When in Doubt: Neural Non-Parametric Uncertainty Quantification for Epidemic Forecasting

Harshavardhan Kamarthi, Lingkai Kong, Alexander Rodriguez, Chao Zhang, B. Aditya Prakash

College of Computing

Georgia Institute of Technology

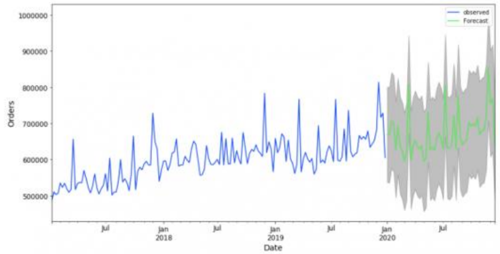
Presented at NeurIPS 2021

Time Series Forecasting

- Important and well studied machine learning problem
- Covers wide-range of domains:
  - economics
  - retail
  - weather forecasting
  - epidemiology
- Given:
  - Historical sequences from past
  - Current sequence
- Predict: Future sequence values

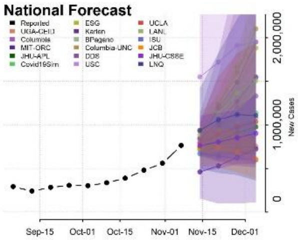
Why forecasting?

- Enable reliable and robust decision making in real world



Sales forecast

Order more supplies?



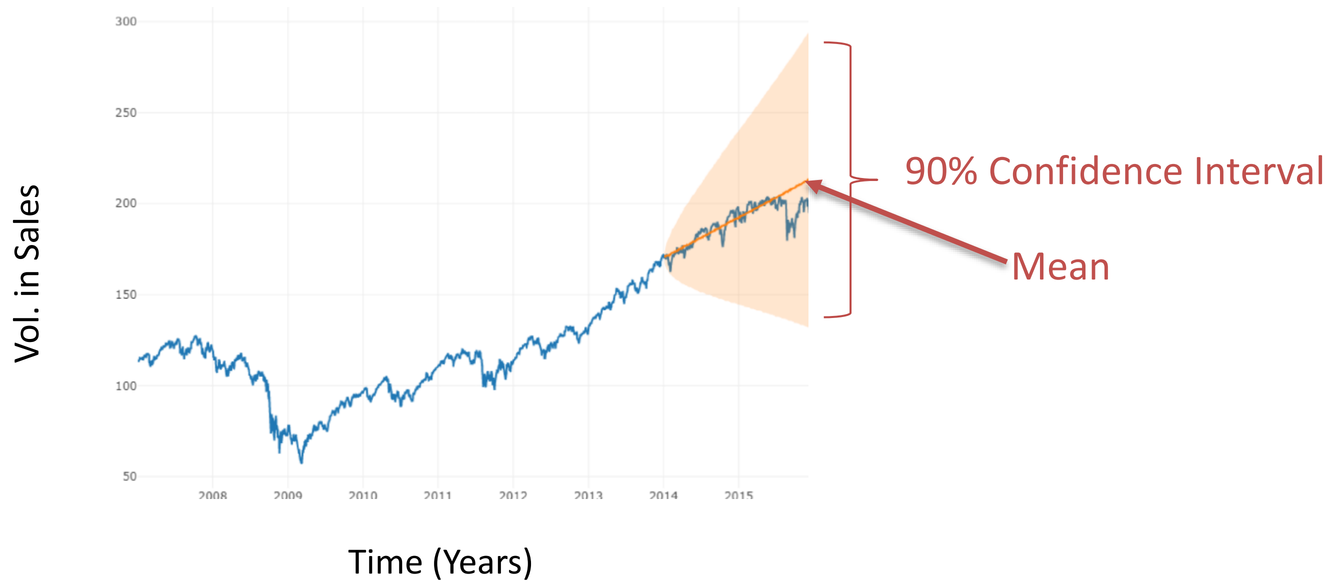
Covid-19 Cases

Lockdown policy

Kamarthi et.al 2021

Probabilistic forecasts

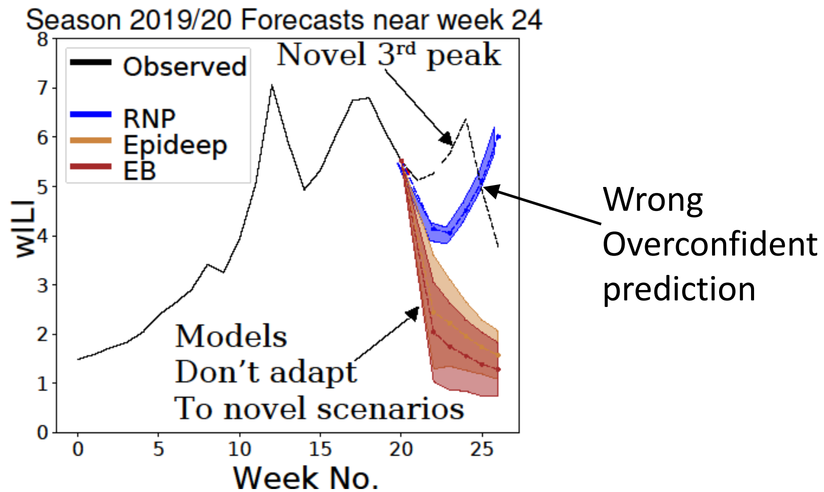
- Predictions with uncertainty
  - Mean: Most probable point estimate
  - Confidence interval: Range around mean where target lies with high confidence



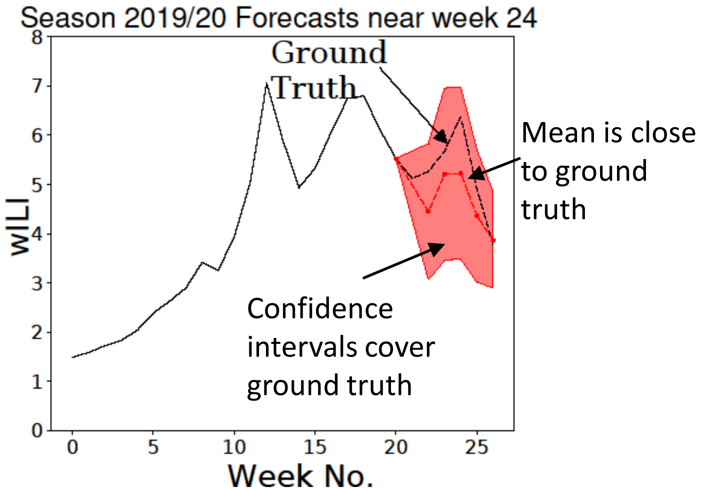
Kamarthi et.al 2021

Why Accurate AND Well-Calibrated forecasts?

