Energy System Modelling
Summer Semester 2020, Lecture 13

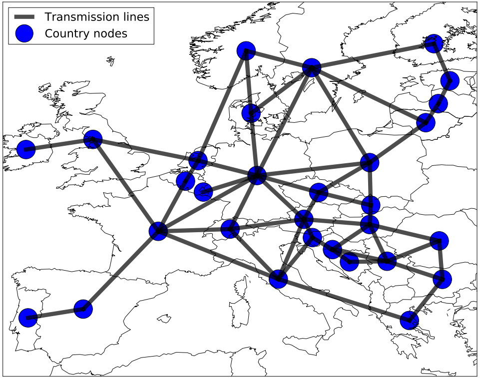
Dr. Tom Brown, tom.brown@kit.edu, https://nworbmot.org/
Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics (IAI)

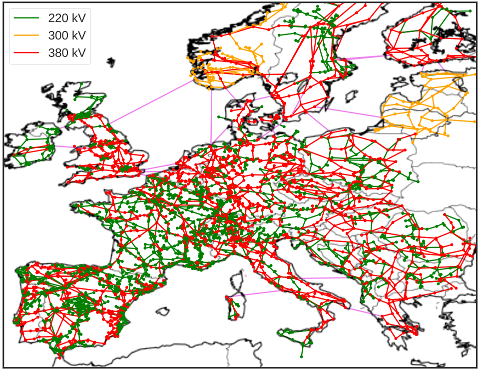
Table of Contents

Workflow management for complex models

Motivation: More detailed data-driven models, more scenarios

When we upgrade from the 30-node model to 5000+ nodes for a detailed grid model, and calculate 100 scenarios with different settings, **data management becomes important**.

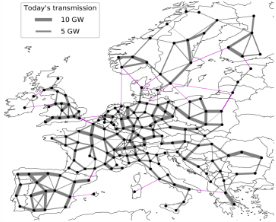




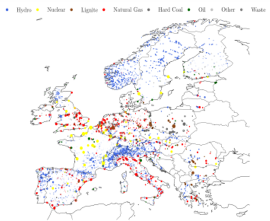
Data-Driven Modelling

Lots of different types of data come together for the modelling...

- Clustered network model

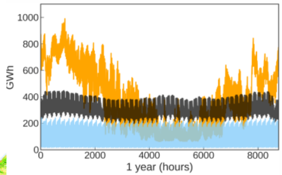


- Power plants & technology assumptions



- Renewable potentials & decades of hourly time series for each point in space

- Demand forecasts & time series



Analysis and optimisation

Problems

- Many different data sources
- Many data sources need cleaning and processing before they can be used
- Many intermediate scripts and datasets
- Many colleagues hack something together in a folder - hard to reproduce later
- Often dependencies are not clear (both data and software)
- Data and code change over time
- Want to run many parametric scenarios for same model

Problems

- Many different data sources
- Many data sources need cleaning and processing before they can be used
- Many intermediate scripts and datasets
- Many colleagues hack something together in a folder - hard to reproduce later
- Often dependencies are not clear (both data and software)
- Data and code change over time
- Want to run many parametric scenarios for same model

What we need is a **workflow management tool**.

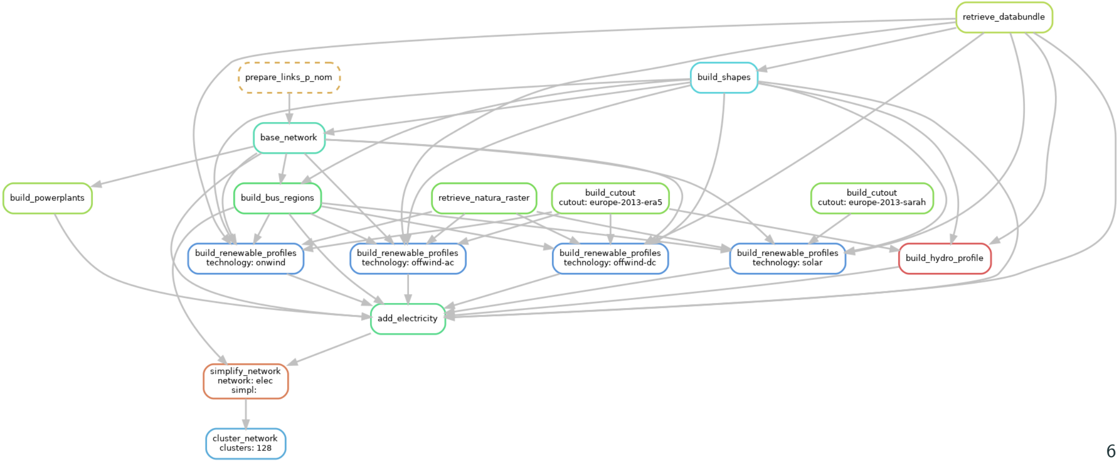
Workflow Management Tool: Snakemake

"The Snakemake workflow management system is a tool to create **reproducible and scalable data analyses**. Workflows are described via a human readable, Python based language. They can be seamlessly scaled to server, cluster, grid and cloud environments, without the need to modify the workflow definition. Finally, Snakemake workflows can entail a description of required software, which will be automatically deployed to any execution environment."

