Development of hierarchical terron workflow based on gridded data – A case study in Denmark

Yannik E. Roell, Yi Peng, Amélie Beucher, Mette B. Greve, Mogens H. Greve

PII: S0098-3004(19)30140-2
DOI: https://doi.org/10.1016/j.cageo.2020.104454
Reference: CAGEO 104454

To appear in: Computers and Geosciences

Received Date: 13 February 2019
Revised Date: 11 December 2019
Accepted Date: 20 February 2020


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.
Abstract:
Defining homogeneous zones by soil, climate, and landscape is a first step in creating terroir units. Applications in delineating homogeneous zones are commonly interested in all land use types (urban, forest, agriculture, etc.) or vegetation specific. Terron mapping is one method for creating non-vegetation homogeneous zones specifically for agricultural management and environmental assessment. Previously, terrons were defined as areas similar in soil and landscape. In this study, we implement climate as a key component in terron delineation and a workflow is developed that automates the necessary steps in generating and mapping terron classes. The workflow is flexible by allowing the user to define different input variables, method of modeling, output resolution, and hierarchical terrons. To assess the workflow, terron classes are modeled and mapped in Denmark: 1) using fuzzy c-means with various soil and climate gridded covariates to identify national regions, and 2) using k-means to develop regional terrons within the national regions based on soil and landscape gridded covariates. The elbow method was applied to determine the optimal number of national terron classes, dividing Denmark into
three national regions. Each national region was further divided into nine regional terrons. The resulting regional terron map is comprised of 27 terrons. National and regional terrons are defined at 304 m and 30.4 m resolutions, respectively. A dendrogram produced from regional terron centroids was used to generalize terrons into groups of three to five terrons. From the generalization, seven groups are used to determine terrons that are most promising for crop production. The regional terron map constitutes a useful tool enabling future land-use management decisions and the development of terroirs for Danish crops.

Keywords: terroir, terron, clusters, fuzzy c-means, k-means

Authorship statements:

Yannik E. Roell collected and analyzed data, designed the analysis, and wrote the manuscript.

Yi Peng contributed to analyzing data and critically revised the manuscript.

Amélie Beucher contributed to analyzing data and critically revised the manuscript.

Mette B. Greve contributed to collecting data and critically revised the manuscript.

Mogens H. Greve supervised research project and critically revised the manuscript.

All authors give their final approval of the manuscript version to be submitted.

Declarations of interest: none

Link to code: https://github.com/yroell/terron_workflow_code
1 Introduction

Terron is an entity to quantitatively describe regions similar in soil and landscape (Carré and McBratney, 2005). Grouping physical characteristics in zones, as done in terron mapping, is a crucial step in creating a terroir (Malone et al., 2014). The concept of terroir, initially started with wine in France, has been linked with environment and quality in a variety of crops (Vaudour et al., 2015). Terroir refers to geographic areas where soil and microclimate are relatively homogenous and impact distinctive qualities of a food product (Wilson, 1998; Barham, 2003; Dougherty, 2012).

The quality of various agro-food products has been linked to ecology and culture of specific places through the concept of terroir. This ranges from beverages (e.g. tequila: Bowen and Zapata, 2009; coffee: Silva et al., 2014; tea: Besky, 2014) to foods (fruits, vegetables, cheese, honey, olive oil: Trubek, 2008; Jacobsen, 2010). Distinguishing between the origins of products has three main steps (Bonfante et al., 2011):

1) Create homogeneous zones of climate, topography, geology, and soil.
2) Analyze products and production techniques quantitatively and qualitatively in each homogeneous zone.
3) Organize homogeneous zones into a hierarchy system with respect to the product.

Development of terroir requires time and different techniques. Delimiting terroir units is typically executed in two stages (Carey et al., 2008): mapping environmental features to obtain homogeneous zones (i.e. step 1 above) and studying crops in the homogeneous zones (i.e. steps 2 and 3 above). Analysis of products in a qualitative way (e.g. panel taste test) is challenging and not objective.
The explosion of spatial data availability and sensing technologies provided several possibilities to enhance the process of delimiting terroir units in a more objective way (Vaudour et al., 2010; Urretavizcaya et al., 2013). Therefore, many studies focus on creating homogeneous zones based on available spatial data (Arnó et al., 2009; Pedroso et al., 2010).

Several definitions have been proposed regarding regions similar in soil, climate, and landscape depending on how the homogeneous zones are applied. Natural terroir units and basic terroir units are similar regions of the biosphere characterized by topography, climate, substrate, and soil (Laville, 1993; Morlat, 1989). Agroecozones are regions of crop production that are similar in biophysical characteristics, such as soil, landform, and climate (Anderson et al., 1999). Land systems are units in an area that have a pattern of topography, soil, and vegetation (Christian and Stewart, 1953). Each of the previous homogeneous zones are either created for a specific vegetation or use vegetation in the classification.

Soil and Terrain (SOTER) units are non-vegetation specific homogeneous zones derived from soil and landscape polygons (World Soil Resources Reports, 1993). Homogeneous zones derived from polygon maps lack the ability to interpret transition zones between attributes (Goodchild and Gopal, 1989). Furthermore, using polygons to map continuous data, such as soil texture, assumes a large area is uniform resulting in fewer approaches to analyze the spatial variation (Burrough, 1996). A vast array of approaches exist to analyze spatial variation when mapping continuous data in grids (Chrisman, 1991).

Terron mapping analyzes spatial variation using continuous gridded data by clustering regions with similar features. Two terron mapping approaches have been developed in the digital soil mapping (DSM) community. As a pioneer study of digital terron mapping, Carré and McBratney (2005) successfully classified and mapped 18 terron classes in La Rochelle region in
France based on 1113 soil profiles and digital elevation model (DEM) derivatives. These 18 terron classes were carried out by fuzzy c-means clustering based on a combination matrix including soil and landscape factors. Terron classes were interpolated using regression-kriging with landscape attributes as covariates to produce the final terron map. Instead of using soil profile data, Malone et al. (2014) used gridded data to define 12 terron classes in the Lower Hunter Valley in Australia. Fuzzy c-means with eight pedological and eight landscape variables were used to generate the final terron map.

