KSE | Центр досліджень продовольства та землекористування

Policy brief

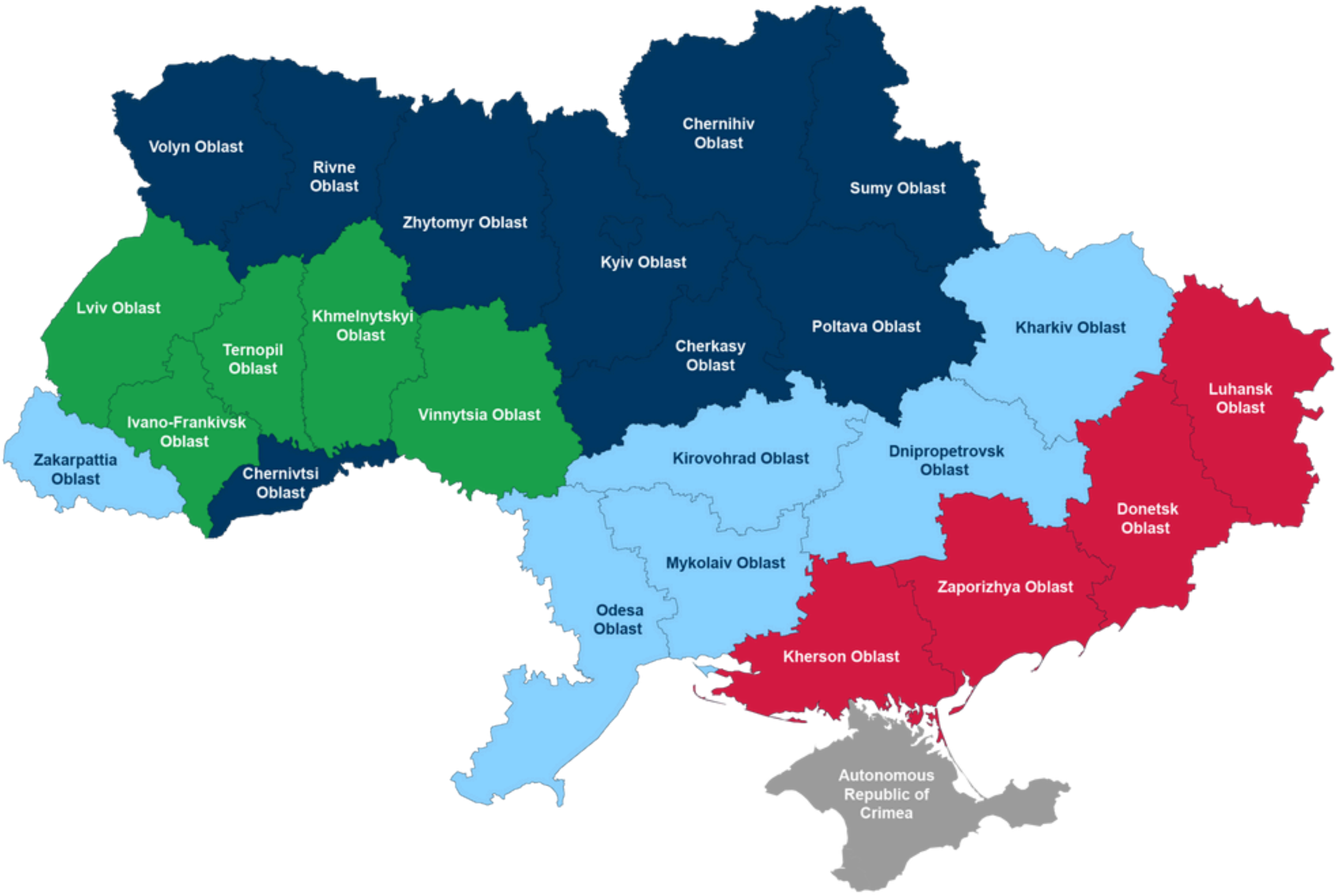
Regionalization of agricultural production in Ukraine

P. Pyvovar,
H.Stolnikovych

Summary

This brief presents the regionalization in Ukraine's agricultural sector in the AGMEMOD model. Main motivation for the analysis is to address the structural differences in patterns of agricultural production in different Ukrainian regions, thus improving the accuracy and precision of the forecasts. K-means clustering method was applied to the dataset of regional averages of yield values. The proposed regionalization is presented on the Figure 1. Average yield values for main crops across clusters are presented on the Table 1.

Figure 1. Proposed regionalization



Source: *Own estimation*

Table 1. Average yield values for main crops across clusters, 100 kg/ha

Cluster	Wheat	Corn	Barley	Soybeans	Rapeseed	Sunflower
1	51.45	85.59	46.62	26.09	31.70	29.71
2	47.11	75.39	40.03	21.89	28.70	26.91
3	38.24	52.78	33.31	18.82	22.79	21.34

Source: *Own calculations based on State Statistics Service of Ukraine data*

Introduction and motivation

Regionalization, in the context of agricultural production, refers to the division of a country into several regions based on various factors, which have an impact on production characteristics of the regions. Recognizing and accounting for these regional differences allows for more precise and detailed projections to be made, which would depict region-specific patterns and challenges.

Previously, regionalization in Ukraine's component of AGMEMOD model was implemented by Bogonos and Stepaniuk (2017). The proposed division of Ukraine into five regions (Mixed Forest, Forest Steppe, Steppe, Crimea, Donbas) was based on the estimation of crops yields differences in the different climatic regions.

The need for an updated regionalization stems from two aspects. Firstly, in context of the Russian full-scale invasion to Ukraine, and the implications it has brought to the agricultural sector, it is assumed that the new structural differences between the administrative regions arise. Damages, losses, and disruption in supply chains are expected to have a prolonged effect that would last for years, so we think that this component should be incorporated in the updated regionalization. Secondly, while yield differences encompass a significant amount of information about structural differences between the regions' agricultural production, there is still a large portion of variance that is not explained by a one-variable approach, as yield differences could be caused by different factors. In this analysis we aim to account for these factors as well, thus resulting in a more precise division.

In the following sections of this policy brief, we will discuss the approach which is used, selection of the indicators, and present the results of the analysis with the overview of proposed regions. To avoid the confusion, in the further text administrative regions (oblasts) are referred to as "regions", and the proposed regionalization units are referred to as "clusters".

Approach and selection of indicators

The goal of regionalization is to improve the accuracy of forecasts and test different development scenarios for the agricultural sector.

The main methods for conducting regionalization are K-means, standard deviation of the yields of the main agricultural crops in Ukraine (wheat, corn, barley, soybean, rapeseed, sunflower).

