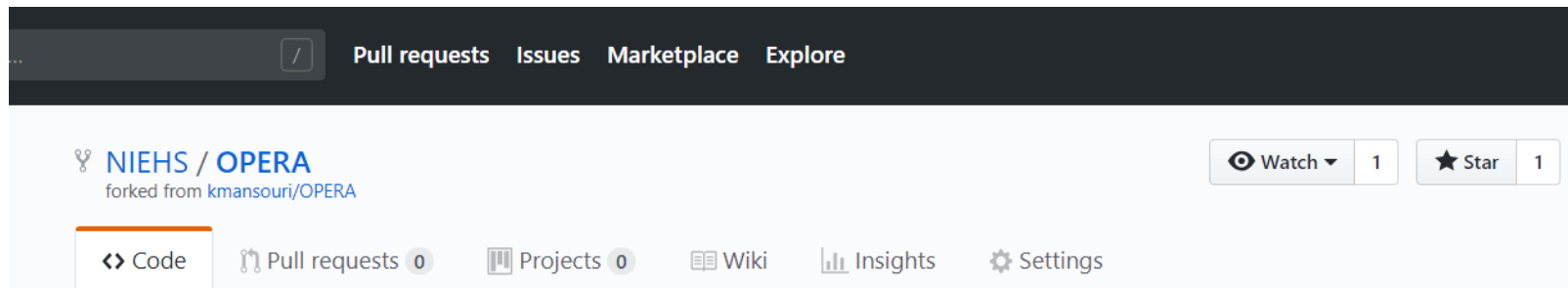




Windows version installation and quick run guide
Command line & GUI

OPERA Github repo:



The screenshot shows the top navigation bar of a Github repository. It includes links for 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. Below this, the repository name 'NIEHS / OPERA' is displayed, along with a note 'forked from kmansouri/OPERA'. On the right, there are buttons for 'Watch' (1) and 'Star' (1). A secondary navigation bar contains links for 'Code', 'Pull requests' (0), 'Projects' (0), 'Wiki', 'Insights', and 'Settings'.

Free and open-source application (command line and GUI) providing QSAR models predictions as well as applicability domain accuracy assessment for physicochemical properties, environmental fate and toxicological endpoints.

=====>Download the latest compiled version from the "releases" tab and run the executable installer.

[Manage topics](#)









A horizontal bar displaying repository statistics: '134 commits', '1 branch', '12 releases' (circled in red), and '1 contributor'. A blue arrow points from the text 'Go to the releases to download the latest version' to the '12 releases' link.



A row of buttons for repository actions: 'Branch: master' (dropdown), 'New pull request', 'Create new file', 'Upload files', 'Find File', and a green 'Clone' button.

This branch is 11 commits ahead of kmansouri:master. [Pull request](#)

 kmansouri Merge pull request #9 from kmansouri/master ... Latest commit d234

 Icon.png	OPERA 1.2 icon	
 LICENSE	Initial commit	
 Logo.png	Added logo and icon	
 OPERA1.5_Source_code.zip	MATLAB source code for OPERA1.5	3
 OPERA2.0_Source_code.zip	MATLAB source code for OPERA 2.0	2

Go to the releases to download the latest version

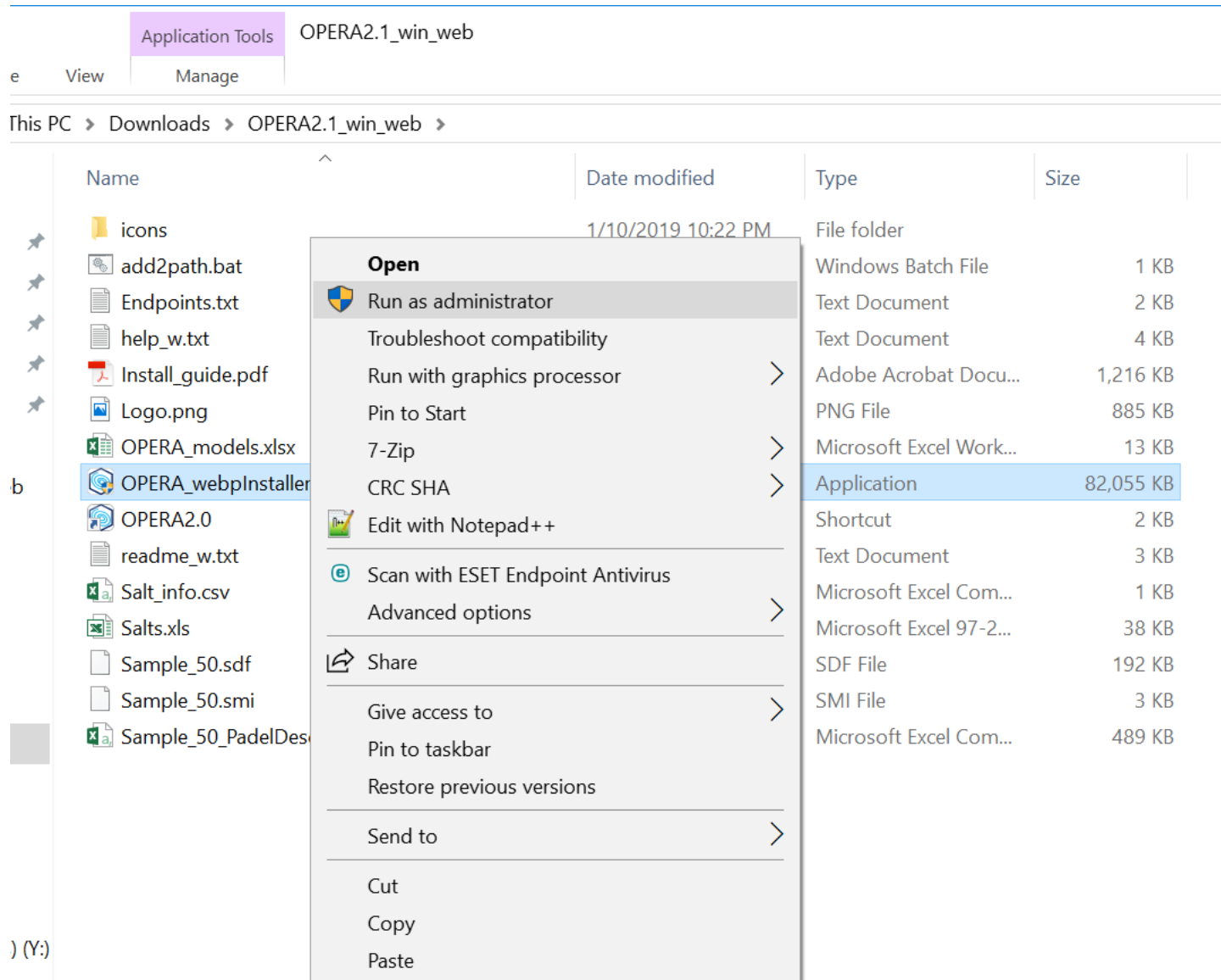
Download the installer from “Releases”

The screenshot shows the GitHub interface for the repository **NIEHS / OPERA**, which is a fork of **kmansouri/OPERA**. The top navigation bar includes a search field and links to Pull requests, Issues, Marketplace, and Explore. Below the repository name, there are tabs for Code, Pull requests (0), Projects (0), Wiki, Insights, and Settings. The **Releases** tab is selected, showing a list of releases. The latest release, **v2.0-beta7-windows**, is highlighted with a green box and labeled "Latest release". It has 551,294 downloads and is marked as "Verified". To the right of the release list, the title **OPERA 2.0 Windows** is displayed, followed by a note that **kmansouri** released it 27 days ago with 2 commits to master since. Below this, there are 4 assets listed: **OPERA2.0_win_offline.zip**, **OPERA2.0_win_web.zip**, **Source code (zip)**, and **Source code (tar.gz)**.

Read the provided information about the different installers available and select one of the installer options:

- OS: windows, Linux
- Command line, GUI
- online installer, offline installer

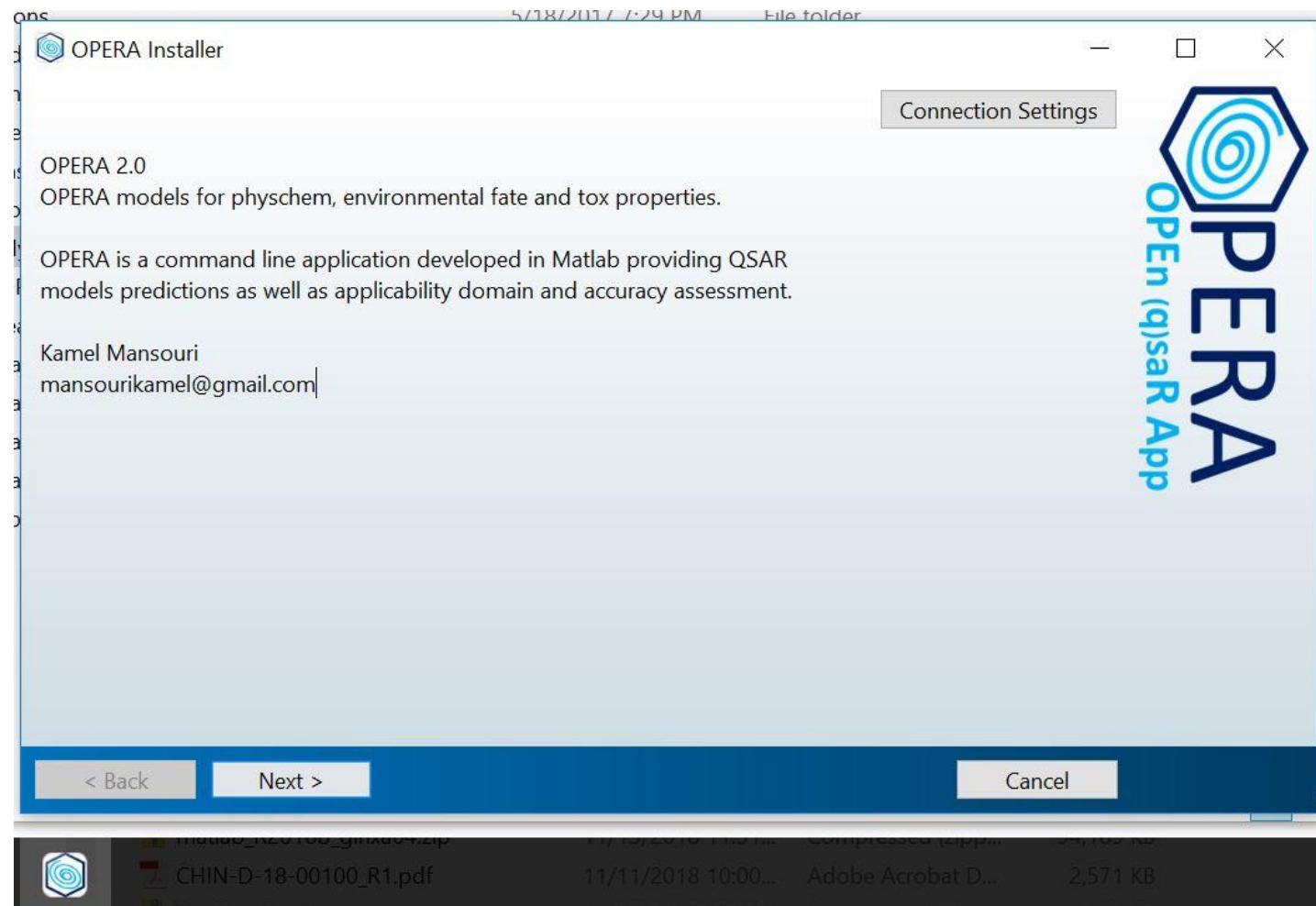
Unzip and run OPERA_Installer as administrator



