Solving stochastic unit commitment in a high performance computing environment

AASS Workshop – IMPA

Anthony Papavasiliou
Ignacio Aravena

Center for Operations Research and Econometrics,

Université catholique de Louvain

1966 - 2016 CORE BRIDGING GAPS

March 31st, 2016

Renewables Making Headlines


BBC News Sport Weather Capital Culture

NEWS EUROPE

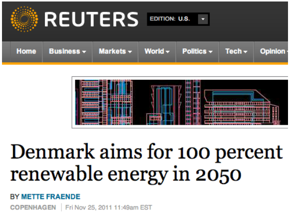
Home UK Africa Asia Europe Latin America Mid-East US & Canada Business Health

30 May 2011 Last updated at 12:25 GMT

Germany: Nuclear power plants to close by 2022

COMMENTS (542)

Germany saw mass anti-nuclear protests in the wake of the Fukushima disaster

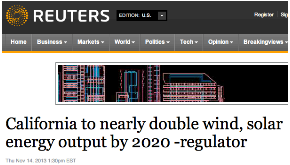

REUTERS EDITION: U.S.

Home Business Markets World Politics Tech Opinion

Denmark aims for 100 percent renewable energy in 2050

BY METTE FRAENDE

COPENHAGEN | Fri Nov 25, 2011 11:49am EST

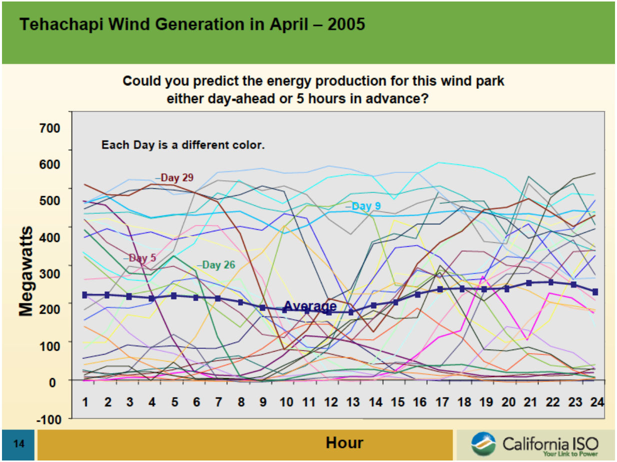

REUTERS EDITION: U.S. Register

Home Business Markets World Politics Tech Opinion Breakingviews

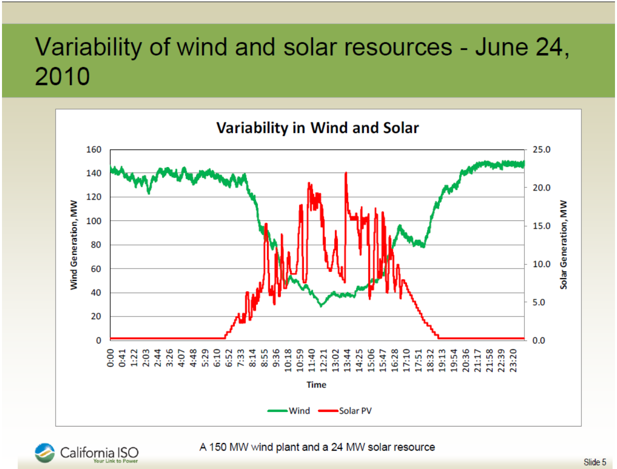
California to nearly double wind, solar energy output by 2020 -regulator

Thu Nov 14, 2013 1:30pm EST

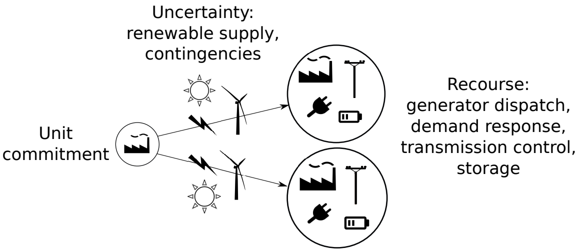
Uncertainty



Variability



Unit Commitment under Uncertainty



Appropriate for modeling various balancing options:

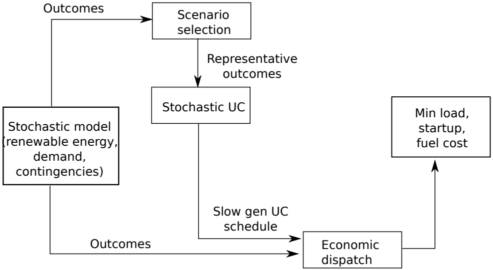
- Demand (deferrable, price responsive, wholesale)
- Storage (pumped hydro, batteries)
- Transmission control (FACTS, tap changers, switching)

