Temporal Disaggregation of U.S. State Natural Gas Data

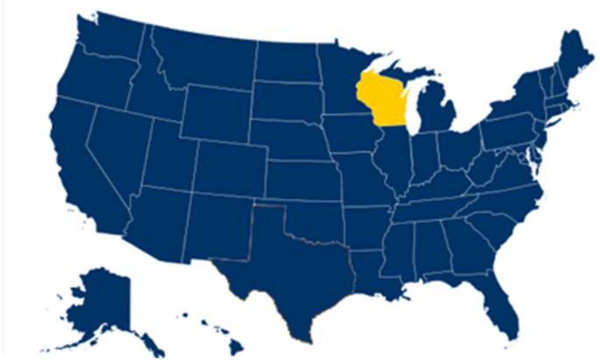
International Symposium on Forecasting 2021
Colin Quinn
Richard Povinelli

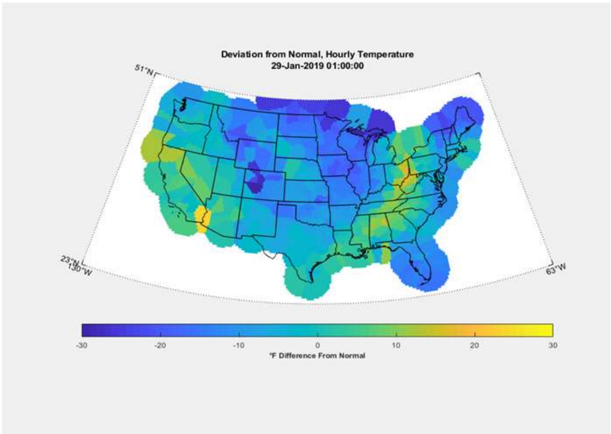
6/28/2021

Speaker Introduction

- Milwaukee, Wisconsin
  - Marquette University
    - Computer Science, Applied Statistics
    - GasDay Laboratory

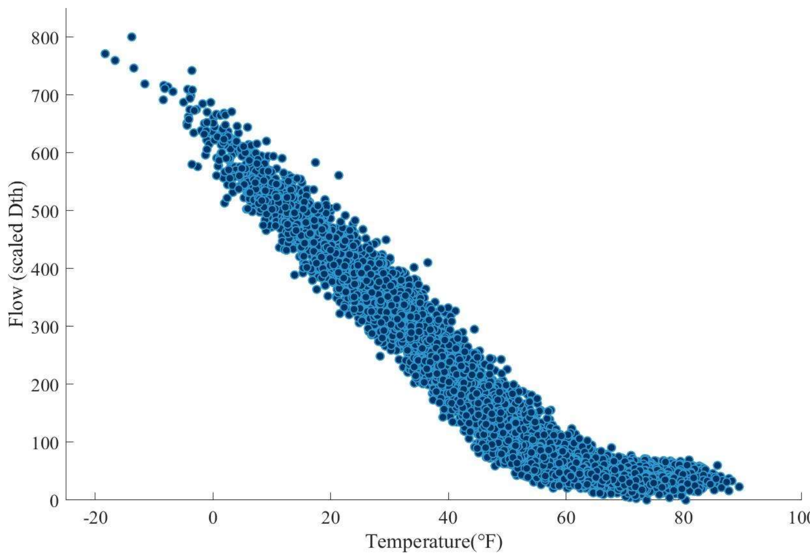
- Research domain: natural gas load forecasting
  - Disaggregation of natural gas demand series





Agenda

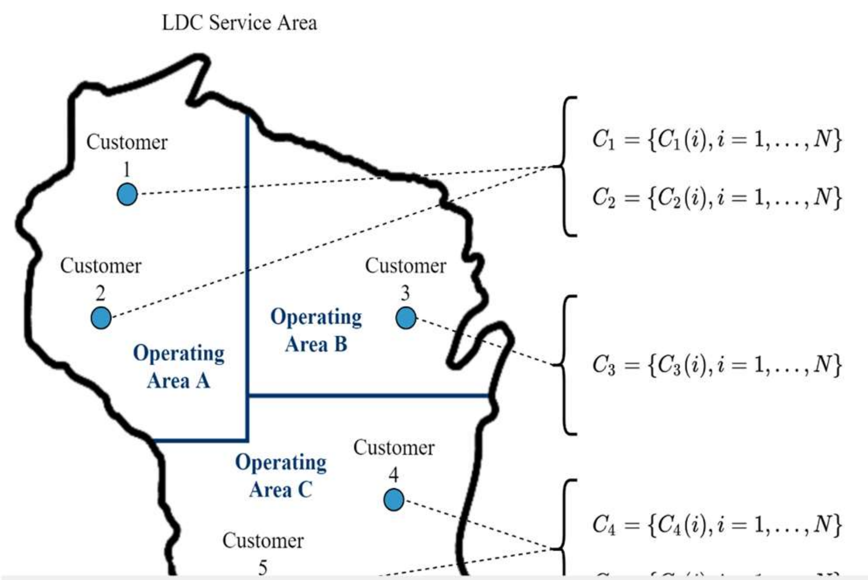
- Introduction to natural gas consumption, demand, and forecasting
  - Problem statement
    - Load vs demand
    - Consumption drivers
- Natural gas data
  - Industry evolution
    - Data availability and quality
    - Smart Meter data
- Methods
  - Disaggregation
    - Applied:
      - Flow shifting algorithm
      - Time series reconstruction algorithm



