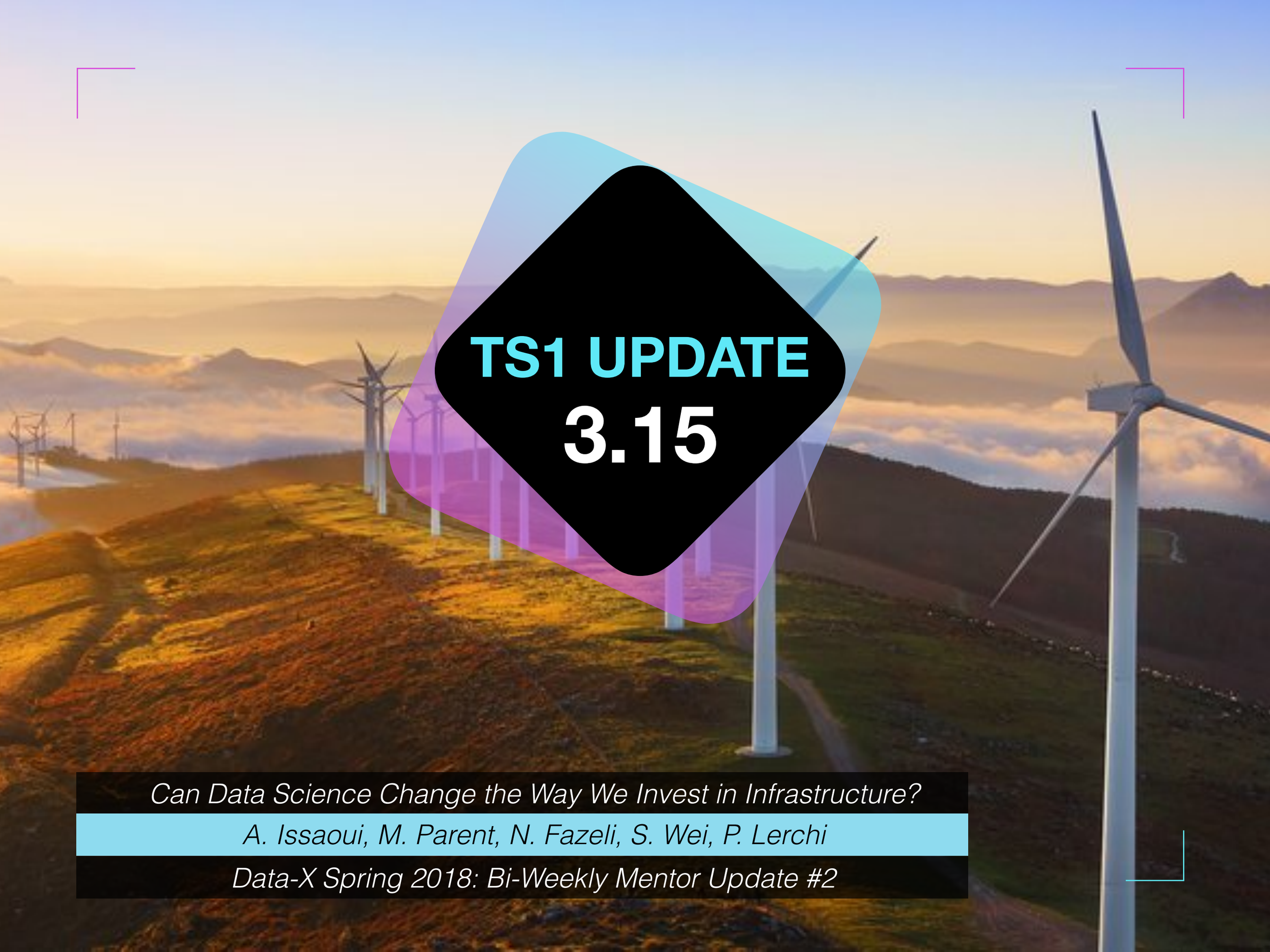


TS1 UPDATE 3.15

Can Data Science Change the Way We Invest in Infrastructure?

