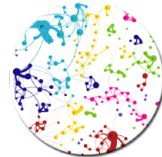




CNRS - INP - UT3 - UT1 - UT2J

Institut de Recherche en Informatique de Toulouse

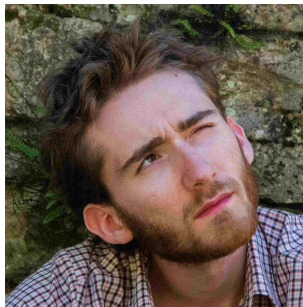


Digital Sufficiency in Data Centers

Maël Madon
ROADEF'24 @ Amiens
March 7, 2024

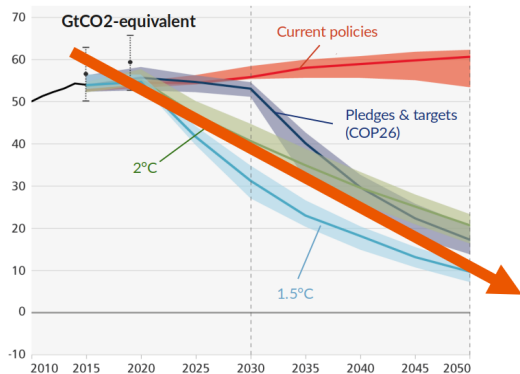
- Maël Madon, 2nd year PhD student
- SEPIA team at IRIT (Toulouse)
- Supervisors: Georges Da Costa and Jean-Marc Pierson

SEPIA



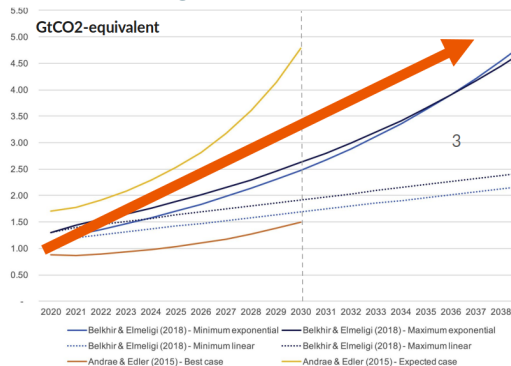
Context: growing greenhouse gases emissions

What we need to do:



Source: Valérie Masson-Delmotte, data from IPCC

ICT industry =
2-4% global emissions

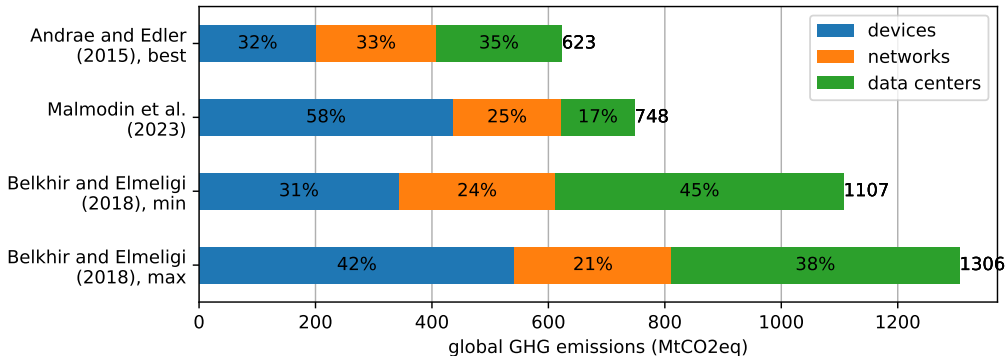


Source: Freitag et al. 2021



Where do ICT emissions come from?

Global GHG emission estimates for ICT industry in 2020:





Footprint reduction in data center

Traditional techniques:

- **Energy efficiency** (virtualization, workload consolidation, DVFS...)
- **Use of renewable energies** (workload adaptation to power envelope, geographic load shifting, ...)
- **Data center environment** (cooling management, waste heat utilization, use of batteries)



www.pexels.com



Efficiency techniques

Before



After





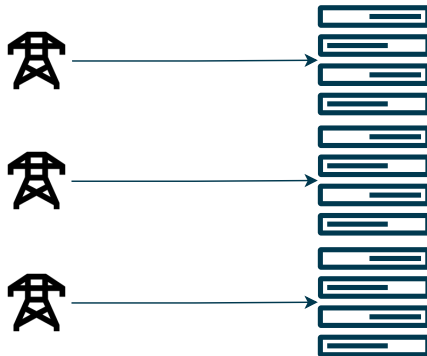
Rebound effect?

