0.1 Uncertainty in the Droplet Radius

If we look at the equation **??** we can see that in the case we are dealing with there are three main unknown parameters if we consider that we are acquiring the velocity from the data space supplied by the experimental part. These parameters are the radius, the density and the drag coefficient. Note also that ,additionally to the velocity, we are getting information of the droplet radius in each sample.
The main idea of this first scenario is setting the drag coefficient as a constant value. With this approach the model is more restrictive in the sense that we are allowing only uncertainties in two parameters: the radius and the density.

0.1.0.1 Drag Coefficient approximation

Taking this into account, the inverse problem resolution will be based on the unknown two parameters. At this point is important to approximate the fixed variable (drag coefficient) to the most certain value. Usually the drag coefficient is obtained balancing the forces actuation over a body But in this case, we don't have any information regarding to the forces, so we need to approximate the drag coefficient somehow. One option is using tabulated values that comes from experiments but the problem is that the majority of this experiments are based on air or in solid bodies. For that reason in this specific problem, we are going to consider a polynomial expression from **??** which expresses the drag coefficient in function of the Reynolds Number:

$$C_d = \frac{24}{Re} + \frac{2.6\left(\frac{Re}{5.0}\right)}{1+\left(\frac{Re}{5.0}\right)^{1.52}} + \frac{0.411\left(\frac{Re}{2.63\dot{1}0^5}\right)^{-7.94}}{1+\left(\frac{Re}{2.63\dot{1}0^5}\right)^{-8.00}} \tag{1}$$

Where the Reynolds Number is calculated in function of the droplet radius r_d , the water density ρ and the kinematic viscosity μ:

$$Re = \frac{\rho r_d U_d}{\mu} \tag{2}$$

Thus, using the experimental data, the drag coefficient has been calculated for each sample and taking the average of those values we get an approximation for the drag coefficient which is approximately 0.77 it seems quite large for the typical value of a sphere in a fluid (around 0.5) but will be our first approximation. Knowing the exact vale for the droplet density we can approximate this value using the physical model [**??**]. In this case we get a value over 1.25. As we have two dissimilar values we will solve the inverse problem for those values and two more in order to study their influence in the posterior distribution.

0.1.0.2 Inferring Radius Results

Once we know the unknown and the fixed parameters, we can study the inverse problem and solve it using the application. As it has been described before, for the resolution, QUESO needs two inputs files.The *queso.in* describes basically the QUESO resolution options and the $model_input.in$ which describes the main characteristics of the scenario:

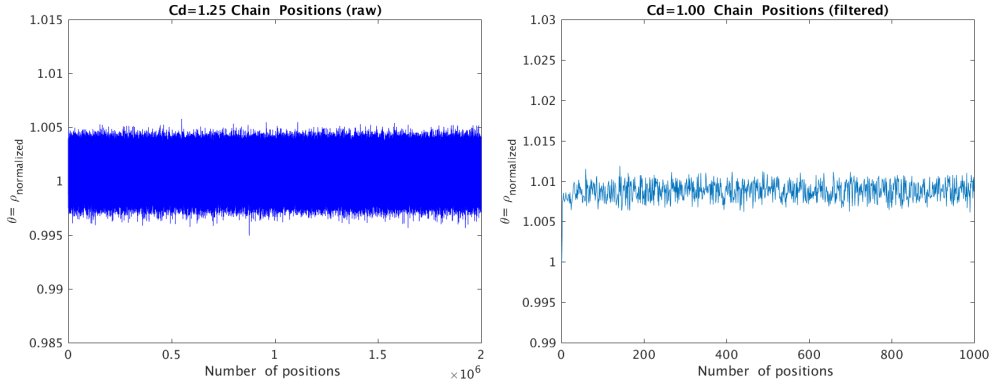
1. For establishing the prior, the application needs information *a priori* concerning the parameters which consists in a minimum and a maximum value for the parameters domain. Remark that the variables will be studied in their normalized value. Thus if we want the exact value we must multiply by the nominal value.

2. In this scenario, we will give the specific value to the drag coefficient.

3. The other variables of the physical model are the water density and the gravity which are constant values for all scenarios.

After setting the input file, we can solve the inverse problem and plot the posterior information using MATLAB. The idea behind the different scenarios is to study the influence of the parameters in the posterior distribution of the density and thus validate the application. The results are validated considering

that we already know the value of the droplet density (The value of the castor oil density is $961 kg/m^3$) and then a good approximation would be the one which is centered around this value.

