Supply Chain and
Logistics Institute

Online Detection Of Supply Chain Network Disruptions Using Sequential Change-Point Detection for Hawkes Processes

Khurram Yamin, Haoyun Wang, Benoit Montreuil, **Yao Xie**

School of Industrial and Systems Engineering
Georgia Institute of Technology

Monitoring logistic networks and detecting anomalies

LOGISTICS REPORT

Southern California Ports Struggle to Trim Cargo Backlog as Omicron Surges

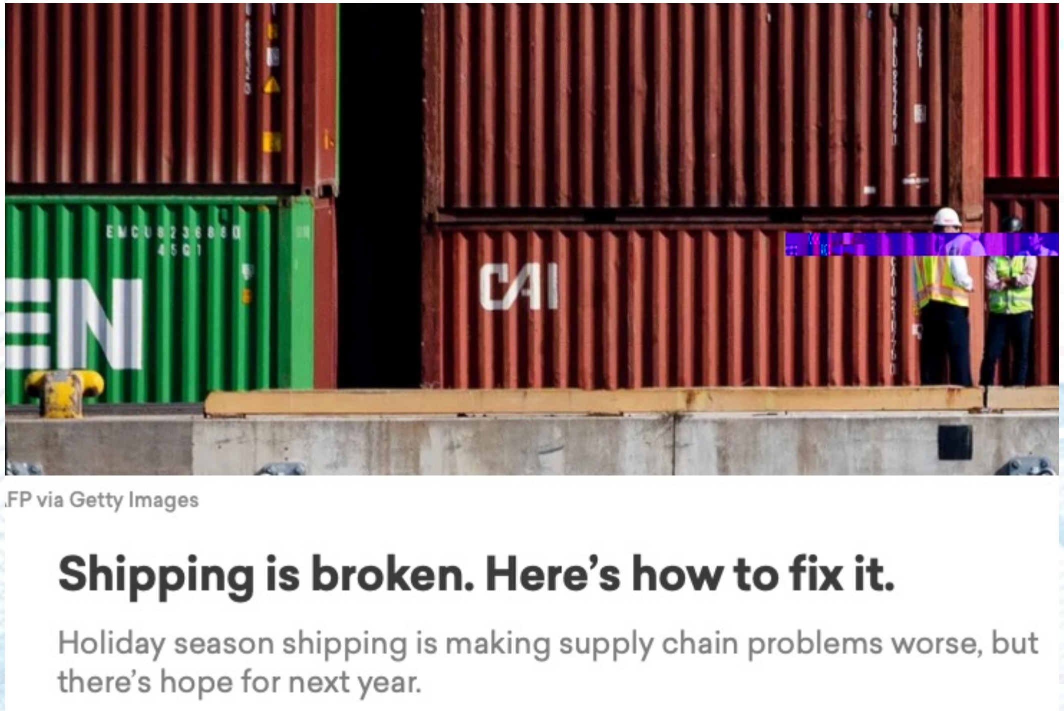
Covid-related absences sidelined about 800 dockworkers at the ports of Los Angeles and Long Beach this week