The efficiency indicator of regionalization is the minimum deviation of the yield of the main agricultural crops combined into clusters compared to the yield deviation without regionalization.

Input data:

- Yield: wheat, corn, barley, soybean, rapeseed, sunflower.
- Yield taken on average for 2019, 2020, 2021, 2023 years (2022 is omitted, because of the outstanding production circumstances in the first year of the full-scale invasion).

- By types of producers: all producers, all enterprises, farms, households;
- By regions of Ukraine: – 20 oblasts (excluding ARC, Kherson, Zaporizhia, Donetsk, Luhansk).

Crimea was excluded from the analysis due to the unavailability of reliable recent data. However, it is assumed that Crimea should be placed in a separate region in the proposed regionalization as it was done in previous regionalization due to its distinct climatic and economic characteristics.

Kherson, Zaporizhia, Donetsk, and Luhansk were singled out into a separate cluster due to a complex of specific conditions including military actions that significantly affect all aspects of life, including agriculture. Constant conflicts lead to infrastructure destruction making normal agricultural business operations practically impossible. Many agricultural lands are mined, making them dangerous for cultivation and use. This significantly reduces the area available for agriculture and complicates mechanized land processing. Infrastructure objects, including irrigation systems, roads, warehouses, and agricultural enterprises, are largely destroyed. These factors make these regions unique in the context of agricultural activity and require a special approach in planning and managing agricultural resources.

Various methods of regionalization and clustering that can be applied in scientific research are considered in the literature. For example, in the work "Regionalization and Clustering in Applied Geography" by Aydin, O., Janikas, Mark. V., Assunção, R. M., & Lee, T. H., hierarchical (agglomerative and divisive) and non-hierarchical (k-means) clustering methods for forming regions based on object similarity are described. Nearest neighbor analysis and spatial autocorrelation help to identify spatial patterns and the degree of connection between variable values in neighboring locations. In the book "Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment," edited by N. Janardhana Raju, geostatistical methods such as kriging and variographic analysis for interpolation and modeling of spatial data, as well as multivariate geostatistics for analyzing interrelated variables, are considered. Geospatial analysis methods, including Geographic Information Systems (GIS) and remote sensing (RS), are important tools for collecting, storing, analyzing, and visualizing geospatial data. Given the variety of methods, the K-means method proves to be the most suitable for regionalization due to its ability to efficiently cluster data based on a predefined number of clusters and provide high accuracy and consistency of results.

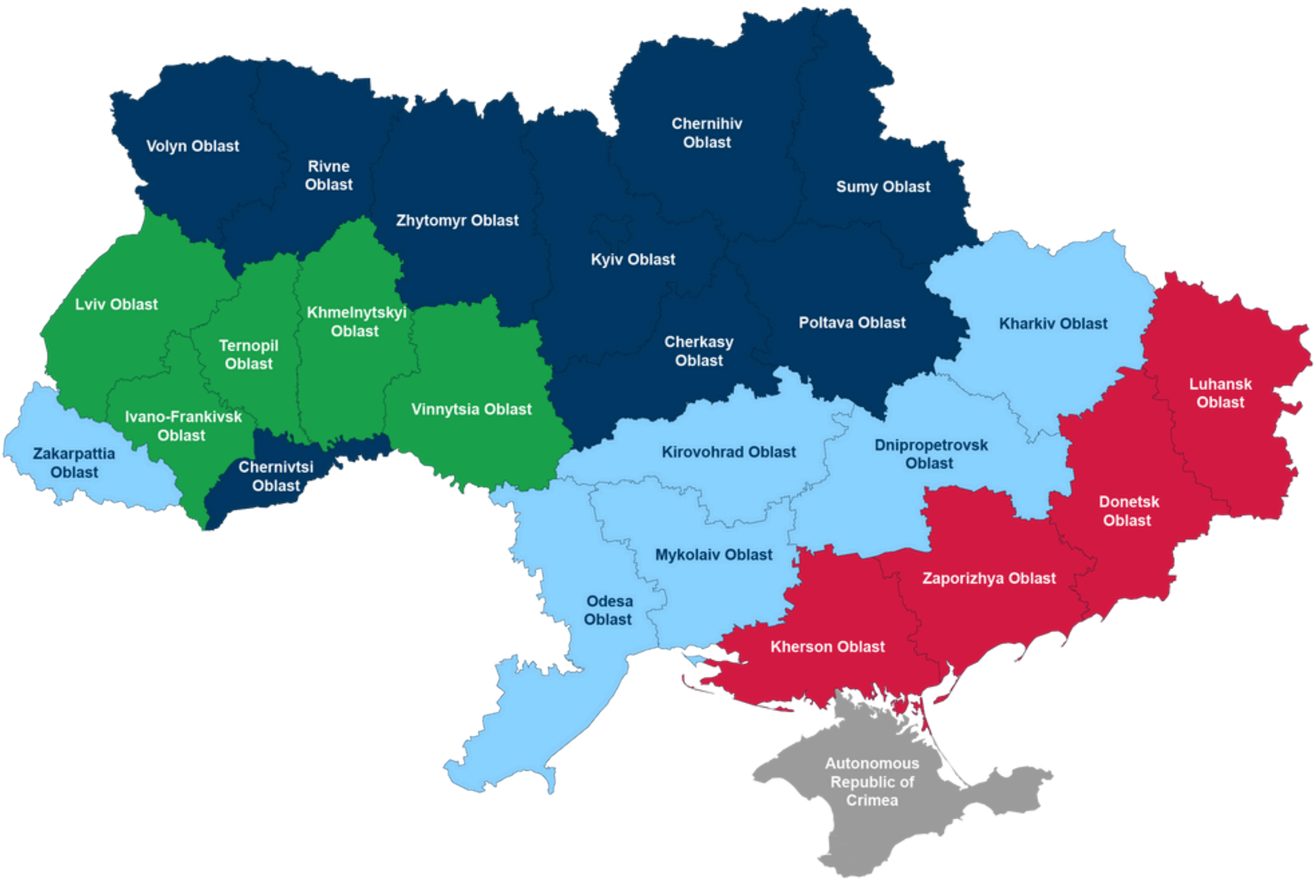
Approaches to Clustering:

1. Clustering based on natural and climatic conditions. Input data: the area of forested territories, grassy cenoses, and cultivated lands (data obtained using the Dynamic World V1 algorithm for classifying Sentinel-2 satellite images), average temperature for each month, and monthly precipitation based on information obtained from the ERA5 satellite (Daily Aggregates). The data for analysis were taken for the year 2021. The results are presented in Appendix 1.

