Project 3. *Supply enough clean drinking water.*

System Modeling for Climate Response
Department of Civil & Environmental Engineering
Duke University
Henri P. Gavin

due: December 5, 2025

1 Stochastic Optimization

The performance of systems designed to operate within natural environments depends (to a greater or lesser extent) upon the variability of their environments. A system is said to have *robust performance* with respect to *variability in environmental conditions* if its performance is not critically sensitive to changes in its environmental conditions. As you have seen, systems that are finely tuned to a particular (assumed) set of conditions can fail *almost surely* ("a.s.") under slightly different conditions.

The design and optimization of systems for robust performance with respect to uncertain environments must therefore inherently involve a model for the variability in the system's operating environment. Unlike *deterministic optimization*, in which all attributes within any analysis are presumed to be known precisely and in which a unique value of the objective can be computed precisely for a particular operating environment, *stochastic optimization* can be applied to systems in which the system objective and its constraints depend on random aspects of the problem, and can therefore be estimated statistically. Methods that are highly efficient for deterministic optimization problems (i.e., methods that rely on accurate gradient and Hessian computations) tend to not perform as efficiently on stochastic problems. On the other hand, stochastic optimization methods will usually find a good solution to stochastic problems, but will be less efficient in the solution of deterministic optimization problems.

Stochastic optimization methods generally require only approximate measures of the gradient of the objective with respect to the design variables. Stochastic optimization methods can be *tuned* to converge quickly for a particular problem, and the convergence rate can be very sensitive to the tuning. In most stochastic optimization problems, the set of design variables at iteration i, $\mathbf{x}^{(i)}$ is updated through the addition of a set of random variables $\mathbf{d}^{(i)}$,

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + \mathbf{d}^{(i)}$$

in which $\mathbf{d}^{(i)}$ has a particular probability density function, $f_D(\mathbf{d}^{(i)})$. In most algorithms the pdf of D changes from one iteration to the next, usually by reducing its variance. The updated set of design variables is analyzed and if it provides better performance, then it is assigned to the set of optimal design variables. Various stochastic optimization methods differ in the details of how the sequence of perturbation vectors $\mathbf{d}^{(i)}$ evolves.

One class of *stochastic optimization* methods requires the simultaneous and interrelated evolution of several initial guesses to a particular solution. Such methods are highly computationally intensive since they involve the parallel analyses of multiple potential solutions. Such evolutionary methods go by names such as *genetic algorithms*, *particle swarm*, and *ant colony* optimization.

Constraints in stochastic optimization problems are usually incorporated via a penalty method.

The *optimized step size random search* (OSSRS) method of B.V. Sheela (*CMAME* 19(1): 99-106 (1979)), and the *Nelder-Mead* method (*CJ* 7(4): 308-313 (1965)) can be applied to problems in stochastic optimization, since neither method requires precise computation of the gradients of the objective and the constraints with respect to the design variables. In applying these methods to stochastic optimization problems, it can be helpful to characterize both the expected value of the performance *and* its variability. To that end, the performance of a trial design in iteration (i) (quantified by values of the design variables $\mathbf{x}^{(i)}$), will be evaluated in terms of its sample statistics (the average and sample standard deviation). Specifically, for each trial design ($\mathbf{x}^{(i)}$), the scalar objective function $f(\mathbf{x}^{(i)})$ and the m inequality constraints $\mathbf{g}(\mathbf{x}^{(i)})$ are analyzed N times, each time using a different simulation for the random aspects of the problem. For each of the N analyses, a the penalized cost function is computed

$$f_{\mathsf{A}}(\mathbf{x}^{(i)}) = f(\mathbf{x}^{(i)}) + P \left[\sum_{j=1}^{m} \left\langle g_j(\mathbf{x}^{(i)}) \right\rangle \right]^q ,$$

where P is a positive penalty factor and q is a positive penalty exponent. The sample average $\mathsf{avg}(f_{\mathsf{A}}(\mathbf{x}^{(i)}))$ and the sample coefficient of variation $\mathsf{cov}(f_{\mathsf{A}}(\mathbf{x}^{(i)}))$ will be computed from the sample of N values of f_{A}. The quantity to be minimized for the stochastic optimization in this project is

$$\min_{\mathbf{x}} f_{\mathsf{S}} = \mathsf{avg}(f_{\mathsf{A}}(\mathbf{x}))\,(1 + \mathsf{cov}(f_{\mathsf{A}}(\mathbf{x})))$$

This objective is the 84$^{\text{th}}$ percentile of the distribution of costs. There are three benefits of quantifying the stochastic performance like this.

- This performance metric inherently recognizes that the performance can be known only in terms of its statistics.
- This performance metric increases with the mean (expected) performance *and* its variability. Systems with volatile performance have a higher cost.
- If the system is to be optimized for conditions corresponding to a specific hazard level, only operating environments exceeding a specific extreme should be considered. In such cases the 84$^{\text{th}}$ quantile of the *sampling distribution of the mean* response corresponding to the extreme hazard is a good performance criteria; so the cov of the performance should be divided by $\sqrt{N}$. This performance metric decreases with the number of analyses (N). By analyzing the behavior of the trial designs more thoroughly, the confidence in the estimate of the mean increases, effectively reducing the impact of variability.

  In this project the full range of hazard levels will be considered, and so the perfance metric is the estimated mean plus standard deviation of the cost function.

2 Problem Statement Abstract

In this project, you will design the capacities and controls for a drinking water supply system. The water supply system is to be designed to serve a growing community for fifty years. The performance of the system depends on its design (prescribed by you) and the naturally random operating environment (specifically, precipitation, temperature, and population). In this project the system will be modeled as a *dynamic system*, in which the state of the system changes from day to day as a function of the flows into and out-of three subsystems: an aquifer, a reservoir, and a water treatment plant. The *state* of the system on any given day is quantified in terms of the volumes of water and the concentrations of contaminants (biological, sediments, petro-chemical) in the three sub-systems. Many of these flows (precipitation, evaporation, water consumption) are random, or depend on random quantities (temperature, population, human behavior). The random quantities in this problem can be simulated from a set of probability distributions that capture the statistical nature of these random quantities. Part of the problem involves optimizing the capacities of the system, and part of the problem involves determining how measurements of the state of the system can be used to control how much water is processed on any given day, and how aggressively the water needs to be decontaminated. Three treatment types (chlorine, filters, and activated carbon) have different degrees of effectiveness on the three pollutants. The performance of the system will be monetized in terms of a capital cost (that depends on the optimized capacities) and an operating cost (that depends on the water plant control method, and the randomness of the operating environment).

3 Model for a Drinking Water Supply and Treatment System

The idealized drinking water supply system modeled in this study consists of a watershed and an aquifer which supply water to a reservoir which, in turn, feeds water to a water treatment plant. The water treatment plant consists of a tank of untreated water which feeds water through three treatment processes to a treated water tank, where the water is held until the community draws from it. This system is shown schematically in Figure 1.

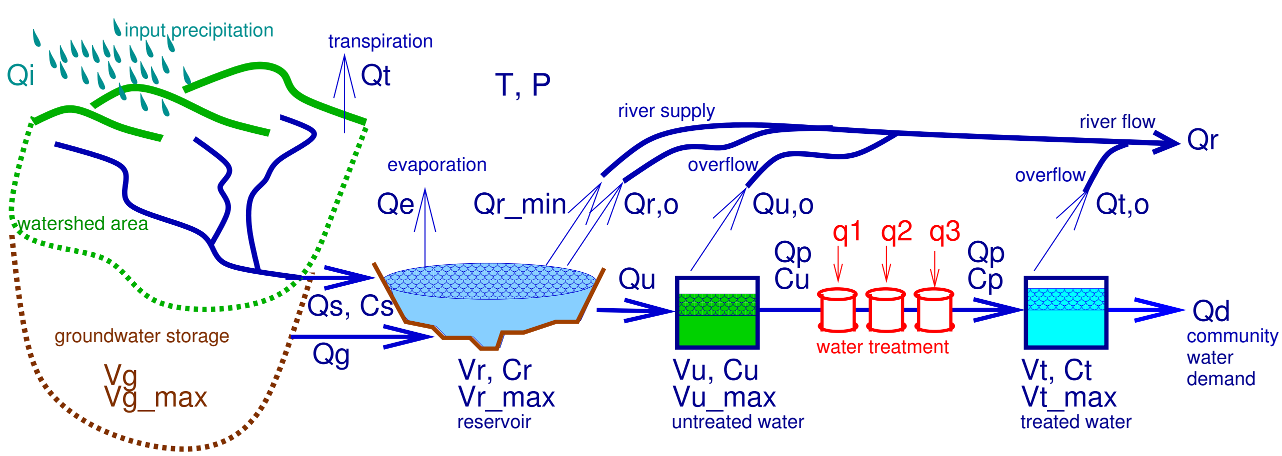

Figure 1. Schematic of a water treatment system.

This idealized water supply and treatment system is modeled using mass balance equations, with the assumption that the pollutants are always evenly mixed within each volume, and that the flow is incompressible. If at a point in time t, a tank contains a volume $V(t)$, with pollutant concentration $C(t)$, receives an average inflow of $Q_{\mathsf{in}}(t)$, with concentrations $C_{\mathsf{in}}(t)$, and delivers an average outflow of $Q_{\mathsf{out}}(t)$, (Mgal/day), then the rate of change of the volume in the tank is

$$\frac{d}{dt}V(t) = Q_{\mathsf{in}}(t) - Q_{\mathsf{out}}(t) \ , \tag{1}$$

with contaminant mass

$$\frac{d}{dt}m(t) = C_{\mathsf{in}}(t)Q_{\mathsf{in}}(t) - C(t)Q_{\mathsf{out}}(t) \tag{2}$$

where $C(t) = m(t)/V(t)$.

4 Precipitation, Transpiration, Groundwater flow, Stream flow, and Evaporation

4.1 Rainfall Simulation

Precipitation falling into the watershed is transpired through vegetation, is stored as ground water, or is otherwise passed directly to a reservoir through streams.

The total rainfall on a day is the product of the two random variables, W, whether or not there is rain on that day ($W \in (0, 1)$), and R, the amount of rain on that day, in inches. These two random variables will be modeled based on measured observations of local precipitation.

Data from the USGS raingage station at the Falls Lake Dam, (USGS 02087182), are shown in Figure 2. From 1998-01-01 to 2023-04-10 there were 2822 days with rain out of a total of 9232 days, giving 3.27 days between rainfalls (on average). This is called the *return period* T_{r} for a day with rain.

$$T_{\mathsf{r}} = \frac{\text{total number of days}}{\text{total number of days with rain}} = 3.27 \text{ days between days with rain, on average} \ . \tag{3}$$

To model the random occurances of precipitation, it is convenient to assume these occurances to be statistically independent of the occurance of precipitation on previous or subsequent days. With these assumptions, we can model the probability of rain on any given day as

$$\text{Prob[a wet day]} = \text{Prob}[W = 1] = \frac{1}{T_{\mathsf{r}}} \exp\left(-\frac{1}{T_{\mathsf{r}}}\right) = \text{Prob}\left[\mathcal{U} \le \frac{1}{T_{\mathsf{r}}} \exp\left(-\frac{1}{T_{\mathsf{r}}}\right)\right] \ , \tag{4}$$

where $\mathcal{U}$ is a standard uniformly-distributed variable between 0 and 1.

