Groundwater Pollution in Denmark

Spatial Analytics Exam 2021

Emil Trenckner Jessen
AU ID: au604547

Johan Kresten Horsmans
AU ID: au618771

School of Communication and Culture,
Aarhus University
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

10^{th} of June, 2021

Abstract

The quality of Denmark's groundwater supplies is under threat due to the use of nitrate in agriculture. In the last few years there has been a cultural shift in public opinion, where conventional agriculture is progressively being seen as more damaging to the environment, compared to its organic counterpart. We wish to assess the scientific validity of this belief. Using kriging, linear modeling and coefficient t-tests, we carry out a spatial analysis comparing nitrate concentration levels in Denmark for conventional- and organic agricultural farming. Here, we find significantly higher nitrate concentrations for organic fields, with organic- and conventional fields having mean nitrate concentrations of 24.51 mg/L and 23.48 mg/L, respectively. Followingly, we discuss potential confounding factors in our analysis and their impact on our results. Furthermore, follow-up questions and prospects for relevant future research are discussed.

Keywords: Nitrate Pollution, Agriculture, Geospatial Analytics, Kriging, Denmark, Groundwater Quality, RStudio.

Character count: 27,281

Regarding initials and code:

ETJ stands for Emil Trenckner Jessen
JKH stands for Johan Kresten Horsmans

This PDF contains the report for our exam in Spatial Analytics. All required sections are marked in relation to who has been the main contributor. Nevertheless, it is important to underline that we regard the assignment as a completely collaborative effort since each section has been written and extensively edited by both authors. Moreover, the coding has been produced by both authors in tandem.

For the accompanying R-code, please refer to the main GitHub repository at: https://github.com/emiltj/groundwater_pollution_dk. Here you will find a knitted .md-file with all the code and produced outputs.

To replicate our analysis, you need to go through the following steps to download the needed data and scripts.

NOTE: The code below will only work for RStudio.

1. Clone the following repository to a directory of your own choice: https://github.com/emiltj/groundwater_pollution_dk.
2. Run the following commands in your bash terminal from RStudio:

NOTE: There may be slight variations depending on the terminal and operating system you use. The following example is designed for *Git Bash* on *Windows 10*:

```
cd {directory\where\repository\has\been\cloned}
bash data_download.sh
```

You are now ready to run the code.

Contents

1 Introduction

1.1 Nitrate, water quality and health

Historically, the Danish groundwater has been readily drinkable without much processing and has generally been of high quality. However, the quality of the underground water has been under threat from the agricultural sector, due to its great use of nitrate for fertilization of the fields (Di Cameron, 2002). Although heavily used in agriculture, nitrate is also a naturally occurring compound produced during decomposition of organic materials, typically at terrain level. During periods of precipitation, nitrate is leached through the soil and eventually reaches and diffuses into bodies of groundwater (Myrold Tiedje, 1985). The time period between nitrate infusion at ground level and the diffusion of nitrate into groundwater reservoirs might be of up to 20 years but is greatly dependent on various factors such as precipitation, depth of groundwater reservoirs, soil density and soil type. The majority of the nitrate is, as time progresses, however, transformed into other substances, through the naturally occurring process of denitrification. This means that nitrate concentrations decrease at depth (Knowles, 1983; WHO, 1993). The nitrate content in groundwater is exceedingly small in areas of wild, uncultivated nature. However, given human interference, areas may be exposed to unnaturally high nitrate concentrations at terrain level. In such instances, the process of denitrification does not adequately transform nitrate and the groundwater may end up with high levels of nitrate. If the EU regulations of 50 mg/L are exceeded, the water is said to be contaminated and consumption of such water may cause illnesses such as methemoglobinemia that may end up being fatal (EEC, 1991; Ward et al., 2005).

1.2 Agricultural nitrate pollution and public opinion

The agricultural industry is the largest contributor to nitrate pollution of the environment and groundwater, due to its use of fertilizers that are essential to the growth of crops (Lindinger Scheidleder, 2004; Wick et al., 2012). As a result of the nitrate pollution from farming, Denmark and large parts of Europe have groundwater with nitrate levels exceeding the limitations of the European Union (EU) (Hansen et al., 2012; Strebel et al., 1989). This poses a drinking water problem in both the near and distant future. Farming is usually divided up into conventional– and organic farming. The two differ in their use of pesticides and fertilization, with organic farming avoiding the use of pesticides and resorting to livestock manure for fertilization as opposed to inorganic fertilizers produced artificially. The literature suggests that organic farming has certain environmental advantages over conventional farming. Furthermore, organic farming yields less of a decrease in biodiversity compared to conventional farming while also diminishing the threat that pesticides pose to ecosystems and to human health (Fuller et al., 2005; Sánchez-Bayo et al., 2002).

The consumer demand for organic crops in Denmark has seen a rapid increase during the last few years (Aaberg, 2020). The recent popularity spurt has largely been due to a cultural shift in public opinion towards being more environmentally aware. This cultural trend has been further solidified by the marketing angle of organic products (Aaberg, 2020; Bezawada Pauwels, 2013; Pearson Henryks, 2008; Willer Lernoud, 2019). Apart from a tendency to highlight an increase in biodiversity, Some organic trade associations and nature conservation organizations also seem to emphasize a positive effect in terms of nitrate pollution (Danmarks Naturfredningsforening, 2021; Ladegaard, 2019). While the immediate benefits in terms of biodiversity are irrefutable, evidence for less nitrate pollution seems to be unclear, with research pointing in different directions (Danmarks Naturfredningsforening, 2021; Raadet for Groen Omstilling, n.d.) Regardless of whether fertilizers have been artificially produced or not, they largely consist of nitrate. As such, both types of agricultural farming omit nitrate that eventually leaches into the soil and diffuses into groundwater bodies.

