The Promise of Supercomputing for Optimal Management of Energy Systems

Mihai Anitescu

Mathematics and Computer Science Division
Argonne National Laboratory
anitescu@mcs.anl.gov

Emil Constantinescu (Wilkinson), Victor Zavala (ANL Director’s Postdoc)
Matt Rocklin, Sangmin Lee, Theodore Krause,
Supercomputing 09
November, 2009
Challenges of the Next-Generation Power Grid

- Major Adoption of Renewable Resources (20-30%)
- Highly Decentralized Generation and Demand
Electricity Supply - Billions $/Yr Market

ON/OFF Power Levels, day ahead.

System Operator

Spot Market, 1hr

Transmission/Distribution

Weather

Demand

Generation

Consumer

Anticipating Weather (In 1hr, 1 day, at 1km) Minimizes Reserve Cost
Objective 1: Uncertainty quantification in weather forecast: A test case with real data

- Importance in Energy Systems Management
  - Weather influences both energy supply (wind/solar/thermal) and demand
  - Accurate weather/climate forecast leads to an efficient proactive resource management (reduces needed reserves and cost, though not consumption directly).

- Weather forecasting at the renewable energy (RE) system scale (1km) is a grand challenge:
  - Very large-system, difficult to simulate
  - Chaotic,
  - Incomplete/missing physics
  - Unknown initial conditions, uncertain forcings
  - Lack of information makes deterministic prediction impossible, we must quantify and compute uncertainty in weather forecast (i.e. its probability distribution).

- Objective 1: Is there hope of predicting the probability distribution of weather at RE scale operationally? (24 hour ahead in 1 hour at 1 km)?
Objective 2: What is the economic impact of using weather (and demand/pricing) forecast in energy management

- The impact strongly depends on the decision system and of the way it accounts for risk.
- We posit the decision problem under uncertainty as a stochastic programming problem.
- Note that we do not expect to reduce consumption, (that is design, not management), but we expect to be able to reduce cost, and implicitly
  - Reduce the risk of not meeting demand
  - Reduce the peak requirements on power grid and this increase resilience by optimally answering to the appropriate incentives (such as electricity pricing).
- Question 2: For realistic Independent System Operator Problems, Building Systems, and Photovoltaic Systems, what are the expected cost reductions?
Forecast and Uncertainty
Uncertainty in dynamical systems: 1. Data

- Assume a time-discretized process with imperfect initial state and forcing information and noisy measurements.

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

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

(1)

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

(2)

where

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

and

$$V_k \approx N(\bar{x}_k, R_k^{-1}).$$

We want find $D(x_0^{in}, \cdots, x_k^{in})$'s mean and variance.
Uncertainty in dynamical systems: 2. the posterior.

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

\[
P(x_k^{in}, x_{k-1}^{in}, \ldots, x_0^{in} | z_0^{obs}, z_1^{obs}, z_2^{obs}, \ldots, z_k^{obs}) = C_k \tilde{C}_k \frac{\exp\left(-\frac{1}{2} f(X^{in}, Z^{obs})\right)}{P(Z^{obs})}.
\]

\[
f(X^{in}, Z^{obs}) = \sum_{i=0}^{k} (x_i^{in} - \bar{x}_i - \bar{y}(t_{i-1}, x_{i-1}^{in}))^T Q_i^{-1}(x_i^{in} - \bar{x}_i - \bar{y}(t_{i-1}, x_{i-1}^{in}))
+ \sum_{i=n}^{k} (z_i^{obs} - h_i(\bar{y}^\perp(t_i, x_i^{in})))^T R_i^{-1}(z_i^{obs} - h_i(\bar{y}^\perp(t_i, x_i^{in}))).
\]

- 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.
Step 1: Moving Horizon Best State Estimation

\[ y(t) \]

Measured Outputs

\[ y_{\ell-N+1} \quad y_{\ell-N} \quad y_{\ell-1} \quad y_{\ell} \quad y_{\ell+1} \]

\[ z(t) \]

Unmeasured States

\[ z(0) \quad z_0 \]

Current Time

\[ t_{\ell-N} \quad t_{\ell-N+1} \quad t_{\ell-1} \quad t_{\ell} \quad t_{\ell+1} \]

\[
\min_{p(t), z_0} \sum (y(t_k) - y_{\ell-k+N})^T V_{y}^{-1} (y(t_k) - y_{\ell-k+N})
\]

\[
\frac{dz}{dt} = f(z(t), p(t), u(t)) \\
y(t) = g(z(t), p(t), u(t)) \\
z(0) = z_0 \quad \text{Uncertain}
\]

WRF Model

\[
\min_{p(t), z_0} \sum (y(t_k) - y_{\ell-k+N+1})^T V_{y}^{-1} (y(t_k) - y_{\ell-k+N+1})
\]

\[
\frac{dz}{dt} = f(z(t), p(t), u(t)) \\
y(t) = g(z(t), p(t), u(t)) \\
z(0) = z_0
\]

Uncertainty in Current State \( x_{\ell} \)

Needed To Quantify Future Forecast

\[ \Pi_{\ell} \rightarrow 0 \]
Step 2: Estimate the prior covariance matrix.

