IMPERIAL

Optimization-informed Linear Pooling of Probabilistic Forecasts

Akylas Stratigakos, Salvador Pineda, Juan M. Morales
44th International Symposium on Forecasting, Dijon, France
03/07/2024

OASYS Optimization and Analytics for Sustainable energY Systems

LEVERHULME TRUST

Agenda

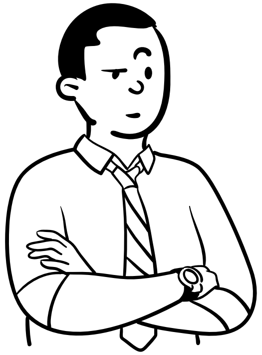
Motivation

Forecast Combination for Constrained Optimization

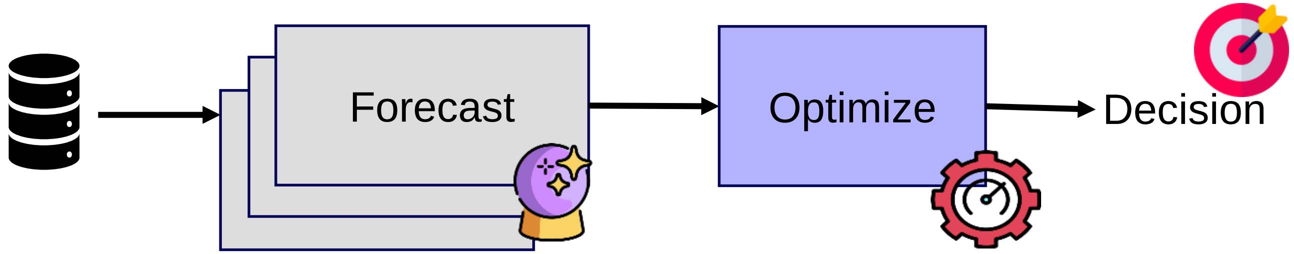
From Data to Decisions: *Forecast-then-Optimize* Paradigm

Almost all real-world decisions are made under parameter uncertainty

- Demand, renewable production, market prices, etc.
- Forecasting (point, probabilistic, scenarios) is a critical step



Does increased forecast accuracy translate into forecast value?



Cost to minimize

Engineering & physical constraints, risk & comfort level

Forecasting models

Contextual information: Weather data, market data, historical data

Multiple Forecasts: Good or Bad?

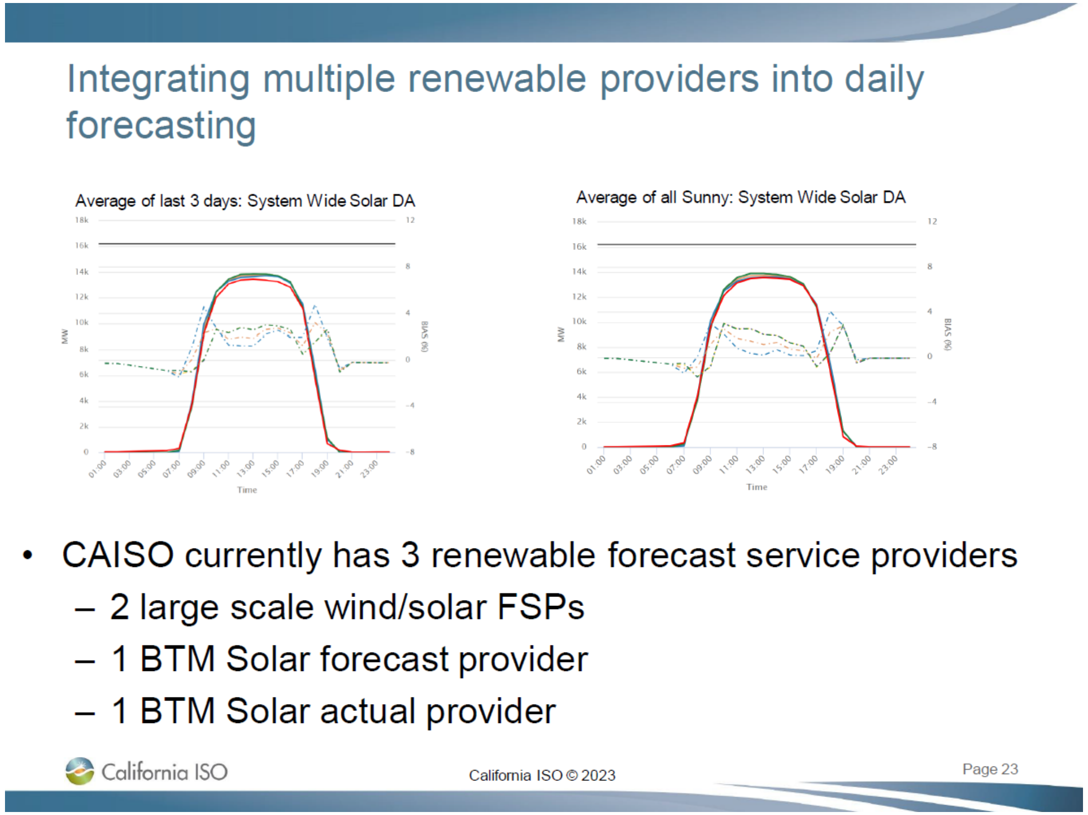
In many real-world settings, decision-makers receive multiple forecasts from external vendors

