Environmental Modelling and Software 108 (2018) 8–13

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

An open-source Python implementation of California's hydroeconomic optimization model

Mustafa S. Dogan[a,b], Max A. Fefer[a,b], Jonathan D. Herman[a,b,*], Quinn J. Hart[c], Justin R. Merz[c], Josue Medellín-Azuara[d], Jay R. Lund[a,b]

[a] Department of Civil & Environmental Engineering, University of California, Davis, USA
[b] Center for Watershed Sciences, University of California, Davis, USA
[c] Department of Land, Air, and Water Resources, University of California, Davis, USA
[d] School of Engineering, University of California, Merced, USA

ARTICLE INFO

Keywords:
Hydroeconomic modeling
Network flow optimization
Multi-reservoir operations

ABSTRACT

This short communication describes a new open-source implementation of the CALVIN model (CALifornia Value Integrated Network), a large-scale network flow optimization model of California's water supply system. The model is cross-platform, uses common data formats, and connects to several freely available linear programming solvers. Given inputs including hydrology, urban/agricultural demand curves, and variable operating costs, the model minimizes the systemwide cost of water scarcity and operations including surface and groundwater reservoirs, wastewater reuse, desalination, environmental flow requirements, and hydropower. Key outputs include water shortage costs and marginal economic values of water and infrastructure capacity. We benchmark the scalability of different solvers up to roughly 5 million decision variables, using shared-memory parallelization on a high performance computing cluster. Runtimes are reduced by two orders of magnitude relative to the original model when no initial solution is provided, in addition to the benefits such as accessibility and transparency that come with an open-source platform. While this model is specific to California, the data and model structure are separated, so a similar framework could be used in any system where water allocation has been formulated as a network flow problem.

1. Background

1.1. Hydroeconomic modeling

Hydroeconomic models combine water resources systems engineering and economics, where water allocations are driven by economic value, and conversely, economic costs and benefits are impacted by hydrology (Cai, 2008; Harou et al., 2009; Booker et al., 2012). Optimization models incorporating economic objectives have been used extensively to evaluate water resources planning and management decisions over the past decades (Labadie, 2004). Draper et al. (2003) first introduced CALVIN (CALifornia Value Integrated Network), a hydroeconomic model describing California's water supply infrastructure including surface and groundwater reservoirs, urban and agricultural demands, environmental flow requirements, hydropower production, wastewater reuse and desalination facilities, plus urban and agricultural water conservation. The model has since been used in numerous studies, with topics ranging from climate adaptation to groundwater overdraft. Subsequent authors have updated the structure and parameters of the model to reflect changes in water demands and environmental requirements. The economic optimization framework is unique among statewide California models and provides results such as willingness-to-pay for additional water delivery and marginal value of increased conveyance and storage capacities. Other large-scale water resources studies in California have included contributions to groundwater modeling (Dogrul et al., 2016), agricultural economics (Howitt et al., 2012; Winter et al., 2017), and multi-objective analysis (Quinn et al., 2004; Yang et al., 2015).

The vast improvements in computing power since CALVIN's inception and its ability to explore potential scenarios in California water supply provide an opportunity to move the network structure and data to a new optimization platform. This short communication describes the outcome of this effort, following some general design goals:

* Corresponding author. Department of Civil & Environmental Engineering, University of California, Davis, USA.
E-mail addresses: msdogan@ucdavis.edu (M.S. Dogan), mafefer@ucdavis.edu (M.A. Fefer), jdherman@ucdavis.edu (J.D. Herman), qjhart@ucdavis.edu (Q.J. Hart), jrmerz@ucdavis.edu (J.R. Merz), jmedellin-azuara@ucmerced.edu (J. Medellín-Azuara), jrlund@ucdavis.edu (J.R. Lund).

https://doi.org/10.1016/j.envsoft.2018.07.002
Received 17 February 2018; Received in revised form 12 June 2018; Accepted 5 July 2018
Available online 12 July 2018
1364-8152/

Software availability

Name of Software CALVIN (Python version)
Description An open-source hydroeconomic model that optimizes water allocation to agricultural and urban users in California, originally developed by Draper et al. (2003). Built on top of the Pyomo library (Hart et al., 2012). Storage and demand nodes are connected in a network structure; the dataset is hosted here: (https://github.com/ucd-cws/calvin-network-data)
Developers M. Dogan, M. Fefer, J. Herman, Q. Hart, J. Merz
Funding Source Development was partially supported by the California Energy Commission 4th Climate Change Assessment, Award EM4CD3-04A, and by NSF CyberSEES, Award 1539593. High-performance computing resources provided by the UC Davis College of Engineering
Source Language Python 2.7, 3.4
Supported Systems Unix, Linux, Windows, Mac
License MIT License
Availability https://github.com/ucd-cws/calvin

- *Cross-platform*: Model should run on Windows/OSX/Linux
- *Open data formats*: Input and output data should use only non-proprietary formats such as CSV and JSON
- *Freely available*: Programming language and solvers should be free and open-source. Several solvers described in this work are cost-free only for academic use, but they are not strictly required.
- *Separation of model and data*: The HOBBES project (https://hobbes.ucdavis.edu) stores the network dataset independent from any particular model, allowing this work to use multiple state-of-the-art solvers or models.