Terron modeling, proposed by Carré and McBratney (2005), is a function of soil, landscape, and the interaction between the two. Previously, terron modeling has only been applied to small areas and was restricted to regions with a homogeneous climate. Depending on the scale of the study area or availability of microclimate data, climate should be considered as a part of terron modeling for agriculture and other land-use management purposes. Incorporating climate as an additional soil forming factor (Jenny, 1941), applications for terron modeling could be extended to larger regions with varying climates. The initial intent of combining soil and landscape for terron mapping was to support management decisions for agricultural land. Topography is an important component in assessing areas that are prone to leaching (Tiktak et al., 2002) or erosion (Mitasova et al., 1996), but climate is a necessary parameter in environmental risk models. Therefore, the following model incorporates climate as an essential component in the terron concept:

\[
\text{Terron} = f(\text{soil, climate, landscape, soil*climate*landscape})
\]

(1)

The present study has two main objectives: 1) to develop a generic automated system for terron classification at different scales based on gridded data, and 2) to classify and map terron classes in Denmark incorporating soil, climate, and landscape information at two different scales.
Creating an easily executable and versatile workflow involves allowing the user to set different input parameters, such as variables to cluster, method of modeling, output resolution, and number of scales (i.e. hierarchical terrons). Using Denmark as a case study, the workflow will be tested to see how efficiently data is handled at two different scales. Data will include low-resolution climate and soil gridded covariates at a national scale and high-resolution soil and landscape gridded covariates at a regional scale. The purpose for terrons in Denmark is to generate terrons that indicate areas which are best suited for crop production.

2 Methodology

A workflow was developed for flexibility when creating terrons. Four main parts are in the workflow (Figure 1): data handling, number of clusters, clustering algorithm, and creating hierarchical terrons. The four parts have parameters that the user can set to customize the terron creation. Two outputs are produced: 1) a map of terrons with an additional membership map if fuzzy c-means is used for clustering and 2) a file with raw location data to allow for future summary statistics. Each hierarchical level generates these two outputs. The script for the workflow was created in R 3.4 (R Core Team, Vienna, Austria).

2.1 Data Handling

Data handling has two parameters (Figure 1): variables for clustering and a resolution value (i.e. cell size). Terron classes are generated from variables collected in one folder (only TIFF files). The first file in the folder is used as the reference to determine the coordinate system and units that the resolution is defined in. Values from the TIFF files will be extracted from a set of points created from the reference file (one point for the centroid of each cell). The dataset
created from combining TIFF values is used for clustering and creating a terron map. For example, a 3 by 3 grid with four variables to cluster would generate a 9 by 4 dataset: nine points (one point for the centroid of each cell) each with four values (one for each variable). The second parameter is the resolution in map units from the coordinate system. The resolution value is used in two ways: to determine how many points are created from the reference file and to set the output map resolution.

2.2 Number of Clusters

The number of clusters used to cluster the dataset can be determined automatically using the elbow method or set manually (Figure 1). The elbow method determines the optimal number of clusters by running the clustering algorithm with an increasing number of clusters. By increasing the number of clusters, within-cluster dispersion will decrease. The elbow, or optimal number of clusters, is where the within-cluster dispersion does not change drastically with the addition of another cluster (Thorndike, 1953; Milligan and Cooper, 1985). Datasets do not always produce clear elbows; thus, a complementary approach to the elbow method is used. The data is randomized and clustering is performed. The within-cluster dispersion for the random data is noted as the baseline for data with no structure. The cluster number with the maximum difference between the actual data within-cluster dispersion and randomized data within-cluster dispersion is considered as the elbow (Kintigh and Ammerman, 1982; Kintigh, 1990) and used in the clustering algorithm. Otherwise, a number can be chosen manually for the number of clusters.

2.3 Clustering Algorithm
Two clustering algorithms are used in the workflow (Figure 1): k-means or fuzzy c-means. K-means, initially proposed by MacQueen (1967), is a popular and simple partitional unsupervised learning algorithm that is referred to as hard clustering since each point is assigned to only one cluster (Jain, 2010). To determine where the centroids of the clusters are, k-means aims to minimize the sum of squared error for all the clusters,

\[ J(C) = \sum_{j=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_j\|^2 \]  

(2)

where \( i \) is a set of n-dimensional points to be clustered into \( K \) clusters and cluster \( c_j \) has a mean of \( \mu_j \). K-means consists of three main steps: 1) generate initial \( K \) cluster centers within points, 2) assign each point to nearest cluster center, and 3) compute new cluster centers (Jain and Dubes, 1988). The second and third steps are repeated until clusters stabilize by minimizing the squared error for all \( K \) clusters.

Fuzzy c-means, initially proposed by Dunn (1973), is an extension of k-means that uses fuzzy set logic (Zadeh, 1965). Fuzzy c-means is referred to as soft clustering since each point is assigned a membership value for each cluster, with all memberships summing up to 1. The difference between k-means and fuzzy c-means is the addition of \( u_{ij}^m \) as the degree of membership of \( x_i \) for cluster \( j \) in Eq. 2, where

\[ u_{ij} = \frac{1}{\sum_{k=1}^{K} \left( \frac{\|x_i - \mu_j\|}{\|x_i - \mu_k\|} \right)^{2/(m-1)}} \]  

(3)

and \( m \) determines cluster fuzziness with a larger \( m \) indicating fuzzier clusters. When \( m = 1 \), memberships are either 0 or 1, resulting in hard clustering (k-means). Membership values range from 0 to 1 with 0 indicating not a member of the cluster and 1 indicating a complete member of the cluster.
The output from the clustering algorithms is a map of final clusters which are the terron classes. For fuzzy c-means, each location is assigned a terron class using the highest membership value. A membership map is produced by using the highest membership value for each location. Lower membership values indicate that the location does not fall clearly into any terron class and typically appears between terron class boundaries.

