Mapping Gas Permeability of Sustainable Packaging Materials to Link Food Barrier Needs by Clustering algorithms

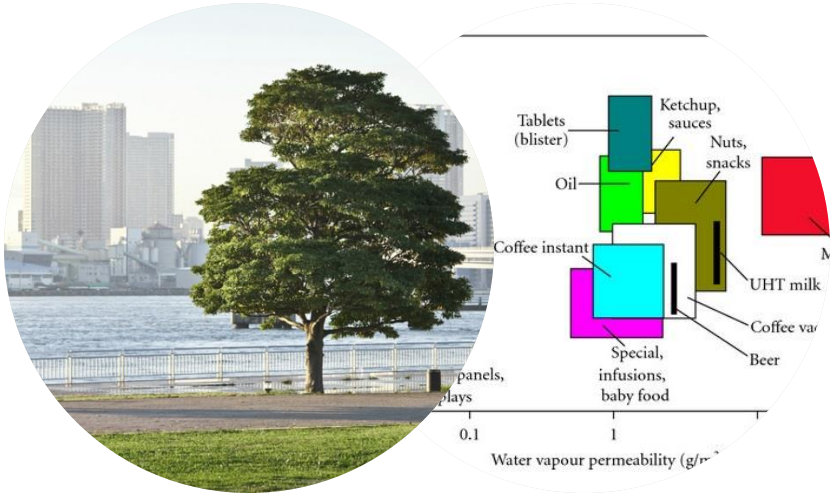
Final Colloquial

Chair Group : Food Quality & Design

MSc Student : Windy Yeh

Supervisor : Deniz Turan Kunter

Date : Feb 17th



WAGENINGEN
UNIVERSITY & RESEARCH

Background

Food products package in Europe

Others

40%

plastics

95%

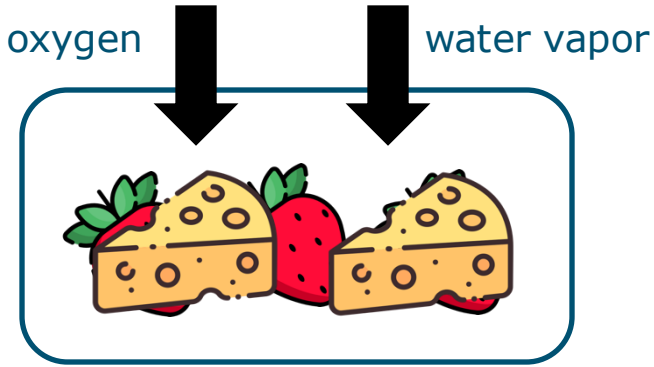
petroleum-based

Sustainable Materials
PLA, PHB, chitosan

WAGENINGEN
UNIVERSITY & RESEARCH

Two problems in shifting to sustainable packaging materials

- Inferior gas barrier properties
- Different food barrier needs



(Guillard et al., 2018)

What is the knowledge gap?

Sustainable Materials

Chitosan

Cellulose

Paperboard

Glass

polylactic acid

Polyethylene + corn starch

Chitosan - Cellulose

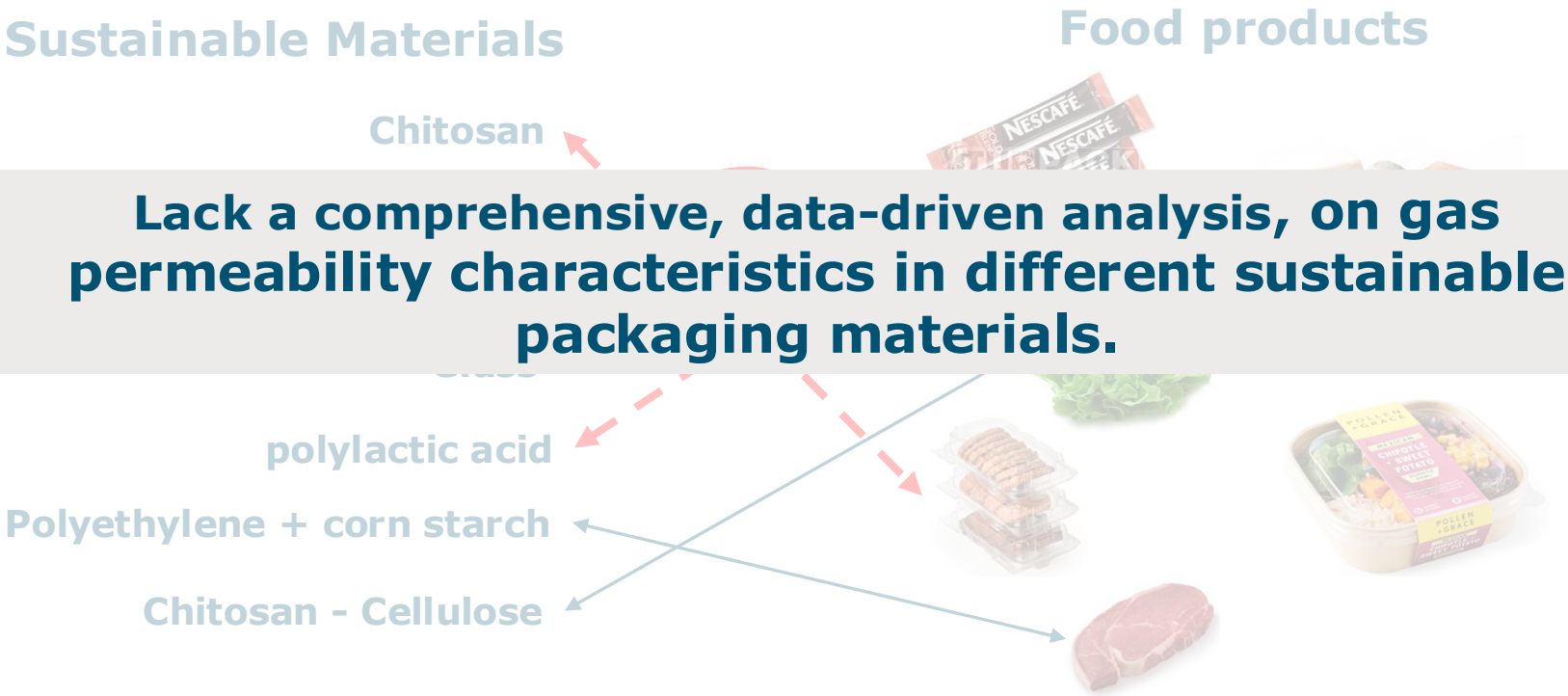
?

