Computational experiments in Science

Horse wrangling in the digital age

Mathieu Lagrange and Mathias Rossignol

September 19, 2016

http://www.irccyn.ec-nantes.fr/~lagrange

Outline

Outline

Outline

Outline

Horse taxidermy

- public horse
- hidden horses
- Potemkin Villages

Horse taxidermy

- public horse
- hidden horses
- Potemkin Villages

Horse taxidermy

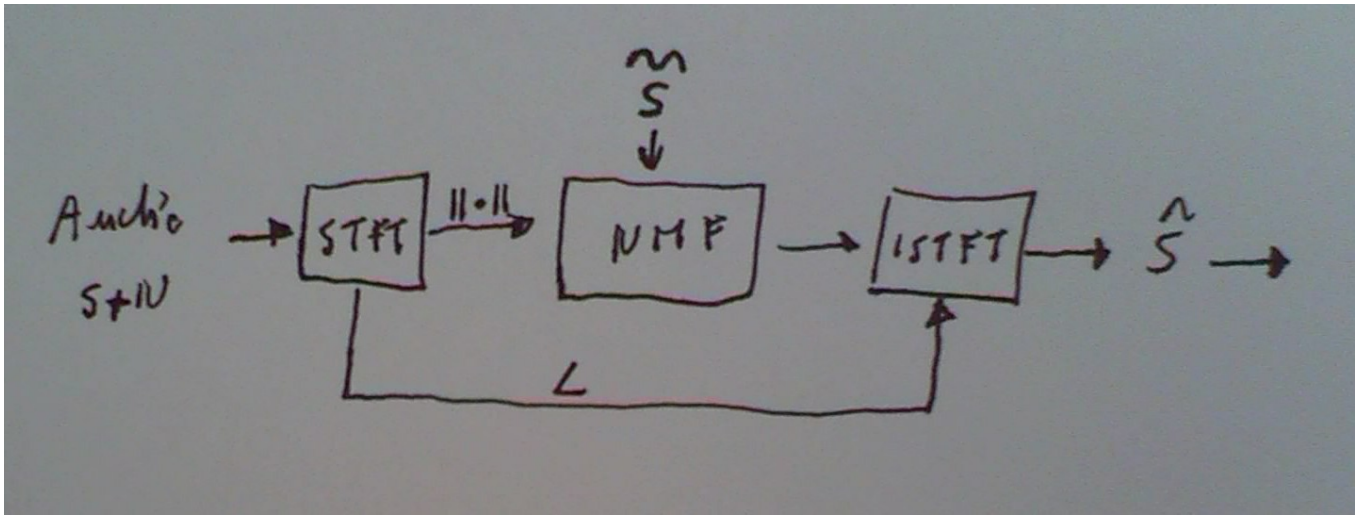
- public horse
- hidden horses
- Potemkin Villages