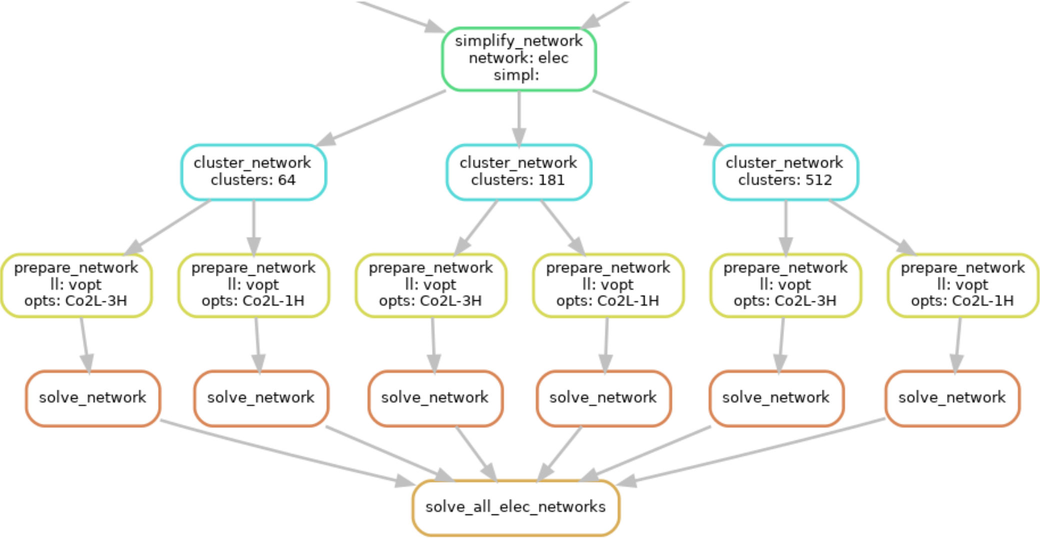
See **Snakemake presentation**.

Snakemake for PyPSA-Eur: Building Model

The **Snakemake** workflow management system is a tool to create **reproducible and scalable data analyses**. Dependency graph: **nodes** for scripts, **directed edges** map outputs to inputs.



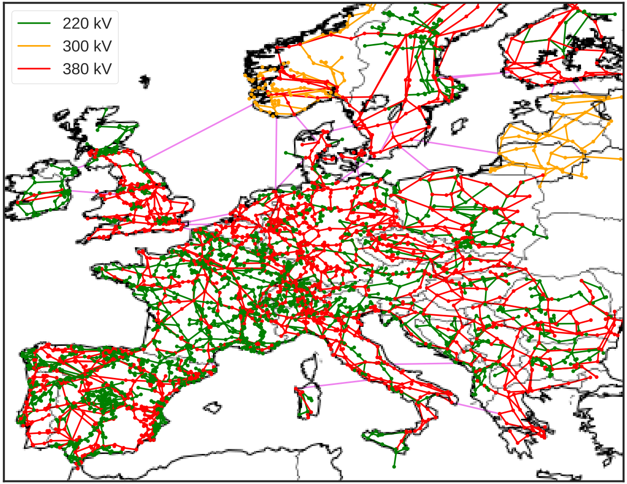
Snakemake for PyPSA-Eur: Managing Scenarios



PyPSA-Eur

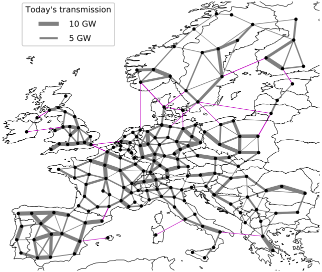
See **PyPSA-Eur slidedeck** for breakdown of PyPSA-Eur snakemake rules.

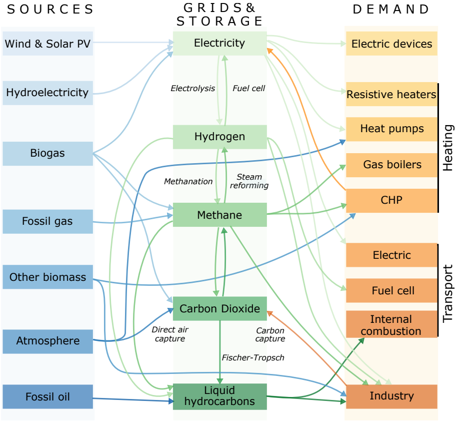
Checkout also the **PyPSA-Eur GitHub repository**.



PyPSA-Eur-Sec builds on PyPSA-Eur

- **PyPSA-Eur-Sec** includes PyPSA-Eur as a **snakemake subworkflow**.
- PyPSA-Eur builds the clustered power grid and renewable generators.
- Then PyPSA-Eur-Sec adds other sectors.



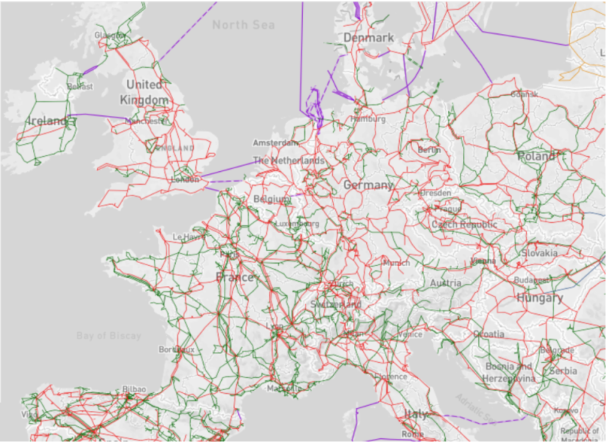


Effect of spatial scale on results of energy system optimisations

Motivation: Transmission bottlenecks

Many of the results we've examined so far have aggregated countries to a single node.