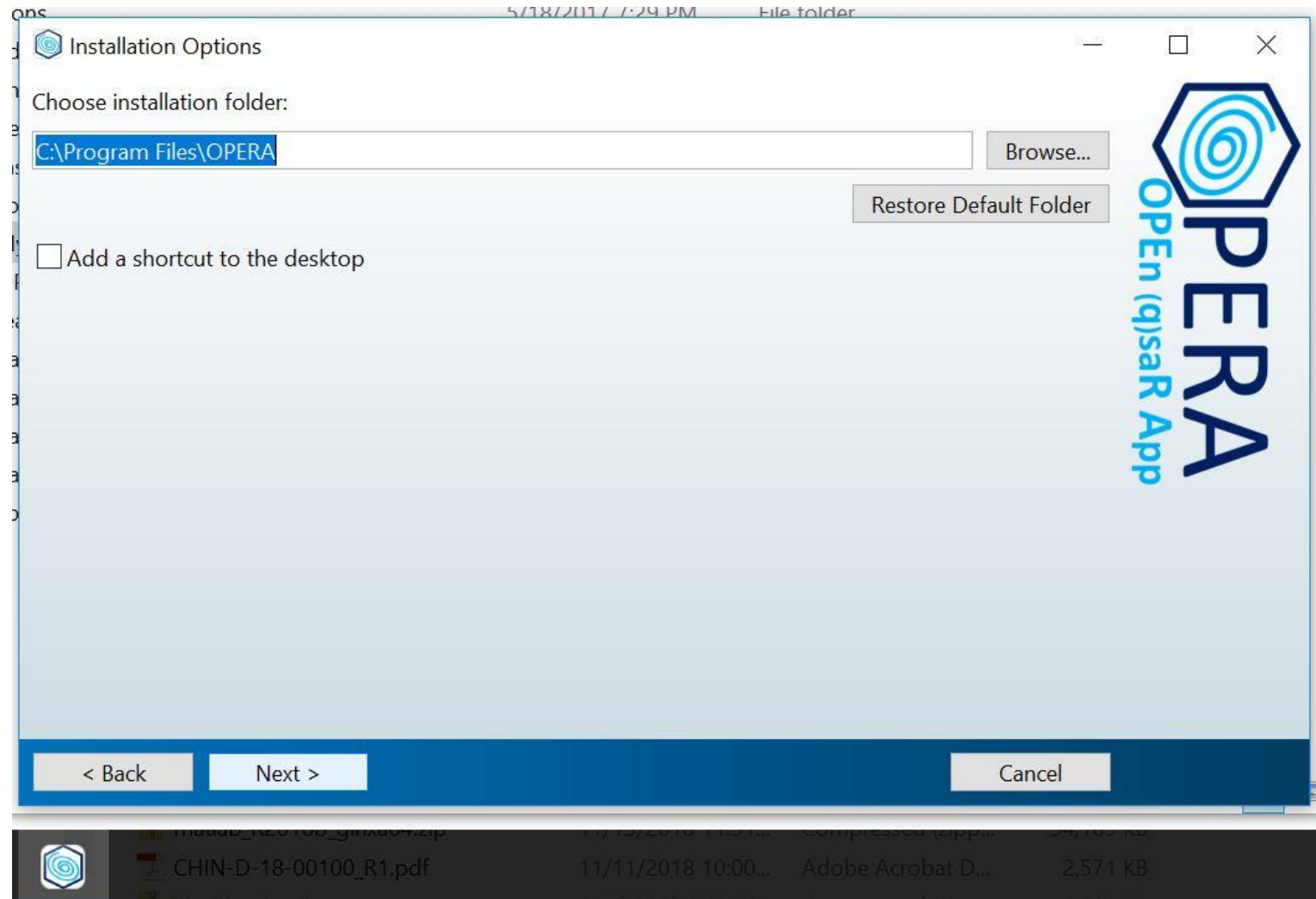
The installer will unpack and initialize quickly



First intro to OPERA, click next when ready



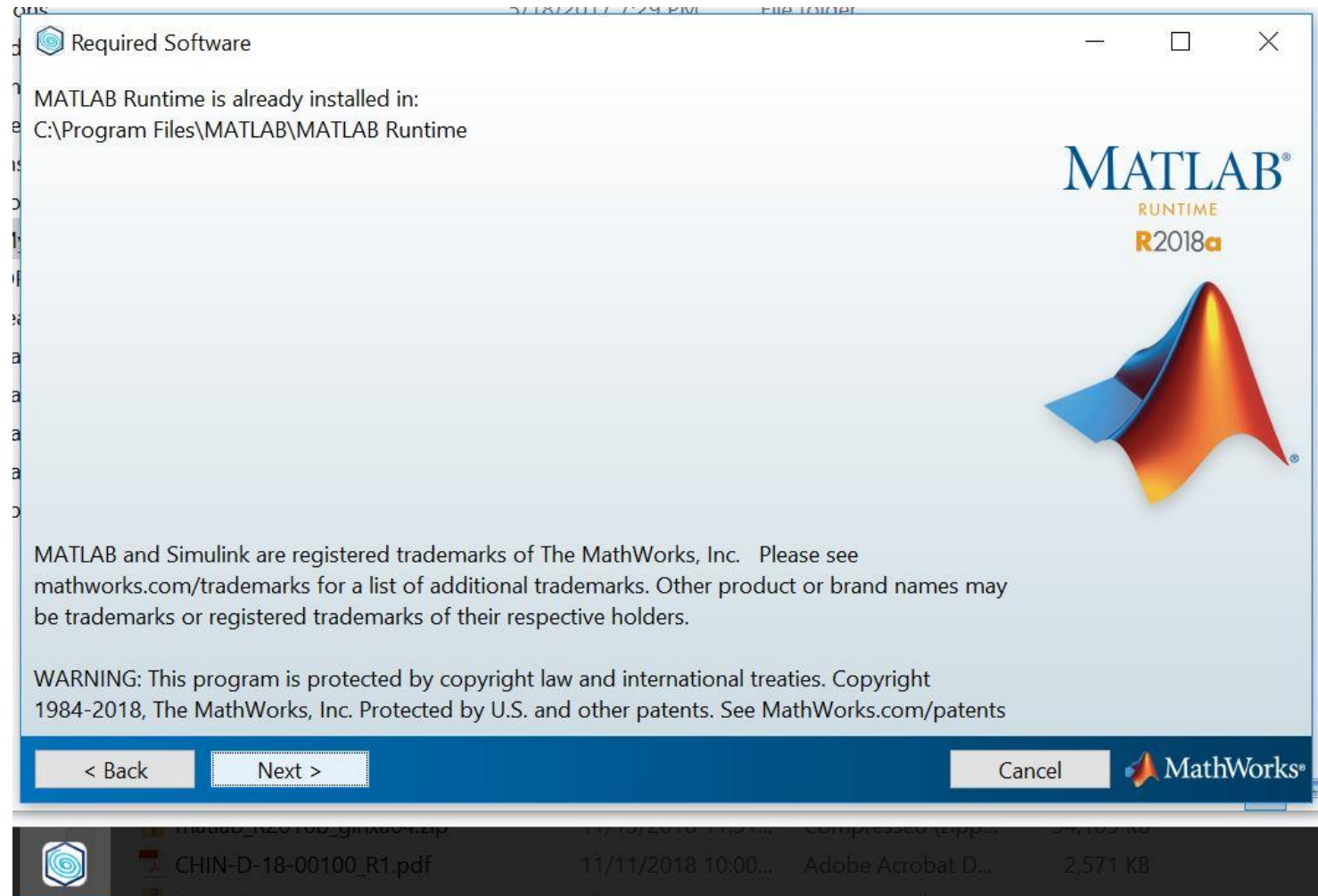
Recommended install options.



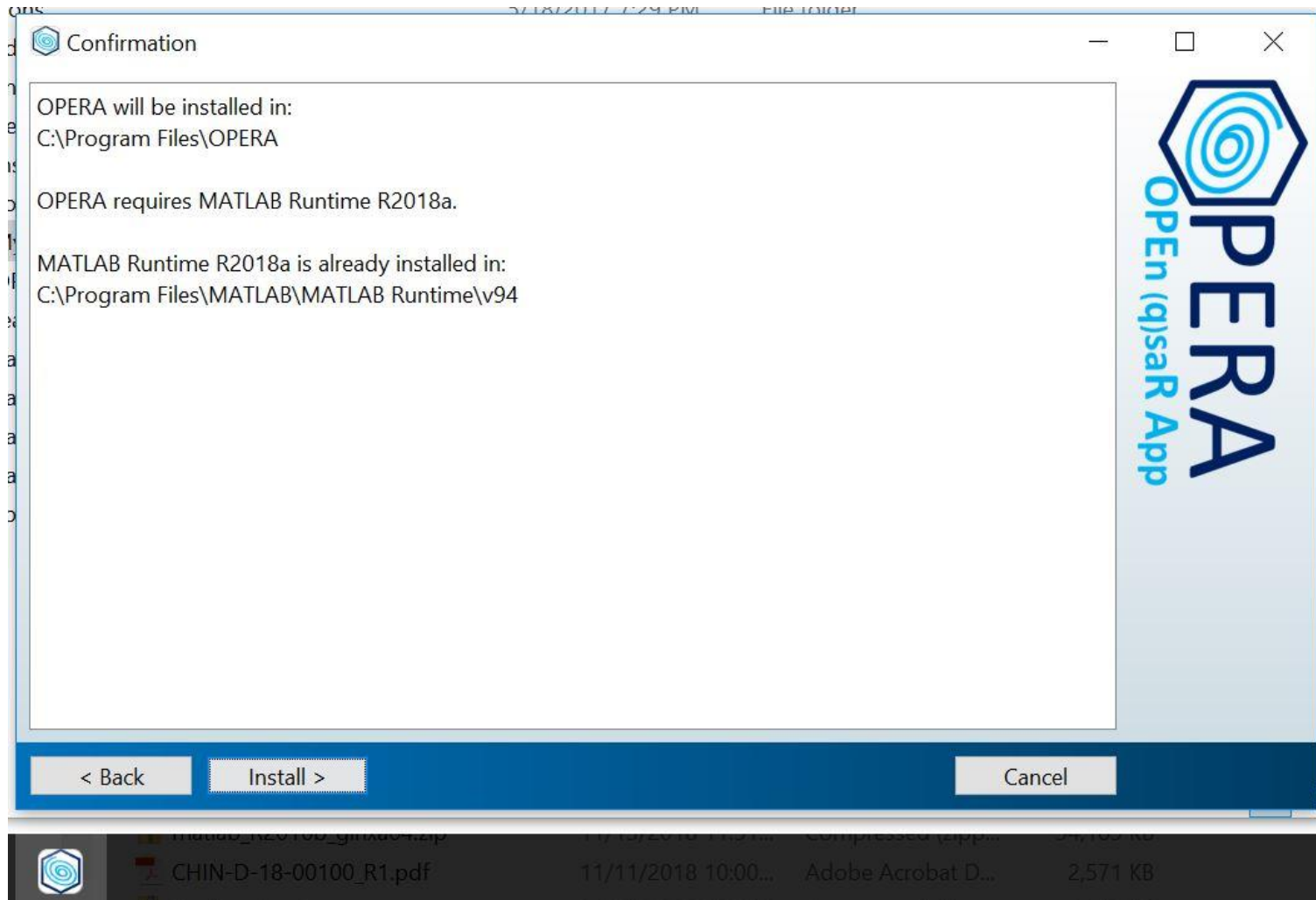
It is important to

- For GUI, select add shortcut
- For command line, the shortcut is provided in the unzipped file. Just copy paste it to the desktop.
- Note that if the default installation folder is modified:
- The GUI will adapt automatically
- The command line shortcut location needs to be updated and further instructions will be provided during the first run.