2. Clustering based on the yield and area of main agricultural crops. The results are presented in Appendix 2.
3. Clustering based on the hydrothermal coefficient. The hydrothermal coefficient (HTC) is an indicator used in agronomy and meteorology to assess the relationship between temperature and precipitation in a given region over a specified period, usually the growing season. It provides insight into hydrothermal conditions that are critically important for the development of agricultural crops. HTC is calculated using the formula: HTC = R / (0.1 × T), where: R – total precipitation for a given period (in millimeters), T – total daily average temperatures for the same period (in degrees Celsius). High HTC values indicate wet conditions, while low values indicate arid conditions. HTC is typically used to assess the moisture availability of an area, which is important for planning agronomic measures []. The results are presented in Appendix 3.
4. Clustering based on the yields of main agricultural crops. Comparison of the average standard deviation of main agricultural crop yields based on the four methodological approaches presented above using the K-means method shows that the fourth clustering approach - based on the yield of main agricultural crops - is the most appropriate for further regionalization.

Comparison of the average standard deviation of main agricultural crop yields based on the four methodological approaches presented above using the K-means method shows that the fourth clustering approach - based on the yield of the main agricultural crops - is the most appropriate for further regionalization.

Figure 2. Proposed 4-cluser division



Source: *Own estimation*

Results and overview of clusters

The proposed 4-cluster division is presented on the Figure 2. The obtained clusters are:
North-Center cluster: Chernihiv, Sumy, Poltava, Kyiv, Zhytomyr, Cherkasy, Volyn, Rivne, Chernivtsi oblasts.
West-South cluster: Lviv, Ternopil, Khmelnytskyi, Vinnytsia, Ivano-Frankivsk oblasts.
Center-East cluster: Kharkiv, Dnipropetrovsk, Kirovohrad, Odesa, Mykolaiv, Zakarpattia oblasts.
South-East cluster: Luhansk, Donetsk, Zaporizhzhia, and Kherson oblasts.

Comparison of the average standard deviation of the yields of the main agricultural crops across all regions and across clusters

	Wheat	Corn	Barley	Soybeans	Rapeseed	Sunflower
Average yield standard deviation across all regions						
All producers	6.8	18.8	5.5	4.3	4.0	4.2
All enterprises	7.2	15.4	6.8	3.9	4.1	4.1
Enterprises (except farms)	7.5	14.3	7.6	3.4	4.0	4.1
Farms	6.6	18.0	6.1	5.1	4.5	4.0
Households	5.2	9.7	4.5	3.2	3.6	3.6
Average yield standard deviation across three clusters						
All producers	4.3	8.3	2.2	3.1	2.5	2.5
All enterprises	4.2	8.3	3.1	2.8	1.9	2.4
Enterprises (except farms)	4.4	8.1	3.9	2.4	2.1	2.6
Farms	4.0	8.9	2.8	4.5	2.5	2.4
Households	3.8	6.1	3.4	2.2	2.8	2.8

Regionalization significantly reduces the standard deviation of yields for all categories of producers and all main crops, indicating an increase in yield stability (in terms of variation from the mean) within each cluster. For clarity, let's consider specific examples. For wheat, among all producers, the standard deviation decreased from 6.8 to 4.3; among enterprises – from 7.2 to 4.2; among enterprises (excluding farms) – from 7.5 to 4.4; among farms – from 6.6 to 4.0; among households – from 5.2 to 3.8. For corn, among all producers, the standard deviation decreased from 18.8 to 8.3; among enterprises – from 15.4 to 8.3; among enterprises (excluding farms) – from 14.3 to 8.1; among farms – from 18.0 to 8.9; among households – from 9.7 to 6.1. For barley, among all producers, the standard deviation decreased from 5.5 to 2.2; among enterprises – from 6.8 to 3.1; among enterprises (excluding farms) – from 7.6 to 3.9; among farms – from 6.1 to 2.8; among households – from 4.5 to 3.4. For soybeans, among all producers, the standard deviation decreased from 4.3 to 3.1; among enterprises – from 3.9 to 2.8; among enterprises (excluding farms) – from 3.4 to 2.4; among farms – from 5.1 to 4.5; among households – from 3.2 to 2.2. For rapeseed, among all producers, the standard deviation decreased from 4.0 to 2.5; among enterprises – from 4.1 to 1.9; among enterprises (farms) – from 4.0 to 2.1; among farms – from 4.5 to 2.5; among households – from 3.6 to 2.8. For sunflower, among all producers, the standard deviation decreased from 4.2 to 2.5; among enterprises – from 4.1 to 2.4; among enterprises (farms) – from 4.1 to 2.6; among farms – from 4.0 to 2.4; among households – from 3.6 to 2.8.

The highest effectiveness of clustering is observed for crops such as wheat, corn, and barley, where the standard deviation significantly decreases. For wheat, the deviation decreases from 6.8 to 4.3 among all producers, indicating better grouping of regions based on their agricultural characteristics. For barley, the deviation decreases from 5.5 to 2.2, indicating significant improvement in yield stability within clusters.

The greatest reduction in yield deviation is observed in groups of all producers and households. For example, for corn, among all producers, the deviation decreases from 18.8 to 8.3, indicating a significant increase in yield stability and predictability. This more than doubling reduction means that regionalization allows reducing the impact of adverse factors, such as weather conditions and uneven distribution of resources. Among households, the deviation decreases from 9.7 to 6.1, also indicating the effectiveness of regionalization. As a result, farms have the opportunity to better plan their resources, distribute labor, and optimize the use of fertilizers and plant protection products.

The results show that using K-means methods and yield standard deviation analysis for regionalizing Ukrainian regions is an effective approach to increasing forecast accuracy and improving the stability of agricultural production. This enables better planning of agricultural strategies, adaptation to climate change, and improved resource management. Moreover, increased yield stability contributes to the growth of regional economic stability and provides more predictable incomes for farmers, which, in turn, positively affects the overall development of agriculture in Ukraine.

References:

Aydin, O., Janikas, M. V., Assunção, R. M., & Lee, T. H. (2021). A quantitative comparison of regionalization methods. International Journal of Geographical Information Science, 35(11), 2287-2315.

Bogonos, M., O. Stepaniuk (2017): Agricultural Outlook Ukraine 2017-2030. Baseline: projection of development of the agricultural sector in current economic and political frameworks and absent monetary state support. Agricultural Policy Report APD/APR/06/2017, German-Ukrainian Agricultural Policy Dialogue, Kyiv.

Brown, C.F., Brumby, S.P., Guzder-Williams, B. et al. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci Data 9, 251 (2022). doi:10.1038/s41597-022-01307-4

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), (date of access), https://cds.climate.copernicus.eu/cdsapp#!/home

Raju, N. J. (2016). Geostatistical and geospatial approaches for the characterization of natural resources in the environment. Cham: Springer, 969.

