Windows version installation and quick run guide

OPERA 2.1

Download the installer from "Releases"

- OPERA2.0_win_web.zip: the installer will automatically download all dependencies. If Matlab runtime is already installed, it will just deploy the app.

- OPERA2.0_win_offline.zip: the installer is self-contained with all dependencies. No internet connection required during installation.

Unzip and run OPERA_Installer as administrator

The installer will unpack and initialize quickly

First intro to OPERA2, click next when ready

Recommended install options.
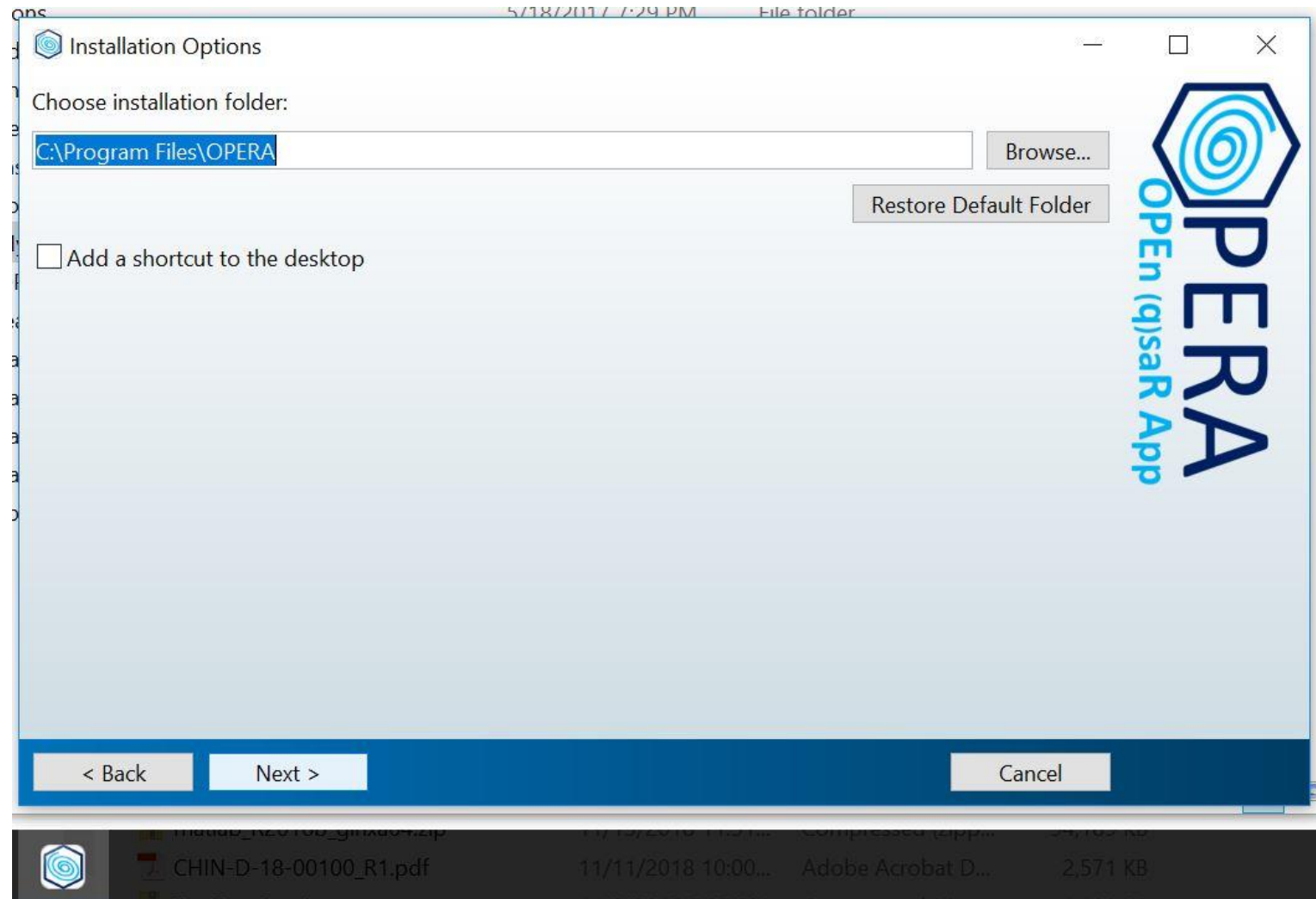
It is important to
- Keep the recommended installation folder
- No need to create the shortcut at this step. It will be created later.

