Scalable Nonlinear Model Predictive Control for Energy Applications

Mihai Anitescu*

Mathematics and Computer Science Division

Argonne National Laboratory

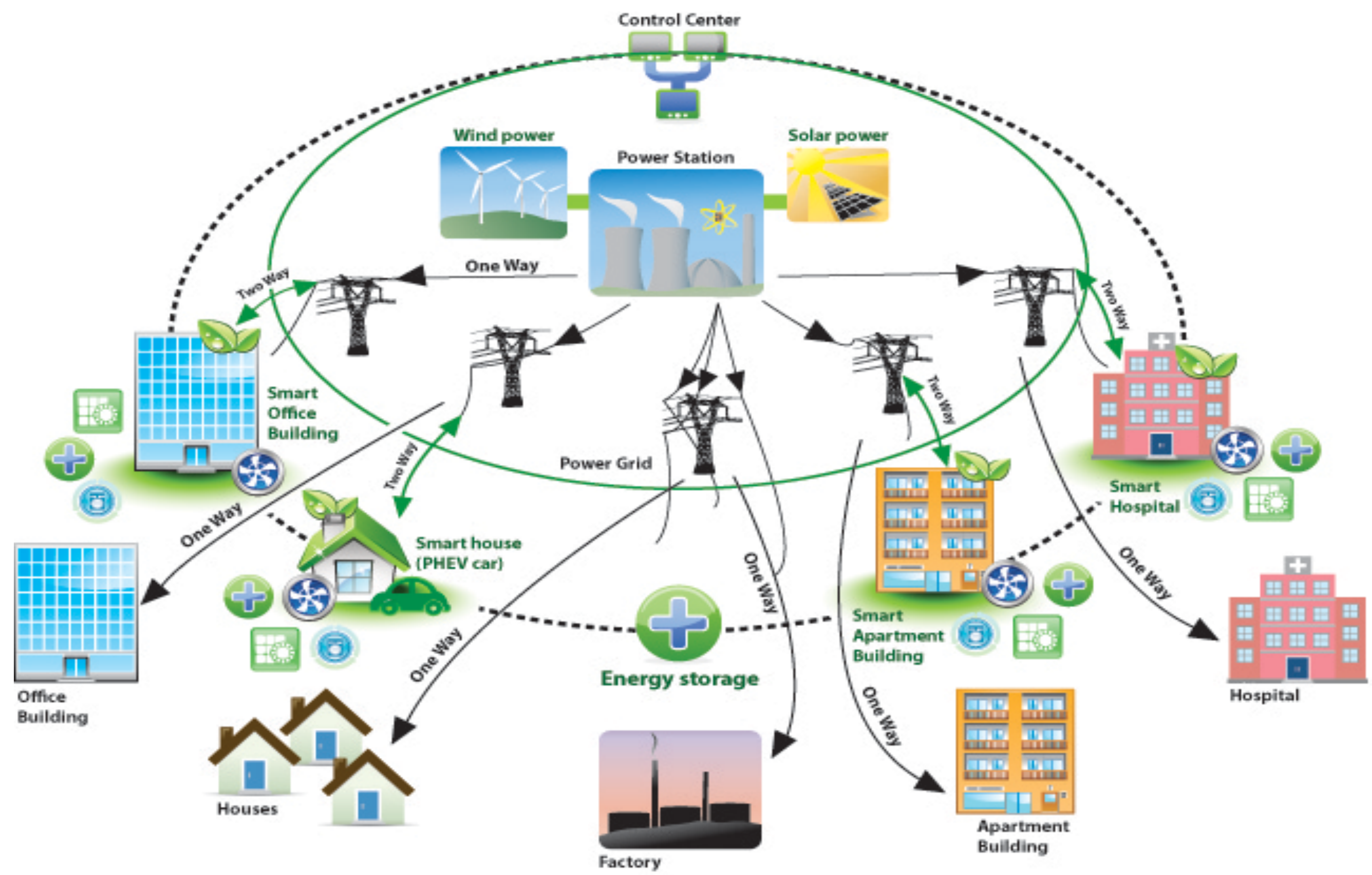
With V. Zavala, E. Constantinescu*, C. Petra, M. Lubin, S. Lee, Matt Rocklin, T. Krause

Bucharest CSCS 2011. – (* PUB Graduates)

Motivation: Management of Energy Systems under Ambient Conditions Uncertainty

Motivation 1. Complexity of energy systems

Next-Generation Power Grid



- Increased Complexity:
  - Financial constraint scales approach physical constraints scales
  - Loads can become "active" with more complicated dynamical behavior.
  - Shorter time responses from smart meters, gas plants (which are faster).

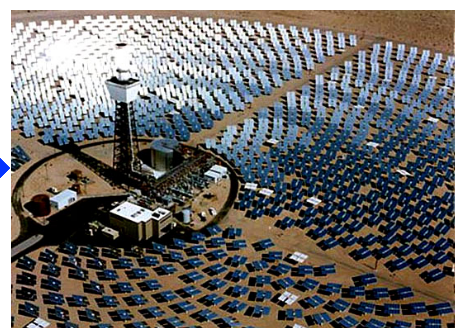
Motivation 2. Uncertainty effects of energy systems

Ambient Condition Effects in Energy Systems

- Operation of 90% of Energy Systems is Affected by Ambient Conditions

- Power Grid Management: **Predict Demands** *(Douglas, et.al. 1999)*
- Power Plants: **Production Levels** *(General Electric)*
- Petrochemical: **Heating and Cooling Utilities** *(ExxonMobil)*
- Buildings: **Heating and Cooling Needs** *(Braun, et.al. 2004)*
- **(Focus) Next Generation Energy Systems assume a major renewable energy penetration: Wind + Solar + Fossil** *(Beyer, et.al. 1999)*

- But increased reliance on renewables must account for their variability …

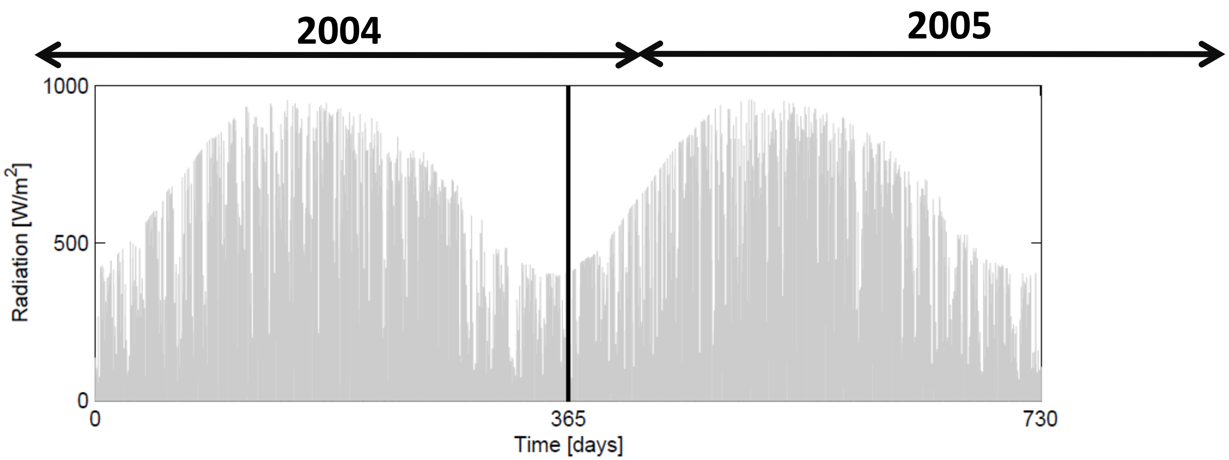


Variability/Uncertainty in Ambient Conditions

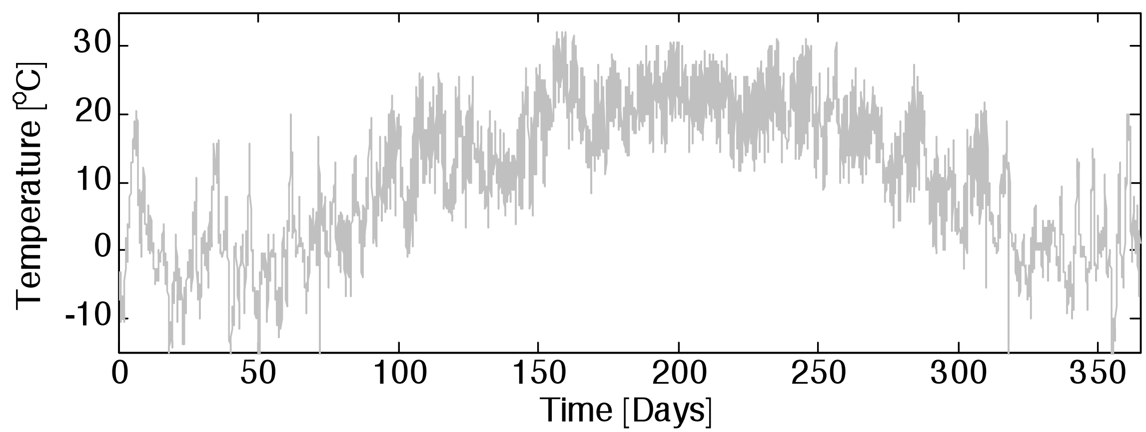
Weather Conditions (Temperature, Wind Speed, Humidity …) have high variability/uncertainty

- Complex Physico-Chemical Phenomena, Spatio-Temporal Interactions
- Inherently Periodic (Day-Night, Seasonal)

Total Ground Solar Radiation
Chicago, IL

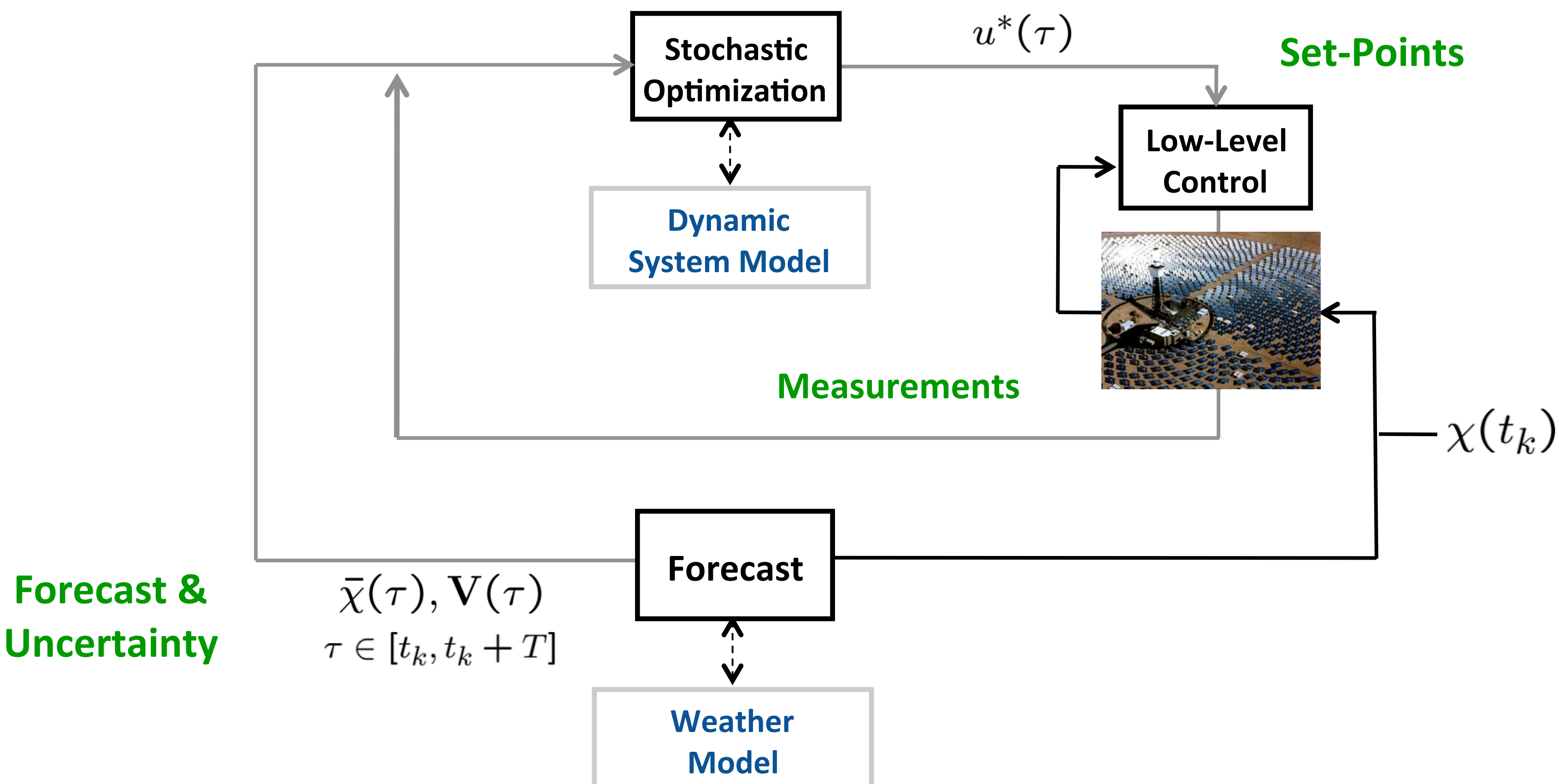


Ambient Dry-Bulb Temperature
Pittsburgh, PA



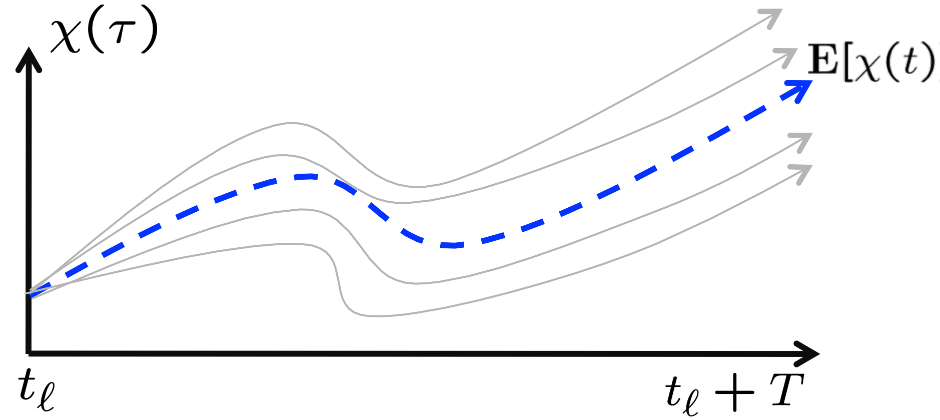
Motivation 3. Nonlinear Model Predictive Control for representing system complexity and uncertainty.

