Scenario Building

Weather Year Selection Methodology

Method overview

- Select range of weather scenarios for each target year (2030, 2035, 2040, 2050)
- Calculate key figures from climate variables (variability and average) and normalize
- Consider regions according to their relevance (capacity, load)
- Reduce dimensions (PCA)
- Apply k-means clustering to find most representative series representing weather scenarios

Data used

Wind power generation profiles (PECD*, hourly)

- Onshore (existing technologies)
- Offshore (existing technologies)

Solar generation profiles (PECD, hourly)

Hydro generation (PECD, weekly):

- Reservoir
- Run of River

HDD/CDD (heating and cooling degree days) derived from Temperature (PECD) as a proxy for Load** (daily)

Renewables installed capacities (PEMMDB) and Load (ETM) (per target year)

*PECD 4.2 SSP 2.45 climate models: CMR5, ECE3, MEHR. For information on the PECD please see: https://cds.climate.copernicus.eu/datasets/sis-energy-pecd?tab=overview & https://confluence.ecmwf.int/pages/viewpage.action?pageId=439598955 the PECD 4.2 will be uploaded on the website soon.
** uses threshold for cooling and heating from DFT, with exception of Italy which provided own values

Details Methodology

HDD/CDD Calculation Method

Heating Degree Days (HDD)

```
HDD = t_threshold_heating - t_average
(when t_threshold_heating > t_average, otherwise 0)
```

Cooling Degree Days (CDD)

```
CDD = t_average - t_threshold_cooling
(when t_threshold_cooling < t_average, otherwise 0)
```

Example Thresholds

Austria (AT): Heating threshold = 12.3°C, Cooling threshold = 25.5°C

t_average = daily average temperature

Range of years used for selection

Target Years & Windows

- TY2030: 2025-2034
- TY2035: 2040-2049
- TY2040: 2035-2044
- TY2050: 2045-2054

Climate Models & Candidates

Climate Models: CMR5, MEHR, ECE3

Total Candidates per Target: 3 models × 10 years = 30 climate-year series

Selection Outcome: 3 representative series per target year

Methodology Overview

1

Calculate Average Values

Yearly averages & cumulative anomalies

→

2

Overall Statistics

Mean, std dev, and deltas

→

3

Weighting Factors

Regional capacity weighting

→

4

Normalize Parameters

Standardization for comparison

→

5

PCA & K-Means

Dimensionality reduction & clustering

Step 1: Calculate Yearly Average Values

Annual Mean Calculation

$$\mu_{v,r,y,m} = \frac{1}{T} \sum_{t=1}^{T} x_{v,r,y,m}(t)$$

$\mu_{v,r,y,m}$: annual mean of variable v in region r for year y and weather scenario m;
T: number of time steps in the year;
$x_{v,r,y,m}(t)$: value of variable v at time t.

Cumulative Anomaly Calculation

$$CA_{v,r,y,m} = \sum_{t=1}^{T} \left[x_{v,r,y,m}(t) - \overline{x_{v,r}}(t) \right]$$

$CA_{v,r,y,m}$: cumulative anomaly for variable v;
$x_{v,r,y,m}(t)$: variable value at time t;
$\overline{x_{v,r}}(t)$: multi-model average at time t.

Step 2: Overall Statistics & Deltas

Aggregate Mean

$$\mu_{v,r}^{tot} = mean_{y,m}(\mu_{v,r,y,m})$$

$\mu_{v,r}^{tot}$: overall mean of μ_{v,r,y,m} across all years y and weather scenarios m.

Aggregate Standard Deviation

$$\sigma_{v,r}^{tot} = std_{y,m}(\mu_{v,r,y,m})$$

$\sigma_{v,r}^{tot}$: standard deviation of $\mu_{v,r,y,m}$ across all years y and weather scenarios m.

