

Contents

| | | |
|---------|--|----|
| 1 | ABC-Flow | 3 |
| 1.1 | Introduction | 3 |
| 1.2 | Methods | 3 |
| 1.3 | Results | 5 |
| 1.3.1 | Distance Calculations | 5 |
| 1.3.2 | Comparing 1D and 2D distances | 5 |
| 1.3.2.1 | Normal distribution | 6 |
| 1.3.2.2 | Uniform distribution | 12 |
| 1.3.2.3 | Comparing uniform and normal distributions . . . | 12 |
| 1.3.2.4 | Bimodal distributions | 14 |
| 1.3.2.5 | Comparing bimodal and normal distributions . . . | 16 |
| 1.3.3 | Applying ABC-Flow to simulated Gardner data | 16 |
| 1.3.4 | Applying ABC-Flow to experimental toggle switch data . . . | 19 |
| 1.3.4.1 | ATc induction | 19 |
| 1.3.4.2 | IPTG induction | 20 |
| 1.3.5 | Conclusions | 20 |

1 ABC-Flow

1.1 Introduction

1.2 Methods

Algorithm 1 ABC-Flow

```

1: Read input file
2: if ABC-Rejection then
3:   Sample from priors
4:   Simulate model
5:   Convert signal to intensity
6:   Measure distance to data
7:   Reject particles if  $d > \epsilon$ .
8:   if number of accepted particles == number of particles then
9:     Exit
10:  else
11:    Return to step 3.
12:  end if
13: end if

14: if ABC-SMC then
15:   Initialise  $\epsilon$ 
16:   population  $p \leftarrow 1$ 
17:   if  $p = 1$  then
18:     Sample particles ( $\theta$ ) from priors
19:   else
20:     Sample particles from previous population
21:     Perturb each particle by  $\pm$  half the range of the previous population (j)
    to obtain new perturbed population (i).
22:   end if
23:   Simulate model
24:   Convert signal to intensity
25:   Measure distance to data
26:   Reject particles if  $d > \epsilon$ .
27:   Calculate weight for each accepted  $\theta$ 
28:   
$$w_t^{(i)} = \begin{cases} 1, & \text{if } p = 0 \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N w_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}, \theta_t^{(i)})}, & \text{if } p \geq 0. \end{cases}$$

29:   Normalise weights
30:   repeat steps 17 - 29
31:   until  $\epsilon \leq \epsilon_T$ 
32: end if

```

1.3 Results

1.3.1 Distance Calculations

Algorithm 2 Distance calculation

- 1: $\text{Grid} \leftarrow \text{min}(\text{data}):\text{max}(\text{data}):\text{ngrid}$
 - 2: $\text{kD} = \text{kernel density estimation}(\text{data})$
 - 3: $\text{kS} = \text{kernel density estimation}(\text{simulations})$
 - 4: $\text{fD} = \text{kD}(\text{xx})$
 - 5: $\text{fS} = \text{kS}(\text{xx})$
 - 6: $d = \sum ((fD - fS)^2)$
-

1.3.2 Comparing 1D and 2D distances

In order to compare the 1D and 2D fitting to the data in ABC-Flow, we must first find out how comparable the distance measures are. Here we simulate two normal distributions, with identical mu and sigma, and calculate the distance between the two using the distance measure used in ABC-Flow. Doing this 1000 times, we then plot the distribution of epsilon. By doing that we can calculate the variance of the epsilon distribution, and find out the error that can be expected when measuring the distance in ABC-Flow. By doing that in 1D and in 2D we can compare the epsilon variances.

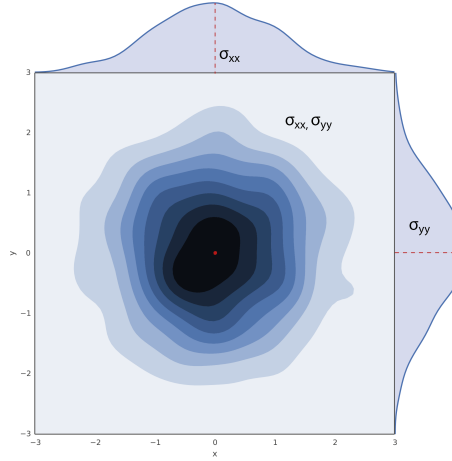


Figure 1.1 Comparing 1D and 2D distributions.

