

Agent-based modelling and time series inference of filamentous yeast colonies

Kai Li¹, Alexander K. Y. Tam², Jennifer M. Gardner³, Joanna F. Sundstrom³, Vladimir Jiranek^{3,4,*}, J. Edward F. Green¹, Benjamin J. Binder¹, and Andrew J. Black¹

¹School of Computer and Mathematical Sciences, University of Adelaide, Adelaide SA 5005, Australia

²UniSA STEM, The University of South Australia, Mawson Lakes SA 5095, Australia

³Discipline of Wine Science, Waite Campus, University of Adelaide, Urrbrae SA, 5064, Australia

⁴School of Biological Sciences, The University of Southampton, Southampton, SO17 1BJ, United Kingdom

*Correspondence: V.Jiranek@soton.ac.uk

Abstract: The baker's yeast *Saccharomyces cerevisiae* can form invasive filamentous colonies with non-uniform spatiotemporal patterns. We aimed to better understand how individual cellular actions give rise to colony-scale patterns. We used an off-lattice agent-based model to simulate colony growth, and used a neural likelihood estimator (NLE) to infer parameters for experimental photographs. Li et al. [1] used approximate Bayesian computation (ABC) to infer parameters using coarse summary statistics obtained from a single time point averaged across experimental replicates. The NLE overcomes the computational expense of ABC, allowing us to infer the parameters of individual colonies from a full time series of experimental photographs. To demonstrate the capabilities of our model and inference technique, we tested extensively on synthetic data and then predicted yeast growth under three different experimental conditions. As before, the proportion of total colony growth above which pseudohyphal growth is permitted was a key parameter that contributed to colony morphology. Since our NLE-based approach incorporates time series data, it yielded better parameter estimates and more accurate predictions compared to our ABC-based method. This updated approach improved understanding of how the probability that a stated cell produces a pseudohyphal cell influences colony morphology. In this way, the model also has the potential to generate hypotheses, which can be tested through biological experiments to increase the understanding of the basis for different growth patterns in yeast.

Keywords: *Saccharomyces cerevisiae*, off-lattice model, neural likelihood estimator, parameter inference.

1 Introduction

Yeasts are among the most versatile microorganisms studied in biology. Due to their low cost and ease of culturing, yeasts are often used as a model organism for other eukaryotic cells in scientific research [2]. With over 1,200 known species, yeasts also play a vital role in human life. For example, the yeast *Candida albicans* is an opportunistic pathogen that can cause serious infections in humans [3]. These infections are underpinned by yeast's ability to invade surrounding tissue. Yeasts also have positive impacts. For example, the baker's yeast *Saccharomyces cerevisiae* is used in the production of baked goods and alcoholic beverages [4]. Other positive applications of yeasts include in the biotechnology industry for biofuels [5], in the health industry for drug discovery [6] and in advancing genome research [7]. There is also ongoing research in optimising and developing new strains for producing specialised wines [8], for example the AWRI 796 and Simi White strains involved with this work. Hence, there is great motivation in studying yeast.

Filamentous *S. cerevisiae* colonies consist of two cell types [9], which we term sated and pseudohyphal. Sated cells are ellipsoids with average major axis lengths of 4.2 μm and minor axis lengths of 3.0 μm . Pseudohyphal cells are also ellipsoidal, but are elongated compared to sated cells, with axis lengths of 6.7 μm and 1.9 μm [1, 9]. Sated cells bud from sites located near both ends of the major axis, whereas pseudohyphal cells bud from the distal pole only [10]. Budding is an asexual process driven by nutrients. In filamentous colony experiments, nutrient diffusion is much faster than cell proliferation [11]. Consequently, the nutrient concentration is effectively uniform in space, but depletes over time due to consumption by cells. Yeast cells are non-motile, and since nutrients are spatially uniform, they do not undergo directed growth [11], as the roughly symmetrical colonies of Figure 1.1 indicate. During early growth, when nutrients are abundant, *S. cerevisiae* colonies exhibit spatially uniform Eden-like circular growth. This circular growth consists of sated cells only, and is shown in Figure 1.1(a)–(b). However, when nutrients deplete the colony produces elongated pseudohyphae (Figure 1.1(c)–(h)) in an attempt to access more nutrients. Pseudohyphal growth also occurs underneath the colonies, where cells invade their substratum [12]. This invasive phenomenon is important clinically, but in this work we restrict ourselves to two-dimensional surface growth.

In this paper, we use agent-based modelling to simulate colony-level morphologies arising under different experimental conditions. Agent-based models have been used extensively in biology to study cancer [13–15], tissue growth [16–18], molecular biology [19–21], as well as yeast [11, 22, 23]. Li et al. [1]

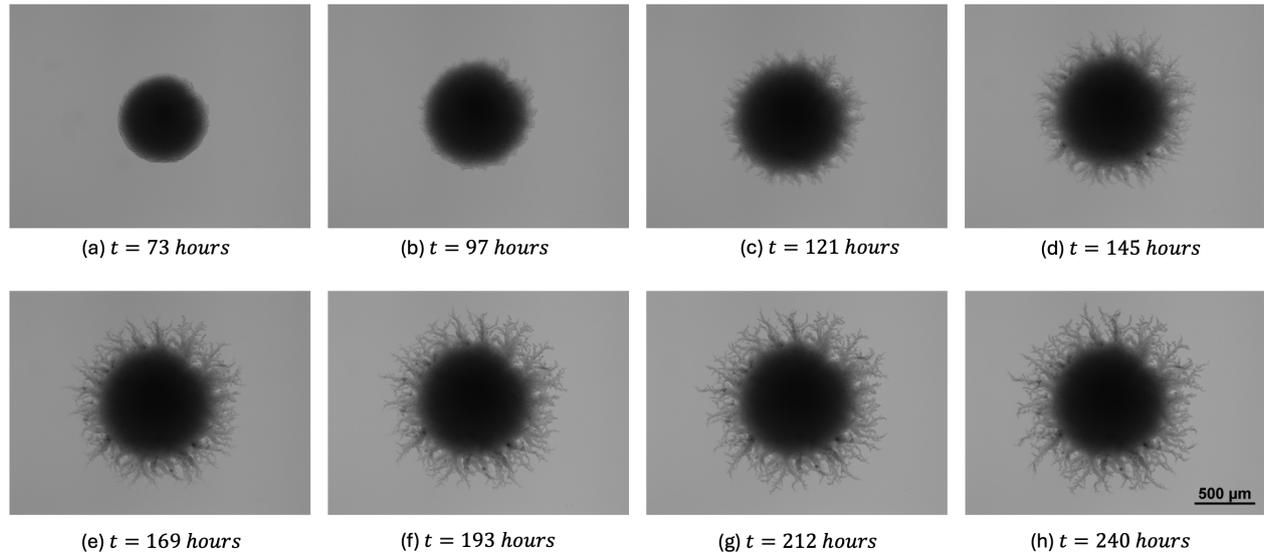


Figure 1.1: Time series images of a growing yeast colony of AWRI 796 50 μm showing the transition from Eden-like circular growth to filamentous growth.

used an off-lattice agent-based model, where the agent represents a cell, to capture filamentous yeast growth. Li et al. [1] confirmed the important role of a cell phenotype switch from sated to pseudohyphal in colony morphology. In our model, this switch occurs when the colony attains a threshold size, which promotes filamentation by depleting surrounding nutrients, and requiring more resources to sustain the larger colony. We return to the model of Li et al. [1] in this paper, with the key extension of using time series data for individual colonies to parameterise the model. This time series data consist of a series of 8 photographs of individual colonies at different stages of development. Incorporating the additional data extends the work of Li et al. [1], who used data from a single time point at the end of the experiment, and summary statistics averaged over experimental replicates. Analysing the time series data reveals that the probability that a sated cell produces a pseudohyphal cell is a key determinant of colony morphology, which was not fully appreciated using data from only a single time slice.

Since approximate Bayesian computation (ABC) [24] is expensive and scales poorly to large data sets, our time series data requires a more computationally efficient inference method. Hence, in this work we perform inference using neural likelihood estimation (NLE) [25–27]. NLE is an inference method that uses a conditional density approximation parameterized by a neural network to approximate the likelihood function. A key advantage of neural likelihood methods is that they require significantly fewer simulations to train the model compared to doing inference with ABC [25], and can more

easily exploit parallel computing resources. NLE is most useful where the likelihood is intractable and model simulation is expensive, and has found applications in economics [28], cosmology [26] and neuroscience [29]. Since neural methods drastically speed up parameter inference compared to ABC, in this paper we explore the variability of parameter values within an entire set of 14 individual experiments for a given yeast strain and initial nutrient level. This process required less than a day of computational time, whereas with the ABC method this would have taken many months due to the high computational cost of simulation, image processing, and parameter inference.

2 Materials and Methods

Our approach combines yeast growth experiments, mathematical modelling, image processing, and parameter inference. Subsequent subsections outline our methods for each of these steps.

2.1 Yeast Growth Experiments

We use the same experimental set-up as described by Li et al. [1]. The experiments are performed using the AWRI 796 strain at two initial ammonium sulfate nutrient concentrations (AWRI 796 50 μM and 500 μM), and the Simi White strain at 50 μM . These are two commercially available strains of *S. cerevisiae* used in wine production. While colonies are growing, we take a photograph of each daily, from Day 3 (73 hours) until Day 10 (233 hours). Example photographs from one replicate are shown in Figure 1.1. We have photographs for 14 individual colonies for each set of experimental conditions. When referring to a set of experiments, we name them according to yeast strain and nutrient level. For example, AWRI 796 50 μM refers to experiments of the yeast strain AWRI 796, with an initial nutrient concentration of 50 μM .

2.2 Mathematical Model

We use the two-dimensional off-lattice agent-based model for a filamentous yeast colony developed by Li et al. [1]. The model contains two cell types: sated and pseudohyphal, and assumes that all cells of a given type are ellipses with the same dimensions: axis lengths of 4.2 μm and 3.0 μm for sated cells, and 6.7 μm and 1.9 μm for pseudohyphal cells. In practice, we approximate both cell types as dodecagons. We simulate colony growth one cell division event at a time starting from a single cell, until the number of cells in the colony, n , reaches a maximum value prescribed in advance, n_{max} . This

value n_{\max} is chosen to approximate the number of cells in the experimental or synthetic colony that we seek to simulate. Since we do not model nutrients explicitly, this assumption enables us to neglect time, removing the need to have unknown proliferation rates as additional parameters to be inferred. Sated cells proliferate from one of 4 possible budding sites located at both ends of the cell, whereas pseudohyphal cells can only bud from one of 2 sites located at prescribed angles from the distal pole. A schematic of the two cell types and all possible cell proliferation events in the model is shown in Figure 2.1. Sated cells are indicated in green, and pseudohyphal cells are indicated in purple. Our time-free model contains five key parameters that influence filamentous yeast growth, each taking values on the unit interval $[0, 1]$. In vector form, we write the parameters as $\theta = (n^*, p_a, p_{sp}, p_{ps}, \gamma)$, and describe their meanings below.

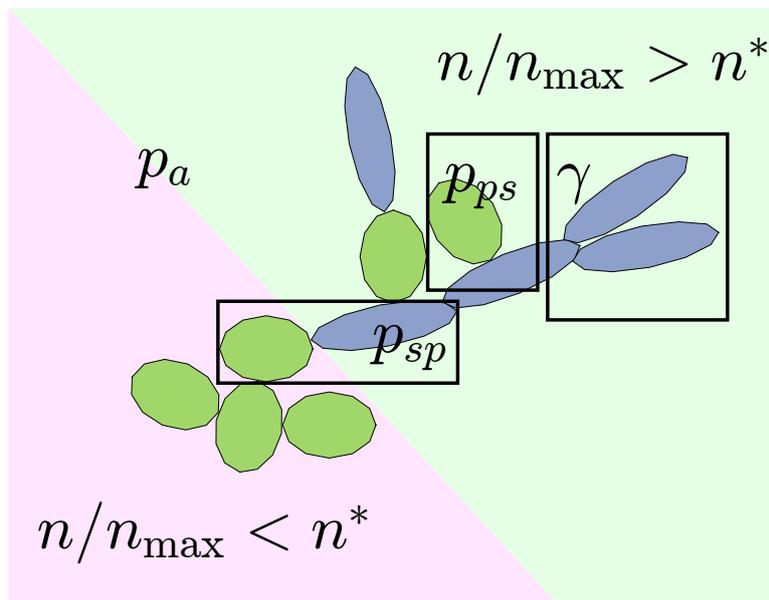


Figure 2.1: Budding patterns in the agent-based model. All cells are sated (green dodecagons) until the threshold $n > n^*n_{\max}$ is attained. After this threshold, the first pseudohyphal cell (blue dodecagons) can be produced. Then, more sated and pseudohyphal cells are produced via a set of proliferation events.

For each individual colony simulation, we assume a threshold number of cells, n^*n_{\max} , below which all cells are sated and no pseudohyphal growth occurs. This sated-only growth is shown in the pink region of Figure 2.1. If the current number of cells in the simulation $n > n^*n_{\max}$, pseudohyphal growth becomes possible. We then implement a series of steps to determine the next cell to proliferate, and the type of daughter cell produced. If one or more pseudohyphal cells are present, then the next cell to proliferate will be a sated cell with probability p_a , or a pseudohyphal cell with probability $1 - p_a$. Once a cell type is chosen, the proliferating cell is chosen at random from all cells of that type. If

a sated cell is chosen, it will produce a pseudohyphal daughter cell with probability p_{sp} , or a sated cell with probability $1 - p_{sp}$. If a pseudohyphal cell is chosen and does not have any daughter cells, then a pseudohyphal daughter cell is produced from the distal pole. However, if the pseudohyphal cell chosen to proliferate already has an existing pseudohyphal daughter cell, its behaviour is controlled by two parameters, γ and p_{ps} . If the selected pseudohyphal cell has a pseudohyphal daughter, we abort proliferation with probability γ and instead select a new pseudohyphal mother cell with no pseudohyphal daughter to proliferate. Larger values of the parameter γ limit forking, which is the scenario where multiple pseudohyphal branches emanate from the same pseudohyphal mother. Proliferation proceeds with probability $1 - \gamma$, and the daughter cell is sated with probability p_{ps} , and another pseudohyphal cell with probability $1 - p_{ps}$. If at any stage in the simulation the cell chosen to proliferate has no budding site available or if the centre of the daughter cell would lie within the boundary of another cell, proliferation is aborted, and the procedure begins again from the start.

2.3 Image Processing and Summary Statistics

Image processing and quantification enabled us to compare experimental and simulation results. The first step was to convert grayscale photographs and simulation plots to binary images. We obtained the binary images using the Tool for Analysis of the Morphology of Microbial Colonies (TAMMiCol) [30], which uses automatic thresholding and interpolation to obtain binary images. We saved simulation images at the same pixel resolution as experimental photographs. Computing summary statistics for the spatial patterns in images enabled us to compare simulated and experimental images. We compute six summary statistics per image: colony area, maximum radius, mean radius, branch length, filamentous area, and compactness. These summary statistics quantify the size, extent of filamentous growth, and spatial pattern of the filaments.

The colony area (I_A) is the total number of occupied pixels in the binary image, which is readily obtained in MATLAB. To compute the maximum and mean radii, we first used MATLAB to locate the colony centroid. The mean radius ($I_{R_{\text{mean}}}$) is the mean distance between the centroid and any pixel located on the colony perimeter, and the maximum radius ($I_{R_{\text{max}}}$) is the maximum distance from the centroid to an occupied pixel. The mean radius indicates the amount of the colony occupied by the central circular region of Eden-like growth. We achieved better results using the mean radius (see Figure 2.2(b)) to quantify the size of the circular region than the complete spatial randomness [31] and minimum [1] radii, which we have used in our previous work.

The filamentous area (I_{AF}) is the number of occupied pixels in the annular region between the mean radius and the maximum radius. Filamentous area quantifies the extent of filamentation but also accounts for the density of filaments. To obtain the branch length (I_B), we summed the length of branches of a skeletonised experimental image, an example of which is shown in Figure 2.2c. MATLAB's `bwmorph()` function provides a count of the number of branches in a skeletonised image of the filamentous region. Due to the computational cost of skeletonisation, we only counted a quarter of the colony's branches. Finally, compactness (I_C) is the colony's perimeter-to-area ratio, P/A , normalised by the perimeter-to-area ratio of a circle with area A . Compactness therefore quantifies the irregularity of the colony shape, such that a perfect circle has the smallest possible value $I_C = 1$, and deviations from the circle result in $I_C > 1$. Using the colony perimeter (see Figure 2.2(a)) and area obtained from MATLAB, we compute the compactness using $I_C = P^2/(4\pi A)$.

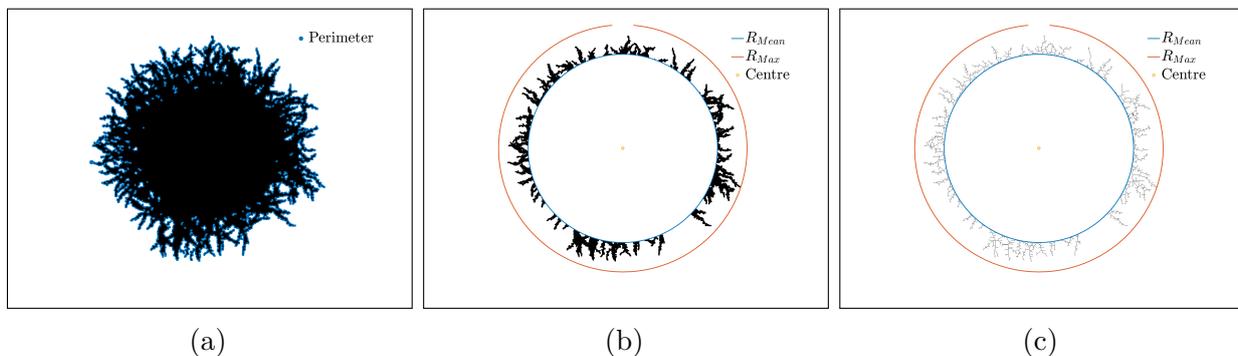


Figure 2.2: Visualisation of summary statistics for a binary image from a simulated colony. (a) Binary image with perimeter overlaid. (b) Processed binary image indicating the colony centroid, maximum radius, mean radius, and filamentous area. (c) Skeletonized binary image of the filamentous region, which was used to compute the branch length.

