Assignment 4: Exploratory Model Analysis

Matthijs Brouns 1505157 18th of December, 2014

Hélène van Heijningen 4076850

*Analyze the model you have built carefully or choose to use the 2-area epidemic model supplied. What are the main uncertain factors (7 to 15 factors) in this model? What are justifiable ranges/values for these factors? Use the sensitivity analysis tools in Vensim to explore the behavior of your model over these ranges. Analyze the results and discuss the policy implications of your findings.*

*In your write up of the assignment, clarify which uncertain factors you have chosen and why, what ranges you have used and why, motivate your sampling technique, and your analysis of the results. The write up should be such that your analysis can be replicated by me, in theory.*

*Tips:*

* *You can export the data from the sensitivity analysis to tab separated or comma separated files for further analysis. Read the Vensim manual for details*
* *Many of the discussed techniques (PRIM, CART) are available in open source implementations. You could even use the code of the exploratory modeling workbench if desired.*
* *Ranges should be plausible and grounded in information.*
* *Reference the theory material.*

# Approach

In this Section we will explain our approach to exploratory model analysis of our version of the Kirkwood water crisis model.

We start off by identifying the most uncertain yet important parameters in our model. We identified these parameters by looking at the structure of the modeling, using the causal diagram shown in figure X, as well as by using the results from the sensitivity analysis that has been performed earlier. After going through the different parameters and structures that are uncertain in the model, we have chosen to focus on 10 factors. Why we chose for to take into account certain factors and why we decided to keep others out will be explained in the next Section.

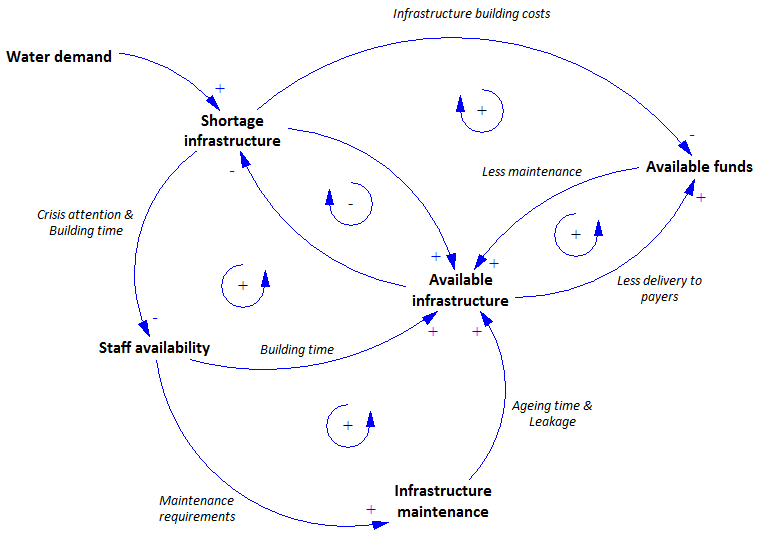
After the factors were chosen, their ranges, or values, had to be determined. These values have to be realistic, as in that these values might occur in the real system even if they happen in uncommon scenarios. Doing this for all the different factors was difficult as it involves a lot of guessing, but we will explain more on that in the Section on factors and their ranges.

Next we determined what output variable we wanted to do the multivariate analysis for. This is discussed in the Section about classifying model behaviour. The run setup was the following step that needed to be addressed. Different types of multivariate analysis were compared and a realistic yet statistically relevant amount of runs were chosen.

The multivariate analysis was then run and the data was loaded into R to do further analysis. We wanted to classify the different types of behaviour produced by the multivariate analysis to then use this to find out what values the factors had to adopt so that the model would give the behaviour we were looking for; Kirkwood where no crisis happens or the crisis gets solved by the end of the run.

Then by analyzing the potential policies that were possible to implement in the Kirkwood water system, given the values the factors need to adopt so that water crisis are solved or do not happen, these policies were put back into the model.

This has been done to test the robustness of the different policies to come to a final conclusion on what policy could be implemented that is the most effective and robust for the Kirkwood area.



# Factors and their ranges

When determining the factors that were uncertain and needed further exploration with regards to the possible effects of these factors on the model behaviour, we identified 10 factors. Of course there are many more factors in our model and we will now explain why we chose to take these into account and keep other factors out of this exploratory model analysis.

## Staff availability

From the sensitivity analysis earlier performed we knew that factors influencing the staff availability had a relatively large effect on the model behaviour, especially on whether a crisis would occur or not. The factors that influence the staff availability are the number of staff hired, hiring time period and the staff service time. From these factors we chose to take into account the number of staff hired per time period (so the hiring time period stays the same, but we change the number of people hired) and the staff service time because those are the main causes of changes in staff availability. In terms of policies we also thought these two factors would be interesting as a policy could consist of both hiring more people, or educating more people so there are more suitable workers, and of making a staff member’s service time longer by providing a more attractive working environment and agreements.