1.3.2.1 Normal distribution

Here, we simulate two normal distributions, with $\mu = 0$ and $\sigma = 1$ 1000 times and measure the distance. In the 2D case, we simulate two multivariate normal distributions, with $\mu = 0$ and covariance=[0.5 0, 0 0.5]. In the 1D case the median of the epsilon distribution was 0.00196 and the variance 0.0016. In the 2D case the median was 3.41e-14 and the variance 1.97e-07.

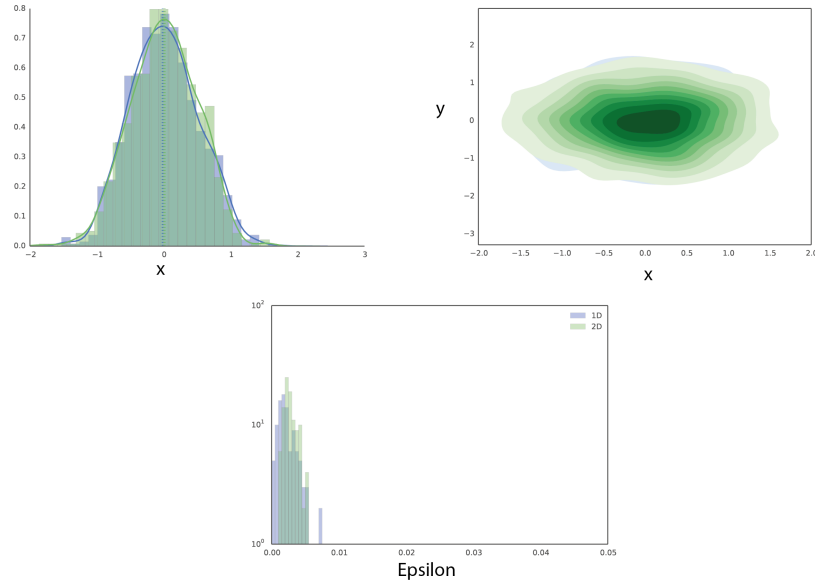


Figure 1.2 Epsilon distribution for 1D (blue) and 2D (green) distances.

Next we study the effect that the number of data points have on the variance of the epsilons.

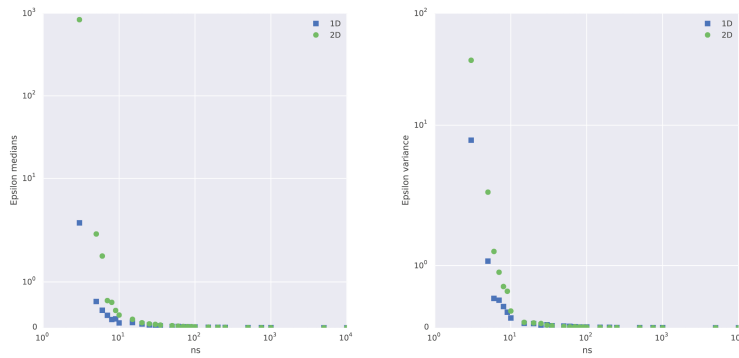


Figure 1.3 Epsilon distribution medians and variance over number of data points.

As the number of data points increases, the distance calculation becomes more precise and more accurate. The median distance approaches zero, and the variance of the epsilon distribution decreases to zero. From Figure 1.3 we conclude that the minimum number of data points that need to be used to calculate the distance between the distributions in ABC-Flow is 100.

The next parameter to be optimised is the bin size used in the distance calculation. In both 1D and 2D, the space is divided into bins, and the distance between corresponding bins in the data and the simulated data is calculated. The overall distance between two distributions is the sum of the distances between corresponding bins.

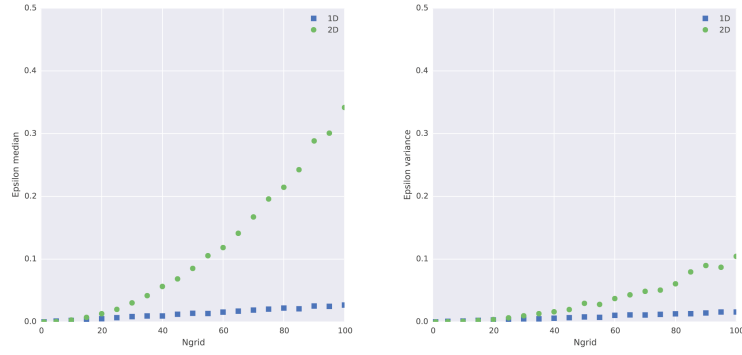


Figure 1.4 Epsilon distribution medians and variance vary with the size of bins used.