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

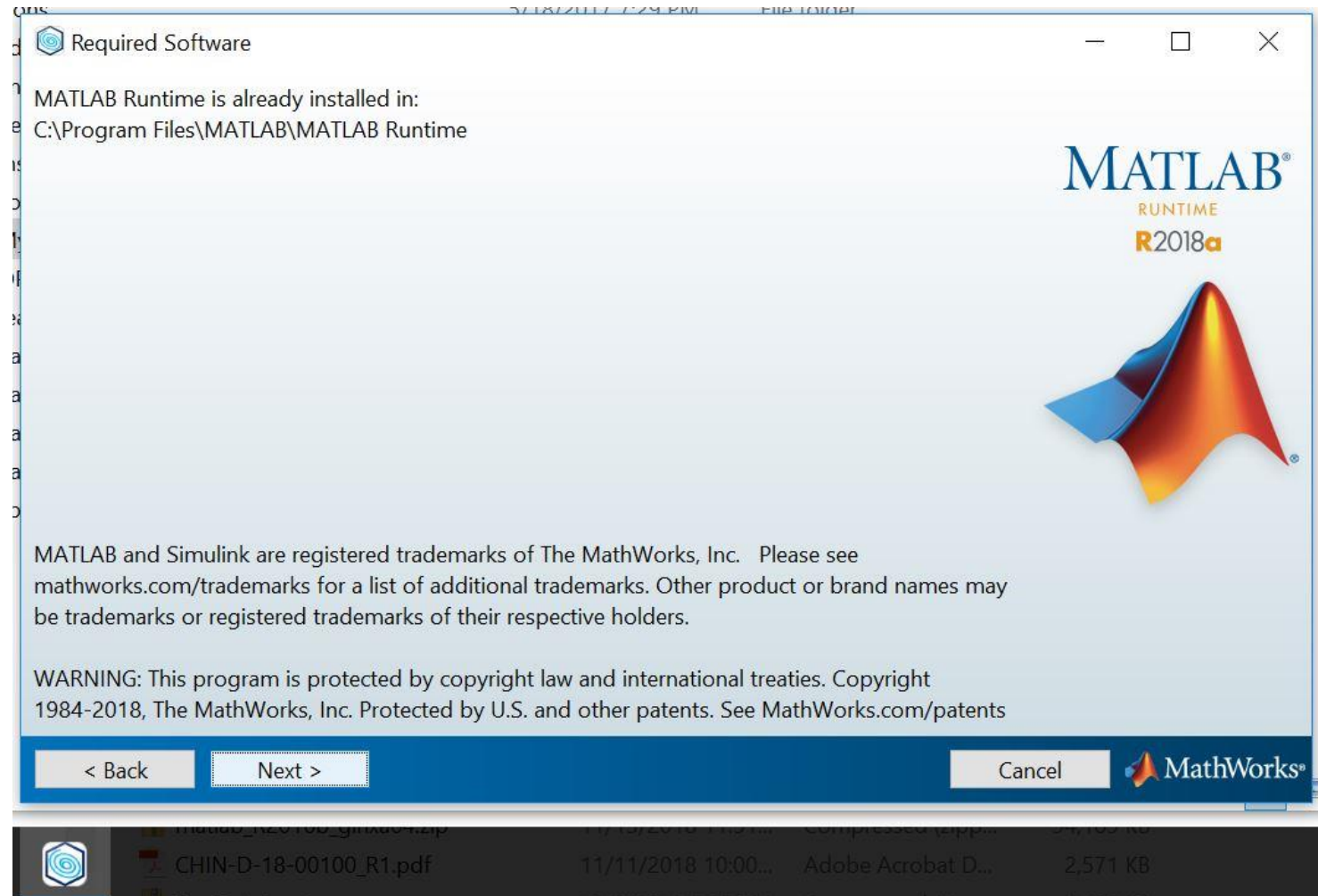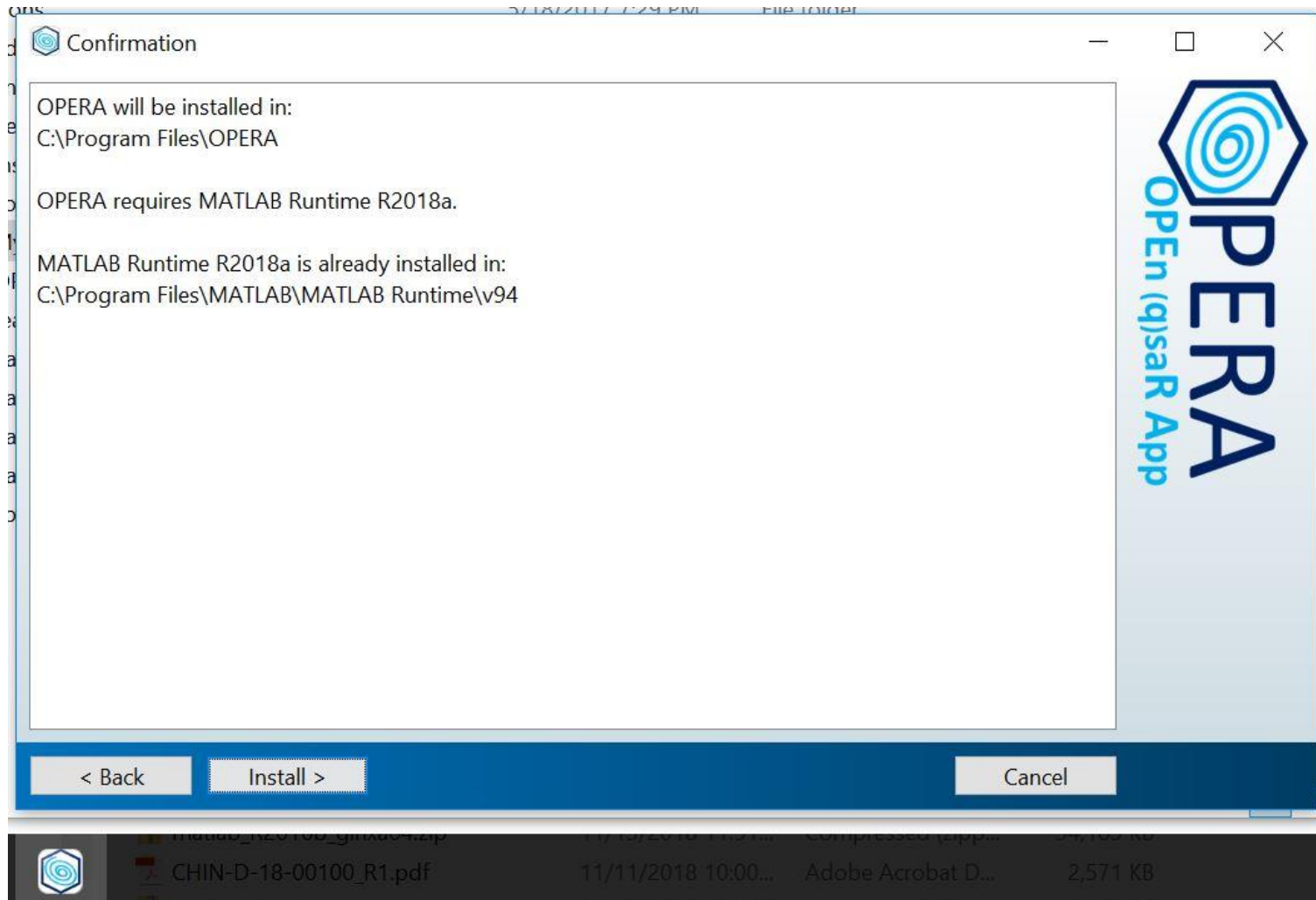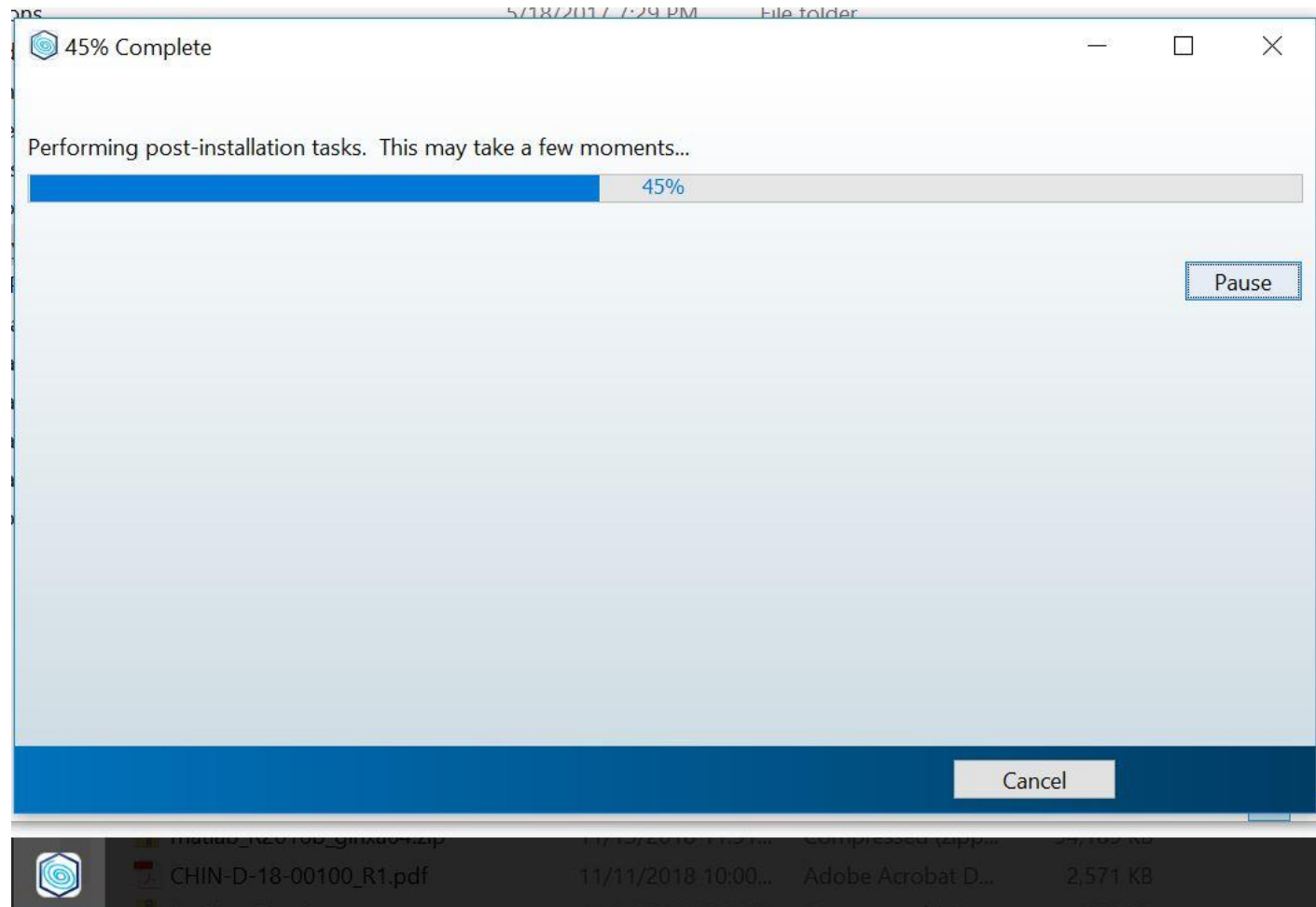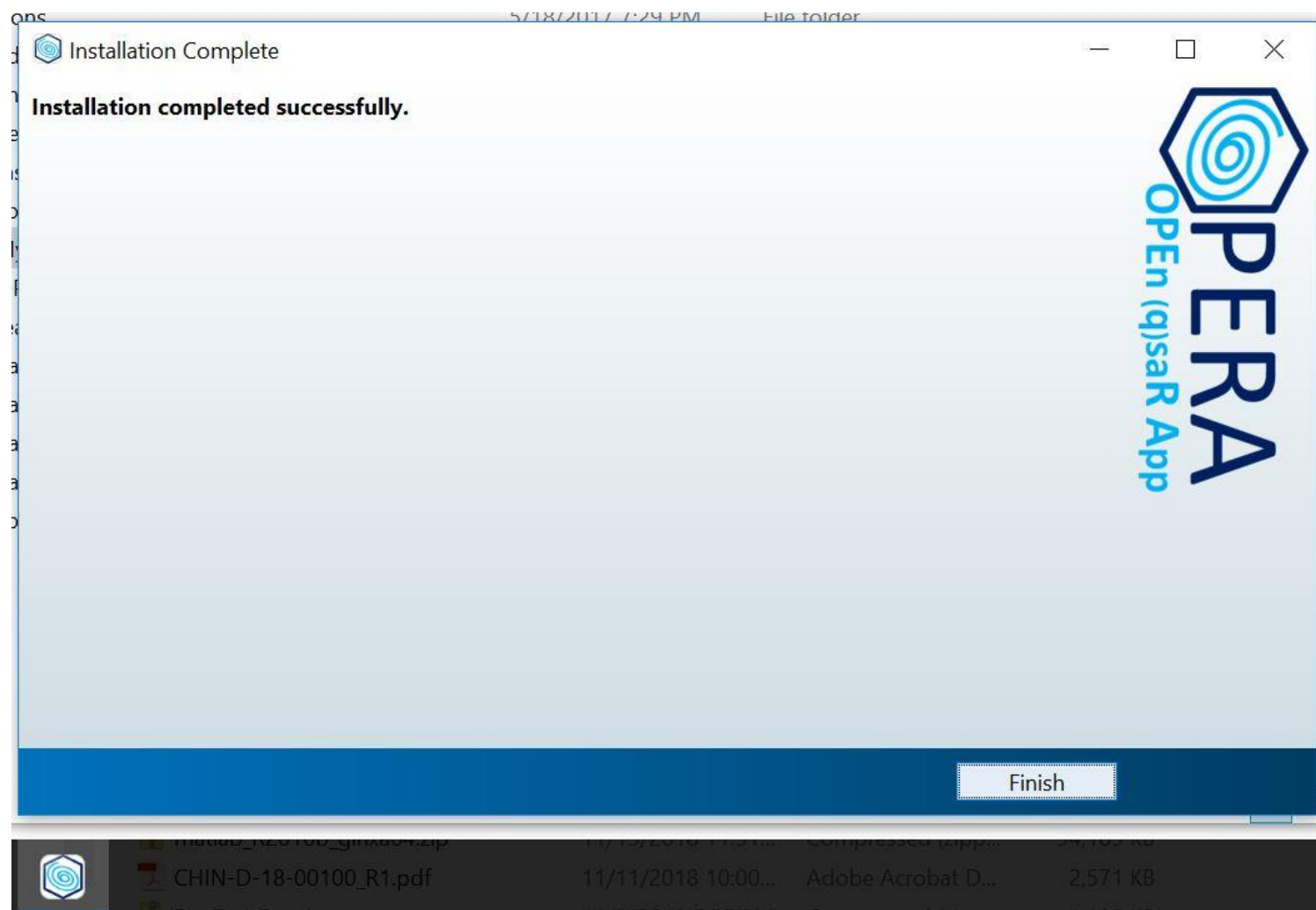| Name	Type	Size
CDKDescUI-2.0.jar	Executable Jar File	11,672 KB
desc_fp.xml	XML File	1 KB
icon.ico	Icon	7 KB
OPERA.exe	Application	19,279 KB
padel-full-1.00.jar	Executable Jar File	30,943 KB
splash.png	PNG File	72 KB