Клименко, Т. В., & Трембіцька, О. І. (2021). Зміна температурного режиму повітря та гідротермічного коефіцієнта (ГТК) вегетаційного періоду у зоні Полісся України. Sciences of Europe, (78-2), 5-7.

Annex 1

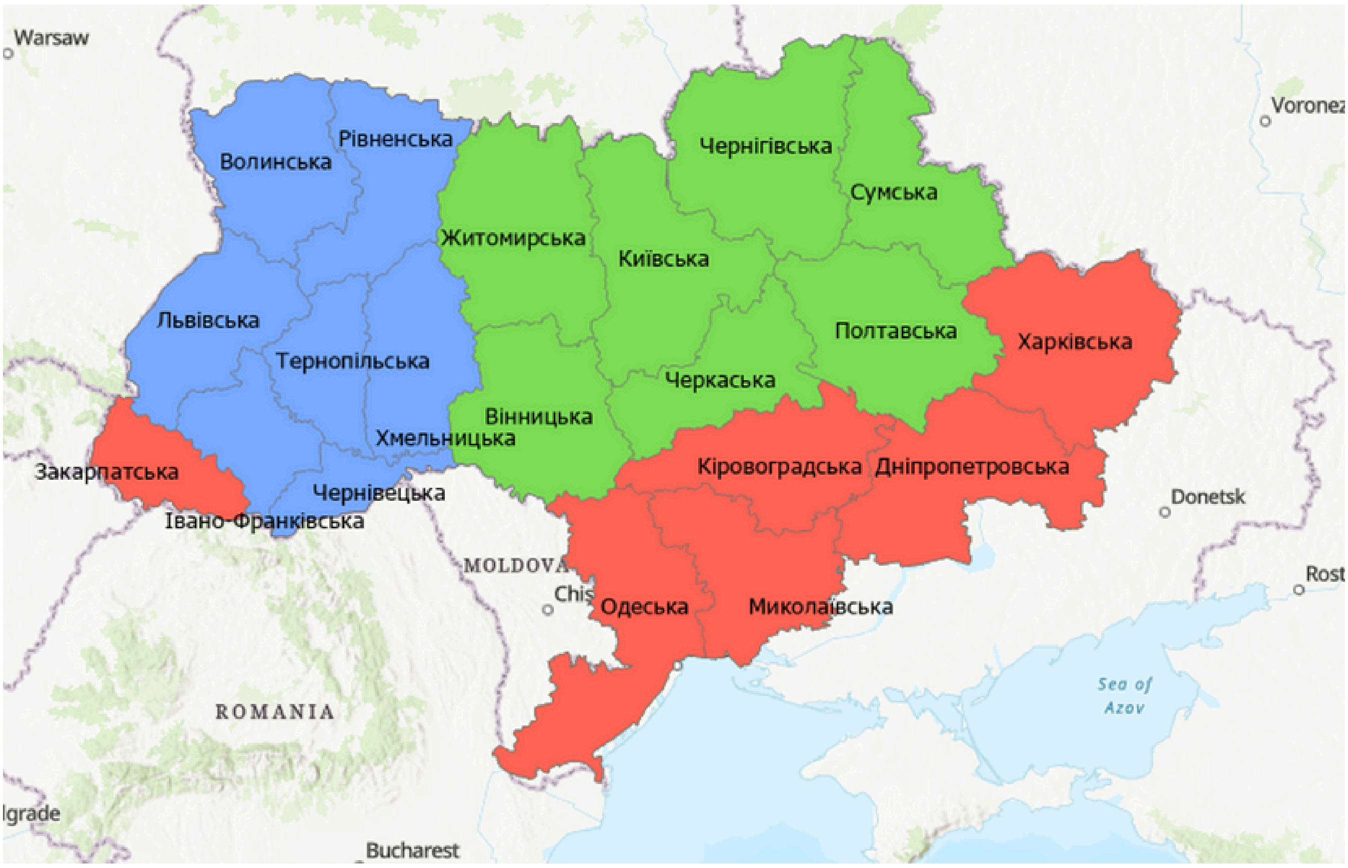
1.Clustering based on natural and climatic conditions

Average standard deviation of the yields of main agricultural crops

	Wheat	Corn	Barley	Soybeans	Rapeseed	Sunflower
Average yield standard deviation across all regions						
All produsers	6.8	18.8	5.5	4.3	4.0	4.2
All enterprises	7.2	15.4	6.8	3.9	4.1	4.1
Enterprises (except farms)	7.5	14.3	7.6	3.4	4.0	4.1
Farms	6.6	18.0	6.1	5.1	4.5	4.0
Households	5.2	9.7	4.5	3.2	3.6	3.6
Average yield standard deviation across three clusters						
All producers	6.0	11.7	4.3	2.8	2.7	2.8
All enterprises	6.1	11.4	4.3	3.9	4.4	2.1
Enterprises (except farms)	5.0	13.9	5.1	4.7	4.3	2.8
Farms	6.4	10.5	4.9	2.4	3.6	3.4
Households	6.1	14.6	5.2	3.3	2.8	3.3