Food products

WAGENINGEN
UNIVERSITY & RESEARCH

(Makino & Hirata, 1997) (Strantz & Zottola, 1992)

What is the knowledge gap?



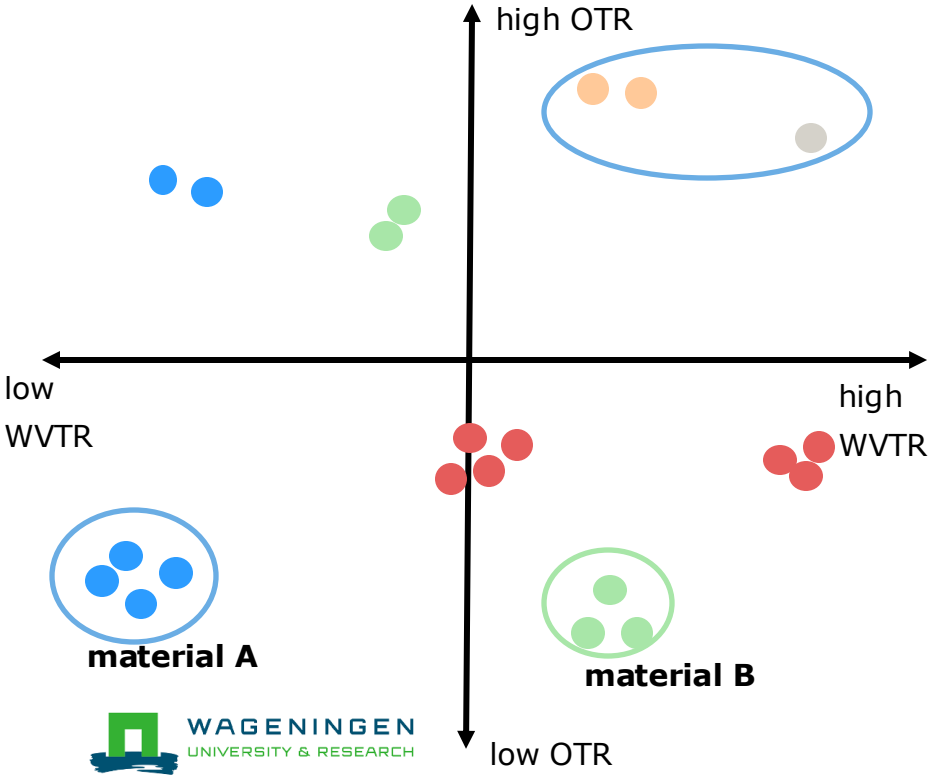
Lack a comprehensive, data-driven analysis, on gas permeability characteristics in different sustainable packaging materials.

(Makino & Hirata, 1997) (Strantz & Zottola, 1992)

What is the aim of the project

OTR: oxygen transmission rate
WVTR : water vapor transmission rate

Sustainable Materials



Food products

What is the aim of the project

Sustainable Materials

high OTR

low WVT

low OTR

material A

material B

Food products

high OTR

low OTR

Research Questions

1. Which **clustering algorithms** most accurately classify sustainable packaging materials based on their gas permeability?