- Forecast combination has been known to improve accuracy since the 1960's [1]

Power & Energy systems:

- How to utilize probabilistic forecasts [2]?
- Multivendor optimization

My forecast: System Operators (SOs) will <u>definitely</u> catch up by 2060



Snapshot from Motley, A., 2023. CAISO: Advances in the use of wind and solar forecasting, https://www.esig.energy/event/g-pst-esig-webinar-series-advances-in-the-use-of-wind-and-solar-forecasting/

[1] Bates, J.M., Granger, C.W., 1969. The combination of forecasts. Journal of the operational research society 20, 451–468.
[2] https://globalpst.org/wp-content/uploads/042921G-PST-Research-Agenda-Master-Document-FINAL_updated.pdf

Combining Probabilistic Forecasts: What, How, and Why

Evaluating probabilistic forecasts
- Utility, cost, regret (problem-dependent)
- Proper scoring rules: CRPS, QS (universal ranking) [3]
- Specific regions of interest? [4]

Many moving pieces:
- *What* to combine: PDFs, CDFs, or quantiles?
- *How* to combine: Linear or nonlinear
- *Why* are we combining: Utility or CRPS?

Forecast combinations should aim to be "sophisticatedly simple." [5]

Probabilistic forecast combination with *optimization-informed* linear pooling
- Weights optimized for downstream decisions
- Adaption to contextual information

[3] Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association 102, 359–378.
[4] Gneiting, T., Ranjan, R., 2013. Combining predictive distributions.
[5] Wang, X., Hyndman, R.J., Li, F., Kang, Y., 2023. Forecast combinations: An over 50-year review. International Journal of Forecasting 39, 1518–1547.

Methodology

Mathematical Background

Contextual Stochastic Optimization with Multiple Forecasts

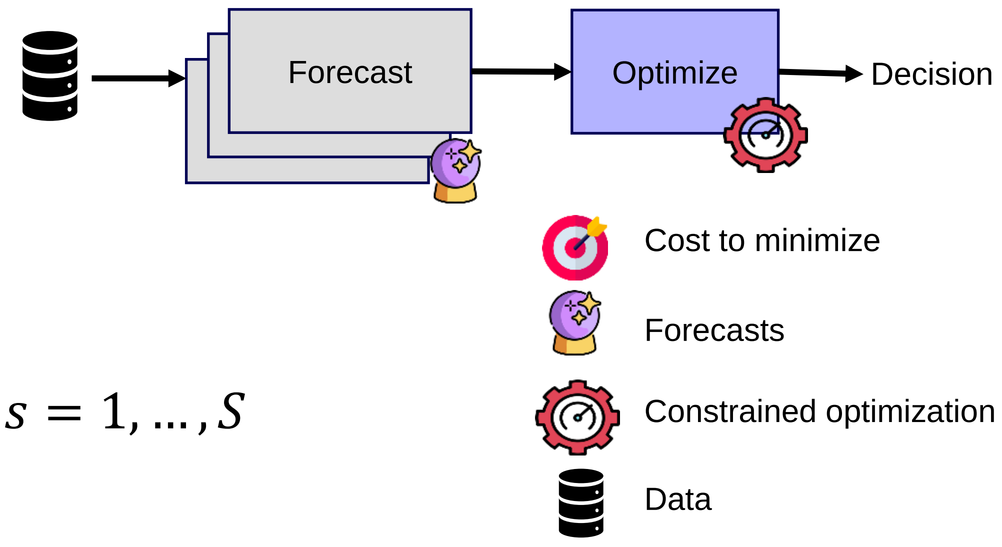
$$\min_{z \in Z} \mathbb{E}_Y[c(z;Y) \mid X = x_0] = \min_{z \in Z} \mathbb{E}_{Y \sim p^a}[c(z;Y)]$$

$\mathrm{c(z)}$: Convex cost

z: Decisions within feasible set Z

Y: Uncertainty following conditional distribution $Y \sim p^a$

$X = x_0$: contextual information (weather, etc.)