1.3 Motivation and aim of the paper

Given knowledge of the health detriment that nitrate pollution from farms induces, we deem it compelling and relevant to investigate the relationship between organic- and conventional agricultural land use and nitrate pollution. Apart from the strictly environmental relevance, it is also interesting to investigate from a cultural perspective. When seen from this point of view, it is worth knowing whether the public opinion on organic agriculture as the environmentally safe alternative really holds when it comes to nitrate pollution. Especially given knowledge that some marketing bureaus and organizations emphasize reduced nitrate pollution, although the research on the subject appears inconclusive.

With this in mind, we specifically want to investigate the following question:

What is the relationship between nitrate levels and type of land use (conventional vs. organic farming)?

To assess this question, we first use existing nitrate concentration measurements to estimate nitrate concentrations in all of Denmark. This interpolation is carried out through the use of universal kriging. Using existing spatial data, we overlay layers of conventional- and organic fields and extract nitrate concentrations from the respective field types. Subsequently, we use linear modeling to test for nitrate differences between organic and conventional farming and to look at what these differences are. Finally, we discuss and scrutinize our results and methodological choices to shed light on potential confounding factors and to assess the ecological validity of our spatial analysis.

2 Methods

2.1 Software framework

To answer our research question, we utilized a variety of freely available software tools. First and foremost, the entire analysis framework was carried out using R 4.0.3 and RStudio 1.2.1335 (R Core Team, 2020). For working with spatial data in both vector- and raster format, we used the following packages: *sf*, *raster*, *sp*, *dplyr*, *lubridate* and *tidyverse* (Hijmans et al., 2015, 2015; Pebesma et al., 2012; Spinu, 2016; Wickham et al., 2019; Wickham Wickham, 2020). These packages were utilized to load, preprocess and transform the data throughout the analysis. For our statistical modeling, we utilized *gstat* (Pebesma et al., 2015) for variogram modeling and kriging and *lmtest* (Hothorn et al., 2015) for regression coefficient t-testing. All plotting was done with a mix of *raster*, *ggplot2*, *ggthemes* and *tmap* (Arnold Arnold, 2015; Hijmans et al., 2015; Tennekes, 2018; Wickham, 2011, p. 2).

All the code utilized for our analysis can be found in groundwater_pollution_dk.md in the main GitHub repository for the analysis. Here you will also find in-depth commenting for each step throughout the entirety of the employed code. The code was run and written on a Dell XPS-15, 16 Gb RAM, i7-6700HQ CPU @ 2.60 GHz running Windows 10, 64 bit. If you wish to replicate our analysis, the commented R-markdown code can be found here: groundwater_pollution_dk.Rmd. To run the code, you will need to download the supplied data to your working directory. This can be done by cloning the main GitHub repository and then running the data_download.sh script from a bash terminal. For more information on the structure of the supplied GitHub repository and how to run the code, please see the main README.

2.2 Data acquisition

The data used in the analysis is all publicly available and has been acquired through the online portals of the respective ministries and software companies that are in possession of the data. Note that the portals' links and data availability may be deprecated in the future. However, in such instances, access to the data can nonetheless still be achieved through our script data_download.sh located in our Github repository. The four datasets that were used in our analysis are:

1. Point data with samples of nitrate levels in Denmark. This dataset contains 14,350 measurements of nitrate concentrations at different geographic locations in Denmark from 1900 to March 2021. The dataset was provided by courtesy of *De Nationale Geologiske Undersøgelser for Danmark og Grønland* (GEUS). The included variables used in our analysis are: coordinates, measurement date, nitrate concentration and measurement depth. The data was retrieved from their webpage (GEUS, n.d.).

2. A shapefile containing polygons of all current agricultural fields in Denmark as of April 2021. This dataset is provided by the *Ministry of Food, Agriculture and Fisheries of Denmark* (Danish Agricultural Agency). The dataset was retrieved from their webpage (LBST, n.d.).

3. Shapefiles containing polygons of all organic agricultural fields in Denmark, registered each year between the period of 2012 to 2020. This dataset is provided by the Danish Agricultural Agency. The dataset was supposed to be publicly available through the Danish Agricultural Agency's website, but since this was not the case, we were sent the following link to the dataset in a mail correspondence with the agency (LBST, n.d.).

4. A shapefile containing a polygon in the shape of Denmark, courtesy of the software company *IGIS MAP*. The shapefile was retrieved from their webpage (IGIS MAP, n.d.).

2.3 Processing

The following section details the processing steps carried out in our analysis with an emphasis on decision-making bottlenecks. For the exact functions utilized, please refer to the groundwater_pollution_dk.md-file found in the main GitHub repository.

We started by loading the nitrate data as a dataframe and renaming the relevant columns. We then proceeded to remove outliers from the data. We discarded all measures with a nitrate concentration above 200 mg/L and, furthermore, also all measures with a measurement depth below 0 (i.e. above ground level). The negative depths were removed since they were obvious measurement errors. The measurements with a concentration of nitrate above 200 mg/L were removed since the literature suggests that such concentrations are only found in the case of severe accidents (Schullehner et al., 2014). Therefore, we deemed that these high-concentration measurements would be confounding noise in our analysis since the accidental nitrate spills would cloud our investigation of the systemic differences in agricultural land use. Regular expressions were employed to remove excess non-coordinate characters from the coordinate column, which was then split into longitude- and latitude coordinates. Using the preprocessed coordinates, the nitrate data was converted to an '*sf*'-object.