Horse taxidermy

- public horse
- hidden horses
- Potemkin Villages

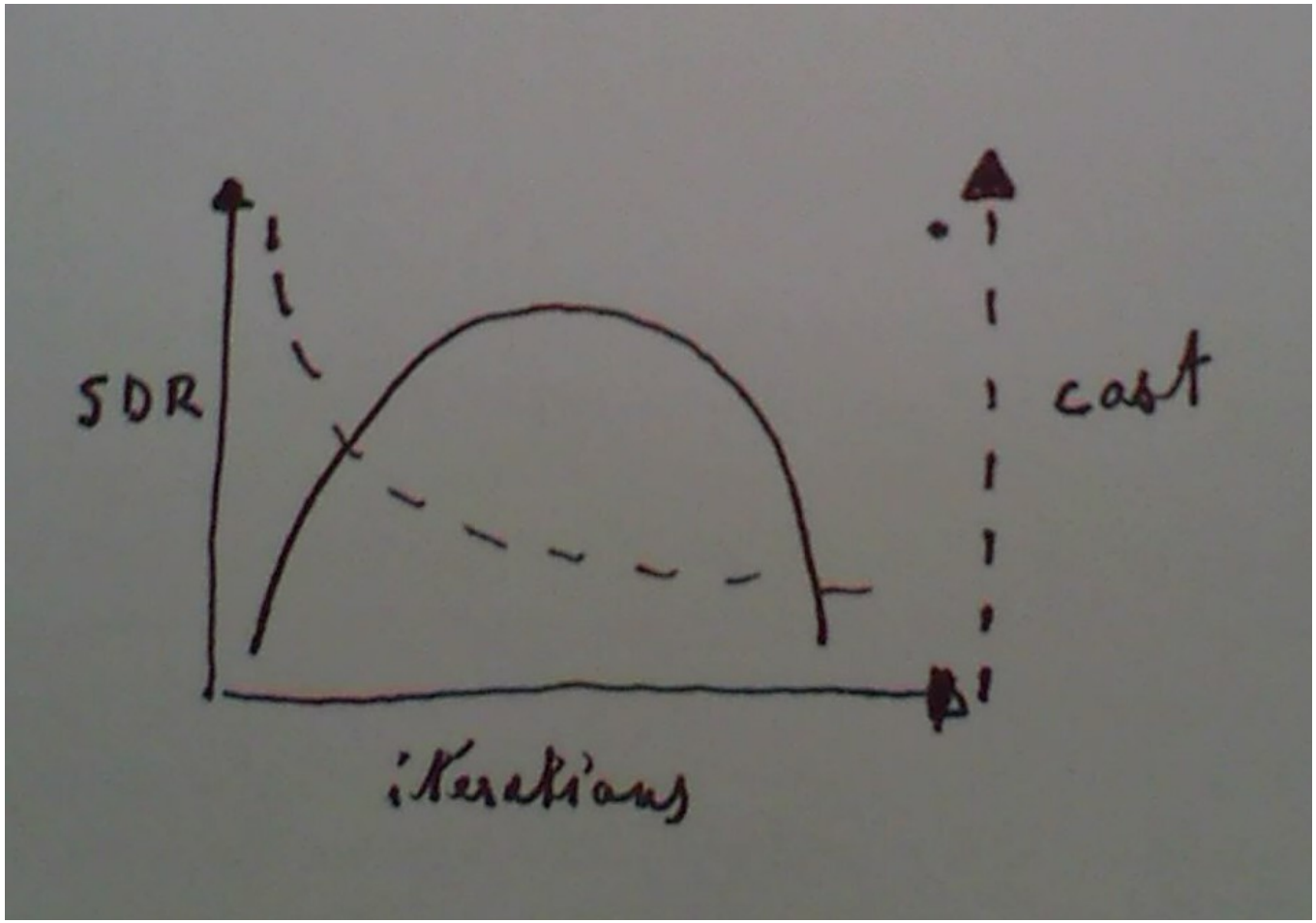
Public horses for Source separation

classic NMF Source Separation scheme



NMF is an iterative scheme, do you set the number of iterations for balancing computational cost and performance ?

Public horses for Source separation



Public horses for Source separation

the use of NMF is an example of a "public" horse

- quality-of-fit optimized criterion
- target criteria (SDR, SAR, SIR)
- heavy use of "early stop" to ensure the non divergence of the target criteria

Public horses for Source separation

the use of NMF is an example of a "public" horse

- quality-of-fit optimized criterion
- target criteria (SDR, SAR, SIR)
- heavy use of "early stop" to ensure the non divergence of the target criteria

Public horses for Source separation

the use of NMF is an example of a "public" horse

- quality-of-fit optimized criterion
- target criteria (SDR, SAR, SIR)
- heavy use of "early stop" to ensure the non divergence of the target criteria

Public horses for Source separation

the use of NMF is an example of a "public" horse

- quality-of-fit optimized criterion
- target criteria (SDR, SAR, SIR)
- heavy use of "early stop" to ensure the non divergence of the target criteria

Why hidden horses

Digital data analysis is still in infancy (less than half a century), so uncontrolled behaviors are

- part of research
- and everyone is more or less happy with it

Why hidden horses

Digital data analysis is still in infancy (less than half a century), so uncontrolled behaviors are

- part of research
- and everyone is more or less happy with it

Why hidden horses

Digital data analysis is still in infancy (less than half a century), so uncontrolled behaviors are

- part of research
- and everyone is more or less happy with it

Why hidden horses

That said, spurious correlations are a well known matters in MANY, MANY, MANY research fields and the fact that the engineering community put so few effort on addressing those matters is childish.
There are many ways our community is severely harming itself

- badly designed tasks
- too small and biased datasets
- unrealistic and non reproducible results

Peter Norvig: Warning Signs in Experimental Design and Interpretation

Why hidden horses

That said, spurious correlations are a well known matters in MANY, MANY, MANY research fields and the fact that the engineering community put so few effort on addressing those matters is childish.
There are many ways our community is severely harming itself

- badly designed tasks
- too small and biased datasets
- unrealistic and non reproducible results

Peter Norvig: Warning Signs in Experimental Design and Interpretation

Why hidden horses

That said, spurious correlations are a well known matters in MANY, MANY, MANY research fields and the fact that the engineering community put so few effort on addressing those matters is childish.
There are many ways our community is severely harming itself

- badly designed tasks
- too small and biased datasets
- unrealistic and non reproducible results

