Convex Optimization in Renewable Energy Management

Giray Ogut
University PhD Dissertation Defense
Department of Electrical Engineering

Advisor: Stephen Boyd

Stanford University, February 13 2025

PhD overview

- **research overview:**
  - developing data-driven, interpretable forecasting and control models using convex optimization to address the challenges of integrating renewable energy sources into the power grid
- **research areas:**
  - copula models for photovoltaic systems
  - interpretable net load forecasting
  - time dilated analysis of photovoltaic output
  - continuous ranked probability score regression
  - single node energy management
  - aging-aware battery control
- **research focus:**
  - interpretability and auditability
  - robustness with missing data and outliers
  - computational efficiency
  - tractability and scalability

Papers

- **G. Ogut**, B. Meyers, S. Boyd, "PV Fleet Modeling via Smooth Periodic Gaussian Copula", *2023 IEEE 50th Photovoltaic Specialists Conference (PVSC)*, San Juan, PR, USA, 2023, pp. 1-8.
- K. Johansson, **G. Ogut**, M. Pelger, T. Schmelzer, S. Boyd, "A simple method for predicting covariance matrices of financial returns", *Foundations and Trends in Econometrics*, 2023.
- **G. Ogut**, B. Meyers, S. Boyd, "Time Dilated Bundt Cake Analysis of PV Output", *2024 IEEE 52nd Photovoltaic Specialists Conference (PVSC)*, Seattle, WA, USA, 2024, pp. 1-7.
- **G. Ogut**, B. Meyers, S. Boyd, "Interpretable net load forecasting using smooth multiperiodic features", *IEEE Transactions on Power Systems*, under review.
- O. Nnorom Jr., **G. Ogut**, S. Boyd, P. Levis, "Aging-aware battery control via convex optimization", *Working paper*.
- **G. Ogut**, B. Meyers, S. Boyd, "Single node energy management via ADP and MPC", *Working paper*.
- **G. Ogut**, S. Boyd, "CRPS regression via convex optimization", *Working paper*.
- B. Meyers, A. Dufour, **G. Ogut**, "Anomaly detection in PV fleet data via interpretable machine learning", *Working paper*.

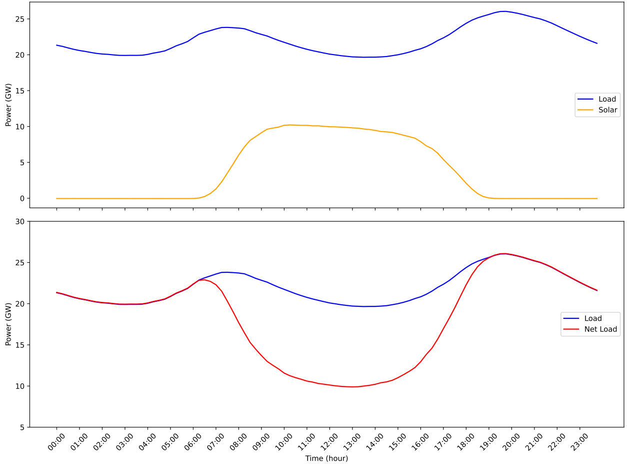
Today's talk

- **G. Ogut**, B. Meyers, S. Boyd, "PV Fleet Modeling via Smooth Periodic Gaussian Copula", *2023 IEEE 50th Photovoltaic Specialists Conference (PVSC)*, San Juan, PR, USA, 2023, pp. 1-8.
- K. Johansson, **G. Ogut**, M. Pelger, T. Schmelzer, S. Boyd, "A simple method for predicting covariance matrices of financial returns", *Foundations and Trends in Econometrics*, 2023.
- **G. Ogut**, B. Meyers, S. Boyd, "Time Dilated Bundt Cake Analysis of PV Output", *2024 IEEE 52nd Photovoltaic Specialists Conference (PVSC)*, Seattle, WA, USA, 2024, pp. 1-7.
- **G. Ogut**, B. Meyers, S. Boyd, "Interpretable net load forecasting using smooth multiperiodic features", *IEEE Transactions on Power Systems*, under review.
- O. Nnorom Jr., **G. Ogut**, S. Boyd, P. Levis, "Aging-aware battery control via convex optimization", *Working paper*.
- **G. Ogut**, B. Meyers, S. Boyd, "Single node energy management via ADP and MPC", *Working paper*.
- **G. Ogut**, S. Boyd, "CRPS regression via convex optimization", *Working paper*.
- B. Meyers, A. Dufour, **G. Ogut**, "Anomaly detection in PV fleet data via interpretable machine learning", *Working paper*.

