While this manuscript addresses a highly relevant and important topic—adjoint-free data assimilation for land surface model parameter estimation using atmospheric CO_2 concentrations, there is a fundamental conceptual misunderstanding regarding the data assimilation framework employed. The authors repeatedly describe their method as a "4DEnVar" approach. However, after careful review of the methodology and experimental design, it is clear that the implemented framework aligns more closely with a 3DEnVar method rather than a true 4DEnVar. I will provide more specific comments below detailing the evidence for this classification error and offering suggestions for how to appropriately revise the manuscript.

1. The authors incorrectly label their method as "4DEnVar." In classical data assimilation terminology, the key distinction between 3D and 4D variational methods lies in the incorporation of assimilation windows. A three-dimensional variational (3DVar or 3DEnVar) method assimilates observations as a function of space only, without explicitly considering the time evolution of the model states. In contrast, a four-dimensional variational (4DVar or 4DEnVar) method introduces a temporal dimension by defining an assimilation window, allowing the model to evolve dynamically and best fit the observations distributed across time within that assimilation window. In this manuscript, there is no explicit mention of an assimilation window, nor is there any evidence that the model trajectory evolves and interacts with observations at multiple times during a window. Instead, observations appear to be treated statically, consistent with a 3DEnVar framework. Therefore, the method used in this study should be accurately referred to as 3DEnVar, not 4DEnVar. The manuscript must be revised to correct this misclassification throughout, including the title, abstract, methods, results, and discussion sections.

2. The description at L56–58 defines general variational assimilation, not 4DVar specifically. 4DVar uniquely involves an assimilation window and time-evolving model trajectory. Please correct this definition.

3. Data assimilation can generally be categorized into full-field assimilation and anomaly assimilation. Full-field assimilation adjusts the model state towards observed absolute values, while anomaly assimilation only incorporates the observed anomalies. The authors must clearly specify which approach is used. If full-field assimilation is applied, the impact on climatology should be explicitly evaluated and presented.

4. The cost function shown in Equation (5) corresponds to the standard 3DVar formulation, as it lacks any temporal dimension or model trajectory integration. A true 4DVar cost function should involve the evolution of the model state over time:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}\sum_{i=0}^{n}(H_i M_{t_0 \to t_i} x - y_{obs}^i)^T R_i^{-1}(H_i M_{t_0 \to t_i} x - y_{obs}^i)$$

Please revise both the equation and the method name accordingly.

5. Lines 310-315: The R matrix represents the observation error covariance (not both the model/observation error) and should account for spatial heterogeneity in observation uncertainty. Setting all diagonal elements to 0.01 ppm is overly simplistic. Even in a simple setup, R could be statistically computed based on the variance of the observation. Please consider a more realistic design or justify this simplification.

6. Lines 364-366: Please clearly define how RMSD is computed, and indicate whether systematic bias is removed prior to calculation.

7. The evaluation relies almost entirely on RMSD and MAD. Please consider providing additional diagnostic metrics, such as correlation coefficients between posterior and true fields, or skill scores, to offer a more complete picture of assimilation performance.

8. The differences in RMSD reductions between different methods and configurations are discussed, but no statistical tests are provided. Please include simple significance tests (e.g., paired t-tests) to assess whether the differences in RMSD reductions are statistically meaningful across stations.

9. The comparison between 4DEnVar and ε-4DVar is repeatedly emphasized, but ε-4DVar is not equivalent to standard 4DVar. Please emphasize earlier and more clearly

that the ε-4DVar results are only a rough approximation and that conclusions should not be generalized to comparisons with a full 4DVar system.

10. All evaluations are performed against the same synthetic dataset used for assimilation. For robustness, a portion of the synthetic observations should be withheld during assimilation and used for independent validation.