Team Control Number

49238

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

Problem Chosen

E

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2016
MCM/ICM
Summary Sheet
(Your team's summary should be included as the first page of your electronic submission.)

Type a summary of your results on this page. Do not include
the name of your school, advisor, or team members on this page.

Water scarcity is a major world problem; in this paper we examine the effect of hydraulic fracturing (hydrofracking) on the clean water supply of Pennsylvania in the US. The mathematical models in this paper examine both environmental and fiscal aspects of the clean water problem: pollution of ground water due to hydrofracking, revenue for the state gained from hydrofracking, and effective distribution of potable water. The first model, a parabolic PDE, details the propagation of pollution in the ground water of Pennsylvania. It models the movement of the contamination plumes emitted by hydrofracking violations. With the pollution propagating through Pennsylvania, certain ground water wells will be affected and possibly rendered unusable. The model's output is the certain ground water sources that are contaminated beyond the EPA-regulated safety levels. The second model, a linear optimization program, manages the financial cost of this polluted water to the state of Pennsylvania as well as whether or not cities will face water-shortages. Together, these two models give complementary short-term and long-term perspectives on problematic hydraulic fracturing in Pennsylvania.

1 Summary

As a case study of how environmental and social factors can obstruct the access to water resources, this report focuses on the phenomenon of ground water pollution from hydraulic-fracturing (”hydrofracking” or just ”fracking”) in Pennsylvania in the Northeastern United States. We begin in Section 2 by introducing background on fracking in Pennsylvania and isolating the goals for our models. In the following section, Section 3, we enumerate the assumptions for our models, both of which are introduced and compared in Section 4. Section 5 details the first model for long-term analysis, a Partial Differential Equations Model, and some of its model testing. Next, we outline a full Linear Optimization Program for short-term analysis and its sensitivity analysis in Section 6. Finally, we summarize our findings and conclusions in Section 7

2 Introduction/Restatement

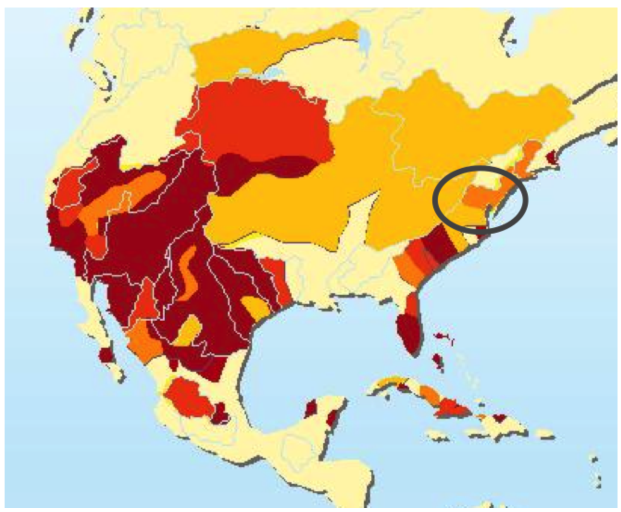

Figure 1: Map of water stress in the United States (Source: [12]).

More and more, humans are inhibiting their own access to natural resources, including fresh water. While many places around the world are subject to natural phenomenon such as drought, many industrialized nations are suffering water scarcity due to their own over-population or mismanagement of this precious resource. In the United States, for instance, much of the water stress on the west coast (see Figure 1) is caused by not only its dry climate, but also overuse of water and general overpopulation. Meanwhile, the east coast’s water resources are used up largely by hydrofracking which can take up to 10 million gallons of water per well. This industry, hydrofracking, is entirely voluntary and yet horribly detrimental to the environment. Despite its roast in the media spotlight and the great risk it poses to its citizens and environmental resources, the state of Pennsylvania continues to condone fracking as a serious industry. Because of this peculiar social dynamic, we decided to choose Pennsylvania as our region of interest.

Increased hydraulic fracturing has increased not only in Pennsylvania but also in the rest of the tri-state area, which includes New York and New Jersey. Hydrofracking is the method of high pressure injection of "fracking fluid" (primarily water that also contains sand and other proppants suspended with the aid of thickening agents) to create cracks in the layers of oil shale. See Figure 2. In doing so, natural gas flows more freely and can then be harvested. Fracking has been a hot topic in the news over past years, and we were interested in modeling both the environmental and financial aspects surrounding this industry. A second benefit to studying fracking is that it is easily isolated as a pollutant of ground water, i.e. we can model the environmental violations of fracking facilities without needing to take into account any other sources of pollution.

