MSc Thesis in Mathematics

Development of an Extended Bathtub Model Capturing Heterogenous Traffic States

Pim Kloet

May 28th 2026

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Mathematics.

The work in this thesis was carried out in the Mathematical Physics group at Delft University of Technology.

Supervisors:	Dr. Havva Yoldaş
	Dr. ir. Irene Martínez
	Prof.dr. Henk Schuttelaars
Co-reader:	Dr. Victor Knoop

Abstract

Highway congestion costs European road networks an estimated €100 billion annually. Highway traffic is characterised by congestion shock waves, arising when density exceeds a critical threshold density. These shock waves cause a heterogeneous density distribution across the network, reducing its overall efficiency. Crucially, this heterogeneity is path-dependent: the network operates less efficiently during recovery from congestion than during its build-up, a phenomenon known as hysteresis. Current network-level traffic models cannot reproduce this, as they rely on a memoryless speed–density relation.

This thesis extends the classical bathtub model — a network-level traffic model driven by average density — with an explicit, path-dependent congestion variable representing the spatial extent of congestion in the network. To calibrate the model, loop detector data from 124 weekday mornings on the A10 ring road in Amsterdam (January–June 2018) were used to construct empirical congestion measures. Three candidate congestion measures were compared — density spread, an unweighted congestion fraction, and a density–weighted variant — on their ability to characterise the network congestion state. The density–weighted measure was selected based on superior fit to observed network speeds ($R^2 = 0.960$). Empirical analysis established a critical density threshold $\rho_{\text{crit}} \approx 17.2$ veh km^{-1} above which congestion onset and recovery follow asymmetric paths, captured by a dynamical system with separate build-up and recovery rates. Calibration confirmed that congestion builds approximately 31% faster than it dissipates. Forward simulation of the calibrated model reproduces the hysteresis loops observed on the A10.

The model offers a tractable basis for traffic management applications such as ramp metering, variable speed limits, and congestion pricing, enabling prediction of the asymmetric dynamics of morning rush congestion. A current limitation is that inflow is treated as exogenous, producing an unrealistically sharp gridlock sensitivity. Incorporating observed inflow data is identified as the most important direction for future work.

Acknowledgements

Writing this thesis has been a journey — one that would not have been possible without the support of many people around me.

First and foremost, I would like to thank my daily supervisors Havva and Irene for giving me the opportunity to work on this project. I have learned an enormous amount from both of you, and your guidance has shaped this thesis in more ways than I can count. I would like to thank Henk for stepping in whenever I got stuck. I always left our meetings with renewed confidence. I would also like to thank Victor for being part of my committee and for his genuine interest in my work. A special thanks goes to Irene's research group for the monthly meetings, the interesting presentations, and the feedback on my work and presentation; these sessions made me feel part of something larger than just my own thesis. To my family and friends: thank you for being there. Thanks for celebrating the good moments and for keeping me going through the harder ones.

I would also like to acknowledge my use of Claude (Anthropic) as an AI assistant for brainstorming ideas and getting feedback on my writing.

Contents

1 Introduction

1.1 Motivation and Societal Relevance

Traffic congestion costs European road networks an estimated €100 billion annually — more than one per cent of EU GDP [Christidis et al., 2012]. Congestion generates substantial losses through delays to freight, excess fuel consumption, and degraded air quality [Treiber et al., 2008; Golob and Regan, 2001; Bedoya-Maya et al., 2022]. These costs are growing: urban populations continue to expand, while land scarcity and environmental constraints severely limit the scope for new road capacity on most European networks [European Commission, 2011]. Efficient use of existing infrastructure has therefore become a policy priority.

A key feature of highway congestion is its tipping-point behaviour. Below a critical vehicle density, traffic flows freely and speeds are near the speed limit. Once that threshold is exceeded, shock waves form — abrupt transitions between fast and slow traffic that propagate backwards through the network — and speeds drop sharply [Lighthill and Whitham, 1955; Richards, 1956]. What makes this particularly consequential is that recovery is not the reverse of onset: once congestion has established itself, a substantially larger drop in demand is needed before the network returns to free flow. This asymmetry between build-up and recovery is an instance of hysteresis — a phenomenon, familiar from physics and economics, in which the state of a system depends not only on current conditions but on its history. In traffic, this means that the same density can correspond to different speeds depending on whether the network is loading or recovering [Geroliminis and Sun, 2011; Knoop et al., 2015; Buisson and Ladier, 2009].

In practice, a large fraction of daily road use is concentrated in two short windows: the morning and evening rush hours. These are precisely the conditions under which the critical density threshold is most likely to be crossed, making the rush hour the primary context in which hysteresis occurs. Management tools such as ramp metering, variable speed limits, and dynamic congestion pricing can be deployed on motorway networks with the aim of keeping density below this threshold [Li et al., 2014; De Palma and Lindsey, 2011]. The effectiveness of any such intervention depends on the quality of the underlying traffic model: a controller must be able to predict how density and speed evolve over the course of the rush hour, and in particular how the network responds once the critical threshold has been exceeded.

Accurately describing these dynamics is therefore both a scientific and a practical priority, and capturing them in network-level models remains an open problem in traffic flow theory. The relevant background and prior work are reviewed in Chapter 2.

1.2 Network-level Traffic Modelling

Predicting network-level congestion dynamics requires a model that is both tractable and calibratable from available data. Detailed traffic simulation models can reproduce individual vehicle behaviour in great detail, but they require origin-destination data, route choice assumptions, and explicit network geometry — data that are often not available and can be costly to collect [Daganzo, 2007]. A more practical alternative is the *network-level* approach, which represents the entire network by a small number of aggregate variables: the average vehicle density ρ and the average network speed v.

The empirical foundation of network-level modelling is the Macroscopic Fundamental Diagram (MFD): a reproducible single-valued relationship between network density and network speed, first found at network level by Godfrey [1969] and later confirmed by Geroliminis and Daganzo [2008]. The MFD can be estimated directly from loop detector measurements, without any knowledge of trip origins, destinations, or routes.

Given an MFD, the dynamics of the network reduce to a single differential equation — the bathtub model. Vehicles enter the network at an inflow rate $f(t)$ and complete their trips at a rate determined by the current density and the MFD. The density $\rho(t)$ then evolves as the balance between inflow and this trip-completion outflow [Vickrey, 1969; Daganzo, 2007; Jin, 2020]. The model requires only loop detector data for calibration and a small number of interpretable parameters. For example, it only uses the average trip distance rather than a full trip distance distribution as required by more general formulations such as the generalised bathtub model [Jin, 2020]. These properties make it well suited as a basis for real-time traffic management on motorway networks.

1.3 Hysteresis in Highway Networks

The oversimplification of the single-valued MFD is most clearly visible in the velocity–density diagram of a congested morning rush. Figure 1.1 shows data from the A10 ring road in Amsterdam between 06:00–10:00, where each point represents the average network density and velocity over a 5 minute period. Rather than a single-valued function, the observations trace a loop: during the loading phase, as density increases, the network follows one branch. During recovery, as density decreases, it follows a different branch at systematically lower velocity. The same density corresponds to two distinct speeds depending on whether the network is loading or unloading.

The mechanism underlying this loop was analysed by Gayah and Daganzo [2011]. During loading, congestion is concentrated at a small number of persistent bottlenecks and the rest of the network remains near free flow. During recovery, however, congestion has spread unevenly across the network, and heavily loaded links clear more slowly than lightly loaded ones. This spatial heterogeneity during recovery reduces aggregate network flow below what the same density produced during loading. This is precisely because, as Cassidy et al. [2011] showed, mixed-regime conditions push observations below the MFD. The result is a clockwise loop in the velocity–density plane, first observed at the network level by Buisson and Ladier [2009] and subsequently confirmed on multiple motorway networks [Geroliminis and Sun, 2011; Knoop et al., 2015].

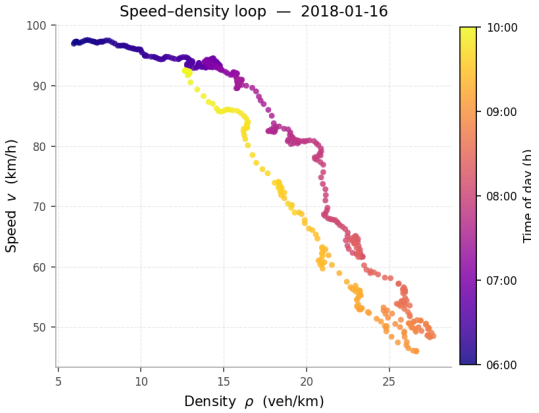


Figure 1.1: Speed–density diagram for the A10 ring road in Amsterdam during the morning rush (06:00–10:00) on 16 January 2018, covering the clockwise direction. Each point represents the network average density and velocity over a 5 minute period. Colour indicates the time of day. The loading and unloading branches trace a clockwise hysteresis loop in the speed–density plane, with the unloading branch lying below the loading branch for the same density. The dataset is described in Chapter 3.

1.4 Knowledge Gap

Despite this well-documented phenomenon, standard network-level models such as the bathtub model are not equipped to reproduce it. Because they assign a single speed to every density value via the MFD, the loading and unloading phases must follow identical paths — there is no mechanism for the network to remember whether it is recovering from congestion or congestion is building.

One attempt to address this limitation is the Generalised Macroscopic Fundamental Diagram (GMFD) [Knoop et al., 2015], which uses the spatial spread of density σ, being the degree of variation in density in the network. It represents the amount of congestion in the network. This density spread is used as a second argument of the MFD, $v = V(\rho, \sigma)$, to create the GMFD. This GMFD can describe hysteresis, because the hysteresis loops are present in the density spread. Despite this advance, the GMFD is a static relationship: it specifies what speed a given (ρ, σ) pair produces, but provides no dynamic equation governing how σ evolves over time. Without such an equation, the GMFD cannot be embedded in a predictive bathtub model.

What is missing is a path-dependent congestion variable with a corresponding dynamic equation that evolves as part of the system state. If congestion were simply a function of current density, the speed–density relation would again be single-valued and hysteresis could not be captured. So if it depends on density, it should do so in an indirect, path-dependent manner. To the best of our knowledge, no existing bathtub-type model fills this gap.

1.5 Research Questions and Objectives

The central research question of this thesis is:

> *How can morning rush hysteresis in highway networks be captured by a bathtub model extended with a path-dependent congestion state variable, and what are the structural and empirical properties of such a model when calibrated to A10 Amsterdam data?*

This question is addressed through four sub-questions.

SQ1. *What are the empirical structural properties of morning rush congestion on the A10 that should constrain the design of a path-dependent bathtub model?*

Before proposing any model, we first establish what the data actually show. We examine whether there is a well-defined onset threshold during loading, whether the recovery phase follows a different trajectory from loading, and whether density spread is correlated with a lagged value of density. These empirical findings directly constrain the structure of the model developed as an answer to SQ2.

SQ2. *What two-state dynamical model with density and a path-dependent congestion variable is consistent with the empirical properties identified in SQ1?*

Given the structural properties established empirically, this sub-question asks how they can be translated into a tractable mathematical model. The challenge is to formulate congestion dynamics that are path-dependent by construction, while remaining consistent with the bathtub framework and simple enough to be calibrated from loop detector data.

SQ3. *What parameter values does the model take when calibrated to the A10 data, and what do they reveal about the structure of morning rush congestion dynamics?*

The model proposed in SQ2 contains parameters governing the speed function and the congestion dynamics. This sub-question concerns their identification from the A10 loop detector data and the interpretation of the resulting values. In particular, we ask what the calibrated parameters reveal about the relative rates of congestion build-up and recovery, and whether these are consistent with the empirical observations from SQ1.

SQ4. *Does this model reproduce the hysteresis loops observed in the A10 data?*

The calibrated model is simulated forward under a stylised trapezoidal inflow profile representing a morning rush hour, and the resulting trajectory is compared qualitatively to the empirical properties concluded from the A10 data. The sensitivity of the model to the peak inflow level is also examined, exploring the transition between scenarios that recover and scenarios that reach gridlock.

1.6 Thesis Outline

Chapter 2 provides the theoretical and empirical background, introducing traffic modelling theory from the LWR model through the MFD and bathtub framework, and reviewing the literature on hysteresis and prior attempts to capture it dynamically. Chapter 3 addresses SQ1 by analysing A10 loop detector data to establish three structural properties of morning

rush congestion: a free-flow regime, a loading threshold, and path-dependent recovery. These findings motivate and constrain the model proposed in Chapter 4, which addresses SQ2 by introducing a two-state dynamical system with an explicit, path-dependent congestion variable. Chapter 5 addresses SQ3 by calibrating the model parameters from the A10 data and interpreting the results. Chapter 6 addresses SQ4 by simulating the calibrated model under an artificial morning rush inflow and demonstrating that it reproduces the clockwise hysteresis loops observed empirically. Finally, Chapter 7 synthesises the findings, answers the central research question, reflects on limitations and outlines directions for future work.

2 Background and Literature

2.1 Fundamental Relations of Traffic: From Link Fundamental Diagram to the MFD

2.1.1 Single-road Traffic Flow and the Fundamental Diagram

The theoretical foundation of macroscopic traffic modelling is the *Lighthill–Whitham–Richards* (LWR) model [Lighthill and Whitham, 1955; Richards, 1956], which treats traffic as a one-dimensional compressible fluid. The primary quantity is the vehicle density $k(x,t) : [0,L] \times (0,\infty) \to \mathbb{R}_+$, measured in vehicles per kilometre, which varies continuously over space $x \in [0,L]$ and time $t \geq 0$. Conservation of vehicles requires

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0, \tag{2.1}$$

where $q(x,t) = k(x,t)\, v(x,t)$ is the traffic flow (vehicles per hour) and $v(x,t)$ is the mean vehicle speed at position x and time t. Equation (2.1) alone is underdetermined: it contains two unknowns, k and q, and must be closed by a relationship between them. This is the *fundamental diagram* $q = Q(k)$, or equivalently the *speed–density relation* $v = V(k)$, so that $Q(k) = k\,V(k)$.

The fundamental diagram encodes the empirical observation that, in equilibrium, vehicle speed decreases monotonically with density. Drivers reduce their speed as the road becomes more crowded, and at some maximum *jam density* $k_{\max}$ traffic comes to a standstill. The minimum requirements for any physically admissible speed–density relation are therefore

(i) $V'(k) \leq 0$ (speed is non-increasing in density),

(ii) $V(0) = v_{\max}$ (free flow at zero density),

(iii) $V(k_{\max}) = 0$ (standstill at jam density).

A variety of functional forms satisfying these conditions have been proposed. The earliest is the linear *Greenshields* relation [Greenshields et al., 1935],

$$V(k) = v_{\max}\left(1 - \frac{k}{k_{\max}}\right), \tag{2.2}$$

which yields the parabolic flow–density curve $Q(k) = v_{\max}\, k\,(1 - k/k_{\max})$, with capacity (maximum flow) achieved at $k = k_{\max}/2$. More commonly used in practice is the triangular fundamental diagram, in which flow increases linearly with density up to a critical density k_c, then decreases linearly to zero at jam density [Newell, 1993]. The linear free-flow branch corresponds to vehicles travelling at (or close to) the speed limit; the congested branch corresponds to queued traffic moving at a much lower speed. Together, the two branches

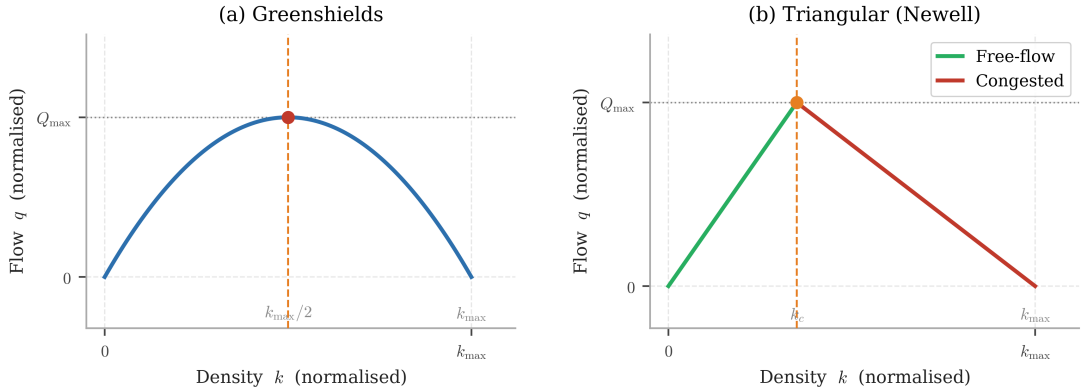


Figure 2.1: Two commonly used fundamental diagrams in the flow–density plane. (a) The Greenshields relation yields a parabolic curve with capacity $Q_{\max}$ achieved at $k = k_{\max}/2$. (b) The triangular fundamental diagram [Newell, 1993] has a linear free-flow branch with slope $v_{\max}$ up to critical density k_c, and a linear congested branch back to jam density $k_{\max}$. For the same average density, a heterogeneous network with links in both regimes produces less flow than a homogeneous one at the same density, a consequence of the concavity of both diagrams.

reproduce the characteristic bimodal structure observed in single-road loop detector data [Cassidy, 1998]. Figure 2.1 illustrates both functional forms.

Regardless of the specific form chosen for the fundamental diagram, the LWR model can predict the formation and propagation of *shock waves*. These arise when downstream density exceeds upstream density, for instance, when freely flowing traffic meets a queue. This creates an abrupt transition between the two regimes. This discontinuity travels at the Rankine–Hugoniot speed

$$w_s = \frac{Q(k_+) - Q(k_-)}{k_+ - k_-}, \tag{2.3}$$

where k_- and k_+ denote the upstream and downstream densities respectively [Hugoniot, 1887; Rankine, 1870]. Since $k_+ > k_-$ implies $Q(k_+) < Q(k_-)$ on a concave fundamental diagram, the numerator is negative and $w_s < 0$: the shock propagates upstream against the flow of traffic.

2.1.2 The Macroscopic Fundamental Diagram

The LWR model describes conditions at a single stretch of road. For the purpose of network-level management this is too complex. What is needed is not a spatially resolved description but an aggregate characterisation of the state of an entire road network. The *Macroscopic Fundamental Diagram* (MFD) plays this role at the network level, in analogy with the fundamental diagram at the link level.

The MFD was conjectured by Godfrey [1969] and explored theoretically by Daganzo [2007], but its empirical existence was first confirmed by Geroliminis and Daganzo [2008] using loop detector data from a 10 km^2 district of central Yokohama. By averaging flow and density

across all detectors over five-minute intervals, they found a remarkably well-defined and reproducible relationship between the network-averaged density and the network-averaged flow: the MFD. The network quantities used throughout this thesis are defined analogously. For a network with N loop detectors, each representing a road segment of length l_s, the *network density* (average density) and *network production* (average flow) are

$$\rho(t) = \frac{\sum_{s=1}^{N} l_s \, k_s(t)}{\sum_{s=1}^{N} l_s}, \qquad P(t) = \frac{\sum_{s=1}^{N} l_s \, q_s(t)}{\sum_{s=1}^{N} l_s}, \tag{2.4}$$

where k_s and q_s are the density and flow at detector s. The MFD is then the relationship $P = P(\rho)$, or equivalently the speed–density relation $v = V_{\text{MFD}}(\rho) = P(\rho)/\rho$. The two representations carry identical information; this thesis works primarily with the speed–density form $v = V(\rho)$, as it enters directly into the bathtub model developed in Section 2.2.

