Interpretable Dynamics Models

For Data-Efficient Reinforcement Learning

Markus Kaiser, Clemens Otte, Thomas A. Runkler, Carl Henrik Ek
markus.kaiser@siemens.com

April 24, 2019

Siemens AG, Technical University of Munich, University of Bristol

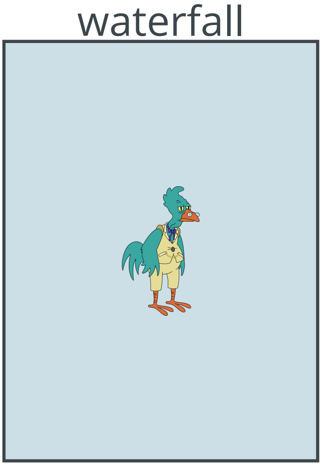
Wet-Chicken Benchmark[1]

[1]Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]

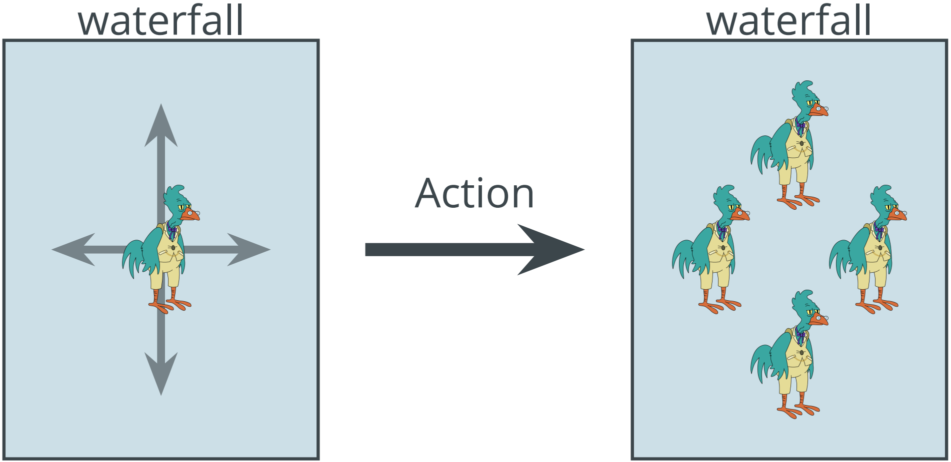
[1] Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]



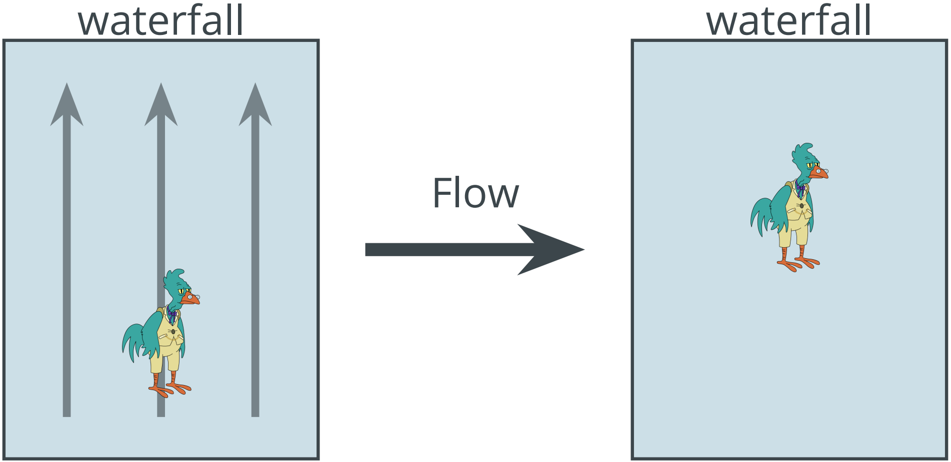
[1] Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]



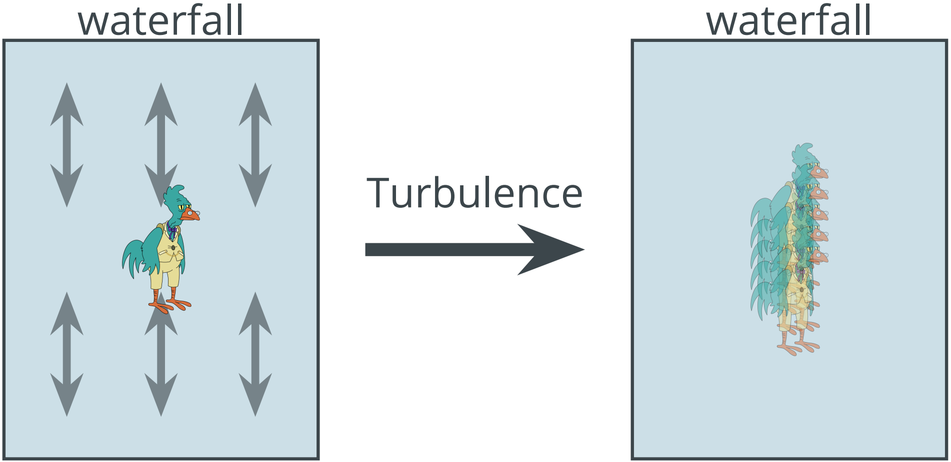
[1] Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]



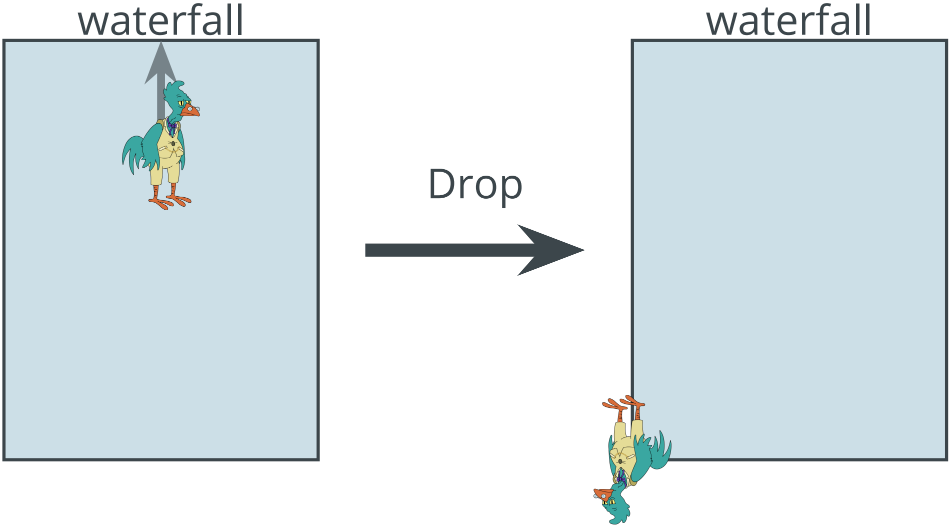
[1] Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]



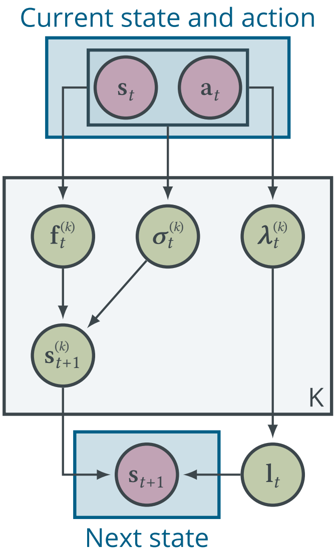
[1]Tresp 1994; Hans and Udluft 2009.

Wet-Chicken Benchmark[1]

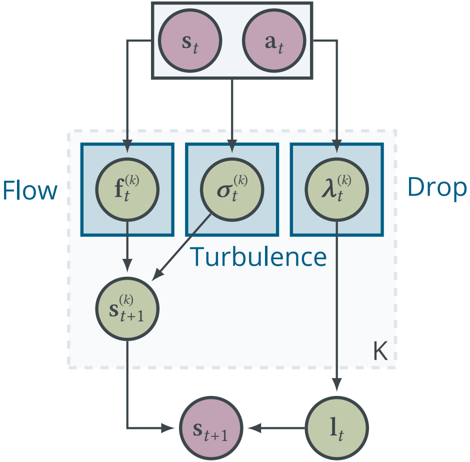


[1] Tresp 1994; Hans and Udluft 2009.

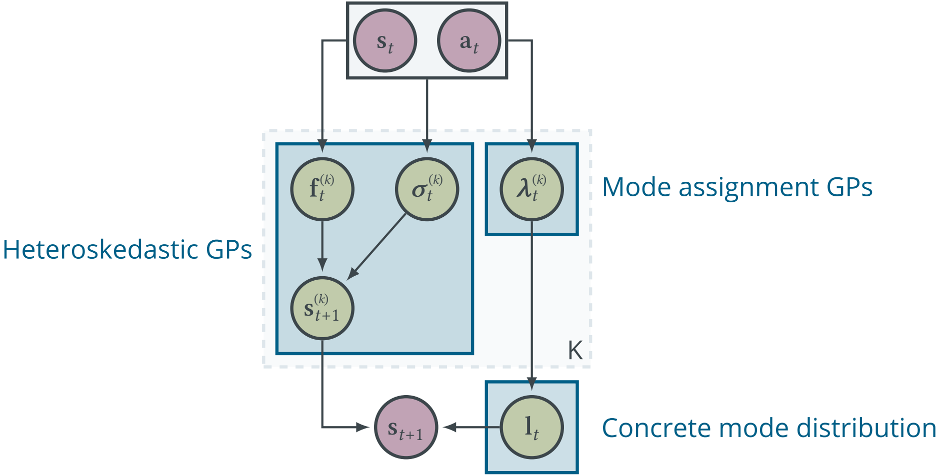
Dynamics: Graphical Model



Dynamics: Graphical Model



Dynamics: Graphical Model



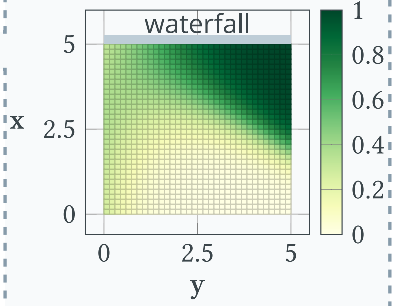
Dynamics: Posterior

$$\begin{aligned} \mathrm{p}(\Delta\mathbf{x}_{t+1}) = \; & \mathrm{p}(\Delta\mathbf{x}_{t+1}|\text{drop}) \cdot \mathrm{p}(\text{drop}) \\ & + \mathrm{p}(\Delta\mathbf{x}_{t+1}|\text{no drop}) \cdot \mathrm{p}(\text{no drop}) \end{aligned}$$

Dynamics: Posterior

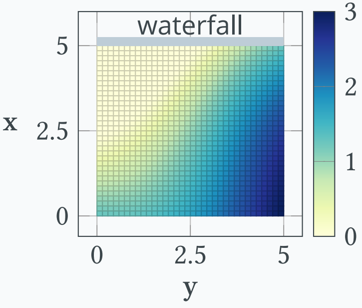
$$p(\Delta \mathbf{x}_{t+1}) = p(\Delta \mathbf{x}_{t+1} \mid \text{drop}) \cdot p(\text{drop}) + p(\Delta \mathbf{x}_{t+1} \mid \text{no drop}) \cdot p(\text{no drop})$$

p(drop)



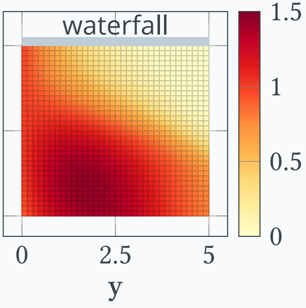
Flow

$\mathbb{E}[\Delta \mathbf{x}_{t+1} \mid \text{no drop}]$

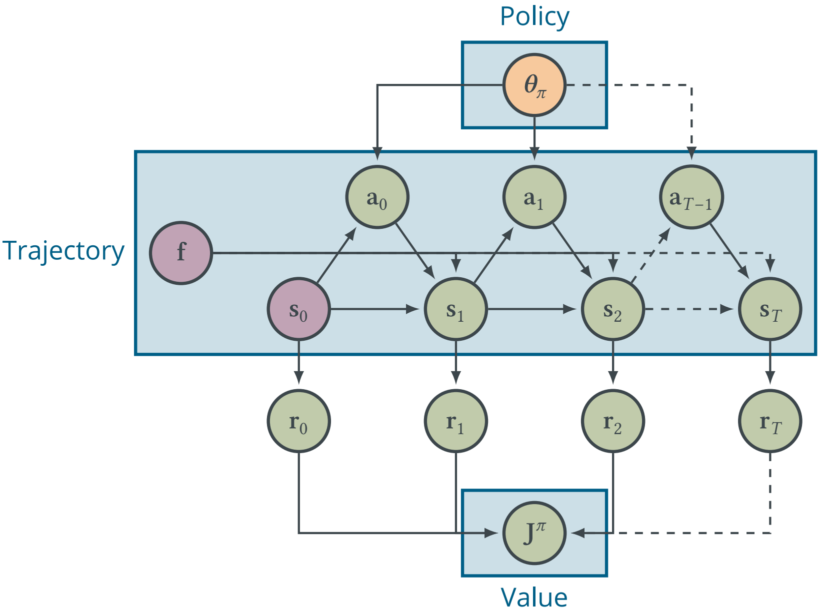


Turbulence

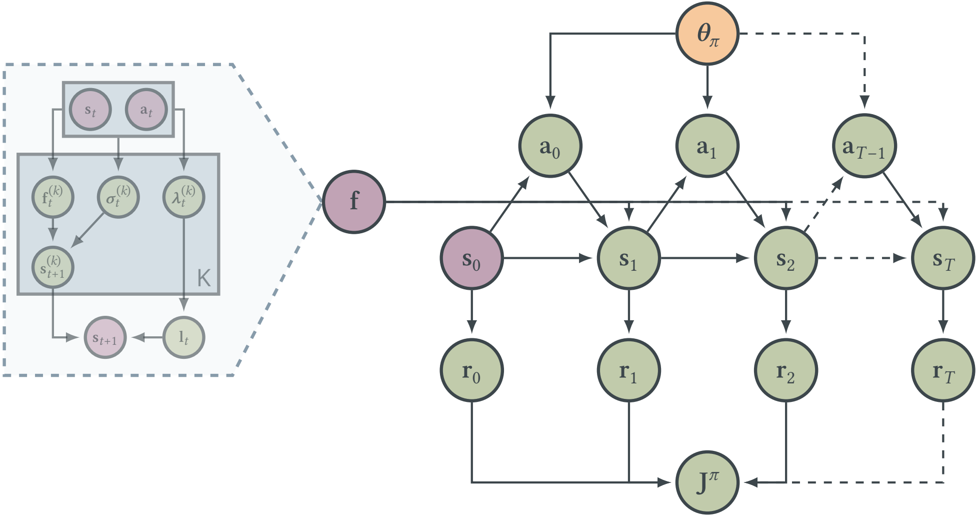
$\sigma[\Delta \mathbf{x}_{t+1} \mid \text{no drop}]$



Policy: Graphical Model



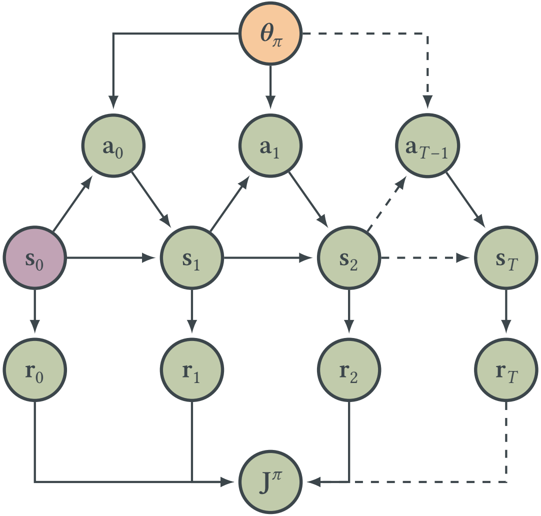
Policy: Graphical Model



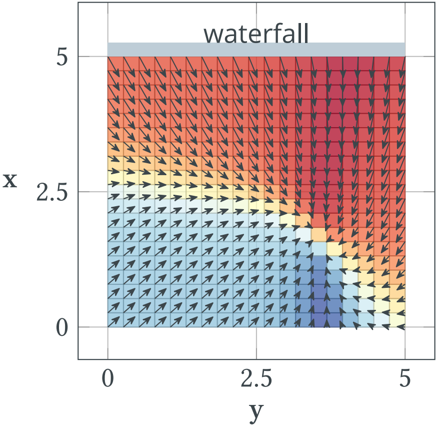
$$\mathbb{E}[\mathbf{J}^{\pi}(\boldsymbol{\theta}_{\pi})] = \sum_{t=0}^{T} \gamma^{t}\, \mathbb{E}_{\mathrm{p}(\mathbf{s}_t|\boldsymbol{\theta}_{\pi})}[\mathbf{r}_t]$$

$$\approx \frac{1}{P} \sum_{p=1}^{P} \sum_{t=0}^{T} \gamma^{t} r_t^{(p)}$$

$$\nabla \mathbf{J}^{\pi}(\theta_{\pi}) \approx \frac{1}{P} \sum_{p=1}^{P} \sum_{t=0}^{T} \gamma^{t} \nabla_{\theta_{\pi}} r_t^{(p)}$$