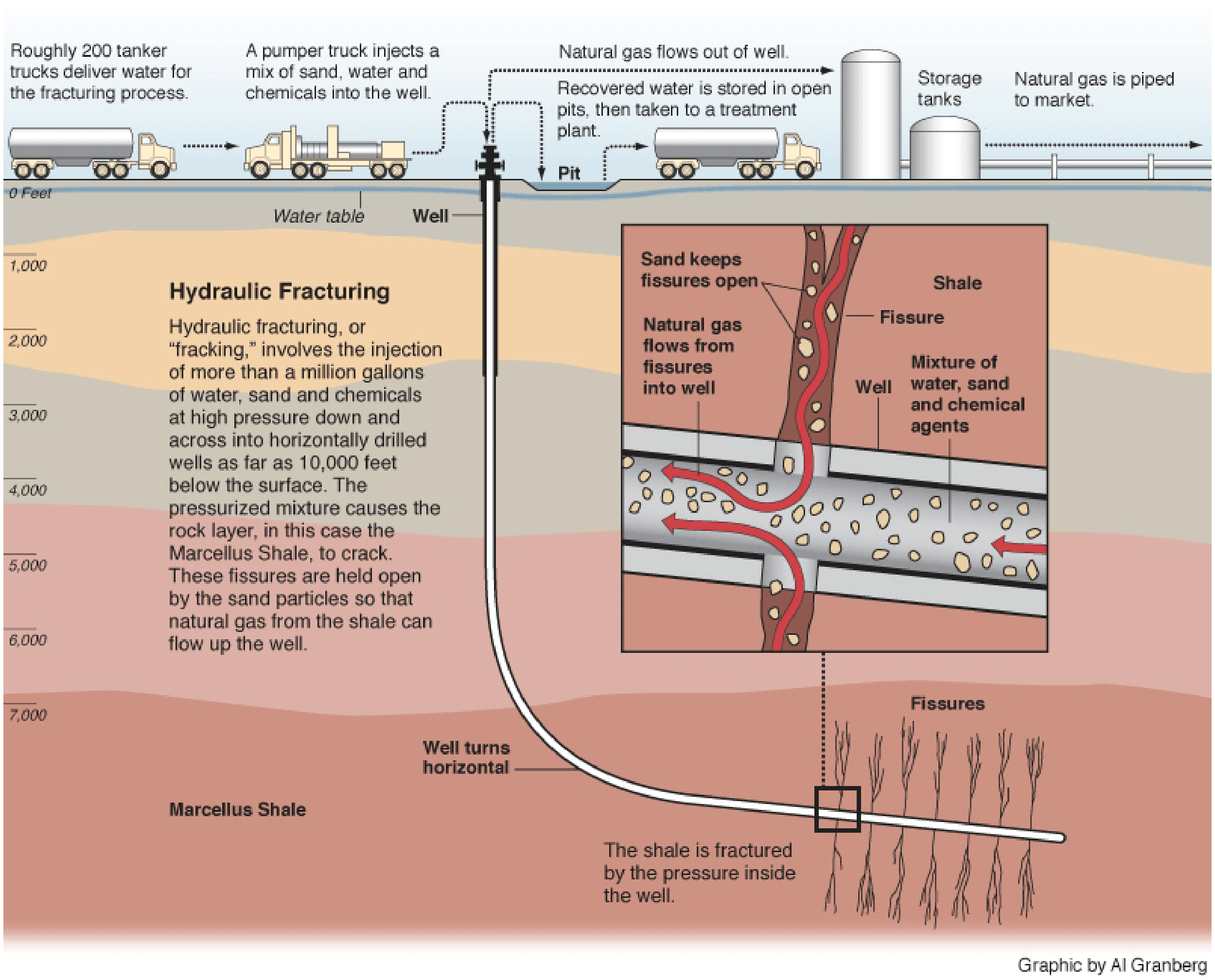

Figure 2: Graphic detailing hydraulic fracturing (Source: [7]).

Besides the loss of chemicals in the actual fracturing process, most waste from hydrofracking comes from the leaking of storage tanks in which the used fracking fluids are stored. These fluids then flow into the soil and make their way into the ground water aquifers which can lead to the contamination of water. Our first model will track the motion of these contamination plumes and attempt to quantify the long-term effects that this pollution has on available ground water.

The easiest way to stop this not only wasteful but also directly environmentally detrimental process would be to immediately stop hydraulic fracturing; however, the fracking industry is now deeply ingrained in the economy of the region. The industry that pollutes the water of thousands of citizens in the region is the same industry that feeds thousands of its workers' families. This is the difficulty of the entrenched economic situation that must be considered and studied, as well. The goal of our second model is to economize both the water resources as well as the financial resources in the region in order to determine if decreasing hydrofracking could benefit both aims.

3 Assumptions

- Both models require moderate discretization. To aid in organizing this process, we first discretize the region proportionally to its rectangular shape (Figure 3).

- Pennsylvania has both large rural areas as well as concentrated areas of major cities; this dichotomy complicates the problem of modeling the water pollution and economics of the entire state. Thus, we chose to not represent the rural families who drink from private wells in our models; instead, we focus on the activity of larger, more centralized populations. We also ignore surface water sources, such as Lake Erie, and instead focus on ground water and its pollution.

- Choosing fracking facilities indiscriminantly to mimic pollution would be unfair because it implies that all fracking pollutes equally. In fact, some fracking stations perform much worse than others, as shown by their long records of environmental violations [1]. Thus, we chose to rely on the number of violations while choosing fracking locations because these are the sites that are undoubtedly causing pollution.

- We assumed multiple geological properties of the underlying aquifers in Pennsylvania. After reviewing [5] and [6] , we determined the majority of Pennsylvania's aquifers were either sandstone or shale. We also determined that a transitivity of 147 $\frac{ft^2}{day}$ would be an appropriate average while an average vertical height of the aquifer of $50ft$ correlates to a storativity coefficient of $6 * 10^{-4}$.

- We use many simplifying assumptions to model the water purification process from its source to its final destination. For instance, we focus on the top water sources, purification facilities, destinations, and fracking facilities ("top" by different metrics, see Figure 4). This assumes that instead of hundreds of each kind of location that there would be a centralized few that are most important to the system.

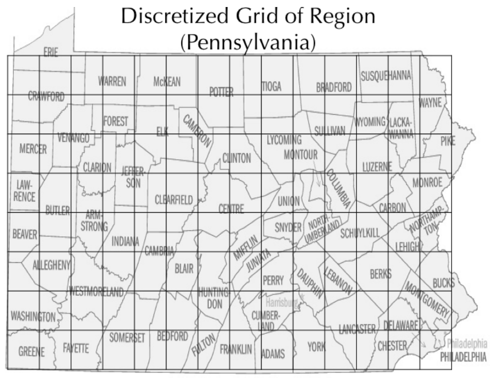

Figure 3: Discretization of the region. The matrix used in our program is 280×160, but above we only display 14×8. Therefore every square represents 20×20 units.

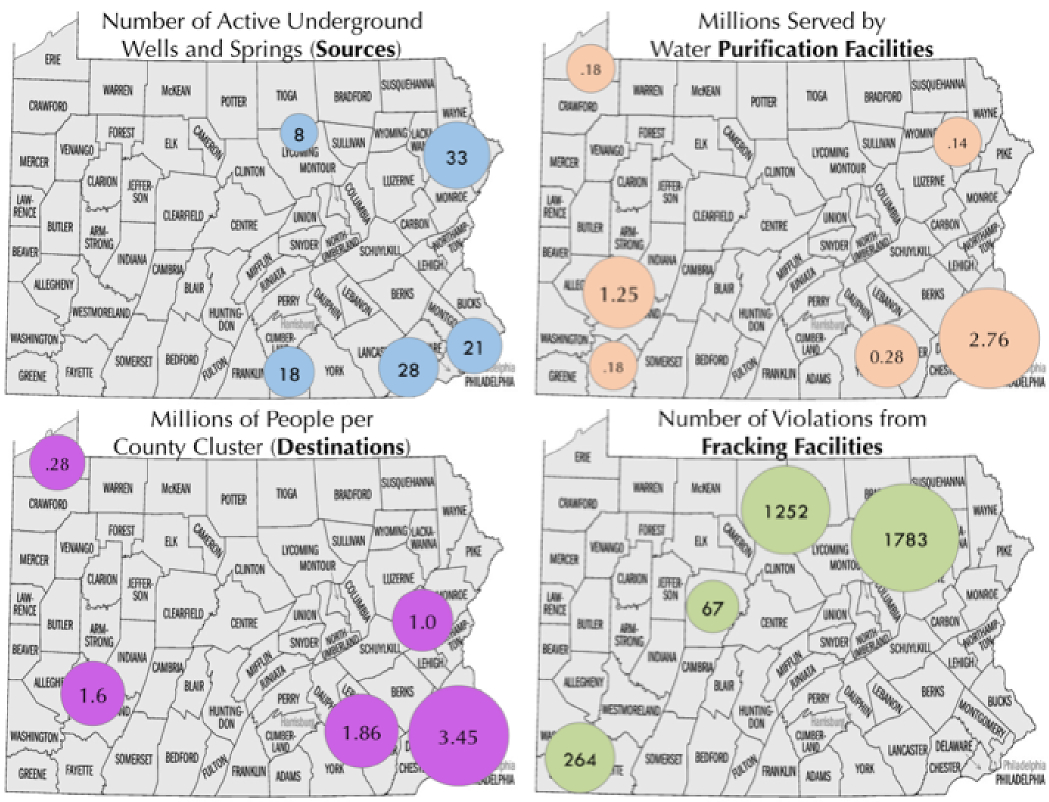

Figure 4: **Top Left:** The largest groups of underground water sources in Pennsylvania where water is pulled from the ground and transported to purification facilities [9]. **Top Right:** Areas with the purification facilities serving the largest populations [2]. **Bottom Left:** Areas of Pennsylvania with the the top fifteen most populous counties and therefore the highest demand for potable water [8]. **Bottom Right:** The top ten counties with the most fracking violations (such as waste spills) [1].

