

# A Data-Driven Method for Robust Water Allocation under Uncertainty

David Love and Güzin Bayraksan

## Abstract

We investigate the application of techniques from distributionally robust optimization to water allocation in under future uncertainty and apply it to a southeastern portion of Tucson, AZ. Tucson, like most of the southwestern United States, faces considerable uncertainty in its ability to provide water for its citizens in the future. The sources of this uncertainty include (1) the future population growth of the region, (2) the availability of water from the Colorado River, and (3) the effects of climate change and how it relates to water usage. This paper presents a new data-driven approach for integrating forecasts for all these sources of uncertainty in a single optimization model and provide insight on the value of constructing additional treatment facilities to mitigate the future risks.

**Keywords:** Data-driven optimization, water allocation, climate change

## 1 Introduction and Motivation

The continued availability of water resources is in question in many parts of the world, especially in light of an uncertain climate variability and population increase. To combat these influences, methods of allocating water resources must adapt. The Lower Colorado River Basin supplies water to 25 million people in California, Nevada and Arizona, in addition to Mexico, making the management of this resource very important. In Tucson, Arizona the Colorado River supplies more than 60% of the water through the Central Arizona Project (CAP), the largest and most expensive aqueduct system in the United States. This paper studies allocation of CAP water in a developing portion of Tucson under future climate and population uncertainty, creating a single integrated model using projections for each of these sources of uncertainty.

Climate variability is one of the most important sources of uncertainty in evaluating the sustainability of water resources. A wide variety of general circulation models (GCMs) have been developed for evaluating the effects of climate change. The results from these models are used as drivers for Regional Climate Models (RCMs) to generate temperature and precipitation predictions on finer scales. To gain insight on still smaller regions, Bias Correction and Spatial Downscaling (BCSD) techniques have been applied to the results of the RCMs to generate predictions on even finer scales [18]. Methods of comparing and integrating the results of these differing models are important parts of water allocation planning. However, these predictions contain significant uncertainties in themselves, with potential modeling errors in every step of the prediction process. This poses an important challenge in decision making when these predictions are used.

In this paper, we present a new model for sustainable water allocation to handle uncertainties from various data sources in a robust manner. This modeling approach, recently proposed and investigated by Ben-Tal et al. [2] and Love and Bayraksan [14], is particularly attractive in its ability to handle situations where the future uncertainty model is itself uncertain, like the climate

models used to predict future water use. It allows for using the data directly, integrating those scenarios of interest into a single modeling formulation. These scenarios can come from direct observation, results of simulation, or from expert opinion that the decision maker would especially like to be robust against. In this water allocation problem, scenarios are generated from projections of population growth, Colorado River water supply, and climate-driven per-capita demand for potable and nonpotable water use.

## 1.1 Related Literature

Models of optimization under uncertainty, especially when the uncertainty is represented stochastically, have a long history in water allocation problems. One early example is given by Hall and Howell [9], which formulates a stochastic linear programming model for water rights allocation in arid and semiarid agriculture. Bishop et al. [3] provide an early stochastic model of a water allocation system, including wastewater treatment and reuse. More recent work includes Higgins et al. [11], which provides a multistage nonlinear stochastic formulation of water resource allocation and applies it to South East Queensland in Australia. Other studies have used interval-stochastic programming to formulate the model, e.g., [13]. Shao et al. [21] combine inexact stochastic programming with uncertainty measured by the Conditional Value at Risk (CVaR). This is particularly similar to the work presented here, because CVaR can be represented in the form presented in this paper [14].

Another optimization model for water distribution, developed by Wang et al. [23], considers the entire South Saskatchewan River Basin. This study applies two multi-objective optimization models to generate water management solutions that are both equitable and economically efficient using the Cooperative Water Allocation Model.

Most similar to the water allocation model considered here is the CALVIN (CALifornia Value Integrated Network) model that has been used to study water distribution in California [8, 12]. CALVIN is a multi-period generalized network flow model developed to minimize the total economic cost of water distribution in California. Harou et al. [10] integrated sustained drought scenarios into the CALVIN model based on historical records of long-term droughts in California. Later, Connell-Buck et al. [7] integrate the GFDL-CM3 climate model, one of the climate models used in this work, into CALVIN as a method of evaluating the impact of a warmer and dryer climate on water availability in the future.

Several studies have been directed towards evaluating the relative benefits of the GCMs and associated RCMs. NARCCAP, the North American Regional Climate Change Assessment Program, compares several models by showing how they perform using historical observations and computing climate scenarios in the 21st century. Mearns et al. [15] compare the results of six RCMs using observational data over much of the United States and Canada from 1980-2004. Brekke et al. [4] study 17 climate models from the World Climate Research Programme’s (WCRP’s) Coupled Model Intercomparison Project phase 3 (CMIP3) and determine model credibility criteria for Northern California. This comparison helped to inform our choices of climate models included in the present work.

## 1.2 Contributions

This paper integrates multiple projections on future uncertainty to produce a unified model for water allocation under uncertainty. We pair the water allocation model developed by Woods et al. [25] with the uncertainty modeling technique described in [14] to develop a novel data-driven water allocation problem. This formulation is able to combine projections of future uncertainty, including

population growth, water availability and climate variation, into a single model. When solved, we obtain estimates of the severity and likelihood of future water shortages, the expected operating cost of the allocation system and a weighing of the relative desirability of each possible scenario. In addition, we calculate the effect of several alternative infrastructure scenarios that may decrease both the operating cost and severity of the future shortages.

### 1.3 Organization

The rest of the paper is organized as follows. Section 2 presents the water allocation model. Section 3 briefly describes the  $\phi$ LP-2 approach from [14] and reviews some of the key features. Section 4 describes the application to Tucson, AZ and Section 5 presents the results and a discussion on water allocation and infrastructure decisions and recommendations as a result of this study. We end in Section 6 with our conclusions.

## 2 Generalized Network Water Allocation Model

A generalized network model of water allocation is defined by a set of nodes and directed arcs  $(N, A)$ . The nodes represent available water supply inputs, water treatment plants, reservoirs, interconnection points, and water demand sites. The arcs represent the conveyance system (pipes, etc.) that carry water between the nodes. Water can be stored in between time periods in reservoirs to meet future demands. The model aims to find the minimal cost water flows considering energy, treatment, storage, and transportation costs over the planning period.

### 2.1 Deterministic Model

Water flows on arc  $(i, j) \in A$  during time period  $t = 1, \dots, P$  are represented by decisions  $x_{ijt}$ . Each arc  $(i, j) \in A$  and time period  $t$  has a unit cost  $c_{ijt}^x$ , loss coefficient  $0 \leq a_{ijt} \leq 1$  to account for evaporation, leakage from the pipes, etc. and bounds on the flow  $l_{ijt}^x \leq x_{ijt} \leq u_{ijt}^x$ . Each node  $j \in N$  has a supply/demand for time period  $t$ ,  $b_{jt}$ . Nodes representing reservoirs are able to store water between time periods. Nodes capable of storing water are denoted as members of the set  $S$ . Stored water available at node  $j$  at the end of time period  $t$  is  $s_{jt}$ , with associated cost  $c_{jt}^s$  and bounds  $l_{jt}^s \leq s_{jt} \leq u_{jt}^s$ . Finally, water released into the environment from node  $j$  in period  $t$  is given by  $r_{jt}$ , with bounds  $l_{jt}^r \leq r_{jt} \leq u_{jt}^r$ . The deterministic model is a multi-period generalized network flow model of the form

$$\begin{aligned}
\min_{x,s,r} \quad & \sum_{(i,j) \in A} \sum_{t=1}^P c_{ijt}^x x_{ijt} + \sum_{j \in N} \sum_{t=1}^P c_{jt}^s s_{jt} \\
\text{s.t.} \quad & \sum_{j:(j,i) \in A} a_{jit} x_{jit} = \sum_{j:(i,j) \in A} x_{ijt} + r_{jt} + b_{jt} \quad \forall i \notin S, t \\
& \sum_{j:(j,i) \in A} a_{ji,t-1} x_{ji,t-1} + s_{j,t-1} = \sum_{j:(i,j) \in A} x_{ijt} + s_{jt}, \quad \forall i \in S, t \\
& l_{ijt}^x \leq x_{ijt} \leq u_{ijt}^x, \quad \forall i, j, t \\
& l_{jt}^s \leq s_{jt} \leq u_{jt}^s, \quad \forall j, t \\
& l_{jt}^r \leq r_{jt} \leq u_{jt}^r, \quad \forall j, t.
\end{aligned}$$

To compute the first-year constraints, we assume that  $x_{ij0} \equiv 0$  and that initial storage levels  $s_{j0}$  are given.

