# 40 Years of…

# Leveraging Machine Learning for Network Traffic Forecasting

Author,
**Diane Onguetou, PhD**
Independent Consultant
onguetou@umich.edu

Co-Author,
**Achintha Maddumabandara, P. Eng.**
Principal Data Analyst
Rogers Communications

Co-Author & Presenter,
**Jeffrey Lee**
Manager, Data Acquisition & Reporting
Rogers Communications

SCTE CABLE-TEC EXPO 23
OCTOBER 16–19, 2023 • DENVER, CO
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

## Introduction

## Case study

- The Network
- The Dataset

## CAGR approach

- Global CAGR
- Global vs. local CAGR estimates
- Predictions from global vs. local CAGR

## Machine Learning

- Seasonal decomposition
- Time series forecasting models
- Time series vs. CAGR forecasting

## Key Insights

- Main improvements
- Challenges when applying ML
- Summary and recommendations

## Network Traffic Forecasting

- Traffic analysis
  [historical patterns, peak demand periods, seasonal variations]
- Growth projections
- Capacity planning
- Performance optimization
- Technology planning

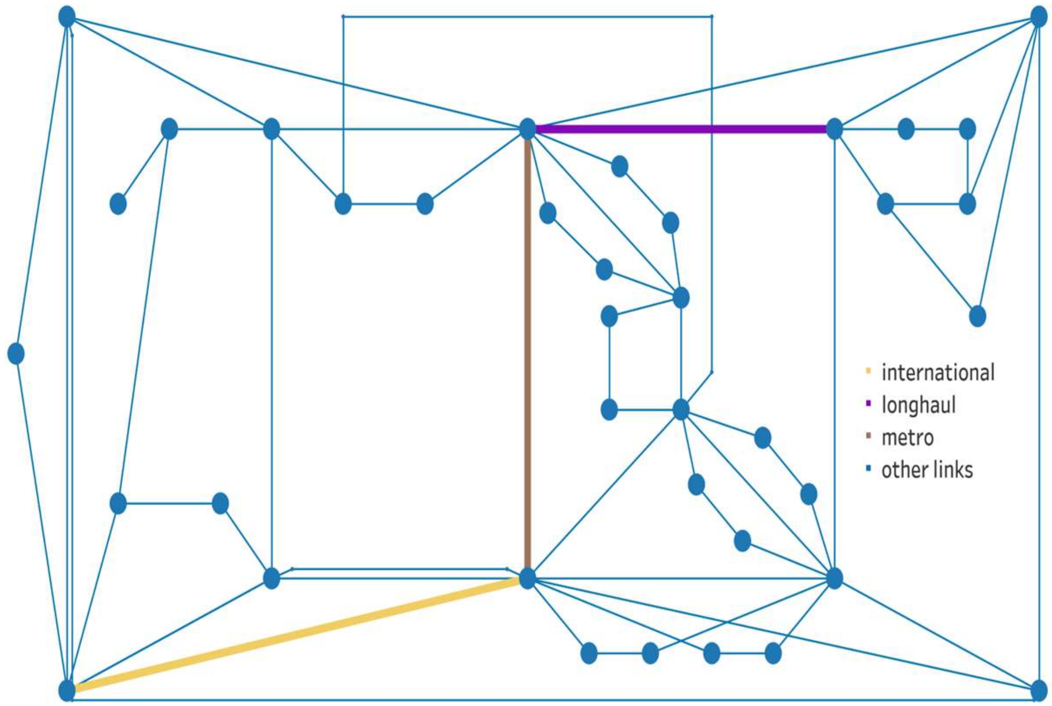
## Leveraging ML for Network Traffic Forecasting

- Captures complex patterns and dynamics in the data, leading to more effective resource management and improved network performance
- Automate iterative calculations and model attributes such as trends and seasonality, failure events, subsequent interactions between the primary and failover links, and network burst patterns
- Computational power to scale the analysis of multiple variables and high granularity

