Particle Filter on Discrete Lorenz63 Model upon bidimensional non-linear observations

Miguel Franco Pérez
Dept. Statistics and Operations Research [Intern]
Polithecnical University of Catalunya (UPC)
Barcelona, Spain
miguel.franco@upc.edu
franper.miguel@gmail.com
OrcID: 0000-0001-8205-9679

***Abstract*—In this mini-project, we will implement the Particle Filter algorithm (Sequential Importance Sampling Resampling) on a state ruled by Lorenz63 chaotic dynamic by observing just a set of non-linear observations arising from that state. This will allow us to really understand how and why the particle system works for estimating the latent state even over complex dynamic systems with non-linear and noisy observations.**

***Index Terms*—Particle Filter, Sequential Importance Sampling Resampling, Lorenz Model, Dynamic Systems, Bayesian Non-Linear Inference**

I. Introduction

Particle Filter is an advanced tool for dealing with stochastic (thus, non-deterministic) latent states estimation. It is based on an extension of Importance Sampling procedure, embedded as a consequence as a MonteCarlo technique.

Lorenz Model is a well-known chaotic dynamical system. Small perturbations on the initial point of the system leads to major changes on the final result. Lorenz dynamic propagation is non-linear.

Furthermore, the available observations will arise from that latent (non-observed) state through a non-linear bi-dimensional observation.

The capability of particle filter for dealing with such a complex framework shows its powerful potential for solving intrincated systems as long as we have their inner dynamics well-characterized.

II. Data Generation

A. State: Discrete Lorenz Model

Lorenz model is a dynamical continuous model. Here we will work under the Euler-Maruyama discretization of that system with time-step h. The state dynamic is characterized as a three-dimensional vector $X_t = (X_{1,t},\ X_{2,t},\ X_{3,t})$ whose components follow the dynamic

$$\begin{aligned}
X_{1,t} &= X_{1,t-1} - hs(X_{1,t-1} - X_{2,t-1}) + \sigma\sqrt{h}Z_{1,t} \\
X_{2,t} &= X_{2,t-1} + h(rX_{1,t-1} - X_{2,t-1} - X_{1,t-1}X_{3,t-1}) + \sigma\sqrt{h}Z_{2,t} \\
X_{3,t} &= X_{3,t-1} + h(X_{1,t-1}X_{2,t-1} - bX_{3,t-1}) + \sigma\sqrt{h}Z_{3,t} \\
&\text{Where } Z_{i,t} \sim \mathcal{N}(\mu_Z = 0, \sigma_Z = 1) \\
&t = 1, 2, ..., T
\end{aligned} \tag{1}$$

We may try to synthesize the behavior by stating that the future of each coordinate of the state $(X_{1,t},\ X_{2,t},\ X_{3,t})$ depends on the previous value of that coordinate plus an interaction with the other coordinates (except for $X_{1,t}$, which will depend on the interaction of its past with $X_{2,t}$, not with $X_{3,t}$). Finally, a random identical and independently distributed noise (iid) white noise is added to create a stochastic dynamic (non-deterministic).

This behavior is tuned by a set of parameters $\{h,\ s, r, b\}$. $\{h\}$ parameter is common to all coordinates and results from the discretizations of the continuous time; it acts a multiplier of the effect of the interaction with the other coordinates. The remaining parameters $\{s, r, b\}$ just affect a single future coordinate $\{X_{1,t+1}, X_{2,t+1}, X_{3,t+1}\}$ respectively.

