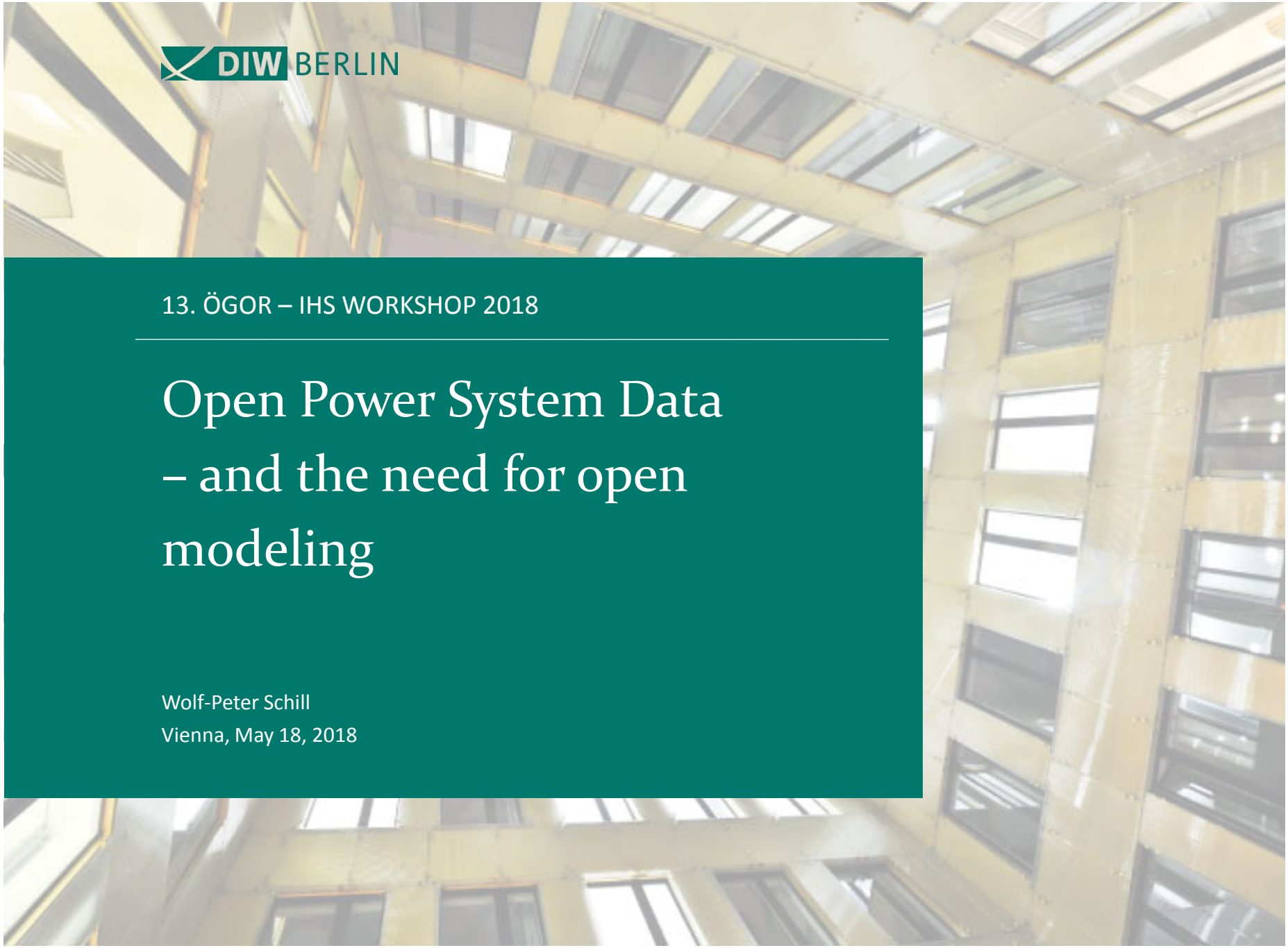


DIW BERLIN

13. ÖGOR – IHS WORKSHOP 2018

Open Power System Data – and the need for open modeling

Wolf-Peter Schill

Vienna, May 18, 2018

Overview

1. The Open Power System Data platform
2. More is needed: open modeling / open science
3. A recent example: Sinn (2017) vs. Zerrahn, Schill, Kemfert (2018)
4. Conclusions

1 OPSD in a nutshell

- We started 2015 with a project funded by BMWi
  - Initial team: neon, Uni Flensburg, TU Berlin, DIW Berlin
- Background
  - Energy research and policy advice based on numerical models
  - These are data-intensive: generation capacities, time series...
  - Data often available, but hard to find, poorly documented, tedious to process, partly subject to restrictive terms of use
- The idea
  - An online platform: https://open-power-system-data.org/
  - Aggregating publicly available European data
  - Free of charge

→ Avoid redundant work and improve quality of data

1 IT solution: lightweight, decentralized, open

- Lightweight, decentralized and scalable
- Built on open standards and open software:
  - Open Knowledge's *Data Package* standard (CSV, JSON)
  - Jupyter Notebooks (Python)
    → code and documentation in one file
  - GitHub → version control
- File formats for download:
  - CSV: de facto standard
  - Excel: preferred by some users
  - SQLite: queryable, for experts

1 Data aggregation and validation – no data creation

- We collect existing, publicly available data
- We draw on official sources
  - Statistical offices, Eurostat
- …or on semi-official sources
  - TSOs, regulators, public agencies, sector associations
- Data only carefully modified / supplemented

1 Some careful modifications

- Correction of obvious errors
  - Missing commas
  - Alphabetical characters in numerical data fields
  - GW-scale rooftop solar plants
  - We also communicate these errors back to the sources
- Translation into English
- Technology category alignment
- We add geolocations of generators
- Time series issues
  - Provision of UTC & EC(S)T timestamps
  - Solving inconsistencies with daylight savings time, leap years
  - Interpolation of short gaps (max 2 hours)

1 Data is organized in five data packages

Data package	Description
Conventional power plants	List of conventional power plants in Germany and European countries
National generation capacity	Aggregated generation capacity by technology and country
Renewable power plants	List of renewable energy power stations
Time series	Load, wind and solar, prices in hourly resolution
Weather data	Script for the download of MERRA-2 weather data

→ We started with Germany and then expanded the platform to other countries

1 Data package „Conventional power plants“(as of end 2017)

- Conventional power plants in Germany and Europe (actual plants, no aggregates)
  - Nuclear, coal, natural gas, oil
  - Run-of-river, reservoirs, pumped storage
- Power plant data based on publicly available sources
  - Germany: BNetzA and UBA
  - Neighbors: Mostly TSOs, ministries, and associations
- Data covered
  - Installed capacity
  - Main energy source, type of technology
  - Where available: CHP capability and commissioning year
  - Own supplements: geo-coordinates, efficiency, EIC code