2. Which **sustainable materials** are best suited for packaging **different food categories**?

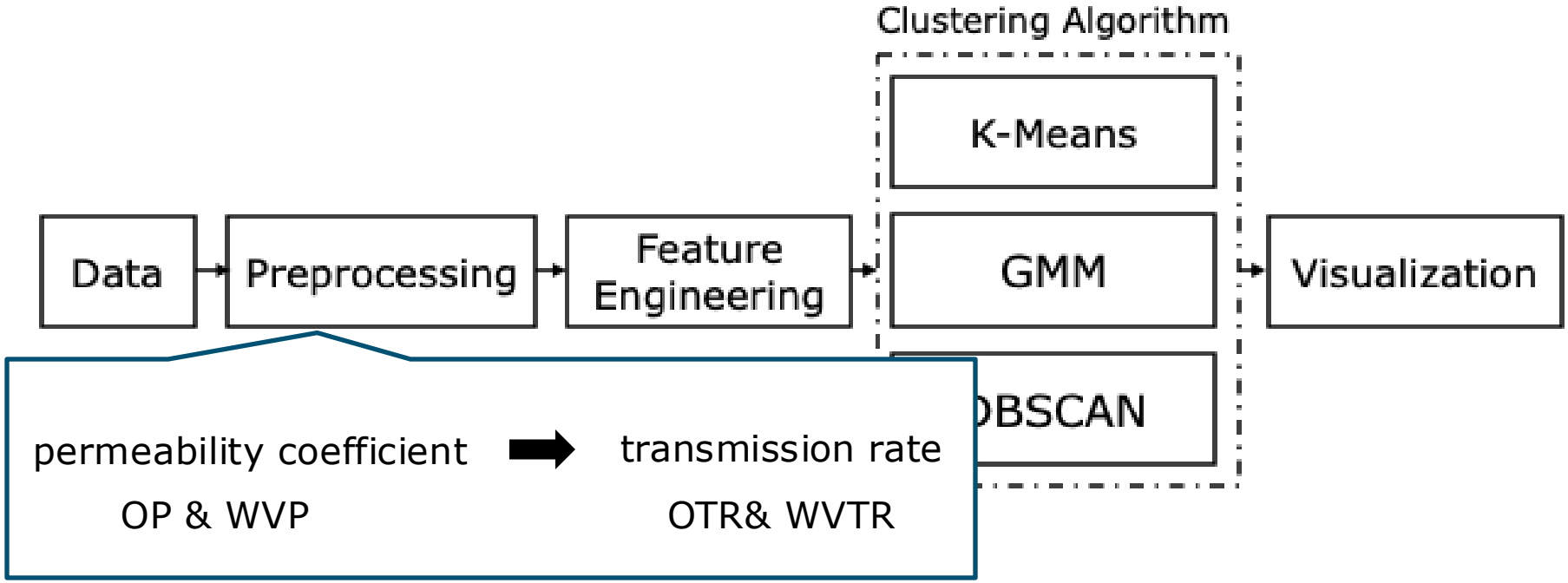
WAGENINGEN UNIVERSITY & RESEARCH

Materials

Variable	Explanation	Values
Doc	the article title	Barrier and surface properties of chitosan-coated greaseproof paper
DOI	Digital Object Identifier	https://doi.org/10.1016/j.carbpol.2006.02.005
Target	generic concept represented	Permeability
Type	the ontology concept category, symbolic, quantitative or addimensional	QUANTITY
Original_Value	a list of annotated tokens for symbolic data, two lists of annotated tokens for quantitative data	(['1.6'], ['cm', '^', '3', 'mm', '/', '(', 'm', '^', '2', '24', 'h', 'atm', ')'])
Attached_Value	the list of annotated tokens to disambiguate a measure unit when necessary for quantitative data. None for symbolic data.	['OP']
Annotator	annotator id	1

(Lentschat et al., 2021)