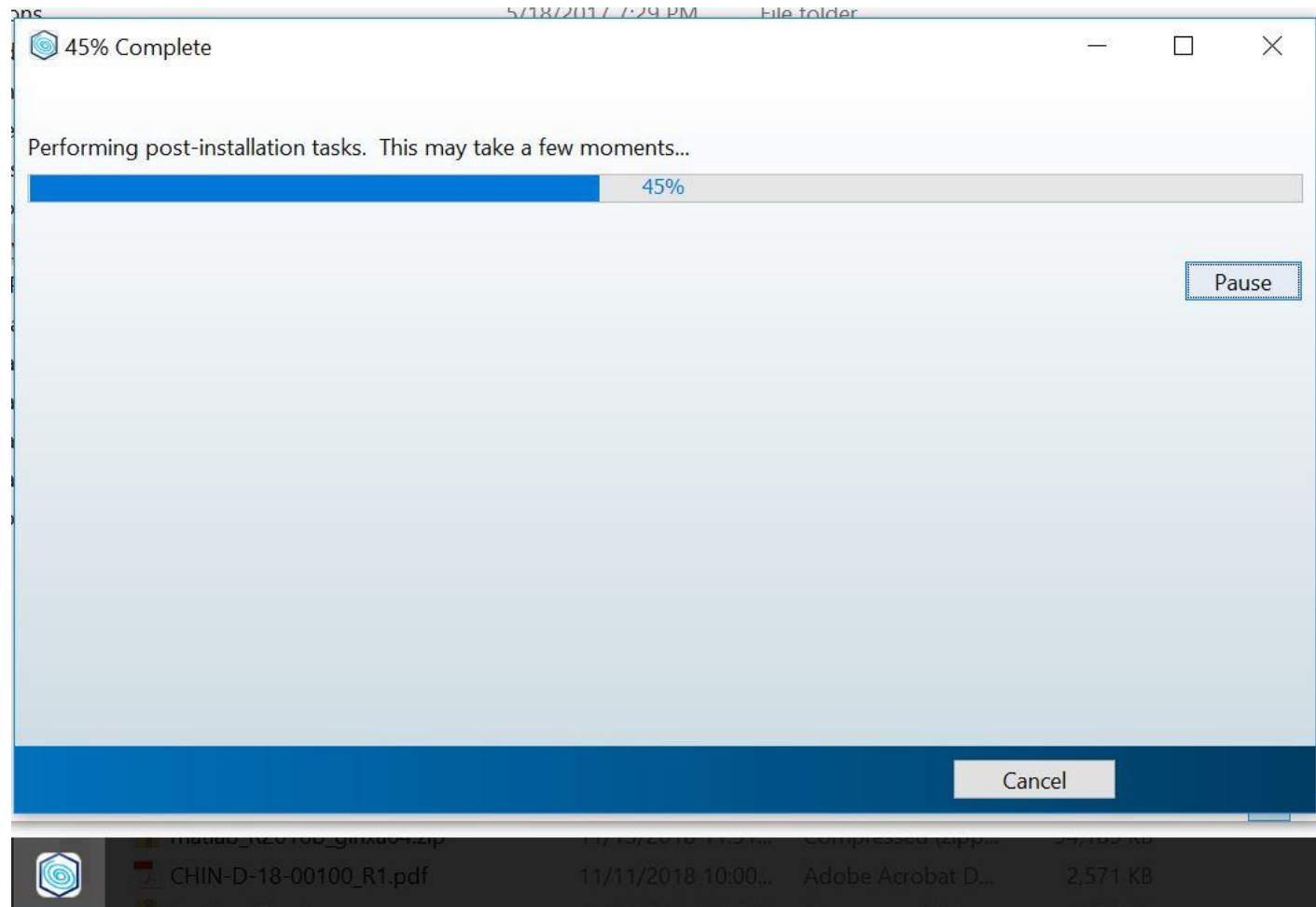
Just click next



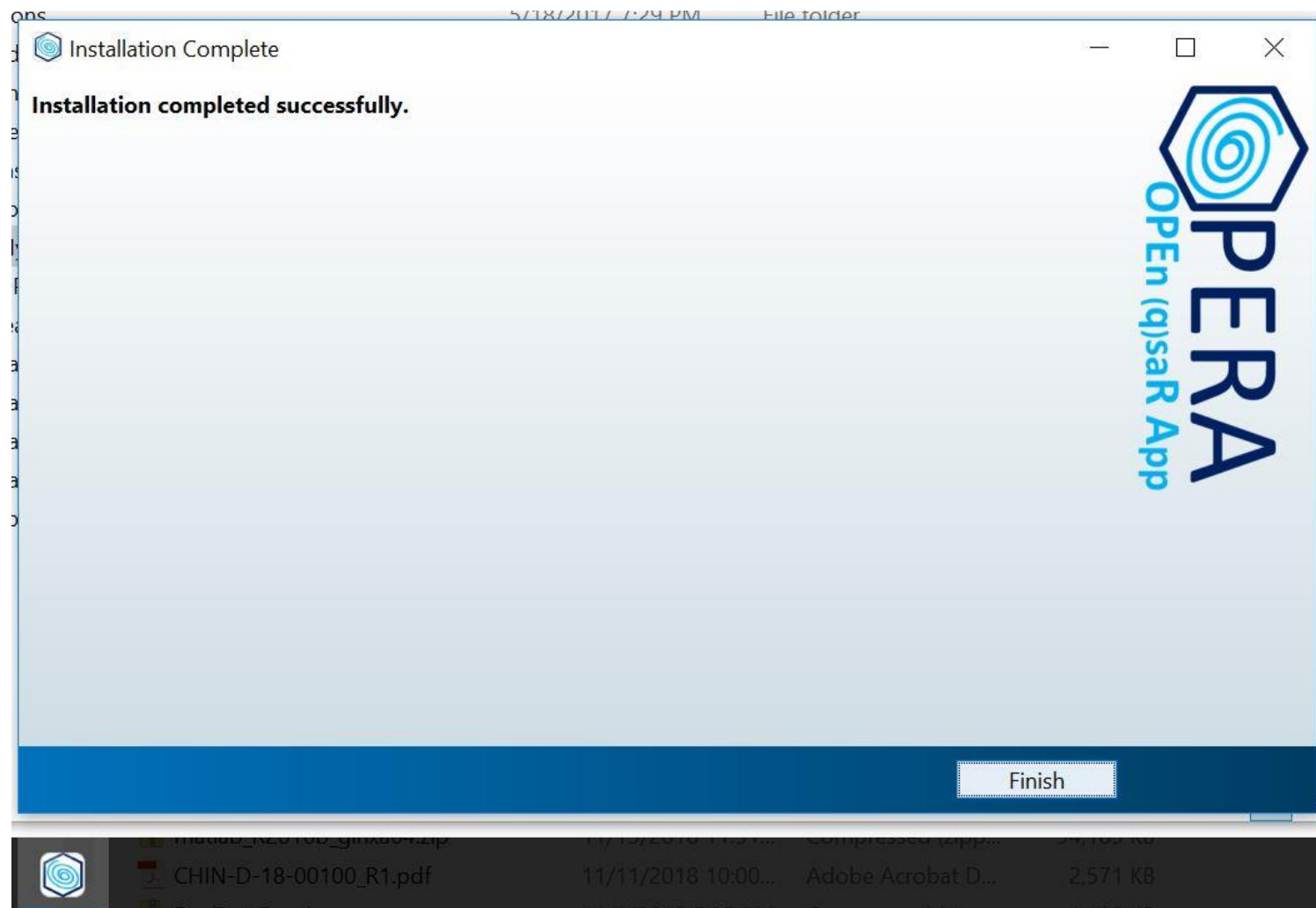
Confirm and install.



Wait while downloading and/or installing the runtime. (next time will go faster)









That's it! IT's OPERAtional.



Installed files

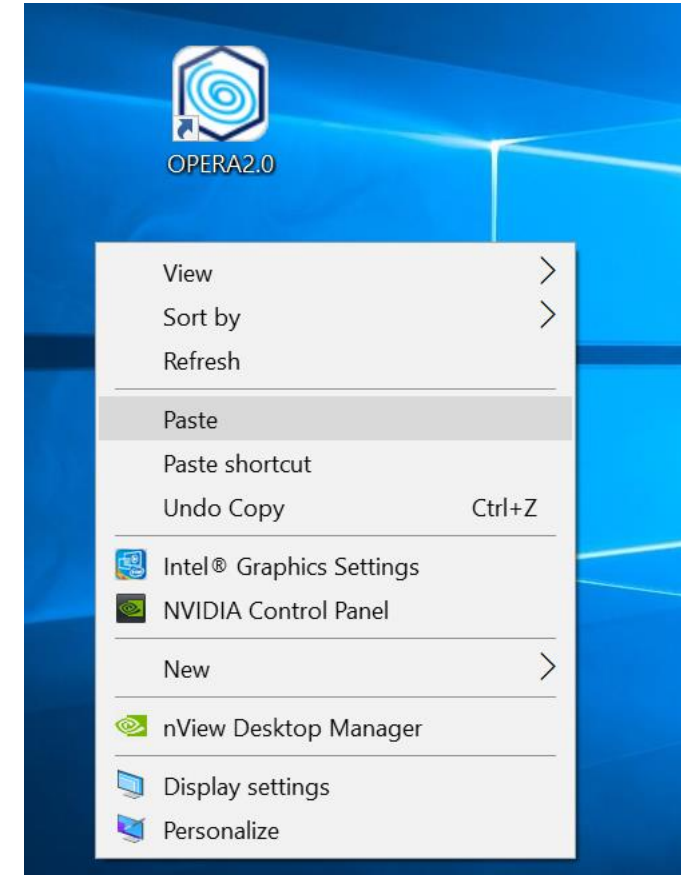
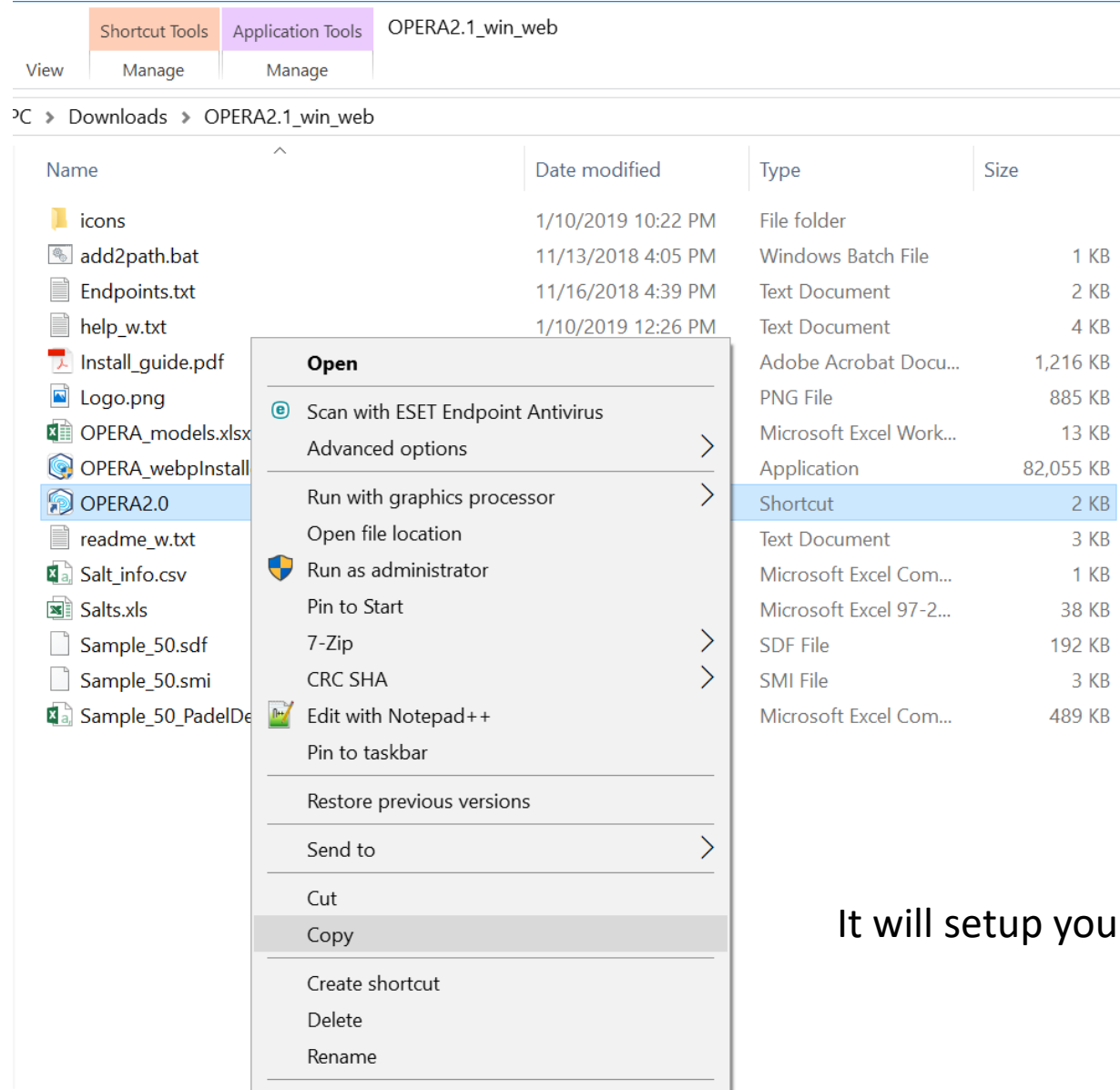
OPERA files