Natural gas consumption, demand, and forecasting

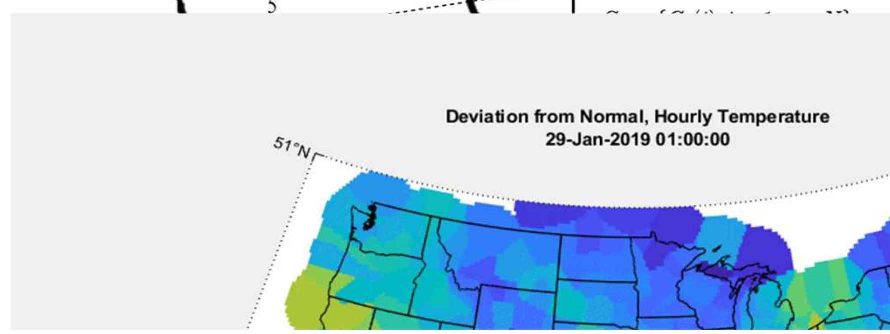
- Natural gas is a fossil energy source extracted for sale and consumption
  - Residential, commercial, and industrial uses

$$C = \{\, C(i), i = 1, \dots, N\}.$$

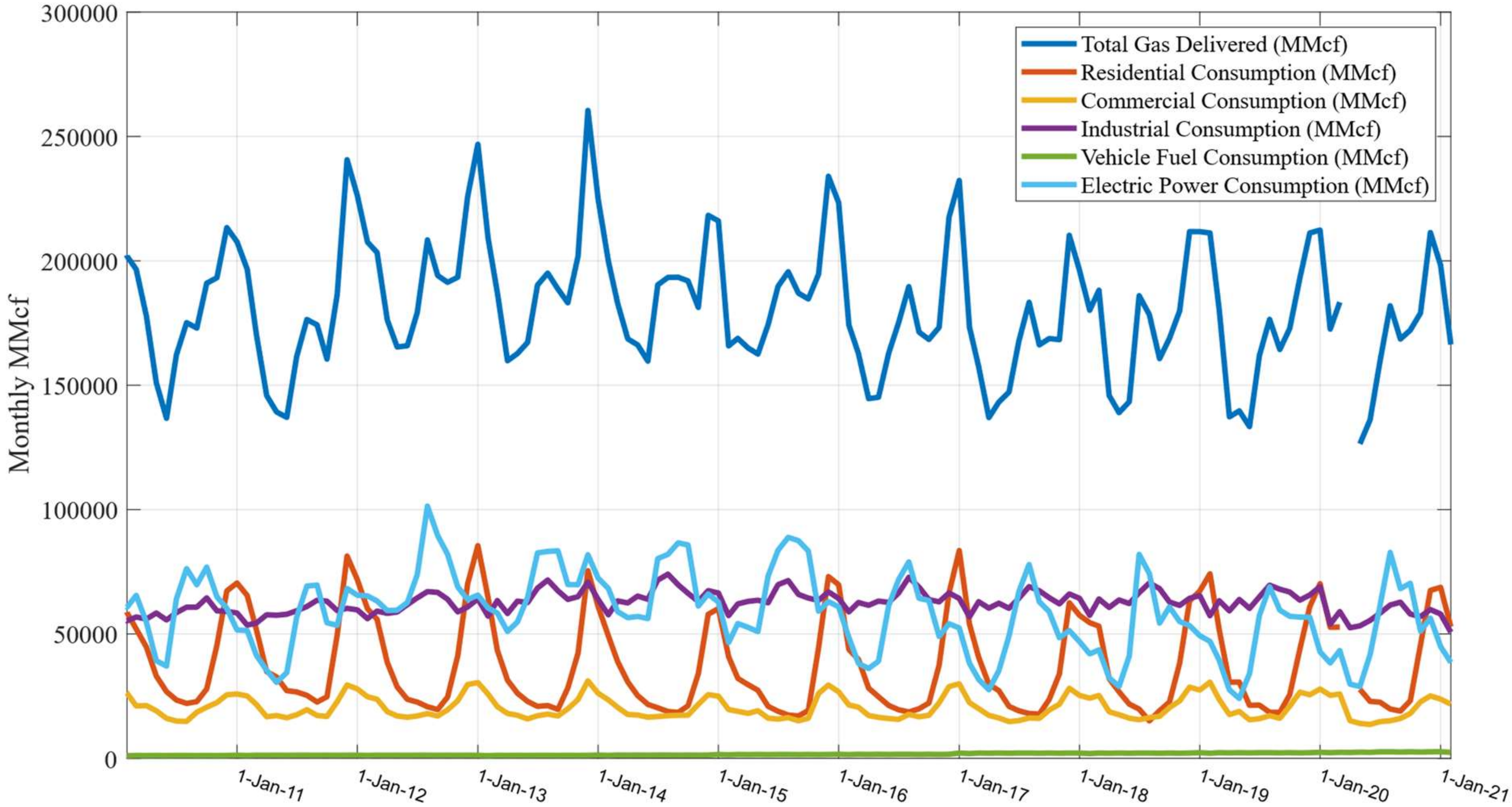


- Natural gas consumption
  - Index *i* commonly measured monthly
    - Billing cycles

- Uses, availability, and quality of time series *C*
  - U.S. Energy Information Administration
    - Public data source
  - Local Distribution Companies (LDC)
    - Introduction to Smart Meter data (daily)



Wisconsin: Consumption by Sector (E.I.A. Data)