The deterministic model has the obvious shortcoming that many of the parameters are unknown. Often, the unknowns are represented by solving distinct scenarios that cover the range of possibilities. However, even in this case, the models provide no insight on how much weight each result should be given. To compensate for these problems, we formulate a stochastic water allocation model in the next section.

## 2.2 Stochastic Model

The model is converted to a two-stage stochastic model with  $P_1$  periods in the first stage and  $P - P_1$  stages in the second stage. We assume that the supplies and demands are uncertain, as well as available water treatment capacities (as an effect of relative population growth).

$$\begin{aligned}
\min_{(x,s,r) \in L^1} \quad & \sum_{(i,j) \in A} \sum_{t=1}^{P_1} c_{ijt}^x x_{ijt} + \sum_{j \in N} \sum_{t=1}^{P_1} c_{jt}^s s_{jt} + \sum_{\omega=1}^n q_{\omega} h_{\omega}(s) \\
\text{s.t.} \quad & \sum_{j:(j,i) \in A} a_{jit} x_{jit} = \sum_{j:(i,j) \in A} x_{ijt} + r_{jt} + b_{jt} \quad \forall i \notin S, 1 \leq t \leq P_1 \\
& \sum_{j:(j,i) \in A} a_{ji,t-1} x_{ji,t-1} + s_{j,t-1} = \sum_{j:(i,j) \in A} x_{ijt} + s_{jt}, \quad \forall i \in S, 1 \leq t \leq P_1,
\end{aligned} \tag{1}$$

where  $\{q_{\omega}\}_{\omega=1}^n$  is the probability distribution of the uncertain second stage problem  $h_{\omega}(s)$ , given by

$$\begin{aligned}
h_{\omega}(s) = \min_{(x,s,r) \in L_{\omega}^2} \quad & \sum_{(i,j) \in A} \sum_{t=P_1+1}^P c_{ijt}^x x_{ijt} + \sum_{j \in N} \sum_{t=P_1+1}^P c_{jt}^s s_{jt} \\
\text{s.t.} \quad & \sum_{j:(j,i) \in A} a_{jit} x_{jit}^{\omega} = \sum_{j:(i,j) \in A} x_{ijt}^{\omega} + r_{jt}^{\omega} + b_{jt}^{\omega}, \quad \forall j, P_1 + 1 \leq t \leq P, \\
& \sum_{j:(j,i) \in A} a_{ji,t-1} x_{ji,t-1} + s_{j,t-1} = \sum_{j:(i,j) \in A} x_{ijt}^{\omega} + s_{jt}^{\omega}, \quad \forall i \in S, t = P + 1, \\
& \sum_{j:(j,i) \in A} a_{ji,t-1} x_{ji,t-1}^{\omega} + s_{j,t-1}^{\omega} = \sum_{j:(i,j) \in A} x_{ijt}^{\omega} + s_{jt}^{\omega}, \quad \forall i \in S, P_1 < t \leq P,
\end{aligned} \tag{2}$$

and  $L^1$  and  $L_{\omega}^2$  represent the feasible regions defined by the lower and upper variable bounds.

## 2.3 Motivation for a Data-Driven Distributionally Robust Model

Although the stochastic model removes most of the problems with uncertain parameters, it has introduced the probability distribution  $q_1, \dots, q_n$  that itself needs to be estimated. The method of distributionally robust optimization helps to counteract this problem. In the distributionally robust formulation, the distribution is replaced by a “nominal” distribution that is extracted directly from the data and the optimization takes place over a set of distributions that are “similar” to the nominal distribution. The advantage is that both the nominal distribution and the level of similarity are extracted from the data and thus no parameters need to be estimated. The sequel describes the distributionally robust method presented in [14].

## 3 Review of Data-Driven Optimization using $\phi$ -Divergences

We now provide background on our modeling approach. This discussion is largely based on [14].

Divergence	$\phi(t)$	$\tilde{\phi}(t)$	$\phi(t), t \geq 0$	$I_\phi(p, q)$
Kullback-Leibler	$\phi_{kl}$	$\phi_b$	$t \log t - t + 1$	$\sum p_\omega \log \left( \frac{p_\omega}{q_\omega} \right)$
Burg Entropy	$\phi_b$	$\phi_{kl}$	$-\log t + t - 1$	$\sum q_\omega \log \left( \frac{q_\omega}{p_\omega} \right)$
$\chi^2$ -Distance	$\phi_{\chi^2}$	$\phi_{m\chi^2}$	$\frac{1}{t}(t-1)^2$	$\sum \frac{(p_\omega - q_\omega)^2}{p_\omega}$
Modified $\chi^2$ -Dist.	$\phi_{m\chi^2}$	$\phi_{\chi^2}$	$(t-1)^2$	$\sum \frac{(p_\omega - q_\omega)^2}{q_\omega}$
Variation Distance	$\phi_v$	$\phi_v$	$ t-1 $	$\sum  p_\omega - q_\omega $

Table 1: Definitions of some common  $\phi$ -divergences and their adjoints.

### 3.1 Introduction to $\phi$ -Divergences

$\phi$ -divergences are used in statistics to measure the “distance” between two distributions [16]. In the discrete and finite case, they measure the distance between two discrete distributions  $p$  and  $q$  given by vectors representing probability mass functions  $p = (p_1, \dots, p_n)^T$  and  $q = (q_1, \dots, q_n)^T$ , satisfying  $\sum_{\omega=1}^n p_\omega = \sum_{\omega=1}^n q_\omega = 1$ . The  $\phi$ -divergence is defined by

$$I_\phi(p, q) = \sum_{\omega=1}^n q_\omega \phi \left( \frac{p_\omega}{q_\omega} \right),$$

where  $\phi(t)$ , called the  $\phi$ -divergence function, is a convex function on  $t \geq 0$  such that  $\phi(t) \geq 0$  and  $\phi(1) = 0$  and with the additional interpretations that  $0\phi(a/0) = a \lim_{t \rightarrow \infty} \frac{\phi(t)}{t}$  and  $0\phi(0/0) = 0$ .

$\phi$ -divergences are not, in general, metrics. For example, most  $\phi$ -divergences do not satisfy the triangle inequality and many are not symmetric in the sense that  $I_\phi(p, q) \neq I_\phi(q, p)$ . One exception is the Variation distance (see Table 1), which is equivalent to the  $L^1$ -distance between the vectors.

A  $\phi$ -divergence has an adjoint, defined by

$$\tilde{\phi}(t) = t\phi \left( \frac{1}{t} \right), \quad (3)$$

which satisfies all criteria for a  $\phi$ -divergence [1] and has the property that  $I_{\tilde{\phi}}(p, q) = I_\phi(q, p)$ . Divergences that are symmetric with respect to the input vectors are known as self-adjoint.

Table 1 lists some common examples of  $\phi$ -divergences, along with their adjoints. We can see that the Kullback-Leibler divergence and the Burg entropy are adjoints. The  $\chi^2$ -distance is related to the famous  $\chi^2$  test [16] and its adjoint is the Modified  $\chi^2$ -distance.

