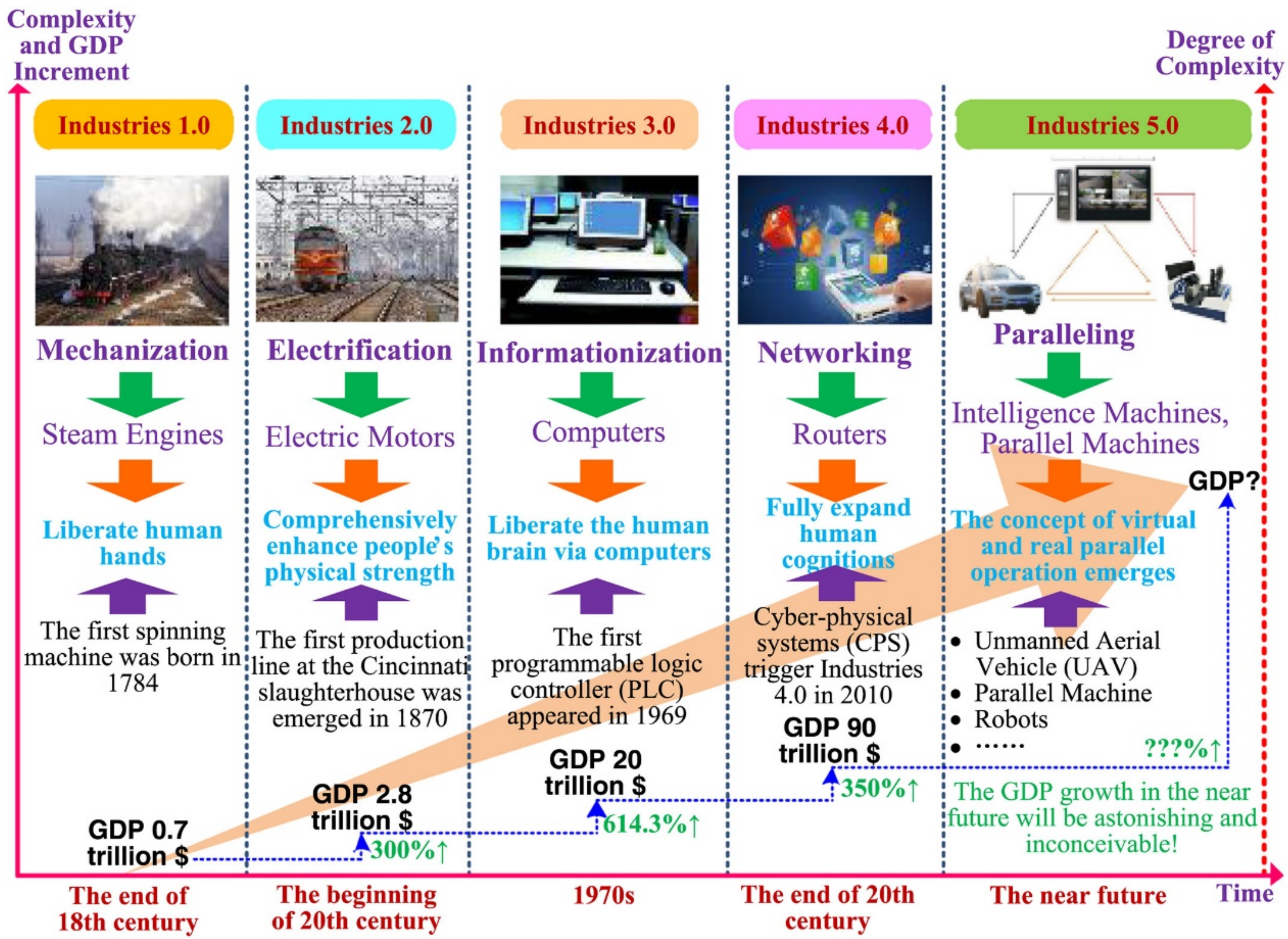
Power System Applications of Artificial Intelligence (AI)

Ian A. Hiskens
Vennema Professor of Engineering
Professor, Electrical Engineering and Computer Science
University of Michigan, Ann Arbor

MichiganEngineering

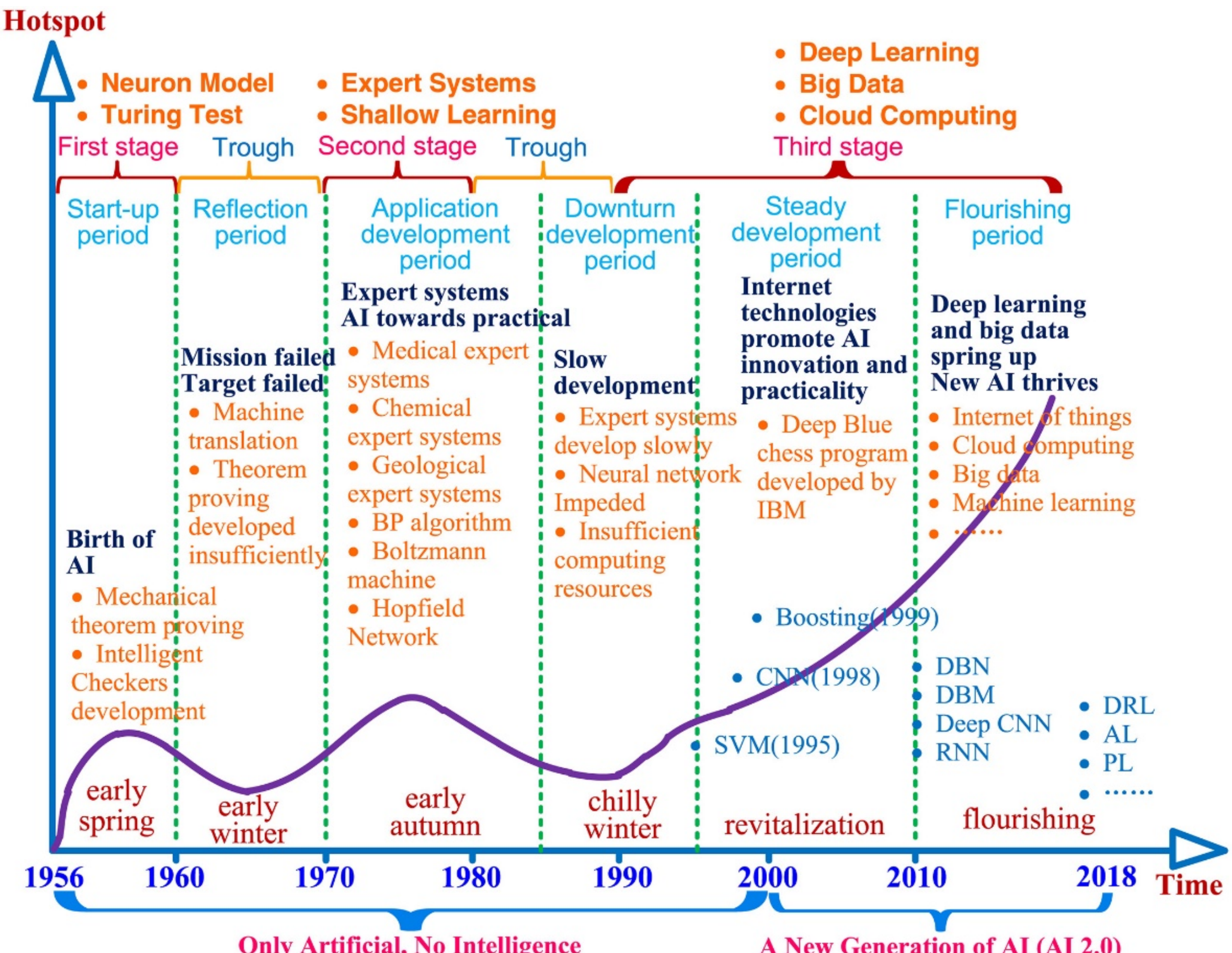
Adelaide Power System
Summer School
February 10-13, 2020

