Staffing and scheduling flexible call centers by two-stage robust optimization

Sara Mattia
IASI-CNR
sara.mattia@iasi.cnr.it

joint work with

Fabrizio Rossi, Stefano Smriglio
DISIM, Università di L'Aquila

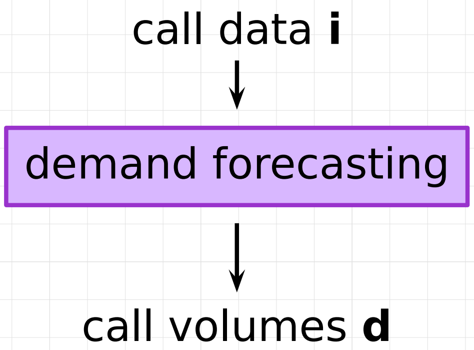
Mara Servilio
IASI-CNR

this talk

robust shift scheduling in call centers

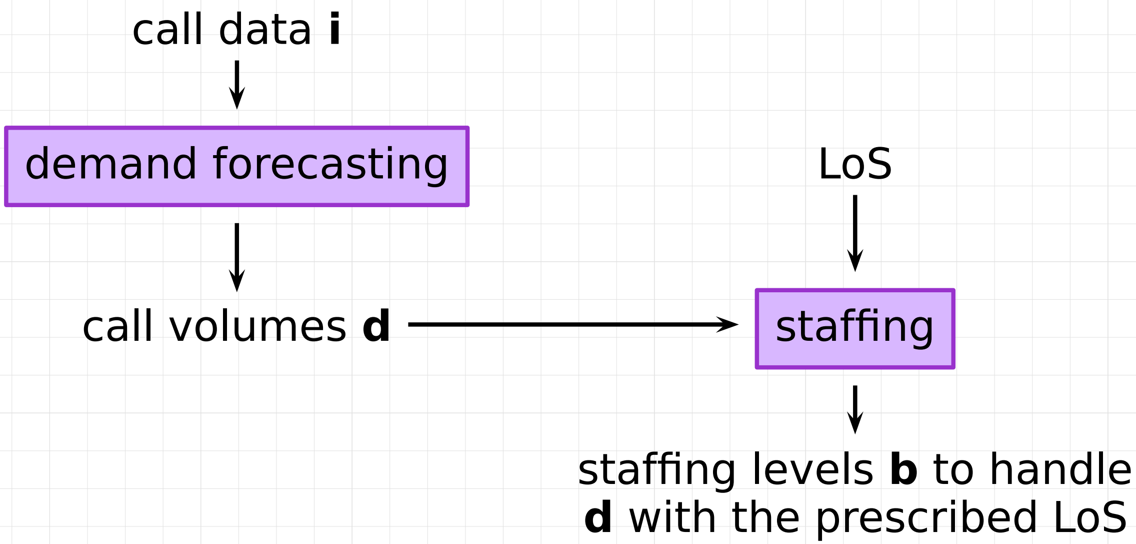
this talk

robust shift scheduling in call centers



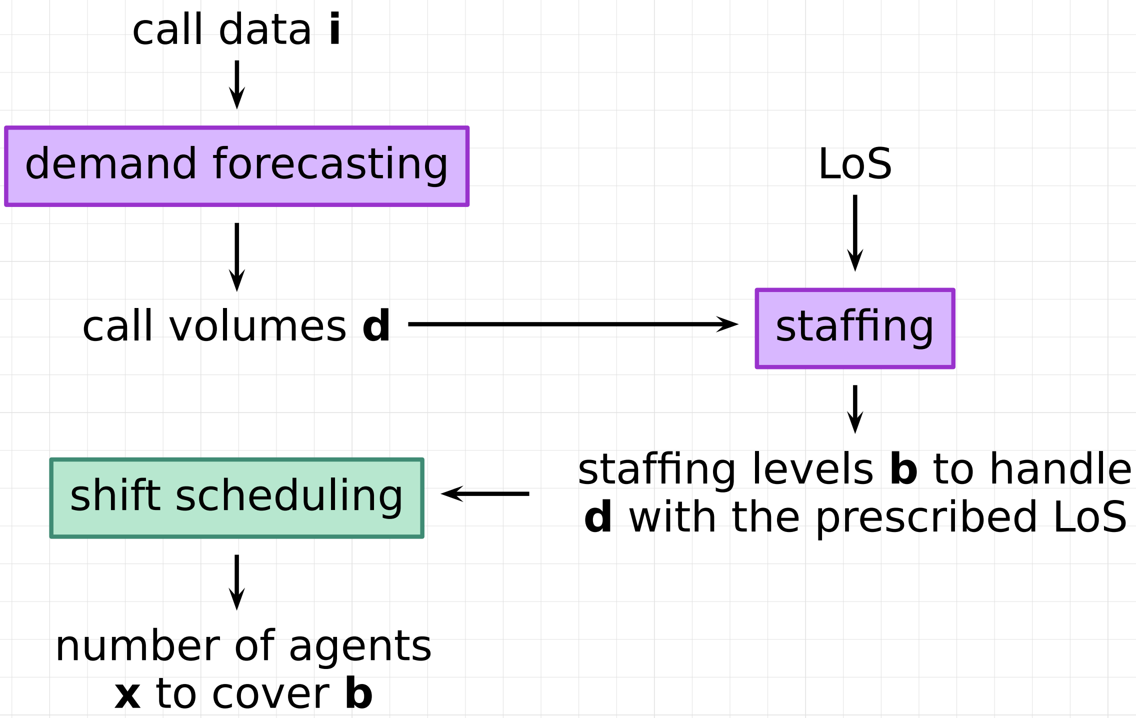
this talk

robust shift scheduling in call centers



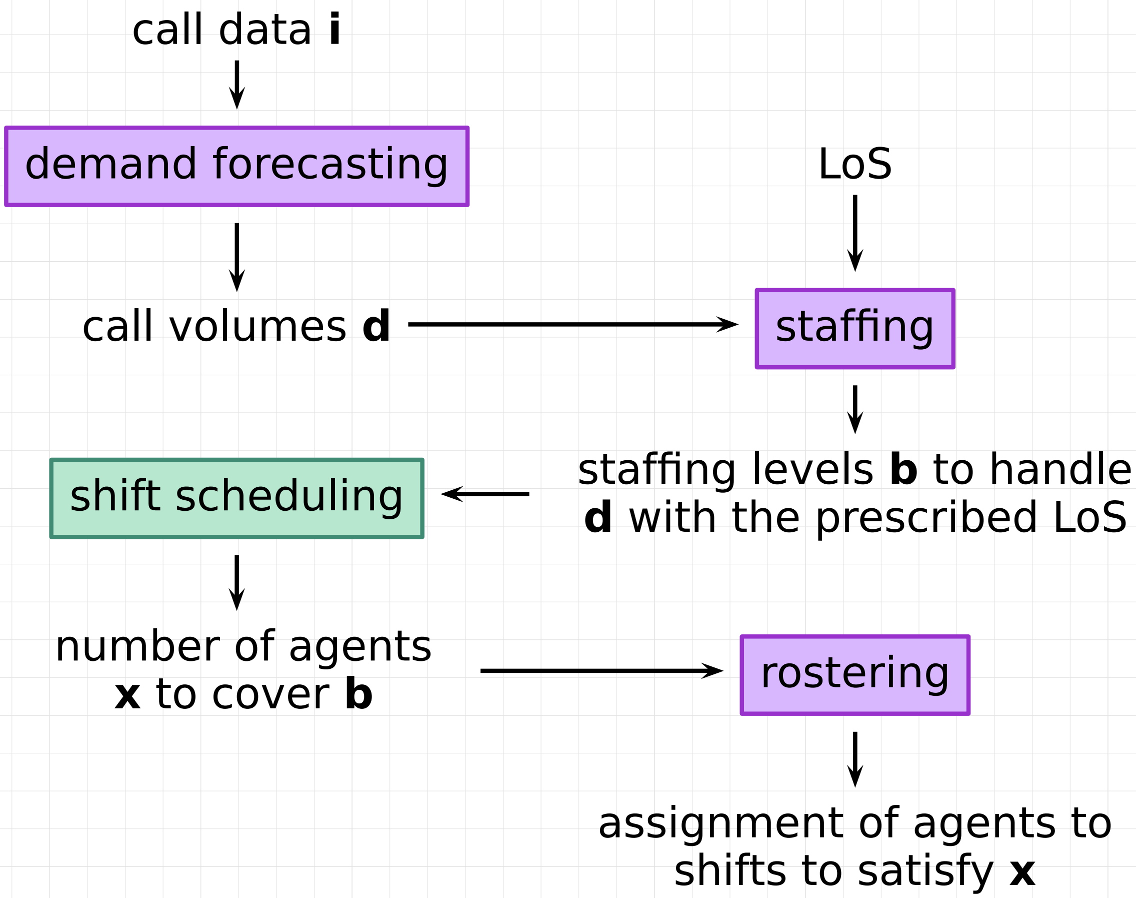
this talk

robust shift scheduling in call centers



this talk

robust shift scheduling in call centers



input

- time horizon T
- staffing levels b_t, $t \in T$
- shifts J

shift scheduling problem

choose the number of employees to be assigned to the shifts to cover the demand at minimum cost