- Accurate: Mean of forecast distribution close to ground truth
- Well-Calibrated: Confidence intervals cover ground truth
  - Especially during uncertain or anomalous scenarios
  - When point forecasting is harder



Forecast mean and confidence intervals of top forecasting models



Forecast mean and confidence intervals of our model (EpiFNP)

Example: Flu Forecasting

- Predict influenza incidence for next 4 weeks
- Important for public health policy, intervention planning, etc.

CENTERS FOR DISEASE CONTROL AND PREVENTION

Iowa health officials warns of "twindemic" as flu season approaches

by Deion Broxton | Friday, October 15th 2021

TOP STORY

Public health officials encourage flu vaccinations as 2021-22 season looms

Richard Craver Sep 26, 2021 0

BUSINESS

Hand Sanitizer Sales Jumped 600% in 2020. Purell Maker Bets Against a Post-Pandemic Collapse

Gojo Industries adds second factory, expecting new hygiene habits will remain after Covid-19 crisis fades

CDC's wILI Dataset

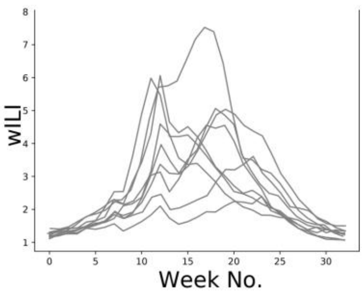
- CDC's ILINet surveillance system collects publicly available *wILI* (**weighted Influenza-like estimates**) :
  - anonymized aggregated indicator of out-patient cases with flu-like symptoms
- wILI signals for US and 8 HHS regions
- Weekly wILI available for 2004 - present

2003\04 2004\05 2018\19 2019\20

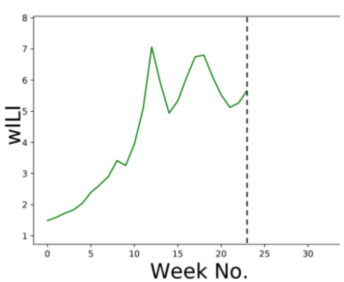
Historical wILI sequences

Real-time Forecasting Setup

1. Given Data

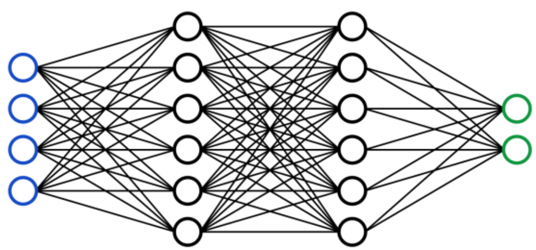


Data of historical sequences

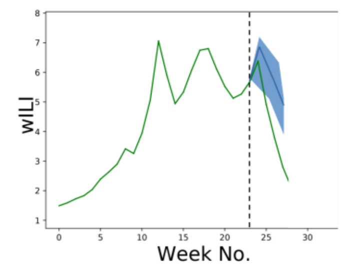


Sequence till current week

2. Train Model

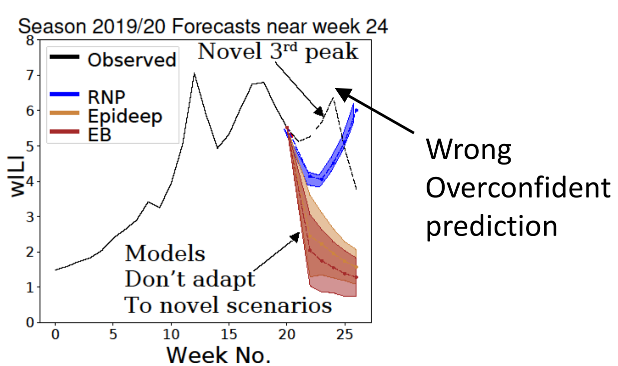


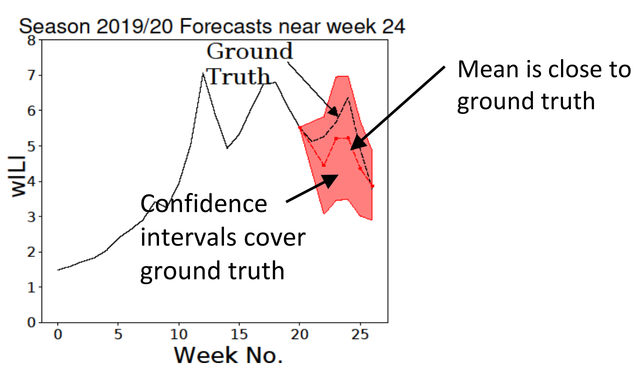
3. Forecast future incidence distribution (1-4 weeks in future)



Current flu forecasting models

- Most methods focus on point-predictions [Reich+ PNAS 2018]
- Other state-of-art probabilistic methods like Empirical Bayes [Brooks+ Comp Bio 2015], Delta Density [Brooks+ PLOS 2018], Gaussian Process [Zimmer+ NIPS 2020], EpiDeep [Adhikari+ KDD 2019]
  - Do not focus on producing **well-calibrated forecasts.**
  - Can't adapt or provide reliable forecast confidence on novel patterns





Kamarthi et.al 2021

Our Goal

- Develop deep probabilistic model for accurate and well-calibrated time-series forecasting
- Explainable forecasts from complex temporal similarities with historical season – also enables sound decision making

Outline

- Motivation
- Overall Idea & Approach
- EpiFNP Framework Details
- Experiments
- Conclusion

Deep Sequential models

- Leverage Neural models like GRU, LSTM, RNN, Epideep
  - Captures long term patterns
  - Widely successful for point-forecasts
- But …
  - Doesn’t learn probability density of prediction

Recent deep learning approaches for calibrated forecasting

- Bayesian Deep Learning [McKay Neuro. Comp 1992, Louizos+ ICML 2017]:
    - Difficult to set useful priors, intractable inference
- Deep ensembles [Lakshminarayan+ NIPS 2017]:
    - compute intensive
- Hard to estimate uncertainty very well [Kong+ ICML 2020]

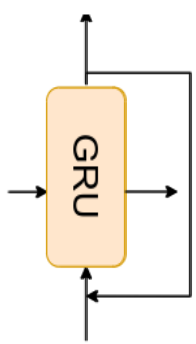
Georgia Tech

Non-parametric GP approach

- Gaussian process based non-parametric model
- Directly leverage similarity with training sequences as part of functional of distribution
  - Quantify uncertainty based on similarity with previously seen patterns
- But …
  - Need to capture complex long-term sequential patterns
  - Inefficient for high-dimensional data

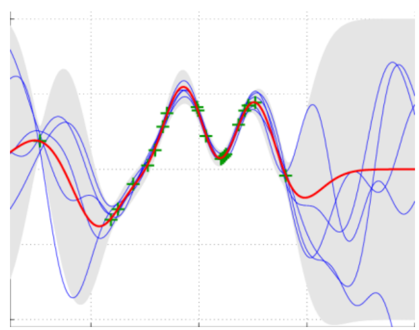
Our approach: Neural Gaussian Process

- Marry two approaches:
  - Leverage Deep sequential models to capture complex temporal patterns as low-dimensional representation
  - Use flexible non-parametric modelling to learn well-calibrated and accurate predictions