### 3.2 Formulation

We can simplify the notation for the first-stage (1) and second-stage (2) problems as follows. In the first stage, decision variables  $\{x_{ijt}\}$ ,  $\{s_{jt}\}$  and  $\{r_{jt}\}$  become the vector  $\mathbf{x}$ , costs  $\{c_{ijt}^x\}$  and  $\{c_{jt}^s\}$  are written as the row vector  $\mathbf{c}$ , the supply/demand parameters  $b_{jt}$  become the vector  $\mathbf{b}$  and the constraint matrix is written as  $A$ . In the second stage, we denote the decisions as  $\mathbf{y}^\omega$ , the costs as  $\mathbf{k}^\omega$ , the supply/demands as  $\mathbf{d}^\omega$ , and the constraint matrices multiplying  $\mathbf{y}^\omega$  and  $\mathbf{x}$  as  $D^\omega$  and  $B^\omega$ , respectively.

We begin with a two-stage stochastic linear program with recourse (SLP-2). The SLP-2 is designed to model a situation where some decisions (occurring in the first stage) must be made

under some amount of uncertainty. This uncertainty will later reveals itself, after which the decision-maker can make additional choices (called *recourse decisions*) in light of this new information. See [19] for an overview of two-stage models in stochastic programming.

Let  $\mathbf{x}$  be a vector of first-stage decision variables with cost vector  $\mathbf{c}$ , constraint matrix  $\mathbf{A}$  and right-hand side  $\mathbf{b}$ . We assume a finite distribution given by  $q_\omega$  with scenarios indexed by  $\omega = 1, \dots, n$ . The SLP-2 is

$$\min_{\mathbf{x}} \left\{ \mathbf{c}\mathbf{x} + \sum_{\omega=1}^n q_\omega h_\omega(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \right\}, \quad (4)$$

where

$$h_\omega(\mathbf{x}) = \min_{\mathbf{y}^\omega} \{ \mathbf{k}\mathbf{y}^\omega : \mathbf{D}\mathbf{y}^\omega = \mathbf{B}\mathbf{x} + \mathbf{d}^\omega, \mathbf{y}^\omega \geq 0 \}. \quad (5)$$

We assume *relatively complete recourse*; i.e., the second-stage problems  $h_\omega(\mathbf{x})$  are feasible for every feasible solution  $\mathbf{x}$  of the first-stage problem; and that the second-stage problems  $h_\omega(\mathbf{x})$  are dual feasible for every feasible solution  $\mathbf{x}$  of the first-stage problem. For convenience, we denote  $X = \{ \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \}$ .

The SLP-2 formulation assumes that the distribution  $\{q_\omega\}_{\omega=1}^n$  is known. However, in many applications, like the water application in this paper, the distribution is itself unknown. One technique to deal with this is to replace the known distribution with a set of distributions that is believed to contain the true distribution. In this paper, we construct this set of distributions by considering all distributions whose  $\phi$ -divergence from the nominal distribution  $q$  is sufficiently small. We focus on a data-driven setting and assume that  $q$  is generated from observations, where scenario  $\omega$  has been observed  $N_\omega$  times, with  $N = \sum_{\omega=1}^n N_\omega$  total observations, although  $q$  can be obtained in other ways. In SLP-2, this data-driven setting would correspond to the probability of scenario  $\omega$  to be  $q_\omega = \frac{N_\omega}{N}$ . In our setting, the observations come from projections of climate, population and water availability, making  $N_\omega \equiv 1$  the most natural choice. However, some  $N_\omega$  can be increased if certain scenarios are believed to be more likely than others.

By replacing the specific distribution in SLP-2 with a set of distributions sufficiently close to the nominal distribution with respect to  $\phi$ -divergence, we create the model from [14]. In this model, the objective function is minimized with respect to the worst-case distribution selected from the set of distributions in an effort to be robust with respect to these uncertainties. The resulting minimax formulation is

$$\min_{\mathbf{x} \in X} \max_{p \in \mathcal{P}} \left\{ \mathbf{c}\mathbf{x} + \sum_{\omega=1}^n p_\omega h_\omega(\mathbf{x}) \right\}, \quad (6)$$

where the set of distribution  $\mathcal{P}$  is given by

$$\mathcal{P} = \left\{ \sum_{\omega=1}^n q_\omega \phi \left( \frac{p_\omega}{q_\omega} \right) \leq \rho, \right. \quad (7)$$

$$\left. \sum_{\omega=1}^n p_\omega = 1, \right. \quad (8)$$

$$\left. p_\omega \geq 0, \forall \omega \right\}. \quad (9)$$

Constraint (7) works to ensure that only distribution sufficiently close to  $q$  are selected. Constraints (8) and (9) simply ensure that  $p_1, \dots, p_n$  is itself a probability distribution. We discuss how to determine  $\rho$  in (7) in Section 3.3.

The problem (6) can be formulated in alternate ways that make it computationally tractable. Most notable is the fact that (6) is a convex problem, for which a wide variety of algorithms are available. For further discussion on computational methods, we direct the interested reader to [2] and [14].

### 3.3 The Level of Robustness

The literature on  $\phi$ -divergences provides some insight on choosing a reasonable asymptotic value of  $\rho$  in the data-driven setting. When  $\phi$  is twice continuously differentiable around 1 with  $\phi''(1) > 0$ , Pardo [16, Theorem 3.1] shows that the statistic  $T_N^\phi(q^N, q^{\text{true}}) = \frac{2N}{\phi''(1)} \sum_{\omega=1}^n q_\omega^{\text{true}} \phi\left(\frac{q_\omega^N}{q_\omega^{\text{true}}}\right)$  converges in distribution to a  $\chi^2$ -distribution with  $n - 1$  degrees of freedom, where  $q^N$  denotes the empirical distribution ( $q_\omega^N = N_\omega/N$ ) and  $q^{\text{true}}$  denotes the underlying true distribution. Most  $\phi$ -divergences in Table 1 satisfy this differentiability condition. Ben-Tal et al. [2] then use this result to suggest the asymptotic value

$$\rho = \frac{\phi''(1)}{2N} \chi_{n-1, 1-\alpha}^2, \quad (10)$$

where  $\chi_{n-1, 1-\alpha}^2$  is the  $1 - \alpha$  percentile of a  $\chi_{n-1}^2$  distribution, which produces an approximate  $1 - \alpha$  confidence region on the true distribution. For corrections for small sample sizes and more details, we refer the readers to [16] and [2].

### 3.4 $\phi$ -Divergences Selected for this Study

We selected three  $\phi$ -divergences for the computations in this study, based on the classification of  $\phi$ -divergences presented in [14]: Modified  $\chi^2$ -distance, Kullback-Leibler divergence and Burg entropy.

Because the data in this study comes in the form of projections of possible future scenarios, the decision maker may be interested in knowing whether some scenarios should be given probability  $p_\omega = 0$  in the worst-case distribution. That is, some of the scenarios may be considered to be overly optimistic and the decision maker may want a formulation capable of making this distinction. We refer to determining  $p_\omega = 0$  as *suppressing* scenario  $\omega$ .

Both the Modified  $\chi^2$ -distance and the Kullback-Leibler divergence are capable of suppressing scenarios, but they do so in different ways. The problem (6) formulated with the Modified  $\chi^2$ -distance is capable of choosing whether to suppress any scenario individually, thus generating a wide variety of possible model output. In contrast, when the Kullback-Leibler divergence is used, the only possible results are: (a) no scenarios will be suppressed (i.e.,  $p_\omega > 0$  for every scenario  $\omega$ ), or (b) all but one scenario will be suppressed (i.e., some scenario will have  $p_\omega = 1$ , while all others will have probability 0).

Unlike the two  $\phi$ -divergences discussed above, the Burg entropy is not capable of suppressing scenarios and thus the result will always show  $p_\omega > 0$  for every scenario. The Burg entropy may be used if the decision maker wants to ensure that no scenarios are suppressed. See [14] for a derivation of these behaviors.

### 3.5 Value of Data

With a data-driven formulation such as (6), it is natural to ask how the optimal value and solution changes as more data is gathered. In particular, one might be concerned about being overly conservative in the problem formulation and thus missing the opportunity to find a better solution to the true distribution. For (6), this means that the initial model is likely to be more conservative

in an effort to be robust, while the new information could make the model less conservative because new information removes the current worst-case distribution from the set of distributions. Below, we present a simple method of determining if taking an additional sample will eliminate the old worst-case distribution and allow for better optimization; i.e., a lower-cost solution.