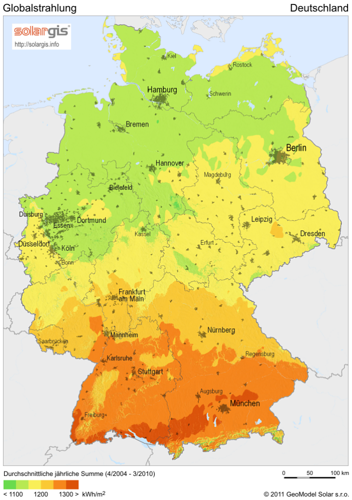
However, there are also transmission network bottlenecks **within** countries (e.g. North to South Germany).

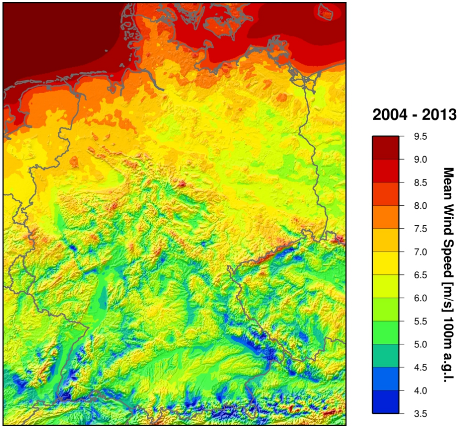


Source: ENTSO-E

Motivation: Wind and solar resource variation

There is also considerable variation in wind and solar resources...



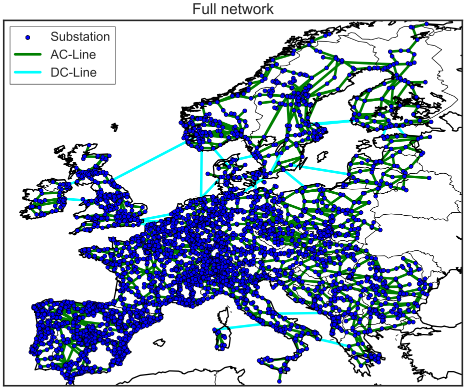


Spatial resolution

We need spatial resolution to:

- capture the **geographical variation** of renewables resources and the load
- capture **spatio-temporal effects** (e.g. size of wind correlations across the continent)
- represent important **transmission constraints**

BUT we do not want to have to model all 5,000 network nodes of the European system.



Source: Own representation of Bart Wiegman's GridKit extract of the online ENTSO-E map, https://doi.org/10.5281/zenodo.55853

Clustering: Many algorithms in the literature

There are lots of algorithms for clustering networks, particularly in the engineering literature:

- k-means clustering on (electrical) distance
- k-means on load distribution
- Community clustering (e.g. Louvain)
- Spectral analysis of Laplacian matrix
- Clustering of Locational Marginal Prices with nodal pricing (sees congestion and RE generation)
- PTDF clustering
- Cluster nodes with correlated RE time series

The algorithms all serve different purposes (e.g. reducing part of the network on the boundary, to focus on another part).

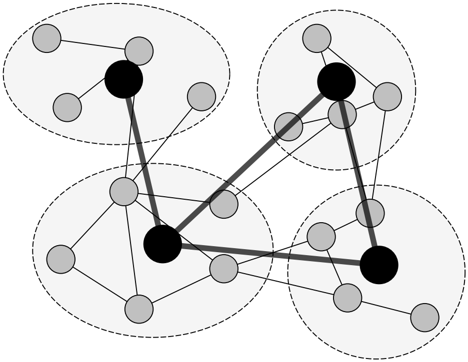
Not always tested on real network data.

k-means clustering on load & conventional generation

Our **goal**: maintain main transmission corridors of today to investigate highly renewable scenarios with no grid expansion. Since generation fleet is totally rebuilt, do not want to rely on current generation dispatch (like e.g. LMP algorithm).

Today's grid was laid out to connect big generators and load centres.

Solution: Cluster nodes based on spatial distribution using k-means, with a weighting to sites with higher average load and conventional generation capacity.



k-means clustering on load & conventional generation

Suppose the N nodes i have spatial coordinates (x_i, y_i). The k-means algorithm works by partitioning them into $k \leq N$ sets N_c for $c = 1, \ldots k$ such that the sum of squared distance to the centroid (x_c, y_c) (mean point inside each set) is minimised:

$$\min_{\{(x_c, y_c)\}} \sum_{c=1}^{k} \sum_{i \in N_c} w_i \left\| \begin{pmatrix} x_c \\ y_c \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right\|^2$$

Each node i is weighted w_i by the average load and the average conventional generation there.

Use the centroid as the location of the new clustered node.