Deep Sequential Models

+



Gaussian Process Framework

=

Accurate and Well-calibrated neural forecasting model

Recent work

- Functional Neural Process [Louizos+ NeurIPS 2020]
  - Non-parametric modeling framework
  - Used static datasets
- Recurrent Neural Process [Qin+ 2020]
  - Based on Neural Process [Garnelo+ NeurIPS 2020]
  - Uses attention over training sequences

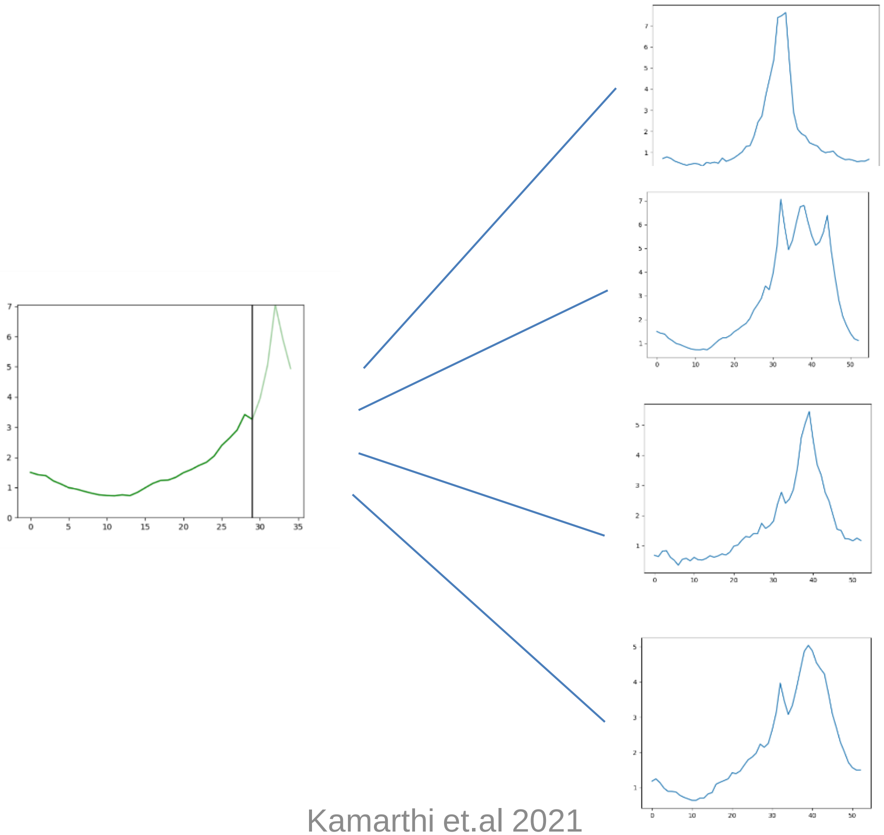
Our contribution: New FNP framework for sequential data leveraging complex stochastic latent correlations with training data to derive the predictive distribution

Outline

- Motivation
- Overall Idea & Approach
- EpiFNP Framework Details
- Experiments
- Conclusion

Training and Reference Set

- EpiFNP models predictive distribution based on similarity to sequences seen in past

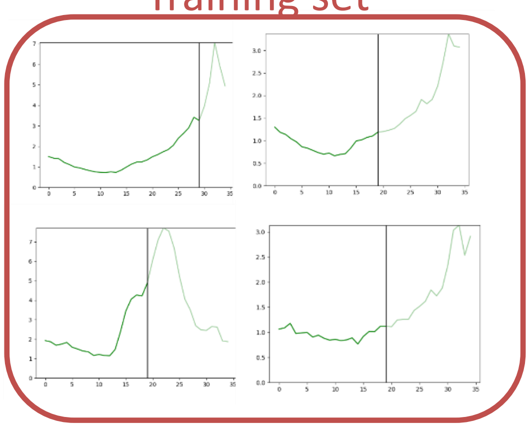


Kamarthi et.al 2021

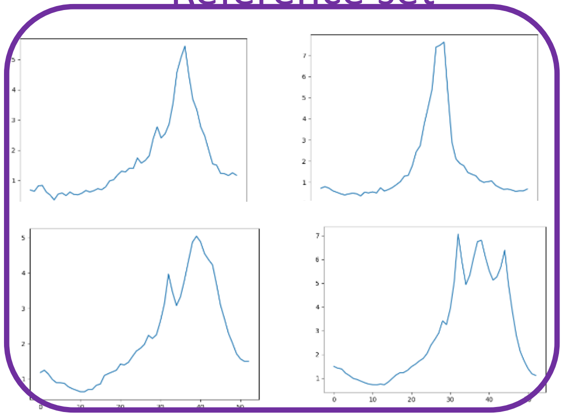
Training and Reference Set

- Sequences on which we model similarity – **Reference sequences/set**
- Input sequence on which we train/forecast future sequences – **Training sequences/set**

Training set



Reference set



Georgia Tech

EpiFNP – Overview (I)

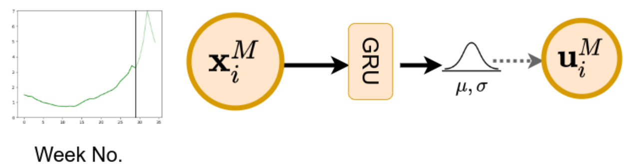
- Gaussian process based Neural Process architecture
  - Three components

Georgia Tech

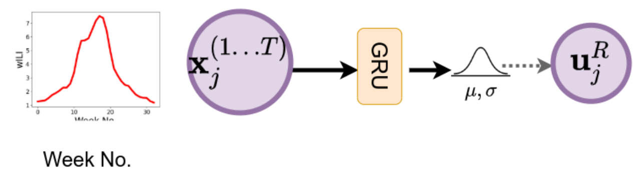
EpiFNP – Overview (II)

- Uses Deep Sequential Modules (GRU) to stochastically model sequences in latent space – **Probabilistic Neural Sequence Encoder**

