Optimal policies for Bayesian olfactory search in a turbulent flow

R. A. Heinonen[1], F. Bonaccorso[1], L. Biferale[1], A. Celani[2], and M. Vergassola[3]

[1]Dept. Physics and INFN, University of Rome, "Tor Vergata"
[2]The Abdus Salam International Center for Theoretical Physics
[3]Dept. Physics, Ecole Normale Supérieure

Supported by the European Research Council under grant No. 882340

erc European Research Council
Established by the European Commission
Supporting top researchers from anywhere in the world

Olfactory search problem

Introduction: searching for an odor source

- Insects often need find source (usually upwind) of an odor or other cue advected by the atmosphere
- E.g. mosquito drawn to human by CO_2; moth drawn to mate by pheromones
- Source may be ~ 100 m away(!)
- N.B. also applications to aquatic animals, robotics

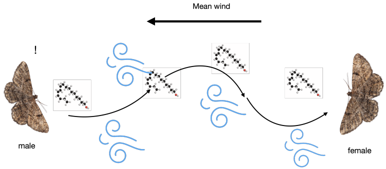


Figure Artist's conception of a moth searching for a mate via pheromone cues.

The effect of turbulence

- Classical search strategy is chemotaxis, i.e. just go up the concentration gradient
- But: turbulence mixes cue into **stochastic, intermittent** landscape. Gradient estimation is unfeasible

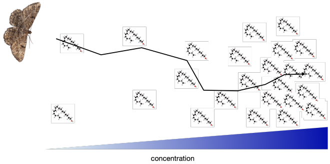


Figure Artist's conception of chemotaxis strategy.

Figure A turbulent environment leads to a patchy odor landscape with intermittent detections.

Concentration intermittency from experiment

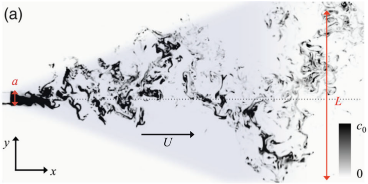


Figure Concentration field from jet flow experiment [Villermaux and Innocenti, 1999]. Fig taken from [Celani et al., 2014]

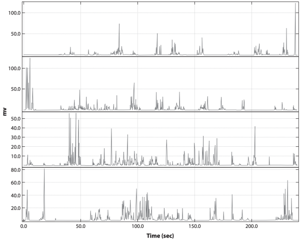


Figure Time series from experiment showing concentration signal 50 m from a propylene source over 16 minutes. From [Yee et al., 1993]

Real moth trajectory

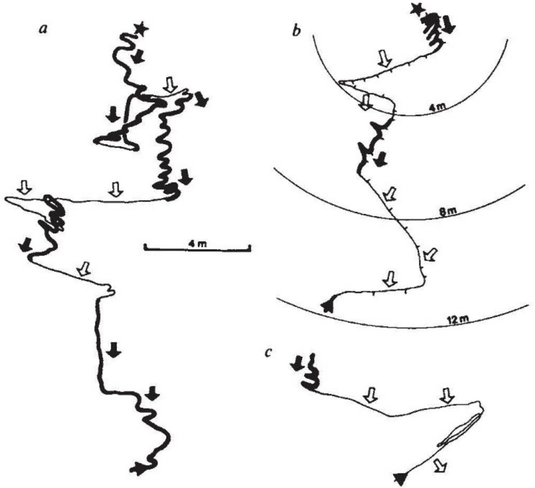


Figure Trajectory of gypsy moth from experiment [David et al., 1983] as it tracks sex pheromone source, showing upwind surging when in the plume and crosswind casting when out of the plume

A first heuristic: cast-and-surge

- [Balkovsky and Shraiman, 2002] introduced "cast-and-surge" heuristic policy based on observed insect behavior
- Agent has internal clock τ that counts timesteps since last detection
- Agent zigzags toward the source, with length of crosswind excursions increasing with τ
- Model-free approach (no knowledge of the statistics)