1 Data package „National generation capacity“ (as of end 2017)

- National generation capacity for European countries
  - Aggregated statistics of installed capacity
- Harmonized classification of generation capacity
  - Type: nuclear, fossil fuel, renewable energy source, other
  - Energy source: hard coal, natural gas, hydro, wind, etc.
  - Technology: gas turbine, combined cycle, wind on/offshore etc.
- Temporal coverage as provided by individual data source
  - Annual data covering 1990 – 2017, majority of data 2013 - 2015
- Data statistics
  - 1504 country statistics from 25 national and 4 international sources
  - ≈4 statistics for each country differing by year / data source

1 Data Package “Renewable power plants” (as of end 2017)

- Data
  - Solar, wind, geothermal, bioenergy, hydro (> 1.8 mio units)
  - Renewable capacity time series
- Resolution
  - Units (DE, DK, CH)
  - Aggregated unit groups (PL, FR)
- Data covered
  - Energy source & technology
  - Electrical capacity [MW]
  - Georeference [latitude, longitude]
  - Data source
  - Partly: TSO, DSO, address, hub height, rotor diameter, EEG-key…

1 Data Package "Time series" (as of end 2017)

- Data covered:
  - Load
  - Day-ahead spot prices
  - Wind & solar generation
- Temporal resolution:
  - 60 min (all data) and 30 min / 15 min (some data)
- Geographic resolution:
  - Countries / bidding zones / control areas
- Temporal coverage
  - Some data back to 2005, but largely from 2015 onwards
- Total number of observations:
  - ~ 9.5 mio (60 min), ~ 0.6 mio (30 min), ~ 10.5 mio (15 min)

1 OPSD continues to improve

- Platform successfully established during first project phase 2015-2017
  - Many site visits, data referenced in a range of papers
  - Platform was awarded with two prices in 2017

SH Schleswig-Holstein
Der echte Norden

Deutschland
Land der Ideen
Ausgezeichneter Ort 2017

- Ongoing updates and extension

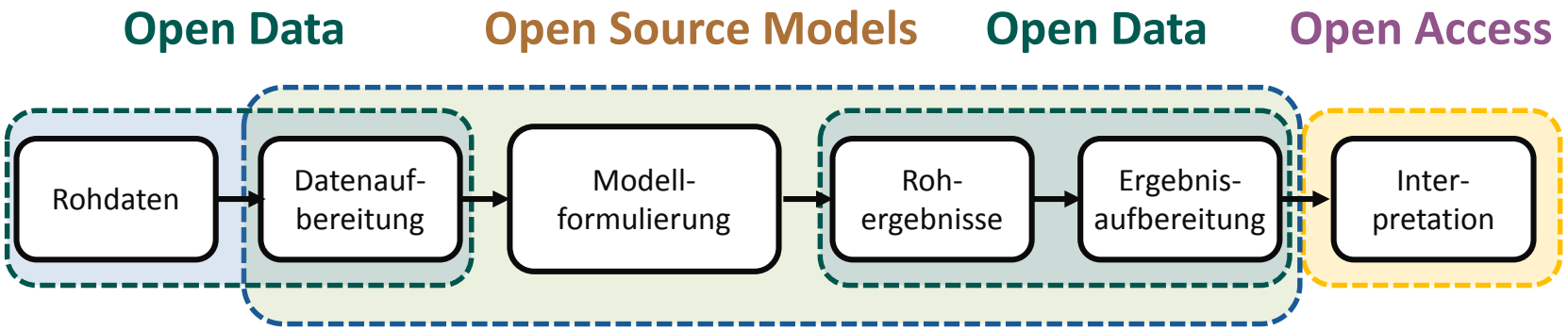
Open modeling / open science

- Open (input) data is good – but not sufficient
- More is needed: open modeling / open science
  - Make model-based research accessible, understandable, replicable and reviewable
  - Necessary for academic research
  - And probabaly even more so for applied / policy-relevant work

- Open modeling community recently gained momentum
  - Check out http://www.openmod-initiative.org

Open modeling / open science

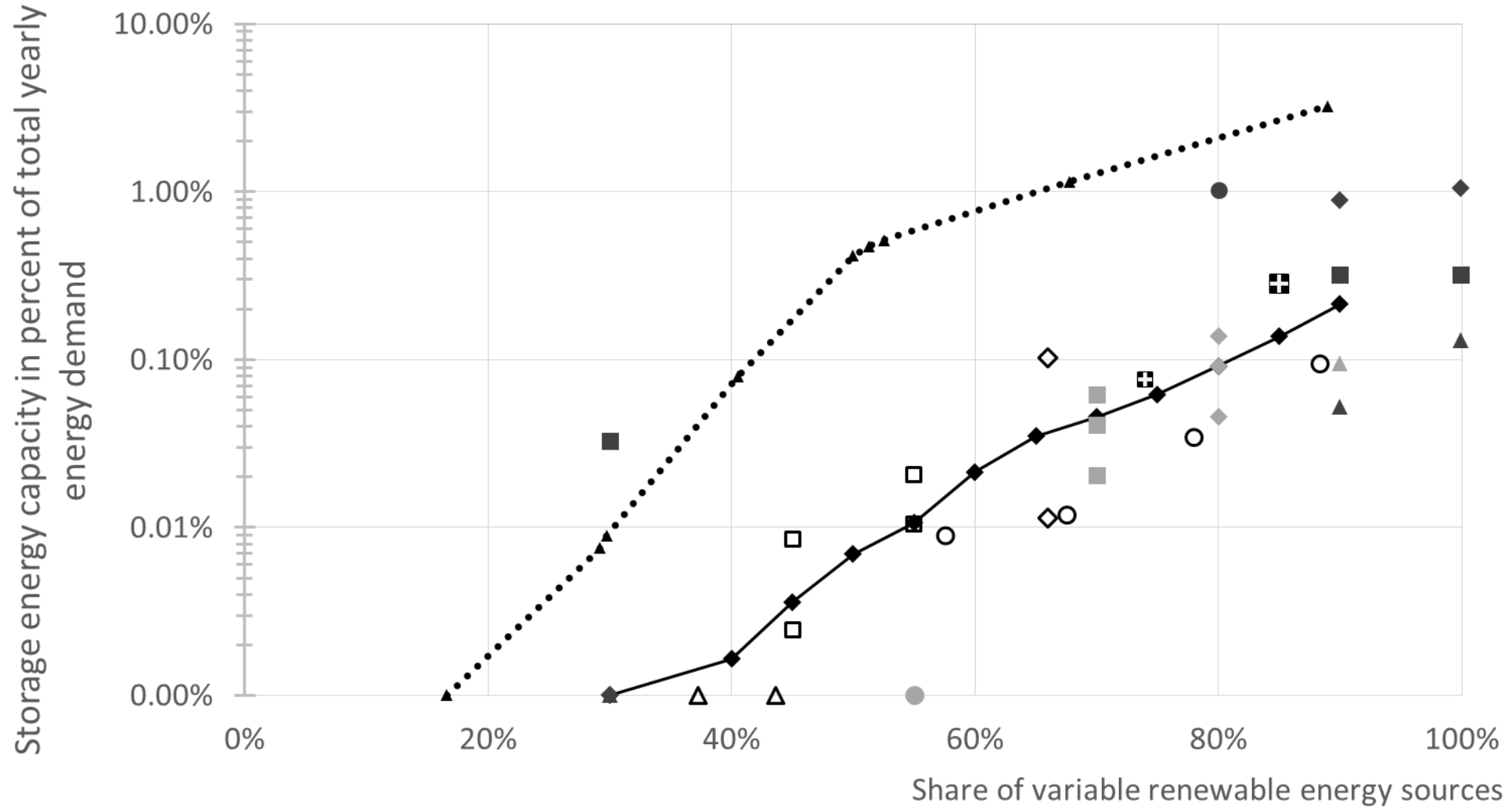
- What does it mean for energy modelers?
  - Use open data, wherever possible
  - Make input data and code available → GitHub, Zenodo etc.
  - Cp. also FAIR data principles (DOI: 10.1038/sdata.2016.18 1)
  - Ideally, also provide all (major) results openly in suitable form
  - Open access version of papers (at least „green“)
    → personal homepages, arXiv etc.



