IODINE IN DANISH GROUNDWATER AND DRINKING WATER

INTERDISCIPLINARY STUDY ON HYDROGEOCHEMISTRY AND HUMAN HEALTH

Denitza D. Voutchkova
PhD progress report
12/14/2012
Evaluation Committee

Søren Munch Kristiansen (supervisor)
Department of Geoscience, Aarhus University

Birgitte Guldberg Hansen (co-supervisor)
Geological Survey of Denmark and Greenland (GEUS)

Bo Holm Jacobsen (internal examiner)
Department of Geoscience, Aarhus University

Torben Sigsgaard (external examiner)
Department of Public Health – Institute of Environmental and Occupational Medicine, Aarhus University

Qualifying Exam (part A)

January, 28th 2013

Department of Geoscience
Aarhus University
Høegh-Guldbergs Gade 2
DK-8000 Aarhus C

E-mail: ddv@geo.au.dk
This progress report presents the work done throughout the last 17 months and the plans for the rest of the period of my PhD studies. The progress report and the following oral presentation are part of the midterm evaluation on my research at the Department of Geoscience, Aarhus University, where I am currently employed as PhD fellow. My PhD project is part of a 3 year Geocenter research project “Iodine in the hydrological cycle in Denmark: implications for human health”, financed jointly by GEUS and Aarhus University. While the frame of the project includes the whole hydrological cycle of iodine (I), my work so far is focused on a small part of it – I concentrations and speciation in the groundwater and drinking water in Denmark.

I would like to use the chance to thank my supervisors Søren Munch Kristiansen (AU) and Birgitte Guldberg Hansen (GEUS) as well as GSST for giving me the chance to work with this fascinating topic and to be part of a great research team. The work done until now and presented in this report is result of the close collaboration and the fruitful discussions with Søren, Birgitte and the other project participants from GEUS: Vibeke Ernstsen who is project supervisor and responsible for the Inorganic Chemistry Lab at GEUS and without her and Christina Lynge measuring I species would be impossible; Kim Esbensen who is also project supervisor and opened the door to the multivariate data analysis universe and taught me how to use PCA and PLS; Brian Sørensen whose experience with Jupiter database provided significant help in the extraction of historical groundwater chemical data; and last but not least Lasse Gudmundsson for the assistance with the groundwater sampling and the help with organizing it. I would like to thank all of them also for the comments, recommendations and ideas throughout the whole process of analysing the historical I data and the preparation of the draft of my first paper.

Denitza D. Voutchkova
(Деница Д. Вучкова)
CONTENTS
Preface .................................................................................................................................................. ii
Objectives and Scope of the PhD ....................................................................................................... 1
Geochemistry and Geology .................................................................................................................. 2
  Danish Geology and Groundwater .................................................................................................. 2
  From Groundwater to Tap Water .................................................................................................. 3
  The Iodine Cycle in the Terrestrial Environment: Concentrations and Speciation .................... 4
Epidemiology and Nutrition ................................................................................................................ 5
  Iodine Intake and Human Health ................................................................................................. 5
  Short Historical Overview on the Iodine Enrichment in Denmark .............................................. 7
  Methods for Estimating the Iodine Intake ...................................................................................... 8
    Dietary Iodine Intake in Denmark ............................................................................................... 8
    Iodine Intake Based on UI excretion in Denmark ................................................................... 9
    Connection and Comparison between Dietary Intake and UI .................................................. 10
  Other Factors .................................................................................................................................. 11
Summary ............................................................................................................................................. 12
1st Part of the PhD: Multivariate Analysis of Historical Groundwater Data .................................... 12
2nd Part of the PhD: Case Studies ..................................................................................................... 13
  Site Description ............................................................................................................................ 13
  Objectives ...................................................................................................................................... 14
3rd part: National Study ..................................................................................................................... 14
  Criteria for Choosing Sampling Locations .................................................................................. 15
  Data Analysis ................................................................................................................................ 17
Plan for the Remaining Study Period ............................................................................................. 17
References: ........................................................................................................................................ 19
Appendix A: Activities During Part A ............................................................................................. 21
  Completed Courses: ....................................................................................................................... 21
  Teaching Assistance ....................................................................................................................... 21
  Conferences and Meetings .......................................................................................................... 21
Appendix B: Submitted for Publication Paper ................................................................................. 21
OBJECTIVES AND SCOPE OF THE PhD

The general goal of this PhD project is to investigate the spatial variation of iodine (I) concentrations in the Danish ground- and drinking water. More specifically, we are interested in elucidating the sources and processes influencing this variability by focusing on the geological and geochemical causes. Further, we expect to look deeper into the geo-availability, bio-accessibility and bioavailability of I and the human health effects.

Interdisciplinary knowledge and collaboration are the prerequisites for successful completion of this project. We have the ambition to establish a link between specialists from the fields of geoscience (geography, geology, geochemistry etc.) and public health (nutrition, epidemiology, etc.) in order to understand the complicated nature of I in its entirety which would result in high quality medical geology\textsuperscript{1} research in Denmark.

The PhD project is divided in three parts (see Figure 1). The first part consists of a study based on historical data on I in Danish groundwater with main focus on the geochemistry and geology. The second part is a logical continuation, where we have chosen four locations for further small scale studies with the possibility for connection with existing epidemiological data. In this part of the project we conduct ground- and drinking water sampling. The analytical methods are developed in the Inorganic Chemistry Lab of GEUS for this project. The third part is a national study on I in the drinking water and the health effects at population level, including a sampling campaign with around 180 sampling points, which is currently being designed.

1. Multivariate analysis of historical data
   - Submitted paper
   - Hydrogeochemistry

2. Case Studies: Skagen, Randers, Ishøj, Stevns
   - Lab analyses soon finished
   - Human Health Effects

3. National study
   - Planning started
   - Drinking water

FIGURE 1 SCHEMATIC PRESENTATION OF THE PhD PROJECT AND ITS STATUS

Throughout the next two chapters “Geochemistry and Geology” and “Epidemiology and Nutrition” follows a brief introduction and a literature review with stress on the gaps of knowledge or the need of further research is given. After the background information, the three parts of the PhD are presented in greater detail. Last but not least, a plan for the rest of the PhD is proposed as a starting point for discussion at the qualifying exam.

\textsuperscript{1} Medical geology is a rapidly growing discipline dealing with the influence of geology on the distribution of health problems (Selinus et al. 2011)
The present Danish landscape is shaped mainly during the last ice age. Figure 2 presents the geological strata, below the Quaternary glacial deposits, where most of the primary aquifers are located. Aquifers are permeable geological units (sand, gravel or fractured rock) that can yield water in usable quantities when tapped by a well (Spitz 1996).

Danish primary and secondary aquifers (except on Bornholm) consist mainly of Quaternary or Miocene sand and/or gravel, Palæocene to Late Cretaceous chalk/limestone (Kelstrup et al. 1982). Bornholm, on the other hand, has a more diverse geology and the following aquifers can be found: Late Cretaceous limestone, sandstone and sand; Jurassic and Lower Cretaceous gravel and sand; Middle Cambrian to Silurian shale and limestone; Lower Cambrian sandstone and quartzite; and pre-Cambrian gneiss and granite (Kelstrup et al. 1982).

The groundwater composition may vary substantially from region to region depending on the sources and the processes affecting it and on its age. The groundwater is in constant interaction with sediments, where the oxidation/reduction processes, weathering and dissolution of minerals, ion exchange reactions, mixing of waters with different water qualities, and the anthropogenic activities result in change of the groundwater composition. As rainwater is the source of most of the groundwater, the rain composition, the evapotranspiration and the leaching through the soils are also contributing to the diversity of the groundwaters.

In Denmark almost all of the drinking water is with groundwater origin; for 2010 and 2011 only one waterwork has reported surface water abstractions of about 0.1mio m$^3$/year (Heraldsted lake, Ringsted) (Jupiter, status December 2012).
FROM GROUNDWATER TO TAP WATER

Table 1 gives an overview on the number of waterworks and their type, as well as the reported groundwater abstraction volumes for 2010 and 2011. This information is based on data extracted from Jupiter database in December 2012. Some more abstraction volumes for 2011 might be reported in the future. As it is seen from Table 1, there are 5250 waterworks registered in Jupiter, a little bit more than half of which are reporting abstraction of about 398 mio m$^3$ for 2010. To be precise, the small water supplies ("Husholdninger, 3-9 husstande") should not be included when talking about waterworks, because treatment facilities might be missing, the annual abstraction volumes are not bigger than 0.06 mio m$^3$/year and the water quality control is low. Taking this into consideration, the “real” waterworks reporting groundwater abstractions for 2010 are 2585 and yield about 397 mio. m$^3$/year.

<table>
<thead>
<tr>
<th>Type of organisation</th>
<th>Count all (by ID number)</th>
<th>Number of waterworks with reported abstractions</th>
<th>Reported abstraction volumes [mio m$^3$/year]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>for 2010</td>
<td>for 2011</td>
</tr>
<tr>
<td>Small supply system (3-9 households)</td>
<td>772</td>
<td>249</td>
<td>271</td>
</tr>
<tr>
<td>Public water supply system</td>
<td>679</td>
<td>337</td>
<td>326</td>
</tr>
<tr>
<td>Private water supply system</td>
<td>3785</td>
<td>2246</td>
<td>2120</td>
</tr>
<tr>
<td>Water supply (different code in Jupiter)</td>
<td>14</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Sum</td>
<td>5250</td>
<td>2834</td>
<td>2718</td>
</tr>
</tbody>
</table>

Most of those waterworks abstract water from more than one well or well site. In some cases groundwater from different aquifers is pumped up and mixed together before it gets to the treatment facility (e.g. Copenhagen area). The raw groundwater undergoes physical and/or chemical treatment at the waterworks after which it is stored as pure drinking water in buried in the ground tanks or pumped to water towers. Part of the pure drinking water is transported by system of pipelines to houses and working places, where it is ready to be used as tap water. Other part of the pure drinking water might go to certain industries or to other waterworks. The groundwater treatment in most of the Danish waterworks is considered as simple treatment, consisting of only aeration and sand filtration. However, in some places a more sophisticated treatments might be used, e.g. Skagen waterwork.

