# Carbon Lifecycle Analysis for Scientific Computing

**NeIC NT1 Manager**
**Mattias Wadenstein**
**<maswan@ndgf.org>**

2024-12-12
WLCG Sustainability Workshop
CERN

NORDIC E-INFRASTRUCTURE COLLABORATION

# Overview

- Question
- Methodology
- Assumptions
- Calculations
- Conclusions

Authors: Prof. Wim Vanderbauwhede (University of Glasgow)
Mattias Wadenstein (NeIC NT1/Umeå University)

# Question

- What is the optimal replacement cycle for scientific computing hardware?
  - From a total $CO_2e$ emissions point of view
- Requires us to make a carbon lifecycle analysis
  - Manufacture
  - Transport
  - Operations
  - Scrapping
- Is changing to modern more efficient computers good to minimize emissions?

# Methodology

- We have made a model for total emissions over time taking into account the following factors:
  - periodical replacement of compute hardware
  - time evolution of electricity carbon intensity
  - time evolution of compute hardware energy efficiency
  - time evolution of compute hardware embodied carbon
  - load
  - PUE
  - heat reuse and other emissions
  - cluster expansion

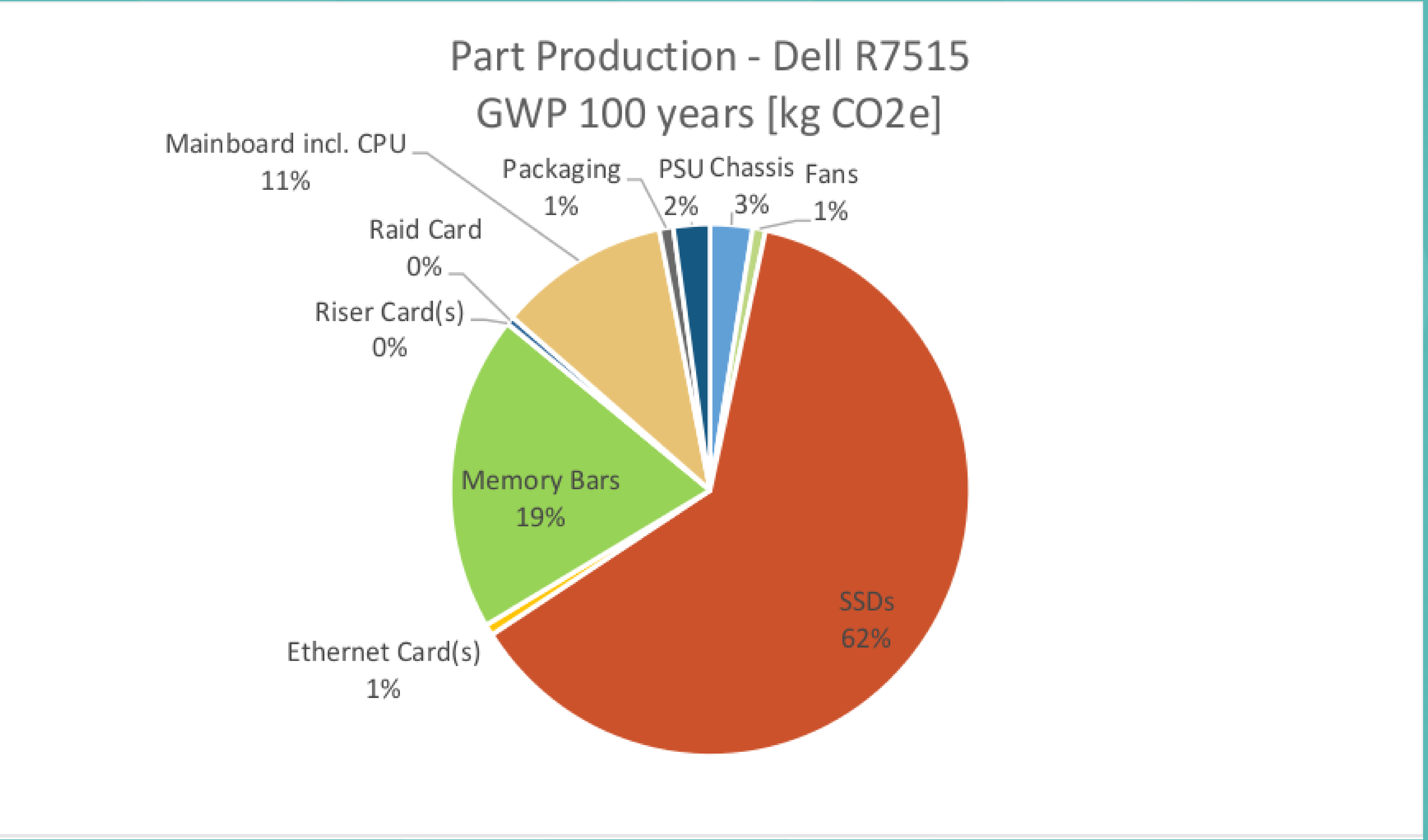
# Assumptions

- The data on equipment comes from vendors
  - Dell, HPE, Lenovo all have published impact of a few select models
  - Covering: Manufacture, transport, recycling
  - Totally dominated by manufacturing!
  - Of this chip manufacturing (CPU, RAM, SSD) is 80+%
  - Some assumptions needed to map general purpose servers to HPC
- The vendor documents also have ops numbers
  - But these are not very applicable to scientific computing
  - Assuming low CPU load, low efficiency compute rooms, dirty power
- We can find real load and facility numbers!

# Assumptions

- We base embodied carbon on Dell R7515
  - Technical report from Dell
  - This model has less RAM and cores, but more SSD than typical compute nodes
  - We assume (based on deatails of CPU, RAM, SSD carbon intensity) that this works out to roughly the same per node
  - Not a big selection of server models with embodied carbon data
- Alternative approach would be own estimates
  - Based on die area, industry average emissions for a given process node, etc
- Or actual vendor data (currently lacking)