### Staff hired / time period

Range: 0 – 25 staff members hired per 2 years

Distribution: Vector (0;25;1)

The number of staff hired per time period (2 years) has a base value in the model of 5 members per 2 years. We have set this range in the multivariate analysis from 0 to 25 staff members hired per 2 years. 25 per 2 years may be too much for the current system to be able to do in the future, but we wanted to be sure that we would grasp the effect of this very important factor in its entirety, even if this meant that the model might only show no crisis when the staff hired per 2 years had to increase to 25 members. This has been modeled as a vector incremented by 1, since you cannot have 3.4 staff members for example.

### Staff service time

Range: 0 – 25 years

Distribution: Uniform (0;25)

The staff service time has a base model value of 10 years (it takes 10 years on average for a staff member to leave). We have set the range for this factor from 0 to 25 years as this seemed a realistic value to where the staff service time could increase, by for example implementing a policy that provides better work agreements and a better environment for the workers so they will stay longer. The distribution that has been used is a uniform distribution between 0 and 25 so that all possible time periods are taken into account.

## Distribution of staff attention

The other part of the staffing system involved dividing the attention of available staff over the different activities such as maintenance, planning, constructing and refurbishing activities. Factors that influence this system are the average production rate of a staff member, the distribution of activities between primary and secondary activities, the distribution of activities between planning and construction tasks (secondary activities) and finally the number of staff required for one tanker. We have decided to take the last three factors into account and leave the first one, the average production rate per staff member out of this analysis. This is because the average production rate per staff member is not an uncertain value and could only theoretically be altered by making the work of the staff members more simple so they can do more in the same time period. However, we believe this is not a policy that is easily comprehensible or implementable and above all not realistic as there is no money available in Kirkwood to set up a more efficient working environment. The other factors however are very interesting as they control how much staff attention is given to different tasks. By altering this distribution we are curious to see if this could have a positive impact on the system.

### Attention rate for primary / secondary staff activities

Range: Proportionally; Primary activities first; Secondary activities first

Distribution: Vector (0;1;0.5)

The attention rate for primary / secondary staff activities is currently being distributed proportionally to the amount of maintenance production required and secondary staff production required. This distribution could however be changed by altering the model structure. In the multivariate analysis we include this factor by including a variable in the model which changes the model structure when it takes on one of the following three values: 0, 0.5 or 1. We distinguish between three possible structures:

* Divide staff by ratio between primary and secondary activities
* Do all maintenance tasks first before giving attention to secondary activities
* Do all secondary tasks first before giving attention to primary activities

### Attention rate between planning and construction tasks

Range: Between 1% planning, 99% refurbishment and 99% planning, 1% refurbishment

Distribution: Uniform (0.01;0,99)

This factor has a base model value that distributes the planning and refurbishment tasks as follows: 25% of the attention of secondary tasks is given to planning and 75% of the attention is given to refurbishment activities. To see what effect this uncertain factor could have on the model behaviour it has been taken into account with a uniform distribution of values between 1% and 99% to make sure that at least some attention is given to all activities at all times.

### Staff required for tankers

Range: 0 – 1

Distribution: Vector (0;1;1)

In our model we made the assumption that part of the staff that works on infrastructure would have to work on supplying water to the households when tankers were in play in case of a crisis. This means that when there is a crisis and tankers are needed, a large part of the entire group of available staff members is taken away from infrastructure jobs and towards the tankers. Since this is an assumption on our part, this is an uncertain factor which could influence how crisis would evolve during time. Therefore it was very important to take this factor into account. The range of this factor is set to either 0 staff members being needed to tankers or 1 staff member needed, equal to the current model setup.

## Funding system

The funding system is the next system in which several uncertain factors can be found. The main factors influencing this system are the annual financial bail-out, the percentage of income used for maintenance, the price per unit of infrastructure, the maintenance costs per unit infrastructure and the percentage of delivery costs recovered from households and other urban users. The annual financial bail-out had already been established as a very uncertain factor as the government was already contemplating removing this subsidy in the future. Because of this, this factor is taken into account to get a grasp of the real importance of this factor in the system. Furthermore, the percentage of income used for maintenance is taken into account to see if a different budget allocation of the Kirkwood municipality might influence the system for the better. The maintenance and infrastructure unit costs are not taken into account as these factors are not uncertain. Lastly, the percentage of delivery costs recovered has not been taken into account

### Annual financial bail-out

0 – 200

UNIFORM

### Percentage of income used for maintenance

0.2 – 0.6

UNIFORM

## Infrastructure system

#### Infrastructure aging time

20 – 40

UNIFORM

#### Max infrastructure pushing

0 – 0.5

UNIFORM

#### Target percentage infrastructure maintenance

5 - 25

UNIFORM

# Run setup

Number of runs: 2000

Say we want at least 20 levels per variable

20 ^10 -> 1,024 e13 runs for full factorial -> completely unrealistic

Latin Hypercube needs less samples since more samples are not needed for more input dimensions. Max number of combinations for a cube can be determined using following formula with M divisions and N dimensions.