The total cumulative precipitation from 1998-01-01 to 2024-04-06 was 965.4 inches or 0.094 inches per day (on average). But the amount of any given rainfall is random and the

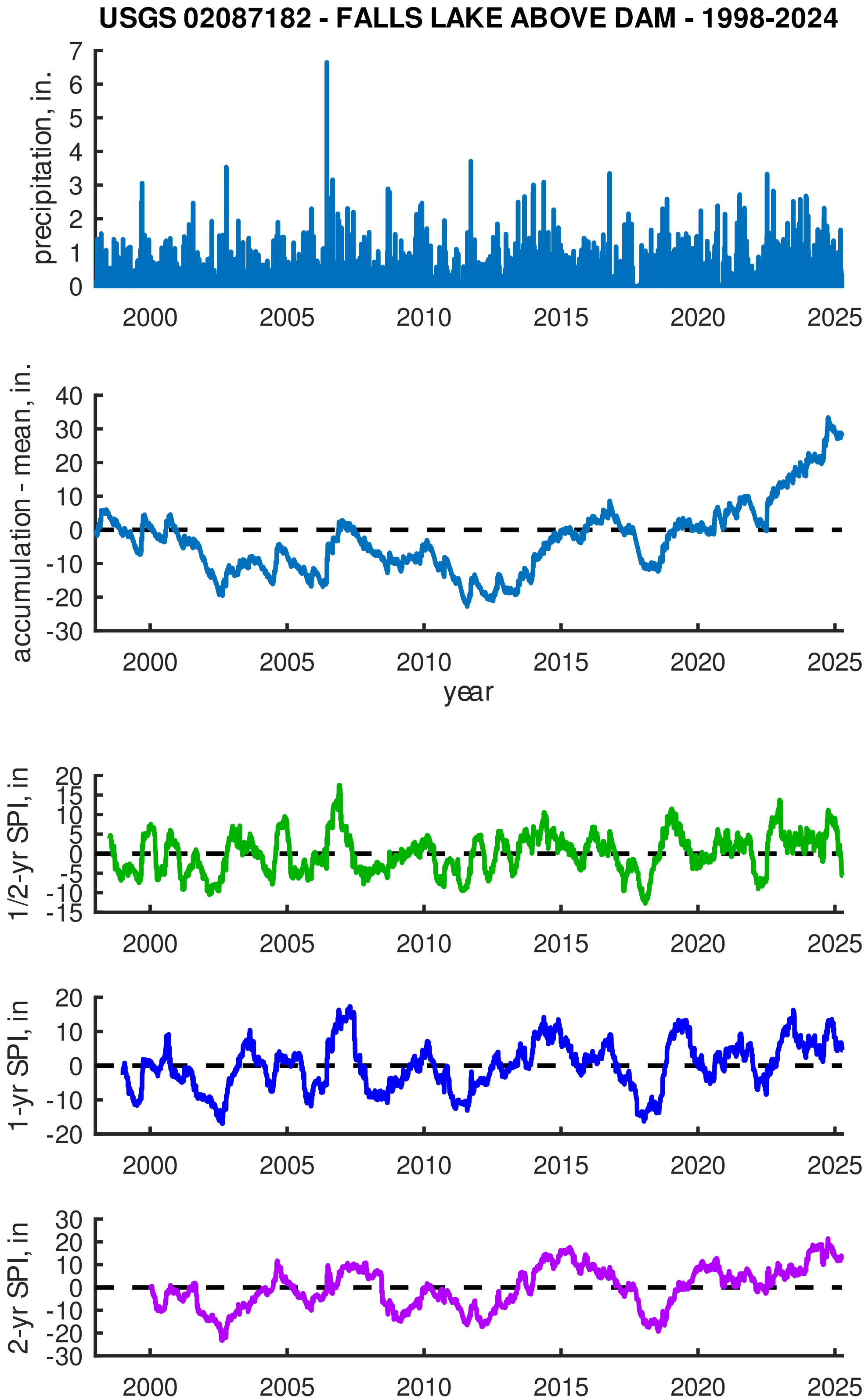

Figure 2. Rainfall data and specific precipitation indices, USGS 02087182, 1998-2025

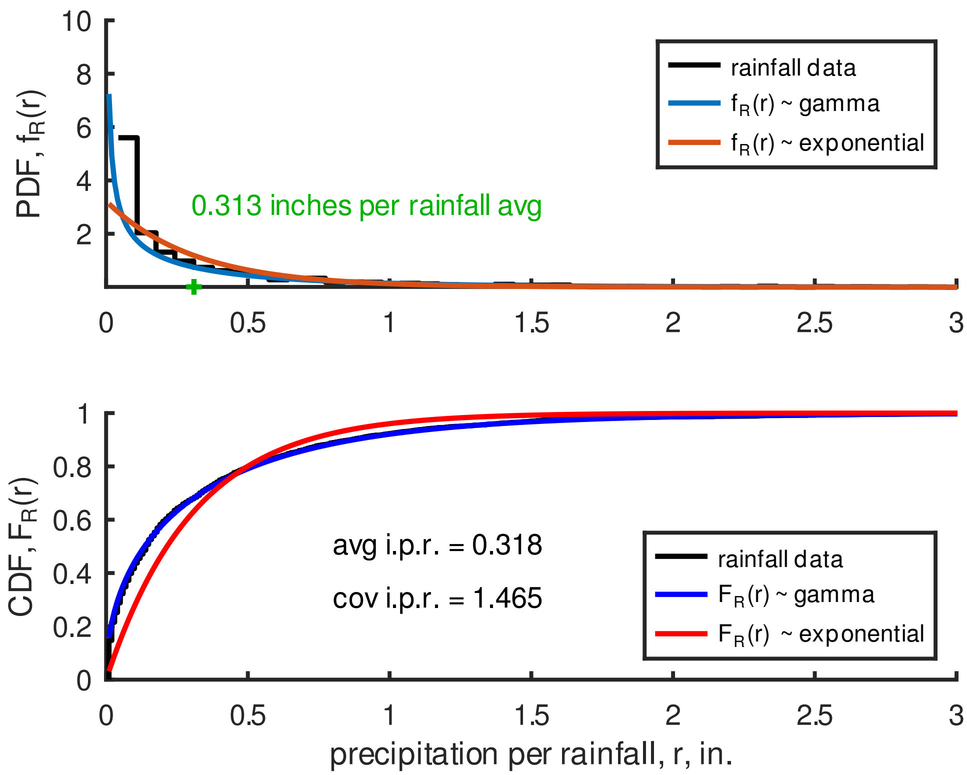

Figure 3. Rainfall probability distributions, USGS 02087182, 1998-2025

distribution of rainfall amounts can be deduced from records of daily rainfall. For the data shown in Figure 2, the average amount of rain per rainfall can be calculated as

$$\bar{r} = (964.4 \text{ cumulative inches of rain})/(3031 \text{ days with rain}) = 0.304 \text{ inches per rainfall} . \tag{5}$$

The coefficient of variation for the amount of rain per rainfall from this data is 1.47. Figure 3 plots the PDF and CDF of this rainfall data along with the exponential distribution[1] and the gamma distribution[2]. Note that the exponential distribution underestimates the likelihood of very small rainfalls and very large rainfalls, but the gamma distribution captures the variability in rainfall amounts very well. The exponential distribution has only one parameter (the mean rainfall rate), whereas the gamma distribution has two design variables (the mean $\bar{r}$ and the coefficient of variation c_r). We will use the gamma distribution to simulate daily rainfall amounts in this project.

The deviation of rainfall from the yearly average can be reported as the cumulative sum of rainfall for the last 365 days minus the average yearly rainfall. This quantity is referred to as the *one-year specific precipitation index (SPI)* and has units of inches of rainfall. The one-half-year, one-year and two-year SPI's for the USGS Falls Lake rainguage are plotted in Figure 2. Note that prior to the "drought of 2007-2008", this region had exceptionally high

[1] http://en.wikipedia.org/wiki/Exponential_distribution
[2] http://en.wikipedia.org/wiki/Gamma_distribution

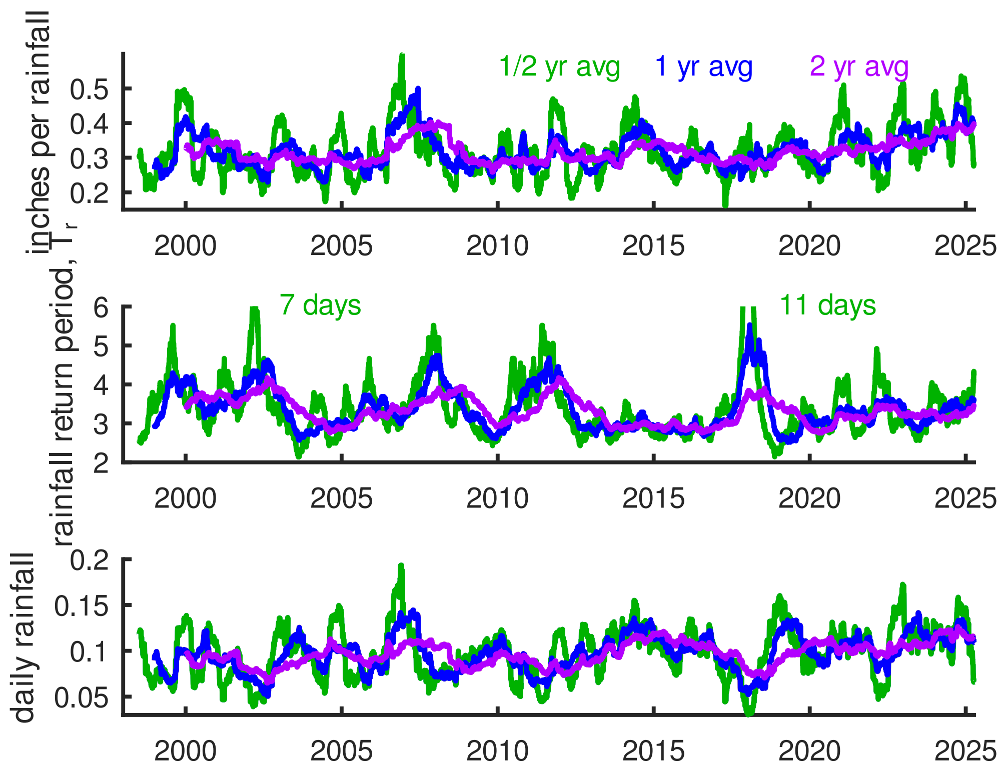

Figure 4. Rainfall trends, USGS 02087182, 1998-2025

rainfall.