Industrial development



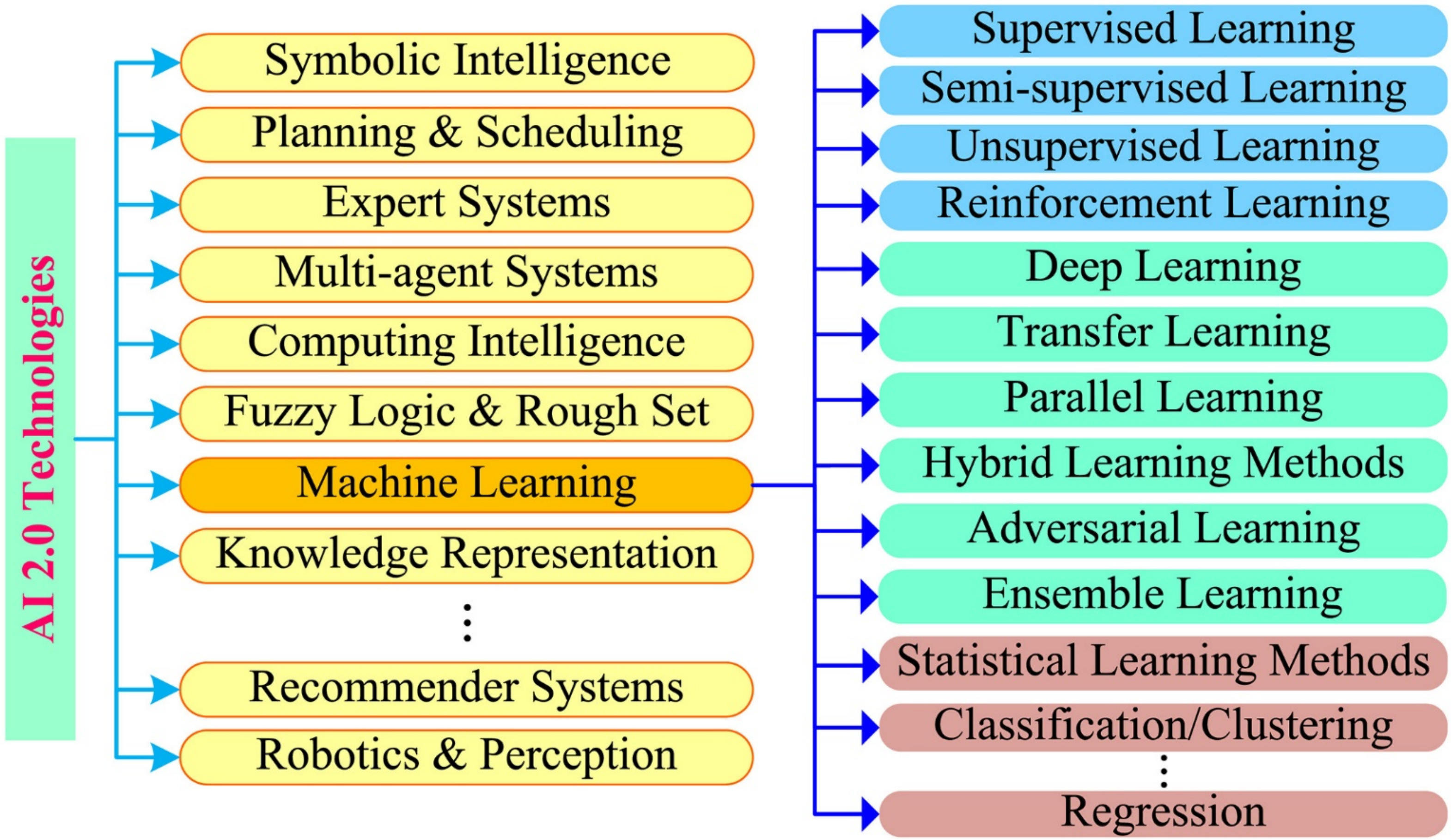
From Cheng, Yu

The evolution of AI



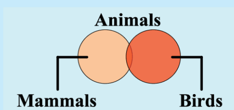
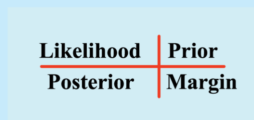
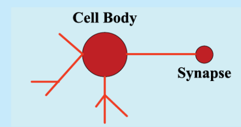
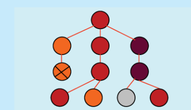
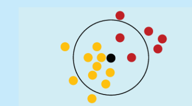
From Cheng, Yu

Current research in AI



From Cheng, Yu

Basis of machine learning

	Tribes				
Items	**Symbolists**	**Bayesians**	**Connectionists**	**Evolutionaries**	**Analogizers**
Origin	Logic theory, philosophy	Statistics	Neuroscience	Evolutionary biology	Psychology
Graphical representation					
Central idea	Cognition is computing use symbols, rules, and logic to represent knowledge and draw logical inference (prediction of results).	Assess the likelihood of occurrence for probabilistic inference, modify the occurrence probability, and make optimal decisions.	Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons (simulate the brain).	Generate variations and then assess the fitness of each for a given purpose with genetic algorithm (GA) and genetic programming.	Optimize a function in light of constraints (similarity between old and new knowledge).
Research issue	Knowledge structure	Uncertainty	Credit assign	Structural discovery	Similarity
Favored algorithm	Rules and decision trees	Naive Bayes or Markov	Neural networks	Genetic programs	Support vectors
Representative algorithm	Inverse deduction algorithm	Probabilistic reasoning	BP algorithm and deep learning	Genetic programming	Kernel machine, nearest neighbor algorithm
Representative application	Knowledge map	Anti-spam, probabilistic prediction	Machine vision, speech recognition	Starfish robot	Netflix recommender system
Representative figure	Tom Mitchell, Steve Muggleton, and Ross Quinlan	David Heckerman, Judea Pearl, and Michael Jordan	Yann LeCun, Geoff Hinton, and Yoshua Bengio	John Koda, John Holland, and Hod Lipson	Peter Hart, Vladimir Vapnik, and Douglas Hofstadter

From Cheng, Yu

Power system applications of AI (1)

- Load forecasting (short- and long-term).
- Short-term forecasting of wind and solar.
- Fault diagnosis and condition monitoring.
  - Generators, transformers, high-voltage circuit breakers, power electronic converters.
- Security/reliability assessment.
  - Moving beyond *N-1* reliability.
  - Decision-making at scale and high complexity.
  - Stability assessment.
- Identifying cyber and physical attacks.
  - Intrusion detection in power system information networks.
- Electricity markets.
  - Determine bidding strategies for energy markets.
  - Assess market power.

Power system applications of AI (2)

- Feedback controls.
  - Generator and motor controls.
  - Voltage/reactive power controls.
  - Automatic generation control (AGC).
  - Microgrids and multi-energy systems.
  - Estimating ensemble characteristics.
  - Baseline estimation.
- Consumers.
  - Thermostats that learn occupancy patterns.
  - Other applications and opportunities?

Power system data

- Vast numbers of participants create vast amounts of highly distributed data.
- Heterogeneous (multiple forms, multiple time scales).
- Asynchronous.
- Noisy.
- Incomplete.
- Low value density.

- Difficulties in collecting, storing, processing, mining.
- Privacy concerns.

Case studies

- **Reliability assessment.**
- Baseline estimation.
- Estimating ensemble characteristics.
- Cluster-based chance-constrained optimization.
- Corrective model predictive control.
  - Not actually AI but a competing technology.

Reliability assessment (1)

- This section is based on notes from Louis Wehenkel.
- Requirement:
  - At sub-second temporal resolution, balance generation, consumption and storage, whilst satisfying network constraints, in spite of various threats.
- Threats:
  - Unanticipated variations of generation and/or demand, weather conditions.
  - Component failures, human errors, adversarial attacks.
- Problems to avoid:
  - Component overloads, voltage and/or frequency excursions.
  - Cascading overloads, instabilities, blackouts.
- Aim:
  - Optimization and control closer to real-time.

Reliability assessment (2)

Every 5 minutes, for the real-time system state x_{rt} , assess the risk induced by contingencies.

- Based on data and models:
  - $C(x_{rt}), \pi_c(x_{rt}, c)$: set of contingencies and their probabilities.
  - $f_{cr}(x_{rt}, c)$: measure of the severity of contingency c in state x_{rt}.
- Assess the expected impact of possible contingencies:
  - $\mathbb{E}\{f_{cr}|x_{rt}\} = \sum_{c \in C(x_{rt})} \pi_c(x_{rt}, c) f_{cr}(x_{rt}, c)$

    - expected cost of service interruptions.
  - $\mathbb{P}\{f_{cr} > \eta | x_{rt}\} = \sum_{c \in C(x_{rt})} \pi_c(x_{rt}, c) 1\big(f_{cr}(x_{rt}, c) > \eta\big)$

    - probability of large service interruptions.

Machine learning

- Background:
  - The evaluation of the contingency response function $f_{cr}(x_{rt}, c)$ is generally computationally expensive.
  - Even so, this function is evaluated as often as possible by the TSO, yielding growing datasets $D = \{(x_{rt}^i, c^i), f_{cr}(x_{rt}^i, c^i)\}_{i=1}^{\cdots}$
- Supervised machine learning:
  - From a sample of input-output pairs $\{(z^i, y^i)\}_{i=1}^{n}$, we can learn a function $h(\cdot)$ such that $|h(z) - y|$ is small on average.
- Application to real-time reliability assessment:
  - Learn a regression proxy: $h_{regr}(x_{rt}, c) \approx f_{cr}(x_{rt}, c)$
  - Learn a classifier proxy: $h_{class}(x_{rt}, c) \approx 1\big(f_{cr}(x_{rt}, c) \geq \eta\big)$
- The underlying assumptions are:
  - h-proxies are much faster to evaluate than $f_{cr}(x_{rt}, c)$.
  - It is possible to learn sufficiently accurate h-proxies.

Machine learning considerations

- How often to apply ML to refresh the proxies.
  - On the fly in real-time.
  - Ahead of time.
- How to gather the datasets used for learning the proxies.
  - Passively, by exploiting data generated by the EMS.
  - Actively, by using Monte-Carlo approaches.
- How to best use the available ML techniques.
  - Interpretability.
  - Computational performance (learning and prediction).
- How to use the learned proxies $h_{r,c}$.
  - Stand-alone mode.
  - Together with “exact” simulator of f_{cr} .

