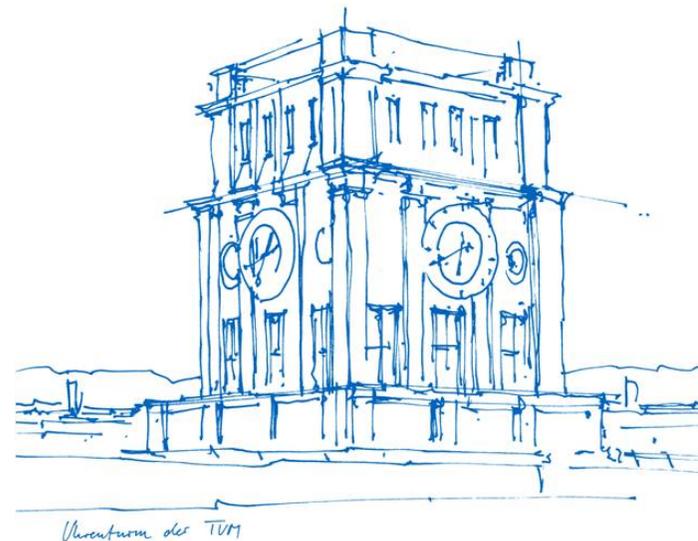


Same Same but Different?

Clusterung von Länderdaten zu Metaregionen mittels Machine Learning Methoden

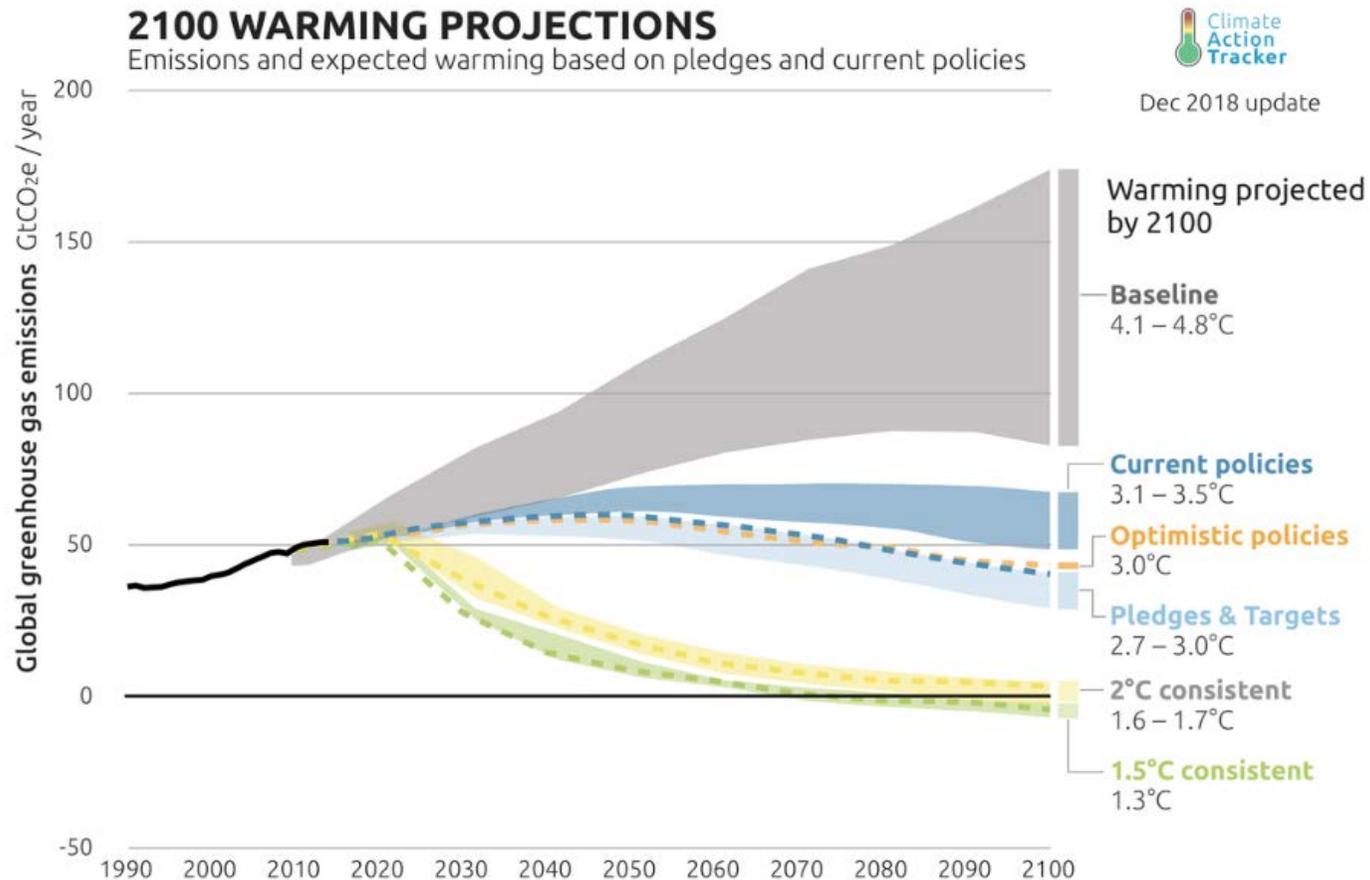
Clara Orthofer

Wien, 14. Februar 2019



Globale Energiesystemmodelle – Wer bekommt wieviel vom Kuchen?

Für das 1.5°C Ziel bedarf es drastischer Veränderung aller Energiesysteme der Welt. Wollen wir das Zeit erreichen müssen erneuerbare Energien die fossilen Energieträger ablösen.



Traditionelle Modellregionenbildung

Globale Energiesystemmodelle teilen die Welt traditionell in rund 10-20 Regionen auf, die v.a. die geographischen Gegebenheiten sowie nicht quantifizierbare qualitative "soft facts" berücksichtigen.

Modell	Version	Regionen
TIMES -TIAM	UCL	16
	NZ	16
	ETSAP	15
OSeMOSYS	GENeSYS-MOD	10
POLES	JRC	66
MESSAGE	GEA	11

Herleitung:

- v.a. historisch bedingt
 - Berücksichtigung geografischer Nähe & politischer Ausrichtung
- Viel über Erfahrung der Modellierer

Einflussfaktoren:

- Staatsform und Staatshistorie
- Mitgliedschaft in Handels- und/oder politischen Unionen
- Ökonomische Situation
- Geschwindigkeit des technologischen Wandels
- Demografische Entwicklung
- politische Ambitionen und Ziele

Traditionelle Modelregionenbildung

Die Welt abbildende Energiemodelle teilen die Welt traditionell in rund 10-20 Regionen auf, die v.a. die geographischen Gegebenheiten sowie nicht quantifizierbare qualitative "soft facts" berücksichtigen.