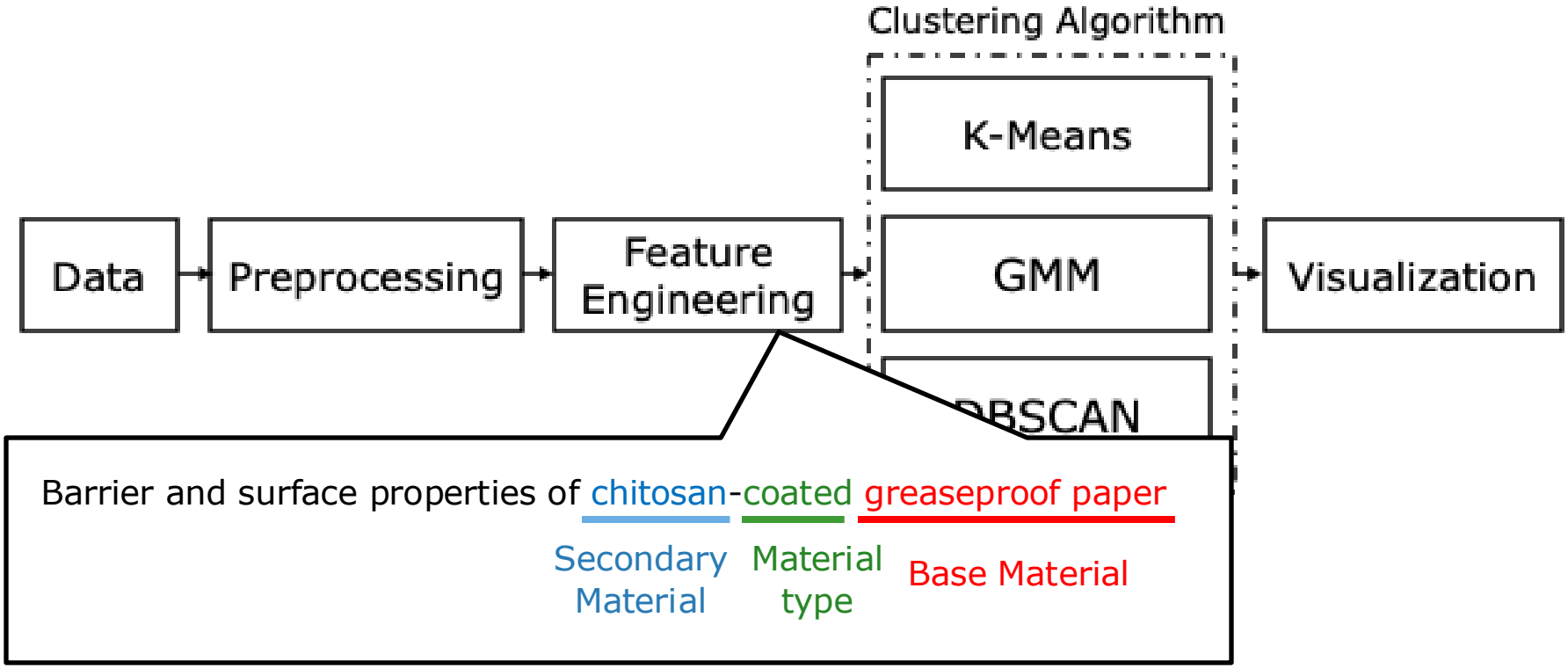
Methods



$$PC = \frac{TR \times thickness}{Pressure}$$

(Siracusa, 2012)

Methods



Natural Language Processing

Methods

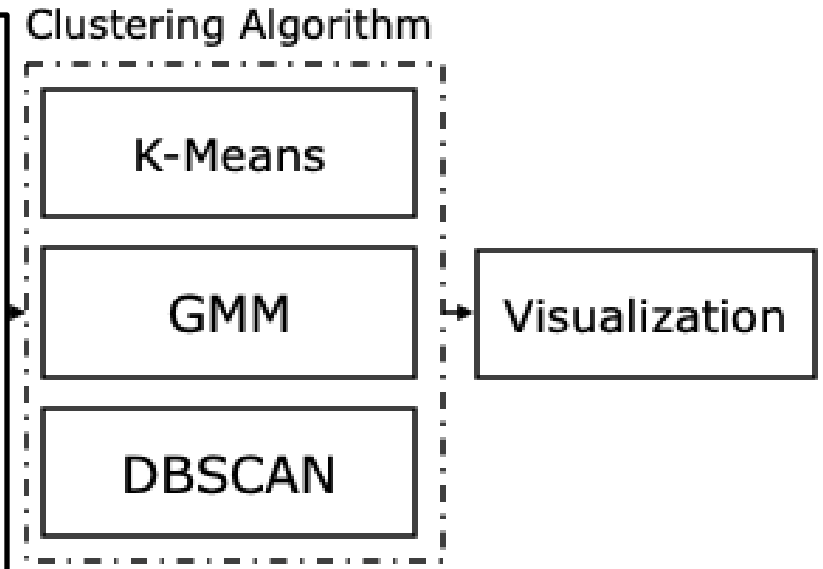
GMM : Gaussian Mixture Model

DBSCAN : Density-Based Spatial Clustering of Applications with Noise

Partition-based algorithm , basic and widely used
e.g. Food Risk Prediction

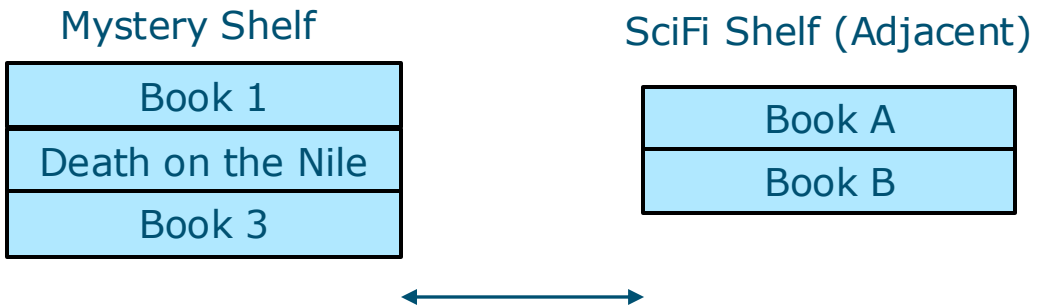
Model-based algorithm flexible for arbitrary shapes

Density-based algorithm, no perquisites number needed
e.g. Nutritional Classification



Which model most accurately classify sustainable packaging materials based on their gas permeability?

Silhouette Score & Davies-Bouldin Index



compact & large distance
-> **low Davies-Bouldin Index**

a : How difference is Death on the Nile to Book 1 & Book 3?
b : How difference is Death on the Nile to books on SciFi shelf?

b high and a low -> **high Silhouette Score**

Which model most accurately classify sustainable packaging materials based on their gas permeability?

Metric	Algorithm	Before optimization	After optimization
Silhouette Score	K-Means	0.583	0.680
	DBSCAN	0.558	**0.900**
	GMM	0.539	0.705
Davies- Bouldin Index	K-Means	0.673	0.433
	DBSCAN	0.520	**0.388**
	GMM	0.772	0.428

OTR 23℃, 50%RH

WVTR 25℃, 50%RH

DBSCAN can most accurately classify sustainable packaging materials based on their gas permeability?

