INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Kerr L, Sproul D, Grima R. 2022 Cluster mean-field theory accurately predicts statistical properties of large-scale DNA methylation patterns. *J. R. Soc. Interface* **19**: 20210707.
https://doi.org/10.1098/rsif.2021.0707

Received: 8 September 2021
Accepted: 13 December 2021

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
biomathematics, computational biology

Keywords:
DNA methylation, stochastic modelling, master equations, epigenetics

Author for correspondence:
Ramon Grima
e-mail: ramon.grima@ed.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Cluster mean-field theory accurately predicts statistical properties of large-scale DNA methylation patterns

Lyndsay Kerr[1], Duncan Sproul[2] and Ramon Grima[3]

[1]MRC Institute of Genetics and Cancer,[2]MRC Human Genetics Unit and CRUK Edinburgh Centre, Institute of Genetics and Cancer, and[3]School of Biological Sciences, University of Edinburgh, Edinburgh, UK

RG, 0000-0002-1266-8169

The accurate establishment and maintenance of DNA methylation patterns is vital for mammalian development and disruption to these processes causes human disease. Our understanding of DNA methylation mechanisms has been facilitated by mathematical modelling, particularly stochastic simulations. Megabase-scale variation in DNA methylation patterns is observed in development, cancer and ageing and the mechanisms generating these patterns are little understood. However, the computational cost of stochastic simulations prevents them from modelling such large genomic regions. Here, we test the utility of three different mean-field models to predict summary statistics associated with large-scale DNA methylation patterns. By comparison to stochastic simulations, we show that a cluster mean-field model accurately predicts the statistical properties of steady-state DNA methylation patterns, including the mean and variance of methylation levels calculated across a system of CpG sites, as well as the covariance and correlation of methylation levels between neighbouring sites. We also demonstrate that a cluster mean-field model can be used within an approximate Bayesian computation framework to accurately infer model parameters from data. As mean-field models can be solved numerically in a few seconds, our work demonstrates their utility for understanding the processes underpinning large-scale DNA methylation patterns.

1. Introduction

DNA methylation is a repressive epigenetic mark [1] which is primarily found on the cytosines of CpG dinucleotides in mammals. Double-stranded CpG dyads can be unmethylated or methylated on both strands (u and m, respectively) or methylated on only one strand (hemimethylated, h). DNA methylation is largely erased from the genome during early mammalian development [2]. It is then re-established by the de novo DNA methyltransferases DNMT3A and DNMT3B [3] resulting in a landscape where 70–80% of CpGs are methylated in most human cells [4]. Regulatory elements such as promoters and enhancers often remain methylation free [1]. During DNA replication, the nascent strand is synthesized with unmethylated cytosines and methylation patterns are copied by the maintenance methyltransferase, DNMT1 [5]. Failure to maintain DNA methylation at a locus results in passive DNA demethylation. Methylation can also be removed actively through transient modification by ten eleven translocation (TET) enzymes and subsequent DNA repair [6].

Waves of demethylation and remethylation take place during early development and the generation of germline cells [2]. Changes in DNA methylation patterns also occur during development and cellular differentiation, resulting in cell type-specific methylation patterns [7]. The correct establishment of DNA methylation patterns is vital for normal development. Mutations in DNMTs cause Mendelian disorders in humans [8–10] and mice knockouts

© 2022 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

die before or shortly after birth [3,5]. Widespread alterations in DNA methylation patterns occur in cancer and ageing [11,12], but the significance of these changes is unclear. It has been observed that globally hypomethylated mice expressing a single hypomorphic DNMT1 allele develop cancer, suggesting that altered DNA methylation can cause cancer [13]. However, the mechanisms underpinning DNA methylation changes remain unclear preventing the robust delineation of their role in development and disease.

Mathematical models are powerful tools for understanding complex biological processes, including DNA methylation. The importance of interactions between CpGs in maintaining DNA methylation patterns was first postulated through modelling [14]. Specifically, the authors modelled collaborative interactions where CpGs within a region of the genome can influence the state of other CpGs, e.g. through enzyme recruitment. Models including such collaborativity were subsequently found to explain experimental measurements of methylation maintenance *in vitro* and *in vivo* more closely than those that did not include it [15,16]. A recent study also suggests that collaborativity mediated by neighbour-guided error correction through DNMT1 is important for maintaining DNA methylation [17]. Deterministic models, non-spatial stochastic models and spatial stochastic models have all been used to describe DNA methylation [18]. Deterministic models are based on rate equations while stochastic models are based on Fokker–Planck equations or chemical master equations (CMEs). CMEs are ideal because they take into account the inherent discreteness of molecular fluctuations [19] which is well known to play an important role in cellular dynamics [20]. The CME of simple non-spatial stochastic models can be solved exactly in closed-form [21], but this is often not possible for spatial stochastic models. Rather in this case, stochastic simulations are used to model the individual reaction processes described by the CME. Various types of stochastic models have been used to describe collaborative methylation systems (e.g. [22–26]). To date, such mathematical models have been applied to understand methylation patterns on a kilobase scale. However, megabase-scale alterations to DNA methylation patterns occur in development, cancer and ageing [27]. These include the formation of megabase-sized partially methylated domains in cancer which have been linked to genome instability [27,28]. Existing models rely on simulations that are too computationally expensive to run for such large genomic regions.

Here, we test the idea that large-scale steady-state methylation patterns can be modelled in a tractable manner using mean-field (MF) models. By comparison to synthetic data generated from stochastic simulations, we demonstrate that a type of cluster MF model can predict the statistical properties of large-scale methylation patterns. In §2, we introduce a nearest-neighbour collaborative model for DNA methylation and describe the process used to simulate data from this model. We describe the three MF models we test in §3. In §4, we compare the ability of each MF model to predict statistics associated with methylation patterns resulting from the simulations. We find that a type of cluster MF model provides excellent predictions and demonstrate that this model can be used within an approximate Bayesian computation (ABC) framework to infer parameters underpinning large-scale methylation systems. Finally, in §5, we discuss the implications of our findings.

2. Nearest-neighbour collaborative model

2.1. Model set-up

We consider the reaction system in figure 1, where some reactions are non-collaborative (involving only the 'target' CpG, whose methylation state changes during the reaction), while others are collaborative (involving both a target CpG and a 'mediator' CpG). The role of the mediator is to encourage the reaction to occur, e.g. via the recruitment of methylase or demethylase enzymes. This system, and reduced versions, have previously been used to examine small-scale methylation patterns [14,29,30]. Previously, collaborative and non-collaborative reactions have been referred to as positive-feedback and noisy reactions in a study modelling histone modifications [31].

We make the following assumptions:

(i) *CpGs can only influence the methylation state of their nearest neighbours* (see figure 2). While both experimental and modelling studies have demonstrated the importance of CpGs being influenced by surrounding CpGs [14,15], the extent and range of such influence are unknown. We therefore consider only interactions between nearest neighbours.

(ii) *There are no direct transitions between the unmethylated and methylated states.* We justify this with the observation that methylase and demethylase enzymes act on single DNA strands [6,32]. This implies that hemimethylation is a necessary transition state between unmethylated and fully methylated CpGs.

(iii) *The system has reached a steady state.* Here, we assume that there are no long-term effects of DNA replication on methylation patterns. This assumption is supported by the observation that the DNA methylation patterns of cycling and arrested cells are similar [33].

We do not impose an upper bound on the system size, $N \in \mathbb{N}$, allowing large-scale methylation patterns to be considered. We also assume that the rates k_i, $i = \{1, 2, \ldots, 12\}$ are of the form given in table 1, where x measures the strength of collaborativity between CpGs ($x < 1$ indicates that non-collaborative reactions dominate, while $x > 1$ indicates that collaborative reactions dominate). The parameter y measures the strength of methylation versus demethylation ($y < 1$ corresponds to demethylation dominating and $y > 1$ corresponds to methylation dominating). The parameter $a > 0$ scales the reaction rates. While the values of $x \geq 0$, $y \geq 0$ determine which reactions are most probable (see table 1 legend), all reactions in figure 1 are possible for any $\{x, y\}$.

2.2. Simulations of nearest-neighbour collaborative system

In the simple case of two CpGs, the system can be in six possible states: *mm*, *uu*, *hh*, *um*, *hm* and *uh*. For such a small system, all possible transitions between states (via reactions in figure 1) can be identified and the evolution of the system can be described exactly by six mathematical equations, one for each state. However, for large systems, it is infeasible to identify all possible states and transitions between states meaning that equations describing the exact evolution of the system cannot be formulated. While

non-collaborative reactions:

$$\underline{\text{methylation}} \qquad\qquad \underline{\text{demethylation}}$$
$$u \xrightarrow{k_1} h, \quad h \xrightarrow{k_2} m, \qquad\qquad m \xrightarrow{k_3} h, \quad h \xrightarrow{k_4} u.$$

collaborative reactions:

$$\underline{\text{methylation}}$$
$$u + h \xrightarrow{k_5} h + h, \quad u + m \xrightarrow{k_6} h + m, \quad h + h \xrightarrow{k_7} m + h, \quad h + m \xrightarrow{k_8} m + m,$$

$$\underline{\text{demethylation}}$$
$$m + h \xrightarrow{k_9} h + h, \quad m + u \xrightarrow{k_{10}} h + u, \quad h + h \xrightarrow{k_{11}} u + h, \quad h + u \xrightarrow{k_{12}} u + u.$$

Figure 1. System of reactions under consideration. Here u, h and m represent unmethylated, hemimethylated and methylated CpGs, respectively. Non-collaborative reactions involve only one CpG, while collaborative reactions involve two CpGs. For each collaborative reaction, the second reactant (the mediator, see text) recruits an enzyme that changes the methylation state of the first reactant (the target). For example, the reaction $u + h \rightarrow h + h$ involves a hemimethylated CpG at one site recruiting a methylase enzyme which changes the state of a CpG at another site from unmethylated to hemimethylated. Reaction rates are k_i, $i = \{1, \ldots, 12\}$. Note that only one CpG site changes methylation state during each reaction.

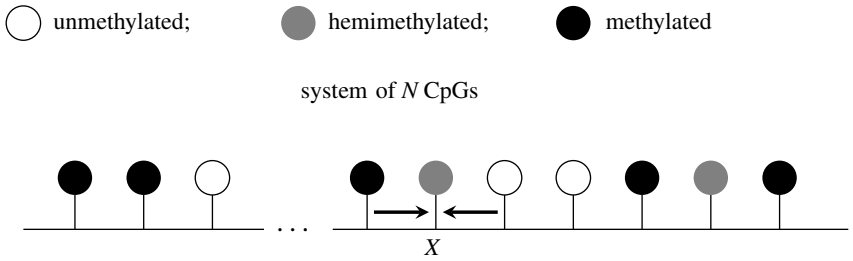

Figure 2. Collaborative interactions that can influence a target X under the nearest-neighbour collaborative model. Individual CpGs are represented by 'lollipops', with their colour indicating their methylation status (white: unmethylated; grey: hemimethylated; black: methylated). Collaborative methylation and demethylation reactions can only occur between neighbouring CpGs (potential influences on CpG X are indicated by arrows).

Table 1. Reaction rates for the model in figure 1. Here, $a > 0$, $x \geq 0$, $y \geq 0$ so that all reaction rates are non-negative. The values of x and y determine the probability associated with each type of reaction. For example, consider the case $x < 1$, $y < 1$. Then $k_3 = k_4 = a$ is the largest reaction rate and so non-collaborative demethylation reactions are most probable, while $k_5 = k_6 = k_7 = k_8 = axy$ is the smallest reaction rate and so collaborative methylation reactions are least probable.

	demethylation	methylation
non-collaborative	$k_3 = k_4 = a$	$k_1 = k_2 = ay$
collaborative	$k_9 = k_{10} = k_{11} = k_{12} = ax$	$k_5 = k_6 = k_7 = k_8 = axy$

stochastic simulations are computationally expensive, they are the ground truth of the nearest-neighbour collaborative system to which we compare our MF models and so here we describe the process underlying these simulations.

We focus on the steady-state case so that u, h and m levels fluctuate around some fixed steady-state values. For a system of N CpGs, we simulate the nearest-neighbour collaborative system using Gillespie's algorithm [34]—see figure 3 for an illustration of how this algorithm simulates a 4-CpG system. Essentially, in each step of Gillespie's algorithm, a timepoint is chosen and a single reaction selected to occur. This reaction changes the methylation state of a single CpG, i.e. all other CpGs remain unchanged. Throughout the simulations, we monitor the proportions of u, h and m sites, declaring steady state to be attained when these fluctuate around fixed values, after which we sample the methylation pattern, containing N methylation states, at constant time intervals. For each parameter set, we sample at $T = 10^6/N$ timepoints to obtain a dataset of 10^6 steady-state methylation states.

2.3. Analysis of simulated data

To facilitate our analysis, we define a sequence u^t, $t \in \{1, \ldots, T\}$, where

$$u_i^t = \begin{cases} 1, & \text{if CpG } i \text{ is in the } u \text{ state at timepoint } t, \\ 0, & \text{otherwise.} \end{cases}$$

We define h^t and m^t similarly; see figure 4. We also define z^t via

$$z^t = u^t + 2h^t + 3m^t. \tag{2.1}$$

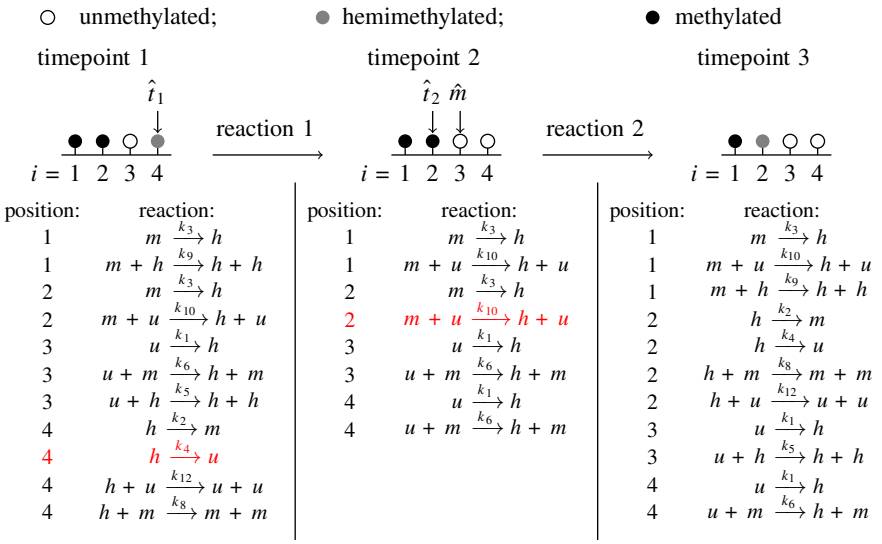

Figure 3. Stochastic simulations of the nearest-neighbour collaborative system. For simplicity, a 4-CpG system at 3 timepoints is considered here. We fix a, x and y and impose periodic boundary conditions so that the first and final CpG can interact. After all potential non-collaborative and collaborative reactions are identified, Gillespie's algorithm chooses a timepoint and a single reaction to occur at this timepoint. This reaction changes the methylation state of a single CpG and the list of potential reactions is updated to account for this change. This process is repeated to generate dynamical behaviour. In the example shown, we start with the system on the left. All possible reactions are listed and at the first timepoint (chosen by Gillespie's algorithm) a $h \xrightarrow{k_4 u}$ reaction is chosen to occur at target position $\hat{t}_1$. The methylation state at the target position is changed accordingly (all other CpGs remain unchanged) and the list of potential reactions is updated to account for this change (middle). At the next timepoint (chosen by Gillespie's algorithm), a $m + u \xrightarrow{k_{10}} h + u$ reaction is chosen to occur at target position, $\hat{t}_2$, with CpG $\hat{m}$ acting as mediator. Again, the methylation state at the target position is changed accordingly (all other CpGs remain unchanged) and the list of potential reactions is updated (right). This process is repeated until the system reaches steady state.