Parallel Computing Literature in Power Systems

- Monticelli et al. (1987): Benders decomposition algorithm for SCOPF
- Pereira et al. (1990): Various applications of parallelization including SCOPF, composite (generator, transmission line) reliability, hydrothermal scheduling
- Falcao (1997): Survey of HPC applications in power systems
- Kim, Baldick (1997): Distributed OPF
- Bakirtzis, Biskas (2003) and Biskas et al. (2005): Distributed OPF

Full Model

- Application: stochastic unit commitment for large-scale renewable energy integration
- Two-stage model representing DA market (first stage) followed by RT market (second stage)



Unit Commitment Model

- Domain $\mathcal{D}$ represents min up/down times, ramping rates, thermal limits of lines, reserve requirements
- Generator set partitioned between fast (G_f) and slow (G_s) generators

$$(UC): \min \sum_{g\in G} \sum_{t\in T} (K_g u_{gt} + S_g v_{gt} + C_g p_{gt})$$

$$s.t. \sum_{g\in G_n} p_{gt} = D_{nt}$$

$$P_g^- u_{gt} \leq p_{gt} \leq P_g^+ u_{gt}$$

$$e_{lt} = B_l(\theta_{nt} - \theta_{mt})$$

$$(\mathbf{p}, \mathbf{e}, \mathbf{u}, \mathbf{v}) \in \mathcal{D}$$

Stochastic Unit Commitment Model

$$(SUC): \min \sum_{g \in G} \sum_{s \in S} \sum_{t \in T} \pi_s (K_g u_{gst} + S_g v_{gst} + C_g p_{gst})$$

$$s.t. \sum_{g \in G_n} p_{gst} = D_{nst},$$

$$P^-_{gs} u_{gst} \leq p_{gst} \leq P^+_{gs} u_{gst}$$

$$e_{lst} = B_{ls}(\theta_{nst} - \theta_{mst})$$

$$(\mathbf{p}, \mathbf{e}, \mathbf{u}, \mathbf{v}) \in \mathcal{D}_s$$

$$u_{gst} = w_{gt}, v_{gst} = z_{gt}, g \in G_s$$

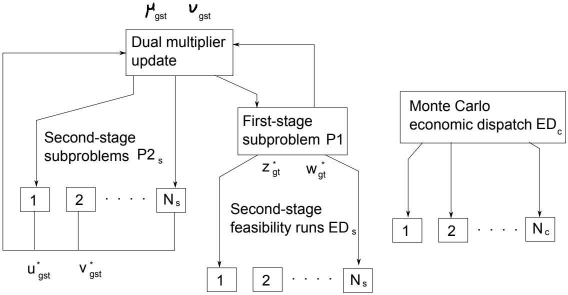
Lagrangian Decomposition Algorithm

- Past work: (Takriti et al., 1996), (Carpentier et al., 1996), (Nowak and Römisch, 2000), (Shiina and Birge, 2004)
- Key idea: relax non-anticipativity constraints on both unit commitment and startup variables
  1. Balance size of subproblems
  2. Obtain lower and upper bounds at each iteration

  Lagrangian:

$$\mathcal{L} = \sum_{g \in G} \sum_{s \in S} \sum_{t \in T} \pi_s (K_g u_{gst} + S_g v_{gst} + C_g p_{gst})$$
$$+ \sum_{g \in G_s} \sum_{s \in S} \sum_{t \in T} \pi_s (\mu_{gst}(u_{gst} - w_{gt}) + \nu_{gst}(v_{gst} - z_{gt}))$$

Parallelization

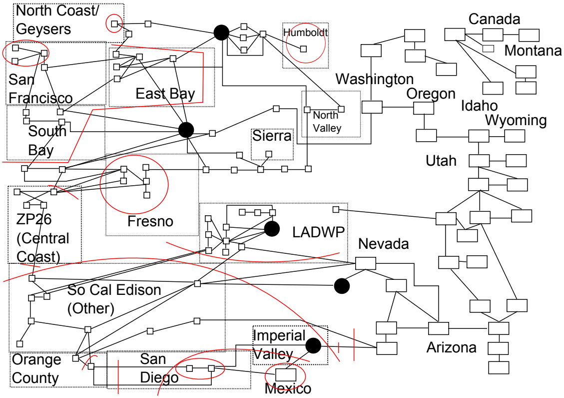


- Lawrence Livermore National Laboratory Hera cluster: 13,824 cores on 864 nodes, 2.3 Ghz, 32 GB/node
- MPI calling on CPLEX Java callable library

