

A close-up photograph of a person's finger pointing at a digital screen. The screen displays a line graph with a red trend line and blue data points. The background is dark and out of focus.

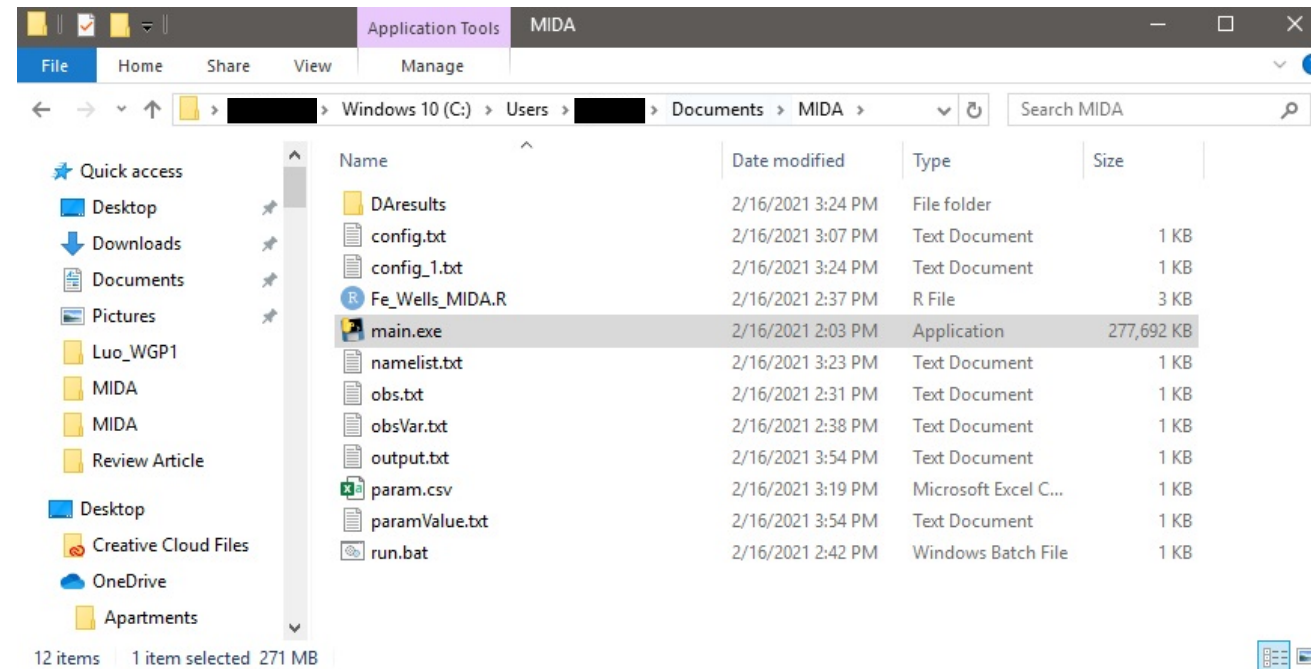
# Using MIDA for Data Assimilation

# MIDA setup

MIDA will need the following files to operate:

1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable

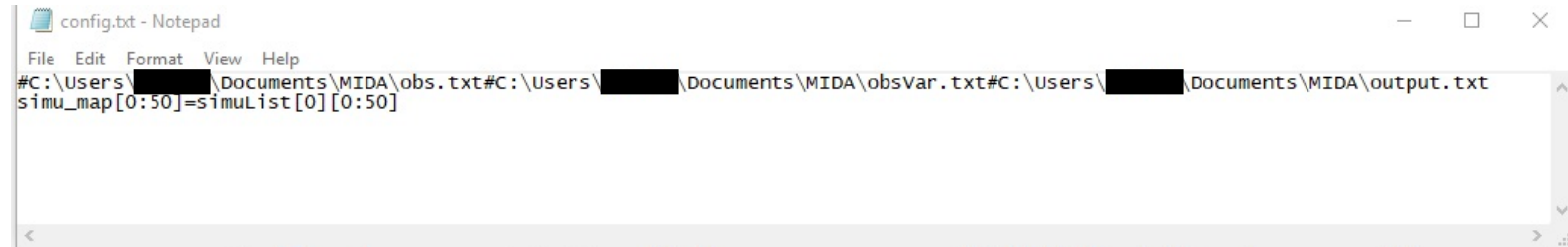
Everything else in the folder is created by MIDA



# MIDA config file

MIDA will need the following files to operate:

1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable



```
config.txt - Notepad
File Edit Format View Help
#C:\Users\...\Documents\MIDA\obs.txt#C:\Users\...\Documents\MIDA\obsVar.txt#C:\Users\...\Documents\MIDA\output.txt
simu_map[0:50]=simuList[0][0:50]
```

There are three # symbols in the first line (no spaces):

First #: file location of observations

Second #: file location of observation variance

Third #: file location of output file from executable model

The second line maps your observations to your output:

- When using the above example, the length of obs.txt and output.txt must be the same
- The observations in each text file should be in same order
- Other more complicated mapping is possible

To think about:

- The obs.txt and obsVar.txt are created by hand from your data
- The output.txt file should be output from your executable code
- Make sure they align in length and observations if this simple mapping is used

Simu\_map is where you would develop your mapping function(s)

# MIDA Parameter csv

MIDA will need the following files to operate:

1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable

To think about:

- \* indicates a column that must be filled
- Min/Max can be created arbitrarily (e.g. default\*0.5) or based on your knowledge
- Is used by MIDA, but not used by your model executable
- File must be saved as MSDOS-type csv (not UTF-8 or others)

	A	B	C	D	E	F	G	H	I
1	No.	*Name	*DA or no	*Default	*Min	*Max	Full Name	Unit	Reference
2	1	c1	1	1	1.00E-01	2.00E+00			
3	2	c2	1	0	0.00E+00	0.00E+00			
4	3	c3	1	0.29	2.90E-02	5.80E-01			
5	4	c4	1	0.7	7.00E-02	1.40E+00			
6	5	c5	1	0.00008	8.00E-06	1.60E-04			
7	6	c6	1	0.04	4.00E-03	8.00E-02			
8	7	c7	1	0.000001	1.00E-07	2.00E-06			
9	8	c8	1	0.005	5.00E-04	1.00E-02			
10	9	c9	1	0.009	9.00E-04	1.80E-02			
11	10	c10	1	0.000001	1.00E-07	2.00E-06			
12	11	c11	1	0.75	7.50E-02	1.50E+00			
13	12	c12	1	0.95	9.50E-02	1.90E+00			
14	13	c13	1	0.7	7.00E-02	1.40E+00			
15	14	c14	1	1	1.00E-01	2.00E+00			