The empirical and theoretical significance of the MFD is substantial. It allows an entire urban or motorway network to be described by a single scalar state variable, the network density $\rho(t)$, and a corresponding scalar output, the production $P(t)$. No origin-destination data, no route assignment, and no detailed network geometry are required.

Theoretically, Daganzo [2007] showed that a well-defined MFD exists whenever vehicles are distributed homogeneously across the network, in the sense that all links operate at the same level of saturation. Under this condition, network-level production is uniquely determined by network-level density. So the price paid for the simplicity is this homogeneity assumption.

2.2 The Bathtub Model

2.2.1 Vickrey's Formulation

The MFD makes it possible to describe an entire road network by a single scalar state variable, the network density $\rho(t)$ (vehicles per network kilometre). The dynamical model that governs how $\rho(t)$ evolves over time is known as the *bathtub model* or *accumulation-based model*, a term coined by Vickrey [2019] and later popularized by Daganzo [2007]. It captures the analogy between a network filling up with vehicles and water rising in a bathtub.

The model treats the road network as a single reservoir, with a single speed. Vehicles enter at an exogenous inflow rate $f(t)$ (vehicles per hour per network kilometre) and complete their trips at an outflow rate $g(t)$ that depends on current network conditions. Vehicle conservation then gives

$$\frac{d\rho}{dt} = f(t) - g(t), \tag{2.5}$$

where $\rho(t)$ is the average vehicle density in the network.

To close the system, an expression for the outflow $g(t)$ is needed. This closure rests on two ingredients: the MFD, which gives the average network speed $v = V(\rho)$, and an assumption about trip lengths. In Vickrey's formulation, the average trip length of all the vehicles in the network B remains constant. A vehicle travelling at speed v completes a distance B in time

B/v. Aggregating over all vehicles in the network, the rate at which trips are completed, and vehicles therefore leave the network, is

$$g(t) = \frac{\rho(t)\, V(\rho(t))}{B}. \tag{2.6}$$

The numerator $\rho\, V(\rho)$ is the network production per unit road length: the total vehicle-kilometres travelled per hour per kilometre of network. Dividing by the average trip length B converts this into a vehicle departure rate in vehicles per hour per kilometre of network, which when integrated over the network gives the total outflow $g(t)$. Note that this definition of $g(t)$ is a modelling assumption and thus a simplification of reality. Substituting $g(t)$ into (2.5) yields the complete bathtub model,

$$\frac{d\rho}{dt} = f(t) - \frac{\rho(t)\, V(\rho(t))}{B}. \tag{2.7}$$

This is a single ordinary differential equation in the scalar state variable $\rho(t)$, driven by the exogenous inflow profile $f(t)$. It requires only three inputs: the MFD $V(\rho)$, the average trip length B, and the inflow $f(t)$. The MFD and inflow can be estimated directly from loop detector data, while B is difficult to identify directly from loop detector data alone.

2.2.2 Assumptions and Limitations

The bathtub model rests on two assumptions that deserve attention.

1. The first is the *homogeneity assumption* inherited from the MFD: the model assigns a single speed $V(\rho)$ to the entire network, implicitly treating all vehicles as if they were distributed uniformly. As discussed in Section 2.1.2, a well-defined MFD only exists when this condition holds. In reality, link densities are heterogeneous: some links are near free flow while others are heavily congested. As will be shown in Section 2.3.1, this heterogeneity reduces network productivity below what $V(\rho)$ would predict, and the bathtub model has no mechanism to represent it. This is the fundamental reason it cannot reproduce hysteresis.
2. The second is the *fixed average trip length assumption*. By treating B as a constant, the model assumes that the average distance travelled by vehicles does not vary over time and is the same for all trips. As pointed out by Jin [2020], this prevents the model from distinguishing between short and long trips, which may respond differently to congestion.

2.2.3 The Generalized Bathtub Model

Jin [2020] addressed the fixed trip length assumption by introducing the Generalized Bathtub Model. Instead of tracking only the network density $\rho(t)$, the generalized model tracks the full distribution of remaining trip distances through the variable $K(t, x)$, defined as the number of vehicles with a remaining distance of at least x at time t. Vehicle conservation in the (t, x) plane then yields a partial differential equation governing the evolution of $K(t, x)$, which can accommodate any trip distance distribution.

The generalized model is more realistic, but also substantially more complex. Its state is a function rather than a scalar, and its calibration requires data on the distribution of trip lengths,

which is difficult to obtain. For the purpose of this thesis — where the primary objective is to incorporate congestion dynamics rather than a general trip distance distribution — the added complexity is not warranted. We therefore work with the simpler Vickrey formulation throughout, noting that the congestion extension developed in Chapter 4 could in principle be combined with the generalized bathtub framework in future work.

2.3 Heterogeneity and Hysteresis

2.3.1 Spatial Heterogeneity and the Limits of the MFD

The MFD and the bathtub model rest on the assumption that vehicles are distributed homogeneously across the network. In practice, this condition is rarely satisfied. Congestion is not a uniform phenomenon: it originates at specific bottlenecks — on-ramps, lane reductions, sections with reduced speed limits — and spreads as backward-propagating shock waves through the surrounding road segments. The network therefore operates simultaneously in two distinct regimes: a congested zone close to the bottleneck, where densities are high and speeds are low, and a free-flow zone elsewhere, where traffic moves near the speed limit. This coexistence of regimes is what we refer to as spatial heterogeneity.

The consequences for the MFD are quantifiable. As shown analytically by Cassidy et al. [2011] for networks with triangular fundamental diagrams on individual links, the network-averaged production falls below the MFD whenever the network is in a mixed-regime state. The reason is that the MFD is a concave function of density: the average of a high-density (congested) point and a low-density (free-flow) point at the same mean density lies below the curve. In other words, for the same number of vehicles in the network, a heterogeneous distribution produces less throughput than a homogeneous one.

Empirically, Buisson and Ladier [2009] demonstrated this effect for a motorway network in Toulouse, finding that the scatter in the flow–density diagram increased substantially on days with spatially uneven congestion. Geroliminis and Sun [2011] confirmed the same pattern across multiple networks, showing that the sharpness of the MFD is directly related to the degree of spatial homogeneity.

2.3.2 Hysteresis: the Network-Level Loop

The most systematic manifestation of spatial heterogeneity during a morning rush is *hysteresis*: the observation that the network follows different trajectories in the speed–density plane depending on whether vehicle density is increasing or decreasing. The same average density corresponds to different average speeds during the loading and unloading phases, tracing out a loop rather than a single curve. This phenomenon was first documented at the link level by Treiterer and Myers [1974] and subsequently observed at the network level by Geroliminis and Sun [2011] and Knoop et al. [2015].

The mechanism behind the network-level loop can be described as follows. During the loading phase, demand rises and average density increases. At low densities, traffic is homogeneous and the network follows the free-flow branch of the MFD. Once density exceeds a critical level, small disturbances — merging conflicts, or sudden braking events — can trigger local instabilities that amplify into self-sustaining stop-and-go waves. These

events are called *nucleations* [Kerner, 2004], more on them in the next section. As a nucleation propagates upstream, it creates a region of high local density and low speed called a *shock wave*, while the rest of the network remains near free flow. The network has become spatially heterogeneous.

Once congestion is established, it tends to persist even as demand begins to fall. During loading, congestion is concentrated at a small number of bottlenecks while the rest of the network remains near free flow. During recovery, however, congestion has spread spatially across many links, and this heterogeneous distribution does not dissolve immediately as demand decreases. Gayah and Daganzo [2011] formalised this as the *persistence* of congestion: the network does not return to free flow until demand has fallen substantially below the loading threshold. During this recovery phase, the spatial heterogeneity — and hence the suppressed production — remains in place at densities where none was present during loading. This is the hysteresis loop.

The direction of the loop in the flow–density plane is clockwise: during loading, production is high relative to density; during recovery, production is lower for the same density. This is consistent with the empirical observations on the A10 shown in Figure 1.1 and with the findings of Geroliminis and Sun [2011] across multiple networks. Clockwise loops are the most commonly reported form, though other shapes — including counter-clockwise loops and figure-eight patterns — have been observed and attributed to the hierarchical structure of certain networks and routing effects [Xu et al., 2023]. This thesis focuses exclusively on the clockwise case as observed on the A10.

2.3.3 Highway networks and Nucleation Dynamics

Hysteresis has been documented in both urban and highway networks, though the evidence is considerably more abundant for the latter. Empirical evidence for pronounced hysteresis loops in urban networks remains scarce [Mahmassani et al., 2013]. In urban networks, sources of local disturbance — traffic lights, roundabouts, pedestrian crossings, and narrow sections — are structurally guaranteed and affect virtually all vehicles. Their cumulative effect on the speed–density relation is relatively predictable, producing a tight MFD [Geroliminis and Daganzo, 2008] in which the loading and unloading branches are difficult to distinguish.

On highway networks, by contrast, hysteresis is a well-documented and recurring phenomenon [Geroliminis and Sun, 2011; Buisson and Ladier, 2009]. The mechanism is rooted in the sparse and localised nature of highway bottlenecks. Under low-to-moderate demand, vehicles travel at or near the speed limit and bottlenecks — on-ramps, lane drops, sections with reduced speed limits — remain dormant: local densities are too small to trigger a breakdown. Once demand rises above a critical level, however, a bottleneck can be activated by a small local fluctuation, a process Kerner refers to as *nucleation* [Kerner, 2004]. The activated bottleneck generates a growing congested cluster that propagates upstream as a shock wave, spreading congestion far beyond the bottleneck itself and into road segments that were previously in free flow.

The critical asymmetry — and the origin of hysteresis — lies in the difference between the activation and dissolution thresholds. Activating a bottleneck requires only a small fluctuation at sufficiently high density, but once the congested cluster is established it is self-sustaining: the outflow from the congested zone is lower than the inflow that created it, and the cluster persists even as demand begins to fall. The network does not return to

free flow until density has dropped substantially below the level at which congestion formed [Kerner, 2004]. This asymmetry between onset and recovery is precisely what produces the hysteresis loop in the speed–density plane.

The standard bathtub model cannot reproduce this behaviour. Because it assigns a single speed $V(\rho)$ to every density value through the MFD, the loading and unloading phases will follow the same path in the speed–density plane, leaving no room for path-dependent dynamics.

2.4 The Generalised Macroscopic Fundamental Diagram

2.4.1 Definition and Empirical Support

Spatial heterogeneity underlies both MFD scatter and hysteresis, yet the standard MFD treats it as noise rather than as a meaningful quantity. The natural modelling response is to make heterogeneity an explicit argument of the speed function. This is the idea behind the *Generalised Macroscopic Fundamental Diagram* (GMFD), introduced by Knoop et al. [2015].

The GMFD extends the standard speed–density relation by adding the spatial spread of density $\sigma(t)$ as a second argument:

$$v = V(\rho, \sigma), \tag{2.8}$$

where $\sigma(t)$ is the standard deviation of the local density distribution across the network sensors,

$$\sigma(t) = \sqrt{\frac{1}{N}\sum_{s=1}^{N}(k_s(t) - \rho(t))^2}, \tag{2.9}$$

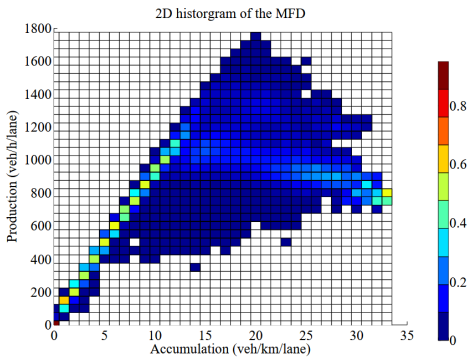
with $k_s(t)$ the density at sensor s and $\rho(t)$ the network-average density as defined in Equation (2.4).

The GMFD satisfies two conditions, both empirically motivated and theoretically grounded in the mixed-regime analysis of Cassidy et al. [2011]:

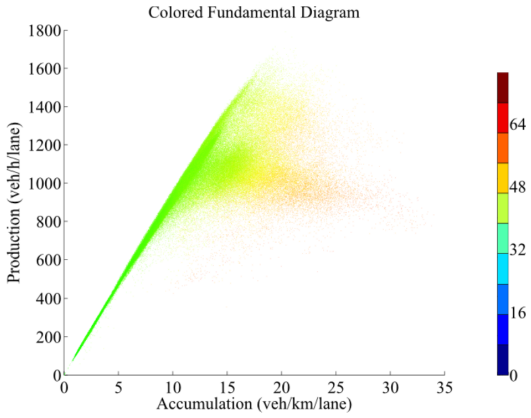
(i) $\partial V/\partial \rho \leq 0$: average speed decreases with average density, as in the standard MFD;

(ii) $\partial V/\partial \sigma \leq 0$: average speed decreases with density spread, because a more heterogeneous network has more vehicles caught in congested segments at lower speed.

Empirical support for the GMFD on the A10 ring road is shown in Figure 2.2, adapted from Knoop and Hoogendoorn [2013]. The left panel shows the standard production–density relation: a broad relationship exists, but with substantial scatter at intermediate densities — precisely the range where mixed-regime conditions are most likely. The right panel colours the same observations by density spread σ: the scatter is largely accounted for by heterogeneity, with higher σ corresponding to systematically lower production for the same average density. This demonstrates that the GMFD captures meaningful structure that the standard MFD cannot.



(a) Histogram of the measurements in the accumulation production plane. The color shows which fraction of the measurements at a certain accumulation has the corresponding production.



(b) The MFD in the accumulation production plane. The color shows the spatial spread of density (vehicles/km/lane). It shows that the higher the accumulation, the higher this spead. Furthermore, *for a certain accumulation* the lower the production the higher the spread.

Figure 2.2: Empirical MFD and spatial density spread for the A10 ring road in Amsterdam. The scatter in the standard production–density relation (left) is largely explained by the spatial spread of density (right): for the same average density, higher density spread corresponds to lower production. Adapted from Knoop and Hoogendoorn [2013].

2.4.2 The GMFD as a Static Relationship and its Limitations

The GMFD is a significant advance over the standard MFD. By explicitly representing heterogeneity, it can account for congested network states that lie below the free-flow curve and explain the scatter observed in empirical data. However, it has a fundamental limitation: it is a *static* relationship. It specifies what speed a given pair (ρ, σ) produces, but says nothing about how σ itself evolves over time.

This limitation becomes apparent when one attempts to use the GMFD as the basis for a predictive bathtub model. The bathtub framework is a dynamical system in which the state variable $\rho(t)$ evolves according to a differential equation. To produce predictions of how the network trajectory unfolds over the course of a morning rush, a second differential equation governing the evolution of $\sigma(t)$ is needed, and the GMFD alone does not provide one. Moreover, such an equation cannot simply relate σ to the current value of ρ through a static function $\sigma = \sigma(\rho)$. Hysteresis requires that the same average density correspond to different network speeds depending on the history of the system; within the GMFD framework this is only possible if σ takes different values at the same ρ depending on that history. Any dynamic equation for σ must therefore be path-dependent — its evolution must depend on whether the network is currently loading or unloading, not only on the instantaneous value of ρ.

A further challenge is that the bathtub model contains no spatial information by construction. The density spread σ is a measure of variation across link-level densities, but the bathtub framework represents the network by aggregate scalars only. The dynamics of σ must therefore be described entirely in terms of aggregate quantities — current density and the rate of density change — without access to the underlying spatial distribution that generates σ.

2.5 Synthesis and Research Gap

The preceding sections have established the theoretical and empirical context for this thesis. Three conclusions emerge that together define the research gap.

First, the standard bathtub model cannot reproduce hysteresis. Because it assigns a single speed $V(\rho)$ to every density value through the MFD, the loading and unloading phases follow identical paths in the speed–density plane. The model has no memory: it cannot distinguish whether the network is building up towards congestion or recovering from it.

Second, the GMFD provides the right conceptual framework — representing speed as a function of both density and heterogeneity — but remains a static relationship. It specifies what speed a given (ρ, σ) pair produces, but provides no dynamic equation governing how σ evolves over time. Without such an equation, the GMFD cannot be embedded in a predictive bathtub model.

Third, any dynamic equation for σ must be path-dependent by construction. If σ were simply a function of current density, the speed–density relation would again be single-valued and hysteresis could not be captured. The equation must therefore distinguish between the loading and unloading phases — precisely the asymmetry that the nucleation dynamics of Section 2.3.3 suggest is physically grounded.

To the best of our knowledge, no existing bathtub-type model incorporates a path-dependent congestion variable together with a dynamic equation governing its evolution. This thesis addresses that gap by extending the classical bathtub model with an explicit congestion state variable $c(t)$ whose dynamics are governed by a direction-dependent differential equation. The empirical properties that any such variable must satisfy are established in Chapter 3, the model is formulated in Chapter 4, and its parameters are calibrated to the A10 data in Chapter 5. The model is tested using an artificial inflow in Chapter 6 and Chapter 7 discusses the results and concludes.

3 Empirical Analysis of the A10

3.1 Dataset Description and Preprocessing

Data Source, Study Period and Spatial Domain

The empirical analysis in this chapter is based on loop detector measurements collected from the A10 ring road surrounding Amsterdam. The A10 is a motorway ring road of approximately 32.3 km that encircles the city (Figure 3.1). The data were provided by the research group Traffic and Planning at Delft University of Technology and originate from the National Data Warehouse for Traffic Information (NDW). The study period runs from 1 January to 30 June 2018, restricted to the morning rush window of 06:00–10:00. This four-hour window captures the build-up and, in most cases, the full recovery of morning congestion on the A10. Table 3.1 summarises the main characteristics of the dataset.

The analysis covers the clockwise direction of the A10. The spatial domain begins at kilometre marker 0, in-between exit 17 and 18, and covers 157 virtual detector locations at 200 m spacing, corresponding to approximately 31.4 km of the 32.3 km ring. All weekdays in the study period are included. Weekend days and Dutch public holidays are excluded, as their demand profiles differ systematically from working-day commuting patterns. The 9 public holidays falling within the January–June 2018 window are listed in Appendix A. Days on which one or more detectors recorded sensor malfunction values (flow or speed equal to 99 999) are removed in their entirety; this affected only 30 June 2018. The resulting dataset comprises 124 usable weekday mornings.

Figure 3.1: Map of the A10 ring road surrounding Amsterdam [Blokker and Blokker, 2013].

Table 3.1: Overview of the A10 Amsterdam dataset.

Property	Value
Time period	1 January – 30 June 2018
Time window per day	06:00–10:00
Measurement interval	30 seconds
Number of virtual detectors	157
Detector spacing	200 m
Total road length covered	31.4 km
Speed limit	100 $\mathrm{km\,h^{-1}}$ (80 $\mathrm{km\,h^{-1}}$ on $\approx$ 5 km western section)
Variables recorded	Vehicle count (flow), average speed
Data source	NDW (pre-processed, see below)
Direction	Clockwise

Raw Measurements and Preprocessing

The data are recorded by double loop detectors. At each detector location two quantities are recorded at 30-second intervals: the number of vehicles passing the detector during that interval q, already converted to $\mathrm{veh\,h^{-1}}$ (vehicles per hour), and the average speed v [$\mathrm{km\,h^{-1}}$] of those vehicles. The reported speed is the *time-mean speed*, defined as the arithmetic average of the instantaneous speeds of all vehicles recorded during the interval [Nationaal Dataportaal Wegverkeer, 2023].

The speed limit on the A10 is 100 $\mathrm{km\,h^{-1}}$ for most of the ring. An exception is a section of approximately 5 km on the western side, between Knooppunt De Nieuwe Meer and exit 2 (Amsterdam–Sloterdijk), where the speed limit is 80 $\mathrm{km\,h^{-1}}$.