→ ▾ ↑ 📁 > This PC > Local Disk (C:) > Program Files > OPERA > application			
▼ 📁 OPERA	^		
> 📁 appdata			
📁 application			
> 📁 sys			
> 📁 uninstall			
> 📁 Realtek			
> 📁 STMicroelectron			
	Name ^	Type	Size
	 CDKDescUI-2.0.jar	Executable Jar File	11,672 KB
	 desc_fp.xml	XML File	1 KB
	 icon.ico	Icon	7 KB
	 OPERA.exe	Application	19,279 KB
	 padel-full-1.00.jar	Executable Jar File	30,943 KB
	 splash.png	PNG File	72 KB

Runtime files

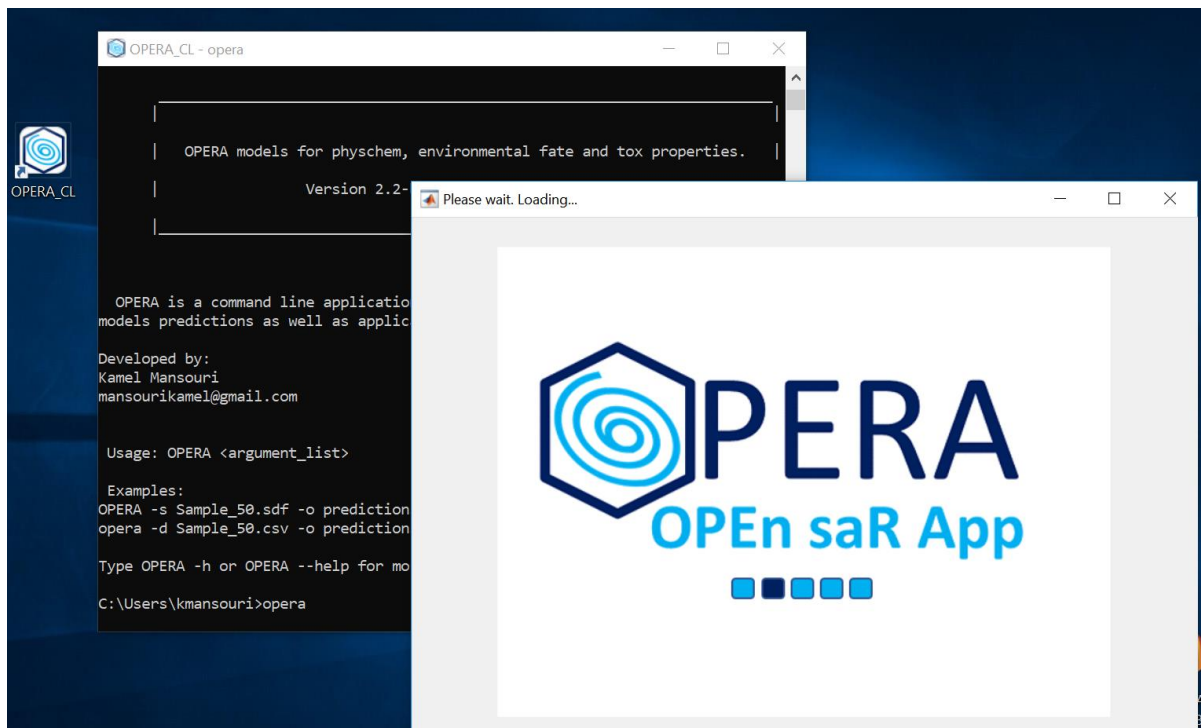
C > Local Disk (C:) > Program Files > MATLAB > MATLAB Runtime > v94		
	^	
Name	Date modified	Type
📁 appdata	12/21/2018 10:37...	File folder
📁 bin	12/21/2018 11:06...	File folder
📁 etc	12/21/2018 10:37...	File folder
📁 extern	12/21/2018 10:37...	File folder
📁 help	12/21/2018 11:06...	File folder
📁 java	12/21/2018 10:36...	File folder
📁 lib	12/21/2018 11:06...	File folder
📁 mcr	12/21/2018 10:36...	File folder
📁 polyspace	12/21/2018 11:03...	File folder
📁 remote	12/21/2018 11:05...	File folder
📁 resources	12/21/2018 11:06...	File folder
📁 runtime	12/21/2018 10:37...	File folder
📁 settings	12/21/2018 10:36...	File folder
📁 src	12/21/2018 11:05...	File folder
📁 sys	12/21/2018 11:06...	File folder
📁 toolbox	12/21/2018 11:06...	File folder
📁 ui	12/21/2018 11:05...	File folder
📁 uninstall	12/21/2018 10:37...	File folder
📄 MCR_license.txt	2/3/2015 2:40 PM	Text Document
📄 patents.txt	1/16/2018 3:26 PM	Text Document
📄 trademarks.txt	12/28/2013 2:08 ...	Text Document
📄 VersionInfo.xml	2/23/2018 1:39 PM	XML Document

For command line only: copy the provided shortcut to your desktop.



It will setup your environment variable and run the application in command line.

Double click the shortcut to run the app

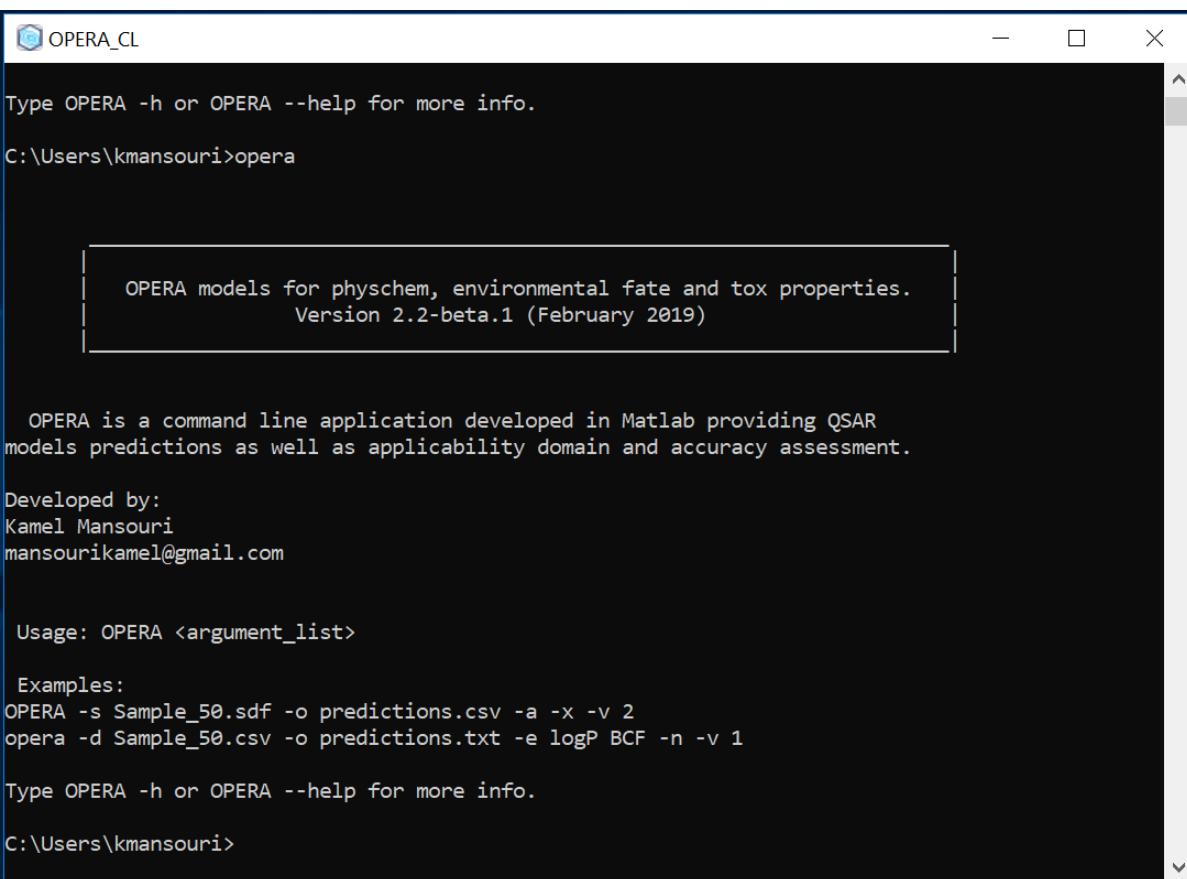


Command line



GUI

Ready to run the models



OPERA_CL

Type OPERA -h or OPERA --help for more info.

C:\Users\kmansouri>opera

OPERA models for physchem, environmental fate and tox properties.
Version 2.2-beta.1 (February 2019)

OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment.

Developed by:
Kamel Mansouri
mansourikamel@gmail.com

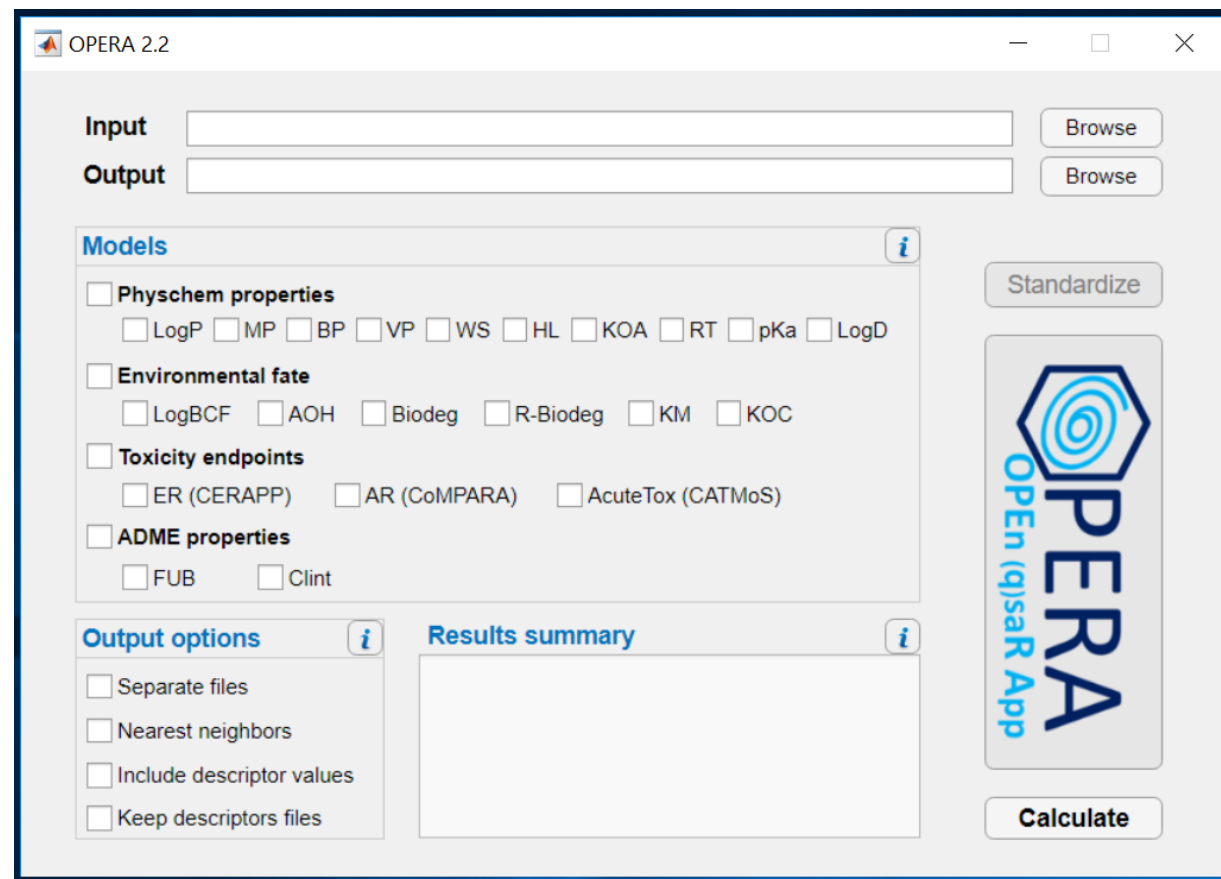
Usage: OPERA <argument_list>

Examples:
OPERA -s Sample_50.sdf -o predictions.csv -a -x -v 2
opera -d Sample_50.csv -o predictions.txt -e logP BCF -n -v 1

Type OPERA -h or OPERA --help for more info.