B. Observation: Bidimensional Non-linear

From the previous state dynamic, we will observe a bi-dimensional realization $Y_t = (Y_{1,t},\ Y_{2,t})$ that, as we introduced, will arise from another non-linear mechanism:

$$\begin{aligned}
Y_{1,t} &= 0.1\ X_{1,t}X_{2,t} + \sigma_u U_{1,t} \\
Y_{2,t} &= 0.1\ X_{1,t}X_{3,t} + \sigma_u U_{2,t} \\
&\text{Where } U_{i,t} \sim \mathcal{N}(\mu_U = 0, \sigma_U = 1) \\
&t = 1, 2, ..., T
\end{aligned} \tag{2}$$

Again, we may try to synthesize the behavior by stating that each coordinate of the observation $(Y_{1,t},\ Y_{2,t})$ depends on the interaction of the state's first coordinate $X_{1,t}$ with the second and third coordinates respectively. That interaction is controlled by a common fixed factor of 0.1. Finally, a random identical and independently distributed noise (iid) white noise is added to create a stochastic (non-deterministic) observation procedure.

C. Partial Observation

For this project, we will consider that observations Y_t will just be available on the initial observation $(t = 1)$ and later on each B discrete steps. Thus, our final observation set will follow the structure of

$$Y_t = \begin{cases} (Y_{1,t} \ , Y_{2,t}) & \text{if } t \in \{1, Bi\} \quad \forall i \in \mathbb{N} \\ (\text{NA}, \text{NA}) & \text{else} \end{cases} \tag{3}$$

D. Parameters & initial state selection

For this project, as instructed, we determined to use the following set of parameters:

$$\begin{aligned} T &= 20000 \\ \\ h &= 10^{-3} = 0.001 \\ s &= 10 \\ r &= 28 \\ b &= \frac{8}{3} \\ \sigma &= 1 \\ \\ B &= 10 \\ \sigma_u &= 2 \end{aligned} \tag{4}$$

This particular selection of $\{s = 10, r = 28, b = \frac{8}{3}\}$ is what leads to exactly the Lorenz68 model (discretized).
Finally, as initial state for $t = 1$ we set a point in the attractor of this Lorenz63 model:

$$X_1 = (X_{1,1} = -5.9165, \ X_{2,1} = -5.5233, \ X_{3,1} = 24.5723) \tag{5}$$

E. Resulting Data

The resulting data is a matrix that for each time t until T stores the value for the three components of X_t and the two components of Y_t (when available) like [1]

$$\begin{bmatrix} X_{11} & X_{21} & X_{31} & | & Y_{11} & Y_{21} \\ X_{12} & X_{22} & X_{32} & | & Y_{12} = NA & Y_{22} = NA \\ \vdots & \vdots & \vdots & | & \vdots & \vdots \\ X_{1t} & X_{2t} & X_{3t} & | & Y_{1t} & Y_{2t} \\ \vdots & \vdots & \vdots & | & \vdots & \vdots \\ X_{1T} & X_{2T} & X_{3T} & | & Y_{1T} & Y_{2T} \end{bmatrix} \tag{6}$$

from which we will just observe the second half, the observations and just for those t when they are available. Even if not observable on real life, the first half will be useful for simulation controlling and understanding of what is happening later.

[1]We apologize for not respecting the consensus of writing the sub-indexes as $A_{row,column}$. We are using rows as t and columns as coordinates following a data.frame wide format that we consider far more reader-friendly even if less mathematically consensual.

F. Generated Data Visualization

For a far more visual understanding of the model and data produced we plot here the evolution of the state and the evolution of the available observations.

The state evolves following the Lorenz chaotic but deterministic dynamic, but we still observe how the stochastic behavior provokes slight moves that spread through the dynamical leading to major changes on the state (as described on equation 1). The blue line shows the whole set of observations while the orange dots identify those states that will lead to an observation (each $B = 10$ steps). This plotting represent a latent state that we can monitor just because we are under a simulation context. On a real problem this is the target we are aiming without seeing it directly.

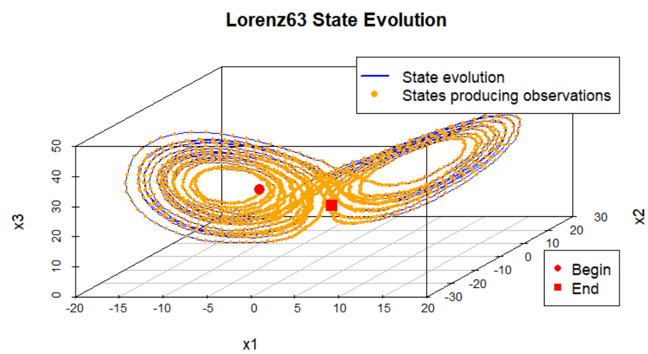

Fig. 1. Latent State Evolution.