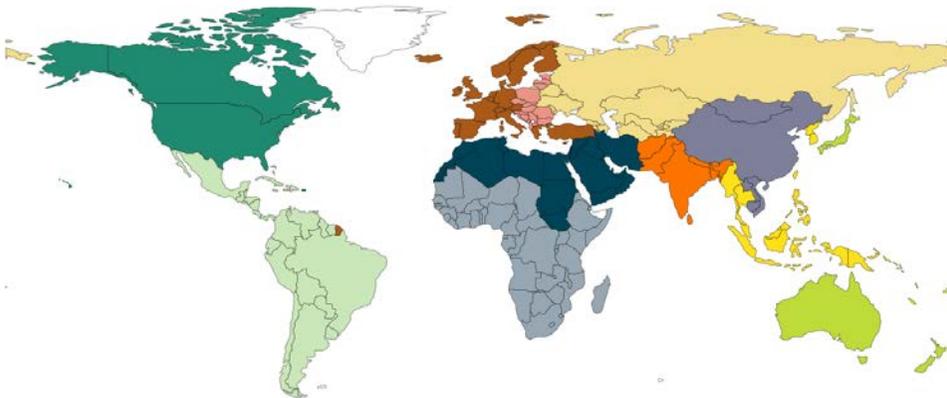


Abb.: Die 11 MESSAGE Regionen

Herleitung:

- v.a. historisch bedingt
- Berücksichtigung geografischer Nähe & politischer Ausrichtung
- Viel über Erfahrung der Modellierer

Einflussfaktoren:

- Staatsform und Staatshistorie
- Mitgliedschaft in Handels- und/oder politischen Unionen
- Ökonomische Situation
- Geschwindigkeit des technologischen Wandels
- Demografische Entwicklung
- politische Ambitionen und Ziele

Explorative Datenanalyse: Identifikation & Erzeugung relevanter Features

Visuelle und quantitative Analyse mehrere dutzend Parameter zur Identifikation eines nachvollziehbaren Featuresets zur Abbildung von EE-getriebener Energiesystementwicklung.

Untersuchte Features

Datenpunkte je Jahr und Land

- Primärenergiebedarf	1
- Endenergieverbrauch	1
- Umwandlungseffizienz	1
- Detaillierte aufgeschlüsselte erneuerbare Energien-Potentiale	>150
- aufgeschlüsselte fossile Ressourcen und Reserven	bis zu 10
- spezifische Primär- und Endenergieverbräuche pro BIP und pro Kopf	2
- ...	
- Bruttoinlandsprodukt (absolut und pro Kopf)	2
- Moodies Rating	1
- Human Development Index (HDI)	1
- Elektrifizierungsgrad	1
- Urbanisierungsgrad	1
- Anteil erneuerbarer Energien am Energiemix	10
- ...	

Explorative Datenanalyse: Reduktion der EE-Potentialbeschreibung

Zur Reduktion der komplexen Beschreibung der erneuerbaren Energien Potentiale wurde der Solar- und der Windenergiefaktor als Maß der Potentialgröße eingeführt.

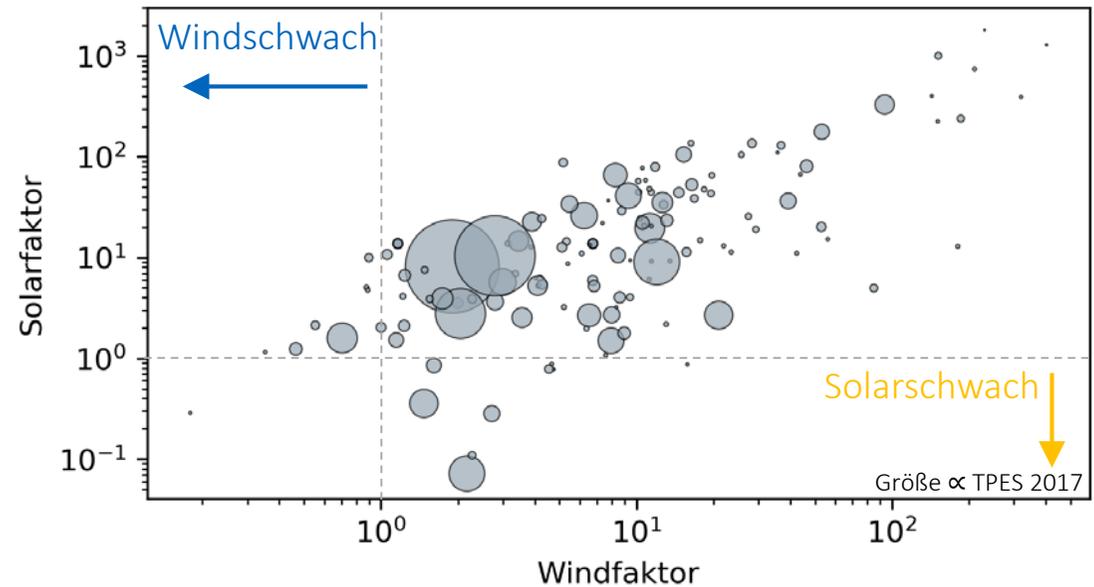
$$\text{Potential Faktor}_L = \frac{\sum_{d,k}^{D,K} \text{Potential}_{d,k,l}}{\text{TPES}_l}$$

l, L ... Land

k, K ... Kapazitäts – Kategorie

d, D ... Distanz zum Verbraucher

TPES ... Primärenergieverbrauch



Explorative Datenanalyse: Reduktion der EE-Potentialbeschreibung

Zur Bewertung der Potentiale und zur Unterscheidung ihrer Güte wurden die gemittelten Stromgestehungskosten sowie die gemittelten Volllaststunden definiert und berechnet.

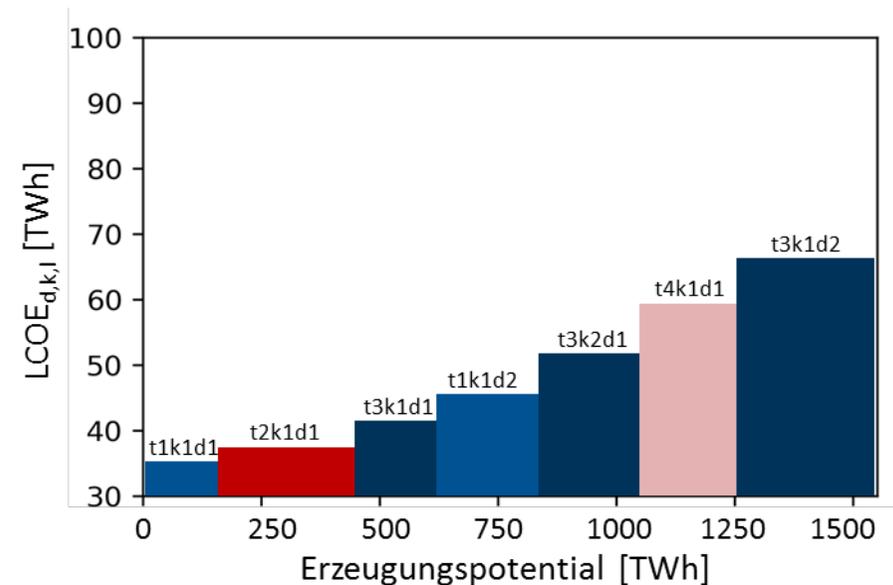
Stromgestehungskosten je Distanz- und Kapazitätskategorie:

$$LCOE_{d,k,l} = \frac{\text{inv. Kosten}_l + \text{Anschlusskosten}_d}{\text{Lebensdauer} * VLh_{d,k,l}} + \frac{\text{fix. Kosten}_l}{VLh_{d,k,l}} + \text{var. Kosten}_l$$

l, L ... Land

k, K ... Kapazitäts – Kategorie

d, D ... Distanz zum Verbraucher



Explorative Datenanalyse: Reduktion der EE-Potentialbeschreibung

Zur Bewertung der Potentiale und zur Unterscheidung ihrer Güte wurden die gemittelten Stromgestehungskosten sowie die gemittelten Volllaststunden definiert und berechnet.