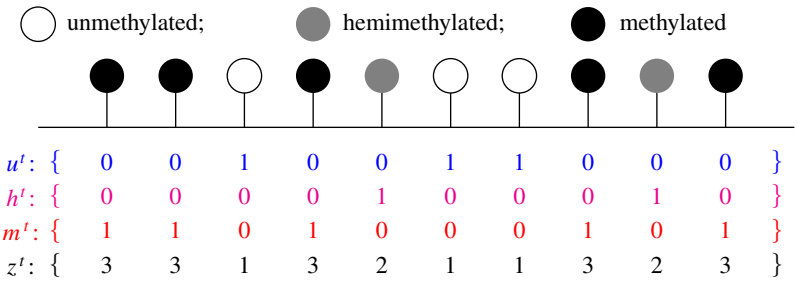

Figure 4. Construction of the sequences u^t, h^t, m^t and z^t. For simplicity, only 10 CpGs are shown for a single timepoint, t. A vector u^t is created, where $u_i^t = 1$ if CpG i is unmethylated at time t and $u_i^t = 0$ otherwise. Vectors h^t and m^t are constructed similarly. Finally, a vector z^t is created via $z^t = u^t + 2h^t + 3m^t$.

The steady-state mean u, h and m levels, μ_{us}, μ_{hs}, μ_{ms}, are obtained via

$$\mu_{us} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} u_i^t, \quad \mu_{hs} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} h_i^t \quad \text{and}$$
$$\mu_{ms} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} m_i^t. \tag{2.2}$$

Throughout, the subscript 's' denotes steady-state statistics.

For each t, we also calculate the mean and variance over z^t (μ_z^t, $\sigma^2(z^t)$, respectively) and average over all $t \in \{1, \ldots, T\}$ to obtain an overall steady-state mean and variance, μ_{zs} and $\sigma^2(z_s)$, which are given by

$$\mu_{zs} = \frac{1}{T}\sum_{t=1}^{T}\mu_z^t = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} z_i^t \tag{2.3}$$

and

$$\sigma^2(z_s) = \frac{1}{T}\sum_{t=1}^{T}\sigma^2(z^t) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{(z_i^t - \mu_z^t)^2}{N-1}. \tag{2.4}$$

We then define sequences v^t and w^t by

$$v^t = \{z_1^t, z_2^t, \ldots, z_{N-1}^t, z_N^t\} \quad \text{and}$$
$$w^t = \{z_2^t, z_3^t, \ldots, z_N^t, z_1^t\}, \quad t \in \{1, \ldots, T\}. \tag{2.5}$$

For each $t \in \{1, \ldots, T\}$, the sequence v^t is identical to z^t, while w^t is a shifted version of z^t (i.e. $w_i^t = z_{i+1}^t$ for $i = \{1, \ldots, N-1\}$ and $w_N^t = z_1^t$). For each $i = \{1, \ldots, N\}$, comparing v_i^t and w_i^t provides information regarding the methylation state of two neighbouring CpGs at time $t \in \{1, \ldots, T\}$. We calculate the covariance between v^t and w^t for each t, Covar(v^t, w^t), averaging over $t \in \{1, \ldots, T\}$ to obtain an overall steady-state covariance between neighbouring sites:

$$\text{Covar}(z_s) := \text{Covar}(v_s, w_s) = \frac{1}{T}\sum_{t=1}^{T}\text{Covar}(v^t, w^t)$$
$$= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(v_i^t - \mu_v^t)(w_i^t - \mu_w^t), \tag{2.6}$$

where $\mu_v^t = \mu_w^t = \mu_z^t$. Finally, the steady-state correlation between neighbouring sites, $\rho(z_s) := \rho(v_s, w_s)$, is calculated

via

$$\rho(z_s) := \rho(v_s, w_s) = \frac{\mathrm{Covar}(v_s, w_s)}{\sqrt{\sigma^2(v_s)\sigma^2(w_s)}} = \frac{\mathrm{Covar}(v_s, w_s)}{\sigma^2(z_s)}. \quad (2.7)$$

3. Mean-field models for DNA methylation maintenance

For large CpG systems, the CME describing the nearest-neighbour collaborative model cannot be easily solved and stochastic simulations are computationally expensive. In contrast, it is often the case that models simplified using the MF approximation can be computationally solved in a time-efficient manner and we aim to test whether they accurately approximate the nearest-neighbour collaborative model for DNA methylation. To this end, we construct three MF models (see §§3.1–3.3). These models consider an infinite system of CpGs and so, by design, their ability to accurately describe a genomic region increases with the size of the region. In these models, nearest-neighbour interactions are approximated by considering the mean state of the system. In the first model, nearest-neighbour interactions are entirely approximated by considering the probability that two states are adjacent (one-site MF model). The second model describes distinct pairs of CpGs (distinct pairs MF model). Interactions occurring within a pair are directly accounted for and other nearest-neighbour interactions are approximated by considering the probability that two paired states are adjacent. In the third model, we consider overlapping pairs of CpGs (overlapping pairs MF model). Interactions occurring within a pair are again directly accounted for, but now other nearest-neighbour interactions are approximated by considering the probability that two paired states overlap. The remainder of this section is devoted to mathematical descriptions of these models.

Note that while the computational cost of simulations increases with sequence length, numerical solutions of the MF models discussed in §§3.1–3.3 are independent of sequence length. Hence these models describe arbitrarily large genomic regions.

3.1. One-site mean-field model

We define the proportion of sites in the u, h, m states to be the mean u, h, m levels, μ_u, μ_h, μ_m, respectively. Here we construct a one-site MF model, where changes in the system are influenced by μ_u, μ_h, μ_m, rather than nearest-neighbour interactions; see figure 5.

Consider the reaction $u + h \xrightarrow{k_5} h + h$ in figure 1. Since the h mediator is unchanged by the reaction, we can write this as an effective first-order reaction $u \xrightarrow{2k_5\mu_h} h$, where the h mediator is absorbed into the effective reaction rate by making it proportional to μ_h. The factor of two accounts for the h mediator being on either side of the u target. Similarly, $u + m \xrightarrow{k_6} h + m$ can be written as $u \xrightarrow{2k_6\mu_m} h$. Thus the $u \xrightarrow{k_1} h$, $u + h \xrightarrow{k_5} h + h$, $u + m \xrightarrow{k_6} h + m$ reactions in figure 1 can be written as a single effective first-order reaction $u \xrightarrow{r} h$, with $r = k_1 + 2k_5\mu_h + 2k_6\mu_m$. We thus write the system in figure 1 as the effective first-order system

$$m \xrightarrow{a_1} h, \quad h \xrightarrow{a_2} m, \quad u \xrightarrow{a_3} h, \quad h \xrightarrow{a_4} u, \quad (3.1)$$

where

$$\begin{aligned} a_1 &:= a_1(\mu_u, \mu_h) = k_3 + 2k_9\mu_h + 2k_{10}\mu_u, \\ a_2 &:= a_2(\mu_h, \mu_m) = k_2 + 4k_7\mu_h + 2k_8\mu_m, \\ a_3 &:= a_3(\mu_h, \mu_m) = k_1 + 2k_5\mu_h + 2k_6\mu_m \\ \text{and} \quad a_4 &:= a_4(\mu_u, \mu_h) = k_4 + 4k_{11}\mu_h + 2k_{12}\mu_u, \end{aligned}$$

and $\mu_u + \mu_h + \mu_m = 1$. The k_7 and k_{11} terms have an additional factor of two since their associated reactions involve two h reactants, and either of these can change state during the reaction.

Let L_u, L_h, L_m be the 'level' (proportion) of u, h, m at a single CpG, respectively. A CpG can only be in one state at any time and so the state of a CpG at position j can be described by a state vector of the form

$$\begin{aligned} &(L_u, L_h, L_m)_j = (1, 0, 0) \quad \text{or} \quad (L_u, L_h, L_m)_j = (0, 1, 0) \\ &\quad \text{or} \quad (L_u, L_h, L_m)_j = (0, 0, 1) \end{aligned} \quad . \quad (3.2)$$

The probabilities associated with these three states sum to one and so we only need to consider the probabilities associated with $(L_u, L_h, L_m)_j = (1, 0, 0)$, and $(L_u, L_h, L_m)_j = (0, 1, 0)$. Using (3.1) we construct the CME describing the probability that a site is in the (1, 0, 0) or (0, 1, 0) state:

$$\left.\begin{aligned} &\frac{\mathrm{d}P(1,0,0)}{\mathrm{d}t} = -a_3P(1,0,0) + a_4P(0,1,0) \\ \text{and} \quad &\frac{\mathrm{d}P(0,1,0)}{\mathrm{d}t} = a_1\big(1 - P(1,0,0) - P(0,1,0)\big) - a_2P(0,1,0) \\ &\qquad\qquad + a_3P(1,0,0) - a_4P(0,1,0). \end{aligned}\right\} \quad (3.3)$$

We can also obtain moment equations (see appendix C of [35]) for the statistics of L_u, L_h, L_m. The mean values are $\mu_u = \langle L_u \rangle$, $\mu_h = \langle L_h \rangle$, $\mu_m = \langle L_m \rangle$, where the angled brackets denote the expected value. Due to the conservation law $L_m = 1 - L_u - L_h$, we again need only consider equations for u and h. From equation (C 1) in [35] with

$$\begin{aligned} &R = 4, \quad n = \{L_u, L_h\}, \quad S = \begin{bmatrix} 0 & 0 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \\ &\hat{f} = \{a_1L_m, a_2L_h, a_3L_u, a_4L_h\}, \end{aligned} \quad (3.4)$$

the means are described by the equations

$$\begin{aligned} &\frac{\mathrm{d}\mu_u}{\mathrm{d}t} = -a_3\mu_u + a_4\mu_h \quad \text{and} \\ &\frac{\mathrm{d}\mu_h}{\mathrm{d}t} = a_1(1 - \mu_u - \mu_h) - a_2\mu_h + a_3\mu_u - a_4\mu_h. \end{aligned} \quad (3.5)$$

Note that these equations are identical to (3.3) with $P(1, 0, 0) = \mu_u$ and $P(0, 1, 0) = \mu_h$. Setting

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}t} = \frac{\mathrm{d}\mu_h}{\mathrm{d}t} = 0,$$

leads to implicit equations for the steady-state means,

$$\begin{aligned} \mu_{us} &= \frac{a_{1s}a_{4s}}{a_{1s}a_{3s} + a_{2s}a_{3s} + a_{1s}a_{4s}} \quad \text{and} \\ \mu_{hs} &= \frac{a_{1s}a_{3s}}{a_{1s}a_{3s} + a_{2s}a_{3s} + a_{1s}a_{4s}}, \end{aligned} \quad (3.6)$$

where $a_{1s} = a_1(\mu_{us}, \mu_{hs})$, $a_{2s} = a_2(\mu_{hs}, 1 - \mu_{us} - \mu_{hs})$, $a_{3s} = a_3(\mu_{hs}, 1 - \mu_{us} - \mu_{hs})$, $a_{4s} = a_4(\mu_{us}, \mu_{hs})$. Since equation (3.6) is independent of a, the means depend only on x and y. For fixed parameters, we can solve equation (3.6) numerically using the Mathematica function NSolve to obtain values for μ_{us} and μ_{hs}.

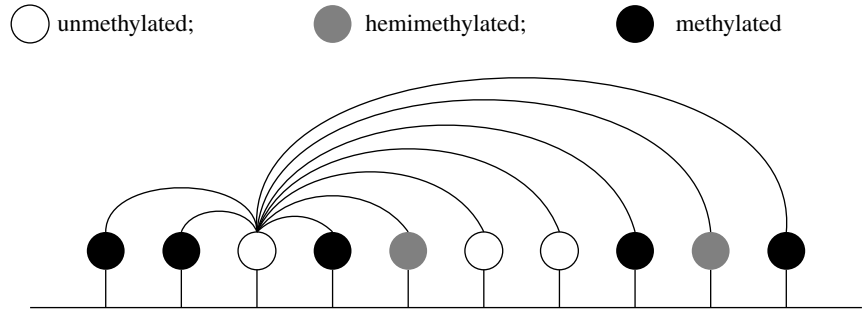

Figure 5. Schematic of the one-site MF model. CpGs are influenced by the mean of the system rather than nearest-neighbour interactions.