Before



Data center = 1
Energy = 2

After

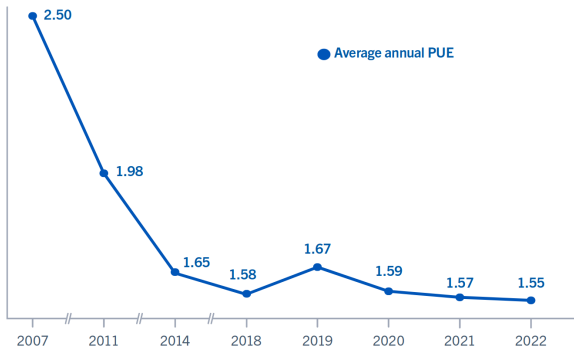


Data center = 3
Energy = 3



Rebound effect

What is the average annual PUE for your largest data center? (n=669)



Source: *Uptime Institute Global Data Center Survey 2022*

Global trends in digital and energy indicators, 2015-2022

	2015	2022	Change
Internet users	3 billion	5.3 billion	+78%
Internet traffic	0.6 ZB	4.4 ZB	+600%
Data centre workloads	180 million	800 million	+340%
Data centre energy use (excluding crypto)	200 TWh	240-340 TWh	+20-70%
Crypto mining energy use	4 TWh	100-150 TWh	+2300-3500%
Data transmission network energy use	220 TWh	260-360 TWh	+18-64%

Source: *International Energy Agency*



- Efficiency is not enough: **sufficiency**

Sufficiency policies (IPCC, 2022)

A set of measures and daily practices that **avoid demand** for energy, materials, land and water **while delivering human well-being** for all within planetary boundaries.

- **What would "sufficiency" mean for data centers?**
=> voluntary limitation, empower and involve the user



Data center user?

Data center



Submission interface (IaaS, PaaS)



*job submissions,
VM reservations*



Direct users
(developer, cloud provider,...)

Cloud solution (SaaS)



User interface App, mail client, web browser...

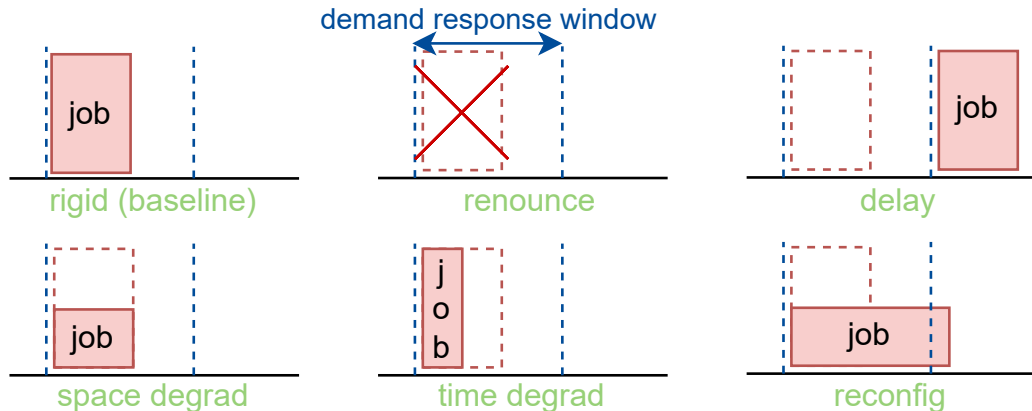


Indirect users
(professional or private use)

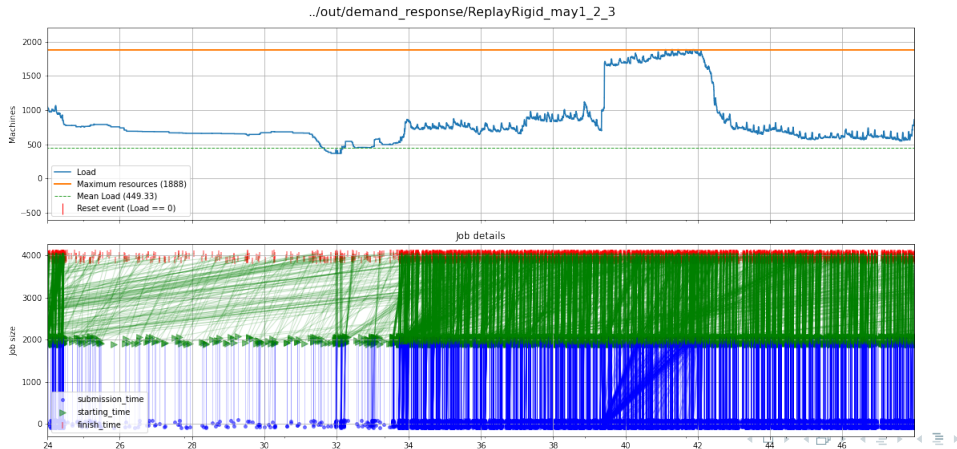
(cloud context)