Over the next century rainfall return periods T_r and the amount of rain per rainfall $\bar{r}$ are expected to increase so that the total rate $\bar{r}/T_r$ is relatively unchanged, at roughly 0.1 inches per day.[3] The time scale for over which these climactic changes may transpire is quantified in this project as the *climate change time scale*, $CCTS$ (in years). For this project, the rainfall return period T_r over a several-decade period will be assumed to increase linearly with time according to the relationship

$$T_r(t) = 3.28\ (1 + (t/(CCTS \times 365)))\ . \tag{6}$$

In other words, the return period for days with rain doubles in $CCTS$ years. Because the cumulative annual rainfall is projected to remain constant, the quantity of rain per rainfall must increase, so for this project,

$$(\bar{r}(t) \text{ inches per rainfall }) = (0.094 \text{ inches per day})\ (T_r(t) \text{ days between rainfalls}) \tag{7}$$

Trends in rainfall from the USGS Falls Lake rain-gauge are plotted in Figure 4. The 1/2-year, 1-year and 2-year running averages of the rainfall return period T_r, the amount of rain per rainfall, $\bar{r}$, and the average daily rainfall $\bar{r}/T_r$ are plotted. Periods of infrequent rain

[3]citation needed.

(droughts in 2002 and 2008) are clearly visible from this time series. Note that just prior to the 2008 drought the rain-gauge recorded a period of very large rainfalls. For this single rain-gauge, there are no discernible trends over the last twenty-two years. If this rain-gauge is representative of regional rainfall patterns, then the climate change time scale for regional rainfall effects is probably around 100 years, or more.

Simulated data from this statistical model qualitatively matches the observed rainfall, as shown in Figures 5 and 6.

The daily input flow of precipitation from the atmosphere to the ground surface, Q_i, is the daily rainfall WR multiplied by the rainfall area, A_r. The rainfall area is assumed to be a lognormally distributed random variable, with a median value equal to 120 square miles and a coefficient of variation of 0.90. The watershed area A_w is 250 square miles and the rainfall area cannot exceed the watershed area. The rainfall area is assumed to be un-correlated with the rainfall amount. The rainfall flow is given by $Q_\mathsf{i} = A_\mathsf{r}\, WR$, and is converted to units of Mgal/day.

4.2 Temperature Simulation

In this study the temperature has annual variation a seven year variation and a long-term warming trend. The daily high temperature is modeled with

$$T(t) = \bar{T} - T_1 \cos(2\pi(t-15)/365) + T_7 \sin(2\pi t/(7 \cdot 365) + T_\mathsf{c} t/(365 \times CCTS) + \delta T \ , \quad (8)$$

where $\bar{T}$ is the mean temperature, T_1 is the yearly temperature variation, T_7 is a level of a seven-year temperature variation, T_c represents the change in temperature over a time period of $CCTS$, d is the day, and δT is the random day-to-day temperature variation. An example of a simulation of daily temperature variation is shown in Figure 7.

4.3 Population Simulation

The population grows according to

$$P(t) = P_0 + P_1(t/365) + P_2(t/365)^2 + \delta P \ , \quad (9)$$

where P_0 is the initial population and P_1 and P_2 are population growth coefficients. The coefficient P_2 is treated as a normally distributed random variable in this study (with a mean of 54 and coefficient of variation of 0.08), and results in a population growth from 250,000 in 2010 to somewhere between 550,000 and 600,000 (with 99% confidence) fifty years later. The term δP represents a small random day-to-day variation in population.

The population growth model is calibrated to data for Durham County from the U.S. Census Bureau.

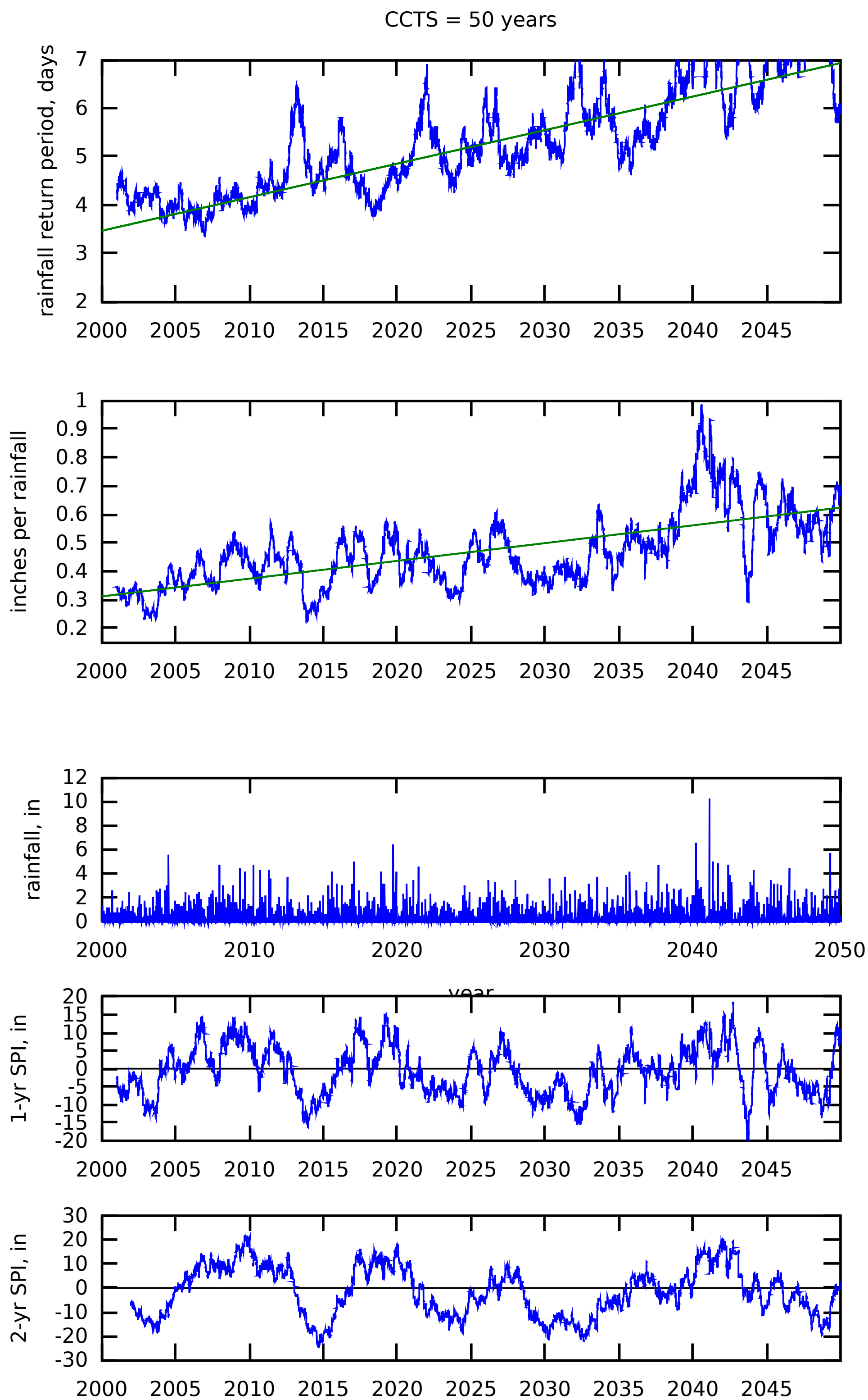

Figure 5. Modeled changes in rainfall frequency and intensity as a non-stationary compound random process.

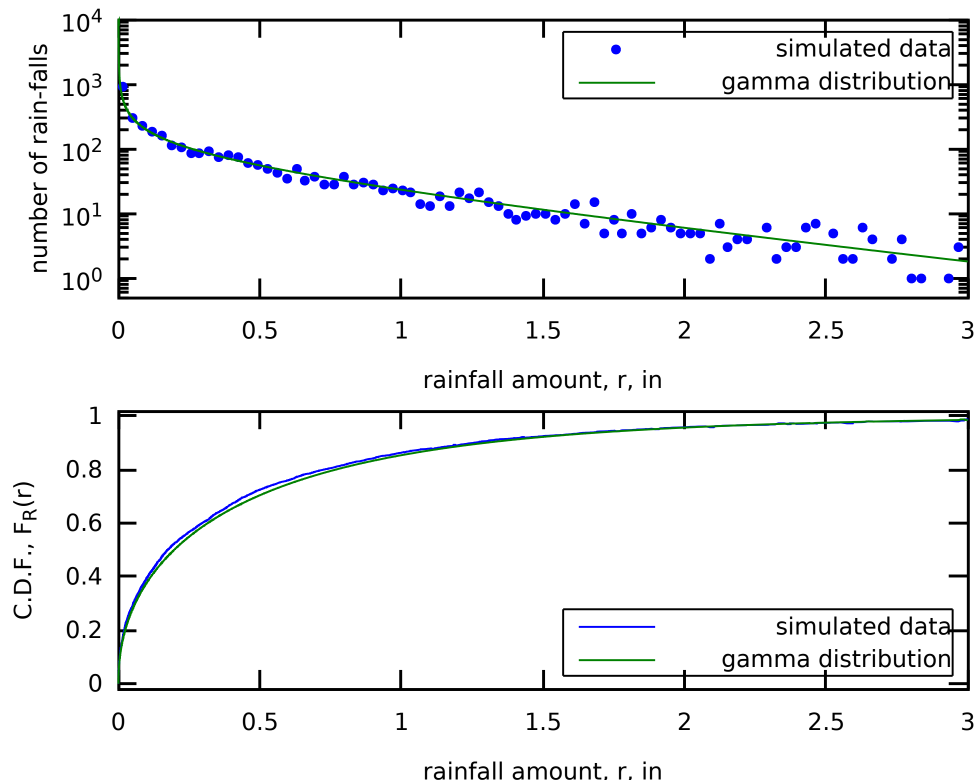

Figure 6. pdf and cdf of precipitation per rainfall as a gamma-distributed random variable.

4.4 Flow modeling

Rainwater flows into streams, is transpired through vegetation, or flows into the ground, as described in the following. The flow into the ground is the daily precipitation less the daily transpired water, the daily stream flow, and the daily groundwater flow,

$$\frac{d}{dt} V_{\mathsf{g}}(t) = Q_{\mathsf{i}}(t) - Q_{\mathsf{t}}(t) - Q_{\mathsf{g}}(t) - Q_{\mathsf{s}}(t) \ , \tag{10}$$

The flow of transpired water, $Q_{\mathsf{t}}(t)$, depends on temperature, $T(t)$, and the ground moisture,

$$Q_{\mathsf{t}}(t) = (\alpha_{\mathsf{t}} + \beta_{\mathsf{t}} T(t)) \ V_{\mathsf{g}}(t)/V_{\mathrm{g,max}} \ , \tag{11}$$

where α_{t} and β_{t} are constant coefficients.

