Using High Performance Computing for the Management under Uncertainty of Energy Systems

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

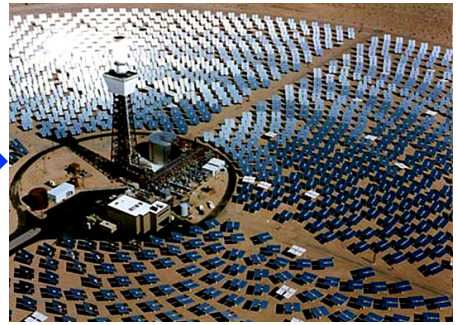
Princeton, Oct 2010

Motivation: Management of Energy Systems under Ambient Conditions Uncertainty

Ambient Condition Effects in Energy Systems

- **Operation of 90% of Energy Systems is Affected by Ambient Conditions**
  - Power Grid Management: **Demands** *(Douglas, et.al. 1999)*
  - Power Plants: **Production Levels** *(General Electric)*
  - Petrochemical: **Heating and Cooling** *(ExxonMobil)*
  - Buildings: **Heating and Cooling** *(Braun, et.al. 2004)*
  - **(Focus) Next Generation Energy Systems assume a major renewable energy penetration: Wind + Solar + Fossil** *(Beyer, et.al. 1999)*
- **Efficiency (reduction of waste) becoming a major concern: Focus on optimal management, not only optimal design**

Question 1: What are the benefits versus the costs of <u>anticipating</u> weather conditions?

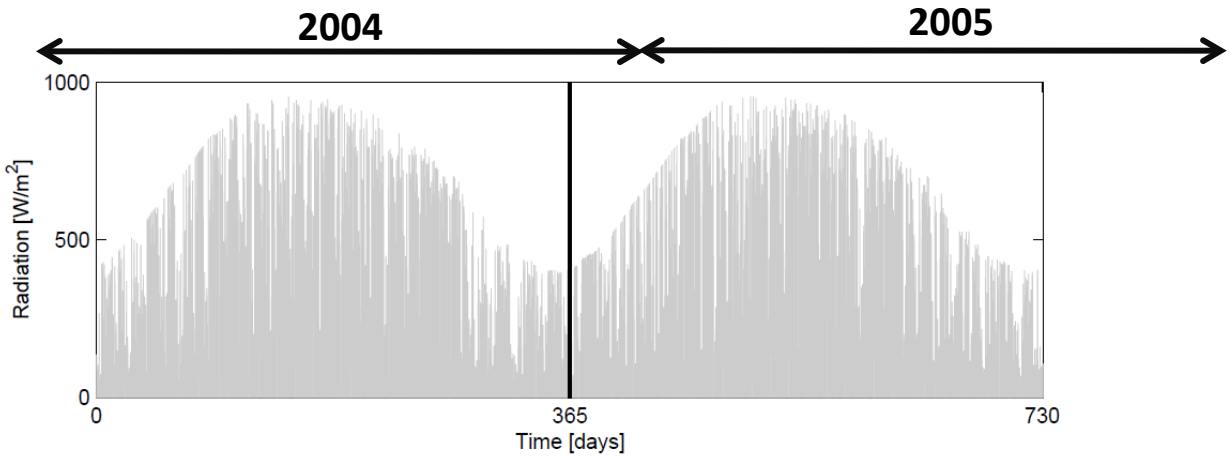


Variability/Uncertainty in Ambient Conditions

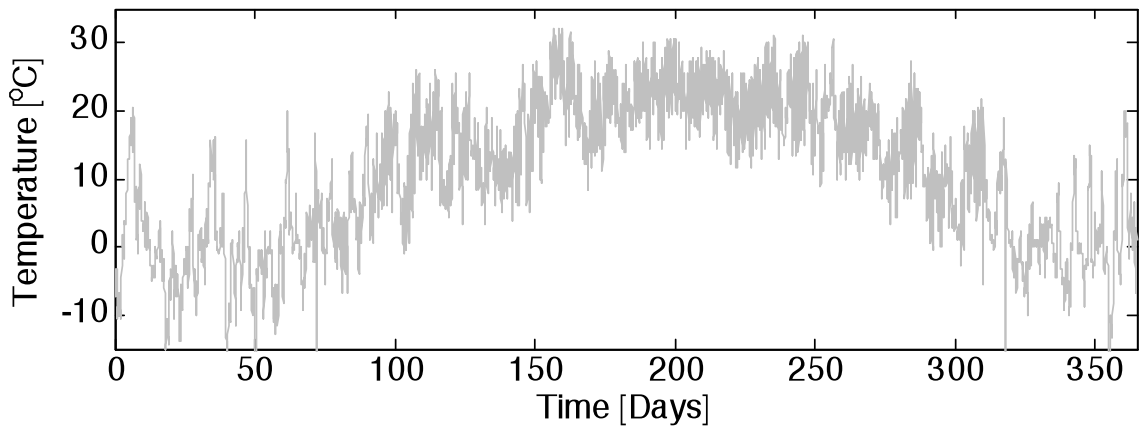
Weather Conditions (Temperature, Wind Speed, Humidity …) have high variability/uncertainty

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

Total Ground Solar Radiation
Chicago, IL

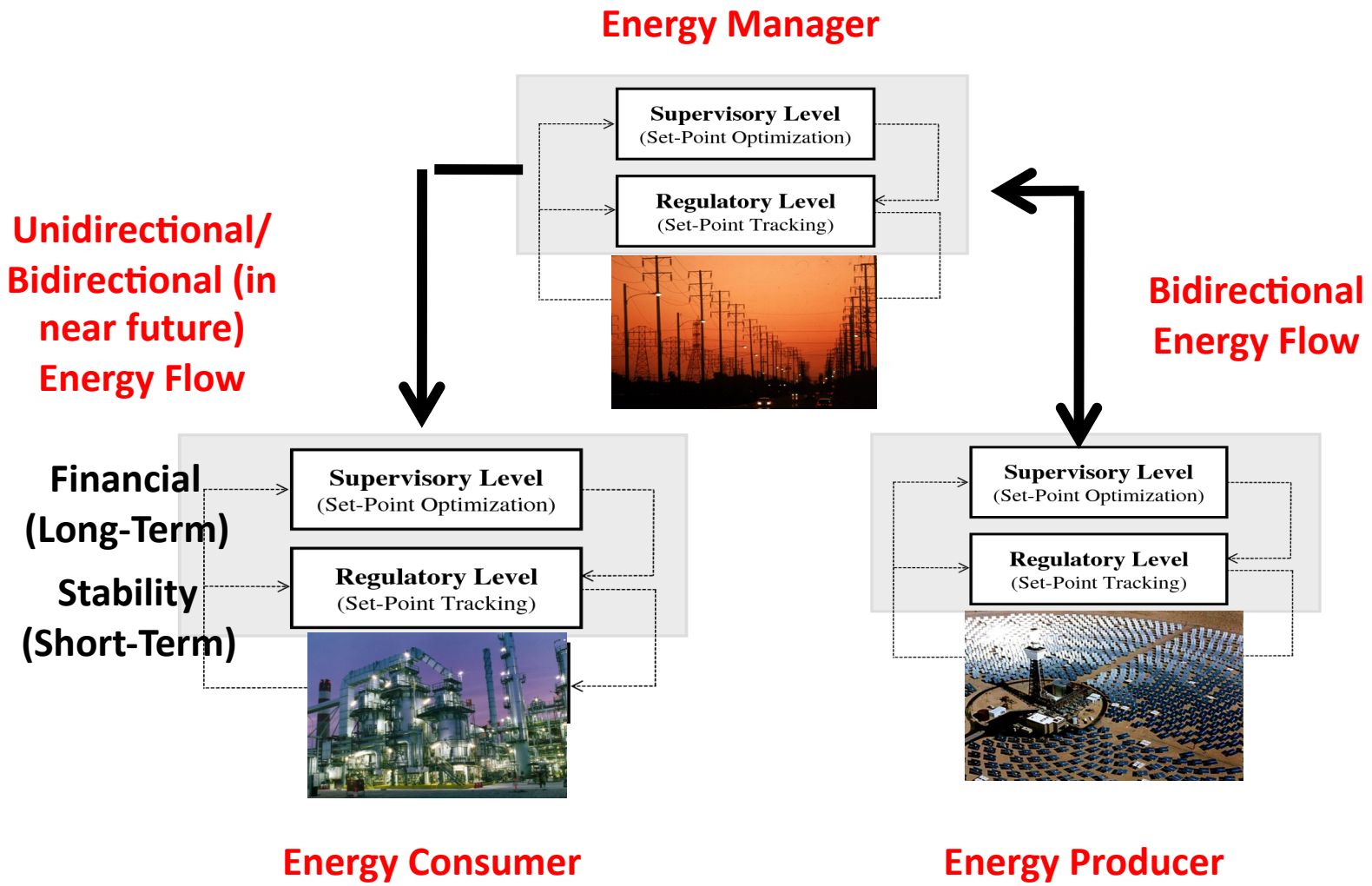


Ambient Dry-Bulb Temperature
Pittsburgh, PA



Question 2: How to Handle Uncertainty?

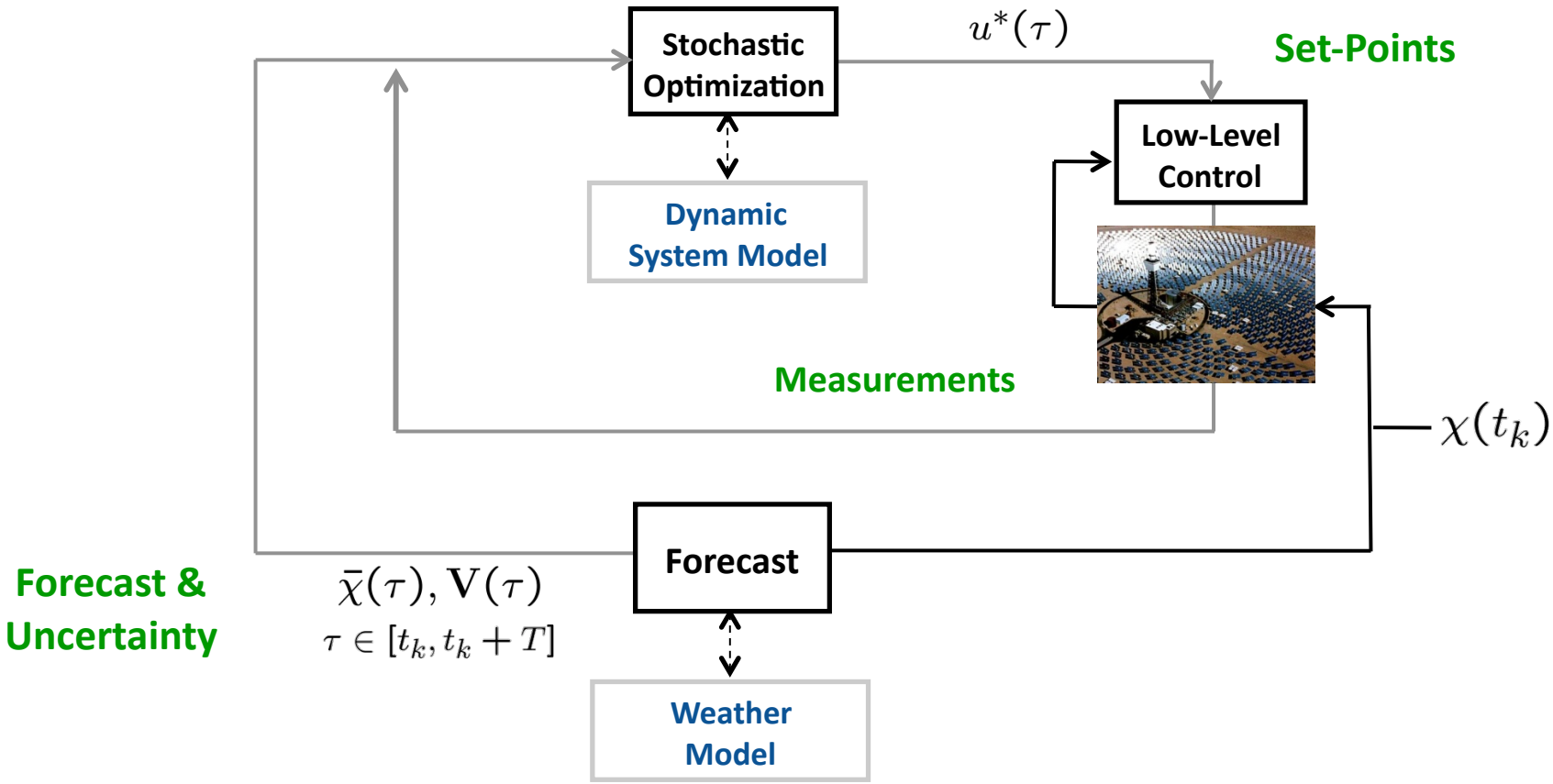
Hierarchical Management in Energy Systems



Optimization Traditionally <u>Reactive</u> (short-term) but econ is long-term ….

Question 3: How to handle the management of energy systems under uncertainty?

Abstraction: Basic Operational Setting



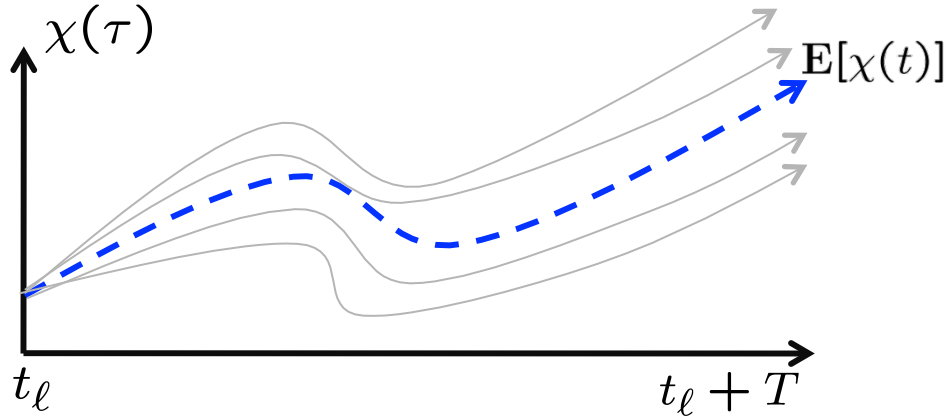
Dynamic Uncertainty: Receding Horizon Optimization

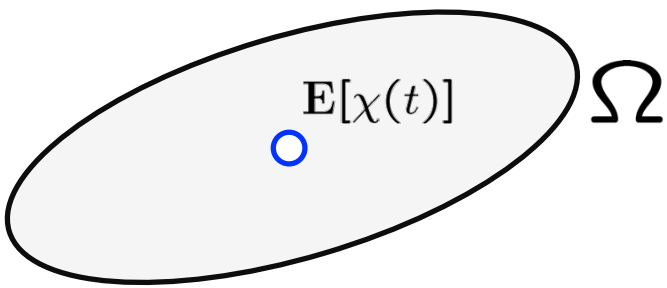
Benefits: Accommodate Forecasts, Constraint Handling, Financial Objectives, Complex Models

Deterministic

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

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





Stochastic Programming

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

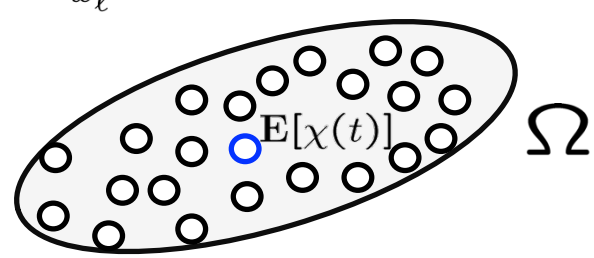
$$\left.\begin{aligned} \frac{dz}{dt} &= \mathbf{f}(z(t), y(t), u(t), \chi(t)) \\ 0 &= \mathbf{g}(z(t), y(t), u(t), \chi(t)) \\ 0 &\geq \mathbf{h}(z(t), y(t), u(t), \chi(t)) \end{aligned}\right\} \quad \forall \chi(t) \in \Omega$$

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

Complexity (Solution Time)

1,000 – 10,000 Differential-Algebraic Equations

100-1000 Scenarios



Digression about the Features of Stochastic Programming for Energy Systems w Renewables

SAA Stochastic Programming Approximation.

$$
\begin{aligned}
\min_{\mathbf{u}} \quad & \frac{1}{S}\sum_{k=1}^{S} \varphi(z_k, y_k, \mathbf{u}, \chi_k) \\
\text{s.t.} \quad & \mathbf{c}(z_k, y_k, \mathbf{u}, \chi_k) = 0 \\
& \mathbf{h}(z_k, y_k, \mathbf{u}, \chi_k) \leq 0 \\
& k = 1, \ldots, S
\end{aligned}
$$

- Stochastic programming/SAA is a well-studied paradigm of operations research.
- Nevertheless, in energy systems, the distribution – at least of weather forecast -- is *not* given. It has to be modeled from data using some knowledge of the application (and sometimes, distribution modeling *is more computing intensive than obtaining the optimal decision*).
- Energy Systems have a *significant* physical layer. Therefore one scenario can have a significant number of degrees of freedom.