To show how these summary statistics vary throughout experiments, we plotted each summary statistic from Day 3 (~ 73 hours) until Day 10 (~ 233 hours). An example of these six summary statistics evaluated on time series data of AWRI 796 $50\ \mu\text{M}$ colonies is given in Figure 2.3. Each semi-transparent line represents an experimental colony, with 14 colonies in total. We also overlay these with box plots at each recorded time point to show the spread of the data, with the red crosses indicating outliers.

2.4 Parameter inference

Given observed data \mathbf{t}' from a series of experimental images, our goal was to estimate the posterior distribution of the parameters $p(\boldsymbol{\theta}|\mathbf{t}') \propto p(\mathbf{t}'|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta})$ is the prior distribution and $p(\mathbf{t}'|\boldsymbol{\theta})$ is the likelihood. The simulation described in the previous section gives a way of sampling from the

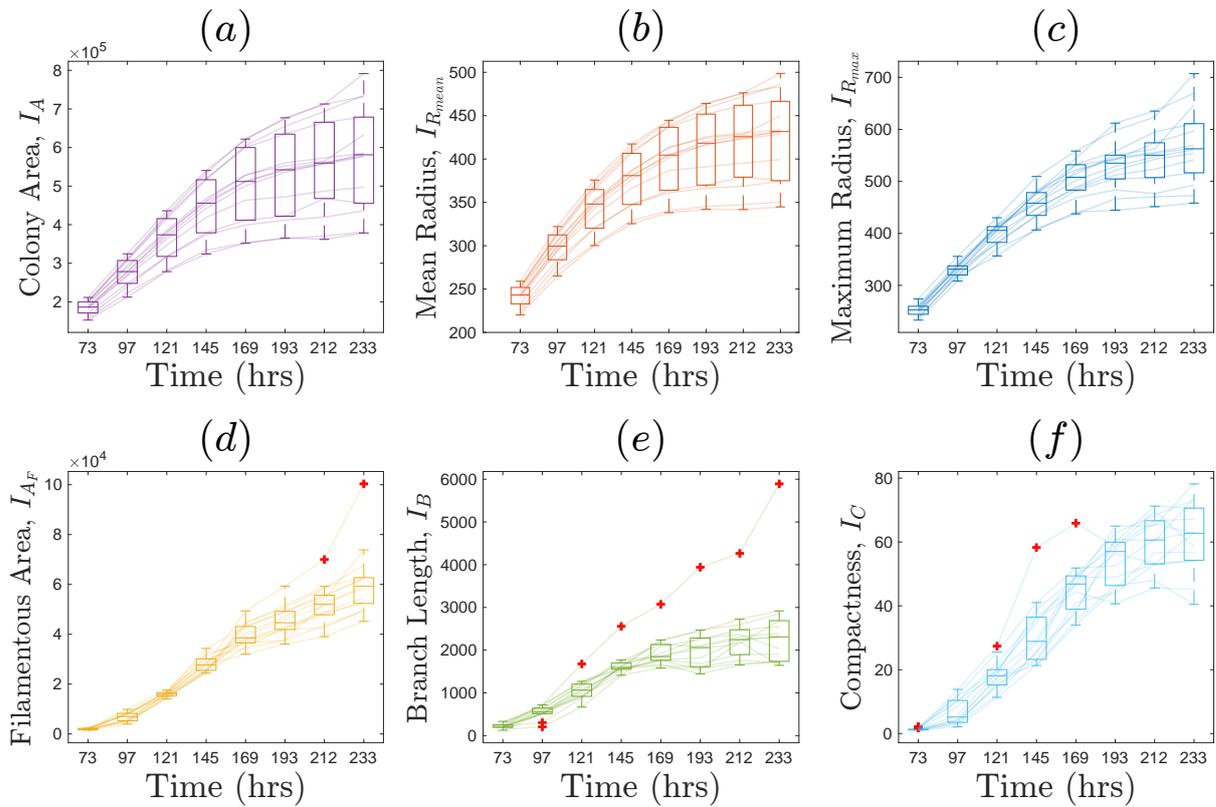


Figure 2.3: Summary statistics used to quantify AWRI 796 50 μm colonies over time. (a) Colony area, I_A . (b) Mean radius, $I_{R_{mean}}$. (c) Maximum radius, $I_{R_{max}}$. (d) Filamentous area, I_{A_F} . (e) Branch length, I_B . (f) Compactness, I_C .

distribution $p(\mathbf{t}|\boldsymbol{\theta})$, but actually evaluating it is intractable. NLE instead takes advantage of this ability to simulate the model to learn an approximation to the likelihood function. We can then sample from the posterior using standard Markov chain Monte Carlo (MCMC) methods, by replacing the true likelihood with the approximation.

Starting from a set of samples $\{\boldsymbol{\theta}_i, \mathbf{t}_i\}_{i=1:N}$ that are obtained by $\boldsymbol{\theta}_i \sim \tilde{p}(\boldsymbol{\theta})$ and $\mathbf{t}_i \sim p(\mathbf{t}|\boldsymbol{\theta}_i)$, we trained a conditional neural density estimator $q_w(\mathbf{t}|\boldsymbol{\theta})$, where w are the weights of a neural network. The proposal distribution $\tilde{p}(\boldsymbol{\theta})$ controls where the data is generated, and therefore how accurate the approximation is. This introduces some circularity to the problem as we would like $\tilde{p}(\boldsymbol{\theta})$ to concentrate in areas of high posterior density, but we do not know this *a priori*. To address the issue of circularity, current state-of-the-art methods take a sequential approach to generating training data and refining the proposal [25]. Our testing revealed that this sequential approach was unnecessary for our problem, and good results could be obtained from a single round of training on a sufficiently large dataset. This is probably due to the simple shape and limited support of the posterior distributions in our problem.

The conditional neural density estimator that we use in this work is a Gaussian mixture density network (MDN). The conditional density takes the form

$$q_w(\mathbf{t}|\boldsymbol{\theta}) = \sum_{j=1}^m \omega_j(\boldsymbol{\theta}; w) q_j(\mathbf{t}|\boldsymbol{\theta}; w), \quad (2.1)$$

where $\omega_j(\boldsymbol{\theta}; w)$ are the mixture weights such that $\sum_j \omega_j = 1$ and $q_j(\mathbf{t}|\boldsymbol{\theta}; w)$ are multivariate normal densities $q_j(\mathbf{t}|\boldsymbol{\theta}; w) = \mathcal{N}(\boldsymbol{\mu}_j(\boldsymbol{\theta}; w), \Sigma_j(\boldsymbol{\theta}; w))$. The weights, means and covariances are all functions of the parameters $\boldsymbol{\theta}$, which are in turn determined by the weights of the neural network w [32]. We implement the MDN using the Julia package `MixtureDensityNetwork.jl`. Below are detailed steps for how we trained and sampled from the models.

The parameter vector $\boldsymbol{\theta}$ has five elements, and each parameter takes values on the unit interval. We generate the training dataset from independent uniform distributions, *i.e.* $\tilde{p}(\boldsymbol{\theta}) = 1$, $\boldsymbol{\theta} \in [0, 1]^5$. The domain of the MDN is \mathbb{R}^5 , so we apply a logit transform to the samples such that the transformed parameters take values on an unrestricted domain when training the model. Note that the Metropolis-Hastings (MH) sampling discussed later is also performed on this unrestricted space, and then samples are transformed back to $[0, 1]^5$ to visualise the final posteriors. This requires the addition of Jacobian terms in the calculation of the MH acceptance probabilities that are detailed in Appendix A.1.

We define the components of the observed data \mathbf{t} as

$$\mathbf{t}_{(i,j)} = \left(I_A, I_{R_{\text{mean}}, I_{R_{\text{max}}}}, I_{AF}, I_B, I_C \right)_j. \quad (2.2)$$

where i indexes the six different summary statistics $[I_A, I_{R_{\text{mean}}, I_{R_{\text{max}}}}, I_{AF}, I_B, I_C]$ and $j = 1, \dots, T_F$ is over the number of time points. Typically, we set $T_F = 8$ as there is a maximum of eight time course images per colony. For example, to interpret the component of the vector \mathbf{t} , we take $\mathbf{t}_{(1,8)}$, which is the area of the colony at time point eight. As the summary statistics vary by different orders of magnitude, as shown in Figure 2.3, we normalised within each summary statistic by dividing by the maximum value of each time series summary statistic from the simulations of the prior. This ensures all summary statistics are close to the unit interval while preserving the trend of the time series data. Note that for a set of simulations, we simulate up to a target cell number which is prescribed in advance based on the experimental colony area. Hence, the summary statistics I_A will be approximately the same for all simulations and the target colony.

The architecture of the neural networks and the number of mixtures were determined by splitting the data into a 90/10 training/test set and performing a number of experiments monitoring the log likelihood. From this we found that training the MDN using 600 epochs, a batch size of 16, two hidden layers of depth 64 each, and a mixture of six yielded good results across all experimental and synthetic conditions, so this architecture was fixed. Throughout we use the Adam optimiser with a learning rate of 1×10^{-3} . The MDN model was trained on 1000 simulated data points. This number of simulations was chosen because they could be completed on the available HPC resources in less than 24 hours. This lent itself to a convenient workflow for running and analysing results without requiring a sequential method, which although probably a more efficient use of resources requires more care in tuning [25].

After learning the likelihood function, we sample from the posterior distribution using a basic MH algorithm. The proposals were independent Normal distributions with a standard deviation of 1. We took 200,000 samples to estimate the posterior distribution and remove the first 10,000 samples of burn-in. Thinning was also applied to the chains by taking every tenth parameter value to reduce autocorrelation. This part of the procedure is very fast, so it did not require any substantial tuning. As discussed above, the final step was to apply an inverse logit transform to the samples to return them to the unit interval.

3 Results

3.1 Synthetic data set

To test how well our inference algorithm could uncover the true parameter values, we simulated synthetic datasets and performed inference to recover the model parameters. We show one example here in the main text and another two in the Appendix. Instead of initialising the synthetic test colony from random parameters, we used the parameter values obtained by Li et al. [1] for AWRI 796 50 μM . We did this because non-physical colony morphologies will result from many combinations of random parameters. Since our summary statistics were designed with the specific aim of capturing the morphology of filamentous yeast colonies, they work best when the colonies are similar to the experimental colonies. Therefore, non-physical colonies are not suitable for the analysis. For the example shown here, the parameters used were $n^* = 0.67$ $p_{ps} = 0.58$ $p_{sp} = 0.25$, $\gamma = 0.12$, $p_a = 0.14$. We simulate the synthetic colony up to the same area as observed in those experiments for AWRI 796 50 μM sample one (see Figure 1.1) and we use all six summary statistics at all eight time points, providing 48 summary statistics per colony.

A pair plot of the posterior distribution is shown in Figure 3.1, where the ground truth parameters are shown with blue lines. The proportion of total colony growth above which we permit pseudohyphal growth, n^* , could be recovered with a small variance. This parameter has an important impact on colony morphology, as it determines when filamentation can begin. For other parameters, such as p_{ps} and p_{sp} , the true values lie in the tails of the distribution. The forking parameter γ and the probability of biasing sated cell proliferation, p_a , lie near the mode of the posterior. These parameters could be recovered with some uncertainty due to the large variance in the distributions.

To check the performance of the inference we also simulated colonies by sampling parameters from the prior and posterior distributions and created ribbon plots of the summary statistics over time. As expected, the synthetic datasets lie within both the prior (Figure 3.2a) and posterior (Figure 3.2b) predictive intervals as the realisation is drawn from the prior distribution. After approximately 125 hours, filamentous area, branch length, and compactness increased sharply. These increases probably occur because $n > n^*_{n_{\max}}$ after this time, initiating filamentous growth. Moreover, the posterior prediction distribution is centred around the true summary statistics and has much less variability than the prior predictive distribution. This demonstrates that the inference is performing as expected,

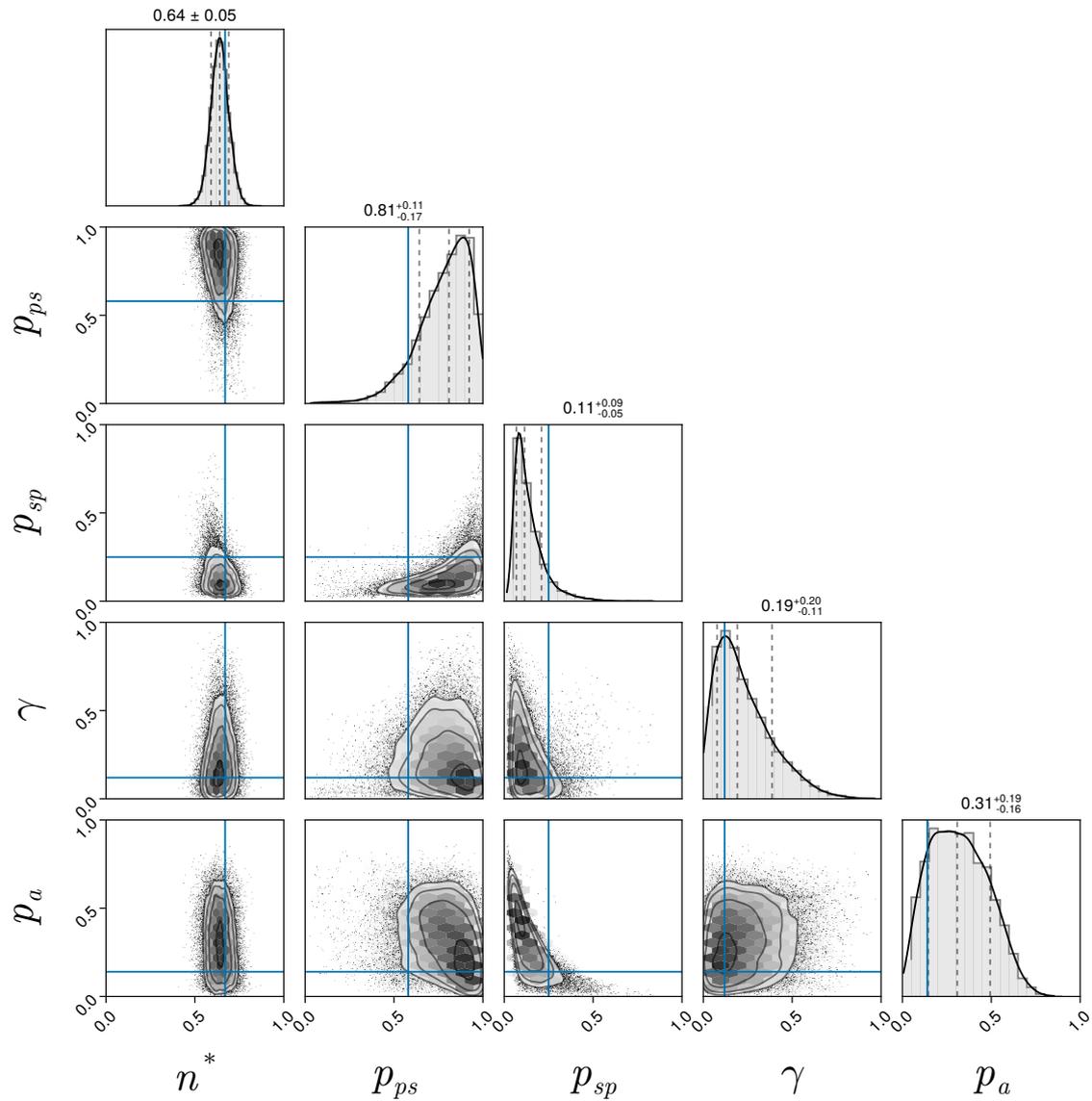


Figure 3.1: Pair plots of posterior distribution for the synthetic data. The blue line represents the ground truth parameter values. $n^* = 0.67$ $p_{ps} = 0.58$ $p_{sp} = 0.25$, $\gamma = 0.12$, $p_a = 0.14$.

even though some of the parameters are only weakly inferred.

To further validate the parameter inference, we randomly sampled 10 sets of parameters from the posterior distribution and plotted images of the simulated colonies through time to provide a visual comparison with the colony used for the inference. The 10 synthetic colonies shared very similar morphologies to the original data (see Appendix Figure A.5). The yeast colony model is stochastic so there is some natural variability in the summary statistics for a fixed set of parameters. This can be seen in Figure A.2, which shows ten realisations grown to the same area for the same fixed parameter values. We investigated how this variability affects the inferred parameters by rerunning the inference on these trajectories, and the posteriors were broadly similar.

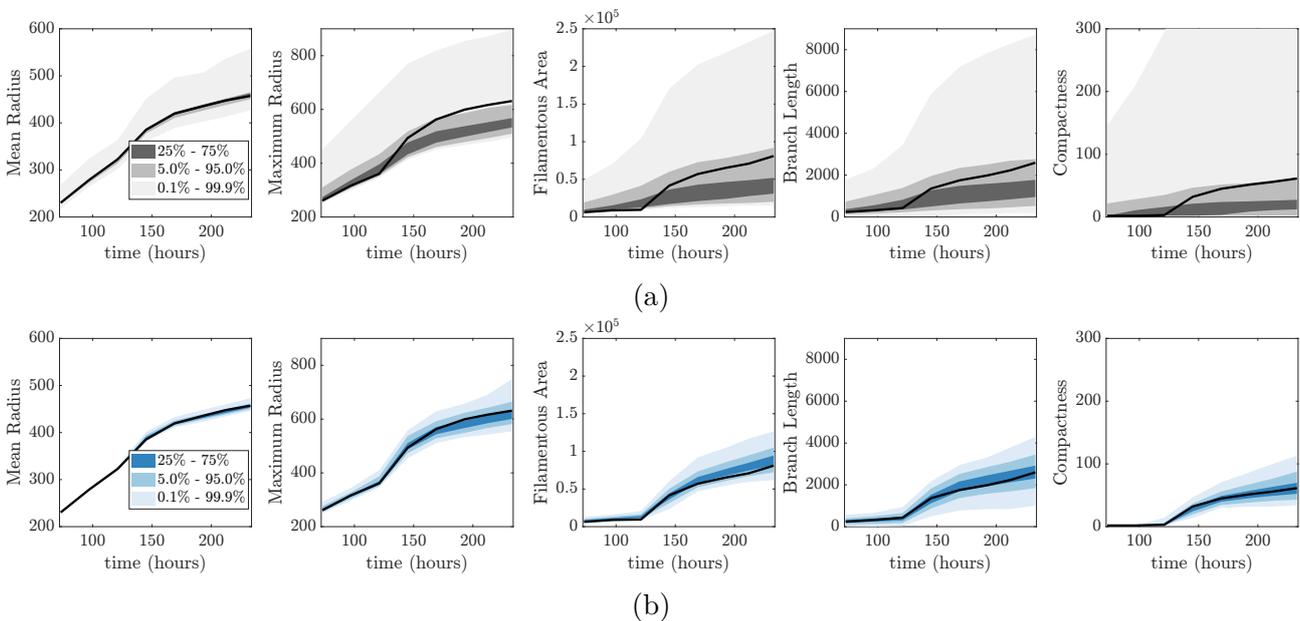


Figure 3.2: Ribbon plots showing the prior (a) and posterior (b) predictive distribution for the synthetic dataset. The black lines shows the realisation used for the inference.