Flow from the ground water to the reservoir, Q_{g}, increases with greater soil moisture.

$$Q_{\mathsf{g}}(t) = \alpha_{\mathsf{g}} V_{\mathsf{g}}(t)/V_{\mathrm{g,max}} \ . \tag{12}$$

The daily flow into streams depends on how wet the ground is. If $V_{\mathrm{g,max}}$ is the capacity of the ground to hold water, and $V_{\mathsf{g}}(t)$ is the current volume of water in the ground, the flow into the streams, Q_{s}, is modeled by

$$Q_{\mathsf{s}}(t) = \alpha_{\mathsf{s}} V_{\mathsf{g}}(t)/V_{\mathrm{g,max}} + \langle V_{\mathsf{g}}(t) - V_{\mathrm{g,max}} \rangle /4 \tag{13}$$

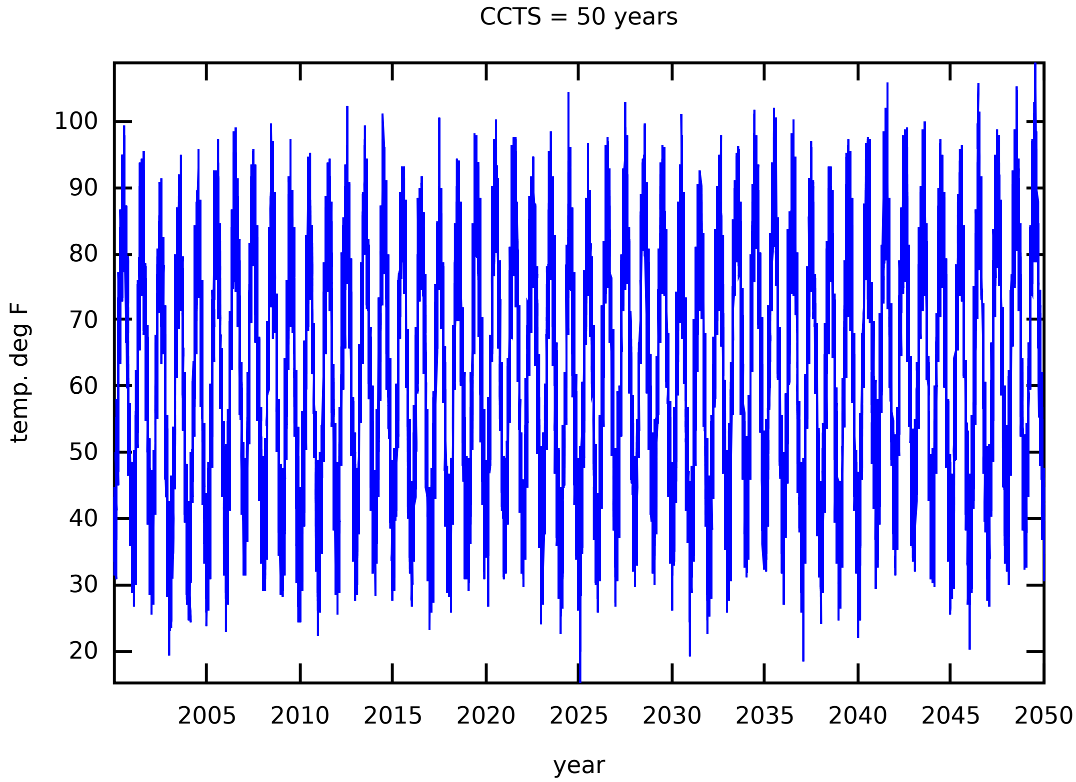

Figure 7. Simulation of daily temperature variations.

where the Macaulay brackets mean that $\langle V_{\mathsf{g}}(t) - V_{\mathrm{g,max}} \rangle = (V_{\mathsf{g}}(t) - V_{\mathrm{g,max}})$ if $(V_{\mathsf{g}}(t) - V_{\mathrm{g,max}}) > 0$ and $\langle V_{\mathsf{g}}(t) - V_{\mathrm{g,max}} \rangle = 0$ otherwise.

The daily stream flow into the reservoir carries three kinds of pollutants, micro-organisms, suspended solids, and petro-chemicals. The stream flow concentrations of these pollutants increase with the regional population, $P(t)$, and with the stream flow $Q_{\mathsf{s}}(t)$.

$$C_{\mathsf{s}}(t) = \bar{C}_{\mathsf{s}} + c_{\mathsf{p}} P(t) + c_{\mathsf{s}} Q_{\mathsf{s}}(t) \ , \tag{14}$$

where $\bar{C}_{\mathsf{s}}$ and c_{p} and c_{s} are constant coefficients.

The reservoir is fed by stream flow, Q_{s}, and ground water, Q_{g} The capacity of the reservoir is $V_{\mathrm{r,max}}$ and the current volume of water in the reservoir is $V_{\mathsf{r}}(t)$, The reservoir supplies water to a river and to the water treatment plant. The rate of change of the volume of water within the reservoir is

$$\frac{d}{dt} V_{\mathsf{r}}(t) = Q_{\mathsf{s}}(t) + Q_{\mathsf{g}}(t) - Q_{\mathsf{e}}(t) - Q_{\mathsf{u}}(t) - Q_{\mathrm{r,min}} - \langle V_{\mathsf{r}}(t) - V_{\mathrm{r,max}} \rangle / 9. \tag{15}$$

where the flow to the river is $Q_{\mathrm{r,min}} + \langle V_{\mathsf{r}} - V_{\mathrm{r,max}} \rangle / 9$, and the term $\langle V_{\mathsf{r}} - V_{\mathrm{r,max}} \rangle / 9$ represents water overflowing the reservoir. The daily evaporation from the reservoir, Q_{e}, depends on the temperature and the volume held within the reservoir,

$$Q_{\mathsf{e}}(t) = (\alpha_{\mathsf{e}} + \beta_{\mathsf{e}} T(t)) \ V_{\mathsf{r}}(t) / V_{\mathrm{r,max}} \ . \tag{16}$$

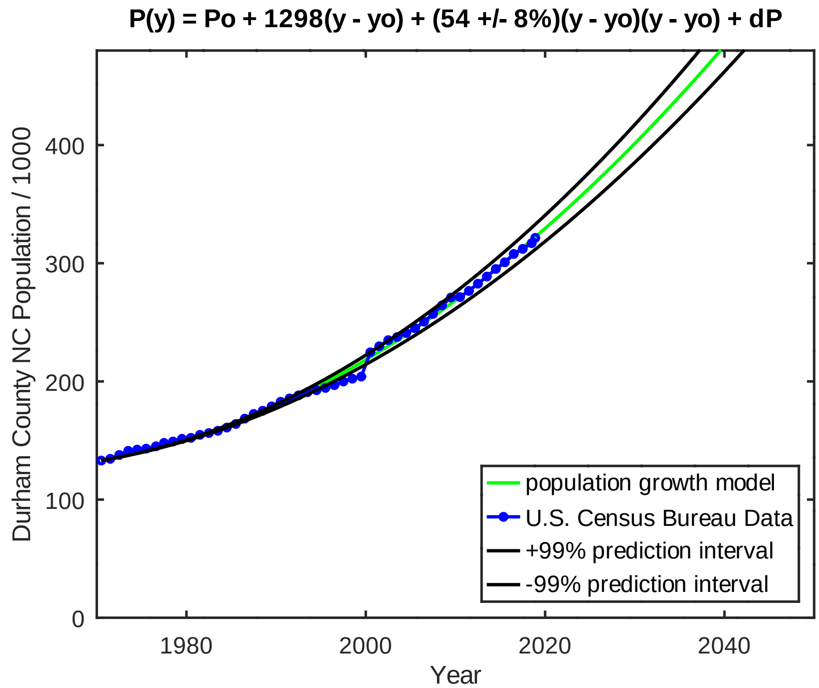

Figure 8. Durham County population from 1970 to 2019, and a model for future population growth.

The flow to the untreated water tank, Q_{u}, is controlled by the plant operator. The plant operator can not take out more water than remains in the reservoir at the end of any given day. A record of measured reservoir levels is plotted in Figure 9. Note that during droughts reservoir levels tend to drop over a period of months and that reservoir stores are replenished quickly once a few heavy rainfalls arrive.

Pollutant concentrations are expressed in terms of volume fractions (volume of contaminant per volume of water). If the pollutant concentrations in the stream flow, $C_{\mathsf{s}}(t)$ and the pollutant concentrations in the reservoir are $C_{\mathsf{r}}(t)$, then the rate of change of the pollutant mass in the reservoir is

$$\frac{d}{dt} m_{\mathsf{r}}(t) = C_{\mathsf{s}}(t) Q_{\mathsf{s}}(t) - C_{\mathsf{r}}(t) Q_{\mathsf{u}}(t) - C_{\mathsf{r}}(t) Q_{\mathrm{r,min}} - C_{\mathsf{r}}(t) \langle V_{\mathsf{r}}(t) - V_{\mathrm{r,max}} \rangle / 9 \ , \tag{17}$$

from which $C_{\mathsf{r}}(t) = m_{\mathsf{r}}(t)/V_{\mathsf{r}}(t)$ is easily found.

5 Water Treatment and Consumption

At the treatment plant, water is held for processing in an untreated water tank of capacity $V_{\mathrm{u,max}}$ and flows through the treatment process at a controlled flow rate Q_{p}. The current volume of water in the untreated water tank is $V_{\mathsf{u}}(t)$, and it has pollutant concentrations $C_{\mathsf{u}}(t)$. The rate of change of the volume of water within the untreated tank is

$$\frac{d}{dt} V_{\mathsf{u}}(t) = Q_{\mathsf{u}}(t) - Q_{\mathsf{p}}(t) - \langle V_{\mathsf{u}}(t) - V_{\mathrm{u,max}} \rangle . \tag{18}$$

where the term $\langle V_{\mathsf{u}} - V_{\mathrm{u,max}} \rangle$ represents water overflowing the untreated tank to the river. The rate of change of pollution mass in the untreated tank is

$$\frac{d}{dt} m_{\mathsf{u}}(t) = C_{\mathsf{r}}(t) Q_{\mathsf{u}}(t) - C_{\mathsf{u}}(t) Q_{\mathsf{p}}(t) - C_{\mathsf{u}}(t) \langle V_{\mathsf{u}}(t) - V_{\mathrm{u,max}} \rangle \ , \tag{19}$$

from which the concentrations $C_{\mathsf{u}}(t) = m_{\mathsf{u}}(t)/V_{\mathsf{u}}(t)$ are readily found.