\left(\prod_{n=0}^{M-1} (M-n)\right)^{N-1} = (M!)^{N-1}

This still results in 1e7 possibilities. To keep analysis feasible we run 10.00 runs.

# Model behavior classification

How do we define a crisis: maybe simply whether there is infrastructure shortage?

#### Classify behavior on type of crisis which occurs

Problems:

Crises happen in pairs and follow up on eachother soon

No distinguishable difference between runs where crisis is never solved or crisis is solved

#### Classify behavior on timeslot where crisis first occurs

Problem: no feedback on whether crisis gets solved

#### Classify behavior on number of years of crisis

No feedback on type of crisis, however, not a problem because crises happen together mostly

Problem: No idea when crisis occurs, only how long the crisis occurs

No idea on severity of the crisis

#### Classify behavior on number of years of crisis + timeslot where crisis occurs

Solves problem with previous

However: much more classification categories (previous category \* number of timeslots) -> decreased interpretability and increased curse of dimensionality (requires a lot more runs to statistically make sense)

#### Classify behavior on infrastructure shortage \* period (area under the curve)

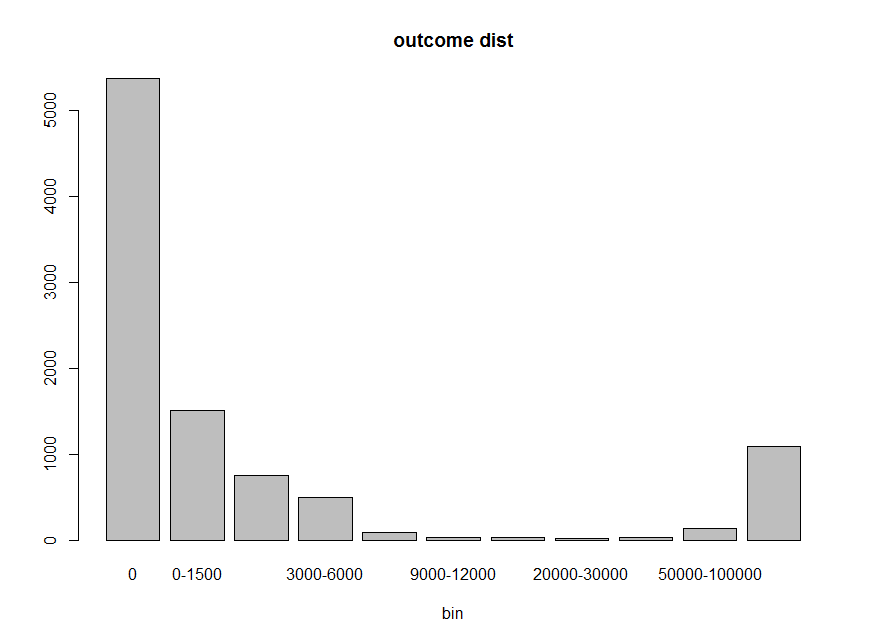
Gives very accurate, high resolution result

