Synthesis of Parametric Hybrid Automata from Time-Series

Miriam García Soto, Thomas A. Henzinger & Christian Schilling

ATVA 2022 - October 2022

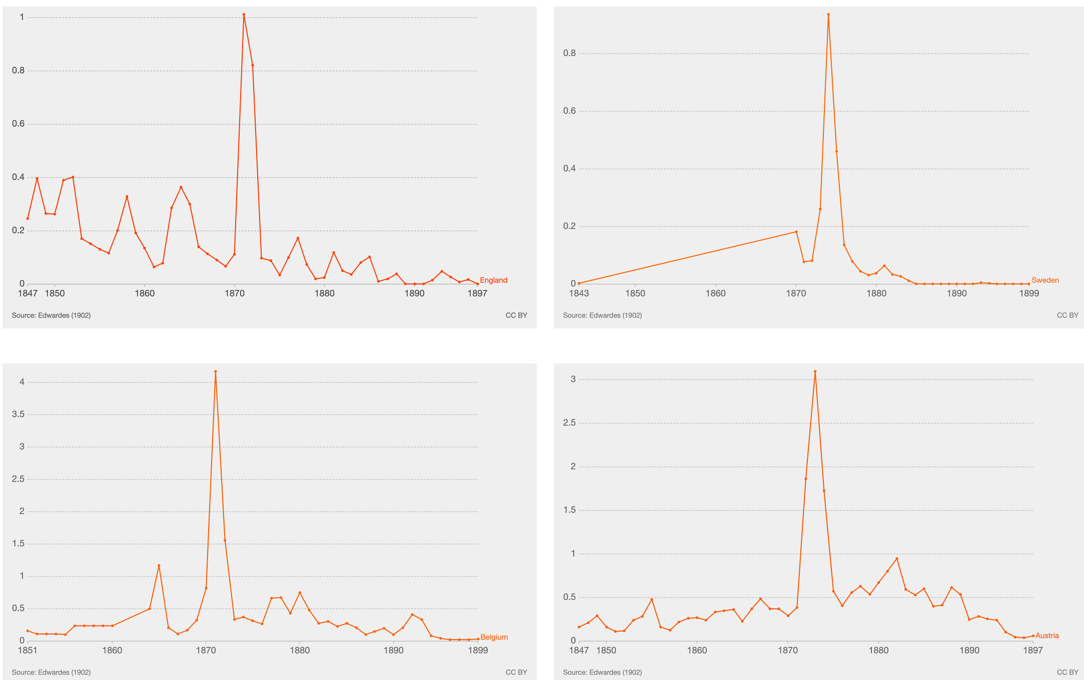
UNIVERSIDAD COMPLUTENSE MADRID

IST AUSTRIA Institute of Science and Technology

AALBORG UNIVERSITET

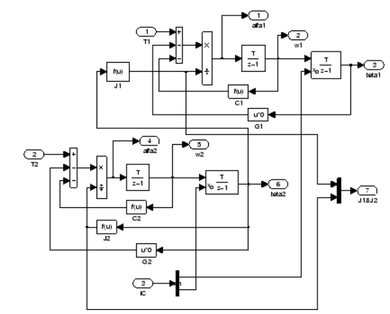
Motivation

Main goal of many sciences is to create a model from a real system



Experimental data

Analysis

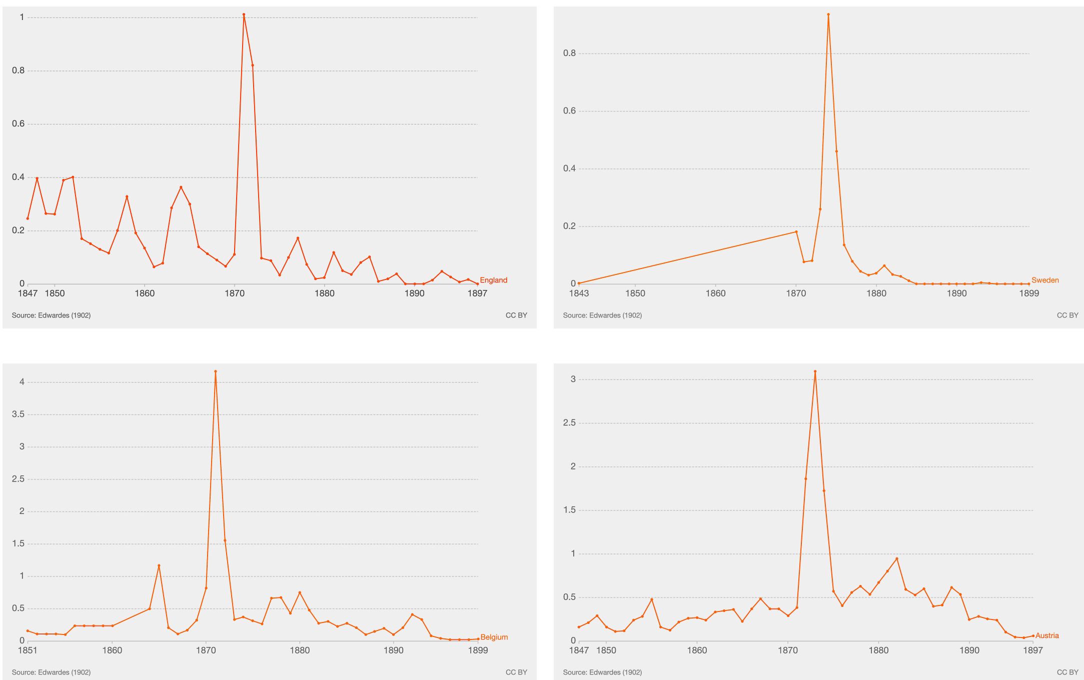


Model

Challenge

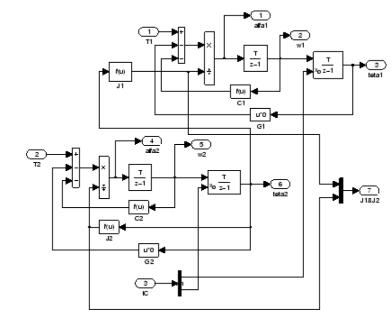
How to automatically create a model?

Main goal of many sciences is to create a model from a real system



Experimental data

Synthesis



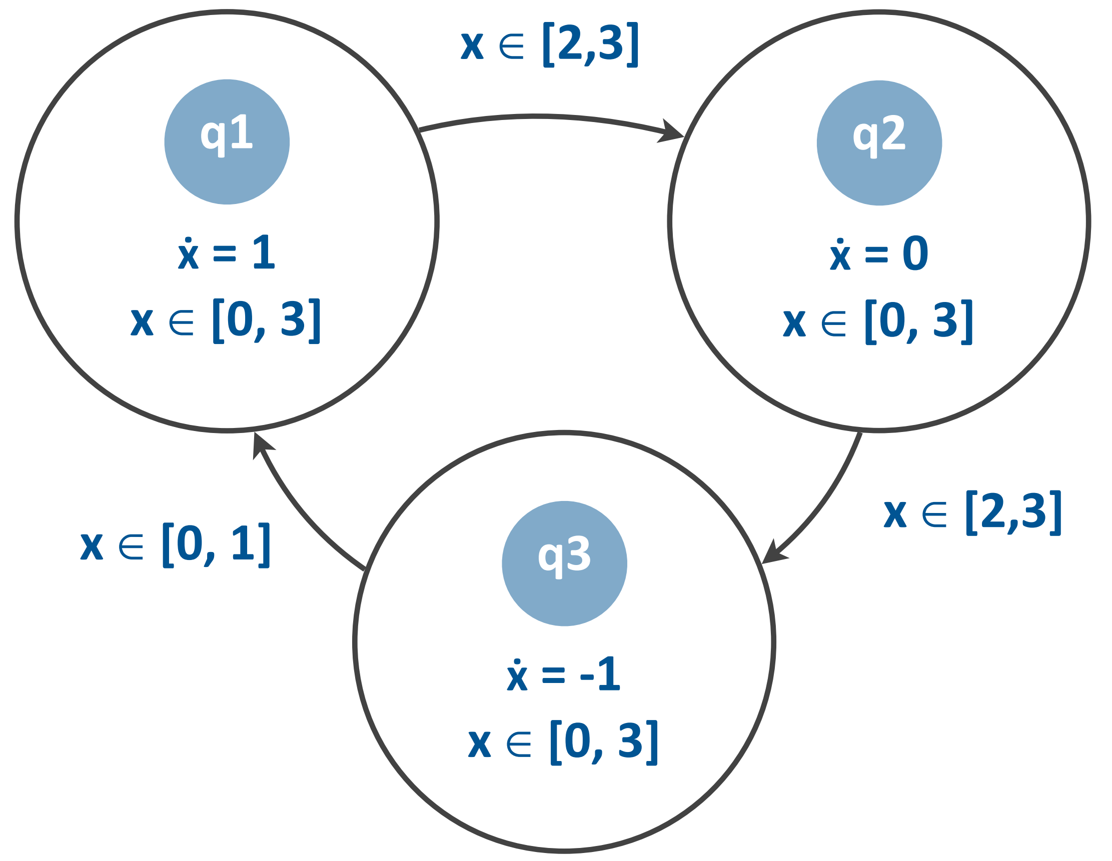
Model

- Correctness Specifications - **Time-series** data
- Synthesis Algorithm - **Two phases** based on clustering and polyhedral construction
- Model - linear **hybrid automaton** with constant dynamics

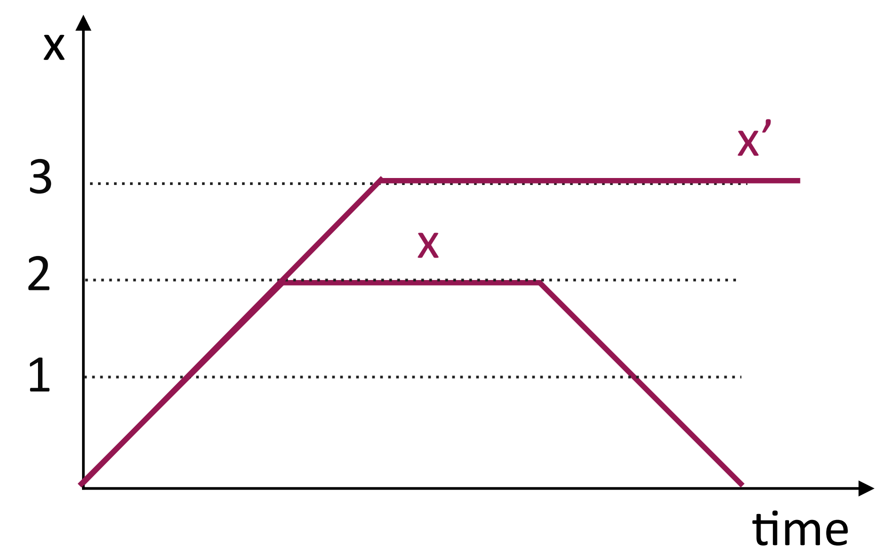
Model

Hybrid Automata capture the mixed **continuous** and **discrete** behaviour.

Linear hybrid automaton (LHA)



Non deterministic mode changes



Piecewise linear (PWL) function x(time)

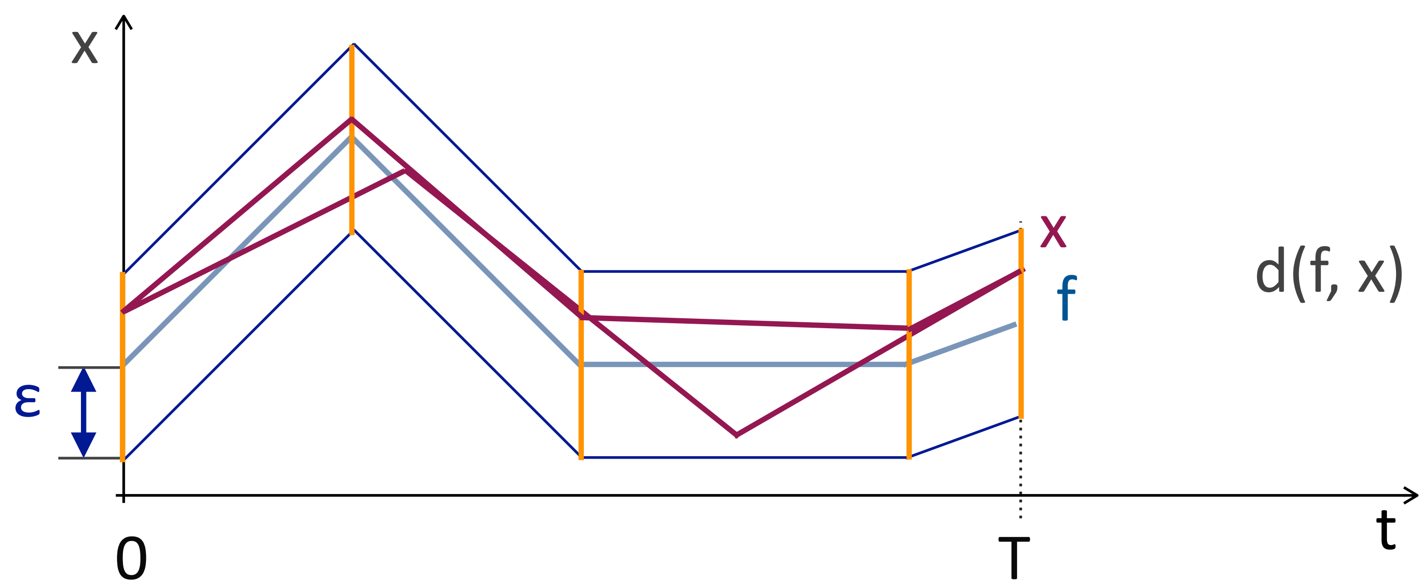
$x_i(\text{time}) = m_i$

Problem: find a LHA **model** that is **close to** the **data**

Problem

ε-capturing

An LHA *H* ε-captures a PWL function f if there exists an execution x in *H* with $d(f, x) \leq \varepsilon$



$$d(f, x) = \max_{t \in [0,T]} \| f(t) - x(t) \|$$

Synthesis problem

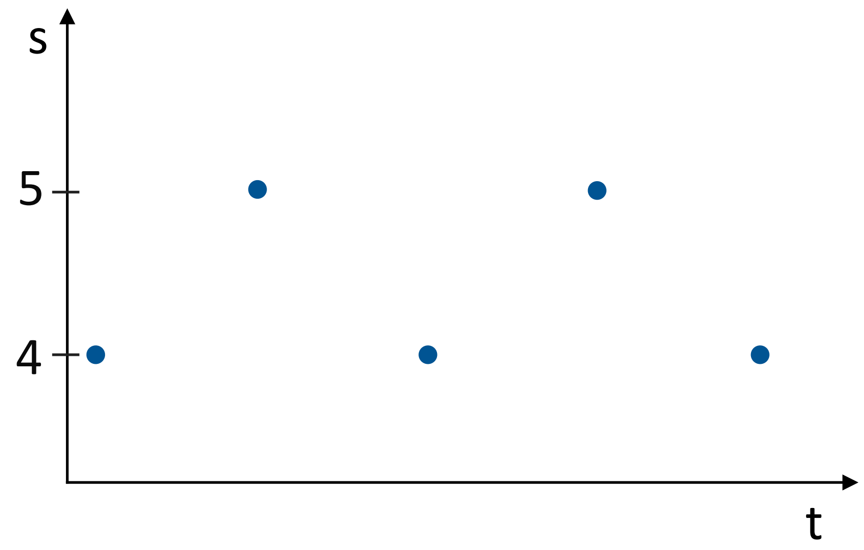
Given a set of PWL functions *F* and $\varepsilon \geq 0$, construct an LHA *H* that ε-captures all $f \in F$

Synchronous time **switchings** for f and x

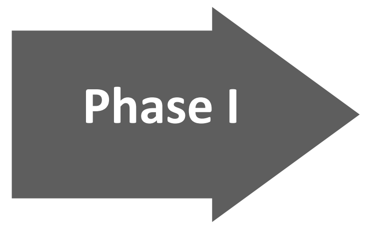
Synthesis algorithm

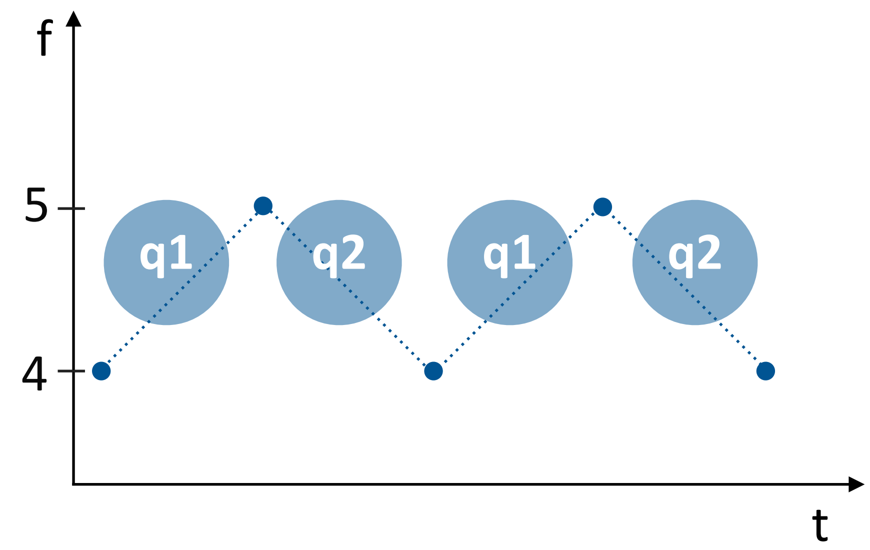
- **Offline** synthesis algorithm
- **Finite** set of time-series
- **Family of** linear hybrid automata **models**
- **Parameter polyhedron** represents all the model solutions
- **ε** is not given but **minimised**

Synthesis algorithm

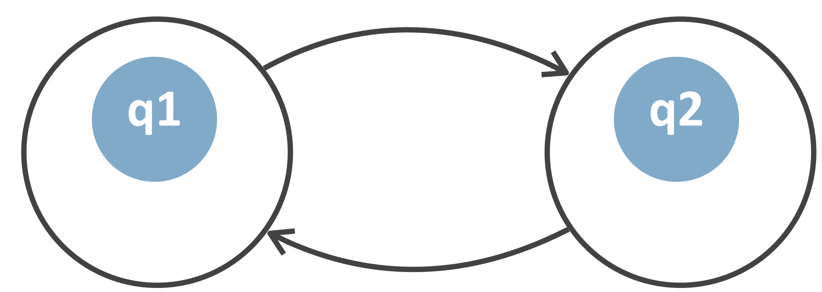


Time-series s(t)

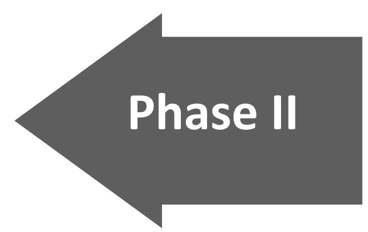


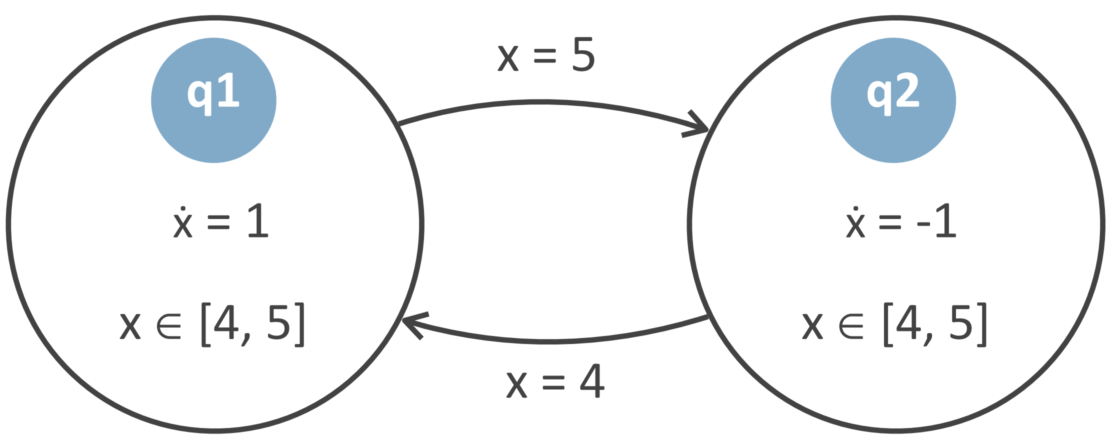


Piecewise linear (PWL) function f(t)



Discrete structure H_d





Hybrid automaton H

For every **time-series s(t)** there exists an **execution x(t)** in H such that **distance(s(t), x(t)) ≤ ε** at every time t

Phase I: time-series to PWL function

Time series s(t)

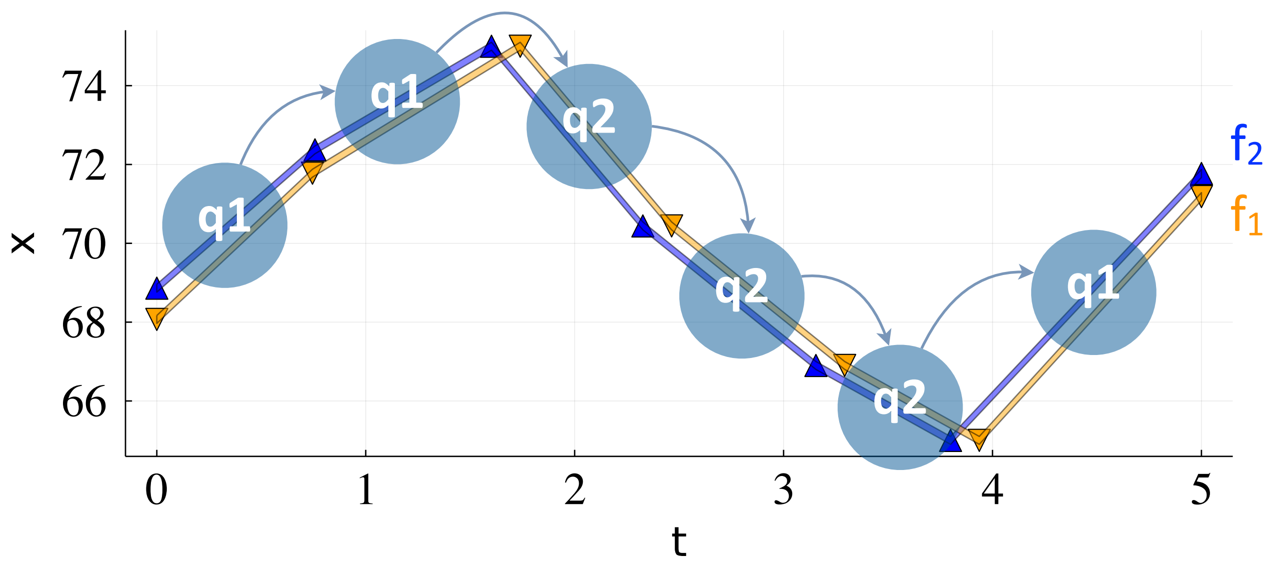
+

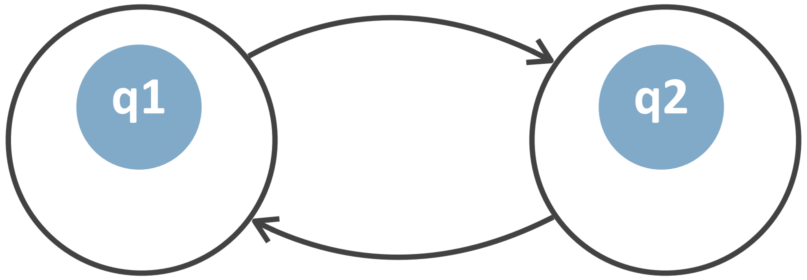
δ ≥ 0

Piecewise linear (PWL) function f(t)

- **Simplification** of time-series
- **Linear interpolation**
- Variant of **Ramer-Douglas-Peucker** algorithm

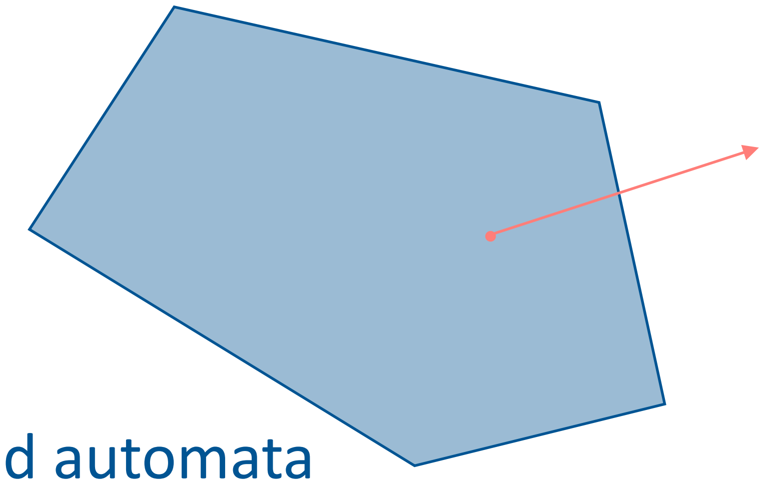
Phase I: symbolic synthesis





- Send slope vectors of pieces to **clustering algorithm** (k-means)
- Clustering cost for different **numbers of clusters** k, together with relative improvement compared to $k-1$
- **Mapping** from **pieces** in the PWL function **to clusters**

Phase II: flow polyhedron



Instantiated linear hybrid automaton

Family of linear hybrid automata

number of symbolic locations

number of time-series

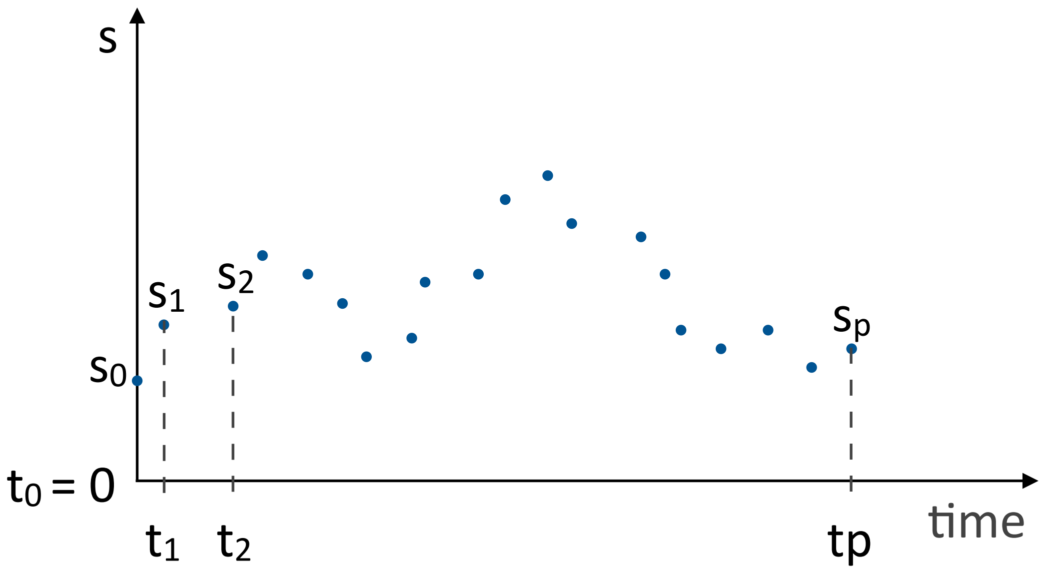
$$\lambda \quad n + r \quad n + 1$$

dimension of time series data

ε value

Intersection of the flow polyhedra for each time-series

Phase II: parameter polyhedron



- time-series $\mathbf{s} = s_0, s_1, \ldots, s_p$ with p linear pieces
- with time instants $t_0, \ldots, t_p$
- $\mathbf{m}_1, \ldots, \mathbf{m}_\lambda$ as parameters for slopes
- Initial state $\mathbf{x}_0$ of execution $\mathbf{x}$
- $M: \{1, \ldots, p\} \longmapsto \{1, \ldots, \lambda\}$

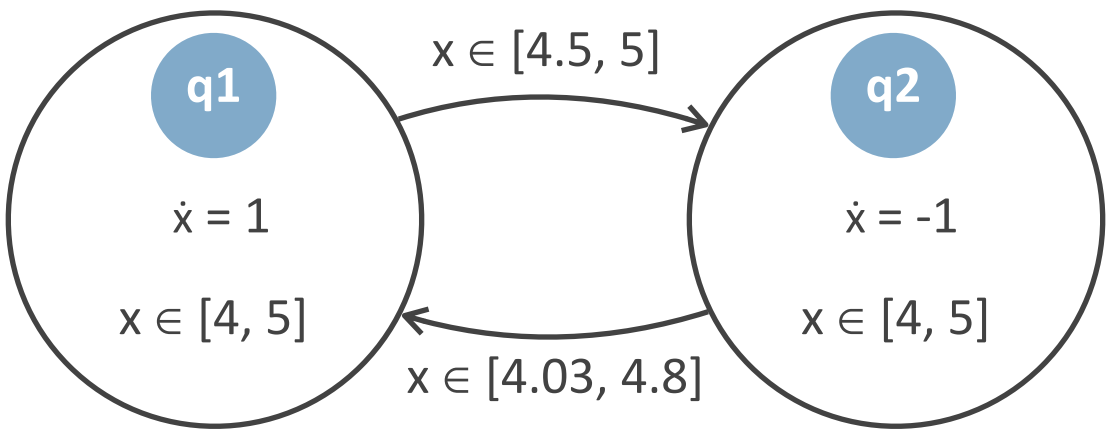
Flow polyhedron for time-series s

$$
\begin{aligned}
P_s = \{(\mathbf{m}_1,\ldots,\mathbf{m}_\lambda,\mathbf{x}_0,\varepsilon) \in \mathbb{R}^{\lambda n + rn} \times \mathbb{R}^{\geq 0} \mid \mathbf{x}_0 \in B_\varepsilon(s_0), \\
\mathbf{x}_0 + (t_1 - t_0)\mathrm{m}_{M(1)} \in B_\varepsilon(s_1), \\
\mathbf{x}_0 + (t_1 - t_0)\mathrm{m}_{M(1)} + (t_2 - t_1)\mathrm{m}_{M(2)} \in B_\varepsilon(s_2), \\
\mathbf{x}_0 + (t_1 - t_0)\mathrm{m}_{M(1)} + \ldots + (t_p - t_{p-1})\mathrm{m}_{M(p)} \in B_\varepsilon(s_p)\}.
\end{aligned}
$$

$$\mathbf{P} = \bigcap_{\mathbf{s} \in \textit{Data}} \mathbf{P}_\mathbf{s}$$

Phase II: hybrid automaton model

- Choose a point **p** in the parameter polyhedron **P**
- **Minimisation** of **ε** provides a point **p** = ($\mathbf{m}_1$, …, $\mathbf{m}_\lambda$, $x_0^{(1)}$, …, $x_0^{(r)}$, **ε**)
- $\mathbf{m}_1$, …, $\mathbf{m}_\lambda$ define the dynamics for each of the hybrid automaton modes ($\dot{x} = m_i$)
- The **invariant** of each mode is the ε-bloated convex hull around all data points associated with the mode
- The **guard** of each transition is the ε-bloated convex hull around all data points associated with the transition



Correctness

Synthesis problem

Given a finite set of time-series and a discrete structure, find the minimal value $\varepsilon \geq 0$ and a hybrid model H with the given discrete structure such that H ε-captures each time-series.

Theorem
Phase II solves problem in polynomial time

Implementation

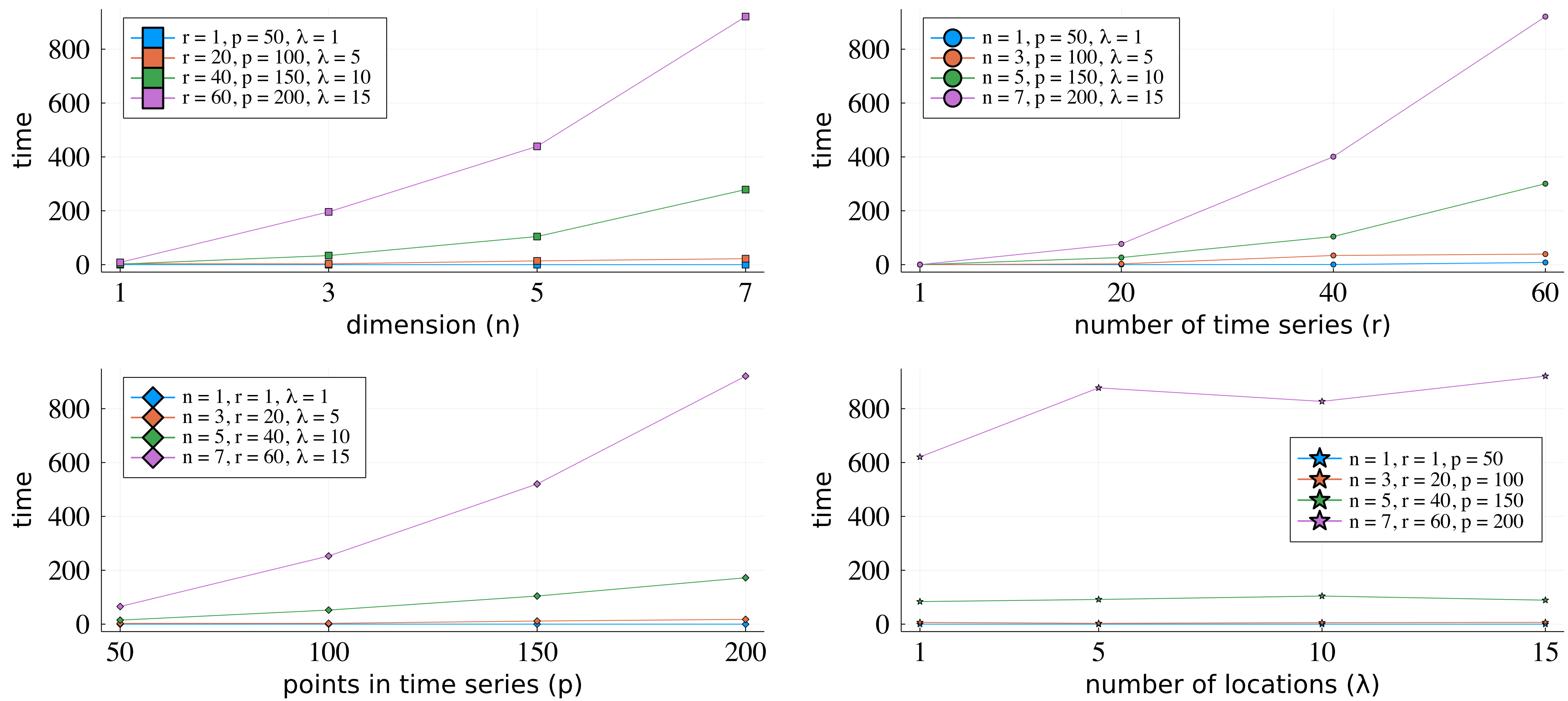
HySynthParametric

- **Julia** programming language
- Evaluation in **two case studies**
- Synthetic data for the **thermostat** and **cell-cycle** models
- **Scalability** evaluation for the thermostat
- **Synthesis** of mammalian cell-cycle from biological model