- We also assume that some quantities of water become unusable through either pollution or the purification process. In reality, there are dozens of water sources in a region, and at each one if the water is too contaminated, then the whole well of water is unusable. However, this model discretizes the regions such that one source is actually the accumulation of several water collection facilities. Therefore it is reasonable to believe that from this "pooled" source, some water is potable and some is not.

- Finally, the net worth of any single fracking facility is estimated conservatively as a *profit* from the state's perspective. We assume that the benefits of attracting workers, employing citizens, and accruing taxes and fines from violations (financially) overcome the costs of increased health care, decreased water supply, lawsuits, and other ramifications of fracking.

4 Model Design

We now introduce our two models whose complementary analyses give us a better understanding of the environmental and economic situation of fracking in Pennsylvania.

The first model employs partial differential equations as a tool for the long-term analysis of the environmental availability of groundwater. By quantifying the spread of pollutants from fracking locations, this model makes it possible to identify locations where the contamination exceeds the EPA standard for potable water. This flexible tool is also ideal for predicting pollution over many decades and a vast geographical region.

The second model utilizes optimization techniques from operations research in order to advise the organization of available resources. This linear program incorporates *both* economic and environmental factors into a classic cost-benefit analysis of the use of hydrofracking. In this way, our model optimizes the decisions over water and financial resources, providing an intervention plan for mitigating the environmental harm caused by fracking. Unlike the PDE model, the optimization model is best for analysis in a single moment or "slice" in time. Figure 5 summarizes these two models with respect to time.

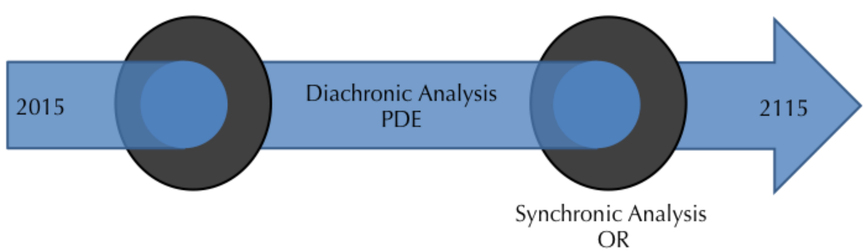

Figure 5: Our two models displayed graphically.

5 Partial Differential Equations Model

As a diachronic analysis, we developed a PDE in order to model the environmental effects on water availability in the region. This model incorporates environmental factors such as sedimentation and stratigraphy (The study of the type of sediment and the strata of sedimentary layers). We used a relatively standard hyperbolic PDE to model the flow of contamination from each fracking site [11]:

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = \frac{s}{T}\frac{\partial h}{\partial t} \tag{1}$$

where T represents transmisitivity (the rate of flow in an aquifer), S is the storativity (the volume of water held in a unit area of the aquifer) coefficient and h is our function: $h(x, y, t)$.

In order to solve this equation, we used the numerical methods technique of finite differences; we used the standard numerical approximation of the first derivative and second derivatives [4]:

$$\frac{\partial h}{\partial t} = \frac{h(x, y, t + \Delta t) - h(x, y, t)}{\Delta t}$$
$$\frac{\partial^2 h}{\partial t^2} = \frac{h(x, y, t + \Delta t) - 2h(x, y, t) + h(x, y, t - \Delta t)}{(\Delta t)^2}.$$

We used the notation $U^t_{x,y}$ to indicate a function evaluation at x, y, and t. To indicate the movement of a $\Delta x, \Delta y,$ or Δt, we added or subtracted 1 from the index (e.g. at time $t + \Delta t$ would be represented by $U^{t+1}_{x,y}$). Finally, we assumed that our step size in the x direction is the same as our step size in the y direction. Putting this all into our differential equation we attained

$$\frac{U^t_{x+1,y} - 2U^t_{x,y} + U^t_{x-1,y}}{(\Delta x)^2} + \frac{U^t_{x,y+1} - 2U^t_{x,y} + U^t_{x,y-1}}{(\Delta x)^2} = \frac{s}{T}\frac{U^{t+1}_{x,y} - U^t_{x,y}}{\Delta t}, \tag{2}$$

which reduced nicely to

$$U^{t+1}_{x,y} = \frac{T\Delta t}{(\Delta x)^2 s}[U^t_{x+1,y} - 4U^t_{x,y} + U^t_{x-1,y} + U^t_{x,y+1} + U^t_{x,y-1}] + U^t_{x,y}. \tag{3}$$

After completing a stability analysis on the discretized PDE, we determined a stability coefficient of

$$\Delta t = \frac{(\Delta x)^2 s}{4T}.$$

We made sure to keep this in mind while determining the step sizes for time in our model by ensuring our step sizes were small enough compared to the distances.

We discretized the state of Pennsylvania into a grid 160 miles by 280 miles. We then assigned a pollution level to certain sources we deemed consistent violators of hydro fracking regulations [1]. Finally, we allowed the PDE to evolve given the aforementioned initial conditions. The PDE model was simulated using python (see Appendix 8.1 for the full code). We were able to follow the pollution plumes with some consistency; however, due to the slow speed of our computers and the built-in error of the method for such a large discretization we lost some potency. Yet, we were still able to determine reasonably accurate pollution level estimates. Thus we deem our model to be reasonable given the time and computing constraints we were working under.

From our plots and table we can clearly see the plumes migrate considerably after a reasonable length of time (Figure 7). However, due to our assumptions about the placement of fracking wells, placement of pumping wells, and initial contamination levels, only one of the pumping wells was overly contaminated for the entire evolution (Figure 6). Still, we can deduce that given a few more years (10-20) contamination levels would have reached a critical level in pump 2. When we ran the simulations for a single contamination event, rather than continual replenishment , the plumes started to dissipate to under-critical numbers within 30 years.

Figure 6: The relative levels of pollution (2.4 being the critical contamination level). Note that only source 1 is above the critical level.

	15 Years	20 Years	40 Years	60 Years	80 Years	100 Years
Source 1	5.764	5.629	5.082	4.461	3.685	2.631
Source 2	0.046	0.101	0.396	0.689	0.915	1.034
Source 3	0.000	0.000	0.000	0.000	0.001	0.005
Source 4	0.000	0.000	0.000	0.000	0.000	0.004
Source 5	0.000	0.000	0.000	0.000	0.000	0.003