# Assumptions



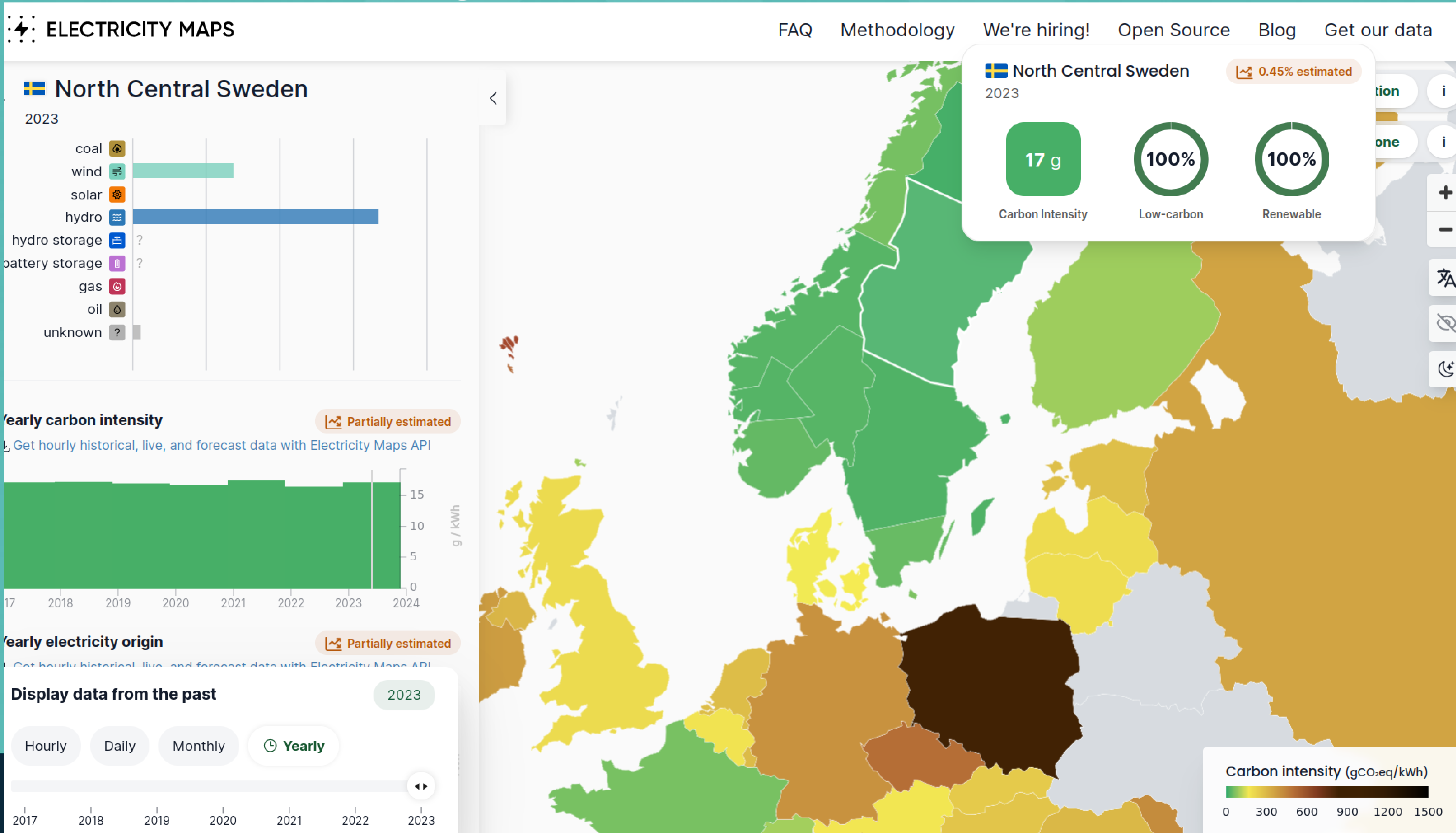
# Assumptions

- The carbon intensity in manufacturing will likely decrease somewhat per effective unit of computing
  - We have a parameter for this, but it is set fairly conservatively
  - No good data on this that we know of
- We assume constant emissions per electricity area, but these are likely to improve over time in most places
  - For political and non-political reasons

# Compute node assupmtions

- Spring 2024 best in class AMD Bergamo CPU nodes with 2x128 cores, 4GB ram/core
  - Current best in class might be AMD Turin, with 14000 HS23 and probably less than 2200W avg power draw on a dual 9965, we lack power measures though
- Benchmark load: HEPScore23
  - Known to scale well with LHC computing
  - Node score 7500
  - Average node power draw during benchmark: 1200 W
  - Or, 160 W/kHS23
  - Numbers from D. Britton, HEPiX Fall 2023
- Assuming 80% of the benchmark as lifetime usage
  - Batch system fill over time, downtimes, IO bottlenecks, etc

# Power generation emissions



# Power generation emissisons

- Data from: app.electricitymaps.com
- Looking at 2023 average for electricity production and imports in that area
- Does not reflect green power purchasing
  - Facilities buying green power look a bit worse here than what they are paying for
- Does not reflect marginal generation or exports
- There are higer emission areas than any of our sample data centers

# Marginal power consumption

- If HPC2N in North Central Sweden draws 1 kWh more from the grid, what's the impact?
  - 1 kWh more generated in the same power mix
  - 1 kWh more generated with different power mix
  - 1 kWh less exported to other power areas
- Range:
  - 1 g/kWh (non-fossil production, as per contract)
  - 11 g/kWh (margin power likely hydro, CO2e mostly land use effect)
  - **17 g/kWh (long-term average) ← this is what this paper uses**
  - 460 g/kWh (more gas burned in Denmark)
  - 1100 g/kWh (more coal burned in Poland)

# Facility cooling

- Mostly additional electricity to drive fans, pumps, and compressors, measured as PUE
- One facility has explicit $CO_2e$/kWh from use of district cooling
- Heat reuse in cold locations explored in one scenario

# Facility numbers – real world

- HPC2N, Sweden
  - 17 gCO2e/kWh electricity
  - PUE 1.03 + 3.6 gCO2e/kWh district cooling
- Vega, Slovenia
  - 247 gCO2e/kWh electricity
  - PUE 1.13
- BNL, USA
  - 282 gCO2e/kWh electricity
  - PUE 1.35
- ASGC, Taiwan
  - 535 gCO2e/kWh electricity
  - PUE 1.62

# Facility numbers - hypothetical

- HPC2N, Sweden – heat reuse
  - Assuming we run heat pumps to heat the university campus for the cold months of the year
  - Higher PUE (compressors drawing electricity): 1.33
  - Leading to estimated reduction of 40 gCO2e/kWh yearly average emissions due to offsetting district heating
  - -17.4 gCO2e/kWh
- French Vega
  - Assuming identical computer facility as Vega
  - But running on the French power mix of mostly nuclear power
  - 53 gCO2e/kWh