input

- time horizon T
- staffing levels b_t, $t \in T$
- shifts J, with costs $\mathbf{c}$
- $A \in \{0,1\}^{|T| \times |J|}$, $a_{tj} = 1$ if j covers t, zero otherwise

classic model [Segal 1974]

$$\min \sum_{j \in J} c_j x_j$$

$$\sum_{j \in J} a_{tj} x_j \geq b_t \quad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^n$$

classic model [Segal 1974]

$$\min \sum_{j \in J} c_j x_j$$

$$\sum_{j \in J} a_{tj} x_j \geq b_t \quad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}$$

advantages

- well-studied model
- network structure (C1P)

classic model [Segal 1974]

$$\min \sum_{j \in J} c_j x_j$$

$$\sum_{j \in J} a_{tj} x_j \geq b_t \quad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}$$

advantages

- well-studied model
- network structure (C1P)

assumptions

- no understaffing and overstaffing costs
- **b** perfectly matches the real demand

flexible model

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

- u_t, o_t under and overstaffing in $t \in T$, with costs w_t^u and w_t^o

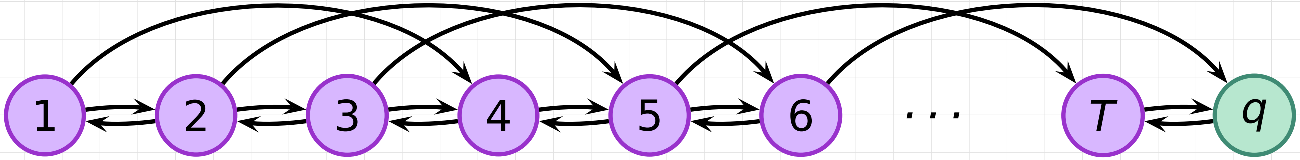
flexible model

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

- u_t, o_t under and overstaffing in $t \in T$, with costs w_t^u and w_t^o
- network structure (C1P)



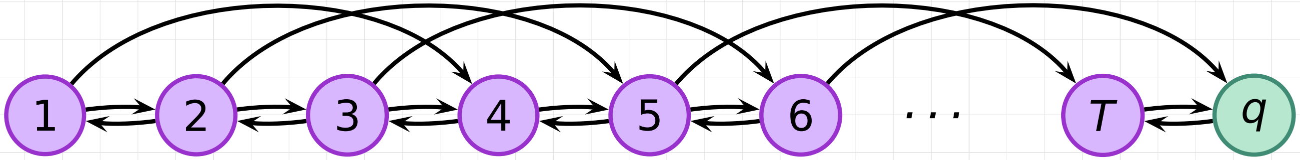
flexible model

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

- u_t, o_t under and overstaffing in $t \in T$, with costs w_t^u and w_t^o
- network structure (C1P)



- $o_t u_t = 0$, $t \in T$ for any optimal solution

staffing levels and real demands

computed staffing levels may differ from the real demands

staffing levels and real demands

computed staffing levels may differ from the real demands

- errors in call volume estimation

staffing levels and real demands

computed staffing levels may differ from the real demands

- errors in call volume estimation
- approximations made at the staffing stage (SIPP model)
- significant correlation between consecutive slots occurs in practice [Avramidis et al. 2009]

staffing levels and real demands

computed staffing levels may differ from the real demands

- errors in call volume estimation
- approximations made at the staffing stage (SIPP model)
- significant correlation between consecutive slots occurs in practice [Avramidis et al. 2009]

how variations from real demands are handled

- simulation-based cutting plane methods [Atlason et al. 2008, Avramidis et al. 2004]

staffing levels and real demands

computed staffing levels may differ from the real demands

- errors in call volume estimation
- approximations made at the staffing stage (SIPP model)
- significant correlation between consecutive slots occurs in practice [Avramidis et al. 2009]

how variations from real demands are handled

- simulation-based cutting plane methods [Atlason et al. 2008, Avramidis et al. 2004]
- stochastic programming [Bodur and Luedtke 2016, Liao et al. 2012, Thomas and Harrison 2010]

staffing levels and real demands

computed staffing levels may differ from the real demands

- errors in call volume estimation
- approximations made at the staffing stage (SIPP model)
- significant correlation between consecutive slots occurs in practice [Avramidis et al. 2009]

how variations from real demands are handled

- simulation-based cutting plane methods [Atlason et al. 2008, Avramidis et al. 2004]
- stochastic programming [Bodur and Luedtke 2016, Liao et al. 2012, Thomas and Harrison 2010]
- managers react to front-end understaffing by moving people from/to the back-office

this talk

- an exact method matching the managers practice

this talk

- an exact method matching the managers practice ⇒ two-stage optimization

this talk

- an exact method matching the managers practice $\Rightarrow$ two-stage optimization
- **b** uncertain

this talk

- an exact method matching the managers practice ⇒ two-stage optimization
- **b** uncertain ⇒ merge staffing aspects and shift scheduling

this talk

- an exact method matching the managers practice ⇒ two-stage optimization
- **b** uncertain ⇒ merge staffing aspects and shift scheduling
- uncertainty set

this talk

- an exact method matching the managers practice ⇒ two-stage optimization
- **b** uncertain ⇒ merge staffing aspects and shift scheduling
- uncertainty set
  - cardinality

    deviations typically occur only in a limited number of time periods

this talk

- an exact method matching the managers practice ⇒ two-stage optimization
- **b** uncertain ⇒ merge staffing aspects and shift scheduling
- uncertainty set
  - cardinality

    deviations typically occur only in a limited number of time periods
  - correlation

    deviations at consecutive time periods are often not independent

shift scheduling problem

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

robust optimization

shift scheduling problem

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

robust optimization

- single stage
  - single stage with RHS only uncertainty is easy [Minoux 2008]
  - no single stage with equality constraints

shift scheduling problem

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

robust optimization

- single stage
- two-stage
  - two stage with RHS only uncertainty is difficult [Minoux 2011]
  - branch-and-cut

shift scheduling problem

$$\min \sum_{j \in J} c_j x_j + \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$\sum_{j \in J} a_{tj} x_j + u_t - o_t = b_t \qquad t \in T$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}, \mathbf{u}, \mathbf{o} \in \mathbb{Z}_+^{|T|}$$

robust optimization

- single stage
- two-stage
- affinely adjustable
  - no second stage full flexibility
  - compact reformulation if U computationally tractable

robust problem

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U} \begin{array}{ll} \min \sum_{t\in T}(w_t^o o_+ w_t^u u_t) & \\ u_t - o_t = b_t - \sum_{j\in J} a_{tj} x_j & t \in T \\ \mathbf{u}, \mathbf{o} \in \mathbb{R}^{|T|} & \end{array} \right\}$$

robust problem

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U} \begin{array}{ll} \min \sum_{t\in T} (w_t^o o_+ w_t^u u_t) & \\ u_t - o_t = b_t - \sum_{j\in J} a_{tj} x_j & t \in T \\ \mathbf{u}, \mathbf{o} \in \mathbb{R}^{|T|} & \end{array} \right\}$$

robust problem

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U} \begin{array}{ll} \min \sum_{t\in T}(w_t^o o_+ w_t^u u_t) & \\ u_t - o_t = b_t - \sum_{j\in J} a_{tj} x_j & t\in T \\ \mathbf{u},\mathbf{o}\in\mathbb{R}^{|T|} & \end{array} \right\}$$

robust problem

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U} \begin{array}{ll} \min \sum_{t\in T}(w_t^o o_+ w_t^u u_t) & \\ u_t - o_t = b_t - \sum_{j\in J} a_{tj} x_j & t \in T \\ \mathbf{u}, \mathbf{o} \in \mathbb{R}^{|T|} & \end{array} \right\}$$

personnel reallocation (PR) cost

$$\min \sum_{t\in T}(w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = \bar{b}_t - \sum_{j\in J} a_{tj}\bar{x}_j \quad t \in T$$

$$\mathbf{u}, \mathbf{o} \in \mathbb{R}^{|T|}$$

$$\max \sum_{t\in T}(\bar{b}_t - \sum_{j\in J} a_{jt}\bar{x}_j) y_t$$

$$-w_t^o \leq y_t \leq w_t^u \quad t \in T$$

robust problem

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U} \begin{array}{ll} \min \sum_{t\in T}(w_t^o o_+ w_t^u u_t) & \\ u_t - o_t = b_t - \sum_{j\in J} a_{tj} x_j & t \in T \\ \mathbf{u},\mathbf{o} \in \mathbb{R}^{|T|} & \end{array} \right\}$$

personnel reallocation (PR) cost

$$\min \sum_{t\in T}(w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = \bar{b}_t - \sum_{j\in J} a_{tj}\bar{x}_j \quad t \in T$$

$$\mathbf{u},\mathbf{o} \in \mathbb{R}^{|T|}$$

$$\max \sum_{t\in T}(\bar{b}_t - \sum_{j\in J} a_{jt}\bar{x}_j) y_t$$

$$-w_t^o \le y_t \le w_t^u \quad t \in T$$

Benders-like reformulation

$$\min_{\mathbf{x}\in\mathbb{Z}_+^{|J|}} \left\{ \sum_{j\in J} c_j x_j + \max_{\mathbf{b}\in U, -w_t^o \le y_t \le w_t^u,\, t\in T} \sum_{t\in T}(b_t - \sum_{j\in J} a_{jt} x_j) y_t \right\}$$

