Unsupervised Concept Drift Detection based on Parallel Activations of Neural Network

Supplementary material

1 Hyperparameter selection

The first experiment aimed to select appropriate hyperparameters of the proposed method. Of the five available hyperparameters only the two most critical were optimized: *alpha* and *threshold*. The remaining ones were fixed:

- the number of network outputs e was 12,
- the number of statistical test replications r was 12,
- the sample size s was 50.

A neural network with single hidden layer containing ten neurons and a *ReLU* activation function was used.

For the *alpha* parameter, 15 values from the range 0.03 to 0.2 were tested, and for the *threshold* parameter, ten values from the range 0.1 to 0.3. The operation of the method with the indicated configurations was tested for the streams with ten drifts. The result of this experiment should indicate the range of values of these two parameters for which the method effectively recognizes drifts. Since both examined parameters indicate sensitivity to changes, their relationship should be visible.

The results of the three drift detection error measures have been processed for simplified interpretation, as shown in Figure 1. The heat map's vertical axis describes the alpha parameter's values, and the horizontal axis describes the threshold parameter. The results for streams with gradual drifts are presented on the left, and for sudden drifts on the right side of a figure, the rows of sub-figures present various dimensionalities of the data.

The results from the three measures were normalized to the range 0-1. For visualization purposes, cases for which the errors were not possible to calculate (no detection by the method) were filled with the values of the highest calculated ones for each of their type. After such normalization, three types of errors were presented for each stream as a heat map, where subsequent color channels corresponded to subsequent measures. The red channel shows the D1 error, the green channel D2, and the blue channel R error. After such arrangement, the lowest errors in all three criteria will be marked by colors close to black.

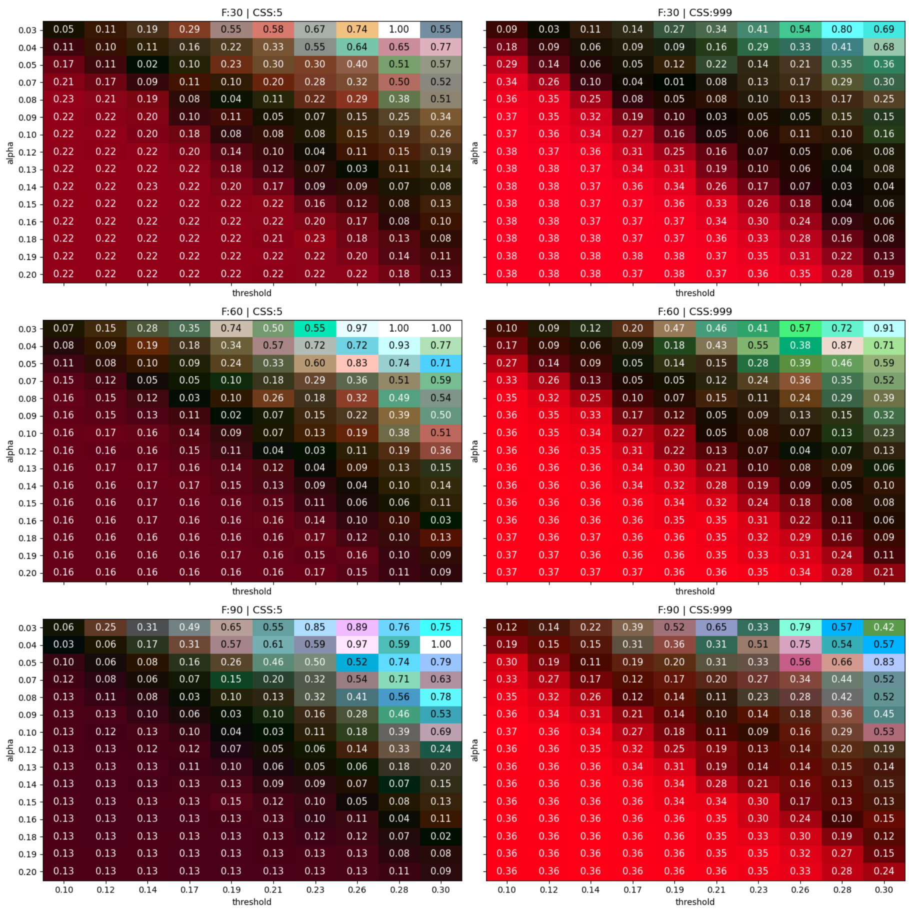

Figure 1: The combination of drift detection error measures, normalized to a range 0-1 in each of three measures, presented as a RGB image. The text labels are presenting results averaged across all three errors. Black cells indicate the lowest error across all channels.

Additionally, the text label in each cell of the heat map describes the average values of the three measures after normalization.