Problem statement

1. Given a customer's natural gas consumption series C, we require estimate $\hat{C}$

$$C = \{\, C(i), i = 1, \dots, N\}$$

$$\hat{C} = F\big(f_1(X_1), f_2(X_2), \dots f_m(X_m)\big).$$

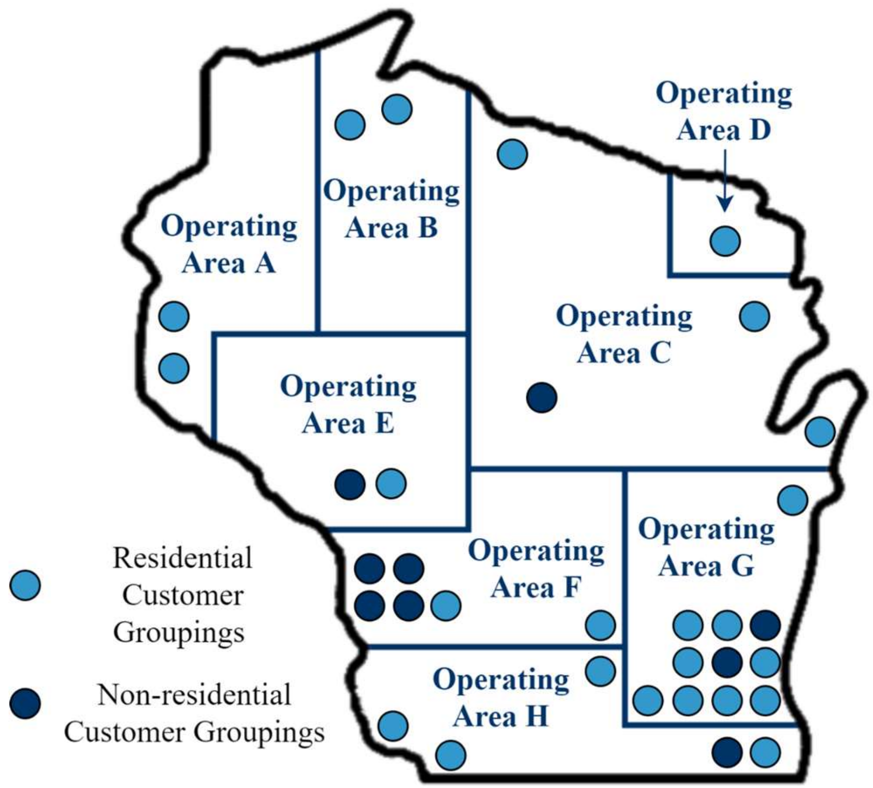
2. Given a time series C, we require an underlying time series, c, of higher frequency to forecast at a more granular level

$$c = \{c(j), j = 1, \dots, K\}$$

$$\hat{c} = \mathcal{A}^{-1}(C).$$

Customer class, geographical ranges, historical trends

LDC Service Area

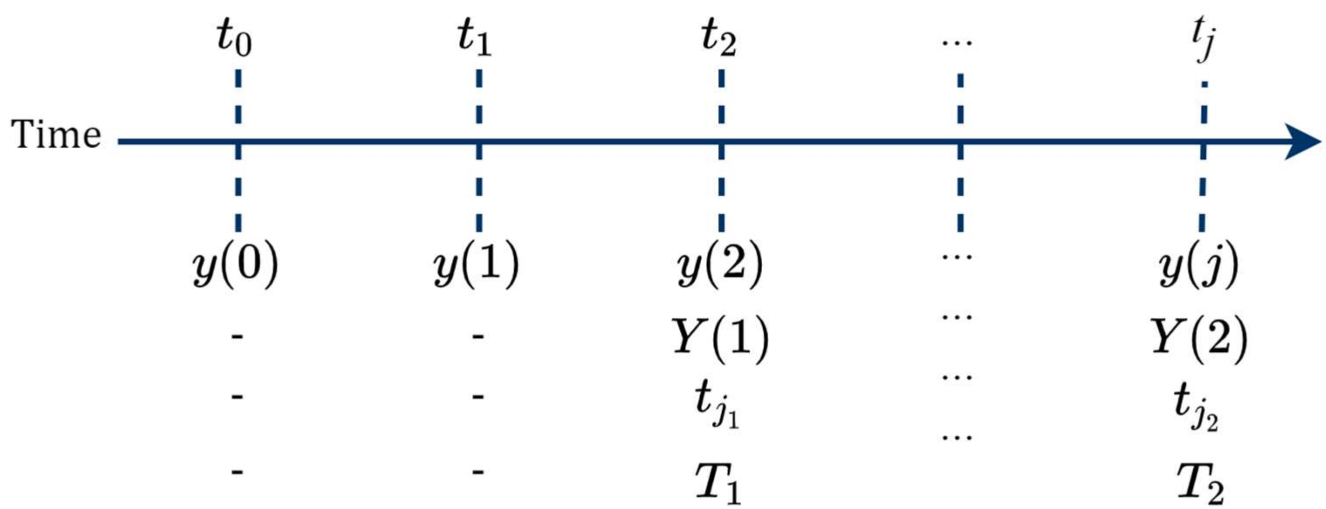


Operating Area Code	Number of Customers	Number of Bills
A	82,438	4,946,280
B	25,823	1,549,380
C	147,325	8,839,500
D	50,683	3,040,980
E	13,659	819,540
F	234,752	14,085,120
G	316,166	18,969,960
H	129,154	7,749,240

