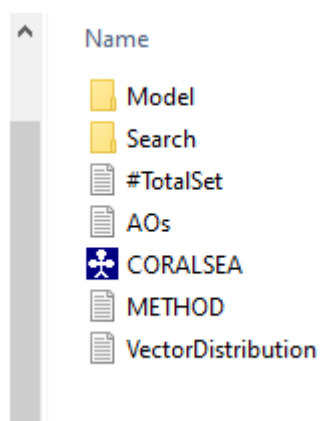
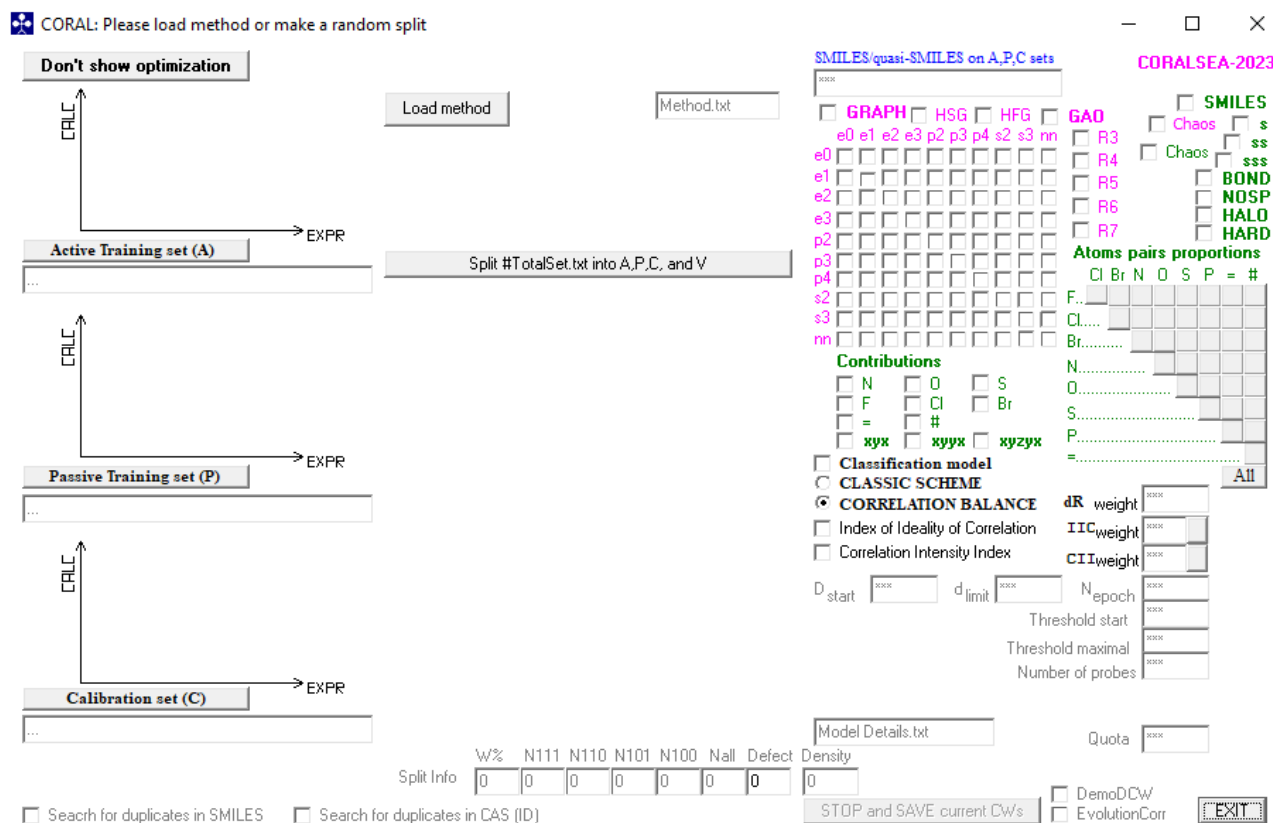


TOY EXPERIMENT WITH CORAL



Click (run) CORALSEA.exe



File #TotalSet.txt contains toy dataset for simple experiments.