OUTLINE

- **Question 1: What are the benefits versus the costs of <u>anticipating</u> weather conditions?**
- **Question 2: How do we handle ambient conditions uncertainty?**
- **Question 3: How to handle the management of energy systems under uncertainty?**

- Sampling from the distribution of the ambient conditions (Q2).
  - How do I manipulate this large uncertainty space?
- Benefits of considering uncertainty in energy systems (Q3, Q1)
  - 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? (Q3)
  - … Since they have a large physical layer.
  - How do I obtain scalable algorithms for solving them

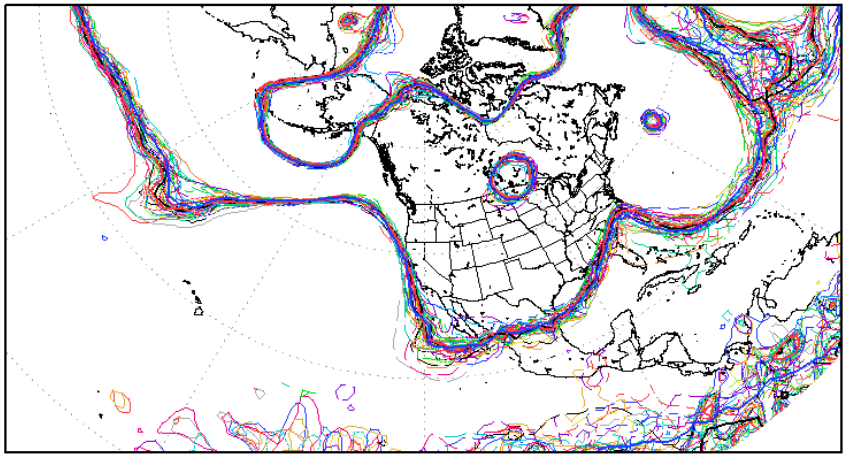
1. Sampling from the distribution of ambient conditions ("uncertainty quantification" of ambient conditions). – Question 2

Physical Model-Based Statistical Model of Ambient Conditions

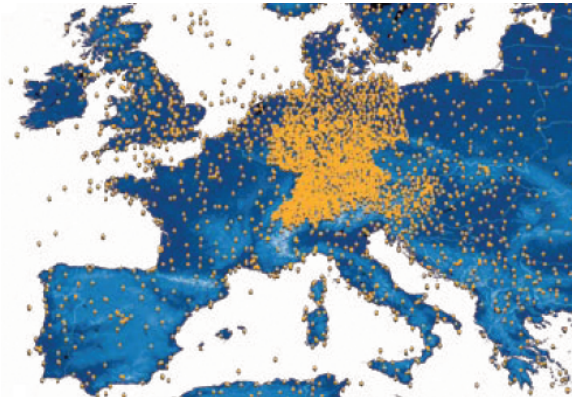
- Advanced Meteorological Models (WRF)
  - Detailed, First Principles, Physico-Chemical Phenomena
  - High Complexity 4-D Fields (10^6 States)

- Model Reconciled to Measurements From Multiple Stations

- Reconciliation Techniques:
  - 3-D Var *Courtier, et.al. 1998*
  - 4-D Var (Moving Horizon Estimation) *Navon et.al., 2007*
  - Extended and Ensemble Kalman Filter *Eversen, et.al. 1998*



http://www.emc.ncep.noaa.gov/gmb/ens/



http://www.meteomedia.com/

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 $k = 0, \cdots, K$

$$x_k^{in} = M(x_{k-1}^{in}) + W_k, \quad (1)$$

$$z_k^{obs} = H(x_k^{in}) + V_k, \quad (2)$$

where

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

and

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

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 …

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

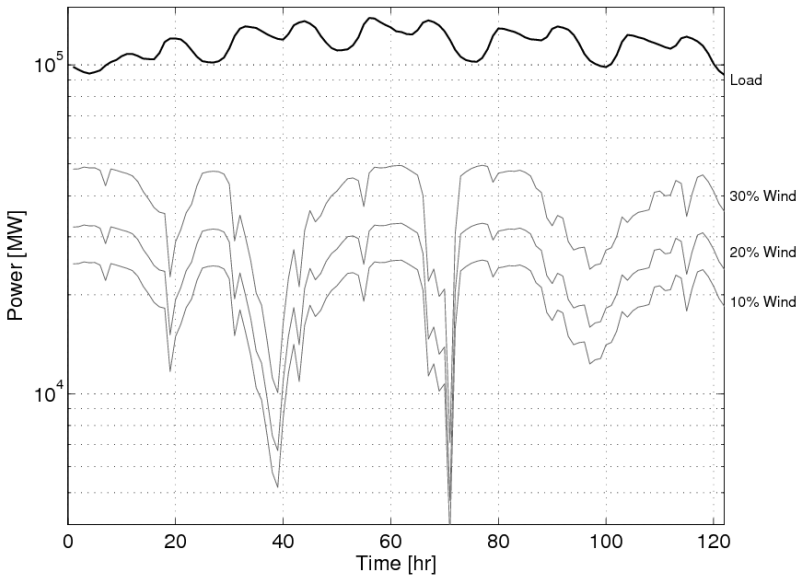
$$f(\boldsymbol{X}^{in}, \boldsymbol{Z}^{obs}) = \sum_{i=0}^{k} (\boldsymbol{x}_i^{in} - \bar{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}(t_{i-1}, \boldsymbol{x}_{i-1}^{in}))^T Q_i^{-1} (\boldsymbol{x}_i^{in} - \bar{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}(t_{i-1}, \boldsymbol{x}_{i-1}^{in}))$$
$$+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

- 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.

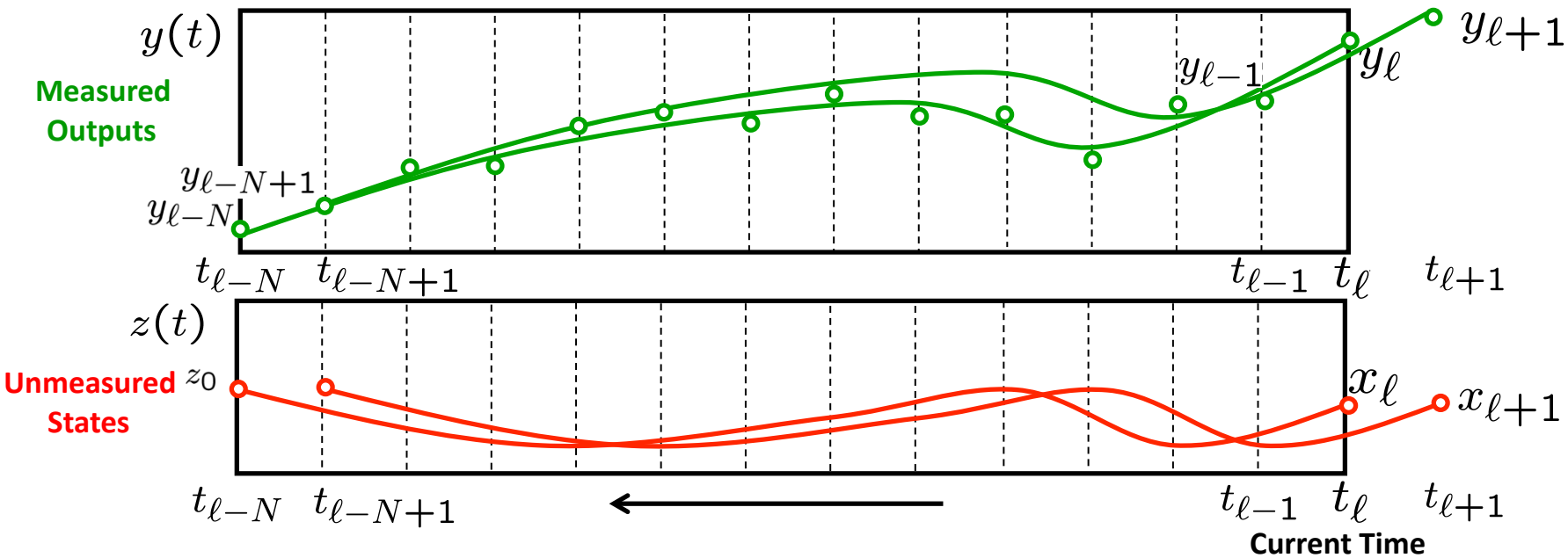
Do I really need to do it by myself?

- 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).



Step 1: Moving Horizon Best State Estimation

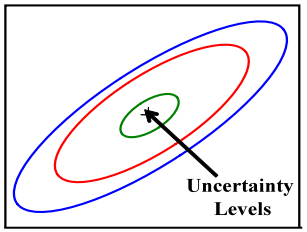


$$\min_{p(t),\, z_0} \sum (y(t_k) - \underline{y_{\ell-k+N}})^T \mathbf{V}_y^{-1} (y(t_k) - y_{\ell-k+N})$$

WRF Model

$$\begin{aligned} \frac{dz}{dt} &= \mathbf{f}\left(z(t), p(t), u(t)\right) \\ y(t) &= \mathbf{g}(z(t), p(t), u(t)) \\ z(0) &= z_0 \end{aligned}$$

Uncertain



$\mathbf{\Pi}_\ell \to 0$

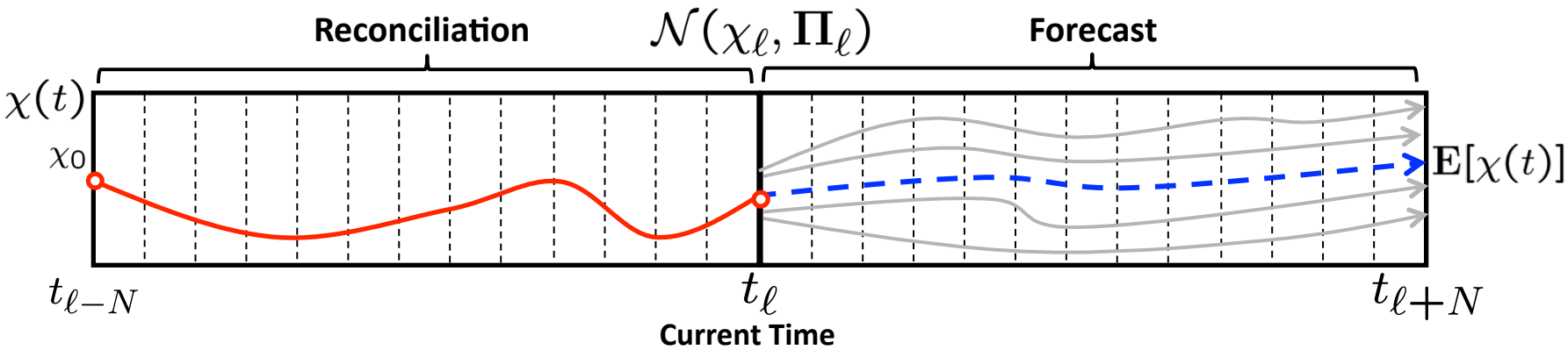
$$\min_{p(t),\, z_0} \sum (y(t_k) - y_{\ell-k+N+1})^T \mathbf{V}_y^{-1} (y(t_k) - y_{\ell-k+N+1})$$

$$\begin{aligned} \frac{dz}{dt} &= \mathbf{f}\left(z(t), p(t), u(t)\right) \\ y(t) &= \mathbf{g}(z(t), p(t), u(t)) \\ z(0) &= z_0 \end{aligned}$$

Uncertainty in Current State x_ℓ

Needed To Quantify <u>Future</u> Forecast

Step 2: Estimate the prior covariance matrix.



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

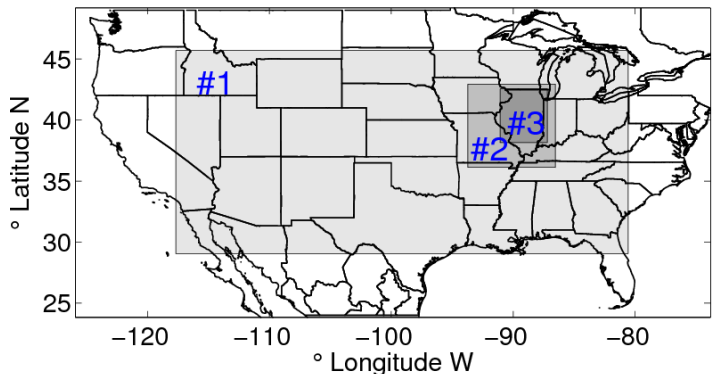
$$\mathbf{V}_{ik} \approx \mathbf{d}\mathbf{d}^{\mathbf{T}} = \sum_j d_{ij} d_{kj}^T = \epsilon_i \cdot \epsilon_k = \left[\begin{array}{c|c|c|c} \epsilon_0 \cdot \epsilon_0 & \epsilon_1 \cdot \epsilon_0 & \cdots & \epsilon_n \cdot \epsilon_0 \\ \epsilon_0 \cdot \epsilon_1 & \epsilon_1 \cdot \epsilon_1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

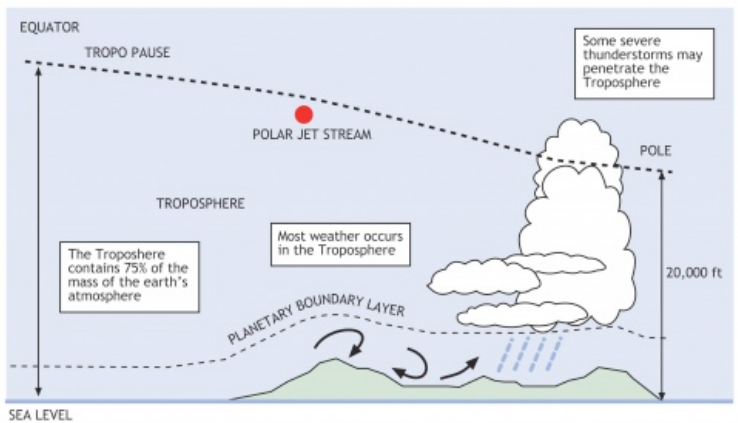
•Our uncertainty model for multi-period forecasting:

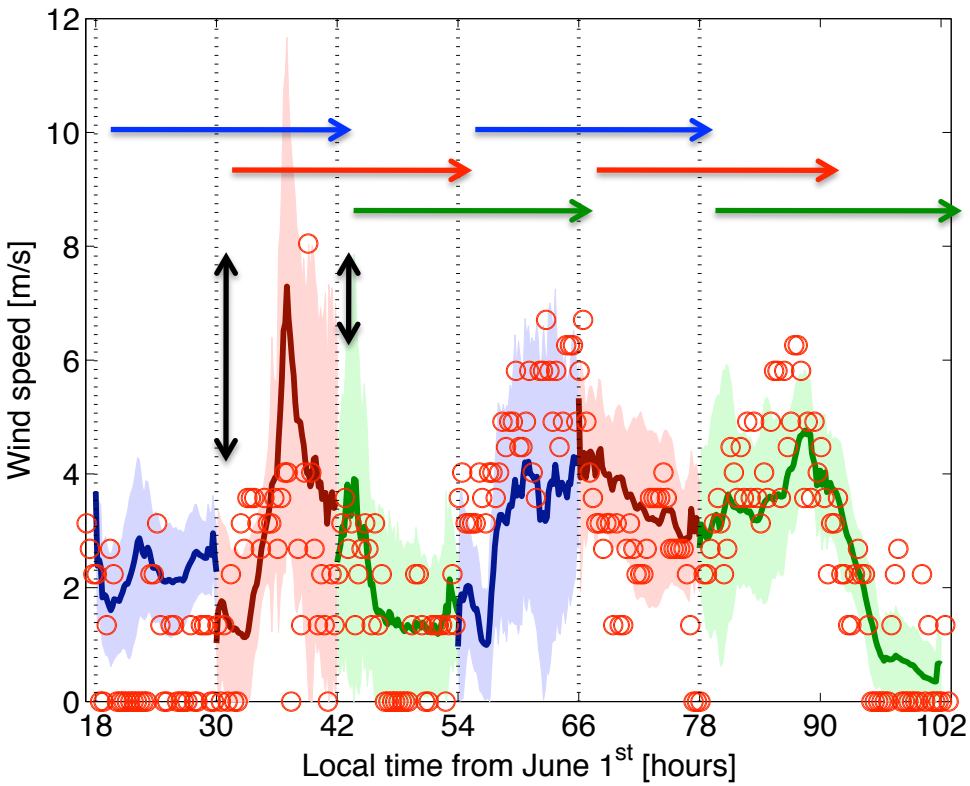
- •Mean: Best Estimate.
- •Covariance: Fixed correlations at current moment, variance adjusted from physical data and best estimates update size.
- •Sampling: Take samples, propagate through the model = empirical ensemble

Short-term wind speed prediction with uncertainty using numerical weather prediction models

- Designed a multiscale **UQ framework** to generate forecasts with confidence intervals