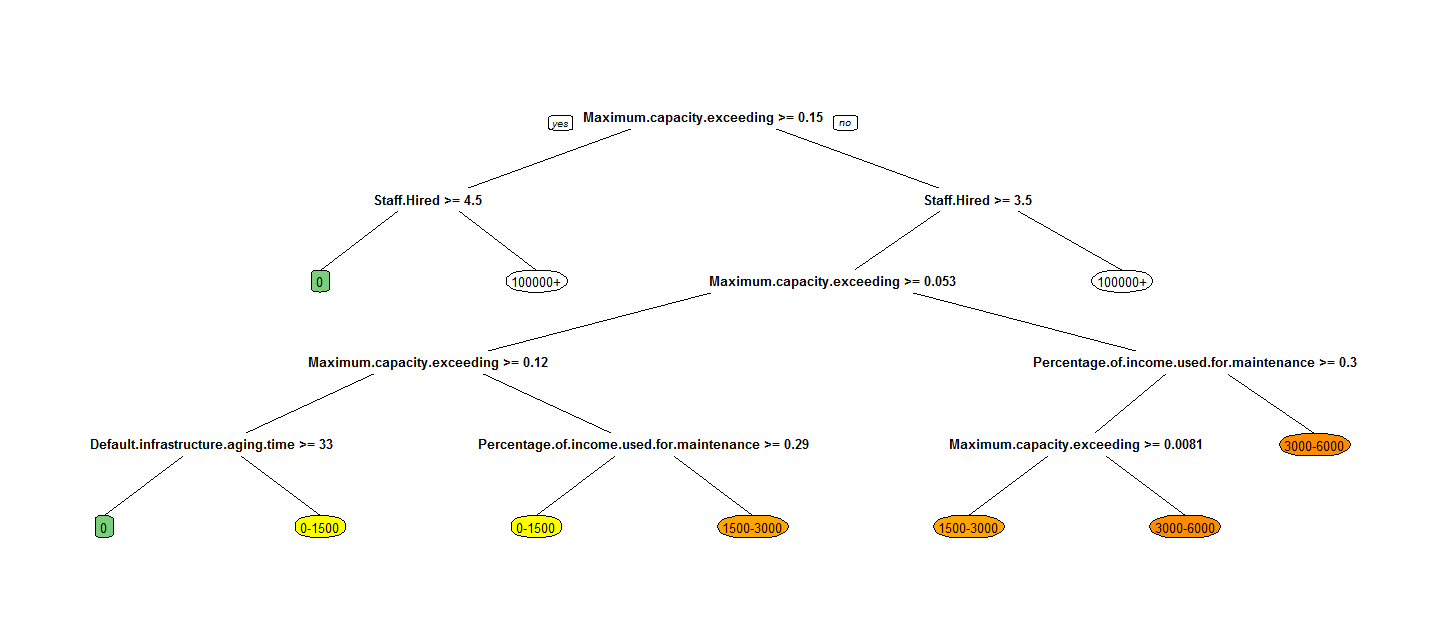
Lower interpretability

To increase interpretability we can divide the infrastructure shortage (in ML / year) by the average household water consumption to get household\*years of shortage which is a more relatable measure.

Keep in mind that this means that a low number of households with a long period of shortage is classified the same as many households with short shortages. This may or may not be desirable for all purposes but we think it’s fine for this.

# Results





# Policy

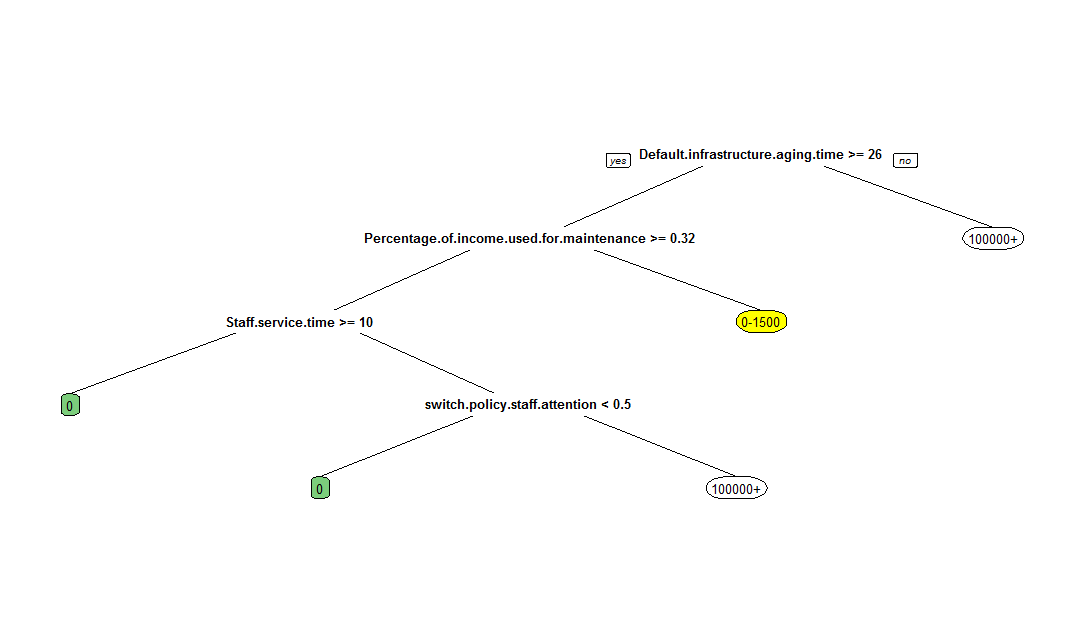
Staff hiring increase to 5 per two years. Keep a minimum of 15% infrastructure pushing available to act as a buffer

Staff hiring increase to 4 per two years while keeping at least 12% infrastructure pushing available and use infrastructure which has a slightly longer default lifespan

# Policy robustness

To test policy a sensitivity analysis is performed

Policy 1:



Policy 2