Case studies

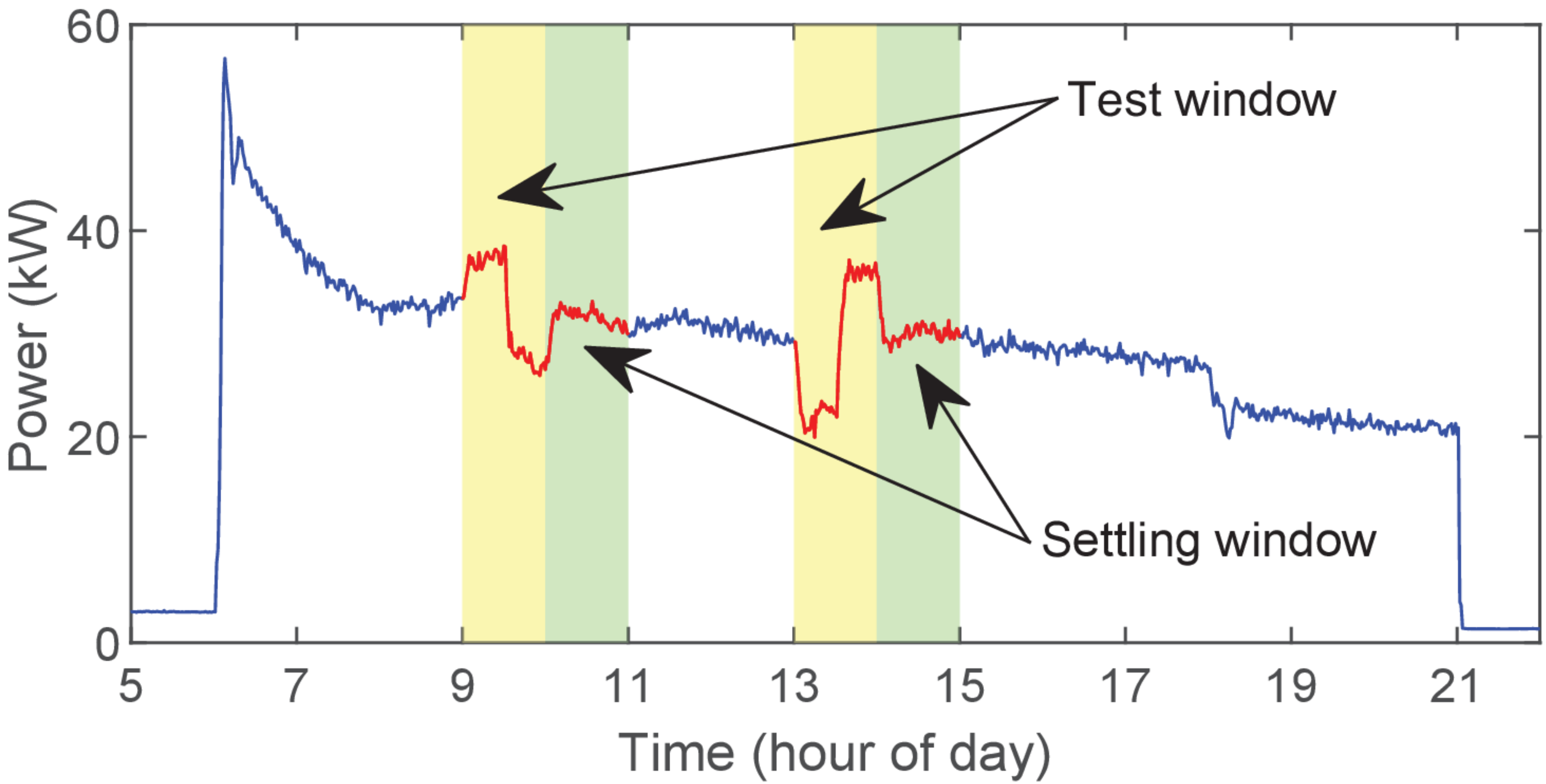
- Reliability assessment.
- **Baseline estimation.**
- Estimating ensemble characteristics.
- Cluster-based chance-constrained optimization.
- Corrective model predictive control.

Baseline estimation

- The thermal mass of commercial buildings allows their heating, ventilation and air conditioning (HVAC) systems to be perturbed (slightly) without affecting occupant comfort.
- This variability can be exploited to provide power system ancillary services such as (slow) regulation.
- A critical challenge in implementing such demand response (DR) is the estimation of the power consumption that would have occurred if there had been no DR action.
  - This is needed to determine if the requested control action was enacted, for financial settlement.
  - Referred to as the counterfactual baseline.

Fan power control

- Our focus in on controlling the HVAC fan power.

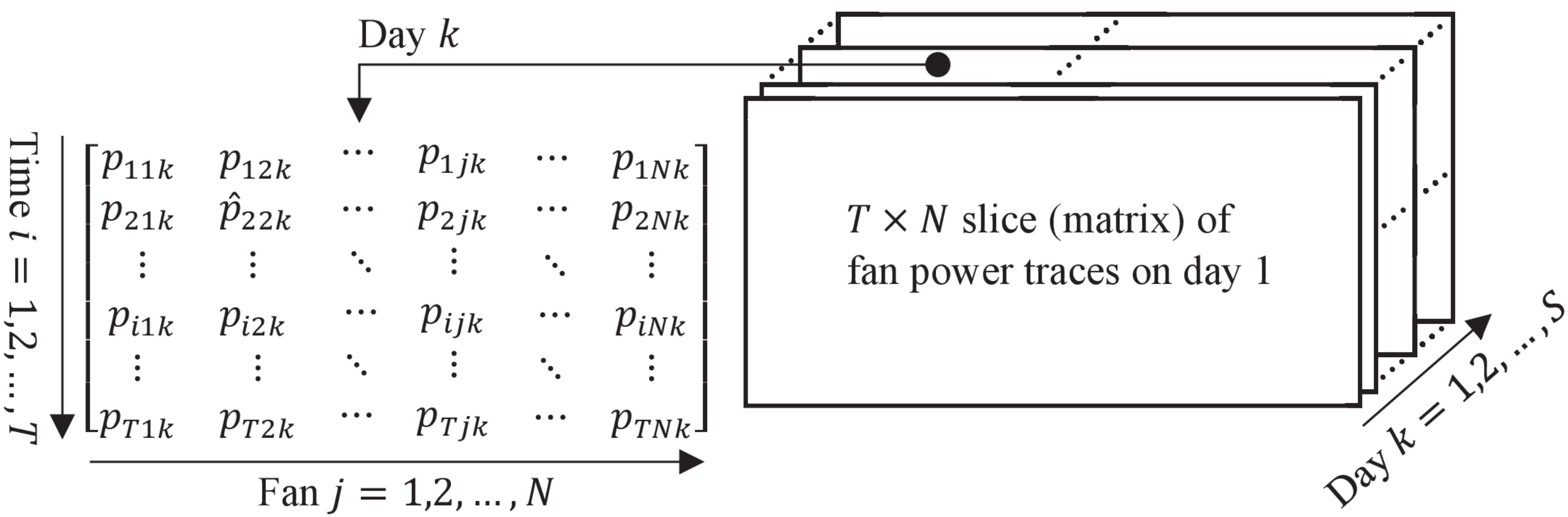


Tensor completion

- Tensor decomposition can be used for baseline estimation.
  - Tensor decomposition is an unsupervised data analysis method that can find dominant patterns across multiple dimensions, e.g. time, fan and day.
  - Tensor completion is the closely related problem of imputing missing or unobserved entries of a tensor.
  - Our approach to tensor completion uses generalized canonical polyadic (GCP) tensor decomposition, but other approaches may be just as useful.

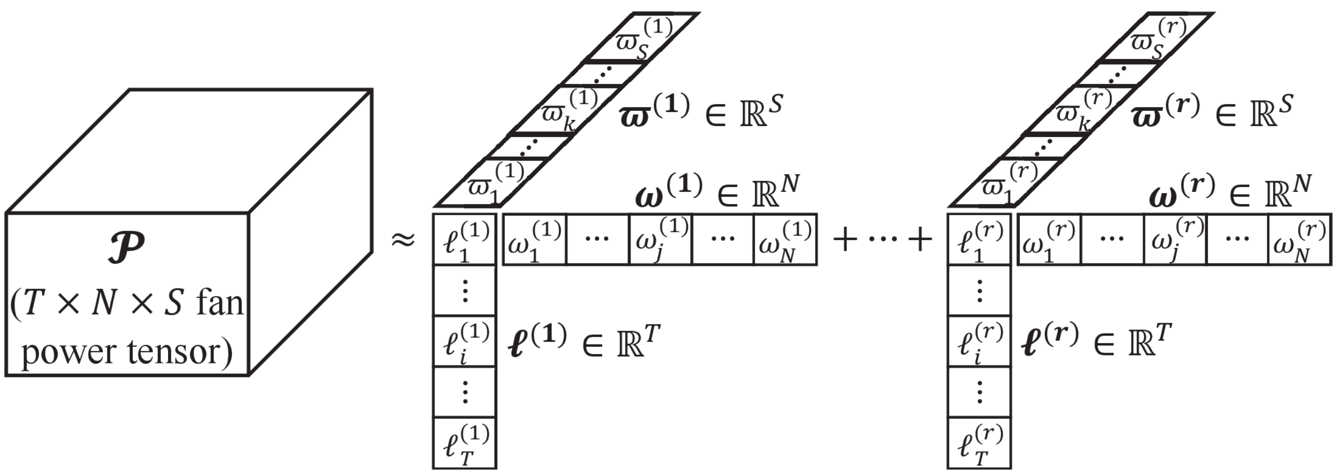
Tensor formation

- Fan power is arranged in a three-way $\mathrm{time} \times \mathrm{fan} \times \mathrm{day}$ tensor.
  – Correlations across fans and days become naturally expressed as patterns across each of the three modes (time, fan and day) and can be captured by tensor decomposition.