Love and Bayraksan [14, Theorem 1 and Corollary 1] provide sufficient conditions for an additional sample to rule out the worst-case distribution, which are summarized here for the  $\phi$ -divergences used in this study. An additional sample of scenario  $\hat{\omega}$  will result in a decrease in the worst-case expected cost of the  $\phi$ LP-2 if the following condition is satisfied for:

**Burg entropy:**  $\frac{p_{\hat{\omega}}}{q_{\hat{\omega}}} < \frac{N}{N+1}$ ,

**Kullback-Leibler:**  $\sum_{\omega} q_{\omega} \left(\frac{p_{\omega}}{q_{\omega}}\right)^{\frac{N}{N+1}} \log \left(\frac{p_{\omega}}{q_{\omega}}\right)^{\frac{N}{N+1}} + 1 > \left(\frac{p_{\hat{\omega}}}{q_{\hat{\omega}}}\right)^{\frac{N}{N+1}}$

**Modified  $\chi^2$ :**  $2 \sum_{\omega} p_{\omega} \frac{p_{\omega}}{q_{\omega}} > \left(\frac{p_{\hat{\omega}}}{q_{\hat{\omega}}}\right)^2 + \left(\frac{N+1}{N}\right)^2$ .

The conditions listed above provide a simple test for determining which scenarios would decrease the expected cost if they were given additional weighting in the problem. In our problem described in Section 4, scenarios can be categorized as part of two groups: low population and high population. Our analysis of the value of data in Section 5 shows that the low population scenarios satisfy the above conditions, as would be expected. Further details are available in Section 5.

## 4 Application to Tucson, AZ

In this section, we provide details on the generalized network model (1) describing water allocation in a developing region of Tucson, AZ shown in Figure 1.

The resulting model has a total of  $P = 41$  time periods, representing years 2010–2050. For each time period, the network has 62 nodes representing demand for potable and nonpotable (reclaimed) water, pumps, water treatment plants, and the available water supply from the Colorado River. Costs for all time periods are brought into present value by applying a 4% discount rate per year. The network in each time period has 102 arcs, representing the pipe network carrying the water between the nodes physically and connecting the network to the five reservoirs that connect the time stages in the model. We use  $P_1 = 5$  time periods for the first stage.

Uncertainty in the model comes from population growth, water availability, and climate-driven per-capita demand. The remainder of this section discusses our methods for modeling the problem uncertainty.

### 4.1 Network Details

The majority of Tucson’s water comes from the Colorado river, brought in by the Central Arizona Project (CAP) canal, which is shown in blue on the left side of Figure 1. This water may then be treated and sent to customers, or pumped underground to one of five aquifers to be saved for future use. One aquifer recharge facility, the Central Avra Valley Storage And Recovery Project (CAVSARP) is shown in Figure 1, along with two water treatment plants: the Tucson Water Reclamation Facility (TWRP) and the Haden-Udal Water Treatment Plant (HUWTP).

The southeastern portion of Tucson, outlined in purple in Figure 1, is being increasingly developed. This area is split into different elevation zones, shown by the green lines, the demands in each of which will be served by a dedicated reservoir and pump. That is, the elevation lines in Figure 1 also split the region into several pressure zones. Given the capacity of the existing treatment

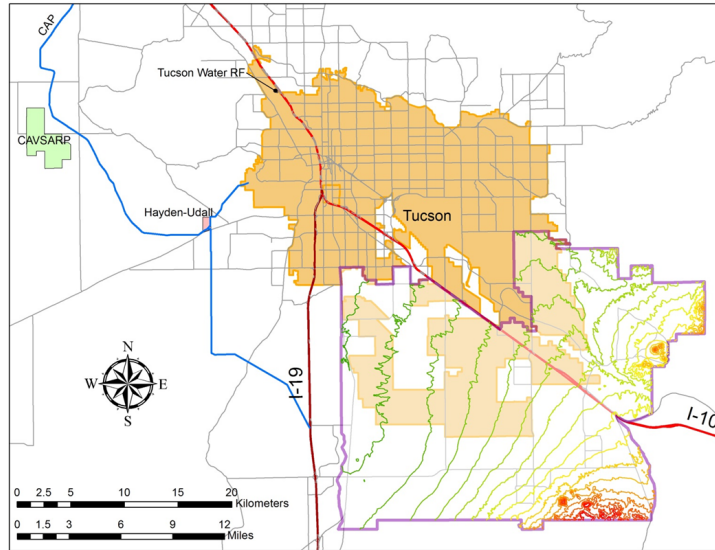


Figure 1: A map of Tucson, AZ, showing the RESIN area outlined in purple in the southeastern part. The green lines indicate elevation changes, denoting the difference between the pressure zones considered in this study.

plants and the energy cost of pumping the water through a series of pressure zones, the governing agencies in Tucson are interested in the possibility of building additional treatment facilities in this area, herein known as the RESIN (RESilient and Sustainable INfrastructures) area.

A more detailed view of Tucson’s water treatment system is shown in Figure 2. The blue nodes and arcs indicate the potable water treatment, storage and distribution system, which takes CAP water as an input and distributes it to the pressure zones shown in Figure 3. The purple nodes and arcs indicate non-potable water which can be used to satisfy some (especially agricultural and industrial) demands. The black nodes and arcs indicate wastewater that is returned from potable water use.

The pressure zones in the RESIN area, shown in Figure 3, have a periodic structure, with the possible exception of additional infrastructure as described below. Each pressure zone contains a potable and non-potable demand node. Note that potable water, being of higher quality, can be used to meet either type of demand. Each zone also contains a potable-water reservoir and pump, which is used to supply potable to the next pressure zone and a booster station that provides the conveyance for the non-potable water. Finally, each zone has a wastewater return, which is transported back to the centralized processing facilities via a gravity flow.

The quantity of water demand in each zone is determined as a multiple of the projected population in each year. Potable demand makes up 80% of the total demand and nonpotable demand is the remaining 20%.

A dummy node, not shown in the figures, capable of supplying water to the demand sites in the event of a water shortage is also included in the model. The cost of this extra supply is set at \$800/acre-foot.

## 4.2 Population & Water Supply Estimates

Estimates of the population in each pressure zone were compiled from two population studies in Pima County in Arizona: the Water & Wastewater Infrastructure, Supply & Planning Study [24]

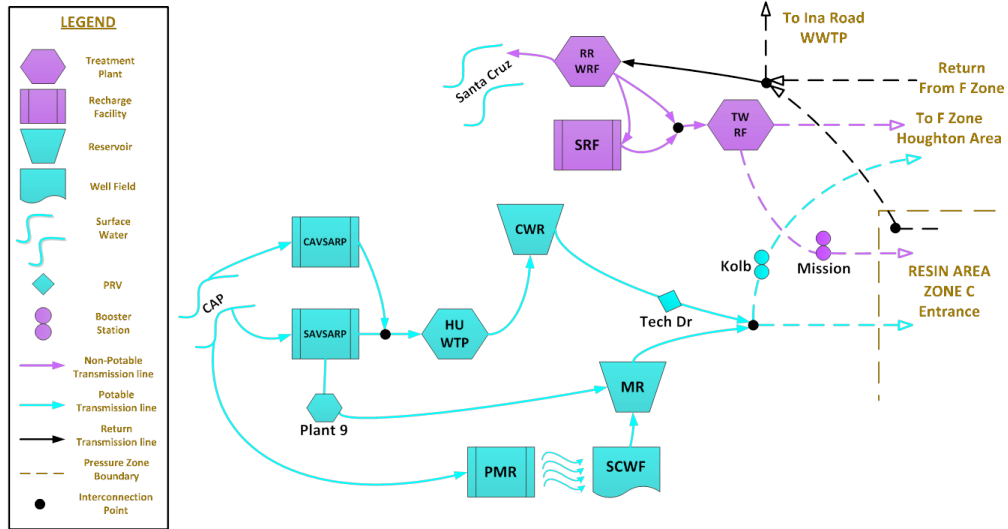


Figure 2: A schematic of the existing water treatment and recharge facilities in Tucson. Blue indicated potable water, purple indicates non-potable (reclaimed) water, and black represents waste water.