Using equation (C2) in [35] with (3.4) the second-moment equations are given by

$$\left.\begin{aligned}
\tfrac{\mathrm{d}\langle L_uL_u\rangle}{\mathrm{d}t} &= -2a_3\langle L_uL_u\rangle + 2a_4\langle L_uL_h\rangle + a_4\mu_h + a_3\mu_u,\\
\tfrac{\mathrm{d}\langle L_uL_h\rangle}{\mathrm{d}t} &= -a_1\Big(\langle L_uL_u\rangle + \langle L_uL_h\rangle\Big) - a_2\langle L_uL_h\rangle\\
&\quad + a_3\Big(\langle L_uL_u\rangle - \langle L_uL_h\rangle\Big)\\
&\quad - a_4\Big(\langle L_uL_h\rangle - \langle L_hL_h\rangle\Big) + a_1\mu_u - a_3\mu_u - a_4\mu_h\\
\text{and } \tfrac{\mathrm{d}\langle L_hL_h\rangle}{\mathrm{d}t} &= -2a_1\Big(\langle L_uL_h\rangle + \langle L_hL_h\rangle\Big)\\
&\quad -2a_2\langle L_hL_h\rangle + 2a_3\langle L_uL_h\rangle - 2a_4\langle L_hL_h\rangle\\
&\quad + a_1(1-\mu_u+\mu_h) + a_2\mu_h + a_3\mu_u + a_4\mu_h.
\end{aligned}\right\} \tag{3.7}$$

From equation (3.2), we expect $L_uL_u = L_u$, $L_hL_h = L_h$ and $L_uL_h = 0$ at any CpG. Solving equation (3.7) in steady state leads to

$$\left.\begin{aligned}
\langle L_uL_u\rangle_s &= \langle L_u\rangle_s = \mu_{us}, \quad \langle L_uL_h\rangle_s = 0,\\
\langle L_hL_h\rangle_s &= \langle L_h\rangle_s = \mu_{hs}.
\end{aligned}\right\} \tag{3.8}$$

Variances, $\sigma^2(L_{us})$ and $\sigma^2(L_{hs})$, can then be obtained via

$$\left.\begin{aligned}
\sigma^2(L_{us}) &= \langle L_uL_u\rangle_s - \mu_{us}^2 = \mu_{us} - \mu_{us}^2\\
\text{and}\quad \sigma^2(L_{hs}) &= \langle L_hL_h\rangle_s - \mu_{hs}^2 = \mu_{hs} - \mu_{hs}^2,
\end{aligned}\right\} \tag{3.9}$$

along with the covariance $\mathrm{Covar}(L_{us}, L_{hs})$, which is given by

$$\mathrm{Covar}(L_{us}, L_{hs}) = \langle L_uL_h\rangle_s - \mu_{us}\mu_{hs} = -\mu_{us}\mu_{hs}. \tag{3.10}$$

The mean, variance and covariances associated with the m state can now be obtained using

$$\left.\begin{aligned}
\mu_{ms} &= 1 - \mu_{us} - \mu_{hs},\\
\sigma^2(L_{ms}) &= \sigma^2(1 - L_{us} - L_{hs}) = \sigma^2(L_{us}) + \sigma^2(L_{hs})\\
&\quad + 2\mathrm{Covar}(L_{us}, L_{hs}),\\
\mathrm{Covar}(L_{us}, L_{ms}) &= -\sigma^2(L_{us}) - \mathrm{Covar}(L_{us}, L_{hs})\\
\text{and}\quad \mathrm{Covar}(L_{hs}, L_{ms}) &= -\mathrm{Covar}(L_{us}, L_{hs}) - \sigma^2(L_{hs}).
\end{aligned}\right\} \tag{3.11}$$

We calculate the steady-state mean and variance, μ_{zs} and $\sigma^2(zs)$, associated with the variable $z = L_u + 2L_h + 3L_m$ via

$$\mu_{zs} = \mu_{us} + 2\mu_{hs} + 3\mu_{ms}$$

and

$$\sigma^2(z_s) = \sigma^2(L_{us}) + 4\sigma^2(L_{hs}) + 9\sigma^2(L_{ms}) + 2\Big(2\mathrm{Covar}(L_{us}, L_{hs}) + 3\mathrm{Covar}(L_{us}, L_{ms}) + 6\mathrm{Covar}(L_{hs}, L_{ms})\Big).$$

Note that the superscript t was only used in §2.2 to differentiate between samples at different timepoints. Here, we simply have a single z. Since no spatial information is obtained from the one-site MF model, the covariance and correlation between neighbouring sites cannot be extracted.

3.2. Distinct pairs mean-field model

We next construct a two-site MF model, where we consider 'clusters' of two adjacent CpGs. Such cluster MF models have been successfully used to study vehicular traffic and driven-diffusive gas models [36,37]. We define the mean level (proportion) of pairs in the six possible states,

$$mm, \quad uu, \quad hh, \quad um\ (:=mu), \quad hm\ (:=mh), \quad uh\ (:=hu), \tag{3.12}$$

to be μ_1, μ_2, μ_3, μ_4, μ_5, μ_6, respectively. Here, μ_4 is the proportion of pairs containing u and m, irrespective of order. Similarly, μ_5 is the proportion of pairs containing h and m, and μ_6 is the proportion of pairs containing u and h, irrespective of order.

In the distinct pairs MF model (DPMF model), CpGs within a pair are allowed to interact directly, preserving some nearest-neighbour interactions. The influence of the nearest-neighbour CpGs flanking the pair is then approximated by considering the probabilities that an adjacent pair is in each of the six possible states; see figure 6. Here, each CpG belongs to only one pair and each pair of sites is a single reactant. As with the one-site model, we consider an effective first-order reaction system, given by

$$\left.\begin{aligned}
&mm \xrightarrow{\hat{a}_1} hm, \quad uu \xrightarrow{\hat{a}_2} uh,\\
&hh \xrightarrow{\hat{a}_3} uh, \quad hh \xrightarrow{\hat{a}_4} hm, \quad um \xrightarrow{\hat{a}_5} hm, \quad um \xrightarrow{\hat{a}_6} uh,\\
&hm \xrightarrow{\hat{a}_7} um, \quad hm \xrightarrow{\hat{a}_8} mm, \quad hm \xrightarrow{\hat{a}_9} hh,\\
&uh \xrightarrow{\hat{a}_{10}} hh, \quad uh \xrightarrow{\hat{a}_{11}} uu, \quad uh \xrightarrow{\hat{a}_{12}} um,
\end{aligned}\right\} \tag{3.13}$$

where the effective rates are given by

$$\left.\begin{aligned}
\hat{a}_1 &= 2k_3 + k_9(2\mu_3 + \mu_5 + \mu_6) + k_{10}(2\mu_2 + \mu_4 + \mu_6),\\
\hat{a}_2 &= 2k_1 + k_5(2\mu_3 + \mu_5 + \mu_6) + k_6(2\mu_1 + \mu_4 + \mu_5),\\
\hat{a}_3 &= 2k_4 + 2k_{11} + k_{11}(4\mu_3 + \mu_5 + \mu_6)\\
&\quad + k_{12}(2\mu_2 + \mu_4 + \mu_6),\\
\hat{a}_4 &= 2k_2 + 2k_7 + k_7(4\mu_3 + \mu_5 + \mu_6) + k_8(2\mu_1 + \mu_4 + \mu_5),\\
\hat{a}_5 &= k_1 + k_6 + k_5\Big(\mu_3 + \tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_6\Big(\mu_1 + \tfrac{\mu_4}{2} + \tfrac{\mu_5}{2}\Big),\\
\hat{a}_6 &= k_3 + k_{10} + k_9\Big(\mu_3 + \tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_{10}\Big(\mu_2 + \tfrac{\mu_4}{2} + \tfrac{\mu_6}{2}\Big),\\
\hat{a}_7 &= k_4 + k_{11}\Big(\mu_3 + 2\tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_{12}\Big(\mu_2 + \tfrac{\mu_4}{2} + \tfrac{\mu_6}{2}\Big),\\
\hat{a}_8 &= k_2 + k_8 + k_7\Big(\mu_3 + 2\tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_8\Big(\mu_1 + \tfrac{\mu_4}{2} + \tfrac{\mu_5}{2}\Big),\\
\hat{a}_9 &= k_3 + k_9 + k_9\Big(\mu_3 + \tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_{10}\Big(\mu_2 + \tfrac{\mu_4}{2} + \tfrac{\mu_6}{2}\Big),\\
\hat{a}_{10} &= k_1 + k_5 + k_5\Big(\mu_3 + \tfrac{\mu_5}{2} + \tfrac{\mu_6}{2}\Big) + k_6\Big(\mu_1 + \tfrac{\mu_4}{2} + \tfrac{\mu_5}{2}\Big),\\
\hat{a}_{11} &= k_4 + k_{12} + k_{11}\Big(\mu_3 + \tfrac{\mu_5}{2} + 2\tfrac{\mu_6}{2}\Big) + k_{12}\Big(\mu_2 + \tfrac{\mu_4}{2} + \tfrac{\mu_6}{2}\Big)\\
\text{and}\quad \hat{a}_{12} &= k_2 + k_7\Big(\mu_3 + \tfrac{\mu_5}{2} + 2\tfrac{\mu_6}{2}\Big) + k_8\Big(\mu_1 + \tfrac{\mu_4}{2} + \tfrac{\mu_5}{2}\Big),
\end{aligned}\right\} \tag{3.14}$$

and $\sum_{i=1}^{6}\mu_i = 1$. We describe the construction of $\hat{a}_1$ in appendix A. Essentially, $\hat{a}_1$–$\hat{a}_{12}$ are constructed by considering all possible ways that each reaction in (3.13) can occur via a reaction from figure 1 taking place. Such reactions can occur within the reactant pair or can take place between a site within the pair and a site from an adjacent pair. Terms that are associated with the k_7 and k_{11} reaction rates and involve interactions between two *hh*, two *hm* or two *uh* pairs have an additional factor of two since either pair can change state during these reactions. While we can, in principle, calculate the distribution of pairs in (3.13) [38], we restrict our attention to obtaining moments of the system.

Let $L_1, L_2, L_3, L_4, L_5, L_6$ be the level (proportion) of each of the paired states at a single pair of CpGs. Since a pair of CpGs can only be in one state at any time, the state of a single pair of CpGs can be described by one of six vectors S_j, $j = 1, \ldots, 6$, where

$$S_j = (L_1, L_2, L_3, L_4, L_5, L_6) \quad \text{such that } L_i = 1, \text{ for } i = j \text{ and } L_i = 0 \text{ for } i \neq j. \tag{3.15}$$

The u, h, m levels within a pair of CpGs, $\hat{L}_u$, $\hat{L}_h$, $\hat{L}_m$, are then given by

$$\hat{L}_u = L_2 + \frac{L_4}{2} + \frac{L_6}{2}, \quad \hat{L}_h = L_3 + \frac{L_5}{2} + \frac{L_6}{2} \quad \text{and} \quad \hat{L}_m = L_1 + \frac{L_4}{2} + \frac{L_5}{2}. \tag{3.16}$$

Using (3.13), we construct the CME for the system along with the first and second moment equations for L_i, $i = \{1, \ldots, 6\}$; see appendix C. The first moment equations describe $\mu_i = \langle L_i \rangle$, the mean values of L_i for $i = \{1, \ldots, 6\}$. For fixed parameters, solving these equations numerically in steady state gives the steady-state means, μ_{is}, $i = \{1, \ldots, 6\}$. Note that these means are independent of a.

From the second moment equations, we obtain the steady-state expected values of L_iL_j, $\langle L_iL_j \rangle_s$, for $i, j \in \{1, 2, \ldots, 6\}$. As expected from equation (3.15), $\langle L_iL_i \rangle_s = \langle L_i \rangle_s = \mu_{is}$, $\langle L_iL_j \rangle_s = 0$ for all $i \neq j$, $i, j \in \{1, 2, \ldots, 6\}$. Variances and covariances are then obtained using

$$\sigma^2(L_{is}) = \langle L_iL_i \rangle_s - \mu_{is}^2 = \mu_{is} - \mu_{is}^2, \quad i = 1, 2, \ldots, 6$$

and

$$\mathrm{Covar}(L_{is}, L_{js}) = \langle L_iL_j \rangle_s - \mu_{is}\mu_{js} = -\mu_{is}\mu_{js}, \quad i \neq j, i, j = 1, 2, \ldots, 6.$$

Once again, these are independent of the parameter a.

We now have statistics for the paired states in (3.12). The steady-state means for the u, h, m levels in a pair are then given by

$$\mu_{us} = \mu_{2s} + \frac{\mu_{4s}}{2} + \frac{\mu_{6s}}{2}, \quad \mu_{hs} = \mu_{3s} + \frac{\mu_{5s}}{2} + \frac{\mu_{6s}}{2} \quad \text{and} \quad \mu_{ms} = \mu_{1s} + \frac{\mu_{4s}}{2} + \frac{\mu_{5s}}{2}. \tag{3.17}$$

The pair-to-pair variance in u level is given by

$$\sigma^2(\hat{L}_{us}) = \sigma^2(L_{2s}) + \frac{1}{4}\sigma^2(L_{4s}) + \frac{1}{4}\sigma^2(L_{6s}) + 2\left(\frac{1}{2}\mathrm{Covar}(L_{2s}, L_{4s}) + \frac{1}{2}\mathrm{Covar}(L_{2s}, L_{6s}) + \frac{1}{4}\mathrm{Covar}(L_{4s}, L_{6s})\right),$$

and similarly for the variances associated with h and m, $\sigma^2(\hat{L}_{hs})$ and $\sigma^2(\hat{L}_{ms})$. Covariances are given by

$$\begin{aligned}\mathrm{Covar}(\hat{L}_{us}, \hat{L}_{ms}) = {} & \mathrm{Covar}(L_{1s}, L_{2s}) + \frac{1}{2}\mathrm{Covar}(L_{2s}, L_{4s}) \\ & + \frac{1}{2}\mathrm{Covar}(L_{2s}, L_{5s}) + \frac{1}{2}\mathrm{Covar}(L_{1s}, L_{4s}) \\ & + \frac{1}{4}\sigma^2(L_{4s}) + \frac{1}{4}\mathrm{Covar}(L_{4s}, L_{5s}) \\ & + \frac{1}{2}\mathrm{Covar}(L_{1s}, L_{6s}) + \frac{1}{4}\mathrm{Covar}(L_{4s}, L_{6s}) \\ & + \frac{1}{4}\mathrm{Covar}(L_{5s}, L_{6s})\end{aligned}$$

and analogously for $\mathrm{Covar}(\hat{L}_{hs}, \hat{L}_{ms})$, $\mathrm{Covar}(\hat{L}_{us}, \hat{L}_{hs})$.

Note that the statistics obtained so far relate to u, h, m levels within a pair of CpGs. The mean level of a state within a pair is the same as the mean level of the state at each site. However, this is not the case for higher moments. For example, $\sigma^2(\hat{L}_{us})$, $\sigma^2(\hat{L}_{hs})$, $\sigma^2(\hat{L}_{ms})$ are pair-to-pair variances, rather than site-to-site variances.

We aim to obtain statistics relating to $z = L_u + 2L_h + 3L_m$, where L_u, L_h, L_m are the single-site u, h, m levels. We define $\hat{z} = \hat{L}_u + 2\hat{L}_h + 3\hat{L}_m$, noting that $\hat{z}$ contains information regarding pairs of CpGs. Essentially,

$$\hat{z} = \frac{v + w}{2}, \tag{3.18}$$

where v and w are as in equation (2.5) and we again do not require the superscript t.