Peter Norvig: Warning Signs in Experimental Design and Interpretation

Why hidden horses

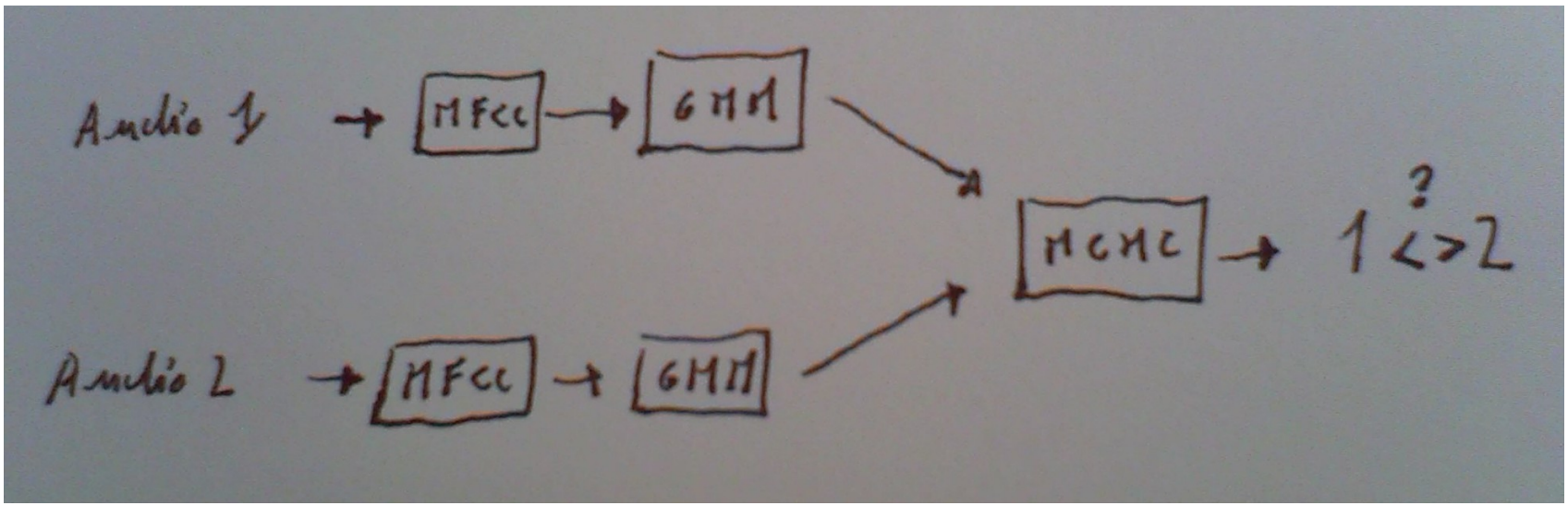
That said, spurious correlations are a well known matters in MANY, MANY, MANY research fields and the fact that the engineering community put so few effort on addressing those matters is childish.
There are many ways our community is severely harming itself

- badly designed tasks
- too small and biased datasets
- unrealistic and non reproducible results

Peter Norvig: Warning Signs in Experimental Design and Interpretation

An example

Acoustic Scene Similarity Retrieval (ASSR)



An example

A Potemkin Village in ASSR

- 2007: near perfect results, solved case
- but strong issue in database design, similar to the album effect in artist / genre recognition
- 2015: new figure, GMMs useless and performance only slightly over chance

---

J.J. Aucouturier & al The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music The Journal of the Acoustical Society of America, 2007, vol. 122, no 2, p. 881-891

M. Lagrange & al The bag-of-frames approach: a not so sufficient model for urban soundscapes The Journal of the Acoustical Society of America, 2015, 138(5), EL487 - EL492

An example

A Potemkin Village in ASSR

- 2007: near perfect results, solved case
- but strong issue in database design, similar to the album effect in artist / genre recognition
- 2015: new figure, GMMs useless and performance only slightly over chance

---

J.J. Aucouturier & al The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music The Journal of the Acoustical Society of America, 2007, vol. 122, no 2, p. 881-891

M. Lagrange & al The bag-of-frames approach: a not so sufficient model for urban soundscapes The Journal of the Acoustical Society of America, 2015, 138(5), EL487 - EL492

An example

A Potemkin Village in ASSR

- 2007: near perfect results, solved case
- but strong issue in database design, similar to the album effect in artist / genre recognition
- 2015: new figure, GMMs useless and performance only slightly over chance

---

J.J. Aucouturier & al The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music The Journal of the Acoustical Society of America, 2007, vol. 122, no 2, p. 881-891

