HAL
open science

Realis'2014: Reproductibilité expérimentale pour l'informatique en parallélisme, architecture et système

Lucas Nussbaum, Olivier Richard

To cite this version:

Lucas Nussbaum, Olivier Richard. Realis'2014: Reproductibilité expérimentale pour l'informatique en parallélisme, architecture et système. ComPAS : Conférence d'informatique en Parallélisme, Architecture et Système, Apr 2014, Neuchatel, Suisse. hal-01011401

HAL Id: hal-01011401

https://inria.hal.science/hal-01011401v1

Submitted on 23 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Realis'2014: Reproductibilité expérimentale pour l'informatique en parallélisme, architecture et système

Lucas Nussbaum - Olivier Richard

ComPAS'2014 - 24/04/2014

Reproductibilité

- Extraits de Wikipédia (04/14)
  - La reproductibilité d'une expérience scientifique est **une des conditions** qui permettent d'**inclure les observations** réalisées durant cette expérience dans le processus d'**amélioration perpétuelle des connaissances scientifiques**.
  - Le **critère de reproductibilité** est une des conditions sur lesquelles le philosophe **Karl Popper** distingue le **caractère scientifique d'une étude**.
  - Reproducibility is the ability of **an entire experiment or study to be reproduced**, either by **the researcher** or **by someone else working independently.**

Différents types

- Différents types de reproductibilité [1]
  - **Replications that vary little or not at all with respect to the reference experiment**
    - même méthode, environnement, params → même résultat
  - **Replications that do vary but still follow the same method as the reference experiment**
    - même méthode, mais {env, params} différents → même conclusion
  - **Replications that use different methods to verify the reference experiment results**
    - méthode différente → même conclusion

1. Omar S. GÓMEZ , Natalia JURISTO et Sira VEGAS . "Replications types in experimental disciplines". Dans : ESEM'10. 2010.

Pourquoi se préoccuper de la reproductibilté maintenant ?

- Quelles motivations ?
  - **Aucune**, c'est **une nécessité, un besoin**
  - **Expliciter et « solidifier »** les processus expérimentaux, l'analyse des résultats, l'archivage . . .
- **Complexité des objets d'études**
  - Architecture, système, applicatif
- **Complexité des bancs d'expérimentation**
  - Complexité des plates-formes pour l'expérimentation matérielles/logiciels
    - **Grid'5000 , Emulab, Senslab, PlanetLab. . .**
    - **Piles logicielles**, chaine de compilation, benchmarks, injecteur de charges, recueil de traces. . .
    - **Interactions** difficillement maitrisables.

Devient un **objet d'étude en soi**

- Besoins d'outils spécifiques
  - Outils de **controle d'expérience**
  - Outils de **description**, de trace, de construction d'environnement
  - **Tests** de la reproductibilité partielle ou totale
  - *Litterate Programming*
  - Article exécutable
  - Maitrise et **optimisation des plans d'experiences**
  - ...
- **Pas d'approche / de solution globale**
- **Des niveaux de préoccupation différents**
  - Production de résulats d'expériences
    - complexité, cout, disparition du matériel
  - Analyses des résultats
    - article reproductible (cf Présentation Arnaud Legrand / Luka Stanisic)

Différents mouvements autour de la reproductibilité

- Regain d'intérêt dans **plusieurs disciplines**
  47 out of 53 medical research papers on the subject of cancer were irreproducible[2]
- **Reproducible Research**
- **Open Science / Open data**
  - Partager les données et les outils
- **Développements d'outils**
- **Développements des plate-formes pour l'expérimentation**
- **Archivage et pérennisation**
  - figshare, myexperiment, RunMyCode...

2. Begley, C. G. ; Ellis, L. M. (2012). « Drug development : Raise standards for preclinical cancer research ». Nature 483 (7391) : 531–533.

Realis'2014

Objectifs :

- Évaluer la reproductibilité des articles soumis à ComPAS
- Sensibiliser la communauté à la reproductibilité expérimentale
- Se confronter aux difficultés pratiques :
  Comment décrire une expérience pour en permettre la reproductibilité ?

Realis'2014

Objectifs :

- Évaluer la reproductibilité des articles soumis à ComPAS
- Sensibiliser la communauté à la reproductibilité expérimentale
- Se confronter aux difficultés pratiques :
  Comment décrire une expérience pour en permettre la reproductibilité ?