Jan 11, 2022
WSJ

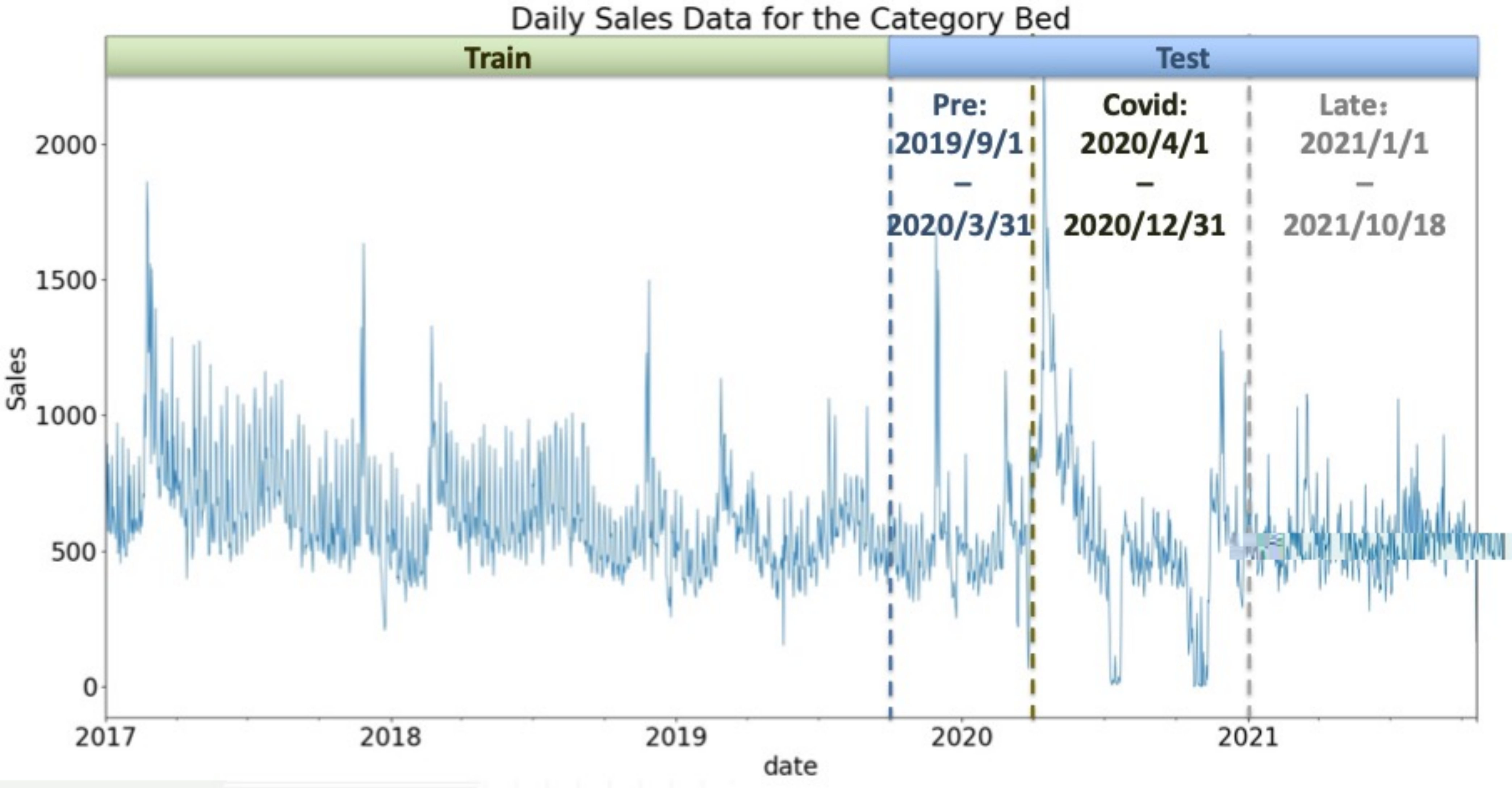
Monitoring logistic networks

- Can we make accurate predictions?
- Can we quantify uncertainties?

As of Friday morning, approximately 70 ships filled with cargo were anchored outside the ports of Los Angeles and Long Beach, which are the points of entry for more than [40 percent of US imports](). This backlog is a clear reminder that there aren't enough workers or facilities to take in all the products that are being shipped to the United States right now.

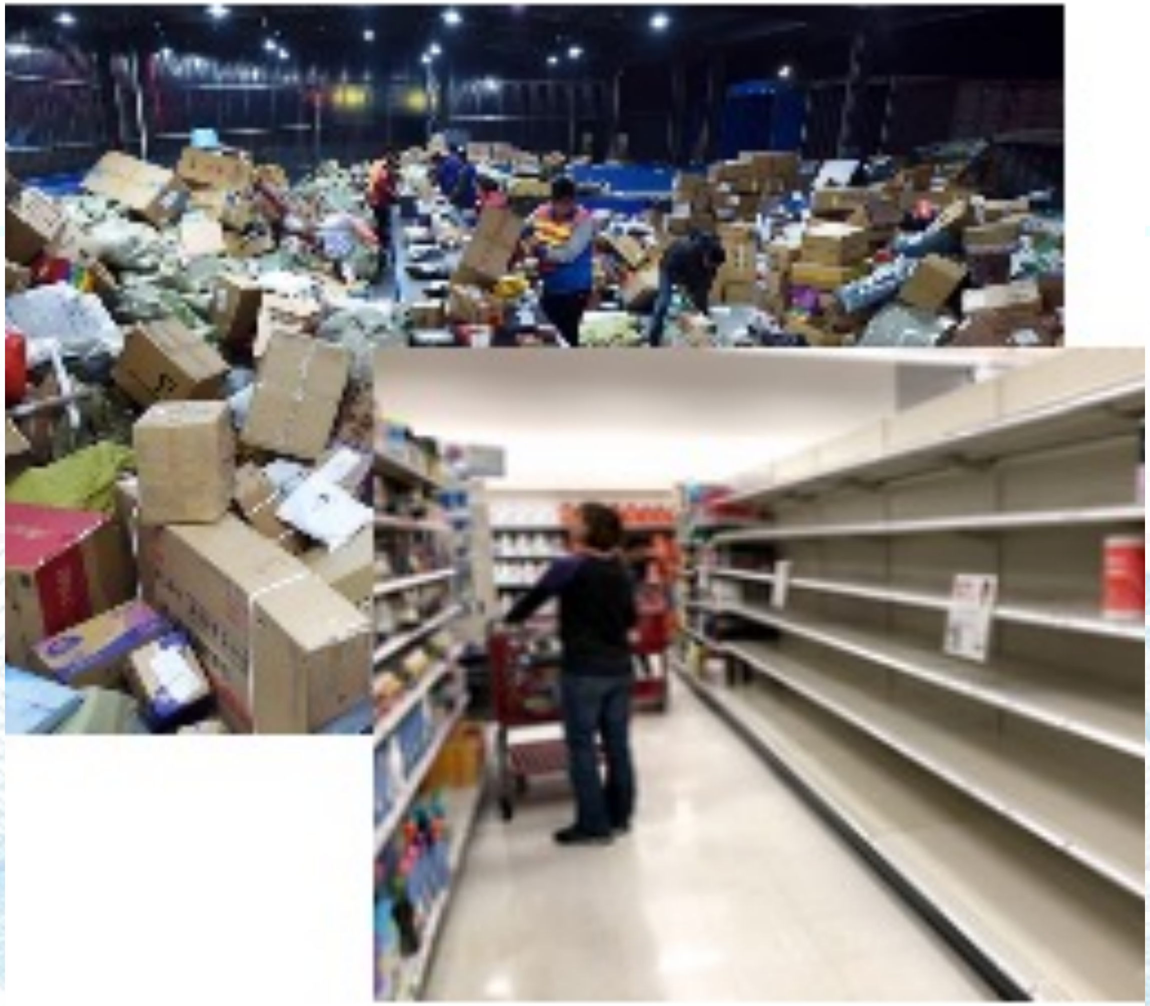


Detecting anomalies and react to it



Demand uncertainty in supply chain management

- Capacity, demand, and cost are often assumed to be known in traditional supply chain problems (Gholizadeh et al., 2018)
- In reality, they are unknown.
- When we try to estimate and predict, uncertainty cannot be ignored
    - Varying customer's demand
    - Change of pattern
    - Pandemic
    - Shortage of labor
    - Holiday effect