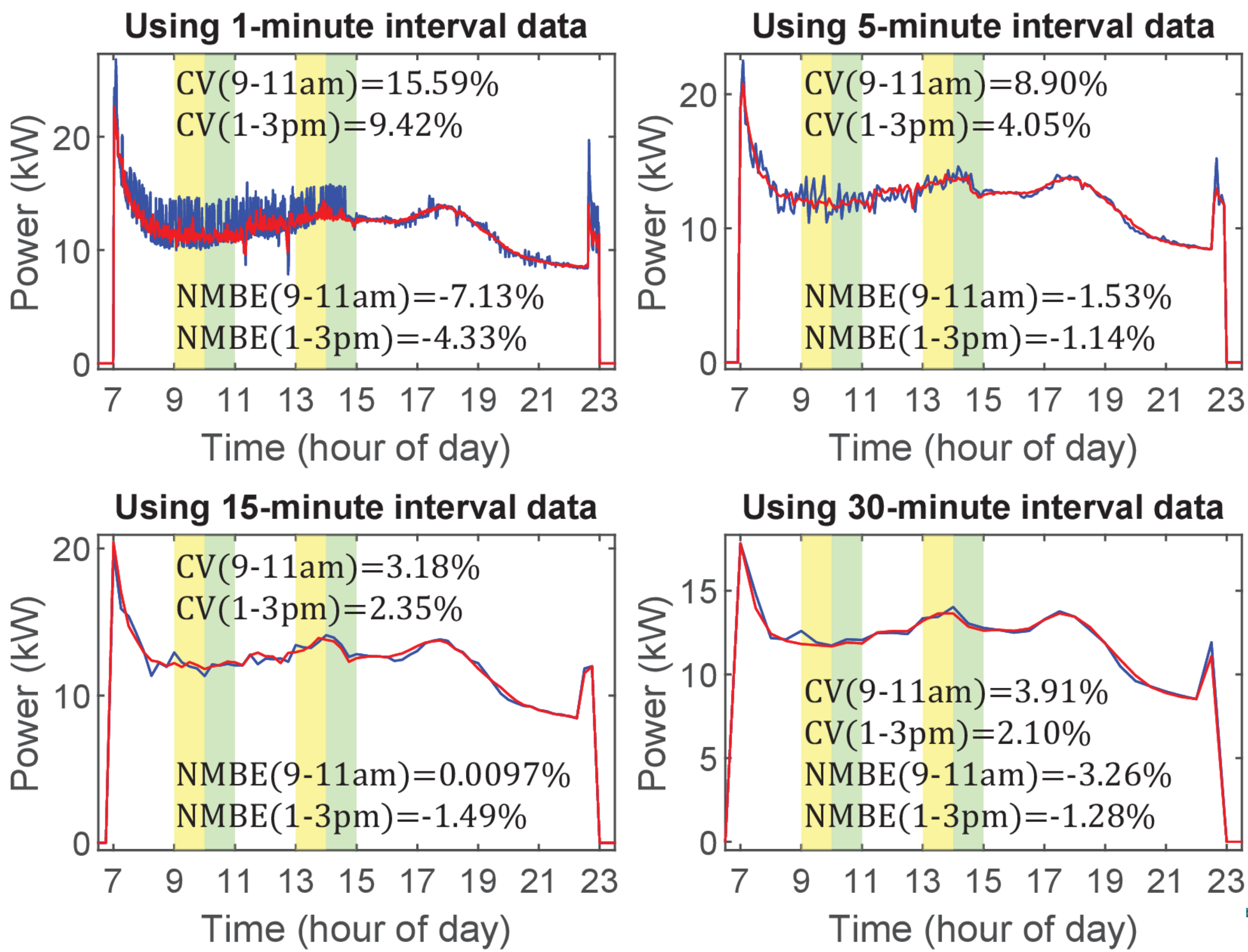
Low rank approximation

- Data from $S-1$ baseline days and 1 event day are included in the tensor.
- The baseline power within the event window of the event day (to be estimated) is treated as missing measurements.
- These tensor entries are imputed by approximating the known entries with a low-rank tensor.
  - This low-rank approximation is the sum of *r* outer products.



Example

- The vectors forming the approximation capture the dominant underlying patterns.



Case studies

- Reliability assessment.
- Baseline estimation.
- **Estimating ensemble characteristics.**
- Cluster-based chance-constrained optimization.
- Corrective model predictive control.

Disaggregation of demand

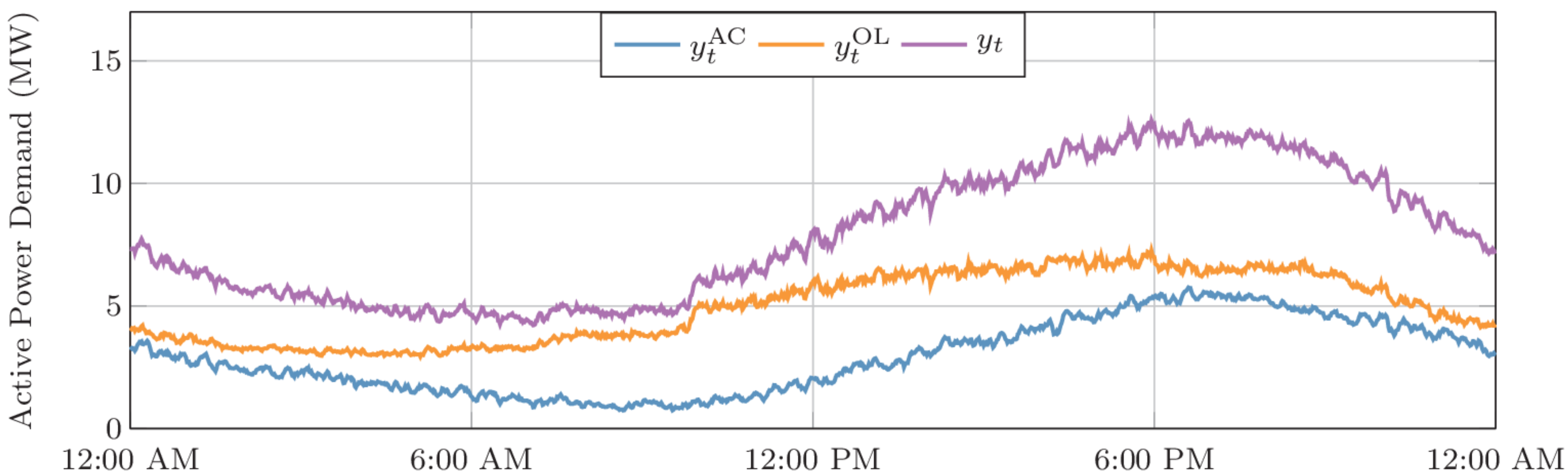
- This section is based on Greg Ledva's PhD thesis.
- What is the total aggregate power consumption of air conditioners (ACs) on a distribution feeder?
  - This information is necessary for load control schemes that use ACs.
    - AC load availability is a gain in the control loop.
  - Could provide a warning of vulnerability to "fault induced delayed voltage recovery" (FIDVR).
  - Knowing the weather-dependent proportion of the feeder load may help better predict response to weather changes.

Measurements

- Total AC power could be obtained by placing sensors on each device.
  - Expensive.
  - Requires extensive communications.
  - Privacy issues.
- Better to use existing substation SCADA metering and some knowledge of the physical system.

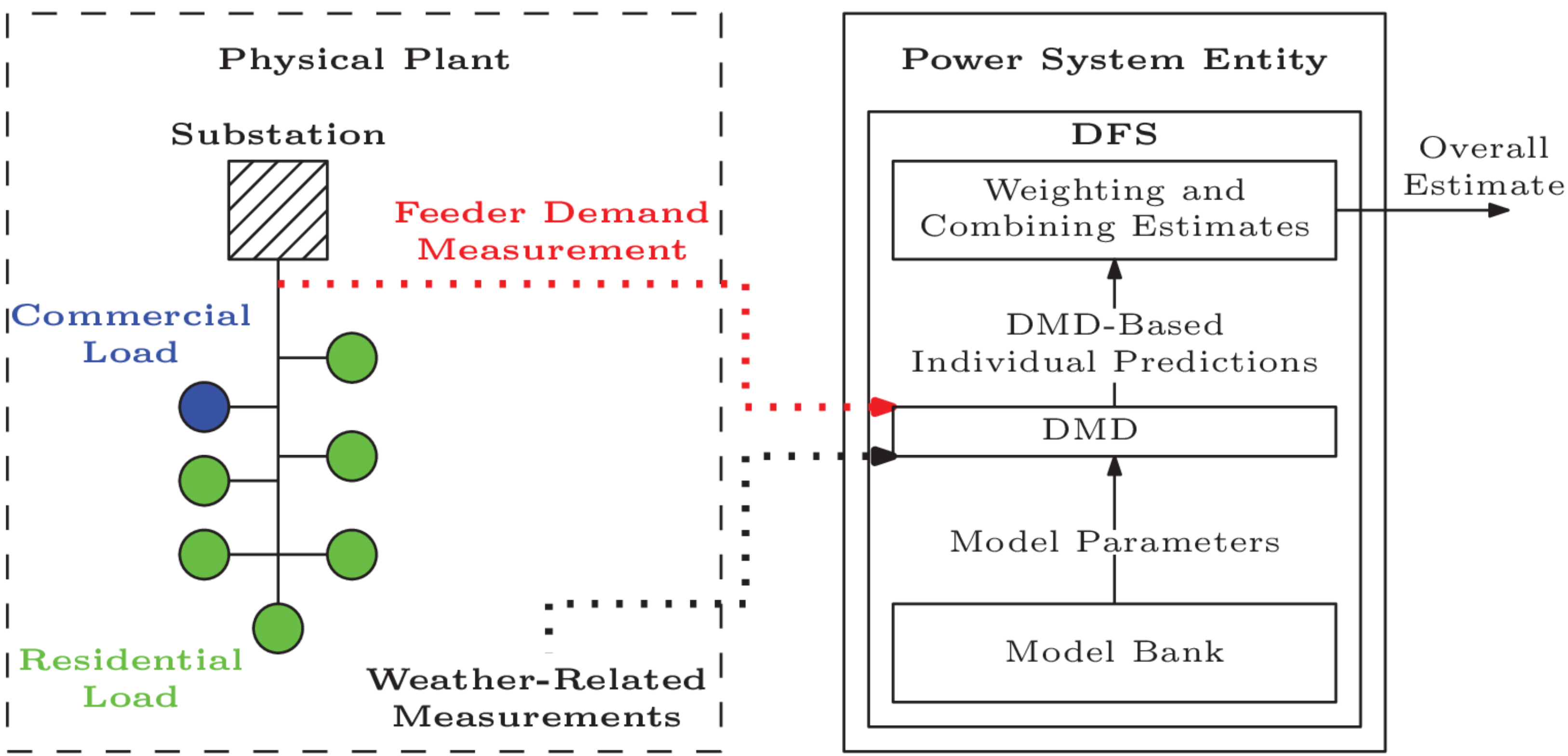
Disaggregation

- Dynamic fixed share (DFS) is used to separate feeder active power into total residential AC load and all other load (OL).
  - This feeder-level AC load is composed of a large number of small loads, each undergoing its thermostat cycling.