S vendors provide us with probabilistic forecasts $p_s, \ s = 1, \dots, S$

- Linear pool: $p^{\text{comb}} = \sum_s \lambda_s \, p^s$, where $\lambda_s \geq 0, \sum_s \lambda_s = 1$

Data: $\left\{\left(\mathrm{y_i}, \mathrm{p_i^1}, \dots, \mathrm{p_i^S}, \mathrm{x_i}\right)\right\}_{\mathrm{i=i}}^{n}$

Forecast evaluation:

- Quality: Average CRPS over n observations
- Regret: Incurred decision cost due to error

$R(z(p), y) = c(z(p), y) - c(z^*, y)$, where z^* the perfect foresight decision

Linear Pooling: Does it Work for Decisions?
Hedging Against the Worst-case

Motivating the linear pooling [6] (wisdom of crowds):

- For any combination weights and any distribution, we get an upper bound on our disappointment for any distribution

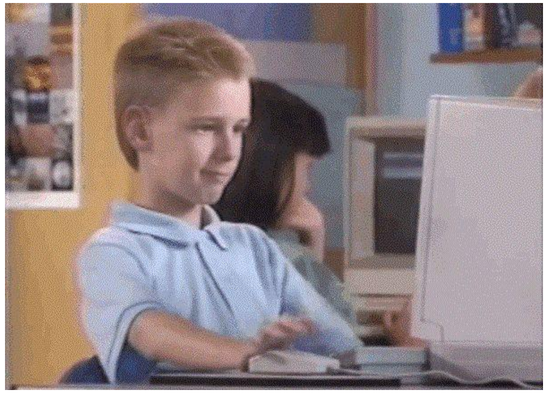
$$\mathbb{E}\left[R\left(z\left(p^{\text{comb}}\right); p^{a}\right)\right] \leq \max_{s=1,\ldots,S} \mathbb{E}[R(z(p^{s}); p^{a})]$$

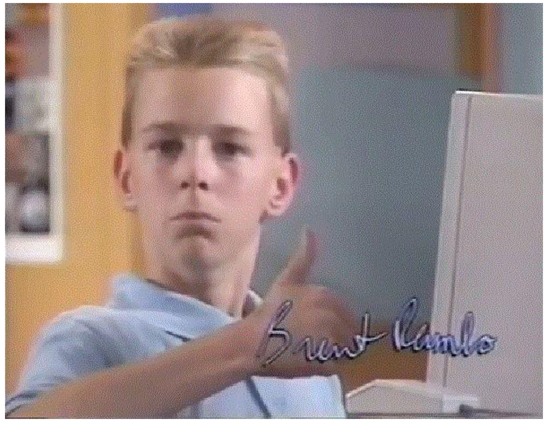
- Expected regret of $z\left(p^{\text{comb}}\right)$ is smaller than the average regret of $z(p^{s})$

What does this mean:

- Hedge decision risk against the **worst-case** component forecast
- Models are often (awfully) wrong
- Linear pooling is particularly useful when models are misspecified

Q: How to select the combination weights λ_s?





[6] Stone, M., 1961. The linear opinion pool. Ann. Math. Statist 32, 1339–1342.