We further tested our inference on two more synthetic data sets with different parameter values to test the robustness of our inference algorithm. For these data sets we used smaller values of n^* (0.28 and 0.39) as this will induce filamentation earlier compared to the first synthetic data set; other parameters were also varied. The posterior pair plots for both are shown in Figure A.1a. We also present their respective predictive distribution in Figure A.3. Again, we could retrieve n^* to a high degree of accuracy, while the other four parameters can also be somewhat inferred with the ground truth near the mode of the posterior distributions. For both $n^* = 0.28$ and $n^* = 0.39$, the ground truth summary statistic values (solid black line) lie within the posterior predictive intervals. Moreover, due to the lower variance of posterior distributions compared to the example shown above, we see

much narrower posterior predictive intervals compared to the prior predictive intervals. For a smaller n^* , the other parameters are more identifiable. This is because a smaller value of n^* triggers earlier filamentation and since many of the summary statistics quantify filamentation, this makes the other parameters more identifiable.

3.1.1 Increasing the number of summary statistics improves parameter inference

We next investigated how the addition of time-series data improved the parameter inference over just using a single image at the end of the growth period as in Li et al. [1]. We reran the inference on the synthetic data using data from just the last time point (using the notation defined in Section 2.4 this is $\mathbf{t}_{(i,8)}$, for all $i = 1, \dots, 6$), two time points ($\mathbf{t}_{(i,4)}$ and $\mathbf{t}_{(i,8)}$), and all eight ($\mathbf{t}_{(i,j)}$, for all $j = 1, \dots, 8$). The marginal posteriors for this experiment are shown in Figure 3.3.

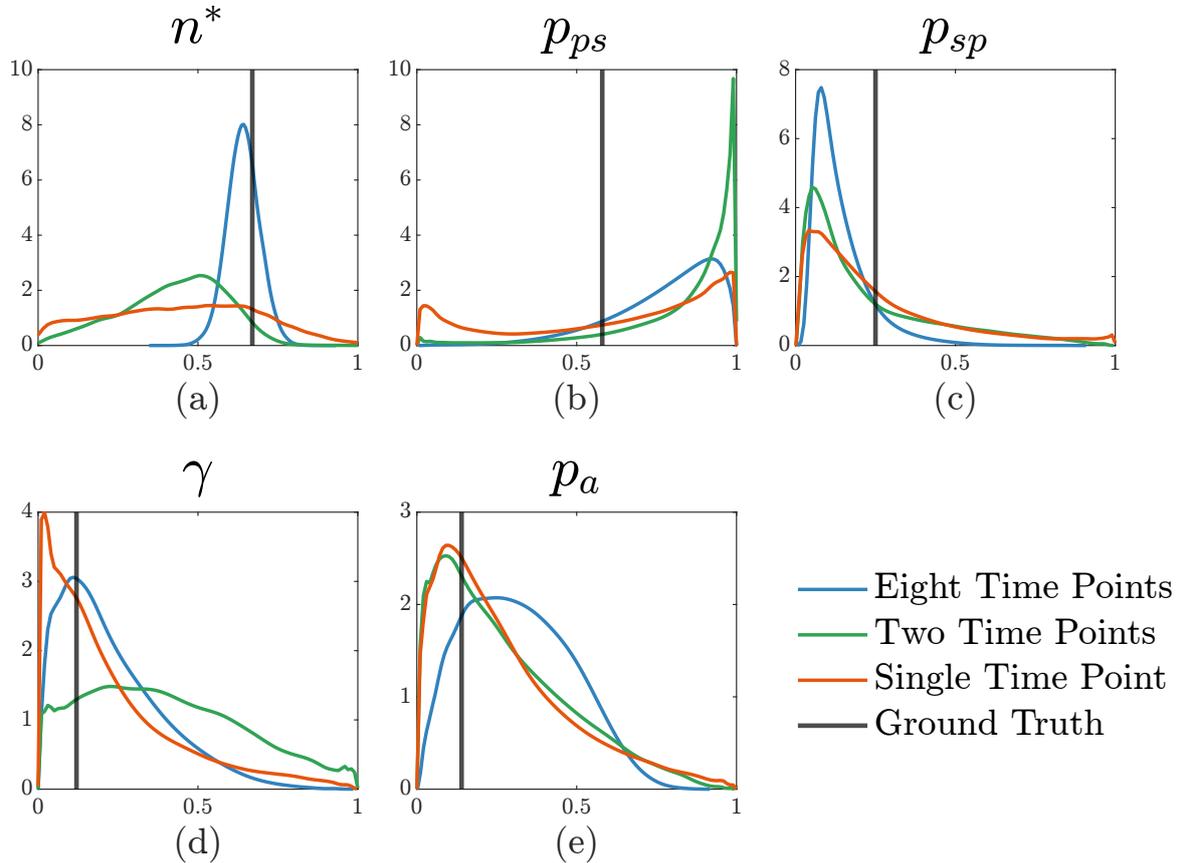


Figure 3.3: Marginal posterior distributions from performing inference using data from an increasing number of time points.

The most clear result is that the posterior for n^* contracts sharply with more data becoming much more informative. Finer resolution data allows the model to narrow down when filamentation begins, allowing for a much better estimate of n^* . Similarly, γ is more identifiable as the number of summary

statistics increases. This is evident by a decrease in the variance and the ground truth (blue vertical line) being situated around the mode of the distributions. For p_{sp} , there was little change in the mode of the distribution, with the true value still lying in the tail of the distribution. Moreover, p_a increased in uncertainty as the number of summary statistics increased. This could be because learning the higher-dimensional distribution is harder, or could be just due to the stochasticity in the time series. The parameter p_{ps} had large variability across all three types of summary statistics, suggesting this parameter is unidentifiable, which is consistent with the results of Li et al. [1].

Figure A.7 shows the posterior prediction ribbon plots for these experiments. These clearly show that more data improves the accuracy of the posterior predictions across all time points of the simulation. The effect of a tighter posterior distribution on n^* can be seen visually in Figure A.8. More data can accurately constrain the time at which filamentation begins producing better agreement near the start of the time series. Hence, we use all eight images when running the inference.

3.2 Experimental modelling and inference

Having tested our inference algorithm on synthetic datasets, we turn to the experimental colonies for the three experimental conditions. This allowed us to gain further insight into the biological mechanism contributing to the colony morphology development of *S. cerevisiae* at different stages in time. Overall, the model performs well in capturing the colony morphology. Figure 3.4 compares the experimental images with simulated colonies where parameters are randomly sampled from the corresponding posteriors. We also present an additional ten simulations in the Appendix A.4. Visually, the simulated colonies match very closely to the experimental colonies at all time points, indicating that our model predictions effectively capture the colony morphology.

We plot the prior and posterior predictive distributions as ribbon plots in Figure A.9. The posterior predictives encompass a majority of the experimental data within the 5 to 95% intervals. There are small discrepancies between the simulated realisations and experimental data that could indicate model misspecification [33]. For example, the maximum radius and filamentous area are on the boundary of the posterior predictive distributions at some time points. However, overall, the model is able to capture the experimental morphology for most summary statistics, with the posterior prediction residing closely around the true values of the summary statistics.

Figure 3.5 compares the marginal posterior distributions for the three strains. Pairplots of the posteriors

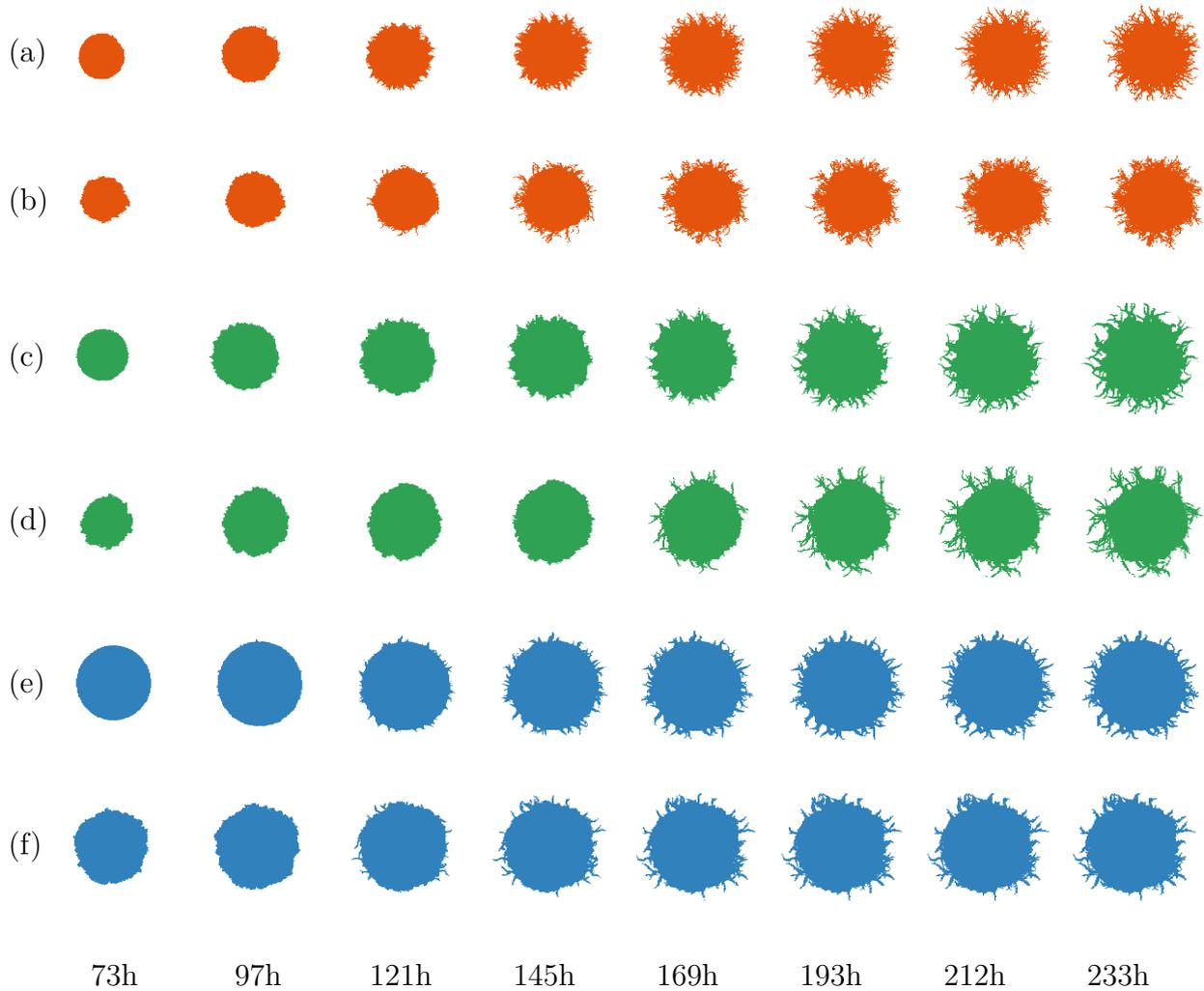


Figure 3.4: Colony simulation comparison between experiment and posterior prediction. (a) Experiment of AWRI 796 50 μM Sample 1 filamentous yeast. (c) Experiment of Simi White Sample 7 50 μM filamentous yeast. (e) Experiment of AWRI 796 500 μM Sample 9 filamentous yeast. (b,d,f) Respective simulations with parameter values sampled from the posterior distribution in Figure 3.5 with colours distinguishing between different experiments.

can be seen in Appendix A.6. As anticipated, n^* and p_{sp} were the parameters with the most informative

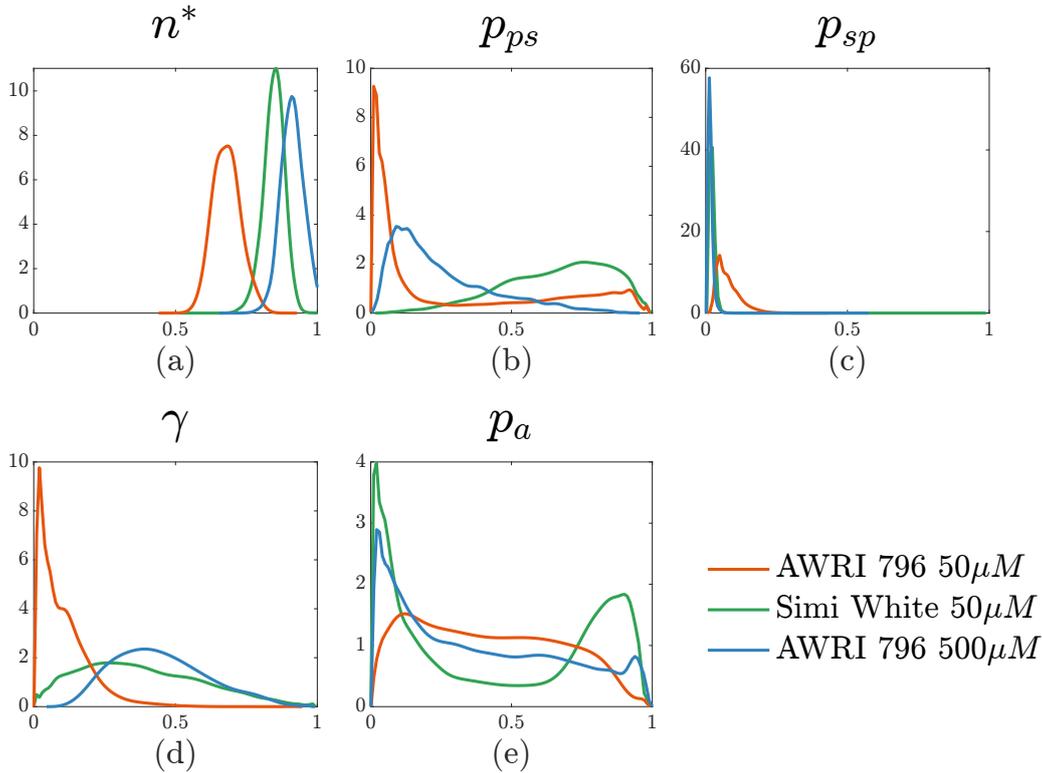


Figure 3.5: Comparison of the marginal posteriors for three experimental colonies. The colonies used here are AWRI 796 50 μM sample 1, AWRI 796 500 μM sample 8, and Simi White 50 μM sample 7.

posteriors. For example, comparing AWRI 796 50 μM with AWRI 796 500 μM , we saw a big difference in n^* which is expected as filamentation occurs earlier in the AWRI 796 50 μM than AWRI 796 500 μM due to the smaller initial nutrient level. Similarly, we expect p_{sp} to be small as the experimental colonies have a limited number of distinct filaments. Since the number of filaments is controlled by the number of pseudohyphal cells emerging from sated cells, smaller p_{sp} values will result in fewer filaments. Differences in p_{sp} between different experiments can be explained by the amount of filamentation in each experiment. AWRI 796 50 μM has more filaments than AWRI 796 500 μM , and thus the AWRI 796 50 μM colony has a larger p_{sp} . Therefore, although analysis of the synthetic data suggested that the NLE method may not recover the true value for p_{sp} , it does identify relative changes in this parameter due to experimental conditions. These differences in p_{sp} across experimental conditions were not detected using the ABC-based method and inference using data from a single time point [1].

The posterior for the forking control parameter, γ , is wide for Simi White and AWRI 796 500 μM , but right-skewed for AWRI 796 50 μM . This can be explained by the small values of p_{sp} for the Simi White and AWRI 796 500 μM , which limits the number of pseudohyphal cells being produced. Since

forking only applies to pseudohyphal cells, with fewer pseudohyphal cells the effect of γ is diminished. When p_{sp} takes larger values, such as in the AWRI 796 50 μm colonies with more filaments than other experiments, the forking parameter is more identifiable. Compared to Simi White 50 μm and AWRI 796 500 μm , the AWRI 796 50 μm colonies have more thin filament branches and increased forking. Therefore, the parameter γ , which limits forking, is lower for AWRI 796 50 μm than for the other two experimental conditions.

The posterior distribution for p_{ps} , the probability of a pseudohyphal cell producing a sated cell, is difficult to interpret, suggesting that this parameter is unidentifiable. To test how p_{ps} impacts morphology, we simulated colonies with varying the values of p_{ps} in Appendix Figure A.13. These show that, at least for this combination of parameters, changing p_{ps} had minimal impact on the overall colony morphology. Another uninformative parameter is p_a . This parameter controls the probability of choosing a sated cell to proliferate over a pseudohyphal cell after filamentation has been triggered. Hence, these two parameters p_{ps} and p_a affect the thickness of the filaments of the simulated colonies. However, currently our summary statistics do not capture the filament thickness, which could be the reason why these parameters are uninformative.

3.3 Variability within replicates of the same experimental conditions

Due to large variability in final colony area across replicates with the same experimental conditions (see Appendix Figure A.18), we also investigated the impact of colony area on the inferred posteriors. We selected three colonies with the smallest, median and maximum area from the AWRI 796 50 μm experiments and performed inference on these. We observed from the posterior distributions (see Figure 3.6) that different areas do impact our estimates of the parameters for our model. The mean value of n^* increases as the size of the colony increases. Moreover, the distributions of p_{sp} and γ are similar between colonies while p_{sp} and p_a are uninformative.

