Examination TPM

Delft University of Technology - Faculty of Technology, Policy and Management

Course name:	**Advanced Simulation**	**Course code:**	**epa1352**
Date:	**15 April 2021**	**Time:**	**09:00-12:00**
Module manager:	**Dr. Y. Huang**		

Examination questions:

Number of open questions:	**3 (*)** questions
Number of multiple choice questions:	**0** questions
Max. number of points:	**90 points**

☐ **all questions have the same weight**

☒ **the questions have different weights** (indicated per question)

Total number of pages (incl. cover page): **5+** pages

Use of tools and information sources:

During the examination, the use of any tools or information sources (this includes mobile phones, smartphones or any devices with similar functions) is strictly forbidden unless stated below.

Permitted tools and information sources:

☒ **books** ☒ **notes** ☒ **dictionaries** ☐ **readers** ☐ **formulae sheets**

☐ **calculator** ☒ **computer** ☒ **slides, papers, all course materials**

Additional instructions:

(*) 30 points each, total 90 points + 10 points = 100 points.

This mark contributes for 50% to your final mark, and has to get a mark ≥ 5.8 to be combined with the lab assignments (40%) and reading assignments (10%) for the overall mark.

Do not edit page 1 of this file.

Write your name and student number on page 2 of this file.

An exam file without honours pledge will not be graded.

Final marking date:
(the maximum marking period is 15 working days)
10 May 2021

To be submitted to BrightSpace EPA1352 "Assignment – Online Exam":

☐ Examination work with name and student number on each page.

☐ Examination documents

☒ This file

Any suspicion of fraud or any breach of the exam rules will be immediately reported to the Board of Examiners

For more information about fraud:
TU Delft Student portal> TPM> Rules and Guidelines

Name	**FirstName LastName**
Student Number	**1234567**

In answering the following questions, relate your answers to the EPA1352 lectures, readings, discussions and assignments.

---

1) **Systems Modelling and Simulation (30 points)**

a) Explain and reflect on the *levels of systems knowledge* and their relations to modelling and simulation. Use examples in your explanation. (~200 words, 10 points)

System knowledge is defined using the 5 levels described in the hierarchy of epistemological levels of systems (slide 13 Week 2; Klir).

The basic level (level 0: source system) is the level of primitive understanding. Lets take for example the EPA1352 course. At level 0 we look at who is teaching the course, who are the TA's.

The next level(level 1: data system) is the source system supplemented with data. So it is known what we are interested in but we gather data about it. So in the example this would be data from course evaluations and feedback from students.

The next level in the hierarchy (level 2: generative system) defines one overall characterizations of the constraints of the variables. So for example when looking at this course, we observe the course over time and see how grades and evaluations change.

The next level (level3: structure system) combines certain generative systems and look how these interact with each other. Here we try to relate different generative systems to each other. For example, this would imply other taking other courses into account as well and looking at which courses are related, or which course should be prerequisites for another.

The last level of system knowledge is that of Meta-systems (level 4). This level describes changes of system traits that are defined as invariant at lower levels. For example, when we are looking at the EPA course and the change over time we also include variables like percentage of online education, changing courses in the program and changing of teachers.

b) When used appropriately, simulation models can serve the society extremely well. This, however, is not necessarily true. List and explain three cases when models might not be used appropriately. Explain how to address the challenge(s) in each case. (~200 words, 10 points)

1. When models are built without due consideration of modeller bias. In this case a modeller may or may not have an agenda or reason for a particular outcome. There are many ways these biases can impact model results, from unconscious structural changes to more explicit 'data-hacking'. If these biases are not removed or accounted for, then the final model outputs will reflect the bias and produce skewed results. To help prevent this, data-based modelling techniques could be used such as data driven modelling and simulation, whereby data is used to refine a model. Adjusting the model on data alone can help to prevent this bias, however if there is bias in the underlying data, then this too would be reflected in the model.
2. When model outputs are used but the inherent limitations of the model are ignored to support an existing agenda, without properly conveying the underlying uncertainty about the accuracy or confidence in model outputs. In this case, politicians, for example, may use the model as evidence that their suggested course of action is correct or supported by 'science'. In reality it is very easy to skew or 'frame' the outputs of a model to fit your own agenda. To address this challenge, it is important that modellers adequately explain the limitations of all models and communicate the uncertainties. It would also be helpful to utilise standard practices or

benchmarks in model explanation to help prevent other politicians becoming convinced of another interpretation of the model.