The steady-state means of z, v, w and $\hat{z}$ coincide and

$$\mu_{zs} = \mu_{\hat{z}s} = \mu_{vs} = \mu_{ws} = \mu_{us} + 2\mu_{hs} + 3\mu_{ms}.$$

Also, $vw = L_2 + 2L_6 + 3L_4 + 4L_3 + 6L_5 + 9L_1$ and so the steady-state expected value of vw is

$$\langle vw \rangle_s = \mu_{2s} + 2\mu_{6s} + 3\mu_{4s} + 4\mu_{3s} + 6\mu_{5s} + 9\mu_{1s}.$$

From this, we calculate the steady-state covariance between neighbouring sites as

$$\mathrm{Covar}(z_s) = \mathrm{Covar}(v_s, w_s) = \langle vw \rangle_s - \langle v \rangle_s \langle w \rangle_s = \langle vw \rangle_s - \mu_{zs}^2.$$

The steady-state variance associated with $\hat{z}$ can be calculated via,

$$\begin{aligned}\sigma^2(\hat{z}_s) &= \sigma^2(\hat{L}_{us} + 2\hat{L}_{hs} + 3\hat{L}_{ms}) \\ &= \sigma^2(\hat{L}_{us}) + 4\sigma^2(\hat{L}_{hs}) + 9\sigma^2(\hat{L}_{ms}) \\ &\quad + 2\Big(2\mathrm{Covar}(\hat{L}_{us}, \hat{L}_{hs}) + 3\mathrm{Covar}(\hat{L}_{us}, \hat{L}_{ms}) \\ &\quad + 6\mathrm{Covar}(\hat{L}_{hs}, \hat{L}_{ms})\Big).\end{aligned}$$

However, $\sigma^2(\hat{z}_s)$ is the pair-to-pair variance. Using $\sigma^2(z_s) = \sigma^2(v_s) = \sigma^2(w_s)$, we obtain

$$\begin{aligned}\sigma^2(\hat{z}_s) &= \sigma^2\left(\frac{v_s + w_s}{2}\right) = \frac{1}{4}\left(\sigma^2(v_s) + \sigma^2(w_s) + 2\mathrm{Covar}(v_s, w_s)\right) \\ &= \frac{1}{2}\left(\sigma^2(z_s) + \mathrm{Covar}(v_s, w_s)\right),\end{aligned}$$

leading to the steady-state site-to-site variance

$$\sigma^2(z_s) = 2\sigma^2(\hat{z}_s) - \mathrm{Covar}(v_s, w_s).$$

The steady-state correlation between neighbouring pairs is obtained via

$$\rho(z_s) = \rho(v_s, w_s) = \frac{\mathrm{Covar}(v_s, w_s)}{\sqrt{\sigma^2(v_s)\ \sigma^2(w_s)}} = \frac{\mathrm{Covar}(v_s, w_s)}{\sigma^2(z_s)}.$$

To summarize, the statistics of primary interest from our calculations are: the means μ_{us}, μ_{hs}, μ_{ms}, μ_{zs}, the variance $\sigma^2(z_s)$, the covariance $\mathrm{Covar}(z_s)$ and the correlation $\rho(z_s)$.

3.3. Overlapping pairs mean-field model

Similarly to the DPMF model, there are also six possible states for a pair of CpGs in the overlappling pairs MF model (OPMF model); see (3.12). The OPMF model also incorporates direct interactions within a pair. However, each CpG now belongs to two pairs, one with its left-hand neighbour and one with its right-hand neighbour, leading to a system of overlapping pairs. The influence of CpGs flanking a pair is now approximated by considering the conditional probability that the pair overlaps with another pair of a certain state; see figure 7.

As before, we consider an effective first-order reaction system, given by

$$\left.\begin{aligned} &mm \xrightarrow{\tilde{a}_1} hm, \quad uu \xrightarrow{\tilde{a}_2} uh, \quad hh \xrightarrow{\tilde{a}_3} uh, \quad hh \xrightarrow{\tilde{a}_4} hm,\\ &um \xrightarrow{\tilde{a}_5} hm, \quad um \xrightarrow{\tilde{a}_6} uh,\\ &hm \xrightarrow{\tilde{a}_7} um, \quad hm \xrightarrow{\tilde{a}_8} mm,\\ &hm \xrightarrow{\tilde{a}_9} hh, \quad uh \xrightarrow{\tilde{a}_{10}} hh, \quad uh \xrightarrow{\tilde{a}_{11}} uu, \quad uh \xrightarrow{\tilde{a}_{12}} um, \end{aligned}\right\} \tag{3.19}$$

where the effective rates are given by

$$\left.\begin{aligned} \tilde{a}_1 &= 2k_3 + 2k_9\left(\tfrac{\mu_5/2}{\mu_1+\mu_4/2+\mu_5/2}\right) + 2k_{10}\left(\tfrac{\mu_4/2}{\mu_1+\mu_4/2+\mu_5/2}\right),\\ \tilde{a}_2 &= 2k_1 + 2k_5\left(\tfrac{\mu_6/2}{\mu_2+\mu_4/2+\mu_6/2}\right) + 2k_6\left(\tfrac{\mu_4/2}{\mu_2+\mu_4/2+\mu_6/2}\right),\\ \tilde{a}_3 &= 2k_4 + 2k_{11} + 2k_{11}\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + 2k_{12}\left(\tfrac{\mu_6/2}{\mu_3+\mu_5/2+\mu_6/2}\right),\\ \tilde{a}_4 &= 2k_2 + 2k_7 + 2k_7\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + 2k_8\left(\tfrac{\mu_5/2}{\mu_3+\mu_5/2+\mu_6/2}\right),\\ \tilde{a}_5 &= k_1 + k_6 + k_5\left(\tfrac{\mu_6/2}{\mu_2+\mu_4/2+\mu_6/2}\right) + k_6\left(\tfrac{\mu_4/2}{\mu_2+\mu_4/2+\mu_6/2}\right),\\ \tilde{a}_6 &= k_3 + k_{10} + k_9\left(\tfrac{\mu_5/2}{\mu_1+\mu_4/2+\mu_5/2}\right) + k_{10}\left(\tfrac{\mu_4/2}{\mu_1+\mu_4/2+\mu_5/2}\right),\\ \tilde{a}_7 &= k_4 + k_{11}\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + k_{12}\left(\tfrac{\mu_6/2}{\mu_3+\mu_5/2+\mu_6/2}\right),\\ \tilde{a}_8 &= k_2 + k_8 + k_7\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + k_8\left(\tfrac{\mu_5/2}{\mu_3+\mu_5/2+\mu_6/2}\right),\\ \tilde{a}_9 &= k_3 + k_9 + k_9\left(\tfrac{\mu_5/2}{\mu_1+\mu_4/2+\mu_5/2}\right) + k_{10}\left(\tfrac{\mu_4/2}{\mu_1+\mu_4/2+\mu_5/2}\right),\\ \tilde{a}_{10} &= k_1 + k_5 + k_5\left(\tfrac{\mu_6/2}{\mu_2+\mu_4/2+\mu_6/2}\right) + k_6\left(\tfrac{\mu_4/2}{\mu_2+\mu_4/2+\mu_6/2}\right),\\ \tilde{a}_{11} &= k_4 + k_{12} + k_{11}\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + k_{12}\left(\tfrac{\mu_6/2}{\mu_3+\mu_5/2+\mu_6/2}\right)\\ \text{and}\ \ \tilde{a}_{12} &= k_2 + k_7\left(\tfrac{\mu_3}{\mu_3+\mu_5/2+\mu_6/2}\right) + k_8\left(\tfrac{\mu_5/2}{\mu_3+\mu_5/2+\mu_6/2}\right). \end{aligned}\right\} \tag{3.20}$$

We detail the construction of $\hat{a}_1$ in appendix B.

An identical approach to that in §3.2 leads to the CME and first and second moment equations for the system, see appendix C, and from these we obtain means, variances and covariances associated with the paired states and with the pair-to-pair u, h, m levels. Again, these depend only on x and y. The statistics associated with z are obtained as for the DPMF model; see §3.2.

4. Model comparison and parameter inference

To test whether MF models are capable of modelling large-scale methylation patterns, we now compare model predictions to synthetic data generated using nearest-neighbour collaborative simulations. The statistical properties obtained from our models are independent of a and so we fix $a = 0.2$. Since demethylation dominates when $y < 1$ and methylation dominates when $y > 1$, we hypothesize that a sharp change in the behaviour of the system may occur at $y = 1$. To capture this potential transition for different collaborativity strengths, we consider $x = \{0.1, 1, 5, 50\}$, $y = \{0.1, 0.2, \ldots, 2\}$. Since we approximate the nearest-neighbour collaborative system by MF models, which consider an infinite system of CpGs, finite-size effects cause discrepancies between simulations and MF model predictions as x increases. We counteract this by increasing the number of simulated sites and so simulate $N = 200$ CpGs when $x = \{0.1, 1, 5\}$ and $N = 500$ CpGs when $x = 50$.

For each parameter set, we simulate $n = 10$ replicate datasets using the Gillespie algorithm (see §2.2). In particular, in each simulation, we take steady-state samples at $T = 10^6/200 = 5000$ timepoints when $x = \{0.1, 1, 5\}$ and at $T = 10^6/500 = 2000$ timepoints when $x = 50$. Each of the replicate datasets therefore contains 10^6 states, from which we calculate the statistics of interest (see §2.3). We then calculate the means and standard errors over the 10 datasets to obtain overall summary statistics.

4.1. Mean-field models capture steady-state methylation levels

We first compare the mean u, h, m levels (μ_{us}, μ_{hs}, μ_{ms}) from the MF models to those from the simulations (figure 8). Considering the simulated data first, we observe that u and m dominate when $y < 1$ or $y > 1$, respectively. h is an intermediate state between u and m and peaks at $y = 1$, where there is also a sharp transition between u- and m-dominant states.

While all models capture the qualitative behaviour of the means as x and y are varied, we observe that predictions of methylation levels from the OPMF model are closest to those observed in the simulated data (figure 8). All three models predict the mean u and m levels reasonably well; however only the OPMF model accurately predicts the mean h level for large x and for y close to one. All predictions from the OPMF model are within the error of the simulated data and it successfully captures the transition observed at $y = 1$ for all x considered. Conversely, the predictions of the other models deviate from the simulations at the transition point when x is large. The one-site MF model deviates to the greatest extent for 87% of the parameter sets and, for each mean, the average percentage error across y is higher for the one-site MF model compared to the other models for all x except μ_{hs} when $x = 5$ (see appendix D, table 3). This suggests that the one-site model has the worst predictive power and we exclude it from further analysis.

4.2. The overlapping pairs mean-field model accurately predicts associations between neighbouring sites

To test whether MF models can predict associations between neighbouring CpGs, we consider $z = \{z_1, z_2, \ldots, z_N\}$, where $z_i = \{1, 2, 3\}$ if CpG i is in the u, h, m state, respectively.

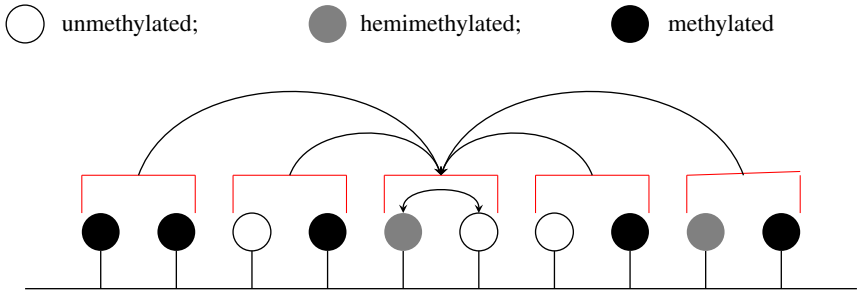

Figure 6. Schematic of the distinct pairs MF model. The two CpGs within a pair can interact directly with each other and the pair is also influenced by the mean state of pairs in the system. In the figure, the *uh* pair can change state due to interactions between the *u* and *h* within the pair, and due to the mean state of pairs in the system.

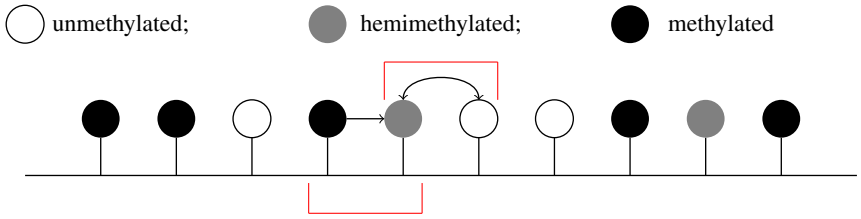

Figure 7. Schematic of the overlapping pairs MF model. A pair of CpGs interact directly with each other and the effect of CpGs flanking the pair is approximated by considering the conditional probabilities that a flanking site is in the *u*, *h* or *m* state, given the state of its neighbour within the pair.

From z, we calculate the mean and variance associated with the methylation state, and the covariance and correlation in methylation state between neighbouring sites. In the simulated data (figure 9), we again observe a transition in these statistics when the methylation and demethylation strengths are equal ($y = 1$). Our results counterintuitively suggest that neighbouring sites are most correlated here (the peak $\rho(z_s)$ occurs when $y = 1$).

To gain insight into this observation, we examine the patterns that evolve in the stochastic simulations for $x = 50$ (figure 10). When y is small, large u clusters form and we intuitively expect neighbouring sites to be highly correlated. However, these large clusters are interspersed with infrequent, isolated occurrences of h and m which have low correlations with their neighbours. Moreover, ten m (or h) sites appearing in isolation will result in smaller u clusters than the ten sites appearing as a single cluster. The overall effect is to reduce the correlation when y is small. A similar rationale explains the low correlation when y is large. u and m cluster sizes are most similar when methylation and demethylation are equally strong, resulting in u and m sites correlating equally with their neighbours and the overall correlation peaking.

We next consider predictions of these statistics from the MF models. Predictions from the OPMF model lie within the error observed in the simulated data for all parameters considered (figure 9). Conversely, predicted statistics from the DPMF model show large deviations from the corresponding simulated statistics when x is large and y is close to one demonstrating that it has lower predictive power. The OPMF model again shows superior performance over the DPMF model when the percentage errors are considered. For each z statistic in figure 9, the average percentage error across y is higher for the DPMF model compared to the OPMF model for all x except μ_{zs} when $x = 0.1$, where the percentage errors for the two models are the same to 2 decimal places (see appendix D, table 3).