Delta Calculations

$$\Delta\mu_{v,r,y,m} = \mu_{v,r,y,m} - \mu_{v,r}^{tot}; \Delta\sigma_{v,r,y,m} = CA_{v,r,y,m} - \sigma_{v,r}^{tot}$$

$\Delta\mu_{v,r,y,m}$: deviation of annual mean from aggregate mean;

$\mu_{v,r}^{tot}$: overall mean of μ_{v,r,y,m} across all years y and weather scenarios m

$\sigma_{v,r}^{tot}$: standard deviation of $\mu_{v,r,y,m}$ across all years y and weather scenarios m.

$CA_{v,r,y,m}$: cumulative anomaly for variable v;

Purpose: These deltas represent deviations from the overall baseline, enabling comparison across different scenarios and time periods.

Step 3: Calculate Weighting Factors

Regional Weighting Factor

$$w_{r,y} = \frac{B_{r,y}}{\sum_r B_{r,y}}$$

$w_{r,y}$: weight for region r in year y;

$B_{r,y}$: base metric (e.g., capacity or load) for region r.

Weighted Deltas

$$\Delta\mu^{w}_{v,r,y,m} = w_{r,y}\Delta\mu_{v,r,y,m}; \Delta\sigma^{w}_{v,r,y,m} = w_{r,y}\Delta\sigma_{v,r,y,m}$$

Weighted deviations: multiply deltas by regional weight $w_{r,y}$.

Important Considerations

Use installed capacity data for the correct target year

Exclude regions not in analysis (Turkey, Ukraine, Moldova)

Consider country-level data if regional data unavailable

Step 4: Normalize Parameters

Standardization Formula

$$Z_i = \frac{\Delta_i^w - mean(\Delta^w)}{std(\Delta^w)}$$

Z_i: standardized weighted delta for series i; $mean(\Delta^w)$, $std(\Delta^w)$: mean and standard deviation of weighted deltas across all series.

Purpose of Normalization

Brings all parameters to a similar scale, facilitating meaningful PCA and K-Means analysis. Essential for comparing variables with different units and ranges.

Process Steps

- Subtract overall average from each target year's value
- Divide by standard deviation
- Repeat for all parameters

Step 5a: Principal Component Analysis (PCA)

Input Data Structure

Year	a_Var1	SD_Var1	a_Var2	SD_Var2	...	a_Var12	SD_Var12
WY01	-0.12	0.45	...	...	...	...	...
WY02	0.08	-0.34	...	...	...	...	...
...	...	...	...	...	...	...	...
WY30	1.21	-0.77	...	...	...	...	...

PCA Output: Component Loadings

Loadings show how much each original variable contributes to principal components:

- **PC1:** Largely driven by a_Var1, SD_Var1, and a_Var2
- **PC2:** Influenced by SD_Var2 and other variables
- **Dimensionality Reduction:** From 6+ variables to 2-3 components

