.2 SDs)
- 1448 (3.4 SDs)
- 1289 (2.1 SDs)
- 7224 (4.5 SDs)
- 5876 (6.7 SDs)
- 6237 (2.8 SDs)
- 1412 (2.4 SDs)
- 310 (4.6 SDs)
- 6693 (3.7 SDs)

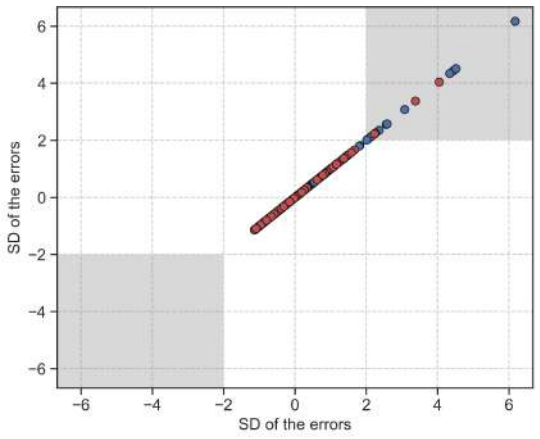
Test: 3 outliers out of 78 datapoints (3.8%)
- 5904 (2.0 SDs)
- 7104 (2.5 SDs)
- 6413 (3.7 SDs)



PFI (only important descriptors):

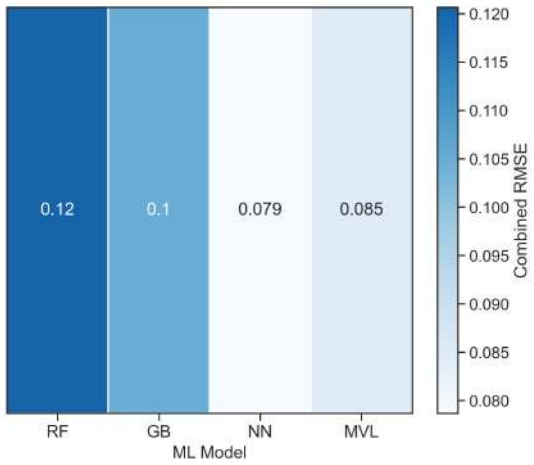
Outliers (max. 10 shown)
Train: 16 outliers out of 312 datapoints (5.1%)
- 1838 (2.4 SDs)
- 2929 (2.0 SDs)
- 612 (2.2 SDs)
- 609 (2.0 SDs)
- 1448 (2.0 SDs)
- 5531 (2.6 SDs)
- 5575 (2.1 SDs)
- 7224 (4.4 SDs)
- 5259 (2.1 SDs)
- 7159 (2.6 SDs)

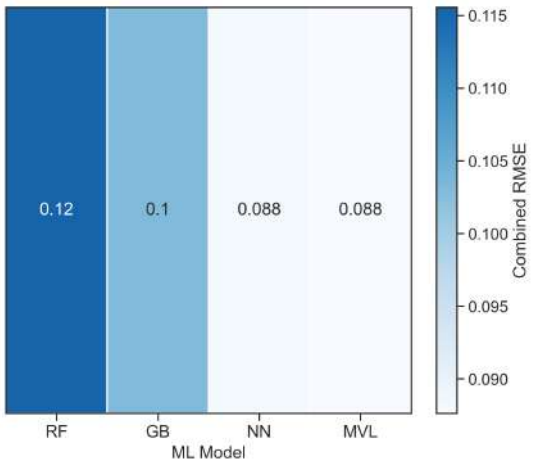
Test: 3 outliers out of 78 datapoints (3.8%)
- 7535 (2.2 SDs)
- 7104 (3.4 SDs)
- 6413 (4.0 SDs)



Section F. Model Screening

This section compares different combinations of hyperoptimized algorithms and partition sizes. The combined error is calculated as the product of the training error, validation error, and cross-validation error.





Section G. Reproducibility

This section provides all the instructions to reproduce the results presented.

1. Download these files *(the authors should have uploaded the files as supporting information!)*:

- CSV database (dbptrans_gen_6_QDESCP_interpret_descriptors.csv)

2. Install and adjust the versions of the following Python modules:

- Install ROBERT and its dependencies: conda install -y -c conda-forge robert
- Adjust ROBERT version: pip install robert==2.0.2
- scikit-learn-intelex: not installed

(if scikit-learn-intelex is installed, slightly different results might be obtained)

3. Run ROBERT using this command line in the folder with the CSV database:

python -m robert --csv_name "dbptrans_gen_6_QDESCP_interpret_descriptors.csv" --y "S1/T1" --names "code_name" --ignore "['Dipole module', 'G of H-bonds H2O', 'MolLogP']"

4. Execution time, Python version and OS:

Originally run in Python 3.10.18 using Windows 10.0.26200

Total execution time: 289.5 seconds *(the number of processors should be specified by the user)*

Section H. Transparency

This section contains important parameters used in scikit-learn models and ROBERT.

1. Parameters of the scikit-learn models (same keywords as used in scikit-learn):

No PFI (standard descriptor filter):

sklearn model: MLPRegressor
hidden_layer_1: 2
hidden_layer_2: 0
max_iter: 450
alpha: 0.08003410758548654
tol: 8.830109334221373e-05
random_state: 0
solver: lbfgs

PFI (only important descriptors):

sklearn model: LinearRegression

2. ROBERT options, including prediction type (REG or CLAS), folds and repeats used for CV, etc:

No PFI (standard descriptor filter):

type: reg
kfold: 5
repeat_kfolds: 10
seed: 0
error_type: rmse

PFI (only important descriptors):

type: reg
kfold: 5
repeat_kfolds: 10
seed: 0
error_type: rmse

Section I. Abbreviations

Reference section for the abbreviations used.

ACC: accuracy
ADAB: AdaBoost
CSV: comma separated values
CLAS: classification
CV: cross-validation
F1 score: balanced F-score
GB: gradient boosting
GP: gaussian process
KN: k-nearest neighbors
MAE: root-mean-square error
MCC: Matthew's correl. coefficient
ML: machine learning
MVL: multivariate lineal models
NN: neural network
PFI: permutation feature importance
R2: coefficient of determination
REG: Regression
RF: random forest
RMSE: root mean square error
RND: random
SHAP: Shapley additive explanations
VR: voting regressor

Miscellaneous

General tips to improve the models and instructions to predict new values.

Some general tips to improve the score

1. Adding meaningful datapoints might help to improve the model. Also, using a uniform population of datapoints across the whole range of y values usually helps to obtain reliable predictions across the whole range. More information about the range of y values used is available in Section C.

2. Adding meaningful descriptors or replacing/deleting the least useful descriptors used might help. Feature importances are gathered in Section D.

How to predict new values with these models?

1. Create a CSV database with the new points, including the necessary descriptors.

2. Place the CSV file in the parent folder (i.e., where the module folders were created)

3. Run the PREDICT module as 'python -m robert --predict --csv_test FILENAME.csv'.

4. The predictions will be shown at the end of the resulting PDF report and will be stored in the last column of two CSV files called MODEL_SIZE_test(_No)_PFI.csv, which are in the PREDICT folder.