In Figure A.17, we present the prior and posterior predictive distributions. For the smallest colony, the mean radius, branch length and compactness lie within the 25–75% prediction intervals, while the maximum radius and filamentous area are less accurately predicted. As previously noted, this could be due to model misspecification. However, an interesting trend is that the uncertainty in the predictive intervals decreases for the larger colonies, with the intervals being more centred around the experimental summary statistics (black lines). We also visualise colonies drawn from the posterior

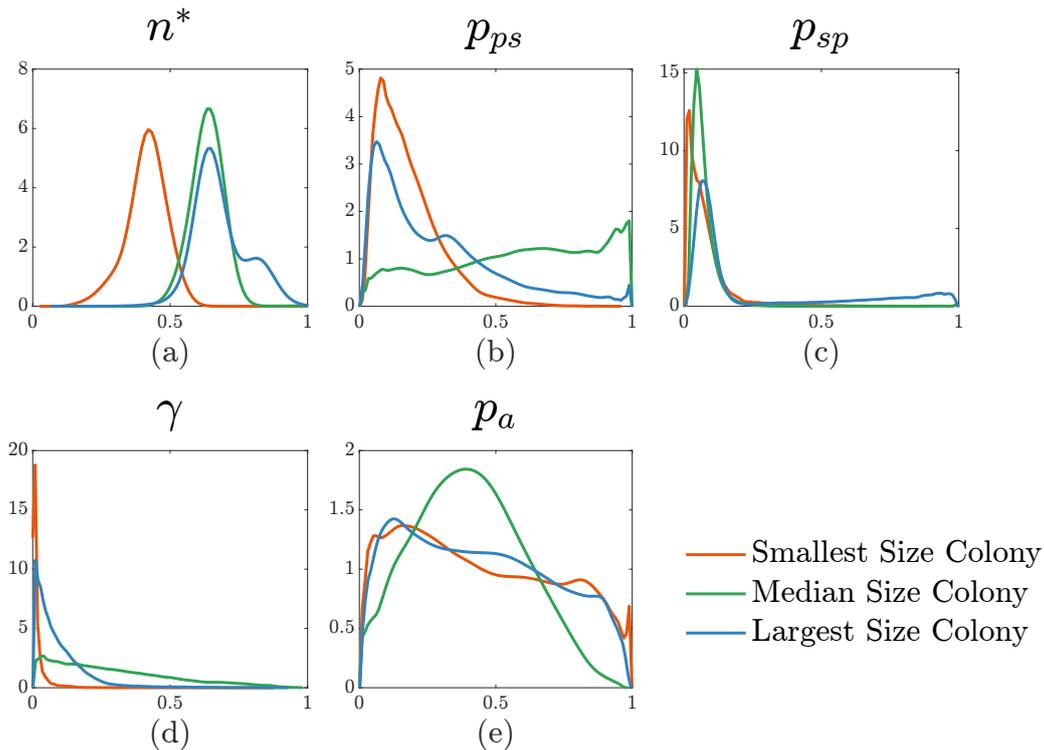


Figure 3.6: Comparison of posterior distributions of AWRI 796 50 μM colonies of different final areas. The smallest colony was sample 12, the median size was sample 14, and the largest was sample 9.

distributions as shown in Figure 3.7. As reflected in the predictive distributions, the smallest colonies appear the least similar to their experimental counterparts, while the simulations match best for the largest colony.

To further investigate the variability in parameter values within a set of experimental conditions, we performed inference on all 14 individual colonies of AWRI 796 50 μM . We then took the mean of each approximate posterior and compared these point estimates with those derived using ABC by Li et al. [1]. Box plots of the means are shown in Figure 3.8 along with the ABC results. The most impactful parameter on colony morphology, n^* , is similar to the value observed in the ABC inference. The variability in this parameter is also small, as expected. Similarly, the other parameters are in good agreement. It is important to note that the variability in the parameter values is expected and is due to the large variability of colony area across replicates of the same experimental conditions, as shown in Appendix A.8. In addition, we do not expect the two estimates of the parameter means to be the same due to different techniques being used, and the different priors used. That is, in the ABC, due to computational complexity, we chose weakly informative priors, whereas in the NLE, we used uniform priors. However, as both techniques gave similar results, it suggests the assumptions made by Li et al.

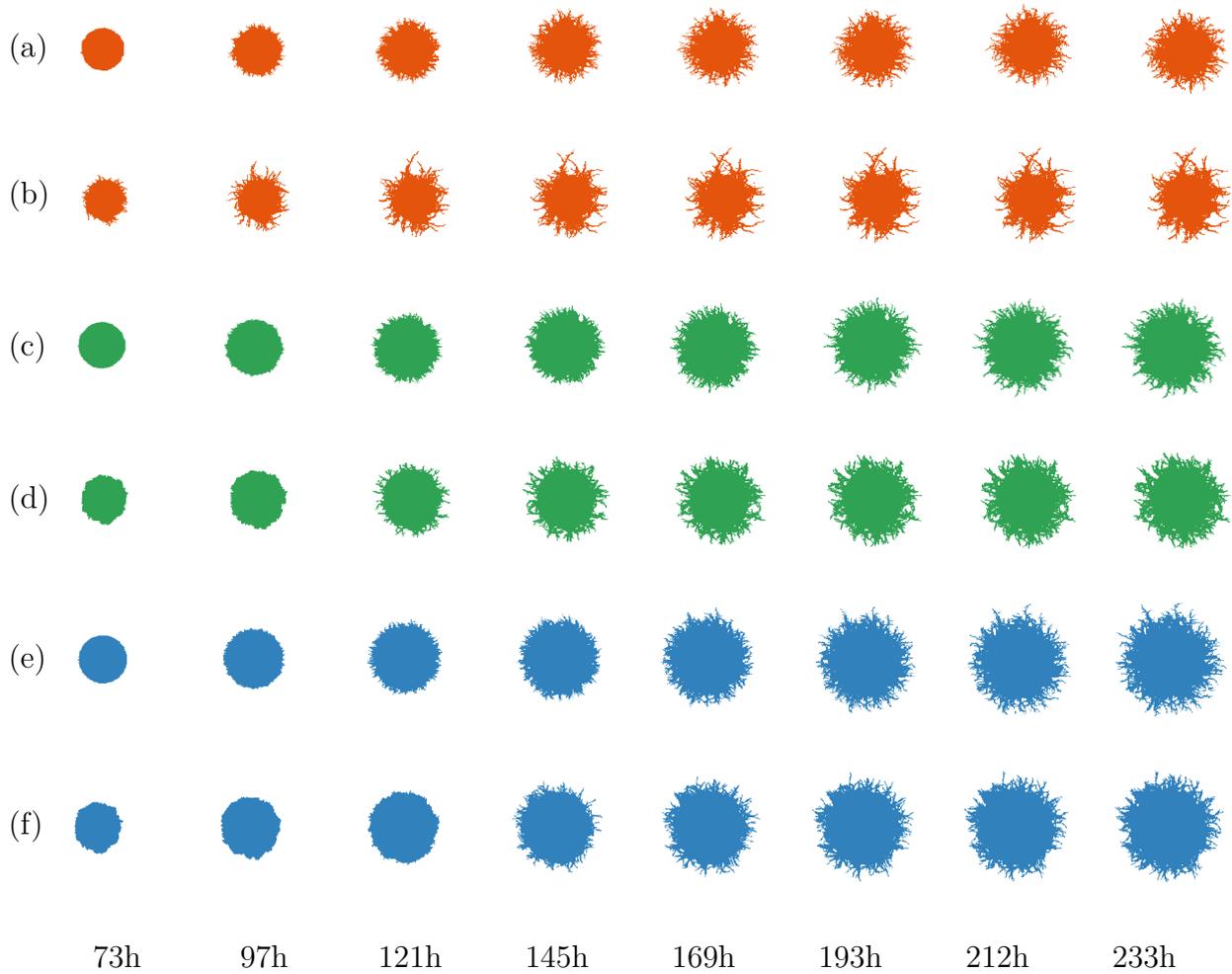


Figure 3.7: Simulation drawn from the posterior distribution in Figure 3.6 for different area colonies of AWRI 796 $50 \mu\text{m}$. (a) Sample 12, smallest size colony. (b) Simulation of Sample 12. (d) Experiment of Sample 14, medium-sized colony. (d) Simulation of sample 9 (e) Experiment of Sample 9, largest size colony. (f) Simulation of Sample 9.

[1] were reasonable.

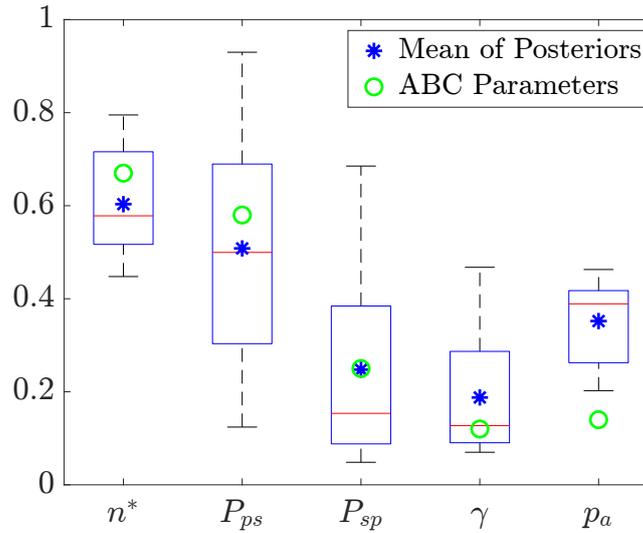


Figure 3.8: Boxplots derived from the means of the posterior distributions inferred from 14 experiments of AWRI 796 50 μM . The star represents the average of the means, the red bar represents the median value, and the circle markers represent the ABC mean parameters.

In this work, we have been comparing independent inferences that assume the parameters are different within the same experiment. We also investigated using a joint posterior approach to combine individual colony parameters within an experimental condition. This approach assumes that the parameter values are the same within a set of replicates of the same experimental condition. From the posterior predictive distribution, we found that the joint posterior approach performed poorly at capturing the colony morphologies compared to the individual inference shown in Figure A.17. This suggests that colonies within the same set of experiments are subject to random effects that change the individual parameters for each colony.

4 Discussion and Conclusion

We used neural likelihood estimation and a time series of photographs to infer parameters of an agent-based model for filamentous yeast colonies. We first verified our inference on synthetic data and found that using a full time series of 48 summary statistics (8 photographs, 6 summary statistics per photograph) provided the most informative posterior distributions. We showed our model could successfully capture the morphology of colonies grown in three experimental conditions. The most important determinants of colony morphology were n^* , the threshold size for filamentous growth, and p_{sp} , the probability of a sated cell producing a pseudohyphal cell.

We used a basic neural likelihood estimation (NLE) rather than sequential neural likelihood estimation (SNLE) that trains the model over a number of rounds. Initial testing using SNLE was performed where convergence was determined using the Euclidean distance between simulated (synthetic) parameter values and means of the posterior distributions. Convergence happened within a single round. Most likely, this was due to a relatively large (1,000 data points) sample size. However, if our sample size was small (100 data points), then our results would have benefited from SNLE. Another added benefit of using NLE was that simulation and inference took roughly one day compared to weeks for the ABC method used by Li et al. [1]. NLE also allows us to reuse simulations for independent colonies and run them up to specific areas, compute respective summary statistics, and thereby further reduce computational costs. Furthermore, the analysis in Section 3.3 required training 14 different models. Although this was not as computationally expensive as it first seems because some data could be reused, it would be better to train a single model that also takes the area as a parameter. Additionally, a different likelihood model, such as an autoregressive conditional density model [34], may better capture the dependencies in the time-series data than the Gaussian mixture model.

Some parameters of our model were unidentifiable, suggesting the possibility of simplifying the model to have fewer parameters while maintaining the ability to produce similar morphologies. Alternatively, we would need to add new summary statistics to better capture the aspects of the morphology controlled by these parameters. For example, the current statistics lack the ability to capture branch thickness. One potential improvement could be to use a hierarchical random effects model [35], which is most appropriate for analysis of multiple colonies from the same experiment. Fitting such a model is beyond the scope of this work, as it requires properly accounting for time and differences in initial conditions. There were small discrepancies between the predicted and experimental colonies, perhaps due to model misspecification. This possible misspecification suggests that we can make improvements to our technique, such as including a summary statistic that quantifies the thickness of individual filaments and an angular metric to quantify the spread of the filaments [31]. Future work could include developing a three-dimensional model of filamentous growth with growth in the z -direction invading the agar. Cells adhering to and invading the agar medium are of great interest to biologists because of their importance in pathogenic infections.

Acknowledgements

KL acknowledges funding from the Australian Government through a Research Training Programme Scholarship. BJB, JEEFG, AKYT, and VJ acknowledge funding from the Australian Research Council (Grant numbers DP230100406, DE240100097). This work was supported with supercomputing resources provided by the maths1 High Performance Computing service at the University of Adelaide.

Data, Code and Materials

All the data, code and materials can be found on [GitHub](#).

References

- [1] K. Li, J. E. F. Green, H. Tronnolone, A. K. Y. Tam, A. J. Black, J. M. Gardner, J. F. Sundstrom, V. Jiranek, and B. J. Binder, “An off-lattice discrete model to characterise filamentous yeast colony morphology”, *PLOS Computational Biology* 20 (2024), e1012605.
- [2] D. Botstein, S. A. Chervitz, and M. Cherry, “Yeast as a model organism”, *Science* 277 (1997), pp. 1259–1260.
- [3] D. L. Moyes and J. R. Naglik, “Mucosal immunity and *Candida albicans* infection”, *Journal of Immunology Research* 2011 (2011), p. 346307.
- [4] S. Dashko, N. Zhou, C. Compagno, and J. Piškur, “Why, when, and how did yeast evolve alcoholic fermentation?”, *FEMS yeast research* 14 (2014), pp. 826–832.
- [5] U. Petrovič, “Next-generation biofuels: a new challenge for yeast”, *Yeast* 32 (2015), pp. 583–593.
- [6] T. R. Hughes, “Yeast and drug discovery”, *Functional & Integrative Genomics* 2 (2002), pp. 199–211.
- [7] X.-X. Shen, D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. Haase, J. H. Wisecaver, M. Wang, D. T. Doering, et al., “Tempo and mode of genome evolution in the budding yeast subphylum”, *Cell* 175 (2018), pp. 1533–1545.
- [8] V. Jiranek, “The prospect of superior yeast for winemaking: recent successes through bioprospecting”, *Current Opinion in Biotechnology* 90 (2024), p. 103200.
- [9] C. J. Gimeno, P. O. Ljungdahl, C. A. Styles, and G. R. Fink, “Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS”, *Cell* 68 (1992), pp. 1077–1090.
- [10] T. Freedman, A. Porter, and B. Haarer, “Mutational and hyperexpression-induced disruption of bipolar budding in yeast”, *Microbiology* 146 (2000), pp. 2833–2843.
- [11] H. Tronnolone, A. Tam, Z. Szenczi, J. Green, S. Balasuriya, E. L. Tek, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, et al., “Diffusion-limited growth of microbial colonies”, *Scientific Reports* 8 (2018), p. 5992.

- [12] J. G. Cook, L. Bardwell, S. J. Kron, and J. Thorner, “Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast *Saccharomyces cerevisiae*.”, *Genes & Development* 10 (1996), pp. 2831–2848.
- [13] A. L. Jenner, C.-O. Yun, P. S. Kim, and A. C. Coster, “Mathematical modelling of the interaction between cancer cells and an oncolytic virus: insights into the effects of treatment protocols”, *Bulletin of Mathematical Biology* 80 (2018), pp. 1615–1629.
- [14] X. Wang, A. L. Jenner, R. Salomone, D. J. Warne, and C. Drovandi, “Calibration of agent based models for monophasic and biphasic tumour growth using approximate Bayesian computation”, *Journal of Mathematical Biology* 88 (2024), p. 28.
- [15] J. J. Klowss, A. P. Browning, R. J. Murphy, E. J. Carr, M. J. Plank, G. Gunasingh, N. K. Haass, and M. J. Simpson, “A stochastic mathematical model of 4D tumour spheroids with real-time fluorescent cell cycle labelling”, *Journal of the Royal Society Interface* 19 (2022), p. 20210903.
- [16] B. J. Binder, K. A. Landman, M. J. Simpson, M. Mariani, and D. F. Newgreen, “Modeling proliferative tissue growth: a general approach and an avian case study”, *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 78 (2008), p. 031912.
- [17] P. J. Brown, J. E. F. Green, B. J. Binder, and J. M. Osborne, “A rigid body framework for multicellular modeling”, *Nature Computational Science* 1 (2021), pp. 754–766.
- [18] R. J. Murphy, P. Buenzli, R. Baker, and M. Simpson, “A one-dimensional individual-based mechanical model of cell movement in heterogeneous tissues and its coarse-grained approximation”, *Proceedings of the Royal Society A* 475 (2019), p. 20180838.
- [19] G. An, Q. Mi, J. Dutta-Moscato, and Y. Vodovotz, “Agent-based models in translational systems biology”, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1 (2009), pp. 159–171.
- [20] T. E. Gorochofski, “Agent-based modelling in synthetic biology”, *Essays in biochemistry* 60 (2016), pp. 325–336.
- [21] M. Soheilypour and M. R. Mofrad, “Agent-based modeling in molecular systems biology”, *BioEssays* 40 (2018), p. 1800020.

- [22] H. Tronnolone, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, and B. J. Binder, “Quantifying the dominant growth mechanisms of dimorphic yeast using a lattice-based model”, *Journal of The Royal Society Interface* 14 (2017), p. 20170314.
- [23] A. J. Gallo, H. Tronnolone, J. E. F. Green, and B. J. Binder, “Modelling uniaxial non-uniform yeast colony growth: comparing an agent-based model and continuum approximations”, *Journal of Theoretical Biology* 523 (2021), p. 110715.
- [24] S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of approximate Bayesian computation*, CRC press, 2018.
- [25] G. Papamakarios, D. Sterratt, and I. Murray, “Sequential neural likelihood: fast likelihood-free inference with autoregressive flows”, *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 837–848.
- [26] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, “Fast likelihood-free cosmology with neural density estimators and active learning”, *Monthly Notices of the Royal Astronomical Society* 488 (2019), pp. 4440–4458.
- [27] A. Zammit-Mangion, M. Sainsbury-Dale, and R. Huser, “Neural methods for amortized inference”, *Annual Review of Statistics and Its Application* 12 (2024).
- [28] J. Dyer, P. Cannon, J. D. Farmer, and S. Schmon, “Black-box Bayesian inference for economic agent-based models”, *arXiv preprint arXiv:2202.00625* (2022).
- [29] A. Fengler, L. N. Govindarajan, T. Chen, and M. J. Frank, “Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience”, *Elife* 10 (2021), e65074.
- [30] H. Tronnolone, J. M. Gardner, J. F. Sundstrom, V. Jiranek, S. G. Oliver, and B. J. Binder, “TAMMiCol: Tool for analysis of the morphology of microbial colonies”, *PLoS Computational Biology* 14 (2018), e1006629.
- [31] B. J. Binder, J. F. Sundstrom, J. M. Gardner, V. Jiranek, and S. G. Oliver, “Quantifying two-dimensional filamentous and invasive growth spatial patterns in yeast colonies”, *PLoS Computational Biology* 11 (2015), e1004070.
- [32] C. M. Bishop, *Mixture Density Networks*, Technical Report, Birmingham: Aston University, 1994, pp. 1–25.