Stromgestehungskosten je Distanz- und Kapazitätskategorie:

$$LCOE_{d,k,l} = \frac{\text{inv. Kosten}_l + \text{Anschlusskosten}_d}{\text{Lebensdauer} * VLh_{d,k,l}} + \frac{\text{fix. Kosten}_l}{VLh_{d,k,l}} + \text{var. Kosten}_l$$

l, L ... Land

k, K ... Kapazitäts – Kategorie

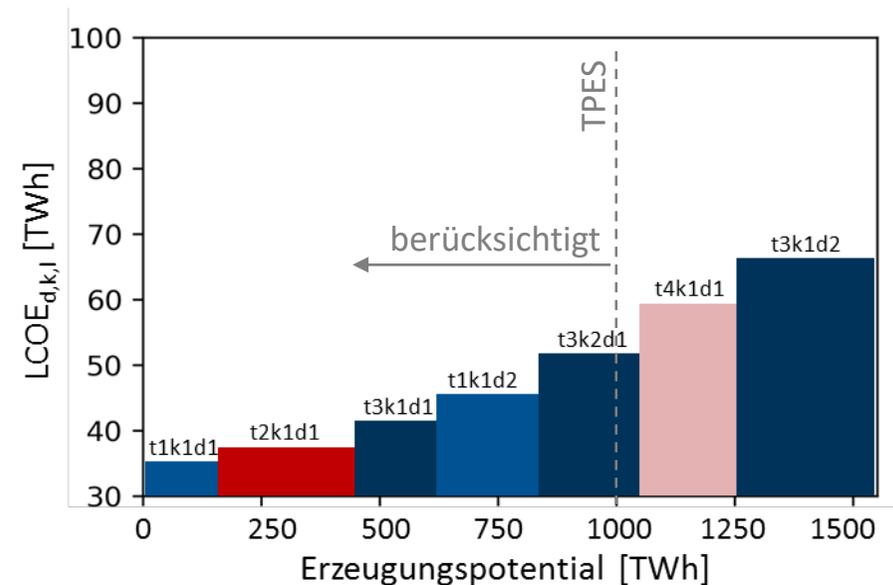
d, D ... Distanz zum Verbraucher

Gewichtete Kosten- & VLh der kostengünstigsten Kategorien:

$$\emptyset LCOE_l = \frac{\sum_{d,k}^{D,K} LCOE_{d,k,l} * \text{Potential}_{d,k,l} * VLh_{d,k,l}}{TFC_l}$$

$$\emptyset VLh_l = \frac{\sum_{d,k}^{D,K} VLh_{d,k,l} * \text{Potential}_{d,k,l}}{TFC_l}$$

$$\forall k, d \in \sum_{d,k}^{D,K} \langle \text{Potential}_{d,k,l} * VLh_{d,k,l} \rangle \leq TFC_l$$



Explorative Datenanalyse: Reduktion der EE-Potentialbeschreibung

Zur Bewertung der Potentiale und zur Unterscheidung ihrer Güte wurden die gemittelten Stromgestehungskosten sowie die gemittelten Volllaststunden definiert und berechnet.

Stromgestehungskosten je Distanz- und Kapazitätskategorie:

$$LCOE_{d,k,l} = \frac{\text{inv. Kosten}_l + \text{Anschlusskosten}_d}{\text{Lebensdauer} * VLh_{d,k,l}} + \frac{\text{fix. Kosten}_l}{VLh_{d,k,l}} + \text{var. Kosten}_l$$

l, L ... Land

k, K ... Kapazitäts – Kategorie

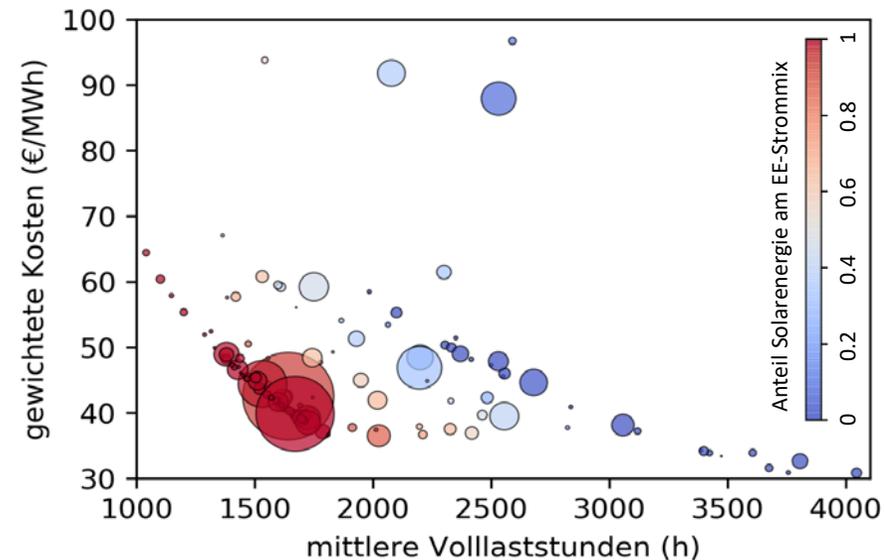
d, D ... Distanz zum Verbraucher

Gewichtete Kosten- & VLh der kostengünstigsten Kategorien:

$$\emptyset LCOE_l = \frac{\sum_{d,k}^{D,K} LCOE_{d,k,l} * \text{Potential}_{d,k,l} * VLh_{d,k,l}}{TFC_l}$$

$$\emptyset VLh_l = \frac{\sum_{d,k}^{D,K} VLh_{d,k,l} * \text{Potential}_{d,k,l}}{TFC_l}$$