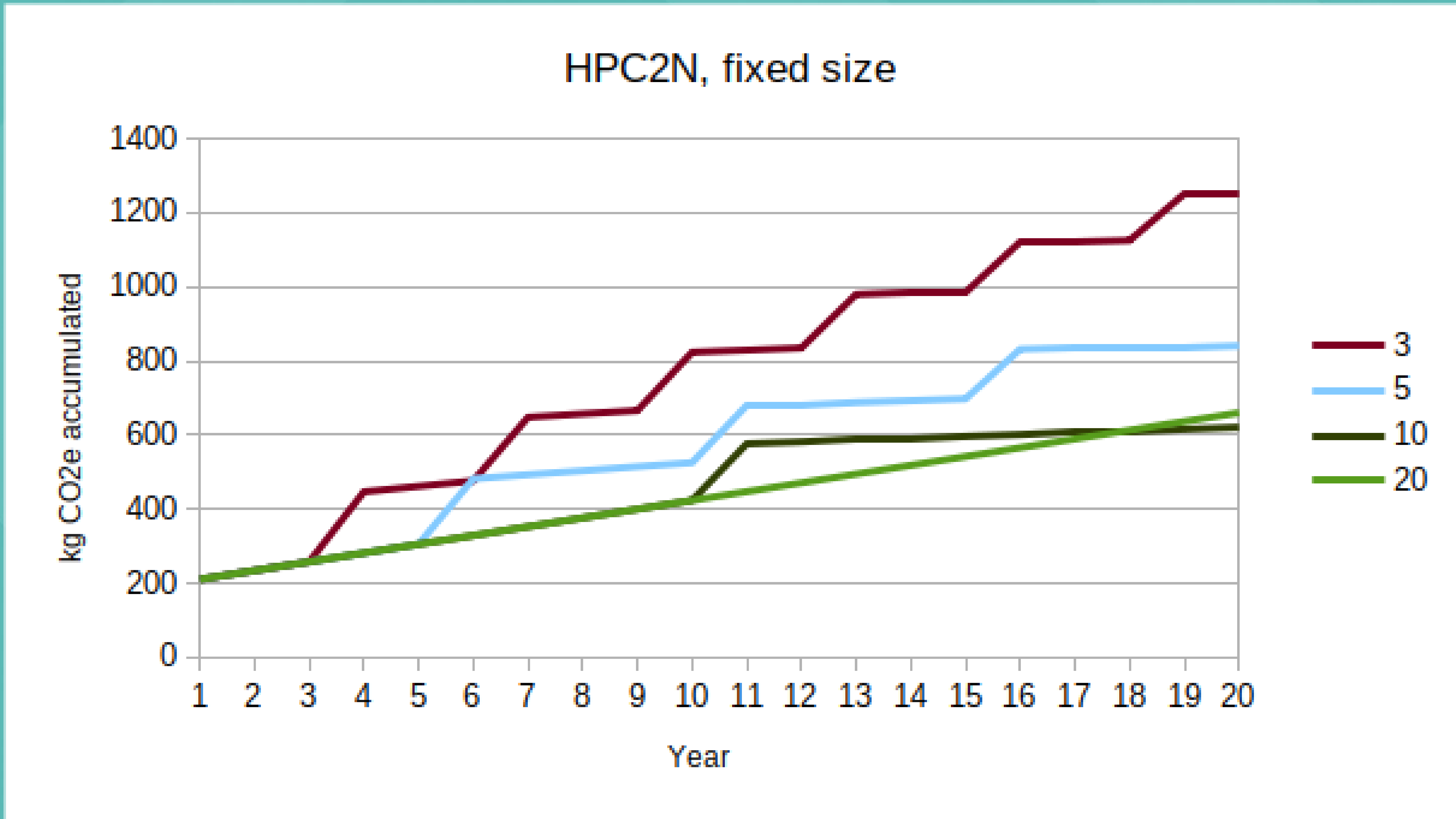
# Scenario

- A new scientific computing site contributing to LHC computing, installing 1 kHS23 on day 1
  - This is to get reasonable kg numbers in the graphs, the numbers are identical for tons per MHS23
- Then increasing by 15% each year
  - Alternative scenario with no yearly increase also provided
- Replacing old hardware after 3, 5, 10, or >20 years
- How large are the total emissions over 20 years?

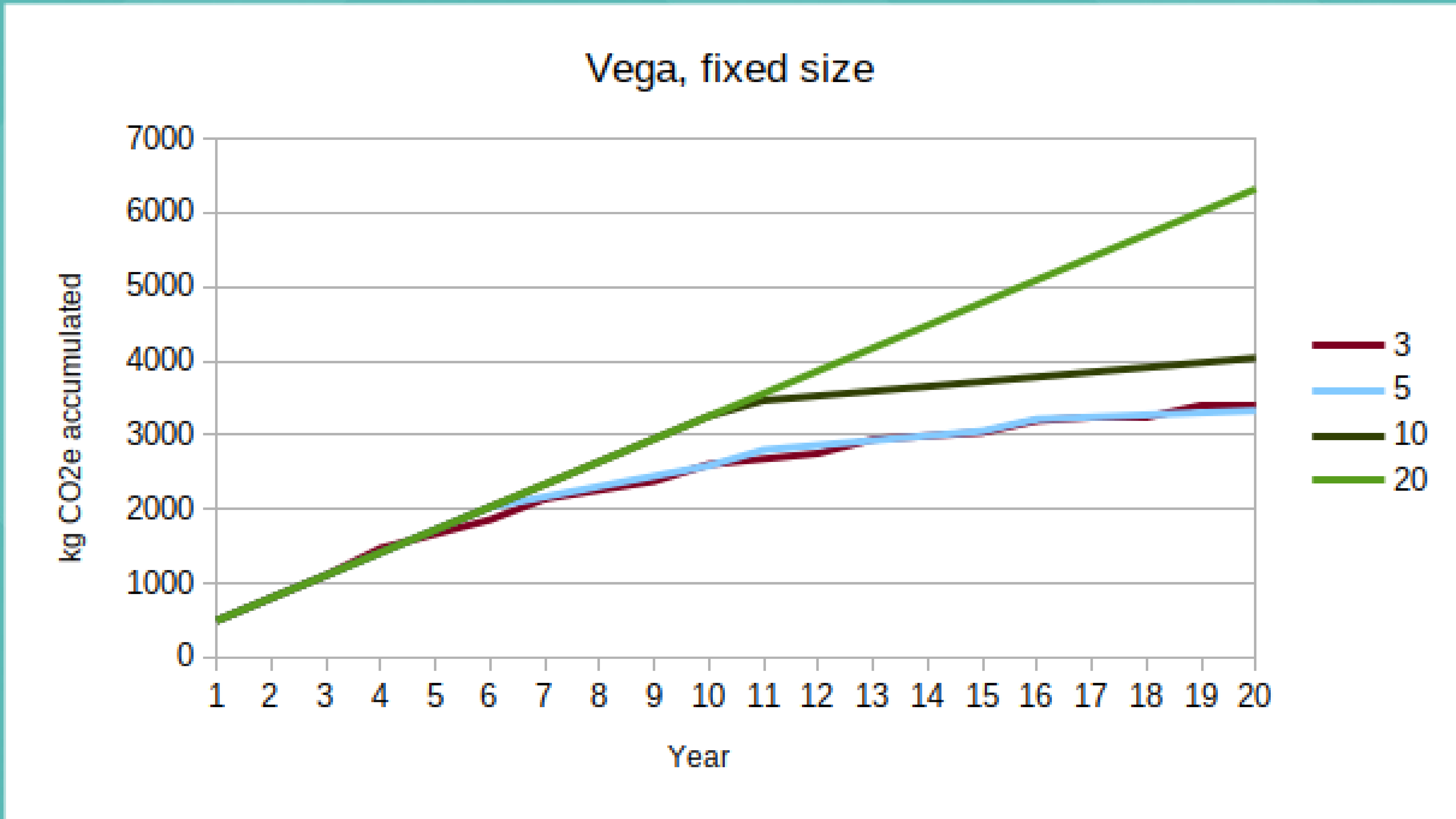
# Calculations

- Way too many graphs, but I leave them in the presentation for reference
- I’ll talk about some interesting features in some of them
- The Y-scales are different!
  - Putting HPC2N and ASGC in the same plot will have all HPC2N lines flat at approximately 0
- Accumulated lifetime emissions
- In the 15% growth scenario the share of embodied carbon ranges from 6% to 92%
  - Depending on how green the electricity production is