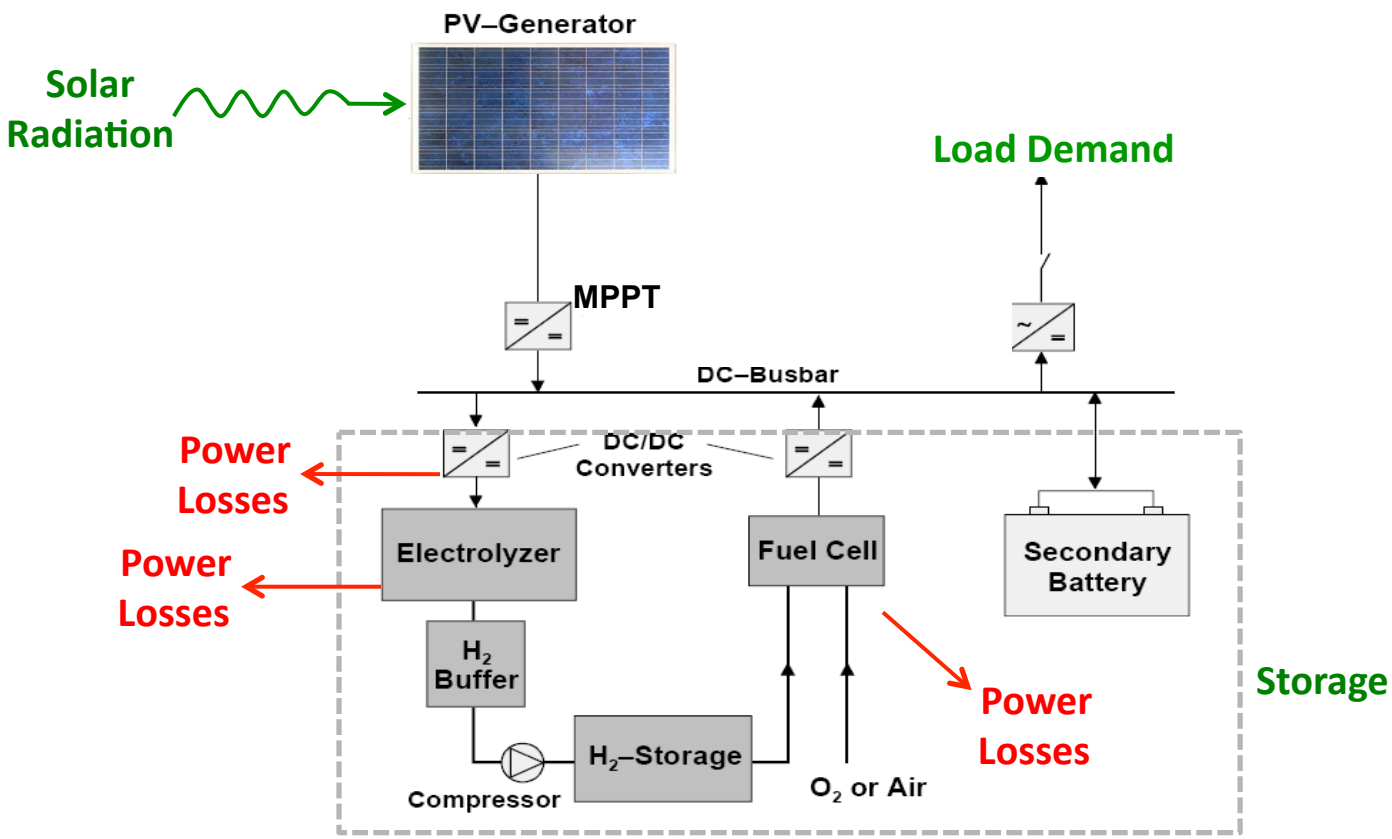




2. Benefits of considering uncertainty in energy systems.

Question 1

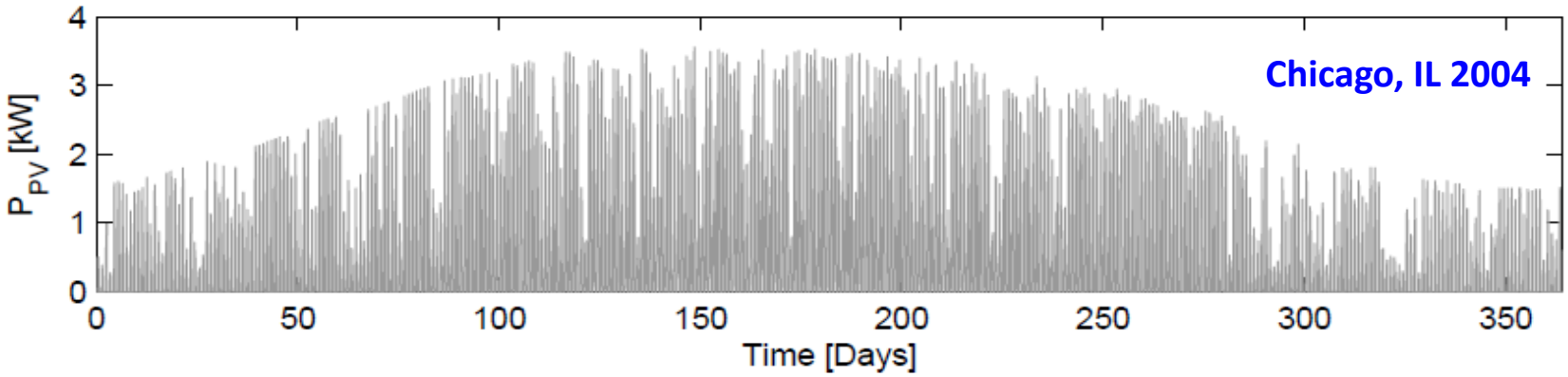
Hybrid Photovoltaic-H_2 System



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

Hybrid Photovoltaic-H_2 System

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



True Future Radiation

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

Minimize Operating Costs + Maximize H_2 Production

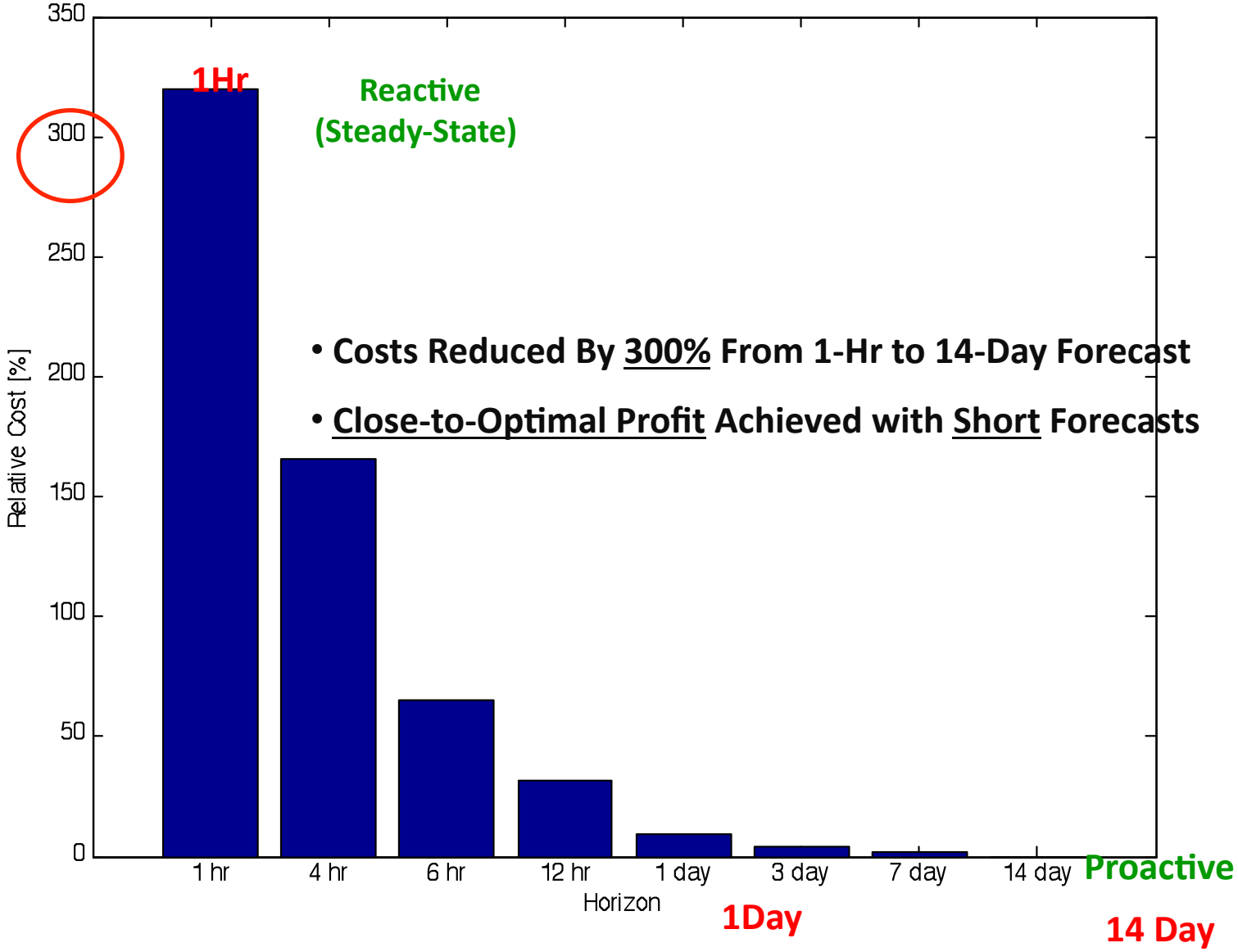
$$\begin{aligned} \frac{dz}{dt} &= \mathbf{f}(z(t), y(t), u(t), \chi(t)) \\ 0 &= \mathbf{g}(z(t), y(t), u(t), \chi(t)) \\ 0 &\geq \mathbf{h}(z(t), y(t), u(t), \chi(t)) \\ z(0) &= x_\ell \end{aligned}$$

Energy Balances

State-of-Charge, Fuel Cell and Electrolyzer Limits

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

Hybrid Photovoltaic-H_2 System

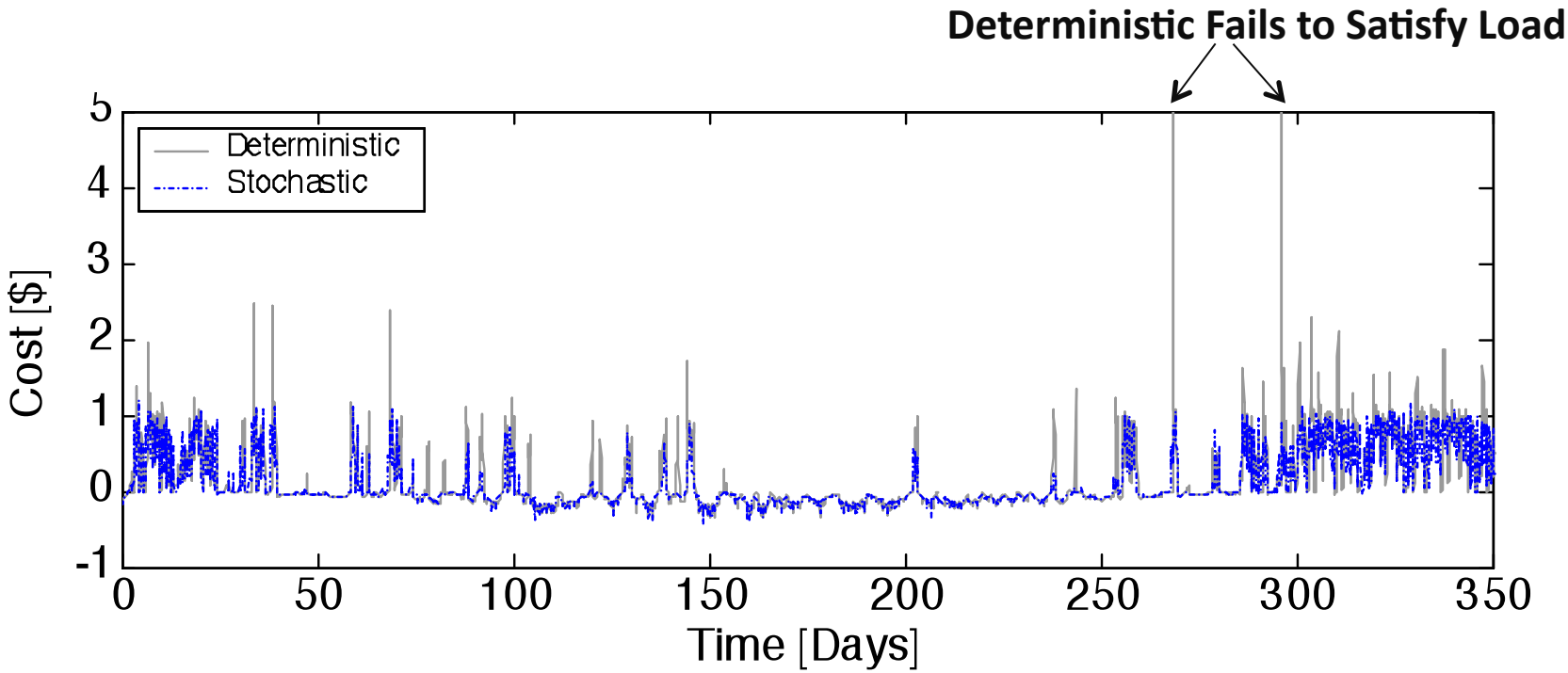


- Costs Reduced By 300% From 1-Hr to 14-Day Forecast
- Close-to-Optimal Profit Achieved with Short Forecasts

Forecast Matters!

Hybrid Photovoltaic-H_2 System

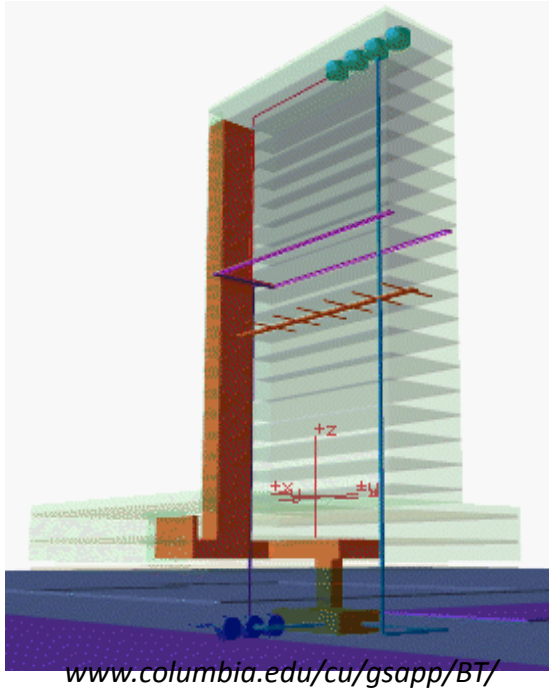
Load Satisfaction Deterministic ("Optimization on Mean") vs. Stochastic



Therefore, the alternative to stochastic programming can turn out infeasible – considering uncertainty matters for robustness !!

Handling Stochastic Effects Particularly Critical in Grid-Independent Systems

Thermal Management of Building Systems



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

Minimize Annual Heating and Cooling Costs

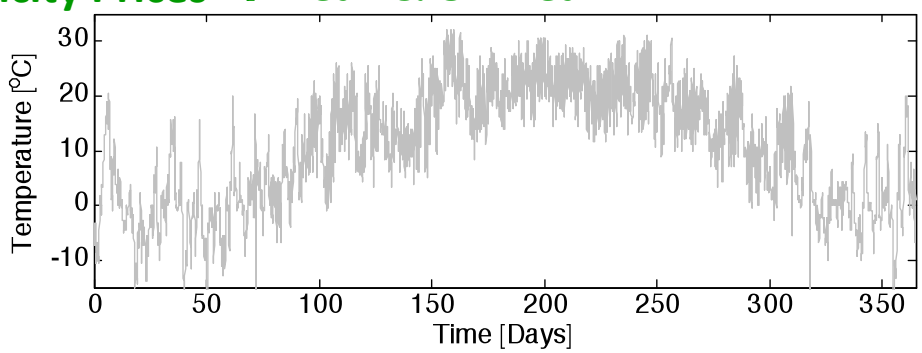
$$\min_{u(t)} \quad \int_{t_\ell}^{t_{\ell+N}} \left[C_c(t)\varphi_c(t) + C_h(t)\varphi_h(t)\right] dt$$

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

Energy Balances

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

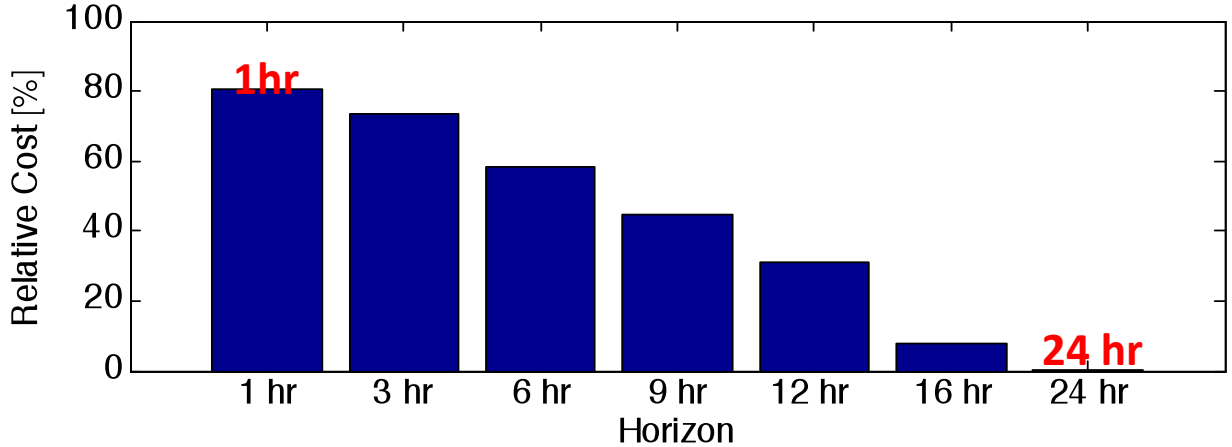
Time-Varying Electricity Prices → Peak & Off-Peak