C:\Users\kmansouri>

Command line



OPERA 2.2

Input Browse

Output Browse


Models i

- ☐ **Physchem properties**
☐ LogP ☐ MP ☐ BP ☐ VP ☐ WS ☐ HL ☐ KOA ☐ RT ☐ pKa ☐ LogD
- ☐ **Environmental fate**
☐ LogBCF ☐ AOH ☐ Biodeg ☐ R-Biodeg ☐ KM ☐ KOC
- ☐ **Toxicity endpoints**
☐ ER (CERAPP) ☐ AR (CoMPARA) ☐ AcuteTox (CATMoS)
- ☐ **ADME properties**
☐ FUB ☐ Clint

Output options i

- ☐ Separate files
- ☐ Nearest neighbors
- ☐ Include descriptor values
- ☐ Keep descriptors files

Results summary i



OPERA
Open (q)sar App

Standardize

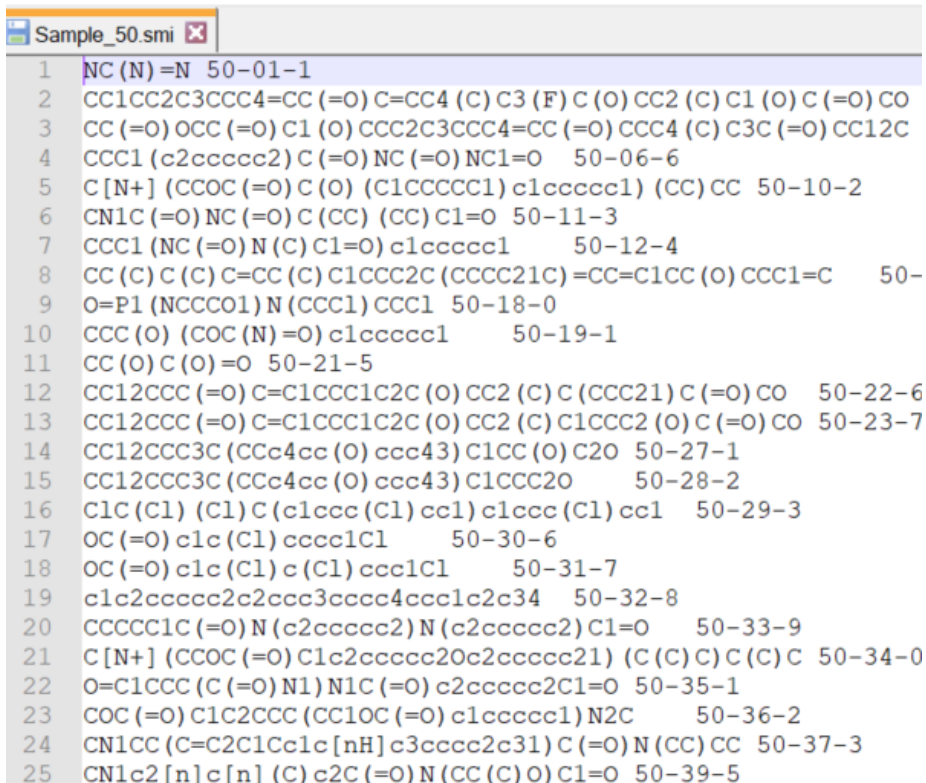
Calculate

GUI

Accepted Input files are QSAR-ready structures in:

SMILES file:

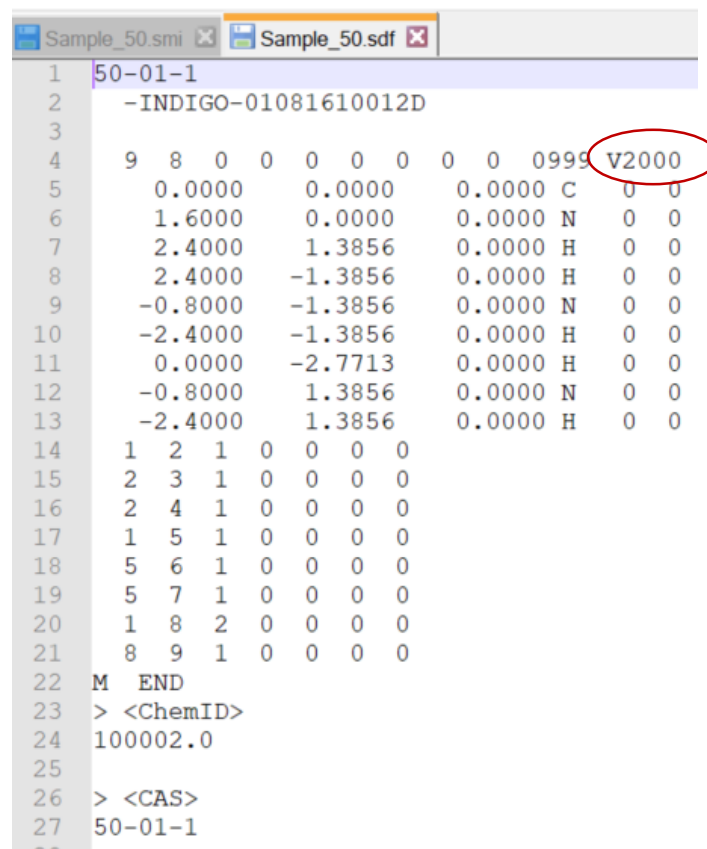
- Extension .smi
- Tab delimited text file (structure \t ID)
- No headers & no empty lines



```
1 NC(N)=N 50-01-1
2 CC1CC2C3CCC4=CC(=O)C=CC4(C)C3(F)C(O)CC2(C)C1(O)C(=O)CO
3 CC(=O)OCC(=O)C1(O)CCC2C3CCC4=CC(=O)CCC4(C)C3C(=O)CC12C
4 CCC1(c2ccccc2)C(=O)NC(=O)NC1=O 50-06-6
5 C[N+](CCOC(=O)C(O)(C1CCCC1)c1ccccc1)(CC)CC 50-10-2
6 CN1C(=O)NC(=O)C(CC)(CC)C1=O 50-11-3
7 CCC1(NC(=O)N(C)C1=O)c1ccccc1 50-12-4
8 CC(C)C(C)C=CC(C)C1CCC2C(CCCC21C)=CC=C1CC(O)CCC1=C 50-
9 O=P1(NCCCC1)N(CCC1)CCC1 50-18-0
10 CCC(O)(COC(N)=O)c1ccccc1 50-19-1
11 CC(O)C(O)=O 50-21-5
12 CC12CCC(=O)C=C1CCC1C2C(O)CC2(C)C(CCC21)C(=O)CO 50-22-6
13 CC12CCC(=O)C=C1CCC1C2C(O)CC2(C)C1CCC2(O)C(=O)CO 50-23-7
14 CC12CCC3C(CC4cc(O)ccc43)C1CC(O)C2O 50-27-1
15 CC12CCC3C(CC4cc(O)ccc43)C1CCC2O 50-28-2
16 ClC(Cl)(Cl)C(c1ccc(Cl)cc1)c1ccc(Cl)cc1 50-29-3
17 OC(=O)c1c(Cl)cccc1Cl 50-30-6
18 OC(=O)c1c(Cl)c(Cl)ccc1Cl 50-31-7
19 c1c2ccccc2c2ccc3cccc4ccc1c2c34 50-32-8
20 CCCCC1C(=O)N(c2ccccc2)N(c2ccccc2)C1=O 50-33-9
21 C[N+](CCOC(=O)C1c2ccccc2Oc2ccccc21)(C(C)C)C(C)C 50-34-0
22 O=C1CCC(C(=O)N1)N1C(=O)c2ccccc2C1=O 50-35-1
23 COC(=O)C1C2CCC(CC1OC(=O)c1ccccc1)N2C 50-36-2
24 CN1CC(C=C2C1Cc1c[nH]c3ccccc2c31)C(=O)N(CC)CC 50-37-3
25 CN1c2[n]c[n](C)c2C(=O)N(CC(C)O)C1=O 50-39-5
```

SDF/Mol file (v2000):

- Extension .sdf/.mol
- Different blocks
- Single or multiple molecules



```
1 50-01-1
2 -INDIGO-01081610012D
3
4 9 8 0 0 0 0 0 0 0 0 0999 V2000
5 0.0000 0.0000 0.0000 C 0 0
6 1.6000 0.0000 0.0000 N 0 0
7 2.4000 1.3856 0.0000 H 0 0
8 2.4000 -1.3856 0.0000 H 0 0
9 -0.8000 -1.3856 0.0000 N 0 0
10 -2.4000 -1.3856 0.0000 H 0 0
11 0.0000 -2.7713 0.0000 H 0 0
12 -0.8000 1.3856 0.0000 N 0 0
13 -2.4000 1.3856 0.0000 H 0 0
14 1 2 1 0 0 0 0
15 2 3 1 0 0 0 0
16 2 4 1 0 0 0 0
17 1 5 1 0 0 0 0
18 5 6 1 0 0 0 0
19 5 7 1 0 0 0 0
20 1 8 2 0 0 0 0
21 8 9 1 0 0 0 0
22 M END
23 > <ChemID>
24 100002.0
25
26 > <CAS>
27 50-01-1
```

Descriptors files:

- Extension .csv
- Calculated by PaDEL and/or CDK
- Continuous and/or fingerprints

	A	B	C	D	E	F	G
Name	nAcid	ALogP	ALogP2	AMR	apol	naAromAtn.	
100001	0	0	0	0	43.20152	20	
100002	0	0	0	0	48.0551	22	
100003	0	0	0	0	39.68152	18	
100004	0	0	0	0	14.56076	6	
100005	0	1.8631	3.471142	29.4773	19.8951	0	
100006	0	0.1248	0.015575	9.7308	27.78793	10	
100007	0	0	0	0	31.30793	14	
100008	0	0.4562	0.208118	4.601	29.54793	12	
100009	0	0.642	0.412164	5.5021	26.02793	10	
100010	0	0	0	0	22.93434	10	
100011	0	0.642	0.412164	5.5021	26.02793	10	
100012	0	0	0	0	27.78793	12	
100013	0	0.789	0.622521	15.0732	21.17434	6	
100014	0	0.5906	0.348808	7.7935	20.98076	9	
100015	0	1.284	1.648656	11.0042	20.74793	6	
100016	0	1.926	3.709476	16.5063	23.84152	6	
100017	0	2.568	6.594624	22.0084	26.9351	6	
100018	0	0.0686	0.004706	23.2621	18.56152	0	
100019	0	0.5729	0.328214	14.2678	23.84152	6	

QSAR-ready structures:

- Option 1: download the QSAR-ready SMILES from the EPA CompTox Dashboard:
https://comptox.epa.gov/dashboard/dsstoxdb/batch_search

United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Step Six: Click "Download"

Select Output Format: Excel Download

Customize Results

☐ Select All
☐ Select All in Lists

Chemical Identifiers

☒ DTXSID
☒ Chemical Name
☐ DTXCID
☒ CAS-RN
☐ InChIKey
☐ IUPAC Name

Structures

☐ Mol File
☐ SMILES
☐ InChI String
☐ MS-Ready SMILES
☒ QSAR-Ready SMILES

Intrinsic And Predicted Properties

☐ Molecular Formula
☐ Average Molecular Weight

Presence in Lists:

- ☐ 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities
- ☐ AEGLs: Acute Exposure Guideline Levels
- ☐ ANDROGEN: Androgen Receptor Chemicals
- ☐ ARTICLE: Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)
- ☐ ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances
- ☐ ATSDR: Toxic Substances Portal Chemical List
- ☐ California Office of Environmental Health Hazard Assessment
- ☐ CERAPP: Collaborative Estrogen Receptor Activity Prediction Project
- ☐ CHEMINV: ToxCast/Tox21 Chemical inventory available as DMSO solutions (20181123)
- ☐ CHEMINV: EPA Chemical Inventory for ToxCast
- ☐ CHEMINV: EPA ToxCast CHEMINV list of volatiles
- ☐ CHEMINV: EPA ToxCast Cheminventory chemicals with stability problems
- ☐ EPA ToxCast Cheminventory DMSO Insolubles
- ☐ EPA ToxCast Cheminventory List of Reactives
- ☐ PubBank database from the University of Alberta
- ☐ ANTIBIOTIC list of antibiotics
- ☐ DRUGS: Pharmaceutical List with EU, Swiss, US Consumption Data

QSAR-Ready SMILES are the SMILES representations of the desalted, de-isotoped, stereo-neutral forms of chemical structures associated with particular chemical substances. One QSAR-Ready SMILES can map to multiple individual chemical substances. They differ from MS-Ready SMILES based on a more restricted set of salt counterions primarily.