Scenario Selection

- Past work: (Gröwe-Kuska et al., 2002), (Dupacova et al., 2003), (Heitsch and Römisch, 2003), (Morales et al., 2009)
- Scenario selection algorithm inspired by importance sampling
  1. Generate a sample set $\Omega_S \subset \Omega$, where $M = |\Omega_S|$ is adequately large. Calculate the cost $C_D(\omega)$ of each sample $\omega \in \Omega_S$ against the best deterministic unit commitment policy and the average cost $\bar{C} = \sum_{i=1}^{M} \frac{C_D(\omega_i)}{M}$.
  2. Choose N scenarios from Ω_S, where the probability of picking a scenario ω is $C_D(\omega)/(M\bar{C})$.
  3. Set $\pi_s = C_D(\omega)^{-1}$ for all $\omega^s \in \hat{\Omega}$.

WECC Model



Eight day types: one per weekday/weekend × season

Unit Characteristics

Type	No. of units	Capacity (MW)
Nuclear	2	4,499
Gas	94	20,595.6
Coal	6	285.9
Oil	5	252
Dual fuel	23	4,599
Import	22	12,691
Hydro	6	10,842
Biomass	3	558
Geothermal	2	1,193
Wind (deep)	10	14,143
Fast thermal	88	11,006.1
Slow thermal	42	19,225.4

Model Size

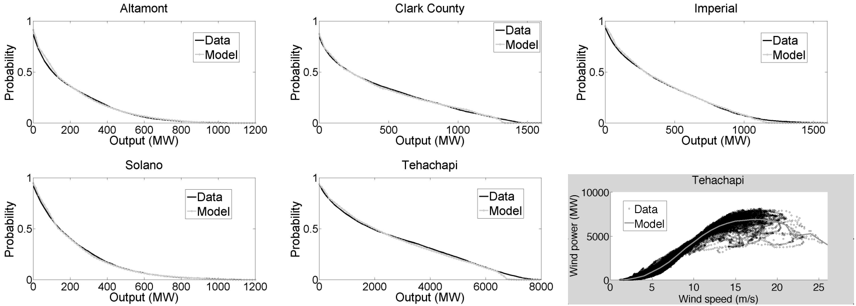
Model	Gens	Buses	Lines	Periods	Scens.
CAISO1000	130	225	375	24	1000
WILMAR	45	N/A	N/A	36	6
PJM	1011	13867	18824	24	1
CWE120	656	679	1037	96	120

Model	Integer var.	Cont. var.	Constraints
CAISO1000	3,121,800	20,643,120	66,936,000
WILMAR	16,000	151,000	179,000
PJM	24,264	833,112	1,930,776
CWE120	1,152,768	53,337,600	64,728,720

Wind Production Model

- Relevant literature: (Brown et al, 1984), (Torres et al., 2005), (Morales et al, 2010)
- Calibration steps
  1. Remove systematic effects:
$$y^S_{kt} = \frac{y_{kt} - \hat{\mu}_{kmt}}{\hat{\sigma}_{kmt}}.$$
  2. Transform data to obtain a Gaussian distribution:
$$y^{GS}_{kt} = N^{-1}(\hat{F}_k(y^S_{kt})).$$
  3. Estimate the autoregressive parameters $\hat{\phi}_{kj}$ and covariance matrix $\hat{\Sigma}$ using Yule-Walker equations.

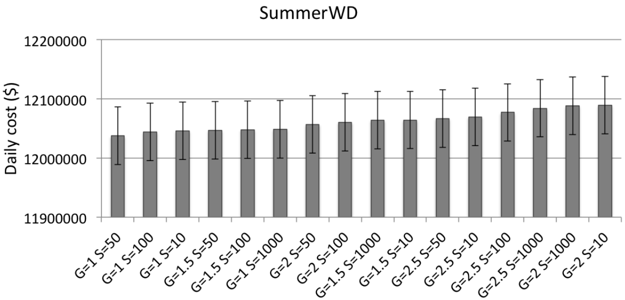
Data Fit


Altamont
Clark County
Imperial
Solano
Tehachapi
Tehachapi
Probability
Output (MW)
Data
Model
Wind power (MW)
Wind speed (m/s)

Number of Scenarios Versus Optimality Gap

- A large number of scenarios:
  - results in a more accurate representation of uncertainty
  - increases the amount of time required in each iteration of the subgradient algorithm
- A smaller optimality gap implies that the relaxation is ‘closer’ to an optimal solution
- Given a time budget (a few hours at best in day-ahead operations), do we want to solve a more representative problem less accurately or a less representative problem more accurately?