Step 5b: K-Means Clustering

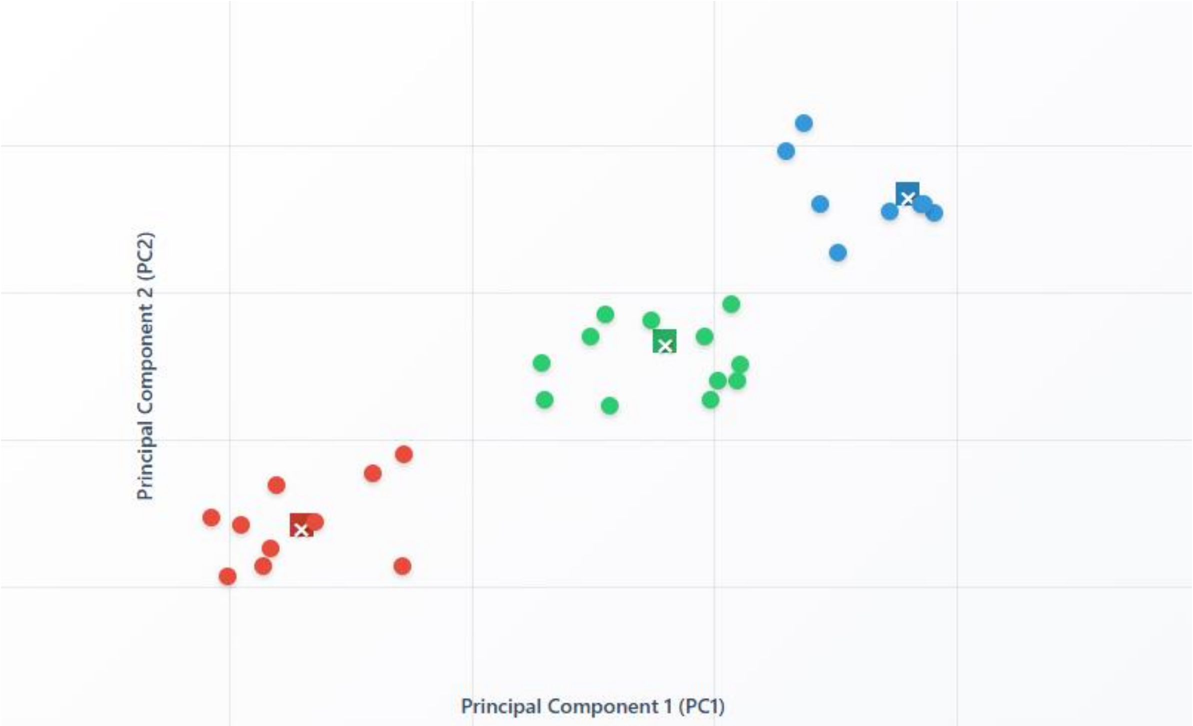
Algorithm Execution

1.Choose k=3 (for 3 representative scenarios)

2.Initialize three centroids (k-means++ or random)

3. Iterate until convergence:

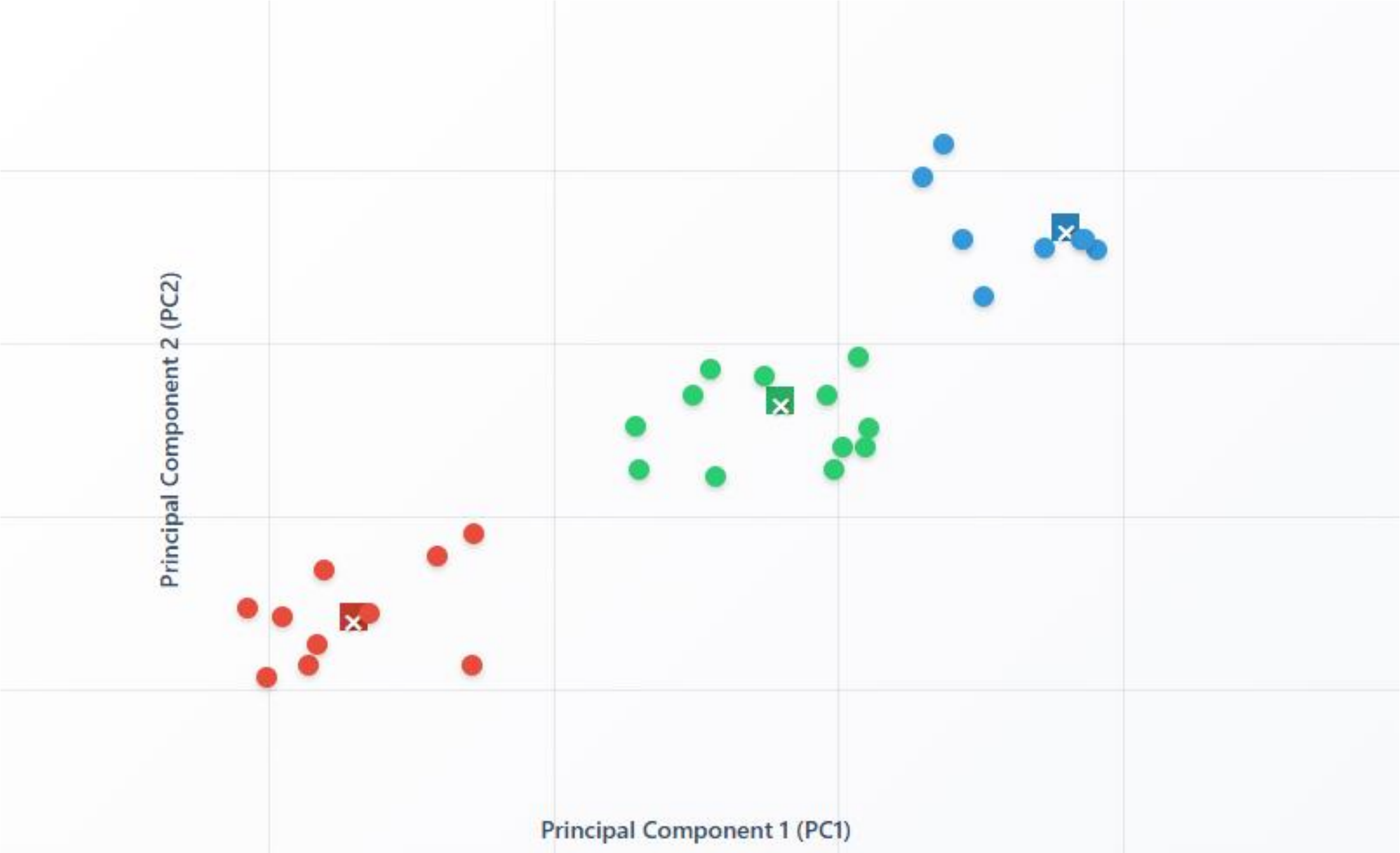
- Assignment step: assign each point to nearest centroid
- Update step: recompute centroids as mean of assigned points