A recent example: An article by Hans-Werner Sinn (*European Economic Review* 99, 2017)

- Sinn (2017): analysis of electrical storage requirements
  - Data-driven approach – but neither data nor code provided
  - Description of methodology not very clear
  - Result: wind and solar energy imply excessive storage needs in Germany: 2.1 (5.8, 16.3) TWh for 50% (68%, 89%) RES
  - Strong conclusion: storage major barrier for RES expansion
- We do an „open science“ replication
  - We use OPSD input data
  - We provide all input data and all tools / code on Zenodo
    - https://doi.org/10.5281/zenodo.1170554
  - Open-access preprint on arXiv (Zerrahn, Schill, Kemfert):
    - https://arxiv.org/abs/1802.07885

3 Sinn vs. literature



What Sinn does – and what we do

Sinn's general approach

- Hourly 2014 time series of demand and (onshore) wind and PV capacity factors
- Scale up combined RES until given annual renewables share in demand is reached

Storage heuristic

- Store hourly RES surplus and release as soon as residual demand is positive again
- Storage size determined as minimum GWh to integrate all variable renewables

Questionable implicit assumptions

- Full RES integration by electrical storage, no curtailment
- No economic objective function
- Germany only, no other flexibility options, no sector coupling, …

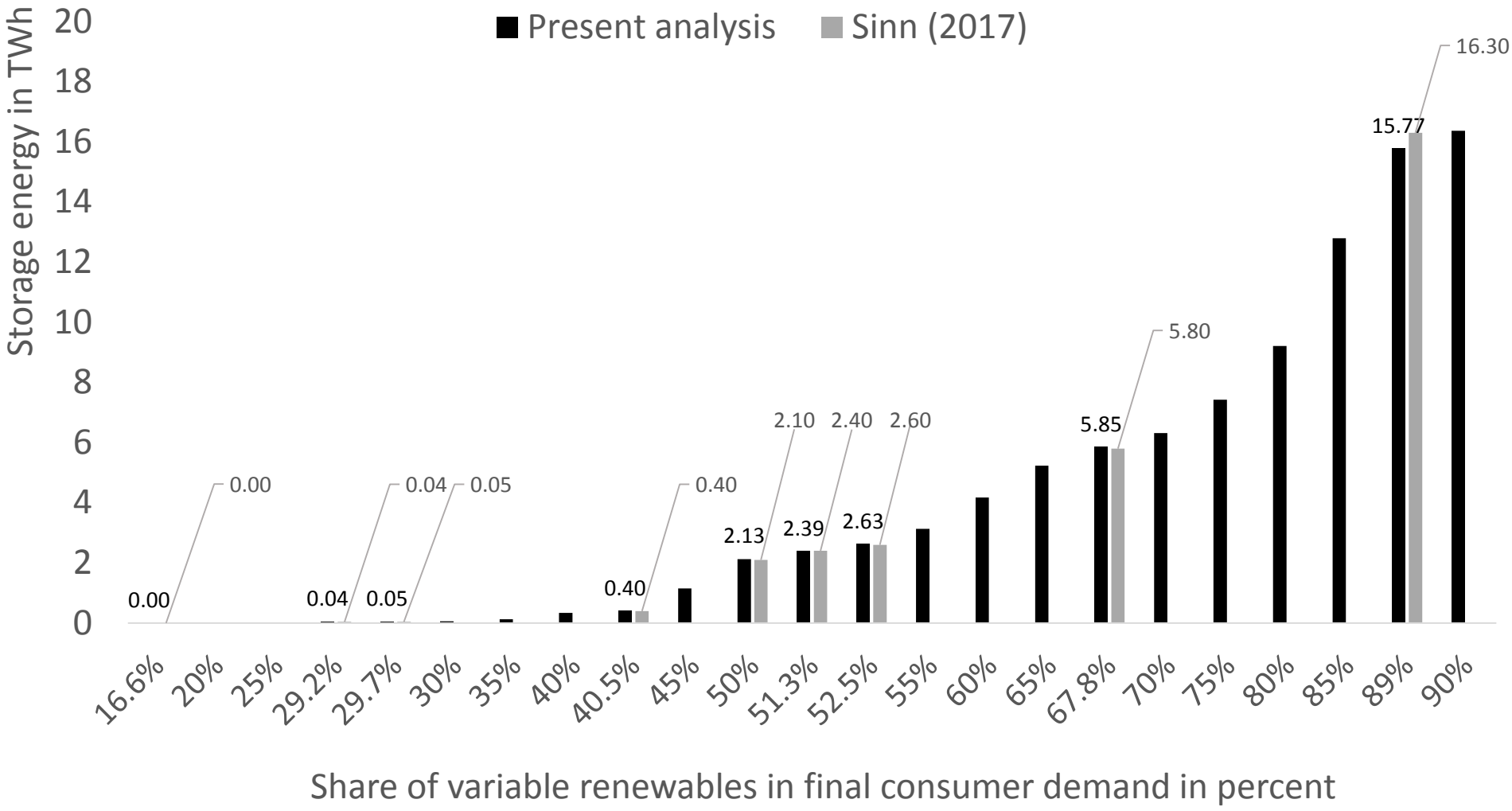
Our contribution

- Provide open-source replication of Sinn's approach
- Address questionable implicit assumptions

