Study of Common Online Changepoint Detection Algorithms

Zhymir Thompson [1, 2], Gurcan Comert [3], Joud N. Satme [1], Austin R.J. Downey [1], Jason D. Bakos [2]

1 Department of Mechanical Engineering
2 Department of Computing and Engineering
University of South Carolina, Columbia SC

3 Department of Engineering and Computer Science
Benedict College, Columbia SC

UNIVERSITY OF SOUTH CAROLINA

Outline

1) Background information
2) Experimental setup
   i. Dataset generation
   ii. Models
3) Model theory & experiment results
4) Conclusion

South Carolina

Background

University of South Carolina

Motivation

- A system or structure under normal conditions may suddenly shift to a state of shock
- Examples of unexpected shock:
  - A car driving down the street
  - A projectile impacting a structure
  - A structure in an earthquake
- It's important to know when a structure is in shock
  - Earlier detection means faster response
  - Responding too late could mean permanent structural damage
  - Responding too early can also be detrimental

Plane launching projectiles [1]

Collapsed building [2]

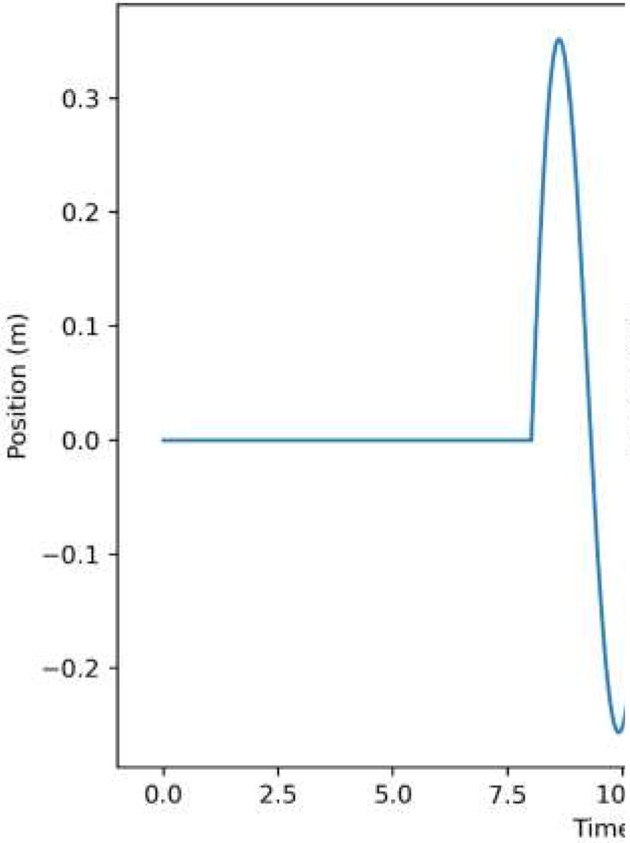
Damaged road [3]

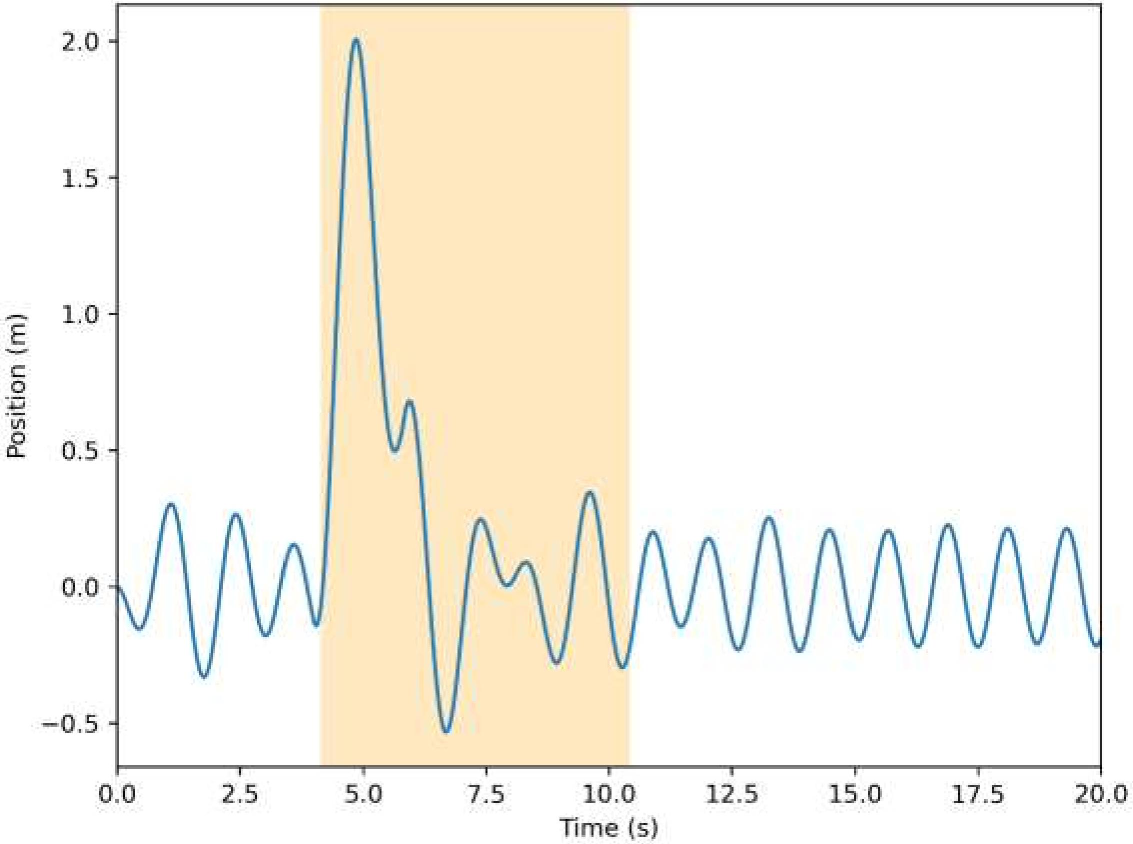
1. https://www.flickr.com/photos/sdasmarchives/4564390777/
2. Photo by samimibirfotografci : https://www.pexels.com/photo/rescue-team-at-collapsed-building-15533288/
3. Photo by Zülfü Demir📸: https://www.pexels.com/photo/damaged-asphalt-road-20518249/