Partial Sequence till current week

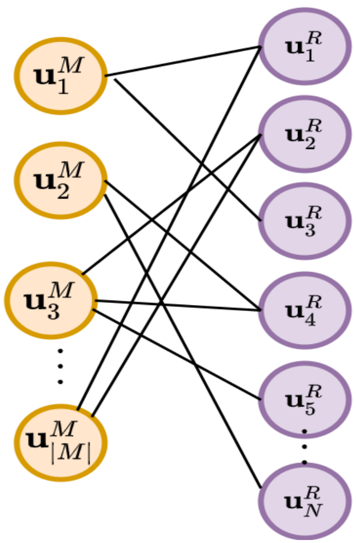


Full sequence of past seasons



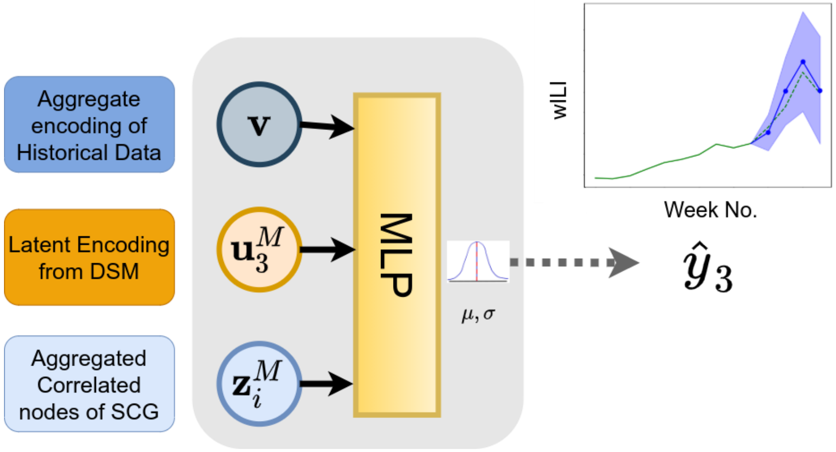
EpiFNP – Overview (III)

- Induces correlations with past sequences using similarity in latent space – **Stochastic Correlation Graph**



EpiFNP – Overview (IV)

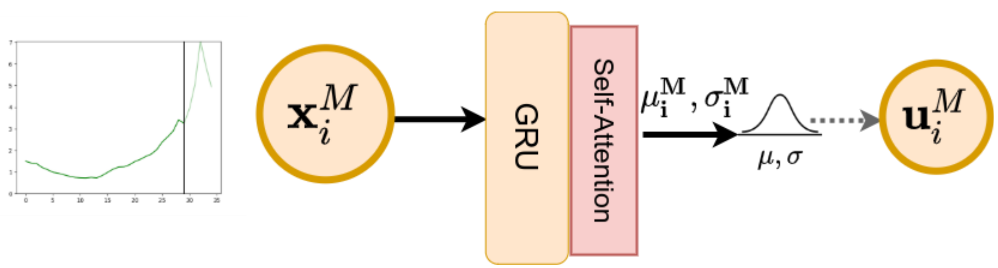
- Uses representations of correlated sequences from SCG, current season's representation to output predictive distribution - **Predictive Distribution Parameteriztion**



Probabilistic Neural Encoder

- Encodes current week and historical sequences into a latent vector distribution – Multi-variate Gaussian
- Quantify **uncertainty of input sequence**
- GRU with self-attention over hidden states to get a deterministic embedding of mean and variance
- Sample from the Gaussian as latent representation of sequence

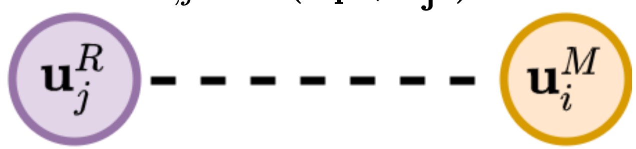
$$p(\mathbf{u_i^M}|\mathbf{x_i^M})$$
$$p(\mathbf{u_i^R}|\mathbf{x_i^{(1...T)}})$$



Stochastic Correlation Graph

- Leverages similarity between current (training/test set) and historical sequences (reference set) in latent space
- Samples edges proportional to similarity in latent space using RBF kernel

$$d_{i,j} = \kappa(\mathbf{u_i^M}, \mathbf{u_j^R})$$

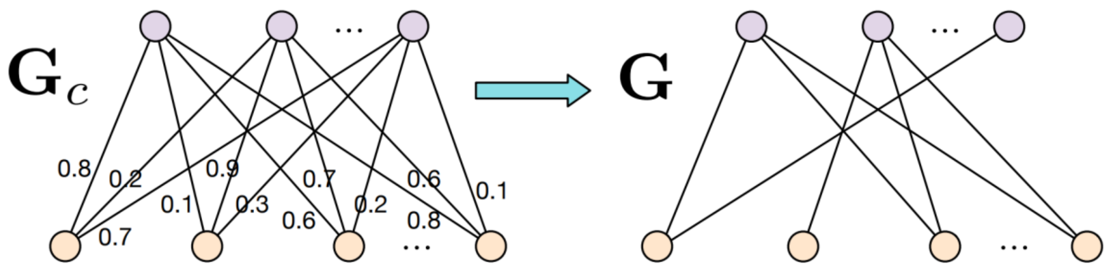


Connect an edge with probability $d_{i,j}$

Stochastic Correlation Graph

- Construct a bipartite network between historical sequences and training/input sequences

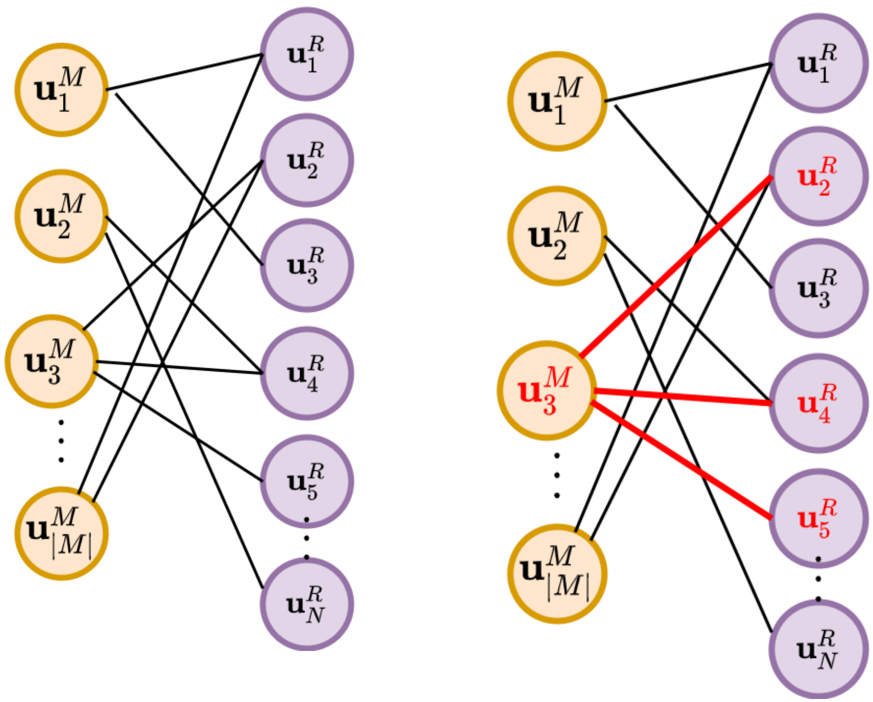
$$P(\mathbf{G} | \{d_{i,j}\}_{i \in M, j \in R})$$