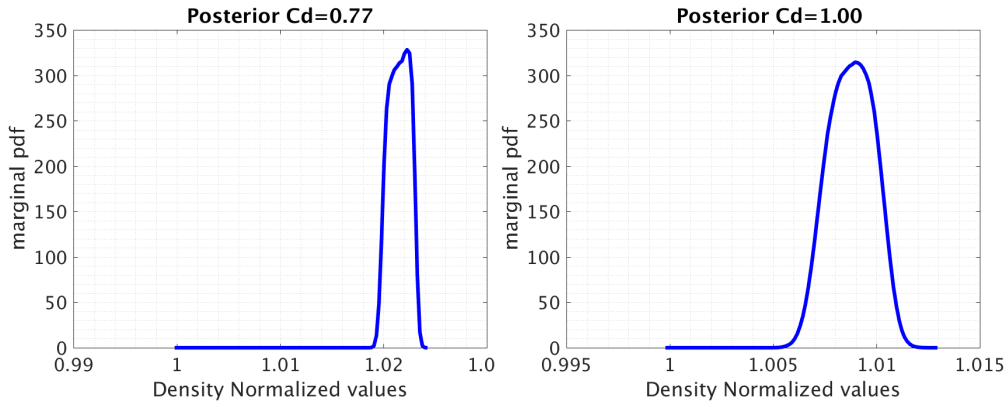


(a) Raw Chain

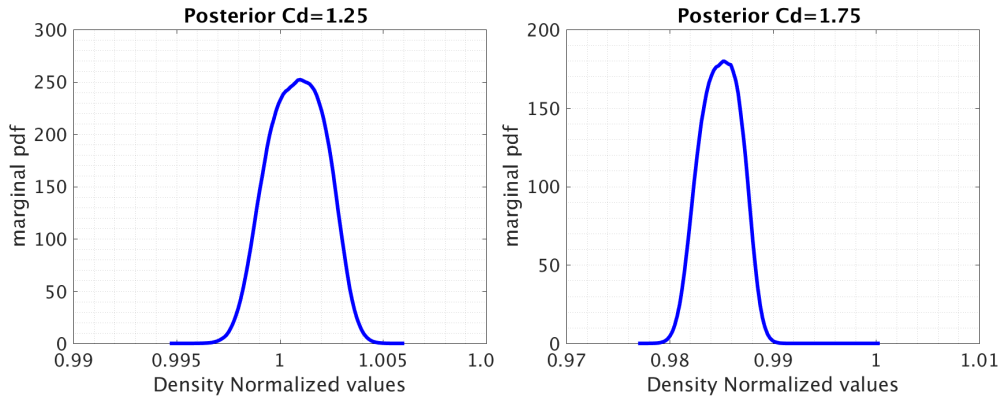
(b) Filtered Chain

Figure 0.1: Markov Chain

Previous to show the main results, we need to make give some points from the data we get from the output files. Basically in the output files we get the samples of the Markov Chain and using this data we get the information related to the parameters such as the histograms, the density distribution,...Looking at this chains, one can appreciate if the samples fits with the interval set for the prior distribution. In the figure (0.1) we can see the shape of either the raw chain and the filtered. Concerning to the second one it will be shown later that aims to have the same information than the whole chain.