Metric	Algorithm	Before optimization	After optimization
Silhouette Score	K-Means	0.583	0.680
	DBSCAN	0.558	**0.900**
	GMM	0.539	0.705
Davies- Bouldin Index	K-Means	0.673	0.433
	DBSCAN	0.520	**0.388**
	GMM	0.772	0.428

OTR 23℃, 50%RH

WVTR 25℃, 50%RH

WAGENINGEN
UNIVERSITY & RESEARCH

```
def print_cluster_sizes(labels, algorithm_name):
    cluster_counts = pd.Series(labels).value_counts()
    print(f"\n{algorithm_name} Cluster Sizes:")
    for cluster, count in cluster_counts.items():
        cluster_name = "Noise" if cluster == -1 else f"Cluster {cluster}"
        print(f"{cluster_name}: {count} data points")

# K-Means
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_labels = kmeans.fit_predict(data)
kmeans_silhouette = silhouette_score(data, kmeans_labels)
kmeans_davies_bouldin = davies_bouldin_score(data, kmeans_labels)
print(f"K-Means: Silhouette Score = {kmeans_silhouette}, Davies-Bouldin Index = {kmeans_davies_bouldin}")
print_cluster_sizes(kmeans_labels, "K-Means")

# DBSCAN
dbscan = DBSCAN(eps=eps, min_samples=min_samples)
dbscan_labels = dbscan.fit_predict(data)
# Exclude noise points (-1) for silhouette score
valid_dbscan = dbscan_labels != -1
dbscan_silhouette = silhouette_score(data[valid_dbscan], dbscan_labels[valid_dbscan]) if valid_dbscan.any() else -1
dbscan_davies_bouldin = davies_bouldin_score(data[valid_dbscan], dbscan_labels[valid_dbscan]) if valid_dbscan.any() else float('inf')
print(f"\nDBSCAN: Silhouette Score = {dbscan_silhouette}, Davies-Bouldin Index = {dbscan_davies_bouldin}")
print_cluster_sizes(dbscan_labels, "DBSCAN")

# Gaussian Mixture Model
gmm = GaussianMixture(n_components=n_clusters, random_state=42)
gmm_labels = gmm.fit_predict(data)
gmm_silhouette = silhouette_score(data, gmm_labels)
gmm_davies_bouldin = davies_bouldin_score(data, gmm_labels)
print(f"\nGMM: Silhouette Score = {gmm_silhouette}, Davies-Bouldin Index = {gmm_davies_bouldin}")
```

Discussion 1

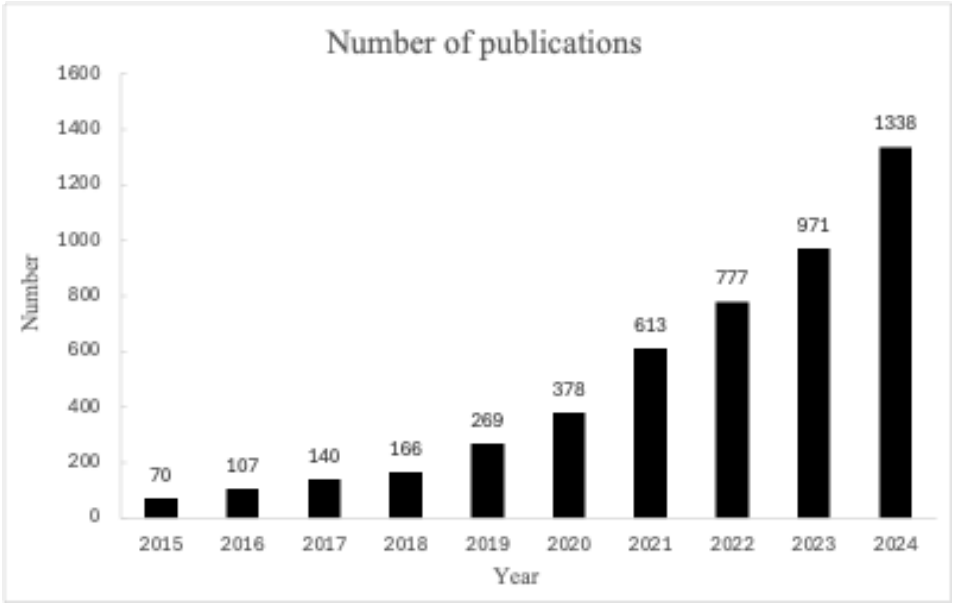
How clustering models were used in gas permeability?

- No clustering model was used in the food packaging material gas permeability
- Predicting model by machine learning (Phan et al., 2024)
- Gas permeability is influenced by multiple factors (De Almeida & Raquel, 2018; Freeman, 1999)