Basic Operational Setting



NLMPC Receding Horizon Optimization

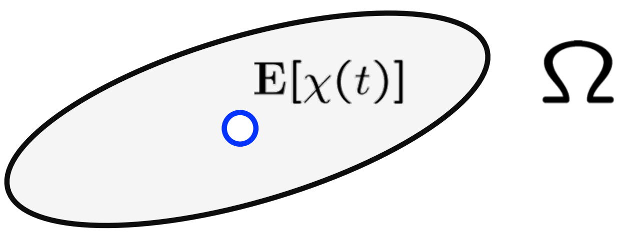
Benefits: Accommodate Forecasts, Constraint Handling, Financial Objectives, Complex Models



Deterministic NLMPC

$$\min_{u(t)} \int_{t_\ell}^{t_{\ell+N}} \varphi(z(t), y(t), u(t), \mathbf{E}[\chi(t)])dt$$

$$\begin{aligned}
\frac{dz}{dt} &= \mathbf{f}(z(t), y(t), u(t), \mathbf{E}[\chi(t)]) \\
0 &= \mathbf{g}(z(t), y(t), u(t), \mathbf{E}[\chi(t)]) \\
0 &\geq \mathbf{h}(z(t), y(t), u(t), \mathbf{E}[\chi(t)]) \\
z(0) &= x_\ell
\end{aligned}$$

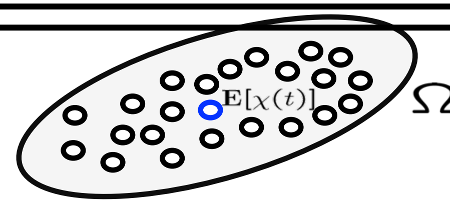


Stochastic NLMPC

$$\min_{u(t)} \mathop{\mathbf{E}}_{\chi(t)\in\Omega} \left[\int_{t_\ell}^{t_{\ell+N}} \varphi(z(t), y(t), u(t), \chi(t))dt \right]$$

$$\left.\begin{aligned}
\frac{dz}{dt} &= \mathbf{f}(z(t), y(t), u(t), \chi(t)) \\
0 &= \mathbf{g}(z(t), y(t), u(t), \chi(t)) \\
0 &\geq \mathbf{h}(z(t), y(t), u(t), \chi(t))
\end{aligned}\right\} \forall \chi(t) \in \Omega$$

$$z(0) = x_\ell$$



Complexity (Solution Time)

1,000 – 10,000 Differential-Algebraic Eqns

100-1000 Scenarios

First entry of control law implemented ➔ recede horizon ➔ restart

CHALLENGES

- Do I have to solve a nonlinear program at each step … that sounds expensive.
- Benefits of considering uncertainty in energy systems
  - Use photovoltaic, building systems, energy dispatch problems as test cases.
  - Does forecast matter?
  - Does stochasticity matter?
- How do we solve the resulting stochastic programming problems?
  - … Since they have a large physical layer.
  - How do I obtain scalable algorithms for solving them
- Sampling from the distribution of the ambient conditions.
  - How do I create and manipulate this large uncertainty space?

OUTLINE

- 1: Sequential Quadratic Programming for NLMPC
- 2: Benefits of Considering Uncertainty Information and Long Forecast Horizons.
- 3: Scalable Stochastic Programming for NLMPC
- 4: Sampling from Spatio-Temporal Uncertainty

1. SEQUENTIAL QP FOR NLMPC

An abstract view of the issues

- **Rolling horizon optimal control:** $F(w,t)=0 \Rightarrow w=w(t)$ **the optimal control/state manifold.**
  - We already wrote the optimality conditions to get it here
  - F can be an operator that includes differential equations for dynamics, which can be discretized somehow.
  - w includes state variables, control variables and Lagrange multipliers
- **The variable** w **cannot be computed instantly, so we must allow it a time** Δt **.**
  - The problem becomes $F(w(t^k),t^k)=0; \quad F(w(t^{k+1}),t^{k+1})=0; \quad t^{k+1}=t^k+\Delta t$
- **Better, but we cannot guarantee that we find a solution in** Δt **even now. What if we solve the subproblem inexactly, e.g only its linearization or an inexact linearization?**

$$F(w^k,t^k)+\nabla_w F(w^k,t^k)(w^{k+1}-w^k)+\nabla_t F(w^k,t^k)\Delta t+r^k=0;$$

- **Could it work? Yes, if we can track the manifold (stability):**

$$\boxed{\left\|w^k-w(t^k)\right\| \le O\left(\left(\Delta t\right)^p\right)}$$

Nonlinear Programming

- How do we prove this? We have to deal with inequality constraints, so not so easy.
- So, write the optimality conditions at time t for NLMPC

$$\begin{aligned}\min_x \;& f(x,t)\\ \text{s.t. } & c(x,t)=0\\ & x\geq 0\end{aligned}$$

Active-Set SQP

$$\begin{aligned}\nabla_x f(x,t) + \nabla_x c(x,t)\,\lambda - \nu &= 0\\ c(x,t) &= 0\\ x^{(i)} &= 0, \forall i \in \mathcal{A}\\ \nu^{(i)} &= 0, \forall i \notin \mathcal{A}\end{aligned}$$

Interior Point

$$\begin{aligned}\nabla_x f(x,t) + \nabla_x c(x,t)\,\lambda - \nu &= 0\\ c(x,t) &= 0\\ X \cdot V &= \mu e\end{aligned}$$

- But still very difficult to work with with inexact linearizations …
- An alternative, convenient framework for the case of inequality constraints: generalized equations (Robinson).

$$F(x,t) = \begin{pmatrix} \nabla_x f(x,t) + \nabla_x c(x,t)^T \lambda \\ c(x,t) \end{pmatrix}; \quad -F(x,t) \in \mathcal{N}_{\mathbb{R}^{n+}\times\mathbb{R}^m}(x)$$

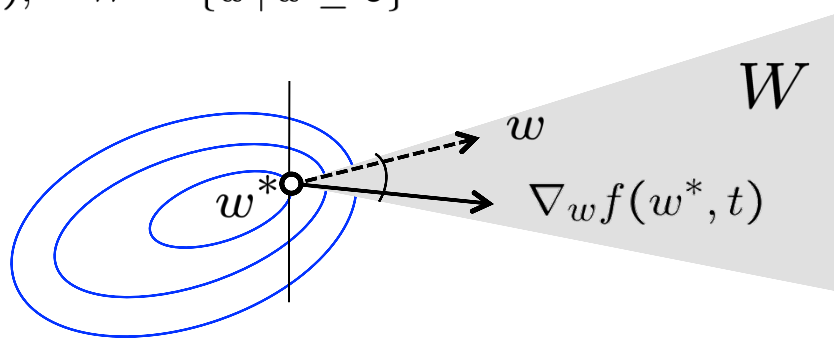
Generalized Equations

Generalized Equations (GE) *Robinson, 1977, 1980*

$-F(w,t) \in \mathcal{N}_W(w) \leftarrow$ **Normal Cone Operator (compare with NLE)**

First-Order KKT Conditions of $\quad \min\limits_{w \in W} \; f(w,t), \quad W = \{w \,|\, w \geq 0\}$

$$-\nabla_w f(w^*,t)^T (w^* - w) \geq 0, \; \forall w \in W$$



Canonical Linearized Generalized Equation (LGE)

$$\delta \in F(w^*_{t_0}, t_0) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w) \qquad w(\delta) = \psi^{-1}[\delta] \leftarrow$$ **Solution Operator**

Definition *(Robinson, 1977)*: **LGE is Strongly Regular at** $w^*_{t_0}$ $\quad \exists\, L_\psi \geq 0$ s.t. $\|w(\delta) - w^*_{t_0}\| \leq L_\psi \|\delta\|$

Theorem: ψ^{-1} **is Lipschitzian if:**

$$\mathrm{M} = \nabla_w F(w^*_{t_0}, t_0) = \begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \mathrm{M}_{13} \\ \mathrm{M}_{21} & \mathrm{M}_{22} & \mathrm{M}_{23} \\ \mathrm{M}_{31} & \mathrm{M}_{32} & \mathrm{M}_{33} \end{bmatrix} \qquad \hat{\mathrm{M}} = \begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} \\ \mathrm{M}_{21} & \mathrm{M}_{22} \end{bmatrix}$$

Active $\quad$ Degenerate $\quad$ Inactive

1. M_{11} **Non-Singular**
2. $\mathrm{M}_{22} - \mathrm{M}_{21}\mathrm{M}_{11}^{-1}\mathrm{M}_{12}$**s Positive Definite**

Generalized Equations

Context of NLP $\min_{x \in X} f(x,t)$, s.t. $c(x,t) = 0$

Solution of Perturbed LGE $\bar{w}_t = [\bar{x}_t \; \bar{\lambda}_t]$ **round** $w^*_{t_0}$

$$\boxed{0 \in F(w^*_{t_0}, t) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w)} \Longleftrightarrow$$

KKT Conditions of Perturbed QP

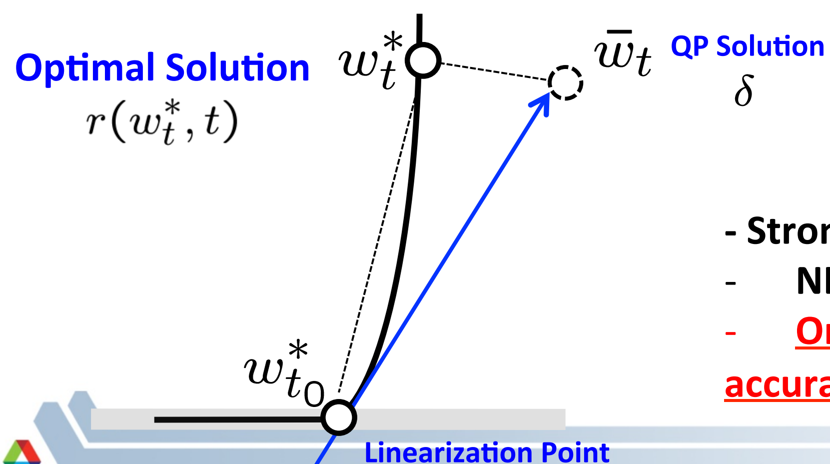
$$\boxed{\begin{array}{ll} \min & \nabla_x f(x^*_{t_0}, t)^T \Delta x + \frac{1}{2} \Delta x^T \nabla_{xx} \mathcal{L}(w^*_{t_0}, t_0) \Delta x \\ \text{s.t.} & c(x^*_{t_0}, t) + \nabla_x c(x^*_{t_0}, t_0)^T \Delta x = 0 \\ & \Delta x \geq -x^*_{t_0} \end{array}}$$

Canonical Form $\Updownarrow$

$$\delta \in F(w^*_{t_0}, t_0) + \nabla_w F(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_W(w)$$ **With** $$\delta = F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t)$$

From Lipschitz Continuity and Mean Value Theorem

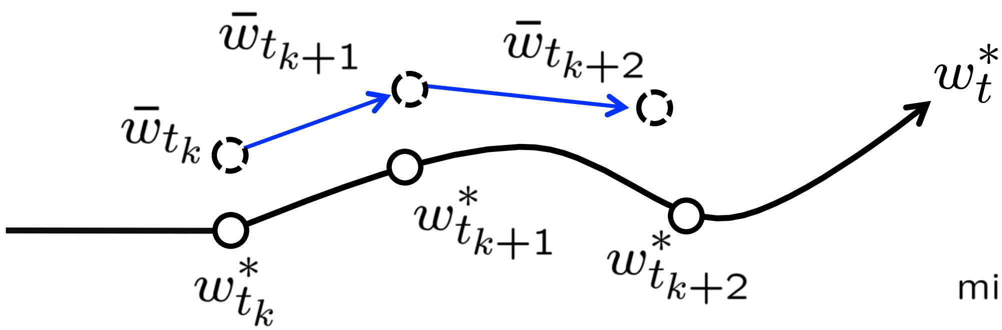
$$\begin{aligned} \|w^*_t - \bar{w}_t\| &\leq L_\psi \|r(w^*_t, t) - \delta\| \\ &\leq L_\psi \| \left(F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t) \right) - \left(F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t) \right) \| \\ &\leq L_\psi \| F_w(w^*_{t_0}, t_0)(w^*_t - w^*_{t_0}) - F(w^*_t, t) + F(w^*_{t_0}, t) \| \\ &\leq L \Delta t^2 \end{aligned}$$



- **Strong Regularity Requires SSOC and LICQ**
- **NLP Error is Bounded by LGE <u>Perturbation</u>**
- **<u>One QP solution from exact manifold is second-order accurat</u>**

Generalized Equations

But I am never EXACTLY on the manifold: Stability of uncentered NLP Error



Time-Dependent QP

$$\begin{aligned} \text{min} \quad & \nabla_x f(\bar{x}_{t_k}, t_{k+1})^T \Delta x + \tfrac{1}{2} \Delta x^T \nabla_{xx} \mathcal{L}(\bar{w}_{t_k}, t_k) \Delta x \\ \text{s.t.} \quad & c(\bar{x}_{t_k}, t_{k+1}) + \nabla_x c(\bar{x}_{t_k}, t_k)^T \Delta x = 0 \\ & \Delta x \geq -\bar{x}_{t_k} \end{aligned}$$

Theorem

- **A1: LGE is Strongly Regular at** $w_{t_k}^*$
- **A2:** $\bar{w}_{t_k}$ **Exists in Neighborhood and** $\exists\, \delta_r \geq 0$ s.t. $\|\bar{w}_{t_k} - w_{t_k}^*\| \leq L_\psi \|r(\bar{w}_{t_k}, t_k)\| \leq L_\psi \delta_r$

For sufficiently small Δt, I can track the manifold stably, solving 1 QP per step

$$\boxed{\|\bar{w}_{t_k} - w_{t_k}^*\| \leq L_\psi \delta_r \Rightarrow \|\bar{w}_{t_{k+1}} - w_{t_{k+1}}^*\| \leq L_\psi \delta_r}$$

Stability Holds Even if QP Solved to $O(\Delta t^2)$ Accuracy .. Can use iterative methods.

Summary NLMPC GE

- We can maintain stability of NLMPC with inequality constraint while solving only one quadratic program (with linear constraints) per step.
- This work extends work by Ohtsuka (for equality constraints) and Diehl et al. (which provided no proof).
- The key analytical development: generalized equations (Robinson).
- We can solve this QP inaccurately (incompletely) as long as we do it within $O(\Delta t^2)$ e.g. by using an augmented Lagrangian. Very suitable for real-time (no matrices are stored, and constant progress is made). Ideal for building systems.
- Reference: Zavala and Anitescu, SIAM Journal of Control 2011.
- However, in this work we will solve it at high accuracy at the moment, we aim for much larger systems.

2. Benefits of considering uncertainty in energy systems

Q: What is the benefit of long forecast horizons?

Q: What is benefit of uncertainty

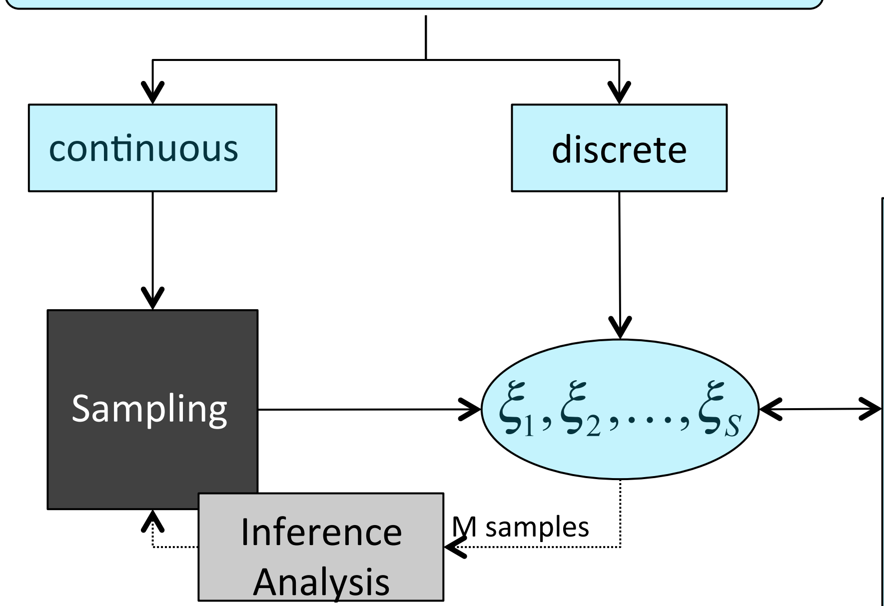
2.1. Benefits/Formulation

Managament under uncertainty paradigm: stochastic programming.