M. Lagrange & al The bag-of-frames approach: a not so sufficient model for urban soundscapes The Journal of the Acoustical Society of America, 2015, 138(5), EL487 - EL492

An example

A Potemkin Village in ASSR

- 2007: near perfect results, solved case
- but strong issue in database design, similar to the album effect in artist / genre recognition
- 2015: new figure, GMMs useless and performance only slightly over chance

---

J.J. Aucouturier & al The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music The Journal of the Acoustical Society of America, 2007, vol. 122, no 2, p. 881-891

M. Lagrange & al The bag-of-frames approach: a not so sufficient model for urban soundscapes The Journal of the Acoustical Society of America, 2015, 138(5), EL487 - EL492

One problem

DATA

Audio Scene Analysis

We need

- lots of data
- well annotated data
- public domain data
- controllable complexity

Audio Scene Analysis

We need

- lots of data
- well annotated data
- public domain data
- controllable complexity

Audio Scene Analysis

We need

- lots of data
- well annotated data
- public domain data
- controllable complexity

Audio Scene Analysis

We need

- lots of data
- well annotated data
- public domain data
- controllable complexity

Audio Scene Analysis

We need

- lots of data
- well annotated data
- public domain data
- controllable complexity

Data simulation

IMHO

- simulated data is ok,
- as long as the aim is to gain knowledge and not to go to production
- simulation is not real data, but also not synthesized data
- tricky part is to choose the right level of abstraction.

Data simulation

IMHO

- simulated data is ok,
- as long as the aim is to gain knowledge and not to go to production
- simulation is not real data, but also not synthesized data
- tricky part is to choose the right level of abstraction.

Data simulation

IMHO

- simulated data is ok,
- as long as the aim is to gain knowledge and not to go to production
- simulation is not real data, but also not synthesized data
- tricky part is to choose the right level of abstraction.

Data simulation

IMHO

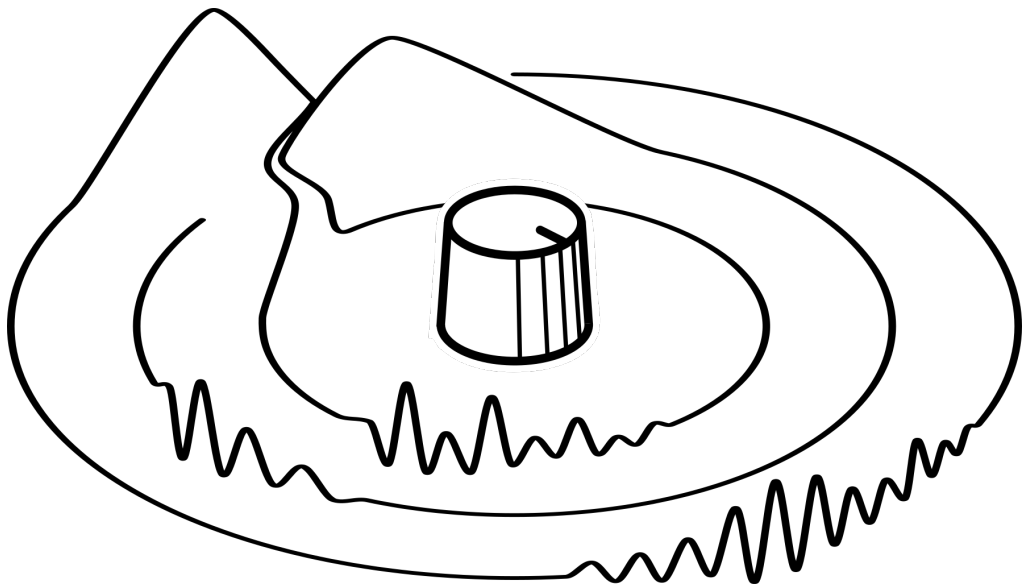
- simulated data is ok,
- as long as the aim is to gain knowledge and not to go to production
- simulation is not real data, but also not synthesized data
- tricky part is to choose the right level of abstraction.

Data simulation

IMHO

- simulated data is ok,
- as long as the aim is to gain knowledge and not to go to production
- simulation is not real data, but also not synthesized data
- tricky part is to choose the right level of abstraction.

SimScene



In a nutshell

SimScene is an acoustic scene simulator

- built as a sequencer
- with abstracted scheduling parameters
- where events are defined as a set of recording samples
- the outcome follows the "skeleton of events on a bed of texture paradigm"

Nelken, I., & de Cheveigné An ear for statistics Nature neuroscience, 16(4), 381-382.

In a nutshell

SimScene is an acoustic scene simulator

- built as a sequencer
- with abstracted scheduling parameters
- where events are defined as a set of recording samples
- the outcome follows the "skeleton of events on a bed of texture paradigm"

Nelken, I., & de Cheveigné An ear for statistics Nature neuroscience, 16(4), 381-382.