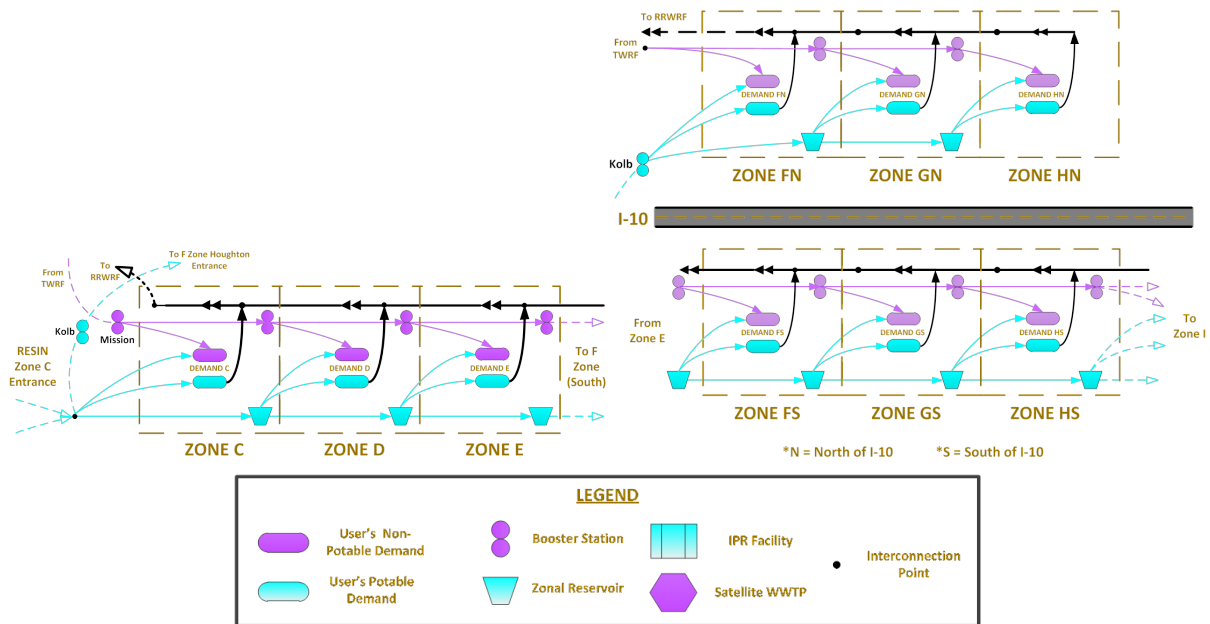


Figure 3: A schematic illustration of the pressure zones in the RESIN area. Each zones contains potable and non-potable demand nodes, a reservoir and booster station for transporting the potable and non-potable water, respectively, and waste water return pipes.

Estimate	population	
	2010	2050
Pima Association of Governments [17]	53,028	460,000
WISP [24]	53,028	740,000
City of Tucson [5]	750,000	1,300,000

Table 2: Population estimates for the RESIN area and for the Tucson Water service area.

	Estimate	Std. Error	<i>t</i> value	<i>p</i>
(Intercept)	2.026e+05	7.145e+03	28.3 6	$< 2 \cdot 10^{-16}$
Single Family	2.543e+00	4.171e-02	60.98	$< 2 \cdot 10^{-16}$

Table 3: Results of the linear regression of population on number of single family units in Tucson Water’s service counts

and the Traffic Analysis Zone (TAZ) [17]. Pima Association of Governments [17] provides a lower estimate of growth in the RESIN area, while [24] provides the higher estimate. The RESIN area represents a subset of the population served by Tucson Water, the population of which is estimated from [5]. Sample values can be seen in Table 2.

Tucson Water has a yearly CAP water allocation of 144,000 acre-feet, which may be reduced in the event of an extended drought reducing the flows of the Colorado river. The Arizona Department of Water Resources believes that a 10% reduction to 130,000 acre-feet is the maximum realistic reduction by 2050 [20]. The water allocation to the RESIN area is determined by the proportion of the Tucson Water service population in the RESIN area, e.g., under the [24] growth scenario with the full CAP allocation, the RESIN area is allocated about 81,970 acre-feet in 2050.

A reduction in Colorado River water allotment will occur if the elevation of Lake Mead falls below 1025 feet, according to the Colorado River Compact 2007 interim guidelines. For the nominal distribution, we estimate the probability of this event as 0.2, based on [22]. For this purpose, we will select the “number” of observations as  $N_\omega = 1.6$  for any scenario without an allotment restriction and  $N_\omega = 0.4$  for any scenario with an allotment restriction. These numbers are chosen so that any pair of scenarios that differ only by water allotment are given a total of 2 observations, while maintaining weighting desired.

### 4.3 Water Demand Regression

We next need to estimate the weather-dependent water demand in the RESIN area. To estimate demand separately in each elevation zone, we need a statistical model of the per-capita water demand to be used in conjunction with the zonal population estimates. For this task, data on monthly demand for water in Tucson from 1991–2011 was used, along with the recorded amount of rain and the average daily high temperature in each month.

A direct measure of population was only available on a yearly basis, but service counts were available for each month, from which population estimates were inferred in order to estimate the per-capita demand for water. We performed a linear regression on the total population against the monthly service count data to estimate the population for each month. The linear regression found that only the number of single family residences was a statistically significant predictor of population size. The results are shown in Table 3. The counts of single family residences accounted for the vast majority of the variation in the population data, demonstrated by an adjusted  $R^2 = 0.9946$ .

Given the monthly estimates on population, the water demand per capita in gallons per capita per day (GPCD) were calculated for every month in the 1991–2011 data. A linear regression of these GPCD values on climate-related variables like the mean daily high temperature and average precipitation rate for each month and the year of the observation, was then calculated. The year of the observation is included because the average GPCD water demand began dropping near the beginning of the 21st century, from over 170 GPCD in 1996 to under 140 GPCD in 2011. The linear fit, shown in Table 4 has an adjusted  $R^2 = 0.7266$ . We see that both temperature and precipitation are statistically significant and temperature seems to explain more of the variation in the data.

This fit, with the year included, projects water demand to be around 145–155 GPCD in 2010,

	Estimate	Std. Error	<i>t</i> value	<i>p</i>
(Intercept)	2721.2539	355.5267	7.654	$4.33 \cdot 10^{-13}$
Temperature	3.5830	0.1446	24.774	$< 2 \cdot 10^{-16}$
Precipitation	-2.6329	1.1337	-2.322	0.021
Year	-1.3296	0.1776	-7.486	$1.24 \cdot 10^{-12}$

Table 4: Results of the linear regression of water demand in GPCD on average daily high temperature, precipitation rate, and year

	Estimate	Std. Error	<i>t</i> value	<i>p</i>
(Intercept)	9254.0201	904.3062	10.233	$< 2 \cdot 10^{-16}$
Temperature	3.5913	0.1345	26.694	$< 2 \cdot 10^{-16}$
Precipitation	-2.2570	1.0510	-2.148	0.0327
Bounded Year	-4.5847	0.4509	-10.167	$< 2 \cdot 10^{-16}$

Table 5: Results of the linear regression of water demand in GPCD on average daily high temperature, precipitation rate, and and the bounded year.

which decreases over time to 100–110 GPCD by 2050. Tucson Water uses a typical estimate of 120–145 GPCD in their projections. The range specified by this fit is similar to this for many of the years considered. Because the water demand is linear in the year, we refer to this as the linear fit.