FIGURE 3 GROUNDWATER – DRINKING WATER – TAP WATER SCHEME

As Figure 3 implies, there is a difference between the terms groundwater, drinking water and tap water. The tap water quality is result of the combined effect of the groundwater quality, the treatment and the transport procedures. Almost 100% of the drinking water in Denmark is with groundwater origin, however not all groundwater serves drinking water purposes. The drinking water from one storage facility might be supplying houses in one city, neighbourhood or couple of small villages through different pipelines, which might also affect the quality of the tap water due to the material or the age of the supply pipeline. The water quality is monitored at the abstraction wells (groundwater quality), when leaving the water treatment facility (drinking water quality), and also at chosen consumers (tap water quality). The frequency of the inspection and its type depends on the produced/supplied water volumes, e.g. waterworks producing <35000m$^3$/year have to check the quality at the waterwork once each two years, however if the produced drinking water is 3.5-7mio m$^3$/year then it should be done at least four times per year for some of the monitored chemical elements (6 times for micro organic pollutants). The quality requirements both for
the pure drinking water and the tap water are given in (Miljøministeriet 2011). Iodine is not part of the monitoring programs of the Danish waterworks and there are no international water standards for it either (Reimann and Birke 2010). Reimann and Birke (2010) explain the lack of standards for I with the ideology of the regulations: the focus is normally on contamination and eventual toxicity of certain element and no tradition of setting deficiency warnings exists.

THE IODINE CYCLE IN THE TERRESTRIAL ENVIRONMENT: CONCENTRATIONS AND SPECIATION

The I geochemical cycle is a spatially and temporally dynamic system, where different biotic and abiotic processes are taking a part in. In the hydrogeochemical cycle, I is predominantly in oxidation state -1 (iodide, I\(^-\)) and +5 (iodate, IO\(_3^+\)); and, next to these inorganic forms, I is also found in different organic species, which can also contribute a significant fraction of the total I (Hu et al. 2009). The water composition together with I concentration and speciation can be used to study the origin/the source of I and the processes influencing its geographical distribution. It is possible to group samples with similar composition and to explain the grouping by different factors, e.g. geology, landuse, depth of extraction etc. The associations between I and other elements in the groundwater and drinking water is important not only for geochemical studies, but also when studying the combined effect of I and potential goitrogenic elements from the drinking water on the human health.

Iodine concentrations found in different media (sea-, rain-, bottled mineral and spring water, as well as sediments) are summarised in the paper introduction (Appendix B). Contents of I in Danish aquifer sediments or in the soil and their geographical distribution have not been studied at all.

Existing data on I in Danish drinking water are limited to studies based on very few sampling locations: tap water from 55 (K. M. Pedersen et al. 1999), 41 (Rasmussen et al. 2000), and 47 locations (Saxholt et al. 2008), and drinking water from 22 waterworks (Andersen et al. 2002). This contemporary data set shows that the I concentration in the drinking water in Denmark varies from 0.7 µg/l to 140 µg/l. However there is one order of magnitude difference between the highest values reported by K. M. Pedersen et al. (1999), Andersen et al. (2002), Saxholt et al. (2008) and the one found by Rasmussen et al. (2000) for Jutland. According to Andersen and Lauberg (2009) there is a distinguishable difference between East and West Denmark with respect to I content in drinking water. Even though such regional variations of I content in the Danish drinking water are known, no detailed geochemical studies related to I exist.

As part of our first study we found that the I concentration in the groundwater varies from below detection limit to 1220 µg/l (median: 5.4 µg/l, average: 13.83 µg/l). From Figure 2 (Appendix B) can be seen the geographical distribution of the samples and their I concentrations. As in the cited studies above, more samples with higher concentrations are found on Sealand than on Jutland. However, this trend is not that clear, because both high and low I concentrations are clustered together.

As with the concentration, the I speciation in different media varies. A lot of studies have been done on I speciation of sea water. It is known that I is found as I\(^-\), IO\(_3^-\), dissolved organic I (DOI), volatile organic carbons (CH\(_2\)I and other forms), I\(_2\), and particulate I, but the most abundant is IO\(_3^-\) (Ito and Hirokawa 2009). IO\(_3^-\) is converted to I\(^-\) by biological processes in surface water and also near the sediments under anoxic conditions, ibid. The species distribution varies with depth and geographical location, ibid. V. Hansen et al. (2011) studied IO\(_3^-\) and I\(^-\) spatial distribution in the Baltic Sea, Kattegat and Skagerak and linked it to a great extend to the water transport and the environmental conditions.

Iodide is the most common I form in rain making up to 50% of rainfall-I, whereas IO\(_3^-\) is the second major component (Fuge 2005). Opposite, for Denmark Hou et al. (2009) measured IO\(_3^-\), Ti and non-ionic I in the precipitation from Roskilde in the period 2001-2006 and found that the dominating species is IO\(_3^-\).

No measurements on the I species (I\(^-\), IO\(_3^-\), organic bound I and TiI) in the Danish drinking and groundwater have been reported. However, Andersen et al. (2002) stated that I in Danish groundwater is bound in humic substances, because
they found association between the TI and the humic substances for samples from 6 places (Skagen, Ringsted, Nykøbing, København, Samsø, Nakskov) by using HPLC size exclusion. However, neither organic matter (DOC, TOC, NVOC etc.), nor I species were measured directly. It is of great interest (both geochemical but also epidemiological) to find a more convincing proof, so further investigation on the I and its form in the Danish drinking and groundwater is desirable.

The variation in I concentration (also speciation) in Danish groundwater can be influenced by one or combination of processes and sources: desorption from clay and/or high organic matter sediments; seawater intrusion caused by abstractions close to the coastline; leaching of I from the topsoil; dissolution of carbonate rocks and release of saline pore water; upward transport of highly mineralised water to the groundwater etc. One of the most widespread hypothesis is that I originates from recent marine deposits, however no comprehensive geochemical and geological study on the I origin in the Danish groundwater exists.

EPIDEMIOLOGY AND NUTRITION

IODINE INTAKE AND HUMAN HEALTH

Iodine is an essential component of human thyroid hormones regulating the metabolic processes in cells and playing a role in the early development of most organs, including the brain (WHO 2007). The recommended daily nutrient intake (RNI) differs from age group to age group: 0-59 months is 90 µg, 6-12 years - 120 µg, adolescence and adults - 150 µg, and during pregnancy and lactation - 250 µg, ibid. Insufficient or excessive I intake can both cause different health effects (listed further), however the focus falls mainly on the deficiency (ID). The epidemiological criteria for assessment of the severity of this problem on population level are normally either based on goitre prevalence (not discussed here) or on median urinary I excretion (UI). The serum TSH concentrations and thyroid hormones further reflect the effects of I deficiency (WHO 2007), however no literature on that topic has been reviewed as part of this report.

According to the WHO (2007) only one country in Europe is with moderate ID (20-49 µg/l UI), ten with mild (50-99 µg/l), 20 with optimal I nutrition (100-199 µg/l), and only one with more than adequate (200-299 µg/l), Figure 4. Denmark is amongst the ten countries with mild ID. The map (Figure 4) is compiled from different recent studies; for countries where data on national level is not present, e.g. Denmark and 13 other, a subnational surveys are used (WHO 2007).

The WHO (2007) has grouped the wide spectrum of I deficiency disorders (IDD) occurring if the nutritional recommendations are not met in a population on age principle: fetus – abortions, stillbirths, birth defects, perinatal mortality, endemic cretinism; neonate – hypothyroidism, endemic mental retardation; child and adolescent – goitre, hypo- and hyperthyroidism, impaired mental function, slower physical development; adult – goitre with its complications, hypo- and hyperthyroidism, impaired mental function. Additionally, increased vulnerability of the thyroid gland to nuclear radiation is characteristic despite the age (WHO 2007).

---

2 “RNI refers to the daily intake which meets the nutrient requirements of almost all (97.5%) apparently healthy individuals in an age- and sex-specific population group. It is set at the estimated average requirement + 2 standard deviations.” (WHO 2007)

3 Enlarged thyroid gland
Laurberg et al. (2001) (Figure 5) and Laurberg et al. (2010) (Figure 6) connect the epidemiological criteria with the associated health risk. Figure 6 presents a slightly different overview than Figure 5 which might be partly due to the difference in the used marker: I vs. I' or the conversion between µg/24hours to µg/L. The levels from Figure 6 correspond better to the epidemiological criteria of the WHO (2007). These intake levels (from both examples) are defined on a population or subpopulation basis and might differ on individual scale based on factors like age, gender, exposure to goitrogens (Laurberg et al. 2001), fluctuations in the intake, other nutritional deficiencies, interaction between I intake and genetics (Laurberg et al. 2010).

![FIGURE 4 IODINE NUTRITION IN EUROPE BASED ON UI (DATA FROM 1994-2006) (WHO 2007)](image)

<table>
<thead>
<tr>
<th>Median 24-hour urinary iodine excretion of population or subpopulation</th>
<th>Risk associated with iodine intake level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe iodine deficiency</td>
<td>&lt;25 µg</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate iodine deficiency</td>
<td>25–60 µg</td>
</tr>
<tr>
<td>Mild iodine deficiency</td>
<td>60–120 µg</td>
</tr>
<tr>
<td>Optimal iodine intake</td>
<td>120–220 µg</td>
</tr>
<tr>
<td>Mild iodine excess</td>
<td>220–400 µg</td>
</tr>
<tr>
<td>400–800 µg</td>
<td>Hypothyroidism in the middle aged and elderly</td>
</tr>
<tr>
<td>&gt;800 µg</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Severe iodine excess</td>
<td></td>
</tr>
</tbody>
</table>

If the 24-hour urine volume is around 1.5 liter, optimal iodine intake gives a median nonfasting spot urine concentration of ~80–150 µg/L. Values are for populations of adults. Intakes should be ~50 µg per 24 hours higher in pregnant and lactating women (4).