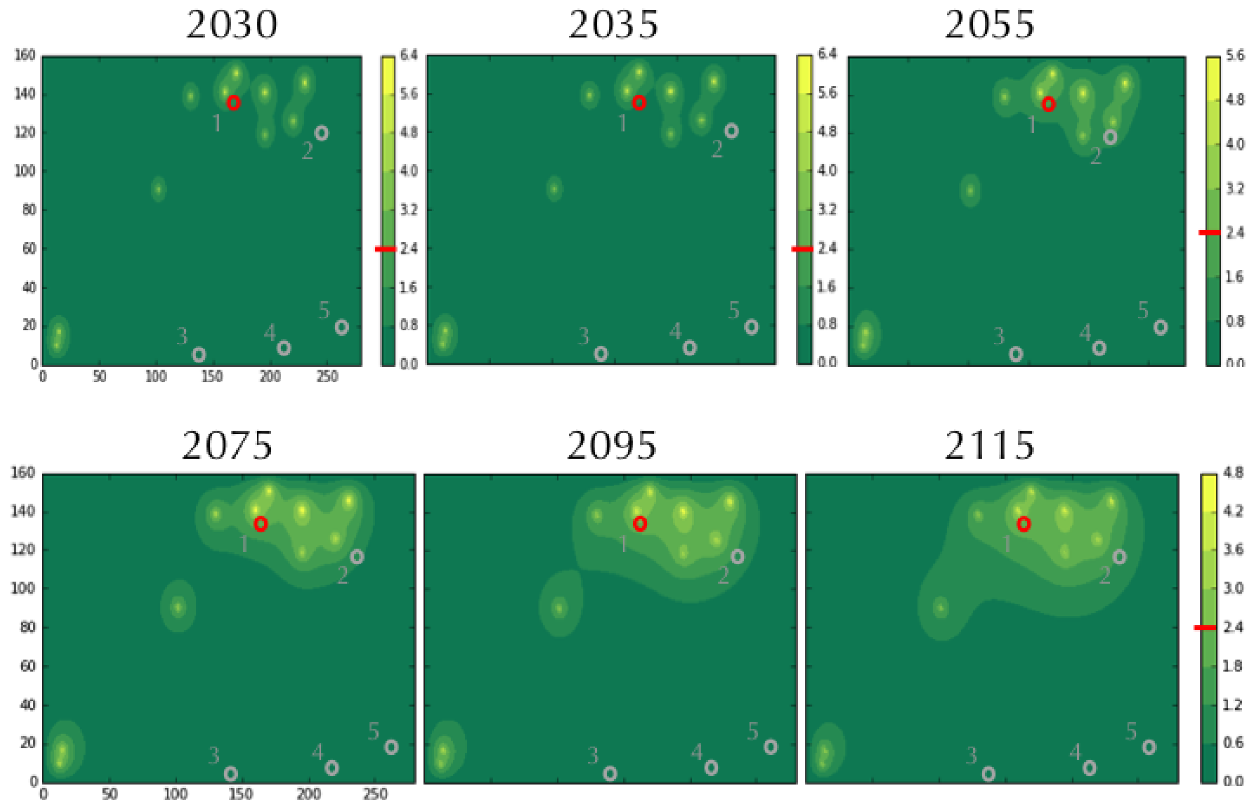

Figure 7: The contour maps show the pollution plumes evolving over time from the ten fracking facilities.(Task 2) The five underground water sources are located and labeled; the source that becomes too contaminated for safe use is colored red. Note that a critical contamination level is correlated to 2.4 on all scales. This model assumes that nearly the same amount of pollution is being dumped from the fracking location with a general decreasing trend. The data was logarithmically scaled in order for the lower levels of contamination to be more visible. The color-scale is the same for the three lower contour maps (from year 2075 to 2105).

6 Linear Program Optimization Model

This model aims at optimizing the economic factors surrounding water availability after collection from its source. Through linear programming, we can take into consideration the mechanisms of water transportation and distribution, associated costs, as well as the contradictory forces of profit from fracking and its detrimental effect on potable water. See Figure 8 for the full model of the process. In the following sections, we detail the construction of our model (see Appendix 8.2 for the full model).

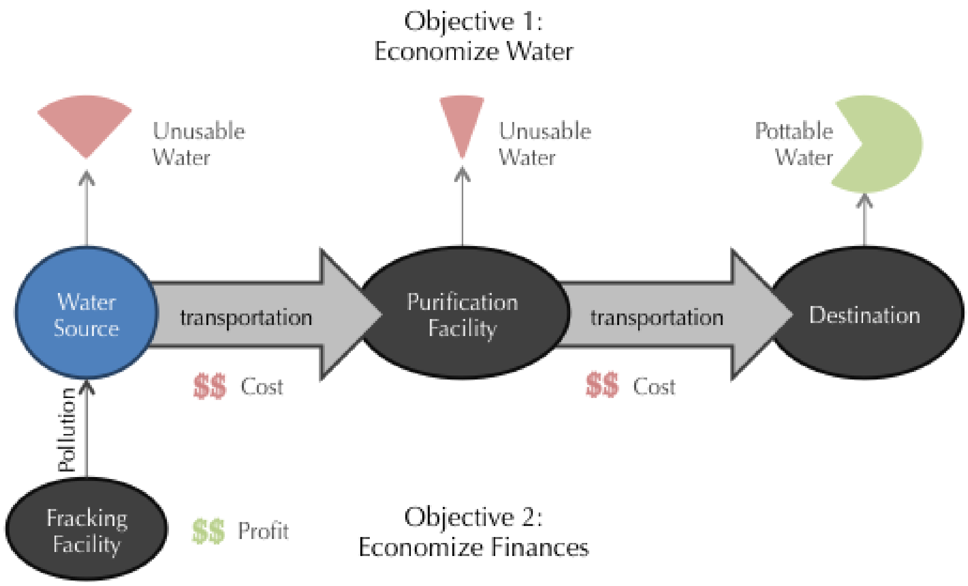

Figure 8: This linear program naturally models the practice of water purification from its source to one's tap. At each stage, some water becomes unusable, so our first objective is to maximize the amount of potable water (per capita). However, we cannot ignore the financial costs of transportation and the profits from fracking that overlay this system, therefore our second objective is to maximize net profit.

6.1 LP: Parameters and Decision Variables

Below, we present the parameters used to represent the 5 sources for water, 6 purification facilities, 5 destinations, and 10 fracking stations. It will be crucial for experts to carefully research these parameters in order to guarantee accurate and reliable results from the optimization model. The table shown in Figure 9 summarizes these parameters.

- Note that from the PDE model's output, we have that each source i has at least V_i tons of unusable water and at most W_i tons of usable water.

- At each purification station j, as an inevitable result of the purification process, some proportion θ_j of the water becomes unusable. These individual efficiency quotas will effectively pit the purification facilities "against each other" in the model so that they compete to win more business, which is appropriate for our economic model.

- Each destination k has a minimum volume of water required to satisfy the community's demands; also, the destination's population δ_k will be used to quantify water-per-capita distribution as an objective function in Section 6.4.

- It is crucial for our model to fairly represent both the environmental and economic impacts of hydrofracking in order to make an informed decision moving into the future.

Therefore, we model each fracking station l, while active, as inducing both a profit π_l to the state as well as proportionally reducing the amount of potable water at source i by ϕ_{il}. This proportional reduction of potable water is intentionally different than the input parameter V_i: while V_i captures the idea of *past pollution* from fracking behavior (which will only dissipate but never go away), the proportional loss of usable water represents *ongoing pollution* from fracking facilities that remain active. By introducing such a parameter, we will incentivize the program to discontinue fracking in order to preserve water resources. However, as discussed in Section 3, not every fracking station has the same environmental impact; some have many more violations than others. Thus, it is important to consider each facility independently according to its individual impact on the system.

- Transporting water at each stage of the process comes with a cost, which may be accumulated from real-world factors such as maintenance of the pipes, valve replacement, facility costs, distance traveled, etc. In order to represent these costs, we let a_{ij} be the cost to transport one unit (one thousand gallons) of water from source i to purification facility j. Similarly, we define b_{jk} as the cost for transporting one unit of water from purification facility j to destination k. This cost will act as a linear weight for $\boldsymbol{\beta}_{jk}$.