4.3. Overlapping pairs mean-field model can infer the parameters underpinning large-scale methylation patterns

While thorough inference exploration is not the objective of the current work, we carry out a short proof-of-concept study to demonstrate that the OPMF model could, in principle, be used to infer collaborativity and methylation strengths from data. We generate synthetic data for selected model parameters, see §2.2, and then infer these parameters back using the OPMF model.

Methylomes are typically assayed by whole-genome bisulfite sequencing [39]. A variant of this, hairpin-bisulfite sequencing, can be used to assay both strands of each DNA molecule [40]. In both cases, the resulting data are composed of short reads. Each read assays few CpGs and we do not know if reads originate from the same cell or DNA molecule. Simulated datasets from previous sections do not provide a good reflection of bisulfite sequencing since all of the CpGs in the simulated system were sampled at the same timepoints, the equivalent of the CpGs originating from the same molecule. To obtain short-read data, we instead simulate data for $N = 1000$ CpGs. After steady state is reached, we sample the system at 10 000 different timepoints. This is equivalent to sampling 10 000 molecules in steady state at a single timepoint. For each CpG, we take the methylation state at 30 timepoints, randomly chosen from the original 10 000. The timepoints chosen for each CpG site are independent of those chosen for other CpGs. This emulates hairpin-bisulfite sequencing data with coverage of 30 reads per CpG. We combine the sample states for all CpGs into a single dataset, X, and consider $z = \{z_1, z_2, \ldots, z_{1000}\}$, where $z_i = \{1, 2, 3\}$ if X_i corresponds to a u, h, m state, respectively. The mean and variance of z are calculated and used for inference.

There are numerous well-established methods for conducting inference. For example, in cases where a likelihood function is available, inference can be conducted using

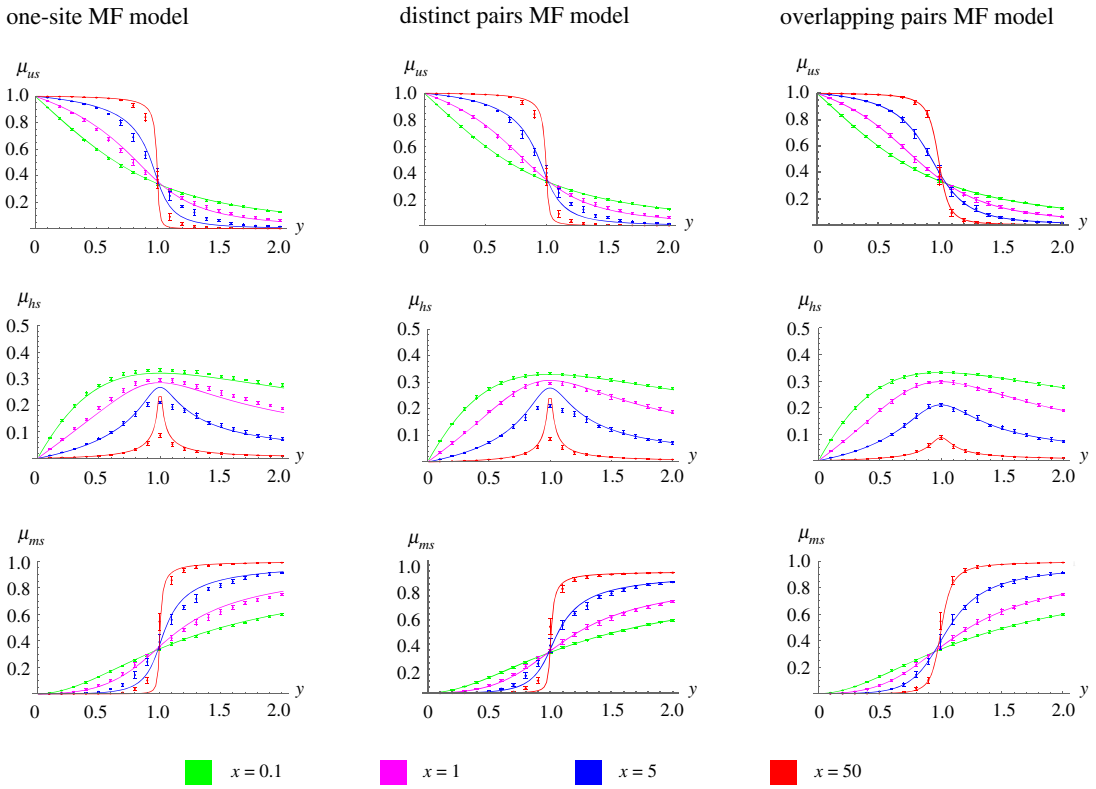

Figure 8. The OPMF model accurately predicts the average behaviour of large-scale methylation patterns. The mean u, h and m levels are plotted against the methylation strength y for the different models and collaboration strengths, x (left: one-site MF model; middle: DPMF model; right: OPMF model). For each model, the predictions of the means are shown by solid lines (with different colours denoting predictions for different values of x). The points and error bars correspond to the ground-truth means obtained from $n = 10$ replicate simulations of the full nearest-neighbour model (the mean results obtained from the ten simulations are shown by points with error bars denoting the standard error in this estimate across the ten simulations).

Table 2. Inferred parameters for the nearest-neighbour collaborative model using the OPMF model and the ABC SMC algorithm.

	$y = 0.3$	$y = 1$	$y = 1.7$
$x = 0.1$	$(x, y) = (0.197, 0.329)$	$(x, y) = (0.066, 0.997)$	$(x, y) = (0.101, 1.701)$
$x = 1$	$(x, y) = (0.676, 0.251)$	$(x, y) = (1.011, 0.990)$	$(x, y) = (0.923, 1.730)$
$x = 5$	$(x, y) = (6.302, 0.338)$	$(x, y) = (5.287, 1.000)$	$(x, y) = (6.432, 1.613)$
$x = 50$	$(x, y) = (59.571, 0.329)$	$(x, y) = (50.963, 0.999)$	$(x, y) = (65.347, 1.599)$

maximum-likelihood estimation [41], which provides point estimates for model parameters. In our case, this likelihood is the solution of the CME for the full nearest-neighbour collaborative model and thus is not available to us. We therefore use the approximate Bayesian computation sequential Monte Carlo algorithm (ABC SMC; see [42] for a comprehensive review) which is an alternative likelihood-free inference approach that allows us to use the moments derived from the OPMF model. Using a Bayesian inference approach also has the advantage of providing us with a distribution of the estimate value from which we can calculate confidence intervals associated with our inferred parameter values [43–46]. We use uniform priors, $U(0, 100)$ and $U(0, 2)$, for x and y, respectively. We also define the distance, d, between the simulations and model prediction to be the sum of the absolute relative errors of the mean and variance, i.e.

$$d = \left|\frac{\mu_{\text{model}} - \mu_{\text{data}}}{\mu_{\text{data}}}\right| + \left|\frac{\sigma^2_{\text{model}} - \sigma^2_{\text{data}}}{\sigma^2_{\text{data}}}\right|,$$

where μ_{model}, μ_{data} are the means of z from the model and data, respectively, and σ^2_{model}, σ^2_{data} are the variances associated with the model and data, respectively. To rapidly select appropriate tolerances, we calculate the true distances between the simulated data and model predictions at the parameter values of interest.

As in previous sections, we examine $x = \{0.1, 1, 5, 50\}$. For each x, we infer for $y = \{0.3, 1, 1.7\}$ using the GpABC Julia package [47]. Accepted x, y values from the final ABC SMC population make up the posterior distributions for x and y, with the means taken to be the inferred parameter values. 95% confidence intervals were calculated by removing the lowest 2.5% and highest 2.5% from the posteriors.

We find that the inferred parameter values are always of the same order of magnitude as the true values (table 2), with the most successfully inferred parameters being inferred within 1% of the true values (figure 11*a*,*b*). There are only two cases where the true parameter values lie outwith the inferred 95% confidence intervals (e.g. figure 11*c*). However, we obtain wide posteriors for large x (e.g. figure 11*d*), indicating more uncertainty in the inference.

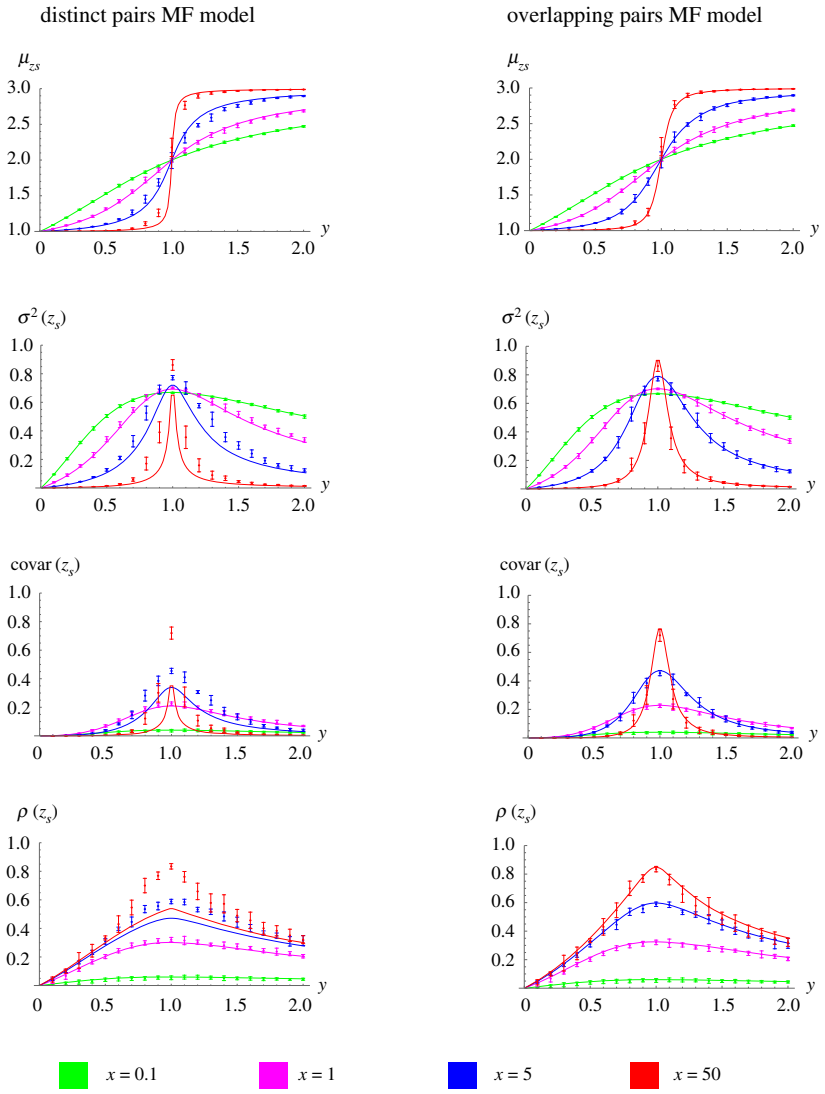

Figure 9. The OPMF model accurately predicts associations between neighbouring CpGs. Predictions of the means (μ_{zs}), variances ($\sigma^2(z_s)$), covariances (covar(z_s)) and correlations ($\rho(z_s)$) are plotted against y for the DPMF model (left) and OPMF model (right). For each model, the predictions of the statistics are shown by solid lines (with different colours denoting predictions for different values of x). The points and error bars correspond to the ground-truth statistics obtained from $n = 10$ replicate simulations of the full nearest-neighbour model (the mean results obtained from the ten simulations are shown by points, with the error bars denoting the standard error of this estimate across the ten simulations). Note that since z is determined by x and y, the statistics plotted here are implicit functions of x and y.

The accuracy of inference is highly dependent on the model sensitivity to parameters, with ease of inference increasing as sensitivity to the parameters increases. To test the sensitivity of our model to model parameters, we calculate the relative sensitivity [48] of the OPMF model to the parameters x and y, for $x = \{0.1, 0.2, ..., 99.9, 100\}$ and $y = \{0.1, 0.2, ..., 1.9, 2\}$. For all parameters considered, the x-sensitivity divided by the y-sensitivity is strictly less than one, indicating that the model shows more sensitivity to y than x. This means that y will be inferred more accurately than x.

5. Discussion

Genomic DNA methylation patterns vary between cell types, across differentiation and in disease. The mechanisms underpinning this variation remain unclear but can be better understood using mathematical models. Current approaches are limited by their inability to feasibly model large systems of CpGs and thus understand known large-scale features of methylomes. Here, we show that a cluster MF model, based around overlapping pairs of CpGs, can predict DNA methylation patterns generated under a nearest-neighbour collaborative model. This suggests that MF models are a valuable tool for understanding large-scale DNA methylation features.

Previous studies have used mathematical modelling to gain insight into the mechanisms regulating the establishment and maintenance of DNA methylation patterns. In particular, the requirement of collaborativity between CpGs to maintain DNA methylation patterns was postulated through modelling [14] before being observed experimentally [15,17]. Previous models of DNA methylation rely on stochastic simulations. However, their computational expense limits their use to the study of promoter-scale DNA methylation (regions around 1 Kb in size). The CPU-time taken to run simulations here on 200–500 sites (corresponding to approx. 20–50 Kb) varies and is parameter-dependent. For example, simulations of ten replicates with 200 CpGs take under one minute when $x = 0.1$, $y = 0.1$, but over 20 min when $x = 5$, $y = 0.9$. For $x = 50$, $y = 1$, simulations of 10

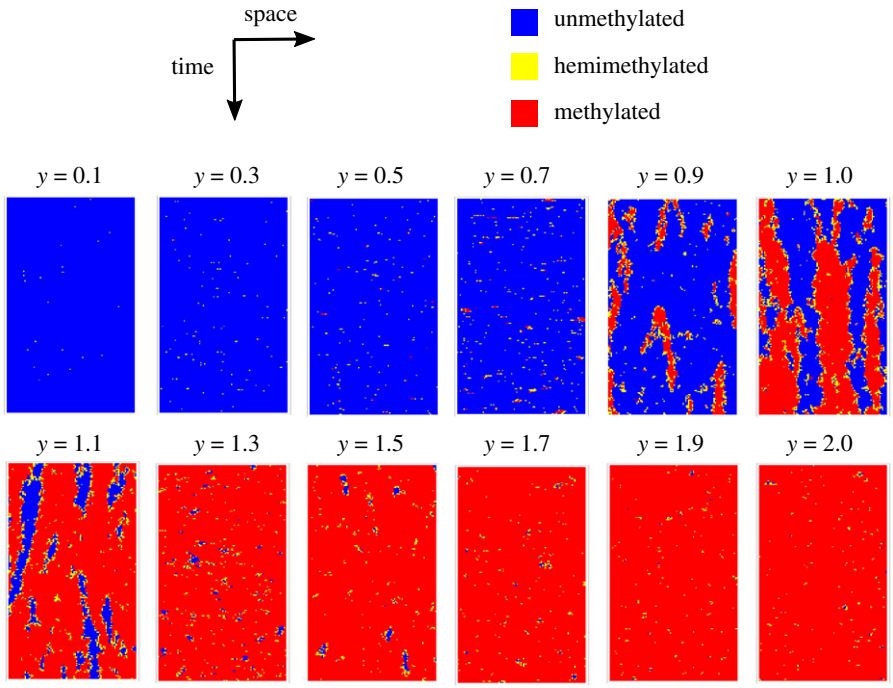

Figure 10. Size of unmethylated (u) clusters and methylated (m) clusters are most similar when $y = 1$. Visualizations of stochastic simulations over time for different values of y. For each y, the horizontal axis shows simulated methylation patterns for 100 CpGs when $x = 50$ from the full nearest-neighbour collaborative system (see §2.2). The vertical axis shows patterns obtained at successive time points after steady state has been reached where timepoints are taken at constant time intervals.

replicates with 500 CpGs take approximately 6 h. There are an average of 10 000 CpGs per megabase in the human genome making these models infeasible for such large genomic regions due to computational expense. This is particularly true in the case of parameter inference where such simulations would need to be rerun many times. By contrast, the OPMF model can be applied to arbitrarily large systems of CpGs and solved numerically in seconds for any parameters to give accurate predictions for statistics of interest. Our model is based upon the same, or similar, reaction systems used in previous stochastic modelling studies [14,29,30], but can be used to study larger systems of CpGs than previously considered. This makes our MF model far better suited to understanding the mechanisms underpinning megabase-sized variations in DNA methylation observed in development, ageing and cancer, which occur at a scale three orders of magnitude larger than promoters [27].