3 Replication

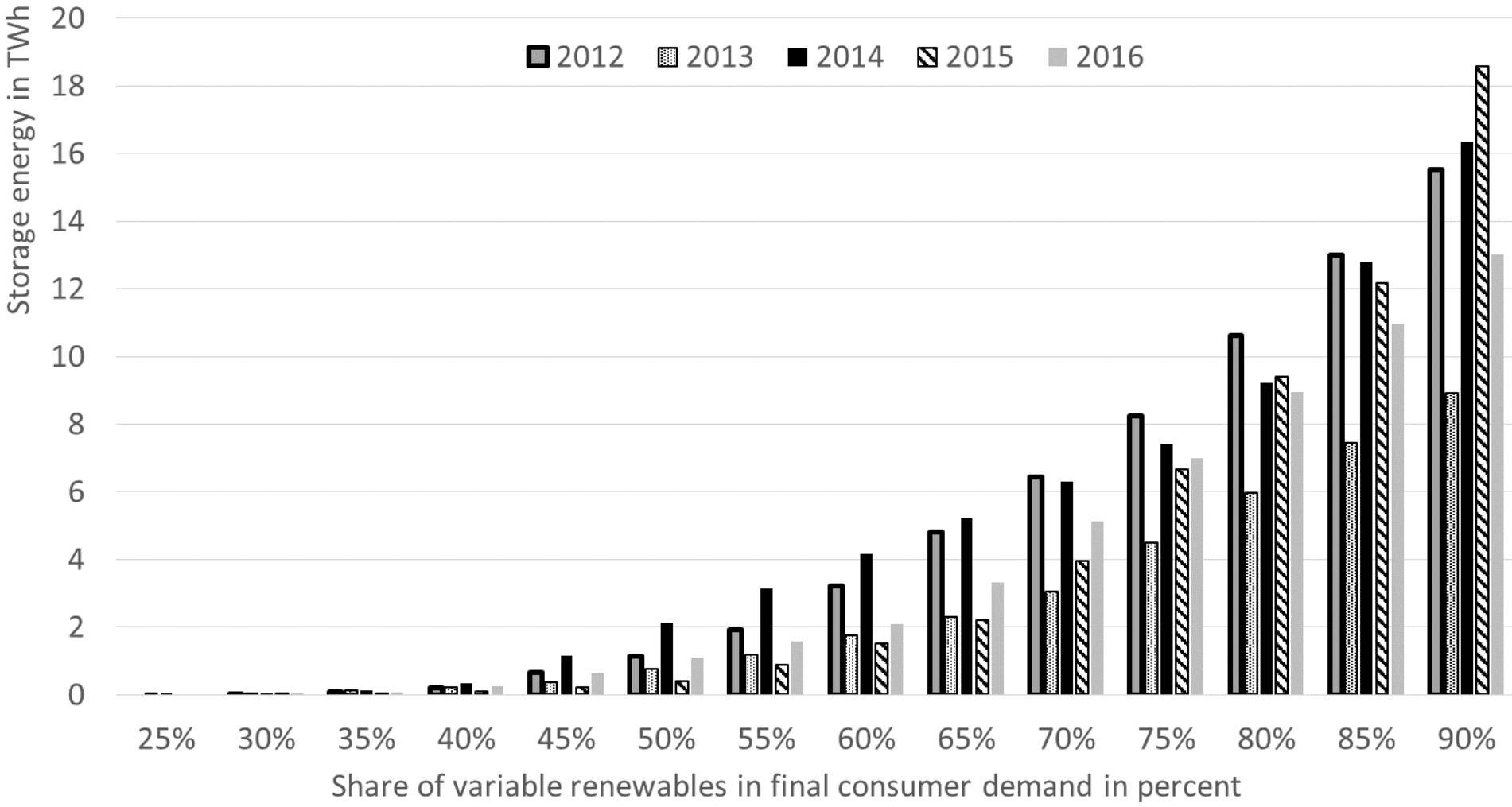
We can replicate it!

- Open input data, open Excel tool (https://doi.org/10.5281/zenodo.1170554)



3 (Non-)Robustness → easy with OPSD inputs

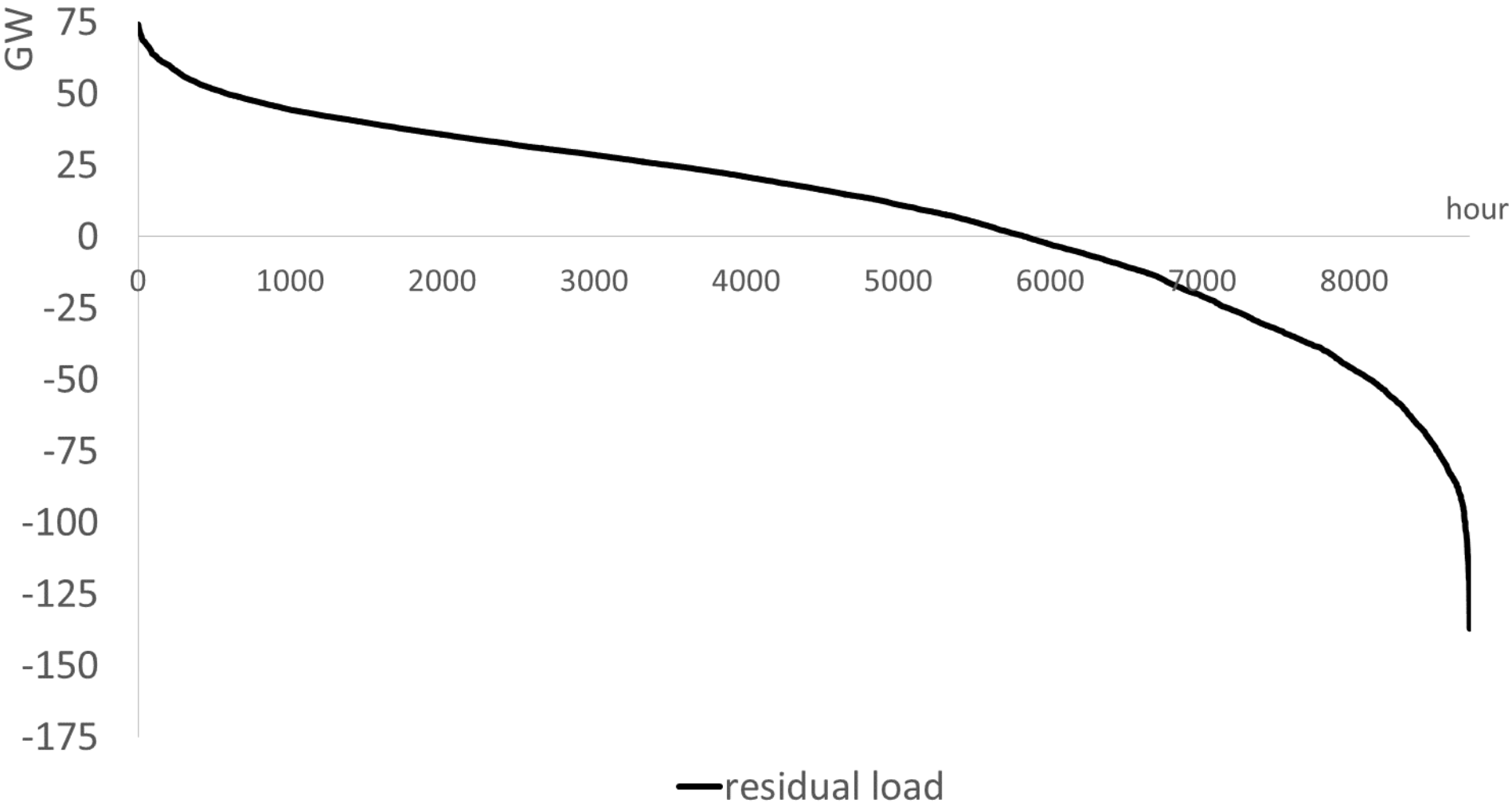
Replication: other base years than 2014 lead to different outcomes