Cost Ranking: Summer Weekdays



- $S = 1000$ corresponds to sample average approximation algorithm
- Average daily cost and one standard deviation for 1000 Monte Carlo outcomes

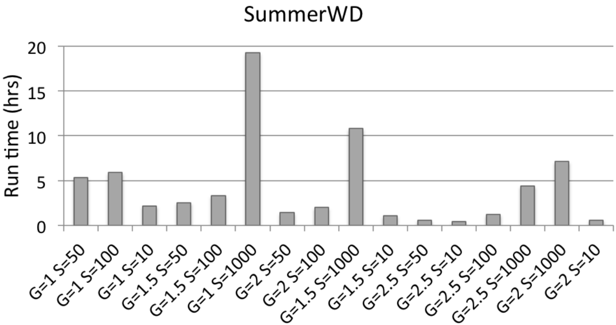
Influence of Duality Gap

- Among three worse policies in summer, $S = 1000$ with G = 2%, 2.5%
- Best policy for all day types has a 1% optimality gap ($S = 1000$ only for spring)
- For all but one day type the worst policy has $G = 2.5\%$
- For spring, best policy is $G = 1, S = 1000$
- For spring, summer and fall the worst policy is the one with the fewest scenarios and the greatest gap, namely $G = 2.5, S = 10$

Validation of Scenario Selection Policy

- Top performance for winter, summer and fall is attained by proposed scenario selection algorithm based on importance sampling
- For all day types, the importance sampling algorithm results in a policy that is within the top 2 performers
- Satisfactory performance (within top 3) can be attained by models of moderate scale (S50), provided an appropriate scenario selection policy is utilized

Run Time Ranking: Summer Weekdays

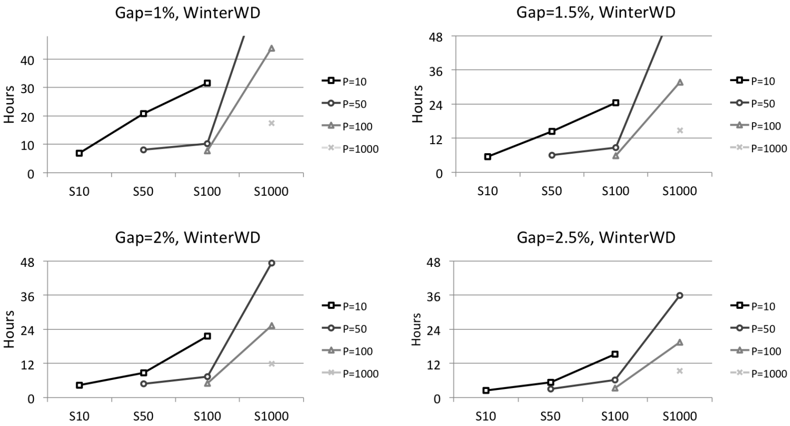


- Best-case running times ($S = P$)

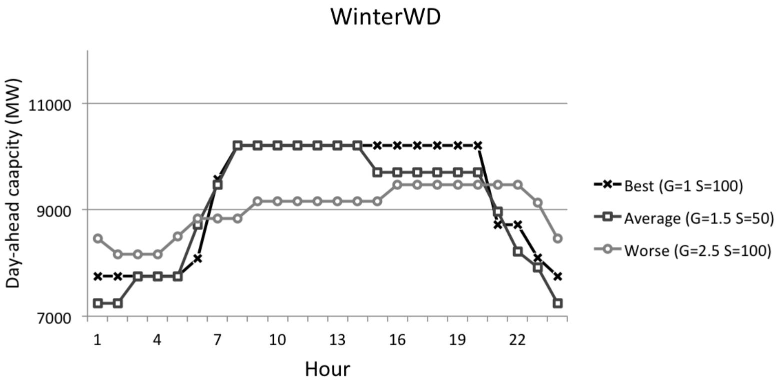
How Many Scenarios?

