Assignment 1b

Matthijs Brouns 1505157  
Hélène van Heijningen 4076850

# Question 1

*Generate three data series, each having 150 data points, using the following functions;*

*f(t)= 40+SIN(t/5)\*50+5\*t*

*g(t)= 55+SIN(90+t/5)\*30+5\*t*

*h(t)= 10+6\*t*

*Assume that data series generated by f(t) represents measured data from a real system, and that the data series g(t) and h(t) are model-generated outputs. Without using the formulae above in your argumentation, discuss the behavioral validity of the model-generated data based on quantitative evidence. Please refer to the behavior reproduction tests in Sterman's chapter on validation (see Reading Material on Model Building, Data and Validation on Blackboard).*

We have used Excel to generate the values for f(t), g(t) and h(t) for 150 data points. Next the measures shown in table 1 were calculated to determine whether g(t) or h(t) is a better behavioral fit compared to the original data set f(t).

|  |  |  |
| --- | --- | --- |
| **Measures** | **g(t) compared to f(t)** | **h(t) compared to f(t)** |
|  | 0.95 | 0.97 |
| MAE | 43.77 | 60.91 |
| MSE | 2487.36 | 5426.77 |
| RMSE | 49.87 | 73.67 |
| MAPE | 14.64 % | 23.59 % |
| MAE/Mean | 10.15 % | 13.16 % |
| Bias () | 0.06 | 0.36 |
| Unequal variation () | 0.00091 | 0.36 |
| Unequal covariation () | 0.94 | 0.28 |

R-squared is theme coefficient of determination, measuring the fraction of the variance in the data ‘explained’ by the model. However, R-squared will not be used here to determine the behavioural match between both models as R-squared can have very different values on a common trend with the same absolute error values. Therefore this measure is not that useful.

MAE (mean absolute error), MAPE (mean absolute error as a percent of the mean), MAE/Mean and (R)MSE ((root) mean square error) all provide measures of the average error between the simulated and actual series. MAE weights all errors linearly while (R)MSE weights large errors much more heavily than small ones. Both measure the error in the same units as the variable itself. MAPE should not be used since the data series includes points close to zero. In this case, MAE/Mean provides a more robust dimensionless measure.

The Theil inequality statistics give a decomposition of the error by dividing the MSE into three components: Bias, Unequal variation, Unequal covariation. A model has a high bias when the model output and data have different means. Unequal variation indicated that the variances of the two series differ. Unequal covariation means that the model and data are imperfectly correlated, that is, they differ point by point. These three fractions together are equal to 1. The Theil inequality statistics are thus a great way to get insight into the behavioural validity of a model.

The behavioural validity will thus be based on the values of MAE, (R)MSE, MAE/Mean and Theil’s Inequality Statistics.

## Behavioural validity

Since the unequal covariation is very high (0.94) and the bias (0.06) and unequal variation (0.00091) are very low, the formula g(t) captures the mean and trends in the data very well. However, this does imply a possible systemic error when phasing is important for the purpose of the model and if the model is driven by historical data. The error can be characterized as unsystematic if the model is driven by random noise. Cases like these can indicate a fairly constant phase shift of a cyclical mode otherwise reproduced well. More likely, a large indicates the presence of noise or cyclical modes in the data series not captured by the model. When is large, the majority of the error is unsystematic; a model should not be faulted for failing to match the random component in the data.

Large errors and large bias or unequal variation fractions indicate systematic error and should lead to questions about the assumptions of the model. Therefore a model with high and low and is an ideal model, minimizing systemic errors and matching the behaviour of the real system very well.

Furthermore, the MAE, RMSE and MAE/Mean values are lower than the values for h(t) which indicates that g(t) is a better behavioural fit than h(t). All in all we conclude g(t) has a higher behavioural validity.

# Question 2

*Justify your formulation of a LOOKUP function in the Kirkwood water crisis model. Please refer to the theory (and additional) material on table functions.*

We will discuss the formulation of a LOOKUP function in our infrastructure subsystem. This LOOKUP function is called the ‘Effect of maintenance on leakage’. As the name mentions, it describes the effect of low levels of maintenance on the amount of water wasted through leakages.