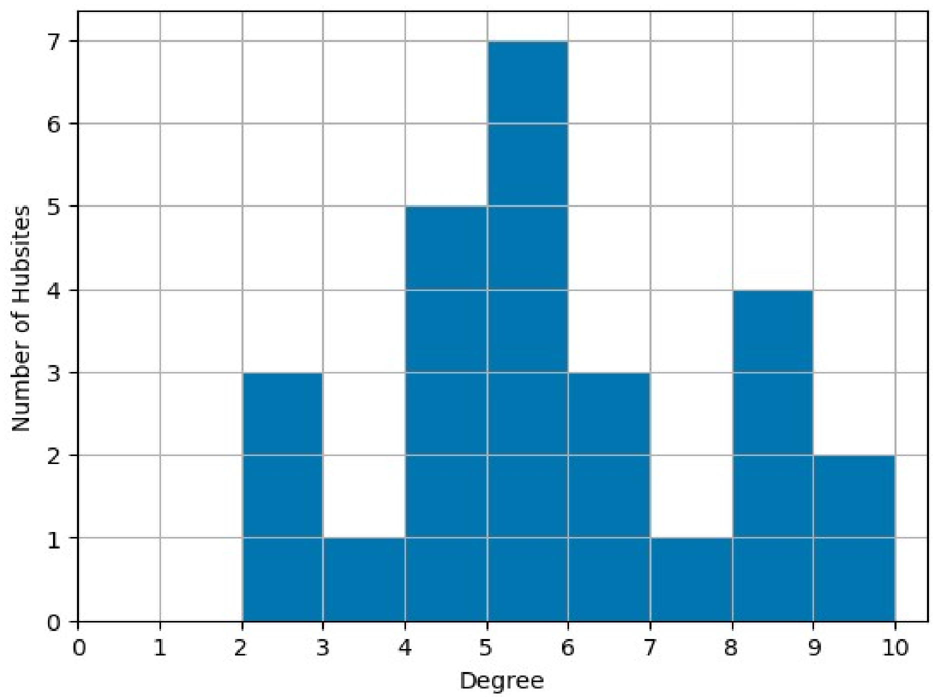
## Main Objective

- Consider a large service provider network and sample traffic for 5 consecutive years
- Analyzes traffic patterns in the first four years and use insights and findings to predict traffic in the fifth year