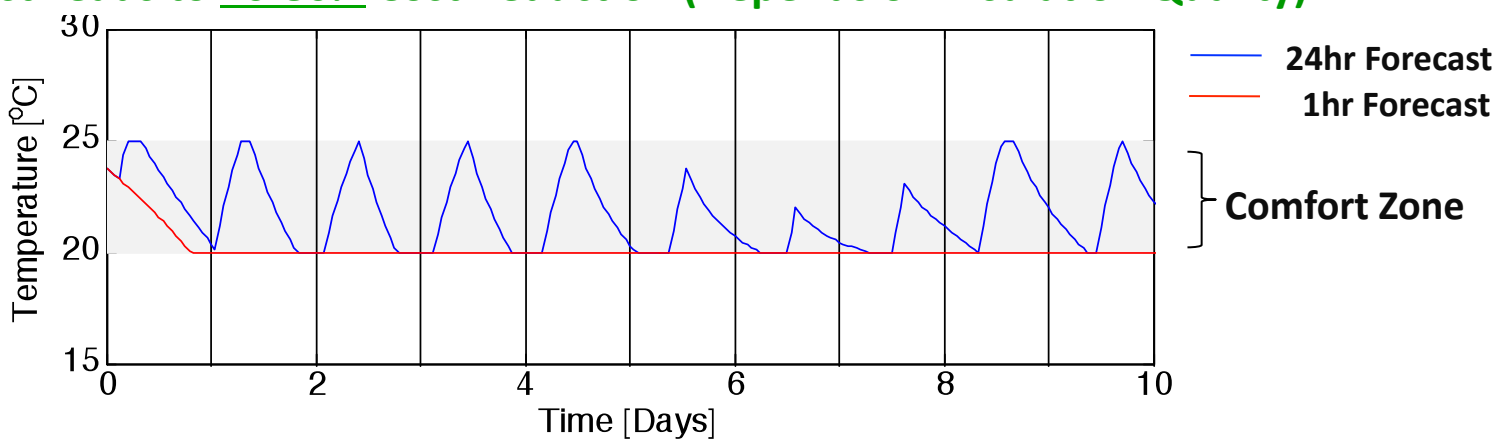
Pittsburgh, PA 2006

Thermal Management of Building Systems

Effect of Forecast on Energy Costs



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

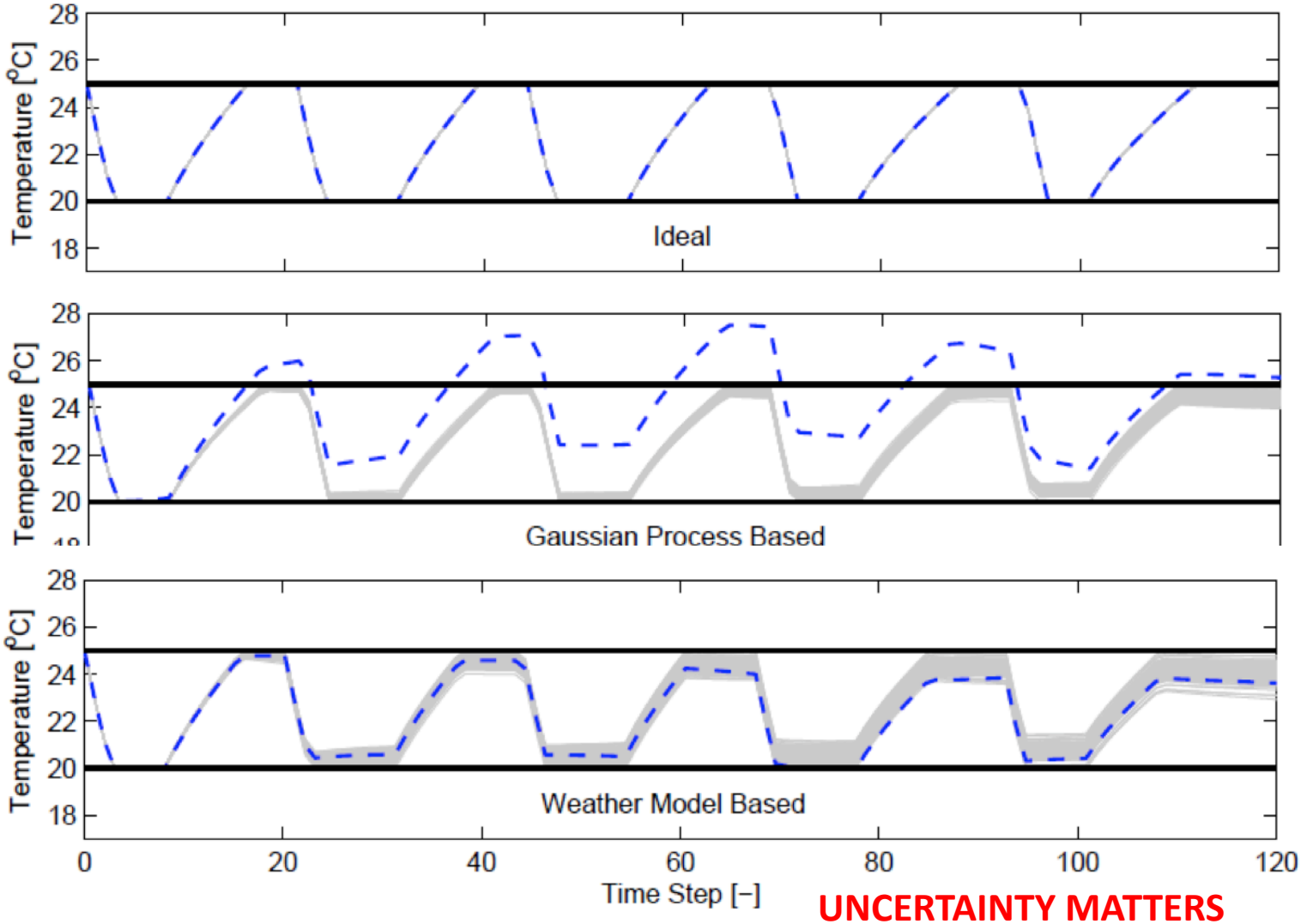


Exploit Comfort Zone and Weather Info to Heat/Cool when Cheaper *Braun, 1990*

Forecast Matters!

Thermal Management of Building Systems

Performance Optimizer using WRF and GP Model Forecasts



Perfect Forecast

Gaussian Process Model

WRF Model

UNCERTAINTY MATTERS

Stochastic Unit Commitment with Wind Power (SAA)

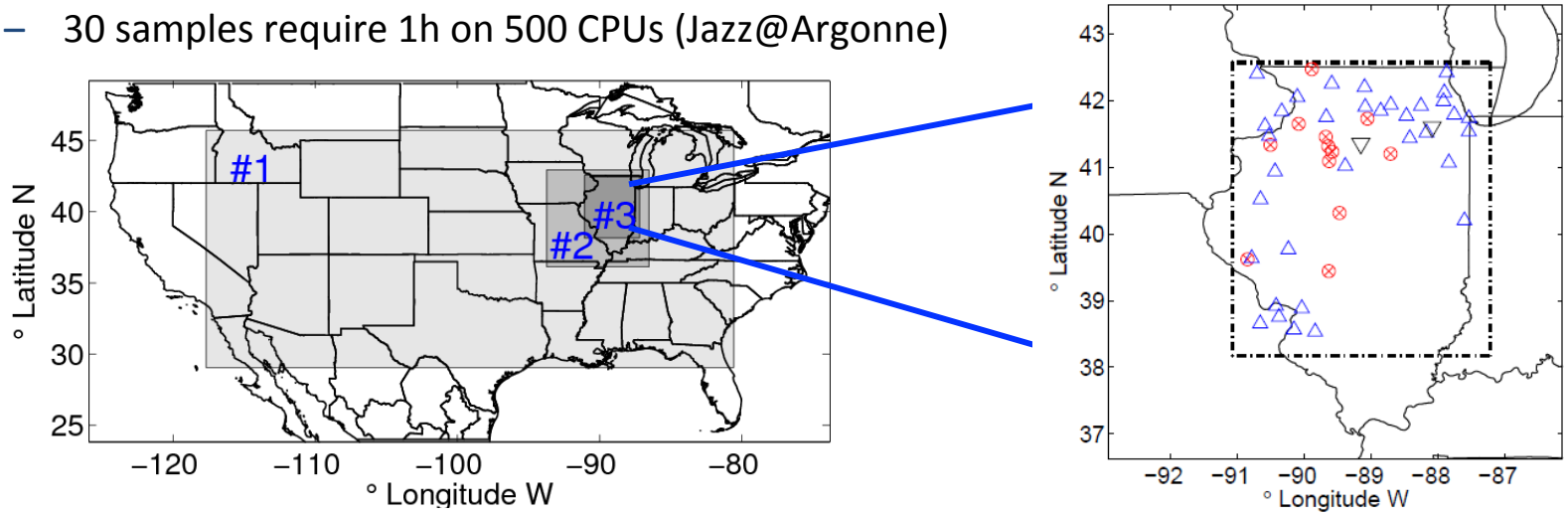
$$\min \quad \text{COST} = \frac{1}{N_s}\sum_{s\in\mathcal{S}}\left(\sum_{j\in\mathcal{N}}\sum_{k\in\mathcal{T}} c^p_{sjk} + c^u_{jk} + c^d_{jk}\right)$$

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

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

ramping constr., min. up/down constr.

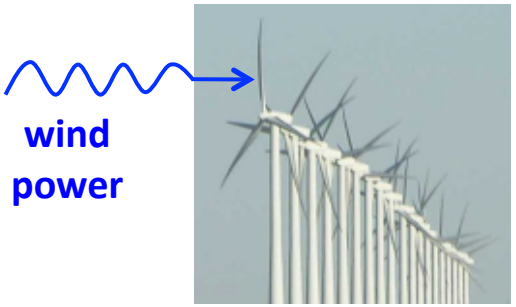
- Wind Forecast – WRF(Weather Research and Forecasting) Model
  - Real-time grid-nested 24h simulation
  - 30 samples require 1h on 500 CPUs (Jazz@Argonne)



Zavala & al 2010.

Wind power forecast and stochastic programming

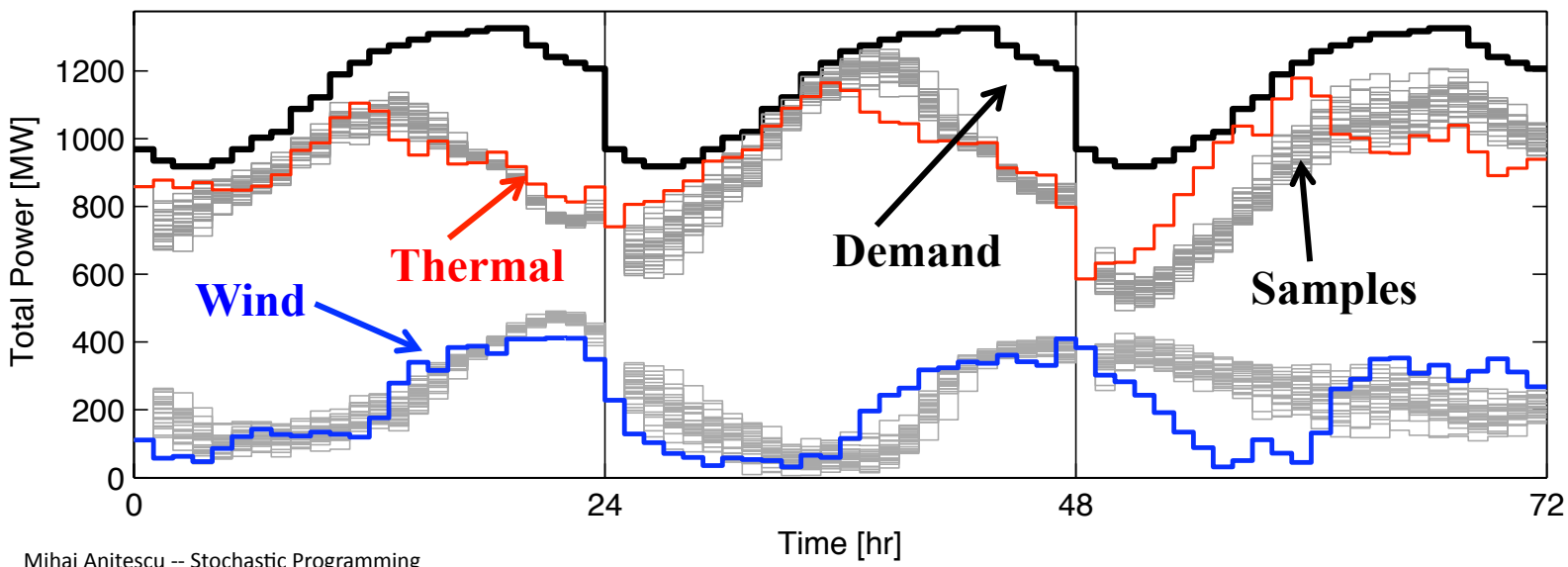
- Unit commitment & energy dispatch with uncertain wind power generation for the State of Illinois, assuming 20% wind power penetration, using the same windfarm sites as the one existing today.



- 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%.

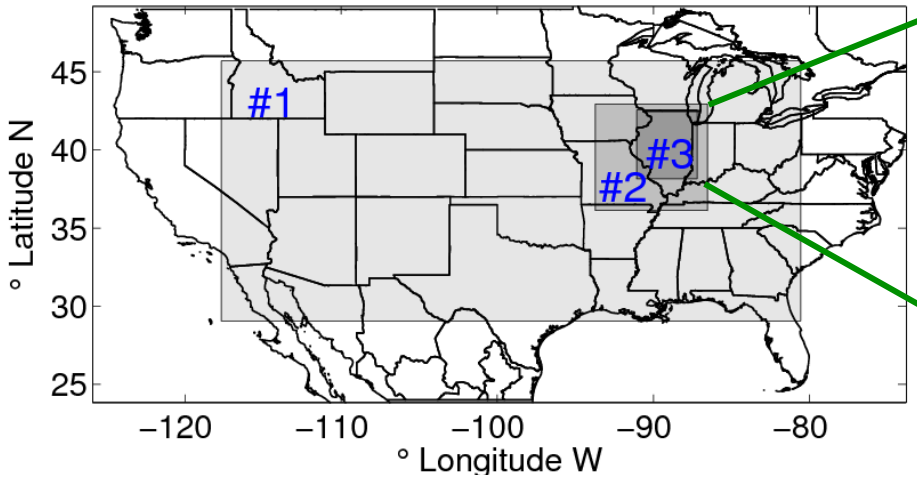
Uncertainty Matters



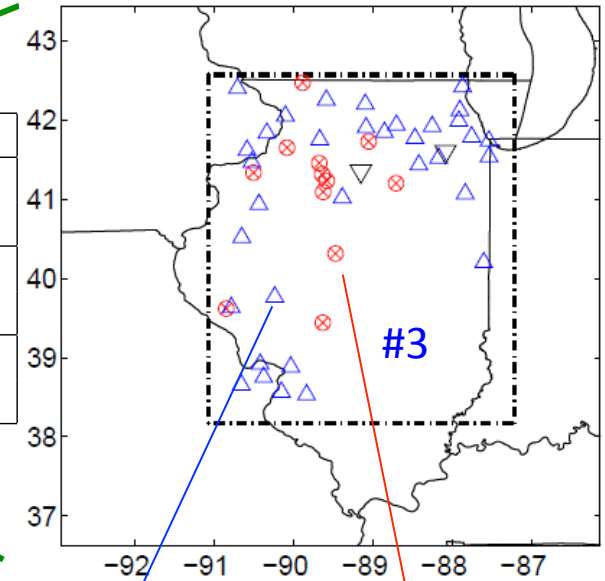
We answered Q1 (uncertainty and forecast matters), how about Q2 (how do I sample from ambient uncertainty?)

Ensemble forecast and uncertainty quantification with WRF on Jazz

- Computational domain setup



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

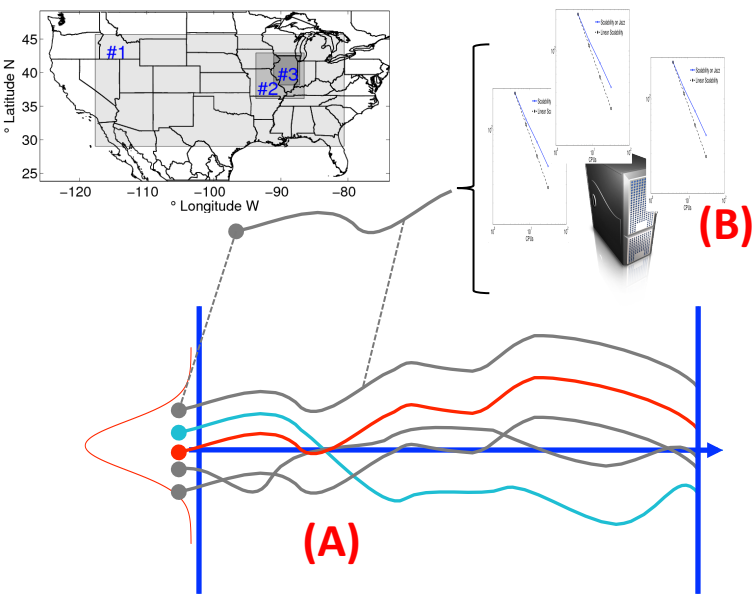


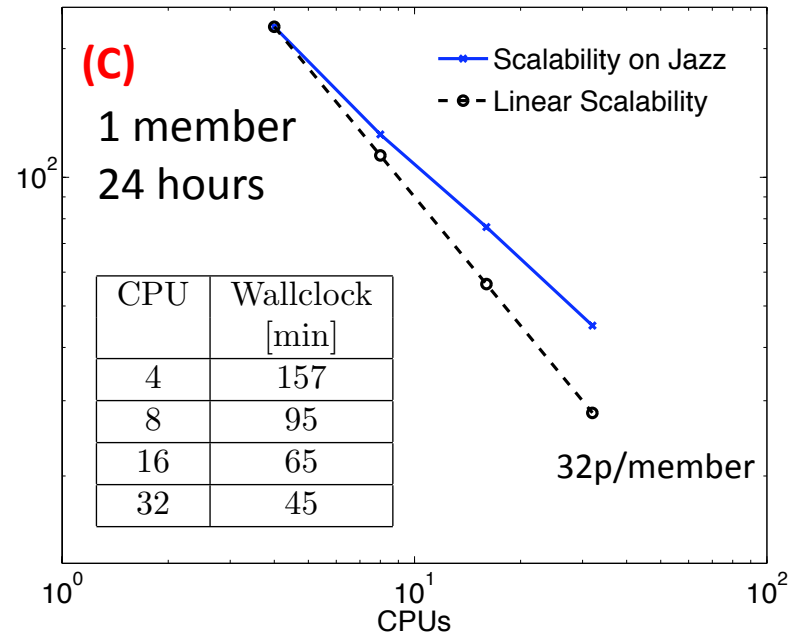
- **Jazz:** 350 nodes; Intel Pentium IV Xeon@ 2.4GHz; 1 or 2 Gb RAM per node; Myrinet 2000 @ 0.25 GB/s, 6-8 µsec latency

- **24 hours** [simulation time] **in one hour** [real time] on Jazz with 30 members on **500 processors**;

WRF scalability on Jazz

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





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

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

- 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 (Question 3)

INTERIOR POINT METHODS FOR STOCHASTIC PROGRAMMING

Managament under uncertainty paradigm: stochastic programming.