Click Split #TotalSet.txt into A,P,C, and V

CORAL: Please load method or make a random split

Don't show optimization

Load method

Method.txt

P.C. and V

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

	Input	Output
1.	#TotalSet.txt	Load
2.	Do distribution	
3.	Save files	

Dispersion Limit = 0.01

	Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.25	0
Invisible training set	-	0.25	0
Calibration set	#	0.25	0
#ValidationSet.txt	*	0.25	0

Loading of vector of Distribution

Save vector of Distribution

EXIT

SMILES/quasi-SMILES on A,P,C sets

CORALSEA-2023

GRAPH HSG HFG GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

R3 R4 R5 R6 R7

SMILES Chaos s ss sss BOND NOSP HALO HARD

Atoms pairs proportions Cl Br N O S P = #

Contributions

N O S F Cl Br = # xyx xyxx xyzyx

Classification model CLASSIC SCHEME CORRELATION BALANCE

Index of Ideality of Correlation

Correlation Intensity Index

d start d limit Nepoch

Threshold start

Threshold maximal

Number of probes

Model Details.txt

Quota

STDP and SAVE current CW's

DemoDCW EvolutionCorr

EXIT

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
0	0	0	0	0	0	0	0

Search for duplicates in SMILES

Search for duplicates in CAS (ID)

Click Load

CORAL: Please load method or make a random split

Don't show optimization

Load method

Method.txt

P.C. and V

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

	Input	Output
+1.	#TotalSet.txt	Load
2.	Do distribution	
3.	Save files	

Dispersion Limit = 0.010

	Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	0
Invisible training set	-	0.250	0
Calibration set	#	0.250	0
#ValidationSet.txt	*	0.250	0

Loading of vector of Distribution

Save vector of Distribution

EXIT

SMILES/quasi-SMILES on A,P,C sets

CORALSEA-2023

GRAPH HSG HFG GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

R3 R4 R5 R6 R7

SMILES Chaos s ss sss BOND NOSP HALO HARD

Atoms pairs proportions Cl Br N O S P = #

Contributions

N O S F Cl Br = # xyx xyxx xyzyx

Classification model CLASSIC SCHEME CORRELATION BALANCE

Index of Ideality of Correlation

Correlation Intensity Index

d start d limit Nepoch

Threshold start

Threshold maximal

Number of probes

Model Details.txt

Quota

STDP and SAVE current CW's

DemoDCW EvolutionCorr

EXIT

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
0	0	0	0	0	0	0	0

Search for duplicates in SMILES

Search for duplicates in CAS (ID)

The screenshot displays the CORAL software interface during the preparation of a dataset split. A central dialog box titled "Preparation of split April 13, 2022" is open, detailing the steps and parameters for creating training and validation sets.

Dialog Box Content:

- Title Bar:** Preparation of split April 13, 2022
- Instructions:**
 - The first action is "Load". File "#TotalSet.txt" must exist in your folder.
 - The second action should be "Do distribution".
 - The third action should be "Save files".
- Action Steps:**
 - +1.** Input: #TotalSet.txt | Load | Output: #TotalSet.txt | N = 295
 - +2.** Do distribution
 - 3.** Save files
- Dispersion Limit:** 0.010
- Distribution Table:**

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	72	0.244
Invisible training set	-	0.250	74	0.251
Calibration set	#	0.250	74	0.251
#ValidationSet.txt	*	0.250	75	0.254
- Buttons:** Loading of vector of Distribution, Save vector of Distribution, EXIT

Main Interface Elements:

- Top Menu:** CORAL: Please load method or make a random split
- Left Panel:** Contains buttons like "Don't show optimization", "CALC", "Act...", "Pas...", "CALC", "CALC".
- Right Panel:** Displays various settings and results:
 - Method.txt:** Method selection area.
 - P.C. and V:** Parameters for correlation and variance.
 - SMILES/quasi-SMILES on A,P,C sets:** Checkboxes for GRAPH, HSG, HFG, GAO, R3, R4, R5, R6, R7, nm, e0, e1, e2, e3, p2, p3, p4, s2, s3.
 - CORALSEA-2023:** Checkboxes for SMILES, Chaos, BOND, NOSP, HALO, HARD.
 - Atoms pairs proportions:** Cl Br N O S P = #
 - Contributions:** Checkboxes for N, O, S, F, Cl, Br, xyx, xyxx, xyzy.
 - Classification model:** Radio buttons for CLASSIC SCHEME (selected), CORRELATION BALANCE, Index of Ideality of Correlation, Correlation Intensity Index.
 - Weights:** dr weight, IIC weight, CII weight.
 - Epochs:** Threshold start, Threshold maximal, Number of probes.
 - Model Details.txt:** Model details section.
 - Quota:** Quota setting.
 - Bottom Status Bar:** Search for duplicates in SMILES, Search for duplicates in CAS (ID), STOP and SAVE current CWs, DemoDCW/EvolutionCorr, EXIT.