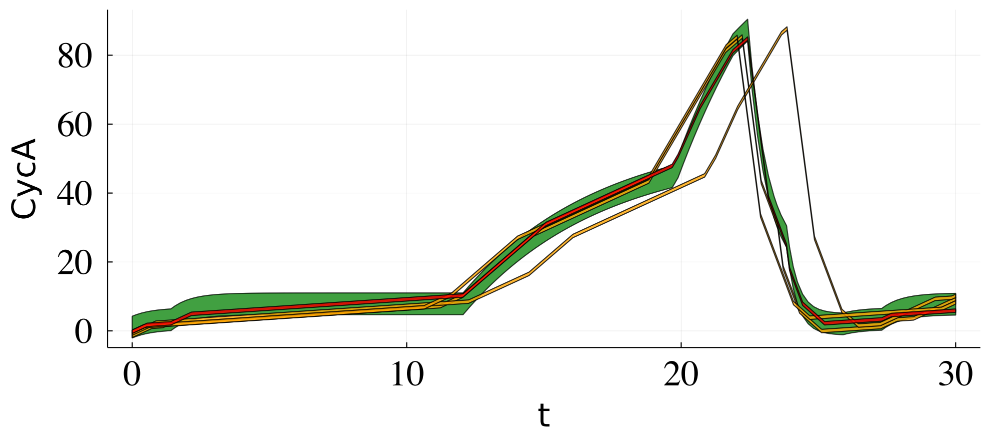
https://github.com/HySynth/HySynthParametric

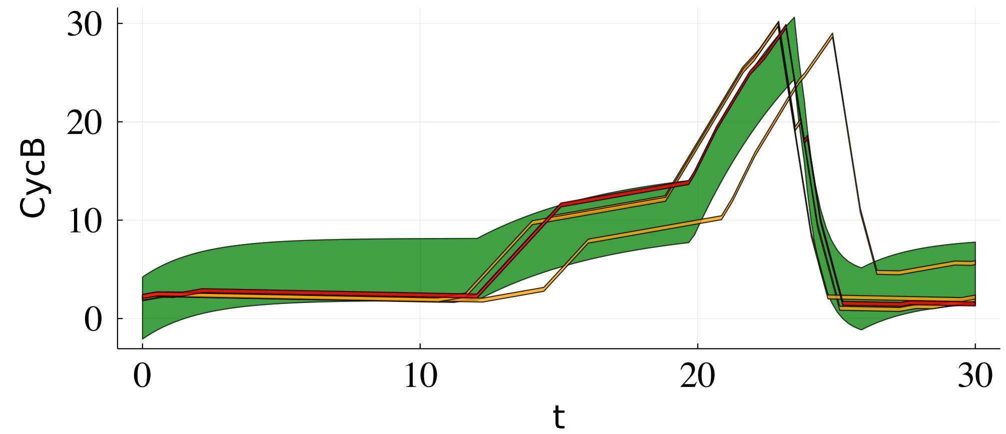
Evaluation: scalability

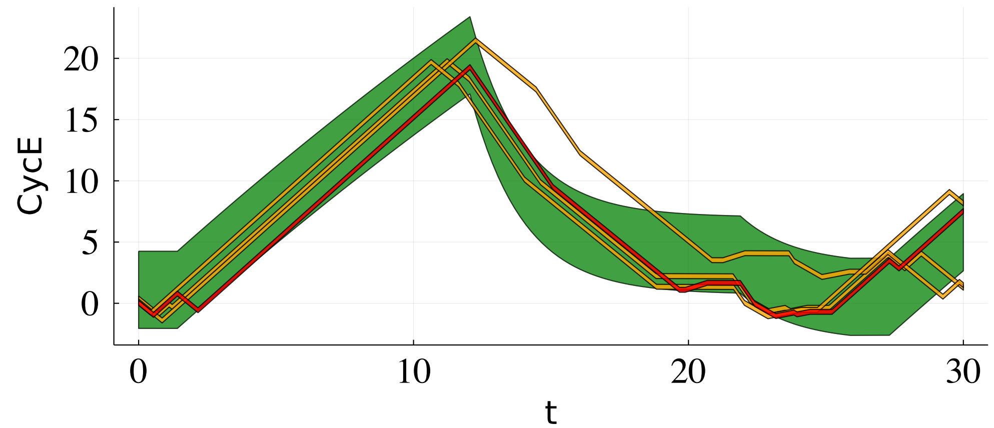
- *r* time series with *p* data points, *n* dimensions and λ locations



Evaluation: cell cycle regulation







Summary

- **Automatic synthesis** of linear hybrid automaton from time-series data
- **Two** algorithmic **phases**
- Phase I: **discrete structure** of the hybrid automaton
- Phase II: **parameter space** of all the possible models
- Selects a model by solving a **linear program**
- The **model** is **ε-close** to the time-series
- **ε** is **minimal** for the discrete structure chosen in the phase I
- **Algorithm** is **polynomial** and scales to thousands of data points