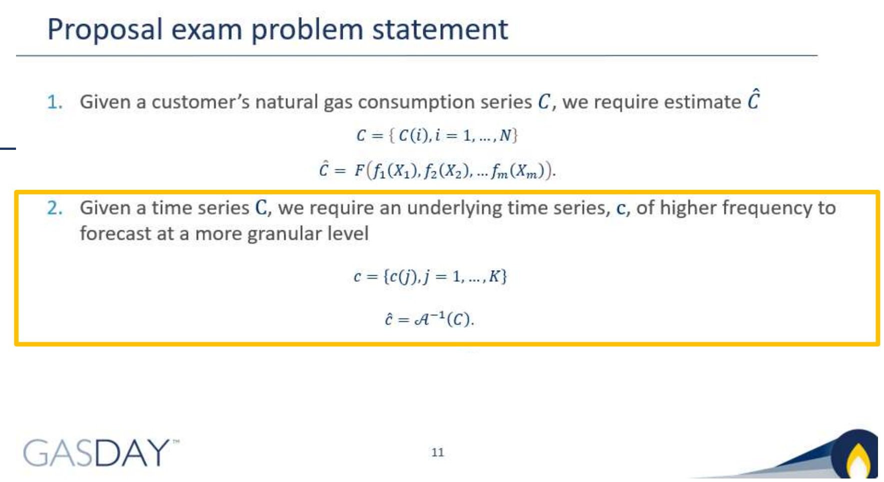
Temporal disaggregation

Series	First interval	General interval
$Y = \{\mathrm{Y(i)}, i = 1, \dots, N\}$	$T_1 = (t_0, t_{j_1}]$	$T_i = (t_{j_{i-1}}, t_{j_i}]$
$y = \{\mathrm{y(j)}, j = 1, \dots, K\}$	$t_1 = (t_0, t_1]$	$t_j = (t_{j-1}, t_j]$

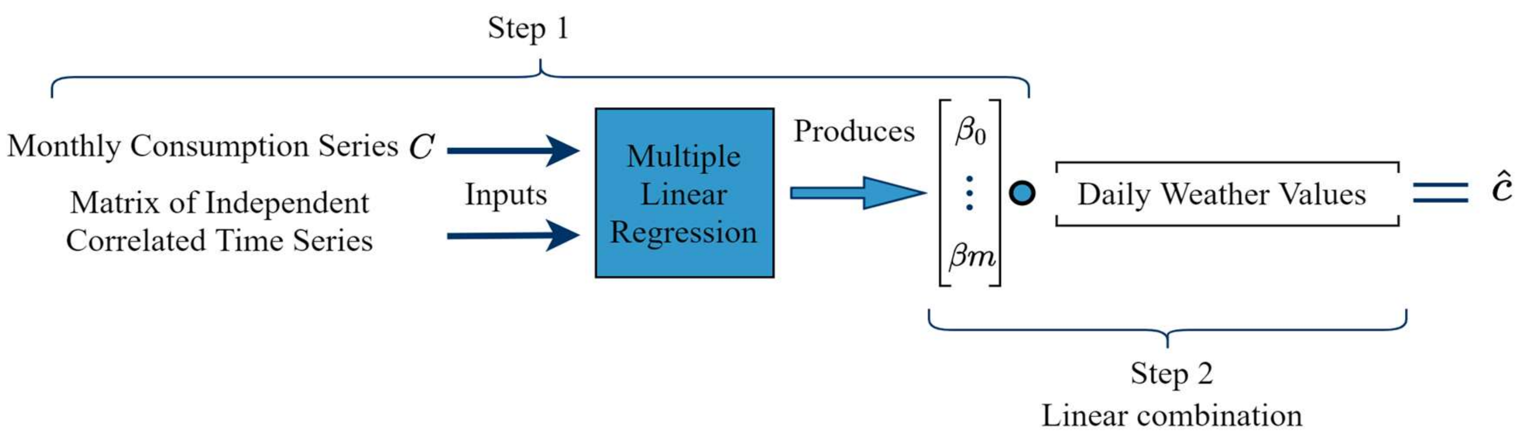
- Let i be the index of aggregated times T_i and low-frequency time series values $Y(i)$
- Let j be the index of disaggregated times t_i and high-frequency time series values $y(j)$



Methods of solution



- Temporal disaggregation from monthly consumption to daily
  - Time series reconstruction (TSR) algorithm
  - Inputs: Aggregated data C and underlying correlated variables X
  - Output: Underlying estimate $\hat{c}$

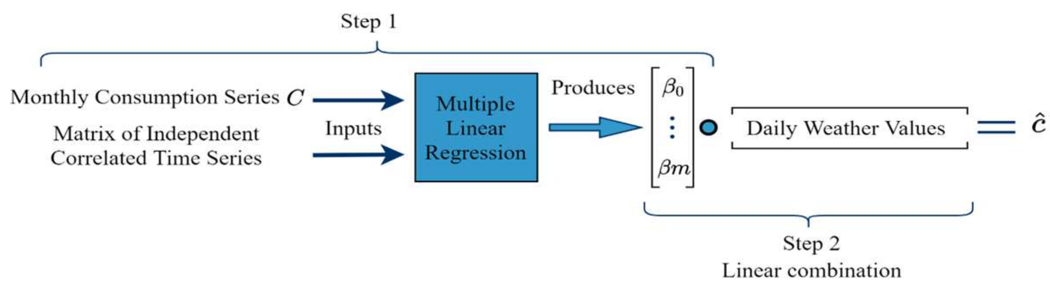


Method of solution

$$\mathrm{HDDW}_{55} = \begin{cases} \mathrm{HDD}_{55} \cdot \left(\frac{152+WS}{160}\right) & WS \leq 8 \\ \mathrm{HDD}_{55} \cdot \left(\frac{72+WS}{80}\right) & WS > 8, \end{cases}$$