Decision-focused Linear Pooling

Finding Optimal Weights that Minimize Expected Regret

Regularized loss

$$\min_{\theta} \frac{1}{n} \sum_{i \in [n]} R\left(z_i^{\text{comb}}; y_i\right) + \gamma \text{CRPS}(p_i^{\text{comb}}, y_i)$$

$$\text{s.t.} \quad z_i^{\text{comb}} \in \text{argmin}_{z \in Z} \mathbb{E}_{p_i^{\text{comb}}}[c(z; Y)], i \in [n]$$

Induced decisions

$$p_i^{\text{comb}} = \sum_s \lambda_{s,i} p_i^s, i \in [n]$$

Linear pool

$$\boldsymbol{\lambda}_i = f(x_i; \theta), i \in [n]$$

Adapting weights to contextual information

Model parameters

Decision-focused Linear Pooling
Solution Methodology

Optimal weights:

$$\min_{\theta} \frac{1}{n} \sum_{i \in [n]} R\left(z_i^{\text{comb}}; y_i\right) + \gamma \text{CRPS}\left(p_i^{\text{comb}}, y_i\right)$$

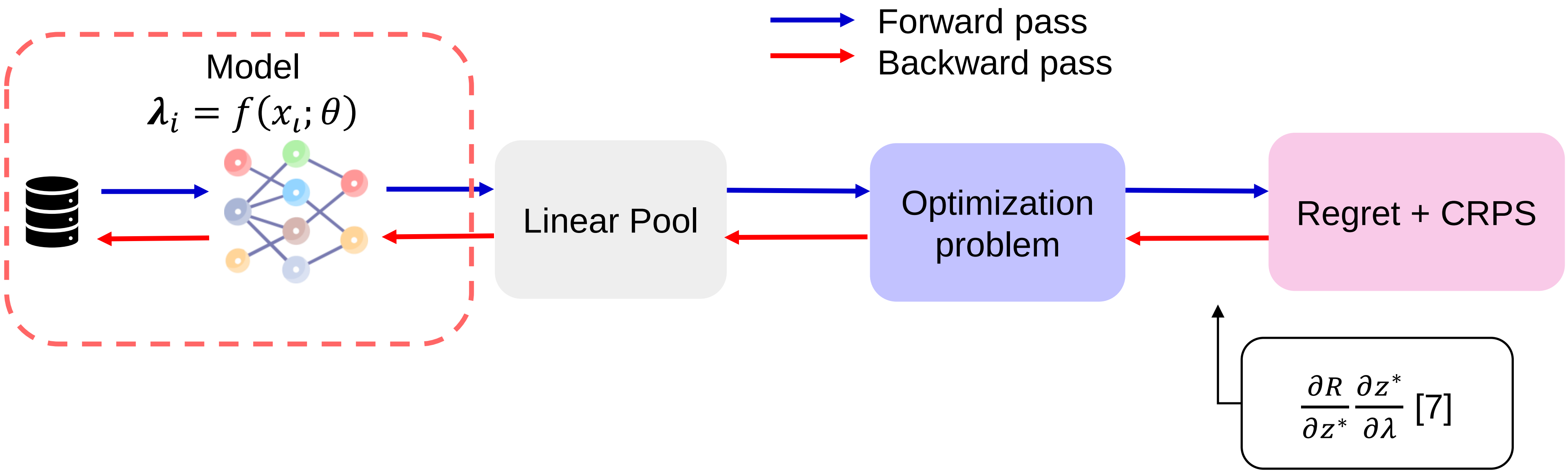
$$\text{s.t.} \quad z_i^{\text{comb}} \in \operatorname{argmin}_{z \in Z} \mathbb{E}_{p_i^{\text{comb}}}[c(z; Y)], i \in [n]$$

$$p_i^{\text{comb}} = \sum_s \lambda_{s,i} p_i^s, i \in [n]$$

$$\boldsymbol{\lambda}_i = f(x_i; \theta), i \in [n]$$

Performance-based Inverse Weighting:

- Find average, in-sample regret r_s
- Set $\lambda_s = \frac{r_s}{\sum r_s}$



[7] Agrawal, Akshay, et al. "Differentiable convex optimization layers." *Advances in Neural Information Processing Systems* 32 (2019).

Numerical Experiments & Conclusions

Numerical experiments
Experimental Setup

Component forecasts:
- Non-parametric ML methods to weight historical observations y_i
- Weighted Sample Average Approximation: $\min_{z \in Z} \sum_{i \in [n]} \omega_i(x_0)\, c(z; y_i)$
- Models: $k\mathrm{NN}, \mathrm{CART}, \mathrm{Random\ Forest\ (RF)}$

Combination methods:
- OLP: Ordinary linear pool with uniform weights
- CRPSL: Weights that minimize CRPS
- $\mathrm{DFL} - \gamma$: Decision-focused weights, with γ regularization
- invW: Inverse weighting based on in-sample regret

Evaluation metrics:
- CRPS (quality)
- Regret (value)