Benders-like reformulation

$$\min \sum_{j \in J} c_j x_j + \lambda$$

$$\lambda \geq \sum_{t \in T} (b_t - \sum_{j \in J} a_{jt} x_j) y_t \quad \mathbf{b} \in U, \mathbf{y} \in dual(PR)$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}$$

Benders-like reformulation

$$\min \sum_{j \in J} c_j x_j + \lambda$$

$$\lambda \geq \sum_{t \in T} (b_t - \sum_{j \in J} a_{jt} x_j) y_t \quad \mathbf{b} \in U, \mathbf{y} \in dual(PR)$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}$$

the separation problem

$$\max_{\mathbf{b} \in U} \quad \min \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = \bar{b}_t - \sum_{j \in J} a_{tj} \bar{x}_j \quad t \in T$$

$$\mathbf{o}, \mathbf{u} \in \mathbb{R}_+^{|T|}$$

Benders-like reformulation

$$\min \sum_{j \in J} c_j x_j + \lambda$$

$$\lambda \geq \sum_{t \in T} (b_t - \sum_{j \in J} a_{jt} x_j) y_t \quad \mathbf{b} \in U, \mathbf{y} \in dual(PR)$$

$$\mathbf{x} \in \mathbb{Z}_+^{|J|}$$

the separation problem

$$\max_{\mathbf{b} \in U} \quad \min \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = \bar{b}_t - \sum_{j \in J} a_{tj} \bar{x}_j \quad t \in T$$

$$\mathbf{o}, \mathbf{u} \in \mathbb{R}_+^{|T|}$$

$$\max \sum_{t \in T} (b_t - \sum_{j \in J} a_{jt} \bar{x}_j) y_t$$

$$-w_t^o \leq y_t \leq w_t^u \quad t \in T$$

$$\mathbf{b} \in U$$

the uncertainty set

- cardinality

  deviations typically occur only in a limited number of time periods

- correlation

  deviations at consecutive time periods are often not independent

the uncertainty set

- cardinality

  deviations typically occur only in a limited number of time periods

- correlation

  deviations at consecutive time periods are often not independent

$$U = \left\{ \mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T \right.$$

the uncertainty set

- cardinality

  deviations typically occur only in a limited number of time periods

- correlation

  deviations at consecutive time periods are often not independent

$$U = \Big\{\mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T$$
$$\boldsymbol{\pi} \in \{0,1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \le \pi_t, t \in T$$

the uncertainty set

- cardinality

  deviations typically occur only in a limited number of time periods

- correlation

  deviations at consecutive time periods are often not independent

$$U = \Big\{ \mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T \\ \boldsymbol{\pi} \in \{0,1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \leq \pi_t, t \in T \\ \sum_{t \in T} \pi_t \leq \Gamma$$

the uncertainty set

- cardinality

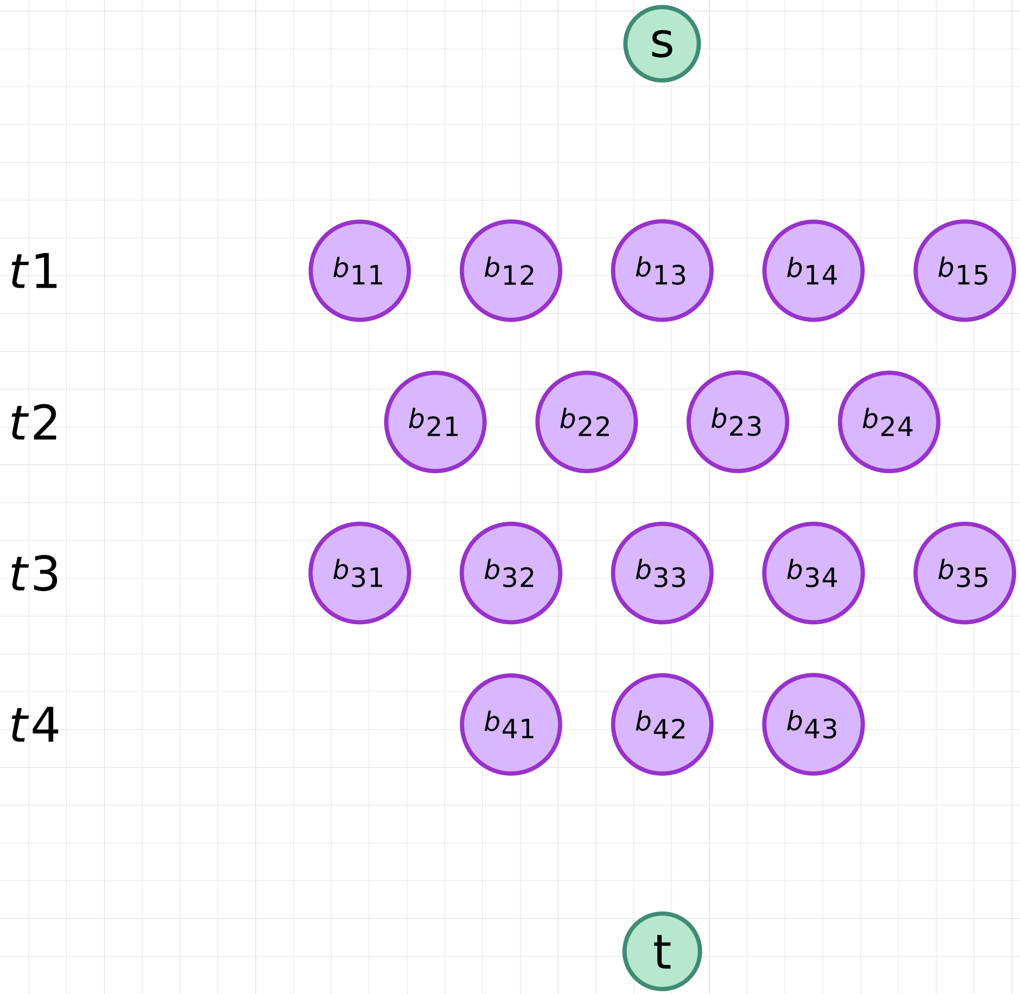
  deviations typically occur only in a limited number of time periods

- correlation

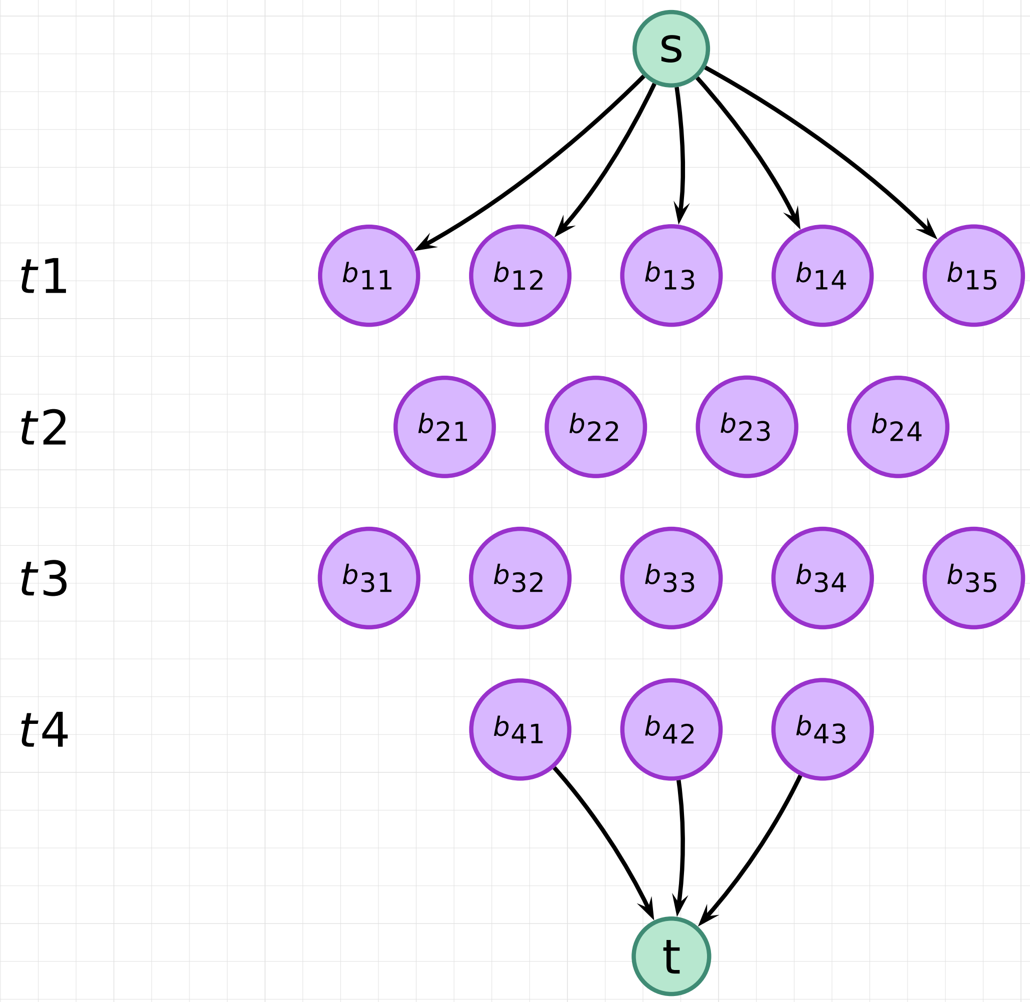
  deviations at consecutive time periods are often not independent

$$
\begin{aligned}
U \quad = \quad & \Big\{ \mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T \\
& \boldsymbol{\pi} \in \{0,1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \leq \pi_t, t \in T \\
& \sum_{t \in T} \pi_t \leq \Gamma \\
& |D_t p_t - D_{t-1} p_{t-1}| \leq \Delta_t,\ t \in T \setminus \{1\} \Big\}
\end{aligned}
$$

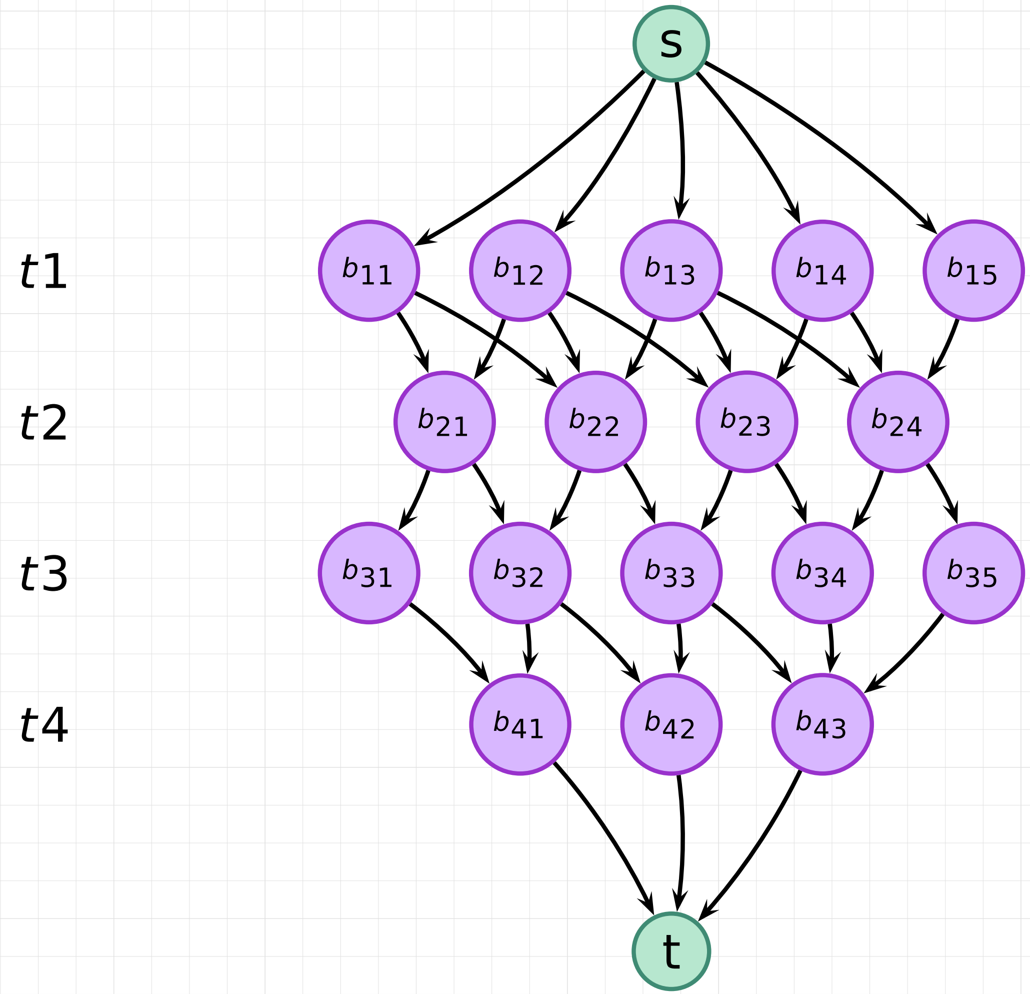
separation by (single) resource constrained longest path



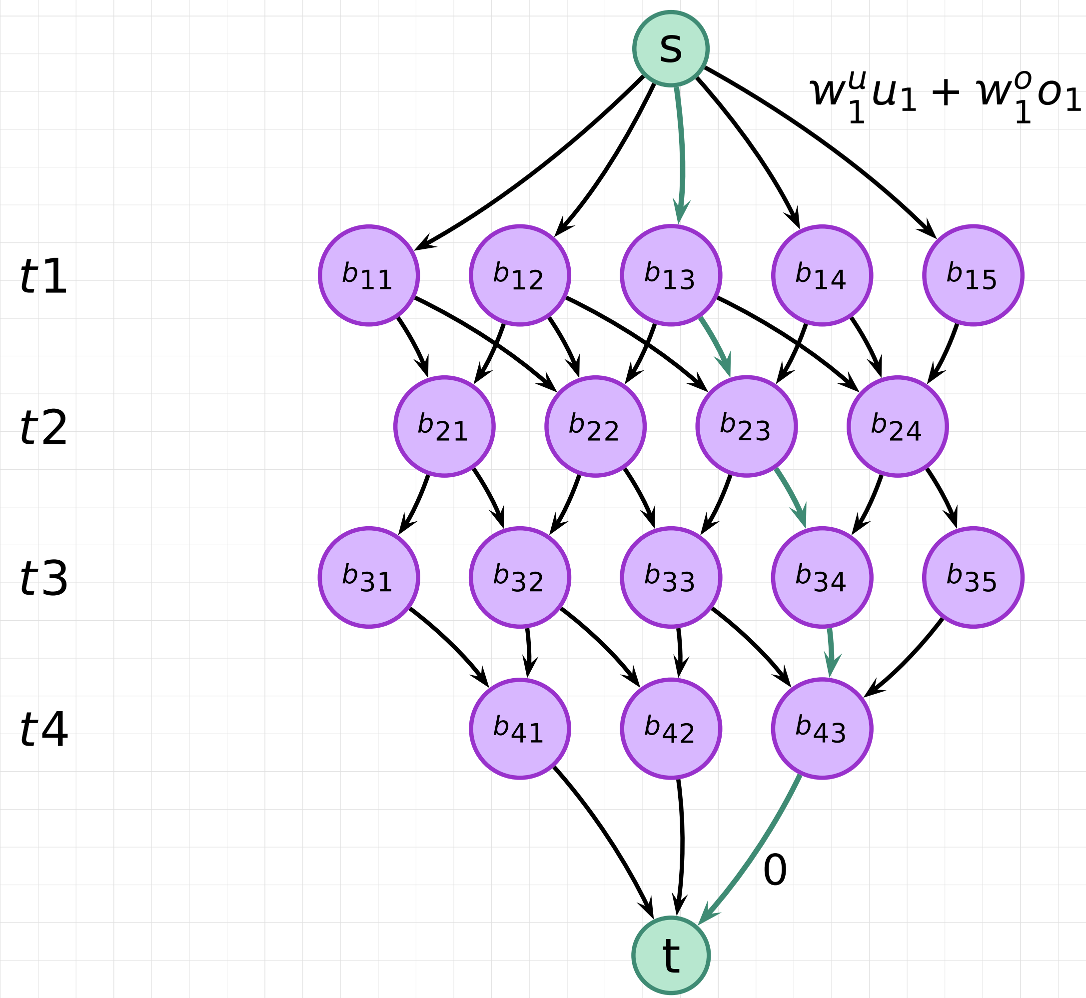
separation by (single) resource constrained longest path



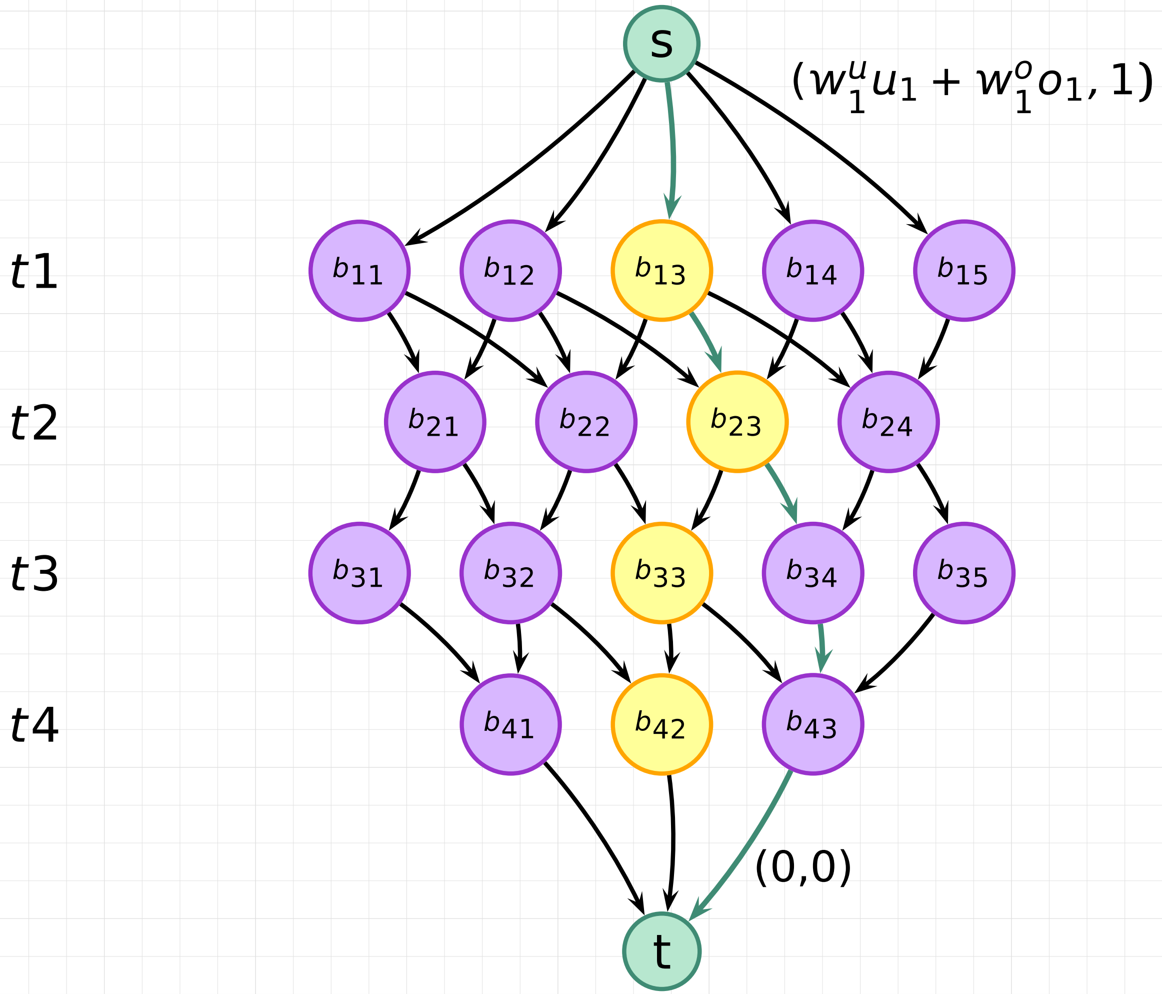
separation by (single) resource constrained longest path



separation by (single) resource constrained longest path



separation by (single) resource constrained longest path



the MIP separation problem

$$\max_{\mathbf{b} \in U} \quad \min \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = b_t - \sum_{j \in J} a_{tj} \bar{x}_j \quad t \in T$$

$$\mathbf{o}, \mathbf{u} \in \mathbb{R}_+^{|T|}$$

the MIP separation problem

$$\max \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = b_t - \sum_{j \in J} a_{tj} \bar{x}_j \qquad t \in T$$

$$o_t \leq M \alpha_t \qquad t \in T$$

$$u_t \leq M(1 - \alpha_t) \qquad t \in T$$

$$\mathbf{b} \in U,\ \mathbf{o}, \mathbf{u} \in \mathbb{R}_+^{|T|},\ \boldsymbol{\alpha} \in \{0, 1\}^{|T|}$$

the MIP separation problem

$$\max \sum_{t \in T} (w_t^o o_t + w_t^u u_t)$$

$$u_t - o_t = b_t - \sum_{j \in J} a_{tj} \bar{x}_j \qquad t \in T$$

$$o_t \leq M\alpha_t \qquad t \in T$$

$$u_t \leq M(1-\alpha_t) \qquad t \in T$$

$$\mathbf{b} \in U,\ \mathbf{o}, \mathbf{u} \in \mathbb{R}_+^{|T|},\ \boldsymbol{\alpha} \in \{0,1\}^{|T|}$$

MIP vs RCSP

- MIP vs dynamic programming
- $O(|T|)$ right-hand-sides vs $O(\bar{D}^2|T|)$ edge costs to update
- MIP solver can be stopped before optimality

istances

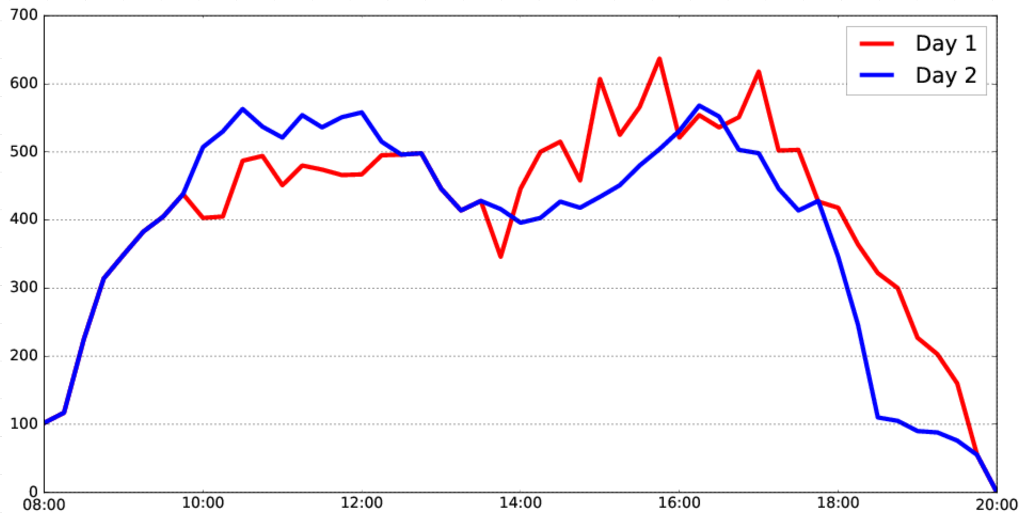
- call center of an Italian Public Agency (2008)

istances

- call center of an Italian Public Agency (2008)
- 49 time slots of 15 minutes, 4-,6- and 8-hour shifts

istances

- call center of an Italian Public Agency (2008)
- 49 time slots of 15 minutes, 4-,6- and 8-hour shifts
- two sampling days
  - Day1: working day
  - Day2: day before a holiday



istances

$$
\begin{aligned}
U \quad = \quad & \Big\{ \mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t, \ t \in T \\
& \boldsymbol{\pi} \in \{0, 1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \leq \pi_t, t \in T \\
& \sum_{t \in T} \pi_t \leq \Gamma \\
& |D_t p_t - D_{t-1} p_{t-1}| \leq \Delta_t, \ t \in T \setminus \{1\} \Big\}
\end{aligned}
$$

istances

$$
\begin{aligned}
U \quad = \quad & \Big\{\mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T \\
& \boldsymbol{\pi} \in \{0,1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \leq \pi_t, t \in T \\
& \sum_{t \in T} \pi_t \leq \Gamma \\
& |D_t p_t - D_{t-1} p_{t-1}| \leq \Delta_t,\ t \in T \setminus \{1\}\Big\}
\end{aligned}
$$

- $\Gamma \Rightarrow$ cardinality
- $\Delta \Rightarrow$ correlation
- $D \Rightarrow$ confidence in staffing values

istances

$$
\begin{aligned}
U \quad = \quad & \Big\{\mathbf{b} \in \mathbb{Z}^{|T|} : b_t = \tilde{b}_t + D_t p_t,\ t \in T \\
& \boldsymbol{\pi} \in \{0,1\}^{|T|}, \mathbf{p} \in \mathbb{R}^{|T|}, |p_t| \le \pi_t, t \in T \\
& \sum_{t \in T} \pi_t \le \Gamma \\
& |D_t p_t - D_{t-1} p_{t-1}| \le \Delta_t,\ t \in T \setminus \{1\}\Big\}
\end{aligned}
$$

- six Γ values $\Rightarrow$ cardinality
- two Δ values $\Rightarrow$ correlation
- three D values $\Rightarrow$ confidence

branch-and-cut statistics

	max	avg	min
time	1420	191	6
nodes	813	132	1

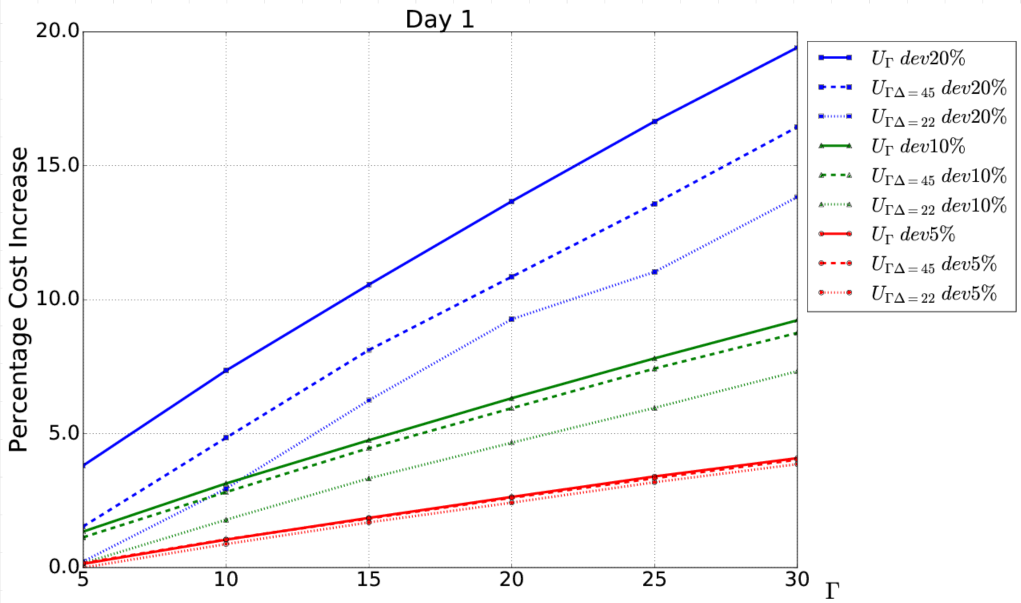
branch-and-cut statistics

	max	avg	min
time	1420	191	6
nodes	813	132	1
gap at root	0.57%	0.12%	0.02%

branch-and-cut statistics

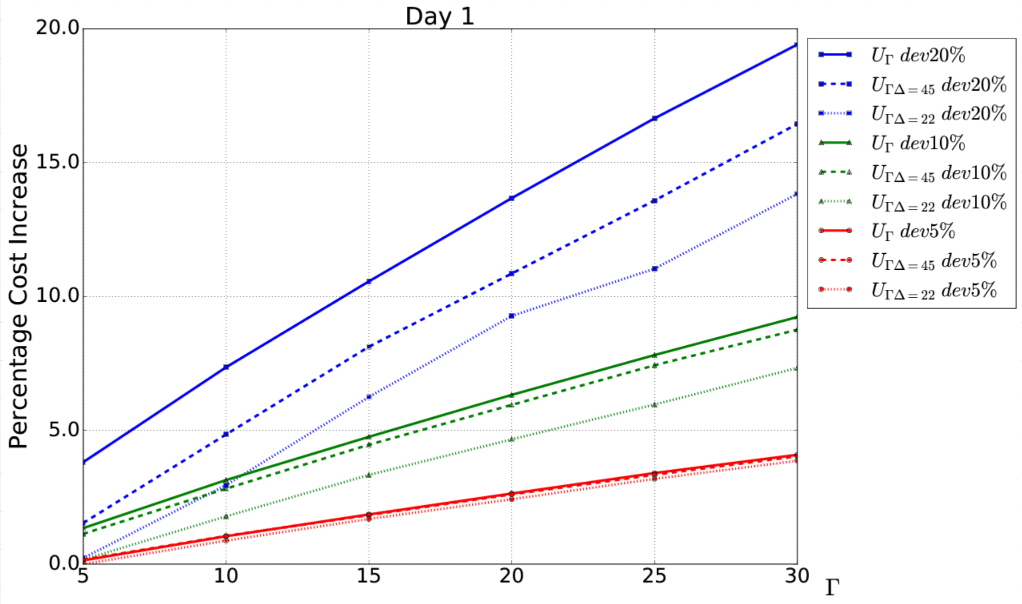
	max	avg	min
time	1420	191	6
nodes	813	132	1
gap at root	0.57%	0.12%	0.02%
sep time	1400	185	5
cuts	199	128	89

correlation vs SIPP model



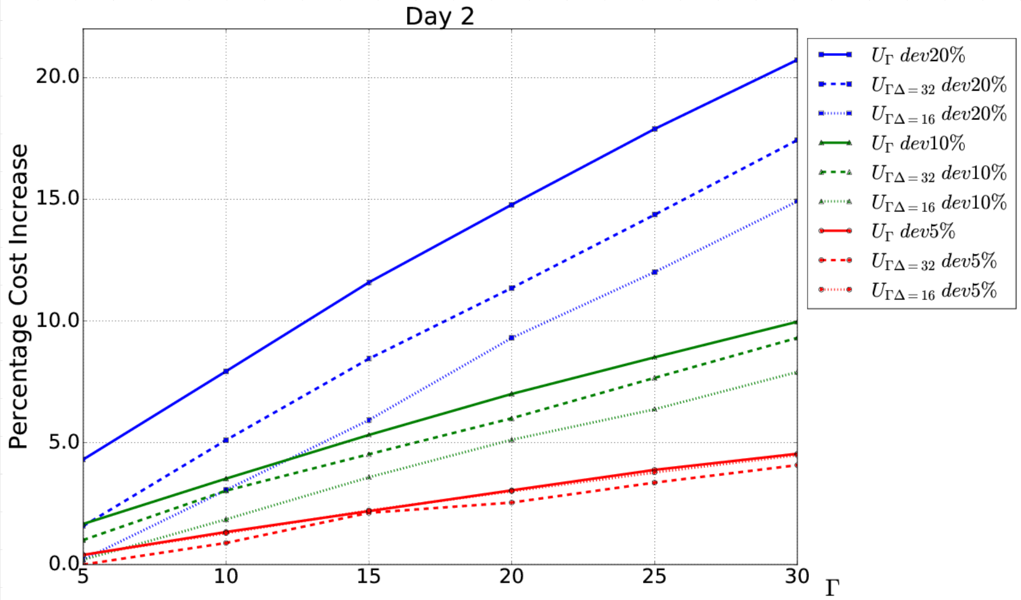
- solid lines ⇒ only cardinality
  dotted and dashed lines ⇒ cardinality plus correlation
- different line styles ⇒ different Δ values
- different colors ⇒ different D values

correlation vs SIPP model



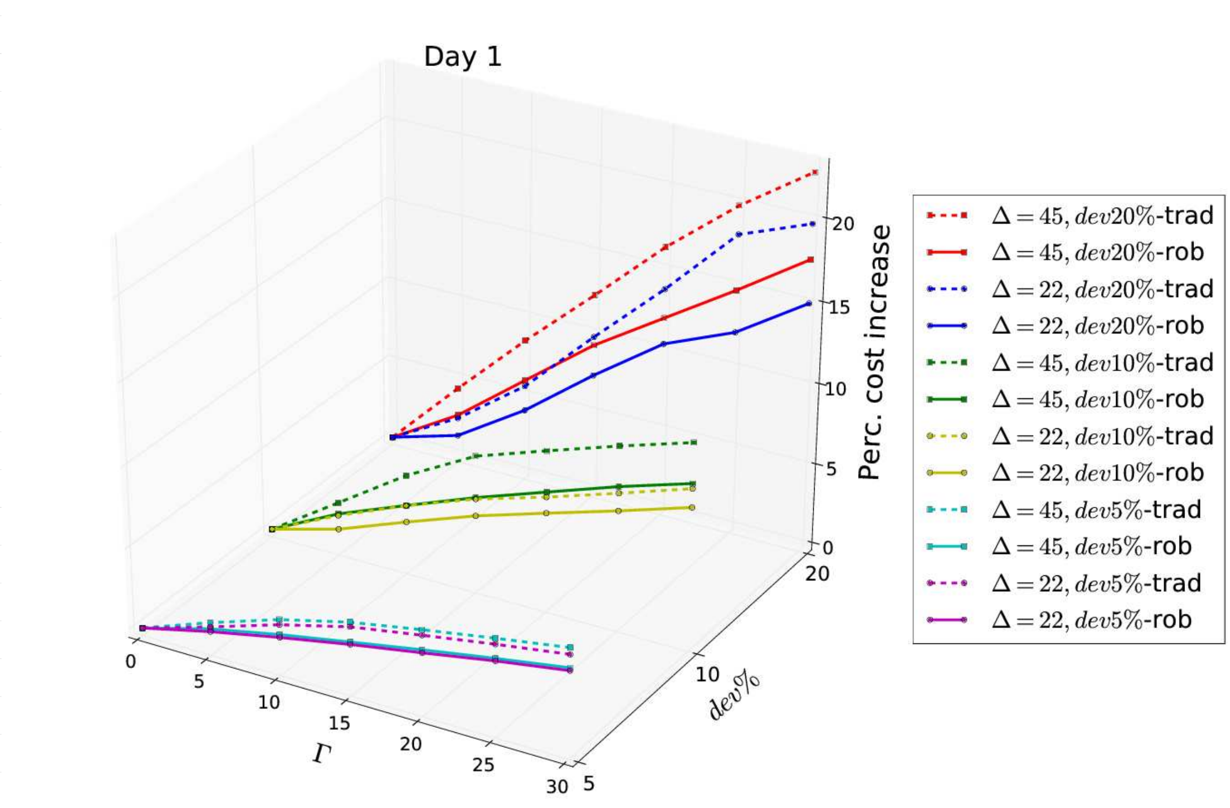
- no correlation ⇒ larger costs
- maximum difference is about 5% (5000 €, per day)

correlation vs SIPP model



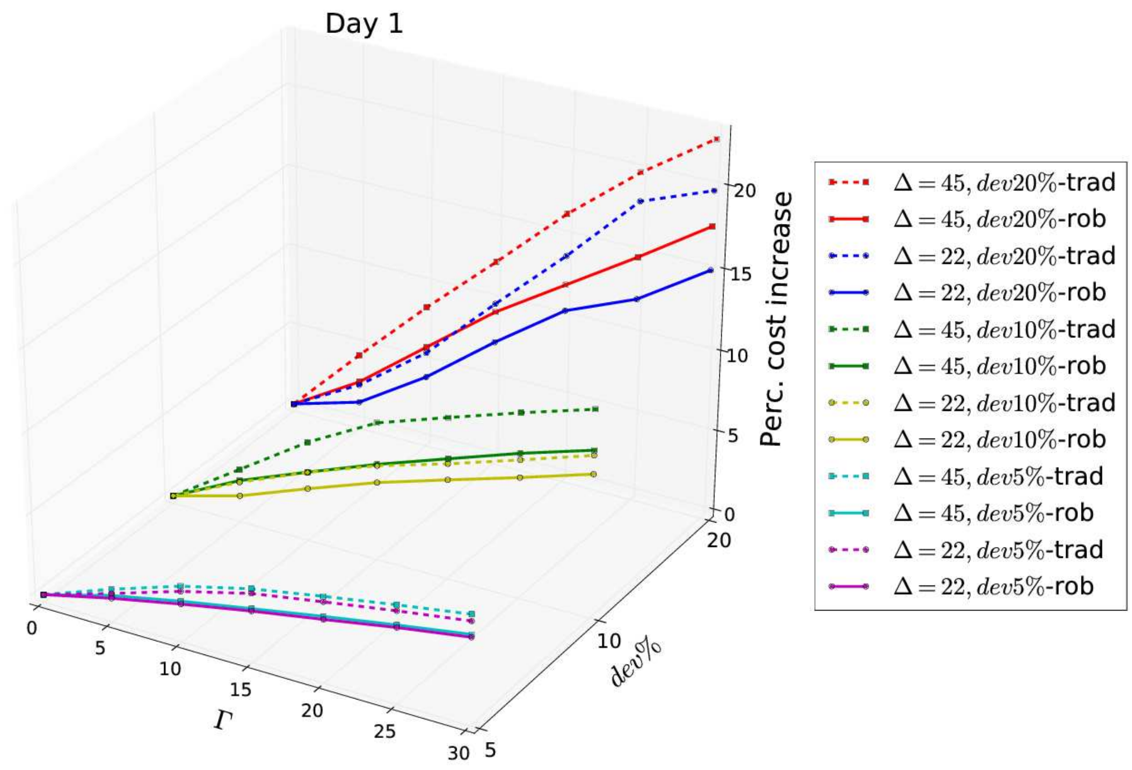
- no correlation ⇒ larger costs
- maximum difference is about 5% (5000 €, per day)

optimization vs managers: costs



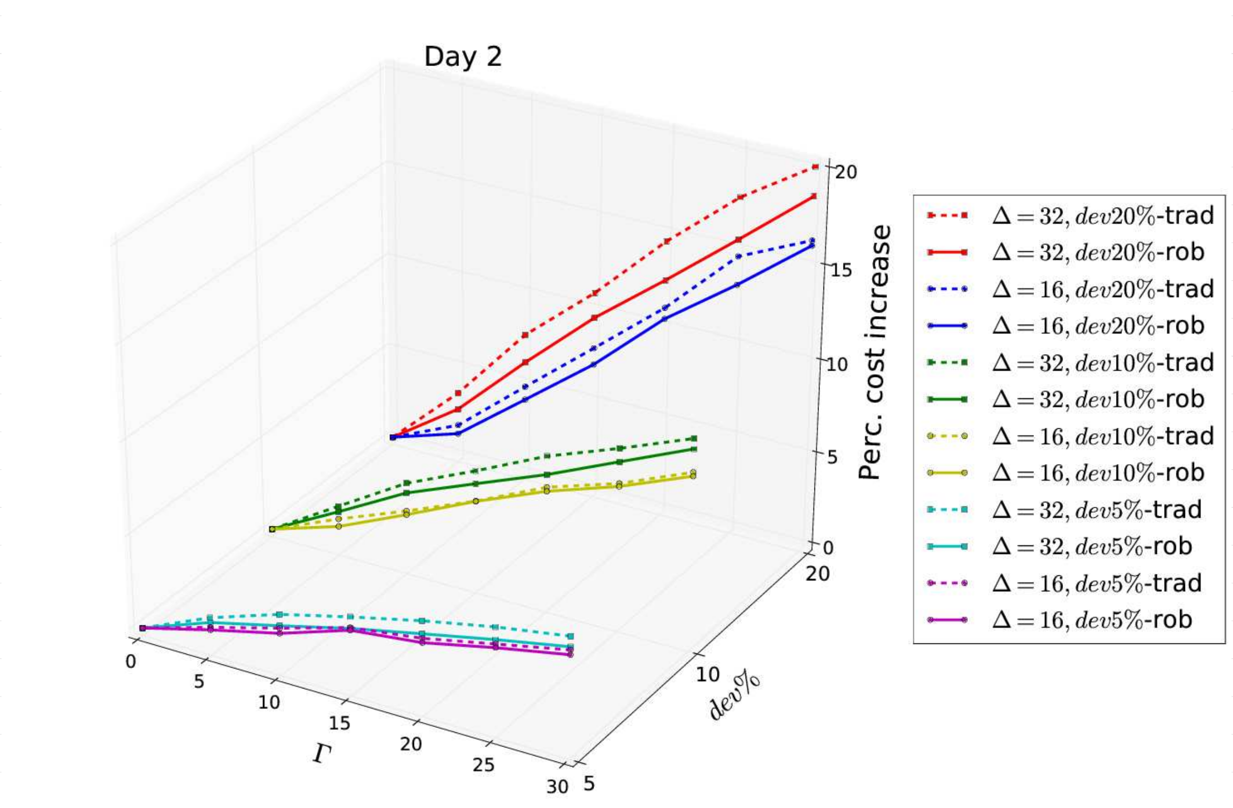
- solid lines ⇒ optimization
  dotted lines ⇒ managers
- same group ⇒ same D
  same color ⇒ same Δ

optimization vs managers: costs



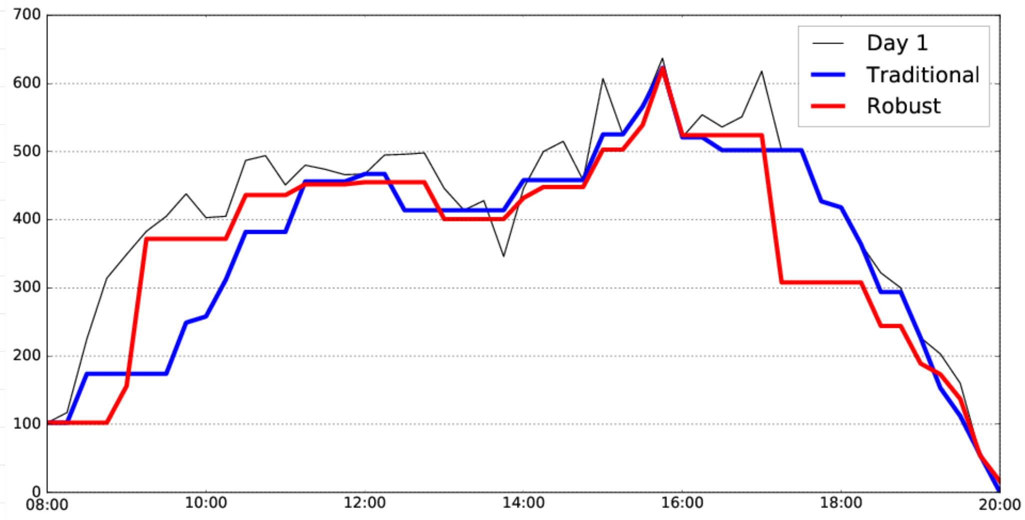
- optimization ⇒ smaller costs
- maximum difference is about 5% (5000 €, per day)