Annex 2

2. Clustering based on yields and sown areas of the main agricultural crops

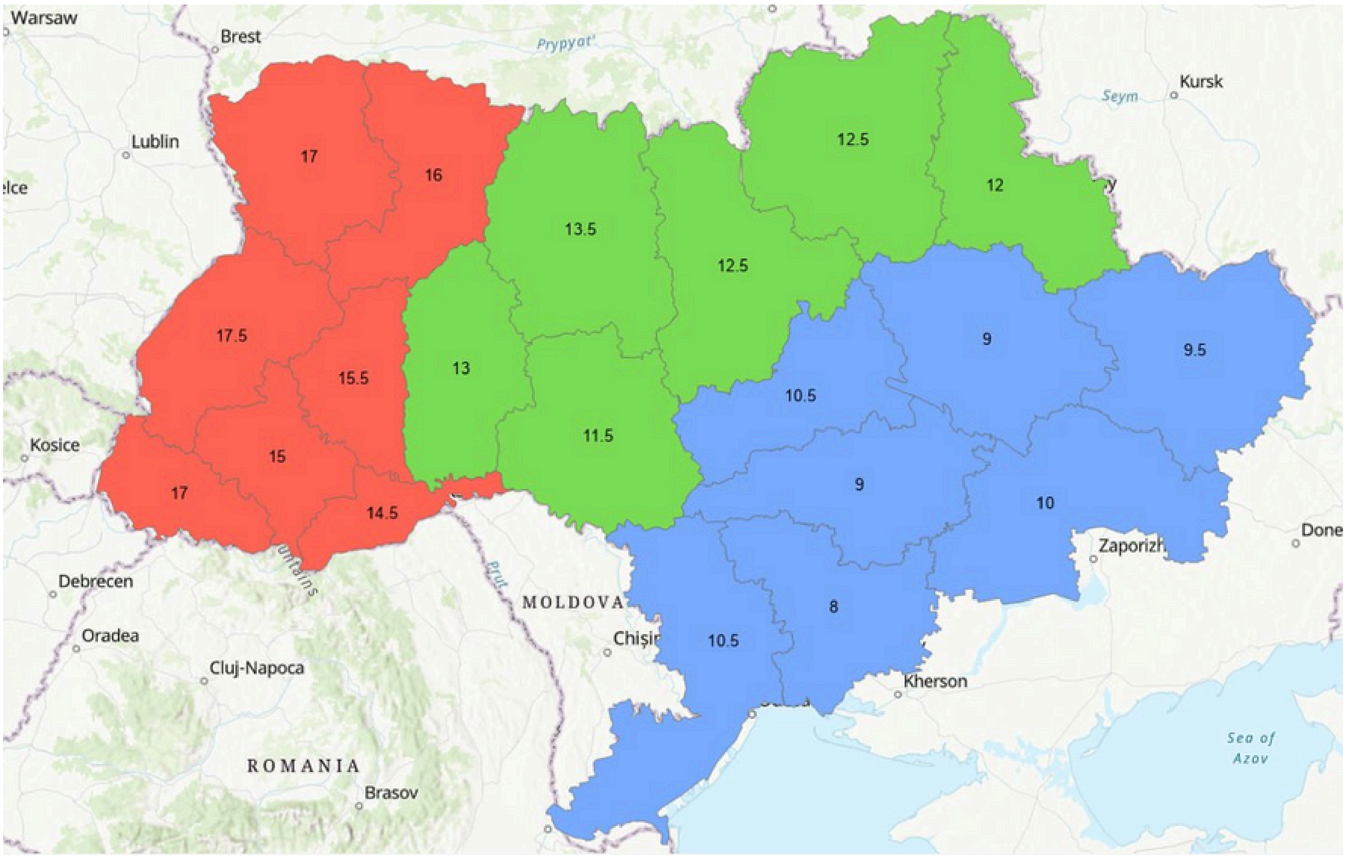


Average standard deviation of the yields of main agricultural crops

	Wheat	Corn	Barley	Soybeans	Rapeseed	Sunflower
Average yield standard deviation across all regions						
All produsers	6.8	18.8	5.5	4.3	4.0	4.2
All enterprises	7.2	15.4	6.8	3.9	4.1	4.1
Enterprises (except farms)	7.5	14.3	7.6	3.4	4.0	4.1
Farms	6.6	18.0	6.1	5.1	4.5	4.0
Households	5.2	9.7	4.5	3.2	3.6	3.6
Average yield standard deviation across three clusters						
All producers	6.1	11.9	4.9	2.7	2.8	3.0
All enterprises	6.6	12.1	4.9	3.9	4.5	2.3
Enterprises (except farms)	5.6	14.2	5.5	4.7	4.4	3.0
Farms	6.5	11.1	5.4	2.7	2.7	3.8
Households	5.6	12.1	4.5	3.3	3.4	3.2

Annex 3

3. Clustering based on the values of hydrothermal coefficient



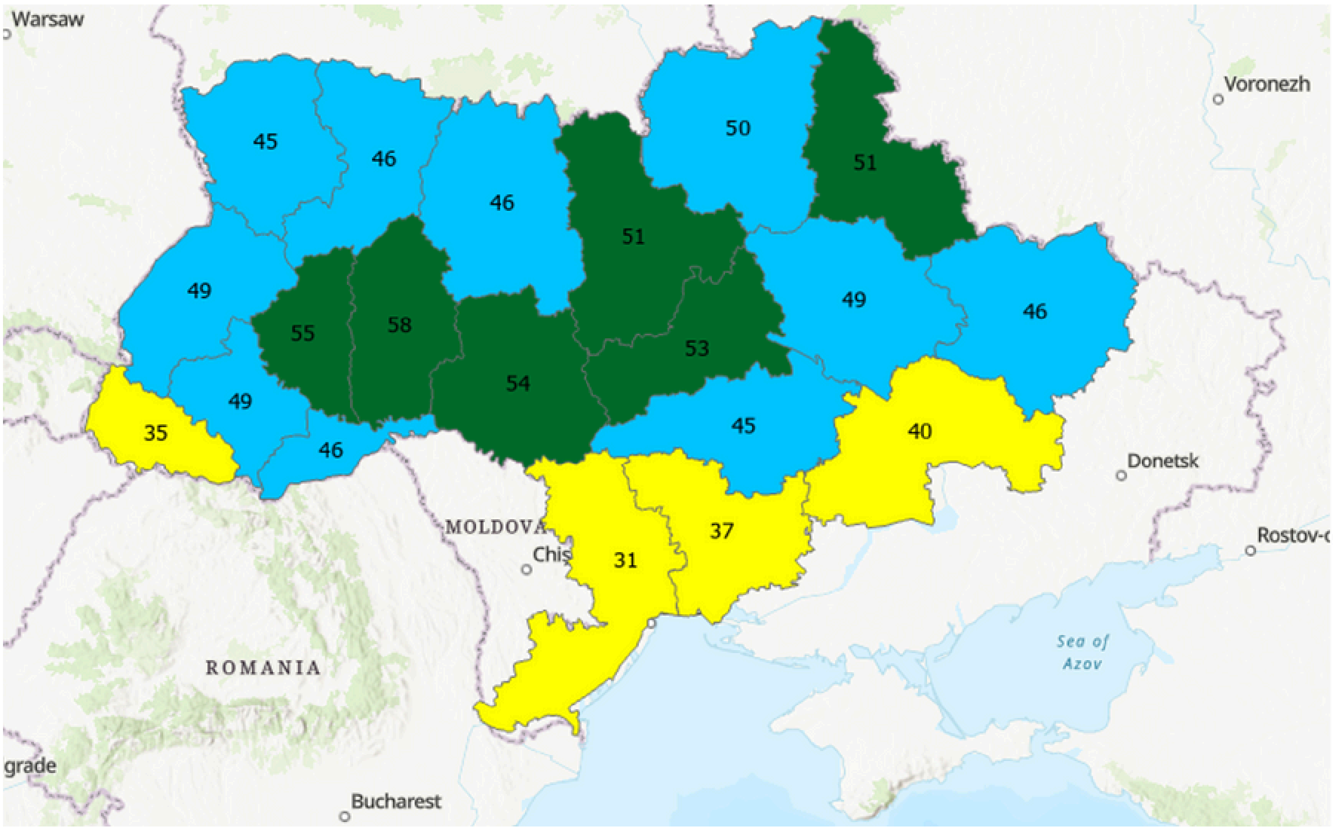
Average standard deviation of the yields of main agricultural crops