To our knowledge, the largest system previously examined mathematically contained 10^5 CpGs [15]. However, here simulations were conducted for only a single model parameter set. Running large-scale simulations of this type for many parameter sets will result in computational bottlenecks, meaning that such simulations cannot be used for inference. A previous study has proposed a method, based on the generalized method of moments, for rapid inference using DNA methylation patterns [49]. However, the largest system tackled with this approach contains 10 CpGs. Here, we show that our OPMF model can, in principle, be used for accurate, time-efficient inference when modelling arbitrarily large genomic regions. 'Divide and conquer' strategies have also previously been used to reduce the computational expense of large-scale simulations [50,51]. Here, a large simulation is split into smaller batches and batch results are aggregated to give overall estimates. Such an approach could make simulations of the full nearest-neighbour collaborative model applicable to larger systems of CpGs by reducing the computational burden. However, we have observed finite-size effects in our stochastic simulations suggesting that splitting them into smaller batches would result in discrepancies between estimates obtained via this approach and those obtained via a single large-scale simulation.

Our results show that the success of a MF model depends on the way in which it is constructed. Of the three MF models considered here we observe that the one-site MF model and OPMF model provide the worst and best approximations to the full nearest-neighbour model, respectively. In the full nearest-neighbour collaborative system, the probability that a CpG is in a particular state depends on the state of its nearest neighbours on the left and right. Since the one-site MF model contains no spatial information, it least reflects the full model. The DPMF and OPMF models both incorporate spatial information by allowing direct nearest-neighbour interactions to occur within pairs of CpGs. However, in the DPMF model, non-overlapping pairs are considered and the probability that a pair is in a certain state depends only on the mean state of pairs in the system, rather than on the states of its adjacent pairs. Conversely, in the OPMF model, overlapping pairs are considered and the state of a pair depends on the state of pairs it overlaps. Hence the OPMF model best reflects the full model.

When used for inference, the OPMF model has a higher sensitivity to methylation strength (y) than collaborativity strength (x) explaining why the former is generally better inferred than the latter. However, some posteriors obtained in §4.3 are very wide and/or the true parameter values lie outside the inferred 95% CIs, indicating that there is scope for inference to be improved. Since our model shows

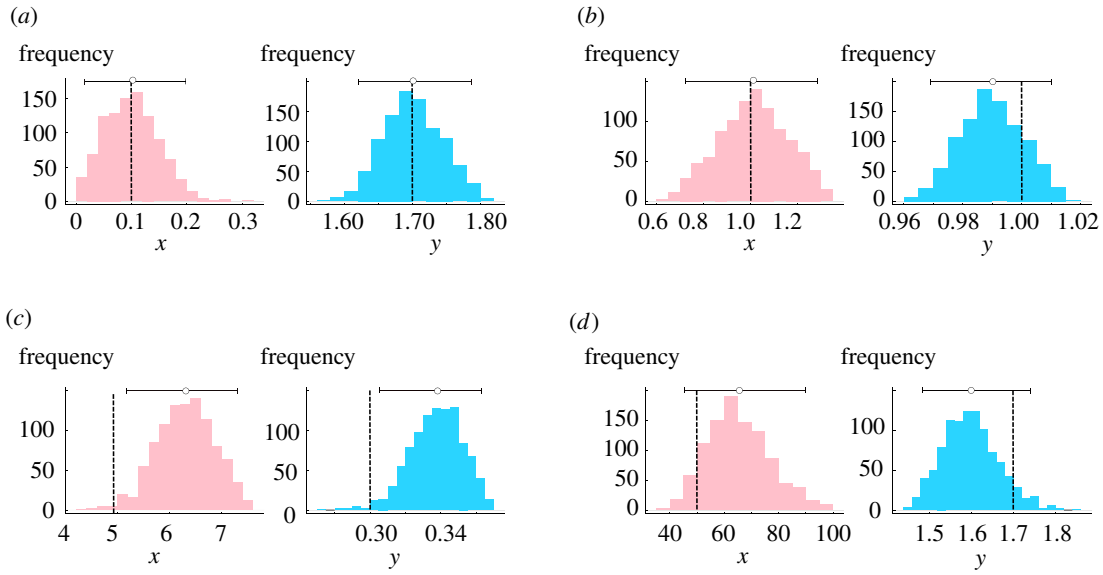

Figure 11. The OPMF model can be used to infer collaborativity and methylation strengths. Example posteriors from inference are shown, with (*a–d*) each corresponding to a different parameter set. True parameter values are denoted by dashed lines, inferred parameter values are shown as dots, with 95% confidence intervals shown as horizontal bars.

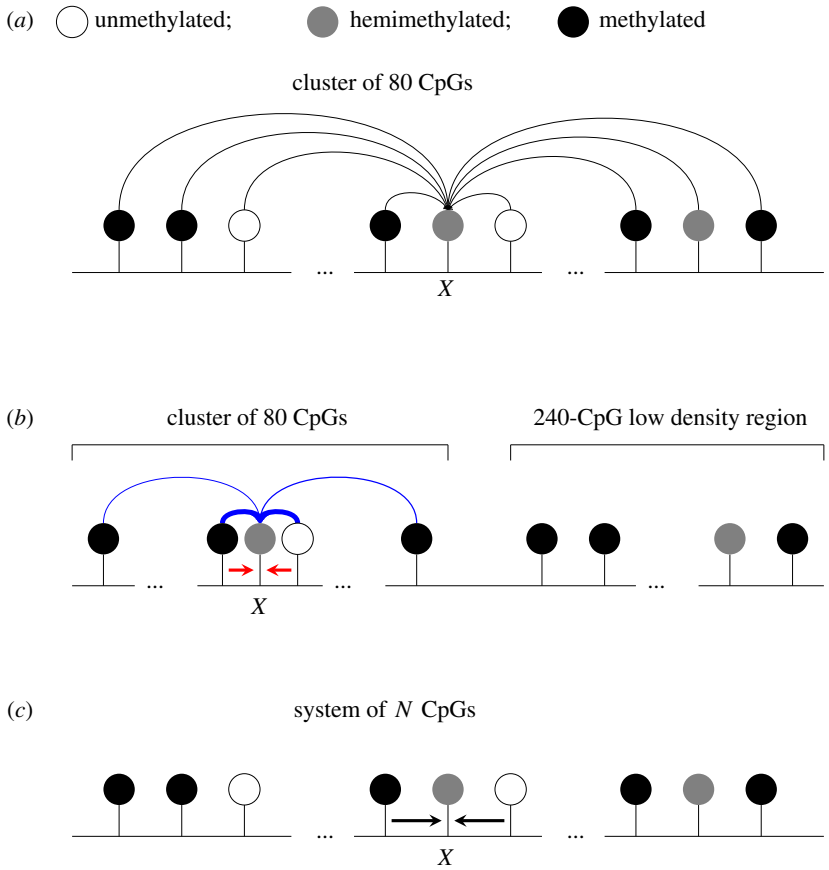

Figure 12. Potential collaborative interactions that can influence a target, *X*, under the models in [14] and the model considered here. (*a*) A cluster of 80 CpGs is first considered in [14], where a CpG can collaborate with any other CpG in the system with equal probability. (*b*) A high-density cluster (of 80 CpGs) adjacent to a highly methylated low-density region (of 240 CpGs) is then considered in [14], where there is nearest-neighbour collaborative methylation (red arrows) and the probability of collaborative demethylation occurring due to interaction between two sites decays as the distance between them increases (blue arrows; decay in reaction probability shown by narrowing width of arrows). Note that collaborative demethylation is restricted to the 80-CpG cluster. (*c*) In this paper, collaborative reactions only occur between neighbouring CpGs and there is no upper bound on the system size, allowing large-scale patterns to be considered.

impressive performance in forward prediction, discrepancies between true and inferred parameters are likely due to insufficient data or the inference technique used. ABC algorithms are sensitive to the summary statistics, distance and tolerance used in the acceptance criteria for parameters [52]. Moreover, regardless of the method used, inference will always be difficult for systems of CpGs that are either ~100% methylated or ~100% unmethylated. This is because a very large x would

result in ~100% unmethylation for almost all $y < 1$ and ~100% methylation for almost all $y > 1$. Despite these challenges, our analysis demonstrates that the OPMF model can in principle be used for inference. The accuracy of inference may be improved in future studies by experimenting with different inference techniques and sample sizes. In addition, the summary statistics that can be reliably calculated from real data are dependent on the technology used to generate the data. Most DNA methylation data are currently derived from technologies that produce short reads containing between one and three CpGs. Statistics involving associations between consecutive CpGs from a single molecule cannot therefore be accurately estimated. This restricts inference based on such data to statistics such as the mean and standard deviation. Long-read technologies which can capture hundreds of CpGs on a single read have recently been applied to assay DNA methylation patterns [53]. Data obtained from such technologies could allow higher order statistics to be accurately estimated and potentially improve inference.

Here, we assume that the processes governing the creation of methylation patterns *in vivo* are described well by our nearest-neighbour collaborative model. It is possible that collaborativity *in vivo* can occur between non-nearest-neighbours, something which is not explicitly accounted for in our nearest-neighbour collaborative model. Collaborative methylation interactions are likely determined by the properties of the DNA methylation machinery. DNMT1 and DNMT3B both methylate processively along DNA strands whereas DNMT3A methylates in a distributive manner but can form multimers along the DNA fibre [32]. However, the range and strengths across which these interactions occur are currently unclear so we focus on nearest-neighbour interactions. We note that our OPMF model does capture interactions beyond nearest neighbours because the mean state of pairs in the system influences the change in state of CpGs. Previous modelling studies have also considered different forms of collaborative interactions between CpGs in a system (figure 12*a*,*b*). In [14], collaboration between any CpGs in the system and nearest-neighbour collaborative methylation alongside distance-dependent collaborative demethylation were both demonstrated to produce stable CpG clusters that were either methylated or unmethylated. Stable clusters were also observed under a distance-dependent collaborative model, where collaborative demethylation dominates over short ranges and collaborative methylation dominates over long ranges [29].

Here, we have also assumed that the system of CpGs we model reaches a steady state. This means that we consider either non-dividing cells, or dividing cells which settle down to steady state between replication events such that DNA replication has no long-term effects on DNA methylation patterns. Whether or not a cell satisfies the latter case is dependent on the real magnitudes of k_i, $i = \{1, \ldots, 12\}$, and the time between replication events. Experimental studies show that arrested cells have similar DNA methylation patterns to those that are cycling, supporting the assumption that DNA replication has no long-term effect on DNA methylation [33]. Furthermore, an analysis of DNA methylation patterns on newly synthesized DNA suggests that re-methylation occurs within 20 min of replication [54]. However, another analysis of DNA methylation following replication suggests re-methylation is often delayed [55]. At present, it is unclear whether this delay is sufficient to have an effect on methylation patterns during the following cell cycle.

Our assumption that k_i, $i = \{1, \ldots, 12\}$ take the form in table 1 could be violated in reality. For example, DNMT1 shows a strong preference for h sites over u sites [32], meaning that methylation reactions with an h target may have higher reaction rates than those with a u target. Future work could relax rate assumptions to account for such factors. Preliminary investigations confirm that the OPMF model provides a good approximation to the nearest-neighbour collaborative system when k_i, $i = \{1, \ldots, 12\}$ are considered as twelve independent parameters (data not shown). However, the difficulty of parameter inference increases with the number of parameters, meaning that relaxing parameter assumptions will likely decrease inference quality. Nonetheless, our model suggests that it is the parameters x and y that determine steady-state methylation patterns, rather than individual reaction rates. This is supported by a study where modelling of experimental data suggested that the ratio between methylation and demethylation rates determines steady-state methylation levels at single CpGs [56].

Here, we demonstrate that MF models can accurately predict the behaviour of large CpG systems subjected to nearest-neighbour collaboration. Our study presents the first mathematical modelling approach that can be applied to arbitrarily large systems of CpGs. The future application of this approach will facilitate the delineation of the methylation dynamics that underpin the formation of large-scale methylation patterns in developmental and disease contexts.

Data accessibility. The code used to simulate the full nearest-neighbour collaborative system is available at https://github.com/lkerr34/kerr_mean_field_paper_2021.

Authors' contributions. L.K.: formal analysis, writing—original draft, writing—review & editing; D.S.: conceptualization, supervision, writing—original draft, writing—review & editing; R.G.: conceptualization, supervision, writing—original draft, writing—review & editing.

Competing interests. We declare we have no competing interests.

Funding. L.K. is a cross-disciplinary post-doctoral fellow supported by funding from the University of Edinburgh and Medical Research Council (MC_UU_00009/2). D.S. is a Cancer Research UK Career Development fellow (reference C47648/A20837), and work in his laboratory is also supported by an MRC university grant to the MRC Human Genetics Unit. R.G. is supported by Leverhulme Trust research awards (RPG-2018-423 and RPG-2020-327).