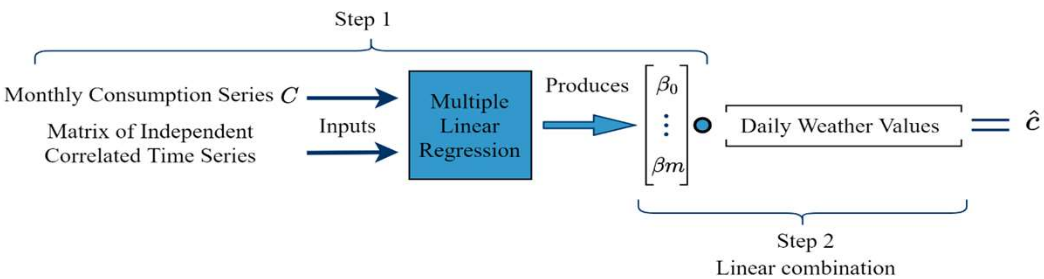
$$\mathrm{HDDW}_{65} = \begin{cases} \mathrm{HDD}_{65} \cdot \left(\frac{152+WS}{160}\right) & WS \leq 8 \\ \mathrm{HDD}_{65} \cdot \left(\frac{72+WS}{80}\right) & WS > 8. \end{cases}$$

$$\mathrm{CDD}_{65} = \max\left(0, T - 65\right).$$



- Step 3: Piecewise linear optimization
  - Inputs: TSR $\hat{c}$ (from step 2)
  - Outputs: An alternative $\hat{c}$ that maintains the consistency between the aggregated data and the sum of the estimated data within an aggregated time step.
  - Loses variability, but maintains similar shape to TRS output

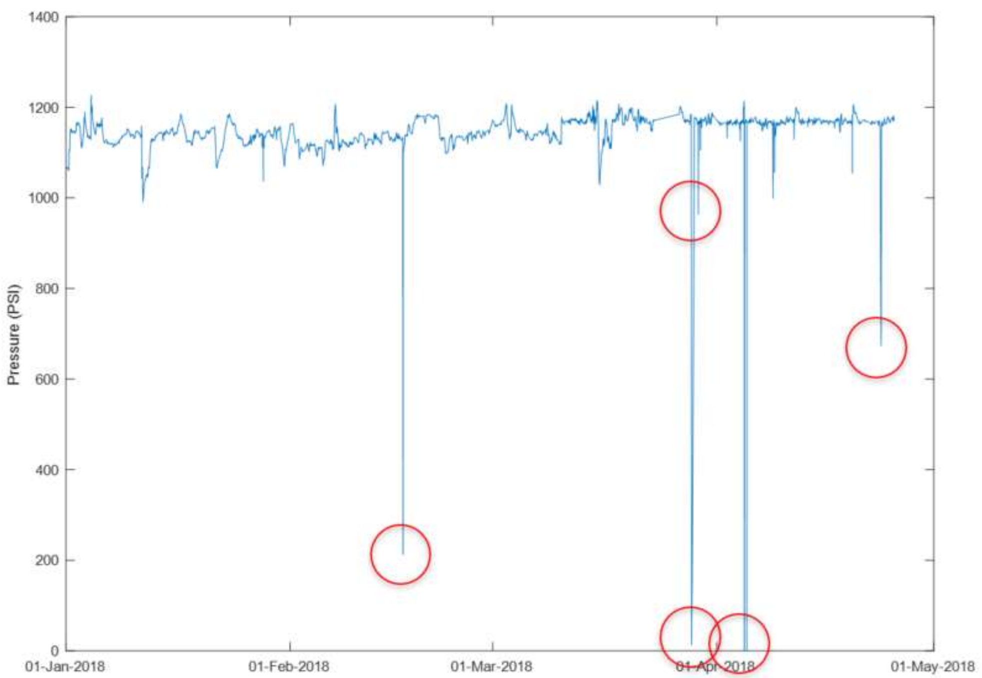
Method of solution



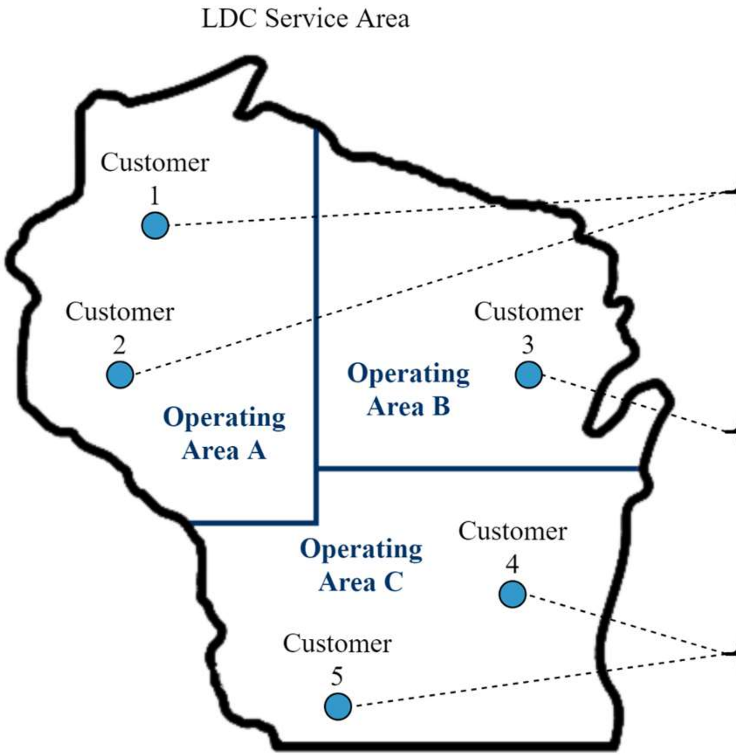
- Step 4:Resample, interpolate
  - Inputs: $\hat{c}$ (from step 3)
  - Outputs: A new underlying $\hat{c}$ that is used in step 1:N iterations