The screenshot displays the CORAL software interface. A dialog box titled "Preparation of split April 13, 2022" is open, showing instructions and a table of distribution parameters. The main application window is visible in the background, showing various settings and a table of atom pairs proportions.

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.
 The second action should be "Do distribution".
 The third action should be "Save files".

	Input	Output	
+1.	#TotalSet.txt	Load	#TotalSet'.txt N = 295
+2.	Do distribution		
+3.	Save files		

Dispersion Limit = 0.010

	Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	72 0.244
Invisible training set	-	0.250	74 0.251
Calibration set	#	0.250	74 0.251
#ValidationSet.txt	*	0.250	75 0.254

Buttons: Loading of vector of Distribution, Save vector of Distribution, EXIT

CORAL: Please load method or make a random split

Buttons: Don't show optimization, Load method, Method.txt

SMILES/quasi-SMILES on A,P,C sets

Buttons: ☐ GRAPH, ☐ HSG, ☐ HFG, ☐ GAO, ☐ SMILES, ☐ Chaos, ☐ s, ☐ ss, ☐ sss, ☐ BOND, ☐ NOSP, ☐ HALO, ☐ HARD

Atoms pairs proportions

Cl Br N O S P = #

Contributions

☐ N, ☐ O, ☐ S, ☐ F, ☐ Cl, ☐ Br, ☐ =, ☐ #, ☐ yx, ☐ yxyx, ☐ xyzyx

Classification model

☐ CLASSIC SCHEME, ☒ CORRELATION BALANCE

☐ Index of Ideality of Correlation, ☐ Correlation Intensity Index

dR weight, **IIC weight**, **CII weight**, **d start**, **d limit**, **N epoch**, **Threshold start**, **Threshold maximal**, **Number of probes**

Model Details.txt, **Quota**

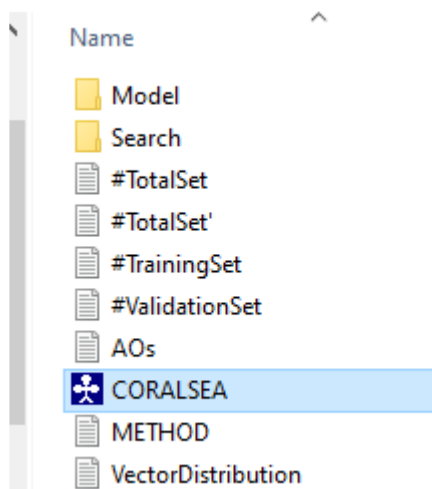
W%, **N111**, **N110**, **N101**, **N100**, **Nall**, **Defect**, **Density**

Split Info

☐ Search for duplicates in SMILES, ☐ Search for duplicates in CAS (ID)

STOP and SAVE current CWs, **DemoDCW**, **EvolutionCorr**, **EXIT**

You have obtained in the folder the following files



The TrainingSet.txt contains Active training set; Passive training set; Calibration set; and Validation set.

Each the above subset has a defined task. The task for the active training set is to calculate correlation weights, which give as large as possible target function (in the first approximation it is a correlation) between experimental and predicted endpoint for the active training set. The task for the passive training set is inspection: whether these data give reasonable correlation coefficient for the similar compounds in the passive training set. The task of the calibration set is to detect the overtraining. The task for the validation set is the final estimation of the predictive potential of the model.

Now, click Load method

CORAL: select Phase (1 or 2); or change and save method; or import current model

☐ Don't show optimization

Active Training set (A) → EXPR

Passive Training set (P) → EXPR

Calibration set (C) → EXPR

Method.txt (Load method | Save method)

Split #TotalSet.txt into A,P,C, and V

Building up a model $E = C_0 + C_1 * DCW(Threshold, Nepochs)$

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nm

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☐ Chaos ☐ \$ ☐ ss ☐ sss

☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0
#	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model ☒ CLASSIC SCHEME

☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation

☒ Correlation Intensity Index

Model Parameters:

D_{start} 0.5 d_{limit} 0.1 Nepoch 15

Threshold start 5

Threshold maximal 5

Number of probes 1

Model Details.txt

Quota 15

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
0	0	0	0	0	0	0	0

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

☒ DemoDCW ☒ EvolutionCorr

STOP and SAVE current CWs **EXIT**

If you want change method do it but after click Save method.