In a nutshell

SimScene is an acoustic scene simulator

- built as a sequencer
- with abstracted scheduling parameters
- where events are defined as a set of recording samples
- the outcome follows the "skeleton of events on a bed of texture paradigm"

Nelken, I., & de Cheveigné An ear for statistics Nature neuroscience, 16(4), 381-382.

In a nutshell

SimScene is an acoustic scene simulator

- built as a sequencer
- with abstracted scheduling parameters
- where events are defined as a set of recording samples
- the outcome follows the "skeleton of events on a bed of texture paradigm"

Nelken, I., & de Cheveigné An ear for statistics Nature neuroscience, 16(4), 381-382.

In a nutshell

SimScene is an acoustic scene simulator

- built as a sequencer
- with abstracted scheduling parameters
- where events are defined as a set of recording samples
- the outcome follows the "skeleton of events on a bed of texture paradigm"

Nelken, I., & de Cheveigné An ear for statistics Nature neuroscience, 16(4), 381-382.

Use

- Matlab
- open source:
  `https://bitbucket.org/mlagrange/simscene`
- used in DCASE 2013 and 2016 editions
- produced datasets on archive

Mathieu Lagrange, Grégoire Lafay, Mathias Rossignol, Emmanouil Benetos, Axel Roebel An evaluation framework for event detection using a morphological model of acoustic scenes IEEE TASLP v24-10, 1854-1864

Use

- Matlab
- open source:
  `https://bitbucket.org/mlagrange/simscene`
- used in DCASE 2013 and 2016 editions
- produced datasets on archive

Mathieu Lagrange, Grégoire Lafay, Mathias Rossignol, Emmanouil Benetos, Axel Roebel An evaluation framework for event detection using a morphological model of acoustic scenes IEEE TASLP v24-10, 1854-1864

Use

- Matlab
- open source:
  `https://bitbucket.org/mlagrange/simscene`
- used in DCASE 2013 and 2016 editions
- produced datasets on archive

Mathieu Lagrange, Grégoire Lafay, Mathias Rossignol, Emmanouil Benetos, Axel Roebel An evaluation framework for event detection using a morphological model of acoustic scenes IEEE TASLP v24-10, 1854-1864

Use

- Matlab
- open source:
  `https://bitbucket.org/mlagrange/simscene`
- used in DCASE 2013 and 2016 editions
- produced datasets on archive

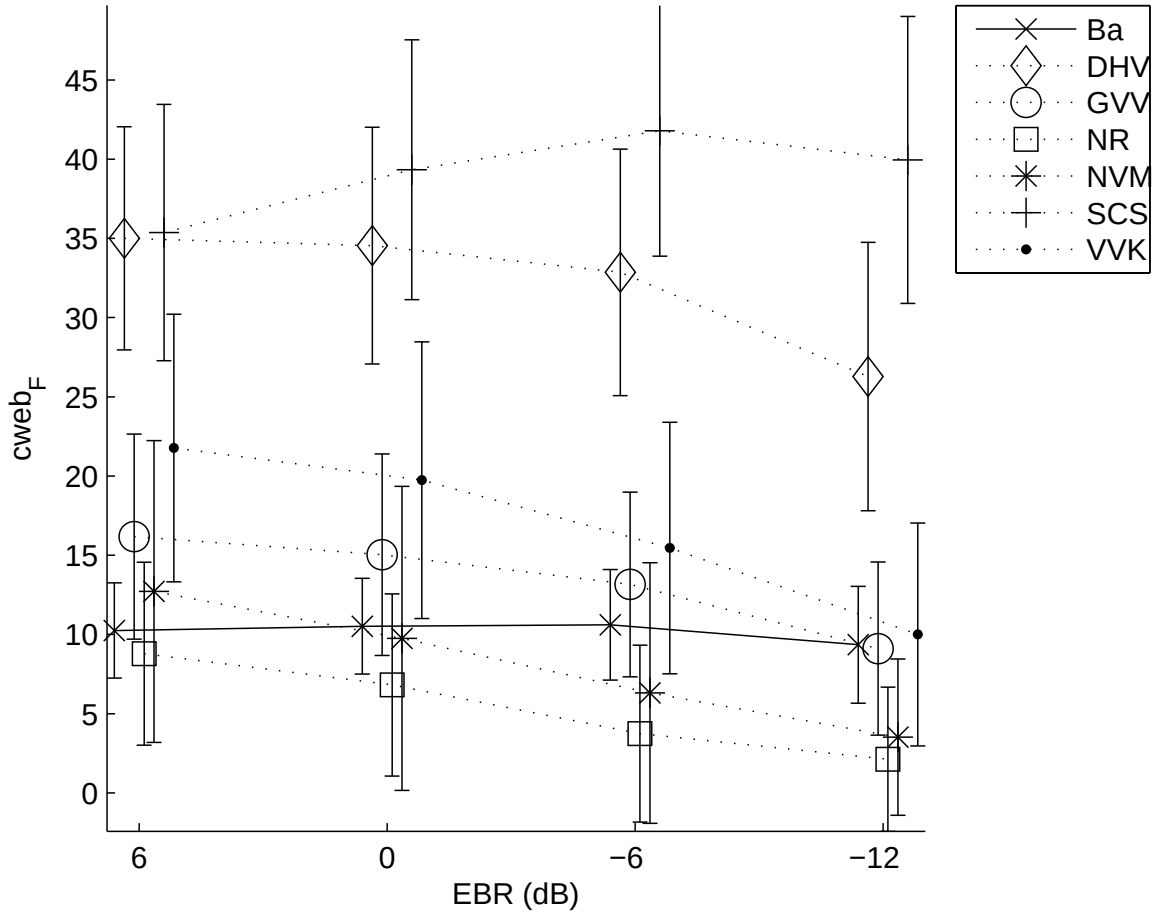
Mathieu Lagrange, Grégoire Lafay, Mathias Rossignol, Emmanouil Benetos, Axel Roebel An evaluation framework for event detection using a morphological model of acoustic scenes IEEE TASLP v24-10, 1854-1864