Figure 9: Summary of parameters.

5 Sources		
V_i	i=1..5	Minimum unpottable water from source i (in thousands of gallons)
W_i	i=1..5	Maximum pottable water from source i (in thousands of gallons)
6 Purification Facilities		
θ_j	j=1..6	Percentage loss of usable water due to purification station j being active
5 Destinations		
δ_k	k=1..5	Population of destination k (in millions)
μ_k	k=1..5	Minimum water supply for destination k (in thousands of gallons)
10 Fracking Facilites		
π_l	l=1..10	Profit incurred by activating fracking facility l (in thousands of dollars)
$\varphi_{i,l}$	i=1..5, l=1..10	Percentage loss of usable water from source i due to fracking station l being active
Transportation Costs		
$a_{i,j}$	i=1..5, j=1..6	Cost of transporting one thousand gallons of water from source i to purification facility j (in thousands of dollars)
$b_{j,k}$	j=1..6, k=1..5	Cost of transporting one thousand gallons of water from purification facility j to destination k (in thousands of dollars)

In addition to these (input) unchanging parameters, we define the main decision variables of our program. The main decisions relate to transportation of water, the activity of facilities, and the volumes of (un)usable water.

First, we define $\boldsymbol{\alpha}_{ij}$ as the amount of water transported from source i to purification facility j. Secondly, $\boldsymbol{\beta}_{jk}$ is defined as the amount of water transported from purification facility j to destination k.

$$\boldsymbol{\alpha}_{i,j} = \text{Amount of water transported from source } i \text{ to purification facility } j \quad \forall i, j \ \text{(continuous)}. \tag{1}$$

$$\boldsymbol{\beta}_{j,k} = \text{Amount of water transported from purification facility } j \text{ to destination } k \quad \forall i, j \ \text{(continuous)}. \tag{2}$$

Next, for each facility we define a binary decision variable to represent it as either active or inactive. Each purification facility j's activity is modeled by p_j, and each fracking facility l's activity is represented by f_l. The linear program then has control over inactivating any facility in the case that it would improve the system's performance at large, e.g. if either one purification facility were horribly inefficient ($\theta \approx 1$) or one of the fracking facilities polluted an obscene amount ($\phi \approx 1$).

$$\boldsymbol{p}_j = \begin{cases} 1, \text{ if purification facility } j \text{ is active} \\ 0, \text{ if purification facility } j \text{ is inactive} \end{cases} \quad \forall j \ \text{(binary)}. \tag{3}$$

$$\boldsymbol{f}_l = \begin{cases} 1, \text{ if fracking facility } l \text{ is active} \\ 0, \text{ if fracking facility } l \text{ is inactive} \end{cases} \quad \forall l \ \text{(binary)}. \tag{4}$$

Finally, we define decision variables to aid in tracking the volumes of water that are *unusable* at particular stages of the process. At the first stage the amount of unusable water at source i is represented by v_i^1, and at the second stage the amount of unusable water at purification facility j is represented by v_j^2.[1] In Section 6.3 we show how to construct constraints with these decision variables to enforce a reliable "flow" between water source to destination.

$$\boldsymbol{v}_i^1 = \text{Amount of unusable water at source } i \quad \forall i \ \text{(continuous)}. \tag{5}$$

$$\boldsymbol{v}_j^2 = \text{Amount of unusable water at purification facility } j \quad \forall j \ \text{(continuous)}. \tag{6}$$

[1] We assume that no water becomes unusable while transported to or stored at its final destination.

6.2 LP: Simple Constraints

We will now enumerate all the constraints that are necessary to appropriately fit this linear program to the problem. We begin with the simple constraints, defining discreteness, nonnegativity, etc.

Both of the transportation decision variables, $\boldsymbol{\alpha}_{ij}$ and $\boldsymbol{\beta}_{jk}$, are nonnegative, continuous variables. Functional upper bounds are described in Section 6.3.

$$\boldsymbol{\alpha}_{ij} \geq 0 \qquad \forall i, j \tag{7}$$
$$\boldsymbol{\beta}_{jk} \geq 0 \qquad \forall j, k \tag{8}$$

Both of the facility activity decision variables, $\boldsymbol{p}_j$ for purification facilities and $\boldsymbol{f}_l$ for fracking facilities, are binary.

$$\boldsymbol{p}_j \in \{0, 1\} \qquad \forall j \tag{9}$$
$$\boldsymbol{f}_l \in \{0, 1\} \qquad \forall l \tag{10}$$

Similar to the transportation variables, the variables representing unusable water at the sources, $\boldsymbol{v}_i^1$, and at the purification facilities, $\boldsymbol{v}_j^2$, are nonnegative. Functional upper limits are defined in the following section.

$$\boldsymbol{v}_i^1 \geq 0 \qquad \forall i \tag{11}$$
$$\boldsymbol{v}_j^2 \geq 0 \qquad \forall j \tag{12}$$

6.3 LP: Flow constraints

We will now enumerate the flow constraints which ensure that all of the water resources are transported from the sources to some destination as either unusable or potable.

We begin by identifying the amount of water at each source i that it is unusable, $\boldsymbol{v}_i^1$. Note that from the PDE model, the parameter V_i is initially given to quantify the amount of polluted water; we *add to* this amount any water that is polluted from *continued use* of fracking stations. Thus, any active fracking station l pollutes a proportion ϕ_{il} of the water at source i, making it unusable (which we want to incorporate into $\boldsymbol{v}_i^1$).

$$V_i + \sum_l \phi_{il} \boldsymbol{f}_l = \boldsymbol{v}_i^1 \qquad \forall i \tag{13}$$

Then we want to require for the first stage of transportation that the sum of all water volumes leaving a source i, i.e. $\sum_j \boldsymbol{\alpha}_{ij}$, is exactly equal to the amount of potable water at that source, which is now the amount of potable water originally at source i, W_i minus the polluted water, v_i. With a strict equality, we then have that no potable water is wasted.

$$\sum_{j=1} \boldsymbol{\alpha}_{i,j} = W_i - \boldsymbol{v}_i^1 \qquad \forall i \tag{14}$$

At the second stage, we derive the amount of water that becomes unusable at purification station j by its inefficiency proportion, θ_j. This variable is defined as $\boldsymbol{v}_j^2$:

$$\boldsymbol{v}_j^2 = \theta_j \sum_{i=1..} \boldsymbol{\alpha}_{i,j} \qquad \forall j \tag{15}$$

We then want that the sum of water volumes entering purification plant j minus the water that becomes unusable through the purification process, $\boldsymbol{v}_j^2$, to equal the sum of water volumes leaving the plant.

$$\sum_{i} \boldsymbol{\alpha}_{i,j} - \boldsymbol{v}_j^2 = \sum_{k} \boldsymbol{\beta}_{j,k} \qquad \forall j \tag{16}$$

At the final stage, we require that the final water volumes reaching each destination sum up to at least its minimal requirement, μ_k. The final water volume at station k is the summed inflow from all purification stations.

$$\sum_{j=1} \boldsymbol{\beta}_{j,k} \geq \mu_k \qquad \forall k \tag{17}$$

Remember that $\boldsymbol{p}_j$ is a binary variable that regulates whether a purification facility is active (1) or inactive (0). If inactive, then no water should be transported to or from this facility. Thus we define the following constraint to *suppress* all water transportation through facility j if the variable is set to zero. The big M represents a constant large enough to allow for any feasible amount of water to pass through the facility while it is active (an M equal to twice the amount of total water in the system should suffice).

$$\sum_{i} \boldsymbol{\alpha}_{\boldsymbol{ij}} + \sum_{k} \boldsymbol{\beta}_{jk} \leq M\boldsymbol{p}_j \qquad \forall j \tag{18}$$

6.4 LP: Objective Functions

We have two main objectives for the mechanics of the water purification system: to economize both the water availability and the financial net profit. We begin with the first objective of efficiently managing the water resources from source to destination by considering both the potable water that reaches each destination and the amount of unusable water that is essentially *wasted*.