3 Intuition: What drives Sinn's results?

Full integration of all vRES drives storage requirements

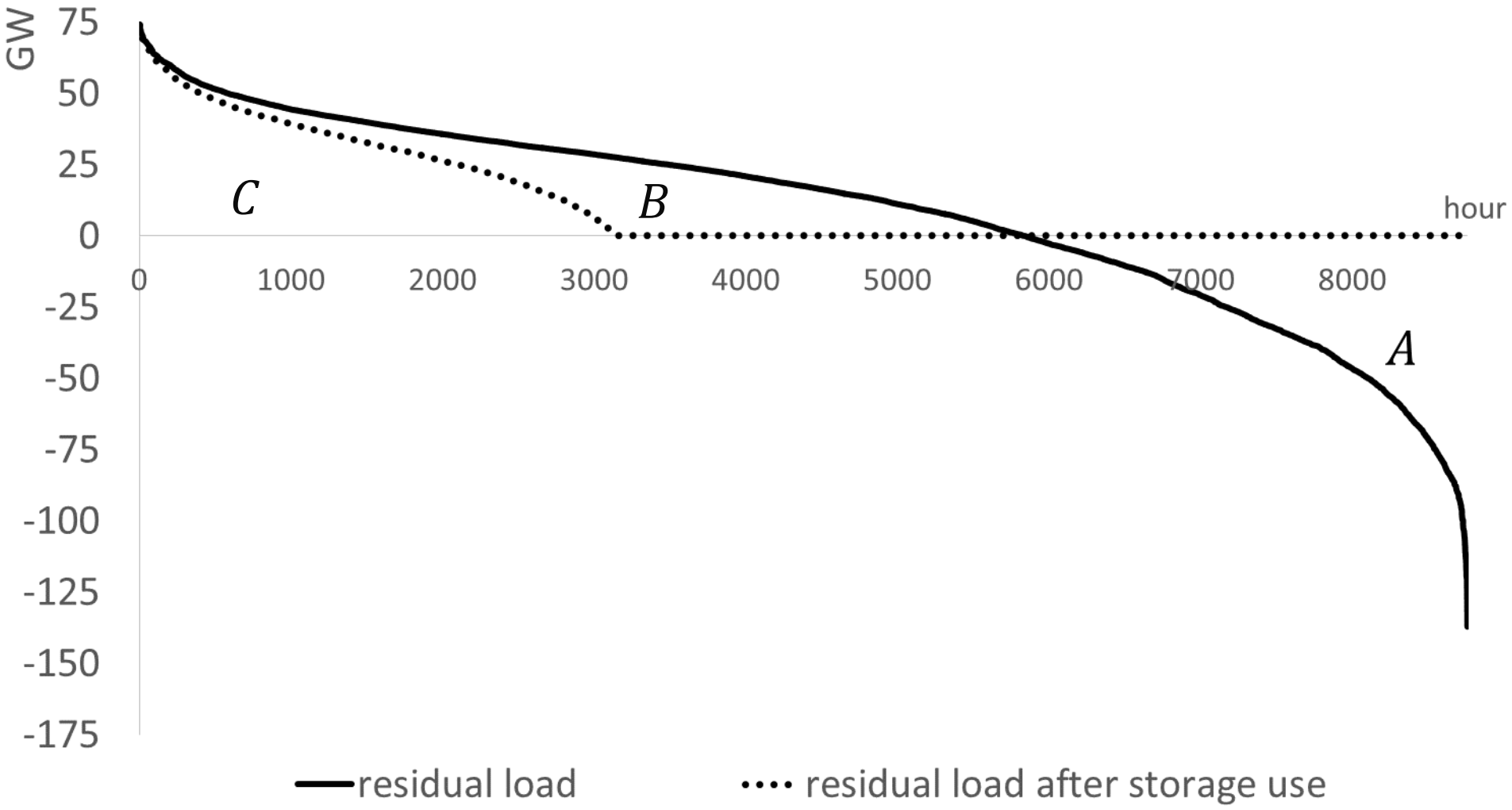
- Residual load duration curve (80% vRES)