Processus :

1. Les auteurs soumettent leur description d'expérience à Realis (article enrichi, article supplémentaire, archive)
   $\Rightarrow$ 3 Parallélisme, 2 Architecture, 1 Système
2. Les auteurs reproduisent les expériences d'un autre article ($\rightsquigarrow$ rapport)
   Objectifs :
   - Réussir à reproduire (à l'identique) l'ensemble des résultats originaux, sans avoir à contacter les auteurs
   - Évaluer la qualité (flexibilité) de l'approche pour la reproductibilité : changer les paramètres ?

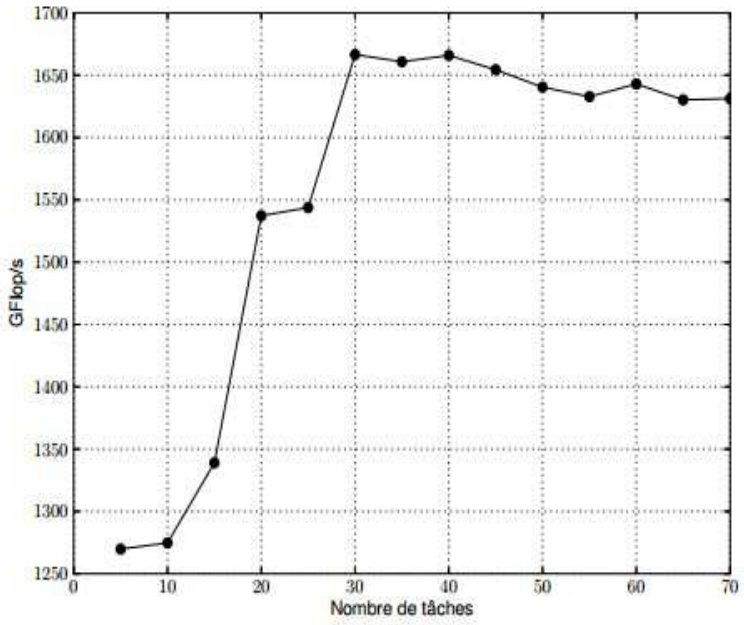
   Si nécessaire, contacter les auteurs

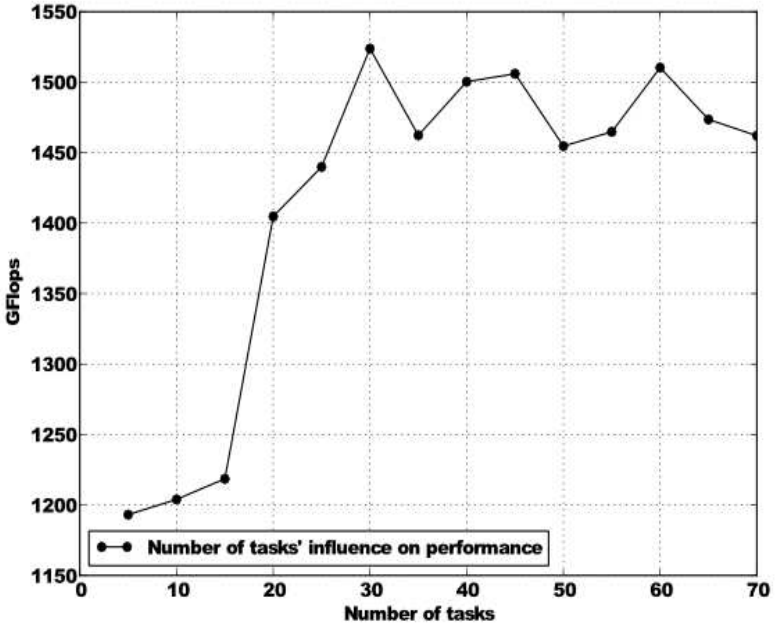
Six articles (plus ou moins) reproductibles

- Parallélisme :
  - Karaoui Moahmed Lamine, Quentin Meunier, Franck Wajsburt and Alain Greiner. *Mécanisme de synchronisation scalable à plusieurs lecteurs et un écrivain*
  - Beat Wolf, Pierre Kuonen and Thomas Dandekar. *POP-Java : Parallélisme et distribution orienté objet*
  - Marc Sergent and Simon Archipoff. *Modulariser les ordonnanceurs de tâches : une approche structurelle*
- Architecture :
  - Kathy Porada, David Parello and Bernard Goossens. *Analyse et réduction du chemin critique dans l'exécution d'une application*
  - Rahmoune Djallal, David Parello and Bernard Goossens. *Exécution en parallèle*
- Système :
  - Emmanuel Quemener. *Déduplication extrême d'OS avec SIDUS : un petit pas pour la reproductibilité ?* (poster)

