Review 3

We would like to thank the reviewer for their efforts in improving our manuscript. We have edited the text in response to the reviewer's comments as follows.

1) The parameters adopted in the different experiments are introduced at page 4 "…in the case of the running example we used the following argument values: n = 7, m = 4, k = 2, dk = dv = 2, p = 16" but they are not described. Perhaps a table explaining them and later on a short motivation for their selection in the different cases could help the readers understand how they should select them.
We now specify in the text that the transformer configuration was obtained by means of the Optuna hyperparameter setting package, except for the autoregressive model which used the built-in pmdarima optimizer. All edited sections are in blue.

2) Perhaps adding convolutional neural networks with attention could be interesting to the study. [1]

[1] Temporal Convolutional Attention Neural Networks for Time Series Forecasting Yang Lin, Irena Koprinska, Mashud Rana Code: https://github.com/YangLIN1997/TCAN-IJCNN2021

We also included a computational comparison against the suggested contribution.