- [33] R. P. Kelly, D. J. Warne, D. T. Frazier, D. J. Nott, M. U. Gutmann, and C. Drovandi, “Simulation-based bayesian inference under model misspecification”, *arXiv preprint arXiv: 2503.12315* (2025).
- [34] G. Papamakarios, T. Pavlakou, and I. Murray, “Masked Autoregressive Flow for Density Estimation”, *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.
- [35] L. M. Sullivan, K. A. Dukes, and E. Losina, “An introduction to hierarchical linear modelling”, *Statistics in Medicine* 18 (1999), pp. 855–888.

A Supplementary Material

A.1 Logit transform of parameters

When sampling out the posterior, we include a Jacobian term in the priors because of the logit transformation performed on the parameters θ . Beginning with the logit transform,

$$z = \log\left(\frac{\theta}{1-\theta}\right). \quad (\text{A.1})$$

The inverse logit transform of Equation (A.1) is

$$\theta = \frac{1}{1 + \exp(-z)}. \quad (\text{A.2})$$

The probability density function, $P(\theta)$ of the Beta distribution defined on a unit interval is given as

$$p(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}. \quad (\text{A.3})$$

When we transform θ by applying the logit transform, we need to account for the change using the derivative $d\theta/dz$. Using the transformation of the variable, the transformed density becomes,

$$p(z) = p(\theta) \cdot \frac{d\theta}{dz}, \quad (\text{A.4})$$

where $d\theta/dz$ is the Jacobian. Taking the derivative of Equation (A.2) with respect to θ we get,

$$\frac{d\theta}{dz} = \frac{\exp(-z)}{(1 + \exp(-z))^2}. \quad (\text{A.5})$$

Substituting Equations (A.3) and (A.5) into Equation (A.4) to get the transformed density,

$$p(z) = \frac{1}{B(\alpha, \beta)} \left(\frac{1}{1 + \exp(-z)}\right)^{\alpha-1} \left(1 - \frac{1}{1 + \exp(-z)}\right)^{\beta-1} \frac{\exp(-z)}{(1 + \exp(-z))^2}. \quad (\text{A.6})$$

A.2 Synthetic data

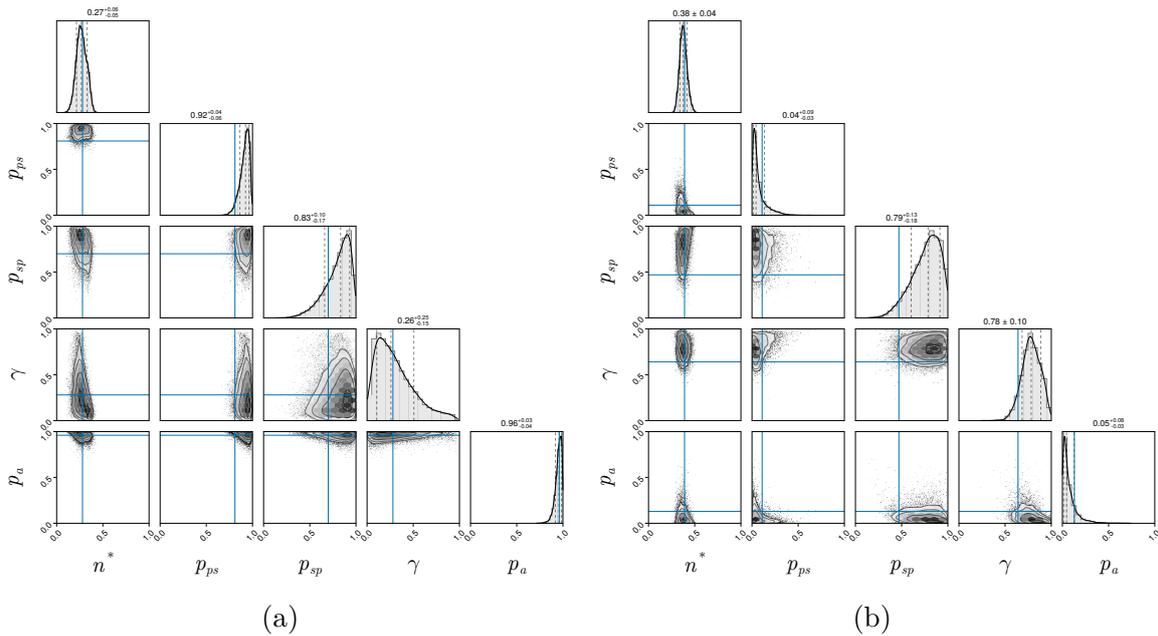


Figure A.1: Pair plots of synthetic posterior distribution. The blue line represents the ground truth parameter values. (a) $n^* = 0.28$, $p_{ps} = 0.81$, $p_{sp} = 0.70$, $\gamma = 0.28$, $p_a = 0.96$. (b) $n^* = 0.39$, $p_{ps} = 0.11$, $p_{sp} = 0.47$, $\gamma = 0.64$, $p_a = 0.13$.

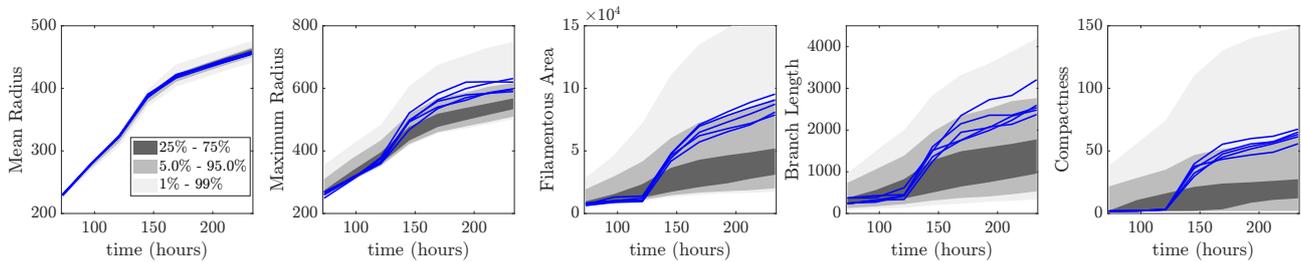


Figure A.2: Summary statistics from simulations of the same parameters simulated up to the same area. This shows the variability within a simulation for a fixed set of parameters.

We sample parameters from the posterior of the synthetic dataset to show the variability in colony morphology of ten realisations. The parameter values are reported in the left corner of each simulation with the ground truth in panel (a). As n^* is the most important parameter in our model, we vary the ground truth colony n^* from $n^* = 0.28$ (Figure A.4), $n^* = 0.39$ (Figure A.5), to $n^* = 0.67$ (Figure A.6). All three sets of simulated results sampled from the posterior align well with the ground truth.

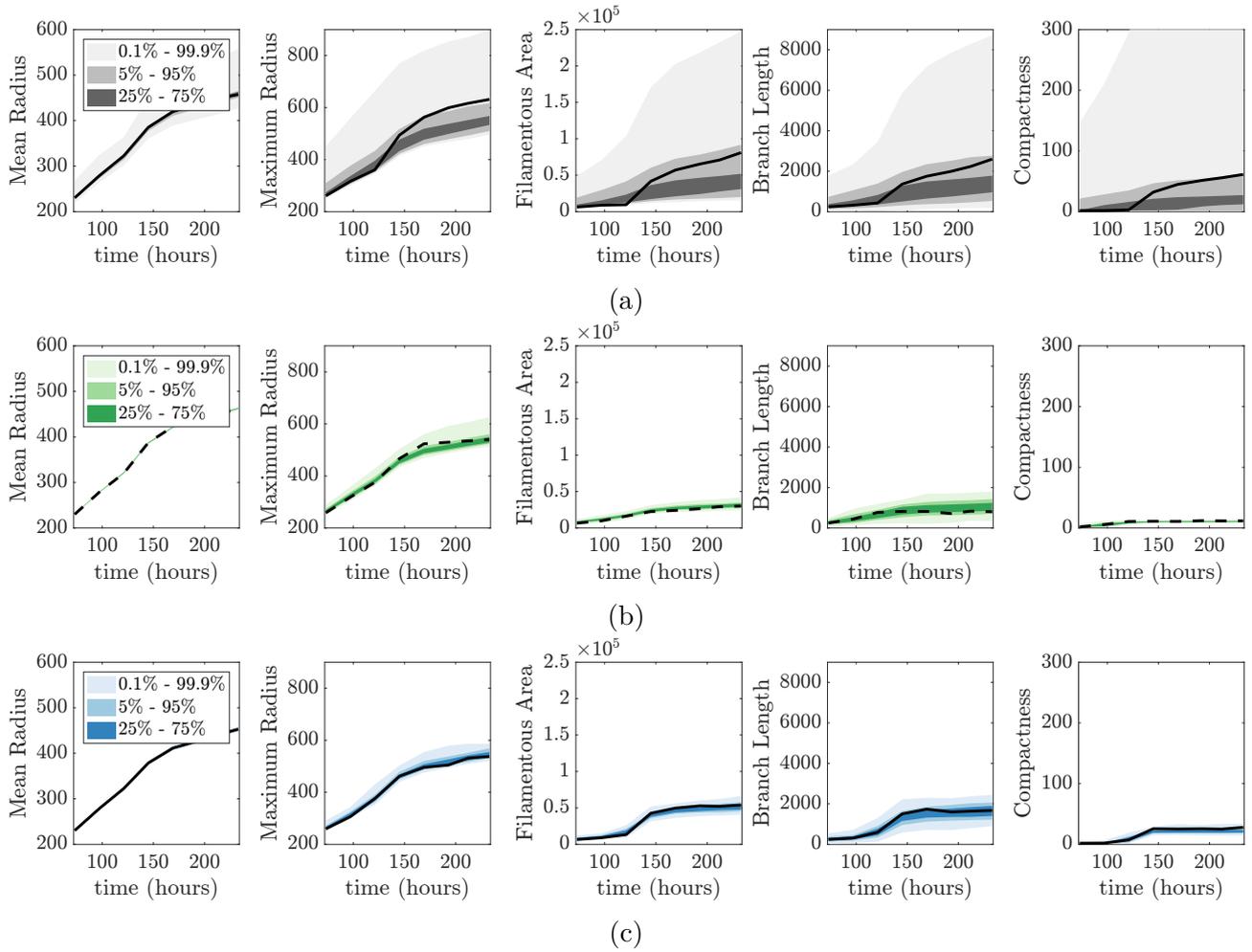
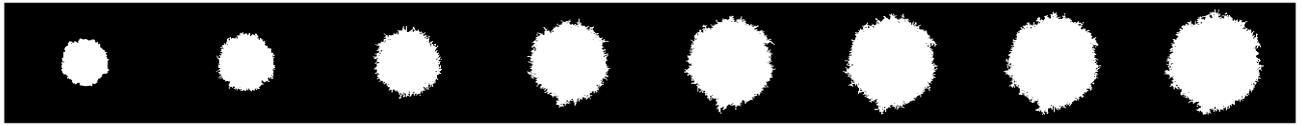
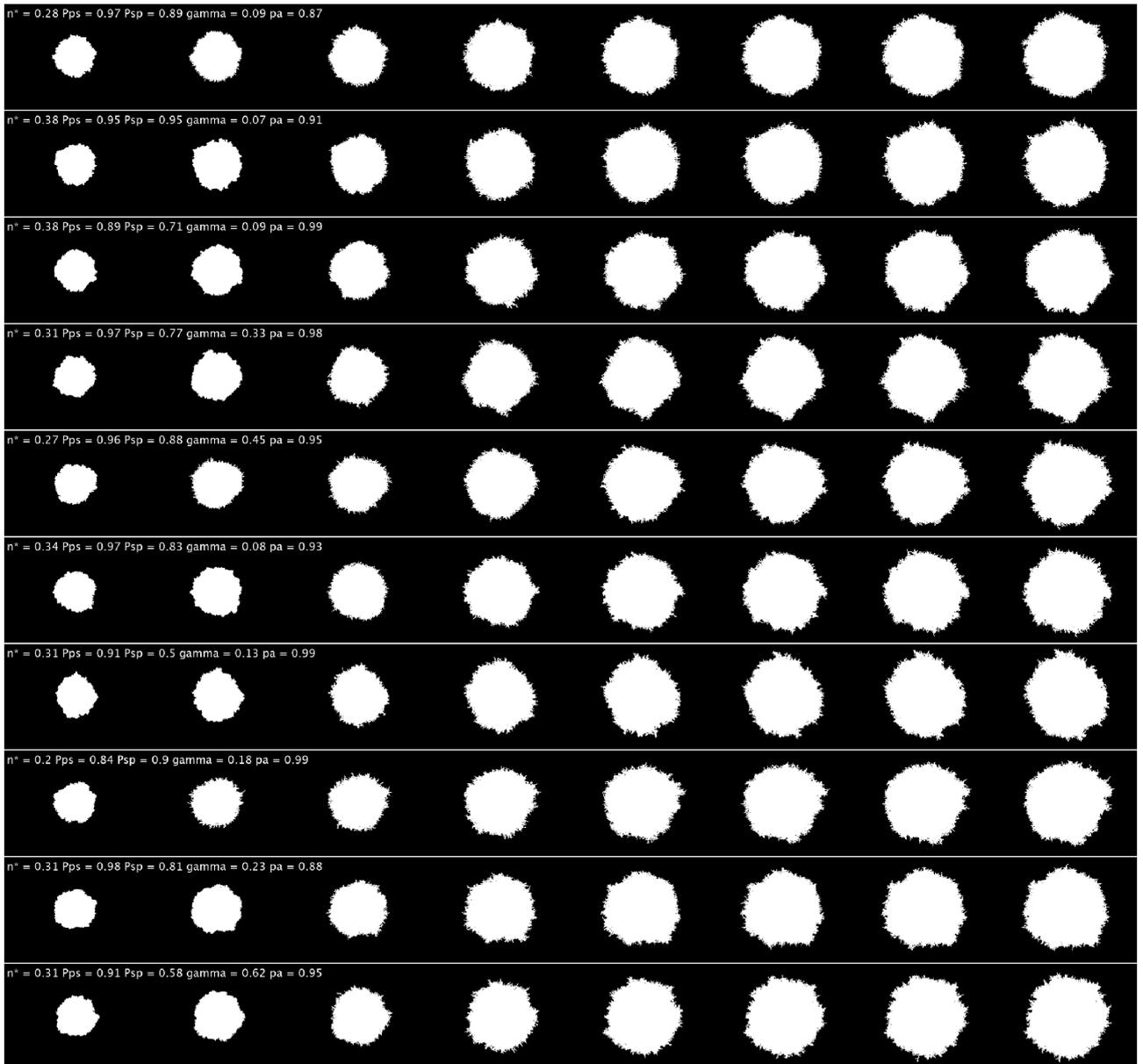


Figure A.3: Predictive distribution for the synthetic datasets. (a) Prior predictive distribution with the black lines representing the experimental data. The following are the posterior predictive distributions with parameters (b) $n^* = 0.28$, $p_{ps} = 0.81$, $p_{sp} = 0.70$, $\gamma = 0.28$, $p_a = 0.96$. (c) $n^* = 0.39$, $p_{ps} = 0.11$, $p_{sp} = 0.47$, $\gamma = 0.64$, $p_a = 0.13$.

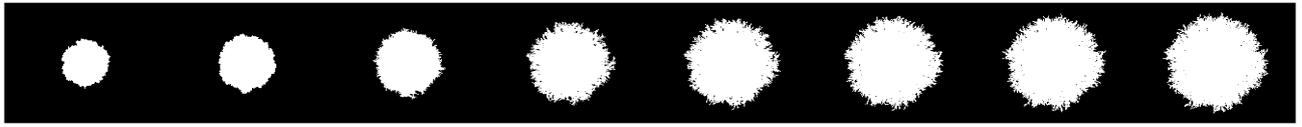


(a) Synthetic ground truth filamentous yeast simulations with $n^* = 0.28$, $p_{ps} = 0.81$, $p_{sp} = 0.70$, $\gamma = 0.28$, $p_a = 0.96$.

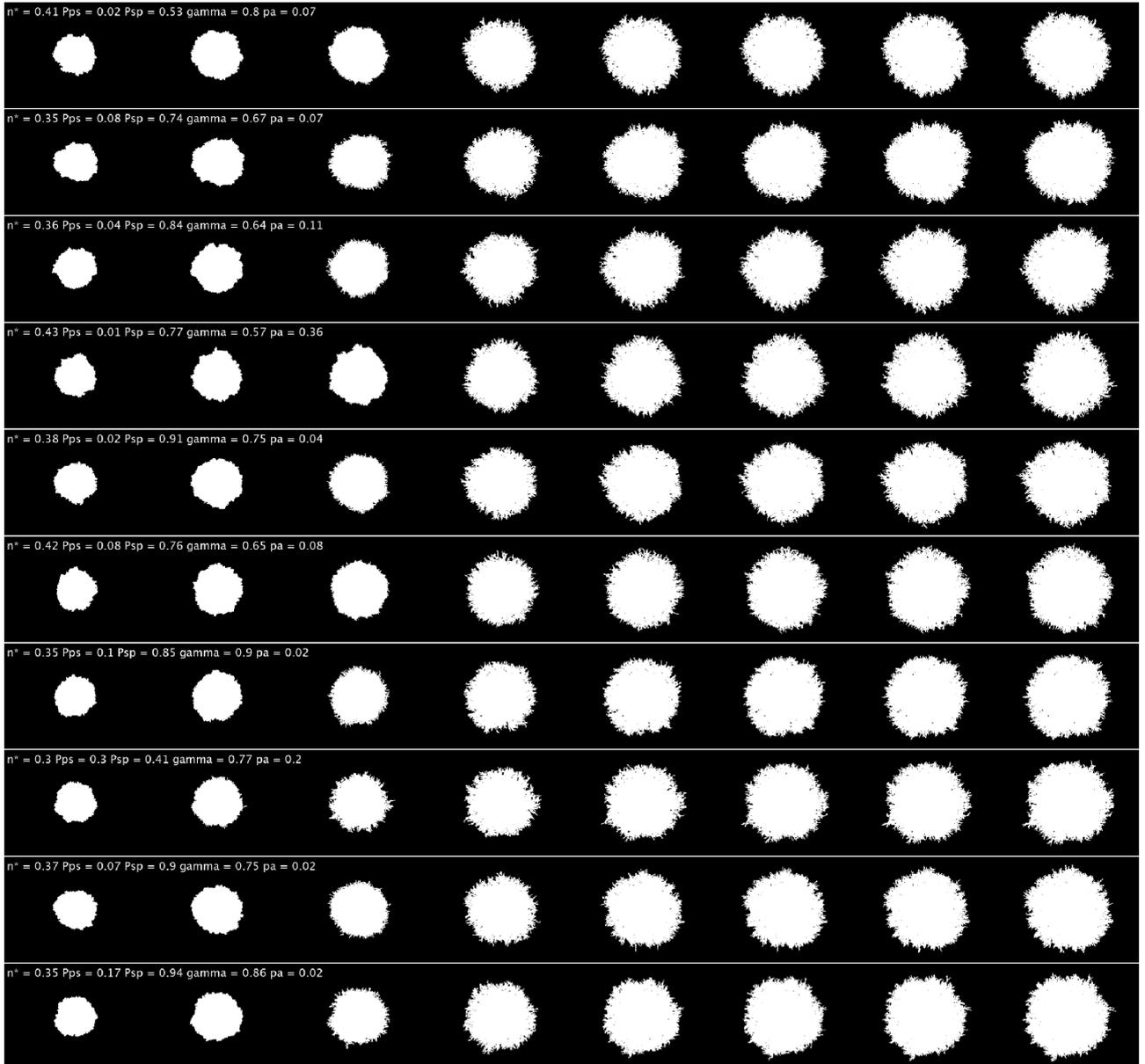


(b)

Figure A.4: Colony simulation comparison between (a) Ground truth and (b) NLE prediction of ten realisations.