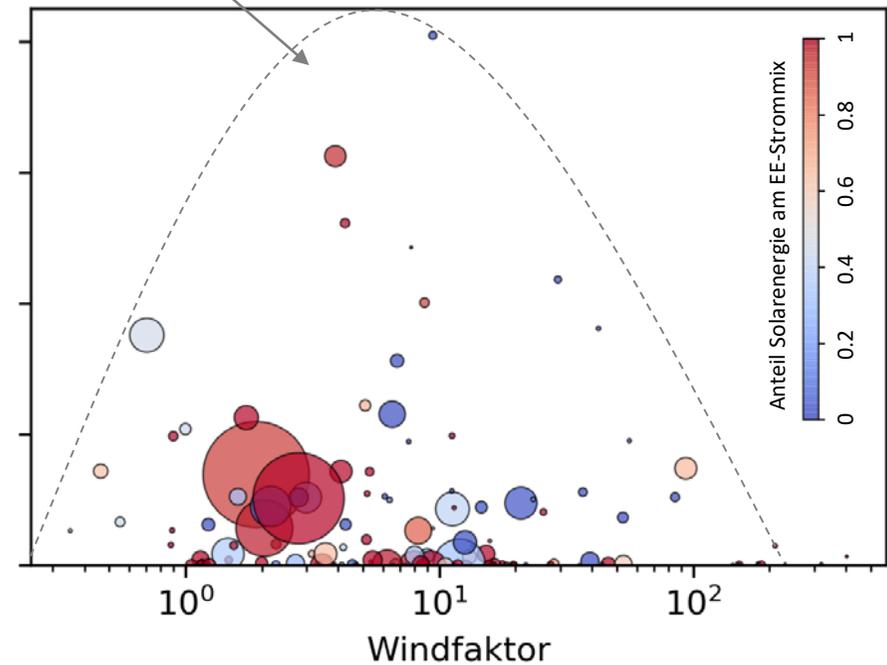
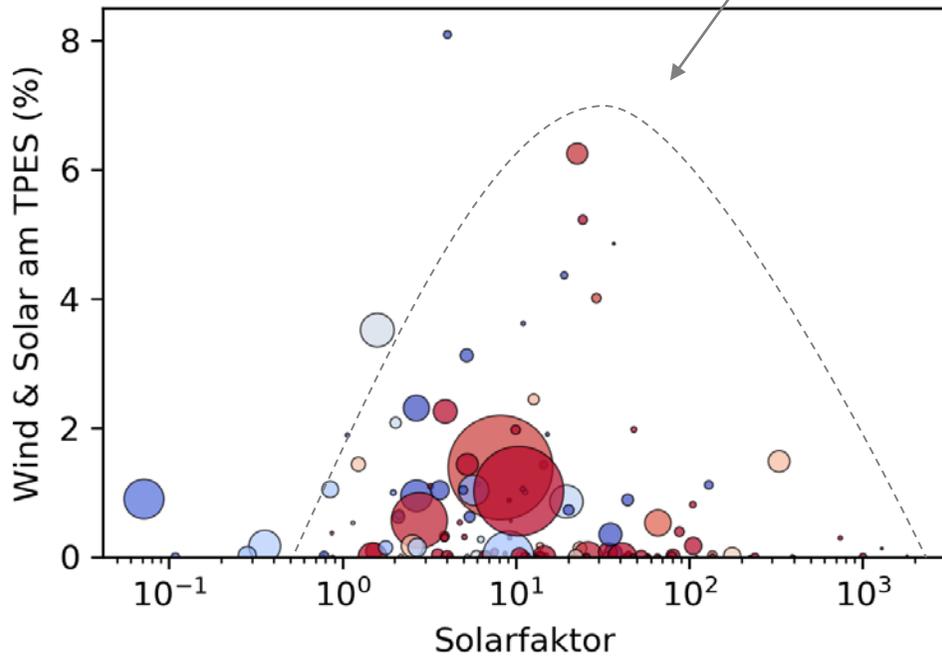
$$\forall k, d \in \sum_{d,k}^{D,K} \langle \text{Potential}_{d,k,l} * VLh_{d,k,l} \rangle \leq TFC_l$$



Feature Auswahl: Definition der Beschreibenden Parameter

Welche Parameter beeinflussen die Nutzung erneuerbarer Energien heute? Welche in der Zukunft?

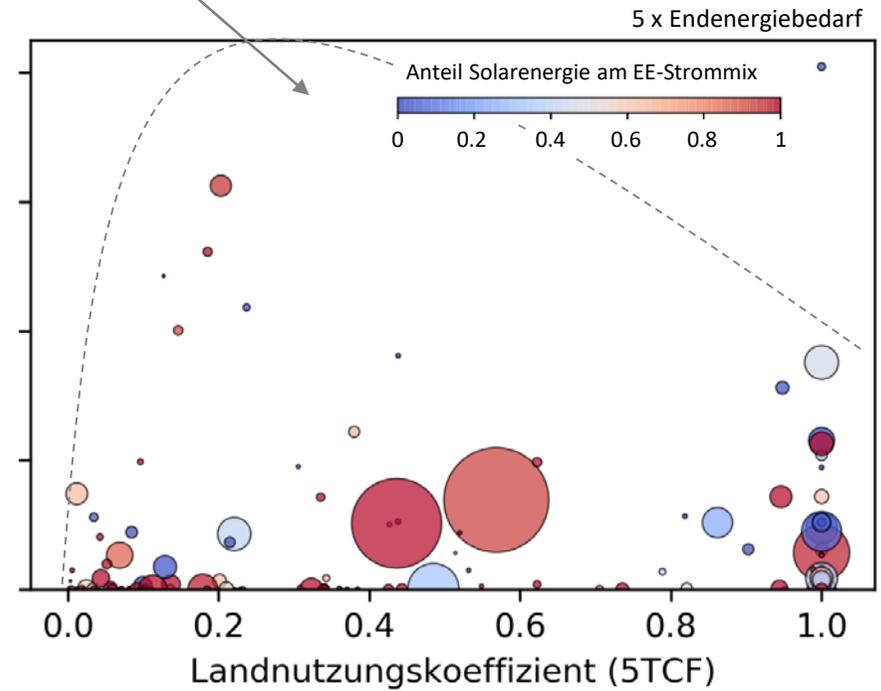
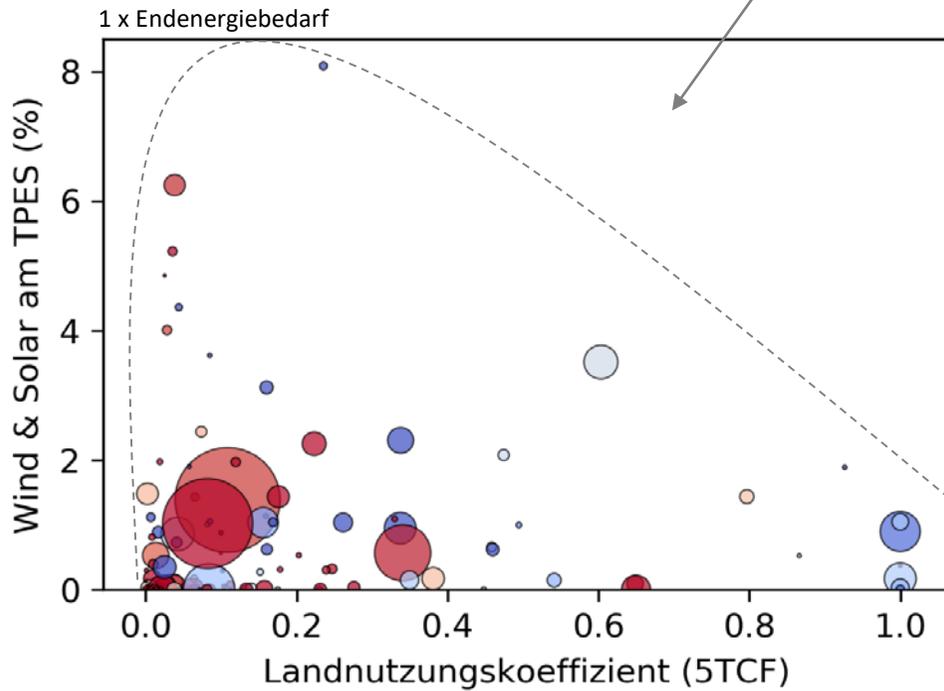
Heute noch kaum Einfluss –
Zukünftig aber ev. limitierender Faktor