Runtime files

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

Now copy that shortcut to your desktop.
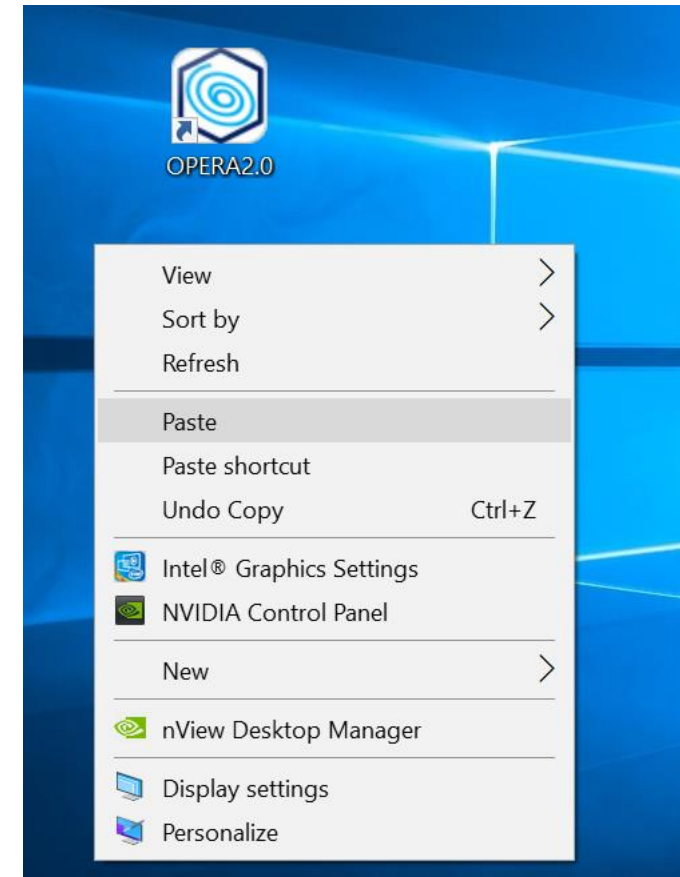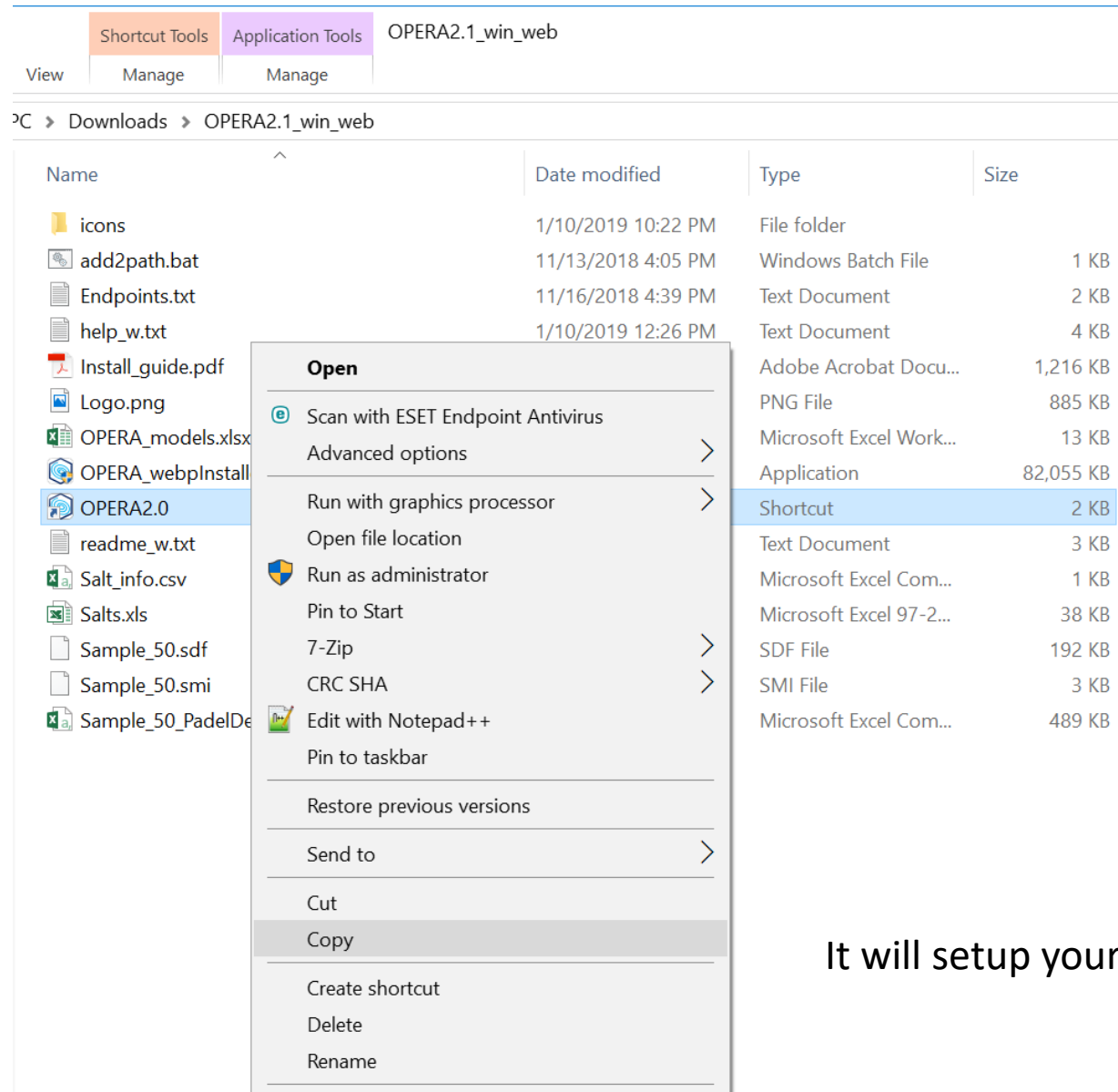
It will setup your environment variable and run the application in command line.