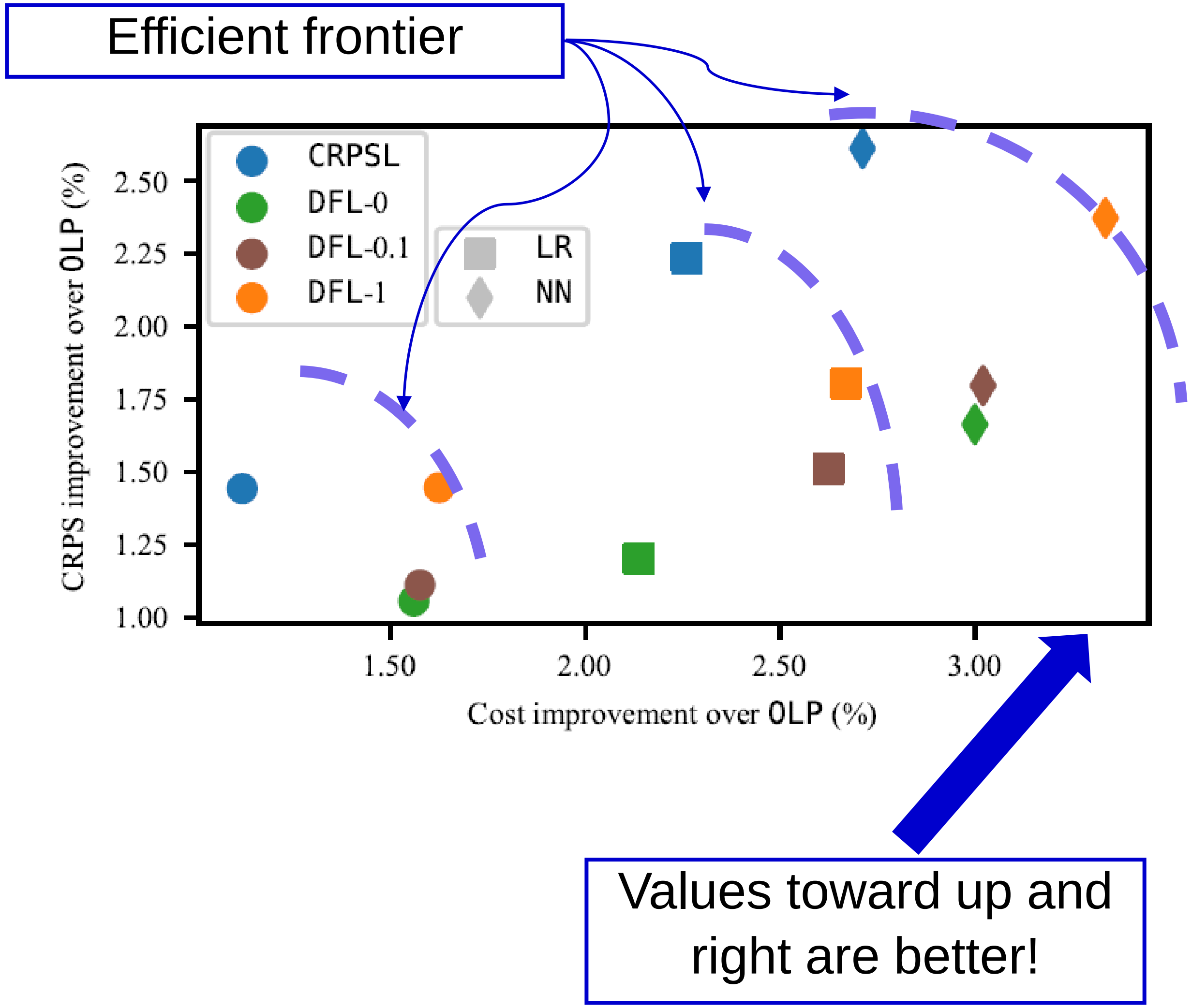
Numerical experiments
Solar Forecasting and Trading in Electricity Markets

Setting: Solar aggregator participates in a day-ahead electricity market
Problem: Regularized newsvendor

$$c(z;Y) = (1-\rho)\max\left(\frac{\tau}{1-\tau}(Y-z), z-Y\right) + \rho(Y-z)^2$$

Key results:

- $\mathrm{DFL}-\gamma$ leads to lower regret (trading costs), CRPSL leads to higher CRPS
- A combination ($\gamma = 1$) brings the best of both worlds
- Adapting the linear pooling weights improves static combinations
- Relative ranking of combination methods holds for conditional combinations