Stochastic correlation graph

- Derive **local latent variable** from connected nodes of SCG

$$P(\mathbf{z_i^M} | \mathbf{G})$$

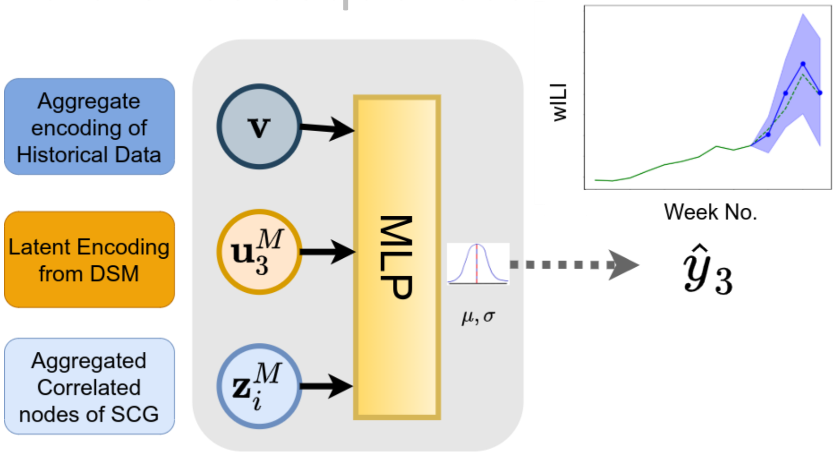


$$\mu(\mathbf{z}_3^M), \sigma(\mathbf{z}_3^M) = \sum_{j \in \{2,4,5\}} h(\mathbf{u_j^R})$$

$$\mathbf{z}_3^M \sim \mathcal{N}(\mu(\mathbf{z}_3^M), \sigma(\mathbf{z}_3^M))$$

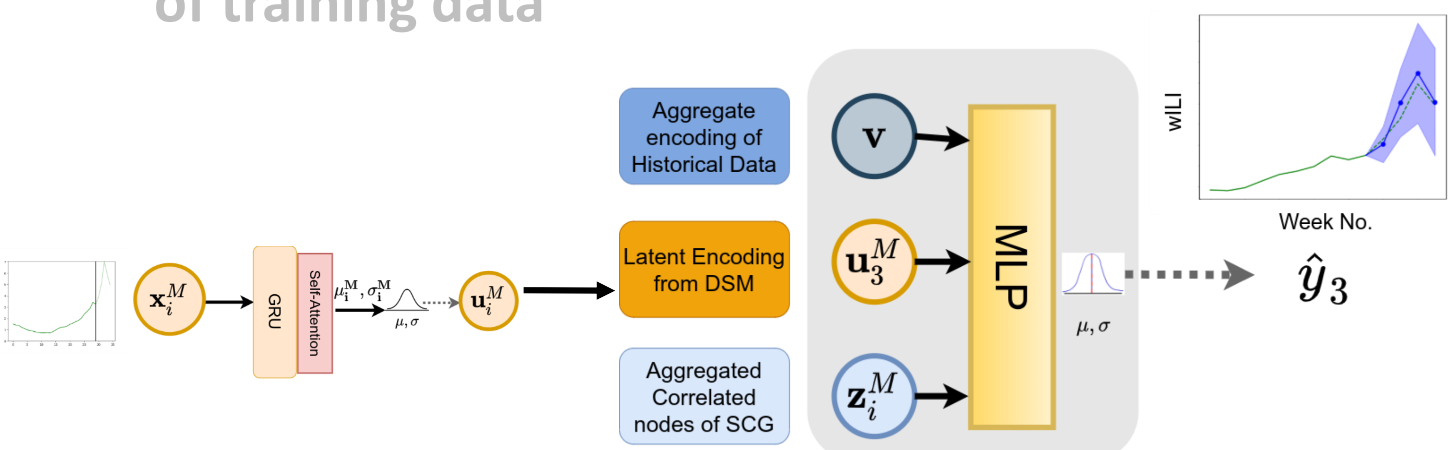
Deriving Prediction distribution

- Combines uncertainty from different perspectives to parameterize predictive distribution:
  - Sequence embedding distribution for current sequence – **input specific** temporal patterns and uncertainty
  - Local latent embedding from SCG – relation and uncertainty based on **correlation with training data**
  - Combination of all reference sequences – **global uncertainty of training data**



Deriving Prediction distribution

- Combines uncertainty from different perspectives to parameterize predictive distribution:
  - Sequence embedding distribution for current sequence – **input specific** temporal patterns and uncertainty
  - Local latent embedding from SCG – relation and uncertainty based on **correlation with training data**
  - Combination of all reference sequences – **global uncertainty of training data**

