

CALVIN & PyVIN Shortcourse Handout

Mustafa S. Dogan *

April 1, 2017

Contents

1 Prerequisites	3
1.1 Command line	4
1.1.1 Windows	4
1.1.2 Mac OS	5
2 Introduction and background	7
3 HOBES database	9
3.1 Data flow overview	9
4 Updated version: PyVIN	10
4.1 Model architecture	10
4.2 Debug mode	11
4.3 PyVIN model	11
5 Your first model run	13
5.1 Postprocessing results	13
6 References and useful links	14

*Graduate Student, Dept. of Civil and Env. Eng., Univ. of California, Davis, 1 Shields Avenue, Davis, CA 95616. E-mail: msdogan@ucdavis.edu.



Agenda and Topics

- | | |
|---------------|--|
| 15 min | – Introduction and set-up |
| 1 hr | – CALVIN theory and model introduction |
| 15 min | – Break |
| 15-20 min | – HOBBES database |
| 15 min | – Data flow overview |
| 1 hr | – Break |
| 20-25 min | – PyVIN updates and model architecture |
| 15 min | – A PyVIN example |
| 15 min | – Break |
| 15-20 min | – Required software and installation |
| 20 min | – Your first PyVIN run |
| 20-25 min | – Postprocessing results |
-

1 Prerequisites

Please bring your **laptop** with below software dependencies installed if you want a hands-on PyVIN experience. Install following software in advance since some of them, such as Anaconda, takes long time to download and install. If you have any question, e-mail msdagan@ucdavis.edu

- **Python version 3+ with Anaconda**

Anaconda is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment (Source: Wikipedia).

Link: <https://www.continuum.io/downloads>

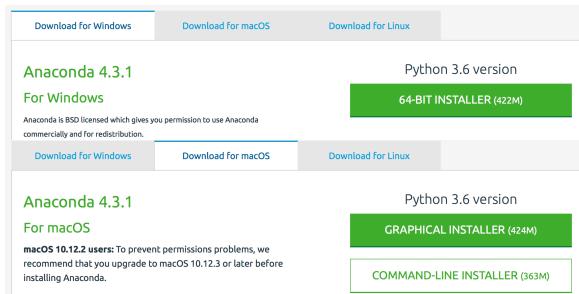


Figure 1: Anaconda package for Windows and Mac OS

Note: Download Python v3+ (not v2.7) because Pyomo command line installation works only with Python v3+

- **GitHub**

GitHub is a web-based Git or version control repository and Internet hosting service. It offers all of the distributed version control and source code management functionality of Git as well as adding its own features. It provides access control and several collaboration features such as bug tracking, feature requests, task management, and wikis for every project (Source: Wikipedia).

Link: <https://desktop.github.com>

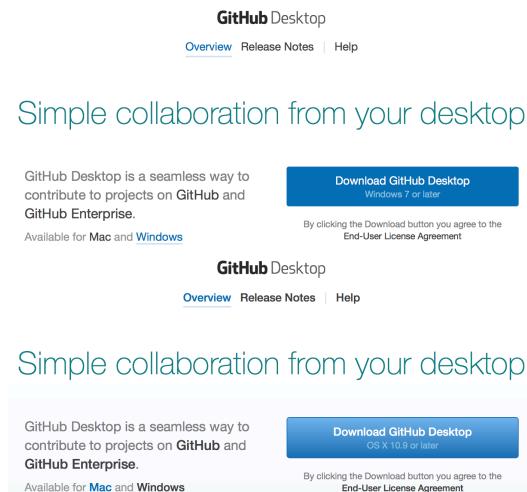


Figure 2: GitHub installation for Windows and Mac OS

- **Pyomo**

Pyomo is a Python-based, open-source optimization modeling language with a diverse set of optimization capabilities. The easiest way to install Pyomo is to type following commands to your command line tool.

Command: `conda install -c conda-forge pyomo`

Command: `conda install -c conda-forge pyomo.extras`

Command: `conda install -c conda-forge glpk`

Note: if you don't know how to use command line, please refer to description below and then type the commands.