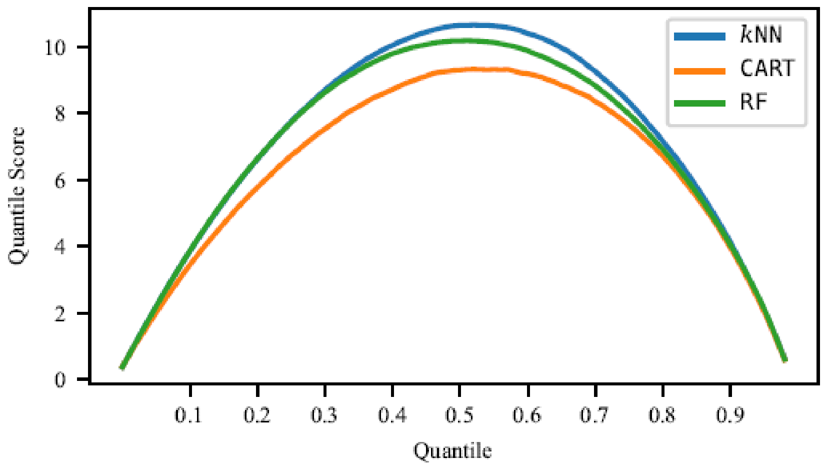


Numerical experiments
Wind Forecasting and Grid Scheduling

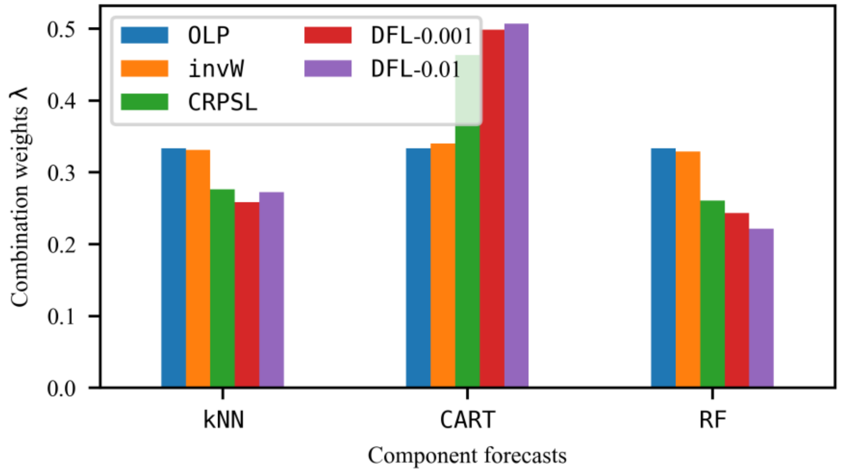
Setting: Schedule G generators under net demand uncertainty

Problem: Two-stage LP with fixed recourse

$$
\begin{aligned}
\min_{\mathbf{z},\mathbf{z}_k^u,\mathbf{z}_k^d} \quad & \mathbf{c}^\top \mathbf{z} + \sum_{k\in[K]} p_k(\mathbf{c}^{u\top}\mathbf{z}_k^u - \mathbf{c}^{d\top}\mathbf{z}_k^d), & \\
\text{s.t.} \quad & \mathbf{1}^\top(\mathbf{z} + \mathbf{z}_k^u - \mathbf{z}_k^d) = \xi_k, & k \in [K], \\
& \mathbf{0} \le \mathbf{z} \le \overline{\mathbf{z}}, & \\
& \mathbf{0} \le \mathbf{z}_k^u \le \min(\overline{\mathbf{z}}^u, \overline{\mathbf{z}} - \mathbf{z}), & k \in [K], \\
& \mathbf{0} \le \mathbf{r}_k^d \le \min(\overline{\mathbf{z}}^d, \mathbf{z}), & k \in [K].
\end{aligned}
$$