We find that the 2D distance is more sensitive to the bin size than the 1D distance. From Figure 1.4 we conclude that for a sample size of 100 data points, the optimal bin size for 1D and 2D distance calculation is 10. The above optimizations have been made by using standard normal distributions, of $\mu = 0$ and $\sigma = 1$. Here we investigate whether the distance calculation depends on the μ and σ of the distributions.

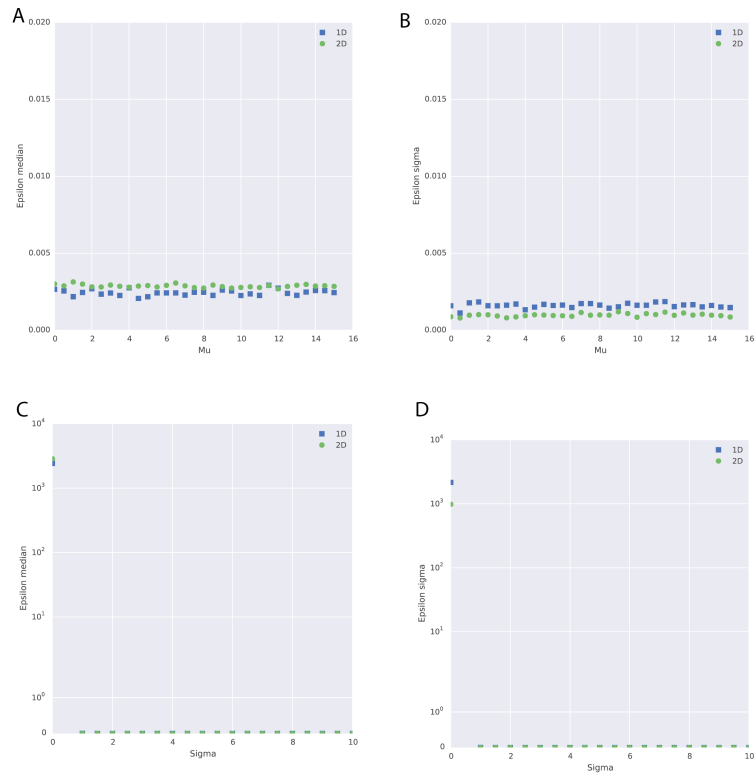


Figure 1.5 The distance calculations do not depend on the μ and σ of the distributions. The median (A) and variance (B) of the epsilons remains constant with increasing μ (C,D) as well as with increasing σ .

Next, by varying the amount of μ and σ by which the two normals differ, we can find out the dynamical range of the epsilons. Whether two distributions are identical or vary by a large amount, we can get an estimate of how much the epsilon will vary from one to the other, both in 1D and 2D.