Three treatment processes, chlorine, filters, and activated carbon, draw the untreated water at a controlled daily rate Q_{p}. De-contaminants are injected into the three treatment processes at daily rates of q_1, q_2, and q_3 ($q_j \ll Q_{\mathsf{p}}$). The pollution concentrations after the three treatments are C_{p}.

$$[C_{\mathsf{p}}(t)] = \exp\left[- [\mathbf{R}] \left[q(t)/Q_{\mathsf{p}}(t) \right] \right] \ [C_{\mathsf{u}}(t)]. \tag{20}$$

The terms R_{ij} in the matrix $\mathbf{R}$ give the effectiveness of process j in treating pollutant i, so that the processed water has the concentrations given in eqution (20). The reductions in concentrations increases with $\mathbf{R}(q(t)/Q_{\mathsf{p}}(t))$.

$$\mathbf{R} = \begin{array}{cc} & \begin{array}{ccc} \text{chlorine} & \text{filters} & \text{activated carbon} \end{array} \\ \begin{array}{c} \text{micro-organisms} \\ \text{suspended solids} \\ \text{petro-chemical} \end{array} & \begin{bmatrix} 1000 & 1000 & 250 \\ 50 & 2500 & 50 \\ 50 & 250 & 1000 \end{bmatrix} \end{array}$$

The treated water is then added to a treated water tank of capacity $V_{\mathrm{t,max}}$ and containing a volume of treated water V_{t} with pollutant concentrations C_{t}. The rate of change

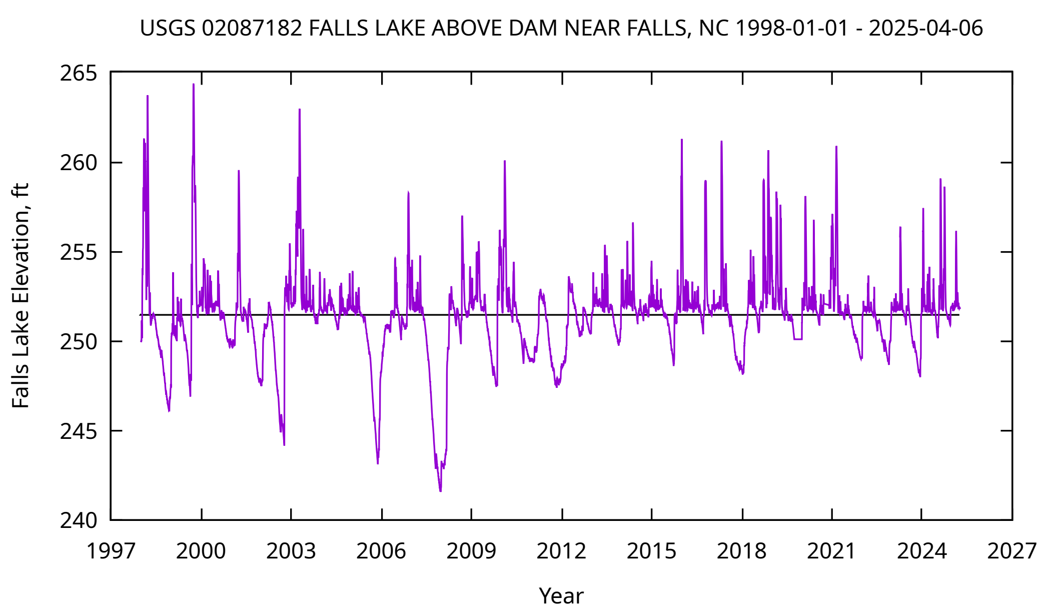

Figure 9. Falls Lake elevations, 1998-present, USGS 02087182.

of the volume of water in the treated tank is

$$\frac{d}{dt}V_{\mathsf{t}}(t) = Q_{\mathsf{p}}(t) - Q_{\mathsf{d}}(t) - \langle V_{\mathsf{t}}(t) - V_{\mathrm{t,max}}\rangle \tag{21}$$

where the term $\langle V_{\mathsf{t}} - V_{\mathrm{t,max}}\rangle$ represents water overflowing the treated tank to the river. The rate of change of the pollution mass the treated water tank is

$$\frac{d}{dt}m_{\mathsf{t}}(t) = C_{\mathsf{p}}(t)Q_{\mathsf{p}}(t) - C_{\mathsf{t}}(t)Q_{\mathsf{d}}(t) - C_{\mathsf{t}}(t)\langle V_{\mathsf{t}}(t) - V_{\mathrm{t,max}}\rangle \ , \tag{22}$$

from which the concentrations in the treated water, $C_{\mathsf{t}}(t) = m_{\mathsf{t}}(t)/V_{\mathsf{t}}(t)$ are readily found.

The daily water demand, $Q_{\mathsf{d}}(t)$, increases with temperature and population. If the reservoir falls below 50% capacity, then water conservation regulations are enacted and the daily consumption is reduced a factor C_{c}. This factor is a random variable in the simulations. A record of Durham water consumption along with drought severity indices are plotted in Figure 10.

$$Q_{\mathsf{d}}(t) = P(t)(100 + 0.4(T(t) - \bar{T}))\,(1 - C_{\mathsf{c}}\langle(0.50)V_{\mathrm{r,max}} - V_{\mathsf{r}}(t)\rangle/(0.50V_{\mathrm{r,max}} - V_{\mathsf{r}}(t))) + \delta Q_{\mathsf{d}} \tag{23}$$

The flow in the river is the sum of the inflows into the river

$$Q_{\mathsf{r}}(t) = Q_{\mathrm{r,min}} + \langle V_{\mathsf{r}}(t) - V_{\mathrm{r,max}}\rangle + \langle V_{\mathsf{u}}(t) - V_{\mathrm{u,max}}\rangle + \langle V_{\mathsf{t}}(t) - V_{\mathrm{t,max}}\rangle \ . \tag{24}$$

6 Performance and Costs

The performance of the water supply and treatment system is measured in terms of the total cost to build and operate the system for fifty years.

The cost of the reservoir ($V_{\mathrm{r,max}}$) is \$0.01/gallon. The installation cost for the untreated water tank ($V_{\mathrm{u,max}}$) is \$0.5M+\$0.10/gal and for the treated water tank ($V_{\mathrm{t,max}}$) is \$0.5M+\$0.20/gal. The installation costs for for the treatment system ($Q_{\mathrm{p,max}}$) is \$0.10/gal/-day.

The operating costs for the chlorine, filters, and activated carbon are \$0.03/gal, \$0.01/gal, and \$0.05/gal of decontaminant, respectively. If V_{t} falls below $0.1V_{\mathrm{t,max}}$, then a penalty of \$0.05 per gallon per day is charged for the water consumed that day. A penalty of \$0.10 per gallon per day is charged for days in which the pollutant concentrations in the treated water exceed their respective limits. And a penalty of \$10M per day is charged for days in which the river floods the downstream communities ($Q_{\mathsf{r}} > 5000$ Mg/d).

7 System Design

In this project, the design of the plant is specified by the reservoir capacity $V_{\mathrm{r,max}}$, the water treatment tank capacities $V_{\mathrm{u,max}}$ and $V_{\mathrm{t,max}}$, end the maximum allowable flow rate through the treatment paths, $Q_{\mathrm{p,max}}$. To get a sense of scale, an Olympic sized pool holds

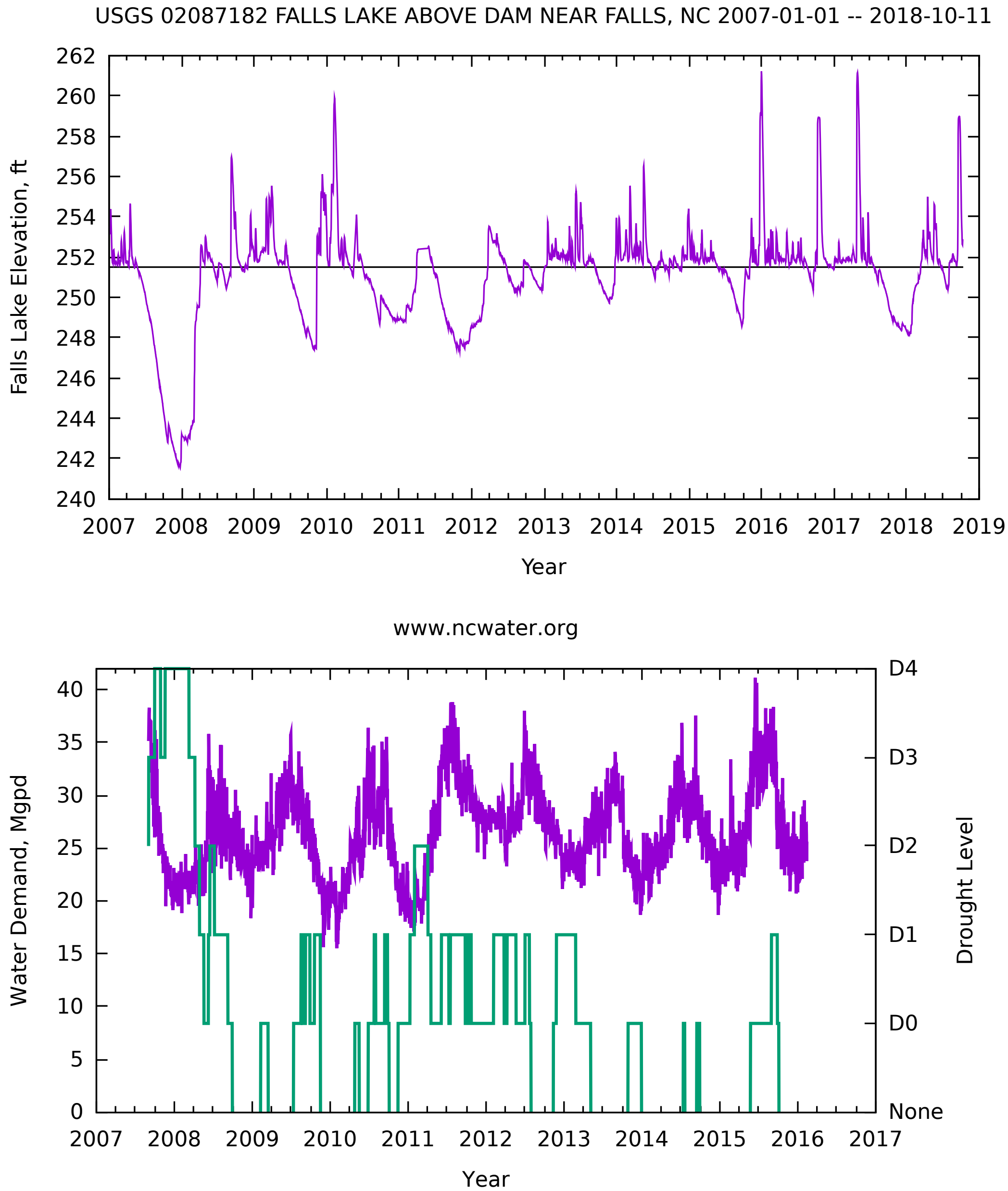

Figure 10. Falls Lake elevation, drought index, and Durham County water demand from 2007 to 2016. Note that there is practically no correlation between the drought index and the water demand.

about 0.5 Mgal. The operation of the plant is fully specified by a method of utilizing the measured condition of the system [$V_\mathsf{r}(t)$, $V_\mathsf{u}(t)$, $V_\mathsf{t}(t)$, $C_\mathsf{u}(t)$, and $C_\mathsf{t}(t)$] to regulate the flows [$Q_\mathsf{u}(t)$, $Q_\mathsf{p}(t)$, $q_1(t)$, $q_2(t)$, $q_3(t)$] in order to ensure that enough clean water is available for the municipality in a cost-effective manner. In order to succeed in meeting these requirements, the plant must operate in an uncertain dynamic environment, in which environmental conditions and water consumption vary from day to day with long-term trends, and with random fluctuations.

These issues can be systematically represented as a standard control problem, as illustrated in Figure 11. This kind of a control system is called a "two-input, two-output" system because each of the two inputs to the plant (the environment, w, and the controls, u) and each of the two outputs (the cost, z, and the measurements, y) have a qualitatively different interpretations.