![FIGURE 5 HEALTH RISKS ASSOCIATED WITH POPULATION I INTAKE (24 HOURS UI MEDIAN) (FROM LAURBERG ET AL. 2001)](image)
Short Historical Overview on the Iodine Enrichment in Denmark

A working group, formed to re-evaluate the need of I enrichment program in Denmark, published their decision in 1996 (Rasmussen et al. 1996). The conclusion, based on the existing at that time data, was that the I intake in Denmark is below the recommended values and there is need of increasing it, ibid. However, Denmark is not considered to be I deficient, nor to have endemic goitre; it “…must be regarded as area with marginal iodine deficiency, and in the elderly population, especially among women, goitre is present with high frequency” (Rasmussen et al. 1996).

Iodized salt could not be bought legally in Denmark until 1998 when a voluntary program for using iodized salt started (Rasmussen et al. 2002; Laurberg et al. 2006). The target of this program was to increase the I intake with 50µg/day (Laurberg et al. 2006). Two years later the voluntary program was revised and found insufficient; as a result a mandatory one was to be introduced in the period July 2000-April 2001, ibid. From that period I was added to the salt used in households, and in production of cakes and bread, ibid.

Before implementing the salt fortification, a monitoring program “The Danish Investigation of Iodine Intake and Thyroid Diseases” (DanThyr) was developed (Laurberg et al. 2006). DanThyr consists of three main parts: 1) cohort studies in two areas in Denmark: Aalborg and Copenhagen before and after the fortification program; 2) a register of new cases of hyper- and hypothyroidism in Aalborg and Copenhagen; 3) a central (national) register for surgical, medical, and radiiodine treatment (Laurberg et al. 2006).

It must be noted that the cohort study locations were chosen so they “…cover the main difference in levels of iodine intake in Denmark caused by different levels of iodine in groundwater…” (Laurberg et al. 2006). The levels they refer to are from (K. M. Pedersen et al. 1999), where tap water samples were collected at 55 lab locations. The representativity of these samples with respect to the population’s I intake from drinking water is questionable, however at that time these were the only present data on I in drinking water.

Iodine Nutrition Table

<table>
<thead>
<tr>
<th>Iodine Nutrition</th>
<th>Median Urinary Iodide Concentration µg/l</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe ID</td>
<td>&lt;25</td>
<td>Cretinism, Goitre, Hypothyroidism</td>
</tr>
<tr>
<td>Moderate ID</td>
<td>25–50</td>
<td>Low IQ, Goitre, Hypothyroidism, Hyperthyroidism</td>
</tr>
<tr>
<td>Mild ID</td>
<td>50–100</td>
<td>Goitre, Hypothyroidism</td>
</tr>
<tr>
<td>Optimal</td>
<td>100–200</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>More than adequate</td>
<td>200–300</td>
<td>Hypothyroidism, Early Graves’ disease?</td>
</tr>
<tr>
<td>Excessive</td>
<td>&gt;300</td>
<td>Hypothyroidism, Goitre, Early Graves’ disease?</td>
</tr>
</tbody>
</table>

The exact boundaries of iodine intake associated with the different diseases depends on intake of goitrogens and other nutritional deficiencies. Subgroups of the population may have iodine intake levels that differ from the main population.

FIGURE 6 HEALTH RISKS ASSOCIATED WITH POPULATION I INTAKE (MEDIAN URINARY IODIDE) (FROM LAURBERG ET AL. 2010)
METHODS FOR ESTIMATING THE IODINE INTAKE

There are two methods for estimating I intake: by assessing the dietary intake or by measuring the UI excretion (Rasmussen et al. 2002). UI excretion represents around 90% of the I intake (Laurberg et al. 2001).

DIETARY IODINE INTAKE IN DENMARK

Drinking water is generally not considered a significant major contributor to the dietary I intake as it usually provides close to 10% only (Fuge 2005). However, in Denmark, where drinking water originates from groundwater, 25% of the I intake in the average diet was derived from drinking water and other beverages (some containing high proportions of processed groundwater) prior to recent mandatory I fortification of table salt (Rasmussen et al. 2000). Additionally, as part of DanThyr cohort study (Rasmussen et al. 2002) a Food Frequency Questionnaire (FFQ) (n=4366) and diet records (n=269 from which 254 also completed FFQ) were used for estimating the dietary I intake of the participants from Aalborg and Copenhagen. Rasmussen et al. (2002) found that milk and other beverages contribute about 68% of the I intake (44% from milk and milk products) before the implementation of salt fortification program.

A more recent study by A. N. Pedersen et al. (2010) on the dietary habits in Denmark (n=4431) shows that after the fortification, I from drinking water and other beverages represents 14% of the dietary intake, Figure 7. The I from salt accounts for 19%, whereas the highest contribution is from milk and milk products – 35%, *ibid.*

![Diagram showing contribution from different food groups to the dietary iodine](image)

**FIGURE 7 CONTRIBUTION TO THE DIETARY I FROM 12 FOOD CATEGORIES: MILK & MILK PRODUCTS, VARIOUS, BREAD & CORN, BEVERAGES, FISH, CHEESE, EGGS, POTATOES, VEGETABLES, FRUITS, MEAT, AND SUGAR & CANDY (DATA FROM (A. N. PEDERSEN ET AL. 2010))**

---

5 Including water, tea, coffee, juice, soft drinks, beer and wine (Rasmussen et al. 2002);
6 All kind of milk (low, fat, etc.) and various curdled milk and cream products;
7 Spices, including salt;
8 All kind of bread, grain products including rice, pasta and breakfast cereals, flour;
9 Diverse group spanning from water to snaps; including coffee, tea, soft drinks, wine, spirits, but not milk and juice;
10 Fresh, frozen and smoked fish, canned fish and shellfish;
11 All kind of cheese including quark naturel;
12 Eggs and eggs used for cooking (omelette, gratin, pastry, etc.);
13 Potatoes, potato flour, French fries;
14 Fresh and processed vegetables, excluding potatoes, including frozen vegetables, canned, dried legumes and ketchup;
Next to the geographical variation in I concentration in drinking water and milk, a seasonal variation in I from milk was found with higher concentrations in March than in June (A. N. Pedersen et al. 2010). The long term temporal variation in I from drinking water has not been studied on a national level. Rasmussen et al. (2000) found no significant difference in I concentration from 41 tap water samples collected in January and June 1997, respectively. Additionally, K. M. Pedersen et al. (1999) found that the day-to-day variation of I concentration from tap water at two labs measured in 10 consecutive days varies little: 1) Aalborg 4.2-6.7 µg/l, and 2) Skagen 138-151 µg/l. Further and more detailed studies should be conducted to be able to conclude on the temporal variability of I in the Danish drinking- and groundwater, and consequently the effect on the population’s dietary I intake.

**IODINE INTAKE BASED ON UI EXCRETION IN DENMARK**

The I intake estimated from UI excretion can be based on 24-hour or spot urine samples. The most precise estimate is from collecting 24-hour samples; however this method is troublesome in epidemiological studies and this is the reason why the I concentration from spot urine samples is widely used as estimate for I excretion on a population level (Knudsen 2005). I concentrations shouldn’t be compared to 24-hour I excretions directly due to issues with the dilution and the variation of 24-hour urine volume, *ibid*. To adjust the I concentration from spot urine a urine creatinine concentration is used. Knudsen (2005) presents a comparison between these different ways of estimating UI excretion, Figure 8, and an adjusted I/creatinine ratio taking into account the age and sex differences in the creatinine excretion.

| Four ways of expressing iodine excretion in epidemiological surveys. Percentiles and comparison between 24-hour urine values and three estimates based on corresponding casual urine samples (n=112). |
|---------------------------------|---|---|---|---|---|
| 24-hour iodine excretion (µg/day) | 5th | 25th | median | 75th | 95th |
| Adjusted iodine/Cr ratio (µg/day) | 58 | 92 | 126 | 161 | 319 |
| Iodine / gram creatinine (µg/g) | 41 | 55 | 77 | 105 | 198 |
| Iodine concentration (µg/l) | 39 | 68 | 87 | 133 | 230 |

$r =$ regression coefficient compared to 24-hour iodine excretion. Data from (IV).

**FIGURE 8 COMPARISSE O F DIFFERENT WAYS TO ESTIMATE UI EXCRETION (KNUDSEN 2005)**

K. M. Pedersen et al. (1997) summarised the existing data from Danish studies on I intake (based on UI excretion) for the period 1969 to 1995 before the start of DanThyr. From this summary can be seen that the only study (24 hours UI) covering whole Denmark has been performed on 4580[15] male gymnasium students in the age group 17-20 (abstract published by Munkner (1969)). Areas where regional studies have been conducted throughout 1969-1995 are Copenhagen (n=111, n=143, n=30), Fyn (n=505), Randers (n=20, n=54, n=423, n=29), Holbæk (n=391), Aarhus (n=30), Aalborg (n=33), and Ringkøbing (n=30), (K. M. Pedersen et al. 1997).

The effect of the salt fortification program on the UI excretion was evaluated as part of DanThyr. Two cross-sectional studies with participants from Aalborg and Copenhagen were carried out: before (C1 cohort, n=4649, period 1997-1998) and after the fortification (C2 cohort, n=3570 participants from C1 not included, period 2004-2005), (Rasmussen et al. 2008). Iodine excretion was expressed both as I concentration (µg/l) and as 24-hour estimate (µg/day) based on the measured I concentration, creatinine ratio and expected daily creatinine excretion. The median UI value found before and after the fortification is 61 vs. 101µg/l (Aalborg) or 94 vs. 145µg/day (Copenhagen), respectively (Rasmussen et al. 2008). Table 2 shows the medians of the 24-hour UI estimated values before and after the fortification for participants taking and not taking I containing supplements.