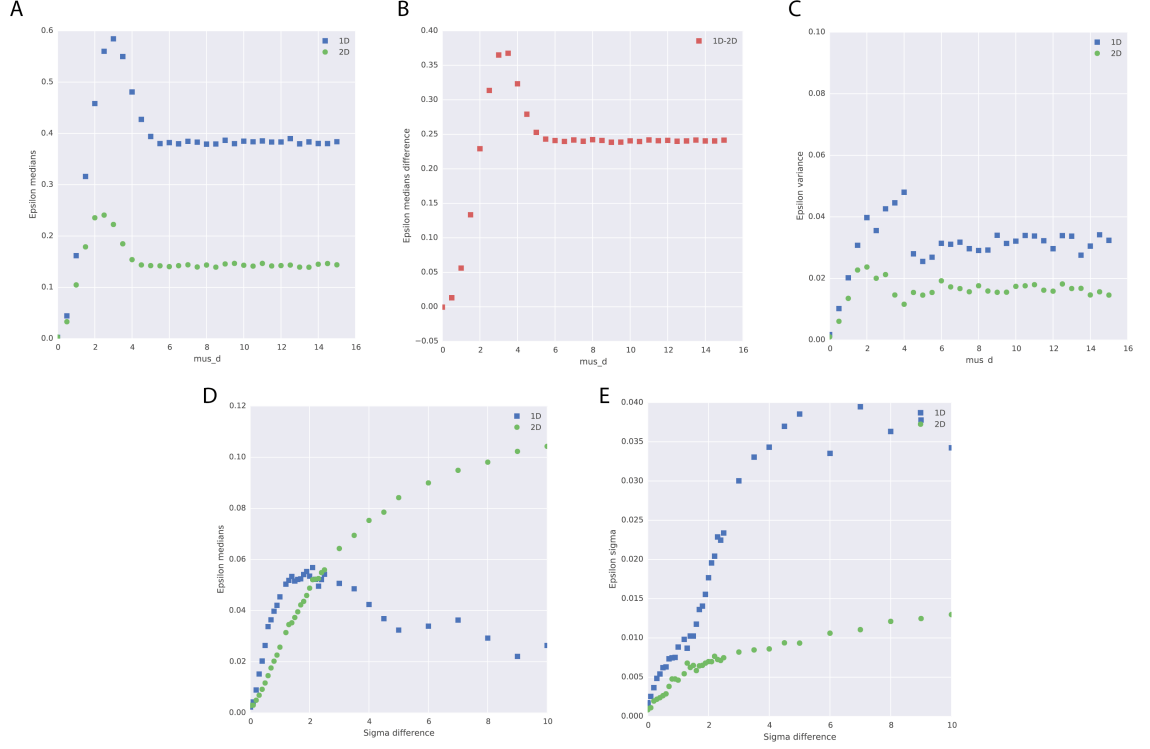


Figure 1.6 (A) The range by which epsilon varies as the difference between the μ of the distributions increases. (B) The difference between the epsilon distribution medians in the 1D and 2D case is not constant with increasing μ difference between the data sets. (C) The variance of the epsilon distributions remains relatively constant with increasing μ . (D) The median of the epsilon distributions varies by a small amount with increasing difference in the σ of the distributions, but the variance (E) remains relatively constant.

From Figure 1.6 we find that as the difference in μ increases the epsilon medians reach a plateau. We find that beyond a difference of 4 in μ , the distance calculation cannot further separate the distances. This can be caused by the fact that when first dividing the space into bins, the range of the data is used to define the grid. If all the simulated data is located outside that grid, the algorithm can no longer distinguish between them, and will only allocate them as outside the range. The variance of the epsilon distributions does not vary significantly with increasing difference in μ . As the difference in σ increases between the distributions, we find that the median in the 2D distance calculation increases but not in the 1D. Note, that the range of the difference in the epsilon medians is small and thus we conclude that differences in the μ of a distribution are much better detected than the differences in the σ . The variance of the epsilon distribution when σ is varied does not change significantly with increasing σ difference.

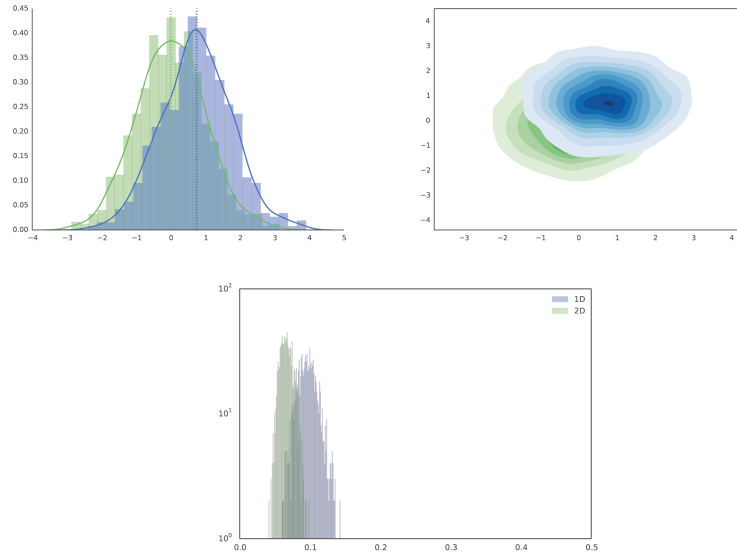


Figure 1.7 The difference in distributions when epsilon median is smaller than 0.1 in 1D and 2D

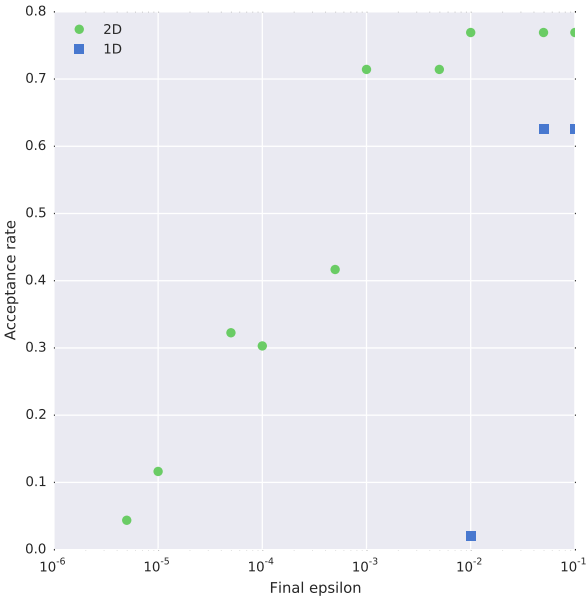


Figure 1.8 Acceptance rate drops rapidly in the 1D case

1.3.2.2 Uniform distribution

We study the epsilon distribution variance in the uniform distribution, $[0, 1]$.