(a) Cd=0.77 (b) Cd=1.00



(c) Cd=1.25 (d) Cd=1.75

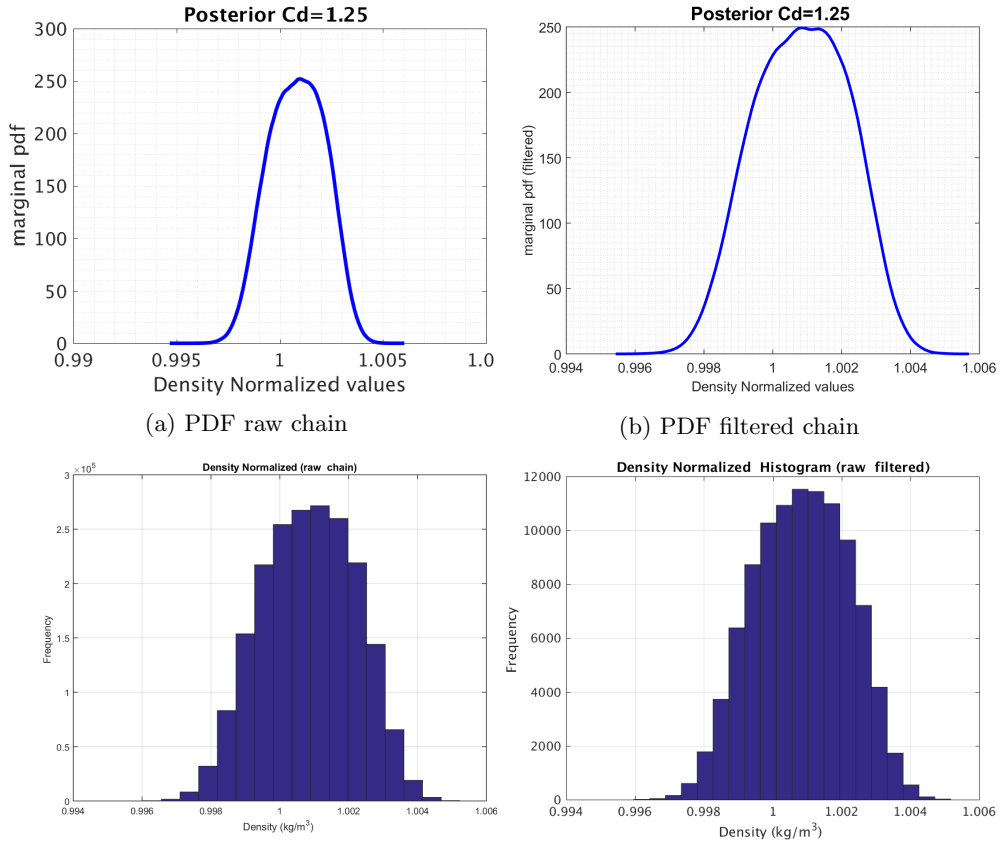
Figure 0.2: Droplet density PDF for different CD's .

In the previous section we have demonstrate that one possible value of the drag coefficient for our experiment was approximately 0.77 but analyzing the figure 0.2 we can see that this value gives a posterior distribution quite separated from the nominal value (which is represented as the value of 1). For the figure (b) we have the same problem but in this case we can note that we are getting close to the right value which may be $Cd = 1.25$ as the posterior distribution is practically centered at 1. Finally it can be seen in the figure (c) that if we give high values to the drag coefficient the density tends to decrease which may lead us to conclude that the density is inversely proportional to the drag coefficient which is not true so we need to take care of this kind of conclusions an be aware at all moments the problem we are analyzing. In the same way is important analyze the density distribution in order to avoid values higher than the water density as in this case the problem has no sense.
At this point we

we can use the histograms to see in a best visual way the change in distribution Then regarding your last comment ([0.6,0.8]—data)([0.6,0.8]—data) represents the belief is in [0.6,0.8][0.6,0.8] after having observed the data. Conversely the highest posterior density (HPD) interval (which will be something close to [0.6,0.8] in your case), gives you the smallest interval covering a given subset (typical 95

There is another interesting point regarding the different output files we can get from QUESO. Is the case of the filtered chains. The interested in the filtered chain is due to the fact that for many problems is important to retain the information of the posterior distribution. However storing all the information can very expensive in terms of memory and processing. Is when the filtered chain makes sense because this chain is not essential in this problem, but can be very important for sequential processing where new data is processed upon arrival in a short period of time and is also needed save as much memory as possible.
In thh figure 0.3 it can be seen the PDf and the histogram for the whole chain (raw chain) and the filtered one. As the figures show, the confidence interval is very similar for two chains considering that

the number of samples are 2 million for the raw chain and only 1000 samples for the filtered chain.