QSAR-ready structures:

- Option 2: run the KNIME QSAR-ready standardization workflow:
<https://github.com/kmansouri/QSAR-ready>

kmansouri / QSAR-ready

<> Code 0 Issues 0 Pull requests 0

Standardization workflow for QSAR-ready chemical structures

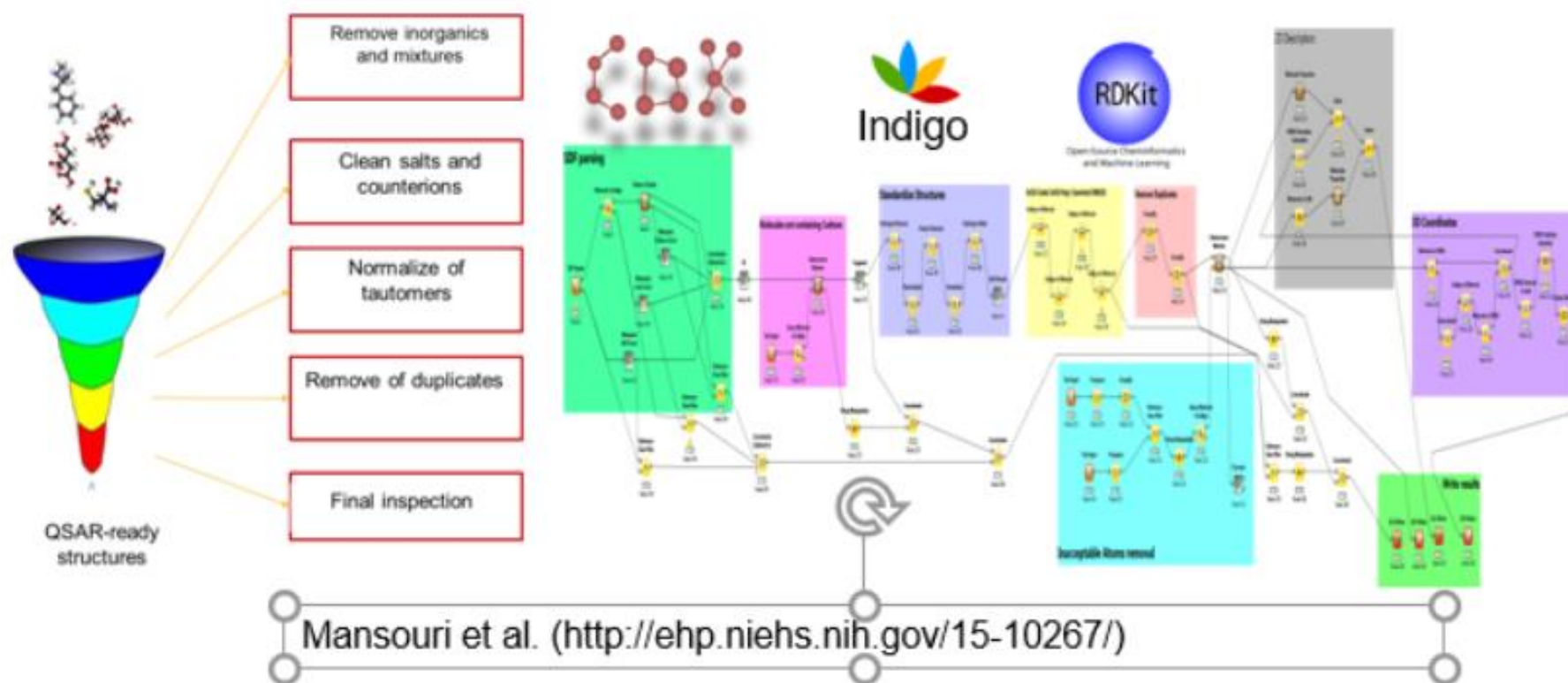
Manage topics

24 commits 1 branch

Branch: master New pull request

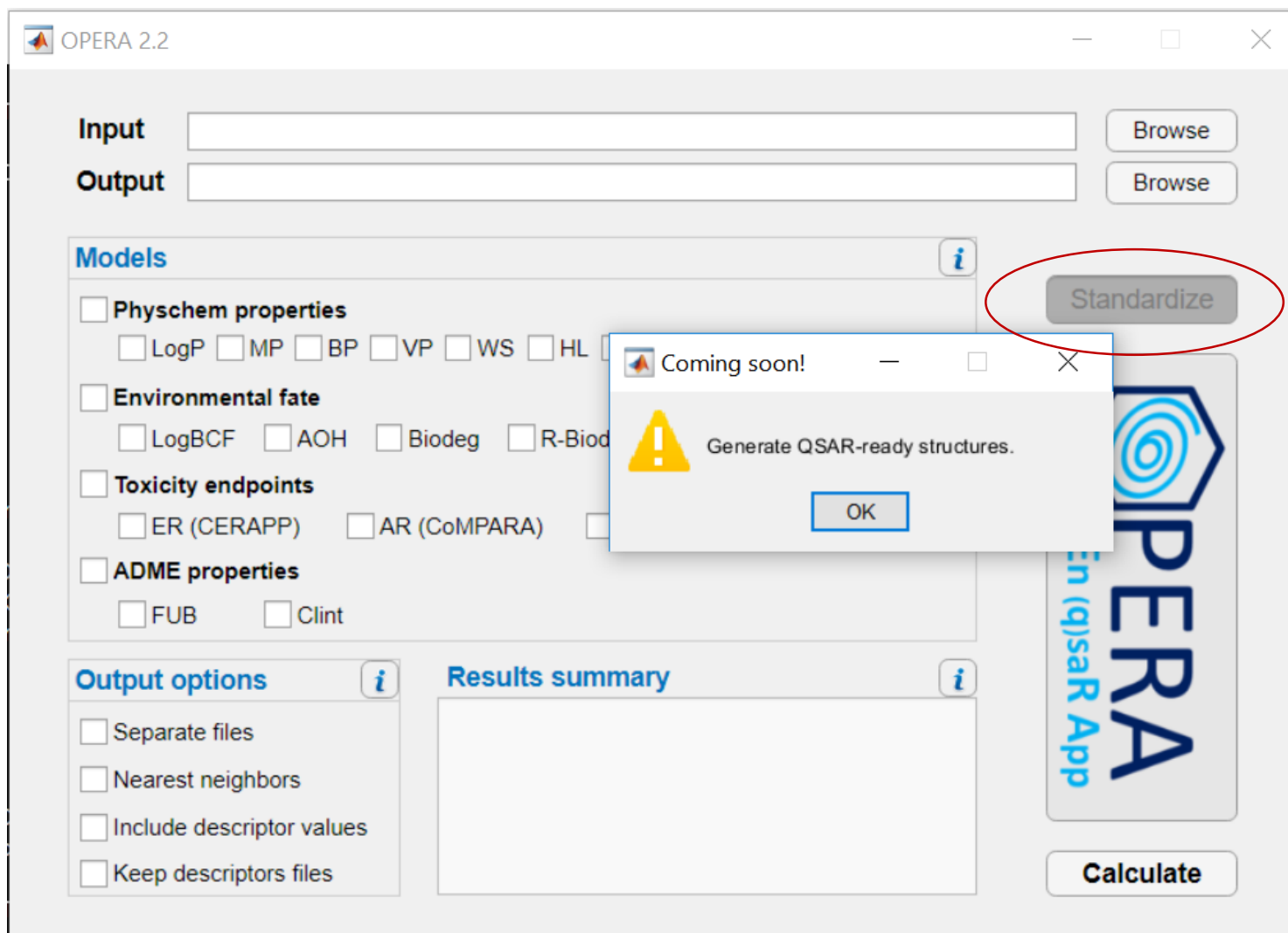
kmansouri Sample output files

- LICENSE
- QSAR-ready.knwf**
- README.md
- Sample_input.zip
- Sample_output.zip
- Transformations.zip



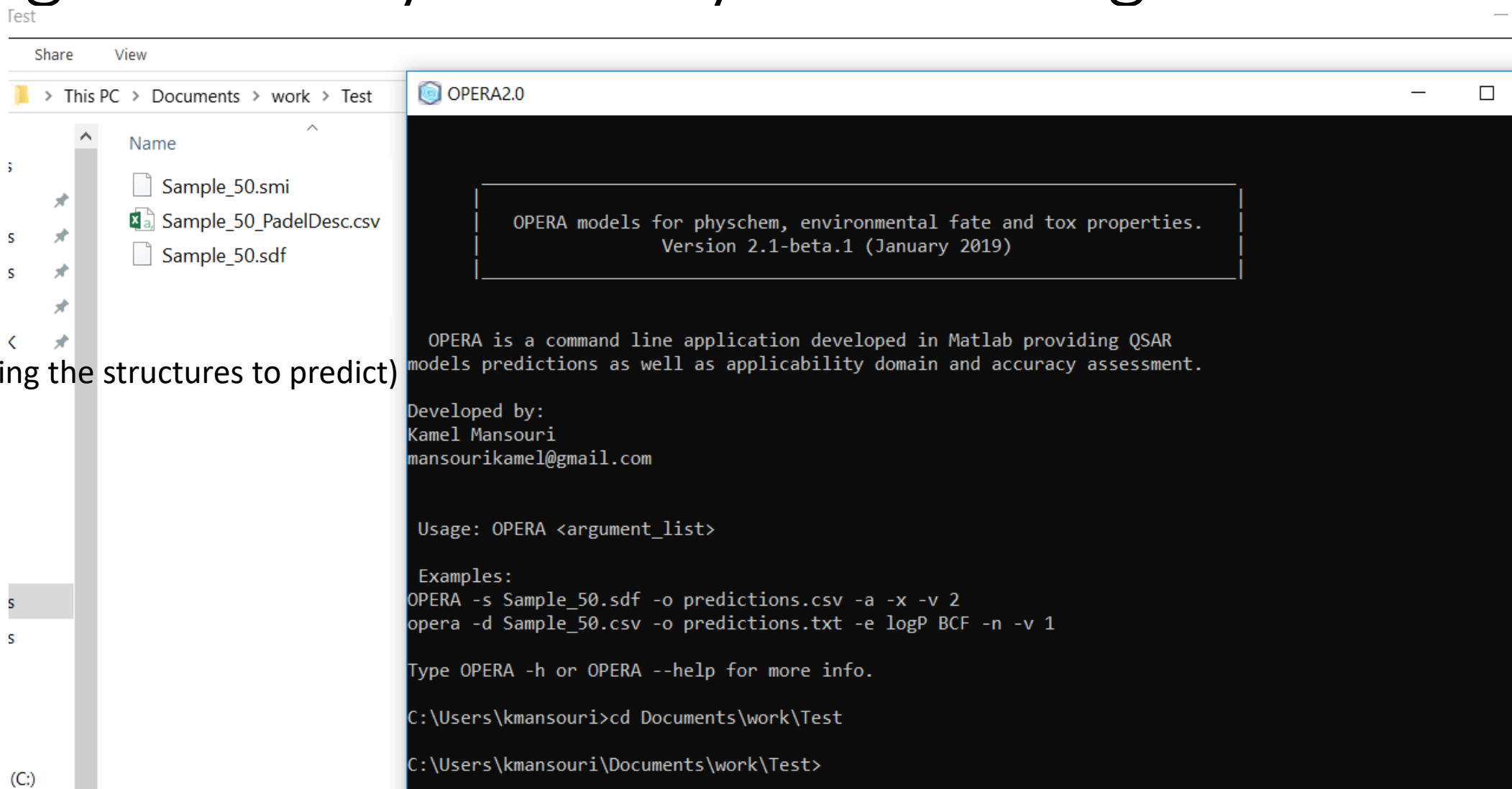
QSAR-ready structures:

- Option 3 (in the Future): Use the standardize option in OPERA



Command line: Change directory “cd” to your working folder

(Folder containing the structures to predict)



Use the commands in the help file or type: “opera -h” for help

```
C:\Users\kmansouri\Documents\work\Test>opera -h
```

```
OPERA models for physchem, environmental fate and tox properties.  
Version 2.1-beta.1 (January 2019)
```

OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment. All models are built on curated data from public domain. Molecular descriptors are calculated using PaDEL and CDK software.