K-means is executed from the cluster package within R. Fuzzy c-means using Euclidean as the distance metric is either executed from the e1071 package or from the fuzme package, based on the stand alone FuzMe software (Minasny and McBratney, 2002). A detailed description of the fuzzy cluster analysis can be found in McBratney and Grujter (1992). Using the fuzme package will allow the clustering to be performed many times to optimize and fine-tune the results but drastically increases computation time. Three methods of clustering are included depending on the scale and resolution of the study area and the purpose of the terrons. If a membership map is needed, fuzzy c-means should be used. If the study area is large and high-resolution variables are used, either k-means or fuzzy c-means from the e1071 package should be used to reduce computation time.

2.4 Creating Hierarchical Terrons

A telescoping technique is used to create hierarchical sets of terrons. Terrons can be created in levels at different scales depending on the purpose. Each of the different levels allow for separate steps for data handling, number of clusters, and clustering algorithm. The parameters set in a level will be applied to each terron created in the previous level. This process continues until there are no more levels available. For example, if three terrons are created in the first level, the parameters set for the second level will be implemented three times. After each level,
cleaning is possible to remove the salt and pepper effect. This effect is mainly between cluster
boundaries and cleaning will create more continuous terrons before going to the next level. If the
user wants to smooth the terron boundaries before proceeding to the next level, the workflow
stops and the user can clean the data. Once the cleaning is finished, the user runs the code and
the workflow will continue just after the cleaning step.

3 Application to Denmark

3.1 Study Area

Denmark lies between 54.5 – 57.8°N and 8.0 – 15.2°E with a temperate climate and
covers over 43,500 km² of area. Mean annual temperature is 9°C and precipitation is 770 mm
with western regions receiving more rain than eastern regions (Wang, 2012). The soils of
Denmark are mainly loamy Luvisols in eastern and central regions of the country, while sandy
Podzols dominates the western regions (Jacobsen, 1984). Due to multiple glaciations during the
last glacial period (Weichselian Glaciation), soil distribution and formation are highly variable
(Schou, 1949). Eastern and western regions were developed in moraine landforms with western
regions being glaciofluvial flood plains and older eroded moraine landforms; the northern region
consists of raised seabeds. Topography is relatively flat, being formed by many glacial advances,
with a mean elevation of 32 m and a max elevation of 171 m. The majority of Denmark’s land
use is agriculture.

Denmark was divided into national and regional levels for this study using hierarchical
terrons to utilize both low-resolution climate and soil data and high-resolution soil and landscape
data. Variable selection was based on factors affecting general crop production and chosen based
on available data and expert knowledge. National variables and terrons have a 304 m resolution
and regional variables and terrons have a 30.4 m resolution. All variables were normalized between 0 and 1 before starting the workflow. Variables used for each level can be seen in Table 1.

3.2 National Terron

The six variables used for clustering the national terrons were divided into two categories: climate and soil (Figure 2). The four climate variables consisted of: number of frost days (below 0°C) during April 1 – 15 and October 1 – 15, annual number of growing days above 10°C, global solar radiation during the growing season (between April 1 and October 31), and precipitation during the growing season. Temperature and solar radiation data came from Aarhus University’s meteorological database (AGRO Climate, 2018). Downloaded data were from 1985 – 2013 at a resolution of 40-km grids and consisted of daily mean and minimum temperature in Celsius at 2 m above the ground and global solar radiation in MJ/m². Kriging was performed on the 40-km grids to generate maps of frost, growing days, and global solar radiation. Precipitation is from monthly averages between 1961 – 1990 at a resolution of 40 km (Scharling, 2000).

The soil variables correspond to the first and second principal component (PC) score values from topsoil (0 – 20 cm) spectra. The PC scores were generated from 578 topsoil visible-near infrared spectra. The PC analysis was performed in R using non-linear iterative partial least squares and mapped using kriging. The first two PCs explain 96% of variation in the spectra data (PC 1 explains 87% and PC 2 explains 9%). See Knadel et al. (2013) for a more detailed explanation for generating PC scores from topsoil spectra in Denmark.

National terrons were defined using fuzzy c-means from the fuzme package with five runs to optimize the results. The elbow method was applied to determine how many classes
should be selected (Figure 3). Clustering was performed with cluster numbers between 2 and 12 to obtain the optimal number of clusters. The number of clusters decided upon from the elbow method was three. Three classes were used to perform fuzzy c-means. Boundary cleaning was accomplished in ArcGIS using the majority value for statistical type within Focal Statistics with a 10 by 10 neighborhood.

3.3 Regional Terron

The six variables used for clustering the regional terrons were divided into two categories: soil and landscape (Figure 4). The four soil variables were: clay percentage between 100 – 200 cm, pH between 60 – 100 cm, organic carbon content in g/kg between 0 – 30 cm, and the root zone capacity in mm. Clay, pH, and organic carbon were acquired from Adhikari et al. (2013a, 2014a, 2014b). Root zone capacity was developed using the methods from Madsen et al. (1998). Wetlands were excluded from the root zone capacity calculation but were later given a value of 1, once the dataset was normalized.

The two landscape variables were slope (in degrees) and Saga Wetness Index (WI). Both slope and Saga WI were generated from a 30.4 m resolution digital elevation model (Adhikari et al., 2013b).

Regional terrons were defined using k-means. Fuzzy c-means was not performed on the regional terrons due to computation time. The number of regional terron classes was manually selected as the number of soil classes in each national terron (Adhikari et al., 2013b). This method of determining the number of terron classes was utilized by Carré and McBratney (2005) and Malone et al. (2014). Using the number of soil classes resulted in nine regional terrons in each national terron for a total of 27 regional terrons.
4 Application Results and Discussion

4.1 National Terrons

The three terrons are represented as North (national terron 1), West (national terron 2), and East (national terron 3). The percent variance explained (ratio between within-cluster sum of squares and total sum of squares) is 70% using three clusters for the six national variables. The national terron map and associated highest membership are represented in Figure 5.