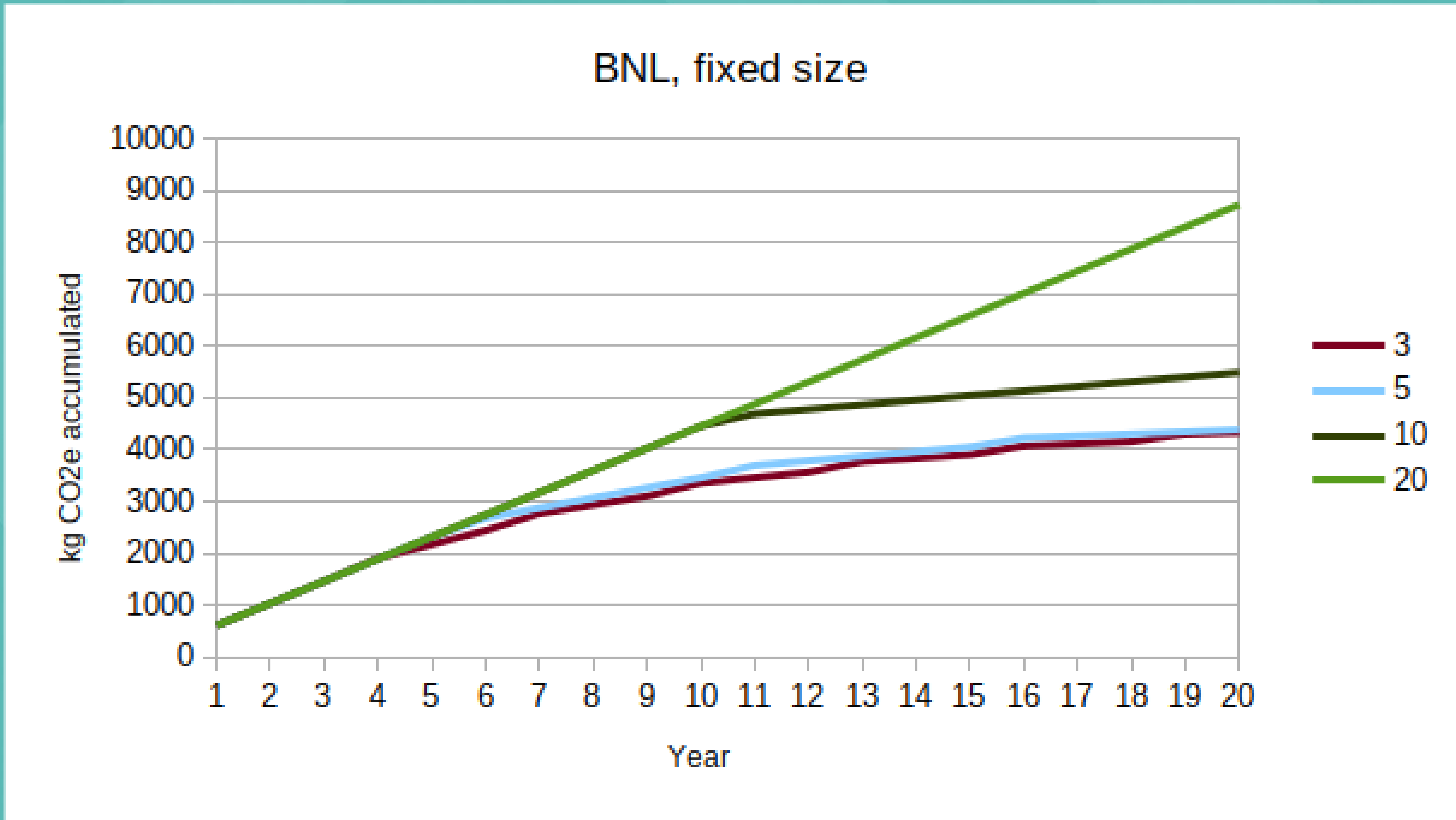
# HPC2N, fixed capacity



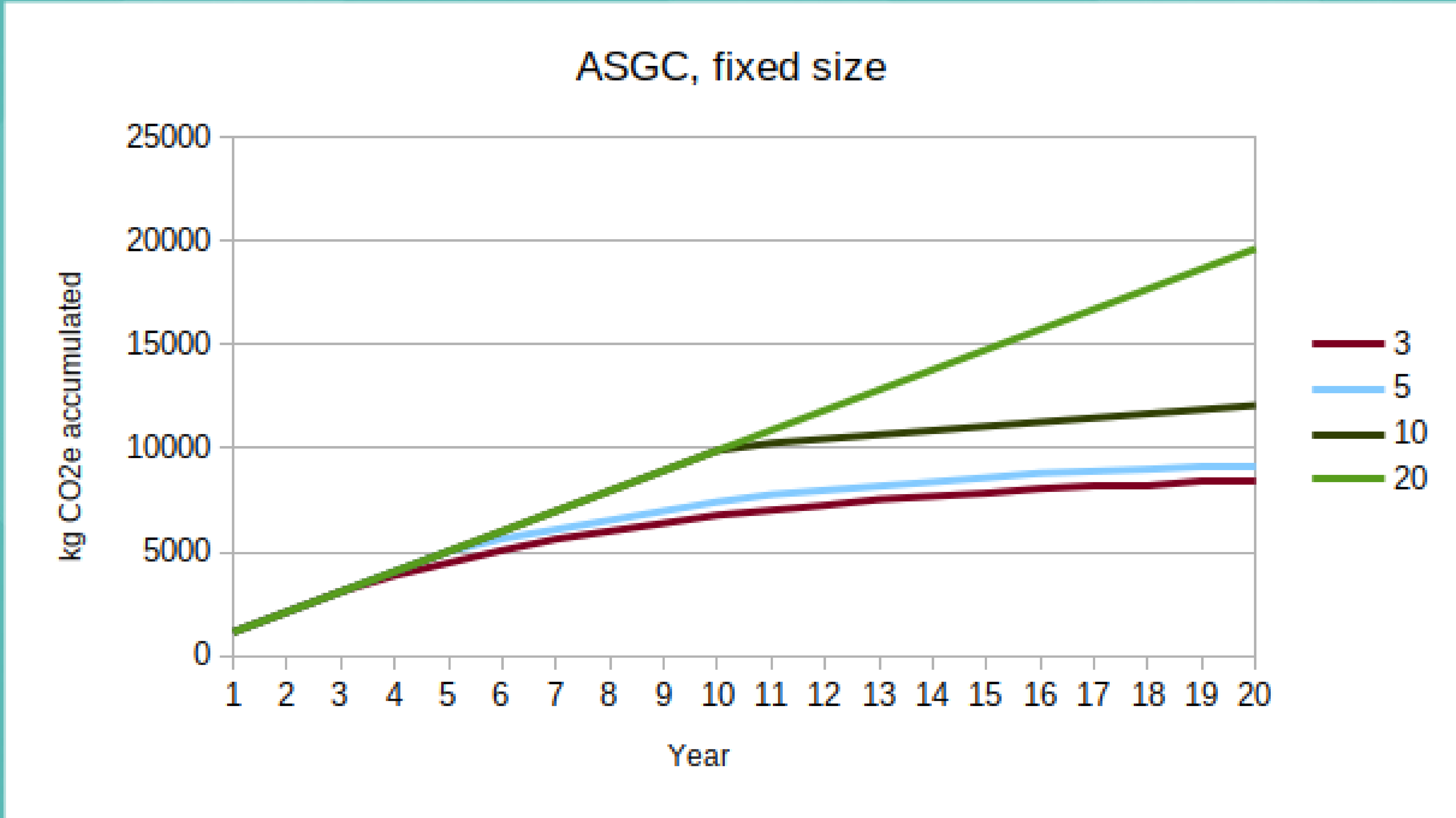
# Vega, fixed capacity



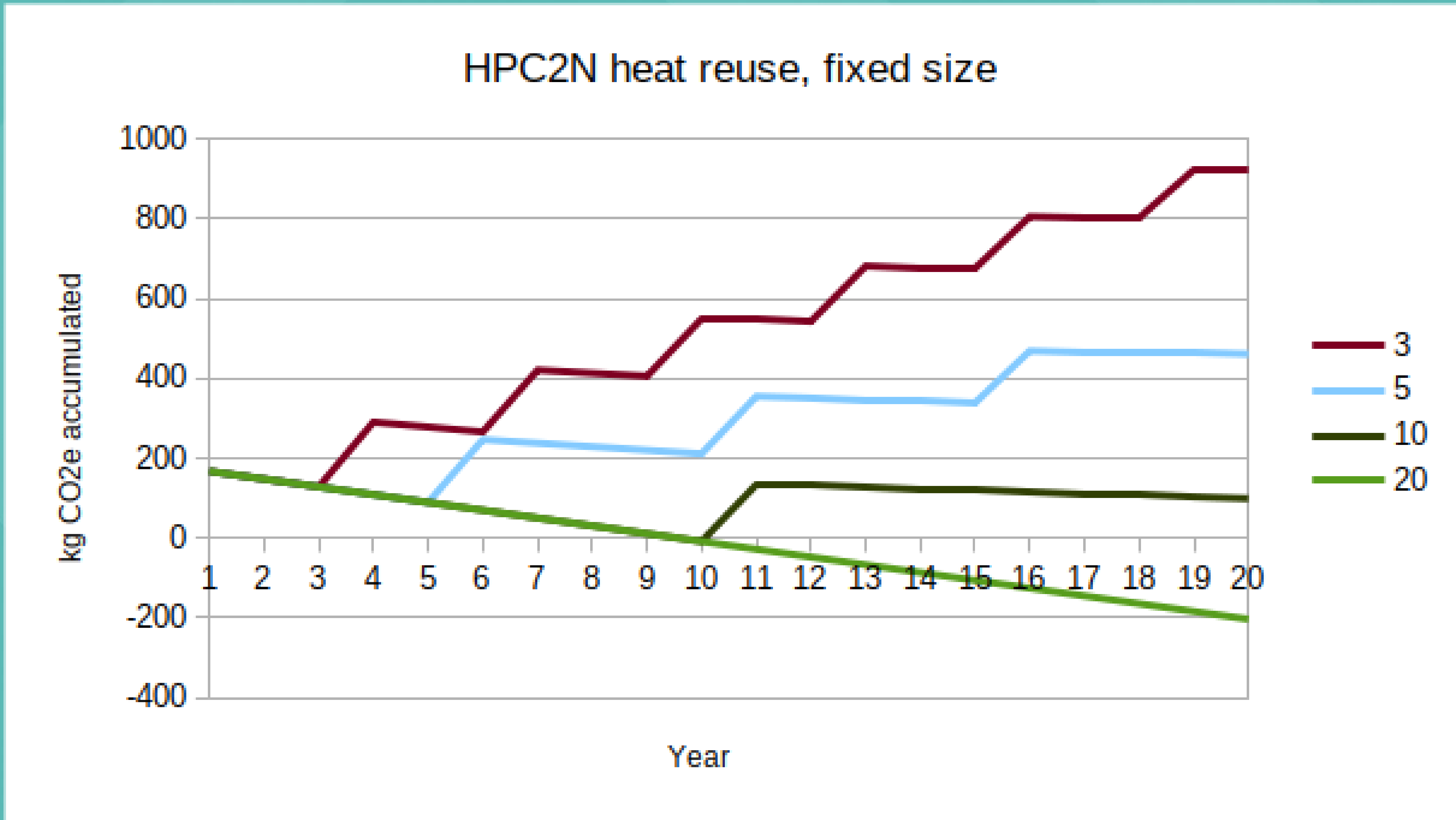
# BNL, fixed capacity



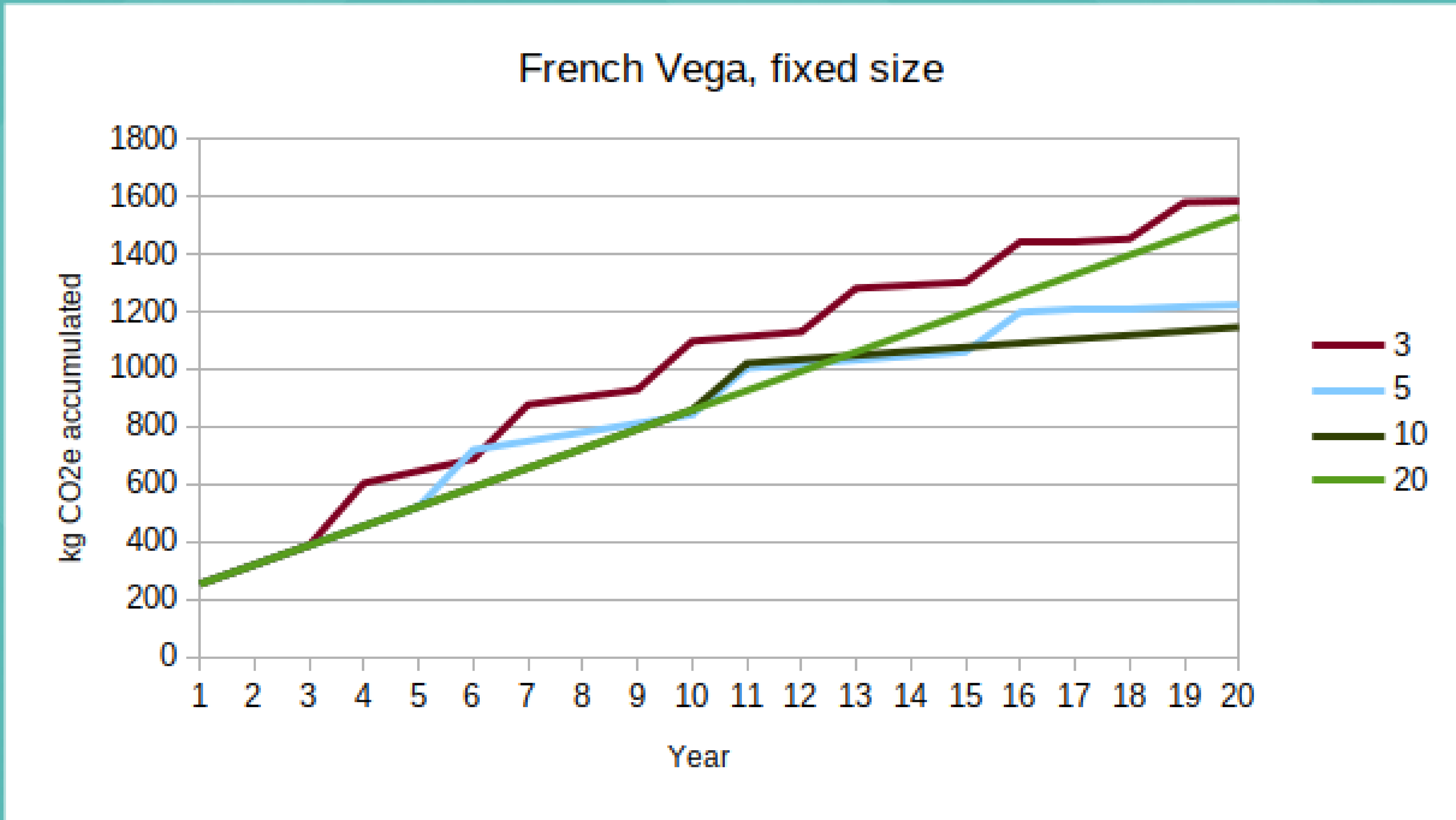
# ASGC, fixed capacity



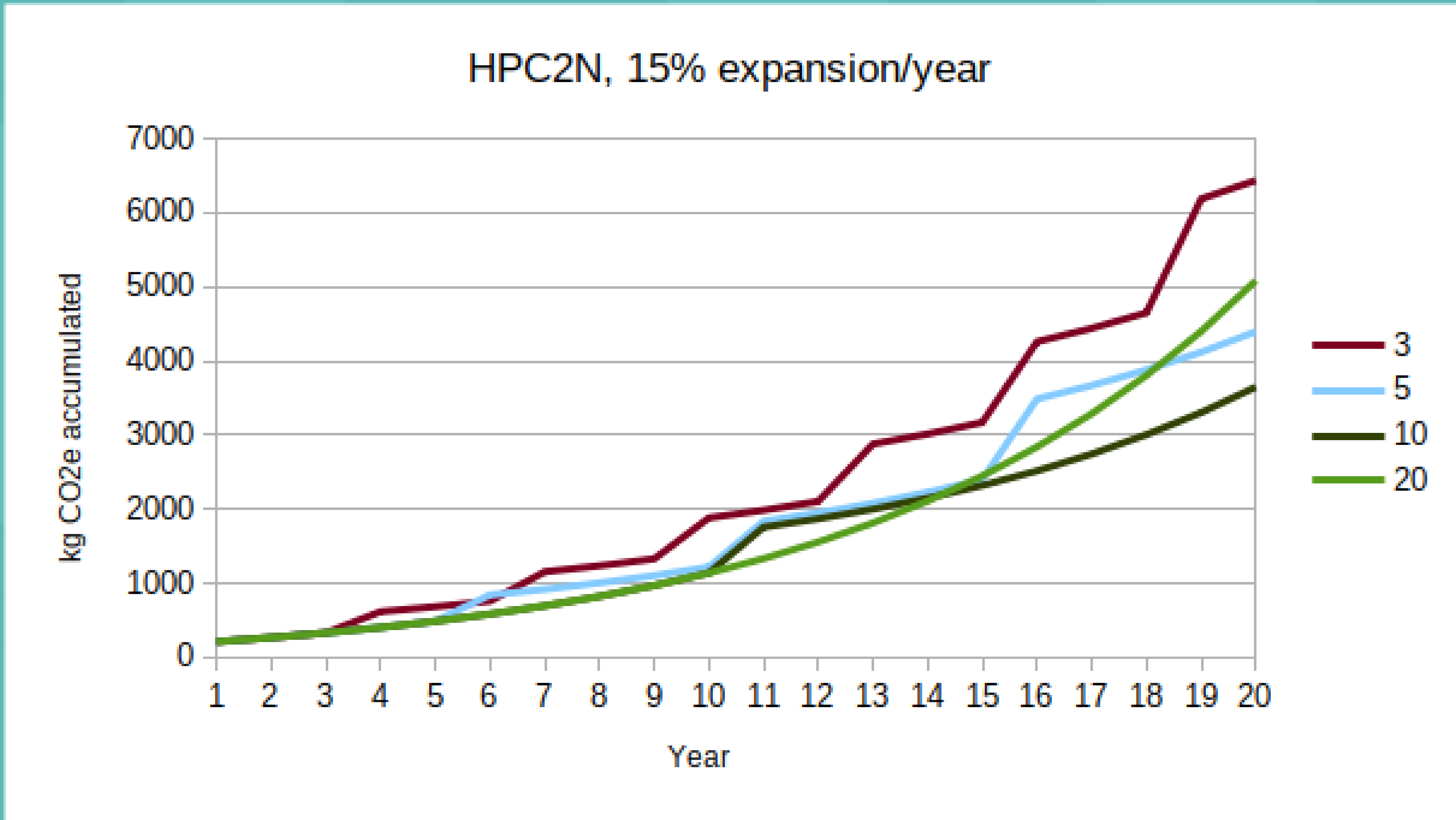
# “HPC2N heat reuse”, fixed



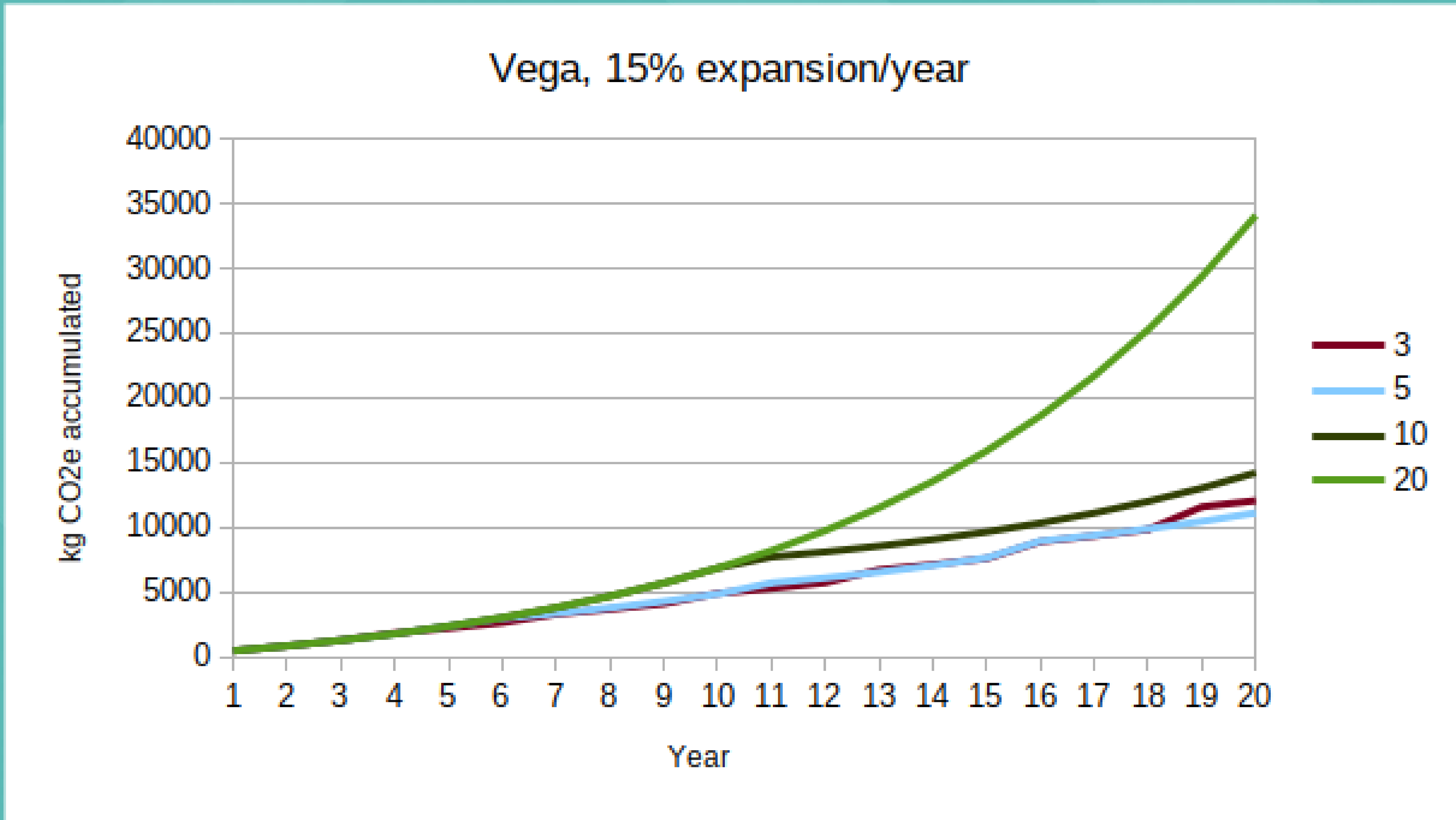