	Wheat	Corn	Barley	Soybeans	Rapeseed	Sunflower
Average yield standard deviation across all regions						
All produsers	6.8	18.8	5.5	4.3	4.0	4.2
All enterprises	7.2	15.4	6.8	3.9	4.1	4.1
Enterprises (except farms)	7.5	14.3	7.6	3.4	4.0	4.1
Farms	6.6	18.0	6.1	5.1	4.5	4.0
Households	5.2	9.7	4.5	3.2	3.6	3.6
Average yield standard deviation across three clusters						
All producers	5.8	12.2	4.6	2.3	2.9	3.7
All enterprises	7.0	11.9	4.6	3.6	4.7	3.9
Enterprises (except farms)	5.9	12.0	5.7	4.0	4.4	3.7
Farms	7.2	11.2	6.0	2.8	3.4	3.5
Households	6.1	13.1	4.3	3.0	3.5	4.0

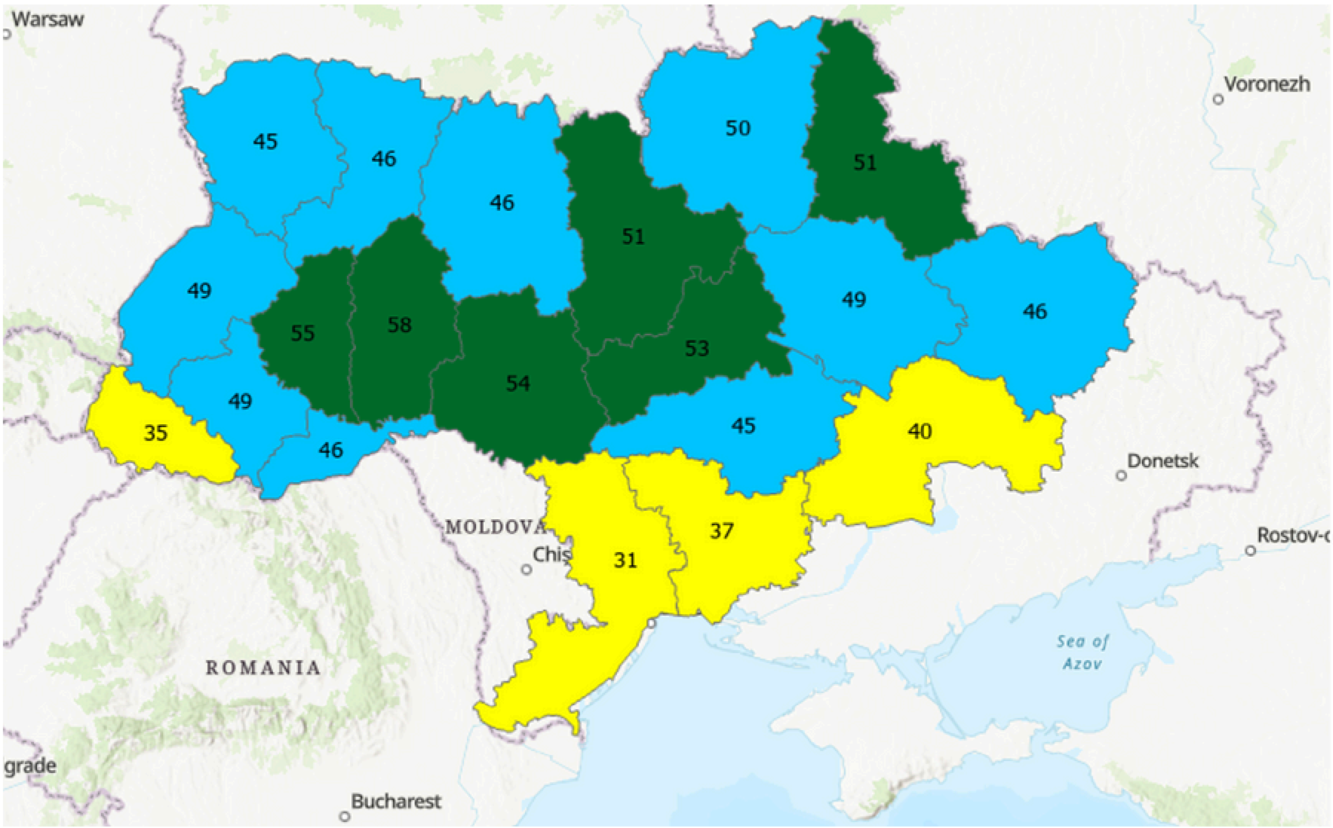
Annex 4

4. Clustering of regions based on the yield of each crop individually.

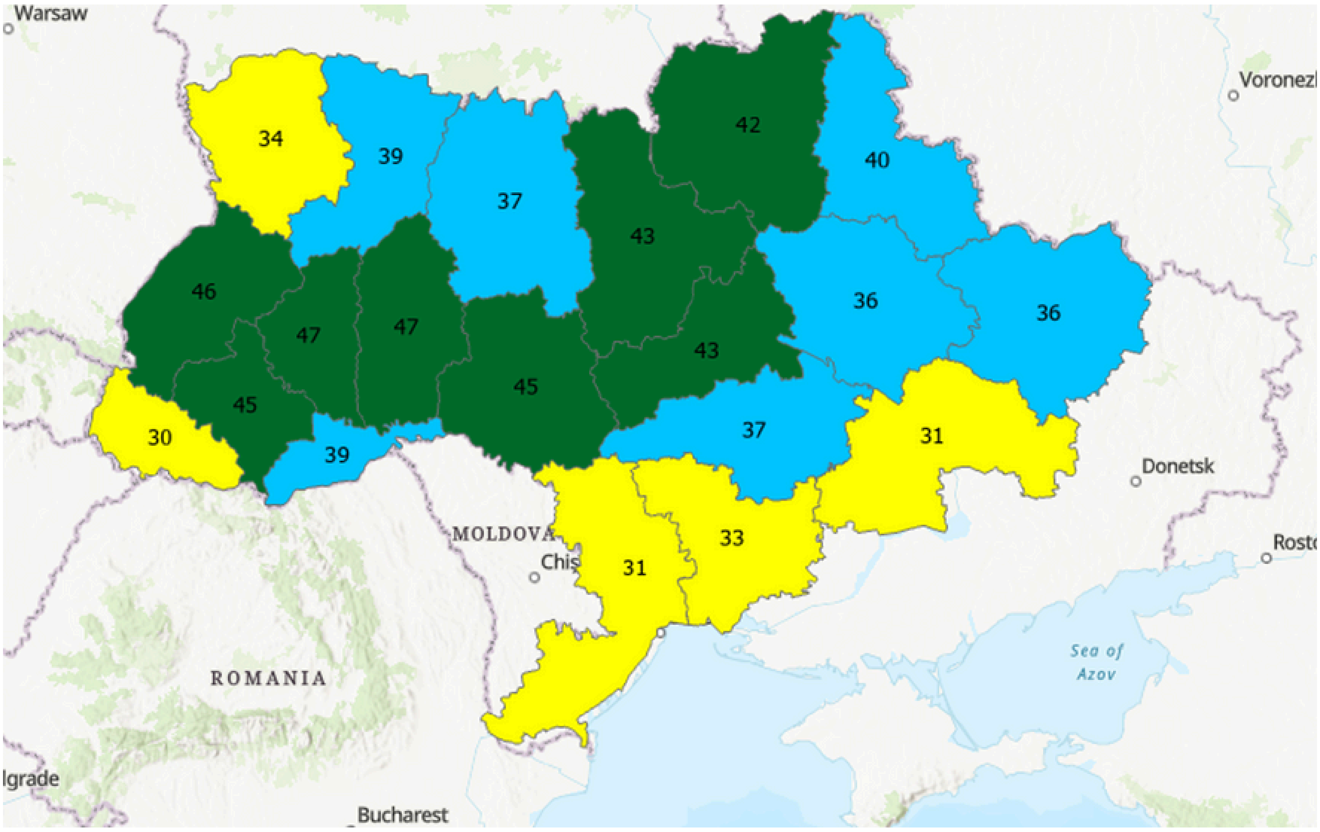
Clustering based on the wheat yield values



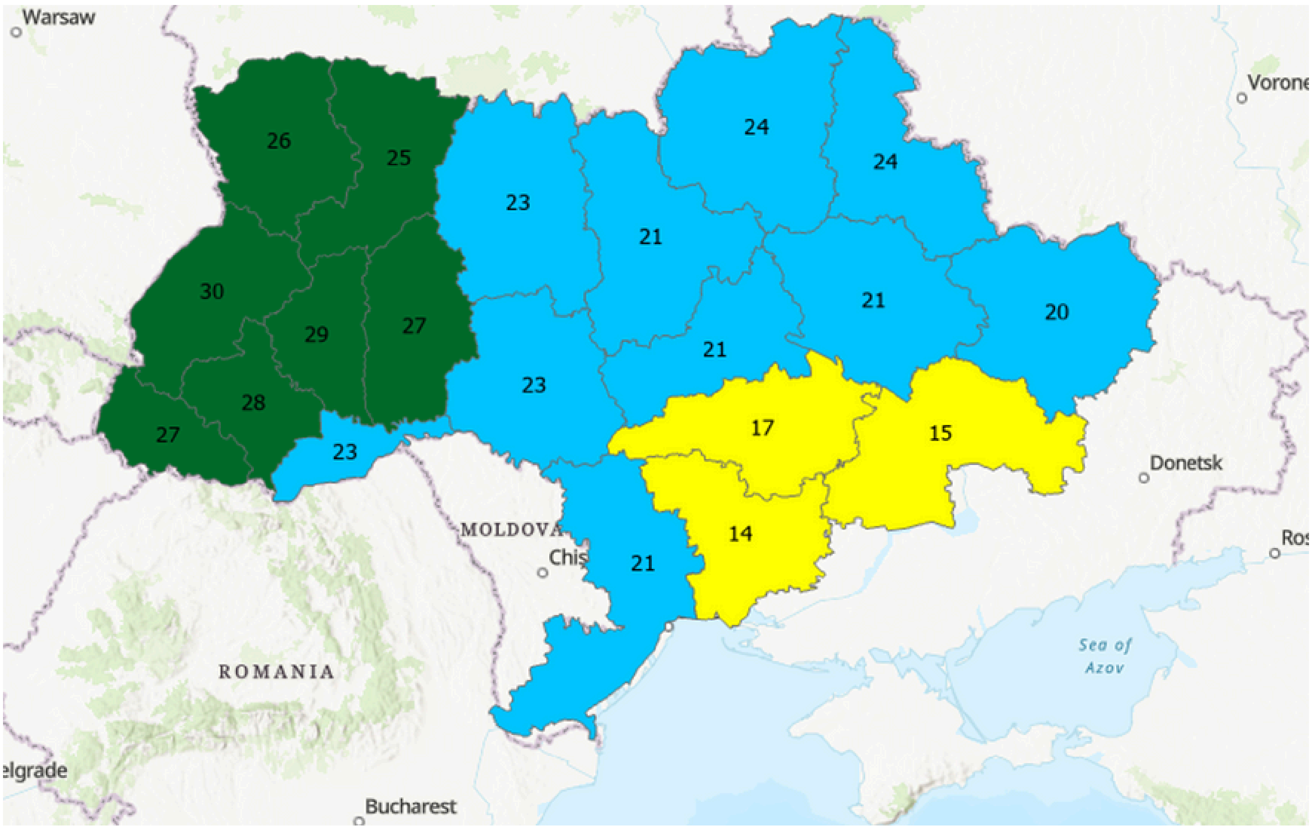
Clustering based on the corn yield values