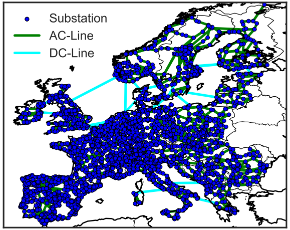
Reconstitution of network

Once the partition of nodes is determined:

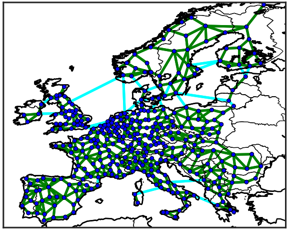
- A new node is created to represent each set of clustered nodes
- Hydro capacities and load is aggregated at the node; VRE (wind and solar) time series are aggregated, weighted by capacity factor; potentials for VRE aggregated
- Lines between clusters replaced by single line with length 1.25 $\times$ crow-flies-distance, capacity and impedance according to replaced lines
- $n-1$ blanket safety margin factor grows from 0.3 with ≥ 200 nodes to 0.5 with 37 nodes (to account for aggregation)

k-means clustering: Networks

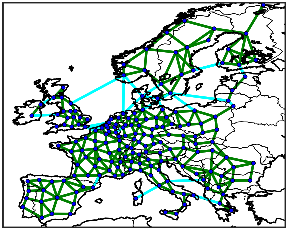
Full Network



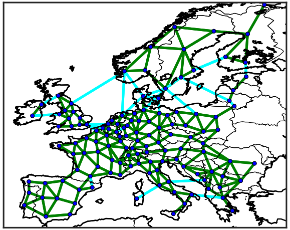
Network with 362 clusters



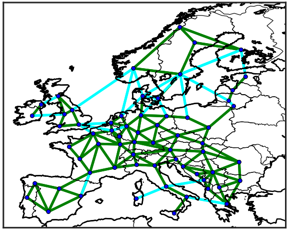
Network with 181 clusters



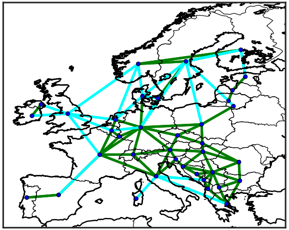
Network with 128 clusters



Network with 64 clusters



Network with 37 clusters



Question of spatial resolution

How is the overall minimum of the cost objective (building and running the electricity system) affected by an increase of spatial resolution in each country?

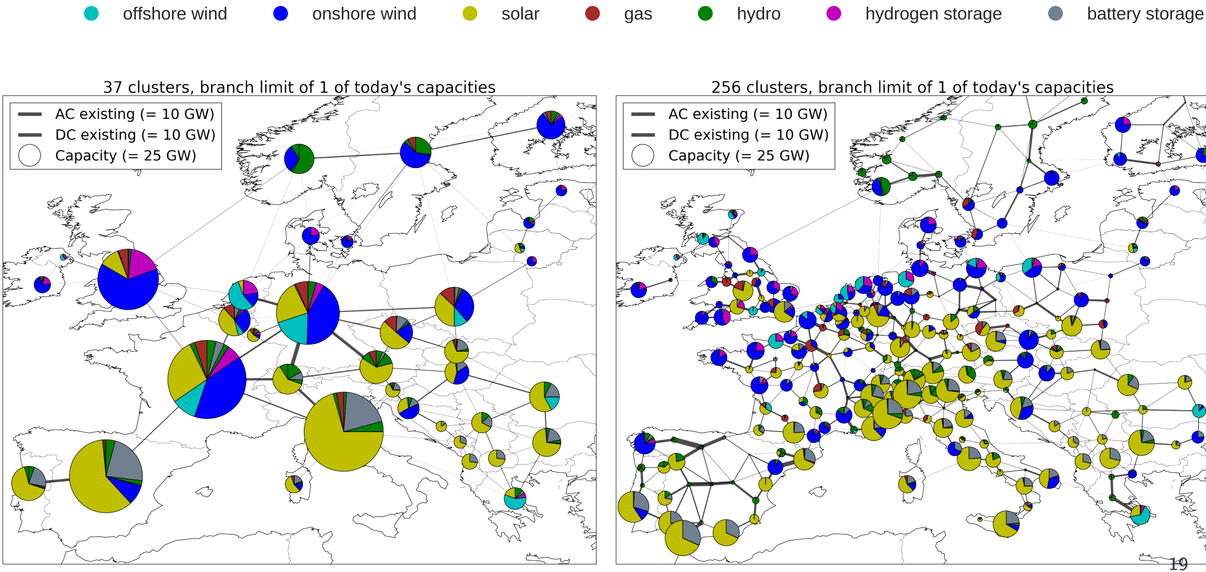
We expect

- A better representation of existing internal bottlenecks will prevent the transport of e.g. offshore wind to the South of Germany.
- Localised areas of e.g. good wind can be better exploited by the optimisation.

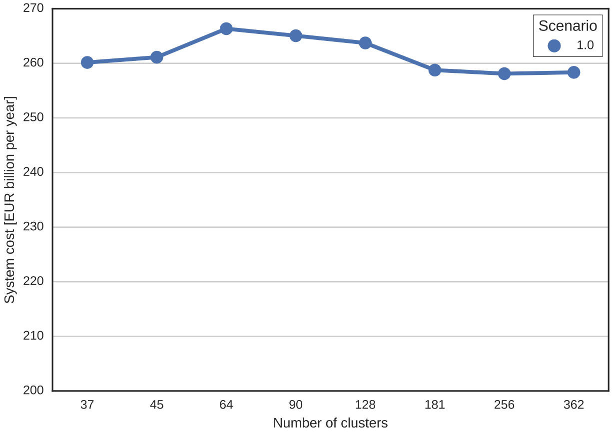
Which effect will win?

