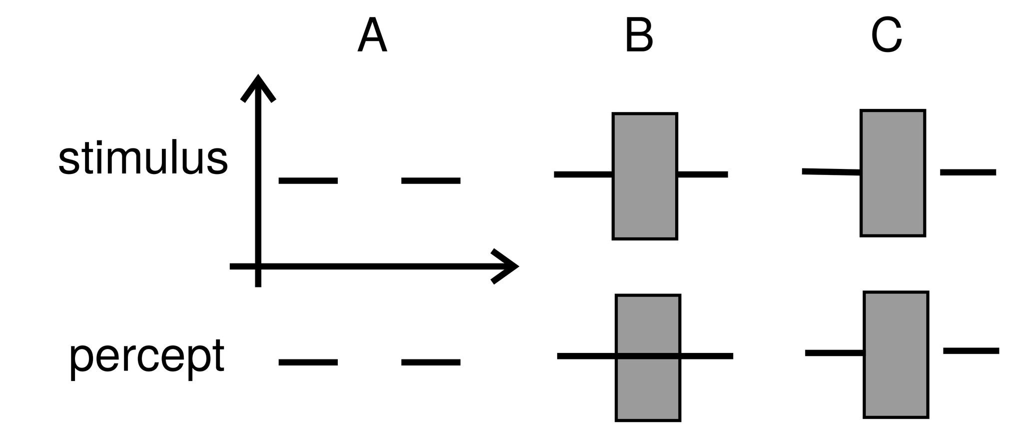
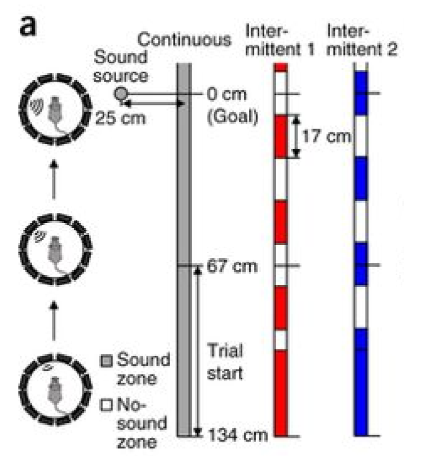
Neural recognition and postdiction by temporal distributed distributional code

Li Kevin Wenliang, Maneesh Sahani
Gatsby Computational Neuroscience Unit, University College London kevinli@gatsby.ucl.ac.uk

Introduction

- Accurate and flexible state representation is crucial
- Experiments suggest optimal "cue combination" [2, 4]
- **Postdiction** is common in dynamic environments [5]
- Examples: continuity illusions [1], localisation [3]





