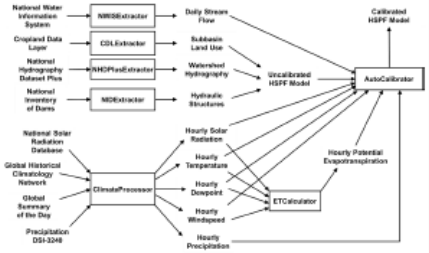
A Python Environment for Automated Integration of Hydrologic Data into Watershed Models

Mitchell Sawtelle & David Lampert
School of Civil & Environmental Engineering– Oklahoma State University

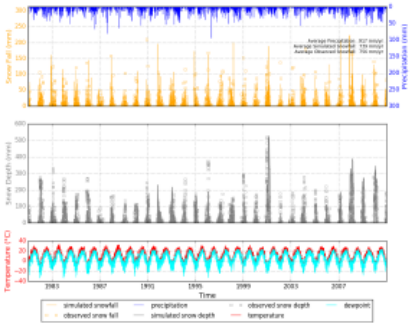
Overview

Watershed models such as the Hydrological Simulation Program in Fortran (HSPF) are used for water resources planning and management. The construction of HSPF models is a highly data and time intensive procedure. Python, which is a high-level, open source programming language, provides an environment for automating the acquisition of the necessary data as well as construction the input files. Recently developed software (PyHSPF) can be used to build HSPF input files, run simulations, and postprocess simulation results using the Python Programming Language. The combination of tools in Python, public datasets on the Internet and the PyHSPF extensions enable rapid development of long-term, reproducible, and sophisticated models for new hydrologic insight. The figure below illustrates the online data sources and automatic model construction procedure.



PyHSPF Information Flow Diagram.

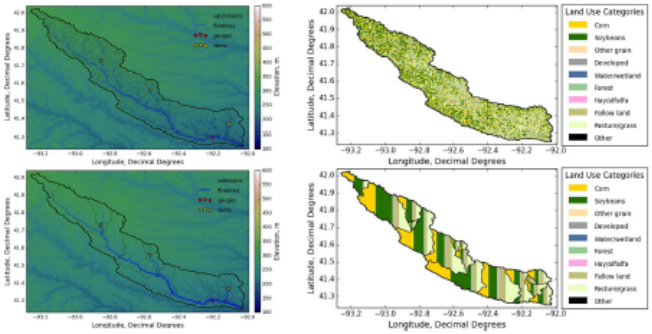
The climate data are extracted from the National Solar Radiation Database, the Global Historical Climatology Network, the Global Summary of the Day Database, and the Hourly Precipitation Database (3240). The data are then aggregated at each time step to generate time series of snowfall, precipitation, solar radiation, dew point, temperature, and wind speed as shown below.



Auto-Generated Climate Time Series.

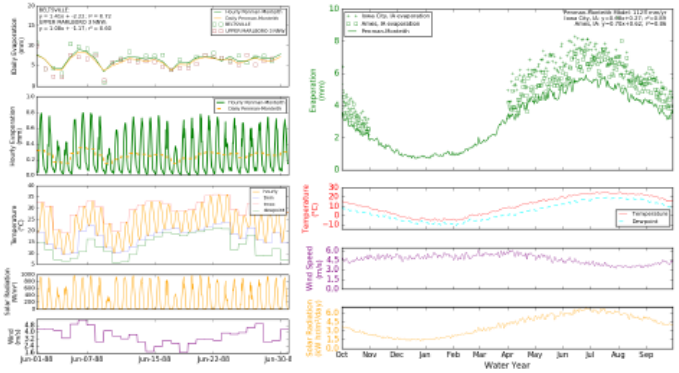
Data Processing

The National Hydrography Dataset Plus Version 2 provides high-resolution data on stream reaches and catchments as shown below. The raw data can be aggregated to develop a model that incorporates hydraulic structures that can be determined from the National Water Information Service and the National Inventory of Dams. The Cropland Data Layer contains 30-m resolution land use data for the entire United States since 2008. Python tools have been developed to automate the extraction of data from these sources.



National Hydrography Dataset Plus Version 2 Raw Data (Upper Left), Processed Data (Lower Left), Cropland Data Layer Raw Data (Upper Right), and Processed Data (Lower Right).

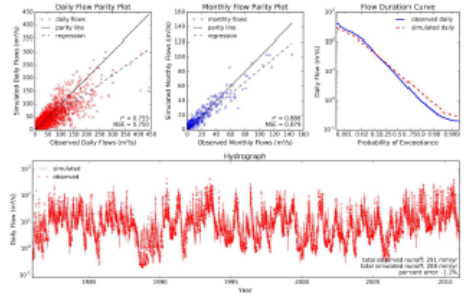
The raw climate time series are used to calculate hourly potential evapotranspiration time series using the United Nations Food and Agriculture Organization's Penman-Monteith Method. The procedure is entirely automated. The figures below illustrate the calculation across a short time period and show a comparison with observed pan evaporation on an annual plot.



Automatically Generated Reference Evapotranspiration Time Series for a One-Week Period (Left) and Across Multiple Years (Right).

Calibration

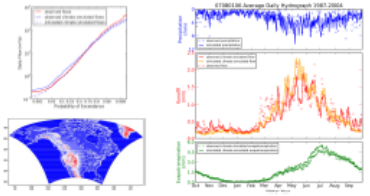
The processed datasets are used to construct the land segments and stream reaches that are the foundation of the HSPF model. The climate time series are used to provide the external forcing for the model simulations. The hydrological process parameters are then fitted through a calibration procedure.



Calibration Results Including Daily and Monthly Parity Plots (Upper Left and Center), Flow Duration Curves (Upper Right), and Hydrograph (Lower).

Climate Change Application

Following calibration, the model can be used to answer research questions including changes in climate and land use.



Future Research Ideas

Future applications made possible with these tools include:

- Larger-scale models
- Data mining (correlations between hydrological processes and human activities)
- Time-variable land use
- Pesticides
- Nutrients

Software Availability

The software is freely available at:

https://github.com/djlampert/PyHSPF
https://github.com/Msawtelle/PyNHDPLUS