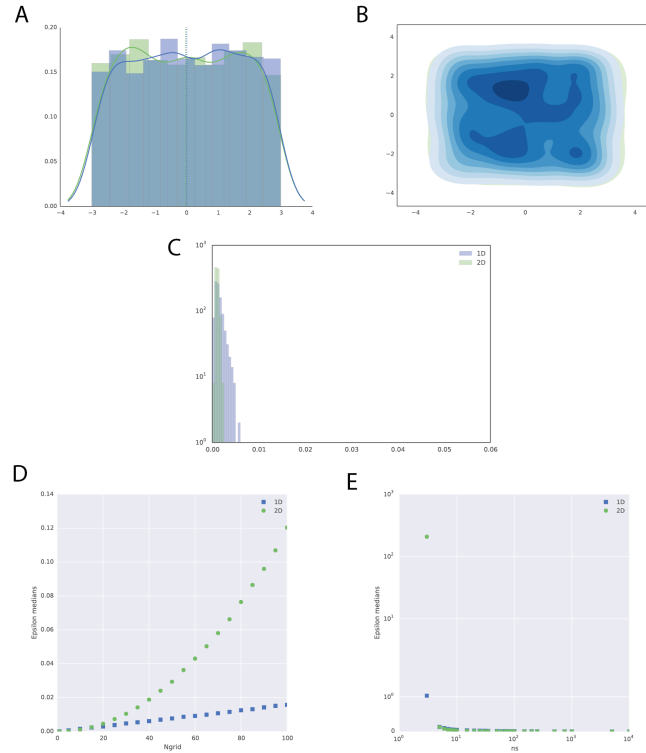


Figure 1.9 Comparing the 1D and 2D distances between uniform distributions. (A) and (B) show samples of the uniform distributions compared in 1D and 2D respectively. (C) The epsilon distributions in the 1D and 2D distances are equivalent (D,E) Epsilon medians and variance varies with how dense the grid is in the 2D case.

1.3.2.3 Comparing uniform and normal distributions

When comparing a normally distributed data set to a uniformly distributed data set, we don't see a great difference between the 1D and 2D epsilons.

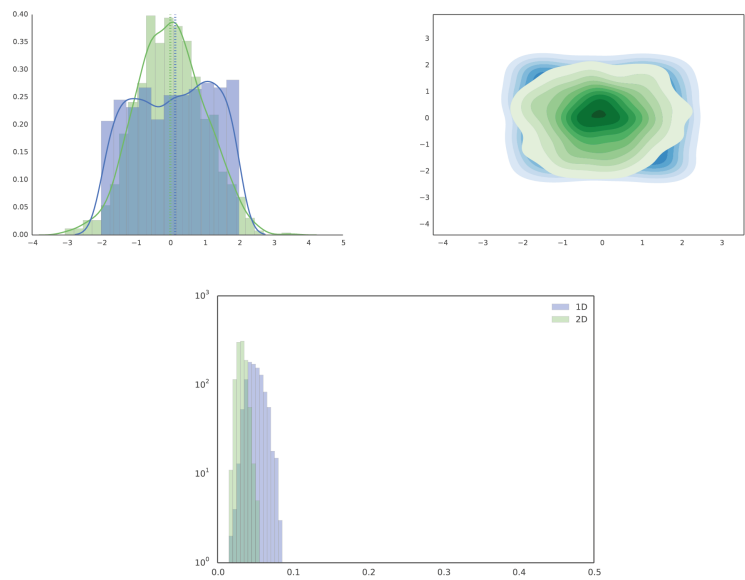


Figure 1.10 Comparing normally distributed data to uniformly distributed simulations.

1.3.2.4 Bimodal distributions

Another type of distribution that is commonly found in Flow cytometry experiments, is a bimodal distribution. Here we test whether the 1D and 2D distances are equivalent when measuring distances in this type of distributions.