(a) PDF raw chain

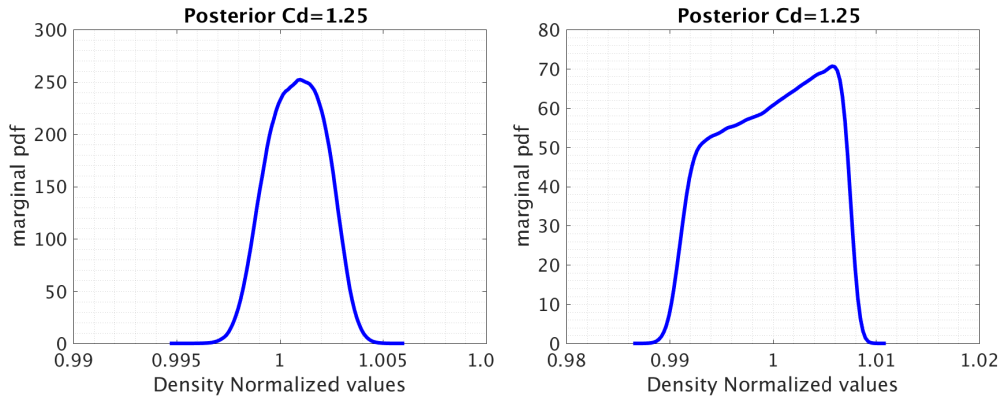
(b) PDF filtered chain

(c) Histogram raw chain

(d) Histogram filtered chain

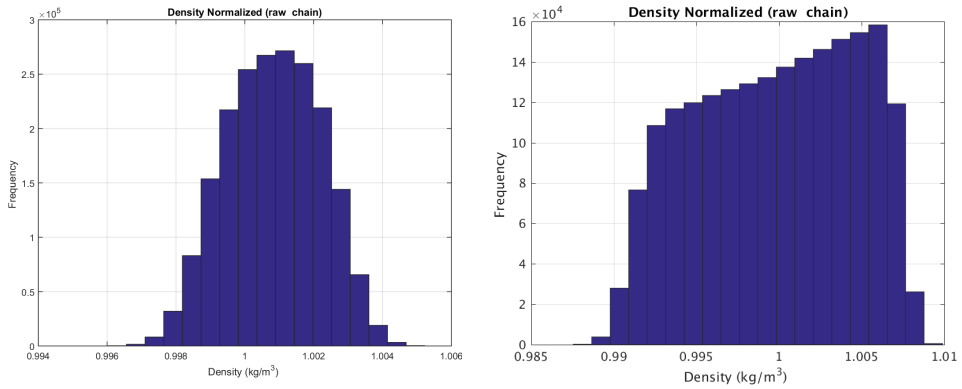
Figure 0.3: PDF and Histogram for the raw and filtered chain

Additionally, is also interesting analyzing the effect that the uncertainty has over the density distribution. It has been explained that in the input file the prior information is set. So, for the case of the radius parameter an interval has been chosen based on the experiment. But what happens if the prior interval of the radius gets longer. In this case the uncertainty of knowing the radius is higher and this change affects to the posterior distribution (figure 0.4) .



(a) PDF R_d in [0.95-1.05]

(b) PDF R_d in [0.8-1.2]



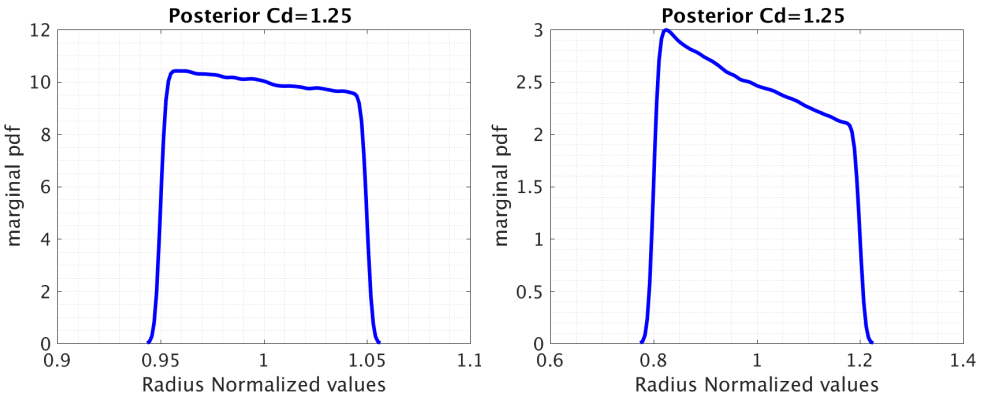
(c) HistogramR_d in [0.95-1.05]

(d) Histogram R_d in [0.8-1.2]

Figure 0.4: PDF and Histogram varying the prior radius interval.