## Logical View

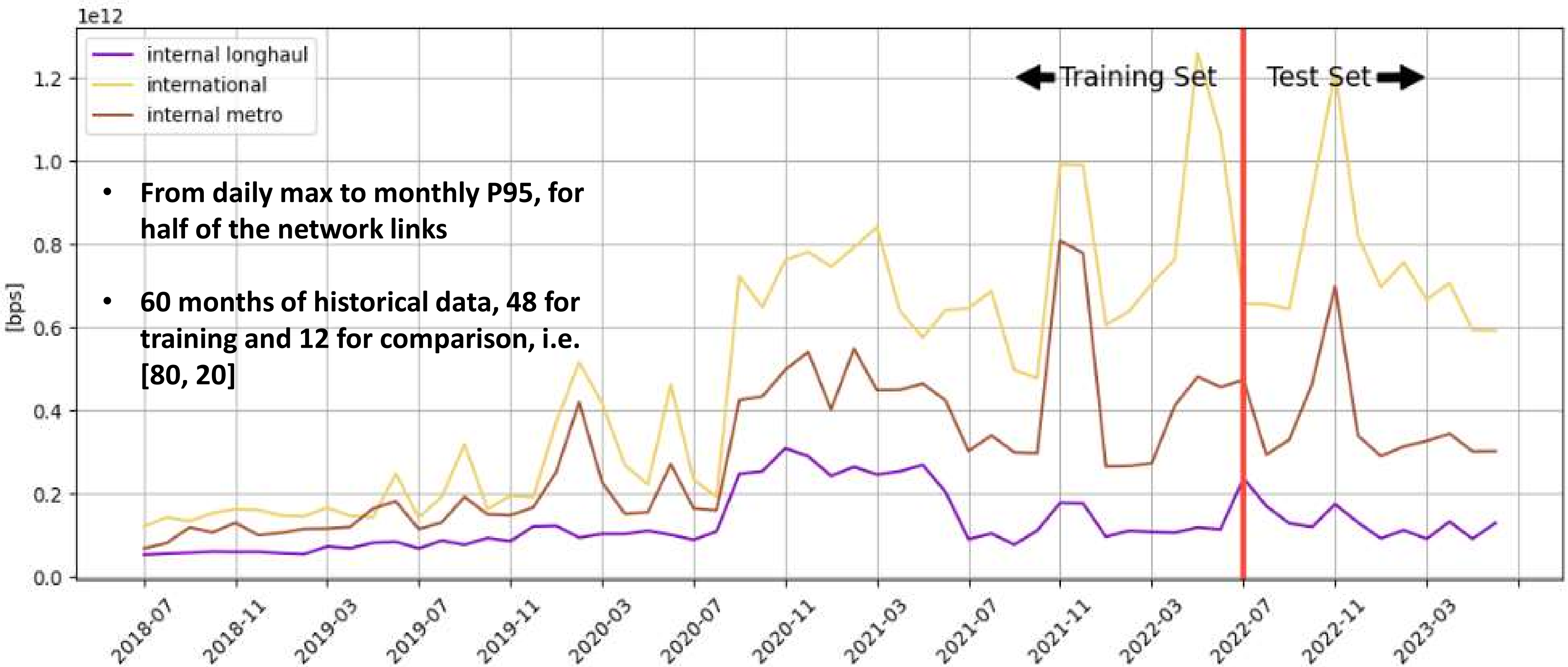



## Distribution of Connectivity at a Hubsite



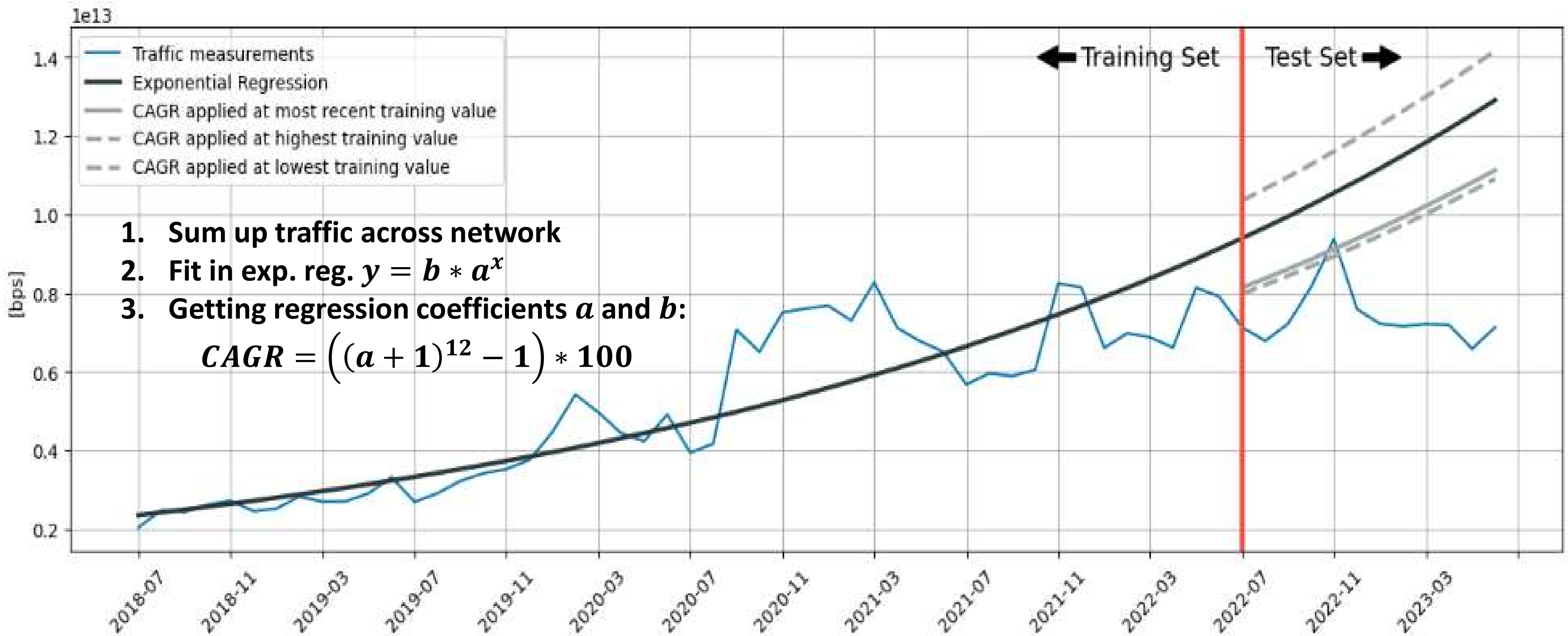
# CASE STUDY – The Dataset

- **From daily max to monthly P95, for half of the network links**
- **60 months of historical data, 48 for training and 12 for comparison, i.e. [80, 20]**



SCTE CABLE-TEC EXPO 23
OCTOBER 16–19, 2023 • DENVER, CO

# How to determine global CAGR?




1. **Sum up traffic across network**
2. **Fit in exp. reg.** $y = b * a^x$
3. **Getting regression coefficients** $a$ **and** $b$:

$$CAGR = \left((a+1)^{12} - 1\right) * 100$$

SCTE CABLE-TEC EXPO 23 OCTOBER 16–19, 2023 • DENVER, CO

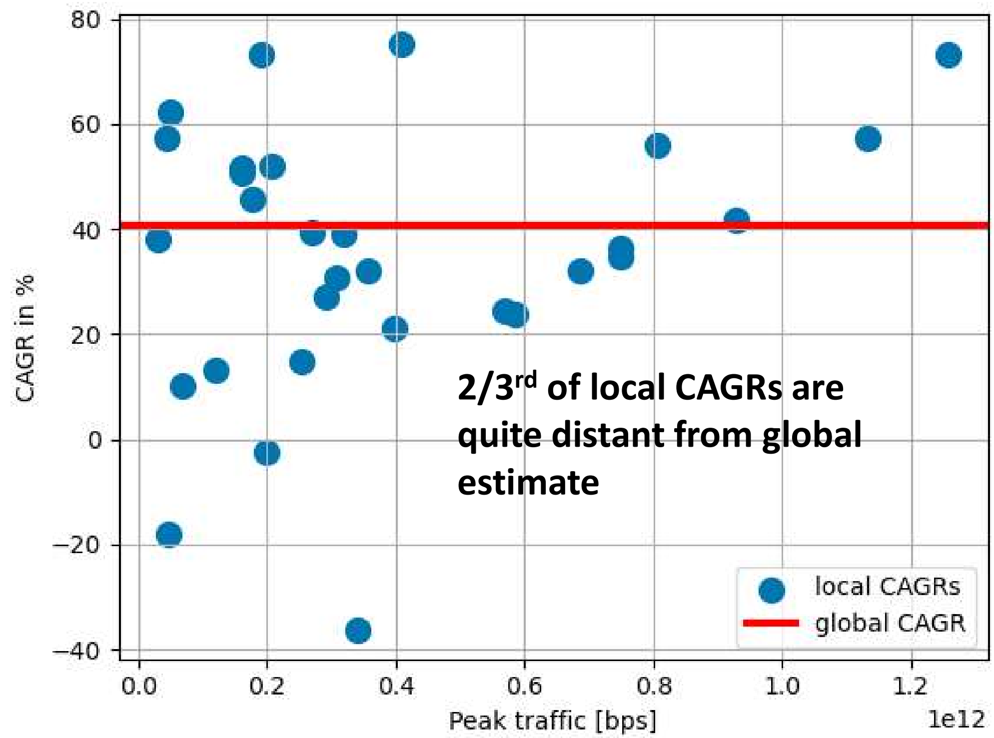
## Limitations

i. Logic behind the choice of reference month
   - Most recent measurement on hands
   - vs. min/average/max over past 12 months

ii. CAGR model accuracy

iii. Applying a total-traffic CAGR to individual contributors
   - Global vs. local CAGR estimates
   - Predictions from global vs. local CAGR approaches (next slide)

## Global vs. Local CAGR



SCTE CABLE-TEC EXPO 23
OCTOBER 16–19, 2023 • DENVER, CO

# CAGR is sensitive to period choice.

## Seasonal Decomposition (STL)

- Breaks down a time series data into its fundamental components (i.e. trend, seasonality, and residual) to better understand underlying patterns

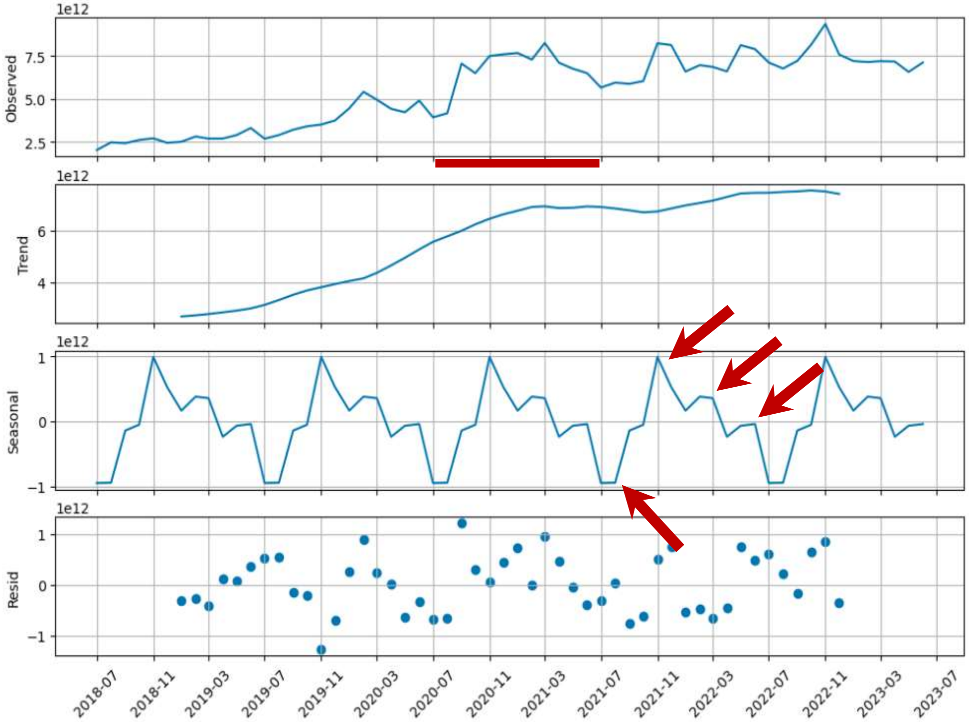
## Seasonal autoregressive integrated moving average (SARIMA)

- Combines autoregressive, differencing, and moving average components along with seasonal components to handle data with both non-seasonal and seasonal patterns