On the other hand, the observations are the unique data we will be seeing on a real case. In this case, as expected, it collapses the three-dimensional chaotic behavior into a bi-dimensional (similarly chaotic) behavior following the rule settled at equation 2. The defy now is, from this observations, reconstruct the plausible latent state.

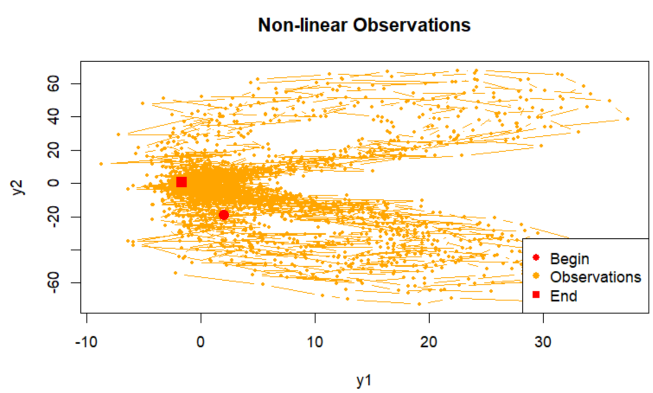

Fig. 2. Observation Evolution.

III. Method

Particle Filter or Sequential Importance Sampling Resampling

After this introduction and getting closer to the dynamical system defined and which kind of observations it yields we can move to the main procedure on this project, the Particle Filter or Sequential Importance Sampling Resampling.

Since it was deep covered on the course, we will not review the whole theory behind Importance Sampling or Kalman Filtering. Nevertheless, we consider relevant to review the key steps on Particle Filter mainly thinking on a future in which we return to this content and project and we will need a refresh on the main ideas of the process.

A. Proposal q(x) - Heuristic Prior

The first step is creating a prior on the state coordinates at time $t = 1$. This will be required for sampling the first particles. Later, will no longer be needed, since we will assume that the proposal on Importance sampling $q(X)$ and the prior on the coordinates $p(X)$ are equivalent (that is why we are speaking of them in-distinctively). Thus, we require a function such

$$q(x) \equiv p(x). \tag{7}$$

We need for that function that it assigns a relevant probability (non-null) to the true value of the state (on the contrary we are under a degenerated prior)... but we do not have those states [2]. Here we will combine an exploration of the observed data (table I) and the knowledge we have on how those observations are produced from those states (equation 2). Since

Statistic	$Y_{1,t}$	$Y_{2,t}$
Min.	-8.674	-72.590
Median	3.017	-4.296
Mean	6.327	-4.824
Max.	37.377	67.734

TABLE I
EXPLORATORY DATA ANALYSIS (EDA) ON $Y_{1,t}$ & $Y_{2,t}$.

$X_{2,t}$ is just associated with $Y_{1,t}$ it seems reasonable to focus on $Y_{1,t}$ for settling the prior on $X_{2,t}$. If we imagine a extreme case in which $X_{1,t} \approx 1$ then all the observation relevance come from $X_{2,t}$ meaning that $X_{1,t} \approx 1 => X_{2,t} \approx 10Y_{1,t}$. So it seems reasonable to create a prior that covers those values. For considering the negative symmetry (induced for any negative sign on $X_{1,t}$) we opted for a normal centered at 0 (thus we cover any smaller values of $X_{2,t}$ caused by bigger in absolute value $X_{1,t}$) with a standard deviation of $10max(|Y_{1,t})|$. A symmetrical reasoning was made for $X_{3,t}$ in association with $Y_{2,t}$

For $X_{1,t}$ we could not find other helpful reasoning. However, if we repeat the reasoning for each observation, we could inverse the roles, assuming $X_{2,t}$ (or $X_{3,t}$) being approximately 1 and, subsequently, the whole $Y_{1,t}$ (or $Y_{2,t}$) depending entirely on $X_{1,t}$. So we repeated here the most variable prior of the previous priors.