It can be noted that the fact of increasing the uncertainty in the radius interval makes a change in the posterior distributions. So, the effect of the uncertainty leads to an increase of the confidence interval for droplet density posterior although the distribution is practically centered to the real value (the mean is over 1).

Finally, in the figure 0.5 the results for the radius posterior distribution are plotted. Basically, for both cases there is no variation respect to the prior interval, which means that there is no extra information got from the data space (in this case we just get the information of the radius for each sample). For getting results more accurate we would need more information in the data space.



(a) Radius PDF R_d in [0.95-1.05] (b) Radius PDF R_d in [0.8-1.2]

Figure 0.5: PDF and Histogram varying the prior radius interval.

0.1.1 Uncertainty in the Drag coefficient

This scenario is similar to the previous one in the number of unknown parameters. Now is proposed to set the radius as a constant value and the drag coefficient as the second parameter to infer. Concerning the scenario configuration this means that the radius is excluded from the parameter domain of QUESO. It does not mean that the radius is a fixed value for all the realizations. In fact this is one of the differences respect to the previous scenario as meanwhile the drag coefficient was completely constant in all moments for the first case, here the radius changes for each sample data (from the experimental vector, we get the radius, droplet position and the time as we have described before) Therefore, in this case we don't need to include any value of the radius in the input file allowing little more uncertainty.
Thus, the idea behind this scenario is to study the effect of the uncertainty in the droplet density posterior. In this case, we will study the effect of the drag coefficient prior:

- **Case 1**: This scenario is a first approximation where the drag coefficient is set to a medium range interval centered at the more probable value (1.25 which was shown in the previous section).
- **Case 2**: In this one, the interval gets longer.
- **Case 3**: The aim of this case is to see if the nominal value of the drag (which is decreased) coefficient greatly affects density posteriors.
- **Case 4**: This case aims to study the effect of setting a large uncertainty in the drag coefficient parameter.
- **Case 5**: Similar to the previous one but in this case the density interval becomes longer.
- **Case 6**: The last case studies the inverse problem when there is not much uncertainty allowed.

In the table 1 the different values are shown:

Name/Parameter	C_d **Nominal** $[-]$	C_d **Prior**	ρ **Nominal** kg/m^3	ρ **Prior**
Case 1	1.25	$[0.9 - 1.1]$	961	$[0.9 - 1.1]$
Case 2	1.25	$[0.8 - 1.2]$	961	$[0.9 - 1.1]$
Case 3	1.00	$[0.8 - 1.2]$	961	$[0.8 - 1.2]$
Case 4	1.25	$[0.5 - 3.0]$	961	$[0.8 - 1.2]$
Case 5	1.25	$[0.9 - 1.1]$	961	$[0.8 - 1.2]$
Case 6	1.25	$[0.995 - 1.005]$	961	$[0.9 - 1.1]$

Table 1: Characteristics for the different cases studied.

In the figure 0.6 the posterior distributions for the droplet density in the different cases are plotted. The first conclusion we can get from these plots is that the uncertainty in the priors leads to a higher uncertainty in the posteriors (if we look at the density distributions, the confidence interval is longer for those cases where the prior interval is longer). This situation seems to be within normalcy. In the same way, it can be noted that the posterior are, in general, no centered to 1. In the table 2 are gathered the mean for the different posteriors and one appreciate that we are getting densities below the nominal value. In addition to this remark, and analyzing the figure 0.7 it can be noted that the posterior tends to give more probability to a higher values of the drag coefficient and lower values to the density.