The original detector stations are distributed at irregular intervals along the road. To obtain a uniform spatial grid suitable for network-level aggregation, the raw measurements were mapped onto 157 *virtual detector* stations at equal intervals of 200 m. This was carried out prior to this study using the Adaptive Smoothing Method (ASM) developed by Treiber and Helbing [2002]; Schreiter et al. [2010]. The ASM reconstructs a smooth, continuous traffic state field from sparse raw observations by computing, at each output point, a weighted average of neighbouring measurements. The weights are determined by an exponential kernel that is skewed along the characteristic wave directions of traffic: a forward-propagating kernel for free-flow conditions (wave speed ≈ 80 $\mathrm{km\,h^{-1}}$) and a backward-propagating kernel for congested conditions (wave speed ≈ -18 $\mathrm{km\,h^{-1}}$). The output at each grid point is a blend of these two regime-specific estimates, with the blending weight determined by local speed. The processed dataset is therefore not a simple spatial interpolation of detector readings but a physically informed reconstruction of the traffic field that respects the wave structure of congestion. Throughout this chapter the 157 virtual stations are treated as the primary spatial units of observation.

Derived Variables

From the measured flow $q_s(t)$ and speed $v_s(t)$ at virtual detector s, the local traffic density is derived via the fundamental relation

$$k_s(t) = \frac{q_s(t)}{v_s(t)}. \tag{3.1}$$

Each virtual detector is taken to represent a road segment of length $L_s = 200$ m. Since all segments have equal length, spatial averaging over detectors reduces to a simple arithmetic mean.

The following network-level variables are derived from the detector measurements and used throughout this chapter.

Network density $\rho(t)$ [veh km^{-1}]: the spatially averaged density across all $N = 157$ detectors,

$$\rho(t) = \frac{1}{N} \sum_{s=1}^{N} k_s(t). \tag{3.2}$$

This quantity represents the overall loading of the network at each instant.

Network Average Velocity $v(t)$ [km h^{-1}]: the density-weighted spatial average speed,

$$v(t) = \frac{\sum_{s=1}^{N} k_s(t) v_s(t)}{\sum_{s=1}^{N} k_s(t)} = \frac{\sum_{s=1}^{N} q_s(t)}{\sum_{s=1}^{N} k_s(t)}. \tag{3.3}$$

This represents the average vehicle speed of the network.

Network production $P(t)$ [veh h^{-1}]: the spatially averaged flow,

$$P(t) = \frac{1}{N} \sum_{s=1}^{N} q_s(t). \tag{3.4}$$

This represents the aggregate throughput of the network.

The following variables are proposed as candidate measures of the network congestion state; their suitability is evaluated in Section 3.5.

Spatial spread of density $\sigma(t)$ [veh km^{-1}]: the standard deviation of local densities across detectors,

$$\sigma(t) = \sqrt{\frac{1}{N} \sum_{s=1}^{N} (k_s(t) - \rho(t))^2}. \tag{3.5}$$

Low values indicate a spatially homogeneous traffic state, while high values indicate localised congestion coexisting with free-flow conditions elsewhere in the network.

Unweighted congestion fraction $c_{\text{unw}}(t)$: the fraction of detectors currently operating below a critical speed threshold,

$$c_{\text{unw}}(t) = \frac{1}{N} \sum_{s=1}^{N} \mathbf{1}[v_s(t) < f_{\text{crit}}\, v_{\text{lim},s}]\,, \tag{3.6}$$

where $v_{\text{lim},s}$ is the speed limit at detector s and $f_{\text{crit}} \in (0,1)$ is a threshold fraction. A sensor is classified as congested when its speed falls below f_{crit} times its local speed limit. The threshold is expressed as a fraction of the local limit rather than an absolute value to account for the two different speed limits present on the A10. This measure is the fraction of road segments currently congested.

Density-weighted congestion fraction $c_{\text{w}}(t)$: the density-weighted fraction of detectors operating below the critical speed,

$$c_{\text{w}}(t) = \frac{\sum_{s=1}^{N} k_s(t)\, \mathbf{1}[v_s(t) < f_{\text{crit}}\, v_{\text{lim},s}]}{\sum_{s=1}^{N} k_s(t)}. \tag{3.7}$$

By weighting each detector by its local density, this measure captures the fraction of *vehicles* currently located in a congested area, rather than the fraction of road segments. A heavily loaded congested link therefore contributes more than a lightly loaded one.

Both congestion fractions take values in $[0,1]$, where 0 indicates a fully free-flowing network and 1 indicates that all road segments (or all vehicles) are congested. The two measures differ during spatially heterogeneous conditions: c_{w} places greater weight on links with many vehicles.

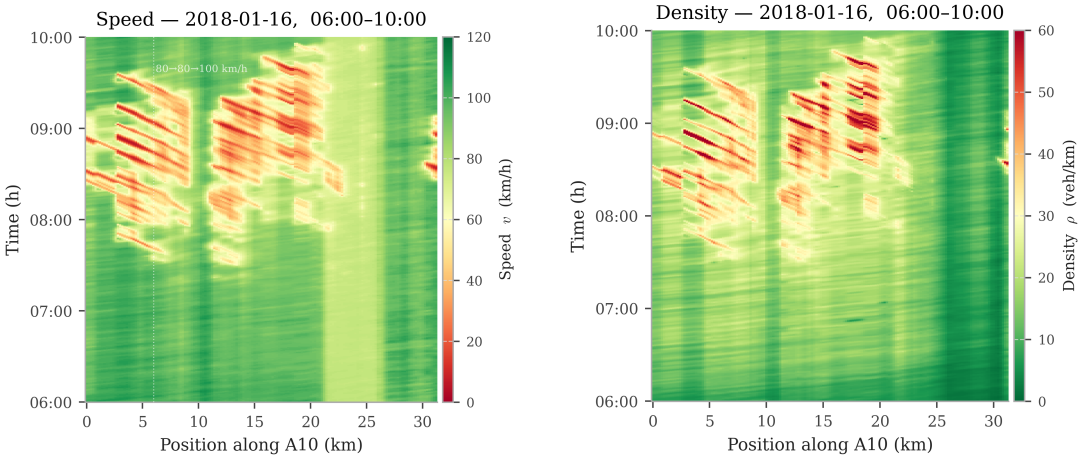
3.2 Local Analysis: Congestion Structure

Before aggregating to the network level, it is instructive to examine congestion at the level of individual virtual detector locations. Working with the spatially resolved data preserves information about the spatial structure of congestion that is lost in network averages, and allows the physical mechanism to be identified directly.

A space–time (x–t) diagram represents the measured traffic state at every detector location and every point in time as a two-dimensional field, with position x along the ring on the horizontal axis and time t on the vertical axis. The colour at each point encodes a traffic variable, so that the evolution of congestion across the full spatial extent of the ring can be read at a glance.

Figure 3.2 shows speed and density space–time diagrams for 16 January 2018 side by side. Both use a red–yellow–green colour scale (low speed = red, free flow = green). The light green vertical band between kilometre 21 and 26 in the velocity diagram is the region with 80 km h^{-1} speed limit.

The most salient feature of Figure 3.2 is a set of diagonal bands of low speed and high density, the shock waves. These bands run consistently from lower-right to upper-left in the x–t plane,



(a) Speed space–time diagram. Green regions indicate free-flow conditions at or near the applicable speed limit. Red and yellow bands indicate congestion propagating upstream (lower-right to upper-left).

(b) Density space–time diagram on a logarithmic colour scale. High-density regions (dark red) align precisely with the low-speed bands, confirming that the diagonal structures are genuine congestion waves.

Figure 3.2: Space–time diagrams for the A10 (clockwise direction), 16 January 2018, 06:00–10:00. Horizontal axis: position along the ring from kilometre marker 0; vertical axis: time of day.

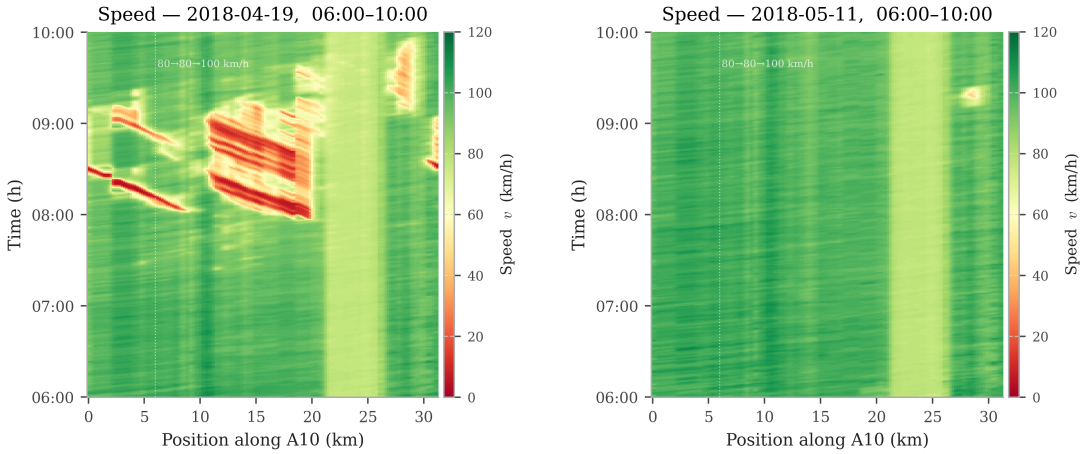
indicating that they move upstream — counter to the direction of vehicle travel — at an approximately constant rate. The slope of the bands corresponds to a backward propagation speed of 15–25 $\mathrm{km\,h^{-1}}$, consistent with the kinematic wave speeds predicted by the LWR model for typical motorway conditions [Lighthill and Whitham, 1955; Richards, 1956].

A closer inspection reveals that the waves do not initiate at random positions. They originate at a small number of specific locations along the ring — recurring positions corresponding to geometric features of the road such as on-ramps, weaving sections, or lane reductions. These are persistent bottlenecks: on days with high demand they activate early and generate long-lived waves that propagate far upstream; on lighter days the same locations produce only short-lived disturbances.

Figure 3.3 shows speed space–time diagrams for two contrasting mornings: 19 April 2018, a second heavily congested day from a different month of the study period, and 11 May 2018, a day on which the network remained in free flow throughout the morning rush.

On the congested day, backward-propagating shock waves are again clearly visible, initiating at similar bottleneck locations as on 16 January and propagating upstream at comparable speeds. The timing and intensity of individual waves differ between the two congested days, but the qualitative structure is consistent. This reproducibility across months confirms that the pattern observed on 16 January is not exceptional but reflects a systematic feature of the A10 network under high demand.

The free-flow day presents a qualitatively different picture. No diagonal structures are visible; speed is near-uniform across the full spatial extent of the ring for the entire four-hour window. This contrast makes concrete what distinguishes a congested morning from a free-flow one



(a) 19 April 2018: speed space–time diagram on another congested morning. Backward- propagating shock waves initiate at the same bottleneck locations as in Figure 3.2a, demonstrating the systematic character of congestion on the A10.

(b) 11 May 2018: speed space–time diagram on a free-flow morning. Speed is near-uniform across the ring throughout the window; no backward-propagating structures are present.

Figure 3.3: Speed space–time diagrams for two contrasting mornings on the A10 (clockwise direction), 06:00–10:00. Left: a heavily congested day; right: a day in which the network remains in free flow throughout.

at the local level: it is the presence or absence of shock wave formation, not a gradual, network-wide reduction in speed.

3.3 Network-level Overview: the Speed–Density Relation

The space–time diagrams above established that congestion on the A10 takes the form of backward-propagating shock waves originating at persistent bottlenecks. Whether a morning exhibits hysteresis, however, depends not on individual waves but on how many form, how far they propagate, and how long they persist. To examine this, we aggregate the 157 detector measurements into the network-level variables and study the relation between network-averaged speed $v(t)$ and network density $\rho(t)$ across all weekday mornings.

Figure 3.4 shows the network density $\rho(t)$ as a function of time of day, averaged across all 124 weekday mornings, with individual days shown as faint traces. The overall shape follows the expected morning rush pattern: density rises from a low free-flow baseline around 06:00, peaks near 08:30–09:00, and then falls as demand subsides. A closer inspection of the mean trace suggests a slight sub-hourly modulation during the plateau phase, with density tending to be marginally higher near the quarter-hour marks. One plausible explanation is commuter scheduling pressure: departures clustering around fixed work start times may produce a weak periodic pulse in network demand with a 30-minute period. The effect is subtle, but it motivates the oscillatory inflow scenario explored in Section 6.2.

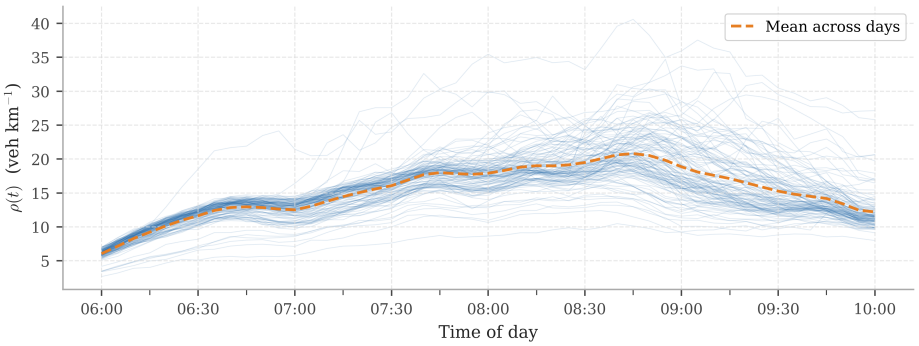


Figure 3.4: Network density $\rho(t)$ as a function of time of day. Faint blue traces show individual weekday mornings; the orange line is the mean across all 124 days. A systematic 30-minute oscillation is visible in the mean during the plateau phase (approximately 07:00–09:00), with peaks near the quarter-hour marks. Density is averaged over 5-minute periods.

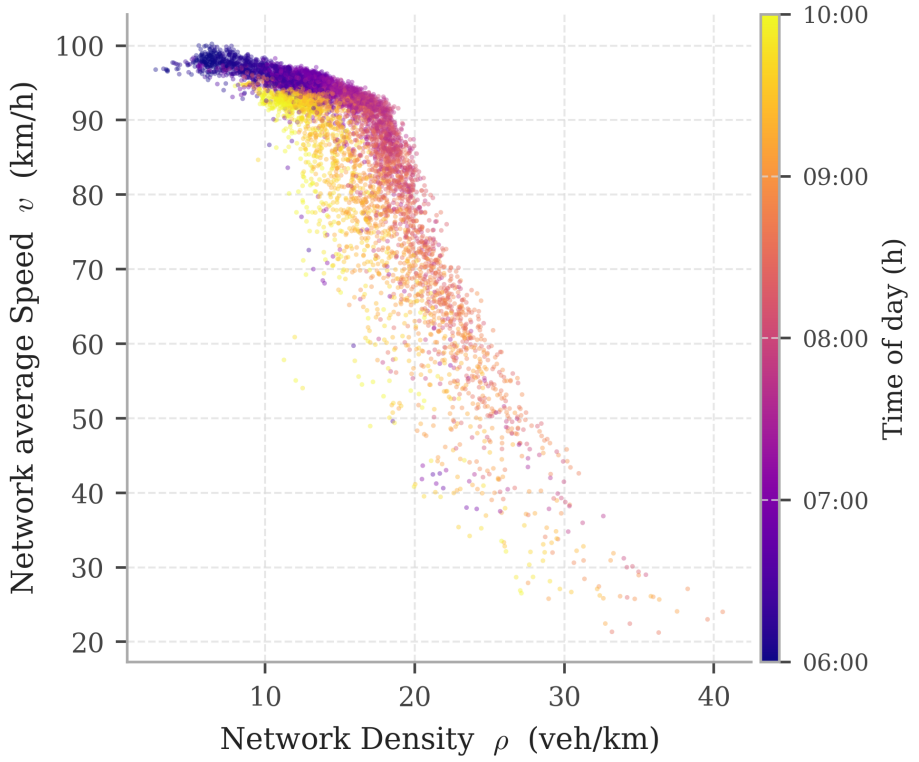


Figure 3.5: Speed–density scatter for all weekday mornings (excluding holidays listed in Appendix A) on the A10, 06:00–10:00, coloured by time of day from 06:00 (dark) to 10:00 (light). Each point represents one 5-minute average.

Figure 3.5 shows the v–ρ scatter for all 124 weekday mornings, with each point representing a 5-minute average coloured by time of day. It shows two features clearly. The colour pattern traces a clockwise loop in the v–ρ plane: darker points (early morning) lie at higher

speeds than lighter points (late morning) at the same density, confirming that loading and recovery follow distinct branches. At the same time, there is a clear density level around 17–19 veh km^{-1} beyond which speed drops sharply. This abrupt transition is consistent with the critical density threshold documented in the highway traffic literature [Kerner, 2004; Gayah and Daganzo, 2011]: below a certain demand level the network remains in a homogeneous free-flow state, but once exceeded, bottlenecks activate and congestion can form.

The threshold suggests that the speed–density relation can be single-valued when not exceeding this critical density, since no shock waves are present in that regime. To verify this directly, we isolate observations that were recorded before the threshold was first crossed on each day. Formally, for each day d we define

$$t_d^* = \inf\{t : \rho_d(t) > \rho_{\text{crit}} = 17\}$$

and retain only observations with $t < t_d^*$. This truncation removes all post-threshold dynamics, including any recovery-phase observations, while preserving the early-morning loading phase that is common to all days. We use $\rho_{\text{crit}} = 17$ veh km^{-1} as the truncation threshold; the formal estimation of this value is carried out in Chapter 5.

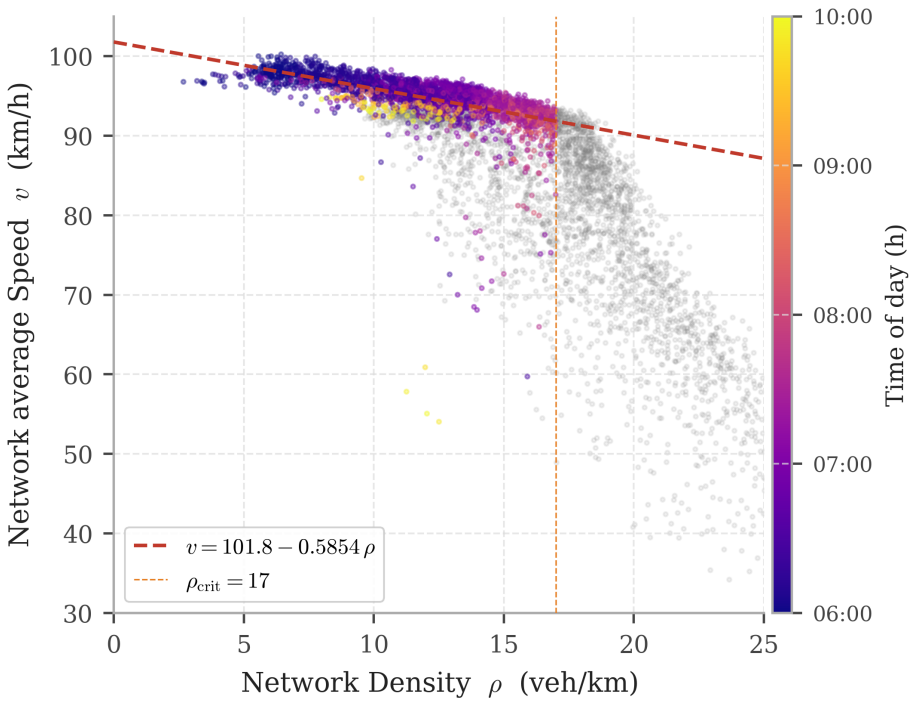


Figure 3.6: Speed–density relation for pre-critical observations only (coloured points), overlaid on the full dataset (light gray). Only observations recorded before the network density first exceeded $\rho_{\text{crit}} = 17$ veh km^{-1} are retained. The pre-critical points collapse onto a single well-defined curve close to the free-flow speed, well described by a linear fit $v = v_{\max} - \alpha\,\rho$ (red line). The multi-valued scatter visible in the gray background arises entirely from post-threshold dynamics.

The pre-critical observations collapse onto a single, well-defined curve with minimal vertical spread, close to the free-flow reference line (see Figure 3.6). To make the single-valued character of this regime precise, we fit a linear speed–density relation

$$v = v_{\max} - \alpha\,\rho, \tag{3.8}$$

to the pre-critical observations, shown as the red line in Figure 3.6. The tight agreement between the fit and the data confirms that a simple linear function adequately describes the pre-critical regime. This demonstrates that speed can be assumed uniquely determined by density prior to the first threshold crossing, and the free-flow speed–density relation is therefore treated as single-valued before exceeding ρ_{crit}.

Modelling Assumption #1

Single-valued free-flow regime. Before exceeding ρ_{crit}, the speed–density relation is single-valued: speed is uniquely determined by density, independently of the time of day. Any admissible model must reduce to a standard fundamental diagram in this regime, with no active congestion dynamics before threshold density is exceeded.

3.4 Delayed Density as a Predictor of Congestion

The v–ρ scatter of Figure 3.5 shows that congestion is not uniquely determined by current density. A natural first hypothesis is therefore whether it could instead be predicted from a time-delayed density signal, without introducing an independent state variable.

Specifically, if the congestion state at time t is correlated with the network density at some earlier time $t - T$, one could model congestion dynamics simply as $\xi(t) \approx h(\rho(t - T))$ for some function h and fixed delay T, without introducing an independent differential equation for $\xi(t)$. This section tests that hypothesis directly and explains why it was ultimately rejected.

To test this hypothesis, we compute the Pearson cross-correlation between each congestion measure and the lagged network density. The Pearson correlation coefficient measures the strength of the linear association between two variables, taking values in $[-1, 1]$ where ± 1 indicates a perfect linear relationship and 0 indicates no linear association. It is a natural choice here because we are asking a specific and testable question: is there a particular lag T at which density becomes a stronger linear predictor of congestion? A sharp peak in the correlation profile at some $T > 0$ would support the delayed-density hypothesis; a flat or monotonically declining profile would refute it.

For each of the three candidate congestion measures introduced in Section 3.1 — the density spread σ, the unweighted congestion fraction c_{unw}, and the density-weighted fraction c_{w} — we compute the Pearson cross-correlation with the lagged network density $\rho(t - T)$ as a function of the lag T. Denoting any one of the three candidate measures generically as $\xi(t)$, the cross-correlation at lag T is:

$$r(T) = \text{Corr}(\xi(t),\, \rho(t - T)), \tag{3.9}$$

evaluated by pooling all 30-second observations across all 124 weekday mornings. Lags are evaluated over the range $T \in [0, 15]$ min at 30-second resolution. The calculation of the Pearson correlation coefficient is given in Appendix B. The results are shown in Figure 3.7.

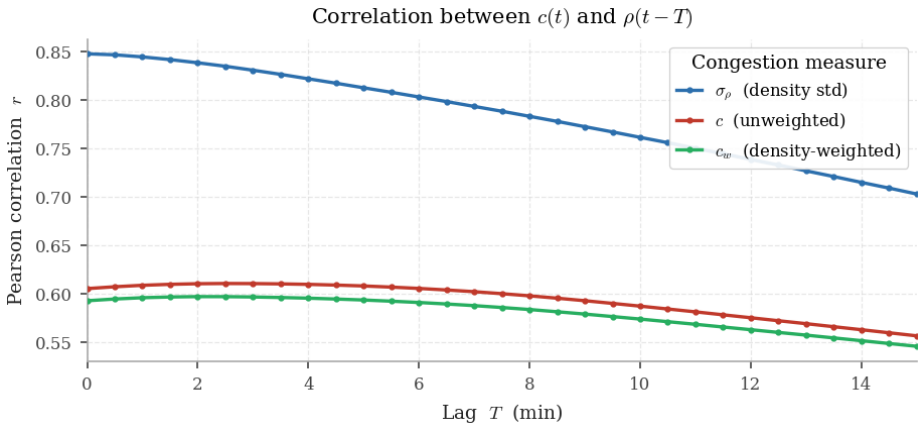


Figure 3.7: Pearson cross-correlation $r(T)$ between each congestion measure and the lagged network density $\rho(t-T)$, as a function of lag $T \in [0, 15]$ min, pooled across all 124 weekday mornings. The density spread σ achieves the highest contemporaneous correlation ($r \approx 0.85$ at $T = 0$) but declines monotonically with lag, peaking at zero delay. The unweighted fraction c_{unw} and the density-weighted fraction c_{w} both start near $r \approx 0.60$ and exhibit only a negligible, diffuse maximum near $T \approx 3$–5 min before declining. None of the three measures shows a sharp, dominant peak at any particular lag.

All three measures are positively correlated with lagged density across the full range of lags shown. The critical question is whether the correlation peaks sharply at a specific lag — a precondition for any fixed-delay model to be viable. The figure gives a clear negative answer. For σ, the correlation is highest at $T = 0$ and declines monotonically: introducing a lag makes the predictor strictly worse, not better. For c_{unw} and c_{w}, the correlation is nearly flat for small lags and shows at most a marginal, broadly distributed maximum near $T \approx 3$–5 min before similarly declining. In neither case does a single lag emerge as a reliable predictor that would justify modelling congestion as a deterministic function of delayed density.

The absence of a dominant lag implies that a model of the form $\xi(t) = f(\rho(t-T))$ cannot provide an adequate description of congestion dynamics on the A10. This modelling direction was therefore abandoned. The delay differential equation model produced before reaching this conclusion is documented in Appendix C for completeness.

3.5 Congestion Metrics Comparison

The cross-correlation analysis of Section 3.4 established that none of the three congestion measures can be adequately described as a deterministic function of delayed density. We therefore adopt an alternative approach and treat congestion as an *independent state variable* with its own dynamics, rather than a derived quantity. Three candidate variables were introduced in Section 3.1: the spatial spread of density σ, the unweighted congestion fraction c_{unw}, and the density-weighted congestion fraction c_{w}.

For a congestion variable ξ to serve as a useful second state variable in an extended bathtub model, it must satisfy three requirements.

1. The relation between v and ξ must be sufficiently sharp and monotone such that a velocity function $v(\rho, \xi)$ is well defined and identifiable from data. If v and ξ are weakly or noisily related, the congestion variable adds little explanatory power to the model.
2. The relation between ρ and ξ must exhibit clear threshold behaviour near ρ_{crit}, with $\xi \approx 0$ below threshold. This ensures that the measure captures genuine congestion onset. Its dynamics should moreover be sufficiently structured that they can be approximated by a tractable differential equation.
3. The dynamics of ξ must be asymmetric: congestion should build faster than it dissipates. This is motivated by the hysteresis observed in Figure 3.5 and is consistent with the physical mechanism of shock wave formation and dissolution documented in the traffic literature [Kerner, 2004]. A measure that does not exhibit this asymmetry in the loading/unloading analysis cannot serve as the basis for a hysteresis model.

Figure 3.8 shows the v–ξ and Figure 3.9 shows the ξ–ρ scatter for each of the three candidate variables.

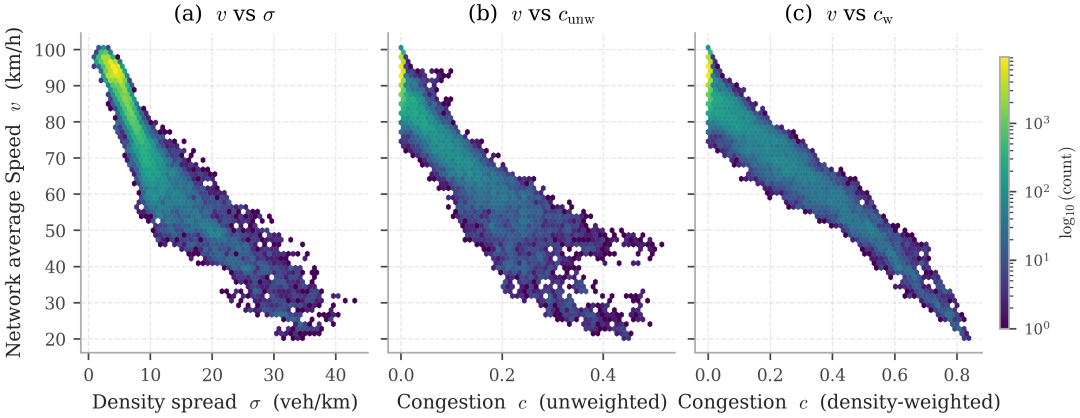


Figure 3.8: Speed–congestion scatter for the three candidate variables: density spread σ (left), unweighted congestion fraction c_{unw} (centre), and density-weighted congestion fraction c_{w} (right). In each panel the candidate variable is on the horizontal axis. Colour encodes the log-count of the number of observations within that specific hexagon. All 30 second observations are used from all weekdays excluding holidays.

In the v–σ panel, speed decreases with increasing σ. The relation is non-linear and shows considerable scatter at higher values of σ. The σ–ρ panel shows that σ begins rising from low densities, well below ρ_{crit}, with no pronounced change in slope around the threshold density.

The v–c_{unw} panel shows an approximately linear relation: speed decreases steeply and approximately linearly with increasing c_{unw}. Scatter increases for increasing congestion. This increasing scatter might be caused by the measure not taking into account the number of vehicles per link, so quiet links are weighted the same as dense links. The c_{unw}–ρ panel shows zero or near-zero values for low values of density. Congestion rises predominantly above the threshold density.

The v–c_{w} panel shows a sharp and mostly linear relation. This relation is more well-defined than for the other congestion measures. The c_{w}–ρ panel shows a similar situation to the

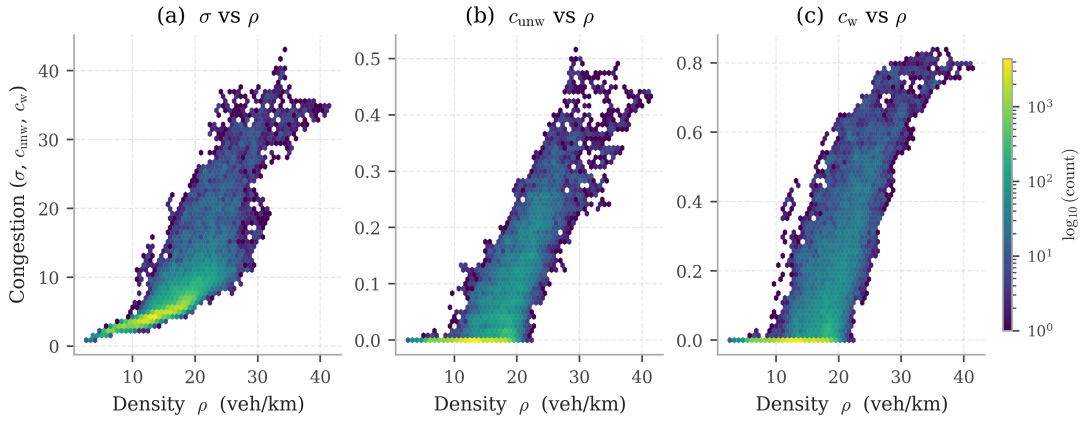


Figure 3.9: Congestion–density scatter in hexagonal bins for the three candidate variables: density spread σ (left), unweighted congestion fraction c_{unw} (centre), and density-weighted congestion fraction c_{w} (right). In each panel density is on the horizontal axis. Colour encodes the log-count of the number of observations within that specific hexagon. All 30 second observations are used from all weekdays excluding holidays.

unweighted congestion, with the additional feature of saturation at high densities, where it deviates from the mostly linear increase of congestion relative to density.

Both c_{unw} and c_{w} satisfy Requirement 1: they show a sharp v–ξ relation. The density spread σ does not satisfy Requirement 2 due to its non-zero values at low densities. For c_{unw} and c_{w}, the ξ–ρ panels suggest near-zero values at low densities, though this is partly obscured by the pooled loading and unloading observations. The remaining requirements are evaluated in the following section, where the two phases are separated.

3.6 Asymmetric Dynamics: Loading vs. Unloading

The v–ρ scatter in Section 3.3 already showed that the loading and unloading branches are distinct: the network speed is lower during recovery than during build-up at the same density. If this asymmetry originates in the congestion variable, it should be directly visible in the ρ–ξ plane: for the same network density, congestion should be higher during unloading than during loading. Splitting the loading and unloading might give more insight.

Each weekday morning, excluding holidays, is split into two non-overlapping phases at the time of the daily density maximum. Let t_d^{peak} denote the time at which $\rho(t)$ reaches its maximum on day d. Observations with $t \leq t_d^{\text{peak}}$ are assigned to the *loading phase*; observations with $t > t_d^{\text{peak}}$ are assigned to the *unloading phase*.

Figure 3.10 shows hexagon bin plots of the ρ–ξ relation, split by phase (columns) and congestion definition (rows). Several features are worth noting.

The top row of Figure 3.10 shows the ρ–σ relation during loading and unloading. Unlike the congestion fraction measures discussed below, σ begins rising from low densities, well below any identifiable ρ_{crit}. There is a small but not pronounced threshold visible. This confirms

the finding from Section 3.5: a model built on σ would require congestion dynamics to be active even during free-flow, or would need an explicit correction term to account for this baseline spread. This is undesirable from a modelling perspective and disqualifies σ as the dynamic congestion variable. The remaining discussion therefore focuses on c_{unw} and c_{w}.

In both loading panels for c_{unw} and c_{w} (middle and bottom rows, left column), the congestion fraction is zero or near-zero for densities below approximately $\rho_{\text{crit}} \approx 17$ veh km^{-1}. Above this level, congestion rises steeply and approximately proportionally with excess density. The main cluster of observations follows a single, well-defined branch with a steep slope, consistent with the threshold mechanism that is inactive in free-flow and activates sharply upon reaching the critical density.

Modelling Assumption #2

Congestion onset at a critical density. During the loading phase, both c_{unw} and c_{w} are negligible below ρ_{crit} and rise steeply above it. Any admissible model must incorporate a critical density ρ_{crit} at which congestion is initiated, with congestion dynamics that are inactive before exceeding this threshold.

In the unloading panels (middle and bottom rows, right column), the picture changes considerably. Congestion values are systematically higher than their loading-phase counterparts at the same density. The distribution is wider and shifted upward: observations that would correspond to low or moderate congestion during loading are instead associated with high congestion during unloading. At lower densities, where loading-phase congestion was near zero, the unloading-phase panels still contain many observations with measurable congestion, reflecting the persistence of shock waves after the peak density has passed.

The dominant branch in the unloading panels has a shallower slope in the c–ρ plane than the loading branch. This is consistent with the following physical picture: as density decreases during recovery, congestion dissipates, but at a slower rate relative to density than it built up. The network sheds congestion more gradually than it accumulated it.

The asymmetry between the loading and unloading panels confirms the hypothesis at the beginning of this section: at any given density, the congestion state of the network depends on whether it is loading or recovering, not on the instantaneous density alone.

Modelling Assumption #3

Asymmetric recovery. During unloading, congestion is systematically higher at any given density than during loading, and dissipates at a slower rate relative to density than it accumulated. Consequently, network speed is lower during recovery than during build-up at the same density, producing the clockwise hysteresis loop. Any admissible model must incorporate direction-dependent congestion dynamics, with a build-up rate that exceeds the recovery rate.

3.7 Empirical Assumptions for Model Development

The analyses in this chapter have identified three structural properties of morning rush congestion on the A10 that are robust across the full six-month dataset. Each property

was stated as a modelling assumption at the point in the chapter where the evidence was presented. The three empirically concluded assumptions are summarised here.

Assumption 1 — Single-valued free-flow regime. Before exceeding ρ_{crit}, the speed–density relation is single-valued: speed is uniquely determined by density, independently of the time of day. Any admissible model must reduce to a standard fundamental diagram in this regime, with no active congestion dynamics before threshold is exceeded.

Assumption 2 — Congestion onset at a critical density. During the loading phase, congestion is negligible before exceeding ρ_{crit} and rises steeply above it. Any admissible model must incorporate a critical density ρ_{crit} at which congestion is initiated, with congestion dynamics that are inactive before exceeding this threshold.

Assumption 3 — Asymmetric recovery. During unloading, congestion is systematically higher at any given density than during loading, and dissipates at a slower rate relative to density than it accumulated. Consequently, network speed is lower during recovery than during build-up at the same density, producing the clockwise hysteresis loop. Any admissible model must incorporate direction-dependent congestion dynamics, with a build-up rate that exceeds the recovery rate.

In the next chapter, a two-state dynamical model is proposed that encodes each of these assumptions mathematically, extending the classical bathtub model with an independent congestion equation aimed at reproducing the observed hysteresis.

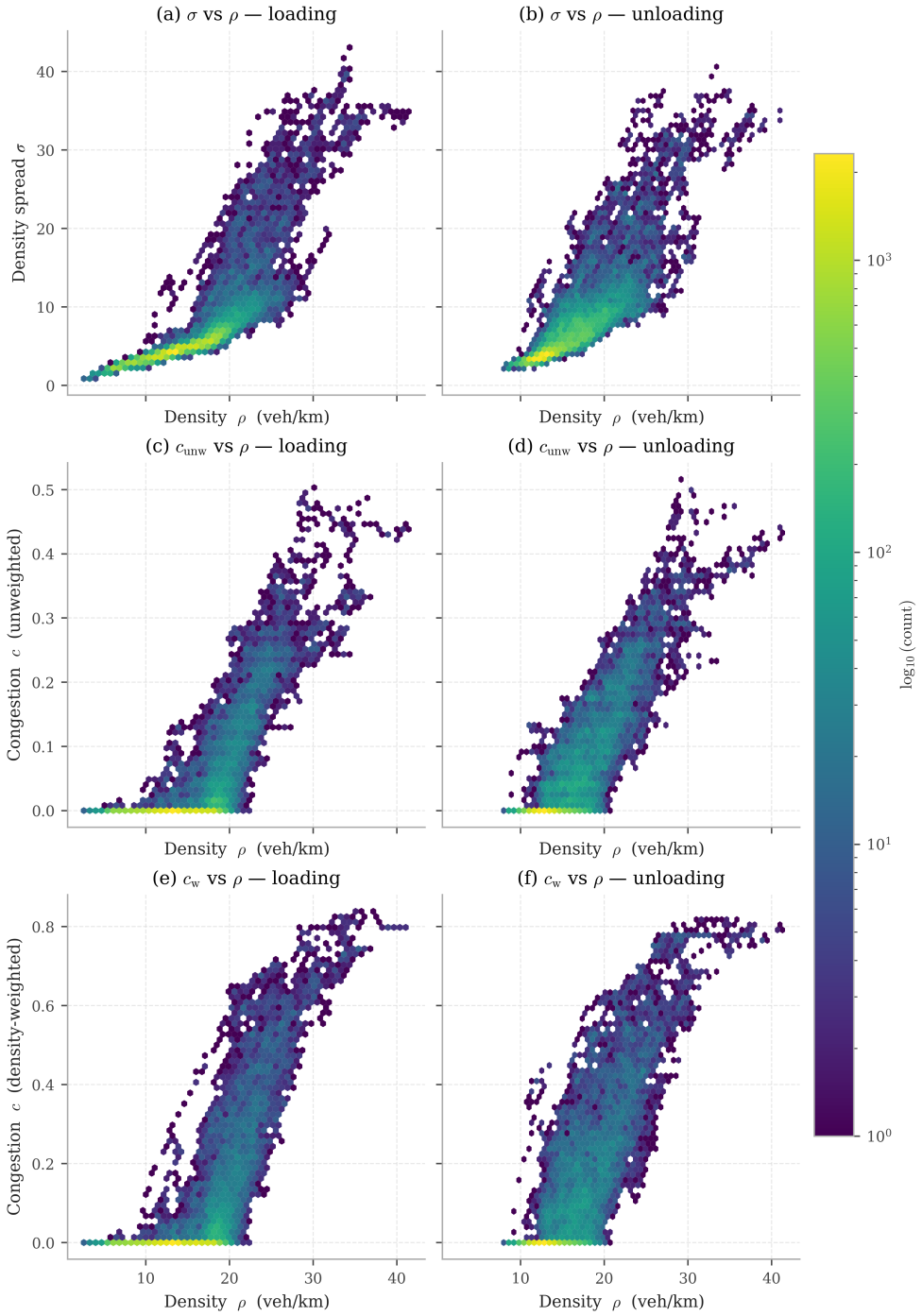


Figure 3.10: Hexagon bin plots of the congestion–density relation, split by phase (loading: left column; unloading: right column) and congestion definition (density spread σ: top row; unweighted fraction c_{unw}: middle row; density-weighted fraction c_w: bottom row). Colour encodes log-count of the number of observations within that specific hexagon. All 30 second observations are used from all weekdays excluding holidays.

4 Model Formulation

The model developed in this chapter contains two state variables, two initial conditions, eight parameters and one input function, summarised in Tables 4.1 and 4.2.

Table 4.1: Model variables, their descriptions, units, and roles in the dynamical system. The two state variables $\rho(t)$ and $c(t)$ are governed by differential equations; $v(t)$ is derived algebraically from the state variables; and $f(t)$ is an exogenous input.

Symbol	Description	Unit	Type
t	Time	h	Independent variable
$\rho(t)$	Network density	$\mathrm{veh\,km^{-1}}$	State variable
$c(t)$	Congestion level	–	State variable
$v(t)$	Average network speed	$\mathrm{km\,h^{-1}}$	Derived (Equation 4.4)
$f(t)$	Inflow per unit length	$\mathrm{veh\,km^{-1}\,h^{-1}}$	Forcing term (input)

Table 4.2: Model parameters, their descriptions, units, and the equation in which they appear.

Symbol	Description	Unit	Equation
$v_{\max}$	Free-flow speed	$\mathrm{km\,h^{-1}}$	Extended MFD (4.4)
α	Density sensitivity	$\mathrm{km^2\,veh^{-1}\,h^{-1}}$	Extended MFD (4.4)
β	Congestion sensitivity	$\mathrm{km\,h^{-1}}$	Extended MFD (4.4)
ρ_{crit}	Critical density	$\mathrm{veh\,km^{-1}}$	Congestion Dynamics (4.6)
γ	Congestion build-up rate	$\mathrm{km\,veh^{-1}}$	Congestion Dynamics (4.6)
η	Congestion recovery rate	$\mathrm{km\,veh^{-1}}$	Congestion Dynamics (4.6)
B	Average trip length	km	Conservation of Cars (4.3)
f_{crit}	Critical speed threshold	–	Congestion Definition (3.6), (3.7)
$\rho_0 = \rho(0)$	Density initial condition	$\mathrm{veh\,km^{-1}}$	
$c_0 = c(0)$	Congestion initial condition	–	

4.1 Modelling Philosophy and State Variables

The classical bathtub model describes the network by a single scalar state variable, the network density $\rho(t)$. The average speed is then a deterministic function of density alone, $v = V(\rho)$. As argued in Chapter 2, this single-state structure excludes hysteresis: if v is uniquely determined by ρ, the loading and unloading branches must coincide.

Chapter 3 confirmed this failure empirically and, crucially, provided guidance on what must replace it. The cross-correlation analysis of Section 3.4 ruled out models in which congestion

is predicted by a fixed lag of density, based on A10 highway data introduced in Section 3.1. The comparison of Section 3.5 and the loading/unloading analysis of Section 3.6 showed that the congestion fraction $c(t)$ captures threshold-like onset during loading and persistent elevated values during recovery — properties that density alone does not.

This motivates a two-state formulation. We describe the network by the pair $(\rho(t), c(t))$, where

- $\rho(t)$ [veh km^{-1}] is the **network density**: the spatially averaged number of vehicles per kilometre of road, per lane. It represents the overall loading of the network at each instant.
- $c(t) \in [0, 1]$ (dimensionless) is the **congestion level**: the spatial extent of congestion in the network, independently of the current density. A value of $c = 0$ indicates a fully free-flowing network, in which no road segment is operating below the critical speed. A value of $c = 1$ indicates that the entire network is congested. Intermediate values reflect the fraction of the network that is currently experiencing congested conditions. It is weighted by vehicles or road segments, depending on the definition chosen.

Given the comparison of congestion measures in Chapter 3, two possible congestion definitions are

1. **Unweighted congestion fraction** $c_{\text{unw}}(t)$: the fraction of the network currently operating below a critical speed threshold. This definition answers the question: what fraction of road segments are currently congested?
2. **Density-weighted congestion fraction** $c_{\text{w}}(t)$: the density-weighted fraction of the network operating below the critical speed. This definition answers the question: what fraction of vehicles are currently experiencing congestion?

Note that these definitions are mainly for comparison with and calibration to empirical loop-detector data. In the model itself there is no spatial component, so congestion $c(t)$ is a network-level term evolving according to a dynamical equation depending only on other network-level terms.

Definitions of their respective empirical measures are given in Section 3.1. In the derivation of these measures a critical speed threshold f_{crit} is needed, which enters the model as a parameter. The choice between the unweighted and density-weighted definitions of $c(t)$ is deferred to Chapter 5, where both are estimated and compared.

The key modelling idea is that $c(t)$ is *not* a deterministic function of $\rho(t)$. It has its own dynamics, governed by a differential equation that depends on the direction of density change. This is precisely what allows the same density to correspond to different congestion levels — and hence different speeds — depending on the entire density path instead of exclusively current density.

4.2 Density Dynamics

The evolution of $\rho(t)$ follows directly from conservation of vehicles. Vehicles enter the network at an exogenous inflow rate $f(t)$ [veh km^{-1} h^{-1}] and exit at an outflow rate $g(t)$,

giving

$$\dot{\rho}(t) = f(t) - g(t). \tag{4.1}$$

To close the system, an expression for $g(t)$ is required. We adopt the trip-completion closure of the Vickrey bathtub model [Vickrey, 1969]: the outflow equals the total vehicle-kilometres travelled per unit time divided by the average trip length B [km]:

$$g(t) = \frac{\rho(t)\, v(t)}{B}. \tag{4.2}$$

The product $\rho(t)\, v(t)$ is the network production: the number of vehicle-kilometres completed per unit time per unit length of road. Dividing by B converts this into a trip completion rate. Note that this outflow closure is a simplifying assumption: it assumes that the average trip length of vehicles in the network is constant over time. Substituting (4.2) into (4.1) then yields the density equation:

$$\dot{\rho}(t) = f(t) - \frac{\rho(t)\, v(\rho(t), c(t))}{B}. \tag{4.3}$$

The average trip length B is not identifiable from the available loop detector data, as neither the inflow $f(t)$ nor the outflow $g(t)$ is directly observed. A geometric estimate is derived in Chapter 5.

4.3 Extended MFD (Velocity Function)

The average network speed is modelled as a linear function of the two state variables:

$$v(\rho, c) = v_{\max} - \alpha\, \rho - \beta\, c. \tag{4.4}$$

This relation is referred to as the Extended MFD (EMFD). The density term follows the Greenshields model, the simplest linear speed–density relation. The congestion term is linear in c based on the empirical relationship shown in Figure 3.8. The parameters have the following interpretation.

- $v_{\max}$ [$\mathrm{km\,h^{-1}}$] is the free-flow speed: the network speed in the absence of vehicles and congestion.
- α [$\mathrm{km^2\,veh^{-1}\,h^{-1}}$] is the density sensitivity: the reduction in average speed per additional vehicle per kilometre. This term is present even in free-flow conditions, reflecting the mild friction between vehicles at sub-critical densities.
- β [$\mathrm{km\,h^{-1}}$] is the congestion sensitivity: the additional speed reduction associated with each unit increase in the congestion fraction. Since $c \in [0, 1]$, a value of $\beta = 100$ implies that a fully congested network ($c = 1$) travels 100 $\mathrm{km\,h^{-1}}$ slower than a fully free-flowing one at the same density.

The density term $\alpha\rho$ and the congestion term βc play qualitatively different roles. Below the critical density, congestion is absent ($c \approx 0$) and speed decreases gently with density through $\alpha\rho$ alone. Above the critical density, c grows rapidly and the βc term dominates, producing the sharp speed drop observed in Figure 3.5.

The admissible state space is restricted to $v(\rho, c) \geq 0$, ensuring physical meaningfulness.

4.4 Congestion Dynamics

The empirical analysis of Section 3.6 showed that the congestion–density relation is not the same during loading and unloading: congestion builds steeply above ρ_{crit} during loading, and dissipates more gradually during recovery. These observations motivate a direction-dependent evolution equation for $c(t)$.

Define the switching function $\Phi(\rho, \dot{\rho}, c)$ as follows:

$$\Phi(\rho, \dot{\rho}, c) = \begin{cases} 0, & \dot{\rho} > 0, \quad \rho < \rho_{\text{crit}}, \\ \gamma\, \dot{\rho}, & \dot{\rho} > 0, \quad \rho \geq \rho_{\text{crit}}, \\ \eta\, \dot{\rho}, & \dot{\rho} < 0 \text{ and } c > 0, \\ 0, & \dot{\rho} < 0 \text{ and } c = 0, \end{cases} \tag{4.5}$$

where $\gamma, \eta > 0$ are measured in [km veh^{-1}]. The congestion dynamics are

$$\dot{c}(t) = \Phi(\rho(t), \dot{\rho}(t), c(t)), \quad \text{with} \quad 0 \leq c(t) \leq 1. \tag{4.6}$$

Each case has a direct empirical motivation.

Free-flow loading ($\dot{\rho} > 0$, $\rho < \rho_{\text{crit}}$). Below the critical density, the network is spatially homogeneous and no shock waves are present (Constraint 1 of Section 3.7). Density increases do not generate additional congestion, so $\dot{c} = 0$.

Congested loading ($\dot{\rho} > 0$, $\rho \geq \rho_{\text{crit}}$). Above the threshold, shock waves nucleate and spread as density rises. Congestion grows proportionally to the rate of density increase, with build-up rate γ (Constraint 2).

Unloading ($\dot{\rho} < 0$). As density falls, shock waves gradually dissolve. Congestion decreases proportionally to the rate of density decrease, with recovery rate η. The case $c = 0$ is treated separately to prevent congestion from becoming negative: once the network has fully recovered, no further decrease is possible.

The empirical finding of Section 3.6 is that the loading branch has a steeper slope than the unloading branch in the c–ρ plane. This is captured directly by the condition $\gamma > \eta$: congestion builds faster than it dissipates.

4.5 Complete Model

The Dynamical System

Combining equations (4.3)–(4.6) with the linear velocity function (4.4) gives the complete two-state model (4.7), subject to $\rho(t) \geq 0$ and $0 \leq c(t) \leq 1$.

$$\begin{cases} \dot{\rho}(t) = f(t) - \dfrac{\rho(t)}{B}(v_{\max} - \alpha\,\rho(t) - \beta\,c(t))\,, \\ \dot{c}(t) = \begin{cases} 0, & \dot{\rho} > 0, \quad \rho < \rho_{\text{crit}}, \\ \gamma\,\dot{\rho}, & \dot{\rho} > 0, \quad \rho \geq \rho_{\text{crit}}, \\ \eta\,\dot{\rho}, & \dot{\rho} < 0 \text{ and } c > 0, \\ 0, & \dot{\rho} < 0 \text{ and } c = 0. \end{cases} \end{cases} \tag{4.7}$$

5 Parameter Calibration

The extended bathtub model including a congestion state variable, presented in Chapter 4 contains eight parameters. In order for the model to represent the actual A10 network, these parameters are calibrated based on the empirical data.

5.1 Velocity Function Parameter Calibration

The velocity function of the model expresses the average network speed as a linear function of the two state variables,

$$v(\rho, c) = v_{\max} - \alpha\,\rho - \beta\,c. \tag{5.1}$$

Three parameters must be calibrated from data: the free-flow speed $v_{\max}$ [km h^{-1}], the density sensitivity α [km^2 veh^{-1} h^{-1}], and the congestion sensitivity β [km h^{-1}]. For the congestion variable c, both definitions introduced in Section 3.1 are considered: the unweighted fraction c_{unw} and the density-weighted fraction c_{w}.

Throughout this chapter, a critical speed threshold of $f_{\text{crit}} = 0.5$ is used to classify detectors as congested: a sensor is congested when its measured speed falls below half its locally applicable speed limit. This parameter will not be calibrated or optimized. All 124 usable weekday mornings in the dataset are used in this chapter, aggregated into 5 minute observation blocks, in order to reduce the influence of transient states and measurement noise. This corresponds to $N = 6076$ data points.

5.1.1 Estimation Method

The parameters $v_{\max}$, α, and β are calibrated by minimising the residual sum of squares (RSS) between the observed network speed v_i and the model prediction $v(\rho_i, c_i)$ across all N observations:

$$\text{RSS} = \sum_{i=1}^{N} (v_i - v(\rho_i, c_i))^2 = \sum_{i=1}^{N} (v_i - v_{\max} + \alpha\,\rho_i + \beta\,c_i)^2. \tag{5.2}$$

Because the velocity function (5.1) is linear in its three parameters, the minimisation of (5.2) reduces to an ordinary least squares (OLS) problem with a closed-form solution. Stacking the N observations into a matrix

$$\mathbf{X} = \begin{pmatrix} 1 & \rho_1 & c_1 \\ \vdots & \vdots & \vdots \\ 1 & \rho_N & c_N \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix},$$

the parameter vector $\hat{\boldsymbol{p}} = (v_{\max}, \alpha, \beta)^{\top}$ that minimises the RSS is given by the normal equations

$$\hat{\boldsymbol{p}} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{v}. \tag{5.3}$$

The covariance matrix of the estimator is

$$\operatorname{Cov}(\hat{\boldsymbol{p}}) = \hat{\sigma}^2\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}, \qquad \hat{\sigma}^2 = \frac{\mathrm{RSS}}{N-3}, \tag{5.4}$$

where $N - 3$ is the number of degrees of freedom [Seber and Lee, 2003]. Standard errors for each parameter are the square roots of the diagonal entries of (5.4). In practice, (5.3) is evaluated via the `curve_fit` routine of SciPy [Virtanen et al., 2020], which solves the general nonlinear least-squares problem using the Levenberg–Marquardt algorithm [Levenberg, 1944; Marquardt, 1963]; since the model is linear in its parameters, this algorithm converges in a single step to the OLS solution above.

The quality of fit is summarised by two standard statistics. The root mean squared error $\mathrm{RMSE} = \sqrt{\mathrm{RSS}/N}$ gives the typical prediction error in $\mathrm{km\,h^{-1}}$, while the coefficient of determination

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}, \qquad \mathrm{TSS} = \sum_i (v_i - \bar{v})^2, \tag{5.5}$$

measures the fraction of total speed variance explained by the model.

5.1.2 Results

Table 5.1 reports the fitted parameters and fit statistics for both congestion definitions. Figure 5.1 shows the corresponding speed–density diagrams stratified by congestion level c.

Table 5.1: Velocity function parameter estimates for both congestion definitions at $f_{\mathrm{crit}} = 0.5$. Uncertainties are one standard error (1σ) from the covariance matrix returned by `curve_fit`.

	Unweighted c_{unw}		Density-weighted c_{w}	
Parameter	Estimate	$\pm 1\sigma$	Estimate	$\pm 1\sigma$
$v_{\max}$ [$\mathrm{km\,h^{-1}}$]	104.4	0.2	104.2	0.2
α [$\mathrm{km^2\,veh^{-1}\,h^{-1}}$]	0.90	0.02	0.87	0.01
β [$\mathrm{km\,h^{-1}}$]	134	1	67.0	0.3
R^2	0.926		0.960	
RMSE [$\mathrm{km\,h^{-1}}$]	3.82		2.79	

Both congestion definitions yield a well-fitting linear velocity function. The density-weighted variant, however, performs substantially better than the unweighted one: $R^2 = 0.960$ versus 0.926, and an RMSE of 2.79 $\mathrm{km\,h^{-1}}$ versus 3.82 $\mathrm{km\,h^{-1}}$.

The estimated free-flow speed of $v_{\max} \approx 104\ \mathrm{km\,h^{-1}}$ is slightly above the posted speed limit of 100 $\mathrm{km\,h^{-1}}$. This is not unusual: the speed limit is a legal maximum rather than a physical ceiling.

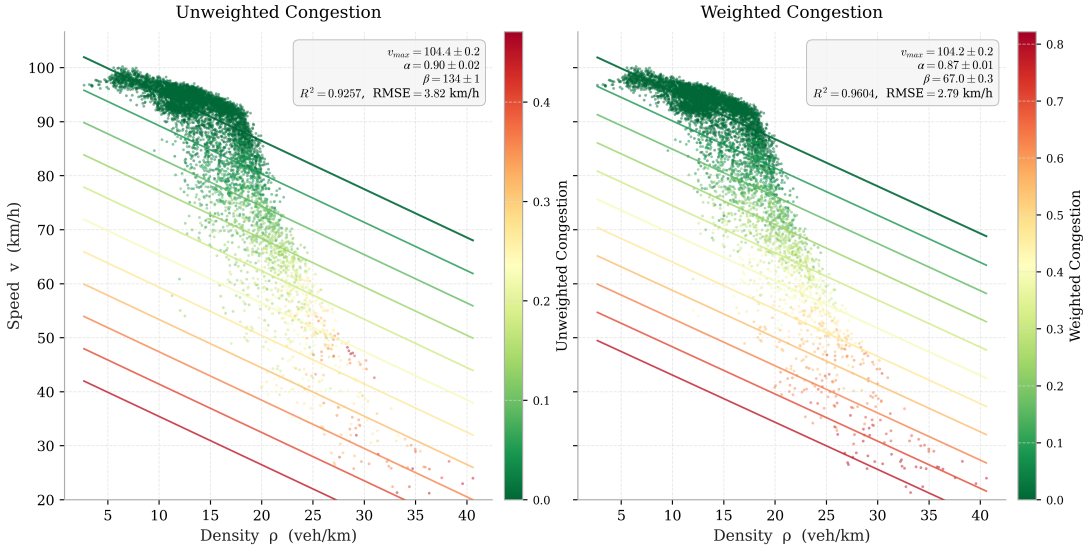


Figure 5.1: Velocity fit for $v(\rho, c) = v_{\max} - \alpha\rho - \beta c$ using the unweighted congestion fraction (left) and the density-weighted congestion fraction (right). Each point represents one 5 minute aggregated network-level observation. Colour encodes the congestion level c, from free-flow (green, $c \approx 0$) to heavily congested (red, high c). Lines show projections of the fitted velocity surface onto the speed–density plane for fixed values of c at equal spacing.

The density sensitivity α is small, reflecting that density alone has a small effect on speed under free-flow conditions. The dominant speed reduction mechanism is the congestion term. Since $c \in [0, 1]$, the congestion sensitivity $\beta = 67.0\ \mathrm{km\,h^{-1}}$ for the weighted definition implies that a 0.1 increase in congestion results in a 6.7 $\mathrm{km\,h^{-1}}$ decrease in network velocity. The much larger β for the unweighted definition (134 $\mathrm{km\,h^{-1}}$) partly compensates for the fact that c_{unw} tends to be smaller than c_{w} for the same traffic state.

5.2 Congestion Dynamics Parameter Calibration

The congestion dynamics in the model were described in Equation 4.6. This equation governs the time evolution of the congestion level $c(t)$ as a function of density. Three parameters must be calibrated: the critical density ρ_{crit} [veh $\mathrm{km^{-1}}$], the congestion build-up rate γ [km $\mathrm{veh^{-1}}$], and the congestion recovery rate η [km $\mathrm{veh^{-1}}$].

5.2.1 Simulation-Based Objective

Unlike the velocity function, the congestion dynamics cannot be calibrated by fitting a static relationship between observed variables. The congestion level $c(t)$ is a path-dependent quantity: its value at any given moment depends on the entire history of $\rho(t)$ up to that point, not only on the instantaneous density. A pair of observations (ρ_i, c_i) cannot therefore be treated as independent data points linked by a simple function.

The calibration must instead proceed by forward simulation. Given a candidate parameter vector $(\gamma, \eta, \rho_{\text{crit}})$ and the observed density time series ρ_t for a single day, the congestion dynamics are integrated forward in discrete time steps of 5 minutes according to the discretized version of Equation 4.6:

$$c_{t+1} = \begin{cases} c_t, & \Delta\rho_t > 0, \quad \rho_t < \rho_{\text{crit}}, \\ c_t + \gamma\,\Delta\rho_t, & \Delta\rho_t > 0, \quad \rho_t \geq \rho_{\text{crit}}, \\ \max\{0,\ c_t + \eta\,\Delta\rho_t\}, & \Delta\rho_t < 0, \\ c_t, & \Delta\rho_t = 0, \end{cases} \tag{5.6}$$

where $\Delta\rho_t = \rho_t - \rho_{t-1}$ is the observed density increment over one time step of $\Delta t = 5$ min, and the $\max\{0, \cdot\}$ enforces the non-negativity constraint on c. Starting from $c(0) = 0$ at 06:00, this produces a simulated congestion trajectory $\hat{c}(t)$ for that day. The procedure is repeated for every day in the dataset, and the total residual sum of squares between the simulated and observed congestion is used as the objective:

$$\mathcal{L}(\gamma, \eta, \rho_{\text{crit}}) = \sum_{d} \sum_{t} \left(c_t^d - \hat{c}_t^d(\gamma, \eta, \rho_{\text{crit}})\right)^2, \tag{5.7}$$

where the outer sum runs over all days d and the inner sum over all 5 minute time steps within the morning window of each day.

Because each evaluation of $\mathcal{L}$ requires a full simulation pass over the entire dataset, gradient-based methods are not straightforward to apply: the switching structure of Φ (Equation 4.5) renders the objective $\mathcal{L}$ non-smooth, and numerical gradients are unreliable near the switching boundaries. A derivative-free global optimisation method is therefore used instead.

5.2.2 Optimisation Method: Differential Evolution

The parameters are identified by minimising $\mathcal{L}$ using *differential evolution* [Storn, 1996], a population-based global optimisation algorithm that does not require gradient information. The algorithm maintains a population of candidate solutions and iteratively generates new candidates by combining existing ones through mutation and crossover. Each candidate is accepted if it improves on its predecessor, gradually driving the population towards the global minimum.

Differential evolution is well suited to this problem for two reasons. First, it does not assume smoothness or convexity of the objective, making it robust to the non-smooth switching structure of Φ. Second, its global search behaviour reduces the risk of converging to a local minimum, which is a concern when fitting simulation-based objectives that may have multiple regions of near-optimal parameter values.

The search is conducted over the parameter bounds $\gamma, \eta \in (0.001, 0.060)$ and $\rho_{\text{crit}} \in (15.0, 19.0)$ veh km^{-1}, which were chosen to encompass the range of physically plausible values. The algorithm is run with a population size of 15 per parameter, up to 2000 generations. The implementation uses the `differential_evolution` routine from SciPy [Virtanen et al., 2020]. Standard errors are estimated from the numerical Hessian of $\mathcal{L}$ evaluated at the optimum.

5.2.3 Results

Table 5.2 reports the fitted parameters and fit statistics for both congestion definitions. Figure 5.2 shows a scatter of the simulated versus observed congestion across all observations. Figure 5.3 shows the simulated and observed time series for the twelve most congested days.

Table 5.2: Congestion dynamics parameter estimates from differential evolution, for both congestion definitions. Uncertainties are one standard error (1σ) from the numerical Hessian at the optimum.

	Unweighted c_{unw}		Density-weighted c_{w}	
Parameter	Estimate	$\pm 1\sigma$	Estimate	$\pm 1\sigma$
γ [km veh^{-1}]	0.023	0.003	0.047	0.004
η [km veh^{-1}]	0.019	0.005	0.036	0.004
ρ_{crit} [veh km^{-1}]	17.0	0.1	17.21	0.04
R^2	0.806		0.769	
RMSE	0.033		0.075	

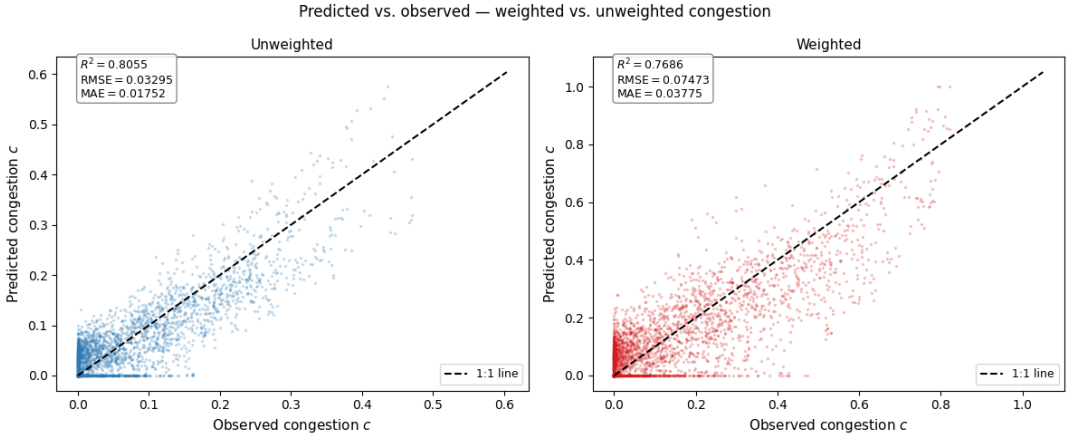


Figure 5.2: Simulated versus observed congestion for the unweighted (left) and density-weighted (right) congestion definitions. Each point represents one 5 minute aggregated network-level observation. The dashed line is the 1:1 reference. Points clustering tightly along this line indicate that the simulated congestion trajectory closely tracks the observed one.

Both fits explain a substantial fraction of the variance in observed congestion, with R^2 values of 0.806 and 0.769 for the unweighted and density-weighted definitions respectively. The unweighted model achieves a better fit on both metrics despite operating on a variable with a smaller range, suggesting that c_{unw} is more faithfully reproduced by the simple proportional dynamics of Φ.

The critical density is estimated at $\rho_{\text{crit}} \approx 17$ veh km^{-1} for both definitions, with tight standard

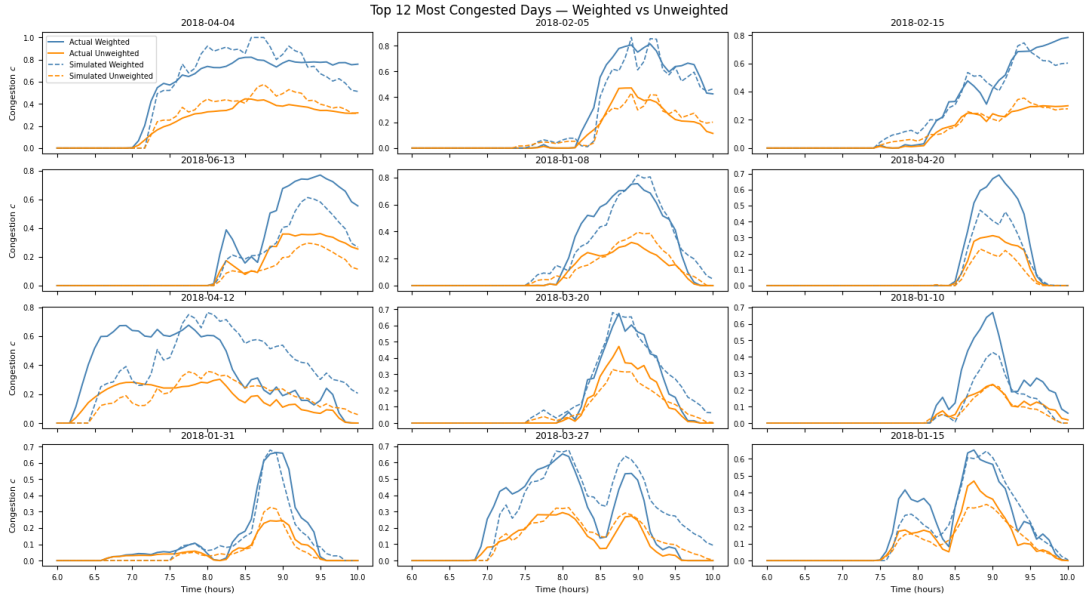


Figure 5.3: Observed and simulated congestion time series for the twelve days with the highest peak congestion. Solid lines show observed congestion; dashed lines show the simulation. Blue corresponds to the density-weighted measure c_{w}; orange to the unweighted measure c_{unw}.

errors. This is consistent with the threshold identified visually in the speed–density scatter of Section 3.3 and provides independent quantitative confirmation of that observation.

The central empirical finding of this section concerns the relationship between γ and η. For both congestion definitions, $\gamma > \eta$: congestion builds at a faster rate per unit density increase than it dissipates per unit density decrease. For the unweighted measure, $\gamma/\eta \approx 1.21$; for the density-weighted measure, $\gamma/\eta \approx 1.31$.

Although the unweighted model achieves a marginally better fit for the congestion dynamics ($R^2 = 0.806$ versus 0.769), the density-weighted definition is selected for the remainder of the analysis. The velocity function fit favours c_{w} more decisively ($R^2 = 0.960$ versus 0.926, RMSE 2.79 versus 3.82 $\mathrm{km\,h^{-1}}$), and the speed–density diagram of Figure 5.1 shows that the weighted congestion level produces a smoother and more coherent colour gradient across the speed–density plane: under c_{w}, colour transitions from free-flow to congested conditions progressively and consistently, whereas under c_{unw} the colours become disordered at high densities, suggesting that the unweighted measure is a noisier indicator of the true traffic state. The congestion dynamics fit is based on a first version of the dynamics equation, which leaves room for further refinement; the stronger performance of c_{w} in the velocity fit therefore justifies its selection as the preferred congestion measure.

5.3 Average Trip Length

The parameter B [km] is the average trip length of vehicles using the A10. It appears in the density dynamics equation as the factor converting network production into a trip-completion

rate:

$$\dot{\rho}(t) = f(t) - \frac{\rho(t)\, v(\rho(t), c(t))}{B}. \tag{5.8}$$

Identifying B from data would require knowledge of both the right-hand side of this equation and the inflow $f(t)$. The outflow term $\rho(t)\, v(t)/B$ can be computed once B is known, but $f(t)$ — the rate at which vehicles enter the A10 per unit length — is not observed in the loop detector dataset. Without $f(t)$, the conservation equation cannot be used to recover B, and no alternative identification strategy is available from the data at hand.

A geometric argument provides a rough estimate. The A10 is a ring road of total length approximately 32 km. A vehicle choosing to travel between two points on the ring will take the shorter of the two arcs. The maximum distance a vehicle can travel on the ring while still following the shorter arc is half the circumference, giving a maximum trip length of approximately 16 km. Under a uniform distribution of entry and exit points around the ring, the expected trip length of a vehicle on the shorter arc is half of this maximum, giving

$$B \approx \frac{16}{2} = 8 \text{ km}. \tag{5.9}$$

It should be interpreted as an order-of-magnitude estimate rather than a calibrated value: in reality, trip length distributions on ring roads are not uniform, and many vehicles use the A10 as part of a longer journey that extends beyond the ring.

Table 5.3 collects all calibrated parameters of the model. Values for c_{w} are used in the remainder of the analysis.

Table 5.3: Summary of all model parameters. Uncertainties are one standard error (1σ). The parameter B is not identifiable from the available data and is set by geometric argument.

Parameter	Description	c_{unw}	c_{w}
$v_{\max}$ [km h^{-1}]	Free-flow speed	104.4 ± 0.2	104.2 ± 0.2
α [km^2 veh^{-1} h^{-1}]	Density sensitivity	0.90 ± 0.02	0.87 ± 0.01
β [km h^{-1}]	Congestion sensitivity	134 ± 1	67.0 ± 0.3
ρ_{crit} [veh km^{-1}]	Critical density	17.0 ± 0.1	17.21 ± 0.04
γ [km veh^{-1}]	Build-up rate	0.023 ± 0.003	0.047 ± 0.004
η [km veh^{-1}]	Recovery rate	0.019 ± 0.005	0.036 ± 0.004
f_{crit} [–]	Critical speed threshold		0.5
B [km]	Average trip length		≈ 8

6 Model Analysis

The preceding chapters established the empirical properties of morning rush congestion on the A10 (Chapter 3), translated those properties into a two-state dynamical model (Chapter 4), and calibrated the model parameters from loop detector data (Chapter 5). This chapter addresses Sub-question 4: *Does this model reproduce the hysteresis loops observed in the A10 data?*

Section 6.1 demonstrates that the calibrated model produces a clockwise hysteresis loop under a trapezoidal morning rush inflow. Section 6.2 extends this to a demand profile that includes 30-minute periodic oscillations and examines how sub-hourly demand variation affects the hysteresis loop. Section 6.3 examines how the model behaviour changes as peak demand approaches the gridlock boundary.

All results in this chapter use the density-weighted calibration parameters from Table 5.3.

6.1 Forward Simulation under Trapezoidal Inflow

To demonstrate the model's behaviour under a controlled morning rush scenario, the network is loaded with a trapezoidal inflow profile $f(t)$ over the window 06:00–10:00. The trapezoid has four phases: a linear ramp-up from 06:00 to 07:00, a flat plateau from 07:00 to 08:30, a linear ramp-down from 08:30 to 09:30, and a return to the baseline level thereafter. This shape is a stylised representation of a morning rush hour: demand rises as commuters enter the network, sustains at a peak level during the core of the rush, and then falls as the peak period ends. The inflow profile is illustrated in Figure 6.1.

The complete model (4.7) is integrated numerically using a fixed-step forward Euler scheme with a time step of $\Delta t = \frac{1}{120}$ h (30 seconds), matching the temporal resolution of the empirical A10 data. At each step i, the state is updated as:

$$\rho_{i+1} = \max\left(\rho_i + \Delta t \left[f(t_i) - \frac{\rho_i\, v(\rho_i, c_i)}{B}\right],\ 0\right), \tag{6.1}$$

$$c_{i+1} = \mathrm{clip}(c_i + \Delta t\, \Phi(\rho_i,\, \dot{\rho}_i,\, c_i),\ 0,\ 1), \tag{6.2}$$

where $\dot{\rho}_i = f(t_i) - \rho_i\, v(\rho_i, c_i)/B$ is evaluated at the current state, and the max and clip operations enforce the physical constraints $\rho \geq 0$ and $c \in [0, 1]$.

The switching function Φ in Equation (4.5) is evaluated pointwise at each step using the sign of $\dot{\rho}_i$: no special event-detection is required because the regime boundaries ($\dot{\rho} = 0$ and $\rho = \rho_{\text{crit}}$) are naturally crossed between steps at the 30-second resolution. The simulation is initialised at $\rho(0) = \rho_0 = 6$ veh km^{-1} and $c(0) = 0$, corresponding to a free-flow steady state.

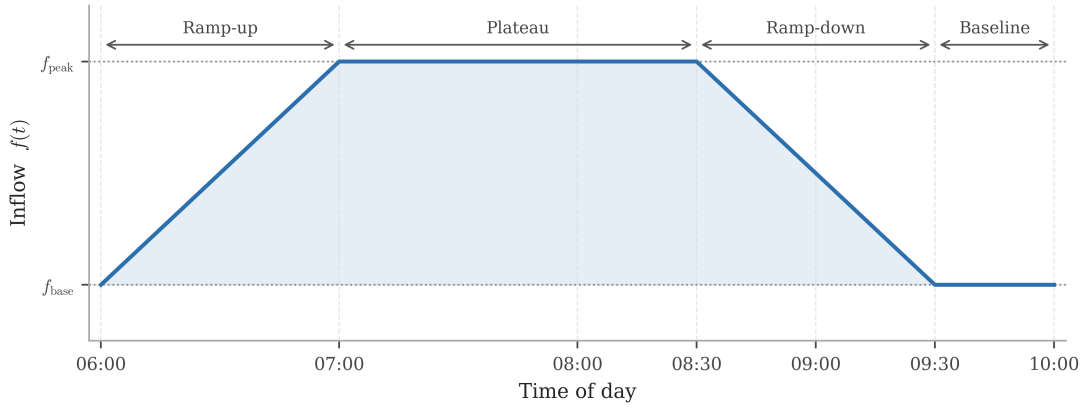


Figure 6.1: Schematic of the trapezoidal inflow profile $f(t)$ used in the forward simulation. The profile consists of four phases: a linear ramp-up from 06:00 to 07:00, a flat plateau from 07:00 to 08:30, a linear ramp-down from 08:30 to 09:30, and a return to the baseline level f_{base} thereafter.

Gridlock is defined as the first time step at which $v = 0$. When gridlock occurs, the simulation is terminated at that step.

The baseline inflow f_{base} is set to the equilibrium value consistent with the initial density, which on weekdays is around $\rho_0 = 6$ veh km^{-1}:

$$f_{\text{base}} = \frac{\rho_0}{B}\, v(\rho_0, 0) = \frac{\rho_0}{B}\,(v_{\max} - \alpha\,\rho_0) \approx 74.2 \text{ veh km}^{-1}\,\text{h}^{-1}, \tag{6.3}$$

so that $\dot{\rho}(0) = 0$, ensuring the network begins in a free-flow steady state.

The peak inflow f_{peak} is the single free parameter of the scenario. It controls how heavily the network is loaded during the plateau phase. To maximise the hysteresis area while avoiding gridlock, f_{peak} is set just below the gridlock boundary f^*:

$$f_{\text{peak}} = f^* - \varepsilon, \tag{6.4}$$

where ε is a small offset. This places the system at its operational limit: congestion builds as heavily as possible during the plateau phase, yet the network retains just enough outflow capacity to recover once demand falls.

The critical boundary f^* is identified by binary search, with f_{base} as a known safe lower bound. An analytical upper bound that guarantees gridlock is given by the maximum possible congestion-free outflow, obtained by maximising $\frac{\rho}{B}(v_{\max} - \alpha\rho)$ over ρ:

$$f_{\max} = \frac{v_{\max}^2}{4\alpha B} \approx 378 \text{ veh km}^{-1}\,\text{h}^{-1}. \tag{6.5}$$

Since no outflow can ever exceed $f_{\max}$, any inflow at this level guarantees gridlock. The binary search interval $[f_{\text{base}}, f_{\max}]$ is halved iteratively until the boundary is located to a precision of 0.1 veh km^{-1} h^{-1}, yielding

$$f^* \approx 198.7 \text{ veh km}^{-1}\,\text{h}^{-1}. \tag{6.6}$$

Setting $\varepsilon = 0.1$ veh km^{-1} h^{-1} gives $f_{\text{peak}} = 198.6$ veh km^{-1} h^{-1}, producing the most congested trajectory that still recovers. The sensitivity of the model to peak inflow is examined in Section 6.3.

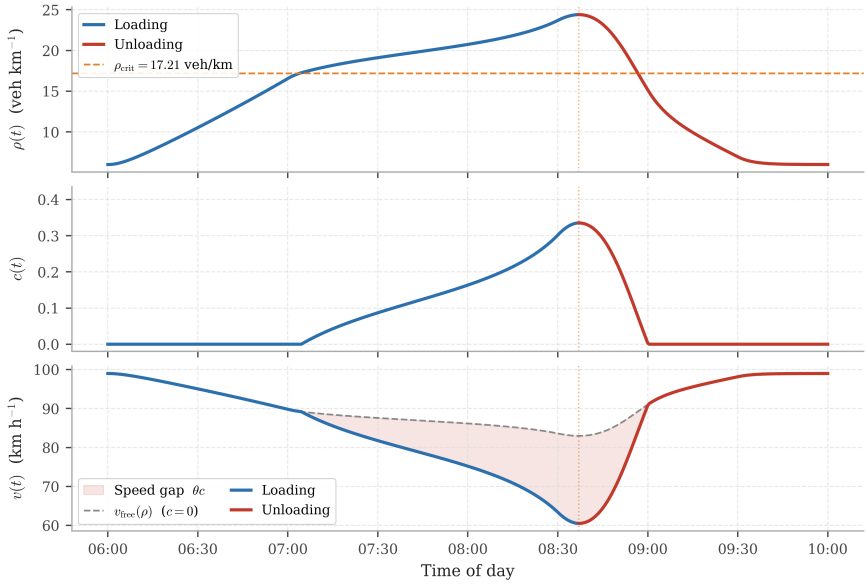


Figure 6.2: Simulated time series under the critical trapezoidal inflow ($f_{\text{peak}} = 198.6$ veh km^{-1} h^{-1}). *Top:* network density $\rho(t)$; the orange dashed line marks $\rho_{\text{crit}} = 17.21$ veh km^{-1}, and the dotted vertical line indicates the density peak. *Middle:* congestion level $c(t)$, which remains zero until ρ crosses ρ_{crit} and then builds continuously throughout the plateau and early ramp-down. *Bottom:* network speed $v(t)$ (solid) against the free-flow reference $v_{\text{free}}(\rho) = v_{\max} - \alpha\rho$ (dashed grey); the shaded region is the congestion-induced speed penalty $\beta\, c(t)$. Blue: loading phase ($\dot{\rho} > 0$); red: unloading phase ($\dot{\rho} < 0$).

The simulation, given the described inflow profile, produces two views of the model behaviour, shown in Figures 6.2 and 6.3. Figure 6.2 traces the three state variables as functions of time, revealing how the system evolves over the course of the morning rush. Figure 6.3 projects the same trajectory onto the (ρ, c) phase plane and the speed–density diagram, making the hysteresis loop directly visible.

Density crosses $\rho_{\text{crit}} = 17.21$ veh km^{-1} shortly after 07:00 as the ramp-up steepens. Below this threshold the congestion level c remains identically zero and speed follows the free-flow curve $v_{\text{free}}(\rho)$ exactly, so the trajectory in Figure 6.3 moves along the $c = 0$ axis. Once ρ_{crit} is exceeded, c begins to grow and speed falls below v_{free}. This behaviour directly reflects the switching structure of Φ in Equation (4.5): the congestion dynamics are inactive in free flow, as established empirically in Section 3.6.

During the loading phase $c(t)$ rises steadily at a rate proportional to $\gamma\,\dot{\rho}$. Once the density peaks and $\dot{\rho}$ turns negative, congestion begins to dissipate at a rate proportional to $\eta\,|\dot{\rho}|$. Because $\gamma > \eta$ (Table 5.3), recovery is slower relative to the density change than build-up. In the phase portrait this manifests as the loading branch rising more steeply than the unloading branch returns, leaving a non-zero enclosed area between them — the geometric signature of hysteresis.

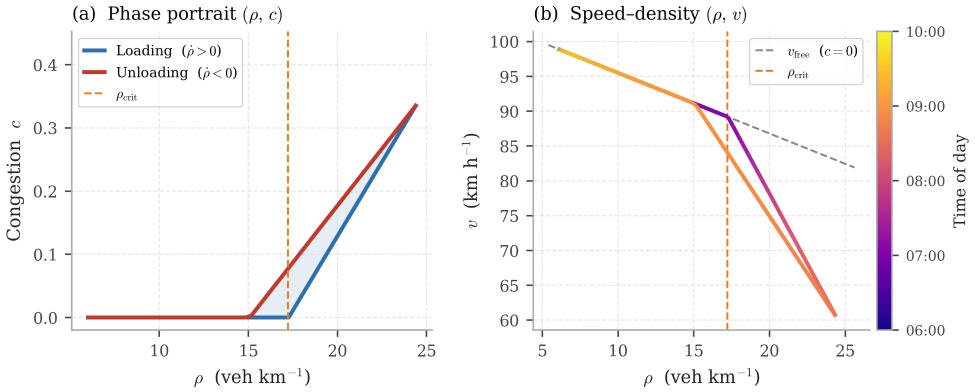


Figure 6.3: Hysteresis loops produced by the calibrated model under the critical trapezoidal inflow. *Left:* phase portrait in the (ρ, c) plane. The loading branch (blue) and unloading branch (red) trace distinct paths, with the loading branch rising more steeply. The shaded region between the two branches represents the area of the hysteresis loop. *Right:* speed–density diagram, with the trajectory coloured by time of day. The dashed grey line is the free-flow reference $v_{\text{free}}(\rho)$ $(c = 0)$. In both panels the orange dashed line marks ρ_{crit}.

The shaded region in the bottom panel of Figure 6.2 is the congestion-induced speed penalty $\beta\, c(t)$, which grows during loading and narrows during recovery. Because c is higher during unloading than loading at the same density, the unloading branch lies below the loading branch in the speed–density diagram of Figure 6.3, tracing a clockwise loop. The network recovers to free-flow speeds only once density has fallen well below ρ_{crit} and c has fully dissipated. This clockwise loop in the (ρ, v) plane is qualitatively identical to the empirical loops observed on the A10 and reproduced in Figure 1.1.

6.2 Effect of Demand Oscillations on Hysteresis

The trapezoidal inflow of Section 6.1 is a stylised scenario: demand rises and falls smoothly, with no variation. Empirical density time series on the A10, however, show a systematic sub-hourly oscillation: network density tends to peak near the quarter-hour marks (:15 and :45) and trough near the half-hour marks (:00 and :30), see Figure 3.4 in Section 3.3. A plausible explanation is scheduling pressure — commuters coordinating arrivals around fixed work start times cluster their departure choices, producing a periodic pulse in demand with a 30-minute period.

This section examines whether such demand oscillations materially affect the hysteresis produced by the model. The trapezoidal envelope is preserved, but the inflow is now modulated by a sinusoidal component:

$$f_{\text{osc}}(t) = f_{\text{trap}}(t)\left[1 + A\sin\left(\frac{2\pi t}{T_{\text{osc}}} + \varphi\right)\right], \tag{6.7}$$

where $T_{\text{osc}} = 0.5$ h is the oscillation period and the phase φ is chosen so that inflow peaks land at the quarter-hour marks :15 and :45. The amplitude $A = 0.05$ gives oscillatory peaks 5% above the envelope. Because these instantaneous peaks are higher than the smooth

trapezoidal case, the gridlock boundary for the envelope is reached at a lower value; a fresh binary search yields a critical envelope peak of $f^*_{\text{osc}} = 197.0$. With $\epsilon = 0.1$ this gives $f^{\text{osc}}_{\text{peak}} = 196.9$.

Figure 6.4 shows the resulting time series, with the trapezoidal baseline overlaid as dashed lines for reference. The 30-minute rhythm in the inflow (top panel) propagates into the density: rather than the smooth monotone rise of the baseline, $\rho(t)$ climbs in a staircase pattern, reaching a higher peak of approximately 25 veh km^{-1}. The congestion level $c(t)$ inherits this structure through a ratchet-like mechanism: each inflow pulse above ρ_{crit} increments c, while the subsequent dip only partially erodes it, driving the peak to $c \approx 0.39$ against ≈ 0.34 for the smooth baseline. Speed mirrors this in reverse, dipping lower during each pulse and recovering slightly during each trough.

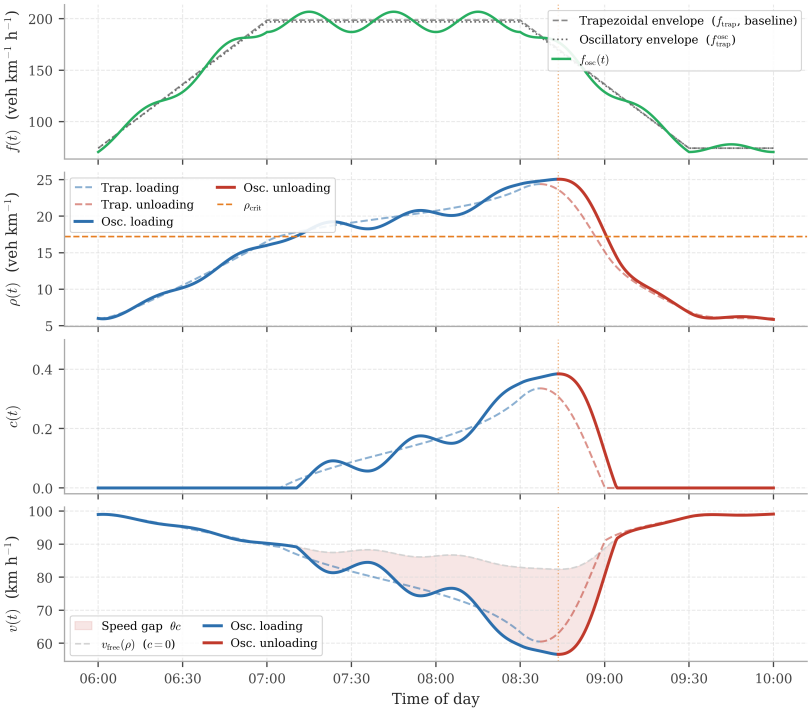


Figure 6.4: Simulated time series under the oscillatory inflow $f_{\text{osc}}(t)$ (Equation (6.7), $A = 0.05$, $T_{\text{osc}} = 0.5$ h); dashed lines show the trapezoidal baseline for reference. Blue: loading phase; red: unloading phase.

Figure 6.5 shows the phase portrait and speed–density diagram. The (ρ, c) loop has the same qualitative shape as the trapezoidal case: c remains zero below ρ_{crit} and the loading and unloading branches diverge above it, enclosing a finite area. The ratchet mechanism shifts the peak congestion upward relative to the smooth baseline, but the loop structure is otherwise preserved, confirming that the hysteresis mechanism is robust to sub-hourly demand variation.

In summary, the 30-minute demand oscillation amplifies hysteresis relative to the smooth trapezoidal scenario via a ratchet-like congestion build-up, while leaving the qualitative loop structure unchanged.

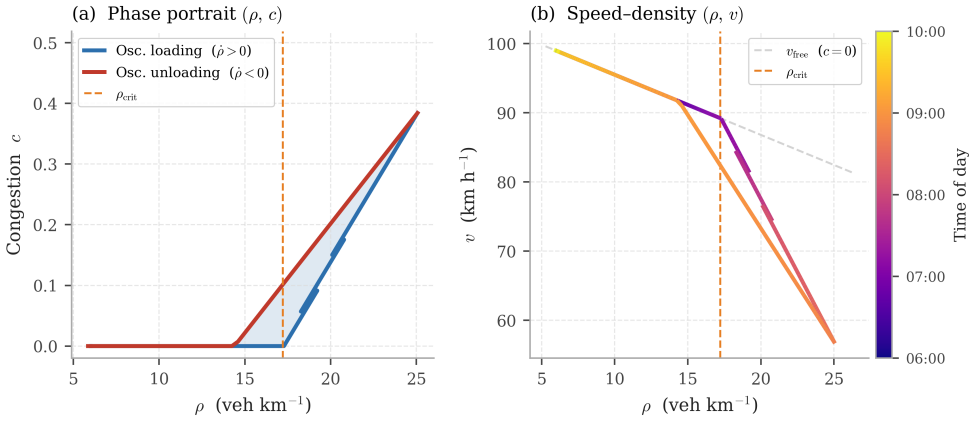


Figure 6.5: Hysteresis loops under the oscillatory inflow. *Left:* phase portrait (ρ, c). *Right:* speed–density (ρ, v), coloured by time of day.

6.3 Gridlock Sensitivity

This section examines how the model behaviour changes as the peak inflow approaches and crosses the critical threshold f_{peak} identified in Section 6.1. Four scenarios are simulated around f_{peak}: demand levels of -2%, -1%, f^* and $+1\%$ relative to the critical inflow.

Figure 6.6 shows the resulting trajectories in the (ρ, c) phase plane and the speed–density diagram. Figure 6.7 shows the corresponding speed time series $v(t)$.

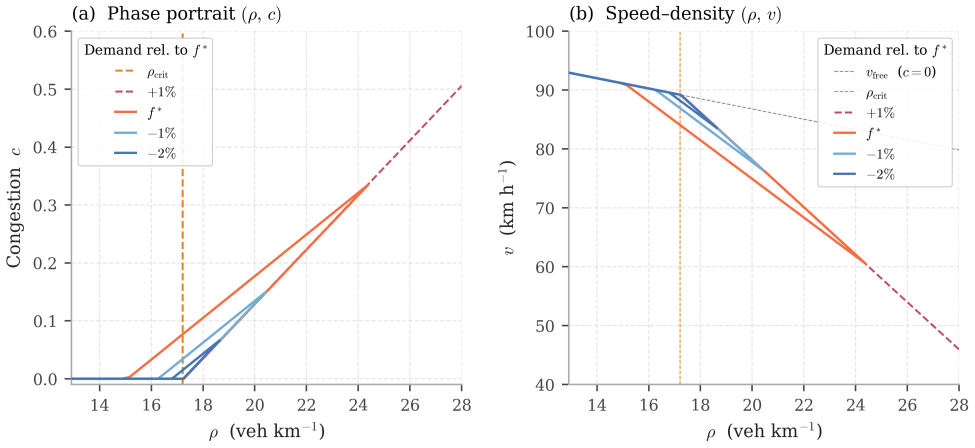


Figure 6.6: Sensitivity of hysteresis loops to peak inflow demand. Four scenarios around the critical inflow f^*_{peak}. Solid lines: sub-critical scenarios that recover; dashed lines: super-critical scenarios that reach gridlock ($v = 0$), terminated at that point and marked with a dot. *Left:* phase portrait (ρ, c). *Right:* speed–density diagram (ρ, v); the dashed grey line is the free-flow reference $v_{\text{free}}(\rho)$. Orange dashed vertical line: ρ_{crit}.

The difference between the -2% and $+1\%$ scenarios — a change of only 6 veh km^{-1} h^{-1} in peak inflow — separates full gridlock from a sub-critical scenario with moderate congestion

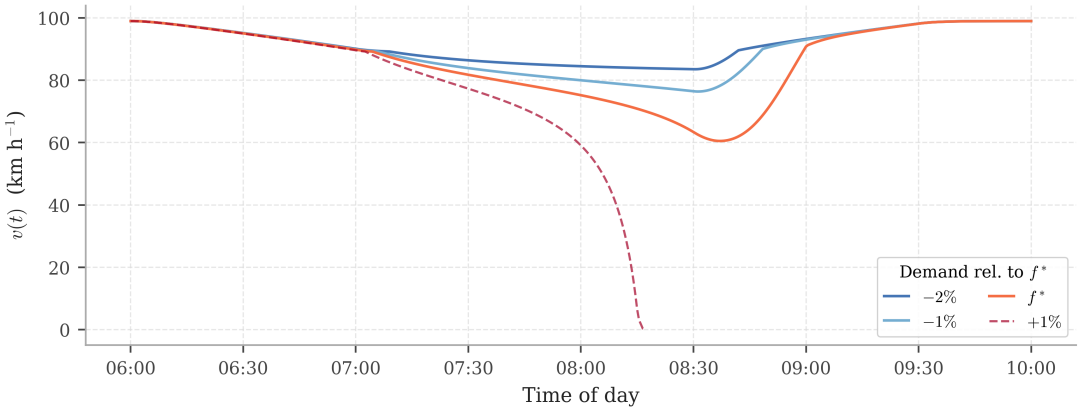


Figure 6.7: Speed time series $v(t)$ for the four gridlock sensitivity scenarios. Solid lines: sub-critical scenarios; dashed lines: super-critical scenarios, terminated at gridlock.

that fully recovers. In the phase portrait, the sub-critical scenario reaches a lower peak density and returns along a recovery branch, while the super-critical scenario continues loading until gridlock.

Among the three sub-critical scenarios, the area enclosed by the hysteresis loop in the (ρ, c) plane grows monotonically with f_{peak}. Higher demand drives density further above ρ_{crit}, allowing more congestion to accumulate during loading. Because the recovery rate η is fixed, the higher the peak value of c, the longer the unloading branch before c returns to zero and the wider the loop. The same effect is visible in the speed–density diagram: the gap between the loading and unloading branches in (ρ, v) space increases with demand.

The three simulations in this chapter collectively answer SQ4. The calibrated model reproduces a clockwise hysteresis loop in both the (ρ, c) and (ρ, v) planes under a artificial morning rush inflow, consistent with the empirical loops observed on the A10. This result is robust to sub-hourly demand oscillations, which amplify congestion through a ratchet mechanism without altering the qualitative loop structure. Near the gridlock boundary, the model exhibits a sharp transition between full recovery and gridlock, with hysteresis loop area growing monotonically with peak demand below the critical threshold.

7 Discussion and Conclusion

This thesis investigated whether morning rush hysteresis on the A10 can be captured by a macroscopic bathtub model extended with an explicit, path-dependent congestion variable. Empirical analysis of loop detector data established three modelling assumptions that motivated a two-state dynamical system, calibrated directly from loop detector data. Forward simulation confirmed that the calibrated model reproduces the clockwise hysteresis loops observed empirically. The following sections discuss how these contributions position this work relative to the existing literature, reflect on the limitations of the approach, and provide a detailed conclusion addressing the research questions.

7.1 Contribution and Positioning

The empirical analysis contributes a methodology for identifying the three modelling assumptions that any hysteresis-capturing bathtub model should satisfy. A key part of this analysis is the introduction of two novel congestion measures — an unweighted congestion fraction c_{unw} and a density-weighted variant c_{w} — alongside the density spread σ previously used by Knoop and Hoogendoorn [2013]. Both measures are designed to better satisfy the structural requirements of a hysteresis-capturing bathtub model, and the systematic comparison provides a transparent methodology for selecting an empirical congestion measure, with c_{w} identified as the most suitable basis for calibration.

The central modelling contribution is a bathtub-type model with a path-dependent congestion variable and explicit dynamics. It supplies the dynamic equation for heterogeneity evolution that the GMFD [Knoop and Hoogendoorn, 2013; Knoop et al., 2015] did not specify. The calibration procedure demonstrates that all model parameters can be identified from standard loop detector data alone. Together, these contributions offer a tractable path toward traffic models that can predict not just steady-state congestion, but the asymmetric dynamics of build-up and recovery that dominate the morning rush.

7.2 Limitations

Despite these contributions, the model rests on several simplifying assumptions that bound the scope of its conclusions.

First, the switching function Φ is specified to match the modelling assumptions of Chapter 3, but it does not derive from any micro-level description of how shock waves form and propagate.

Second, the model treats inflow $f(t)$ as an exogenous forcing term, with no feedback from network conditions to demand. In practice, drivers respond to congestion through route

choice, departure time adjustment, and modal shift, effectively capping inflow before gridlock is reached. The absence of this feedback is likely the reason the model exhibits unrealistically sharp sensitivity to gridlock near the critical threshold f^*, where a change of only a few vehicles per kilometre per hour separates full recovery from complete network breakdown.

Third, the outflow formulation inherits the fixed average trip length assumption of the Vickrey bathtub framework. By defining outflow as $\rho v / B$ with constant B, the model assumes the average trip length does not vary over time or with congestion conditions. This is a simplifying assumption that cannot be validated against observed inflow and outflow data independently, since $f(t)$ is unobserved.

Fourth, the congestion variable $c(t)$ depends on the threshold $f_{\text{crit}} = 0.5$, which is fixed rather than calibrated. This choice directly determines which detectors are classified as congested at any given time step, and therefore shapes the empirical data that the model is fitted to satisfy. The sensitivity of the results to this threshold has not been examined.

Fifth, the loop detector data used in this thesis were preprocessed using the Adaptive Smoothing Method [Treiber and Helbing, 2002], which reconstructs a smooth traffic state field from sparse raw observations. The congestion dynamics estimated here are therefore partly a property of this reconstruction rather than of the raw measurements, and results may differ if alternative preprocessing methods are applied.

Finally, the empirical analysis and calibration are based on a single corridor — the clockwise direction of the A10 ring road in Amsterdam — over a six-month window in 2018. Whether the three modelling assumptions and the calibrated parameter values generalise to other networks, directions, or time periods remains an open question.

7.3 Future Directions

The most important direction for future work is the measurement of the inflow $f(t)$. Access to reliable inflow data would unlock several improvements simultaneously: the average trip length B could be identified from the conservation equation rather than estimated geometrically; the model could be validated against independent flow observations; and the exogeneity assumption could be relaxed by allowing $f(t)$ to respond to current network conditions, which would also resolve the unrealistic gridlock sensitivity identified above.

A second priority is validation on other highway networks. The A10 is an urban ring road with specific geometric and demand characteristics that may not be representative of Dutch highways more broadly. Applying the methodology to other corridors would establish whether the three modelling assumptions — free-flow singularity, onset threshold, and asymmetric recovery — are a general feature of morning rush congestion or specific to the A10.

Third, the congestion dynamics function Φ introduced in this thesis is a first attempt. Deriving a similar formulation from microscopic traffic flow theory would place the model on a firmer theoretical foundation. Alternative functional forms for Φ could also be explored and compared against the current specification.

Two more possible extensions are worth noting. The critical speed threshold f_{crit}, currently fixed at 0.5, could be incorporated as a calibrated parameter, with a sensitivity analysis

establishing how strongly the results depend on this choice. Additionally, the linear velocity function could be replaced by alternative specifications to test whether the modelling conclusions are robust to this assumption.

Finally, the practical value of the model for real-time traffic management remains to be demonstrated. The bathtub framework was motivated partly by its potential as a prediction engine for control applications. A natural next step is to assess whether it produces actionable predictions under interventions such as ramp metering, variable speed limits, or congestion pricing.

7.4 Conclusion

The central research question of this thesis asked how morning rush hysteresis in highway networks can be captured by a bathtub model extended with a path-dependent congestion state variable, and what the structural and empirical properties of such a model are when calibrated to A10 Amsterdam data. The four sub-questions are answered below.

Empirical analysis of loop detector data from 124 weekday mornings on the A10 ring road established three structural model assumptions. Below a critical density of $\rho_{\text{crit}} \approx 17.2$ veh km^{-1}, the speed–density relation is single-valued and no congestion dynamics are active. Above this threshold, congestion onset is sharp and approximately proportional to excess density. During recovery, congestion dissipates more gradually than it accumulated, producing lower network speeds at the same density. Additionally, a systematic comparison of three candidate congestion measures — density spread σ, unweighted congestion fraction c_{unw}, and density-weighted fraction c_{w} — identified c_{w} as the most suitable empirical measure for characterising the network congestion state, outperforming both alternatives in fit to observed network speeds.

The three modelling assumptions motivated a two-state dynamical system in which an explicit congestion variable $c(t)$ evolves alongside network density $\rho(t)$. The congestion dynamics are governed by a switching function Φ that is inactive below ρ_{crit}, increases $c(t)$ at rate γ when density is increasing above the threshold, and decreases it at rate η during recovery. This construction captures all three empirical findings by design: the free-flow regime is preserved below threshold, onset is sharp and proportional, and the asymmetry $\gamma > \eta$ ensures that recovery is slower than build-up. The average network speed is modelled as a linear function of both $\rho(t)$ and $c(t)$, extending the classical MFD with an explicit congestion term.

The model parameters were identified from loop detector data. The velocity function achieved a strong fit to observed network speeds ($R^2 = 0.960$, RMSE = 2.79 km h^{-1}), confirming that the linear extended MFD captures the speed–density relationship well across both free-flow and congested conditions. Calibration of the congestion dynamics confirmed $\gamma/\eta \approx 1.31$: congestion builds approximately 31% faster than it dissipates. This asymmetry is consistent with the empirical observation and provides a concrete measure of the hysteresis phenomenon on the A10.

Forward simulation of the calibrated model under a artificial morning rush inflow reproduced the clockwise hysteresis loops observed empirically on the A10. The loading and unloading branches trace distinct paths in both the (ρ, c) and (ρ, v) planes, with the unloading branch lying at lower speeds for the same density. This result proved robust to sub-hourly demand oscillations, which increase congestion without altering the qualitative loop structure. Near

the gridlock boundary, the model reveals a sharp transition from full recovery to complete network breakdown.

Together, these findings demonstrate that morning rush hysteresis can be captured by expanding the classical bathtub model with a path-dependent congestion variable, calibrated from loop detector data. To the best of our knowledge, this is the first bathtub-type model to supply an explicit dynamic equation for the evolution of network heterogeneity, filling the gap left by the static GMFD. The model offers a tractable foundation for traffic management applications such as ramp metering, variable speed limits, and congestion pricing, that must predict not just steady-state congestion, but the asymmetric dynamics of build-up and recovery that dominate the morning rush.

A Holidays

Date	Holiday
2018-01-01	New Year's Day
2018-03-30	Good Friday
2018-04-01	Easter Sunday
2018-04-02	Easter Monday
2018-04-27	King's Day
2018-05-05	Liberation Day
2018-05-10	Ascension Day
2018-05-20	Whit Sunday
2018-05-21	Whit Monday

Table A.1: Public holidays in the Netherlands in the period January-June 2018.

B Pearson Cross-Correlation Calculation

The Pearson cross-correlation coefficient between a congestion measure $c(t)$ and the lagged network density $\rho(t-T)$ at lag T is defined as

$$r(T) = \frac{\sum_{i=1}^{n}\big(c(t_i)-\bar{c}\big)\big(\rho(t_i-T)-\bar{\rho}\big)}{\sqrt{\sum_{i=1}^{n}\big(c(t_i)-\bar{c}\big)^2}\sqrt{\sum_{i=1}^{n}\big(\rho(t_i-T)-\bar{\rho}\big)^2}}, \tag{B.1}$$

where $t_1, \ldots, t_n$ are the observation times, $\bar{c}$ and $\bar{\rho}$ are the sample means of c and ρ respectively, and n is the total number of observations. The coefficient takes values in $[-1, 1]$, with $r = 1$ indicating a perfect positive linear relationship, $r = -1$ a perfect negative linear relationship, and $r = 0$ no linear association.

In practice, the observations are discrete 30-second intervals. For a given lag $T = m \cdot \Delta t$ (with $\Delta t = 30$ s and $m = 0, 1, 2, \ldots$), the lagged pairs are formed by aligning $c(t_i)$ with $\rho(t_{i-m})$, discarding the first m observations for which the lagged density is unavailable. Equation (B.1) is then evaluated over the $n - m$ remaining pairs.

The pooling across mornings is performed by concatenating all valid $(c(t_i),\, \rho(t_{i-m}))$ pairs from all 124 weekday mornings into a single vector before computing $r(T)$. This ensures that the correlation estimate is based on a large sample ($n \approx 124 \times 480 = 59{,}520$ observations per lag value at $\Delta t = 30$ s over a four-hour morning window), reducing the influence of day-to-day variability on the estimate. The means $\bar{c}$ and $\bar{\rho}$ are therefore computed over the pooled sample rather than per morning.

Note that the Pearson coefficient measures only *linear* association. A low value of $r(T)$ does not exclude the possibility of a nonlinear relationship between $c(t)$ and $\rho(t-T)$. However, given that the primary purpose of this analysis is to assess whether a fixed-delay model is viable — which would require a strong and consistent predictive relationship — the linear correlation is a natural and conservative diagnostic: if even the linear association is weak or diffuse across lags, a fixed-delay model is unlikely to be useful regardless of functional form.

C Delay Differential Equation Model

As described in Section 3.4, an early modelling direction explored whether congestion could be represented as a deterministic function of a time-delayed density signal, without introducing an independent state variable. This appendix documents the model that was developed and analysed before this direction was abandoned in favour of the switching-function approach of Chapter 4.

Model Formulation

The starting point is the hypothesis that density spread $\sigma(t)$ is proportional to a delayed value of density,

$$\sigma(t) = \alpha\, \rho(t - T_d), \tag{C.1}$$

where $T_d > 0$ is a fixed delay and $\alpha > 0$ a proportionality constant. The physical motivation is that congestion does not form instantaneously: high density at time $t - T_d$ leads to increased heterogeneity at time t as shock waves nucleate and propagate. Combining this with a GMFD in which speed depends only on density spread,

$$v(\rho, \sigma) = v_{\max}\left(1 - \frac{\sigma}{\sigma_{\max}}\right) = v_{\max}\left(1 - \frac{\rho(t - T_d)}{\rho_{\max}}\right), \tag{C.2}$$

and substituting into the Vickrey outflow closure yields a single delay differential equation (DDE) for network density:

$$\frac{d\rho}{dt} = f(t) - \frac{\rho(t)}{B}\, v_{\max}\left(1 - \frac{\rho(t - T_d)}{\rho_{\max}}\right). \tag{C.3}$$

Non-dimensionalisation

To simplify analysis, we introduce dimensionless density $x = \rho/\rho_{\max}$, dimensionless influx $\mu = fB/(v_{\max}\rho_{\max})$, and rescale time so that the delay becomes unity, $t \leftarrow t/T_d$. This yields the dimensionless DDE

$$\frac{dx}{dt} = T\mu(t) - T\, x(t)\big[1 - x(t-1)\big], \tag{C.4}$$

where $T = T_d v_{\max}/B$ is a dimensionless parameter. An initial history $x(t) = \varphi(t)$ on $t \in [-1, 0]$ is required to complete the problem.

Equilibria

For constant influx $\mu(t) \equiv \mu$, equilibria $x(t) \equiv x_{\text{eq}}$ satisfy

$$0 = \mu - x_{\text{eq}}(1 - x_{\text{eq}}), \tag{C.5}$$

which gives

$$x_{\text{eq}} = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - 4\mu}. \tag{C.6}$$

Two equilibria exist when $0 < \mu < \frac{1}{4}$, exactly one when $\mu = \frac{1}{4}$, and none when $\mu > \frac{1}{4}$. The lower equilibrium $x_{\text{eq}}^{-} < \frac{1}{2}$ is physically meaningful as it corresponds to a sub-critical free-flow state.

Stability Analysis

Linearising around an equilibrium x_{eq} via the substitution $x(t) = x_{\text{eq}} + \varepsilon(t)$ gives

$$\frac{d\varepsilon}{dt} = a\,\varepsilon(t) + b\,\varepsilon(t-1), \tag{C.7}$$

where $a = -T(1 - x_{\text{eq}})$ and $b = T x_{\text{eq}}$. The characteristic equation is

$$\lambda = a + b\,e^{-\lambda}. \tag{C.8}$$

The roots of (C.8) can be expressed using the Lambert W function as $\lambda = a + W_n(b\,e^{-a})$, $n \in \mathbb{Z}$. Since all complex roots have real part strictly less than the unique real root λ_{real}, stability is determined by the sign of λ_{real} alone. One can show that $\lambda_{\text{real}} < 0$ if and only if $x_{\text{eq}} < \frac{1}{2}$. It follows that the lower equilibrium x_{eq}^{-} is asymptotically stable and the upper equilibrium x_{eq}^{+} is unstable for all $T > 0$.

Reason for Abandonment

Despite producing hysteresis under periodic inflow, the DDE model was ultimately abandoned for two reasons. First, the cross-correlation analysis of Section 3.4 showed that no single fixed delay T_d emerges as a reliable predictor of congestion across the 124 weekday mornings in the A10 dataset: the correlation between $\sigma(t)$ and $\rho(t - T)$ declines monotonically with lag rather than peaking at a characteristic value. Second, the delay formulation does not naturally encode the loading threshold ρ_{crit} identified empirically in Section 3.3, making it inconsistent with Assumption 2 of Section 3.7.

Bibliography

Bedoya-Maya, F., Calatayud, A., and González Mejía, V. (2022). Estimating the effect of road congestion on air quality in Latin America. *Transportation Research Part D: Transport and Environment*, 113:103510.

Blokker, B. B. V. P. B. and Blokker, B. B. V. P. B. (2013). Bouwen doen we voortaan buiten de A10. *NRC*.

Buisson, C. and Ladier, C. (2009). Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. *Transportation Research Record*, 2124(1):127–136.

Cassidy, M. J. (1998). Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological*, 32(1):49–59.

Cassidy, M. J., Jang, K., and Daganzo, C. F. (2011). Macroscopic Fundamental Diagrams for Freeway Networks: Theory and Observation. *Transportation Research Record*, 2260(1):8–15.

Christidis, P., Rivas, J. N. I., et al. (2012). Measuring road congestion. *Institute for Prospective Technological Studies (IPTS), European Commission Joint Research Centre. Retrieved from http://ipts. jrc. ec. euro pa. eu/publications/pub. cfm*.

Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.

De Palma, A. and Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6):1377–1399.

European Commission (2011). White paper: Roadmap to a single european transport area — towards a competitive and resource efficient transport system. Technical report, European Commission.

Gayah, V. V. and Daganzo, C. F. (2011). Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability. *Transportation Research Part B: Methodological*, 45(4):643–655.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.

Geroliminis, N. and Sun, J. (2011). Hysteresis Phenomena of a Macroscopic Fundamental Diagram in Freeway Networks. *Procedia - Social and Behavioral Sciences*, 17:213–228.

Godfrey, J. W. (1969). THE MECHANISM OF A ROAD NETWORK. *Traffic Engineering & Control*, 8(8).

Golob, T. F. and Regan, A. C. (2001). Impacts of highway congestion on freight operations: Perceptions of trucking industry managers. *Transportation Research Part A: Policy and Practice*, 35(7):577–599.

Greenshields, B. D., Bibbins, J. R., Channing, W. S., and Miller, H. H. (1935). A STUDY OF TRAFFIC CAPACITY. *Highway Research Board Proceedings*, 14.

Hugoniot, H. (1887). Memoir on the propagation of movements in bodies, especially perfect gases (first part). *J. de l'Ecole Polytechnique*, 57(3).

Jin, W.-L. (2020). Generalized bathtub model of network trip flows. *Transportation Research Part B: Methodological*, 136:138–157.

Kerner, B. S. (2004). *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Understanding Complex Systems. Springer, Berlin.

Knoop, V. L. and Hoogendoorn, S. P. (2013). Empirics of a generalized macroscopic fundamental diagram for urban freeways. *Transportation research record*, 2391(1):133–141.

Knoop, V. L., van Lint, H., and Hoogendoorn, S. P. (2015). Traffic dynamics: Its impact on the Macroscopic Fundamental Diagram. *Physica A: Statistical Mechanics and its Applications*, 438:236–250.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.

Li, D., Ranjitkar, P., and Ceder, A. A. (2014). Integrated Approach Combining Ramp Metering and Variable Speed Limits to Improve Motorway Performance. *Transportation Research Record*, 2470(1):86–94.

Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the royal society of london. series a. mathematical and physical sciences*, 229(1178):317–345.

Mahmassani, H. S., Saberi, M., and K., A. Z. (2013). Urban Network Gridlock: Theory, Characteristics, and Dynamics. *Procedia - Social and Behavioral Sciences*, 80:79–98.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441.

Nationaal Dataportaal Wegverkeer (2023). Dataformaten: DATEX2 v2.3 verkeersgegevens. NDW Technical Documentation. Accessed: May 5th 2026.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, 27(4):289–303.

Rankine, W. J. M. (1870). Xv. on the thermodynamic theory of waves of finite longitudinal disturbance. *Philosophical Transactions of the Royal Society of London*, 160:277–288.

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research*, 4(1):42–51.

Schreiter, T., van Lint, H., Treiber, M., and Hoogendoorn, S. (2010). Two fast implementations of the adaptive smoothing method used in highway traffic state estimation. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 1202–1208.

Seber, G. A. and Lee, A. J. (2003). *Linear regression analysis*. John Wiley & Sons.

Storn, R. (1996). On the usage of differential evolution for function optimization. In *Proceedings of North American fuzzy information processing*, pages 519–523. Ieee.

Treiber, M. and Helbing, D. (2002). Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooper@ tive Tr@ nsport@ tion Dyn@ mics*, 1(3):3–1.

Treiber, M., Kesting, A., and Thiemann, C. (2008). How much does traffic congestion increase fuel consumption and emissions? applying a fuel consumption model to the ngsim trajectory data. In *87th Annual Meeting of the Transportation Research Board, Washington, DC*, volume 71, pages 1–18.

Treiterer, J. and Myers, J. A. (1974). THE HYSTERESIS PHENOMENON IN TRAFFIC FLOW. In *Transportation and Traffic Theory, Proceedings*, volume 6.

Vickrey, W. (2019). Types of Congestion Pricing Models. *Economics of Transportation*, 20:100140.

Vickrey, W. S. (1969). Congestion Theory and Transport Investment. *The American Economic Review*, 59(2):251–260.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Xu, G., Zhang, P., Gayah, V. V., and Hu, X. (2023). Opposing Hysteresis Patterns in Flow and Outflow Macroscopic Fundamental Diagrams and Their Implications. *Transportation Research Record*, 2677(8):100–117.