The linear fit projects a continual increase in water efficiency out to 2050. To combat this, a second model was developed to capture the decrease in water usage in the data, but not extrapolate this trend. For this, we choose to conduct the regression on a bounded value of the year, i.e., choose the value in a set  $\{Y_l, Y_l + 1, \dots, Y_u\}$  closest to the actual year. First, the upper bound  $Y_u = 2011$  was chosen so that no decrease in water demand would be projected beyond the scope of the data. The value  $Y_l = 2004$  was then selected to generate the best fit to the data. This piecewise fit is shown in Table 5 and has an adjusted  $R^2 = 0.7634$ .

The piecewise fit typically predicts demands of 135–145 GPCD throughout 2010–2050, increasing through the time period, which is in-line with Tucson Water’s predictions.

#### 4.4 Climate Models and Greenhouse Emissions Paths Used

A list of the climate models used in this paper is shown in Table 6. Bias-Corrected and Spatially Downscaled (BCSD) data from Coupled Model Intercomparison Project: Phase 5 (CMIP5) was obtained from [6]. Additional information on the downscaled data can be found in [18]. Each model included predictions of the variables `tasmax` and `pr`, the average daily high temperature ( $^{\circ}C$ ) and the average precipitation rate (mm/day) in each month, respectively.

Each model had output associated with a given path for future greenhouse gas emissions. Our analysis includes four paths: RCP2.6, RCP4.5, RCP6.0 and RCP8.5. The greenhouse emissions path RCP2.6 is an optimistic case where emissions are drastically reduced by mid-century. The paths RCP4.5 and RCP6.0 show stabilization of emissions before and after 2100, respectively. Finally, RCP8.5 is the case where emissions continue to grow quickly throughout the remainder of the century.

A comparison of the model output of the average daily high temperature for 2040–2050 for emissions path RCP8.5 is shown in Table 7. We can see that average daily high temperature is consistently 1–3 $^{\circ}C$  higher in every month, with smaller variations between the models. Table 8

Institution	Model
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-mk-3-6-0
Geophysical Fluid Dynamics Laboratory	GFDL-CM3 GFDL-ESM2M
Met Office Hadley Centre	HadGEM2-ES
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies and Japan Agency for Marine-Earth Science and Technology	MIROC5 MIROC-ESM

Table 6: A list of climate models used in this analysis

Month	Historic	GFDL-CM3	GFDL-ESM2M	CSIRO	HadGEM2-ES	MIROC-ESM-CHEM	MIROC5
Jan	19.1	21.1	19.5	21.2	19.6	21.8	20.6
Feb	20.7	23.3	22.5	21.8	22.5	23.7	22
Mar	24	26.1	23.6	24.9	24.5	25.8	25.2
Apr	28	31.7	30	30	31	31.9	31.1
May	33.5	36	33.9	34.2	35.3	35.8	35.9
Jun	38.2	41.5	39.7	40.1	40.5	40.1	39.9
Jul	38.1	41	39.6	39.8	40	40.2	39.7
Aug	37	40.3	38.4	39	39.6	39	38.8
Sep	35.5	38.6	36.5	36.6	37.3	37.4	37.6
Oct	30.1	33.5	32.6	32.1	34.1	31.8	32.6
Nov	23.7	26.4	24.3	25.5	26.2	25.4	26
Dec	18.6	20.6	19.4	21.1	21	21	21.1

Table 7: A comparison of projected ( $\geq 2040$ ) average daily high temperatures ( $^{\circ}C$ ) with the historical record (1991-2011) in the RESIN area. All values come from models using emissions path RCP8.5.

compares the mean precipitation per month (mm) with each RCP8.5 projection. We note that the projections typically predict more rain than the historic record, especially during the summer monsoon in July–September. CSIRO seems to have an especially large precipitation prediction during these months.

#### 4.5 Scenarios and Infrastructure Configurations

In total, each problem has scenarios generated from the following considerations, with the number of possibilities given in parentheses:

- Population projections (2),
- Water allotment scenarios (2),
- Per capita demand regression model (2),
- Climate models (6),
- Emissions paths (4),

which results to a total of 192 scenarios, for which we define  $N = 192$  total observations. In the nominal distribution, the weighting of each scenarios depends only on the water allotment as described in Section 4.2. With these weights, the nominal distribution is given by  $q_{\omega} = 0.008\bar{3}$

Month	Historic	GFDL-CM3	GFDL-ESM2M	CSIRO	HadGEM2-ES	MIROC-ESM-CHEM	MIROC5
Jan	22	19	31	13	37	12	25
Feb	24	23	41	23	23	13	26
Mar	17	21	22	22	17	20	19
Apr	8	2	12	9	3	4	5
May	5	3	7	4	3	4	3
Jun	4	4	6	8	7	8	6
Jul	52	75	58	85	57	71	58
Aug	58	55	74	66	54	66	55
Sep	33	32	35	50	32	45	35
Oct	14	47	26	42	14	53	26
Nov	14	27	20	10	20	15	9
Dec	23	20	25	31	30	25	25

Table 8: A comparison of projected ( $\geq 2040$ ) average precipitation per month (mm) with the historical record (1991-2011) in the RESIN area. All values come from models using emissions path RCP8.5.

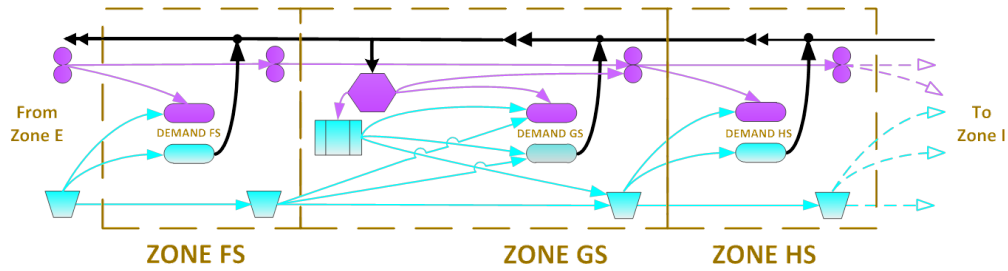


Figure 4: A satellite wastewater treatment plant is shown here, build in pressure zone GS.

for the 96 scenarios with no allotment reduction and  $q_\omega = 0.00208\bar{3}$  for the 96 scenarios with a reduction in water allotment.

In addition to the scenarios outlines, there are four options for additional infrastructure that may be constructed in the RESIN area. These possibilities are:

- **None:** No additional infrastructure is constructed.
- **WWTP:** A satellite wastewater treatment plant (WWTP) is constructed, capable of treating wastewater up to a nonpotable quality, for satisfying demands in its own zone and and higher zones.
- **IPR:** In addition to the WWTP, an indirect potable reuse (IPR) facility can be constructed, which further treats water from the WWTP up to potable quality.
- **RO:** The IPR facility described above also includes a reverse osmosis (RO) capability in this option.

Figure 4 gives an illustration of an additional WWTP and IPR constructed in zone GS.

## 5 Results and Discussion

The water allocation model was solved using the decomposition algorithm described in [14]. We solved the model for different  $\phi$ -divergences, using values of  $\rho$  equivalent to asymptotic confidence regions of 90%, 95% and 99%.

$\phi$ -divergence	Operating Cost (\$1MM)		
	90%	95%	99%
Modified $\chi^2$	295.45	296.28	297.84
Kullback-Leibler	297.61	298.56	300.36
Burg	297.88	298.79	300.50

Table 9: Total worst-case expected operating cost of the water system over the 2010-2050 time span without additional treatment infrastructure. Columns represent worst-case costs at the 90%, 95% and 99% confidence levels.

	RCP2.6	RCP4.5	RCP6.0	RCP8.5	(all)
CSIRO	0.0381	0.0416	0.0354	0.0404	0.1554
GFDL-CM3	0.0473	0.0522	0.0406	0.0428	0.1829
GFDL-ESM2M	0.0411	0.0406	0.0367	0.0427	0.1611
HadGEM2-ES	0.0407	0.041	0.0429	0.0436	0.1683
MIROC5	0.0353	0.0402	0.0344	0.0426	0.1524
MIROC-ESM-CHEM	0.0395	0.0432	0.0466	0.0505	0.1799
(all)	0.2421	0.2587	0.2366	0.2626	1

Table 10: Marginal worst-case probabilities of each climate model and emissions scenario as computed by the Kullback-Leibler Divergence.

