1 Supplementry Materials

1.1 Pseudo-code for OEC algorithm Algorithm 3 shows the update step of our OEC algorithm at time $k+1$.

Input : $\boldsymbol{x}_{k+1}$, Set of normal cluster boundaries C, State-tracker T, Consequent anomaly buffer B
Output: Updated C, S, B
Note : $|.|$ shows the cardinality of a set
$\gamma_1 = 0.99, \gamma_2 = 0.999$;
$C' = \{\}$; `/* Clusters to be updated */`
$\Omega = \{\}$; `/* Set of weights */`
Anomaly=true;
foreach g *in* C **do**
  $t = (\boldsymbol{x}_{k+1} - m_{k,g})^T S_{k,g}^{-1} (\boldsymbol{x}_{k+1} - m_{k,g})$;
  if $t < (\chi^2)_p^{-1}(\gamma_1)$ **then** *Anomaly*=false ;
  if $t > (\chi^2)_p^{-1}(\gamma_2)$ **then** continue ;
  $C \leftarrow C - g$; $C' \leftarrow C' \cup g$;
  $\Omega \leftarrow \Omega \cup e^{\frac{-t}{2}}$;
end
if *Anomaly* **then** $B \leftarrow B \cup \boldsymbol{x}_{k+1}$; **else** $B = \{\}$;
$\Omega = \Omega / \sum_{\omega \in \Omega} \Omega$;
foreach g *in* C' **do**
  Update g using (4.4) and (4.5) with corresponding weight in Ω;
end
Update T using (4.7) and (4.8);
$C \leftarrow C \cup C'$;
if $|B| > p+1$ **then**
  if *T is 2-separated with respect to all $g \in C$* **then**
    g_{new} = create new cluster using B;
    $C \leftarrow C \cup g_{new}$; $B = \{\}$;
  end
end

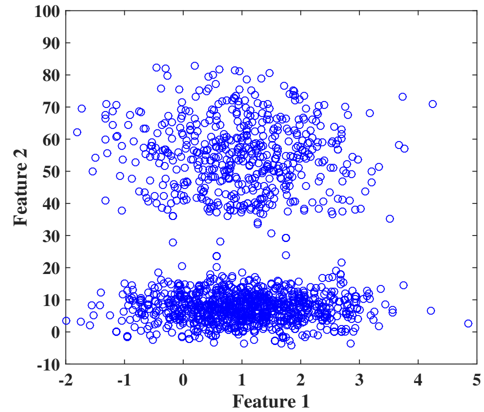
Algorithm 1: Online Elliptical Clustering (OEC) algorithm for input $\boldsymbol{x}_{k+1}$

1.2 Synthetic dataset generation The synthetic dataset $S1$ has been generated using two modes, M_1 and M_2, with different dynamic functions and input signals $(x_t, t = 2, 3, \ldots)$. M_1 follows the difference function in (1.1) and M_2 the difference function in (1.2).

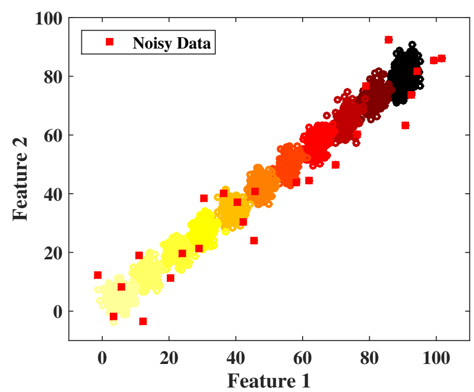
$$y_t - 1.801 y_{t-1} + 0.8187 y_{t-2} = 1.018 x_{t-1} \tag{1.1}$$

$$y_t - 1.5 y_{t-1} + 0.7 y_{t-2} = x_{t-1} + 0.5 x_{t-2} \tag{1.2}$$

A Gaussian with $\mu = 1$ and $\sigma = 1$ is used as input x_t to both modes. To build $S1$, we considered 4 mode changes between the two modes at uniform random intervals between 200 and 500 samples starting with M_1. Instead of a sudden shift between the dynamics of each mode, we change the individual parameters of one mode to the other mode in 5 equal steps, during which we generate 10 samples in each intermediate mode. Fig. 1(a) shows the scatter plot of the $S1$ dataset with the input (Feature 1 $= x$) and output (Feature 2 $= y$).



(a) S1: Synthetic Locally Linear Processes



(b) S2: Shifting Gaussian Distribution

Figure 1: The scatter plot of the synthetic datasets used in the evaluations.

The second synthetic dataset (shown in Fig. 1(b)), is generated by considering two modes, M_1 and M_2, with different normal distributions $N(\Sigma_1, \mu_1)$ and $N(\Sigma_2, \mu_2)$ and 9 intermediate modes. The parameter values of the modes M_1 and M_2 are: $\Sigma_1 = \begin{pmatrix} 3.8418 & -2.6474 \\ -2.6474 & 4.8478 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 1.5239 & -0.5390 \\ -0.5390 & 1.6467 \end{pmatrix}$, and $\mu_1 = (95, 75)$ and $\mu_2 = (5, 5)$. M_1 is the initial mode, and M_2 is the final mode. M_1 is transformed as follows. First, 500 samples $\{k = 1 \ldots 500\}$ are drawn from M_1. Sampling continues as each individual value in the covariance matrix and the mean is changed in 10 equal steps. After the first step, 200 samples $\{k = 501 \ldots 700\}$ are taken from the new normal distribution. After each new step 200 more samples are added to the dataset. The final step ends at mode M_2. The stars show 1% of the samples at each normal

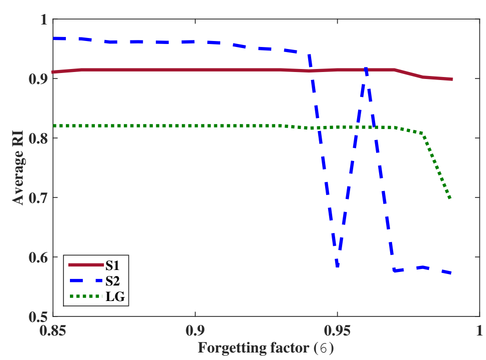
Table 1: List of the parameters of the OEC algorithm.

Parameter	Range	Suggested value	Effect/Purpose
γ_1	$[0.85,\ 0.99]$	0.99	Minimal/It is used mainly for anomaly detection.
γ_2	$[0.999,\ 1)$	0.999	Minimal/Protecting the algorithm against large-valued outliers, the effect increases on very noisy datasets.
n_s	$[10,\ 100]$	20	Minimal/A period where the incremental algorithm needs to stabilize, higher values are better when the sampling rate is high and the two consequent data samples are very close.
c-separation	$[1.5,\ 3]$	2	High/This parameter corresponds to overlap between clusters. For the suggested value, *at time of detection* a new cluster should be a separate cluster to the existing clusters. Lower values allows for more overlap.
λ	$[0.85\ 1)$	0.9	Medium/With higher forgetting (lower values of λ), OEC can detect smaller changes but it might result in higher than expected number of clusters.

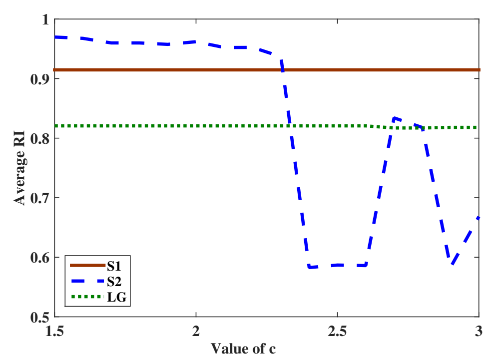
distribution, which are perturbed by uniform noise from $[-10,\ 10]$. A small level of noise is added to this dataset to investigate how the algorithms react to noise.

1.3 Sensitivity Analysis The proposed OEC algorithm has multiple input parameters. However, most of the input parameters can be selected using known concepts with OEC being robust to small variations in the parameters. Table 1 shows all the parameters of the model. We investigate the effects of the two most influential parameters, i.e., the value c for c-separation and λ, on the RI values calculated over the datasets. Fig. 2 shows the changes in the mean of RI values over the experiments for different values of these two parameters. There is no significant change in RI values obtained over any of the three datasets for λ in the range $[0.85\ 0.94)$. In $S2$ from $\lambda = 0.95$, in the other two datasets from $\lambda = 0.98$, we observe changes in the performance of the algorithm. We observe higher sensitivity in the $S2$ and LG datasets. This is due to the fact that small changes in these datasets and higher λ values reduce the tracking capability of the state-tracker. This situation is not observed for $S1$ as it has more prominent changes.

Dasgupta (2000) showed that 2-separation is obtained when we have two well separated Gaussian distributions. Therefore, we recommend 2-separation to be used in the OEC algorithm. However, changing this value can change the outcome of the algorithm. Increasing the 2-separation to 3-separation may lead the algorithm to lose its capacity to detect small changes in the algorithm. In fact, this occurs for dataset $S2$. Fig.2(b) shows how increasing this value reduces the average RI of OEC in this dataset.



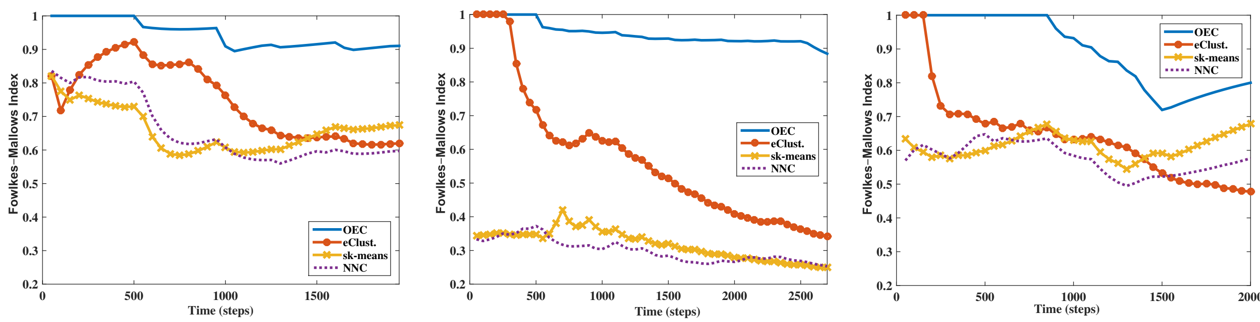
(a) Effects of changing λ



(b) Effects of changing c for the c-separation value

Figure 2: Effect of changes in different input parameters of the OEC algorithm.

1.4 Fowlkes-Mallows Index Similar to RI index, the results obtained for this index shows that OEC outperforms the other methods. The graphs are shown



(a) S1: Synthetic Locally Linear Processes (b) S2: Shifting Gaussian Distribution (c) The real-life dataset: LG dataset

Figure 3: The scatter plot of the synthetic datasets used in the evaluations.

in Fig. 3.