The background reading stated that on average 30% of water is wasted through leakages. We assumed that this percentage of leakage occurs at the baseline of 8% of infrastructure maintained per year. The input of this lookup is the percentage of maintained infrastructure annualy which ranges from 0 to 100%. The output is the percentage of leakage which occurs. We assume that when all infrastructure receives regular maintenance, leakages can be reduced to 0%. However, when no infrastructure is maintained, leakages will increase to 50%.

The lookup is modelled as shown in figure 1.

The lookup graph is modified to have 0-slopes at both ends in order to be able to properly extrapolate the lookup outside of the given boundaries. The “effect of maintenance on leakage” variable puts the value of percentage infrastructure maintained annually in a lookup extrapolate with the lookup and multiplies it by the normal: “leakage at 0perc maintenance”.

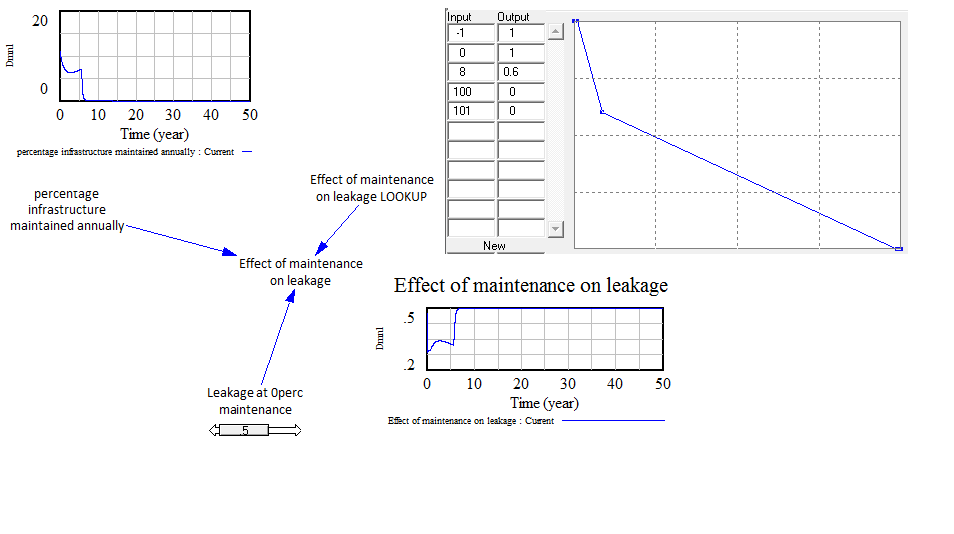


Figure 1: Effect of maintenance on leakage

This setup makes sure that the lookup is rigid because extrapolation is done correctly and that it is dimensionally consistent and robust (bron!) due to all input and output of the lookup being dimensionless.

# Question 3

*Validate your Kirkwood water crisis model using three different types of validation test excluding sensitivity analysis (this is the subject of assignment 3). For instance, you could use extreme value analysis or behavior pattern tests, amongst others.*

A model must be tested thoroughly before policies can be designed and implemented. Model validity is one of the most important tests that need to be passed by a model. Model validity, or credibility, is used to establish that thee model is an acceptable description of the real system with respect to the dynamic problem of interest. Model validity can be done by two types of tests as described by Barlas (2007).

* Structure tests: is the structure of the model a meaningful description of the real relations that exist in the problem of interest?
* Behaviour tests: are the dynamic patterns generated by the model close enough to the real dynamic patterns of interest?

When validating our model, we will focus only on structure tests where we check the relations within the model and how the shown behaviour is consistent with the structure of the model. Behaviour tests will not be done since we do not have real system data to be able to check the behaviour of our model.  
Three validation methods will be used (Sterman, 2000)

* Theoretical direct structure test: Structure assessment
* Structure-oriented behaviour test: Extreme conditions
* Structure-oriented behaviour test: Behaviour reproduction

The structure assessment test is used to test whether the model structure is consistent with relevant descriptive knowledge of the system. One of the methods of doing this test is drawing causal diagrams of the system (Sterman, 2000)   
The extreme conditions test is used to check whether equations still make sense when its inputs take on extreme values. This test can include tests which focus on whether the model responds plausibly when subjected to extreme policies, shocks or parameters. The test is executed by inspecting each equation and testing the response to extreme values of each input, alone and in combination.  
The behaviour reproduction test is used to determine whether the model generates various modes of behaviour that are observed in the real system. It is executed by examining the behaviour of the model in comparison with the behaviour of the real system.