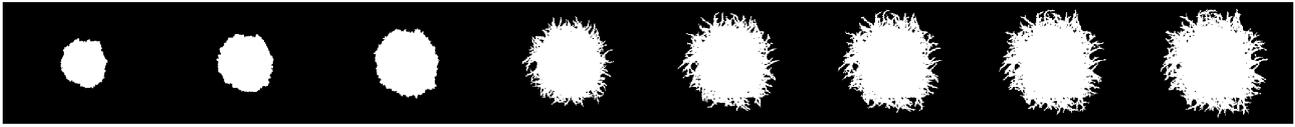


(a) Synthetic ground truth filamentous yeast simulations with $n^* = 0.39$, $p_{ps} = 0.11$, $p_{sp} = 0.47$, $\gamma = 0.64$, $p_a = 0.13$.

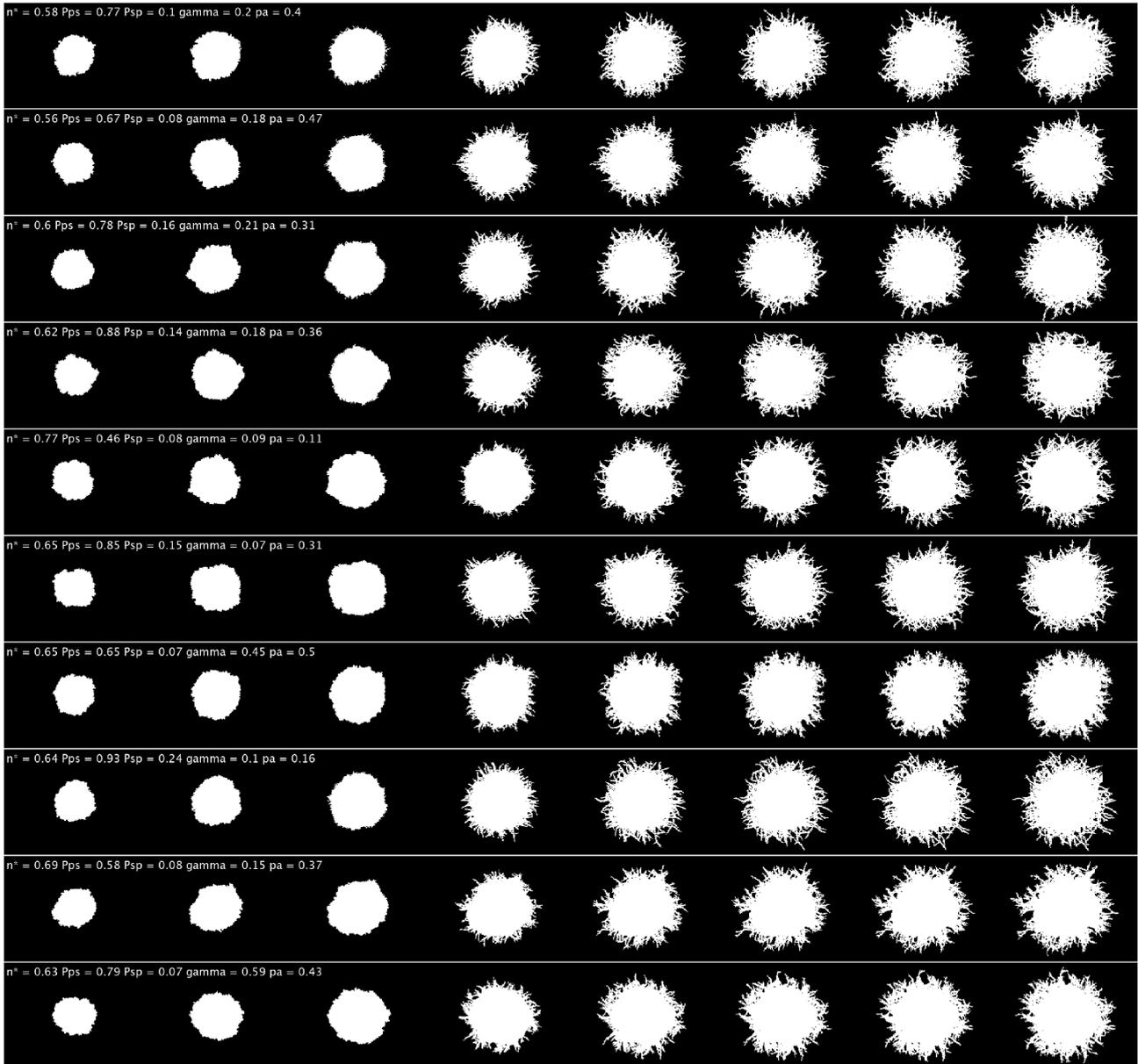


(b)

Figure A.5: Colony simulation comparison between (a) Ground truth and (b) NLE prediction of ten realisations.



(a) Synthetic ground truth filamentous yeast simulations with $n^* = 0.67$, $p_{ps} = 0.58$, $p_{sp} = 0.25$, $\gamma = 0.12$, $p_a = 0.14$.



(b)

Figure A.6: Colony simulation comparison between (a) Ground truth and (b) NLE prediction of ten realisations.

A.3 Synthetic data using different numbers of summary statistics

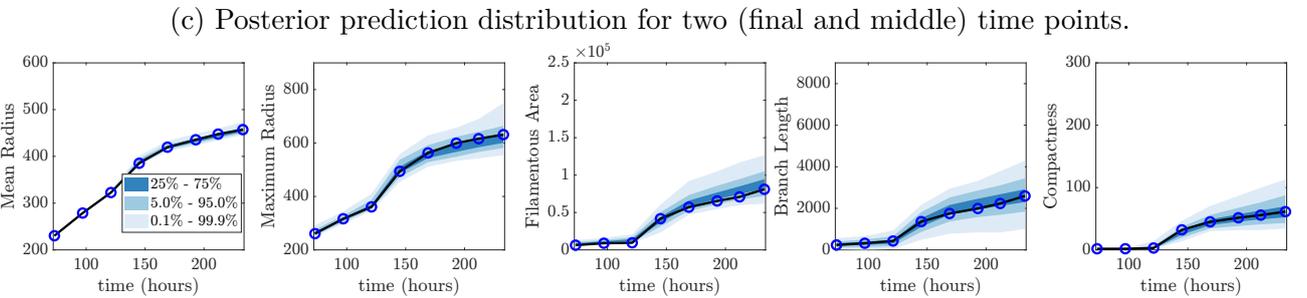
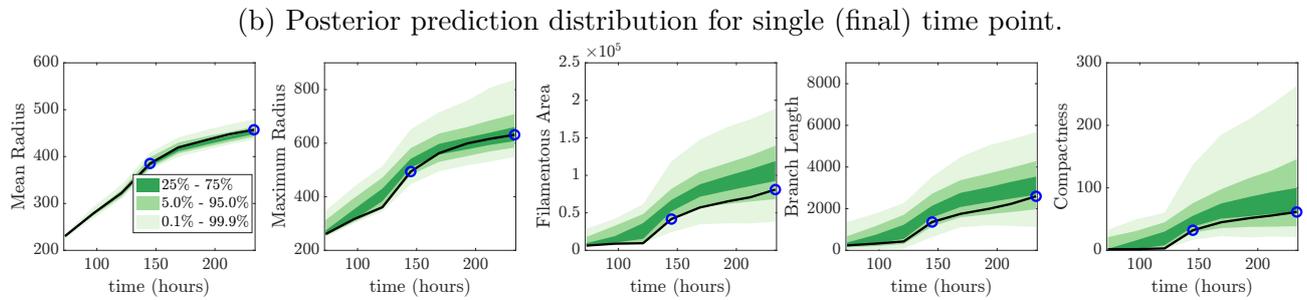
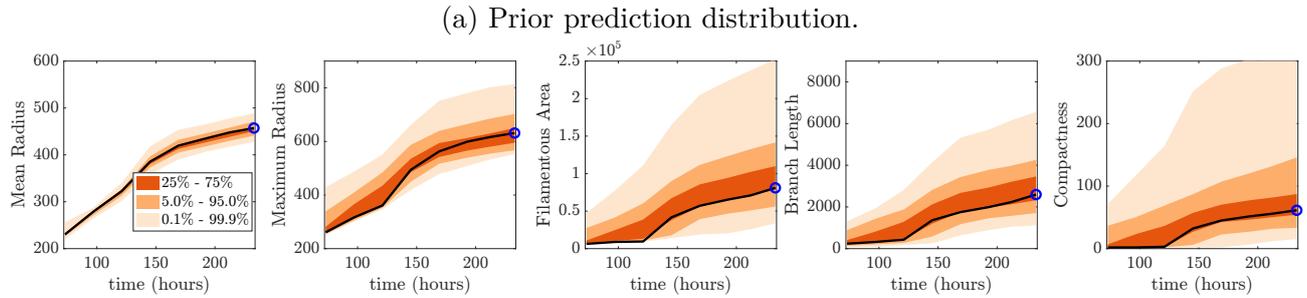
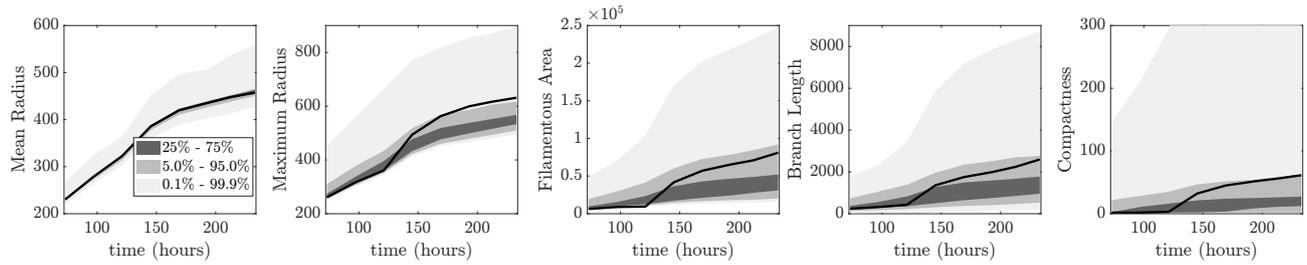


Figure A.7: (a) Prior predictive distribution. Posterior predictive check for (b) eight time points, (c) two time points and (d) a single time point. The blue markers indicate the time points where we evaluate the summary statistics. (b,c,d) share the same prior predictive distribution shown in (a).

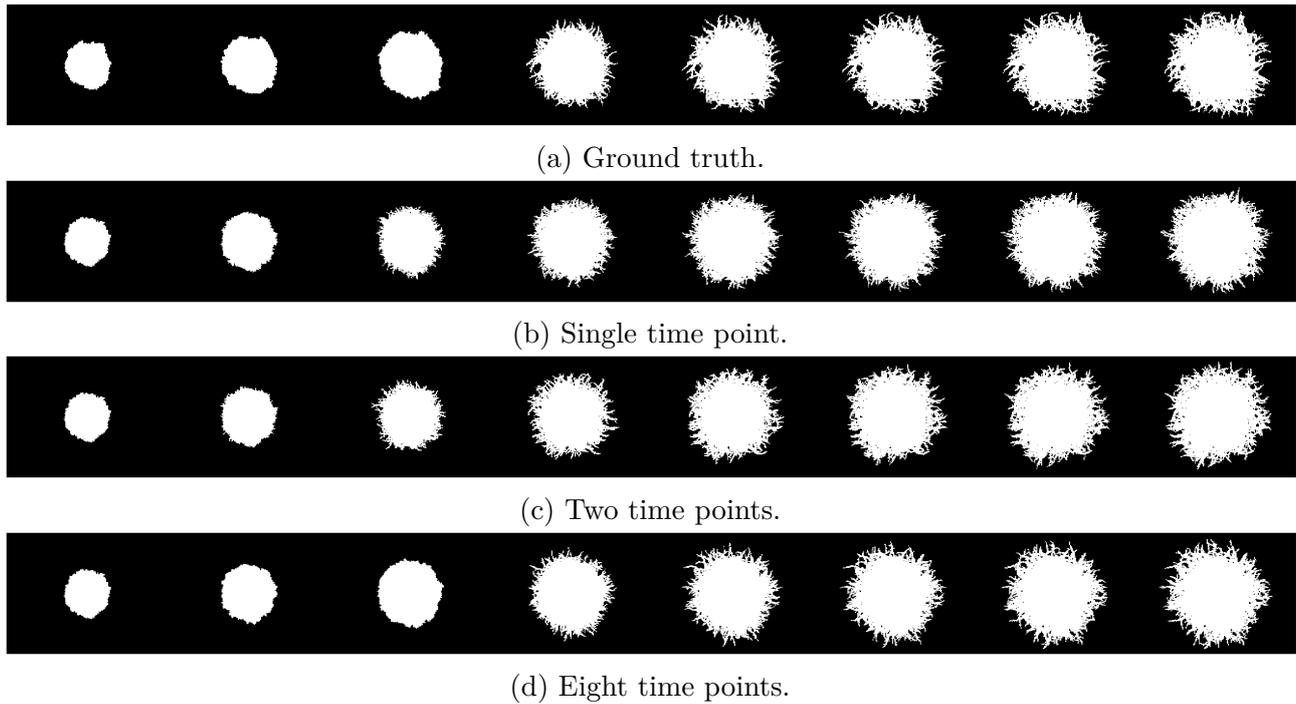
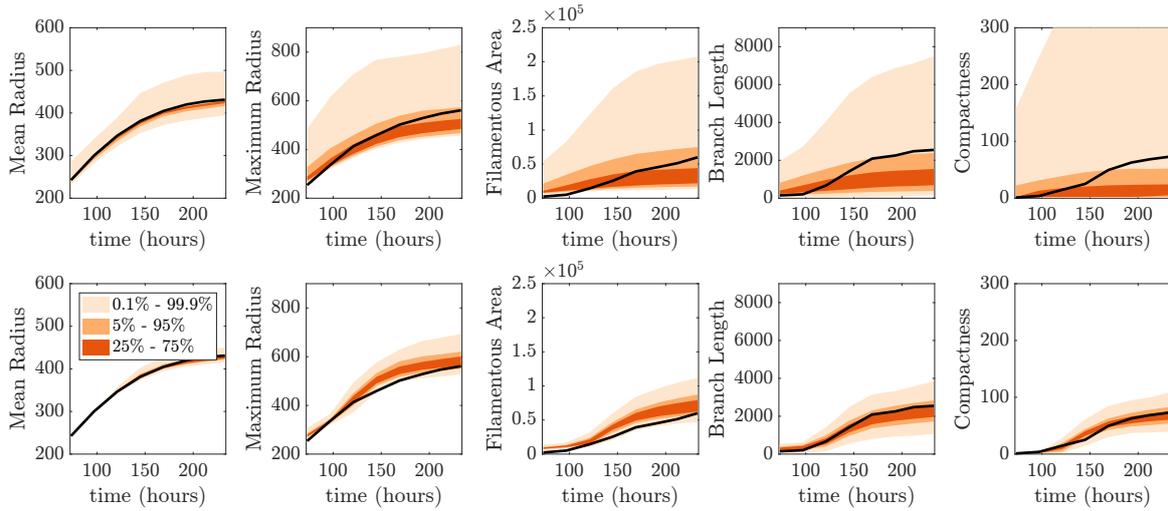


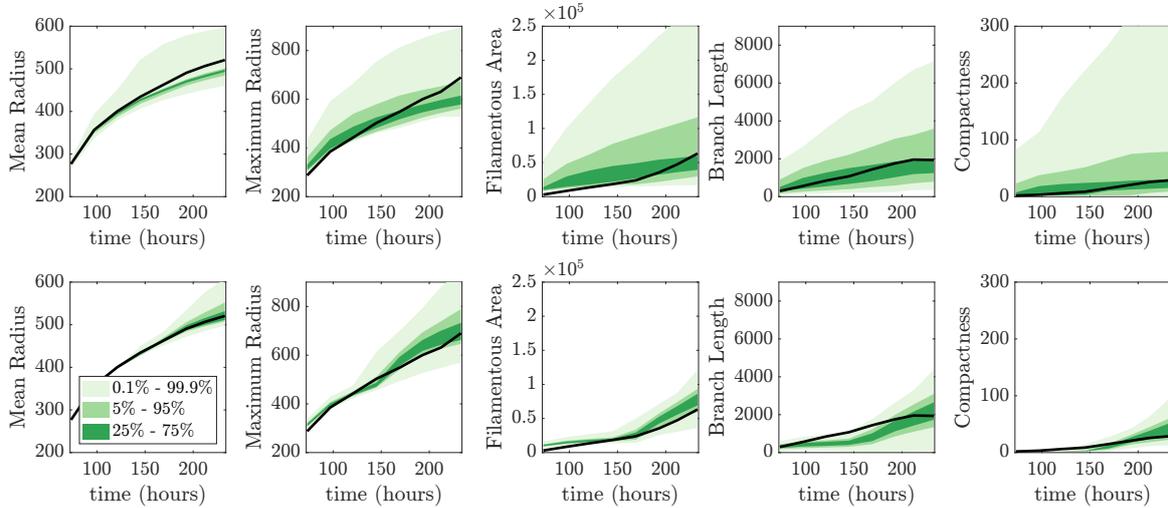
Figure A.8: Comparison of simulations generated from posterior distributions inferred using different numbers of summary statistics. (a) The ground truth is plotted for comparison against a colony inferred from (b) a single time point (final time), (c) two time points (middle and final time) and (d) all time data available (eight time points).