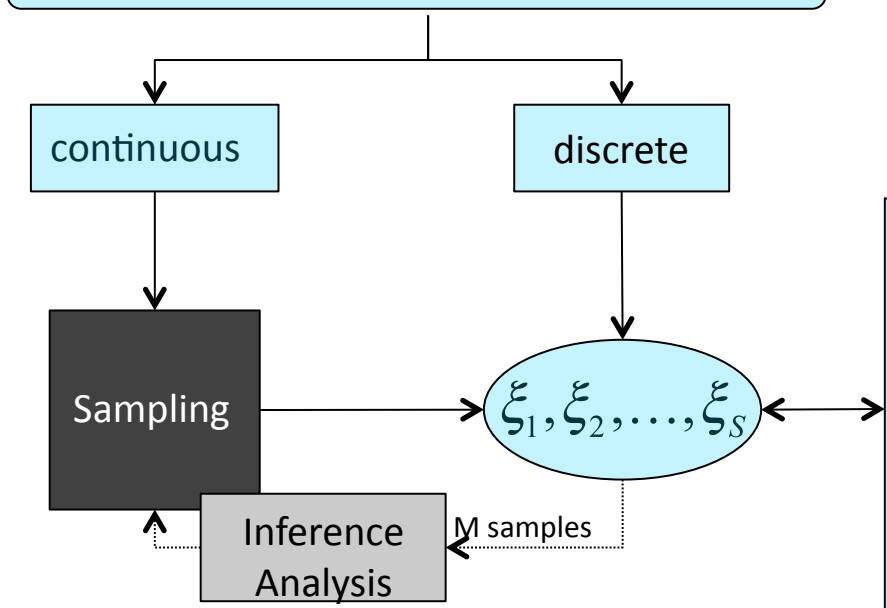
- Two-stage stochastic programming with recourse ("here-and-now")

$$\underset{x_0}{Min}\left\{ f_0(x_0) + \mathbb{E}\left[\underset{x}{Min}\, f(x,\omega) \right] \right\}$$

subj. to.

$$\begin{aligned} A_0 x_0 & & = & \quad b_0 \\ A(\omega)x_0 + & \quad B(\omega)x & = & \quad b(\omega) \\ x_0 \geq 0, & \quad x(\omega) \geq 0 & & \end{aligned}$$

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



Sample average approximation (SAA)

$$\underset{x_0, x_1, x_2, \ldots, x_S}{Min} \quad f_0(x) + \frac{1}{S}\sum_{i=1}^{S} f_i(x_i)$$

subj. to.

$$\begin{aligned} A_0 x_0 & & = & \quad b_0 \\ A_k x_0 + & \quad B_k x_k & = & \quad b_k, \\ x_0 \geq 0, & \quad x_k \geq 0, & & \quad k = 1, \ldots, S \end{aligned}$$

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)

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

The Direct Schur Complement Method (DSC)

- Uses the arrow shape of H

$$\begin{bmatrix} H_1 & & & & G_1^T \\ & H_2 & & & G_2^T \\ & & \ddots & & \vdots \\ & & & H_S & G_S^T \\ G_1 & G_2 & \dots & G_S & H_0 \end{bmatrix} = \begin{bmatrix} L_1 & & & & \\ & L_2 & & & \\ & & \ddots & & \\ & & & L_S & \\ L_{10} & L_{20} & \dots & L_{S0} & L_c \end{bmatrix} \begin{bmatrix} D_1 & & & & \\ & D_2 & & & \\ & & \ddots & & \\ & & & D_N & \\ & & & & D_c \end{bmatrix} \begin{bmatrix} L_1^T & & & & L_{10}^T \\ & L^T{}_2 & & & L_{20}^T \\ & & \ddots & & \vdots \\ & & & L^T{}_S & L_{S0}^T \\ & & & & L^T{}_c \end{bmatrix}$$

- Solving Hz=r

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

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

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

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

Diagonal solve

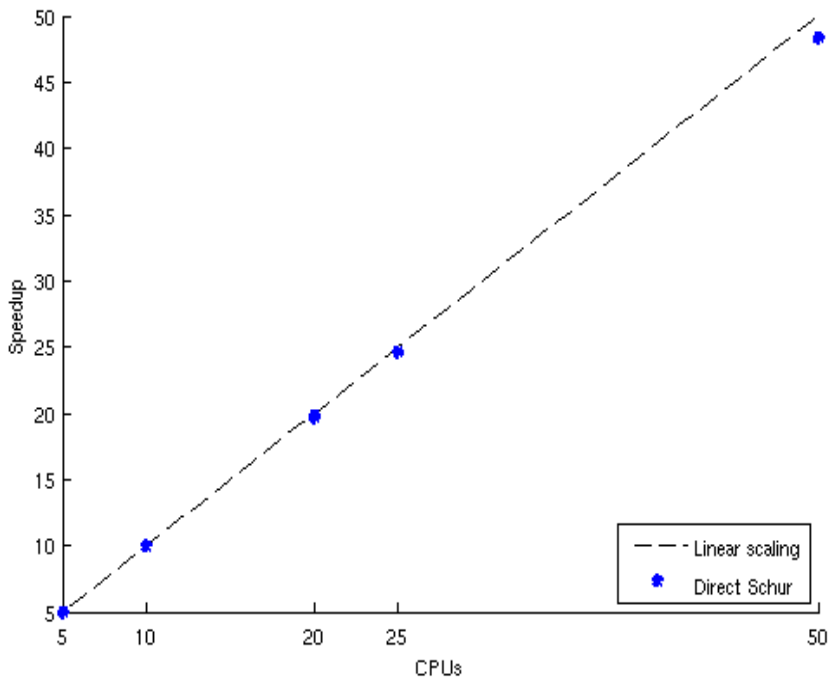
$$\boxed{z_0 = L_c^{-1} v_0}$$

$$z_i = L_i^{-T} \left(v_i - L_{i0}^T z_0 \right), \quad i = 1, \dots S.$$

Forward substitution

High performance computing with DSC

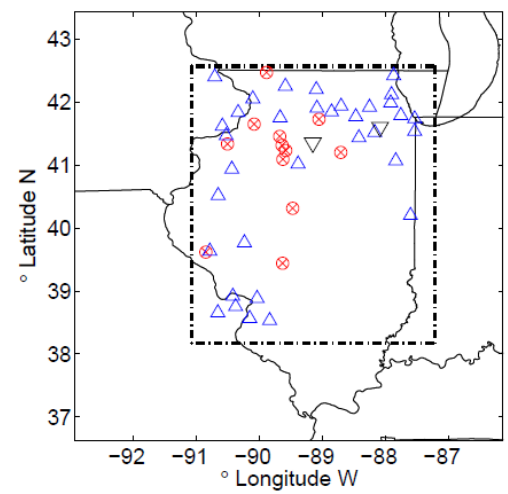
- Gondzio (OOPS) 6-stages 1 **billion** variables
- Zavala et.al., 2007 (in IPOPT)
- Our experiments (PIPS) – **strong** scaling is investigated
  - Building energy system
    - Almost linear scaling



on Fusion @ Argonne

- Unit commitment
  - Relaxation solved
  - Largest instance

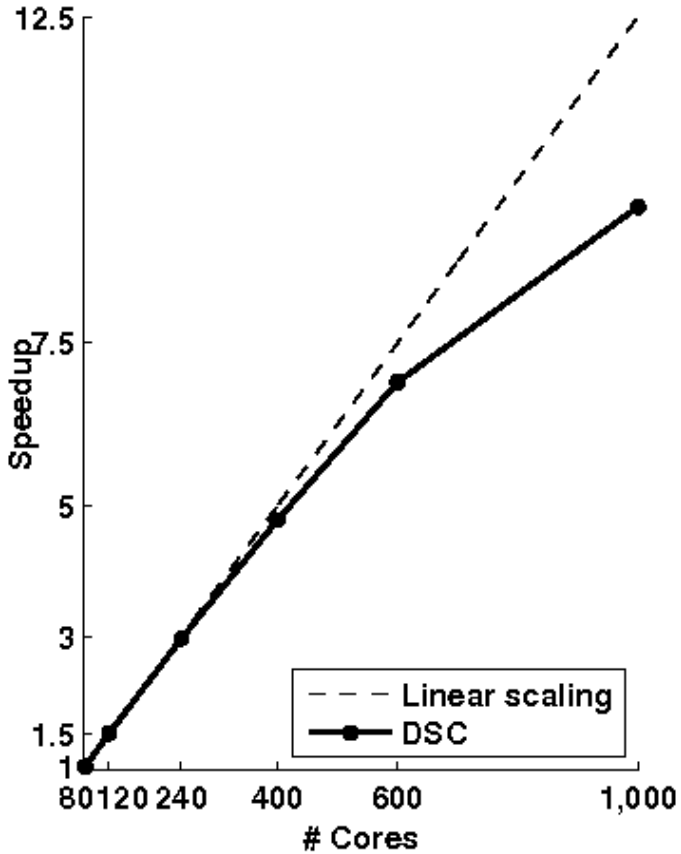
12 x



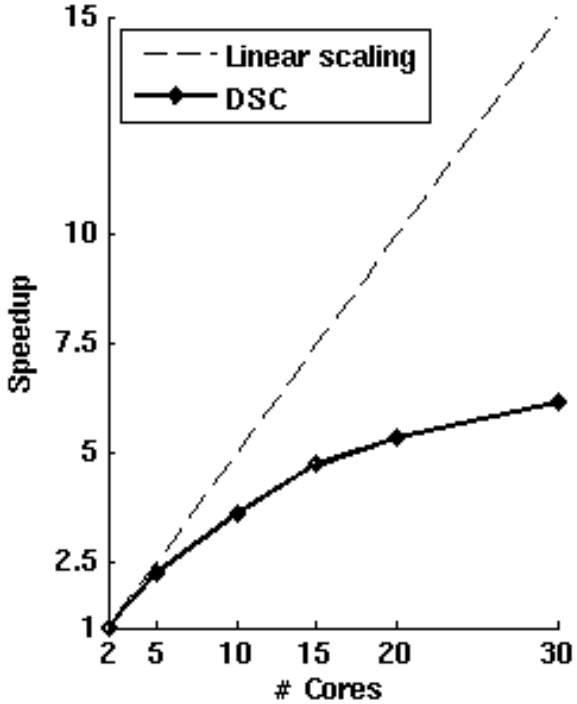
- 28.9 millions variables
- **1000 cores**

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)

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

$$L_M D_M L_M^T = M$$

(separate process)

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

$$v_i = D_i^{-1} w_i, \; i = 0, \ldots, N$$

$$z_0 = \mathrm{Krylov}(C, M, \tilde{r}_0)$$

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

The Stochastic Preconditioner

- The exact structure of C is

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

- 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

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

$$\underbrace{}_{p} \quad \underbrace{}_{r}$$

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

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

- 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 $KerA_0$.

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

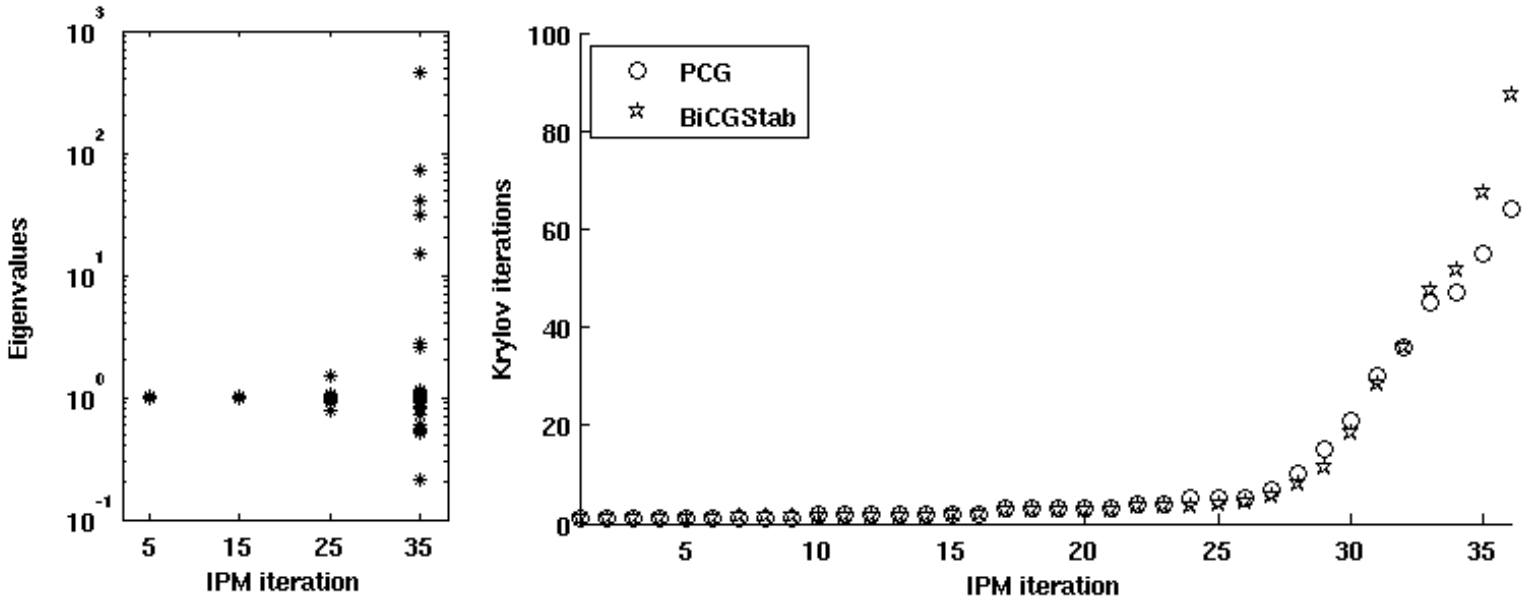
  - Does not compute the basis Z_0 for $KerA_0$. Instead,

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

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

Performance of the preconditioner

- Eigenvalues clustering & Krylov iterations



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

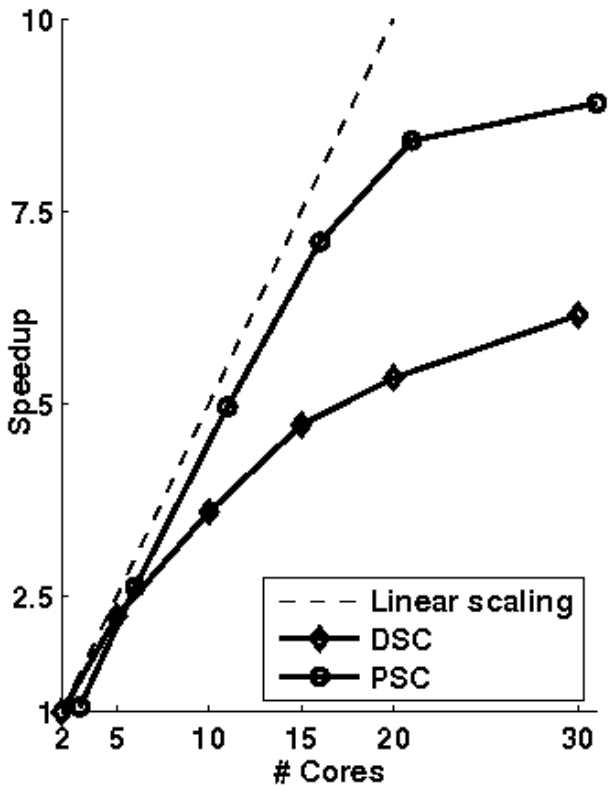
$$S_n \approx S_N \text{ and } \mathrm{S}_N \approx \mathbb{E}\left[S(\omega)\right], \text{ where}$$

$$S(\omega) = \left(Q_0 + D_0\right) + \left[A^T(\omega)\left(B(\omega)\left(Q(\omega) + D(\omega)\right)^{-1} B^T(\omega)\right)^{-1} A(\omega)\right]$$

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

- $$\begin{bmatrix} \tilde{Q} & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} L & 0 \\ AL^{-T} & \bar{L} \end{bmatrix} \begin{bmatrix} I & \\ & -I \end{bmatrix} \begin{bmatrix} L^T & L^{-1}A^T \\ 0 & \bar{L}^T \end{bmatrix}, \text{ where } LL^T = \tilde{Q},\ \bar{L}\bar{L}^T = A\tilde{Q}^{-1}A^T$$
- 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

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

- 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.