The structure assessment test is a validation test which falls under the direct structure test category. This means the validation tests strictly focuses on validating the structure of the model and not the behaviour. Theoretical direct structure tests are validation tests which have to be done in the earlier steps of validating a model according to Barlas' sequence of steps of model validation (1996). Structure-oriented behaviour tests are performed afterwards. The structure assessment test will be performed only for the entire model and not for its sub-systems since the model is not that large. The structure-oriented behaviour tests will be done for all the sub-systems and for the entire model.

## Overall model validation

### Structure assessment

We have performed this test by drawing a causal diagram of the system as it was communicated to us and then checking this against how we modelled the relations in the actual model. Figure 2 shows the causal model of the Kirkwood water infrastructure system.

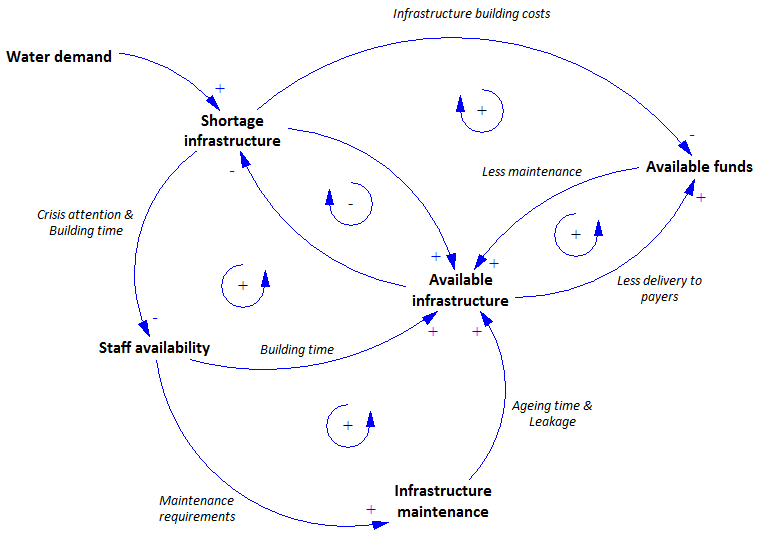


Figure 2: Causal model

The model has been checked on the implementation of all of these described feedback loops and we found that all have been implemented.

## Infrastructure System Validation

The infrastructure system will be validated by performing behaviour reproduction and extreme conditions testing.

### Behaviour reproduction

When the system is not constrained by funds or staff constraints it is expected that the practical infrastructure capacity will meet the total water demand relatively easy and smooth in about 5 years (the infrastructure building time). Introducing a large step change in the water demand in the system will result in a large peak which will also be solved relatively smoothly.

Figure 3: Baseline system behaviour

### Extreme conditions

When we drastically reduce the time required to build and plan infrastructure we expect the shortage to be solved sooner than in our base case scenario. However, when we increase the planning times we expect that the shortage cannot be solved and instead remain on a constant level over time. Also, the oscillations around the target infrastructure capacity will be much lower when the delays are smaller. This is shown in Figure 3.

Figure 4: Extreme value testing on building times and water demand

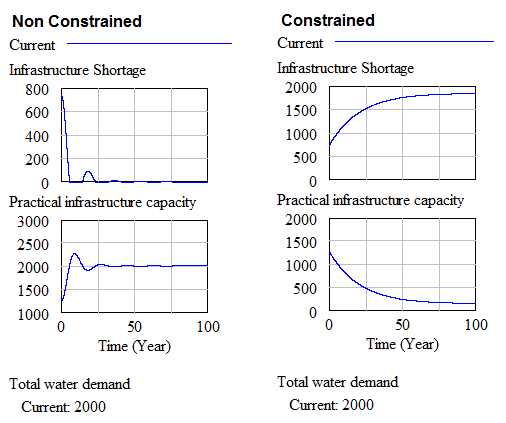
When performing extreme values testing on the variables constraining the building and maintaining of the infrastructure we expect to see that in the non-restricted cases the infrastructure shortage gets solved smooth and easy. However, when heavily constrained we expect the shortage not to be solved and instead remain on a constant level over time. This is shown in figure 4.