A.4 Experimental data

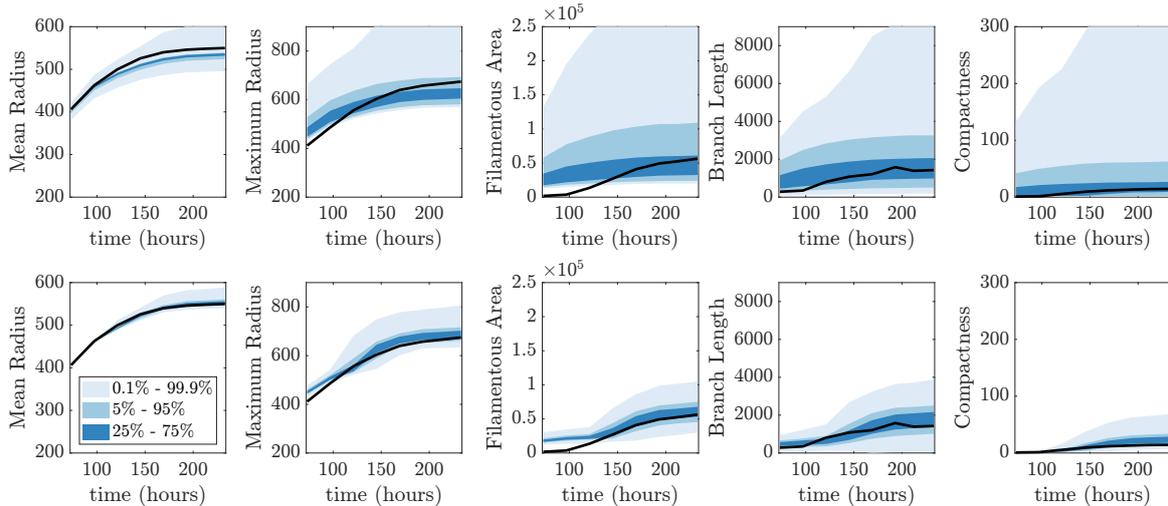
Here we present the experimental data, predictive distribution, and ten realisations drawn from the posterior distribution.



(a) AWRI 796 50 μM

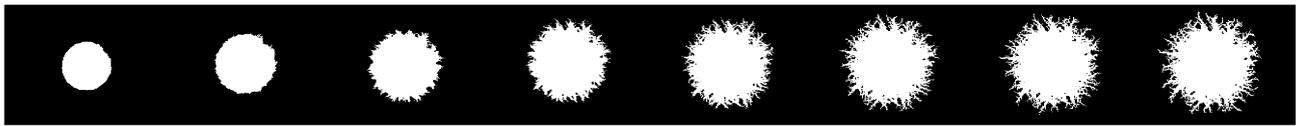


(b) Simi White 50 μM

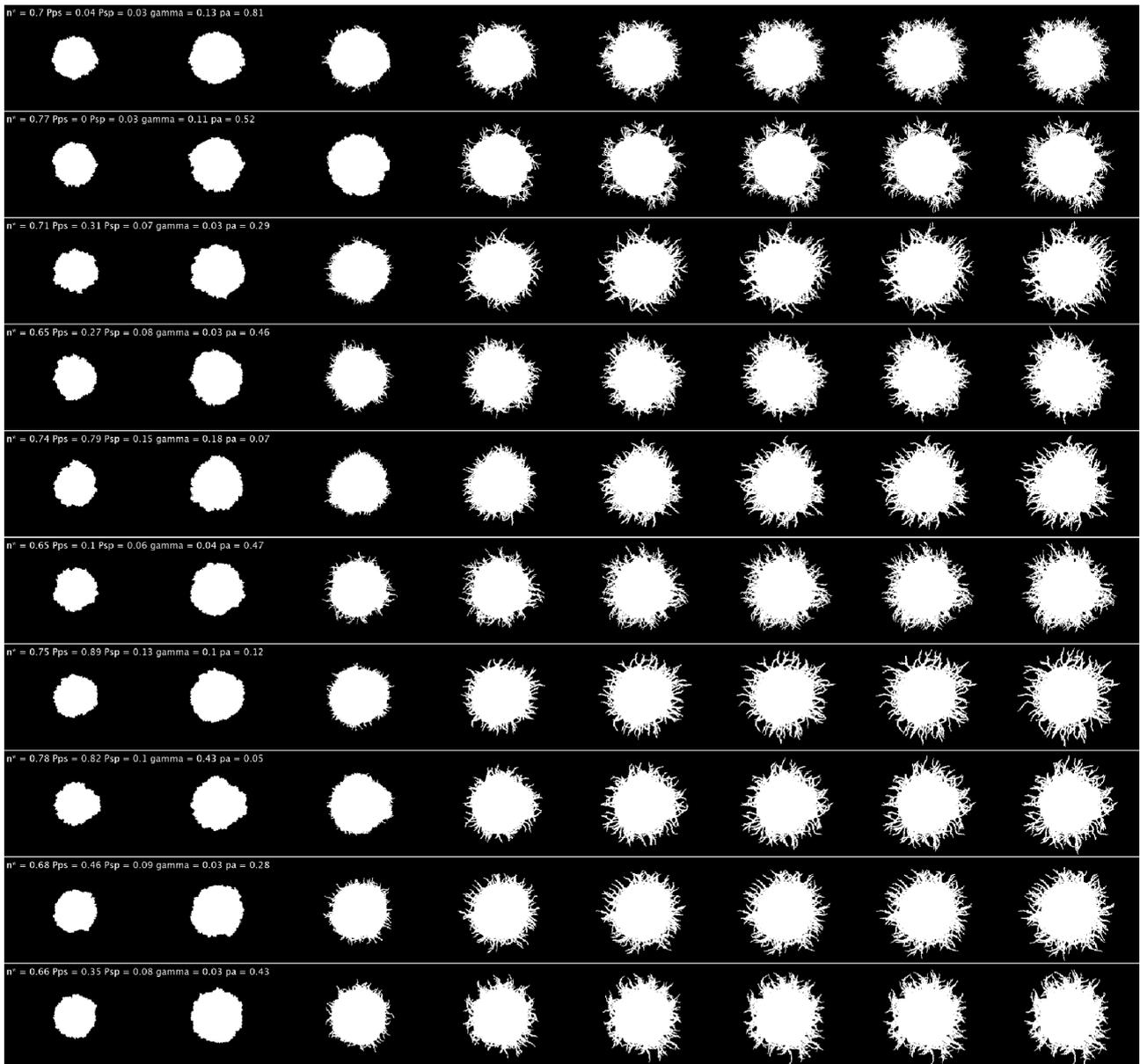


(c) AWRI 796 500 μM

Figure A.9: Prior predictive plots (top row) and their corresponding posterior predictive distribution (bottom row) for each experimental colony. (a) AWRI 796 50 μM (b) Simi White 50 μM and (c) AWRI 796 500 μM .



(a) Experimental filamentous yeast colony for AWRI 796 50 μm

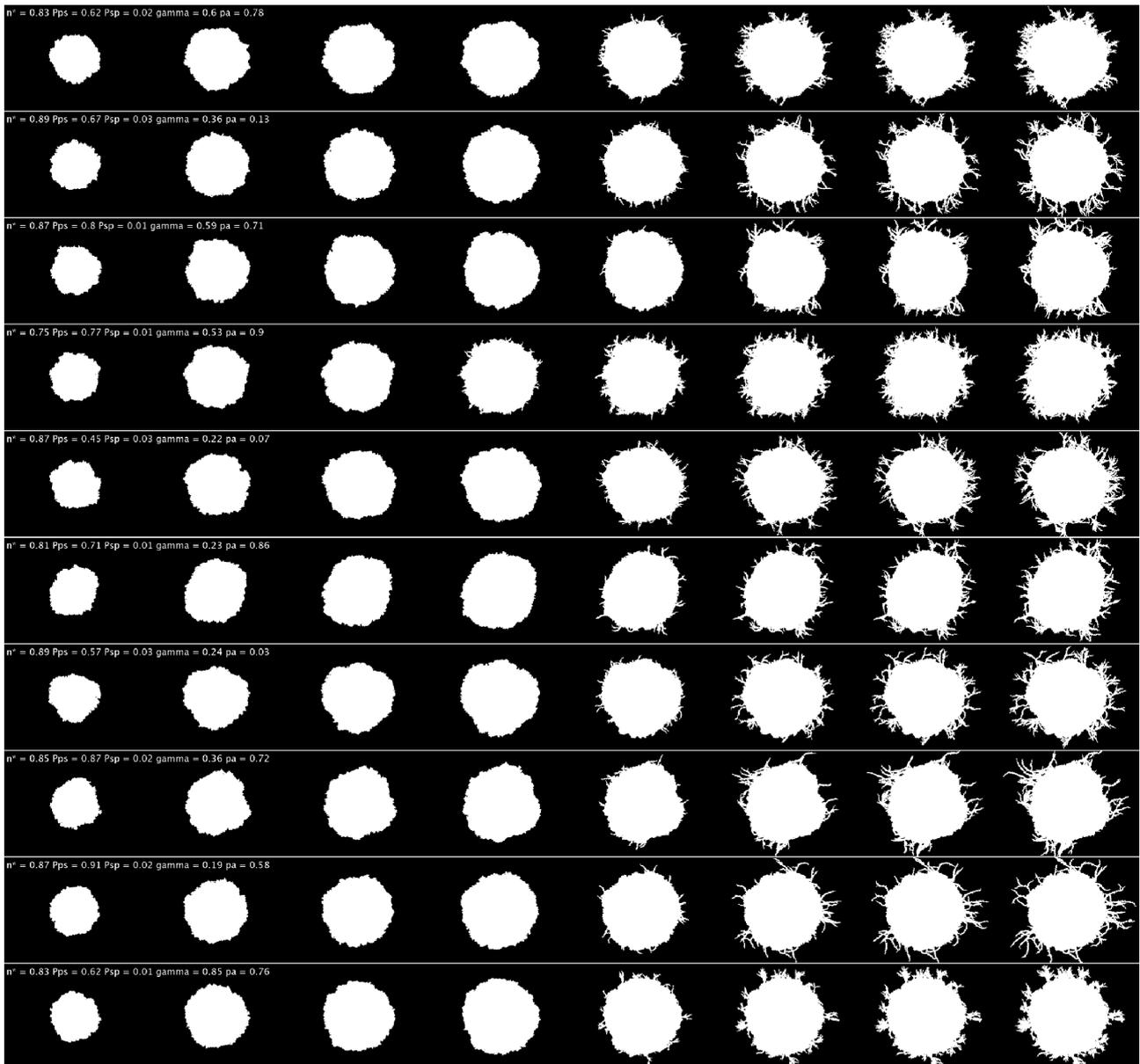


(b) Ten realization from the drawn from the posterior distribution.

Figure A.10: Colony comparison between (a) experimental AWRI 796 50 μm and (b) ten realisations drawn from the posterior distribution.



(a) Experimental filamentous yeast colony for Simi White 50 μm

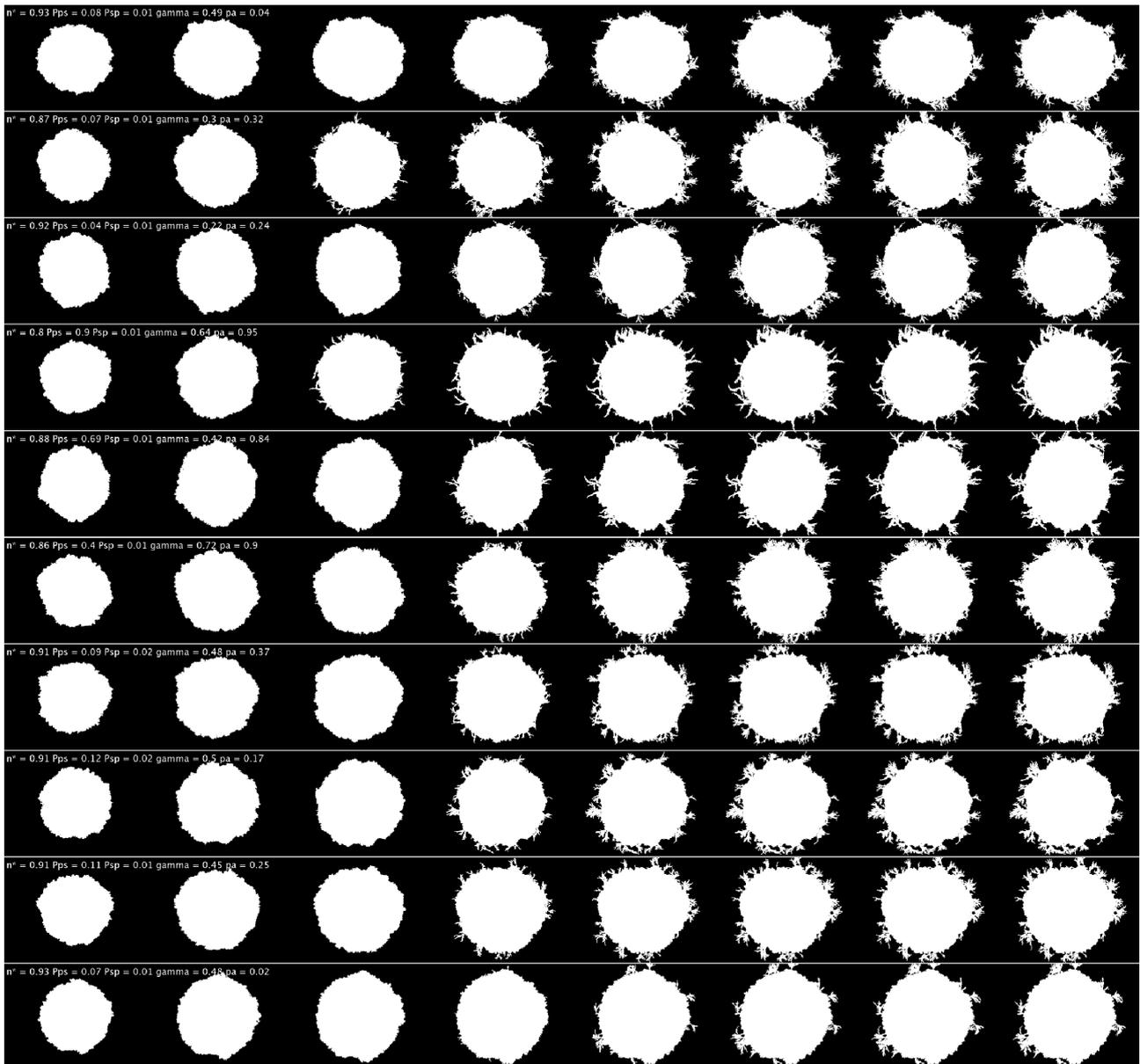


(b) Ten realization from the drawn from the posterior distribution.

Figure A.11: Colony comparison between (a) experimental Simi White 50 μm and (b) ten realisations drawn from the posterior distribution.



(a) Experimental filamentous yeast colony for AWRI 796 500 μm

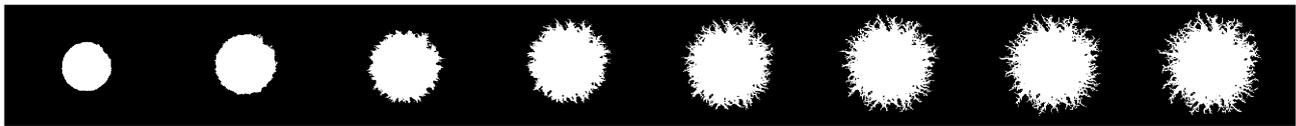


(b) Ten realization from the drawn from the posterior distribution.

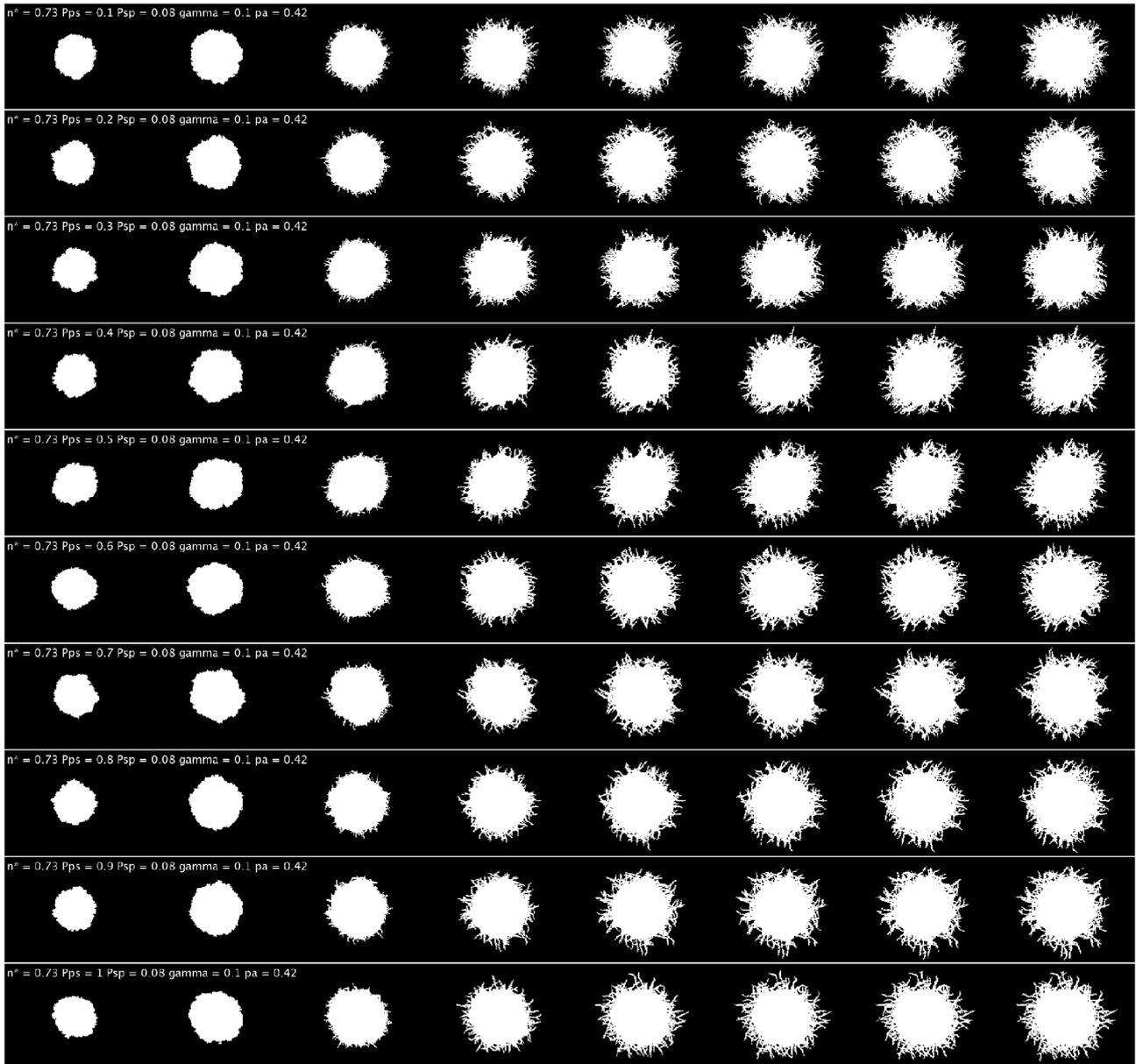
Figure A.12: Colony comparison between (a) experimental AWRI 796 500 μm and (b) ten realisations drawn from the posterior distribution.