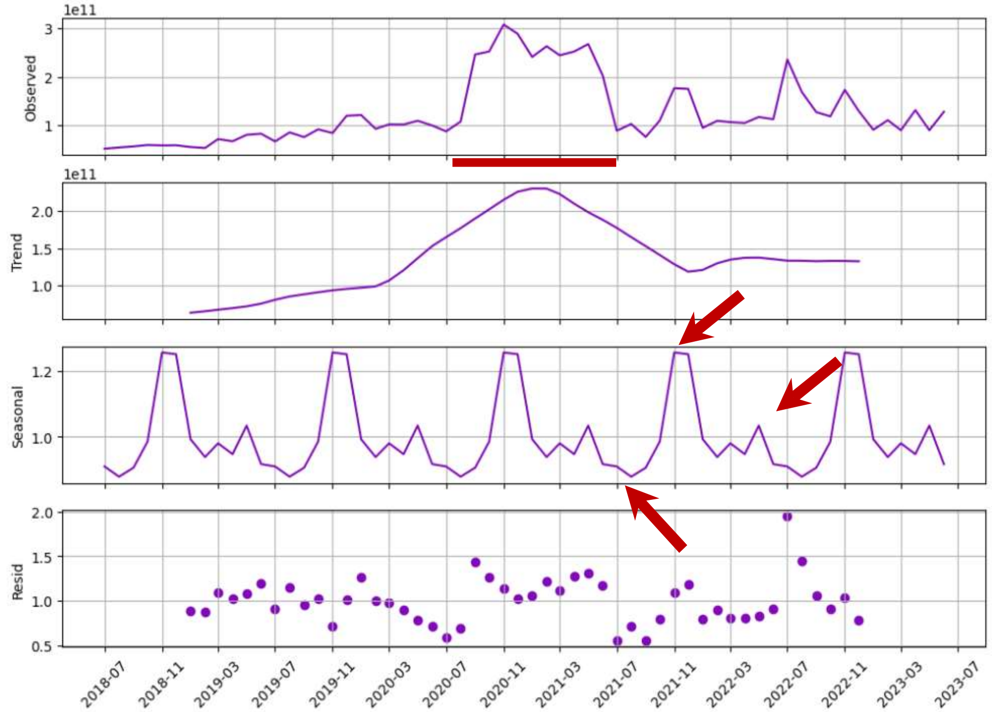
## Holt-Winters/Exponential Smoothing (ETS)

- Uses weighted average of past observations to predict future values, considering trend, seasonality, and level components to make accurate predictions for time series data