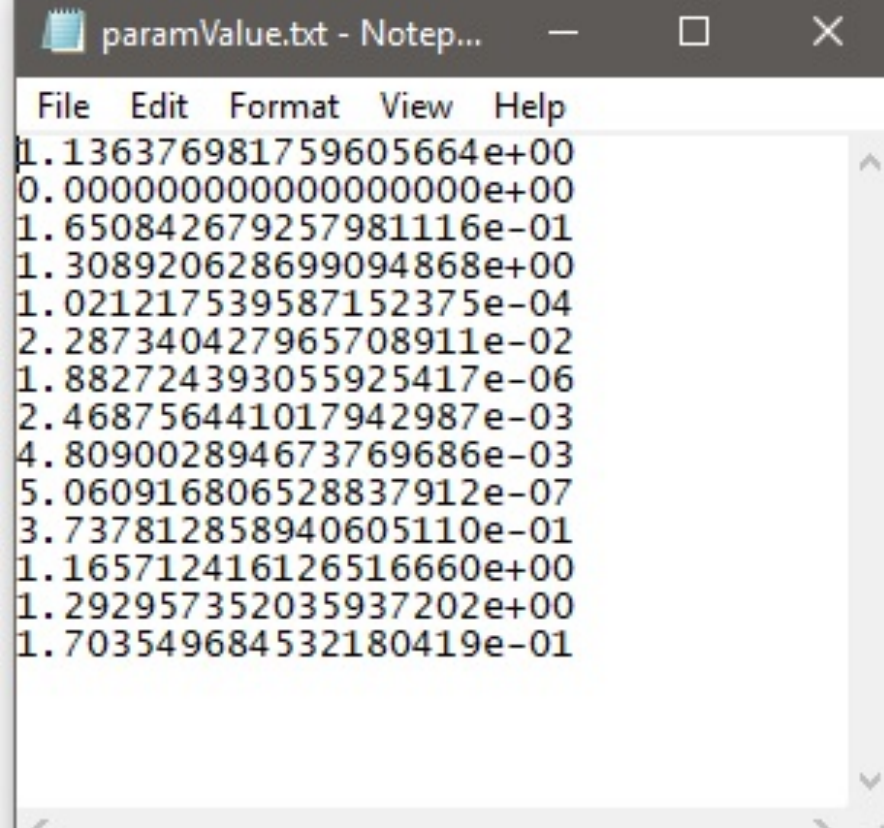
The param.csv file describes your parameters:

1. No.: The arbitrary parameter number (should align with executable)
2. Name: Name of parameter (can be arbitrary)
3. DA: Controls if parameter is included in DA – 1 for yes, 0 for no
4. Default: The starting parameter values for your model
5. Min: Minimum allowable parameter value
6. Max: Maximum allowable parameter value
7. Fullname: full description of parameter
8. Unit
9. Reference

# MIDA Parameter txt

MIDA will need the following files to operate:

1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable



```
File Edit Format View Help
1.136376981759605664e+00
0.000000000000000000e+00
1.650842679257981116e-01
1.308920628699094868e+00
1.021217539587152375e-04
2.287340427965708911e-02
1.882724393055925417e-06
2.468756441017942987e-03
4.809002894673769686e-03
5.060916806528837912e-07
3.737812858940605110e-01
1.165712416126516660e+00
1.292957352035937202e+00
1.703549684532180419e-01
```

To think about:

- Must align with parameters in model executable
- Should be simple list, no header, no column names
- When running, MIDA outputs parameters to this file that are then ran through your model executable

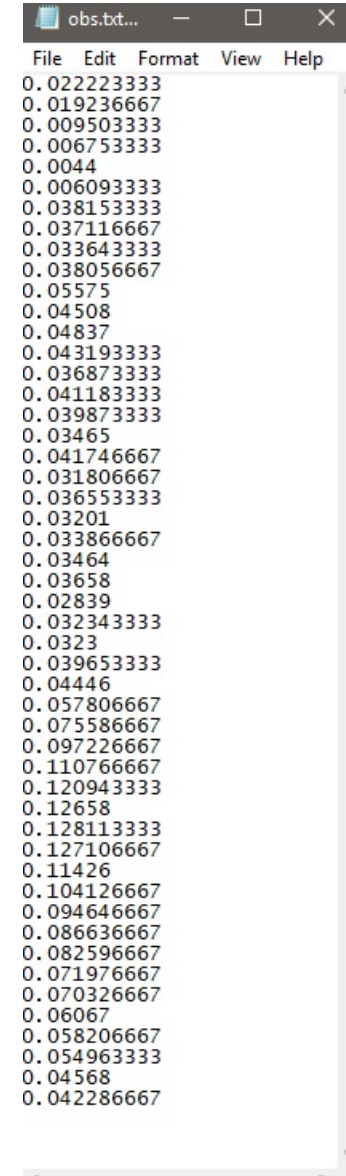
# MIDA Observation files

MIDA will need the following files to operate:

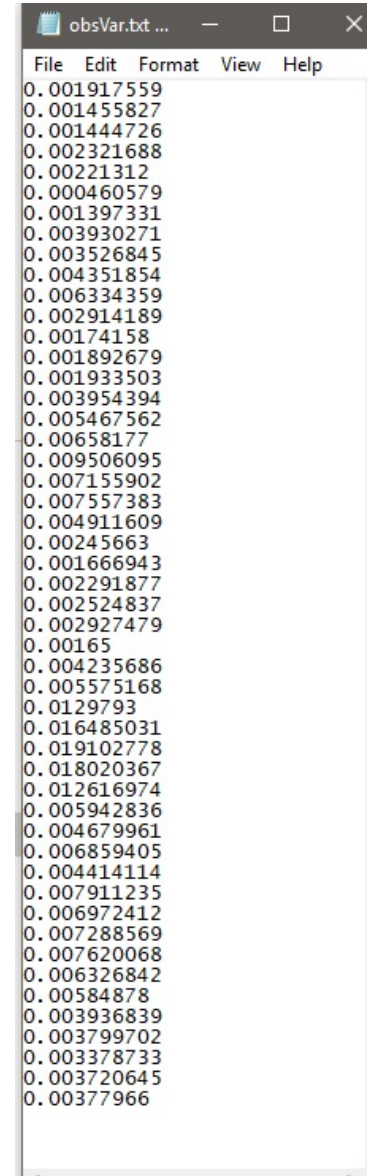
1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable

To think about:

- obs and obsVar must align
- Should be simple list, no header, no column names
- When running, MIDA compares obs.txt and obsVar.txt with model executable results (output.txt)



```
File Edit Format View Help
0.022223333
0.019236667
0.009503333
0.006753333
0.0044
0.006093333
0.038153333
0.037116667
0.033643333
0.038056667
0.05575
0.04508
0.04837
0.043193333
0.036873333
0.041183333
0.039873333
0.03465
0.041746667
0.031806667
0.036553333
0.03201
0.033866667
0.03464
0.03658
0.02839
0.032343333
0.0323
0.039653333
0.04446
0.057806667
0.075586667
0.097226667
0.110766667
0.120943333
0.12658
0.128113333
0.127106667
0.11426
0.104126667
0.094646667
0.086636667
0.082596667
0.071976667
0.070326667
0.06067
0.058206667
0.054963333
0.04568
0.042286667
```



```
File Edit Format View Help
0.001917559
0.001455827
0.001444726
0.002321688
0.00221312
0.000460579
0.001397331
0.003930271
0.003526845
0.004351854
0.006334359
0.002914189
0.00174158
0.001892679
0.001933503
0.003954394
0.005467562
0.00658177
0.009506095
0.007155902
0.007557383
0.004911609
0.00245663
0.001666943
0.002291877
0.002524837
0.002927479
0.00165
0.004235686
0.005575168
0.0129793
0.016485031
0.019102778
0.018020367
0.012616974
0.005942836
0.004679961
0.006859405
0.004414114
0.007911235
0.006972412
0.007288569
0.007620068
0.006326842
0.00584878
0.003936839
0.003799702
0.003378733
0.003720645
0.00377966
```

# MIDA model executable

MIDA will need the following files to operate:

1. A config file (e.g. config.txt)
2. A parameter description file (e.g. param.csv)
3. A parameter file for execution (e.g. paramValue.txt)
4. A list of observation values (e.g. obs.txt)
5. A list of observation variance (e.g. obsVar.txt)
  - This is optional
6. Model executable file (e.g. Fe\_Wells\_MIDA.R)
7. If using R, a .bat file (e.g. run.bat)
  - The bat file runs the Rscript as an executable

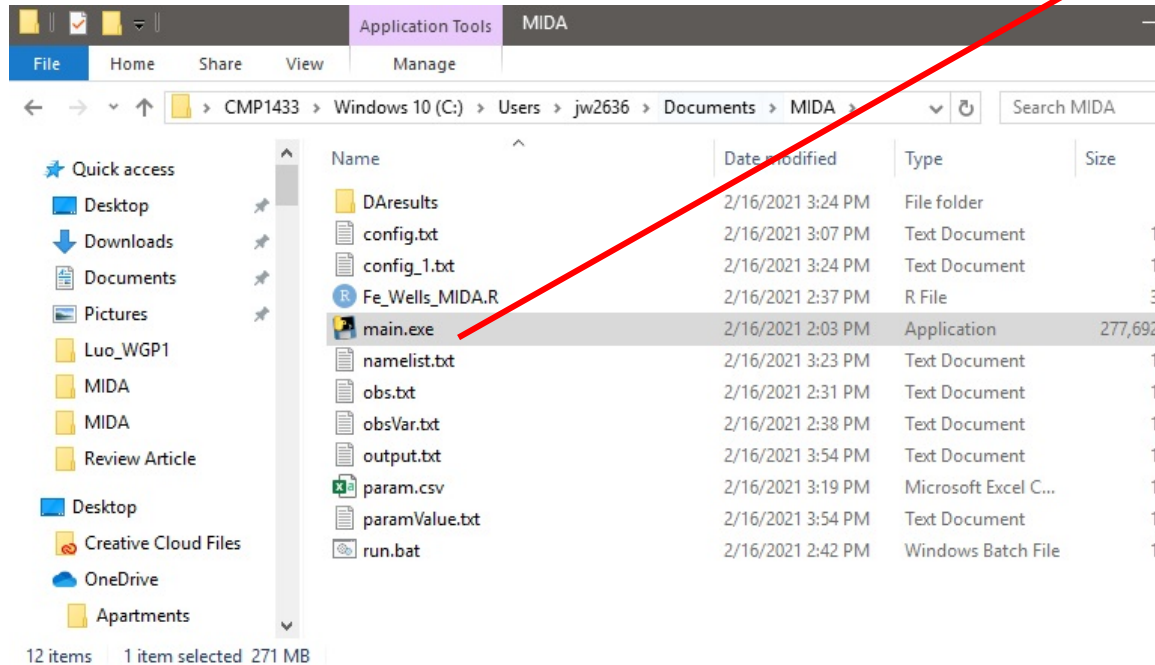
To think about:

- Model executable must take par input and output model results
- Rscript is called by a run.bat (create run.txt file and then change file name)
- Bat file should contain "Rscript.exe XXXXXX.R" where x is file name
- When running, MIDA sends parameters to model and receives the modeled output from the model executable

```
Fe_Wells_MIDA.R
1 #read parameters
2 pars <- read.table("./paramValue.txt")
3 pars <- pars$V1
4
5 #build empty arrays for data
6 c1 <- array(NA, dim=c(330,1))
7 c2 <- array(NA, dim=c(330,1))
8 c3 <- array(NA, dim=c(330,1))
9 f1 <- array(NA, dim=c(330,1))
10 f2 <- array(NA, dim=c(330,1))
11 f3 <- array(NA, dim=c(330,1))
12 R <- array(NA, dim=c(329,1))
13
14 #start and end days
15 start_day <- 1
16 end_day <- 329
17
18 #set initial C
19 cp1 = pars[1]
20 cp2 = pars[2]
21 cp3 = 1-cp1-cp2
22 total_lignin = 38
23 c1[1]=cp1*total_lignin
24 c2[1]=cp2*total_lignin
25 c3[1]=cp3*total_lignin
26
27 #set initial Fe
28 fp1 = pars[3]
29 fp2 = pars[4]
30 fp3 = 1-fp1-fp2
31 total_fe = 5.1
32 f1[1]=fp1*total_fe
33 f2[1]=fp2*total_fe
34 f3[1]=fp3*total_fe
35
36 #Parameters
37 k1=pars[5]
38 k2=pars[6]
39 k3=pars[7]
40 r1=pars[8]
41 r2=pars[9]
42 r3=pars[10]
43 a21m=pars[11]
44 a21f=pars[12]
45 a32=pars[13]
46 b32=pars[14]
47
48 #model to iterate
49 fe_mod <-function(start_day,end_day,
50                   c1,c2,c3,f1,f2,f3,R,
51                   k1,k2,k3,r1,r2,r3,
52                   a21m,a21f,a32,b32){
53
54   #calculate matrix at each step
55   for (i in start_day:end_day){
56
57     #linking functions
58     T21=a21m*k1+a21f*r1*f1[i]
```

```
run.txt - Notepad
File Edit Format View Help
Rscript.exe Fe_Wells_MIDA.R
```

# MIDA Run



DAModule - A Generic Module for Data Assimilation

Help

### Preparation of Data Assimilation

The number of simulations:  Select Work Path:

	min	max	default
1			
2			
3			
4			
5			
6			

Load Files:

Observation File List

	file name
1	
2	
3	
4	
5	
6	

Observation Variance File List

	file name
1	
2	
3	
4	
5	
6	

Simulation Output File List

	file name
1	
2	
3	
4	
5	
6	

(Optional) Gelman-Rubin convergence test:

### Execution of Data Assimilation

Load Namelist File:

Choose variables to be print in DA: ☒ total mismatch ☒ acceptance rate ☐ delta\_mismatch ☐ mismatch for each obs ☐ obs var

# MIDA Run

Preparation of Data Assimilation (only needed once to create namelist.txt file):

1. Assign number of simulations
2. Find location of MIDA folder
3. Load parameter description (csv file)
4. Load config.txt file location
5. Save to namelist file

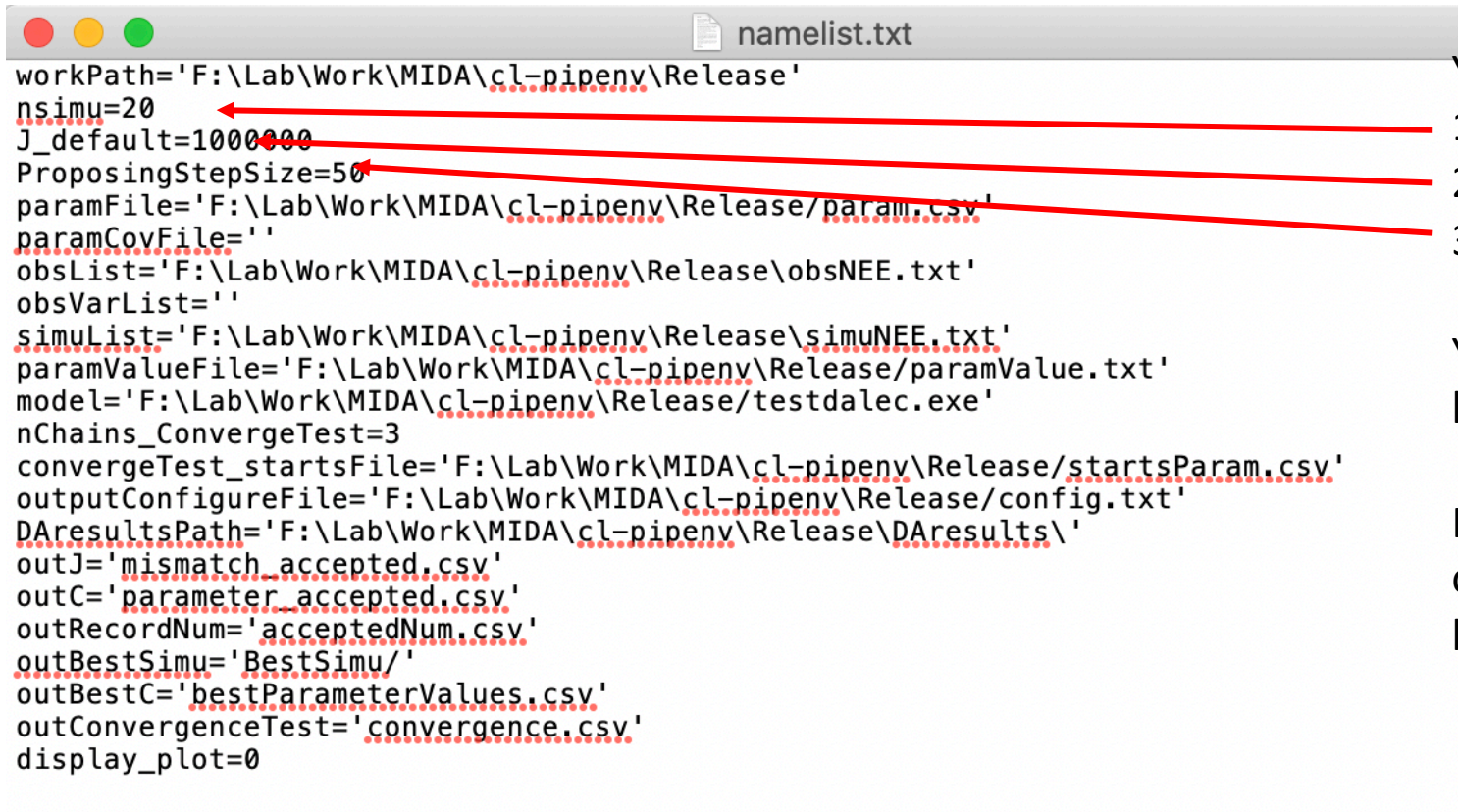
Execution of Data Assimilation:

1. Select namelist file (created by previous steps)
2. Run Data Assimilation

The screenshot shows the DAmodule interface with the following sections and fields:

- Preparation of Data Assimilation**
  - The number of simulations**: A text input field with a red arrow pointing to it from step 1.
  - Select Work Path**: A text input field with a red arrow pointing to it from step 2.
  - Choose A Directory**: A button.
  - Load Parameter Range**: A table with columns 'min', 'max', and 'default'. A red arrow points to it from step 3.
  - (Optional) Load Parameter Covariance**: A text input field.
  - Load Model Executable File**: A text input field.
  - Load Output Configuration File**: A button. A red arrow points to it from step 4.
- Observation File List**: A table with column 'file name'.
- Observation Variance File List**: A table with column 'file name'.
- Simulation Output File List**: A table with column 'file name'.
- (Optional) Gelman-Rubin convergence test**: A section with a **Choose Different Startpoints** button and a text input field.
- 0. Save to Namelist File**: A button.
- Execution of Data Assimilation**
  - Load Namelist File**: A text input field with a red arrow pointing to it from step 1.
  - Choose variables to be print in DA**: Checkboxes for ☒ total mismatch, ☒ acceptance rate, ☐ delta\_mismatch, ☐ mismatch for each obs, and ☐ obs var.
  - 1. Run Data Assimilation**: A button with a red arrow pointing to it from step 2.
  - 2. Generate Plots**: A button.

# MIDA controls



```
workPath='F:\Lab\Work\MIDA\cl-pipenv\Release'
nsimu=20
J_default=1000000
ProposingStepSize=50
paramFile='F:\Lab\Work\MIDA\cl-pipenv\Release\param.csv'
paramCovFile=''
obsList='F:\Lab\Work\MIDA\cl-pipenv\Release\obsNEE.txt'
obsVarList=''
simuList='F:\Lab\Work\MIDA\cl-pipenv\Release\simuNEE.txt'
paramValueFile='F:\Lab\Work\MIDA\cl-pipenv\Release\paramValue.txt'
model='F:\Lab\Work\MIDA\cl-pipenv\Release\testdalec.exe'
nChains_ConvergeTest=3
convergeTest_startsFile='F:\Lab\Work\MIDA\cl-pipenv\Release\startsParam.csv'
outputConfigureFile='F:\Lab\Work\MIDA\cl-pipenv\Release\config.txt'
DAresultsPath='F:\Lab\Work\MIDA\cl-pipenv\Release\DAresults\'
outJ='mismatch_accepted.csv'
outC='parameter_accepted.csv'
outRecordNum='acceptedNum.csv'
outBestSimu='BestSimu/'
outBestC='bestParameterValues.csv'
outConvergenceTest='convergence.csv'
display_plot=0
```

You can edit the namelist directly to control DA:

1. Change MCMC iterations
2. Change max accepted value from cost function
3. Change MCMC parameter step size

You could also manually change file location/names of files used by MIDA if necessary

If changes are made to the namelist while MIDA is open it is recommended to restart MIDA's main.exe before data assimilation.

# MIDA plot graphing

DAmodule - A Generic Module for Data Assimilation

Help

### Preparation of Data Assimilation

The number of simulations  Select Work Path

Load Parameter Range

	min	max	default
1			
2			
3			
4			
5			
6			

Load Files:

(Optional) Load Parameter Covariance

Load Model Executable File

Load Output Configuration File

Observation File List

	file name
1	
2	
3	
4	
5	
6	

Observation Variance File List

	file name
1	
2	
3	
4	
5	
6	

Simulation Output File List

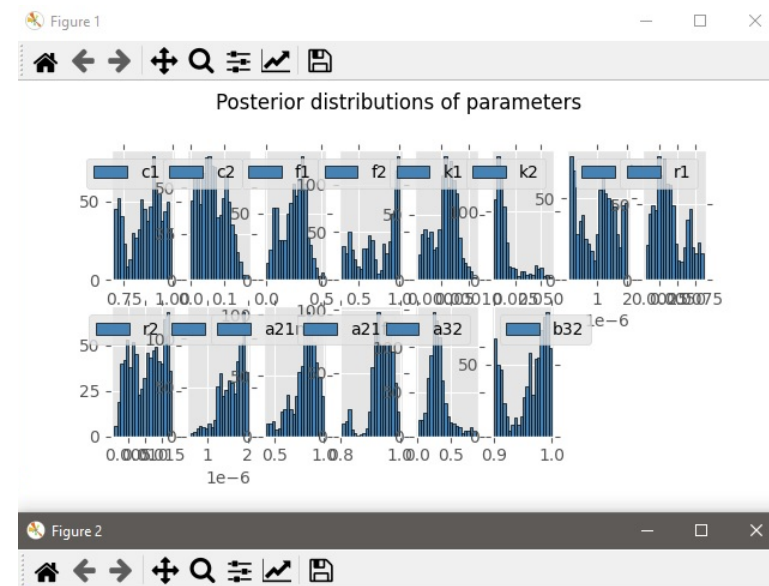
	file name
1	
2	
3	
4	
5	
6	

(Optional) Gelman-Rubin convergence test ☐ Choose Different Startpoints

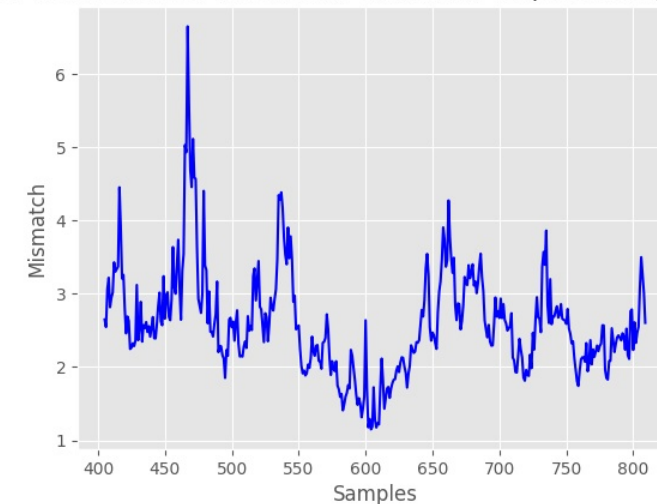
### Execution of Data Assimilation

Load Namelist File:

Choose variables to be print in DA: ☒ total mismatch ☒ acceptance rate ☐ delta\_mismatch ☐ mismatch for each obs ☐ obs var



hes between observations and simulation outputs during MCMC

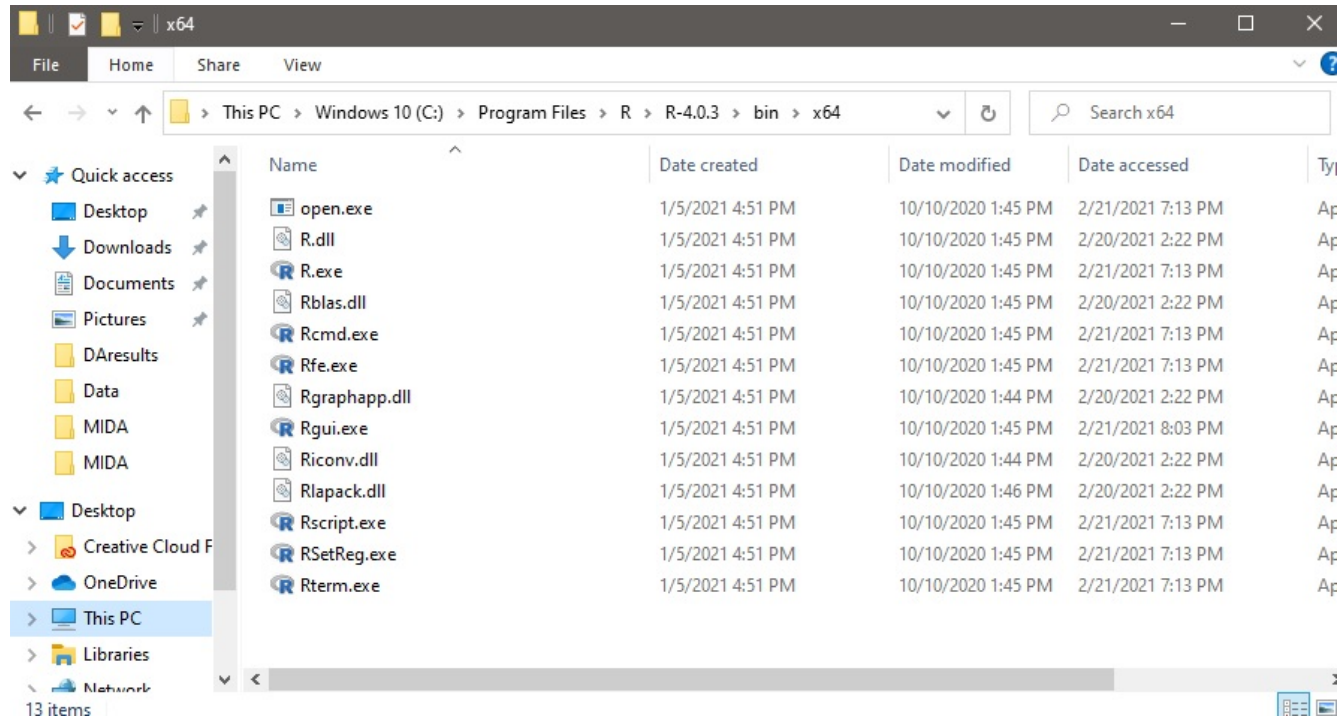


After DA, simply click Generate plots to look at the posterior distribution

A close-up photograph of a person's finger pointing at a digital screen. The screen displays a blue-toned line chart with several data points connected by lines. The background is dark and out of focus.

WG1 example setup

# Setup R path on Windows



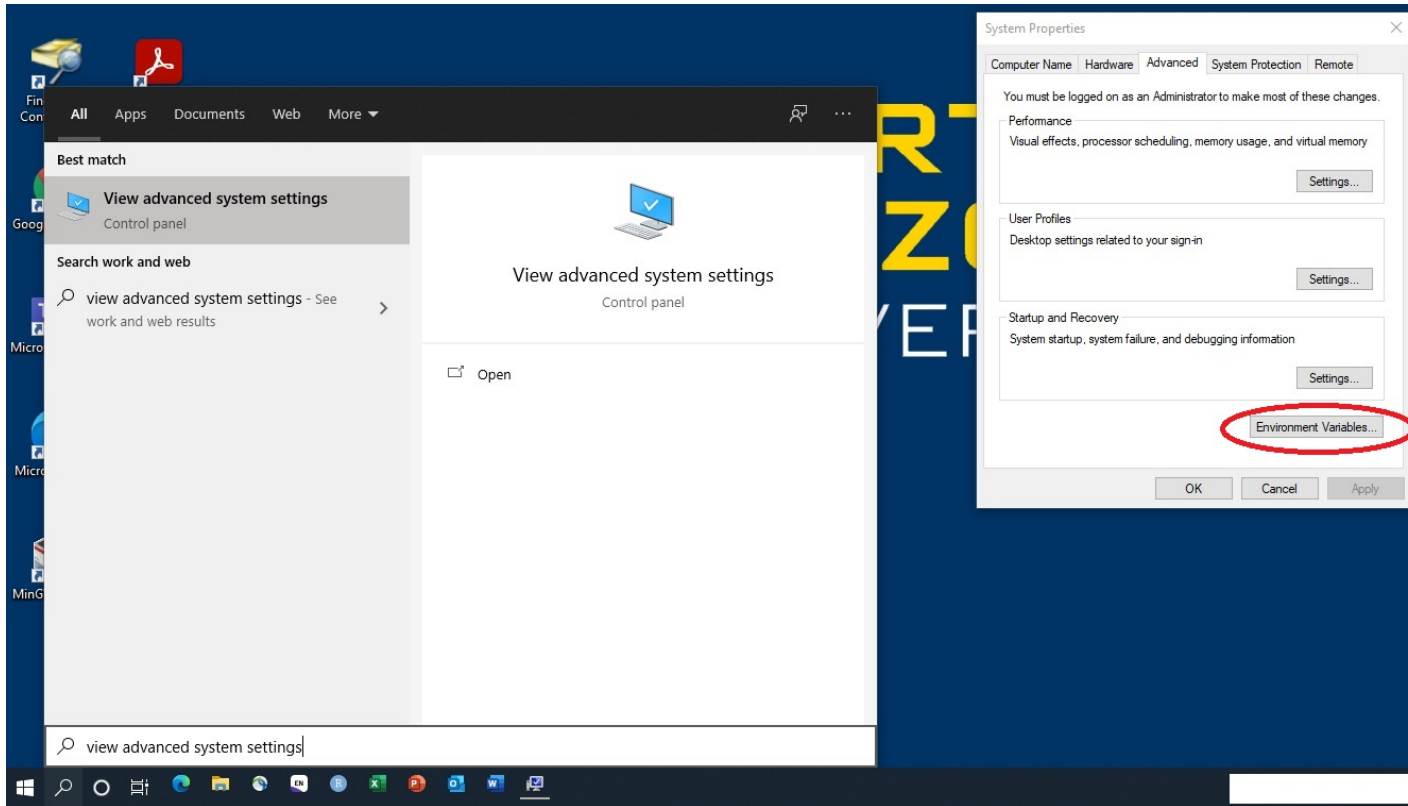
First find the x64 folder that contains your R install

