Quicker Quality Scan:

A Multi-Variable Evaluation and Diagnosis Method for Vehicle-Actuated Traffic Signal Controllers

M.M.C.J. Machielsen

TU Delft

Vialis

Quicker Quality Scan

A Multi-Variable Evaluation and Diagnosis Method for Vehicle-Actuated Traffic Signal Controllers

Houten, Monday, 20 May 2019
Version 3.1

To obtain the degree of *Master of Science* in *Civil Engineering* at Delft University of Technology, to be defended publicly on Wednesday 29 May 2019 at 11.00h.

Author	**M.M.C.J. (Martijn) Machielsen, BBE** mmcj.machielsen@outlook.com 4490940	
Graduation Committee	**Prof.dr.ir. S.P. Hoogendoorn**	Delft University of Technology
	Dr.ir. A. Hegyi	Delft University of Technology
	Dr.ir. A.M. Salomons	Delft University of Technology
	Dr. M.T.J. Spaan	Delft University of Technology
	Drs. G. Stern	VIALIS

An electronic version of this thesis is available at https://repository.tudelft.nl/.

Preface

Over the past few months, every time I drove to Delft, I could not help myself but counting down the number of times I will drive past Breda and Rotterdam, and I have to stop at the traffic signals at Kruithuisplein, on my way to TU Delft. I counted the number of times in anticipation – every time, one step closer to obtaining the degree of *Master of Science* in *Civil Engineering* – but also with some sort of sadness – the journey I made in just under four years is about to end. I have to come to think about that, to let it sink in, and to look back at my TU Delft journey, every time I was standing in traffic jams on the motorway, or waiting for a red light – which both happened more often than I want to admit. Along the way, I have learned a lot, mastered all kinds of new skills, and met many new and very interesting people and friends. Even though I doubted a long time about starting my education at TU Delft four years ago, I am truly very happy that I did. And now, with this Master thesis, my time at TU Delft will be formally concluded, and this great journey will come to an end. But before presenting my thesis, I would like to thank the following people, who helped me, and supported me in different ways to complete my Master thesis project.

First of all, I would like to give special thanks to the members of my committee. Andreas Hegyi, and Maria Salomons, as daily supervisors from TU Delft, with the questions they asked me, and the feedback, tips, and information they provided me with, they really helped me improving the quality of this thesis. Moreover, I would like to thank them for their support, and understanding during the whole process, in which I combined my Master thesis project with an Additional Graduation Work project. I would like to thank Serge Hoogendoorn, for supervising the committee, and providing me with feedback and comments, and providing me with new ideas on things to investigate further in this thesis. I would like to thank Matthijs Spaan, for the interesting insights, which I would not have gotten elsewhere. And from VIALIS, I would like to thank George Stern for his continuous support as well. His questions, feedback, information, and so on, triggered me to keep improving my thesis. I am thankful to say that it was my pleasure to work with all of the committee.

Secondly, I would like to thank my colleagues at VIALIS for creating a nice working environment, and the help they gave me along the way, in particular the colleagues from Verkeerskunde. I would like to give special thanks to both George Stern, and Jeroen Hakvoort for offering me the chance to do both my Master thesis project, and Additional Graduation Work project at VIALIS, in particular on subjects that I find very interesting, and that intrigued me from the start. Moreover, I would like to thank them for offering me the chance to start my professional career at VIALIS in July.

Thirdly, I would like to thank my friends, of whom I have met several along the way at TU Delft, for their support. With their help, the many (board) game nights we organised, and the many barbecues filled with home-made ribs, I was able to get my mind of my thesis, and just relax.

Lastly, and I am going to do this in Dutch, wil ik mijn ouders en zus enorm bedanken. Ik weet dat ik in stressvolle periodes niet de leukste thuis ben (anders ook niet, zullen ze nu lachend zeggen), maar ondanks dat, bleven jullie achter mij staan, mij steunen en in mij geloven. Ik weet dat jullie hoe dan ook trots op mij zijn, maar ik wil jullie, hier, op deze plaats, toch echt van harte bedanken voor jullie support, voor jullie geloof in mij de afgelopen jaren, voor alles. Ik zal jullie daarvoor, en dus ook voor nog veel meer, meer om hier te benoemen, altijd dankbaar zijn.

M.M.C.J. Machielsen
Rijen, 19 May 2019

Summary

To evaluate traffic signal controllers, and vehicle-actuated traffic signal controllers in particular, in terms of how they are performing with respect to the road authority's policies on traffic flow and accessibility, traffic safety, and environmental factors, several methods are developed in practice. For instance, VIALIS uses the Instant Quality Scan (IQS), and Quick Quality Scan (QQS) methods, which use a rating instrument, the so-called BI-tool (*Beoordeling Instrument tool*), to check whether the road authority's policy demands are met, given a set of various performance indicators, e.g. delay, queue length, etc. Based on the BI-tool results, the traffic engineer tries to find the cause of any performance issues, diagnose the problem, and propose countermeasures to mitigate them. This evaluation assesses performance indicators independent of each other. It turns out that in scientific literature, the assessment of individual performance indicators is still common practice as well. Moreover, the literature shows that only a limited number of studies focused on how the traffic signal control performance could be improved, rather than solely improvements on the investigated method, algorithm, system, etc. For instance, Bullock & Day (2009) discussed the development of a data collection method that can be used to evaluate the traffic performance of traffic signal controllers. Lavrenz, Day, Smith, Sturdevant, & Bullock (2016) stated that a periodic traffic performance evaluation of traffic signal controllers improves the traffic performance, although they did not provide a framework to do so. Radivojevic & Stevanovic (2017b) did provide a framework, although they did not consider any countermeasures to improve the traffic performance. Methods similar to IQS and QQS evaluations, including the BI-tool are not found in literature. Instead, it turns out that the BI-tool is a decision support system. Decision support systems are used in various fields, such as medicine, etc., though with little to no attention to traffic signal controller evaluation decision support systems: e.g. Moalla, Elkosantini, & Darmoul (2013) focused on the development of traffic signal control algorithm based on a decision support system, including an evaluation component as a trigger for actions, though without formally assessing the traffic signal controller, nor diagnosing the causes of insufficient performance. This implies that a functional, and integral evaluation and diagnosis method for traffic signal controllers, based on a multi-variable assessment, is currently lacking. Therefore, the objective of this thesis is to develop, and present such an integral method, which detects inefficiencies in terms of traffic performance functioning, scores the vehicle-actuated traffic signal controller, diagnoses the cause of the detected inefficiency, and propose countermeasures to improve the traffic performance functioning of the vehicle-actuated traffic signal controller, based on a multi-variable assessment.

The few scientific studies that did focus on the integral evaluation of traffic signal controllers, as listed above, are used as starting point for this development, alongside the current practice at VIALIS. The resulting integral traffic signal controller evaluation and diagnosis method consists of a five-step process, which basically represent a more elaborate BI-tool, and is therefore considered as decision support system as well, though without meeting all formal requirements (e.g. a user interface is not developed in this thesis):

1. *Selection of performance indicators*: some of the key aspects of this selection are that (multi-variable) performance indicators should be selected that (a) offer a complete overview of the traffic performance, (b) can be used to diagnose problems, and (c) can be defined in terms of a reference performance (the computed performance that is expected, and considered as good performance). The selection of multi-variable performance indicators, as tested in this thesis, is based on the analysis of various IQS and QQS evaluation reports of VIALIS (n.d. [a]), yielding four multi-variable performance indicators: (i) degree of saturation, (ii) delay, (iii) phase failure (ending a green phase without fully serving the queue), and (iv) queue length.
2. *Calibration*: the calibration is used to distinguish the inaccuracies of the reference performance models (as used to compute the reference performance) from inefficiencies of the traffic signal controller performance, by defining an error term, and bandwidth. The calibration uses data of an (assumed) problem-free traffic

signal controller. The calibration is performed exclusively for the reference performance models of multi-variable performance indicators that include such inaccuracies, in this case delay, and queue length.

3. *Inefficiency detector*: those periods with inefficiencies (a deviation of the generated performance (measured performance) from the reference performance) are detected. The module describes a five-step procedure in which (i) a list of analysis periods with inefficient performance, (ii) performance statistics, (iii) inefficiency ratios (a measure for the time during which the performance was inefficient), and (iv) evaluation scores (a score for the efficiency of the controller on a scale from 1 (very bad) to 10 (excellent)) are generated.
4. *Diagnosis module*: the problems that caused the detected inefficiencies are diagnosed by comparing the generated performance with examples of inefficient performance from a database. In this thesis, the tested database consists of simulation data of the tested signalised intersection. Also, in the diagnosis module, it is checked whether the proposed countermeasures were effective, by implementing the proposed countermeasures (intermediate optimisation of the controller), and re-running the inefficiency detector, and diagnosis module. That way, the integral method is tested as a cyclic process, which enables the diagnosis of multiple problems for the same traffic signal controller.
5. *Optimisation*: the diagnosis module results in an optimised traffic signal controller, which can be used as input for a re-calibration, as the (assumed) problem-free traffic signal controller, or as a starting point to re-do the whole evaluation and diagnosis process (re-evaluation) after several years.

Before testing the integral evaluation and diagnosis method in a case study, it is investigated what the accuracy is of the reference performance models for delay, and queue length, given three different arrival patterns: (i) uniform, (ii) random, and (iii) platoon arrivals. It is concluded that the calibration of the reference performance models is sufficient to account for inaccuracies caused by using random arrivals, instead of uniform, and platoon arrivals as assumed in the reference performance models.

The testing of the integral evaluation and diagnosis method in a case study, is done by deliberately implementing problems in a traffic signal controller, and then assess these faulty controllers. To limit the number of potential problems that can be tested, four problems are selected based the analysis of various IQS and QQS evaluation reports of Vialis (n.d. [a]): (i) incorrect gap times of long detector, (ii) incorrect maximum green time settings, (iii) deactivated alternative realisations, and (iv) inadequate geometric intersection design. The last problem is included as an implicit problem: if the inefficiencies cannot be diagnosed as one of the other three problems, the problem is assumed to be inadequate geometric intersection design.

The case study is performed as a half-blind case study: for several tested alternatives, the implemented problems are known beforehand to the author of this thesis, and in several alternatives, this is unknown. The testing of the integral traffic signal controller evaluation and diagnosis method showed that the method is able to detect inefficiencies, and assign them to the problem(s) that caused the detected inefficiencies. It must be noted that this included the assumption that the problem is inadequate geometric intersection design if the inefficiencies cannot be diagnosed as one of the other three problems. This is in particularly relevant when assessing the base case alternative (the assumed problem-free controller), which already showed some inefficiencies.

Therefore, the conclusion is that the inefficiencies on the traffic performance functioning, as tested in this thesis, of a vehicle-actuated traffic signal controller can be indicated, and mitigated by applying the integral evaluation and diagnosis method presented in this thesis. Although the presented method is not perfect yet (e.g. the method is not yet a formal decision support system), its potential is clear. For future work, it is therefore recommended to formalise, and automate the presented method, as well as its components. Furthermore, it is recommended to test the method on different types of signalised intersections, and investigate the use of data measured in practice, also with respect to (filling) the database with examples of inefficient performance.

SAMENVATTING

Het evalueren en beoordelen van verkeersregelinstallaties (VRI's), en voertuigafhankelijke VRI's (VA-VRI's) in het bijzonder, ten opzichte van de beleidswensen van de wegbeheerder, in termen van doorstromings-, verkeersveiligheids- en leefomgevingseffecten, wordt gedaan aan de hand van beoordelingsmethodieken vanuit de praktijk. Voorbeelden daarvan zijn de Instant Quality Scan (IQS) en Quick Quality Scan (QQS) zoals ontwikkeld en gebruikt door VIALIS. In de IQS- en QQS-evaluaties wordt gebruik gemaakt van een Beoordeling Instrument tool (BI-tool) om te controleren of de performance van de VRI overeenkomt met het beleid van de wegbeheerder, bijvoorbeeld ten aanzien van wacht- en verliestijden, wachtrijlengtes, etc. De verkeerskundige gebruikt de resultaten van de BI-tool om de exacte performance issues te vinden, de oorzaak te vinden en maatregelen voor te stellen. De BI-tool beoordeling gebeurt op basis van op zichzelf staande performance indicatoren – de relaties tussen performance indicatoren worden uitsluitend door de verkeerskundige gemaakt. De wetenschappelijke literatuur laat eenzelfde beeld zien, namelijk dat de relaties niet expliciet beoordeeld worden. Bovendien, in de literatuur is slechts weinig aandacht voor de beoordeling van de performance van een VRI op zichzelf: doorgaans worden de methoden, systemen, algoritmes, etc. van en voor VRI's beoordeeld en waar mogelijk verbeterd. Zo beschrijven Bullock & Day (2009) een methode om VRI-data te verzamelen, die gebruikt kunnen worden om de VRI te beoordelen, Lavrenz, Day, Smith, Sturdevant, & Bullock (2016) concludeerden dat een periodieke beoordeling van de VRI noodzakelijk is om de performance op peil te houden, zonder dat ze daarbij een beoordelingsmethode voorschrijven, en hoewel Radivojevic & Stevanovic (2017b) wel een dergelijke methode ontwikkelden, stellen zij geen maatregelen voor om de performance te verbeteren, indien nodig. Methoden vergelijkbaar met de IQS en QQS van VIALIS, al dan niet inclusief de BI-tool, zijn niet gevonden in de literatuur. Desalniettemin is gevonden dat de BI-tool in het bijzonder een beslissingsondersteunend systeem is. Dergelijke systemen worden veel gebruikt in de praktijk, onder meer in de medische zorg, etc., maar wederom zelden in de beoordeling van VRI's: bijvoorbeeld Moalla, Elkosantini, & Darmoul (2013) gebruikten een beslissingsondersteunend systeem als VRI-algoritme, inclusief een zekere mate van performance beoordeling, doch met noch een formele beoordeling, noch een diagnose van de oorzaken van eventuele problemen. De literatuurreview toont dus aan dat een integrale beoordelings- en diagnosemethodiek voor VA-VRI's, gebaseerd op een multivariabele beoordeling, momenteel niet bestaat. Daarom is het doel van dit onderzoek om een dergelijke methodiek te ontwikkelen en te presenteren. De methodiek moet inefficiënties van de performance kunnen detecteren, de oorzaken kunnen diagnosticeren, en maatregelen voorstellen om deze te mitigeren.

De weinige wetenschappelijk studies die relateren aan dit onderwerp, alsook de huidige praktijk bij VIALIS met IQS- en QQS-evaluaties zijn gebruikt als startpunt. Dit resulteerde in wat in feite een uitgebreide BI-tool is. De methode is daarom ook geclassificeerd als een beslissingsondersteunend systeem, doch zonder te voldoen aan de formele voorwaarden van een dergelijk systeem (zo is de user interface niet ontwikkeld in dit onderzoek, bijvoorbeeld). De methode bestaat uit vijf stappen:

1. *Selectie van performance indicatoren*: voor de selectie van de (multivariabele) performance indicatoren is het van belang dat de geselecteerde (multivariabele) performance indicatoren (a) een compleet overzicht van de performance van de VRI geven, (b) gebruikt kunnen worden om problemen te diagnosticeren en (c) gedefinieerd kunnen worden als een referentieperformance (de berekende performance die verwacht wordt en die gezien wordt als goede performance). Voor deze thesis is deze selectie gebaseerd op de analyse van verscheidene IQS- en QQS-evaluaties van VIALIS (n.d. [a]), hetgeen resulteerde in de volgende vier multivariabele performance indicatoren: (i) verzadigingsgraad, (ii) vertraging (verliestijd), (iii) overstaan (phase failures; de groenfase beëindigen zonder de volledige wachtrij afgewikkeld te hebben) en (iv) wachtrijlengte.
2. *Kalibratie*: de kalibratie dient om onnauwkeurigheden van de referentieperformancemodellen (rekenmodellen om de referentieperformance te bepalen) te onderscheiden van VRI-inefficiënties, door een foutterm en

bandbreedte te definiëren. Daartoe is data nodig van een (aangenomen) "perfecte" VRI, zonder problemen. De kalibratie is uitsluitend relevant voor de referentieperformancemodellen met mogelijke onnauwkeurigheden. In deze thesis betreft dit de modellen voor de vertraging en wachtrijlengte.
3. *Inefficiëntie detector*: de momenten met inefficiënties (een afwijkende gegeneerde performance, zoals gemeten, ten opzichte van de referentieperformance worden gedetecteerd, waarna (i) een lijst met analyseperioden met inefficiënties, (ii) de performance statistieken, (iii) de inefficiëntieratio (een maat voor de tijd gedurende de performance was inefficiënt) en (iv) de evaluatiescores (een score voor de efficiëntie van de VRI op een schaal van 1 (zeer slecht) tot 10 (uitstekend)) worden bepaald.
4. *Diagnose module*: de oorzaken van de inefficiënties worden gediagnosticeerd als zijnde problemen, door de gegeneerde performance te vergelijken met voorbeelden van inefficiënte performances per probleem. Deze voorbeelden komen uit een database, die in deze thesis bestaat uit simulatiedata voor de onderzochte kruising. Daarnaast worden maatregelen voorgesteld en wordt gecontroleerd of deze effectief waren, door ze te implementeren en de inefficiëntie detector en diagnose module opnieuw te doen. Dit maakt de algemene methode een cyclisch proces die de diagnose van meerdere problemen voor eenzelfde VRI mogelijk maakt.
5. *Optimalisatie*: de diagnose module levert een geoptimaliseerde VRI op, die gebruikt kan worden als (aangenomen) "perfecte" VRI voor herkalibratie, of voor herevaluatie na enkele jaren.

Voordat de hierboven beschreven methode getest kan worden, is de nauwkeurigheid van de referentieperformancemodellen voor vertraging en wachtrijlengte onderzocht, gegeven drie aankomstenpatronen: (i) uniform, (ii) random en (iii) peloton aankomsten. De conclusie is dat de kalibratie van de referentieperformancemodellen afdoende rekening houden met onnauwkeurigheden door het gebruik van random aankomsten, in plaats van de in de modellen aangenomen uniforme en peloton aankomsten.

Middels een casestudie is de methode getest. De casestudie gebruikt opzettelijk inefficiënte VRI's door problemen opzettelijk te implementeren. Om het aantal te testen problemen te beperken, zijn vier problemen geselecteerd op basis van de analyse van verscheidene IQS- en QQS-evaluaties van Vialis (n.d. [a]): (i) incorrecte hiaattijden van de lange lus detector, (ii) incorrecte maximum groentijden, (iii) gedeactiveerde alternatieve realisaties en (iv) inadequaat kruispuntontwerp. Inadequaat kruispunt ontwerp is dan een impliciet probleem: als de inefficiënties niet toegeschreven kunnen worden aan een of meerdere van de andere problemen, wordt aangenomen dat inadequaat kruispunt het probleem is.

De casestudie is gedaan als een halfblinde test: van sommige alternatieven was bij de auteur van deze thesis bekend wat de geïmplementeerde problemen waren, en van sommige alternatieven was dit onbekend. De test van de methode heeft aangetoond dat de voorgestelde integrale beoordelings- en diagnosemethodiek in staat is om inefficiënties te detecteren, ze toe te wijzen aan de correcte oorzaak en maatregelen voor te stellen om deze gediagnosticeerde problemen op te lossen. Daarbij dient wel opgemerkt te worden dat de hierboven genoemde aanname met betrekking tot het probleem van inadequaat kruispunt ontwerp hoofdzakelijk van belang is gebleken in het basecase alternatief (de aangenomen "perfecte" VRI) waarin reeds een zekere mate van inefficiënte performance werd waargenomen.

Daarom is de conclusie dat de performanceproblemen van een VA-VRI, zoals getest in dit onderzoek, met succes gedetecteerd kunnen worden als inefficiënties, gediagnosticeerd kunnen worden als problemen, en gemitigeerd kunnen worden door het voorstellen van bijbehorende maatregelen door gebruik te maken van de in deze thesis voorgestelde integrale beoordelings- en diagnosemethodiek voor VA-VRI's. Ofschoon de methodiek nog niet perfect is (zo is de methodiek nog geen formeel beslissingsondersteunend systeem), toont het wel zijn potentieel aan. Daarom wordt voorgesteld voor toekomstig onderzoek om de methodiek en zijn componenten te formaliseren en te automatiseren. Bovendien is het een aanbeveling om de methodiek uitvoeriger te testen op meerder kruispunten en daarbij ook gebruik te maken van praktijk data. Dit relateert eveneens aan (het vullen van) de gebruikte database met voorbeelden van inefficiënte performance.

Table of contents

NOTATION AND DEFINITIONS

Symbol	Unit	Math. domain	Description
$\boldsymbol{B}$	-[1]	-[1]	Bandwidth
$\boldsymbol{d}$	s/pce	$\mathbb{R}$	Delay
$\boldsymbol{d}_{\mathbf{ad}}$	s/pce	$\mathbb{R}$	Acceleration-deceleration delay
$\boldsymbol{g}$	-	$\mathbb{R} \in [1,10]$	Evaluation score
$\boldsymbol{H}^{\boldsymbol{Q}}$	m	$\mathbb{R}$	Maximum headway in queue
$\mathbf{IR}$	-	$\mathbb{R}$	Inefficiency ratio
$\boldsymbol{i}$	-	-	Performance indicator index
$\boldsymbol{j}$	-	$\mathbb{N}$	Signal group index
$\boldsymbol{L}$	pce	$\mathbb{R}$	Queue length
$\boldsymbol{L}_{\mathbf{max}}$	m	$\mathbb{R}$	Maximum queue length
$\boldsymbol{n}^{\mathbf{ineff}}$	#	$\mathbb{N}$	Total number of analysis periods with inefficient performance
$\boldsymbol{n}_{\boldsymbol{C}}$	#	$\mathbb{N}$	Number of cycles
$\boldsymbol{n}_{\boldsymbol{\tau}}$	#	$\mathbb{N}$	Total number of analysis periods
$\boldsymbol{P}$	-	$\mathbb{R}$	Proportion of vehicles arriving during effective green
$\boldsymbol{p}$	-[1]	-[1]	Generated performance
$\boldsymbol{p}^{*}$	-[1]	-[1]	Base case generated performance
$\boldsymbol{q}$	pce/h	$\mathbb{R}$	Traffic flow volume
$\boldsymbol{r}$	-[1]	-[1]	Reference performance
$\boldsymbol{s}$	pce/h	$\mathbb{R}$	Saturation flow
$\boldsymbol{T}_{\boldsymbol{A}}$	s	$\mathbb{R}$	Amber time
$\boldsymbol{T}_{\boldsymbol{C}}$	s	$\mathbb{R}$	Cycle time
$\boldsymbol{T}_{\boldsymbol{G}}$	s	$\mathbb{R}$	Green time
$\boldsymbol{T}_{\boldsymbol{G},\mathbf{eff}}$	s	$\mathbb{R}$	Effective green time
$\boldsymbol{T}_{\boldsymbol{R}}$	s	$\mathbb{R}$	Red time
$\boldsymbol{T}_{\boldsymbol{R},\mathbf{gar}}$	s	$\mathbb{R}$	Guaranteed red time
$\boldsymbol{v}^{\boldsymbol{Q}}_{\mathbf{max}}$	km/h	$\mathbb{R}$	Maximum speed in queue
$\boldsymbol{v}^{\boldsymbol{Q}}_{\mathbf{min}}$	km/h	$\mathbb{R}$	Minimum speed in queue
$\boldsymbol{X}$	-[1]	-[1]	Undefined describing variable
$\boldsymbol{x}$	-	$\mathbb{R}$	Degree of saturation
$\boldsymbol{\varepsilon}$	-[1]	-[1]	Error term
$\boldsymbol{\lambda}_{\mathbf{1}}$	s	$\mathbb{R}$	Green time start lag
$\boldsymbol{\lambda}_{\mathbf{2}}$	s	$\mathbb{R}$	Green time end lag, also known as utilised amber time
$\boldsymbol{\sigma}$	-[1]	-[1]	Standard deviation
$\boldsymbol{\tau}$	-	$\mathbb{N}$	Analysis period index
$\boldsymbol{\phi}$	#	$\mathbb{N}$	Phase failure

[1] The unit, and mathematical domain depend on the unit, and mathematical domain of considered multi-variable performance indicator, and its describing variables.

1. Introduction

For decades, traffic signal control has been, and still is a proven and often used method with respect to accessibility, and traffic safety, especially in urbanised regions. Traffic signal control is especially relevant on intersections where a lot of traffic converges, crosses, or changes its direction. Here, a trade-off must be made between flow and accessibility, traffic safety, and environmental factors.

A traffic signal controller controls traffic by separating traffic flows in time, rather than in space only. This means that by using a traffic signal controller, conflicting traffic streams make use of the same infrastructure, though at different moments in time. This implies that the functioning of the traffic signal controller is not only a matter of traffic safety (preventing crashes by separating traffic in time), but also of throughput, and environmental effects, in terms of delays, cycle times, etc. The latter introduces the principle of efficient time management: in the past, the question was when a direction gets its turn, and optimising time management, in terms of total delay (CROW, 2006), while nowadays, it is more about finding an optimal solution for the use of the scarce time, and space at intersections, due to the growth of traffic in the past decennia. Efficient time management affects the evaluation of traffic signal controller as well, because the objective has become to reduce the delays.

The latter introduces the term "delay". In literature, ambiguous definitions are given for delay. The definition of delay used in this research is that delay denotes the additional travel time a road user experiences, due to waiting (stop delay), plus the time one loses to brake (deceleration delay) from, and accelerate (acceleration delay) to the desired or free flow speed, or restricted speed (e.g. when driving in a platoon). In the Dutch literature, this definition corresponds to the term *verliestijd* (CROW, 2006). A more detailed definition of this term, and other terms used in this research, are given in appendix A. If applicable, the relation to the definitions often used in Dutch literature are given as well.

1.1. Research motivation

Road authorities in the Netherlands often request market parties, such as Vialis, to evaluate signalised intersections, and the traffic signal controllers in particular, on their technical performance, and traffic performance. Therefore, several methods have been developed to meet this demand. Two of those methods are the Instant Quality Scan (IQS), and Quick Quality Scan (QQS), see also appendix A. Although the structure of both the IQS, and QQS is identical, the QQS is a more elaborate method than the IQS, implying that an IQS-evaluation is less time-consuming. The objective of these evaluations is to provide road authorities with data, and knowledge on how their traffic signal controllers are functioning. The main focus is on (i) the functioning of the traffic signal control system itself, (ii) the throughput of traffic, whereas modes as motorised traffic, active modes (i.e. bicyclists, pedestrians, etc., see also appendix A), and Public Transport are distinguished, and (iii) traffic safety. Furthermore, several traffic flow characteristics are included as well, such as degree of saturation. The result of the evaluation is, besides a description on how the traffic signal controller is currently functioning, an advice to improve the functioning. This proposal is given in relation to the constraints, provided by the road authority, in terms of policy demands. Although the names of the IQS, and QQS suggests that they are rather quick methods, they can be quite time-consuming, because the traffic engineer has to check each relevant performance indicator, and/or variable manually on the given constraints, even though the data collection, and processing software used in the evaluations produces the values of the performance indicators automatically. The data collection is done using the KWC (*KWaliteitsCentrale* (Quality Centre), a software program developed by Vialis to evaluate signalised intersections, see also appendix A) (Vialis, 2017). The output of the KWC is used by the BI-tool (*Beoordeling Instrument tool*; Rating Instrument Tool, see also appendix A), which is an assessment tool in the KWC in which for several performance indicators, such as oversaturation, and red-light running, static threshold and reference values are given. The BI-tool checks whether, and when these values are exceeded, and thus whether there were performance problems (Vialis, n.d. [a]). The output of the BI-tool is used by the traffic engineer to analyse the phase-log data (data on the detector states, and green/amber/red phases, see Figure 1-1, and appendix A (Vialis, 2017)), and other KWC outputs (e.g. peak diagrams, and graphs, see Figure 1-2) to find the cause of the performance issues found by the BI-tool, which is relevant for which countermeasures that are proposed. It must be noted that the BI-tool only states that there are performance issues. It does not diagnose these issues, which means that the BI-tool does not yet state what is the cause of the found performance issues. That means that the traffic engineer has to diagnose the traffic signal controller by hand. Because the outputs of the KWC, and BI-tool are given as individual performance indicators, the relations between the performance indicators are only implicitly considered as "expert judgement" of the traffic engineer.

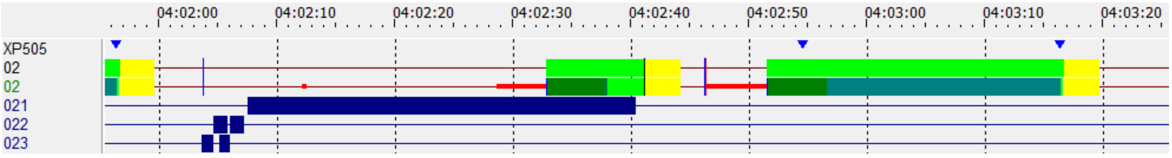

Figure 1-1 | Visualisation of phase-log data in the KWC, with the phase of a signal group (red, green, amber), and the state of detectors (active, blue) (Vialis, n.d. [a]).

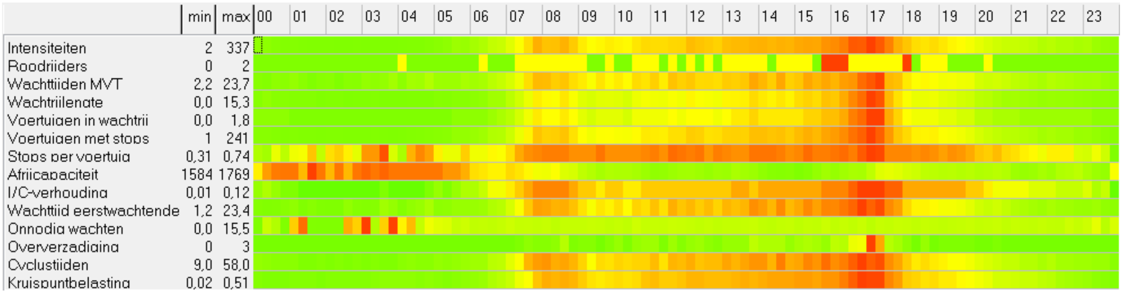

Figure 1-2 | Peak diagram generated by the KWC: a simplified representation of several performance indicators, such as stops, delays, volume-to-capacity ratio, cycle times, etc., per 15 minutes for an average workday. The scale (green-amber-red) denotes the qualitative performance of that performance indicators, whereas green is low, and red is high. Here, it can be seen that the evening peak hour (17.00h) has in general high values for delays (*wachttijden MVT*), queue length (*wachtrijlengte*), stops (i), volume-to-capacity ratio (*I/C-verhouding*), cycle times (*cyclustijden*), etc., and low values for e.g. unnecessary waiting (*onnodig wachten*) (Vialis, n.d. [a]).

The latter introduces the fact that the evaluation in practice are based on individual performance indicators. The literature review showed that in scientific literature, this is also the case, as will be discussed in chapter 2. This means that currently multiple performance indicators are assessed independent of each other, e.g. the cycle time is assessed separately from delay, thus potential relations between performance indicators are not included: multi-variable assessment is only done implicitly included as "expert judgement" of the traffic engineer. In an internal memo at Vialis of (De Leeuw, 2010), it is proposed to formalise multi-variable assessment as a way to better evaluate traffic signal controllers, for instance using the BI-tool. With such an explicit multi-variable assessment, one could assess for instance whether a high delay was "justified": either the delay is the result of oversaturation, and therefore acceptable, or the delay might be caused by problems within the traffic signal controller, which is not acceptable. In other words, in the case of oversaturation, the traffic signal controller is assumed to function at the best of its abilities, and that because of this, the high delays are not the result of the traffic signal controller itself, but rather as a consequence of insufficient geometric intersection design, for instance. In the other case, the delays are indeed the result of inefficient functioning of the traffic signal controller, e.g. by giving an insufficient green time to a signal group. This principle can be applied on other performance indicators, and variables as well. Therefore, such a multi-variable assessment is an interesting aspect to investigate.

Also, the QQS and IQS evaluations are examples of semi-standardised evaluation methods: the general evaluation steps are defined, but not formalised. This introduces the influence of the road authority, and the underlying objective of the evaluation for that road authority: what does the road authority want to know? This results in variations in the way traffic signal controllers are evaluated, as found in examples of QQS and IQS reports, and as stated by traffic engineers at VIALIS. Furthermore, the evaluation itself is, in its current state, very specialised. This means that for every intersection, and for every traffic signal controller, different issues are found, using the phase-logging files. The result is then that although the evaluation process is generic, the evaluation result, and used performance indicators are in all cases specifically related to that particular intersection. This impedes the comparison of the results of one intersection to the results of other intersections. Besides, the evaluations of the traffic signal controller (the system), and the signalised intersection (the infrastructure) are intertwined, as stated traffic engineers. This implies that the current evaluation method is ambiguous in terms of what is exactly assessed, and how this should be done.

1.2. Knowledge gap

As concluded in the previous section, the functional, and integral evaluation of signalised intersections is a semi-standardised method in practice. Also, in literature, not much attention is given to the functional, and integral evaluation of signalised intersections, which implies that such assessment/evaluation methods are currently lacking, as will be discussed more in detail in chapter 2 as well. Although VIALIS has a semi-standardised generic traffic signal control evaluation method, an evaluation of multiple performance indicators with respect to each other is still not considered, at least not formally, and explicitly. At the moment, only isolated assessment is done, while multi-variable, and simultaneous assessment is only performed implicitly as "expert judgement" of the traffic engineer. Formalising this might give better insight in the cause of, for instance, high delays. This might also enable proposing better countermeasures, since it is expected that there is knowledge on the cause of the identified problems. Also, the analysis tool that is currently is used (BI-tool), only states that there are performance issues, but does not identify the cause of these problems, and thus does not diagnose the traffic signal controller. Therefore, the following knowledge gaps are identified:

- A consistent assessment, and evaluation method for traffic signal controllers, including the diagnosis of performance issues;
- Understanding of the interaction between performance indicators, in relation to the evaluation and diagnosis of performance issues of traffic signal controllers;
- Possibilities for a using a formal, and explicit assessment on multiple performance indicators, and variables with respect to each other, as a way to evaluate and diagnose traffic signal controllers.

1.3. Research objective and questions

The objective of this study is to develop, and present an integral evaluation and diagnosis method for traffic signal controllers, including a simultaneous assessment of multiple performance indicators, which detects inefficiencies in terms of traffic performance functioning, scores the traffic signal controller, diagnoses the cause of the detected inefficiency, and propose countermeasures to improve the traffic performance functioning of the traffic signal controller, by gaining understanding in the various performance indicators and their interaction, how these performance indicators can be used to detect inefficiencies, and how the detected inefficiencies relate to certain problems, in order to diagnose the problems and propose countermeasures. The resulting evaluation and diagnosis method is then basically a system which (i) detects the error, (ii) diagnoses the system, and (iii) propose countermeasures to solve the found error.

In order to reach the research objective, the following main research question and corresponding sub-questions are formulated:

How can an integral evaluation and diagnosis method, based on an assessment of multiple performance indicators, indicate inefficiencies on the traffic performance functioning of a traffic signal controller, and propose countermeasures to mitigate those inefficiencies?

- *What are the motives and methods used for the evaluation of traffic signal controllers?*
- *Which performance indicators must be selected?*
  - What are the traffic signal controller issues and corresponding inefficiencies?
  - Which performance indicators are used to diagnose these traffic signal controller problems?
  - How to create multi-variable performance indicators from single-value performance indicators?
- *How can the multi-variable performance indicators be used to detect inefficiencies?*
  - What is the definition of an inefficiency in relation to the multi-variable performance indicators?
  - What distinguishes inefficiencies in the traffic signal controller from inaccuracies in the inefficiency detection method?
  - How do inefficiencies present themselves via the multi-variable performance indicators?
- *How can detected inefficiencies be assigned to traffic signal controller problems?*
  - How do the traffic signal controller problems present themselves as inefficiencies?
- *What are the steps of an integral evaluation and diagnosis method for traffic signal controllers?*
  - Which evaluation method can be used?

1.4. Research scope

The evaluation and diagnosis method developed in this thesis, is designed for isolated signalised intersections, thus no signalised intersections as part of a network. Also, only vehicle-actuated traffic signal controllers are considered. This corresponds to the common Dutch practice regarding traffic signal control, because fixed time traffic signal controllers are not usually applied. The evaluation method should include multiple of performance indicators, and variables (which performance indicators, and variables are to be considered, is part of the research). Still, the research will focus on traffic performance variables, implying that technical inefficiencies, and/or malfunctions are not part of this research.

1.5. Thesis outline

The thesis is structured based on the aforementioned research questions. The thesis starts with an introduction in chapter 1. In chapter 2, the literature review is discussed, where the question is answered regarding what the motives and methods are for the evaluation of traffic signal controllers, both in literature, and in practice. Also, this chapter discusses literature on decision support systems as part of the literature review on evaluation methods. Based on these finding, the integral traffic signal controller evaluation and diagnosis method is presented in chapter 3. This chapter discusses the definitions that apply regarding inefficiency detection and diagnosis, and relevant decision support system components. Also, the chapter discusses the integral evaluation and diagnosis method step-by-step, including the development and discussion of the methods to detect inefficiencies, and diagnose the problem based on the detected inefficiencies. Next, in chapter 4, the development of the method is discussed in more detail, thereby focusing on how the method is tested in this thesis, in terms of scope (selection of problems, and multi-variable performance indicators, etc.), additional definitions, and an assessment of the accuracy of the used computational models, using a simulation study in VISSIM. In chapter 5, the application of the integral evaluation and diagnosis method is tested in a case study, also using a simulation study in VISSIM. Lastly, chapter 6 concludes the thesis by explicitly answering the research questions as listed above, and proposing recommendations regarding further implementation of the method, and for future work.

2. Traffic signal control evaluation in literature

Traffic signal controllers are evaluated for various reasons. In literature, this is not different. Indeed, the reasons to evaluate traffic signal controllers range from testing a new traffic signal control algorithm, to the evaluation of traffic signal controllers as part of a maintenance cycle. Each reason includes its own performance indicators, and uses its own methods. The results in terms of proposed countermeasures differ per evaluation reason as well. However, as will be discussed below, these countermeasures are usually related to the algorithm, model, method, etc. that is evaluated, rather than the traffic performance and technical functioning of the traffic signal controller. Consequently, the methods to evaluate traffic signal controller are related to the objective of the evaluation. Nonetheless, some general methods are found, including the use of expert systems, and decision support systems.

All these reasons, performance indicators, countermeasures, and methods are discussed in this chapter. That way, the question is answered what the motives for, and methods used in traffic signal control evaluations are. First, the evaluation motives are discussed, followed by the methods. The latter relates to expert systems, and decision support systems as well. The chapter concludes with some general remarks on the aforementioned topics in relation to the traffic signal control evaluation and diagnosis method presented in this thesis.

2.1. Evaluation motives

As stated, there are multiple motives, and reasons to evaluated signalised intersections, and traffic signal controllers in particular. Based on the literature review discussed below, it is found that the main reasons are as follows:

i. Evaluation of (new) traffic signal control aspects, such as algorithms, systems, models, etc.;
ii. Traffic signal control performance evaluation;
iii. Traffic signal control evaluation as part of periodic maintenance.

Of these three main reasons, the first one is most commonly found in literature.

2.1.1. Evaluation of new traffic signal control aspects

The evaluation of (new) traffic signal control algorithms, systems, models, etc. focus on assessing whether the newly proposed algorithms, systems, models, etc. perform as expected. This could be done using a field test evaluation. In a field test evaluation, the evaluation focuses on the performance of a specific traffic signal control algorithm or program, often in a FOT (Field Operation Test, see also appendix A) environment. The considered performance indicators are based on the algorithm that is evaluated. In literature, most of these evaluations focus on the assessment of urban network control algorithms, whereas consequently network-wide performance indicators are used. For instance, the evaluation of the Split Cycle, and Offset Optimisation Technique (SCOOT, see also appendix A) algorithm, included performance indicators that were related to the objective of SCOOT: minimise congestion by optimising the green split, cycle times, and offset of multiple signalised intersection in a region, based on the departure of an upstream intersection (Martin & Hockaday, 1995). This was evaluated using a FOT in Anaheim, California, U.S.A., by comparing a group of signalised intersections with the algorithm, with a group of signalised intersections without, on network-wide travel times and delays, stops, and traffic flow volume (Moore II, Mattingley, MacCarley, & McNally, 2005). In a similar way, Kosmatopoulos, et al. (2006) assessed the functioning of another urban network control algorithm in three different regions in a FOT before-after study, with the average mean speed in the network, as function of the total time spent, and total distance travelled, as performance indicator. The implementation of the Sydney Coordinated Adaptive Traffic System (SCATS, see also appendix A) algorithm – an algorithm comparable to SCOOT in terms of its objective – in Las Vegas, Nevada, U.S.A. was also assessed in a FOT before-after study, on travel times, and stops per route (Tian, Ohene, & Hu, 2011). However, in all these field test evaluations, the evaluation ends with describing the results of the performance assessment, implying that no countermeasures are proposed to potentially the improve the assessed urban network control algorithm, other than further finetuning of algorithm parameters.

Another example of traffic signal control evaluation in relation to how a specific method functions, is the evaluation of data source and processing methods at signalised intersections. Here, the other ways to collect and process data are assessed. For instance, Hu, Fontaine, Park, & Ma (2016) assessed how private-sector probe data can be used for a corridor of adaptive traffic signal controllers. Their performance indicators were corridor delay, and travel time reliability. In a similar way, the use of other data sources, including Bluetooth data, is evaluated when assessing an urban network control algorithm on travel time and delays, and queue lengths, again on a network-level (Lidbe, Tedla, Hainen, & Jones Jr., 2017). Another type of data source and processing evaluation is performed by Huang, et al. (2018), who evaluated adaptive traffic signal control schemes by introducing intelligent performance indicators, with the objective to improve the data quality of existing automated performance indicators, such as traffic flow volume, cycle times, and degree of saturation. In all three of the aforementioned cases, a FOT-study was used. Also, given the scope of these evaluations, they were not able to identify traffic signal control related improvements for the system. Indeed, the countermeasures proposed in these publications focus on the data collection method, and/or analysis system rather than the traffic signal controller itself.

The development of new systems is also used as motive to evaluate traffic signal controllers. E.g., Wu, Hunter, Lee, & Rodgers (2011) proposed a method to describe the performance of a traffic signal controller using the two-fluid model theory, which assumes that travel time, and stop delay are related to each other. On the other hand, Balke, Charara, & Parker (2005) developed a new performance indicator scoring system, including the hardware needed to collect the data. The performance indicator scores can be used to identify potential issues in the controller in relation to several performance indicators, such as cycle times, green, amber, and red phase realisations, time to service (time-to-green), queue service time (time needed to clear the queue from the start of the green phase of a signal group until the queue is dispersed), phase failure rate (at the start of the red phase, the queue from the preceding green phase is not yet fully dispersed, see also appendix A), and red-light running. In a similar way, Balasha & Toledo (2015) developed a mesoscopic traffic simulation model to evaluate signalised intersections, and to propose countermeasures to improve the traffic signal controller, though only considering delay as traffic signal control related performance indicator. Also, a macroscopic model for traffic signal controllers has been developed for the German state of Bavaria with the objective to identify, and rank signalised intersections with (potentially) a poor performance in terms of delays, both on a network level, and for local intersections Wünsch, Bölling, Von Dobschütz, & Mieth (2015). The aforementioned evaluations used real-life data to calibrate, and assess the proposed models.

The development of a new performance indicator is also an example of system development-oriented traffic signal control evaluation. Wood, Palmer, & Bretherton (1994) proposed wasted capacity as performance indicator, defined as the product of saturation flow (capacity during the green phase, see also appendix A), and blocked time during a green phase, due to spillback of a downstream intersection. Other performance indicators that were developed, though without including countermeasures to improve the performance of the assessed signalised intersection(s), are given, for instance, by Teply (1993), who proposed a new performance indicator: overload factor, denoting the probability of overload at signalised intersections, based on the degree of saturation. Matsoukis (2005) proposed a new integral performance indicator, denoted as the weighted sum of the scores of the performance indicators traffic safety (e.g., number of crashes), queue length, delay, blockage, cycle time, and unused pedestrian (green) phases. The result is a score similar to a Level Of Service (LOS, see also appendix A) for a signalised intersection, which can be used to prioritise intersections in term of where actions might be needed to improve the functioning. Bullock & Day (2009) proposed another performance indicator that integrates volume-to-capacity ratio, and degree of saturation, to assess the performance of urban traffic signal controllers. However, they did not propose countermeasures, other than relocating capacity in the urban network. Dakic, Mladenović, Stevanovic, & Zlatkovic (2018) developed approach delay, and average arrivals on green ratio, thereby integrating arrivals on green, platoon ratio (a measure for how platoons arrive, see also appendix A), cycle times, green times, traffic flow volume, delay, and queue length. Although the assessed their new performance indicators on how they could be used in evaluating traffic signal controllers, they did not mention potential improvements on behalf of the traffic signal controller, but rather on the performance indicator models themselves. The same goes for the model proposed by So, Stevanovic, & Koonce (2016), who proposed an automated method that collects data to estimate the parameters for a newly formulated volume-delay function (an analytical function that describes the impact traffic flow volume has on delay at a macroscopic level).

2.1.2. Evaluating traffic signal control performance

As introduced, there is another motive to evaluate traffic signal controllers, namely evaluating with the objective to assess the traffic performance, and propose countermeasures to improve this performance. It is found that this is not studied much in scientific literature. However, one study in particular stands out. Lavrenz, Day, Smith, Sturdevant, & Bullock (2016) stated that the traffic performance, and signal timing degrade over the years, due to a growth of traffic flow volumes. Because this affects various aspects of the traffic performance, such as travel time (reliability), they proposed a traffic signal control evaluation system. Their system works towards a rather basic countermeasure: updating timing plans in terms of the green, amber, and red phase time settings, cycle time settings, and block sequence (sequence of conflicting green phases, given as phases, or blocks, see also appendix A). Moreover, they stated that their traffic signal control evaluation system could be part of a maintenance cycle, since they tested it in a five-year long before-after FOT study. This introduces the evaluation of traffic signal controllers as part of maintenance cycles.

2.1.3. Evaluation as part of a maintenance cycle

Evaluation for maintenance denotes the overall and integral evaluation of signalised intersections on multiple performance indicators and variables at once in one method, in relation to the periodic maintenance of traffic signal control systems. An example of such a maintenance-oriented traffic signal controller evaluation study, is the study of Krajzewicz, et al. (2014), who developed a scheme to consistently evaluate traffic signal control algorithms based on travel time, stops, queue lengths, cycle time, delay, degree of saturation, traffic safety indicators, delay-based LOS, and route distribution. These performance indicators were already introduced by Blokpoel, Krajzewicz, & Nippold (2010). Alternatively, Sunkari (2004) discusses the benefits of signal retiming, similar to the study of Lavrenz, et al. (2016). He describes a general approach, including the performance indicators geometric intersection design, traffic flow volumes, current timing settings (cycle time, green time, etc.), and collision data, whereas field observations should be considered as well to contribute to a better signal retiming proposal. Sunkari (2004) concluded that periodic signal retiming is a relatively easy, and cost-efficient method to maintain the traffic performance.

A more integral maintenance-oriented traffic signal controller evaluation method is proposed by Radivojevic & Stevanovic (2017b). They proposed a framework that is meant to be used as an annual evaluation tool. Their framework includes several aspects, such as monitoring, user satisfaction, equipment, among others, in addition to traffic performance indicators, grouped in multiple sections: management, traffic signal operation, signal timing practices, traffic monitoring, and maintenance. The framework assigns partial grades times a weight set per section to derive a section-specific grade. These section-specific grades are confronted with an overall weight set, that in turn will determine the overall LOS of a traffic signal controller. They concluded that their framework enables an unbiassed, widely applicable evaluation of signalised intersections, due to the numerical aspects, and weight sets. In the technical report by Radivojevic & Stevanovic (2017a), on which the publication of Radivojevic & Stevanovic (2017b) is based, some countermeasures are mentioned to improve the performance of the assessed signalised intersection, based on the partial grades. These countermeasures focus mainly on parameter adjustments in terms of signal timing (green, amber, and red phase time settings, cycle time settings, and block sequence), regardless of the type of traffic signal control, i.e. fixed time traffic signal control, vehicle-actuated traffic signal control, etc. Other improvements relate to the framework itself, for instance including more performance indicators.

Besides scientific literature, there are guidelines available on traffic signal controller evaluation. In the U.S.A., the National Transportation Operations Coalition (2007) presented the Traffic Signal Audit Guide, in which a stepwise approach for assessing a traffic signal controller is proposed. The Traffic Signal Audit Guide includes a list of all the audit items that should be considered. The result of the audit is a formal statement on the performance of the traffic signal controller, accompanied with appropriate recommendations for improvements, although examples of such improvements are not given in the publication. In the Netherlands, a generic guideline for the traffic performance evaluation of signalised intersections is developed by 't Hoen, Vanhuysse, & Los (2013). In this guideline, a functional evaluation method is presented, making use of the multiple performance indicators. However, those performance indicators are not explicitly stated, since the guideline states that the relevant performance indicators depend on the used evaluation hard- and software, e.g. the KWC (*KWaliteitsCentrale*, see also appendix A) (Vialis, 2017). Furthermore, they defined four types of countermeasures:

- Countermeasures for direct implementation: (i) updating the signal timing plan in terms of green, amber, and red phase time settings, cycle time settings, and clearance times, and (ii) detector parameter adjustments (e.g. gap times);
- Short term (0-2 years): (iii) updating the signal timing plan in terms of green, amber, and red phase time settings, cycle time settings, and (iv) updating geometric design elements of the intersection (e.g. road markings, and traffic signs);
- Medium long term (2-5 years): (v) updating the signal timing plan in terms of block sequence, and consequently the green, amber, and red phase time settings, cycle time settings, etc., (vi) changing the geometric design of the intersection in terms of lane configuration, and (vii) adding or removing a signal group;
- Long term (5-10 years): (viii) changing the geometric design of the intersection in terms of reconstruction, or adding or removing an approach, or adjusting detector configuration, (ix) development plans, and (x) replacement of the traffic signal controller.

Alternatively, general traffic signal controller guidelines propose some other interesting performance indicators that are relevant in the traffic performance, and functional evaluation of signalised intersections, even though they do not focus explicitly on the evaluation of traffic signal controllers. For instance, guidelines from various countries all consider delay as performance indicator (AWV, 2009; AWV, 2011; Andersson, 2011; CROW, 2006; CROW, 2012; FGSV, 2010; Koonce, et al., 2008; Planath, et al., 2003; Statens Vegvesen, 2007; TRB, 2000; TRB, 2012; Vejdirektoratet, 2012; Vägverket & Svenska Kommunförbundet, 2004). Additionally, the American guidelines consider the Queue Storage Ratio (QSR, see also appendix A). The queue storage ratio is a performance indicator that denotes the ratio of the back-of-queue to the available vehicle storage length, and is therefore a measure for the likelihood that blockage of the lane will occur: if $QSR < 1.0$, blockage will not occur during the analysis period. Factors such as (maximum) queue length, acceleration, and deceleration affect the ratio (TRB, 2012). The Dutch guidelines mention Vehicle Lost Hours as a performance indicator. Vehicle Lost Hours are a traffic quantity which expresses the cost of delay in terms of lost time (hour) that the equivalent of one vehicle experiences (CROW, 2006), see also appendix A. This performance indicator is also used in practice (VIALIS, n.d. [a]). In Sweden, and Norway, a similar performance indicator is used, though expressed as the monetary cost of delay per vehicle, and/or per day or year (Planath, et al., 2003; Statens Vegvesen, 2007). Lastly, it must be noted that environmental effects, for instance in terms of emissions, of a signalised intersection are explicitly considered as a performance indicator Germany, and Sweden. The guidelines of these countries stated that indicators such as fuel consumption, and emissions should be included, given the impact of the traffic system on the environment (FGSV, 2010; Planath, et al., 2003; Vägverket & Svenska Kommunförbundet, 2004).

2.2. Evaluation methods

As stated in the previous section, there is little attention paid to integral evaluations of traffic signal controllers. That is, most traffic signal evaluations in literature are focused on evaluating new algorithms, systems, models, etc., rather that the traffic performance and technical function. The methods that are used to perform such evaluations relate directly to the objective of those studies, for instance using before-after studies, simulation studies, etc. The literature that did focus on the (integral) evaluation of traffic signal controller performance, did not show one method. For instance, Lavrenz, et al. (2016) used a before-after FOT study, while e.g. Radivojevic & Stevanovic (2017b) provided a method in which multiple sections are graded step by step to come to a final LOS. Similar methods are found in other literature. This implies that if a traffic signal controller is evaluated on its traffic performance, the conclusion is whether or not this is a good performance, which means that no further research is done on what the causes might be of insufficient performance. However, in practice, the causes of the insufficient performance are identified. As introduced in section 1.1, the practice at VIALIS with Instant Quality Scan (IQS) evaluations, and Quick Quality Scan (QQS) evaluations, see also appendix A. These evaluations start by finding a first identification of the traffic signal controller performance using the results of the BI-tool (*Beoordeling Instrument tool*; Rating Instrument Tool, see also appendix A). The BI-tool is an assessment tool in the KWC in which for several performance indicators, such as oversaturation, and red-light running, static threshold and reference values are given, whereas the BI-tool checks whether, and when these values are exceeded, and thus whether there were performance problems (VIALIS, n.d. [a]). Thus, the results of the BI-tool only indicate whether or not the traffic signal control performance is in line with the policies of the road authority. Next, the traffic engineer uses the output of the BI-tool to identify the cause of the potentially found insufficient performance, for instance by analysing phase-log data (see Figure 1-1), or using road user complaints[1]. This emphasises the fact that the BI-tool does not identify the cause of insufficient performance that might be found. In literature, systems or tools similar to the BI-tool are not explicitly mentioned. On the other hand, systems similar to the Dutch KWC system are mentioned. As discussed in section 2.1.1, a system to collect the necessary data is introduced by Balke, et al. (2005). However, it must be noted that Balke, et al. (2005) only focused on the data collection system, rather that the data processing tool. Nonetheless, the use of a data processing tool such as the BI-tool in traffic signal control evaluation introduces the use of expert systems and decision support systems.

2.2.1. Expert systems and decision support systems

An expert system is an automated system that focuses on problem-solving. Usually, expert systems are specialised to solve a specific type problem, using specialised knowledge, and skills. That way, it copies the way human experts solve similar specialised problems (Ford, 1985; Turban & Watkins, 1986). In other words, expert systems

[1] In some QQS and IQS evaluations, the BI-tool is not used: in those evaluations, there is no automated system used to identify insufficient performance, and thus the performance of a traffic signal controller is assessed solely by using the direct output of the KWC (e.g. phase-log data), road user complaints, etc. (VIALIS, n.d.).

aim at solving complex problems. On the other hand, decision support systems aim at helping making decisions: decision support systems are automated systems that aim at helping in the process of decision-making, for instance when selecting the appropriate countermeasure, strategy, diagnosis, etc. (Bal, Fatih Amasyali, Sever, Kose, & Demirhan, 2014; Ford, 1985; O'Sullivan, Fraccaro, Carson, & Weller, 2014; Turban & Watkins, 1986). Although such systems are automated systems, they do not automate the decision-making process. Instead, they are to be used a tool provided to the decision-maker to help him, or her making the right decision (Bolman, Jak, & Van Hoof, 2018; Power, 2002; Sprague Jr. & Carlson, 1982).

This means that expert systems, and decision support systems are two different systems. As stated by Turban & Watkins (1986), an expert system might be qualified as a decision support system, but a decision support system cannot be qualified as an expert system. This is best visualised by their respective architectures. Both systems consist of four components, as stated by Ford (1985), and Power (2002), with in both systems a user interface. In expert systems, the other components represent the task-specific data (the database), domain-specific data (knowledge base), and control (the inference component, in which a solution is found by applying the model on the database), see Figure 2-1. A decision support system also includes a database, and a model component (consisting of one or more models), whereas the database consists of data used to make decisions. The data in the database might be expert information, but also empirical data, etc. This data is used in the model component, alongside the data provided by the user, to come to a decision. That is, in the model component, rules, analytical models, simulation models, etc., are used to determine which decision must be made. The result is communicated to the user via the communication/dialogue component, see Figure 2-1. Additionally, to the communication component, Bal, et al. (2014) state that in some decision support systems, a so-called explanation module is present which aims at validating the results generated in the other components. In both expert systems, and decision support systems, the user interface is that part of the system that is directly visible for the user, the other components remain "hidden" for the user. Given the use of expert systems, and decision support systems in various fields (as will be discussed in section 2.2.2), the user is usually an expert. This emphasises the way e.g. a decision support system helps the expert making a decision.

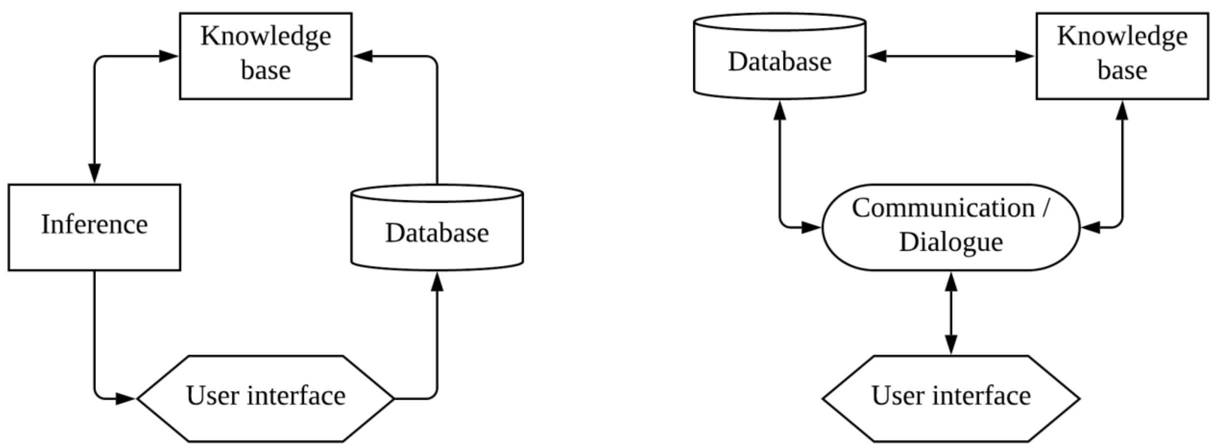

Figure 2-1 | Basic architecture of expert systems (left), and decision support systems (right), as given by Ford (1985), and Power (2002).

The basic architecture, and underlying principles of both systems imply that they might be related. Indeed, Turban & Watkins (1986) state that expert systems might be integrated in decision support systems. In most decision support systems developed in recent years, expert systems are integrated implicitly: solving a complex specialised problem has become part of the decision-making process (Bal, et al., 2014; Bolman, et al., 2018; O'Sullivan, et al., 2014). In these decision support systems, the expert system is part of the model component, and, if applicable, the explanation module. That way, decision support systems with integrated expert systems fulfil the demands of their use in practice, since Turban & Watkins (1986) concluded that, originally, decision support systems are used to solve unique problems, while expert systems are used to solve repetitive problems. The integration thus enables solving repetitive problems with a decision support system. Because of their integration in recent years, only decision support systems are discussed further below, thereby assuming that expert systems are integrated into the decision support systems.

2.2.2. Types of decision support systems

Decision support systems are widely used in various fields: the world of finance, retail, medical services, transport scheduling, retail, policy making (e.g. traffic safety policies), etc. (Bolman, et al., 2018; Hsu, 2018; Legato & Mazza, 2018; Loureiro, Miguéis, & Da Silva, 2018; Martensen, et al., 2019; O'Sullivan, et al., 2014). Of all these fields, the medical decision support systems are generally assumed to be the most used decision support systems in practice (Hsu, 2018). Even though the objective (which decision-making process is influenced) of the systems used in the various fields differ, the underlying principle remains the same.

Because decision support systems are used in such a variety of fields, there are multiple types of decision support systems. For instance, O'Sullivan, et al. (2014) state that medical decision support systems are either passive, semi-active, or active, with an increasing role of automated triggers: passive systems react on requests of the decision maker, while active systems provide decisions pro-actively. Also, they state that the complexity of the decision support system plays a role, in particular in how well it can be explained to the user on which the made decision is based. For instance, simple systems only check whether the given input is within a pre-defined allowed range, and whether countermeasures are possible if this is not the case. As soon as prognoses are made as well, the system becomes more complex. If artificial intelligence is also considered, the system is considered as complex. In complex systems, the integration of expert systems in decision support systems in recent years is emphasised, since nowadays these complex decision support system needs specialised knowledge and skills to make such prognoses. In the case of complex decision support systems, there are multiple models possible to make predictions or prognoses, of which some are as enumerated below (Bal, et al., 2014; Hsu, 2018; Loureiro, et al., 2018; Qatawneh, Alshraideh, Almasri, Tahat, & Awidi, 2019):

- *Decision trees*: decision trees are paths that consist of conditions, which can be visualised as a tree. Given the input provided by the user, it is checked whether certain conditions are met, for instance based on the database, as a way to come to the final outcome. These conditions might be rule-based, e.g. using “if-then” conditions. This type of model is found to be easy to interpret.
- *Neural networks*: neural networks make use of artificial intelligence, and mimic the neurons in the human brain, in particular the connections between the neurons. Neural networks are usually considered in combination with multi-layer perception models, which consist of at least three layers: an input layer, one or more so-called hidden layers, and an output layer. The dataflow is unidirectional, thus from input to output. This kind of network is able to classify, and process large datasets in order to come to a conclusion.
- *Support vector machines or regression*: a model based on statistical principles, that aims at estimating the values for certain parameters. This model is considered as a data mining model, and can be used to recognise patterns (Vapnik, 2000; Witten, Frank, Hall, & Pal, 2017).
- *(Naïve) Bayesian networks*: a (naïve) Bayesian network is a model that uses conditional probabilities to predict whether e.g. a diagnosis is correct. To do so, it uses Bayes’ theory, stating that the probability of a conclusion (e.g. a disease) given an observation (e.g. a symptom) depends on the probability of that conclusion (based on data regarding how often that conclusion is made, e.g. how often a disease is diagnosed), and the

probability that a certain observation is made (e.g. how often that symptom is found). This type of model performs best when large datasets are used (Lindgaard, Pyper, Frize, & Walker, 2009; Liu, Lu, Ma, Chen, & Qin, 2016).

The fact that different types of decision support systems are possible, is also stated by Power (2002), although he mainly focused on the tools or components that provide the most important functionality in the decision support system. Some of these types that focus on using numerical data are mentioned below:

- *Data-driven*: the main focus lays on processing large amounts of data, e.g. as used in spatial decision support systems;
- *Model-driven*: the dominant functionality is related to the models used, and how they can be manipulated, e.g. optimisation models;
- *Knowledge-driven*: also known as "management expert system", uses predefined knowledge, e.g. as a tool for pattern recognition.

2.2.3. Decision support systems and traffic signal control evaluation

As stated above, decision support systems are widely used in a variety of fields. Traffic engineering, and in particular traffic signal control, are rarely mentioned in literature as a field in which decision support systems are used. Nonetheless, a neural network-based decision support system is developed to monitor traffic flow at signalised intersections in terms of fundamental diagrams (Messai, Thomas, Lefebvre, & El Moudni, 2015). In relation to traffic signal control, Wen (2008) discusses an expert system used as a dynamic and automated traffic signal control algorithm, and Elkosantini & Ahmed (2014) developed a traffic signal control algorithm based on a decision support system including data fusion. Although these systems aim at improving the performance of a traffic signal controller, the systems do not discuss a traffic signal control evaluation decision support system. Furthermore, although Moalla, Elkosantini, & Darmoul (2013) also developed a traffic signal control algorithm based on a decision support system, they did include a traffic signal control evaluation component. Their evaluation component is based on an analogy to biology, namely an artificial immune system. A "disease", or "antigen" is defined as a delay, and/or queue length exceeding a predefined threshold value, the "medicine", or "antibody" is a traffic signal control action aimed at reducing the delay, and/or queue length, whereas they considered one traffic signal control action, namely retiming the cycle time, and green time. However, this traffic signal control evaluation component in the decision support system of Moalla, et al. (2013) is only used as a way to determine whether or not the cycle time, and green time should be retimed: the traffic signal control system did not identify the cause why the delay, and/or queue length exceeded the threshold value. Altogether, this implies that decision support systems aimed at traffic signal evaluation are quite rare in literature.

On the other hand, in practice decision support systems are used to evaluate traffic signal controllers. As briefly introduced above, an example of such an application of decision support systems, is the current practice at VIALIS with QQS and IQS evaluations, in particular regarding the use of the BI-tool. Because the output of the BI-tool is only information whether or not there were traffic signal control issues (did the performance indicators show that the static threshold and reference values for the given performance indicators were exceeded?), it does not state the cause of these issues. Indeed, this output is used by the traffic engineer to find the cause of the performance problems, and propose countermeasures to mitigate these problems. Although the BI-tool does not provide the cause of the performance problems, it does help the traffic engineer finding the problems, and selecting the correct countermeasures to mitigate the found problems. Therefore, it is concluded that the BI-tool is an example of a decision support system. That way, one could compare the BI-tool to a medical decision support system, where the traffic signal controller is the patient: the patient shows some unidentified symptoms, which are analysed and structured by the BI-tool. The output of the BI-tool (decision support system) is then a classification of the symptoms (what are the exact symptoms?). This classification is used by the doctor (traffic engineer) to find the corresponding disease (problem). Because of this, the BI-tool is a rather simple decision support system. Moreover, given the classifications of O'Sullivan, et al. (2014), and Power (2002), it is concluded that the BI-tool is a (i) passive, (ii) simple, (iii) data-driven decision support system, because (i) it is only activated by a request of the user, (ii) the system only checks whether the inputs are according to the pre-defined threshold and reference values, without any prognoses or whatsoever, and (iii) it processes large amounts of data, without using any models, or recognising patterns, for instance. Because it is a simple decision support system, the type of model (decision trees, neural networks, support vector machines, and (naïve) Bayesian networks) the BI-tool uses is not relevant, since it does not make predictions or prognoses.

The conclusion that the BI-tool is a decision support system, is useful when developing a traffic signal controller evaluation and diagnosis method. Indeed, the traffic signal controller evaluation and diagnosis method is considered as an extension of the existing BI-tool, which does include the diagnosis of the cause of the found performance problems, as discussed earlier. Using the analogy of a medical decision support system, this means that the traffic signal controller evaluation and diagnosis method aims at identifying the disease (problem) as well, as addition to the current practice of only identifying the exact symptoms of the traffic signal controller. Because of this, the classification of the BI-tool as decision support system is used as starting point, since it already includes parts of the integral traffic signal control evaluation and diagnosis method presented in this thesis.

2.3. Main findings

Based on the literature review discussed in this chapter, the main findings are as listed below:

- The literature showed that traffic signal controllers are evaluated for various reasons: (i) assessment of new traffic signal control aspects (e.g. algorithms, systems, models, etc.), (ii) traffic performance evaluation, and (iii) evaluation as part of periodic maintenance. The literature reviewed for all of these three main motives showed that little to no attention is given to the integral evaluation of traffic signal controllers. That is, only a limited number of studies focused on how the traffic signal control performance could be improved.
- The study of Bullock & Day (2009) discussed the development of a data collection method that can be used to evaluate the traffic performance of traffic signal controllers. Their method is rather similar to the Dutch method using the KWC. Lavrenz, et al. (2016), and Sunkari (2004) stated that traffic signal retiming, and thus a periodic traffic performance evaluation of traffic signal controllers, improves the traffic performance of these traffic signal controllers, although did not provide a framework to perform such an evaluation, nor did they list other countermeasures that could improve the traffic performance. A framework for a periodic traffic performance evaluation of traffic signal controllers as part of periodic maintenance is proposed however by Radivojevic & Stevanovic (2017b), although they did not consider any countermeasures to improve the traffic performance. Also, they included process and management related performance indicators. Altogether, this implies that integral traffic signal controller evaluation methods, with the objective to evaluate and improve the traffic performance, are seldom investigated in literature.
- Additionally, in literature there is little to no attention paid to automated traffic signal controller evaluation methods: the use of e.g. expert systems, and decision support systems as a way to evaluate and diagnose traffic signal controllers is not discussed widely in literature. However, the use of decision support systems in other fields, such as medicine, is widely discussed, as well as the integration of expert systems in decision support system. Although traffic signal control decision support system examples are found in literature, for instance Wen (2008), Elkosantini & Ahmed (2014), and Moalla, et al. (2013), they all focused on the development of traffic signal control algorithm based on a decision support system, rather than the evaluation of a

traffic signal controller. The traffic signal control algorithm of Moalla, et al. (2013) did include an evaluation component as a trigger for actions, though it did not formally assess the traffic signal controller, nor did it diagnose the causes of insufficient performance.

- On the other hand, in practice, the so-called BI-tool is classified as a decision support system, because it helps the traffic engineer diagnosing the traffic performance issues of the traffic signal controller. Moreover, given the classifications of O'Sullivan, et al. (2014), and Power (2002), it is concluded that the BI-tool is a (i) passive, (ii) simple, (iii) data-driven decision support system, because (i) it is only activated by a request of the user, (ii) the system only checks whether the inputs are according to the predefined threshold and reference values, without any prognoses or whatsoever, and (iii) it processes large amounts of data, without using any models, or recognising patterns, for instance. Because of this, the BI-tool as decision support system is used as starting point in developing the integral traffic signal control evaluation and diagnosis method presented in this thesis. This also relates to the fact that the BI-tool already includes several interesting components, such as determining whether the traffic signal control performance is sufficient.

3. Traffic signal controller evaluation and diagnosis method

As concluded in the previous chapter, little attention is given in literature to an integral method to evaluate and diagnose traffic signal controller. Therefore, such a method is developed in this thesis. The integral traffic signal controller evaluation and diagnosis method is presented in this chapter. This chapter gives the outlines of the method. The details of the method, and the testing of the method in a case study will be discussed in following chapters.

The traffic signal controller evaluation and diagnosis method consists of two major modules: the inefficiency detector, and the diagnosis module. Besides these two modules, several other steps are included, yielding a total of five steps. The integration of these modules and steps into the integral evaluation and diagnosis method, basically represents a more elaborate BI-tool. That way, the integral traffic signal controller evaluation and diagnosis method is also a decision support system, although the evaluation and diagnosis method takes it one step further: the evaluation and diagnosis method does diagnose the traffic signal controller problems. In other words, an extra step is added to the BI-tool, implying that not only the traffic signal controller problems are found as values of performance indicators exceeding a predefined threshold or reference value, but these values are used to assign the correct cause to these problems. Additionally, the traffic signal controller evaluation and diagnosis method makes use of multi-variable performance indicators.

Thus, in short, this chapter discusses the integral traffic signal controller evaluation and diagnosis method, including its major modules. To this end, the relevant definitions related to the modules, and decision support system components are discussed are discussed first. Next, the evaluation and diagnosis method is discussed step by step, including the underlying procedures of the major modules. The chapter concludes with the main findings and how this is relevant for the development and testing of the method in the next chapters.

Additionally, for reference purposes, the variables, introduced in this chapter, are defined in *Notation and definitions* on page ix of this thesis as well.

3.1. Evaluation and diagnosis method definitions

The traffic signal controller evaluation and diagnosis method makes use of several definitions, in particularly regarding the modules of the method (the inefficiency detector, and the diagnosis module). These definitions enable the evaluation, and diagnosis process. Also, it shows how the method is designed as a decision support system.

3.1.1. Definitions used in the modules

The traffic signal controller evaluation and diagnosis method thus includes an inefficiency detector, and a diagnosis module. The former introduces the term "inefficiency". The method uses the following definition of an inefficiency:

An inefficiency is a difference between the generated, measured performance, and the computed reference performance.

This definition introduces the terms "generated performance" and "reference performance". The generated performance is defined as follows:

The generated performance denotes the performance of the traffic signal controller, in terms of the considered multi-variable performance indicators, that is measured, e.g. in practice, thus representing what the performance has been.

This means that, basically, the generated performance is the measured, actual performance. On the other hand, the definition of the reference performance implies that the reference performance is a measure for the expected performance of the traffic signal controller:

The reference performance denotes the computed performance of the traffic signal controller, using multi-variable performance indicators, thus representing what performance should have been.

The definition of an inefficiency, with respect to definitions of the generated, and reference performance implies that the reference performance is considered as efficient performance. This means that the reference performance represents not only the expected performance, but also good performance. However, the exact definition of good performance also depends on the inputs used to compute the reference performance. The inputs depend on the describing variables of the multi-variable performance indicators used in the evaluation and diagnosis method. This will be explained in more detail in section 4.2, as well a more precise definition of good performance. In this chapter, the definition of the reference performance as good performance suffices.

The definition of the reference performance as good performance in relation the definition of an inefficiency also implies that if the generated performance is “better” than the reference performance (e.g. the generated delay is less than the computed delay), this is still considered as an inefficiency in the method presented in this thesis. This done to simplify the formulation, and testing of the integral method. This means that if an inefficiency is given as a “better” generated performance than the reference performance, a different method might be needed to assess this type of inefficiency. By assuming that all differences between the generated performance, and the reference performance are inefficiencies, regardless of whether the generated performance might be considered as a “better” generated performance than the reference performance, the integral evaluation and diagnosis method is simplified. However, it is recommended for future work to investigate how a “better” generated performance with respect to the reference performance can assessed differently, and what this means for the steps of the integral evaluation and diagnosis method.

Furthermore, it must be noted that the definition of an inefficiency is based on the current practice in the BI-tool, namely comparing the measured traffic performance of a traffic signal controller to predefined static threshold or reference values. In the traffic signal controller evaluation and diagnosis method, these static threshold or reference values are made dynamic, using reference performance models, as will be discussed below. Therefore, the definition of an inefficiency explicitly states that the difference between the generated performance, and the reference performance plays a crucial role. This means that the inefficiency detector compares the generated performance with the reference performance as a way to detect potential inefficiencies, emphasising the relation to the BI-tool. Also, the use of a reference performance shows resemblance to the before-after studies found in literature, see section 2.1. In a before-after study, the results of the “before study” are used as reference to compare with the results of the “after study”. However, in this traffic signal controller evaluation and diagnosis method, the “before study” is the computed reference performance, and the generated performance represents the “after study”. Thus, because of the similarities with the BI-tool, and the before-after studies in literature, the definition of an inefficiency is reference-based.

The reference performance is computed using computational models, the so-called reference performance models. These models will be discussed in section 4.1.5. Because these models use several assumptions, as will be discussed in section 4.3, the traffic signal controller evaluation and diagnosis method includes a calibration step to calibrate the reference performance models in relation to these assumptions, as will be discussed in sections 3.2.2, and 4.4.

3.1.2. Decision support system components

As stated above, the traffic signal controller evaluation and diagnosis method is a decision support system. This is because the method helps the traffic engineer (user) to diagnose the traffic performance problems of the assessed traffic signal controller. The traffic signal controller evaluation and diagnosis method as a decision support system is classified as a passive, medium-complex, knowledge-driven system, using an informal naïve Bayesian network combined with a decision tree, given the classifications as discussed in section 2.2.2. It must be noted that this is based on the principle that the traffic signal controller evaluation and diagnosis method is basically a more elaborated BI-tool. Also, this classification is based on the fact that the traffic signal controller evaluation and diagnosis method as a decision support system is (i) only activated by a request of the user (thus passive), (ii) it only checks whether or not inefficiencies are found, though using an informal Bayesian network (thus mid-complex, given the aforementioned definition of an inefficiency, and use of a Bayesian network), and (iii) it uses predefined, specialised knowledge (thus knowledge-driven, since reference performance models are used). Although the traffic signal controller evaluation and diagnosis method uses probabilities to assign an inefficiency (a symptom) to a traffic signal controller issue, these probabilities are not formalised, hence the use of an informal naïve Bayesian network. Based on the probabilities, a decision tree is used to select the correct countermeasures. This will be discussed in more detail in section 3.2.4.

However, it must be noted, that the traffic signal controller evaluation and diagnosis method, as presented in this thesis, does not fulfil all the requirements of a formal decision support system (the method presented here is not fully automated, as will be discussed in section 3.2). Also, the user interface, and communication component are developed, implying that only the model component, and database are developed in this thesis. Therefore, the traffic signal controller evaluation and diagnosis method is not a ready-to-use decision support system, but rather is considered as a proto-type, or framework of a decision support system. Nonetheless, the objective of the traffic signal controller evaluation and diagnosis method remains the same, namely to help traffic engineers in diagnosing traffic signal controllers. In future work, it is recommended to formalise the traffic signal controller evaluation and diagnosis method as a decision support system.

3.2. Evaluation and diagnosis method step by step

As introduced, the traffic signal controller evaluation and diagnosis method is an integration of the two major modules (inefficiency detector, and diagnosis module) into a five-step method. The steps are as enumerated below:

1. The first step describes the selection of multi-variable performance indicators that are used in the evaluation and diagnosis of the considered traffic signal controller. The selection of multi-variable performance indicators is based on the problems that one wants to diagnose. This is closely related to what a road authority considers as a problem, which in turn relates to the policy of the road authority on signalised intersections. That way, the selection of multi-variable performance indicators is based on being able to find the problems. The selection of problems, and performance indicators as tested in this thesis, will be discussed in section 3.2.1. Furthermore, the selected multi-variable performance indicators must be defined in such a way that a reference performance can be determined. In other words, the definitions of the selected multi-variable performance indicators must be formulated in such a way that it is possible to determine good performance, given as the reference performance. The output of this first step are the selected multi-variable performance indicators.
2. Given the selected multi-variable performance indicators, the corresponding reference performance models are calibrated. This calibration aims at adjusting the reference performance models of the considered multi-variable performance indicators in terms of how it handles certain assumptions. Therefore, this only relates

to the performance indicators of which the definitions of the reference performance models include such assumptions. The calibration step uses measured data on the base case generated performance and describing variables needed to determine the reference performance, and computed data on the reference performance. This data originates from an (assumed) problem-free traffic signal controller, which is considered as the base case. An (assumed) problem-free traffic signal controller is a traffic signal controller for which it is concluded that the traffic performance can be used as base case performance. This means that the used data must originate from a period in which it is known, or assumed that the traffic signal controller performed good, hence the term "problem-free". This can be data from e.g. shortly after the first implementation of the controller in practice, or from a period during which the road authority has concluded that the traffic performance was good. In the latter case, it is possible that there are problems present, hence the notation of an "assumed" problem-free traffic signal controller. The result of the calibration step is then that the reference performance models are adjusted, to cancel out inaccuracies of the reference performance models. The exact outputs, and steps of the calibration will be discussed in section 3.2.2. The assumptions, and multi-variable performance indicators that are considered in this step, will be discussed in sections 4.1, and 4.3.

3. Next, the first major module is initiated. The first major module is the inefficiency detector. The objective of the inefficiency detector is to detect analysis periods, or time intervals with inefficient performance, and thus to detect inefficiencies. The exact process of the inefficiency detector, as well as the outputs of the inefficiency detector, will be discussed in section 3.2.3. The output of this step is used in the second major module (diagnosis module) in the next step.
4. In this fourth step, the diagnosis module is initiated. The diagnosis method aims at identifying the cause of inefficient performance, using the inefficiency detector outputs, thus diagnosing the problems of the traffic signal controller. The diagnosis module explicitly includes the aforementioned informal naïve Bayesian network in combination with a decision tree to diagnose the problems of the assessed traffic signal controller. Thus, the diagnosis of the problems, based on the inefficiency detector output, uses informal probabilities in combination with "if-then" rules (hence the decision tree) to find the cause of the detected inefficiencies. The use of the informal naïve Bayesian network implies that the diagnosis module uses data on the problems, and the way they cause inefficiencies. The diagnosis module can diagnose multiple problems at the same traffic signal controller. Based on the diagnosed problems, countermeasures are implemented. In this step, it is also checked whether the proposed countermeasures mitigated the diagnosed problems successfully. Therefore, a feedback loop is included back to step (3), the inefficiency detector. The exact steps of the diagnosis module will be discussed in section 3.2.4. If the countermeasures are implemented successfully, the last step is considered. This implies that the output of this step is a list of the diagnosed problems, and a list of proposed countermeasures.
5. It is assumed that the successful implementation of the countermeasures resulted in an optimised traffic signal controller. Therefore, this last step represents the result of the traffic signal controller evaluation and diagnosis method. This step is, however, also the starting point for a re-evaluation. Thus, the output of this step might be used to define a new (assumed) problem-free traffic signal controller, as used in the calibration in step (2). Also, given the changes in the traffic flow volumes, etc., the optimised traffic signal controller might become a faulty traffic signal controller over the years, e.g. as stated by Lavrenz, et al. (2016) in section 2.1.2, which is then used as input for a new evaluation and diagnosis process. However, it must be noted that even with changes in the traffic flow volumes, etc. over the years, the generated performance might still be considered as efficient in the future, due to the use of multi-variable performance indicators in the integral method presented here, implying that policies of the road authority might change as well over the year. Also, new control methods might yield similar results, i.e. a changed performance. Still, both aspects can be accounted for with this integral evaluation and diagnosis method.

The steps as discussed above are visualised in Figure 3-1. The blue boxes represent the steps, the lined boxes represent the main aspects per steps, and the dotted lines represent the data flow of inputs, and outputs per step.

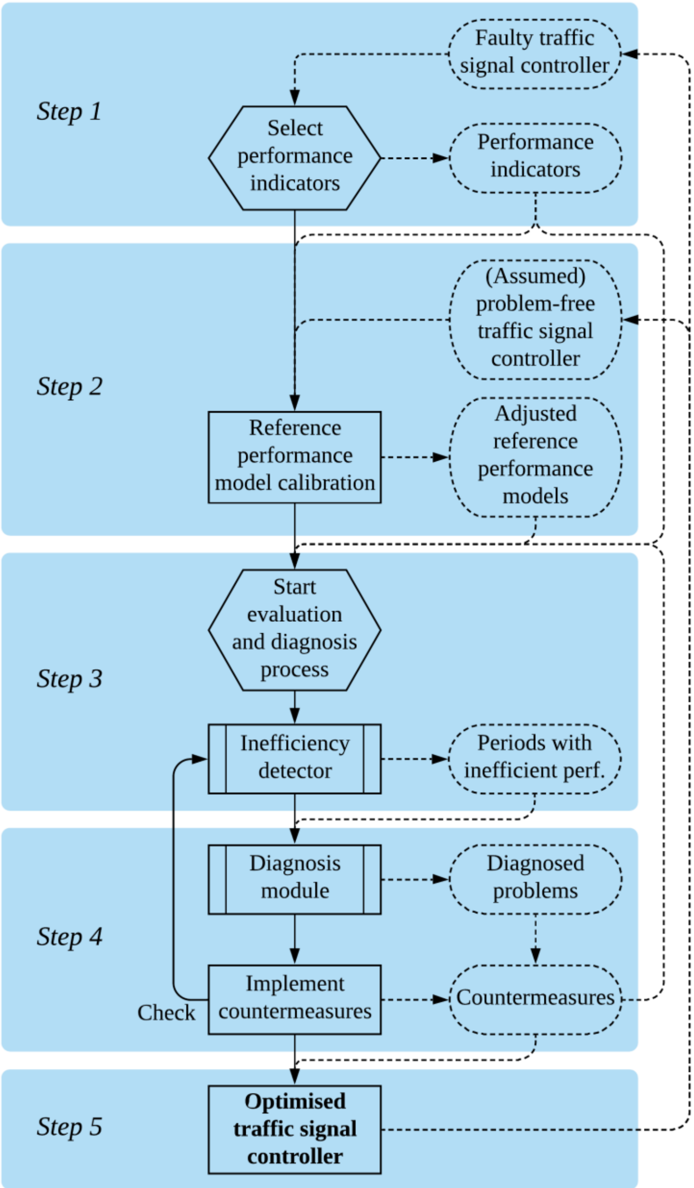

Figure 3-1 | Schematic overview of the traffic signal controller evaluation and diagnosis method.

3.2.1. Selection of multi-variable performance indicators

The selection of the multi-variable performance indicators is based on the policy constraints of the road authority, and the problems that one wants to diagnose, as discussed above. This basically outlines the scope of the integral traffic signal controller evaluation and diagnosis method.

The scope of the traffic signal controller evaluation and diagnosis method, as tested in this thesis, is mainly based on the problems of a traffic signal controller. The problems are found in literature, and evaluation reports from practice. Indeed, in literature, and in practice (e.g. in QQS and IQS evaluation reports) a wide variety of traffic signal control issues, and performance indicators are listed. Since it would cost too much time to include all, and thus test all, a selection is made. Besides, the use of a limited number of traffic signal control issues, and performance indicators to develop the traffic signal controller evaluation and diagnosis method is considered to fit best with the research objective, in particularly given the limited attention given to such a method in literature.

The scope of the method as tested in this thesis will be discussed in more detail in section 4.1. In that section, the criteria for the selected multi-variable performance indicators, and definitions are discussed. Nonetheless, the selected multi-variable performance indicators are as follows:

- Degree of saturation;
- Delay;
- Phase failure;
- Queue length.

For these four multi-variable performance indicators it holds that the reference performance models for delay, and queue length have to be calibrated, because these reference performance models make use of several assumptions, as will be discussed in section 4.3.

3.2.2. Reference performance model calibration

As briefly introduced, the objective of the calibration step is to account for the inaccuracies of the reference performance models that follow from using certain assumptions in said performance models. This means that the error of the computed reference performance is reduced with respect the generated performance of the base case (assumed) problem-free traffic signal controller. Thus, to calibrate the reference performance models, data is needed of an (assumed) problem-free traffic signal controller. As stated above in section 3.2.1, this is only relevant for the reference models for delay, and queue length, as tested in this thesis.

3.2.2.1. Result of the calibration

The result of the calibration is two-sided: at the one hand, a bandwidth is defined which describes the range in which the differences between the reference performance, and generated base case performance of the (assumed) problem-free traffic signal controller are the result of reference performance model inaccuracies. On the other hand, the reference performance models are adjusted in order to account for an over- or underestimation of the reference performance with respect to the base case performance of the (assumed) problem-free traffic signal controller.

The adjustment of reference performance models is done by adding an error term ε to the reference performance model, whereas the error term is defined per signal group j. The error term is defined as the statistical average of the difference between reference performance r [-][1], and generated base case performance p^* [-][1] of the (assumed) problem-free traffic signal controller. Since the reference performance is computed using a reference performance model, the reference performance r is defined as a function of the describing variables of the reference performance model X_1 to X_n, for n describing variables X [-][1]. These describing variables are measured. In mathematical terms, this comes down to:

$$\varepsilon_{i,j} = \frac{1}{n_\tau} \cdot \sum_{\tau} \left(r_{i,j}^{\tau}\left(X_{i,j,1}^{\tau}, \dots, X_{i,j,n}^{\tau}\right) - p_{i,j}^{*,\tau} \right) \tag{3-1}$$

[1] The unit of the reference performance, and generated performance depend on the unit of the multi-variable performance indicator of which the respective performances are given. For the units of the describing variables, it also holds that they depend on the units of the describing variables of the considered multi-variable performance indicator.

for multi-variable performance indicator i, and with the number of analysis periods n_τ [#], and analysis period index τ.

The result of the adjusted reference performance models is then that the differences between the reference performance, and generated performance, found when applying the adjusted reference performance models on the traffic signal controller that is evaluated, are not the result of inaccuracies in the reference performance models, but rather are caused by problems of the traffic signal controller.

The bandwidth represents the range in which the deviations of the reference performance with respect to the generated performance are assumed to be caused by inaccuracies in the reference performance models. A generated performance that lays in the bandwidth is considered as good, efficient performance. Only when the generated performance p [-][1] of a signal group j, for a multi-variable performance indicator i, during an analysis period τ, lays outside the bandwidth B [-][1], it is concluded that this is caused by problems of the traffic signal controller. Thus, when it holds that:

$$p_{i,j}^{\tau} \notin B_{i,j} \tag{3-2}$$

the generated performance is considered as inefficient. The bandwidth is thus an interval. The interval is defined using the statistical standard deviation of the difference between the reference performance, and generated performance of the base case (assumed) problem-free traffic signal controller:

$$B_{i,j} = \left[r_{i,j}\left(X_{i,j,1}, \dots, X_{i,j,n}\right) - 2\sigma_{i,j} \ , \ r_{i,j}\left(X_{i,j,1}, \dots, X_{i,j,n}\right) + 2\sigma_{i,j}\right] \tag{3-3}$$

whereas:

$$\sigma_{i,j} = \sqrt{\frac{1}{n_\tau} \cdot \sum_{\tau} \left(\left(r_{i,j}^{\tau}\left(X_{i,j,1}^{\tau}, \dots, X_{i,j,n}^{\tau}\right) - p_{i,j}^{*,\tau}\right) - \varepsilon_{i,j}\right)^2} \tag{3-4}$$

with signal group index j, multi-variable performance indicator i, bandwidth B [-][1], standard deviation σ [-][1], error term ε [-][1], number of analysis periods n_τ [#], and analysis period index τ, reference performance $r(X_1, \dots, X_n)$ [-][1] as function of n measured describing variables X [-][1], and generated base case performance p^* [-][1].

The bandwidth is thus defined using the value of two times the standard deviation, 2σ. The choice for 2σ is rather arbitrary. In the pre-testing of the method, in particular the inefficiency detector module, it was found that using only σ resulted in a quite high number of detected inefficiencies for the (assumed) problem-free traffic signal controller. Since it is considered as an (assumed) problem-free traffic signal controller, the number of detected inefficiencies should be low. It was found that this is the case when using 2σ. In future research, it is recommended to investigate how an alternative definition of the bandwidth yields the best results.

---

[1] The units depend on the unit of considered multi-variable performance indicator, and its describing variables.

3.2.2.2. Calibration process

The calibration step itself is a three-step process. The steps are identical for each reference performance model that needs to be calibrated. In this case, this means that the steps are identical for the calibration of the reference performance models for delay, and queue length. The steps are as enumerated below

1. First, data is collected for the base case generated performance of the (assumed) problem-free traffic signal controller. Additionally, data is collected on the describing variables needed to compute the reference performance. The output of this step is data on the generated performance, and on the describing variables needed to determine the reference performance.
2. Given the data on the describing variables, the reference performance is computed. This is also the output of this step.
3. The data of both the generated base case performance, and reference performance is used to determine the error term, and bandwidth. The error term is used to adjust the reference performance model. The result is then that the reference performance models of the multi-variable performance indicators are calibrated, and are thus ready to use in the inefficiency detector, and diagnosis module.

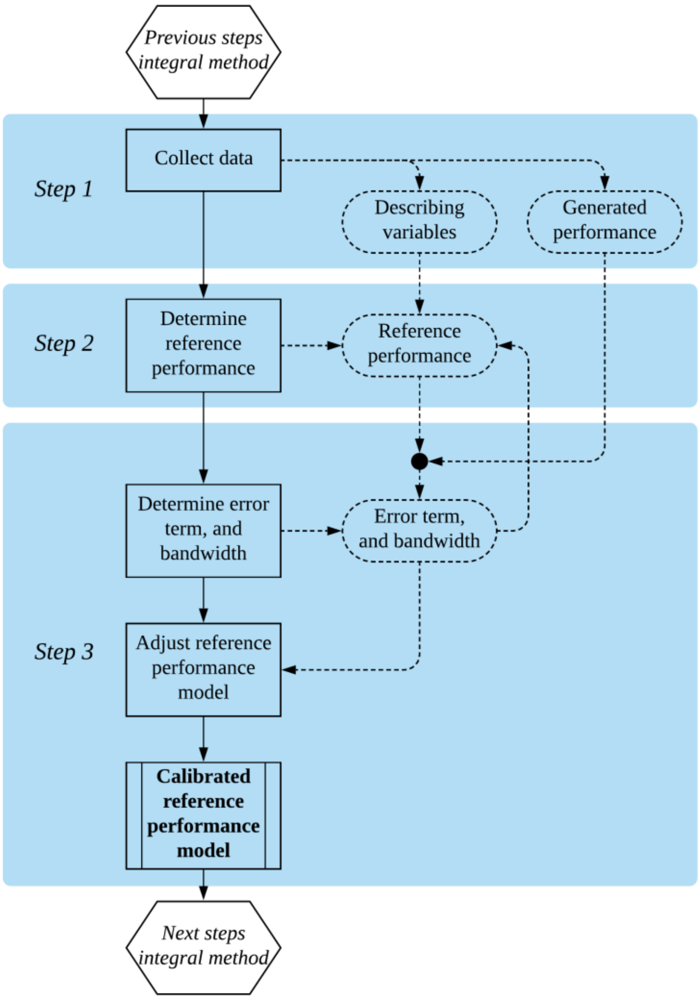

Figure 3-2 | Schematic overview of the process of the reference performance calibration used to determine the adjusted reference performance models, including the error terms, and bandwidths.

The steps of the reference performance calibration are visualised in Figure 3-2 on the previous page. The blue boxes represent the steps, and the lined boxes represent the actions that are performed within a step, as well as the sequence of actions, given as arrows. The dotted boxes, and lines represent the data flow.

It must be emphasised that the calibration step is only relevant for the reference performance models of the multi-variable performance indicators that make use of certain assumptions that may lead to inaccuracies of the reference performance model, as will be discussed in section 4.3. Thus, the calibration step may not be performed for all selected multi-variable performance indicators, implying that not for every multi-variable performance indicator an error term, and a bandwidth is determined. For future work, it is recommended to investigate whether the reference performance models of other multi-variable performance indicators need to be calibrated as well.

3.2.3. Inefficiency detector

The basic principle of the inefficiency detector is to identify those analysis periods during which the generated performance deviates from the reference performance. This relates to the definition of an inefficiency as given in section 3.1.1. If for a multi-variable performance indicator a bandwidth is defined in the calibration step, it holds that an inefficiency is only found when Equation 3-2 is true. Otherwise, an analysis period τ had inefficient performance if it holds that:

$$p_{i,j}^{\tau} \neq r_{i,j}^{\tau}\left(X_{i,j,1}^{\tau}, \ldots, X_{i,j,n}^{\tau}\right) \tag{3-5}$$

with generated performance p [-][1], and reference performance $r(X_1, \ldots, X_n)$ [-][1] as function of n measured describing variables X [-][1], for multi-variable performance indicator i, and signal group j.

In other words, the inefficiency detector aims at detecting inefficiencies per analysis period, which implies that if during such an analysis period, no inefficiencies are detected, the conclusion is that during that period, the traffic signal controller performed efficiently.

3.2.3.1. Inefficiency detector outputs

Therefore, the basic output of the inefficiency detector is a list of analysis periods during which inefficiencies are detected, thus during which Equations 3-2, and/or 3-5 are true. In addition to this list, several other outputs are generated, as enumerated below:

- *Performance statistics*: the performance statistics are the statistical averages, and standard deviations of the differences between the reference performance, and the generated performance.
- *Inefficiency ratio*: the inefficiency ratio is defined in this thesis as the ratio of the number of analysis periods during which inefficient performance was found, with respect to the total number of analysis periods. That way, the inefficiency ratio is a measure for how often the performance was found to be inefficient, and thus how long the performance was considered as inefficient. The inefficiency ratio IR is defined per multi-variable performance indicator i, and signal group j:

$$\mathrm{IR}_{i,j} = \frac{n_{i,j}^{\mathrm{ineff}}}{n_{\tau}} \tag{3-6}$$

[1] The units depend on the unit of considered multi-variable performance indicator, and its describing variables.

with inefficiency ratio IR [-], number of periods with inefficient performance $n_{i,j}^{\text{ineff}}$ [#], and total number of analysis periods n_τ [#]. The number of periods with inefficient performance are thus based on whether Equations 3-2, and/or 3-5 are true. If the inefficiency ratio equals 1.00, the performance was inefficient during the complete analysis period.

- *Evaluation score*: the evaluation score is a measure on a scale from 1 to 10, representing how efficient the performance was, whereas 1 is the most inefficient, and 10 is the most efficient. That way, the evaluation score is a measure for how efficient the performance of a multi-variable performance indicator was when compared to another multi-variable performance indicator. The definition of the evaluation score as given in this thesis implies that the computation of the evaluation score can only be done for calibrated reference performance models. Thus, for the multi-variable performance indicators of which the reference performance models did not have to be calibrated, no evaluation score can be computed. Therefore, this means that the evaluation scores in this thesis are computed for only delay, and queue length, as mentioned in section 3.2.1. The evaluation score is computed via:

$$g_{i,j} = \frac{1}{n_\tau} \cdot \sum_\tau \left(10 - 4.5 \cdot \frac{\left| r_{i,j}^\tau \left(X_{i,j,1}^\tau, \dots, X_{i,j,n}^\tau \right) - p_{i,j}^\tau \right|}{2\sigma_{i,j}} \right) \tag{3-7}$$

$$g_{i,j} \in [1,10] \tag{3-8}$$

where g [-] equals the evaluation score, n_τ [#] the total number of analysis periods, τ the analysis period index, σ [-][1] the standard deviation as defined above in Equation 3-4 (hence the limitation that the evaluation score can only be computed for multi-variable performance indicators with calibrated reference performance models), p [-][1] the generated performance, and $r(X_1, \dots, X_n)$ [-][1] the reference performance as function of n measured describing variables X [-][1], for multi-variable performance indicator i, and signal group j. The computation of the evaluation score shows again that if the generated performance is "better" than the reference performance (e.g. the generated delay is less than the reference delay), this treated the same as a generated performance that is "worse" than the reference performance (e.g. the generated delay is more than the reference delay), by using the absolute difference between the reference performance, and generated performance. As explained in section 3.1.1, this relates to the definition of an inefficiency, as a way to simplify the formulation, and testing of the integral method.

Although the evaluation scores can only be computed for multi-variable performance indicators with calibrated reference performance models, the evaluation scores are considered as a valuable inefficiency detector output, because it is measure for how (in)efficient the performance was. The inefficiency ratio only gives information on whether or not, and how often the performance was inefficient. It takes an analysis period with a slightly inefficient performance into account the same as it takes an analysis period with an extremely inefficient performance into account. The evaluation score is defined to help giving additional information for those analysis periods where the performance was slightly inefficient, for instance. The result might then be that even though the performance was quite inefficient as function of time (e.g. the inefficiency equalled IR $= 60\%$), while the gradation of inefficient performance, given as the evaluation score, might have been just below a sufficient score, e.g. $g = 5.3$ on average during those inefficient periods. This enables a more profound decision of which countermeasures are to be implemented. For instance, from the point of view of the road authority, which might have a limited budget, the road authority might conclude to implement low-cost countermeasures only, since the evaluation score is slightly insufficient.

[1] The units depend on the unit of considered multi-variable performance indicator, and its describing variables.

3.2.3.2. Procedure of inefficiency detector

Altogether, the inefficiency detector module consists of a four-step procedure, as enumerated below:

1. First, data is collected on the generated performance, and the describing variables of the reference performance models. The data is collected for the traffic signal controller that is assessed, and might have problems, thus not for the (assumed) problem-free traffic signal controller used in the calibration step.
2. The data on the describing variables is used to determine the reference performance. This is also the output of this step.
3. Next, the performance statistics are computed, per multi-variable performance indicator, per signal group, using the difference between the reference performance, and generated performance. Also, the inefficiencies are detected. The generated performance from step (1) is compared to the reference performance from step (2). The definition of an inefficiency, as given in section 3.1.1, is leading. In other words, an analysis period is only considered as an inefficient period, if that definition holds. In mathematical terms, it thus holds that an analysis period had inefficient performance if Equations 3-2, and/or 3-5 are true for the considered multi-variable performance indicator, and signal group. Otherwise, the generated performance is considered as efficient. The output of this step is then the basic output of the inefficiency detector procedure: a list of analysis periods at which the traffic signal controller performed inefficiently.
4. Lastly, the other inefficiency detector outputs are computed: the inefficiency ratio, and evaluation scores. This completes the output of the inefficiency detector procedure, thereby including the list of analysis periods during which inefficiencies are detected, and performance statistics as well.

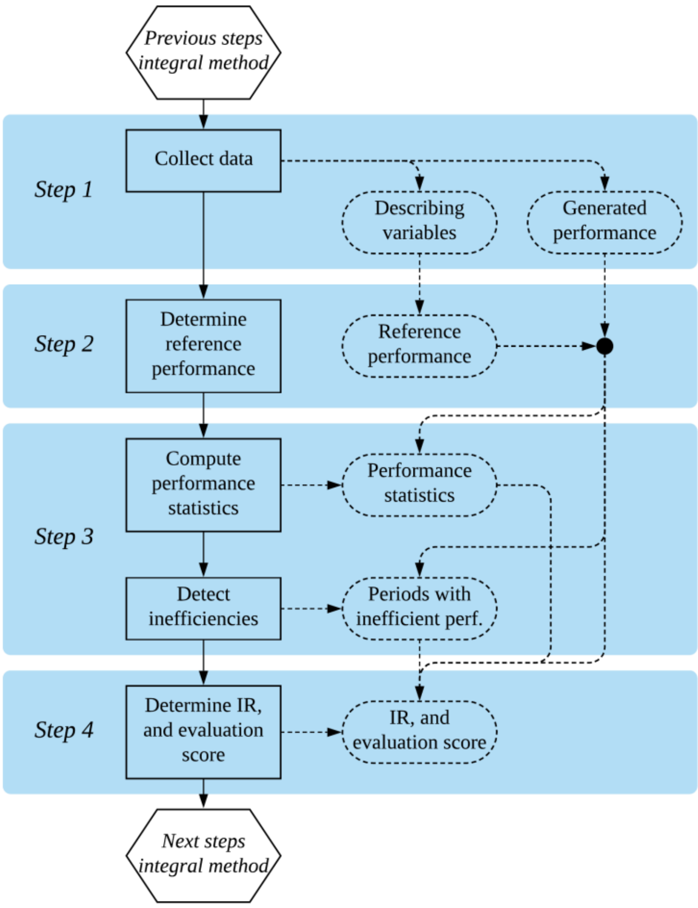

Figure 3-3 | Schematic overview of the inefficiency detector procedure, with inefficiency ratio IR.

The aforementioned steps are visualised in Figure 3-3 on the previous page, where the blue boxes represent the steps. The lined boxes represent the actions that are performed within a step, as well as the sequence of actions, given as arrows. The dotted boxes, and lines represent the data flow.

3.2.4. Diagnosis module

The inefficiency detector results are used in the diagnosis module to assign the correct cause to the detected inefficiencies. That is, the traffic signal control problems that caused inefficient performance are identified. Also, the diagnosis module includes a proposal of countermeasures that aim at mitigating the found problems.

3.2.4.1. Diagnostic principles

The diagnosis module shows resemblance to the plan-do-check-act cycle, developed to standardise the process of change. It can also be used to improve a process, or system, and consolidate those improvements. The cycle consists of four steps: (i) plan, (ii) do, (iii) check, and (iv) act. The cycle consecutively (i) plans the opportunity to improve the considered system, (ii) tests the improvement, (iii) checks whether the tested improvement yielded the desired result, and (iv) implements the improvement if proven effective (American Society for Quality, 2019). The similarity between the plan-do-check-act cycle, and the diagnosis, is that both are continuous, and repetitive processes, aiming at improving the system. Although the explicit steps of procedure of the diagnosis module, as will be discussed in section 3.2.4.2, cannot be translated literarily to the four steps of the plan-do-check-act cycle, the same principle holds. That way, a continuous, repetitive cycle is used to ensure that the countermeasures did mitigate the found inefficiencies, and thus that the countermeasures were effective.

Furthermore, the diagnosis module explicitly relates to the decision support system components, because in the diagnosis module, the medium-complex, knowledge-driven characteristics of the integral evaluation and diagnosis method are emphasised, which means that, as introduced in section 3.1.2, the informal naïve Bayesian network combined with a decision tree is used to select the correct countermeasures to mitigate the found problems. The diagnosis module uses a database filled with examples of inefficient performance caused by various problems, whereas the examples are given as data on the generated performance, reference performance, and inefficiency detector outputs. The database thus includes the data on (i) which generated performance is measured for a given inefficiency and problem, (ii) the number of times that inefficiency is found at a traffic signal controller, and (iii) the number of times that problem is diagnosed at a traffic signal controller. Altogether, this is used to find the probability that a found generated performance, in relation to the inefficiency detector outputs, is caused by the given problem. These probabilities are part of the naïve Bayesian network, as stated by Bal, et al. (2014), Hsu (2018), Lindgaard, et al. (2009), and Liu, et al. (2016). However, these probabilities are not formalised, hence the use of the term “informal” naïve Bayesian network. Therefore, it is recommended for future work to formalise the naïve Bayesian network.

Additionally, the informal probabilities of the naïve Bayesian network are used in a decision tree to select the correct countermeasures. Because the probabilities are not formalised, a conceptual pattern recognition is used: if the pattern (the generated performance, reference performance, and inefficiency detector outputs) of a found inefficiency at the assessed traffic signal controller is similar to the pattern of an inefficiency in the database, the inefficiency is diagnosed as caused by that inefficiency from the database. Thereto it is tested that a generated pattern is similar to a pattern in the database if the generated pattern shows the same trend as the pattern in the database – i.e. a high inefficiency ratio for a specific (set of) signal group(s), approximately equal performance statistics, etc. Because this implies a somewhat informal, and arbitrary approach, it is recommended to formalise the pattern recognition in future work. Besides, this pattern recognition process is not automated. Therefore, it is recommended for future work to automate the pattern recognition process in this diagnose module.

The conceptual pattern recognition implies that the decision tree includes as many decision steps as that there are problems possible. To limit this number of decision steps, this thesis uses a limited number of selected problems, as will be discussed in depth in section 4.1. Also, because the pattern recognition in the decision tree, as discussed in this thesis, is not automated yet, the pattern recognition is considered as a proto-type, hence the term "conceptual" pattern recognition. The informal naïve Bayesian network in combination with the decision tree, including the inefficiency detector output, is tested as the model part of the decision support system. It must be noted that, as mentioned in section 3.1.2 as well, the communication component, and user interface of a formal decision support system are not developed in this thesis. Also, as introduced in section 3.1.2 as well, it is recommended for future work to formalise the diagnosis module as a way to formalise the integral method as decision support system.

3.2.4.2. Diagnosis module procedure

Altogether, the diagnosis module consists of four-step procedure. The steps are as discussed below:

0. The diagnosis module explicitly uses the output of the inefficiency detector module. Therefore, the inefficiency detector is included as step (0).
1. The diagnosis module starts with comparing the inefficiency detector outputs (the generated output) with the database. The database includes examples of inefficient performance: the data in the database consists of the data on (i) which generated performance is measured for a given inefficiency and problem, (ii) the number of times that inefficiency is found at a traffic signal controller, and (iii) the number of times that problem is diagnosed at a traffic signal controller. These inefficient performance examples are in fact the informal probabilities of the naïve Bayesian network. The comparison of the generated inefficiency detector outputs, and inefficient performance examples, is the output of this step.
2. Next, patterns are recognised using a decision tree. This step thus relates to the aforementioned conceptual pattern recognition. The patterns recognition compares the generated inefficiency detector outputs to each inefficient performance example. Each example relates to a specific problem, and countermeasure. The basic principle of the pattern recognition uses "if-then" rules, hence the decision tree: if a generated inefficiency detector output is similar to an example of inefficient performance, thus if the patterns are similar, as explained earlier in section 3.2.4.1, then the generated inefficiency is assigned to that problem. The number of "if-then" rules equals the number of problems that are possible, or included in the diagnosis module. The output of this step is that the generated inefficiency is diagnosed as the result of a problem.
3. Given the diagnosed problem(s), the diagnosis module proposes the corresponding countermeasure(s). This is also the output of this step.
4. In the fourth step, the proposed countermeasure(s) is/are implemented, thereby aiming at mitigating diagnosed problem is mitigated. The effect of the implementation of the proposed countermeasure(s) is also checked. This check is used to verify whether the problem was indeed mitigated, and whether this improved the performance of the traffic signal controller. If this is not the case, the check is used to propose new countermeasure(s). The check is done by using the new generated performance of the traffic signal controller with the proposed countermeasure(s) implemented to detect inefficiencies in the inefficiency detector. In fact, the whole inefficiency detector procedure, and diagnosis module procedure are repeated. The output might then be that either (a) no more inefficiencies were found, and thus that the proposed countermeasures were effective, or (b) that other inefficiencies are found. If the latter is the case, new countermeasures might be proposed. These new countermeasures are again tested on their effectiveness by redoing the inefficiency detector procedure, and diagnosis module procedure. If no (new) inefficiencies are found, or when

the performance of the traffic signal controller with the implemented countermeasure(s) complies to the road authority's policy demands, the diagnosis module is finished, and the next steps of the integral method are considered, as listed in section 3.2.

The steps of the diagnosis module, as discussed above, are visualised in Figure 3-4. The lined boxes represent the actions that are performed within a step, as well as the sequence of actions, given as arrows. The dotted boxes, and lines represent the data flow. The decision tree in step (2) consists of n decisions, and thus that n problems are included in the diagnosis module shown in Figure 3-4.

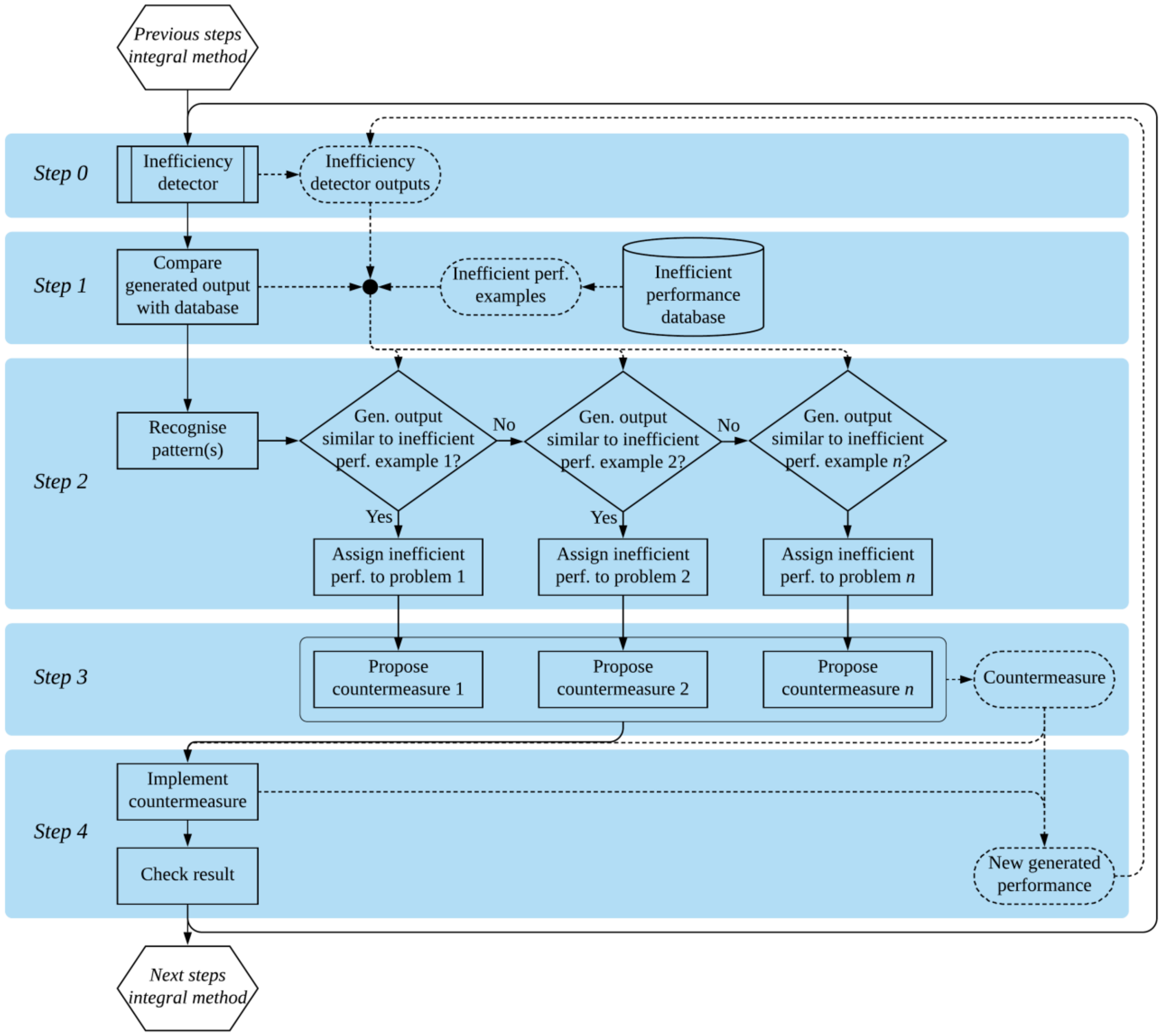

Figure 3-4 | Schematic overview of diagnosis module.

Furthermore, the cyclic approach of this diagnosis module ensures that potential "hidden" problems are found, and thus that the traffic signal controller is improved effectively. The "hidden" problems are problems that are being overshadowed by other problems. For instance, a problem (a) on a signal group might have a (much) larger effect on the traffic performance of the assessed traffic signal controller than problem (b). Using the database of inefficient performances, problem (a) might be diagnosed, and mitigated first. By re-doing the diagnosis module procedure, though now with problem (a) mitigated, problem (b) can be diagnosed, and mitigated. Thus, by checking the effectiveness of implementing the corresponding countermeasure, the traffic signal controller is evaluated again, and now the other, “hidden”, problem might be found, since there might be still inefficiencies found. Nonetheless, it must be noted that the diagnosis module proposes one countermeasure at a time: one iteration (steps (2) and (3)) of the diagnosis module results in one countermeasure that is proposed. This one proposed countermeasure does not have to be the countermeasure that is most effective. This implies that any specific order in which the proposed countermeasure(s) – as a result of multiple iterations of the diagnosis module – are to be implemented, is not considered in the diagnosis module as presented above. Therefore, it is recommended for future work to investigate how the diagnosis module can propose multiple countermeasures at a time, and in which order these proposed countermeasures can be implemented. Also, it is recommended to investigate whether a specific order of implementation is needed.

Furthermore, the diagnosis of the so-called “hidden” problems emphasises the resemblance to the plan-do-check-act cycle. Although, as stated before, the steps of the diagnosis method do not correspond explicitly with the four steps of this cycle, it can be seen that step (1) correspond to the step (i) plan of the cycle. Steps (2) and (3) relate to the cycle step (ii) do. Finally, step (4), thus implementation and restarting at step (0), corresponds to steps (iii) check, and (iv) act of the cycle.

3.3. Main findings

The main findings regarding the integral traffic signal controller evaluation and diagnosis method are:

- The integral traffic signal controller evaluation and diagnosis method is the integration of two major modules (the inefficiency detector, and diagnosis module) in a five-step method. The method is tested as a passive, medium-complex, knowledge-driven decision support system, using an informal naïve Bayesian network in combination with a decision tree. This means that the probabilities of the naïve Bayesian network are not formalised, and are used to define the “if-then” rules of the decision tree. Also, the method does not comply to the formal definition of a decision support system, because the communication component, and user interface are not developed in this thesis. Therefore, it is recommended to formalise the method, as well as its components, in future work.
- The first step focuses on the selection of multi-variable performance indicators, whereas an important aspect is that the selected multi-variable performance indicators can be defined in terms of a reference performance, or good performance.
- In the second step of the method, the reference performance models are calibrated to account for the inaccuracies of the reference performance models that follow from using certain assumptions in these performance models. This implies that the calibration of the reference performance models might not be needed for some multi-variable performance indicators. The calibration itself is three-step process in which the reference performance models are adjusted, using an error term, and bandwidth. To do so, the calibration uses data of an (assumed) problem-free traffic signal controller, which is defined as a traffic signal controller of which the overall performance is considered as good, and thus efficient. This means that the used data must originate from a period in which it is known, or assumed that the traffic signal controller performed good (problem-free). This can be data from e.g. shortly after the first implementation of the controller in practice, or from a period during which the road authority has concluded that the traffic performance was good. In the latter case, it is possible that there are problems present, hence the notation of an assumed problem-free traffic signal controller.
- The basic principle of the inefficiency detector (the third step of the evaluation and diagnosis method) is to identify those periods during which the generated performance deviated from the reference performance, which is defined as a generated performance that lays outside the bandwidth, or that is not equal to the reference performance when no bandwidth is computed. The module consists of a five-step procedure in

which step by step, the inefficiency detector outputs are generated. The procedure yields (i) a list of analysis periods during which the performance is considered as inefficient, (ii) the performance statistics, (iii) the inefficiency ratio (the ratio between the number of periods with inefficient performance with respect to the total number of analysed periods), and (iv) the evaluation score (a measure for the level of inefficient performance, given as a score on a scale from 1 to 10). That way, the inefficiency ratio is a measure for the period during which the performance was inefficient, as opposed to the evaluation score which is a measure for how inefficient the performance was. In general, it holds that the higher the inefficiency ratio, the lower the evaluation score is. However, this does not imply a fixed relationship: an extremely high inefficiency ratio does not necessarily mean an extremely low evaluation score, the evaluation score might be found to be slightly insufficient, thus slightly lower than 5.5.

- The diagnosis of problems is performed in the fourth step of the evaluation and diagnosis method, and relates to the diagnosis module. In the diagnosis method the informal naïve Bayesian network in combination with a decision tree is evident. That is, in the diagnosis module, given as a four-step procedure, a database with examples of inefficient performance (informal naïve Bayes) is used to define "if-then" rules (decision tree) as a way to diagnose problems. The database is thus used to recognise whether the observed pattern at the assessed traffic signal controller is similar to the patterns found in the database. If that is the case, the inefficiency can be assigned to the corresponding problem. The number of "if-then" rules equals the number of possible problems of the traffic signal controller. Because the pattern recognition is somewhat arbitrary, it is recommended to formalise the pattern recognition in future work.
- Consequently, countermeasures are proposed in the fourth step as well. The implementation of the countermeasures is part of the diagnosis module, and the integral method in general. That way, it can be checked whether the proposed countermeasures did successfully mitigate the diagnosed problems. That way, the diagnosis module, and the integral method is tested as a cyclic process, which enables the diagnosis of multiple problems for the same traffic signal controller. Because of this, problems of which the resulting inefficient performance are "overshadowed" by the inefficient performance of other problems, are found with this method as well, due to this cyclic process. However, the diagnosis module proposes one countermeasure per iteration, implying that it is recommended for future work to investigate how the diagnosis module can propose multiple countermeasures per iteration, as well as an order of implementation of the proposed countermeasures.
- The fifth step of the integral traffic signal controller evaluation and diagnosis method represents the starting point for a re-evaluation, as part of a periodic traffic signal controller maintenance policy. The output of this step can be used as input for the (assumed) problem-free traffic signal controller, as used in the calibration step, provided that data is collected shortly after the successful mitigation of problems. On the other hand, the traffic signal controller might also be re-evaluated after several years, because the changes in the traffic flow volumes, etc., over the years might result in a decreased performance of the traffic signal controller, provided that the settings are not adjusted accordingly. This emphasises the cyclic process that is included in the traffic signal controller evaluation and diagnosis method as well.

4. Method development

The integral traffic signal controller evaluation and diagnosis method presented in the previous chapter, represent the outlines of the method. That is, in the given steps, there are components that have to be discussed more in depth. For instance, the definitions of the reference performance models are not given in the previous chapter, nor are the problems that are diagnosed, which relates explicitly to the number of "if-then" rules in the diagnosis module. These components are discussed in this chapter as part of the development of the integral method.

First, the scope of the method is discussed, which relates to the selected problems, and multi-variable performance indicators and the limitations that follow from the selected problems, and multi-variable performance indicators. Next, the definitions of the reference performance models are discussed, followed by an introduction of the inaccuracies of the reference performance models. Fourthly, the calibration of the used the reference performance models is discussed, using a simulation study. The chapter concludes with some general remarks.

Additionally, for reference purposes, the variables, introduced in this chapter, are defined in *Notation and definitions* on page ix of this thesis as well.

4.1. Scope of the method

The scope of the method relates to the selected problems, and multi-variable performance indicators, as well as the limitations that follow from the selected multi-variable performance indicators in terms of applicability of the integral traffic signal controller evaluation and diagnosis method presented in the previous chapter. The selection of problems, and multi-variable performance indicators is made because the integral method includes a decision tree with as many decision rules ("if-then" rules) as there are performance issues at a traffic signal controller, see section 3.2.4.2. To limit the number of decision rules, a limited number of problems is selected. Furthermore, in literature, and in practice (e.g. in QQS and IQS evaluation reports) a wide variety of traffic signal control issues, and performance indicators are listed. Since it would cost too much time to include all, and thus test all, a selection is made. Besides, the use of a limited number of traffic signal control issues, and performance indicators to develop the traffic signal controller evaluation and diagnosis method is considered to fit best with the research objective, in particularly given the limited attention given to such a method in literature.

Given the limited number of studies on traffic signal control evaluation in terms of general traffic performance, as concluded in section 2.3, the selection of performance issues at a traffic signal controller is mainly based on findings in practice.

4.1.1. Traffic signal control performance issues in practice

As briefly introduced, the literature as discussed in section 2.1, and QQS and IQS evaluation reports provided by Vialis (n.d. [a]), for instance, consider a wide variety of performance indicators to evaluate traffic signal controllers. In the evaluation reports from practice, these performance indicators are used to identify and diagnose an equally varying set of traffic signal control issues, and problems. Based on the analysis of the evaluation reports from practice, it is concluded that there are three categories of problems in general, see Figure 4-1 on the next page: (1) incorrect control design (i.e. parameter settings, etc.), (2) technical issues, and (3) geometric intersection design. Then, (1) relates to the internal settings of the traffic signal controller, and detection, e.g. parameter settings. This control design relates directly to how the traffic signal controller controls traffic in terms of how much green is given per signal group, and when. The detection settings consist of the settings of detector parameters, e.g. function, and gap times. These settings might be incorrect for the considered traffic signal controller. For instance, the time-of-day programs are not initiated on the correct moment, or the gap times of the detector are incorrect. Secondly, (2) the technical issues relate to either detection, or other hardware, such as lamp posts, lights, etc. The former describes the issues a detector might have, other than parameter settings, e.g. detector overperformance. The other hardware issues relate to hardware other than the detectors, thus e.g. a signal light malfunction, or a damaged lamp post, etc. Lastly, (3) the geometric intersection design describes those issues that are caused by an inadequate geometric intersection design, e.g. a faulty intersection type, or too few, or too many lanes on a signal group.

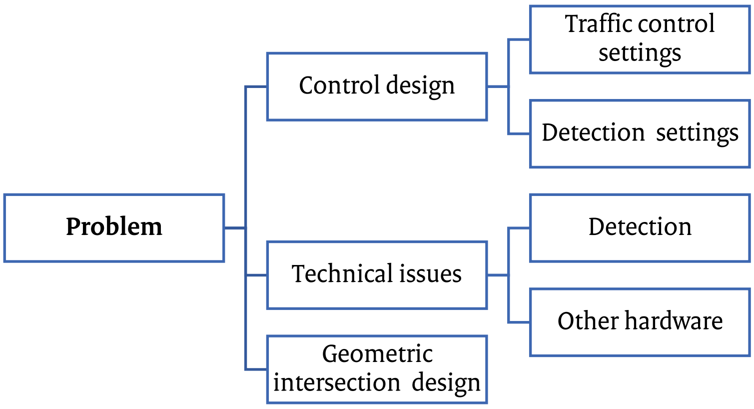

Figure 4-1 | Problem categories.

4.1.2. Selection of problems

Each of the problem categories, and corresponding sub-categories, as mentioned in Figure 4-1, relate to a set of potential issues, or problems. These potential issues are listed in Table 4-1 on the next page. The potential issues are based on the countermeasures as found in literature, and evaluation reports from practice. Given the aforementioned problem categories, and potential problems, a selection is made of problems that are investigated further in this thesis. This selection is based on the criteria as listed below:

a. The problems for which a countermeasure is most often mentioned in evaluation reports from practice, are included, based on the top-10 of most mentioned problems, as given in appendix B;
b. If two or more problems are comparable to each other (e.g. the problems both affect the way the green phase is lengthened), only one problem is selected, based on the other criteria;
c. Only problems that can be quantitatively measured in terms of the corresponding performance indicators are selected;
d. The limitations of computational reference performance models (as will be discussed in section 4.1.6) limit the problems that are selected as well;
e. The scope of the research (see section 1.4), thus e.g. only problems of vehicle-actuated traffic signal controllers are selected;
f. The available time for this research, that is, no more than four problems are selected, based on the other aforementioned criteria.

Given these criteria, the four problems printed in italic in Table 4-1 are selected. These problems are also further specified to relate them to a specific setting:

- Deactivated *alternative realisations* (green phase realisations other than primary realisations, see also appendix A);
- Maximum *green time settings*;
- *Detector gap times* of long detector;
- Inadequate *geometric intersection design*.

Table 4-1 | Overview of potential problems, with problems investigated in this research in *italic*.

Problem category	Potential issue/problem	Problem description
Control design: traffic control settings	*Alternative realisations* Coupled request/realisations[1] *Green time settings* Block sequence Wait-in-red/green[1] Time of day program (Urban) network control Congestion program Synchronised-/pre-start Clearance time settings Conflict handling	- Activated/deactivated program/setting (e.g. activated/deactivated alternative realisations, coupled request/realisation, etc.) - Incorrect timer settings (e.g. too high/low setting for maximum green time, synchronised-/pre-start, etc.) - Suboptimal settings (e.g. suboptimal block sequence, permitted/protected conflict handling, etc.)
Control design: detection settings	*Gap times* Extension function Request function	- Incorrect timer settings (e.g. too high/low setting for gap time) - Activated/deactivated extension/request function of detector(s) - Suboptimal function-specific settings (i.e. incorrect function on a detector, etc.)
Technical issues: detection	Inactive detector Sensitivity	- Detector malfunction - Detector oscillation[1] - Detector overperformance[1] - Detector underperformance[1]
Technical issues: other hardware	Various (e.g. hardware malfunction, etc.)	- E.g. traffic signal light malfunction
Geometric intersection design	Insufficient capacity	- E.g. insufficient number of lanes - Limited/no potential for (further) optimisation of traffic signal controller

It must be noted that geometric intersection design problems are included as potential problems because they represent the problems that lay "outside" of the traffic signal controller: the geometric intersection design problems represent the problems that limit e.g. the potential of the traffic signal controller to be optimised further. Nonetheless, these problems fit with the aforementioned selection criteria. Indeed, (a) considering the top 10 of most often mentioned countermeasures in evaluation reports from practice (see appendix B), all of the selected problems are in this top 10. Secondly, other problems, such as coupled request/realisation, time-of-day program, and detection functions, are excluded due to (b) the comparability of the problems, e.g. the setting for a coupled request/realisation is similar to the setting of an alternative realisation, that is, both affect the request for, and realisation of a green phase. Thirdly, (c) several problems cannot directly be measured using the considered performance indicators. Although they affect those performance indicators, the problem is not found based on data for these performance indicators. E.g., the setting for wait-in-red/green cannot easily be measured in terms of delay, for instance, during peak-hours, implying that it can be measured best during off-peak hours. In turn, this means that multiple time-of-day programs should be considered, which will cost more

[1] See appendix A for a clarification of the term.

time to investigate, thereby relating to (f) the available time for this thesis as well. Furthermore, conflict handling, among others, is not included because (d) the used computational reference performance models (as will be discussed in section 4.1.5) are only valid for signalised intersections with protected conflicts, for instance. Lastly, (e) this means that the potential issues due to technical issues are not investigated in this research, for technical inefficiencies, and/or malfunctions are not part of this thesis. In the same way, settings for (urban) network control are excluded, because only isolated intersections are investigated.

The specification of the problems as a way to relate them to one specific setting is done because e.g. "green time settings" is a rather wide definition. First, regarding the problem of detector gap time settings, only the gap times of the long detector are considered, given the IVER 2002 detector configuration given by Siteur, et al. (2002) (the detection configuration is shown in section 5.2.2.2). For the green time settings, the maximum green time is considered, because this is a setting that is not bound to guidelines. E.g. for the minimum green time, CROW (2006) provided guidelines with recommended minimum green times. Although they also provided a formula to determine the maximum green time, the formula can be interpreted, and adjusted to fit the road authority's demands, implying that the maximum green time is more "adjustable". Indeed, VIALIS uses a different maximum green time formula, as listed in appendix C. Because of this, the maximum green time is considered. Thirdly, for alternative realisations, it is already discussed that the problem is that alternative realisations are deactivated, because by default, alternative realisations are activated, as found in the evaluation reports from practice (VIALIS, n.d. [a]). Regarding the geometric intersection design, the problem is not specified: inadequate geometric intersection design. This is because the problem of geometric intersection design is given as insufficient capacity, for which many countermeasures (infrastructural adjustments) are possible, e.g. adding lanes. Therefore, the specific geometric intersection design elements are not considered.

4.1.3. Selected countermeasures

The selected problems are mitigated by countermeasures. However, based on an analysis of the evaluation reports from practice, it is found that one problem might have more than one countermeasure to mitigate the problem. Indeed, e.g. regarding alternative realisations, the problem might be that alternative realisations for a signal group are for instance deactivated, or the timer settings for the conditions for an alternative realisation of a signal group are sub-optimal. Each problem regarding alternative realisation relate to another countermeasure, for instance activating alternative realisations, or adjust alternative realisation timer settings. The same goes for the other problems. Therefore, a selection of countermeasures is made as well. The number of selected countermeasures equals the number of selected problems, thus four countermeasures are considered. In short, the following countermeasures are considered:

- Activate alternative realisations;
- Adjust maximum green time settings according to the computed maximum green time, as computed using the formula used by VIALIS, see appendix C;
- Adjust the long detector gap time to the correct setting (0.0 s[1]), as specified by Siteur, et al. (2002), given an IVER 2002 detection configuration;
- Adjust the geometric design of intersection.

Since for the adjustments of the maximum green time, and long detector gap time there are various incorrect settings, "degrees of incorrectness", or variants, possible, it is tested in a simulation study which specific

[1] In practice, road authorities might deviate from the guidelines, and use another gap time as the "correct" gap time, given that particular situation. However, for simplicity, the settings as given in the guidelines of Siteur, et al. (2002) are considered in this thesis.

incorrect setting is included, as a way to further specify the countermeasures. The simulation study compared the inefficient performance of traffic signal controllers caused by various variants of the selected problems with each other, and with the base case generated performance of the (assumed) problem-free controller, as explained in appendix D. For the long detector gap times, and maximum green time settings, these variants are most relevant, since here multiple variants are tested, see Figure 4-2. The problem of alternative realisations is tested to be deactivated alternative realisations, implying that only one variant is tested, see Figure 4-2. Although for the adjustments of the geometric design of intersection, there are also various countermeasures possible, this is not tested, see Figure 4-2. Instead, the general countermeasure to improve the allocation of capacity at the intersection by adjusting the geometric intersection design is considered. The selected incorrect settings, and corresponding countermeasures are listed in Table 4-2.

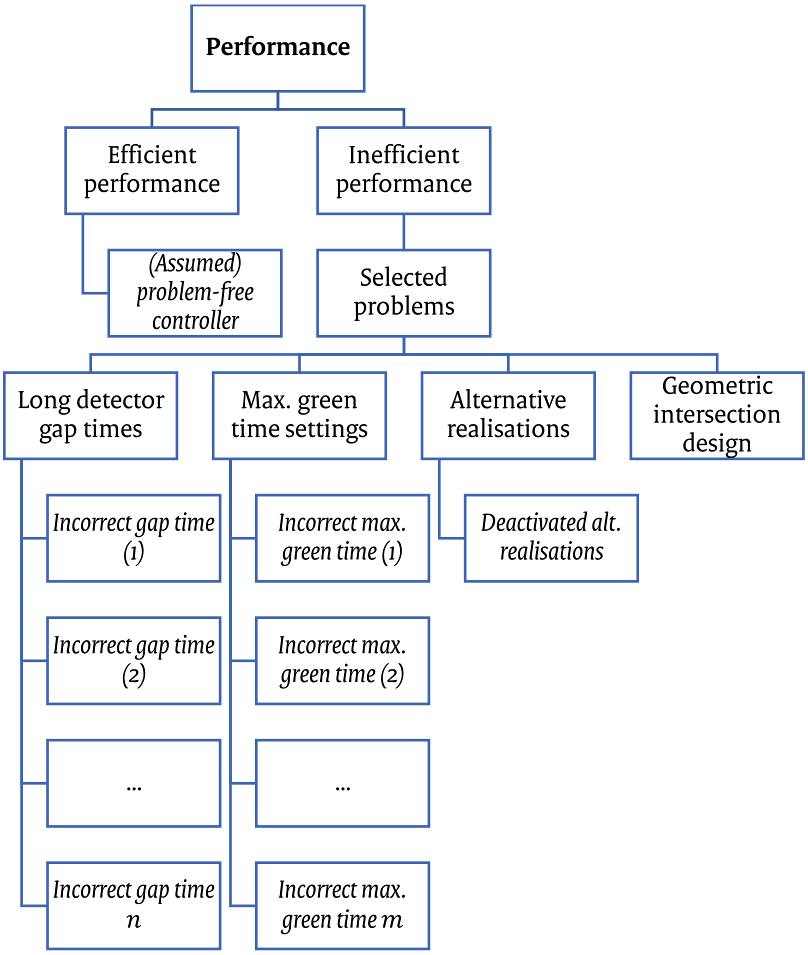

Figure 4-2 | Variants tested in the simulation study in appendix D, printed in *italic*, and the relation of the variants to the selected problems, and the overarching performance.

Table 4-2 | Selected problems with corresponding incorrect setting, and specified countermeasure.

Problem	Incorrect setting	Countermeasure
Long detector gap times	1.5 s	Adjust long detector gap times to 0.0 s
Maximum green time settings	50% of computed maximum green time	Adjust maximum green time settings to computed maximum green time
Alternative realisations	Deactivated	Activate alternative realisations
Geometric intersection design	Insufficient capacity	Adjust geometric design of intersection

4.1.4. Considered performance indicators

Based on the evaluation reports from practice, it is found that multiple performance indicators are used to find the problems, and in particular the traffic performance issues these problems cause. The considered potential issues present themselves as symptoms which are measured using the performance indicators. The used performance indicators in literature, and evaluation reports from practice relate to an overarching symptom: inefficient use of green. This conclusion is the result of the analysis of various evaluation reports from practice. The inefficient use of green consists of multiple aspects, whereas the core is that road users – vehicles, bicyclists, and pedestrians – do not utilise the green time as efficiently as possible. This relates to both undersaturated, and oversaturated conditions, per signal group. In the case of undersaturation, inefficient use of green is the result of unused green, while in the case of oversaturation it is the result of an inefficient green split. These are two different aspects, though with the same consequence. This means that during undersaturated conditions, thus when there are no vehicles (no queue left) at the start of red, too much green is given to one signal group, causing e.g. an "empty" intersection and thus additional delay on a conflicting signal group. This also implies that inefficient use of green might be the result of e.g. an inefficient block sequence. On the other hand, during oversaturated conditions, thus when there is a queue of one or more vehicles at the start of red, e.g. additional delay occurs due to oversaturation. Although oversaturation implies that all of the available green time of a signal group might be used, it also means that it is not enough green time. This might be the result of an exceedingly high traffic flow demand, or an inefficient allocation of the green time. The latter implies that the green time could have been allocated more efficiently, hence inefficient use of green in oversaturated conditions. It must be noted that in oversaturated conditions, the extent to which there is oversaturation plays a crucial role in the consequences for the considered signal group, and other (conflicting) signal groups, as well as the potential effect of a different allocation of the green time. Furthermore, as stated before, the symptoms are caused by different problems. Hence, the problems introduce the statement that symptoms are caused by either the traffic signal controller, and/or the geometric design of the intersection. The latter means that a traffic signal controller might be performing at the best of its abilities, and thus in terms of control as efficient as possible, but still resulting in inefficiencies (symptoms) in terms of traffic performance. These symptoms might be caused by the geometric design of the intersection, implying that the problem is this geometric intersection design. Therefore, this problem is included in this research, as discussed in section 4.1.3.

The finding that the overarching symptom of inefficient use of green can be found via various performance indicators, introduces that inefficient use of green presents itself as multiple other symptoms. In general, the two most mentioned performance indicators, and thus symptoms, are found to be delay, and unnecessary waiting. Unnecessary waiting is a measure for the credibility of the traffic signal controller, and how drivers "experience" the signalised intersection (see also appendix A): did drivers have to wait for "nothing", thus an (partial) "empty" intersection? Although some part of waiting for nothing is justified, because the intersection might be "empty" for traffic safety reasons, such as clearance times, the underlying principle is that unnecessary waiting should not occur on a signalised intersection. Furthermore, unnecessary waiting is part of the delay at an intersection. For delay it also holds that some delay is justified, because there is always the chance that one or more vehicles arrive during red, which have to stop, and wait for green, and thus stand in queue and experience delay. However, for both delay and unnecessary waiting, it holds that less is better. Both performance indicators are then a measure for the (in)efficient performance of signalised intersection. Also, other performance indicators – queue length, red-light running, etc. – follow from, and/or are related to delay, and/or unnecessary waiting. However, testing all possible performance indicators is considered as too time-consuming. Therefore, a selection is made of four performance indicators that are used to develop the integral evaluation and diagnosis method.

This selection is thus based on the time available for this thesis, but also on whether quantifiable definitions were found in literature: do the formal definitions in literature provide definitions that can be quantified and thus measured? If that is the case, the performance indicator is included. Also, those performance indicators are selected that show (formal) dependencies, or relations to other performance indicators. In other words, multi-variable performance indicators are selected. Lastly, because a simulation study is used to test the traffic signal controller evaluation and diagnosis method, performance indicators that cannot (easily) be measured in a simulation study are excluded. This is in particular true for red-light running. Given these criteria, the following four multi-variable performance indicators are selected:

- Degree of saturation;
- Delay;
- Phase failure;
- Queue length.

The formal definitions, as will be discussed in section 4.1.5, already show that all four multi-variable performance indicators depend on the single-value performance indicators traffic flow volume, cycle time, and green time. For each multi-variable performance indicator, a short definition will be discussed in section 4.1.5 as well, in relation to the reference performance models.

4.1.5. Reference performance model definitions

The definitions of each of the various multi-variable performance indicators are discussed in this section. The definitions include formulas that are used as reference performance models. The reference performance models all use traffic flow volume, cycle time, and green time as describing variables, among other multi-variable performance indicator specific describing variables. The elaborate formulas are listed in appendix C, since for delay, and queue length in particular, rather complicated computational models are used in this thesis.

4.1.5.1. Degree of saturation

The degree of saturation x [-] is a measure for the number of vehicles that are actually served by the traffic signal controller (CROW, 2006). The degree of saturation is computed per signal group j, and is thus a function of the measured traffic flow volume q [pce[1]/h], cycle time T_C [s], and green time T_G [s]:

$$x_j(q_j, T_{C,j}, T_{G,j}) = \frac{q_j T_{C,j}}{(T_{G,j} - \lambda_1 + \lambda_2)s_j} \tag{4-1}$$

with the saturation flow s [pce/h], green time start lag λ_1 [s], and green time end lag λ_2 [s] as fixed values per signal group. The green time lags are used to compute the effective green time $T_{G,\text{eff},j}$ [s], see also Figure 4-3:

$$T_{G,\text{eff},j} = T_{G,j} - \lambda_1 + \lambda_2 \tag{4-2}$$

However, in practice, this definition for the degree of saturation is not usually applied (VIALIS, n.d. [a]). Instead, the degree of saturation is expressed as oversaturation, which in this research corresponds to the definition of phase failure, as will be discussed in section 4.1.5.3.

[1] A passenger car equivalent (pce) is a converted unit, denoting the number of passenger cars that could pass a given point, or road section at an intersection instead of the given vehicles (bus, truck, etc.) in the time that vehicle uses (CROW, 2006), instead of vehicles per hour [veh/h], see also appendices A and C.

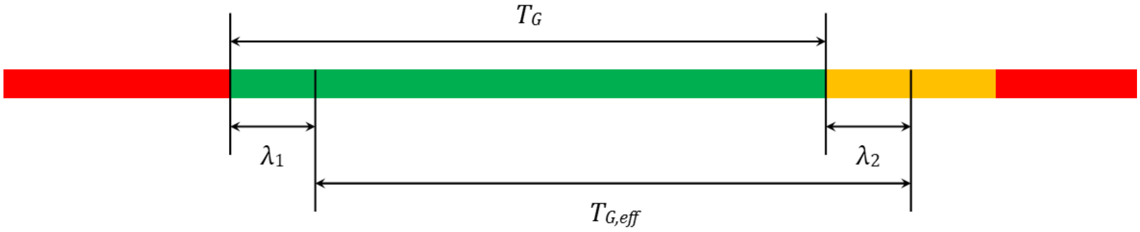

Figure 4-3 | Green time T_G and effective green time $T_{G,\text{eff}}$.

4.1.5.2. Delay

The delay d [s/pce] denotes the additional travel time a vehicle, or road user experiences when travelling through a network, or, in this case, when passing a signalised intersection, with respect to the free flow travel time. Regarding delays, an important distinction is to be made, as shortly discussed in chapter 1, namely between delay on the one hand, and stop delay on the other hand. The latter denotes the waiting time – the delay a vehicle experiences due to standing still –, whereas the former includes stop delay, acceleration delay, and deceleration delay (CROW, 2006; Dion, Rakha, & Kang, 2004), see Figure 4-4.

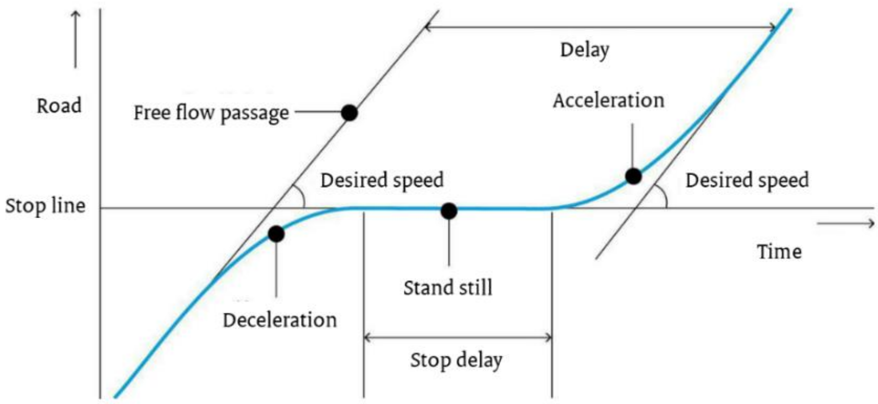

Figure 4-4 | Delay, deceleration delay, stop delay, and acceleration delay (CROW, 2006).

In literature, and in guidelines, various formulas are given to compute, or estimate the delay. The traffic signal controller evaluation and diagnosis method uses the delay formulas given in the Highway Capacity Manual 2000 (HCM2000) (TRB, 2000). The formulas are given in appendix C.

4.1.5.3. Phase failure

A phase failure denotes a green phase is terminated without fully serving the queue (Balke, et al., 2005). If a traffic signal controller is assumed to have no issues, this implies that during oversaturated conditions, the maximum green time is reached, thereby terminating the green and amber phase without fully serving the queue. Also, this means that at the start of the red phase, there is still a queue standing from the previous cycle: the initial queue (CROW, 2006). That way, an initial queue is a measure for congestion, given as oversaturation (VIALIS, n.d. [a]), and thus for phase failures as well. Given the aforementioned definition, it is tested that phase failures may only occur during oversaturation. If a phase failure occurs while there is no oversaturation, thus when $x \geq 1$, it is tested that the phase failure is a measure for a traffic signal control issue.

Phase failures ϕ [#] are measured using the detector states. If the detector at the stop line becomes active – there is a vehicle driving over, or standing on the stop line detector, and is thus driving past, or waiting for the stop line – within the first 2 seconds of the red phase, thus during the guaranteed red time $T_{R,\text{gar}} = 2$ s, a phase failure occurs, regardless of whether the vehicle joined the queue during amber, for instance. Some examples of phase failures are visualised in Figure 4-5. There it can be seen that during the first red phase T_R^1, the stop line detector becomes active shortly after the guaranteed red time $T_{R,\text{gar}}^1$. Then, there is no phase failure, thus $\phi^1 = 0$. However, during T_R^2, the stop line detector does become active during $T_{R,\text{gar}}^2$, thus resulting in $\phi^2 = 1$. In the following green phase, the queue is not completely served. This can be seen by the vehicle that stops for the stop line during the amber phase. That way, the stop line detector is already active before, and thus during T_R^3. Again, this is a phase failure, thus $\phi^3 = 1$, because the stop line detector is active during $T_{R,\text{gar}}^3$. Lastly, during T_R^4, the stop line detector becomes active again, shortly after the start of T_R^4, thus $\phi^4 = 1$.

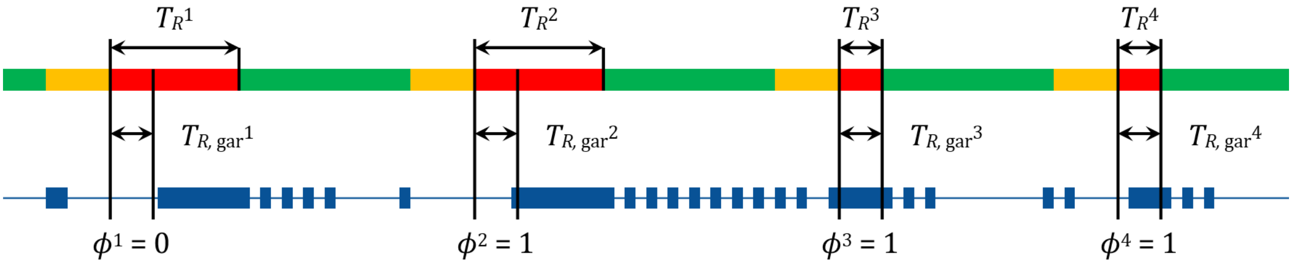

Figure 4-5 | Schematic overview of situations in which a phase failure occurs. The blue boxes and line represent the state of the stop line detector (inactive when there is a line, active when there is a block).

As introduced above, there are two categories of phase failures: phase failures during oversaturation, and other phase failures. This categorisation affects the reference performance for phase failures, because the phase failures during oversaturation are scored differently from other phase failures. That means that if a signal group is oversaturated, the reference performance for phase failures is equal to:

$$1 \leq r_{\phi,j}^{\tau} \leq n_{C,j}^{\tau} \qquad \text{if } x_j^{\tau} \geq 1\text{, otherwise } r_{\phi,j}^{\tau} = 0 \tag{4-3}$$

with reference performance for phase failures r_ϕ [#] of signal group j, number of cycles n_C [#] during the analysis period τ, and degree of saturation x. It must be noted that oversaturation is here thus defined as a degree of saturation equal to or higher than 1. Also, both the reference performance for phase failures, and the number of cycles are defined as counters (natural numbers). The number of cycles n_C^τ during an analysis period τ is the number of times a cycle time is measured during τ, and thus number of times the degree of saturation x is computed. If there is structural oversaturation, with $x \geq 1$ during each cycle in τ_1 (and thus an initial overflow queue during each cycle in τ_1), the reference performance for phase failures equals $r_\phi^{\tau_1} = n_C^{\tau_1}$. If during that another analysis period τ_2 with $n_C^\tau \geq 2$ cycles, only once oversaturation is measured, the reference performance for phase failures equals $r_\phi^{\tau_2} = 1$. It must be noted that this definition is formulated in this thesis.

4.1.5.4. Queue length

The queue length L [pce] denotes the number of vehicles standing in the queue at a signalised intersection. Just as with delay, various formulas to compute, or estimate the queue length are given in literature, and in guidelines. The traffic signal controller evaluation and diagnosis method uses the delay formulas given in the Highway Capacity Manual 2010 (HCM2010) (TRB, 2012). The formulas are given in appendix C.

4.1.6. Limitations

The integral traffic signal controller evaluation and diagnosis method has some limitations regarding the applicability of the method. The limitations mainly relate to the limitations of the reference performance models, and thus the definitions of the multi-variable performance indicators, as discussed in the previous section. The limitations are therefore considered as the conditions, and assumptions under which the method is valid. In short, the following limitations are considered, in alphabetical order:

- *Arrival pattern*: the formal definitions of delay, and queue length, in particular the formulas given in the HCM2000 (TRB, 2000), and HCM2010 (TRB, 2012) listed in appendix C, assume that the vehicles arrive either according to a uniform arrival pattern, or as platoon. Therefore, caution is needed when using the method as presented in this thesis on signalised intersections with other arrival types, for instance random arrivals. Because the case study in chapter 5 uses VISSIM, which simulates random, Poisson distributed arrivals by default (PTV AG, 2016), it is tested in section 4.4 how this alternative arrival pattern causes inaccuracies, and how these can be accounted for.
- *Coordinated signalised intersections and urban networks*: the scope of this thesis is to only consider isolated signalised intersections, whereas coordinated signalised intersections, and urban networks of signalised intersections not included, see section 1.4. Therefore, the evaluation and diagnosis method presented here is only valid for single, isolated signalised intersections. Future research is needed to make the method applicable for more complex signalised intersections, and urban networks of signalised intersections.
- *Conflict types*: at signalised intersections, different conflict types are considered. The formal definitions of delay, and queue length, as given by TRB (2000), and TRB (2012), are only valid for signal groups with protected conflicts (see appendices A, and C). Although the models can be altered to account for permitted conflicts (see appendices A, and C), this is not considered as part of the scope of this research because it does not show a different approach for the development of the traffic signal controller evaluation and diagnosis method. Therefore, the traffic signal controller evaluation and diagnosis method is concluded to be valid only when considering protected conflicts.
- *Evaluation scores*: the evaluation scores can only be computed for multi-variable performance indicators that use calibrated reference performance models, as stated in section 3.2.3.1. Because the reference performance models for delay, and queue length, as tested in this thesis, are the only reference performance models that are calibrated, it is concluded that only for delay, and queue length evaluation scores can be computed, thus for phase failures, and the degree of saturation no evaluation scores are computed.
- *Metering effect neighbouring (signalised) intersection*: the effect of the metering behaviour of a neighbouring (signalised) intersection is not investigated, given the scope of the thesis in section 1.4. This implies that the method may only be used for isolated intersections. In future research, the performance of the method can be investigated in a network of signalised intersections, thereby including the metering effect of an upstream (signalised) intersection.
- *Modes*: the formal definitions of delay, and queue length are based on the American practice regarding traffic signalisation. This implies that the formal definitions for these multi-variable performance indicators are only valid for signal groups with motorised traffic, thus no signal groups for active modes (TRB, 2000; TRB, 2012). Although it is not mentioned in relation to the definitions of the degree of saturation, and number of phase failures, it is also concluded that these definitions cannot be used for signal groups for active modes. Also, in practice, it is found to be practically impossible to collect reliable data for active modes in terms of traffic flow volume in particular (VIALIS, n.d. [a]). Therefore, the method presented in this thesis is only valid for signal groups with motorised traffic only. In future work, it is recommended to develop methods to collect more reliable data on active mode signal groups to make the integral evaluation and diagnosis method also applicable for active mode signal groups.

- *Reference performance models*: the formulas listed in appendix C are used as reference performance models of the selected multi-variable performance indicators. However, it must be noted that for the degree of saturation, the formula, given in Equation 4-1 in section 4.1.5.1, is not used as reference performance model, but rather as formula to determine the generated degree of saturation. This is due to the fact that the degree of saturation is included as a multi-variable performance indicator as result of a preliminary testing of the evaluation and diagnosis method. Therefore, the degree of saturation is considered as an integral part of the evaluation and diagnosis method, though without a reference performance model. As a consequence, inefficiencies cannot be detected using only the degree of saturation, at least not given the definitions presented in this thesis, and thus no inefficiency detector outputs are generated for the degree of saturation, limiting the applicability of this multi-variable performance indicator. Nevertheless, the degree of saturation is used to interpret the inefficiency detector results in the diagnosis module. It is recommended for future work to develop a reference performance model for the degree of saturation, as a way to include the degree of saturation explicitly in the inefficiency detector as well.

4.2. Definition of good performance

As introduced in section 3.1.1, the reference performance denotes the performance as computed, using the reference performance models, as explained in section 4.1.5, while the generated performance denotes the performance that is measured in practice. In section 3.1.1, it is also introduced that the reference performance represents good performance. This definition is further specified, given the aforementioned reference performance models.

The reference performance models all make use of the same describing variables, namely traffic flow volume, cycle time, and green time, as stated in section 4.1.5. However, the values of these describing variables can either be measured (e.g. in practice), or be prognosed by computation, based on the traffic flow volume – only prognoses for the cycle time, and green time may be made. In particular, this means that the cycle times, and green times can either be measured, or be prognosed. Both types of cycle time, and green time (measured, and prognosed) relate to a different definition of good performance: the computed reference performance using the measured cycle time, and green time might yield a different performance than when the reference performance is computed using prognosed cycle time, and green time. This is because the reference performance that follows from using the measured values for the cycle time, and green time represents the best performance a signalised intersection can generate, thus representing the performance of the signalised intersection at the best of its abilities in practice. On the other hand, the reference performance based on the prognoses for the cycle time, and green time represents the “utopian” performance of the signalised intersection. This is the reference performance if all the reference performance model inputs were computed based on the formulas, and definitions in theory, implying that is based on the results of an optimised, ideal traffic signal control model. However, the reference performance based on the prognosed cycle time, and green time can be a rather strict reference performance, in particular when compared to a reference performance using the measured values for the cycle time, and green time. At vehicle-actuated traffic signal controllers, the green time depends on the presence of vehicles, and is lengthened as long as there are vehicles approaching, until the maximum green time is reached, or a prioritised conflicting signal group has a request. Because of this, the cycle time varies per cycle as well (CROW, 2006). This may cause that, given the formulas for green time, and cycle time, the measured green time, and cycle time are longer than computed. This might be the result of the measured traffic flow volume, for instance when only one vehicle is present at a signal group (the minimum green time might be longer than necessary to serve this one vehicle, implying that the computed green time needed to serve one vehicle is shorter than the minimum green time). In some cases, measured the green time, and cycle time may be shorter than computed, e.g. in the case vehicles arrive as a platoon. In both cases, this implies that the reference performance based on the prognosed cycle time, and green time may not be able to handle several variable aspects of the traffic flow in practice. In other words, the prognosed cycle time, and green time may need certain assumptions as well, besides the assumptions that are already used in the reference performance models. Because this might cause extra inaccuracies in the computation of the reference performance in general, it is concluded to use the measured cycle time, and green time. Also, as stated above, this is considered as the "fairest" reference performance.

Altogether, this introduces the definition of good performance. As explained in section 3.1.1, the reference performance is defined as the expected performance, and is used to compare with the measured generated performance in order to detect inefficiencies. That way, the reference performance, thus as computed using the reference performance models, is defined as good performance. To determine what good performance is, and thus what the reference performance is, the measured cycle time, and green time are used, alongside the measured traffic flow volume. That way, the reference performance models use the measured values for the describing variables of the selected multi-variable performance indicators, as explained in section 3.2. If calibrated reference performance models, as discussed in section 3.2.2, are used, good performance includes the bandwidth as well. Therefore, taking the aforementioned considerations into account, the definition of good performance, as tested in this thesis, is formulated:

> **Good performance is defined as the reference performance (including the bandwidth) as computed with the (calibrated) reference performance model, using the measured traffic flow volume, cycle time, and green time as describing variables of the reference performance.**

It must be noted that this is *a* definition of good performance: the definition of good performance might differ if other choices were made, for instance using the cycle time, and green time prognoses to determine the reference performance, or with another evaluation method (e.g. other than using a reference performance). Given the limited time available, and the relation of the aforementioned BI-tool of VIALIS and the integral evaluation method presented in this thesis, the definition as formulated above is used to test the integral traffic signal controller evaluation and diagnosis method. This implies that the definition given here leaves room for future research: how can good performance be defined in another way? It is also recommended to combine this future work with the further development of the integral method.

Furthermore, the definition of good performance, as formulated above, includes the calibrated reference performance models. This introduces the frequency at which the reference performance must be calibrated. As discussed in section 3.2, the integral evaluation and diagnosis method calibrates several reference performance models, using data of an (assumed) problem-free traffic signal controller. It is stated that the data used to calibrate represent the base case generated performance, as generated by an (assumed) problem-free traffic signal controller. This (assumed) problem-free traffic signal controller might then be the traffic signal controller after a first evaluation using the method presented in this thesis. Using the data from this (successfully) evaluated traffic signal controller, the reference performance models are re-calibrated in a re-evaluation. This implies that re-calibration is done every time a traffic signal controller is evaluated. If the traffic conditions have changed over the years (e.g. a growth of the traffic flow volumes) while the traffic signal control settings did not change, the re-evaluation is expected to identify, and diagnose which problems this has caused, and which countermeasures are needed to improve the traffic signal control performance. In other words, re-calibration is only done as part of a re-evaluation, whereas data is used of the previously successfully evaluated traffic signal controller. This is based on the assumption that the calibration is mainly relevant to account for inaccuracies of the reference performance models in terms of the assumptions they use, as will be explained in section 4.3.

Nonetheless, it is possible to re-calibrate the reference performance intermediately, thus between evaluations, if the traffic conditions have changed drastically over the years. However, if such an intermediate re-calibration is done, it must be noted that the intermediate re-calibration explicitly assumes that the traffic signal controller is problem-free at that moment. This is closely related to what the road authority considers as efficient performance. In future work, it is recommended to investigate what the effect is of different frequencies of re-calibration.

4.3. Inaccuracies of reference performance model

As briefly introduced in sections 3.2.2, 4.1.5, and 4.2, the reference performance models use various assumptions. The assumptions might affect the way inefficiencies are detected in the inefficiency detector module. The underlying question is then what distinguishes inefficiencies in the traffic signal controller from inaccuracies in the reference performance model? To answer this question, it is important to emphasise that the inefficiency detector identifies an inefficiency as a deviation of the generated performance with respect to the reference performance. This deviation may be caused by either problems of the traffic signal controller (see section 4.1.2), or by inaccuracies of the reference performance model. This means that the assumptions of the reference performance model might not represent the traffic processes in practice correctly.

Besides the inaccuracies caused by the reference performance models, there are inaccuracies related to system observation. System observation relates to the way data is collected – the way the generated performance is measured. For instance, in practice, the delays are estimated based on loop detector data, rather than actually measured. Because simulation data is used in this thesis, the inaccuracies following from system observation do not a play role, although it is recommended for future research to assess these inaccuracies as well when using data from practice, or other data that might suffer from inaccuracies in the system observation.

Regarding the inaccuracies caused by the assumptions used in the reference performance models, it holds that if the reference performance model uses a different assumption than the data collection method, a discrepancy arises. This discrepancy might result in that the generated performance differs from the reference performance, which is defined as an inefficiency in section 3.1.1. However, it is assumed that if both the reference performance model, and the data collection method use the same assumption, the found inefficiencies are not related to the reference performance model, but are the result of the traffic signal controller, given as the problems discussed in section 4.1.2.

In short, this means that the assumptions used in the reference performance models affect the detected inefficiencies. Therefore, the inaccuracies of the reference performance model are defined as the deviations of the generated performance with respect to the reference performance that are caused by the use of (different) assumptions. In this research, several assumptions, as used in the reference performance models, are relevant. These assumptions are enumerated below:

- *Arrival pattern*: this relates to which arrival pattern is assumed in the reference performance model, especially regarding the reference performance models for delay and queue length, as introduced in section 4.1.6 as well. However, this also relates to how the generated delays, and queue lengths are estimated in practice using cumulative curves, and how they are measured exactly in a simulation environment. For instance, the reference performance models for delay, and queue length, given by TRB (2000), and TRB (2012) assume a uniform arrival pattern, while in VISSIM, the arrival pattern is random (Poisson distributed) by default (PTV AG, 2016);

- *Green time lags*: this relates to the different values that can be used for the green time start, and end lags, respectively λ_1, and λ_2, see also Figure 4-3 in section 4.1.5.1, whereas the end lags are most relevant. Indeed, for the λ_1, there is consensus in literature, while for λ_2 – also known as utilised amber time –, various values can be used, based on the road authority's policy: a shorter gap time leads to a higher utilisation of the amber time, and thus a higher green time end lag (Krol, Prinsen, & Misdom, 2018). The value of λ_2 is then affected by the detection gap time settings. It must be noted that the end lag is not a setting that can be influenced, but is the result of such an adjustable setting, in this case the detector gap time. For λ_1, it is mentioned in Dutch guidelines that this value depends on the considered type of movement of a signal group: for through-going movements, $\lambda_1 = 3$ s, while for turning movements $\lambda_1 = 5$ s, at most (Grontmij, 2001). Because of simplicity, $\lambda_1 = 3$ s is assumed for all signal groups. For the green time end lag, no value is assumed. Instead, the green time end lag is measured in the simulation model;
- *Block sequence dependent cycle time*: a different sequence of blocks results in a different cycle time, due to differences in the internal lost time between two consecutive blocks. Nonetheless, it is assumed that the block sequence of the traffic signal controller tested in the case study in chapter 5, is the most efficient block sequence, and thus that the cycle time is also as efficient as possible when considering the internal lost time. This implies that for the traffic flow volume input sets, as discussed in chapter 5 as well, the block sequence is identical, and is assumed to be optimal in both cases. This assumption is based on an analysis in COCON (a software package used to design (fixed time) traffic signal controllers, which is widely used in the Netherlands, see also appendix A) for the studied intersection, using different traffic flow volume input sets;
- *Acceleration-deceleration delay*: the computation of the reference queue length includes the acceleration-deceleration delay explicitly, as explained in appendix C. The acceleration-deceleration delay is computed as well (TRB, 2012). This computation uses various describing variables, with assumed values. The values are based on the default settings in VISSIM, as used in the simulation studies in the section 4.4, and chapter 5. As a result, the acceleration-deceleration has an assumed value as well. The assumed acceleration-deceleration delay equals $d_{\text{ad}} = 6.29$ s/pce.

Of these four assumptions, the assumption on the arrival patterns is investigated more in depth in the following section 4.4. The other assumptions are not investigated in detail. The assumption on the green time lags is not investigated in depth, because the assumption is affected by the detector gap time settings, whereas the value of the green time end lag is measured, rather than assumed. Also, the detector gap time setting is one of the problems that is tested in this thesis, as discussed in section 4.1. The assumption on the block sequence relates to both the generated performance, and reference performance. Since in both performances the same assumption is used, it is concluded that the detected inefficiencies are not caused by this assumption, but rather by the problems as listed in section 4.1. The same holds for the assumption on the acceleration-deceleration delay, because the reference performance model uses the same assumptions as the way the generated performance is measured. Nonetheless, it must be noted that other values for the acceleration rate, deceleration rate, and threshold speed of a stopped vehicle might be chosen. For simplicity, and given the limited time for this research, the values as used by default in the simulation model are used. In future work, it is recommended to investigate different values for the acceleration-deceleration delay.

4.4. Calibration: assessment of accuracy of reference performance models

In the previous section, it is stated that the reference performance models include several inaccuracies, in particular the reference performance models for delay, and queue length, as tested in this thesis. The inaccuracies are caused by assumptions, of which the assumption on arrival patterns is investigated more elaborately. This assumption relates to the arrival pattern assumed in the HCM2000, and HCM2010 models to compute respectively the delay, and queue length. The arrival pattern describes the way vehicles arrive at the intersection. In particular, this is about the process, or statistical distribution that describes best the arrival pattern. Usually in practice, the arrival pattern cannot be described completely as a specific (statistical) distribution, because the fluctuations in the arrival pattern are too big, or unpredictable. In the reference performance models of delay, and queue length, a uniform arrival pattern is assumed (TRB, 2000; TRB, 2012). However, in practice, a random (Poisson distributed) arrival pattern is more commonly assumed. Indeed, because this assumed to resemble practice better, VISSIM uses random arrivals by default (PTV AG, 2016). Therefore, a discrepancy exists between practice, and the reference performance models regarding this particular assumption. The effect of this discrepancy is investigated in this section, as part of the assessment of the accuracy of the reference performance models. It must be noted that this assessment is in fact the calibration step as explained in section 3.2.2. This implies that the steps of the calibration step are used to assess the accuracy of the reference performance models. The calibration discussed in this section, uses data of a simulation study.

4.4.1. Simulation study objective

The objective of the simulation study is to calibrate the reference performance models for delay, and queue length, given by TRB (2000), and TRB (2012), assessed in relation to the assumption on the arrival pattern. This is basically a comparison of the outcomes of the reference performance models using the theoretically assumed arrival pattern, versus the assumed arrival pattern in practice. A total of three arrival patterns are considered:

- Uniform arrivals;
- Random arrivals (Poisson-distributed arrivals);
- Platoon arrivals.

For every arrival pattern, the error terms, and bandwidths are computed for delay d, and queue length L, using the equations given in section 3.2.2.1:

$$\varepsilon_{d,j} = \frac{1}{n_\tau} \cdot \sum_{\tau} \left(r_{d,j}^{\tau}\left(q_j^{\tau}, T_{C,j}^{\tau}, T_{G,j}^{\tau}\right) - p_{d,j}^{\tau,*} \right) \tag{4-4}$$

$$\varepsilon_{L,j} = \frac{1}{n_\tau} \cdot \sum_{\tau} \left(r_{L,j}^{\tau}\left(q_j^{\tau}, T_{C,j}^{\tau}, T_{G,j}^{\tau}\right) - p_{L,j}^{\tau,*} \right) \tag{4-5}$$

$$B_{d,j} = \left[r_{d,j}\left(q_j, T_{C,j}, T_{G,j}\right) - 2\sigma_{d,j} \;\;,\;\; r_{d,j}\left(q_j, T_{C,j}, T_{G,j}\right) + 2\sigma_{d,j} \right] \tag{4-6}$$

$$B_{L,j} = \left[r_{L,j}\left(q_j, T_{C,j}, T_{G,j}\right) - 2\sigma_{L,j} \;\;,\;\; r_{L,j}\left(q_j, T_{C,j}, T_{G,j}\right) + 2\sigma_{L,j} \right] \tag{4-7}$$

with error terms ε_d [s/pce] and ε_L [pce], and bandwidths B_d [s/pce] and B_L [pce] for respectively delay and queue length. The reference performances are given as r_d [s/pce] for delay and r_L [pce] for queue length, both as function of the measured describing variables traffic flow volume q [pce/h], cycle time T_C [s], and green time T_G [s], the generated base case performances as p_d^* [s/pce] and p_L^* [pce] for respectively delay and queue length, and the standard deviations σ_d [s/pce] and σ_L [pce] for delay and queue length, all per signal group j, and with the number of analysis periods n_τ, and analysis period index τ.

4.4.2. Simulation study set-up

The general simulation set-up is as discussed in appendix E. However, for this simulation study, a couple of alterations are relevant. These alterations relate to the network lay-out, the simulation period, and model inputs in terms of traffic flow volumes and traffic signal control settings.

4.4.2.1. Network lay-out

The three arrival patterns are simulated on two networks. The uniform arrival pattern, and random arrival pattern are simulated on the network as will be explained in section 5.2.1, and visualised in Figure 5-1. For the platoon arrivals, a second network is simulated, see Figure 4-6 on the next page. It must be noted that the network in Figure 4-6 includes elements that are implemented in the network in Figure 5-1 as well, to be able to create an uniform arrival pattern. The network in Figure 4-6 consists of a single link, with one lane, on which in total three stop lines for two traffic signal controllers are placed. The traffic signal controllers in the are mainly used to create alternative arrival patterns. This is due to the fact that VISSIM simulates random, Poisson distributed arrivals by default (PTV AG, 2016). Moreover, this default setting cannot be changed, implying the need of the aforementioned uniform arrival creator, given as traffic signal controller (a) in Figure 4-6. Additionally, to create platoon arrivals, a second stop line (b) is used which simulates the exact same traffic signal controller as the controller simulated at stop line (c), see Figure 4-6. That way, platoons of vehicles depart at stop line (b), and arrive as platoon at stop line (c). The controller at stop line (c) is thus used to assess the arrival pattern. Also, by varying the distance (d) between stop lines (b) and (c), various proportions of arrivals during green are created, enabling a more thorough investigation.

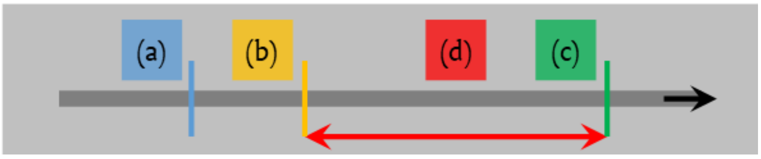

Figure 4-6 | Network lay-out to simulate platoon arrivals, with (a) uniform arrival creator, (b) uniform arrival traffic signal controller, and (c) platoon arrival traffic signal controller with (d) varying distance of (c) with respect to (b).

As stated, the uniform arrival creator is also implemented in the network discussed in section 5.2.1, to simulate uniform arrivals. The uniform arrival creators in that network are implemented per signal group, and are located directly downstream of the *Vehicle Inputs* (the location where vehicles are put into the network).

Lastly, it must be noted that the uniform arrival creator did not result in a stable arrival pattern, implying that an exactly uniform arrival pattern could not be made. It is observed that at some moments, a vehicle did not depart at the uniform arrival creator because it reacted to late. Also, it is found that it is quite difficult to fix the speed, and thus the headways of the vehicles in VISSIM, causing fluctuations in the arrival pattern. Therefore, a uniform arrival pattern is approached, rather than created precisely.

4.4.2.2. Simulation period

The networks are simulated using 10 runs, whereas each run of the simulation consists of 4200 simulation seconds (70 simulation minutes). In the data collection, and data analysis, only the middle hour is used, thus the first, and the last 300 simulation seconds (5 simulation minutes each) are not included. That way, potential disturbances in the data due to the filling, and emptying of the network are excluded from the data analysis. This means that data is collected, and analysed for a simulation period of 3600 seconds (60 simulation minutes). Also, this enables a data analysis during the period when the signalised intersection is fully functional.

The number of runs determines the reliability of the evaluation results. The number of runs is computed using a limited number of pilot runs, and is based on the standard deviation of the data collected in these pilot runs. This results in the equations given by WSDOT (2014), and listed in appendix E.2. Given these formulas, a maximum number of 10 runs is selected. Using delay data from 10 pilot runs, a standard deviation of the pilot sample of 0.18 s/pce is found. When this standard deviation equals the accepted standard deviation, it was found that the reliability approaches 99%. Otherwise, with an accepted standard deviation of 0.13 s/pce, the corresponding reliability approaches 95%. Therefore, 10 simulation runs are used.

4.4.2.3. Model inputs: traffic flow volumes and traffic signal controller settings

The traffic flow volume input sets are limited to how uniform arrivals are created in VISSIM, because the resulting traffic flow volume in [pce/h] departing at the uniform arrival creator – stop line (a) in Figure 4-6 – depends on the cycle time of this fixed time traffic signal dummy-controller. Then, in total three traffic flow volume input sets are simulated, see Table 4-3. It must be noted that these traffic flow volume input sets are only used in the network as given in Figure 5-1. In the network shown in Figure 4-6, only one traffic flow input set, i.e. 600 pce/h, is simulated. This done because in this network platoons are assessed. In order to create platoons of sufficient size, only one traffic flow volume is considered.

Table 4-3 | Traffic flow volume input sets considered in the simulation study, with corresponding traffic flow rates, and cycle times, as used in the uniform arrival creator.

Traffic flow volume q [pce/h]	Traffic flow rate q' [pce/s]	Cycle time T_C [s]
600	0.167	6
300	0.083	12
150	0.042	24

This introduces the use specific traffic signal controller settings for the controllers used to create uniform arrivals, and platoon arrivals. This means that in the network depicted in Figure 4-6, different traffic signal controller settings are used. Indeed, these controllers are fixed time controllers. The assessed traffic signal controller in the network, as given in Figure 5-1, is a vehicle actuated traffic signal controller which uses the settings as will be mentioned in section 5.2.2.2.

In short, for the uniform arrival creator, and platoon arrival creator, the settings are as enumerated below:

- The fixed time traffic signal controller used to create uniform arrivals is designed with a fixed green time of 1 second, with no amber time. The cycle time is a measure for the created traffic flow volume, see Table 4-3;
- The fixed time traffic signal controller used to create, and assess platoon arrivals is designed with a cycle time of 100 seconds ($T_C = 100$ s), a green time of 40 seconds ($T_G = 40$ s), and amber time of 3 seconds ($T_A = 3$ s). It is assumed that the green time start, and lags are equal, thus $\lambda_1 = \lambda_2$, thereby yielding that the effective green time $T_{G,\text{eff}}$ equals the green time, thus $T_G = T_{G,\text{eff}} = 40$ s.
- To assess an exceptional good platoon arrival pattern, the distance (d) between the stop lines (b) and (c) in Figure 4-6 equals the distance a vehicle uses for a travel time equal to the cycle time, thus 100 seconds, whereas the distance also depends on the speed. To create other platoon ratios, the distance (d) is varied equal to a varying travel time within the range of 50-150 seconds, with a step size of 1 second.

4.4.3. Results

The results of the simulation study are plotted in Figure 4-7, and Figure 4-8 for uniform arrivals, and random arrivals, and in Figure 4-9 for platoon arrivals, on the next pages. It must be noted that the plots in Figure 4-7, and Figure 4-8 are based on the simulation results from the network as shown in Figure 5-1, and that the plots in Figure 4-9 are based on the network as depicted in Figure 4-6.

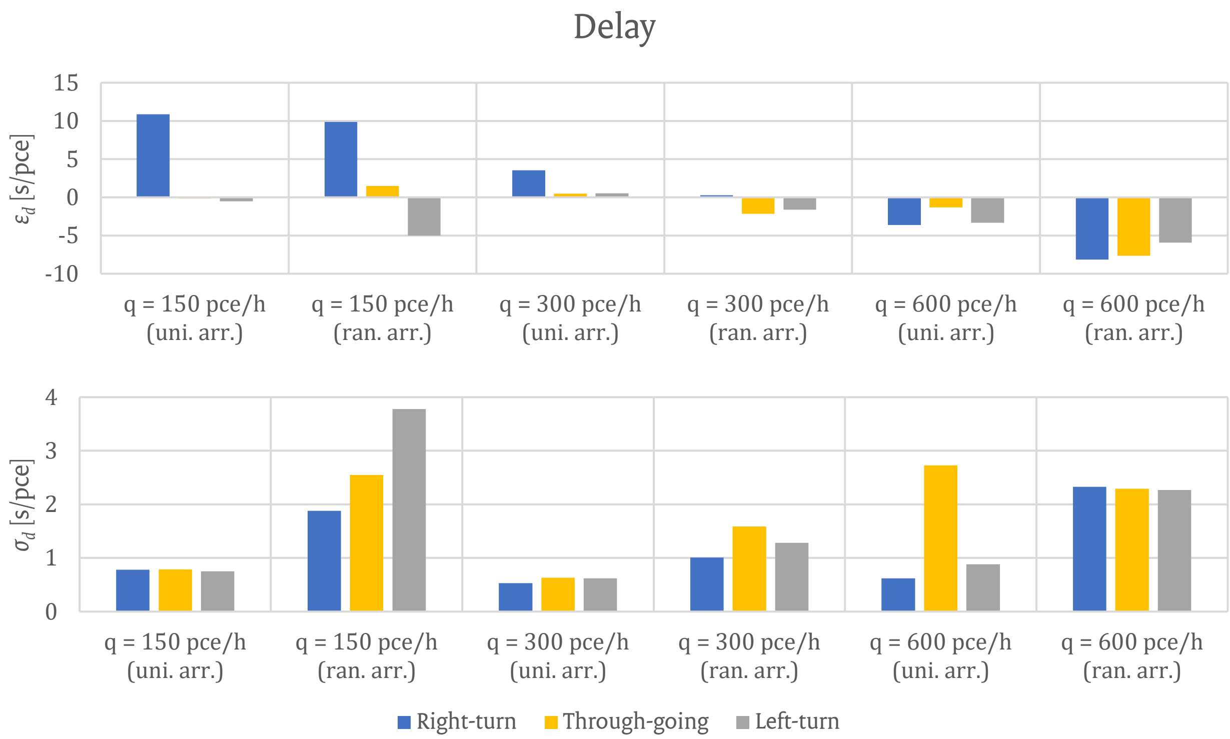

Figure 4-7 | Simulation study results: average error terms for delay ε_d (top), and standard deviation (st. dev.) for delay σ_d (bottom), for uniform arrivals (uni. arr.), and random arrivals (ran. arr.), per traffic flow volume q.

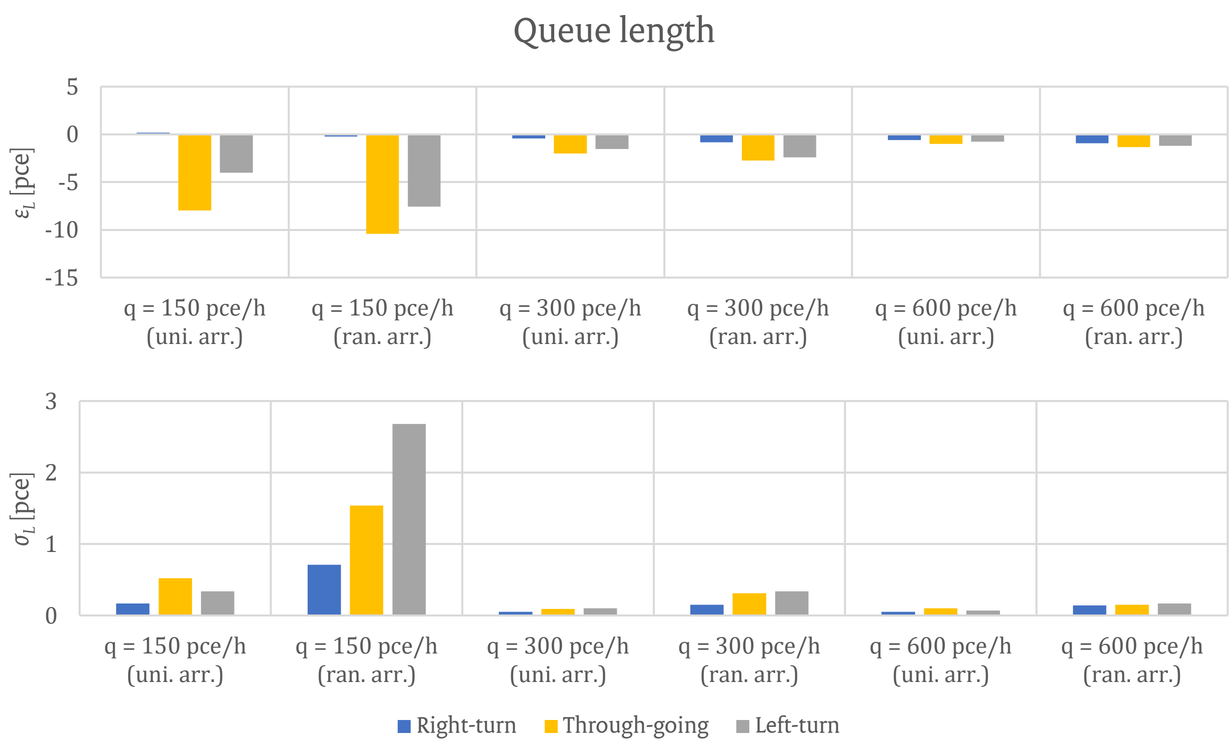

Figure 4-8 | Simulation study results: average error terms for queue length ε_L (top), and standard deviation (st. dev.) for queue length σ_L (bottom), for uniform arrivals (uni. arr.), and random arrivals (ran. arr.), per traffic flow volume q.

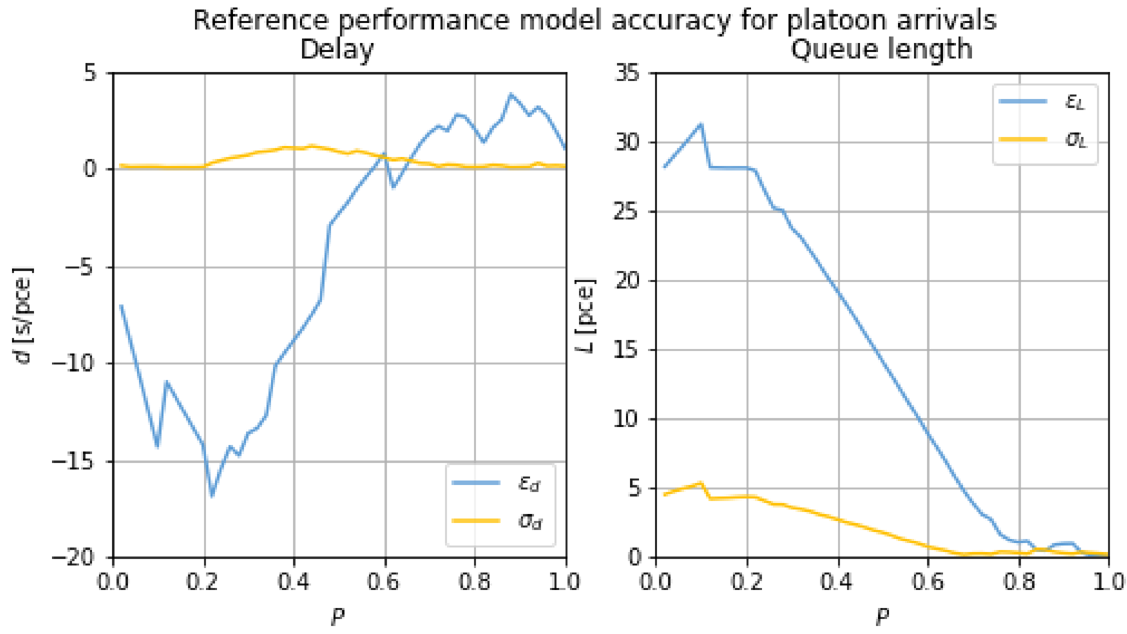

Figure 4-9 | Simulation study results: error terms ε, and standard deviations σ of delay d (left), and queue length L (right), for platoon arrivals, as function of proportions of arrivals during green P [-][1].

The results in Figure 4-7, and Figure 4-8 are given per direction, thus right-turning, through-going, and left-turning, rather than per individual signal group. This means that the results for e.g. signal groups 01, 04, 07, and 10 are averaged to determine the result for right-turning signal groups. The results per signal group are given in appendix F.1.

In the plots in Figure 4-7, Figure 4-8, and Figure 4-9, a positive error terms for delay, and/or queue length (e.g. $\varepsilon_d > 0$) can be observed, as well as a negative error terms (e.g. $\varepsilon_d < 0$). Both are considered as errors caused by inaccuracies of the reference performance models. A positive error term represents the situation in which reference performance model overestimates the generated performance (e.g. reference delay is larger than generated delay), and a negative error term represents an underestimation (e.g. reference queue length is smaller than generated queue length).

4.4.4. Conclusions

The results of the simulation study show that the different arrival patterns do result in different deviations of the delay, and queue length, in terms of the respective error terms. For instance, see Figure 4-7, the standard deviations of the difference in delay of uniform arrivals are smaller than those of random arrivals. The same trend is observed for queue lengths, implying smaller bandwidths. Also, the error terms are smaller for uniform arrivals when compared to random arrivals. Therefore, it is concluded that the assumed arrival pattern in the reference performance models – uniform arrivals – perform quite good, in particular with respect to random arrivals. Furthermore, it can be seen that each direction – right-turning, through-going, and left-turning – yields a different result in terms of average error terms. When looking at the results for platoon arrivals, it can be seen that the reference performance models tend to be more accurate for higher proportions of vehicles arriving during green. That means that the more vehicles arrive during green (higher value for proportions of arrivals

[1] The variable P [-] denotes the proportion of vehicles arriving during effective green ($T_{G,\text{eff}} = T_G - \lambda_1 + \lambda_2$), as defined in section 4.1.5.1, and thus denotes the number of vehicles q_G [pce/h] that arrived during the green phase with respect to the total traffic flow volume q [pce/h]. The proportion of vehicles arriving during effective green is defined as function of the effective green time $T_{G,\text{eff}}$ [s], and cycle time T_C [s] as well: $P = q_G T_{G,\text{eff}} / q T_C$, see also appendix C (TRB, 2012).

during green[1] P), the error terms for delay, and queue length, as well as the standard deviations, become smaller. Although platoon arrivals are not assumed in the reference performance models by default, it can be seen that the variables that account for platoon arrivals (see appendix C) do result in a fairly well estimated delays, and queue lengths. In other words, the reference performance are quite well able to compute a reference delay, and queue length, for both uniform, and platoon arrivals.

Regarding the results for uniform arrivals, it can be seen that the reference performance models do still result in a deviation between the reference performance, and generated performance. Although this might be caused by inaccuracies unrelated to the arrival pattern, it must be noted that it might also be caused by the fact that an exact uniform arrival pattern was not created. Indeed, as stated in section 4.4.2.1, a uniform arrival pattern is approached, due to the limitations of VISSIM with respect to arrival patterns other than random arrivals. This means that the found error terms for uniform arrivals, might be the result of the fluctuations in the approached uniform arrival pattern.

Therefore, given the results of the calibration presented in Figure 4-7, Figure 4-8, and Figure 4-9, it is concluded that uniform arrivals, and platoon arrivals – the arrival patterns assumed in the reference performance models for delay and queue length – tend to yield smaller error terms, and standard deviations, as used to define the bandwidth: a uniform arrival pattern results in the smallest differences between the reference performance, and generated performance for both delay, and queue length when compared to random arrivals, or even platoon arrivals. However, given the fact that fluctuations in the uniform arrivals are observed, due to the unstable uniform arrival creator, it is concluded to use random arrivals instead. When using random arrivals, it is important to determine error terms, and bandwidths to account for the inaccuracies in the reference performance models, since these models assume a different arrival pattern. Lastly, it must be noted that even though isolated signalised intersections are used in this thesis, as stated in sections 1.4 and 4.4.2.1, the reference performance models are able to handle platoon arrivals rather well.

4.5. Main findings

The main findings of this chapter are as follows:

- The scope of the method relates to the selected problems, and multi-variable performance indicators, as used to develop, and test the method from section 3.2 in this thesis. This selection is based on findings in literature (section 2.1), and evaluation reports from practice (VIALIS, n.d. [a]). Given these sources, a total of four problems is selected: deactivated alternative realisations, incorrect maximum green times, incorrect detector gap times, and inadequate geometric intersection design. For each problem, several countermeasures are possible, as well as various variants ("degrees of incorrectness"). This implies that the selected problems are further specified, as discussed in section 4.1. These problems are found using the four selected multi-variable performance indicators degree of saturation, delay, queue length, and phase failures, as defined in section 4.1.5.
- The use of the selected multi-variable performance indicators introduces several limitations, which mainly relate to the formal definitions of the selected multi-variable performance indicators, that are used to define the reference performance models. These limitations are listed in section 4.1.6, and include the assumed arrival pattern, intersection type, and modes, among others.
- The definition of good performance is based on the reference performance. As defined in sections 3.1.1, and 4.2, the reference performance denotes the expected performance, and is considered as efficient performance given the definition of an inefficiency in section 3.1.1 as well. However, the models to compute the reference performance use inputs (i.e. describing variables) that can be measured, or computed. Using

measured inputs yield different reference performance results than using computed inputs. To minimise the effect of assumptions on the reference performance, the definition of good performance is formulated as the reference performance as computed using measured inputs. It must be noted that different definition of good performance are possible as well. It is therefore recommended for future work to investigate in which other ways good performance can be defined. Besides, the definition of good performance includes the calibrated reference performance models, for which it is recommended to investigate the effect of different frequencies of re-calibration of these reference performance models as future work.

- The inaccuracies of the reference performance models relate to the assumptions used in the reference performance, as opposed to the assumptions used to determine the generated performance. This means that using the reference performance models, periods of inefficient performance might be the result of either the traffic signal controller (the selected problems in section 4.1), or the reference performance model itself. In short, the following assumptions are found to affect the accuracy of the reference performance models: (i) arrival pattern, (ii) green time lags, (iii) block sequence dependent cycle time, and (iv) acceleration-deceleration delay. The first assumption is investigated in more detail in section 4.4.
- The accuracy of the reference performance models with respect to the assumption of arrival patterns, is tested using a simulation study. This simulation study resembles the calibration step of the integral traffic signal controller evaluation and diagnosis method, as explained in section 3.2.2. Three arrival patterns are investigated: (i) uniform arrivals, (ii) random arrivals, and (iii) platoon arrivals, whereas it must be noted that the investigated reference performance models have parameters to handle uniform, and platoon arrival patterns. The simulation study results showed that it is for these two arrival patterns, the accuracy of the reference performance models is best. However, since it was also found that the simulation of uniform arrivals resulted in fairly unstable results, it is concluded to use random arrivals instead. Therefore, it is also concluded that when using random arrivals, it is important to calibrate the reference performance models by defining the error terms, and bandwidths.

5. Case study

The test of the integral traffic signal controller evaluation and diagnosis method is discussed in this chapter, using a case study in a simulation environment, in which various alternatives of an intersection are simulated with a faulty traffic signal controller with different problems implemented deliberately. The test focuses on whether the method presented in this thesis is able to detect the inefficiencies caused by the problems, and diagnose the problems correctly. Also, it is tested whether the diagnosed problems are mitigated successfully. Furthermore, the test of the method aims at identifying any potential improvements, and alike that can be investigated in future research regarding this method.

Thus, in short, this chapter discusses the testing of the evaluation and diagnosis method. The test makes use of a simulation study. Therefore, the simulation study objective is discussed first. Next, the simulation study set-up is discussed. Thirdly, a summary of the results is given. A more elaborated overview of the results is given in appendix F.2. Fourthly, the application of the traffic signal controller evaluation and diagnosis method is discussed in depth, by discussing the step-by-step use of the evaluation and diagnosis method for several alternatives, whereas some general findings are briefly discussed as well. Finally, the chapter concludes with a summary of the main findings, thereby also discussing some aspects that might be investigated in future work.

Additionally, for reference purposes, the variables, introduced in this chapter, are defined in *Notation and definitions* on page ix of this thesis as well.

5.1. Case study objective

As briefly introduced above, the objective of the case study is to test the traffic signal controller evaluation and diagnosis method, as defined in section 3.2. This is done by implementing the selected problems from section 4.1 in a traffic signal controller, for different signal groups, thus simulating faulty traffic signal controllers. By simulating these faulty controllers, and evaluating the output, it is tested whether or not the implemented problems can be traced back using the inefficiency detector, and diagnosis module, as parts of the integral evaluation and diagnosis method. That way, the testing of the evaluation and diagnosis method in this simulation study is thus meant to check whether or not the considered problems are indeed identifiable, and can thus be diagnosed. Additionally, the objective is also to assess the developed traffic signal controller evaluation and diagnosis method, and identify potential improvements to the integral method.

5.2. Case study set-up

The simulation study set-up is as discussed in appendix E, where the simulation period, number of runs, and data collection and processing methods are discussed.

5.2.1. Network lay-out

The case study uses a simulated network of a signalised intersection with four approaches. The network used for this signalised intersection is based on an existing signalised intersection in Tilburg, the Netherlands, although with some alterations to the distribution of lanes. These alterations are made to create a signalised intersection with an equal number of lanes per direction, per approach. Each approach has three signal groups yielding a total of twelve signal groups, as shown in Figure 5-1. The through-going signal groups have two lanes, and the left- and right-turning signal groups have one lane each. The traffic flow volumes per signal group are discussed in section 5.2.2.1. On this intersection, a vehicle actuated traffic signal controller is applied. The settings of this controller are discussed in section 5.2.2.2. The simulated arrival pattern is the default arrival pattern in VISSIM: random, Poisson-distributed arrivals. The speed limit on this intersection is 80 km/h, which is equal to the speed limit on the existing signalised intersection on which the network is based.

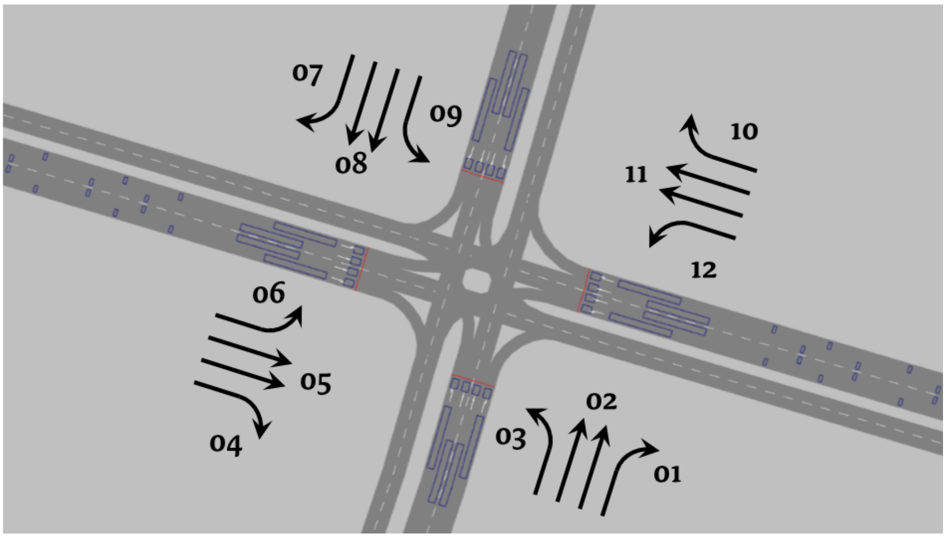

Figure 5-1 | Network lay-out of the simulated signalised intersection, including signal groups numbers, and number of lanes per signal group.

5.2.2. Model inputs

The inputs of the simulation model relate to the different traffic flow volumes that are investigated, and the traffic signal control settings.

5.2.2.1. Traffic flow volumes

The case study only considers passenger cars, implying that other vehicle types, e.g. trucks, are not simulated. This means that the traffic flow volume in [veh/h] equals the traffic flow volume in [pce/h]. Two traffic flow volume sets are considered: equal flows, and major-minor flows. The former represents the situation where each signal group has an identical traffic flow volume, in this case study 300 pce/h. The major-minor traffic flow volume set represents the situation where the through-going directions, given as signal groups 02 and 08, are the major road with the highest traffic flow volume (600 pce/h). All other signal groups are considered as minor directions with 150 pce/h. This traffic flow volume set is used to account for a different distribution of green time over all considered signal groups. Furthermore, it must be noted that the load ratios (the ratio of the traffic flow volume, and saturation flow, see also appendices A, and C) differ per signal group, because the through-going signal groups have two lanes, and thus a higher saturation flow with respect to the left- or right-turning signal groups.

5.2.2.2. Traffic signal controller settings

As stated in sections 1.4, 4.4.2, and 5.2.1, a vehicle actuated traffic signal controller is simulated. The traffic signal controller makes use of the detection configuration according to the IVER 2002 guidelines, given by Siteur, et al. (2002). This detection configuration uses a total of five detectors per lane. Directly upstream of the stop line, there is the stop line detector (a), followed by the long detector (b), and three distant detectors (c), (d), and (e), see Figure 5-2. The properties, and settings per detector, are listed in Table 5-1 on the next page. It must be noted that these settings are according to the definitions of Siteur, et al. (2002). In practice, location-specific settings might be used that deviate from these settings. However, for simplicity, this thesis assumes that the properties, and settings per detector according to the definitions of Siteur, et al. (2002) are correct, and sufficient.

Figure 5-2 | Detection configuration according to the IVER 2002 guidelines, given by Siteur, et al. (2002), with the stop line detector (a), followed by the long detector (b), and three distant detectors (c), (d), and (e).

Table 5-1 | Properties per detector according to the IVER 2002 detection configuration, including adjusted distances to the stop line for signal groups including/for turning movements, as defined by Siteur, et al. (2002) for a maximum speed of 80 km/h.

Detector	Detector length [m]	Distance to stop line[1] [m] (through)	Distance to stop line[1] [m] (turning)	Gap time [s]
Stop line detector (a)	3	4	4	2.5 (through) 3.1 (turning)
Long detector (b)	20	40	30	0
Distant detector 1 (c)	1	71	61	2
Distant detector 2 (d)	1	89	79	3
Distant detector 3 (e)	1	115	105	2

The settings regarding the green time, amber time, and red time, per signal group, are given in Table 5-2. It must be noted that the maximum green times per traffic flow volume set, as discussed in section 5.2.2.1, are as computed with the formula used by VIALIS (see appendix C), and are thus considered as the correct maximum green times. The other settings are based on guidelines provided by CROW (2006), for which it must be noted that amber times are based on a signalised intersection with a speed limit of 80 km/h, as stated in section 5.2.1.

Table 5-2 | Traffic signal controller timer settings per signal group: amber time minimum red time, minimum green times, and maximum green times (according to the equation of VIALIS, see appendix C, per traffic flow volume set).

Signal group	Amber time [s]	Min. red time [s]	Min. green time [s]	Max. green time: equal flows [s]	Max. green time: major-minor [s]
01	4	2	5	25	15
02	5	2	5	15	30
03	4	2	5	30	15
04	4	2	5	25	15
05	5	2	5	15	15
06	4	2	5	30	15
07	4	2	5	30	15
08	5	2	5	15	30
09	4	2	5	30	15
10	4	2	5	25	15
11	5	2	5	15	15
12	4	2	5	30	15

[1] The distance to the stop line differs for signal groups with turning movements, due to a lower speed at the stop line for turning signal groups with turning movements.

Furthermore, a fixed block sequence is applied. The block sequence for this signalised intersection is given in Table 5-3. The block sequence is based on the critical conflict group of this intersection, consisting of signal groups 02, 05, 09, and 12, as found in the analysis of the signalised intersection in COCON. It must be noted that with alternative realisations, a signal group might also have a green phase in any block other than the block listed in Table 5-3.

Table 5-3 | Block sequence.

Block I	Block II	Block III	Block IV
02, 08	03, 04, 09, 10	05, 11	01, 06, 07, 12

5.2.3. Data collection and processing

The data is collected using a duration of the analysis period of 300 s (5 minutes). This implies that a total of 120 analysis periods is considered, given the duration of the simulation period (36000 s, see appendix E.1). The data collection focuses on the multi-variable performance indicators, and their describing variables, as selected in section 4.1.4. Therefore, the following data is collected:

- Delays;
- Queue lengths;
- Detector states to determine phase failures;
- Traffic flow volumes, and arrivals per cycle phase (green, amber, red);
- Green times, amber times, red times, and cycle times;
- Utilised amber time.

The data collected as listed above is used to determine the generated performance, e.g. for delay. The data on traffic flow volumes, cycle time, and cycle time are used to compute reference performance as well. The evaluation reports from practice of VIALIS (n.d. [a]) indicate that all data listed above can be measured in practice as well, though the data of the detector states is then also used to estimate the delays, queue length, and utilised amber time.

The data processing is done using a PYTHON script. Because most of the aforementioned data can only be extracted from VISSIM as raw data, a data processing tool was made to generate useful data. Indeed, phase failures can only be found by combining the raw detector state data, with the raw traffic signal data. The same holds for data on arrivals per cycle phase, and utilised amber time. Moreover, it must be noted that instead of the default raw detector state data in VISSIM, the raw data of *data collection points* is used. These *data collection points* were placed on the exact same location as a detector, and are used because of the better accuracy of these data points – data per 0.01 s using *data collection points*, versus data per 1.00 s using raw detector state data – to resemble the detector state data collection in practice better, e.g. using the KWC and phase-log data (Figure 1-1), where detector state data is collected with an accuracy of 0.10 s (VIALIS, n.d. [a]), which is less accurate than the data collection points, but more accurate than raw detector state data from VISSIM.

It must be noted that the *queue counter measurements* in VISSIM make use of conditions. These conditions describe the situation for which it holds that a queue is present. The conditions relate to the speed in the queue, the headway, and maximum queue length, see Table 5-4 on the next page. As long as all these conditions are met, the *queue counter measurement* measures a queue. Thus, e.g., as soon as the speed reduces to less than 5.0 km/h, it is assumed that there is queue, which is not dispersed until the speed increases to more than 10.0 km/h. This means that, according to VISSIM, a queue might suddenly disappear if vehicles start driving faster than 10.0km/h, while this might still be considered as a queue in practice. This is best illustrated with the following example: if the last vehicle in the queue is relatively far upstream of the stop line, the queue length is high, but becomes zero as soon as the speed of the queue exceeds 10.0 km/h, while there is still a moving queue. Although different values for the aforementioned conditions might yield results that resemble queue lengths as measured in practice better, it is assumed that the default settings of VISSIM suffice. The queue length is then averaged per analysis period. This implies that the average queue length also includes the parts of the analysis period during which no queue was present.

Table 5-4 | *Queue counter measurement* conditions for which it holds that a queue is present.

Condition	Definition
Minimum speed in queue $v^Q_{\min}$ [km/h]	$v^Q_{\min} < 5.0$
Maximum speed in queue $v^Q_{\max}$ [km/h]	$v^Q_{\max} > 10.0$
Maximum headway in queue H^Q [m]	$H^Q = 20.0$
Maximum queue length $L_{\max}$ [m]	$L_{\max} = 500.0$

5.2.4. Alternatives

First, the set-up of the case study in terms of the alternatives, is that the case study is a half-blind study: for some alternatives, to author of this thesis it is known at the start of the application of the traffic signal controller evaluation and diagnosis method which problem is deliberately implemented on which signal group. For some alternatives, the so-called blind alternatives, this is unknown to the author. In the case of unknown problems, the objective is to detect inefficiencies, and diagnose the problem using only the output data (generated performance of the multi-variable performance indicators, and the describing variables used in the reference performance models) of the traffic signal controller, with respect to the reference performance. In other words, for the alternatives with unknown problems, the problems are to be found using only the data, and thus not by checking all the settings of the traffic signal controller manually. The problems in these blind alternatives are implemented by colleagues at VIALIS. That way, the problems are unknown initially to the author of this thesis, and are therefore only found using the integral evaluation and diagnosis method.

Secondly, alternatives with two problems are tested, or one problem on two signal groups. This means that the interaction of the problems is also tested. That way, it is investigated whether it the system can also identify more than one problem. This relates to the feedback loop in the diagnosis module, as discussed in section 3.2.4. The implementation of two problems enables a test of the concept of a cyclic diagnosis procedure.

It must be noted that although four problems are selected in section 4.1, only three are implemented deliberately. In other words, the problem of inadequate geometric intersection design is not tested explicitly. Instead, this problem is used as an exit condition of the feedback loop in the diagnosis module: if the detected inefficiencies cannot be assigned to one of the three other problems, the inefficiencies are assumed to be caused by inadequate geometric intersection design. This implies that this problem is a sort of "last resort" problem. Also, it implies that if inefficiencies are found in the base case traffic signal controller, representing the problem-free traffic signal controller, the inefficiencies are assumed to be caused by inadequate geometric intersection design as well.

Altogether, this resulted in a total of ten alternatives. Each alternative has a different traffic flow volume input set, as discussed in section 5.2.2.1, thus either equal flows, or major-minor flows. Three alternatives are blind alternatives, six alternatives have two problems implemented on purpose, and one base case alternative representing the problem-free traffic signal controller used in the calibration step. For the base case alternative, and blind alternatives, both traffic flow input sets are simulated. In the case of the blind alternatives, this has to do with the fact that e.g. the problem regarding the maximum green time relates to different maximum green times per traffic flow volume input set, whilst it is unknown whether this problem is implemented, and if so, for which traffic flow volume input set. An overview of the ten alternatives is given in Table 5-5. It must be noted that this table lists the alternatives, and problems as known at the start of the case study.

Table 5-5 | List of alternatives, with corresponding traffic flow volume input set, implemented problem(s), and signal group(s) with the considered problem(s), as known at the start of the case study.

Alternative	Traffic flow volume input set	Problem(s)	Signal group(s) with problem(s)
0 (**base case**)	Equal flows and major-minor flows	N/A	N/A
1	Equal flows	Alternative realisations	02 and 03
2	Major-minor flows	Maximum green time	04 and 06
3	Equal flows	Gap time	02 and 07
4	Major-minor flows	Alternative realisations Maximum green time	04 10
5	Equal flows	Maximum green time Gap time	05 09
6	Major-minor flows	Alternative realisations Gap time	07 11
7 (**blind alternative**)	Equal flows and major-minor flows	Unknown	Unknown
8 (**blind alternative**)	Equal flows and major-minor flows	Unknown	Unknown
9 (**blind alternative**)	Equal flows and major-minor flows	Unknown	Unknown

5.2.5. Database with examples of inefficient performance

To be able to evaluate and diagnose the traffic signal controllers as given in the alternatives discussed in section 5.2.4, a database with examples of inefficient performance is needed, as part of the diagnosis module, and as a formal part of the decision support system, as explained in sections 2.2.1, 3.1.2, and 3.2.4. The database is filled with data from examples of inefficient performance, as explained in section 3.2.4.1, for the signalised intersection, as given in section 5.2.1. Thus, the database consists of data of the same signalised intersection that is studied in this case study. The characteristics of the signalised intersection (e.g. traffic flow volume sets, (correct) traffic signal controller settings, etc.) in the database are identical. Moreover, data is used from the simulation study as used to select the countermeasures in section 4.1.3, and as discussed more elaborately in appendix D. This implies that the database is rather limited, for it only consists of examples of inefficient performance on signal groups 01, 02, and 03.

It must be emphasised that the database used in this case study is a fairly limited database with examples of inefficient performance, because (i) the examples are only given for signalised intersection given in section 5.2.1, (ii) it only includes examples for signal groups 01, 02, and 03, and (iii) it only includes examples of three selected problems. This means that the database with examples of inefficient performance used in this case study is not generic. However, it can be used to fill such a generic database. This is also true for the tested alternatives, and results: the data of the tested alternatives in this case study may be used to fill a more generic, and widely applicable database. Therefore, it is recommended for future work to create a more generic, and widely applicable database with examples of inefficient performance, to be used in future evaluations.

5.3. Results

The results of the case study are discussed first for the calibration. Next, the results per module, and per alternative are presented.

5.3.1. Calibration

The calibration of the reference performance models relate to the reference performance models for delay, and queue length explicitly. The resulting error terms, and bandwidths are given in Figure 5-3, and Figure 5-4 for respectively delay, and queue length. It must be noted that these results are identical to the results for random arrivals as discussed in section 4.4, because in both cases, the same problem-free traffic signal controller is used.

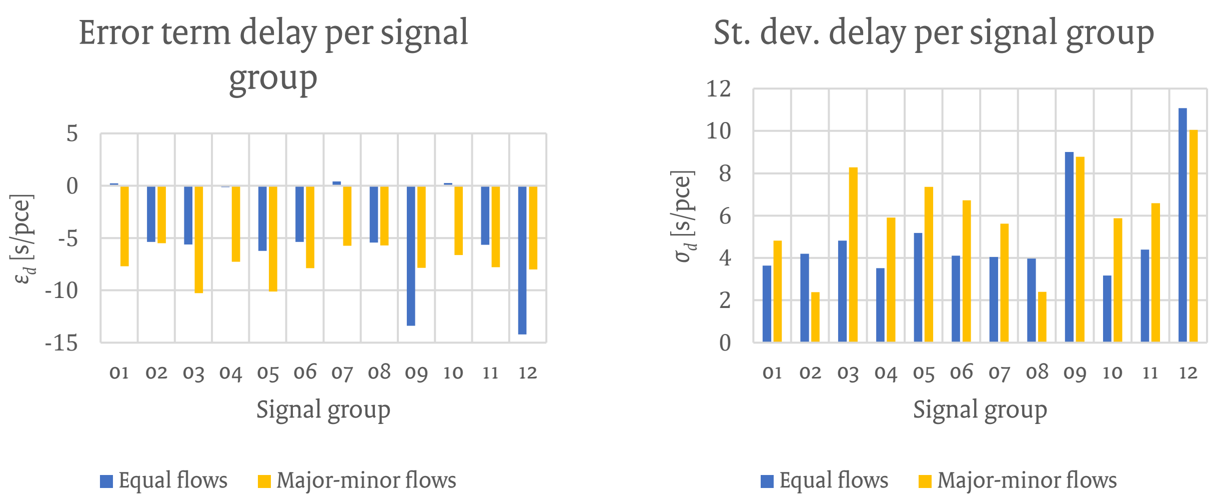

Figure 5-3 | Calibration results for delay, with error terms ε_d (top), and standard deviations σ_d (st. dev., bottom) as used to define the bandwidth.

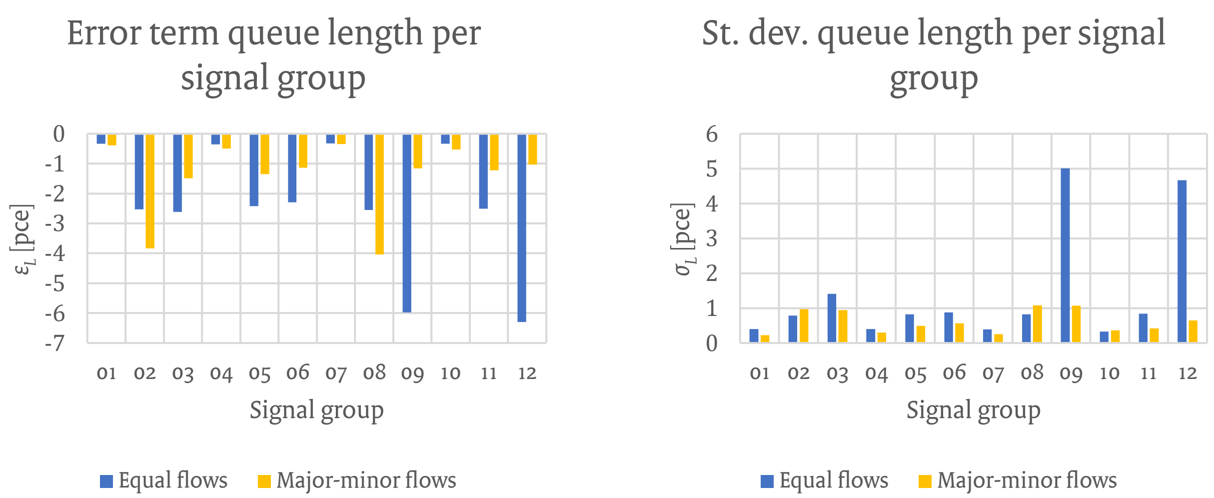

Figure 5-4 | Calibration results for queue length, with error terms ε_L (top), and standard deviations σ_L (st. dev., bottom) as used to define the bandwidth.

5.3.2. Inefficiency detector

The results of the inefficiency detector are given in terms of the overall inefficiency ratios, and average evaluation scores for the whole intersection, thus not per signal group, unless explicitly stated otherwise. That way, the amount of data presented in this section is limited. In appendix F.2, a detailed overview of the simulation study results are given, per alternative, per signal group. In the appendix, the performance statistics are given as well. Also, it is emphasised that the results presented here do not include inefficiency ratios, and evaluation scores for the degree of saturation, even though the degree of saturation is considered as an integral part of the evaluation and diagnosis method, because the degree of saturation is included as a multi-variable performance indicator as result of a preliminary testing of the evaluation and diagnosis method, as stated in section 4.1.6. Also, no evaluation scores are given for phase failures, since the reference performance model for this multi-variable performance indicator did not have to be calibrated, as concluded in section 4.1.6 as well.

Altogether, the inefficiency ratios per alternative, and the evaluation scores per alternative are given in respectively Figure 5-5, and Figure 5-6 on the next page. Given these results, the following findings stand out:

- The results for the base case alternative show that the (assumed) problem-free traffic signal controller does already suffer from a certain level of inefficient performance, e.g. the inefficiency ratio for the base case alternative is approximately 0.049 for both delays, and queue lengths, and 0.042 for phase failures in the equal flows scenario, see Figure 5-5. However, it is assumed that the observed inefficiency ratios in the base case alternative are due to an inadequate geometric intersection design, and thus part of the constraints of the traffic signal controller. The results per signal group, as will be discussed in section 5.3.3 for several examples, and as given in more detail in appendix F.2, show more significant differences. Also, it must be emphasised that the observed inefficiency ratios for the base case scenario could be the result of other traffic signal controller problems, e.g. a minimum green time that is too long for the through-going signal groups. However, those other problems are not selected in this thesis, instead it is assumed that the observed inefficiency ratios for the base case scenario are the result of an inadequate geometric intersection design. In future work, it is recommended to include more problems in the testing of the integral evaluation and diagnosis method to develop more insight in the causes of detected inefficiencies. This also relates to the creation of a more generic, and widely applicable database, as introduced in section 5.2.5.
- The results in terms of evaluation scores show a rather constant evaluation score on average, thereby varying between around an approximate evaluation score of 8.2. Only a few alternatives show evaluation scores lower than 8.0.
- In the base case alternative 0 with equal flows, the overall inefficiency ratio for phase failures is approximately 0.042, while other alternatives with equal flows, such as alternatives 3, and 5 (see Table 5-5) have an overall inefficiency ratio for phase failures of respectively approximately 0.018, and 0.038. This means that in these alternatives, with problems implemented on purpose, the performance improved in terms of less phase failures, for the whole signalised intersection. It is assumed that the relatively high inefficiency ratio for phase failures in alternative 0 are the result of the random fluctuations in the arrival pattern, resulting in phase failures due to an unfavourable termination of the green phase. This assumption is based on the findings that in alternatives 3, and 5, as well as other alternatives, the inefficiency ratios for delays, and queue lengths are higher than in alternative 0. Indeed, e.g. the inefficiency ratio for queue length in alternative 0 is approximately 0.049, as opposed to 0.065, and 0.107 in respectively alternatives 3, and 5, see Figure 5-5. Consequently, it means that the problems implemented on purpose do not affect phase failures as negatively as delay, and queue length, because the random fluctuations are present in all alternatives, and thus that the use of various multi-variable performance indicators provide more information on the inefficiencies the diagnosed problems did cause. Besides, it is recommended for future work to investigate this finding more in-depth, in particular with respect to the various traffic flow volume input sets. This future work might then also focus on the question whether a value judgement should be added to the phase failure performance, i.e. assessing one phase failure as more severe than another, e.g. in terms of moment it occurs with respect to guaranteed red time (see Figure 4-5).

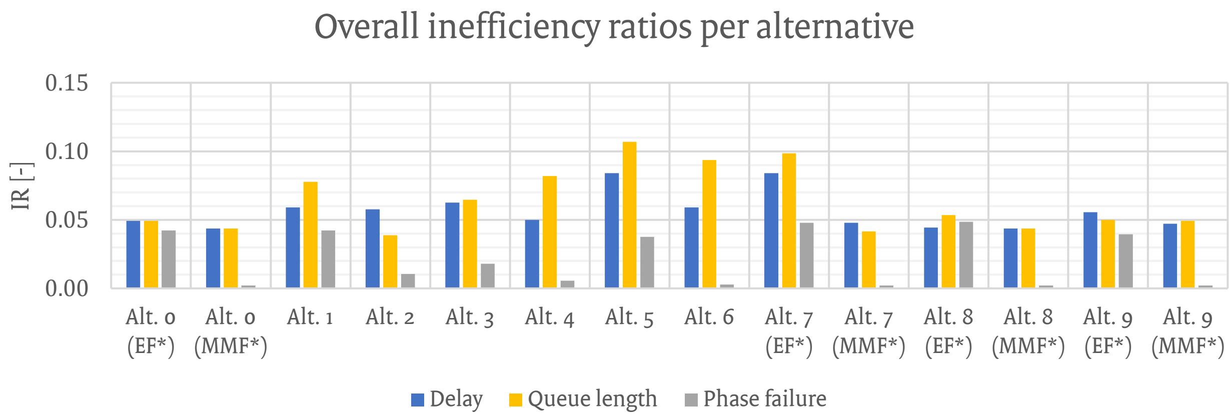

Figure 5-5 | Inefficiency ratios IR [-] for delay, queue length, and phase failure, for all signal groups combined, per alternative. *: EF = equal flows; MMF = major-minor flows.

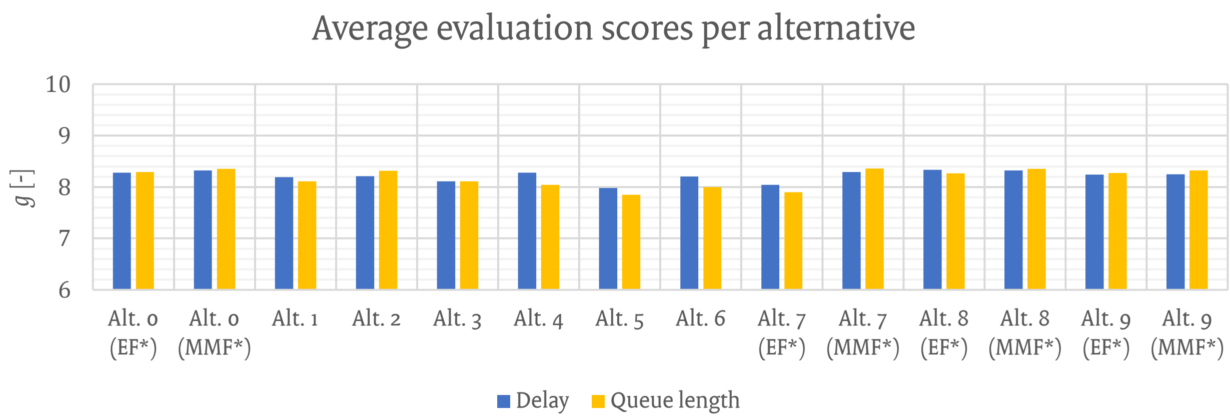

Figure 5-6 | Evaluation scores g [-] for delay, and queue length, averaged for all signal groups combined, per alternative. *: EF = equal flows; MMF = major-minor flows.

5.3.3. Diagnosis module

The testing of the diagnosis module focuses on testing whether the problems can be diagnosed using the inefficiency detector outputs. The test of the diagnosis module for the blind alternatives 7, 8, and 9 is considered as the "true test case", for it is not yet known which problems are implemented deliberately. Therefore, the testing is discussed in two parts: first for the known problems (alternatives 1 to 6), and then for the unknown problems (blind alternatives 7 to 9). For all alternatives, it holds that results that are not presented in this section, are given in appendix F.2.

5.3.3.1. Double problems: alternatives 1 to 6

The alternatives with two problems implemented, showed that the cyclic approach of the diagnosis module does indeed result in the diagnosis of both problems. It was found that, in general, one problem presents itself more clearly than the second problem: one problem is initially found, which is then mitigated. When this first problem is mitigated, the traffic signal controller is evaluated again, after which the inefficiency detector output results show that another problem is present. In this second iteration, the second problem is found, and successfully mitigated.

The testing of the diagnosis module emphasised that the inefficiency detector outputs for delays, queue lengths, and phase failures are not always sufficient to diagnose a problem: the use of the degree of saturation as multi-variable performance indicator as integral part of the traffic signal controller evaluation method has proven its added value in this case study. Indeed, it was found that in the case of incorrect gap times for the long detector, e.g. in alternative 5, the inefficiency detector outputs for delays, queue lengths, and phase failures are not sufficient to diagnose this problem. However, the degree of saturation showed that on the signal groups with the incorrect long detector gap time, the degree of saturation is lower, see Figure 5-7. This is due to the fact that the same traffic flow volumes are being served by those signal groups, while using more green time. That is, due to the increased gap time, the green phase is terminated later than should have been the case with a correct gap time, thus resulting in a longer green time. Given the definition of the degree of saturation ($x = qT_C/sT_{G,\text{eff}}$), as discussed in section 4.1.5.1, it can be seen that for constant traffic flow volume q, saturation flow s, and cycle time T_C on the one hand, and an increased green time T_G on the other hand, the degree of saturation decreases.

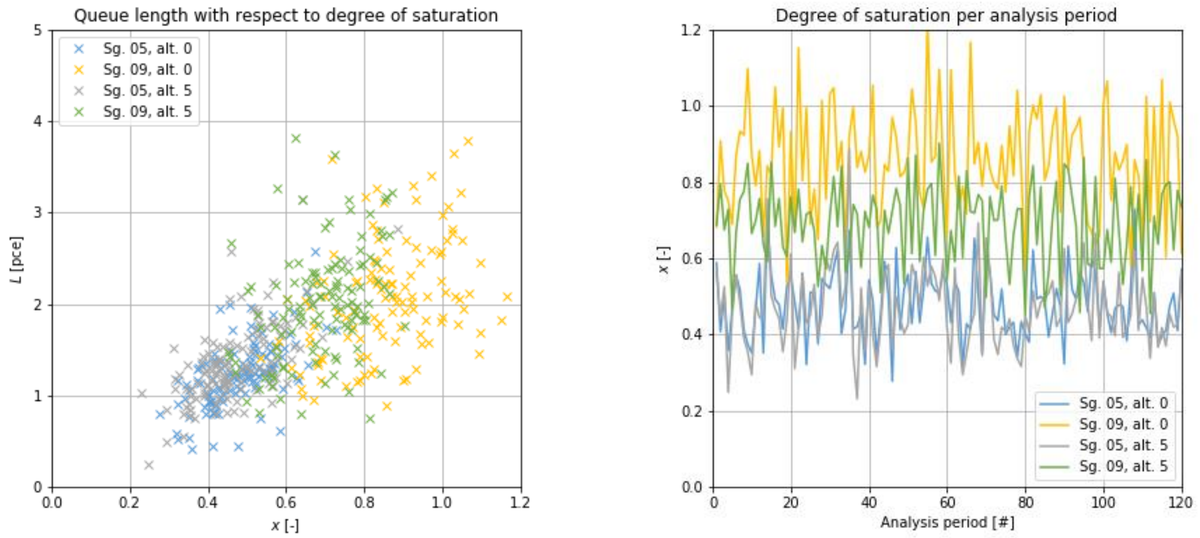

Figure 5-7 | Generated queue length *L* with respect to degree of saturation *x* (left), and degree of saturation per analysis period (right), for signal groups (sg.) 05 and 09, for alternative (alt.) 0, and 5.

Furthermore, it must be emphasised that the incorrect long detector gap time did not result in significantly different inefficiency detector outputs. This might raise the question whether an incorrect long detector gap time is a problem, or more specifically, whether the tested incorrect long detector gap time is a problem, because it did not result in clearly higher delays, or queue lengths, as can be seen in Figure 5-7 as well. Therefore, it is recommended for future work to investigate which problems are "notable" problems, thus problems with clearly higher delays, queue lengths, etc., also with respect to the inefficiency detector outputs. This implies that it is recommended to investigate which problems should be included in the database the integral method uses, as discussed in section 5.2.5 for this case study. In the same way it could be investigated how less "notable" problems (problems that result in less significant inefficiencies, but are still considered as problems, e.g. in practice) can be detected, and diagnosed using the integral evaluation and diagnosis method presented in this thesis, for instance by adding other multi-variable performance indicators, stricter bandwidths, etc.

The other double problem alternatives did not show other significant issues regarding the application of the diagnosis module, other than the issues mentioned above. The result was that the problems are diagnosed, and mitigated successfully for these alternatives.

5.3.3.2. Blind alternative 7

Because the overall inefficiency ratios, and average evaluation score of alternative 7, per traffic flow volume input set, as given in Figure 5-5, and Figure 5-6, showed that alternative 7 with equal flows resulted in a more significant deviation of the overall inefficiency ratio, and average evaluation score with respect to the base case alternative 0. Therefore, alternative 7 with equal flows is discussed further.

The inefficiency ratios, and evaluation scores per signal group, for alternative 7 with equal flows, are given in Figure 5-8. These inefficiency detector output results show a clear, and quite significant inefficient performance of signal group 06. Indeed, the inefficiency ratio for, e.g., queue length is over 0.50, implying that during more than half of the simulation period, the generated queue length was outside the reference bandwidth for queue length. A similar trend is seen for the other multi-variable performance indicators, and evaluation scores.

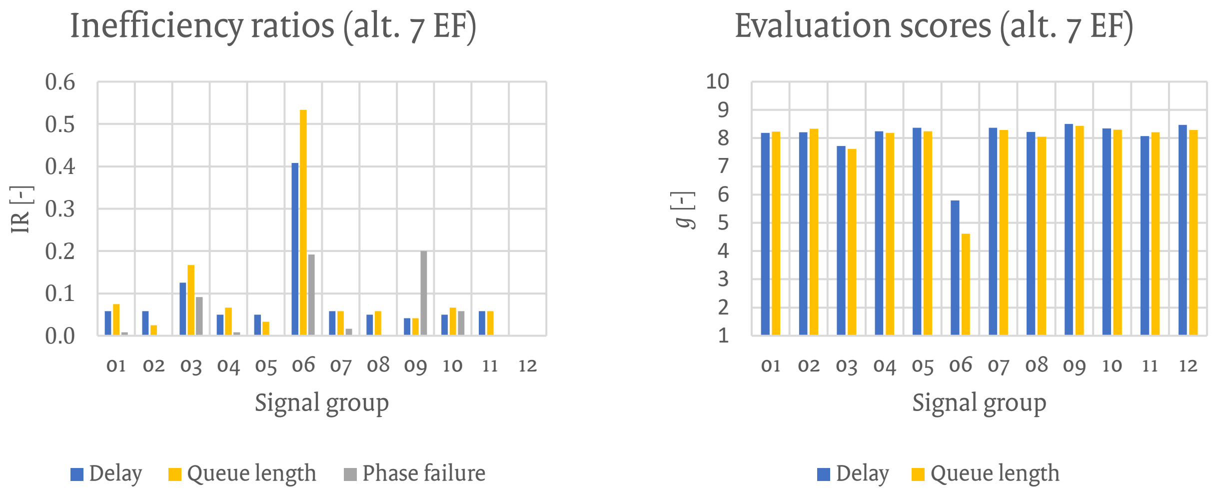

Figure 5-8 | Inefficiency detector output results per signal group, for alternative (alt.) 7 with equal flows (EF).

The found inefficiency detector output results are known to be assignable to a single problem, as introduced in section 5.2.4. Given the level of inefficient performance for signal group 06 in particular, it is concluded that the implemented problem is that alternative realisations are deactivated. Thus, given the generated inefficiency detector output in Figure 5-8, with respect to examples of trends of inefficient performance in the database (as discussed in section 5.2.5), in particular in terms of a signal group with deactivated alternative realisations, the conclusion is that the generated trend is similar to those in the database. This implies that the diagnosed problem is "deactivated alternative realisations of signal group 06". Therefore, the proposed countermeasure is to activate the alternative realisations of signal group 06. This is checked by implementing this countermeasure, and rerunning the diagnosis module, where it was found that the proposed countermeasure did indeed mitigate the diagnosed problem.

5.3.3.3. Blind alternative 8

The results show that the problem implemented in alternative 8 only presents itself in the equal flows traffic flow volume input set, because the overall inefficiency ratios, and average evaluation scores of alternative 8 with major-minor flows are identical to those of the base case alternative 0 with major-minor flows, see Figure 5-5, and Figure 5-6. Therefore, only alternative 8 with equal flows is discussed. The inefficiency detector output results of this alternative, per signal group, are given in Figure 5-9.

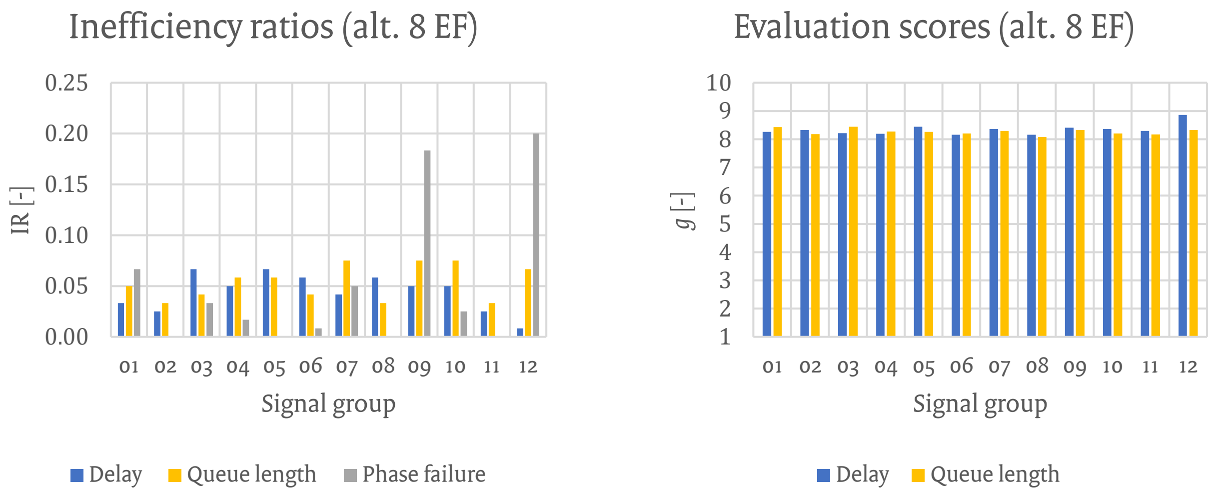

Figure 5-9 | Inefficiency detector output results per signal group, for alternative (alt.) 8 with equal flows (EF).

Again, it is known that the inefficiency detector output results are assignable to a single problem. The results show relatively high inefficiency ratios for phase failures at signal groups 09, and 12, though with rather good results for delays, and queue lengths on this signal group. However, the performance statistics, in particular with respect to the database, did show the observed trend in alternative 8 is not related to the inefficiency ratios for phase failures at signal groups 09, and 12. Instead, in the database, an example of inefficient performance at this particular intersection, with identical traffic flow volumes, showed almost identical inefficiency detector outputs, see Figure 5-10, and Figure 5-11 on the next page. The example of inefficient performance in the database is related to a traffic signal controller with incorrect maximum green time on signal group 01. Therefore, the countermeasure is proposed to adjust the maximum green time of signal group 01. It is checked whether the proposed countermeasure did successfully mitigate the diagnosed problem. The results of the traffic signal controller evaluation and diagnosis method for the traffic signal controller with the implemented countermeasure, showed that this was the case. Therefore, the conclusion is that in alternative 8 with equal flows, the maximum green time of signal 01 was incorrect.

It must be noted that this problem was quite difficult to diagnose, because the problem of a reduced maximum green time on a right-turning signal group does not present itself very often. In other words, a right-turning signal group does not usually reach its maximum green time, at least not given the tested traffic flow volumes in this tested (see section 5.2.2.1), mainly due to the alternative realisations it receives. This results in the fact that the periods of inefficient performance are relatively scarce, which impedes the diagnosis procedure. However, this may be also due to the limited database that is used. This means that if a more elaborate, and generic database would be used, the diagnosis procedure would be easier, for more examples of inefficient performance as a result of an incorrect maximum green time on a right-turning signal group would be included. This emphasises the recommendation for future work given in section 5.2.5, namely to create a more generic, and widely applicable database filled with examples of inefficient performance, thereby including examples of other problems than those selected in this thesis as well.

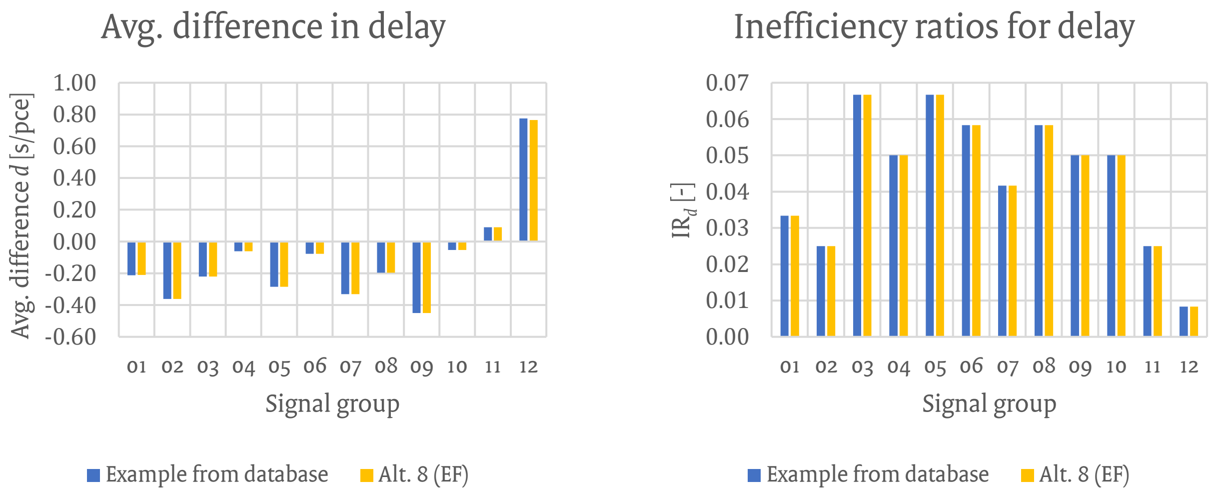

Figure 5-10 | Comparison of inefficiency detector outputs for delay of an example from the database with respect to alternative (alt.) 8 with equal flows (EF), with average (avg.) difference in delay d ($(\sum_\tau r_d^\tau - p_d^\tau)/n_\tau$ [s/pce]), and inefficiency ratios for delay (IR_d [-]).

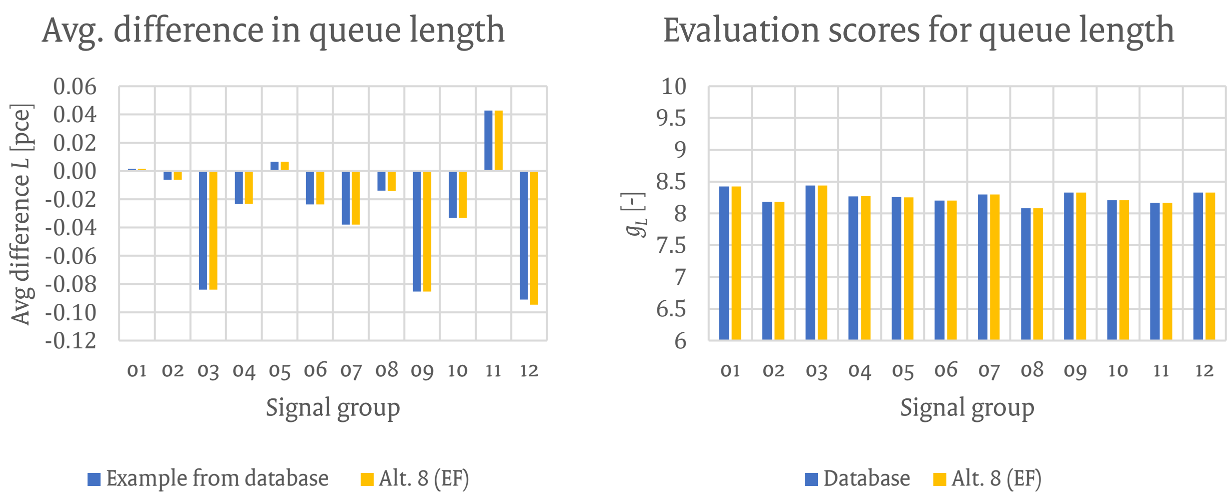

Figure 5-11 | Comparison of inefficiency detector outputs for queue length of an example from the database with respect to alternative (alt.) 8 with equal flows (EF), with average (avg.) difference in queue length L ($(\sum_\tau r_L^\tau - p_L^\tau)/n_\tau$ [pce]), and evaluation scores for queue length (g_L [-]).

5.3.3.4. Blind alternative 9

For the last alternative, no significant differences were found in the overall inefficiency ratios, and average evaluation scores, to determine which traffic flow volume input set (equal flows, or major-minor flows) yields the clearest results. Therefore, the inefficiency detector outputs per signal group for both traffic flow volume input are considered. Again, it is known that the inefficiency detector output results are caused by a single problem.

The inefficiency detector outputs did not show results similar to results of examples in the database. Therefore, the degree of saturation is assessed, see Figure 5-12. It can be seen that on various signal groups, the degree of saturation has increased, or decreased in alternative 9 with respect to the base case alternative 0. The differences are in general rather small, e.g. for signal group 03 the degree of saturation in alternative 0 with equal flows was 0.70, while in alternative 9 with equal flows, the degree of saturation was 0.71. However, for only one signal group, the degree of saturation decreased for both traffic flow volume sets, namely signal group 08. Although the changes are again rather small (0.01 lower for equal flows, and 0.08 lower for major-minor flows), the conclusion was that the problem in alternative 9 is an incorrect gap time on the long detector of signal group 08. Therefore, the proposed countermeasure is to adjust this long detector gap time. Next, it is checked whether it did mitigate the diagnosed problem. This was the case.

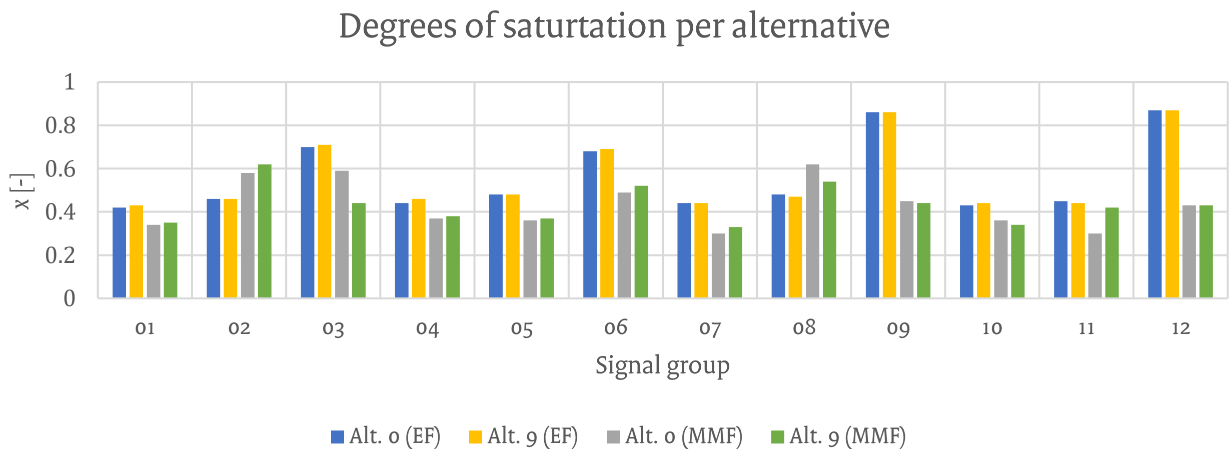

Figure 5-12 | Degrees of saturation x, per signal group, for alternatives (alt.) 0, and 9, for both equal flows (EF), and major-minor flows (MMF).

Several notations must be made regarding this alternative. First, the conclusion that the gap time of the long detector on signal group 08 is based solely on the finding that for both traffic flow volume sets, the degree of saturation has decreased. This implies that the size of the difference in degree of saturation did not play a role. Indeed, even though the differences were rather small, the degree of saturation only decreased for both traffic flow volume sets on signal group 08. Therefore, it is recommended for future work to investigate whether the size of the difference should play a role in future evaluations. This research should focus on which other inefficiencies, problems, or other factors might cause the degree of saturation to decrease. That way, an additional condition could be added to assign the found decreased degree of saturation to the problem of an incorrect long detector gap time.

Secondly, it was found that the incorrect gap time on the long detector on signal group 08, was only implemented on one of the two long detectors on this signal group – signal group 08 has two lanes, with on each lane a long detector. Because in the examples in the database the problem of the incorrect gap time is implemented on both long detectors, if the considered signal group had two lanes, which impeded the diagnosis procedure. Moreover, the problem was implemented on the long detector on the left lane, for which it is assumed that less traffic uses this lane, thus that on the right lane, the traffic flow volume is higher. As a consequence, the importance of the left lane long detector is lower, i.e. it has less influence on the termination of the green phase. This results in a more complicated diagnosis procedure as well, because the effect of the problem in terms of inefficient performance, becomes less clear.

5.3.4. Conclusion

Altogether, all problems were diagnosed, and mitigated successfully, using the integral traffic signal controller evaluation and diagnosis method. Even for the blind alternatives, the problems that were implemented on purpose, were diagnosed, and mitigated. However, during this testing, several remarks were made on the evaluation and diagnosis process. These remarks, as well as the diagnosed problem(s) per alternative, are listed in Table 5-6.

Table 5-6 | Overview of diagnosed problem(s) per alternative, per signal group (sg.), including remarks, as conclusion of the case study.

Alternative	Diagnosed problem(s)	Signal group(s) with diagnosed problem(s)	Remarks
1	Alternative realisations	02 and 03	N/A
2	Maximum green time	04 and 06	N/A
3	Gap time	02 and 07	No second iteration needed
4	Alternative realisations Maximum green time	04 10	N/A
5	Maximum green time Gap time	05 09	N/A
6	Alternative realisations Gap time	07 11	N/A
7 (blind alternative)	Alternative realisations	06	N/A
8 (blind alternative)	Maximum green time	01	Impeded diagnosis process due to limited number of inefficient performance periods
9 (blind alternative)	Gap time	08	N/A

5.4. Main findings

The testing of the integral traffic signal controller evaluation and diagnosis method showed that, for the considered alternatives, all of the problems that were implemented deliberately, were identified, diagnosed, and mitigated with success. Furthermore, some remarks were made. In short, the main findings of the testing of the integral evaluation and diagnosis method are as enumerated below:

- The inefficiency detector output results of the base case alternative (0) already show some level of inefficient performance. However, this is assumed to be due to an inadequate intersection design. This principle is also applied when in other alternatives, no cause is found for the inefficient performance.
- Besides the inefficient performance of the base case alternative, it is found that for the alternatives with problems implemented on purpose, the average evaluation scores are relatively high: even for faulty traffic signal controllers with one, or more problems implemented, the average evaluation scores lay above 7.5, which implies a fairly good performance. Only when the evaluation scores per signal group are considered, lower scores are found. This is caused by the implementation of a problem on, in this thesis, one, or two signal groups. For these signal groups, the evaluation scores might be lower, but since the other signal groups still score relatively high, the lower evaluation scores are "lost" in the intersection-wide average evaluation score. This emphasises that a signal group-based evaluation score provides more information on the performance of the traffic signal controller. Nonetheless, it is recommended for future work to investigate whether intersection-wide evaluation scores, or a conflict group-based approach can provide valuable insights as opposed to the signal group-based approach tested in this thesis. This future work might also include the use of data of signalised intersections with more than two problems implemented.
- Given the set-up of the case study, it is found that the inefficiency detector output in terms of a list of those periods during which the performance was found to be inefficient, is not crucial for a correct functioning of the integral traffic signal controller evaluation and diagnosis method. However, in practice, it is expected that this output is crucial, because a problem might only be implemented in a specific time-of-day program of the traffic signal controller, for instance, implying that this list of periods of inefficient performance can be used to find that specific time-of-day program where the problem is relevant. It is recommended for

future work to examine the exact value of the inefficiency detector output result in terms of periods of inefficient performance, in relation to how it is used by traffic engineers in practice.

- The application of the integral traffic signal controller evaluation and diagnosis method in the case study discussed in this chapter has proven to be able to identify, diagnose, and mitigate several problems, based on the inefficiency detector output results in terms of inefficiency ratios, evaluation scores, and performance statistics. The resulting proposed countermeasures were found to mitigate these problems effectively. However, the diagnosis module makes use of this pattern-recognising approach, implying that a database of examples of inefficient performance is needed, to be able to perform this type of evaluation. The database itself must be filled with data as well. In this case study, data is used for the same signalised intersection, though for only a limited number of signal groups, and a limited number of problems. Although this database sufficed to test the integral method in this case study, the use of a larger database, including more complex problems, and combinations of problems, is recommended in future work.
- As discussed in section 3.1.2, the integral evaluation and diagnosis method is considered as an informal decision support system, because it includes components of such a system, which are not formalised, or automated. This is in particular true for the pattern-recognising step of the diagnosis module, because this pattern-recognition is done manually in this case study. Therefore, the recommendation for future work on the formalisation of the integral method, as given in section 3.3, is emphasised here as result of the case study.
- The test of the traffic signal controller evaluation and diagnosis method showed that the step in which it is checked whether the proposed countermeasure did mitigate the diagnosed problem successfully, is a very valuable step, for it enables the diagnosis of multiple problems at one traffic signal controller. This also emphasises the added value of a cyclic, iterative approach. Indeed, via iterations, it was found that more than one problem can be diagnosed. Although this is done for two problems in this thesis, it is hypothesised that the same principle holds if more problems are considered. This is recommended to investigate for future work.
- Because this thesis used a controlled test environment (it was known how many problems were implemented, and the base case alternative was known to be problem-free in terms of the selected problems), there is no exit condition defined for this iterative process. In practice, such an exit condition might be needed, e.g. for it is not known whether the traffic signal controller used in the pilot study was indeed a problem-free controller. In that case, an exit condition might be used as a measure for when one is satisfied with the evaluation and diagnosis method results, e.g. a maximum number of iterations. Therefore, it is recommended for future work to (i) use data from an uncontrolled test environment (e.g. from practice), and (ii) investigate whether a, and which exit condition is preferred.
- Lastly, it must be noted that the case study emphasised the added value of using the degree of saturation as multi-variable performance indicator, as integral part of the traffic signal controller evaluation and diagnosis method. Using this multi-variable performance indicator, the problem of an incorrect gap time can be diagnosed. The conclusion is therefore that including the degree of saturation as multi-variable performance indicator is an important integral aspect of the evaluation and diagnosis method. This introduces the recommendation for future work to investigate more in-depth how certain problems can be found in terms of multi-variable performance indicators, and the resulting deviations of the generated performance with respect to the reference performance, as part of a further development of the traffic signal controller evaluation and diagnosis method.

6. Conclusions and recommendations

In the first chapter of this thesis, the research objective with corresponding research questions were formulated. These research questions are answered in this chapter. Also, recommendations are given on both further implementation of the traffic signal controller evaluation and diagnosis method, and future work on this method, and related topics.

6.1. Conclusions

The objective of this study was to develop, and present an integral evaluation and diagnosis method for traffic signal controllers, including a simultaneous assessment of multiple performance indicators, which detects inefficiencies in terms of traffic performance functioning, scores the traffic signal controller, diagnoses the cause of the detected inefficiency, and propose countermeasures to improve the traffic performance functioning of the traffic signal controller, by gaining understanding in the various performance indicators and their interaction, how these performance indicators can be used to detect inefficiencies, and how the detected inefficiencies relate to certain problems, in order to diagnose the problems and propose countermeasures. This resulted in the following research question: *How can an integral evaluation and diagnosis method, based on an assessment of multiple performance indicators, indicate inefficiencies on the traffic performance functioning of a traffic signal controller, and propose countermeasures to mitigate those inefficiencies?*

To be able to answer this research question, the sub-questions are answered first:

What are the motives and methods used for the evaluation of traffic signal controllers?

Traffic signal controllers are evaluated for various reasons, and motives: (i) assessment of new traffic signal control aspects (e.g. algorithms, systems, models, etc.), (ii) traffic performance evaluation, and (iii) evaluation as part of periodic maintenance. However, it is found in section 2.1 that little to no attention is given to the integral evaluation of traffic signal controllers. Indeed, in the literature discussed in section 2.1, it is found that traffic signal controllers are evaluated seldomly to assess traffic performance evaluation. Some examples that relate to this are found, although they mainly discuss the methods, and systems for an integral evaluation of traffic signal controllers. In these studies, methods are proposed to collect data to evaluate traffic signal controllers (Bullock & Day, 2009), or a framework for a periodic traffic performance evaluation of traffic signal controllers as part of periodic maintenance (Radivojevic & Stevanovic, 2017b), or it is noted that such periodic maintenance is needed to adjust the traffic signal timing settings, and thus to improve the traffic performance of traffic signal controllers (Lavrenz, et al., 2016; Sunkari, 2004). Moreover, these studies focused mainly on the system or method they proposed, and did therefore not consider how the traffic performance of traffic signal controllers is affected by various problems, and how this can be mitigated.

In practice, a method does exist to evaluate the traffic performance of traffic signal controllers, namely the Quick Quality Scan (QQS), or Instant Quality Scan (IQS), in combination with the BI-tool at VIALIS. In these QQS, and IQS evaluations, the objective is to evaluate various performance indicators using the BI-tool to find potential problems, and propose countermeasures to mitigate the problems. That way, the BI-tool is a decision support system. Although such a decision support system is used in practice, the use of decision support systems as a way to evaluate and diagnose traffic signal controllers is not discussed widely in literature, at least not as an integral evaluation method: e.g. Wen (2008), Elkosantini & Ahmed (2014), and Moalla, et al. (2013) focused on the development of traffic signal control algorithm based on a decision support system, rather than the evaluation of a traffic signal controller. Nonetheless, the use of decision support systems in other fields showed promising results regarding the detection, and diagnosis of problems.

Which performance indicators must be selected?

The selection of performance indicators is based on the scope of the integral evaluation and diagnosis method presented in this thesis. This implies that the presented method may include various performance indicators. The selected performance indicators must be related the policy constraints of the road authority, and the problems that one wants to diagnose.

In this thesis, the selection is meant to outline the scope of the development, and testing of the method. This selection is based on findings in literature (section 2.1), and evaluation reports from practice (Vialis, n.d. [a]). Given these sources, a total of four problems is selected: deactivated alternative realisations, incorrect maximum green times, incorrect long detector gap times, and inadequate geometric intersection design. These problems relate to various symptoms (inefficient performance of various performance indicators), with as overarching symptom an “inefficient use of green.” This overarching symptom presents itself mainly in terms of delay, and unnecessary waiting. Because unnecessary waiting is part of the delay at an intersection, only delay is considered. In addition, three other performance indicators are selected, to give a more complete overview of the performance of the traffic signal controller, yielding the following performance indicators:

- Degree of saturation;
- Delay;
- Phase failure;
- Queue length.

These four performance indicators are, in general, rather commonly used in the evaluations in practice. Furthermore, the formal definitions, as given in section 4.1.5, showed that these four performance indicators are in fact multi-variable performance indicators, with traffic flow volume, cycle time, and green time as describing variables.

How can the multi-variable performance indicators be used to detect inefficiencies?

The method presented in this thesis, and as explained in section 3.2, consists of two major modules: the inefficiency detector, and the diagnosis module. The inefficiency detector focuses on identifying periods with inefficient performance, whereas an inefficiency is defined as “*a difference between the generated, measured performance, and the computed reference performance*.” This definition is based on the current practice in the BI-tool, namely comparing the measured traffic performance of a traffic signal controller to predefined static threshold or reference values, although these threshold or reference values are made dynamically (the reference performance of a multi-variable performance indicator) in the method presented in this thesis. Both the generated, and reference performance represent are expressed in terms of the aforementioned multi-variable performance indicators, and their describing variables. The reference performance denotes the computed performance of the traffic signal controller, for the multi-variable performance indicators, using a reference performance model.

Such a reference performance model is a computational model to determine the reference performance. However, the reference performance models use various assumptions, which might affect the way inefficiencies are detected in the inefficiency detector module, which means that the assumptions might cause inaccuracies in the computation of the reference performance. Because the reference performance is used to the detect inefficiencies, it is important to account for these inaccuracies. Several assumptions play a role: (i) arrival pattern, (ii) green time lags, (iii) block sequence dependent cycle time, and (iv) acceleration-deceleration delay. The first assumption is tested in particular in section 4.4. The calibration step, as performed in section 4.4 for the inaccuracies of the reference performance models due to the assumptions on the arrival pattern, showed that the inaccuracies of the reference performance models can be accounted for, using the calibration. In this calibration, an error term is used to adjust the reference performance model. In addition, a bandwidth is defined, which represents the range of the reference performance in which deviations of the exact reference performance are the result of the inaccuracies in the reference performance models. This implies that the inaccuracies of the reference performance models are distinguished from traffic signal controller inefficiencies by checking whether the detected inefficiency (given the aforementioned definition of an inefficiency) lays within the bandwidth: if that is the case, it is an inaccuracy, otherwise it is a traffic signal controller inefficiency.

The testing of the integral method, as discussed in chapter 5, showed that the selected problems present themselves via different inefficiencies: the inefficiency detector output results, given as performance statistics, inefficiency ratios, and evaluation scores per multi-variable performance indicator, as explained in section 3.2.3.1, differ per simulated problem (deactivated alternative realisations, incorrect maximum green times, incorrect long detector gap times, and inadequate geometric intersection design). This implies that each problem, as selected in this thesis, has its own pattern in terms inefficiency detector output results, per multi-variable performance indicator.

How can detected inefficiencies be assigned to traffic signal controller problems?

As shortly introduced above, each problem has its own pattern in terms inefficiency detector output results. That means that the pattern of one problem differs from another problem. That way, a pattern recognition approach can be used to assign the found the inefficiency detector output pattern to a traffic signal controller problem. The method presented in this thesis uses a decision tree, based on an informal naïve Bayesian network, to perform this pattern recognition. This means that a database with examples of inefficient performances is used to compare the measured inefficiency detector output results, per multi-variable performance indicator with. The database includes the informal probabilities (hence “informal” naïve Bayes) that a measured symptom (inefficiency detector output) is related to a given problem. Per given problem in the database, it is checked whether the measured pattern is similar to the pattern(s) in the database, using “if-then” rules (hence decision tree). Thereto it is tested that a generated pattern is similar to a pattern in the database if the generated pattern shows the same trend as the pattern in the database – i.e. a high inefficiency ratio for a specific (set of) signal group(s), approximately equal performance statistics, etc. If that is the case, the inefficiency can be assigned to the corresponding problem. The number of “if-then” rules equals the number of possible problems of the traffic signal controller.

What are the steps of an integral evaluation and diagnosis method for traffic signal controllers?

As introduced above, the traffic signal controller evaluation and diagnosis method consists of two major modules: the inefficiency detector, and the diagnosis module. This resulted in a five-step process, as depicted in Figure 6-1. In short, the steps are summarised as follows (the elaborate discussion of the steps is given in section 3.2):

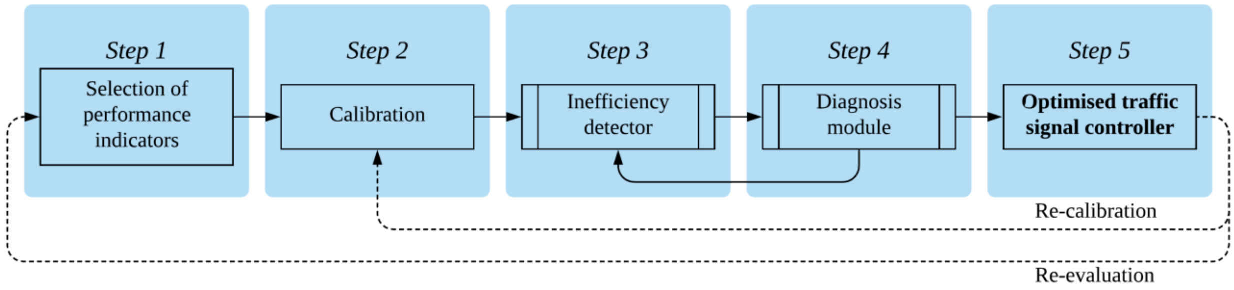

Figure 6-1 | Simplified, schematic overview of the integral traffic signal controller evaluation and diagnosis method.

1. *Selection of performance indicators*: first, the relevant performance indicators are selected, for which data is collected. This relates to the multi-variable performance indicators, as well as their describing variables. Some of the key aspects here are that (multi-variable) performance indicators should be selected that (i) offer a complete overview of the traffic performance, (ii) can be used to diagnose problems, and (iii) can be defined in terms of a reference performance. The output is a list of relevant (multi-variable) performance indicators.
2. *Calibration*: the adjustment of the reference performance models is used to cancel out inaccuracies of the reference performance models, by defining error terms, and bandwidths: the outputs of the calibration. The calibration is performed exclusively for the reference performance models of multi-variable performance indicators that include such inaccuracies. It must be noted that the calibration uses data of an (assumed) problem-free traffic signal controller.
3. *Inefficiency detector*: the basic principle of the inefficiency detector is to identify those periods during which the generated performance deviated from the reference performance, which is defined as a generated performance that lays outside the bandwidth, or that is not equal to the reference performance when no bandwidth is computed. The module consists of a five-step procedure in which step by step, the inefficiency detector outputs are generated: (i) a list of analysis periods during which the performance is considered as inefficient, (ii) the performance statistics, (iii) the inefficiency ratio, and (iv) the evaluation score.
4. *Diagnosis module*: the diagnosis of problems focuses on assigning the correct problem to the detected inefficiencies, and propose countermeasures to mitigate the diagnosed problems. The diagnosis is based on an informal decision support system, using an informal naïve Bayesian network in combination with a decision tree. Also, in this step, it is checked whether the proposed countermeasures were effective, by implementing the proposed countermeasures, thus optimising the traffic signal controller. That way, the diagnosis module, and the integral method is tested as a cyclic process, which enables the diagnosis of multiple problems for the same traffic signal controller, hence the feedback from step (4) to step (3) in Figure 6-1, since this check is done by re-running the inefficiency detector, and diagnosis module. However, the diagnosis module proposes one countermeasure per iteration, at least as tested in this thesis.
5. *Optimisation*: the results of the diagnosis module is the final optimisation of the traffic signal controller, yielding an optimised traffic signal controller. Therefore, it is assumed that the diagnosis module diagnosed all problems, and mitigated them by proposing, and implementing countermeasures. The results can be used as input for a re-calibration, as the (assumed) problem-free traffic signal controller, or as a starting point to re-do the whole evaluation and diagnosis process (re-evaluation). The latter is especially relevant given the gradual decrease of performance of the traffic signal controller due to the changes in the traffic flow volumes, etc. (Lavrenz, et al., 2016). This emphasises the cyclic process that is included in the traffic signal controller diagnosis and evaluation method as well.

How can an integral evaluation and diagnosis method, based on an assessment of multiple performance indicators, indicate inefficiencies on the traffic performance functioning of a traffic signal controller, and propose countermeasures to mitigate those inefficiencies?

The integral traffic signal controller evaluation and diagnosis method presented in this thesis, includes an assessment of multi-variable performance indicators, as part of a procedure to detect inefficiencies caused by various traffic signal controller problems. Based on the detected inefficiencies, the problems that caused the inefficiencies are diagnosed, and countermeasures are proposed to mitigate the diagnosed problems. This process of evaluating and diagnosing the traffic signal controller is summarised in five steps. Altogether, the presented method is classified as an informal decision support system, based on the current practice at VIALIS, with the QQS, and IQS evaluations using the BI-tool.

The testing of the integral traffic signal controller evaluation and diagnosis method showed that the method is able to detect inefficiencies, and assign them to the problem(s) that caused the detected inefficiencies. Although the testing in a half-blind case study included alternatives for which it was known beforehand what the problems were, the blind alternatives indicated that if the problem is not known beforehand – just as an evaluation in practice – the problem can successfully be diagnosed, and mitigated with the method. Indeed, using the pattern-recognition approach, the selected problems were successfully identified, and diagnosed. The resulting proposed countermeasures were found to mitigate these problems effectively. However, it must be noted that this includes the diagnosis of the problem of inadequate geometric intersection design if the detected inefficiencies could not be assigned to any of the other three selected problems (deactivated alternative realisations, incorrect maximum green times, and incorrect long detector gap times). This was found to be especially relevant when assessing the problem-free traffic signal controller. Nonetheless, the test of the traffic signal controller evaluation and diagnosis method showed that the step in which it is checked whether the proposed countermeasure did mitigate the diagnosed problem successfully, is a very valuable step, for it enables the diagnosis of multiple problems at one traffic signal controller.

Therefore, the conclusion is as follows:

The inefficiencies on the traffic performance functioning of a vehicle-actuated traffic signal controller, at least those selected in this thesis, can be indicated, and mitigated by applying the integral traffic signal controller evaluation and diagnosis method presented in this thesis. The integral evaluation and diagnosis method consecutively detects inefficiencies, diagnoses the problems that caused these inefficiencies, and proposes countermeasures to mitigate the diagnosed problems, using an assessment of various multi-variable performance indicators. That way, the presented method is decision support system that helps the traffic engineer to identify the moments of inefficient performance, diagnose the problems that caused them, and propose countermeasures to mitigate the diagnosed problems.

6.2. Recommendations

The conclusion that the integral traffic signal controller evaluation and diagnosis method (the "Quicker Quality Scan") is able to detect inefficiencies, diagnose the corresponding problems, and mitigate the diagnosed problems, is paired with several recommendations. As discussed throughout the thesis, various recommendations are given for future work. This future work also relates to improvements on the integral evaluation and diagnosis method, and the implementation in practice. These recommendations are formulated in this section.

6.2.1. Improvements on the integral evaluation and diagnosis method

As stated, various recommendations for future work are given throughout this thesis, whereas several recommendations are mainly related to potential improvements on behalf of the integral traffic signal controller evaluation and diagnosis method. These recommendations are listed below, in chronological order based on the sections where the recommendation is given. For a more elaborate explanation about the listed recommendations, please refer to the respective sections, as listed below as well.

- *Formalisation of decision support system* (sections 3.1.2, and 5.4): the integral evaluation and diagnosis method is developed as a decision support system. However, not all formal components of a decision support system (see section 2.2.1) are developed. For instance, no user interface was developed. Therefore, it is recommended to formalise the integral evaluation and diagnosis method as a decision support system.
- *Bandwidth definition* (section 3.2.2.1): the definition of the bandwidth is rather arbitrary, implying that in future work alternative definitions might be investigated.
- *Formalisation of diagnostic principles* (sections 3.2.4.1, 4.2, and 5.4): the diagnosis module procedure uses an informal naïve Bayesian network, in combination with a decision tree. The fact that the probabilities in the naïve Bayesian network are not formalised, and that the pattern-recognition process in the decision tree is based on a somewhat arbitrary approach, it is recommended to formalise both diagnostic principles. This could be done by collecting more data of examples of inefficient performance, to fill the database, and to determine the formal probabilities for the naïve Bayesian network, and by stating explicit conditions for the pattern-recognition process as used in the decision tree. Furthermore, it is recommended to investigate other definitions of good performance as input for the diagnosis of problems.
- *Automation of diagnostic principles* (section 3.2.4.1): the formalisation of the diagnostic principles also enables the automation of the diagnosis module. This relates to the formalisation of the decision support system as well.
- *Multi-problem diagnosis* (sections 3.2.4.2, and 5.4): the presented diagnosis module diagnoses one problem per iteration, because the used database only included examples of inefficient performance caused by a single problem. By formalising the diagnostic principles (as listed above), the database could also be filled with multi-problem examples of inefficient performance, as a way to diagnose multiple problems per iteration.
- *Countermeasure implementation order* (section 3.2.4.2): because presented diagnosis module diagnoses one problem per iteration, the order of countermeasures is not explicitly investigated. It is recommended to investigate in which order the proposed countermeasures can be implemented, for the diagnosed problems, and a specific order of implementation is needed.
- *Complex intersections* (section 4.1.6): the integral evaluation and diagnosis method is tested for a single, isolated signalised intersection, with signal groups for motorised traffic only. In future work, the method is recommended to be elaborated to include more complex intersections (i.e. active modes, coordinated intersections, urban networks, nearby intersections, etc.).
- *Degree of saturation reference performance* (section 4.1.6): although the degree of saturation is selected as an integral aspect of the traffic signal controller evaluation and diagnosis method, no reference performance model was developed. It is recommended to investigate whether this is needed, and if that is the case, to implement it in the method.
- *Re-calibration* (section 4.2): the calibration step is based on the inaccuracies of the reference performance model, and is assumed to be performed once. However, it is possible that re-calibration of the reference performance model is needed. It is recommended to investigate whether this is the case, and if so, what the desired frequency of re-calibration should be.
- *Database* (sections 5.2.5, and 5.3.3.3): the database used in the case study is rather limited database, since it is filled with a limited number of examples of inefficient performance. Also, the examples originate from one

signalised intersection as well. In future work, this database could be expanded, for instance using data for multiple problems, and various geometric intersection designs, thereby including data from practice (e.g. data from QQS, and IQS evaluations) as well.

- *Other problems* (section 5.3.2): the case study considered only four problems, whereas the problem of inadequate geometric intersection design is diagnosed as the cause of the inefficiencies if the cause was not found to be any of the three problems. However, the detected inefficiencies might be the result of other problems that were not selected, and tested in this thesis, e.g. a minimum green time that is too long for the given traffic flow volume. Therefore, it is a recommendation to include more problems in future development of the method.
- *Notable problems* (sections 5.3.3.1, and 5.4): not all selected problems resulted in notable inefficiency detector outputs (i.e. distinct inefficiency ratios, evaluation scores, and performance statistics for one problem with respect to another problem), raising the question how serious a traffic signal controller problem must be until it is considered as a problem. For future work, it is recommended to answer this question, in order to fill the database with useful examples of inefficient performance.
- *Size of change of degree of saturation* (section 5.3.3.4): no explicit decision rule was used regarding the size of the change of the degree of saturation to diagnose certain problems. Therefore, it is recommended for future work to investigate whether the size of the difference should play a role in future evaluations. This research should focus on which other inefficiencies, problems, or other factors might cause the degree of saturation to decrease.
- *Intersection-wide, or conflict group-based evaluation approach* (section 5.4): the method uses a signal group-based approach, while another approach might give insightful information as well.
- *Exit condition* (section 5.4): in the case study, alternatives with a maximum two problems implemented deliberately are considered, implying that more complex alternatives are not investigated. When this would be done, it is recommended to investigate whether an exit condition in the diagnosis module is desired, to prevent an infinite (feedback) loop. This is also related with the formalisation of the method, and the multi-problem recommendations as discussed above.

6.2.2. Implementation in practice

Besides the recommendations of the further development of the integral traffic signal controller evaluation and diagnosis method, several recommendations are given with respect to the implementation in practice. Although these recommendations are also part of a further development, it is in particular relevant to how the method might be implemented in practice, especially with respect to the current practice of QQS, and IQS evaluations at VIALIS.

Indeed, as recommended in section 6.2.1, the formalisation of the integral method enables the development of a "true" decision support system. This emphasises the relation to the current BI-tool even more, for the BI-tool does, in fact, meet the formal definition of a decision support system. By formalising the method, including its modules, the BI-tool could be replaced, or developed further by including the method, as a way to perform the QQS, and IQS evaluations even quicker.

However, an important notion is that the method treats all inefficiencies equally: if the generated performance is significantly "better" than the reference performance, it is still treated as an inefficiency, the same way as if it were as significantly "worse." This approach in which all inefficiencies are treated the same is chosen to simplify the development, and testing of the method. However, as indicated by various experts at VIALIS, and as introduced in section 3.1.1, it might be more logical to treat "better" inefficiencies differently from the other inefficiencies. This is based on the idea that if e.g. (far) less delay is measured than computed (reference delay), the traffic signal controller might have outperformed itself. When implementing the presented method in practice, it is recommended to address this issue, and thus to to investigate how a "better" generated performance with respect to the reference performance can assessed differently, and what this means for the steps of the integral evaluation and diagnosis method.

Furthermore, before implementation in practice, it is recommended to examine exact value of the inefficiency detector output result in terms of periods of inefficient performance, in relation to how it is used by traffic engineers in practice. Additionally, it is recommended investigate how the integral traffic signal controller evaluation and diagnosis method performs when using data from practice (e.g. data from the KWC), rather than simulation data. Also, this enables a pilot study in which the integral method is tested on a real case as part of a QQS evaluation, alongside the current practice of a QQS evaluation.

Altogether, it is recommended to invest time, and effort in the further development of the integral traffic signal controller evaluation and diagnosis method, given its promising test results of the case study discussed in this thesis.

6.2.3. Future research

The recommendations on the further development of the integral traffic signal controller evaluation and diagnosis method address mostly the shortcomings of the integral method as presented in this thesis. However, the integral method introduced several aspects that require further research in a more general sense, as listed below:

- Because of the assumptions used in the models to compute the reference performance, it was concluded that the models needed to be calibrated. In this thesis, this related to the reference performance models for delay, and queue length. However, it is possible that the reference performance models of other multi-variable performance indicators, including multi-variable performance indicators that were not selected in this thesis (e.g. red-light running), must be calibrated as well. Therefore, it is recommended to investigate, per potential multi-variable performance indicator, whether the corresponding reference performance model has to be calibrated. Furthermore, it is recommended to investigate whether other computational models for e.g. delay, and queue length are available, or can be developed that does not have to be calibrated, at least not regarding the assumption on e.g. arrival pattern.
- The tested integral evaluation and diagnosis method did not include a reference performance model for the degree of saturation. As stated in section 6.2.1, it is recommended to investigate whether this is needed, and if that is the case, to implement it in the method. This research could focus on how a reference performance for the degree of saturation could be defined in the first place, and which inputs it should use. The latter introduces a more general recommendation, namely to investigate whether the measured cycle time, and green time are sufficient to compute any reference performance, thereby relating to the definition of good performance, as discussed in section 4.2. Indeed, the use of a computed cycle time, and green time might yield other insightful results, and even enable the computation of a reference degree of saturation. Therefore, future research could focus on other definitions of good performance, how this affects the computation of reference performances, in particular for the degree of saturation, and how this affects potential recalibration of the reference performance models.
- Together with the testing of the integral method using data from practice, as mentioned in section 6.2.2, it is recommended to investigate how this might affect the inaccuracies of the reference performance models. This relates to the inaccuracies of the system observation. Future research could then, for instance, focus on how the assumptions made in the system observation, and reference performance models amplify, or maybe cancel out inaccuracies.

- The inaccuracies of the reference performance models are explicitly assessed with respect to the assumption on the arrival pattern. However, other potential causes of inaccuracies are identified as well in section 4.3, e.g. green time lags, and acceleration-deceleration delay. In future research, it is recommended to investigate the effect of these potential inaccuracies. Special attention to acceleration-deceleration delay is recommended, in particular with respect to the considered reference performance model(s) that might use this describing variable.
- Regarding the case study results, it is recommended to assess the presented integral traffic signal controller evaluation and diagnosis method more elaborately. Although the results are promising (using the method, it was possible to correctly diagnose the problems in the blind alternatives), the results are based on one signalised intersection, for which only two traffic flow volume sets are simulated. Moreover, the settings of the simulation case study limit the reliability of the results (e.g. the configuration of the intersection in terms of number of approaches, and number of lanes per approach). To come to more reliable, and decisive results, the method has to be tested on more signalised intersections – with different configurations as well – with more traffic flow volume sets, and using more (simulation) data. This could also be data from practice (e.g. QQS and IQS data), as introduced is section 6.2.2.
- Lastly, it is recommended to investigate the way the policies of road authorities are formulated. At the moment, the policies dictate a static threshold or reference value that is used in the BI-tool. However, the traffic system consists of many relationships that affect the performance of e.g. a signalised intersection, that might not be captured in the current formulation of policies for signalised intersections with static threshold or reference values. Therefore, it is recommended to investigate whether this way of formulating policies is still adequate. This also relates to the innovations, and developments currently taking place in the world of traffic engineering, and traffic signalisation in particular, for instance with the rise of intelligent traffic signal controllers which control traffic based on other aspects than solely the presence of vehicles, but rather based on vehicle arrivals, etc., such as look-ahead traffic signal control algorithms.

References

Akçelik, R. (1980). *Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals*. Vermont South, Australia: Australian Road Research Board.

Akçelik, R. (1981). *Traffic Signals: Capacity and Timing Analysis. Australian Road Research Board*. Research Report ARR No. 123.

American Society for Quality. (2019). *Plan-Do-Check-Act Cycle*. Retrieved March 28, 2019, from American Society for Quality: https://asq.org/quality-resources/pdca-cycle

Andersson, M. (2011, May 17). *Maintenance of traffic signal system performance: Handbok Trafiktekniskt underhåll (TTU)*. Trafikverket.

AWV. (2009). *Veilige Wegen en Kruispunten*. Brussels, Belgium: Agentschap Wegen en Verkeer (AWV).

AWV. (2011). *Verkeerslichtengeregelde kruispunten*. Brussels, Belgium: Agentschap Wegen en Verkeer (AWV).

Bal, M., Fatih Amasyali, A., Sever, H., Kose, G., & Demirhan, A. (2014). Performance Evaluation of the Machine Learning Algorithms Used in Inference Mechanism of a Medical Decision Support System. *The Scientific World Journal* (2014). doi:10.1155/2014/137896

Balasha, T., & Toledo, T. (2015). MESCOP: A Mesoscopic Traffic Simulation Model to Evaluate and Optimize Signal Control Plans. *Transportation Research Record: Journal of the Transportation Research Board* (2488), pp. 1-9. doi:10.3141/2488-01

Balke, K. N., Charara, H., & Parker, R. (2005). *Development of a Traffic Signal Performance Measurement System (TSPMS)*. Texas Transportation Institute, Texas Department of Transportation, and the Federal Highway Administration (FHA). College Station, Texas, U.S.A.: Texas Transportation Institute.

Blokpoel, R. J., Krajzewicz, D., & Nippold, R. (2010). Unambiguous Metrics for Evaluation of Traffic Networks. *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1277-1282). IEEE. doi:10.1109/ITSC.2010.5625135

Bolman, B., Jak, R. G., & Van Hoof, L. (2018). Unravelling the myth – The use of Decisions Support Systems in marine management. *Marine Policy* (87), 241-249. doi:10.1016/j.marpol.2017.10.027

Bullock, D. M., & Day, C. M. (2009). Performance measures for managing urban traffic signal systems. *Urban Transport XV: Urban Transport and the Environment 15*(107), 167-175. doi:10.2495/UT090161

CROW. (2006). *Handboek Verkeerslichtenregelingen* (Vol. Publication 213). Ede, the Netherlands. ISBN: 9066284447.

CROW. (2012). Bijlage 6. Analyseren van een voertuigafhankelijke regeling. In CROW, *Verkeersmanagement - Verkeerskundig beheer van regel- en informatiesystemen*. Ede, the Netherlands: CROW. Retrieved from http://kennisbank.crow.nl/KennisModule

CROW. (2014). *Handboek verkeerslichtenregelingen 2014* (Vol. Publication 343). Ede, the Netherlands.

Dakic, I., Mladenović, M. N., Stevanovic, A., & Zlatkovic, M. (2018). Upgrade Evaluation of Traffic Signal Assets: High-Resolustion Performance Measurement Framework. *Promet – Traffic & Transportation* (30), pp. 323-332. doi:10.7307/ptt.v30i3.2518

De Leeuw, M. (2010). *Memo Kruispuntbeoordeling*. Houten, the Netherlands.

Dion, F., Rakha, H., & Kang, Y.-S. (2004). Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research Part B*, *38*, pp. 99-122. doi:10.1016/S0191-2615(03)00003-1

DTV Consultants. (2017). *COCON Help versie 9.0.b*. Breda, the Netherlands.

Elkosantini, S., & Ahmed, F. (2014). Decision fusion for signalized intersection control. *Kybernetes*, *44*(1), pp. 57-76. doi:10.1108/K-08-2013-0185

FGSV. (2010). *Richtlinien für Lichtsignalanlagen (RiLSA): Lichtzeichenanlagen für den Straßenverkehr*. Forschungsgsellschaft für Straßen- und Verkehrswesen (FGSV), Arbeitsgruppe Verkehrsmanagement, Cologne, Germany.

Ford, F. N. (1985). Decision Support Systems and Expert Systems: A Comparison. *Information & Management*, 8(1), pp. 21-26. doi:10.1016/0378-7206(85)90066-7

Grontmij. (2001). *Afwikkeling eerste deel wachtrij*. Grontmij, Verkeer & infrastructuur, De Bilt, the Netherlands.

Hsu, W.-Y. (2018). A decision-making mechanism for assessing risk factor significance in cardiovascular diseases. *Decision Support Systems*, *115*, pp. 64-77. doi:10.1016/j.dss.2018.09.004

Hu, J., Fontaine, M. D., Park, B., & Ma, J. (2016). Field Evaluations of an Adaptive Traffic Signal—Using Private-Sector Probe Data. *Journal of Transport Engineering*, *142*(1). doi:10.1061/(ASCE)TE.1943-5436.0000806

Huang, T., Poddar, S., Aguilar, C., Sharma, A., Smaglik, E., Kothuri, S., & Koonce, P. (2018). Building Intelligence in Automated Traffic Signal Performance Measures with Advanced Data Analytics. *Journal of the Transportation Research Board: Transportation Research Record*, pp. 1-13. doi:10.1177/0361198118791380

Koonce, P., Rodegerdts, L., Lee, K., Qualye, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., Urbanik, T. (2008). *Signal Timing Manual*. Portland, Oregon, U.S.A.: Federal Highway Administration (FHA).

Kosmatopoulos, E., Papageorgiou, M., Bielefeldt, C., Dinopoulou, V., Morris, R., Mueck, J., Richards, A., Weichenmeier, F. (2006). International comparative field evaluation of a traffic-responsive signal control strategy in three cities. *Transportation Research Part A*(40), pp. 399-413. doi:10.1016/j.tra.2005.08.004

Krajzewicz, D., Blokpoel, R. J., Niebel, W., Hebenstreit, C., Härri, J., Milano, M., Bellini, A.-C., Stützle, T. (2014). Towards a unified Evaluation of Traffic Light Algorithms. *Proceedings of the 5th Conference: Transport Solutions from Research to Deployment Transport Research Arena*. Paris, France: Transport Research Arena 2014.

Krol, L., Prinsen, L., & Misdom, L. (2018). *Onderzoek detectieconfiguratie en signaalgroepafhandeling*. IVER, Goudappel Coffeng, and Vialis, Deventer, the Netherlands.

Lavrenz, S. M., Day, C. M., Smith, W. B., Sturdevant, J. R., & Bullock, D. M. (2016). Assessing Longitudinal Arterial Performance and Traffic Signal Retiming Outcomes. *Transportation Research Record: Journal of the Transportation Research Board* (2558), pp. 66-77. doi:10.3141/2558-07

Legato, P., & Mazza, R. M. (2018). A decision support system for integrated container handling in a transshipment hub. *Decision Support Systems*, *108*, pp. 45-56. doi:10.1016/j.dss.2018.02.004

Lidbe, A. D., Tedla, E. G., Hainen, A. M., & Jones Jr., S. L. (2017). Analytical Techniques for Evaluating the Implementation of Adaptive Traffic Signal Control Systems. *Journal of Transport Engineering, Part A: Systems*, *143*(5). doi:10.1061/JTEPBS.0000034

Lindgaard, G., Pyper, C., Frize, M., & Walker, R. (2009). Does Bayes have it? Decision Support Systems in diagnostic medicine. *International Journal of Industrial Ergonomics*, *39*, pp. 524-532. doi:10.1016/j.ergon.2008.10.011

Liu, X., Lu, R., Ma, J., Chen, L., & Qin, B. (2016). Privacy-Preserving Patient-Centric Clinical Decision Support System on Naïve Bayesian Classification. *IEEE Journal of Biomedical and Health Informatics, 20*(2), pp. 655-668. doi:10.1109/JBHI.2015.2407157

Loureiro, A. L., Miguéis, V. L., & Da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems, 114*, pp. 81-93. doi:10.1016/j.dss.2018.08.010

Martensen, H., Diependaele, K., Daniels, S., Van den Berghe, W., Papadimitriou, E., Yannis, G., Van Schagen, I., Weijermars, W., Wijnen, W., Filtness, A., Talbot, R., Thomas, P., Machata, K., Aigner Breuss, E., Kaiser, S., Hermitte, T., Thomson, R., Elvik, R. (2019). The European road safety decision support system on risks and measures. *Accident Analysis and Prevention, 125*, pp. 344-351. doi:10.1016/j.aap.2018.08.005

Martin, P. T., & Hockaday, S. L. (1995). SCOOT - An Update. *ITE Journal*, pp. 44-48.

Matsoukis, E. C. (2005). Calculating the performance index for isolated and coordinated traffic signal controlled intersections. *WIT Transactions on The Built Environment 77*(77), pp. 681-690.

Messai, N., Thomas, P., Lefebvre, D., & El Moudni, A. (2015). Neural networks for local monitoring of traffic magnetic sensors. *Control Engineering Practice, 13*, pp. 67-80. doi:10.1016/j.conengprac.2004.02.005

Moalla, D., Elkosantini, S., & Darmoul, S. (2013). An artificial immune network to control traffic at a single intersection. *5th International Conference on Industrial Engineering and Systems Management*, (pp. 842-848). Rabat, Morocco.

Moore II, J. E., Mattingley, S. P., MacCarley, C. A., & McNally, M. G. (2005). Anaheim Advanced Traffic Control System Field Operations Test: A Technical Evaluation of SCOOT. *Transportation Planning and Technology, 28*(6), 465-482. doi:10.1080/03081060500515622

Mouwen, A., Weiland, P. I., & Quirijns, G. (2004). *Effecten TOVERgroen*. AGV Adviseurs in mobiliteit, and Arane Adviseurs in verkeer en vervoer. 's-Hertogenbosch, the Netherlands: Provincie Noord-Brabant.

National Transportation Operations Coalition. (2007). *Traffic Signal Audit Guide*. Washington, D.C., U.S.A.

O'Sullivan, D., Fraccaro, P., Carson, E., & Weller, P. (2014). Decision time for clinical decision support systems. *Clinical Medicine, 14*(4), pp. 338-341. doi:10.7861/clinmedicine.14-4-338

Planath, S., Johansson, C.-H., Lindqvist, E., Eliasson, J., Lind, G., & Marton, M. (2003). *ITS Effektsamband, Uppdatering av Effektsamband 2000 med avseende på ITS*. Borlänge, Sweden: Vägverket.

Power, D. J. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Cedar Falls, Iowa, U.S.A.: Quorum Books. ISBN: 156720497X.

PTV AG. (2016). *PTV VISSIM 8 User Manual*. Karlsruhe, Germany: PTV AG.

Qatawneh, Z., Alshraideh, M., Almasri, N., Tahat, L., & Awidi, A. (2019). Clinical decision support system for venous thromboembolism risk classification. *Applied Computing and Informatics, 15*, pp. 12-18. doi:10.1016/j.aci.2017.09.003

Radivojevic, D., & Stevanovic, A. (2017a). *Manual on Performance of Traffic Signal Systems: Assessment of Operations and Maintenance*. Florida Atlantic University. Tallahassee, Florida, U.S.A.: Florida Department of Transportation.

Radivojevic, D., & Stevanovic, A. (2017b). Framework for Quantitative Annual Evaluation of Traffic Signal Systems. *Transportation Research Record: Journal of the Transportation Research Board*(2619), pp. 20-35. doi:10.3141/2619-03

Siteur, C., Van Run, H., Mak, W., Mens, J., Van Liere, P., & Van Vught, L. (2002). *Signaalgroepafwikkeling en Detectieconfiguratie*. IVER werkgroep "Signaalgroepafwikkeling". Initiatiefgroep Verkeersregeltechnici Rijkswaterstaat en Provincies (IVER).

So, J., Stevanovic, A., & Koonce, P. (2016). Estimating performance of traffic signals based on link travel times. *Journal of Advanced Transportation* (50), pp. 786-801. doi:10.1002/atr.1375

Sprague Jr., R. H., & Carlson, E. D. (1982). *Building effective decision support systems*. Prentice Hall Professional Technical Reference. ISBN: 0130862150.

Statens Vegvesen. (2007). *Håndbok 142: Trafikksignalanlegg: Planlegging, drift og vedlikehold*. Oslo, Norway: Statens Vegvesen Vegdirektoratet. ISBN: 9788272076084.

Sunkari, S. R. (2004). The Benefits of Retiming Traffic Signals. *ITE Journal, 74*(4). ISSN: 01628178.

't Hoen, W., Vanhuysse, S., & Los, R. (2013). *Handleiding verkeerskundige evaluaties: Handvatten voor wegbeheerders om de prestatie van hun VRI's in kaart te brengen*. Ministerie van Infrastructuur en Milieu, Groene Golf Team. Delft, the Netherlands: Ministerie van Infrastructuur en Milieu.

Teply, S. (1993). Probability of Overload at Signalized Intersections. *Transportation Research Board* (1398), pp. 101-110.

Tian, Z., Ohene, F., & Hu, P. (2011). Arterial Performance Evaluation on an Adaptive Traffic Signal Control System. *6th International Symposium on Highway Capacity and Quality of Service* (pp. 230-239). Stockholm, Sweden: Procedia Social and Behavioral Sciences 16. doi:10.1016/j.sbspro.2011.04.445

TRB. (2000). *Highway Capacity Manual (HCM) 2000*. Washington, DC, U.S.A.: Transportation Research Board (TRB) of the National Sciences. ISBN: 0309066816.

TRB. (2012). *Highway Capacity Manual (HCM) 2010, Volume 4: Applications Guide, Chapter 31: Signalized Intersections: Supplemental*. Washington, DC, U.S.A.: Transportation Research Board (TRB) of the National Sciences.

Turban, E., & Watkins, P. R. (1986). Integrating Expert Systems and Decision Support Systems. *MIS Quarterly, 10*(2), pp. 121-136. doi:10.2307/249031

Vägverket & Svenska Kommunförbundet. (2004). *Vägar och gators utformning: Trafiksignaler*. Borlänge, Sweden: Vägverket. ISSN: 14019612.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory* (Second edition). Springer. ISBN: 9781441931603.

Vejdirektoratet. (2012). *Bedre trafiksignaler: Konsekvenser for trafikanter og samfund*. Copenhagen, Denmark: Vejdirektoratet. ISBN: 9788770606721.

VIALIS. (2017). *Kwaliteitscentrale, versie 4.8.3.1*. IT&M. Vialis.

VIALIS. (n.d. [a]). *Various Internal QQS/IQS Reports*. Elst/Houten, the Netherlands.

VIALIS. (n.d. [b]). *Quick Quality Scan*. Retrieved August 28, 2018, from Vialis.nl: https://www.vialis.nl/nl/portfolio/advies/quick-quality-scan

Webster, F. V. (1958). Traffic Signal Settings. *Road Research Laboratory* (Technical Paper No. 39).

Wen, W. (2008). A dynamic and automatic traffic light control expert system for solving the road congestion problem. *Expert Systems with Applications, 34*, pp. 2370-2381. doi:10.1016/j.eswa.2007.03.007

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques* (Fourth edition). Morgan Kaufmann. ISBN: 9780128042915.

Wood, K., Palmer, J. P., & Bretherton, R. D. (1994). Congestion Analysis and Diagnosis in UTC Networks. *7th International Conference on Road Traffic Monitoring and Control*, (pp. 172-176).

WSDOT. (2014). *Protocol for VISSIM Simulation*. Washington State Department of Transportation (WSDOT), Olympia, Washington, U.S.A. Retrieved from http://www.wsdot.wa.gov/NR/rdonlyres/378BEAC9-FE26-4EDA-AA1F-B3A55F9C532F/0/VISSIMProtocol.pdf

Wu, S. K., Hunter, M. P., Lee, C., & Rodgers, M. O. (2011). Evaluation of Traffic Signal Control System Using a System-wide Performance Measure under Two-Fluid Model Theory. *KSCE Journal of Civil Engineering*, 15(2), pp. 395-403. doi:10.1007/s12205-011-1087-y

Wünsch, G., Bölling, F., Von Dobschütz, A., & Mieth, P. (2015). Bavarian Road Administration's Use of Probe Data for Large-Scale Traffic Signal Evaluation Support. *Transportation Research Record: Journal of the Transportation Research Board* (2487), pp. 88-95. doi:10.3141/2487-08

A. List of definitions

This appendix lists the definitions of the terms, terminology, and abbreviations used in the thesis. If applicable, the Dutch translation or definition is given as well. The table with definitions continues on the next pages.

Term/abbreviation	Definition
Active mode	Pedestrians, bicyclists, etc., as a transport mode.
Alternative realisation	Green phase realisations of signal groups outside the pre-defined block sequence. A signal group (a) receives an alternative realisation if a conflicting, primary signal group (b) in a given block is already served, or has no request, whereas signal group (a) is not a primary signal group in that block (CROW, 2006).
BI-tool	Assessment tool in the KWC used for evaluation reports. In the BI-tool, threshold and reference values are given. The BI-tool checks whether, and when these values are exceeded. This is input for QQS and IQS evaluations (Vialis, n.d. [a]). Dutch definition: *Beoordeling Instrument tool*.
Block sequence	The follow-up of conflicting signal groups, based on the critical conflict group, given as predefined blocks. In vehicle actuated traffic signal controllers, a block sequence is a basic setting, usually fixed to fit the demand of a given period. In intelligent traffic signal controllers, the block sequence might change cycle to cycle. Also known as phase plan (CROW, 2006).
COCON	*COherent CONglomeraat van verkeersregeltechnische software*: software package developed by the Dutch company DTV consultants (2017), used to design (fixed time) traffic signal controllers. COCON is quite commonly used in practice in the Netherlands, as part of the design process of a (vehicle-actuated) traffic signal controller, e.g. to design a block sequence.
Coupled request/ realisation	Request for a (parallel signal) group with no traffic detected. If on one of two (parallel) signal groups traffic is detected (signal group (a)), and no traffic on the other signal group (b), but the green time, amber time, and clearance time are approximately equal, or even less for signal group (b) with respect to signal group (a), virtual traffic is used to make a coupled request for signal group (b) together with signal group (a) with real traffic. In the same way, a (parallel) signal group might receive a coupled realisation under the same conditions (CROW, 2006). Dutch definition: *mee-aanvraag/-realisatie*.
Credibility	The credibility of a signalised intersection describes how a road user experiences the signalised intersection. E.g., a signalised intersection is not credible if road users frequently have to wait unnecessary (CROW, 2006). In this research, credibility also includes other related aspects: (i) clearance lost time, (ii) red-light running, and (iii) block sequence: do the settings for these aspects seem justified, and/or logical for a road users, from the point of view of said road user? If not, the settings are considered as not credible.
Delay	Stop delay, plus the time a vehicle loses to brake (deceleration delay) from, and accelerate (acceleration delay) to the desired or free flow speed, or restricted speed (e.g. when driving in a platoon) (CROW, 2006; Dion, et al., 2004). Dutch definition: lost time (*verliestijd*).
Detector oscillation	When a detector is oscillating, it repeatedly becomes active-inactive for extremely short periods of time while the detector should be active constantly in that total period of time (Vialis, n.d. [a]). Dutch definition: *jutteren*.
Detector overperformance	Overperformance of a detector implies that the detector is detecting “more traffic than there is driving”. In practice, this comes down to a detector that is detecting traffic on an adjacent lane, for instance (Vialis, n.d. [a]). Dutch definition: *bovengedrag*, or *overspraak*.
Detector underperformance	Underperformance of a detector implies that the detector is detecting “less traffic than there is driving”. In other words, the detector “does not see all traffic”, even if traffic is passing the detector (Vialis, n.d. [a]). Dutch definition: *ondergedrag*.
FOT	Field Operation Test: testing of a scheme, project, algorithm, etc., in practice.

Term/abbreviation	Definition
IQS	Instant Quality Scan: a semi-standardised traffic signal controller evaluation method, used by VIALIS, similar to, and with the same objective as QQS, though without certain aspects of the QQS, thereby reducing the time needed to perform the quality scan (VIALIS, n.d. [b]).
KAR	*Korte Afstand Radio* (Short Distance Radio): a wireless system, functioning as a target detector system, used to realise prioritised green phases for public transport, and/or emergency services (CROW, 2006).
KWC	*Kwaliteitscentrale* (Quality Centre): a software program to evaluate signalised intersections, and used to collect data on traffic performance (e.g. traffic flow volumes, red-light-running, etc.), phases, and functioning of the detectors (VIALIS, 2017).
LOS	Level Of Service: a measure for the performance of, in this case, traffic signal controllers. A high LOS represents a good performance.
pce-value	Passenger Car Equivalent value: converted unit for traffic flow volume, which expresses various vehicle classes, and/or modes as passenger cars. More precisely, for each vehicle class, a pce-value exists which represent the number of passenger cars that could pass a given point, or road section at an intersection instead of the given vehicles (bus, truck, etc.) in the time that vehicle uses (CROW, 2006). Dutch definition: *personenauto-equivalent (pae)*.
Permitted conflict	Conflict between traffic signal groups at a signalised intersection that can have green and/or amber at the same moment. The conflict is solved using the regular traffic rules (CROW, 2006). Dutch definition: *deelconflict*.
Phase failure	A phase failure occurs when consecutively green and amber phase is terminated before the queue is fully dissolved (Balke, et al., 2005).
Phase-log data	Data on the detector states, and green phases, including green sub-phases, amber phases, and red phases, including red sub-phases. The data is used to assess the traffic signal controller. Thereto, the phase-log data also includes data regarding special signals, such as bridge openings, etc., if applicable to the assessed signalised intersection (VIALIS, 2017; VIALIS, n.d. [a]).
Platoon ratio	A measure for how platoons arrive, computed as the ratio between vehicles arriving during the green indication, and the total traffic flow volume. The higher the platoon ratio, the more favourable the platoon arrives at the signalised intersection, i.e. all vehicles arrive during green (TRB, 2000).
Primary realisation	Green phase realisations of signal groups within the pre-defined block sequence (CROW, 2006).
Protected conflict	Conflict between traffic signal groups at a signalised intersection that cannot have green and/or amber at the same moment. It is customary that signal groups that cross perpendicularly are signal groups with protected conflicts (CROW, 2006). A signalised intersection with only protected conflicts is also called a conflict-free intersection. Dutch definition: *conflictvrij*.
QQS	Quick Quality Scan: a semi-standardised traffic signal controller evaluation method, used by VIALIS, with the objective to provide road authorities with data, and knowledge on how their traffic signal controllers are functioning, both in technical and traffic performance terms (VIALIS, n.d. [b]).
QSR	Queue Storage Ratio: a measure for the likelihood that blockage of a lane will occur, denoted as the ratio of the back-of-queue to the available vehicle storage length. Blockage will occur if QSR ≥ 1.0 (TRB, 2012).
Saturation flow	Capacity during the green phase. In other words, the maximum amount of traffic one lane on a signalised intersection can facilitate under given traffic circumstances, including traffic signal control program, road design, and traffic flow composition, if the traffic signal control would give that lane green for one hour (CROW, 2006). Dutch definition: *afrijcapaciteit*.

Term/abbreviation	Definition
SCATS	Sydney Coordinated Adaptive Traffic System: an urban network control algorithm, developed in Sydney, Australia, with the objective to optimise the green split, cycle times, and offset of multiple signalised intersection in a region, using feedback from measurements at the stop lines (Tian, et al., 2011).
SCOOT	Split Cycle, and Offset Optimisation Technique: an urban network control algorithm, developed in the United Kingdom, with the objective to optimise the green split, cycle times, and offset of multiple signalised intersection in a region, based on the departure of an upstream intersection (Martin & Hockaday, 1995).
Stop delay	Delay due to standing still (CROW, 2006; Dion, et al., 2004). Dutch definition: waiting time (*wachttijd*).
Tovergroen	*Toepassen voorzieningen voor vrachtverkeer* (Application of services for trucks): a priority setting for trucks, by prolonging the green phase of a stream to prevent the need for a truck to brake to standstill, and re-accelerate. The underlying principle is that Tovergroen reduces the delays of, and emissions by trucks (Mouwen, Weiland, & Quirijns, 2004).
Unnecessary waiting	Stop delay experienced by a vehicle during a red phase, while it could have had a green phase (Vialis, n.d. [a]). Unnecessary waiting on signal group (a) occurs when a conflicting signal group (b) has a green phase, though without vehicles using that conflicting signal group (b), implying that vehicles on signal group (a) are waiting for "nothing". Unnecessary waiting does not include the lost time due to the clearance time from signal group (a) to (b), because the clearance time is mandatory for traffic safety reasons.
VECOM	Vehicle Communications: a road-side based system, functioning as a target detector system, used to realise prioritised green phases for public transport, and/or emergency services (CROW, 2006).
VLH	Vehicle Lost Hour: number of hours of delay with respect to the free flow travel time, in relation to the traffic flow volume (CROW, 2006). One (1) VHL equals one (1) vehicle that experienced a delay of one (1) hour, or e.g. sixty (60) vehicles that experienced one (1) minute delay each. Dutch definition: *voertuigverliesuur*, *VVU*.
Wait-in-green	Traffic signal control tactic. When a traffic signal controller has wait-in-green as set tactic, the traffic signal controller will keep giving green to a (set) of non-conflicting signal group(s). The maximum green time timer for such a signal group is not active. If a vehicle approaches the intersection on a signal group that is in conflict with the wait-in-green signal group(s), a request is placed, and the green phase(s) of the wait-in-green signal group(s) is terminated as soon as possible, enabling a green phase for the requested signal group. When the requested signal group returns to red, the wait-in-green signal group(s) return to green, even if there is no traffic approaching on those signal groups (CROW, 2006). Dutch definition: *wachtstand groen*.
Wait-in-red	Traffic signal control tactic. When a traffic signal controller has wait-in-red as set tactic, the traffic signal controller will return to all-red (all signal groups have red) if there is no traffic. When a vehicle approaches on one signal group of the intersection, green will be given for that signal group only. If multiple vehicles approach the intersection simultaneously on conflicting signal groups, green will be given according to the block sequence, or other relevant settings with a similar effect (CROW, 2006). Dutch definition: *wachtstand rood*.

B. Overview of QQS and IQS evaluation report analysis

As part of the literature study discussed in chapter 2, various evaluation reports from practice (QQS, and IQS evaluation reports of VIALIS (n.d. [a])) are analysed. These reports offer insight in the current practice at VIALIS regarding traffic signal controller evaluations. Besides, they offer additional information on traffic signal controller evaluations at points where the scientific literature was found to be insufficient, e.g. the countermeasures possible to improve the functioning of a traffic signal controller. Therefore, the QQS, and IQS evaluation reports from practice are analysed, whereas a total of 214 evaluation reports are considered (142 (66%) QQS reports, and 72 (34%) IQS reports). The results are presented for three main sections of these reports: (i) policy constraints, as formulated by the road authority[1], (ii) performance indicators, as analysed by the traffic engineer to assess the traffic signal controller, and (iii) countermeasures, as proposed by the traffic engineer, though without specifying the exact countermeasure if applicable (e.g. "time settings for green phases" is mentioned instead of "increasing maximum green time"). The results are given as tabled lists, with the most often mentioned, or included policy constraints, performance indicators, or countermeasures are listed on top. The least often mentioned, or included aspects are given at the bottom of each list. This ranking is based on the number of times it is mentioned in both the QQS, and IQS reports.

Table B-1 | Policy constraints in QQS reports.

Policy constraint	Mentioned in QQS reports [#]	Given in QQS reports by road authorities [%][2]
Maximum accepted (stop) delay active modes	12	92.31%
Maximum accepted unnecessary waiting	12	92.31%
Preventing double stops	11	84.62%
Maximum accepted cycle time on-peak	10	76.92%
Maximum accepted red-light running	9	69.23%
Maximum accepted (stop) delay motorised traffic	6	46.15%
Maximum accepted cycle time	5	38.46%
Average accepted (stop) delay active modes	3	23.08%
Average accepted (stop) delay motorised traffic	3	23.08%
Maximum accepted cycle time off-peak	3	23.08%
Maximum accepted periods with oversaturation	3	23.08%
PT priority	3	23.08%
Guaranteed coordination between intersections	2	15.38%
Maximum accepted (stop) delay PT	2	15.38%
Traffic safety	2	15.38%
Flexibility	1	7.69%
Maximal allocation of green time	1	7.69%

[1] Policy constraints are exclusively used in QQS reports (VIALIS, n.d.).
[2] 100% = 142 QQS reports.

Table B-2 | Performance indicators in QQS, and IQS reports.

Performance indicators	Mentioned in QQS reports [#]	Mentioned in QQS reports [%][1]	Mentioned in IQS reports [#]	Mentioned in IQS reports [%][2]
Detection and detectors	142	100.00%	72	100.00%
Traffic flow volume	138	97.18%	72	100.00%
Stop delay	136	95.77%	46	63.89%
Unnecessary waiting	138	97.18%	42	58.33%
Time settings for green, amber, and red phases	112	78.87%	32	44.44%
Red-light running	137	96.48%	4	5.56%
Cycle time	129	90.85%	9	12.50%
Degree of (over)saturation	90	63.38%	44	61.11%
Selective detection (e.g. PT detection)	74	52.11%	59	81.94%
Double) stops	122	85.92%	4	5.56%
Time-of-day program(s)	105	73.94%	20	27.78%
Queue length	109	76.76%	1	1.39%
Average amount of stops	99	69.72%	2	2.78%
Saturation flow	92	64.79%	1	1.39%
V/C-ratio	91	64.08%	1	1.39%
Block sequence	58	40.85%	26	36.11%
Intersection load ratio	83	58.45%	0	0.00%
Utilisation of green phase	75	52.82%	1	1.39%
Urban network control	50	35.21%	11	15.28%
Geometric design of intersection	31	21.83%	24	33.33%
Clearance times	52	36.62%	0	0.00%
KAR[3]/VECOM[3] (e.g. emergency services, PT, etc.))	44	30.99%	8	11.11%
Duration of green	47	33.10%	4	5.56%
Prioritised PT realisation	38	26.76%	10	13.89%
Prioritised congestion realisation (congestion intervention)	34	23.94%	13	18.06%
Permitted conflicts[3] versus protected conflicts[3]	41	28.87%	3	4.17%
Acoustic signals	29	20.42%	2	2.78%
Waiting time predictors	16	11.27%	1	1.39%
VLH	14	9.86%	1	1.39%
Unused capacity	9	6.34%	0	0.00%
Speed	4	2.82%	2	2.78%
Tovergroen[3]	3	2.11%	3	4.17%
Congestion/traffic management outputs	2	1.41%	2	2.78%
Delay	4	2.82%	0	0.00%

[1] 100% = 142 QQS reports.
[2] 100% = 72 IQS reports.
[3] See appendix A for a clarification of the abbreviation, or term.

Table B-3 | Countermeasures in QQS, and IQS reports. The top-10 of most mentioned countermeasures is given in *italic*, whereas countermeasures that relate to problems outside the scope of the research are excluded (table continues on next page).

Countermeasures	Mentioned in QQS reports [#]	Mentioned in QQS reports [%][1]	Mentioned in IQS reports [#]	Mentioned in IQS reports [%][2]
Detection/detectors: Technical functioning	113	79.58%	52	72.22%
Time settings for green phases	*120*	*84.51%*	*35*	*48.61%*
Geometric design of intersection: Revision geometric design (incl. adding lanes, new signal posts, revisioning road markings, etc.)	*75*	*52.82%*	*12*	*16.67%*
Selective detection: Technical functioning	43	30.28%	39	54.17%
Detection/detectors: Gap time	*53*	*37.32%*	*22*	*30.56%*
Detection/detectors: Occupancy time	*40*	*28.17%*	*15*	*20.83%*
Geometric design of intersection: Revision technical drawing	39	27.46%	8	11.11%
Alternative realisations	*29*	*20.42%*	*7*	*9.72%*
Coupled request/realisation	*24*	*16.90%*	*10*	*13.89%*
Time-of-day program(s)	*21*	*14.79%*	*11*	*15.28%*
Synchronised- or pre-start	*24*	*16.90%*	*1*	*1.39%*
Wait-in-green/wait-in-red	*21*	*14.79%*	*3*	*4.17%*
Urban network control	19	13.38%	4	5.56%
Selective detection: Exit loop location	17	11.97%	5	6.94%
Prioritised congestion realisation	*17*	*11.97%*	*4*	*5.56%*
Detection/detectors: Sensitivity	14	9.86%	6	8.33%
Time settings for amber phases	20	14.08%	0	0.00%
Acoustic signals	19	13.38%	0	0.00%
Detection/detectors: Request function	5	3.52%	13	18.06%
Selective detection: Entrance loop location	10	7.04%	8	11.11%
Detection/detectors: Location	15	10.56%	1	1.39%
PT priority	13	9.15%	3	4.17%
Block sequence	10	7.04%	3	4.17%
Congestion detection	10	7.04%	2	2.78%
Congestion/traffic management outputs	8	5.63%	3	4.17%
Coupled intersections	8	5.63%	3	4.17%
Time settings for red phases	11	7.75%	0	0.00%
Waiting time predictors	10	7.04%	1	1.39%
Clearance times	10	7.04%	0	0.00%

[1] 100% = 142 QQS reports.
[2] 100% = 72 IQS reports.

Countermeasures	Mentioned in QQS reports [#]	Mentioned in QQS reports [%][1]	Mentioned in IQS reports [#]	Mentioned in IQS reports [%][2]
Detection/detectors: Coupling	6	4.23%	4	5.56%
Permitted conflicts versus protected conflicts	7	4.93%	1	1.39%
Extra realisation active modes	7	4.93%	0	0.00%
Credibility	5	3.52%	0	0.00%
Detection/detectors: Process management	4	2.82%	1	1.39%
Detection/detectors: Software	4	2.82%	1	1.39%
Active mode priority	2	1.41%	1	1.39%
Tovergroen[3]	1	0.70%	2	2.78%
Flexibility	2	1.41%	0	0.00%
New cables	1	0.70%	0	0.00%

[1] 100% = 142 QQS reports.
[2] 100% = 72 IQS reports.
[3] See appendix A for a clarification of the abbreviation, or term.

C. Elaborate and additional formulas

In the report, the definitions of several multi-variable performance indicators, and their describing variables are introduced, without discussing the detailed formulas. In this appendix, these detailed formulas are discussed, as well as some additional formulas, and definitions, such as the cycle time, and maximum green time.

Additionally, for reference purposes, the variables, introduced in this appendix, are defined Table C-1 as well.

Table C-1 | Notation and definitions in appendix C (table continues on next page).

Symbol	Unit	Math. domain	Description
a_{acc}	m/s²	$\mathbb{R}$	Acceleration rate
a_{dec}	m/s²	$\mathbb{R}$	Deceleration rate
C	pce/h	$\mathbb{R}$	Capacity
d	s/pce	$\mathbb{R}$	Delay
d_1	s/pce	$\mathbb{R}$	Uniform control delay
d_2	s/pce	$\mathbb{R}$	Incremental delay
d_3	s/pce	$\mathbb{R}$	Initial queue delay
d_{ad}	s/pce	$\mathbb{R}$	Acceleration-deceleration delay
f	-	$\mathbb{R}$	Uniform delay progression adjustment factor
f_p	-	$\mathbb{R}$	Supplemental adjustment factor
j, m	-	$\mathbb{N}$	Signal group index
I	-	$\mathbb{R}$	Upstream metering adjustment factor
k	-	$\mathbb{R}$	Incremental delay factor
L	pce	$\mathbb{R}$	Queue length
L_1	pce	$\mathbb{R}$	First queue length term
L_2	pce	$\mathbb{R}$	Second queue length term
L_3	pce	$\mathbb{R}$	Third queue length term
L_s	pce	$\mathbb{R}$	Initial queue length
n	-	$\mathbb{N}$	Number of lanes
P	-	$\mathbb{R}$	Proportion of vehicles arriving during effective green
q	pce/h	$\mathbb{R}$	Traffic flow volume
q'	pce/s	$\mathbb{R}$	Traffic flow volume rate: $q' = q/3600$
q_G	pce/h	$\mathbb{R}$	Traffic flow volume
q'_G	pce/s	$\mathbb{R}$	Traffic flow volume rate arriving during effective green: $q'_G = q_G/3600$
q_R	pce/h	$\mathbb{R}$	Traffic flow volume
q'_R	pce/s	$\mathbb{R}$	Traffic flow volume rate arriving during effective red: $q'_R = q_R/3600$
R	-	$\mathbb{R}$	Platoon ratio
s	pce/h	$\mathbb{R}$	Saturation flow
s'	pce/s	$\mathbb{R}$	Saturation flow rate: $s' = s/3600$
T_C	s	$\mathbb{R}$	Cycle time
T_G	s	$\mathbb{R}$	Green time
$T_{G,\text{eff}}$	s	$\mathbb{R}$	Effective green time
$T_{G,\text{max}}$	s	$\mathbb{R}$	Maximum green time
$T_{G,\text{min}}$	s	$\mathbb{R}$	Minimum green time
T_R	s	$\mathbb{R}$	Red time

Symbol	Unit	Math. domain	Description
$T_{R,\text{eff}}$	s	$\mathbb{R}$	Effective red time
$\mathcal{T}$	s	$\mathbb{R}$	Duration of analysis period
$\mathcal{T}_a$	s	$\mathbb{R}$	Adjusted duration of unmet demand during $\mathcal{T}$
$\mathcal{T}_u$	s	$\mathbb{R}$	Duration of unmet demand during $\mathcal{T}$
u	-	$\mathbb{R}$	Fraction effective green per cycle: $u = T_{G,\text{eff}}/T_C$
v_a	km/h	$\mathbb{R}$	Average speed
v_s	km/h	$\mathbb{R}$	Threshold speed defining a stopped vehicle
y	-	$\mathbb{R}$	Load ratio: $y = q/s$
x	-	$\mathbb{R}$	Degree of saturation: $qT_C/sT_{G,\text{eff}}$
x_{usi}	-	$\mathbb{R}$	Degree of saturation of upstream signalised intersection
α	-	$\mathbb{R}$	Scaling parameter
λ_1	s	$\mathbb{R}$	Green time start lag
λ_2	s	$\mathbb{R}$	Green time end lag (utilised amber)
τ	-	$\mathbb{N}$	Analysis period index
ω	-	$\mathbb{R}$	Delay parameter

C.1. Cycle time

The cycle time T_C [s] is the time needed to give every signal group at a signalised intersection at least one possibility for a green phase. This definition implies that the cycle time is an intersection wide variable. Also, it implies that it is not necessary that every signal group actually gets a green phase. This is especially relevant for vehicle-actuated traffic signal controllers, where the cycle time is computed afterwards, and is thus in fact measured, usually per signal group – thus instead of an intersection-based cycle time, a signal group-based cycle time is considered. In these types of traffic signal controller, the actual green time that is to be realised is not known – only the minimum, and maximum green times are known – because the length of the green phase is determined by the presence of vehicles. Then, at least the minimum green time is given, and based on the presence of vehicles, or traffic, the green phase is lengthened until the maximum green time is reached (CROW, 2006). Besides, it is possible that signal groups receive multiple green phases during the same cycle, for instance due to alternative realisations, again emphasising a signal group-based cycle time. This is also accounted for when measuring the cycle time by indicating whether a green phase was a primary realisation – thus no alternative realisation. The cycle time per signal group is measured as the time between the start-of-green moments of two consecutive primary green phase realisations. On the other hand, a different definition of measured cycle time per signal group could be considered, namely that the measured cycle time is the time from the first request, until the end of the first following amber phase. That way, the sub-phase of wait-in-red is removed from the measurements, thereby reducing the cycle time in (highly) undersaturated conditions (e.g. during night time). However, given the formulas for delay, and queue length, as will be discussed in the following sections C.3, and C.4, these two definitions introduce several problems. That is, these definitions, in particular the first one, with a signal group-based cycle time measured as the time between the start-of-green moments of two consecutive primary green phase realisations, result in the fact that the queue that is built up during the red phase of cycle (i) is served in the green phase of the next cycle (ii). Therefore, a third definition for the cycle time per signal group is introduced, namely that the (measured) cycle time per signal group is the time between two consecutive start-of-red moments. That way, the cycle time covers the red phase and the first following green phase, thereby ensuring that vehicles that arrived during red, and pass the stop line in the consecutive green phase, are accounted for in the same cycle. It must be noted that this definition treats primary realisations, alternative realisations, and other non-primary realisations the same, thus all as primary realisations. Again, this is to account for the vehicles arriving during red, and pass the stop line in the consecutive green phase during the same cycle. The cycle time is measured in this thesis as visualised in Figure C-1.

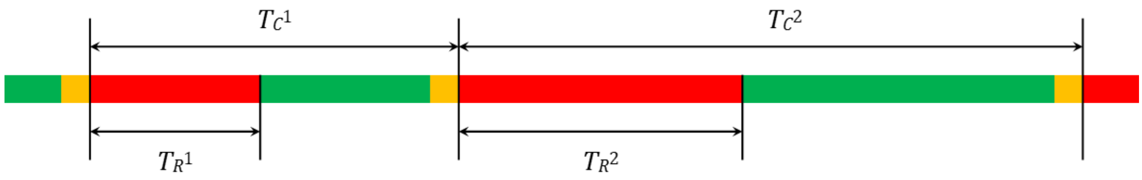

Figure C-1 | Cycle time T_C^j based on the red time for two consecutive red phases, T_R^1, and T_R^2.

C.2. Maximum green time

The green time T_G [s] is already shortly introduced in the definition of the cycle time, and in section 4.1.5.1. As stated, in vehicle-actuated traffic signal controllers, the green time depends on the presence of vehicles, or traffic. In evaluations, and other formulas the effective green time $T_{G,\text{eff}}$ [s] is commonly used instead of the green time T_G. The effective green time denotes the time that traffic is being served, thereby excluding the start-up lost time (start lag λ_1 [s]), and including that part of the amber phase in which traffic is still passing the stop line (utilised amber time, or end lag λ_2 [s]), see Figure C-2. In mathematical terms, this comes down to:

$$T_{G,\text{eff},j} = T_{G,j} - \lambda_1 + \lambda_2 \tag{C-1}$$

with signal group index j.

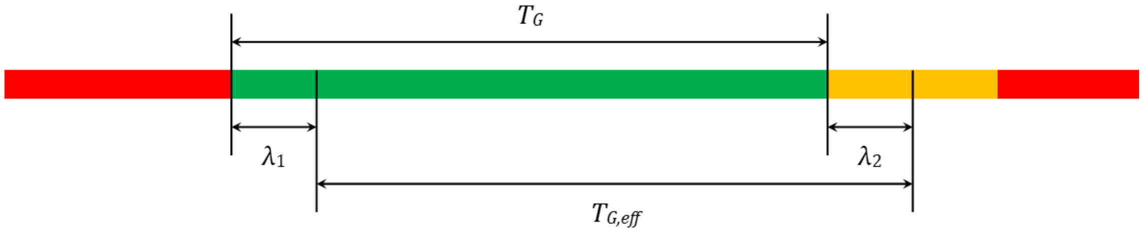

Figure C-2 | Green time T_G and effective green time $T_{G,\text{eff}}$.

When the green time is known, the red time T_R [s] can be found as the time left in the cycle, minus the amber time:

$$T_{R,j} = T_C - T_{G,j} - T_{A,j} \tag{C-2}$$

In a similar way, the effective red time $T_{R,\text{eff}}$ [s] is found via:

$$T_{R,\text{eff},j} = T_C - T_{G,\text{eff},j} \tag{C-3}$$

Furthermore, the minimum green time $T_{G,\text{min}}$ [s] denotes the minimum duration of the green time, and is also known as guaranteed green time. The setting for the minimum green time is given by the Dutch guidelines, which state a minimum green time ranging from 3 to 7 s is to be applied, depending on the applied detection configuration regardless of the movements on that signal group (CROW, 2014). On the other hand, for the maximum green time $T_{G,\max}$ [s], there is no fixed value available. Instead, the maximum green time is computed per signal group j. In the Netherlands, the computation of the maximum green time makes use of the green time $T_{G,i}$ as designed in the most optimal fixed time traffic signal controller:

$$T_{G,\max,j} = \alpha T_{G,j} \tag{C-4}$$

with scaling parameter α, for which the value ranges between 1.2, and 1.4, depending on the number of blocks in the block sequence, and with $T_{G,j} \geq T_{G,\min,j}$ (CROW, 2014). Because the value for α is quite arbitrarily found, and could result in low maximum green times, VIALIS defined another method to find the maximum green time. Again, the green time $T_{G,i}$ as designed in the most optimal fixed time traffic signal controller is used:

$$T_{G,\max,i} = \max\left\{15, \left\lceil \frac{120 \cdot \left(x_j T_{G,j} / \max x_m\right)}{T_C} \right\rceil\right\} \tag{C-5}$$

with the degree of saturation $x = qT_C/sT_{G,\text{eff}}$ [-] (see section 4.1.5.1), and the maximum degree of saturation x_m [-] of all signal groups m at the intersection, including j. Note that now the minimal maximum green time is 15 s. Although this minimal maximum green time is quite arbitrary, just as the value for scaling parameter α in Equation C-4, it is considered as a better alternative, for the arbitrary value is now not used to compute the maximum green time, but is rather a boundary condition, thereby reducing the effect it has on the computed maximum green time. Therefore, the computation of the maximum green time is done using Equation C-5.

C.3. Delay

The delay denotes the additional travel time a vehicle, or road user experiences when travelling through a network, or, in this case, when passing a signalised intersection, with respect to the free flow travel time. Regarding delays, an important distinction is to be made, as shortly discussed in chapter 1, namely between delay on the one hand, and stop delay on the other hand. The latter denotes the waiting time – the delay a vehicle experiences due to standing still –, whereas the former includes stop delay, acceleration delay, and deceleration delay (CROW, 2006; Dion, et al., 2004), see Figure 4-4.

In the Dutch guidelines, as well as in other international guidelines, several formal delay formulas are discussed. The delay formulas of Webster (1958), and Akçelik (1981) are mentioned by the Dutch guidelines (CROW, 2006). However, these formulas have several limitations. First, the formulas assume that vehicles arrive according to respectively a uniform distribution, and a Poisson distribution. Secondly, the formulas are calibrated with respect to fixed time controllers. Although it is stated that the formulas could still be applied to estimate delays of vehicle- and/or traffic actuated traffic signal controllers, the Dutch guidelines also state that caution is needed because the exact effects are not known (CROW, 2006). On the other hand, the HCM2000 introduced an alternative delay formula, which addresses the aforementioned shortcomings, since the HCM2000 delay formula is able to handle (i) other arrival patterns, for instance with platoons, and (ii) vehicle- and/or traffic actuated traffic signal controllers, as will be discussed below. Because of this, the HCM2000 delay formula is preferred, and therefore used in this research. Therefore, the delay formulas defined by Webster (1958), and Akçelik (1981) are not discussed further here. Please, refer to the respective references for more information about the delay formulas of Webster (1958), and Akçelik (1981)

The HCM2000 delay formula is a sum of three terms, and expresses the delay d in seconds per pce[1] [s/pce] (TRB, 2000):

$$d = d_1 + d_2 + d_3 \tag{C-6}$$

The three terms represent respectively (1) the uniform control delay d_1 [s/pce], (2) the incremental delay d_2 [s/pce], and (3) the initial queue delay d_3 [s/pce]. The delay terms, and thus the delay itself, are computed per analysis period, with duration $\mathcal{T}$ [h].

C.3.1. Uniform control delay

The computation of the uniform control delay d_1 is split up in two parts, depending on the degree of saturation $x = qT_C/sT_{G,\text{eff}}$ [-] (see section 4.1.5.1). Both parts relate to the cycle time T_C [s], and effective green time $T_{G,\text{eff}}$ [s], and include the progression adjustment factor f [-], given as:

$$f = \frac{(1 - Ru)f_p}{1 - u} \tag{C-7}$$

with fraction effective green per cycle $u = T_{G,\text{eff}}/T_C$ [-], platoon ratio R [-], and supplemental adjustment factor f_p [-], whereas f_p depends on the range of R, as given in Table C-2. The platoon ratio is given as:

$$R = \frac{q_G}{q} \tag{C-8}$$

where q_G [pce/h] equals the traffic flow volume arriving on the green indication, and q [pce/h] the (total) traffic flow volume.

Table C-2 | Values for supplemental adjustment factor f_p depending on the range of platoon ratio R (TRB, 2000).

Arrival type	Progression quality	Range of R	Default R	Supplemental adjustment factor f_p
1	Very poor	$R \leq 0.50$	0.333	1.00
2	Unfavourable	$0.50 < R \leq 0.85$	0.667	0.93
3	Random arrivals	$0.85 < R \leq 1.15$	1.000	1.00
4	Favourable	$1.15 < R \leq 1.50$	1.333	1.15
5	Highly favourable	$1.50 < R \leq 2.00$	1.667	1.00
6	Exceptional	$R > 2.00$	2.000	1.00

Furthermore, the computation of the uniform control delay d_1 includes the duration of duration of unmet demand $\mathcal{T}_u$ [h], representing the part of the analysis period during which the demand exceeded capacity, which is computed using the degree of saturation. Note that the formulas also take the initial queue length L_s [pce] into account as a condition:

[1] See appendix A for a clarification of the abbreviation, or term.

$$\mathcal{T}_u = 0 \qquad \text{if } L_s = 0 \tag{C-9}$$

$$\mathcal{T}_u = \mathcal{T} \qquad \text{if } L_s > 0 \text{ and } x \geq 1 \tag{C-10}$$

$$\mathcal{T}_u = \min\left\{\mathcal{T}, \frac{L_s}{(1-x)s}\right\} \qquad \text{if } L_s > 0 \text{ and } x < 1 \tag{C-11}$$

Altogether, the uniform control delay d_1 is then computed via:

$$d_1 = \frac{(T_C - T_{G,\text{eff}})\mathcal{T}_u}{2\mathcal{T}} + \frac{T_C(1-u)^2}{2(1-y)} \cdot \frac{(\mathcal{T} - \mathcal{T}_u)f}{\mathcal{T}} \qquad \text{for } x < 1 \tag{C-12}$$

$$d_1 = \frac{(T_C - T_{G,\text{eff}})\mathcal{T}_u}{2\mathcal{T}} + \frac{T_C - T_{G,\text{eff}}}{2} \cdot \frac{(\mathcal{T} - \mathcal{T}_u)f}{\mathcal{T}} \qquad \text{for } x \geq 1 \tag{C-13}$$

with load ratio $y = q/s$ [-].

C.3.2. Incremental delay

The second term d_2 shows resemblance to the overflow queue formula of Akçelik (1981), though with some adaptations. For instance, the incremental delay term d_2 includes factors that account for (i) random arrivals (nonuniform arrivals, and random delay), (ii) overflow queues (delay due to (over)saturation), and (iii) type of signal control into account. The (iii) type of signal control is accounted for with incremental delay factor k [-]. Then, the value for k depends on the degree of saturation, and the unit extension, see Table C-3. The unit extension is the minimum gap size (in seconds) between two successive vehicles for which the green phase is terminated, and is defined per signal group. For unit extension values higher than 5.0, the corresponding value for k is to be found with extrapolation, whereas the maximum value is $k = 0.50$. Also, it is assumed that for degrees of saturation exceeding 1.00, the value for $k = 0.50$, since such oversaturated vehicle-actuated signalised intersections tend to behave as a fixed time traffic signal controller, for which a fixed value $k = 0.50$ is applied as well. Alternatively, the value for k can be computed via

$$k = k_{\min} + (1 - 2k_{\min})(x - 0.5) \qquad k_{\min} \leq k \leq 0.5 \tag{C-14}$$

with a minimum incremental delay factor $k_{\min}$ [-], which equals the value for k for $x = 0.50$, given the unit extension of the analysed signal group. Then, the computed value for k may not exceed $k = 0.50$.

Table C-3 | Values for incremental delay factor k (TRB, 2000).

Degree of saturation → ↓ Unit extension [s]	**≤ 0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**≥ 1.00**
≤ 2.0	0.04	0.13	0.22	0.32	0.41	0.50
2.5	0.08	0.16	0.25	0.33	0.42	0.50
3.0	0.11	0.19	0.27	0.34	0.42	0.50
3.5	0.13	0.20	0.28	0.35	0.43	0.50
4.0	0.15	0.22	0.29	0.36	0.43	0.50
4.5	0.19	0.25	0.31	0.38	0.44	0.50
5.0	0.23	0.28	0.34	0.39	0.45	0.50
≥ 5.5	Extrapolate					
Fixed time controller	0.50					

Furthermore, an upstream metering adjustment factor I [-] is included to account for the effect of upstream signalised intersections, for instance on the arrival pattern (e.g. platoons), and therefore depends on the degree of saturation of the upstream signalised intersection x_{usi} [-]:

$$I = 1.0 - 0.91x_{\text{usi}}^{2.28} \qquad \text{for } x_{\text{usi}} \leq 1.0 \tag{C-15}$$

This equation only accounts for upstream signalised intersections within a radius 1.6 kilometres of the analysed signalised intersection. If the nearest upstream signalised intersection is further away, a value $I = 1.000$ is to be applied.

This results in the formula used to compute the incremental delay d_2:

$$d_2 = 900\mathcal{T}\left((x-1) + \sqrt{(x-1)^2 + \frac{8kIx}{s\mathcal{T}}}\right) \tag{C-16}$$

C.3.3. Initial queue delay

The initial queue delay d_3 is used to account for the additional delay road users experience due to the fact that the queue of the previous green phase is not completely served. Thus, at the start of the red phase, there is already a queue waiting. The initial queue delay is 0 if there is no initial queue. Otherwise, the initial queue delay is computed via:

$$d_3 = \frac{1800L_s\mathcal{T}_u(1+\omega)}{s\mathcal{T}} \tag{C-17}$$

The computation uses three factors: (i) the initial queue length L_s, (ii) the duration of unmet demand $\mathcal{T}_u$, and (iii) a delay parameter ω [-]. All these factors include the degree of saturation x. Then, (i) computes the queue length at the start of the red phase due to oversaturation in a preceding analysis period, whereas the computation takes the initial queue length, and degree of saturation of the previous analysis period into account:

$$L_s^\tau = \max\{0, L_s^{\tau-1} + s^{\tau-1}\mathcal{T}(x^{\tau-1} - 1)\} \tag{C-18}$$

Secondly, (ii) the duration of unmet demand $\mathcal{T}_u$ represents the part of the analysis period during which the demand exceeded capacity, which is computed using the degree of saturation. Note that the formulas also take the initial queue length L_s into account as a condition.

Again, it must be noted that the formulas given in Equations C-7 to C-18, are rewritten with respect to the literal formulas given in the HCM2000. The formulas are rewritten to account for oversaturated conditions, due to the mathematical issues that arose when applying $x \geq 1$ in the literal formulas in the HCM2000.

Thirdly, (iii) delay parameter ω is used as a sort of weight factor for how much delay has built up: the longer the initial queue L_s, the larger the delay parameter becomes. the delay parameter is only relevant if the duration of unmet demand $\mathcal{T}_u$ equals the duration of the analysis period $\mathcal{T}$. Also, although the value of the delay parameter increases for an increasing initial queue length, this happens at a decreasing rate:

$$\omega = 1 - \frac{s\mathcal{T}(1-x)}{L_s} \qquad \text{if } \mathcal{T}_u = \mathcal{T}, \text{ otherwise } \omega = 0 \tag{C-19}$$

C.4. Queue length

The queue length denotes the number of vehicles standing in the queue at a signalised intersection.

In literature, several definitions and formulas for the queue length are given. For instance, Akçelik (1980) defined the queue length in relation to (i) the overflow queue, (ii) the queue that built up during red, and (iii) the queue that built up during the first portion of the green time. However, this formula has the same limitations as the delay model of Akçelik (1981), for it includes the same underlying variables, and models. Because the delay model of Akçelik (1981) is considered as inadequate, the same is concluded for the queue length formula. Therefore, an alternative formula is used, as given by the HCM2010 (TRB, 2012).

This queue length estimation model computes the queue length L [pce] based on three terms:

$$L = L_1 + L_2 + L_3 \tag{C-20}$$

C.4.1. First queue length term

C.4.1.1. Proportion of vehicles arriving during green

The first term L_1 [pce] represents the number of fully stopped vehicles, due to the traffic signal phasing. This term uses the converted variables for traffic flow volume rate $q' = q/3600$ [pce/s], and saturation flow rate $s' = s/3600$ [pce/s]. Also, it includes the traffic flow arrival rate during red q'_R [pce/s]:

$$q'_R = \frac{(1-P)T_C q'}{T_{R,\text{eff}}} \tag{C-21}$$

In this definition of q'_R, the variable P [-] is included. This variable represents the proportion of vehicles arriving during green, and thus denotes the number of vehicles that arrived during the green phase with respect to the total traffic flow volume. Note that this is similar to the definition of the platoon ratio R, as given in section C.3.1. The function of this variable is to account for different arrival types. This implies that it is originally used to compute the traffic flow arrival rate q'_G during effective green (TRB, 2012):

$$q'_G = \frac{P T_C q'}{T_{G,\text{eff}}} \tag{C-22}$$

However, the HCM2010 does not give the formula for the proportion of vehicles arriving during effective green P. Nevertheless, by rewriting the formula for q'_G, it was found that:

$$P = \frac{T_{G,\text{eff}} q'_G}{T_C q'} = u \cdot \frac{q'_G}{q'} \tag{C-23}$$

Previously, the platoon ratio R is defined as $R = q'_G/q'$, implying that $P = Ru$. In fact, this is in line with the original definition of the platoon ratio as given in HCM2000 (TRB, 2000), namely:

$$R = \frac{P T_C}{T_{G,\text{eff}}} = \frac{P}{u} \tag{C-24}$$

On the other hand, when rewriting Equation C-21, another definition for P is found:

$$P = 1 - (1 - u) \cdot \frac{q'_R}{q'} \tag{C-25}$$

Combining Equations C-23 and C-25, yields the formal definition for the traffic flow volume rate q' [pce/s]:

$$q' = q'_G u + q'_R (1 - u) \tag{C-26}$$

This definition implies that the total traffic flow volume rate equals the sum of the arrivals during effective green ($q'_G u$), and arrivals during effective red ($q'_R(1 - u)$).

Therefore, in this research, the proportion of vehicles arriving during effective green is defined as:

$$P = Ru \tag{C-27}$$

If, and only if, vehicles arrive according to a uniform distribution, regardless of the indicated phase, thus when $q'_G = q'$, and $R = 1.00$, it holds that $P = u$.

C.4.1.2. Acceleration-deceleration delay

Furthermore, the first term of the HCM2010 queue length estimation model, uses the acceleration-deceleration delay d_{ad} [s/pce] as variable, and condition. This variable describes the delay a vehicle experiences due to acceleration, and deceleration in the queue. This delay is computed using the average acceleration rate a_{acc} [m/s²], and deceleration rate a_{dec} [m/s²] of vehicles at the intersection, and the average speed v_a [km/h]. Also, a threshold value is used to define a stopped vehicle v_s [km/h]. In general, $v_s > 0$, because vehicles may drive slowly in queue. Therefore, a value of $v_s = 8.05$ km/h is used, which corresponds to the value as used in HCM2010. Furthermore, a correctional term is used, which is not given in the original formula in HCM2010 (TRB, 2012). This correction term is added to the formula to account for the transition from imperial units to metric units, because originally, the formula considered imperial units. When imperial units are applied, the correction term equal to $1/5.28$ should be set to 1.00. Altogether, the acceleration-deceleration delay is computed as:

$$d_{\text{ad}} = \frac{1}{5.28} \cdot \left(\frac{1}{a_{\text{acc}}} + \frac{1}{a_{\text{dec}}}\right)\left(\frac{\left(1.47(v_a - v_s)\right)^2}{2 \cdot 1.47 v_a}\right) \tag{C-28}$$

C.4.1.3. Number of fully stopped vehicles

Then, the first term of the HCM2010 queue length estimation model is computed via:

$$L_1 = q'_R T_{R,\text{eff}} + q'_G \left(\frac{q' T_C \left(1 - P - \frac{P d_{\text{ad}}}{T_{G,\text{eff}}}\right)}{s'(1 - P \cdot \min\{1, x\})} - d_{\text{ad}}\right) \qquad \text{if } d_{\text{ad}} \leq x T_{G,\text{eff}}(1 - P) \tag{C-29}$$

$$L_1 = q'_R \left(T_{R,\text{eff}} - d_{\text{ad}} + \frac{q' T_C (1 - P)\left(T_{R,\text{eff}} - d_{\text{ad}}\right)}{s'\left(T_{R,\text{eff}} - T_{G,\text{eff}}(1 - P) \cdot \min\{1, x\}\right)}\right) \qquad \text{if } d_{\text{ad}} > x T_{G,\text{eff}}(1 - P) \tag{C-30}$$

C.4.2. Second queue length term

The second term L_2 [pce] relates to the random fluctuations in demand causing random phase failures, and overflow queues. Therefore, this term L_2 uses the capacity C [pce/h], the incremental delay d_2, as given in section C.3.2, and the number of lanes n [#] of the considered signal group:

$$L_2 = \frac{C d_2}{3600 n} \tag{C-31}$$

C.4.3. Third queue length term

The third term L_3 [pce] is used to account for the initial queue L_s, as given by Equation C-18, and excludes the aforementioned overflow queue:

$$L_3 = \frac{L_s}{n} \qquad \text{if } q \geq C \tag{C-32}$$

$$L_3 = \frac{1}{n\mathcal{T}}\left(\mathcal{T}_a L_s + \frac{\mathcal{T}_a^2 (q - C)}{2}\right) \qquad \text{if } q < C \tag{C-33}$$

The definition of L_3 includes an alternative definition for the duration of unmet demand: the adjusted duration of unmet demand $\mathcal{T}_a$ [h]:

$$\mathcal{T}_a = \frac{L_s}{C - q} \leq \mathcal{T} \qquad \text{if } q < C\text{, otherwise } \mathcal{T}_a = \mathcal{T} \tag{C-34}$$

This queue length estimation model is based on time-space diagrams of signalised intersection approaches. Lastly, the queue length estimation model of HCM2010 is used explicitly in the computation of the queue storage ratio.

Furthermore, the queue length estimation model as given in Equations C-20 to C-34, is based on, and may only be applied on signal groups with protected phasing. In other words, the model presented here may not be applied for signal groups with permitted conflicts (TRB, 2012). Although the model could be used when considering e.g. permitted conflicts by making several alterations to the model, this is not considered in this research. In other words, this research focuses on signalised intersections with only protected conflicts. Therefore, the altered model is neither relevant, nor discussed in this research.

D. Problem and countermeasure selection simulation study

This thesis considered four traffic signal controller problems, and their corresponding countermeasures in particular to test the integral traffic signal controller evaluation and diagnosis method, as introduced in section 4.1.3:

- Activate alternative realisations;
- Adjust maximum green time settings according to the computed maximum green time, as computed using the formula used by VIALIS, see appendix C;
- Adjust the long detector gap time to the correct setting (0.0 s[1]), as specified by Siteur, et al. (2002), given an IVER 2002 detection configuration;
- Adjust the geometric design of intersection.

Since for the adjustments of the maximum green time, and long detector gap time there are various incorrect settings, "degrees of incorrectness", or variants, possible, it is tested in a simulation study which specific incorrect setting is included, as a way to further specify the countermeasures. The simulation study compared the inefficient performance of traffic signal controllers caused by various variants of the selected problems with each other, and with the base case generated performance of the (assumed) problem-free controller, as explained in this appendix. For the long detector gap times, and maximum green time settings, these variants are most relevant, since here multiple variants are tested. The problem of alternative realisations is tested to be deactivated alternative realisations, implying that only one variant. Although for the adjustments of the geometric design of intersection, there are also various countermeasures possible, this is not tested. Instead, the general countermeasure to improve the allocation of capacity at the intersection by adjusting the geometric intersection design is considered. An overview of these different variants is given in Figure 4-2 in section 4.1.3.

D.1. Simulation study set-up

The general simulation study set-up (simulation period, number of runs, and data collection and processing) is discussed in appendix E.

The simulation study uses a total of four alternatives, including a base case alternative, based on the problems listed above, see Table D-1. In this table it can be seen that the problem on inadequate geometric intersection design is not tested in this simulation study, because of the fact that this problem is only included to be able to assign inefficiencies to a problem if they cannot be assigned to any other problem, as explained in section 5.2.4. For the alternatives with the incorrect gap time (1), and incorrect maximum green time (2), multiple variants are tested. These variants relate to different variations in how incorrect the setting is of the corresponding problem. It holds that regarding (1) incorrect long detector gap times, a total of three variants are considered, see Table D-2. The first incorrect gap time is set on 1.0 s, the second is set on 1.5 s, and the third on 2.0 s. Gap times shorter than 1.0 s are not considered because the effect of those settings are expected to be too small to detect, since the correct gap time is 0.0 s. For instance, the effect of an incorrect gap time of 0.1 s with respect to a correct gap time of 0.0 s is expected to be too small to be detectable. On the other hand, given the limited time available, gap times larger than 2.0 s are not considered. Also, it is hypothesised that higher gap times, exceeding 2.0 s, have a too extreme effect. The same principle is used when defining the scenarios of alternative (2) incorrect maximum green time. Here, it is assumed that the correct maximum green time is more than sufficient, implying that maximum green time is only reached on rare occasions, and thus that a reduced maximum

[1] In practice, road authorities might deviate from the guidelines, and use another gap time as the "correct" gap time, given that particular situation. However, for simplicity, the settings as given in the guidelines of Siteur, et al. (2002) are considered in this thesis.

green time has a better detectable effect on the deviations of the generated performance with respect to the reference performance. To reduce the maximum green time relatively equal for all signal groups, the maximum green time reductions are given as percentages, starting at −10%, up to −50%, with steps of 10%. Additionally, a more extreme scenario of −75% is considered, yielding a total number of scenarios of six, see Table D-2.

Table D-1 | List of alternatives, with relevant settings corresponding to each problem per alternative.

Problems → ↓ Alternatives	Long detector gap time setting	Maximum green time setting	Alternative realisations
Alternative 0 (base case)	Correct gap time	Correct maximum green time	Activated
Alternative 1	Incorrect gap time	Correct maximum green time	Activated
Alternative 2	Correct gap time	Incorrect maximum green time	Activated
Alternative 3	Correct gap time	Correct maximum green time	Deactivated

Table D-2 | Variants for alternatives (alt.) 1, and 2, with for alternative (1) incorrect gap time settings [s], and for alternative (2) reduced maximum green times [%].

Variant	0 (correct)	1	2	3	4	5	6
Alt. 1	0.0 s	1.0 s	1.5 s	2.0 s	N/A	N/A	N/A
Alt. 2	0%	-10%	-20%	-30%	-40%	-50%	-75%

Furthermore, this simulation study uses deliberately faulty traffic signal controllers, which are traffic signal controllers with problems implemented on purpose. These problems might be applied on all signal groups at once, or on a specific set of signal groups. In this appendix, the problems are applied on three signal groups individually, depending on the corresponding direction, implying that each alternative consists of three scenarios: a problem on a right-turning signal group, a problem on a through-going signal group, and a problem on a left-turning signal group. To limit the time needed to perform this simulation study, only signal groups 01, 02, and 03 are considered for the signalised intersection discussed in section 5.2.1, and visualised in Figure 5-1. These signal groups are respectively a right-turning, through-going, and left-turning signal group. This results in three scenarios per alternative, per variant, as listed in Table D-3.

Table D-3 | Scenarios per alternative, per variant, with corresponding signal group, and direction.

Scenario	Signal group with problem implemented on purpose	Direction
1	01	Right-turn
2	02	Through
3	03	Left-turn

Lastly, it must be noted that this simulation study applies the integral evaluation and diagnosis method presented in this thesis, though only partly: only the inefficiency detector module is used.

D.2. Results

It must be noted that the results are only given for the maximum conflict group, plus the signal groups on which the scenarios are based, thus signal groups 01, 02, 03, 05, 09, and 12. That way, the amount of data presented is limited.

The results of the alternatives, and in particular per variant, show that certain problems have only a limited impact on the outcome of the inefficiency detector. For instance, the inefficiency ratios for delay, and queue length in alternative (1) regarding incorrect gap times do not show substantial changes per variant, as can be seen in Figure D-1, and Figure D-2 (with the inefficiency ratio as a percentage). Even more constant are the evaluation scores (Figure D-3, and Figure D-4), which lay between 8.0 and 9.0. However, for alternative (2), regarding the maximum green time, a clear increase in the inefficiency ratios, and decrease of the evaluation scores can be seen in Figure D-5 to Figure D-8. In particular in the cases where the maximum green time was reduced with 50% or more, the inefficiency ratios increase strongly (from approximately 18% to 52%, for the inefficiency ratio for delay for signal group 02 in alternative 2, scenario 2, with equal flows, top middle plot in Figure D-5). The same can be seen for the evaluation scores. Alternative (3), with deactivated alternative realisations, shows increased inefficiency ratio, and lowered evaluation scores for both delay, and queue length, in particular for the signal group on which the problem is implemented on purpose, see Figure D-9, and Figure D-10. Moreover, it can be seen that the higher the inefficiency ratio is, the lower the evaluation score is. This effect is best visible on signal group 01, which is observed to have quite a lot alternative realisations, in particular in block I (with signal groups 02, and 08, see Table 5-3 in section 5.2.2.2). This is due to the fact that signal group 01 has very few conflicts, and is thus able to have more alternative realisations.

Altogether, the results show that a more incorrect setting does not necessarily result in a higher inefficiency ratio, and/or lower evaluation score. This is true for the detector gap time settings in particular. For the maximum green times, a more incorrect setting does result in a clearly higher inefficiency ratio, and/or lower evaluation score, though mostly in the scenarios where the maximum green time is decreased most, i.e. −50%, and −75%.

Because for the incorrect gap time, no decisive results are found (one incorrect gap time does not necessarily result in a more inefficient performance than another incorrect gap time), the incorrect setting of 1.5 s is selected. For the incorrect maximum green time it is found that if the green time decreases with 50% or more, the performance of the traffic signal controller reduces strongly. Since 75% yields a too extreme result, 50% is selected. In short, the incorrect settings, and corresponding countermeasures are selected as given in Table D-4.

Table D-4 | Selected problems with corresponding incorrect setting, and specified countermeasure.

Problem	Incorrect setting	Countermeasure
Long detector gap times	1.5 s	Adjust long detector gap times to 0.0 s
Maximum green time settings	50% of computed maximum green time	Adjust maximum green time settings to computed maximum green time
Alternative realisations	Deactivated	Activate alternative realisations
Geometric intersection design	Insufficient capacity	Adjust geometric design of intersection

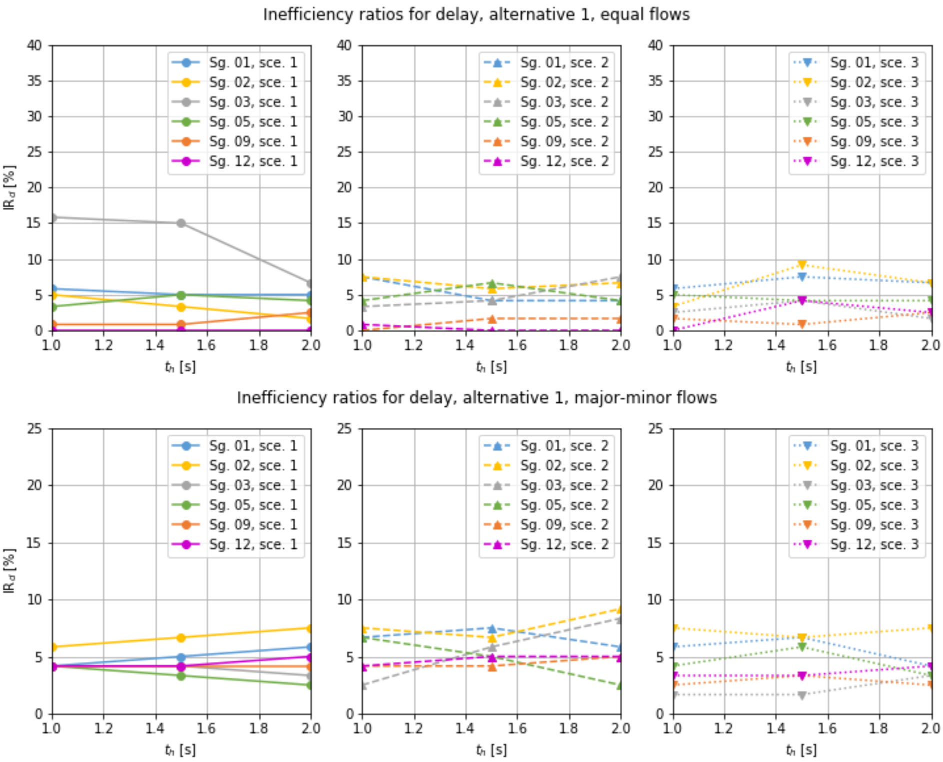

Figure D-1 | Inefficiency ratios for delay IR$_d$ [%] for alternative (1) incorrect gap time t_h [s], per variant of t_h [s], per scenario (sce.; as given in Table D-3), per signal group (sg.).

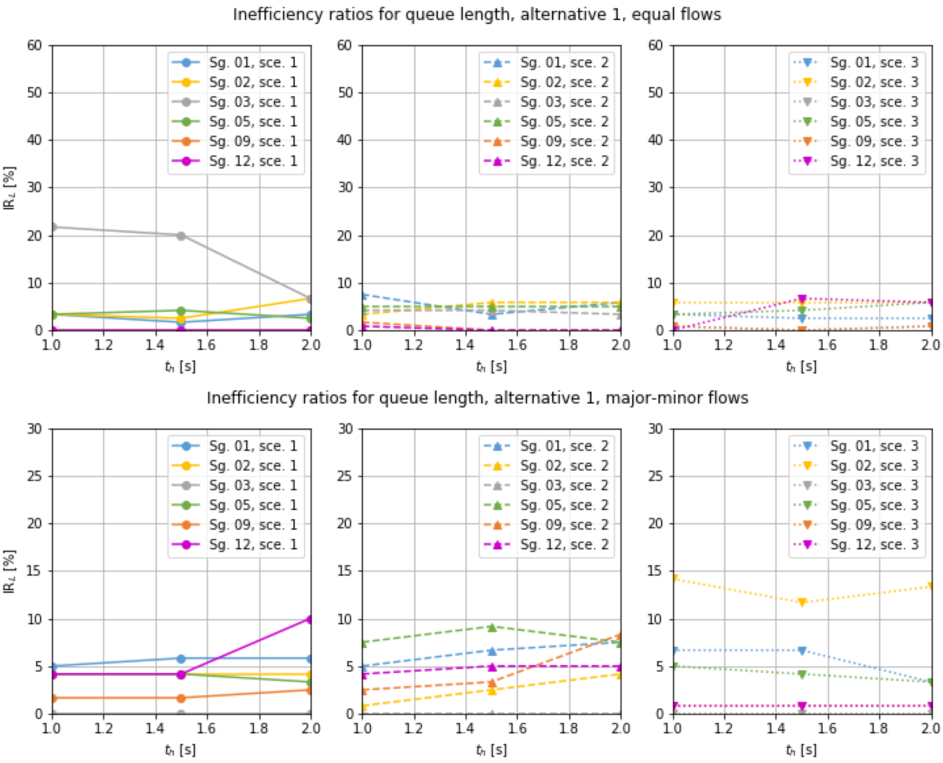

Figure D-2 | Inefficiency ratios for queue length IR$_L$ [%] for alternative (1) incorrect gap time t_h [s], per variant of t_h [s], per scenario (sce.; as given in Table D-3), per signal group (sg.).

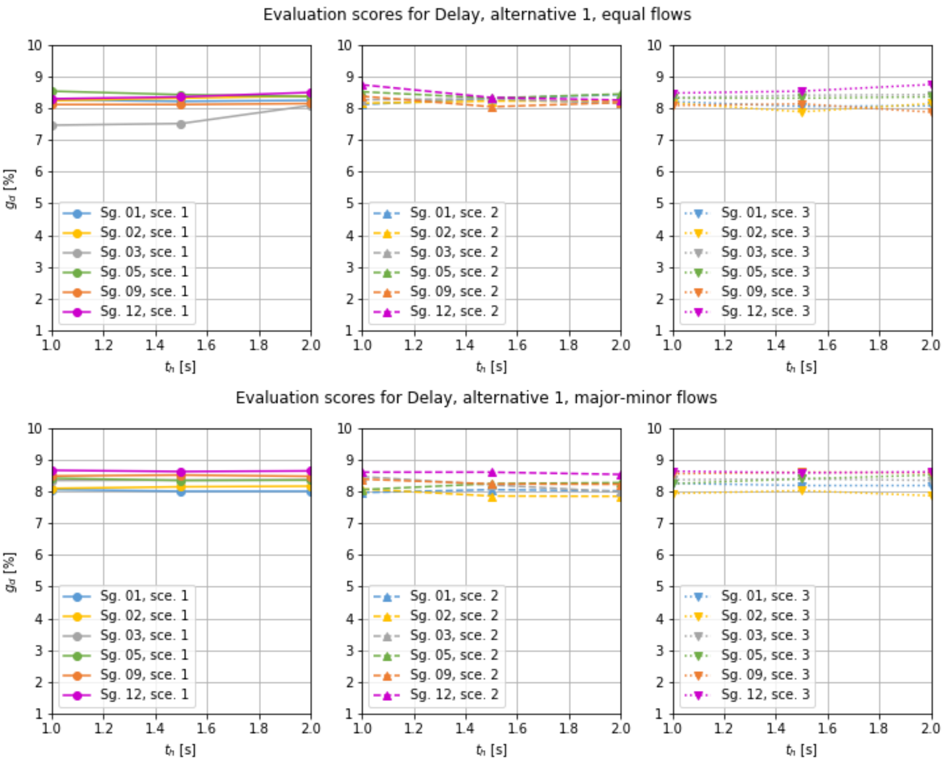

Figure D-3 | Evaluation scores for delay g_d [-] for alternative (1) incorrect gap time t_h [s], per variant of t_h [s], per scenario (sce.; as given in Table D-3), per signal group (sg.).

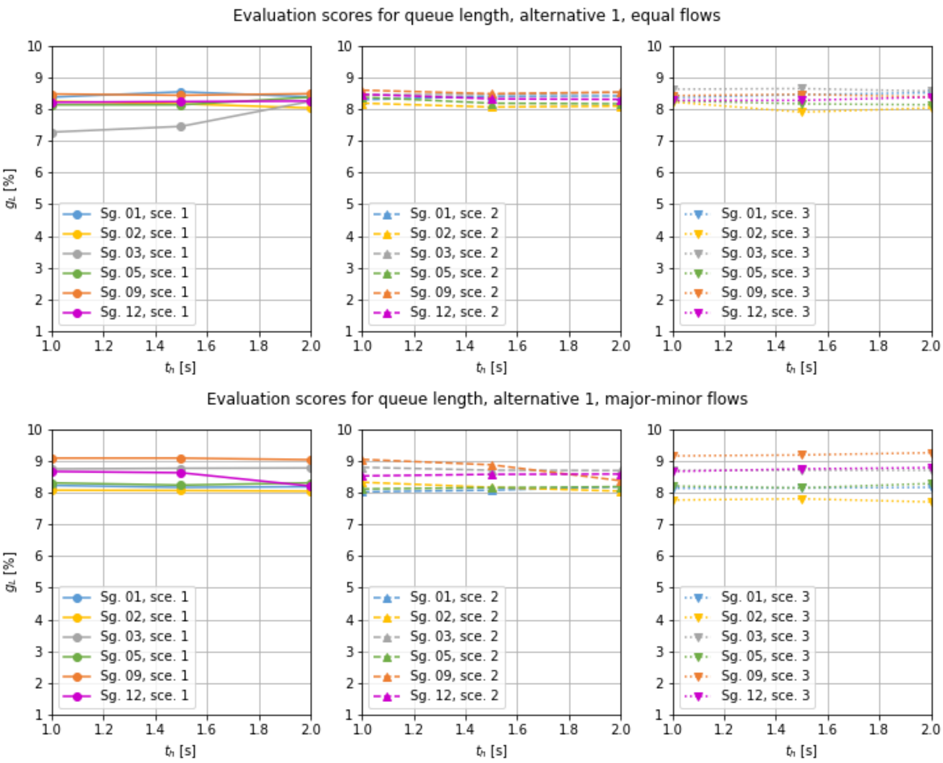

Figure D-4 | Evaluation scores for queue length g_L [-] for alternative (1) incorrect gap time t_h [s], per variant of t_h [s], per scenario (sce.; as given in Table D-3), per signal group (sg.).

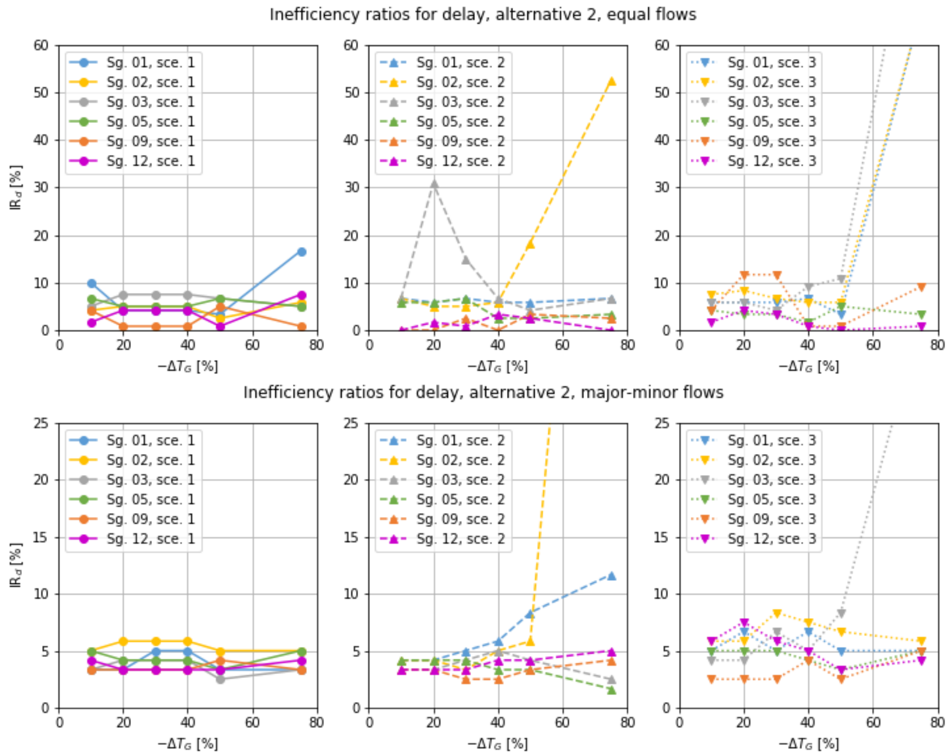

Figure D-5 | Inefficiency ratios for delay IR_d [%] for alternative (2) incorrect maximum green time T_G [s], per variant of $-\Delta T_G$ [%], per scenario (sce.; as given in Table D-3), per signal group (sg.).

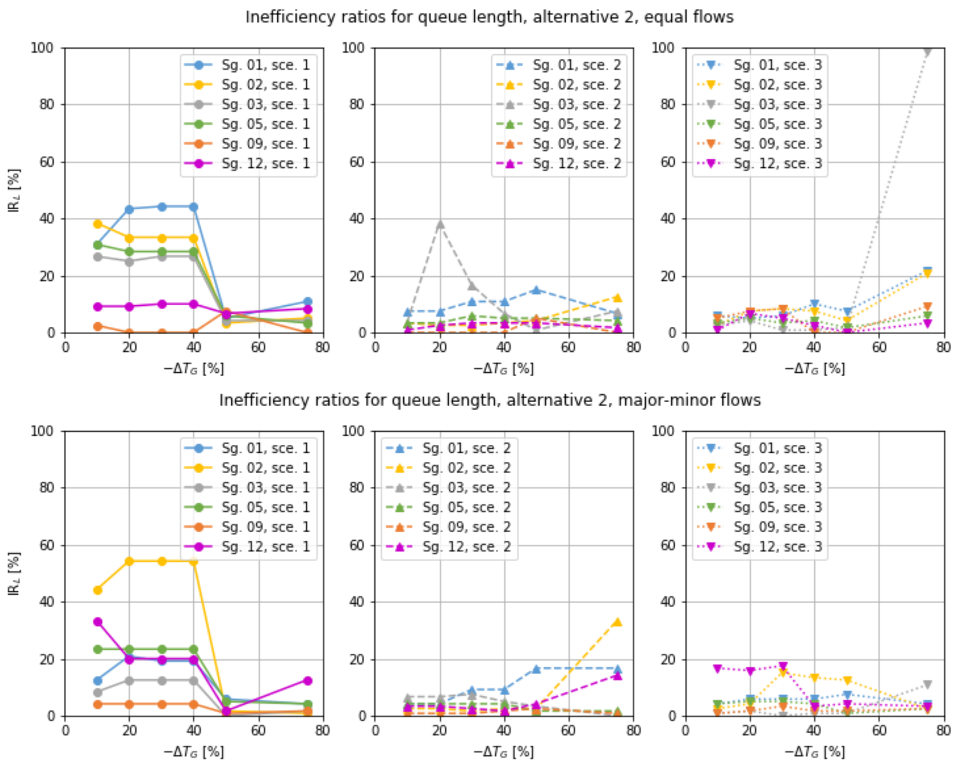

Figure D-6 | Inefficiency ratios for queue length IR_L [%] for alternative (2) incorrect maximum green time T_G [s], per variant of $-\Delta T_G$ [%], per scenario (sce.; as given in Table D-3), per signal group (sg.).

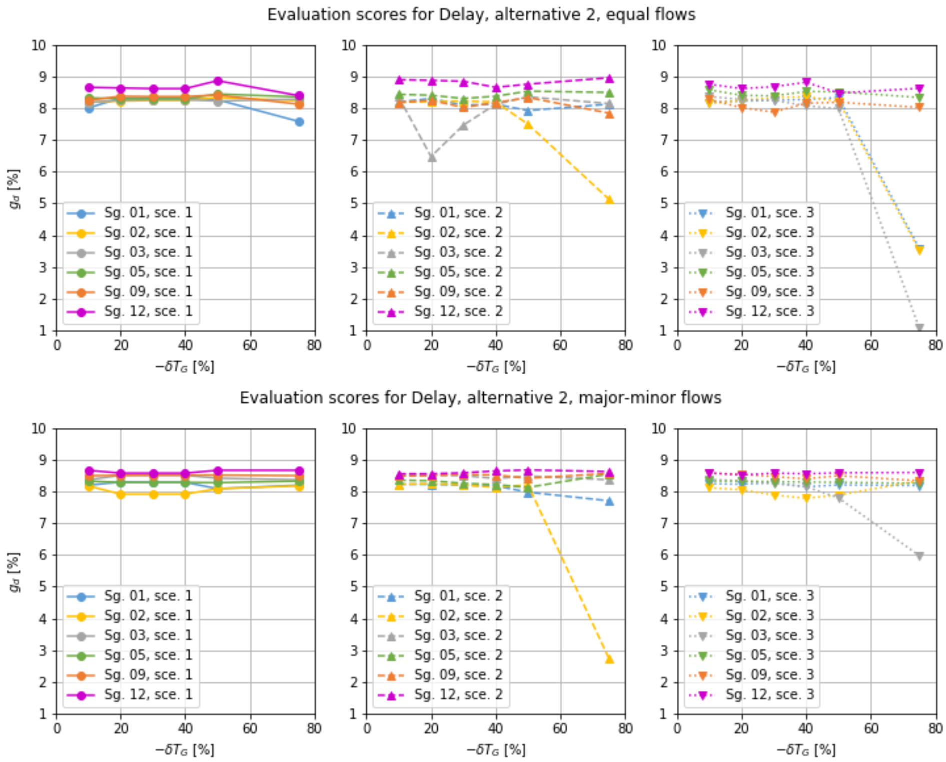

Figure D-7 | Evaluation scores for delay g_d [-] for alternative (2) incorrect maximum green time T_G [s], per variant of $-\Delta T_G$ [%], per scenario (sce.; as given in Table D-3), per signal group (sg.).

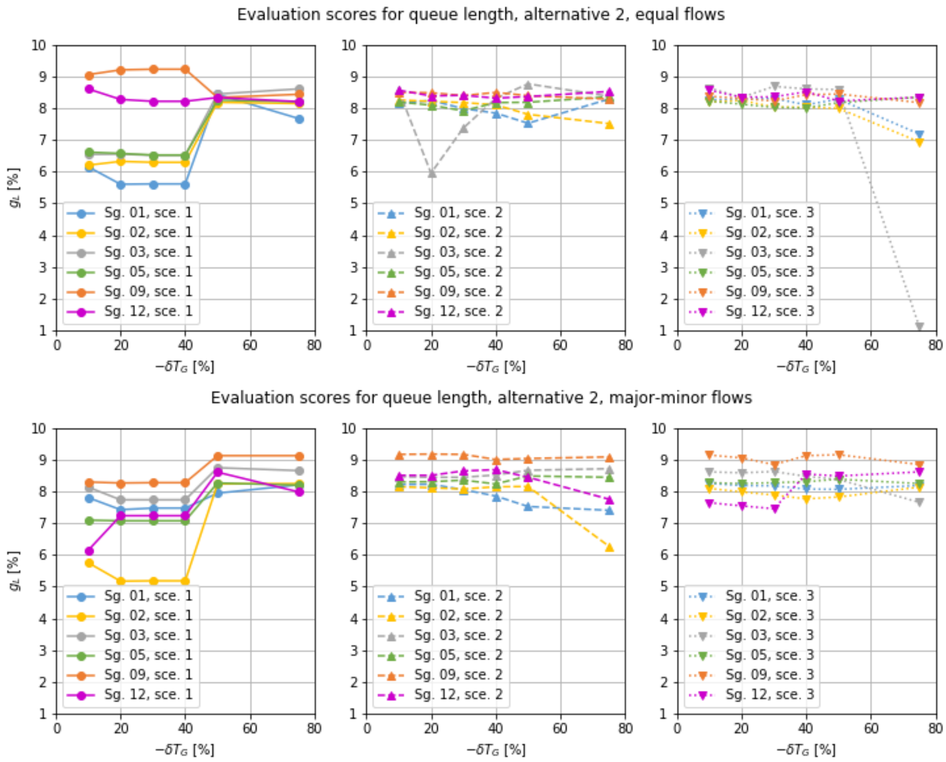

Figure D-8 | Evaluation scores for queue length g_L [-] for alternative (2) incorrect maximum green time T_G [s], per variant of $-\Delta T_G$ [%], per scenario (sce.; as given in Table D-3), per signal group (sg.).

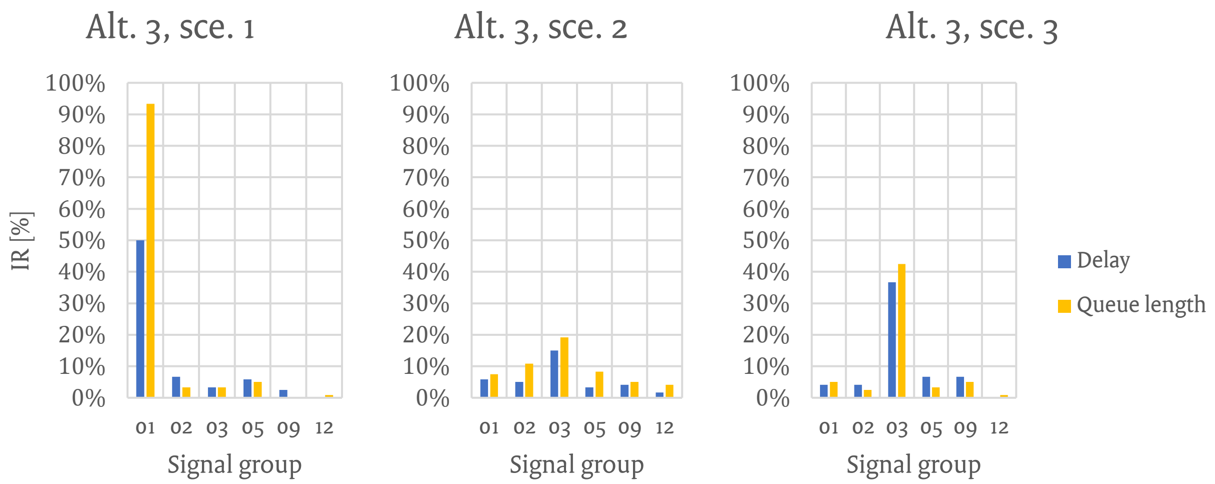

Figure D-9 | Inefficiency ratios IR [%] for delay, and queue length, for alternative (alt.) 3, per scenario (sce.; as given in Table D-3), for the signal groups in the critical conflict group, for equal flows.

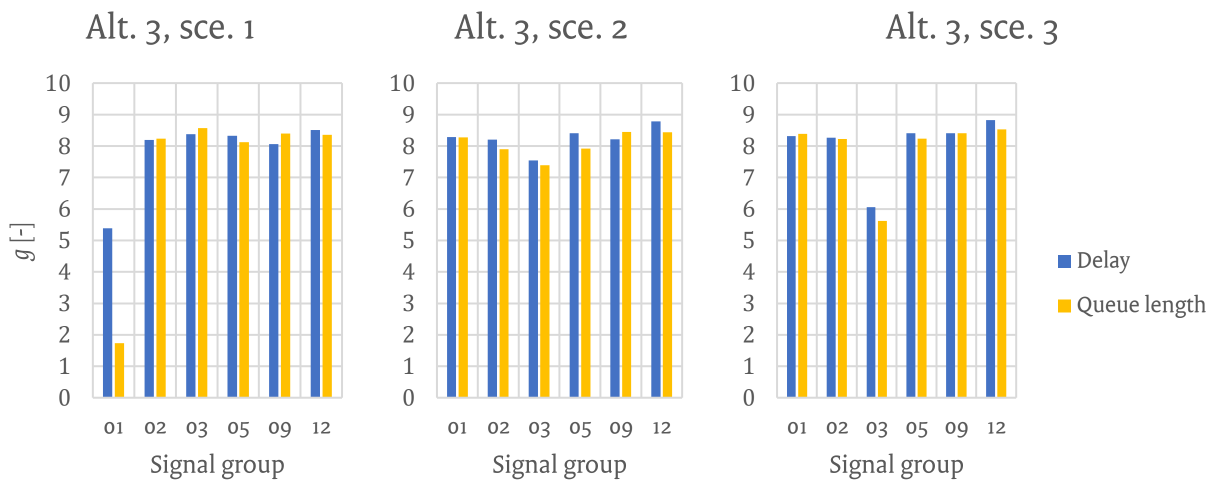

Figure D-10 | Evaluation scores *g* [-] for delay, and queue length, for alternative (alt.) 3, per scenario (sce.; as given in Table D-3), for the signal groups in the critical conflict group, for equal flows.

E. General simulation study set-up

In the thesis report, simulation studies are performed using VISSIM. Because the set-ups of the simulation studies are mainly equal – the main aspects are the same for each of the simulation studies, whereas only a limited number of study-specific alterations apply – the general set-up of the simulation studies is discussed in this appendix.

E.1. Simulation period

E.1.1. Calibration: assessment of accuracy of reference performance models

Each run of the simulation consists of 4200 simulation seconds (70 simulation minutes). In the data collection, and data analysis, only the middle hour is used, thus the first, and the last 300 simulation seconds (5 simulation minutes each) are not included. That way, potential disturbances in the data due to the filling, and emptying of the network are excluded from the data analysis. This means that data is collected, and analysed for a simulation period of 3600 seconds (60 simulation minutes). Also, this enables a data analysis during the period when the signalised intersection is fully functional.

E.1.2. Case study

Each run consists of a total of 10 hours, plus 10 minutes, thereby resulting in a total simulation period of 36600 seconds. Of these 36600 simulation seconds, the first, and last 300 seconds (600 seconds in total) are excluded. That way, potential disturbances in the data due to the filling, and emptying of the network are excluded from the data analysis. This means that data is collected, and analysed for a simulation period of 36000 seconds (600 simulation minutes, thus 10 simulation hours).

This long simulation period is used to fit with the objective of the case study, namely to be able to detect inefficient performance, and assign this to a specific problem. By collecting data over a single period of 10 hours, there is no need to average the results of individual hours, thereby resulting in one consistent data collection period. Moreover, by using a period of 10 hours, more data is collected, and can be used to detect inefficient performance, and assign this to a specific problem.

E.2. Number of runs

The number of runs determines the reliability of the evaluation results. The number of runs is computed using a limited number of pilot runs, and is based on the standard deviation of the data collected in these pilot runs. This results in the following equation given by WSDOT (2014):

$$N \geq z^2_{\frac{\alpha}{2},n-1}\left(1+\frac{\xi^2}{2}\right)\left(\frac{\sigma_s}{\sigma_a}\right)^2 \tag{E-1}$$

In this equation, α represents the reliability [%], N the number of simulation runs [#], σ_s, and σ_a represent respectively the standard deviation of the pilot sample, and the accepted standard deviation [-], ξ equals the normal distribution excess value [-], and $z_{\alpha/2,n-1}$ is the value of student-t distribution for a given reliability α [-], and pilot sample size n [#]. For the normal distribution excess value ξ, it holds that $\xi = 0$ when the average values are used to compute the pilot sample standard deviation, therefore yielding:

$$N \geq z^2_{\frac{\alpha}{2},n-1}\left(\frac{\sigma_s}{\sigma_a}\right)^2 \tag{E-2}$$

This equation could be reduced further, if the accepted standard deviation σ_a would be set equal to the sample standard deviation σ_s, thus $\sigma_a = \sigma_s$, yielding:

$$N \geq z^2_{\frac{\alpha}{2},n-1} \tag{E-3}$$

Now, the number of simulation runs can be computed, given a value for $z_{\alpha/2,n-1}$, and thus a given value for reliability α, and pilot sample size n. However, given the limited time for this research, this process is reversed. Instead, a number of runs N is selected, after which the corresponding value for $z_{\alpha/2,n-1}$, and thus reliability α, is computed, given a pilot sample size n.

However, in the case of the case study as discussed in chapter 5, only one run is simulated in the simulation study. This is done because the objective of this simulation study is to collect data to be able to assign an inefficiency to a certain problem. Therefore, the reliability of the data with respect to the number of runs is considered as irrelevant at this point.

On the other hand, in the case of the simulation study as will be discussed in section 4.4, the reliability of the results does play a role. Therefore, the number of runs for this particular simulation pilot study is computed. This done by selecting a number of runs, and computing the corresponding reliability. Then, a maximum number of 10 runs is selected. Using delay data from 10 pilot runs, a standard deviation of the pilot sample of 0.18 s/pce is found. When this standard deviation equals the accepted standard deviation, it was found that the reliability approaches 99%. Otherwise, with an accepted standard deviation of 0.13 s/pce, the corresponding reliability approaches 95%. Because this is a rather high reliability, 10 simulation runs are performed.

E.3. Data collection and processing

The data is collected using a duration of the analysis period of 300 seconds (5 minutes). The data collection focuses on the multi-variable performance indicators, and their describing variables, as considered in this thesis. Therefore, the following data is collected:

- Delays;
- Queue lengths;
- Detector states to determine phase failures;
- Traffic flow volumes, and arrivals per cycle phase (green, amber, red);
- Green times, amber times, red times, and cycle times;
- Utilised amber time.

This data is then used to compute the other describing variables, and performance indicators, e.g. degree of saturation.

The data processing, and computation of the other describing variables is done using a PYTHON script. Because most of the aforementioned data can only be extracted from VISSIM as raw data, a data processing tool was made to generate useful data. Indeed, phase failures can only be found by combining the raw detector state data, with the raw traffic signal data. The same holds for data on arrivals per cycle phase, and utilised amber time. Moreover, it must be noted that instead of the default raw detector state data in VISSIM, the raw data of data collection points is used. These data collection points were placed on the exact same location as a detector, and are used because of the better accuracy of these data points – data per 0.01 seconds using data collection points, versus data per 1.00 second using raw detector state data.

Lastly, it must be noted that an arrival is defined as the front of a vehicle arriving at the stop line. In relation to arrivals during red, this might cause issues because vehicles that join the queue do not arrive at the stop line during red, but during the consecutive green phase, for instance, according to this definition. Therefore, the arrivals are estimated. Given an average speed of 80 km/h, it is assumed that if at moment t a vehicle passes a given point 133.33 m upstream of the stop line, that vehicles arrives during e.g. green if at moment $t + 6$ the traffic signal is green for that signal group. If the queue length approaches 133.33 m, vehicles that join the queue are slowing down at this point. This is however not included in the estimation of the arrivals. Although this might result in incorrect estimations of the arrivals, it is assumed that this does not have a significant effect, because it is expected that the queue length does not reach 133.33 m. Nonetheless, more accurate predictions of the arrivals, for instance by tracking individual vehicles, is expected to yield better results. Arrivals in the queue are assumed to be vertical queues with respect to arrivals per cycle phase (i.e. green, amber, red). Additionally, the arrivals per cycle phase, although they are random, are converted into (computed) uniform arrivals, whereas the uniform arrival rate differs per cycle phase.

F. Elaborate results of simulation study, and case study

In the main report of the thesis, only a limited amount of data of the simulation study results of sections 4.4, and 5.3 are presented. In this appendix, a detailed overview of these results is presented. The results are presented in tables, and figures with results per signal group, and per traffic flow volume input set.

F.1. Assessment of accuracy of reference performance models results

In section 4.4, the results for uniform, and random arrivals are given as averaged results for right-turning, through-going, and left-turning movements. This appendix presents per signal group, per traffic flow volume input set.

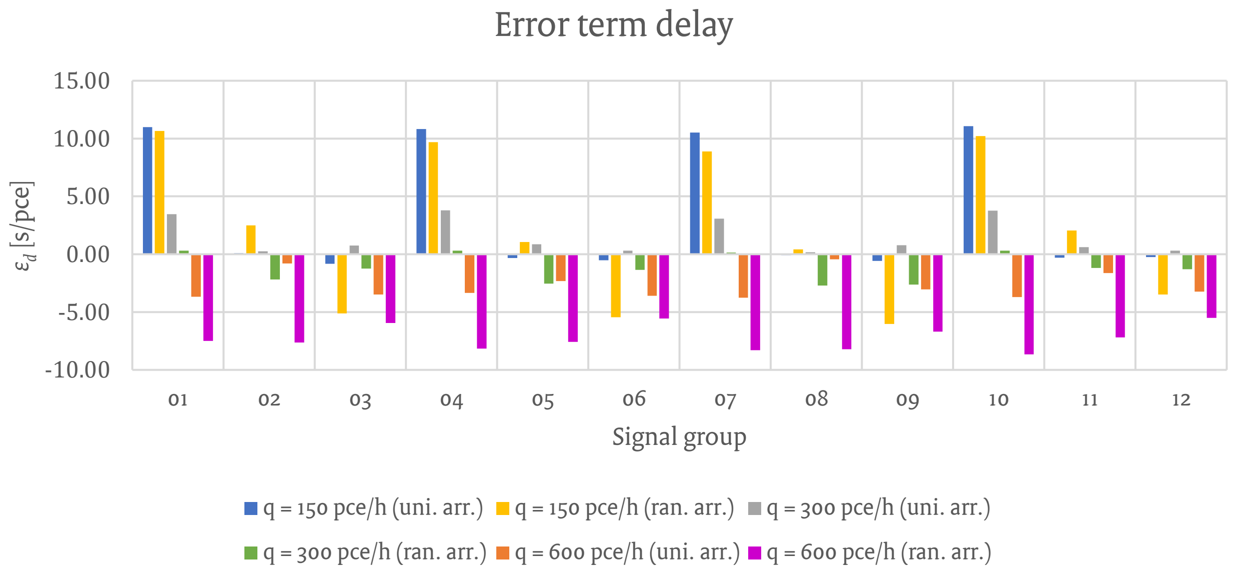

Figure F-1 | Error term for delay ε_d [s/pce], for uniform (uni.) and random (ran.) arrivals (arr.), per traffic flow volume set (q), per signal group.

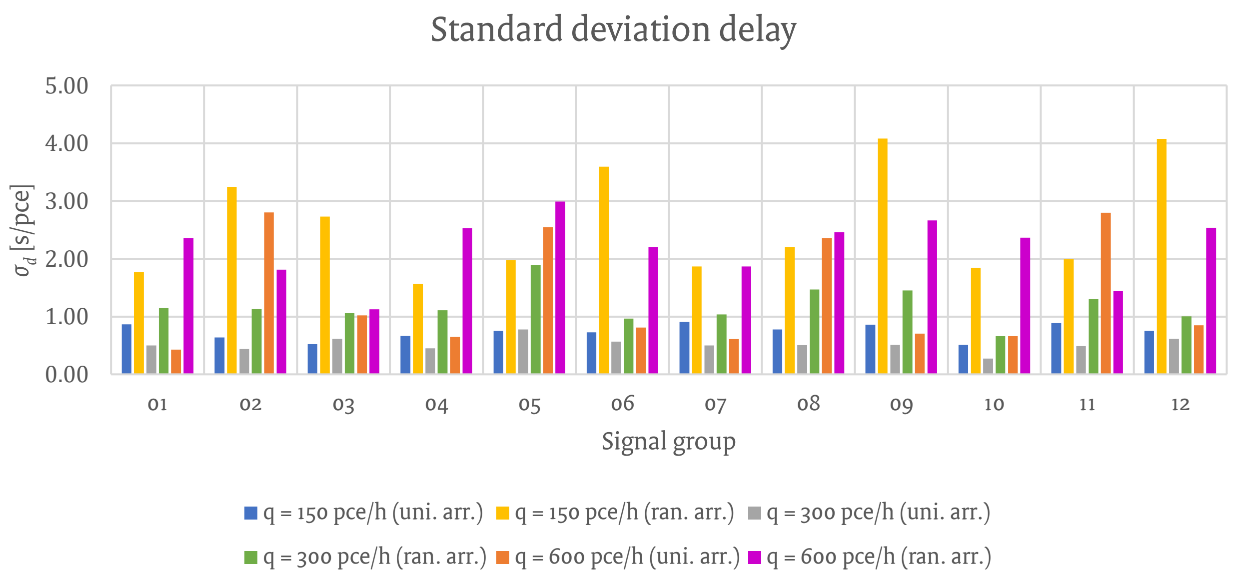

Figure F-2 | Standard deviation for delay σ_d [s/pce] (to define the bandwidth), for uniform (uni.) and random (ran.) arrivals (arr.), per traffic flow volume set (q), per signal group.

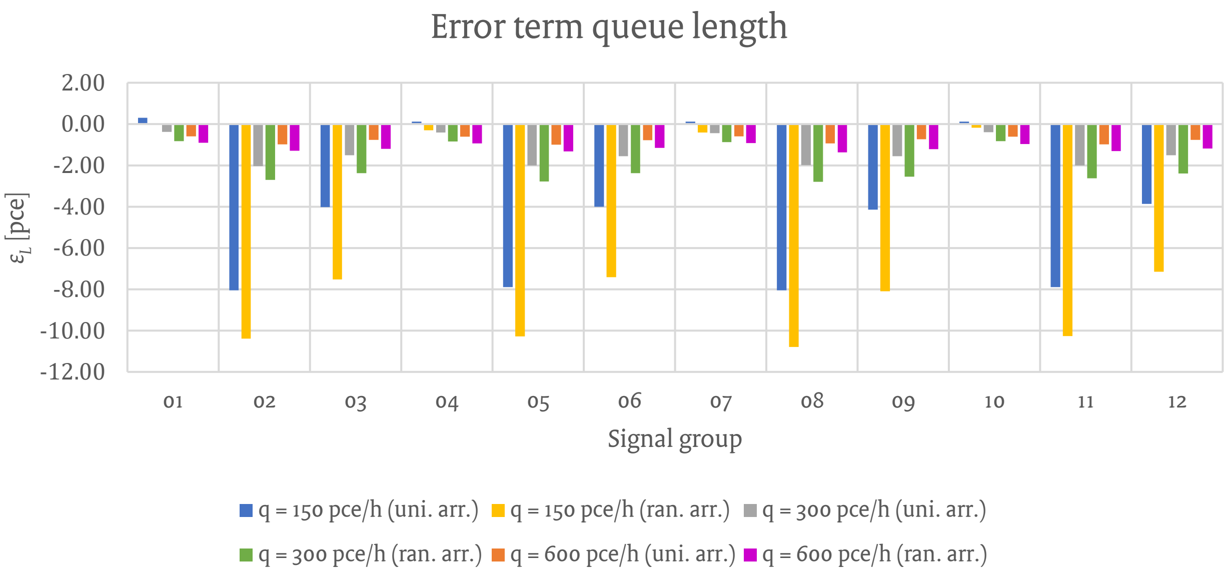

Figure F-3 | Error term for queue length ε_L [pce], for uniform (uni.) and random (ran.) arrivals (arr.), per traffic flow volume set (q), per signal group.

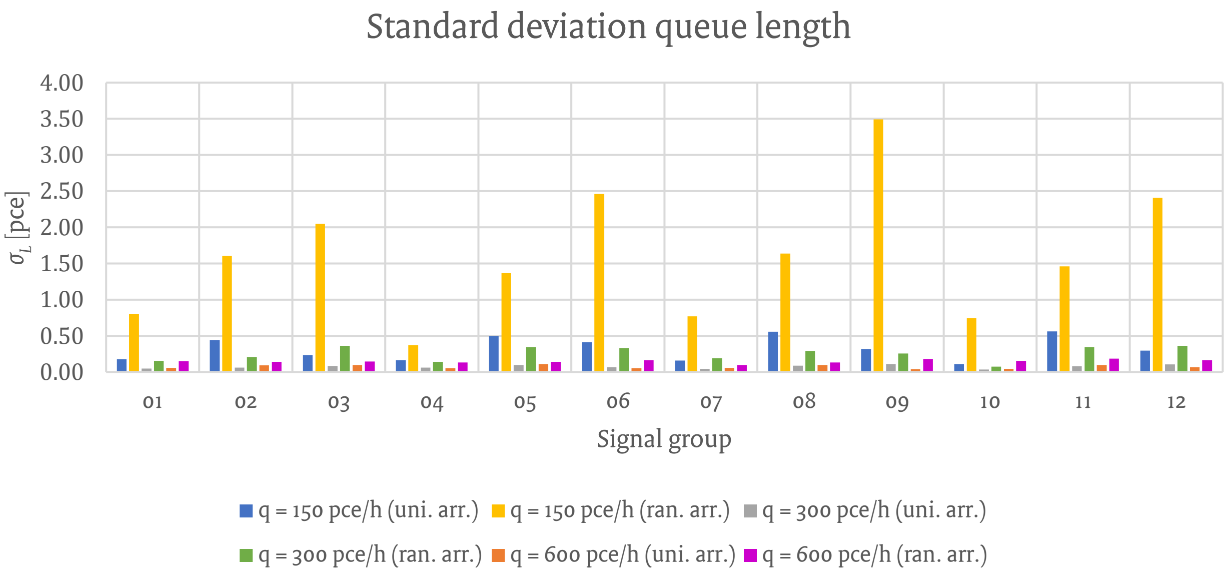

Figure F-4 | Standard deviation for queue length σ_L [pce] (to define the bandwidth), for uniform (uni.) and random (ran.) arrivals (arr.), per traffic flow volume set (q), per signal group.

F.2. Case study results

In section 5.3, only a few examples of the case study simulation results are presented. This appendix shows the more elaborate results, per signal group. Also, all the inefficiency detector outputs (inefficiency ratios, evaluation scores, and performance statistics, as defined in section 3.2.3.1) are presented. The case study results are presented per alternative. In short, the results as listed in Table F-1 are presented. The results are given for both traffic flow volume input sets (as defined in section 5.2.2.1), if applicable according to the tested alternatives. For reference purposes, Table F-2 lists the tested alternatives, with corresponding traffic flow volume input set, implemented problem(s), and signal group(s) with the considered problem(s), as known at the start of the case study.

Table F-1 | Presented results for the case study per alternative (alt.).

Alt.	Error term	Standard deviation (bandwidth)	Inefficiency ratio[1]	Evaluation scores[2]	Performance statistics[1]	Degree of saturation
0[3]	▪	▪	▪	▪		▪
1			▪	▪	▪	▪
2			▪	▪	▪	▪
3			▪	▪	▪	▪
4			▪	▪	▪	▪
5			▪	▪	▪	▪
6			▪	▪	▪	▪
7[4]			▪	▪	▪	▪
8[4]			▪	▪	▪	▪
9[4]			▪	▪	▪	▪

Table F-2 | List of alternatives (alt.), with corresponding traffic flow volume input set, implemented problem(s), and signal group(s) with the considered problem(s), as known at the start of the case study.

Alt.	Traffic flow volume input set	Problem(s)	Signal group(s) with problem(s)
0[3]	Equal flows and major-minor flows	N/A	N/A
1	Equal flows	Alternative realisations	02 and 03
2	Major-minor flows	Maximum green time	04 and 06
3	Equal flows	Gap time	02 and 07
4	Major-minor flows	Alternative realisations Maximum green time	04 10
5	Equal flows	Maximum green time Gap time	05 09
6	Major-minor flows	Alternative realisations Gap time	07 11
7[4]	Equal flows and major-minor flows	Unknown	Unknown
8[4]	Equal flows and major-minor flows	Unknown	Unknown
9[4]	Equal flows and major-minor flows	Unknown	Unknown

[1] Only for delay, queue length, and phase failures.
[2] Only for delay, and queue length.
[3] Base case alternative.
[4] Blind alternative.

F.2.1. Base case alternative: calibration

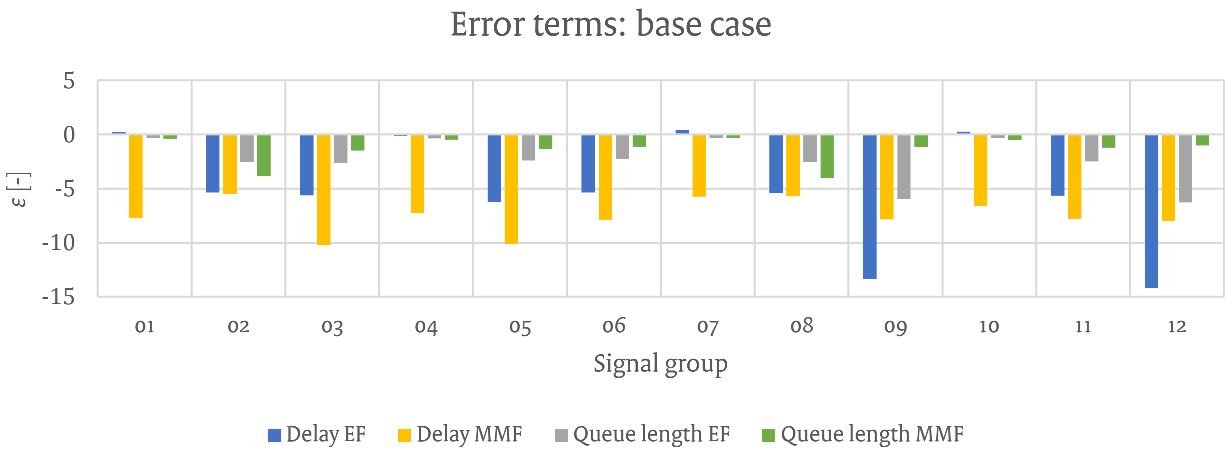

Figure F-5 | Error terms for the base case alternative, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

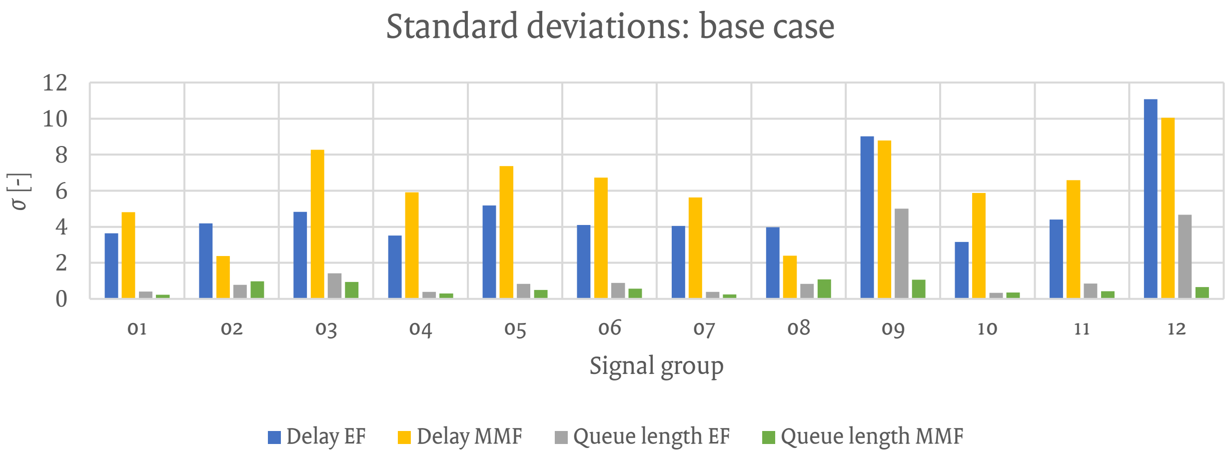

Figure F-6 | Standard deviations for the base case alternative, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

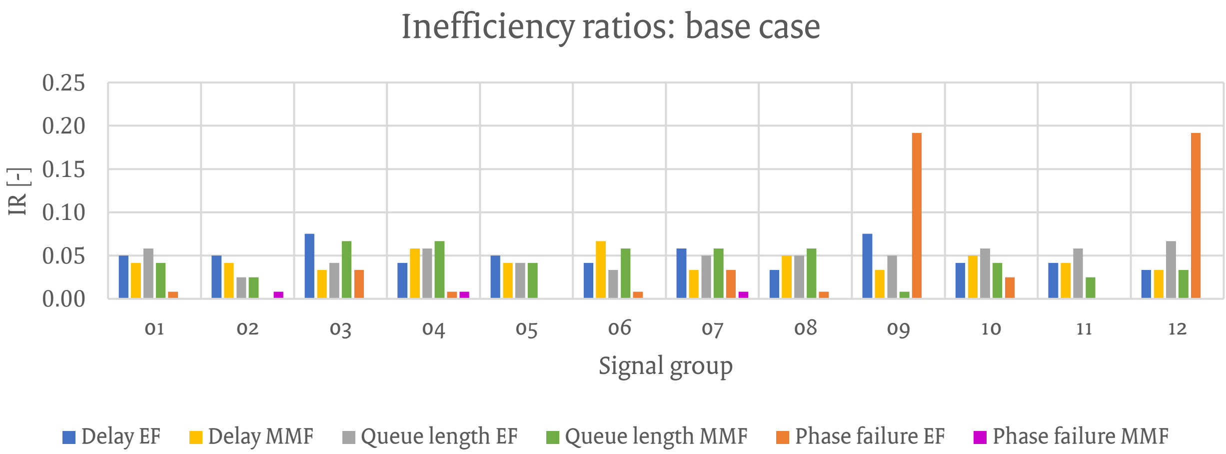

Figure F-7 | Inefficiency ratios for the base case alternative, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

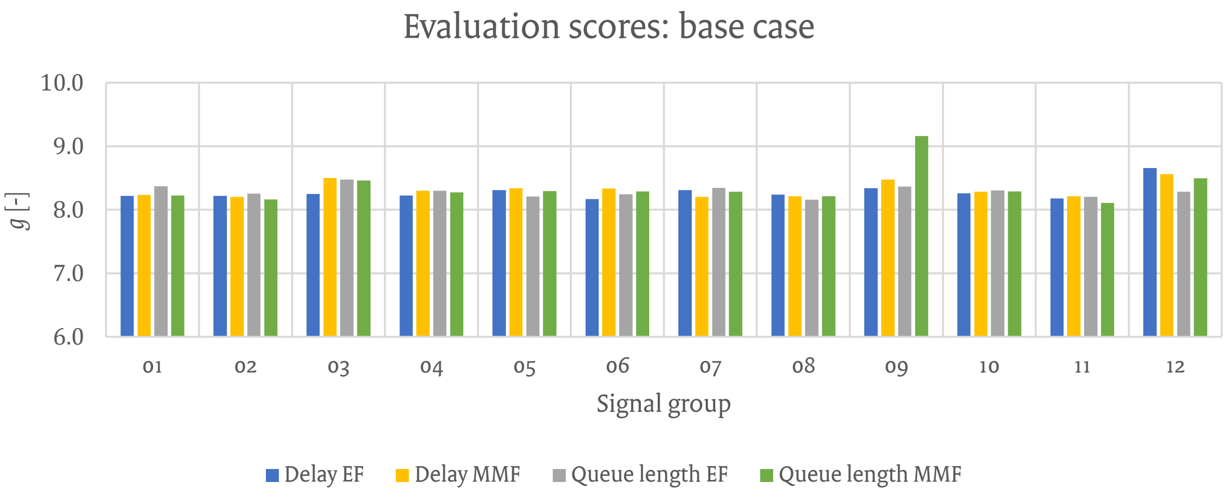

Figure F-8 | Evaluation scores for the base case alternative, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

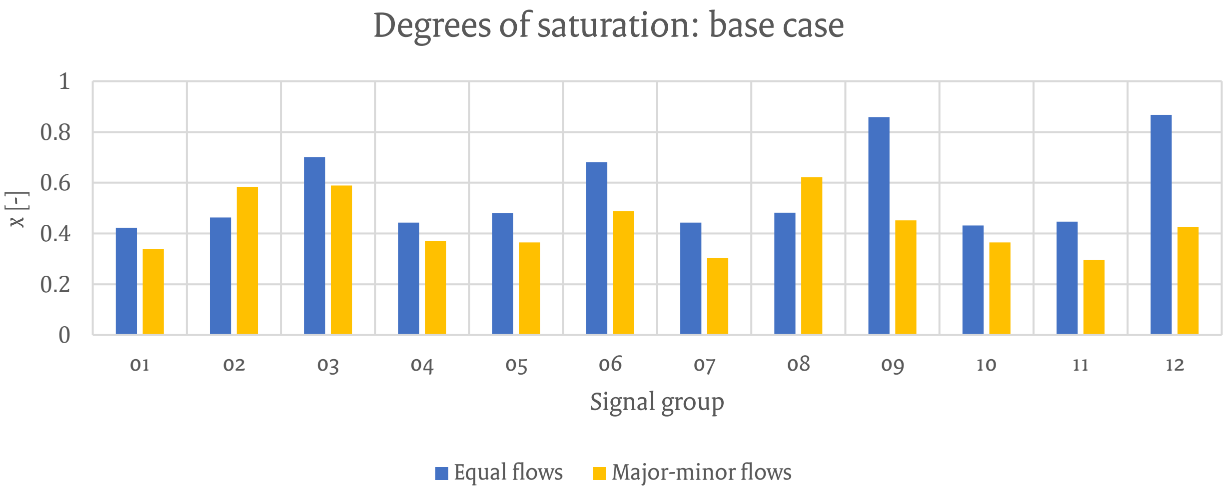

Figure F-9 | Degrees of saturation for the base case alternative.

F.2.2. Alternative 1

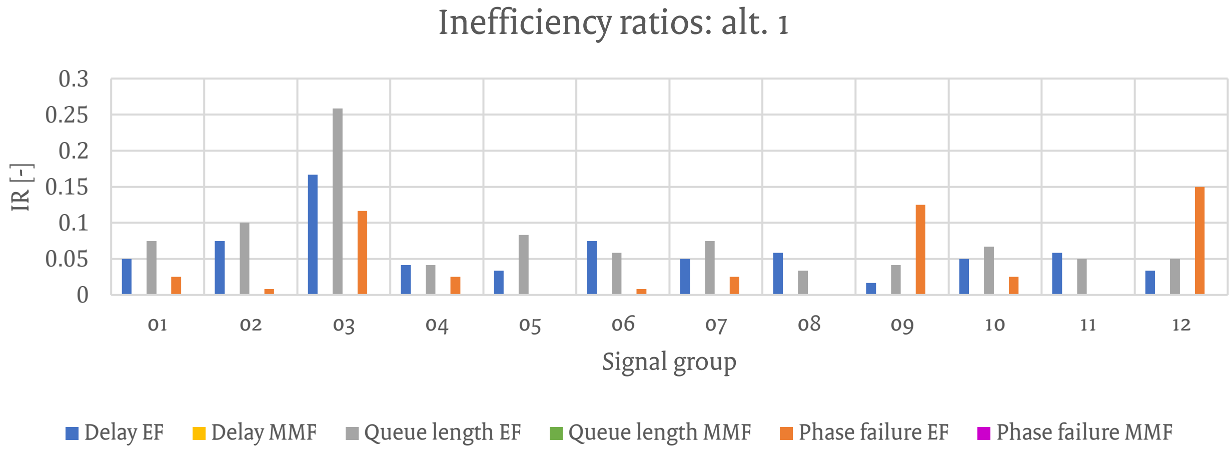

Figure F-10 | Inefficiency ratios for alternative (alt.) 1, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

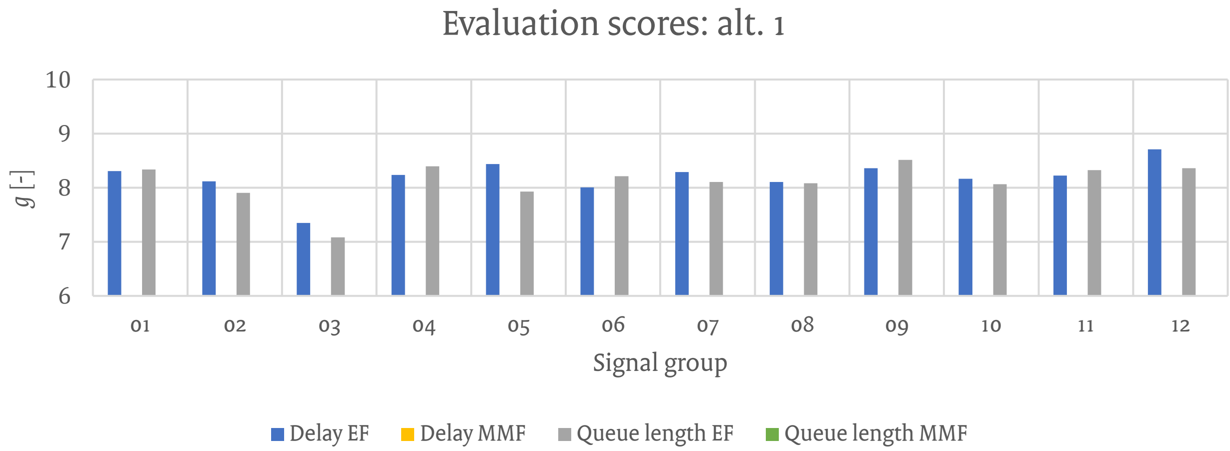

Figure F-11 | Evaluation scores for alternative (alt.) 1, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

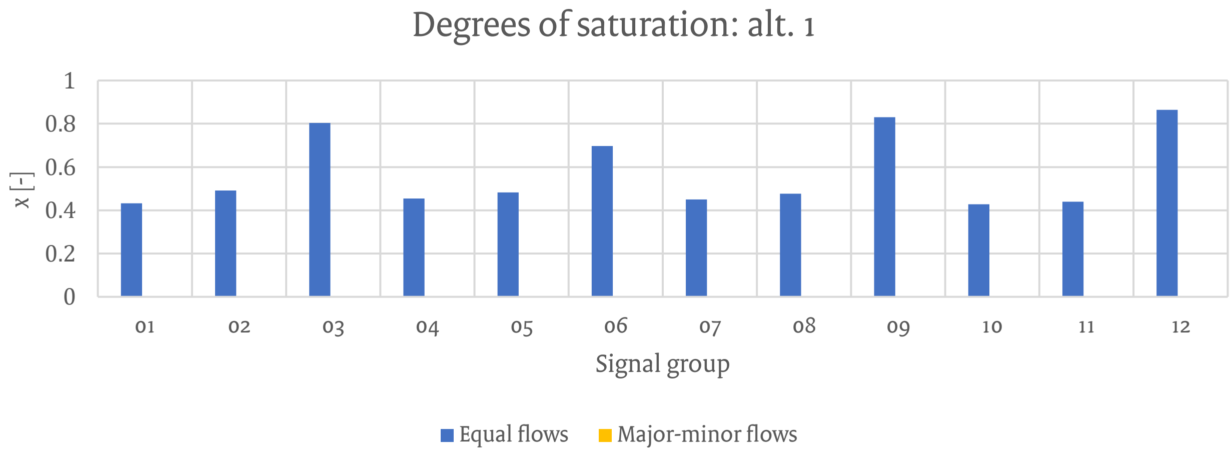

Figure F-12 | Degrees of saturation for alternative (alt.) 1, per traffic flow volume input set.

Table F-3 | Performance statistics (averages) for alternative (alt.) 1, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	-0.14	N/A	-0.02	N/A	0.03	N/A
02	-0.49	N/A	-0.41	N/A	0.01	N/A
03	-3.67	N/A	-1.97	N/A	0.01	N/A
04	0.08	N/A	-0.01	N/A	0.03	N/A
05	0.36	N/A	-0.05	N/A	0.00	N/A
06	-0.28	N/A	-0.07	N/A	0.01	N/A
07	-0.45	N/A	-0.05	N/A	0.03	N/A
08	-0.17	N/A	0.01	N/A	0.00	N/A
09	2.51	N/A	1.10	N/A	0.01	N/A
10	0.54	N/A	0.06	N/A	0.03	N/A
11	0.21	N/A	0.04	N/A	0.00	N/A
12	0.26	N/A	0.28	N/A	0.00	N/A

Table F-4 | Performance statistics (standard deviations) for alternative (alt.) 1, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	3.52	N/A	0.41	N/A	0.16	N/A
02	4.41	N/A	0.95	N/A	0.09	N/A
03	7.76	N/A	3.96	N/A	0.09	N/A
04	3.40	N/A	0.36	N/A	0.16	N/A
05	4.73	N/A	0.96	N/A	0.00	N/A
06	4.63	N/A	0.93	N/A	0.09	N/A
07	4.05	N/A	0.46	N/A	0.16	N/A
08	4.25	N/A	0.85	N/A	0.00	N/A
09	7.92	N/A	4.21	N/A	0.09	N/A
10	3.28	N/A	0.34	N/A	0.16	N/A
11	4.48	N/A	0.80	N/A	0.00	N/A
12	11.12	N/A	4.47	N/A	0.00	N/A

F.2.3. Alternative 2

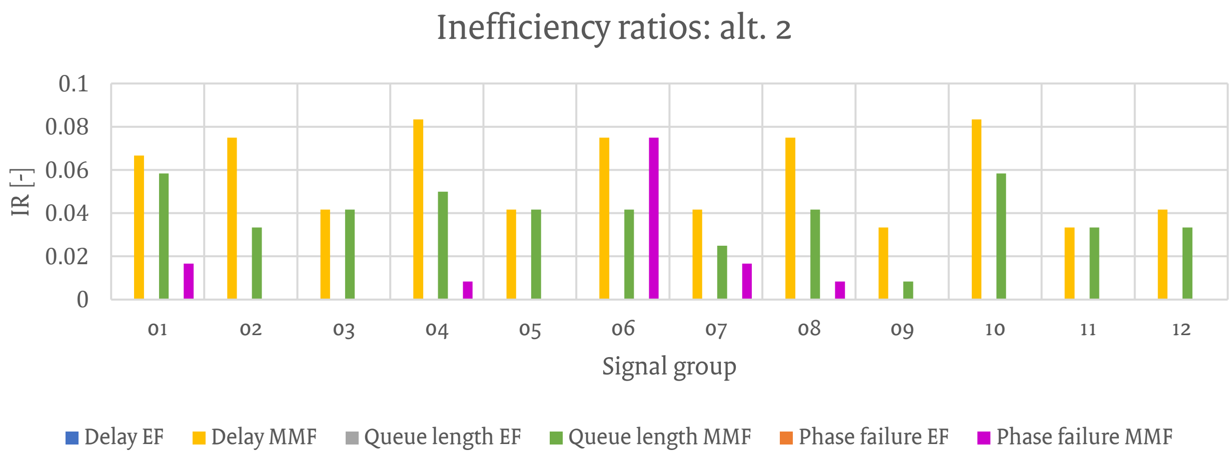

Figure F-13 | Inefficiency ratios for alternative (alt.) 2, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

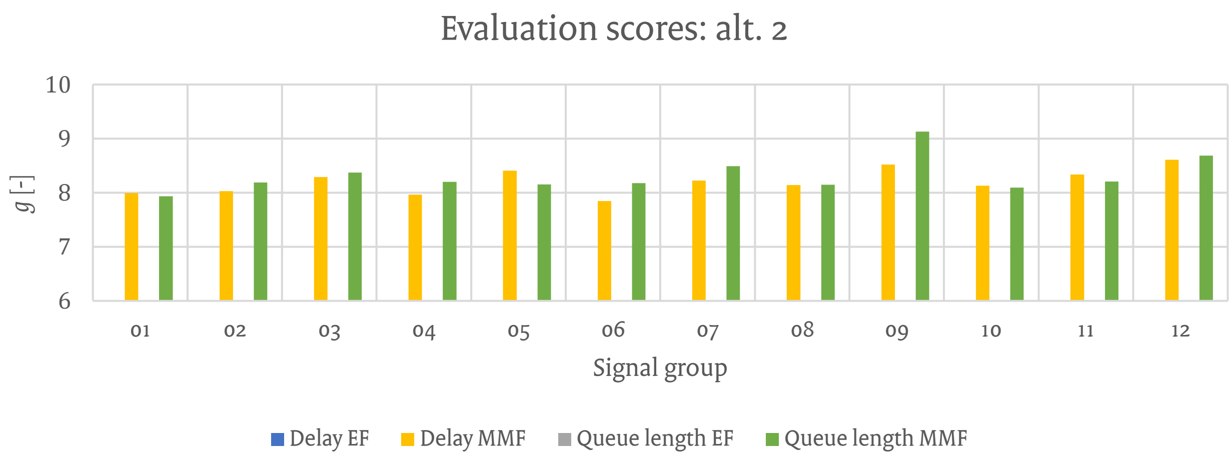

Figure F-14 | Evaluation scores for alternative (alt.) 2, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

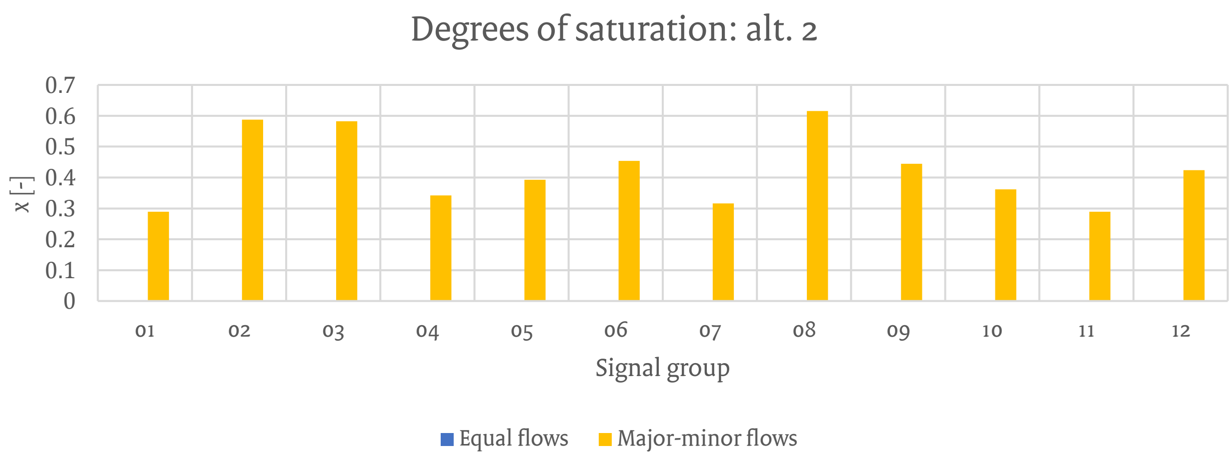

Figure F-15 | Degrees of saturation for alternative (alt.) 2, per traffic flow volume input set.

Table F-5 | Performance statistics (averages) for alternative (alt.) 2, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	1.34	N/A	0.04	N/A	0.02
02	N/A	0.32	N/A	0.04	N/A	0.00
03	N/A	0.52	N/A	0.05	N/A	0.00
04	N/A	1.83	N/A	0.07	N/A	0.01
05	N/A	0.07	N/A	0.01	N/A	0.00
06	N/A	3.24	N/A	0.15	N/A	0.08
07	N/A	-0.44	N/A	0.01	N/A	0.02
08	N/A	0.20	N/A	0.18	N/A	0.01
09	N/A	-0.56	N/A	-0.01	N/A	0.00
10	N/A	-0.47	N/A	-0.02	N/A	0.00
11	N/A	0.16	N/A	0.06	N/A	0.00
12	N/A	1.08	N/A	-0.01	N/A	0.00

Table F-6 | Performance statistics (standard deviations) for alternative (alt.) 2, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	5.13	N/A	0.25	N/A	0.13
02	N/A	2.74	N/A	0.95	N/A	0.00
03	N/A	8.87	N/A	0.91	N/A	0.00
04	N/A	6.56	N/A	0.31	N/A	0.09
05	N/A	7.22	N/A	0.52	N/A	0.00
06	N/A	7.28	N/A	0.57	N/A	0.26
07	N/A	5.71	N/A	0.22	N/A	0.13
08	N/A	2.54	N/A	1.06	N/A	0.09
09	N/A	8.68	N/A	1.09	N/A	0.00
10	N/A	6.44	N/A	0.40	N/A	0.00
11	N/A	6.19	N/A	0.40	N/A	0.00
12	N/A	9.03	N/A	0.54	N/A	0.00

F.2.4. Alternative 3

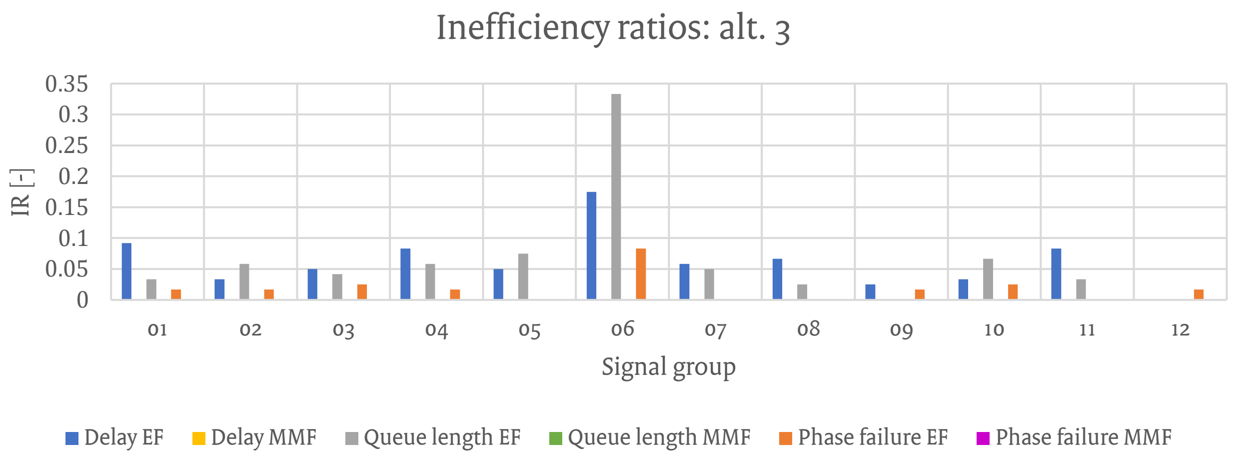

Figure F-16 | Inefficiency ratios for alternative (alt.) 3, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

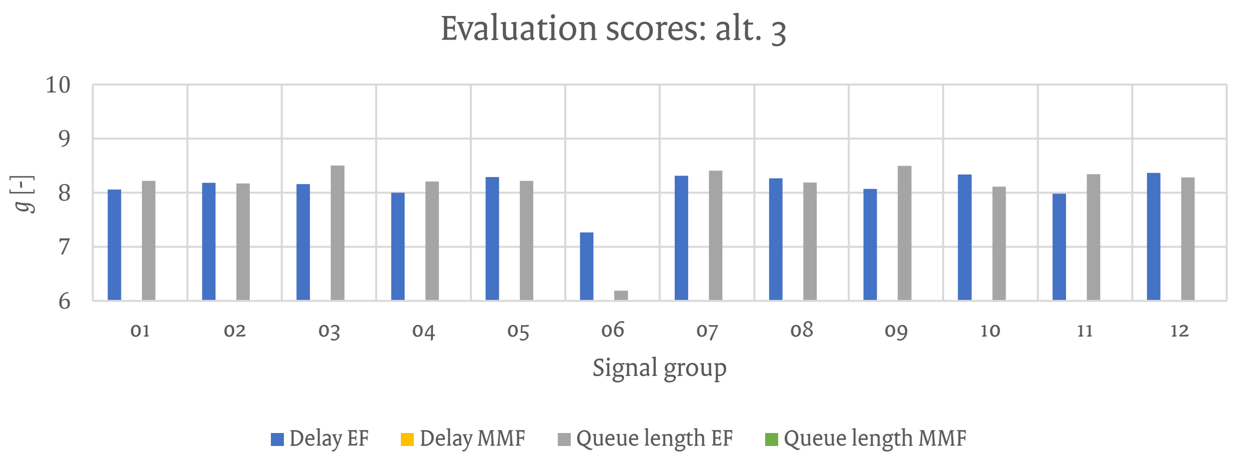

Figure F-17 | Evaluation scores for alternative (alt.) 3, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

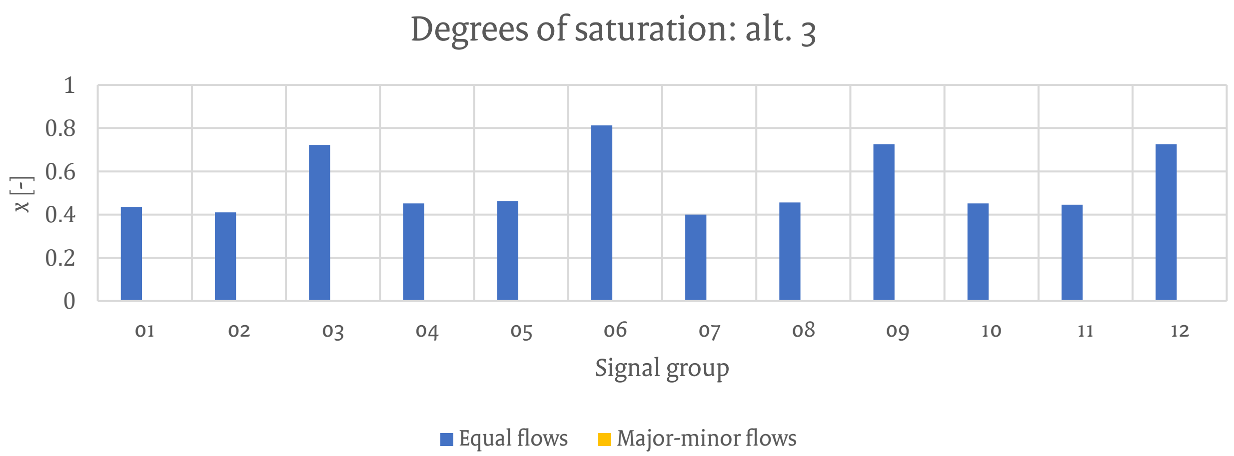

Figure F-18 | Degrees of saturation for alternative (alt.) 3, per traffic flow volume input set.

Table F-7 | Performance statistics (averages) for alternative (alt.) 3, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	0.22	N/A	0.00	N/A	0.02	N/A
02	1.01	N/A	0.11	N/A	0.02	N/A
03	-0.63	N/A	-0.22	N/A	0.02	N/A
04	0.20	N/A	0.00	N/A	0.02	N/A
05	0.00	N/A	-0.07	N/A	0.00	N/A
06	-3.42	N/A	-2.06	N/A	0.00	N/A
07	0.11	N/A	0.02	N/A	0.00	N/A
08	0.43	N/A	0.18	N/A	0.00	N/A
09	6.85	N/A	3.10	N/A	0.00	N/A
10	-0.13	N/A	-0.05	N/A	0.03	N/A
11	-0.05	N/A	-0.03	N/A	0.00	N/A
12	7.79	N/A	3.39	N/A	0.01	N/A

Table F-8 | Performance statistics (standard deviations) for alternative (alt.) 3, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	4.12	N/A	0.46	N/A	0.13	N/A
02	4.15	N/A	0.85	N/A	0.13	N/A
03	4.80	N/A	1.51	N/A	0.13	N/A
04	3.93	N/A	0.41	N/A	0.13	N/A
05	5.19	N/A	0.83	N/A	0.00	N/A
06	6.05	N/A	3.17	N/A	0.00	N/A
07	3.94	N/A	0.37	N/A	0.00	N/A
08	4.13	N/A	0.81	N/A	0.00	N/A
09	5.54	N/A	1.70	N/A	0.00	N/A
10	3.18	N/A	0.40	N/A	0.16	N/A
11	5.01	N/A	0.78	N/A	0.00	N/A
12	5.13	N/A	1.59	N/A	0.09	N/A

F.2.5. Alternative 4

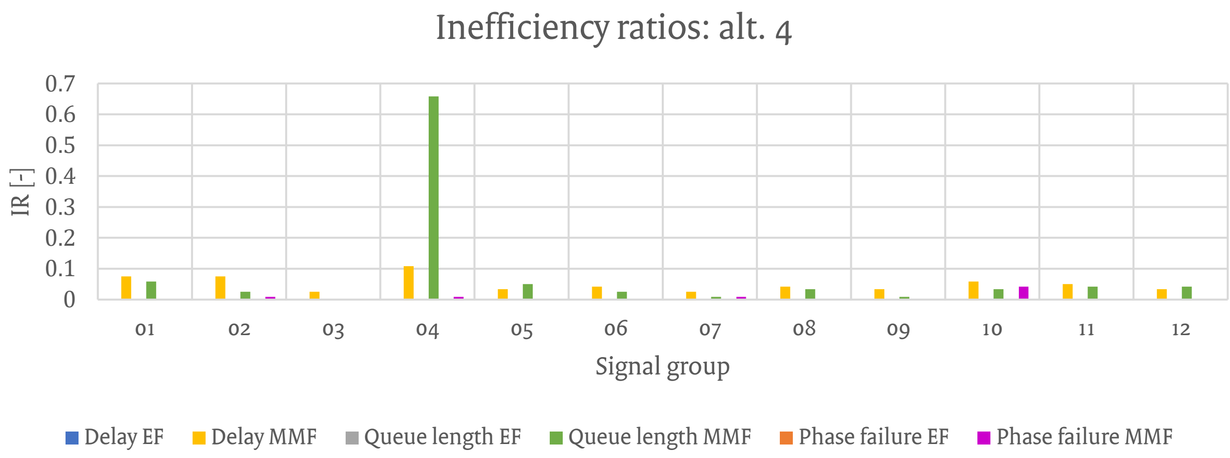

Figure F-19 | Inefficiency ratios for alternative (alt.) 4, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

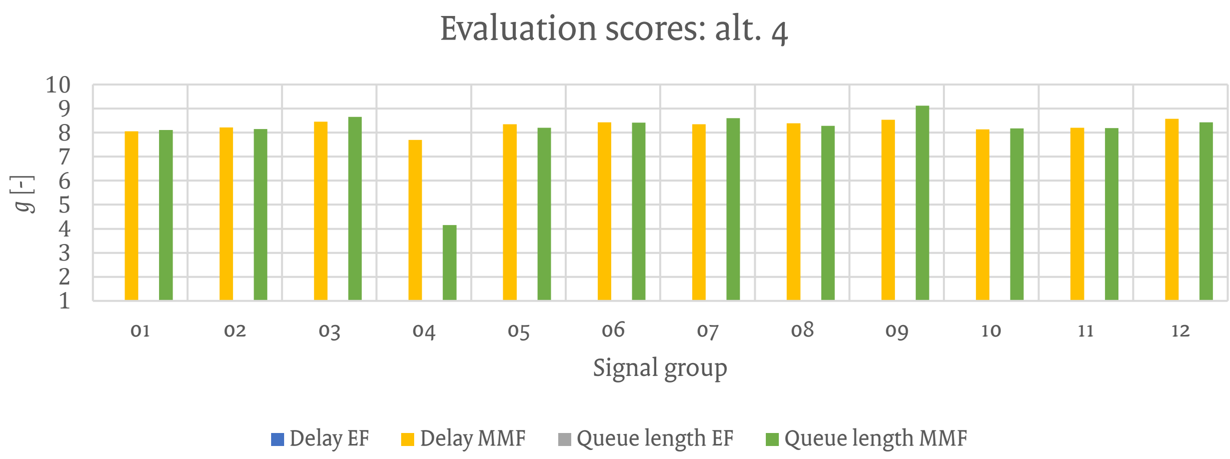

Figure F-20 | Evaluation scores for alternative (alt.) 4, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

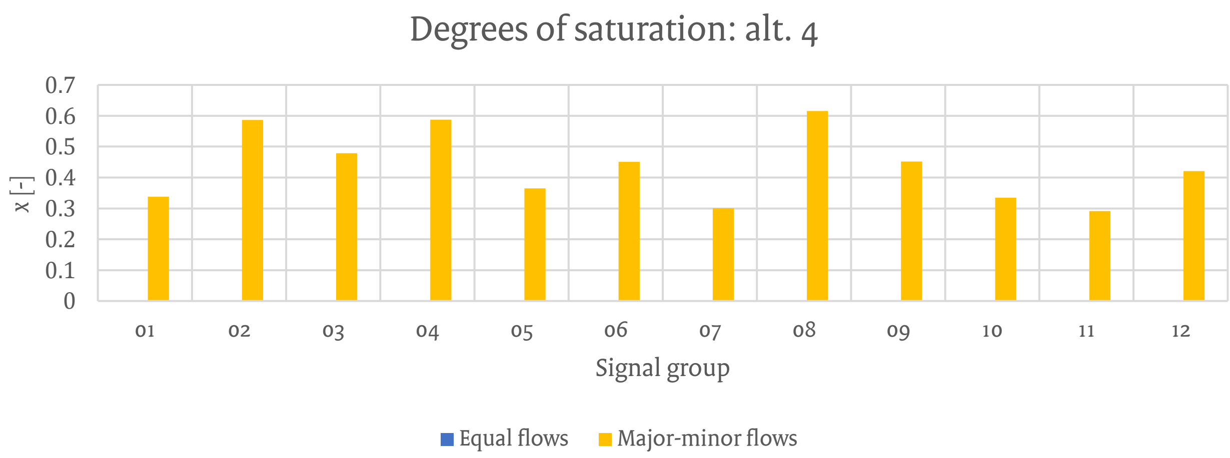

Figure F-21 | Degrees of saturation for alternative (alt.) 4, per traffic flow volume input set.

Table F-9 | Performance statistics (averages) for alternative (alt.) 4, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	-0.26	N/A	0.00	N/A	0.00
02	N/A	0.19	N/A	0.01	N/A	0.01
03	N/A	0.86	N/A	0.30	N/A	0.00
04	N/A	-1.75	N/A	-0.99	N/A	0.01
05	N/A	0.57	N/A	0.02	N/A	0.00
06	N/A	0.18	N/A	0.04	N/A	0.00
07	N/A	0.08	N/A	0.03	N/A	0.01
08	N/A	-0.08	N/A	0.09	N/A	0.00
09	N/A	0.15	N/A	0.00	N/A	0.00
10	N/A	0.63	N/A	0.07	N/A	0.04
11	N/A	-1.07	N/A	0.06	N/A	0.00
12	N/A	0.77	N/A	0.04	N/A	0.00

Table F-10 | Performance statistics (standard deviations) for alternative (alt.) 4, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	5.54	N/A	0.25	N/A	0.00
02	N/A	2.53	N/A	0.98	N/A	0.09
03	N/A	7.91	N/A	0.61	N/A	0.00
04	N/A	7.30	N/A	0.85	N/A	0.09
05	N/A	7.51	N/A	0.50	N/A	0.00
06	N/A	6.12	N/A	0.51	N/A	0.00
07	N/A	5.81	N/A	0.19	N/A	0.09
08	N/A	2.19	N/A	1.02	N/A	0.00
09	N/A	8.35	N/A	1.07	N/A	0.00
10	N/A	6.23	N/A	0.38	N/A	0.20
11	N/A	6.76	N/A	0.42	N/A	0.00
12	N/A	9.93	N/A	0.66	N/A	0.00

F.2.6. Alternative 5

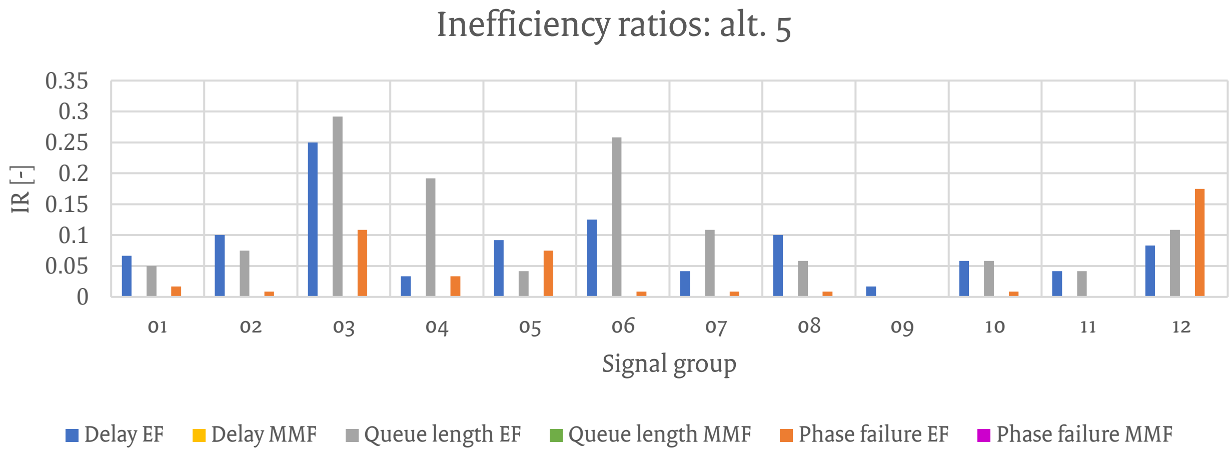

Figure F-22 | Inefficiency ratios for alternative (alt.) 5, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

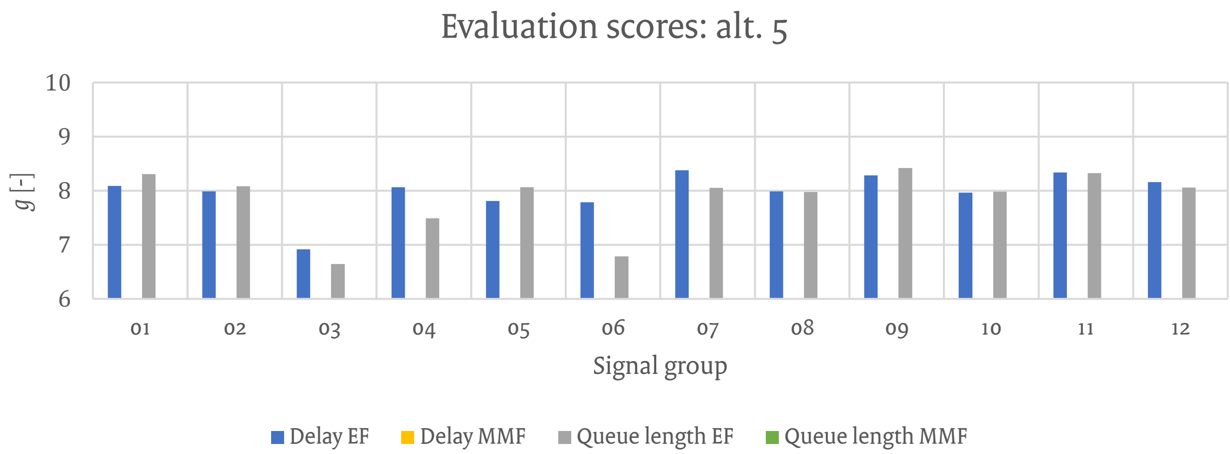

Figure F-23 | Evaluation scores for alternative (alt.) 5, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

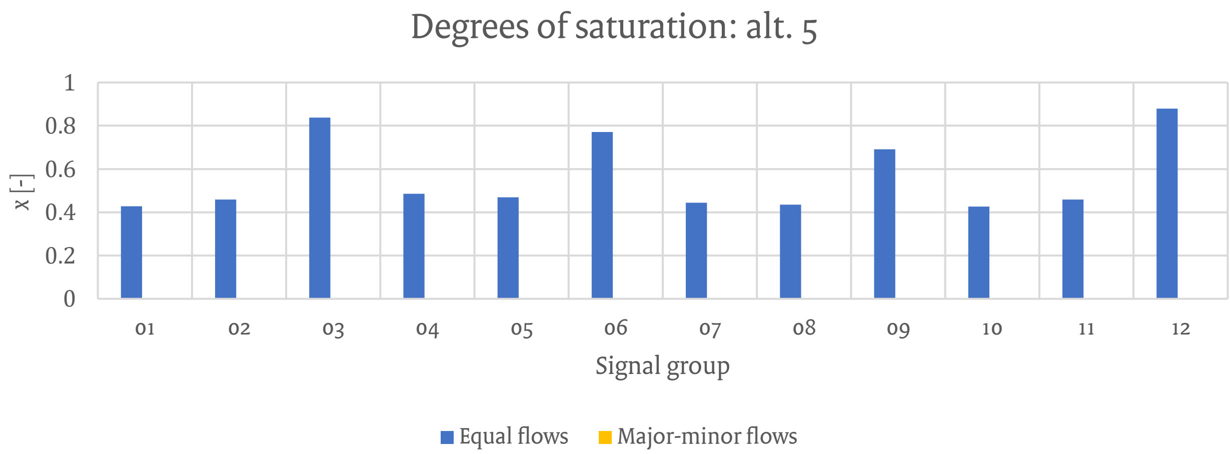

Figure F-24 | Degrees of saturation for alternative (alt.) 5, per traffic flow volume input set.

Table F-11 | Performance statistics (averages) for alternative (alt.) 5, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	-0.08	N/A	-0.01	N/A	0.02	N/A
02	0.47	N/A	-0.05	N/A	0.01	N/A
03	-6.13	N/A	-2.96	N/A	0.01	N/A
04	-0.10	N/A	-0.24	N/A	0.03	N/A
05	3.66	N/A	0.12	N/A	0.08	N/A
06	-2.61	N/A	-1.14	N/A	0.00	N/A
07	-0.31	N/A	-0.02	N/A	0.01	N/A
08	0.23	N/A	0.14	N/A	0.01	N/A
09	6.55	N/A	3.52	N/A	0.00	N/A
10	0.75	N/A	0.11	N/A	0.01	N/A
11	-0.53	N/A	-0.10	N/A	0.00	N/A
12	-2.22	N/A	-0.83	N/A	0.00	N/A

Table F-12 | Performance statistics (standard deviations) for alternative (alt.) 5, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	3.86	N/A	0.39	N/A	0.13	N/A
02	4.87	N/A	0.86	N/A	0.09	N/A
03	11.01	N/A	5.76	N/A	0.09	N/A
04	3.75	N/A	0.53	N/A	0.18	N/A
05	5.40	N/A	0.94	N/A	0.26	N/A
06	4.62	N/A	2.02	N/A	0.00	N/A
07	4.18	N/A	0.45	N/A	0.09	N/A
08	4.68	N/A	0.89	N/A	0.09	N/A
09	4.47	N/A	0.95	N/A	0.00	N/A
10	3.44	N/A	0.35	N/A	0.09	N/A
11	4.12	N/A	0.80	N/A	0.00	N/A
12	17.31	N/A	6.76	N/A	0.00	N/A

F.2.7. Alternative 6

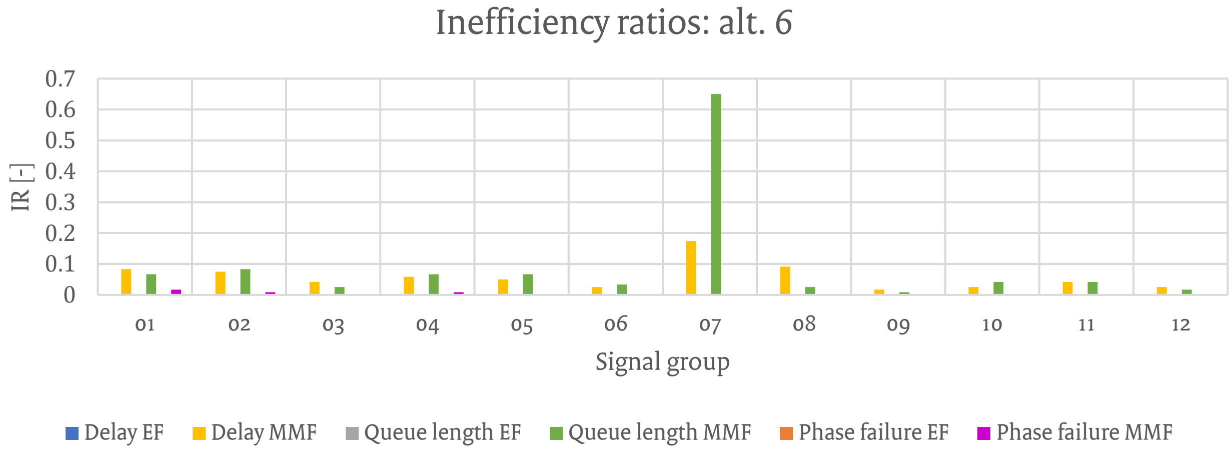

Figure F-25 | Inefficiency ratios for alternative (alt.) 6, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

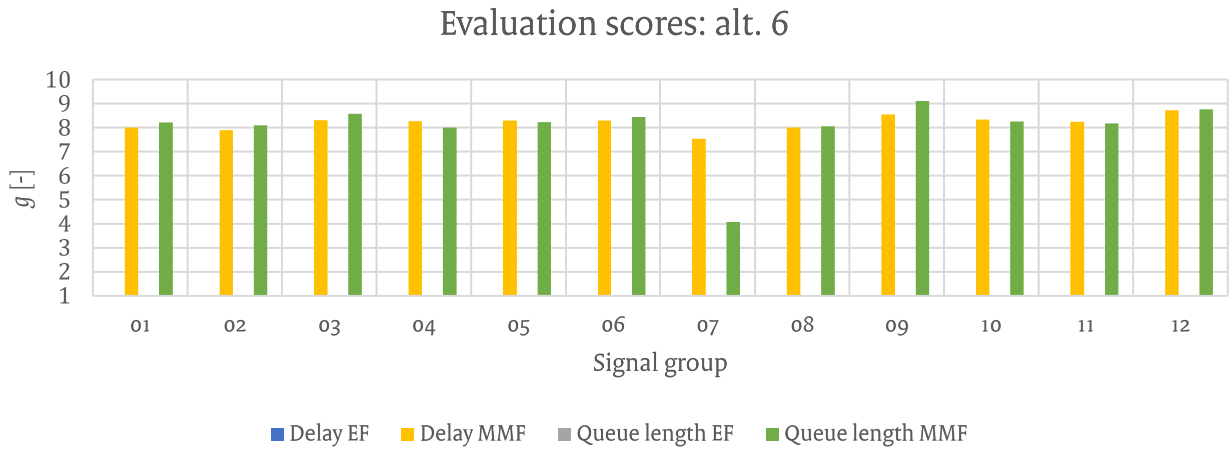

Figure F-26 | Evaluation scores for alternative (alt.) 6, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

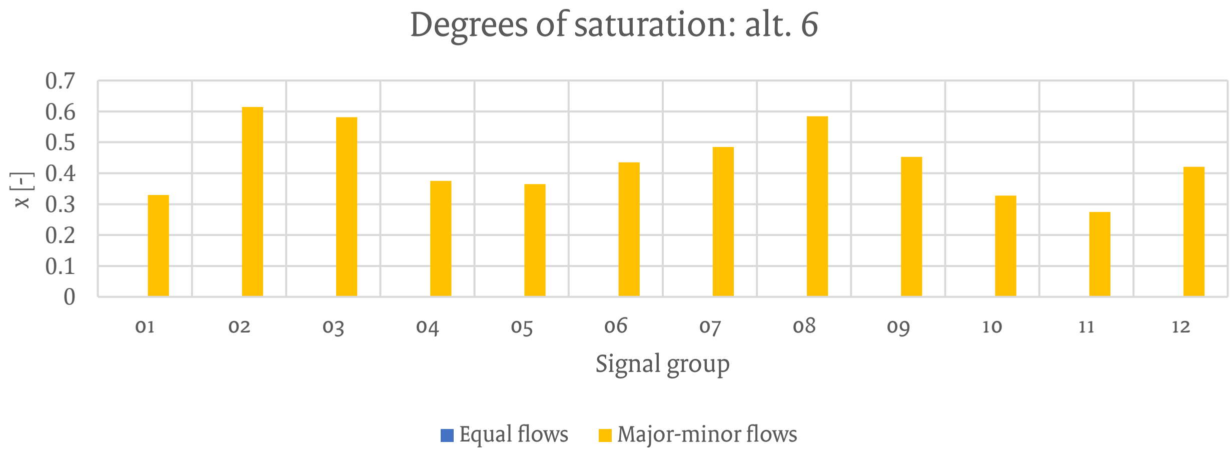

Figure F-27 | Degrees of saturation for alternative (alt.) 6, per traffic flow volume input set.

Table F-13 | Performance statistics (averages) for alternative (alt.) 6, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	0.13	N/A	0.01	N/A	0.02
02	N/A	-0.38	N/A	-0.08	N/A	0.01
03	N/A	0.14	N/A	0.07	N/A	0.00
04	N/A	-0.76	N/A	-0.07	N/A	0.01
05	N/A	0.16	N/A	0.01	N/A	0.00
06	N/A	1.15	N/A	0.10	N/A	0.00
07	N/A	-2.32	N/A	-0.90	N/A	0.00
08	N/A	0.58	N/A	0.24	N/A	0.00
09	N/A	-0.20	N/A	-0.05	N/A	0.00
10	N/A	0.78	N/A	0.02	N/A	0.00
11	N/A	-0.03	N/A	-0.03	N/A	0.00
12	N/A	0.43	N/A	0.03	N/A	0.00

Table F-14 | Performance statistics (standard deviations) for alternative (alt.) 6, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	N/A	7.62	N/A	0.24	N/A	0.13
02	N/A	2.77	N/A	1.01	N/A	0.09
03	N/A	8.98	N/A	0.75	N/A	0.00
04	N/A	5.94	N/A	0.35	N/A	0.09
05	N/A	7.97	N/A	0.58	N/A	0.00
06	N/A	6.14	N/A	0.49	N/A	0.00
07	N/A	8.76	N/A	0.80	N/A	0.00
08	N/A	2.83	N/A	1.07	N/A	0.00
09	N/A	8.10	N/A	1.07	N/A	0.00
10	N/A	5.21	N/A	0.35	N/A	0.00
11	N/A	6.58	N/A	0.42	N/A	0.00
12	N/A	8.97	N/A	0.53	N/A	0.00

F.2.8. Alternative 7

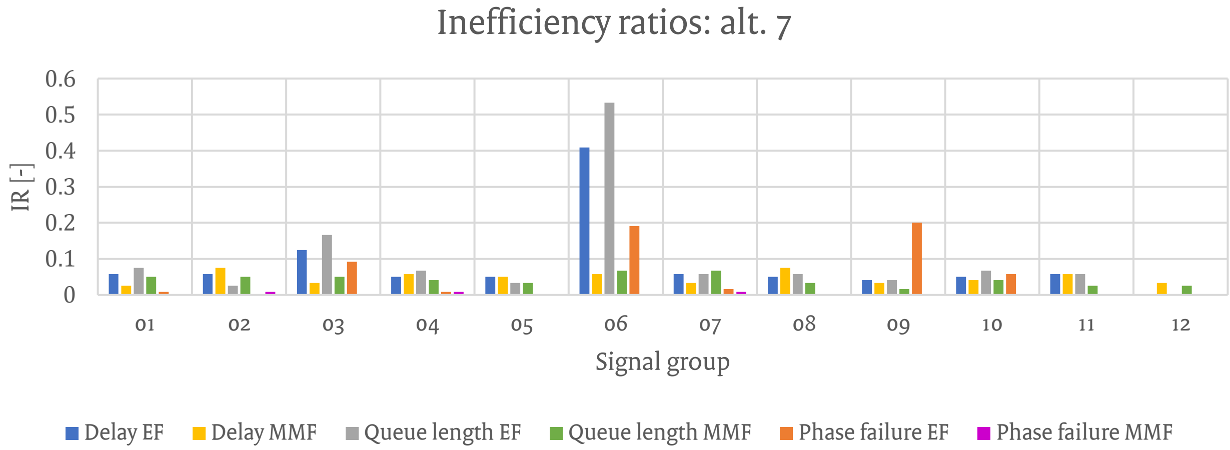

Figure F-28 | Inefficiency ratios for alternative (alt.) 7, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

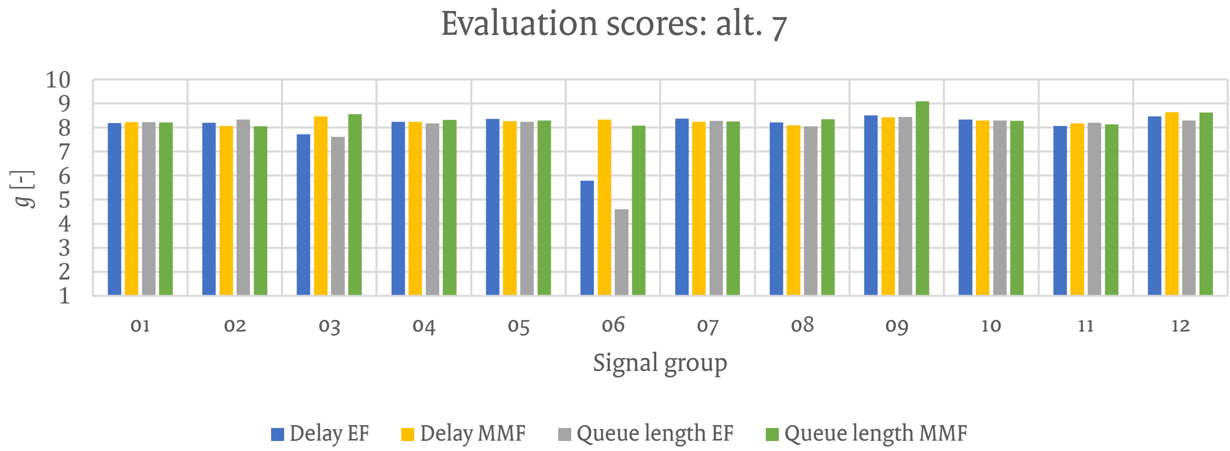

Figure F-29 | Evaluation scores for alternative (alt.) 7, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

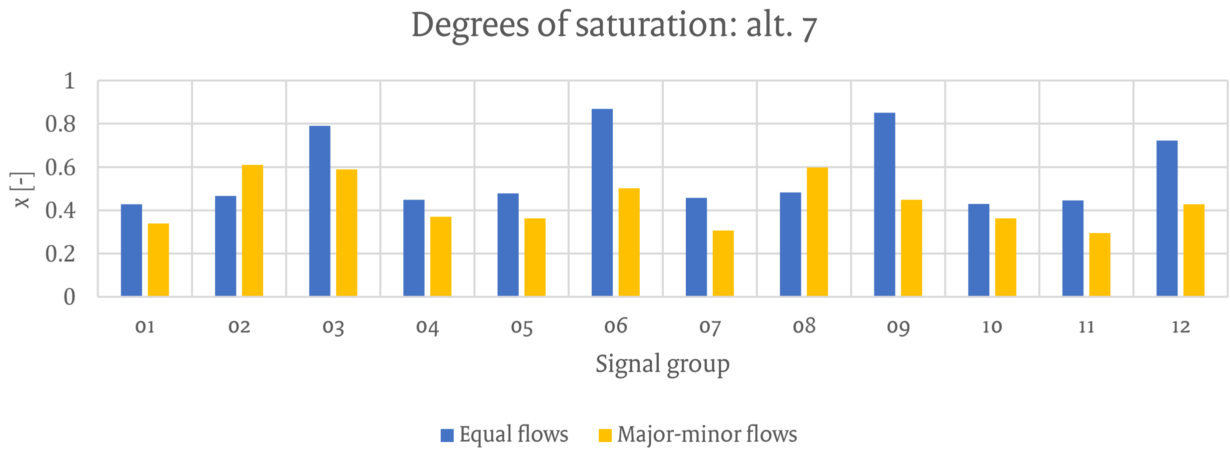

Figure F-30 | Degrees of saturation for alternative (alt.) 7, per traffic flow volume input set.

Table F-15 | Performance statistics (averages) for alternative (alt.) 7, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	0.04	0.21	0.00	0.00	0.01	0.00
02	-0.03	-0.38	-0.06	-0.12	0.00	0.01
03	-2.51	0.37	-1.35	0.01	0.02	0.00
04	0.17	-0.07	-0.01	0.00	0.01	0.01
05	-0.27	-0.26	0.00	0.01	0.00	0.00
06	-8.01	0.07	-3.72	-0.12	0.01	0.00
07	-0.28	-0.19	-0.04	-0.02	0.02	0.01
08	-0.26	0.48	-0.03	0.07	0.00	0.00
09	1.19	-0.37	0.48	0.00	0.03	0.00
10	0.04	-0.11	0.01	0.01	0.06	0.00
11	0.14	0.00	0.04	0.00	0.00	0.00
12	7.16	0.44	3.51	0.01	0.00	0.00

Table F-16 | Performance statistics (standard deviations) for alternative (alt.) 7, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	3.72	4.78	0.42	0.23	0.09	0.00
02	4.29	2.56	0.85	1.01	0.00	0.09
03	6.32	7.98	2.99	0.88	0.13	0.00
04	3.52	6.19	0.42	0.30	0.09	0.09
05	5.08	7.60	0.81	0.49	0.00	0.00
06	8.65	6.67	4.48	0.65	0.09	0.00
07	3.99	5.53	0.41	0.26	0.13	0.09
08	3.95	2.51	0.86	1.00	0.00	0.00
09	7.96	8.99	4.29	1.09	0.16	0.00
10	3.11	5.85	0.33	0.33	0.23	0.00
11	4.72	6.77	0.84	0.42	0.00	0.00
12	4.97	9.77	1.24	0.59	0.00	0.00

F.2.9. Alternative 8

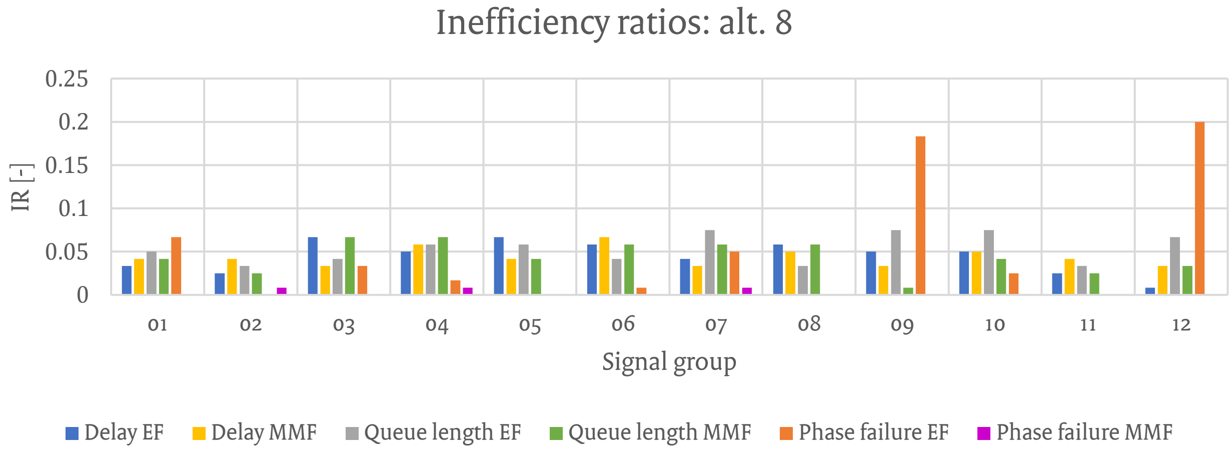

Figure F-31 | Inefficiency ratios for alternative (alt.) 8, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

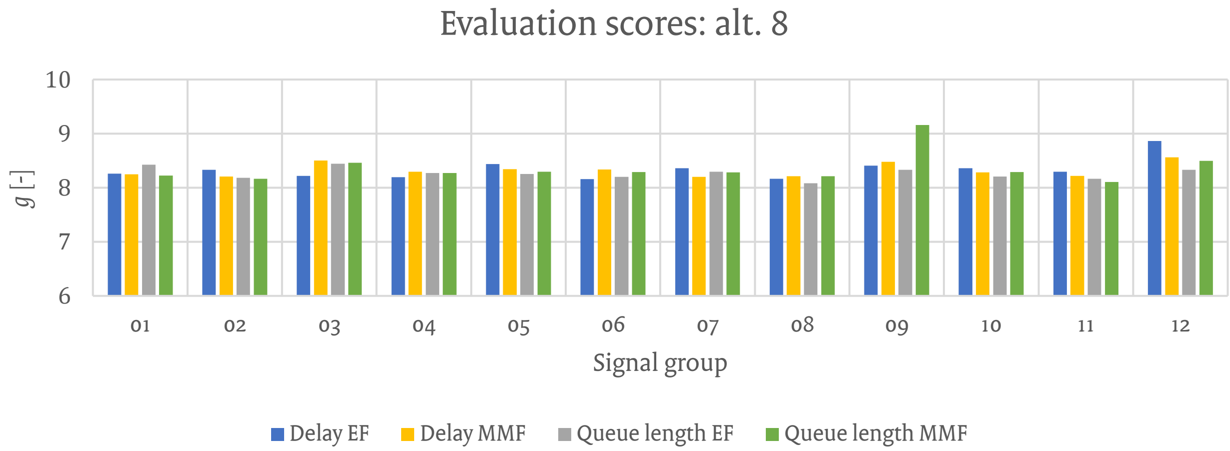

Figure F-32 | Evaluation scores for alternative (alt.) 8, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

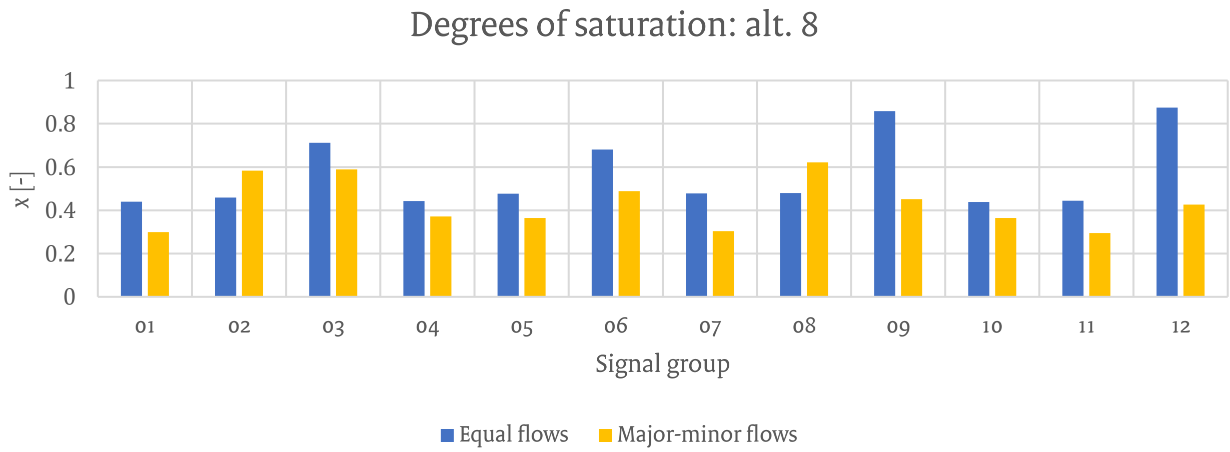

Figure F-33 | Degrees of saturation for alternative (alt.) 8, per traffic flow volume input set.

Table F-17 | Performance statistics (averages) for alternative (alt.) 8, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	-0.2111	0.7270	0.0015	0.0212	0.0667	0.0000
02	-0.3619	-0.0010	-0.0063	0.0001	0.0000	0.0083
03	-0.2204	0.0005	-0.0840	0.0001	0.0250	0.0000
04	-0.0618	-0.0007	-0.0232	0.0000	0.0167	0.0083
05	-0.2856	-0.0011	0.0064	0.0000	0.0000	0.0000
06	-0.0771	0.0005	-0.0236	-0.0003	0.0083	0.0000
07	-0.3309	0.0000	-0.0379	-0.0001	0.0500	0.0083
08	-0.1962	-0.0005	-0.0141	-0.0001	0.0000	0.0000
09	-0.4501	-0.0001	-0.0852	0.0001	0.0083	0.0000
10	-0.0536	0.0000	-0.0332	-0.0001	0.0250	0.0000
11	0.0894	-0.0006	0.0427	0.0001	0.0000	0.0000
12	0.7645	-0.0008	-0.0945	-0.0001	0.0000	0.0000

Table F-18 | Performance statistics (standard deviations) for alternative (alt.) 8, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	3.5500	4.6554	0.3685	0.2207	0.2494	0.0000
02	3.8812	2.3837	0.8057	0.9724	0.0000	0.0909
03	4.8220	8.2739	1.4072	0.9461	0.1561	0.0000
04	3.5209	5.9068	0.3961	0.3010	0.1280	0.0909
05	5.1565	7.3591	0.8200	0.4956	0.0000	0.0000
06	4.3015	6.7222	1.0002	0.5645	0.0909	0.0000
07	3.7879	5.6196	0.3908	0.2491	0.2179	0.0909
08	4.1565	2.3938	0.8640	1.0755	0.0000	0.0000
09	8.7094	8.7855	5.3578	1.0671	0.0909	0.0000
10	3.1320	5.8743	0.3695	0.3591	0.1561	0.0000
11	4.1397	6.5952	0.8495	0.4165	0.0000	0.0000
12	8.8243	10.0531	4.5636	0.6489	0.0000	0.0000

F.2.10. Alternative 9

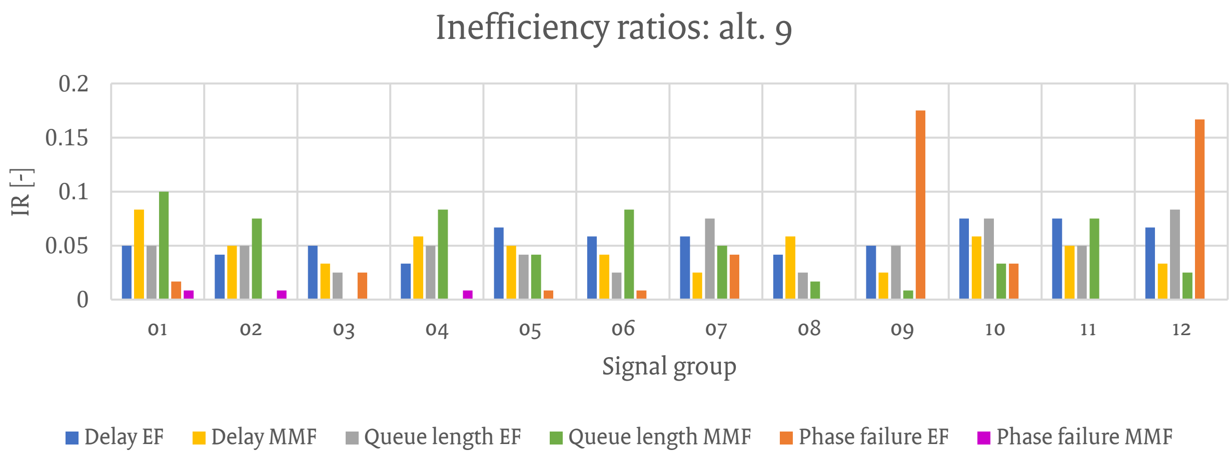

Figure F-34 | Inefficiency ratios for alternative (alt.) 9, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

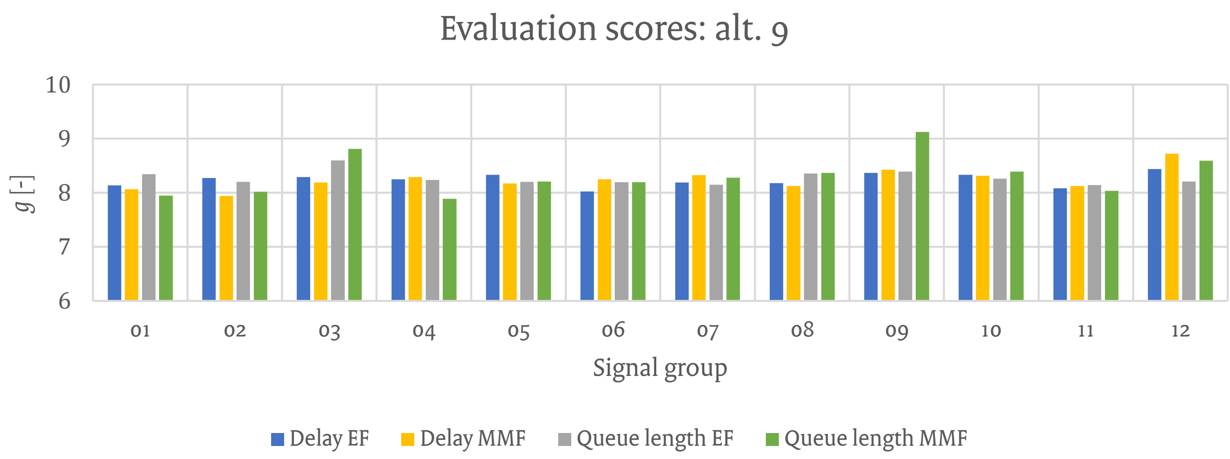

Figure F-35 | Evaluation scores for alternative (alt.) 9, for delay, and queue length, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

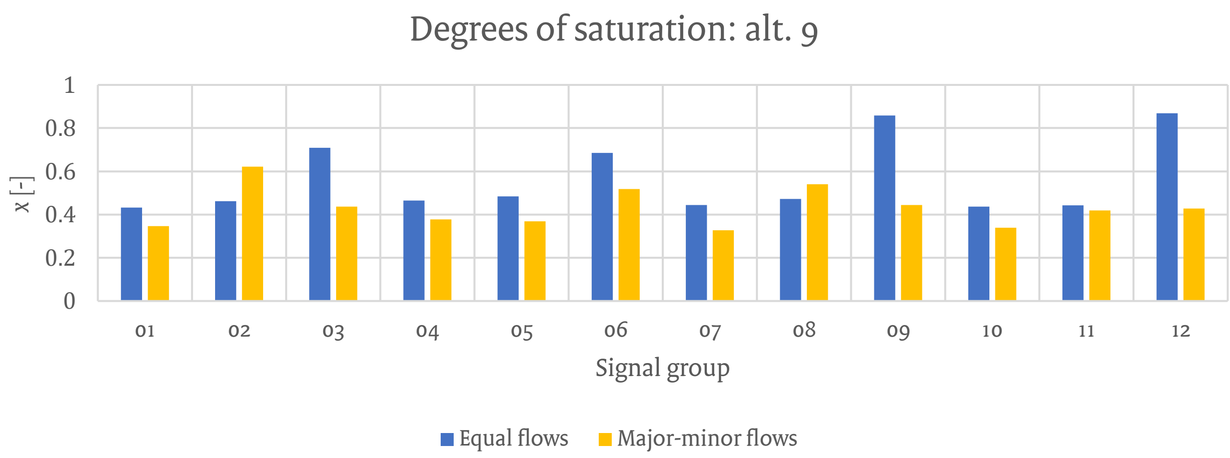

Figure F-36 | Degrees of saturation for alternative (alt.) 9, per traffic flow volume input set.

Table F-19 | Performance statistics (averages) for alternative (alt.) 9, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	-0.05	-0.09	0.00	-0.01	0.02	0.01
02	-0.40	0.00	0.03	-0.08	0.00	0.01
03	0.00	1.71	0.01	0.35	0.02	0.00
04	0.27	-0.51	-0.02	-0.08	0.00	0.01
05	0.30	-0.13	0.00	0.01	0.01	0.00
06	0.32	0.13	-0.04	-0.04	0.01	0.00
07	-0.29	-0.48	-0.06	-0.01	0.04	0.00
08	-0.28	0.13	0.00	0.24	0.00	0.00
09	1.02	-0.01	0.10	0.04	0.01	0.00
10	-0.21	0.39	-0.02	0.01	0.03	0.00
11	0.16	-1.92	0.04	-0.10	0.00	0.00
12	-1.19	0.05	-0.22	-0.01	0.00	0.00

Table F-20 | Performance statistics (standard deviations) for alternative (alt.) 9, for delay, queue length, and phase failure, per traffic flow volume input set (EF = equal flows; MMF = major-minor flows).

Signal group	Delay EF [s/pce]	Delay MMF [s/pce]	Queue length EF [pce]	Queue length MMF [pce]	Phase failure EF [#]	Phase failure MMF [#]
01	3.75	5.65	0.41	0.30	0.13	0.09
02	3.97	2.74	0.80	1.06	0.00	0.09
03	4.70	8.21	1.36	0.49	0.13	0.00
04	3.39	6.14	0.41	0.35	0.00	0.09
05	5.47	7.39	0.85	0.49	0.09	0.00
06	4.40	6.61	0.90	0.61	0.09	0.00
07	4.25	7.03	0.51	0.29	0.20	0.00
08	4.09	2.73	0.76	0.94	0.00	0.00
09	8.23	10.46	4.64	1.02	0.09	0.00
10	3.23	5.59	0.35	0.32	0.18	0.00
11	4.77	6.63	0.87	0.47	0.00	0.00
12	12.48	9.11	5.43	0.59	0.00	0.00