1.1 Command line

This shortcourse is not intended to explain command line use. But I will walk you through steps enough to install Pyomo and run PyVIN model. If you like to learn basics of command line, here is a nice crash course:

<https://learnpythonthehardway.org/book/appendixa.html>

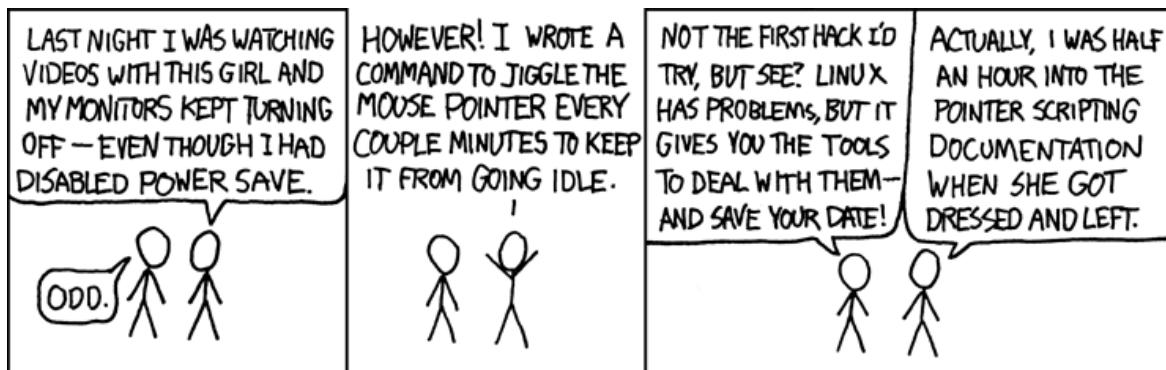


Figure 3: A command line comic from xkcd, source: <https://xkcd.com/196/>

1.1.1 Windows

After installing GitHub, double click on "Git Shell" and then type the command above to install Pyomo and GLPK solver.

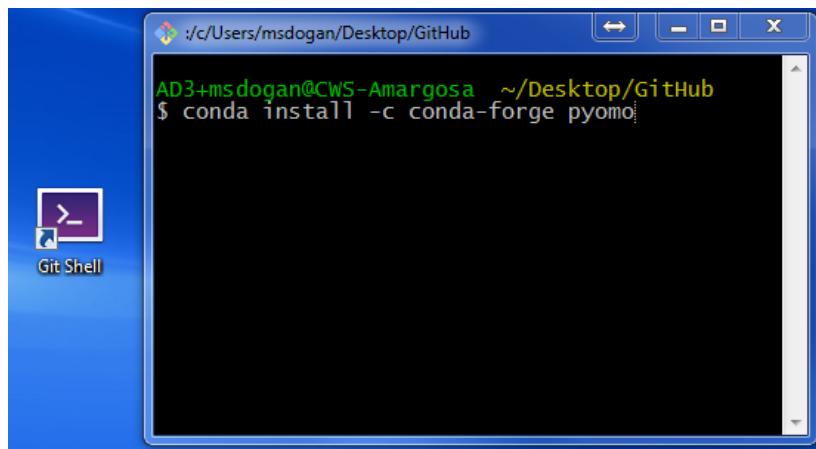


Figure 4: Command line (Git Shell) for Windows

1.1.2 Mac OS

Search "terminal" in Spotlight Search and then double click. Type the command above to install Pyomo and GLPK solver.