National terrons are summarized using the 25th to 75th percentile values for each national variable (Table 2). Coarse sand and clay from Adhikari et al. (2013a) at a depth of 0 – 30 cm are added to the table for comparison with the PC values. The north terron, encompassing 35% of Denmark, has fewer growing days above 10°C per year. The west terron, encompassing 27% of Denmark, has more precipitation during the growing season. The east terron, encompassing 38% of Denmark, has less precipitation and more solar radiation. The North and West have similar number of frost days while the West and East have similar number of growing days. The topsoil texture of Denmark correlates with the topsoil PCs (Greve et al., 2007; Knadel et al., 2013). A high PC 1 value corresponds to a higher percentage of coarse sand, as seen in the West, and a low PC 2 value corresponds to a higher percentage of clay, as seen in the East. Combining the PC maps, the glacial boundary from the last glaciation is visible. This boundary is the major division between the west and east terrons.

4.2 Regional Terrons

The map of all 27 regional terrons is in Figure 6. Regional terrons have been given a unique value to differentiate between national terrons. A unique terron value is a combination of
the national terron (North = 1, West = 2, East = 3) and the regional terron (1 to 9). For example, in the east national terron, the first regional terron is identified as terron 31.

Regional terrons are summarized using 25th to 75th percentile values for each variable and the two most prevalent soil classes (Adhikari et al., 2013b) in the terron (Table 3). To determine the similarities between each of the terrons, a Euclidean distance matrix from terron centroids of the six regional variables was created. The matrix was generated using the `hclust` function in R to produce a dendrogram (Figure 7). The 27 terrons were divided into seven groups for generalization. The seven groups are described below.

4.2.1 Group 1 – Wetland Areas

Group 1 are terrons 13, 15, 21, 27, and 32. The terrons are associated with the highest root zone capacity values and groundwater close to the soil surface, indicating wetland areas. The North and West have two terrons for wetland areas. The North is divided into two wetland areas: high organic carbon content (terron 13), indicating raised bogs in the far North, and low organic carbon content (terron 15), for the rest of the wetland areas. The West is divided into two wetland areas: low clay content (terron 21) and high clay content (terron 27). The low clay content makes terron 15 and 21 similar and the high clay content makes terron 27 and 32 similar. The raised bogs with high organic carbon in the North makes terron 13 the most distinct wetland area.

4.2.2 Groups 2-7 – Non-Wetland Areas

Groups 2-7 are different from group 1 because these areas have no groundwater table in the root zone. On average, all these groups have similar organic carbon content and Saga WI
values. Each group is assigned the dominating soil classes according to FAO (Adhikari et al., 2013b) and the most similar Danish soil classification number (i.e. JB; Greve and Breuning-Madsen, 1999) to determine the major soil type for the group, and described according to crop production.

4.2.2.1 Group 2 – Loamy Soils in East

Group 2 are terrons 31, 33, and 35. These terrons have higher pH (> 7.5), clay percentage, and root zone capacity compared to the rest of the non-wetland groups. The first and second soil classes are Luvisols and Cambisols, respectively, and group 2 is associated with sandy clayey soil (i.e. JB 6). Group 2 is the best for most crops in Denmark. The most fertile soils in Denmark are in the East (Rothe, 1844); thus, no north or west terrons are present in this group.

4.2.2.2 Group 3 – Loamy Sands

Group 3 are terrons 19, 23, and 36. Terrons in group 3 have high pH (~ 7.0) and are relatively sandy with little clay. Terrons 19 and 36 have a higher root zone capacity, making these terrons better for crops than terron 23. The first or second soil class in these areas are Arenosols. Group 3 is associated with clayey sandy soil (i.e. JB 3 and 4) and has average growing conditions for crops.

4.2.2.3 Group 4 – Fluvial Sands in West

Group 4 are terrons 16, 17, 24, and 26. The soil is sandy with low pH and a small root zone. The terrons are relatively low in all variables but terrons 17 and 26 are both steeper than
terrons 16 and 24. The first and second soil class in all these terrons are Podzols and Arenosols, respectively. This group is associated with coarse sandy soil (i.e. JB 1). Group 4 is the worst for most crops in Denmark.

4.2.2.4 Group 5 – Sandy Loams

Group 5 are terrons 11, 22, 28, and 37. These terrons have high clay content and root zone capacity. The West has two terrons in this group with terron 28 being associated with steeper areas and higher root zone capacity values. The first or second soil class in these areas are Podzols. Group 5 is associated with clayey sandy soil and sandy clayey soil (i.e. JB 4 and 5, respectively), resulting in average growing conditions for crops.

4.2.2.5 Group 6 – Steepest Slopes

Group 6 are terrons 12, 25, and 38. The terrons in this group are the steepest areas within each of the national terrons. The soils are low in pH and have intermediate clay content. The first or second soil class in these areas are Podzols. Group 6 is not assigned a JB value because it mainly borders sloping areas of river valleys and is not associated with a particular soil type and not evaluated for crop production because these terrons represent a small portion all across Denmark.

4.2.2.6 Group 7 – Loamy Soils in Central and West

Group 7 are terrons 14, 18, 29, 34, and 39. These terrons have high levels of clay and larger root zone capacities. Both north and east national terrons have two terrons in this group, with terrons 18 and 34 having lower clay content and steeper slopes than terrons 14 and 39. The
first or second soil class in these areas are Luvisols and are associated with clayey sandy soil and
sandy clayey soil (i.e. JB 4 and 6, respectively). Group 7 is the second best for crop production
and compares well to the Danish soil fertility map of 1844 (Rothe, 1844).