Feature Auswahl: Definition der Beschreibenden Parameter

Welche Parameter beeinflussen die Nutzung erneuerbarer Energien heute? Welche in der Zukunft?

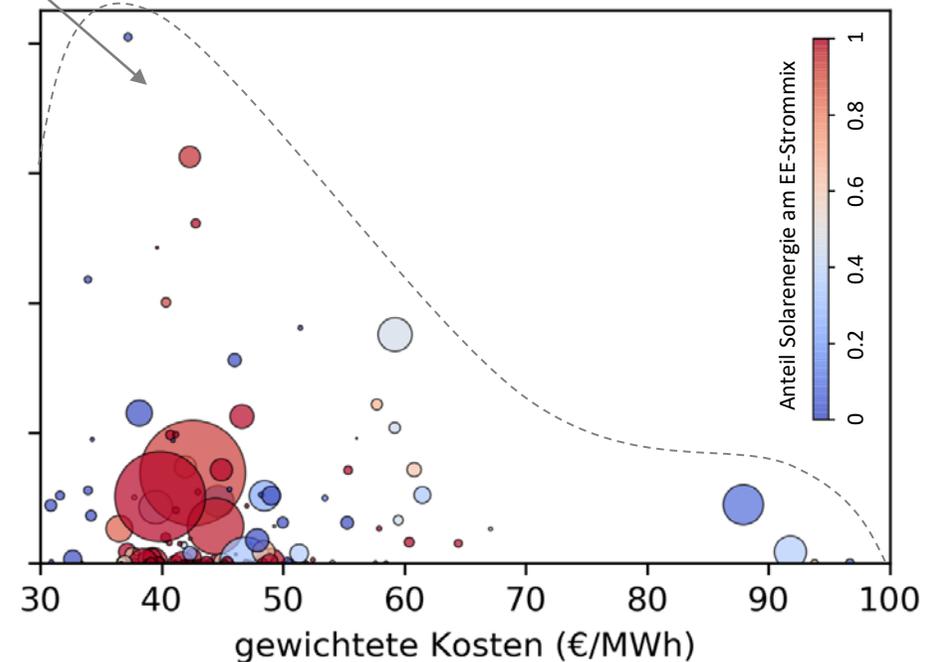
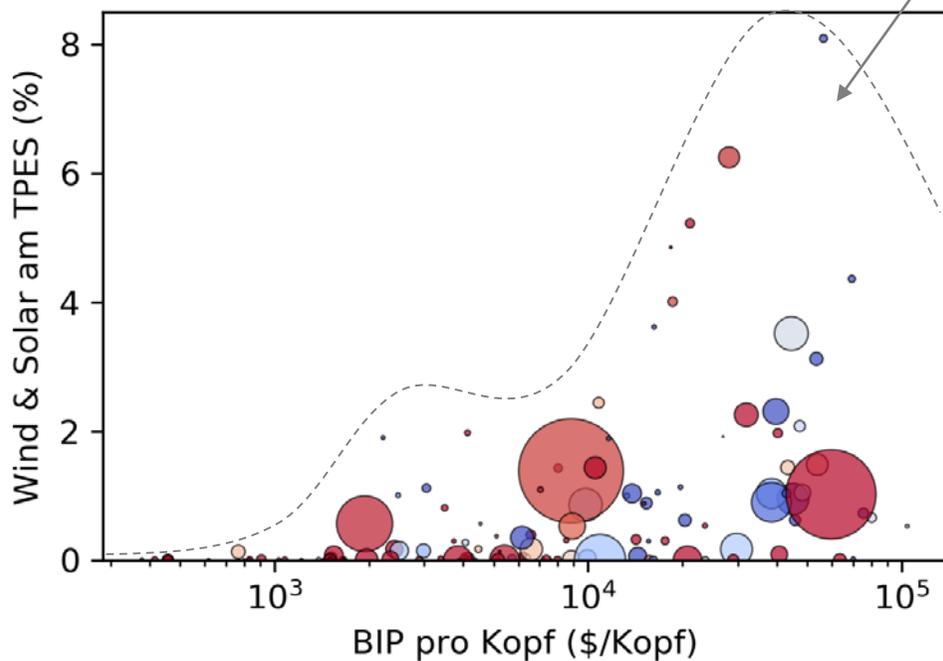
Verfügbare Landfläche ist heute schon relevant –
Zukünftig noch relevanter?



Feature Auswahl: Definition der Beschreibenden Parameter

Welche Parameter beeinflussen die Nutzung erneuerbarer Energien heute? Welche in der Zukunft?

BIP & Kosten sind heute schon relevant –
Bleibt das so?