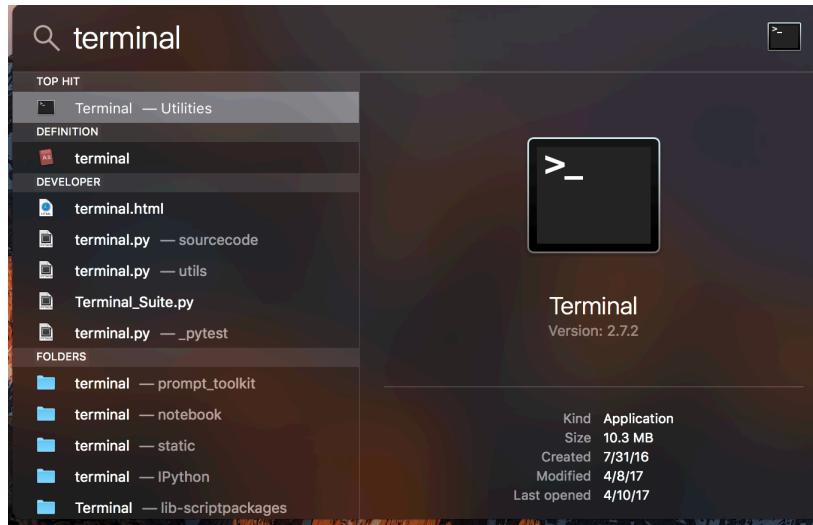


Figure 5: Command line (terminal) for Mac OS

Successful "pyomo" installation:

```

1 Mustafas-MacBook-Pro:~ msdogan$ conda install -c conda-forge pyomo
2 Fetching package metadata .....
3 Solving package specifications: .
4
5 Package plan for installation in environment /Users/msdogan/anaconda:
6
7 The following NEW packages will be INSTALLED:
8
9     appdirs:    1.4.3-py27_0  conda-forge
10    pyomo:      5.1.1-py27_0  conda-forge
11    pyutilib:   5.4.1-py27_0  conda-forge
12
13 The following packages will be SUPERSEDED by a higher-priority channel:
14
15     conda:      4.3.16-py27_0          --> 4.2.13-py27_0  conda-forge
16     conda-env:   2.6.0-0              --> 2.6.0-0        conda-forge
17
18 Proceed ([y]/n)? y
19
20 conda-env-2.6. 100% ##### Time: 0:00:00 1.11 MB/s
21 appdirs-1.4.3- 100% ##### Time: 0:00:00 50.42 kB/s
22 conda-4.2.13-p 100% ##### Time: 0:00:02 155.17 kB/s
23 pyutilib-5.4.1 100% ##### Time: 0:00:01 217.86 kB/s
24 pyomo-5.1.1-py 100% ##### Time: 0:00:07 294.78 kB/s

```

Successful "pyomo.extras" installation:

```

1 Mustafas-MacBook-Pro:~ msdogan$ conda install -c conda-forge pyomo.extras
2 Fetching package metadata .....
3 Solving package specifications: .....
4
5 Package plan for installation in environment /Users/msdogan/anaconda:
6
7 The following packages will be downloaded:
8
9     package           | build

```

```

10      _____
11      |          py27_0          117 KB
12      |          py27_1          146 KB
13      |          py27_0          14 KB  conda-forge
14      |          py27_2          109 KB  conda-forge
15      |          py27_0           1 KB  conda-forge
16
17      _____
18      |          Total:        387 KB
19
20
21 The following NEW packages will be INSTALLED:
22
23      pymysql:    0.7.9 - py27_0
24      pyodbc:     3.0.10 - py27_1
25      pyomo.extras: 3.2 - py27_0  conda-forge
26      pyro4:       4.43 - py27_2  conda-forge
27      serpent:    1.12 - py27_0  conda-forge
28
29 Proceed ([y]/n)? y
30
31 Fetching packages ...
32 pymysql-0.7.9 - 100% ##### Time: 0:00:00 312.08 kB/s
33 pyodbc-3.0.10 - 100% ##### Time: 0:00:00 288.39 kB/s
34 serpent-1.12-p 100% ##### Time: 0:00:00 116.83 kB/s
35 pyro4-4.43-py2 100% ##### Time: 0:00:00 183.68 kB/s
36 pyomo.extras-3 100% ##### Time: 0:00:00 1.33 MB/s
37 Extracting packages ...
38 [      COMPLETE      ##### 100%
39 Linking packages ...
40 [      COMPLETE      ##### 100%

```

Successful "GLPK" solver installation:

```

1 Mustafas-MacBook-Pro:~ msdogan$ conda install -c conda-forge glpk
2 Fetching package metadata .....
3 Solving package specifications: .....
4
5 Package plan for installation in environment /Users/msdogan/anaconda:
6
7 The following packages will be downloaded:
8
9      package          |          build
10
11      gmp-6.1.2          |          0          716 KB  conda-forge
12      glpk-4.61          |          0          1.0 MB  conda-forge
13
14      _____
15      |          Total:        1.7 MB
16
17
18 The following NEW packages will be INSTALLED:
19
20      glpk: 4.61-0  conda-forge
21      gmp: 6.1.2-0  conda-forge
22
23 Proceed ([y]/n)? y
24
25 Fetching packages ...
26 gmp-6.1.2-0.ta 100% ##### Time: 0:00:02 248.51 kB/s
27 glpk-4.61-0.ta 100% ##### Time: 0:00:03 290.45 kB/s
28 Extracting packages ...
29 [      COMPLETE      ##### 100%
30 Linking packages ...
31 [      COMPLETE      ##### 100%
32 Mustafas-MacBook-Pro:~ msdogan$

```

2 Introduction and background

Developed in early 2000s, **CALifornia Value INtegrated model (CALVIN)** combines ideas from economics and engineering optimization with advances in software and data to suggest more integrated management of water supplies regionally and throughout California. CALVIN is an hydro-economic optimization model for California's advanced water infrastructure that integrates the operation of water facilities, resources, and demands, and it aims to optimize surface and groundwater deliveries to agricultural and urban water users. It allocates water to maximize statewide agricultural and urban economic value, considering physical and policy constraints. It replicates water market operations transferring water from users with lower willingness-to-pay (WTP) to users with higher WTP. CALVIN uses historical hydrology and 2050 water demand projections for its operations. Figure 6 regions and coverage of the model.