4.3 Application Discussion

Identifying areas best suited for particular crops and monitoring areas with high
environmental risks is a necessity for Denmark since over 60% of the area is cultivated (Danish
Agriculture and Food Council, 2016). For this reason, Denmark was a strategic location for
applying the new workflow. Two major environmental concerns in Denmark are erosion and
pesticide leaching. Although relatively flat, erosion due to farming and wind erosion is a problem
for certain areas in Denmark (Vejhe et al., 2003). With drinking water being drawn from aquifers
and not being chemically treated before drinking, leaching is a major concern in Denmark.
Degradation products of chemicals being applied to fields have been detected in the groundwater
resulting in a threat to the quality of drinking water (Rosenbom et al., 2017). If terrons were
developed with variables known to cause erosion or leaching, the resulting terrons could be
organized by environmental risk severity to allow decisions concerning land management to
prioritize higher risk areas.

The concept of terroir has been established in Denmark for many crops, such as apples,
carrots, cheese, and herring (Mithril et al., 2012). Wine cultivation has also become a larger
activity in cold climates that have once been considered unsuitable. In particular, Gustafsson and
Mårtensson (2005) have shown the potential for extending wine production in Scandinavia. The
current wine regions in Denmark from Fødevarestyrelsen (2018) indicate there are four regions.
These regions appear to be administrative boundaries and do not pertain to regions with similar
soil, climate, and topography. The biggest region, Jutland, represents the largest differences in
landscape type (Madsen et al., 1992) and the largest range in soil textures (Adhikari et al.,
2013a); thus, combining Jutland into one wine region would have a wide variation in the
resulting products. Due to the incorporation of soil, climate, and topography in generating the
new regions, the national terrons would be a suitable alternative to the current wine regions.

Generating verbal descriptions for terrons is crucial if the terrons are to be utilized in the
future. Due to the large number of terrons, the terron groups were produced to simplify the map
and to be able to create more meaningful descriptions (Figure 8). Descriptions revolved around
determining how successful each group would be for crop production (Section 4.2.2.1 – 2.6).
Crop production was not determined for groups 1 and 6 since these groups represent wetlands
and steep slopes; both groups were not addressed because they represent small areas throughout
Denmark and are not highly utilized in crop production.

Validation of the terron groups was accomplished by using what is presumed the natural
growth potential of the soil. During the 1680’s, a land assessment was completed for all of
Denmark to determine the hard grain potential for taxation purposes (Pedersen, 1975; Madsen et
al., 1992). Each settlement was assigned a hard grain value based on the soil quality and higher
hard grain values were associated with higher yields. Hard grain values for over 12,000
settlements can be downloaded from HisKis (Dam, 2007); see Pedersen (1975) and Madsen et al.
(1992) for further understanding of King Christian V’s Great Danish Land Register of 1688.
Hard grain per hectare is mapped for each parish from 1688 in Figure 8. Hard grain values for
terron groups are averaged to compare the differences between natural growth potential of each
group (Table 4). All groups are statistically different when comparing average hard grain values
5 General Discussion

The importance of creating this workflow was to introduce the novelty of generating hierarchical terrons that can be easily reproducible. The simplicity of changing input data and receiving clustered maps will allow the work to be implemented by a wide range of researchers.

As seen from the case study, the workflow operates well when the input parameters are chosen wisely. The workflow can be easily repeated for any location, but two caveats need to be addressed: deciding on the number of clusters and variable selection.

The number of clusters selected will determine how the terron maps look. Using manual selection to choose the number requires previous knowledge of the location and potentially subjective decisions from experts. Objective decisions, based on statistical methods and accurate data, are beneficial in many situations, but expert opinion can be crucial in complex environmental systems (Krueger et al., 2012). Subjective decisions have been utilized in land-use decision-making and combined with objective modeling (Bantayan and Bishop, 1998). Manual selection can be useful to compare results to previous maps, as done in Carré and McBratney (2005). However, the input variables might have an optimal number of clusters that should be selected.

The variable selection affects the number of clusters in two main ways. First, as the number of variables to be clustered increases, the chance of having an optimal number of clusters decreases since a larger possibility of using irrelevant variables exists (Alelyani et al., 2013). Distinguishing between clusters with irrelevant variables is difficult due to the added noise.
Second, apparent clusters in n-dimensions might result in a different number of clusters in m-dimensions (Dy and Brodley, 2004). For example, clusters might be obvious when grouping five variables, but the clusters might not be obvious using either four or six variables. Thus, the number of clusters depends on the variables selected.

Automating the selection process for the number of clusters assumes an optimal number of clusters in the data. The target of the workflow is to reach an optimal number objectively, using the elbow method. However, the optimal number is subjective because other approaches may lead to different values. Besides the elbow method, another common approach to determine the optimal number when clustering is the silhouette method. The silhouette method measures the quality of a cluster by determining how well each object lies within the cluster (Rousseeuw, 1987). Tibshirani et al. (2001) demonstrated, using six different methods, that the estimated number of clusters is different depending on the method used. Incorporating multiple methods to quantify the optimal number of clusters would lead to difficulties in the automating process. Thus, only the elbow method was incorporated to simplify the workflow.

Variable selection depends on the terron purpose. If terrons are created for wine production, climate and soil have a large impact on taste (van Leeuwen et al., 2004), and extensive research has shown what affects wine flavor (salinity: Lanyon et al., 2004; water levels: White et al., 2007; soil depth, pH, and drainage: Burns, 2012). Knowing the physiology of a particular crop, variables can be selected to depict necessary components of growing conditions. For example, growing the grape vine (Vitis vinifera L.) in Europe, the warmest month should exceed a mean temperature of 18.9°C and coldest month should not drop below -1.1°C (Prescott, 1965). Thus, two variables reflecting mean monthly temperature during winter and summer is important. A variable reflecting solar radiation is crucial for crops, such as quinoa,
where daylength is an important factor (Oelke et al., 1992). Some crops are not as sensitive to temperature but require specific soil types for best growing conditions. Well-drained soils, such as deep sandy loams, are necessary for hops, but other crops need wetlands, such as wild rice (Oelke et al., 1992). This indicates that variable selection of soil properties is important for specific crops. The use of terrons is critical to consider before executing the workflow to determine the best variables for the application.