---

15. Fresh and processed fruit, dried fruit, jam, fruit concentrates and juice, nuts and other oil seeds
16. All kind of charcuterie and offal, excluding poultry
17. Chocolate, marzipan, honey, syrup, powdered sugar and sugar from cakes, does not include the sugar from soft drinks
18. Munkner (1969) reports n=6000 (including Faroe Islands);
The results show an increase in the I intake in all investigated groups (close to the planned 50μg/day), fewer individuals had low I intake after the fortification (9% of the samples <70 μg/day), and none of the groups can be regarded as having “excessive” intake (Rasmussen et al. 2008). Further Rasmussen et al. (2008) discuss the need of continuous monitoring in various population groups and the possible existence of groups with insufficient intake even in populations with sufficient median I excretion.

A follow up observational study on the C1 cohort (n=2465, 59.1% of the participants in the first survey 1997-1998) was conducted in the period 2008-2010 and recently reported by Bjergved et al. (2012), Table 3. Iodine intake increased in both Copenhagen and Aalborg to mild deficiency, however the values found for C1 follow up cohort are significantly lower than the one for C2 cohort, which suggests that the I intake has decreased in the recent years (Bjergved et al. 2012). The authors noted that the relative low participation rate might have introduced selection bias in the results, so additional studies are needed in order to evaluate this decrease, ibid.

**Table 3 Median UI excretion concentrations for the C1 cohort, results for the baseline study before the fortification program 1997-1998 and 11 years follow up 2008-2010 (based on data from Bjergved et al. (2012))**

<table>
<thead>
<tr>
<th>µg/l</th>
<th>C1 baseline</th>
<th>C1 follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not taking I supplements</td>
<td>52</td>
<td>75</td>
</tr>
<tr>
<td>All</td>
<td>61</td>
<td>84</td>
</tr>
</tbody>
</table>

Recently comparative population-based study with participants from Randers (n=212, born 1920) and Skagen (n=218, born 1918-1923) was published by Andersen et al. (2012). The study has been carried out prior to the beginning of the salt fortification program (1997-1998, (Andersen et al. 2008)) and shows that the UI excretion is in the recommended range for Skagen, while the participants from Randers were mild to moderately deficient (Table 4), ibid.

**Table 4 Median of the estimated 24-hour UI excretion before the fortification program (age and gender corrected creatinine adjustment) (based on data by Andersen et al. (2012) and Andersen et al. (2008))**

<table>
<thead>
<tr>
<th>µg/day</th>
<th>Randers</th>
<th>Skagen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not taking I supplements</td>
<td>50</td>
<td>177</td>
</tr>
<tr>
<td>All</td>
<td>74</td>
<td>184</td>
</tr>
</tbody>
</table>

**Connection and Comparison Between Dietary Intake and UI**

In his PhD thesis Knudsen (2005) notes that registration of food intake is rarely used in thyroid epidemiological studies, because even though it is a straightforward approach for measuring the I intake, knowledge about the I concentration and variation in the food as well as the long term dietary habits of each individual is crucial and might introduce bias. Rasmussen et al. (2001) evaluated the difference between estimated dietary I intake (determined by FFQ and from 4 day dietary record) and the measured 24 hour UI for Aalborg and Copenhagen separately. The comparison between FFQ (including only the most I-rich foods) and the dietary record (covering the whole diet) showed no significant difference, however the intake from FFQ was higher than the intake from dietary records for Aalborg, but not for Copenhagen (Rasmussen et al. 2001). UI excretion (24 hour sample) was found to be significantly
lower than I intake: for Copenhagen (n=23) 115 vs 121 µg/day; for Aalborg (n=50) 81 vs. 99 µg/day; and for all samples (n=73) 91 vs. 106 µg/day, respectively (Rasmussen et al. 2001).

K. M. Pedersen et al. (1999) studied partially the connection between dietary I intake and UI. They correlated the I content of tap water with the average 24-hour UI (data from Munkner (1969)) for 41 Danish cities and found significant correlation (r=0.68, P<0.001) with regression line y=1.7x+43.2.

**OTHER FACTORS**

Gaitan (1990) reviews studies on the etiology of endemic goitre and discusses evidences that not only I deficiency, but also other environmental factors may cause that condition. Some of these environmental factors are naturally occurring compounds (goitrogens and antithyroid agents) present in food products and water (Gaitan 1990). Goitrogens (sulfurated organics, flavonoids, pyridines) are found in milk, cassava, millet, pinon nut, babassu, seaweeds, the vegetables from *Cruciferae* family, onion and garlic, *ibid*. Goitrogens in water are associated mainly with bacteria or shale- and coal derived pollutants: phenolics, sulfurated organics, phthalates, pyrirdines, PAH, other hydrocarbons, and organochlorines, *ibid*. There are no studies from Denmark evaluating the presence/absence of these goitrogens in combination with the geographical variation of I in the drinking water. However for most of the cited goitrogens in the water, quality controls at the water supply systems exist and are given in (Miljøministeriet 2011). In the review by Gaitan (1990) few other chemical elements are mentioned as possible goitrogenic factors in water: high Ca and Mg salts, Li. The effect of smoking on the goitre prevalence has been debated, however based of DanThyr study, Knudsen et al. (2002) concluded that smoking is associated with increased thyroid volumes and this effect is stronger in areas with I deficiency. It is suggested that thiocyanate\textsuperscript{19} is a competitive inhibitor of I uptake and organification, *ibid*.

Laurberg et al. (2003) reviewed literature on goitrogens and humic substances in drinking water and the epidemiology of thyroid disease and concluded that humic substances in water may be goitrogenic. The most likely reason is release of chemical substances inhibiting thyroid hormone production during the intestinal degradation, however more studies are needed in order to delineate the association between different types of humic substances and thyroid disease (Laurberg et al. 2003). Andersen et al. (2008) estimated 85% bioavailability of the I associated with humic substances from Skagen\textsuperscript{20} by using a very simplified algorithm:

\[
\text{bioavailability} (\%) = \frac{(I_{\text{Skagen}} - I_{\text{Randers}}) \cdot (UI_{\text{Skagen}} - UI_{\text{Randers}})^{+1.21+0.9}}{(UI_{\text{Skagen}} - UI_{\text{Randers}})} \cdot 100\% , \text{where}
\]

\( (I_{\text{Skagen}} - I_{\text{Randers}}) \) is the difference in the I concentration found in the tap water of Skagen and Randers;

\( (UI_{\text{Skagen}} - UI_{\text{Randers}}) \) is the difference in the median UI excretion.

The tap water I difference is multiplied by average daily water intake (1.2l) and by 0.9, because the UI excretion is approximately 90% of the total I excretion. The studied population is the same as in (Andersen et al. 2012): n=212, born 1920 for Randers and n=218, born 1918-1923 for Skagen, and for this calculation participants taking I supplements were excluded.

Even though calculated on a population level, the used algorithm is omitting the I obtained by milk consumption (and other food products) and completely neglects the possibility of goitrogens present in the milk or in other food products. This, together with the limitations listed by the authors, is the reason why the title-statement\textsuperscript{21} and the

\textsuperscript{19} A degradation product of cyanide from tobacco smoke (Knudsen et al. 2002)

\textsuperscript{20} Andersen et al. (2002) stated that I is associated with humic substances

\textsuperscript{21} “Naturally occurring iodine in humic substances in drinking water in Denmark is bioavailable and determines population iodine intake”
outcome of this study should be cited carefully. Further investigation is needed in order to conclude about the I bioavailability from water.

SUMMARY
Based on the presented overview, few topics were found to be perspective for future research due to lack or very little existing data:

- I speciation data for Denmark and spatial variability in both concentration and speciation: patterns of distribution and influential factors;
- Temporal variation of I concentration and speciation; long term time series are needed in order to evaluate the stability of I throughout the year(s) and how different environmental factors are influencing it (e.g. rainfall events); how the iodine intake is influenced by the combination of eventual I variability and milk seasonal variability;
- Bioavailability of I and I species;
- Combined effect of I and other chemical elements in the drinking water and the implications for human health (e.g. water hardness, Li, Fe, Se, etc.);
- Recognising population groups with potential insufficient intake based on the variability of I concentration and speciation;
- Correlation between I intake from water and registered cases of IDD on municipality scale for whole Denmark;

The background information presented in this report is far from complete review of all I relevant sub-topics, therefore the list with topics for further research is incomplete. However, such small introduction was needed in order to serve as bridge between geoscience and public health fields and to facilitate the discussion of the future work on this PhD project. Summary of the project parts, the present moment status, as well as some plans and the reasoning behind them are given next.

1ST PART OF THE PHD: MULTIVARIATE ANALYSIS OF HISTORICAL GROUNDWATER DATA

Aim: The aim of this investigation of I in Danish groundwater is to study spatial variations including covariance with other geochemical variables, due to multiple sources and/or other governing processes. The main two objectives are to identify 1) which are the factors (both I sources and key processes) governing I content in Danish groundwater and 2) in what way the I content is influenced by these factors.

Materials and Methods: Groundwater data from the Danish public nationwide geological and hydrological database, Jupiter (GEUS 2011) was used. We explored and quantified the specific relations between I and 19 other major and trace elements and quantitative well characteristics such as geology, distance to coastline and distance to major faults by using combination of multivariate data analysis (Principle Component Analysis and Partial Least Squares regression) and different traditional tools (redox type; saturation indexes; Na/Cl ratio; and Cl/Br ratio). For further details see Appendix B.

Results, Discussion, and Conclusion are presented in full detail in the submitted for publication manuscript (Appendix B).

Based on the performed analysis and on the complexity of the I variation in the groundwater it was reasonable to conclude that there may be more than one source or process contributing to the total I concentration in the
groundwater even in one single location. Studying the I variation on a smaller scale would most probably give better results.

2nd PART OF THE PhD: CASE STUDIES

The general objective of this part of the project is to continue our previous study and to find which processes and sources are governing the I content in Danish groundwater and drinking water and in what way. Skagen, Randers, Ishøj, and Stevns have been selected as potentially interesting areas. The main difference between the previous study and this one is the existence of better control over the data quality – the sampling and the lab analysis are conducted as part of this project. The smaller scale, compared to the previous national scale study, will allow a more comprehensive analysis. And most importantly, I speciation is included in the analytical program. The sampling campaign was during October-November 2012. Lab results are expected in the very near future. Here only short site descriptions as well as specific goals for the chosen sites are presented. The case studies from Randers and Skagen share similar specific goals, while for Ishøj, and Stevns two different hypothesis are going to be tested.