Results precursor: data quality

- Data quality
  - E.I.A. State Disaggregation
    - Source: Energy Information Administration
      - Aggregate monthly consumption from select LDCs throughout the state
    - Source: LDC data
      - Real world data, inconsistent intervals
      - Nondisclosure agreements, avoiding PII



Introduction to natural gas demand forecasting



$$C_1 = \{C_1(i), i = 1, \ldots, N\}$$

$$C_2 = \{C_2(i), i = 1, \ldots, N\}$$

$$OpArea_A = \sum_{i=1}^{N} C_1\,(i) + C_2(i)$$

$$C_3 = \{C_3(i), i = 1, \ldots, N\}$$

$$OpArea_B = \ C_3$$

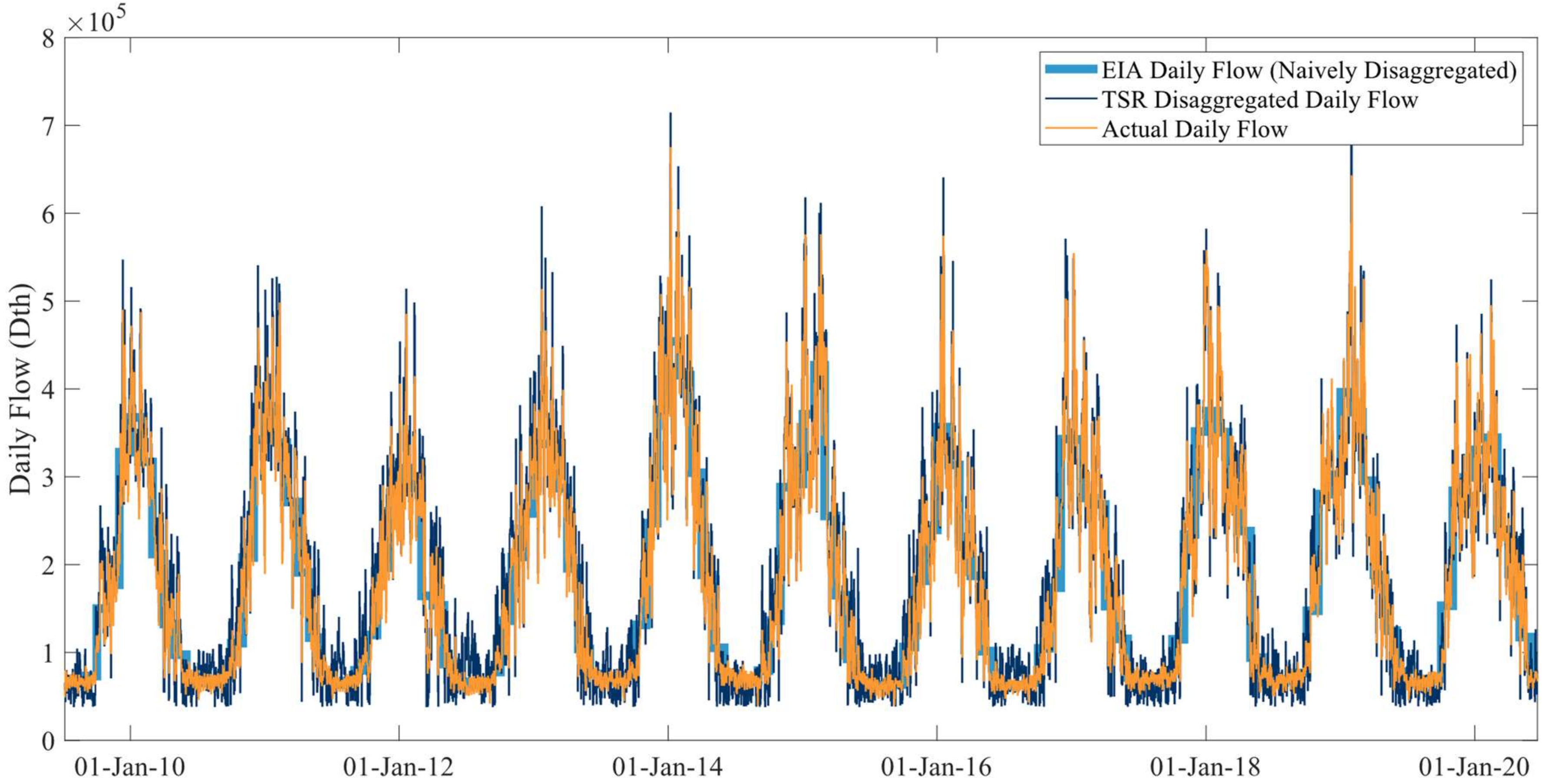
$$C_4 = \{C_4(i), i = 1, \ldots, N\}$$

$$C_5 = \{C_5(i), i = 1, \ldots, N\}$$

$$OpArea_C = \sum_{i=1}^{N} C_4\,(i) + C_5(i)$$

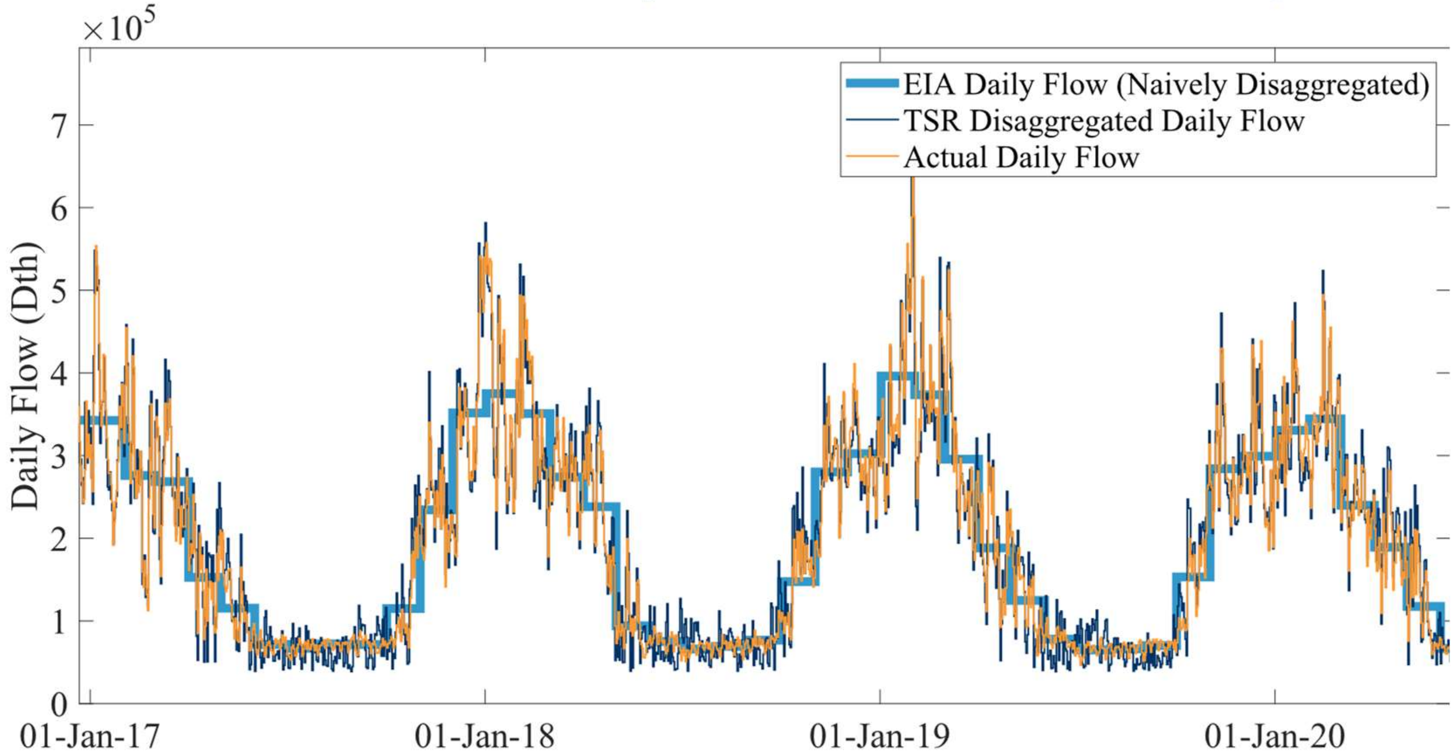
$$LDC = \sum_{i=1}^{N} OpArea_A(i) + \ OpArea_B(i) \ + OpArea_C(i)$$

Results – Wisconsin total consumption disaggregated



Results – Wisconsin total consumption disaggregated

- Training set: State of Wisconsin consumption data from 2010-2016 (all sectors)
- Test set: State of Wisconsin consumption data from 2017-2020 (all sectors)



Results – Wisconsin total consumption disaggregated

- Training set: State of Wisconsin consumption data from 2010-2016 (all sectors)
- Test set: State of Wisconsin consumption data from 2017-2020 (all sectors)

- Root mean square error (RMSE) $= \sqrt{\frac{\sum_{i=1}^{T}(\hat{c}(t)-\mathrm{c}(\mathrm{t}))^2}{T}}$
  - Naïve Disaggregation RMSE: 306 MMcf
  - TRS RMSE: 201 MMcf
- Mean absolute percentage error (MAPE) $= \frac{1}{T}\sum_{i=1}^{T}\frac{|(\hat{c}(t)-\mathrm{c}(\mathrm{t}))|}{(\mathrm{c}(\mathrm{t}))}$
  - Naïve Disaggregation MAPE: 22.6%
  - TSR MAPE: 15.8%

- Algorithm benefits:
  - Able to handle arbitrary aggregated time steps and non-uniform
  - Mechanisms in place to drive aggregated sub-intervals towards the original interval value

Questions?

- Colin O. Quinn
  - colin.o.quinn@marquette.edu
- Richard J. Povinelli
  - richard.povinelli@marquette.edu