First we only optimize the gas, wind and solar generation capacities, the long-term and short-term storage capacities and their economic dispatch including the available hydro facilities **without grid expansion**.

Nodal energy shares per technology (w/o grid expansion)

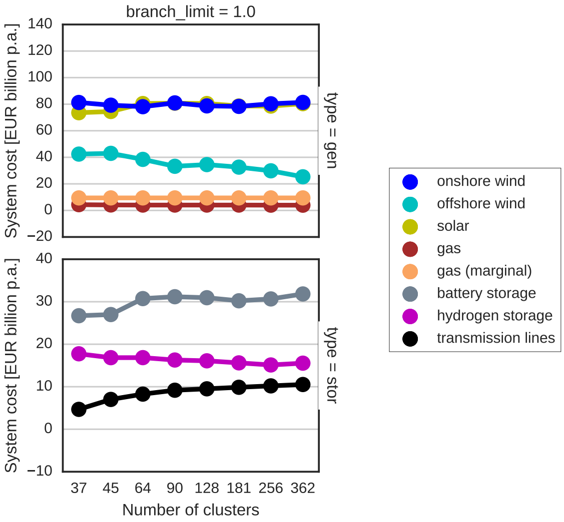


Costs: System cost w/o grid expansion



- Steady total system cost at € 260 billion per year
- This translates to € 82/MWh (compared to today of € 50/MWh to € 60/MWh)

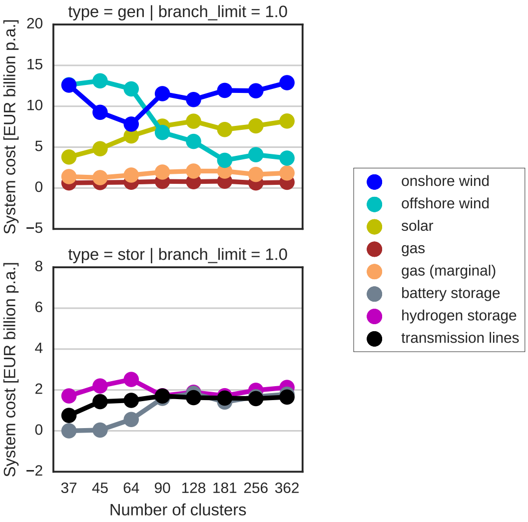
Costs: System cost and break-down into technologies (w/o grid expansion)



If we break this down into technologies:

- 37 clusters captures around half of total network volume
- Redistribution of capacities from offshore wind to solar
- Increasing solar share is accompanied by an increase of battery storage
- Single countries do not stay so stable

Costs: Focus on Germany (w/o grid expansion)



- Offshore wind replaced by onshore wind at better sites and solar (plus batteries), since the represented transmission bottlenecks make it impossible to transport the wind energy away from the coast
- the effective onshore wind capacity factors increase from 26% to up to 42%
- Investments stable at 181 clusters and above

Interaction between network expansion and spatial scale

6 different scenarios of network expansion by constraining the overall transmission line volume in relation to today's line volume $\mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}}$, given length d_ℓ and capacity F_ℓ of each line ℓ:

$$F_\ell \geq F_\ell^{\mathrm{today}} \tag{1}$$

$$\sum_\ell d_\ell F_\ell \leq \mathrm{CAP}_{\mathrm{trans}} \tag{2}$$

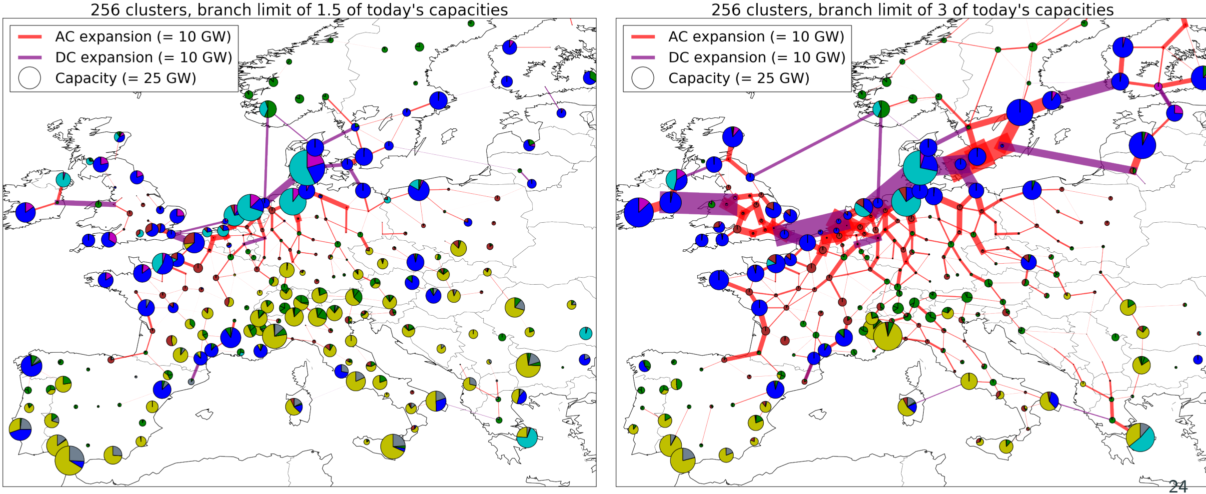
where

$$\mathrm{CAP}_{\mathrm{trans}} = x\,\mathrm{CAP}_{\mathrm{trans}}^{\mathrm{today}} \tag{3}$$