Remember that one of our constraints is that each destination receives its minimal amount of water supply from all incoming transportation (Constraint 17), so we are not concerned with considering a shortage in our objective. Remember that the amount of potable water reaching each destination is represented as the sum of incoming transports from purification stations, i.e. $\sum_j \boldsymbol{\beta}_{jk}$. In order to calculate the surplus, subtract from this value the minimum requirement for water required (μ_k). The surpluses should not, however, be considered as an unweighted sum because a city with a small population and a large surplus is a misuse of resources when an overpopulated city barely has its minimal water supply. Therefore, we want to represent the water surplus *per capita* by weighting the surplus by the inverse of the destination's population, π_k.

$$\text{Water Surplus per Capita} = \sum_k \frac{1}{\delta_k}\left[\left(\sum_j \boldsymbol{\beta}_{jk}\right) - \mu_k\right]$$

Future research could explore the effectiveness of more sophisticated weights that employ a square (δ_k^2), logarithm ($ln(\delta_k)$), etc. For now, this degree of specificity is beyond the scope of our study.

Next, we include the sum of unusable water from the first two stages in the process, i.e. at the source, $\boldsymbol{v}_i^1$, and the purification facility, $\boldsymbol{v}_j^2$. Each gallon of unusable water requires its own disposal, which is costly, and ultimately represents lost opportunity for a marketable product.

$$\text{Total Unusable Water} = \left(\sum_i \boldsymbol{v}_i^1\right) + \left(\sum_j \boldsymbol{v}_j^2\right).$$

Therefore, we define our first objective function in order to economize the water supply, $E(W^*)$ as the water surplus per capita minus the total unusable water:

$$E(W^*) = \sum_k \frac{1}{\delta_k}\left[\left(\sum_j \boldsymbol{\beta}_{jk}\right) - \mu_k\right] - \left(\sum_i \boldsymbol{v}_i^1\right) - \left(\sum_j \boldsymbol{v}_j^2\right). \tag{19}$$

Now we turn to the financial situation. Remember that each active fracking facility returns a profit, π_l. Therefore, we sum across all facilities (any inactive facilities will have $\boldsymbol{f}_l = 0$, so their profits will not be considered).

$$\text{Profit from Fracking Facilities} = \sum_l \pi_l \boldsymbol{f}_l.$$

We also consider the costs of all stages of transportation. The decision of how much to transport along each possible route is captured by $\boldsymbol{\alpha}_{ij}$ and $\boldsymbol{\beta}_{jk}$, and there costs are represented (per unit) by a_{ij} and b_{jk}, respectively.

$$\text{Cost of Transportation} = \left(\sum_i \sum_j \boldsymbol{\alpha}_{ij} a_{ij}\right) + \left(\sum_j \sum_k \boldsymbol{\beta}_{jk} b_{jk}\right).$$

Thus, we define our second objective function in order to economize the financial net profit, $E(F^*)$ as the profit from fracking facilities minus the cost of transportation:

$$E(F^*) = \sum_l \pi_l \boldsymbol{f}_l - \left(\sum_i \sum_j \boldsymbol{\alpha}_{ij} a_{ij} \right) - \left(\sum_j \sum_k \boldsymbol{\beta}_{jk} b_{jk} \right). \tag{20}$$

We now have two operational objective functions that well capture our ideals for economizing water and financial resources. Note that either or both of these functions can be used independently from the rest of the program; for example, any interested person could use these *summary functions* as a way to describe how a region is performing in its water management before deciding to use the optimization program (Task 1).

What is the benefit of bothering to use the objective functions along with the costly optimization of the program? Once optimized, the final decision variables should be considered as advisory tools to inform a region's leaders what decisions should be made as relating to the use of fracking facilities, purification facilities, and transportation routes. In other words, the program's results offer an *intervention plan* for the at-risk water supply (Task 4).

When deciding to use the two objective functions for optimization, one could either combine both functions into a single objective function or define this as a *multi-objective* program. Multiple objectives would establish clear prioritization, which often well represents the political negotiation of such matters. A single (combined) objective may be considered more intuitive as a way to share the perspectives, but it also lacks the guarantee of optimality from the multi-objective approach [10]. We describe the strengths and weaknesses of each approach below.

- Single Objective:

  $$\text{maximize } c_1 E(W^*) + c_2 E(F^*)$$

  This model would allow the user to experiment with weights, $c_1, c_2 > 0$, in order to represent their own *non-exclusive* preferences–e.g. 40% financial and 60% environmental–which can be intuitive. The resulting intervention plan would then be determined from a shared perspective of both concerns. On the other hand, the c weights would ultimately be arbitrary and difficult to substantiate under critical cross-examination. However, as a single-run algorithm, it would be the most computationally effective approach.

- Multi-objective, prioritizing Environmental Resources:

  $$(1) \text{ maximize } E(W^*)$$
  $$(2) \text{ maximize } E(F^*)$$

  This model would not require arbitrary weights, and it would focus first on optimizing the use of water resources. Considering that each fracking facility has a detrimental effect on nearby water sources, it is inevitable that this program would *deactivate all of the fracking facilities.* On the other hand, it would activate all purification facilities in order to maximize the clean water moved through the system. This would challenge the second round of the algorithm to efficiently manage the transportation of the maximum water supply. This is of course most ideal for the environmentally-conscious citizens, but the costs may not be as convincing for state representatives and certainly not for business executives.

- Multi-objective, prioritizing Financial Resources :

$$(1)\text{ maximize } E(F^*)$$

$$(2)\text{ maximize } E(W^*)$$

  This model would also not require arbitrary weights, but it would focus first on optimizing the financial profit from the system. Since each fracking facility accrues profit, the program would most likely activate all of the fracking facilities. On the other hand, the transportation of water accumulates cost, so the program may deactivate a purification facilities in the case that a more cost-efficient route of transportation were available. It would also likely only meet the minimal water supply level at each destination, possibly eliminating any surpluses. While definitely in the best interest of the fracking industries, this approach would clearly devalue the water supply and the citizens' health. In addition, the results from our PDE model show that this would be an unsustainable option for moving into the future.

What we have discussed is how *both the researchers and the audience* should be considered while determining how to apply these objective functions. Most audiences would agree to either the single objective model or the multi-objective model prioritizing the environmental resources. Ideally, these two models would both optimize by decreasing the use of at least some fracking stations. Evidence like this might be most convincing in pushing the case for mitigating the activity of fracking facilities and thereby their pollution.