We proceeded to specify the *Coordinate Reference System* (CRS). The CRS in question is the map projection ETRS89, UTM zone 32 N (EPSG: 25832). ETRS89 refers to the geodetic reference system and is quite precise within the boundaries of Europe (Altamimi, 2018). UTM stands for Universal Transverse

Mercator and refers to the map projection system, which in this case is a cylindrical secant projection. The distances are therefore slightly distorted, but within the Danish borders only up to a maximum of 40 cm per km (Geodætisk Institut, 2020). The data was subsequently filtered to only contain measurements between 2012 and 2020 so it would only contain data from the same time period as our organic fields. All points with duplicate geometries were removed. For a graphic depiction of the nitrate data after the described processing, see *figure 1*:

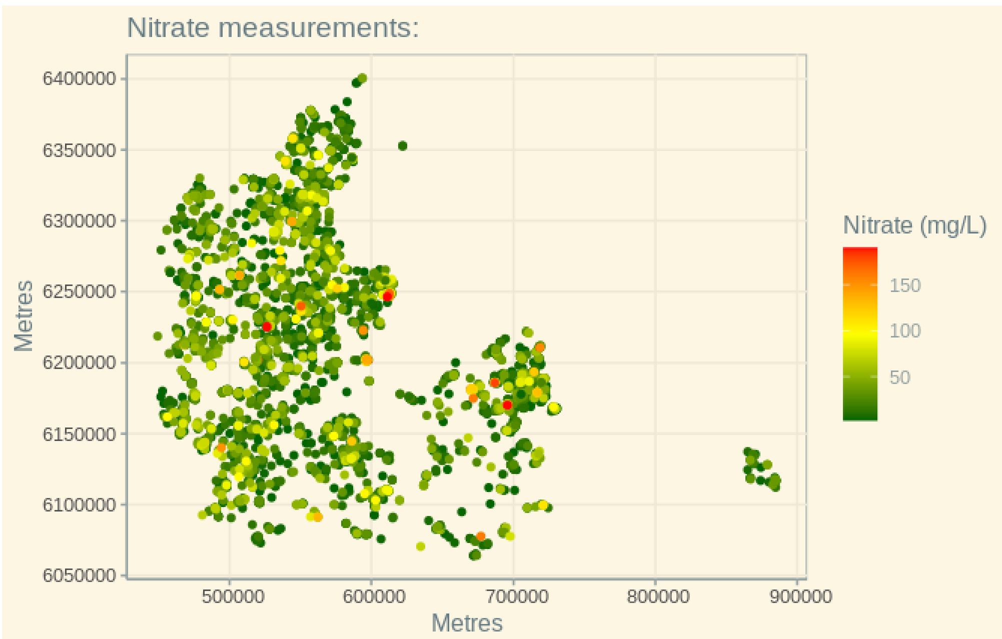


Figure 1: Measurements of nitrate concentration (mg/L).

All files with agricultural field polygons were loaded. The datasets had the same CRS as our nitrate data, but for the sake of good practice, we explicitly set the CRS as EPSG: 25832. The organic field data consisted of polygons of all the new organic fields registered between 2012 and 2020. The geometries of these individual spatial polygons were merged into a single '*SpatialPolygons*'-object, named "*organic_fields*" containing 12,560 polygons. The file containing all field polygons was also converted to a 'SpatialPolygons'-object named "*all_fields*" containing 470,000 polygons. Followingly, a new variable called "*conventional_fields*" was defined, where the polygons in "all_fields" that overlapped with the polygons in "organic_fields" were removed. A total of 9,749 polygons were discarded in this process. Since there were 12,560 organic fields in our dataset, one would expect that a total of 12,560 fields would be removed. Since the dataset that contained all fields was incomplete, this was not the case. For plots of the conventional- and organic fields, see appendix *figure A* and *figure B*, respectively. To reduce computational load and processing time, the "conventional_fields"-object was subset to contain as many polygons as "organic_fields". We did this by sampling 12,560 random polygons from the object.

2.4 Data exploration - measurement depth

It is clear in the literature that nitrate concentrations in samples are highly dependent on the measurement depth, as nitrate transforms into other organic matters when it seeps into the ground (Knowles, 1983). As such, given two fields with the same amount of nitrate pollution, we might see differences in pollution in our data, due to the nitrate samples having different measurement depths. As our research interest lies in the relationship between nitrate concentration and agricultural land use rather than measurement depth of samples, we wanted to take this effect into account. We investigated the possibility of removing the effect by modeling out the variance that measurement depth explains in our data. This could be implemented by adding a penalty to the nitrate concentration based on measurement depth - i.e. samples near terrain level would be applied a small penalty, while deep samples would be penalized harder. To make an informed decision on whether to include a depth penalty or not, we decided to inspect the data:

We started by creating separate columns for the nitrate measurements' X- and Y coordinates and appending the measurement coordinates to a matrix which was then converted to a "SpatialPoints"-object. Variables containing all coordinates for conventional- and organic field polygons were created. Variables called "conventional_overlap" and "organic_overlap" were created with a function that went through all points and polygons in the polygon lists. These variables were then processed using various functions to make them indexable. The points overlapping with organic polygons were assigned the class "organic". Oppositely, the points overlapping with conventional polygons were assigned the class "conventional". The rest of the points were assigned the class "out of bounds". We then created a density plot with the three categories (see *figure 2*):