UofSC South Carolina

What do we mean by shock?

- Shock is a sudden departure from the expected





South Carolina

Change Point Detection

- **Definition**: Change point Detection - "The identification of abrupt changes in the generative parameters of sequential data"
- Change point detection involves identifying when a point signifies a change in the features of a sequence of data
- We considered an event a 'change point' when a change is beyond our expectations of the assumed underlying distribution
- Examples:
  - Change in mean level
  - Shift in amplitude

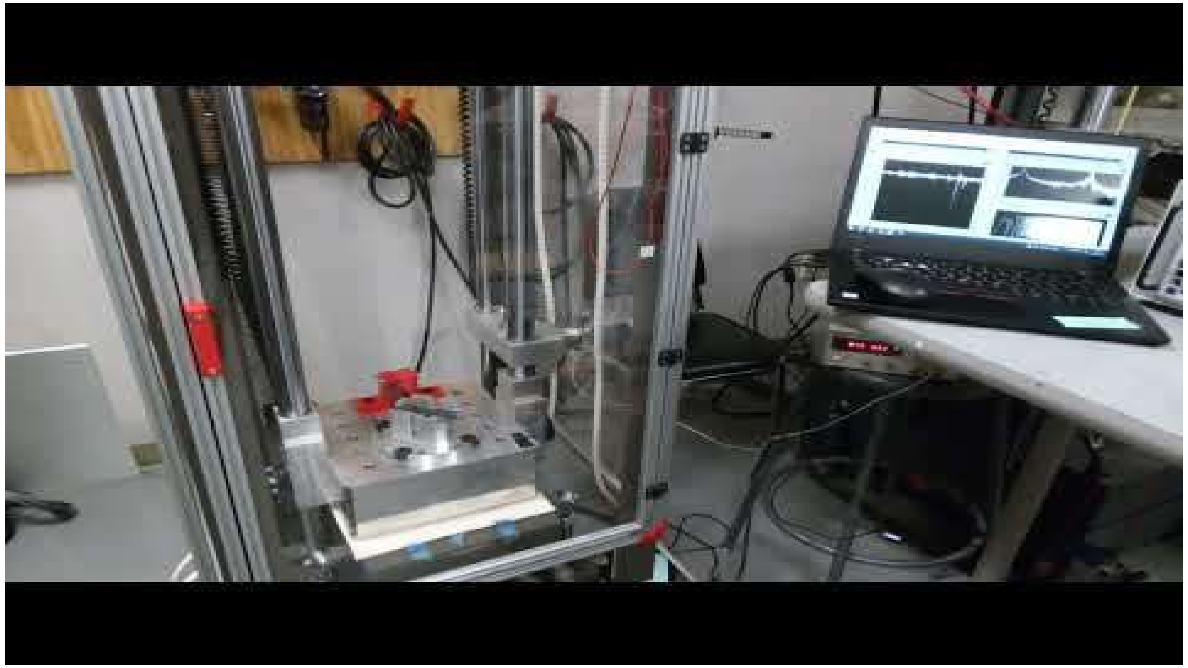
Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.

UofSC South Carolina

Experimental Setup

Purpose

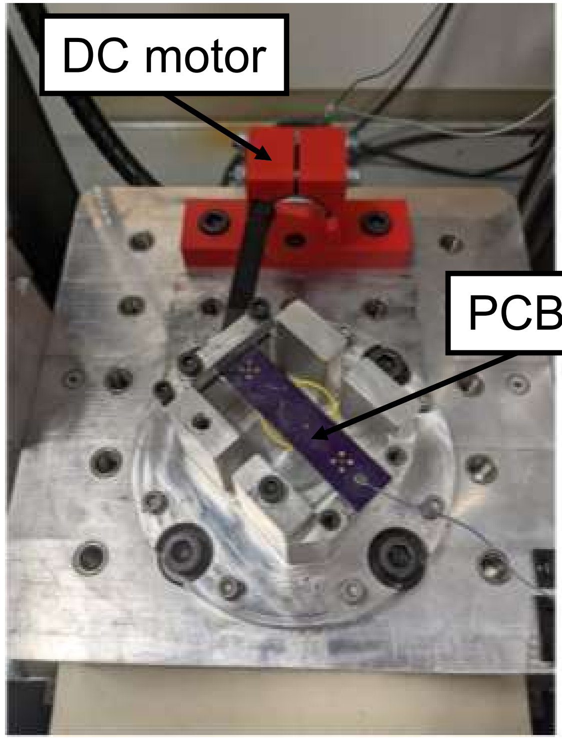
- Change point detection used in aberration detection
- Our goal was to measure likelihood of system being in a state of shock
  - Given a set of acceleration data, when do we think the response is abnormal
  - Abnormal response could mean a damage-inducing event for the electronic system