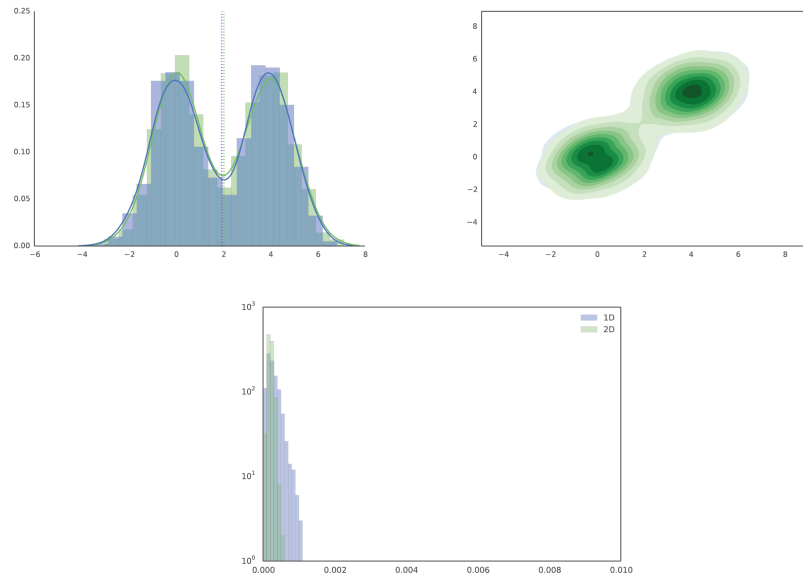


Figure 1.11 Comparing the 1D and 2D distances between bimodal distributions.

We find that the two distances are comparable when dealing with this kind of distribution. Next, we examine how the distance between bimodal distributions with different μ varies in the 1D and 2D cases. We find a similar behaviour like the one observed when comparing normal distributions with increasinglt different μ (Figure 1.6).

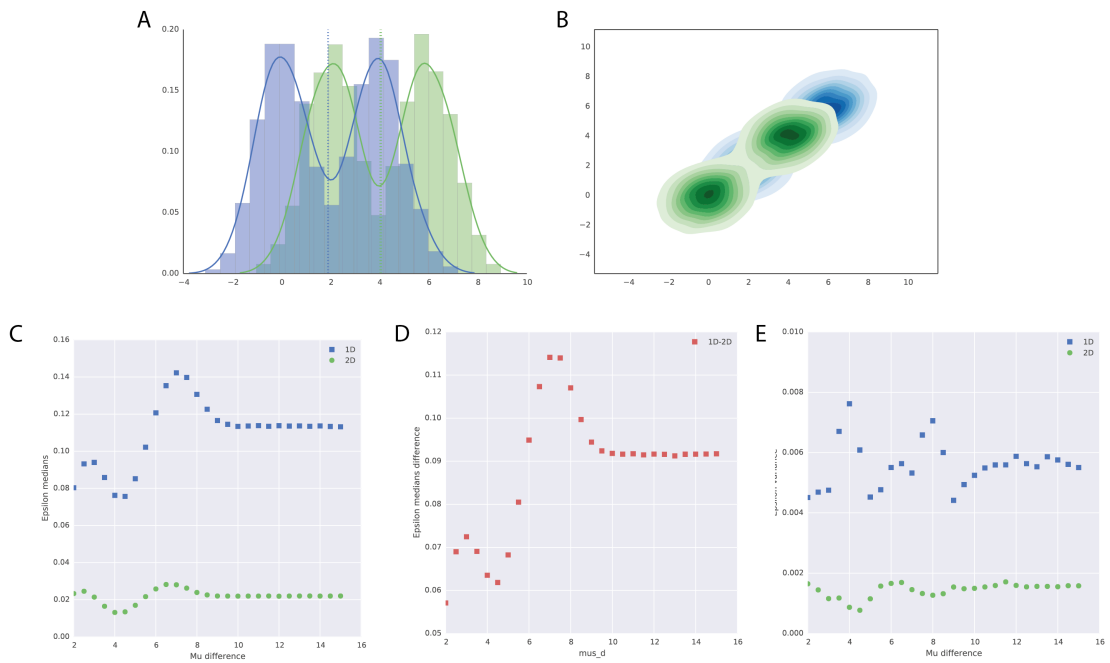


Figure 1.12 Comparing the 1D and 2D distances between bimodal distributions. (A) and (B) show samples of the bimodal distributions compared in 1D and 2D respectively with a μ difference of 4 between simulations and data. (C,D) The range by which epsilon median and variance varies as the difference between the μ of the distributions increases.

1.3.2.5 Comparing bimodal and normal distributions

Finally, we study how the distances perform when comparing a bimodal with a normal distribution. We test the distances by using a bimodal distribution and a series of normal distributions with increasing μ , in 1D and 2D. We find that epsilon is the lowest when the μ of the normal distribution corresponds to the μ of one of the two peaks in the bimodal distribution and the highest when there is no overlap between the distributions.