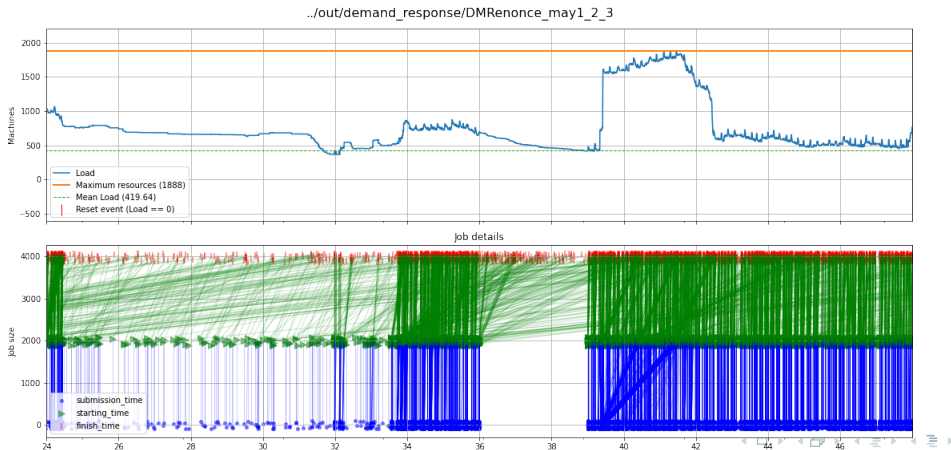
“Sufficiency behaviors” for direct data center users



■ Behavior during demand response window: **rigid**



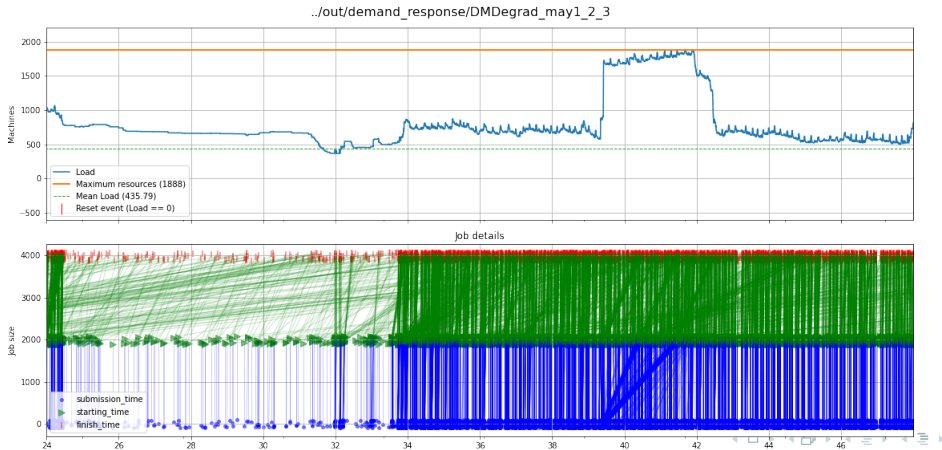
■ Behavior during demand response window: **renounce**



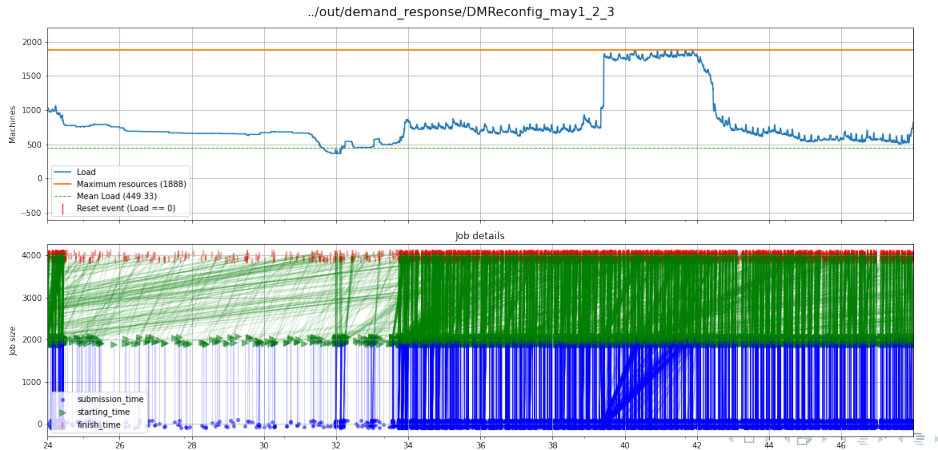
■ Behavior during demand response window: **delay**