- Two-stage stochastic programming with recourse ("here-and-now")

$$\begin{aligned} \underset{x_0}{Min} & \left\{ f_0(x_0) + \mathbb{E}\left[\underset{x}{Min}\, f(x,\omega) \right] \right\} \\ \text{subj. to.} \quad & A_0 x_0 \qquad\qquad\qquad = \; b_0 \\ & A(\omega)x_0 + \quad B(\omega)x \;\; = \; b(\omega) \\ & x_0 \geq 0, \qquad x(\omega) \geq 0 \end{aligned}$$

- $\xi(\omega) := (A(\omega), B(\omega), b(\omega), Q(\omega), c(\omega))$

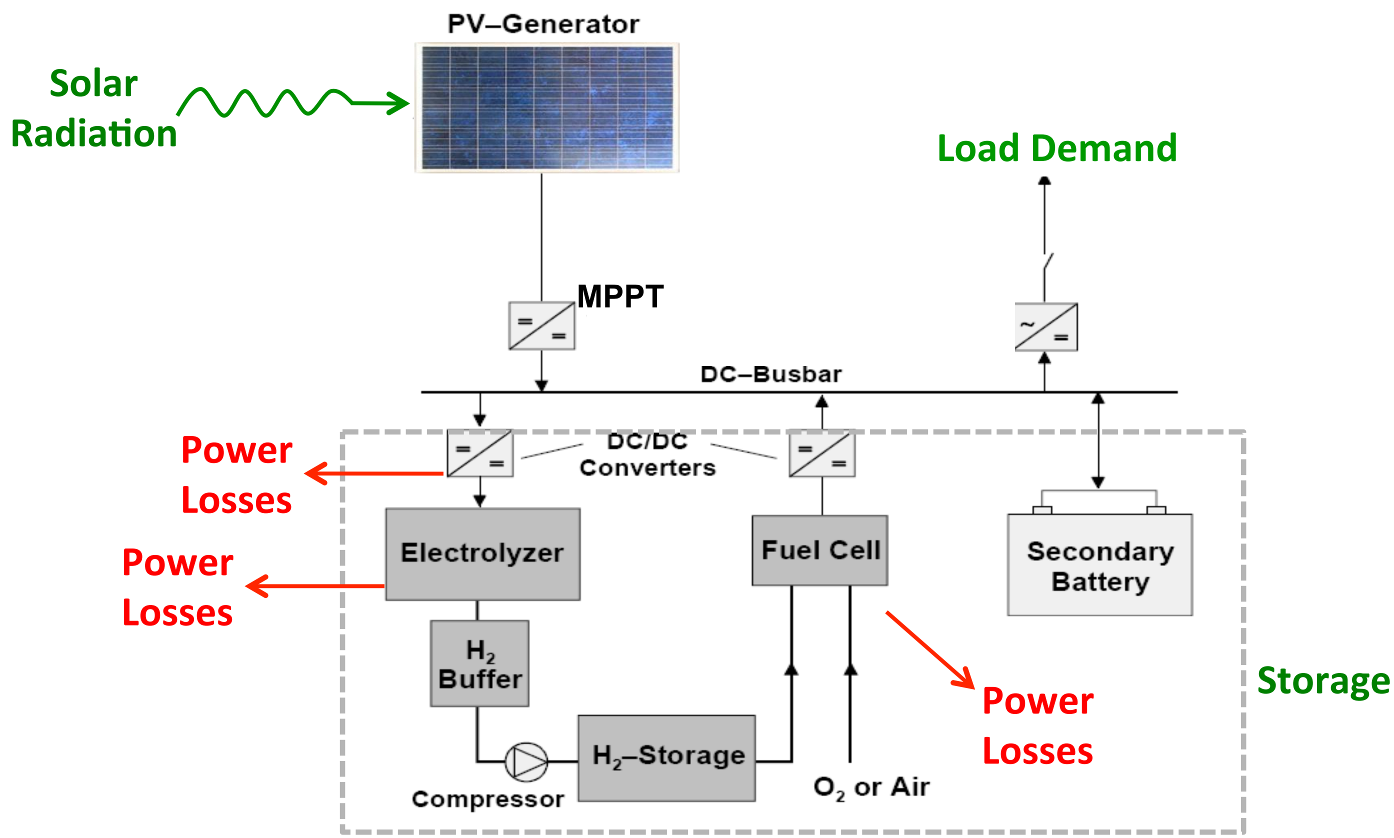


Sample average approximation (SAA)

$$\begin{aligned} \underset{x_0, x_1, x_2, \ldots, x_S}{Min} \quad & f_0(x) + \frac{1}{S}\sum_{i=1}^{S} f_i(x_i) \\ \text{subj. to.} \quad & A_0 x_0 \qquad\qquad\quad = \quad b_0 \\ & A_k x_0 + \quad B_k x_k \;\; = \quad b_k, \\ & x_0 \geq 0, \quad x_k \geq 0, \qquad k = 1, \ldots, S \end{aligned}$$

2.2. Benefits/Photovoltaics

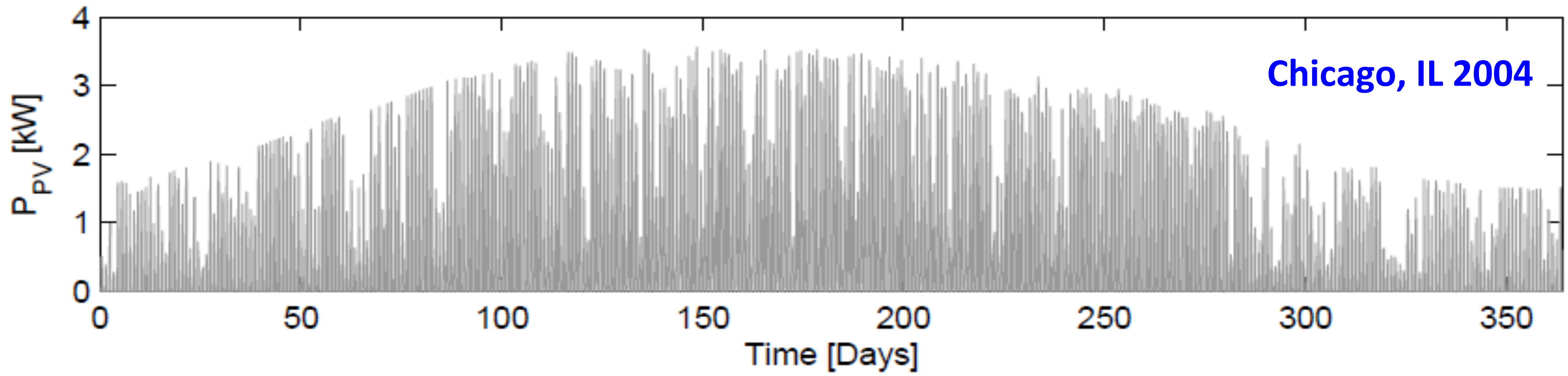
Hybrid Photovoltaic-H_2 System



- **Operating Costs Driven by <u>Uncertain</u> Radiation** *Ulleberg, 2004*
- **Performance Deteriorates by Multiple <u>Power Losses</u>**

Hybrid Photovoltaic-H_2 System

Effect of Forecast on Economics *Z., Anitescu, Krause 2009*



<u>True</u> Future Radiation

$$
\begin{aligned}
\min_{u(t)} \quad & \int_{t_\ell}^{t_{\ell+N}} \varphi(z(t), y(t), u(t), \chi(t))dt \\
\frac{dz}{dt} &= \mathbf{f}(z(t), y(t), u(t), \chi(t)) \\
0 &= \mathbf{g}(z(t), y(t), u(t), \chi(t)) \\
0 &\geq \mathbf{h}(z(t), y(t), u(t), \chi(t)) \\
z(0) &= x_\ell
\end{aligned}
$$

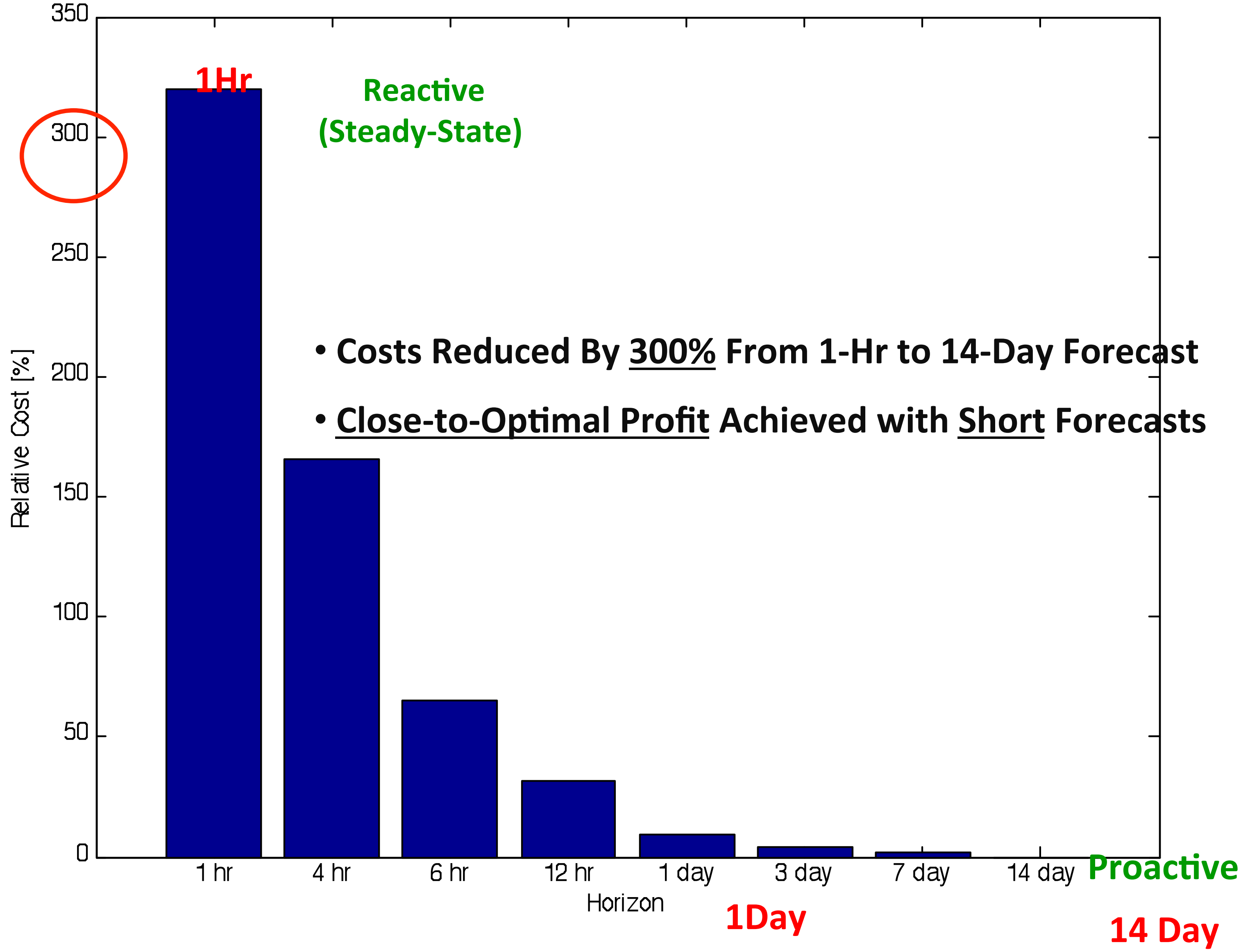
Minimize Operating Costs + Maximize H_2 Production

Energy Balances

State-of-Charge, Fuel Cell and Electrolyzer Limits

- **Forecast Horizon of <u>One Year</u>** – **Highest Achievable Profit**
- **Receding-Horizon with <u>1hr, 1 Day, …,14 Days</u> Forecast** - **8,700 Problems in Each Scenario**

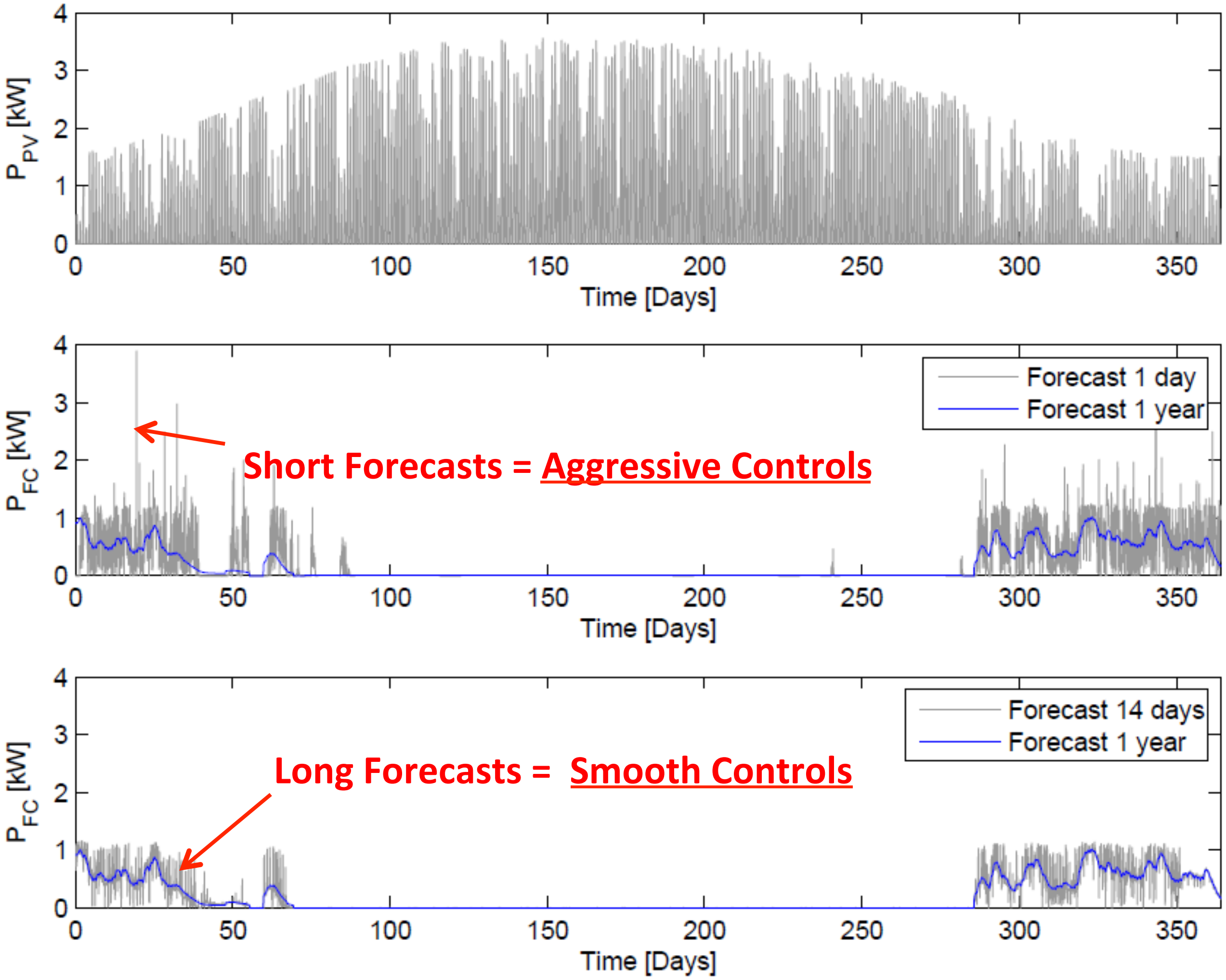
Hybrid Photovoltaic-H_2 System



- Costs Reduced By <u>300%</u> From 1-Hr to 14-Day Forecast
- <u>Close-to-Optimal Profit</u> Achieved with <u>Short</u> Forecasts

Hybrid Photovoltaic-H_2 System

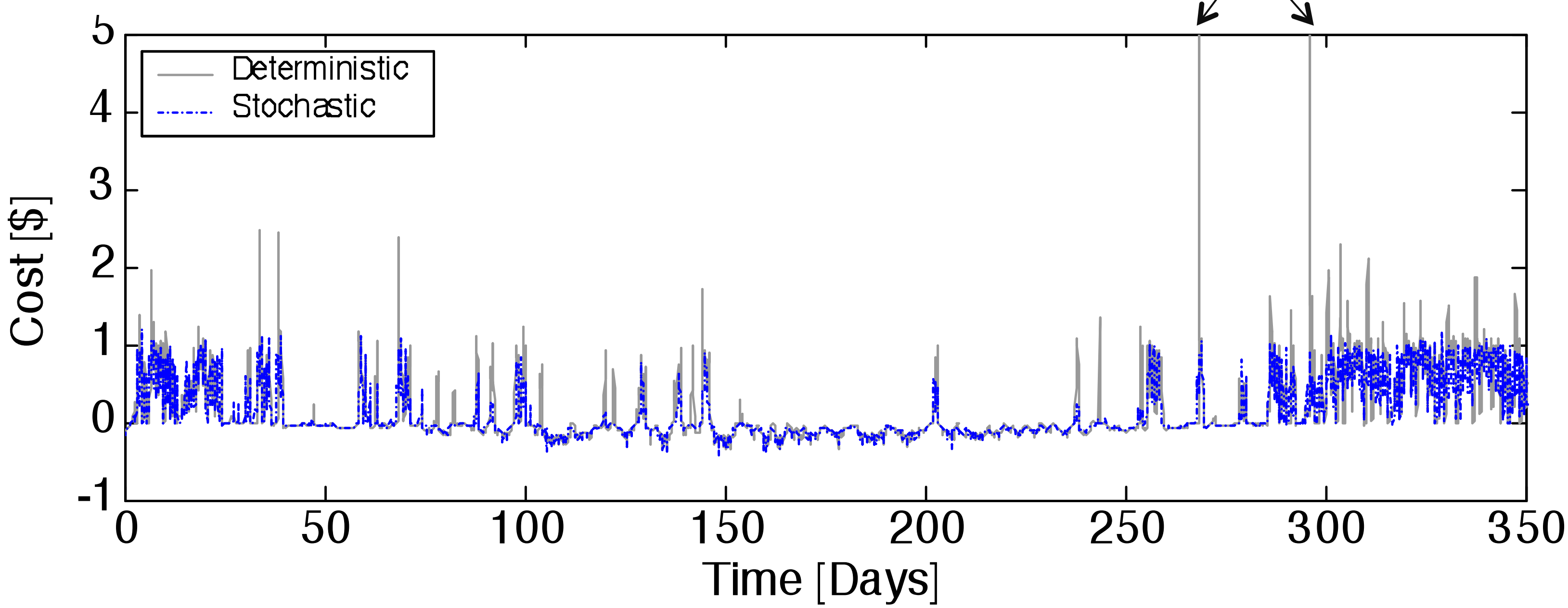
Profiles of Fuel Cell Power



Hybrid Photovoltaic-H_2 System

Load Satisfaction Deterministic ("Optimization on Mean") vs. Stochastic

Deterministic Fails to Satisfy Load

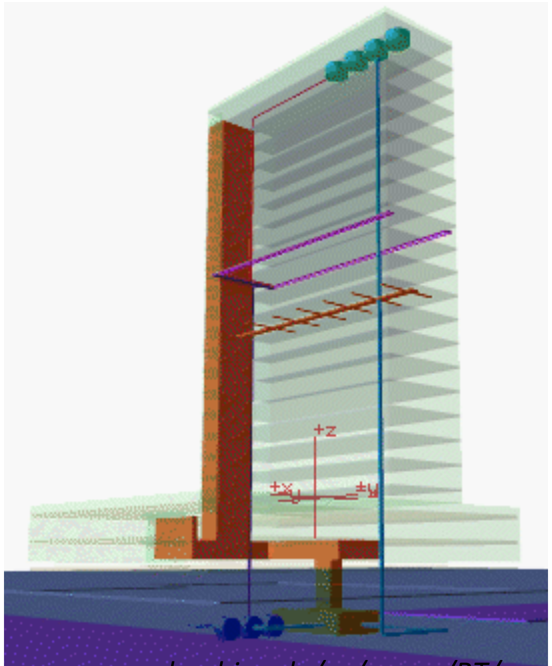


Therefore, the alternative to stochastic programming can turn out infeasible !!

Handling Stochastic Effects Particularly Critical in Grid-Independent Systems

2.3 Benefits/Buildings

Thermal Management of Building Systems



www.columbia.edu/cu/gsapp/BT/LEVER/

Minimize Annual Heating and Cooling Costs

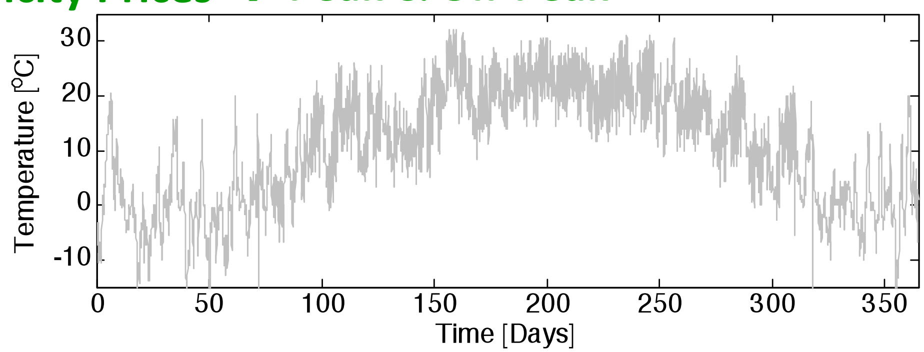
$$\min_{u(t)} \int_{t_\ell}^{t_\ell+N} [C_c(t)\varphi_c(t) + C_h(t)\varphi_h(t)]\, dt$$

$$\begin{aligned}
C_I \cdot \frac{\partial T_I}{\partial \tau} &= \varphi_h(\tau) - \varphi_c(\tau) - S \cdot \alpha' \cdot (T_I(\tau) - T_W(\tau, 0)) \\
\frac{\partial T_W}{\partial \tau} &= \beta \cdot \frac{\partial^2 T_W}{\partial x^2} \\
\alpha' (T_I(\tau) - T_W(\tau, 0)) &= -\mathbf{k} \cdot \left.\frac{\partial T_W}{\partial x}\right|_{(\tau,0)} \\
\alpha'' (T_W(\tau, L) - T_A(\tau)) &= -\mathbf{k} \cdot \left.\frac{\partial T_W}{\partial x}\right|_{(\tau,L)} \\
T_I(0) &= T_I^\ell \\
T_W(0, x) &= T_W^\ell(x)
\end{aligned}$$

Energy Balances

NLP with 100,000 Constraints & 20,000 Degrees of Freedom

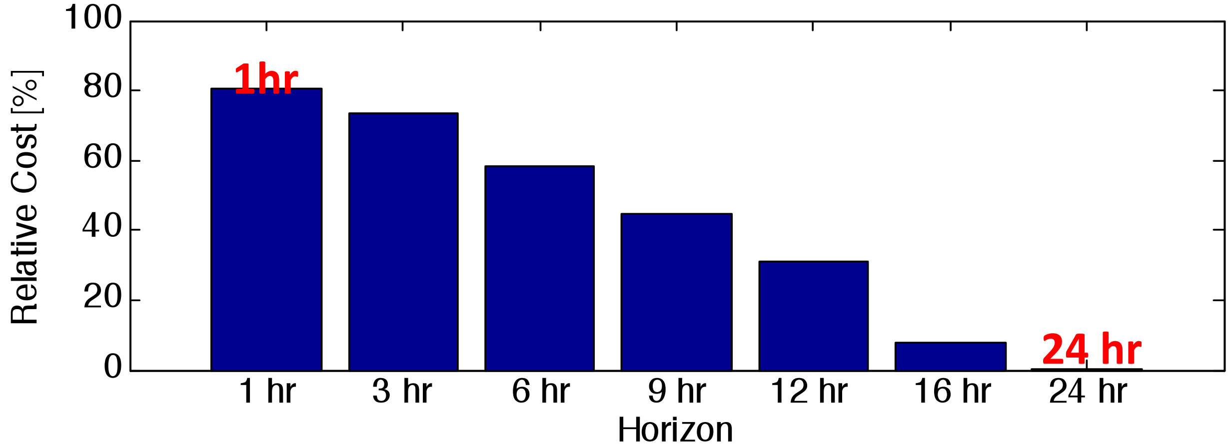
Time-Varying Electricity Prices → Peak & Off-Peak



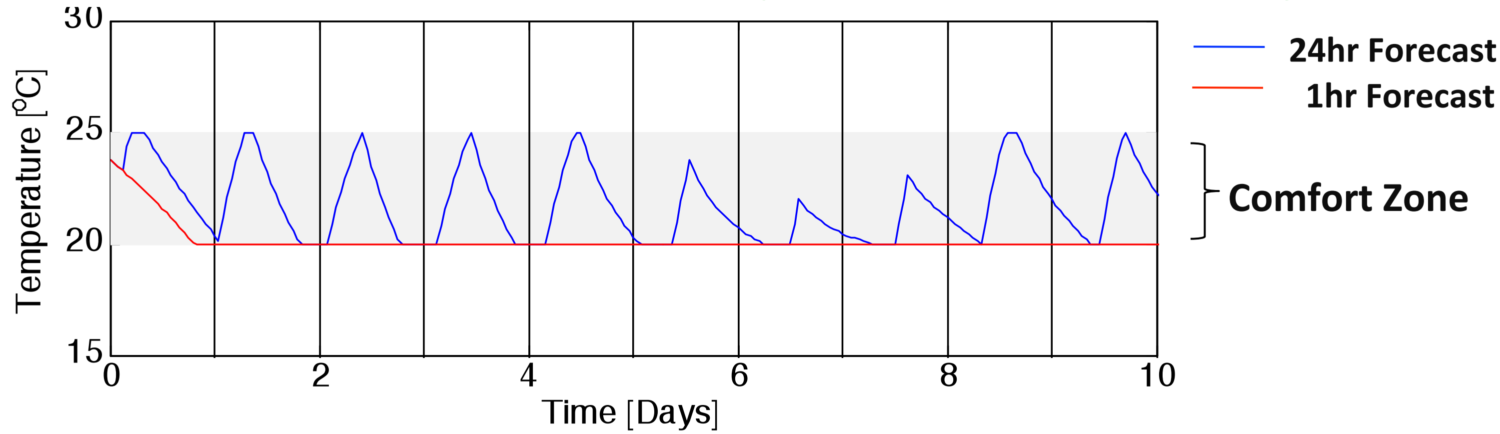
Pittsburgh, PA 2006

Thermal Management of Building Systems

Effect of Forecast on Energy Costs



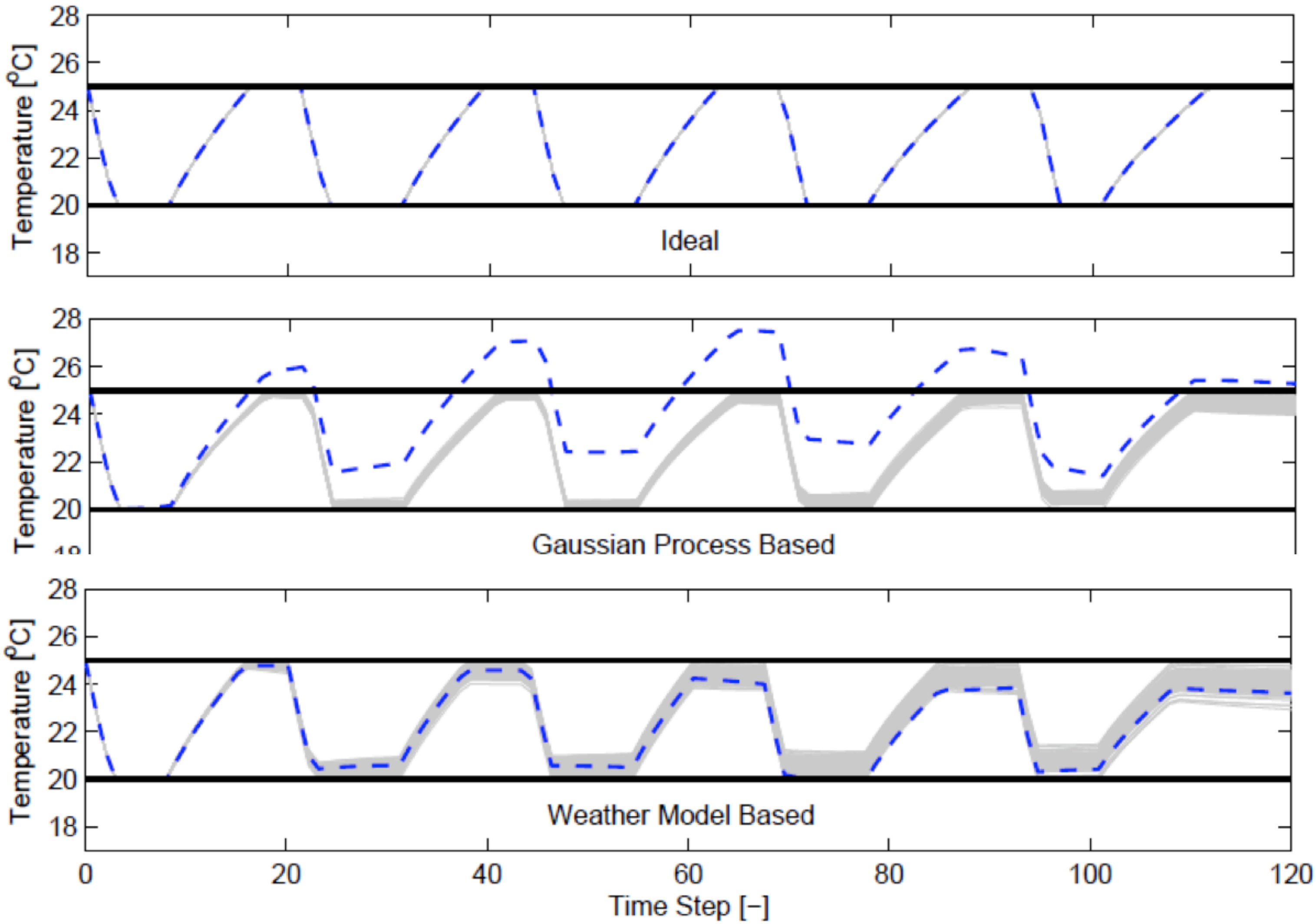
Forecast Leads to 20-80% Cost Reduction (Depends on Insulation Quality)



Exploit Comfort Zone and Weather Info to Heat/Cool when Cheaper *Braun, 1990*

Thermal Management of Building Systems

Performance Optimizer using WRF and GP Model Forecasts



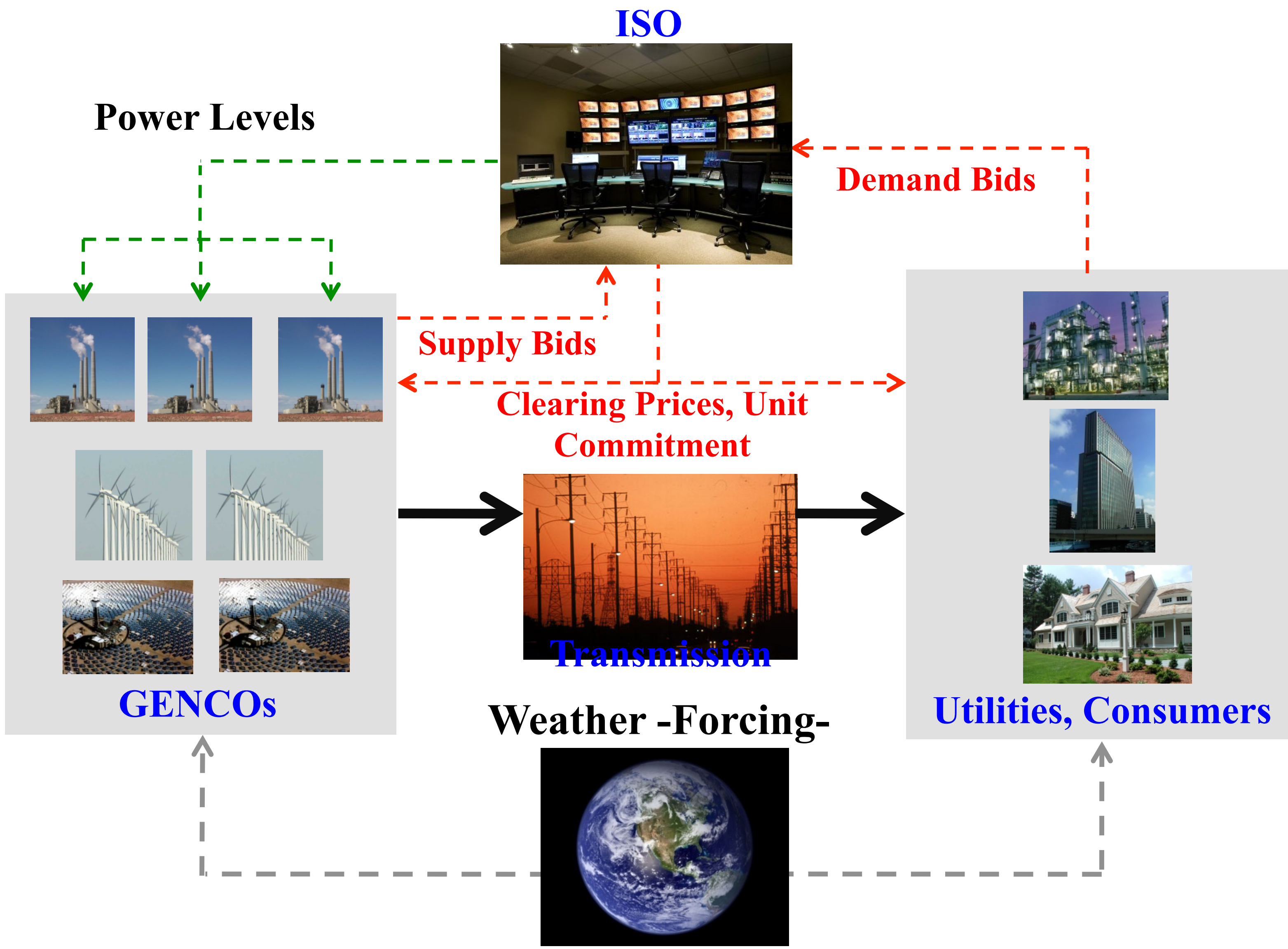
Perfect Forecast

Gaussian Process Model

WRF Model

2.4. Benefits/Regional Energy Systems

Electricity SYSTEM in the US



Dynamic & Uncertain <u>Forcing Factors</u> -Weather- Drive Markets

<u>Volatility</u> Due to Market Friction: (Generation Ramping, Congestion)

Stochastic Unit Commitment with Wind Power (SAA)

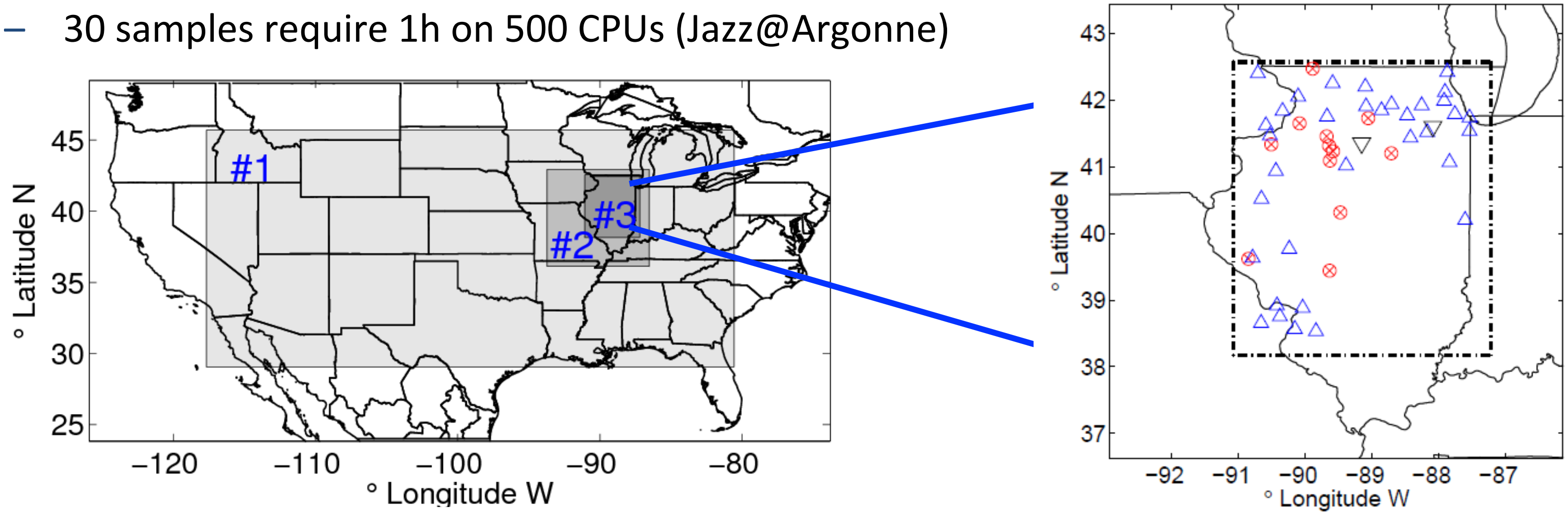
$$\min \quad \text{COST} = \frac{1}{N_s}\sum_{s\in\mathcal{S}}\left(\sum_{j\in\mathcal{N}}\sum_{k\in\mathcal{T}} c^p_{sjk} + c^u_{jk} + c^d_{jk}\right)$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{N}} p_{sjk} + \sum_{j\in\mathcal{N}_{wind}} p^{wind}_{sjk} = D_k, s\in\mathcal{S}, k\in\mathcal{T}$$

$$\sum_{j\in\mathcal{N}} \overline{p}_{sjk} + \sum_{j\in\mathcal{N}_{wind}} p^{wind}_{sjk} \geq D_k + R_k, s\in\mathcal{S}, k\in\mathcal{T}$$

ramping constr., min. up/down constr.

- Wind Forecast – WRF(Weather Research and Forecasting) Model
  - Real-time grid-nested 24h simulation
  - 30 samples require 1h on 500 CPUs (Jazz@Argonne)

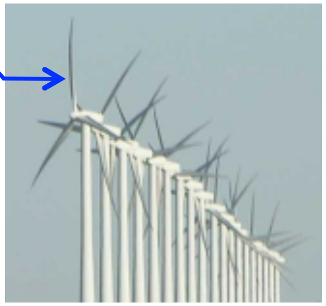


Zavala & al 2010.

Wind power forecast and stochastic programming

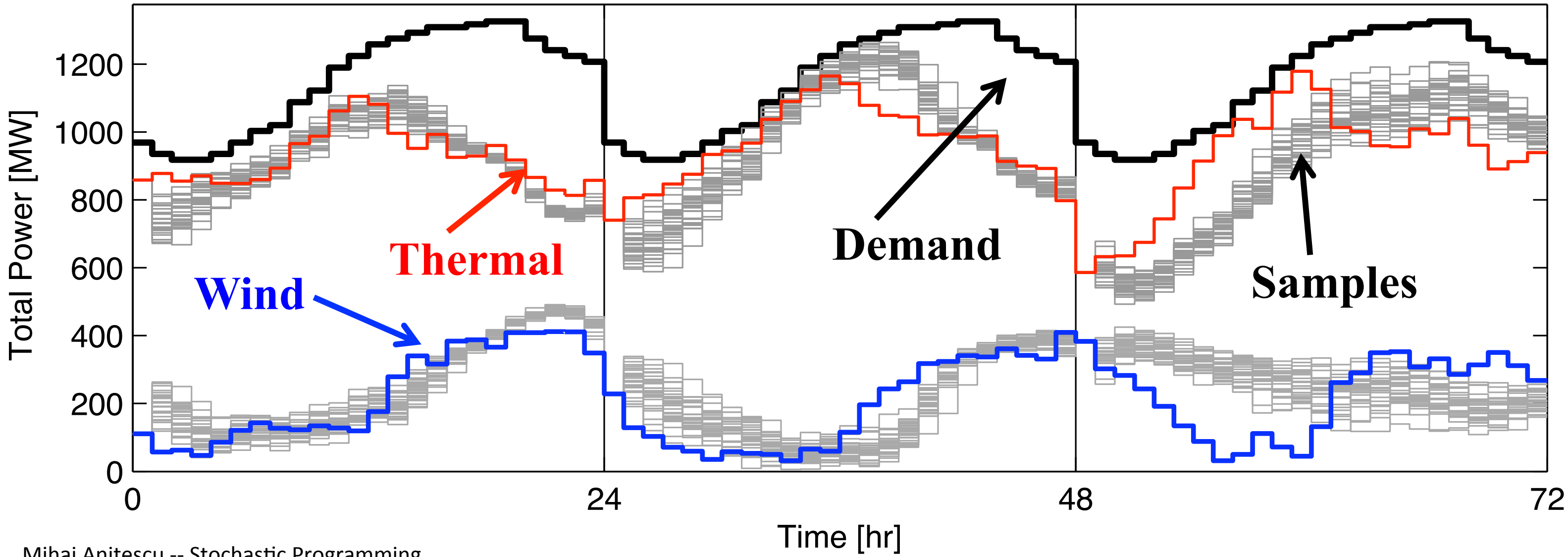
- Unit commitment & energy dispatch with uncertain wind power generation for the State of Illinois, assuming 20% wind power penetration, using the same windfarm sites as the one existing today.

wind power



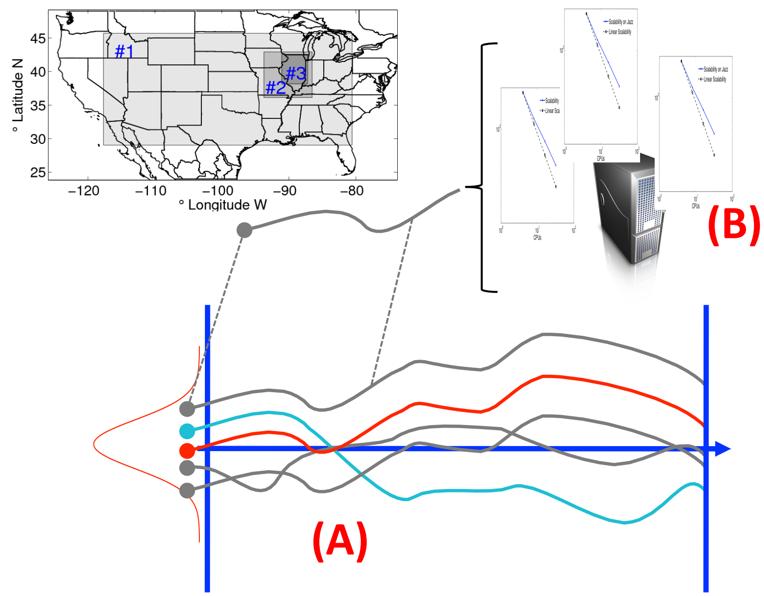
- Full integration with 10 thermal units to meet demands. Consider dynamics of start-up, shutdown, set-point changes

- The solution is only 1% more expensive then the one with exact information. Solution on average infeasible at 10%.

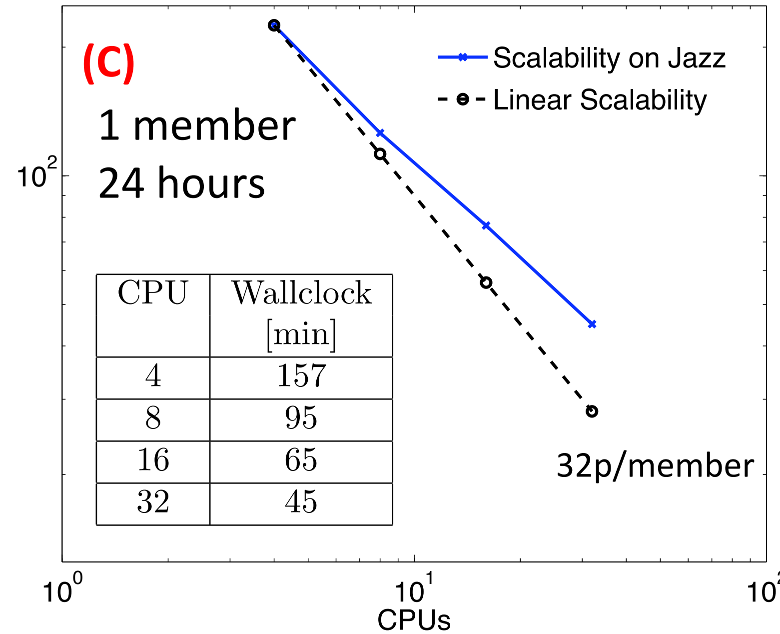


WRF scalability on Jazz

- Two-level parallelization scheme – very scalable: **(A)** realizations are independent, **(B)** each is parallelized, and **(C)** explicit



	Grid	Size
US:	#1 - 32 km^2	130×60
	#2 - 6 km^2	126×121
Illinois:	#3 - 2 km^2	202×232



CPU	Wallclock [min]
4	157
8	95
16	65
32	45

- 24 hours [simulation time] -> one hour [real time] on Jazz with 30 members; [2 km]; (almost) linear scalability with area **(C)**
  - ✔ Illinois [2km]: 500 processors
  - ☐ US [2 km]: ~50,000 processors
  - ☐ US [1 km]: ~400,000 processors

3. How do we solve the resulting stochastic programming problems –(by interior point with special linear algebra)

Linear Algebra of Primal-Dual Interior-Point Methods

Convex quadratic problem

$$\text{Min } \frac{1}{2}x^T Q x + c^T x$$

$$\text{subj. to. } Ax = b$$

$$x \geq 0$$

IPM Linear System

$$\begin{bmatrix} Q+\Lambda & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = rhs$$

Multi-stage SP

Two-stage SP

nested arrow-shaped linear system (via a permutation)

$$\begin{bmatrix}
\boxed{\widetilde{H}_1} & B_1^T & & & & & & & 0 & 0 \\
B_1 & 0 & & & & & & & A_1 & 0 \\
 & & \boxed{\widetilde{H}_2} & B_2^T & & & & & 0 & 0 \\
 & & B_2 & 0 & & & & & A_2 & 0 \\
 & & & & \ddots & & & & \vdots & \vdots \\
 & & & & & \boxed{\widetilde{H}_S} & B_S^T & & 0 & 0 \\
 & & & & & B_S & 0 & & A_S & 0 \\
0 & A_1^T & 0 & A_2^T & \dots & 0 & A_S^T & & \widetilde{H}_0 & A_0^T \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & & A_0 & 0
\end{bmatrix}$$

The Direct Schur Complement Method (DSC)

- Uses the arrow shape of H

$$
\begin{bmatrix} H_1 & & & & G_1^T \\ & H_2 & & & G_2^T \\ & & \ddots & & \vdots \\ & & & H_S & G_S^T \\ G_1 & G_2 & \dots & G_S & H_0 \end{bmatrix} = \begin{bmatrix} L_1 & & & & \\ & L_2 & & & \\ & & \ddots & & \\ & & & L_S & \\ L_{10} & L_{20} & \dots & L_{S0} & L_c \end{bmatrix} \begin{bmatrix} D_1 & & & & \\ & D_2 & & & \\ & & \ddots & & \\ & & & D_N & \\ & & & & D_c \end{bmatrix} \begin{bmatrix} L_1^T & & & & L_{10}^T \\ & L_2^T & & & L_{20}^T \\ & & \ddots & & \vdots \\ & & & L_S^T & L_{S0}^T \\ & & & & L_c^T \end{bmatrix}
$$

- Solving Hz=r

$$L_i D_i L_i^T = H_i, \; L_{i0} = G_i L_i^{-T} D_i^{-1}, \quad i = 1, \dots S,$$

$$C = H_0 - \sum_{i=1}^{S} G_i H_i^{-1} G_i^T, \; \boxed{L_c D_c L_c^T = C.}$$

Implicit factorization

$$w_i = L_i^{-1} r_i, \quad i = 1, \dots S,$$

$$\boxed{w_0 = L_c^{-1}} \left(r_0 - \sum_{i=1}^{S} L_{i0} w_i \right)$$

Back substitution

$$v_i = D_i^{-1} w_i, \; i = 0, \dots, S$$

Diagonal solve

$$\boxed{z_0 = L_c^{-1} v_0}$$

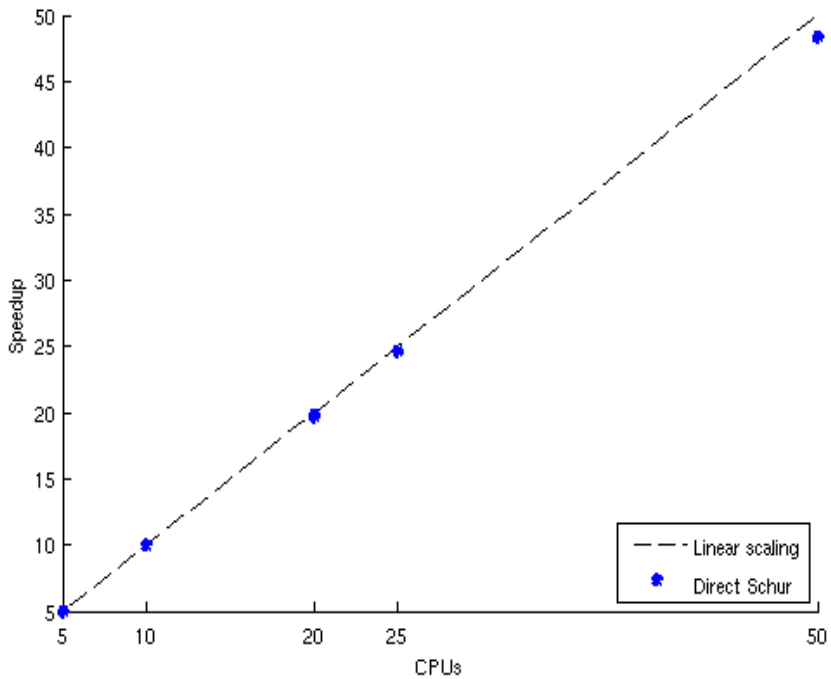
$$z_i = L_i^{-T} \left(v_i - L_{i0}^T z_0 \right),$$

$$i = 1, \dots S.$$

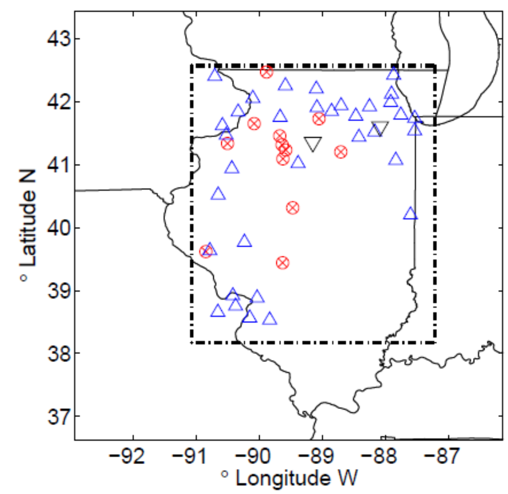
Forward substitution

High performance computing with DSC

- Gondzio (OOPS) 6-stages 1 **billion** variables
- Zavala et.al., 2007 (in IPOPT)
- Our experiments (PIPS) – **strong** scaling is investigated
  - Building energy system
    - Almost linear scaling
  - Unit commitment
    - Relaxation solved
    - Largest instance



12 x

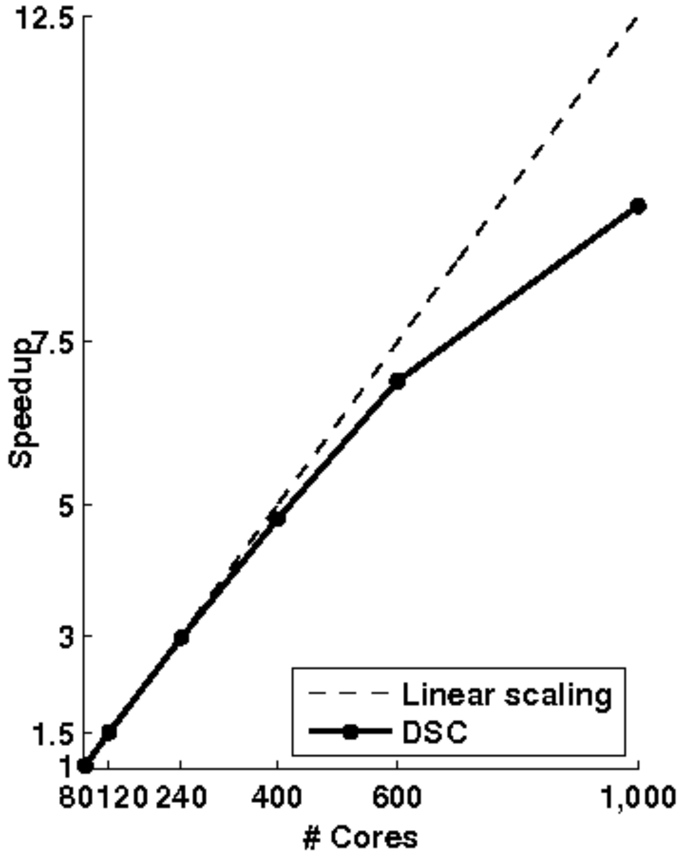


- 28.9 millions variables
- **1000 cores**

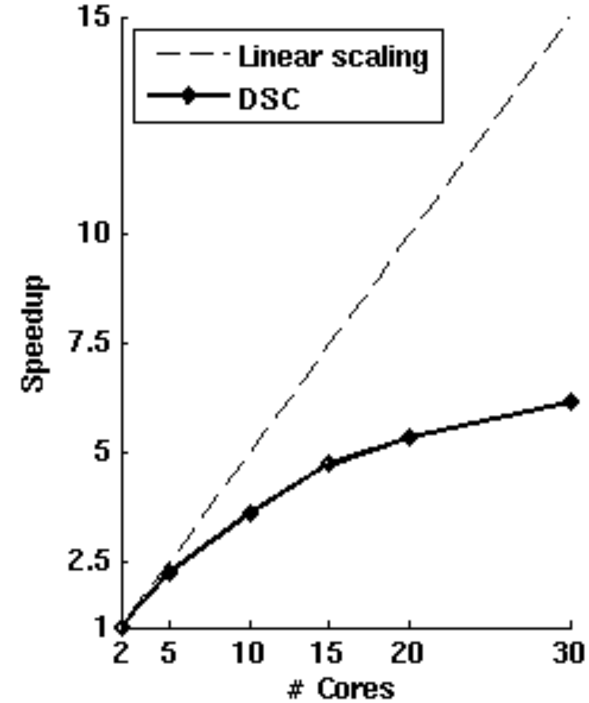
on Fusion @ Argonne

Scalability of DSC

Unit commitment
76.7% efficiency …



…but not always the case, since first stage calculations can keep everyone blocked



Large number of 1st stage variables: **38.6%** efficiency

on Fusion @ Argonne

BOTTLENECK SOLUTION 1: STOCHASTIC PRECONDITIONER

Preconditioned Schur Complement (PSC)

$$L_i D_i L_i^T = H_i,\ L_{i0} = G_i L_i^{-T} D_i^{-1}, \quad i = 1, \dots N,$$
$$C = H_0 - \sum_{i=1}^{N} G_i H_i^{-1} G_i^T$$

$$L_M D_M L_M^T = M$$

(separate process)

$$w_i = L_i^{-1} r_i$$
$$\tilde{r}_0 = \left(r_0 - \sum_{i=1}^{N} L_{Ni} w_i \right)$$

$$v_i = D_i^{-1} w_i,\ i = 0, \dots, N$$
$$z_0 = \mathrm{Krylov}(C, M, \tilde{r}_0)$$

$$z_i = L_i^{-T} \left(v_i - L_{i0}^T z_0 \right)$$

The Stochastic Preconditioner

- The exact structure of C is

$$C = \begin{bmatrix} \underbrace{\widetilde{Q}_0 + \frac{1}{S}\sum_{i=1}^{S}\left[A_i^T \left(B_i \widetilde{Q}_i^{-1} B_i^T\right)^{-1} A_i \right]}_{S_S} & A_0^T \\ A_0 & 0 \end{bmatrix}.$$

- IID subset of n scenarios: $\mathcal{K} = \{k_1, k_2, \ldots, k_n\}$

- The **stochastic preconditioner** (Petra & Anitescu, 2010)

$$S_n = \widetilde{Q}_0 + \frac{1}{n}\sum_{i=1}^{n}\left[A_{k_i}^T \left(B_{k_i} \widetilde{Q}_{k_i}^{-1} B_{k_i}^T\right)^{-1} A_{k_i} \right].$$

- For C use the constraint preconditioner (Keller *et. al.*, 2000)

$$M = \begin{bmatrix} S_n & A_0^T \\ A_0 & 0 \end{bmatrix}.$$

Quality of the Stochastic Preconditioner

$$S_n = \widetilde{Q}_0 + \frac{1}{n}\sum_{i=1}^{n}\left[A_{k_i}^T \left(B_{k_i} \widetilde{Q}_{k_i}^{-1} B_{k_i}^T \right)^{-1} A_{k_i} \right] \qquad S_S = \widetilde{Q}_0 + \frac{1}{S}\sum_{i=1}^{S}\left[A_i^T \left(B_i \widetilde{Q}_i^{-1} B_i^T \right)^{-1} A_i \right]$$

- **"Exponentially" better preconditioning** (Petra & Anitescu 2010)

$$\Pr(|\lambda(S_n^{-1}S_S) - 1| \geq \varepsilon) \leq 2p^4 \exp\left(-\frac{n\varepsilon^2}{2p^4 L^2 \|S_S\|_{max}^2} \right)$$

- **Proof**: Hoeffding inequality (p is dim on S; L is a bound on data)

- Assumptions on the problem's random data
  1. Boundedness
  2. Uniform full rank of $A(\omega)$ and $B(\omega)$

  not restrictive (=> L)

Quality of the Constraint Preconditioner

$$M = \begin{bmatrix} S_n & A_0^T \\ A_0 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} S_S & A_0^T \\ A_0 & 0 \end{bmatrix}$$

$$\begin{array}{cc} \underbrace{}_{p} & \underbrace{}_{r} \end{array}$$

- $M^{-1}C$ has an eigenvalue 1 with order of multiplicity $2r$.
- The rest of the eigenvalues satisfy

$$0 < \lambda_{min}(S_n^{-1}S_S) \leq \lambda(M^{-1}C) \leq \lambda_{\max}(S_n^{-1}S_S).$$

- Proof: based on Bergamaschi *et. al.*, 2004.

The Krylov Methods Used for $Cz_0 = r_0$

$$\begin{bmatrix} S_S & A_0^T \\ A_0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} r_0^1 \\ r_0^2 \end{bmatrix}$$

- **BiCGStab** using constraint preconditioner M
- **Preconditioned Projected CG (PPCG)** (Gould *et. al.,* 2001)
  - Preconditioned projection onto the $KerA_0$.

    $$P = Z_0 \left(Z_0^T S_n Z_0 \right)^{-1} Z_0^T$$

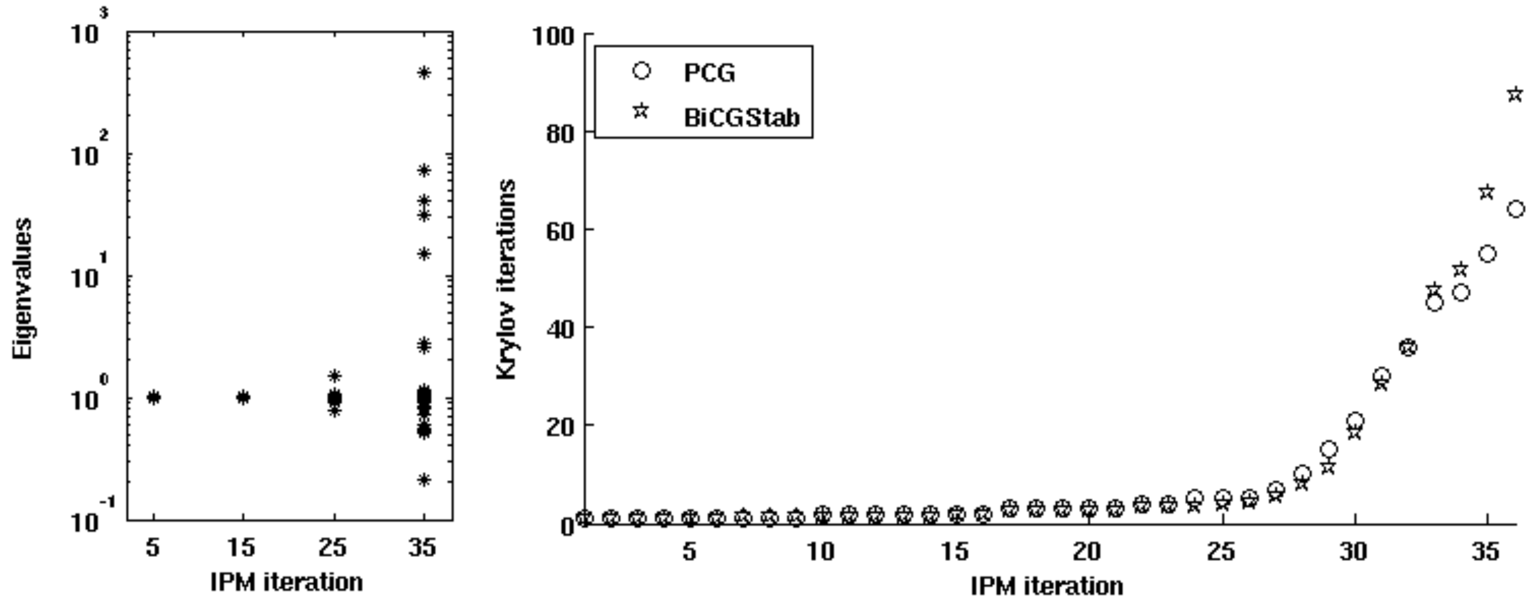
  - Does not compute the basis Z_0 for $KerA_0$. Instead,

    $g = \Pr$ is computed from $\begin{bmatrix} S_n & A_0^T \\ A_0 & 0 \end{bmatrix} \begin{bmatrix} g \\ u \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$.

  - $y_0 = (A_0 A_0^T)^{-1} A_0 \left(r_0^1 - S_N x_0 \right)$

Performance of the preconditioner

- Eigenvalues clustering & Krylov iterations

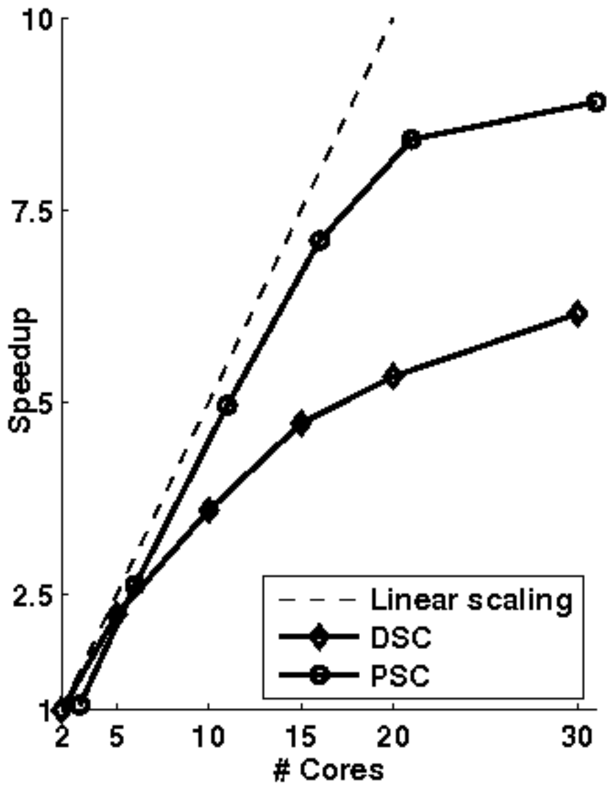


- Affected by the well-known ill-conditioning of IPMs.

The "Ugly" Unit Commitment Problem; PSC gets further

- DSC on P processes vs PSC on P+1 process

Optimal use of PSC – **linear scaling**



- 120 scenarios - # cores used for preconditioner
- Conclusion: PSC hides the latency well, but it eventually hits a memory wall as well.

Factorization of the preconditioner can not be hidden anymore; we need to accelerate it as well; cannot solve larger problems where improvement would likely be larger

SOLUTION 2: PARALELLIZATION OF STAGE 1 LINEAR ALGEBRA

Parallelizing the 1st stage linear algebra

- We distribute the 1st stage Schur complement system.

$$C = \begin{bmatrix} \tilde{Q} & A_0^T \\ A_0 & 0 \end{bmatrix}, \quad \tilde{Q} \text{ dense symm. pos. def.,} \quad A_0 \text{ sparse full rank.}$$

- C is treated as dense.
- Alternative to PSC for problems with large number of 1st stage variables.
- Removes the memory bottleneck of PSC and DSC.
- We investigated ScaLapack, Elemental (successor of PLAPACK)
  - None have a solver for symmetric indefinite matrices (Bunch-Kaufman);
  - LU or Cholesky only.
  - So we had to think of modifying either.

Cholesky-based LDL^T -like factorization

- Can be viewed as an "implicit" normal equations approach.
- In-place implementation inside Elemental: no extra memory needed.
- Idea: modify the Cholesky factorization, by changing the sign after processing p columns.
- It is much easier to do in Elemental, since this distributes elements, not blocks.
- Twice as fast as LU
- Works for more general saddle-point linear systems, *i.e.,* pos. semi-def. (2,2) block.

Distributing the 1st stage Schur complement matrix

- **All** processors contribute to **all** of the elements of the (1,1) dense block

$$\tilde{Q} = \widetilde{Q}_0 + \frac{1}{S}\sum_{i=1}^{S}\left[A_i^T \left(B_i \widetilde{Q}_i^{-1} B_i^T \right)^{-1} A_i \right]$$

- A large amount of inter-process communication occurs.
- Possibly more costly than the factorization itself.
- Solution: use buffer to reduce the number of messages when doing a *Reduce_scatter.*
- LDL^T approach also reduces the communication by half – only need to send lower triangle.

A Parallel Interior-Point Solver for Stochastic Programming (PIPS)

- Convex QP SAA SP problems
- Input: users specify the scenario tree
- Object-oriented design based on OOQP
- Linear algebra: tree vectors, tree matrices, tree linear systems
- Scenario based parallelism
  - tree nodes (scenarios) are distributed across processors
  - inter-process communication based on MPI
  - dynamic load balancing
- Mehrotra predictor-corrector IPM
- We investigated scaling up to 130K processors.

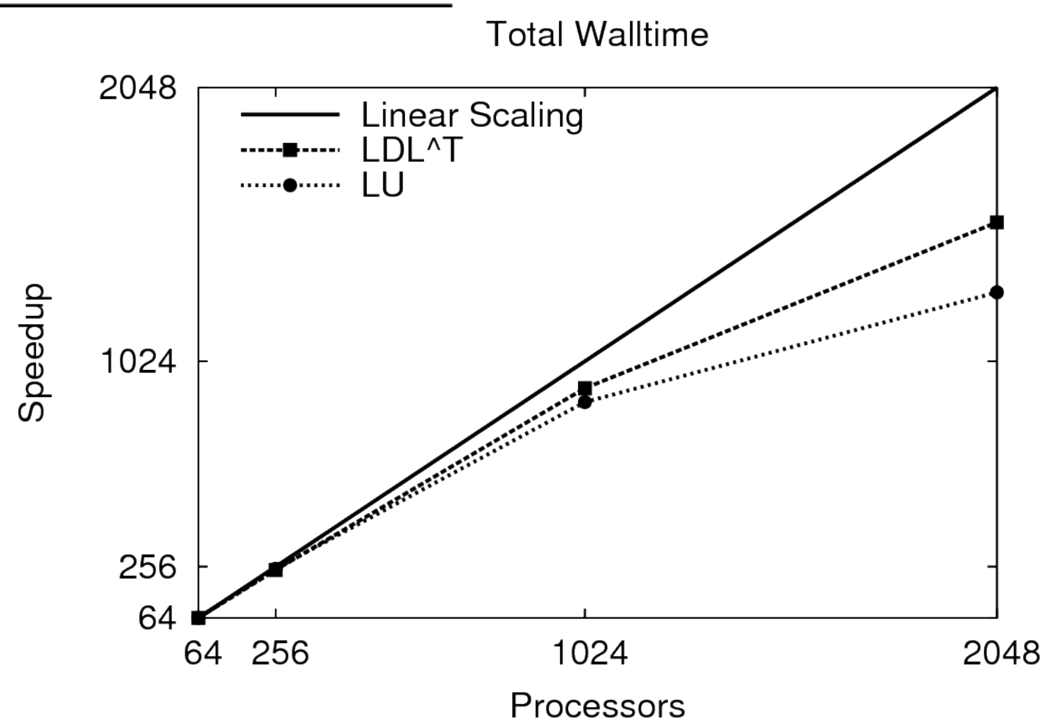
Large-scale performance

- Comparison of ScaLapack (LU), Elemental(LU), and LDL^T (1024 cores)

Units	1st Stage Size $(Q+A)$	Factor (Sec.) LU(S)	LU(E)	LDL^T	Reduce (Sec.) LU	LDL^T
300	23436+1224	16.59	20.04	**6.71**	54.32	**26.35**
640	49956+2584	60.67	83.24	**36.77**	256.95	**128.59**
1000	78030+4024	173.67	263.53	**90.82**	565.36	**248.22**

SAA problem:
189 million variables

- Strong scaling
  - **90.1%** from 64 to 1024 cores;
  - 75.4% from 64 to 2048 cores.
  - > 4,000 scenarios.



Argonne Leadership Computing Facility (ALCF) – BG/P system

BG/P Surveyor System

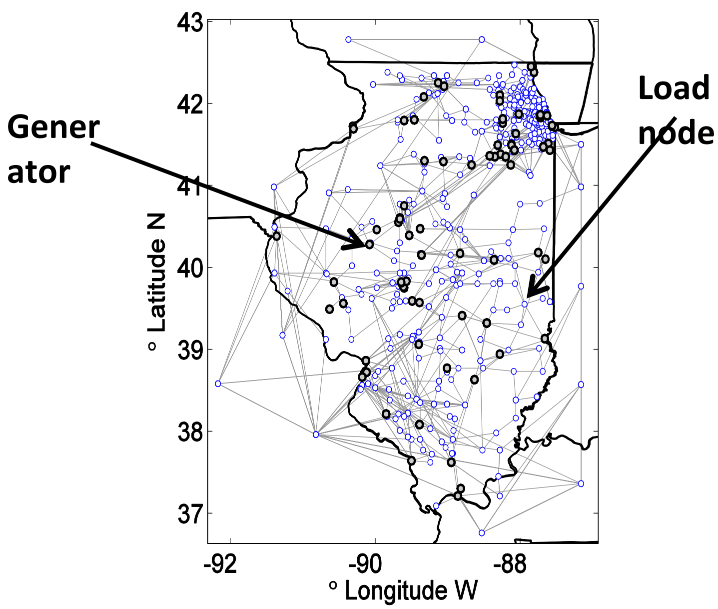
13.6 TF/s 1 rack BG/P
1024 compute nodes
(4096 CPUs)

BG/P Intrepid System

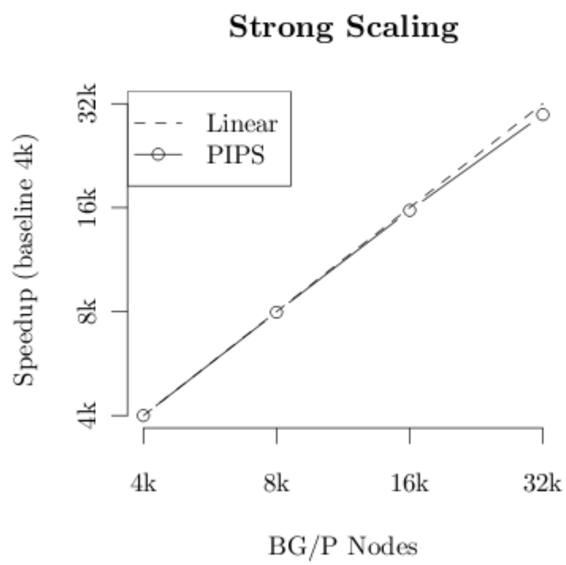
557.1 TF/s 40 rack BG/P
40960 compute nodes
(163840 CPUs)

Results on BG/P

- Now include transmission.



- 3B variables, 100K primal variables, 32K scenarios -- one problem solved in about 10 hours – we are working on “real-time” – 1 hour.



Using BG/L for energy dispatch

- Is it really worth using a supercomputer for this task?
- Let's look at the most pressing item of Supercomputing usage: power.
- BG/P needs about 20MW of power.
- The Midwest US has 140GW of power installed, and the peak demands runs up to 110GW.
- We will never reduce power consumption, but we will make it more reliable, and cheaper. (we estimate 1-5%)
- This is worth on the order of 100-500MW, far above what BG/P costs in power consumption.

4. Sampling from the distribution of ambient conditions (“uncertainty quantification” of ambient conditions).

Uncertainty representation in weather:

- The gold standard in weather forecast: Hidden Markov Model
- Assume a time-discretized process with imperfect initial state and forcing information and noisy measurements.

The dynamic model is depicted as for $k = 0, \cdots, K$

$$
\begin{aligned}
x_k^{in} &= M(x_{k-1}^{in}) + W_k, & (1) \\
z_k^{obs} &= H(x_k^{in}) + V_k, & (2)
\end{aligned}
$$

where

$$W_k \approx N(\bar{x}_k, Q_k^{-1})$$

and

$$V_k \approx N(\mathbf{0}, R_k^{-1}).$$

We want find $D(x_0^{in}, \cdots, x_K^{in})$'s mean and variance.

- M: physical model, a multi-variable, multi-dimensional, time-dependent partial differential equation (state size for 1 time step: 10^4—10^12)

Uncertainty in dynamical systems: 2. the posterior.

- Under the typical 4D Var assumptions (normality of noise and input) we can write down the posterior …

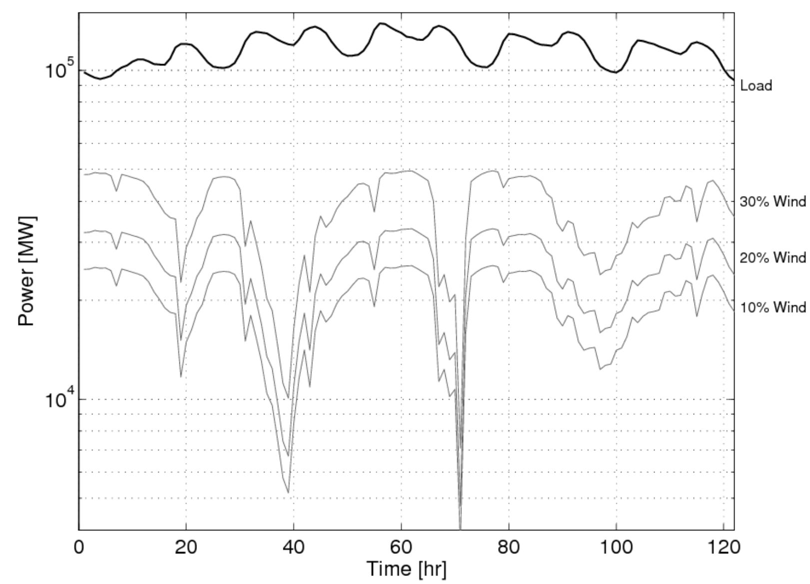
$$P(\boldsymbol{x}_k^{in}, \boldsymbol{x}_{k-1}^{in}, \cdots, \boldsymbol{x}_0^{in} | \boldsymbol{z}_0^{obs}, \boldsymbol{z}_1^{obs}, \boldsymbol{z}_2^{obs}, \cdots, \boldsymbol{z}_k^{obs}) = C_k \tilde{C}_k \frac{\exp\left(-\frac{1}{2} f(\boldsymbol{X}^{in}, \boldsymbol{Z}^{obs})\right)}{P(\boldsymbol{Z}^{obs})}.$$

$$f(\boldsymbol{X}^{in}, \boldsymbol{Z}^{obs}) = \sum_{i=0}^{k} (\boldsymbol{x}_i^{in} - \bar{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}(t_{i-1}, \boldsymbol{x}_{i-1}^{in}))^T Q_i^{-1} (\boldsymbol{x}_i^{in} - \bar{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}(t_{i-1}, \boldsymbol{x}_{i-1}^{in})) + \sum_{i=0}^{k} (\boldsymbol{z}_i^{obs} - \boldsymbol{h}_i(\tilde{\boldsymbol{y}}^{\perp}(t_i, \boldsymbol{x}_i^{in})))^T R_i^{-1} (\boldsymbol{z}_i^{obs} - \boldsymbol{h}_i(\tilde{\boldsymbol{y}}^{\perp}(t_i, \boldsymbol{x}_i^{in})))$$

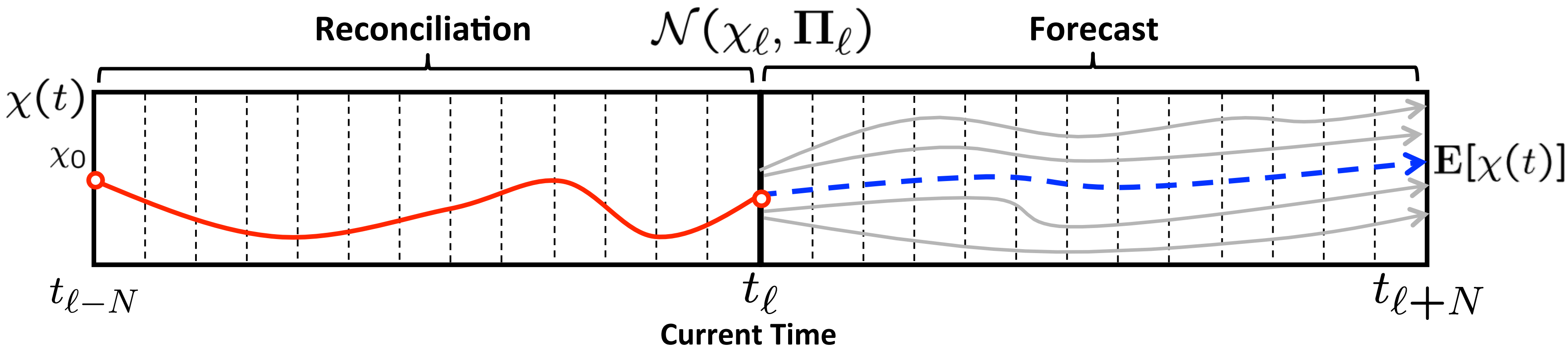
- A very difficult distribution to sample from.
- Solution: first, find the best estimate of the state.
- Then, approximate the prior covariance by an ergodic/Gaussian Process method.

Do I really need to do it by myself?

- NWS gives excellent estimates for large scale patterns, but:
- NWS gives me too coarse of a resolution forecast for energy apps. I need resolution because wind has enormous variability
- NOAA/NCEP gives me an even coarser resolution for uncertainty forecast (200km).
- Data at heights relevant for wind applications (100m) is most often not reported (emphasis on surface).



Idea : best estimate + covariance estimation + ensembles



•Use some form of an ergodic hypothesis. Take $d_{ij} \in \mathbb{R}^{N\times(2\times 30\text{days})}$,

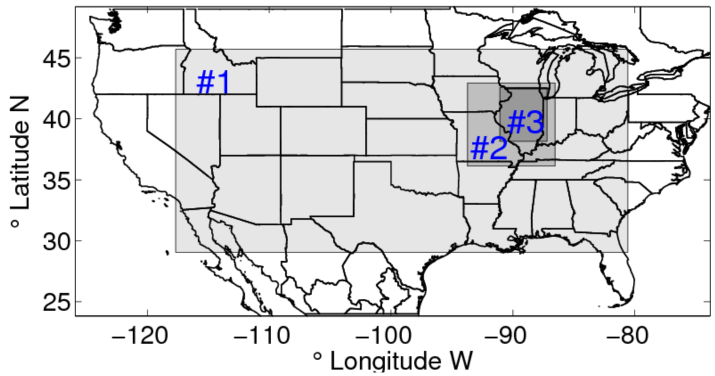
$$\mathbf{V}_{ik} \approx \mathbf{d}\mathbf{d}^{\mathbf{T}} = \sum_j d_{ij} d_{kj}^T = \epsilon_i \cdot \epsilon_k = \left[\begin{array}{c|c|c|c} \epsilon_0\cdot\epsilon_0 & \epsilon_1\cdot\epsilon_0 & \cdots & \epsilon_n\cdot\epsilon_0 \\ \epsilon_0\cdot\epsilon_1 & \epsilon_1\cdot\epsilon_1 & \cdots & \epsilon_n\cdot\epsilon_1 \\ \cdots & \cdots & \cdots & \cdots \\ \epsilon_0\cdot\epsilon_n & \epsilon_1\cdot\epsilon_n & \cdots & \epsilon_n\cdot\epsilon_n \end{array}\right], \quad \mathbf{C}_{ik} = \frac{\epsilon_i\cdot\epsilon_k}{|\epsilon_i||\epsilon_k|}.$$

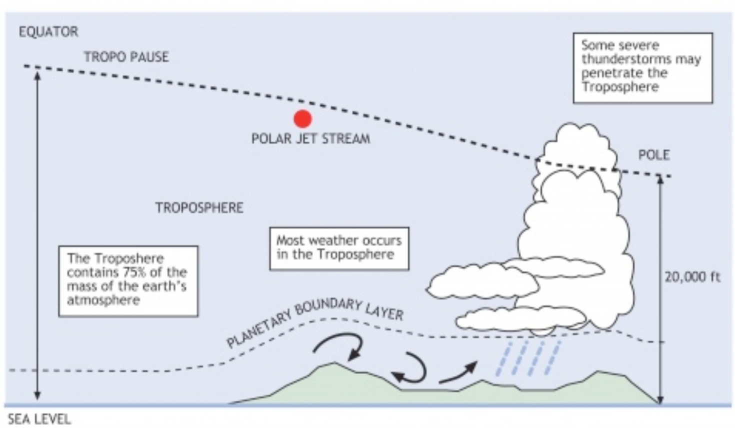
•Our uncertainty model for multi-period forecasting:

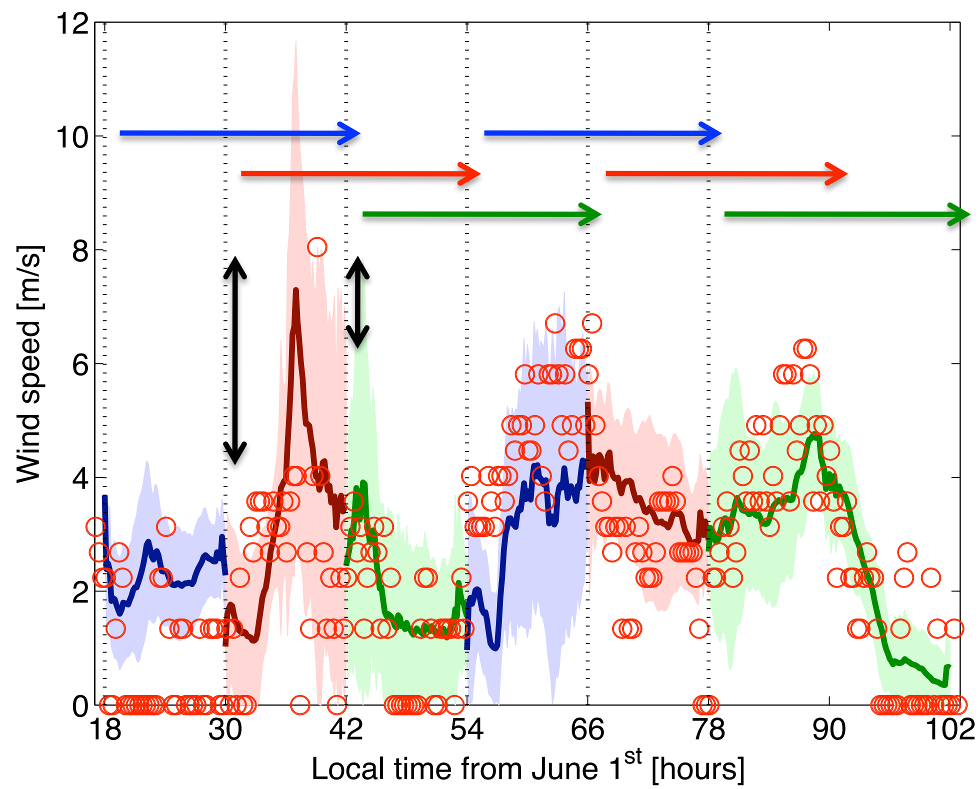
- •Mean: Best Estimate using a variational methods.
- •Covariance: Fixed correlations at current moment, variance adjusted from physical data and best estimates update size.
- •Sampling: Take samples, propagate through the model = empirical ensemble

Short-term wind speed prediction using numerical weather prediction models

- Designed an **UQ framework** to generate forecasts with confidence intervals

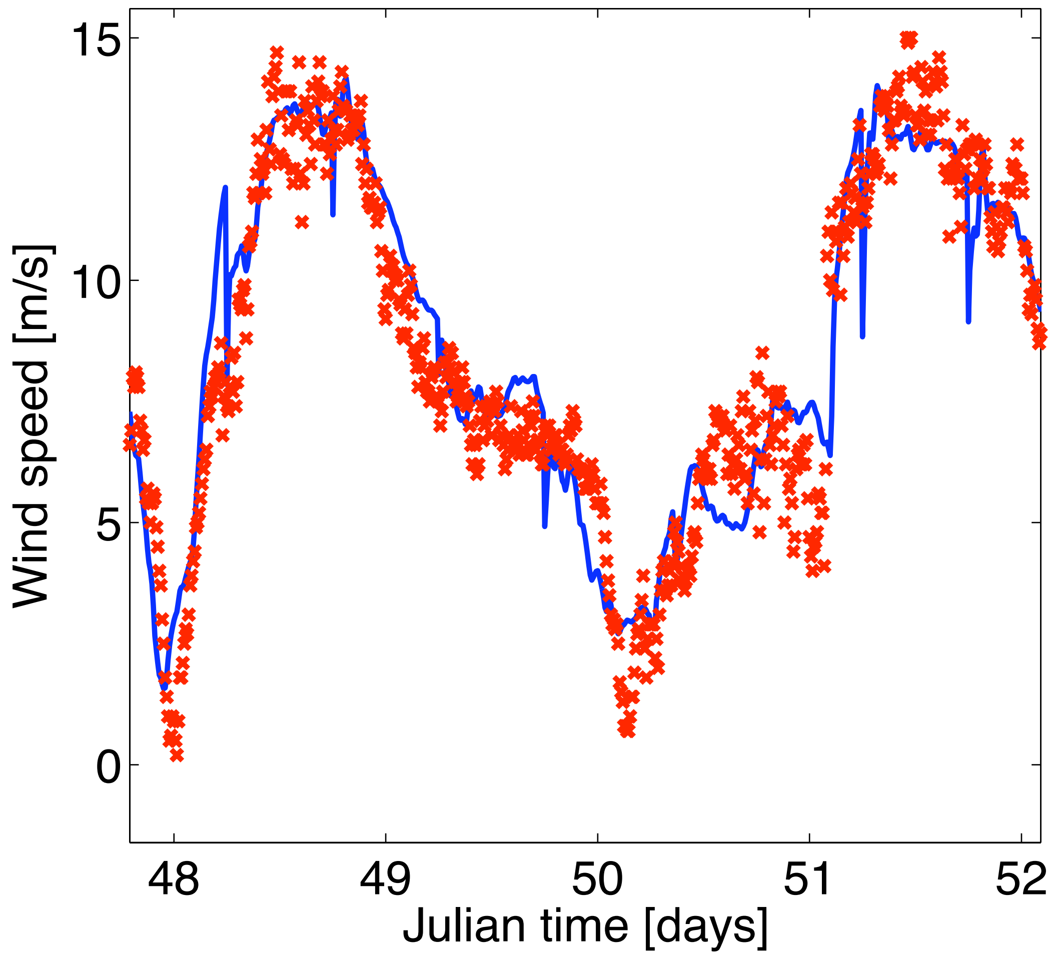






Long-term predictions: wind speed

- **Wind 2009**: at 80m above ground



Conclusions

- NLMPC can model complex energy systems. Physical layer, financial and economic effects, commodity markets.
- We proved that sequential quadratic programming is a good solution for NLMPC; relaxations can be used for real-time applications.
- We have demonstrated that forecast and stochastic programming matters for energy applications; the costs do not increase by much but we are always feasible – so benefit of considering uncertainty is in the reliability.
- We presented several techniques for scalable stochastic quadratic program used in NLMPC. We investigated – successfully – scaling to 130K processors for problems with 3B variables.

Future work

- Better uncertainty models for weather forecast (how do we eliminate the nagging alpha parameter and have a more solidly founded uncertainty approach).
- Other linearization approaches for NLMPC with even cheaper subproblems.
- Do the conclusions hold when I have to add more details in the simulation (e.g transmission lines, model the airflow in the building, etc …)
- New math / stat for stochastic programming
  - Asynchronous optimization
  - SAA error estimate
- New scalable methods for a more efficient software
  - Better interconnect between (iterative) linear algebra and sampling
    - importance-based preconditioning
    - multigrid decomposition
  - Target: emerging exa architectures

References

- Victor Zavala, Emil Constantinescu, Theodore Krause, and Mihai Anitescu." On-line economic optimization of energy systems using weather forecast information." *Journal of Process Control*, Volume 19, Issue 10, 2009, Pages 1725-1736
- Emil Constantinescu, Victor Zavala, Matthew Rocklin, Sangmin Lee, and Mihai Anitescu. A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment with Wind Power Generation. IEEE Transactions on Power Systems, 26(1), pp. 431--441 (2011)
- Victor Zavala and Mihai Anitescu. On-Line Nonlinear Programming as a Generalized Equation. *SIAM J. Control Optim.* Volume 48, Issue 8, pp. 5444-5467 (2010).
- Cosmin Petra and Mihai Anitescu. "A preconditioning technique for Schur complement systems arising in stochastic optimization". Preprint ANL/MCS-P1748-0510. Submitted to *Computational Optimization and Applications*.
- Miles Lubin, Cosmin G. Petra, Mihai Anitescu, and Victor Zavala.Scalable Stochastic Optimization of Complex Energy Systems. Preprint ANL/MCS-P1858-0311