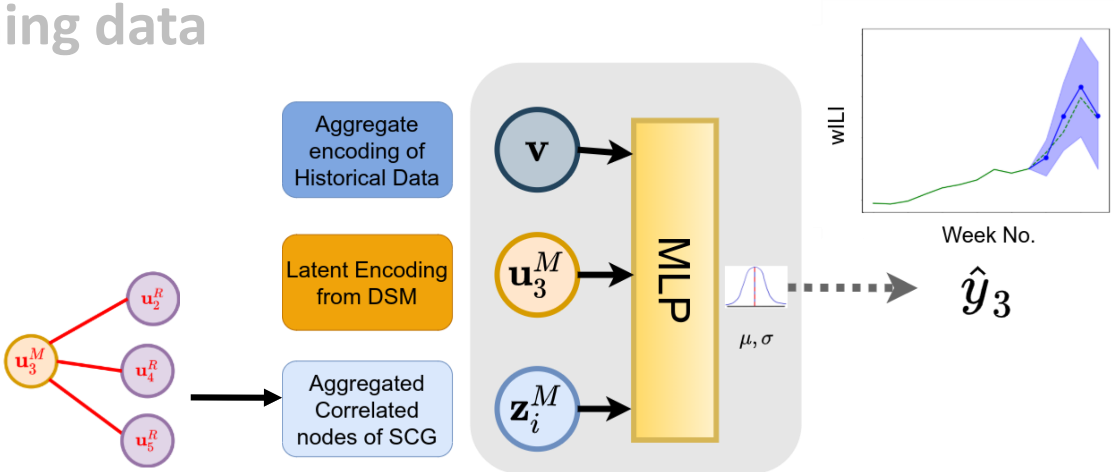


Kamarthi et.al 2021

Georgia Tech

Deriving Prediction distribution

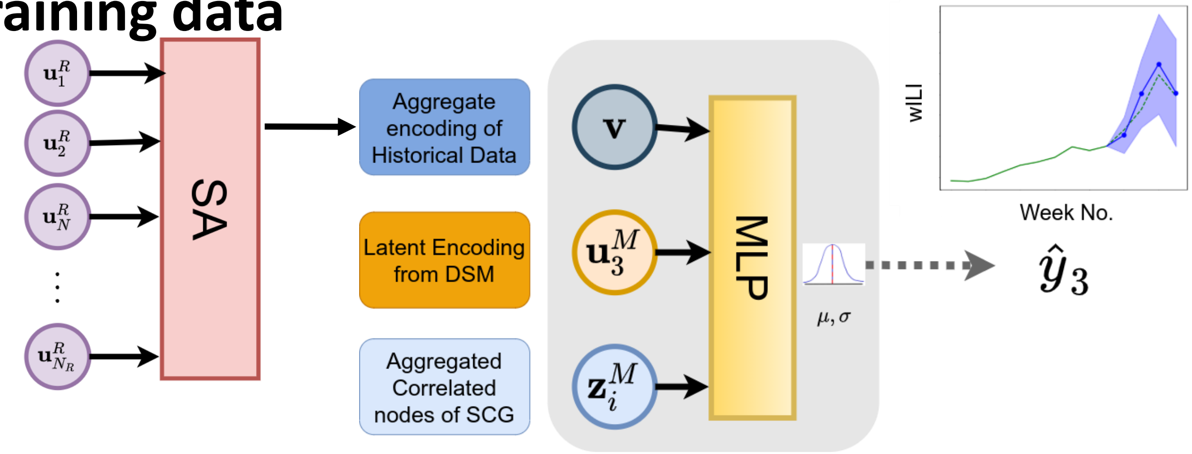
- Combines uncertainty from different perspectives to parameterize predictive distribution:
  - Sequence embedding distribution for current sequence – **input specific** temporal patterns and uncertainty
  - Local latent embedding from SCG – relation and uncertainty based on **correlation with training data**
  - Combination of all reference sequences – **global uncertainty of training data**



Kamarthi et.al 2021

Deriving Prediction distribution

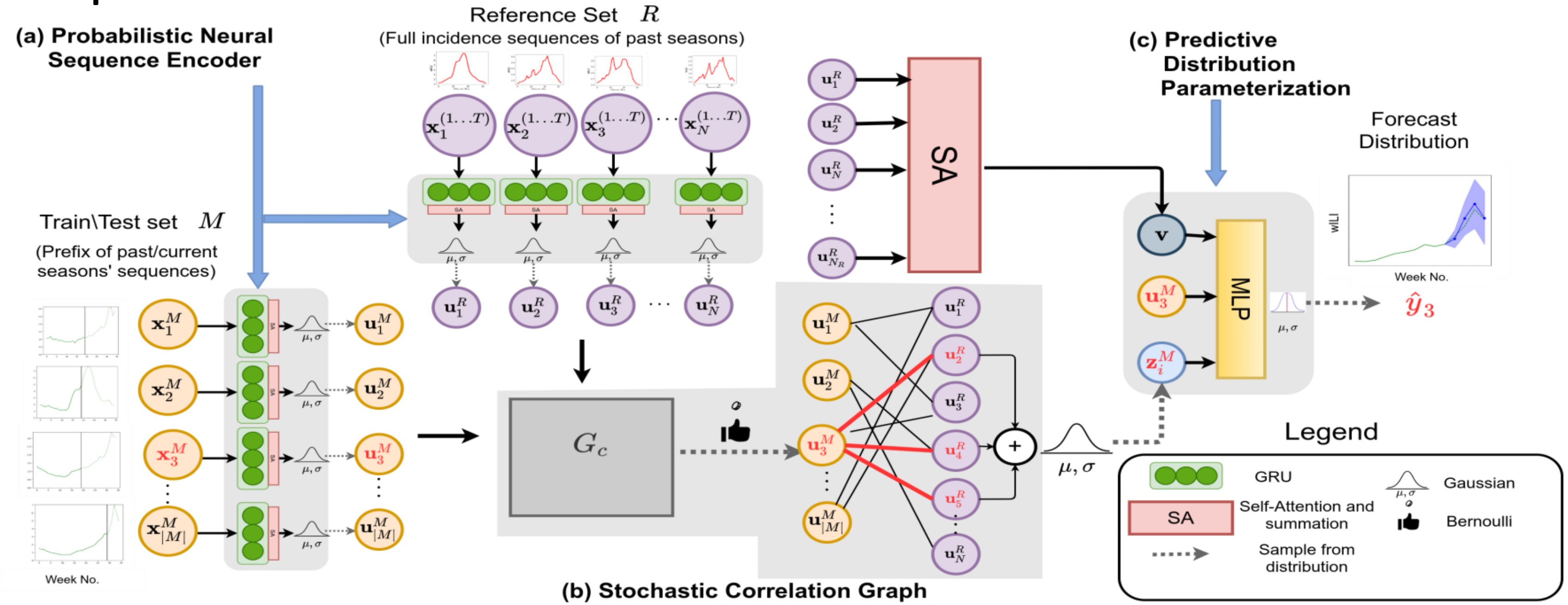
- Combines uncertainty from different perspectives to parameterize predictive distribution:
  - Sequence embedding distribution for current sequence – **input specific** temporal patterns and uncertainty
  - Local latent embedding from SCG – relation and uncertainty based on **correlation with training data**
  - Combination of all reference sequences – **global uncertainty of training data**



Kamarthi et.al 2021

EpiFNP training

- All components are trained end-to-end on training set of past data
- Input data for training formed from prefix of historical sequences



Variational Inference

- To overcome intractable marginalization over latent random variables:
- Model variational distribution of **local latent variable** for all sequences

  $\prod_{i \in M} [q(\mathbf{z_i^M} | \mathbf{x_i^M})]$ Approximates $\prod_{i \in M} \left[p(\mathbf{u_i^M} | \mathbf{x_i^M}) \left(\prod_{j} p(\mathbf{u_j^R} | \mathbf{x_j^R}) \right) p(\mathbf{G} | \mathbf{u_i^M}, \{\mathbf{u_j^R}\}_j) p(\mathbf{z_i} | \mathbf{G}) \right]$

- Variational ELBO loss used to update parameters via Stochastic Gradient Descent based training.