A.5 Testing the impact of p_{ps} on colony morphology

To demonstrate that the parameter p_{ps} is uninformative, we fixed all other parameters constant and varied p_{ps} from 0.1 by increments of 0.1 up to 1 and simulated the associated colonies presented in Figure A.13. As expected, we see the simulated colony morphology align closely with the experimental colony, with very little difference between simulations of varying values of p_{ps} .



(a) Experimental images of filamentous yeast colony for AWRI 796 50 μm sample 1.



(b)

Figure A.13: Colony comparison between (a) experiment of AWRI 796 50 μm and (b) each realisation is a different value of p_{ps} while hold all other parameters constant (values reported in Figure 3.5) demonstrating this parameter is non-informative.

A.6 Pairplots of experimental posterior distributions

Pairplot of posteriors of experimental colonies for AWRI 796 50 μM (Figure A.14), Simi White 50 μM (Figure A.15) and AWRI 796 500 μM (Figure A.16).

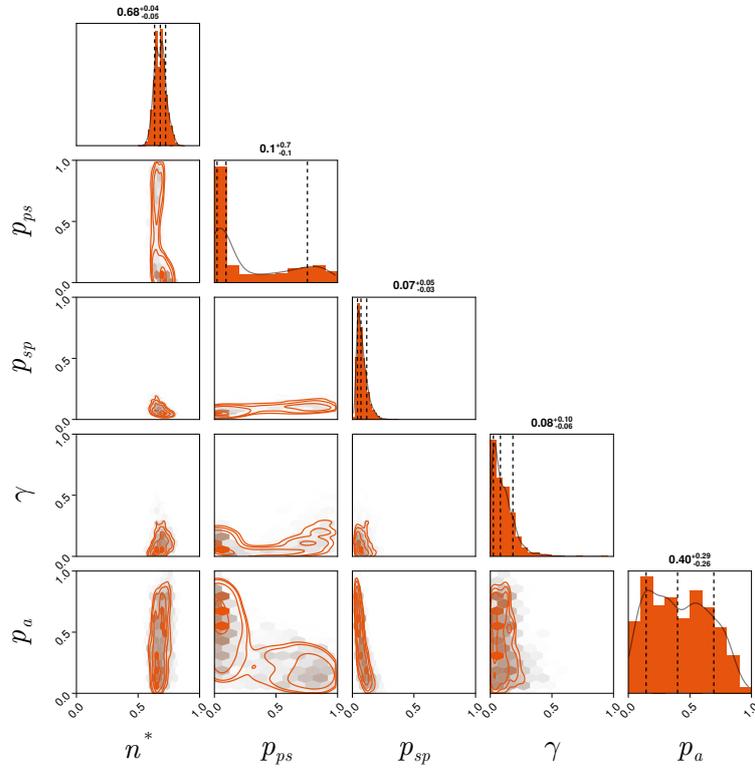


Figure A.14: Pair plot of AWRI 796 50 μM posterior distribution.

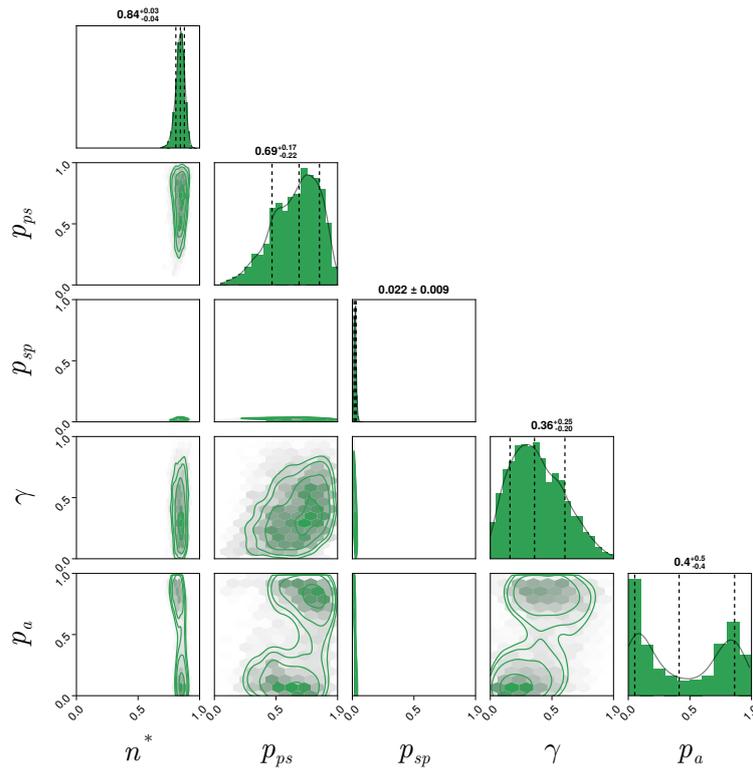


Figure A.15: Pair plot of Simi White 50 μm approximate posterior distribution.

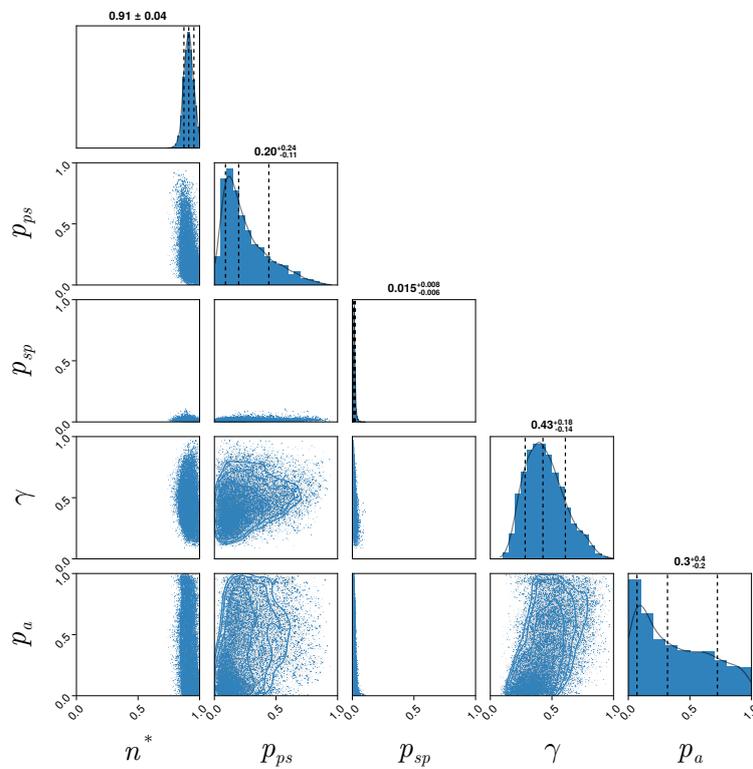
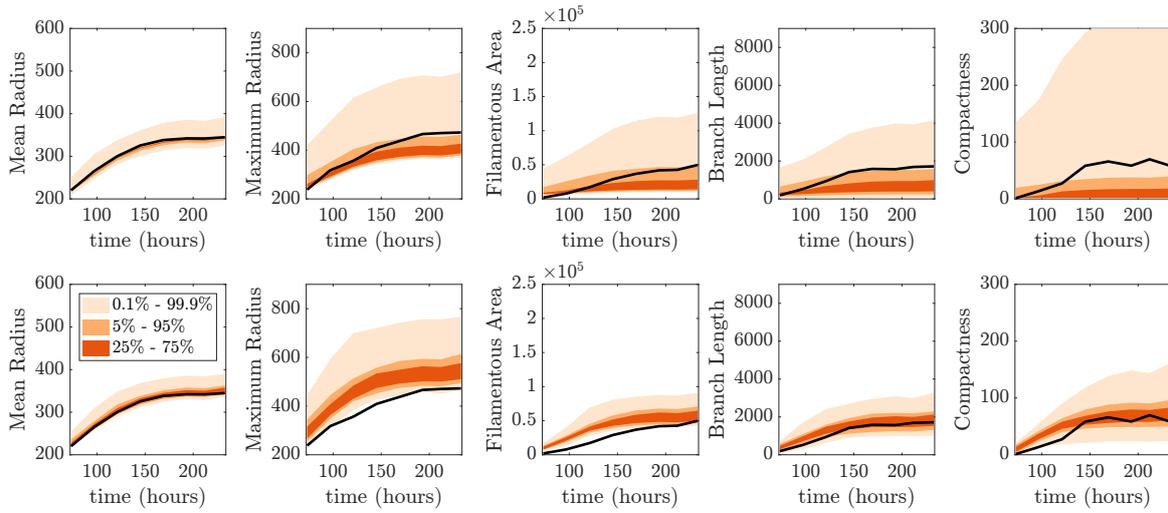
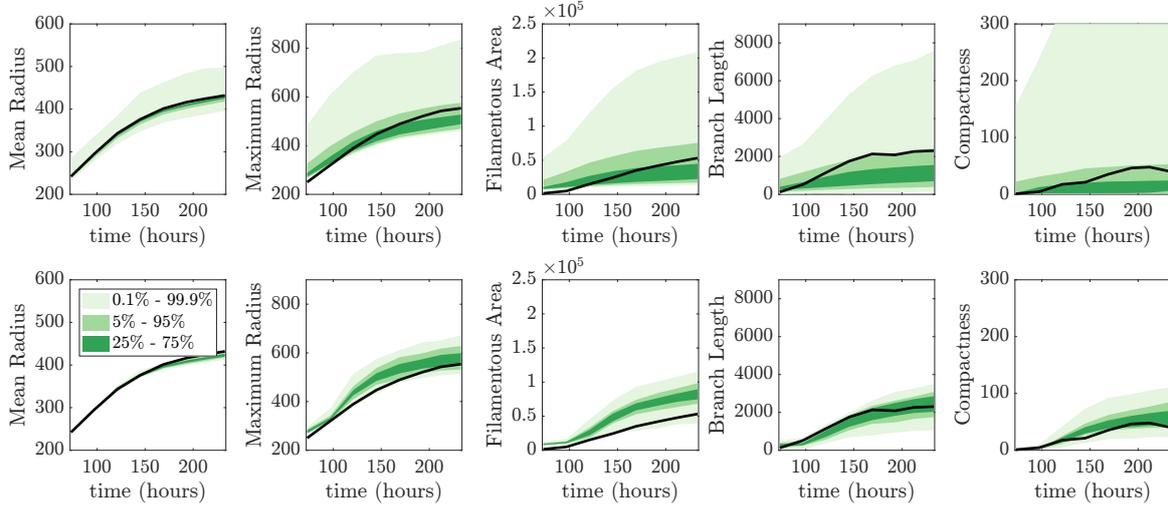


Figure A.16: Pair plot of AWRI 796 500 μm approximate posterior distribution.

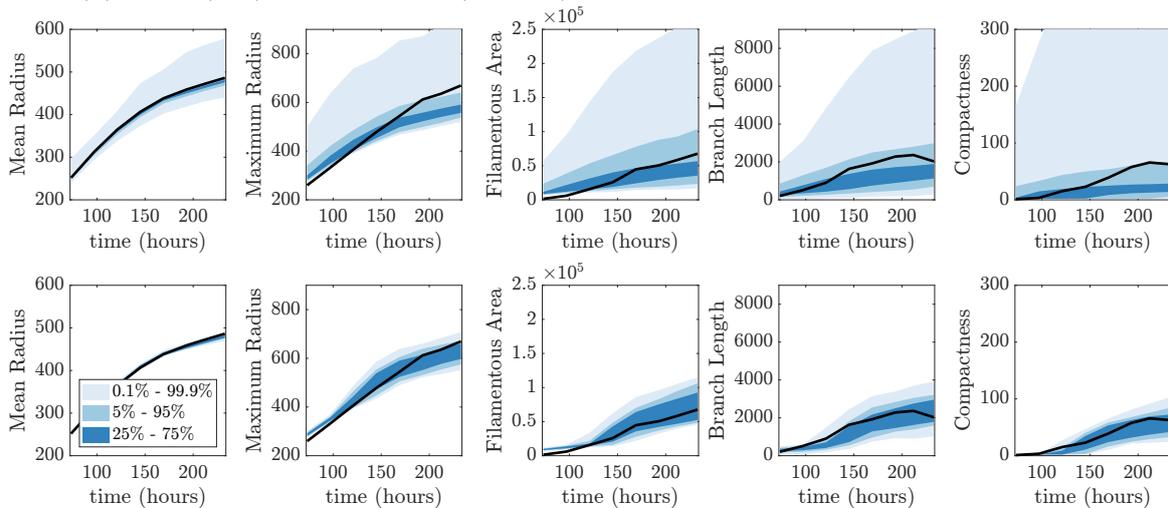
A.7 Posterior predictive checks from varying area



(a) Prior (top) and posterior (bottom) prediction check for the smallest colony size.



(b) Prior (top) and posterior (bottom) prediction check for median colony size.



(c) Posterior prediction check for largest colony size.

Figure A.17: Prior (top) and posterior (bottom) prediction check for (a) smallest colony, (b) median colony and (c) largest colony based on area.

A.8 Variability of area among replicates of the same experimental conditions

We present the colonies at the final experimental time to show the variability within an individual experiment of AWRI 796 50 μm as shown in Figure A.18.

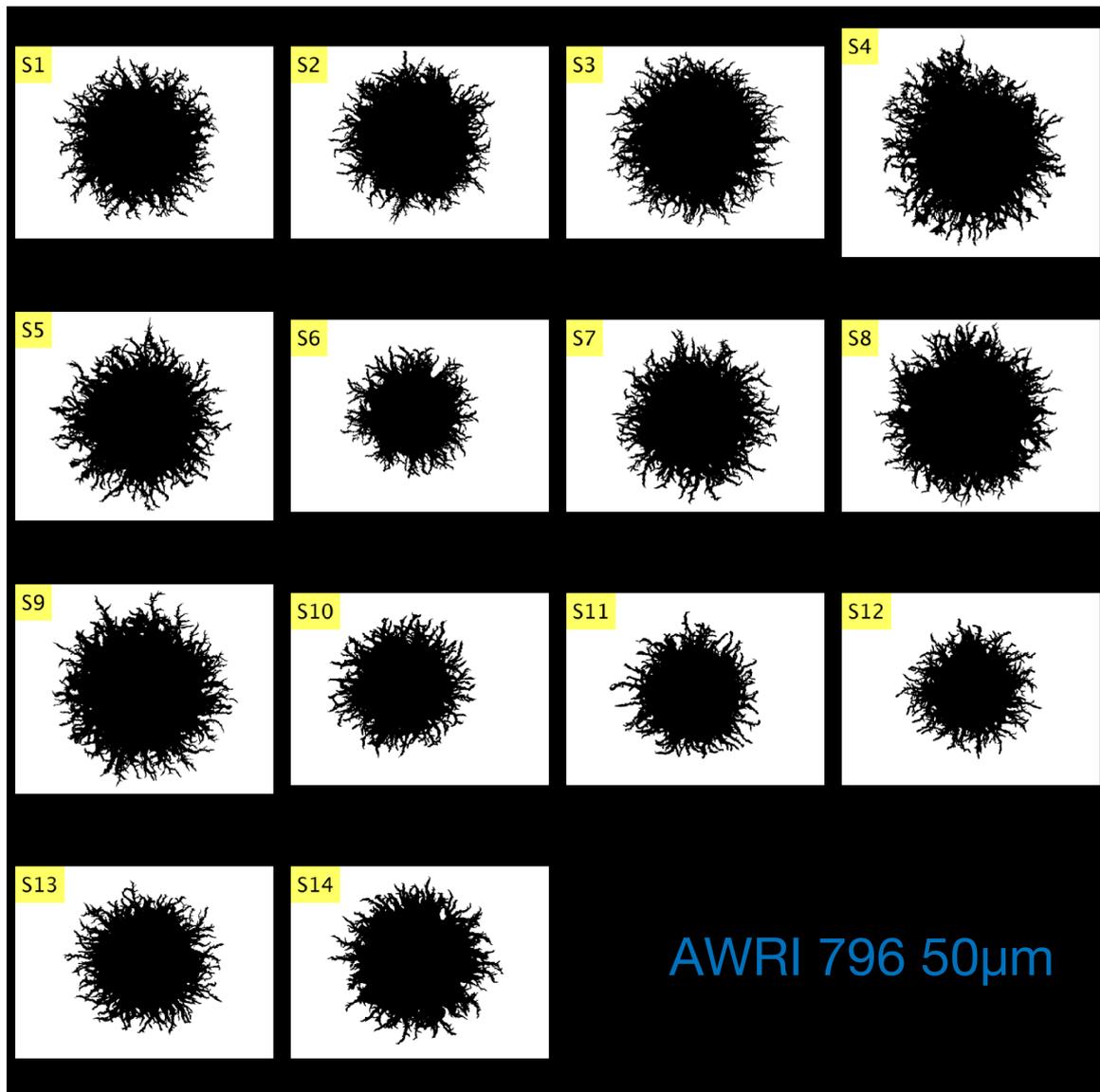


Figure A.18: Variability of experimental image across the AWRI 796 50 μm experiments at final time (233 hours). The smallest colony was sample 12, the median size was sample 14, and the largest was sample 9.