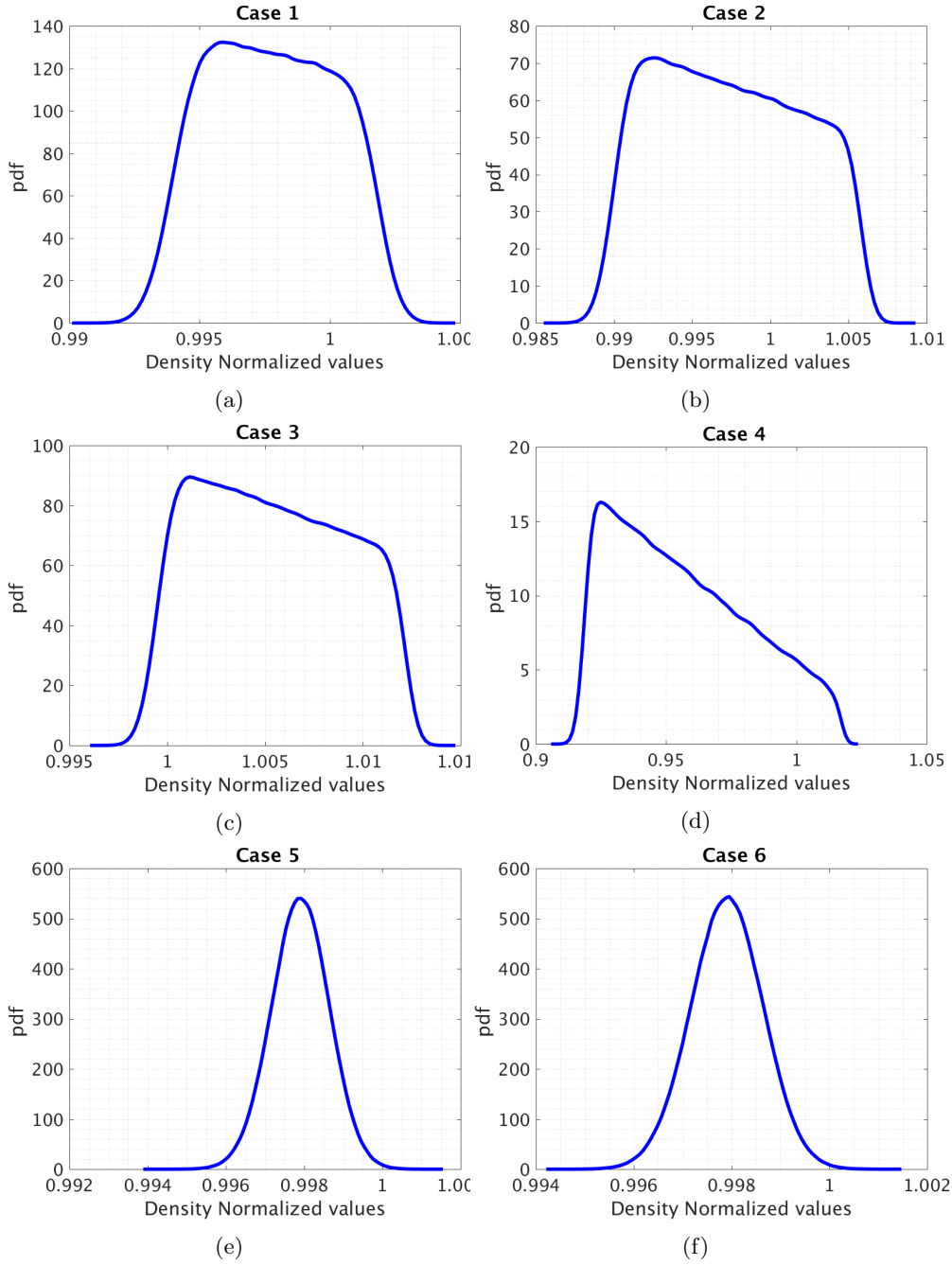