■ Behavior during demand response window: **degrad**



■ Behavior during demand response window: **reconfig**



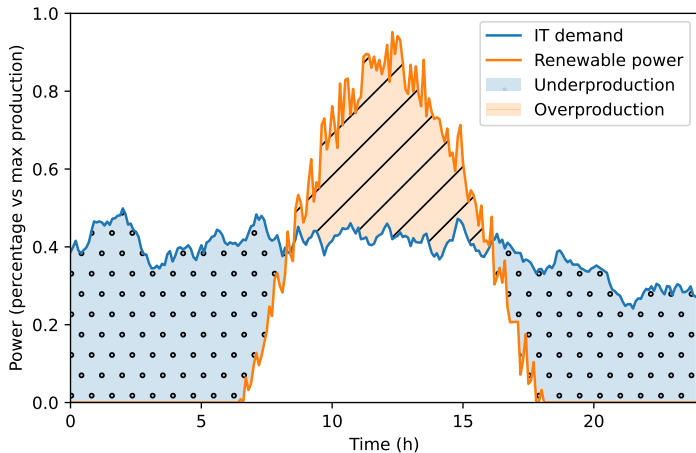
- Pros and cons of each behavior:

behavior	energy in	energy overall	sched. metrics	"acceptability"
renounce	1st	1st	1st	5th
delay	1st	5th	5th	2nd
space degrad	3rd	3rd	2nd	3rd
reconfig	3rd	4th	4th	1st
time degrad	5th	2nd	2nd	3rd

- All the details in the article¹
- Open-source software and reproducible experiments:
gitlab.irit.fr/sepia-pub/open-science/demand-response-user

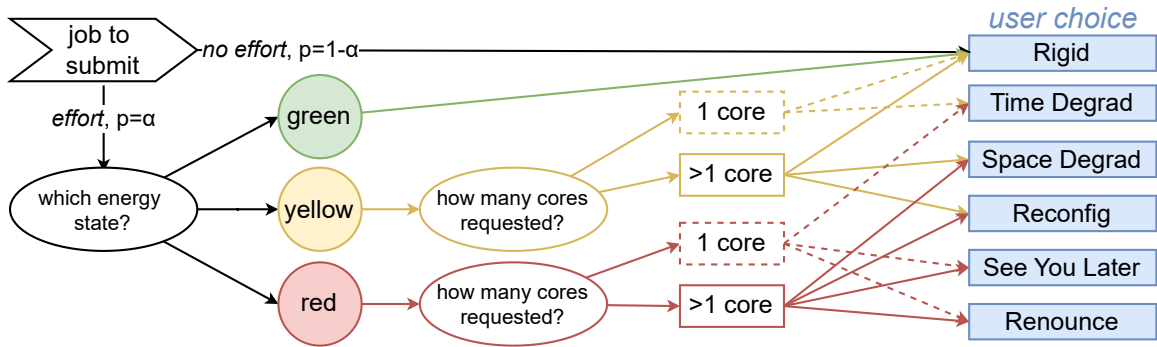
¹M. Madon, G. Da Costa, and J.-M. Pierson, "Characterization of Different User Behaviors for Demand Response in Data Centers," in Euro-Par 2022, doi: [10.1007/978-3-031-12597-3_4](https://doi.org/10.1007/978-3-031-12597-3_4)