Georgia Tech

Outline

- Motivation
- Overall Idea & Approach
- EpiFNP Framework Details
- Experiments
- Conclusion

Baselines – Epidemic Forecasting

- Epidemiological Baselines: Previous widely used and SOTA epidemic forecasting baselines
  - SARIMA – Classical Time series forecasting model
  - Empirical Bayes (EB) – Won previous flu forecasting challenges
  - Delta Density (DD) – widely used top non-parametric model
  - EpiDeep (ED) – A top-performing deep Learning model that leverages sequence similarity
  - Gaussian Process (GP) – A recent top performing model

Baselines – General ML Time-series

- General Deep Probabilistic Baselines: Previous deep probabilistic methods for general sequence prediction tasks
  - Monte Carlo Dropout (MCDP) on GRU
  - Bayesian Neural Network (BNN)
  - Recurrent Neural Process (RNP) – Modification of Vanilla Neural Process on sequences

Evaluation metrics – Accuracy

- Accuracy metrics
  - Root Mean Squared Error (**RMSE**) [Adhikari+ KDD 2018]
  - Mean Absolute Percentage Error (**MAPE**) [Reich+ PNAS 2018]
  - Log Score (**LS**): widely used in epidemic forecasting literature [Reich+ PNAS 2018]. Measure Log likelihood of prediction in small interval around ground truth

Evaluation metrics - Calibration

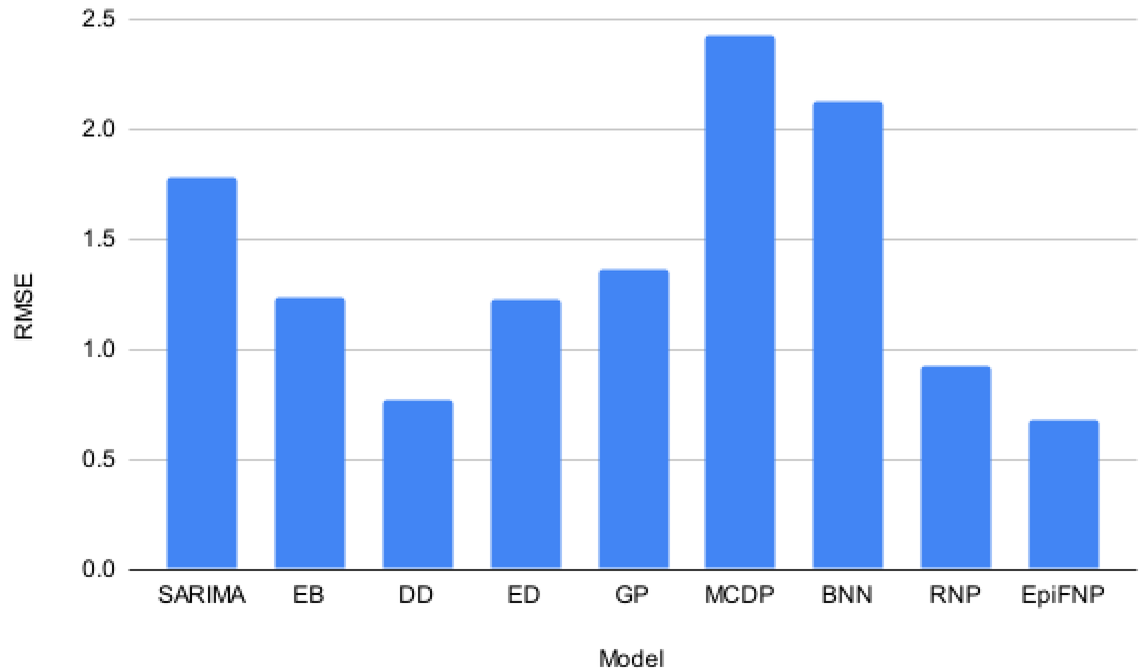
- Calibration Score (**CS**):
  - k(c) as fraction of ground truth predictions that fall within confidence level c of prediction distribution.
  - CS measures absolute difference between c and k(c) [Kuleshov+ ICML 2018]

$$CS = \int_0^1 |k(c) - c| dc$$

Georgia Tech

Obs 1: EpiFNP provides accurate point-predictions

- 13% and 42% better in RMSE and MAPE scores respectively



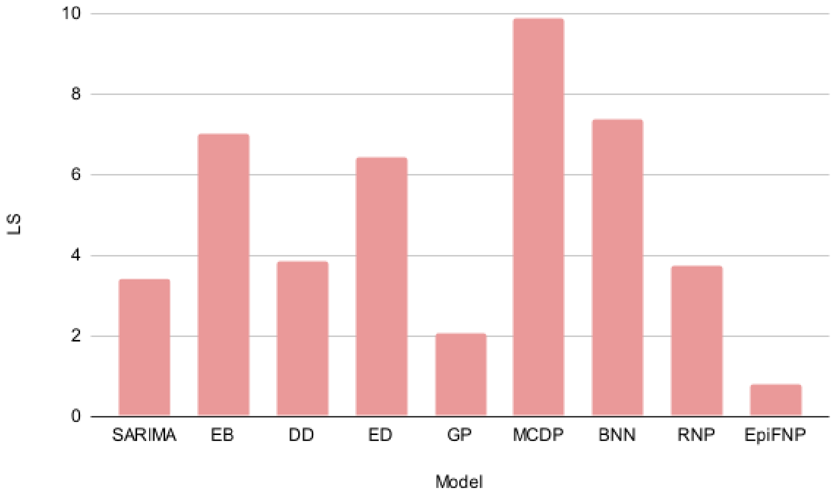
Avg. RMSE of EpiFNP and Baselines
(over US National and 8 HHS regions)