Noise in gridded data causes problems during clustering (Dave, 1991). The layers used in this study were mainly generated from interpolation (e.g. kriging), resulting in an uncertainty at each location that will accumulate the more layers used. Combining soil and landscape properties would result in transition zones between the unit boundaries, not sharp boundaries produced from hard clustering methods (de Bruin and Stein, 1998). Fuzzy sets have been incorporated in other classification techniques to deal with boundary transitions and to determine the degree a location is within all classes (remote sensing pattern recognition: Fischer and Benedikt, 1997; agroecozones: Liu and Samal, 2002). Fuzzy c-means assesses the vagueness of terron development through the membership map. Vagueness is linked to the challenges of delimiting the terron classes (Fischer and Benedikt, 1997). Thus, the membership map is crucial to determine how close each location is to being assigned the optimal terron. However, computation time drastically increases compared to k-means. Generating regional terrons in Denmark using k-means reduced computation time but lacks an associated membership map to evaluate how well each location is assigned to the right cluster.

Future work could improve the workflow for further applications. Currently, the workflow only incorporates one file type and one method for automatically deciding the optimal number of clusters. The workflow should incorporate all types of gridded data and more methods.
for automating the number of clusters, such as the silhouette method. Hierarchical terrons should be compared to terrons created at a single level to determine the differences generated at each level.

6 Conclusion

The workflow developed in this study generates hierarchical terrons at multiple scales from variables at different resolutions. The novelty of hierarchical terrons is a crucial aspect of this work and with climate being a new factor of terron modeling, the application of terrons to new regions will increase. Either to start defining terroir units for crops or helping deal with environmental risks, terrons are versatile and applicable in many areas. The workflow was developed to create terrons from gridded data and can be utilized in any field using gridded data for classification. In particular, users making agriculture and environmental risk decisions can apply this workflow for land management at different scales. The ability to determine fertile lands for a particular crop or areas prone to environmental risks using the same tool is important for cultivated lands. Overall, terron development is an informative approach to support land-use management decisions.

We used the workflow to make a generic terron classification using variables important for crop production in Denmark. The implementation of the workflow on Denmark produced a national terron map that could be used as the three new wine regions, as opposed to the current four wine regions based on administrative boundaries. The regional terrons can differentiate between wetlands, steep sloping areas, fertile soils, and poor soils. This study demonstrated that terrons from this workflow were useful in distinguishing the different regions in Denmark based on soil, climate, and landscape.
Acknowledgements

The study was supported by the ProvenanceDK Project with funding from the Danish Innovation Foundation.

Computer Code Availability

https://github.com/yroell/terron_workflow_code


10*10, 20*20 & 40*40 km temperatur og potentiell for dampning 20*20 & 40*40 km. Teknisk Rapport, Danish Meteorological Institute.


Environmental Science & Policy 6, 37–50.


Table 1. Characteristics of variables used for each level: national and regional.

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>VARIABLES</th>
<th>RANGES</th>
<th>UNITS</th>
<th>DESCRIPTION</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATIONAL</td>
<td>Frost</td>
<td>2.4 – 4.5</td>
<td>Days</td>
<td>Number of days below 0°C during April 1-15 and October 1-15</td>
<td>AGRO</td>
</tr>
<tr>
<td></td>
<td>Growing Days</td>
<td>146.8 – 158.9</td>
<td>Days</td>
<td>Number of days above 10°C during the year</td>
<td>AGRO</td>
</tr>
<tr>
<td></td>
<td>Precipitation</td>
<td>269.3 – 481.4</td>
<td>Millimeters</td>
<td>Precipitation between April 1 and October 31 in millimeters</td>
<td>Scharling 2000</td>
</tr>
<tr>
<td></td>
<td>Solar Radiation</td>
<td>2970 –</td>
<td>MJ/m²</td>
<td>Global solar radiation between April 1 and October 31 in MJ/m²</td>
<td>AGRO</td>
</tr>
<tr>
<td></td>
<td>Principal Component 1</td>
<td>-0.09 –</td>
<td></td>
<td>First principal component from topsoil spectra data</td>
<td>Knadel et al. 2013</td>
</tr>
<tr>
<td></td>
<td>Principal Component 2</td>
<td>-0.08 –</td>
<td></td>
<td>Second principal component from topsoil spectra data</td>
<td>Knadel et al. 2013</td>
</tr>
<tr>
<td>REGIONAL</td>
<td>Clay</td>
<td>0 – 57.1</td>
<td>Percent</td>
<td>Percent clay in soil between 100-200 cm</td>
<td>Adhikari et al. 2013a</td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>2.4 – 9.6</td>
<td></td>
<td>pH in soil between 60-100 cm</td>
<td>Adhikari et al. 2014a</td>
</tr>
<tr>
<td></td>
<td>Carbon</td>
<td>0 – 60.9</td>
<td>g/kg</td>
<td>Organic carbon content between 0-30 cm in g/kg</td>
<td>Adhikari et al. 2014b</td>
</tr>
<tr>
<td></td>
<td>Root Zone Capacity</td>
<td>9.2 – 1581.7</td>
<td>Millimeters</td>
<td>Weighted sum of rooting depths using 5 depths in millimeters</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0 – 90</td>
<td>Degrees</td>
<td>Slope generated from a 30.4 m digital elevation model in degrees</td>
<td>Adhikari et al. 2013b</td>
</tr>
<tr>
<td></td>
<td>Saga Wetness Index</td>
<td>6.9 – 19.1</td>
<td></td>
<td>Saga WI generated from 30.4 m digital elevation model</td>
<td>Adhikari et al. 2013b</td>
</tr>
</tbody>
</table>

1 = raw data from weather station reference and used kriging on points to obtain map

2 = obtained digital elevation model from Adhikari et al. 2013b and generated slope and Saga WI maps
Table 2. Percentile (25th to 75th) values for national variables by national terron (North, East, or West). Precip is precipitation and TS Clay and TS C. Sand are topsoil (0 – 30 cm) clay and coarse sand, respectively. Area is percentage of area that terron occupies in Denmark.