In the process of meeting these goals, a few specific features are added, as described in Section 2.

1.2. Network flow optimization

CALVIN is a network flow optimization model seeking to minimize statewide operating and water scarcity cost, subject to physical and regulatory constraints. Network flow problems commonly appear in energy and transportation problems in addition to water; references such as Bazaraa et al. (1977) provide detailed explanation and solution methods for network flow programming. The physical system is represented by a set of nodes $\mathscr{N}$ and links $\mathscr{A}$. Links are defined by $(i, j, k) \in \mathscr{A}$, where i is the source node, $j, j \neq i$ is the destination node, and k is the index of the piecewise linear component for links. Each link has the following properties: flow X_{ijk} (the decision variable), unit cost c_{ijk}, lower bound l_{ijk}, upper bound u_{ijk}, and amplitude (loss factor) a_{ijk}.

The objective function and constraints are:

$$\min_X z = \sum_i \sum_j \sum_k c_{ijk} X_{ijk} \tag{1}$$

subject to:

$$X_{ijk} \leq u_{ijk}, \forall\ (i, j, k) \in \mathscr{A} \tag{2}$$

$$X_{ijk} \geq l_{ijk}, \forall\ (i, j, k) \in \mathscr{A} \tag{3}$$

$$\sum_i \sum_k X_{jik} - \sum_i \sum_k a_{ijk} X_{ijk} = 0, \forall\ j \in \mathscr{N} \tag{4}$$

The objective function (Equation (1)) is a summation over all links i, j, k representing the total cost of flow conveyed in the network. The constraints (Equations (2)–(4)) enforce the upper and lower bounds on each link, and the mass balance at each node, respectively. The lower bound constraints include non-negativity: $l_{ijk} \geq 0\ \forall\ (i, j, k)$. The network flow formulation is a suitable model for large-scale water supply operations, but many other approaches have been proposed to model multi-reservoir systems (e.g., Labadie, 2004; Matrosov et al., 2011; Li et al., 2015; Giuliani et al., 2016).

The version of the model published by Draper et al. (2003) and used in subsequent studies employs the HEC-PRM solver created by the U.S. Army Corps of Engineers Hydrologic Engineering Center. This solver uses the out-of-kilter method for network flow optimization, and stores data in the binary HEC-DSS format. It is available for the Windows operating system. The CALVIN model requires approximately one week to solve using this software platform without an approximate initial solution.

2. Model description

2.1. California network dataset: HOBBES

To implement the model (Equations (1)–(4)), a set of nodes, links, and their properties must be provided. As part of this work, the database for California's water supply network has been migrated to the HOBBES platform (https://hobbes.ucdavis.edu). Unlike previous large-scale water resources optimization models, the HOBBES web-based framework separates the network data from the model-specific optimization method. This approach is intended to facilitate data standardization and documentation, increase transparency, reduce overhead development costs, and enhance modeling collaborations between academic, industry, government agencies, and non-governmental organizations (Medellín-Azuara et al., 2013). Similar benefits have been observed across a range of water resources web applications (Swain et al., 2015).

HOBBES combines a web visualization of the network (https://hobbes.ucdavis.edu/cwn) with the underlying database and metadata in JSON format (https://github.com/ucd-cws/calvin-network-data). The platform also includes automated scripts to export the database to a CSV format compatible with most solvers. The CALVIN GitHub repository provides links to CSV files that have been exported using this tool. Each row of the exported CSV file represents one link; the first two columns contain strings representing the names of the source and destination nodes i, j. The following columns contain the piecewise index k, cost c, amplitude a, lower bound l, and upper bound u.

The dataset currently covers the period 1922–2003 on a monthly timestep. Water scarcity costs are represented by piecewise linear functions for projected urban and agricultural demands in 2050 (Medellin-Azuara et al., 2007; Dogan et al., In Review). The physical network contains approximately 1000 spatial nodes and 60,000 links, selected types of which are shown in Fig. 1. The network model contains a copy of each node for each timestep, where links between timesteps represent storage at surface and groundwater storage nodes. Therefore, the size of the optimization problem increases linearly with the length of the model period. A 1-year model run contains approximately 60,000 decision variables (links), while the full 82-year run contains over 5 million decision variables, including piecewise linear water shortage and operating costs. These values also include optional "debug links" to address potential infeasibilities in the optimization, which are described in Section 2.3 and account for roughly 40% of the links in the model.

2.2. Dependencies and installation

The model requires the Python programming language and its standard scientific libraries (NumPy, SciPy, pandas); we recommend the Anaconda Distribution (https://www.continuum.io/downloads) to easily download and maintain these standard packages. Beyond this, the Pyomo library (Hart et al., 2012) is also required, and is not included with the standard packages in the Anaconda distribution. Pyomo