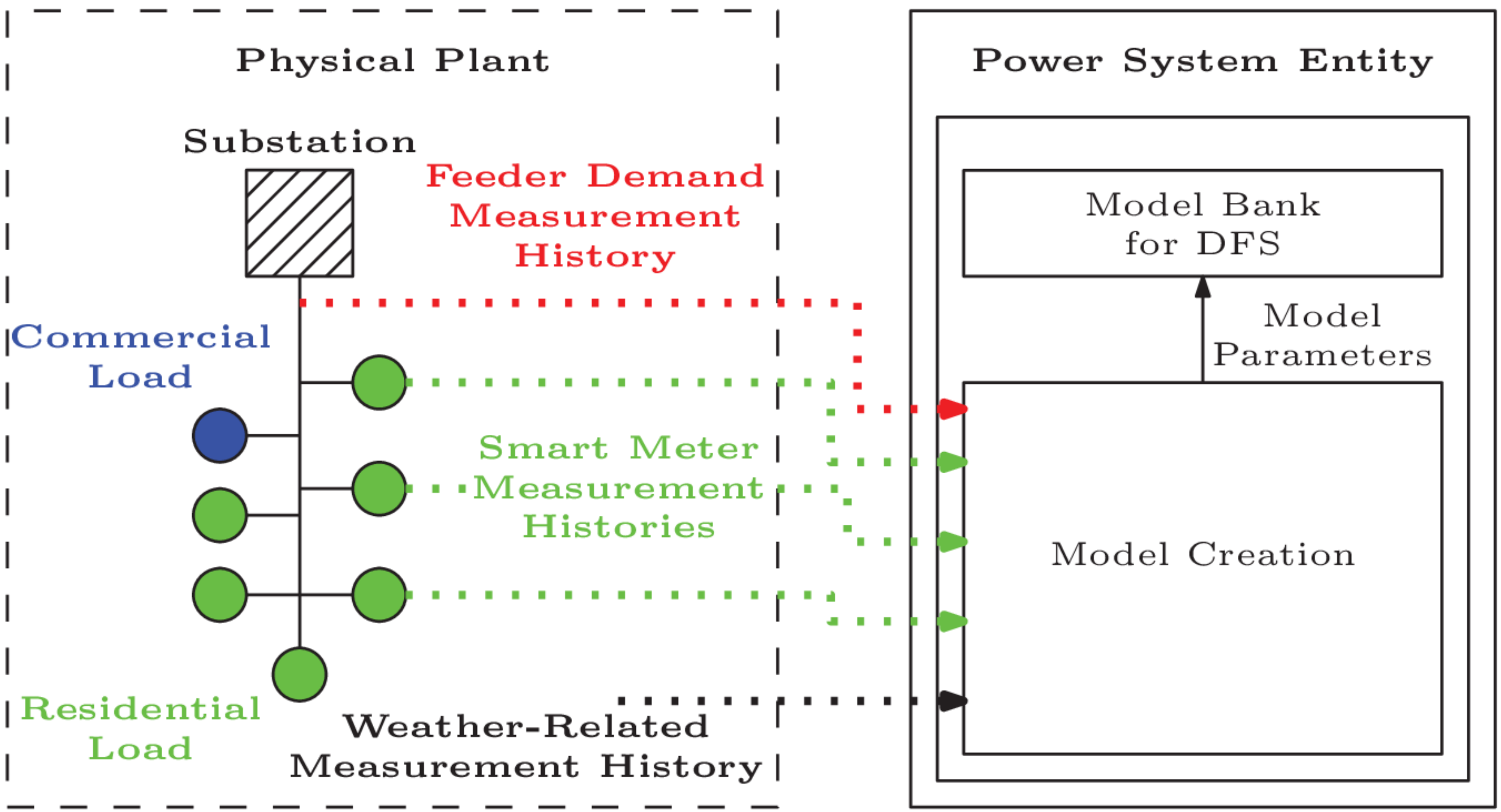
Online disaggregation

- Feeder-level disaggregation is performed online using measurement of active power that is time-averaged over 1-minute intervals.



Offline model construction

- Assume historical load data are available, and used offline to construct models.



Models

- The models can be created using a variety of techniques.
  - Use historical, device-level measurements along with historical feeder and weather data.
- Two types of regression models were used to predict OL demand: time-of-day regression and multiple linear regression.
  - The time-of-day regression is based on smoothing the OL demand of a previous day.
  - The multiple linear regression uses time of the week and outdoor temperature.

AC demand models

- Three types of models were used to predict AC demand:
  - A multiple linear regression model.
  - Linear time-invariant (LTI) system models.
  - Linear time-varying (LTV) system models.

- The multiple linear regression uses time of the week and lagged outdoor temperature raised to the powers 1…4. (Five features in total.)

LTI models

- The LTI models have the form:

$$\hat{x}_{t+1}^{LTI,m} = A^{LTI,m}\hat{x}_t^{LTI,m}$$

$$\hat{y}_t^{AC,LTI,m} = C^{LTI,m}\hat{x}_t^{LTI,m}$$

  - The state $\hat{x}_t$ is 2-dimensional, one state gives the proportion of ACs that are on and the other the proportion that are off.
  - The state transition matrix $A^{LTI,m}$ is a transposed Markov transition matrix which captures the proportion of ACs that maintain their current state or transition to the other state.
  - Different models correspond to different ambient temperatures.

LTV models

- The LTV models have the form:

$$\hat{x}_{t+1}^{LTV,m} = A_t^{LTV,m} \hat{x}_t^{LTV,m}$$

$$\hat{y}_t^{AC,LTV,m} = C_t^{LTV,m} \hat{x}_t^{LTV,m}$$

- Two models are used, based on different ways of varying A_t^{LTV} and C_t^{LTV} with (delayed) ambient temperature.

Online learning algorithm

- The dynamic fixed share (DFS) algorithm uses the dynamic mirror descent (DMD) algorithm.
- DMD uses a single model to generate predictions of the total demand, a loss function to penalize errors between the predicted and measured total demand, and a convex optimization to adjust the model.
- DFS applies DMD separately to each model and uses a weighting algorithm to associate a weight with each model, then combines the predictions into an overall estimate.

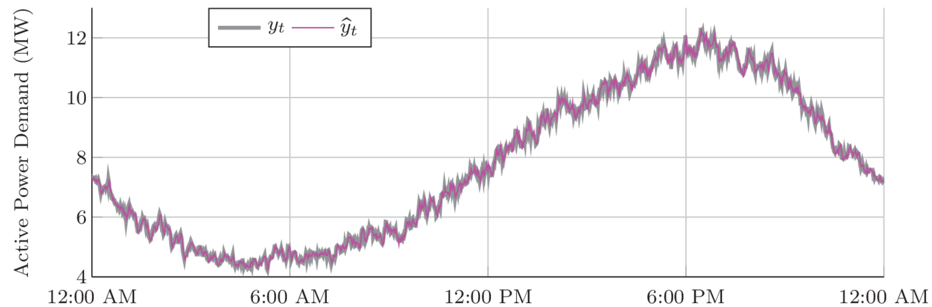
$$\tilde{\theta}_t^m = \arg\min_{\theta \in \Theta} \eta^s \langle \nabla \ell_t(\hat{\theta}_t^m), \theta \rangle + D(\theta \| \hat{\theta}_t^m)$$

$$\hat{\theta}_{t+1}^m = \Phi^m(\tilde{\theta}_t^m)$$

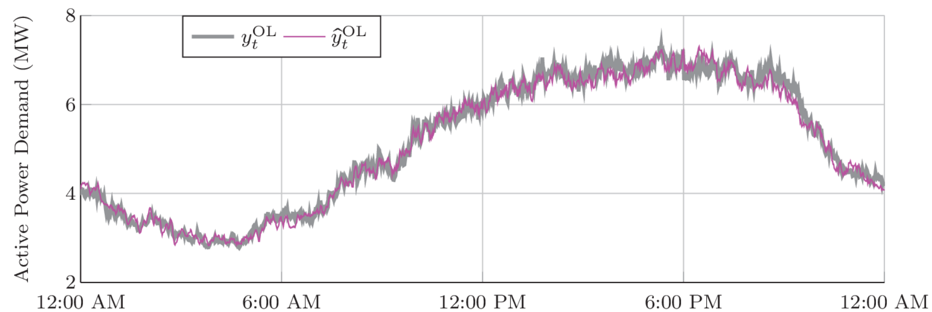
$$w_{t+1}^m = \frac{\lambda}{N^{mdl}} + (1-\lambda) \frac{w_t^m \exp\big(-\eta^r \ell_t(\hat{\theta}_t^m)\big)}{\sum_{j=1}^{N^{mdl}} w_t^j \exp\big(-\eta^r \ell_t(\hat{\theta}_t^j)\big)}$$

$$\hat{\theta}_{t+1} = \sum_{m \in \mathcal{M}} w_{t+1}^m \hat{\theta}_{t+1}^m$$

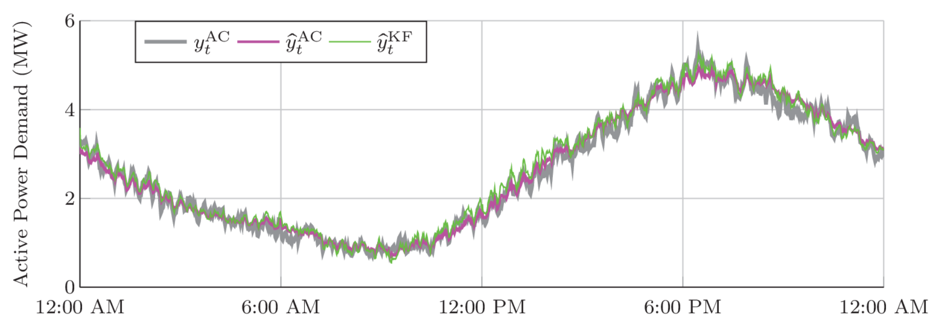
Example (1)