Acknowledgements. We thank Chris Ponting, Diego Oyarzun and members of the Grima and Sproul laboratories for helpful discussions about the manuscript.

Appendix A. Derivation of effective reaction rates for the distinct pairs mean-field model

Here, we illustrate how $\hat{a}_1$, the reaction rate associated with $mm \longrightarrow hm$ in (3.13), is constructed. We first consider all reactions in figure 1 that can occur *within* the mm pair, resulting in conversion to an hm pair. Clearly, $mm \longrightarrow hm$ occurs if either m within the mm undergoes $m \xrightarrow{k_3} h$. Hence $\hat{a}_1$ must contain a $2k_3$ term. However, no collaborative reaction involves two m CpGs interacting to produce an h and an m. Thus $mm \longrightarrow hm$ cannot arise from collaboration between the two CpGs within the mm.

Next, we consider all $mm \longrightarrow hm$ reactions that can occur due to an m within the mm interacting with a site from an adjacent pair. For example, an m could collaborate with a u from an adjacent pair via $m + u \xrightarrow{k_{10}} h + u$, i.e. we can have

$$\begin{aligned} &mm + uu \longrightarrow hm + uu, \quad mm + um \longrightarrow hm + um, \\ &mm + uh \longrightarrow hm + uh. \end{aligned} \tag{A 1}$$

The first reaction can be written as an effective first-order reaction, $mm \xrightarrow{2k_{10}\mu_2} hm$, where the uu is absorbed into the reaction rate by making it proportional to μ_2, and the factor of two allows for either m within the mm to undergo this reaction. The second reaction in (A 1) can be written as $mm \xrightarrow{k_{10}\mu_4} hm$, where the um is absorbed into the reaction rate and either m can undergo the reaction, giving us a factor of two. However, this reaction requires the u within the um to be directly adjacent to the mm. The probability of having um in this particular order is $\frac{1}{2}$, giving an effective reaction rate of $2k_{10}\frac{1}{2}\mu_4 = k_{10}\mu_4$. Similarly, the third reaction in (A 1) can be written as $mm \xrightarrow{k_{10}\mu_6} hm$.

Finally, an m within the mm can interact with an h from an adjacent pair, via $m + h \xrightarrow{k_9} h + h$, i.e. we can have

$$\begin{aligned} &mm + hh \longrightarrow hm + hh, \quad mm + hm \longrightarrow hm + hm, \\ &mm + uh \longrightarrow hm + uh. \end{aligned} \tag{A 2}$$

Using similar arguments as above, we can write each of these three reactions as the effective first-order reaction $mm \longrightarrow hm$, with rates $2k_9\mu_3$, $k_9\mu_5$ and $k_9\mu_6$, respectively.

Hence,

$$\hat{a}_1 = 2k_3 + k_9(2\mu_3 + \mu_5 + \mu_6) + k_{10}(2\mu_2 + \mu_4 + \mu_6). \tag{A 3}$$

Appendix B. Derivation of effective reaction rates for the overlapping pairs mean-field model

We now construct $\tilde{a}_1$, the reaction rate associated with $mm \longrightarrow hm$ in (3.19). As with the DPMF model, reactions occurring within the mm pair contribute a $2k_3$ term to $\tilde{a}_1$.

Now, $mm \longrightarrow hm$ can occur due to a $m + u \xrightarrow{k_{10}} h + u$ reaction if one of the m sites in the mm pair is also in a pair with a u site, i.e. we can have

$$mm + um \longrightarrow hm + um,$$

where now the um and mm reactants share a common m. This can be written as the effective first-order reaction $mm \longrightarrow hm$ with rate

$$2k_{10}\left(\frac{\mu_4/2}{\mu_1 + \mu_4/2 + \mu_5/2}\right), \tag{B 1}$$

where $\mu_4/2/(\mu_1 + \mu_4/2 + \mu_5/2)$ is the probability that an m within the mm also forms a um with its other neighbour, i.e. it is the conditional probability that a pair is in the um state given that we know a particular site in the pair is m. Recall that μ_4 gives the proportion of um and mu pairs, while μ_5 gives the proportion of hm and mh pairs. The factors of $1/2$ in (B 1) account for the fact that the u and h in the um and hm must be on a particular side of the m (since an m is on its other side). The factor of two at the front of (B 1) allows for either m in the mm to undergo the reaction.

Similarly, $mm \longrightarrow hm$ can occur due to a $m + h \xrightarrow{k_9} h + h$ reaction if one of the m sites in the mm is also in a pair with a h site, i.e. we can have

$$mm + hm \longrightarrow hm + hm,$$

where the reactant hm and mm share a common m. This reaction can again be written as $mm \longrightarrow hm$, where the rate

$$2k_9\left(\frac{\mu_5/2}{\mu_1 + \mu_4/2 + \mu_5/2}\right)$$

is derived in a similar way as above.

Hence, we have

$$\tilde{a}_1 = 2k_3 + 2k_9\left(\frac{\mu_5/2}{\mu_1 + \mu_4/2 + \mu_5/2}\right) + 2k_{10}\left(\frac{\mu_4/2}{\mu_1 + \mu_4/2 + \mu_5/2}\right). \tag{B 2}$$

Appendix C. Master equations and moment equations for the cluster mean-field models

Here, we construct the CME for the DPMF and OPMF models, along with the first and second moment equations for the systems. Note that, for $i = 1, \ldots, 12$, $a_i = \hat{a}_i$ in the DPMF model, where $\hat{a}_i$ is defined in (3.14), and $a_i = \tilde{a}_i$ in the OPMF model, where $\tilde{a}_i$ is defined in (3.20). The CME is given by

$$\begin{aligned} \frac{\mathrm{d}P(S_1)}{\mathrm{d}t} &= -a_1P(S_1) + a_8P(S_5), \\ \frac{\mathrm{d}P(S_2)}{\mathrm{d}t} &= -a_2P(S_2) + a_{11} \\ &\quad \big(1 - P(S_1) - P(S_2) - P(S_3) - P(S_4) - P(S_5)\big), \\ \frac{\mathrm{d}P(S_3)}{\mathrm{d}t} &= -(a_3 + a_4)P(S_3) + a_9P(S_5) + a_{10} \\ &\quad \big(1 - P(S_1) - P(S_2) - P(S_3) - P(S_4) - P(S_5)\big), \\ \frac{\mathrm{d}P(S_4)}{\mathrm{d}t} &= -(a_5 + a_6)P(S_4) + a_7P(S_5) + a_{12} \\ &\quad \big(1 - P(S_1) - P(S_2) - P(S_3) - P(S_4) - P(S_5)\big) \end{aligned}$$

and

$$\begin{aligned} \frac{\mathrm{d}P(S_5)}{\mathrm{d}t} &= a_1P(S_1) + a_4P(S_3) \\ &\quad + a_5P(S_4) - (a_7 + a_8 + a_9)P(S_5), \end{aligned}$$

where for $j = 1, \ldots, 6$, $S_j = (L_1, L_2, L_3, L_4, L_5, L_6)$, with $L_j = 1$ and $L_i = 0$ for $i \neq j$, and $P(S_j)$ is the probability associated with the state S_i. Note that interaction between pairs is incorporated into the CME via the effective reaction rates a_1–a_{12} (see appendix A and appendix B). Following the approach in appendix C of [35], we obtain first moment equations for the DPMF and OPMF models, where μ_1–μ_6 are the mean levels of each paired state:

$$\begin{aligned} \frac{\mathrm{d}\mu_1}{\mathrm{d}t} &= -a_1\mu_1 + a_8\mu_5, \\ \frac{\mathrm{d}\mu_2}{\mathrm{d}t} &= -a_2\mu_2 + a_{11}(1 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5), \\ \frac{\mathrm{d}\mu_3}{\mathrm{d}t} &= -(a_3 + a_4)\mu_3 + a_9\mu_5 + a_{10}(1 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5), \\ \frac{\mathrm{d}\mu_4}{\mathrm{d}t} &= -(a_5 + a_6)\mu_4 + a_7\mu_5 + a_{12}(1 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5) \end{aligned}$$

and

$$\frac{\mathrm{d}\mu_5}{\mathrm{d}t} = a_1\mu_1 + a_4\mu_3 + a_5\mu_4 - (a_7 + a_8 + a_9)\mu_5.$$

Note again that we can obtain μ_6 via $\mu_6 = 1 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5$. The second moment equations, describing the

evolution of $\langle L_iL_j\rangle$ for $1 \le i, j \le 5$, are given by

$$\frac{\mathrm{d}\langle L_1L_1\rangle}{\mathrm{d}t} = a_1\mu_1 + a_8\mu_5 + 2\Big(a_1\langle L_1L_1\rangle + a_8\langle L_1L_5\rangle\Big),$$

$$\frac{\mathrm{d}\langle L_2L_2\rangle}{\mathrm{d}t} = a_2\mu_2 + a_{11}(1-\mu_1-\mu_2-\mu_3-\mu_4-\mu_5) + 2\Big(-a_2\langle L_2L_2\rangle + a_{11}(\mu_2 - \langle L_1L_2\rangle - \langle L_2L_2\rangle - \langle L_2L_3\rangle - \langle L_2L_4\rangle - \langle L_2L_5\rangle)\Big),$$

$$\frac{\mathrm{d}\langle L_3L_3\rangle}{\mathrm{d}t} = (a_3+a_4)\mu_3 + a_9\mu_5 + a_{10}(1-\mu_1-\mu_2-\mu_3-\mu_4-\mu_5) + 2\Big(-(a_3+a_4)\langle L_3L_3\rangle + a_9\langle L_3L_5\rangle + a_{10}\big(\mu_3 - \langle L_1L_3\rangle - \langle L_2L_3\rangle - \langle L_3L_3\rangle - \langle L_3L_4\rangle - \langle L_3L_5\rangle\big)\Big),$$

$$\frac{\mathrm{d}\langle L_4L_4\rangle}{\mathrm{d}t} = (a_5+a_6)\mu_4 + a_7\mu_5 + a_{12}(1-\mu_1-\mu_2-\mu_3-\mu_4-\mu_5) + 2\Big(-(a_5+a_6)\langle L_4L_4\rangle + a_7\langle L_4L_5\rangle + a_{12}\big(\mu_4 - \langle L_1L_4\rangle - \langle L_2L_4\rangle - \langle L_3L_4\rangle - \langle L_4L_4\rangle - \langle L_4L_5\rangle\big)\Big),$$

$$\frac{\mathrm{d}\langle L_5L_5\rangle}{\mathrm{d}t} = a_1\mu_1 + a_4\mu_3 + a_5\mu_4 + (a_7+a_8+a_9)\mu_5 + 2\Big(a_1\langle L_1L_5\rangle + a_4\langle L_3L_5\rangle + a_5\langle L_4L_5\rangle - (a_7+a_8+a_9)\langle L_5L_5\rangle\Big),$$

$$\frac{\mathrm{d}\langle L_1L_2\rangle}{\mathrm{d}t} = -(a_1+a_2)\langle L_1L_2\rangle + a_8\langle L_2L_5\rangle + a_{11}\big(\mu_1 - \langle L_1L_1\rangle - \langle L_1L_2\rangle - \langle L_1L_3\rangle - \langle L_1L_4\rangle - \langle L_1L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_1L_3\rangle}{\mathrm{d}t} = -(a_1+a_3+a_4)\langle L_1L_3\rangle + a_8\langle L_3L_5\rangle + a_9\langle L_1L_5\rangle + a_{10}\big(\mu_1 - \langle L_1L_1\rangle - \langle L_1L_2\rangle - \langle L_1L_3\rangle - \langle L_1L_4\rangle - \langle L_1L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_1L_4\rangle}{\mathrm{d}t} = -(a_1+a_5+a_6)\langle L_1L_4\rangle + a_7\langle L_1L_5\rangle + a_8\langle L_4L_5\rangle + a_{12}\big(\mu_1 - \langle L_1L_1\rangle - \langle L_1L_2\rangle - \langle L_1L_3\rangle - \langle L_1L_4\rangle - \langle L_1L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_1L_5\rangle}{\mathrm{d}t} = -a_1\mu_1 - a_8\mu_5 + a_1\langle L_1L_1\rangle + a_4\langle L_1L_3\rangle + a_5\langle L_1L_4\rangle + a_8\langle L_5L_5\rangle - (a_1+a_7+a_8+a_9)\langle L_1L_5\rangle,$$

$$\frac{\mathrm{d}\langle L_2L_3\rangle}{\mathrm{d}t} = -(a_2+a_3+a_4)\langle L_2L_3\rangle + a_9\langle L_2L_5\rangle + a_{10}\big(\mu_2 - \langle L_1L_2\rangle - \langle L_2L_2\rangle - \langle L_2L_3\rangle - \langle L_2L_4\rangle - \langle L_2L_5\rangle\big) + a_{11}\big(\mu_3 - \langle L_1L_3\rangle - \langle L_2L_3\rangle - \langle L_3L_3\rangle - \langle L_3L_4\rangle - \langle L_3L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_2L_4\rangle}{\mathrm{d}t} = -(a_2+a_5+a_6)\langle L_2L_4\rangle + a_7\langle L_2L_5\rangle + a_{11}\big(\mu_4 - \langle L_1L_4\rangle - \langle L_2L_4\rangle - \langle L_3L_4\rangle - \langle L_4L_4\rangle - \langle L_4L_5\rangle\big) + a_{12}\big(\mu_2 - \langle L_1L_2\rangle - \langle L_2L_2\rangle - \langle L_2L_3\rangle - \langle L_2L_4\rangle - \langle L_2L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_2L_5\rangle}{\mathrm{d}t} = a_1\langle L_1L_2\rangle - (a_2+a_7+a_8+a_9)\langle L_2L_5\rangle + a_4\langle L_2L_3\rangle + a_5\langle L_2L_4\rangle + a_{11}\big(\mu_5 - \langle L_1L_5\rangle - \langle L_2L_5\rangle - \langle L_3L_5\rangle - \langle L_4L_5\rangle - \langle L_5L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_3L_4\rangle}{\mathrm{d}t} = -(a_3+a_4+a_5+a_6)\langle L_3L_4\rangle + a_7\langle L_3L_5\rangle + a_9\langle L_4L_5\rangle + a_{10}\big(\mu_4 - \langle L_1L_4\rangle - \langle L_2L_4\rangle - \langle L_3L_4\rangle - \langle L_4L_4\rangle - \langle L_4L_5\rangle\big) + a_{12}\big(\mu_3 - \langle L_1L_3\rangle - \langle L_2L_3\rangle - \langle L_3L_3\rangle - \langle L_3L_4\rangle - \langle L_3L_5\rangle\big),$$

$$\frac{\mathrm{d}\langle L_3L_5\rangle}{\mathrm{d}t} = -a_4\mu_3 + a_9\mu_5 + a_1\langle L_1L_3\rangle + a_4\langle L_3L_3\rangle + a_5\langle L_3L_4\rangle - (a_3+a_4+a_7+a_8+a_9)\langle L_3L_5\rangle + a_9\langle L_5L_5\rangle + a_{10}\big(\mu_5 - \langle L_1L_5\rangle - \langle L_2L_5\rangle - \langle L_3L_5\rangle - \langle L_4L_5\rangle - \langle L_5L_5\rangle\big)$$

and

$$\frac{\mathrm{d}\langle L_4L_5\rangle}{\mathrm{d}t} = -a_5\mu_4 - a_7\mu_5 + a_1\langle L_1L_4\rangle + a_4\langle L_3L_4\rangle + a_5\langle L_4L_4\rangle - (a_5+a_6+a_7+a_8+a_9)\langle L_4L_5\rangle + a_7\langle L_5L_5\rangle + a_{12}\big(\mu_5 - \langle L_1L_5\rangle - \langle L_2L_5\rangle - \langle L_3L_5\rangle - \langle L_4L_5\rangle - \langle L_5L_5\rangle\big).$$