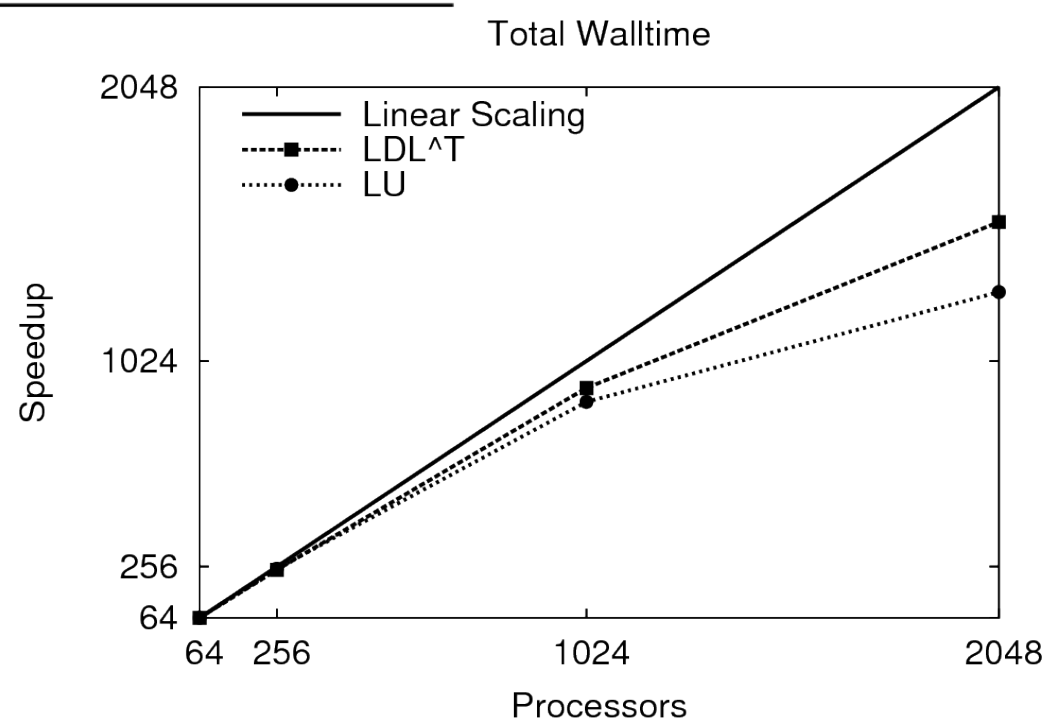
Large-scale performance

- 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.



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; scales well to 2000, port to BG in progress
  - tree nodes (scenarios) are distributed across processors
  - inter-process communication based on MPI
  - dynamic load balancing
- Mehrotra predictor-corrector IPM

Conclusions

- Forecast and accommodation of uncertainty of ambient conditions is crucial to the economic performance AND reliability of energy systems.

- To achieve the accuracy of forecast with uncertainty required by energy systems, we need high spatio-temporal resolution and high performance computing (~ 1 mil cores …)

- Stochastic programming is effective at providing optimal management decisions in real time, if it used on ~thousands of processors.

Future work

- Better uncertainty models for weather forecast (how do we eliminate the nagging alpha parameter and have a more solidly founded uncertainty approach).

- 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

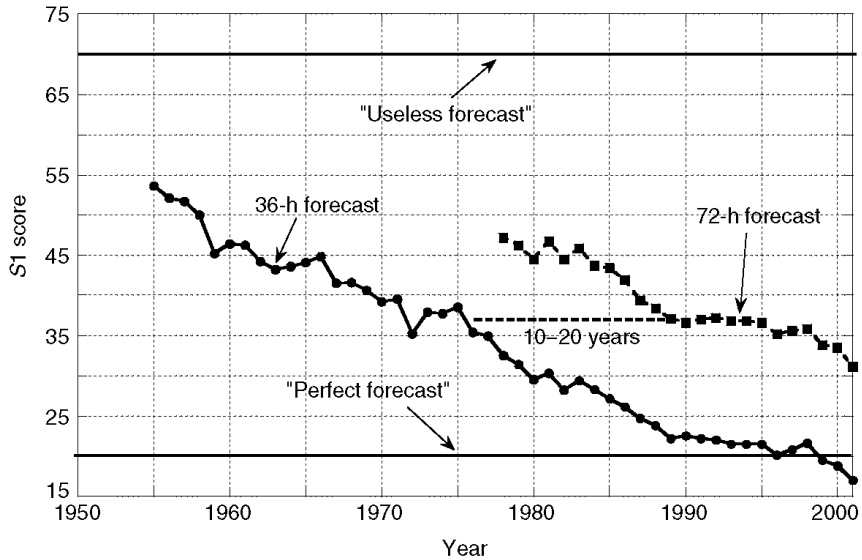
References

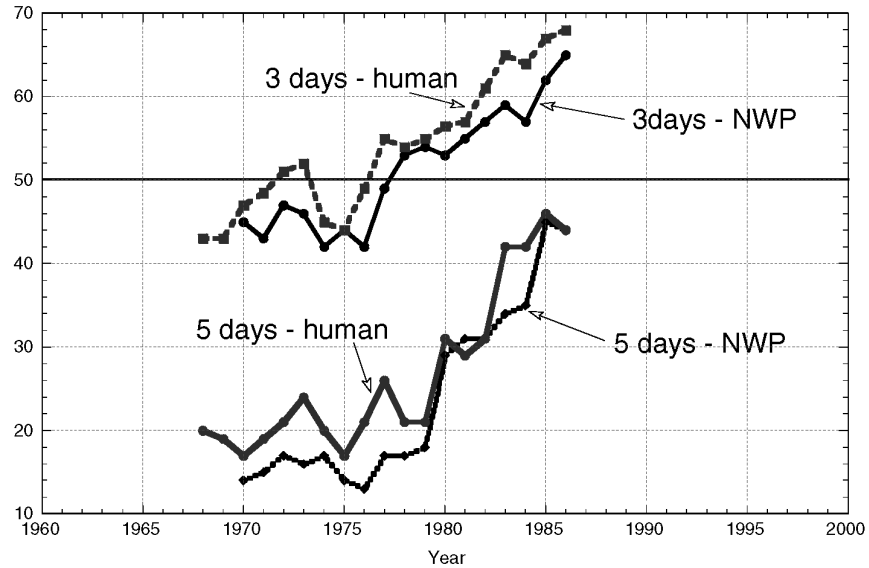
- Victor Zavala, Emil Constantinescu, Theodore Krause, and Mihai Anitescu." On-line economic optimization of energy systems using weather forecast information." *Journal of Process Control*, Volume 19, Issue 10, 2009, Pages 1725-1736
- Emil Constantinescu, Victor Zavala, Matthew Rocklin, Sangmin Lee, and Mihai Anitescu. A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment with Wind Power Generation. To appear in IEEE Transactions on Power Systems. Preprint ANL/MCS-P1681-1009.Electronic Version : DOI:10.1109/TPWRS.2010.2048133
- 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 Petra, and Mihai Anitescu "On the parallel solution of dense saddle-point linear systems arising in stochastic programming". Preprint ANL/MCS-P1798-1010.

Weather Forecast; State of the Art

- Problem: given current and historical information from satellite and met towers (and others), and physics of the atmosphere, estimate current state of the atmosphere and forecast the future state.
- State of the atmosphere: described by a partial differential equation (with several observables).
- Issues:
  - Incomplete (and, actually, very sparse) spatial and spatial frequency coverage. For example, wind data at 100 meters from surface is virtually inexistent. That is one of the issues for which the model is most useful for; to correlate past observations with unobservable spatial and spectral states of the atmosphere.
  - A complex, but *reasonably* well understood physical model. For operational WRF forecasts to run the US at 1 km^2 in for 24 hrs with 30 “particles” in 1hr we need 400K processors. For many decisions lower resolution is sufficient.
- Incidentally, very useful in management of power grid, and energy apps if high renewable penetration is considered.

Weather Forecast: Progress due to Increased Model Resolution of Numerical Weather Prediction (NWP)

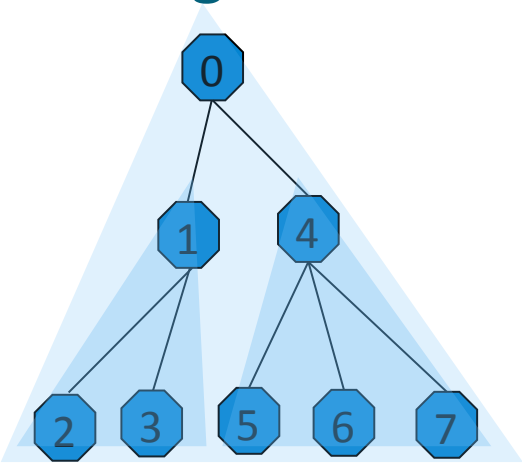




Source: (Kalnay 2003)

- S1 skill factor (SB; error in the gradient of pressure at 500hPa) and SAC skill factor (LB; correlation in latitudinal sea pressure variability) have improved markedly due to model.
- Human is human + NWP
- Conclusion: model matters a lot in state estimation of weather forecast.

Multi-stage SAA SP Problems – Scenario formulation



- Depth-first traversal of the scenario tree
- **Nested** half-arrow shaped Jacobian
- Block separable obj. func.

$$Min \frac{1}{2}\sum_{i=0}^{7} x_i^T Q_i x_i + \sum_{i=0}^{7} c_i^T x_i$$

s.t.

$$\begin{array}{lllllllll}
A_0 x_0 & & & & & & & = & b_0 \\
A_1 x_0 + & B_1 x_1 & & & & & & = & b_1 \\
 & A_2 x_1 + & B_2 x_2 & & & & & = & b_2 \\
 & A_3 x_1 + & & B_3 x_3 & & & & = & b_3 \\
A_4 x_0 + & & & & B_4 x_4 & & & = & b_4 \\
 & & & & A_5 x_4 + & B_5 x_5 & & = & b_5 \\
 & & & & A_6 x_4 + & & B_6 x_6 & = & b_6 \\
 & & & & A_7 x_4 + & & & B_7 x_7 & = b_7 \\
x_0 \geq 0 & x_1 \geq 0 & x_2 \geq 0 & x_3 \geq 0 & x_4 \geq 0 & x_5 \geq 0 & x_6 \geq 0 & x_7 \geq 0 &
\end{array}$$

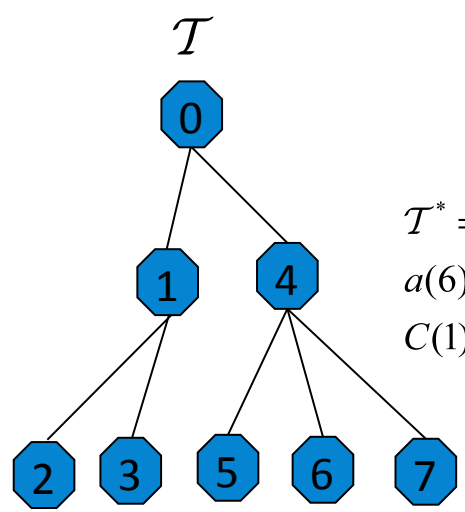
Tree Linear Algebra – Data, Operations & Linear Systems

- Data
  - Tree vector: b, c, x, etc
  - Tree symmetric matrix: Q
  - Tree general matrix: A

$$\text{Min } \frac{1}{2}x^T Q x + c^T x$$
$$\text{subj. to. } Ax = b$$
$$x \geq 0$$

- Operations

$$(u \odot v)_n = u_n \odot v_n, \forall n \in \mathcal{T}$$
$$(Qx)_n = Q_n x_n, \forall n \in \mathcal{T}$$
$$(Ax)_n = A_{a(n)} x_{a(n)} + B_n x_n, \forall n \in \mathcal{T}^*$$
$$(A^T y)_n = B_n^T y_n + \sum_{c \in C(n)} A_c^T y_c$$

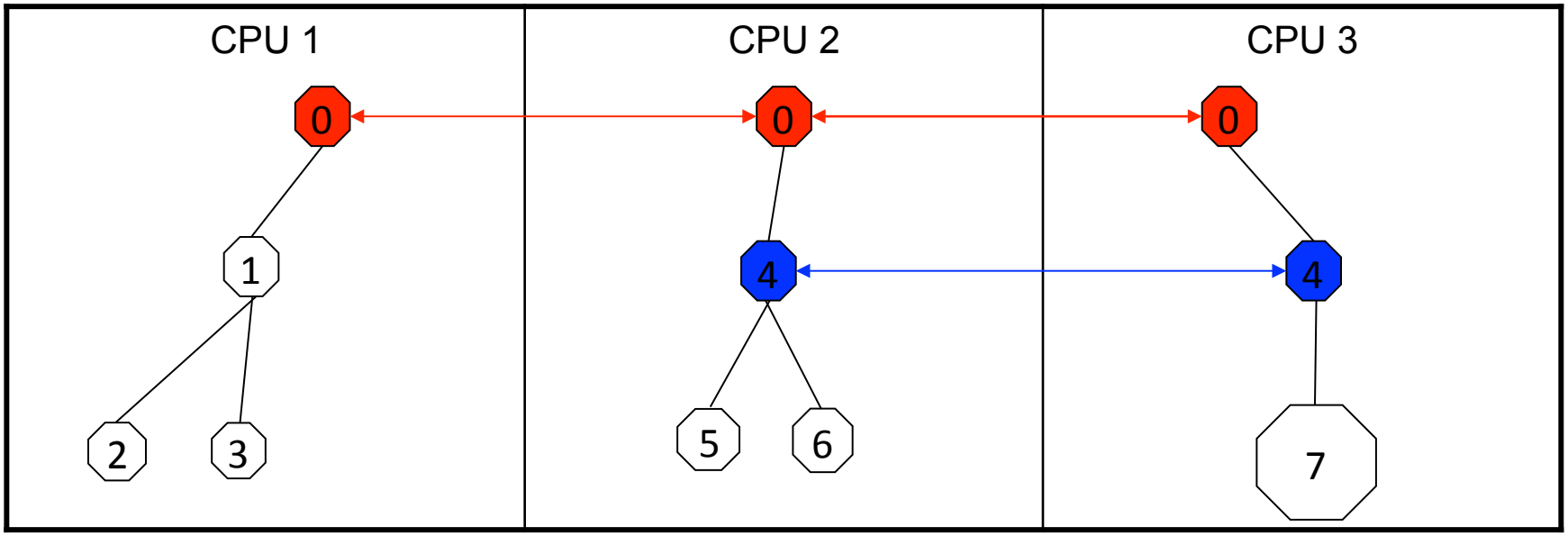


$$\mathcal{T}^* = \mathcal{T} - \{0\}$$
$$a(6) = 4 \text{ (ancestor map)}$$
$$C(1) = \{2,3\} \text{ (children map)}$$

- Linear systems: for each non-leaf node a two-stage problem is solved via Schur complement methods as previously described.

Parallelization – Tree Distribution

- The tree is distributed across processes.
- Example: 3 processes



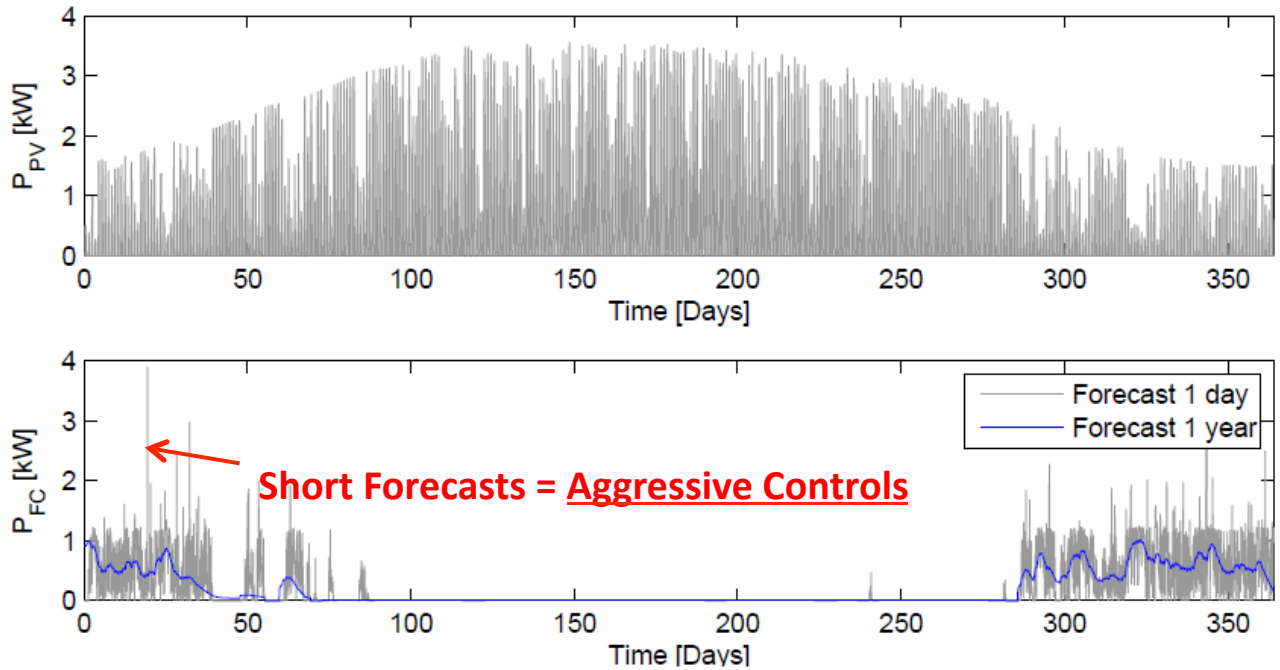
- Dynamic load balancing of the tree
  - Number partitioning problem --> graph partitioning --> METIS

Extra Slides

- HMM is foundation
- I need to deal with the minus infinity window, so I approximate by means of a background. Of course now the background has to be estimated.
- Now you need to estimate covariance background.
- Particle filters are one of the most common techniques. We need derivatives if we want to do a good job. They do not do operationally Kalman or other particle filters.
- 3DVar is more common, needs an uncertainty on top of it. We are emulating the operation method.
- If 3Dvar alone. Satelite data, and reason data.