[1] B. Gal, “The Challenges of Time-Series Forecasting in Retail,” *Retail Technology*, 2017. https://cb4.com/blog/time-series-forecasting-challenges/ (accessed Mar. 20, 2021).
[2] S. R. Vitullo, G. F. Corliss, M. Adya, F. Nourzad, and R. H. Brown, “An algorithm for disaggregating temporal natural gas consumption,” *Can. Appl. Math. Q.*, vol. 21, no. 3, pp. 391–410, 2013.
[3] W. Wei and D. Stram, “Disaggregation of Time Series Models,” *Wiley R. Stat. Soc.*, vol. 52, no. 3, pp. 453–467, 1990, [Online]. Available: https://www.jstor.org/stable/2345669.
[4] F. Moauro and G. Savio, “Temporal disaggregation using multivariate structural time series models,” *Econom. J.*, vol. 8, no. 2, pp. 214–234, 2005, doi: 10.1111/j.1368-423x.2005.00161.x.
[5] A. Dotis-Georgiou, “Autocorrelation in Time Series Data,” *DZone: A Devada Media Property*, 2019. https://dzone.com/articles/autocorrelation-in-time-series-data (accessed Jan. 02, 2020).
[6] H. Akaike, “Fitting Autoregressive Models for Prediction,” *Ann. Inst. Stat. Math.*, vol. 21, no. 1, pp. 243–247, 1969, doi: 10.1007/BF02532251.
[7] J. Durbin, “The Fitting of Time-Series Models,” *Int. Stat. Inst.*, vol. 28, no. 3, pp. 233–244, 1960, [Online]. Available: https://www.jstor.org/stable/1401322.
[8] V. Flovik, “How (not) to use Machine Learning for time series forecasting: Avoiding the pitfalls,” *Towards Data Science*, 2018. https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424 (accessed Jan. 07, 2020).
[9] A. Harvey, “Forecasting with Unobserved Components Time Series Models,” in *Handbook of Economic Forecasting*, vol. 1, G. Elliott, C. Granger, and A. Timmermann, Eds. Elsevier, 2006, pp. 327–412.
[10] C. Chafield, *Time-Series Forecasting*, 1st ed., no. 1. Boca Raton, Florida: Chapman & Hall, 2000.
[11] “Time Series Analysis: Concept, Additive and Multiplicative Models,” *Theintactone*, 2020. https://theintactone.com/2019/08/28/bsa-u2-topic-1-time-series-analysis-concept-additive-and-multiplicative-models/ (accessed Apr. 02, 2021).
[12] T. Hong, “Energy Forecasting: Past, Present, and Future,” *Foresight Int. J. Appl. Forecast.*, vol. 32, pp. 43–48, 2013.
[13] B. Soldo, “Forecasting natural gas consumption,” *Appl. Energy*, vol. 92, pp. 26–37, 2012, doi: 10.1016/j.apenergy.2011.11.003.
[14] S. Makridakis and R. L. Winkler, “Averages of Forecasts: Some Empirical Results,” *Manage. Sci.*, vol. 29, no. 9, pp. 987–996, Sep. 1983, doi: 10.1287/mnsc.29.9.987.
[15] B. E. Flores, “A pragmatic view of accuracy measurement in forecasting,” *Omega*, vol. 14, no. 2, pp. 93–98, 1986, doi: https://doi.org/10.1016/0305-0483(86)90013-7.
[16] R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications With R Examples*, 4th ed. Springer, 2017.
[17] P. Balestra and M. Nerlove, “Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas,” *Econometrica*, vol. 34, no. 3, pp. 585–612, 1966, [Online]. Available: https://www.jstor.org/stable/1909771.
[18] C. Chow and A. Lin, “Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series,” *Rev. Econ. Stat.*, vol. 53, no. 4, pp. 372–375, 1971.
[19] S. H. Mohr and G. M. Evans, “Long term forecasting of natural gas production,” *Energy Policy*, vol. 39, no. 9, pp. 5550–5560, 2011, doi: 10.1016/j.enpol.2011.04.066.
[20] S. R. Vitullo, R. H. Brown, G. F. Corliss, and B. M. Marx, “Mathematical Models for Natural Gas Forecasting,” *Can. Appl. Math Q.*, vol. 17, no. 4, pp. 1–13, 2010, [Online]. Available: papers2://publication/uuid/5E8CDA33-908F-4B19-8F90-02A304882841.
[21] S. Fan and R. J. Hyndman, “Short-term load forecasting based on a semi-parametric additive model,” *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, 2012, doi: 10.1109/TPWRS.2011.2162082.
[22] E. Elattar, J. Y. Goulermas, and Q. H. Wu, “Electric Load Forecasting Based on Locally Weighted Support Vector Regression,” *IEEE Trans. Syst. Man. Cybern.*, vol. 40, no. 4, pp. 438–447, 2010.
[23] W. S. Cleveland and S. J. Devlin, “Locally weighted regression: An approach to regression analysis by local fitting,” *J. Am. Stat. Assoc.*, vol. 83, no. 403, pp. 596–610, 1988, doi: 10.1080/01621459.1988.10478639.
[24] J. G. Tamba *et al.*, “Forecasting natural gas: A literature survey,” *Int. J. Energy Econ. Policy*, vol. 8, no. 3, pp. 216–249, 2018.
[25] I. Moghram and S. Rahman, “Analysis and Evaluation of Five Short-Term Load Forecasting Techniques,” *IEEE Power Eng. Rev.*, vol. 9, no. 11, pp. 42–43, 1989, doi: 10.1109/MPER.1989.4310383.
[26] C. N. Babu and B. E. Reddy, “A Moving-average Filter Based Hybrid ARIMA-ANN Model for Forecasting Time Series Data,” *Appl. Soft Comput. J.*, vol. 23, pp. 27–38, 2014, doi: 10.1016/j.asoc.2014.05.028.
[27] J. W. Taylor, “Short-term electricity demand forecasting using double seasonal exponential smoothing,” *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, 2003, doi: 10.1057/palgrave.jors.2601589.
[28] S. J. Huang and K. R. Shih, “Short-term load forecasting via ARMA model identification including non-Gaussian process considerations,” *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, 2003, doi: 10.1109/TPWRS.2003.811010.
[29] H. Spliid, “A Fast Estimation Method for the Vector Autoregressive Moving Average Model with Exogenous Variables,” *J. Am. Stat. Assoc.*, vol. 78, no. 384, pp. 843–849, Dec. 1983, doi: 10.1080/01621459.1983.10477030.