3. Some simulation models are too complex. This not only hinders understanding but also usefulness as only few people will be able to grasp the idea and uncertainty builds up with more parameters. Hence, sometimes building less complex, smaller simulation models might help, to understand what the model actually does, and that information is not lost in complexity. Sometimes, less is more. This calls for simplification and reduction. A simulation model is created for a reason. However, sometimes that reason/the question tried to answer is lost in the process of simulating. The modeller should always question him/herself whether the model is fit for purpose, and strive for a balance between complexity and simplification.

c) List and explain three distinct data quality criteria that troubled you most in completing the lab assignments. What were the challenges? How did you solve them? If not, why? (~300 words, 10 points)

1. Semantic accuracy – How well the data conforms to the real-world value (Huang, 2013). In the lab the most obvious issue were large latitudinal and longitudinal spikes in roads. These were caused by erroneous lat/lon values in the data which resulted in incorrect LRP positions. Because the positions obviously did not conform to the actual position of the road LRPs they can be considered issues of semantic accuracy. To solve this issue we changed the data by interpolating the values of the erroneous LRPs based on a rolling median of the surrounding 7 LRP data points.
2. Pragmatic completeness – The degree to which the data is of sufficient breadth for the project (Huang, 2013). The key issue in this regard was the fact that certain road data was entirely missing from the provided dataset. In some cases, it was obvious from a line of consecutive bridges, that a road should be present. Inspecting the area manually from google maps also showed that a road was indeed present. Yet there were no road LRPs in some cases. This was a problem because it prevented a full understanding of a road network and would have implications on future simulation studies on the road network. To solve this issue we tried to identify bridges which were labelled with roads, that were not present in the data. We created a substitute road that connected these bridges. The road location was obviously not a reflection of the real system (semantically inaccurate) , however, by including these bridges and the road link in the network we thought that it would improve the overall pragmatic quality of the model.
3. Mapping consistency – This relates to issues where there is not uniformity in data values representing the same instance (Huang, 2013). A common problem we found was that when combining the road and bridge datasets, the location data for certain bridges did not match up for LRPs that were supposed to represent the same point. This caused issues whereby bridges did not actually appear on a road. We did not manage to solve this issue due to time constraints limiting our efforts, however one conceived solution was to assume accuracy of one dataset (road or bridge) and use that throughout the model and replace the location data of the areas with mapping consistency problems. This solution would still contain some errors; however it would help to solve the issue of mapping consistency, and help prevent confusion in the final model.

2) Data-driven Modelling and Simulation (30 points)

a) Explain and compare (1) data modelling, (2) simulation modelling (in general) and (3) data-driven simulation modelling. Include their principles as well as pros and cons in your answer. Use examples in your explanation. (~300 words, 15 points)

Data modelling, simulation modelling and data-driven simulation modelling have been covered in lecture 4 by Y. Huang. Furthermore, Keller&Hu (2019) explain data-driven simulation modelling. The table below gives an overview of the three approaches as explained in the lecture and the paper.

What?	Pro	Con

Data modelling: model represents relation/correlation between different sets of data Example: machine learning	- Good at showing correlations	- Correlation does not mean causality - Cannot deal well with anomalies and changing circumstances - Data quality influences output heavily
Simulation modelling: Converting expert knowledge into models simulating to understand more about the system, based on physical/operation laws and logic (also known as knowledge-driven simulation) Example: most models, SD campylobacter model created by the different groups in the Advanced System Dynamics course	- Combines system knowledge from expert and desired behaviour - Explains what causes what and why → ♉represents causality	- Includes bias of expert/modeller →only for specific aspects of the system, might omit certain parts (un-) consciously - System knowledge hard to acquire →often limited and easily outdated
Data-driven simulation modelling: framework that discovers/creates simulation models in an (semi-)automated way (including components and generative behaviour) Example: see paper by Keller&Hu (2019)	- Reduced bias from modeller as s/he only gives minor input and the model creates itself more or less →new insights - Only way to create and work with large scale simulations	- Often reproduces historic data very well →however future is not an exact replication of history - Often computationally demanding

As depicted in the table, data modelling focusses mostly on correlation while simulation modelling does so on causal relations. Data-driven simulation modelling aims to combine the two, removing the bias from the simulation modelling, which is often not wanted, so giving a solution which can be explained in an objective way.

b) List three different types of data-driven modelling and simulation. Explain and compare them. Include their principles and challenges in your answer. Use examples in your explanation. (~300 words, 15 points)

There are several approaches to data-driven modelling and simulation.

Darema in Yilmiz et al. 2014 calls his approach “Data-driven application systems (DDDAS)”. The idea behind this approach is to continuously update a model and the corresponding simulation through incorporating additional data dynamically. The model input data can be directly connected to measurements, so e.g., with covid new findings (number of positive cases) can be incorporated consistently. Therefore, the model is always up to date. This enables constant improvements in accuracy and augments the model continuously. However, it also requires models and algorithms that are stable and converge with guarantee.

A second approach is presented by Keller&Hu (2019). In comparison to the first approach explained, here not the data incorporation is automated, but the model generation: a model space is defined from which various models can be generated (data and scope so taking covid again: infection data and where infections happen for certain town). Then, a genetic algorithm searches through this model space creating a population (numerous agents (people) created with different interaction that explains the data), fitness functions evaluate how close the behaviour is to the desired one (which created agents/behaviour match best?). Last, it is evaluated how well the approach can be applied to specific domain (how well does this approach be used → robust, composable, flexible?). In this approach, the modeller is hardly present, resulting in a reduced bias.

Lastly, component-based modelling can be seen as data-driven modelling when model structure is generated automatically aggregating predefined model components. The focus is not set on using the newest data nor letting the model “run free”, but on creating it in a way in which its parts can be reused best. If components are designed in a compatible way, they can be combined (semi-)

automatically for modelling and simulation (Hofmann, 2004). This is closely linked to multi-resolution modelling as explained by David&Tolk (2007). The main advantage here is that the different components can be reused, and complex systems broken into manageable parts that can be combined again later. So, taking Covid as an example again, the model can be broken down into different submodels and it can be focussed e.g. on transmission in classrooms or how aerosols interact in the air.

All three approaches are computationally quite demanding: incorporating newest data at all times means that the model is always simulating, the approach by Keller&Hu creates numerous different

---

3) **Networks and Experimental Design (30 points)**

a) Explain the notion of centrality in networks. Several centrality metrics have high correlation coefficient values. Are these linked to common underlying dimensions like functionality or ethical considerations? Explain how. (~300 words, 10 points)

Note: The topics related to this question were discussed in Lecture 5 and the paper,

Jafino, B. A., Kwakkel, J., & Verbraeck, A. (2020). Transport network criticality metrics: a comparative analysis and a guideline for selection. Transport Reviews, 40(2), 241-264

Centrality metrics can be used to calculate the relative importance of nodes in a network, with this 'importance' being dependent on the values of the decision makers setting the metrics. The three main measures of centrality as examined in Lecture 5 of this course were degree centrality, closeness centrality and betweenness centrality.

So why do centrality metrics have high correlation coefficient values? In the Jafino et al (2020) paper this was due to centrality metrics tending to return high values for the same few nodes/links in the network. This is likely due to the fact that these metrics share common properties like having a high number of neighbours increasing the likelihood of also having a high closeness centrality, for example.

The underlying functional and ethical dimensions influence these in a number of ways:

- Degree centrality and connectivity: Jafino et al defined functional connectivity as 'the availability of connections among all locations of interest in a city' (Jafino et al, 2020, pp 248). As degree centrality is a measure of the weighted number of neighbours of a given node, it will be inherent to measures of connectivity in a network. However, it is worth noting that degree centrality doesn't inherently capture cascade effects (like how many neighbours a node's neighbours have), making it somewhat less realistic as a functional measure of connectivity.
- Closeness centrality and accessibility: Jafino et al defined accessibility as a functional dimension in terms of the ease with which network users can move between nodes in different locations. Correspondingly, closeness centrality is a measure of how short the 'distance' between a node and all other nodes in a network (mathematically, it is the inverse of the sum of the shortest paths for all nodes). If a node has high closeness centrality, then users will be able to reach other nodes from that point more quickly, making that node inherently more accessible.
- Betweenness centrality and traffic flow: Traffic flow here can be assumed to be a proxy for the ethical dimensions (utilitarian and egalitarian), where prioritising links with high traffic flow might indicate utilitarian preferences, while prioritising low traffic flow indicates more egalitarian approaches. Betweenness centrality, as a measure of the extent to which a node contains some number of shortest paths through a network. Preferencing nodes with a higher betweenness centrality might be a feature of metrics with a utilitarian ethical dimension, while preferencing nodes with lower betweenness centrality might indicate egalitarian approaches (as you are inherently focusing on underserved nodes/areas that connect the disconnected).

Case description: In India, every year flooding results in hundreds of kilometres of road closures. Consider the officials from the road transportation sector are approaching you to understand which roads to reinforce to withstand the impacts of flooding. Each road reinforcement has a cost attached to it. Their goal is to identify which roads to reinforce in order to minimise the impact of road closures on the daily life of citizens. They have a limited budget for reinforcement. Assume you have access to a simulation model that gives you an estimate of the *impacts* of every road closure in the country. You can also define additional impacts if you want.

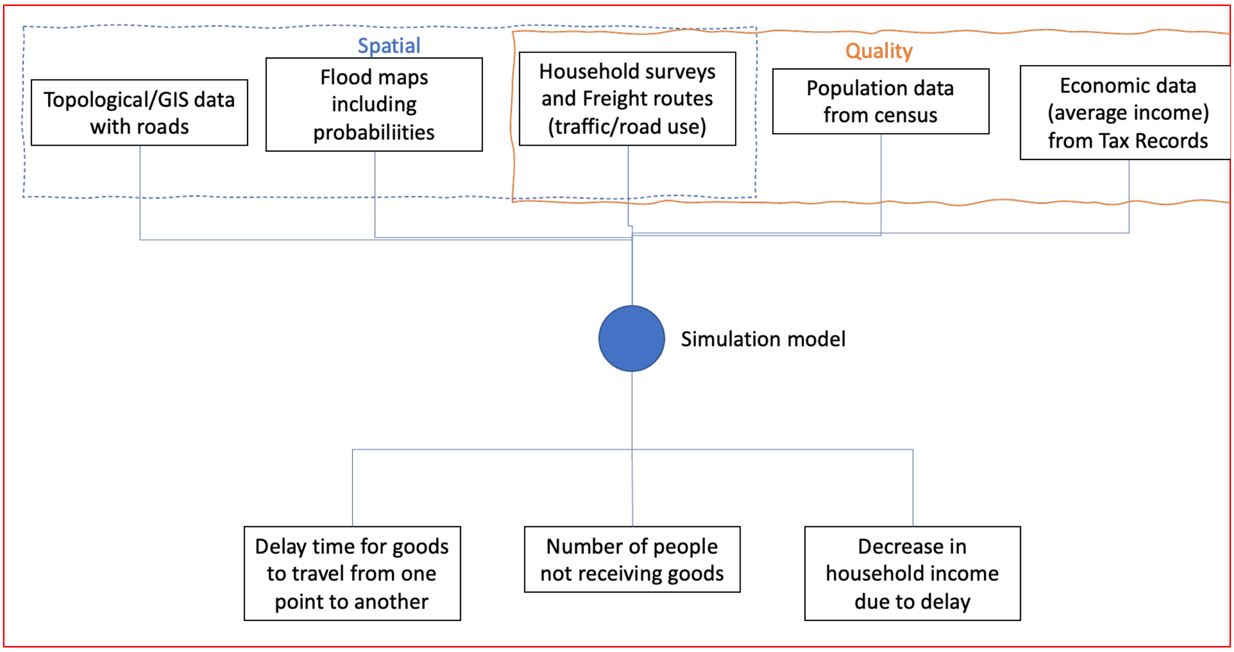
Impact1 = delay in goods reaching a region.
Impact2 = number of people not receiving goods in a region.
Goods = Food, medicine, clothes, essentials, etc.

Note: The topics related to questions 3b) and 3c) were discussed in Lecture 6 and the paper,

He, Y., Thies, S., Avner, P., & Rentschler, J. (2020). The Impact of Flooding on Urban Transit and Accessibility

You may look at Fig 7. in the paper for inspiration on a high-level block diagrams of analysis.

b) Read the case description above. Use the theory on *fundamentals of network criticality analysis* to sketch a high-level block diagram of your analysis so you can present it to the road transportation sector. You can also use a pen and paper and paste the image here. (10 points)


Spatial
Quality
Topological/GIS data with roads
Flood maps including probabiliities
Household surveys and Freight routes (traffic/road use)
Population data from census
Economic data (average income) from Tax Records
Simulation model
Delay time for goods to travel from one point to another
Number of people not receiving goods
Decrease in household income due to delay

c) For the case description above, specify and describe the input and output (decision) variables, constraints, objective functions if any, and the structure of solutions (~ 200 words, 10 points)

In this simulation, you are trying to identify the most cost-effective means of upgrading the transport network to reduce vulnerability to floods. The scenarios will compare impacts of upgrading one road link at a time.

Fundamental of analysis	**Description**
Decision variables	Input: Top layer of the above diagram Output: Bottom layer of the above diagram All alternative for the system should be represented in terms of their delay time, number of people not receiving goods, and impact on household income.

Constraints	Available budget – this informs the number of nodes which can be upgraded in the system.
Objective function	The simulation should contain functions which aim to: - Minimise delay time - Minimise number of people not receiving goods - Minimise the decrease in household income
Solution	The solution presented will contain a ranked list of road segments which should be upgraded to reduce vulnerability to flooding based on the metrics described in 3b.

An exam file without honours pledge will not be graded.

Your honours pledge HERE.