(a) Total Demand



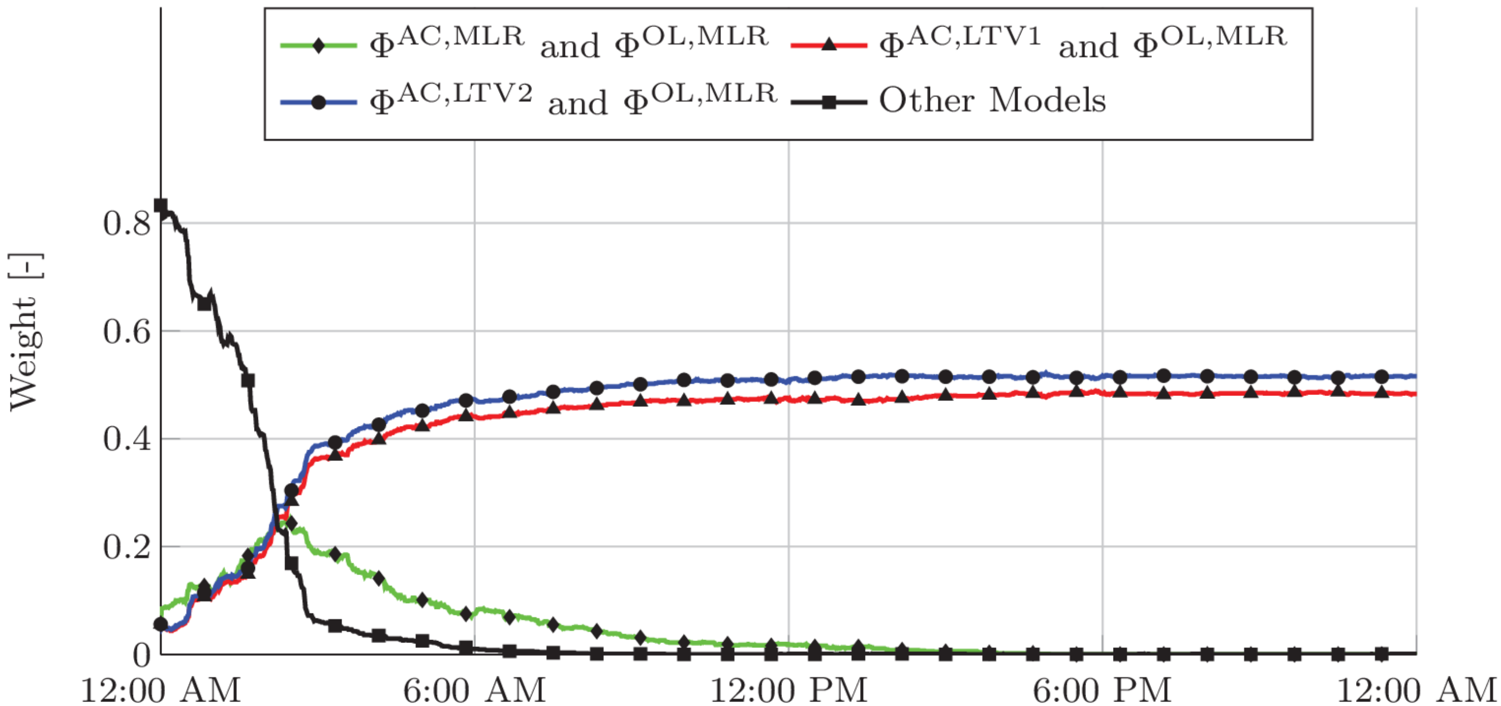
(b) OL Demand



(c) AC Demand

Example (2)

- Better prediction-measurement matching leads to larger weighting and more influence in the overall prediction.



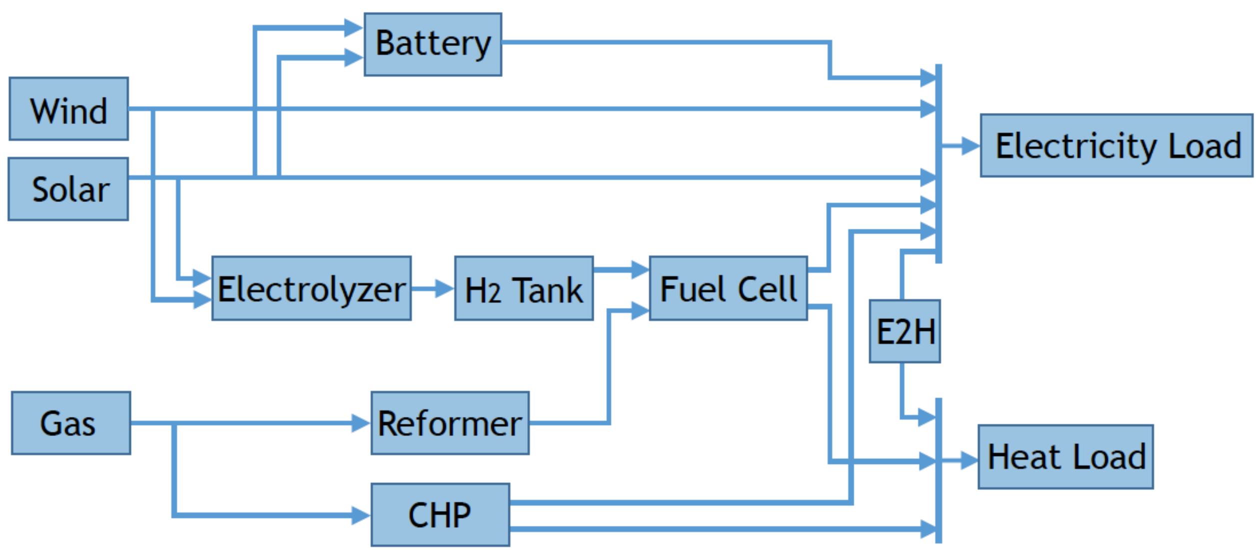
MichiganEngineering

Case studies

- Reliability assessment.
- Baseline estimation.
- Estimating ensemble characteristics.
- **Cluster-based chance-constrained optimization.**
- Corrective model predictive control.

Energy hub background

- Consider energy hubs that incorporate electricity, heat, natural gas and hydrogen.
  - No electrical connection to the distribution grid.
  - Self-powered by solar and wind resources.
  - Minimal natural gas purchase under adverse weather conditions.
  - Two forms of energy storage, batteries and hydrogen, the latter being in conjunction with electrolysis and fuel-cell conversion.



Optimization background

- The energy hub planning problem is formally expressed as a chance-constrained (CC) optimization problem, which explicitly takes into account the stochasticity of renewable generation and loads.
- The chance-constrained optimal planning problem is subsequently reformulated as a robust counterpart problem. This reformulation allows battery charging/discharging complementarity to be expressed via an equivalent linear representation.
- A cluster-based energy-hub design approach is proposed to achieve more flexible control and to better manage the trade-off between conservativeness and reliability.

CC formulation

- The uncertainty intrinsic in renewable resources requires careful treatment. To explicitly take into account such stochasticity, the optimal capacity design problem is formulated as a CC optimization problem of the form:

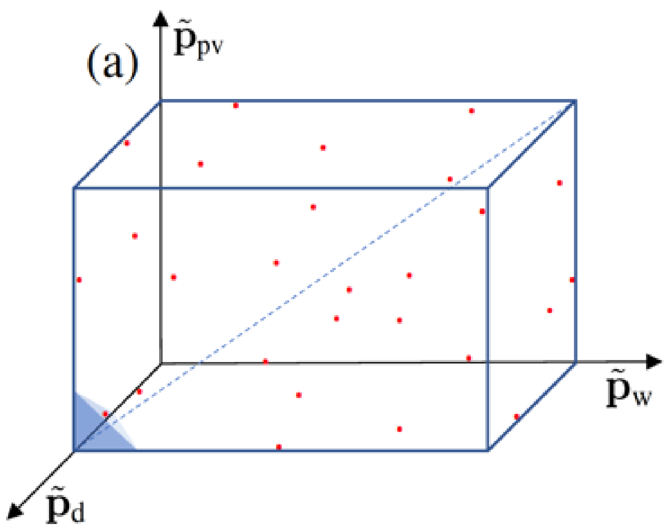
$$(\mathcal{P}0) \quad \min_{x \in \mathcal{X} \subseteq \mathbb{R}^{n_x}} \quad J(x)$$

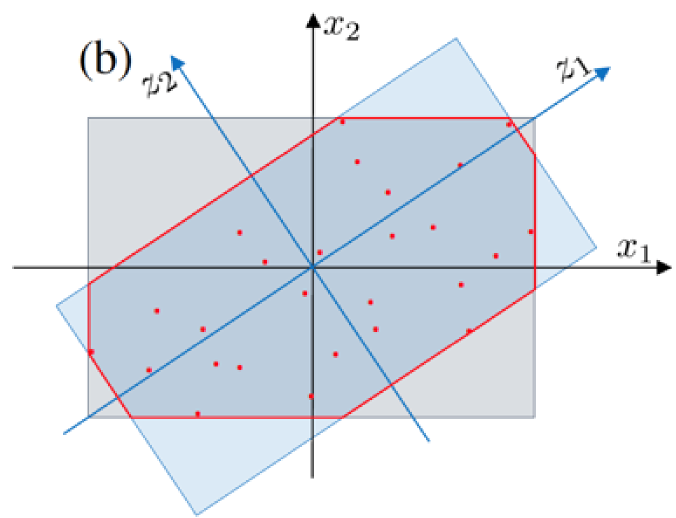
$$\text{subject to} \quad \Pr\left(\delta \in \Delta \middle| \max_{j=1,\ldots,m} g_j(x,\delta) \leq 0\right) \geq 1-\epsilon,$$

- δ : random variables.
- x: decision variables.
- ϵ : pre-specified maximal probability of violation.