Interpretable net load forecasting using smooth multiperiodic features

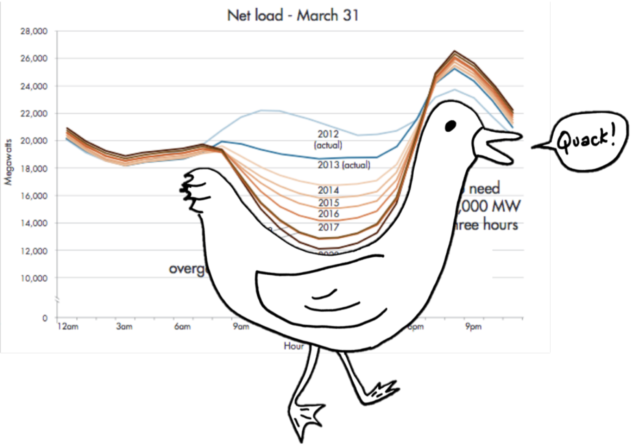
Motivation

- net load is total demand minus behind-the-meter generation (*e.g.* solar)
- figure shows the total daily electricity load and the total solar generation for California, averaged over March 2021, in GW (gigawatts)



Motivation continued

- if you squint your eyes, the load and net load curves look like a duck, hence the name 'duck curve'
- steep rise in net load around 5-7pm can be a challenge for generators, which work better and more efficiently when generating constant or smoothly varying power



(courtesy of the California ISO)

Setting

- consider a real-valued time series, possibly with missing data,

$$y = (y_1, \dots, y_T) \in (\mathbf{R} \cup \{?\})^T,$$

  where $y_t = ?$ means that the value y_t is missing
- let $\mathcal{T} = \{t \mid y_t \in \mathbf{R}\}$ denote the set of known values
- $p_t = (y_t, \dots, y_{t-M+1}), \quad t = M, \dots, T$, vector of the M past values at time t
- $f_t = (y_{t+1}, \dots, y_{t+H}), \quad t = 1, \dots, T - H$, vector of the H future values at time t
- focus on three related forecasting tasks involving predicting f_t from p_t

Tasks

- **point forecast:** estimate f_t with forecasts denoted

$$\hat{f}_t = (\hat{y}_{t+1}, \ldots, \hat{y}_{t+H}) \in \mathbf{R}^H$$

  use average absolute error (AAE) as metric

Tasks

- **point forecast:** estimate f_t with forecasts denoted

$$\hat{f}_t = (\hat{y}_{t+1}, \ldots, \hat{y}_{t+H}) \in \mathbf{R}^H$$

  use average absolute error (AAE) as metric

- **marginal quantile forecast:** estimate the $0 \leq \eta_1 < \cdots < \eta_Q \leq 1$ quantiles of the entries of f_t denoted as

$$q_{t,j} \in \mathbf{R}^H, \quad j = 1, \ldots, Q$$

  use continuous ranked probability score (CRPS) as metric

Tasks

- **point forecast:** estimate f_t with forecasts denoted

$$\hat{f}_t = (\hat{y}_{t+1}, \dots, \hat{y}_{t+H}) \in \mathbf{R}^H$$

  use average absolute error (AAE) as metric

- **marginal quantile forecast:** estimate the $0 \leq \eta_1 < \cdots < \eta_Q \leq 1$ quantiles of the entries of f_t denoted as

$$q_{t,j} \in \mathbf{R}^H, \quad j = 1, \dots, Q$$

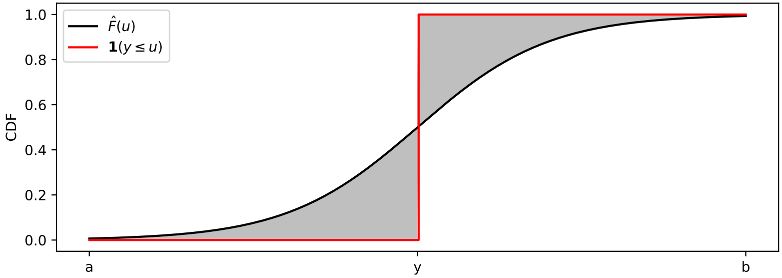
  use continuous ranked probability score (CRPS) as metric

- **generate samples of the future:** generate R plausible realizations of the future values, denoted

$$f_{t,j} \in \mathbf{R}^H, \quad j = 1, \dots, R$$

  use log-likelihood and visual inspection as metrics

CRPS



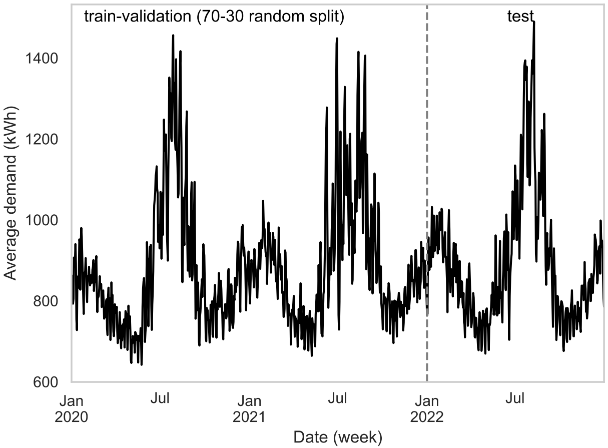
- CRPS is one of many ways to judge a probabilistic prediction and is defined as

$$\mathcal{C}(\hat{F}, y) = \int_a^b (\hat{F}(u) - \mathbf{1}(y \leq u))^2 \, du$$