Hybrid Photovoltaic-H_2 System

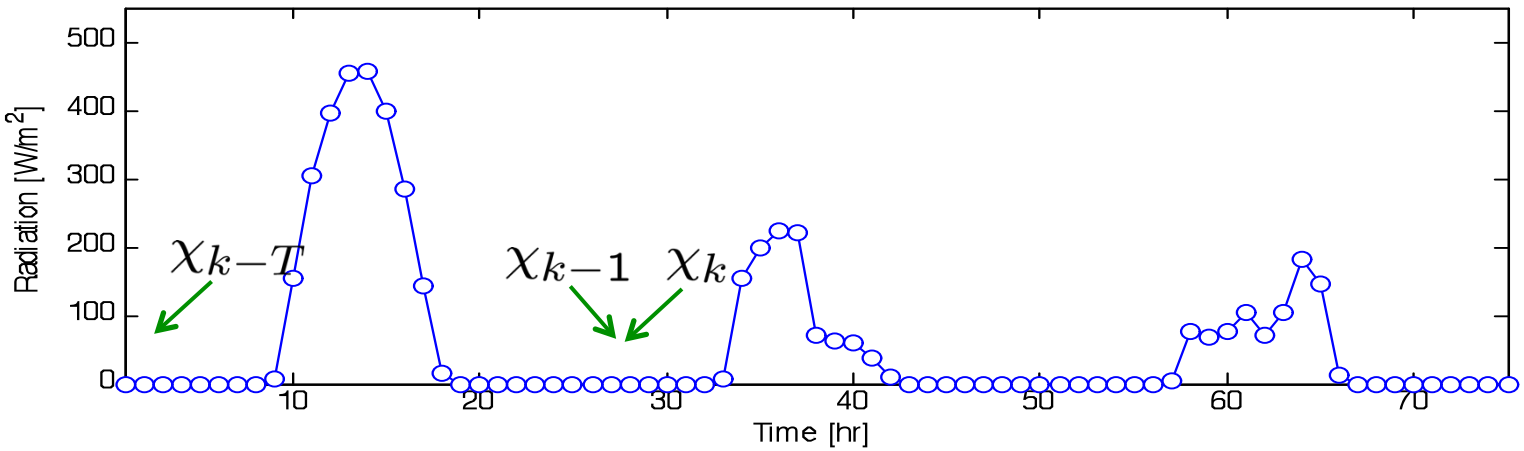
Profiles of Fuel Cell Power



Possible data-based model of ambient conditions

Quantifying Model Uncertainty (Data-Based (Time-Series) vs. Physics-Based)

Solar Radiation Forecast with Gaussian Process (GP) Modeling *Zavala & A, 2008*



1. Input-Output Data Sets: $\mathbf{Y}_j := \chi_k \qquad \mathbf{X}_j := [\chi_{k-1},\ \chi_{k-T}]$

2. Covariance Structure : $\mathbf{V}(\mathbf{X}_j, \mathbf{X}_i, \eta) := \eta_0 + \eta_1 \cdot \exp\left(-\frac{1}{\eta_2}\|\mathbf{X}_j - \mathbf{X}_i\|^2\right)$

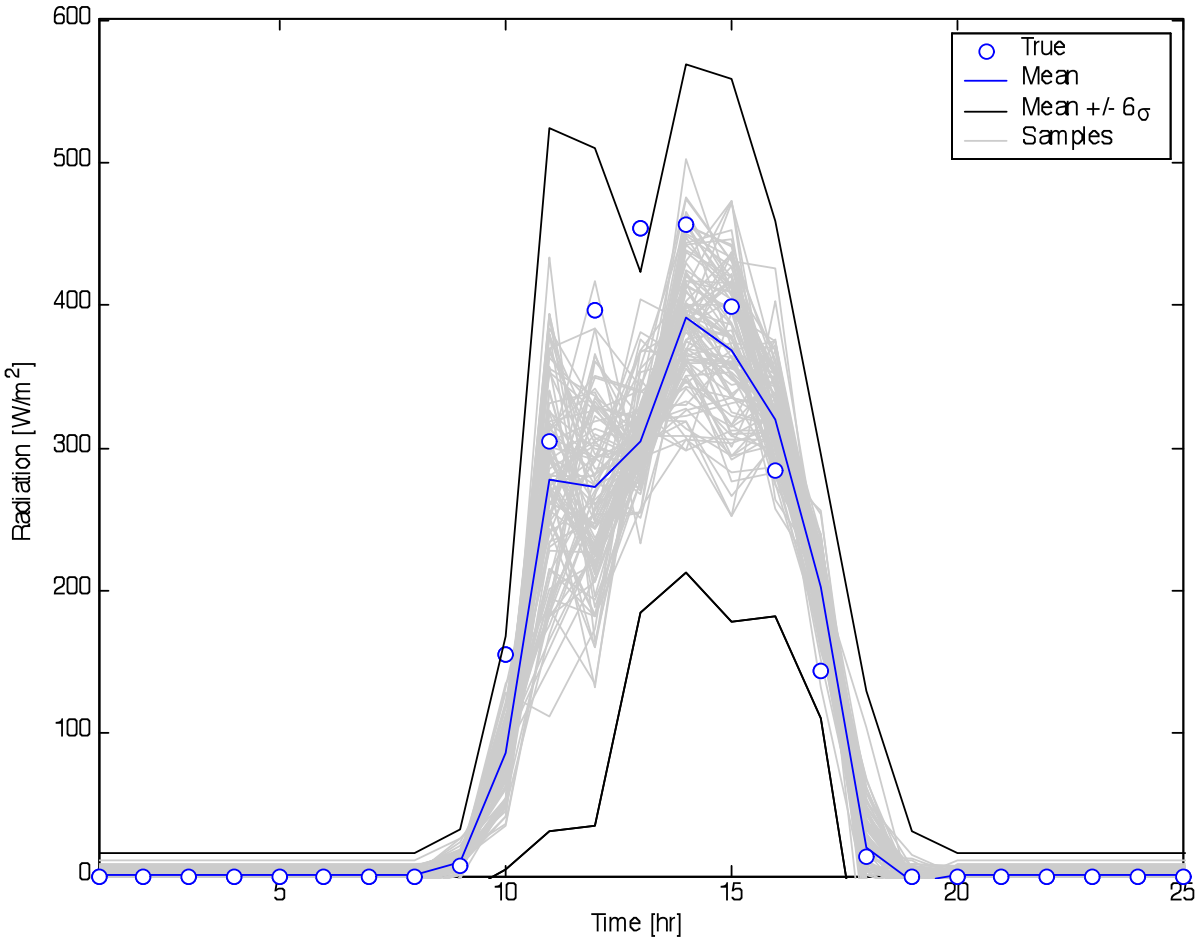
3. Apply Maximum Likelihood: $\log p(\mathbf{Y}|\eta) = -\frac{1}{2}\mathbf{Y}\mathbf{V}^{-1}(\mathbf{X},\mathbf{X},\eta)\mathbf{Y} - \frac{1}{2}\log\det(\mathbf{V}(\mathbf{X},\mathbf{X},\eta))$

4. Posterior Distribution: $\mathbf{Y}^P = \mathbf{V}(\mathbf{X}^P, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X},\mathbf{X},\eta^*)\mathbf{Y}$ **Forecast Mean**

$\mathbf{V}^P = \mathbf{V}(\mathbf{X}^P, \mathbf{X}^P, \eta^*) - \mathbf{V}(\mathbf{X}^P, \mathbf{X}, \eta^*)\mathbf{V}^{-1}(\mathbf{X},\mathbf{X},\eta^*)\mathbf{V}(\mathbf{X}, \mathbf{X}^P, \eta^*)$ **Covariance**

Data-based statistical model of ambient conditions

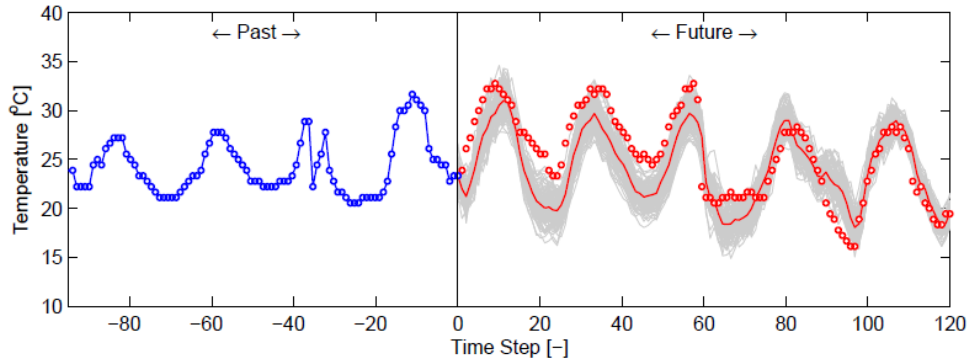
One-Day Ahead Forecast and Samples from Posterior Distribution



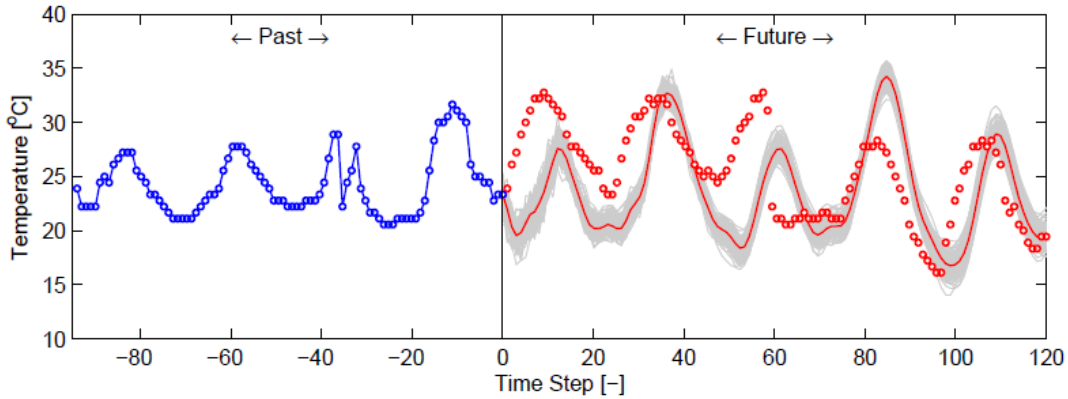
Covariance Structure *Sort of* Makes Physical Sense, Wide Uncertainty Bounds

Data-based statistical model of ambient conditions

Ambient Temperature Forecast with GP Modeling *Zavala, A, et.al. 2009*



One Hour Ahead



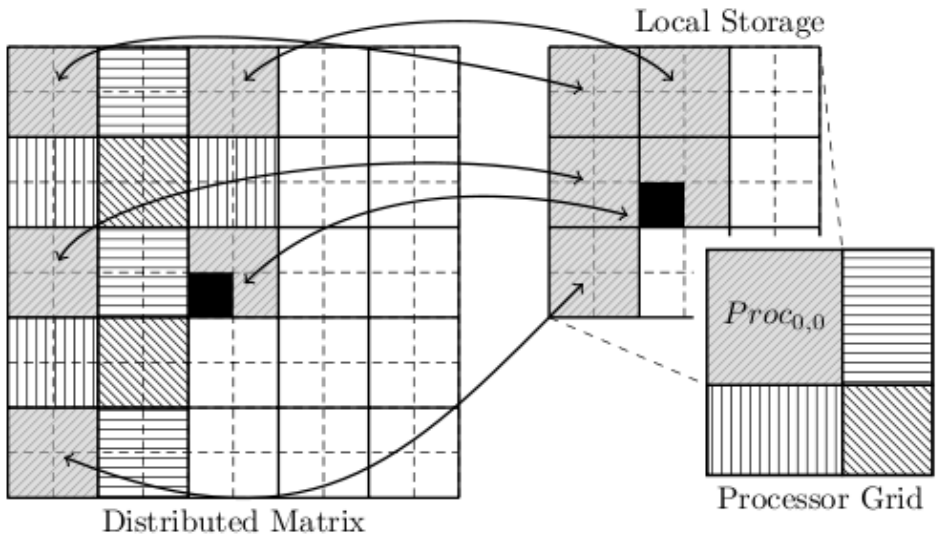
5 Days Ahead

Time-Series Cannot Capture Physical Effects (Spatial), Inconsistent Uncertainty Bounds

GP (and all data-oriented approaches) Provide Accurate <u>Interpolations</u> but Poor Extrapolations (e.g. Geostatistics)

ScaLapack

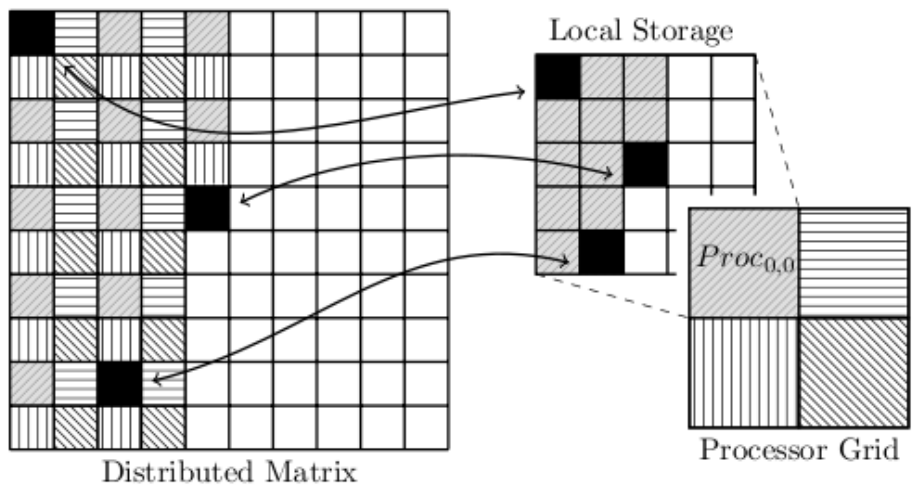
- Classical block distribution of the matrix



- Blocked "down-looking" Cholesky - algorithmic blocks
  - **Size of algorithmic block = size of distribution block!**
- For cache-performance - **large** algorithmic blocks
- For good load balancing - **small** distribution blocks
- Must trade off cache-performance for load balancing
- Communication: basic MPI calls
- Inflexible in working with sub-blocks

Elemental

- Unconventional “elemental” distribution: blocks of size 1.