In the lower left corner of each subfigure, areas for a high alpha parameter and a low threshold parameter combinations are marked in red. With this configuration, the method will signal multiple redundant detections due to the sensitivity overdrive. Conversely, for low alpha and high threshold values, visible in the upper right corner of the subfigures and marked in green, blue or teal are configurations for which no drifts were signaled or there were very few detections. The best configuration of the method will be parameter values lying between these two extremes. It is worth noting that the method's effectiveness is highly dependent on the selection of these two parameters, and the lack of precise selection may result in excessive detection or failure to recognize drifts.

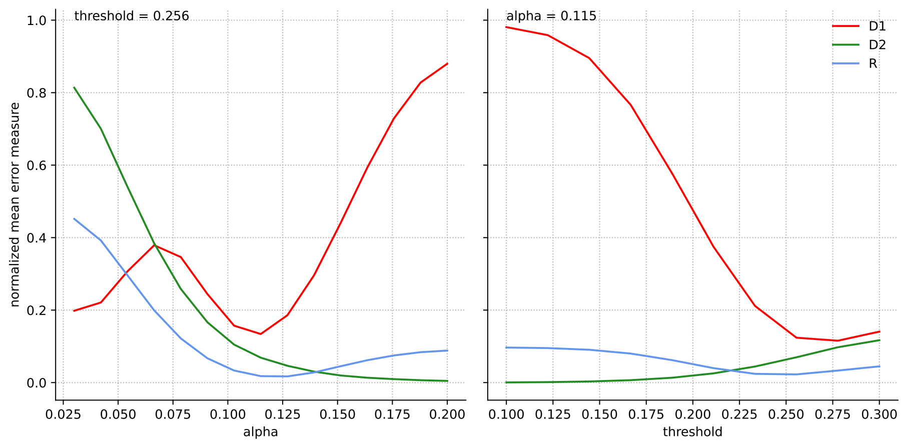

Figure 2: Normalized values of drift detection error measures, for fixed values of threshold and alpha hyperparameters, smoothed with gaussian filter. The regions with low error values (for all three types of measures) are showing the best method calibration.

Additionally, to present a direct interpretation of results, Figure 2 was prepared to show how normalized errors change for a fixed value of *alpha* and *threshold* parameters. The specific colors present the three types of errors, according to color channels in figure 1 – red curve shows D1, green D2, and blue R measure.

Ultimately, the following parameter combinations were selected for subsequent experiments: for gradual drifts – alpha equal to 0.13 and threshold equal to 0.26; for sudden drifts – alpha equal to 0.07 and threshold 0.19.

2 Method comparison

This experiment compared the performance of the proposed approach with reference methods. The results are shown in Figures 3-6 for 3, 5, 10 and 15 drifts, respectively. The columns of each Figure indicate the results for gradual (left) and sudden drift (right), and the rows presents results for different numbers of features. On the horizontal axis of each graph, successive chunks of the data stream are visible, while the central moments of the actual drift are marked with ticks and grid. Each detection is marked with a single point. For emphasis, the proposed approach is shown in red. The consecutive lines show the results from subsequent replications for a given detector.

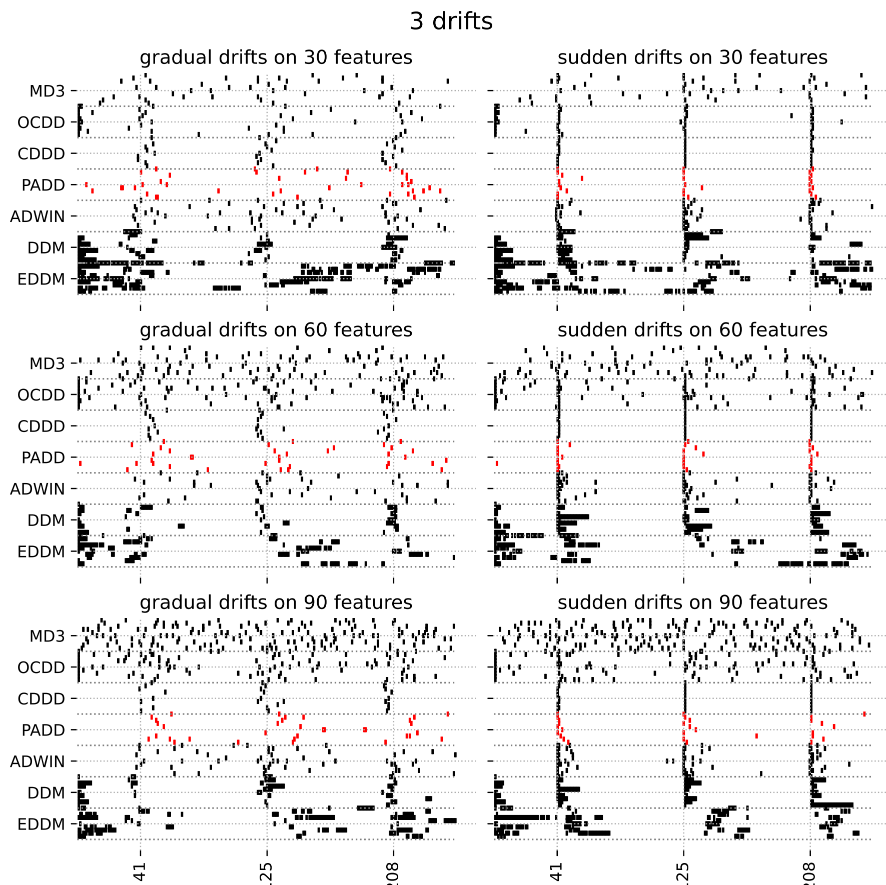

Figure 3: Detection moments of all evaluated methods for streams with 3 drifts. The real drifts are marked with ticks on horizontal axis and particular methods's detections in ten stream replications with black or red (in case of the proposed approach) points.

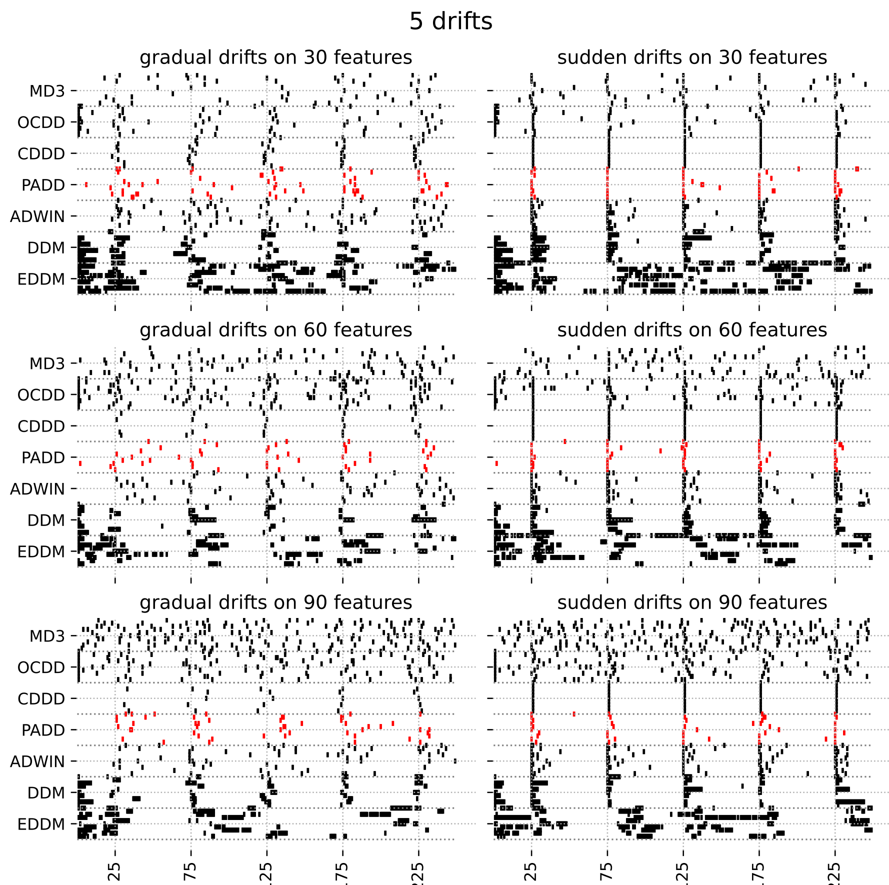

Figure 4: Detection moments of all evaluated methods for streams with 5 drifts. The real drifts are marked with ticks on horizontal axis and particular methods's detections in ten stream replications with black or red (in case of the proposed approach) points.

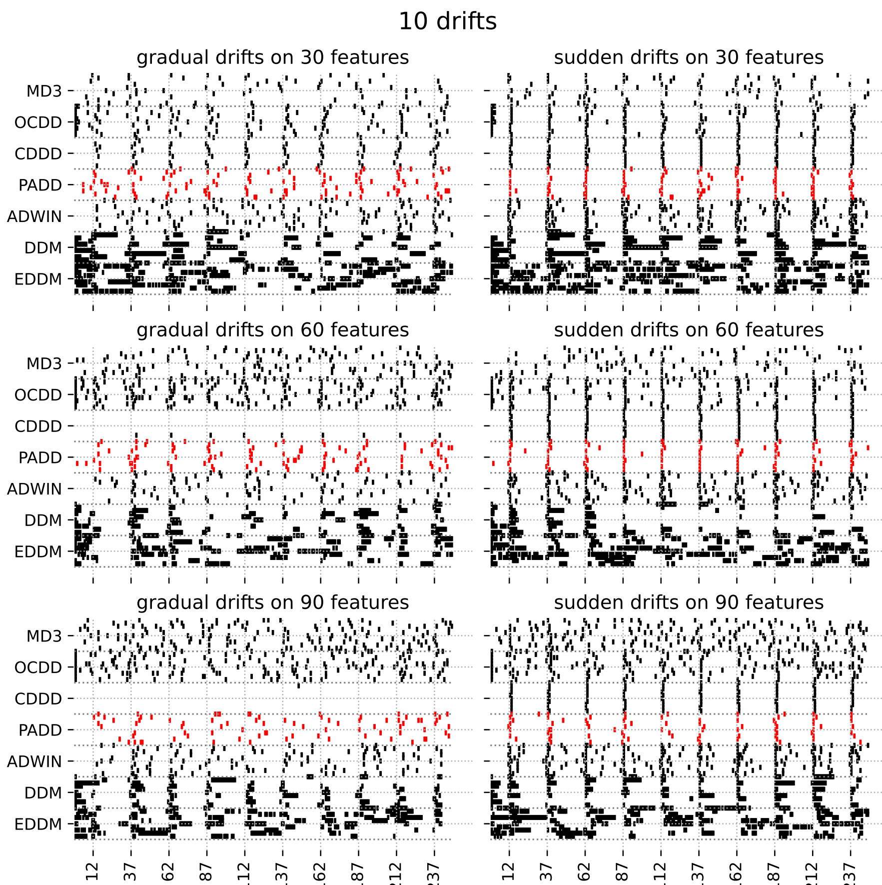

Figure 5: Detection moments of all evaluated methods for streams with 10 drifts. The real drifts are marked with ticks on horizontal axis and particular methods's detections in ten stream replications with black or red (in case of the proposed approach) points.

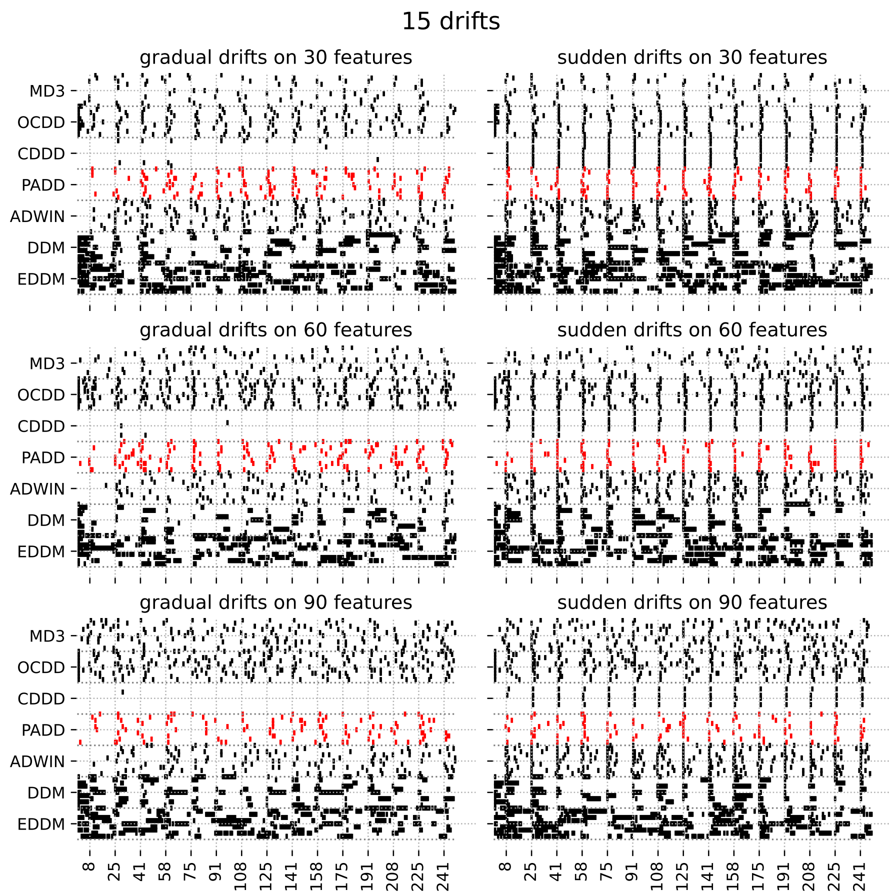

Figure 6: Detection moments of all evaluated methods for streams with 15 drifts. The real drifts are marked with ticks on horizontal axis and particular methods's detections in ten stream replications with black or red (in case of the proposed approach) points.