MS-No.: egusphere-2022-1362

Version: Revision

Title: Data-driven methods to estimate the committor function in conceptual ocean models

Author(s): Valérian Jacques-Dumas and co-authors

Point-by-point reply to reviewer #1

March 20, 2023

We thank the reviewer for their careful reading and for the useful comments and will adapt the manuscript accordingly. Below is a point-by-point reply with the referee's comments in bold font, our reply in italic font and the changes in manuscript in normal font.

1. **- p10: Equation (21) is unclear to me: u and v are time-series. The inner product $<,>$ is, by context, to be interpreted over the state-space and not time? If so, the index i on the rhs is the temporal index. Why is there a sum over the time index? Is this a time-integral? If so, the rhs is a vector and the lhs a scalar? Fortunately, it seems that equation (21) is not used anywhere later on, but arguably the notation in the section could be clarified.**

   We agree that the notation was confusing and we changed it (lines 261-267*). Here, u and v are functions from the state-space to $\mathbb{R}$ and the inner product $<,>$ is defined as an integral over state-space. Equation (21) corresponds to a Monte-Carlo estimate of this inner product using the available time series.*

   We will add more detail about this inner product and the steps that lead to Equation (21) on page 10.

2. **- p22: The explanation around lines 555 to 565 about 'aborted transitions' is not very satisfactory. In particular, the amount of 'aborted trajectories' is quantified by the committor itself by definition. There cannot be a large fraction of trajectories that reach q=0.5 and then abort. Instead, the fraction of trajectories reaching q=0.5 but not transitioning is exactly 0.5, and the same is true for any other value of the committor (for example, 10% of all trajectories reaching q=0.9 eventually abort). It is therefore unclear how one model can show more aborted transitions than another, or what an aborted transition even is. Maybe I am misunderstanding the intention of this paragraph.**

   We thank the reviewer about this point. Considering the definition of the committor, "aborted transition" was indeed poor phrasing. The intention of this paragraph was to explain why the logarithm score is decreasing for RC as N_T increases while the difference score increases, and to show that the trend of the logarithm score may be misleading when comparing two committor estimates. We also attempted to explain why one score decreases while the other increases. In the double-gyre model, the noise has a larger effect than in the AMOC box model and trajectories take longer to reach either an on or off-state. This causes the average logarithm score of the Monte-Carlo estimate of the committor to decrease in this model, as the trajectory explores larger areas of the phase space and the committor oscillates before reaching 0 *or* 1*.*

   On page 23, we will rewrite this paragraph, remove the expression "aborted transition" and provide a better explanation of the difference between both scores.