Formulation

- Control policies.
  - Control policies are required to ensure hub components respond appropriately to realizations of random renewable generation infeed and load profiles.
- Robust counterpart problem formulation
  - We propose two methods to improve the robust set formulation:
    - Cutting-based approach.
    - Principal component analysis (PCA)-based approach.





Principal component analysis (PCA)

- PCA searches for a linear coordinate transformation of the original random variables, and converts the data into a new set of coordinates, i.e., the principal components (PCs).
- These PCs are uncorrelated and arranged in a descending sense, such that the first few PCs capture most of the variations in the data whereas the last few PCs describes near constant relationships in the data.
- PCA is usually used for data reduction and reconstruction, by neglecting most of the small PCs. However, we use PCA to extract directions of PCs, and use this information to guide the reshaping of the robust set and the clustering of random trajectories in the multi-policy design framework.

Multi-policy design (1)

- Conservativeness: The robust counterpart approach is conservative.
- The main reason is that the control policy for battery dispatching is not sufficiently flexible, since the affine policy has to cope with a wide range of possible realizations of the random variables.
- We explore a multi-policy design based on clustering the random trajectories to reformulate the CC problem.

Multi-policy design (2)

- Subdivide the total probability space into k disjoint clusters (using PCA), and design for each cluster a different nominal forecast trajectory and affine control policy. The problem (P0) is transformed into:

$$(\mathcal{P}3) \quad \min_{x \in \mathcal{X} \subseteq \mathbb{R}^{n_x}} \quad J(x)$$

subject to

$$\Pr\Big(\max_{j=1,\dots,m} g_j^1(x,\delta) \le 0 \mid \delta \in \Delta_1\Big) \ge 1-\epsilon,$$
$$\Pr\Big(\max_{j=1,\dots,m} g_j^2(x,\delta) \le 0 \mid \delta \in \Delta_2\Big) \ge 1-\epsilon,$$
$$\cdots$$
$$\Pr\Big(\max_{j=1,\dots,m} g_j^k(x,\delta) \le 0 \mid \delta \in \Delta_k\Big) \ge 1-\epsilon.$$

Multi-policy design (3)

- We enforce that every specific scenario (realization of the random trajectory) belongs to exactly one cluster:

$$\Pr(\delta \in \Delta_1) + \Pr(\delta \in \Delta_2) + \cdots + \Pr(\delta \in \Delta_k) = 1.$$

- The total probability of constraint satisfaction is given by:

$$\sum_{i=1}^{k} \Pr\left(\max_{j=1,\ldots,m} g_j^i(x,\delta) \leq 0 \mid \delta \in \Delta_i \right) \cdot \Pr(\delta \in \Delta_i) \geq 1 - \epsilon,$$

where the $1-\epsilon$ lower bound is a direct consequence of the structure of the chance-constraints in (P3).

Example

- The process of clustering results in a more flexible control structure, which leads to less conservative capacity design.

Type	4-Cluster	2-Cluster	1-Cluster
$\overline{p}_{\mathrm{w}}$(kW)	393.70	422.99	459.45
$\overline{p}_{\mathrm{pv}}$(kW)	0.73	1.35	0.00
$\overline{p}_{\mathrm{b}}$(kW)	15.04	7.13	8.32
$\overline{e}_{\mathrm{b}}$ (kWh)	65.27	37.79	41.31
$\overline{p}_{\mathrm{elz}}$(kW)	6.37	9.26	12.07
$\overline{M}_{\mathrm{H_2}}$(kg)	1.4818	2.1695	2.98
$\overline{p}_{\mathrm{fc}}$(kW)	11.51	14.86	12.52
$\overline{p}_{\mathrm{rfm}}$(kW)	15.00	15.00	15.00
$\overline{p}_{\mathrm{chp}}$(kW)	15.00	15.00	15.00
$\overline{p}_{\mathrm{e2h}}$(kW)	8.16	8.57	7.83
$\overline{p}^{\mathrm{e}}_{\mathrm{sh}}$(kW)	35.00	35.00	35.00
Cost (Million)	\$2.08	\$2.18	\$2.26

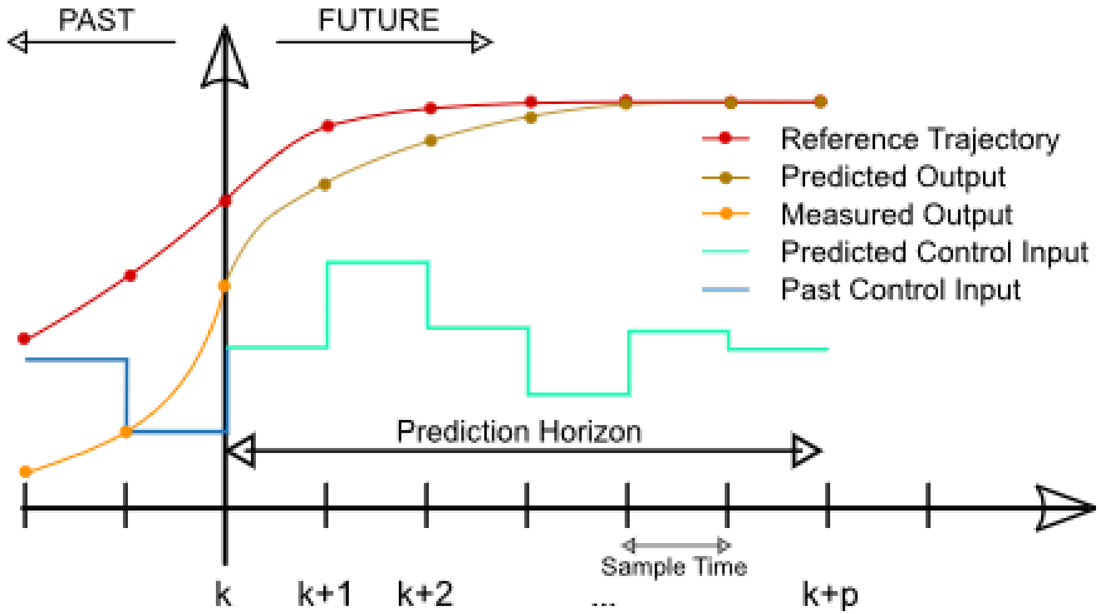
Case studies

- Reliability assessment.
- Baseline estimation.
- Estimating ensemble characteristics.
- Cluster-based chance-constrained optimization.
- **Corrective model predictive control.**

Aim

- Address longer-term instability/viability concerns.
  - Consider processes that evolve over 15-45 minutes.
    - Transmission line and transformer overloading (thermal limits).
    - Voltage collapse.
  - Control updates occur every 2-5 minutes.
- Assume the system is transiently stable (recovers from an initial fault).
- Assume control of generation (conventional and renewable), FACTS (active and reactive power), energy storage, load control, phase-shifting transformers.
- Require computational tractability for large power systems, ~5000 nodes.

Model predictive control



- Measurement of the system state is provided by state estimation.
- Economic dispatch provides the reference trajectory.
- Dynamics are introduced by:
  - Conductor temperature.
  - Generation (the control variable is the change in generation).
  - Energy storage state of charge.
  - Transformer tapping.
  - Long-term load recovery dynamics.

Quadratic program formulation

- Standard optimal control form of objective function:

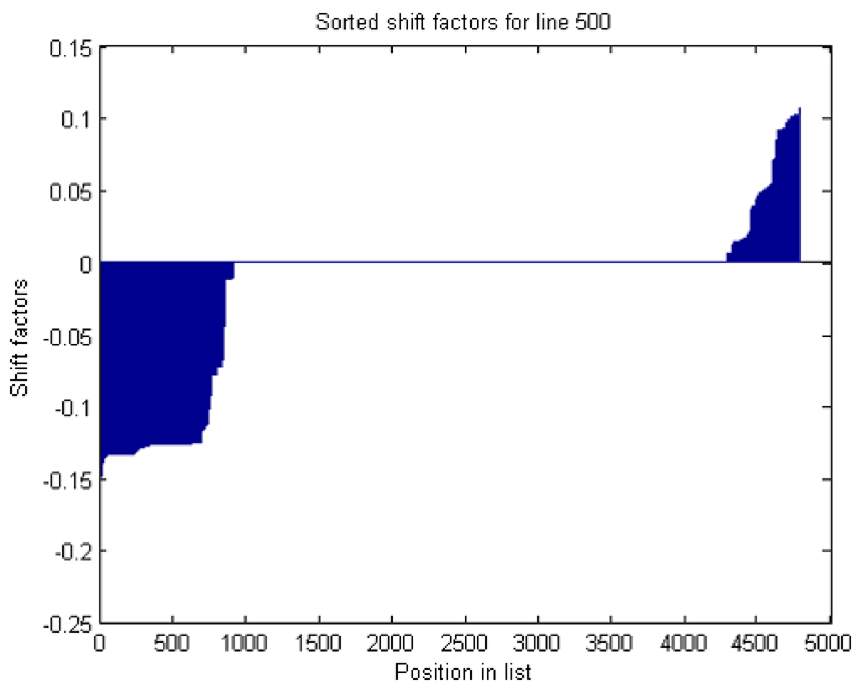
$$\min_{u[l],l=0,\ldots,M-1} \left\|x[M]-x^{sp}_{k+M}\right\|_{S_M} + \sum_{l=0}^{M-1} \left\{\|x[l]-x^{sp}_{k+l}\|_Q + \|u[l]-u^{sp}_{k+l}\|_R\right\}$$

- Penalize:
  - High conductor temperatures.
  - Voltages outside (high and low) limits