CALVIN forces quantitative understanding of integrated water and economic system. Motivation for the CALVIN effort include:

- making better sense of integrated system and operations
- seeking ways to improve system management
- quantifying user willingness to pay for additional water
- finding insights into changes in physical capacities and policies

With the recent updates, CALVIN is evolved to open-source PyVIN model, optimizing water resources in a short amount of time with state-of-the-art solvers and modeling platform. PyVIN has the same input data and objective as CALVIN but it is modeled on Pyomo, a Python-based high level algebraic modeling language. PyVIN also uses HOBES database, which allows better documentation, collaboration and communication between models as well as modelers.

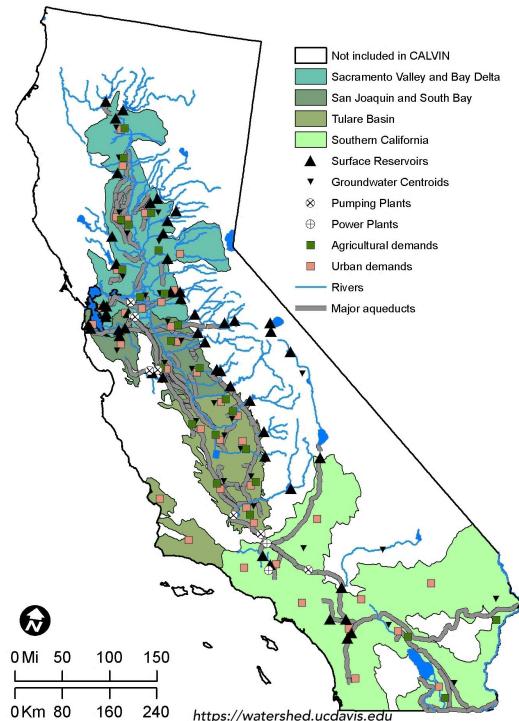
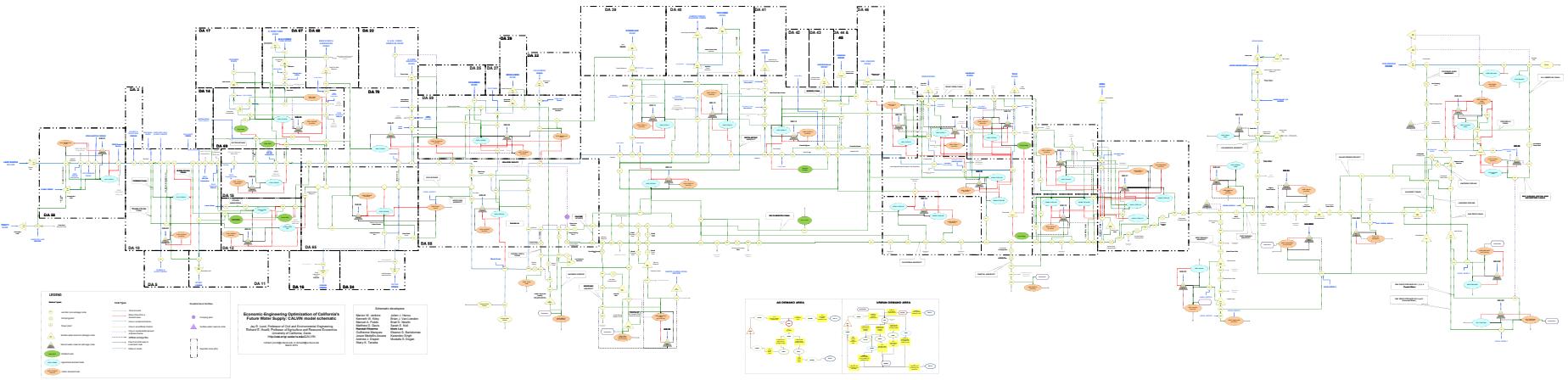