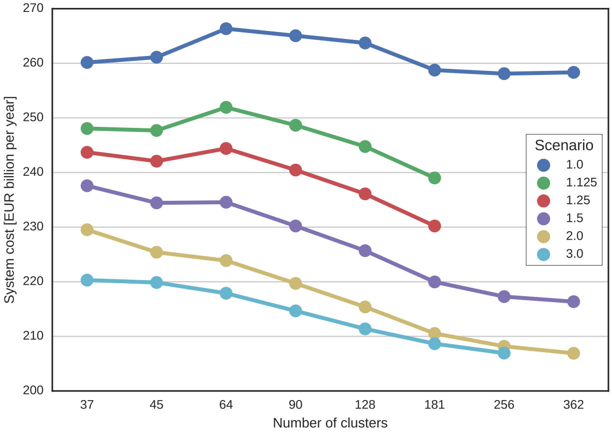
for $x = 1$ (today's grid) $x = 1.125, 1.25, 1.5, 2$, $x = 3$ (optimal for overhead line at high number of cluster).

With expansion

offshore wind · onshore wind · solar · gas · hydro · hydrogen storage · battery storage

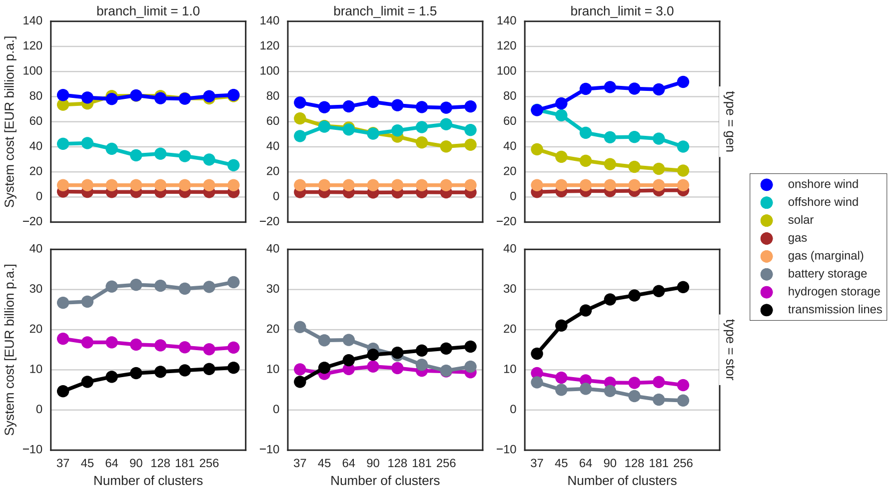


Costs: Total system cost

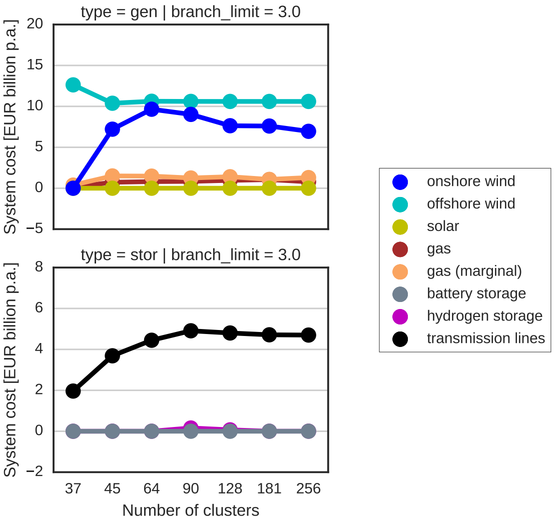


- Steady cost for No Expansion (1)
- For expansion scenarios, as clusters increase, the better expoitation of good sites decreases costs faster than transmission bottlenecks increase them
- Decrease in cost is v. non-linear as grid expanded (25% grid expansion gives 50% of optimal cost reduction)
- Only a moderate $20-25\%$ increase in costs from the Optimal Expansion scenario (3) to the No Expansion scenario (1).

Costs: Break-down into technologies

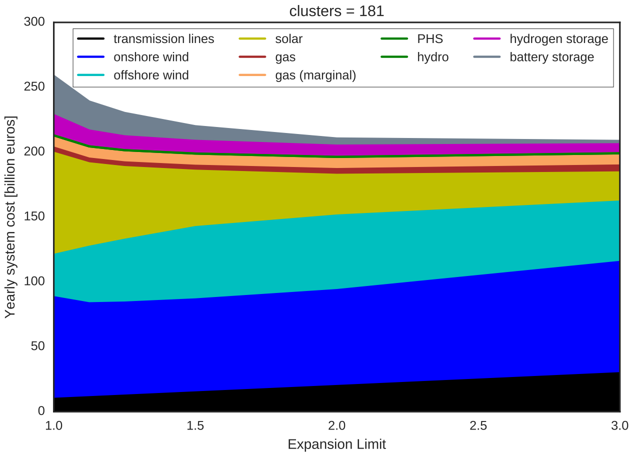


Costs: Focus on Germany (CAP = 3)



- Investment reasonably stable at 128 clusters and above
- System consistently dominated by wind
- No solar or battery for any number of clusters