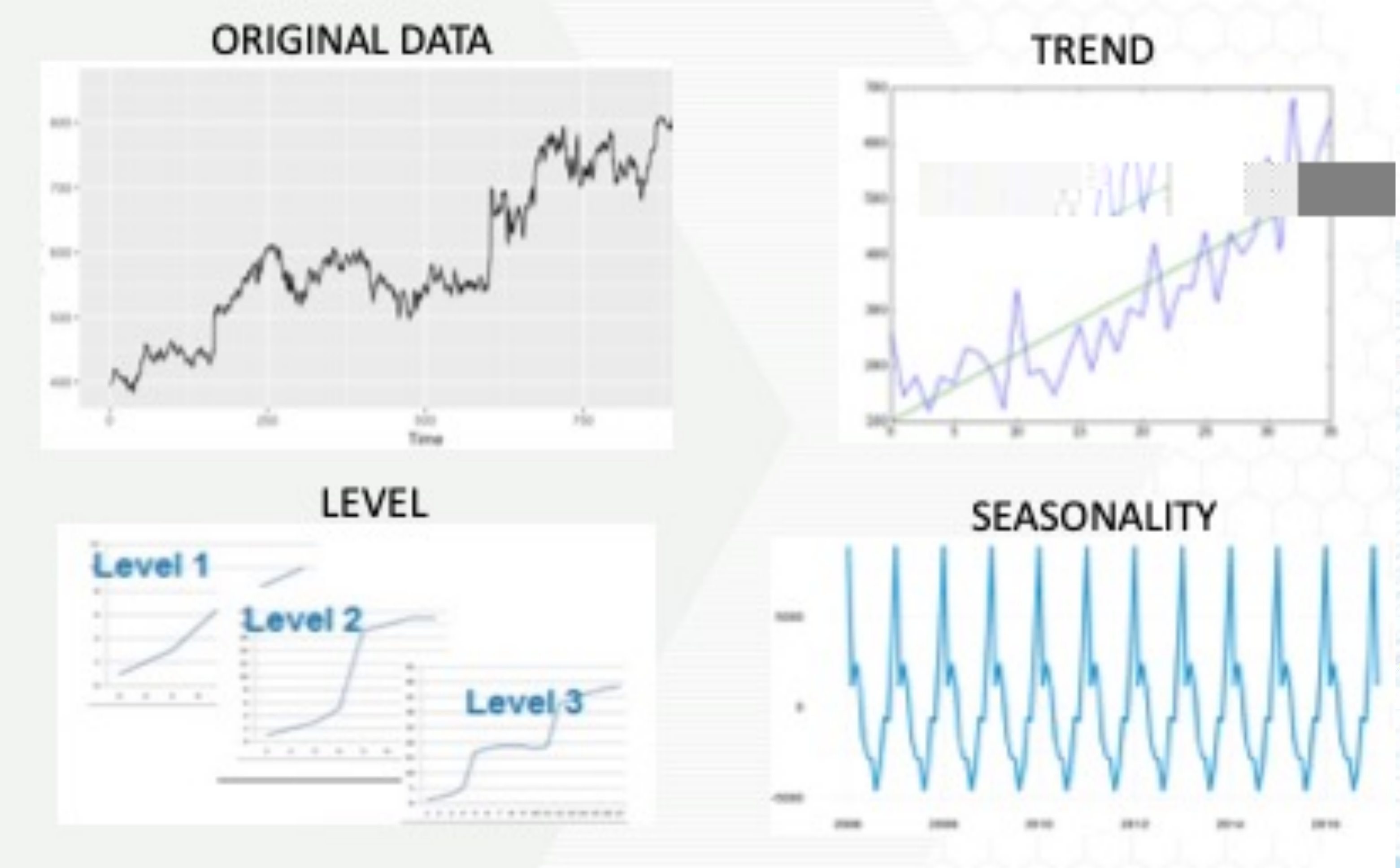
Prediction in supply chain

- Retailer and third-party logistic service providers (3PLs) reply on efficient order fulfilment process because timeliness and frequency are crucial performance factors (Leung et al. 2020)
- **Normal**: Prediction (considering seasonality and variability)
- **Abnormal**: Detect and react as quickly as possible

- E-commence sales increased by 44% yearly due to pandemic
- Demand for peak-season parcel delivery services in US ~ 4.7million/day in 2021

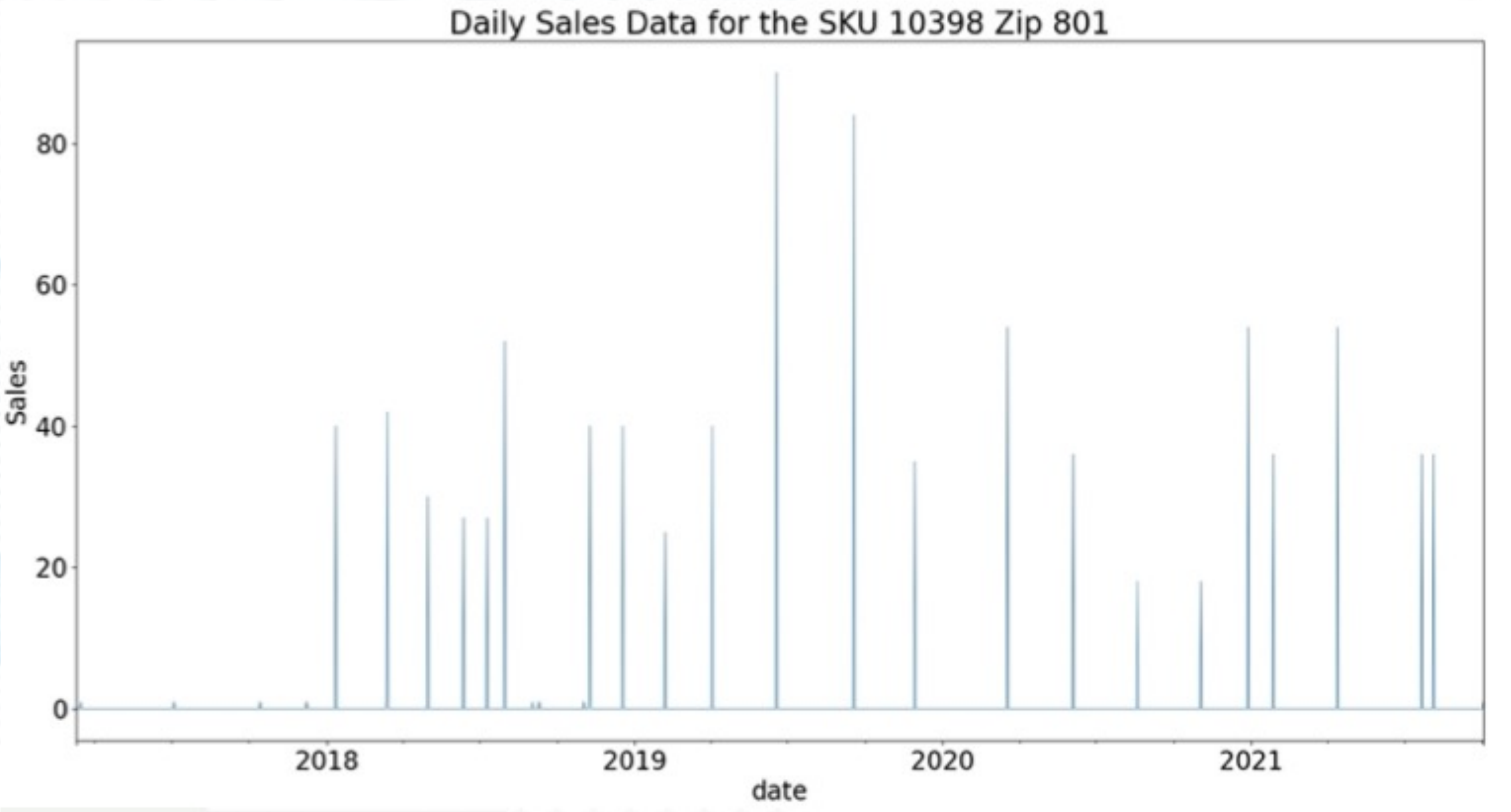
Traditional prediction: low-dimensional

Most retail demands forecasts only rely on time-series model of served sales data



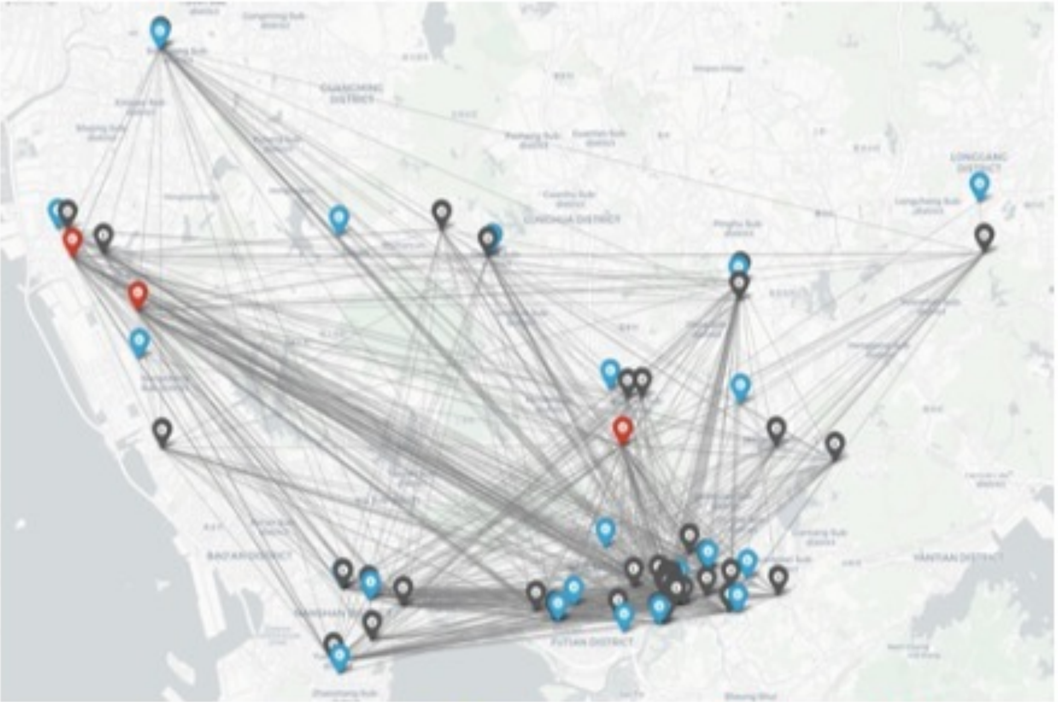
Challenge: Sparse data

Daily sale for a type of “bed” at a zip code is a sparse data
We need a different type of data modeling and prediction



Modern high-dimensional high-resolution prediction

Challenge: Complexity



40 access hubs 21 local hubs 3 gateway hubs

- Geographical Level:
  - **6** facilities
  - **67** sold-to states, **73,116** sold-to cities
  - **892** 3-digit zip-codes in USA, **2,872** 3-digit zip-codes in CA
- Product level:
  - **8** groups: Living Room, Master Room, Dining Room, …
  - **29** types of products: Armoire, Bed, Bookcase, …
  - **2,193** SKUs in 2019, **2,481** SKUs in 2020, **1,812** SKUs in 2021

High-resolution prediction

- Can we achieve prediction for sales at a location, time, category, and even a specific product?

Three Levels of Forecasts

- Sales per day per category
- Sales per day per SKU
- Sales per day per SKU per 3-digit Zip-Code

Forecasting Output of Category-Bed

Forecast Date	prediction	Start Date	actual	Sigma	Level - Category	Level - SKU	Level - ZipCode
9/3/19	752	9/2/19	691	0	Bed		
9/4/19	672	9/2/19	583	0	Bed		
9/5/19	600	9/2/19	531	0	Bed		
9/6/19	620	9/2/19	429	0	Bed		

Forecasting Output of SKU-7250758

Forecast Date	prediction	Start Date	actual	Sigma	Level - Category	Level - SKU	Level - ZipCode
9/3/19	54	9/2/19	88	0	Bed	7250758	
9/4/19	43	9/2/19	54	0	Bed	7250758	
9/5/19	48	9/2/19	40	0	Bed	7250758	
9/6/19	50	9/2/19	32	0	Bed	7250758	

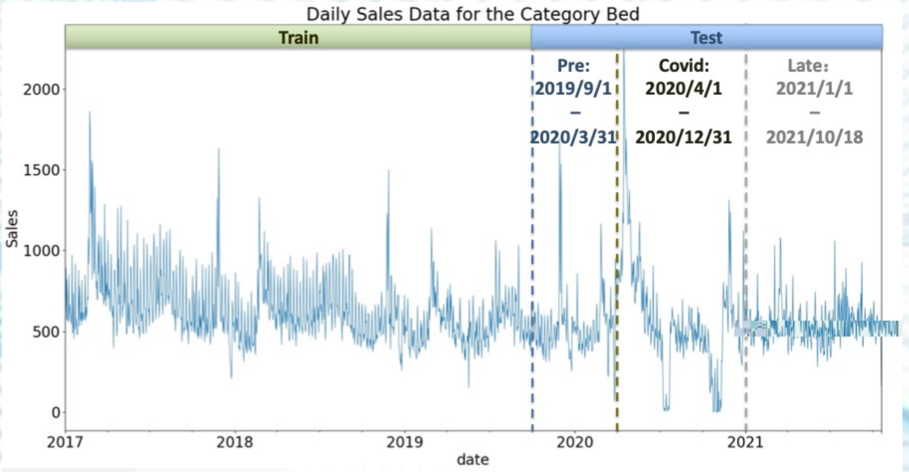
Conformal prediction interval for dynamic time-series. Chen Xu, and Yao Xie. ICML 2021 (Long presentation, top 3%).

Inferring serial correlation with dynamic backgrounds. Song Wei, Yao Xie, Dobromir Rahnev. ICML 2021.

Sequential adversarial anomaly detection for one-class event data. S. Zhu, H. Shaowu, M. Zhang, Y. Xie. Major Revision, INFORMS Journal on Data Sciences.

Objective and Context

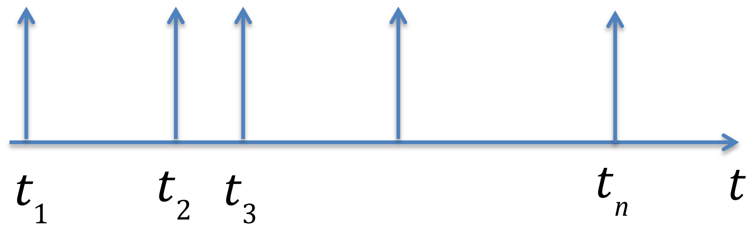
- Our paper seeks to detect an inflection or change-point resulting from the Covid-19 pandemic on supply chain data received from a large furniture company
- Covid-19 created new needs, and it is logical to question whether supply chains were affected as people looked for new products to satisfy those needs.
- Such a question is normally extremely difficult to answer because of the lack of publicly available, up to date, and robust supply chain data



IPIC 2023

How to model sparse and asynchronous time series data?

- **Hawkes processes** (Hawkes 1971)
- Point-process: a sequence of random events at times $\{t_1, t_2, \dots\}$



history

$$\lambda(t)dt = P\{\text{event in } [t, t + dt) | H^t\}$$

$$\lambda(t|H^t) = \lim_{\Delta t \to 0} \frac{E[N(t + \Delta t)|H_t]}{\Delta t}$$

A natural framework for prediction.

Hawkes, Alan G. "Spectra of some self-exciting and mutually exciting point processes." *Biometrika* 58, no. 1 (1971): 83-90.

Alan Hawkes

Hawkes process

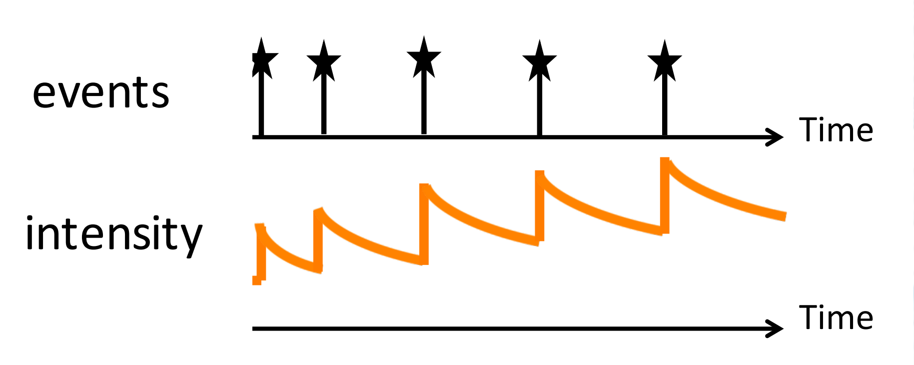
- A sequence of events at times $\{t_1, t_2, \ldots\}$

$$\lambda(t) = \mu(t) + \alpha \sum_{t_k < t} \phi(t - t_i)$$

Baseline intensity

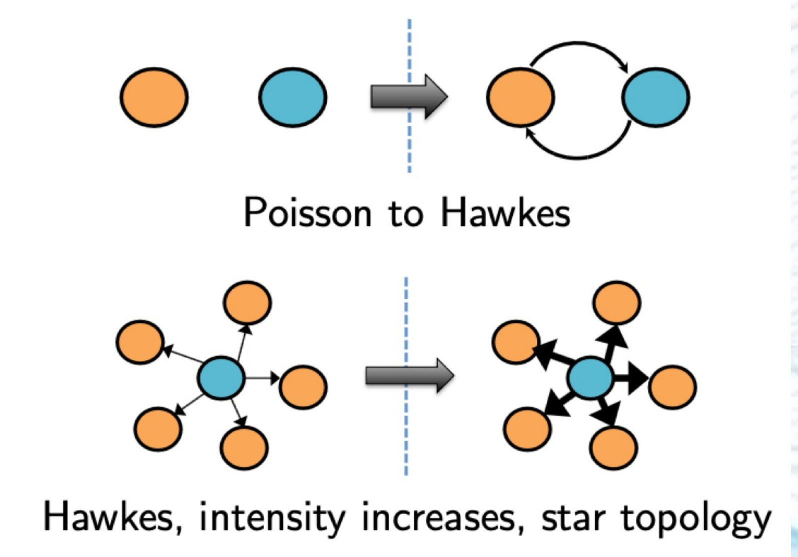
Self-influence

Influence kernel function



Background – Change Point Detection

- Data: event occurrence time and location
- Unknown change-point κ , where after the change-point the background intensities and triggering effects between nodes change.



Cumulative sum (CUSUM)

- When the post-change parameters can be estimated accurately, CUSUM is a computational and memory efficient detecting procedure

$$S_t^{\mathrm{CUSUM}} = \sup_{0 \le \nu \le t} \ell_{\nu,t}$$

- where $\ell_{\nu,t}$ is the log-likelihood ratio up to time t between the post-change and pre-change intensity functions as if ν is the true change-point, and can be computed recursively if the data are discrete and i.i.d.

$$S_{t+1}^{\mathrm{CUSUM}} = \max\left\{ S_t^{\mathrm{CUSUM}} + \log \frac{f_1(x_{t+1})}{f_0(x_{t+1})},\ 0 \right\}$$

Generalized likelihood ratio (GLR)

- When the post-change parameters are unknown, we can compute the generalized likelihood ratio in a sliding window to reflect the difference between the current data and the pre-change mode

$$S_t^{\mathrm{GLR}} = \sup_{\boldsymbol{\mu}_1, \boldsymbol{A}_1} \ell_{t-w,t,\boldsymbol{\mu}_1,\boldsymbol{A}_1}$$

- where $\ell_{t-w,t,\boldsymbol{\mu}_1,\boldsymbol{A}_1}$ is the log-likelihood ratio up to time t as if t − w is the true change-point, and GLR takes the maximum over all potential post-change scenarios

Data

- The dataset contains the location and times of product orders.
- We investigate the orders of the most popular product – the work desk.
- First, we examine orders for the US; then we narrow down orders to California.
- On the national level, we use states as nodes for the Hawkes Process
- On the state level, we use counties as nodes.
- group the orders into these levels

Experimental Setup

- We use the March 2018 to March 2019 data to train the Maximum Likelihood Estimates of the pre-change parameters of the Hawkes Process Network

- We would expect the CUSUM and GLR statistics to remain small until roughly March 2020 when the WHO declared Covid-19 a global pandemic and raise significantly after that

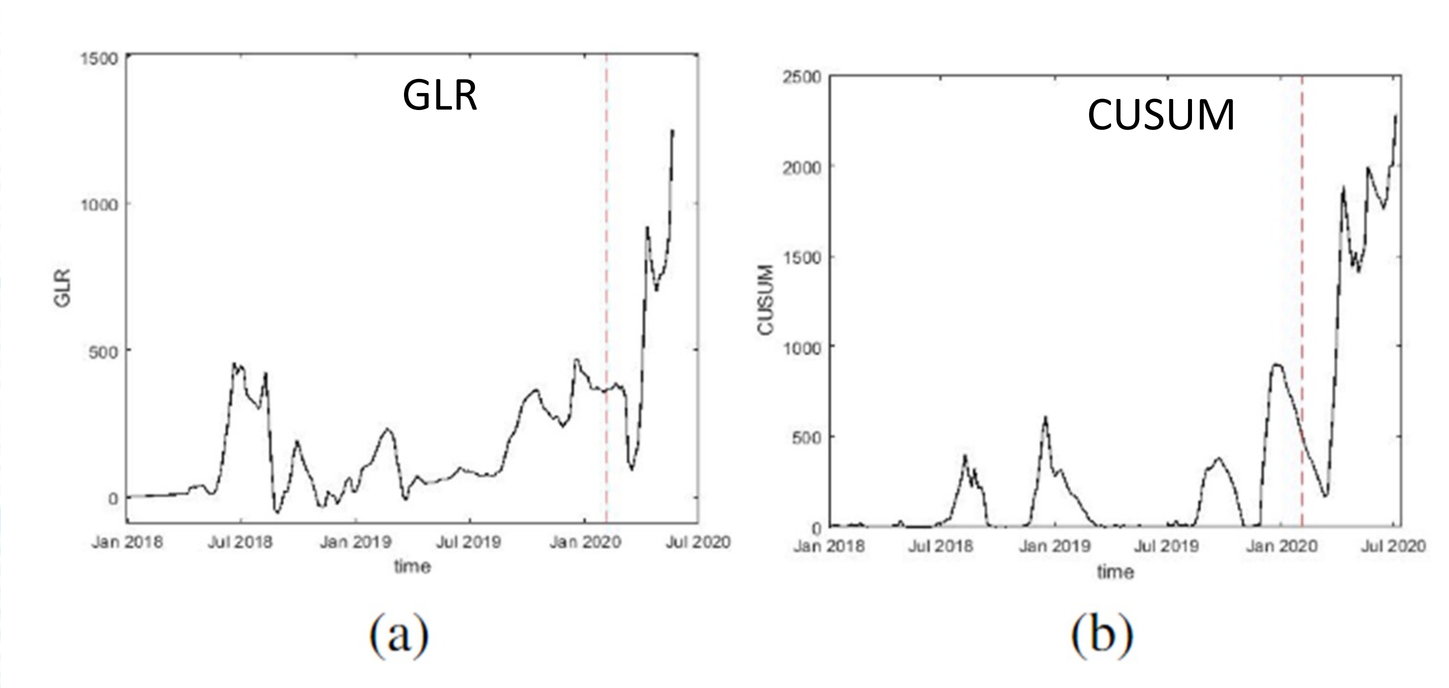
Experimental Setup Continued

- For GLR, we design a window length $w = 100$ days based on life experiences

- For CUSUM, the post-change parameter μ_1 can be set to $2\mu_0$ or $0.5\mu_0$ to detect a change, either a surge or a downfall, in average demand

Work Desk – United States

- (a) GLR and (b) CUSUM statistic over time for national orders.
- The x-axis is in days starting from January 21st, 2018.
- The vertical line marks March 1st, 2020, when Covid was declared a pandemic



Work Desk – United States continued

- Width of the directed edges corresponds to the interstate influences,
- Size of the node is proportional to the background intensity.

Pre-change

(a)

Post-change

(b)

IPIC 2023

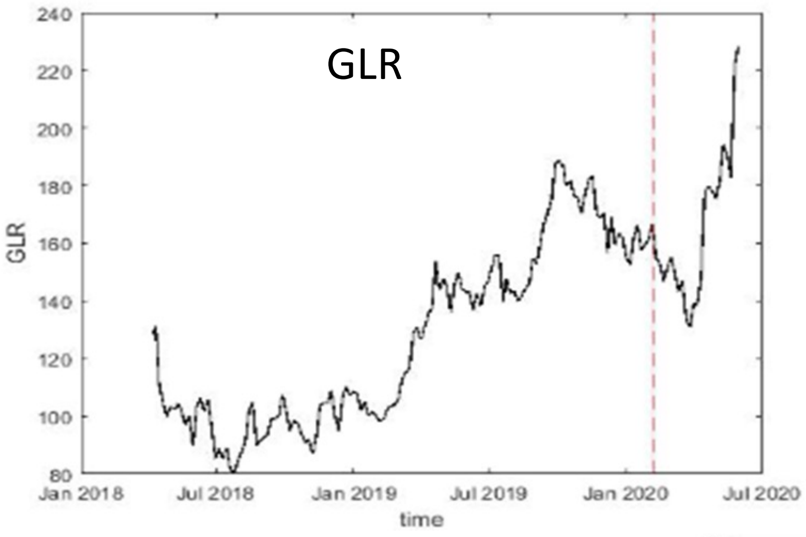
Results – United States

- In the national case, the **GLR score spikes** after March of 2020 in a way that it never does between March of 2019 and March of 2020

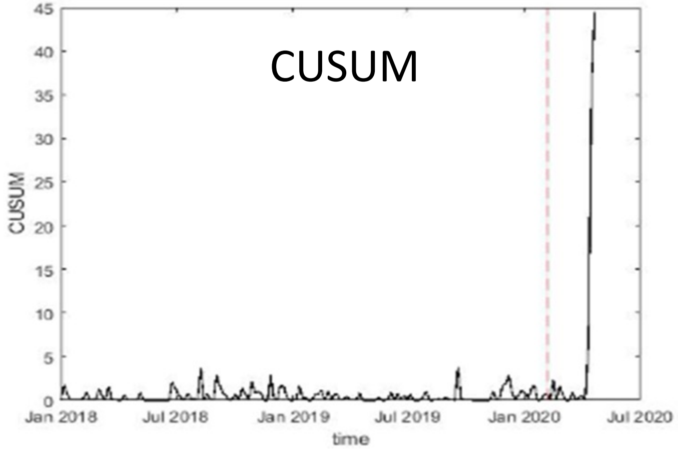
- In the visualization of the Hawkes Process Model, we can see very strong causal effects between states that change between the pre-change and post-change model

- We successfully capture the disruption in the distribution of orders caused by Covid-19 by doubling μ_0 in the CUSUM Model

Work Desk – California

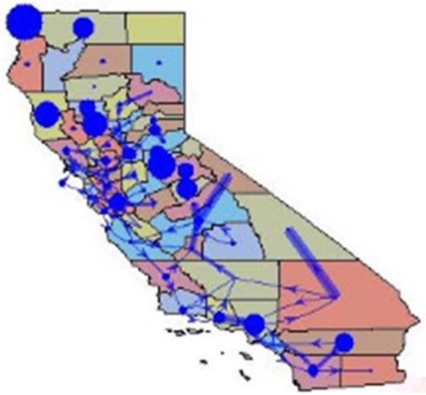


(a)



(b)

Pre-change



(a)

Post-change

(b)

Results - California

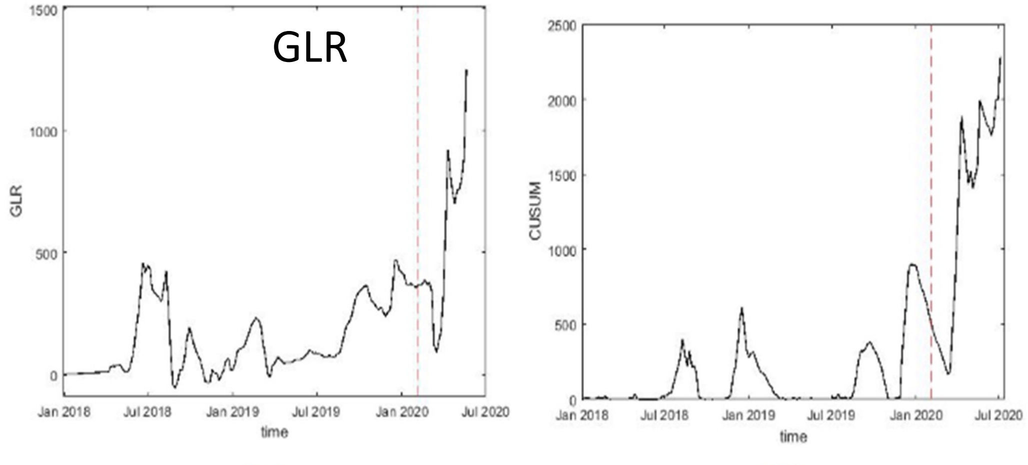
- In the California case, GLR spikes post March 2020 but is not extreme in relative magnitude compared to the spikes that came between March 2019 and March 2020, or to the relative magnitude of spike that we saw in the GLR score of the national case

- Several interesting patterns in the model such as **counties with small populations in the middle part of California** still exhibiting some influence on surrounding counties

- CUSUM can detect the surge in demand

Summary

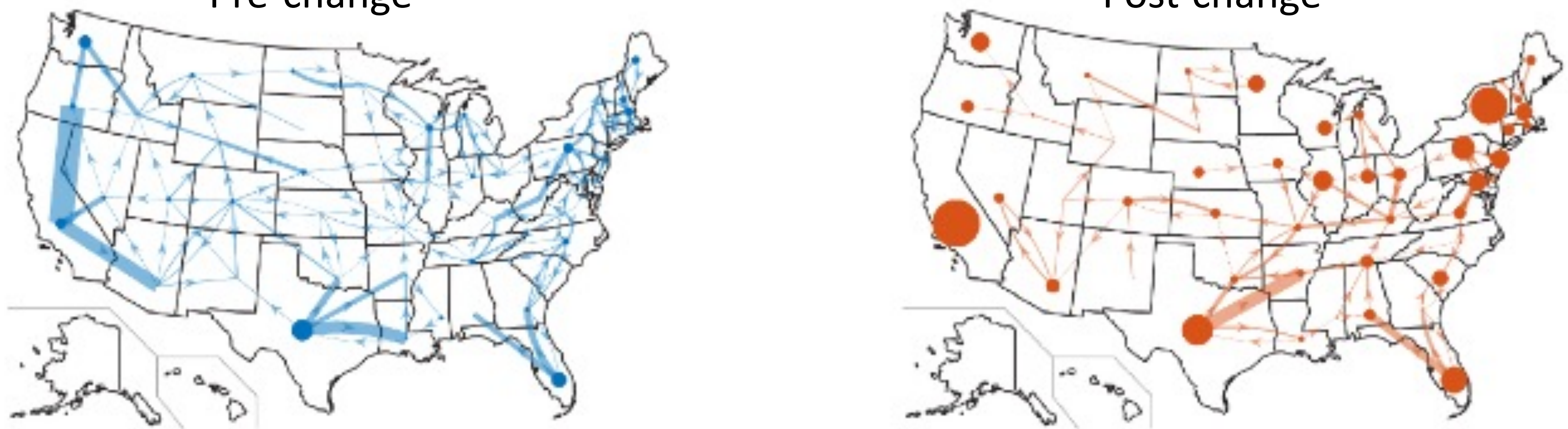
- First application of sequential change-point detection to real supply chain data in peer-reviewed literature
- CUSUM performance better than GLR, and more efficient in computation and memory
- Interpretable results: Influence networks



(a) (b)

Pre-change

Post-change



IPIC 2023

Next Steps

- More extensive analysis could be done to determine how sensitive CUSUM is to changes in post-change parameters, and across all states, what tolerance ranges we could be confident that CUSUM would perform better than GLR
- References

- Sequential change-point detection for mutually exciting point processes over networks. Haoyun Wang, Liyan Xie, **Yao Xie**, Alex Cuozzo, Simon Mak. Technometrics, Vol. 65, No. 1, pp. 44-56, 2023.
- Online Detection of Supply Chain Network Disruptions Using Sequential Change-Point Detection for Hawkes Processes. Khurram Yamin, Haoyun Wang, Benoit Montreuil, **Yao Xie. IPIC 2023.**