Using $L_6 = 1 - L_1 - L_2 - L_3 - L_4 - L_5$, we obtain

$$\begin{aligned}
\langle L_1L_6\rangle &= \mu_1 - \langle L_1L_1\rangle - \langle L_1L_2\rangle - \langle L_1L_3\rangle + \langle L_1L_4\rangle + \langle L_1L_5\rangle,\\
\langle L_2L_6\rangle &= \mu_2 - \langle L_1L_2\rangle - \langle L_2L_2\rangle - \langle L_2L_3\rangle + \langle L_2L_4\rangle + \langle L_2L_5\rangle,\\
\langle L_3L_6\rangle &= \mu_3 - \langle L_1L_3\rangle - \langle L_2L_3\rangle - \langle L_3L_3\rangle + \langle L_3L_4\rangle + \langle L_3L_5\rangle,\\
\langle L_4L_6\rangle &= \mu_4 - \langle L_1L_4\rangle - \langle L_2L_4\rangle - \langle L_3L_4\rangle + \langle L_4L_4\rangle + \langle L_4L_5\rangle,\\
\langle L_5L_6\rangle &= \mu_5 - \langle L_1L_5\rangle - \langle L_2L_5\rangle - \langle L_3L_5\rangle + \langle L_4L_5\rangle + \langle L_5L_5\rangle
\end{aligned}$$

and

$$\langle L_6L_6\rangle = 1 - 2(\mu_1+\mu_2+\mu_3+\mu_4+\mu_5) + \langle L_1L_1\rangle + \langle L_2L_2\rangle + \langle L_3L_3\rangle + \langle L_4L_4\rangle + \langle L_5L_5\rangle + 2\Big(\langle L_1L_2\rangle + \langle L_1L_3\rangle + \langle L_1L_4\rangle + \langle L_1L_5\rangle + \langle L_2L_3\rangle + \langle L_2L_4\rangle + \langle L_2L_5\rangle + \langle L_3L_4\rangle + \langle L_3L_5\rangle + \langle L_4L_5\rangle\Big).$$

Appendix D. Percentage error tables

See table 3.

Table 3. Percentage errors associated with figures 8 and 9 for the MF models and values of x considered. For each x, the relative error was calculated for the three models at each y, where the relative error associated with a statistic γ is $|\gamma_{\text{sim}} - \gamma_{\text{pred}}|/\gamma_{\text{sim}}$, with γ_{pred} being the model prediction and γ_{sim} being the result obtained from the simulations. Overall percentage errors were obtained for each model by averaging the relative errors across y (for each x) and multiplying by 100. For all three models, percentage errors are shown for μ_{us}, μ_{hs} and μ_{ms}. For the paired MF models, percentage errors for μ_{zs}, $\sigma^2(z_s)$, covar(z_s) and $\rho(z_s)$ are also displayed.

	μ_{us}	μ_{hs}	μ_{ms}	μ_{zs}	$\sigma^2(z_s)$	covar(z_s)	$\rho(z_s)$
$x = 0.1$							
one-site MF	1.46%	3.16%	1.58%	n.a.	n.a.	n.a.	n.a.
DPMF	0.78%	0.77%	0.95%	0.17%	0.45%	25.70%	25.38%
OPMF	0.70%	0.38%	0.90%	0.17%	0.42%	24.32%	24.05%
$x = 1$							
one-site MF	6.11%	6.60%	6.46%	n.a.	n.a.	n.a.	n.a.
DPMF	4.27%	1.91%	3.91%	0.73%	3.34%	7.58%	5.90%
OPMF	1.06%	0.812%	1.65%	0.23%	0.96%	5.22%	5.64%
$x = 5$							
one-site MF	26.5%	6.03%	22.21%	n.a.	n.a.	n.a.	n.a.
DPMF	19.13%	6.99%	14.91%	1.93%	16.52%	26.54%	13.60%
OPMF	2.60%	2.05%	1.63%	0.29%	2.67%	4.37%	3.30%
$x = 50$							
one-site MF	46.00%	21.84%	42.65%	n.a.	n.a.	n.a.	n.a.
DPMF	32.62%	19.80%	28.98%	1.78%	34.89%	44.27%	18.85%
OPMF	7.13%	6.30%	7.20%	0.71%	10.36%	12.84%	6.91%

References

1. Schübeler D. 2015 Function and information content of DNA methylation. *Nature* **517**, 321–326. (doi:10.1038/nature14192)
2. Smallwood SA, Kelsey G. 2012 De novo DNA methylation: a germ cell perspective. *Trends Genet.* **28**, 33–42. (doi:10.1016/j.tig.2011.09.004)
3. Okano M, Bell DW, Haber DA, Li E. 1999 DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257. (doi:10.1016/S0092-8674(00)81656-6)
4. Suzuki MM, Bird A. 2008 DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476. (doi:10.1038/nrg2341)
5. Li E, Bestor TH, Jaenisch R. 1992 Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926. (doi:10.1016/0092-8674(92)90611-F)
6. Wu X, Zhang Y. 2017 TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534. (doi:10.1038/nrg.2017.33)
7. Smith ZD, Meissner A. 2013 DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220. (doi:10.1038/nrg3354)
8. Heyn P *et al.* 2019 Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* **51**, 96–105. (doi:10.1038/s41588-018-0274-x)
9. Xu GL *et al.* 1999 Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191. (doi:10.1038/46052)
10. Winkelmann J *et al.* 2012 Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* **21**, 2205–2210. (doi:10.1093/hmg/dds035)
11. Sproul D, Meehan RR. 2013 Genomic insights into cancer-associated aberrant CpG island hypermethylation. *Brief. Funct. Genomics* **12**, 174–190. (doi:10.1093/bfgp/els063)
12. Jones MJ, Goodman SJ, Kobor MS. 2015 DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932. (doi:10.1111/acel.2015.14.issue-6)
13. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. 2003 Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–492. (doi:10.1126/science.1083558)
14. Haerter JO, Lövkvist C, Dodd IB, Sneppen K. 2014 Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* **42**, 2235–2244. (doi:10.1093/nar/gkt1235)
15. Busto-Moner L, Morival J, Ren H, Fahim A, Reitz Z, Downing TL, Read EL. 2020 Stochastic modeling reveals kinetic heterogeneity in post-replication DNA methylation. *PLoS Comput. Biol.* **16**, e1007195. (doi:10.1371/journal.pcbi.1007195)
16. Adam S, Anteneh H, Hornisch M, Wagner V, Lu J, Radde NE, Bashtrykov P, Song J, Jeltsch A. 2020 DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat. Commun.* **11**, 3723. (doi:10.1038/s41467-020-17531-8)
17. Wang Q *et al.* 2020 Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet

flexible epigenetic inheritance. *Nat. Genet.* **52**, 828–839. (doi:10.1038/s41588-020-0661-y)

18. Mc Auley MT, Mooney KM, Salcedo-Sora JE. 2018 Computational modelling folate metabolism and DNA methylation: implications for understanding health and ageing. *Brief. Bioinform.* **19**, 303–317.
19. Gillespie DT. 2007 Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55. (doi:10.1146/physchem.2007.58.issue-1)
20. Johnston I. 2012 The chaos within: exploring noise in cellular biology. *Significance* **9**, 17–21. (doi:10.1111/j.1740-9713.2012.00586.x)
21. Schnoerr D, Sanguinetti G, Grima R. 2017 Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A: Math. Theor.* **50**, 093001. (doi:10.1088/1751-8121/aa54d9)
22. Jenkinson G, Abante J, Feinberg AP, Goutsias J. 2018 An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. *BMC Bioinf.* **19**, 1–23. (doi:10.1186/s12859-018-2086-5)
23. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. 2017 Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat. Genet.* **49**, 719–729. (doi:10.1038/ng.3811)
24. Zhang Y, Wang S, Wang X. 2018 Data-driven-based approach to identifying differentially methylated regions using modified 1D Ising model. *BioMed Res. Int.* **2018**, 1070645. (doi:10.1155/2018/1070645)
25. Lück A, Wolf V. 2020 A stochastic automata network description for spatial DNA-methylation models. In *Int. Conf. on Measurement, Modelling and Evaluation of Computing Systems*, pp. 54–64. Berlin, Germany: Springer.
26. Lück A, Giehr P, Walter J, Wolf V. 2017 A stochastic model for the formation of spatial methylation patterns. In *Int. Conf. on Computational Methods in Systems Biology*, pp. 160–178. Berlin, Germany: Springer.
27. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP. 2018 DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602. (doi:10.1038/s41588-018-0073-4)
28. Du Q *et al.* 2019 Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat. Commun.* **10**, 1. (doi:10.1038/s41467-018-07882-8)
29. Lövkvist C, Dodd IB, Sneppen K, Haerter JO. 2016 DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res.* **44**, 5123–5132. (doi:10.1093/nar/gkw124)
30. Utsey K, Keener JP. 2020 A mathematical model for inheritance of DNA methylation patterns in somatic cells. *Bull. Math. Biol.* **82**, 1–23. (doi:10.1007/s11538-020-00765-4)
31. Dodd IB, Micheelsen MA, Sneppen K, Thon G. 2007 Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129**, 813–822. (doi:10.1016/j.cell.2007.02.053)
32. Jeltsch A, Jurkowska RZ. 2016 Allosteric control of mammalian DNA methyltransferases—a new regulatory paradigm. *Nucleic Acids Res.* **44**, 8556–8575. (doi:10.1093/nar/gkw723)
33. Vandiver AR, Idrizi A, Rizzardi L, Feinberg AP, Hansen KD. 2015 DNA methylation is stable during replication and cell cycle arrest. *Sci. Rep.* **5**, 17911. (doi:10.1038/srep17911)
34. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434. (doi:10.1016/0021-9991(76)90041-3)
35. Grima R. 2012 A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **136**, 04B616. (doi:10.1063/1.3702848)
36. Schreckenberg M, Schadschneider A, Nagel K, Ito N. 1995 Discrete stochastic models for traffic flow. *Phys. Rev. E* **51**, 2939–2949. (doi:10.1103/PhysRevE.51.2939)
37. Chowdhury D, Wang JS. 2002 Flow properties of driven-diffusive lattice gases: theory and computer simulation. *Phys. Rev. E* **65**, 046126. (doi:10.1103/PhysRevE.65.046126)
38. Jahnke T, Huisinga W. 2007 Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.* **54**, 1–26. (doi:10.1007/s00285-006-0034-x)
39. Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. 2015 MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483. (doi:10.1038/nprot.2014.114)
40. Zhao L *et al.* 2014 The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res.* **24**, 1296–1307. (doi:10.1101/gr.163147.113)
41. Rossi RJ. 2018 *Mathematical statistics: an introduction to likelihood based inference*. New York, NY: John Wiley & Sons.
42. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202. (doi:10.1098/rsif.2008.0172)
43. Cao Z, Grima R. 2019 Accuracy of parameter estimation for auto-regulatory transcriptional feedback loops from noisy data. *J. R. Soc. Interface* **16**, 20180967. (doi:10.1098/rsif.2018.0967)
44. Golightly A, Wilkinson DJ. 2005 Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61**, 781–788. (doi:10.1111/j.1541-0420.2005.00345.x)
45. Beaumont MA, Zhang W, Balding DJ. 2002 Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035. (doi:10.1093/genetics/162.4.2025)
46. Williams MJ, Werner B, Heide T, Curtis C, Barnes CP, Sottoriva A, Graham TA. 2018 Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**, 895–903. (doi:10.1038/s41588-018-0128-6)
47. Tankhilevich E, Ish-Horowicz J, Hameed T, Roesch E, Kleijn I, Stumpf MP, He F. 2020 GpABC: a Julia package for approximate Bayesian computation with Gaussian process emulation. *Bioinformatics* **36**, 3286–3287. (doi:10.1093/bioinformatics/btaa078)
48. Wolkenhauer O, Wellstead P, Cho KH, Ingalls B. 2008 Sensitivity analysis: from model parameters to system behaviour. *Essays Biochem.* **45**, 177–194. (doi:10.1042/bse0450177)
49. Lück A, Wolf V. 2019 Generalized method of moments estimation for stochastic models of DNA methylation patterns. (https://arxiv.org/abs/1911.01174)
50. Hadfield S, Papageorgiou A. 2018 Divide and conquer approach to quantum Hamiltonian simulation. *New J. Phys.* **20**, 043003. (doi:10.1088/1367-2630/aab1ef)
51. Subramanian R, Zhang H. 2018 Performance analysis of divide-and-conquer strategies for large scale simulations in R. In *2018 IEEE Int. Conf. on Big Data (Big Data)*, pp. 4261–4267. New York, NY: IEEE.
52. Lintusaari J, Gutmann MU, Dutta R, Kaski S, Corander J. 2017 Fundamentals and recent developments in approximate Bayesian computation. *Syst. Biol.* **66**, e66–e82. (doi:10.1093/sysbio/syw077)
53. Gouil Q, Keniry A. 2019 Latest techniques to study DNA methylation. *Essays Biochem.* **63**, 639–648. (doi:10.1042/EBC20190027)
54. Xu C, Corces VG. 2018 Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science* **359**, 1166–1170. (doi:10.1126/science.aan5480)
55. Charlton J *et al.* 2018 Global delay in nascent strand DNA methylation. *Nat. Struct. Mol. Biol.* **25**, 327–332. (doi:10.1038/s41594-018-0046-4)
56. Ginno PA *et al.* 2020 A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.* **11**, 2680. (doi:10.1038/s41467-020-16354-x)