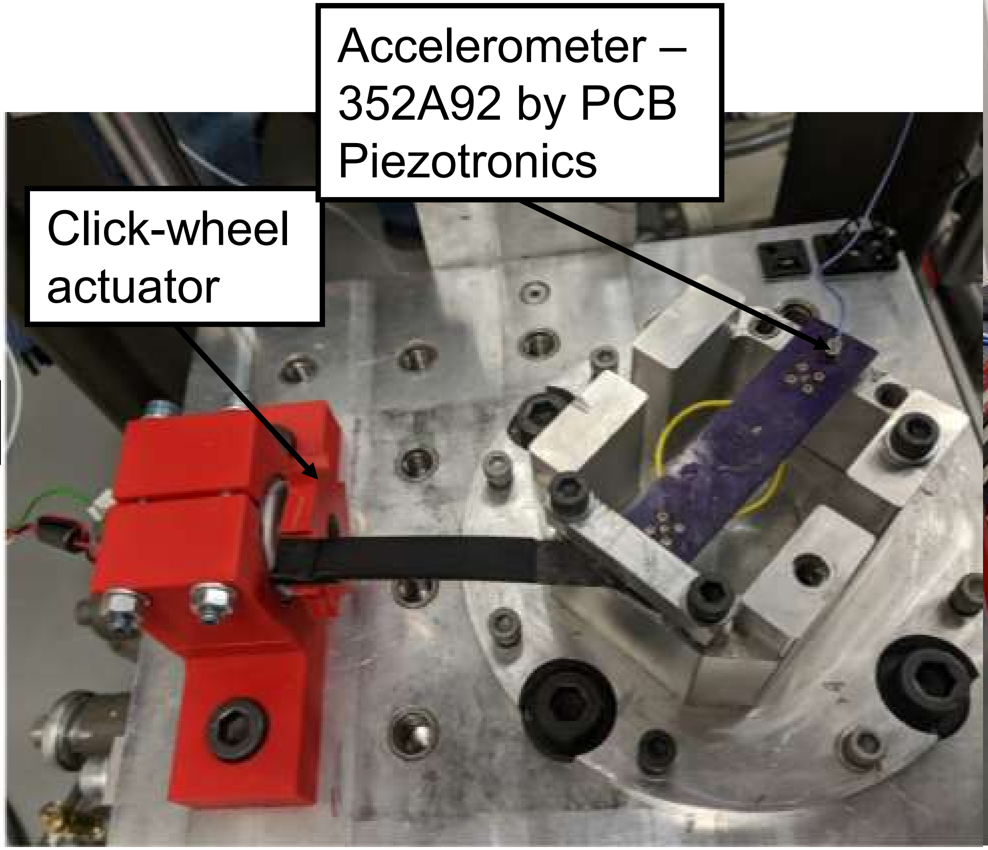


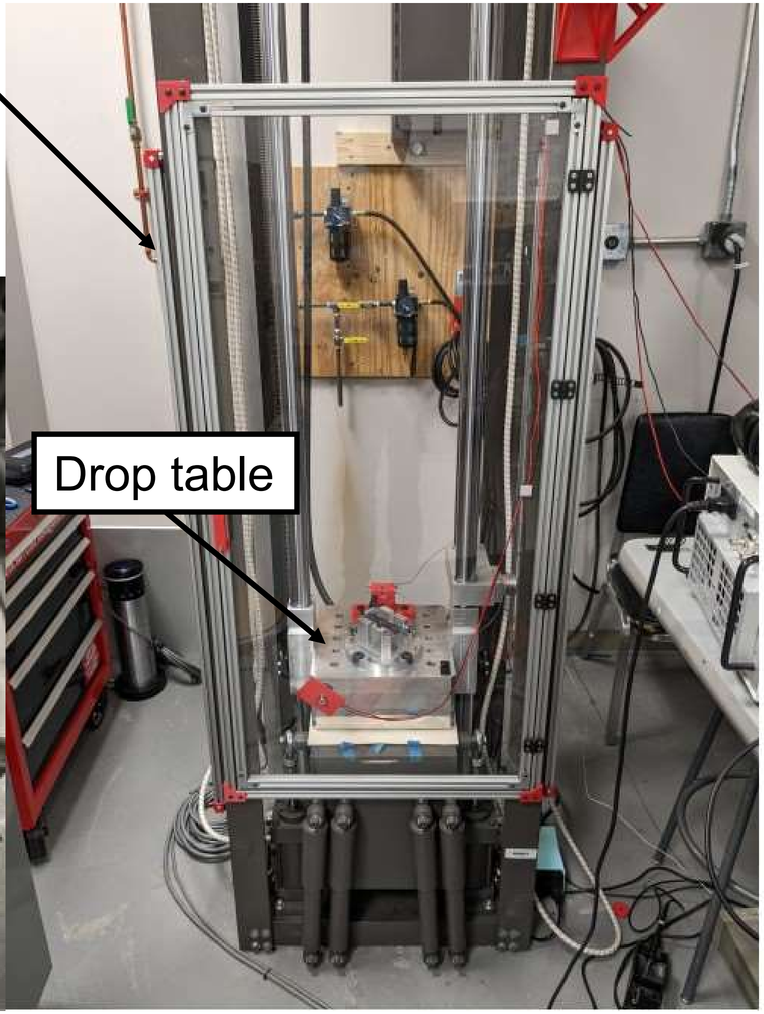
https://www.youtube.com/watch?v=kBaZF9kUQLQ

