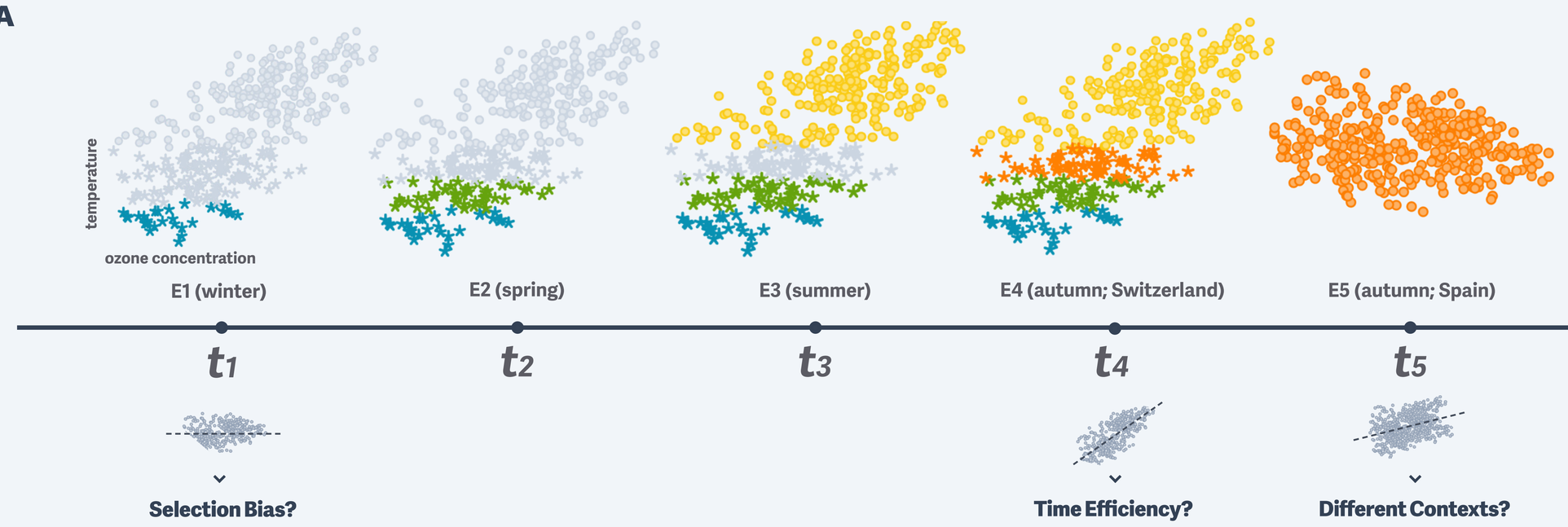
Learning Causal Networks from Episodic Data

Osman Mian, Sarah Mameche, Jilles Vreeken

CISPA Helmholtz Center for Information Security

TL;DR Given biased data arriving in batches, we show how to consistently discover causal networks

EPISODIC DATA



PROBLEM SETTING

Given Data arriving over time t, in episodes $E_1, \ldots E_t, \ldots$

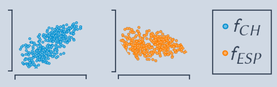
Episodic Bias holds when an episode is a subsample from the population, for multiple subpopulations $S_1, \ldots S_K$

Context Bias can occur when combining episodes E_i, E_j from different contexts with distinct causal models M_i, M_j

ASSUMPTIONS

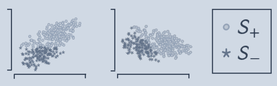
Causal Model consists of a set of causal DAGs $G_1, \ldots, G_R$ over a common set of variables X, and a set of structural equations as follows,

$$X_i^r = f_r(pa_i^r, N_i^r), \quad N_i^r \perp\!\!\!\perp X_i^r$$



Selection Bias in each episode through a categorical variable S that encodes missingness. It depends on X through a noisy selection mechanism

$$S^r = g^r(X^r, N_s^r), \quad N_s^r \perp\!\!\!\perp S^r$$



Ignorability assumes that we obtain sufficently many episodes to allow unbiased estimation of the causal models

Detectability a milder assumption assuming that changes in causal models are reflected in the data distributions

EPISODIC CAUSAL DISCOVERY

Algorithmic Markov Condition posits that the true cause-effect relationships allow a factorization of the data distribution with lowest Kolmogorov complexity

$$K(P(X)) = \sum_r \sum_i K(P(X_i^r \mid pa_i^r))$$

Minimum Description Length (MDL) to approximate the above for a fixed model class, here nonparametric regression models. Here, our objective is

$$\min_{\{G_1,\ldots G_R\}} \sum_r \left(L(G_r) + \sum_i L(X_i^r \mid pa_i^r)\right)$$

Efficient Causal Discovery using an MDL-based score and edge-greedy search (here, using the GLOBE algorithm)

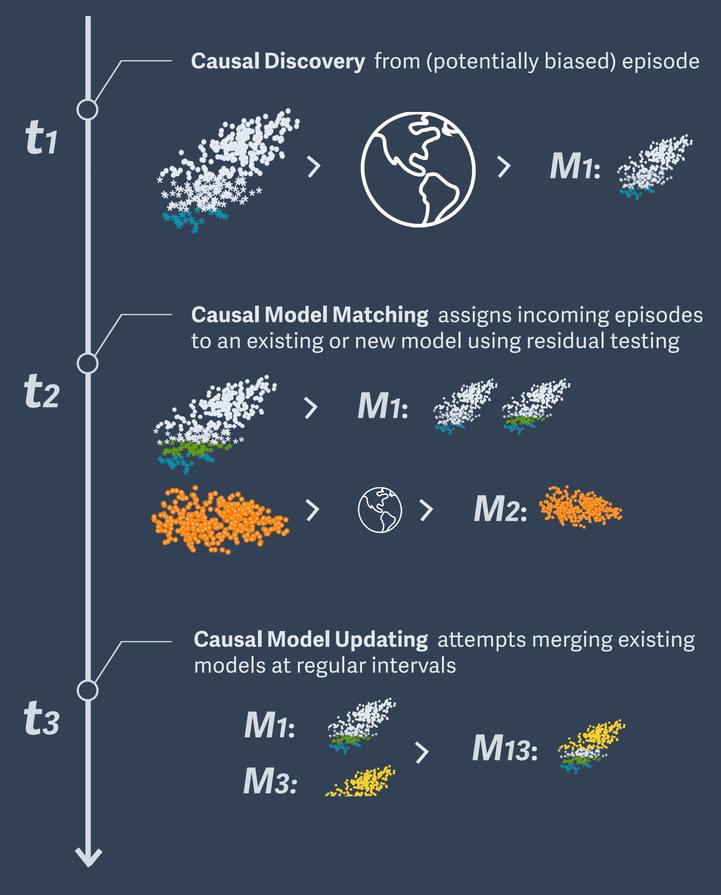
Residual Equality testing to test whether episodes originate from the same structural equation model,

$$H_0 : P(X_i^r \mid pa_i^r) \equiv P(X_i^{r'} \mid pa_i^{r'})$$

(here, using the non-parametric Kolmogorov-Smirnov test).

Thm 1 *(simplified)* Given enough episodes such that ignorability holds, a consistent scoring criterion can consistently identify **multiple** causal networks.

CONTINENT



Thm 2 *(simplified)* Under detectability and using a consistent score, our algorithm is consistent, in particular, it never incorrectly merges **biased** episodes.

EXPERIMENTS