Policy: Posterior

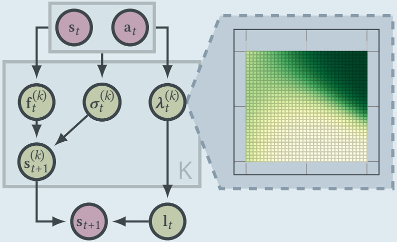


N	NFQ[2]	GP[3]	Ours
100	0.66(16)	**1.41(1)**	1.18(9)
250	1.71(7)	1.54(1)	**2.33(1)**
500	1.60(10)	1.56(1)	**2.25(1)**
1000	1.99(6)	2.13(1)	**2.32(1)**
2500	2.26(2)	1.91(1)	**2.28(1)**
5000	**2.33(1)**	1.91(1)	2.28(1)

[2] Riedmiller 2005
[3] Deisenroth and Rasmussen 2011

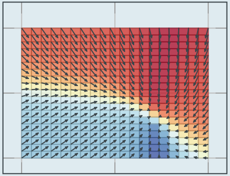
Interaction with domain experts

- Incomplete system knowledge
- Hierarchical priors
- Interpretable sub-models



Trustworthy decision making

- Uncertainty due to incomplete data
- Stochastic systems
- Robust and efficient inference



References

- Deisenroth, Marc and Carl E. Rasmussen (2011). “PILCO: A Model-Based and Data-Efficient Approach to Policy Search”. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 465–472.
- Hans, Alexander and Steffen Udluft (2009). “Efficient Uncertainty Propagation for Reinforcement Learning with Limited Data”. In: *International Conference on Artificial Neural Networks*. Springer, pp. 70–79.
- Kaiser, Markus et al. (Oct. 16, 2018). “Data Association with Gaussian Processes”. In: arXiv: `1810.07158 [cs, stat]`.
- Riedmiller, Martin (2005). “Neural Fitted Q Iteration - First Experiences with a Data Efficient Neural Reinforcement Learning Method”. In: *European Conference on Machine Learning*. Springer, pp. 317–328.
- Tresp, Volker (1994). “The Wet Game of Chicken”. In: *Siemens AG, CT IC 4, Technical Report*.

Variational Bound

$$
\begin{aligned}
\mathscr{L}_{\text{DAGP}} &= \mathbb{E}_{\mathrm{q}(\mathbf{F},\boldsymbol{\lambda},\mathbf{U})}\left[\log \frac{\mathrm{p}(\mathbf{S}', \mathbf{L}, \mathbf{F}, \boldsymbol{\lambda}, \mathbf{U} \,|\, \mathbf{S})}{\mathrm{q}(\mathbf{F}, \boldsymbol{\lambda}, \mathbf{U})}\right] \\
&= \sum_{n=1}^{N} \mathbb{E}_{\mathrm{q}(\mathbf{f_n})}[\log \mathrm{p}(\mathbf{s}'_{\mathbf{n}} \,|\, \mathbf{f_n}, \mathbf{l_n})] + \sum_{n=1}^{N} \mathbb{E}_{\mathrm{q}(\boldsymbol{\lambda}_{\mathbf{n}})}[\log \mathrm{p}(\mathbf{l_n} \,|\, \boldsymbol{\lambda}_{\mathbf{n}})] \\
&\quad - \sum_{k=1}^{K} \mathrm{KL}\left(\mathrm{q}\left(\mathbf{u}^{(\mathbf{k})}\right) \,\middle\|\, \mathrm{p}\left(\mathbf{u}^{(\mathbf{k})} \,\middle|\, \mathbf{Z}^{(\mathbf{k})}\right)\right) - \sum_{k=1}^{K} \mathrm{KL}\left(\mathrm{q}\left(\mathbf{u}_{\boldsymbol{\lambda}}^{(\mathbf{k})}\right) \,\middle\|\, \mathrm{p}\left(\mathbf{u}_{\boldsymbol{\lambda}}^{(\mathbf{k})} \,\middle|\, \mathbf{Z}_{\boldsymbol{\lambda}}^{(\mathbf{k})}\right)\right)
\end{aligned}
$$

Predictive Posterior

$$
\begin{aligned}
\mathrm{q}(\mathbf{s}_{t+1} \,|\, \mathbf{s}_t) &= \int \sum_{k=1}^{K} \mathrm{q}\left(\mathbf{l}_t^{(k)} \,\middle|\, \mathbf{s}_t\right) \mathrm{q}\left(\mathbf{s}_{t+1}^{(k)} \,\middle|\, \mathbf{s}_t\right) \mathrm{d}\mathbf{l}_t^{(1)} \ldots \mathrm{d}\mathbf{l}_t^{(K)} \\
&\approx \sum_{k=1}^{K} \hat{l}_t^{(k)} \hat{\mathbf{s}}_{t+1}^{(k)}
\end{aligned}
$$