Double click the shortcut to run the app
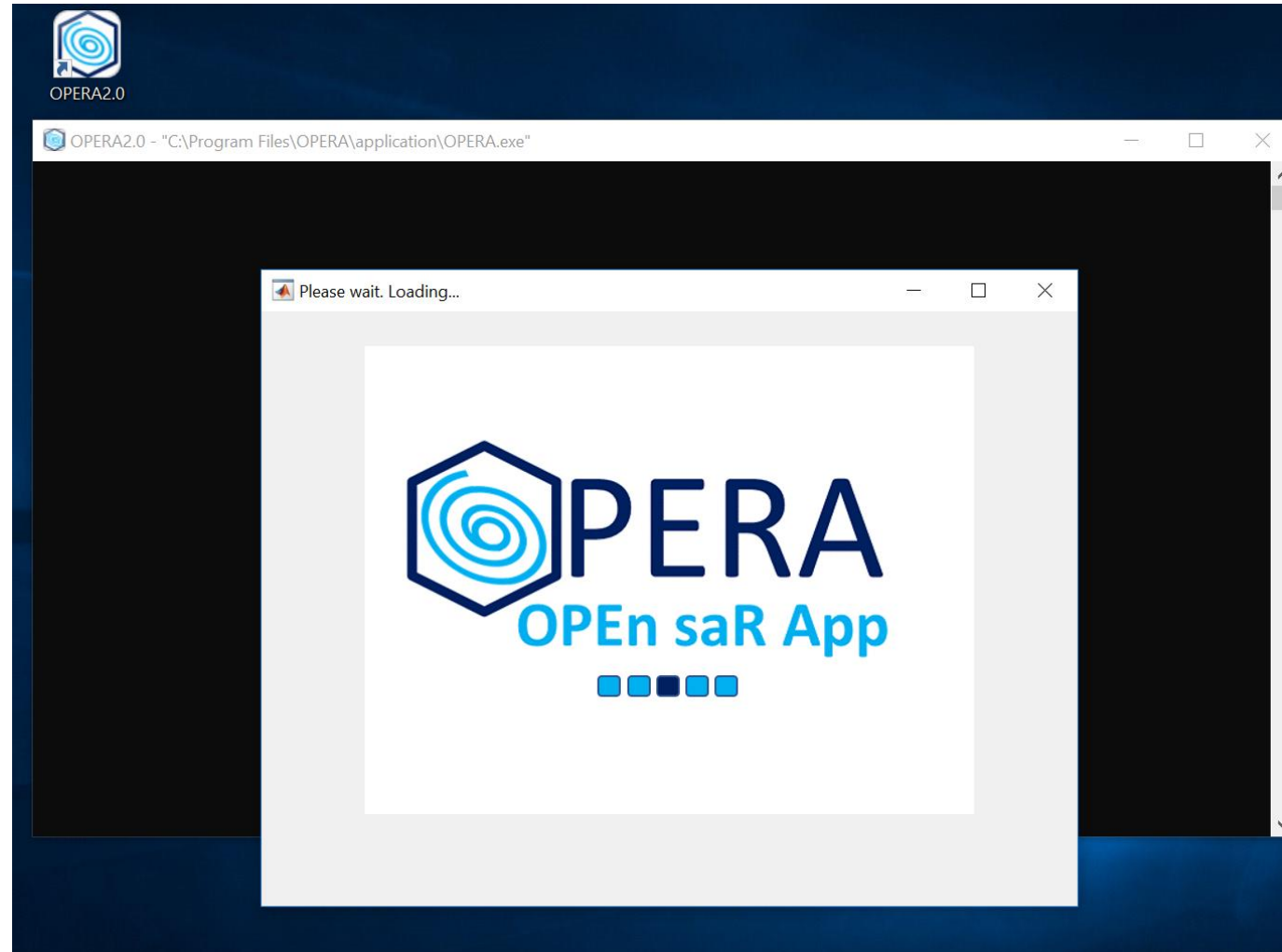
Ready to run the models