South Carolina

Experiment Setup







https://github.com/High-Rate-SHM-Working-Group/Dataset-7-forced-vibration-and-shock

Online vs Offline Detection

Online Detection

- Computes over successive points in a sequence
- Typically lower computation cost
- Faster
- Less accurate
- Change points identified as algorithm runs

Offline Detection

- Computes over complete sequence
- Typically more computationally expensive
- Slower
- More accurate
- Number of change points can be given or guessed

Online Algorithms

- **Bayesian Online Changepoint Detection (BOCPD)** [1]
  - Uses Bayesian statistics to model data over various run lengths
  - Decides if run length is likely to have reset since last change point
- **Expectation Maximization (EM)** [2]
  - Fits Gaussian models to known groups of data
  - Decides which model a given point best fits
- **Grey Systems Modeling (GM)** [3]
  - Transforms window of data two scalar describing behavior
  - Determines if behavior is sufficiently different
- **Cumulative Summation (CUSUM)** [4]
  - “Detects shifts in mean level of sequential data”
  - Determines if a significant shift has occurred in current mean from prior process mean

1. “Bayesian Online Changepoint Detection” Adams & MacKay 2007.
2. Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
3. J. L. Deng. 1989. Introduction to Grey system theory. J. Grey Syst. 1, 1 (1989), 1–24.
4. G. Comert, M. Rahman, M. Islam and M. Chowdhury, "Change Point Models for Real-Time Cyber Attack Detection in Connected Vehicle Environment," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12328-12342, Aug. 2022, doi: 10.1109/TITS.2021.3113675.

UofSC South Carolina

Online Algorithms

- **Bayesian Online Changepoint Detection (BOCPD)**
- **Expectation Maximization (EM)**
- **Grey Systems Modeling (GM)**
- **Cumulative Summation (CUSUM)**

algorithm	movement	compares	decision criterion
BOCPD	Point-by-point	Highest probability of current run length	Most likely run length has decreased
EM	Point-by-point	Point-of-interest to distribution of safe data and distribution of unsafe data	Probability that point is not safe exceeds boundary
GM	Sliding window	Window to reference window	Difference between sequence behaviors exceeds threshold
CUSUM	Point-by-point	Current mean estimate to process mean	Current mean estimate deviates from process mean by more than threshold

South Carolina

Online Algorithms

- **Bayesian Online Changepoint Detection (BOCPD)**
- **Expectation Maximization (EM)**
- **Grey Systems Modeling (GM)**
- **Cumulative Summation (CUSUM)**

algorithm	prediction function	algorithmic time complexity
BOCPD	$\underset{i-1}{\mathrm{argmax}}(p(r_{i-1}, X_{1:i-1})) > \underset{i}{\mathrm{argmax}}(p(r_i, X_{1:i}))$	O(mn), m<=n
EM	normpdf(x, μ_2, σ_2^2) > 0.01	O(mkn), m = set size, k = number of iterations
GM	$\epsilon_{ij} \leq 0.5 \cup \epsilon_{r,ij} \leq 0.5$, where ϵ & ϵ_r are thresholds	O(nw), w = window size
CUSUM	$(C_i^+ > H \cup C_i^- > H)$, where H is a threshold	O(n)

South Carolina

Hyperparameters vs Parameters

- Hyperparameters
  - Initial values chosen by the user
  - <u>Indirectly</u> affect the overall performance of the model
  - Constant throughout model performance
- Parameters
  - Variables modified by the model itself
  - Describe the model's assumptions of the underlying data
  - Updated by the model

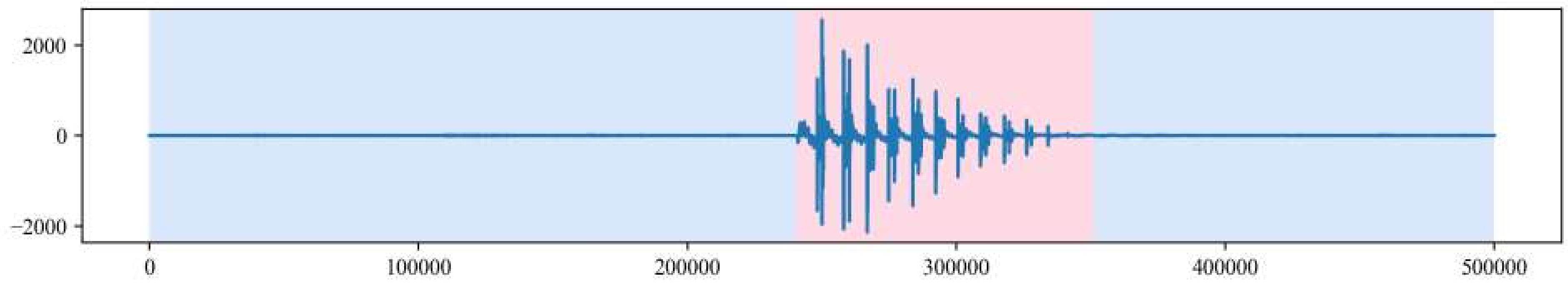
South Carolina

Model Theory and Results

South Carolina

Evaluation - Baseline

- Used an offline method as ground-truth for state of system
  - Python ruptures library
  - Binary segmentation
  - Rank cost function
- This estimate matches our expectations of where shock starts and ends



C. Truong, L. Oudre, N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Evaluation - Metrics

- Confusion Matrix
- Accuracy - Total proportion of correctly guessed values
- Precision – Proportion of positive predictions that were actually positive
- Recall – Proportion of positive values that were correctly predicted
- F1 score – Harmonic mean of precision and recall
- Earliest correct – Earliest correctly predicted positive detection

Confusion Matrix		
	Predicted Negative	Predicted Positive
Actually Negative	TN	FP
Actually Positive	FN	TP

Metric	Equation
Accuracy	$\frac{TP+TN}{(TN+FN+FP+TP)}$
Precision	$\frac{TP}{(TP+FP)}$
Recall	$\frac{TP}{(TP+FN)}$
F1 score	$\frac{TP}{TP+\frac{FN+FP}{2}}$

https://github.com/austindowney/Machine-Learning-for-Engineering-Problem-Solving
Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

Bayesian Online Changepoint Detection (BOCPD): Theory (pt. 1)

- **Definition**: run length – the length of a sequence between two change points
- Run length can only increase by 1 or reset to 0
  - Calculation of future run lengths can be performed recursively
- Computes and tracks distribution of run-length probabilities
- If distribution shifts, then a change point must have occurred

Hyperparameter	interpretation
μ	Prior mean of data
κ	Degrees of freedom
α	Half of degrees of freedom
β	Prior scaling value for the data
lambda	Number of time steps until change point anticipated

"Bayesian Online Changepoint Detection" Adams & MacKay 2007.

UofSC South Carolina

Bayesian Online Changepoint Detection (BOCPD): Theory (pt. 2)

- **Definition**: run length – the length of a sequence between two change points
- Run length can only increase by 1 or reset to 0
  - Calculation of future run lengths can be performed recursively
- Computes and tracks distribution of run-length probabilities
- If distribution shifts, then a change point must have occurred

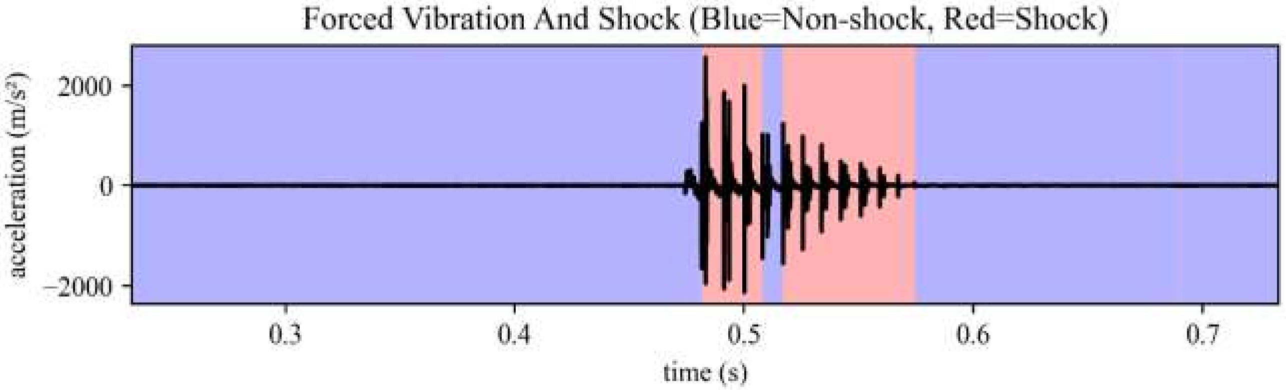
parameter	symbol	update equation
Mean – (prior mean)	μ	$[\mu_0, .., \mu_{t-1}] \rightarrow [\mu, \frac{\kappa_0 \cdot \mu_0 + x_i}{\kappa_0 + 1}, ..., \frac{\kappa_{t-1} \cdot \mu_{t-1} + x_i}{\kappa_{t-1} + 1}]$
Kappa – (degrees of freedom)	κ	$[\kappa_0, .., \kappa_{t-1}] \rightarrow [\kappa, \kappa_0 + 1 .. \kappa_{t-1} + 1]$
Alpha – (1/2 degrees of freedom	α	$[\alpha_0, .., \alpha_{t-1}] \rightarrow [\alpha, \alpha_0 + \frac{1}{2}, .. \alpha_{t-1} + \frac{1}{2}]$
Beta – (average deviation from mean)	β	$[\beta_0, .., \beta_{t-1}] \rightarrow [\beta, \beta_0 + \frac{\kappa_0 * (x_i - \bar{x})^2}{2 * (\kappa_0 + 1)}, ..., \beta_{t-1} + \frac{\kappa_{t-1} * (x_i - \bar{x})^2}{2 * (\kappa_{t-1} + 1)}]$

South Carolina

"Bayesian Online Changepoint Detection" Adams & MacKay 2007.

Bayesian Online Changepoint Detection (BOCPD): Performance

- Very high precision, high recall
- Great accuracy
- F1 score suggests good balance
- Results highly dependent on hyperparameters



Hyperparameter	value
μ	Mean([0:50])
κ	50
α	25
β	50 / var([0:50])
lambda	100

	Predicted False	Predicted True
Actually False	77.9%	0.0%
Actually True	12.2%	9.9%

Accuracy	Precision	Recall	F1 score	Earliest Correct (ms)
0.944	0.985	0.758	0.857	473.784

South Carolina

"Bayesian Online Changepoint Detection" Adams & MacKay 2007.

Expectation Maximization (EM): Theory (pt. 1)

- The EM algorithm has two steps:
  1. **Expectation step** - calculate probability of each item in set belonging to unsafe group
  2. **Maximization step** - update means, variances, and probability for each group
- The algorithm iterates over these two steps until parameters converge to stable values

Hyperparameter	interpretation
μ_1	Prior mean of safe data
μ_2	Prior mean of unsafe data
σ_1	Prior standard deviation of safe data
σ_2	Prior standard deviation of unsafe data
π	Probability of data point being unsafe

UofSC South Carolina

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Expectation Maximization (EM): Theory (pt. 2)

- The EM algorithm has two steps:
  1. **Expectation step** - calculate probability of each item in set belonging to unsafe group
  2. **Maximization step** - update means, variances, and probability for each group
- The algorithm iterates over these two steps until parameters converge to stable values

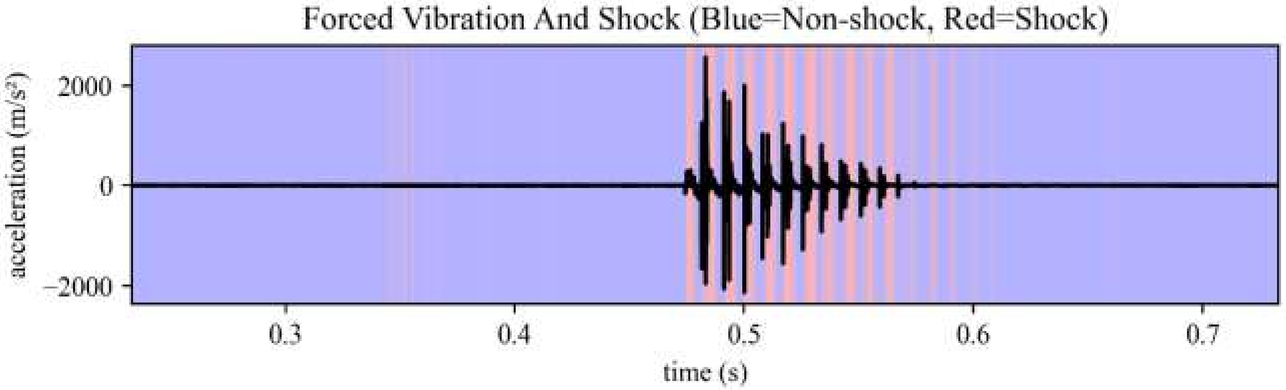
parameter	symbol	update equation
Safe mean	μ_1	$\frac{\sum_{i=1}^{N}(1-(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))\cdot Y_i))}{\sum_{i=1}^{N}(1-p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2)))}$
Change mean	μ_1	$\frac{\sum_{i=1}^{N}(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))\cdot Y_i)}{\sum_{i=1}^{N}(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))}$
Safe variance	σ_2	$\frac{\sum_{i=1}^{N}(1-(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))\cdot(Y_i-\hat{\mu}_2)^2)))}{\sum_{i=1}^{N}(1-p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))}$
Change variance	σ_2	$\frac{\sum_{i=1}^{N}(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))\cdot(Y_i-\hat{\mu}_2)^2)}{\sum_{i=1}^{N}(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2))}$
Attack probability	π	$\frac{1}{N}\sum_{i=1}^{N}(p(Y_i \sim N(\hat{\mu}_2,\hat{\sigma}_2^2)\cdot Y_i)$

South Carolina

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Expectation Maximization (EM): Performance

- Very high precision, low recall



hyperparameter	value
μ_1	Mean([0:100,000])
μ_2	20.0
σ_1	Std([0:100,000])
σ_2	Sqrt(10)
π	0.30

	Predicted False	Predicted True
Actually False	76.8%	1.12%
Actually True	12.7%	9.4%

Accuracy	Precision	Recall	F1 score	Earliest Correct (ms)
0.862	0.880	0.434	0.582	474.623

South Carolina

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Grey Systems Modeling (GM): Theory (pt. 1)

- Based around the idea of extrapolating information from observed data
- Data must be nonnegative, smooth, discrete
- Compares behavior of reference window of observations to window of interest

hyperparameter	interpretation
n	Number of points in window
c	Threshold for difference in behavior
c	Threshold for ratio of behavior

J. L. Deng. 1989. Introduction to Grey system theory. J. Grey Syst. 1, 1 (1989), 1–24.

Grey Systems Modeling (GM): Theory (pt. 2)

- Based around the idea of extrapolating information from observed data
- Data must be nonnegative, smooth, discrete
- Compares behavior of reference window of observations to window of interest

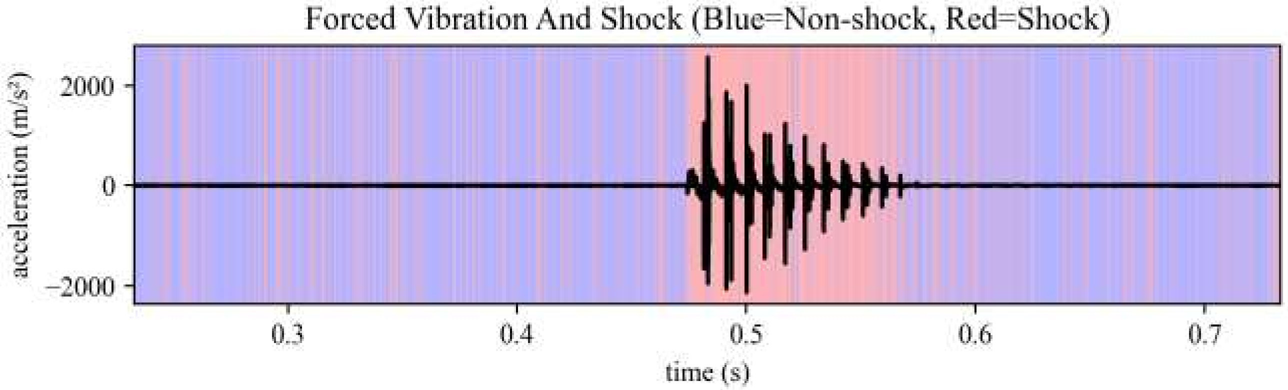
parameter	symbol	update equation
$X^{(0)}$ – (sequence of observations)	$X^{(0)}$	$X^{(0)} = X[k\colon k+n]$
$X^{(1)}$ – (cumulative sum of sequence)	$X^{(1)}$	$X^{(1)} = [X^{(0)}(0),\ldots,\sum_{i=1}^{k} X^{(0)}(i),\ldots,\sum X^{(0)}]$
$Z^{(1)}$ – (running mean of sequence)	$Z^{(1)}$	$Z^{(1)} = [X^{(1)}(0),\ldots,\frac{X^{(1)}(k-1)+X^{(1)}(k)}{2},\ldots,\frac{X^{(1)}(n-1)+X^{(1)}(n)}{2}]$
S_i – (behavioral sequence of sequence)	S_i	$s_i = \sum_{k=1}^{n-1} (\frac{X_i^{(0)}(k)-X_i^{(0)}(1)}{2}) + \frac{X_i^{(0)}(n)-X_i^{(0)}(1)}{2}$
ϵ_{ij} – (absolute degree of grey incidence)	ϵ_{ij}	$\epsilon_{ij} = \frac{1+\lvert s_i\rvert+\lvert s_j\rvert}{1+\lvert s_i\rvert+\lvert s_j\rvert+c\cdot\lvert s_i-s_j\rvert}$

J. L. Deng. 1989. Introduction to Grey system theory. J. Grey Syst. 1, 1 (1989), 1–24.

South Carolina

Grey Systems Modeling (GM): Performance

- Recall better than precision



hyperparameter	value
n	100
c	3.0
C_ratio	0.01

	Predicted False	Predicted True
Actually False	63.1%	14.8%
Actually True	4.85%	17.2%

Accuracy	Precision	Recall	F1 score	Earliest Correct (ms)
0.803	0.538	0.780	0.637	473.858

UofSC South Carolina

J. L. Deng. 1989. Introduction to Grey system theory. J. Grey Syst. 1, 1 (1989), 1–24.

Cumulative Summation (CUSUM): Theory (pt. 1)

- Detects shifts in mean level
- Assumptions:
  - Process follows normal distribution
  - Mean and standard deviation of process is known
- Change point detected if either deviation exceeds threshold

Hyperparameter	interpretation
μ	Assumed mean of the process
α	Weight parameter

G. Comert, M. Rahman, M. Islam and M. Chowdhury, "Change Point Models for Real-Time Cyber Attack Detection in Connected Vehicle Environment," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12328-12342, Aug. 2022, doi: 10.1109/TITS.2021.3113675.

Cumulative Summation (CUSUM): Theory (pt. 2)

- Detects shifts in mean level
- Assumptions:
  - Process follows normal distribution
  - Mean and standard deviation of process is known
- Change point detected if either deviation exceeds threshold

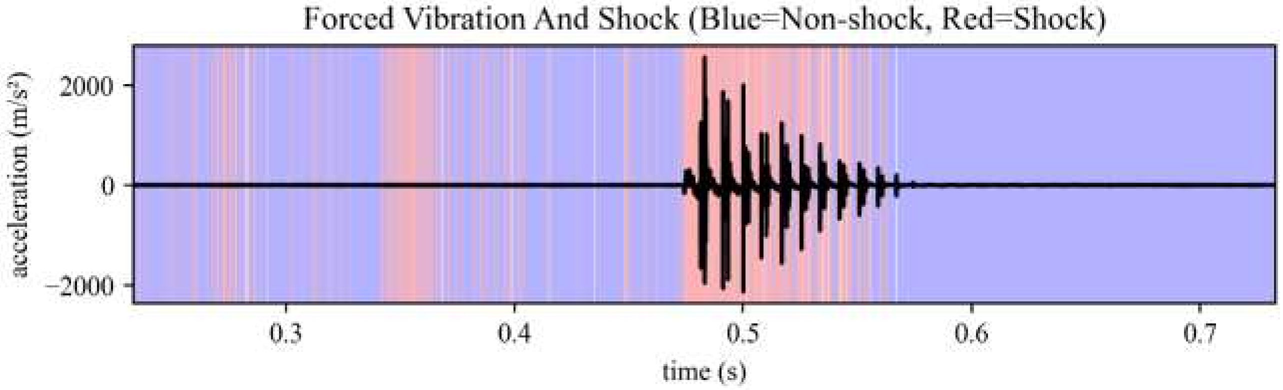
parameter	symbol	update equation
C_i^+ - (positive deviation from target)	C_i^+	$C_i^+ \rightarrow \max(0, C_{i-1}^+ + \frac{\alpha \cdot D_i}{\sigma^2}(Y_i - D_i - \frac{\alpha \cdot D_i}{2}))$
C_i^- - (negative deviation from target)	C_i^-	$C_i^- \rightarrow \max(0, C_{i-1}^- + \frac{\alpha \cdot D_i}{\sigma^2}(Y_i - D_i - \frac{\alpha \cdot D_i}{2}))$
D_i - (difference between weighted average and process mean)	D_i	$D_i \rightarrow \mu_{i-1} - \mu$
μ_i - (weighted average of value and previous average)	μ_i	$\mu_i \rightarrow \alpha \cdot \mu_{i-1} + (1 - \alpha) \cdot X_i$

G. Comert, M. Rahman, M. Islam and M. Chowdhury, "Change Point Models for Real-Time Cyber Attack Detection in Connected Vehicle Environment," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12328-12342, Aug. 2022, doi: 10.1109/TITS.2021.3113675.

Cumulative Summation (CUSUM): Performance

- Performed well overall
- Approximately equal false positive

Hyperparameter	Value
μ	Sample mean
α	0.5



	Predicted False	Predicted True
Actually False	68.7%	9.2%
Actually True	9.7%	12.4%

Accuracy	Precision	Recall	F1 score	Earliest Correct (ms)
0.811	0.573	0.562	0.567	473.772

G. Comert, M. Rahman, M. Islam and M. Chowdhury, "Change Point Models for Real-Time Cyber Attack Detection in Connected Vehicle Environment," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12328-12342, Aug. 2022, doi: 10.1109/TITS.2021.3113675.

UofSC South Carolina

Comparison of Online-CPD Algorithms

- Bayesian Online Changepoint Detection had the best accuracy, precision, and F1 score overall
- CUSUM detected the shock state earliest
- Grey Model has the best recall

	Accuracy	Precision	Recall	F1 score	Earliest Correct (ms)
BOCPD	0.944	0.985	0.758	0.857	473.784
EM	0.862	0.880	0.434	0.582	474.623
GM	0.803	0.538	0.780	0.637	473.858
CUSUM	0.811	0.573	0.562	0.567	473.772

Conclusion

South Carolina

Conclusion

- This paper demonstrates that each algorithm can classify whether a system is in shock
- A proper data transformation can significantly improve model performance
- Model selection is highly dependent on:
  - Process assumptions (what data can be collected)
  - Desired level of performance
  - Acceptable amount of error tolerance
- Propose using a combination of models:
  - A fast model with low false negative rate
  - A more robust model to verify prediction
- Future work:
  - Selecting appropriate transformations to improve algorithm performance
  - Implementing algorithms in hardware-setting

UofSC South Carolina

Dataset 7

Code Repository

THANKS!

This work is supported by:

- The National Science Foundation (NSF)
  - 1937535
  - 1956071
  - 2234921
  - 2237696
- Air Force Office of Scientific Research (AFOSR)
  - FA9550-21-1-0083

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the NSF or U.S.A.F.

UofSC South Carolina

References

- Devon Goshorn, Joud Satme, and Austin Downey. Dataset-7-forced-vibration-and-shock, October 2019. URL: https://github.com/High-Rate-SHM-Working-Group/Dataset-7-forced-vibration-and-shock
- Aminikhanghahi, S., Cook, D.J. A survey of methods for time series change point detection. *Knowl Inf Syst* **51**, 339–367 (2017). https://doi.org/10.1007/s10115-016-0987-z
- C. Jun-fu, Z. Ren-jie and L. Rui, "Grey incidence analysis based on coefficient of determination and its economic application with the data of Central Henan Urban Agglomeration," 2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009), Nanjing, China, 2009, pp. 32-36, doi: 10.1109/GSIS.2009.5408354.
- C. Truong, L. Oudre, N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
- Downey, A. (n.d.). Machine Learning for Engineering Problem Solving.
- Downey, A., & Micheli, L. (n.d.). Open Vibrations.
- G. Comert, M. Rahman, M. Islam and M. Chowdhury, "Change Point Models for Real-Time Cyber Attack Detection in Connected Vehicle Environment," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 12328-12342, Aug. 2022, doi: 10.1109/TITS.2021.3113675.
- Gundersen, G. (2019, August 13). Bayesian Online Changepoint Detection. *gregorygundersen.com*. June 20, 2024, https://gregorygundersen.com/blog/2019/08/13/bocd/
- Gundersen, G. (2019, August 0). Bayesian Inference for the Gaussian. *gregorygundersen.com*. June 20, 2024, https://gregorygundersen.com/blog/2019/04/04/bayesian-gaussian/
- J. L. Deng. 1989. Introduction to Grey system theory. J. Grey Syst. 1, 1 (1989), 1–24.

South Carolina