Formalized and numerically, this resulted on settling as priors[3]

$$q(X) \equiv p(X) = \begin{cases} X_1 \sim \mathcal{N}(\mu_{X_1} = 0, \sigma_{X_1} = 10max(Y_{1,t}, Y_{2,t}) = 670) \\ X_2 \sim \mathcal{N}(\mu_{X_2} = 0, \sigma_{X_2} = 10max(Y_{1,t}) = 370) \\ X_3 \sim \mathcal{N}(\mu_{X_3} = 0, \sigma_{X_3} = 10max(Y_{1,t}) = 670) \end{cases} \tag{8}$$

B. Initial Sampling

Once those priors are settled, we sample M random particles from them. In our project we settled $M = 2500$.

For each $m-th$ particle we obtain a coordinate on each of the state coordinates $(X^{[m]}_{1,t=1},\ X^{[m]}_{2,t=1},\ X^{[m]}_{3,t=1})$ (Figure 3, plots on the left)[4][5]. By combining those three coordinates $m-th$ particle per $m-th$ particle we can recreate the three-dimensional state $X^{[m]}_{t=1}$ (Figure 3, plot on the upper-right).

C. Weighting

Following the previous explanation, from each three-dimensional state $X^{[m]}_{t=1} = (X^{[m]}_{1,t=1},\ X^{[m]}_{2,t=1},\ X^{[m]}_{3,t=1})$ we can also obtain the estimated observation $Y^{[m]}_{t=1} = (Y^{[m]}_{1,t=1},\ Y^{[m]}_{2,t=1})$ just by applying the observation function (equation 2). This is what we obtain on the middle-right plot on the Figure 3.

In this case, we also have the true observation value $Y^{[*]}_{t=1} = (Y^{[*]}_{1,t=1},\ Y^{[*]}_{2,t=1})$. This allows us to compute the likelihood of that true observation $Y^{[*]}_{t=1}$ conditioned on the state per each particle $X^{[m]}_{t=1}$ and we will use it as a non-normalized weight $W^{[m]}$. Since we have two coordinates for the observation which conditioned on X_t are independent (from the structure on observation equation 2) the likelihood will be the product of the likelihood for each coordinate[6]:

$$\begin{aligned} W^{[m]}_{t=1} &= \mathrm{P}(Y^{[*]}_{t=1}|X^{[m]}_{t=1}) = \mathrm{P}(Y^{[*]}_{1,t=1}|X^{[m]}_{t=1}) * \mathrm{P}(Y^{[*]}_{2,t=1}|X^{[m]}_{t=1}) = \\ &= dnorm(x = Y^{[*]}_{1,t=1}|\mu = 0.1\ X^{[m]}_{1,t}X^{[m]}_{2,t}, \sigma = \sigma_u = 2) * \\ &\quad * dnorm(x = Y^{[*]}_{2,t=1}|\mu = 0.1\ X^{[m]}_{1,t}X^{[m]}_{3,t}, \sigma = \sigma_u = 2) \end{aligned} \tag{9}$$

[2]At least theoretically we do not have them. We really have them stored (since simulated data), but we must proceed as if we didn't.

[3]We are working under the empirical Bayesian framework and settling the priors based on the data. For this project, we do not have a real domain for stating priors without observing the data. However, we stated extremely variable priors (extremely uninformative) for declaring honest non-biased empirical bayesian priors.

[4]Since this is a simulated data example, we can plot and trace the real latent values. This is impossible on applied cases.