3 Intuition: What drives Sinn's results?

Full integration of all vRES drives storage requirements

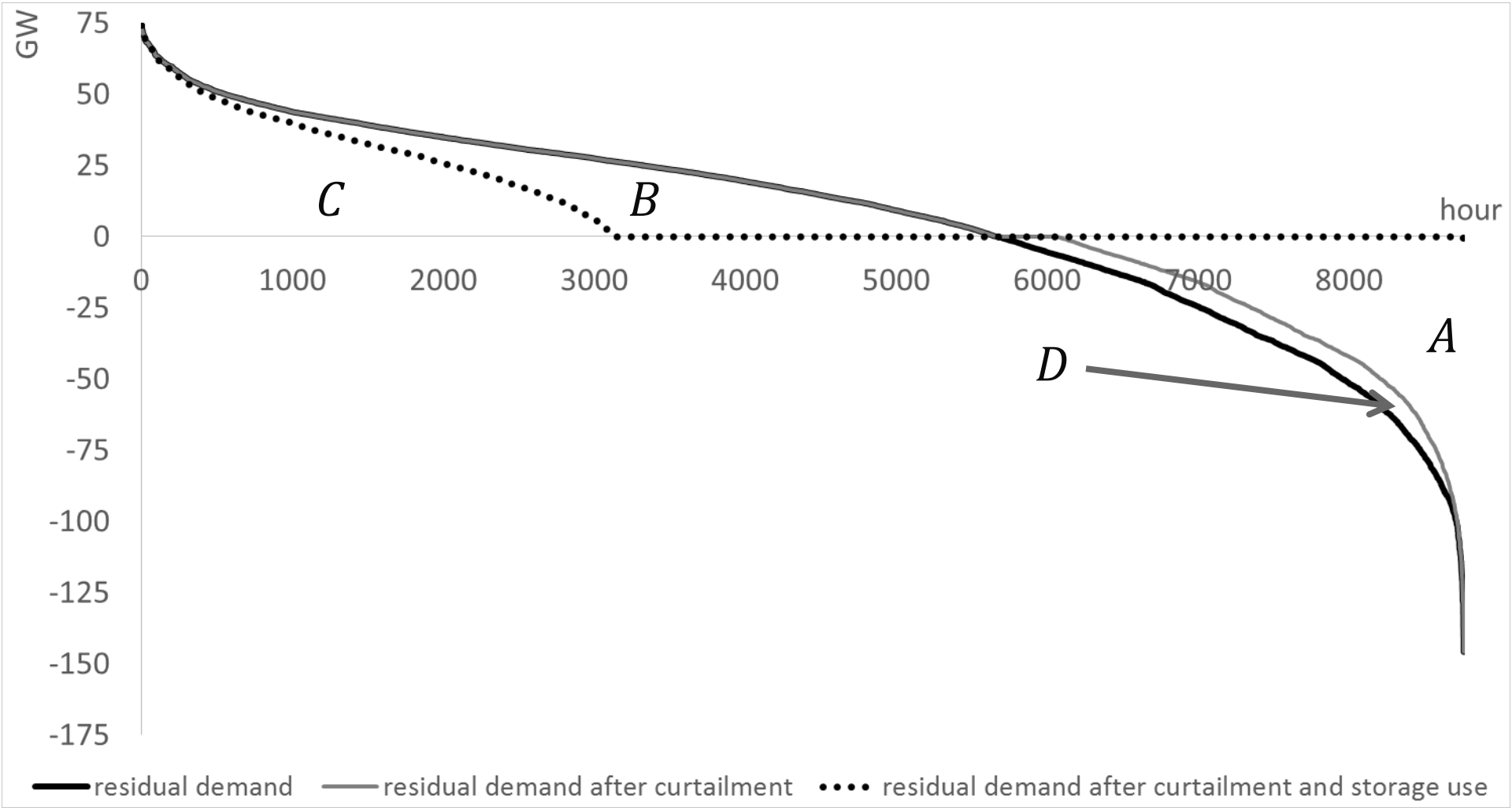
- Residual load duration curve: storage shifts surplus from A to B