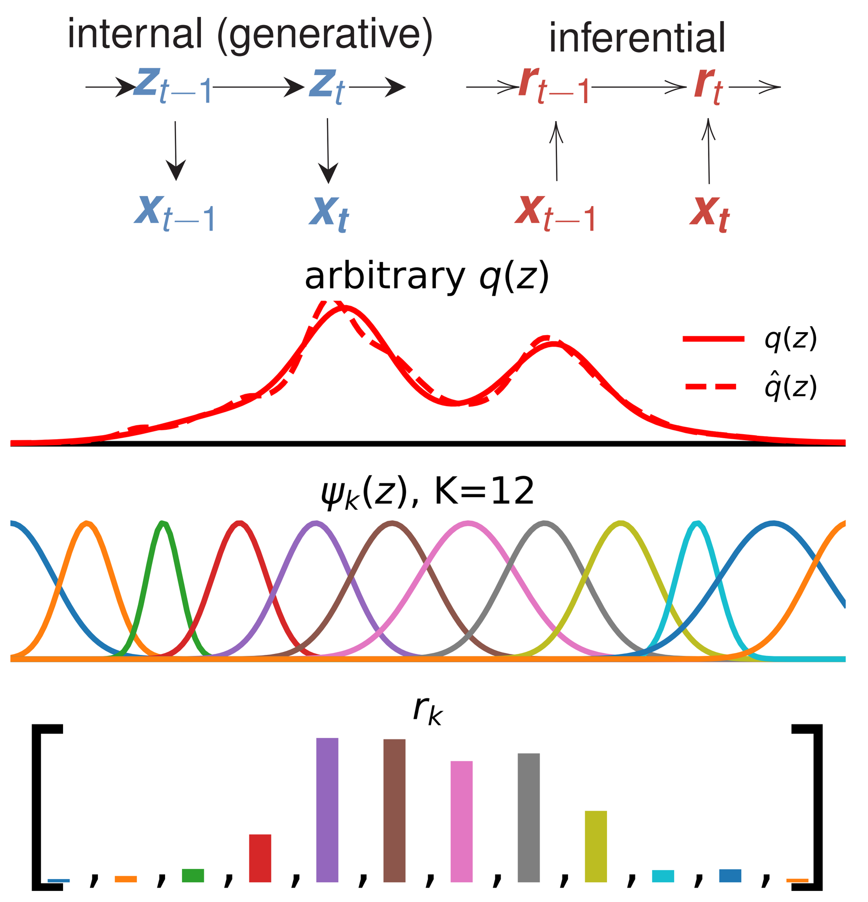
Internal and inferential model

Internal model:

- latent causes $z_t = f(z_{t-1}) + \zeta_z$
- observation $x_t = g(z_t) + \zeta_x$
- No assumptions on ζ

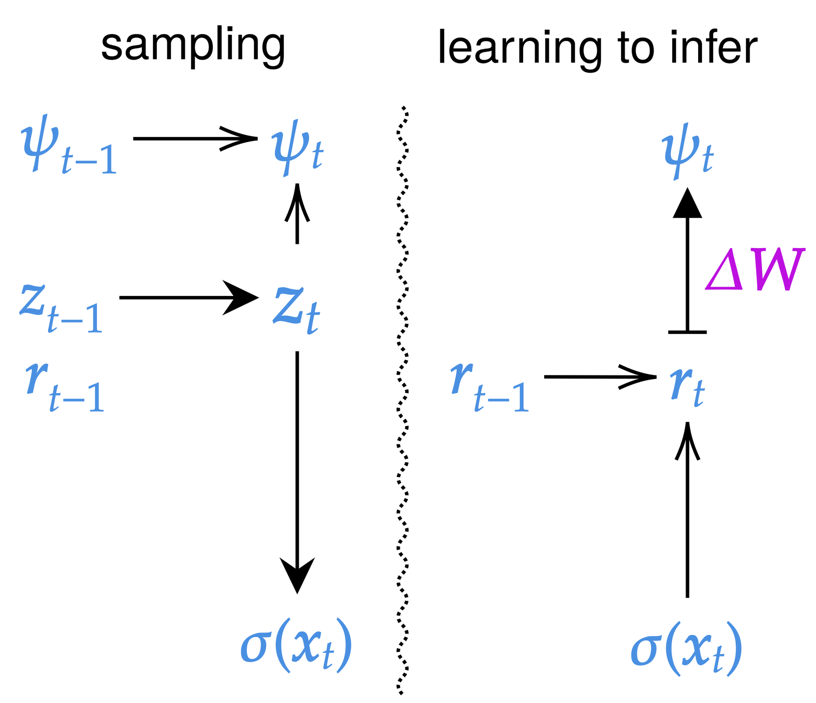
Inferential model

- encode $q(z_{1:t}|x_{1:t})$
- by $r_t = \mathbb{E}_q[\psi(z_{1:t})]$ (DDC [6])
- $\psi_t = k(\psi_{t-1}, z_t)$
- train W: $h_W(r_{t-1}, x_t) \xrightarrow{MSE} \psi_t$ outputs $\mathbb{E}_q[\psi_t]$
- assess by max ent



Learning to infer

1. At each time t, have: r_{t-1}, z_{t-1}, and ψ_{t-1}
2. sample $z_t, x_t \sim p$
3. compute $\psi_t = k(\psi_{t-1}, z_t)$
4. update W to minimise: $\|h_W(r_{t-1}, x_t) - \psi_t\|_2^2$ δ-rule if h is linear/bilinear
5. filter $r_t = k(r_{t-1}, x_t)$
6. readout: find α
$$V(z_{t-\tau:t}) \approx \alpha^\intercal \psi_t$$
$$\mathbb{E}_q[V(z_{t-\tau:t})] \approx \alpha^\intercal r_t$$



Reference

Albert S Bregman. MIT press, 1994.
Marc O Ernst and Martin S Banks. In: *Nature* 415.6870 (2002).
Akihiro Funamizu, Bernd Kuhn, and Kenji Doya. In: *Nature neuroscience* 19.12 (2016).
Konrad P. Körding, Shih-pi Ku, and Daniel M. Wolpert. In: *Journal of Neurophysiology* 92.5 (2004).
Shinsuke Shimojo. In: *Frontiers in psychology* 5 (2014).
Eszter Vértes and Maneesh Sahani. In: *AISTATS* (2018).

Model

internal model: latent $p(z_t|z_{t-1})$ and observation $p(x_t|z_t)$

DDC: encode $q(z_{1:t}|x_{1:t}) \leftrightarrow r_t := \mathbb{E}_q[\psi(z_{1:t})]$

temporal code: $\psi_t = k(\psi_{t-1}, z_t)$

sleep phase: train $h(r_{t-1}, x_t) \to \psi_t$ by MSE (δ-rule)

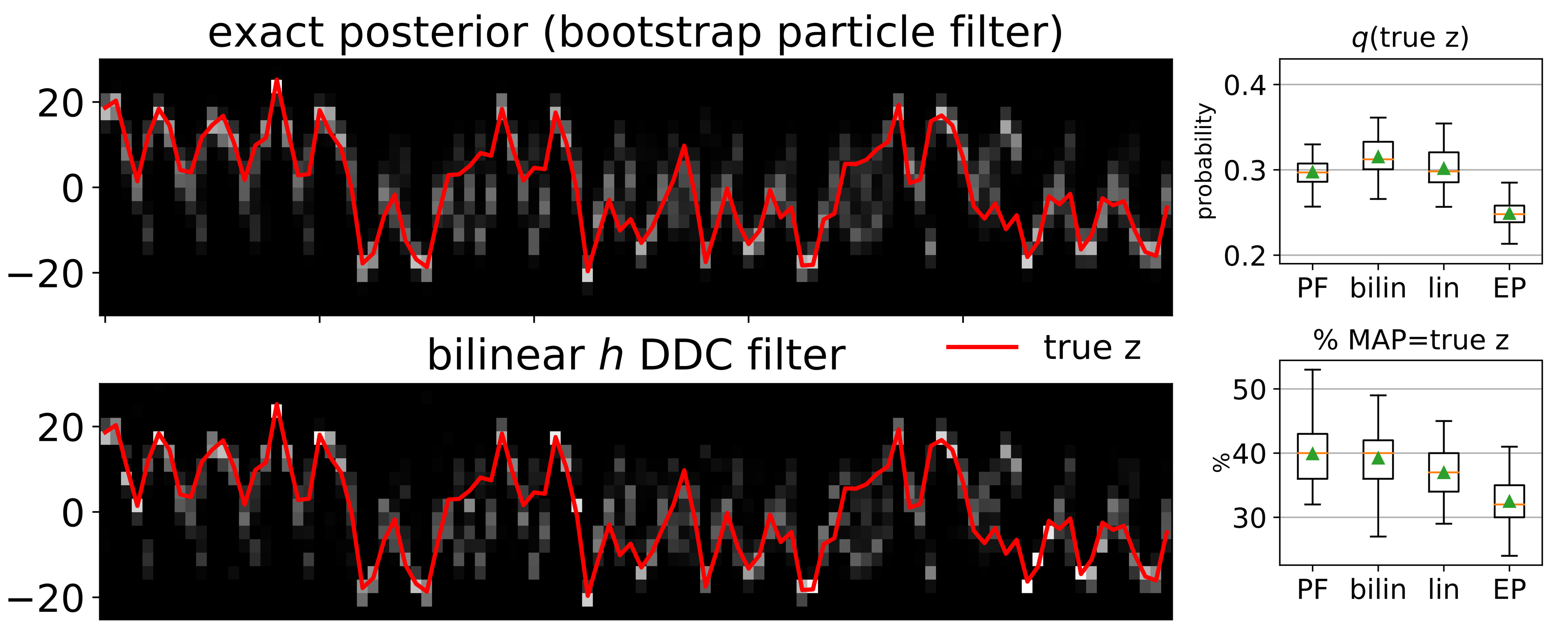
wake phase: predict $\mathbb{E}_q[\psi_t] \approx r_t = h(r_{t-1}, x_t)$

flexible (non-Gaussian) q, postdictive

pre-print

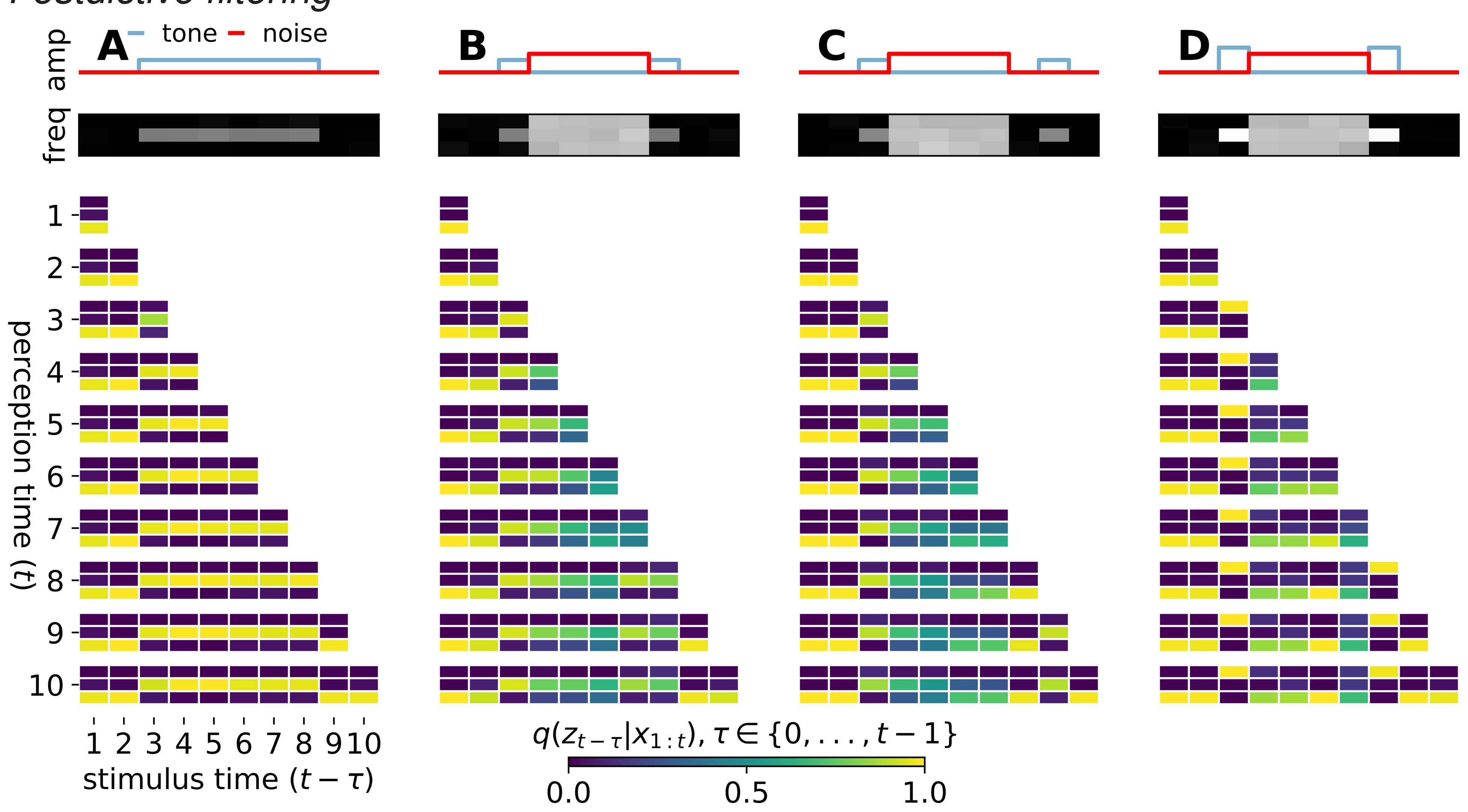
Key results

Bimodal filtering



Auditory continuity illusion

Postdictive filtering



Model details

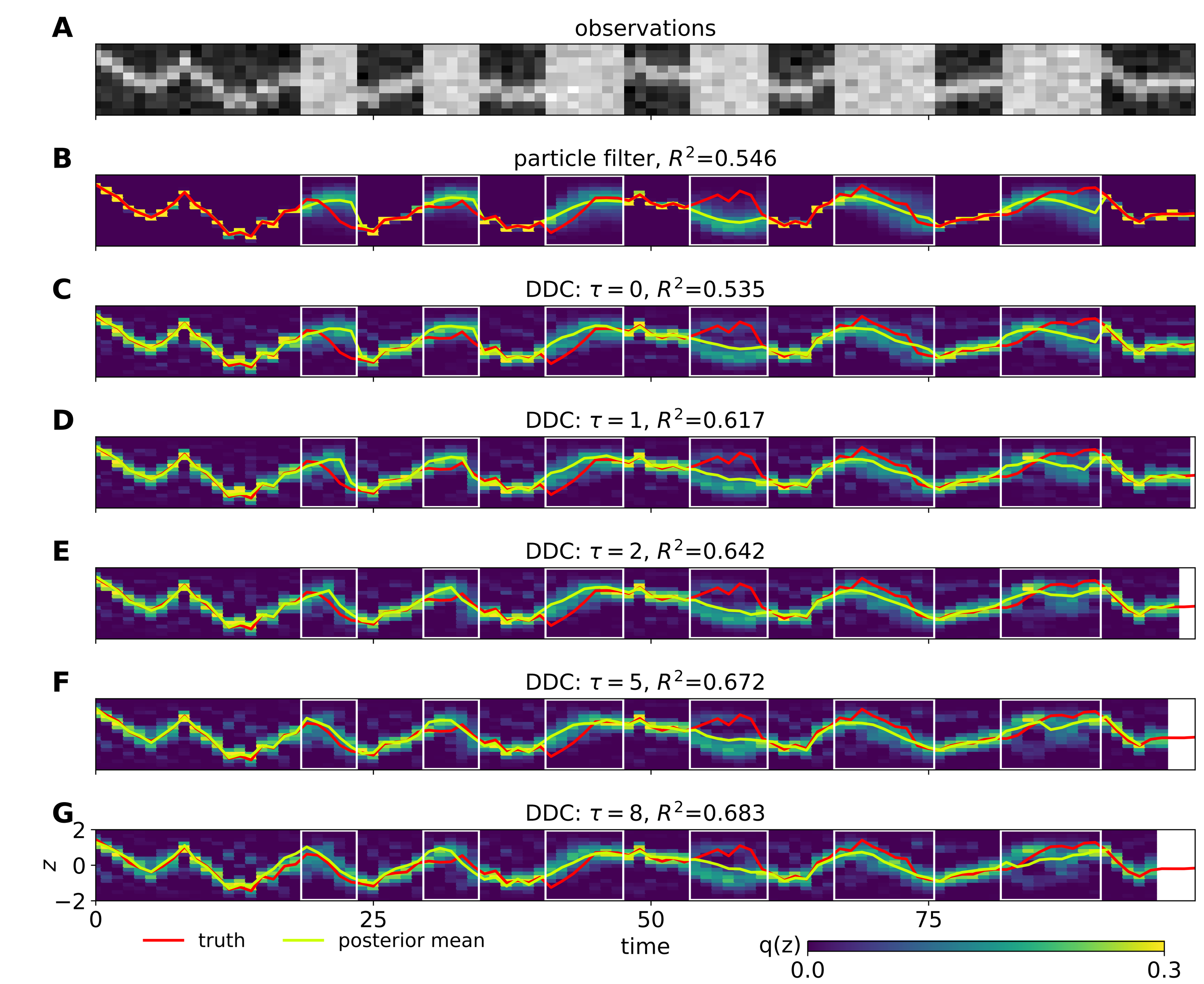
In sleep phase, recognition model solves

$$\min_W \mathbb{E}_{p(z_{1:t}, x_{1:t})}\left[\|h_W(r_{t-1}, x_t) - \psi_t\|_2^2\right] \tag{1}$$

- r_{t-1} is a summary statistics of $x_{1:t-1}$.
- $\psi_t = U\psi_{t-1} + \gamma(z_t)$, random but fixed temporal encoding function
$$\psi_t = \gamma(z_t) + U\gamma(z_{t-1}) + U^2\gamma(z_{t-2}) + \cdots$$
- $h_W^{bil} = W \cdot (r_{t-1}\sigma(x_t)^\intercal)$, or $h_W^{lin} = W[r_{t-1}; \sigma(x_t)]$, $\sigma()$ random but fixed
- $\min \mathbb{E}_p\|h - \psi\|^2 \Leftrightarrow \min \mathrm{KL}[p\|q]$, where $q = \mathrm{ExpFam}(\psi, h_W)$
- Possible to assume h is linear only in $\sigma(x_t)$ and derive a formal solution, albeit with complicated neural implementation
- If the state-space model is stationary, W should converge
- Independent noise in ψ_t, σ and r_t averages out for large population
- Adaptation: follow gradient of variational objective $\nabla_\theta \mathcal{F}_\theta(z, x)$

Additional results

Occluded tracking with DDC



Smoothing flash-lag effect