WAGENINGEN
UNIVERSITY & RESEARCH

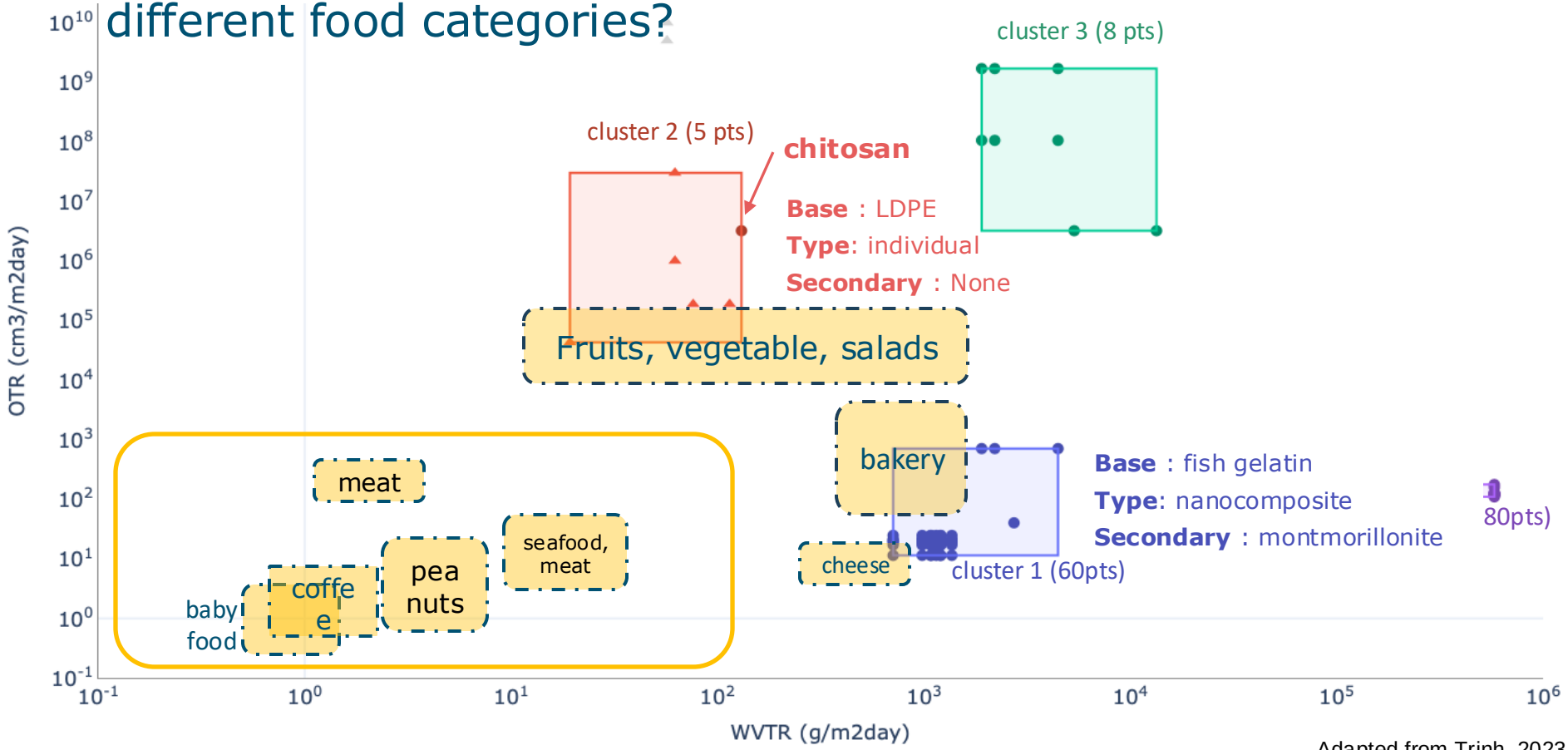
Expanded data volume and **high-quality data** are required

- The data spans from 2000 to 2016
- Mislabel data
- Standardized and transparent reporting of measurement condition



keyword: sustainable packaging material

Which sustainable materials are best suited for packaging different food categories?



Adapted from Trinh, 2023

```
def print_cluster_sizes(labels, algorithm_name):
    cluster_counts = pd.Series(labels).value_counts()
    print(f"\n{algorithm_name} Cluster Sizes:")
    for cluster, count in cluster_counts.items():
        cluster_name = "Noise" if cluster == -1 else f"Cluster {cluster}"
        print(f"{cluster_name}: {count} data points")

# K-Means
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_labels = kmeans.fit_predict(data)
kmeans_silhouette = silhouette_score(data, kmeans_labels)
kmeans_davies_bouldin = davies_bouldin_score(data, kmeans_labels)
print(f"K-Means: Silhouette Score = {kmeans_silhouette}, Davies-Bouldin Index = {kmeans_davies_bouldin}")
print_cluster_sizes(kmeans_labels, "K-Means")

# DBSCAN
dbscan = DBSCAN(eps=eps, min_samples=min_samples)
dbscan_labels = dbscan.fit_predict(data)
# Exclude noise points (-1) for silhouette score
valid_dbscan = dbscan_labels != -1
dbscan_silhouette = silhouette_score(data[valid_dbscan], dbscan_labels[valid_dbscan]) if valid_dbscan.any() else -1
dbscan_davies_bouldin = davies_bouldin_score(data[valid_dbscan], dbscan_labels[valid_dbscan]) if valid_dbscan.any() else float('inf')
print(f"\nDBSCAN: Silhouette Score = {dbscan_silhouette}, Davies-Bouldin Index = {dbscan_davies_bouldin}")
print_cluster_sizes(dbscan_labels, "DBSCAN")

# Gaussian Mixture Model
gmm = GaussianMixture(n_components=n_clusters, random_state=42)
gmm_labels = gmm.fit_predict(data)
gmm_silhouette = silhouette_score(data, gmm_labels)
gmm_davies_bouldin = davies_bouldin_score(data, gmm_labels)
print(f"\nGMM: Silhouette Score = {gmm_silhouette}, Davies-Bouldin Index = {gmm_davies_bouldin}")
```

Discussion 2

Strategies to improve the barrier properties

- Coating
    - e.g. PLA-coated paperboard with an AlOx layer (Peelman et al., 2013)
- Incorporation of nanoparticles
    - e.g. PLA combined with montmorillonite-layered silicate (Arora & Padua, 2010)

Incorporation of nanoparticles

Cluster	1 (n = 60)	2 (n = 6)	3 (n = 8)	4 (n = 180)
OTR (cm^3/m^2day)	22.74 ± 2.30	730000.02±10.60	120000000.00±13.46	126.35±1.13
WVTR (g/m^2day)	1146.83±1.35	66.27±1.97	3548.47±1.98	248416.31±1.40
Dominant Base Material	fish gelatin	LDPE	polylactic acid	carrot puree
Dominant Type	nanocomposite	individual	individual	blend
Dominant Secondary Material	montmorillonite	None	None	CMC

Conclusion

- DBSCAN, is the most effective algorithm for grouping sustainable packaging materials based on gas permeability, with highest the Silhouette Score (0.900) and lowest Davies-Bouldin Index (0.388).