In order to build up a model click Building up a model $E=C_0+C_1...$

CORAL: Building up preferable model

☐ Don't show optimization

Active Training set (A) → EXPR

Passive Training set (P) → EXPR

Calibration set (C) → EXPR

Building up a model $E=C_0+C_1*DCW(Threshold,Nepochs)$

Define preferable threshold and press Continue Continue

$C_0 =$ $C_1 =$

Import of current model

Model Details.txt

Quota

STOP and SAVE current CWs

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

☒ DemoDCW ☒ EvolutionCorr

EXIT

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

☐ e0 ☐ e1 ☐ e2 ☐ e3 ☐ p2 ☐ p3 ☐ p4 ☐ s2 ☐ s3 ☐ nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☐ Chaos ☐ Chaos

☒ s ☐ ss ☐ sss

☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0
#	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyyx ☐ xyzyx

☐ Classification model ☒ CLASSIC SCHEME ☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation ☒ Correlation Intensity Index

D_start d_limit N_epoch

Threshold start Threshold maximal Number of probes

dR weight IIC_clb CII_clb

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
0	0	0	0	0	0	0	0

Insert Threshold value (it can be 1, or 3, or 7, ... max=29) and click Continue

CORAL: Building up preferable model

☐ Don't show optimization

Active Training set (A) → EXPR

Passive Training set (P) → EXPR

Calibration set (C) → EXPR

Building up a model $E=C_0+C_1*DCW(Threshold,Nepochs)$

Define preferable threshold and press Continue Continue

$C_0 =$ $C_1 =$

Import of current model

Model Details.txt

Quota

STOP and SAVE current CWs

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

☒ DemoDCW ☒ EvolutionCorr

EXIT

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

☐ e0 ☐ e1 ☐ e2 ☐ e3 ☐ p2 ☐ p3 ☐ p4 ☐ s2 ☐ s3 ☐ nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☐ Chaos ☐ Chaos

☒ s ☐ ss ☐ sss

☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0
#	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyyx ☐ xyzyx

☐ Classification model ☒ CLASSIC SCHEME ☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation ☒ Correlation Intensity Index

D_start d_limit N_epoch

Threshold start Threshold maximal Number of probes

dR weight IIC_clb CII_clb

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
0	0	0	0	0	0	0	0

Confirm

There are files in "model/*.*" which remain after previous calculations you can delete these files in order to avoid mixture of new files and files which remain after previous calculations. Delete these files?

STOP and SAVE current CWs

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

☒ DemoDCW ☒ EvolutionCorr

EXIT

Click Yes (Click Don't show optimization if you want get result fastly).

CORAL: you should save model now

Don't show optimization

Active Training set (A)
 n=76: R2=0.6792: s=1.06: MAE=0.852: F=157

Passive Training set (P)
 n=73: R2=0.6584: s=1.21: MAE=0.966: F=137

Calibration set (C)
 n=71: R2=0.6510: s=1.16: MAE=0.912: F=129

Selected threshold is 5
 THE CALCULATION IS COMPLETED

Building up a model $E = C0 + C1 * DCW(Threshold, Nepochs)$

Define preferable threshold and press Continue Continue

C0 = C1 =

Insert a SMILES for calculation of DCW and EndPoint

Start of DCW and Endpoint calculation for SMILES from file

Save Model

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
100	38	0	0	0	38	55.10	0.409

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

STOP and SAVE current Cw/s

☒ DemoDCW ☒ EvolutionCorr

EXIT

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☒ GAO

e0 ☐ e1 ☐ e2 ☐ e3 ☐ p2 ☐ p3 ☐ p4 ☐ s2 ☐ s3 ☐ nn

R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☐ Chaos ☒ ss ☐ sss ☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F.....	0	0	0	0	0	0	0	0
Cl.....	0	0	0	0	0	0	0	0
Br.....	0	0	0	0	0	0	0	0
N.....	0	0	0	0	0	0	0	0
O.....	0	0	0	0	0	0	0	0
S.....	0	0	0	0	0	0	0	0
P.....	0	0	0	0	0	0	0	0
=.....	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model ☒ CLASSIC SCHEME ☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation ☒ Correlation Intensity Index

D_start d_limit N_epoch

DemoDCW.txt Threshold start

Threshold maximal

Number of probes

#ValidationSet.txt #ModelForValidationSet.txt

Model Details.txt Quota

Click "Save Model"

In order to check status model for validation set click CORALSEA.exe again

Click Load Method -> click “Import of current model”

