Multiple changepoint detection

Outliers and Constraints

G. Rigaill
with V. Runge, T. Hocking, R. Maidstone, P. Fearnhead

LaMME (Stat & Genome) and IPS2 (Gnet)

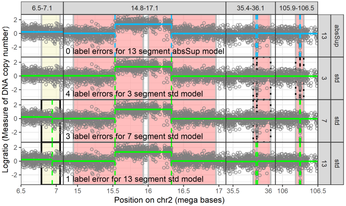
October 2019

Outline

Examples

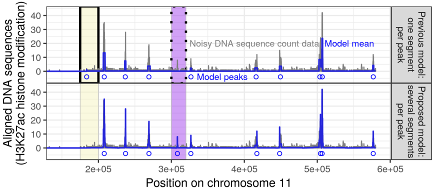
Copy Number data

[Picard *et al.* 2005]



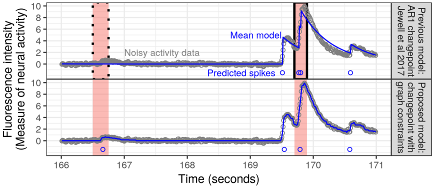
Chip-Seq data

[Hocking *et al.* 2015]



Spike-Train data

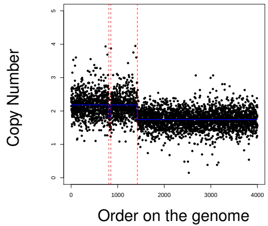
[Jewell *et al.* 2018]



A common yet major problem

According to the National Research Council (US)

- A need for complex models
  - Genomics [Hocking *et al.* 2016, Pierre-Jean *et al.* 2015...]
  - Geology, Finance, Biology, ...



- A recent explosion in methods for detecting changes
  - Univariate Gaussian model: [Harchaoui and Levy-Leduc 2009, Killick *et al.* 2011, Frick *et al.* 2014, Lin et al. 2015, Dette and Wied 2015, Haynes *et al.* 2016, Maidstone *et al.* 2017, Fryzlewicz 2017...]

Methods (exactly) minimizing a (penalized) cost

- Good statistical properties [Yao 1989, Lebarbier 2005, Baraud *et al.* 2009, Arlot *et al.* 2012]
- For which models is the computational burden in $\mathcal{O}(n^2)$ or less ?

A parametric piecewise constant model

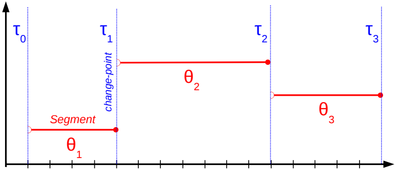
data $X_1, \dots, X_n$

changes $\tau = (\tau_1, .., \tau_D)$

segments $s_d = (\tau_{d-1}, \tau_d]$

parameters $\theta = (\theta_1, .., \theta_D)$

model $X_i \sim \mathcal{F}(\theta_d)$ i.i.d



Model $\mathcal{F}$	$\rightarrow$	Loss γ	$\rightarrow$	Minimize
$X_i \sim \mathcal{N}(\theta_d, \sigma^2)$		$(X_i - \theta_d)^2$		$\sum_{d=1}^{\lvert\tau\rvert} \sum_{\tau_{d-1}+1}^{\tau_d} (X_i - \hat{\theta}_d)^2$

- the number and the position of the changes are not known
- the set of all segmentations, $\mathcal{M}_n$, is of size 2^{n-1}

Maximum likelihood

- Minus the log-likelihood

$$cost(\tau) = \sum_{d=1}^{|\tau|} \min_{\theta}\{ \sum_{\tau_{d-1}+1}^{\tau_d} \gamma(X_i, \theta)\}$$

- Minimizing $cost(\tau)$ would lead to $n-1$ changes
- We need to penalize!

Penalised Maximum likelihood

- Linear penalties, e.g [Yao 1989]

$$pen(D) = 2\sigma^2 D \log(n)$$

- Concave penalites, e.g [Lebarbier 2005]

$$pen(D) = 2\sigma^2 D(2\log(n/D) + 5)$$

- Good statistical properties

[Yao 1989, Lavielle *et al.* 2000, Lebarbier 2005, Garreau *et al.* 2018]

Algorithms

Optimization with a linear penalty: $pen(D) = \beta D$

$$\arg\min_{\tau} \{cost(\tau) + \beta|\tau|\}$$

- Use it several times for a concave penalty [Killick *et al.* 2012, Haynes *et al.* 2016]

Optimization given the number of changepoints

$$\arg\min_{\tau,\ |\tau|=D} \{cost(\tau)\}$$

- Update-rules similar to those for linear penalties
- I will not talk about those in this talk

[Fisher 1958, Bellman 1961, Picard 2005, Rigaill 2010-2015, Maidstone *et al.* 2016]

Outline

Outline

Std. Dynamic Programming

A numerical view

$$C_n = \underset{\tau \in \mathcal{M}_n}{\arg\min} \left\{ cost(\tau) + \beta|\tau| \right\}$$

- Dynamic programming to compute C_n [Fisher 1959, Bellman 1961, Jackson et al. 2005]

$$C_n = \min_{t<n} \left\{ \underbrace{C_t}_{\text{best up to } t} \underbrace{+ \min_{\theta}\{\sum_{t+1}^{n} \gamma(X_i, \theta)\}}_{\text{last segment}} \right\} + \underbrace{\beta}_{\textit{penalty}} \qquad \mathcal{O}(n^2)$$

Std. Dynamic Programming - Generalization

No dependencies

- Justified for many parametric and non-parametric models

  [Auger and Lawrence 1989, Arlot *et al.* 2012, Cleynen and Lebarbier 2013, Celisse *et al.* 2018...]

  - $\mathcal{O}(n^2)$ if an efficient calculation of $\min_{\theta}\{\sum_{t+1}^{n}\gamma(X_i,\theta)\}$ is possible
  - consistency, oracle inequality ...

- Pruning: consider a subset of all $\{\tau < n\}$ (PELT) [Killick *et al.* 2011]

  - $\mathcal{O}(n)$ if the number of "changepoints" is large ($D^* \propto n$)
  - $\mathcal{O}(n^2)$ if no changepoints

What if ?

- What if $n \geq 10^5$ and there are few changepoints ?
- What if it is difficult to calculate $\min_{\theta}\{\sum_{t+1}^{n} \gamma(X_i, \theta)\}$?
- What if there are dependencies in the model ?

Outline

A Viterbi-like DP algorithm

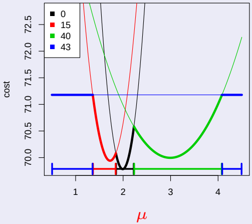
A functional view [Rigaill 2010-2015, Johsnon 2010-2013, Rote 2012, Maidstone *et al.* 2016]

The overall cost is
$$C_n = \min_{\tau \in \mathcal{M}_n, \theta} \left\{ \beta|\tau| + \sum_{d=1}^{|\tau|} \sum_{\tau_{d-1}+1}^{\tau_d} \gamma(X_i, \theta_d) \right\}$$

A functional representation

- conditioning on the last segment parameter

$$\widetilde{C}_n(\mu) = \min_{\substack{\tau,\theta \\ \theta_{|\tau|}=\mu}} \left\{ \beta|\tau| + \sum_{d=1}^{|\tau|} \sum_{\tau_{d-1}+1}^{\tau_d} (X_i - \theta_d)^2 \right\}$$



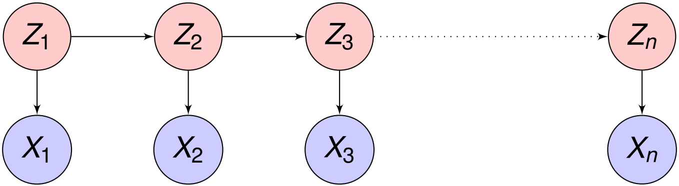
A Viterbi-like DP with an infinite state space

Functional pruning

$$\widetilde{C_{n+1}}(\mu) = \min \left\{ \begin{array}{c} \text{“no change at n”} \\ \widetilde{C_n}(\mu) \\ \\ \min_{\mu'}\{\widetilde{C_n}(\mu')\} + \beta \\ \text{“a change at n”} \end{array} \right\} + (X_{n+1} - \mu)^2$$

- Apply the update-rule per interval using simple calculus
- At worst $2n - 1$ intervals.
- Worst case complexity $\mathcal{O}(n^2)$

Changepoints as a continuous HMM ?



Continuous state space	Z_i in an interval of $\mathbb{R}$
Chain rule	$(X_i \mid Z_i = \mu) \sim \mathcal{N}(\mu, \sigma^2)$
Transition kernel	$k(x, y) \propto \mathbf{I}_{x=y} + e^{-\beta} \mathbf{I}_{x \neq y}$

Update of the functional cost visually

$n = 43$

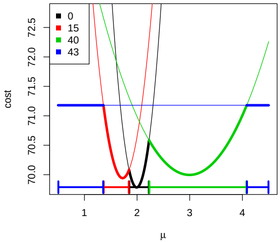
$n = 43 + \text{“}0.5\text{”}$

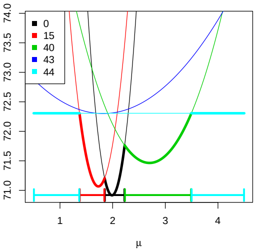
$+(X_{44} - \mu)^2$

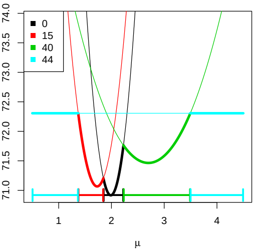
vs. $C_{44} + \beta$

$n = 44$

discard $\tau = 43$



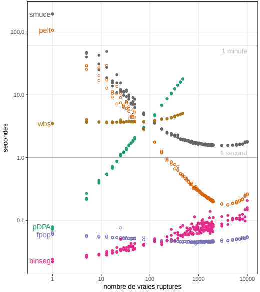




Fpop algorithm in practise

- Quasi-linear on average
  (even without changepoints)
- About 4 seconds for $n = 10^7$

For $n = 10^5$ and varying D^*



Good statistical properties

- Model selection and consistency [Yao 1989, ..., Garreau and Arlot 2017]
- Can be used to be optimise concave penalties
- Top performer on the simulations of WBS [Fryzlewicz 2014]

Extensions of functional pruning

Models solved by Std. DP

- Univariate exponential models [Cleynen *et al.* 2015]
- But not (yet?) efficient for multivariate models . . .

Models not solved by Std. DP

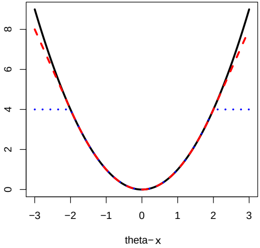
- Non-convex losses to cope with outliers [Fearnhead and Rigaill 2018]
  - No "analytical solution" for $\sum_{\tau+1}^{n} \gamma(X_i, \hat{\theta}_d)$
- Models with dependencies [Maidstone *et al.* 2017, Hocking *et al.* 2018, Jewell *et al.* 2019]
  - Isotonic, peaks, train spikes . . .

Outline

Extension to Robust losses [Fearnhead and Rigaill 2018]

- Robust convex loss for changepoints and sequential tests

[Hušková and Sen 1989; Hušková 1991; Hušková and Picek, 2005; Hušková, 2013]

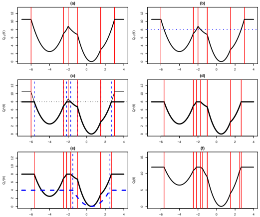


Biweight-loss

$$\gamma(x;\theta) = \begin{cases} (x-\theta)^2 & \text{if } |x-\theta| < K \\ K^2 & \text{otherwise} \end{cases}$$

Functional cost $\widetilde{C}_n$ with the biweight loss

$$\widetilde{C}_n(\mu) = \min_{\substack{\tau,\theta \\ \theta_{|\tau|}=\mu}} \left\{ \beta|\tau| + \sum_{d=1}^{|\tau|} \sum_{\tau_{d-1}+1}^{\tau_d} \gamma(X_i, \theta_d) \right\}$$



with the biweight loss

- $\widetilde{C}_n$ is still a piecewise polynomial function
- The recurrence on the last segment parameter works
- Applied per interval using simple calculus

R-Fpop for the biweight losses

Worst case complexity for the biweight loss

- $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ space

But fast in practise

	Biweight	L_2				
	R-Fpop	Fpop	PELT	BS	WBS	SmuceR
$n = 10^6$	$\sim$ 5 sec	$\sim$ 1 sec	$>$ 10 min	$\sim$ 1 sec	$\sim$ 82 sec	$>>$ 10 min
	Exact			*Heuristic*		

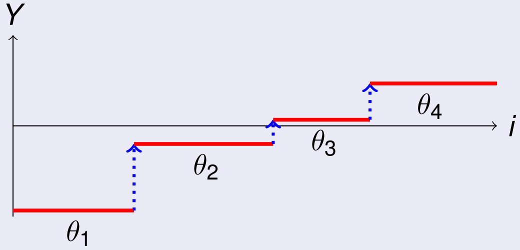
- More simulations in the paper

Outline

Extensions of functional pruning (isotonic)

Isotonic [Hocking *et al.* 2018]

$\theta_1 \leq \theta_2 \leq \theta_3 \ldots$



$$\widetilde{C_{n+1}}(\mu) = \min \left\{ \begin{array}{c} \text{“no change at n”} \\ \widetilde{C_n}(\mu) \\ \\ \underline{\min_{\mu' \leq \mu}} \{\widetilde{C_n}(\mu')\} + \beta \\ \text{“a change at n”} \end{array} \right\} + (X_{n+1} - \mu)^2$$

A graph for complex patterns (an intuitive overview)

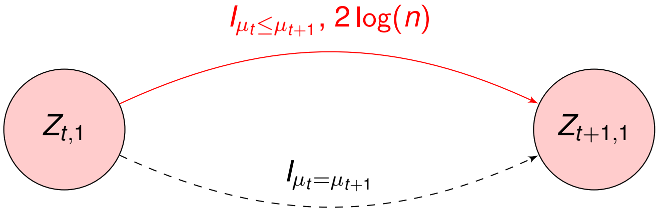
When does it work

- $\mathcal{S}$ a finite set
- A state space in $\mathcal{S} \times \mathbb{R}$
- Describe transitions with a graph $\mathcal{G}$

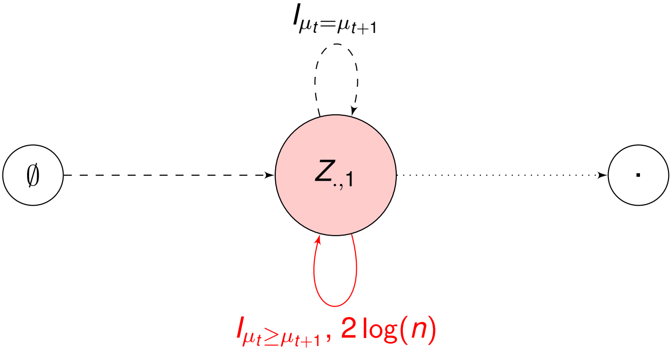
A graph of states and constraints $\mathcal{G}$

- A node (t, s) is associated to state s at time t
- A transition from (t, s) to $(t+1, s')$ is associated with
  - a linear constraints between μ_t and μ_{t+1}
  - a penalty

Isotonic example ($|\mathcal{S}| = 1$)

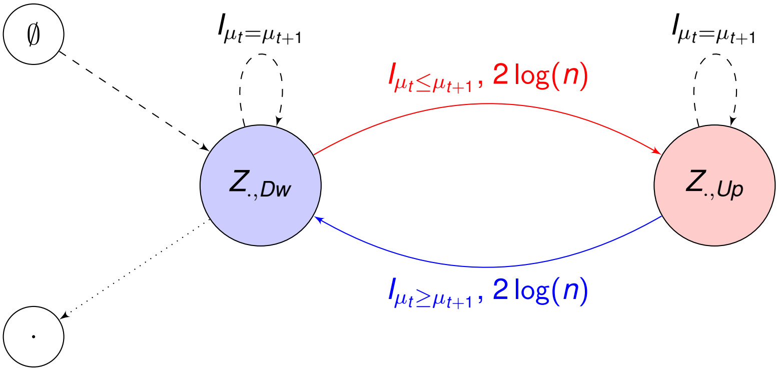


Isotonic example ($|\mathcal{S}| = 1$)

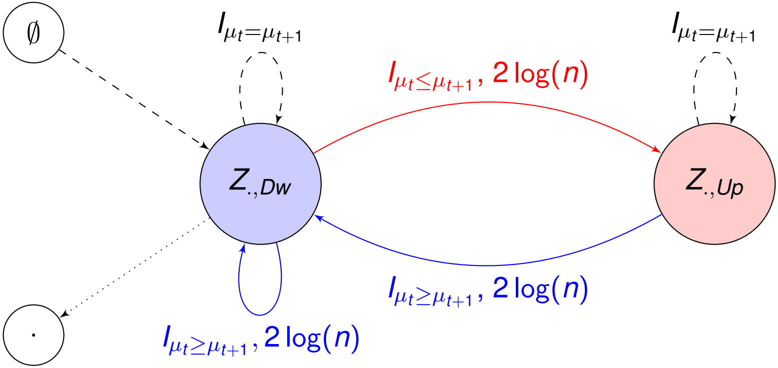


Up-Down pattern ($\mathcal{S} = \{Up, Dw\}$, $|\mathcal{S}| = 2$)

Peaks in Chip-Seq [Hocking *et al.* 2018]



Up-Down* pattern ($\mathcal{S} = \{Up, Dw\}$, $|\mathcal{S}| = 2$)



Same DP-algorithm for many graphs

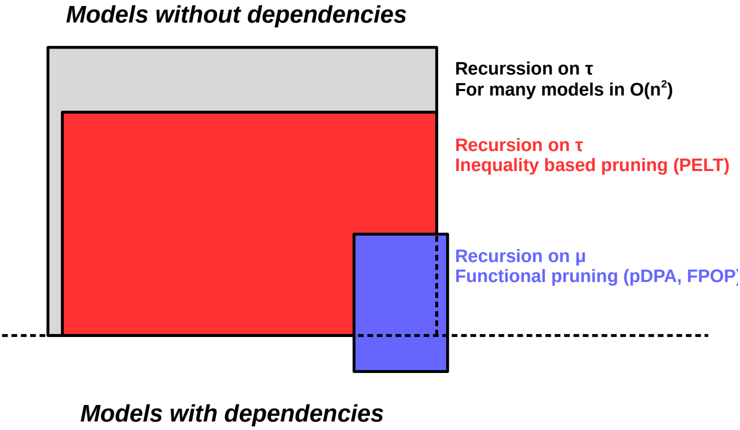
A generic algorithm: gfpop

- It takes as input the data and the graph
- Some applications
  - for peaks [Hocking et al. 2017 and 2018]
  - for spike-trains [Jewell et al. 2019]
- Generic implementation in developpement [Runge et al. in prep]

Pattern	loss	n	time
Up-Down	ℓ_2	10^6	$\sim 13s$
Up-Down	Biweight	10^6	$\sim 40s$

https://github.com/vrunge/gfpop

Segmentation models and exact DP algorithms



Outline

1. Multiple changepoints

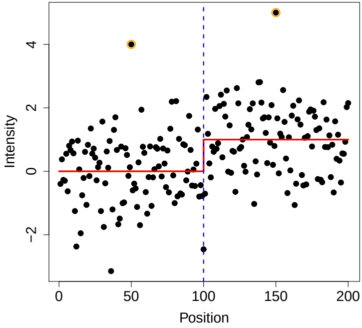
2. Dynamic Programming Algorithms
   - Recurrence on the last change
   - Recurrence on the last segment parameter
   - For outliers
   - For constraints

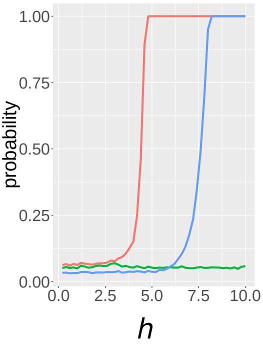
3. Estimation in the presence of outliers
   - Consistency for the number and the position of the changes
   - Some copy number simulations

4. Conclusion

In the presence of outliers

- L_2 loss
- Huber loss
- Biweight loss





Outline

Some notations and assumptions

- D^* true number of changepoints and τ_d^* true changepoint
- Let $\varepsilon_1, ...\varepsilon_n$ be i.i.d. noise random variables
- $X_i = \theta_d + \varepsilon_i$ with $\tau_d < i \leq \tau_{d+1}$

Some notations and assumptions

Define $M(\theta) = E\left[\gamma(\varepsilon_i, \theta)\right]$, with $\arg\min_\theta M(\theta) = 0$

(A1)

There exists constant $c_1 > 0$ and $c_2 > 0$ such that

$$M(\theta) \geq M(0) + \min\{c_1\theta^2, c_2\}$$

This is true if $M(\theta)$ as a positive second derivative around 0

Some notations and assumptions

Define

$$
\begin{aligned}
p &= P\left[\, |\varepsilon_i| \geq K \right] \\
\sigma^2 &= E\left[\, \varepsilon_i^2 \quad | \quad |\varepsilon_i| \ \leq K \right]
\end{aligned}
$$

(A2)

$$K^2(1-2p) - (1-p)\sigma^2 > 0$$

- Assuming ε_i has a finite variance this is true if $K > \sqrt{3}E[\varepsilon_i^2]$
- Assuming unimodality and a mode at 0 this is true if $p < 2/5$

Consistency of the biweight loss

Theorem

There exists constant C_1 and C_2 such that if $\beta_n = C_1 log(n)$

$$Pr\left(\begin{array}{c} \widehat{D}_n = D^* \\ \max_{d \in 1,\ldots D^*} \min_{j \in 1,..\widehat{D}_n} |\tau_d^* - \hat{\tau}_j| \leq C_2 log(n) \end{array} \right) \to 1$$

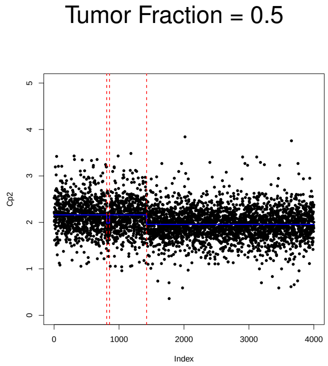
- Idea of the proof:
  - bound the decrease/increase in cost if we add or miss a change
  - γ is bounded and Lipschitz (with constant $2K$)
  - Bernstein's inequality

Outline

Simulating realistic DNA copy profiles [Pierre-Jean *et al.* 2014]

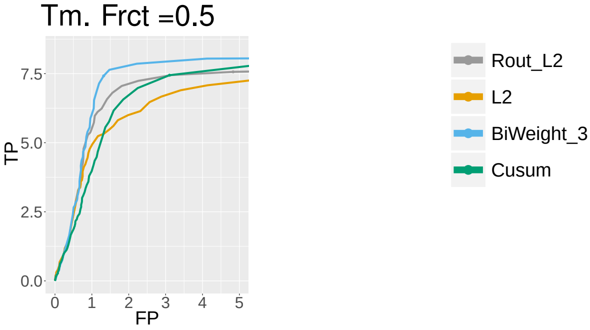
The jointseg package resample profiles with know truth

- Various tumor fractions (1: easy, ... 0.3: difficult)
- Known changepoint positions



Significant improvement in DNA copy number data

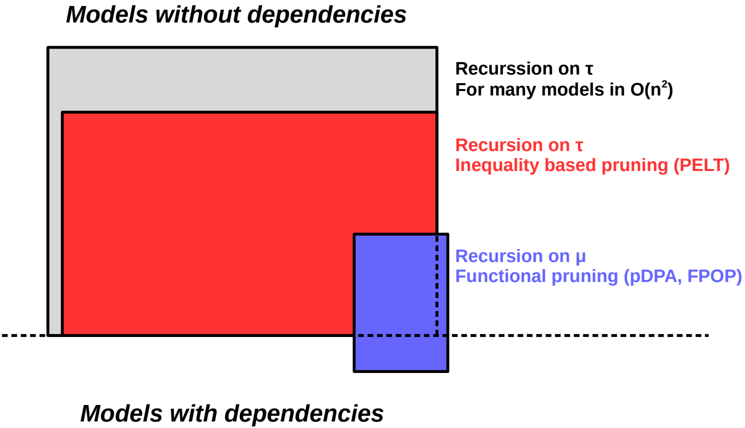
- We vary the value of the penalty (β) and count
  - TP: at least a change in a window of 15 of a true change
  - FP: the number of predicted changes minus the number of TPs



- More simulations in the paper... [Fearnhead and Rigaill 2018]

Outline

Segmentation models and exact DP algorithms



Algorithms minimizing a (penalized) cost

Models without dependencies

- Good statistical properties [Boysen *et al.* 2009, Garreau *et al.* 2018...]
- Computationally efficient: $O(n^2)$ or less

Models with dependencies and constraints

- Statistical properties ?
- Exact algorithms for some models

  [Maidstone *et al.* 2017, Hocking *et al.* 2018, Jewell *et al.* 2019]

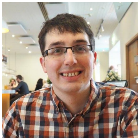
Heuristic optimization

- Necessary to be less than quadratic for complex models
- Some have good statistical properties!

Thank you for listening!

- Vincent RUNGE, Toby HOCKING, Guillaume BOURQUE, Robert MAIDSTONE, Paul FEARNHEAD





- A post-doctoral position is available: <guillem.rigaill@inra.fr>