Pinball loss for component forecasts



Learned weights $\boldsymbol{\lambda}$ of the combination methods

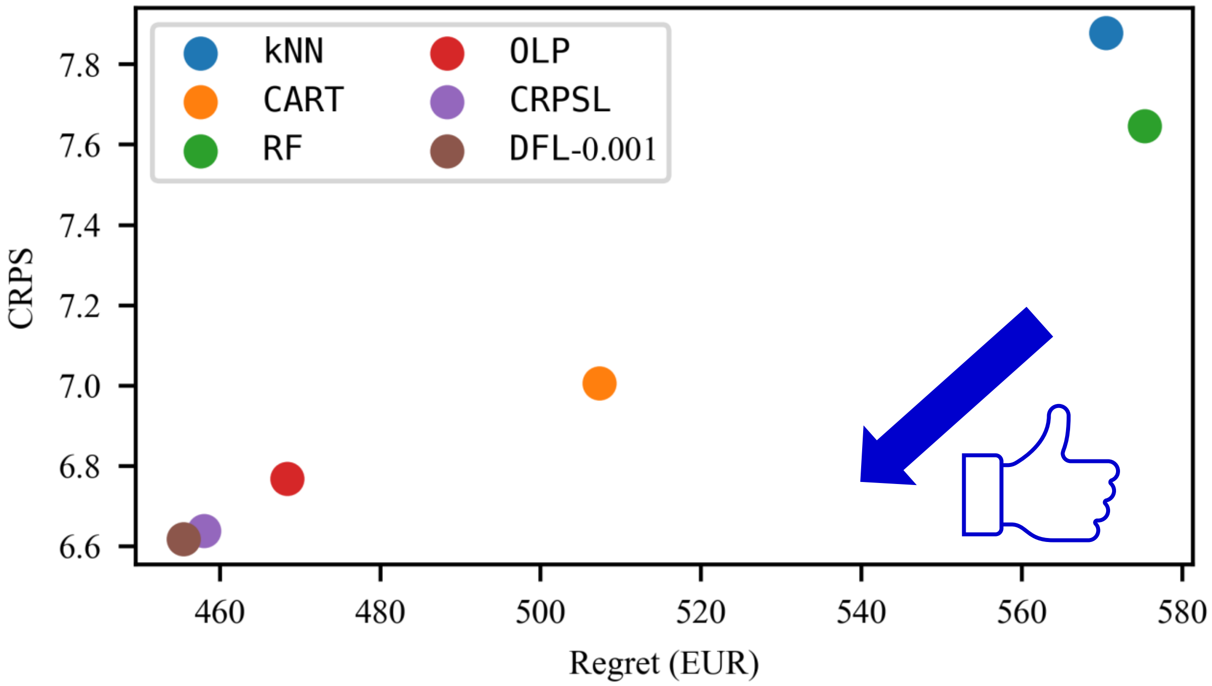
Numerical experiments

Wind Forecasting and Grid Scheduling

Setting: Schedule G generators under net demand uncertainty

Problem: Two-stage LP with fixed recourse

$$
\begin{aligned}
\min_{\mathbf{z},\mathbf{z}_k^u,\mathbf{z}_k^d} \quad & \mathbf{c}^\top \mathbf{z} + \sum_{k\in[K]} p_k(\mathbf{c}^{u\top}\mathbf{z}_k^u - \mathbf{c}^{d\top}\mathbf{z}_k^d), & \\
\text{s.t.} \quad & \mathbf{1}^\top(\mathbf{z}+\mathbf{z}_k^u-\mathbf{z}_k^d) = \xi_k, & k\in[K],\\
& \mathbf{0} \le \mathbf{z} \le \overline{\mathbf{z}}, & \\
& \mathbf{0} \le \mathbf{z}_k^u \le \min(\overline{\mathbf{z}}^u, \overline{\mathbf{z}}-\mathbf{z}), & k\in[K],\\
& \mathbf{0} \le \mathbf{r}_k^d \le \min(\overline{\mathbf{z}}^d, \mathbf{z}), & k\in[K].
\end{aligned}
$$



- Combination $\gg$ component forecasts
- Regret-CRPS combination ($\gamma > 0$) set the best decision-accuracy trade-off

Key Takeaways

Forecast combination to reduce decision costs:

- Approximately 3% reduction in decision costs with decision-focused linear pooling
- Minimizing a combination of decision regret and CRPS combined the best of both worlds

Practical suggestion:

- Generate forecasts for accuracy, combine them for decision-making

Next steps:

- Scalability and computational challenges
- Quantile averaging and multivariate uncertainty

Preprint: https://hal.science/hal-04593114/document
Code: https://github.com/akylasstrat/df-forecast-comb/tree/main
Mail: a.stratigakos@imperial.ac.uk
LinkedIn: https://www.linkedin.com/in/akylas-stratigakos/

Thank you!