SITE DESCRIPTION

Skagen
The Danish town Skagen is supplied with drinking water by Skagen Vandværk (Skagen treatment plant) which is part of Frederikshavn Vand A/S. The drinking water processed in Skagen treatment plant originates entirely from groundwater. Depending on the consumption and on the chosen pumping strategy, the groundwater is pumped from varying number of wells located at two well fields. At the top consumption periods groundwater abstraction is from ten wells at well field East and from 21 wells at well field West. After the abstraction the raw groundwater undergoes aeration, chemical oxidation, pH adjustment, flocculation, mixing, bio-filtration and gravitational sand filtration, and finally, UV treatment. The effluent composition is monitored continuously by on-line temperature, pH, dissolved O2, and turbidity sensors. In order to supply the consumers with water with relatively constant high quality, the pumping rates of the active wells and/or the chemical treatment can be adjusted.

Raw groundwater samples were taken from four wells at field West (DGU 1.267, 1.268, 1.269, 1.270) and one drinking water sample - from the pure water tank.

Randers
Four waterworks in the vicinity of Randers and Randers Fjord were selected: Oust Mølle, Bunkedal, Østrup Skov and Vilstrup, which are part of Verdo Vand A/S. The drinking water originates entirely from groundwater as in Skagen, however no sophisticated treatment is needed in Randers. Each of these four waterworks treats the raw groundwater; Oust Mølle has the simplest system: the groundwater is aerated by injecting high-pressure O2 at the water pipe supplying consumers, whereas the other three have aeration and filtration facilities.

Raw groundwater samples were taken from four wells at Oust Mølle (DGU 68.585, 68.583, 68.541, 68.554), three at Bunkedal (DGU 69.223, 69.318, 69.562), three at Østrup Skov (DGU 68.185, 68.186, 69.207), and four at Vilstrup (DGU 69.213, 69.215, 69.523, 69.558). From each of the four waterworks a drinking water sample was taken too.

Randers and Skagen have been chosen because 1) they represent two very different areas from geological and hydrogeochemical point of view, and 2) some epidemiological (endocrinological) studies for these areas already exist.

Stevns
Stevns-2 (DGU 218.1945) is a borehole located at the abandoned limestone quarry Boesdal. The borehole is drilled to a depth of 350m, penetrating 4m of Danian limestone and approximately 300m of thick Maastrichtian chalk before it terminated in Upper Campanian chalk (Stemmerik et al. 2006). The sampling was designed taking into account data on the gradient of pore water Cl- in depth published by Bonnesen et al. (2009). Significant increase in the salinity starts at
ca. 30mbls\(^{22}\) from where the Cl\(^{-}\) gradient decreases downwards gradually from 200 to 50ppm/m until the depth of ca. 100mbls; bellow that depth the observed concentration is almost at a constant level (Bonnesen et al. 2009).

Groundwater samples were taken from Stevn-2 at three different depths: 130, 75, and 20m below terrain. The sampling depth was restricted by some technical issues with the used pumping equipment.

**Ishøj**

Few wells from the groundwater monitoring program (GRUMO) were chosen at Ishøj municipality located approximately on a transect running from the Køge bay inland. Some of the wells in this area were with the highest I concentrations (>100 µg/l) found in the Jupiter dataset (1\(^{st}\) part of the phd). We measured I\(^{-}\) concentrations at the same locations chosen for this sampling campaign exactly an year earlier for testing the analytical methods. Those very preliminary results showed decrease in the I\(^{-}\) concentration in shore-inland direction (from 84 µg/l to 13 µg/l). It was decided to include Ishøj in this sampling campaign too, because now it is possible to measure I\(^{-}\), IO\(_3\), Tl and all the major ions and cations, which will allow us to better understand the reason for these elevated I concentrations.

Groundwater samples were taken from four wells (DGU 207.2827, 207.2820, 207.2826, 207.3004). Two samples were taken at a different depth from 207.2826.

**OBJECTIVES**

The specific objectives for Skagen and Randers are:

- To compare and explain the I concentration and speciation in the groundwater and its spatial variation on different scales:
  - Between the wells at one waterwork;
  - Between the wells from different waterworks in one area (Randers);
  - Between the wells from different areas (Skagen vs. Randers).
- To compare and explain the differences/similarities between I concentration and speciation in raw groundwater and drinking water (for both Skagen and Randers);
- To elucidate the I sources and the governing processes by using hydrogeochemical tools (and MVDA).

Additionally, the temporal variability of ground and drinking water quality can be evaluated based on the existing historical data from Jupiter for chosen characteristic elements. That way I temporal variability can be assessed indirectly.

Analysing the results from Stevn and Ishøj will lead to an understanding of the factors responsible for the elevated I concentrations in the groundwater. Two hypotheses about the I origin are going to be tested as part of this study: the elevated concentrations in the groundwater are due to 1) seawater intrusion at the aquifer (Ishøj); 2) upward transport (advection) of old saline water (Stevns).

**3\(^{rd}\) PART: NATIONAL STUDY**

The general objective of this national study is to obtain scientifically useful data for the I concentrations and speciation in the Danish drinking water and its spatial distribution and variation. A large sampling campaign will be conducted (around 180 waterworks) and all the lab analysis will be handled by the Inorganic Chemistry Lab at GEUS. According to the plan so far, the sampling will be in the period April-July 2013. At the present moment the focus is on the design of the sampling campaign and choosing the sampling points, so the project goals can be fulfilled in economically optimal way. Here a very preliminary version of the sampling design is presented in order to allow us discussing it as part of

---

\(^{22}\) mbls=meters below sea level; the terrain cote of Stevens-2 is 7mbls, which means that 30mbls=37m below terrain.
the qualifying exam. It should be stressed on the fact that this study will give an overview on the I speciation in Danish drinking water for the first time. It will allow concluding on the spatial variation of I and confirm/oppose previous findings, questioned because of the sampling campaigns design. In general, the I concentration distribution across Denmark in such scale will give a more detailed view and will allow optimizing the estimation of the dietary I intake from water in Denmark. It will be possible to evaluate the correlation with other chemical elements and to study the possible combined effect by distinguishing between types of water or regions where it is present and to assess the relevance for I deficiency.

CRITERIA FOR CHOOSING SAMPLING LOCATIONS

The algorithm for choosing around 180 waterworks from more than 2500 active public and private waterworks can be divided in two steps. First, a dataset with information about all registered public and private waterworks, their addresses, geographic coordinates and annual abstraction volumes was prepared. The data was extracted from Jupiter on 06th of December 2012. The size of the waterwork is evaluated based on the reported to Jupiter annual groundwater abstraction volumes for 2010, not by supplied area or population. This choice has been made because of the lack of information about supplied area or population in the Jupiter database; it is hard or impossible to find this information in internet too. The second step is manual selection of waterworks in ArcMap based on 3 criteria:

a. The largest waterwork in each municipality is selected;

b. A grid cell (20x20km) is prepared and for each cell the largest waterwork is selected (for cells where there were no waterworks selected at a.);

c. Some of the smallest waterworks chosen by criteria a or b can be substituted for the waterworks abstracting more than 1.05 mio m$^3$/year (could be >1.2, 1.5, or 2 mio m$^3$/year too, see further for details) left out of the selection because of the used criteria.

With the design of this sampling campaign we want to make sure that whole Denmark is covered on municipality level, additionally we want to have an even geographical distribution of the sampling points throughout the country. By using the combination of municipality polygons (n=99) and grid cells (n=189, 20x20km) as base for the design, 181 waterworks were selected, Figure 9. From these waterworks 27 have abstraction volumes for 2010 smaller than 0.2 mio m$^3$. On the other hand, if the 3rd criteria is used and all left waterworks with abstraction volumes above 1.05 mio m$^3$/year are selected, 30 more waterworks will be added to the list (n=24 for >1.2, n=13 for >1.5, and n=7 for >2 mio m$^3$). Therefore, the criteria can be optimised after discussion and some of the smallest waterworks can be exchanged with the largest left out of the initial selection due to the first two criteria. Additionally, there are two municipalities (Syddjurs and Stevns), where no waterworks were chosen, because no abstraction volumes have been reported for 2010 (or 2011). In order to fill in this gap, the abstraction volumes from the period before the change in the administrative division in Denmark (around 2007) can be used, which would add 2 more waterworks to the list.
As it is seen from Table 5, the chosen waterworks (by 1st and 2nd criteria) yielded about 40% of the total groundwater abstraction for 2010. This percentage can be increased if the 3rd criteria is used. The selection of waterworks, shown on Figure 9, and the chosen criteria for selection of waterworks are preliminary. After all participants in the project have agreed on the criteria and selection, all chosen waterworks (around 180) will be contacted for a sampling permit. Additionally, a simple questionnaire about the water treatment procedures (type, last change – when, how), the volume of the pure water tank, the proportion produced water supplied to the population (some of the waterworks supply industry and other customers), the supplied areas (neighbourhoods, villages, cities), and the catchment areas will be sent or given to waterworks authority.

The sampling campaign will consists of taking samples from the treated drinking water at the waterwork.
**DATA ANALYSIS**

After the samples are analysed in the lab in GEUS, the data analysis will consist of:

- Mapping I, IO$_3^-$, Ti and organic bound I in Danish drinking water with ArcMap. Can be done on municipality level or by interpolation (Kriging or other), or by taking into account the information about the catchment areas or the populated areas supplied by the waterworks, depending on the outcome of the questionnaire;
- Investigating the connection between I speciation and the aquifer type/geological setting. It does not make sense to do this for waterworks which mix water from very different aquifers; additionally, the connection between I speciation and treatment procedures can be evaluated.
- Correlation analysis (spatial correlations in ArcGIS or/and PCA and PLS) between I and other geochemical data, obtained during the national sampling campaign;
- Investigating the possible correlation between the data obtained by this sampling campaign and epidemiological data for chosen I related diseases (on a municipality level). Epidemiological data is not available at the moment.