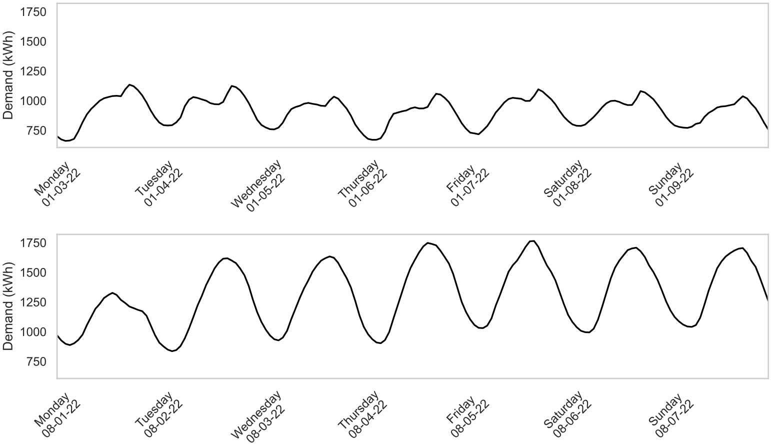
- here $\mathbf{1}(y \leq u) = 0$ when $u < y$ and $\mathbf{1}(y \leq u) = 1$ when $u \geq y$, *i.e.*, it is the CDF of the random variable that takes the value y with probability one

Data

- ISO New England (ISO-NE) net load data for Rhode Island between January 1 2020 and December 31 2022 sampled hourly
- there are 26304 points in total and no missing data
- used data from January 1 2020 to December 31 2021 for in-sample training, and data from January 1 2022 to December 31 2022 for test

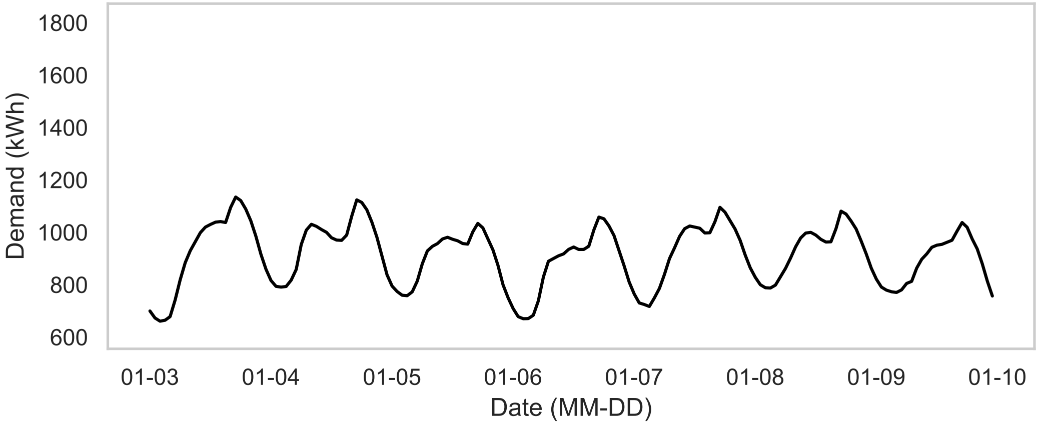


Data continued



- demand is higher during the day, and in summer
- shape of the daily demand changes between seasons

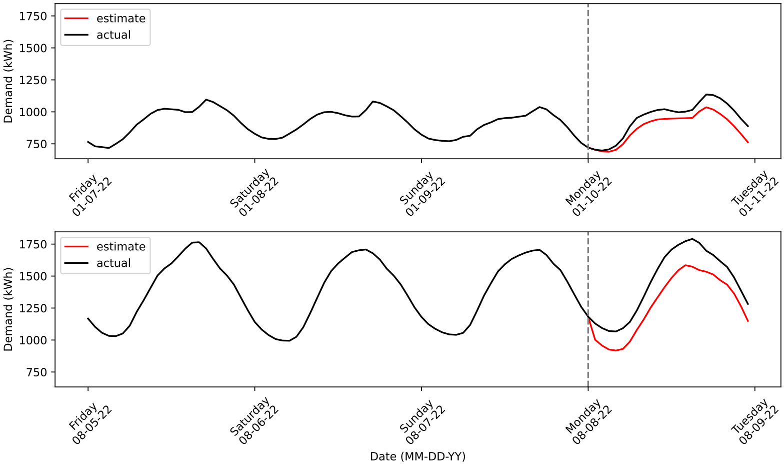
Weekly data for 2022



Baseline

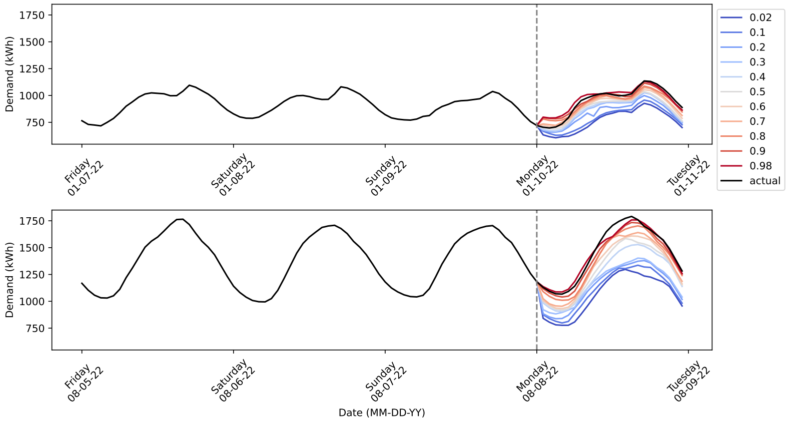
- rolling median forecast looks at the past 14 days of data
- to forecast $t + i$ for $i = 1, \ldots, H$, it uses the median of $(t + i - 24, \ldots, t + i - 336)$
- replace the median with a quantile to obtain marginal quantile forecasts
- used as a baseline in practice as well as the Net Load Forecasting Prize