3 Renewable curtailment (here: energy-oriented strategy)

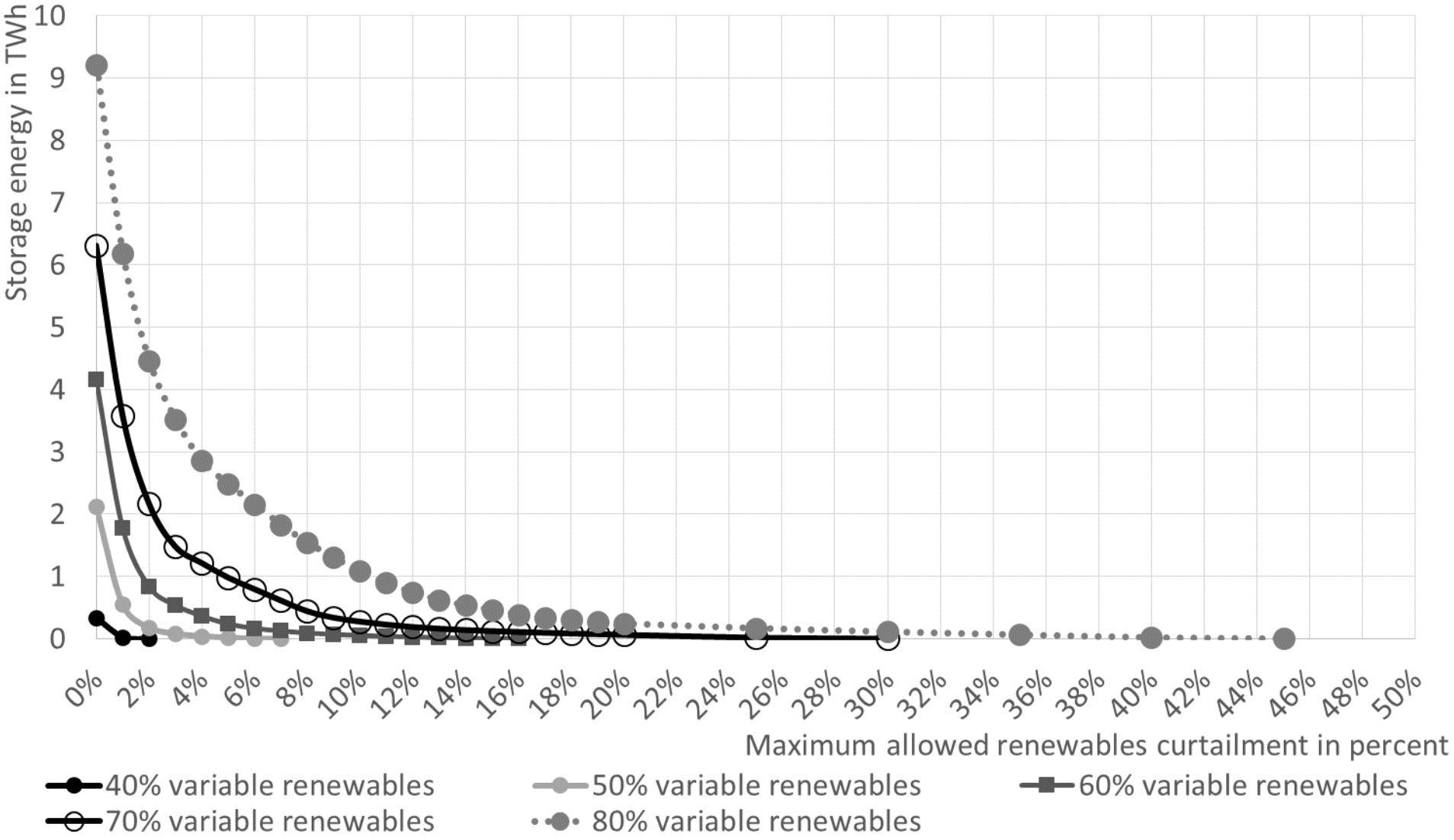
We introduce curtailment

- Renewable surplus curtailed (D) if storage full



3 Renewable curtailment (here: energy-oriented strategy)

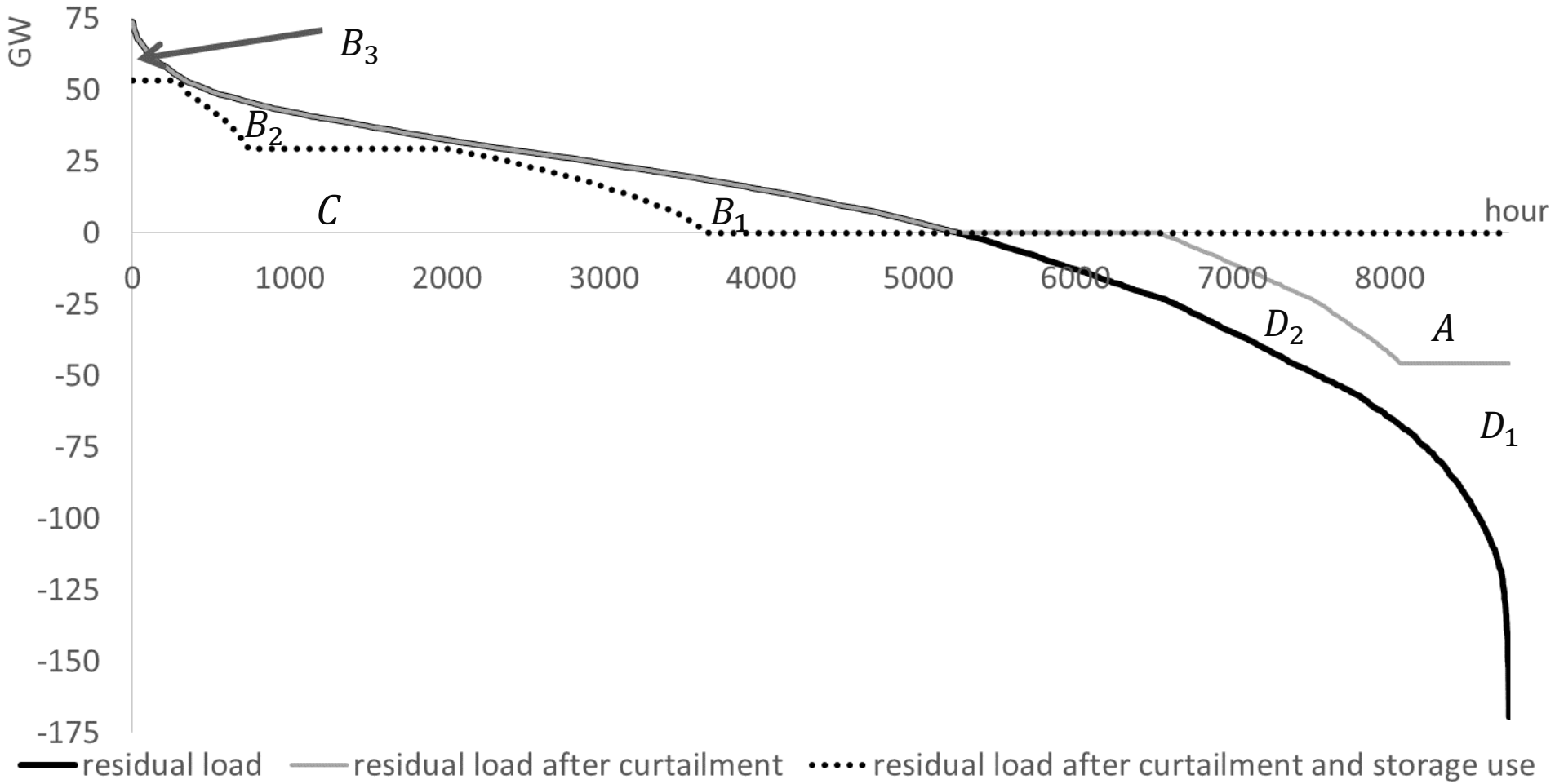
Result: strongly decreasing storage needs