Kamarthi et.al 2021

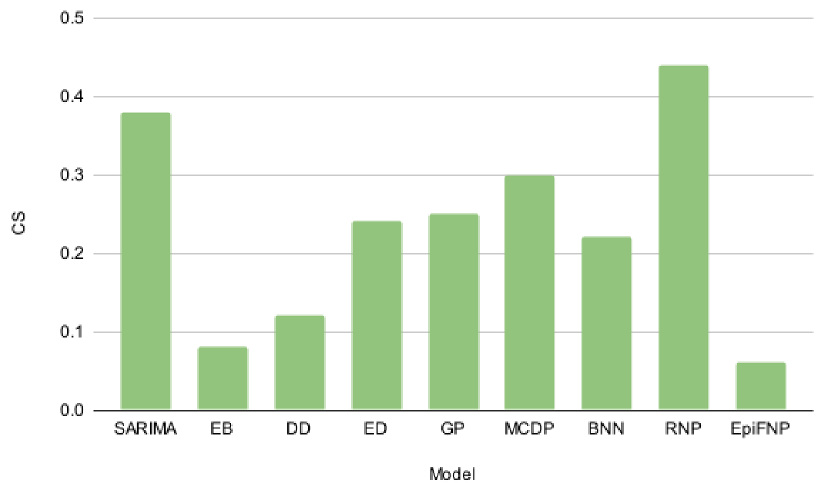
Georgia Tech

Obs 2: EpiFNP provides calibrated predictions

- 2.5 times better LS and 20% better CS compared to best baseline



Avg. LS of EpiFNP and Baselines

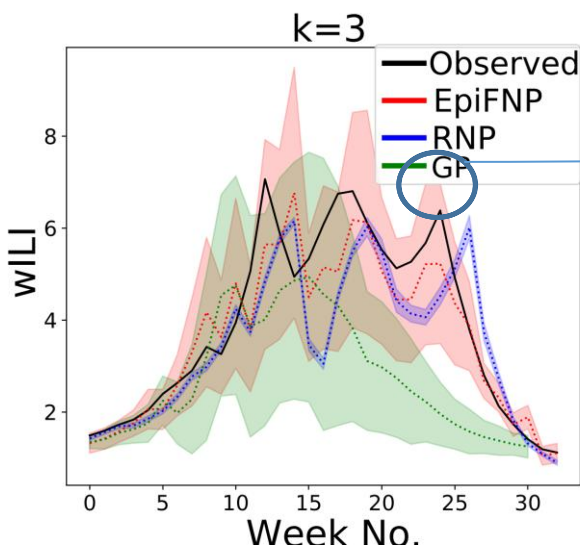


Avg. CS of EpiFNP and Baselines

Kamarthi et.al 2021

Obs 3: Adapting to novel patterns

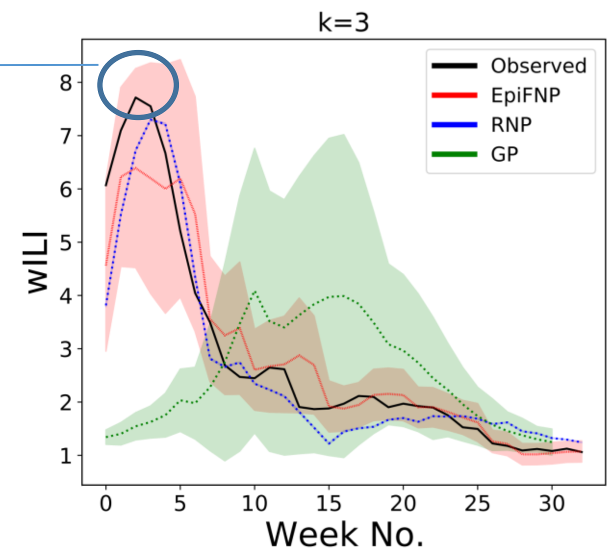
- Evaluate on novel H1N1 (2009/10) and Covid-19 seasons
- Captures unprecedented patterns
- 18-31% improvements in accuracy scores and 3.7 better



Predictions and uncertainty intervals in Covid-19 season

Novel early peak

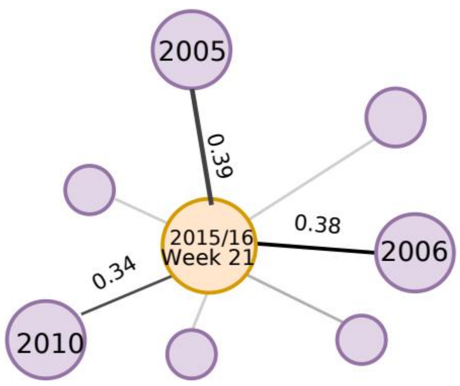
Novel 3rd peak



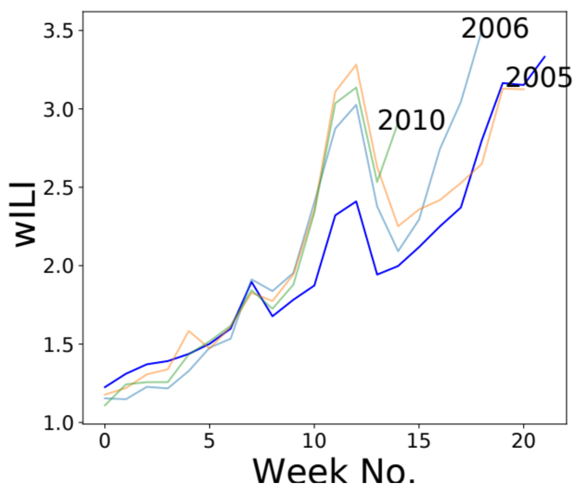
Predictions and uncertainty intervals in H1N1 season

Obs 4: EpiFNP chooses most similar historical seasons

- Observed most frequently sampled historical sequences from SCG
- Automatically identifies similar patterns from historical seasons



Most frequently sampled SCG neighbors of input sequence for Week 21 of 2015/16 season



Most similar seasons chosen by EpiFNP for Week 21 of 2015/16 season

Georgia Tech

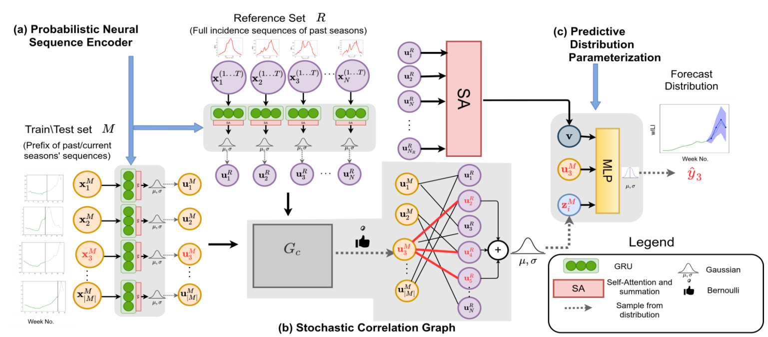
Outline

- Motivation
- Overall Idea & Approach
- EpiFNP Framework Details
- Experiments
- Conclusion

Conclusion

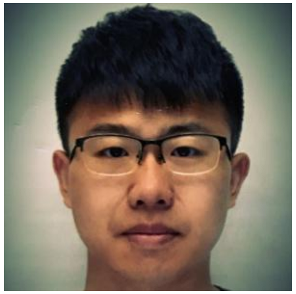
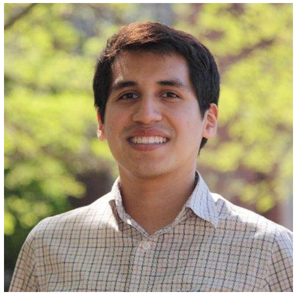
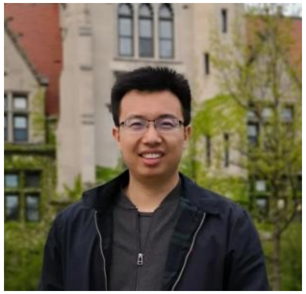
- Introduced **EpiFNP**: novel state-of-art deep probabilistic time series forecasting model for **accurate** and **well-calibrated** prediction
    - DSMs + NGP = accuracy and calibration
    - Flexible probabilistic modelling
    - Leverage similarity with complex patterns in training data
- Setting for flu forecasting: Consistently outperformed top baselines in accuracy and calibration by over 2.5x and 20% respectively
- **Adapted to novel patterns** and provided **explainable prediction** by identifying similar historical patterns and producing **reliable uncertainty**

For Future:
Capture and integrate sources of uncertainty from different data sources



Georgia Tech

Thank You!


Harsha Kamarthi | Lingkai Kong | Alexander Rodriguez | Chao Zhang | B. Aditya Prakash

Research supported in part by the National Science Foundation (Expeditions, IIS Uncertainty, IIS HAI, RAPID, CAREER, CCF, NRT), CDC MIND, ORNL, GTRI, faculty research awards from Facebook, Google and Amazon and funds/computing resources from Georgia Tech.

Email: hkamarthi3@gatech.edu

Code and Dataset: *https://github.com/AdityaLab/EpiFNP*