Rolling median forecasts



- AAE is 70.4 kWh

Rolling quantile forecasts



- CRPS is 24.8 kWh

Method & results

Multiperiodic basis functions

- time-based basis functions $\phi_i : \mathbf{R} \to \mathbf{R}$, $i = 1, \ldots, N$
- time-based feature vectors $\psi_t = (\phi_1(t), \ldots, \phi_N(t)) \in \mathbf{R}^N, \quad t = 1, \ldots, T$

Multiperiodic basis functions

- time-based basis functions $\phi_i : \mathbf{R} \to \mathbf{R}$, $i = 1, \ldots, N$
- time-based feature vectors $\psi_t = (\phi_1(t), \ldots, \phi_N(t)) \in \mathbf{R}^N, \quad t = 1, \ldots, T$
- approximate a smooth Π-periodic function using $2K$ Fourier terms

$$\cos(2\pi kt/\Pi), \quad \sin(2\pi kt/\Pi), \quad k = 1, \ldots, K,$$

  where k is the harmonic number
- for each period Π_i, $i = 1, \ldots, M$ (with $\Pi_P < \cdots < \Pi_1$), use K_i harmonics
- $2\sum_{i=1}^{P} K_i$ basis functions

Interaction terms

- period interactions captured by products of functions from distinct periods
- assign period Π and harmonic k to $\cos(2\pi kt/\Pi)\sin(2\pi \tilde{k}/\tilde{\Pi})$ with $\Pi < \tilde{\Pi}$
- full basis has

$$N = 2\sum_{i=1}^{P} K_i + 4\sum_{1\le i<j\le P} K_i K_j$$

time-based basis functions, denoted $\phi_1, \ldots, \phi_N$, with harmonic numbers $k_1, \ldots, k_N$ (ranging from $\min\{K_1, \ldots, K_P\}$ to $\max\{K_1, \ldots, K_P\}$)

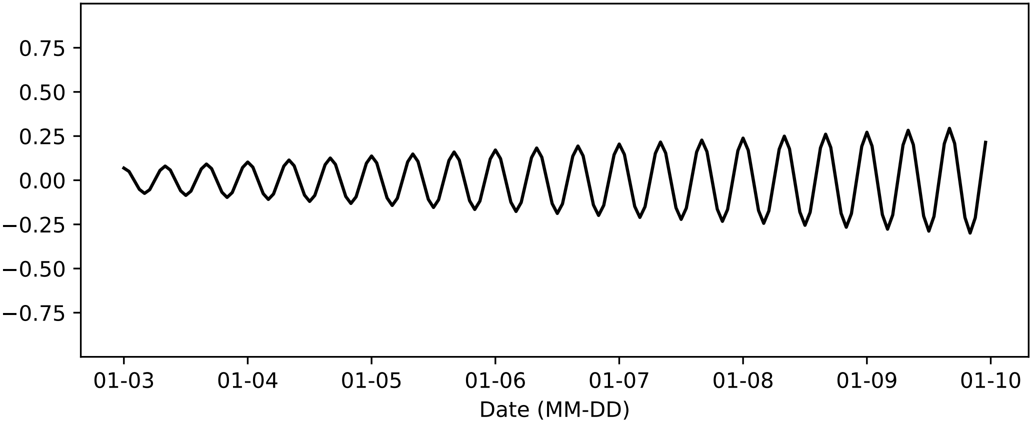
Example

- let t denote hours, set $P = 3$ periodicities $\Pi_1 = 8765.8, \quad \Pi_2 = 168, \quad \Pi_3 = 24$
- use $K_1 = 2$, $K_2 = 3$, $K_3 = 4$ harmonics for annual, weekly, and daily periods
- examples of basis functions

$$\cos(2\pi 3t/24), \quad \sin(2\pi 2t/8765.8), \quad \cos(2\pi 3t/24)\sin(2\pi 2t/8765.8)$$

- $N = 122$ basis functions: 88 daily, 30 weekly, and 4 annual

$\cos(2\pi 3t/24)\sin(2\pi 2t/8765.8)$



Regularization

- define the roughness measure for the P different periods as

$$R_i = \frac{1}{2} \sum_{j \in \mathcal{P}_i} c_j^2 k_j^2, \quad i = 1, \ldots, P,$$

  where $\mathcal{P}_i$ is the set of indices of basis functions associated with period i
- overall roughness $\mathcal{R} = \lambda_1 R_1 + \cdots + \lambda_P R_P$
- $\lambda_1, \ldots, \lambda_P$ are positive hyperparameters associated with the different periods
- $\mathcal{R}$ depends on the coefficients $c_1, \ldots, c_N$, and the hyperparameters $\lambda_1, \ldots, \lambda_P$

Forecast model

- use a linear regression point forecast of the form

$$\hat{f}_t = \theta^{\text{const}} + \theta^{\text{time}}\psi_t + \theta^{\text{past}} p_t,$$

  where $\psi_t \in \mathbf{R}^N$ is a vector of features that are based on time only
- parameters are $\theta^{\text{const}} \in \mathbf{R}^H$, $\theta^{\text{time}} \in \mathbf{R}^{H \times N}$, and $\theta^{\text{past}} \in \mathbf{R}^{H \times M}$
- while the time-based feature vector ψ_t does not contain missing data, the autoregressive feature vector p_t can
- fill in missing entries in each p_t using forward fill and standardize