The sampling and the analysis programs for this national study are not yet finalized. The previously presented ideas and plans reflect our wish to fill in the missing information about I in the water Danes drink and to find out if there is need for optimising the estimates of I intake from water in the Danish population by comparing with previous findings. The regional variation in the I intake from water can be used to estimate the risk for ID on a smaller scale than it was possible before (due to the extensive sampling). Comparing with previous national campaigns in Denmark (K. M. Pedersen et al. 1999; Rasmussen et al. 2000; Saxholt et al. 2008; Andersen et al. 2002) this one has the advantage of better representativity and/or higher resolution.

**PLAN FOR THE REMAINING STUDY PERIOD**

<table>
<thead>
<tr>
<th>Code</th>
<th>Type of task</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Qualifying exam preparation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Logistics plan for National campaign</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Work GEUS 280 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sampling preparations: contact, lab, reservations etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sampling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lab analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teaching: Klim fieldtrip</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Results, Discussion, Conclusions: Case studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MedGeo 2013, poster prep, travelling etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Holiday</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stay abroad 2-3months + phd course 4ECTS needed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Results, Discussion: National campaign</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Writing articles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EGU, prep, travelling, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>phd thesis writing; hand in 31-07-2014</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend**

- Second part of project: Case studies
- Third part of project: National sampling
- Work in other scientific environment, incl. teaching
- Conferences
- Exam preparations
- Holiday

**FIGURE 10 TIME SCHEDULE FOR THE REST OF THE PHD PERIOD**
The work on the 2\textsuperscript{nd} and 3\textsuperscript{rd} part of my project will continue in parallel. First of all, I’ll be working on the design of the national sampling campaign and conducting the sampling itself. After that I’ll be focusing on the analyses of the results, the discussion and conclusions from the 2\textsuperscript{nd} part (the case studies), which should result in a peer-reviewed publication and presentation on the MedGeo 2013 conference. After the lab analyses from the national sampling (3\textsuperscript{rd} part) are ready, I’ll be focusing on results presentation, discussion and conclusion, as well as the writing of a paper. The results from this study and the overall conclusions from my PhD can be presented at EGU 2014. The PhD thesis must be handed not later than 31\textsuperscript{st} of July 2014. It will consist of at least 3 peer-review papers (submitted or in preparation) with working titles:

- “High-Iodine Danish Groundwaters – a Multivariate Data Analytical Survey” – already submitted in *Environmental Geochemistry and Health*;
- “Sources and Speciation of Iodine: Four Danish Case Studies”;
- “Speciation of Iodine in Danish Drinking Water – a Nation-wide Assessment”;
- “Estimation of Dietary I Intake from Water and Assessment of the ID Risk on Municipality Level” (tentative).

As it was mentioned in the *Objectives and Scope of the PhD*, our wish is to establish a link between geoscientist and public health specialists. We are open for collaboration and suggestions on how this plan can be optimised in order our results to be further used as part of medical geology studies. The national campaign is still in its planning phase and it can be adjusted if need be.

Next to the work on my project, in order to successfully finish my PhD I need to have teaching/working experience equal to 280 hours per year. One third of these hours (roughly two months: February and March 2013) will be spend in GEUS on work not related to my PhD project: arsenic in the Danish groundwater, drinking water, and tap water. Additionally, for two weeks in May 2013 I’ll be most probably an instructor at the field course in Klim. A stay abroad is recommended, however it has not been planned yet. It could be combined with a short PhD course or summer course (4 more ECTS points are needed).


**APPENDIX A: ACTIVITIES DURING PART A**

**COMPLETED COURSES:**

- “The Hydrogeochemical Cycle”, Department of Geoscience, Aarhus University – 10 ECTS
- “Geoscience Data Analysis”, GEUS – 5 ECTS
- “Searching databases of scientific literature” GSST, Aarhus University – 1 ECTS
- “Groundwater sampling course”, School of Technology and Business, VIA University College – 1 ECTS
- “Representative sampling of Heterogeneous System in Science, Technology & Industry: Theory of Sampling (TOS)”, Faculty of Energy and Science, Aalborg University (held in GEUS) – 5 ECTS
- “Health and Earth – Building a Safer Environment”, short course on medical geology sponsored by IMGA and SESEH 2012 (held in Galway, Ireland) – 2 ECTS
- “Scientific Writing and Presentation”, GSST, Aarhus University (held in Fuolum) - 2 ECTS
- (planned) Other project specific short course (either advanced ArcGIS, spatial geostatistics, or epidemiology)

**TEACHING ASSISTANCE**

- “Hydrologi, jordbund og dynamisk geomorfologi” – Q3&Q4 2012
- “Hydrologisk og Geomorfologisk Feltkursus, Klim, 2012” – 2 weeks

**CONFERENCES AND MEETINGS**

- *Workshop on Medical Geology*, Copenhagen 7\textsuperscript{th} February 2012 (oral presentation + published abstract)
- *ATV Jord og Grundvand Mødenr. 85*, Gentofte 19\textsuperscript{th} June 2012 (oral presentation + published abstract)
- *SESEH 2012 Sino-European Symposium on Environment and Health*, Galway, Ireland 20-25\textsuperscript{th} August 2012 (oral presentation + published abstract)

**APPENDIX B: SUBMITTED FOR PUBLICATION PAPER**

Denitza D. Voutchkova, Søren M. Kristiansen, Birgitte G. Hansen, Vibeke Ernstsen, Brian L. Sørensen, Kim H. Esbensen

“High-Iodine Danish Groundwaters – a Multivariate Data Analytical Survey”, Journal Environmental Geochemistry and health (on the next page – submitted paper draft)
High-Iodine Danish Groundwaters –

a Multivariate Data Analytical Survey

Denizta Dimitrova Voutchkova1,2, Søren Munch Kristiansen1, Birgitte Guldborg Hansen2, Vibeke Ernstsen3, Brian Lyngby Sørensen2, Kim H. Esbensen3,4

1Department of Geoscience, Aarhus University, Høegh-Guldbergs Gade 2, DK-8000 Aarhus C, Denmark

2Geological Survey of Denmark and Greenland (GEUS), Lyseng. Allé 1, DK-8270 Højbjerg, Denmark

3Geological Survey of Denmark and Greenland (GEUS), Øster Voldgade 10, DK-1350 Copenhagen K, Denmark

4ACABS research group, Aalborg University, campus Esbjerg, DK-6700 Esbjerg, Denmark

Phone: +45 87 15 62 98
Fax: +45 89 42 94 06
e-mail: ddv@geo.au.dk

Abstract

In areas where groundwater is a major source of dietary iodine (I), the I concentration in groundwater is an important factor for public health and epidemiological understandings. The present study performs a multivariate analysis of all available historical Danish I groundwater data from 1933-2011 to investigate source(s) and geochemical processes involved in elevated I concentrations. We here explore and quantify the specific relations between I and 19 other major and trace elements and quantitative well characteristics such as geology, distance to coastline and distance to major faults. Principal Component Analysis on a maximum set of validated co-joint data (506 samples x 20 variables) revealed three distinguishable sample clusters together with a non-descript background population of samples. An augmented analysis (Partial Least Squares Regression, PLS-R) based on the resolved clusters only, with I as Y variable, constitutes our main data analysis; the validated PLS-R model explains 75% of the I variation. Both positive and negative correlations between I and the suite of other influential variables are found. On this basis we interpret four governing geological and geochemical factors (processes) related to the carbonate system, redox conditions, distance to likely I sources, and ion exchange processes. Sample clustering due to geology is not substantial, and it appears likely that more than one source or process is contributing to the total I content in Danish groundwaters. We argue that sediment desorption, sea water intrusion, old saline water diffusion or organic matter breakdown/diagenesis are likely influencing factors.

Keywords: iodine; groundwater; Denmark; multivariate data analysis; PCA; PLS-R;

Introduction

Iodine is an essential component of human thyroid hormones regulating the metabolic processes in cells and playing role in the early development of most organs. Too low I intake, i.e., <150 µg/day for adults, can result in a variety of Iodine Disorder Diseases (IDD) as mental retardation, goitre,
and hypothyroidism, while chronically high intake can also cause health problems such as I
induced hyperthyroidism, elevated goitre, and subclinical hypothyroidism (WHO 2007).

Iodine in human body originates mainly through dietary intake including water; or through
inhalation of atmospheric I (minor contribution), however air I concentration is low: 10-20 ng/m³
(Hou 2009). Drinking water is generally not considered a significant major contributor to the
dietary I intake as it usually provides close to 10% only (Fuge 2005). However, in Denmark,
where drinking water originates from groundwater, 25% of the I intake in the average diet was
derived from drinking water and other beverages (some containing high proportions of processed
groundwater) prior to recent mandatory I fortification of table salt (Rasmussen et al. 2000). In such
areas where groundwater is a major dietary source a thorough understanding of I concentrations in
groundwater is of importance for understanding epidemiology.

The I geochemical cycle is a spatially and temporally dynamic system, where different biotic and
abiotic processes are taking a part (Figure 1). In the hydrogeochemical cycle, I is predominantly in
oxidation state -1 (iodide, I⁻) and +5 (iodate, IO₃⁻); and, next to these inorganic forms, I is also
found in different organic species, which can also contribute a significant fraction of the total I (Hu
et al. 2009).

Figure 1.

Iodine is one of the most abundant micronutrients in seawater, with total concentration of 50-60
µg/l (Ito and Hirokawa 2009) while for coastal waters around Denmark was found to vary
seasonally with averages from 32 to 89 µg/l total I (Truesdale et al. 2003).

Fuge (2005) found 2 µg/l total I in the precipitation from the interior of the UK with up to 5 µg/l
for coastal precipitation. Similar values are found in rain water collected from Germany, averaging
2.2±0.8 µg/l (Gilfedder et al. 2009) and Denmark, range 0.78-2.70 µg/l (Hou et al. 2009).