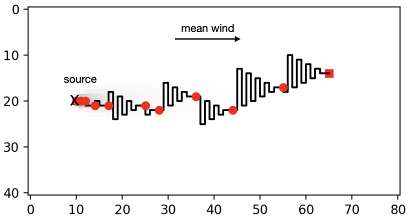


Figure Heuristic cast-and-surge searching in toy environment based on [Balkovsky and Shraiman, 2002]

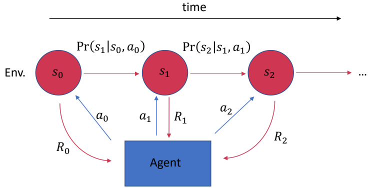
“Optimal” policies?

- Cast-and-surge has good qualitative performance, but one can certainly do better. What is best?
- Idea of this work: what strategy *minimizes* the time of arrival?
- To define this, need some background....

POMDP and optimal policies

Markov decision processes

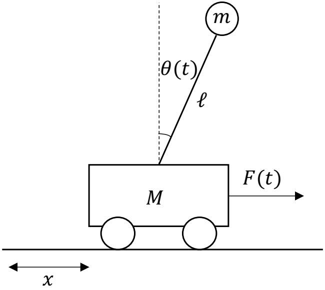
- Agent interacts with environment by taking actions $a \in A$ at each t_i
- Relevant information about system at t_i captured by state $s \in S$. State evolves according to $\Pr(s'|s, a)$
- By assumption: transitions enjoy Markov property. (N.B. extending state to $\tilde{s}_t = \{s_t, s_{t-1}, \ldots, s_{t-k}\}$ captures finite-time memory)
- Agent receives reward R according to $\Pr(R|s, s', a)$ (can be < 0)
- Goal: craft policy $\pi : s \mapsto a$ maximizing $\mathbb{E}_\pi[\sum_{t=0}^{\infty} \gamma^t R_t]$, $0 < \gamma \leq 1$



MDP example: inverted pendulum

- State is $\{\theta, \dot{\theta}, x, \dot{x}\}$. Evolves according to EOM
- Actions: apply over Δt some voltage $V \in [-V_{\max}, V_{\max}]$ to a motor, induces F
- Goal: $\theta \to 0$, minimize power output $\mathcal{P}$
- Motivates reward

$$R_t = -\theta(t)^2 - a\dot{\theta}(t)^2 - b\mathcal{P}(t),\ a, b > 0$$



Partial observability

- In practice, we don't always have access to the state (in fact, we usually don't!)
- Suppose in previous example we only measure $\mathbf{s} = [\theta, \dot{\theta}, x, \dot{x}]$ with uncertainties σ (say Gaussian, uncorrelated)
- System is now *partially observable*
- Measurements are now *observations* $o \in O$, supply *information* about true states s thru likelihood $\Pr(o|s, a)$
- In this example, $\Pr(\mathbf{o}|\mathbf{s}) \propto \prod_i \exp\left[-(o_i - s_i)^2/2\sigma_i^2\right]$

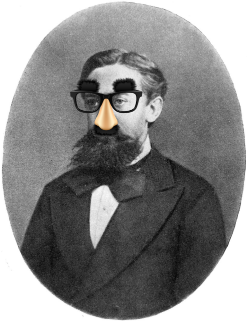


Figure Partially observable Markov

Bayesian inference

- At timestep t, agent has history $(a_1, o_1, a_2, o_2, \ldots, a_{t-1}, o_t)$. What does this say about state?
- Assuming system is Markovian, information can be stored in a probability distribution ("belief") b over $\mathbf{s}$
- Update b after taking a and observing o using Bayes' theorem

$$b(s')_{o,a} = \mathrm{Pr}(o|s', a) \sum_{s} b(s)\mathrm{Pr}(s'|s, a)/Z$$

- *Model-based* approach — need $\mathrm{Pr}(o|s', a)$
- Goal: seek policy $\pi : b \mapsto a$ which maximizes reward

POMDP example: Bernoulli bandits

- Classic decision problem: k slot machines. Each pays out unit reward with unknown probabilities p_i
- Which sequence of levers to pull to maximize total (discounted) reward?
- Tradeoff between exploration (discover the p_i) and exploitation (reap rewards)
- As POMDP: static state $s = \{p_1, \ldots, p_k\}$, actions $a \in \{1, \ldots, k\}$, observations $o_t = R_t \in \{0, 1\}$
- $\Pr(o = 1 | s, a = j) = p_j$; $\Pr(o = 0 | s, a = j) = 1 - p_j$
- N.B. optimal policy can be specified exactly (Gittins 1974)

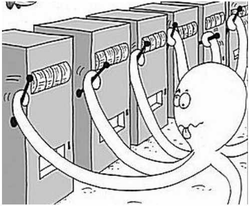


Figure Artist's conception of a multi-armed bandit agent

Optimal POMDP planning: Bellman equation

- Define value function $V_\pi(b)$ as total expected reward under π conditioned on b (assume $R = R(s,a)$):

$$V_\pi(b) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \sum_{s \in S} \gamma^t b_t(s) R(s, \pi(b_t)) \middle| b_0 = b \right]$$

- Optimal value function satisfies Bellman equation

$$V^*(b) = \max_{a \in A} \left[\underbrace{\sum_{s \in S} R(s,a) b(s)}_{\text{immediate expected reward}} + \gamma \underbrace{\sum_{o \in O} \Pr(o|b,a) V^*(b_{o,a})}_{\text{future expected rewards}} \right]$$

- In principle, can be solved exactly, but partial observability makes solution computationally hard. Belief simplex very large (dimension $|S| - 1$)! "Curse of dimensionality"
- Knowing $V^*(b)$ instantly gives you π^*

Olfactory search POMDP

Model search problem

- State: relative position of agent w.r.t. source (unknown) $\mathbf{s} = \mathbf{r} - \mathbf{r}_0$
- Agent makes observation (detection or nondetection) then moves. Assume a strong swimmer (no advection by the flow)
- Try to reach source in as few Δt as possible — give reward γ^T for reaching source in T steps ($0 < \gamma < 1$)
- Key physics input is $\Pr(o|\mathbf{s})$

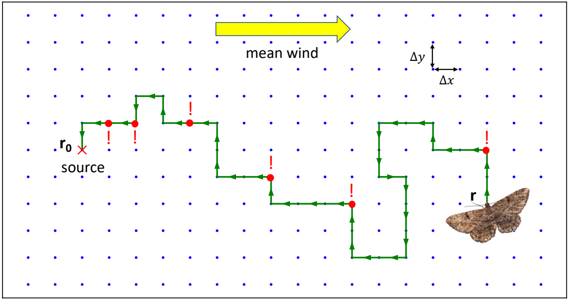
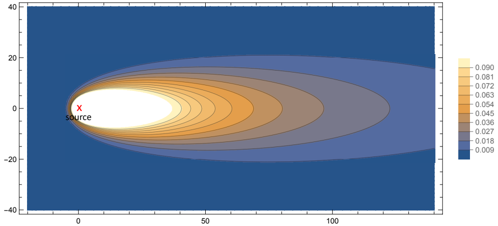


Figure In our setup, agent lives on the gridworld (blue points) and tries to find the source (red ×)

Diffusive model of environment



Advection-diffusion eq.

$$\partial_t c + \underbrace{V}_{\text{mean wind}} \partial_x c = \underbrace{D\nabla^2 c}_{\text{turb. diffusion}} + \underbrace{R\delta(\mathbf{x})}_{\text{point source}} - \underbrace{c/\tau}_{\text{turb. mixing time}} ,$$

stationary solution $+\ 4\pi a D c$ detections/time $\implies$ detection rate

$$h = \frac{aR}{|\mathbf{x}|} \exp\left(\frac{Vx}{2D} - \frac{|\mathbf{x}|}{\lambda}\right), \; p(\mathrm{obs}|\mathbf{x}) = 1 - e^{-h\Delta t}$$

Infotaxis: an important model-based heuristic

- [Vergassola et al., 2007] suggested a policy that seeks to maximize information content of belief

$$\pi(b) = \arg\min_a \sum_o \mathrm{Pr}(o|b,a) H[b_{o,a}]$$

where $H[b] = -\sum_s b(s) \log b(s)$.

- Prioritizes exploration (seek information about source) over exploitation (use information to move towards source)
- Generally performs extremely well, but can improve by adding information about distance from source [Loisy and Eloy, 2022]

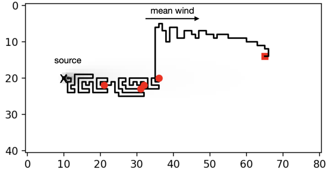


Figure Sample infotaxis trajectory in toy environment.

Optimal policies

- Recent work has demonstrated the present POMDP can be solved effectively using at least three algorithms (Perseus w/ reward shaping, SARSOP, model-based DQN). Can usually beat all available heuristics
    1. Loisy and Eloy *Proc. R. Soc. Lond.* (2022) — DQN in windless setting
    2. RAH, Biferale, Celani, and Vergassola *PRE* (2023) — Perseus in windy setting
    3. Loisy and RAH *EPJE* (2023) — benchmark on Perseus, SARSOP, DQN in windy and windless settings
- But this work done in "toy model" setting (statistics imposed by hand)

Performance of Perseus policies vs. heuristics

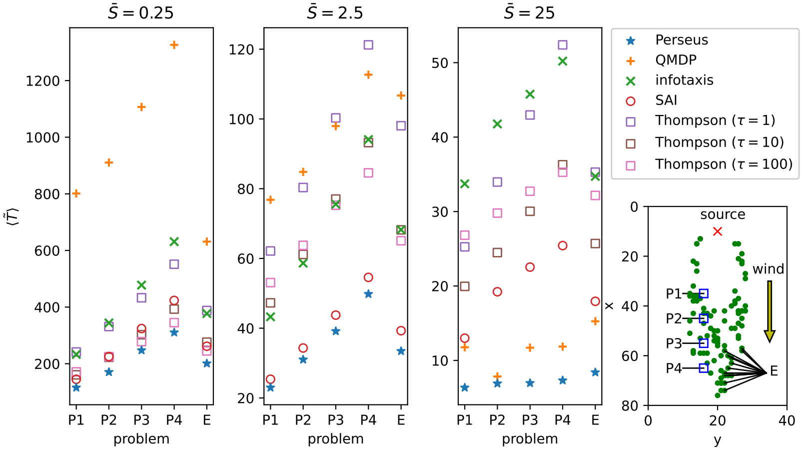
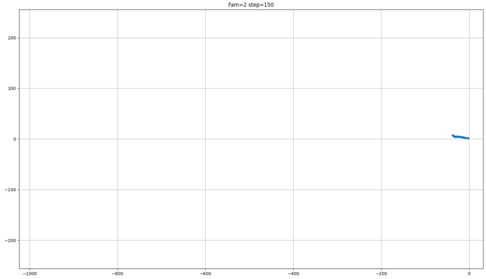


Figure Excess mean arrival times $\langle \tilde{T} \rangle = \langle T \rangle - \langle T_{MDP} \rangle$ for test problems. $\bar{S} = a\Delta tR/\Delta x$ is nondimensional emission rate

Moving to a "real" turbulent flow

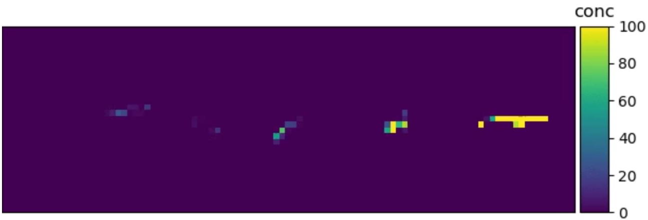
The DNS

- 3-D incompressible Navier-Stokes with mean wind V on $1024 \times 512 \times 512$ grid in turbulent regime $\mathrm{Re}_\lambda \simeq 150$
- Periodic BCs, stochastic large-scale forcing
- Lagrangian particles emitted simultaneously from point sources at 5 locations, data dumped every τ_η ($\sim 4000\tau_\eta$ total)
- Have data for 5 different mean flow speeds ($V/\tilde{v} \simeq 0, 1.5, 3, 6, 9$)



Coarse-graining

- To move to POMDP setting, data are coarse-grained on a quasi-2D slice to obtain 99×33 grid with spacing $\sim 10\eta$
- Particles counted to obtain concentration field



Empirical likelihood

- Define $c_{\rm thr} \gg \langle c | c > 0 \rangle$
- $\Pr(o|\mathbf{s}) \equiv \Pr(c(\mathbf{s}) \geq c_{\rm thr})$ averaged over time and source locations, symmetrized across wind axis
- Use SARSOP to solve for policy using either empirical likelihood or fit to model

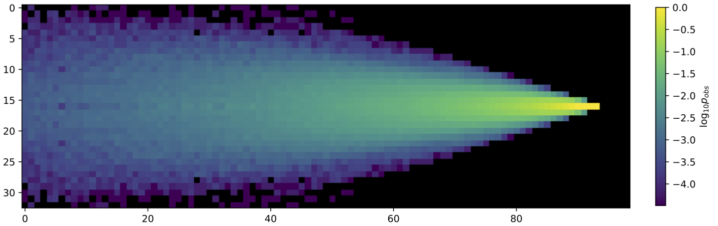
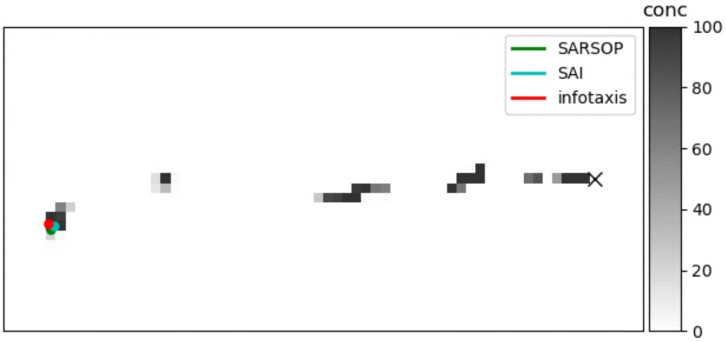


Figure Empirical log10-likelihood of observation for $c_{\rm thr} = 100$, $V/\tilde{v} \simeq 9$

Searching in the DNS: near-optimal vs. heuristics



Arrival time statistics for $V/\tilde{v} \simeq 9$

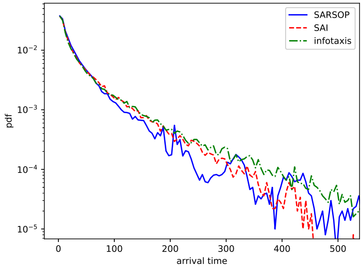


Figure Arrival time pdfs for searching in the source

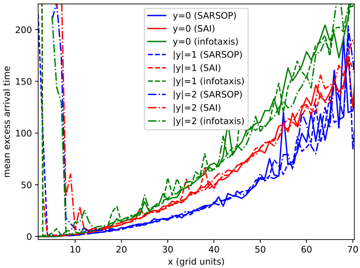


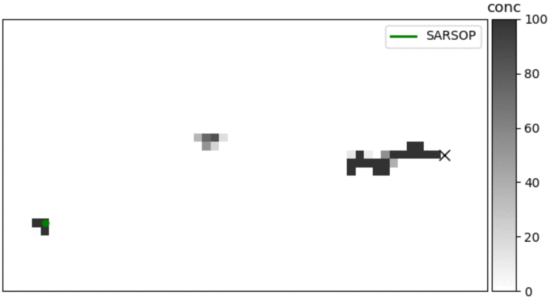
Figure Mean arrival time (minus distance from source) conditioned on starting position

policy	$\mathbb{E}[T \mid T < T_{\max}]$	$\Pr(T \geq 50)$	$\Pr(T \geq 100)$	$\Pr(T \geq T_{\max})$
SARSOP	39.4 ± 0.2	0.223 ± 0.001	0.0951 ± 0.0009	$< 10^{-5}$
SAI	43.0 ± 0.2	0.263 ± 0.001	0.124 ± 0.001	0.0014 ± 0.0001
infotaxis	48.6 ± 0.2	0.277 ± 0.001	0.145 ± 0.001	0.0013 ± 0.0001

Table 1: Arrival time statistics when using the empirical likelihood and searching within the DNS.

Optimal behaviors

Near-optimal policies exhibit behaviors seen in real moths. As time since last encounter grows, agent zigzags cross-wind with increasing amplitude. Eventually turns downwind to avoid missing the source



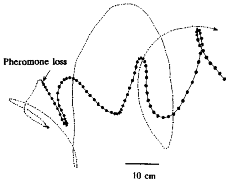
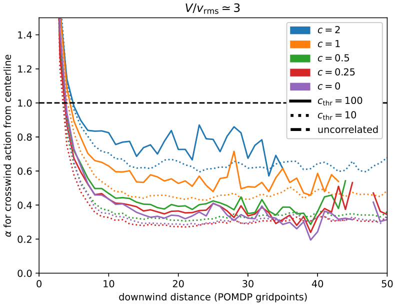
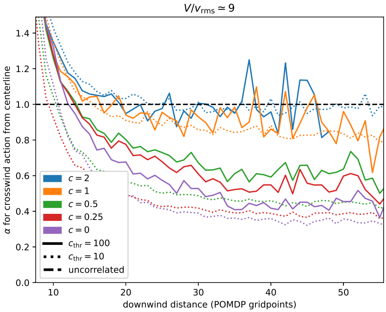


Figure Downwind motion in silkworm flight after odor loss [Willis and Arbas, 1991]

Correlations

- Real flows are not Markovian: due to spatial structure of puffs, consecutive observations usually positively correlated
- Correlation strength sensitive to flow speed, plume shape, c_{thr}
- Define $\alpha \equiv \frac{\log \Pr(o_t=1|o_{t-1}=1,s,a)}{\log \Pr(o_t=1|,s,a)}$ so that $\alpha < 1 \implies$ correlated, $\alpha > 1 \implies$ anticorrelated
- Rescale flow time $t \to ct$



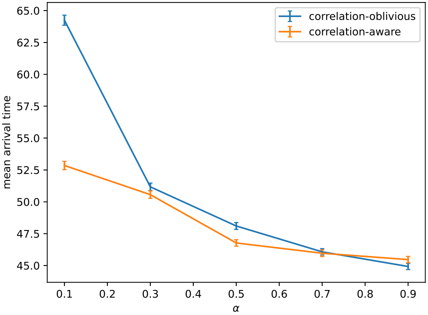


Correlations in the POMDP

- In principle, POMDP can accommodate arbitrary correlations by augmenting state space $s \to s \otimes o_{t-1} \otimes \cdots \otimes o_{t-k}$
- Affects both Bayes inference and optimization (solution of Bellman)
- But makes problem exponentially harder computationally
- Q: does minimal extension ($k = 1$) improve search performance? i.e. exponentially decaying correlations

Artificial correlations

- Control correlations by hand: impose log likelihood ratio α artificially, constant over $\mathbf{x}$ and actions
- Fix unconditioned likelihood to that obtained empirically. Law of total probability $\Pr(A) = \sum_B \Pr(A|B)P(B)$ then sets conditional likelihoods



Takeaway: correlations make searching harder. Only partially mitigated by including them in optimization and Bayesian inference

Searching in a slower flow

- Now, modify correlations by changing flow speed $t_{\text{flow}} \to ct_{\text{flow}}$
- As flow slows down agent has more time to see spatial structure of odor dispersal
- Uncorrelated for $c \to \infty$, frozen flow with strong corr. for $c \to 0$

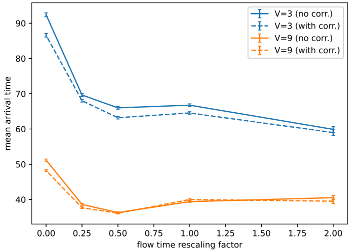


Figure Mean arrival times w/ and w/o correlation sensitivity. Non-monotonic behavior?