CORAL: Calculation of model for external substances

Don't show optimization

Active Training set (A)
n=76: R2=0.6792: s=1.06: MAE=0.852: F=157

Passive Training set (P)
n=73: R2=0.6584: s=1.21: MAE=0.966: F=137

Calibration set (C)
n=71: R2=0.6510: s=1.16: MAE=0.912: F=129

DEMO plots "EXPR vs. CALC" EACH

Split #TotalSet.txt into A,P,C, and V
Building up a model $E = C_0 + C_1 * DCW(Threshold, Nepochs)$

C0 = -4.2255847 C1 = 0.1601975

Insert a SMILES for calculation of DCW and EndPoint

Demo of calculation of DCW and endpoint will be in file

DCW(5,15)= EndPoint =

Start of DCW and Endpoint calculation for SMILES from file

Import of current model

W% N111 N110 N101 N100 Nall Defect Density
Split Info 100 38 0 0 0 38 55.10 0.409

Search for duplicates in SMILES Search for duplicates in CAS (ID) Continue optimization STOP and SAVE current CWs DemoDCW EvolutionCorr EXIT

SMILES/quasi-SMILES on A,P,C sets
#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO
e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES
☐ Chaos ☒ s
☐ Chaos ☒ ss
☐ BOND
☐ NOSP
☐ HALO
☐ HARD

Atoms pairs proportions
Cl Br N O S P = #
F 0 0 0 0 0 0
Cl 0 0 0 0 0 0
Br 0 0 0 0 0 0
N 0 0 0 0 0 0
O 0 0 0 0 0 0
S 0 0 0 0 0 0
P 0 0 0 0 0 0

Contributions
☐ N ☐ O ☐ S
☐ F ☐ Cl ☐ Br
☐ = ☐ #
☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model
☒ CLASSIC SCHEME
☒ CORRELATION BALANCE
☒ Index of Ideality of Correlation
☒ Correlation Intensity Index

dR weight 0.1
IICclb 0.3 0
CIIclb 0.3 0

D_start 0.5 d_limit 0.1 N_epoch 15
Threshold start 5
Threshold maximal 5
Number of probes 1
#ValidationSet.txt #ModelForValidationSet.txt
Model Details.txt Quota 15

In order to consider model for validation set click “Start of DCW and Endpoint...”

CORAL: Calculation of model for external substances

Don't show optimization

Active Training set (A)
n=76: R2=0.6792: s=1.06: MAE=0.852: F=157

Passive Training set (P)
n=73: R2=0.6584: s=1.21: MAE=0.966: F=137

Calibration set (C)
n=71: R2=0.6510: s=1.16: MAE=0.912: F=129

DEMO plots "EXPR vs. CALC" EACH

Split #TotalSet.txt into A,P,C, and V
Building up a model $E = C_0 + C_1 * DCW(Threshold, Nepochs)$

C0 = -4.2255847 C1 = 0.1601975

Insert a SMILES for calculation of DCW and EndPoint

Demo of calculation of DCW and endpoint will be in file

DCW(5,15)= EndPoint =

Start of DCW and Endpoint calculation for SMILES from file

Import of current model

W% N111 N110 N101 N100 Nall Defect Density
Split Info 100 38 0 0 0 38 55.10 0.409

Search for duplicates in SMILES Search for duplicates in CAS (ID) Continue optimization STOP and SAVE current CWs DemoDCW EvolutionCorr EXIT

SMILES/quasi-SMILES on A,P,C sets
#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO
e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES
☐ Chaos ☒ s
☐ Chaos ☒ ss
☐ BOND
☐ NOSP
☐ HALO
☐ HARD

Atoms pairs proportions
Cl Br N O S P = #
F 0 0 0 0 0 0
Cl 0 0 0 0 0 0
Br 0 0 0 0 0 0
N 0 0 0 0 0 0
O 0 0 0 0 0 0
S 0 0 0 0 0 0
P 0 0 0 0 0 0

Contributions
☐ N ☐ O ☐ S
☐ F ☐ Cl ☐ Br
☐ = ☐ #
☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model
☒ CLASSIC SCHEME
☒ CORRELATION BALANCE
☒ Index of Ideality of Correlation
☒ Correlation Intensity Index

dR weight 0.1
IICclb 0.3 0
CIIclb 0.3 0

D_start 0.5 d_limit 0.1 N_epoch 15
Threshold start 5
Threshold maximal 5
Number of probes 1
#ValidationSet.txt #ModelForValidationSet.txt
Model Details.txt Quota 15

Coralsea

DCW/EndPoint calculation for data from #ValidationSet.txt is completed.