# “French Vega”, fixed



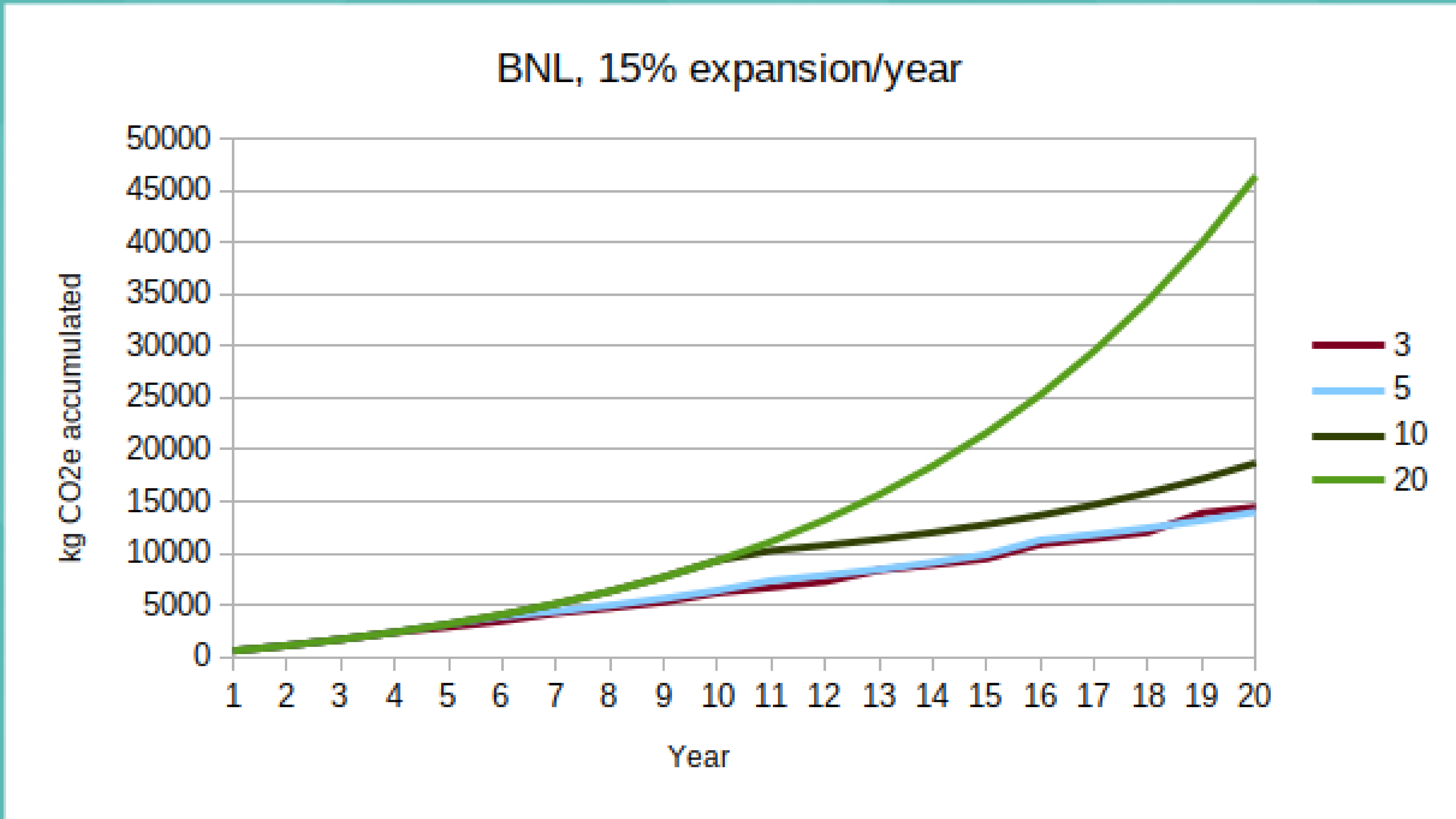
# HPC2N, 15% yearly growth



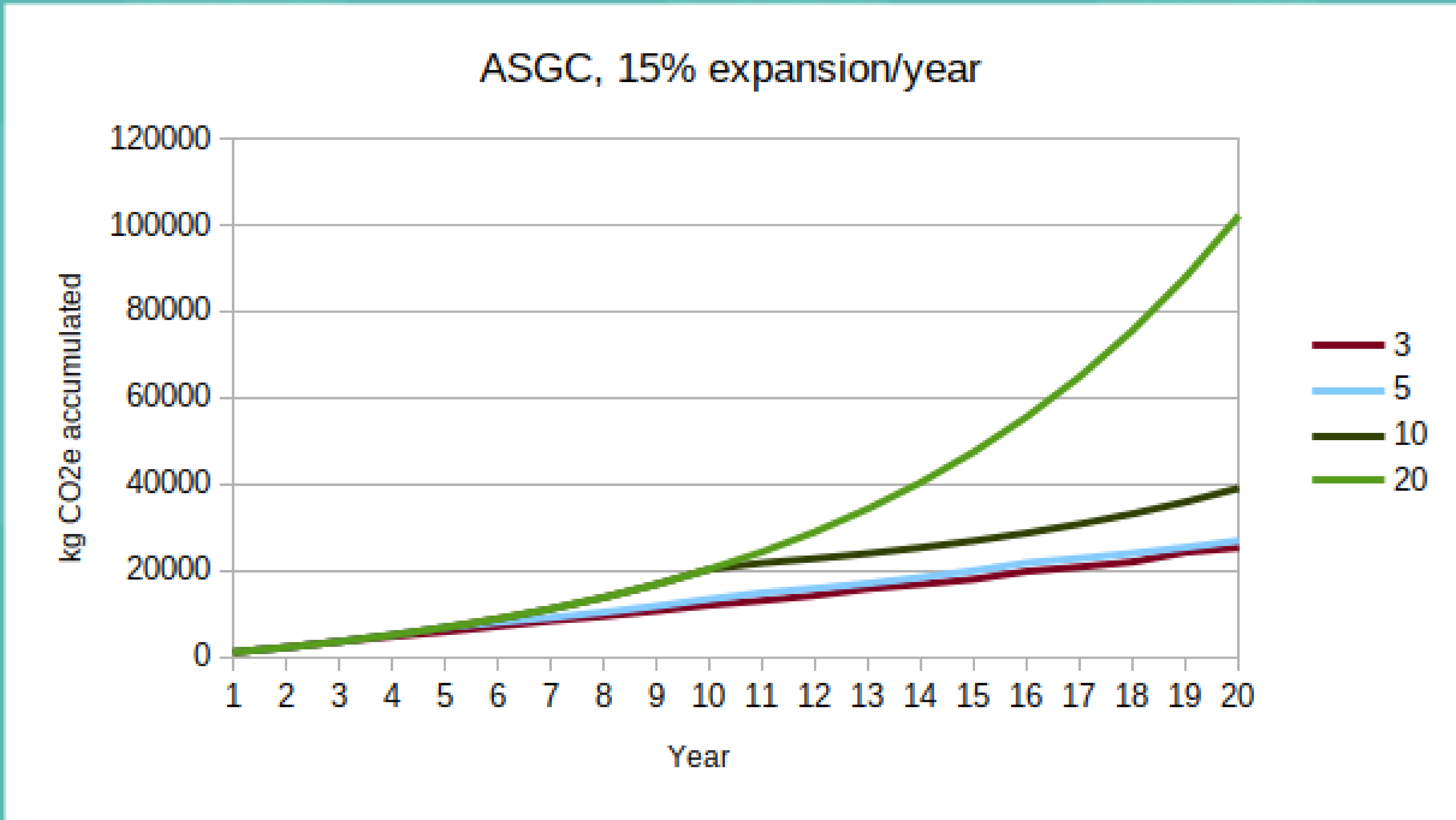
# Vega, 15% yearly growth



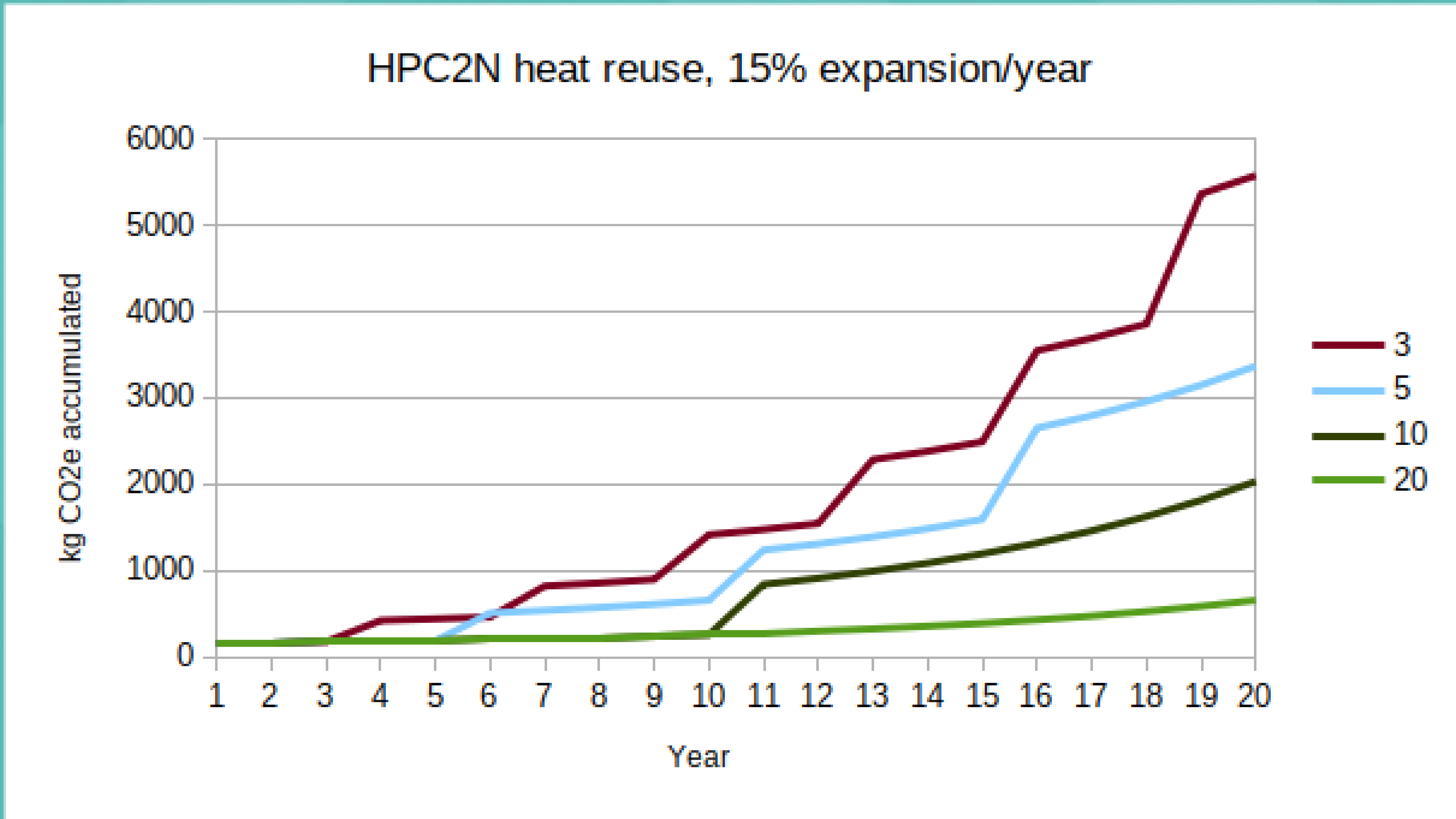
# BNL, 15% yearly growth



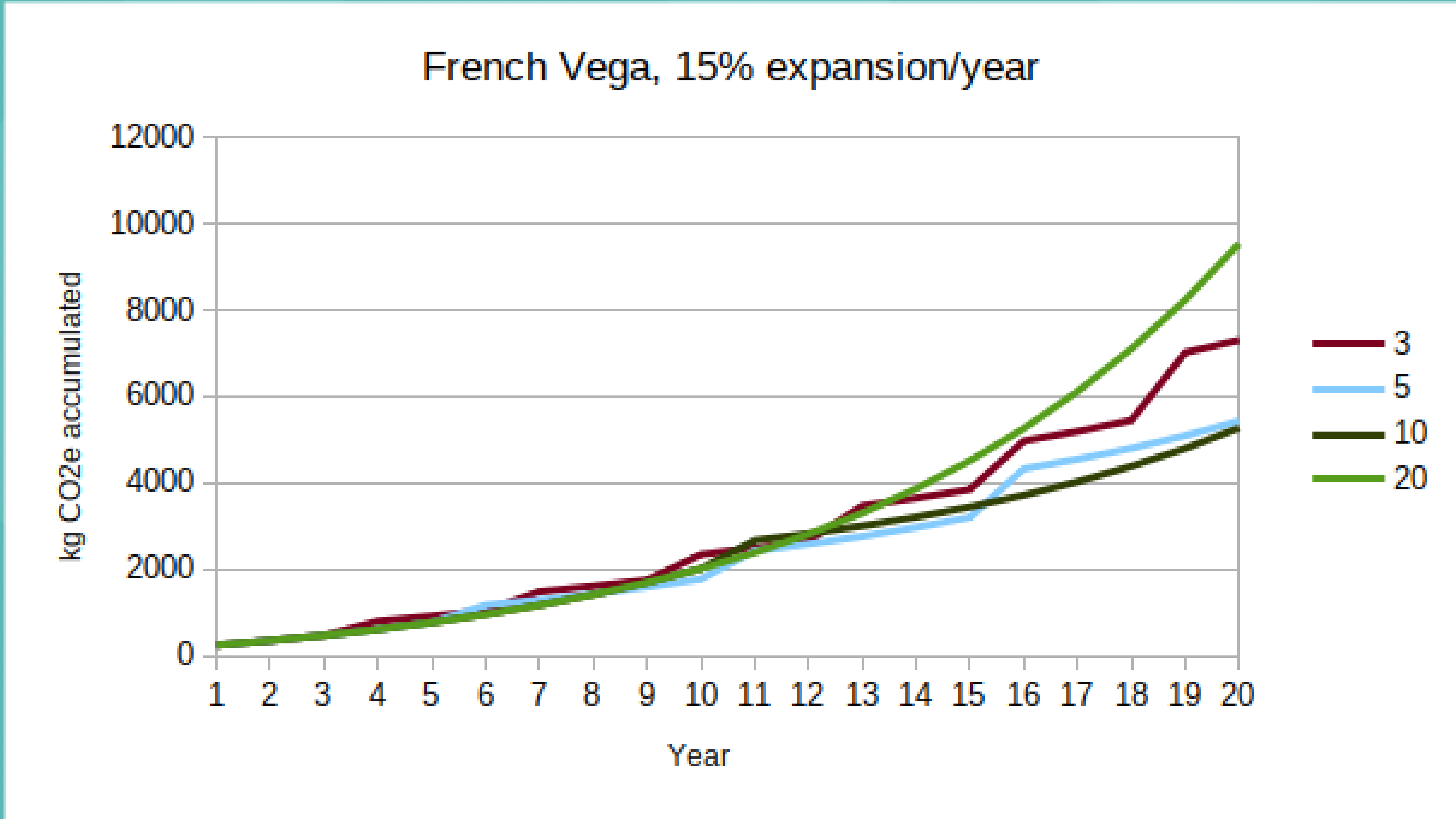
# ASGC, 15% yearly growth



# “HPC2N heat reuse”, 15% growth



# “French Vega”, 15% growth



# Conclusions

- In high emission locations: replace old servers by new as soon as financially viable
  - For heavily loaded scientific computing nodes
  - For general purpose servers or desktops this is less obvious
- In low emission locations: Running old servers for a long time might be better
  - The tradeoffs are emissions vs operating costs (power, staff, parts)
  - At some point you might have to consider emissions for bigger computer rooms too
- Choose low emission locations, if you have a choice

# Conclusions

- Reducing embedded carbon in servers?
  - Don’t buy more SSD or RAM than needed for the workloads
    - 4-8TB SSD is roughly half node manufacturing emissions
- Heat reuse can be a big impact
  - Cold regions with low emission power can even reach negative emissions, depending on what the alternative heat is
  - Comes at a significant financial cost, both investment and running
- Increasing computing needs will increase emissions

# Questions?

NORDIC E-INFRASTRUCTURE COLLABORATION