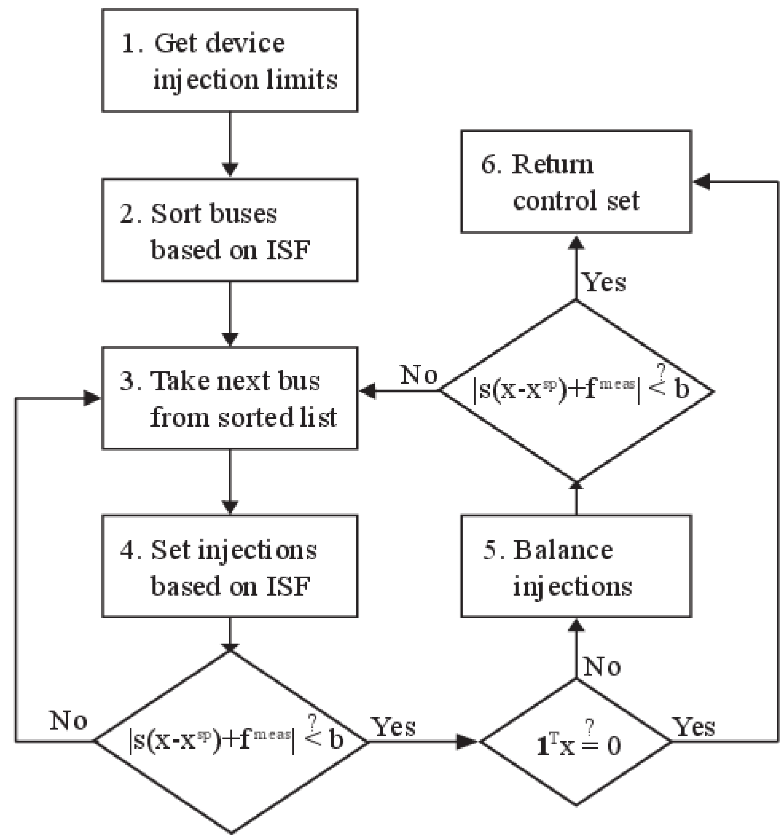
- Constraints:
  - Conventional and renewable generation.
  - Loads.
  - Storage.
  - Power balance (linear active and reactive power balance, variables are $\Delta\theta, \Delta V$).
  - Transformers.
  - Thermal models of line conductor temperature.

Large systems

- Typically only a subset of buses have significant influence over the flow on a specific line.
- Use sensitivity of flow on a line to bus injections to establish relevant controls.



Sensitivities



Selection algorithm

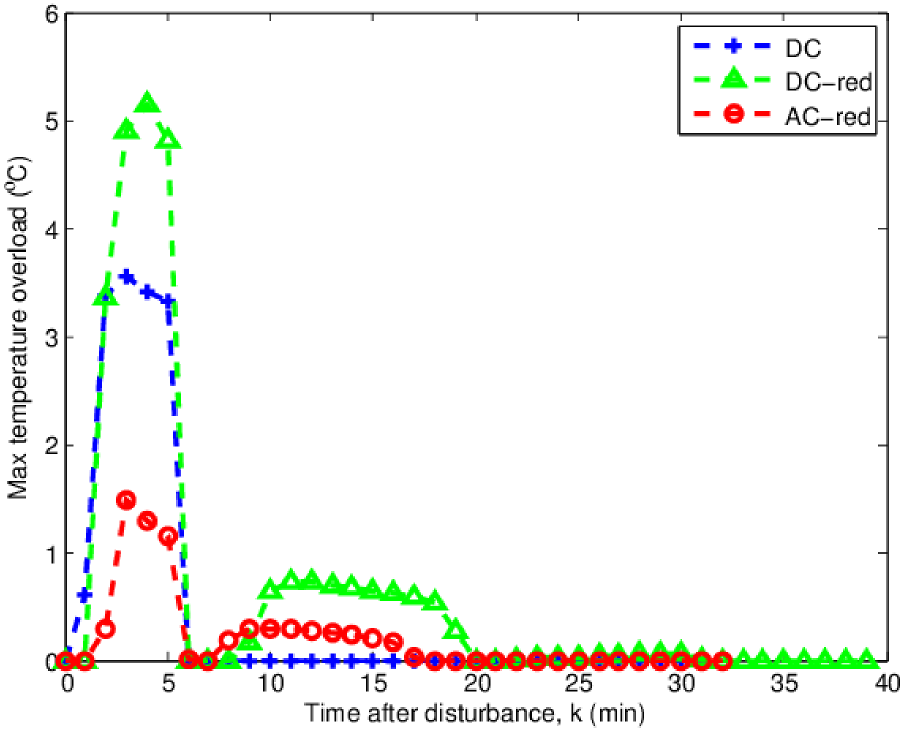
Model reduction

- The only nodes that need to be retained are:
  - Controllable injections.
  - End nodes of overloaded (or potentially overloaded) lines.
  - Locations where voltage magnitudes may deviate outside limits.
- Kron (network) reduction can be used to eliminate unnecessary buses.
- The network size can be changed at each MPC step to capture cascading (or subsiding) effects.

Case study

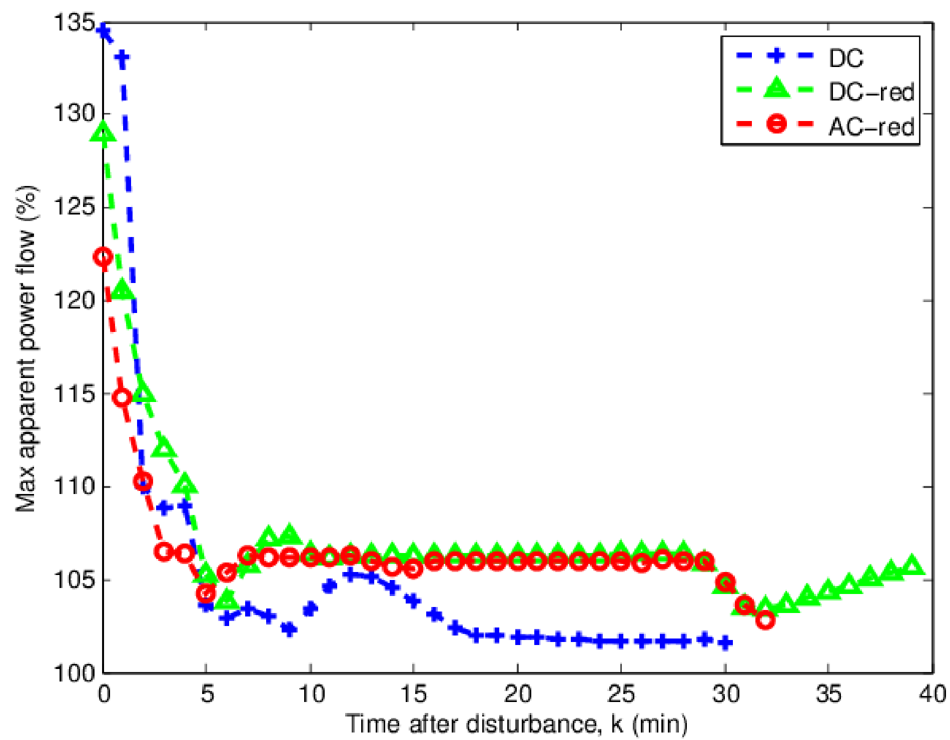
- Californian network: 4259 nodes, 5867 lines/transformers, 2029 conventional and renewable generators, 1443 loads, and 10 grid-scale storage devices.
- Reduced network: 431 buses.
- Controls: 227 conventional generators, 89 renewable generators, 333 loads, and 6 storage devices.
- Thermal models: 6 transmission lines.

Results (1)

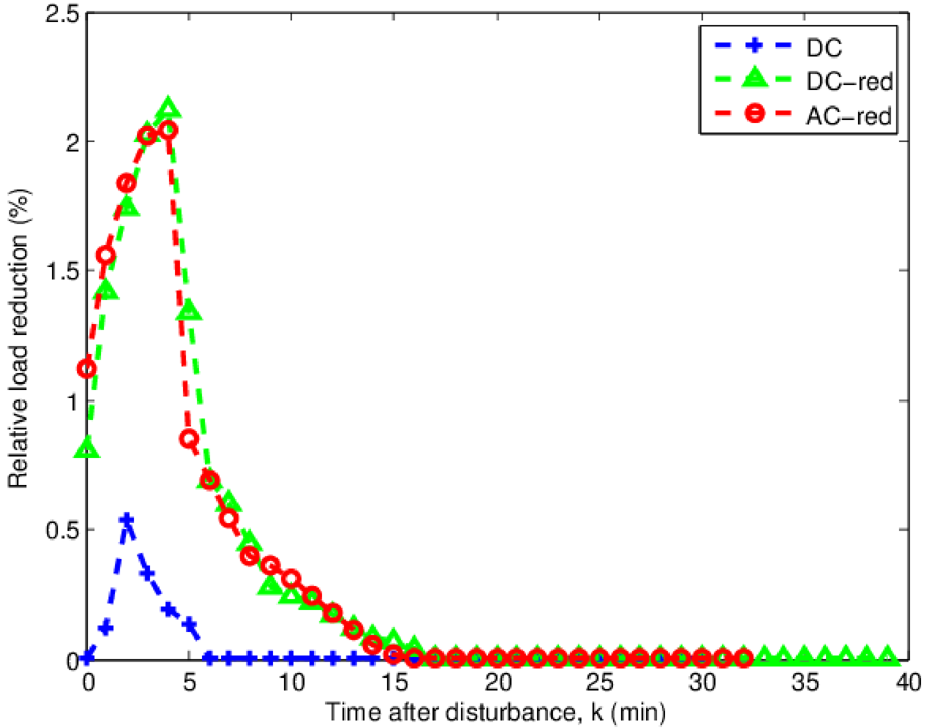


Temperature overload

Maximum apparent power flow



Results (2)



Load reduction

Storage charging