[5]Due to the enormous variability on the prior and the zoom for plotting just relevant areas on the sampling sequence, the initial sampling seems uniform rather than normal, we must clarify that this is just an optical illusion for zooming on almost the mean of a extremely platicurtic normal

[6]Visually, still observable that a higher likelihood is given to those particle points $Y^{[m]}_{t=1}$ closer to the actual observation $Y^{[*]}_{t=1}$ (and subsequently to it associated latent state $X^{[m]}_{t=1}$) and smaller to those further once considered the noise on the observation process (σ_u).

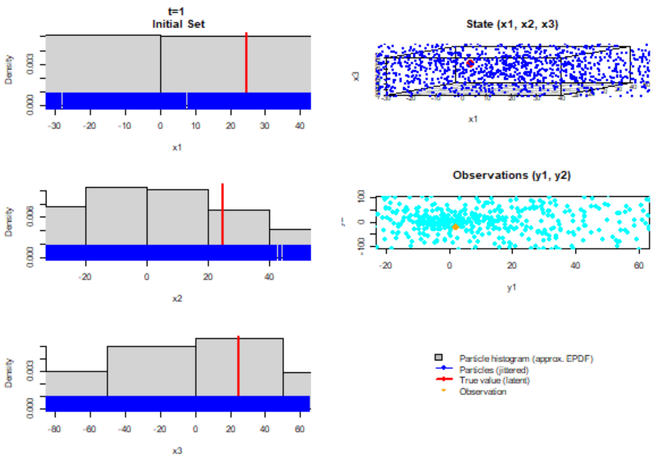

Fig. 3. Initial Sampling (ans subsequent state and observations).

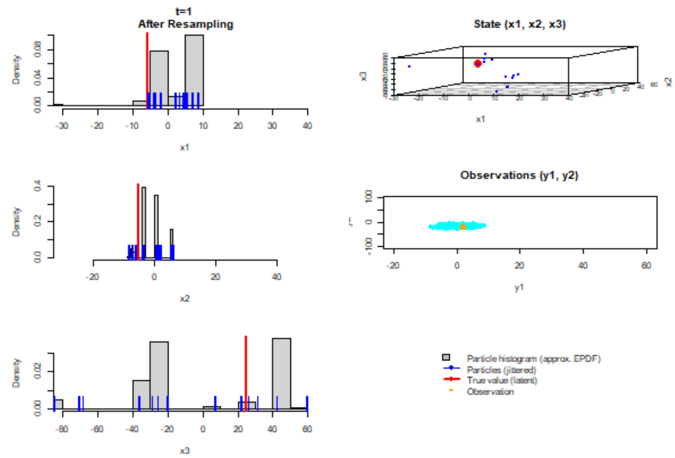

Fig. 4. t=1 After Resampling (and subsequent state and observations).

D. Resampling

Once we have an indicator of how likely and unlikely are our particles from being close to the latent value (assessed for how close they were with the resulting observation) we proceed to resample. This resample step enables good trajectories to deep their search while prevent bad trajectories from futilely occupying our computation time.

However, this sort of "decision" must be assessed somehow. A theoretically good MonteCarlo approach is treating it as a sampling procedure. We just need to assign to each $m-th$ particle an indicator of how well is working (how likely its trajectory is conditioned on the observed data) and resampling with substitution using that performance measure as weight.

The sole change is that we will work with the normalized weights $w_{t=1}^{[m]}$:

$$w_{t=1}^{[m]} = \frac{W_{t=1}^{[m]}}{\sum_{m=1}^{M} W_{t=1}^{[m]}} \tag{10}$$

And now we obtain a new set of M particles $X_{t=1}^{[m]'}$ resampled from the original ones but weighted by their normalized weight (normalized likelihood over the sum of likelihoods). These are the new set of particles that we will be working with, no longer those coming from the initial sampling. As we can see on (figure 4) the extreme variability from the initial set has been extremely reduced.

E. Propagation

Now we must transit from the actual t to $t+1$. This is no more than applying the transition equation (1) to each of our new set of particles and storing them as $X_{t+1}^{[m]}$ [7].