Fitting the point forecast model

- minimize the convex function

$$\sum_{t,i|t+i\in\mathcal{T}^{\text{train}}} \frac{1}{2}|(f_t)_i - (\hat{f}_t)_i| + \mathcal{R}(\theta^{\text{time}}; \lambda^{\text{time}}) + \lambda^{\text{past}}\|\theta^{\text{past}}\|_F^2,$$

  where $\|\cdot\|_F$ is the Frobenius norm
- using an absolute value loss function corresponds to forecasting the median value
- problem above is separable across the entries of f_t, and can be solved for each row of the parameters in parallel
- there are $P+1$ hyperparameters, $\lambda^{\text{time}} \in \mathbf{R}^P$ and $\lambda^{\text{past}} \in \mathbf{R}$ in this model

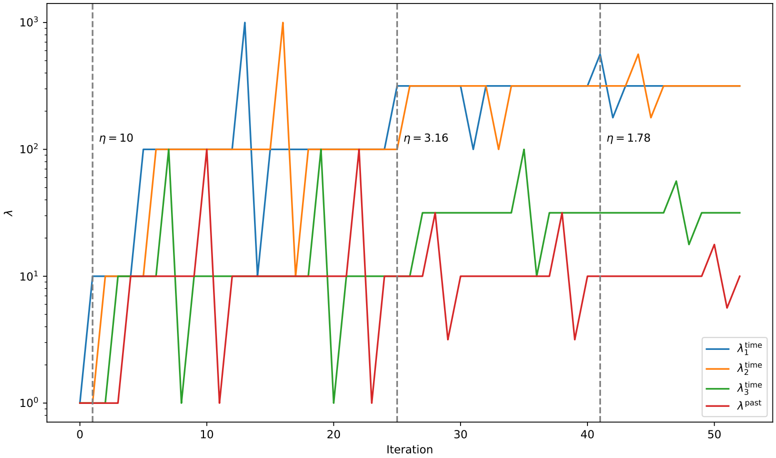
Hyperparamater search

- use cross validation on the in-sample data to choose hyperparameters
- one traditional method, useful when the number of hyperparameters is modest, is complete grid search
- another simple method, found to be effective, uses a cyclical greedy search

Cyclical greedy hyperparameter search

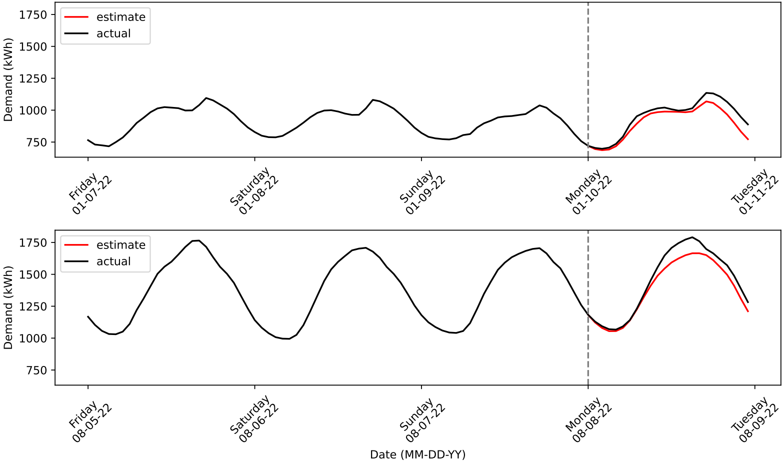
- start with initial hyperparameters, upper and lower bounds
- for each hyperparameter, try increasing by a factor η while keeping others fixed
- if validation error decreases, update it
- otherwise, try decreasing (i.e. dividing by η) and accept if improved
- revert if neither change helps
- stop if no improvement is found over all hyperparameters

Hyperparamater search experiments for point forecast



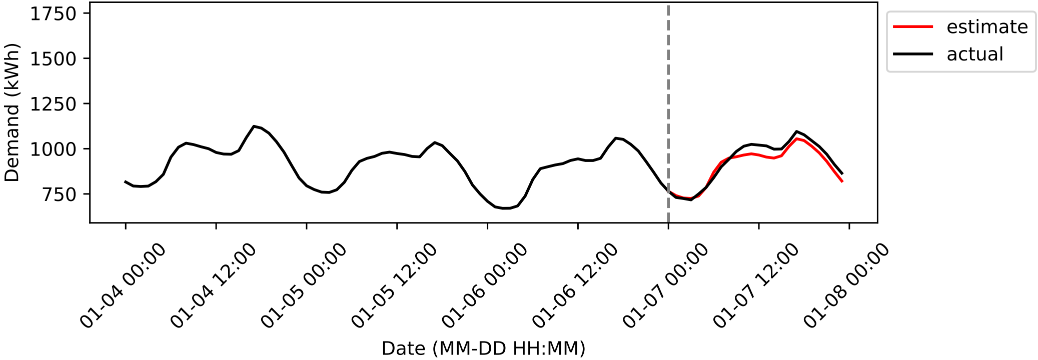
- best hyperparameters:
  $\lambda^{\text{past}} = 10.00, \quad \lambda_1^{\text{time}} = 316.23, \quad \lambda_2^{\text{time}} = 316.23, \quad \lambda_3^{\text{time}} = 31.62$