Contents of I in Danish aquifer sediments have not been studied but marine deposits are generally
found to be enriched in I relative to terrestrial sediments, i.e. I content decreases in the order: deep
sea clays (3.9 mg/kg) > organic C rich shales (near-shore, 0.2-6.2 mg/kg) > limestone (near-shore,
2.5 mg/kg, which is 1/10th of the I found in deep-sea carbonates, (Muramatsu and Wedepohl
1998)) > sandstone (123 µg/kg) > magmatic rocks (4-8 µg/kg), based on Muramatsu and

From the study on 884 European bottled spring and mineral waters, I concentrations were found
to cover three to four orders of magnitude, from <0.2 µg/l to 4030 µg/l with median 4.78 µg/l
(Reimann and Birke 2010). Many of the high values were traced to deep formation waters; while
one I rich bottled water from Norway was influenced by water from a recent marine clay deposit
(Reimann and Birke 2010). In addition, some of the highest I concentrations in groundwaters
reported globally with 129 ± 3 mg/l, 2000 times higher than the seawater concentration level, are
found in Japanese wells influenced by brine (Muramatsu et al. 2001).

Existing data on I in Danish ground water are limited to studies based on very few sampling
locations: tap water from 55 (K. M. Pedersen et al. 1999), 40 (Rasmussen et al. 2000), and 47
locations (Saxholt et al. 2008), and drinking water from 22 waterworks (Andersen et al. 2002).
This contemporary data set shows that I concentration in the drinking water in Denmark varies
from 0.7 µg/l to 140 µg/l. However there is one order of magnitude difference between the highest
values reported by K. M. Pedersen et al. (1999), Andersen et al. (2002), Saxholt et al. (2008) and
the one found by Rasmussen et al. (2000). According to Andersen and Lauberg (2009) there is a
distinguishable difference between East and West Denmark with respect to I content in drinking
water. Even though such regional variations of I content in the Danish drinking water are known
there have been no detailed geochemical studies related to I.
The aim of the present investigation of I in Danish groundwater is to study spatial variations including covariance with other geochemical variables, due to multiple source(s) and/or other governing processes. The main two objectives are to identify 1) which are the factors (both I sources and key processes) governing I contents in Danish groundwater and 2) in what way the I content is influenced by these factors.

A significant historical database exists for Danish groundwater which allowed us to perform extensive data mining and to use a combination between multivariate data analysis (MVDA) and traditional hydrogeochemical analytical tools in order to further our objectives. Principal component analysis (PCA) and partial least squares regression (PLS-R) were used for their suitability for simultaneous multivariate data description and structure exploration, notably for the possibility of discovering grouping of samples and/or variables which may otherwise be swamped in individual sampling and analytical errors if only studied univariately (full-spectrum advantage) (Esbensen 2010); MVDA is a strong approach also because it handles large as well as small data matrices with equal ease, ibid.

**Methods and Materials**

**Data**

*Iodine in groundwater*

The groundwater data used in this study is taken from the Danish public nationwide geological and hydrological database, Jupiter (GEUS 2011). Data was downloaded from Jupiter on 24th of November 2011. Groundwater data containing I analyses (n = 2562) was from the period 1933 to 2011; the I concentration varies from below detection limit to 1220 µg/l (see Figure 2). For 90% of the samples I concentration is below 20 µg/l and only 11 samples have I above 200 µg/l. The median value is 5.4 µg/l and the average I concentration in the groundwater is 13.83 µg/l.

**Figure 2.**

*Other groundwater constituents and well information*

For the purpose of the multivariate study it was interesting to use the available data not only for I, but also for possible co-explanatory geochemically and geologically relevant major and trace elements and parameters. Figure 3 presents all such chemical data available and the number of samples analysed for iodine and the following relevant geochemical parameters: aggressive CO₂, B, Ba, Br, Ca, CH₄, Cl, Fe, F, H₂S, HCO₃⁻, I, K, Li, Mg, Mn, Na, NO₃⁻, O₂, PO₄³⁻, SO₄²⁻, Sr, conductivity, NVOC, pH, and redox values.

Other information used in this study and gathered from Jupiter: location (where the sample was extracted, i.e. well ID number and geographic coordinates, screen number, screen depth, top and bottom); when the samples were collected (date); predominating geological setting at the screen (given with % from the length of the screen and type of geological setting); analytical method used (for each geochemical parameter); samples where concentration is below detection limits (assigned “< x”, where x is a specific detection limit, see further below); objective of analytical determination.

**Figure 3**

*Distance to coastline and to major faults*

Two variables used in the MVDA are not provided by Jupiter: “distance to coast line” and “distance to major faults”. The “Near 3D” tool of ArcMap10.0 (ESRI 1999-2010) was used to calculate the smallest horizontal distance from each sample location (well X,Y coordinates...
provided from Jupiter dataset) to both the coastline and/or to the nearest major geological fault lines. The distances to the present coast line are calculated as distances to the closest sea or fjord, after a specific data control was made, and all derived distance values were manually checked for calculation mistakes due to complex coastline shapes. The distance to faults was measured as distance to the closest of the major fault lines described by Håkansson and Pedersen (1992).

**Dataset preparation**

1st step: Master data set preparation

After the raw Jupiter data were extracted, a master data set (MDS) was prepared which consisted of 28 variables (iodine, 25 elements and parameters, distance to the coast, and distance to major faults) for a total of 2562 objects. Extensive data quality assurance (including detection limit handling) and further dataset preparations (variable and object exclusion) were needed in order to prepare this data set for MVDA, which resulted in necessary reductions (samples/variables).

2nd step: Data quality assessment

The supplementary MDS information (Id well number, sample number, screen number, X and Y coordinates, laboratory methods, sampling date, geology, detection limits, purpose etc.) was used for the initial data quality evaluation. The diversity of I data qualities due to different analytical methods, as well the possible errors in the database entries with respect to I speciation, are discussed in full detail in Supplementary materials 2 and 3. Details about the handling of values below detection limit, which represent 8% of the I analyses (n=216), are given in Supplementary materials 4. Finally, the refined MDS was checked for gross outliers by manual exploration for unusually high values. As a result a few additional errors were also eliminated (Supplementary materials 5).

3rd step: Reduction of MDS

As seen from Figure 3, the number of analyses for the potentially covarying other 25 elements vary from 65 (O₂) to 2131 (B). It is obvious that the requirements for proper MVDA should be that as high a proportion of samples contain non-missing information for all (or mostly all) chemical variables and parameters. Thus the original data set must be reduced so as to present a suitable low proportion of missing values, which should be as randomly distributed over the data matrix as possible. While this configuration is desirable, in practise it is very much the existing parameter coverage in the database that brackets such endeavours. Thus reduction of the MDS is made in two steps: 1) reduction of variables (analytes), i.e. deletion of variables with a too high proportion of adversely distributed missing values (major, tracer elements, and field measurements) followed by 2) reduction of remaining objects (samples) still with a too high proportion of missing values. In the first step, eight variables had to be excluded (Agg.CO₂, CH₄, F, H₂S, NVOC, O₂, PO₄, redox values), based on the insufficient number of co-existing analyses, Figure 3. However there were still too many missing values in the data matrix, so further reduction of the objects also had to be performed, with the primary objective of keeping as many as possible of the high(er) I values. All MDS reductions were made according to well-established general chemometric rules and procedures (Esbensen 2010).

This clean-up results in a reduced MDS (r-MDS), which consists of 20 variables (B, Ba, Br, Ca, Cl, Fe, HCO₃, I, K, Li, Mg, Mn, Na, NO₃, SO₄, Sr, conductivity, distance to coastline, distance to faults, pH) and 505 objects, representing 20% of the original number of objects. Although an apparently dramatic reduction in the number of analyses (2562 → 505) available for MVDA, it is but a necessary reflection of the very disparate analytical coverage displayed in the historical data base and the resulting covariance matrix coverage, Fig. 3. For the purpose of a first foray overview of the fullest multivariate data structure however, r-MDS is the optimal co-joint data set that can be gleaned from the Jupiter data base.
Multivariate data analysis

All column data in r-MDS (variables) were centered and scaled, i.e. centered on the empirical average value, and weighted by the factor 1/STD before analysis; this compound procedure is necessary in order for all variances to influence the correlation equally (Esbensen 2010). All MVDA was performed using the Unscrambler 10.1 software package (CAMO 2011). The Non-linear Iterative Projections by Alternating Least-Squares (NIPALS) algorithm is used because it can handle a reasonable proportion of missing values in the X matrix as well as in [X,Y]. For detailed explanation on the methods NIPALS, PCA, and PLS-R, see (Esbensen 2010) and further background references herein. The MVDA was handled in three-step approach, due to the specifics of the data organisation, as laid out in Figure 4.

1st step: r-MDS

Initially a full Principal Component Analysis (PCA) of the entire r-MDS was made. The objective was to explore the overall data structure, with focus on delineating empirical sample groupings and clusters. Based on this first analysis the r-MDS was divided in two sub-sets, which were further analysed individually. The data structure of this 1st model and the sub-division of r-MDS into two sub-sets (sub-set 1 is the noise part, or so called “background population”, while sub-set 2 is the structured part, consisting of the three significant clusters observed) are discussed in detail in Results.

2nd step: sub-set 1 (Background population)

PCA on sub-set 1 of r-MDS, which now consist of 20 retained variables and 453 objects, was carried out in order to check for the possible existence of “hidden structure” in the background population. This matter is further discussed in Results.

3rd step: sub-set 2 (Significant clusters)

A PLS-R model with 52 objects (samples) and I as the Y variable (X matrix consisting of the complementary 19 variables) was carried out with the purpose to characterise the functional X-Y relationships, i.e. detailed examination of the correlation relationships between X and Y as a means for in-depth hydrogeochemical interpretations. Six samples were identified as outliers on the T-U plots and deleted from the final PLS-R model (Table 3, supplementary materials), and as a result the dataset size was reduced to a final 46 objects (samples), which covered essentially almost all high-I samples and the largest possible suite from the lower concentration levels. These samples are mapped geographically as shown on Figure 5.