6.5 Feasibility

In this section, we consider the feasibility of our linear model. According to its dimensionality, this is a very moderate program. For example, if we analyzed using 5 water sources, 6 purification facilities, 5 destinations, and 10 fracking facilities, the feasible region would exist in 87th-dimensional space. Of course, the real problem faced by a region as large as a state would have incredibly more water sources, facilities, and destinations to consider. In general, if there are S water sources, P purification facilities, D destinations, and F fracking facilities, then the dimensionality is $P(S + D + 2) + F + S$ (before dimension reduction performed by many common linear optimization solvers).

Even a large problem in high dimensional space should be manageable by most optimization solvers since this program is simply linear. Every constraint and objective function is defined with scalar multiplications and summations. The only discrete variables are the binary toggles for activating purification facilities, $\boldsymbol{p}_j$, and fracking facilities, $\boldsymbol{f}_l$; however, experimentation with linear relaxation might find that even treating these as continuous variables from 0 to 1 might result in an optimal solution that would be functionally binary. In this ideal case, given enough time, even the most basic Simplex method should be able to trace the continuous, feasible region.

6.6 Sensitivity Analysis

Sensitivity analysis of linear programs involve asking what would/could happen to the optimal solutions and objective values if parameters in the program are incremented upwards or downwards. This gives us an idea of how the parameters play a role in the optimization and also highlights the importance of carefully researching those that are most influential. Note that once a parameter is adjusted, there are rarely any *guaranteed* increases or decreases in optimality, but careful sensitivity analysis can suggest possible improvements or declines [10].

First, we consider the parameters in the objective equations for economizing water resources, $E(W^*)$. Consider the two parameters in Equation 19: δ_k for the population of destination k and μ_k for the minimum water supply required by destination k. The population has a *loose, inverse* effect on the objective function, i.e. an increase in the population could return the *same or lower* objective value. Similarly, the minimum required water supply is loosely negatively correlated with the objective value, so the more water a city requires, we expect an *equal or decreased* optimal value. These make sense considering that a quickly growing city may stress its current water supply as a sign of poorly managed water resources. Note that these changes in parameters do not affect the optimal solution, itself, only the metric used to summarize its optimality.

We can make similar conclusions about the parameters in Equation 20 for the economizing of financial profit ($E(F^*)$). The first parameter, π_l, represents the profit from each fracking facility. An increase in this parameter may either maintain or improve the optimal value. On the other hand, the two parameters for cost of transportation, a_{ij} and b_{jk}, risk decreasing the optimal value if either is increased.

What would happen if one of the constraints adjusted? *Relaxing* a constraint effectively extends the feasible region, allowing for the program to find a new optimum which could be an improvement. On the other hand, *tightening* a constraint does the opposite: by constricting the feasible region, it risks making the previous optimum infeasible and making a suboptimal point the new optimum [10]. There is only one of the flow constraints from Section 6.3 that we consider in this analysis.[2]

Constraint 17 has just one parameter, μ_k, as in the minimum required water supply for destination k. Because it is on the right hand side of a $\geq$ inequality, then increasing μ_k *tightens* the constraint, which risks reducing the feasible solution to a lower-performing optimal solution. On the other hand, by decreasing μ_k, the constraint is *relaxed*, and a newer, better optimal solution may be found in the extended feasible region. This is consistent with the analysis from the first objective function, $E(W^*)$. Therefore, the μ_k parameter is *doubly* as important as the others since it can affect the optimum in two distinct ways.

Table 1 summarizes the parameters and their potential effects on the optimal solution.

Table 1: Sensitivity analysis of parameters.

Same or better optimal value:	π_l
Same or worse optimal value:	$a_{ij}, b_{jk}, \delta_k, \mu_k(\times 2)$

[2]Constraints 13-16 are all equality constraints, which are more difficult to analyze, and Constraint 18 only has the big M as a constant, which is theoretical and not intended to vary.

7 Conclusions from Models

Although the first model lacks specificity which would come from including a changing aquifer density and material, the PDE model is considered an efficient and reasonably accurate estimator of underground pollution migration in groundwater sources. Our results lead us to the conclusion that our current reliance on hydrofracking as a means of natural gas exploration is environmentally detrimental. We can determine this based on the far-ranging movement of the plumes even in our somewhat inaccurate model which we believe to have underestimated subterranean flow. We believe the model to be an underestimate due to general inaccuracy of the basic numerical differentiation techniques and the lack of a changing aquifer type. This model could easily be generalized to model any type of pollution in ground water, so it could become a powerful tool for predicting the future of underground fresh water resources.

The linear optimization model was originally designed to handle a small subset of the water sources, purification facilities, destinations, and fracking facilities; however, the only limitation to the size of the problem is the computation time required to solve a linear program. There are also many ways to adapt the program to fit any particular researcher's aims. For instance, there are three potential configuration of objective functions, but two of them would be preferred by most who aim to gain evidence against the fracking industry. The resulting output could be considered an intervention plan for how to optimize one's current water resources and how to sustainably pursue a profitable future. Careful research into the parameters will help insure the most accurate and insightful intervention plan; however, the parameter for minimum water supply per destination, μ_k, seems to be more consequential as a parameter than the others.

To summarize these results, we would suggest the following intervention plan (Task 4):

1. First, identify the fracking facilities that are causing the most environmental violations. Their environmental damage far outweighs any profit they are accruing, so shut those down, immediately. If necessary, optimize the linear program with the single constraint to help identify the "weakest links."

2. Experiment with the PDE model to identify which water sources are most at risk for becoming unusable, and ban those, as well. Make sure the remaining fracking facilities have minimal environmental violations and that their continued pollution will be manageable over the course of a reasonable amount of time.

3. Finally, optimize with the linear program with the multiple objectives, prioritizing the water supply, in order to optimally manage the current fresh water resources.

The problem we decided to face–hydrofracking in Pennsylvania–was rather specific, but the models we have developed are quite flexible and even generalizable. The long-term modeling of pollution gives a realistic understanding of the ramifications of certain environmental industries and allow for appropriate planning. Just as important is the ability to well-manage one's resources at any given moment in time. With tools such as these, many regions may be able to find relief from many resource shortages other than water scarcity.

8 Appendix

8.1 PDE Code [3] [13]

```
import numpy as np
import matplotlib.pyplot as plt
T = 0.0019259230725
s = 6*10**(-4)
def time_step(x_length,y_length,y,a,b,n):
    h = (b-a)/n
    for i in range(0,n):
        for i in range(0,len(violations)):
            y[coor[i,1],coor[i,0]] = violations[i]
        for i in range(0,y_length-1):
            for j in range(0,x_length-1):
                y[i,j] = abs(((T*h)/(s))*
        (y[i+1,j]-4*y[i,j]+y[i-1,j]+y[i,j+1]+y[i,j-1])+y[i,j])
    return y
def contour_plot(x_length,y_length,field):
    x_axis = np.linspace(0,x_length,x_length)
    y_axis = np.linspace(0,y_length,y_length)
    fig, ax = plt.subplots()
    X, Y = np.meshgrid(x_axis, y_axis, copy=False, indexing='xy')
    z = plt.contourf(X, Y, field,cmap='summer')
    fig.colorbar(z, ax=ax,ticks=lvls)



x_length = 280
y_length = 160
y = np.matrix(np.zeros((y_length,x_length)))
violations = np.array([np.log(111),np.log(153),np.log(67),np.log(109),np.log(
coor = np.matrix([[13,10],[15,17],[102,90],[130,138],[160,140],[170,150],[195
for i in range(0,len(violations)):
    y[coor[i,1],coor[i,0]] = violations[i]
scalar_field = time_step(x_length,y_length,y,0,15,2000)
scalar_field = np.array(scalar_field)
lvls = [0,.8,1.6,2.4,3.2,4.0,4.8,5.6,6.4]
contour_plot(x_length,y_length,scalar_field)
aquifer_pulls = np.matrix([[140,160],[120,240],[5,140],[10,250],[10,210],[20,
for i in range(0,len(aquifer_pulls)):
    print(scalar_field[aquifer_pulls[i,0],aquifer_pulls[i,1]])
```