- Depends on the amount of available computation time and the number of available computational resources
- No guarantee that a smaller gap for the same instance will deliver a better result (compare, for example, the case of $G = 2$ with the case of $G = 2.5$ for $S = 10$ for winter weekdays). Nevertheless, it is commonly preferable to decrease the duality gap as much as possible

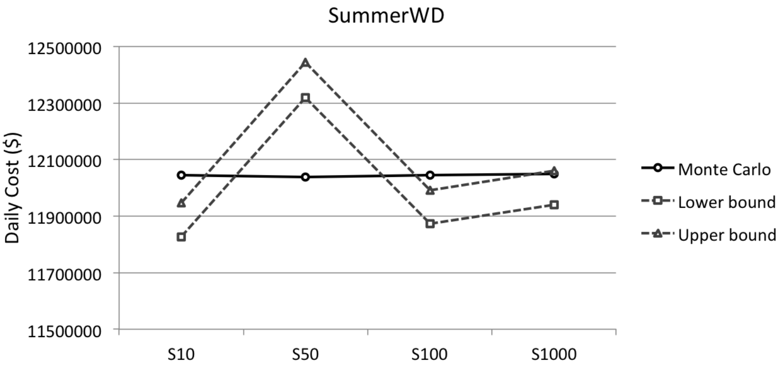
Running Times: Winter Weekdays



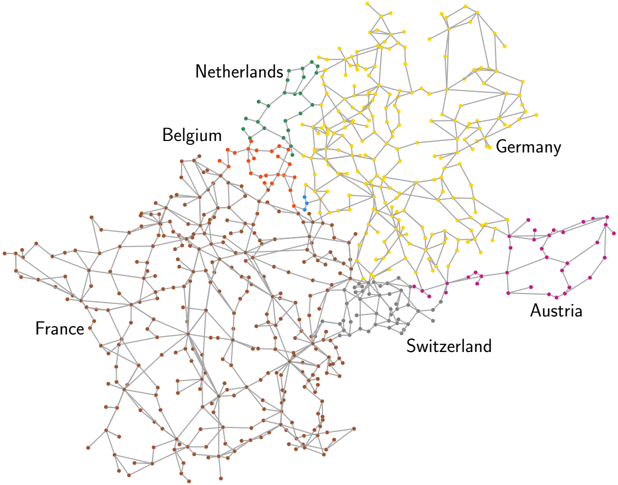
Unit Commitment: Winter Weekdays



Bounds: Summer Weekdays



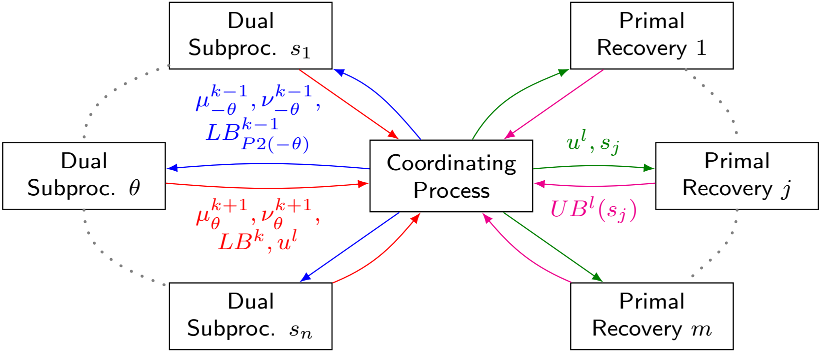
Central Western European Case Study



Asynchronous algorithm

- Synchronous method requires the evaluation of all scenario subproblems for current multipliers μ^k, ν^k in order to perform a new subgradient iteration
- Certain scenario subproblems can take up to 75 times more running time than others (more than 12 hours for hard subproblems compared to 10' for easy subproblems)
- **Idea**: use simpler algorithms for which each iteration requires to evaluate only part of the dual function
- Relevant literature: (Bertsekas & Tsitsiklis, 1989), (Tseng, 2001), (Nedić *et al.*, 2001), (Kiwiel, 2004), (Fercoq & Richtárik, 2013), (Liu *et al.*, 2014)

Proposed scheme



Note: $\mu_\theta^k, \nu_\theta^k$ are maintained within Dual Sub-process θ

Standard coordinate descent iteration

- $k(\theta)$: current iteration in sub-process θ
- Dual Sub-process θ:
  - Evaluates subproblem P2 for scenario θ with current multipliers $\boldsymbol{\mu}_\theta^{k(\theta)}, \boldsymbol{\nu}_\theta^{k(\theta)}$
  - Evaluates P1 with current full multipliers

$$\boldsymbol{\mu} := \left(\boldsymbol{\mu}_{s_1}^{k(s_1)}, \ldots, \boldsymbol{\mu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\mu}_{s_n}^{k(s_n)}\right)$$
$$\boldsymbol{\nu} := \left(\boldsymbol{\nu}_{s_1}^{k(s_1)}, \ldots, \boldsymbol{\nu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\nu}_{s_n}^{k(s_n)}\right)$$

