Karlsruhe Institute of Technology

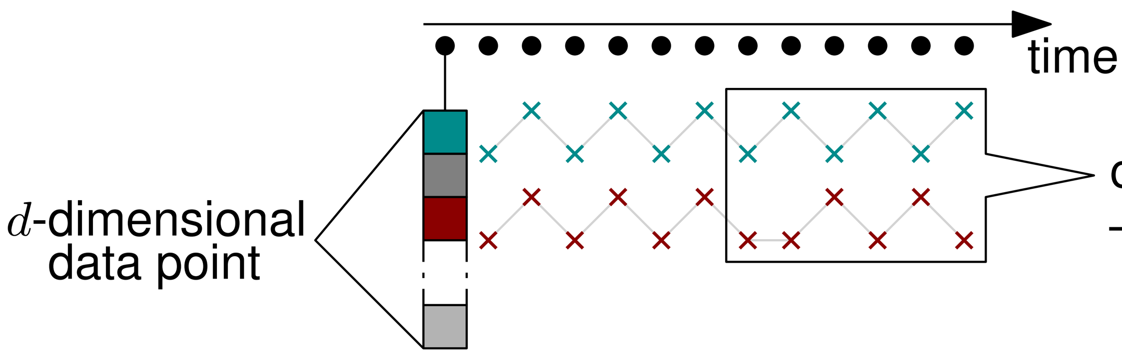
Paper doi.org/10.1007/s10618-023-00999-5
GitHub github.com/heymarco/AdaptiveBernsteinChangeDetector

Adaptive Bernstein Change Detector for High-Dimensional Data Streams

Marco Heyden (marco.heyden@kit.edu), Edouard Fouché, Vadim Arzamasov, Tanja Fenn, Florian Kalinke, and Klemens Böhm

Challenges in high-dimensional data streams

- curse of dimensionality
- changes can affect arbitrary subspaces
- huge number of possible subspaces ($2^d - 1$)



correlation changed between ■ and ■! (change subspace)
– but how severly? (change severity)

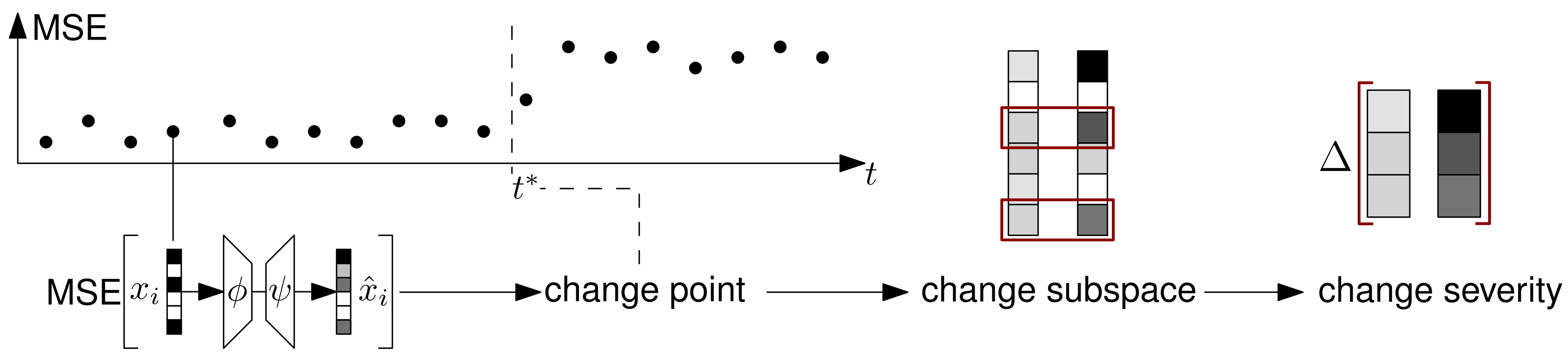
Application

- chemical plants are highly complex
- large number of deployed sensors
- changes in the sensors' readings can hint to undesired changes in the process

credit: KIT

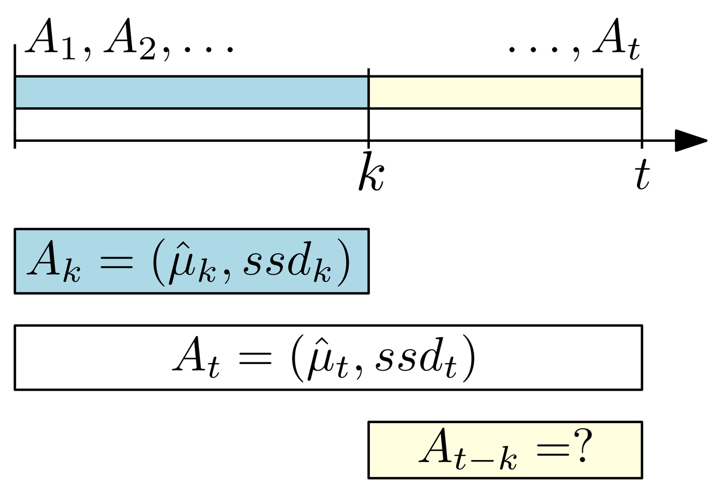
Our algorithm: ABCD

- encodes observations in fewer dimensions (e.g., using PCA, Kernel-PCA, or Autoencoders)
- monitors reconstruction error in adaptive window
- After change:
  - finds change subspace
  - computes severity of drift in subspaces



Adaptive window and stream aggregates

- stream aggregates (based on [1], [2]) allow evaluating multiple possible change points
- efficient variance tracking
- given two aggregates A_k and A_t containing sample mean $\hat{\mu}$ and sum of squared distances ssd with $k < t$, one can derive the aggregate for the time interval $(k, t]$



$$\hat{\mu}_{k+1,t} = \frac{1}{t-k}(t\hat{\mu}_{1,t} - k\hat{\mu}_{1,k})$$

$$ssd_{k+1,t} = ssd_{1,t} - ssd_{1,k} - \frac{k(t-k)}{t}(\hat{\mu}_{1,k} - \hat{\mu}_{k+1,t})^2$$

Change subspace

For each dimension j:

- compute change score in that dimension (the one based on Bernstein's inequality)
- if change score less than τ (external parameter)
  - add j to change subspace

Change severity

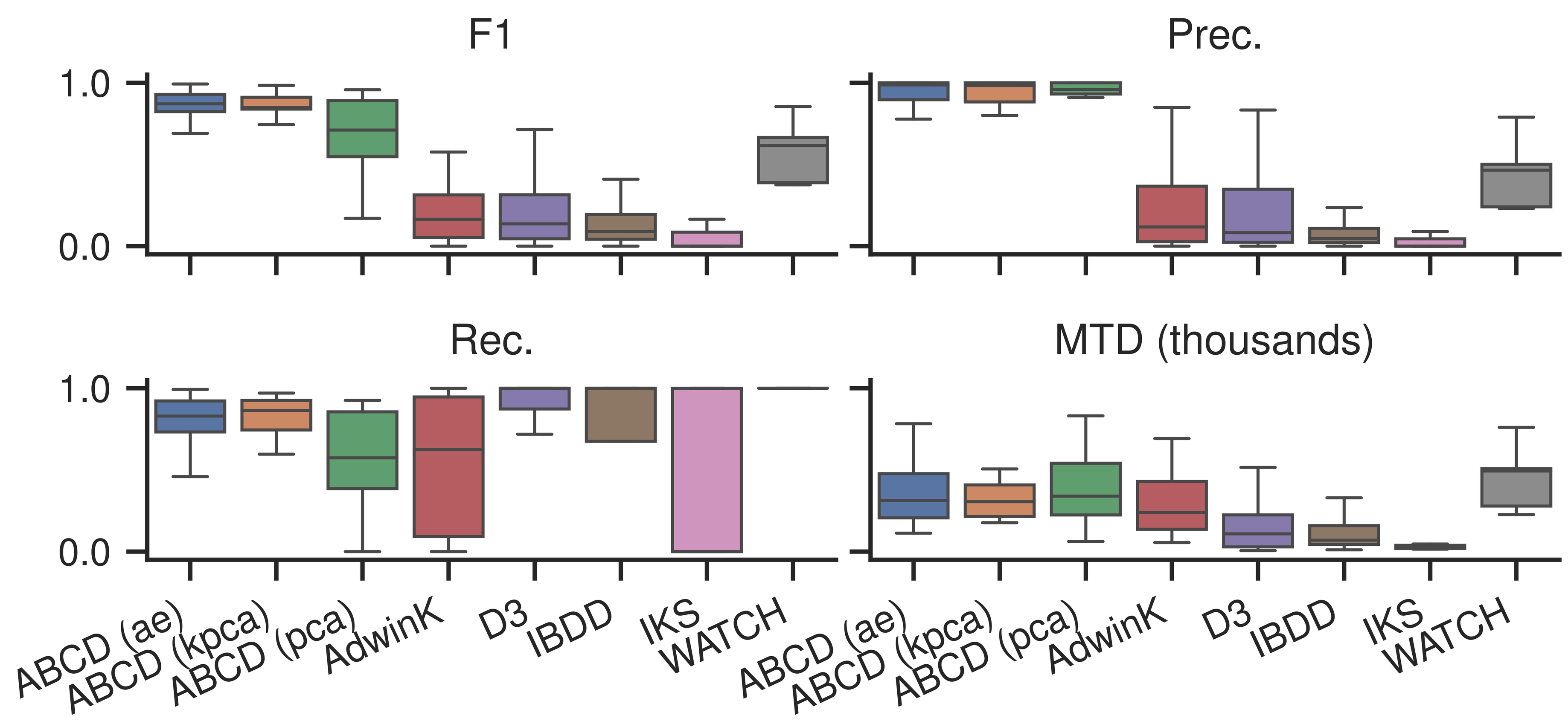
For each dimension j in change subspace:

- standard-normalize the average reconstruction loss $\hat{\mu}^{D^*}_{t^*+1,t}$ observed after the change point t^*

$$\Delta = \frac{\left|\hat{\mu}^{D^*}_{t^*+1,t} - \hat{\mu}^{D^*}_{1,t^*}\right|}{\sigma^{D^*}_{1,t^*}}$$

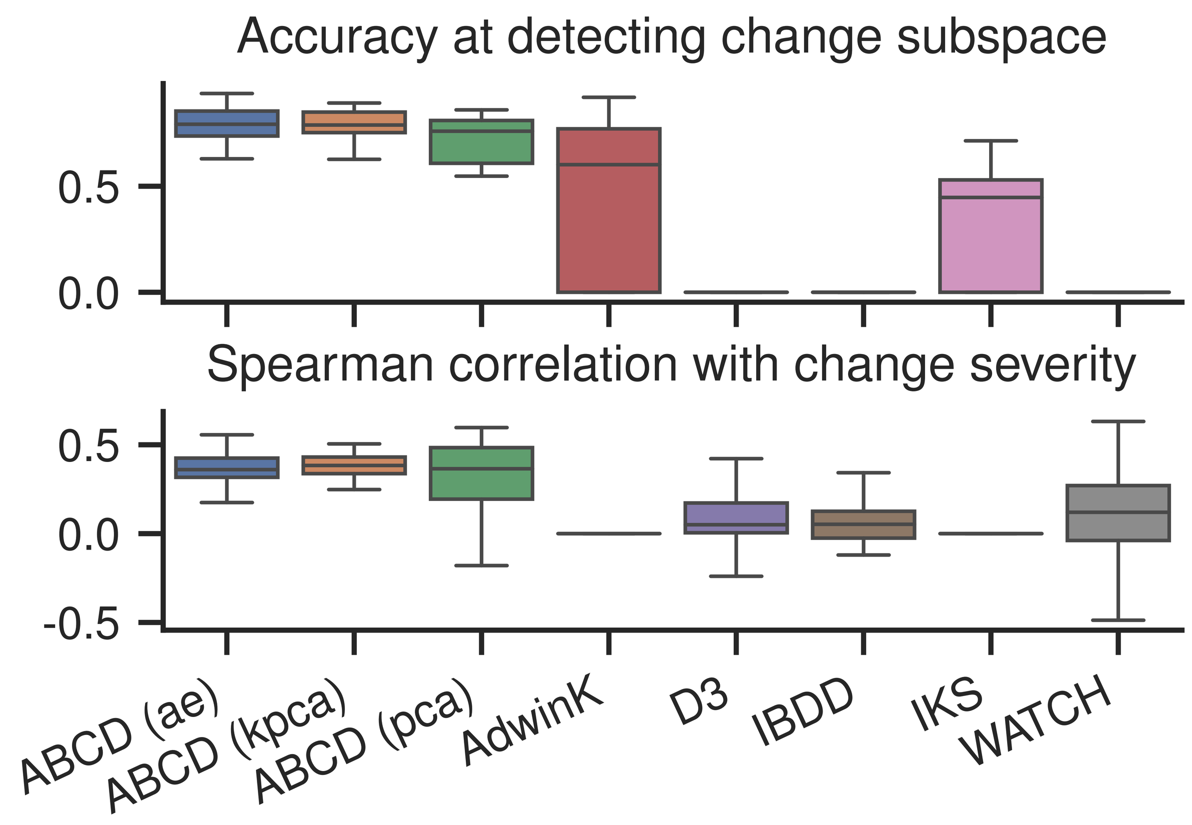
Detecting changes

- Boxes summarize performance of approaches for different hyperparameters and data sets (detailed results in the paper)
- Smaller box → approach is more robust to hyperparameter choice



Characterizing changes

- Top: accuracy at detecting the change subspace; boxes summarize different hyperparameters and data sets
- Bottom: Spearman correlation between severity computed by approach and ground truth



[1] B. P. Welford, "Note on a Method for Calculating Corrected Sums of Squares and Products," *Technometrics*, pp. 419–420, 1962. DOI: 10.1080/00401706.1962.10490022.

[2] T. F. Chan, G. H. Golub, and R. J. LeVeque, "Updating formulae and a pairwise algorithm for computing sample variances," in *COMPSTAT 1982 5th Symposium held at Toulouse 1982*, H. Caussinus, P. Ettinger, and R. Tomassone, Eds., Heidelberg: Physica-Verlag HD, 1982, pp. 30–41, ISBN: 978-3-642-51461-6.