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

\[
V_{ik} \approx dd^T = \sum_j d_{ij} d_{kj}^T = \epsilon_i \cdot \epsilon_k = \begin{bmatrix}
\epsilon_0 \cdot \epsilon_0 & \epsilon_1 \cdot \epsilon_0 & \cdots & \epsilon_n \cdot \epsilon_0 \\
\epsilon_0 \cdot \epsilon_1 & \epsilon_1 \cdot \epsilon_1 & \cdots & \epsilon_n \cdot \epsilon_1 \\
\vdots & \vdots & \ddots & \vdots \\
\epsilon_0 \cdot \epsilon_n & \epsilon_1 \cdot \epsilon_n & \cdots & \epsilon_n \cdot \epsilon_n
\end{bmatrix}, \quad C_{ik} = \frac{\epsilon_i \cdot \epsilon_k}{|\epsilon_i| |\epsilon_k|}.
\]

- “Guess” the diagonal of the variance matrix
Step 2 b: Fit to a Gaussian Process.

Covariance Matrix is Huge and low rank (10^6 x 10^6) But ...
- Spatial Correlations Decay Exponentially *Constantinescu, et.al., 2007*
- Covariance Can be Approximated Using Gaussian Kernels *Zavala, Constantinescu & A, 2009*

\[
\Pi_{i,j} = \exp\left(-\frac{(x_j-x_i)^2+(y_j-y_i)^2}{L_H^2} - \frac{(z_j-z_i)^2}{L_V^2}\right)
\]
Ensemble Forecast Approach

Ensemble Forecast Approach – Use WRF as Black-Box

Sample Prior and Propagate Samples of Posterior Through Model

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

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

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

Validation Results, Pittsburgh Area 2006
5 Day Forecast and +/- 3s Intervals
Building applications

Mathematics and Computer Science

Hour (August 1st-5th)
Uncertainty quantification and forecast using WRF

- **WRF model:** \( x^{t_F} = \mathcal{M}_{t_0 \rightarrow t_F} (x^{t_0}) \)

- **Uncertainties in the initial conditions:** \( x^{t_0}_i = x_{NARR} + L \xi_i \)
  \( \xi_i \sim \mathcal{N}(0, I), \quad i \in [1, N_S] \), \( LL^T = P \), \( C_{ij} = \frac{P_{ij}}{\sqrt{P_{ii}P_{jj}}} \)

- **Evolution of uncertainties through WRF:**
  \( x^{t_F}_i = \mathcal{M}_{t_0 \rightarrow t_F} (x^{t_0}_i) + \eta_i(t) \)
  \( x^{t_0}_i \sim \mathcal{N}(x_{NARR}, P^{t_0}) \), \( \eta_i \sim \mathcal{N}(0, Q) \), \( i \in [1, N_S] \)

- **Uncertainty at the final time:** \( x^{t_F}_i \sim \mathcal{N}(\bar{x}, S^2) \)
  \( \bar{x} = \frac{1}{N_S} \sum_{i=1}^{N_S} x_i \)
  \( S^2 = \frac{1}{N_S - 1} \sum_{i=1}^{N_S} (x_i - \bar{x})(x_i - \bar{x})^T \approx \mathcal{N}(\bar{x}, S^2) \)
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:

\[ x_i \leftarrow \bar{x} + \gamma (x_i - \bar{x}), \; i \in [1, N_S] \]

\[ \gamma = \max (1, \min (\gamma_\sigma, 4)) \]

\[ \gamma_\sigma = \frac{\left| x_{\text{NARR}} - \bar{x} \right|}{\sigma} \]

- Error is underestimated: increase uncertainty
Implementation and Estimation of Necessary Computational Resources
Ensemble forecast and uncertainty quantification with WRF on Jazz

- **Computational domain setup**

  - **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

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

<table>
<thead>
<tr>
<th></th>
<th>Grid</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>#1 - 32 km²</td>
<td>130 × 60</td>
</tr>
<tr>
<td></td>
<td>#2 - 6 km²</td>
<td>126 × 121</td>
</tr>
<tr>
<td></td>
<td>#3 - 2 km²</td>
<td>202 × 232</td>
</tr>
<tr>
<td>Illinois</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Table:

<table>
<thead>
<tr>
<th>CPU</th>
<th>Wallclock [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>157</td>
</tr>
<tr>
<td>8</td>
<td>95</td>
</tr>
<tr>
<td>16</td>
<td>65</td>
</tr>
<tr>
<td>32</td>
<td>45</td>
</tr>
</tbody>
</table>
Validation of the results
Wind power measurements and windmill locations

- Wind turbine locations and weather stations locations in Illinois:
- Real wind and temperature measurements for hindcast (June 2006) ~every 20 min.
- Real wind power measurements for Chicago and Peru, IL:
Wind power is difficult to predict

- Wind power measurements for Chicago

- Deterministic prediction; a new one started every 12 hours

- Forecast with uncertainty
Validation of the wind/temperature forecast and uncertainty

- Wind/temperature validation of uncertainty estimates with real measurements

- Temporal [trends] and spatial [similar outcomes] correlations provide additional info
Optimization under Uncertainty by Stochastic Programming
Benefits: Accommodate Forecasts, Constraint Handling, Financial Objectives, Complex Models

**Deterministic**
\[
\min_{u(t)} \int_{t_\ell}^{t_\ell+N} \varphi(z(t), y(t), u(t), E[\chi(t)]) \, dt
\]
\[
\frac{dz}{dt} = f(z(t), y(t), u(t), E[\chi(t)])
\]
\[
0 = g(z(t), y(t), u(t), E[\chi(t)])
\]
\[
0 \geq h(z(t), y(t), u(t), E[\chi(t)])
\]
\[
z(0) = x_\ell
\]

**Stochastic**
\[
\min_{u(t)} \mathbb{E}_{\chi(t) \in \Omega} \left[ \int_{t_\ell}^{t_\ell+N} \varphi(z(t), y(t), u(t), \chi(t)) \, dt \right]
\]
\[
\frac{dz}{dt} = f(z(t), y(t), u(t), \chi(t))
\]
\[
0 = g(z(t), y(t), u(t), \chi(t))
\]
\[
0 \geq h(z(t), y(t), u(t), \chi(t))
\]
\[
z(0) = x_\ell
\]

Complexity (Solution Time)
- 1,000 – 10,000 Differential-Algebraic Equations
- 100-1000 Scenarios
Stochastic Rolling Horizon 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

\[
\begin{aligned}
\min_{u(t)} & \quad \mathbb{E}_{\chi(t) \in \Omega} \left[ \int_{t_{\ell}}^{t_{\ell}+N} \varphi(z(t), y(t), u(t), \chi(t)) \, dt \right] \\
\begin{aligned}
\frac{dz}{dt} &= f(z(t), y(t), u(t), \chi(t)) \\
0 &= g(z(t), y(t), u(t), \chi(t)) \\
0 &\geq h(z(t), y(t), u(t), \chi(t)) \\
z(0) &= x_{\ell}
\end{aligned}
\end{aligned}
\]

\(\forall \chi(t) \in \Omega\)

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

\[
\begin{aligned}
\min_{u} & \quad \frac{1}{S} \sum_{k=1}^{S} \varphi(z_{k}, y_{k}, u, \chi_{k}) \\
\text{s.t.} & \quad c(z_{k}, y_{k}, u, \chi_{k}) = 0 \\
& \quad h(z_{k}, y_{k}, u, \chi_{k}) \leq 0 \\
& \quad k = 1, \ldots, S
\end{aligned}
\]
Dynamic System Model

Stochastic Optimization

\[ u^*(\tau) \]

Set-Points

Low-Level Control

\[ \chi(t_k) \]

Measurements

Forecast & Covariance

\[ \bar{X}(\tau), V(\tau) \]

\[ \tau \in [t_k, t_k + T] \]

Forecast

Weather Model

Quantifying Uncertainty Key Enabler
One weakness of stochastic programming is that it assumes a distribution is given. In most applications of interest, the distribution has to be modeled from data using some knowledge of the application.

If the uncertainty originates in weather forecast, there is a strong empirical and theoretical basis to create the distribution, or, at least to sample from it.

To our knowledge this is the first time even a moderately complex energy system was managed using stochastic programming with real and operational weather uncertainty.
Sequential Decision Making Under Uncertainty: Stochastic Dynamic Programming

- Stoc DP is the most distinguished framework, though rarely (ever ?) approached from HPC.