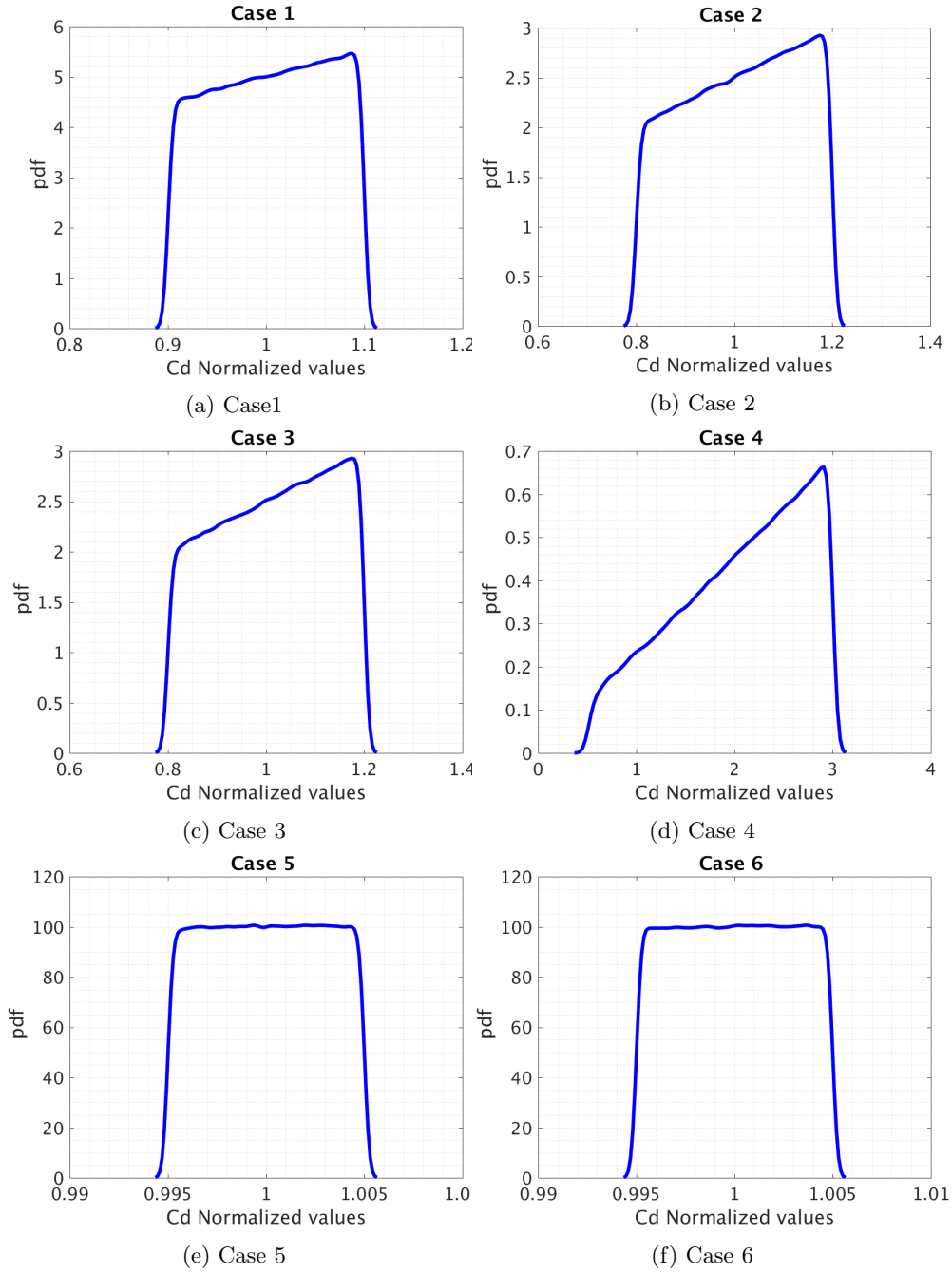