Figure 5: Extreme value analysis on constraining variables

## Population system validation

Since the population system is used mostly as input to the model it is mostly important that it represents the actual population growth in Kirkwood relatively well and that it allows for experimentation with different growth scenarios. Therefore in the validation phase we only do a behavioural reproduction test and not an extreme conditions test.

### Behavioural reproduction

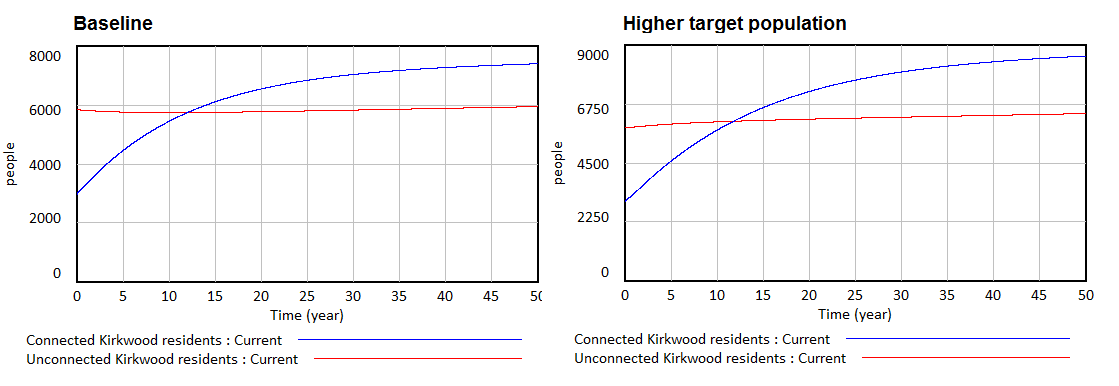
The actual behaviour of the population in Kirkwood was described to us as starting at around 3000 connected and 6000 unconnected residents. The connected Kirkwood residents show a limited growth stabilizing at about 6000-8000 residents. The unconnected household show a small decline but stabilize at around 6000 residents. As figure x. shows, the population system replicates this behaviour rather well. Also, adjusting the target population values shows a similar behaviour, but with larger stabilizing values.

Figure 6: Population system behaviour

## Staff system validation

The staff system will be validated by performing behaviour reproduction and extreme conditions testing.

### Behavioural reproduction



The staff system is expected to show a constrained decrease in available staff members due to the average staff service time being 10 years while only 1 staff member is hired every 2 years. The number of staff members will decrease to 5 over time.

Figure 7: Behavioural reproduction of the staff system

Due to this decrease in staff members, the staff occupancy will rise given a stable water demand and infrastructure. Because of this increase in occupancy the maximum possible production for the staff members will plummet due to stress and tiredness.

## Extreme conditions

The hiring of staff is changed from 1 staff hired every 2 years to 100 staff hired every 2 years. The behaviour which results from this change is to be expected because there is plenty of staff available now to work on all infrastructure related jobs. The staff occupancy is so low that the production rate per person per year will be increased by 50% compared to the normal production rate of 20Ml/(year\*staff).



Figure 8: Extreme conditions test by increasing staff hired

## Funding system validation

The funding system will be validated by performing behaviour reproduction and extreme conditions testing.

### Behavioural reproduction

The funding system behaves as expected; more money is spent on maintenance per year then is received from municipality income. This makes the available funds for maintenance decline over time.