  - Computes block-coordinate subgradient update on $\boldsymbol{\mu}_\theta, \boldsymbol{\nu}_\theta$

- **Problem:** dual function is never fully evaluated $\rightarrow$ impossibility to compute lower bounds

Modified dual iterations

- Dual Sub-process θ:
  - Evaluates subproblem P2 for scenario θ with the current multipliers $\mu_\theta^{k(\theta)}, \nu_\theta^{k(\theta)} \to LB_{P2(\theta)}^{k(\theta)}$
  - Evaluates P1 with **delayed** multipliers $\bar{\mu}, \bar{\nu} \to LB_{P1}^{k(\theta)}$

$$\bar{\mu} := \left(\mu_{s_1}^{k(s_1)-1}, \ldots, \mu_\theta^{k(\theta)}, \ldots, \mu_{s_n}^{k(s_n)-1}\right)$$
$$\bar{\nu} := \left(\nu_{s_1}^{k(s_1)-1}, \ldots, \nu_\theta^{k(\theta)}, \ldots, \nu_{s_n}^{k(s_n)-1}\right)$$

  - Computes block-coordinate subgradient update on μ_θ, ν_θ
  - Computes **lower bound on objective using last evaluations of P2** subproblems for other scenarios,

$$\text{Objective} \geq LB_{P1}^{k(\theta)} + LB_{P2(\theta)}^{k(\theta)} + \sum_{s \neq \theta} LB_{P2(s)}^{k(s)-1}$$

Lower bound initialization

- Certain scenario subproblems can take up to 75 times more than others to be solved $\rightarrow$ **one scenario** can delay the computation of the first “full” lower bound

- Use a relaxation of P2 to obtain an initial lower bound (not useful for updating dual multipliers)
- Which relaxation?
  - Linear relaxation of P2
  - Sequence of OPFs problems

Primal recovery

- Recovering primal candidates (1st stage) from P2 subproblems $\rightarrow$ good quality solutions from first iterations, (Ahmed, 2013)
- Accumulating large number of primal candidates: prune bad candidates if possible
  - Pruning candidates based on cuts from (Angulo *et al.*, 2014)
  - Second stage cost non-increasing function of $\boldsymbol{u}$: $\boldsymbol{u}^i \geq \boldsymbol{u}^j \Rightarrow C_2(\boldsymbol{u}^i) \leq C_2(\boldsymbol{u}^j)$, hence

$$LB(\boldsymbol{u}^{\text{new}}) = C_1(\boldsymbol{u}^{\text{new}}) + \max_{\substack{j \in J \\ u^j \geq u^{\text{new}}}} C_2(\boldsymbol{u}^j)$$

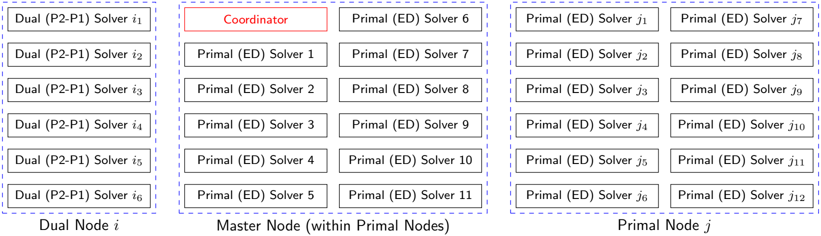
- Asynchronous evaluation of 2nd stage cost for candidates

Implementation details

- Implemented in Mosel using the `mmjobs` module and the XPress solver
- Offline self-tuning of solver parameters for solving P2 subproblems can save up 30% of solution time
- Lawrence Livermore National Laboratory Sierra cluster: 23,328 cores on 1,944 nodes, 2.8 Ghz, 24 GB/node
- Using 10 nodes per SUC instance, multiple subproblems per node

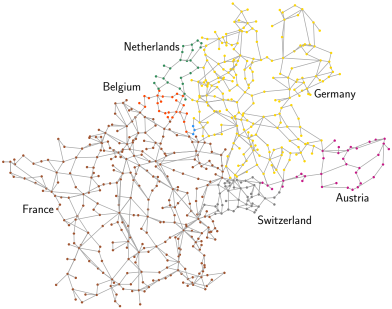
Implementation details

- 5 nodes dedicated to dual iterations / 6 sub-processes per node (due to subproblem P2 memory requirements)
- 5 nodes dedicated to primal recovery / 12 primal recovery scenario sub-problems per node