## 5.1 No Additional Infrastructure

Table 9 gives a breakdown of the expected operating cost in the worst-case distribution identified by each  $\phi$ -divergence at each of the confidence levels considered. We can see that Modified  $\chi^2$  generates a somewhat lower cost and that Burg and Kullback-Leibler produce very similar results.

To compare the models and emissions scenarios, we computed the total probability assigned to each. Results for the Kullback-Leibler divergence can be seen in Table 10. We note that the highest probability is assigned to the highest emissions scenario, RCP8.5, but that the second highest probability is given to RCP4.5, a somewhat lower emissions scenario than RCP6.0. RCP6.0 in fact has the lowest anticipated costs over the 2010-2050 time period. We can also see that GFDL-CM3 results in the highest running cost overall for the water distribution system, followed by MIROC-ESM-CHEM. MIROC5 tends to generate lower-cost simulations and thus is given the lowest probability, followed by CSIRO.

Considering both climate model and emissions scenario together, we find that GFDL-CM3 with the RCP4.5 results in the highest operating cost, while MIROC5 with RCP6.0 has the lowest cost.

We note that the weightings for the climate models and emissions paths are similar with the other  $\phi$ -divergences, although the values differ slightly.

Table 11 lists the maximum total shortage over the 2010-2050 projection for each climate projection. The calculated shortages depend strongly on the physical constraints of the allocation model and little on either the  $\phi$ -divergence or level of robustness chosen.

The major difference between the  $\phi$ -divergences in these results are in the scenarios that are suppressed. Burg entropy is incapable of suppressing scenarios. Kullback-Leibler maintains an “all or nothing” approach to suppressing scenarios, but a confidence level of 99% is not high enough to induce the suppressing behavior. In contrast, the Modified  $\chi^2$ -distance does suppress scenarios for every confidence level tested and it consistently suppressed the low-population linear-fit scenarios. In the next section, we will see that exactly these scenarios tend to generate the smallest shortages

	RCP2.6	RCP4.5	RCP6.0	RCP8.5
CSIRO	568100	536340	570510	626290
GFDL-CM3	580500	572140	575160	614700
GFDL-ESM2M	552490	594280	552290	551120
HadGEM2-ES	535020	571920	538340	530150
MIROC5	569820	587560	513700	532110
MIROC-ESM-CHEM	557210	569890	575670	586880

Table 11: Worst-case shortage for each climate model and emissions scenario, with no additional infrastructure.

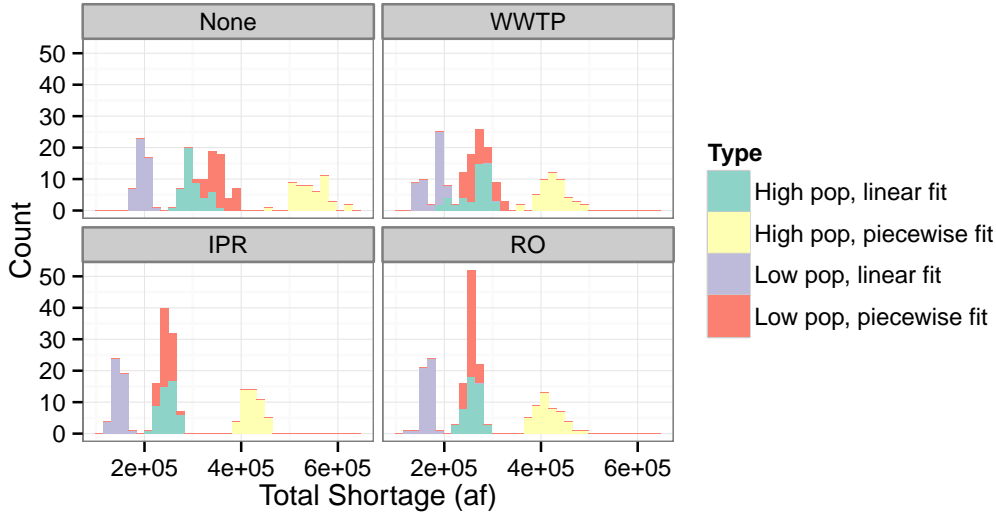


Figure 5: A histogram of total shortage amount over the 41-year study period for each infrastructure configuration

and are thus the least expensive.

## 5.2 Decentralized Treatment Options

Figure 5 shows the total shortage amount over the 41-year study period for each infrastructure configuration. We can see that a satellite wastewater treatment plant provides the most substantial reduction in shortage severity—especially in the high-population, linear fit scenarios—but that the indirect potable recharge facility decreases the worst-case shortages. The reverse osmosis (RO) option offers little benefit over the generic IPR. These shortage amounts are computed with the Kullback-Leibler  $\phi$ -divergence, but are dictated largely by the physical constraints of the system and do not change substantially with different  $\phi$  or  $\rho$ .

Table 12 calculates the worst-case expected operating cost of the water system from 2010-2050, divided by infrastructure type,  $\phi$ -divergence and confidence level. This cost includes \$800 loss per acre-foot of shortage. We can see that the satellite wastewater treatment plant consistently decreases the cost by about \$14 million and the IPR facility reduces the cost by an additional \$24 million over the 41-year time span. The reverse osmosis increases operating costs by about \$10 million and does not decrease shortage amounts as seen in Figure 5, making its addition of

Infrastructure	$\phi$ -divergence	Operating Cost (\$1MM)		
		90%	95%	99%
None	Modified $\chi^2$	295.45	296.28	297.84
	Kullback-Leibler	297.61	298.56	300.36
	Burg	297.88	298.79	300.50
WWTP	Modified $\chi^2$	281.59	282.41	283.82
	Kullback-Leibler	283.52	284.41	286.07
	Burg	283.82	284.67	286.27
IPR	Modified $\chi^2$	257.28	258.07	259.56
	Kullback-Leibler	259.22	260.12	261.82
	Burg	259.43	260.29	261.91
RO	Modified $\chi^2$	267.76	268.57	270.07
	Kullback-Leibler	269.80	270.72	272.46
	Burg	270.06	270.95	272.60

Table 12: Total worst-case expected operating cost of the water system over the 2010-2050 time span divided by type of infrastructure constructed. Columns represent worst-case costs at the 90%, 95% and 99% confidence levels.

questionable value to the system.

### 5.3 Value of Data

Finally, we applied the value of data condition discussed in Section 3.5 to the results of the water model. Recall that the value of data condition is a sufficient condition detailing when an additional observation of a given scenario will rule out the worst-case distribution and guarantee a decrease in the worst-case expected cost. In this problem, several themes were immediately clear from the data: (1) every low population scenario satisfied the condition, (2) every linear fit scenario satisfied the condition, and (3) no high-population, piecewise-fit scenarios satisfied the condition.

These results clearly match what was seen in the breakdown of the projected shortages in Figure 5. In both places, we see that the high population, piecewise fit projection generates substantially worse shortages than any of the other scenarios.

### 5.4 Discussion

Our analyses show that the cost savings from additional wastewater treatment plant and indirect potable recharge facilities would come to about \$38MM over the 41-year time period. The results indicate that the satellite WWTP and IPR, if constructed, would need a capacity of 10 million gallons per day. At that capacity, the estimated cost of these facilities is \$55MM and \$64MM, respectively. Clearly, the additional facilities will not pay for themselves over the planning period given that shortages are assumed to cost \$800/acre-foot. However, by increasing the shortage cost, we find that the satellite WWTP cost breaks even at \$2250/acre-foot of shortage. Comparing the no additional infrastructure setting with the WWTP + IPR setting shows a break-even cost of \$3100/acre-foot of shortage, but much of the savings come from the WWTP itself. When comparing the WWTP setting to the WWTP + IPR setting, we find that the IPR has a break-even cost of \$6500/acre-foot of shortage. The reverse osmosis option adds cost relative to the plain IPR, making it of dubious value. We believe that these break-even costs can provide a guide for deciding on the appropriate level of satellite infrastructure.