Point forecasts



- both for summer and winter predictions are close to actual realizations

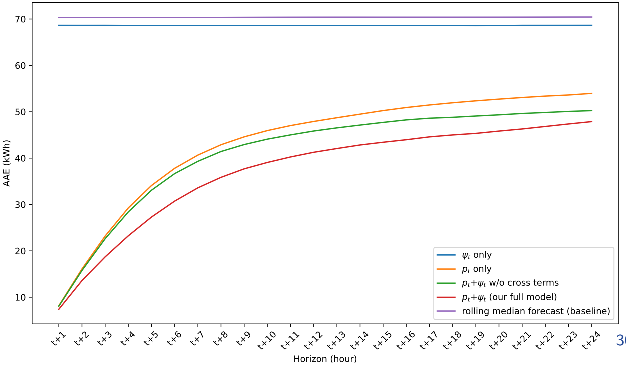
Point forecasts for January 2022



AAE comparison

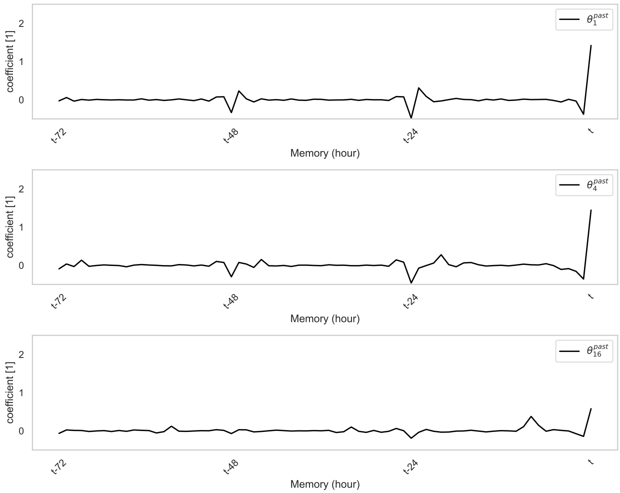
- just using past features gives a more significant improvement on baseline than just using time features
- the lowest AAE is achieved by combining these two types of features
- cross terms give a substantial reduction in AAE

Model	AAE (kWh)
Baseline model	70.4
Time features alone	68.6
Past features alone	43.3
All features except cross terms	41.2
Full features	**37.1**



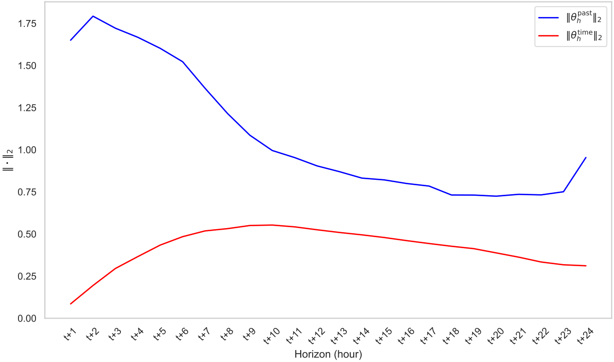
Interpretability

- magnitude of the coefficients can be used to explain which features were important for the forecasted values
- figure shows entries of θ_h^{past} for horizons $h = t+1, t+4, t+16$
- observe that the coefficients are approximately sparse, with most of the entries near zero
- model pays special attention to the previous values 24, 48, and 72 hours ago



Interpretability continued

- relative norms of the rows of θ^{past} and θ^{time} can be used to investigate how important time features are compared to past features for various horizons
- figure shows the ℓ_2 norm of rows of θ^{past} and θ^{time}
- norm of rows of θ^{past} generally decreases with increasing horizon whereas the norm of rows of θ^{time} increases with increasing horizon



Fitting marginal quantile forecast model

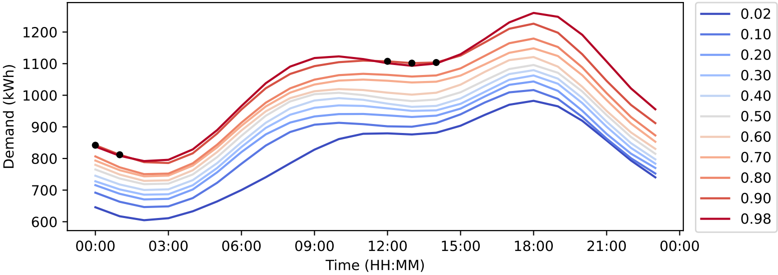
- simply modify the loss function used in point forecast to estimate quantiles other than $\eta = 0.5$
- for the η_j quantile, minimize

$$\sum_{t,i|t+i\in\mathcal{T}^{\text{train}}} \ell_{\eta_j}\left((f_t)_i - (\hat{f}_t)_i\right) + \mathcal{R}(\theta^{\text{time}}; \lambda^{\text{time}}) + \lambda^{\text{past}}\|\theta^{\text{past}}\|_F^2,$$