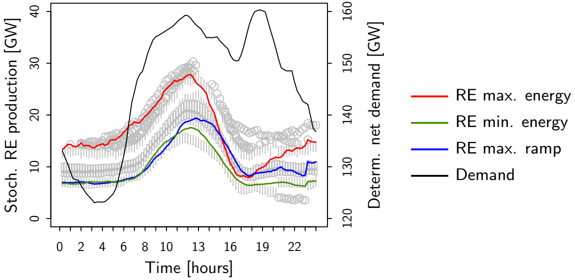
Central Western European Model

- 87 nuclear units (85GW), 144 CHP units (40GW), 272 SLOW units (99GW), 126 FAST units (14GW) and 27 aggregated generators (10GW)
- CWE grid model, Hutcheon & Bialek, 2013 7 countries, 679 nodes, 1073 lines



Renewable energy production profiles

- Multiarea renewable production and demand with 15' time resolution for 2013-2014 collected from national TSOs
- Using typical profiles + forecast errors to generate representative profiles



Simulation setting

- Comparing 5 day-ahead scheduling models:
  - Deterministic UC with secondary and tertiary reserves, **Determ2R**
  - Deterministic UC with primary, secondary and tertiary reserves, **Determ3R**
  - Stochastic UC with 30, 60 and 120 scenarios; **Stoch30**, **Stoch60** and **Stoch120**
- Fixed commitment for NUCLEAR and CHP. No commitment decision associated with AGGREGATED generators.
- 8 day types: 4 seasons $\times$ weekdays/weekends
- Using scenario reduction based on probability metrics, (Heitsch & Römisch, 2007)

Central Western European Model instances

Model	Scenarios	Variables	Constraints	Integers
Determ2R	1	570,432	655,784	9,552
Determ3R	1	636,288	719,213	9,552
Stoch30	30	13,334,400	16,182,180	293,088
Stoch60	60	26,668,800	32,364,360	579,648
Stoch120	120	53,337,600	64,728,720	1,152,768

Running times

Solution statistics over 8 instances (day types).

Model	Nodes	Running time [hours]	Worst final gap [%]
Determ2R	1	1.9 (0.6 – 4.2)	0.95
Determ3R	1	13.6 (6.3 – 27.9)	1.12
Stoch30[1]	10	1.1 (0.7 – 2.2)	0.93
Stoch30i[2]	10	0.8 (0.3 – 1.8)	1.00
Stoch60[1]	10	3.2 (1.1 – 8.4)	1.00
Stoch60i[2]	10	1.5 (0.6 – 4.7)	0.97
Stoch120[1]	10	6.1 (1.6 – 15.0)	1.00
Stoch120i[2]	10	3.0 (1.4 – 10.0)	1.07

[1] Dual initialization using linear relaxation of P2

[2] Dual initialization using sequential OPF

Deterministic or Stochastic UC?

Model	Nodes	Running time [hours]	Worst final gap [%]
Determ2R	1	1.9 (0.6 – 4.2)	0.95
Determ3R	1	13.6 (6.3 – 27.9)	1.12
Stoch60i	10	1.5 (0.6 – 4.7)	0.97

- For a large-scale power system, HPC enables solving SUC within the running time of a state-of-the-art MILP solver for DUC with reserves
- Stochastic UC provides cheaper and more reliable schedules, without the need for exogenous reserve targets
- Good news: **we can choose Stochastic UC!**

Stochastic UC: Optimality Vs. Wall-Time

Solution statistics over 8 instances (day types).

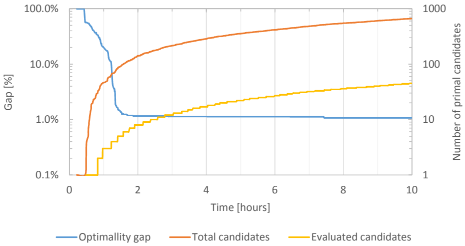
Model	Worst gap [%]			
	1 hour	2 hours	4 hours	8 hours
Stoch30	7.59	1.02	0.93	–
Stoch30i	1.90	1.00	–	–
Stoch60	23.00	5.32	5.22	4.50
Stoch60i	4.60	1.57	1.03	0.97
Stoch120	70.39	31.66	4.61	1.87
Stoch120i	46.69	27.00	1.42	1.07

- Lower bound initialization using sequential OPF demonstrates to be very effective, sometimes avoiding to solve P2 for hard scenarios
- Asynchronous SUC algorithm capable of achieving acceptable optimality gaps within operational time frames

Room for Improvement: Primal Candidate Evaluation

Bounds and primal candidates.