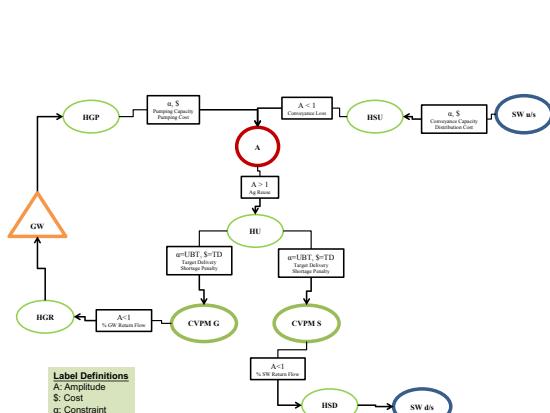
Figure 6: California's water infrastructure and CALVIN's coverage

CALVIN Network Schematic (zoomable)

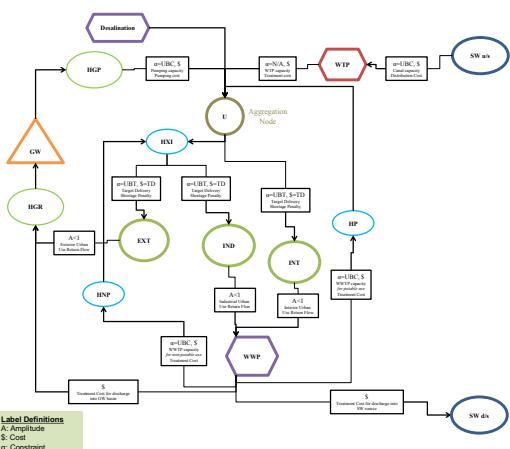


8

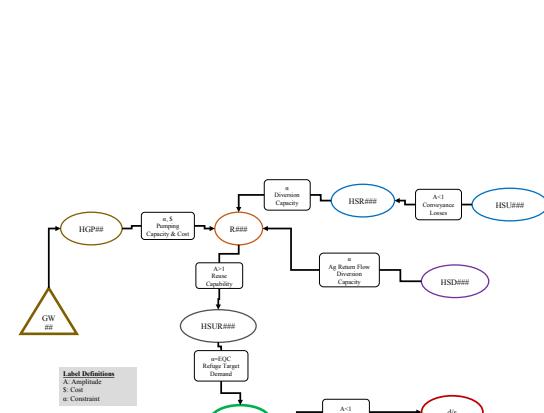
Agricultural Demand



Urban Demand



Wildlife Refuge



updated on April 13, 2017

3 HOBGES database

The HOBGES Project is a bottom up approach to improve and organize the data for water modeling efforts in California. This effort is trying to provide a venue for modelers in California and elsewhere to create an open, organized and documented quantitative representation of the state's intertiered water resources system. Geocoded elements in this database can be interactively converted into tiered networks able to be solved by multiple modeling platforms depending on user preferences, with the appropriate translators. Many HOBGES tools will be web-based with exporting capabilities to the most common analytical and modeling software. HOBGES serves as a cross-platform for data storage, display and documentation. It is a framework for database that aims to better organize data and makes model integration and communication easier by using common format and metadata. Classical approach in modeling is that first model is built and then required data are collected. But HOBGES reverses this order; it serves as a data hub and models are built on top of this database. HOBGES uses GitHub to keep track of changes and documentation. It also has a animation tool to display data.