Figure 0.6: Droplet Posterior distribution for different cases .

For that reason it has been shown that the selection of the priors is not irrelevant as, for instance, in the case (d), when the difference between the minimum and the maximum is the highest value, the posterior gives a fairly high drag coefficient (which is a completely opposite solution we got from the scenario 1). This issue should be due to the high allowed uncertainty and the few information we are acquiring from the experimental data. This lack of information makes also the drag posterior distributions to be quite flat (flatness).



(a) Case1 (b) Case 2

(c) Case 3 (d) Case 4

(e) Case 5 (f) Case 6

Figure 0.7: Drag Coefficient Posterior distribution for different cases .

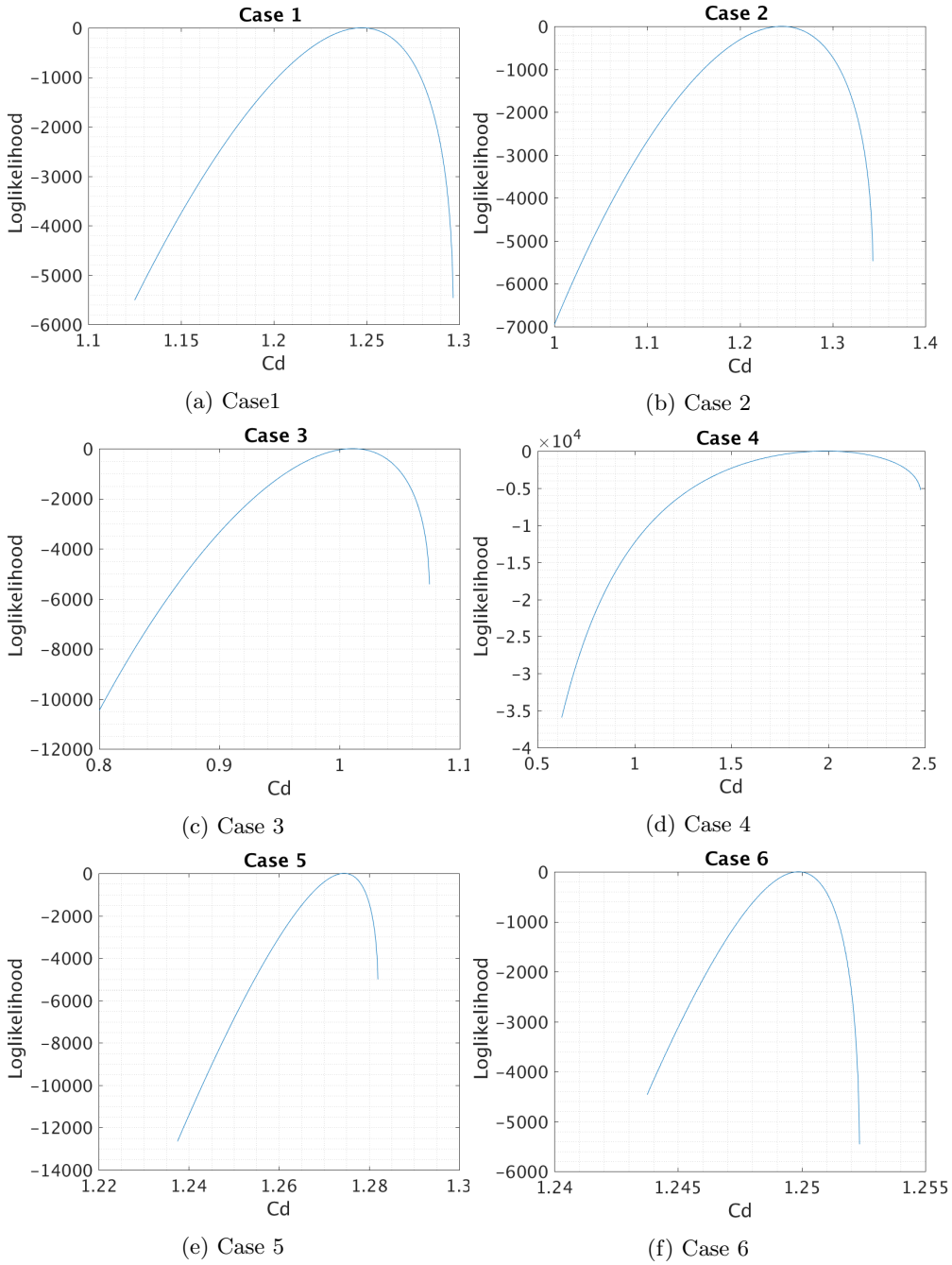
An interesting feature obtained from the table 2 is that the values of the mean for the case *1,2,5,6* are very near so, we can conclude that if the uncertainty is treated properly, we should not have problems for getting accurate values for the density.

Name/Parameter	**Density Normalized Mean**	C_d **Normalized Mean**
Case 1	0.9978	1.003
Case 2	0.9974	1.0132
Case 3	1.0054	1.0132
Case 4	0.9564	2.0491
Case 5	0.9979	1.00
Case 6	0.9979	1.00

Table 2: Results for the Mean Value in the different cases studied.

After analyzing all these results and looking at the influence of the prior, what happens if we don't have any prior information related to the nominal values. The answer can be obtained in the log-likelihood plots. In this kind of graphics we can see graphically around which values is the zone of greatest probability for the parameters. In this scenario is useful to set the interval for the drag coefficient. Thus we can establish the interval in the best way and then, the droplet density posterior can be got more accurately.

In the figure 0.8the drag coefficient log-likelihood is plotted for the different cases. This distribution shows, in general, that inside the prior interval, there is a zone of higher probability. With this information is easier to set the prior in an iterative process and reduce the uncertainty for getting more accurate values.



(a) Case1 (b) Case 2

(c) Case 3 (d) Case 4

(e) Case 5 (f) Case 6

Figure 0.8: Log-likelihood for the density in the different cases .