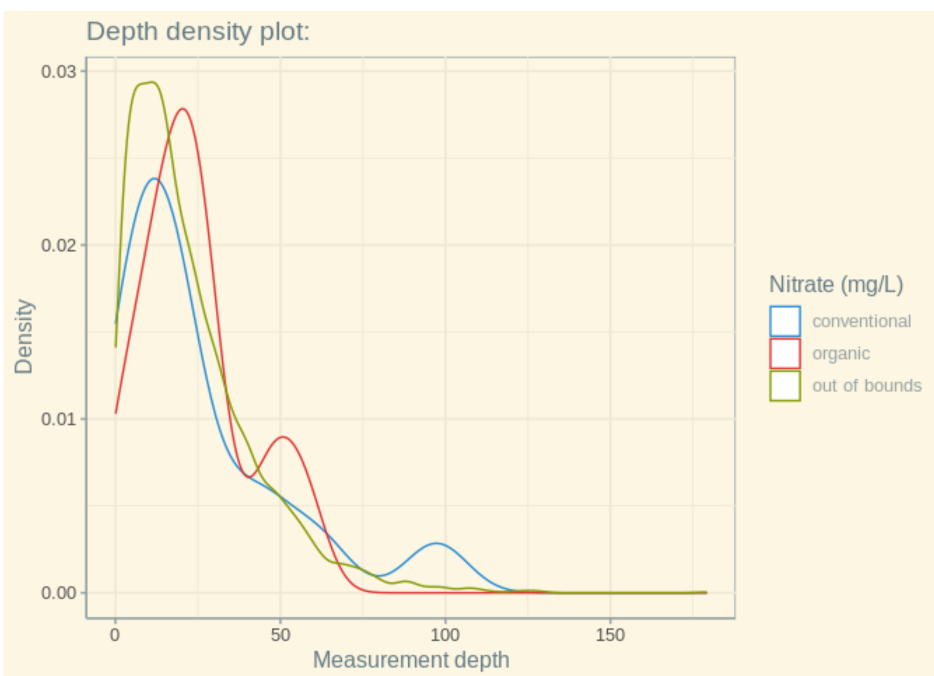


Figure 2: Density plot of measurement depths for different agricultural land types.

Upon investigating the distribution of measurement depths between samples within the different field-types, it appears that there is no relationship between measurement depth and field type. In other words, the variance in measurement depth of nitrate samples is equal across all land types and should thus not be a confounding factor in our analysis. Following our exploration of measurement depth, we decided not to include a depth penalty.

2.5 Kriging

In our depth analysis, it became clear that a small portion of nitrate measurements fell within our fields. As such, in order to carry out the analysis, an interpolated surface with nitrate concentration values for the entirety of Denmark was desired. To create such a surface, we utilized a spatial procedure known as kriging. The kriging procedure works by modeling the correlation between the variance for a given variable and the spatial scatter of its measurements (ArcGIS, 2021). In our case, we wanted to base our kriging on the amount of variance that was shared in nitrate concentration measurements depending on their spatial proximity. We argue that this is a fitting method for our data since it is a reasonable assumption that measurements close to each other should share more similar nitrate concentrations compared to measurements further away. To assess this assumption, we first used our extracted X- and Y coordinates for the nitrate measurements to fit a spatial variogram to the nitrate concentration data with the following formula (written in pseudocode):

$$nitrate\ concentration \sim X + Y$$

For the produced variogram, see *figure 3*:

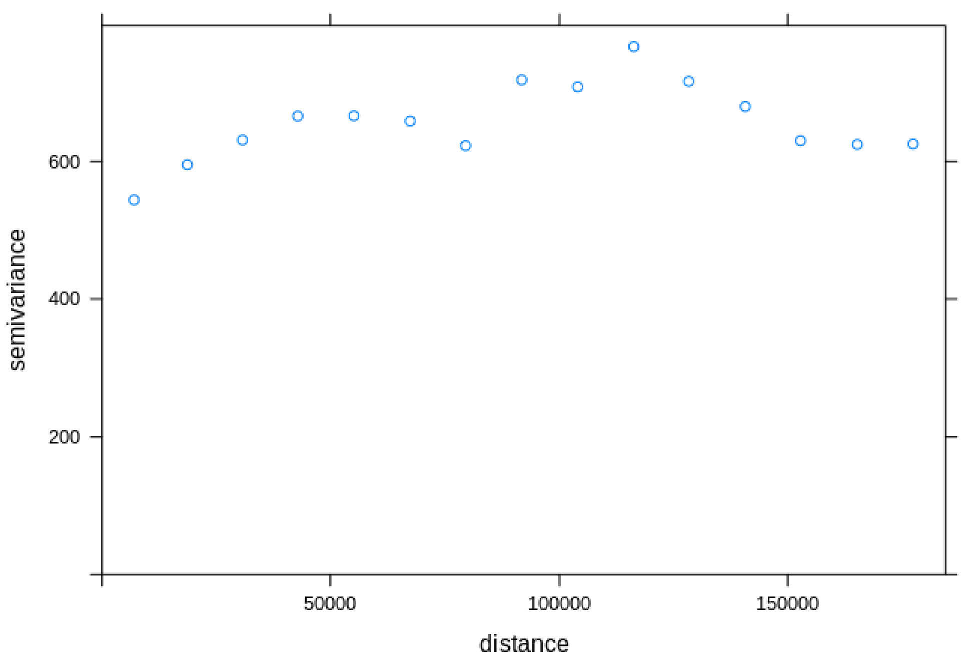


Figure 3: Variogram fitted to the nitrate concentration data.