[7] We will not carry on the "′"sign since this may result on accumulating too much of them across resamplings. Furthermore, since we will not use previous particles mores we do not need to keep for them a distinctive notation.

On this transition, the latent space is producing a "latent observation" but we are not observing it until arriving to a multiple of B where observations are available again.[8]

From here the loop continues by actualizing $t \leftarrow t+1$ and going:

- If on $t+1$ we DO NOT have observation, just state propagation from t+1 to t+2.
- If on $t+1$ we DO have observation, we repeat the likelihood-weights-resampling starting from the likelihood weights.

IV. Results

The whole result of the particle filter can be verified on the animated video that we uploaded on GitHub[9].

A. Particle Filter Dynamic

Since the whole results can be traced, we will explain the main events we can see on them.

After the first resampling a major change is produced, the wide distribution collapses to much gathered distributions. After that, the propagation occurs for 9 times until $t=10$ is reached (when a new observation arrives). On that propagation, particles diverge, slightly covering more alternative combinations on the state and subsequent observation. E.g.: look at the subcloud of particles on the left for the observation at $t=10$ before resampling, figure 9. After the resampling on $t=10$ that tail of particles are cutted down (10). Those particles were associated with extreme values of the state coordinates that also disappear on the resampling step.

This procedure continues on and even if sometimes "incorrect" cloud of particles survive, they are eventually cleaned (filtered) and fastly converge to a cloud around true state parameters (around $t=100$).

[8] However, on the plots we are including it for having a nicely follow of the underneath dynamic. We mark them with different colors for remarking when we are on a true observation and when we are on a "latent observation" if you allow us the oxymoron

[9] Video link on GitHub

It's interesting to verify how even if the state begins to change its position, the system is able to follow it remarkably well.

B. RMSE and Precision Evolution

Finally, we have monitored the evolution of the Root Mean Squared Error (RMSE) and the precision (inverse of the variability) as long as the Particle Filter run.

For the RMSE on the state (X_t) (figure 5) we perceive that it clearly decay on the beginning before fast stabilizing on lower levels on which a permanent latency is not avoidable. $X_{3,t}$ is the coordinate with smallest error level while $X_{2,t}$ has an increased level.

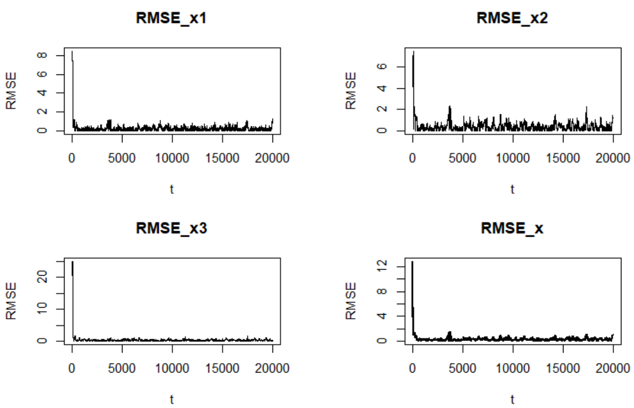

Fig. 5. RMSE on $X_t = (X_{1,t},\ X_{2,t},\ X_{3,t})$ along t.

For the observation, the noise is bigger, coherent since the noise on the observation step is also higher and given that from observation to observation we have a set of B steps in which there is untraced propagation (making that when it appears the error is bigger). The decreasing patter is more clearly visible for $Y_{2,t}$ (figure 6) while for $Y_{1,t}$ the peaks still reach the level of the first error.