3 Alternative approach: optimization with cost-minimizing objective

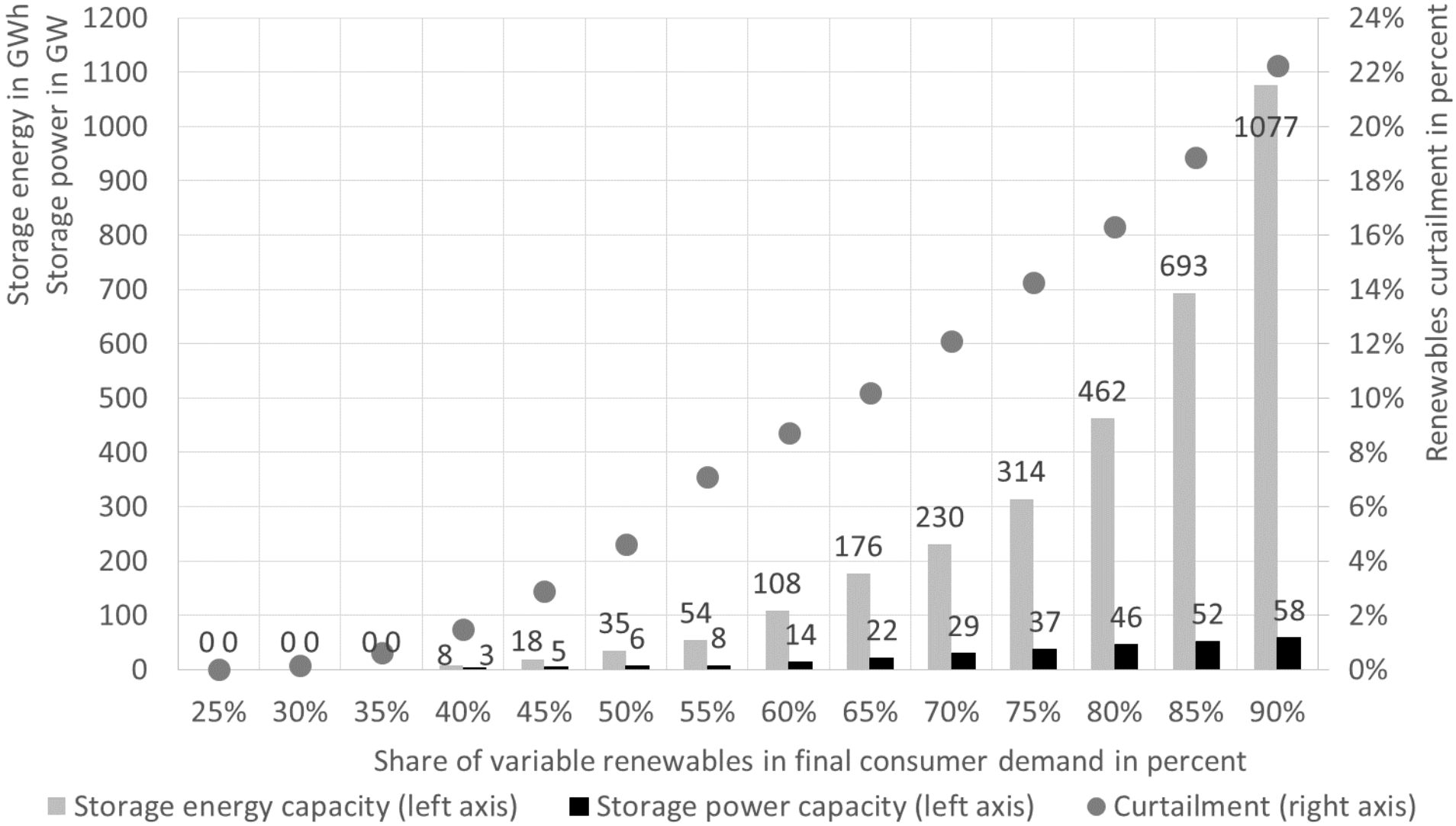
Residual load duration curves

- Storage power (D1) and storage energy (D2) restrictions
- Arbitrage value (B1-B3) and capacity value (B3)



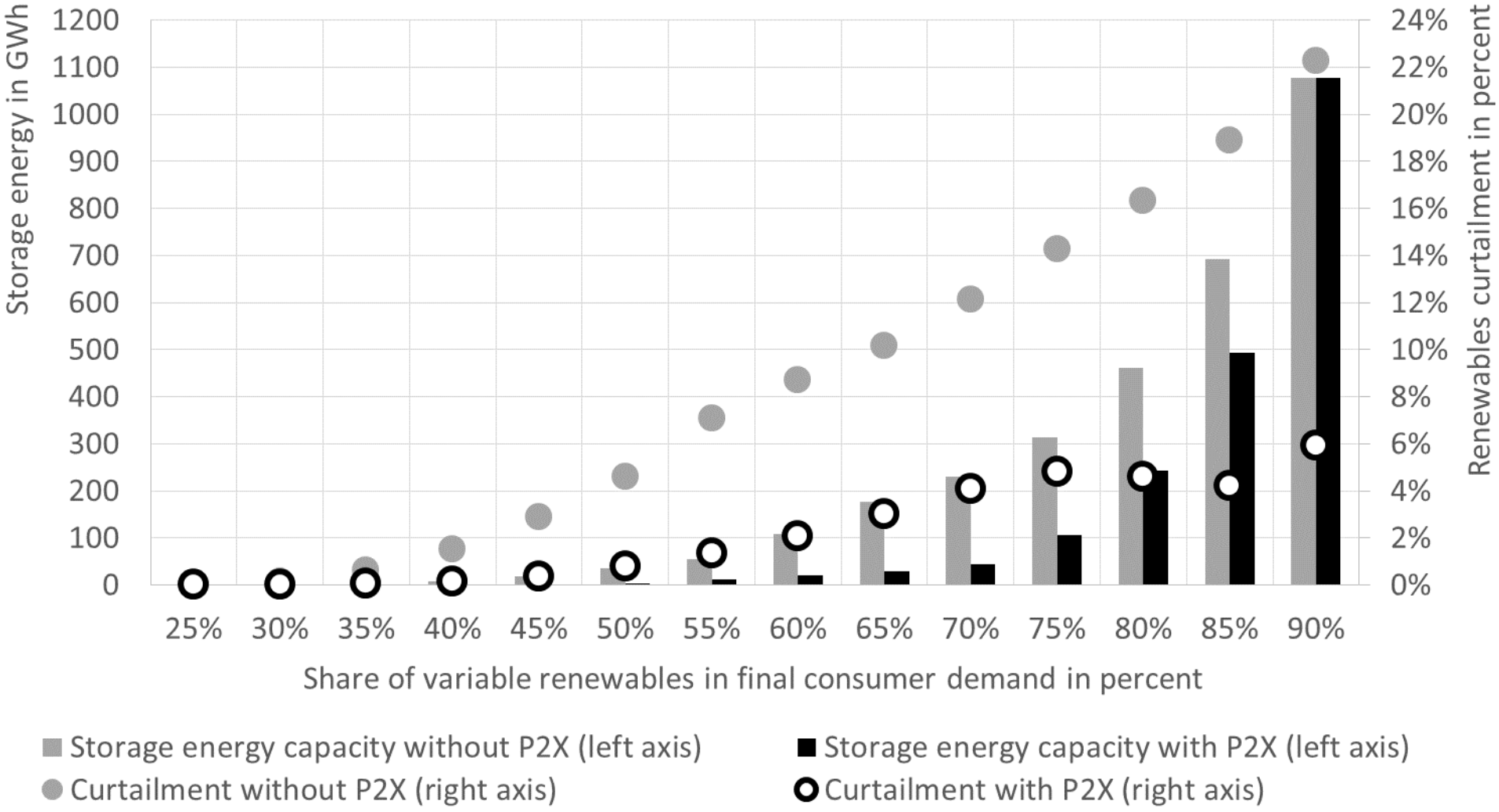
3 Optimization results

Endogenous storage and curtailment: still moderate storage capacities



Result: substantially lower storage and curtailment

- Renewable surplus is not a problem, but can be a valuable resource



4 Conclusions

- OPSD makes it easier to make use of open data
  - OPSD will be continuously updated and expanded
  - If you have feedback, please get in touch!
- Open data is not enough → open modeling / open science
  - Improve credibility and quality of model-based research
  - My view: benefits by far outweigh concerns
- Lessons from a recent example
  - Open replication studies are useful

Thank you for listening

DIW Berlin — Deutsches Institut für Wirtschaftsforschung e.V.
Mohrenstraße 58, 10117 Berlin
www.diw.de

Contact
Wolf-Peter Schill
wschill@diw.de