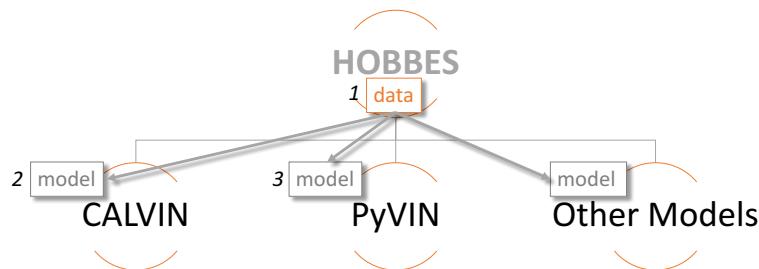


Figure 7: HOBGES database and model integration

3.1 Data flow overview

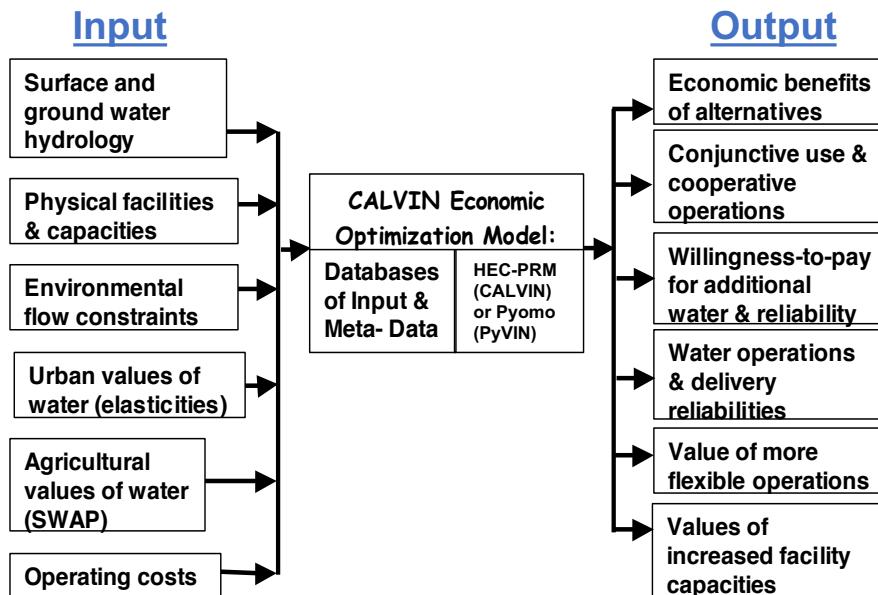


Figure 8: CALVIN (and PyVIN) data overflow with input (left column) and output (right column)

4 Updated version: PyVIN

PyVIN combines CALVIN's knowledge and extensive water infrastructure and hydrology data with a high level algebraic modeling language, Pyomo, and state-of-the-art solvers, such as CPLEX, Gurobi, CBC, and GLPK. Pyomo is an open-source Python based large-scale modeling environment. Its model representation is similar to GAMS and AMPL and it solves the problem with user defined solvers. So, PyVIN is not solver specific and users can choose and install any solver as long as it is compatible with Pyomo.

CALVIN PyVIN	
Large-scale hydroeconomic model	
Optimize water allocation to agricultural and urban users	
Minimize statewide water scarcity and operating costs	
HEC-PRM and VBA based	Pyomo and Python based
Less flexible (limited to HEC-PRM)	More flexible (full LP)
Solver runtime: ~16 hr (depending on initial solution)	Solver runtime: ~1 min (depending on solver)
Requires 32 bit Windows PC	Any computer
HEC-DSS database	Open source: data and source code



Figure 9: CALVIN (and PyVIN) data overflow with input (left column) and output (right column)

4.1 Model architecture

Modeled in different environments, both CALVIN and PyVIN are linear programming network-flow models and solve the same objective function, subject to physical and environmental constraints. The objective is to minimize statewide water scarcity and operating costs (Equation 1). There are three constraints: upper bound (Equation 2), lower bound (Equation 3), and mass balance (Equation 4).

$$\max(z) = \sum_i \sum_j \sum_k c_{ijk} X_{ijk} \quad (1)$$

$$X_{ijk} \leq u_{ijk} \quad (2)$$

$$X_{ijk} \geq l_{ijk} \quad (3)$$

$$\sum_j \sum_i X_{ijk} = \sum_i \sum_j a_{ijk} X_{ijk} + b_j \quad (4)$$

where c represents unit cost, X is flow, u is upper bound, l is lower bound, a is amplitude to represent losses, and b is a local inflow to node j .

4.2 Debug mode

Debug mode adds two more links, `debugsource` and `debugsink`, to each node in the network to prevent infeasibilities. While `debugsource` injects water, `debugsink` removes water from the system if needed at very high cost, \$ 2000000 per acre-foot, which is higher than any other cost in the system. Since the objective is to minimize statewide costs, the model does not use these debug links unless it is really needed, such as mass balance violations. Think of a case where there is a minimum in-stream flow requirement downstream, and your inflow is less than the requirement. The model will try to meet the environmental constraint because it is hard coded and the model cannot change it, so the operation will terminate saying the result is infeasible, and modeler will not know where the problem is because there is hundreds of those requirements. So, debug mode will inject water in that case helping the model find a feasible solution. After that the modeler will look at flows at debug links, and if any of these flows are greater than zero, it means there is mass balance problem. The modeler knows the location and magnitude of the problem and will find a solution.