Use

- Matlab
- open source:
  `https://bitbucket.org/mlagrange/simscene`
- used in DCASE 2013 and 2016 editions
- produced datasets on archive

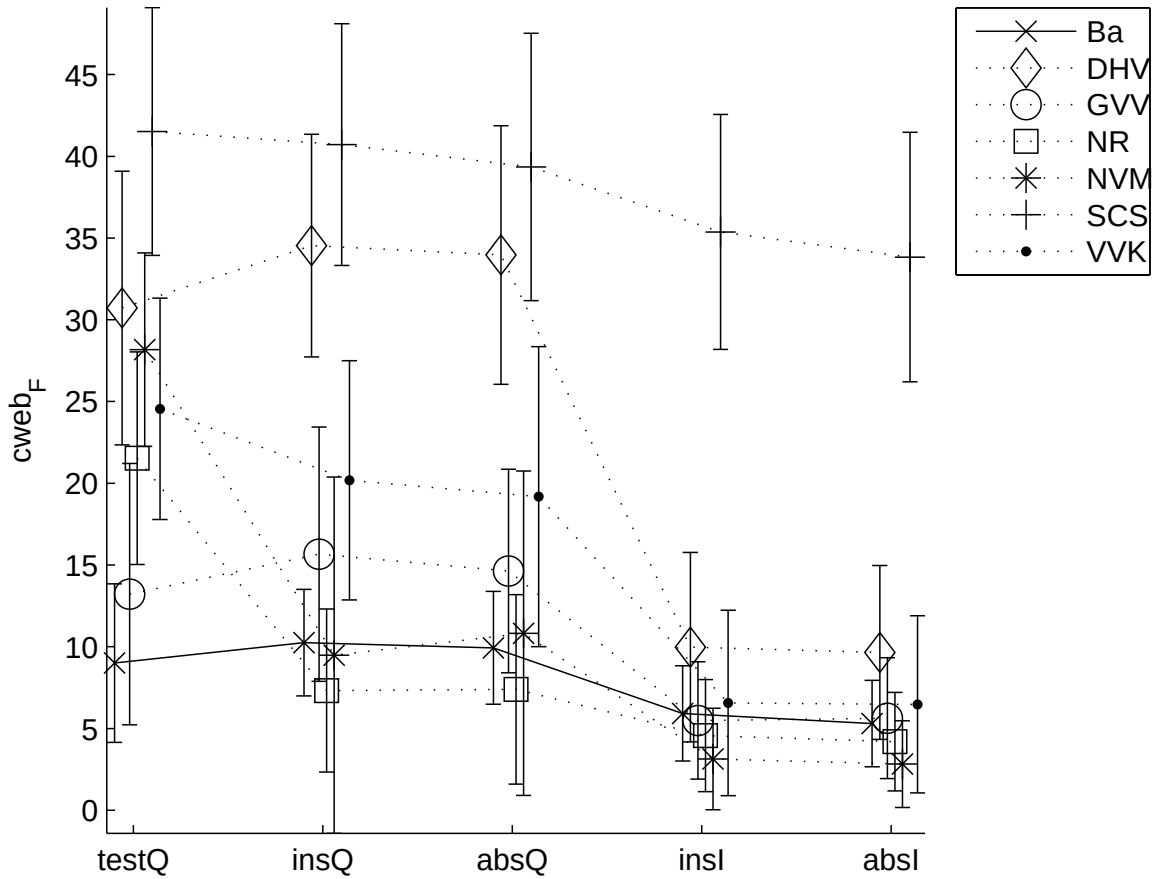
Mathieu Lagrange, Grégoire Lafay, Mathias Rossignol, Emmanouil Benetos, Axel Roebel An evaluation framework for event detection using a morphological model of acoustic scenes IEEE TASLP v24-10, 1854-1864