Feature Auswahl: Definition der Beschreibenden Parameter

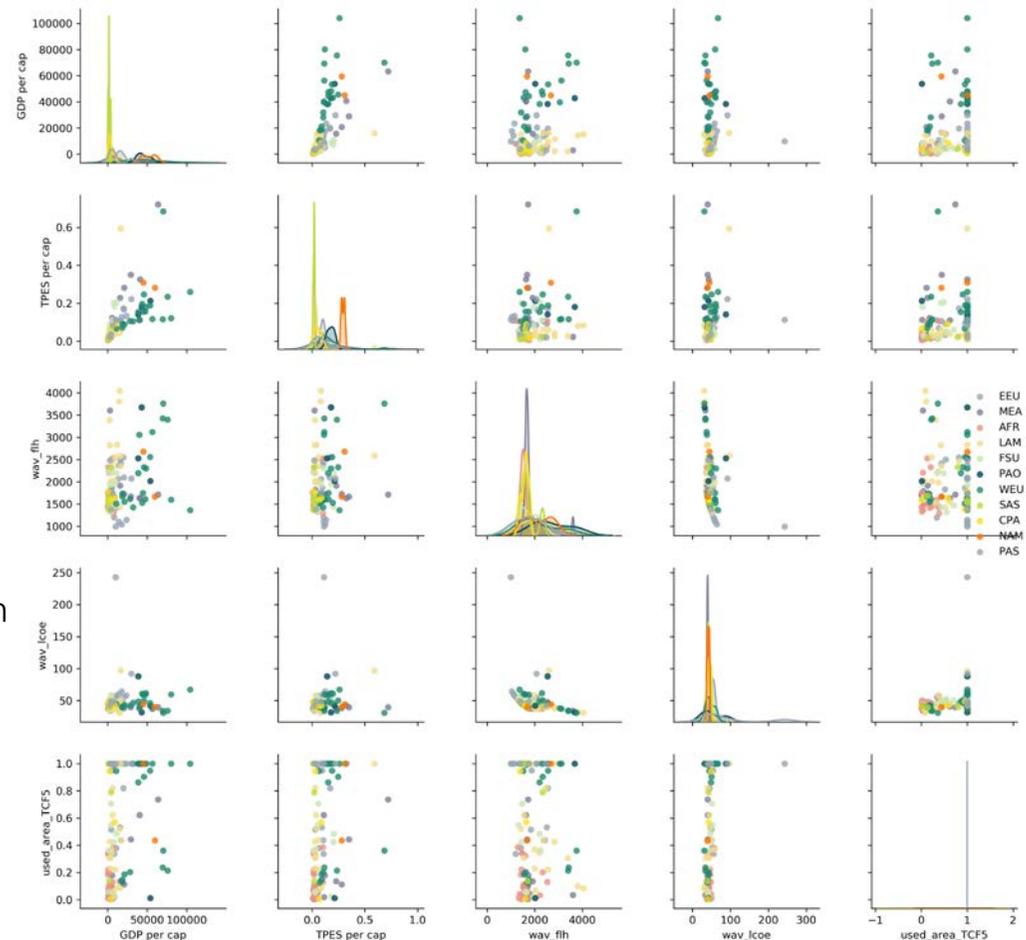
Um die Nachvollziehbarkeit der Clusterergebnisse zu gewährleisten, wurden die, über 200 pro Land und Jahr zu Verfügung stehenden Parametern auf ein Set von fünf Parametern reduziert.

Feature Selektion:

- Erhöht Interpretierbarkeit
- Vermeidung von „Overfitting“
- Berechenbarkeit sicherstellen

Ausgewählte Features:

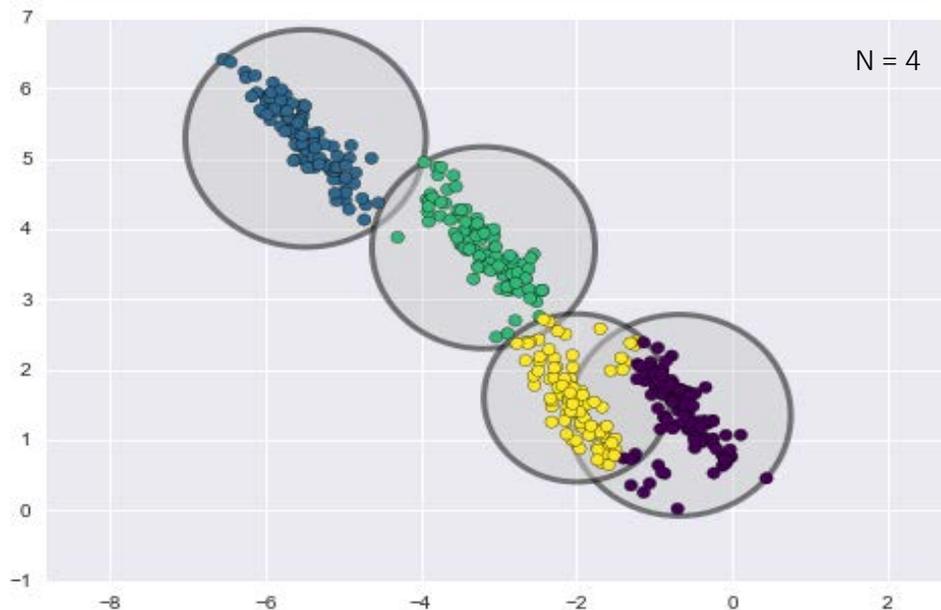
- BIP pro Kopf
- Primärenergiebedarf pro Kopf
- Gewichtete gemittelte Stromgestehungskosten
- Gewichtete gemittelte Volllaststunden
- Landnutzung zur fünffachen Deckung des Endenergiebedarfs (Region)



Clusteranalyse: Vergleich der Clusteralgorithmen

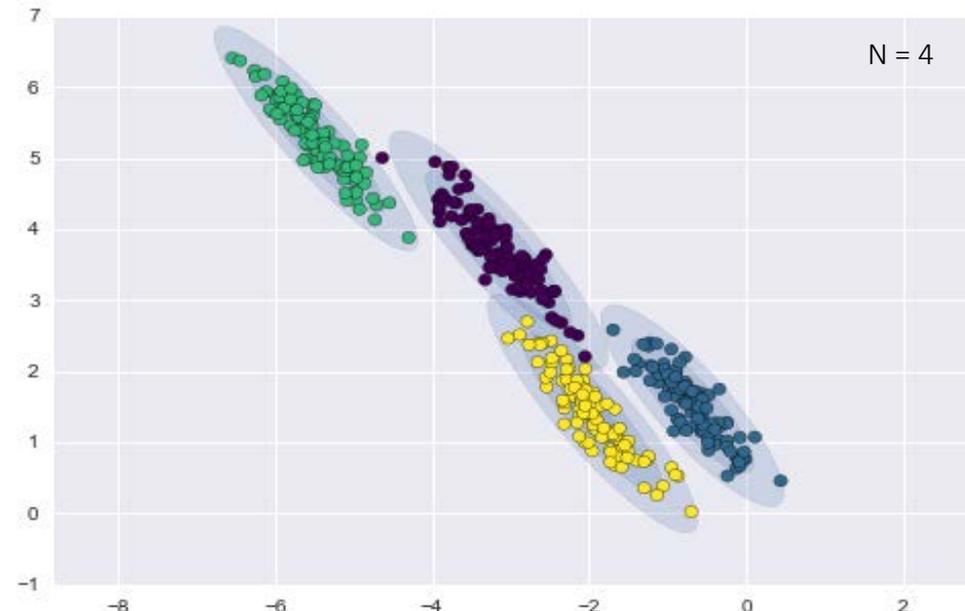
k-Means - deterministisches Modell

- Clusterzentrum ist arithmetische Mittel der Clusterpunkte
- Punkt ist immer näher an eigenem als an anderem Clusterzentrum



GaussianMixture - probabilistisches Modell

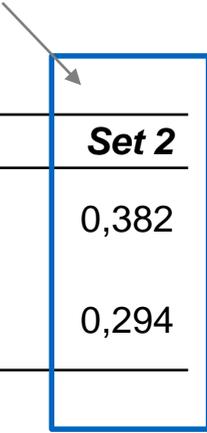
- Punkte bestehen aus Gaußschen Verteilungen unbekannter Parameter
- Zugehörigkeitswahrscheinlichkeit zu Cluster bestimmt Zuordnung



Unüberwachtes lernen - Bewertungsmetrik

Im unüberwachten lernen gibt es kein „richtig“ und „falsch“. Die Bewertungsmetrik, der Silhouettenkoeffizient, bezieht sich daher ausschließlich auf Eindeutigkeit der Zuordnung.