- Size of algorithmic block $\neq$ size of distribution block
- Both cache-performance (large alg. blocks) and load balancing (distrib. blocks of size 1)
- Communication
  - More sophisticated MPI calls
  - Overhead O(log(sqrt(p))), p is the number of processors.
- Sub-blocks friendly
- Better performance in a hybrid approach, MPI+SMP, than ScaLapack

Closed loop simulation using WRF

- 24-hour simulation window – restart from assimilated solution (NARR) every 12 hours

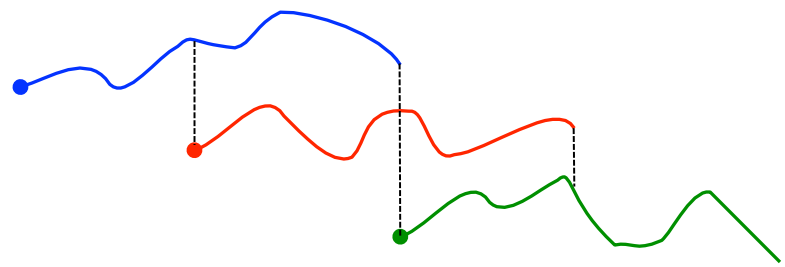
- Restart ensemble with adjusted spread based on error estimates:

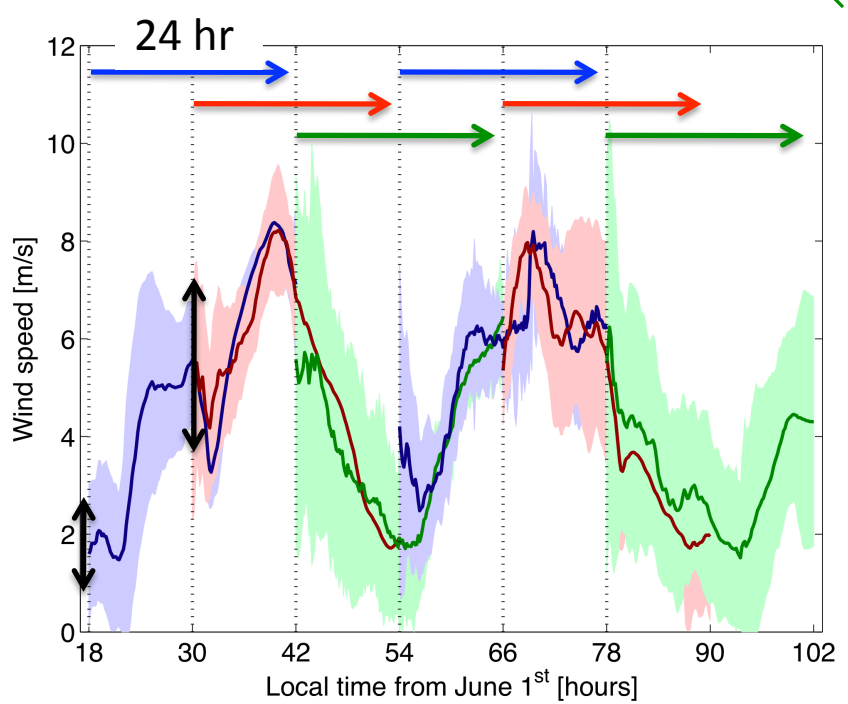
$$x_i \leftarrow \overline{x} + \gamma\left(x_i - \overline{x}\right)\ ,\ \ i \in [1, N_S]$$

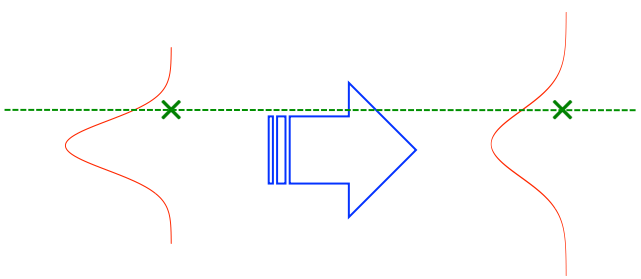
$$\gamma = max\left(1, min\left(\gamma_\sigma, 4\right)\right)$$

$$\gamma_\sigma = \mathrm{mean}_{U,V,T|k=1\ldots5}\left(\frac{|x_{\mathrm{NARR}} - \overline{x}|}{\sigma}\right)$$

- Error is underestimated: increase uncertainty







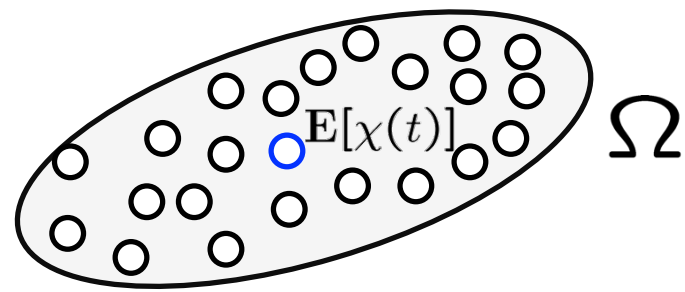
Stochastic Dynamic Optimization

Solution Strategies

- **Dynamic Programming, Taylor Series:** Handling Constraints and Nonlinearity Cumbersome
- **Polynomial Chaos:** Dense Optimization, Multivariable Quadrature
- **Sample Average Approximation (SAA):** Sparse Optimization, Constraints, General Framework

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

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



Nonlinear Programming: Exploit Fine and Coarse Structures at Linear Algebra Level

$$\begin{aligned} \min_{\mathbf{u}} \quad & \frac{1}{S}\sum_{k=1}^{S} \varphi(z_k, y_k, \mathbf{u}, \chi_k) \\ \text{s.t.} \quad & \mathbf{c}(z_k, y_k, \mathbf{u}, \chi_k) = 0 \\ & \mathbf{h}(z_k, y_k, \mathbf{u}, \chi_k) \leq 0 \\ & k = 1, ..., S \end{aligned}$$

$$\begin{bmatrix} \mathbf{K}_1 & & & & Q_1 \\ & \mathbf{K}_2 & & & Q_2 \\ & & \ddots & & \vdots \\ & & & \mathbf{K}_S & Q_S \\ Q_1^T & Q_2^T & \dots & Q_S^T & D_\mathbf{u} \end{bmatrix} \begin{bmatrix} \Delta s_1 \\ \Delta s_2 \\ \vdots \\ \Delta s_S \\ \Delta \mathbf{u} \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_S \\ r_\mathbf{u} \end{bmatrix}$$

MOTIVATION

Optimal management under uncertainty of energy systems

- Applications @Argonne – Anitescu, Constantinescu, Zavala
  - Stochastic Unit Commitment with Wind Power Generation
  - Energy management of Co-generation
  - Economic Optimization of a Building Energy System

- Sources of large dimensional uncertainty in complex energy systems
  - Weather
  - Consumer Demand
  - Market prices

Economic Optimization of a Building Energy System

- **Proactive** management - **temperature forecasting** & **electricity prices**
- Minimize daily energy costs (SAA)

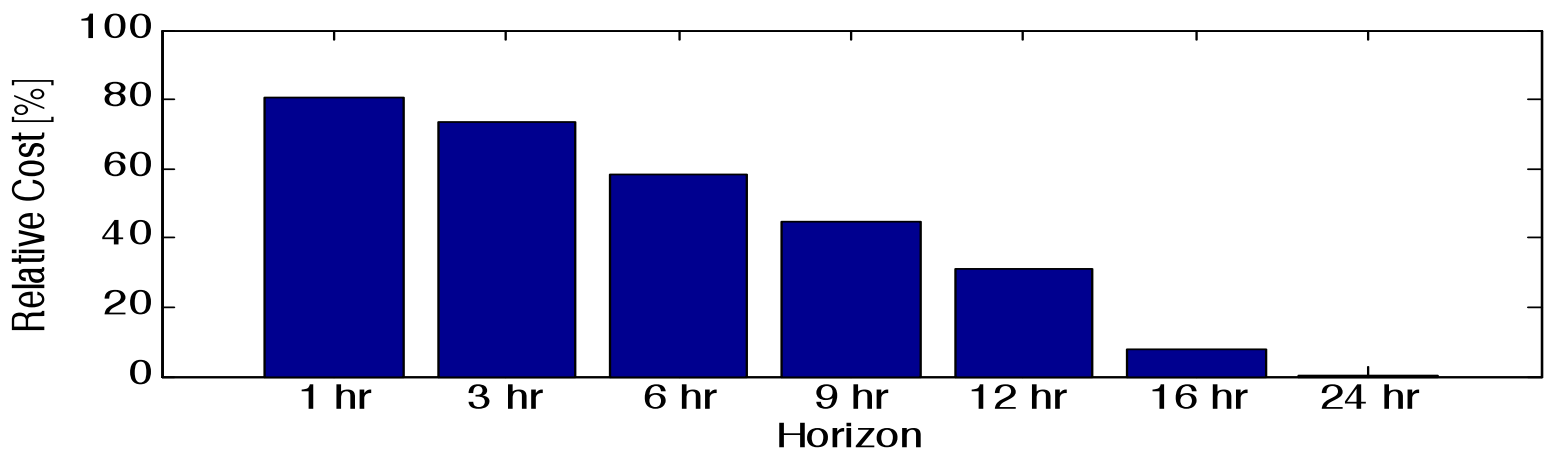
$$\min \quad \text{COST} = \frac{1}{S}\sum_{j=1}^{S}\int_{t}^{t+T}\left[C_{elec}(\tau)\phi_c^{elec}(\tau)+C_{elec}(\tau)\phi_h^{elec}(\tau)+C_{gas}\phi_h^{gas}(\tau)\right]d\tau$$

$$\text{s.t.} \quad C_I\frac{\partial T_I^j}{\partial \tau}=\phi_h^{gas}(\tau)+\phi_h^{elec}(\tau)-\phi_c^{elec}(\tau)-S\alpha'\left(T_I^j(\tau)-T_W^j(\tau,0)\right),\ T_I^j(0)=T_I^0$$

$$\frac{\partial T_W^j}{\partial \tau}=\beta\frac{\partial^2 T_W^j}{\partial x^2},\ \alpha'\left(T_I^j(\tau)-T_W^j(\tau,0)\right)+k\left.\frac{\partial T_W^j}{\partial \tau}\right|_{(\tau,0)}=0,\ \alpha''\left(T_W^j(\tau,L)-T_A^j(\tau)\right)+k\left.\frac{\partial T_W^j}{\partial \tau}\right|_{(\tau,L)}=0$$

$$T_W^j(0,x)=T_W^0(x),\tau\in[t,t+T]$$

$$T_I^{min}\le T_I^j(\tau)\le T_I^{max},\ j\in\{1,\ldots,S\}$$



Zavala & al. 2009

Creating the weather uncertainty

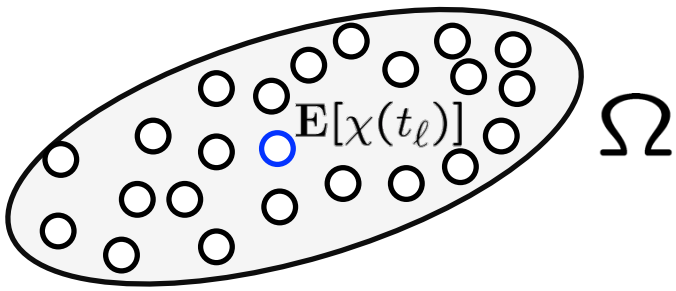
Ensemble Forecast Approach – Use WRF (Weather Research Forecast) as Black-Box

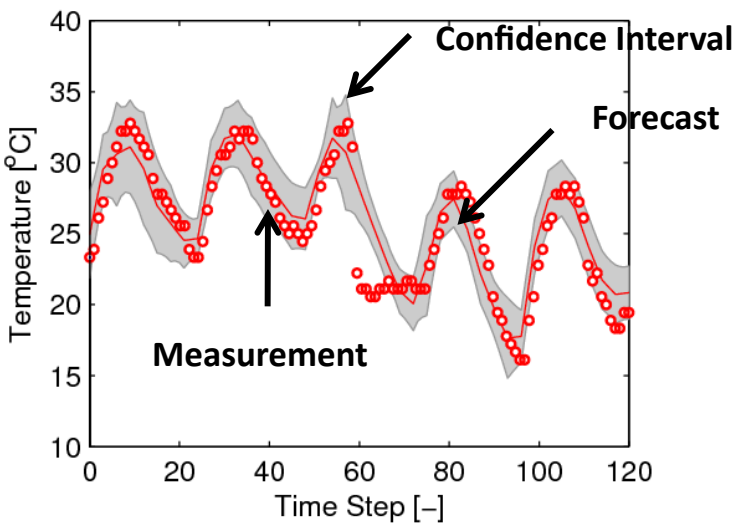
Sample Prior and Propagate Samples of Posterior Through Model

$$\mathbf{Y}_{[i,j]} := \chi_i(t_{\ell+j}) = \underbrace{\mathcal{M}(\mathcal{M}(\ldots\mathcal{M}}_{j \text{ times}}(\chi_i(t_\ell))))$$

$$\mathbf{E}[\mathbf{Y}] \approx \bar{\mathbf{Y}} := \frac{1}{NS}\sum_{i=1}^{NS} \mathbf{Y}_{[i,:]}$$

$$\mathbf{V} \approx \frac{1}{NS-1}\sum_{i=1}^{NS} (\mathbf{Y}_{[i,:]} - \bar{\mathbf{Y}})(\mathbf{Y}_{[i,:]} - \bar{\mathbf{Y}})^T$$





Hours (August 1st-5th)

Validation Results, Pittsburgh Area 2006

5 Day Forecast and +/- 3s Intervals

Building applications

Stochastic Unit Commitment with Wind Power (SAA)

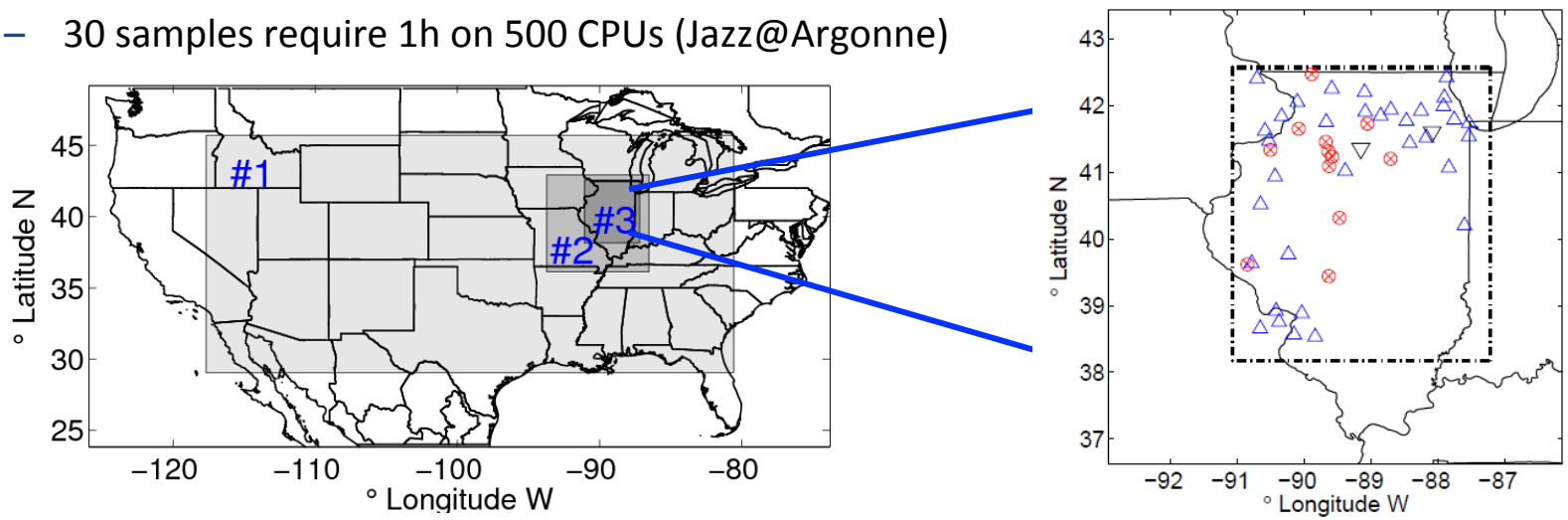
$$\min \quad \text{COST} = \frac{1}{N_s}\sum_{s\in\mathcal{S}}\left(\sum_{j\in\mathcal{N}}\sum_{k\in\mathcal{T}} c^p_{sjk} + c^u_{jk} + c^d_{jk}\right)$$

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

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

ramping constr., min. up/down constr.

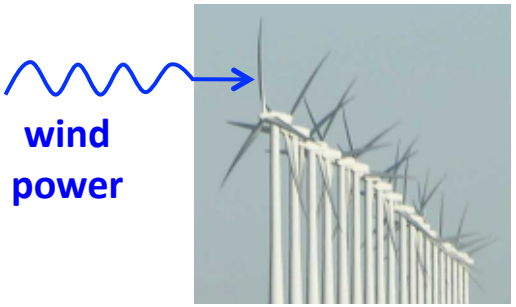
- Wind Forecast – WRF(Weather Research and Forecasting) Model
  - Real-time grid-nested 24h simulation
  - 30 samples require 1h on 500 CPUs (Jazz@Argonne)

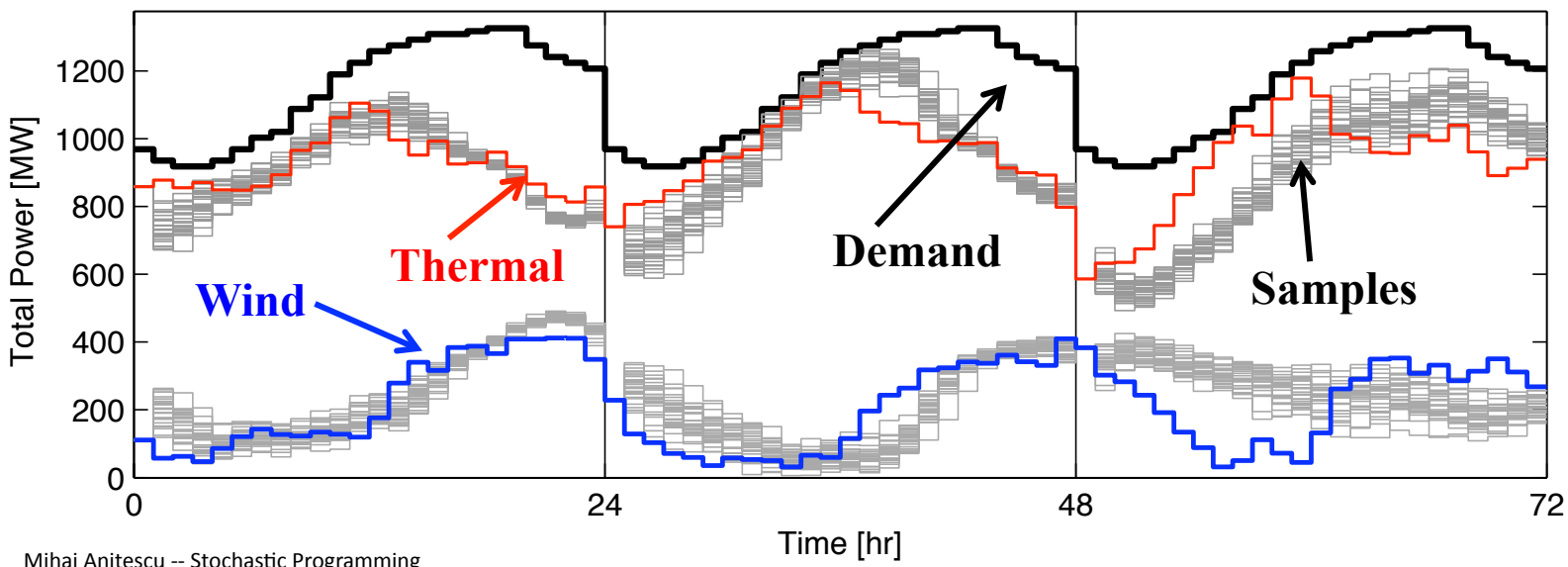


Zavala & al 2010.

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.



- 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%.