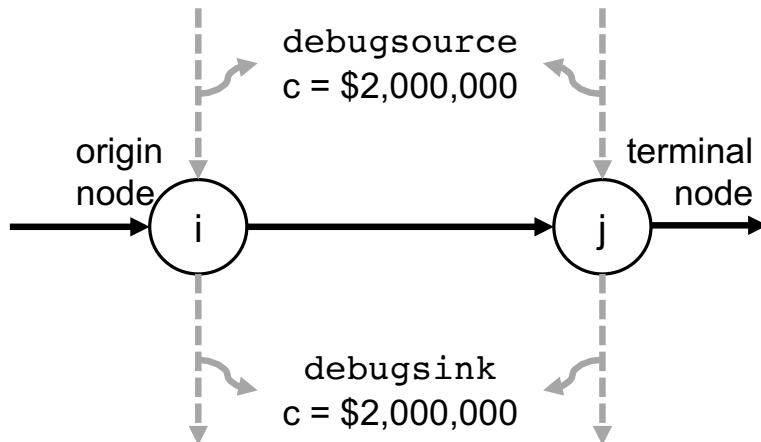


Figure 10: Debug mode links to the system network

4.3 PyVIN model

PyVIN is modeled as an abstract model in Pyomo, separating model structure from data. PyVIN can solve any data size and data do not have to be water resources. As long as it is a network flow problem, such as transportation or transmission, PyVIN can solve it. Also, since it is open-source and has full linear programming features, it has more flexible model representation. Any other constraints can easily be added in addition to three constraint mentioned before. The Pyomo structure code with parameters and equations is shown below. Command to run Pyomo:

```
1 pyomo solve --solver=glpk --solver-suffix=dual pyvin.py data.dat --stream-
   solver --report-timing --json
```

structure file (pyvin.py)

```
1 from __future__ import division
2 from pyomo.environ import *
3 import itertools
4
5 model = AbstractModel()
6
7 # Nodes in the network
8 model.N = Set()
9
```

```

10 # Network arcs
11 model.k = Set()
12
13 model.A = Set(within=model.N*model.N*model.k)
14
15 # Source node
16 model.source = Param(within=model.N)
17 # Sink node
18 model.sink = Param(within=model.N)
19
20 # Flow capacity limits
21 model.u = Param(model.A)
22 # Flow lower bound
23 model.l = Param(model.A)
24 # Link amplitude (gain/loss)
25 model.a = Param(model.A)
26 # Link cost
27 model.c = Param(model.A)
28
29 # The flow over each arc
30 model.X = Var(model.A, within=Reals)
31
32 # Minimize total cost
33 def total_rule(model):
34     return sum(model.c[i,j,k]*model.X[i,j,k] for (i,j,k) in model.A)
35 model.total = Objective(rule=total_rule, sense=minimize)
36
37 # Enforce an upper bound limit on the flow across each arc
38 def limit_rule_upper(model, i, j, k):
39     return model.X[i,j,k] <= model.u[i,j,k]
40 model.limit_upper = Constraint(model.A, rule=limit_rule_upper)
41
42 # Enforce a lower bound limit on the flow across each arc
43 def limit_rule_lower(model, i, j, k):
44     return model.X[i,j,k] >= model.l[i,j,k]
45 model.limit_lower = Constraint(model.A, rule=limit_rule_lower)
46
47 # Enforce flow through each node (mass balance)
48 def flow_rule(model, node):
49     if node in [value(model.source), value(model.sink)]:
50         return Constraint.Skip
51     outflow = sum(model.X[i,j,k]/model.a[i,j,k] for i,j,k in model.A)
52     inflow = sum(model.X[i,j,k] for i,j,k in model.A)
53     return inflow == outflow
54 model.flow = Constraint(model.N, rule=flow_rule)

data file (data.dat)

```

The data file **data.dat** includes list of nodes, and list of links with properties. All links have cost c , amplitude a , lower bound l , and upper bound u . Below is an example data file.

```

1 set N :=
2 INITIAL SR.SHA.1983-10-31 SR.SHA.1983-11-30 FINAL
3 INFLOW.1983-10-31 INFLOW.1983-11-30 ...;
4
5 set k := 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15;
6
7 param source := SOURCE;
8 param sink := SINK;
9
10 param: A: c a l u :=
11 SOURCE INITIAL 0 0 1 0 10000000
12 INITIAL SR.SHA.1983-10-31 0 0 1 3686.84 3686.84
13 SR.SHA.1983-10-31 SR.SHA.1983-11-30 0 -7.003 0.997 630.4 630.4
14 SR.SHA.1983-10-31 SR.SHA.1983-11-30 1 -2.974 0.997 737.4 737.4