where ℓ_{η_j} is the pinball loss associated with jth quantile η_j defined as

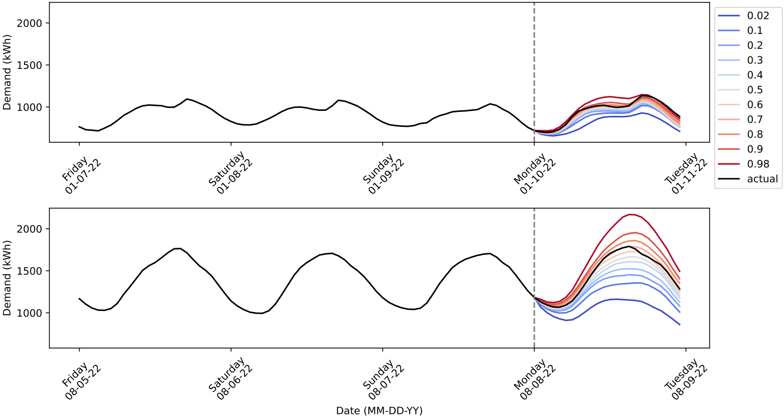
$$\ell_{\eta_j}((f_t)_i, q_{t,j}) = \begin{cases} \eta_j\left((f_t)_i - q_{t,j}\right) & (f_t)_i \geq q_{t,j} \\ (1-\eta_j)\left((f_t)_i - q_{t,j}\right) & (f_t)_i < q_{t,j}. \end{cases}$$

Quantile crossing problem



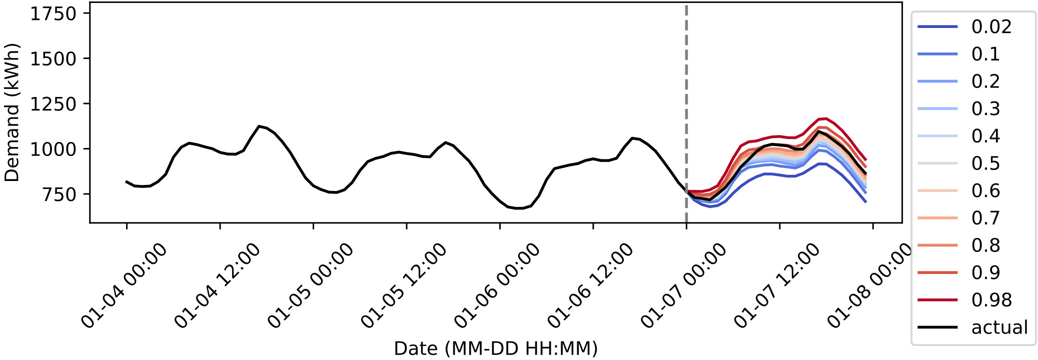
- quantiles can be estimated jointly or separately, and usually there are computational advantages to estimating them separately
- if quantiles are estimated separately, there is the awkward possibility that the predicted quantiles are out of order
- one common method to fix this in practice is to sort the predicted quantiles

Marginal quantile forecasts



- in winter, quantiles form tight bands centered around the actual realization
- in summer, even though the actual realization was close to the estimated median, the interquantile distance is much higher

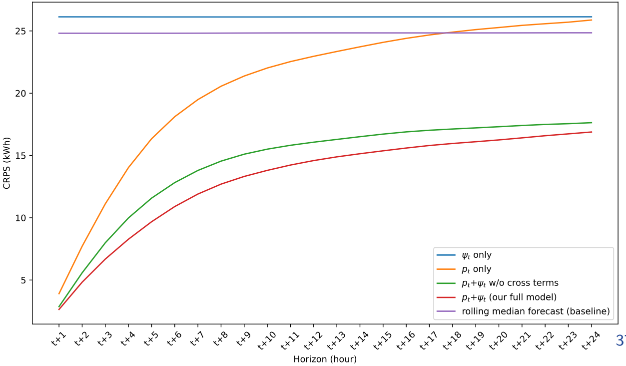
Estimated marginal quantiles for January 2022



CRPS comparison

- results similar to point forecast
- but now it is more important to include time features
- performance of the model with just past features quickly deteriorates with increasing horizon and reaches the baseline level

Model	CRPS (kWh)
Baseline model	24.8
Time features alone	26.1
Past features alone	20.8
All features except cross terms	14.5
Full features	**13.1**



Conditional sample forecasts

- model f_t, conditioned on p_t and ψ_t, as

$$f_t \sim \mathcal{N}(\hat{f}_t + \mu, \Sigma),$$

where $\hat{f}_t$ is point estimate, μ is the point forecast error mean, and $\Sigma \in \mathbf{R}^{H \times H}$ the point forecast error covariance

Conditional sample forecasts

- model f_t, conditioned on p_t and ψ_t, as

$$f_t \sim \mathcal{N}(\hat{f}_t + \mu, \Sigma),$$

  where $\hat{f}_t$ is point estimate, μ is the point forecast error mean, and $\Sigma \in \mathbf{R}^{H \times H}$ the point forecast error covariance
- define the point forecast errors or residuals as $r_t = \hat{y}_t - y_t$
- start with empirical mean

$$\mu_i^{\text{emp}} = \frac{1}{|\mathcal{S}_i|} \sum_{t \in \mathcal{S}_i} (r_t)_i, \quad i = 1, \ldots, H,$$