Copy the folder location

It should look something like this:

`C:\Program Files\R\R-4.0.3\bin\x64`

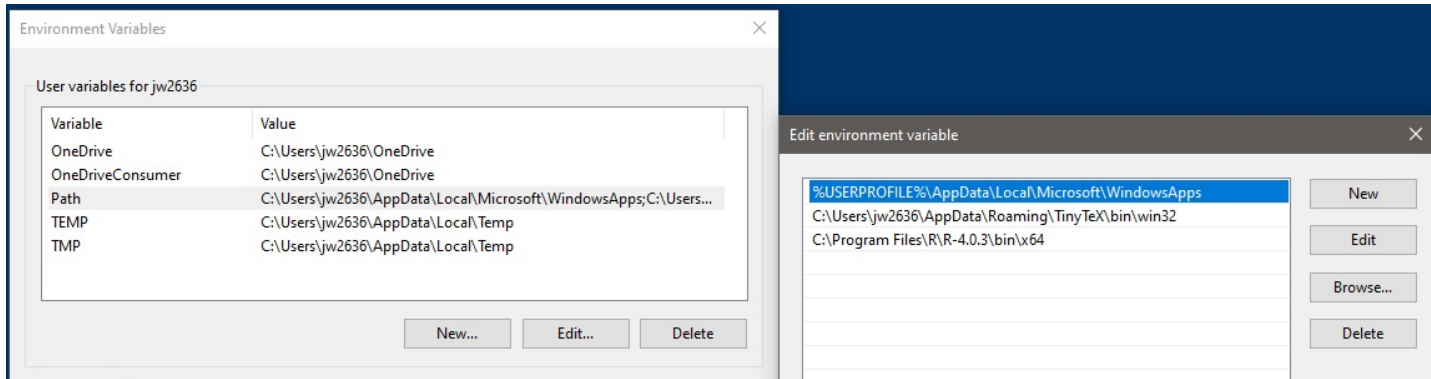
# Setup R path on Windows



Find your environmental variables:

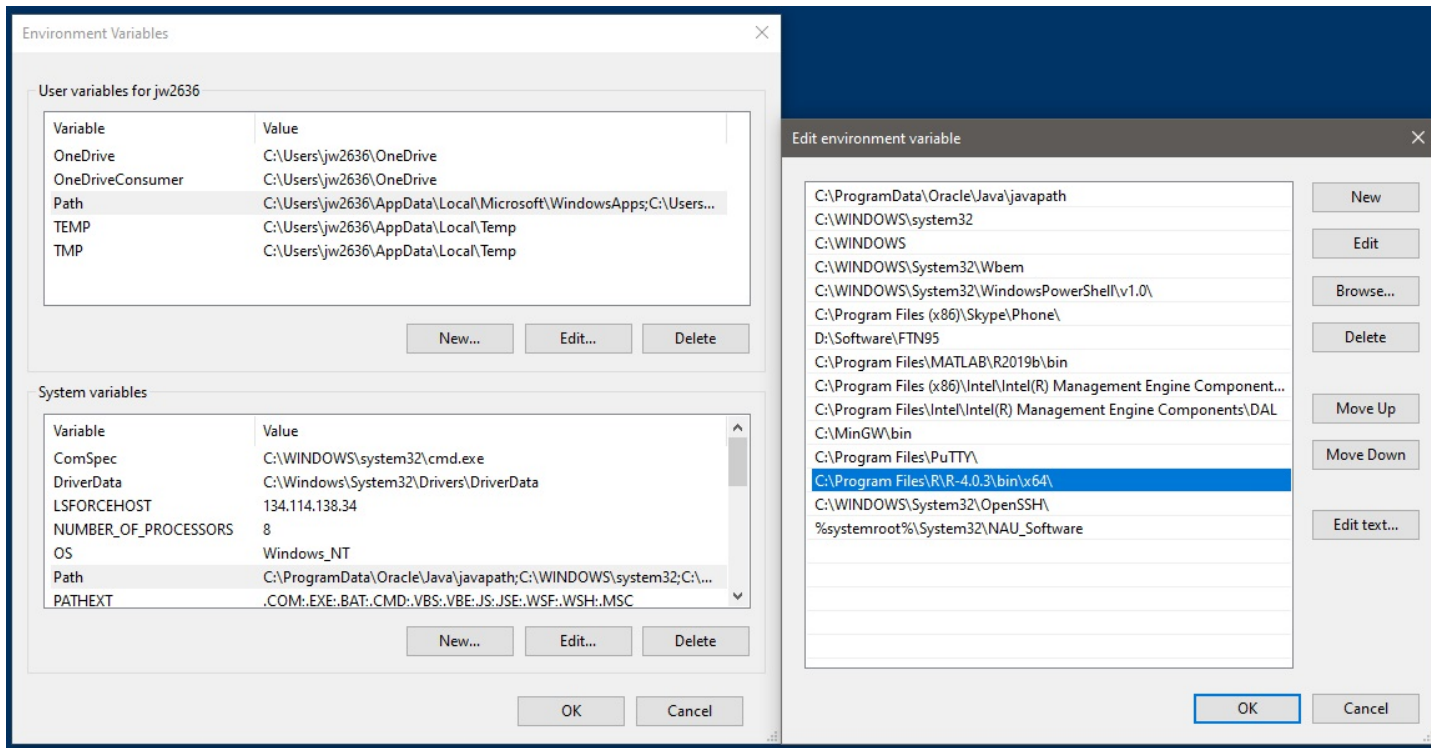
1. In the search enter “view advanced system settings”
2. When opened, click “environmental variables”

