The Figures and analyses should be consistent with how someone would use our approach. We should tell the following story:

Suppose we are an analyst with the bike-sharing outfit, it is currently January, 2018, and we have just collected past data through the end of December, 2017. Our first objective is to conduct a retrospective analysis on the past data to determine if the predictive relationship P(Y|X) is stable over the entire range 2010-Jan through 2017-Dec. If so, we intend to fit our model over the data range 2010-Jan through 2016-Dec as the training data, determine the score-based MEWMA and component EWMA control limits from the data range 2017-Jan through 2017-Dec as the Phase-I data, and then prospectively monitor the data for concept drift beginning 2018-Jan as the Phase-II data. On the other hand, if the retrospective analysis indicates that P(Y|X) is not stable over the entire range 2010-Jan through 2017-Dec, then the objective is to do diagnostics to identify the reasons why and take appropriate actions based on that (e.g., modifying the model or fitting it to a narrower, more recent window of data).

Replace the existing figures with the following:

* A: Retrospective analysis over the over the entire range 2010-Jan through 2017-Dec. The training data will be this same full range, and the Phase-I data for the control limits should be the same as the training data. For this, do both the MEWMA and the component EWMAs. The drift should be more apparent in the component EWMAs. For the figure, show the MEWMA in the top row, and the component EWMAs in the bottom rows.
* B: Refinement of the retrospective analysis over the over the entire range 2010-Jan through 2017-Dec to tighten the control limits and make the concept drift more clear (this is standard procedure in retrospective SPC control charting). To make the drift more apparent in the MEWMA, do another retrospective analysis (MEWMA and component EWMAs) over the over the entire range 2010-Jan through 2017-Dec. But choose the training and Phase-I data to be a smaller window, based on what we see in A. Try this two different ways: B1 - Have the training data and Phase-I data the same smaller window, 2012-Jul through 2015-Jun. And B2 - have the training data the entire retrospective data and the Phase-I data 2012-Jul through 2015-Jun. For either option B-1 or B-2, then just add this figure as the right column of the Analysis A figure. We can make a comment that the approach is flexible, and for the retrospective diagnostic analysis, one can choose the training data to be either the entire retrospective data or only a smaller window. The Phase-I data should always be a smaller window, though, so the control limits are not too wide as in Analysis A.
* C: Based on A and/or B, modify the model as we currently do. Then re-do A (and perhaps B, or we might not need it if it looks stable) with the modified model. Make a single figure like the one for A and B, and contrast it with that figure.
* D: Assuming C looks stable with the modified model, fit our model over the data range 2010-Jan through 2016-Dec as the training data, determine the score-based MEWMA and component EWMA control limits from the data range 2017-Jan through 2017-Dec as the Phase-I data, and then prospectively monitor the data for concept drift beginning 2018-Jan.
* E (not sure if we need to do this): Perhaps repeat the Analysis D but for the original model, to show that the transformed model is more stable than the original model. But maybe we don't need to do this, because Analysis A and B already showed that the original model was not stable.

Other notes for modifying the bike sharing example:

* See my notes in the pdf
* Get rid of the EWMA on the error rate throughout all of the bike sharing examples
* Mention somewhere (maybe also in the paper) that in the above example, there is a one-year lag between the last training data and when we begin using the model in Phase-II. To have better recency, we can take the Phase-I data to be the last year of the training data.