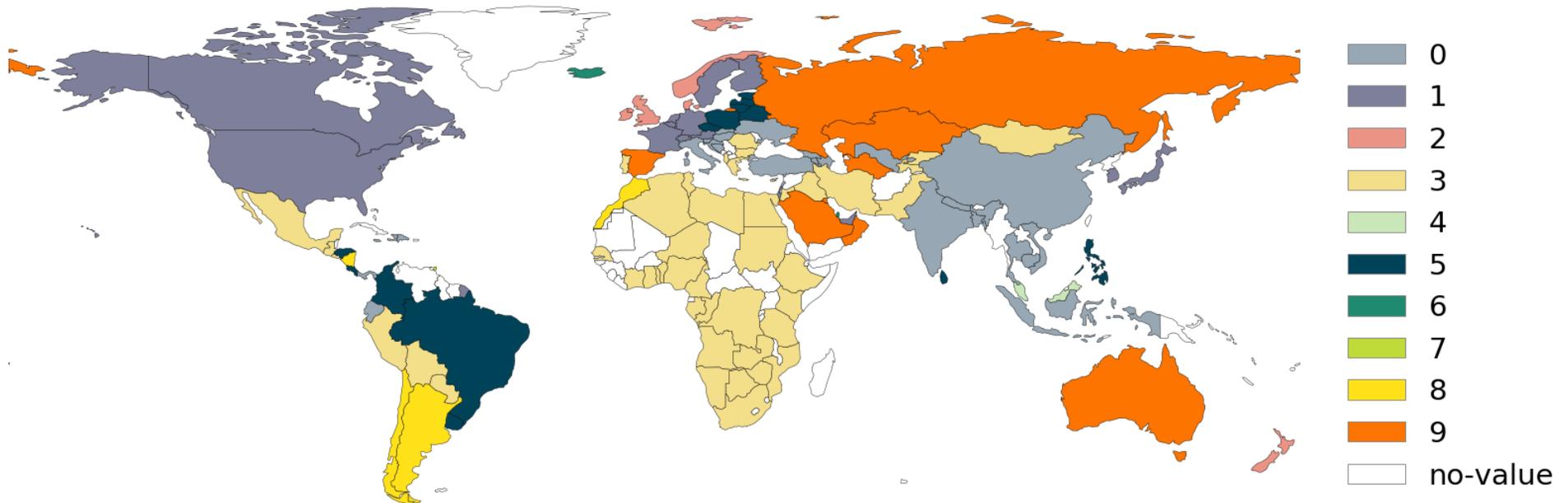
Region ist der dominante Faktor.



Algorithmus	Set 1	Set 2
<i>KMeans</i>	0,457	0,382
<i>GaussianMixture</i>	0,402	0,294

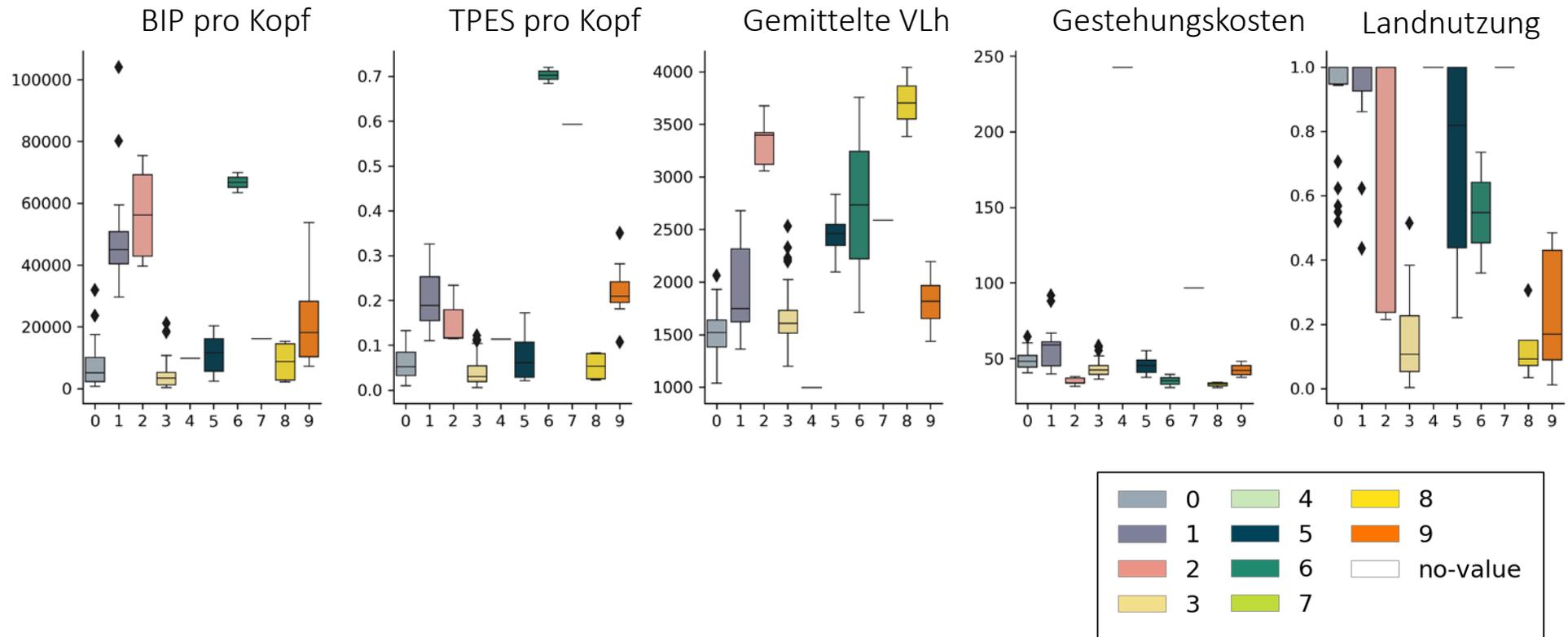
Clusterung: Bewertung der Clusterung mittels *Kmeans* und Set 2

Einige intuitive Cluster – aber auch viele „Ausreißer“ und „alles andere“ Cluster.