## Total Traffic (additive decomposition)



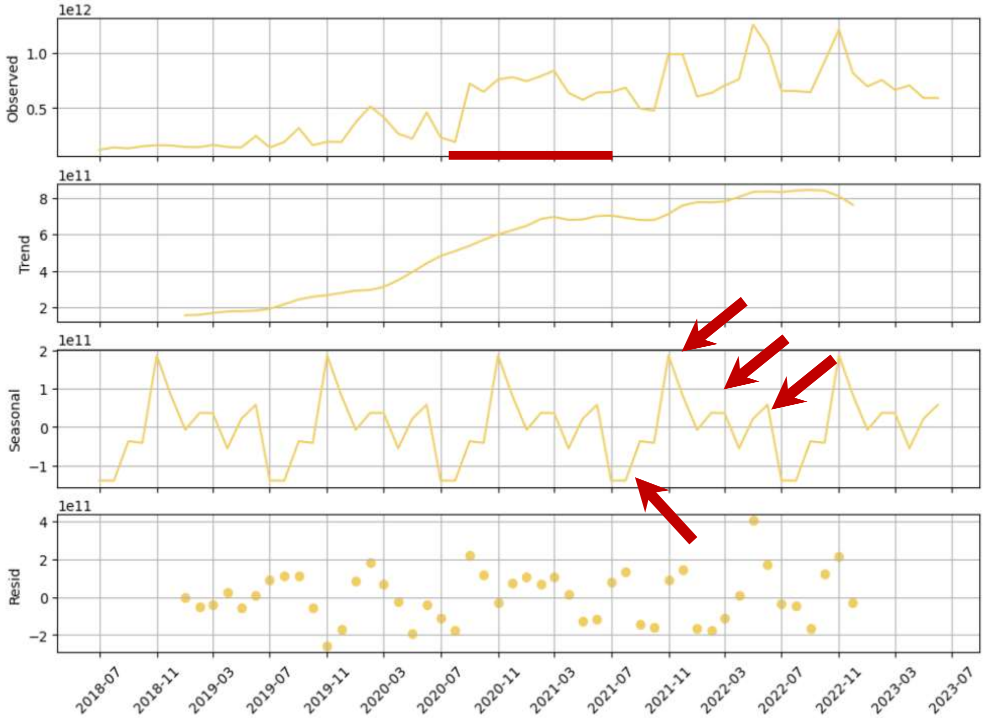

## Example of Internal Longhaul Link (multiplicative)



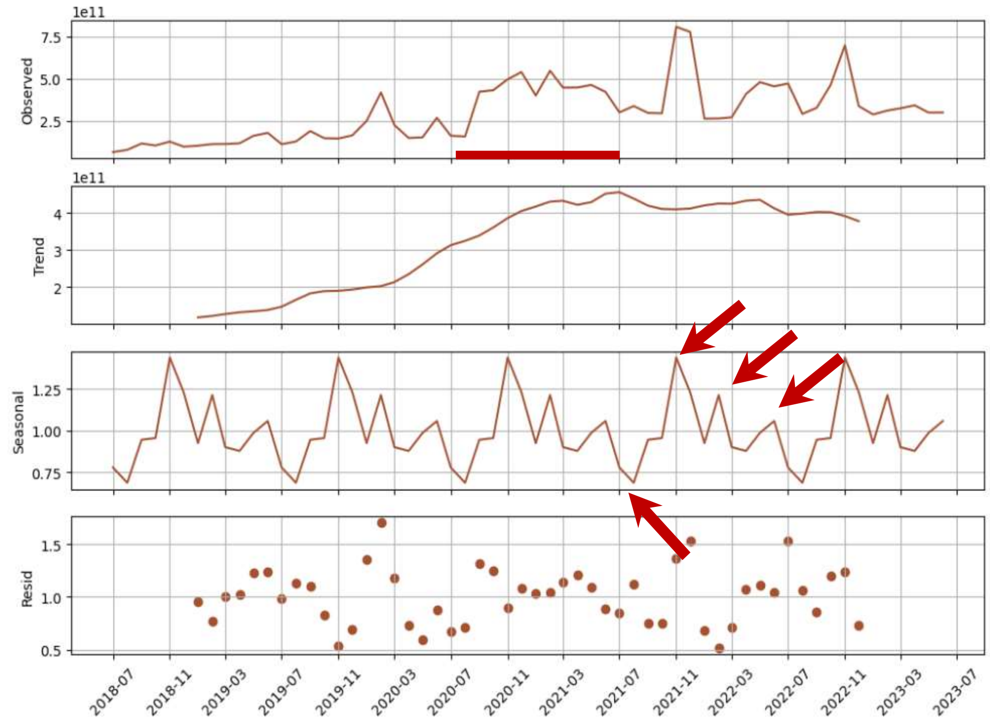

 