- Most sustainable packaging materials have high gas permeability, whereas many food applications require high gas barrier properties.
    - Incorporation of nanoparticle

Recommendation

- Limitation
  - Small dataset
  - Inconsistencies and missing data in key parameters

- Recommendation
  - Expand datasets and improve data quality
  - Including more gas permeability data or additional material properties

Thanks for your listening!

Model Hyperparameters

- K-Means
  - k = 9
- GMM
  - k = 9
- DBSCAN
  - MinPt = 4, epsilon = 1.2

Model Evaluation

- Silhouette Score

$$s = \frac{b - a}{\max(a, b)}$$

where a represents the average distance between a sample and all other points in the same cluster, while b is the average distance between a sample and all other points in the next nearest cluster.

- Davies-Bouldin Index

$$db = \frac{1}{k}\sum_{i=1}^{k} \max\left(\frac{\Delta X_i + \Delta X_j}{\delta(X_i, X_j)}\right)$$

where ΔXi is the distance within the cluster and $\delta(Xi, X\ j)$ is the distance between the cluster i and cluster j

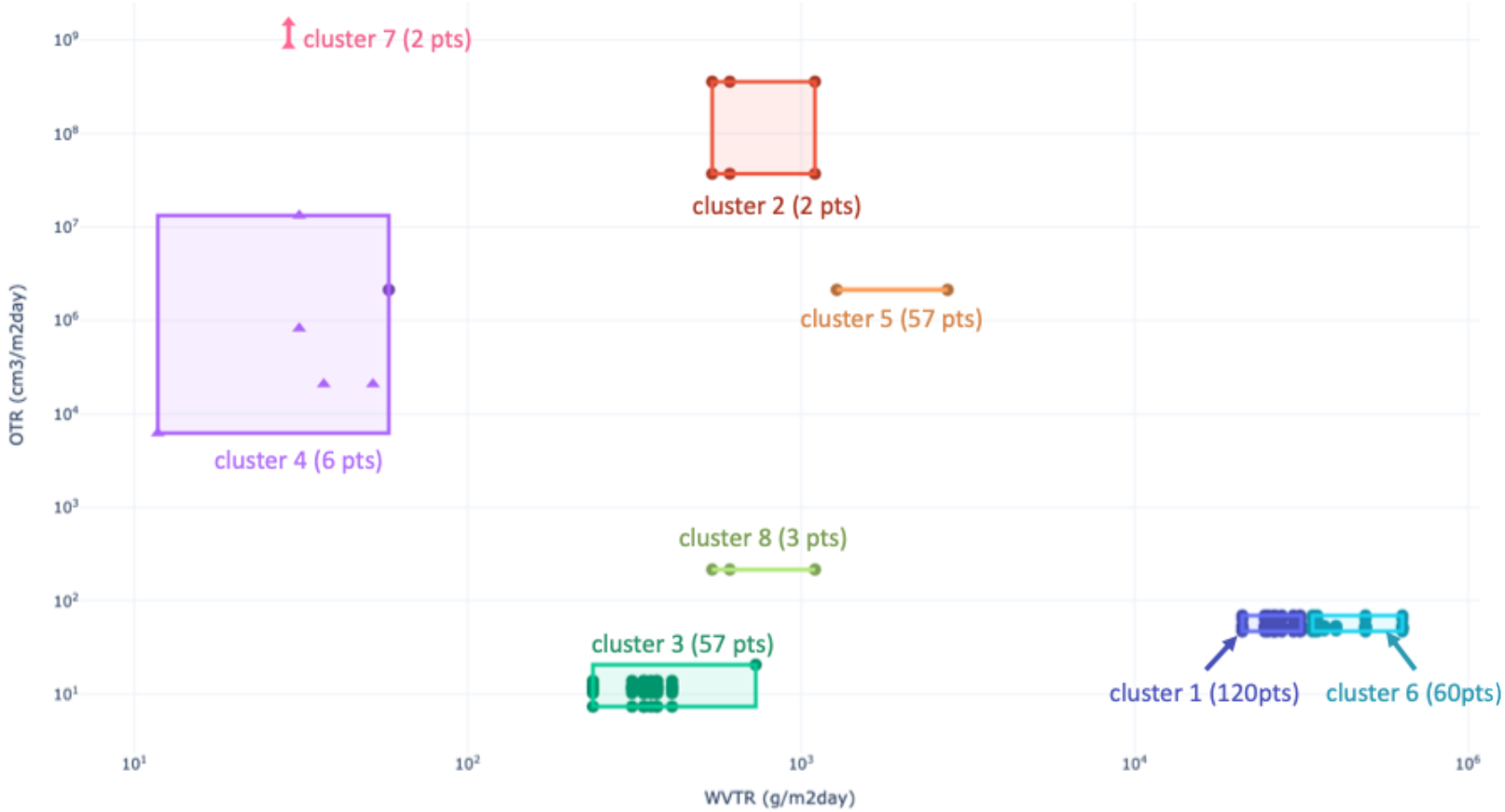
Feature Engineering

Features

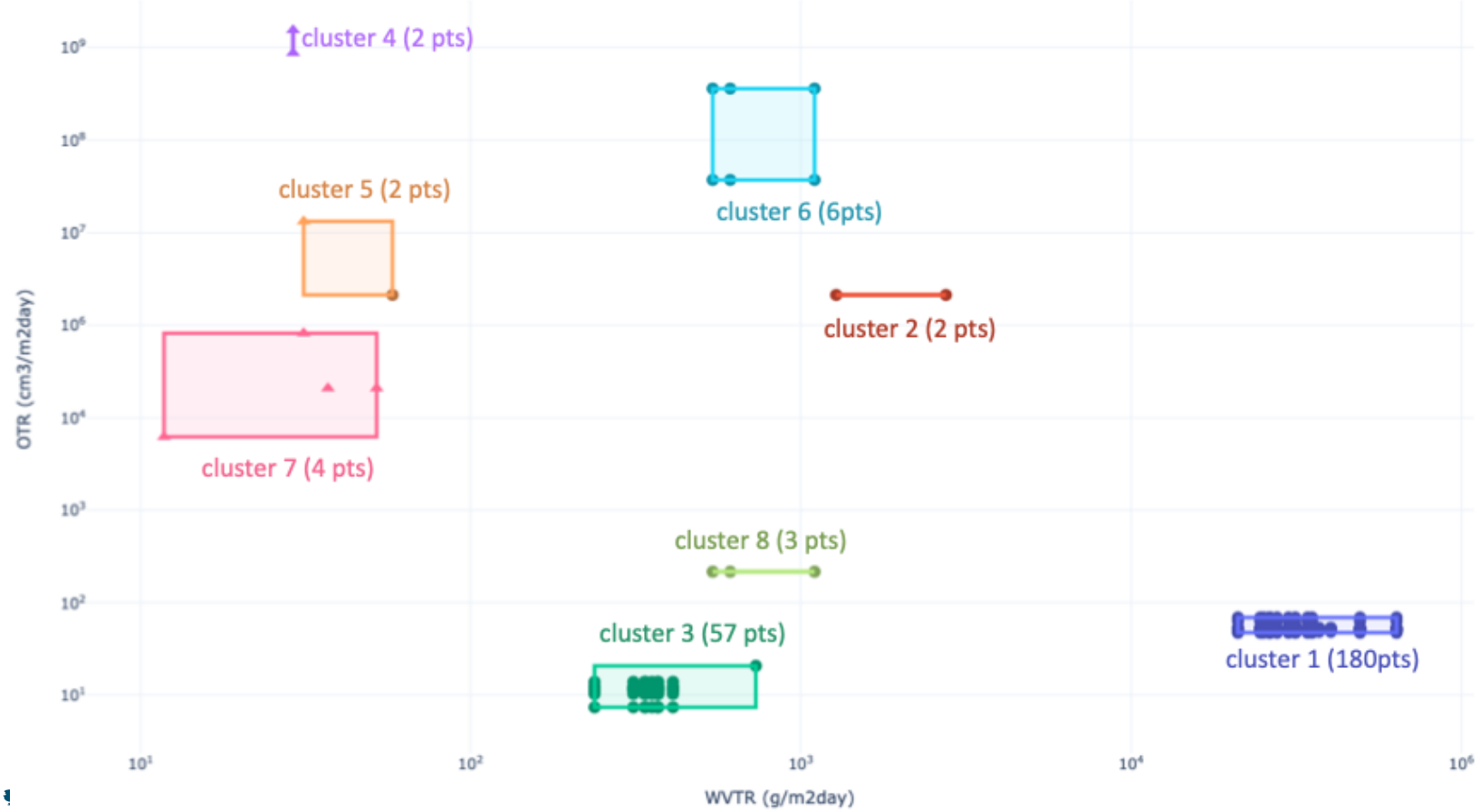
Materials	Based Material	Type	Secondary Material	Composition	OP	OTR	WVP	WVTR	temp	RH
chitosan-coated greaseproof paper	greaseproof paper	coated	chitosan	[2.4]	[0.01, 2]	[0.001, 0.2]	nan	nan	25	95
poly (lactic acid) nanocomposites	polylactic acid	nanocomposite	montmorillonite	[0.15]	nan	nan	$1.160 * 10^{-10}$	$1.160 * 10^{-8}$	38	52
Carrot puree films	Carrot puree	individual	nan	[0.4,0.6]	[20.20, 20.34]	[4.80, 4.84]	nan	nan	25	83

WAGENINGEN
UNIVERSITY & RESEARCH

K-Means Result



GMM Result


cluster 4 (2 pts)
cluster 6 (6pts)
cluster 5 (2 pts)
cluster 2 (2 pts)
cluster 7 (4 pts)
cluster 8 (3 pts)
cluster 3 (57 pts)
cluster 1 (180pts)
OTR (cm3/m2day)
WVTR (g/m2day)