Representative Year Selection

- For each centroid, compute Euclidean distance to every series
- Select climate-year series closest to each centroid
- Three selected indices form the representative years

Step 5b: K-Means Clustering



Methodology Explanation

Data Points: Each point represents one climate-year series from the 30 candidates (3 climate models × 10 years per target period).

Axes: Principal Component 1 and 2 capture the main variance in the normalized, weighted weather parameters.

Clustering: K-means algorithm groups similar weather patterns together based on Euclidean distance in PCA space.

Selection: The climate-year series closest to each centroid (×) becomes the representative weather scenario for that cluster.

Results

Results

Code name	Time period type	SSP scenario	Climate model	Climate year	Study Target Year
WS003	Projection	SSP245	CMR5	2027	2030
WS021	Projection	SSP245	MEHR	2025	2030
WS029	Projection	SSP245	MEHR	2033	2030
WS032	Projection	SSP245	CMR5	2031	2035
WS037	Projection	SSP245	CMR5	2036	2035
WS059	Projection	SSP245	MEHR	2038	2035
WS065	Projection	SSP245	CMR5	2039	2040
WS071	Projection	SSP245	ECE3	2035	2040
WS077	Projection	SSP245	ECE3	2041	2040
WS091	Projection	SSP245	CMR5	2045	2050
WS092	Projection	SSP245	CMR5	2046	2050
WS106	Projection	SSP245	ECE3	2050	2050

Remark: CMR5 overrepresented, but no significance from climate point of view, thus no reason for concerns

Appendix

HDD/CDD calculation

HDD and CDD are calculated for each region as follows:

- $HDD = t_{threshold\_heating} - t_{average} \; \forall \; t_{threshold_{heating}} > t_{average} \; otherwise \; 0$
- $CDD = t_{average} - t_{threshold_{cooling}} \; \forall \; t_{threshold_{cooling}} < t_{average} \; otherwise \; 0$

$t_{threshold\_\_heating}$… temperature threshold, below which heating is necessary, eg. 12.3°C for AT
$t_{threshold\_\_cooling}$… temperature threshold, above which cooling is necessary, eg. 25.5°C[1] for AT
$t_{average}$… average daily temperature

The sum of the HDD and CDD is used as one of the variables. It also serves as an indicator for seasonal patterns.

Thank you