Change directory "cd" to your working 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
```

OPERA2.0

```
|                                                                    |
|     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
```
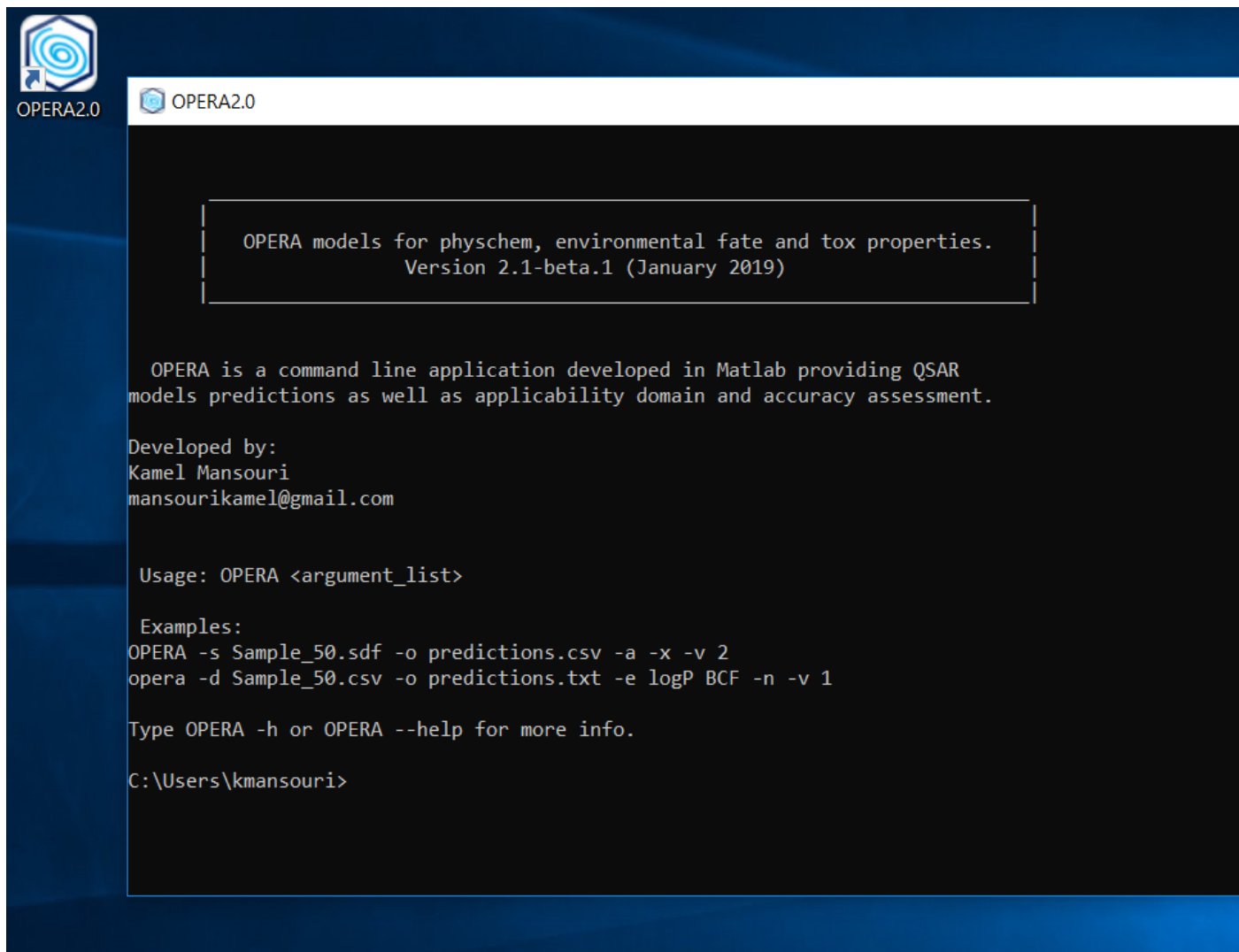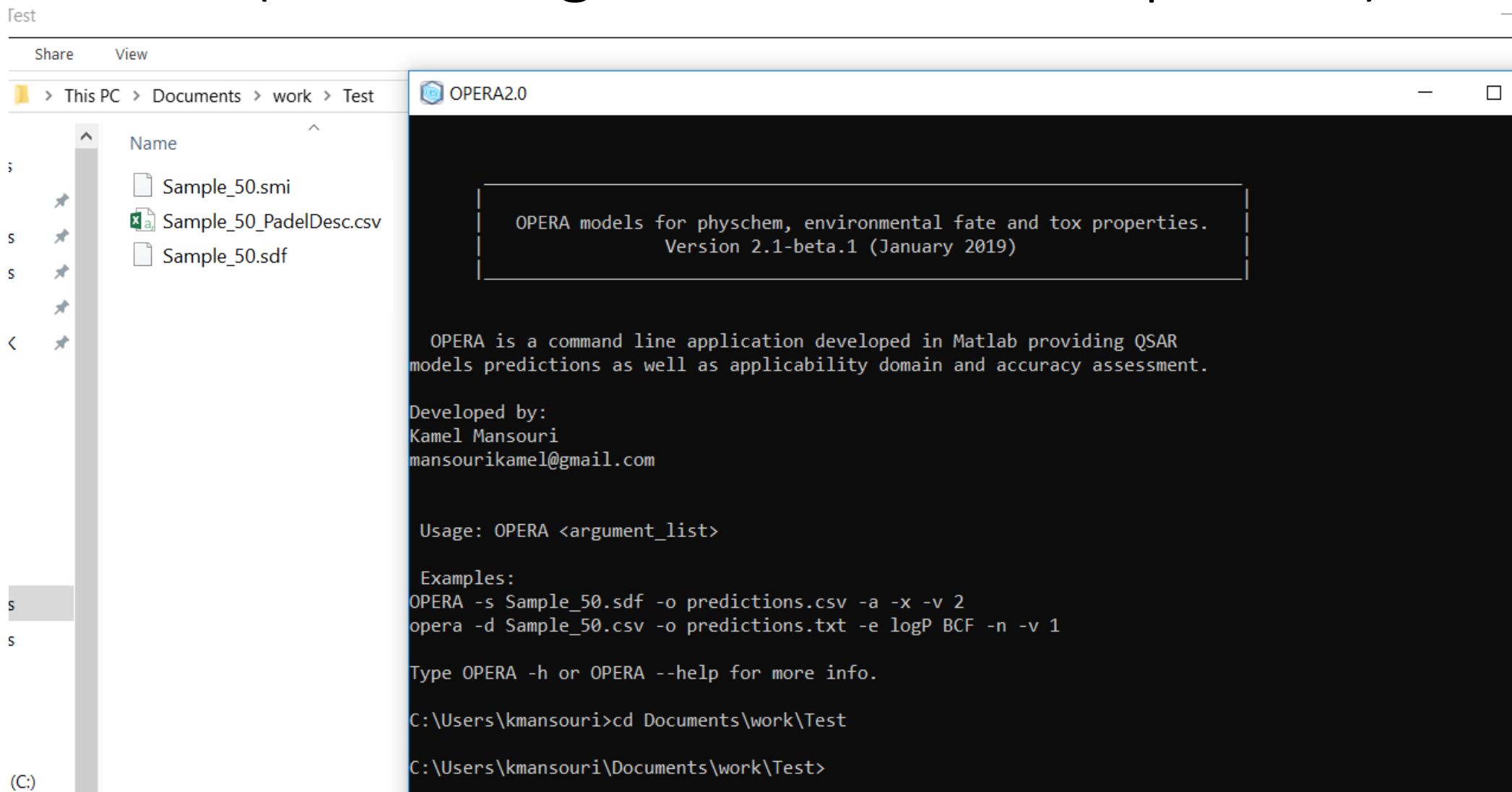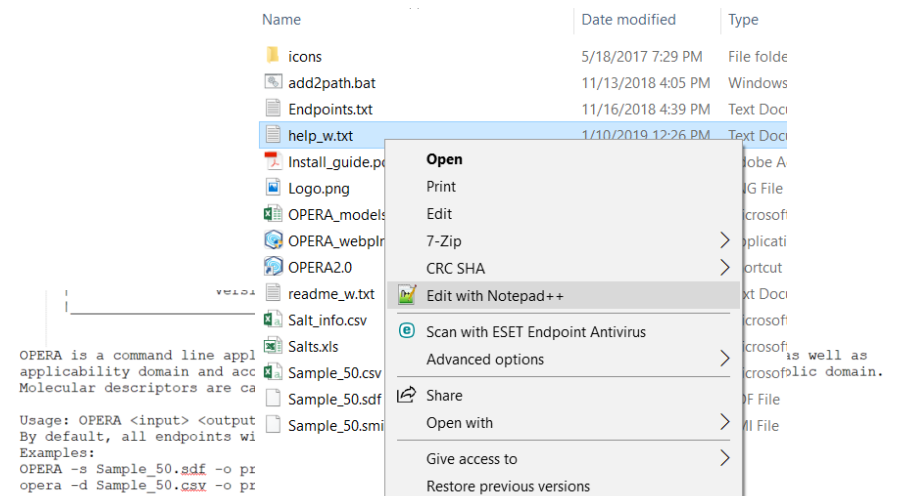
Name	Date modified	Type
icons	5/18/2017 7:29 PM	File folde
add2path.bat	11/13/2018 4:05 PM	Windows
Endpoints.txt	11/16/2018 4:39 PM	Text Doc
help_w.txt	1/10/2019 12:26 PM	Text Doc
Install_guide.p		obe A
Logo.png		G File
OPERA_models		crosof
OPERA_webpln		plicati
OPERA2.0		ortcut
readme_w.txt		t Doc
Salt_info.csv		crosof
Salts.xls		crosof
Sample_50.csv		crosof
Sample_50.sdf		F File
Sample_50.smi		I File

Open
Print
Edit
7-Zip
CRC SHA
Edit with Notepad++
Scan with ESET Endpoint Antivirus
Advanced options
Share
Open with
Give access to
Restore previous versions

Accepted Input files:

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

SDF/Mol file:
- Extension .sdf/.mol
- Different blocks
- Single or multiple molecules

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

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.