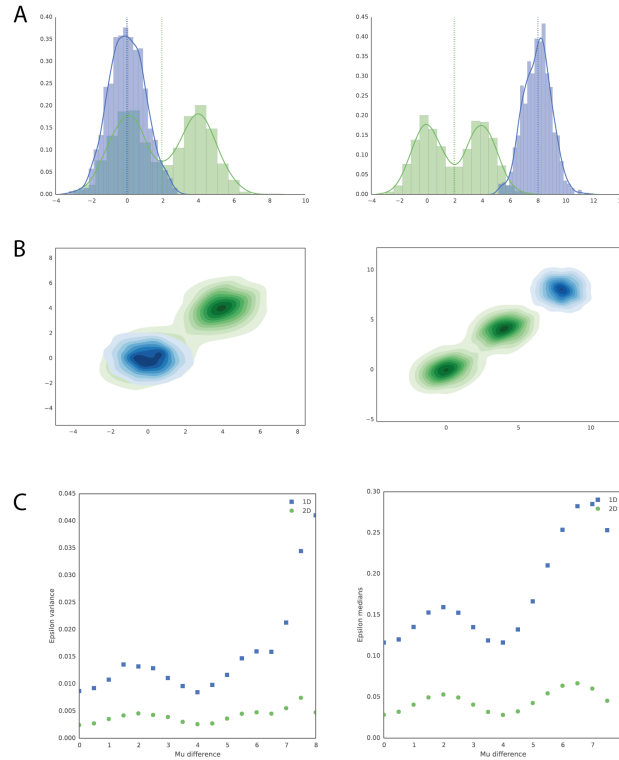


Figure 1.13 Comparing a multimodal to a normal distribution, in 1D and 2D. (A, B) We vary the μ of the normal distribution from equal to the μ of the first peak of the bimodal distribution to beyond the range of the bimodal distribution. (C) We find that epsilon median and variance are at the lowest when the μ of the normal distribution is equal to the μ of one of the peaks of the bimodal distribution.

1.3.3 Applying ABC-Flow to simulated Gardner data

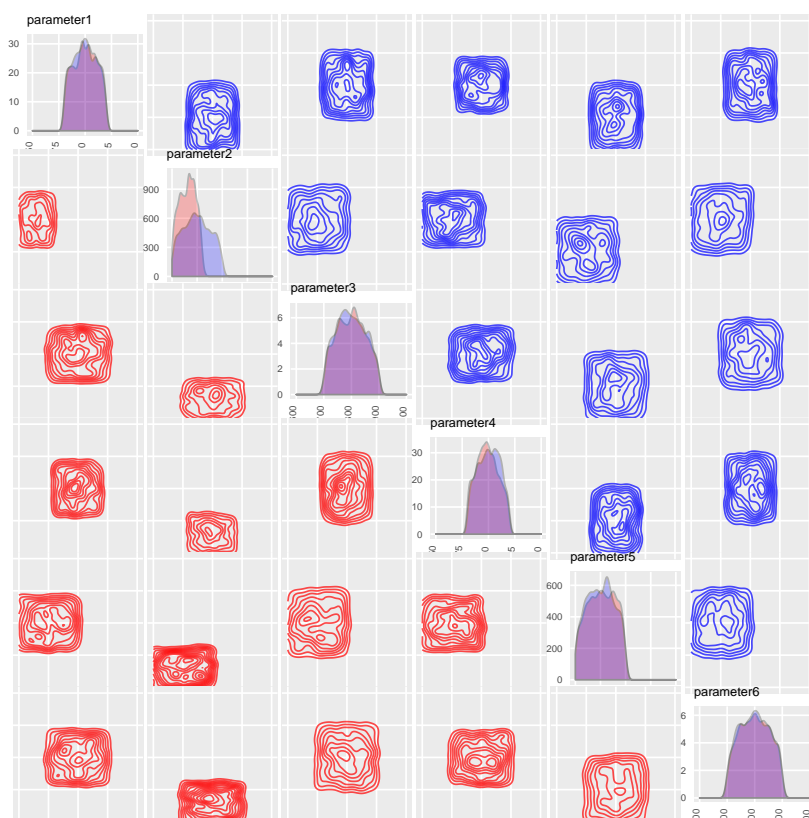


Figure 1.14 Comparing the fit in 1D (red) Vs 2D (blue) in ABC-Flow, when $\epsilon=0.1$.