  where $\mathcal{S}$ is the set of time periods for which $(r_t)_i$ is known

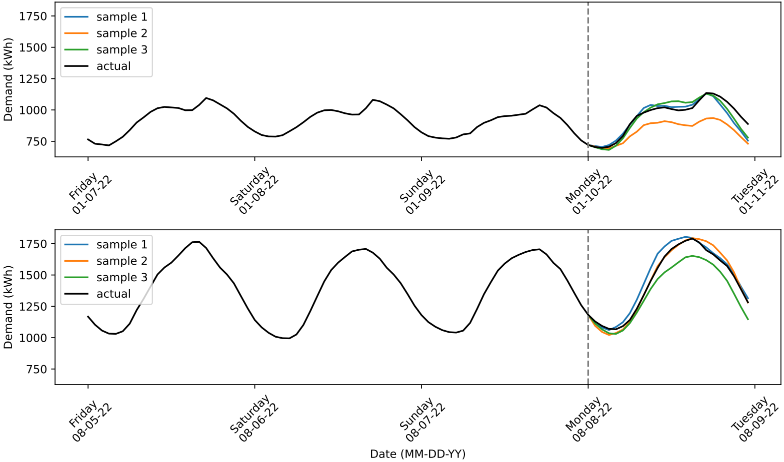
Conditional sample forecasts

- model f_t, conditioned on p_t and ψ_t, as
$$f_t \sim \mathcal{N}(\hat{f}_t + \mu, \Sigma),$$
where $\hat{f}_t$ is point estimate, μ is the point forecast error mean, and $\Sigma \in \mathbf{R}^{H \times H}$ the point forecast error covariance
- define the point forecast errors or residuals as $r_t = \hat{y}_t - y_t$
- start with empirical mean
$$\mu_i^{\text{emp}} = \frac{1}{|\mathcal{S}_i|} \sum_{t \in \mathcal{S}_i} (r_t)_i, \quad i = 1, \ldots, H,$$
where $\mathcal{S}$ is the set of time periods for which $(r_t)_i$ is known
- then form the empirical covariance matrix as
$$\Sigma_{ij}^{\text{emp}} = \frac{1}{|\mathcal{S}_{ij}|} \sum_{t \in \mathcal{S}_{ij}} \left((r_t)_i - \mu_i\right) \left((r_t)_j - \mu_j\right)^T,$$
where $\mathcal{S}_{ij} = \mathcal{S}_i \cap \mathcal{S}_j$ is the set of time periods where both $(r_t)_i$ and $(r_t)_j$ are known

Factor regularization

- regularization can improve the covariance estimate $\tilde{\Sigma}$
- regularization also addresses the issue of Σ^{emp} not being positive semidefinite
- use a factor form, $\Sigma = FF^T + D$, where $F \in \mathbf{R}^{H \times q}$ and $D \in \mathbf{R}^{H \times H}$ is diagonal with nonnegative entries, and $q \leq H$ is number of factors
- take FF^T as the rank q appromixation of $\tilde{\Sigma}$ obtained from an eigendecomposition
- choose D so that $\Sigma_{ii} = \tilde{\Sigma}_{ii}$, $i = 1, \ldots, H$
- number of factors q can be chosen as the number of significant eigenvalues of $\tilde{\Sigma}$

Conditional generated samples



- figure shows 3 conditional generated samples for winter and summer
- generated samples are similar to the actual values of the time series

Conclusion

- developed an interpretable net load forecasting method based on linear regression, which introduces customized time features to model a quasi multiperiodic process
- showed applications such as point forecasts, marginal quantile forecasts, and conditional samples on ISO-NE net load data for Rhode Island
- open source implementation that can be used for any time series that exhibits multiple periodicities, available as `pip install spcqe`

Future work

- include additional features such as weather forecasts
- compare with more advanced, less interpretable models such as XGBoost
- extend the model with other basis functions, analyze other similar processes
- investigate different missing data patterns and imputation methods

Funding

- this material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office (SETO) Award Number 34368

Acknowledgements

Acknowledgements

- my advisor Stephen Boyd

Acknowledgements

- my advisor Stephen Boyd
- my committee members Roger T. Howe, Brad Osgood, Dimitry Gorinevsky, and Mert Pilanci

Acknowledgements

- my advisor Stephen Boyd
- my committee members Roger T. Howe, Brad Osgood, Dimitry Gorinevsky, and Mert Pilanci
- Bennet Meyers and other collaborators listed in the papers

Acknowledgements

- my advisor Stephen Boyd
- my committee members Roger T. Howe, Brad Osgood, Dimitry Gorinevsky, and Mert Pilanci
- Bennet Meyers and other collaborators listed in the papers
- my mother Nurhan Ogut, father Haldun Ogut and little sister Pinar Ece Ogut

Acknowledgements

- my advisor Stephen Boyd
- my committee members Roger T. Howe, Brad Osgood, Dimitry Gorinevsky, and Mert Pilanci
- Bennet Meyers and other collaborators listed in the papers
- my mother Nurhan Ogut, father Haldun Ogut and little sister Pinar Ece Ogut
- current and former cvxgrp members

Acknowledgements

- my advisor Stephen Boyd
- my committee members Roger T. Howe, Brad Osgood, Dimitry Gorinevsky, and Mert Pilanci
- Bennet Meyers and other collaborators listed in the papers
- my mother Nurhan Ogut, father Haldun Ogut and little sister Pinar Ece Ogut
- current and former cvxgrp members
- Utkan Demirci and Gozde Durmus

Thank you!