<table>
<thead>
<tr>
<th>TERRON</th>
<th>FROST</th>
<th>GROWING DAYS</th>
<th>PRECIP</th>
<th>SOLAR</th>
<th>PC 1</th>
<th>PC 2</th>
<th>TS C. SAND</th>
<th>TS CLAY</th>
<th>AREA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORTH (1)</td>
<td>3.5 –</td>
<td>147.4 –</td>
<td>367.2 –</td>
<td>3022 –</td>
<td>-0.008 –</td>
<td>-0.001 –</td>
<td>26.2 –</td>
<td>3.4 –</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>148.3</td>
<td>414.3</td>
<td>3103</td>
<td>0.007</td>
<td>0.024</td>
<td>48.8</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>WEST (2)</td>
<td>3.4 –</td>
<td>150.4 –</td>
<td>431.5 –</td>
<td>3001 –</td>
<td>-0.005 –</td>
<td>-0.008 –</td>
<td>42.4 –</td>
<td>3.0 –</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>153.8</td>
<td>461.1</td>
<td>3038</td>
<td>0.033</td>
<td>0.012</td>
<td>64.0</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>EAST (3)</td>
<td>2.9 –</td>
<td>154.4 –</td>
<td>322.9 –</td>
<td>3096 –</td>
<td>-0.024 –</td>
<td>-0.029 –</td>
<td>24.9 –</td>
<td>9.0 –</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>156.5</td>
<td>352.7</td>
<td>3162</td>
<td>-0.11</td>
<td>-0.017</td>
<td>36.9</td>
<td>13.9</td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Percentile (25th to 75th) values for regional variables by unique regional terron value.

Carb is organic carbon and soil class 1 and class 2 are most prevalent soil classes in each terron with the percentage they represent in parentheses. Area is percentage of area that terron occupies in each national terron.
<table>
<thead>
<tr>
<th>TERRON</th>
<th>CLAY</th>
<th>PH</th>
<th>CARB</th>
<th>ROOT ZONE</th>
<th>SOIL CLASS 1</th>
<th>ARE ( % )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>10.4-14.4</td>
<td>5.71-6.13</td>
<td>1.7-</td>
<td>2.4</td>
<td>216.1-297.6</td>
<td>Luvisols (31.5)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>3.8- 8.8</td>
<td>5.23-5.85</td>
<td>1.3-</td>
<td>2.0</td>
<td>133.6-216.6</td>
<td>Podzols (58.6)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>6.3-14.5</td>
<td>5.28-5.99</td>
<td>15.9-</td>
<td>27.4</td>
<td>1581.7-1581.7</td>
<td>Histosols (63.8)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>17.2-21.8</td>
<td>5.97-6.80</td>
<td>1.6-</td>
<td>2.2</td>
<td>260.8-306.4</td>
<td>Luvisols (67.2)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.0-13.4</td>
<td>5.48-6.66</td>
<td>1.7-</td>
<td>5.3</td>
<td>1581.7-1581.7</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>0-5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1581.7</td>
<td>1581.7</td>
<td>5.48-6.66</td>
<td>1.7-</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Podzols (21.7)</td>
<td>0.7</td>
<td>Arenosols (16.7)</td>
<td>1.2</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Podzols (70.7)</td>
<td>0</td>
<td>Arenosols (70.6)</td>
<td>1.3</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>228.4-291.3</td>
<td>1.9</td>
<td>Luvosols (43.2)</td>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>Podozols (17.1)</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
<td>Arenosols (37.6)</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>2.4-7.6</td>
<td>5.34-5.59</td>
<td>2.9-9.5</td>
<td>1581.7-1581.7</td>
<td>Podzols (63.0)</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>---------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>8.4-12.3</td>
<td>5.41-5.67</td>
<td>1.9-2.8</td>
<td>153.4-195.9</td>
<td>Podzols (65.7)</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1.7-4.1</td>
<td>6.76-7.25</td>
<td>0.5-1.5</td>
<td>30.0-96.9</td>
<td>Arenosols (77.8)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2.0-4.2</td>
<td>5.32-5.53</td>
<td>1.7-2.5</td>
<td>67.4-129.7</td>
<td>Podzols (87.3)</td>
<td></td>
</tr>
<tr>
<td>Layer</td>
<td>Depth</td>
<td>Pfp</td>
<td>pH</td>
<td>EC</td>
<td>TSM</td>
<td>Type</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>--------------------------</td>
</tr>
<tr>
<td>25</td>
<td>6.6-13.9</td>
<td>5.40-6.01</td>
<td>1.2-</td>
<td>1.6</td>
<td>134.7-243.5</td>
<td>Luvisols (33.4)</td>
</tr>
<tr>
<td>26</td>
<td>2.4-5.7</td>
<td>5.26-5.61</td>
<td>1.3-</td>
<td>2.1</td>
<td>75.7-137.7</td>
<td>Podzols (77.6)</td>
</tr>
<tr>
<td>27</td>
<td>14.4-22.3</td>
<td>5.39-6.03</td>
<td>2.7-</td>
<td>6.0</td>
<td>1581.7-1581.7</td>
<td>Fluvisols (44.8)</td>
</tr>
<tr>
<td>28</td>
<td>9.8-14.0</td>
<td>5.57-5.99</td>
<td>1.5-</td>
<td>2.1</td>
<td>162.2-240.4</td>
<td>Luvisols (35.2)</td>
</tr>
<tr>
<td>29</td>
<td>18.0-22.5</td>
<td>5.80-6.26</td>
<td>1.4-</td>
<td>1.9</td>
<td>234.7-271.8</td>
<td>Luvisols (78.2)</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5-16.8</td>
<td>8.8-16.2</td>
<td>17.6-21.0</td>
<td>10.1-16.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.74-8.10</td>
<td>6.28-7.64</td>
<td>7.46-7.96</td>
<td>6.22-6.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3-</td>
<td>3.5-</td>
<td>1.3-</td>
<td>1.2-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>13.3</td>
<td>1.6</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>265.4-307.7</td>
<td>1581.7-1581.7</td>
<td>266.5-299.8</td>
<td>231.4-279.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luvisols (65.0)</td>
<td>Luvisols (29.2)</td>
<td>Luvisols (76.2)</td>
<td>Luvisols (68.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C m b i s o l s (7 4 )</td>
<td>G l e y s o l s (8 5 )</td>
<td>C m b i s o l s (6 9 )</td>
<td>A r e n o s o l s (8 5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>35</td>
<td>10.5-14.4</td>
<td>7.40-7.93</td>
<td>1.2-</td>
<td>1.5</td>
<td>252.2-288.2</td>
<td>Luvisols (74.4)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>.</td>
<td>4</td>
<td>7</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>.</td>
<td>2</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>.</td>
<td>4</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>36</td>
<td>2.3- 6.9</td>
<td>6.65-7.52</td>
<td>1.2-</td>
<td>1.8</td>
<td>159.2-268.5</td>
<td>Luvisols (41.1)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>.</td>
<td>4</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>.</td>
<td>8</td>
<td>3</td>
<td>.</td>
</tr>
<tr>
<td>37</td>
<td>9.9-14.4</td>
<td>6.06-6.53</td>
<td>1.3-</td>
<td>1.9</td>
<td>231.7-283.3</td>
<td>Luvisols (49.1)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>.</td>
<td>6</td>
<td>8</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>.</td>
<td>1</td>
<td>4</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>.</td>
<td>4</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td>38</td>
<td>8.8-17.4</td>
<td>4.18-5.18</td>
<td>1.2-</td>
<td>1.7</td>
<td>184.4-274.8</td>
<td>Luvisols (37.0)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>5</td>
<td>9</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>.</td>
<td>4</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>.</td>
<td>4</td>
<td>7</td>
<td>.</td>
</tr>
<tr>
<td>39</td>
<td>17.9-22.0</td>
<td>6.15-6.64</td>
<td>1.3-</td>
<td>1.7</td>
<td>249.8-286.9</td>
<td>Luvisols (73.1)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>.</td>
<td>7</td>
<td>7</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>.</td>
<td>7</td>
<td>5</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-</td>
<td>.</td>
<td>2</td>
<td>-</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>ls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>(</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.</td>
<td>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