As seen in figure 3, it appears that it is quite hard to find clear spatial correlations in the data. Nonetheless, it appears that measurements within 55,000 meters of each other are correlated, while points at greater distances seem uncorrelated. The fluctuations in variance between points of greater distances (between 60000 and 150000) seem to be without a pattern, leading us to believe that these fluctuations are from random noise. As such, we argue that our data can be used for kriging. Using the variogram, we estimated the following parameters that were to be used for fitting a variogram-model to the data:

Parameter	**Description**	**Estimated value**
Nugget	The y-axis intercept for the variogram	550
Partial sill (psill)	The ceiling (peak semi-variance before the spatial correlation flattens) minus the nugget	70
Range	The distance after which there is no longer any spatial correlation	55,000 metres (55 km)

We fitted a model to the variogram using Matern, M. Stein's parameterization (Zhu Stein, 2006). For a graphic depiction of model fit, see *figure 4*:

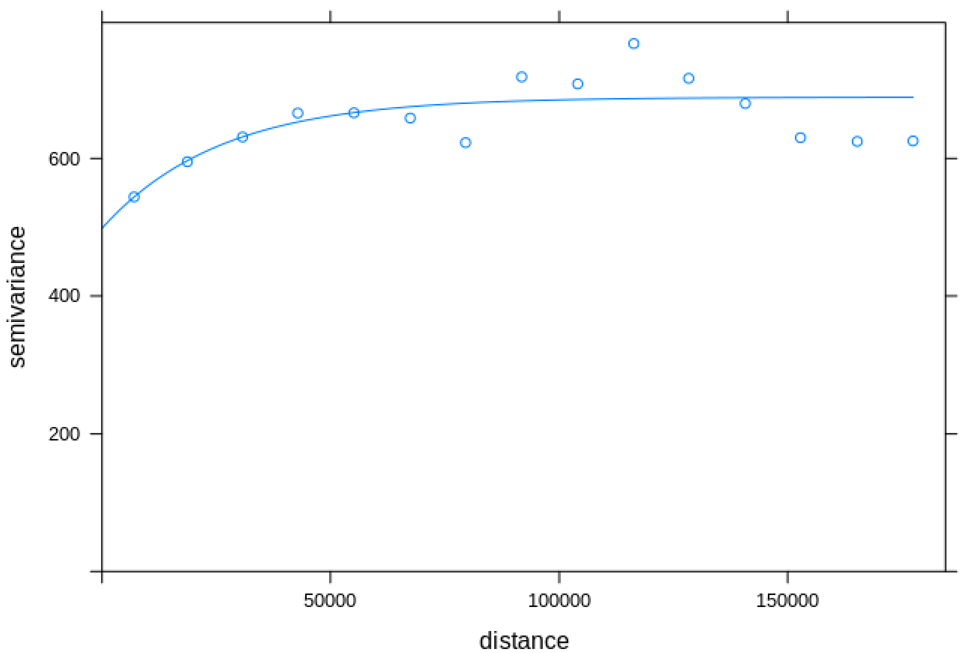


Figure 4: The model fitted to the variogram.

For kriging purposes, we defined a spatial grid that was to be used for creating a raster with interpolated values. We specified that the grid should consist of 44,840 raster cells (i.e. interpolated values), each 2 km in width and height. For visualization purposes, we manually defined that the extent of the grid should be slightly larger than the size of Denmark. We used universal kriging for interpolating the data. The reason for using universal kriging, as opposed to ordinary kriging, is that we have a scientific justification for saying that there is a trend in our data biased towards higher nitrate concentrations surrounding agricultural fields. Universal kriging is known to perform better on datasets with a general trend, such as ours (ArcGIS, 2021). The following formula was specified (written in pseudocode):

kriged surface = nitrate concentration ∼ X + Y, grid = spatial grid with 44,840 cells, model = fitted variogram model

Following the kriging, we now had a raster with values for our entire grid. We wanted to crop the raster to only include interpolated values placed on land since there obviously is no farming in the ocean. To do this, we loaded our retrieved map of Denmark which was transformed to the correct CRS. We cropped the kriged raster and were left with the data seen in *figure 5*:

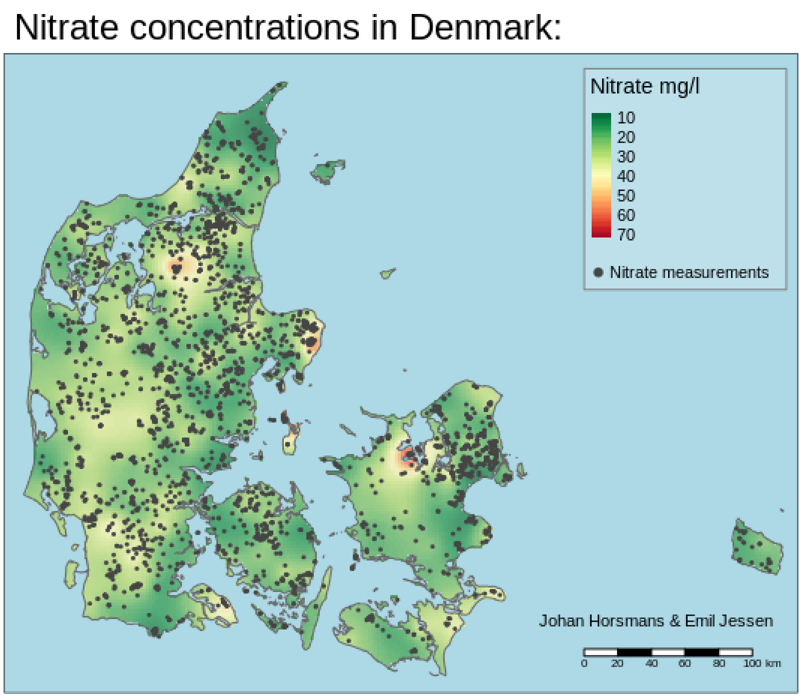


Figure 5: Interpolated nitrate concentrations in Denmark superimposed with the geographical locations of "real" nitrate measurements.

3 Analysis

Following our data processing, we had a raster with interpolated nitrate concentration values and multiple polygons with agricultural fields categorized as either organic- or conventional fields. As such, we were ready to carry out a statistical analysis capable of answering our main research question.