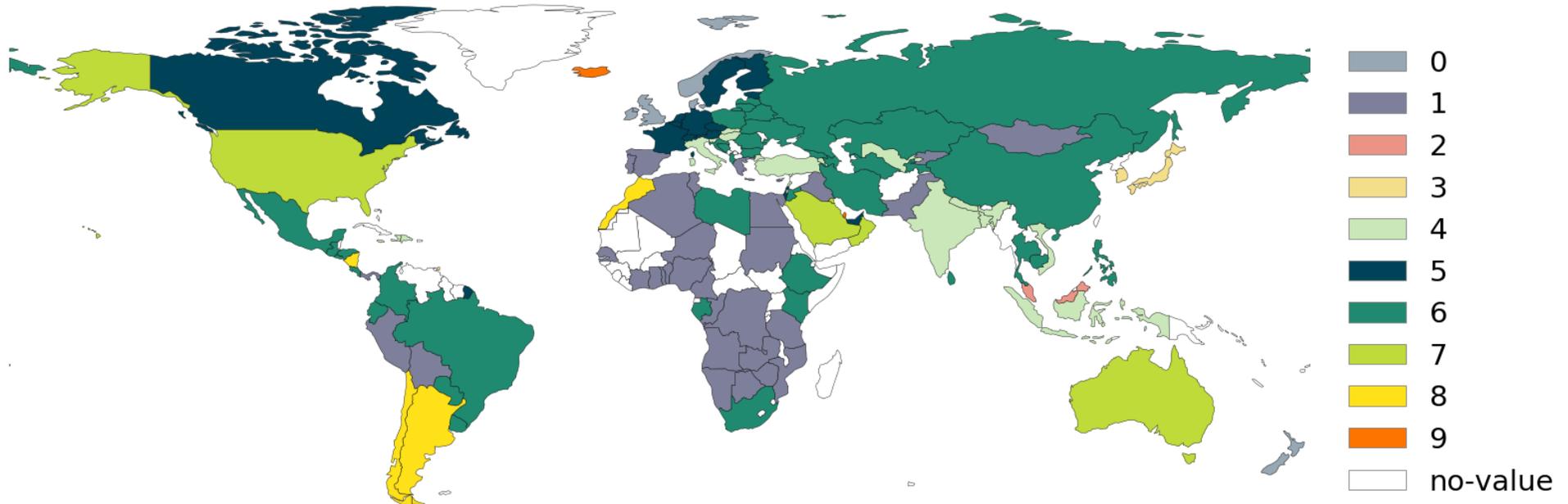
Clustering: Bewertung der Clustering mittels *Kmeans* und *Set 2*

Einige intuitive Cluster – aber auch viele „Ausreißer“ und „alles andere“ Cluster.



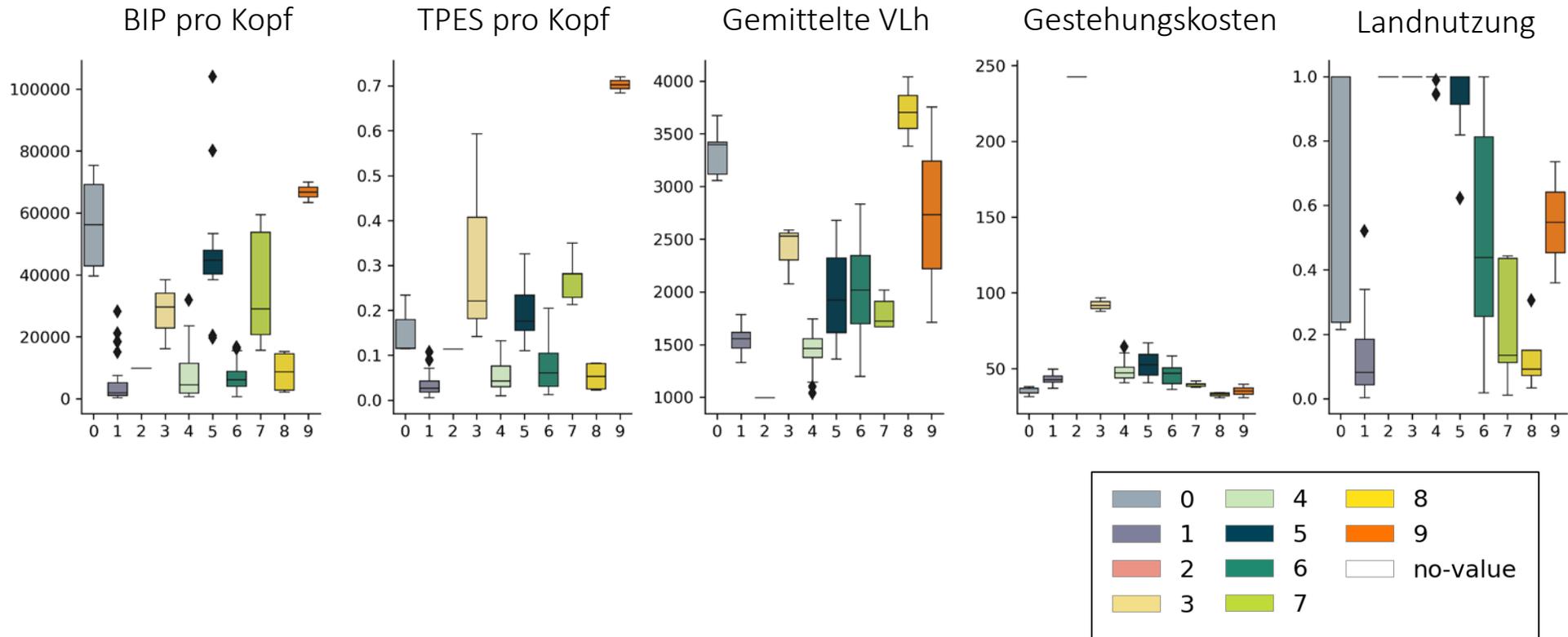
Clustering: Bewertung der Clustering mittels *GaussianMixture* und Set 2

Einige intuitive Cluster – aber auch viele „Ausreißer“ und „alles andere“ Cluster.



Clustering: Bewertung der Clustering mittels *GaussianMixture* und *Set 2*

Einige intuitive Cluster – aber auch viele „Ausreißer“ und „alles andere“ Cluster.



Zusammenfassung, Ausblick und Fazit

Herausforderung: kluge Regionenbildung für globale Energiesystemmodelle

Traditioneller Ansatz: Aggregation auf Expertenbasis (geographische, politisch-historische Ähnlichkeit, ect)

Neuer Ansatz: machine learning mit Fokus auf die Wind- und Solarenergiepotenziale

Ergebnis: schwache strukturelle Abgrenzung (qualitative Informationen fehlen)

Fazit: die Clusteranalyse kann

- den herkömmlichen Ansatz nicht ersetzen
- unterstützendes Tool für die traditionelle Vorgehensweise werden
- der prototypische Bildung von Regionen dienen
- ein erster Startpunkt sein einer Expertenbasierten Analyse sein
- helfen um mögliche Ausreißer zu identifizieren

[Fragen oder Anregungen?](#)

Clara Orthofer

clara.orthofer@tum.de

Thomas Zipperle

thomas.zipperle@tum.de