Usage: OPERA <input> <output> [Options]
By default, all endpoints will be calculated.

Examples:

```
OPERA -s Sample_50.sdf -o predictions.csv -a -x -n -v 2  
opera -d Sample_50.csv -o predictions.txt -e logP BCF -v 1
```

Input:

-s, --SDF, --MOL, --SMI Structure file containing the molecule(s) to be predicted. IDs will be assigned if the file does not contain molecule names. Molecular descriptors will be calculated using PaDEL software.

-d, --Descriptors pre-calculated PaDEL descriptors in csv file. If the first column is not "Name" as the standard PaDEL output, molecule IDs will be assigned.

-m, --Mat, --ascii Matlab matrix or ascii file containing PaDEL descriptors.

-i, --MolID Molecule names in csv file.

-t, --SaltInfo Salt IDs to improve melting point predictions. List provided in Salts.xls

-l, --Labels Descriptor labels. Necessary if the descriptor file does not contain labels or contains more than the 1444 PaDEL 2D descriptors.

Output:

-o, --Output Output file containing the predictions, applicability domain and accuracy information. File extension could be csv or txt. The output will contain by default: Molecule ID, predicted value (pred), Applicability domain (AD), Similarity index (Sim_index) and accuracy estimate (Conf_index).

-n, --Neighbors Add 5 nearest neighbors from training set (CAS, InChIKeys, Observed and predicted values)

-O, --FullOutput Output file containing all prediction details and used descriptors in csv format.

-x, --Seperate Separate output file for each endpoint.

Miscellaneous:

-v, --Verbose Verbose level: 0=silent (default), 1=minimum details, 2=full details.

-a, --All All endpoints to be calculated (default).

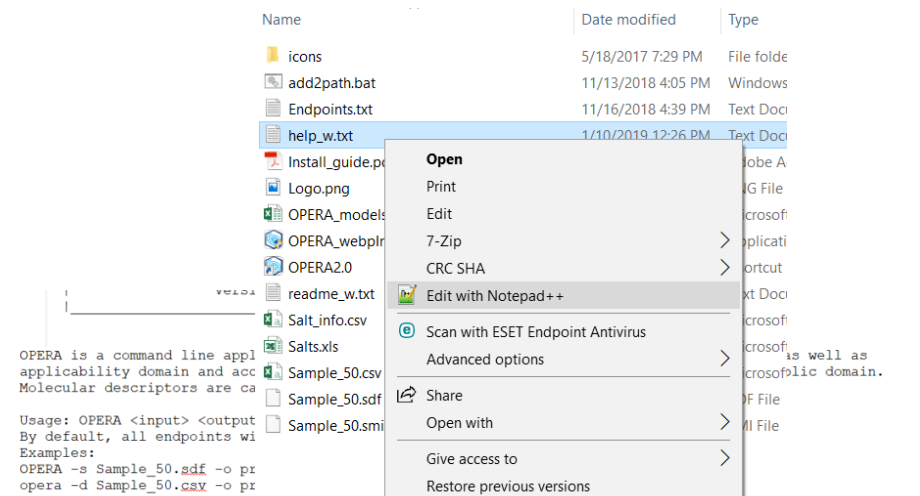
-c, --Clean Clean temporary files (generated during descriptor calculation.)

-logP, -BCF... List endpoints to be calculated (case insensitive). 'BCF'/'logBCF', 'BP', 'logP', 'MP', 'VP'/'logVP', 'WS', 'AOH', 'BioDeg', 'RB'/'ReadyBiodeg', 'HL'/'logHL', 'KM'/'logKM', 'KOA', 'Koc'/'logKoc', 'RT', 'pKa', 'logD', 'CERAPP'/'ER', 'CoMPARA'/'AR', 'CATMoS/AcuteTox'.
Groups of Endpoints: 'StrP' (Structural properties), 'PC'/'Physchem', 'EnvFate'/'EF', 'Tox' (ER, AR, AcuteTox).

-e, --Endpoint List endpoints to be calculated.

-h, --Help Display this help file and exit.

-V, --Version Version of the application



OPERA is a command line application developed in Matlab providing QSAR models predictions as well as applicability domain and accuracy assessment. All models are built on curated data from public domain. Molecular descriptors are calculated using PaDEL and CDK software.

Usage: OPERA <input> <output> [Options]
By default, all endpoints will be calculated.

Examples:

```
OPERA -s Sample_50.sdf -o pr  
opera -d Sample_50.csv -o pr
```

Input:

-s, --SDF, --MOL, --SMI Structure file containing the molecule(s) to be predicted. IDs will be assigned if the file does not contain molecule names. Molecular descriptors will be calculated using PaDEL software.

-d, --Descriptors pre-calculated PaDEL descriptors in csv file. If the first column is not "Name" as the standard PaDEL output, molecule IDs will be assigned.

-m, --Mat, --ascii Matlab matrix or ascii file containing PaDEL descriptors.

-i, --MolID Molecule names in csv file.

-t, --SaltInfo Salt IDs to improve melting point predictions. List provided in Salts.xls

-l, --Labels Descriptor labels. Necessary if the descriptor file does not contain labels or contains more than the 1444 PaDEL 2D descriptors.

Output:

-o, --Output Output file containing the predictions, applicability domain and accuracy information. File extension could be csv or txt. The output will contain by default: Molecule ID, predicted value (pred), Applicability domain (AD), Similarity index (Sim_index) and accuracy estimate (Conf_index).

-n, --Neighbors Add 5 nearest neighbors from training set (CAS, InChIKeys, Observed and predicted values)

-O, --FullOutput Output file containing all prediction details and used descriptors in csv format.

-x, --Seperate Separate output file for each endpoint.

Miscellaneous:

-v, --Verbose Verbose level: 0=silent (default), 1=minimum details, 2=full details.

-a, --All All endpoints to be calculated (default).

-c, --Clean Clean temporary files (generated during descriptor calculation.)

-logP, -BCF... List endpoints to be calculated (case insensitive). 'BCF'/'logBCF', 'BP', 'logP', 'VP'/'logVP', 'WS', 'AOH', 'BioDeg', 'RB'/'ReadyBiodeg', 'HL'/'logHL', 'KM'/'logKM', 'KOA', 'Koc'/'logKoc', 'RT', 'pKa', 'logD', 'CERAPP'/'ER', 'CoMPARA'/'AR', 'CATM Groups of Endpoints: 'StrP' (Structural properties), 'PC'/'Physchem', 'EnvFate'/'EF', 'Tox' (ER, AR, AcuteTox).

-e, --Endpoint List endpoints to be calculated.

-h, --Help Display this help file and exit.

-V, --Version Version of the application

Examples:

Usage: OPERA <argument_list>

General command: "OPERA <input> <output> <options> " no specific order for the arguments.

```
>opera -s Sample_50.sdf -o predictions.csv -logP -v 1
```

Using the sample sdf file (50 chemicals) to calculate logP only in minimum verbose mode (v=1). Endpoints names (only) are case insensitive.

```
>opera -s Sample_50.sdf -o predictions.csv
```

Simplest command specifying only the input and output. By default, OPERA will calculate all endpoints and output them to the same file in silent mode (V=0 by default).

```
>opera -d PadelDesc.csv -o predictions.csv -LOGP -BCF -v 2
```

Running one of the previous commands will generate the descriptor file "PadelDesc.csv" for this list of chemicals that you can use to run other models.

```
>opera -d PadelDesc.csv -o predictions.csv -physchem -v 2 -s Sample_50.sdf
```

This runs a list of models at the same time without running all of them. "physchem" or "pc" will run 10 models.

```
>opera -s Sample_50.smi -o predictions.csv -CATMoS -v 1
```

This will run CATMoS models only using "CATMoS" or "AcuteTox" (case insensitive).

```
>opera -s Sample_50.sdf -o predictions.csv -all -n -x -v 1 -c
```

This will run all models. "-n" to get the nearest neighbors and their experimental and predicted data. "-x" will output each one of them in a separate file. "-c" or "-clean" to delete generated descriptors afterwards.

```
>opera -s Sample_50.sdf -d PadelDesc.csv -o predictions.csv -Tox -n -x -v 1
```

This will run all toxicity endpoints including CERAPP (ER), CoMPARA (AR) and CATMoS (AcuteTox).

Running the provided sample structures

```
OPERA2.0
C:\Users\kmansouri\Documents\work\Test>opera -s Sample_50.sdf -o predictions.csv -v 1

All properties will be calculated:
General structural properties, Physchem, Env. fate and Tox Endpoints (CERAPP, CoMPARA and CATMoS)

===== Molecular Descriptors =====
PaDEL calculating 2D descriptors...
End of descriptors calculation: 50 molecules calculated.
PaDEL generating fingerprints...
End of fingerprint generation: 50 molecules calculated.
CDK 2.0 calculating 2D descriptors...
End of descriptors calculation: 50 molecules calculated.

===== Loading Input Files =====
Loading of PaDEL descriptors file...
The number of input molecules is: 50
Checking loaded variables.
The number of loaded PaDEL descriptors is: 1444
Loading of fingerprints file...
The number of loaded fingerprints bits is: 10145
Loading of CDK descriptors file...
The number of loaded CDK descriptors is: 286

===== Running The Models =====
Generating the general structural properties...
----- PhysChem properties -----
Predicting LogP values (Log10)...
Predicting MP values (Deg. C)...
Predicting BP values (Deg. C)...
Predicting LogVP values (Log10 mmHg)...
Predicting LogWS values (Log10 M)...
Predicting LogHL values (Log10 atm-m3/V)...
Predicting RT values (Mins)...
Predicting LogKOA values (Log10)...
Predicting pKa values (unitless)...
Predicting LogD values (Log10)...
----- Env. Fate Endpoints -----
Predicting LogOH values (Log10 cm3/molecule-sec)...
Predicting LogBCF values (Log10)...
Predicting Biodeg. half-life values (Log10 days)...
Predicting Ready-Biodegradability (Binary 0/1)...
Predicting LogKm half-life values (Log10 days)...
Predicting LogKoc values (Log10 L/Kg)...
----- Toxicity Endpoints -----
Predicting Estrogen Receptor Activity (CERAPP)...
Predicting Androgen Receptor Activity (CoMPARA)...
Predicting Acute Oral Tox. endpoints (CATMoS)...

===== End Of Calculation =====
50 molecules predicted
```

Command line

The screenshot shows the OPERA 2.2 graphical user interface. It features a window with a title bar and standard Windows controls. The interface is divided into several sections:

- Input:** A text box containing the file path `C:\Users\kmansouri\Downloads\Sample_50.sdf` and a "Browse" button.
- Output:** A text box containing the file path `C:\Users\kmansouri\Downloads\Predictions.csv` and a "Browse" button.
- Models:** A section with a blue header and an information icon. It contains three main categories of checkboxes:
 - Physchem properties:** Includes ☒ LogP, ☐ MP, ☐ BP, ☐ VP, ☐ WS, ☐ HL, ☐ KOA, ☐ RT, ☒ pKa, and ☒ LogD.
 - Environmental fate:** Includes ☒ LogBCF, ☒ AOH, ☒ Biodeg, ☒ R-Biodeg, ☒ KM, and ☒ KOC.
 - Toxicity endpoints:** Includes ☐ ER (CERAPP), ☐ AR (CoMPARA), and ☒ AcuteTox (CATMoS).
- ADME properties:** Includes ☐ FUB and ☒ Clint.
- Output options:** A section with a blue header and an information icon. It contains four checkboxes:
 - ☒ Separate files
 - ☒ Nearest neighbors
 - ☐ Include descriptor values
 - ☐ Keep descriptors files
- Results summary:** A section with a blue header and an information icon. It displays a text summary of the calculation process and results.
- Buttons:** A "Standardize" button is located to the right of the Models section. A large "Calculate" button is at the bottom right, next to the OPERA logo.
- Logo:** The OPERA logo, which includes a stylized hexagon with a target-like pattern and the text "OPERA" and "Open (q)sar App".