With our interpolated raster, we started by calculating the mean nitrate concentration for all conventional- and agricultural fields, respectively. Once calculated, the means were categorized as either "organic" or "conventional" and merged into a common dataframe. We then fitted the following linear regression model to the data (written in pseudocode):

$$mean\ nitrate\ concentration \sim land\ type$$

Following this, we did a t-test of regression coefficients to assess the difference between land types. The results can be seen in the section *results*.

4 Results

	Estimate	**Std. Error**	**t-value**	**p-value**
Intercept (conventional farming)	23.484705	0.058407	402.085	2.2e-16 (***)
Slope (organic farming)	1.023951	0.082600	12.396	2.2e-16 (***)

5 Discussion

5.1 Interpretation of results

The main research question for our analysis was *"What is the relationship between nitrate levels and type of land use (conventional vs. organic farming)?"*. From our results, we find a significant difference in nitrate concentrations between measurements from organic fields and measurements from conventional fields. With an intercept of 23.48 (the concentration of nitrate in mg/L for the conventional fields), it becomes evident that the estimated mean nitrate concentration value of conventional fields is substantially lower than the 50 mg/L regulative threshold set by the EU. The slope of the model is 1.02 and represents the increase in mg/L when moving to organic fields, meaning that the estimated mean nitrate concentration for organic fields is 24.51 mg/L. Organic fields are, in other words, found to be correlated with significantly higher nitrate concentrations compared to conventional fields. Although the difference is significant, it is important to note the effect size of the findings. The effect size is relatively small since the mean nitrate concentration increases by 1.02 mg/L when going from conventional- to organic fields, which corresponds to an increase of less than 5%. Nonetheless, from a cultural perspective, we deem this finding to be very relevant since it implies that the opinion of the general public is wrong on the differences in groundwater nitrate pollution between organic- and conventional agriculture. However, it is important to note that, although we find a slight disadvantage in terms of nitrate pollution for organic agriculture, organic farming still holds a considerable advantage over conventional agriculture when it comes to biodiversity (Hole et al., 2005; Maeder et al., 2002).

The obvious follow-up questions to our finding would be: *"Why is organic agriculture found to be more nitrate polluting?"*. Although this question is both important and relevant, it is beyond the scope of this paper. As such, our analysis was not geared towards shedding light on the underlying mechanisms of nitrate pollution. Thus, it can be hard to interpret the causality of the effect found through this study. However, when scrutinizing our results, it becomes clear that there is a chance that our analysis did not take all potential confounding factors into consideration.

5.2 Confounding factors

When looking at the results it is important to be wary of potential limitations, in order to be able to draw the right conclusions. For this study, a few confounding factors might have been present and may therefore also have influenced our analysis and results.

One of these possibly confounding factors is rooted in the fact that fields yield different crops. Although we were unable to retrieve data with the crop types of the fields, this could potentially have contributed greatly to our analysis, since some crops use more fertilizers than others (Finck, 1982). Given this type of information, we would have been able to assess the impact of crop type of a field, as it is not unlikely that this could explain a lot of the variance in nitrate levels. Apart from the relevance of such an assessment, it is also worth noting that crop type could potentially confound the interpretation of our results. If organic fields accommodate more crops with high fertilization needs, such as corn, then it could be a potential driver of the effect that we see in our results. Field type (organic/conventional) may seem to play an important role in nitrate concentrations, but if the variation between the two is due to crop type, then this is crucial information to have when interpreting the results. Likewise, precipitation levels and soil type also influence nitrate concentrations (Dubrovsky et al., 2010; Duyvenbooden, 1983; Fewtrell Lorna, 2004). If the distribution of soil type and precipitation between organic and conventional fields is heterogeneous, then this would also affect the results.

From a statistical point of view, it is also clear that the variance in our variogram (see *figure 4*), is not captured perfectly by our kriging model. Furthermore, it is apparent that even at a theoretical distance of zero, there is still a considerable amount of variance as seen from the fact that the semivariance at the y-axis intercept is approximately 500. As such the values interpolated by kriging do not necessarily accurately depict real-world nitrate concentrations. We, therefore, see this as a clear confounding factor in our analysis that one needs to take into account when interpreting the results. However, we argue that there is no reason to believe that this would have an impact on the finding that organic fields are more polluting when it comes to nitrate. We argue this since we believe that it is highly unlikely that the interpolated values are in any way biased towards systematically assigning higher- or lower values to either organic- or conventional fields. As such, we argue that it is a fair assumption that the noise from the kriging is to a certain extent canceled out since it, theoretically, should have equal influence on both agricultural field types.

When taking these potential limitations of our study into account, it becomes evident that more research ought to be conducted on the matter before coming to any clear-cut conclusions about the relationship between nitrate and agriculture. Investigating nitrate pollution is a non-trivial task and requires taking many factors into consideration, making it a complex area to study. Moreover, even upon establishing a reliable nitrate concentration estimate between types of agriculture, an important key to knowledge about the matter is still missing. For which exact mechanisms cause higher nitrate concentrations? Organic and conventional farming differ on various parameters both in terms of fertilization quantity and type, but also in differences in plowing and the use of fallows - all of which may also impact nitrate concentrations. Although such questions are still left unanswered, we hope this study contributes to the overall understanding of the relationship between nitrate concentrations in the groundwater and agriculture while also contributing to a contextualization of how it is perceived within the Danish culture.

6 Conclusion

We have carried out a spatial analysis on the differences in nitrate pollution in the groundwater between organic- and conventional agricultural fields in Denmark. This is done to investigate whether the public opinion on the matter, indicating less nitrate pollution from organic fields, reflects reality. We carried out the analysis by utilizing kriging to interpolate nitrate concentration values for the entirety of Denmark and using this to compute the differences between the two field types. Here, we found a significant difference showing a mean nitrate concentration of 24.51 mg/L for organic fields compared to 23.48 mg/L for conventional fields. Furthermore, we argue that an important question for future research would be to look into what could potentially explain the causality behind our observations. Lastly, we argue that, due to confounding factors, it is important to interpret our results with some skepticism and use them to contextualize the cultural debate rather than drawing clear-cut conclusions.