Over time the expenditures decline since there is not enough staff to work on infrastructure maintenance, so no money is spent on infrastructure maintenance either. Since no infrastructure is maintained or built, the income from the water supplied stagnates as well, until it reaches the yearly value of the bail-out they receive. This makes the available funds for maintenance increase linearly. In reality this money would also be used, however the other expenditures of the municipality are not taken into account which makes the model behave like this.

Figure 9: Behavioural reproduction of the funding system

### Extreme conditions

Staff hiring is increased from hiring 1 staff member every 2 years to hiring 100 staff members every 2 years. We would expect that this eliminates the staffing constraint on the model, so enough infrastructure can be build and enough money is received from the water supplied to Kirkwood. The model shows this behaviour, however first couple of years the expenditures are low until the expenditures match the income of the municipality. After this point is reached more income from supplied water comes in which is used to build / maintain infrastructure which gives the staff the opportunity to build more infrastructure. This continues until enough infrastructure is built to meet the demand of the population.

Figure 1: Extreme conditions test by increasing bail-out

Figure 10: Extreme conditions test by increasing the staff hired

In the next extreme conditions test the annual financial bail-out is increased from 150kZAR/year to 15000kZAR/year. We would expect that this gives the city enough money to build, however the amount of infrastructure maintained is still constrained by the available staff. This is also shown in the behaviour of the model as for a short time the expenditures increase but then again decline since no infrastructure can be maintained because of the low amount of staff available which leads to a crisis which decreases the amount of staff working on maintenance even more.

## Crisis system validation

The crisis system will be validated by performing behaviour reproduction and extreme conditions testing.

### Behavioural reproduction

In the real system water crisis does occur which means that people are so unhappy about the water delivery system that unrest raises. Also, in a water crisis when water supply is very low tankers are put in place to make sure the people get enough water to survive. The model shows the same kind of behaviour where tankers are indeed put in place to deal with the water shortage in the city and the inhabitants are very unhappy shown by the unrest on a scale from 1 to 10.

Figure 1: Behavioural reproduction of the crisis system

### Extreme conditions

In the extreme conditions testing the staff hiring is again increased from 1 per 2 years to 100 per 2 years. We would expect that this eliminates the staffing crisis, so enough infrastructure is built to meet the demand of the people. The model behaviour shows how this results in no tankers being needed in the system and the people of Kirkwood are not unhappy with the service at all.

Figure 13: Extreme conditions test by increasing the staff hired

A second extreme conditions test is done by increasing the annual financial bail-out from 150kZAR/year to 15000kZAR/year. This will deal with any financial problems, however as expected, since the staffing constraint is still in place a crisis still occurs resulting in the same behaviour as in the base case scenario.

Figure 14: Extreme conditions test by increasing the annual financial bail-out

## Coupled model validation

In this section a simulated run of the model will be used to determine the overall validity of the system.

We see that initially there is no infrastructure shortage. This is correct since the initial infrastructure capacity is higher than the total water demand by the initial population. This also means that the staff productivity is high since they are not overworked.

However, as the total water demand increases and the infrastructure ages, a sharp increase in the infrastructure shortage occurs. When this water infrastructure shortage is combined with the gradual decrease of staff members, the remaining staff members get heavily overworked. This will drop their productivity leading to decreased construction and maintenance of infrastructure. Due to the decreased maintenance, the infrastructure will start aging faster and leaking more, resulting in even larger infrastructure shortages.

Initially, funding of the system goes okay but as less people get water, the billable decrease, therefore decreasing the total available funds.

Figure 15: Total system behaviour

# Unit checks

We ran into several issues when trying to fix all the unit errors. You can check unit errors with several methods, ‘unit check’ in an equation, ‘unit check’ in the model and unit error report in the SDM doc. These three methods however don’t agree and sometimes one will say there is no unit error and the others say that there is.

The model has been made consistent with the unit check Vensim offers, not the unit check done by SDM Doc.

# References

Barlas, Y. (2007). System dynamics: systemic feedback modeling for policy analysis. *SYSTEM*, *1*, 59.

Sterman, J. (2000). TRUTH AND BEAUTY: VALIDATION AND MODEL TESTING. In *Business dynamics: Systems thinking and modeling for a complex world* (1st ed.). Boston: Irwin/McGraw-Hill.