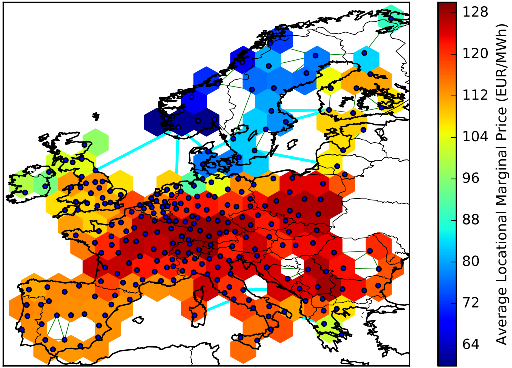
Behaviour as CAP is changed



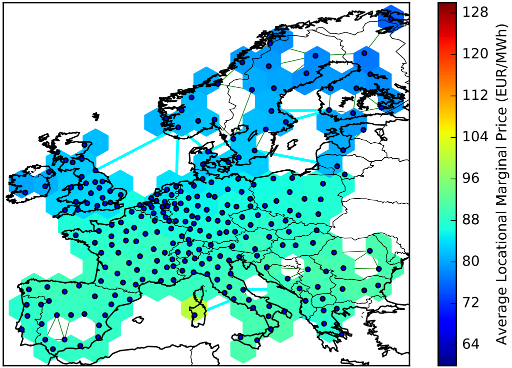
- Same non-linear development with high number of nodes that we saw with one node per country
- Most of cost reduction happens with small expansion; cost rather flat once capacity has doubled, reaching minimum (for overhead lines) at 3 times today's capacities
- Solar and batteries decrease significantly as grid expanded
- Reduction in storage losses too

Locational Marginal Prices CAP=1 versus CAP=3

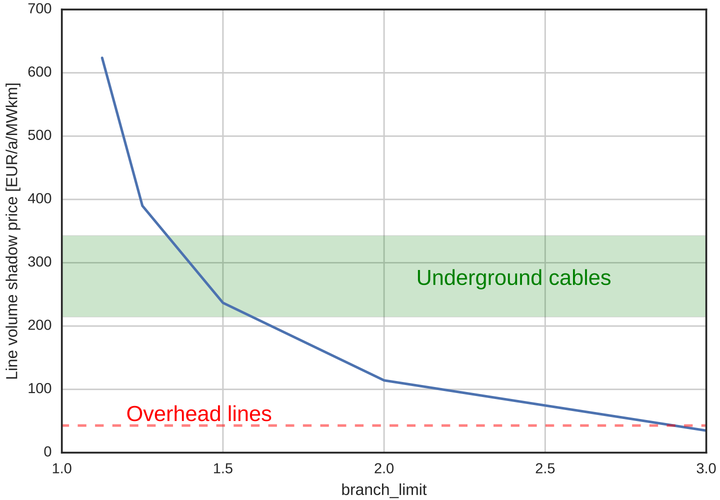
With today's capacities:



With three times today's grid:

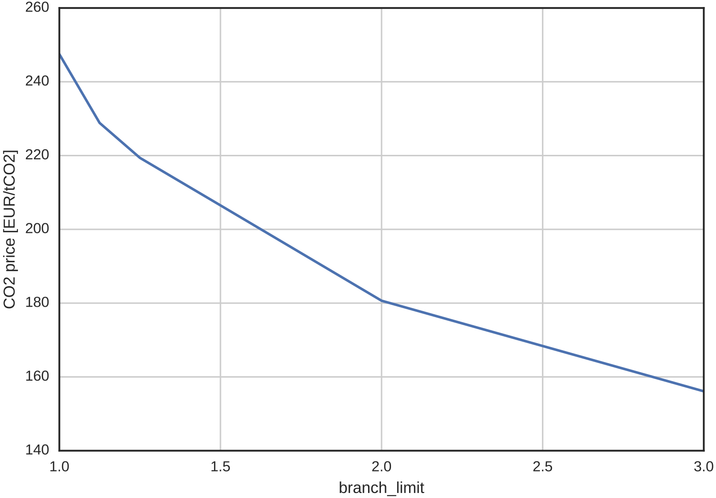


Grid expansion CAP shadow price for 181 nodes as CAP relaxed



- With overhead lines the optimal system has around 3 times today's transmission volume
- With underground cables (5-8 times more expensive) the optimal system has around 1.3 to 1.6 times today's transmission volume

CO2 prices versus line cap for 181 clusters



- CO_2 price of between 150 and 250 €/tCO_2 required to reach these solutions, depending on line volume cap

More Details in Paper

For more details, see the following paper:

- J. Hörsch, T. Brown, "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios," EEM 2017, **link**.

In an upcoming paper with Martha Frysztacki and the same authors, we disentangle the effects of the network resolution from the renewable resource resolution.

