

Comment	Answer
Although most of my major concerns have been properly answered, I still have some comments to make about the use of MAE in some scenarios. I am aware of the difficulties of computing errors when ground truth is 0. However, there are options that I strongly recommend the authors to use to have fully interpretable evaluations. For example, the relative root mean squared error, which is L2-norm / Linf-norm, is a good alternative for your evaluations. I recommend applying this metric or others, first, to the evaluation of the heat flux. Secondly, for the metamodel. Although in the case of the metamodel the MAE can give some guidance to compare the performance of each method, I insist on the use of a relative quantity or giving reference values with appropriate units for interpretation. Finally, the selected metric should be applied to the thermomechanical values, for which I thought the reference was not zero from what I understood in the paper.	Thank you for your feedback. We understand your concern regarding the use of MAE and appreciate the suggestion to use the relative root mean squared error (RRSE). MAE and RMSE are both commonly used error metrics, with MAE providing a straightforward average error magnitude and RMSE giving more weight to larger errors. We chose MAE for evaluating the heat flux due to its simplicity and ease of interpretation, especially given the challenges with zero ground truth values. MAE offers a clear indication of average error magnitude without being disproportionately influenced by outliers. For the metamodels, while MAE can give some guidance to compare the performance of each method, the encoded features do not have direct physical meaning, making the use of relative quantities or reference values less straightforward. We believe MAE provides a robust measure to compare the performance in this context, but we will consider incorporating additional metrics like RMSE in future work to provide a more comprehensive evaluation.
Page 3, line 32/33: It reads " Physics Informed Neural Networksm". Please, correct the typo.	Fixed
Page 3, line 32/33: It reads "finite element simulationsThese findings suggest". There should be a space and a point between "simulations" and "These".	Fixed
Figure 1 is very illustrative, thank you for considering it.	
Section 3.2: The 12000 nodes correspond to the whole discretization of the clutch system(all the components) from my understanding. From text in line 8 of page 12, it seems like the entire clutch system is represented as a 256 x 256 matrix.	The text was adapted to improve the readability. Additionally, to clarify, the entire clutch system is represented as a 256 x 256 matrix in the context of our data encoding and analysis, but this is a transformation from the original FEM node structure for the purposes of applying PCA and metamodeling.
"The process of dimensional reduction and encoding is employed to represent the temperature distribution of the entire clutch system, represented as a 256x256 matrix, with a significantly reduced number of values." However, what is rescaled is each element (the picture is very illustrative and you say it later in the same paragraph clearly). I recommend authors correct the text for readers. Also, then, how many nodes are used to describe each element? This would be helpful to know the real order of magnitude of the rescaling. Finally, it could be interesting to justify why it is valid to rescale to a grid/mesh of 256x256. I suppose higher discretization is required for the FEM, but then the field is sufficiently regular to reduce the size, is that correct?.	Thank you for the feedback. We appreciate your insights and suggestions. To clarify, the process of dimensional reduction and encoding is used to transform the temperature distribution data for each element within the clutch system. Each element is initially represented by a high-resolution FEM mesh consisting of 11,963 nodes in total. The number of nodes is variable and results from the used element size of 0.0004m. As the geometry is kept constant in this case, we have a constant number of nodes. This data is rescaled to a 256x256 matrix, significantly reducing the number of values while preserving essential information. The choice of a 256x256 grid for rescaling is based on the observation that the temperature fields are sufficiently regular, allowing for effective representation with reduced resolution. This balance ensures that the critical features of the temperature distribution are maintained without the computational burden of higher discretization.
The answer of the authors to my question about using PCA over all types of equipment seems incomplete due to a problem with the pdf of the responses. I would appreciate knowing the full response in a future response letter if this happens.	
Reading again your manuscript carefully, and paying attention to the temperature patterns (Figure 5), one could see that patterns are sufficiently similar to apply a linear correlation, but sufficiently different so the method can define also the new, reduced, coordinates as a function of the type of element. Since you only need 3/2 PCA coordinates, maybe a representation could show this clustering property by type. However, it is possible that the clusters are not graphically apparent, and they are only learnable by nonlinear clustering analysis techniques that are not visually illustrative.	Thank you for highlighting the variability in temperature distribution based on the type of element. Indeed, this observation underscores the complexity of the system and the potential for further analysis. In future work, we will explore the possibility of clustering the elements based on their temperature patterns
Section 4: I strongly suggest the authors to include the formula of the MAPE. Is the average performed over all points of all components at once, or is it first applied to each component and then averaged for the total number of components and simulations? This series of combinations could continue, and thus the motivation for the formula of MAPE and a explanation on the application.	Essentially, it organizes the temperature data into a two-dimensional structure where each row represents a sample and each column represents a pixel in the temperature image (256x256). Then, it compares each pixel in the original temperature distribution with its corresponding pixel in the reconstructed distribution across all samples. Finally, it computes the MAPE value, which reflects the average percentage difference between the original and reconstructed temperature values.
In Figure 12, in the first row, it seems like PCA lost the gradient in the reconstruction of the temperature field. Is this critical for these components?	Thank you for your observation. You're correct that the issue with the reconstruction of the gradient in Figure 12 may be related to the transition points between elements, likely influenced by small pressure spikes on the edges. While this discrepancy is notable, it may not be critical for the overall analysis of these components.