7 Acknowledgements

We want to give a special thanks to our instructor for the Spatial Analytics course, Adéla Sobotkova for not only providing guidance for our project but also generally providing an informative and useful course. We also want to express our gratitude towards Lars West Andersen and the Danish Agricultural Agency for providing explanations of the data on agricultural fields.

8 Metadata

Nr	Software metadata description	RStudio pipeline for our analysis
S1	Current software version	*R 4.0.3, RStudio 1.2.1335*
S2	Permanent link to executables of this version (your Github repo URL)	*https://github.com/emiltj/groundwater_pollution_dk*
S3	Legal Software License	*MIT License*
S4	Computing platform / Operating System	*Microsoft Windows 10*
S5	Installation requirements & dependencies for software not used in class	R-packages used: pacman, sf, raster, dplyr, tmap, ggplot2, tidyverse, lubridate, sp, gstat, ggthemes, lmtest
S6	If available Link to software documentation for special software	*No special software utilized.*
S7	Support email for questions	*201810219@uni.au.dk*

Table 1: Software and metadata descriptions

Nr	Metadata description	*Link*
D1	metadata definition file	https://github.com/emiltj/groundwater_pollution_dk/blob/master/groundwater_pollution_dk_metadata.md

Table 2: Link to metadata definition file

9 References

Aaberg, M. (2020, May 21). Forbrugere vilde med økologi: Øko-areal fordoblet før tid *DR. https://www.dr.dk/nyheder/regionale/midtvest/forbrugere-vilde-med-oekologi-oeko-areal-fordoblet-foer-tid*

Altamimi, Z. (2018). *ETRS89.* http://etrs89.ensg.ign.fr/

ArcGIS. (2021, June 7). *How Kriging works — Help | ArcGIS for Desktop.* https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm

Arnold, J. B., Arnold, M. J. B. (2015). *Package 'ggthemes'.*

Bezawada, R., Pauwels, K. (2013). What is Special about Marketing Organic Products? How Organic Assortment, Price, and Promotions Drive Retailer Performance. *Journal of Marketing, 77*(1), 31–51. https://doi.org/10.1509/jm.10.0229

Danmarks Naturfredningsforening. (2021). *Kvælstof—Danmarks Naturfredningsforening.* https://www.dn.dk/vi-arbejder-for/landbrug/kvaelstof/

Di, H. J., Cameron, K. C. (2002). Nitrate leaching in temperate agroecosystems: Sources, factors and mitigating strategies. *Nutrient Cycling in Agroecosystems, 64*(3), 237–256. https://doi.org/10.1023/A:1021471531188

Dubrovsky, N. M., Burow, K. R., Clark, G. M., Gronberg, J. M., Hamilton, P. A., Hitt, K. J., Mueller, D. K., Munn, M. D., Nolan, B. T., Puckett, L. J. (2010). The quality of our Nation's waters—Nutrients in the Nation's streams and groundwater, 1992–2004. *US Geological Survey Circular, 1350*(2), 174.

Duyvenbooden, W. V. (1983). *R Nitrate in the Netherlands: A serious threat to groundwater.*

EEC. (1991). *Directive: 91/676/EEC of 12 December 1991.* Official Journal of European Community; Official Journal of European Community. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31991L0676:EN:HTML

Fewtrell Lorna. (2004). Drinking-Water Nitrate, Methemoglobinemia, and Global Burden of Disease: A Discussion. *Environmental Health Perspectives, 112*(14), 1371–1374. https://doi.org/10.1289/ehp.7216

Finck, A. (1982). *Fertilizers and fertilization: Introduction and practical guide to crop fertilization.* Verlag Chemie. https://www.cabdirect.org/cabdirect/abstract/19821972088

Fuller, R. J, Norton, L. R, Feber, R. E, Johnson, P. J, Chamberlain, D. E, Joys, A. C, Mathews, F., Stuart, R. C, Townsend, M. C, Manley, W. J, Wolfe, M. S, Macdonald, D. W, Firbank, L. G. (2005). Benefits of organic farming to biodiversity vary among taxa. *Biology Letters, 1*(4), 431–434. https://doi.org/10.1098/rsbl.2005.0357

Geodætisk Institut. (2020). *UTM-nettet: Universal Transverse Mercator Grid: opbygning og anvendelse.*

GEUS. (n.d.). *GEUS* Retrieved 8 June 2021, from https://www.geus.dk/

Hansen, B., Dalgaard, T., Thorling, L., Sørensen, B., Erlandsen, M. (2012). Regional analysis of groundwater nitrate concentrations and trends in Denmark in regard to agricultural influence. *Biogeosciences, 9*(8), 3277–3286. https://doi.org/10.5194/bg-9-3277-2012

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., Shortridge, A. (2015). Package 'raster'. *R Package, 734.*

Hole, D. G., Perkins, A. J., Wilson, J. D., Alexander, I. H., Grice, P. V., Evans, A. D. (2005). Does organic farming benefit biodiversity? *Biological Conservation, 122*(1), 113–130. https://doi.org/10.1016/j.biocon.2004.07.018

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., Mitchell, D., Zeileis, M. A. (2015). Package 'lmtest'. *Testing Linear Regression Models. Https://Cran. r-Project. Org/Web/Packages/Lmtest/Lmtest. Pdf. Accessed, 6.*

IGIS MAP. (n.d.). Retrieved 8 June 2021, from https://www.igismap.com/

Knowles, R. (1983). Denitrification. *Microbiological Reviews, 46*(1), 43–70.