GUI

Sample output:

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	LogP_prec	AD_LogP	AD_index	Conf_index	MP_pred	AD_MP	AD_index	Conf_index	BP_pred	AD_BP	AD_index	Conf_index
2	-1.16675	1	0.480419	0.514066	56.96995	0	0.845536	0.591406	257.4774	0	0.148019	0.459055
3	1.865338	1	0.900462	0.860512	240.0609	1	0.886896	0.77831	391.2263	0	0.156367	0.289241
4	2.220271	1	0.908599	0.854532	225.4009	1	0.868502	0.769243	418.5021	0	0.198701	0.29612
5	1.418476	1	0.888307	0.832005	178.3617	1	0.907997	0.792323	255.0721	0	0.102487	0.429512
6	3.263077	1	0.475522	0.511194	113.4288	1	0.423574	0.396192	303.5064	1	0.218644	0.377263
7	1.250174	1	0.88672	0.826925	140.6504	1	0.870612	0.678357	244.3635	0	0.199114	0.408801
8	1.661618	1	0.884391	0.880548	140.7473	1	0.869315	0.704217	313.0569	0	0.227134	0.451173
9	8.521042	0	0.423516	0.451172	109.0241	1	0.918952	0.808039	388.1919	0	0.288908	0.478521
10	0.794911	1	0.916139	0.831594	53.26655	1	0.889102	0.805115	300.1238	0	0.215153	0.397557
11	-1.34435	1	0.542563	0.511082	70.8003	1	0.862021	0.603289	300.3755	0	0.275189	0.513219
12	-0.76893	1	0.880398	0.785205	45.61057	1	0.87657	0.805093	208.6963	1	0.419094	0.402072
13	2.075537	1	0.885388	0.774448	190.2265	1	0.890244	0.776764	400.3502	0	0.227297	0.322959
14	1.653221	1	0.887992	0.897769	221.5952	1	0.91094	0.869202	389.2867	0	0.161782	0.29391
15	2.635143	1	0.863338	0.699103	270.1702	1	0.856718	0.608162	382.0755	0	0.189731	0.315207
16	3.907586	1	0.87718	0.772113	196.6646	1	0.881577	0.794906	372.7094	1	0.273881	0.367613
17	6.91244	1	0.871424	0.810684	104.6093	1	0.899273	0.844186	350.3782	1	0.549401	0.606396
18	2.64288	1	0.919871	0.754499	157.4414	1	0.94492	0.84299	293.8553	1	0.729356	0.715913
19	3.536705	1	0.611748	0.663621	147.8905	1	0.94981	0.796142	300.6033	1	0.552324	0.69184
20	6.128521	1	0.947304	0.88721	223.6441	1	0.978058	0.708651	485.9157	0	0.937575	0.871229

In this example, all endpoints were predicted and the output is written in a single file “predictions.csv”.

For each model, the predicted value is associated with applicability domain and accuracy estimates.

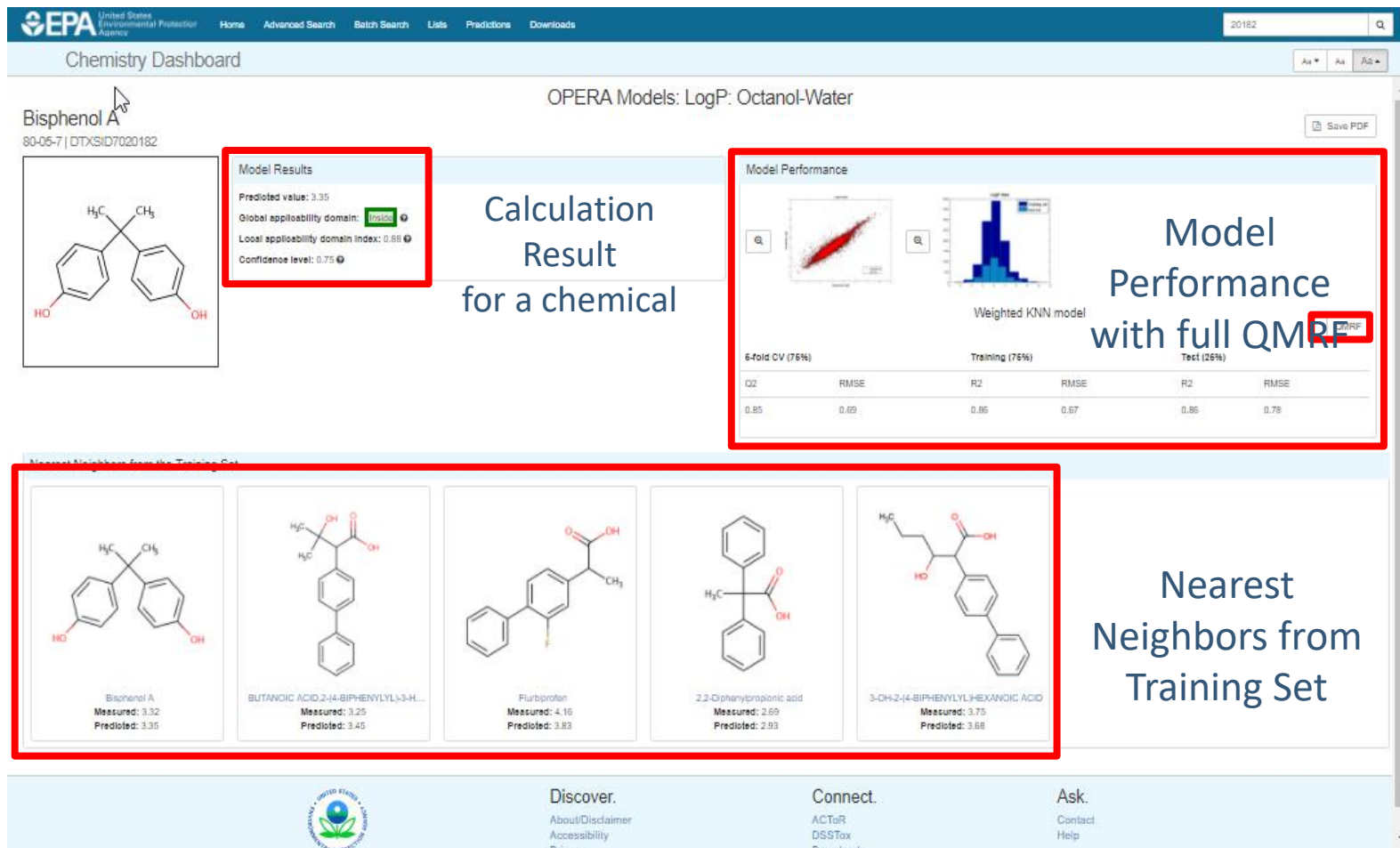
AD_logP (0/1): Global applicability domain considering the whole chemical space of the model.

AD_index [0-1]: Local applicability domain based on the similarity to the 5 nearest neighbors

Conf_index [0-1]: Accuracy estimate based on the predictions of the 5 nearest neighbors.

OPERA prediction report on the EPA Dashboard

<https://comptox.epa.gov/dashboard>



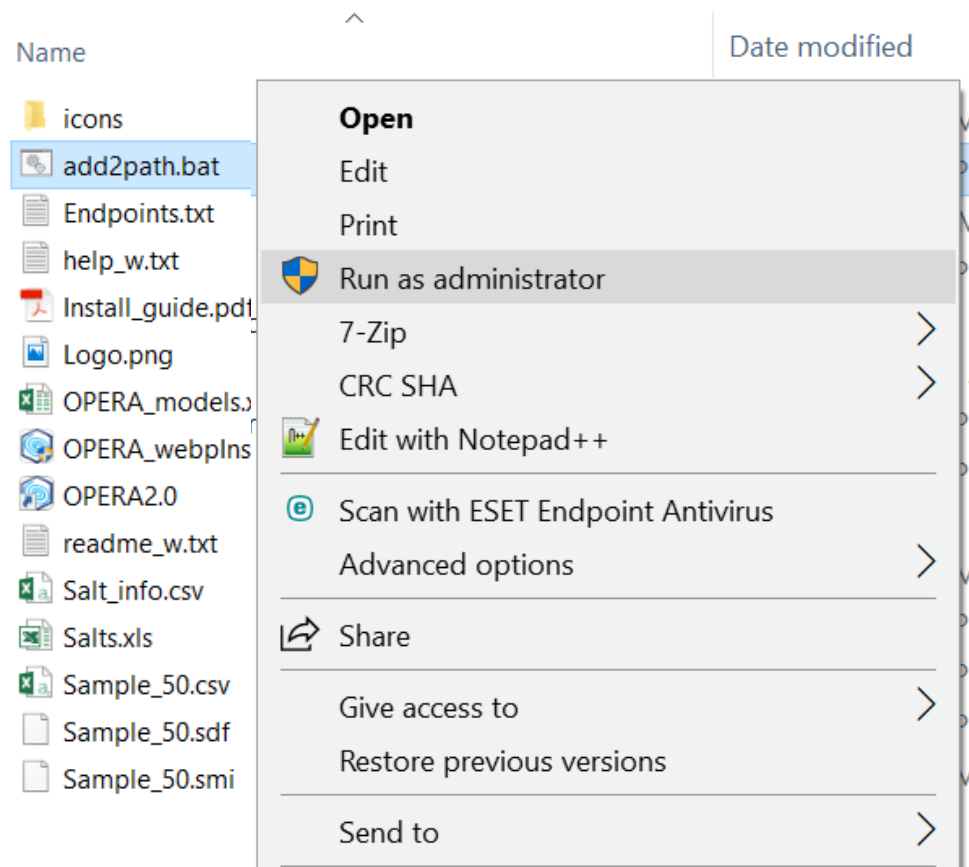
More details in the QMRF reports available on the EPA dashboard, the Github repo and the JRC QSAR database

Optional and only for expert users of command line:

Run add2path.bat as admin to permanently edit system %PATH%.

Only needed to install for all users or to use in a pipeline/ workflow.

Run only ONCE to avoid truncated path!



Thank you!

Register on the GitHub repo for the updates <https://github.com/NIEHS/OPERA>

References:

- [1] Mansouri K. et al. J Cheminform (2018) <https://doi.org/10.1186/s13321-018-0263-1>.
- [2] Mansouri, K. et al. SAR and QSAR in Env. Res. (2016). <https://doi.org/10.1080/1062936X.2016.1253611>
- [3] Williams A. J. et al. J Cheminform (2017) <https://doi.org/10.1186/s13321-017-0247-6>
- [4] The CompTox Chemistry Dashboard (<https://comptox.epa.gov/dashboard>)
- [5] JRC QSAR Model Database <https://qsardb.jrc.ec.europa.eu/qmrf/endpoint>

Your feedback will be appreciated!
Kamel.mansouri@nih.gov
kmansouri@ils-inc.com
mansourikamel@gmail.com