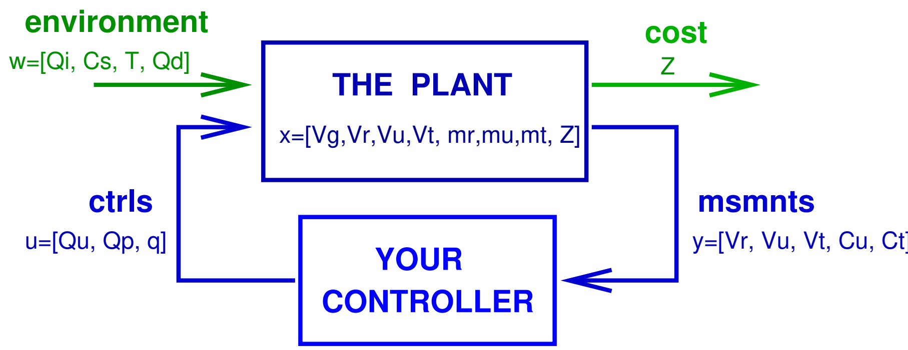

Figure 11. An abstraction of a water treatment plant as a standard control problem.

The mathematical description of this system is a *dynamic system*, represented by a set of first order ordinary differential equations. We can compute anything we might possibly want to know about the system uniquely from the *state*. In this water treatment system the *state vector*, x, contains the volumes of the reservoir and tanks, the concentrations in these volumes, and the current cost. In general, the state equations can be written as

$$\frac{d}{dt}x(t) = f(\ x(t),\ u(t),\ w(t)\ )\ . \tag{25}$$

This function is implemented in the `.m`-function `water_system.m`:[4]

```
function [dxdt,Q] = water_system( t, x, w, ode_constants );
```

This function simply implements the mass-balance and water purification relationships for this plant, and imposes limits on the volumes of water in the reservoir and tanks, and on the flow rate through the water treatment processes. The first four elements of the vector `ode_constants` are the capacities of the system, $[V_{\rm r,max}, V_{\rm u,max}, V_{\rm t,max}, Q_{\rm p,max}]$. You may choose to add additional design variables as you work through the design process. The information in this assignment should be enough for you to completely understand `water_system.m`. You should be able to re-derive any of the mass-balance equations implemented in this `.m`-function.

It is theoretically possible to destabilize the control system by attempting to control flows and volumes too quickly. In order to ensure that the system behaves in a desierably smooth fashion, the tank capacities must be at least as large as the plant processing capacities. Otherwise, the plant could drain the untreated tank in a day or overflow the treated

[4] http://www.duke.edu/~hpgavin/cee251/water_system.m and http://www.duke.edu/~hpgavin/m-files/ode4u.m

tank in a day. These constraints may be written,

$$\boldsymbol{g}(V_{\rm r,max}, V_{\rm u,max}, V_{\rm t,max}, Q_{\rm p,max}) = \left[\begin{array}{c} \phi\ Q_{\rm p,max}/V_{\rm u,max} - 1 \\ \phi\ Q_{\rm p,max}/V_{\rm t,max} - 1 \end{array} \right] \tag{26}$$

where ϕ could be a safety factor greater than 1.

To simulate the behavior and performance of the water treatment plant over it's design life, constants describing the system and the initial state of the system $x(1)$ are first established. The initial cost depends on the tank capacities and the treatment flow capacities. For each time increment, the current plant controls and operating environment are used to compute the state of the plant on the next time increment using the `.m`-function `water_system.m`. This simulation is carried out in the `.m`-function `water_analysis.m`[5].

```
function [cost, constraints] = water_analysis([Vr_max,Vu_max,Vt_max,Qp_max], constants )
```

The four *epistemic* randomly uncertain variables [$CCTS$, $T_{\sf c}$, P_2, $C_{\sf c}$] are assumed to be log-normally-distributed. They are input as members of the set of `constants`.

		median	c.o.v.
climate change time scale	$CCTS$	50 y	0.30
temperature rise	$T_{\sf c}$	4 deg/(365 × $CCTS$).	0.20
population growth	P_2	54 people/day^2	0.10
conservation effectiveness	$C_{\sf c}$	0.15	0.30

Other random variables have normal, Poisson, exponential or lognormal distributions.

temperature	δT	$\mu = 0$, $\sigma = 5$ deg. F	normal
rainfalls	W	$T_r = 3.13(1 + (d/(CCTS \times 365)))$ days	uniform
rainfall amnt.	R	$\bar{r} = 0.096\ T_r$, $c_r = 1.47$	gamma
rainfall area	$A_{\sf r}$	med.= $0.6A_{\sf w}$, c.o.v.=0.90	lognormal
population	δP	$\mu = 0$, $\sigma = 1500$ people	normal
consumption	$\delta Q_{\sf d}$	$\mu = 0$, $\sigma = 5$ gpppd	normal

Simulating random numbers for these variables makes use of *these* `.m`-functions: `randn.m`, `rand.m`, `gamma_rnd.m`,[6] and `logn_rnd.m`.[7]

The simulation generates four sets of plots:

1. Trends in statistical averages of rainfall frequency and rainfall intensity;
   Cumulative precipitation and cumulative transpiration in Mgal; and
   The one-year precipitation index

2. Daily ground water flow and daily evaporation in Mgal/day;
   Volumes $V_{\sf g}$, $V_{\sf r}$, $V_{\sf u}$, and $V_{\sf t}$ normalized to their capacities; and
   Stream flow and river flow $Q_{\sf s}$ and $Q_{\sf r}$.

---

[5] http://www.duke.edu/~hpgavin/cee251/water_analysis.m and http://www.duke.edu/~hpgavin/cee251/water_constants.m

[6] http://www.duke.edu/~hpgavin/m-files/gamma_rnd.m

[7] http://www.duke.edu/~hpgavin/m-files/logn_rnd.m

3. The population, P;
   The daily water demand, Q_{d} expressed in gallons per person per day; and
   The accumulation of the initial costs and operating costs, Z verses time expressed in M$.

4. The concentrations C_{s}, C_{r}, C_{u}, and C_{t} verses time.
   These concentrations are normalized by the respective allowable values.
   A concentration above 1 means that it is in excess of the allowable limit.

The optimization problem is set-up and executed in the `.m`-function `water_opt.m`.[8] Values assigned in `water_opt.m` may be modified.

```
% hpg_water_opt.m -- optimize the water treatment system for
% [ Vr_max,  Vu_max,  Vt_max,  Qd_max ]

water_constants;  % assign numerical values to all the constants in this system

Years = 50     % duration of the analysis
Plots =  0     % 1: draw plots, 0: don't draw plots

analysis_constants{end-1} = Years;

%          Vr,max   Vu,max   Vt,max    Qp,max
%            Mg       Mg       Mg        Mg/day
design_vars_init = [ ???? ;  ???  ; ???  ;    ???     ];

% evaluate the initial guess  ----------------------------------------
 analysis_constants{end} = 1;  % plots on
 f_init = water_analysis( design_vars_init, analysis_constants )

 if (input('   OK to continue? [y]/n : ','s')=='n')  return; end

design_vars_lb = [  ???? ;  ??? ;  ??? ;  ??? ];   % lower bound on design variables
design_vars_ub = [ ????? ; ???? ; ???? ; ???? ];   % upper bound on design variables

% algorithmic constants ...
%             display  tolX  tolF   tolG  MaxEvals  Penalty  Exponent  nMax errJ
  options = [    2     0.10  0.10   1.0   500       1000      2.0       20  0.05];

 analysis_constants{end} = Plots;  % 1: draw plots;   0: don't draw plots ...
[ design_vars_opt, f_opt, g_opt, cvg_hst ] = ...
    NMAopt('water_analysis', design_vars_init, design_vars_lb,design_vars_ub, options, analysis_constants

 plot_cvg_hst ( cvg_hst, design_vars_opt, 20 );

% assess the optimized design ... consider assessing the optimized design a few times
% ... consider assessing the optimized design a few times
 analysis_constants{end} = 1;  % plots on
 f_opt = water_analysis( design_vars_opt, analysis_constants )
```

To evaluate your initial guess, before embarking on an optimization, and in order to quickly track the effects of changes in design variables during the initial design phase, set `years = 10` to just simulate the system for a few years, This random sequence is specified by the value assigned to `my_favorite_number`. Changing `my_favorite_number` will result in a different random sequence. An automated design optimization will run faster if plots are not displayed. Setting `Plots = 0` will suppress plotting. Initially, it may be helpful to try analyses for shorter durations of time by setting `years = 5` or `years = 10`. Ultimately, however, your analysis will need to run for `years = 50`.

[8] http://www.duke.edu/~hpgavin/cee251/water_opt_draft.m

8 Controlling a Drinking Water Treatment System

The means for controlling the drinking water treatment system will be implemented in *your* `.m`-function `my_water_control.m`[9]

```
function u = my_water_control( msmnts, design_vars )
```

This function computes the controls for the plant based on some measurements of the condition of the water supply system, $[V_{\mathsf{r}}, V_{\mathsf{u}}, V_{\mathsf{t}}, C_{\mathsf{u}}, C_{\mathsf{t}}]$ and the design variables. Based on these measurements and the design variables, this function calculates the controls to be implemented, `u`, $[Q_{\mathsf{u}}, Q_{\mathsf{p}}, q_1, q_2, q_3]$. These calculations need to be written in to `my_water_control.m` in the three lines marked in the code. You may choose to add additional design variables as you work through the design process.

```
function u = my_water_control( msmnts, design_vars )
% my_water_control - determine water system control actions
% based on measurements

% YOUR NAME, Civil Eng'g, Duke Univ,  THE DATE

% Resevoir capacity, Mgal
 Vr_max =  design_vars(1);

% volume capacity for the untreated water storage tank, Mgal
 Vu_max =  design_vars(2);

% volume capacity for the   treated water storage tank, Mgal
 Vt_max =  design_vars(3);

% water flow capacity for the water treatment plant, Mgal/day
 Qp_max =  design_vars(4);

 Vr = msmnts(1);                % resevoir volume
 Vu = msmnts(2);                % untreated volume
 Vt = msmnts(3);                %   treated volume
 Cu = msmnts(4:6);              % untreated concentrations
 Ct = msmnts(7:9);              %   treated concentrations

% use the measurements of volumes and concentrations
% (current values and allowable limits)
% to determine flows through the treatment plant

% flow from reservoir into untreated tank
 Qu   =   ?????

% flow processed through the treatment plant
 Qp   =   ??????

% flows for each of the three decontaminants
 q = [ ???? ; ???? ; ???? ];

 u = [ Qu ;  Qp ; q ];

% my_water_control ------------------------------------------ 14 Mar 2022
```

If a tank is nearly full, the control should not call for more water flowing to the tank. If a tank is nearly empty, the control should call for more water flowing to the tank. This kind if inverse relationship is presented in figure 12. You may want to incorporate or build-upon

[9] http://www.duke.edu/~hpgavin/cee251/my_water_control_draft.m

such ideas in your drinking water control system. You may consider a straight (simple) or some kind of curved (complicated) inverse relationship, as shown in the figure.

Which tank volume ($V_r/V_{r,max}, V_u/V_{u,max}, V_t/V_{t,max}$) should be used to determine the daily flow Q_p through the water processing plant?

Which tank volume ($V_r/V_{r,max}, V_u/V_{u,max}, V_t/V_{t,max}$) should be used to determine the daily flow Q_u into the untreated tank?