696

697
Table 4. Mean values and standard deviation of hard grain value by terron groups. Area is percentage of area that terron group occupies in Denmark. All groups are statistically different (one-way ANOVA; Tukey post hoc test: p value < 0.001).

<table>
<thead>
<tr>
<th>Terron Group</th>
<th>Description</th>
<th>Hard Grain Value (Hard Grain / Hectare)</th>
<th>Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wetland</td>
<td>0.075 ± 0.059</td>
<td>16.1</td>
</tr>
<tr>
<td>2</td>
<td>Loamy Soil in East</td>
<td>0.174 ± 0.072</td>
<td>16.5</td>
</tr>
<tr>
<td>3</td>
<td>Loamy Sand</td>
<td>0.098 ± 0.073</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>Fluvial Sands in West</td>
<td>0.038 ± 0.032</td>
<td>22.6</td>
</tr>
<tr>
<td>5</td>
<td>Sandy Loam</td>
<td>0.085 ± 0.062</td>
<td>17.4</td>
</tr>
<tr>
<td>6</td>
<td>Steepest slopes</td>
<td>0.063 ± 0.050</td>
<td>6.4</td>
</tr>
<tr>
<td>7</td>
<td>Loamy Soil in Central and West</td>
<td>0.117 ± 0.060</td>
<td>15.6</td>
</tr>
</tbody>
</table>
Figure 1. Flowchart of hierarchical terron development.
Figure 2. Maps of six normalized variables used for clustering national terrons. A) Average number of days below 0°C during April 1 – 15 and October 1 – 15; B) Average number of days above 10°C during a year; C) Precipitation between April 1 and October 31; D) Average solar radiation between April 1 and October 31; E) Principal component 1 from topsoil spectra; F) Principal component 2 from topsoil spectra.
Figure 3. Within-clusters sum of square from k-means clustering analysis on national variables. Dashed line indicates value chosen as optimal number of cluster for national variables. Circles represent within-cluster dispersion using actual data. Triangles indicate maximum difference between within-cluster dispersion from the actual data and randomized data.
Figure 4. Maps of six normalized variables used for clustering regional terrons. A) Percent clay at a depth of 100 – 200 cm; B) pH of soil at a depth of 60 – 100 cm; C) Organic carbon content at a depth of 0 – 30 cm; D) Root Zone Capacity; E) Slope; F) Saga Wetness Index.
Figure 5. A) Map of national terrons determined by using fuzzy c-means and B) Map of highest membership of each location with a value of 1 indicating complete member to a terron.
Figure 6. Map of nine regional terrons in each national terron for a total of 27 regional terrons.

USE COLOR FOR FIGURE 6

2 COLUMN FITTING IMAGE
Figure 7. Dendrogram of all regional terrons (27 in total). A unique regional terron value is a combination of national terron (North = 1, West = 2, or East = 3) and regional terron (1 to 9).

1 COLUMN FITTING IMAGE
Figure 8. Map of terrons grouped by similarities and map of hard grain value by parish level from 1688 used to validate the terron groups.
Highlights

- A workflow was developed to create and map hierarchical terrons at multiple scales.
- This workflow was applied to Denmark to create 3 national and 27 regional terrons.
- Terrons generated were able to distinguish between fertile and non-fertile lands.
- Wetland areas were identified by one terron showing the workflow versatility.
- Grouped terrons were validated using yield potential values from 1688.
Authorship statements:

Yannik E. Roell collected and analyzed data, designed the analysis, and wrote the manuscript.

Yi Peng contributed to analyzing data and critically revised the manuscript.

Amélie Beucher contributed to analyzing data and critically revised the manuscript.

Mette B. Greve contributed to collecting data and critically revised the manuscript.

Mogens H. Greve supervised research project and critically revised the manuscript.

All authors give their final approval of the manuscript version to be submitted.

Declarations of interest: none

No conflict of interest.