- Example: In Production and Inventory Planning:
  - Recursive Cost at the beginning of stage $t$:
    \[
    f_t(I, \omega_{1:(t-1)}) = \min_{\max(0, d_t-I) \leq x} \min(C, d_t+B-I) \ E_t \left[ c_t(x, \omega_t) + f_{t+1}(I + x - d_t, \omega_{1:t}) \right]
    \]
  - Functional approximation in an T*D (D~1000s) dimensional space !!! It suffers from the curse of dimensionality ... but so does sampling.
  - Could HPC make inroads? We believe so but development cost restricts us to rolling horizon.
Applications
App 1: The operator’s (ISO-IL) problem Unit commitment with wind power generation

- Deterministic problem:

\[
\begin{align*}
\min_{p_{j,k}, \bar{p}_{j,k}} & \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{T}} c_{j,k}^p + c_{j,k}^u + c_{j,k}^d \\
\text{s.t.} & \sum_{j \in \mathcal{N}} p_{j,k} + \sum_{j \in \mathcal{N}_{\text{wind}}} \mathbb{E} \{p_{j,k}^\text{wind}\} = D_k \\
& \sum_{j \in \mathcal{N}} \bar{p}_{j,k} + \sum_{j \in \mathcal{N}_{\text{wind}}} \mathbb{E} \{p_{j,k}^\text{wind}\} \geq D_k + R_k
\end{align*}
\]

- Stochastic program formulation:

\[
\begin{align*}
\min_{p_{s,j,k}, \bar{p}_{s,j,k}} & \frac{1}{N_S} \sum_{s \in \mathcal{S}} \left( \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{T}} c_{s,j,k}^p + c_{j,k}^u + c_{j,k}^d \right) \\
\text{s.t.} & \sum_{j \in \mathcal{N}} p_{s,j,k} + \sum_{j \in \mathcal{N}_{\text{wind}}} p_{s,j,k}^\text{wind} = D_k \\
& \sum_{j \in \mathcal{N}} \bar{p}_{s,j,k} + \sum_{j \in \mathcal{N}_{\text{wind}}} p_{s,j,k}^\text{wind} \geq D_k + R_k
\end{align*}
\]
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 than the one with exact information.

![Wind power forecast and stochastic programming graph](https://via.placeholder.com/150)
Ap2: Thermal management of building in Pittsburgh

- Minimize annual heating and cooling costs
- Time-varying electricity prices (peak/off-peak)
- Forecast with uncertainty leads to 20-80% cost reduction (insulation quality)

[Zavala, Constantinescu, Krause, and Anitescu, 2009]
• Operating Costs Driven by Solar Radiation Ulleberg, 2004
• Performance Deteriorated by Power Losses
Reactive (No Forecast)

- Costs Reduced By 300% From 1-Hr to 14-Day Forecast
- Close-to-Optimal Profit Achieved with 1 Day Forecast
Is it worth using Stochastic Programming and WRF? Or there are simple bypasses?

- The case for forecast seems clear. But how about
  - (Q1) Do I need Stochastic Programming and uncertainty? Maybe if I do deterministic programming on average it is sufficient.
  - (Q2) Do I need WRF to do it? Or can I get by with massaging historical data.
- We Present evidence of Yes on both counts.
App 1: Regional SO commitment

- Deterministic strategy (Programming on average) cannot satisfy demand beyond 10% wind penetration (The reserves help some).
- Evidence for Q1=Yes.
Load Satisfaction Deterministic ("Optimization on Mean") vs. Stochastic

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

Handling Stochastic Effects Particularly Critical in Grid-Independent Systems where no recourse.

Evidence for Q1=Yes.
Performance Optimizer using WRF and GP Model Forecasts. Evidence for Q2=Yes.
Conclusions and Future Work

**Integrative Study of Weather Forecast-Based Optimization**

- Far as we know this is the first integrative study using validated WRF Model for an operational setting (1km, 1 day ahead, in 1hr).
- We showed that stochastic formulation matters hugely for satisfying constraints.
- Weather uncertainty is a hard, important, problem that data-only methods (such as GP) are unlikely to crack.
- We showed that weather forecast inclusions results in 20-80% cost reduction for rolling horizon.

**Future and On-Going Work**

- Better posterior sampling for weather forecast uncertainty.
- Dynamic Programming Formulations and comparisons with rolling horizon.
- High-resolution physics in WRF (feedback from wind farms?).
- Solving larger Stochastic MINLP problems.
- Modeling pricing and demand uncertainty.
- Real time cost-efficient techniques (buildings and PV).
Promise of HPC for Integrated Energy Systems Management

1. Predicting Weather Forecast with Uncertainty Operationally.
   400K Processors would provide in one hour wall clock time 30 WRF Ensemble Members for the next 24 hours at 1 km resolution for the entire US. (Source: Extrapolation of Profiling)

2. Stochastic Dynamic Programming as a resource management strategy for Regional System Operators. ~ 10s of billions of dollars worth of activity per year. (Source: Our educated guess).
Accepted/Published.


Submitted


[PDF Version](http://example.com)