```

```

15 SR.SHA.1983-10-31 SR.SHA.1983-11-30 2 -1.466 0.997 632.2 2032.2
16 SR.SHA.1983-11-30 SR.SHA.1983-12-31 0 -6.972 0.999 609.4 609.4
17 SR.SHA.1983-11-30 SR.SHA.1983-12-31 1 -3.056 0.999 700.8 700.8
18 SR.SHA.1983-11-30 SR.SHA.1983-12-31 2 -1.479 0.999 689.7 1941.7
19 FINAL SINK 0 0 1 0 10000000
20 SR.SHA.1984-09-30 FINAL 0 0 1 2923.297 2923.297
21 SOURCE INFLOW.1983-10-31 0 0 1 0 10000000
22 INFLOW.1983-10-31 SR.SHA.1983-10-31 0 0 1 301.765 301.765
23 SOURCE INFLOW.1983-11-30 0 0 1 0 10000000

```

5 Your first model run

When you type the run command, the command line output (for the one-year run) should look like, showing run status and objective value, which can also be found in `results.json`:

```

1 campus-109-001:Run_Folder msdogan$ pyomo solve --solver=glpk --solver-suffix=
  dual pyvin.py data.dat --json
2 [ 0.00] Setting up Pyomo environment
3 [ 0.00] Applying Pyomo preprocessing actions
4 [ 0.00] Creating model
5 [ 17.23] Applying solver
6 [ 48.14] Processing results
7 Number of solutions: 1
8 Solution Information
9   Gap: 0.0
10  Status: feasible
11  Function Value: -495639639.765
12 Solver results file: results.json
13 [ 52.84] Applying Pyomo postprocessing actions
14 [ 52.84] Pyomo Finished

```

5.1 Postprocessing results

Pyomo generates `results.json` file and puts all output in one single file. However, often we need outputs in time-series format and separate files for flow, storage, evaporation, and dual values. PyVIN's Python-based postprocessor scripts create those separate files and save them as `*.csv` which can be easily shared and used to create figures.

Here is my order to run postprocessing scripts:

- Once you obtain `results.json` file from Pyomo, run `postprocess.py`
- If you have run in "debug mode", then run `debug_flow_finder.py` to see if there is any debug flow.
- Run `aggregate_regions.py` to create regional average results.
- Run `supply_portfolio_hatchedbarplot.py` to plot urban and agricultural water supply portfolios.
- You can always create your own personalized postprocessing scripts based on your needs.

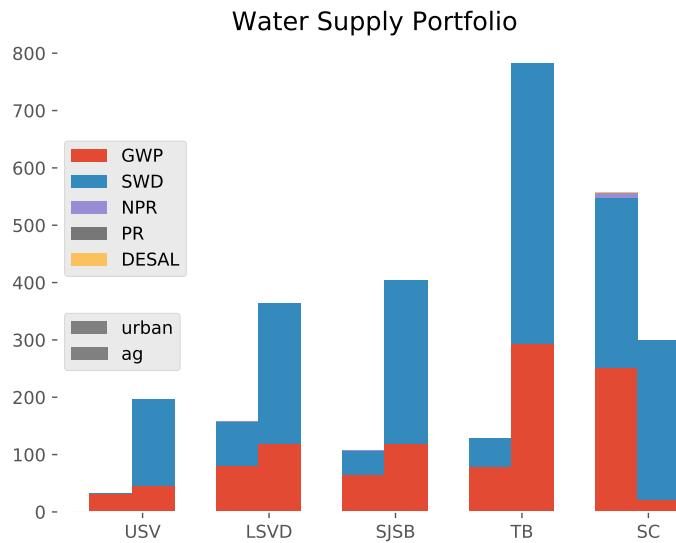


Figure 11: Regional monthly agricultural and urban water supply portfolios

6 References and useful links

CALVIN article that describes the model:

Draper, A. J., Jenkins, M. W., Kirby, K. W., Lund, J. R., & Howitt, R. E. (2003). Economicengineering optimization for California water management. *Journal of water resources planning and management*, 129(3), 155-164.

GitHub repositories: