Anytime Information Cascade Popularity Prediction via Self-Exciting Processes

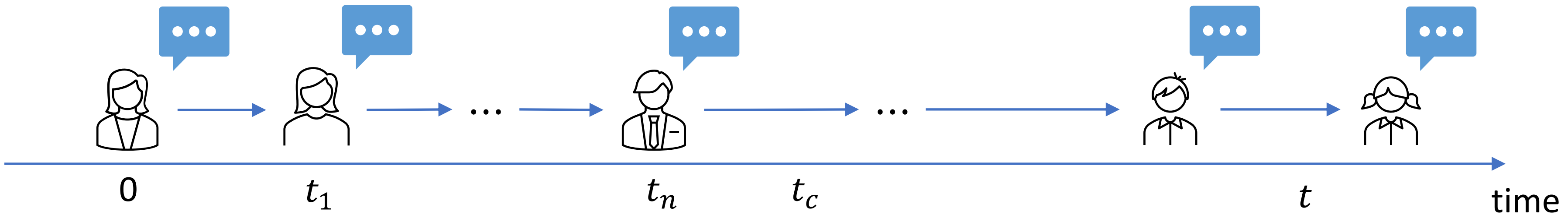
Xi Zhang, Akshay Aravamudan, Georgios C. Anagnostopoulos

Department of Computer Engineering & Sciences, Florida Institute of Technology, Melbourne, FL, USA.

Correspondence to: Xi Zhang <zhang2012@my.fit.edu>

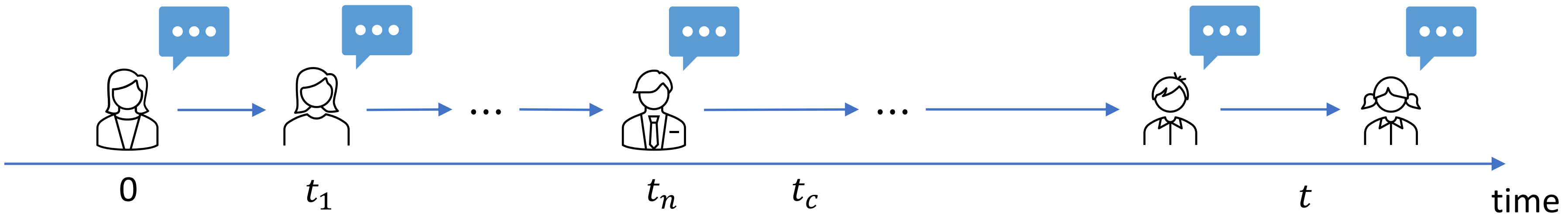
Anytime Cascade Popularity Prediction

Cascade



Anytime Cascade Popularity Prediction

Cascade



Phenomena "rich-get-richer" → heavy-tailed distribution of cascade sizes.

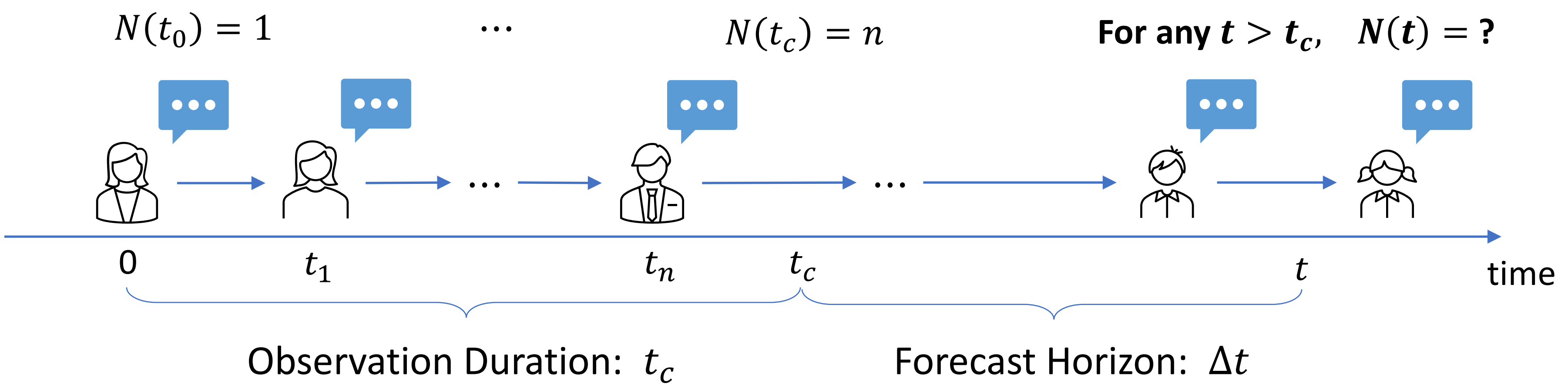
explains ↑

Modeling The Hawkes process is a **self-exciting** process that has been broadly used in modeling cascade dynamics.

Anytime Cascade Popularity Prediction

Popularity

Cascade



Phenomena “rich-get-richer” ➡ heavy-tailed distribution of cascade sizes.

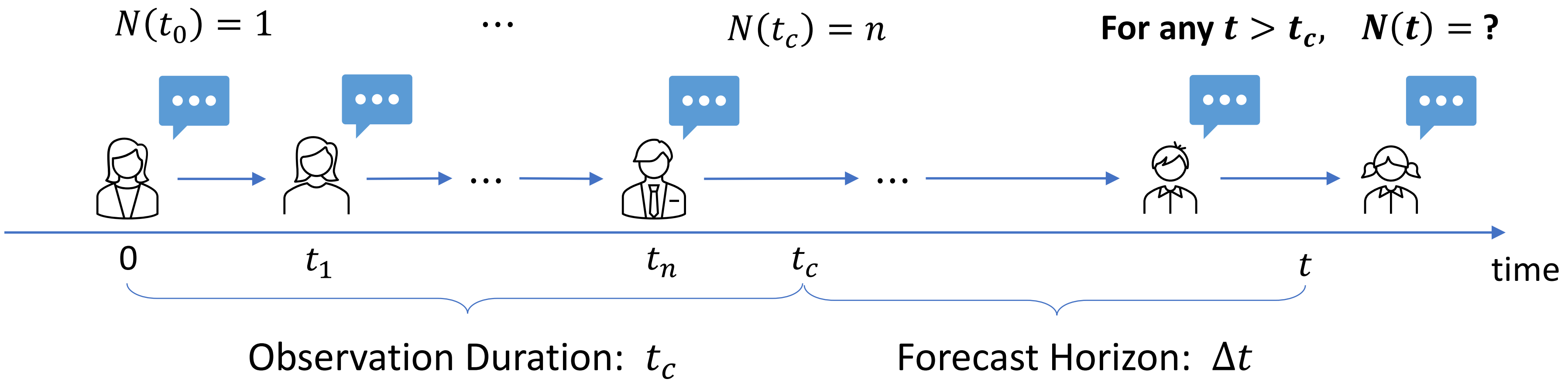
explains

Modeling The Hawkes process is a **self-exciting** process that has been broadly used in modeling cascade dynamics.

Anytime Cascade Popularity Prediction

Popularity

Cascade



Phenomena "rich-get-richer" ➡ heavy-tailed distribution of cascade sizes.

explains

Modeling The Hawkes process is a **self-exciting** process that has been broadly used in modeling cascade dynamics.

Prediction The conditional (on the observed history) mean count of fitted Hawkes process is employed.

Contributions

For general marked Hawkes Point Process(MHPP), we derive **closed-form** expressions for the **conditional** (on the observed history $\mathcal{H}_{t_c}$) mean and variance of its counting process $\mathrm{N}(t)$ at $t \geq t_c$.

For anytime popularity prediction, we propose **Cascade Anytime Size Prediction via self-Exciting Regression model** (**CASPER**), a Hawkes process based **predictive** model, which minimize the prediction error directly, rather than to maximize the generative likelihood value.

Conditional Moments Derivation Procedures

A Hawkes process $N(t)$ can be equivalently viewed as a branching process, s.t.

$$\boldsymbol{N(t)} = \sum_{\boldsymbol{k \geq 0}} \boldsymbol{N_k(t)}$$

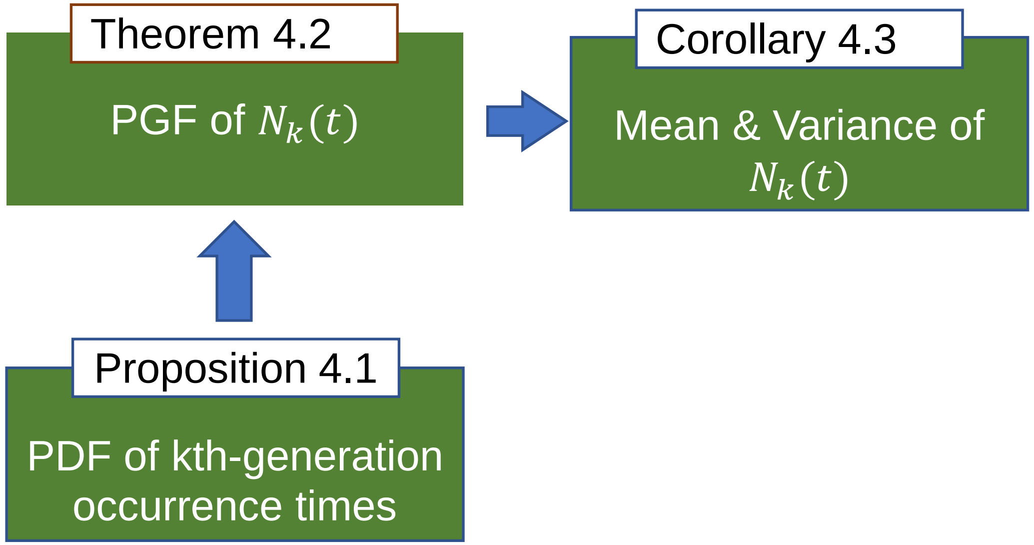
where $N_k(t)$ is k-th generation counting process.

Conditional Moments Derivation Procedures

A Hawkes process $N(t)$ can be equivalently viewed as a branching process, s.t.

$$\boldsymbol{N(t)} = \sum_{\boldsymbol{k \geq 0}} \boldsymbol{N_k(t)}$$

where $N_k(t)$ is k-th generation counting process.



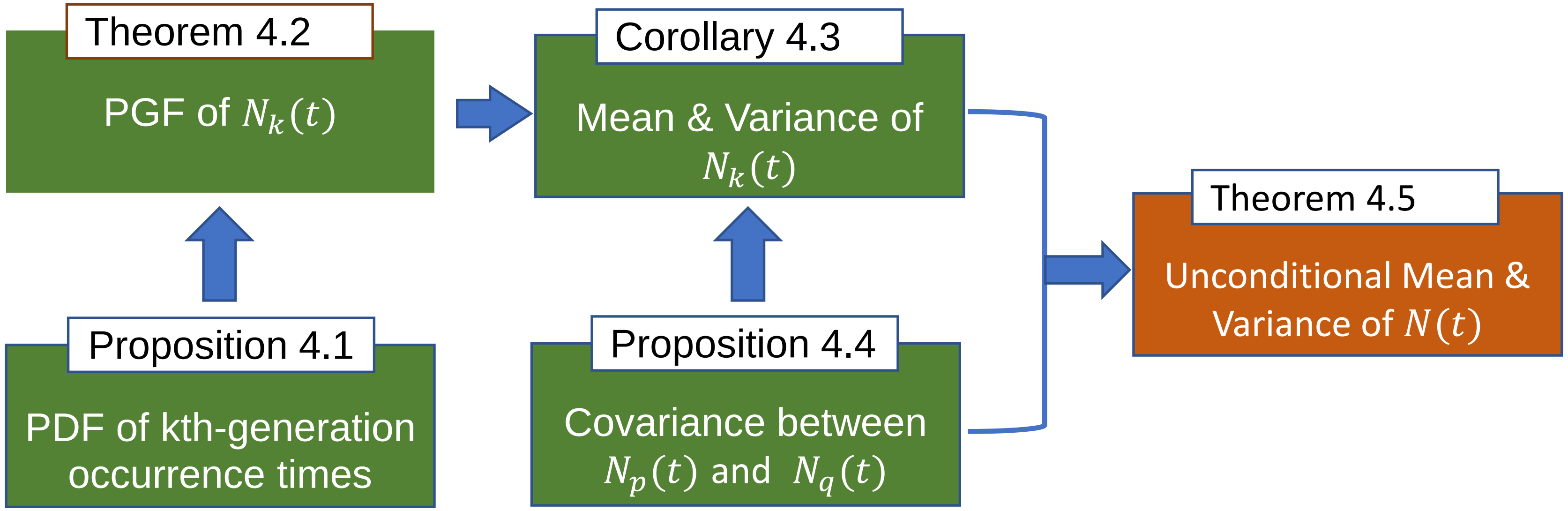
Generational count results

Conditional Moments Derivation Procedures

A Hawkes process $N(t)$ can be equivalently viewed as a branching process, s.t.

$$\boldsymbol{N(t)} = \sum_{\boldsymbol{k \geq 0}} \boldsymbol{N_k(t)}$$

where $N_k(t)$ is k-th generation counting process.



Generational count results

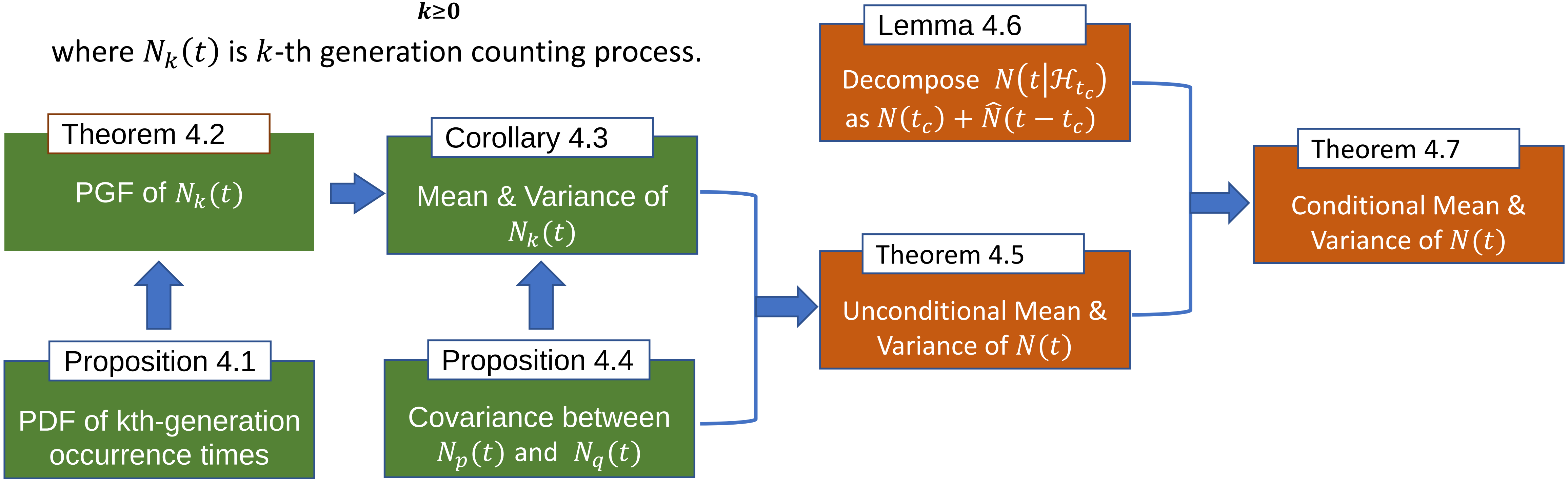
Process count results

Conditional Moments Derivation Procedures

A Hawkes process $N(t)$ can be equivalently viewed as a branching process, s.t.

$$\boldsymbol{N(t)} = \sum_{\boldsymbol{k \geq 0}} \boldsymbol{N_k(t)}$$

where $N_k(t)$ is k-th generation counting process.



Generational count results

Process count results

CASPER

- Given $\mathcal{H}_{t_c}$, the observed history up to time t_c.
- Consider (i, j) s.t. $t_i < t_j \leq t_c$, then

$$\ell_{ij}(\boldsymbol{\theta}) \triangleq (\mathbb{E}\{N(t_j|\mathcal{H}_{t_i};\boldsymbol{\theta})\} - j)^2 \qquad (11)$$

  is the squared loss between the predicted and true count at time t_j given observations up to time t_i.
- Let $\mathcal{S}(t_c) \triangleq \{(i, j) : 0 < t_i < t_j \leq t_c\}$, CASPER's overall loss function is defined as

$$L(\boldsymbol{\theta}|\mathcal{H}_{t_c}) \triangleq \frac{1}{|\mathcal{S}(t_c)|} \sum\sum_{(i,j)\in\mathcal{S}(t_c)} \ell_{ij}(\boldsymbol{\theta}) \qquad (12)$$

Formulation

Assumption: diffusion as MHPP $N(t;\boldsymbol{\theta})$

↓

Train

$\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathcal{H}_{t_c})$

↓

Predict

$N_{\text{pred}}(t) \triangleq \mathbb{E}\{N(t|\mathcal{H}_{t_c};\boldsymbol{\theta}^*)\}$

Results on Synthetic Dataset

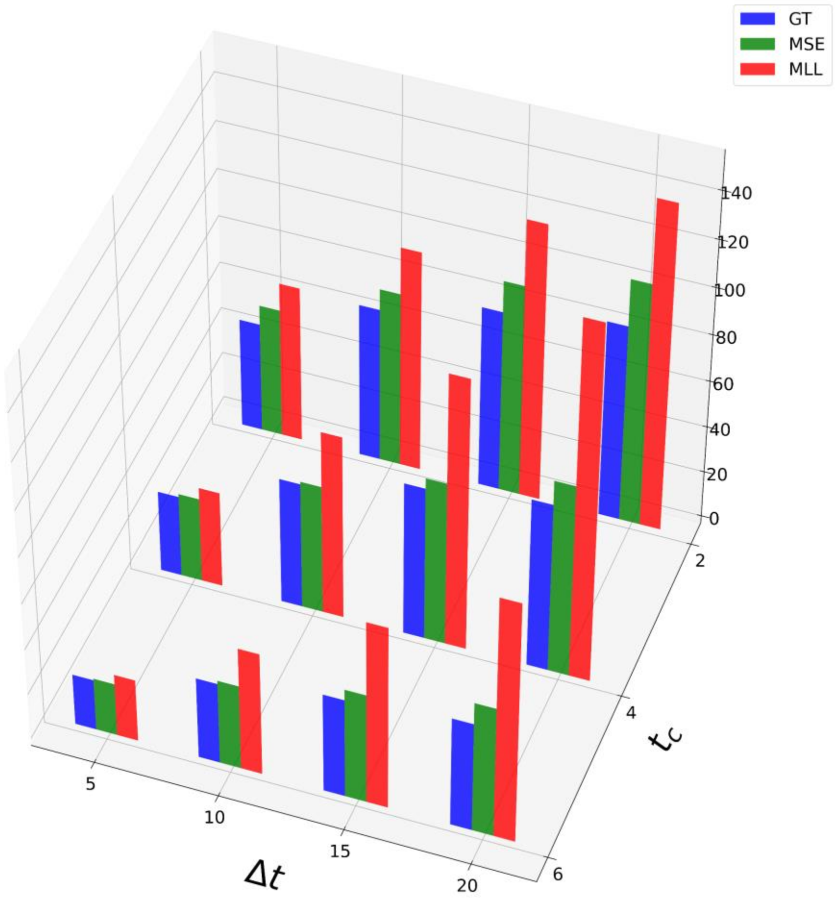


Fig 2: Average APE% on Synthetic Dataset

Predictive vs Generative Learning Approach

Models

- **GT**: ground truth models
- **MSE**: models trained by minimizing the overall loss in Eq (12) – *the predictive learning approach.*
- **MLL**: models trained by maximizing likelihood – the *generative learning approach.*

Conclusion

CASPER's predictive learning approach

- outperforms the generative learning approach.
- highly competitive to the ground truth model.

Results on Real World SEISMIC Twitter Dataset

Training Requires	Algorithms Compared
Observed history	• **MaSEPTiDE**
Additional fully-observed cascades	• **TiDeH**: designed for larger t_c • **Eb-MaSEPTiDE**: aids for smaller t_c

Fig 3a:Short-term prediction with $\Delta t = 4$ hours

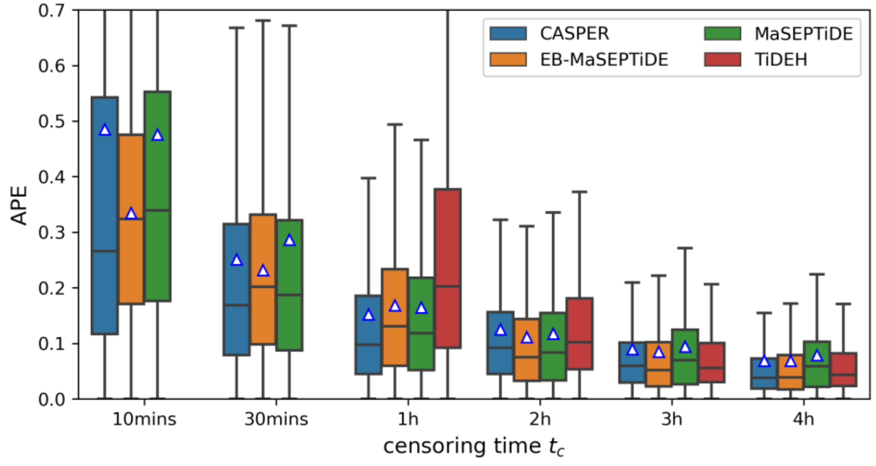
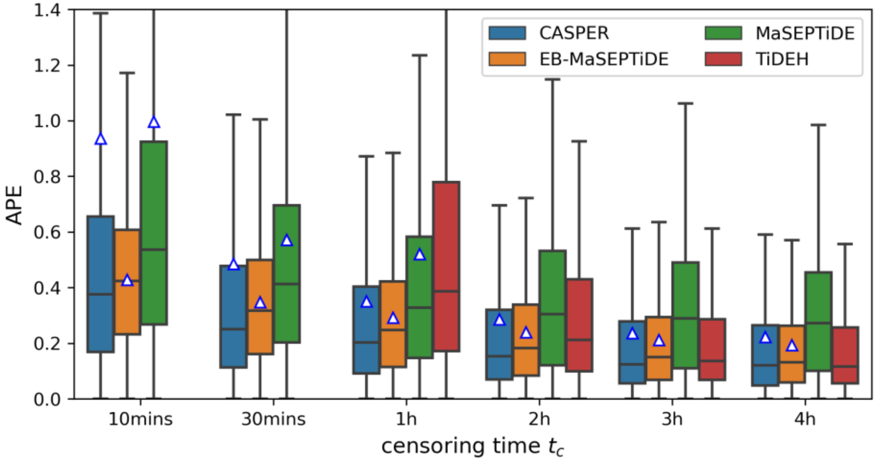


Fig 3b:Long-term prediction with $\Delta t = 4$ days



CASPER exhibits competitive, if not the best performance

- Outperform MaSEPTiDE across all scenarios.
- Competitive to TiDeH for predictions with long observation periods.
- For early-stage prediction, CASPER attains lowest median, but exhibits mean than Eb-MaSEPTiDE.

Thanks for watching