Going deeper in performance analysis



Varying Event to Background Ratio (EBR)

Going deeper in performance analysis



Changing recording location for events and background

Another problem

PROCESSING

The scientific method

- Analysis: Describe problem
- Hypothesis: Specify solution
- Synthesis: Implement solution
- Validation: Compute performance

The scientific method

- Analysis: Describe problem
- Hypothesis: Specify solution
- Synthesis: Implement solution
- Validation: Compute performance

The scientific method

- Analysis: Describe problem
- Hypothesis: Specify solution
- Synthesis: Implement solution
- Validation: Compute performance

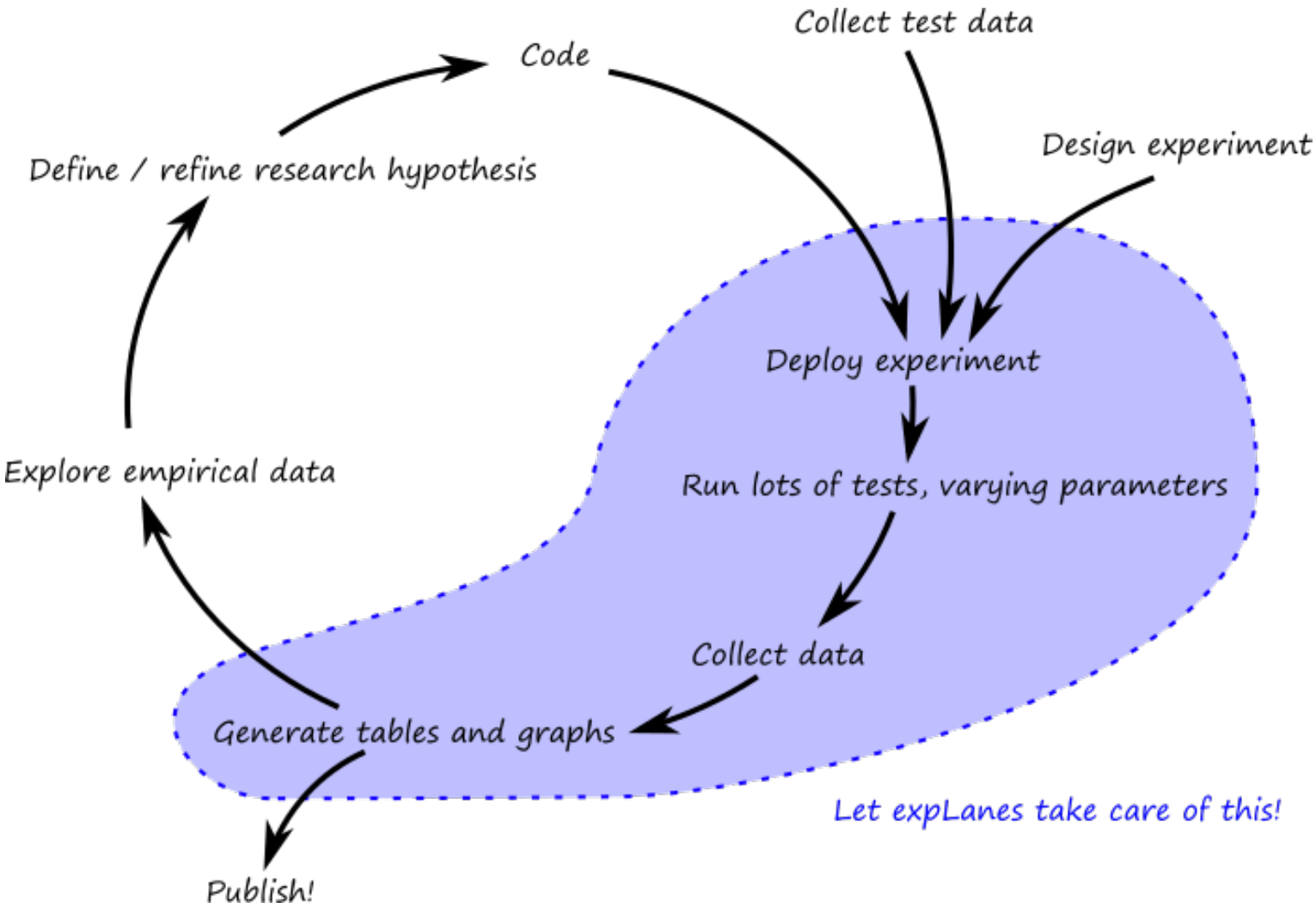
The scientific method

- Analysis: Describe problem
- Hypothesis: Specify solution
- Synthesis: Implement solution
- Validation: Compute performance