Mais de nombreux problèmes

- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :





Mais de nombreux problèmes

- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :

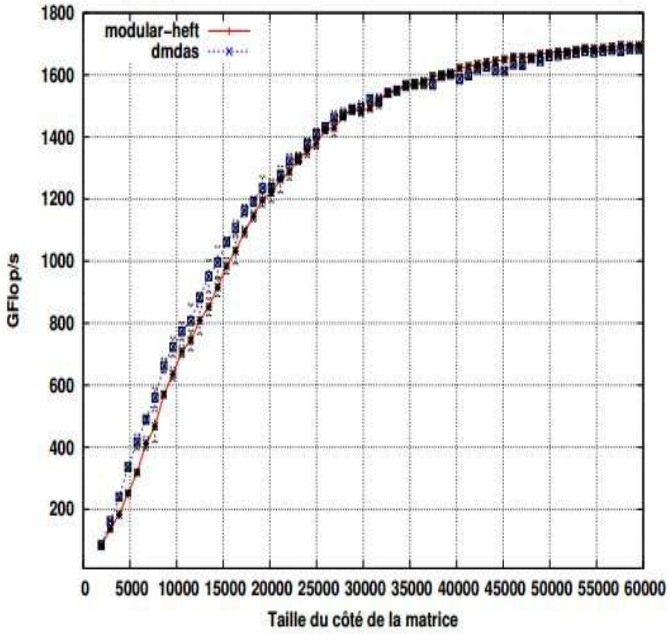


Illustration 3: Figure 4 dans l'article

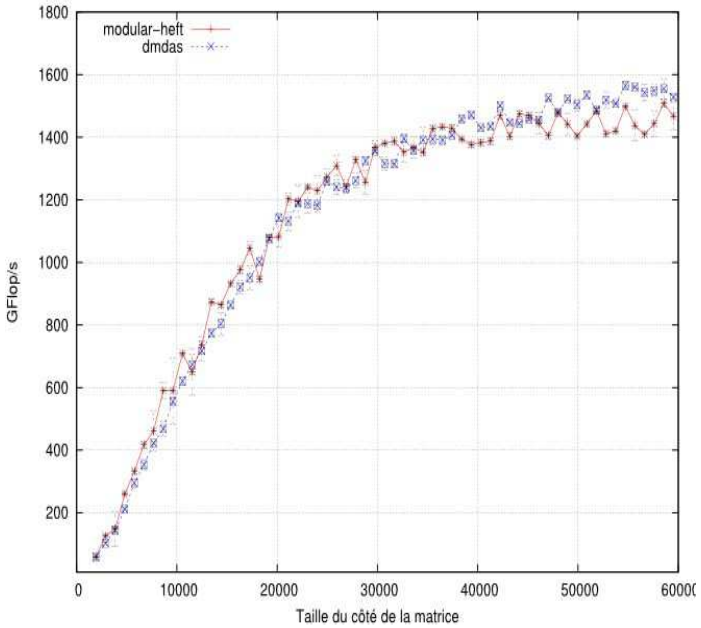
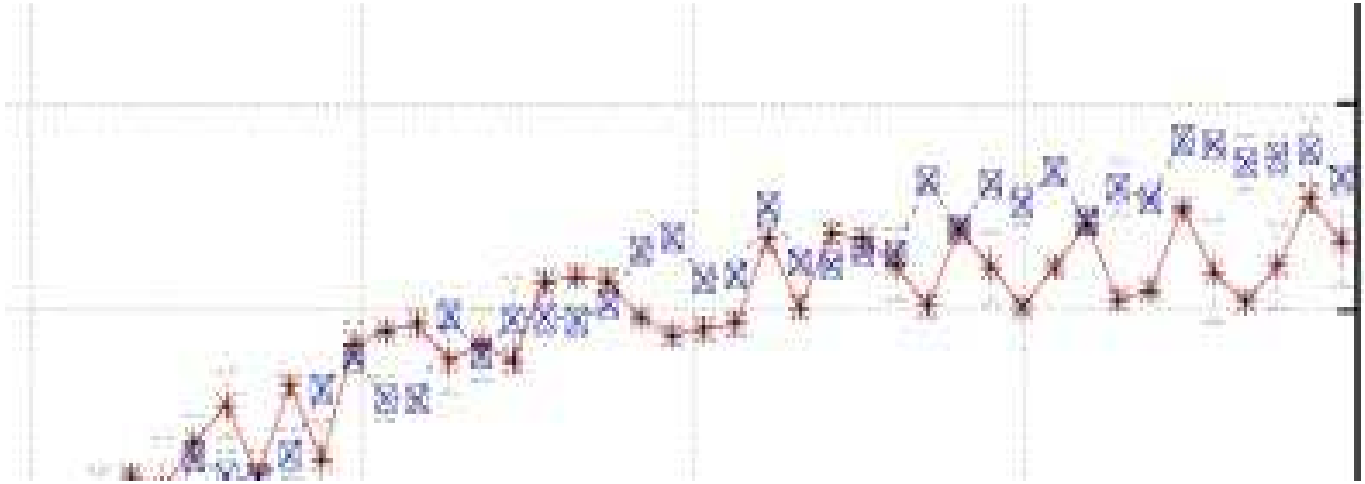