The water purification flows, q_i could be the same from day to day, or they could increase somehow with the pollutant concentrations in the water to be treated and/or the amount of water to be treated.

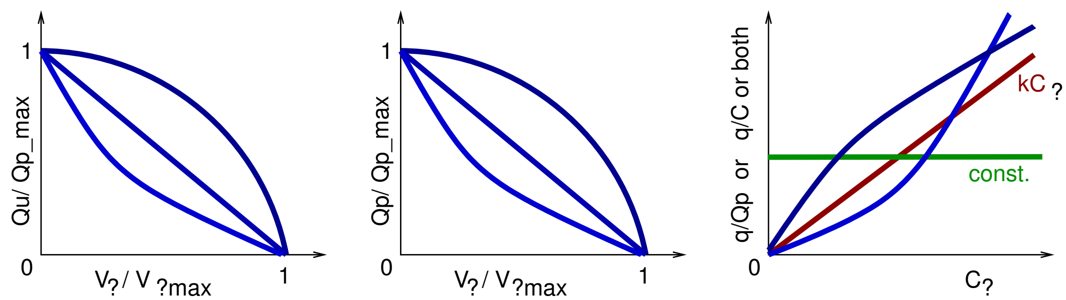

Figure 12. Qualitative representation of some candidate water control functions.

9 Tasks and Design Problems

The following tasks will gude you through the project:

1. How should the plant be operated?

   On any day, t, given measurements of:

   - the current volume of the reservoir storage, $V_{\mathsf{r}}(t)$,
   - the current volume of the untreated storage, $V_{\mathsf{u}}(t)$,
   - the current volume of the treated storage, $V_{\mathsf{t}}(t)$,
   - the untreated contaminant concentrations, $C_{\mathsf{u}}(t)$, and
   - the treated contaminant concentrations, $C_{\mathsf{t}}(t)$,

   determine an economical and appropriate way to regulate the flows (the *controls*)

   - $u = [Q_{\mathsf{u}},\ Q_{\mathsf{p}},\ q_1,\ q_2,\ q_3]$.

   Implement your control policy in the `.m`-function my_water_control.m which calculates $u = [Q_{\mathsf{u}},\ Q_{\mathsf{p}},\ q_1,\ q_2,\ q_3]$ as a function of $y = [V_{\mathsf{r}},\ V_{\mathsf{u}},\ V_{\mathsf{t}},\ C_{\mathsf{u}},\ C_{\mathsf{t}}]$.

2. How big should the water supply system and the treatment plant be?

   - How big should the reservoir, $V_{\mathrm{r,max}}$, be?
   - How big should local storage be at inlet and at outlet, $V_{\mathrm{u,max}}$ and $V_{\mathrm{t,max}}$?
   - How big should the flow capacity in the water treatment plant, $Q_{\mathrm{p,max}}$ be?

   These four constants are the design variables to `water_analysis.m`,

   `param = [ Vr_max, Vu_max, Vt_max, Qp_max ]`

   Optimize these design variables in order to economically meet the following criteria:

   - to not deplete the treated tank storage;
   - to not flood down-stream communities;
   - to meet the water consumption demands of the municipality; and
   - to supply drinking water that meets standards for allowable pollutant concentration levels.

3. What is the performance uncertainty associated with your final design?

   - Once you have determined the answers to questions (1) and (2) above, determine the risk of excessive cost overruns due to uncontrollable environmental factors, such as droughts and population growth.
   - Download the script called `water_montecarlo.m`[10].

[10] http://www.duke.edu/~hpgavin/cee251/water_montecarlo.m

- Enter your best design variables on line 15 of `water_montecarlo.m`
- Run `water_montecarlo.m` and interpret the plots of the resulting probability functions.

```
1  % water_montecarlo.m
2  % Monte-Carlo analysis for cost and sensitivity of a water supply design system
3  % to be run after running truss_opt.m
4  % CEE 201, Duke University, HP Gavin, 2019, 2022
5
6  % on line 9 of water_constants.m . . . Plots = 0;
7
8  %  v(1)  Vr_max  volume of the resevoir                        Mgal
9  %  v(2)  Vu_max  volume of the untreated water tank            Mgal
10 %  v(3)  Vt_max  volume of the treated water tank              Mgal
11 %  v(4)  Qp_max  max. flow through the water treatment plant   Mgal/day
12 %   . . . etc . . . if you have other design variables to include . . .
13
14 %              Vr,max Vu,max Vt,max Qp,max
15 % opt_v = [ ??  ; ??   ; ??   ; ??   ]  %  * PUT YOUR BEST PARAMETERS HERE *
16
17 water_constants;                  % assign numerical values to system constants
18 analysis_constants{end-1} = 50;   % 50 year simulation
19 analysis_constants{end}   = 1;    % show plots
20
21 [ cost , constraints ] = water_analysis(opt_v,analysis_constants)
22 if (input('   OK to continue? [y]/n : ','s')=='n')  return; end
23
24 NP  = length(opt_v);                   % number of design variables
25 NS  = 100;                             % total number of simulations            *
26 NR  = 4;                               % number of random variables in the MCS
27
28 % --- probability models for uncertain quantities
29
30  % a sample of NS observations of NR uncorrelated log normal random variables
31  rv = corr_logn_rnd([CCTS; Tc; P2; Cc], [0.3; 0.2; 0.1; 0.3], eye(NR), NS);
32
33  cost84   = zeros(1,NS);  % 84th percentile cost
34  cost_avg = zeros(1,NS);  % average cost
35  avg_cost = 0;            % average cost
36  ssq_cost = 0;
37
38  figure(10);
39  for sim = 1:NS            % Monte Carlo simulation  (MCS)
40
41      analysis_constants{23}  =  rv(1,sim);    % CCTS
42      analysis_constants{28}  =  rv(2,sim);    %  Tc
43      analysis_constants{30}  =  rv(3,sim);    %  P2
44      analysis_constants{14}  =  rv(4,sim);    %  Cc
45
46      cost(sim) = water_analysis( opt_v, analysis_constants );
47
48      delta_cost    = cost(sim) - avg_cost;
49      avg_cost = avg_cost + delta_cost/sim;
50      ssq_cost      = ssq_cost + delta_cost*(cost(sim) - avg_cost);
51      if(sim>1) cost84(sim)   = avg_cost + sqrt(ssq_cost/(sim-1)); end
52      cost_avg(sim) = avg_cost;
53  end
```

10 Pointers

- Start by downloading the `.m`-files linked in the footnotes of this assignment.
- Before you can run any analysis you will need to come up with a water control method, as described in the previous section, and to enter this method in three (or more?) lines of code in `my_water_control.m` Initially the controls can be constant values, but to make the system run efficiently, you will need to make the controls functions of the measurements.
- Once you have thought of a control strategy and entered it into `my_water_control.m`, you need to think of reasonable volumes for the storage capacities $V_{\mathrm{r,max}}$, $V_{\mathrm{u,max}}$, and $V_{\mathrm{t,max}}$, and the treatment capacity, $Q_{\mathrm{p,max}}$. Use estimates of population size and consumption in gpppd to get rough guesses for these capacities.
- Evaluate your initial guess of the design variables and your water control method by examining the results of a single analysis. Run `water_opt` and if not satisfied with the result, enter `n` in the command window to exit before optimization starts. Try (try) again with another initial guess.
- To find a good initial guess examine the plots and think about how how to improve your system. Take notes on what you notice and try to learn about what is happening inside this system.
  - If a tank is more than full, water is over-flowing, and draining the reservoir unnecessarily.
  - If your tanks are getting close to empty, you might need bigger capacities, V_{max}, Q_{max}.
  - If one of the pollutants is excessively high, increase the decontaminant flow, q, to the process that is most effective for that contaminant.
- Rainfall, temperature, population, stream-flow, concentrations of stream-flow pollutants, and the daily water consumption vary in a systematic fashion but also have random fluctuations. This could be helpful when starting to optimize your design. Ultimately, however, if the design is optimized only for a particular time-series of precipitation, population, etc, it may not perform well under other situations. Since these environmental conditions can not be predicted precisely, the optimization should ultimately be carried out for random environmental conditions. In doing so, it will be helpful to average the results from a few simulations to compute each analysis point. The eighth element of the algorithm `options` vector specifies the number of stochastic analyses that go into a single average.
- To evaluate your initial guess before optimization, try to get good performance for just a few years. You can shorten the time span of the simulation by changing the value for `years` in `water_opt.m`. Try to get a feel for the effects of changing the controls by playing around with the simulation. Running short simulations to start with will help because the shorter simulations take less time to compute, and because you will not

have to contend with long-term and random population and drought trends. When things work well for two years, try for 5 years, 10 years, and finally 50 years.

- Once you have written your function `my_water_control.m`, and you are happy with your initial guess for the four design variables, you may wish to try to optimize your water supply and treatment system using `NMAopt.m` or `ORSopt.m`. When using these optimization routines with `water_analysis.m`, set `Plots = 0`. You can expect the optimization of the 50-year cost to take two to five minutes on a fast computer running MATLAB.
- Once you have optimized your design to be satisfactory for 50-years, run `water_analysis.m` several times in order to get a sense of the effects of uncertain population growth, uncertain climactic conditions, and uncertain effects of water restriction policy.
- The simulation is computed on a day-to-day basis. You can adjust your controls only once a day. So if the tank can hold only one or two days of water, then you could have a problem. In actual water treatment plants, controls are adjusted much more frequently, but in this simulation you can not respond that quickly. Therefore, in this assignment the tank volumes that work well might be larger than what one would design for in a real treatment plant.
- Emily S. Rueb and Josh Cochran, How New York Gets Its Water, NYT Mar. 24, 2016
- Melanie Burford and Greg Moyer, A Marvel of Engineering Meets the Needs of a Thirsty New York, NYT Oct. 16, 2014
- Matt Flegenheimer, After Decades, a Water Tunnel Can Now Serve All of Manhattan, NYT Oct 16, 2013
- Umair Irfan, Eliza Barclay, and Kavya Sukumar, Weather 2050, Vox, 2018-10-31
- USGS Water Use in the United States, USGS, 2020-03-31
- USGS Water Use in North Carolina, 2010 USGS, 2018-03-31
- *Have fun!*

11 Report contents

Each group of two or three people should prepare a report including:

1. A written description of the control volume analysis implemented in lines 65-166 of the `.m`-function `water_system.m` (1 - 2 pg).
2. A written description of the method by which you determined what would be a suitable control policy, a description of any additional design variables you used, a sketch of your water control functions, such as those shown in Figure 12, (1 - 2 pg).
3. A printout of your `.m`-function, `my_water_control.m`.
4. A written description of the method by which you decided what would be good values for the storage capacities, $V_{\mathrm{r,max}}$, $V_{\mathrm{u,max}}$ $V_{\mathrm{t,max}}$, and $Q_{\mathrm{p,max}}$, and your values for these quantities. (0.5 - 1 pg)
5. A printout and discussion of the four sets of plots from an analysis of your final design. (1 pg)
6. A printout and discussion of the four scatter plots and the CDF plot from the Monte-Carlo simulation showing the variability in life-time operating costs and the correlations of the lifetime operating costs with the random quantities assessed, ($CCTS$, dT_{c}, P_2, C_{c}) as determined from question 3. (1 pg)
7. A written description of problems you had in achieving the goals of the assignment, if any. (0.5 pg).
8. A table of results for your final design:

$V_{\mathrm{r,max}}$	$V_{\mathrm{u,max}}$	$V_{\mathrm{t,max}}$	$Q_{\mathrm{p,max}}$	85th percentile cost
MG	MG	MG	MGPD	M$

12 Grading

- 10 points: 85th percentile lifetime operating cost, in comparison with other group projects (from plots generated by `water_montecarlo.m`)
- 20 points: quality of the report: (typed, 11pt or 12 pt, 1.5 space, 1 inch margins, page numbers, figure numbers, figure captions, etc.)
- 30 points: completeness and correctness of descriptions and discussions
- 40 points: including all eight parts above.