8.2 LP Optimization Model

Decision Variables:

$$\boldsymbol{\alpha}_{i,j} = \text{Amount of water transported from source } i \text{ to purification facility } j \quad \forall i,j \text{ (continuous).} \tag{1}$$

$$\boldsymbol{\beta}_{j,k} = \text{Amount of water transported from purification facility } j \text{ to destination } k \quad \forall j,k \text{ (continuous).} \tag{2}$$

$$\boldsymbol{p}_j = \begin{cases} 1, \text{ if purification facility } j \text{ is active} \\ 0, \text{ if purification facility } j \text{ is inactive} \end{cases} \quad \forall j \text{ (binary).} \tag{3}$$

$$\boldsymbol{f}_l = \begin{cases} 1, \text{ if fracking facility } l \text{ is active} \\ 0, \text{ if fracking facility } l \text{ is inactive} \end{cases} \quad \forall l \text{ (binary).} \tag{4}$$

$$\boldsymbol{v}_i^1 = \text{Amount of unusable water at source } i \quad \forall i \text{ (continuous).} \tag{5}$$

$$\boldsymbol{v}_j^2 = \text{Amount of unusable water at purification facility } j \quad \forall j \text{ (continuous).} \tag{6}$$

Simple Constraints:

$$\boldsymbol{\alpha}_{ij} \geq 0 \qquad \forall i,j \tag{7}$$

$$\boldsymbol{\beta}_{jk} \geq 0 \qquad \forall j,k \tag{8}$$

$$\boldsymbol{p}_j \in \{0,1\} \qquad \forall j \tag{9}$$

$$\boldsymbol{f}_l \in \{0,1\} \qquad \forall l \tag{10}$$

$$\boldsymbol{v}_i^1 \geq 0 \qquad \forall i \tag{11}$$

$$\boldsymbol{v}_j^2 \geq 0 \qquad \forall j \tag{12}$$

Flow Constraints:

$$V_i + \sum_l \phi_{il} \boldsymbol{f}_l = \boldsymbol{v}_i \qquad \forall i = 1 \tag{13}$$

$$\sum_{j=1} \boldsymbol{\alpha}_{i,j} = W_i - \boldsymbol{v}_i \qquad \forall i = 1 \tag{14}$$

$$\boldsymbol{v}_j^2 = \theta_j \sum_{i=1..} \boldsymbol{\alpha}_{i,j} \qquad \forall j \tag{15}$$

$$\sum_i \boldsymbol{\alpha}_{i,j} - \boldsymbol{v}_j^2 = \sum_k \boldsymbol{\beta}_{j,k} \qquad \forall j \tag{16}$$

$$\sum_{j=1} \boldsymbol{\beta}_{j,k} \geq \mu_k \qquad \forall k \tag{17}$$

$$\sum_i \boldsymbol{\alpha_{ij}} + \sum_k \boldsymbol{\beta}_{jk} \leq M\boldsymbol{p_j} \qquad \forall j \tag{18}$$

Objective Functions (Maximize):

$$E(W^*) = \sum_k \frac{1}{\delta_k}\left[\left(\sum_j \boldsymbol{\beta}_{jk}\right) - \mu_k\right] - \left(\sum_i \boldsymbol{v}_i^1\right) - \left(\sum_j \boldsymbol{v}_j^2\right). \tag{19}$$

$$E(F^*) = \sum_l \pi_l \boldsymbol{f}_l - \left(\sum_i \sum_j \boldsymbol{\alpha}_{ij} a_{ij}\right) - \left(\sum_j \sum_k \boldsymbol{\beta}_{jk} b_{jk}\right). \tag{20}$$

- Single Objective ($c_1, c_2 > 0$):

$$\text{maximize } c_1 E(W^*) + c_2 E(F^*)$$

- Multi-objective, prioritizing Environmental Resources:

$$\text{(1) maximize } E(W^*)$$

$$\text{(2) maximize } E(F^*)$$

- Multi-objective, prioritizing Financial Resources :

$$\text{(1) maximize } E(F^*)$$

$$\text{(2) maximize } E(W^*)$$

References

[1] Amico, C., DeBelius, D., Detrow, S., and Stiles, M: Pennsylvania Counties With Active Wells. NPR, Washington, D.C., USA. (2011) http://stateimpact.npr.org/pennsylvania/drilling/counties/

[2] Annual Drinking Water Quality Reports for Pennsylvania. Enviromental Protection Agency, Washington, D.C., USA. $http : //ofmpub.epa.gov/apex/safewater/f?p = 136 : 103 ::: NO : RP, 103 : P103_STATE : PA$

[3] Ascher, D., Dubois, P., Hinsen, K., Hugunin, J., and Oliphant, T.: Numerical python. Lawrence Livermore National Laboratory, Livermore, California, USA. (2001) http://www.pfdubois.com/numpy/

[4] Chapra, S. and Canale, R.: Numerical methods for engineers, ed. 5. Boston: McGraw Hill. (2006)

[5] Davis, K. and Putnam, L.: Conceptual and numerical models of groundwater flow in the Ogallala Aquifer in Gregory and Tripp counties, South Dakota. Water Years 19852009, US Geological Survey. (2013)

[6] Fleeger, G.: The geology of Pennsylvania's groundwater. Harrisburg: Pennsylvania Geological Survey. (1999)

[7] Hydraulic Fracturing 101. Earthworks, Washington D.C., USA. $https : //www.earthworksaction.org/issues/detail/hydraulic_fracturing_101\#.Vq6_4sclc9c$

[8] List of Counties in Pennsylvania. Wikipedia: The Free Encyclopedia. (2016) $https : //en.wikipedia.orgwiki/List\_of\_counties\_in\_Pennsylvania$

[9] Pennsylvania Active Water Level Network. United States Geological Survey. (2015) $http : //groundwaterwatch.usgs.gov/StateMap.asp?sa = PA\&sc = 42$

[10] Rardin, Ronald L.: Optimization in operations research. Vol. 166. Upper Saddle River: Prentice Hall. (2010)

[11] Rausch, R.: Groundwater modeling: an introduction to groundwater flow and solute transport modeling with applications. (2009)

[12] United Nations Environment Programme. (2004) $http : //www.unep.org/dewa/vitalwater/jpg/0222 - waterstress - overuse - EN.jpg$

[13] van Rossum, G. and Drake, F. (eds): Python reference manual. PythonLabs, Virginia, USA. (2001) http://www.python.org