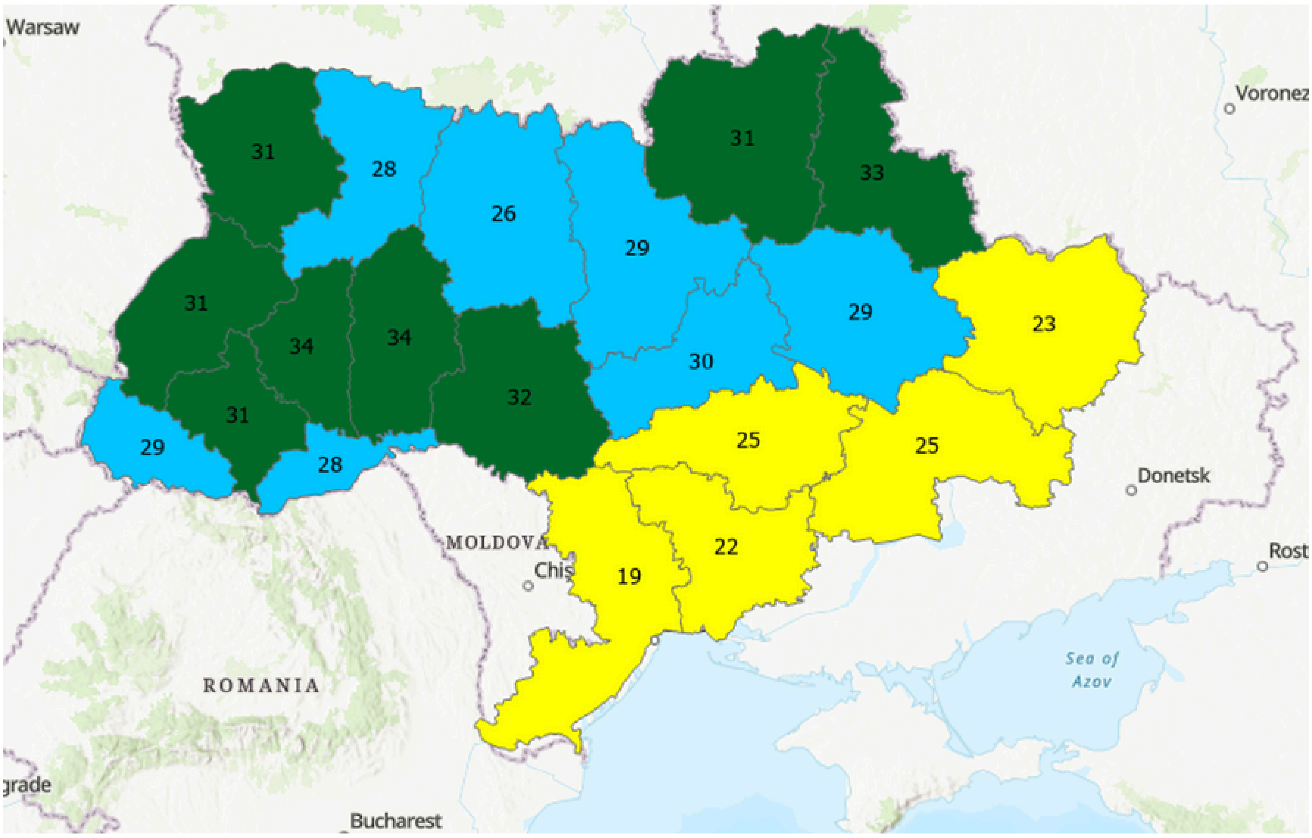
Clustering based on the barley yield values



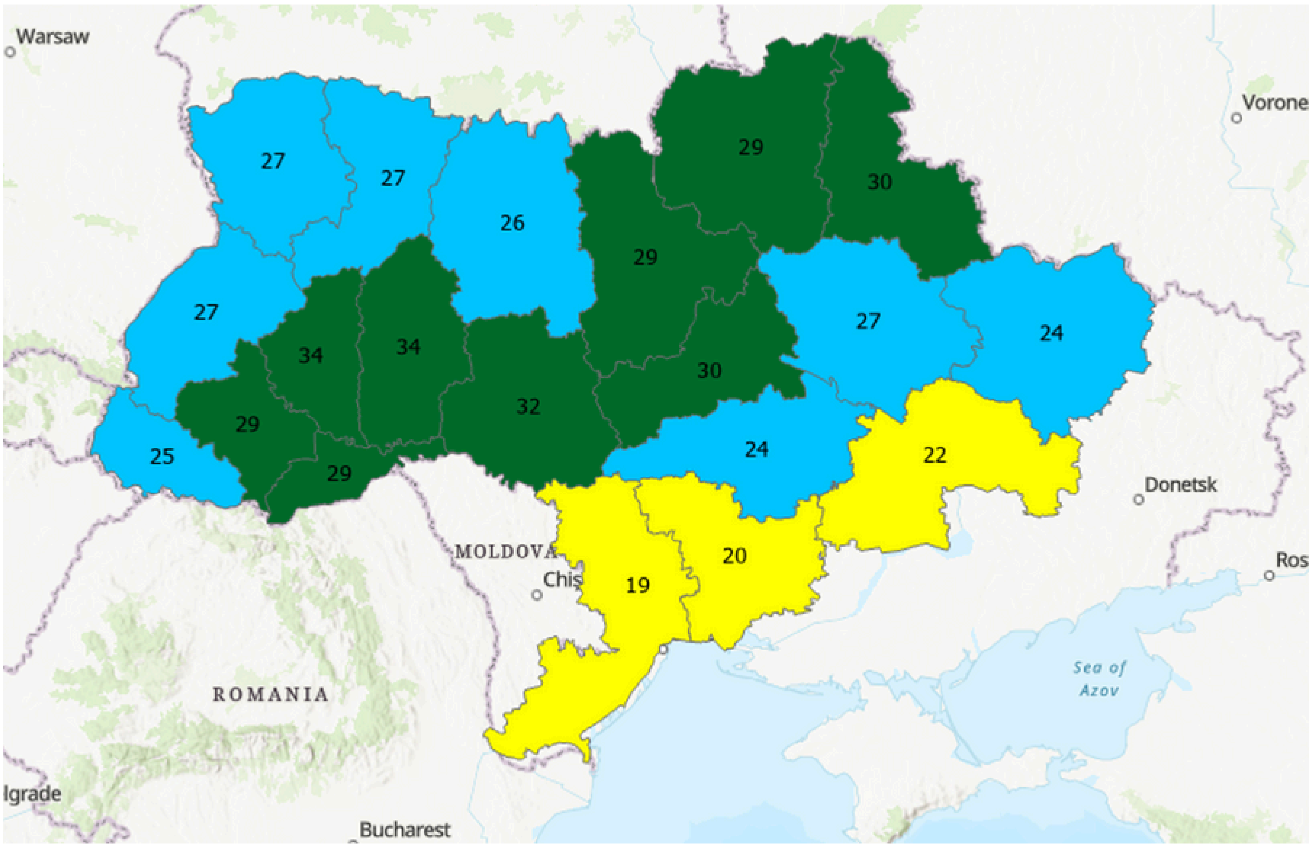
Clustering based on the soybeans yield values



Clustering based on the rapeseed yield values



Clustering based on the sunflower yield values



Contacts

Center for Food and Land
Use Research (KSE Agrocenter)

Kyiv School of Economics
Mykoly Shpaka St, 3

www.agrocenter.kse.ua

agrifood@kse.org.ua

Telegram
KSE Agrocenter

www.kse.ua

KSE | Центр досліджень продовольства та землекористування