## ML | Seasonal Decomposition

### Example of International Link (additive)

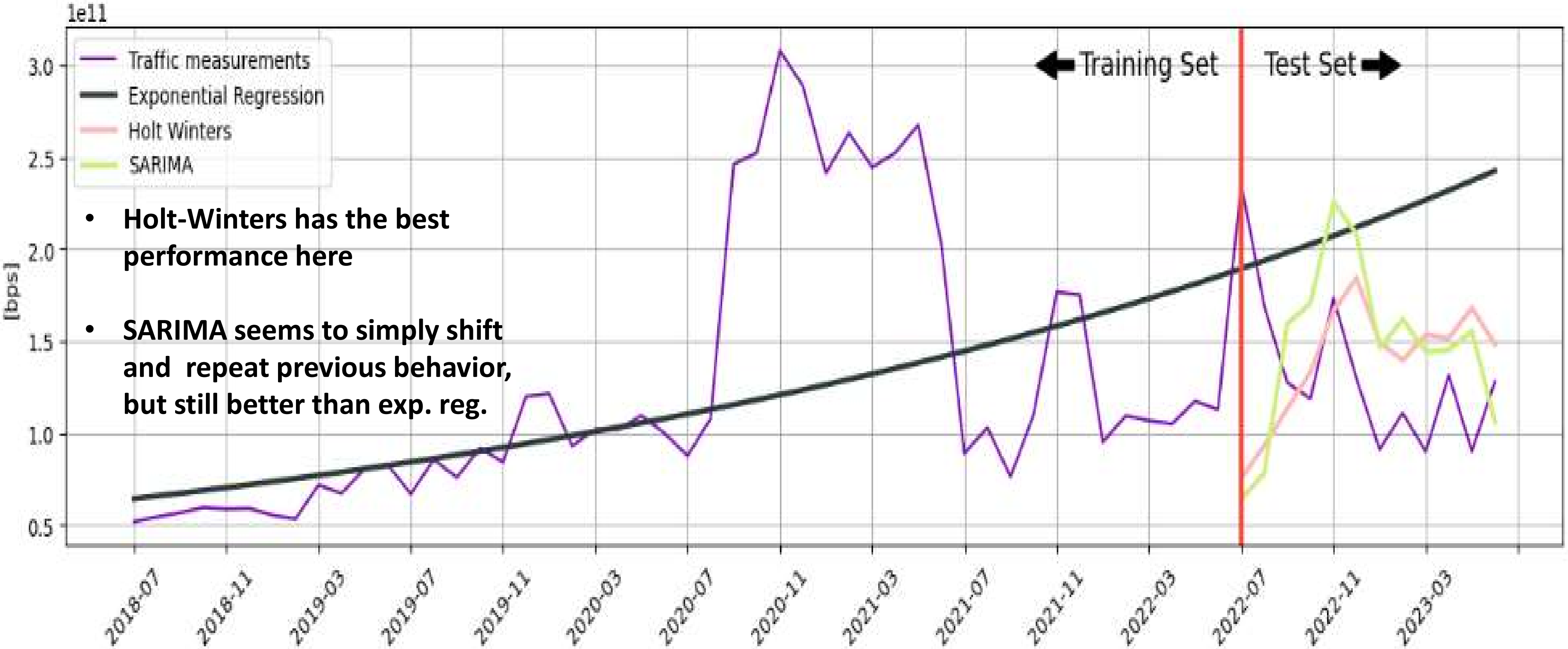



### Example of Internal Metro Link (multiplicative)



# ML | Time Series Forecasting Models [longhaul link]



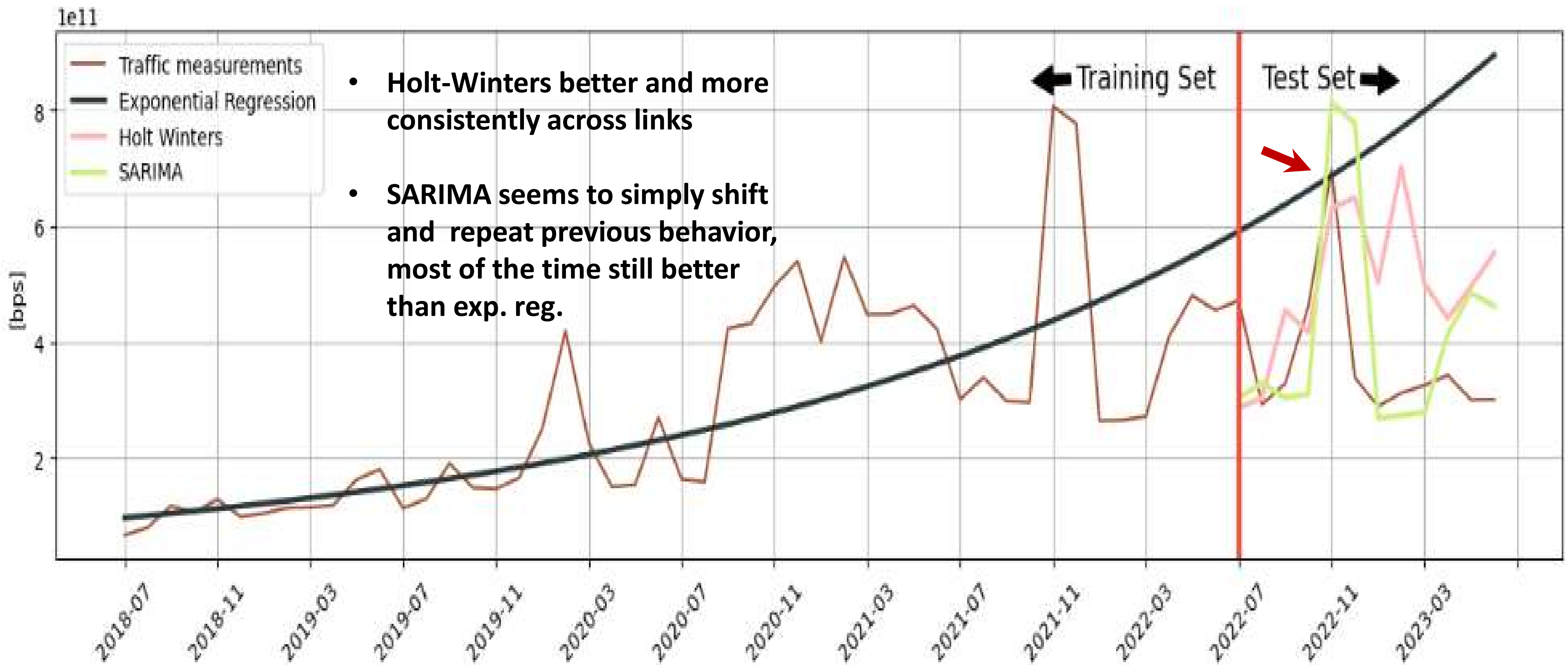



- **Holt-Winters has the best performance here**
- **SARIMA seems to simply shift and repeat previous behavior, but still better than exp. reg.**