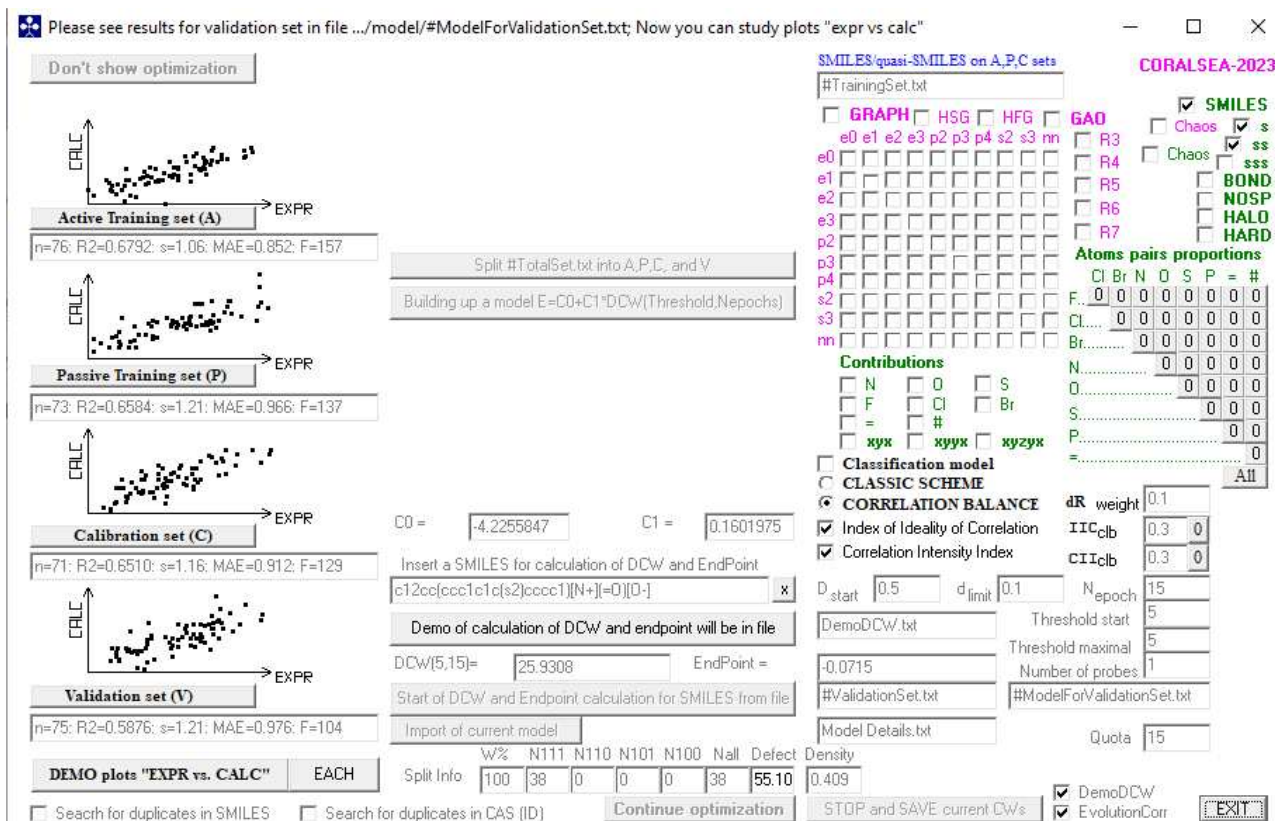
The numerical data is saved in model/#ModelForValidationSet.txt.

Next picture will be graphical representation of the model for Active training set; Passive training set; Calibration set; and Validation set.

File "DOMAIN.txt" contains the Domain of applicability, i.e. data without outliers which are defined according to the statistical defects calculated for SMILES or quasi-SMILES applied to build up the given model; file "DOMAIN0.txt" contains outliers

OK

Click OK



In order to study obtained results read files

W5-1.txt (correlation weights)

M5-1.txt (detail of model for active training set (+), passive training set (-), and calibration set (#).

#ModelForValidationSet\$.txt contains some details on model for validation set.

Now you can repeat these calculation for you data inserted to #MyTotalSet.txt.

Also, you can do copy in folder "MyCORALSEA" and carry out different experiments vary options.

Description of routine detail you can find in ReadMe.pdf as well as in publications (see . <http://www.insilico.eu/coral/PUBLICATIONS.html>).