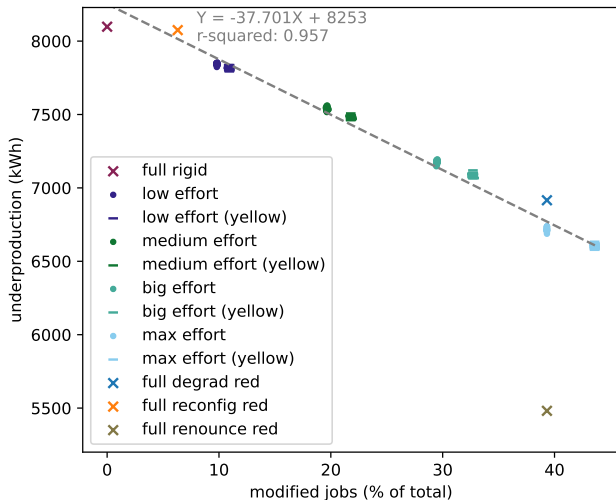
Sufficiency behaviors in a renewable context





Sufficiency behaviors in a renewable context





■ Article in review:
J. Gatt, M. Madon, and G. Da Costa,
*"Digital sufficiency behaviors to deal
with intermittent energy sources in
data center."*



Data center user?

Data center



Submission interface (IaaS, PaaS)



*job submissions,
VM reservations*



Direct users
(developer, cloud provider,...)

Cloud solution (SaaS)



User interface App, mail client, web browser...



Indirect users
(professional or private use)

(cloud context)



Digital Sufficiency in Flexible Work

- **Study goal:** Re-design the use of cloud services for flexible working towards sufficiency
- How much digital interventions are *necessary* and how much is *superfluous*?
- **Method:** Focus groups within companies

What is "flexible work" for you?

Describe the organisation of your working week



work-life balance



working hours



work locations

For each cloud-based daily task
Could you do without?

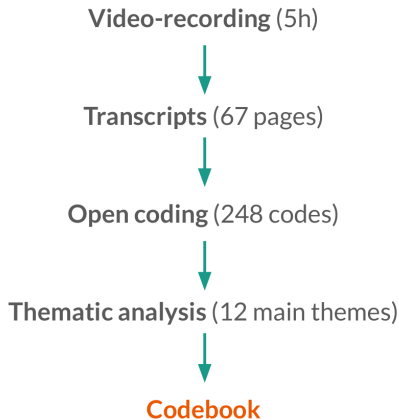
- If yes, how?
- If no, why?

■ Data collection:

- 3 focus groups (Netherlands)
- 2 companies (small IT company / large consulting firm)
- 11 participants (consultants, developers, HR, ...)



■ Data analysis:





Conclusion

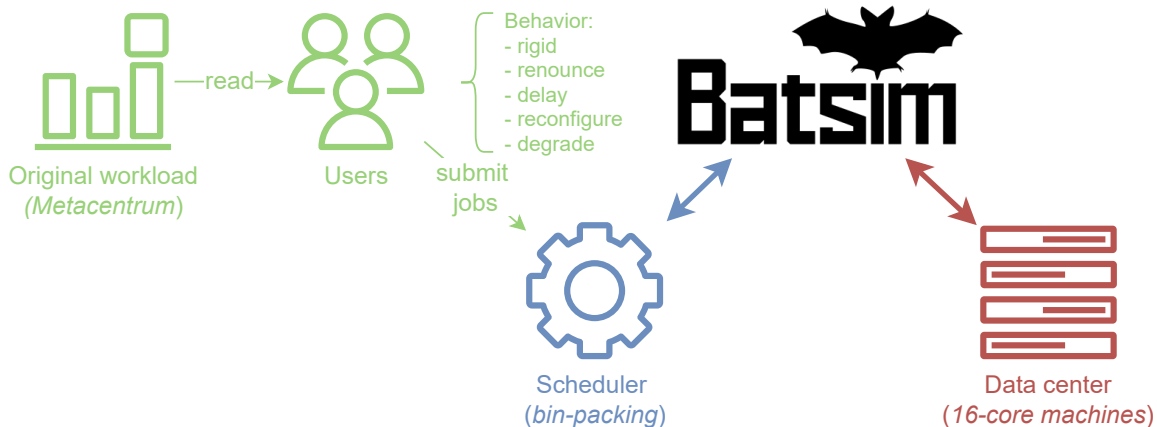
- Going beyond *efficiency*, investigating *sufficiency* for data centers
 - through **simulation** and **qualitative research** methods
- the link between the two works is quite distant
 - simulation: user HPC, direct actor
 - qualitative study: user SaaS, does she have the *agency*?
- Come to my **PhD defense: April 30, 14:00 !!**
- Do not hesitate to contact me :-):
 - www.irit.fr/~Mael.Madon
 - mael.madon@irit.fr



Extra slides



The simulated system





Fluid and residual mass