In order to improve the identifiability of the parameters in the 2D case, we lower the final ϵ : We find increased identifiability for parameters 2 and 5. Next we will

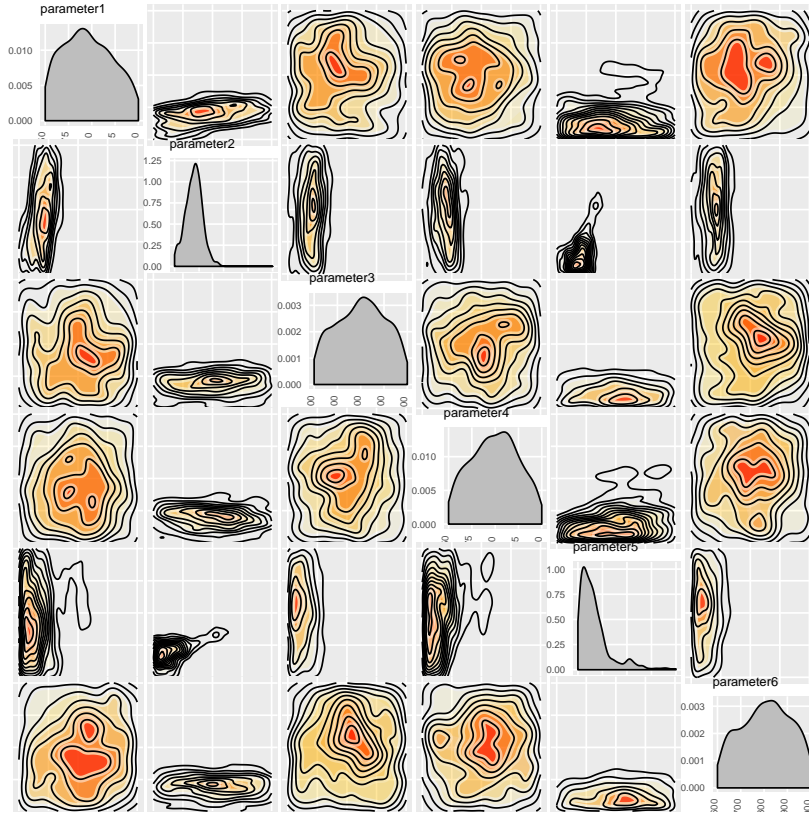


Figure 1.15 $\epsilon=0.00001$.

use this epsilon in the 1D case.

1.3.4 Applying ABC-Flow to experimental toggle switch data

1.3.4.1 ATc induction

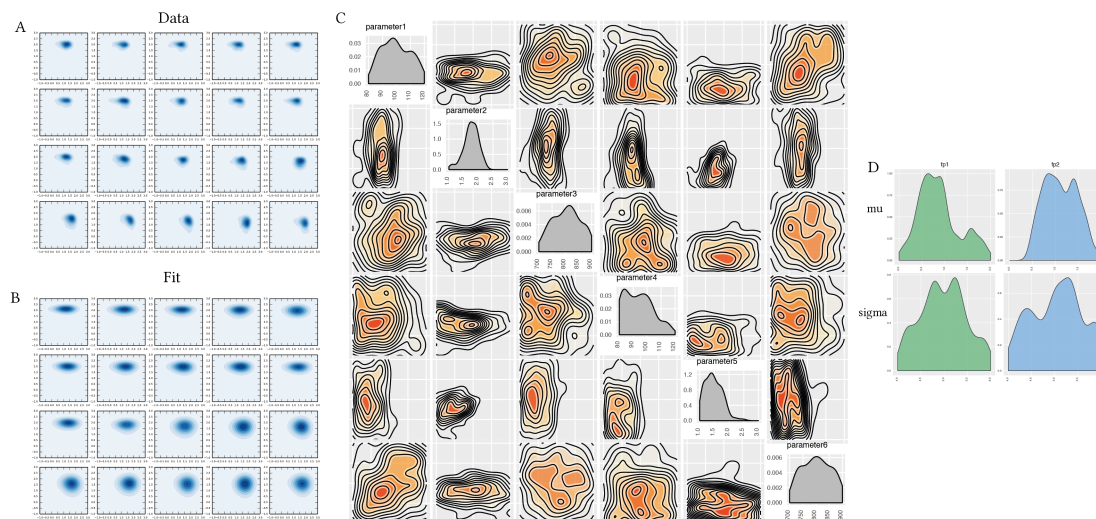


Figure 1.16 Real data 2D fit

1.3.4.2 IPTG induction

1.3.5 Conclusions