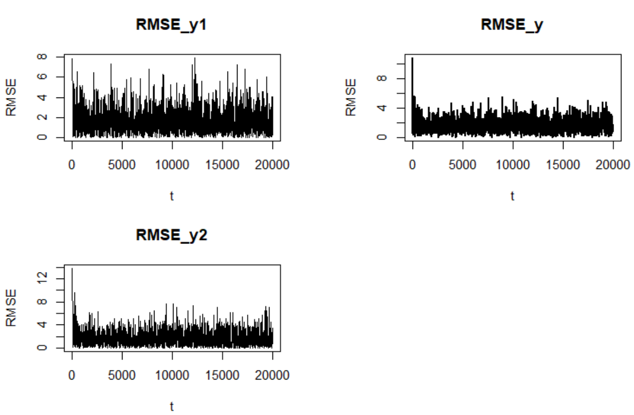

Fig. 6. RMSE on $Y_t = (Y_{1,t},\ Y_{2,t},\ Y_{3,t})$ along t.

On the standard deviation domain, the fast reduction is even clearer, and, as we previously saw, after the first resampling, it shrinks to smaller levels. In this case, both are quite small (figure 7 and 8)

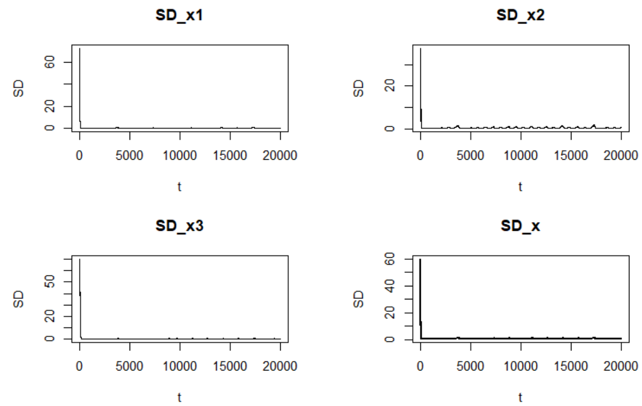

Fig. 7. Standard Deviation on $X_t = (X_{1,t},\ X_{2,t},\ X_{3,t})$ along t.

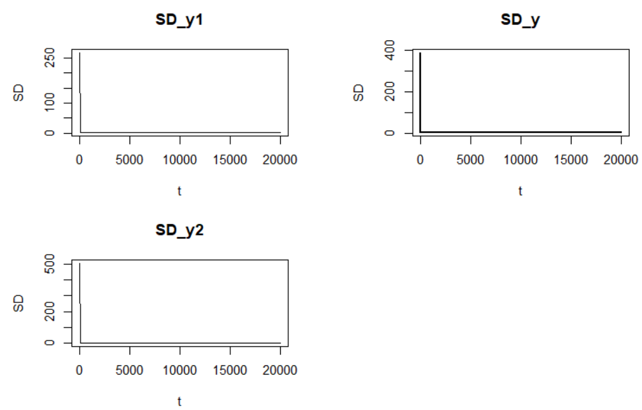

Fig. 8. Standard Deviation on $Y_t = (Y_{1,t},\ Y_{2,t})$ along t.

C. Summary of Results

With this results at hand, we can state that the Particle Filter result is satisfactory, after a small warm-up it can trace acceptably well the latent state even on the complex and slightly noisy non-linear dynamic and non-linear observation procedure.

V. Difficulties

While performing this project, we faced the degenerated sample likelihood problem. Sometimes, upon rolling the Particle Filter, the particles moves consistently away from the true state values and arrive to regions whose likelihood is numerically 0. Breaking the loop.

For passing this problem we find a two side solution:

- Decreasing B. Thus obtaining more frequent actualization of the true observations and leading to a better resampling (bad trajectories are sooner cut down and replaced by other more promising). However we know that this solution may not be possible when on an applied context

(the frequency we receive the observations are usually a given, not a decision).
- Increasing M. Launching more particles initially an which each resampling allows to cover more parametric space and even protect from degeneration as are making less probable not getting the best particles on any new resample particle (which is more likely with few particles). This seems the sole solution if the observation availability B is not modifiable. The drawback is that it increases the computational cost and memory used by tracing all the particles.