Ladegaard, M. (2019). *Drikkevandet er under pres – det økologiske landbrug har en løsning.* Økologisk - nyt om udviklingen. https://nyheder.okologi.dk/politik-og-udvikling/drikkevandet-er-under-pres-det-okologiske-landbrug-har-en-losning

LBST. (n.d.). *Landbrugsstyrelsen.* Retrieved 8 June 2021, from https://lbst.dk/

Lindinger, H., Scheidleder, A. (2004). Indicator fact sheet.(WEU1) Nitrate in groundwater. *European Environmental Assessment.*

Maeder, P., Fliessbach, A., Dubois, D., Gunst, L., Fried, P., Niggli, U. (2002). Soil Fertility and Biodiversity in Organic Farming. *Science, 296*(5573), 1694–1697. https://doi.org/10.1126/science.1071148

Myrold, D. D., Tiedje, J. M. (1985). Establishment of denitrification capacity in soil: Effects of carbon, nitrate and moisture. *Soil Biology and Biochemistry, 17*(6), 819–822. https://doi.org/10.1016/0038-0717(85)90140-3

Pearson, D., Henryks, J. (2008). Marketing Organic Products: Exploring Some of the Pervasive Issues. *Journal of Food Products Marketing, 14*(4), 95–108. https://doi.org/10.1080/10454440801986421

Pebesma, E., Bivand, R., Pebesma, M. E., RColorBrewer, S., Collate, A. A. A. (2012). Package 'sp'. *The Comprehensive R Archive Network.*

Pebesma, E., Graeler, B., Pebesma, M. E. (2015). Package 'gstat'. *Comprehensive R Archive Network (CRAN)* 1–0.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Raadet for Groen Omstilling. (n.d.). Økologisk versus konventionelt. *Rådet for Grøn Omstilling.* Retrieved 4 June 2021, from https://rgo.dk/vi-arbejder-for/okologisk-versus-konventionelt/

Sánchez-Bayo, F., Baskaran, S., Kennedy, I. R. (2002). Ecological relative risk (EcoRR): Another approach for risk assessment of pesticides in agriculture. *Agriculture, Ecosystems Environment, 91*(1), 37–57. https://doi.org/10.1016/S0167-8809(01)00258-4

Schullehner, J., Sigsgaard, T., Hansen, B. (2014). *Nitrat i drikkevandet og vores sundhed.* 20.

Spinu, M. V. (2016). *Package 'lubridate'.* Recuperado el.

Strebel, O., Duynisveld, W. H. M., Böttcher, J. (1989). Nitrate pollution of groundwater in western Europe. *Agriculture, Ecosystems Environment, 26*(3), 189–214. https://doi.org/10.1016/0167-8809(89)90013-3

Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software, 84*(6), 1–39.

Ward, M., deKok, T., Levallois, P., Brender, J., Gulis, G., Nolan, B., VanDerslice, J. (2005). Workgroup Report: Drinking-Water Nitrate and Health—Recent Findings and Research Needs. *Environmental Health Perspectives, 113*(11), 1607–1614. https://doi.org/10.1289/ehp.8043

WHO. (1993). *Guidelines for drinking-water quality.* World Health Organization.

Wick, K., Heumesser, C., Schmid, E. (2012). Groundwater nitrate contamination: Factors and indicators. *Journal of Environmental Management, 111*(3), 178–186. https://doi.org/10.1016/j.jenvman.2012.06.030

Wickham, H. (2011). Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics, 3*(2), 180–185.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.

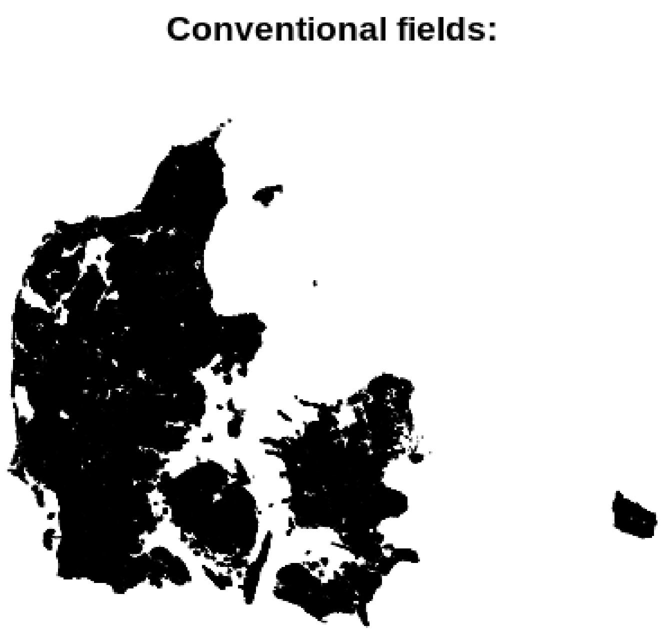
Wickham, H., Wickham, M. H. (2020). Package 'plyr'. *Obtenido Httpscran Rproject Orgwebpackagesdplyrd-plyr Pdf.*

Willer, H., Lernoud, J. (2019). *The World of Organic Agriculture. Statistics and Emerging Trends 2019* (H. Willer J. Lernoud, Eds.; pp. 1–336). Research Institute of Organic Agriculture FiBL and IFOAM Organics International. https://orgprints.org/id/eprint/37018/

Zhu, Z., Stein, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics, 11* (1), 24–44. https://doi.org/10.1198/108571106X99751

10 Appendix

Appendix A - Conventional fields (prior to subsetting):



Appendix B - Conventional fields: