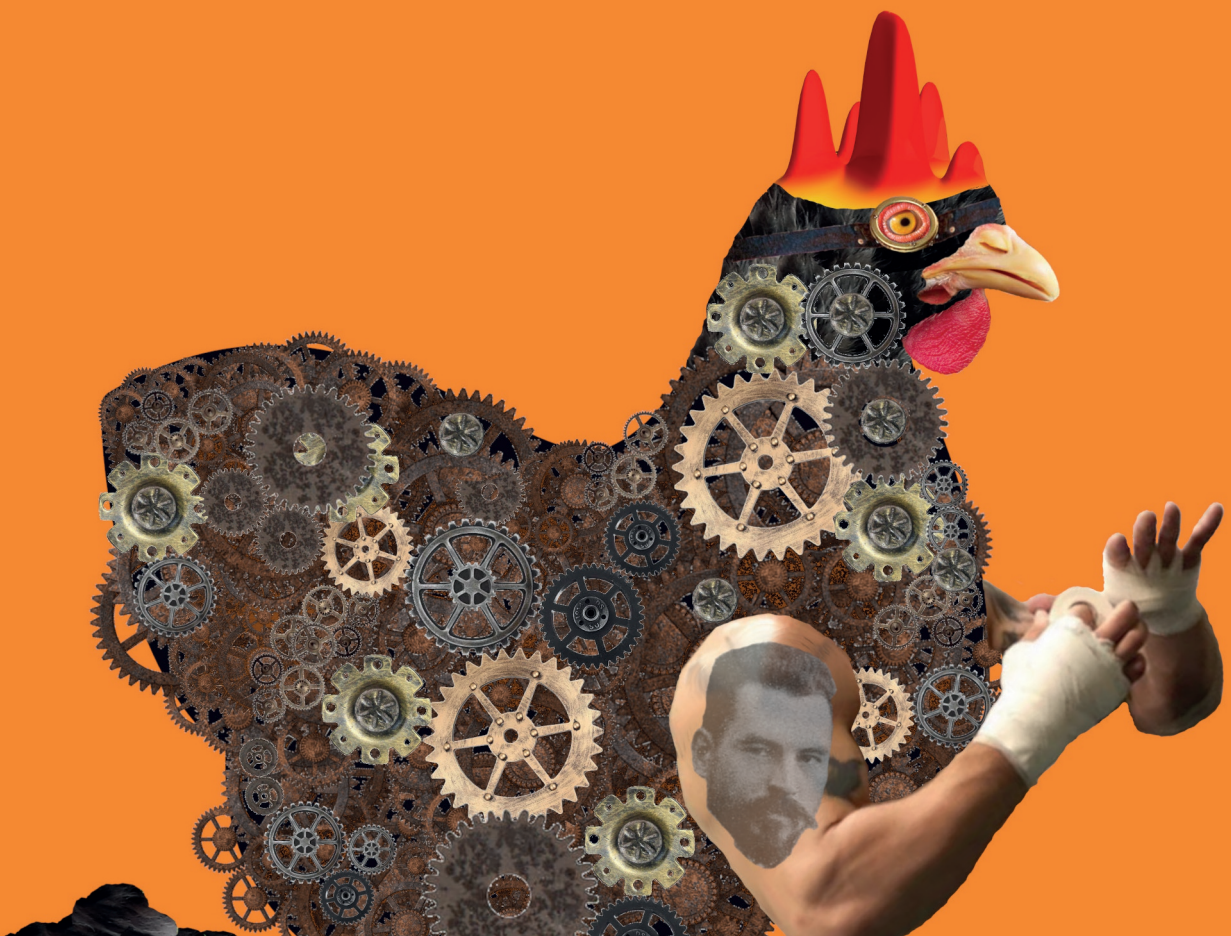
Transmission from the infectious material perspective using spatiotemporal modelling and experimental data

Anna Maria Gamża



Propositions

1. Macroscopic observations can inform about microscopic mechanisms when combined with mechanistic modelling.
   (this thesis)

2. Understanding mechanisms of environmental transmission requires data on transient states of the environmental load of infectious material.
   (this thesis)

3. Every model should be published with a data statement describing what type of data can be used to calibrate and validate it.

4. Contrary to popular belief, it is not Albert Einstein, but Marian Smoluchowski who explained the mechanism underlying Brownian motion.

5. The most important question in science is not “How?” but “Why?”.

6. Focusing on societal relevance in science hinders societal development.

7. Scientists should have managers just like athletes or artists do.

8. Encouraging people to follow their deepest desires is the mechanism underlying manipulation.

Propositions belonging to the thesis, entitled

Transmission from the infectious material perspective using spatiotemporal modelling and experimental data

Anna Maria Gamża

Transmission from the infectious material perspective using spatiotemporal modelling and experimental data

Anna Maria Gamża

Thesis committee

Promotor
Prof. Dr M.C.M. de Jong
Professor of Quantitative Veterinary Epidemiology Group
Wageningen University & Research

Co-promotor
Dr T. J. Hagenaars
Senior scientist, Epidemiology, Bio-informatics & Animal models
Wageningen University & Research

Other members
Dr S. Gubbins, The Pirbright Institute, Pirbright, United Kingdom
Dr A. Hemerik, Wageningen University & Research
Prof. Dr M.E.E. Kretzschmar, University Medical Center Utrecht
Dr F.C. Velkers, Utrecht University

This research was conducted under the auspices of the Graduate School Wageningen Institute of Animal Sciences (WIAS).

Transmission from the infectious material perspective using spatiotemporal modelling and experimental data

Anna Maria Gamża

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 3 April 2023
at 4 p.m. in the Omnia Auditorium.

Anna Maria Gamża
Transmission from the infectious material perspective using spatiotemporal modelling and experimental data,
161 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summary in English

ISBN: 978-94-6447-603-3
DOI: https://doi.org/10.18174/588327

Abstract

As most pathogens are transmitted through environment, a better understanding of the processes underlying environmental transmission is crucial to develop relevant intervention strategies targeted to these processes. In this thesis, I follow the development cycle of both modelling and experimental work designed to study mechanisms of environmental transmission from the infectious material perspective. In Chapter 1, the full modelling framework is described in detail, together with a brief description of the available data types and the model system we use in this thesis, being *Campylobacter jejuni (C. jejuni)* transmission in broilers. The parsimonious model we constructed describes mechanisms of pathogen transmission via the environment using only three parameters: the decay rate parameter (describing survival of infectious material in the environment), the transmission rate parameter (describing jointly shedding rate, absorption rate and probability of infection after absorption) and the diffusion coefficient (describing spatial dispersion of infectious material). In Chapter 2, we present the calibration and validation of the models, using a series of tailor-made experiments which studied *C. jejuni* transmission in broilers, in time and space. As our spatiotemporal model is fully identifiable with data we collected and it satisfactorily describes the experimental observations, we were able to obtain new insights into *C. jejuni* biology. Our results indicate that the environmental decay of infectious forms of *C. jejuni* is much slower than the decay observed for its culturable forms in a separate survival experiment and that spatial dispersion of infectious material with *C. jejuni* is most likely a result of multistep/multi-route dispersion. In Chapter 3, we analyse, in detail, the model identifiability, to inform the design of future transmission experiments. We conduct a comprehensive identifiability analysis, using a combination of methods, including mathematical analysis of the one-dimensional model, and analysis using combinations of simulated and experimental data. In Chapter 4, we demonstrate how our validated spatial modelling framework can further be used to explore the density dependence of environmental transmission. We analyse the density dependence of homogenous mixing, clustering, and distancing of hosts and present a number of scenarios describing uniform mixing, Poisson process mixing and maximal distancing of hosts. We show how constraining of hosts movements reduces transmission, which can be used to quantify intervention strategies targeted on reducing hosts mobility. Moreover, we discuss how host clustering or distancing compares to the random host placement depending on decay and diffusion speed which, when supplemented with relevant data, may be a basis for future studies of the influence of host clustering and/or distancing behaviour on transmission. Overall, in this thesis I presented the development cycle of simultaneous modelling and data gathering (experimental design), during which both new methodology was developed as well as new insights in *C. jejuni* transmission in broilers were obtained.

Table of contents

CHAPTER 1 GENERAL INTRODUCTION 9

CHAPTER 2 UNDERSTANDING ENVIRONMENTAL TRANSMISSION MECHANISMS: A PARSIMONIOUS MATHEMATICAL MODEL VALIDATED WITH INFECTION DATA FROM TAILOR-MADE EXPERIMENTS 29

CHAPTER 3 IDENTIFIABILITY OF ENVIRONMENTAL TRANSMISSION PARAMETERS: QUANTIFYING DISPERSION AND DECAY OF INFECTIOUS MATERIAL USING SPATIOTEMPORAL TRANSMISSION DATA 81

CHAPTER 4 USING SPATIAL MODELLING TO EXPLORE THE HOST-DENSITY DEPENDENCE OF R_0 FOR ENVIRONMENTAL TRANSMISSION 117

CHAPTER 5 GENERAL DISCUSSION 139

REFERENCES 159

SUMMARY 167

ACKNOWLEDGEMENTS 171

CURRICULUM VITAE 173

PUBLICATIONS 175

1

Chapter 1

General introduction

> ***'(...) a description of nature must be natural; it cannot be ad hoc.'***
>
> CHANDRASEKHAR, SUBRAHMANYAN
> "Karl Schwarzschild Lecture: The Aesthetic Base of the General Theory of Relativity." *Mitteilungen der Astronomischen Gesellschaft Hamburg* 67 (1986): 19

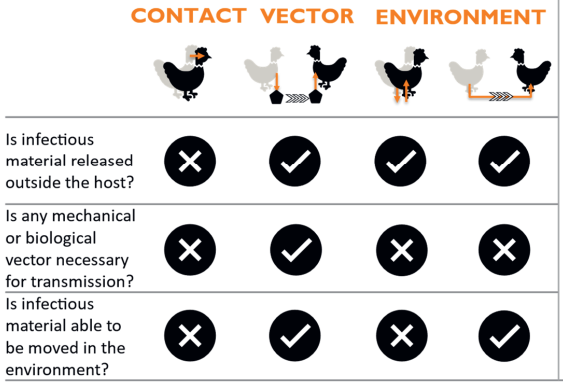
Throughout history "contact" is a key concept in epidemiological modelling. The first mechanistic models of infectious diseases were based on mass action hypothesis, where infectious and susceptible hosts constituting a population were seen as particles that can change their status (chemical or epidemiological properties) when they meet each other; as for chemical particles in solution, the probability of meeting ('collision') between recipient and infectious host is proportional to the product of densities of both (Heesterbeek, 2005).

Surprisingly, even though mathematical modelling of infectious diseases transmission is built on the concept of contact, 'contact' itself is often not well defined. According to the dictionary 'contact' can simply be defined as: "the fact of two people or things touching each other" (Cambridge Dictionary, n.d.). In the epidemiology, the definition of contact that (potentially) results in infection does not always follow this dictionary definition (Hoang et al., 2019), often having wider interpretation (Diekmann et al., 2013).

Definition of (potentially) infectious contact

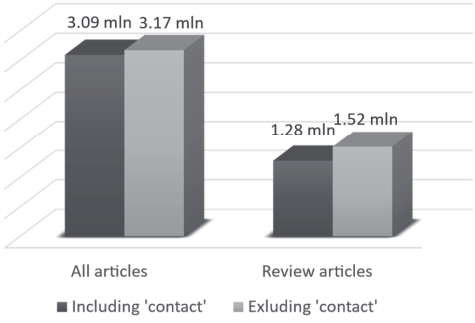
CONTACT *noun* **(TOUCH)**
the fact of two people or things touching each other

Cambridge University Press. (n.d.). Contact. In *Cambridge dictionary*. Retrieved November 20, 2022 https://dictionary.cambridge.org/dictionary/english/contact



GOOGLE SCHOLAR SEARCH TERM

Including 'contact'	Excluding 'contact'
contact (infectious OR infection) (model OR modelling) (transmission OR epidemiology OR epidemiological)	(infectious OR infection) (model OR modelling) (transmission OR epidemiology OR epidemiological)



SAMPLE OF 'CONTACT' DEFINITIONS FROM LITERATURE ON INFECTIOUS DISEASE TRANSMISSION MODELLING

Definition of contact	Source
*"Transmission of infection requires **proximity** between an infectious individual and a susceptible individual. The two individuals being proximate constitutes a "contact event."*	Held, L., Hens, N., D O'Neill, P., & Wallinga, J. (2019). Handbook of infectious disease data analysis. Chapman and Hall/CRC.
*For many diseases transmission can take place when two hosts "contact" each other, **where the meaning of "contact" depends on the context** (...) and may, in fact, sometimes be a little bit vague.*	Diekmann, O., et al. (2013). Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press.
*In other words the occurrence of a new infection depends on **a collision** between an infected and an un-infected animal*	McKendrick, A.G. (1912). The rise and fall of epidemics. Paludism 1, 54-66 (Transactions of the Committee for the Study of Malaria in India).
The nature of these contacts varies **depending on the disease and route (or routes) of transmission**.	Grassly, N., Fraser, C. Mathematical models of infectious disease transmission. Nat Rev Microbiol 6, 477–487 (2008).
A node in a contact network represents an individual host, and an edge between two nodes represents **an interaction** (that has occurred at any point in time for any period) which may allow disease transmission.	Shweta Bansal, Jonathan Read, Babak Pourbohloul & Lauren Ancel Meyers (2010) The dynamic nature of contact networks in infectious disease epidemiology, Journal of Biological Dynamics, 4:5, 478-489,
Physical contacts were consistently defined as involving any sort of skin-to-skin touching (e.g., handshake, hug, kiss, etc). The definition of nonphysical contacts differed somewhat among surveys. Specifically, the majority of surveys using two types of contacts defined a nonphysical contact as a two-way conversation of at least three words at a distance that does not require raising one's voice. In some other surveys, the definition involved close proximity (e.g., verbal communication made within 2 m) without specification of a minimum number of words to be exchanged.	Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, Beutels P, Hens N. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. Epidemiology. 2019 Sep;30(5):723-736.

The ambiguity of 'contact' definition is a direct consequence of the nature of the real-life transmission and processes underlying it. In reality, the vast majority of infections are not transmitted directly. Pathogens transmitted via air, water or fomites spend some time outside of the host, in the environment, before they are absorbed (e.g. inhaled, ingested or contacting skin) by a susceptible individual. For such environmentally transmitted diseases, a contact-based approach that links directly infectious and susceptible hosts is not a natural conceptualisation of transmission as it omits its environmental stage. The contact based direct transmission models such as the commonly used compartmental (SIR type) models are routinely used to describe environmental transmission (Rees et al., 2021). For these models, the rate of transmission is defined as a product of the host-to-host contact rate and the probability of transmission per contact (Real & Biek, 2007). It has been shown that direct transmission models are a good approximation of environmental transmission when pathogens survive in the environment relatively shortly (Breban, 2013) or when the overall environmental pathogen dynamics of the system is fast (Benson et al., 2021; Cortez & Weitz, 2013); as the same studies show, for many systems neglecting the dynamics of infectious material in the environment may produce erroneous results. This is the consequence of two properties of such models: 1) it is assumed that the recipient host can be infected only when the infectious host is present nearby (spatiotemporal correlation of recipient and infectious host); 2) the influence of the environment on pathogens (e.g. their survival or spatial spread) is not included in the model.

Within the contact-based approach, in an attempt to, account for the influence that the environmental stage has on the transmission, some authors introduced the concept of 'indirect contact'. I would like to note that 'indirect contact' is another ambiguous term, used almost exclusively in epidemiology, and that semantically it is an oxymoron. Based on the indirect transmission approach compartmental models can be constructed by adding the environmental load of pathogens as a new compartment (Breban, 2013; Lanzas et al., 2020), as is shown in Fig. 1.1.

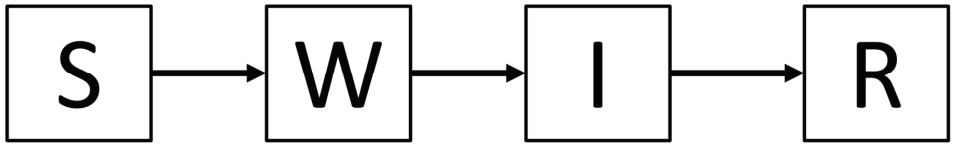
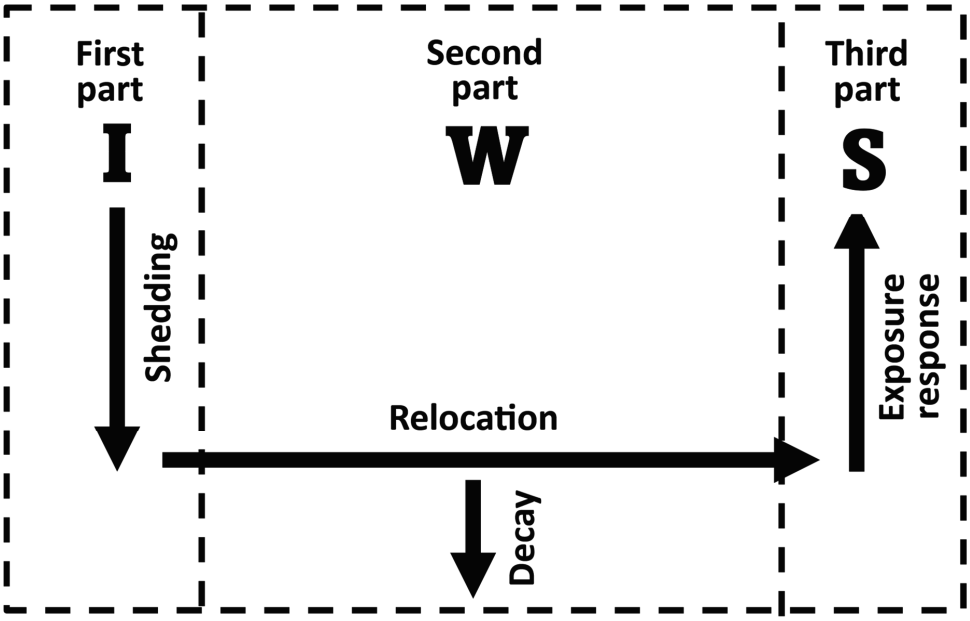


Figure 1.1. Schematic representation of a compartmental model of environmental transmission including the following compartments: S- susceptible (recipient) hosts, W- environmental load of pathogens- pathogen population, I- infectious hosts, R- recovered host.

Such compartmental model can still be interpreted from the contact-based point of view, by defining two separate contact processes: an infectious host contacts the environment to deposit a load of pathogens and a susceptible host contacts the environment to absorb a load (dose) of pathogens. As it was noted in (Brouwer, Weir, et al., 2017), it is difficult to define what is the 'dose' for environmental transmission as it is often not known in what form infectious material comes in (potentially) infectious contact with the hosts and what is the amount transferred into recipient hosts during one contact. Moreover, using a concept of single contact to the environment would correspond to a discrete dose-response model, where infectious material deposited in the environment is in contact with hosts only for short 'single contact' periods of time, which is not a natural representation of environmental transmission as in reality hosts can contact environment (almost) constantly e.g. by breathing, walking and/or interact with it for a prolonged period e.g. while eating or drinking.

Instead of looking at environmental transmission models from contact-based perspective, a more natural approach is to consider an 'infectious material based' perspective. Instead of conceptualising transmission as a point process of discrete events- contacts, the transmission can be seen as constant exposure to infectious material.



Fir 1.2. Schematic representation of the main general processes underlying the environmental transmission between infectious host (I) and susceptible (recipient) host (S) through the environment (W).

If in nature the vast majority of infectious disease is transmitted via environment, why has the contact concept become the go-to interpretation used for the transmission of infectious diseases? The main reason for the contact concept approach to be so prevalent is that the host (and their interaction such i.e. contacts) are easily observable and diagnosable. By observing the presence of pathognomonic clinical symptoms or using diagnostic tests on samples collected from hosts the number of (potentially) infectious hosts in the population can be inferred. At the same time, observing and measuring the concentration of infectious material in the environment is in most cases not possible. Among factors complicating the infectious material detection and pathogen load estimations are heterogeneities of infectious material load that require extremely high sampling intensity, or low concentrations of pathogen that require extremely sensitive detection methods. Moreover, if transmission routes or infectious forms of pathogen are multiple or not well understood it might be difficult to develop an appropriate sampling and detection protocol.

In this thesis, using an infectious material-based approach, I show that host status data indicating the spatiotemporal distribution of hosts can also be used to study processes underlying the environmental transmission of infectious diseases. Further, we explore if a spatial model based on an infectious material approach can be calibrated and validated using host status data only. In Chapter 2, we demonstrate that such data can be used to inform about decay and diffusion of infectious material if provided in certain spatiotemporal resolutions. In Chapter 3, to inform future studies, we explore what spatial resolution is needed to ensure the model identifiability. And in Chapter 4, we explore the inference that can be made with such a model to study density dependence of transmission and intervention strategies aimed at changing said density.

Contact-based vs infectious material-based approach

Because of environmental transmission a clear definition of contact cannot be established in a contact-based transmission modelling approach, a series of problems arise while constructing mathematical models from contact based perspective.

First, the dichotomy of indirect (e.g. via environmental compartment) vs direct transmission is inherent to the contact based modelling. Consequently the modelling decision has to be made whether ‘direct’ or ‘indirect’ or a combination of the two would be a better representation of the system. Such a dichotomy, while often used in modelling, is rarely observed in nature, and choosing the correct model from transmission data alone is often not possible (Cortez & Weitz, 2013). In an infectious material based transmission modelling approach the assumption is made that all infections are transmitted via the environment and exposure to infectious material rather than to infectious host governs the transmission; the so called ‘direct transmission’ can simply be seen as a limit where the production as well as decay of infectious material goes to infinity (Chang & de Jong, 2023) such that for transmission to occur the recipient host has to be at the same time and in the same place as the host that produces infectious material. In practice, this is already achieved when,

in the observation time (typically a day) both shedding and transmission occurs and transmission without shedding in the subsequent days is negligible.

Second, the infectivity of one infectious individual and exposure to infectious material by recipient hosts (used to calculate basic reproduction ratio) are not straightforward for environmental transmission in contact-based approach. For the direct transmission models, the typical (average) infectious individual is often defined only by its infectious period- as the period during which an infectious host is shedding infectious forms of pathogens; in the simplest models the infectivity of the host during its infectious period is assumed to be constant in time (Heesterbeek, 2002). Therefore, it does not account for the accumulation of infectious material in the environment and typically assumes that the recipient host can only be infected when the infectious host is present in the environment. The indirect transmission model does allow for accumulation of infectious material (e.g. in the environmental compartment) but, looking from contact based perspective, it has to be assumed how infectious host contacts the environment to deposit infectious material. In the infectious material-based transmission modelling approach, the infectious individual is described by an infectivity function that accounts for the amount of material they (continuously) deposit in the environment during their infectious period and other dynamic processes that determine how the concentration of the infectious material in the environment changes in space and time. As a consequence, the infectivity function (being the 'environmental load density function' described below) has three stages (modes): 1) a build-up stage, when material accumulates in the environment; 2) a (pseudo) equilibrium stage when amount of produced material equals the amount of material decayed; 3) a decay stage, when no new material is produced, and previously accumulated material is decaying. The infectivity patterns for an infectious material-based model is compared with a contact-based direct transmission model in Fig. 1.3. In Chapter 3, we show the importance of these three stages for the model parameter identifiability. The exposure to infectious material relevant for recipient host is described as continuous process determined by exposure pattern.

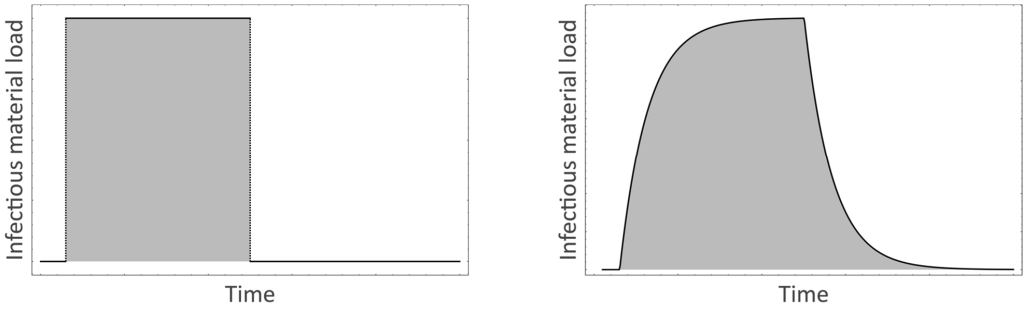


Figure 1.3. Examples of infectivity patterns for a direct transmission model (plot A) and environmental transmission model (plot B).

Third, for the contact-based point of view on transmission the spatial aspect of transmission is not easy to implement in modelling. For direct transmission models, the general assumption is that contact happens when host are in the same place (or really close nearby) at the same time. In other words, infectivity is corelated with position of infectious hosts. For environmental transmission such correlation is not necessarily the case. For indirect transmission models, the spatiotemporal correlation of hosts is not needed for transmission, yet to formulate the models (e.g. individual based models) it needs to be decided where and when host contact the environment to deposit portion of infectious material and how to simulate or mathematically describe such process. To make that decision, the contact process must be well understood, which is not the case for most real-life transmission scenarios. Moreover, resulting model would be highly system specific. In the infectious material-based approach the environmental load is described by a continuous function just by defining the infectivity function of the infectious individual in a spatial context and adding the dispersion mechanism as I show below. This gives us the flexibility to study various spatial configurations of sources as well as exposure areas as shown in Chapters 3 and 4 of this thesis.

Fourth, the scaling with population size and density is not clear for contact-based approaches. Two models were proposed to account for change in population density, namely frequency and density dependent transmission models. The frequency dependent transmission assumes that a contact rate is constant with changing population density (Begon et al., 2002). Consequently, there is increase in competition for contact with a particular host when density increases. This density dependence assumes that contact rate scales linearly with population density (Begon et al., 2002), so there is no increase in competition for contact with a particular host when density increases. As contact is not well defined for most systems, there are neither clear rules about the model choice, nor it is known how transmission would scale with changing density. As a consequence, there is no consensus in the literature, when and how to incorporate density dependence of transmission for both direct (Begon et al., 2002) and indirect (Lanzas et al., 2020) transmission models. This often precludes the comparison of results obtained for the systems where the density (and consequently population size and/or area occupied by hosts) vary. As spatial models can easily be created within our modelling framework based on infectious material-based transmission approach, enabling a more systematic exploration of the density dependence of transmission, in Chapter 4 we compare the scenarios where host numbers and spatial organisation were varied.

Detailed mathematical description of infectious material-based approach

Below I present a (relatively) simple modelling framework based on an infectious material approach providing a basis for modelling environmental transmission of infectious diseases. This is based on a similar model that was published in (van Bunnik et al., 2014). The framework has three main components of infectious material-based approach: 1) an infectivity source pattern describing the source(s) of infection (being the infectious hosts); 2) a process based environmental load dynamics that accounts for decay and dispersion of infectious material deposited in the environment; 3)

exposure response that accounts for absorption of material from the environment by the recipient (susceptible) host and for within host processes in said host.

The spatiotemporal model incorporates all three parts in the probability of infection equation that describes for one susceptible host the probability of being infected during one unit of time. Part of the probability of infection equation is an environmental load density function $W^i\left(\text{t, x, y, } T_1^i, T_2^i, A_{\text{inf}}^i | \rho, \alpha, D\right)$ that describes the spatiotemporal distribution of infectious material produced by one particular infectious host described with an 'infectivity pattern' defined below. Following the principle of parsimony, the framework describes only the main, general processes that are underlying environmental transmission using simple, mechanistic models to describe them. The main equations are presented in Fig. 1.4 while the main parameters describing processes are in Table 1.1. A detailed description of the processes is presented below. Here I initially ignore model parameter identifiability issues for ease of presentation; these issues, (as well as the reduced set of model parameters resulting in practice), will be discussed subsequently.

$$W^i(\mathrm{t, x, y}, T_1^i, T_2^i, A_{\mathrm{inf}}^i|\rho, \alpha, D) = \begin{cases} 0 & t < T_1^i \\ \rho \int_{T_1^i}^{t} \iint_{A_{inf}^i} \frac{1}{4\pi D(t-\tau)} \exp\left[-\frac{(x-x_i)^2+(y-y_i)^2}{4D(t-\tau)}\right] \exp[-\alpha(t-\tau)] dx_i\, dy_i\, d\tau & T_1^i \le t \le T_2^i \\ \rho \int_{T_1^i}^{T_2^i} \iint_{A_{inf}^i} \frac{1}{4\pi D(t-\tau)} \exp\left[-\frac{(x-x_i)^2+(y-y_i)^2}{4D(t-\tau)}\right] \exp[-\alpha(t-\tau)] dx_i\, dy_i\, d\tau & T_2^i < t \end{cases}$$

$$P_{\mathrm{inf}}\left(t_1^r, t_2^r, A_{exp}^r, \{T_1^i, T_2^i, A_{inf}^i\}_I|\xi, \rho, \alpha, D\right) = 1 - \mathrm{Exp}\left[-\xi \int_{t_1^r}^{t_2^r} \iint_{A_{exp}^r} \sum_I W^i(\mathrm{t, x, y}, T_1^i, T_2^i, A_{\mathrm{inf}}^i|\rho, \alpha, D)\, dA_{exp}^r\, dt\right]$$

Diffusion process
Decay process
Infectivity source pattern (shedding process)
Exposure-response process

Figure 1.4 Main equations of the spatiotemporal infectious material based transmission modelling framework: an environmental load density function $W^i(\mathrm{t, x, y}, T_1^i, T_2^i, A_{\mathrm{inf}}^i|\rho, \alpha, D)$ depending on the infectivity pattern determined by one particular infectious host *i* located in the area A_{inf}^i that sheds the infectious material with shedding rate parameter ρ during time period from T_1^i to T_2^i, and on the decay rate parameter α and the diffusion coefficient D of infectious material, and force of infection $P_{\mathrm{inf}}\left(t_1^r, t_2^r, A_{exp}^r, \{T_1^i, T_2^i, A_{inf}^i\}_I|\xi, \rho, \alpha, D\right)$ experienced by one particular recipient host *r* located in the area A_{exp}^r during a time period from t_1^r to t_2^r being exposed with exposure rate parameter ξ to the infectious material produced by a set of infectious hosts I (here we assume that all infectious hosts in the set I have the same shedding rate parameter ρ).

Table 1.1. Main parameters describing processes of the spatiotemporal infectious material-based transmission modelling framework.

Parameter			Definition
ξ	[day^{-1}]	exposure rate parameter	the rate at which the exposed individual becomes exposed to the infectivity unit of the environmental load
ρ	[day^{-1}]	shedding rate parameter	average number of infectivity units that are deposited into the environment per unit of time; assuming homogenous shedding
α	[day^{-1}]	decay rate parameter	the rate at which the infectious material loses its infectivity
D	[m^{2}day^{-1}]	diffusion coefficient	a proportionality constant describing mean square displacement of random walking particle

Infectivity source

When an infectious host is present in the environment and occupies a certain area, it produces new portions of infectious material throughout its infectious period. In the infectious material based model we call these hosts 'source of infection'. The infectivity source pattern $\left\{T_1^i, T_2^i, A_{\text{inf}}^i \,|\rho^i\right\}_I$ (as defined in more detail in Chapter 2) is a set of I infectious hosts each described with its infectious period (T_1^i, T_2^i), an area they occupy A_{inf}^i and a parameter describing source strength (the intensity of the source) - shedding rate parameter ρ^i.

In our modelling framework, to follow the principle of parsimony, we define the infectious period as a period in which a host is shedding infectious material continuously. Therefore, we assume that the new infectious material is produced and shed all the time in a deterministic process.

The exposure-response

For environmental transmission, the exposure response depends on the absorption process, describing how recipient hosts encounters the infectious material in the environment and the within host processes that determine the probability that an infectious particle causes the infection when absorbed by the recipient host. Additionally, it accounts for processes happening in the environment; in the parsimonious modelling framework we define two of such processes: decay and diffusion of the infectious material.

Similarly to source hosts, recipient hosts are described with exposure pattern $\left\{t_1^r, t_2^r, A_{\text{exp}}^r \,|\xi^r\right\}$ that describes exposure period (t_1^r, t_2^r), exposure area A_{exp}^r and an exposure rate parameter ξ^r. In our framework we use a single parameter ξ^r which represents the rate (i.e. per day) at which the recipient individual becomes exposed to the infectious material in the environment, per infectivity unit irrespective of whether it is active or not and is a combination of absorption rate parameter ε and probability of being infected after the infectious particle is absorbed by host p:

$$\xi = p\varepsilon \qquad (1)$$

The 'expected' dose response

To construct the exposure response, one can start with the dose response relationship. The dose response can be constructed following the independent action hypothesis; according to this hypothesis, each unit of infectious material (amount that cannot be further divided, for example 1 egg of macroparasite or 1 bacterium- or rather measurable CFU- or 1 virus particle) can independently infect the host with probability p and the host can escape the infection (potentially caused by this 1 unit) with probability 1-p such that each (potential) transmission event is a Bernoulli trial. When a host absorbs more than 1 unit of infectious material, it must escape from all the units in order to escape from the infection. In this case we have n Bernoulli

trials and the probability of escaping the infection when n units are ingested is the conditional probability of escaping during all n Bernoulli trials:

$$p_{esc} = (1-p)^n \quad (2)$$

Therefore, the probability of being infected after all trials (after being exposed to all n units) is

$$p_{inf} = 1-(1-p)^n. \quad (3)$$

Equation 3 represents the cumulative distribution function (CDF) of the geometric distribution that describes the probability of having less than or equal number of failures (escape from infection) before the first success (infection) occurs.

As in reality, infectious units are always transmitted in a medium, in experimental work the exact number of infectious units (exact dose) cannot be estimated. Rather, the concentration of pathogen in the medium (such as inoculum or gram of faeces) is calculated. The same applies to situation when recipient host is absorbing some volume of infectious material (the dose). Infectious material is therefore defined as the medium that contains infectious units. Knowing only the volume of infectious material and the concentration of infectious units we cannot calculate the exact dose. As the exact number of infectious units in a finite volume of infectious material is Poisson distributed, we can calculate the expected dose. Incorporating the stochasticity of infectious material concentration leads to the well-known exponential dose-response relationship for the expected dose R as follows:

$$P_{\text{inf}} = \sum_{N=0}^{\infty}\left[\text{PDF}_{\text{Poisson}}(N, \mu = R) \times [1-(1-p)^N]\right] = \sum_{N=0}^{\infty}\left[\frac{e^{-R}R^N}{N!}[1-(1-p)^N]\right] = 1-e^{-pR} \quad (4)$$

Generality of the 'expected dose-response'[1]

Interestingly, in any transmission model with a continuous-time infection hazard such as contact based mass action type models, the mathematical expression of the probability of infection across a finite time interval has the same form as the exponential dose-response.

In such transmission models, the probability of infection for a given susceptible satisfies the following Master equation:

$$\frac{dP_{inf}(t)}{dt} = \lambda(t)\left[1 - P_{inf}(t)\right]$$

Where, λ(t) is the infection hazard (force of infection to which the susceptible is exposed).

The solution can be obtained by considering the equation for the probability $1 - P_{\mathrm{inf}}(t)$ of having escaped until time t and reads as follows (assuming start in clean environment- $P_{inf}(t = 0) = 0$):

$$P_{inf}(t) = 1 - e^{-\int_0^t \lambda(t)d\tau}$$

Comparing the expressions derived for expected dose-response $P_{\mathrm{inf}} = 1 - e^{-pR}$ we can conclude that any transmission model with a continuous-time infection hazard has the exponential dose-response built in, with $pR(t) \equiv \int_0^t \lambda(\tau)d\tau$, where $\mathrm{R}(t)$ is the expected cumulative dose up until time t. However, such transmission models leave p and $\mathrm{R}(t)$ implicit as only the product is described (i.e. described as a cumulative infection hazard).

This result is no coincidence, as the exponential expected dose-response relationship follows from the Independent Action Hypothesis, and the model formulation based on an infection hazard corresponds to assuming that consecutive infinitesimal time intervals of exposure represent independent actions; furthermore Poisson variation in the dose arises either through integrating a concentration over a finite volume or through cumulation of a hazard across a finite time interval.

1) Hagenaars, T.J., Personal Communication. 2020.

Following from the generality of the dose response, we can define the exposure response as a continuous process, without a need for dividing the amount of infectious material into discrete doses.

Environmental load dynamics

For the mechanistic model here I present the force of infection that is a part of the exposure response incorporating the exposure response parameter as well as the environmental load dynamics: the main general processes underlying the environmental part of transmission are assumed to be decay and diffusion of infectious material.

The 'expected active' dose

It is known that infectious particles such as bacteria and viruses may be inactivated while spending time outside the host. In other words, if these particles spend enough time in the environment, they are not able to infect hosts even if absorbed and therefore do not contribute to the probability of infection.

To formulate a dose response that accounts for pathogen inactivation (e.g, decay) in the environment we assume that each infectious unit has the same probability of being inactivated (decay) while residing in the environment and this probability does not change in time. Similarly, to the 'expected dose', the 'expected active dose' can also be described with an exponential formula arising from a particular environmental sojourn time period of length (0, t) during which the source was continuously emitting infectious material with shedding rate ρ:

$$W(t|\rho,\ \alpha) = \rho \int_0^t exp[-\alpha(t-\tau)]d\tau, \qquad (5)$$

where α is the decay rate. Therefore, the Equation 5 describes what fraction of the infectious material produced until time t is still active at this time.

The 'expected active local' dose

Transmission experiments (done with animals separated from each other) show that transmission of the infection is possible even when animals do not occupy the same space (Herfst et al., 2012; Holt et al., 1998; van Bunnik, 2014; van Bunnik et al., 2014; Zhou et al., 2018). This indicates that infectious material spreads from the source infectious hosts through environment. Therefore, we assume that in our spatial framework the infectious material is dispersing spatially. To follow the parsimony principle, we use the simplest model that accounts for dispersion of material- the diffusion process that assumes that each particle is moving according to simple random walk (Chandrasekhar, 1943). The diffusion equation, for the one particle regime, describes the probability of finding a particle in certain space, e.g. rectangular described with coordinates, after time t passed since the start of diffusion, and for many particles regime, it describes the concentration of particles in said space. For the one-dimensional case with point source that released the particle at time t=0 from x=0, the diffusion

equation is equal to the probability density function of normal distribution, with a mean in the source position (x=0) a time dependent standard deviation $\sigma = \sqrt{2Dt}$:

$$W(t, x|D) = \frac{1}{\sqrt{4\pi Dt}} exp\left[-\frac{x^2}{4Dt}\right] \quad (6)$$

where D is the diffusion coefficient. Combining equations 5 and 6, we obtain the spatial distribution of the environmental load of infectious particles produced until time t that are still active (not decayed):

$$W(t, x|D, \rho, \alpha) = \rho \int_0^t \frac{1}{\sqrt{4\pi D(t-\tau)}} exp\left[-\frac{x^2}{4D(t-\tau)}\right] exp[-\alpha(t-\tau)]d\tau \quad (7)$$

The two-dimensional diffusion equation for the rectangular continuous source we use in the framework was derived by (van Bunnik et al., 2014).

Connection to data

Epidemiological modelling can be used to study processes underlying transmission only when it is applied with biological (experimental and/or field) data. In mechanistic modelling, the processes are characterised by the parameters describing them, and the biological information about these underlying processes can be obtained when their values are estimated provided there is a confidence in these estimated values and when model fit generated with these values is consistent with biological observations.

Parameter estimation done by fitting the model to the data is called calibration (Rees et al., 2021). For a successful model calibration, the model parameters should be estimated with reasonable uncertainty i.e. have finite value and finite confidence bounds. Therefore, part of the calibration process is the identifiability analysis during which the model is analysed in the context of the data available to examine if all model parameters can be estimated and what data is crucial to obtain reasonable estimates. We present the full identifiability analysis of our spatiotemporal infectious material-based model in Chapter 3 of this thesis.

Apart from model calibration, the quality of model fit must be assessed to gain confidence that parameter estimates are correctly describing biological processes. Any modelling framework describing biological phenomena is useful *only* if it satisfactorily describes the biological observations despite the simplifying assumptions incorporated in the models. During model development, this is ensured by model verification and validation. During verification the structure of the model and its implementation is checked to ensure that the framework implementation works as intended, while during the validation the biological assumptions are challenged by comparing the model outcome with outcome observed in real life. For both verification and validation of the model the final or intermediate output of the model is generated using the parameter set obtained in the model calibration and compared to either *a priori* known parameter values or to the data. As during the verification process it is checked if the model is technically well

implemented, either real life or simulated (artificial) data can be used. The simulation data is preferred, as such data can be created for *a priori* known parameter sets. As during validation process it is checked if model output is consistent with biological observations, data from such observations is crucial for the validation of the model. The model validation using data from transmission experiments in broilers is presented in the Chapter 2 of this thesis and the identifiability analysis of the model done with simulated data presented in Chapter 3 serves also as a verification procedure.

Transmission experiments

To calibrate and validate the model, relevant biological data needs to be available. As previously discussed, in epidemiology the host status data where it is recorded if the host is colonised (and infectious) or non-colonised (and potentially susceptible to infection) are often gathered for many systems either during population screening, field sampling conducted for naturally occurring infections or during transmission experiments. As for the first two cases, the transmission happens naturally, there often is little control of data quality i.e. its spatial and temporal resolution. Due to insufficient temporal resolution, many cases are found positive during the same sampling period, and it is difficult to infer the true transmission chains from the field data. Moreover, for environmentally transmitted diseases, there is little control over environmental conditions influencing the environmental stages of transmission, such like temperature, humidity, or ventilation, which makes the extrapolations, data comparisons or aggregation extremally difficult.

Animal transmission experiments were proposed as a solution to study the transmission of infectious disease in controlled conditions, where the chain of transmission can be established more easily by testing status of all hosts before the start of experiment and choosing appropriate experimental design and sampling protocol to ensure sufficient data quality and quantity (Velthuis et al., 2007). As shown here, in small scale transmission experiments the resolution of status data can be controlled which solves (some of) the problems with model unidentifiability, and the environmental conditions can be standardised to ensure the data can be combined, compared and that measurements taken are suitable for the model in mind.

As modellers often encounter identifiability problems (that are not always directly reported as such), in addition to host data they decide to estimate some parameter separately taking independent measurements from experiments other than host status data (see (Brouwer, Weir, et al., 2017; Colenutt et al., 2020; van Bunnik et al., 2014) for some examples). For the infectious material-based approach, we provided examples of such experiments in Table 1.2.

Table 1.2. Summary of the experimental approaches that can be used to estimate separately the parameters of infectious material-based transmission model

Process	Parameter aimed to be estimated	Experiments to be combined with host status data from transmission experiment	Example from the literature*
Shedding	ρ- shedding rate parameter- average amount of infectious material released by 1 infectious host during 1 time unit	1st Direct measurements of the amount of pathogen that is produced by one infectious host produced per time unit 2nd Indirect indicators, e.g. environmental load estimation from environmental samples	1st (Colenutt et al., 2020) (Collineau et al., 2020) 2nd (Colenutt et al., 2020) (Brouwer, Weir, et al., 2017)
Decay	α- decay rate parameter	measure changes of concentration of pathogens in infectious material in time	(van Bunnik et al., 2014) (Colenutt et al., 2020)
Dose-response	p-probability that the ingested active dose will cause an infection	Inoculation dose response experiment	(van Bunnik et al., 2014) (Lunn et al., 2019)
Diffusion	D- diffusion coefficient	Potentially: Measure fluctuations of concentrations of pathogens in infectious material	Has not been used in epidemiology
Absorption	ε- the absorption rate	Not possible to estimate from a single type of experiment; can be estimated only if we have estimates for all other parameters	

*Examples of the experiments, not necessarily their analysis

The model system: transmission of Campylobacter in broilers

To calibrate and validate the spatiotemporal infectious material model we use *Campylobacter jejuni (C. jejuni)*transmission in broilers as a model system as it was previous done by (van Bunnik et al., 2014). *Campylobacter* spp. are gram negative bacteria from *Campylobacteriacae* family that cause campylobacteriosis, common foodborne disease, in humans (Garrity et al., 2005). These zoonotic bacteria have many other hosts including farm animals and pets (Garrity et al., 2005). As chickens, particularly broilers, are considered the main source of campylobacteriosis, many measures have been taken to mitigate the *Campylobacter* transmission in poultry flocks (Hansson et al., 2018; Wagenaar et al., 2013). Despite these interventions, *Campylobacter* is still highly prevalent in broiler flocks (Mota-Gutierrez et al., 2022) and the reasons for this are not clear. Consequently, a better understanding of *Campylobacter* transmission into and within the flock was identified as an knowledge gap important for design of effective interventions (Hansson et al., 2018). The transmission of *Campylobacter*, together with another identified knowledge gap: better understanding of *Campylobacter* survival in environment (Hansson et al., 2018) can be studied using the spatiotemporal infectious material based approach as presented in (van Bunnik et al., 2014).

In addition to the specific health impact, *C. jejuni* transmission in broilers can also be an appropriate system to gain insights into general mechanisms underlying the transmission between spatially separated hosts in experimental conditions (van Bunnik et al., 2014). Several factors make *C. jejuni* transmission in broilers an excellent experimental system. First, as in the transmission experiments, we use a low pathogenicity *Campylobacter Jejuni* strain, the discomfort of animals used in experiments is minimized. Moreover, as the transmission dynamics in broiler flocks is fast: *C. jejuni* is highly infectious, small dose can cause infection in recipient broilers (Line et al., 2008), it is shed in large amounts in faeces of infectious broilers (Stern & Robach, 2003), and generally it does not survive well in environment (Park, 2002; Smith et al., 2016), especially when compared to other environmentally transmitted bacteria such as *Salmonella* spp (Andino & Hanning, 2015) or *Escherichia coli* (van Bunnik et al., 2014; van Elsas et al., 2011). Due to the fast dynamic, the time of experiments can be reduced, minimizing the discomfort of overgrown broilers and the cost of experiments. As (van Bunnik et al., 2014) have established , the *C. jejuni* transmission in broilers is also a proper system to study the transmission between spatially separated broilers, while using relatively small spatial scale, such that it is practically feasible to house multiple separated areas occupied by broilers in laboratory animal facilities. Here, we use previously published (van Bunnik, 2014; van Bunnik et al., 2014) and newly gathered data from series of *C. jejuni* transmission experiments between broilers separated by various distances to calibrate and validate our spatiotemporal infectious material based model.

2

Chapter 2

Understanding environmental transmission mechanisms: a parsimonious mathematical model validated with infection data from tailor-made experiments

Anna M. Gamża[1,2*], Thomas J. Hagenaars[2*], Miriam G.J. Koene[2], Mart C.M. de Jong[1*]

1) Quantitative Veterinary Epidemiology, Wageningen University & Research, 6708 PB Wageningen, the Netherlands
2) Wageningen Bioveterinary Research, Wageningen University & Research, 8221 RA Lelystad, the Netherlands

Abstract

Although most infections are transmitted through the environment, the processes underlying the environmental stage of transmission are still poorly understood for most systems. Improved understanding of environmental transmission dynamics is important for effective non-pharmaceutical intervention strategies. To study the mechanisms underlying environmental transmission we formulated a parsimonious modelling framework including hypothesised mechanisms of pathogen dispersion and decay. To calibrate and validate the model, we conducted a series of experiments studying distance-dependent transmission of *Campylobacter jejuni (C. jejuni)* in broilers.

We obtained informative simultaneous estimates for all three model parameters: the parameter of *C. jejuni* inactivation, the diffusion coefficient describing pathogen dispersion, and the transmission rate parameter. The time and distance dependence of transmission in the fitted model is consistent with marked spatiotemporal patterns in the experimental observations. These results, for *C. jejuni* in broilers, show that the application of our modelling framework to suitable transmission data can provide mechanistic insight in environmental pathogen transmission.

Introduction

Traditionally, transmission of infectious diseases is modelled as a process that occurs when a susceptible host has direct contact with an infectious host (Heesterbeek, 2005). However, the majority of pathogens are, in fact, transmitted through the environment, i.e. indirectly: through the air, via surfaces and/or via fomites whilst residing in droplets, dust particles or otherwise. For pathogens that spread in this fashion, infectious material is shed into the environment by the infectious host (the source) and it is taken up (e.g. ingested, inhaled, or absorbed through mucosa) by a susceptible recipient host after the material has spent any amount of time in the environment. During this time, infectious material is losing infectivity due to inactivation (decay) of pathogens as these are exposed to environmental conditions that often are not optimal for their survival (Yildiz, 2007). Additionally, part of emitted infectious particles may never produce any exposure to a recipient host, for example by being deposited in locations inaccessible to the hosts. Furthermore, data from indirect transmission experiments confirm that infectious material can also be dispersed through the environment, for example being moved via air flow, water movement, or movements of contaminated objects i.e. fomites (Asadi et al., 2020; Herfst et al., 2012; Holt et al., 1998; van Bunnik et al., 2012; van Bunnik et al., 2014; Zhou et al., 2018). This may cause the material to be moved from areas that are close to the source to parts of the environment more distant from the source, thereby reducing the probability of infection near the source and at the same time facilitating uptake by distant recipient hosts. Depending on the speed of dispersion, material arriving at larger distances from the source is expected to have undergone more inactivation due to a longer travelling time. Thus, the processes of shedding, decay and dispersal interact to shape the overall spatiotemporal rate of pathogen transmission through the environment.

The environmental stage of transmission provides opportunities for non-pharmaceutical interventions aimed at reducing transmission. Some of these, such as separation of hosts (“social distancing”) and hygiene protocols, are applied to the infectious and/or recipient individuals; while other, such as disinfection or ventilation procedures , are applied to the environment occupied by hosts. To develop the best (combination of) intervention strategies and quantify their efficacy, mechanistic mathematical models of transmission are needed that are both calibrated (i.e. having identifiable parameters) and validated.

Here we present an experimentally validated modelling framework to mechanistically model environment dependent processes, namely pathogen decay and dispersion. We use a parsimonious modelling approach, motivated by the fact that detailed transmission mechanisms in host-pathogen-environment systems are generally difficult to observe and measure. Parsimonious models can be used even when limited observations are available. In addition, such models can complement and validate more detailed modelling of very specific hypothesised transmission routes., e.g. based on aerosol physics (Drossinos & Stilianakis, 2020; Wagner et al., 2021)

Mechanistic models of environmental transmission that not only account for pathogen survival but also for movement of infectious material are still under development. As it is described in(Lanzas et al., 2020) two main types of mechanistic models are commonly used to describe environmental transmission: "mean field" compartmental models, versus individual based models; usually these models do not account for movement of infectious material through the environment. The simplest mechanistic description for environmental dispersion of material, that is in our case carrying a certain pathogen load, is as a diffusion process (Chandrasekhar, 1943). In previous epidemiological models, diffusion, or reaction-diffusion, has been mainly used for describing diffusion of hosts, as for example in (Huang et al., 2010). For environmental transmission, diffusion models were also developed, few of which implemented diffusion of pathogens or infectious material for plant (El Jarroudi et al., 2020; Gilligan, 1995; Pielaat & Van Den Bosch, 1998) and animal or human diseases (David et al., 2020; Pang & Xiao, 2019; Wang et al., 2018; Xiao et al., 2020). While the vast majority of transmission models accounting for dispersion of infectious material remains theoretical, for some more complex, simulation models the validation was reported (for airborne (Sørensen et al., 2001; van Leuken et al., 2015) and waterborne (Bidegain et al., 2017) transmission).

More generally, the importance of model calibration and validation for zoonotic environmentally transmitted infections, such as *Campylobacter spp.*, was recently raised in (Rees et al., 2021); authors noted that less than half of the 210 analysed models were validated with real-life data emphasizing the need for modelling that is driven by actual transmission data.

To calibrate and validate an epidemiological model two types of transmission data can be used: field data and data from controlled transmission experiments. The most common information source is field data collected by detecting naturally occurring infections in the areas where the pathogen is prevalent or emerging. As field data represent naturally occurring transmission chains, often there is little control over their quantity and quality. Moreover, linking infection events to infectious source individuals is challenging as the contact structure and chain of infections cannot always be observed and/or verified.

Alternatively, the data can be collected from tailor-made transmission experiments conducted using animal models (Velthuis et al., 2007). In transmission experiments the entire process, i.e. the shedding by infectious hosts, environmental stage, and exposure response of recipient hosts, can be studied in controlled (environmental) conditions. Both source and recipient hosts can easily be identified by starting the experiments in a clean (pathogen free) environment and recording the status of all hosts before and during transmission period. The experiments can be tailored to the spatiotemporal resolution necessary for the system of interest, so that a mathematical model at hand can be calibrated and validated. The last property is especially important for systems where identifiability problems cannot be solved by additional data on specific mechanisms.

An environmental transmission model accounting for pathogen decay and diffusion has been presented previously in (van Bunnik et al., 2014); the four-parameter model was fitted with data obtained in experiments on *Campylobacter jejuni* (in the remainder of this paper abbreviated as *C. jejuni*) and *Escherichia coli* transmission between broilers spatially distanced by a single distance

band (of 0.75 or 1.06 m) combined with a separate survival experiment in which concentration of culturable bacteria in faeces was measured. This separate survival experiment was needed to solve identifiability problems in the model; here we prove that also our three-parameter model describing decay and diffusion of infectious material is not identifiable with the previously published data on *C. jejuni* transmission. To show that the model can be fully identifiable from transmission data only when sufficient spatiotemporal resolution is provided, we conducted new experiments for the same system, varying the distance and the timing of exposure. With the addition of the spatiotemporal data from these new experiments we were able to estimate all parameters, i.e. also the decay rate parameter, with remarkable results when compared to the estimate obtained from the separate survival experiment. Subsequently, we quantitatively validated both time and distance dependence of the model showing that the model fit is consistent with distance dependent delay times and proportions of hosts colonised observed in experiments.

This approach enabled us to simultaneously study pathogen decay and dispersion in the environment using parsimonious modelling and spatiotemporal data from transmission experiments only and hence obtain new insights into mechanisms underlying environmental transmission of *C. jejuni*. One of the insights being that separate experiments counting culturable bacteria in the environment may not provide information representative for the decay rate associated with the infectious environmental stages of the bacteria.

Results:

Model

To study environmental transmission, we developed a spatiotemporal three-parameter model, where each parameter has a precise biological interpretation. A decay rate parameter α describes how fast *C. jejuni* is inactivated in the environment. A diffusion coefficient D describes how the spatial distribution of infectious material in the environment changes over time, as a result of movement/dispersal of this material. A transmission rate parameter β describes probability of infection given one unit of exposure during one time unit. It reflects the joint effects of the remaining host-dependent processes, namely shedding of infectious material by source hosts, exposure to this material of recipient hosts and response (i.e. infection or colonisation) to the dose that the recipient is exposed to.

For a standard compartmental model extended with an environmental reservoir with spatially homogeneous infection load, the probability of infection can be represented as:

$$P_{\text{inf}}(t_1,t_2) = 1 - \exp\left[-\beta \frac{\overline{W}\times(t_2-t_1)}{N}\right], \quad (1)$$

where β is a transmission rate parameter, N the number of hosts, and $\overline{W}$ is the average environmental exposure during a time interval from t_1 to t_2:

$$\overline{W} = \frac{\int_{t_1}^{t_2} W(t)dt}{(t_2-t_1)}. \quad (2)$$

In our spatial model, the probability of infection is given by a spatially non-homogeneous generalization of Equation 1. It represents the exposure response for a given recipient host occupying an exposure area A_{exp} during a time interval t_1 to t_2 and reads as follows:

$$P_{inf}(t_1, t_2, A_{exp}) = 1 - \exp\left[-\beta \int_{t_1}^{t_2} \iint_{A_{exp}} W(t, x, y)\, dx\, dy\, dt\right], \quad (3)$$

where β (again) is a transmission rate parameter and W(t,x,y) is the spatiotemporal density function of the environmental load the recipient is exposed to. W describes how the distribution of the accumulated load changes in time and space, and in our model accounts for continuous shedding, exponential decay and diffusion of material. For further details on the modelling we refer to the Methods section.

Experimental results

To calibrate and validate the model we conducted a series of animal experiments studying transmission of *C. jejuni* between spatially separated broilers. Data on infection were gathered by recording colonization status of the recipient hosts at various locations and at various time points. In earlier transmission experiments on *C. jejuni* broilers were housed in pens separated from the source by a single distance band (of 0.75 or 1.06 m) (van Bunnik et al., 2014) and this narrow distance range proved to be insufficient for estimation of all of the model parameters nor validation of the distance dependence of transmission. In the current study, we designed experiments to compare transmission in pens that were separated by a much broader distance range (0.00 m- 2.00 m). In all those experiments, five broilers were inoculated with *C. jejuni* and placed in an experimental room at day 0. Two types of experiments with slightly different design were conducted: type 1 experiments where exposure of recipients started the same day as the source animals were inoculated and type 2 experiment where the exposure started 20 days after inoculation of the source broilers. This allowed us to validate if the delay in (onset of) transmission across a distance, as reported in (van Bunnik et al., 2014), is shorter when transmission starts in a contaminated environment, as predicted by our spatiotemporal model. Table 2.1 sums up the most important differences between the two types of experiments.

Table 2.1. Summary of experimental design for *C. jejuni* transmission experiments between separated broilers

		Experiment type 1	**Experiment type 2**
Duration	[day]	35	35
Time of inoculation o source hosts	[day]	0	0
Start of exposure of recipient hosts	[day]	0	20
Border to border distance ranges between source and recipient host areas	[m]	0.35- 2.00	0.43- 0.89[1] and 0.00[2]
Number of hosts per recipient pen		1	2
Recipients removed after found positive		Yes	No
Experimental groups		N/A	A, B, C

1) Group A;
2) Groups B & C;

In the type 2 experiment, to gather data on transmission on extremely short distances (0.00 m), beside distanced recipients (group A) we included two additional pairwise groups. Group B consisted of recipients housed together with inoculated animals (source hosts) also from day 20 onwards. Group C consisted of recipients housed with another host that initially was a recipient host and turned into a source host by becoming infected. We note that in group C the recipients were not only exposed to the infectious material of their colonized pen mate but also to material originating from the distanced sources; however, our modelling indicates that the contribution of the latter sources to the total exposure of a group-C recipient is relatively small (see Supplementary Note 4 for details).

Fig. 2.1 shows the outcome of experiments as a function of time from the start of experiment (start of the exposure) and as a function of border-to-border distance between source area and recipient area. Source areas are defined as pens occupied by colonized hosts, while recipient areas are defined as pens occupied by recipient hosts (non-colonised when exposure started). When an area contained both colonized and non-colonized broilers it therefore was both a source and a recipient area.

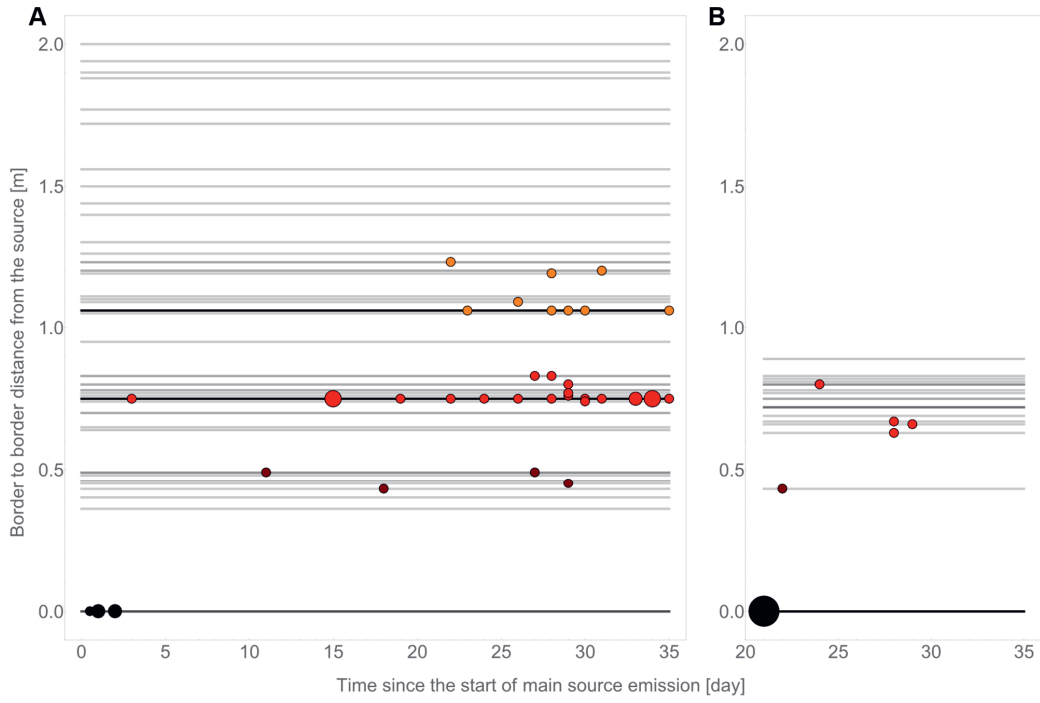


Figure 2.1. Results from experiments on transmission of *C. jejuni* between spatially separated broilers; A) results from all type 1 experiments where exposure started the same day as source broilers (main source) were inoculated and group C from type 2 experiment that were housed with their pen-mate broiler (main source) the day that pen mate started shedding *C. jejuni*; B) results from recipients from type 2 experiment where exposure started 20 days after inoculation of source broilers (main source) for group A (housed in pens distanced from the main source) and B (housed in the same pens as main source broilers); to present the experimental outcome in a clear and intuitive way we simplified and aggregated the data; detailed experimental outcome is provided in supplementary materials; point size is scaled with the case number for that particular time and distance, and colour indicates one of 4 distance bins (black: 0m, dark red: 0.35-0.60 m, red: 0.61-1.00 m, orange: 1.01-1.30 m); grey lines mark the distances of all individual pens used in experiments; pens that housed both source and recipient hosts are shown as distanced by 0 m; line colour is scaled with number of pens for that particular distance (darker for more pens).

For non-zero distances between source and recipient, the experiments displayed a time delay between start of recipient exposure until onset of recipient infections, and this delay time increased with distance. As expected, the delay observed when recipient hosts were placed in an already contaminated environment (type 2 experiment, group A) was shorter than the delay observed for the same distance range when the environment was not contaminated prior to exposure (type 1 experiments). For the pairwise experimental groups (group B and C), where we studied transmission between broilers that were housed together (separated by 0 m distance) no delay in onset of transmission was observed (as expected). For group B that was exposed to an environment previously contaminated by their pen-mates (being the main source) transmission was faster than for group C where there was no previous contamination by the infected pen mate (being the main source), as exposure to the main source started on the same day that this pen mate started shedding.

Parameter estimates

First, we estimated all three parameters (by maximum-likelihood estimation) through a model fit to the data from three previously published type 1 experiments where only a single distance band was studied(van Bunnik et al., 2012; van Bunnik et al., 2014). The estimated decay rate parameter α was 0.000 day^{-1} (CI: 0.000- 0.083), the transmission rate parameter β was 0.008 day^{-1} (CI: 0.005- 0.027), and the diffusion coefficient D was 0.089 m^2day^{-1} (CI: 0.026- 0.826). The profile likelihood plots are provided in Supplementary Note 2. The decay rate parameter α estimated to be 0 day^{-1} means that the decay time is estimated to be indefinitely long, i.e. the parameter is unidentifiable. Therefore, we conclude that the three-parameter model is non-identifiable with the previously published data of only a single distance band. Also, the value α=0 day^{-1} is clearly nonbiological as even for pathogens with strong survival in the environment a finite survival time is expected.

Next, we estimated all three parameters through a model fit to the joint data of the transmission experiments, thus additionally including the new experimental data (one type 1 and one type 2 experiment) studying transmission for varying distance bands. The estimated decay rate parameter α was 0.153 day^{-1} (CI: 0.072- 0.295), the transmission rate parameter β was 0.372 day^{-1} (CI: 0.125- 0.989) and the diffusion coefficient D was 0.013 m^2day^{-1} (CI: 0.008- 0.023). Univariate profile likelihoods for all three parameters are shown in Fig. 2.2. Finite confidence intervals estimated for parameters show that all three parameters were separately identifiable.

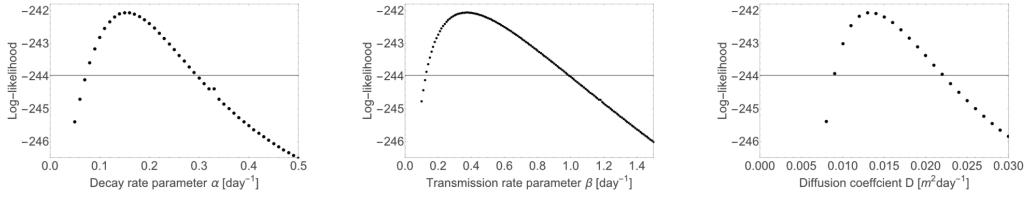


Figure 2.2. Profile likelihoods for model parameters: decay rate parameter α, transmission rate parameter β and diffusion coefficient D, horizontal lines mark the likelihood value for the confidence bounds.

Fit to experimental data

To validate the quality of our model fit to the data we assessed the fit across both the spatial and temporal dimension by using 20 and 5 spatiotemporal bins for type 1 and type 2 experiment, respectively. For each spatiotemporal bin we calculated the total number of positive cases observed during experiments and compared these to the probability mass function for number of cases as predicted by the model. The results for the type 1 experiments are presented in Fig. 2.3, the results for the type 2 experiments are presented on Fig. 2.4 (group A) and 2.5 (group B & C).

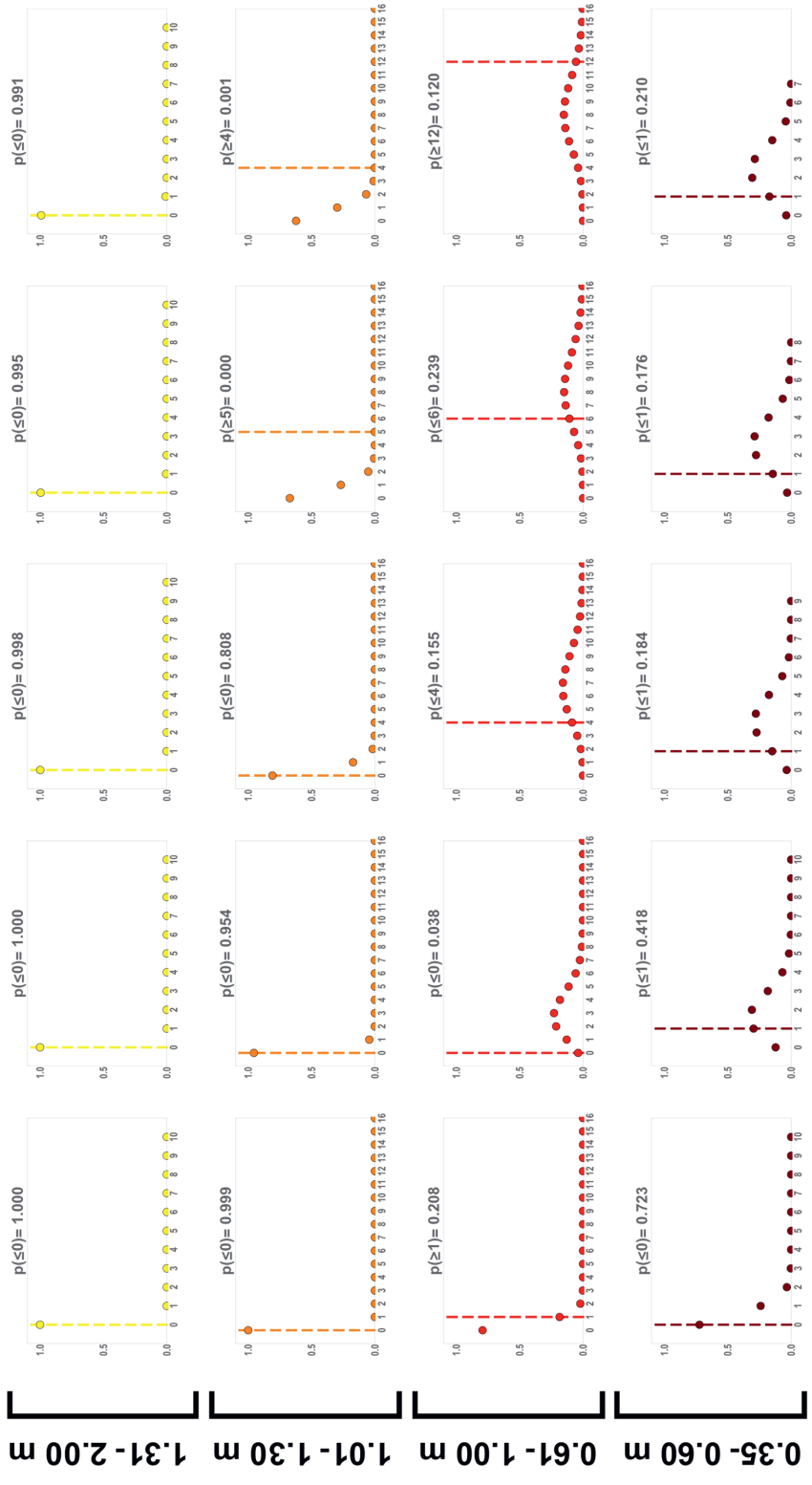


Figure 2.3. Probability mass functions generated from model predictions for type 1 experiments, representing total number of cases per 1 spatiotemporal bin of 1 week for pens grouped into four distance bins: 0.35-0.60 m, 0.61-1.00 m, 1.01-1.30 m, 1.31-2.00 m; on the x axis is the number of positive cases observed during a 1-week interval, and the y axis shows the probability. The vertical line marks the total number of cases observed for the particular bin in the experiments.

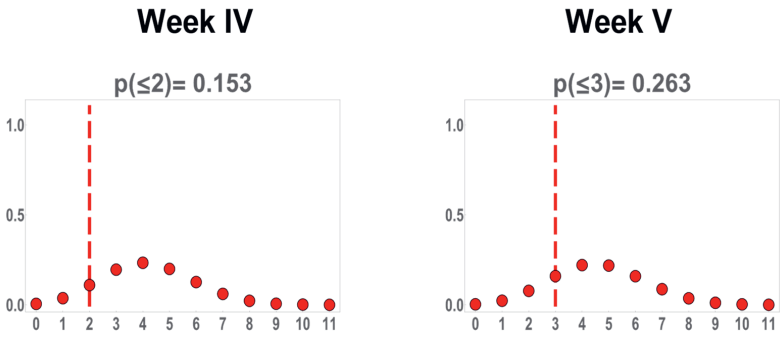


Figure 2.4. Probability mass functions generated from model predictions in type 2 experiment for group A representing total number of cases per week; on the x axis is the number of positive cases observed during a 1-week interval, and the y axis shows the probability. The vertical line marks the total number of cases observed for the particular bin in the experiments.

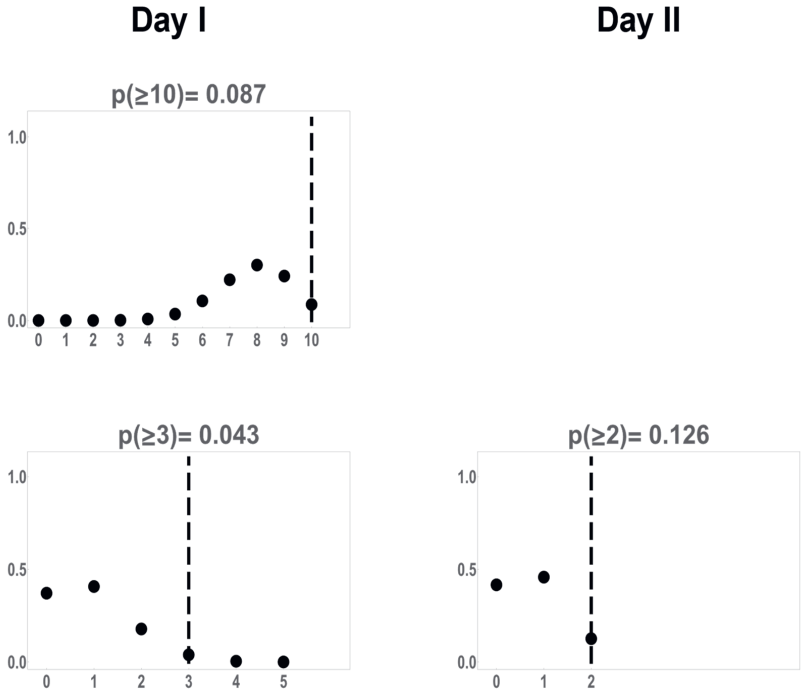


Figure 2.5. Probability mass functions generated from model predictions in type 2 experiment for group B (upper row) & C (lower row), representing total number of cases per day; on each plot the x axis is the number of positive cases observed during a 1-day interval, and the y axis shows the probability. The vertical line marks the total number of cases observed for the particular bin in the experiments.

We calculated a p-value for each spatiotemporal bin as the total model probability for obtaining the observed number of cases or more extreme values. For almost all spatiotemporal bins p-values are higher than 0.025. In type 1 experiments, for distance range 1.01-1.30 m, for the last two spatiotemporal bins significantly more cases were observed in the experiments than the model predicts ($p<0.025$), because of that the Fisher's combined probability test performed for all 25 spatiotemporal bins indicates significant difference between model fit and experimental data ($p=0.000665$). However, once p-value for those two spatiotemporal bins were removed, the combined p-value indicates overall good fit ($p=0.166345$). For a distance bin 0 m (group B &C in type 2 experiment), the fit for individual bins is consistent with experimental data ($p>0.025$), but the observed outcome for all three bins is located in the right tail of distribution and the Fisher's combined probability test performed for them shows a significant difference between data and model for these 3 bins ($p=0.017891$). Detailed results from Fisher's combined probability test are provided in Supplementary Note 1.

Discussion

Here, we presented a mechanistic model, based on the biological hypotheses of pathogen dispersal and decay, that has three parameters which are identifiable with transmission data only. Previously, a four-parameter version of the model was used (of which the decay rate was estimated independently from a survival experiment)(van Bunnik et al., 2014). The fourth parameter called “exposure capacity” was introduced based on the assumption that there is a limiting value of the amount of infectious material to which a host can be exposed. For modelling it means that when the exposure capacity is reached the probability of infection is constant even if accumulated amount of material increases. As there is no clear biological hypothesis underlying this assumption and we decided to start the analysis of distance dependence from the simplest mechanistic model possible, in the current work we used a model without this additional parameter. In the current model the exposure is still constrained, just as an emerging property of the balance between shedding of material on one hand and decay and diffusion on the other.

Collectively, observations from these experiments enabled calibration of the model by obtaining point estimates and finite confidence intervals for all three parameters- decay rate parameter α, diffusion coefficient D and transmission rate parameter β using transmission data only. In the previous analysis (van Bunnik et al., 2014), the decay rate was estimated from a separate survival experiment. As we show here, identifying it together with the other two parameters from transmission data only was not possible because only one distance band was studied. This is an example of a practical unidentifiability (as opposed to structural identifiability), as it arises because the data used for parameter information is insufficient (Wieland et al., 2021).

As a main solution to problems with identifiability it is often recommended to take separate measurements (e.g. for exposure) and/or conduct separate small scale experiments to estimate some

of the parameters independently from transmission data (Eisenberg et al., 2013). While sometimes it is the only possible way to obtain the estimates, it can also be problematic, especially for processes that are poorly understood.

The survival of *C. jejuni* is an example of such a poorly understood process; better understanding of the survival and transmission of *C. jejuni* in poultry was recognised as an important knowledge gap for farm level control of *C. jejuni* (Hansson et al., 2018). As many survival mechanisms could be involved, such as biofilm formation and/or viable but non-culturable states (Murphy et al., 2006), it is difficult to assess the limitations of designing the separate survival experiment to mimic the exact conditions of animal transmission experiments. For that reason, one of the objectives of this study was to estimate decay rate (together with remaining parameters) from transmission data only. Previously, the decay rate was estimated in the separate survival experiment to be α=2.25 day^{-1} by fitting exponential decay curve to enumeration of *C. jejuni* in faeces (van Bunnik et al., 2014). Now, based on transmission data only we find a decay rate estimate that is significantly different: α=0.153 day^{-1} (CI: 0.072-0.295). Also, our current model (when fitted with fixed α=2.25 day^{-1}) is not able to reproduce the observed delay in transmission and its AIC value (503.169) is much higher than the one obtained while fitting all three parameters (490.125) (see Supplementary Note 3 for details). This difference between estimates is consistent with several hypotheses. During the separate survival experiment, the authors mimicked environmental (climate) conditions of transmission experiments and were careful to use the same type of infected animals to produce the infectious material and use the same bacterial strain; however, the estimation was based on enumerations of culturable bacteria made from samples of faeces (van Bunnik et al., 2014). While only culturable forms of *C. jejuni* were measured in the experiment, there is some support for the hypothesis that viable but non-culturable forms can also successfully colonize chickens (Battersby et al., 2016; Bronowski et al., 2014; Cappelier et al., 1999). Moreover, it is not well studied in what form *C. jejuni* contaminated material is dispersing spatially; a faecal sample may not be representative for the actual infectious particles that diffuse in the environment and thus were studied in transmission experiment. Most likely infectious material is dispersed in smaller particles, for example with dust or water droplets. As these small particles are likely to have different micro-conditions than big particles of faeces, the decay rate of *C. jejuni*, even in similar environmental (macro)conditions, may be different for these two situations. Additionally, when fitting an exponential decay curve to survival data, the assumption is made that each particle has the same mean probability for decay. It is possible that particles that contribute significantly to transmission between separated broilers, survived in the environment much longer than the average. Further study is needed to explore which hypotheses (or combination thereof) can explain the observed difference. From a modelling perspective it could be addressed as a biphasic environmental decay (Brouwer, Eisenberg, et al., 2017), while from an

experimental perspective methods for detection of VBNC could be included (Lv et al., 2020) to study decay in various conditions.

The spatial dispersion of infectious material produced with faeces is another poorly understood process that we aimed to understand better. Although for some diseases with faecal-oral transmission dispersion between separate hosts was studied experimentally (Holt et al., 1998; van Bunnik et al., 2012; van Bunnik et al., 2014), the specific mechanisms of pathogen dispersion are not well understood. In our mechanistic model the diffusion coefficient describes how the spatial distribution of particles changes in time as the diffusion is one of the simplest ways to mechanistically describe dispersal. The mechanism underlying standard diffusion is a random walk of individual particles and if many particles are involved, the resulting distribution can be interpreted as a relative amount of material present at each location (Chandrasekhar, 1943). The estimate for the diffusion coefficient we obtained (0.013 $m^2 day^{-1}$; CI: 0.008-0.023) as well as a delay of transmission observed in our experiments (and consistently in the fitted model) indicates that the dispersion of infectious material is relatively slow for *C. jejuni*. Most likely the spatial dissemination of material is a multistep process, influenced by factors such as caretaker movements, birds behaviour (e.g. wing flapping, water spilling), ectoparasites or ventilation. During our experiments, no directional pattern was observed that could be explained by air flow (controlled in our experimental rooms), order of sampling (with inoculated broilers always being sampled last) or other caretaker actions (that were preformed following strict hygienic measures; see detailed description of experiments provided in Methods section). Collectively, our experimental observations indicate that a direct (one-blow) air-borne transfer of *C. jejuni* carrying particles is unlikely.

While both the decay rate parameter α and the diffusion coefficient D used in the current version of the model describe single, well defined biological processes, interpretation of the transmission rate parameter β is less straightforward. It is in fact the combination of parameters describing three host dependent processes: 1) a shedding rate describing the amount of material that is produced by the infectious host; 2) an exposure rate describing how recipient hosts contact the infectious material dispersed and accumulated in the environment, and 3) a probability of becoming colonised per one unit of ingested dose describing dose response. A way to ease the interpretation of the transmission rate parameter in our model, and to aid comparison between systems, is by bringing the infectious unit on the same footing as in direct transmission models. In a direct transmission model the exposure unit is usually defined as the exposure due to one infectious individual across a timestep Δt (usually of one day) and this exposure is taken proportional to $\frac{1\times\Delta t}{N}$ (N being the number of hosts). In models describing transmission through the environment the infectious unit can still be defined such that it corresponds to one infectious individual, according to the following recipe: starting from a well-defined state of the environment (equilibrium or clean) at the beginning of a timestep Δt, one integrates the environmental load originating from a single infectious individual (source host) across the timestep Δt and across the

recipient exposure area. Mathematical formalism of the standardization should be developed further in future studies.

Given that our model has only three estimable parameters, we believe that the model fit demonstrates a very good description of the observations. The delay of transmission observed in the experiments for all distanced ranges is consistent with the model fit. In addition, the model also describes satisfactorily that the delay was much shorter when recipient hosts were placed in already contaminated environment (type 2 experiments, groups A and B) that when experiment started from clean environment (type 1 experiments and group C of type 2 experiment). This all indicates that observed delay is a result of decay and diffusion dynamics rather than a result of overall low transmission rate (constant in time).

In the type 2 experiment we classified animals in direct contact as separated by distance 0m (therefore assuming only indirect transmission) as *C. jejuni* is mainly transmitted via faecal-oral route. Generally, for both experimental groups (A and B) the observed number of cases is slightly higher than the model predicts. This could be an indication that if recipient and source hosts are housed together additional routes of infections are present in comparison to situation when they are housed separately; for example sharing drinkers and feeders may contribute to transmission. We had only small sample size (10 broilers in group B and 5 in group C) for both pairwise groups, so more data should be collected to assess the effect size of potential additional sources on transmission.

Here, we used *C. jejuni* transmission in broilers as a host-parasite model system. The motivation for this was two-fold. First, *C. jejuni* is a zoonotic pathogen transmitted through the environment via the faecal-oral route that remains a major public health problem despite many control programs that have been applied (Efsa Panel on Biological Hazards et al., 2020). To further develop intervention strategies, a better understanding of its transmission dynamics is crucial. Second, *C. jejuni* transmission in poultry is a convenient system to study small scale environmental transmission. Fast transmission dynamics (extremely short incubation period, relatively short survival in environment) reduces time of experiments, clear manifestation (high, consistent shedding) facilitates the detection and sustainability of transmission chains; while low pathogenicity in broilers is important for animal welfare (Newell, 2002). As we show here, this model system is also suitable to study distance dependence of transmission: using practical between-host distances of up to 2 meters enabled us to combine observations ranging from extremely fast transmission (when hosts are housed together), via slower transmission (across middle distances) to absence of transmission (across longer distances).

By estimating all the parameters with transmission data (i.e. exposure and infection data) only, we were able to validate our mechanistic model and gain further insights into environmental transmission of *C. jejuni*. The model fit confirms that a delay (to onset) of distant transmission observed in experiments

can be explained by the dynamics of decay and diffusion of infectious material. The estimated diffusion rate indicates that dispersion of infectious material for faecal-oral diseases is rather slow. Moreover, estimating the decay rate together with other parameters opens new opportunities to study the decay process of pathogens. Generally, insights obtained from transmission data for validated models can be analysed together with the data from separate pathogen survival experiments to study further mechanisms of *C. jejuni* decay in animal systems.

We formulated a methodological framework based on a parsimonious, yet mechanistic model and proved that it can be calibrated and validated with spatiotemporal transmission data. This approach does not require any prior knowledge on detailed transmission routes which are often difficult to determine. Additionally, using our methodological framework the impact of efficient intervention strategies targeted on the environmental stage of transmission, such as cleaning regimens, can be assessed quantitatively through experimental study, supported by mechanistic modelling of transmission, dispersion and decay.

Methods:

Model

We constructed a spatiotemporal, mechanistic, yet parsimonious transmission model that takes into account the dispersion and decay of the assumed infectious material and for the probability of infection the dose response given the exposure (equation 3). Our model can be classified as individual based, where each host is described with the spatial coordinates of the area it occupies and from that the exposure for a particular period and recipient area (occupied by recipient hosts) is calculated based on the infectious period and source area (occupied by infectious hosts). The dispersion of material is described as a diffusion process, which is based on the assumption that particles can be modelled as moving according to a random walk (each particle is making random steps through the environment, where each step has a random direction and the step length is a normally distributed random variable). The decay is described using an exponential function, thus assuming that every infectious unit has the same probability of survival (Crane & Moore, 1986). The instantaneous rate of infection at any location and time is given as a mathematical expression and from that infection probabilities can be calculated in any spatiotemporal resolution of choice (equation 3). This makes the model easily adaptable to transmission systems that vary in spatial organisation or temporal dynamics of exposure.

Thereout these experiments, host position was restrained within pen boundaries, and hosts occupied areas separated from the source by a fixed distance, so the probability of infection can be modelled based on time independent source and recipient areas. Moreover, we assumed in the model that source areas emit infectious material continuously throughout the whole shedding period.

The definition of all parameters and variables is provided in table 2.2.

Table 2.2. Parameters and variables for the spatial model of environmental transmission.

Parameters		
α	[day^{-1}]	Decay rate parameter
β	[day^{-1}]	Transmission rate parameter
D	[m^{2}day^{-1}]	Diffusion coefficient
Configuration parameters		
Q	[m^{-2}day^{-1}]	Source strength
A^i_{inf}	[m^2]	Source area occupied by infectious host(s) *i*
A^r_{exp}	[m^2]	Exposure area occupied by recipient host(s) *r*
Temporal variables		
(T^i_1, T^i_2)	[day]	Emission period of the source area *i*
(t^r_1, t^r_2)	[day]	Exposure period of the recipient area *r*

The density function of environmental load generated by a given source area A^i_{inf} occupied by an infectious individual *i* is described by the following equation:

$$W^i(\text{t, x, y}, T^i_1, T^i_2, A^i_{inf}) =$$

$$\begin{cases} 0 & t < T^i_1 \\ Q \int_{T^i_1}^{t} \iint_{A^i_{\text{inf}}} \frac{1}{4\pi D(t-\tau)} \exp\left[-\alpha(t-\tau) - \frac{(x-x_i)^2+(y-y_i)^2}{4D(t-\tau)}\right] dx_i\, dy_i\, d\tau & T^i_1 \le t \le T^i_2 \\ Q \int_{T^i_1}^{T^i_2} \iint_{A^i_{\text{inf}}} \frac{1}{4\pi D(t-\tau)} \exp\left[-\alpha(t-\tau) - \frac{(x-x_i)^2+(y-y_i)^2}{4D(t-\tau)}\right] dx_i\, dy_i\, d\tau & T^i_2 < t \end{cases} \quad (4)$$

Where the W^i definition distinguishes three cases: 1) if the source is not yet emitting (time of observation is earlier than start of emission) W^i is equal to 0; 2) if the source has started continuous emission (time of observation is somewhere during the emission period) the environmental load is represented as diffusion with continuous source and decay and accounts for all the material that was released from the start of emission to the time of observation; 3) if the source has already stopped emitting (time of observation is after emission period) the W^i accounts for all the material that was released in the past during emission period and its decay and further diffusion up to the time of observation. We assume that the source area is emitting infectious material as soon as at least one host housed inside the area starts shedding. Q is a factor that scales the source strength level to account for differences in density of infectious host per square meter of area, for example in situations when infectious host are occupying areas of various sizes.

For a given recipient area A^r_{exp} that is exposed to multiple sources (more than one source area present), the source term is implemented as a set $\{T^i_1, T^i_2, A^i_{\text{inf}}\}_I$, and equation 3 that describes the probability of infection for particular recipient host *r* turns into:

$$\mathrm{P}_{\text{inf}}\left(t^r_1, t^r_2, A^r_{\text{exp}}, \{T^i_1, T^i_2, A^i_{\text{inf}}\}_I\right) = 1 - \mathrm{Exp}\left[-\beta \sum_I \int_{t^r_1}^{t^r_2} \iint_{A^r_{\text{exp}}} W^i(\mathrm{t, x, y}, T^i_1, T^i_2, A^i_{\text{inf}})\, dA^r_{\text{exp}}\, dt\right] \quad (5)$$

where $\sum_I$ is the sum over the set of source areas.

Experimental data

Spatiotemporal data used to fit the model were obtained in series of experiments on indirect transmission of *C. jejuni* in broilers. In total, data from five experiments, each with multiple groups, were included. The animal experiments and associated procedures were in accordance with the national regulations on animal experimentation and the project licenses were approved by the Dutch Central Authority for Scientific Procedures on Animals (CCD) (permit numbers for previously unpublished experiments: AVD4010020172784 for experiment 4; AVD4010020198586 for experiment 5). Here we follow the ARRIVE guidelines; the detailed description of previously unpublished experiments is provided below (in the *Detailed description of transmission experiments* section).

All experiments lasted 35 days and the status of broilers was recorded daily by testing cloacal swabs for the presence of *C. jejuni*. We used data from four type 1 experiments, where exposure of the recipient from the source hosts started in a clean (i.e. *C. jejuni* free) environment, and one type 2 experiment where recipient hosts were placed in experimental room 20 days after the source animals were inoculated with *C. jejuni*. The schematic spatial organisation of all experimental rooms is provided as supplementary figures (Fig. S2.7 to S2.15).

All type 1 experiments had a similar design. In this analysis we included rooms were five source broilers were orally inoculated with *C. jejuni* on day 0 and placed in one pen situated in the centre of the experimental room. On the same day, *C. jejuni* negative recipient broilers were individually placed in smaller recipient pens spatially separated from the source pen. For experiments 1-3 the spatial setup was similar, the source and recipient pens were approx. 0.75-1.06 m apart in each experimental room; as detailed between-pen distance measurements were not collected during those experiments, coordinates used in the model likelihood calculation were determined based on assumption that all rooms had identical, symmetrical design. For experiment 4 spatial setup was different, border to border distance between pens ranged from 0.35 to 2.00 m and the pens were divided over 6 experimental rooms; for this experiment between-pen distances in all rooms were recorded such that pen coordinates could be calculated for each pen separately for use in the model likelihood calculation.

All rooms in the experiments 1-3 had 10 recipient pens. Experiments 1 and 2 were identical in design and each had animals housed in 2 separate experimental rooms. In experiment 3 animals were housed in 8 experimental rooms; in 4 rooms 5 inoculated broilers were used; in the remaining 4 rooms 20 inoculated broilers were used and these were excluded from analysis presented here. In 2 out of 4 of the included rooms, the broilers in the central pen were also inoculated with *E.coli*. It has been shown in (van Bunnik et al., 2014) that *E.coli* infection is not likely to influence *C. jejuni* results. In the current analysis, we therefore included the *C. jejuni* data of these rooms. In experiment 4 there were 6 experimental rooms, four with 10 recipient pens and two with 4 recipient pens (separated from the source by larger distances). In all type 1 experiments recipient broilers that were detected positive for *C. jejuni* during the experiment were immediately removed from the experimental room to make sure their contribution to the probability of infection was minimal.

The last experiment (experiment 5) was a type 2 experiment with a modified design. The initial conditions were similar to previous experiments: 5 broilers per experimental room were inoculated with *C. jejuni* and placed together in a central pen at day 0. On day 20, the central pen was divided into 5 adjacent sub-pens similar in size, each housing 1 of the inoculated broilers. On the same day, recipient broilers were placed in the experimental room. Three groups of recipients were distinguished: group A consisted of 'distant' recipients placed in ten pens separated from the central pen by 0.43- 0.89 m; each of these pens housed a pair of recipient broilers; Group B was a pairwise type of recipients for which *C. jejuni* free broilers were added to the 5 central sub-pens, so each sub-pen was housing 1 inoculated and 1 recipient broiler; this group was included to study the transmission when the distance between source and recipient host was assumed to be 0 m. Group C, classified after the experiment, was also a pairwise type of recipients, but housed in distanced pens; the recipient was classified to group C after their pen mate started shedding *C. jejuni* during the experiment (becoming the source of the infectious material).

Both experimental rooms in experiment 5 had a similar spatial setup and the size of the central pen was bigger than the size used in type 1 experiments; again all between-pen distances were measured. In each room 5 central sub-pens and 10 recipient pens were present. The source broilers in this experiment were also inoculated with *Salmonella enterica* serovar Enteritidis. Previous transmission experiments, where broilers were inoculated with the same bacterial strains, showed that it is unlikely that inoculation with *Salmonella* had an influence on *C. jejuni* transmission (Heres et al., 2004; Heres et al., 2003). In the current analysis, we used the *C. jejuni* data only, as we estimated only parameters for *C. jejuni* transmission. In contrast to previous experiments, recipient broilers that were detected positive for *C. jejuni* were not removed during the experiment.

Likelihood formulation and parameter estimation

To estimate parameters for the spatiotemporal model (with data from transmission experiments) several assumptions were made. We assumed that once any broiler is detected positive for *C. jejuni* it starts to shed pathogen continuously until the end of the study period. This was based on the fact that, in our experiments, all broilers, detected to be positive for *C. jejuni* once and not removed form experiment, remained positive for the rest of the experimental period. This assumption applies for all inoculated broilers in all experiments and all recipient broilers in experiment 5. For experiments 1-4 recipient broilers who became infected were removed immediately when detected positive; considering the fact that the shedding periods of those individuals were short and diffusion of *C. jejuni* is relatively slow, their contribution to the probability of infection up until the end of the study period was considered to be negligible and not included in likelihood function (see Supplementary Note 4 for details). These approximations allowed us to proceed with a simplified likelihood function.

In the model, sources of infectious material are described in terms of areas that *C. jejuni* shedding hosts occupy. For all experiments, the central pen housing inoculated broilers was considered source area since the day of inoculation, considering the short incubation time of *C. jejuni*; the source strength for all central pen areas (occupied by multiple source broilers) was modelled as constant in time. For the type 2 experiments, where positive recipient host were not removed, newly positive broilers were included into the model as new source areas from the day of detection onwards. As the source areas had a different size in type 1 as compared to type 2 experiments and additional source areas were present in the latter, we standardized the source for all experimental rooms by using the host density (number of broilers per square meter) as a source scaling factor for each particular source area.

In contrast to type 1 experiments, where recipient broilers were housed individually, in type 2 experiments each pen housed a pair of broilers. For model fitting we assumed that these paired broilers do not compete for infectious material and treat them as independent recipients.

To estimate the model parameters, we maximized the likelihood function given by:

$$\prod_{\left\{t_1^r, t_2^r, A_{\text{exp}}^r\right\}_{R\text{esc}}} \left[1 - \mathrm{P}_{\text{inf}}\left(t_1^r, t_2^r, A_{\text{exp}}^r, \left\{T_1^i, T_2^i, A_{\text{inf}}^i\right\}_I\right)\right] \cdot \prod_{\left\{t_1^r, t_2^r, A_{\text{exp}}^r\right\}_{R\text{inf}}} \left[\left(1 - \mathrm{P}_{\text{inf}}\left(t_1^r, t_2^r - 1, A_{\text{exp}}^r, \left\{T_1^i, T_2^i, A_{\text{inf}}^i\right\}_I\right)\right) \cdot \mathrm{P}_{\text{inf}}\left(t_2^r - 1, t_2^r, A_{\text{exp}}^r, \left\{T_1^i, T_2^i, A_{\text{inf}}^i\right\}_I\right)\right]. \quad (6)$$

Where $\left\{t_1^r, t_2^r, A_{\text{exp}}^r\right\}_{R\text{esc}}$ is a set of data of all recipients that escaped from infection, while $\left\{t_1^r, t_2^r, A_{\text{exp}}^r\right\}_{R\text{inf}}$ is a set of data of all recipients that were found positive during the experiment.

The data input system, automatic likelihood formulation, three-step maximization procedure, statistics performed for model validation and all addition calculations were programmed in Mathematica 12.0

(Wolfram Research, 2019); the code is available in Zenodo repository (Gamża, 2021). In detail, the log-likelihood function was formulated jointly for the spatiotemporal *C. jejuni* transmission data from all experimental rooms from all experiments. In the maximization procedure, all spatial integrals (integrals over source areas as well as recipient area) in the expression for the probability of infection were analytically solved using the Integrate function, which significantly reduced computation time. Temporal integrals were solved numerically using the "NIntegrate" function. The log-likelihood function was maximized for all 3 parameters using a three-step procedure. First, a univariate profile likelihood was calculated for the diffusion coefficient D using "NMaximize" function. We used the D profile likelihood as a first step in our optimization, because the maximization with fixed D was the fastest and most prone to find only global maxima. Second, based on the generated D profile likelihood, we manually set suitable constraints for all three parameters within which the "NMaximization" function maximized their values. These constraints were applied to speed up the calculations and prevent the "NMaximization ," function from returning only local maximum. Lastly, the univariate profile likelihoods were obtained for remaining two parameters (α and β), using the same procedure as for the parameter D, including the application of constraints when needed to prevent the "NMaximization" function from returning local maxima. Confidence bounds for parameters were obtained from the corresponding profile likelihoods using the likelihood ratio test; if needed additional points were calculated in regions near the bounds to achieve the desired accuracy of 0.001.

Statistical analysis of model fit

To validate the model, we aggregated model fits and experimental data into spatiotemporal bins. For type 1 experiments, we defined 4 distance bins (0.35- 0.60 m, 0.61- 1.00 m, 1.01- 1.30 m, 1.31- 2.00 m) and these were divided further into 5 spatiotemporal bins of 1 week each, which gave us total of 20 spatiotemporal bins. For type 2 experiment, in group A (distant recipients) there was only one distance bin because all but one of the pens were separated by 0.61- 1.00 m; one pen was separated by 0.43 m but was also included in the same bin for statistical analysis. The distance bin was divided into two spatiotemporal bins of 1 week. For group B and group C (pairwise groups) there was one distance bin of 0 m divided into 1 or 2 spatiotemporal bins of 1 day, as for these groups the dynamics of transmission was much faster.

For each spatiotemporal bin we compared the sum of positive cases observed in experiments to the probability of infection calculated from model fit obtained for parameter values that globally maximized the likelihood (point estimates). The probability of infection for each bin was calculated as probability of being infected anytime during this particular bin. In group A (distant recipients) in type 2 experiment, a pair of broilers occupied each pen, and the probability of infection was defined as probability of observing at least one of the two broilers becoming colonized. We note that the probabilities of being colonised (in general) differed between the individual pens belonging to the spatiotemporal bin, as the

probability varied depending on distance and position to the source and it was calculated using recipient (pen) areas coordinates. Therefore, the probability mass function (PMF) for number of colonised pens in each spatiotemporal bin was calculated using the Poisson Binomial distribution, being the distribution for the sum of independent Bernoulli distributed variables with varying p. In the calculations we used the discrete Fourier transform formula to calculate PMF (Hong, 2013), as in some of the bins the number of cases was (much) bigger than 10 and those were therefore non-computable from exact PMF formula.

Next, from the PMFs we calculated the p-value as the probability of observing the particular experimental outcome or more extreme values; for outcomes occurring in the first half of the distribution this meant integrating over the left tail of the distribution and for outcomes in the second half it meant integrating over the right tail. P-values lower than 0.025 were considered as significant.

To further diagnose the model fit we used Fisher's combined probability test to combine p-values generated for spatiotemporal bins.

Detailed description of transmission experiments

Results from experiments 1-3 were previously published, thus their description can be found in (van Bunnik et al., 2012) (experiments 1 and 2) and (van Bunnik et al., 2014) (experiment 3).

Description of experiment 4

Experimental Design

The experiment was carried in six experimental rooms. Five broilers inoculated with *C. jejuni* were housed together in one pen in the centre of every experimental room (a separate climate-controlled room in an experimental facility of the Wageningen Bioveterinary Research). In four rooms 10 recipient animals and in two rooms 4 recipient animals (all *C. jejuni* negative at the beginning of experiment) were housed individually in pens placed at various distances (see Fig. S2.7-S2.12 for spatial design and pen coordinates) from the central pen.

During the experiment, all source and recipient animals were sampled by means of a cloacae swab. To confirm their Campylobacter status, the swabs were tested for the presence of *C. jejuni*. If a tested recipient animal was found *C. jejuni* positive it was immediately removed from the experiment, euthanised and the cecum was removed and tested for C. jejuni. A broiler was considered infected when both swab and caecal sample were found positive for *C. jejuni*; this was the case for all positive recipients. The experiment ended 35 d post inoculation. All remaining source and recipient animals (that had not been found *C. jejuni* positive until that moment) were euthanized after which the cecum was removed, and a caecal sample was tested for *C. jejuni*.

Animals and housing

One-day-old broilers (type Ross 308, females) were obtained from a commercial hatchery. Only female broilers were used to prevent overgrowing of broilers and health problems attributed to that.

88 chicks were housed together in auxiliary room from day -14 to day -2. At day -6 and -2, cloacal swabs taken from each chick were tested to confirm the absence of *C. jejuni*.

On day -2, 78 chicks were randomly distributed to six experimental rooms for the transmission experiment. Remaining reserve broilers were euthanised. Four rooms contained 5 source animals housed together in one central pen and 10 recipient animals individually housed in 10 pens surrounding the central pen as shown in Fig. S7-S10. The other two rooms contained 5 source animals housed together in one central pen and 4 recipient animals individually housed in 4 pens surrounding the central pen as shown in Fig. S2.11-S2.12. All animals were housed on wood shavings and the drinking water was supplied through a nipple drinking system. In each setup, the drinking nipples in the pens on the long sides of the area were supplied from one common water container, and the central pen had a separate drinking water supply. This precluded transmission via a shared drinking water system. Before the start of the experiment, all experimental rooms were cleaned and disinfected with formaldehyde. Subsequently, samples were taken from inside the room to check for the absence of *C. jejuni*.

Inoculation

At day 0 source broilers were inoculated with 1 ml of the inoculum containing *Campylobacter jejuni*, dose $5.4*10^6$ CFU/ml applied to the crop. For inoculation with *C. jejuni*, the *C. jejuni* strain 356 was used. The strain was cultured in hearth infusion broth (WBVR BM332) microaerobically at 37 °C, overnight, and diluted in buffered peptone water to obtain the intended inoculation dose [$\sim 1 \times 10^6$ colony forming units (CFU)/mL]. The precise concentration (CFU per milliliters) of *C. jejuni* in the administered inoculum was determined by plating serial 10-fold dilutions on modified cefoperazone charcoal deoxycholate agar (mCCDA; WBVR BM322).

Sampling and detection

To check their status all broilers were tested by taking cloacae swabs. The sampling scheme is provided in Table 2.3.

Table 2.3. Sampling scheme for experiment 4

Day post inoculation	
-6	Cloacal swabs of all broilers
-2	Cloacal swabs of all broilers
0	Cloacal swabs of all broilers (before inoculation) Inoculation
0-6	Cloacal swabs of all broilers
7-11	Cloacal swabs of all recipients
12	Cloacal swabs of all broilers
13-18	Cloacal swabs of all recipients
19	Cloacal swabs of all broilers
20-25	Cloacal swabs of all recipients
26	Cloacal swabs of all broilers
27-32	Cloacal swabs of all recipients
33	Cloacal swabs of all broilers
35	Cloacal swabs of all recipients; Caecal samples from all remaining broilers.

Swab samples were collected using sterile swabs (sterile plain dry swabs; Copan Diagnostics, Inc.). For *C. jejuni*, swabs were directly plated on mCCDA (WBVR BM332), incubated microaerobically at 41.5 °C and examined for the presence of *C. jejuni* after 24 and 48 h. After streaking on mCCDA, the swab was placed in Preston enrichment medium [nutrient broth no. 2, Oxoid CM0067 with Campylobacter selective supplement (Oxoid SR0204E) and Campylobacter growth supplement (Oxoid SR0232E)], and incubated microaerobically at 41.5 °C for 24 h. After incubation, it was plated on mCCDA and incubated microaerobically at 41.5 °C, and examined for the presence of *C. jejuni* after 24 and 48 h. Confirmation of suspect colonies was done by MALDI (Biotyper®, Bruker).

At the end of experiment caecal samples were collected from all broilers and tested for *C. jejuni* using the same culturing methods.

Humane endpoints and euthanasia

Disease was not expected to arise as a consequence of infection, because the bacteria which were used, normally do not lead to clinical signs nor to mortality in broilers.

The humane endpoints were defined as follows: 1) not being able to stand up; 2) not being able to eat or drink; 3) severe depression (hardly any response on stimuli); 4) other severe discomfort. The broilers

were observed twice a day. During the experiment no animal reached any of the humane endpoints. Euthanasia methods are listed in table 2.4.

Table 2.4. Euthanasia methods for experiment 4.

Animal	Method
Broilers <250 g	cervical dislocation
Broilers 250-1000 g	sedation (Xylazine and Ketamine) followed by cervical dislocation
Broilers >1000 g	sedation (Xylazine and Ketamine) followed by T61 admission

For a sedation 10 mg/kg Xylazine and 30 mg/kg Ketamine was applied IM.

Hygienic measures

To prevent animal caretakers from acting as a vector of transmission between experimental rooms, strict hygienic measures were used during the entire experiment. Clean coveralls were used at every entry into the experimental rooms. A pair of boots was dedicated to each room, cleaned on entering and exiting it by means of wading through a chlorinated bath. To prevent direct transport from one bird to the next bird, sterile gloves were changed between handling individual animals. The same order of sampling and movement direction within a room during sampling and welfare checks was followed during the full experimental timespan. Inoculated animals were always sampled last.

Description of experiment 5

Experimental Design

Five groups of broilers were used in this experiment; half of each group was housed in experimental room 1 and half in room 2. Ten broilers from source broilers group were inoculated with *C. jejuni* and *Salmonella* Enteritidis. Five of inoculated animals were housed together in one pen in the centre of each experimental room from day 0 to day 20 (a separate climate-controlled room in an experimental facility of the Wageningen Bioveterinary Research). From day 20 onwards, central pen was divide into 5 sub-pens each was housing one broiler form source group and one from non-inoculated (*C. jejuni* negative) direct recipient group (10 in total); in both rooms 20 non-inoculated (*C. jejuni* negative) animals from indirect recipient groups were housed in pairs in 10 pens surrounding this central pen placed at various distances (see Fig. S2.13-S2.14 for spatial design and pen coordinates).

During the experiment, all source and recipient animals were sampled by means of a cloacae swab. To confirm their *C. jejuni* and *Salmonella* Enteritidis status, the swabs were tested for the presence of both bacteria. Recipients positive for *C. jejuni* were not removed from experimental room as also Salmonella transmission was studied during the experiment. Broiler was considered infected by *C. jejuni* when swab samples were found positive for three consecutive days, and assumed to be infected from the day when first positive sample was recorded. The experiment ended 35 d post inoculation. All source

and recipient animals were euthanized and the cecum was removed and caecal sample was tested for *C. jejuni*.

Animals and housing

One-day-old broilers (type Ross 308, females) were obtained from a commercial hatchery. Only female broilers were used to prevent overgrowing of broilers and health problems attributed to that.

After arrival at day -14 all 66 chicks were housed together in auxiliary room in the communal pen. At day -6 they were randomly assigned to experimental group: 10 to source group, 10 to direct recipient groups (I or II), 40 to indirect recipient groups (I or II) and 6 to replacement broiler groups (I or II).

Three rooms were used during experiment: 1 auxiliary room and 2 experimental rooms. The auxiliary room had 3 pens: one communal pen and two marked pens: I and II. The housing scheme for each group is described in table 2.5.

Table 2.5. Housing scheme for experiment 5

Day	
-14	All chicks placed together in auxiliary room in the communal pen
-6	Broilers randomly assigned to experimental group (source broiler, direct recipient I or II, indirect recipient I or II and replacement broiler I or II)
-2	Source broilers moved to 2 experimental rooms (5 broilers per room): housed together in one pen in the center of the room Remaining broilers moved to one of two marked pens in auxiliary room: groups with number I to pen I and groups with number II to pen II
20	The central pen divided into 5 sub-pens through placement of four meshes. The source broilers, housed before in the central pen, are divided into the 5 sub-pens. In each sub-pen a source broiler is housed together with one direct recipient taken from marked pen I or II. Indirect recipients placed in distanced pens in pairs: one indirect recipient from pen I and one from pen II.

All animals were housed on wood shavings and the drinking water was supplied through a nipple drinking system. In each setup, the drinking nipples in the pens on the long sides of the area were supplied from one common water container, and the central pen had a separate drinking water supply, afterday 20 each central sub-pen had separate water supply. This precluded transmission via a shared drinking water system. Before the start of the experiment, all experimental rooms were cleaned and disinfected with formaldehyde. Subsequently, samples were taken from areas inside the room to check for the absence of *C. jejuni* and *Salmonella* Enteritidis.

Additionally, from day 0 to day 20, a small pilot experiment was conducted to study the detection of *C. jejuni* from environmental samples. In both rooms the sampling board covered with wooden shavings

was placed in pseudo-pen between pens 2-3. The board was removed before any recipients were placed in experimental room.

Inoculation

At day 0 source broilers were inoculated with 1 ml of the inoculum which was a mixture (1:1) of *Campylobacter jejuni*, dose $6.0*10^5$ CFU/ml and *Salmonella* Enteritidis, dose $1.2*10^5$ CFU/ml applied to the crop.

For inoculation with *C. jejuni*, *the C. jejuni* strain 356 was used. The strain was cultured in Heart Infusion Broth (WBVR BM332) (microaerobically, 37 °C, overnight) and diluted in buffered peptone water to obtain the intended inoculation dose [$\sim 1 \times 10^6$ colony forming units (CFU)/mL]. The precise concentration (CFU per milliliters) of *C. jejuni* in the administered inoculum was determined by plating on modified cefoperazone charcoal deoxycholate agar (mCCDA; Oxoid CM 793) with selective supplement (Oxoid CM 155) before and after the inoculation of the animals.

For inoculation with *Salmonella* Enteritidis, a nalidixic resistant (MIC >128 mg/l) *Salmonella enterica* serovar Enteritidis phage type 4 was used. The strain was cultured on Heart Infusion Agar with 5% Sheep Blood (HIS, WBVR BM20) (37°C, overnight) and next diluted in buffered peptone water to obtain the intended inoculation dose [$\sim 1 \times 10^5$ colony forming units (CFU)/mL]. The precise concentration (CFU per milliliters) of *Salmonella* Enteritidis in the administered inoculum was determined by plating serial dilutions on HIS (WBVR BM20) before and after the inoculation of the animals.

Sampling and detection

To check their status all broilers were tested by taking cloacae swab. The sampling scheme is provided in Table 2.6.

Table 2.6. Sampling scheme for experiment 5

Day post inoculation	
-6	Cloacal swabs of all broilers
-2	Cloacal swabs of all broilers
0	Inoculation
1-6	Cloacal swabs of all source broilers
14	Cloacal swabs of all source broilers
20	Cloacal swabs of all broilers (before relocation)
21-24	Cloacal swabs of all broilers
25-27	Cloacal swabs of all recipients
28	Cloacal swabs of all broilers
34	Cloacal swabs of all recipients
35	Cloacal swabs of all broilers; Caecal samples from all broilers.

Swab samples were collected using sterile swabs one swab for each bacterium (sterile plain dry swabs; Copan F155CA). For *C. jejuni*, swabs were directly plated on mCCDA (WBVR BM332), incubated microaerobically at 41.5 °C for 48 h and examined for the presence of *C. jejuni* after 24 and 48 hours. After streaking on mCCDA, the swab was placed in Preston enrichment medium [nutrient broth no. 2, Oxoid CM0067 with Campylobacter selective supplement (Oxoid SR0204E) and Campylobacter growth supplement (Oxoid SR0232E)], and incubated microaerobically at 41.5 °C for 24 h. After incubation, it was plated on mCCDA and incubated microaerobically at 41.5 °C, and examined for the presence of *C. jejuni* after 24 and 48 h. Confirmation of suspect colonies was done by MALDI (Biotyper®, Bruker)

For *Salmonella* enteritidis, swabs were incubated in Buffered Peptone Water for 18 hours, from which 0,1 ml was plated on MRSV (WBVR 334), incubated at 37 °C for 24h and plated out on XLD and BGA, both added with 100 ppm naladixic acid (in-house prepared). After 24 hours of incubation, plates were examined for the presence of *Salmonella* Enteritidis. Suspected cultures were confirmed by MALDI typing (Biotyper®, Bruker).

At the end of experiment caecal samples were collected from all the broilers and tested for *Campylobacter* using the same culturing methods.

Humane endpoints and euthanasia

Disease was not expected to arise due to infection, because the bacteria which were used, normally do not lead to severe clinical signs nor to mortality in broilers. The humane endpoints were defined as follows: 1) not being able to stand up; 2) not being able to eat or drink; 3) severe depression (hardly any response on stimuli). The broilers were observed twice a day. During the experiment no animal reached any of the humane endpoints. Euthanasia methods are listed in table 2.7.

Table 2.7. Euthanasia methods for experiment 5

Animal	**Method**
Broilers <250 g	cervical dislocation
Broilers 250-1000 g	sedation (Xylazine and Ketamine) followed by cervical dislocation
Broilers >1000 g	sedation (Xylazine and Ketamine) followed by T61 admission

For a sedation 10 mg/kg Xylazine and 30 mg/kg Ketamine was applied IM.

Hygienic measures

To prevent animal caretakers from acting as a vector of transmission between experimental rooms, strict hygienic measures were used during the entire experiment. Clean coveralls were used at every entry into the experimental rooms. A pair of boots was dedicated to each room, cleaned on entering and exiting it by means of wading through a chlorinated bath. To prevent direct transport from one bird to the next bird, sterile gloves were changed between handling individual animals. The same order of

sampling and movement direction within a room during sampling and welfare checks was followed during the full experimental timespan. Inoculated animals were always sampled last.

Data and code availability

The authors declare that the input experimental data supporting the findings of this study are available within the paper and its supplementary information files, and in Zenodo with the identifier(s) [10.5281/zenodo.5565054](Gamża, 2021).

The authors declare that computer code developed in Mathematica 12 (Wolfram Research, 2019) supporting the findings of this study is available in Zenodo with the identifier(s) [10.5281/zenodo.5565054](Gamża, 2021) .

Acknowledgments and funding sources

We thank the animal caretakers from Wageningen Bioveterinary Research (WBVR) animal facilities for their assistance with the experiments; Jan Cornelissen, Reina van der Heide, Dylano Suanes Lopez, Monique Oosterhof and fellow technicians from Diagnostic Department and the Department of Bacteriology of WBVR for their work on the laboratory analysis; Evelien Germeraad and Jeanet van der Goot for assistance with the experimental design and application for project and experiment licences.

This research was funded through the One Health European Joint Program (OHEJP) on Monitoring the gut Microbiota and Immune Response to Predict, Prevent, and Control zoonoses in humans and livestock in order to minimize the use of antimicrobials (MoMIR-PPC), which received funding from the European Union's Horizon 2020 research. Additional funding was obtained from the Dutch Ministry of Agriculture, Nature and Food Quality (grant WOT-01-005-003) and from an Agri&Food partnership (grant AF-EU-18031).

Additional information

Authors declare no competing interests.

Supplementary Information

Understanding environmental transmission mechanisms: a parsimonious mathematical model validated with infection data from tailor-made experiments

Other supplementary materials for this manuscript include the following: Supplementary dataset 1

Supplementary Note 1: Fisher's combined probability test for spatiotemporal bins

Combined p-value for Fisher's combined probability test for p-values of all the bins from all experiments was: p=0.000637. Combined p-value for Fisher's combined probability test for p-values of all the bins from all experiments, except for p-values smaller than 0.0125 that were removed from the analysis (distance bin 3, week 4 & week 5) was: p=0.166430. P-values for all spatiotemporal bins and combinations of thereof are presented in Table S2.1 and S2.2.

Table S2.1. P-values for all spatiotemporal bins for type 1 experiments and p-values for the Fisher's combined probability test (FCP) calculated for combinations of bins; four distance bins were used: bin 1: 0.35-0.60 m, bin 2: 0.61-1.00 m, bin 3: 1.01-1.30 m, bin 4: 1.31-2.00 m.

	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**FCP**	**FCP with removal[1]**
Bin 4	1.000000	0.999861	0.998330	0.994730	0.990828	**1.000000**	**1.000000**
Bin 3	0.999352	0.954489	0.808321	0.000046	0.001221	**0.000192**	**0.997586**
Bin 2	0.207664	0.038413	0.154957	0.238544	0.119903	**0.024861**	**0.024861**
Bin 1	0.723084	0.418276	0.184279	0.175687	0.209769	**0.260686**	**0.260686**
FCP	**0.875264**	**0.399545**	**0.479563**	**0.000927**	**0.007696**	**0.004953**	
FCP with removal[1]	**0.875264**	**0.399545**	**0.479563**	**0.384611**	**0.286783**		**0.591502**

1) p values smaller than 0.025 removed from analysis (bin 3, week 4 & week 5)

Table S2.2. P-values for all spatiotemporal bins for type 2 experiment with p-values for the Fisher's combined probability test (FCP) calculated for combinations of bins.

	Week 4	**Week 5**	**FCP**
Group A	0.152642	0.263318	**0.169377**
	Day 1	Day 2	**FCP**
Group B	0.086540		**0.086540**
Group C	0.043229	0.125820	**0.033799**
FCP	**0.024648**	**0.125820**	**0.017891**

Supplementary Note 2: Model (un)identifiability from transmission data on a single distance band only

In (van Bunnik et al., 2014) data from *C. jejuni* and *E.coli* transmission experiments between broilers separated by either 0.75 or 1.06 m were used to estimate three out of four parameters for a model version with one additional parameter: exposure capacity. The remaining parameter, the decay rate parameter α was estimated from data obtained in a separate survival experiment in which the concentration of culturable forms of *C. jejuni* and *E.coli* was measured in faeces daily (van Bunnik et al., 2014). Up until the time of counting, the faeces were kept in the same environmental conditions as transmission experiment was conducted in. The diffusion coefficient was assumed to be the same for both bacteria.

To investigate if the three-parameter version of the model, i.e. the model that we present in the main manuscript, is identifiable with data of only a single distance band, we fitted our model to the previously published data of *C. jejuni*. We estimated all three parameters simultaneously from the data published in (van Bunnik et al., 2014) (including only data on *C. jejuni* transmission for experiments that started with 5 inoculated broilers) together with the data from experiments of the same design published earlier in (van Bunnik et al., 2012) (included to increase the sample size). The estimates for parameters were as follows: decay rate parameter α=0.000 day^{-1} (CI: 0- 0.083), the transmission rate parameter β=0.008 day^{-1} (CI: 0.005- 0.027), and the diffusion coefficient D=0.089 m^{2}day^{-1} (CI: 0.026- 0.826). Univariate profile likelihoods are shown in Fig. S2.1.

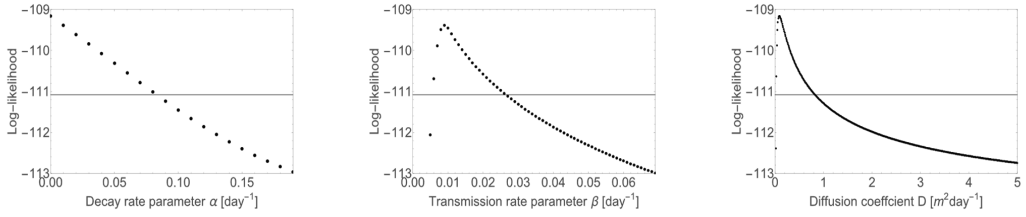


Figure S2.1. Profile likelihoods for model parameters estimated with previously published data on a single distance band only: decay rate parameter α, transmission rate parameter β and diffusion coefficient D, horizontal lines mark the likelihood value for the confidence bounds.

Supplementary Note 3: Model fit and validation for fixed decay rate α=2.25 day^{-1}

A decay rate value α=2.25 day^{-1} was estimated in a separate *C. jejuni* survival experiment, where faeces stored in environmental conditions similar to those used in transmission experiment were sampled and Campylobacter enumeration was performed. The estimate was obtained by fitting an exponential curve to temporal data. See (van Bunnik et al., 2014) for details.

For the model fitting here, based on setting α=2.25 day^{-1} and estimating the two remaining parameters from the transmission experiment data, the same methodology was used as presented in the main text for the estimation of all three model parameters, except that confidence bounds were calculated with lower accuracy.

The likelihood was maximized for two remaining parameters transmission rate was estimated to be β=4.32 day^{-1} [CI: 2.2-7.8] and diffusion rate was estimated to be D=0.15 m^{2}day^{-1} [CI: 0.10-0.23]. Univariate profile likelihood for both parameters are shown in Fig. S2.2.

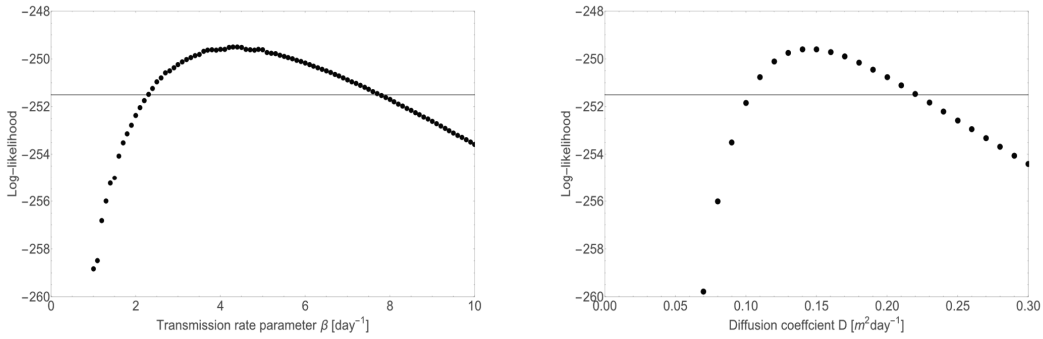


Figure S2.2. Profile likelihoods for model parameters for a model with fixed α=2.25 day-1: transmission rate parameter β and diffusion coefficient D; horizontal lines mark the likelihood value for the confidence bounds.

The Akaike information criterion (AIC) value was calculated form maximum loglikelihood values for this model with fixed α=2.25 day^{-1} (2 parameters; AIC: 503.169) and for original model where all three parameter were estimated together (3 parameters; AIC: 490.125). The difference in AIC is bigger than 10 which indicates that original model where all three parameters were estimated provides a significantly better fit (Burnham, 1998).To assess the fit the same methodology was used as presented in the main text for model with estimation of all three model parameters; the results for the type 1 experiments are presented in Fig. S2.3, the results for the type 2 experiments are presented on Fig. S2.4 (group A) and S2.5 (group B & C).

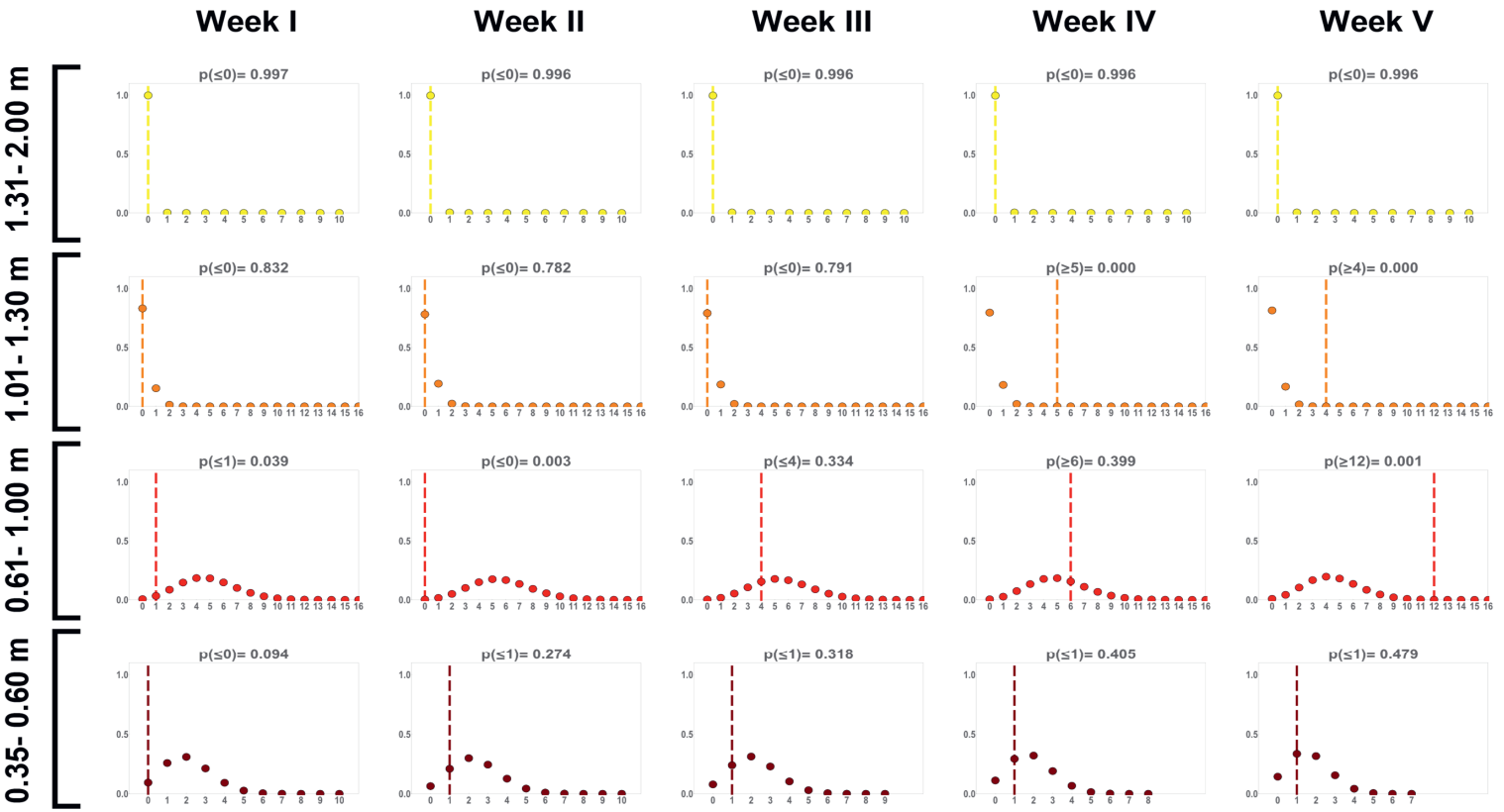


Figure S2.3. Probability mass functions generated from model predictions for model with fixed α=2.25 day^{-1} for type 1 experiments representing total number of cases per 1 spatiotemporal bin of 1 week for pens grouped into four distance bins: 0.35-0.60 m, 0.61-1.00 m, 1.01-1.30 m, 1.31-2.00 m; on the x axis is the number of positive cases observed during a 1-week interval, and the y axis shows the probability. The vertical line marks the total number of cases observed for the particular bin.

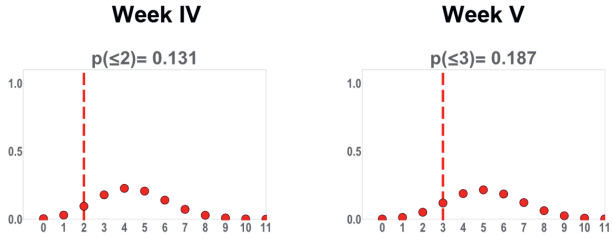


Figure S2.4. Probability mass functions generated from model predictions with fixed α=2.25 day^{-1} in type 2 experiment for group A representing total number of cases per week; on the x axis is the number of positive cases observed during a 1-week interval, and the y axis shows the probability. The vertical line marks the outcome observed in the experiments.

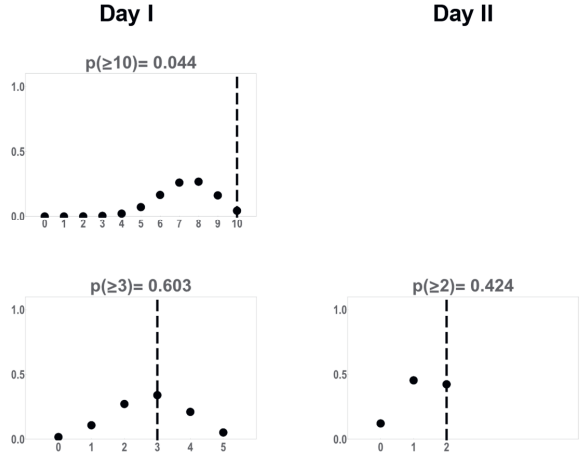


Figure S2.5. Probability mass functions generated from model predictions with fixed α=2.25 day^{-1} in type 2 experiment for group B & C (pairwise groups) representing total number of cases per day; on each plot the x axis is the number of positive cases observed during a 1-day interval, and the y axis shows the probability. The vertical line marks the outcome observed.

P-value for the Fisher's combined probability test for p-values of all the bins form all experiments was: 0.000637. P-values for all spatiotemporal bins and combinations of thereof are presented in Table S2.3 and S2.4.

Table S2.3. P-values for all spatiotemporal bins for type 1 experiments with p values for the Fisher's combined probability test (FCP) for model with fixed α=2.25 day^{-1}

	Week 1	Week 2	Week 3	Week 4	Week 5	FCP
Bin 4	0.997422	0.996178	0.996177	0.996178	0.996178	**1.000000**
Bin 3	0.831535	0.781862	0.791078	0.000003	0.000052	**0.000001**
Bin 2	0.039308	0.002585	0.334424	0.398640	0.001385	**0.000099**
Bin 1	0.094025	0.274093	0.318458	0.405202	0.479225	**0.230265**
FCP	**0.171188**	**0.059063**	**0.762311**	**0.000329**	**0.000035**	**0.000002**

Table S2.4. P-values for all spatiotemporal bins for type 2 experiment with p-values for the Fisher's combined probability test (FCP) for model with fixed α=2.25 day^{-1}

	Week 4	Week 5	FCP
Group A	0.130534	0.186576	**0.114832**
	Day 1	Day 2	**FCP**
Group B	0.044103		**0.044103**
Group C	0.603160	0.423824	**0.604320**
FCP	**0.123053**	**0.423824**	**0.175216**

Supplementary Note 4: Additional analysis

a) Experiment type 2 group C- contributions of main and additional sources

For group C from the type 2 experiment (pairwise transmission in recipient pens), using the fitted model with point estimate values for all three parameters α=0.153 day^{-1}, D=0.013 m^{2}day^{-1} and β=0.372 day^{-1} we evaluated the contribution to the probability of infection from the main source which is the pen mate of the recipient as well as the contribution from all the other sources (central area source and other recipient pens if they became infected before).

Table S2.5. Mean probability of infection estimated from our model with point estimate values of parameters (α=0.153 day^{-1} , D=0.013 m^{2}day^{-1} and β=0.372 day^{-1}) for pens in type 2 experiment: group B (n=10 pens) and group C (n=5 pens); additionally for group C the contribution to probability was estimated separately for main source (their pen mate) and additional sources (separated by distance).

		Day 1[1]	Day 2[1]	Day 3[1]
Group B	from all sources together	0.783854	0.816726	0.859933
Group C	from all sources together	0.179582	0.414533	0.591947
	from main source only	0.138976	0.317660	0.423870
	from additional sources only	0.039521	0.041396	0.043592

1) Day of exposure to the main source

b) Removed (potential) sources contribution for type 1 experiment

For type 1 experiments recipient broilers that were tested positive were removed immediately once cloacal swab collected from them was detected positive for Campylobacter; in modelling those broilers were not included as new sources as they were shedding only for a short period.

To estimate the effect that these additional sources may have on transmission we use point estimate values of the model parameters α=0.153 day^{-1}, D=0.013 m^{2}day^{-1} and β=0.372 day^{-1} to calculate what contribution these sources would make to the probability of transmission.

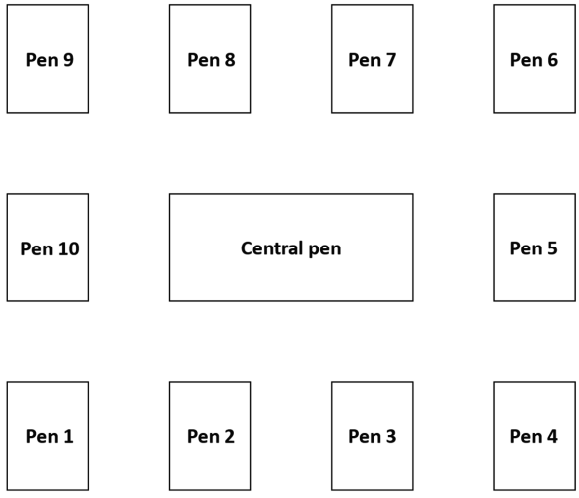


Figure S2.6. Schematic representation of experimental room with named pens from type 1 experiments 1-3.

As an example we consider one of the pens (pen 2, see Fig. S2.6) separated from the central pen and neighboring pens by 0.75 m (border to border distance) and calculate its contribution to the infection probability in the remaining 9 pens and compare to the contribution from the central pen (main source); the results are presented in tables S2.6 and S2.7.

Table S2.6. Probability of being infected by 1 source broiler housed in pen 2 that started shedding on day 1 for all remaining pens for 5 consecutive days.

	Day 1	Day 2	Day 3	Day 4	Day 5
Pen 1	$1.50 \cdot 10^{-10}$	$5.44 \cdot 10^{-07}$	$1.64 \cdot 10^{-05}$	$1.06 \cdot 10^{-4}$	$3.46 \cdot 10^{-4}$
Pen 3	$1.50 \cdot 10^{-10}$	$5.44 \cdot 10^{-07}$	$1.64 \cdot 10^{-05}$	$1.06 \cdot 10^{-4}$	$3.46 \cdot 10^{-4}$
Pen 4	0	0	0	$4.44 \cdot 10^{-16}$	$1.44 \cdot 10^{-13}$
Pen 5	0	0	0	0	$4.44 \cdot 10^{-16}$
Pen 6	0	0	0	0	0
Pen 7	0	0	0	0	0
Pen 8	0	0	0	0	$6.66 \cdot 10^{-16}$
Pen 9	0	0	0	0	0
Pen 10	0	$7.51 \cdot 10^{-12}$	$2.42 \cdot 10^{-09}$	$6.31 \cdot 10^{-08}$	$5.15 \cdot 10^{-07}$

Table S2.7. Probability of being infected by 5 source broilers housed in central pen that started shedding on day 1 for all remaining pens for 5 consecutive days.

	Day 1	Day 2	Day 3	Day 4	Day 5
Pen 1	0	$1.25 \cdot 10^{-11}$	$4.03 \cdot 10^{-09}$	$1.05 \cdot 10^{-07}$	$8.58 \cdot 10^{-07}$
Pen 3	$1.96 \cdot 10^{-10}$	$7.26 \cdot 10^{-07}$	$2.21 \cdot 10^{-05}$	$1.46 \cdot 10^{-04}$	$4.82 \cdot 10^{-04}$
Pen 4	0	$1.25 \cdot 10^{-11}$	$4.03 \cdot 10^{-09}$	$1.05 \cdot 10^{-07}$	$8.58 \cdot 10^{-07}$
Pen 5	$2.50 \cdot 10^{-10}$	$9.07 \cdot 10^{-07}$	$2.73 \cdot 10^{-05}$	$1.77 \cdot 10^{-04}$	$5.76 \cdot 10^{-04}$
Pen 6	0	$1.25 \cdot 10^{-11}$	$4.03 \cdot 10^{-09}$	$1.05 \cdot 10^{-07}$	$8.58 \cdot 10^{-07}$
Pen 7	$1.96 \cdot 10^{-10}$	$7.26 \cdot 10^{-07}$	$2.21 \cdot 10^{-05}$	$1.46 \cdot 10^{-04}$	$4.82 \cdot 10^{-04}$
Pen 8	$1.96 \cdot 10^{-10}$	$7.26 \cdot 10^{-07}$	$2.21 \cdot 10^{-05}$	$1.46 \cdot 10^{-04}$	$4.82 \cdot 10^{-04}$
Pen 9	0	$1.25 \cdot 10^{-11}$	$4.03 \cdot 10^{-09}$	$1.05 \cdot 10^{-07}$	$8.58 \cdot 10^{-07}$
Pen 10	$2.50 \cdot 10^{-10}$	$9.07 \cdot 10^{-07}$	$2.73 \cdot 10^{-05}$	$1.77 \cdot 10^{-04}$	$5.76 \cdot 10^{-04}$

c) Boundary conditions

To keep our model simple and easily computable, we assumed absorbing boundary conditions in our systems, treating each experimental room as an area of infinite size. In experiments, some pens in a room were placed near the wall, so theoretically it is possible that portion of infectious material that reached the wall stayed in the pen (instead of diffusing further) and influenced probability of infection. To assess how the boundary conditions assumption influenced our results, we compared the probability of being infected calculated for point estimate parameter values (α=0.153 day^{-1}, D=0.013 m^2day^{-1} and β=0.372 day^{-1}) for one, chosen recipient area from type 1 experiments 1-3 (pen 2, see Fig. S2.6) to that for a recipient area that is extended (virtually) in the direction of the room wall by an area of the same size (virtual mirror pen).

For both areas the probabilities estimated for day 1, 10, 20 and 30 when exposed to the main source area (central pen) are shown in table S2.8. The difference between these probability estimates is small, which supports our decision to use absorbing boundaries.

Table S2.8. Probability of being infected by 5 source broilers housed in central pen that started shedding on day 1 for one chosen recipient area (pen 2) from type 1 experiment 1-3 and for area of double size extended in the direction of the room wall.

	Day 1	**Day 10**	**Day 20**	**Day 30**
Original recipient area (0.75 m2)	$1.96 \cdot 10^{-10}$	0.006105	0.021043	0.027634
Extended recipient area (1.5 m2)	$1.96 \cdot 10^{-10}$	0.006109	0.021249	0.028208
Difference	0	$4.19 \cdot 10^{-06}$	0.000206	0.000574

Supplementary Figures S2.7 to S2.15: Spatial organization of experimental rooms

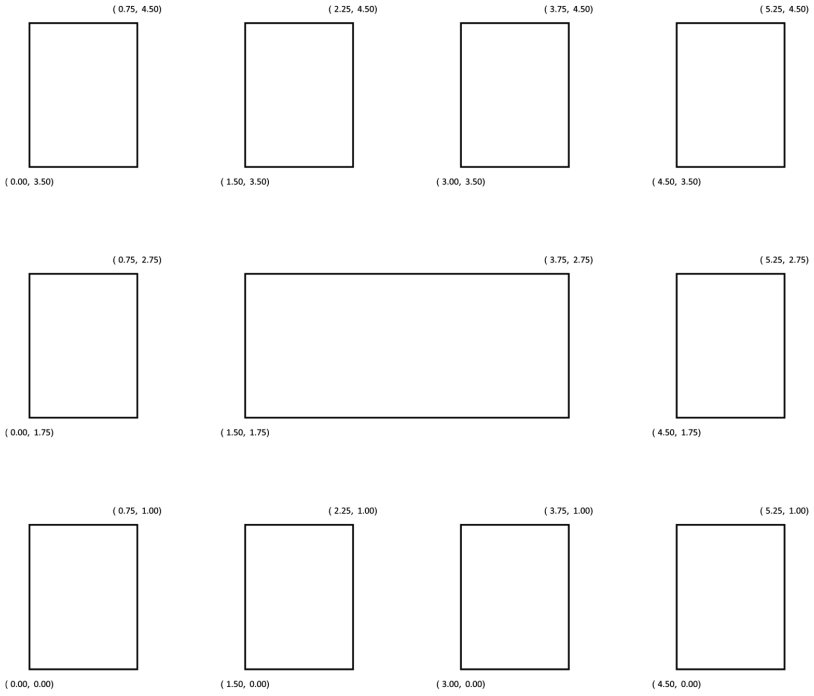


Figure S2.7. Schematic spatial organization for type 1 experiments 1-3, all rooms; as detailed between-pen distance measurements were not collected during those experiments, coordinates were prepared based on the assumption that all rooms had identical, symmetrical design.

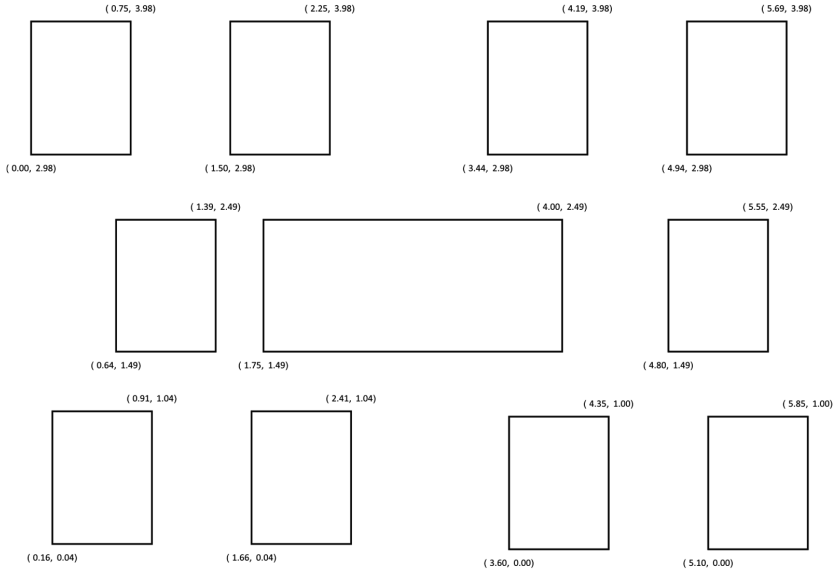


Figure S2.8. Spatial organization for type 1 experiment 4, room 1.

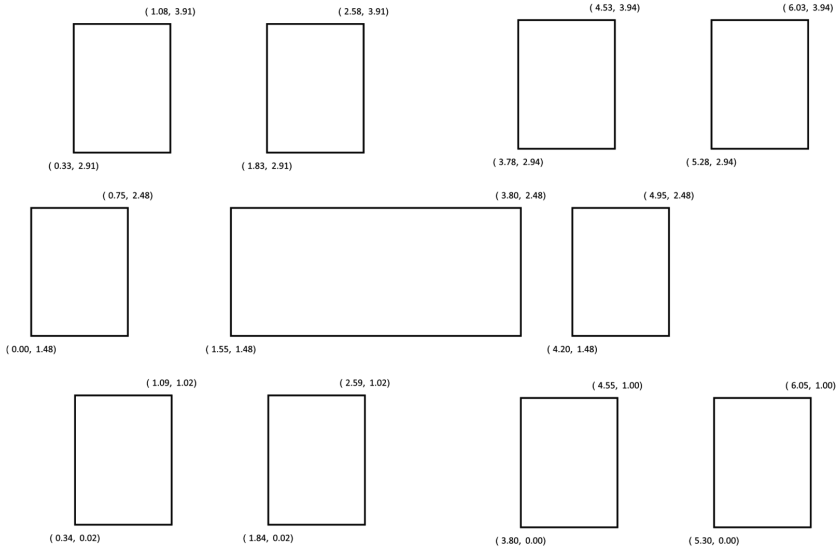


Figure S2.9. Spatial organization for type 1 experiment 4, room 2.

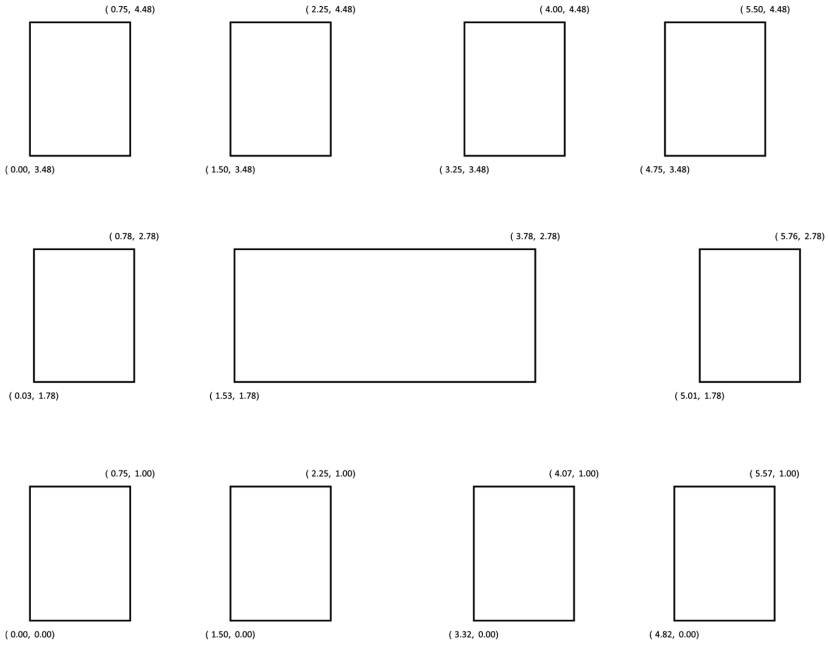


Figure S2.10. Spatial organization for type 1 experiment 4, room 3.

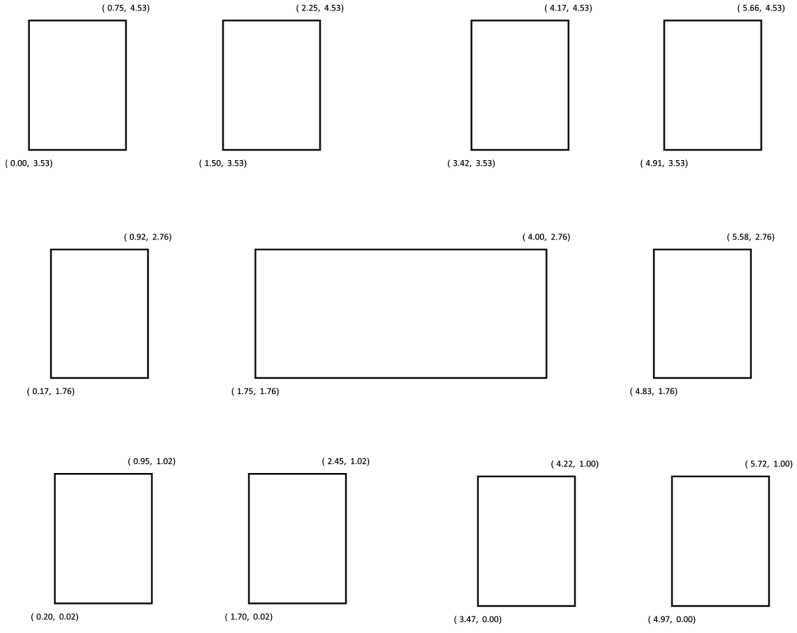


Figure S2.11. Spatial organization for type 1 experiment 4, room 4.

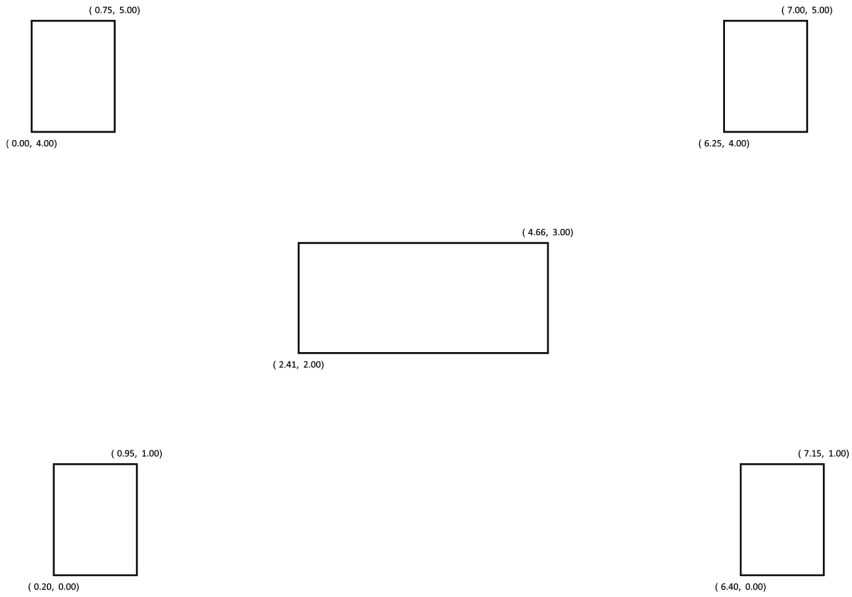


Figure S2.12. Spatial organization for type 1 experiments 4, room 5.

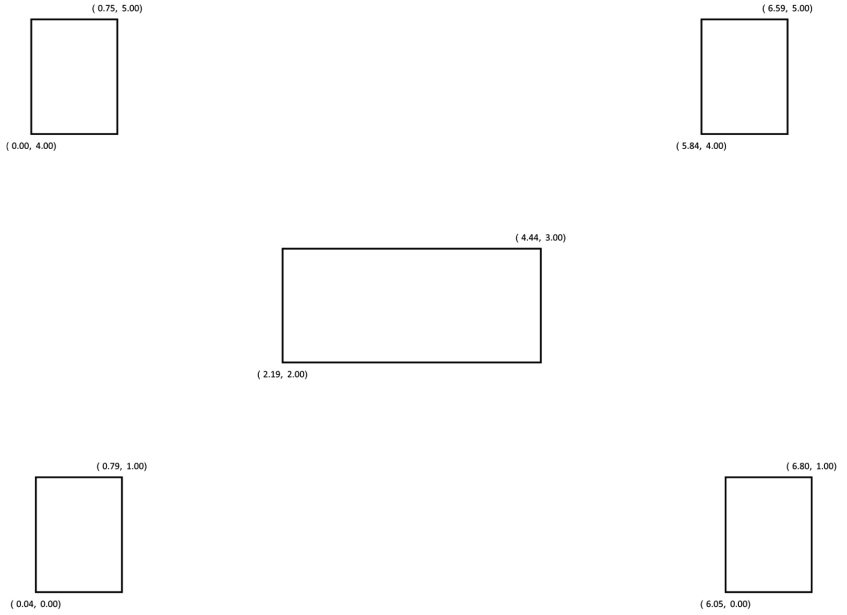


Figure S2.13. Spatial organization for type 1 experiment 4, room 6.

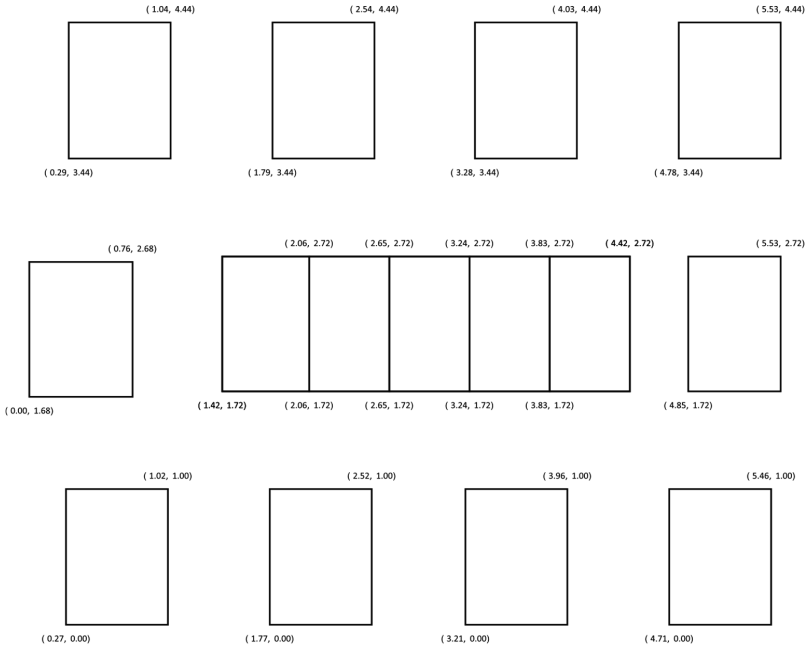


Figure S2.14. Spatial organization for type 2 experiment, room 1.

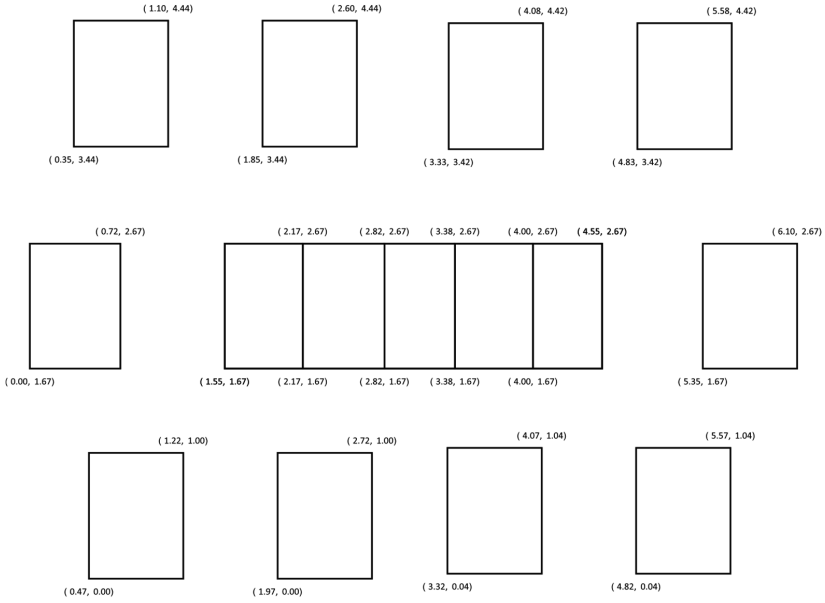


Figure S2.15. Spatial organization for type 2 experiment, room 2.

Dataset S1. Input data for modelling: spatiotemporal data for all recipient and source areas for all experiments: type 1 experiments 1-4 and type 2 experiment groups A-C.

12.10.2021

Contact: anna.gamza@wur.nl; thomas.hagenaars@wur.nl

Dataset coding

	Dataset	Experiment type	Experimental group(s)
	Type1exp_1&2_rec	Type 1	Recipient broilers
	Type1exp_1&2_source	Type 1	Inoculated source broilers
	Type1exp_3_rec	Type 1	Recipient broilers
	Type1exp_3_source	Type 1	Inoculated source broilers
	Type1exp_4_rec	Type 1	Recipient broilers
	Type1exp_4_source	Type 1	Inoculated source broilers
	Type2exp_groupsAB&C_rec	Type 2	Recipient broilers; groups A, B, C
	Type2exp_source	Type 2	Inoculated source broilers

Columns description

	Collumn	Description	Remarks
For recipient datasets:			
	Room nb	Room number	
	Pen nb	Pen number	
	Pen coordinates		
	Exposure start	Start of the exposure period	
	Exposure end	End of exposure period	Exposure period ends when recipient is found positive or removed from experiment
	Removal time		For the type 1 experiments the removal date of recipients was set equal to their infection date, as any shedding of infected recipients between infection and detection and removal was neglected in the model.
	S count	Number of recipient (susceptible) hosts	
	C count	Number of positive cases when exposure period ended	
For source datasets:			
	Room nb	Room number	
	Pen nb	Pen number	
	Pen coordinates		
	Shedding start	Start of the shedding period	For all experiments the start of shedding of inoculated broilers was set to be on the day of inoculation, any delay in shedding was neglected in the model to keep the source strength for those source areas constant in time.
	Shedding end	End of shedding period	
	Removal time		
	I count	Number of inculated broilers	

Further remarks

	Experiment	Remark
	Type1exp_1&2 and Type1_exp_3	Detailed between-pen distance measurements were not collected; coordinates were determined based on assumption that all rooms had identical, symmetrical design
	Type2exp_groupsAB&C_rec	In room 2 pen 10, both recipients were found infected on the same day; as the latent period for Campylobacter in broilers is extremely short, most likely one of the recipients was a source to the other, so we made the assumption that one of these recipients became infected 0.5 days after the other.
	Type1exp_4_source	In room 2 one inoculated broiler died shortly after inoculation; any shedding of this broiler was neglected, i.e. we assumed that throughout the experimental period, 4 instead of 5 inoculated broilers acted as a source in this room.

Type1exp_1&2_rec							
Room nb	Pen nb	Pen coordinates	Exposure start	Exposure end	Removal time	S count	C count
1	1	{0., 0., 0.75, 1.}	0	23	23	1	1
1	2	{1.5, 0., 2.25, 1.}	0	15	15	1	1
1	3	{3., 0., 3.75, 1.}	0	15	15	1	1
1	4	{4.5, 0., 5.25, 1.}	0	28	28	1	1
1	5	{4.5, 1.75, 5.25, 2.75}	0	15	15	1	1
1	6	{4.5, 3.5, 5.25, 4.5}	0	29	29	1	1
1	7	{3., 3.5, 3.75, 4.5}	0	26	26	1	1
1	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
1	9	{0., 3.5, 0.75, 4.5}	0	20	20	1	0
1	10	{0., 1.75, 0.75, 2.75}	0	22	22	1	1
2	1	{0., 0., 0.75, 1.}	0	30	30	1	1
2	2	{1.5, 0., 2.25, 1.}	0	30	30	1	1
2	3	{3., 0., 3.75, 1.}	0	35	35	1	0
2	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
2	5	{4.5, 1.75, 5.25, 2.75}	0	35	35	1	0
2	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
2	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
2	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
2	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
2	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	0
3	1	{0., 0., 0.75, 1.}	0	35	35	1	0
3	2	{1.5, 0., 2.25, 1.}	0	3	3	1	1
3	3	{3., 0., 3.75, 1.}	0	35	35	1	0
3	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
3	5	{4.5, 1.75, 5.25, 2.75}	0	28	28	1	1
3	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
3	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
3	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
3	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
3	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	0
4	1	{0., 0., 0.75, 1.}	0	35	35	1	0
4	2	{1.5, 0., 2.25, 1.}	0	35	35	1	0
4	3	{3., 0., 3.75, 1.}	0	35	35	1	0
4	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
4	5	{4.5, 1.75, 5.25, 2.75}	0	35	35	1	0
4	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
4	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
4	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
4	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
4	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	0

Type1exp_1&2_source						
Room nb	Pen nb	Pen coordinates	Shedding start	Shedding end	Removal time	I count
1	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
2	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
3	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
4	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5

Type1exp_3_rec							
Room nb	Pen nb	Pen coordinates	Exposure start	Exposure end	Removal time	S count	C count
1	1	{0., 0., 0.75, 1.}	0	8	8	1	0
1	2	{1.5, 0., 2.25, 1.}	0	34	34	1	1
1	3	{3., 0., 3.75, 1.}	0	34	34	1	1
1	4	{4.5, 0., 5.25, 1.}	0	12	12	1	0
1	5	{4.5, 1.75, 5.25, 2.75}	0	34	34	1	1
1	6	{4.5, 3.5, 5.25, 4.5}	0	33	33	1	0
1	7	{3., 3.5, 3.75, 4.5}	0	13	13	1	0
1	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
1	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
1	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	1
2	1	{0., 0., 0.75, 1.}	0	35	35	1	1
2	2	{1.5, 0., 2.25, 1.}	0	33	33	1	1
2	3	{3., 0., 3.75, 1.}	0	35	35	1	0
2	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
2	5	{4.5, 1.75, 5.25, 2.75}	0	24	24	1	1
2	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
2	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
2	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
2	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
2	10	{0., 1.75, 0.75, 2.75}	0	33	33	1	1
3	1	{0., 0., 0.75, 1.}	0	35	35	1	0
3	2	{1.5, 0., 2.25, 1.}	0	21	21	1	0
3	3	{3., 0., 3.75, 1.}	0	35	35	1	0
3	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
3	5	{4.5, 1.75, 5.25, 2.75}	0	35	35	1	0
3	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
3	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
3	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
3	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
3	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	0
4	1	{0., 0., 0.75, 1.}	0	35	35	1	0
4	2	{1.5, 0., 2.25, 1.}	0	2	2	1	0
4	3	{3., 0., 3.75, 1.}	0	35	35	1	0
4	4	{4.5, 0., 5.25, 1.}	0	35	35	1	0
4	5	{4.5, 1.75, 5.25, 2.75}	0	35	35	1	0
4	6	{4.5, 3.5, 5.25, 4.5}	0	35	35	1	0
4	7	{3., 3.5, 3.75, 4.5}	0	35	35	1	0
4	8	{1.5, 3.5, 2.25, 4.5}	0	35	35	1	0
4	9	{0., 3.5, 0.75, 4.5}	0	35	35	1	0
4	10	{0., 1.75, 0.75, 2.75}	0	35	35	1	0

Type1exp_3_source						
Room nb	Pen nb	Pen coordinates	Shedding start	Shedding end	Removal time	I count
1	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
2	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
3	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5
4	N/a	{1.5, 1.75, 3.75, 2.75}	0	35	35	5

Type1exp_4_rec							
Room nb	Pen nb	Pen coordinates	Exposure start	Exposure end	Removal time	S count	C count
1	1	{0.16, 0.04, 0.91, 1.04}	0	35	35	1	0
1	2	{1.66, 0.04, 2.41, 1.04}	0	29	29	1	1
1	3	{3.6, 0., 4.35, 1.}	0	11	11	1	1
1	4	{5.1, 0., 5.85, 1.}	0	31	31	1	1
1	5	{4.8, 1.49, 5.55, 2.49}	0	29	29	1	1
1	6	{4.94, 2.98, 5.69, 3.98}	0	35	35	1	0
1	7	{3.44, 2.98, 4.19, 3.98}	0	27	27	1	1
1	8	{1.5, 2.98, 2.25, 3.98}	0	35	35	1	0
1	9	{0., 2.98, 0.75, 3.98}	0	35	35	1	0
1	10	{0.64, 1.49, 1.39, 2.49}	0	35	35	1	0
2	1	{0.34, 0.02, 1.09, 1.02}	0	35	35	1	0
2	2	{1.84, 0.02, 2.59, 1.02}	0	35	35	1	0
2	3	{3.8, 0., 4.55, 1.}	0	35	35	1	0
2	4	{5.3, 0., 6.05, 1.}	0	35	35	1	0
2	5	{4.2, 1.48, 4.95, 2.48}	0	35	35	1	0
2	6	{5.28, 2.94, 6.03, 3.94}	0	35	35	1	0
2	7	{3.78, 2.94, 4.53, 3.94}	0	35	35	1	0
2	8	{1.83, 2.91, 2.58, 3.91}	0	18	18	1	1
2	9	{0.33, 2.91, 1.08, 3.91}	0	35	35	1	0
2	10	{0., 1.48, 0.75, 2.48}	0	35	35	1	0
3	1	{0., 0., 0.75, 1.}	0	35	35	1	0
3	2	{1.5, 0., 2.25, 1.}	0	35	35	1	0
3	3	{3.32, 0., 4.07, 1.}	0	35	35	1	0
3	4	{4.82, 0., 5.57, 1.}	0	35	35	1	0
3	5	{5.01, 1.78, 5.76, 2.78}	0	35	35	1	0
3	6	{4.75, 3.48, 5.5, 4.48}	0	35	35	1	0
3	7	{3.25, 3.48, 4., 4.48}	0	35	35	1	0
3	8	{1.5, 3.48, 2.25, 4.48}	0	35	35	1	0
3	9	{0., 3.48, 0.75, 4.48}	0	35	35	1	0
3	10	{0.03, 1.78, 0.78, 2.78}	0	19	19	1	1
4	1	{0.2, 0.02, 0.95, 1.02}	0	26	26	1	1
4	2	{1.7, 0.02, 2.45, 1.02}	0	30	30	1	1
4	3	{3.47, 0., 4.22, 1.}	0	29	29	1	1
4	4	{4.97, 0., 5.72, 1.}	0	22	22	1	1
4	5	{4.83, 1.76, 5.58, 2.76}	0	28	28	1	1
4	6	{4.91, 3.53, 5.66, 4.53}	0	28	28	1	1
4	7	{3.42, 3.53, 4.17, 4.53}	0	31	31	1	1
4	8	{1.5, 3.53, 2.25, 4.53}	0	29	29	1	1
4	9	{0., 3.53, 0.75, 4.53}	0	35	35	1	0
4	10	{0.17, 1.76, 0.92, 2.76}	0	27	27	1	1
5	1	{0.2, 0., 0.95, 1.}	0	35	35	1	0
5	2	{6.4, 0., 7.15, 1.}	0	35	35	1	0
5	3	{6.25, 4., 7., 5.}	0	35	35	1	0
5	4	{0., 4., 0.75, 5.}	0	35	35	1	0
6	1	{0.04, 0., 0.79, 1.}	0	35	35	1	0
6	2	{6.05, 0., 6.8, 1.}	0	35	35	1	0
6	3	{5.84, 4., 6.59, 5.}	0	35	35	1	0
6	4	{0., 4., 0.75, 5.}	0	35	35	1	0

Type1exp_4_source						
Room nb	Pen nb	Pen coordinates	Shedding start	Shedding end	Removal time	I count
1	N/a	{1.75, 1.49, 4., 2.49}	0	35	35	5
2	N/a	{1.55, 1.48, 3.8, 2.48}	0	35	35	4
3	N/a	{1.53, 1.78, 3.78, 2.78}	0	35	35	5
4	N/a	{1.75, 1.76, 4., 2.76}	0	35	35	5
5	N/a	{2.41, 2., 4.66, 3.}	0	35	35	5
6	N/a	{2.19, 2., 4.44, 3.}	0	35	35	5

Type2exp_groupsAB&C_rec							
Room nb	Pen nb	Pen coordinates	Exposure start	Exposure end	Removal time	S count	C count
1	1	{0.47, 0., 1.22, 1.}	20	35	35	1	0
1	1	{0.47, 0., 1.22, 1.}	20	35	35	1	0
1	2	{1.97, 0., 2.72, 1.}	20	28	35	1	1
1	2	{1.97, 0., 2.72, 1.}	20	29	35	1	1
1	3	{3.32, 0.04, 4.07, 1.04}	20	28	35	1	1
1	3	{3.32, 0.04, 4.07, 1.04}	20	30	35	1	1
1	4	{4.82, 0.04, 5.57, 1.04}	20	35	35	1	0
1	4	{4.82, 0.04, 5.57, 1.04}	20	35	35	1	0
1	5	{5.35, 1.67, 6.1, 2.67}	20	24	35	1	1
1	5	{5.35, 1.67, 6.1, 2.67}	20	25	35	1	1
1	6	{4.83, 3.42, 5.58, 4.42}	20	35	35	1	0
1	6	{4.83, 3.42, 5.58, 4.42}	20	35	35	1	0
1	7	{3.33, 3.42, 4.08, 4.42}	20	35	35	1	0
1	7	{3.33, 3.42, 4.08, 4.42}	20	35	35	1	0
1	8	{1.85, 3.44, 2.6, 4.44}	20	35	35	1	0
1	8	{1.85, 3.44, 2.6, 4.44}	20	35	35	1	0
1	9	{0.35, 3.44, 1.1, 4.44}	20	35	35	1	0
1	9	{0.35, 3.44, 1.1, 4.44}	20	35	35	1	0
1	10	{0., 1.67, 0.72, 2.67}	20	35	35	1	0
1	10	{0., 1.67, 0.72, 2.67}	20	35	35	1	0
1	11	{1.55, 1.67, 2.17, 2.67}	20	21	35	1	1
1	12	{2.17, 1.67, 2.82, 2.67}	20	21	35	1	1
1	13	{2.82, 1.67, 3.38, 2.67}	20	21	35	1	1
1	14	{3.38, 1.67, 4., 2.67}	20	21	35	1	1
1	15	{4., 1.67, 4.55, 2.67}	20	21	35	1	1
2	1	{0.27, 0., 1.02, 1.}	20	35	35	1	0
2	1	{0.27, 0., 1.02, 1.}	20	35	35	1	0
2	2	{1.77, 0., 2.52, 1.}	20	35	35	1	0
2	2	{1.77, 0., 2.52, 1.}	20	35	35	1	0
2	3	{3.21, 0., 3.96, 1.}	20	35	35	1	0
2	3	{3.21, 0., 3.96, 1.}	20	35	35	1	0
2	4	{4.71, 0., 5.46, 1.}	20	35	35	1	0
2	4	{4.71, 0., 5.46, 1.}	20	35	35	1	0
2	5	{4.85, 1.72, 5.53, 2.72}	20	22	35	1	1
2	5	{4.85, 1.72, 5.53, 2.72}	20	24	35	1	1
2	6	{4.78, 3.44, 5.53, 4.44}	20	35	35	1	0
2	6	{4.78, 3.44, 5.53, 4.44}	20	35	35	1	0
2	7	{3.28, 3.44, 4.03, 4.44}	20	35	35	1	0
2	7	{3.28, 3.44, 4.03, 4.44}	20	35	35	1	0
2	8	{1.79, 3.44, 2.54, 4.44}	20	35	35	1	0
2	8	{1.79, 3.44, 2.54, 4.44}	20	35	35	1	0
2	9	{0.29, 3.44, 1.04, 4.44}	20	35	35	1	0
2	9	{0.29, 3.44, 1.04, 4.44}	20	35	35	1	0
2	10	{0., 1.68, 0.76, 2.68}	20	29	35	1	1
2	_S	{0., 1.68, 0.76, 2.68}	20	29.5	35	1	1
2	11	{1.42, 1.72, 2.06, 2.72}	20	21	35	1	1
2	12	{2.06, 1.72, 2.65, 2.72}	20	21	35	1	1
2	13	{2.65, 1.72, 3.24, 2.72}	20	21	35	1	1
2	14	{3.24, 1.72, 3.83, 2.72}	20	21	35	1	1
2	15	{3.83, 1.72, 4.42, 2.72}	20	21	35	1	1

Type2exp_source						
Room nb	Pen nb	Pen coordinates	Shedding start	Shedding end	Removal time	I count
1	N/a	{1.55, 1.67, 4.55, 2.67}	0	35	35	5
2	N/a	{1.42, 1.72, 4.42, 2.72}	0	35	35	5

3

Chapter 3

Identifiability of environmental transmission parameters: quantifying dispersion and decay of infectious material using spatiotemporal transmission data

Anna M. Gamża[1,2], Mart C.M. de Jong[1], Thomas J. Hagenaars[2]

1) Quantitative Veterinary Epidemiology, Wageningen University & Research, 6708 PB Wageningen, the Netherlands
2) Wageningen Bioveterinary Research, Wageningen University & Research, 8221 RA Lelystad, the Netherlands

Abstract:
Traditional mathematical modelling of infection transmission does not explicitly consider the decay and dispersion of infectious material during its environmental stage. However, to improve understanding of both environmental transmission dynamics and the effect of control measures, the explicit inclusion of these aspects in the modelling is desirable for many host-pathogen-environment systems. As including additional processes comes with additional model parameters, it intensifies the challenge of parameter estimation.

Previously, we demonstrated by example that spatiotemporal data on host infection status collected in transmission experiments is sufficient to estimate transmission rate, decay and dispersion parameters. Here, to inform future studies and sampling protocols, we investigate in detail which spatiotemporal resolution of the observations on host infection status is required for full parameter inference. We perform an identifiability analysis of a parsimonious environmental transmission model. The model provides a simple mechanistic representation of the environmental processes, namely decay and diffusion of infectious material. Additionally, the structurally unidentifiable host dependent processes, namely shedding of infectious material by the infectious host and exposure response of recipient hosts, are estimated as a combination by a single transmission rate parameter.

In a structural identifiability analysis, we show that in the presence of a continuous infectivity source, the resulting three parameter model is structurally non-identifiable if the environmental load is in equilibrium, whilst it is identifiable if transient states are observed with sufficient spatiotemporal resolution. These findings are confirmed by a practical identifiability analysis using both simulated and experimental data of varying spatiotemporal resolution. In particular, we show that an informative experimental design can be obtained by tailoring the temporal and spatial resolution of the experiment to designed values of the model parameters. This work provides a systematic basis for future study of environmental transmission dynamics.

Author summary:

In most cases infectious disease transmission involves a transfer of infectious material through the environment between the hosts. During this transfer two important processes occur: 1) infectious material survives in the environment usually only for a finite time after which it no longer is infectious, and 2) the material may be moved (dispersed) from one place to the other, for example with fomites, air or water flow. These survival and dispersal processes provide a basis for non-pharmacological intervention strategies such as “social distancing” and decontamination of surfaces. To evaluate the efficacy of such measures, a suitable mathematical modelling approach for the environmental transmission process is desired, integrated with the collection of observational data that can reliably reflect environmental levels of infectiousness. Such data can be obtained by placing susceptible hosts at different distances from a source with infected hosts, and taking samples from these hosts to check if, when and where the hosts become infected. Here, we study in detail which design of the host placement and sampling strategy enables the quantification of both environmental survival and dispersion parameters as well as a transmission parameter. This study will help to design future studies and further explore the dynamics of environmental transmission.

Introduction

For infections transmitted through the environment, a susceptible host, in order to become infected, must encounter infectious material deposited in the environment rather than to have ”direct contact” with an infectious host. In other words, in this context an infectious individual can be separated in time and/or space from the susceptible host whilst still being a source of infectious material to which the susceptible host is exposed. Indeed, it has been shown in transmission experiments that transmission via the environment is possible when an infectious host is no longer present in the environment (Colenutt et al., 2020) or when the infectious and recipient host are not occupying the same location (Asadi et al., 2020; Herfst et al., 2012; Holt et al., 1998; van Bunnik et al., 2012; van Bunnik et al., 2014; Zhou et al., 2018).

Traditionally, mathematical modelling of infectious diseases is based on the concept of host-to-host contact, i.e. assumes that the exposure of susceptible hosts is determined by the number of infectious hosts at the present moment in combination with mixing patterns in the host population (Heesterbeek, 2005). In contrast, in models that describe environmental transmission this exposure is usually assumed to be determined by the environmental infectivity load, the dynamics of which is determined by the (cumulative) shedding of infectivity by past and present infected hosts and the subsequent decay of this infectivity in time. In the latter models, the environmental load is most often modelled as a single compartment (Breban, 2013; Lanzas et al., 2020), and as a consequence, dispersion of infectious material from one place to another as a mechanism of infection between spatially separated hosts cannot be described. Recently, it has become increasingly clear that for many host-pathogen-environment systems the explicit inclusion of both decay and spatial dispersion of infectious material in the modelling is desired for an improved understanding of both transmission dynamics (van Bunnik et al., 2014) (Chapter 2) and the effect of interventions. In particular, this is the case for non-pharmaceutical interventions based on spatial separation such as host distancing and quarantining.

As including the decay and dispersion of infectious material in the modelling comes with additional model parameters, it intensifies the challenge of model parameter quantification. One approach to this challenge is to supplement data on host status with samples taken from the environment to estimate environmental exposure levels and/or a survival parameter for the agent in the environment (Brouwer, Weir, et al., 2017; Colenutt et al., 2020; van Bunnik et al., 2014). Another approach is to focus on obtaining maximally informative data on host status.

Animal transmission experiments (Velthuis et al., 2007) (Chapter 2) and rapidly developing methods for gathering individualized data (Barrat et al., 2014; Wirth et al., 2020) open the possibility to study mechanisms of transmission using data on host infection status of tailored temporal and spatial resolution. Such data can be used to make inference on decay and dispersion of infectious material, by estimating the model parameters: transmission rate parameter, decay rate parameter and

diffusion coefficient (adopting diffusion as the simplest mechanistic description of dispersion) (Chapter 2). The use of host status information only, means that one can avoid the use of samples of the environmental contamination level, and thereby avoid the use of additional modelling assumptions that are often difficult to validate, in particular the assumption that the environmental samples taken are representative of the true exposure of susceptible hosts. This assumption does not hold when the environmental load is highly heterogenous (such that practical spatial resolutions of sampling points are too sparse to reflect the true distribution of infectious material), when there are different potential transmission routes (such that e.g. it is unclear whether samples from air and/or from surfaces are representative of exposure) or when microbiological testing is not able to account for (only) the infectious forms of pathogens (e.g. PCR detecting genetic material from both live and dead bacteria).

Recently, our team demonstrated that data on host infection status of suitable temporal resolution can indeed be used to jointly estimate the transmission rate parameter and decay rate parameter in a non-spatial compartmental model provided that information on the past and current infectious sources is used in the estimation (Chang & de Jong, 2023). Moreover, in (Chapter 2) we showed by example that data on host infection status of suitable temporal and spatial resolution from indirect transmission experiments can be sufficient for making inference on transmission rate parameter, decay rate parameter and diffusion coefficient in a parsimonious mechanistic model constructed to describe infections resulting from decay and dispersion of infectious material (van Bunnik et al., 2014)(Chapter 2). Moreover, we have shown that using only the host infection status data resulted in much better model fit than when host status data were supplemented with data from separate survival experiment designed to estimate decay rate parameter (Chapter 2). Here, in order to provide a systematic basis for future study of environmental transmission dynamics, we investigate in detail which spatiotemporal resolution of the observations on host infection status is required for full parameter inference.

Our mechanistic model contains only three separately identifiable parameters: a decay rate parameter α describing decay of infectious material, a diffusion coefficient D describing the dispersion of infectious material and a transmission rate parameter β. The transmission parameter β is a combination of individually unidentifiable parameters: the infectious host shedding rate ρ that determines host infectivity, and the recipient's exposure response parameter ξ that determines recipient's susceptibility and depends on pick up of material by the recipient host and its within-host infection dynamics (the dose response):

$$\beta = \rho\xi. \quad (1)$$

In our spatial model, both source and recipient hosts are confined within areas they occupy, while infectious material is not confined and thus can diffuse freely through the whole environment. In

such a model the probability of infection for an exposed host occupying an area A_{exp} during a time period (t_1,t_2) is described as follows:

$$\mathrm{P}_{\text{inf}}\left(t_1,t_2,A_{\text{exp}}\right) = 1 - \exp\left[-\beta \int_{t_1}^{t_2} \iint_{A_{\text{exp}}} W_r\,(t,x,y|\alpha,D)\,dx\,dy\,dt\,\right], \quad (2)$$

where we define

$$W_r(t,x,y|\alpha,D) \equiv \frac{1}{\rho} W(t,x,y|\rho,\alpha,D) \quad (3)$$

and where the environmental load density function $W(t,x,y|\rho,\alpha,D)$ is a continuous, mean-field description of the spatio-temporal distribution of (infectious) material and accounts for three processes occurring simultaneously: 1) shedding, 2) decay (inactivation) and 3) diffusion (dispersal) of infectious material. The density function $W(t,x,y|\rho,\alpha,D)$ is the solution of the two-dimensional reaction-diffusion equation with decay:

$$\frac{\partial W}{\partial t} = D\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) - \alpha W + \rho s, \quad (4)$$

where $s = s(t,x,y)$ is the spatiotemporal infectivity source pattern determined by the location of each infectious host and their shedding period, during which the infectious material is produced continuously with constant shedding rate ρ. When assuming a single point source, located in $(x,y) = (0,0)$, starting at time $\tau = 0$ and terminating at $\tau = T$, the solution reads as follows:

$$\mathrm{W}(t,x,y|\rho,\alpha,D) = \rho \int_0^{\min(t,T)} \frac{\mathrm{e}^{-\frac{x^2+y^2}{4D(t-\tau)} - \alpha(t-\tau)}}{4\pi D(t-\tau)}\,\mathrm{d}\tau. \quad (5)$$

In Equation 2, to account for structural unidentifiability of shedding rate and exposure response (Equation 1), we use the relative environmental load function $W_r(t,x,y|\alpha,D)$ such that the equation for probability of infection only features the (identifiable) transmission rate parameter.

Assuming that the environment is clean at $t = 0$, and the constant infectivity source, being for example a particular infectious host, is active long enough we can distinguish two periods of environmental load: a transient period, during which the environmental load is increasing in time as long as the amount of material produced is larger than the amount of material that decays per time unit; and an equilibrium state when those two amounts have become equal. In case the duration of infectivity source is not infinite, thus when the infectious period is shorter than observation period ($T < t$), a second transient can be identified, during which the infectious material accumulated in the environment is only decaying (and not being replenished anymore) (Chapter 2).

In Chapter 2, it was shown that all three model parameters (the decay rate parameter, the diffusion coefficient and the transmission rate parameter) are identifiable from a specific set of spatiotemporal data from animal transmission experiments. The question remained unanswered what is the minimal amount of data and its spatiotemporal resolution that is required for successful estimation of all three model parameters. This information is needed for future studies, i.e. for the design of new transmission experiments or of sampling strategies in the field.

The model parameter estimation studied here can be viewed as inferring a continuous spatiotemporal environmental load pattern from binary data on host status (where a host can have one of two statuses- infected or non-infected). When designing animal experiments where such binary data is gathered, several trade-offs have to be considered. First, the number of host must be minimized for ethical, financial and/or practical reasons; yet, it must be ensured that the amount and quality of data collected is sufficient for making the desired model parameter inference. Second, for the same reasons the sampling interval (the temporal resolution of the data) has to be maximized; yet, if a suitable temporal resolution is not obtained (e.g. host are not sampled often enough), this may compromise model parameter inference. Third, for practical reasons the spatial setup (spatial resolution of host placement) has to be optimized as well, e.g. conducting animal experiments with hosts separated by long distances requires correspondingly large experimental rooms in the facility; yet, as we have shown in Chapter 2, in absence of sufficient spatial resolution the decay and diffusion parameters cannot be both inferred.

To address these trade-offs and inform about optimizing study design, we aim to define "suitable spatiotemporal resolution" by analysing the identifiability of the parsimonious mechanistic transmission model described above. To obtain a full picture, we study structural and practical identifiability, both in equilibrium and transient states of the system. For the structural identifiability analysis, we combine analytical and numerical solutions of the model. For the practical identifiability analysis, we use simulated data as well as transmission data from animal experiments.

Results

Our analysis is structured as follows. For both structural and practical identifiability analysis we assume, a single constant infectivity source that starts shedding at time τ=0 and is active beyond the observation period, i.e. $\tau = \infty$, and in most cases assume that the environment is clean at the start of the observation period $t = 0$. For clarity of exposition, we assume that when an exposed host is found to have become infected, it will be removed from the experiment so that its contribution to the environmental load may still be neglected (see Chapter 2). For the structural identifiability analysis, first we consider the one-dimensional version of the environmental load density function $W(t, x|\rho, \alpha, D)$ as it is more accessible to formal analysis. We perform an *a priori* identifiability analysis to explore what parameters or combinations thereof can be identified depending on the spatiotemporal resolution of data. To confirm if the findings are applicable to the

two-dimensional model we conduct an *a posteriori* identifiability analysis using data created for a predefined set of parameters. Finally, we perform a practical identifiability analysis for the two-dimensional model using simulated and experimental data of varying spatial resolution.

The model described in Equation 2 that we examine here has three parameters that are presented in Table 3.1. The model is considered to be identifiable if all of its parameters are jointly identifiable (Wieland et al., 2021).

Table 3.1. Parameters for the spatial model of environmental transmission.

Parameter	Unit	Description
α	[day^{-1}]	Decay rate parameter
β	[day^{-1}]	Transmission rate parameter
D	[m^{2}day^{-1}]	Diffusion coefficient

Structural identifiability

One dimensional model

In the one-dimensional case, the spatial distribution of infectious material, as described by the environmental density function $\mathrm{W}(t, x, y|\rho, \alpha, D)$, reads:

$$\mathrm{W}(t, x|\rho, \alpha, D) = \rho \int_0^t \frac{\mathrm{e}^{-\frac{x^2}{4D(t-\tau)}-\alpha(t-\tau)}}{2\sqrt{\pi}\sqrt{\mathrm{D}(t-\tau)}}\,\mathrm{d}\tau \qquad (6)$$

In equilibrium, i.e. when $t_1 \to \infty$, Equation 6 solves to:

$$\mathrm{W}^{\mathrm{eq}}(x|\rho, \alpha, D) = \rho \frac{\mathrm{e}^{-|x|\sqrt{\frac{\alpha}{D}}}}{2\sqrt{D\alpha}}\ . \qquad (7)$$

To analyse the structural identifiability of our three-parameter model we use the function $\beta\ W_r^{\mathrm{eq}}(t, x|\alpha, D)$ with W_r^{eq} being the equilibrium relative environmental density function:

$$\beta\ W_r^{\mathrm{eq}}(t, x|\alpha, D) = \beta \frac{\mathrm{e}^{-|x|\sqrt{\frac{\alpha}{D}}}}{2\sqrt{D\alpha}}\,. \qquad (8)$$

The function $\beta\ W_r^{\mathrm{eq}}(t, x|\alpha, D)$ is plotted in Fig. 3.1; the shape of the function is determined by both decay rate parameter α and diffusion coefficient D, while the transmission rate parameter β is acting as a scaling factor.

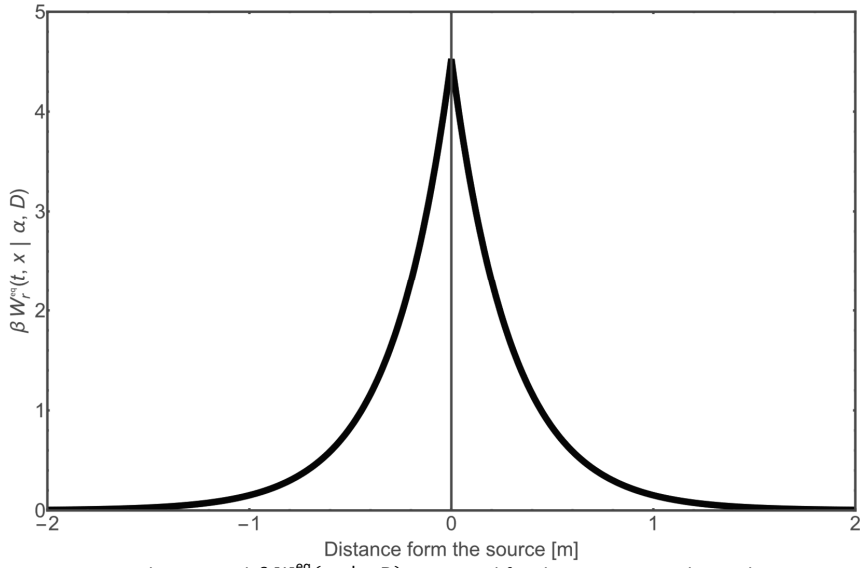


Figure 3.1 One dimensional $\beta\, W_r^{\mathrm{eq}}(t, x|\alpha, D)$ generated for the point source located in point x=0 in function of distance from the source generated for a set of parameters: α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}.

The three-parameter model is structurally non-identifiable in equilibrium if for a particular combination of parameters (α_1, β_1, D_1) we can find at least one other combination of parameters (α_2, β_2, D_2) that gives the same shape of the $\beta\, W_r^{\mathrm{eq}}(t, x|\alpha, D)$ function. To prove that, we just need to find two points, namely x_1 and x_2 for which the function gives the same result for two given sets of parameter values. This requirement yields the following system of equations:

$$\begin{aligned} \beta_1 \frac{e^{-x_1\sqrt{\frac{\alpha_1}{D_1}}}}{2\sqrt{D_1\alpha_1}} &= \beta_2 \frac{e^{-x_1\sqrt{\frac{\alpha_2}{D_2}}}}{2\sqrt{D_2\alpha_2}} \\ \beta_1 \frac{e^{-x_2\sqrt{\frac{\alpha_1}{D_1}}}}{2\sqrt{D_1\alpha_1}} &= \beta_2 \frac{e^{-x_2\sqrt{\frac{\alpha_2}{D_2}}}}{2\sqrt{D_2\alpha_2}} \end{aligned} \tag{9}$$

These conditions are satisfied only in two cases:

1. $x_1=x_2$ and $\beta_2 = \frac{D_2 e^{-x_1\sqrt{\frac{\alpha_1}{D_1}}+x_1\sqrt{\frac{\alpha_2}{D_2}}}\sqrt{D_1\alpha_1}\alpha_2}{D_1\alpha_1\sqrt{D_2\alpha_2}}\ \beta_1$
2. $\frac{D_1}{\alpha_1} = \frac{D_2}{\alpha_2}$ and $\beta_2 = \frac{D_2\alpha_2\sqrt{D_1\alpha_1}}{D_1\alpha_1\sqrt{D_2\alpha_2}}\ \beta_1$

Case 1 is the trivial solution corresponding to the situation when the environmental load is known just for one space point. This solution shows that when data is not provided in any spatial resolution the model is not identifiable as the same probability can be obtained for every α & D combination and accordingly scaled β.

Case 2 is the situation when data is provided with spatial resolution; it shows that the constant source model is still non-identifiable in the equilibrium, but a reduced model with an identifiable combination of parameters can be defined:

$$\beta\, W_r^{\mathrm{eq}}(x|\psi,\omega) = \omega \frac{\mathrm{e}^{-|x|\sqrt{\frac{1}{\psi}}}}{2\sqrt{\psi}}, \tag{10}$$

where $\psi = \frac{D}{\alpha}$ and $\omega = \frac{\beta}{\alpha}$. In contrast to the original three parameters, these two newly defined parameter combinations have a less obvious biological interpretation.

While in equilibrium the amount of infectious material is constant in time, for a transient state, i.e. when the system is not yet in equilibrium (for t<∞), the environmental load density function is both time and space dependent and instead of Equation 8 we have for x>0 (see Methods section for full solution):

$$\beta\, W_r(t,x|\alpha,D) = \frac{\beta}{4\sqrt{D\alpha}}\mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}}\left[1+\mathrm{erf}\left[\frac{-x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]+\mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}}\left(-1+\mathrm{erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]\right)\right]. \tag{11}$$

To analyse the identifiability in this transient state, it is helpful to derive the following condition based on linearization. A model is non-identifiable if we can find a line or surface in parameter space for which the value of $\beta\, W_r(t,x|\alpha,D)$ is the same for (at least) two points in space and (at least) two points in time, so on this line or surface, infinitesimal parameter changes parallel to the line or surface do not change the shape of the $\beta\, W_r(t,x|\alpha,D)$. Therefore, if the model is unidentifiable the following equation is satisfied:

$$\Delta\big(\beta W_r(t,x|\alpha,D)\big) = \sum_{u_i\in(\alpha,\beta,D)} \Delta u_i \frac{\partial(\beta W_r(\vec{u}))}{\partial u_i} = 0. \tag{12}$$

In the Supplementary Note 1 we work out this condition in detail and show that there are no solutions. Thus, we can conclude that the model is structurally identifiable if data on the transient state are available and infectious sources contributing to the environmental load in past and present are known.

Two-dimensional model

To examine the structural identifiability of the two-dimensional model, we performed an *a posteriori* identifiability analysis by estimating parameters with "perfect data". Usually, the likelihood function is based on a finite sample of data, e.g. for one time period of exposure from t_1 to t_2 and two different (types of) exposure areas $A_{\mathrm{exp},1}$, $A_{\mathrm{exp},2}$ the likelihood function is:

$$L = \mathrm{P_{inf}}\big(t_1,t_2,A_{\mathrm{exp1}}\big)^{N_{\mathrm{inf1}}}\Big(1-\mathrm{P_{inf}}\big(t_1,t_2,A_{\mathrm{exp1}}\big)\Big)^{N_{\mathrm{esc1}}}\mathrm{P_{inf}}\big(t_1,t_2,A_{\mathrm{exp2}}\big)^{N_{\mathrm{inf2}}}\Big(1-\mathrm{P_{inf}}\big(t_1,t_2,A_{\mathrm{exp2}}\big)\Big)^{N_{\mathrm{esc2}}}; \tag{13}$$

where $N_{\text{inf},i}$ and $N_{\text{esc},i}$ are the number of new infection cases and the number of escapes respectively that are calculated from data provided from a finite sample. When creating "perfect data" we used the calculated probabilities of infection $\tilde{P}_{\text{inf}}$ calculated from the model for the predefined set of parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}),and a given source location and time, to calculate the true expected proportions of infection days and of escape days:

$$\frac{N_{\text{Inf}}}{N_{\text{Tot}}}=\tilde{P}_{\text{inf}}\left(t_1,t_2,A_{\text{exp}}\right);\ \frac{N_{\text{Esc}}}{N_{\text{Tot}}}=1-\tilde{P}_{\text{inf}}\left(t_1,t_2,A_{\text{exp}}\right). \quad (14)$$

Based on this, a sample of any size N_{Tot} can be created and the corresponding likelihood can be maximized. If the model is identifiable, the predefined set of parameters should be estimated back as for any sample size the maximum likelihood estimator would return true parameter value.

To confirm (for the 2D case) structural unidentifiability when only data for environmental-load equilibrium states is known we maximized the log-likelihood for two space points, each for a time point t=1000 days that was representing the system in equilibrium as confirmed by numerical integration over the model with predefined parameters (see Methods for details). Fig. 3.1 and 3.2 show two-dimensional profile likelihoods obtained for perfect data calculated for the parameter combination α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}.

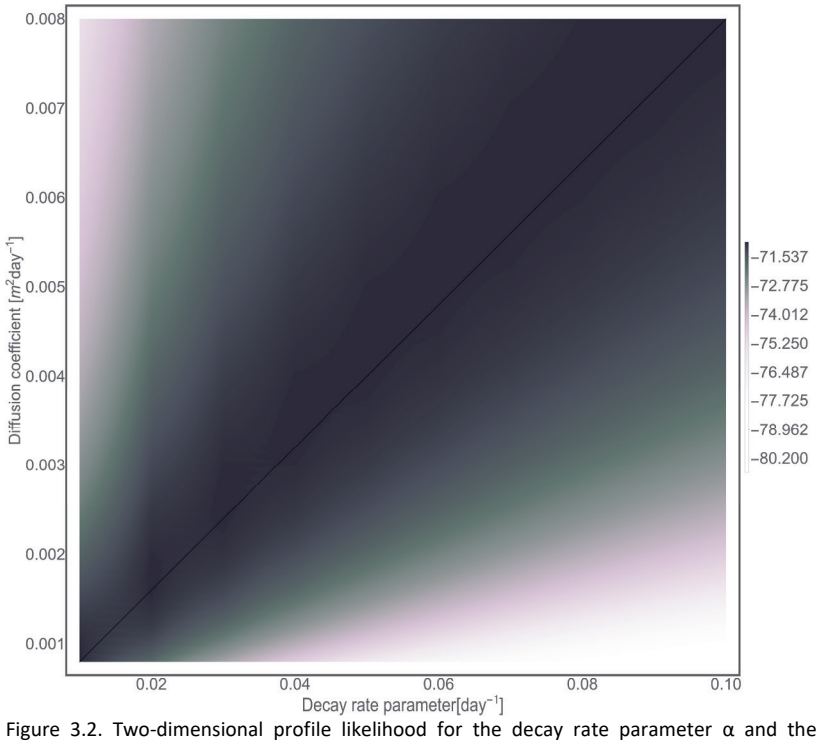


Figure 3.2. Two-dimensional profile likelihood for the decay rate parameter α and the diffusion coefficient D calculated for data from one day in equilibrium created for a predefined set of parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}). The black line marks the maximum value of the log-likelihood.

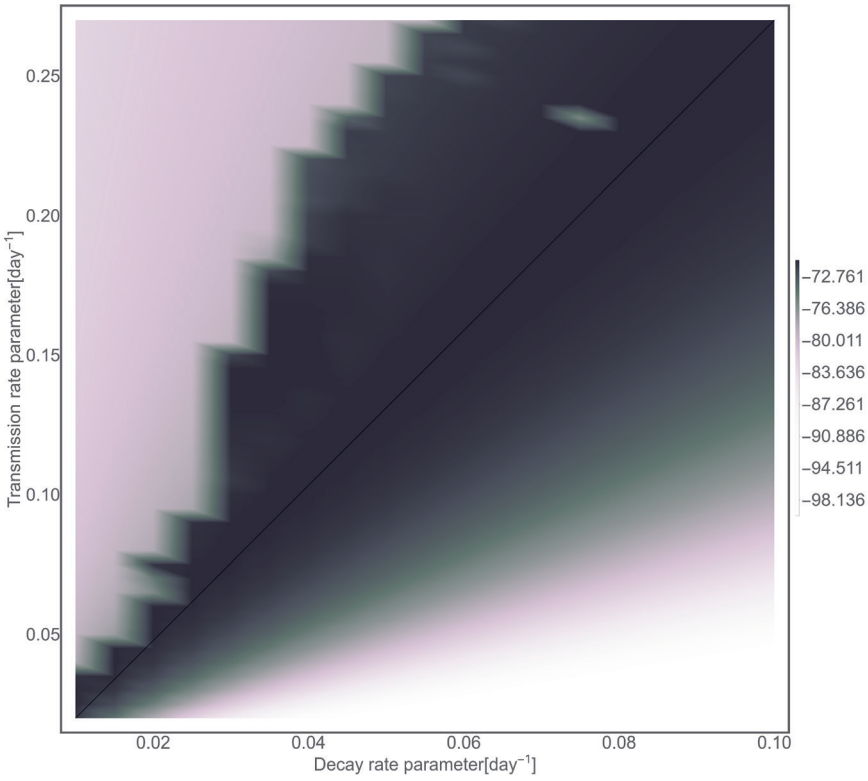


Figure 3.3. Two-dimensional profile likelihood for the decay rate parameter α and the transmission rate parameter β calculated for data from one day in equilibrium created for a predefined set of parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}). The black line marks the maximum value of the log-likelihood, the ragged parts of the profile surface are artifacts from maximization procedure.

The two-dimensional profiles generated for the model in equilibrium illustrate that the likelihood is maximal for any combination of parameters that yields the optimal values for the parameter ratios $\frac{\beta}{\alpha} = 2.667$ and $\frac{D}{\alpha} = 0.087\text{m}^2$. Clearly there is an infinite number of combinations of three parameter values that give the same parameter ratios: $\frac{\beta}{\alpha} = 2.667$ and $\frac{D}{\alpha} = 0.087\text{m}^2$ and thus yield the same maximum log-likelihood value. This shows that the identifiable parameter combinations derived for the one-dimensional system are also valid for the two-dimensional system.

To confirm the structural identifiability for the case that data from transient states are included, we maximized the log-likelihood functions created for the two-dimensional model with perfect data for two scenarios. In both scenarios two space points were sampled. In the first scenario we sampled day 1 and 10 (both representing a transient stage) and in the second scenario we sampled day 1 (representing a transient stage) and day 1001 (representing equilibrium). For both scenarios, the point estimates that maximized the log-likelihood were consistent with the predefined parameter

values: α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}. This confirms the findings from the one-dimensional model analysis that for transient states all three parameters are fully identifiable when data of sufficient spatiotemporal resolution is provided.

Practical identifiability

A structural identifiability analysis yields information on whether parameters are identifiable in principle. In practice, the finite sample size and finite resolution of sampling may still prevent the identification of structurally identifiable parameters, e.g. when all samples are tested negative or when the resolution is insufficient (e.g. all samples test positive at the first sampling moment). We therefore also conducted a practical identifiability analysis, considering various spatiotemporal sampling strategies with finite sample size.

To study practical identifiability, we used a two-dimensional model with one constantly shedding area source and multiple recipient areas as presented in Chapter 2. We considered parameter identifiability for two types of data. First, we used data simulated for predefined parameter values. Second, we used data from *Campylobacter jejuni* (abbreviated further as *C. jejuni*) transmission experiments in broilers previously published in Chapter 2. For both data sets the 'spatial resolution' as well as the 'distance from the source' of recipient areas was varied.

Simulated data

We used the predefined parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}) as a reference and compared results for 12 scenarios, which included 4 scenarios covering one distance band, 4 scenarios covering two distance bands and 1 scenario covering four distance bands, as illustrated in Fig. 3.4a, and additionally 1 scenario was included in which all the grid cells were sampled. In other words, the scenarios differed in the number of distance bands included for which cells were occupied by recipient hosts and/or the number of cells included per distance band ("spatial resolution") and the distance(s) from the source at which the distance band(s) with recipients were placed ("distance to the source"). By including a suitable number of experimental repetitions per scenario we ensured that the scenarios all had the same sample size (a total number of areas n=128 sampled daily) and the same spatiotemporal infectivity source pattern (infectious hosts occupying a single central area source consisting of 4 unit cells and shedding continuously throughout the experimental period). The experimental period was 30 days, with the source shedding continuously all the time (from t=0 to t=30) with source strength 1.0 (1 host/m2) and with recipients exposed, and sampled daily, from day 0 to day 30.

The estimates and confidence bounds obtained for all scenarios are depicted in Fig. 3.4b-d.

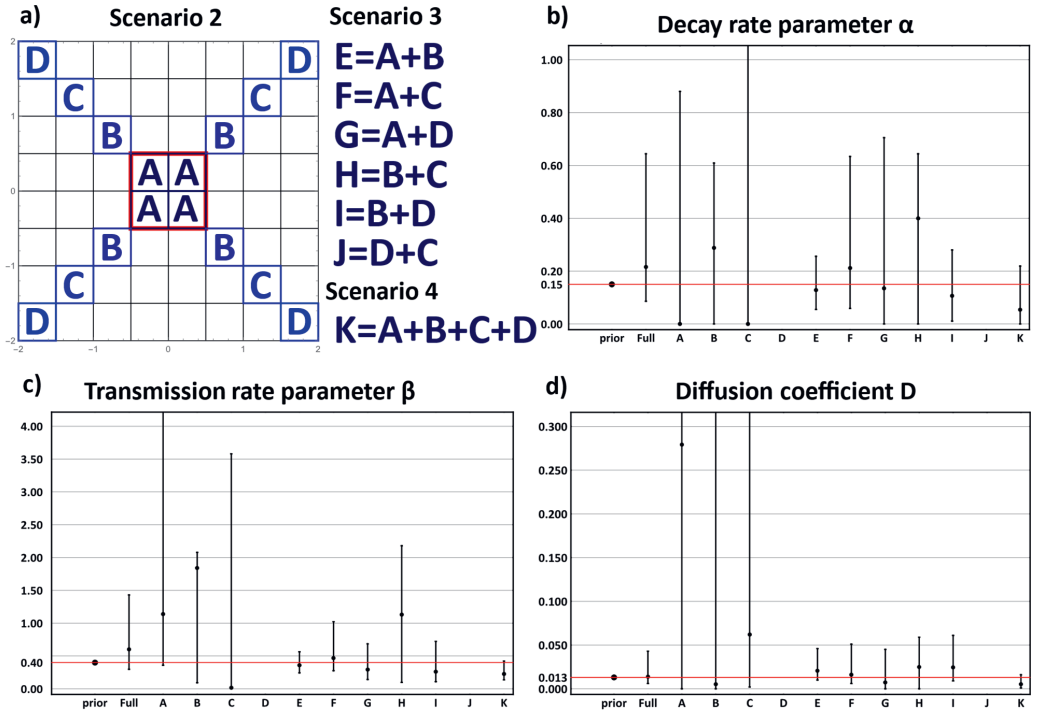


Figure 3.4. Parameter estimates (panels b-d) for a set of 11 experimental design scenarios (panel a), obtained from simulated data using given parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}); a) design of the scenarios indicating the areas sampled around a single area source located in the four central squares (marked in red) and shedding continuously from t=0 until the end of the experiment set at t=30 days. In each labelled grid cell of size 0.5 m x 0.5 m one recipient host was housed, and for each scenario the experiment was repeated until the n=128 recipient areas were sampled, each was sampled daily. In scenario 'Full' (not depicted) the full grid of 64 cells was sampled daily (and there were two experimental repetitions); b) diffusion coefficient estimates for all scenarios; c) decay rate parameter estimates for all scenarios; c) transmission rate parameter estimates for all scenarios.

In scenario 'Full' in which all 64 grid cells were sampled daily in 2 repetitions, such that full spatial resolution was included, the estimates obtained for all three parameters were not significantly different from the a priori parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}), with confidence bounds for the decay rate spanning a wide uncertainty range.

In scenarios A-D, only one distance band with recipient data was included, and the scenarios differed in its distance to the source. The point estimates for all three parameters were different from the predefined values for three scenarios (A, B, C). In all scenarios A–C the range spanned by the confidence bounds was very broad with the bounds often reaching extreme values. For scenario D, (longest distance from the source considered), no estimates were obtained as there were no positive cases in the simulated sample. These results indicate that parameters are not practically identifiable when data on only one distance band data is included and confirm the findings from the structural identifiability analysis that when the observations are lacking spatial resolution the model is not identifiable.

The scenarios E-J were defined as different combinations of two distance bands. For three of these scenarios (E, H, I) all three parameters were estimated satisfactorily: the point estimates were not significantly different from the a priori values and the range spanned by the confidence bounds was narrow. For two scenarios (F, G) satisfactory estimates were obtained for the transmission rate parameter β and the diffusion coefficient D; although the decay rate parameter α estimates are not significantly different from the predefined value, the confidence bounds span a broader range than in the scenarios E, H and I. For one scenario (J), the maximum likelihood cannot be obtained, as the optimization algorithm does not converge, most likely because the simulated number of positive cases (2/128) was too small.

In scenario K, where four distance bands were included, the three parameters were estimated satisfactorily, as the point estimates were not significantly different from the predefined value and the range spanned by the confidence bounds was narrow.

Overall, the best fit was obtained for the scenario E where the two distance bands closest to the source were included. This fit was even better than the fit for scenarios that have a better spatial resolution - scenario 'Full' of samples from the whole grid and scenario K with all four distance bands. Our interpretation of this result is as follows; in the scenarios 'Full' and K, areas far from the source were sampled, and the high probability of escaping from infection across the experimental timespan for these areas resulted in much more escapes than infection cases. As the sample size was the same for all scenarios, the scenario E profited from having a larger number of samples obtained from areas close to the source where probability of being infected anytime during the experiment is similar to the probability of escaping, which made the data from these samples more informative. Overall, these results show that while spatial resolution is crucial to obtain satisfactory estimates, the most informative data is obtained from sampling areas where the probability of being infected during the experiment is similar to the probability of escaping from infection.

Experimental data

To examine practical identifiability using real world data, we used previously published data from *C. jejuni* transmission experiments between spatially separated broilers (van Bunnik, 2014; van Bunnik et al., 2014)(Chapter 2). As a reference we used the estimates obtained when all experimental data was included (Chapter 2). Considering variability in spatial organization (one distance band vs multiple distance bands) and temporal design (starting the exposure of recipients in a clean environment vs starting the exposure of recipients in an already contaminated environment) we divided the experimental data into 5 parts that we will refer to as "experimental data parts". Table 3.2 shows the most important differences in spatio-temporal experimental design between these experimental data parts.

Table 3.2. Spatio-temporal differences in the quality of previously published data from *C. jejuni* transmission experiments between spatially separated broilers

Experimental data part		Distances covered [m][(1)]	Start in clean environment	Proportion of infected hosts	Detailed description
Intermediate-Clean	**(In-Cl)**	0.75 and 1.06	YES	20/80	(van Bunnik, 2014; van Bunnik et al., 2014)
Range-Clean	**(Ra-Cl)**	0.35–2.00	YES	16/48	Chapter 2
Intermediate-Contaminated	**(In-Co)**	0.70–0.90[(2)]	NO	5/40	Chapter 2
Zero-Contaminated	**(Ze-Co)**	0.00[(3)]	NO	10/10	Chapter 2
Zero-Clean	**(Ze-Cl)**	0.00[(3)]	YES[(4)]	5/5	Chapter 2

1) Approximate border to border distance;
2) Except for one recipient area that was separated by 0.43 m;
3) Source and recipient hosts housed together;
4) Start before main source started shedding, possible small contamination beforehand from distance sources.

We created alternative datasets though making different combinations of the data parts, each dataset containing a subset from the five experimental data parts listed in Table 3.2. Parameter estimates for all combinations are depicted in Fig. 3.5-3.7.

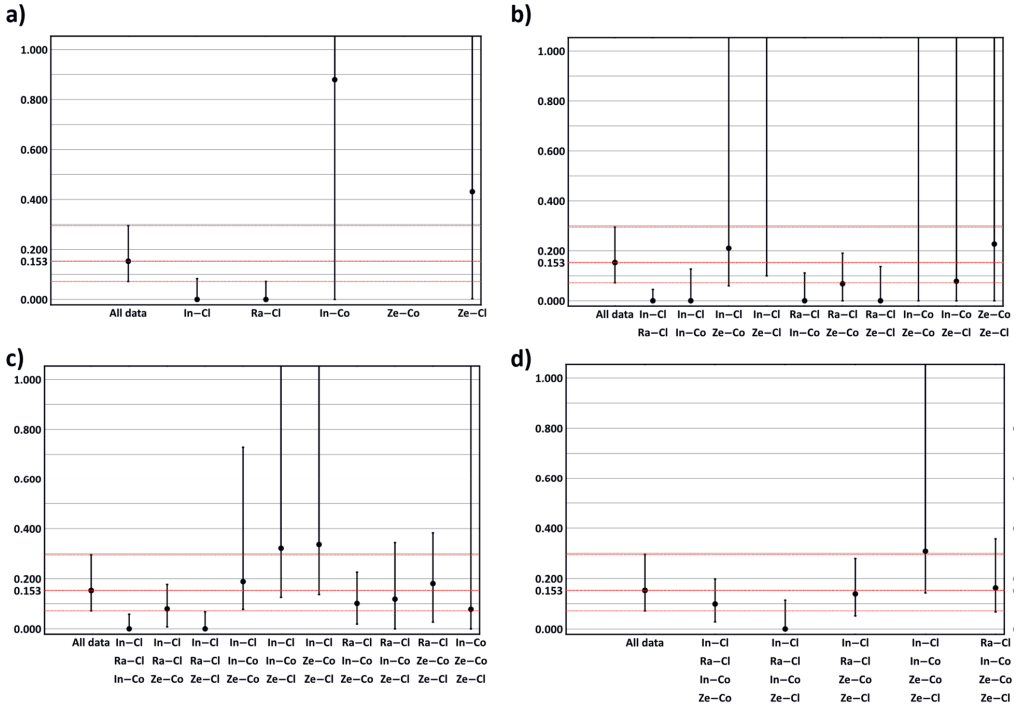


Figure 3.5. Parameter estimates for the decay rate parameter α obtained for various subsets of experimental data from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design; "All exp." represents the estimate obtained using all available experimental data, which was 0.153 day^{-1} (CI: 0.072- 0.295) and is used as reference value (red lines); a) all sub-sets containing a single experimental data part each; b) all sub-sets containing 2 experimental data parts each; c) all sub-sets containing 3 experimental data parts each; d) all sub-sets containing 4 experimental data parts each.

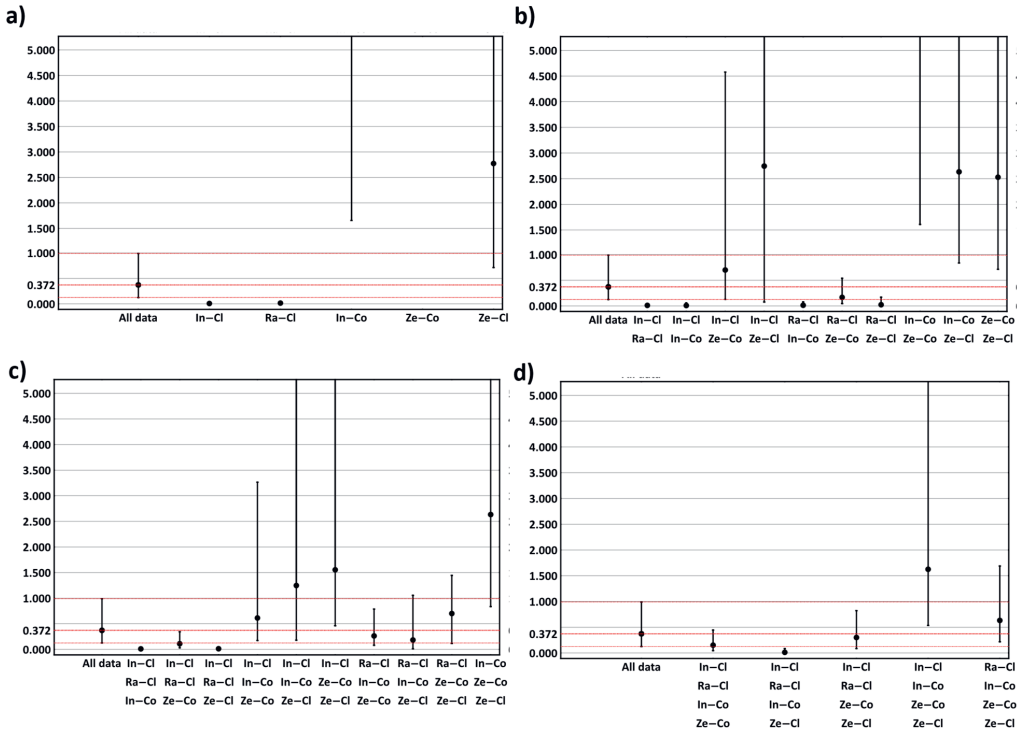


Figure 3.6. Parameter estimates for the transmission rate parameter β obtained for various subsets of experimental data from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design; "All exp." represents the estimate obtained using all available experimental data 0.372 day^{-1} (CI: 0.125- 0.989) and is used as reference value (red lines); a) all sub-sets containing a single experimental data part each; b) all sub-sets containing 2 experimental data parts each; c) all sub-sets containing 3 experimental data parts each; d) all sub-sets containing 4 experimental data parts each.

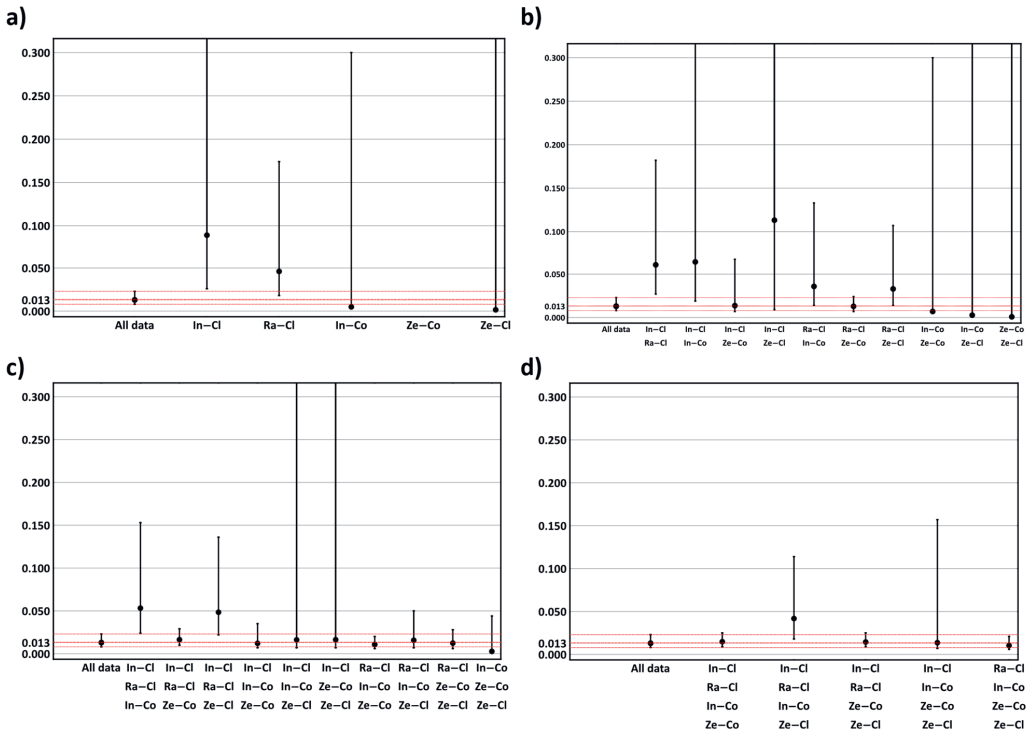


Figure 3.7. Parameter estimates for the diffusion coefficient obtained for various combinations of experimental data parts from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design; "All exp." represents the estimate obtained using all available experimental data 0.013 m2day-1 (CI: 0.008-0.023) and is used as reference value (red lines); a) all sub-sets containing a single experimental data part each; b) all sub-sets containing 2 experimental data parts each; c) all sub-sets containing 3 experimental data parts each; d) all sub-sets containing 4 experimental data parts each.

Parameter estimates for all combinations of experimental data parts were compared to estimates obtained for the complete set of all available experimental data. When only data from one experimental data part was studied the model is clearly not identifiable; in most cases a value of 0 is obtained for the point estimates for all three parameters and/or confidence bounds are reaching extreme values; for one of the experimental data parts (Ze-Co) the maximization does not converge (and this is explained by the observation that all host were infected within one day after exposed to the contaminated environment (Chapter 2). For sub-sets with combinations of two experimental data parts most parameter estimates are still unsatisfactory, with point estimates not being close to the reference value and/or the confidence bounds being wide apart. An exception that provides fairly good estimates for all three model parameters is the combination of data from the experimental data parts Ra-Cl and Ze-Co. For sub-sets with combinations of three experimental data parts, all three parameters were estimated satisfactorily (point estimate not significantly different, not reaching extreme value, confident bounds narrow) for all combinations that contained both data parts Ra-Cl and Ze-Co, and for the combination of Ra-Cl, In-Co and Ze-Cl. For sub-sets with

combinations of four experimental data parts all three parameters were estimated satisfactorily for all combinations that contained both data parts Ra-Cl and Ze-Co.

Overall, we can conclude that including data from experimental data part Ra-Cl is a necessary but not sufficient condition to ensure parameter identifiability. This part contains data from an experiment where the recipient areas were separated from the source area by various distances ranging from 0.35 to 2.00 m (border to border distance). When experimental data part Ze-Cl is also included, the data are sufficient to obtain satisfactory estimates. In the data part Ze-Cl proportion of positive cases was the largest amongst all data parts (10/10 hosts infected).

Discussion

In Chapter 2, we have shown, for a particular example, that with our parsimonious spatial environmental transmission model it is possible to gain insight into transmission resulting from decay and diffusion of infectious material by estimating the three model parameters using transmission data only, provided data with suitable spatiotemporal resolution is available. Here, to inform about the design of future experiments and sampling protocols, we aimed to define the "suitable spatiotemporal resolution" needed to estimate model parameters by performing both structural and practical identifiability analyses for equilibrium and transient states of the environmental load. We showed that in exposure equilibrium the one-dimensional model is structurally unidentifiable; identifiable combinations of parameters can be defined, but those two new parameters, contrary to original three-parameter model (Chapter 2), have a less obvious biological interpretation. For transient states, we showed that the one-dimensional model is structurally identifiable even with minimal spatiotemporal resolution (two points in space each and two points in time).

To study real life scenarios, where sample size is always finite, and address the trade-offs with identifiability for experimental design we conducted a practical identifiability analysis of the two-dimensional model using both simulated and real-life experimental data. Simulated data was analysed to check if a predefined parameter value set can be estimated back from simulated data. Because this part of the study was done *in silico*, we were able to analyse various scenarios and fully control the spatiotemporal resolution and sample size included in the analysis. Analysis of this simulated data confirms the findings from structural identifiability analysis. When only one distance band is studied the model is not identifiable, while if data for a minimum of two distance bands are included the model is identifiable (providing a sufficient number of positive cases). By comparing estimates based on the same sample size but on data varying by distance from the source we showed that the best estimates are obtained when areas are sampled where the probability of being infected anytime during experimental period is comparable to the probability of escaping from infection in fact the optimum is expected when both probabilities are equal to 0.5.

To further explore practical identifiability we have also analysed data from *C. jejuni* transmission experiments between spatially separated broilers (Chapter 2) (van Bunnik, 2014; van Bunnik et al., 2014). The structural identifiability analysis, and the practical identifiability analysis using simulated data, showed that including just two suitable distance bands should be sufficient in principle for ensuring identifiability. However, the practical identifiability analysis using experimental data shows it is not necessarily easy to determine suitable distance bands: although the most valuable data come from an experiment that that covered a wide distance range (Ra-Cl), the parameters became identifiable only after also including the experimental data part Ze-Co for "zero distance" between source and recipients.

While in this study only one set of parameter values was explored, the methodological framework we developed can easily be used for identifiability and power analysis for systems with various parameter sets. Moreover, using Equation 5 we can define a characteristic time 1/α that governs the temporal dependence of the environmental load function (Chang & de Jong, 2023) and a characteristic distance $\sqrt{(D/\alpha)}$ governing the distance dependence. Consequently, for design values of D and α different from the values studied/estimated above, appropriate spatial and temporal resolutions in the experimental design can be readily obtained by correspondingly rescaling the resolutions used in the simulations. In more detail, for the values α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1} we obtain $1/\alpha$= 6.67 days and $\sqrt{(D/\alpha)}$ =0.29 m. Therefore, as we found that a temporal resolution of 1 day = $0.15*1/\alpha$ and a spatial resolution of 0.35 m = $1.2\sqrt{(D/\alpha)}$ were sufficient in this case, for general design values $\alpha = \alpha_{\mathrm{d}}$ and $D = D_{\mathrm{d}}$ a rule thumb is that sufficient temporal and spatial resolutions can be calculated as $0.15*1/\alpha_{\mathrm{d}}$ and $1.2*\sqrt{(D_{\mathrm{d}}/\alpha_{\mathrm{d}})}$, respectively.

We analysed identifiability in the context of spatial resolution of data, assuming that the status of hostS was assessed daily. Such a time series of infections is usually recorded in transmission studies. Our structural identifiability analysis shows that it is crucial to obtain a temporal resolution that is able to capture the transient states of the system. It is possible that for extremely fast system dynamics (fast/high shedding, fast decay and dispersion), the sampling would be needed so frequently that it would not be practically feasible.

Here, we mostly studied the situation where one source area housing infectious hosts was shedding continuously during the whole observation period. Thus, we defined the transient as a state before equilibrium is reached. In (Chapter 2) we showed that in this case the environmental load density function has three modes: 1st mode- before the source starts shedding (no infectious material present), 2nd mode- when source is producing infectious material continuously, 3rd mode- after the source stopped producing infectious material. The second mode is the equivalent of the model we used in this chapter. For the 3rd mode, the source has ceased but infectious material is still present in the environment, where it decays and diffuses further, i.e. it also is a transient state and can be

used for parameter estimation with suitable data. Although this second transient state was beyond the scope of our analysis, our work indicates that the model would be structurally identifiable with data from this state.

We note that in real life situations the environmental load is rarely in equilibrium. Newly infected hosts become new sources and these sources are not removed from the system as it was the case for the *in-silico* data used here and for most of the experiments used here. Nevertheless, analyzing the environmental-load equilibrium is still relevant as it may apply to a very good approximation in certain situations, for example when the incubation period is longer than the remaining observation period such that observed transmission events can be approximated as independent form each other. Our model can be used in situations when sources are changing as we showed when constructing the likelihood based on data from one of our experiments where newly infected recipients were not removed (here experimental data part **In-Co**).

Here, we analysed model identifiability based on host infection status data only. In principle, any binary data that would allow us to estimate the probability of infection in a spatiotemporally resolved manner can be used in the framework we developed. In practice, the detection of colonised hosts remains the best observation strategy available. As an alternative or supplementary strategy one can use environmental monitoring to gather information about the exposure to infectious material deposited in the environment (Brouwer, Weir, et al., 2017; Eisenberg et al., 2013). For infectious disease, environmental monitoring can be defined as detecting pathogens in the environment using direct or indirect indicators. Among direct methods are microbiological sampling and detection, often used to determine contamination of farms, hospital wards or food establishments. Clearly, the detection of pathogens and/or the failure to do so does not necessarily or fully inform about the probability of infection. Perhaps, the most important reason is that for many situations, the relevant exposure mechanisms are not sufficiently understood: e.g. there are multiple transmission routes and their relative importance is poorly known. Also, even if all samples are found negative, infectious material can still be present in the environment, but just escaping from detection. Failure of detection can occur due to detection limits and/or because of inadequate sampling. In particular, if the environmental load is highly heterogenous, it may be very difficult to design an adequate sampling strategy, i.e. one obtaining a set of environmental samples that is representative for exposure. Representativity of environmental sampling results for the actual exposure is also becoming problematic when pathogens detected by environmental monitoring can be of a form that is not infectious to susceptible hosts. For example, PCR detected genetic material of pathogens may not come from live/infectious forms or may be detected in material that is not infectious for the hosts (e.g. pathogens transmitted aerobically detected in water) (Yang & Rothman, 2004).

Indirect indicators for pathogen presence in the environment may also be of interest and can be of two kinds. For specific systems, it is possible to measure some characteristics of environment, for example water turbidity to assess its contamination which is correlated with presence of *Cryptosporidium* (Brouwer, Weir, et al., 2017). Yet, for most systems finding such environmental indicators is not possible. The most generally applicable indirect detection method for infectious material remains the monitoring the infection of sentinel hosts placed in the environment. The observation of a (newly) infected host shows that the infectious material was present in the environment when transmission happened. Moreover, this type of indirect pathogen "detection" by definition, concerns only infectious forms of pathogens and pathogens residing in material that causes the actual (natural) exposure of the hosts.

Here, we showed how the decay and diffusion of infectious material can be studied using transmission data provided in spatialotemporal resolution, i.e. combining host infection status observations with observations on host locations. Such data can be collected in animal transmission experiments or from outbreaks in the field, such as for example cases from seated events, hospital wards and animals housed in separate buildings, pens or fenced pastures.

As our practical identifiability analysis shows that while the spatial resolution of data is crucial for the quality of parameter estimates, increasing the spatial resolution while keeping the same sample size is not beneficial when newly added areas are located in a region where the total probability of being infected during experimental period is much lower than the probability of escaping from infection. Combining data from at least two areas where the probability of being infected anytime during the experiment is not much different from the probability of escape seems to be the best sampling strategy to ensure practical identifiability when sample size is limited. This can be done by calibrating the experimental timespan and the spatial organisation of recipient areas based on the characteristic time and distance of the system for a set of design values for the parameters. Clearly, in many situations even the approximate parameter values may not be known. If suitable estimates of parameters are not available, literature data on time and distance dependence maybe be used to predict areas with suitable probability of infection. If such data is not available for the system at hand, a good option would be to inform the design by a pilot study where a wide distance range is covered, and the experimental timespan is flexible (e.g. observation until 1/3 of recipient hosts is infected or maximum period is reached).

In summary, our analyses show that spatiotemporal sampling protocols based on host infection status and designed to quantify environmental transmission parameters, should take into account three requirements: 1) to sample the transient state(s) of the system; 2) to obtain a specified minimum spatial resolution of data; 3) to sample areas where the design value of the total probability of infected during experimental period is not much different from the probability of escape. If these requirements are fulfilled, inference can be made on the microscopic processes of

decay and dispersion of infectious material, together with estimate describing host dependent processes, using only host infection status data.

Methods

System

The model we analysed accounts for decay and diffusion of infectious material and for the relevant host dependent processes. It uses areas occupied by source and recipient hosts as configuration parameters. The mathematical details were previously described in detail in Chapter 2 and derived in (van Bunnik et al., 2014). The environmental load density function published in Chapter 2 has three modes. Here, we analysed the second mode of this function (Equation 5) therefore assuming that in our system, the source is producing infectiously material continuously throughout the experimental period.

To examine the structural identifiability based on observations for the equilibrium and transient states, we first analytically explored the one-dimensional version of the model. Next, we confirmed the findings for the two-dimensional version by performing numerical calculations. To examine the practical identifiability, we estimated parameters for various combination of simulated and experimental data using a constraint maximization algorithm we developed in Mathematica 12 (Wolfram Research, 2019).

Structural identifiability

One dimensional model

We calculated the time integrals in $\mathrm{W}(t,x|\rho,\alpha,D)$ (Equation 6) in 2 steps. First, the integral from $t-\Delta t$ was solved:

$$\mathrm{W}(t-\Delta t,x|\alpha,D,\rho)=\lambda\int_0^{t-\Delta t}\frac{\mathrm{e}^{-\frac{x^2}{4D(t-\tau)}-\alpha(t-\tau)}}{2\sqrt{\pi}\sqrt{\mathrm{D}(t-\tau)}}\,\mathrm{d}\tau=$$

$$\lambda\frac{e^{-\sqrt{\frac{\alpha}{D}}|x|}\left(\mathrm{Erf}\left[\frac{2t\sqrt{D\alpha}-|x|}{2\sqrt{dt}}\right]-\mathrm{Erf}\left[\frac{2\sqrt{D\alpha}\Delta\mathrm{t}-|x|}{2\sqrt{D\Delta\mathrm{t}}}\right]+e^{2\sqrt{\frac{\alpha}{D}}|x|}\left(\mathrm{Erf}\left[\frac{2t\sqrt{D\alpha}+|x|}{2\sqrt{Dt}}\right]-\mathrm{Erf}\left[\frac{2\sqrt{D\alpha}\Delta\mathrm{t}+|x|}{2\sqrt{D\Delta\mathrm{t}}}\right]\right)\right)}{4\sqrt{D}\sqrt{\alpha}};\qquad(15)$$

Next, the limit $\Delta t\to 0$ was taken:

$$\mathrm{W}(t,x|\alpha,D,\rho)=\begin{cases}\lambda\dfrac{\mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}}\left(1+\mathrm{Erf}\left[\frac{-x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]+\mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}}\left(\mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]-1\right)\right)}{4\sqrt{D\alpha}} & x>0\\ \lambda\dfrac{\mathrm{e}^{x\sqrt{\frac{\alpha}{D}}}\left(1+\mathrm{e}^{-2x\sqrt{\frac{\alpha}{D}}}\left(\mathrm{Erf}\left[\frac{-x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]-1\right)+\mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]\right)}{4\sqrt{D\alpha}} & x<0\end{cases}\qquad(16)$$

where the first term (for x>0) is used to derive Equation 6 presented in the Results section.

To solve the model for equilibrium, we took the limit $t \to \infty$, as follows:

$$\mathrm{W}^{\mathrm{eq}}(x|\alpha, D, \rho) = \lim_{t\to\infty} \mathrm{W}(t, x|\alpha, D, \rho) = \begin{cases} \lambda \frac{\mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}}}{2\sqrt{D\alpha}} & x > 0 \\ \lambda \frac{\mathrm{e}^{x\sqrt{\frac{\alpha}{D}}}}{2\sqrt{D\alpha}} & x < 0 \end{cases} \quad (17)$$

Both, for obtaining these solutions and for developing the numerical calculations we used Mathematica 12 (Wolfram Research, 2019).

Two-dimensional model

For the two-dimensional model we performed a numerical analysis of the environmental load function for a predefined parameter set (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}) and for a point source located in the point (0, 0). We used our maximization algorithm (see below) to estimate back the parameters using a likelihood function based on perfect data points for two points in space: (0.5, 0.5) and (1, 1). To study the system in equilibrium we mimicked the equilibrium state by calculating the probability of infection and generating log-likelihood function for day 1001 of observation for both space points. Based on the time profile of the environmental load density function we confirmed numerically that the system is in (close to) equilibrium state at this day for those two points, see Fig. 3.8.

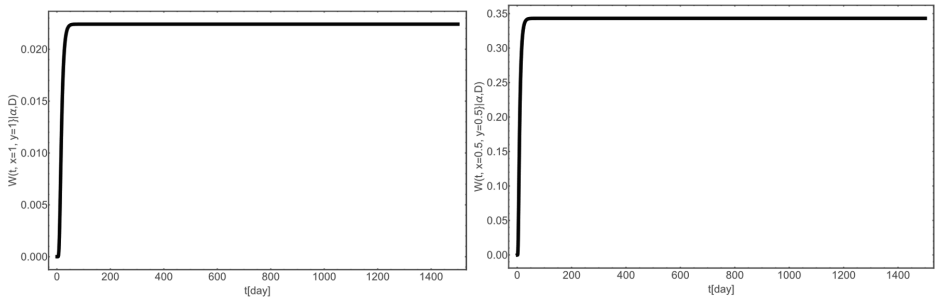


Figure 3.8. Environmental load density function $\beta\, W_r(t, x, y|\alpha, D)$ value for two points in space (0.5,0.5) and (1,1) as a function of time since the start of shedding, calculated for parameters values: α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1} and a single point source located in point (0,0).

For transient states we calculated the probabilities and generated the log-likelihood function for two scenarios. In scenario 1, we used two time points (both transient): day 1 and day 11. In scenario 2, we used different time points: day 1 (transient) and day 1001 (representing the equilibrium).

All log-likelihood functions are provided with a code on Zenodo with the identifier(s) [10.5281/zenodo.7428123](Gamża, 2023a).

Maximization algorithm

We conducted all modelling and analysis in Mathematica 12 (Wolfram Research, 2019).

For numerical analysis we developed an automatic constraint maximization algorithm to obtain point estimates for all three parameters, with respective confidence bounds and univariate profile likelihoods. The code for the algorithm is provided in Zenodo (Gamża, 2023a).

As the log-likelihood functions we maximize are prone to return only local maxima when numerically maximized by the "NMaximize" function without constraints and as we, in addition to the maximum-likelihood estimates, need to obtain confidence bounds based on profile likelihoods, we developed a four-step constraint maximization algorithm. Each step is coded using one fully automatic function, and the schematic description of the algorithm is provided in Fig. 3.9.

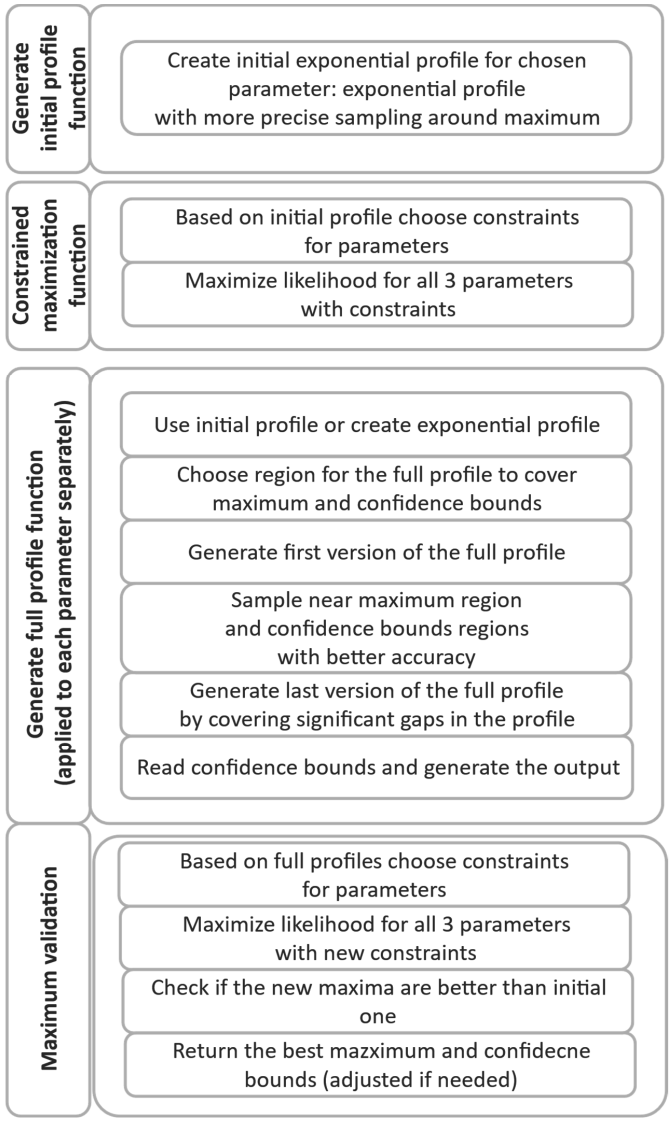


Figure 3.9 Scheme of the constrained maximization algorithm. The detailed purpose-written code in Mathematica can be found in Zenodo [10.5281/zenodo.7428123] (Gamża, 2023a)

In the first step, an initial profile generation was implemented by searching broadly using exponential bins within parameter space of one chosen parameter to find in which region the global maximum is most likely to be found and indicate constraints for the next step - constraint maximization. In the second step, constraint maximization is performed to find constraints for parameters and maximize the likelihood for all three parameters simultaneously. In the third step, the profile likelihood and confidence bounds are produced for each parameter. The profile

likelihood is calculated across a range designed to cover the region spanned by the 95% confidence bounds (based on the likelihood ratio test); to produce smooth profiles, any running parameter point profile values that most likely corresponded to a local maximum (log-likelihood value lower than both neighboring running parameter points) are detected and re-estimated using neighboring estimates as bounding box constraint on the search range; running parameter regions around the maximum-likelihood value and around the confidence bounds are sampled with a preset desired accuracy. The confidence bound values are determined from the profile likelihood. The last step is designed to check if the initially obtained maximum was indeed the global maximum: information from full profiles for all parameters is used to create new constraints to maximize the log-likelihood and compare to the previous maximum. The parameter estimates for the overall maximum are returned with confidence bounds as final output.

Algorithm implementation for transmission data

Bins used to estimate parameter values for both simulated and experimental data are presented in Table 3.3. The exponential profile was generated using an exponential profile bin list while the initial profile was generated by choosing from the initial profile bin list the parameter bins that covered the maximum-likelihood value and the confidence bounds.

Table 3.3. Profile bins provided for the maximization algorithm to estimate parameters from transmission data.

Exponential profile bins lists	{d, {0.0005, 0.005, 0.05, 0.5, 5}}; {α, {0.0005, 0.005, 0.05, 0.5, 5., 50.}}; {β, {0.0005, 0.005, 0.05, 0.5, 5., 50., 500., 5000.}};
Initial profile bins lists[1]	{{d, 0.0001, 0.0009, 0.0001}, {d, 0.001, 0.009, 0.001}, {d, 0.01, 0.09, 0.01}, {d, 0.1, 0.9, 0.1}, {d, 1., 9., 1.}}; {{α, 0.0001, 0.0009, 0.0001}, {α, 0.001, 0.009, 0.001}, {α, 0.01, 0.09, 0.01}, {α, 0.1, 0.9, 0.1}, { α, 1., 9., 1.}, { α, 10., 90., 10.}}; {{β, 0.0001, 0.0009, 0.0001}, {β, 0.001, 0.009, 0.001}, {β, 0.01, 0.09, 0.01}, {β, 0.1, 0.9, 0.1}, {β, 1., 9., 1.}, {β, 10., 90., 10.}, {β, 100., 900., 100.}, {β, 1000., 9000., 1000.}};

1) each bin is in a form {parameter, {min value, max value, step size}}

If confidence bounds were not found within the initial profile bins of the search region (given in Table 3.3), these would still be calculated using the algorithm when located between the following minimal and maximal values of parameters were used: for D min 10^{-5} and max 10 m^2day^{-1}, for α min 0 and max 10 day^{-1}, for β min 10^{-5} and max 1000 day^{-1}. Otherwise, the confidence bound was listed as infinite. All finite confidence bounds values were calculated with a numerical accuracy of 0.001.

Simulated data

We simulated data on a 4x4 m grid with 64 cells each of size 0.5x0.5 m. The (area) source of infection was located in the 4 central cells and was assumed to continuously shed infectious material during the whole experimental period. It was assumed that each cell can become infected only once

(equivalent of having one recipient host in each cell). We created 4 scenarios: 1) all grid cells were sampled; 2) only one distance band with cells on the diagonal of the grid was sampled, repeated for 4 different distance bands; 3) 2 out 4 distance bands were sampled, repeated for all 6 possible combinations of distance bands; 4) all 4 distance bands were sampled. The scheme of areas included in scenarios 2-4 is shown in Fig. 3.4. The cells included in each scenario are presented in Table 3.4. For each repetition in each scenario the same total amount of data points was used: 128 cells sampled daily for 30 days. The function used to generate the simulated data is provided in Zenodo (Gamża, 2023a).

Table 3.4. Recipient cells included in each scenario of simulated data analyzed.

Scenario 1	**All grid cells sampled**	**2 x 64 cells separated by various distances**
Scenario 2	Sampling of four equidistant cells located on diagonal	**A**: 32 x 4 cells separated by 0.00 m **B**: 32 x 4 cells separated by 0.71 m **C**: 32 x 4 cells separated by 1.42 m **D**: 32 x 4 cells separated by 2.12 m
Scenario 3	Sampling of 8 cells separated by 2 different distances; six versions with varying distances	**E**: 16 x 4 cells separated by 0.00 m & 16 x 4 cells separated by 0.71 m **F**: 16 x 4 cells separated by 0.00 m & 16 x 4 cells separated by 1.42 m **G**: 16 x 4 cells separated by 0.00 m & 16 x 4 cells separated by 2.12 m **H**: 16 x 4 cells separated by 0.71 m & 16 x 4 cells separated by 1.42 m **I**: 16 x 4 cells separated by 0.71 m & 16 x 4 cells separated by 2.12 m **J**: 16 x 4 cells separated by 1.42 m & 16 x 4 cells separated by 2.12 m
Scenario 4	Sampling of 16 cells separated by 4 different distances	**K**: 8 x 4 cells separated by 0.00 m & 8 x 4 cells separated by 0.71 m & 8 x 4 cells separated by 1.42 m & 8 x 4 cells separated by 2.12 m

For each combination in each scenario the log likelihood function (being the sum of all log-likelihood functions of all groups in the combination) was automatically constructed and the algorithm described above was used to maximize the function and generate univariate likelihood profiles for all three parameters. The likelihood profiles for all combinations are provided as Supplementary Figure S3.1.

Experimental data

We used previously published data from *C. jejuni* transmission experiments between spatially separated broilers. The details on experimental design, the assumptions made in the likelihood

formulation, and experimental data were published in Chapter 2 and in (van Bunnik, 2014; van Bunnik et al., 2014). To study practical identifiability, we divided the data into 5 experimental data parts as presented in Table 3.2.

We estimated parameters for all 31 combinations of experimental data parts: 1 combination with all 5 parts, 5 combinations with 4 parts, 10 combinations of 3 parts, 10 combinations of 2 parts, 5 contaminations of 1 part. For each combination in each scenario the log likelihood function (being the sum of all log-likelihood functions of all data parts in the combination) was automatically constructed and our algorithm was used to maximize the function and generate univariate generated likelihood profiles for all three parameters. The likelihood profiles for all combinations are provided as Supplementary Figures S3.2-S3.4.

Data and code availability

The authors declare that data and computer code developed in Mathematica 12 (Wolfram Research, 2019) supporting the findings of this study is available in Zenodo with the identifier(s) [10.5281/zenodo.7428123] (Gamża, 2023a).

Supplementary Information

Identifiability of environmental transmission parameters: quantifying dispersion and decay of infectious material using spatiotemporal transmission data.

Supplementary Note 1: Analysis of Equation (12)

We investigate if there are any solutions Δu_i to the condition

$$\Delta\big(\beta W_r(t,x|\alpha,D)\big) = \sum_{u_i\in(\alpha,\beta,D)} \Delta u_i \frac{\partial(\beta W_r(\vec{u}))}{\partial u_i} = 0, \tag{S1}$$

where Δu_i is an infinitesimal displacement along a line or surface in parameter space, and according to Equation (11) of Chapter 3:

$$\beta W_r(\vec{u}) = \beta \frac{1}{4\sqrt{D\alpha}} \mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}} \left(1 + \mathrm{Erf}\left[\frac{-x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] + \mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}} \left(-1 + \mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]\right) \right). \tag{S2}$$

The partial derivatives can be calculated as follows:

$$\frac{\partial(\beta W_r(\vec{u}))}{\partial\alpha} = \beta \frac{1}{8(D\alpha)^{3/2}} \mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}} \left(-D + D\mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}} - x\sqrt{D\alpha} - \mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}} x\sqrt{D\alpha} + \boldsymbol{\frac{4D\mathrm{e}^{-\frac{x^2}{4Dt} - t\alpha + \frac{x\alpha}{\sqrt{D\alpha}}}\sqrt{t\alpha}}{\sqrt{\pi}}} + \left(D + x\sqrt{D\alpha}\right)\mathrm{Erf}\left[\frac{x-2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] + \mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}}\left(-D + x\sqrt{D\alpha}\right)\mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] \right) \tag{S3}$$

$$\frac{\partial(\beta W_r(\vec{u}))}{\partial D} = -\beta \frac{1}{8D^2\sqrt{\alpha}} \mathrm{e}^{-\frac{x\alpha}{\sqrt{D\alpha}}} \left(\sqrt{D} - \sqrt{D}\mathrm{e}^{\frac{2x\alpha}{\sqrt{D\alpha}}} - x\sqrt{\alpha} - \mathrm{e}^{\frac{2x\alpha}{\sqrt{D\alpha}}} x\sqrt{\alpha} + \left(-\sqrt{D} + x\sqrt{\alpha}\right)\mathrm{Erf}\left[\frac{x-2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] + \mathrm{e}^{\frac{2x\alpha}{\sqrt{D\alpha}}}\left(\sqrt{D} + x\sqrt{\alpha}\right)\mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] \right), \tag{S4}$$

$$\frac{\partial(\beta W_r(\vec{u}))}{\partial\beta} = \frac{1}{4\sqrt{D\alpha}} \mathrm{e}^{-x\sqrt{\frac{\alpha}{D}}} \left(1 + \mathrm{Erf}\left[\frac{-x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right] + \mathrm{e}^{2x\sqrt{\frac{\alpha}{D}}} \left(-1 + \mathrm{Erf}\left[\frac{x+2t\sqrt{D\alpha}}{2\sqrt{Dt}}\right]\right) \right). \tag{S5}$$

In Equation (S3) we wrote one term in boldface (and with gray background); this term, which has a different functional form (dependence on $[t,x]$) from all other terms in Equation (S3), has no counterpart in (S4) and/or (S5). For that reason, it is clear that no solutions Δu_i exist for which (S1) is satisfied for (at least) two points in space and (at least) two points in time.

Supplementary Figures S1 to S4: Full profile likelihoods

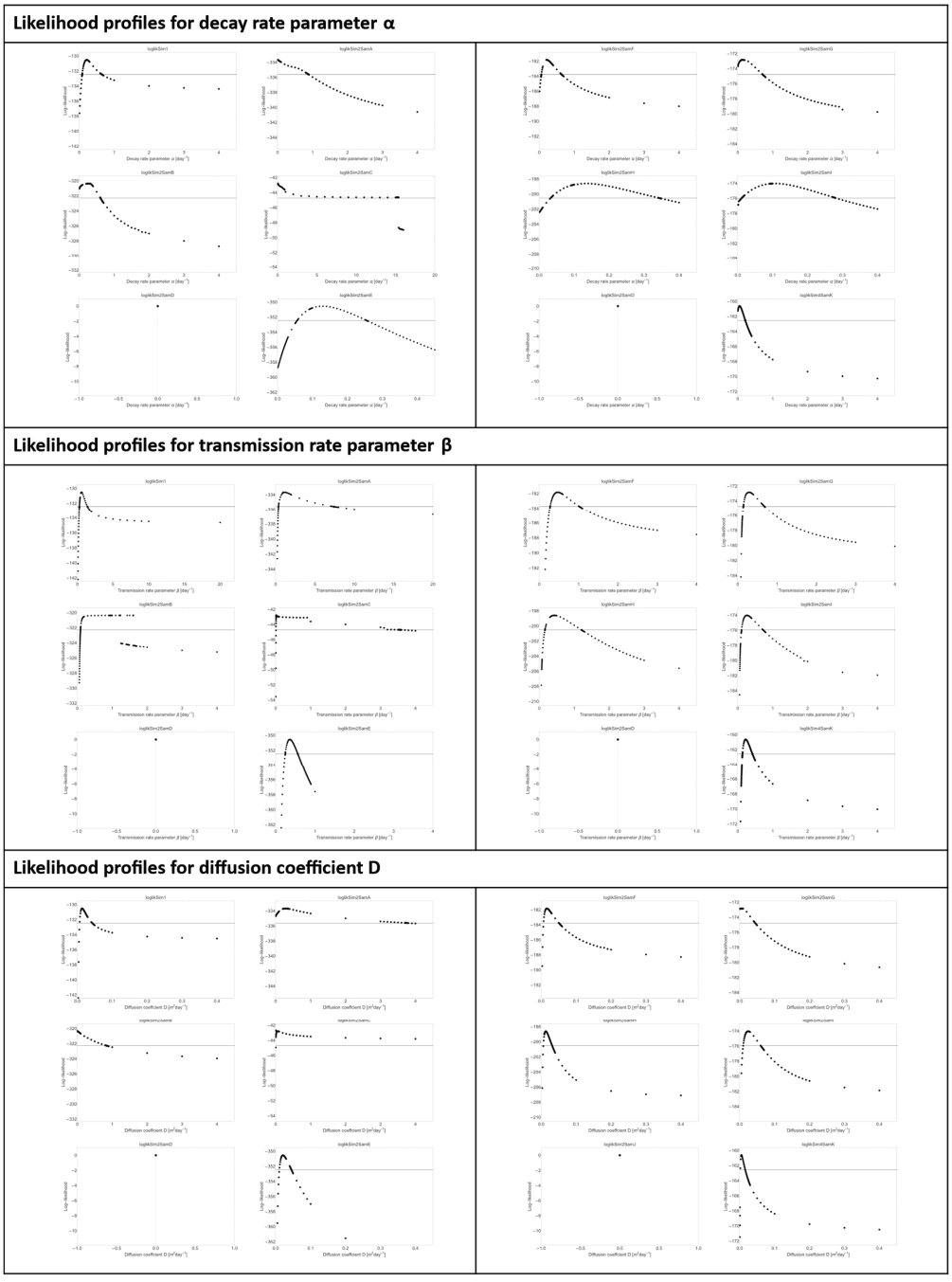


Figure S3.1. Full likelihood profiles for all three model parameters for a set of 12 experimental design scenarios, obtained from simulated data using given parameter values (α=0.15 day^{-1}, β=0.40 day^{-1}, D=0.013 m^{2}day^{-1}); (see main manuscript for details).

Decay rate parameter profile likelihoods for all data part combinations

Experimental data part		Name in the code	Distances covered [m][1]	Start in clean environment	Proportion of infected hosts	Detailed description
Intermediate-Clean	(In-Cl)	**T1exp123**	0.75 and 1.06	YES	20/80	[16, 33]
Range-Clean	(Ra-Cl)	**T1exp4**	0.35–2.00	YES	16/48	Chapter 1
Intermediate-Contaminated	(In-Co)	**T2expgA**	0.70–0.90[2]	NO	5/40	Chapter 1
Zero-Contaminated	(Ze-Co)	**T2expgB**	0.00[3]	NO	10/10	Chapter 1
Zero-Clean	(Ze-Cl)	**T2expgC**	0.00[3]	YES[4]	5/5	Chapter 1

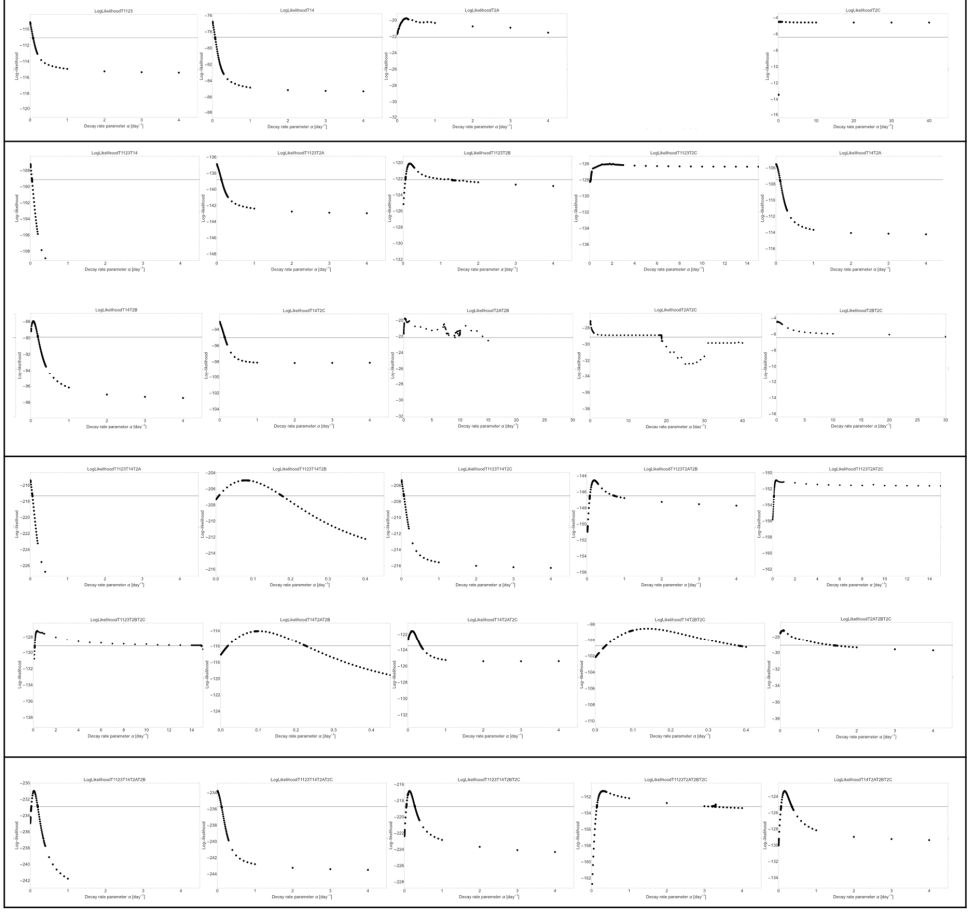


Figure S3.2. Full likelihood profiles for the decay rate parameter α obtained for various combinations of experimental data parts from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design (see main manuscript for details)

Transmission rate parameter profile likelihoods for all data part combinations

Experimental data part		Name in the code	Distances covered [m][1]	Start in clean environment	Proportion of infected hosts	Detailed description
Intermediate-Clean	(In-Cl)	**T1exp123**	0.75 and 1.06	YES	20/80	[16, 33]
Range-Clean	(Ra-Cl)	**T1exp4**	0.35–2.00	YES	16/48	Chapter 1
Intermediate-Contaminated	(In-Co)	**T2expgA**	0.70–0.90[2]	NO	5/40	Chapter 1
Zero-Contaminated	(Ze-Co)	**T2expgB**	0.00[3]	NO	10/10	Chapter 1
Zero-Clean	(Ze-Cl)	**T2expgC**	0.00[3]	YES[4]	5/5	Chapter 1

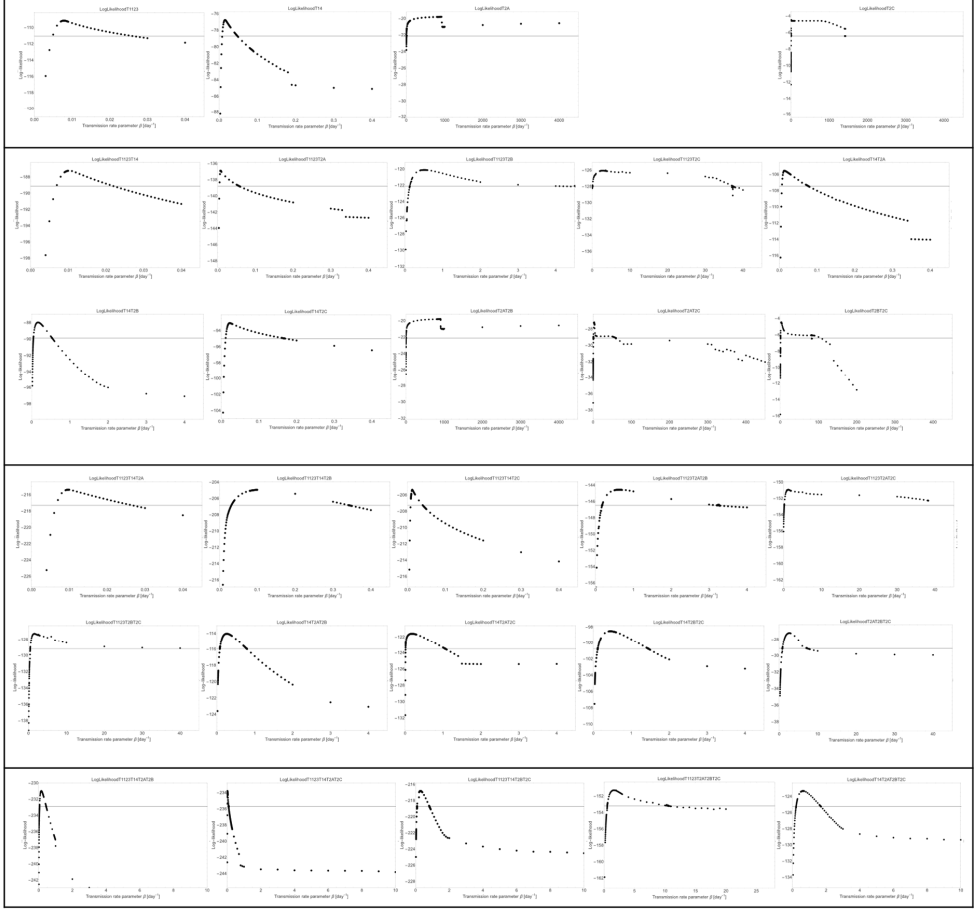


Figure S3.3. Full likelihood profiles for the transmission rate parameter β obtained for various combinations of experimental data parts from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design (see main manuscript for details)

Diffusion coefficient profile likelihoods for all data part combinations

Experimental data part		Name in the code	Distances covered [m][1]	Start in clean environment	Proportion of infected hosts	Detailed description
Intermediate-Clean	(In-Cl)	T1exp123	0.75 and 1.06	YES	20/80	[16, 33]
Range-Clean	(Ra-Cl)	T1exp4	0.35–2.00	YES	16/48	Chapter 1
Intermediate-Contaminated	(In-Co)	T2expgA	0.70–0.90[2]	NO	5/40	Chapter 1
Zero-Contaminated	(Ze-Co)	T2expgB	0.00[3]	NO	10/10	Chapter 1
Zero-Clean	(Ze-Cl)	T2expgC	0.00[3]	YES[4]	5/5	Chapter 1

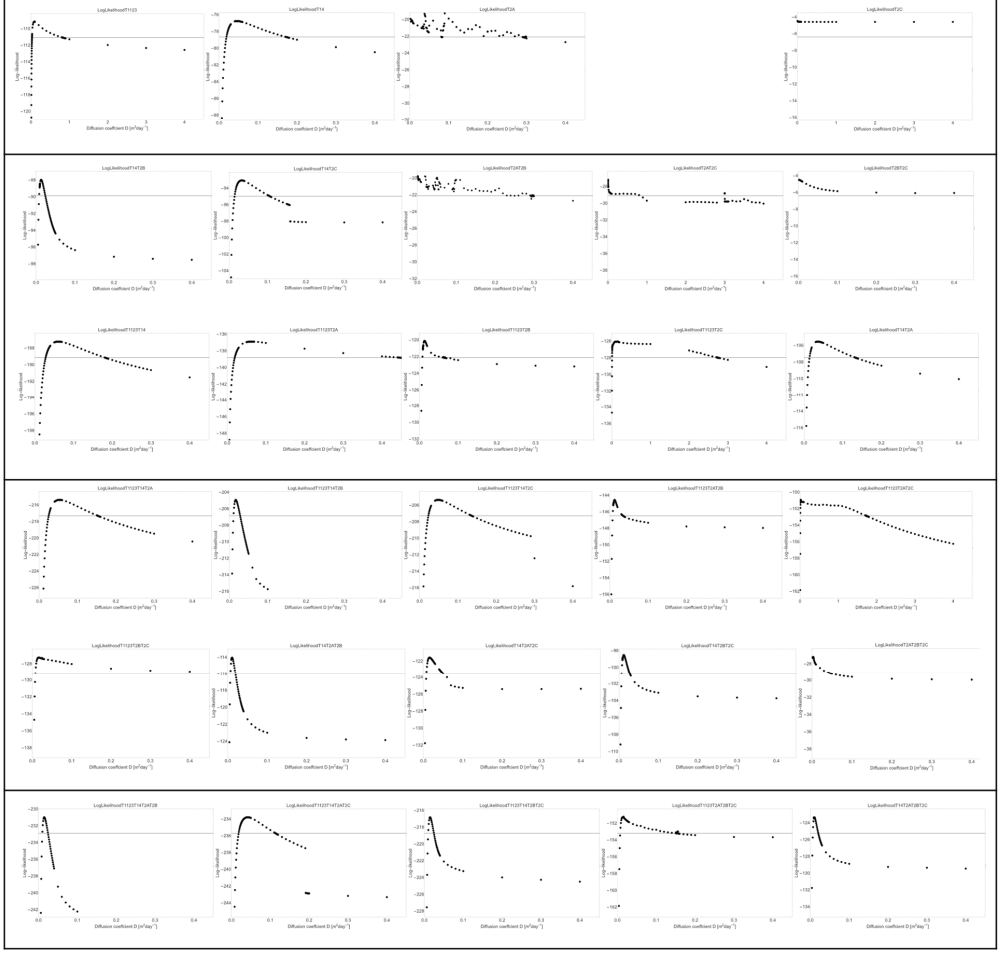


Figure S3.4. Full likelihood profiles for the diffusion coefficient obtained for various combinations of experimental data parts from Campylobacter transmission experiments between spatially separated broilers differing in spatio-temporal design (see main manuscript for details)

4

Chapter 4

Using spatial modelling to explore the host-density dependence of R_0 for environmental transmission

Anna M. Gamża[1,2], Mart C.M. de Jong[1], Thomas J. Hagenaars[2]

1) Quantitative Veterinary Epidemiology, Wageningen University & Research, 6708 PB Wageningen, the Netherlands
2) Wageningen Bioveterinary Research, Wageningen University & Research, 8221 RA Lelystad, the Netherlands

Abstract

Spatial host density can be an important determinant of transmission in host-pathogen-environment systems, particularly in situations where social behaviour of the hosts is strictly confined within a given area such as for a seated audience or for livestock separated in smaller groups. Considering that most infectious agents are transmitted between hosts via the environment, here we explore the use of environmental transmission models to understand how transmission depends on host density. We first provide a generic viewpoint based on analytical arguments and subsequently study a number of specific spatial host mixing scenarios. The numerical study uses a parsimonious spatial modelling approach in which both infectious and recipient (susceptible) hosts are represented by individual areas they occupy, and where the host mixing scenarios are defined by constraints on host movement. Our results suggest that for uniform, unconstrained spatial host mixing, a linear dependence of R0 on host density is obtained, as expected. For a scenario where the movement of each host is constrained within a sufficiently small individual area while these areas can overlap or be adjacent, the linear dependence is still present, although the R0 is much lower than for the unconstrained scenario. As we argue from analytical reasoning, for mixing scenarios with either host clustering behaviour or spatial distancing of hosts, non-linear relationships between R0 and host density are expected. In the host clustering scenario, for a given host density the R0 is bigger than in the uniform mixing scenario, while in the spatial distancing scenario the opposite result is obtained. The deviations from linearity can be substantial, and our numerical results obtained for spatial distancing scenarios provide an indication for which scenarios and parameter settings these deviations cannot be neglected. In such cases, obtaining data on the density dependence of R_0, for example in animal transmission experiments, could potentially allow inference on the host mixing pattern in combination with the spatial range of environmental transmission.

Introduction

Spatial host density can be an important determinant of transmission risk in host-pathogen-environment systems. In particular, when social behaviour of the hosts is strictly confined within a given area such as for livestock housed in small, separated groups or for a seated audience, spatial host density is a parameter that can both be clearly defined as well as be relevant to transmission. The problem of how transmission scales with density of hosts is important for the extrapolation of small-scale animal transmission studies to field situations with a different animal density. Additionally, a difference in host density may be encountered when comparing one husbandry system to another (e.g. intensive and organic animal farming) or when comparing different periods in time for the same system (e.g. before and after thinning in broiler farming). We note that for constant host density, the related scaling issue of transmission with population size is addressed in (Bouma et al., 1995; De Jong et al., 1994)

The notion of host density as a parameter underlying pathogen transmission and control extends of course to pathogens in human populations as is apparent from intervention measures during the recent SARS-CoV-2pandemic that were not only based on host distancing but also implicitly on reducing the local host density (Nightingale et al., 2021; Smith et al., 2021). Among these are the intervention strategies aimed to limit the number of hosts that can be present in a given space (e.g. visitors' or travellers' limits), and the ones assigning spatial locations (seats) to hosts.

Thus far, the host-density dependence of pathogen transmission has not been studied in detail from an environmental transmission viewpoint. Related earlier work is mostly directed towards answering the question how the transmission risk scales with population size in non-spatial models (Bouma et al., 1995; De Jong et al., 1994). While change in population size may result in change of density it is not always the case. Yet the earlier work did consider host density as a determinant, as is manifest from the concept of "density dependent mixing" next to "frequency dependent mixing" (Begon et al., 2002).

As for most host-pathogen systems the pathogen is transmitted from one host to another via the environment we believe that it has a broad relevance to study the effect of host density on transmission from an environmental transmission viewpoint. Environmental transmission may occur when a recipient host is exposed to infectious material at a certain location in the environment, which has accumulated there as a result of previous shedding by infectious hosts, not necessarily at that location, and subsequent dispersal and decay of the infectious material. Therefore, we here use a mathematical model for environmental transmission that takes into account the processes involved: shedding, dispersal, decay and absorption of infectious material, and the response to the dose that is absorbed. The environment in the model is explicitly spatial such that the release of infectious material at the locations of infectious individuals through time can be mathematically "propagated" to the exposure arising at a given time at the location of a

susceptible individual (van Bunnik et al., 2014)(Chapter 2). Using the spatial approach, we discuss density dependence of R0 for uniform mixing, as well as clustering and distancing of hosts. We defined a number of specific spatial host mixing scenarios by imposing constraints on host movement. Next, we explored the host-density dependence of transmission, by calculating how the density of hosts may influence the basic reproduction ratio R_0 for these different scenarios.

Methods

Diffusion-based environmental transmission model:

The spatial environmental transmission model we use was previously published in (van Bunnik et al., 2014)(Chapter 2). The two-dimensional diffusion is a parsimonious model for the dispersal of infectious material through the environment through time, and assumes that the infectiousness of the material declines exponentially with sojourn time in the environment (i.e. time since shedding). Each individual host is described with a location of a movement area of given dimensions - the source area A_{inf}. Shedding of infectious material by infectious hosts is assumed to lead to a uniform contamination of the host movement area. Similarly, the exposure level experienced by susceptible individuals is assumed to be determined by the average infectious material load across their movement area – the exposure area A_{exp}. The infectious material is described in our model by an environmental load density function $W(t, x, y|\rho, \alpha, D)$ which serves as a continuous, mean-field description of the spatio-temporal distribution of infectious material. As described in Chapter 3, the relative density function $W_r(t, x, y|\alpha, D) \equiv \frac{1}{\rho} W(t, x, y|\rho, \alpha, D)$ depends only on two parameters, namely the decay rate parameter α and the diffusion coefficient D, whilst $W(t, x, y|\rho, \alpha, D)$ is the solution of the two-dimensional reaction-diffusion equation with decay:

$$\frac{\partial W}{\partial t} = D\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) - \alpha W + \rho s, \qquad (1)$$

where $s = s(t, x, y)$ is the spatiotemporal infectivity pattern determined by the location and dimensions of the movement areas of each infectious host and their shedding period, during which the infectious material is assumed to be produced continuously with constant shedding rate ρ and uniformly across the host's movement area. For reasons of unidentifiability, the relative environmental load density function is defined ($W_r(t, x, y|\alpha, D)$), and magnitude of the shedding rate ρ is absorbed into the transmission rate parameter β, and its unit of 1/day is absorbed into the infectivity pattern s. Considering an infectivity pattern s consisting of one movement area shared by a number of infectious (i.e. infectivity shedding) hosts, we define the source strength configuration parameter Q for that area as the number of infectious hosts per one day divided by the size of the movement area in m^2.This scaling corresponds to the assumption that when a given host has larger area available, it will spend on average less time in one particular location, resulting

in correspondingly reduced shedding at that location. Then the solution to Equation 1 reads as follows:

$$W_r(t,x,y|\alpha,D) = Q\sum_I \int_{T_1^i}^{\min(t,T_2^i)} \iint_{A_{\text{inf}}^i} \frac{1}{4\pi D(t-\tau)} \exp\left[-\alpha(t-\tau) - \frac{(x-x_i)^2+(y-y_i)^2}{4D(t-\tau)}\right] dx_i\, dy_i\, d\tau, \quad (4)$$

where $\sum_I$ is the sum over the set of source areas, described with source infectivity pattern s. The infection probability across an exposure time interval $[t_1, t_2]$ of a given recipient individual with exposure area A_{exp} is given by:

$$\mathrm{P}_{\text{inf}}(t_1,t_2,A_{\text{exp}}) = 1 - \exp\left[-\beta\, E \int_{t_1}^{t_2} \iint_{A_{\text{exp}}} W_r\,(t,x,y|\alpha,D)\, dx\, dy\, dt\,\right], \quad (5)$$

where E is the exposure dilution configuration parameter defined as the host density of the recipient exposure area (measured as one host per size of the exposure area in m^2). On the one hand, the exposure model is based on the notion that in some unspecified manner, recipient hosts absorb infectious material from the environment. On the other hand, the absorption of infectivity is assumed to be small enough to neglect any resulting reduction in the remaining environmental load. In line with this, in our model there is also no competition between hosts for the infectious material when their movement areas overlap. The definitions of the three model parameters and their units are listed in Table 4.1, spatiotemporal configuration parameters as used in the calculated scenarios are listed in Table 4.2.

Table 4.1. Parameters of the spatial model of environmental transmission.

Parameter	Symbol	Units
Decay rate parameter	α	[day^{-1}]
Transmission rate parameter	β	[day^{-1}]
Diffusion coefficient	D	[$m^2 day^{-1}$]

Table 2. Spatiotemporal configuration parameters as used in the calculated scenarios.

Parameter	Symbol and value	Units
Shedding period of the infected hosts	$[T_1^i, T_2^i]$= (3, 10)	[day]
Exposure period of the recipient hosts	$[t_1, t_2]$= (0, 100)	[day]
Source strength parameter	Q	[$m^{-2} day^{-1}$]
Exposure dilution parameter	E	[$m^{-2} day^{-1}$]

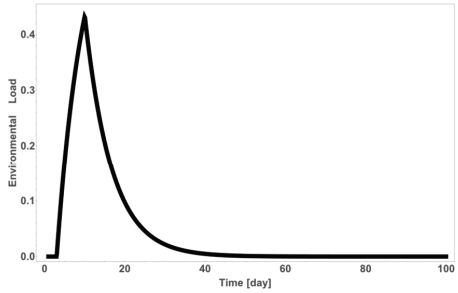
Spatial host mixing scenarios and R_0

First, we will use analytical arguments to derive the expected shapes of the host-density dependence of R_0 for three types of spatial host mixing: spatially uniform mixing, host clustering and host distancing. Second, to confirm the findings from the analysis we use the diffusion-based environmental transmission model to explore several specific homogeneous mixing and distancing scenarios in more detail. This is done by numerically calculating R_0 as the expected value for the number of newly infected recipients by a single primary infective with source area A_{inf} and shedding period $[T_1, T_2]$ in an otherwise susceptible population as follows:

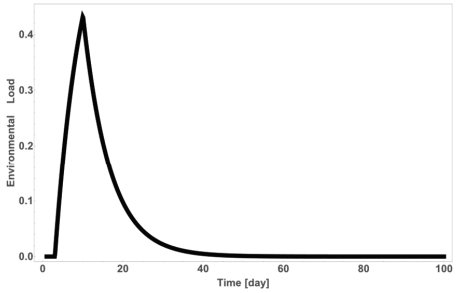
$$R_0 = \sum_r \mathrm{P}_{\text{inf}}\left(t_1, t_2, A^r_{\text{exp}}\right). \quad (6)$$

To keep the computational intensity tractable, this calculation of R_0 neglects local depletion effects that are to be expected if transmission takes place across a sufficiently short spatial range in a spatial setting (Boender et al., 2007; Danon et al., 2011; Diekmann et al., 1998). For all scenarios we assume that all individuals have a movement area of the same size (but in general, each one is located at a different point in space). For ease of calculation, the locations of all individual recipient movement areas are chosen from a regular two-dimensional grid. For scenarios in which movement area configurations were chosen randomly, we performed 100 (scenario III) or 50 (scenario IV) realizations for each host density considered. Full spatial configurations for all realizations are provided with the code in Zenodo [10.5281/zenodo.7430198](Gamża, 2023b). We use absorbing boundary conditions, which means that infectious material can diffuse outside the total environmental area considered. We use a 10m x 10m total area with grid cell size of 0.25m x 0.25m, i.e. consisting of 1600 grid cells. In order to explore how the decay rate parameter and diffusion coefficient influence the host-density dependence of R_0, we considered four sets of parameter values, all with constant transmission parameter β=0.37 day^{-1}, and with values for α and D as listed in Fig. 4.1.

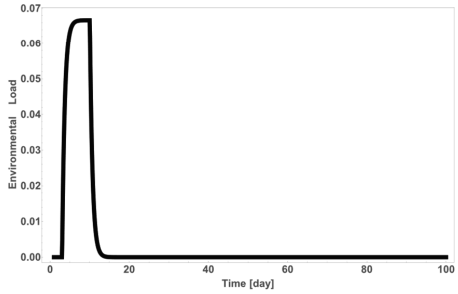
Parameter set 1
α=0.15 day^{-1}, D=0.013 m^{2}day^{-1}



Parameter set 2:
α=0.15 day^{-1}, D=1.3 m^{2}day^{-1}



Parameter set 3:
α=1.5 day^{-1}, D=0.013 m^{2}day^{-1}



Parameter set 4:
α=1.5 day^{-1}, D=1.3 m^{2}day^{-1}

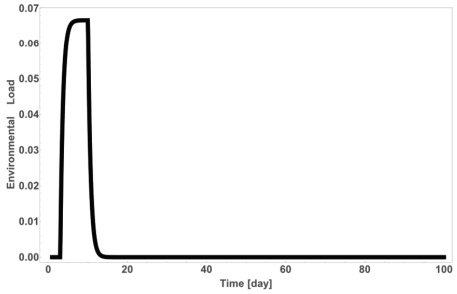


Figure 4.1 Time evolution of the total environmental load integrated across an environmental area of 100 m^2, for a primary infectious host shedding from T_1=3 day to T_1=10 day.

Parameter set 1 corresponds to parameter estimates obtained for Campylobacter in broilers based on data from indirect transmission experiments (Chapter 2). In addition, a stocking density of one broiler per 0.25m x 0.25m grid cell, i.e. 16 broilers per square meter, is representative for stocking densities on broiler farms in the EU (Council Directive 2007/43/EC, 2007). The infectious individual, located in the centre of the grid, was assumed to have a shedding period extending from day 3 (T_1) until day 10 (T_2), and the total exposure time for the recipient hosts was 100 days, which for all parameter sets considered was long enough for the total environmental load to have declined to a negligible level, as can be seen from Fig. 1.

In Fig. 4.2 we show the spatial shape of the time-integrated total environmental load for the four parameter sets considered, illustrating how these parameters sets differ in terms of the spatial range covered by viable infectious material spreading outside the movement area of the shedding host. Whilst for parameter set 2 the spatial range covered by viable infectious material extends for a small part to outside the total environmental area considered, for the other three parameters sets this is not the case and therefore for these three sets no strong border effect is expected to occur in the calculation of R_0.

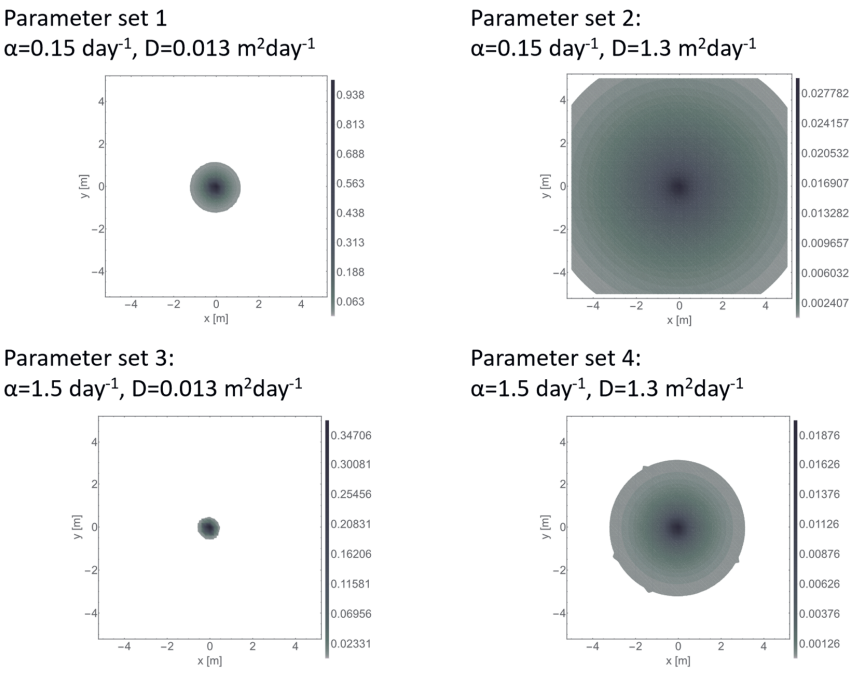


Figure 4.2. Spatial shape of the time-integrated environmental load for a primary infectious host shedding from T_1=3 day to T_1=10 day, integrated across an observation period from t_1 =0 day to t_2 =100 day.

Detailed spatial host mixing scenarios

An overview of the five detailed mixing scenarios is given in Table 4.1.

Table 4.1. Overview of five detailed mixing scenarios. The scenarios I and III are of 'spatially homogeneous mixing' type; the scenarios II, IV and V are of 'host distancing' type.

Scenario	Host movement within local area	Random configurations
Scenario I: Uniform mixing	No	N/A
Scenario II: Uniform host density with non-overlapping movement areas	Yes	No
Scenario III: Poisson-process mixing	Yes	Yes
Scenario IV: Poisson-process mixing with non-overlapping movement areas	Yes	Yes
Scenario V: Maximal distancing	Yes	No

Scenario I: Uniform mixing

All host movement areas fully overlap, all being equal to the full environmental area; this means that all hosts can freely move across the whole environmental area whilst either shedding or absorbing infectious material. This scenario corresponds to a situation where no particular piece of environment is preferred by any particular host, hosts are moving randomly independent of other hosts, and the speed of movements is fast enough for an individual to (to a good approximation) uniformly visit the full environmental area well within the time period of shedding and/or exposure.

Scenario II: Uniform host density with non-overlapping movement areas:

The full environmental area is divided into adjacent host movement areas, with each area being occupied by one host. For this scenario, different values for the host density were obtained by scaling the size of the total area, whilst keeping the total number of 39601 movement areas constant, the big sample size was possible because of the symmetry of the system only a portion of areas had to be sampled to extrapolate over whole area (see Fig. 4.3).

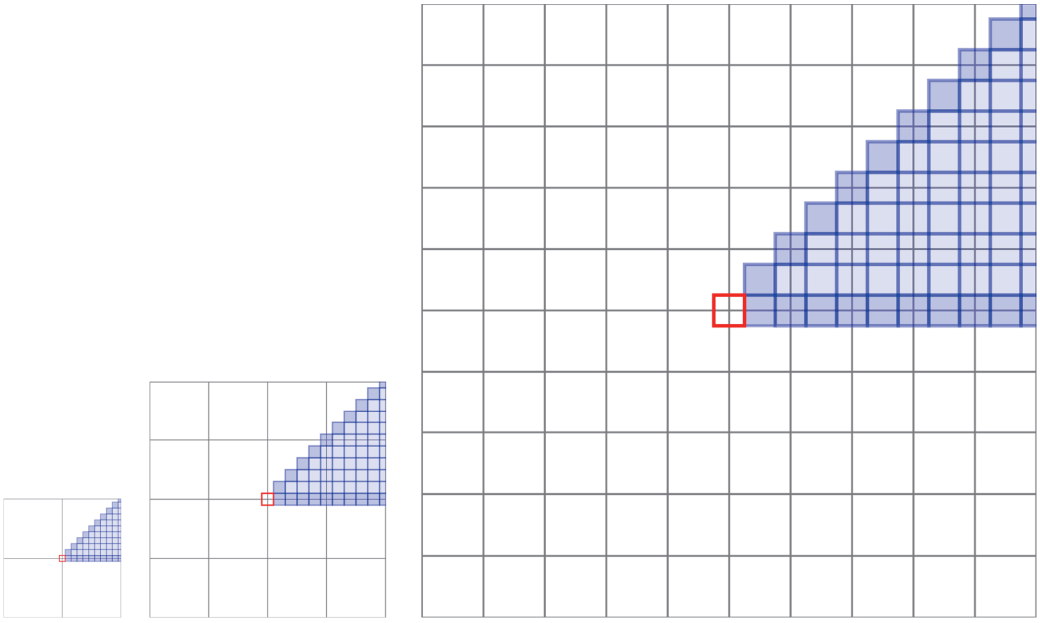


Figure 4.3. Uniform host density with non-overlapping movement areas: three examples of the different area scales considered to obtain results for different host density values; Shown is only a central fragment of the total environmental area; the source of infection localised in the middle (red rectangle) due to the symmetry of the sampled area, probability of infection was sampled only for a fragment of the grid (blue rectangles) and extrapolated to the whole area.

Scenario III: Poisson-process mixing

Individual host movement areas are small compared to the full environmental area, can overlap and are placed randomly according to a two-dimensional Poisson process (see Fig. 4.4 for the example of randomly generated host movement areas). This scenario and all further scenarios below correspond to situations where hosts are moving only through a small piece of the environment within the period of shedding and/or exposure.

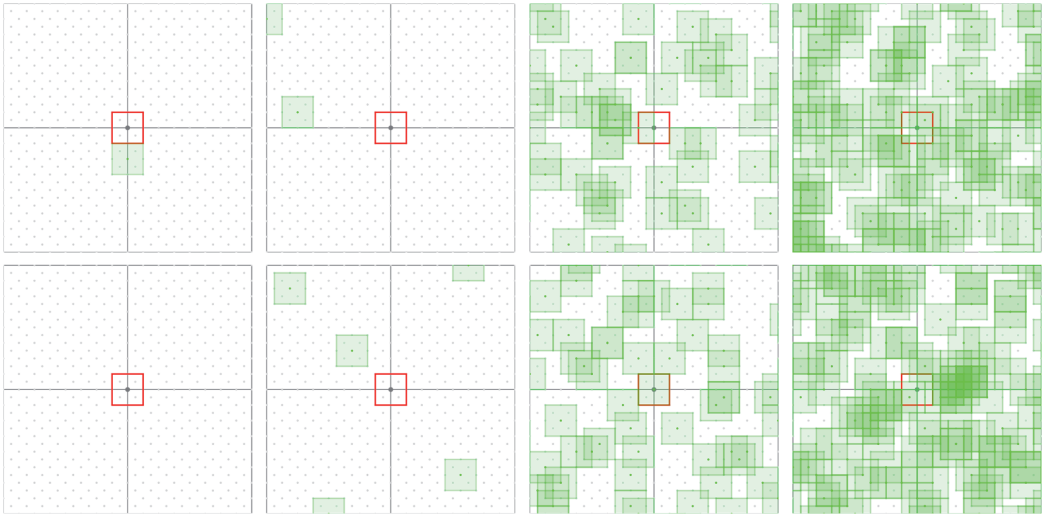


Figure 4.4. Poisson-process mixing: Two examples of randomly generated host movement areas, based on the two-dimensional Poisson process, for each of four different host-density values. Shown is only a central fragment of the total environmental area; the source of infection located in the middle of the grid (red rectangle); the random realizations were constructed by first drawing the total number of hosts from a Poisson distribution with a mean corresponding to the value of the host density considered and then randomly adding the host movement areas; partial or total overlap of areas was allowed.

Scenario IV: Poisson-process mixing with non-overlapping movement areas

Host movement areas cannot overlap and are placed randomly (see Fig. 4.5 for the example of randomly generated host movement areas). This scenario corresponds to situations where hosts have exclusive movement territories.

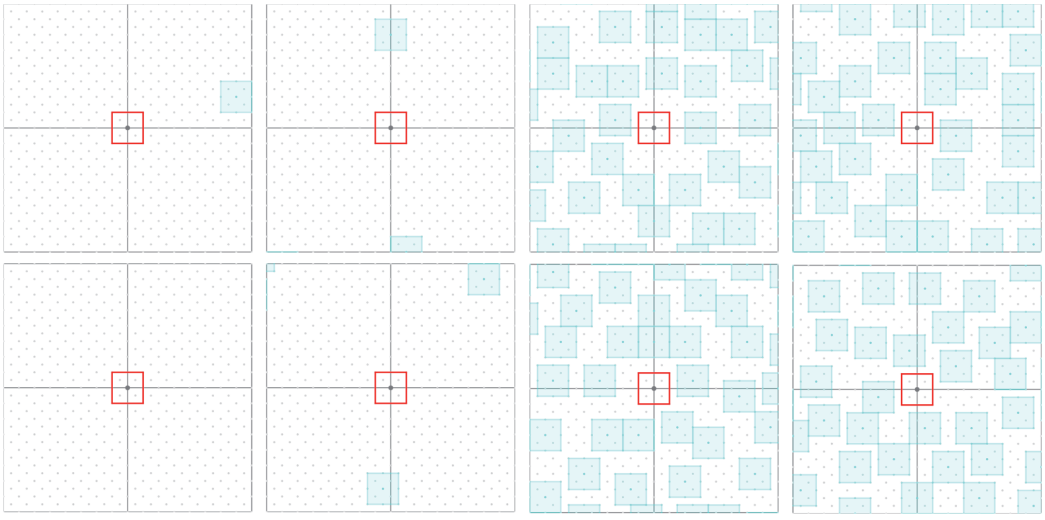


Figure 4.5. Random local mixing with non-overlapping host movement areas: two examples of randomly generated host movement areas for each of four different host-density values. Shown is only a central fragment of the total environmental area; the source of infection located in the middle of the grid (red rectangle); the random realizations were constructed by first drawing the total number of hosts from a Poisson distribution with a mean corresponding to the value of host density considered and then adding the host movement areas one by one to the grid using randomly proposed locations not creating overlap with movement areas already present, until the total number of hosts was reached or until there was no suitable location left (if the second case occurred, for that data point the host density was calculated a posteriori as the mean across the 50 realizations).

Scenario V: Maximal host distancing

Individual host movement areas are placed with maximum distance to their neighbours' (see Fig. 4.6). This model corresponds to situations where host have assigned spots or seats such that they are separated for each other by the largest distance possible for the given host density value, e.g. seating of audiences or placement of cages in animal exhibitions. For high density values we allow host movement areas to overlap.

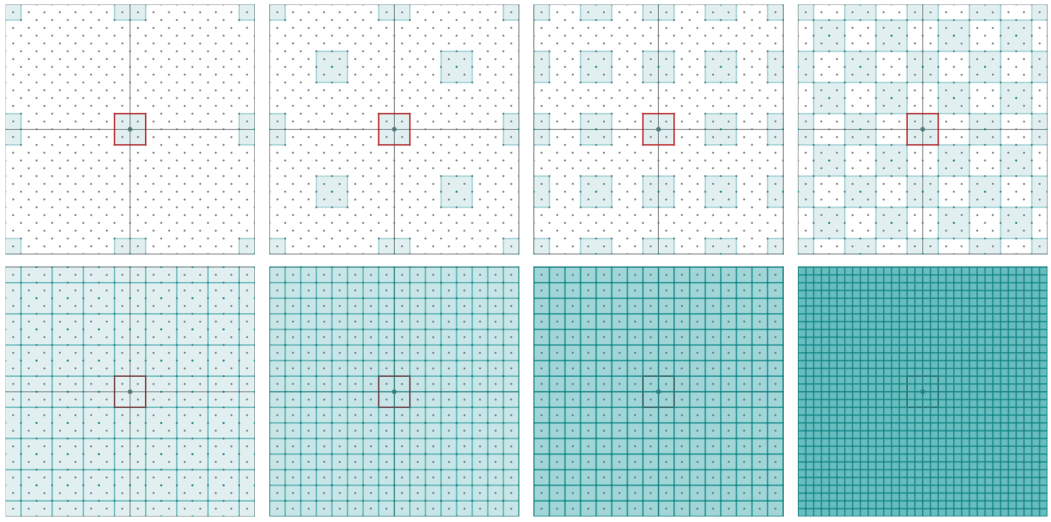


Figure 4.6. Maximal host distancing: configurations of host movement areas used for increasing host density. Shown is only a central fragment of the total environmental area; the source of infection located in the middle of the grid (red rectangle); for each density areas were distributed such that they are separated from each other by the maximum distance possible, considering the fixed environmental area available.

Results

Analytical arguments

We assume a given finite spatial range of transmission around a movement area of an infectious individual beyond which the environmental load is negligible, as illustrated by Fig. 4.2.

Case 1

In the first case, we consider situation when all individual host movement areas coincide with the full environmental area considered (this applies to the uniform mixing scenario I). As the areas where infected individuals shed infectious material and the exposure areas of recipient individuals coincide, the model is reduced to an effectively non-spatial model. Under the assumption, made in the environmental transmission model we consider, that the loss of environmental load due to absorption of infectious material by a recipient host is negligibly small, it is easy to see that the expected number of secondary cases arising from a single infectious host is then proportional to the total number of hosts minus one (discounting the primary infective itself). For a large enough

population this is well approximated by the total number of hosts, and therefore R_0 is, to a good approximation, proportional to the host density.

Case 2

In this case, the individual host movement ranges are smaller than the spatial range of transmission (this applies to mixing scenarios III-V). First, we note that, assuming a given spatial range of transmission (which we denote by r_{T}), we may for any specific location define *the local host density* using a circular area centered at that location and with a radius equal to the range of transmission; namely as $\frac{n}{\pi(r_{\mathrm{T}})^2}$, with n the number of hosts with movement area that is located within the circular area of transmission. Second, we note that the expected number of secondary infections from a given primary infective at any specific location is proportional to the local density of neighboring individuals within the range of transmission, i.e. within the circular area. In other words, this number is proportional to $X-1$, where X is the local number of hosts within the circular transmission area, and the -1 serves to discard the primary infective from the count. R_0 can then be written as proportional to the average value of $X-1$ across all possible selections of the primary infective within and across random spatial host configurations. Considering that the local number of hosts from which the primary infective is randomly selected in the calculation of this average is (again) proportional to the local host density, this average value can be written as being proportional to a weighted spatial average as follows:

$$R_0 \sim \frac{\langle X(X-1)\rangle}{\langle X\rangle} = \langle X\rangle + \frac{\sigma^2(X)}{\langle X\rangle} - 1. \qquad (7)$$

This is reminiscent of a similar result for a "network of acquaintances"(Diekmann et al., 1998), and it states that R_0 is proportional to the sum of the mean host density $\langle X\rangle$ and the host-density variance-to-mean ratio minus one. For Poisson variation in the local host density, the variance-to-mean ratio equals one so that the relationship simplifies to $R_0 \sim \langle X\rangle$, i.e. it states that the transmission intensity scales linearly with host density (corresponding to the 'density-dependent mixing' assumption). If the local density distribution is over-dispersed (variance-to-mean ratio >1), which is the case if hosts tend to cluster (flock together) e.g. due to social behavior or preference for certain parts of environment, Equation (7) predicts that this enhances R_0 relative to a situation with Poisson variation. In contrast, if the local density distribution is under-dispersed (variance-to-mean ratio <1), which is the case if hosts avoid each other e.g. due to social distancing, Equation (7) predicts that this reduces R_0 relative to a situation with Poisson variation. Finally, we note that for high host densities, deviations from Poisson variation in the local host density will start to average out within the local transmission range. This implies that for high host density, the density-dependence curves for Poisson-process mixing, host clustering and host distancing will be approaching each other asymptotically. From these considerations it follows that for host

clustering and host distancing scenarios the scaling of transmission intensity with host density will exhibit the nonlinearities depicted in Fig. 4.7.

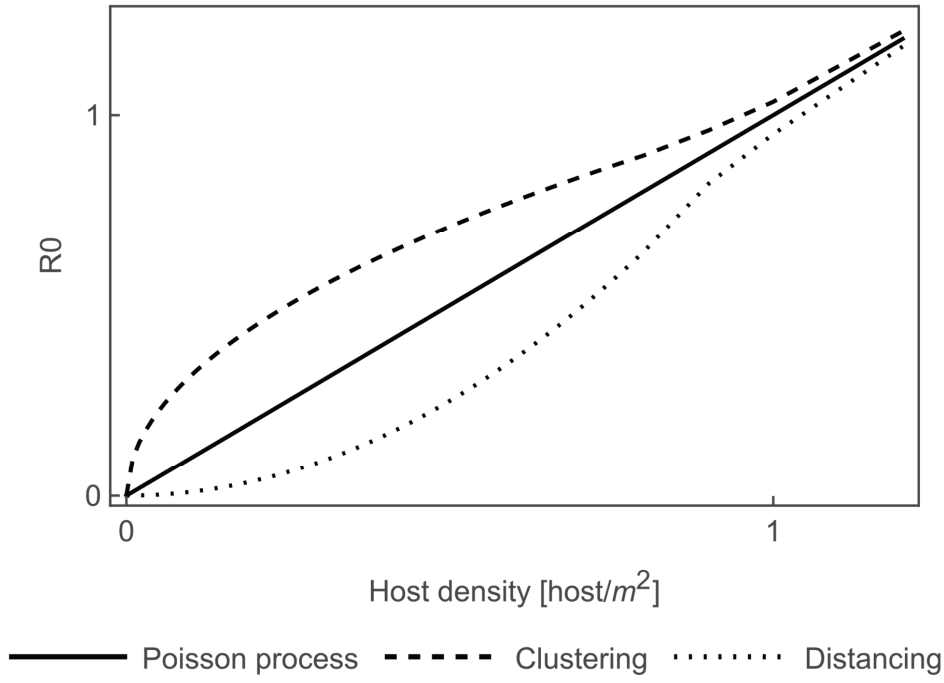


Figure 4.7. General picture for the host-density dependence of R_0 for environmental transmission depending on the spatial host-mixing scenario, as arising from Equation (7) and further considerations.

Results for detailed scenarios

Density dependence curves generated for the first three scenarios and for all parameters sets are presented in Fig. 4.8.

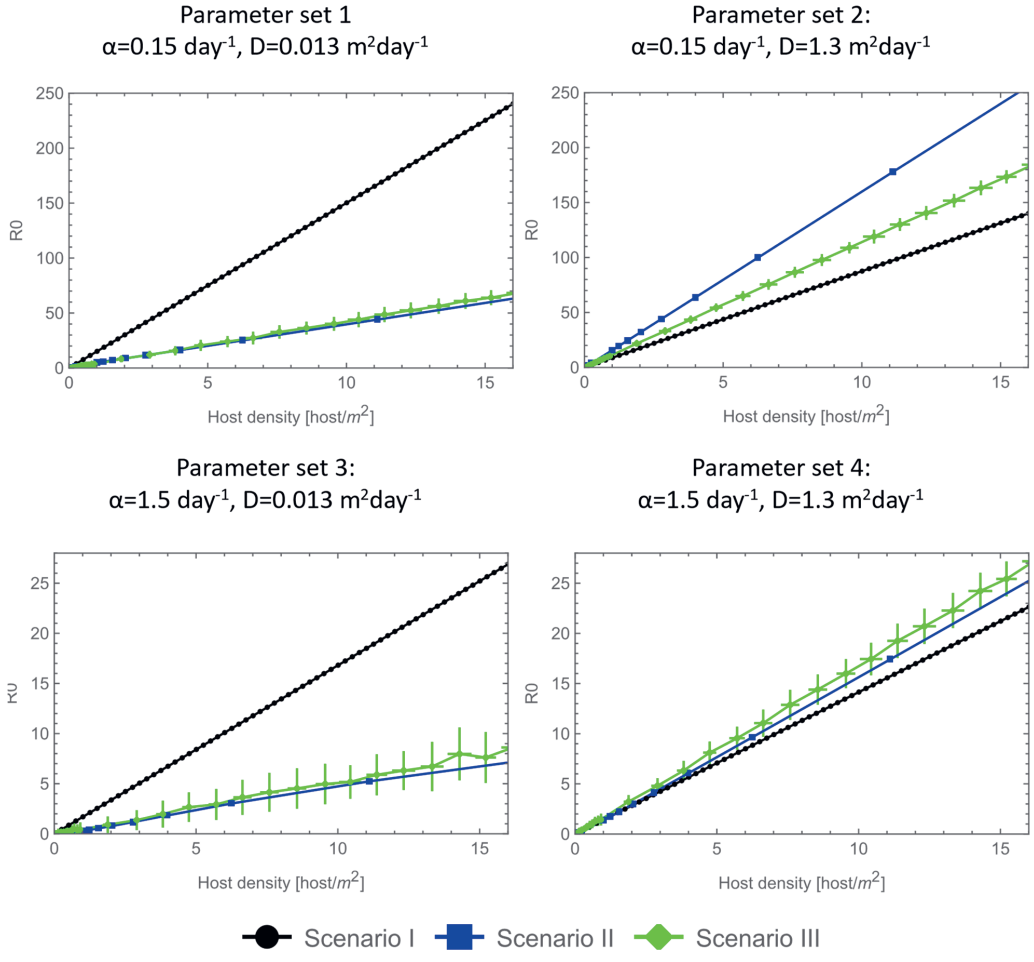


Figure 4.8. Host-density dependence curves for three scenarios: Scenario I (Uniform mixing), Scenario II (Uniform host density with non-overlapping movement areas), Scenario III (Poisson-process mixing) for four parameter sets.

For parameter set 1 and 3, for all the densities the R0 value is much larger for the uniform mixing scenario (Scenario I), where host were allowed to move through the whole environment, than for than for Scenarios II and III where host movements were constrained: uniform host density with non-overlapping movement areas (Scenario II) and Poisson-process mixing (Scenario III). This indicates that intervention strategies aiming at constraining movements of hosts would be effective for systems where the diffusion of infectious material, across a median between-host distance, is slow relative to the infectivity decay. For parameter sets 2 and 4, for all densities the value of R0 is smaller for the uniform mixing scenario than for remaining two scenarios. As these two parameters sets characterise systems with relatively fast diffusion and the environment is assumed to have an absorbing boundary, in the uniform mixing scenario some portion of infectious material is diffusing away from the whole environment. Assuming reflective boundary conditions

(such that all material stays in the environment), for all densities the value of R0 for the uniform mixing scenario is slightly larger than for the remaining two scenarios (see Fig. S4.1).

Density dependence curves obtained for the scenarios III-V and for all parameters sets are presented in Fig. 4.9.

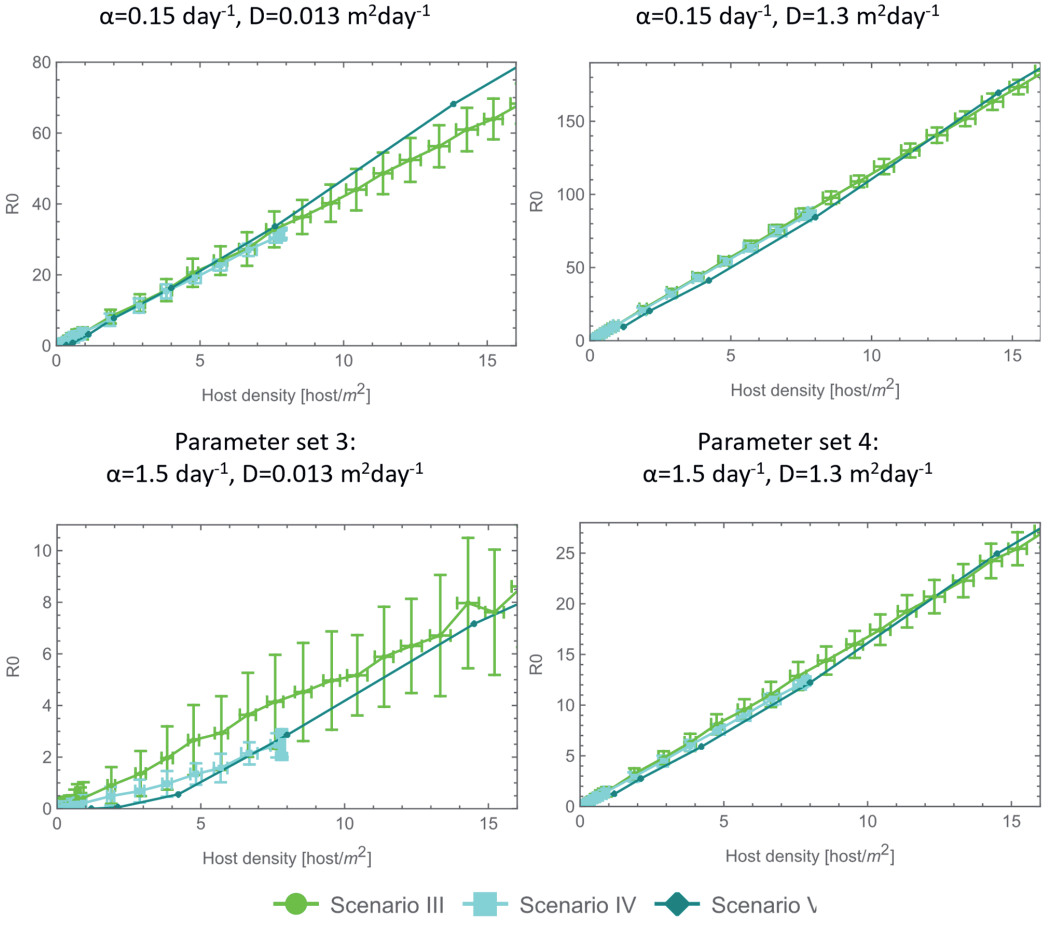


Figure 4.9. Host-density dependence curves for three scenarios: Scenario III (Poisson-process mixing), Scenario IV (Poisson-process mixing with non-overlapping movement areas), Scenario V (Maximal distancing) for four parameter sets.

For the scenarios III-V, the results agree with the pattern predicted in Fig. 4.7 based on our analytical arguments. The Poisson-process mixing yields a linear dependence, and Maximal distancing scenario V shows the expected nonlinearities, which are most pronounced for parameter set 3 which has the shortest transmission range. Scenario IV, Poisson-process mixing with non-overlapping host movement areas, also yields a nonlinear dependence, with R_0 values

below that of the unconstrained Poisson scenario (scenario III). For this scenario, a maximum host density is reached as expected from not allowing host movement areas to overlap.

Discussion

Here, we used an environmental transmission viewpoint on the host-density dependence of R_0. We presented analytical arguments for the expected linear or nonlinear shape of this host-density dependence depending on host spatial mixing characteristics (uniform mixing, clustering and distancing), and demonstrated how a parsimonious spatial environmental transmission model can be used to study this in detail for various mixing scenarios. We considered five such detailed scenarios, including two with different forms of spatial uniformity, and three further ones allowing for spatial variation in local host density.

In scenario I, we assumed uniform mixing, i.e. that hosts can move freely within the whole environment and visit every space point with equal probability. This corresponds to a non-spatial limit of the system, and we observe a linear host-density dependence for each of the considered parameter sets, consistent with expectations based on analytical arguments. When comparing scenario I to scenario II, where host movement areas are non-overlapping, but host are homogenously distributed across the environment, R_0 is reduced in comparison to homogenous mixing (Fig. 4.8) when infectious material spreads slowly. Similar observation can be made when comparing scenario I to scenario IV, where host movements are constrained by a relatively small movement area, but the location of the recipient areas is chosen randomly. This corresponds to the intuition that for the same host density, an event where hosts movements are constrained e.g. have been assigned homogenously distributed spots (e.g. seats) would be significantly safer than event where hosts can move freely for pathogens that spread slowly. Interestingly, when infectious material spreads quickly, R0 is reduced for the uniform mixing scenario as long as it is assumed that infectious material can escape out from the environment. As for our scenarios with constrained hosts the source was located in the middle of environment and thus away from the borders of the environment, only a small portion escaped from the environment, such that hosts were ultimately exposed to more infectious material than in the uniform mixing scenario.

For detailed scenarios where movement ranges of individual hosts are smaller than the spatial range of transmission (scenarios III-V), the results conform to the pattern predicted based on analytical arguments, showing nonlinear behaviour for a maximal distancing scenario that is most pronounced for relatively short-ranged environmental transmission (parameter set 3). For a given host density, the maximal distancing scenario yields a lower R_0 than for the Poisson-process mixing, as is expected for host distancing, cf. safe distance policies, such as those that were implemented during SARS-CoV-2pandemic. A scenario of Poisson-process mixing with non-overlapping host movement areas, also yields a nonlinear dependence, which we interpret as

being a result from the host distancing that is caused by not allowing overlap of host movement areas.

In our model we assumed that recipient hosts are not in competition for infectious material, which means that the exposure experienced by a host at a given location is not influenced by with how many other hosts the access to this location is shared. This assumption may be appropriate e.g. for airborne transmission if the fact that one host inhaled pathogens has negligible influence on the probability of infection for the second exposed host that is located nearby. For systems where the access to infectious material may be reduced when host are sharing the same part of environment, for example because they would need to compete for resources such as for access to food or water source the competition should be implemented in a model by appropriately scaling the force of infection that particular infectious host experiences.

We note that our calculation of R_0 neglects local depletion effects that are to be expected if transmission takes place across a sufficiently short spatial range (Boender et al., 2007; Danon et al., 2011; Diekmann et al., 1998). It remains to be investigated in what way such effects would influence the non-linear host-density dependences observed.

The reduction of R0 for scenarios with constrained hosts indicates that indeed, constraining mobility of hosts is an effective intervention strategy, and the effect size depends on the speed of dispersion (the smaller the diffusion parameter, the more reduction is observed) and the distancing strategy (with the most reduction observed for scenario with maximal distancing).

The deviations from linearity in the density dependence of R0 we observe can be substantial, and our results provide an indication for which scenarios and parameter settings these deviations cannot be neglected. In such cases, obtaining data on the density dependence of R_0 could allow inference on the host mixing pattern combined with the spatial range of environmental transmission.

To explore this type of inference in a controlled way, the density dependence of R0 can be studied in transmission experiments. In the past, transmission experiments were designed to study transmission in populations of varying sizes (Bouma et al., 1995), whilst keeping the density constant. Moreover, experiments with host separated in space were conducted and used previously to validate the environmental transmission model we used here (Chapter 2) (van Bunnik, 2014; van Bunnik et al., 2014). While we studied the distance dependence of transmission in those experiments, the density dependence was not an objective of that work. For human studies, data from small outbreaks in well controlled environments, such as in certain public events or mass transport vehicles can potentially be used to compare the density dependence of R0 for various distancing strategies.

Overall, the methods presented here, once validated with data can be used to quantify the effect of intervention strategies aimed at constraining host movement and/or imposing distancing rules. Moreover, they can be a basis for a systematic study of density dependence needed to compare, aggregate and/or extrapolate the transmission studies.

Data and code availability

The authors declare that data and computer code developed in Mathematica 12 (Wolfram Research, 2019) supporting the findings of this study is available in Zenodo with the identifier(s) [10.5281/zenodo.7430198] (Gamża, 2023b)

Supplementary Information

Using spatial modelling to explore the host-density dependence of R0 for environmental transmission

Supplementary Figure S4.1

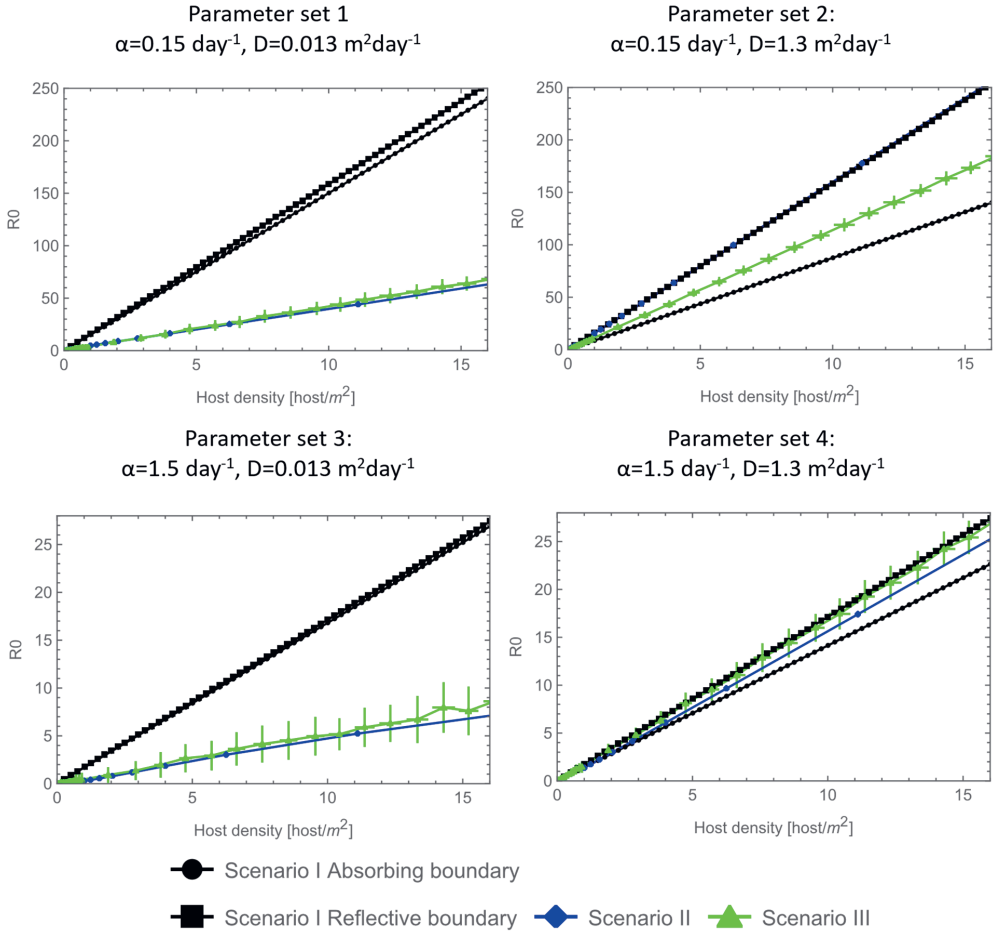


Figure S4.1. Host-density dependence curves for three scenarios: Scenario I (Uniform mixing) for absorbing and reflective boundary conditions, Scenario II (Uniform host density with non-overlapping movement areas), Scenario III (Poisson-process mixing) for four parameter sets.

5

Chapter 5

General discussion

All main mechanistic modelling approaches to infectious disease transmission, such as compartmental mass action based models (Heesterbeek, 2005), network models (Bansal et al., 2010; Danon et al., 2011) or individual based simulation models (Willem et al., 2017) used to study transmission of infectious disease are based on the concept of contact that infectious hosts have with susceptible (recipient) ones. While such approaches are suitable for directly transmitted disease or ones that can be approximated that way, for many host-pathogen-environment systems, the last part: the environment that surrounds hosts should be considered to quantify the transmission and ultimately find and test suitable intervention strategies. While the modelling approaches mentioned are often applied to environmental transmission (for example by introducing the environmental compartment), their contact-based structure hampers the study of the processes underlying the environmental stage of transmission. As these processes are generally not well understood, it is often not known how infectious hosts contact the environment to deposit infectious particles nor how recipient hosts contact the environment to absorb the infectious particles. As has been shown in this thesis and elsewhere (Brouwer, Weir, et al., 2017), those two processes are not separately identifiable during parameter inference using host status data.

In this thesis, I have examined transmission of infection from an infectious material perspective using models based on an environmental load density function that describes the spatiotemporal relationship between the host presence and the probability of transmission. The parsimonious model, presented here, focuses on the two most important processes happening in the environment: the decay and spatial dispersion of infectious material and host dependent process described by joint parameter (transmission rate). In this thesis I have used *C. jejuni* in broilers as a model system to prove that such modelling can be combined with host status data to obtain valuable methodological and biological insights. I went through calibration and validation (Chapter 2, Chapter 3) of our parsimonious environmental transmission model with data from tailor-made transmission experiments and demonstrated how it can be applied to study density dependence of transmission for various systems (Chapter 4).

Modelling and data framework development cycle

To create useful methodology to study the mechanisms underlying poorly known biological systems, such as environmental transmission of infectious diseases, modelling frameworks should be developed together with data frameworks. The development is a cycle of constructing mathematical models and designing new experiments or sampling protocols that can be used to calibrate and validate the models. It usually starts with formulation of the simplest model that can possibly explain the observations using the hypothesised mechanisms. Next, the model is challenged with data to check if it can be calibrated and validated. If calibration fails, e.g. no optimal parameter values are found or the confidence bound are infinite (or too broad to make any biological conclusions) the decision has to be made whether the model needs to be adjusted

or if new data needs to be gathered. If validation fails and there is enough confidence to reject the first model, a subsequent model has to be formulated and tested with the data. The main goal of the cycle is to obtain a mechanistic, (yet as simple as possible) model that can usefully depict the reality, together with data gathering protocols of how to obtain sufficient data quality and quantity to use the model for various systems, e.g. to assess the efficacy of intervention strategies.

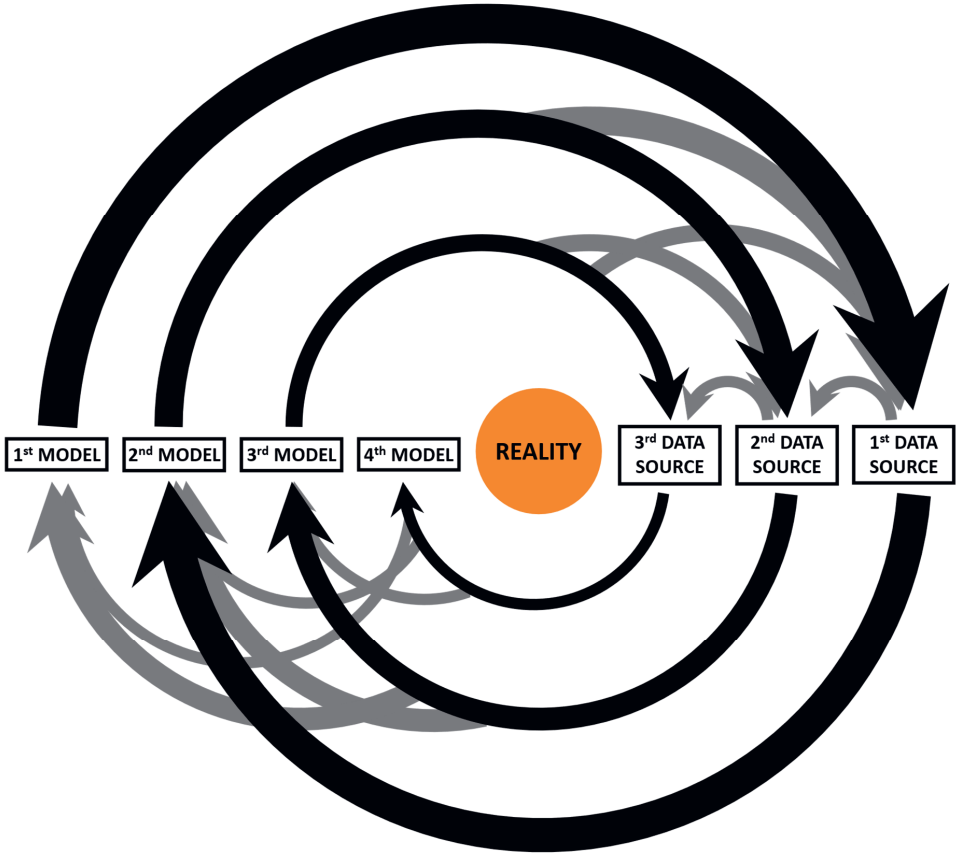


Figure 5.1. Schematic representation of the modelling and data-framework development cycle.

In this thesis, I applied the development cycle to tailor the modelling and data framework to study mechanisms underlying environmental transmission. The spatial model describing environmental transmission with decay and diffusion of infectious material was previously presented in (van Bunnik et al., 2014); compared to the model presented in the thesis, the previous version of the model had one more parameter (carrying capacity). In Chapter 2, we show that with newly gathered transmission data of improved spatial resolution a simpler, three-parameter model version can satisfactorily explain the experimental observations. Therefore, after confronting the model with data, we decided to accept the simpler version of the model to follow the principle of parsimony.

To calibrate the previous model, the available data was from *C. jejuni* transmission experiments where recipient broilers were separated from the infectious ones by a fixed distance band (van Bunnik et al., 2014). The model was not identifiable with these data (see Chapter 2), which in Chapter 3 we confirmed to be due to insufficient spatial resolution. So in (van Bunnik et al., 2014) the additional data from a separate survival experiment was used to estimate the decay rate parameter. As we show in Chapter 2, the estimate from the survival experiment was not representative of the survival of infectious form(s) of Campylobacter contributing to the transmission. We designed and conducted a (new) series of experiments in which the spatial resolution was enhanced – various distance bands were added, including experimental groups designed to study transmission across extremely short distances (infectious and recipient hosts housed together). In Chapter 3, we demonstrate that data from these new experiments was crucial to obtain informative parameter estimates. In Chapter 2, we show that using all (previous and new) experimental data, all three parameters can be estimated using transmission data only. Additionally, we show that the estimated decay rate gives a model fit that is much better than the fit generated using the (significantly different) value for the decay rate estimated from the survival experiment. In summary, we gathered data from new sources – transmission experiments that were tailored to the model in mind and thanks to that we were able to successfully calibrate and validate the model aimed to study mechanisms of the environmental transmission from an infectious material perspective. Consequently, we provided insight into processes underlying environmental transmission of *C. jejuni* in broilers (Chapter 2), sampling strategy protocols for future studies of environmental transmission for other systems (Chapter 3) and provided validated tools to start the study of density dependence on transmission looking from an infectious material perspective (Chapter 4).

The lattice organised spatial models, used to study the density dependence of transmission in Chapter 4, represent the first step of the new cycle of creating modelling and data framework specifically designed to study density of transmission of environmentally transmitted infectious diseases. The next (future) step would be to gather data tailored to validate the presented models and decide if they can be rejected or developed further.

Methodological challenges

Biological models, when developed, need to be supported with data to check their validity and form biologically relevant conclusions. Many mechanistic models are developed to describe transmission of infectious diseases, unfortunately often these models are not connected to data that is or may be acquired nor describe what kind of data can be obtained in the future to validate their findings. In a recent review, it has been shown that for environmentally transmitted zoonotic diseases, such as *Campylobacter* spp., the calibration and validation of the models, while highly recommended, is often not performed (Rees et al., 2021). At the same time, due to the fast development of accessible testing methods more and more data is collected on population as well

as individual level from both humans and animals (Barrat et al., 2014; Jønsson et al., 2016; Wirth et al., 2020). As was exposed during the recent SARS-CoV2 pandemic, connecting various data sources with mathematical modelling comes with several methodological challenges (Kretzschmar et al., 2022).

Complexity challenge

One of the challenges encountered when challenging models with data is managing the balance between model complexity and its identifiability with data that is or can be obtained to glean useful unbiased outcome (Kretzschmar et al., 2022). In this thesis I demonstrate that a strategy in which development of the modelling framework and data acquisition procedures, such as the design of suitable experiments or sampling protocols, are pursued in parallel is a promising strategy to ensure both model parsimony and utility.

Often mechanistic models are describing in detail certain mechanisms that have to be known or hypothesised *a priori*. Our infectious material-based modelling perspective provides a mechanistic, yet parsimonious framework, where general processes of environmental transmission are modelled mechanistically by simple stochastic processes. The mean field approximation of these stochastic processes is used in an environmental load density function that describes the spatiotemporal distribution of infectious material. As such, the framework can be used to study host-pathogen-environment systems for which detailed mechanisms of transmission are not understood, such as *C. jejuni* transmission in broilers, which we used as a model system in our study. Moreover, as in the framework, the underlying stochastic processes used to derive the mean field approximation on the general processes are well defined (see Chapter 1), the model can be adapted to incorporate more specific mechanisms if available data show the need to do so and can be used to estimate the effect.

In our spatiotemporal framework, the infectivity pattern is defined in terms of area(s) that infectious hosts occupy, time periods of the shedding (during which the shedding is assumed to be continuous) and a shedding rate parameter. While we used rectangular shape areas, in fact any source area shape can be incorporated as long as integration over this region is computable. Moreover, here we assumed that an infectious host homogenously visits the whole area. From that, easily a grid model can be created, where the host is able to move from area to area. For example, the grid models presented in Chapter 4 can be used to model different types of movement behaviour of and mixing between hosts. Heterogeneity can also be added as a layer describing the probability for individuals of occupying certain part of the source area. Such heterogeneity is often encountered in biological systems, for example flock animals tend to cluster with each other and many hosts tend to occupy more often certain areas around important parts of the environment they occupy, e.g. near water and food sources (Collins et al., 2011; Febrer et al., 2006).

The framework assumes that during a shedding period, described as part of an infectivity pattern, the shedding of infectious material is constant. Intermitted shedding, reported for pathogens such as *Salmonella* spp (van Immerseel et al., 2004), can be implemented by turning the infectivity pattern of a given host (with given movement area) into a sequence of continuous shedding periods. In our framework, the shedding is described by a shedding rate parameter. As the shedding rate is not identifiable separately from the exposure rate parameter, in the thesis these two parameters are described jointly with the transmission rate parameter. To account for the influence the host density has on the infectious patterns, we used a source strength configuration parameter to scale the source areas with density of hosts assuming that the shedding rate scales linearly with number of hosts occupying a particular area. It is not clear how the source strength would scale for most biological systems. In (van Bunnik et al., 2014) the source with 20 hosts was compared to 5 hosts; when supplemented with more data of suitable spatial distribution this can potentially be used to investigate the source strength for *C. jejuni* in broilers.

In the modelling framework, the survival process is approximated by constant independent decay i.e. an exponential distribution. As it has been shown that within a pathogen population two (or more) subpopulations can exist that differ in survival rate, for example bacteria can be infectious in culturable as well as viable but non-culturable form (Ramamurthy et al., 2014), the framework can further be developed to incorporate biphasic decay. Apart from heterogeneities in the pathogen population, the environmental conditions that influence the survival of pathogens, such as temperature or humidity, may vary both in time and space, in which case the decay process could be modelled with a time and space variable decay rate.

In our implementation, dispersion of infectious material was modelled assuming that the infectious material diffuses through the environment. Diffusion is based on the simple random walk mechanism, where particles are moving in random directions and step size is a Gaussian distributed random variable (Chandrasekhar, 1943). The elegance of using the random walk as underlying mechanism is the existence of a simple mean field solution- the diffusion equation, that in a one-particle regime describes the probability of finding a particle after walking for a particular time starting from the source and in the many particle regime describes the distribution of the particle load (Chandrasekhar, 1943). In this thesis, I did not go beyond this mean field approximation. Potentially, anomalous diffusion, e.g. based on Levy flight mechanisms (Klafter et al., 1996), could be used for example to study systems where material spreads heterogeneously. Such a situation can occur when multiple biological mechanisms, such as dispersion via mechanical vectors and/or via air, are contributing to the dispersion of infectious material.

In the infectious material-based framework, also the exposure process is implemented as a simple mechanism and is characterised by an exposure rate parameter that is constant in time and space. The exposure depends on host behaviour and biology as well as on properties of the specific

environment in mind (which in most cases varies in time and space). Therefore, the exposure can be seen as a system specific feature. For example, the exposure will look differently for broilers and laying hens, the former moving less and eating more than the latter. Generally, caution is needed, as the exposure process in general is not well understood. For many systems, the exact route or multiple routes of exposure that determine the way the pathogen is absorbed by the hosts are not known. The exposure rate parameter is not separately identifiable from the shedding rate (and hence these are described jointly with the transmission rate parameter) when using transmission data only. When considering host densities in Chapter 4, we scaled the exposure with the exposure strength configuration parameter to scale the exposure areas with density of hosts if the exposure rate scales linearly with the area that is available for host. It is not clear how the exposure strength would scale for most biological systems, and most likely it is a system specific feature. In all our experiments on *C. jejuni* transmission, recipient broilers occupied areas of the same (or similar) size. More studies are needed to explore how the available area influences the exposure to infectious material.

For biological systems the level of model complexity should not only be determined by the biology of the processes described mechanistically in the model but also by it's potential to be calibrated (and subsequently validated) with data that is or can be obtained. Such potential is assessed during identifiability analysis. The identifiability analysis for the infectious material-based framework and host status data is presented in Chapter 3.

Identifiability analysis, while well developed for some dynamic models based on differential equations such as systems biology models (Wieland et al., 2021), for the infectious material based spatial model was never done in a systematic manner. Full identifiability analysis consists of structural identifiability analysis, studying if parameters (or combination of thereof) can be estimated for an infinite sample size, and practical identifiability analysis, studying if parameters can be estimated when data are of limited quality or quantity (Wieland et al., 2021). In Chapter 3, we address both structural and practical identifiability of our infectious material based spatiotemporal model using both simulated as well as experimental host status data. In particular, we showed the importance of the spatial resolution on the identifiability of parameters describing the decay and dispersion of the infectious material.

As we show in Chapter 2 and Chapter 3 and as was reported for compartmental models of environmental transmission (Brouwer, Weir, et al., 2017), shedding and exposure are structurally not jointly identifiable when using host data only. This means that even if an infinite number of hosts is sampled with fine-grained resolution, the shedding rate parameter and exposure rate parameter cannot be estimated separately. Therefore, in our infectious material-based modelling framework a joint parameter – transmission rate parameter (being the product of shedding rate parameter and exposure rate parameter) was defined. When some parameters of the model are structurally unidentifiable, the only way to estimate those parameters separately is to use data

that provides additional information. In a transmission study, this means collecting data on measurements different than host status. Such data can come either from the same study, but from different samples or from a separate experiment or field study. In both cases, additional uncertainty is added as every (type of) measurement or experiment comes with its own limitations such as measurement errors, detection limits or biological model assumptions. Sometimes such limitations can significantly influence the model calibration and result in uninformative model estimates, which consequently leads to wrong biological conclusions. Examples can be found in Chapter 2 where we compared the decay rate parameter estimates obtained for two different data sources: 1) host status data of sufficient resolution, and 2) data on survival of culturable forms of *C. jejuni* obtained from environmental samples collected in the same experimental conditions as the transmission data. These estimates were significantly different, and as we have shown the model fit is much better when the decay rate parameter is estimated from transmission data (together with remaining two parameters) than when estimate from the additional survival experiment is used.

To solve the lack of the structural identifiability of the shedding rate parameter with the other parameters of the model, the amount of infectious material per time unit produced by infectious hosts, in excreted material e.g. the faeces (see *Salmonella* model in (Collineau et al., 2020)) can potentially be used. However, for some systems assumption that all pathogen shed is in a form available and infectious for other hosts may significantly influence the model outcome. When little is known about the form of infectious material that the recipient hosts are exposed to, using shedding rate to approximate source strength should be done with caution.

In Chapter 3, we have shown that, in order to obtain informative estimates, the host status data have to be collected in sufficient spatial and temporal resolution. In practice it means that (at least some) measurements have to be taken at time points when the environmental load is not in equilibrium and information about environmental load in areas distanced from the source by various distances have to be collected. When such data is not available, the theory of fluctuations could be considered as a solution to these identifiability problems.

The theory describes how the auto correlation time of fluctuations in the particle load observed in diffusion equilibrium depends on the diffusion coefficient (Chandrasekhar, 1943; Smoluchowski, 1916), and the diffusion coefficient can be estimated from counts of particles observed within one area in time. As we show below, for a model with added decay when only viable (infectious) particles can be measured (e.g. by being detected by hosts that can be infected by them) just one area is not enough; as it is not known if the short auto-correlation time of fluctuations is the result of the particles moving out of the area or dying within, it cannot be determined how fast the particles are spreading. In the future, one can explore if by observing the fluctuations in two neighbouring areas and their corelation with each other, it could be determined how the diffusion and decay influence the fluctuations.

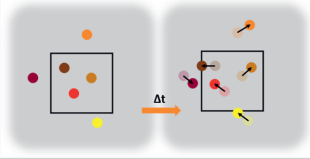
Thoughts on the theory of fluctuations in decaying particles system

SMOLUCHOWSKI'S THEORY OF FLUCTUATIONS

System (**volume or area V**) under diffusion equilibrium is observed periodically at **constant intervals τ**.

Initial assumptions:

I. the motions of individual particles are not mutually influenced and are independent of each other;
II. all positions in the element of volume considered have equal a priori probability.

'SLOW' SYTEM	'FAST' SYSTEM
	
Diffusion coefficient	
Small	Big
Net displacement	
Small	Big
Fluctuations	
Slow	Fast
Observations	
Highly corelated	Slightly corelated

THE SIZE OF FLUCTUATIONS

The frequency W(n) with which different numbers of particles n will be observed in V follow the Poison distribution, with ν being the average number of particles that will be contained in V:

$$P(W=n)=\frac{e^{-\nu}\nu^{n}}{n!}$$

THE SPEED OF FLUCTUATIONS

Probability that in two (or more) consecutive observations we will have outcomes that are somehow dependent of each other.

$A_i^{(n)}$ is a probability that i out of n particles will have emerged from the cage with area V, after time Δt: $A_i^{(n)} = \frac{n!}{i!(n-i)!}P^i(1-P)^{n-i}$ which is a Bernoulli distribution.
E_i is the probability that i particles will have entered the element of volume V during τ.
As we consider situation in equilibrium: $E_i = \langle A_i^{(n)}\rangle_{Av} = \sum_{n=i}^{\infty} W(n)A_i^{(n)}$ which is a Poisson distribution with variance νP. From $A_i^{(n)}$ and E_i transition probabilities can be derived.

PROBABILITY AFTER-EFFECT

"Probability after-effect" P is a probability that a particle somewhere inside area V will have emerged from it during the time τ. The value of P depends on the "precise circumstances of the problem". In case of diffusion (random walk) it will depend on the diffusion coefficient and on the geometry of the V.

P also can be estimated from the experiment, if we have the series of observations we can calculate P with formula: $\langle\Delta^2\rangle_{Av} = \nu P$, where $\langle\Delta^2\rangle_{Av}$ is the mean square of the differences in the numbers of particles observed on consecutive occasions in a long sequence of observations made at constant intervals τ apart and **ν is average number of all the counts observed**

$$1 - P = \frac{1}{4\pi D\tau}\int_{x_1}\int_{x_2}\int_{y_1}\int_{y_2}\exp\left(-\frac{(x_1-x_2)^2+(y_1-y_2)^2}{4D\tau}dy_2dy_1dx_2dx_1\right)$$

DECAYING PARTICLES

For system with decaying particles, P needs to be corrected. If we assume that the system with particles that are emitted and particles that are dying is in equilibrium, we assume that **on average the number of particles that enter the area during the period τ equals the number of particles that leaves OR dies during the period τ.**
Therefore, P_α is a probability that a particle somewhere inside V will have emerged from it OR will have died during the time τ:

$$1 - P_\alpha = \frac{1}{4\pi D\tau}\int_{x_1}\int_{x_2}\int_{y_1}\int_{y_2}\exp\left(-\alpha*\tau-\frac{(x_1-x_2)^2+(y_1-y_2)^2}{4D\tau}dy_2dy_1dx_2dx_1\right)$$

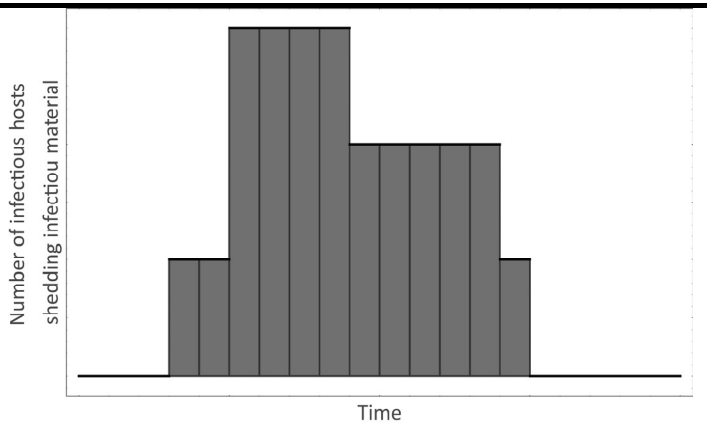
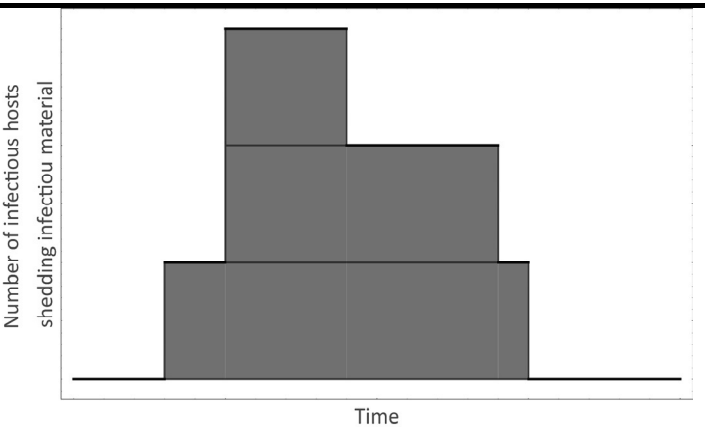
As in transmission experiment we 'detect' only life particles that are present in the area of interest, D and α are unidentifiable when there is no spatial resolution (observations of neighbouring/close - by areas).

Computability challenge

Another challenge when developing a modelling and data framework is ensuring that computation tools are available for the model to be calibrated and validated with data that is gathered. As presented in Chapter 3, we developed a complete methodological framework for calibration of the spatiotemporal infectious material-based model including an input system, automatic log likelihood function generation, log likelihood maximization algorithm and identifiability analysis tools. Our parsimonious model, while conceptually simple, when calibrated with individual level data on host status in time and position, leads to construction of a rather complicated likelihood function containing multiple space and time integrals (see Fig. 1.4 in Chapter 1). Numerical maximization of such function is time consuming and/or prone to errors (e.g. returning a local maximum). To make the likelihoods more tractable, we optimized model implementation (Chapter 2) as well as maximization algorithm (Chapter 3). In the constraint maximisation algorithm, the univariate profile likelihoods are generated for all parameters in a multistep procedure, where the information from the profiles obtained in previous steps is used in current step of maximization to set constraint on remining (running) parameter values. This improves the maximisation performance (reducing the probability of returning local maxima instead of global maxima) and speed (as narrower parameter range is searched during numerical maximisation).

While we have studied mostly experimental data where newly infected host were removed such that a chain of transmission was avoided, our framework can also be used to estimate the parameters when source patterns are dynamic i.e. infectious areas are emerging and disappearing. An example of such use is provided in Chapter 2; to analyse an experiment where newly infected broilers were not removed (and therefore were considered new sources) we developed tools to create infectious patterns for multiple sources. Usually, the data showing the number of infectious individuals in time are decomposed vertically (into discrete time intervals) to construct the input data for parameter estimation. In techniques such as the GLM based estimation, the model is fit to the time series of data where the total environmental load generated by all infectious were present in the environment at the moment and in the past is updated every time interval e.g. a day by using recurrence formula (Chang & de Jong, 2023). In our methodological implementation, we decomposed data using a technique we called “horizontal decomposition”. Rather than creating a time series, a set of the infectivity patterns can be composed, in which every infectious pattern describes the area and period where and when the number of infectious hosts was constant. The details on both approaches are presented in table 5.1.

Table 5.1. Comparison of the decomposition methods used to account for the total contribution of the past and present sources to the present environmental load.

Vertical decomposition	Horizontal decomposition
	
Step-wise integration (over each unite period e.g. day) over shedding and exposure	Integration over the whole period of shedding (T_1^i, T_2^i) for each group of infectious, source hosts described with joint infectivity pattern $\left(\mathrm{t, x, y, \{}T_1^i, T_2^i, A_{inf}^i\}\right)$ over the whole exposure period for each recipient exposed.
Implementation: recurrence formula	Implementation: total environmental load density function being a sum over all distinctive infectivity patterns: $\sum_I W^i\left(\mathrm{t, x, y, \{}T_1^i, T_2^i, A_{inf}^i\}\right)$

Sampling and detection challenge

When considering what data should be used to calibrate and validate environmental transmission models (e.g. aimed to study the decay and dispersion of infectious material), the environmental monitoring indicating the amount of infectious material accumulated in the environment in time and space is often considered. Such monitoring can be classified in two types of methods: measuring direct or indirect indicators. Among direct indicators are microbiological sampling and detection methods, often used to determine the contamination of farms, hospital wards or food establishments. While helpful to assess the general health risk, the detection of pathogens or lack of thereof does not necessarily inform about the risk of transmission. On the one hand, pathogens detected by these methods can be of a form that is not infectious to susceptible hosts. For example, PCR detected genetic material of pathogens may not come from live/infectious forms or it may be detected in material that is not infectious to the hosts (e.g. pathogens transmitted via air detected in food). For example, PCR SARS-CoV-2 tests are extremally helpful to detect (and isolate) new cases but are not a good indicator of infectiousness due to high sensitivity to remnant viral RNA (Johnston et al., 2022; Singanayagam et al., 2020). On the other hand, even if all samples are found negative, infectious material can still be present in the environment, but just not being detected because of inadequate sampling or detection limits. In one of our experiments, where recipient broilers entered experimental room 20 days after the infectious broilers started contaminating the room (see Chapter 2 for details), we conducted small pilot study, where the litter samples from non-occupied pens were sampled just before recipients were placed in the adjacent pen and checked for *C. jejuni*; despite all samples returning negative, few days later the transmission of *C. jejuni* was detected in the same room.

Indirect indicators for pathogen presence in the environment can also be of two kinds. For specific systems, it is possible to measure some characteristics of the environment; for example water turbidity was used to assess its contamination which is correlated with presence of *Cryptosporidium* (Brouwer, Weir, et al., 2017). However, as such indicators are only indirectly connected to the infectious material, their limitations must be addressed really carefully on case-by-case basis. Moreover, for most systems finding such environment indicators is simply not possible. The only generally applicable indirect detection method for infectious material is monitoring the infection of sentinel hosts placed in the environment such as we present in Chapter 2 and Chapter 3. The observation of a (newly) infected host shows that obviously the infectious material was present in the environment when transmission occurred. This latter type of indirect pathogen detection obviously measures only the infectious forms of pathogens and the portion of infectious material that is (naturally) accessible to the hosts.

Methodological perspectives

While Chapters 2 and Chapter 3 focused on validation and calibration of the model, in Chapter 4 we used our infectious material based spatiotemporal model to study the density dependence of transmission. When interpreting transmission from contact-based perspective, the density dependence is defined as the influence of the density on the host-to-host contact rate, defining contacts as potentially infectious contacts that are relevant for transmission. As often those contacts are not well defined, the influence that density may have on the contact rate cannot be modelled satisfactorily.

While some studies estimated the density dependence of R0 from density and prevalence data (Sy et al., 2021), a mechanistic basis for the density dependence was never proposed. From an infectious material perspective, for those infections that are transmitted via environment, such a basis can be developed using our spatiotemporal modelling framework. Homogenous mixing is one of the main assumptions of the non-spatial, compartmental models of transmission. From this assumption, it follows for environmental transmission, that hosts are exposed to infectious material (have contact with environmental compartment) while they are accessing freely the whole environment with equal probability. While some individual based models were proposed to incorporate heterogeneity in visiting the environment, they were constructed for highly specific systems such as a cow barn (Chen et al., 2013). A systematic approach to study density dependence is lacking. In Chapter 4, we attempt to start such a study by constricting a few infectious material-based scenarios with either homogenously mixing or spatially constraint hosts. The models can be used to calculate the density dependence of R0 for various scenarios and indicate how the transmission is reduced when various intervention strategies aimed at constraining hosts movements are considered. As a next step, outcomes of such models should be validated with data to guide following development of the modelling framework.

Insights for Campylobacter transmission in broilers

Being a widespread zoonotic bacterium, *Campylobacter* spp. is one of the main concerns in public health. Despite many efforts, *Campylobacter* is still highly prevalent in broiler flocks, and broilers are considered to be the main source of Campylobacteriosis in human (Mota-Gutierrez et al., 2022; Wagenaar et al., 2013). Despite its ubiquity, there are still important knowledge gaps regarding *Campylobacter* transmission in broilers that hamper the identification of effective intervention strategies (Hansson et al., 2018). This thesis provided insight into a few of them, namely: unknown transmission route, causes of the Campylobacter lag phase observed in the field, and survival of *Campylobacter* in the environment.

Campylobacter spreads in a poultry flock through a contaminated environment. It is well documented that it can be detected in high amounts in faeces of colonised chickens and other animals (Stern et al., 2001; Stern & Robach, 2003) and that chicken become colonised after oral

inoculation (Line et al., 2008). It was shown that chicken can get colonised not only when in contact with faeces, but also in cleaned and disinfected broiler houses (Agunos et al., 2014) or in a clean pen separated from *Campylobacter* positive broilers (van Bunnik, 2014; van Bunnik et al., 2012; van Bunnik et al., 2014). Many hypotheses were raised to explain this route of transmission, e.g. contaminated water, material carried on by workers, visitors or other animals, yet no consensus was reached (Hansson et al., 2018; Sahin et al., 2002). Most likely, material contaminated with *Campylobacter* can be spread via many routes (Frosth et al., 2020). Because this host-pathogen-environment system is not well studied its modelling is challenging. The parsimonious model I present in this thesis and that was first presented by (van Bunnik et al., 2014), does not specify the exact route of *Campylobacter* spatial spreading. Rather, it uses the diffusion process as approximation of all possible routes that might be present. Therefore, the information about transmission can be obtained without modelling all possible routes of transmission. The spatial spread is incorporated in a diffusion equation that is a mean field representation of all the possible transition routes. In experiments presented in Chapter 2 and in (van Bunnik, 2014; van Bunnik et al., 2012; van Bunnik et al., 2014) we show that *C. jejuni* can be transmitted even when broilers do not share environment (nor feeders or water source) with infectious birds. This indicates that contact with contaminated faeces or water source is not the only route of transmission. The relatively low diffusion coefficient estimated for our experimental data indicates that a purely airborne route of *C. jejuni* transmission between separated broilers is unlikely. If an infectious particle were moved from source to recipient area with air flow, the diffusion rate would be much higher, and no distance dependent delay of transmission would be observed. This distance dependence of the delay observed in our experiments, would also not be observed if a fast mechanical vector such as fly or animal caretaker was the main source of transmission. The spread of infectious material most likely happens in slow, multistep process.

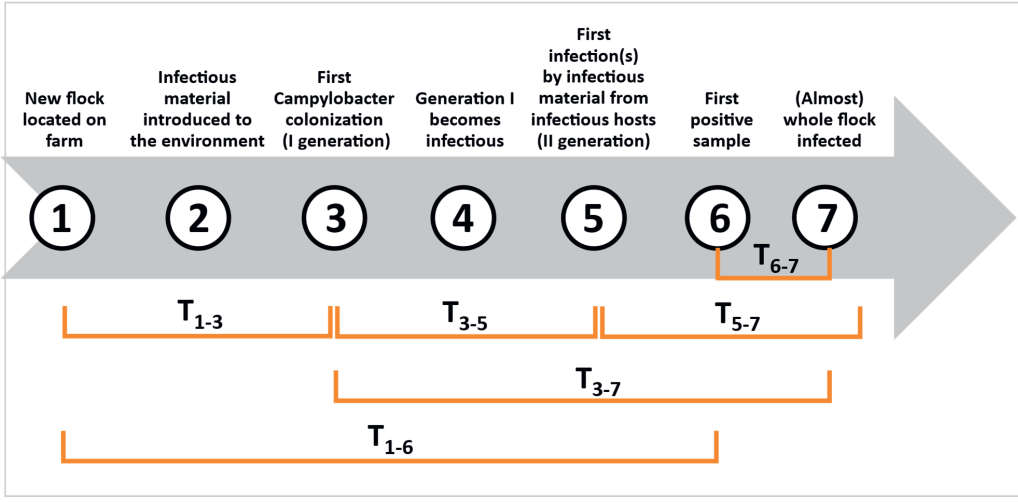
Another poorly understood aspect of the transmission is the survival of *Campylobacter*. While there were experiments studying survival of *C. jejuni* in faeces, used litter or water (Ahmed et al., 2013; Bronowski et al., 2014; Smith et al., 2016; van Bunnik et al., 2014), in a majority of these only the culturable forms of *C. jejuni* were counted. As we show in Chapter 2, the survival characteristics of culturable forms of *C. jejuni* are not necessarily representative for those of all the transmissible forms of *C. jejuni*. Moreover, the estimate of the decay rate obtained in our model, where infectious forms were considered by including only host infection status data, was much lower than for a separate survival experiment studying only culturable forms. This shows that some infectious forms of *C. jejuni* survive better in the environment than culturable forms measured in survival experiment. The reasons of this disparity should be explored further. It was suggested that viable but non culturable forms of bacteria such as *E.coli* may contribute to the transmission (Ding et al., 2017; Li et al., 2014). The transmission potential of viable but non culturable forms of *Campylobacter*

is still not well studied and further research is needed to determine if these forms contribute to transmission (Kassem et al., 2017; Moore, 2001).

In Chapter 2, where we presented all the experimental data, we demonstrated that, as was previously noted by (van Bunnik et al., 2014), when *C. jejuni* transmission between separated broilers is studied experimentally, there is a notable delay between beginning of shedding of the infectious material and first detected positive sample among recipients. As it is shown in Chapter 2, said delay is correlated with distance between source (infectious) hosts and recipient ones, the bigger the distance the delay is longer, being extremally short when infectious and recipient hists are housed together.

Another interesting phenomenon reported for *Campylobacter* is known as the “lag phase”. The lag phase is described in the literature as the period between the placement of a new flock on the broiler farm and the first detected colonisation with *Campylobacter* (Newell, 2002). Therefore, it is a different concept than delay observed experimentally. To explore the possible causes of such a lag phase, it is necessary to consider all the processes that are happening in the beginning of the broiler production cycle (see the box below for details).

Thoughts on Campylobacter lag phase 1/2



EXPERIMENTAL DATA ON DELAY TIMES IN CAMPYLOBACTER TRANSMISSION IN BROILERS

Delay time	Definition	Estimated value [days]	Method	Reference
T_{1-3}	Time between start of production and first infection within a flock	> 21	Epidemiological modelling based on field data	(van Gerwe et al. 2009)
T_{1-6}	Time between first day of production cycle and first positive sample in flock under field conditions	2 flocks: 32, 35	Experiment on farm, broilers sampled weekly	(Shreeve et al. 2000)
		App. 21-28 (earliest 13- latest app. 50)	Field data, sampled weekly	(Jacobs-Reitsma et al. 1995)
		5 farms: 39, 32, 36, 36, 46	Field data, environmental samples from broiler houses, sampled twice/week	(Ring, Zychowska, and Stephan 2005)
T_{3-5}	Time between first infection in generation I and first infection in generation II	<3 (+2)[1]	Transmission experiment with 1 seeder bird and 20 recipients. "+2" means that seeders were inoculated 2 days prior introduction.	(Stern et al. 2001)
		4 experiments: 2.4, 2.7, 5.8. 7.1	Epidemiological modelling based on transmission experiments	(van Gerwe et al. 2005)
T_{3-7}	Time between first infection in generation I and more than 95% of flock infected	<7 (+2) [1]	Transmission experiment with 1 seeder bird and 20 recipients. "+2" means that seeders were inoculated 2 days prior introduction.	(Stern et al. 2001)
T_{5-7}	Time between first infection in generation II and more than 95% of flock infected	4.4-7.2 (for 95% of 20,000 birds flock)	Epidemiological modelling based on field data	(van Gerwe et al. 2009)
		6 (from 5% to 95% of 20,000 birds flock)	Epidemiological modelling based on transmission experiments	(van Gerwe et al. 2005)
T_{6-7}	Time between first positive sample and more than 95% of flock infected	<7 days	Field data, sampled weekly	(Jacobs-Reitsma et al. 1995)
		Two flocks: <7, <7	Experiment on farm, sampled weekly	(Shreeve et al. 2000)

1) seeders were inoculated 2 days prior introduction.

Thoughts on Campylobacter lag phase 2/2

CONSISTENT FINDINGS[1]

- T_{1-6} is most likely 3-5 weeks, no shorter than 2 weeks
- T_{4-7} is less than 1 week, so: 1 week $>T_{4-7}>T_{4-6}>T_{4-5}$
- T_{6-7} is also less than 1 week

SO

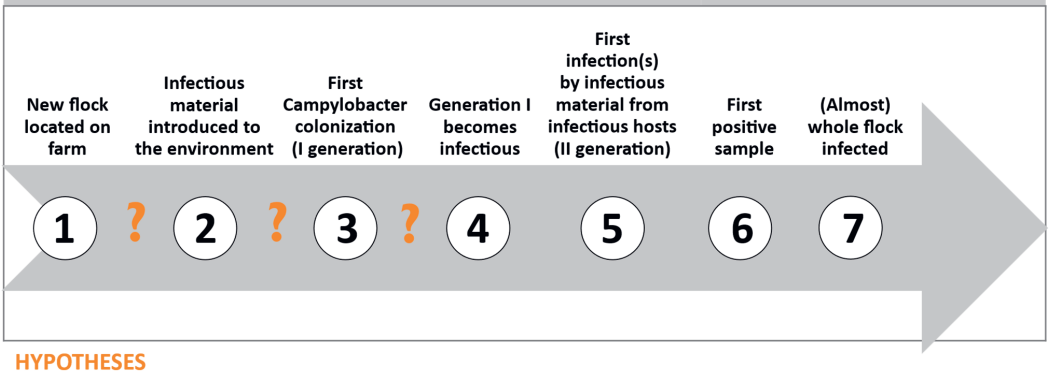
T_{1-4} at least 1 week and at most 4 weeks of unexplained delay

1) determined in multiple experiments

ENVIRONMENT

KNOWLEDGE GAPS

What happens between 1 and 2?	→ Routes of introduction
What happens between 2 and 3?	→ Dose-response
What is the delay time between 3 and 4?	→ Incubation period



HYPOTHESES

I. Delay occurs between 1 and 2 because of environmental load, biosecurity etc, and cumulative probability of introduction is increasing with time

II. Delay occurs between 2 and3 because of dose-response to low-dosages of pathogen (from environmental transmission), and older birds need lower dosage to become infected

III. Delay occurs between 3 and 4 because of some unknown process within a chicken's body; the incubation period is longer for low doses

Note: Hypotheses 2 and 3 require knowledge on how low doses of pathogen influence the hosts (at different ages). When there is a high dose, process 2 -4 is similar to 4-5 so then the delay from 1 to 5 will be strictly environment related.

FUTURE STUDIES

- How does biosecurity level influence delay time? Literature data are inconsistent, more detailed study needed.
- Dose-response for low dosages and age dependence; dose response to low doses experiment is needed.
- If/how the incubation period dependent on dose and age; sxperimental data needed.
- Studies on continuous low-dosage exposure; e.g. experiments on exposure to low doses (e.g. in the environment) needed.

Conclusions

Overall, in this thesis, we use a cycle of modelling and experimental design development to study the mechanisms of environmental transmission from the infectious material perspective. We have developed tools that assist both modelling framework development, including calibration, full identifiability analysis and validation, and design of (future) experiments and/or data sampling protocols. Additionally, we demonstrated that our spatial approach can be used to study density dependence of transmission. The methodology can be used in future studies to compare various systems and quantify effect of intervention strategies. Also, we have provided new insights on *C. jejuni* transmission in broilers. Specifically, we have demonstrated that infectious forms of *C. jejuni* most likely survive longer in the environment than forms culturable in laboratory conditions, and that dispersion of infectious material carrying *C. jejuni* is a slow, most likely multistep process.

References

Agunos, A., Waddell, L., Léger, D., & Taboada, E. (2014). A Systematic Review Characterizing On-Farm Sources of Campylobacter spp. for Broiler Chickens. *PLOS ONE*, *9*(8), e104905. https://doi.org/10.1371/journal.pone.0104905

Ahmed, M. F. M., Schulz, J., & Hartung, J. (2013). Survival of Campylobacter jejuni in naturally and artificially contaminated laying hen feces. *Poultry Science*, *92*(2), 364-369. https://doi.org/10.3382/ps.2012-02496

Andino, A., & Hanning, I. (2015). Salmonella enterica: Survival, Colonization, and Virulence Differences among Serovars. *The Scientific World Journal*, *2015*, 520179. https://doi.org/10.1155/2015/520179

Asadi, S., Gaaloul ben Hnia, N., Barre, R. S., Wexler, A. S., Ristenpart, W. D., & Bouvier, N. M. (2020). Influenza A virus is transmissible via aerosolized fomites. *Nature Communications*, *11*(1), 4062. https://doi.org/10.1038/s41467-020-17888-w

Bansal, S., Read, J., Pourbohloul, B., & Meyers, L. A. (2010). The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, *4*(5), 478-489. https://doi.org/10.1080/17513758.2010.503376

Barrat, A., Cattuto, C., Tozzi, A. E., Vanhems, P., & Voirin, N. (2014). Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, *20*(1), 10-16. https://doi.org/10.1111/1469-0691.12472

Battersby, T., Whyte, P., & Bolton, D. J. (2016). The pattern of Campylobacter contamination on broiler farms; external and internal sources. *J Appl Microbiol*, *120*(4), 1108-1118. https://doi.org/10.1111/jam.13066

Begon, M., Bennett, M., Bowers, R. G., French, N. P., Hazel, S. M., & Turner, J. (2002). A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology and Infection*, *129*(1), 147-153. https://doi.org/10.1017/S0950268802007148

Benson, L., Davidson, R. S., Green, D. M., Hoyle, A., Hutchings, M. R., & Marion, G. (2021). When and why direct transmission models can be used for environmentally persistent pathogens. *PLOS Computational Biology*, *17*(12), e1009652. https://doi.org/10.1371/journal.pcbi.1009652

Bidegain, G., Powell, E. N., Klinck, J. M., Hofmann, E. E., Ben-Horin, T., Bushek, D., Ford, S. E., Munroe, D. M., & Guo, X. (2017). Modeling the transmission of Perkinsus marinus in the Eastern oyster Crassostrea virginica. *Fisheries Research*, *186*, 82-93. https://doi.org/10.1016/j.fishres.2016.08.006

Boender, G. J., Meester, R., Gies, E., & De Jong, M. C. (2007). The local threshold for geographical spread of infectious diseases between farms. *Prev Vet Med*, *82*(1-2), 90-101. https://doi.org/10.1016/j.prevetmed.2007.05.016

Bouma, A., de Jong, M. C. M., & Kimman, T. G. (1995). Transmission of pseudorabies virus within pig populations is independent of the size of the population. *Preventive Veterinary Medicine*, *23*(3), 163-172. https://doi.org/10.1016/0167-5877(94)00442-L

Breban, R. (2013). Role of environmental persistence in pathogen transmission: a mathematical modeling approach. *Journal of Mathematical Biology*, *66*(3), 535-546. https://doi.org/10.1007/s00285-012-0520-2

Bronowski, C., James, C. E., & Winstanley, C. (2014). Role of environmental survival in transmission of Campylobacter jejuni. *FEMS Microbiology Letters*, *356*(1), 8-19. https://doi.org/10.1111/1574-6968.12488

Brouwer, A. F., Eisenberg, M. C., Remais, J. V., Collender, P. A., Meza, R., & Eisenberg, J. N. (2017). Modeling Biphasic Environmental Decay of Pathogens and Implications for Risk Analysis. *Environ Sci Technol*, *51*(4), 2186-2196. https://doi.org/10.1021/acs.est.6b04030

Brouwer, A. F., Weir, M. H., Eisenberg, M. C., Meza, R., & Eisenberg, J. N. S. (2017). Dose-response relationships for environmentally mediated infectious disease transmission models. *PLOS Computational Biology*, *13*(4), e1005481. https://doi.org/10.1371/journal.pcbi.1005481

Burnham, K. P. (1998). Model selection and multimodel inference. *A practical information-theoretic approach*.

Cambridge Dictionary, (n.d.). *Contact*. Cambridge University Press. Retrieved 20.11.2022 from https://dictionary.cambridge.org/

Cappelier, J. M., Magras, C., Jouve, J. L., & Federighi, M. (1999). Recovery of viable but non-culturable Campylobacter jejuni cells in two animal models. *Food Microbiology*, *16*(4), 375-383. https://doi.org/10.1006/fmic.1998.0246

Chandrasekhar, S. (1943). Stochastic Problems in Physics and Astronomy. *Reviews of Modern Physics*, *15*(1), 1-89. https://doi.org/10.1103/RevModPhys.15.1

Chang, Y., & de Jong, M. C. M. (2023). A novel method to jointly estimate transmission rate and decay rate parameters in environmental transmission models. *Epidemics*, *42*, 100672. https://doi.org/10.1016/j.epidem.2023.100672

Chen, S., Sanderson, M. W., White, B. J., Amrine, D. E., & Lanzas, C. (2013). Temporal-spatial heterogeneity in animal-environment contact: Implications for the exposure and transmission of pathogens. *Scientific Reports*, *3*(1), 3112. https://doi.org/10.1038/srep03112

Colenutt, C., Brown, E., Nelson, N., Paton David, J., Eblé, P., Dekker, A., Gonzales José, L., & Gubbins, S. (2020). Quantifying the Transmission of Foot-and-Mouth Disease Virus in Cattle via a Contaminated Environment. *mBio*, *11*(4), e00381-00320. https://doi.org/10.1128/mBio.00381-20

Collineau, L., Phillips, C., Chapman, B., Agunos, A., Carson, C., Fazil, A., Reid-Smith, R. J., & Smith, B. A. (2020). A within-flock model of Salmonella Heidelberg transmission in broiler chickens. *Preventive Veterinary Medicine*, *174*, 104823. https://doi.org/10.1016/j.prevetmed.2019.104823

Collins, L. M., Asher, L., Pfeiffer, D. U., Browne, W. J., & Nicol, C. J. (2011). Clustering and synchrony in laying hens: The effect of environmental resources on social dynamics. *Applied Animal Behaviour Science*, *129*(1), 43-53. https://doi.org/10.1016/j.applanim.2010.10.007

Cortez, M. H., & Weitz, J. S. (2013). Distinguishing between Indirect and Direct Modes of Transmission Using Epidemiological Time Series. *The American Naturalist*, *181*(2), E43-E52. https://doi.org/10.1086/668826

Council Directive 2007/43/EC, (2007). *Consolidated text: Council Directive 2007/43/EC of 28 June 2007 laying down minimum rules for the protection of chickens kept for meat production (Text with EEA relevance)*.

Crane, b. S., & Moore, J. (1986). Modeling enteric bacterial die-off: a review. *Water, Air, and Soil Pollution*, *27*(3), 411-439.

Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., Ross, J. V., & Vernon, M. C. (2011). Networks and the Epidemiology of Infectious Disease. *Interdisciplinary Perspectives on Infectious Diseases*, *2011*, 284909. https://doi.org/10.1155/2011/284909

David, J. F., Iyaniwura, S. A., Ward, M. J., & Brauer, F. (2020). A novel approach to modelling the spatial spread of airborne diseases: an epidemic model with indirect transmission. *Mathematical Biosciences and Engineering*, *17*(4), 3294-3328.

De Jong, M., Diekmann, O., & Heesterbeek, J. (1994). How does infection-transmission depend on population size. *Epidemic Models, Their Structure and Relation to Data, Cambridge University Press, Cambridge*.

Diekmann, O., De Jong, M. C. M., & Metz, J. A. J. (1998). A Deterministic Epidemic Model Taking Account of Repeated Contacts between the Same Individuals. *Journal of Applied Probability*, *35*(2), 448-462. http://www.jstor.org/stable/3215698

Diekmann, O., Heesterbeek, H., & Britton, T. (2013). *Mathematical tools for understanding infectious disease dynamics* (Vol. 7). Princeton University Press.

Ding, T., Suo, Y., Xiang, Q., Zhao, X., Chen, S., Ye, X., & Liu, D. (2017). Significance of Viable but Nonculturable Escherichia coli: Induction, Detection, and Control. *J Microbiol Biotechnol*, *27*(3), 417-428. https://doi.org/10.4014/jmb.1609.09063

Drossinos, Y., & Stilianakis, N. I. (2020). What aerosol physics tells us about airborne pathogen transmission. *Aerosol Science and Technology*, *54*(6), 639-643. https://doi.org/10.1080/02786826.2020.1751055

Efsa Panel on Biological Hazards, Koutsoumanis, K., Allende, A., Alvarez-Ordóñez, A., Bolton, D., Bover-Cid, S., Davies, R., De Cesare, A., Herman, L., Hilbert, F., Lindqvist, R., Nauta, M., Peixe, L., Ru, G., Simmons, M., Skandamis, P., Suffredini, E., Alter, T., Crotta, M., . . . Chemaly, M. (2020). Update and review of control options for Campylobacter in broilers at primary production. *EFSA Journal*, *18*(4), e06090. https://doi.org/10.2903/j.efsa.2020.6090

Eisenberg, M. C., Robertson, S. L., & Tien, J. H. (2013). Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *Journal of Theoretical Biology*, *324*, 84-102. https://doi.org/10.1016/j.jtbi.2012.12.021

El Jarroudi, M., Karjoun, H., Kouadio, L., & El Jarroudi, M. (2020). Mathematical modelling of non-local spore dispersion of wind-borne pathogens causing fungal diseases. *Applied Mathematics and Computation*, *376*, 125107. https://doi.org/10.1016/j.amc.2020.125107

Febrer, K., Jones, T. A., Donnelly, C. A., & Dawkins, M. S. (2006). Forced to crowd or choosing to cluster? Spatial distribution indicates social attraction in broiler chickens. *Animal Behaviour*, *72*(6), 1291-1300. https://doi.org/10.1016/j.anbehav.2006.03.019

Frosth, S., Karlsson-Lindsjö, O., Niazi, A., Fernström, L. L., & Hansson, I. (2020). Identification of Transmission Routes of Campylobacter and On-Farm Measures to Reduce Campylobacter in Chicken. *Pathogens*, *9*(5). https://doi.org/10.3390/pathogens9050363

Gamża, A. M. (2021). Data and code for "Understanding environmental transmission mechanisms: a parsimonious mathematical model validated with infection data from tailor-made experiments". https://doi.org/10.5281/zenodo.5565054

Gamża, A. M. (2023a). Data and code for "Identifiability of environmental transmission parameters: quantifying dispersion and decay of infectious material using spatiotemporal transmission data". https://doi.org/10.5281/zenodo.7428123

Gamża, A. M. (2023b). Data and code for: "Using spatial modelling to explore the host-density dependence of R0 for environmental transmission". https://doi.org/10.5281/zenodo.7430198

Garrity, G. M., Bell, J. A., & Lilburn, T. (2005). Class V. Epsilonproteobacteriaclass. nov. In D. J. Brenner, N. R. Krieg, & J. T. Staley (Eds.), *Bergey's Manual® of Systematic Bacteriology: Volume Two The Proteobacteria Part C The Alpha-, Beta-, Delta-, and Epsilonproteobacteria* (pp. 1145-1194). Springer US. https://doi.org/10.1007/978-0-387-29298-4_4

Gilligan, C. A. (1995). Modelling soil-borne plant pathogens: reaction-diffusion models. *Canadian Journal of Plant Pathology*, *17*(2), 96-108. https://doi.org/10.1080/07060669509500700

Hansson, I., Sandberg, M., Habib, I., Lowman, R., & Engvall, E. O. (2018). Knowledge gaps in control of Campylobacter for prevention of campylobacteriosis. *Transboundary and Emerging Diseases*, *65*(S1), 30-48. https://doi.org/10.1111/tbed.12870

Heesterbeek, J. A. P. (2002). A Brief History of R0 and a Recipe for its Calculation. *Acta Biotheoretica*, *50*(3), 189-204. https://doi.org/10.1023/A:1016599411804

Heesterbeek, J. A. P. (2005). The law of mass-action in epidemiology: A historical perspective. In K. C. a. B. Beisner (Ed.), *Ecological paradigms lost : Routes of theory change.* (pp. 81-105). Academic Press. https://doi.org/10.1016/B978-012088459-9/50007-8

Heres, L., Engel, B., Urlings, H. A. P., Wagenaar, J. A., & van Knapen, F. (2004). Effect of acidified feed on susceptibility of broiler chickens to intestinal infection by Campylobacter and Salmonella. *Veterinary Microbiology*, *99*(3), 259-267. https://doi.org/10.1016/j.vetmic.2003.12.008

Heres, L., Engel, B., Van Knapen, F., Wagenaar, J. A., & Urlings, B. A. P. (2003). Effect of fermented feed on the susceptibility for Campylobacter jejuni colonisation in broiler chickens with and without concurrent inoculation of Salmonella enteritidis. *International Journal of Food Microbiology*, *87*(1), 75-86. https://doi.org/10.1016/S0168-1605(03)00055-2

Herfst, S., Schrauwen, E. J. A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V. J., Sorrell, E. M., Bestebroer, T. M., Burke, D. F., Smith, D. J., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2012). Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets. *Science*, *336*(6088), 1534-1541. https://doi.org/10.1126/science.1213362

Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., Beutels, P., & Hens, N. (2019). A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology (Cambridge, Mass.)*, *30*(5), 723.

Holt, P. S., Mitchell, B. W., & Gast, R. K. (1998). Airborne horizontal transmission of Salmonella enteritidis in molted laying chickens. *Avian Dis*, *42*(1), 45-52.

Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution. *Computational Statistics & Data Analysis*, *59*, 41-51. https://doi.org/10.1016/j.csda.2012.10.006

Huang, W., Han, M., & Liu, K. (2010). Dynamics of an SIS reaction-diffusion epidemic model for disease transmission. *Mathematical biosciences and engineering : MBE*, *7*(1), 51-66. https://doi.org/10.3934/mbe.2010.7.51

Johnston, C., Hughes, H., Lingard, S., Hailey, S., & Healy, B. (2022). Immunity and infectivity in covid-19. *BMJ*, *378*, e061402. https://doi.org/10.1136/bmj-2020-061402

Jønsson, K. A., Tøttrup, A. P., Borregaard, M. K., Keith, S. A., Rahbek, C., & Thorup, K. (2016). Tracking Animal Dispersal: From Individual Movement to Community Assembly and Global Range Dynamics. *Trends in Ecology & Evolution*, *31*(3), 204-214. https://doi.org/10.1016/j.tree.2016.01.003

Kassem, I. I., Helmy, Y. A., Kathayat, D., Candelero-Rueda, R. A., Kumar, A., Deblais, L., Huang, H.-C., Sahin, O., Zhang, Q., & Rajashekara, G. (2017). Nonculturability Might Underestimate the Occurrence of Campylobacter in Broiler Litter. *Foodborne Pathogens and Disease*, *14*(8), 472-477. https://doi.org/10.1089/fpd.2017.2279

Klafter, J., Shlesinger, M. F., & Zumofen, G. (1996). Beyond brownian motion. *Physics today*, *49*(2), 33-39.

Kretzschmar, M. E., Ashby, B., Fearon, E., Overton, C. E., Panovska-Griffiths, J., Pellis, L., Quaife, M., Rozhnova, G., Scarabel, F., Stage, H. B., Swallow, B., Thompson, R. N., Tildesley, M. J., & Villela, D. (2022). Challenges for modelling interventions for future pandemics. *Epidemics*, *38*, 100546. https://doi.org/10.1016/j.epidem.2022.100546

Lanzas, C., Davies, K., Erwin, S., & Dawson, D. (2020). On modelling environmentally transmitted pathogens. *Interface Focus*, *10*(1), 20190056. https://doi.org/10.1098/rsfs.2019.0056

Li, L., Mendis, N., Trigui, H., Oliver, J. D., & Faucher, S. P. (2014). The importance of the viable but non-culturable state in human bacterial pathogens [Review]. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00258

Line, J., Hiett, K., & Conlan, A. (2008). Comparison of Challenge Models for Determining the Colonization Dose of Campylobacter jejuni in Broiler Chicks. *Poultry Science*, *87*(9), 1700-1706. https://doi.org/10.3382/ps.2008-00027

Lunn, T. J., Restif, O., Peel, A. J., Munster, V. J., de Wit, E., Sokolow, S., van Doremalen, N., Hudson, P., & McCallum, H. (2019). Dose–response and transmission: the nexus between reservoir hosts, environment and recipient hosts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1782), 20190016. https://doi.org/10.1098/rstb.2019.0016

Lv, R., Wang, K., Feng, J., Heeney, D. D., Liu, D., & Lu, X. (2020). Detection and Quantification of Viable but Non-culturable Campylobacter jejuni. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02920

Moore, J. E. (2001). Bacterial dormancy in Campylobacter: abstract theory or cause for concern? *International Journal of Food Science & Technology*, *36*(6), 593-600. https://doi.org/10.1046/j.1365-2621.2001.00508.x

Mota-Gutierrez, J., Lis, L., Lasagabaster, A., Nafarrate, I., Ferrocino, I., Cocolin, L., & Rantsiou, K. (2022). Campylobacter spp. prevalence and mitigation strategies in the broiler production chain. *Food Microbiology*, *104*, 103998. https://doi.org/10.1016/j.fm.2022.103998

Murphy, C., Carroll, C., & Jordan, K. N. (2006). Environmental survival mechanisms of the foodborne pathogen Campylobacter jejuni. *Journal of Applied Microbiology*, *100*(4), 623-632. https://doi.org/10.1111/j.1365-2672.2006.02903.x

Newell, D. G. (2002). The ecology of Campylobacter jejuni in avian and human hosts and in the environment. *International Journal of Infectious Diseases*, *6*, S16-S21. https://doi.org/10.1016/S1201-9712(02)90179-7

Nightingale, E. S., Brady, O. J., group, C. C.-w., & Yakob, L. (2021). The importance of saturating density dependence for population-level predictions of SARS-CoV-2 resurgence compared with density-independent or linearly density-dependent models, England, 23 March to 31 July 2020. *Eurosurveillance*, *26*(49), 2001809. https://doi.org/10.2807/1560-7917.ES.2021.26.49.2001809

Pang, D., & Xiao, Y. (2019). The SIS model with diffusion of virus in the environment. *Mathematical Biosciences and Engineering*, *16*(4), 2852-2874.

Park, S. F. (2002). The physiology of Campylobacter species and its relevance to their role as foodborne pathogens. *International Journal of Food Microbiology*, *74*(3), 177-188. https://doi.org/10.1016/S0168-1605(01)00678-X

Pielaat, A., & Van Den Bosch, F. (1998). A model for dispersal of plant pathogens by rainsplash. *Mathematical Medicine and Biology: A Journal of the IMA*, *15*(2), 117-134.

Ramamurthy, T., Ghosh, A., Pazhani, G. P., & Shinoda, S. (2014). Current Perspectives on Viable but Non-Culturable (VBNC) Pathogenic Bacteria [Review]. *Frontiers in Public Health*, *2*. https://doi.org/10.3389/fpubh.2014.00103

Real, L. A., & Biek, R. (2007). Infectious Disease Modeling and the Dynamics of Transmission. In J. E. Childs, J. S. Mackenzie, & J. A. Richt (Eds.), *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission* (pp. 33-49). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70962-6_2

Rees, E. M., Minter, A., Edmunds, W. J., Lau, C. L., Kucharski, A. J., & Lowe, R. (2021). Transmission modelling of environmentally persistent zoonotic diseases: a systematic review. *The Lancet Planetary Health*, *5*(7), e466-e478. https://doi.org/10.1016/S2542-5196(21)00137-6

Sahin, O., Morishita, T. Y., & Zhang, Q. (2002). Campylobacter colonization in poultry: sources of infection and modes of transmission. *Anim Health Res Rev*, *3*(2), 95-105. https://doi.org/10.1079/ahrr200244

Singanayagam, A., Patel, M., Charlett, A., Lopez Bernal, J., Saliba, V., Ellis, J., Ladhani, S., Zambon, M., & Gopal, R. (2020). Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Eurosurveillance*, *25*(32), 2001483. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001483

Smith, S., Meade, J., Gibbons, J., McGill, K., Bolton, D., & Whyte, P. (2016). The impact of environmental conditions on Campylobacter jejuni survival in broiler faeces and litter. *Infection Ecology & Epidemiology*, *6*(1), 31685. https://doi.org/10.3402/iee.v6.31685

Smith, T. P., Flaxman, S., Gallinat, A. S., Kinosian, S. P., Stemkovski, M., Unwin, H. J. T., Watson, O. J., Whittaker, C., Cattarino, L., Dorigatti, I., Tristem, M., & Pearse, W. D. (2021). Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. *Proceedings of the National Academy of Sciences*, *118*(25), e2019284118. https://doi.org/10.1073/pnas.2019284118

Smoluchowski, M. v. (1916). Drei vortrage uber diffusion, brownsche bewegung und koagulation von kolloidteilchen. *Zeitschrift fur Physik*, *17*, 557-585.

Sørensen, J. H., Jensen, C. Ø., Mikkelsen, T., Mackay, D. K. J., & Donaldson, A. I. (2001). Modelling the atmospheric dispersion of foot-and-mouth disease virus for emergency preparedness. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, *26*(2), 93-97. https://doi.org/10.1016/S1464-1909(00)00223-9

Stern, N. J., Cox, N. A., Musgrove, M. T., & Park, C. M. (2001). Incidence and Levels of Campylobacter in Broilers After Exposure to an Inoculated Seeder Bird. *Journal of Applied Poultry Research*, *10*(4), 315-318. https://doi.org/10.1093/japr/10.4.315

Stern, N. J., & Robach, M. C. (2003). Enumeration of Campylobacter spp. in Broiler Feces and in Corresponding Processed Carcasses. *Journal of Food Protection*, *66*(9), 1557-1563. https://doi.org/10.4315/0362-028X-66.9.1557

Sy, K. T. L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *PLOS ONE*, *16*(4), e0249271. https://doi.org/10.1371/journal.pone.0249271

van Bunnik, B. A. (2014). *Mechanisms underlying disease transmission between spatially separated animals*. Wageningen University and Research.

van Bunnik, B. A. D., Hagenaars, T. J., Bolder, N. M., Nodelijk, G., & de Jong, M. C. M. (2012). Interaction effects between sender and receiver processes in indirect transmission of Campylobacter jejuni between broilers. *BMC Veterinary Research*, *8*(1), 123. https://doi.org/10.1186/1746-6148-8-123

van Bunnik, B. A. D., Ssematimba, A., Hagenaars, T. J., Nodelijk, G., Haverkate, M. R., Bonten, M. J. M., Hayden, M. K., Weinstein, R. A., Bootsma, M. C. J., & De Jong, M. C. M. (2014). Small distances can keep bacteria at bay for days. *Proceedings of the National Academy of Sciences*, *111*(9), 3556-3560. https://doi.org/10.1073/pnas.1310043111

van Elsas, J. D., Semenov, A. V., Costa, R., & Trevors, J. T. (2011). Survival of Escherichia coli in the environment: fundamental and public health aspects. *Isme j*, *5*(2), 173-183. https://doi.org/10.1038/ismej.2010.80

van Immerseel, F., de Buck, J., Pasmans, F., Bohez, L., Boyen, F., Haesebrouck, F., & Ducatelle, R. (2004). Intermittent long-term shedding and induction of carrier birds after infection of chickens early posthatch with a low or high dose of Salmonella enteritidis. *Poultry Science*, *83*(11), 1911-1916. https://doi.org/10.1093/ps/83.11.1911

van Leuken, J. P. G., van de Kassteele, J., Sauter, F. J., van der Hoek, W., Heederik, D., Havelaar, A. H., & Swart, A. N. (2015). Improved correlation of human Q fever incidence to modelled C. burnetii concentrations by means of an atmospheric dispersion model. *International Journal of Health Geographics*, *14*(1), 14. https://doi.org/10.1186/s12942-015-0003-y

Velthuis, A. G. J., Bouma, A., Katsma, W. E. A., Nodelijk, G., & De Jong, M. C. M. (2007). Design and analysis of small-scale transmission experiments with animals. *Epidemiology and Infection*, *135*(2), 202-217. https://doi.org/10.1017/S095026880600673X

Wagenaar, J. A., French, N. P., & Havelaar, A. H. (2013). Preventing Campylobacter at the Source: Why Is It So Difficult? *Clinical Infectious Diseases*, *57*(11), 1600-1606. https://doi.org/10.1093/cid/cit555

Wagner, J., Sparks, T. L., Miller, S., Chen, W., Macher, J. M., & Waldman, J. M. (2021). Modeling the impacts of physical distancing and other exposure determinants on aerosol transmission. *Journal of Occupational and Environmental Hygiene*, *18*(10-11), 495-509. https://doi.org/10.1080/15459624.2021.1963445

Wang, X., Zhao, X.-Q., & Wang, J. (2018). A cholera epidemic model in a spatiotemporally heterogeneous environment. *Journal of Mathematical Analysis and Applications*, *468*(2), 893-912. https://doi.org/10.1016/j.jmaa.2018.08.039

Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C., & Timmer, J. (2021). On structural and practical identifiability. *Current Opinion in Systems Biology*, *25*, 60-69. https://doi.org/10.1016/j.coisb.2021.03.005

Willem, L., Verelst, F., Bilcke, J., Hens, N., & Beutels, P. (2017). Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infectious Diseases*, *17*(1), 612. https://doi.org/10.1186/s12879-017-2699-8

Wirth, F. N., Johns, M., Meurers, T., & Prasser, F. (2020). Citizen-Centered Mobile Health Apps Collecting Individual-Level Spatial Data for Infectious Disease Management: Scoping Review. *JMIR Mhealth Uhealth*, *8*(11), e22594. https://doi.org/10.2196/22594

Wolfram Research, I. (2019). Wolfram Research, Inc.

Xiao, Y., Xiang, C., Cheke, R. A., & Tang, S. (2020). Coupling the Macroscale to the Microscale in a Spatiotemporal Context to Examine Effects of Spatial Diffusion on Disease Transmission. *Bull Math Biol*, *82*(5), 58. https://doi.org/10.1007/s11538-020-00736-9

Yang, S., & Rothman, R. E. (2004). PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases*, *4*(6), 337-348. https://doi.org/10.1016/S1473-3099(04)01044-8

Yildiz, F. H. (2007). Processes controlling the transmission of bacterial pathogens in the environment. *Res Microbiol*, *158*(3), 195-202. https://doi.org/10.1016/j.resmic.2006.12.005

Zhou, J., Wei, J., Choy, K.-T., Sia, S. F., Rowlands, D. K., Yu, D., Wu, C.-Y., Lindsley, W. G., Cowling, B. J., McDevitt, J., Peiris, M., Li, Y., & Yen, H.-L. (2018). Defining the sizes of airborne particles that mediate influenza transmission in ferrets. *Proceedings of the National Academy of Sciences*, *115*(10), E2386-E2392. https://doi.org/10.1073/pnas.1716771115

Summary

As most pathogens are transmitted through the environment, the environment plays an important role in sustaining transmission, which is presenting opportunities for developing effective intervention strategies targeted on the environmental stage of transmission. To better understand the processes of environmental transmission and subsequently design and quantify relevant intervention strategies, both formulating and implementing the models as well as designing experiments or sampling protocols to gather data are necessary. In this thesis, I presented the development cycle of modelling and experimental work designed to study mechanisms of environmental transmission from the perspective of the infectious material that accumulates in the environment. One of the advantages of the infectious material perspective is that it avoids unnatural representations of environmental transmission - the perspective of between-host contact. Unnatural perspective can lead to confusion during the communication of findings and therefore hinder the policy development.

In **Chapter 1**, (General introduction), the full modelling framework was described in detail, together with a brief description of the available data types and the model system that I presented here being *C. jejuni* transmission in broilers. The parsimonious model we constructed describes mechanisms of pathogen transmission via the environment using only three parameters: the decay rate parameter, describing decay (inactivation) of infectious material in the environment, the transmission rate parameter, describing jointly shedding of the infectious material by infectious hosts, absorption of (portion of) this material by recipient hosts and probability of infection once the material is absorbed, and the diffusion coefficient describing spatial spread of infectious material.

In **Chapter 2**, we presented the calibration and validation of our model, using a series of experiments on *C. jejuni* transmission between spatially separated broilers (including two unpublished experiments designed for the purpose of the study presented in this thesis). Our analysis showed that the spatiotemporal model is fully identifiable with data we collected and that after calibration the model satisfactorily describes the experimental observations. We prove that data obtained in new experiments, providing data of necessary spatial resolution, was crucial to ensure the model identifiability. Moreover, as we were able to obtain informative estimates of all parameters, new insights into *Campylobacter* biology were obtained. We demonstrated that environmental decay of infectious forms of *C. jejuni* is much slower than decay observed for its culturable form, and discussed the model result that suggests that spatial dispersion of infectious material (containing *C. jejuni*) is most likely a result of multistep/multi-route dispersion process.

In **Chapter 3**, we analysed in more detail model identifiability, to inform the design of future transmission experiments. We conducted both structural and practical identifiability analyses, using a combination of methods, including mathematical analysis of the model, and analysis of the profile likelihoods for numerous combinations of simulated and experimental data. Subsequently, we

proposed a strategy to obtain the most informative experimental design, which is crucial from ethical, economical, and scientific points of view.

In **Chapter 4**, we demonstrated how our validated modelling framework can further be used to explore the density dependence of homogenous mixing, clustering, and distancing of hosts in the context of environmental transmission. As the analysis of density dependence is, by nature, a spatial problem, our spatial framework was used to generate a number of scenarios describing homogenous mixing, Poisson process mixing and strong distancing of hosts. We quantified, how constraining of host movements reduces the basic reproduction ratio, which can be used to assess nonpharmaceutical intervention strategies targeted towards reducing host mobility. Moreover, we explored for which parameter combinations host clustering or distancing, as opposed to random host placement, needs to be considered in the model as they diverge from linear relationship predicted for random mixing models. This study, when supplemented with relevant data, would serve as a basis for future studies exploring hosts-clustering and distancing behaviour and its influence on transmission and population dynamics.

Overall, as it is discussed in **Chapter 5**, in this thesis I presented a development cycle, in which models were developed together with experimental design to obtain useful data and modelling results. During the process, on the one hand, new methodology with broad applicability was developed, and on the other hand, new insights into our model system- *Campylobacter* transmission in broilers- were gained.

Acknowledgements

I would like to thank both my supervisors, Mart de Jong and Thomas Hagenaars for their guidance, support and immense patience.

I would also like to acknowledge all the co-workers I have met during my work in the QVE chair group, in the epidemiology group in Lelystad and during the experimental work for their help and feedback.

Additionally, I would like to thank my friend Ian Boulton for the language check on the final manuscript of this thesis.

Thank you!

Curriculum Vitae

Anna Maria Gamża was born on 7th October 1992 in Tychy, Poland in a Silesian family. In 2018, she obtained her MSc degrees in Veterinary Medicine from Jagiellonian University & University of Agriculture, Poland, and in Biophysics from Jagiellonian University, Poland. The same year she started her PhD in Veterinary Epidemiology at Wageningen University and Research, the Netherlands; under the supervision of Mart de Jong and Thomas Hagenaars, she conducted research on mathematical modelling of environmental transmission which resulted in this thesis. In 2022, she started a position as a Research Assistant at the Roslin Institute, University of Edinburgh, Scotland working on spatial epidemiological modelling and data analysis.

Publications

Peer-reviewed publications

Gamża, A.M., T. J. Hagenaars, M.G.J. Koene and M.C.M. de Jong (2023). "Combining a parsimonious mathematical model with infection data from tailor-made experiments to understand environmental transmission." Scientific Reports. (under review)

Gamża, A.M., M.C.M. de Jong and T.J. Hagenaars (2023). "Identifiability of environmental transmission parameters: quantifying dispersion and decay of infectious material using spatiotemporal transmission data." (in preparation)

Gamża, A.M., M.C.M. de Jong and T.J. Hagenaars (2023). " Using spatial modelling to explore the host-density dependence of R0 for environmental transmission." (in preparation)

Conference proceedings

Gamża, A.M., T. J. Hagenaars and M. C. M. de Jong (2019). Modelling environmental transmission and spatial spreading of common zoonotic bacteria in poultry. WIAS Science Day, Wageningen, the Netherlands. Poster.

Gamża, A. M., T. J. Hagenaars, M. G. J. Koene and M. C. M. de Jong (2019). Spatial Indirect Transmission of Campylobacter Infection in Broilers is Possible Only Within a Certain Distance. One Heath EJP Annual Scientific Meeting 2019, Dublin, Ireland. Oral presentation.

Gamża, A. M., T. J. Hagenaars, M. G. J. Koene and M. C. M. de Jong (2021). Understanding environmental pathogen transmission by combining parsimonious mathematical models and tailor-made animal experiments. Epidemics8 - International Conference on Infectious Disease Dynamics, Online. Oral presentation.

Gamża, A. M., T. J. Hagenaars, M. G. J. Koene and M. C. M. de Jong (2022). Environmental survival of infectious forms of Campylobacter: insights from mathematical modelling and experiments in broilers. WIAS Annual Conference Wageningen, the Netherlands. Oral presentation.

Training and Supervision Plan (TSP)

Graduate School WIAS

A. The Basic Package	year	credits *
WIAS Introduction Day **(mandatory)**	2018	0.3
Ethics and Animal Sciences	2020	0.8
Scientific integrity	2020	0.6
Course on essential skills (Frank Little) *(recommended)*	2018	1.2
Subtotal Basic Package		**2.9**

B. Disciplinary Competences	year	credits *
Preparing own PhD research proposal	2018	6.0
Introduction to Laboratory Animal Science, Utrecht University	2019	3.0
Species-specific LAS course on poultry and other birds (galliformes, passeriformes and columbiformes), Utrecht University	2019	1.4
18.6501x: Fundamentals of Statistics (a course of study offered by MITx, an online learning initiative of the Massachusetts Institute of Technology)	2020	3.0
6.431x: Probability - The Science of Uncertainty and Data (a course of study offered by MITx, an online learning initiative of the Massachusetts Institute of Technology)	2020	3.0
Subtotal Disciplinary Competences		**16.4**

C. Professional Competences	year	credits *
WGS course: Scientific Writing	2021	1.8
WGS course: Effective and efficient communication in academia and beyond	2022	0.8
WIAS Course: The Final Touch	2021	0.6
Subtotal Professional Competences		**3.2**

D. Presentation Skills *(maximum 4 credits)*	year	credits *
< title of presentation, name of conference/seminar, place, date, oral / poster >		
"Spatial Indirect Transmission of Campylobacter Infection in Broilers is Possible Only Within a Certain Distance.", One Heath EJP Annual Scientific Meeting 2019, Dublin, oral	2019	1.0
"Modelling environmental transmission and spatial spreading of common zoonotic bacteria in poultry.", Wias Science Day 2019, poster	2019	1.0
"Understanding environmental transmission of Campylobacter in broilers: models and experiments.", One Health EJP Annual Scientific Meeting 2020, online, oral	2020	1.0
"Understanding environmental pathogen transmission by combining parsimonious mathematical models and tailor-made animal experiments", Epidemics8 - International Conference on Infectious Disease Dynamics, online, oral	2021	1.0
Subtotal presentations		**4.0**

E. Teaching competences *(max 6 credits)*	year	credits *
Supervising practicals-QVE-30806 Management of Infections and Diseases in Animal Populations	2019-2022	4.0
Subtotal Teaching competences		**4.0**

Education and Training Total (minimum 30 credits)*	**30.5**

**One ECTS credit equals a studyload of approximately 28 hours*

The research described in this thesis was financially supported by the One Health European Joint Program (OHEJP) on Monitoring the gut Microbiota and Immune Response to Predict, Prevent, and Control zoonoses in humans and livestock in order to minimize the use of antimicrobials (MoMIR-PPC), which received funding from the European Union's Horizon 2020 research, the Dutch Ministry of Agriculture, Nature and Food Quality (grant WOT-01-005-003) and from an Agri&Food partnership (grant AF-EU-18031).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Anna Maria Gamża
Printed by Digiforce.