Stoch120i, 5 dual – 5 primal nodes, summer weekday.



- Primal candidates pruning is not effective: discards less than 1% of candidates
- Valuable computational resources spent in **detailed** evaluation of sub-optimal candidates

Using More Computational Power...

Solution statistics for Stoch120i over 8 instances.

Nodes	Running time [hours]	Worst gap [%]			
		1 hour	2 hours	4 hours	8 hours
$5D$, $5P$	3.0 (1.4 – 10.0)	46.69	27.00	1.42	1.07
$5D$, $10P$	2.0 (1.3 – 4.1)	46.04	25.51	1.04	1.00

- More cluster nodes dedicated to primal evaluation (P) can significantly reduce running times
- Analogous effect to use a more effective pruning mechanism $\rightarrow$ direction for further research

Application: Policy Analysis on CWE System

- Comparing different policy designs for day-ahead and real-time energy markets in the Central Western European network
  - Zonal day-ahead market and limited real-time coordination between zones, **ZonalDA-LimRT**
  - Zonal day-ahead market and complete real-time coordination between zones, **ZonalDA-ComRT**
  - Centralized deterministic UC in day-ahead and complete real-time coordination, **Deterministic UC**
  - Centralized stochastic UC in day-ahead and complete real-time coordination, **Stochastic UC**
- High levels of renewable energy currently integrated in the system: 51.2 GW of wind power and 47.3 GW solar power

Policy Analysis Results

Expected policy costs and efficiency losses with respect to deterministic UC

Policy	Expected cost [MM€/d]	Efficiency losses [%]	Efficiency losses [MM€/year]
ZonalDA-LimRT	30.49	6.0	631
ZonalDA-ComRT	29.56	2.8	295
Deterministic UC	28.76	–	–
Stochastic UC	28.49	–0.9	–96

- Zonal day-ahead market ignores congestion within zones leading to increased operation costs
- Small efficiency gains of stochastic UC compared to efficiency losses due to zonal day-ahead market design

Conclusions

- **Validation of scenario selection algorithm:** The importance sampling scenario selection algorithm performs favorably relative to SAA with 1000 scenarios
- **Decreasing the duality gap versus increasing the number of scenarios:** Reducing the duality gap seems to yield comparable benefits relative to adding more scenarios
- **Efficiency gains:** All problems solved within 24 hours, given enough processors. Parallelization permits the running time of the studied model to run within acceptable time frames from operations standpoint.

Conclusions

- **Asynchronous algorithm and HPC:** Potential to solve stochastic UC within the same time frame required to solve deterministic UC using a state-of-the-art MILP solver
- **Lower bound initialization:** Sequential OPF provides fast lower bounds, significantly reducing running times.
- **Primal recovery scheme:** Obtaining good primal candidates from first iterations drastically accelerates the convergence of the algorithm. Nevertheless, large scenario instances lead to a large number of sub-optimal candidates that could potentially be pruned.

Perspectives

- Extensions of present model
  - Comparison of alternative relaxations
  - Analysis of duality gap
- Extensions of asynchronous algorithm
  - Pruning and scoring candidates based on bounds for ED subproblems
  - Dynamical queue management for dual and primal processes
  - Multi-stage stochastic UC

References

- A. Papavasiliou, S. S. Oren, *Multi-Area Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network*, Operations Research, vol. 61, no. 3, pp. 578-592, May/June 2013.
- A. Papavasiliou, S. S. Oren, B. Rountree, *Applying High Performance Computing to Multi-Area Stochastic Unit Commitment for Renewable Penetration*, IEEE Transactions on Power Systems, vol. 30, no. 3, pp. 1690-1701, May 2015.
- I. Aravena, A. Papavasiliou, *A distributed asynchronous algorithm for the two-stage stochastic unit commitment problem, IEEE PES General Meeting*, 2015.
- I. Aravena, A. Papavasiliou. *Renewable Energy Integration in Zonal Markets*, under review.

Thank you

Questions?

Contact: anthony.papavasiliou@uclouvain.be

http://perso.uclouvain.be/anthony.papavasiliou/public_html/