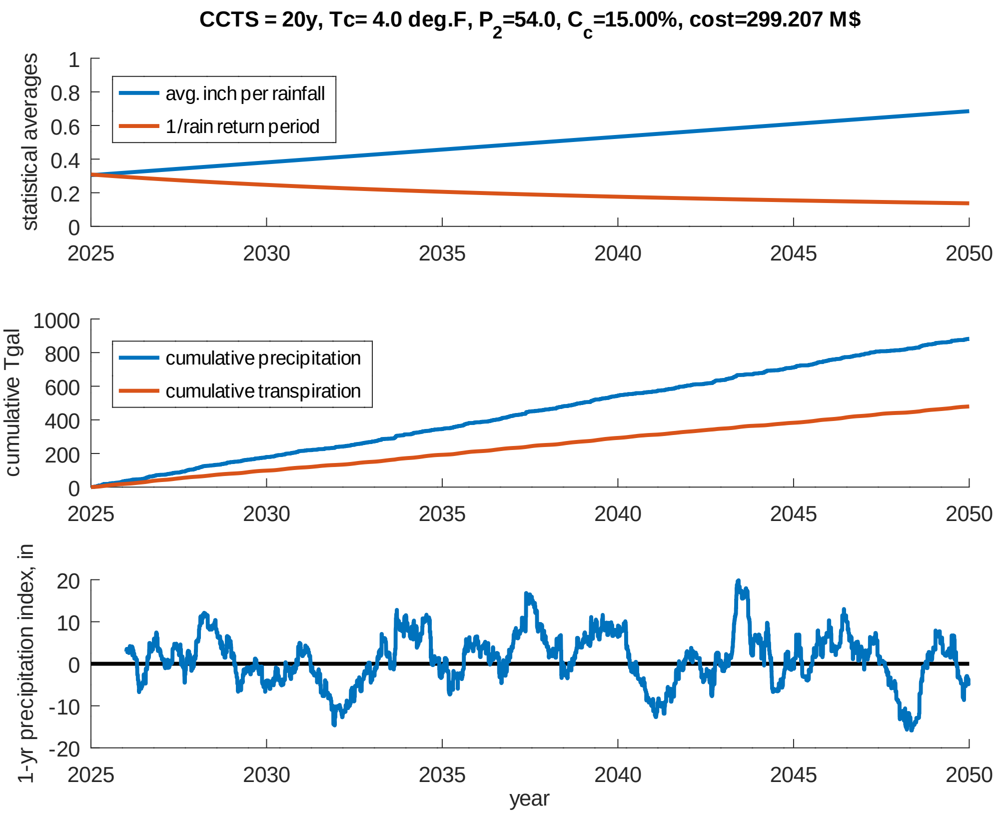

Figure 13. Precipitation and Transpiration

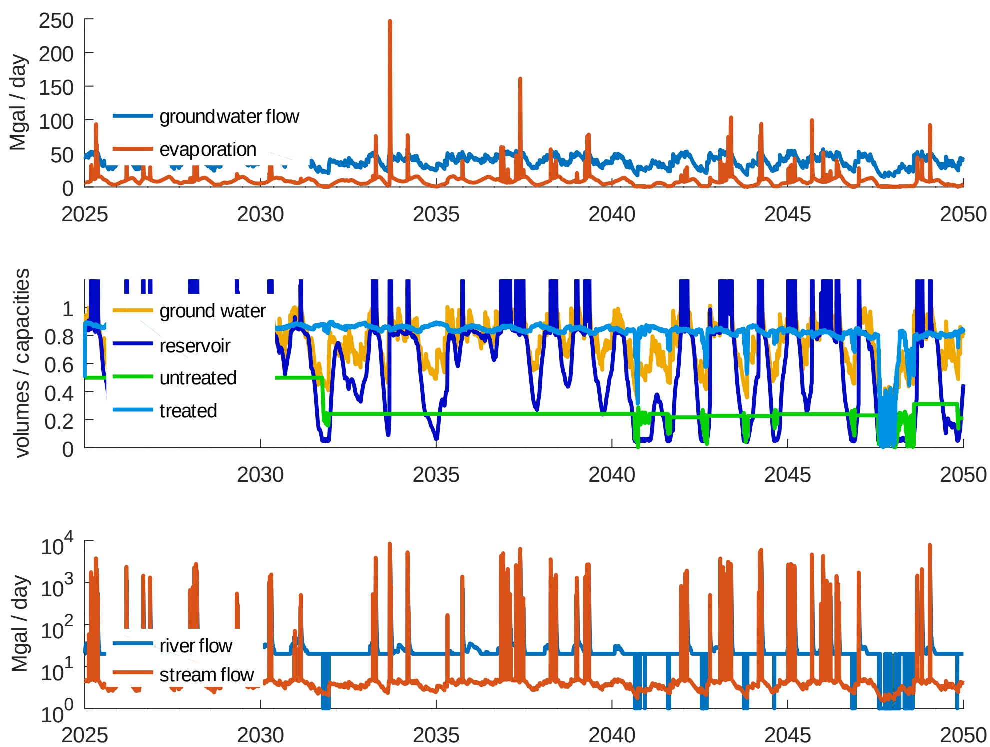

Figure 14. Evaporation, Ground water flow, Volumes, Stream Flow, and River Flow

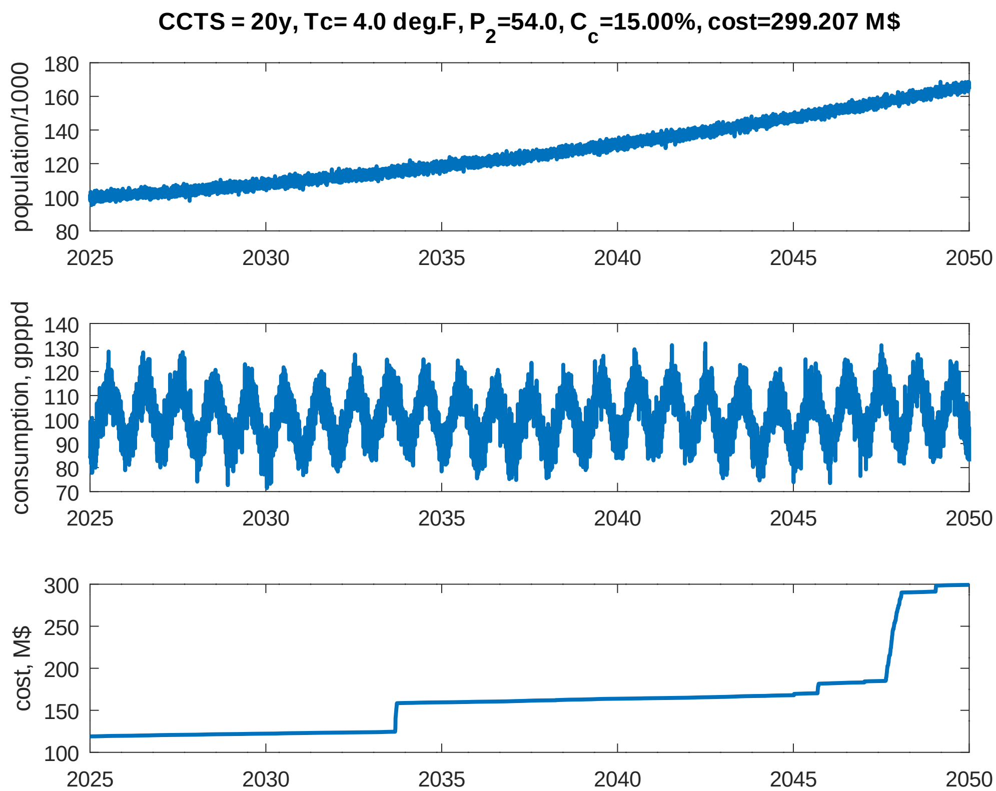

Figure 15. Population, Consumption, Cost

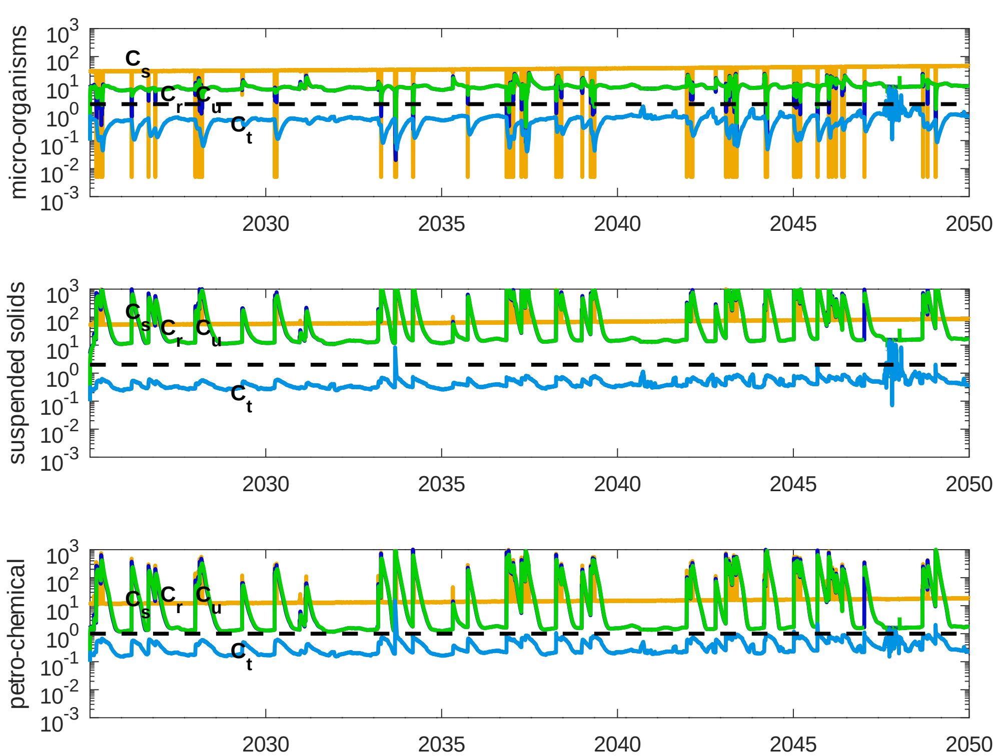

Figure 16. Water Quality