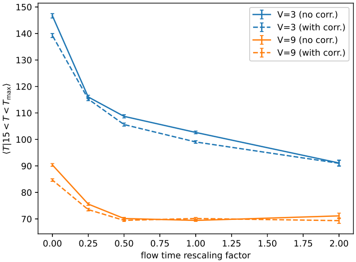


Figure Non-monotonicity disappears if average conditioned on not starting close to the source

Effect of correlations

- What's going on? Correlations impact convergence rate of posterior

$$b(s) = \frac{\exp\left(\sum_{t=1}^{N} \lambda_t\right) b_0(s)}{\sum_s \exp\left(\sum_{t=1}^{N} \lambda_t\right) b_0(s)}$$

where $\lambda_t = \log \ell(o_t|s)$ and ℓ is likelihood under agent's model

- By Law of Large Numbers $\sum_{t=1}^{N} \lambda_t \to NE[\lambda]$
- Can show using Chebyshev's inequality that for $x > 0$

$$\Pr\left(\left|\frac{1}{N}\sum_t \lambda_t - E[\lambda]\right| \geq x\right) \leq \frac{2C}{Nx^2}$$

where $C = \sum_{t=1}^{\infty} \mathrm{Cov}(\lambda_t, \lambda_1)$.

Effect of correlations (cont'd)

- Thus $C = \sum_{t=1}^{\infty} \mathrm{Cov}(\lambda_t, \lambda_1)$ sets typical time to converge (along with spatial structure of λ)
- C increases (decreases) when positive (negative) correlations are turned on and agent is unaware
- If agent is aware, situation is less clear, but frequently find that $C_{uncorr.} < C_{aware} < C_{unaware}$ (for positive corr.)
- This accounts for behavior seen in mean arrival time performance. Nonmonotonic effect explainable by negative correlations close to source, which this argument shows are helpful

Conclusion

- Tracking a source in turbulence is hard because there are no gradients
- POMDP formalizes difficult problem into something we can solve
- Optimal strategies for realistic flows resemble search trajectories observed in real animals
- Correlations in real flows can impede Bayesian search by slowing the convergence of the posterior

References I

Balkovsky, E. and Shraiman, B. I. (2002).
Olfactory search at high reynolds number.
Proceedings of the National Academy of Sciences, 99(20):12589–12593.

Celani, A., Villermaux, E., and Vergassola, M. (2014).
Odor landscapes in turbulent environments.
Physical Review X, 4(4):041015.

David, C., Kennedy, J., and Ludlow, A. (1983).
Finding of a sex pheromone source by gypsy moths released in the field.
Nature, 303(5920):804–806.

Loisy, A. and Eloy, C. (2022).
Searching for a source without gradients: how good is infotaxis and how to beat it.
Proceedings of the Royal Society A, 478(2262):20220118.

Vergassola, M., Villermaux, E., and Shraiman, B. I. (2007).
'Infotaxis' as a strategy for searching without gradients.
Nature, 445(7126):406–409.

Villermaux, E. and Innocenti, C. (1999).
On the geometry of turbulent mixing.
Journal of Fluid Mechanics, 393:123–147.

Willis, M. A. and Arbas, E. A. (1991).
Odor-modulated upwind flight of the sphinx moth, manduca sexta l.
Journal of Comparative Physiology A, 169:427–440.

References II

Yee, E., Kosteniuk, P., Chandler, G., Biltoft, C., and Bowers, J. (1993).
Statistical characteristics of concentration fluctuations in dispersing plumes in the atmospheric surface layer.
Boundary-Layer Meteorology, 65(1):69–109.