Illustration 4: Figure 4 reproduite

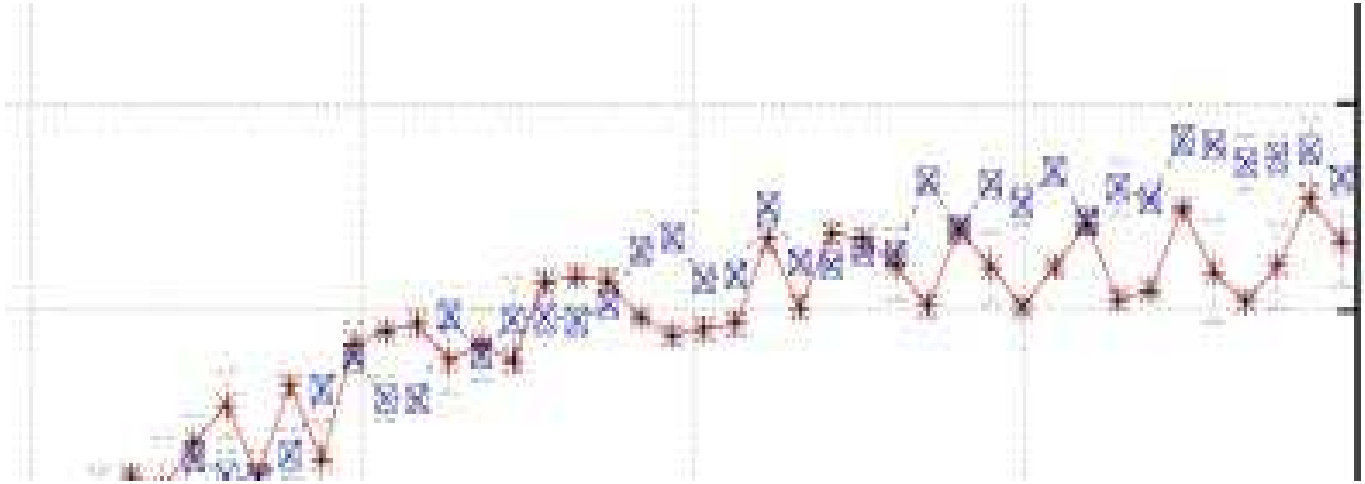
Mais de nombreux problèmes

- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :



Mais de nombreux problèmes

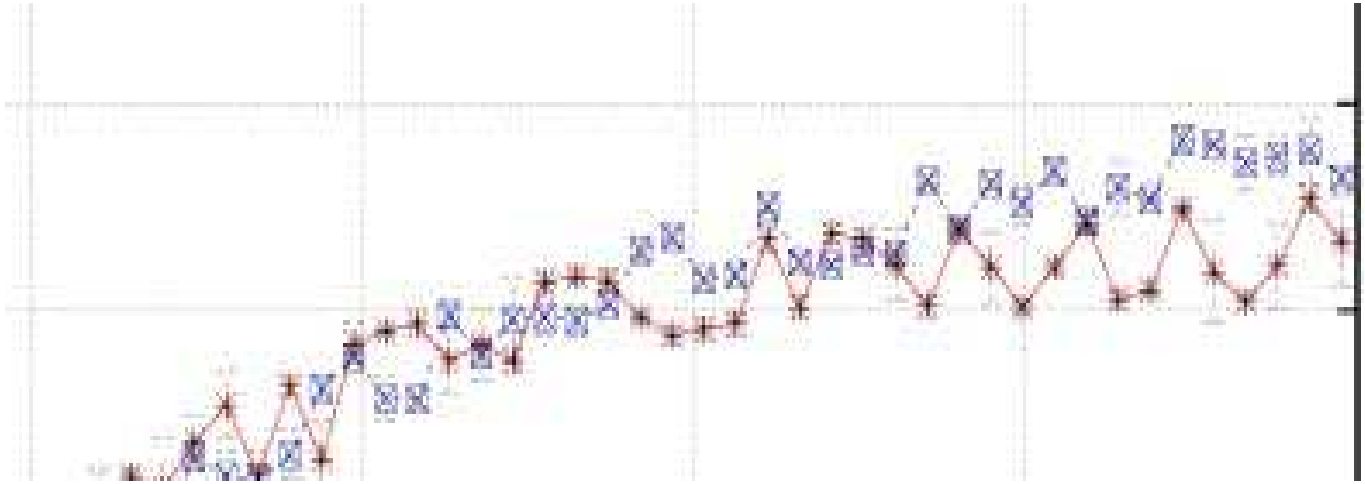
- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :



- Variabilité $\nearrow$, performances $\searrow$

Mais de nombreux problèmes

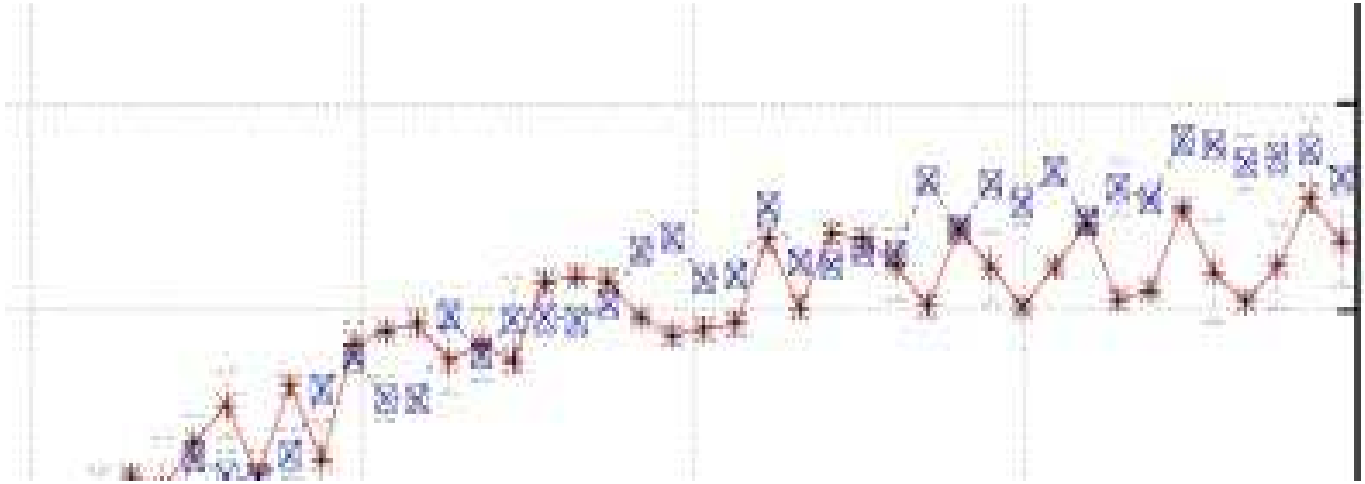
- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :



- Variabilité ↗, performances ↘ mais même machine !
- Version de CUDA ? Problème de refroidissement ?

Mais de nombreux problèmes

- Erreurs de compilation et d'exécution : version de gcc, sh vs bash
- Résultats liés à la version de GCC (?) – non précisée dans l'article
- Résultats exprimés sous forme de mesures de temps d'exécution (spécifique à la machine, mais machine inaccessible)
- Des résultats différents :



- Variabilité ↗, performances ↘ mais même machine !
- Version de CUDA ? Problème de refroidissement ?

Malgré un contexte très favorable !

Des challenges pour l'avenir

- Comment capturer et décrire l'environnement des expériences ?
- Quels formalismes et outils pour décrire et exécuter les expériences ?
- Quel support de la part des outils et plates-formes expérimentales ? (provenance ?)
- Où publier les informations, instructions, codes et données de manière pérenne ? (HAL ?)
- Comment valoriser la reproductibilité ?

Liste de diffusion sur la recherche reproductible :
`http://listes.univ-orleans.fr/sympa/info/recherche-reproductible`