VI. Limitations

As limitations on this project we must state the following ones:

- We work under the assumption that we know clearly (even if stochastic) the state transition equation and the observation equation. However, it is likely that we do not know those models that well or that they depend on certain Θ parameters that we should estimate. This leads to the option of estimating those parameters first by Maximum Likelihood or by Expectation Maximitation and then run the Particle Filtering or more complex solutions for both estimating altogether (at the cost of increased risk of reaching explosive variability scenarios).
- As we stated before, we launched a number of particles enough for preventing the degenerated particle likelihood problem. However, we have not a clear prior rule of how many particles we need to launch. Finding a non-arbitrary rule may be quite interesting.
- A final interesting experiment that we did not arrive to do (and we have already covered too much space) is changing the variability on the transition equation (σ) and on the observation mechanism (σ_u) and verify how the filter behaves under more noisy circumstances. Also, for checking the effect of Particle Filter on Lorenz63 model, we may set different levels of noise for each coordinate and perform a sensibility analysis on what coordinates the increased noise harms more the filter performance.

VII. Conclusion

The Particle Filter has showed as a powerful tool for dealing with complex changing systems in which we are interested on a latent state. At least, under the circumstances of this experiment (knowledge of the procedure for transition and observation and small variability on both of them). Further simulation must be done for assessing the method behavior under those circumstances.

Nevertheless, these results encourage to have good prior expectations on the ability of the method for being capable of dealing with them. Moreover, knowing that there exist alternatives and extensions (like Adaptative Particle Filter or Multiple Importance Sampling) that can empower this basic method for reasonably overpass those defies.

Importance sampling (and its sequential resampling extension) has proven itself as an ”important” technique as well as powerful, which motivates and justifies having studied this course and performing this final project for truly understand it. Based on this understanding, proposals for further applications on solving complex state-space like structures are now possible for us.

VIII. Appendix

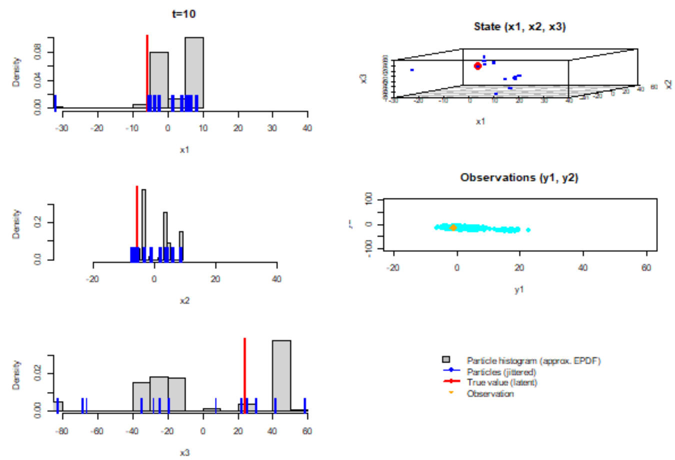

Fig. 9. t=10 Before Resampling (and subsequent state and observations).

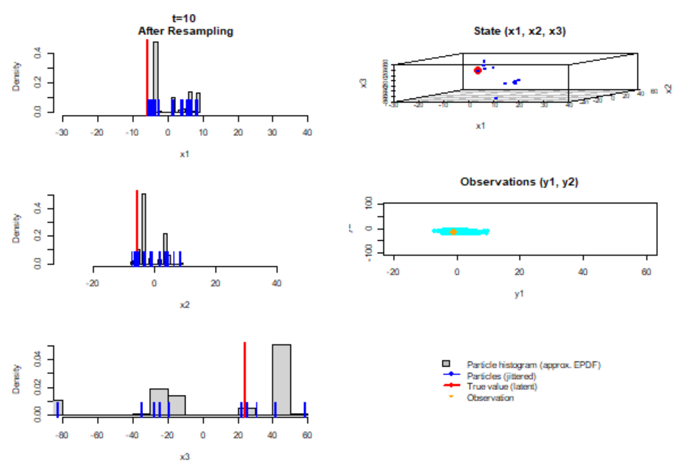

Fig. 10. t=10 After Resampling (and subsequent state and observations).