optimization vs managers: costs



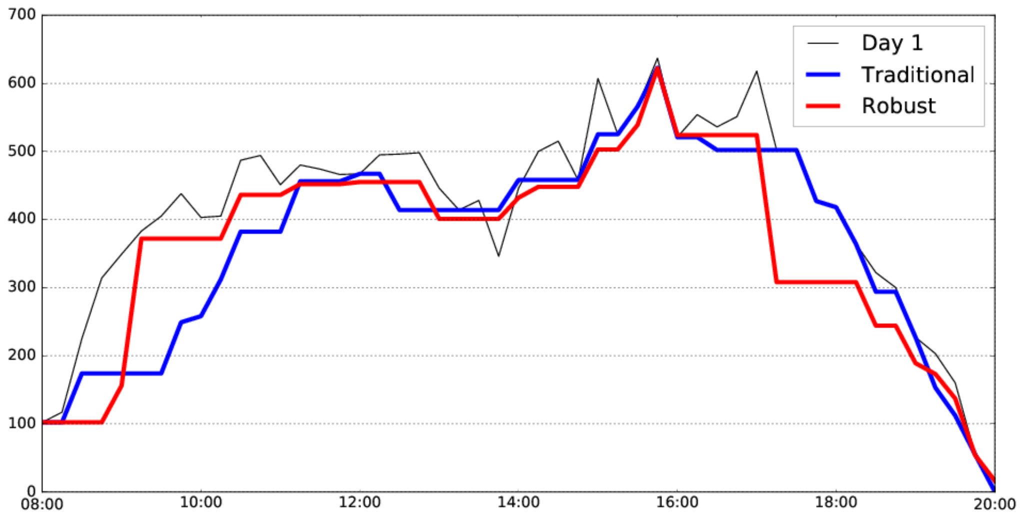
- optimization ⇒ smaller costs
- costs are closer (2% of difference)

optimization vs managers: staffing



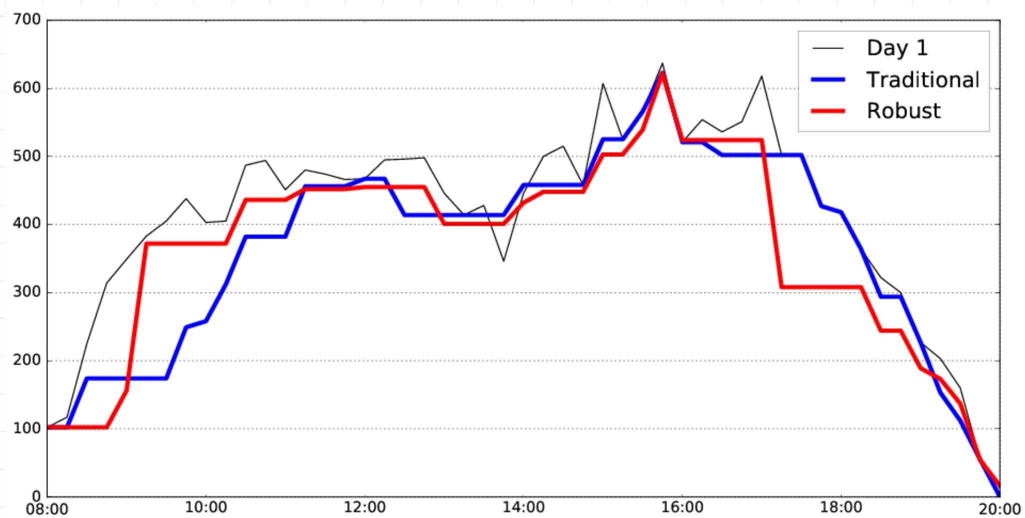
- black line ⇒ demand **b**
- blue line ⇒ staffing **x** according to managers practice
- red line ⇒ staffing **x** computed by optimization

optimization vs managers: staffing



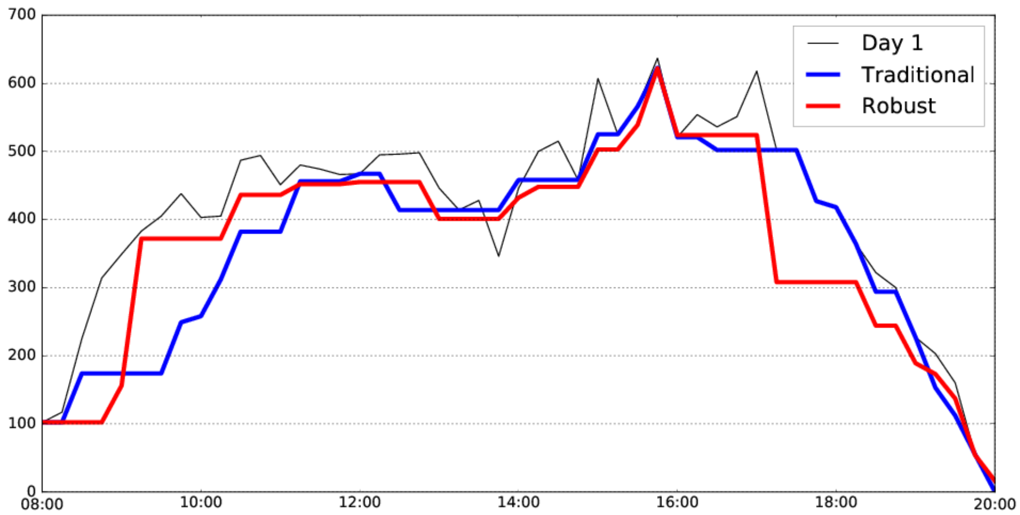
- many **x** with similar cost, but different robustness

optimization vs managers: staffing



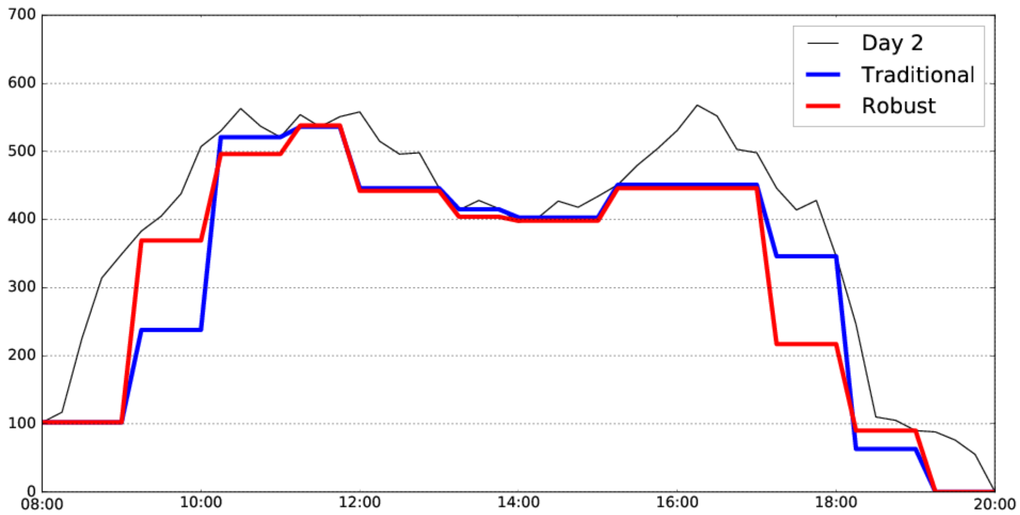
- many **x** with similar cost, but different robustness
- both methods prefer understaffing vs overstaffing

optimization vs managers: staffing



- many **x** with similar cost, but different robustness
- both methods prefer understaffing vs overstaffing
- managers follow the demand variation
- optimization anticipates the demand variation

optimization vs managers: staffing



- many **x** with similar cost, but different robustness
- both methods prefer understaffing vs overstaffing
- managers follow the demand variation
- optimization anticipates the demand

paper

S. Mattia, F. Rossi, M. Servilio, S. Smriglio.
Staffing and scheduling flexible call centers by two-stage robust optimization. Omega, in press.
`http://dx.doi.org/10.1016/j.omega.2016.11.001`

thank you