```
:>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).

Sample output:

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.

```
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)...
---------- Toxcity Endpoints ----------
Predicting Estrogen Receptor Activity (CERAPP)...
Predicting Androgen Receptor Activity (CoMPARA)...
Predicting Acute Oral Tox. endpoints (CATMoS)...

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

	S LogP_prec	T AD_LogP	U AD_index	V Conf_inde	W MP_pred	X AD_MP	Y AD_index	Z Conf_inde	AA BP_pred	AB AD_BP	AC AD_index	AD Conf_inde L
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

predictions

OPERA prediction report on the EPA Dashboard

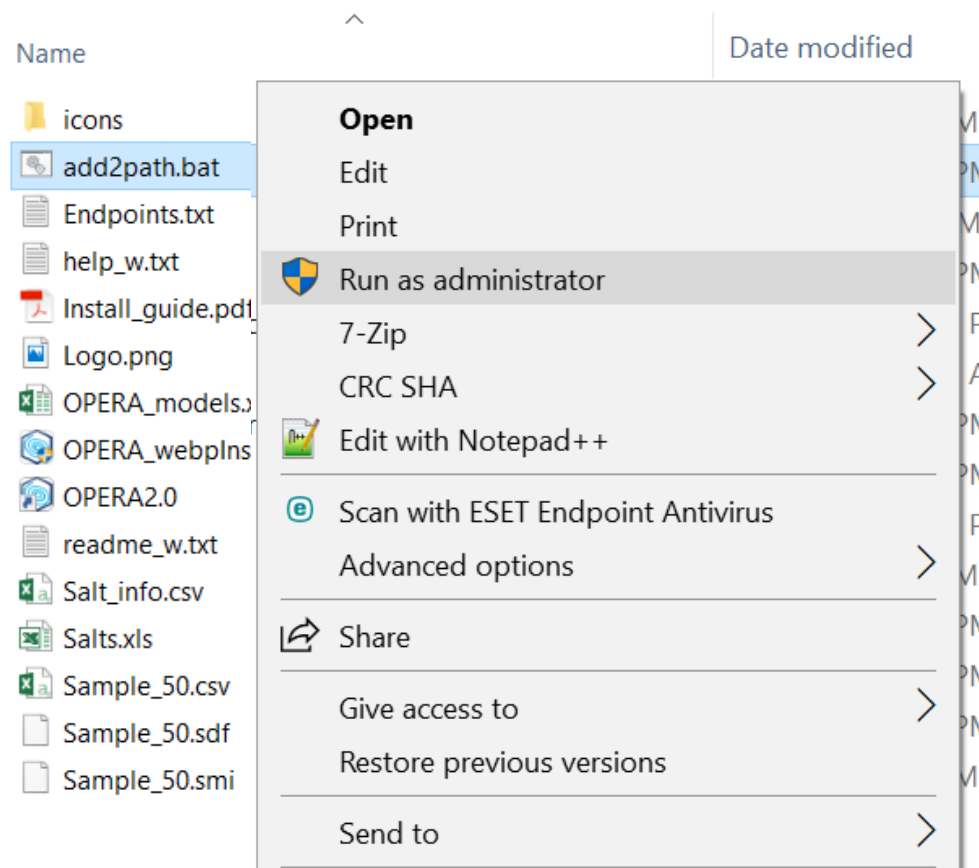**https://comptox.epa.gov/dashboard**
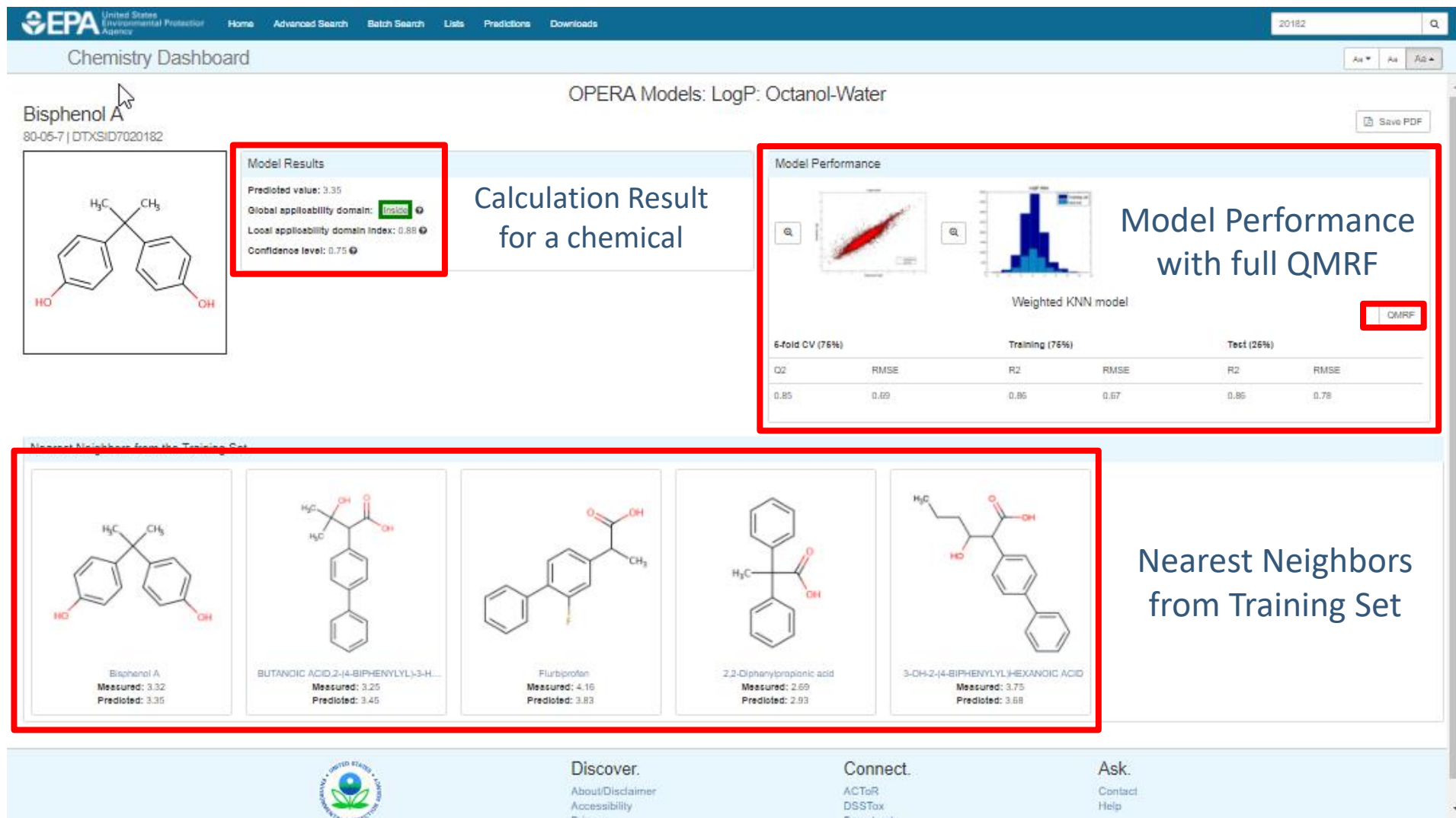
Optional and only for expert users:

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

Your feedback will be appreciated!

kmansouri@ils-inc.com

mansourikamel@gmail.com