# Setup R path on Windows



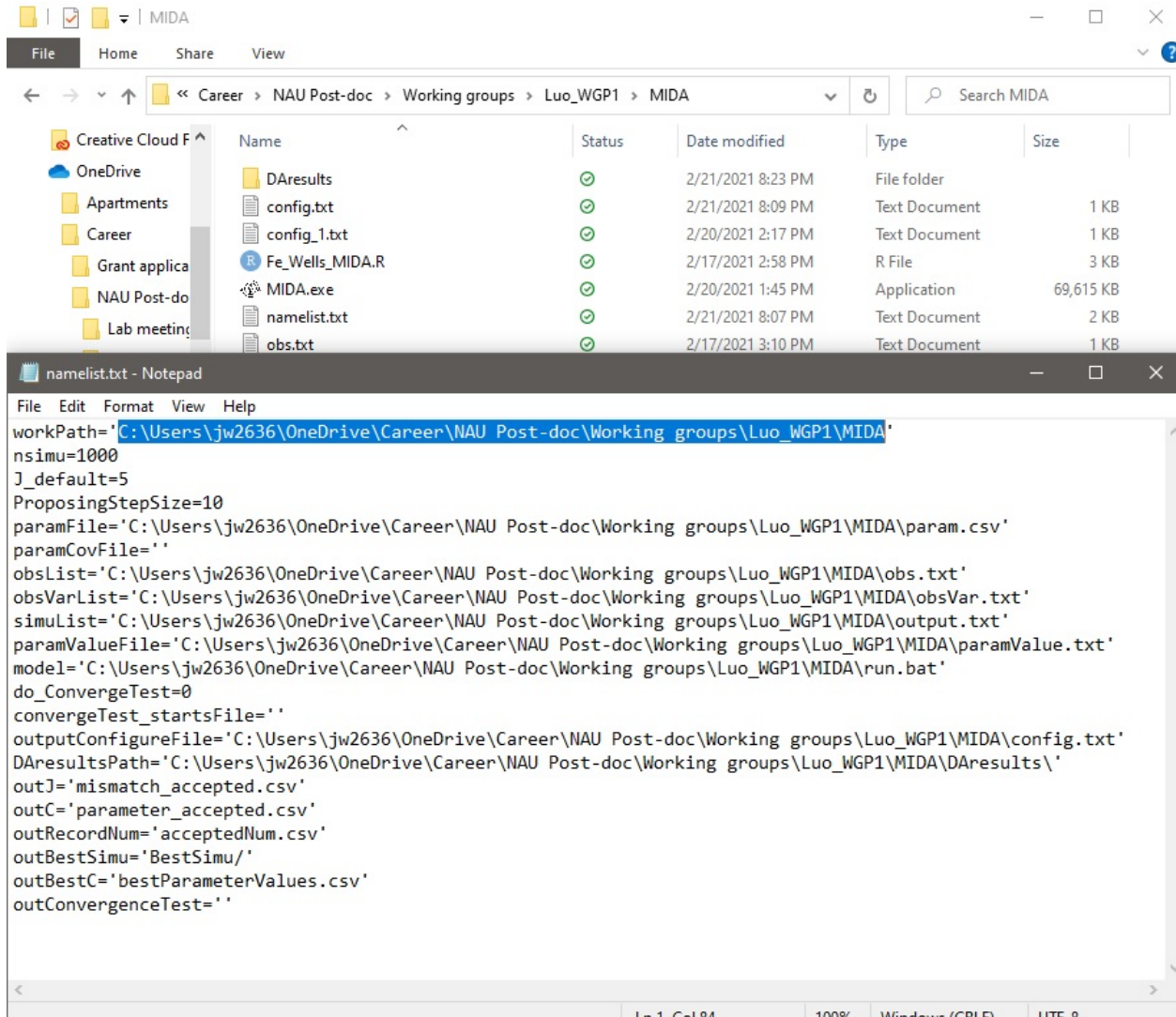
Add the R folder location to your path:

1. Click Path in User variables
2. Click Edit
3. Click New on Edit environment variable window
4. Paste R folder path (e.g. C:\Program Files\R\R-4.0.3\bin\x64)
5. Click OK



Repeat the same procedure for the system variables

# MIDA folder setup



Download WG1 MIDA example from email

1. Save/extract the MIDA folder to your desired file location
2. Open the nameList.txt file with notepad
3. Copy MIDA's new folder location and replace all instances of <code>C:\Users\jw2636\OneDrive\Career\NAU Post-doc\Working groups\Luo\_WGP1\MIDA</code>
4. Alternatively, replace the MIDA folder location in the config.txt file and generate a new nameList file using MIDA (which you'll have to do if you plan on changing the mapping functions, so might as well update the folder locations in both files)

From there you can now open MIDA.exe, select nameList.txt and test DA

You can edit simulation numbers, acceptance cutoffs, and step-size as described in the previous section

You can also change your model executable file to test other model structures. Simply add the R file to the folder, change the file name in nameList.txt, and update the Rscript's name in run.bat

# Initial MIDA results