A. Issaoui, M. Parent, N. Fazeli, S. Wei, P. Lerchi

Data-X Spring 2018: Bi-Weekly Mentor Update #2

PROGRESS REPORT

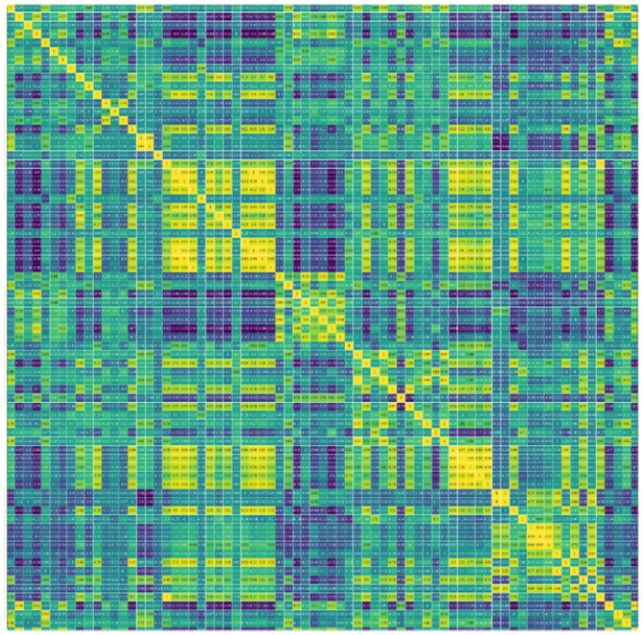
PHASE ONE OBJECTIVES

REVIEW & UNDERSTAND LAST YEAR'S RESULT

BASIC ARIMA TIME SERIES

HANDLING DATA SETS

What have we accomplished thus far?