The scientific method

- Analysis: Describe problem
- Hypothesis: Specify solution
- Synthesis: Implement solution
- Validation: Compute performance

Why using expLanes ?



In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

In a nutshell

- Design: follow strictly the Design of Experiments paradigm (DoE)
- Code: feed forward pipeline
- Process: multi user, multi core, multi host
- Reduce: fast and convenient thanks to factor masking
- Visualize: Matlab, LaTeX, html

More

- website : `http://mathieulagrange.github.io/expLanes`
- demonstrations: clustering, source separation, ...
- feedback welcome !

More

- website : `http://mathieulagrange.github.io/expLanes`
- demonstrations: clustering, source separation, ...
- feedback welcome !

More

- website : `http://mathieulagrange.github.io/expLanes`
- demonstrations: clustering, source separation, ...
- feedback welcome !

More

- website : `http://mathieulagrange.github.io/expLanes`
- demonstrations: clustering, source separation, ...
- feedback welcome !

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

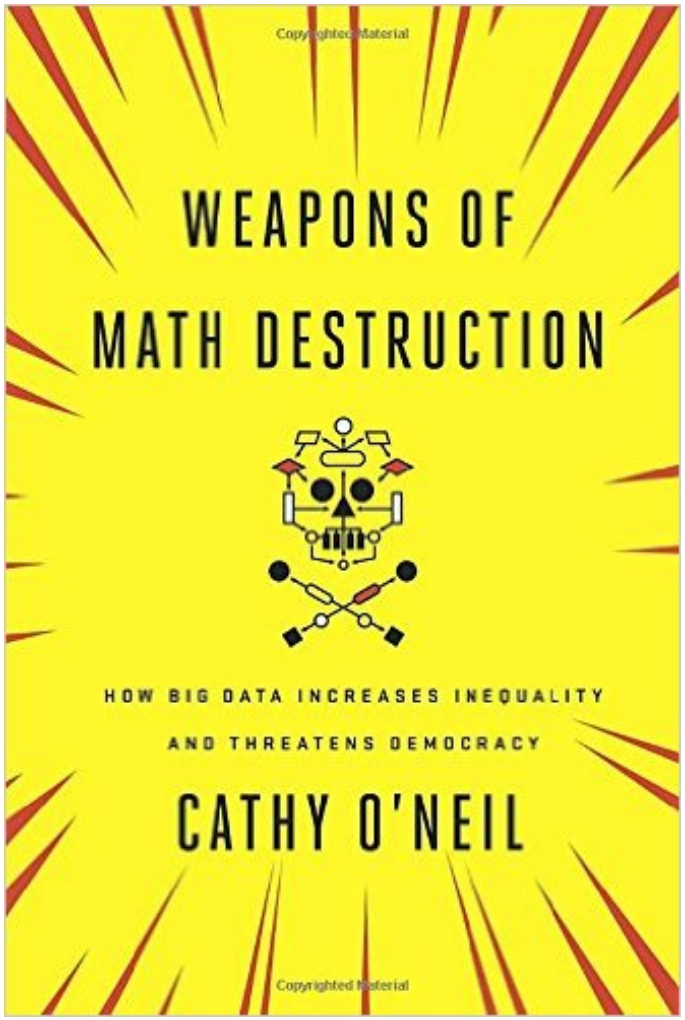
Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Conclusions

- The horse phenomenon is a reality
- Potemkin villages are very common also in emerging fields
- Scientists as individuals are smart humans
- Yet, the social impact is very strong
- As a community, the scientific method must be enforced
- Strict reproducible research is the way to go
- Need tools for that, not only languages

Cathy O'Neil


Copyrighted Material
WEAPONS OF
MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY
CATHY O'NEIL
Copyrighted Material

Lorena Barba

https://figshare.com/articles/Reproducibility_PI_Manifesto/104539

Victoria Stodden

one of three "reproducibility editors" appointed to look over code and data sets submitted by authors to the Journal of the American Statistical Association (JASA)