# ML | Time Series Forecasting Models [international link]

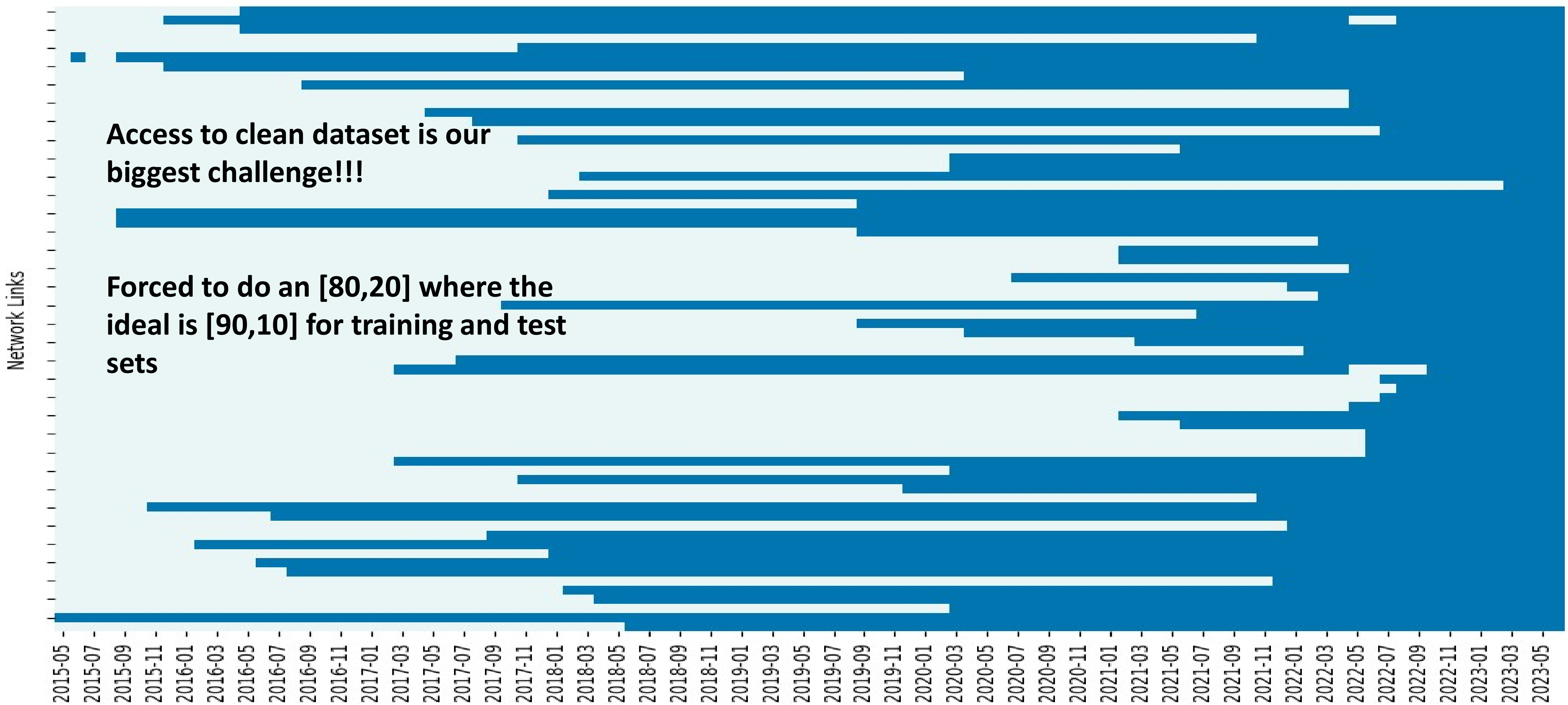

- **Holt-Winters and SARIMA seem to show an equivalent performance here**
- **Exp. reg. still has the worst performance**

# ML | Time Series Forecasting Models [metro link]




- **Holt-Winters better and more consistently across links**
- **SARIMA seems to simply shift and repeat previous behavior, most of the time still better than exp. reg.**

# CHALLENGES WHEN ADOPTING ML-BASED TRAFFIC MODELLING




**Access to clean dataset is our biggest challenge!!!**

**Forced to do an [80,20] where the ideal is [90,10] for training and test sets**

SCTE CABLE-TEC EXPO 23
OCTOBER 16–19, 2023 • DENVER, CO

# Good practices

- Invest in quality datasets
- Choose models appropriate to your needs
- Re-evaluate models in a regular basis and track your progress

# If CAGR simplicity is still appealing and fits current needs

- Plenty of tools allow to easily build ML pipeline that can support distributed CAGR
- Build a strategy behind the reference value to which the growth factor is applied
- CAGR is tied to time series trend and does not consider seasonality that is more likely to be present in your data

# Beyond CAGR…

- Do not be intimidated by the term machine learning.
- Python and R languages offer appropriate libraries that facilitate ML model training and testing
- ML proposes simple and effective models for seasonal decomposition of time series
- ML time series models capture both trend and seasonality for better forecasting
- SARIMA requires parameter tuning to perform
- Holt-Winters is simple enough and shows better results for our use case

Thank you to my co-author:
**Achintha Maddumabandara, P. Eng.**
Principal Data Analyst
Rogers Communications


**Diane Onguetou, PhD**

Independent Consultant
onguetou@umich.edu