Last year's correlation matrix

SKILLS LEARNED

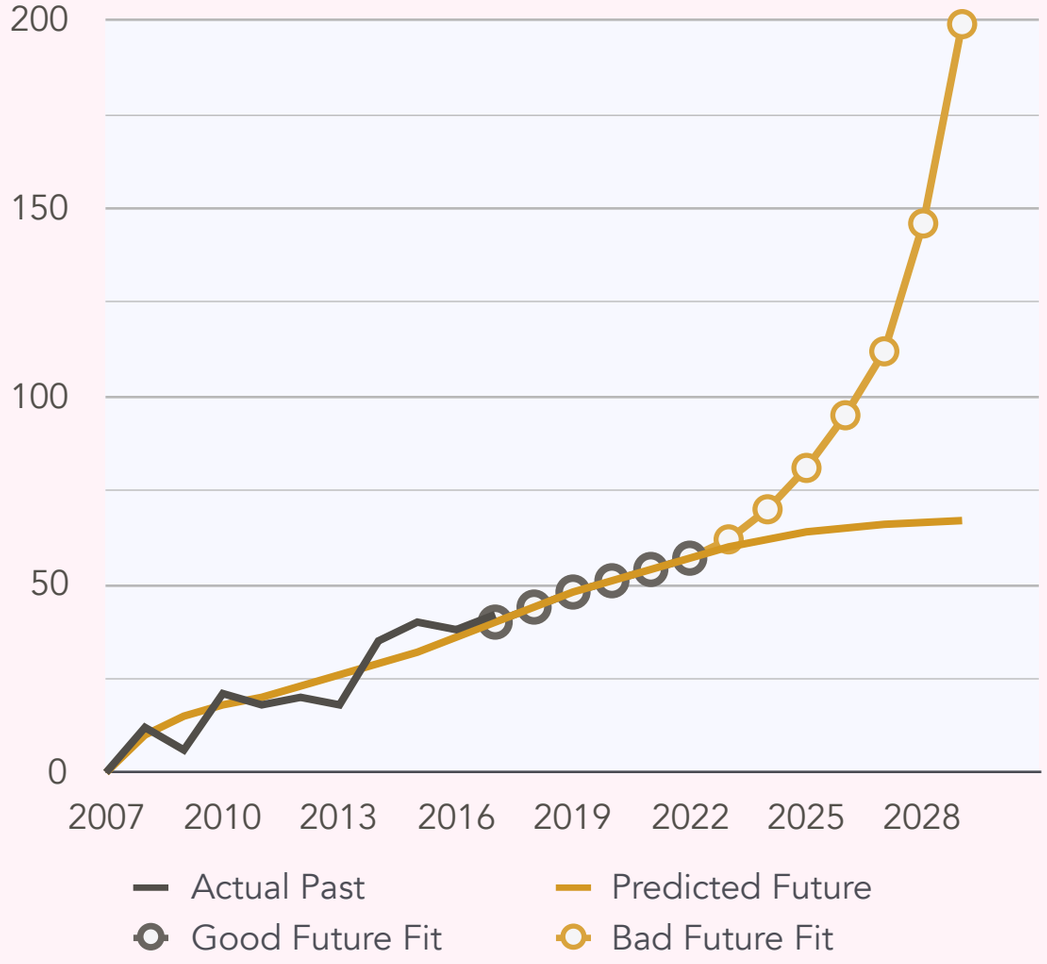
CHECK FOR STATIONARITY

DE-TRENDING DATA

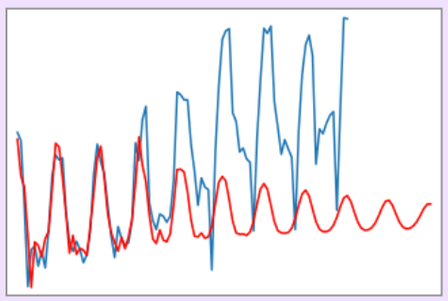
FORECASTING AND RE-TRENDING

In phase one, we focused on getting acquainted with the foundation of last year's project. A summary of our time series coding progress can be found in this notebook, prepared by Shaojin.

FORECASTING



LAST YEAR'S MODEL: DIVERGENCE AFTER THREE YEARS



PROBLEM:

TRADITIONAL TIME SERIES METHODS DO NOT ACCOUNT FOR MARKET DISCONTINUITY

- All data sets are inherently historical
- Can't predict disruptive changes
- Model breaks in the long term

PROPOSAL:

PAIR USER'S INDUSTRY KNOWLEDGE WITH OUR FEATURE LIBRARY TO CREATE A FLEXIBLE FORECASTING PLATFORM

EXAMPLE:

You are a car manufacturer in 1995. What sort of tool would generate reliable 2015 market predictions if there is no existing data about electric vehicles?

OUR IDEAS

How to get around predicting the unpredictable

01 ENABLE CUSTOM FEATURE SELECTION

- ✓ Incorporates intuition
- ✓ Can be time variant
- ✓ Easy user interface
- ✓ Must be paired with performance feedback

02 USER-GENERATED FORECASTING

- ✓ Generates future data
- ✓ More user involvement
- ✓ Rapid analysis
- ✓ Can include multiple future outcomes

03 ADDITIONAL TOOLS FOR ANALYSIS

- ✓ Decision Trees
- ✓ Monte Carlo Simulation
- ✓ Random Forest
- ✓ Pattern Matching

Our goal with this project is NOT to create one trend line that can perfectly forecast the future. Instead, we are trying to build a platform that uses data science to better qualify individual user predictions on future energy markets. The final product should allow anyone with industry-specific experience to create a tailored projection that varies from our baseline ARIMA model.

NEXT STEPS

What should be completed by spring break?

One

FINALIZE APPROACH

In order to set technical objectives, our team needs to confirm that our way of addressing the problem is valid. Will infrastructure investors stand to benefit from using such a tool?

Two

FEATURE COORDINATION

The flexibility of our system depends entirely on the features we decide to track. We will work with the other two groups to understand how important variables influence one another.

Three

INDUSTRY OUTREACH

We are currently in contact with a microeconomics professor at Haas, as well as a data scientist from PG&E. Additional contacts are much appreciated - we need all the help we can get!

THANKS FOR LISTENING!

Your feedback is greatly appreciated

Ahmed Issaoui
Negi Fazeli
Marie Parent
Shaojin Wei
Patrick Lerchi