The  $\phi$ -divergence method demonstrated in this paper provides a way of integrating multiple sources of uncertainty into a single model and deriving results from each considered scenario. The principle advantage of this method is that it produces a measure of the relative importance (i.e., the worst-case distribution) of each scenario considered, which we have used to estimate the expected cost of operation the water allocation system the RESIN area. We have also shown that this worst-case distribution can be used to evaluate the impact of climate scenarios directly, as in the discussion of Table 10.

The two-stage model discussed here forces all uncertainty to be revealed simultaneously. This could be changed in either direction. A multistage model could be adapted in order to give a more refined view of how information comes to be revealed, which could allow for a richer scenario tree or for making multiple infrastructure decisions at multiple points in the model.

On the other hand, the  $\phi$ -divergence technique could be applied to a single-stage model, i.e., multiple simultaneous deterministic models, with only minor alterations. In the setting considered here, every scenario resulted in water shortages, forcing decisions to ultimately be made in a deterministic way. Switching to a single-stage model, in some cases, could decrease computation time without sacrificing model accuracy.

## 6 Summary and Future Work

In this paper, we have integrated projections for future climate, population growth and available water supply into a single integrated water allocation model. These scenarios were combined in a  $\phi$ LP-2 model proposed in [14], for several  $\phi$ -divergences. We applied this method to a developing area in Tucson, Arizona and determined projections for the severity of future water shortages, the expected operating cost of the system, and a worst-case distribution on the future scenarios.

Future extensions of this study have several possible paths. First, we would like to extend this method to the multistage setting, where uncertainty is revealed in multiple stages, rather than all occurring simultaneously. This will allow for an extended treatment of projections of population and water availability. Integrating the climate projections into the estimates of CAP water availability would help to represent the correlation between demand and supply better. Finally, increasing the size of the model to include more of Tucson and the state of Arizona would allow for more detailed future water planning.

## Acknowledgements

Support provided by a Water Sustainability Program Fellowship through the Technology and Research Initiative Fund at the University of Arizona. This work has also been partially supported by the National Science Foundation through grant CMMI-1151226.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP and we thank the climate modeling groups (listed in Table 6 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

## References

- [1] A. Ben-Tal, A. Ben-Israel, and M. Teboulle. Certainty equivalents and information measures: duality and extremal principles. *Journal of Mathematical Analysis and Applications*, 157(1): 211–236, 1991.
- [2] A. Ben-Tal, D. Den Hertog, A. De Waegenare, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59:341–357, 2013.
- [3] A. Bishop, R. Narayanan, S. Pratishtananda, S.L. Klemetson, and W. J. Grenney. Optimization of water allocation, wastewater treatment, and reuse considering non-linear costs, seasonal variations, and stochastic supplies. Technical report, Utah Water Research Library, 04 1975.
- [4] L.D. Brekke, M.D. Dettinger, E.P. Maurer, and M. Anderson. Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments. *Climatic Change*, 89(3-4):371–394, 2008.
- [5] City of Tucson. Update to water plan: 2000-2050. <http://www.tucsonaz.gov/water/docs/wp08-update.pdf>, 2008.
- [6] Downscaled CMIP3 and CMIP5. Downscaled CMIP3 and CMIP5 climate and hydrology projections. [http://gdo-dcp.ucllnl.org/downscaled\\_cmip\\_projections/](http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/), 2013. Last accessed: 2013-10-25.
- [7] C.R. Connell-Buck, J. Medellín-Azuara, J.R. Lund, and K. Madani. Adapting California’s water system to warm vs. dry climates. *Climatic Change*, 109(1):133–149, 2011.
- [8] A.J. Draper, M.W. Jenkins, K.W. Kirby, J.R. Lund, and R.E. Howitt. Economic-engineering optimization for California water management. *Journal of Water Resources Planning and Management*, 129:155–164, 2003.
- [9] W.A. Hall and D.T. Howell. Optimal allocation of stochastic water supply. *Journal of the Irrigation and Drainage Division*, 96(4):395–402, 1970.
- [10] J.J. Harou, J. Medellín-Azuara, T. Zhu, S.K. Tanaka, J.R. Lund, S. Stine, M.A. Olivares, and M.W. Jenkins. Economic consequences of optimized water management for a prolonged, severe drought in California. *Water Resources Research*, 46(5), 2010.
- [11] A. Higgins, A. Archer, and S. Hajkovicz. A stochastic non-linear programming model for a multi-period water resource allocation with multiple objectives. *Water Resources Management*, 22(10):1445–1460, 2008.
- [12] M.W. Jenkins, J.R. Lund, R.E. Howitt, A.J. Draper, S.M. Msangi, S.K. Tanaka, R.S. Ritzema, and G.F. Marques. Optimization of California’s water supply system: Results and insights. *Journal of Water Resources Planning and Management*, 130(4):271–280, 2004.
- [13] Y.P. Li, G.H. Huang, and X. Chen. Multistage scenario-based interval-stochastic programming for planning water resources allocation. *Stochastic Environmental Research and Risk Assessment*, 23(6):781–792, 2009.

- [14] D. Love and G. Bayraksan. Phi-divergence constrained ambiguous stochastic programs. Technical report, Program in Applied Mathematics, University of Arizona, 2013.
- [15] L.O. Mearns, R. Arritt, S. Biner, M.S. Bukovsky, S. McGinnis, S. Sain, D. Caya, J. Correia Jr, D. Flory, W. Gutowski, et al. The north American regional climate change assessment program: overview of phase I results. *Bulletin of the American Meteorological Society*, 93(9):1337–1362, 2012.
- [16] L. Pardo. *Statistical Inference Based On Divergence Measures*. Chapman and Hall/CRC, 2005.
- [17] Pima Association of Governments. Annual traffic count program. <http://www.pagnet.org/RegionalData/Maps/MapsandGISDownloads/tabid/902/Default.aspx>, 2012. Last accessed: November 2, 2013.
- [18] Reclamation. Downscaled CMIP3 and CMIP5 climate and hydrology projections: Release of downscaled cmip5 climate projections, comparison with preceding information, and summary of user needs. Technical report, prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado, 2013.
- [19] A. Ruszczyński. Stochastic programming models. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 483–554. Elsevier, 2003.
- [20] C.A. Scott, C.J. Bailey, R.P. Marra, G.J. Woods, K.J. Ormerod, and K. Lansey. Scenario planning to address critical uncertainties for robust and resilient water–wastewater infrastructures under conditions of water scarcity and rapid development. *Water*, 4(4):848–868, 2012.
- [21] L.G. Shao, X.S. Qin, and Y. Xu. A conditional value-at-risk based inexact water allocation model. *Water Resources Management*, 25(9):2125–2145, 2011.
- [22] U.S. Department of the Interior Bureau of Reclamation. Colorado river basin water supply and demand study: Technical report g — system reliability analysis and evaluation of options and strategies, 2012.
- [23] L. Wang, L. Fang, and K.W. Hipel. Basin-wide cooperative water resources allocation. *European Journal of Operational Research*, 190(3):798–817, 2008.
- [24] WISP. Location of growth, urban form, and cost of infrastructure. [http://www.tucsonpimawaterstudy.com/Reports/Phase2/White\\_paper\\_6.14.09.pdf](http://www.tucsonpimawaterstudy.com/Reports/Phase2/White_paper_6.14.09.pdf), 2009. Last accessed: November 2, 2013.
- [25] G.J. Woods, D. Kang, D.R. Quintanar, E.F. Curley, S.E. Davis, K.E. Lansey, and R.G. Arnold. Centralized versus decentralized wastewater reclamation in the houghton area of tucson, arizona. *Journal of Water Resources Planning and Management*, 139(3):313–324, 2012.