Conclusions

- Generation costs always dominate grid costs, but the grid can cause higher generation costs if expansion is restricted
- Systems with no grid extension beyond today are up to 25% more expensive, but small grid extensions (e.g. 25% more capacity than today) can lock in big savings
- Need at least around 200 clusters for Europe to see grid bottlenecks if no expansion
- Can get away with $\sim$ 120 clusters for Europe if grid expansion is allowed
- This is **no single solution** for highly renewable systems, but a **family of solutions** with different costs and compromises
- Much of the stationary storage needs can be eliminated by sector-coupling: DSM with electric vehicles, thermal storage; this makes grid expansion less beneficial
- Understanding the need for **flexibility at different temporal and spatial scales** is key to mastering the complex interactions in the energy system

Cycle formulations of optimal power flow

Angle-based formulation of linear optimal power flow is slow

The most common way of implementing optimization models with linear power flow is to use the **angle formulation**. We start with energy conservation (KCL)

$$p_i = \sum_\ell K_{i\ell} f_\ell$$

This puts $N-1$ constraints on the L flows f_ℓ (since $\sum_i K_{i\ell} = 0$). For KVL we add N **auxiliary variables** for the voltage angles θ_i with $L+1$ additional constraints on f_ℓ and θ_i:

$$f_\ell = \frac{1}{x_\ell} \sum_i K_{i\ell} \theta_i$$

$$\theta_0 = 0$$

Check totals: $N+L$ variables θ_i, f_ℓ with $N-1+L+1 = L+N$ independent, sparse constraints $\Rightarrow \theta_i, f_\ell$ fully determined by the p_i, which is what we want.

But we don't really care about the angles θ_i and they introduce more variables and constraints.

Is there a better way?

Cycle-based "Kirchhoff" formulation is faster

The cycle-based **Kirchhoff formulation** avoids the auxiliary variables θ_i altogether by implementing the Kirchhoff Voltage Law (KVL) directly on the flows f_ℓ themselves.

We start again with our $N-1$ KCL constraints:

$$p_i = \sum_\ell K_{i\ell} f_\ell$$

and add the $L-N+1$ cycle constraints of KVL from Lecture 4:

$$\sum_\ell C_{\ell c} x_\ell f_\ell = 0$$

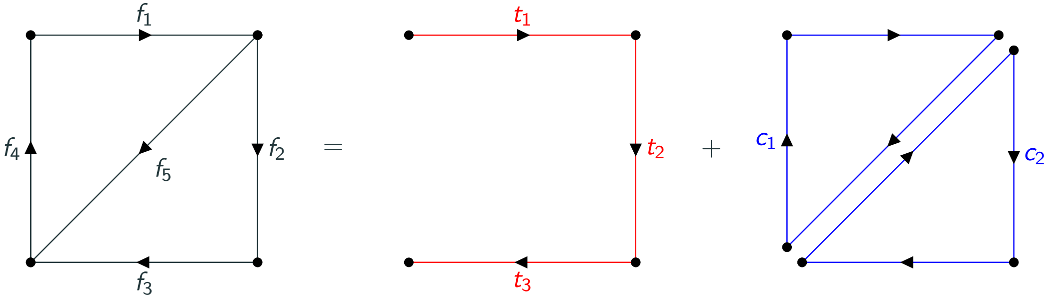
Check totals: L variables f_ℓ with $N-1+L-N+1=L$ independent, sparse constraints $\Rightarrow f_\ell$ fully determined by the p_i, which is what we want.

This has fewer variables and fewer constraints than the angle-based formulation, so we can expect it to perform better.

Cycle formulation of linear power flow

A third **cycle formulation** decomposes the flows in the network into two parts:

1. A flow on a spanning tree of the network, uniquely determined by nodal $\mathbf{p}$ (ensuring KCL)
2. Cycle flows, which don't affect KCL; their strength is fixed by enforcing KVL



$$f_\ell = t_\ell + \sum_k C_{\ell,k} c_k$$

Cycle formulation of linear power flow

The $N-1$ tree flows $\mathbf{t}$ are determined directly from the N nodal powers p_n and the network power balance constraint $\sum_n p_n = 0$.

We solve for the $L-N+1$ cycle flows c_k by enforcing the $L-N+1$ KVL equations:

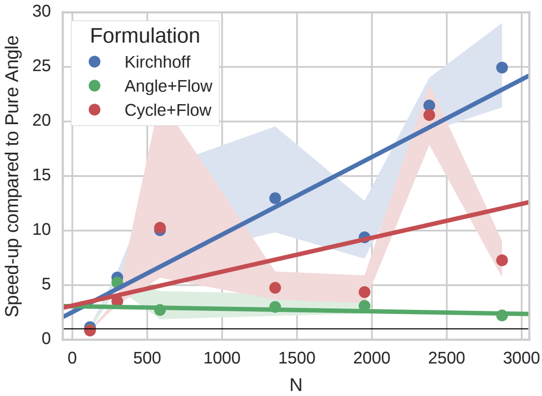
$$C^t X \mathbf{f} = C^t X(\mathbf{t} + C\mathbf{c}) = 0$$

The matrix C is the incidence matrix of the **weak dual graph**, $C^t XC$ is the weighted Laplacian of the dual graph and the above equation becomes a discrete Poisson equation:

$$C^t XC\mathbf{c} = -C^t X\mathbf{t}$$

Now we have only $L-N+1$ variables c_k with $L-N+1$ independent, **semi-dense** constraints.

LOPF speedup with cycle flows



Using **cycle flows instead of voltage angles** we found for generation expansion optimisation (fixed grid):

- A speed-up of up to **200 times**
- Average speed-up of **factor 12**
- Speed-up is highest for **large networks with lots of renewables**

H. Ronellenfitsch, D. Manik, J. Hörsch, **T. Brown, D. Witthaut**, *"Dual theory of transmission line outages,"* 2017, IEEE Transactions on Power Systems, link.

J. Hörsch, H. Ronellenfitsch, **D. Witthaut, T. Brown**, *"Linear Optimal Power Flow Using Cycle Flows,"* 2017, Electric Power Systems Research, link.