Optimal data model test set validation (Esbensen and Geladi 2010) was not possible in this case due to the low number of objects (effectively 46), but since the objective was X-Y interpretation, not prediction, the model could safely be validated using systematic cross-validation (123123) with 2 segments, performed on a Y-sorted matrix, in accordance with the general ground rules of validation (Esbensen and Geladi 2010), (Esbensen 2010). This PLS-R model is also further discussed in Results.

Geological settings

Most danish aquifers consist of Quaternary or Miocene sand and/or gravel, Palaeocene to Late Cretaceous chalk/limestone (Kelstrup et al. 1982). The groundwater samples from sub-set 2 (Figure 5) represent only a relatively few of the Danish aquifers, however, due to the reduction procedures outline above. These samples (n=46, without the six outliers) originate from six
distinct geological settings: Late glacial melt water sand, Campanian-Maastrichtian chalk, Danian chalk, Selandian chalk, Selandian sand, and limestone (age not specified). Further in this study we grouped them in two categories with respect to the geological period (epoch): Quaternary, and Late Cretaceous or Paleocene, hereafter denoted Pre-Quaternary.

**Geochemical indicators**

The results of sub-set 2 PLS-R model were interpreted from a hydrogeochemical point of view with the help of different traditional tools: redox type; saturation indexes; Na/Cl ratio; and Cl/Br ratio. All samples from sub-set 2 were grouped by these parameter categories and visualised on the so called *score plots (SP)* of the PLS-R model (see Results below).

**Redox type**

Redox type is not included as variable in the MVDA, however the redox state of the groundwater samples in this study can be *estimated* based on three other variables: NO₃, Fe and SO₄. This is achieved by using an algorithm for determination of redox water type (B. Hansen et al. 2009). Based on the available data we can differentiate between four classes: 1) “A” and “B” redox type – groundwater from oxic or anoxic nitrate reducing zone (O₂ is not included as variable in r-MDS, thus it can’t be distinguished between oxic and anoxic zones); 2) “C” type – slightly reduced groundwater (iron and sulphate zone); 3) “D” type - strongly reduced (methane and hydrogen sulphide zone); 4) “X” type – mixed water (i.e. samples which do not follow in the first three classes).

**Saturation Indexes (SI)**

Saturation index (SI) of some minerals were calculated for each of the water samples of sub-set 2 (n=46) with the software PHREEQC (Parkhurst and Appelo 1999). Groundwater temperature is fixed at 9ºC and the default pe (pe=4) is used due to a general lack of field measurements. SI was calculated by using the *lnl.dat* database. Included solid phases: aragonite and calcite (CaCO₃); dolomite (CaMg(CO₃)₂); magnesite (MgCO₃); siderite (FeCO₃); strontianite (SrCO₃); witherite (BaCO₃); goethite (FeO(OH)); barite (BaSO₄); celestite (SrSO₄); and gypsum (CaSO₄·H₂O). SI is given by

\[ SI = \log(IAP/K) \] (Appelo and Postma 2005) where IAP is the ion activity product and K is the solubility product of a given mineral. For SI=0 there is equilibrium between the mineral and the solution, SI<0 reflects undersaturation while SI>0 reflects oversaturation (Appelo and Postma 2005).

**Na/Cl ratio**

Na/Cl ratios (molar) were calculated for each of the sub-set 2 samples. By using this ratio, two different conditions are distinguished (B. Hansen et al. 2009), (Kristiansen et al. 2009): 1) reversed ion exchange (Na/Cl ratio < 0.75) when salt water infiltrates a freshwater sediment (as result from sea water intrusion, diffusion from old saline groundwater due to extensive extraction, infiltration of Na-rich water from the ground surface (e.g. road salt)); and 2) ion exchange (Na/Cl ratio > 1.25) when fresh water infiltrates a sediment of marine origin. The water samples with ratio between 0.75 and 1.25 are not representing ion exchange or reverse ion exchange: the authors (Kristiansen et al. 2009) include uncertainty interval of ±0.25.

**Cl/Br ratio**

According to Davis et al. (1998) typically Cl/Br (mass) ratios can point to the source of the groundwater. Davis et al. (1998) have summarised that: 1) for most shallow groundwaters with Cl concentration less than 5 mg/l ratios between 80-160 reflect the local precipitation sources; 2) ratios between 200 and 400 reflect connate water; 3) the sea water influence along the coastline is shown by ratios close to 290; 4) sewage has ratio between 300 and 600; and 5) groundwater affected by halite dissolution has ratios in the range 1 000 to 10 000.
Results and Discussion

Survey PCA model of r-MDS

The PCA model of r-MDS revealed complex data structures, as no less than eight Principal components (PCs) were needed in order to explain 80% of the total variance of the dataset (Figure 6c). This model is not well suited for studying the correlations between the key variables of interest since it includes both a diffusive background population as well as more well-defined structure elements (see below). However, the first three scores yielded a clear and useful delineation of three major data clusters, Figure 6a, 6b. Two clusters groups were discriminated on PC1-PC2 (PC1 and PC2 explain 38% and 10% of the data variance respectively), Figure 6a, while the third was found by PC2-PC3 (PC3 explains an additional 7% of the total variance), Figure 6b. These clusters could be distinguished from a diffuse “background population” comprised of low I concentration samples only. As our intention is to study the high I relationships, this background population is of no further use after discrimination. From Danish geology background knowledge, it is highly likely that different processes, or sources - or combination hereof - are dominating at different geographical locations, which likely contribute to the general character revealed for the full dataset. This is supported by the way samples with very different–low or high–I concentrations are found clustered geographically (Figure 2).

Thus in order to fulfil the data analysis goals r-MDS was subsequently split into two sub-sets consisting of: 1) background population, called sub-set 1; and 2) the three delineated groups of samples together, called sub-set 2; each of these were subsequently analysed separately.

Figure 6.

PCA model of sub-set 1 (background population)

PCA modelling of sub-set 1 didn’t detect discernible low-dimensional data structures. From the residual calibration variance plot (RCVp) it is seen that seven PCs are needed in order to model 70% of the X matrix variance (not presented here). The first 2 PCs represented 29% and 12% of the total X variance only. There was no obvious structure and no grouping of the data in the score plots (SP). This means that there is no simple common data structure in the low I background population (the high(er) I samples are part of sub-set 2). Parts of this model were therefore only used for non-full-spectrum comparison with the results from sub-set 2 (reported in Supplementary materials 6).

PLS-R model of sub-set 2 (significant data groups)

In PLS-R modelling of sub-set 2, the I concentration was designated as the Y-variable. In this fashion the existing X-covariance data structure is forced to correlate as much as possible with Y. This often reveals a different data structure then a PCA model, as the X-space is now decomposed both with respect to variance-maximization as well as this external regression constraint, (Esbensen 2010), the interpretation of which constitutes our major hydrogeochemical study objective.

PLS-R of sub-set 2 (n=46 after outliers exclusion) did indeed show a strong structure. Figure 6f presents the residual variance plot (RVp) both for the calibrated and for the validated model for seven factors (PCs). From this plot it can be seen that the optimal number of PLS-components is three; the model explains close to 75% of the total Y variance. Figure 6e shows the important Predicted vs. Reference Y plot from this model. The correlation coefficient (r²) between the predicted vs. reference I data (Y) is 0.76. Average prediction accuracy of the Y-modelling is measured by the slope of the fitted predicted vs. the reference regression line (0.77). These model quality measures (which both should be as close to 1.00 as possible) must be regarded as satisfactory in the context of the general complex nature of the data set and its quality features.
Based on this promising model we further explored the correlation relationships between X and Y using both inter-variable as well as inter-object (inter-sample) relationships, as detailed below. By setting up a PLS-r model with I concentration of the Y-variable, the model will reveal precisely the variables in the X-data that facilitate the largest correlation between the three PLS-components and the Y-variable; for each X-component this is maximised independently. Subsequently this part of the X-data structure is “deflated”, i.e. projected out of the remaining X-matrix, which is then set up for modelling the next largest X-Y correlation data structure (PLS-component) etc. This sequential regression is carried out only for the number of components delineated in the RVp, i.e. three PLS-components in the present case. Inspection of the X-space can now be carried out based on the first three component scores and their corresponding loading-weight manifestations only, following standard chemometric interpretation principles which are identical whether pertaining to PCA-components or PLS-components, (Esbensen 2010), (Esbensen and Geladi 2010).

This specific data modelling thus corresponds to the questions: How much of the variance in I concentration in the Danish non-background groundwater can be explained (modelled) by the characterising X-variables for a data set made up of the three discriminated clusters? And: Which are the most influential variables in X?

**Loading Weights Plot**

The mutual positive and negative correlations between I and the other 19 variables is visualised on the loading weights plot (LWp) in Figure 6d. The explained variance of the X data corresponds to 79% of the total variance (first factor: 48% and second: 31%). The first factor (PC1), interpreted as an “I factor”, is characterised by: 1) I and its positive correlation to Ba, HCO₃, Br, Na and partially Fe; and 2) I and its negative correlation to distance to coastline, distance to major faults, and pH; Figure 6d. The second factor (PC2) is characterised by several positively correlated variables: Ca, Cl, conductivity (conduct.), K, Mg, and SO₄, which are also weakly correlated along the PC1, Figure 6d.

The groundwater samples in sub-set 2 display pH ranging from 6.49 to 9.8, which compared to the whole dataset, means that the slightly acidic samples (with pH from 4.03 to 6.49, Figure 5 Supplementary materials) are part of the background population and this model therefore represents near neutral to alkaline groundwater buffered by carbonate. In this interpretation it is assumed that I content in the groundwater and I content in the sediment are functionally connected. Adsorption of I by Al and Fe oxides (and bacteriogenic Fe oxides (Kennedy et al. 2011)), as well as clay and organic matter, is pH dependent, with greatest sorption in acidic conditions (Fuge 2005). However, in this model the correlation between I and pH is negative (samples with pH< 6.49 are not included). This negative correlation between pH and I concentration in the groundwater samples does therefore likely not reflect the sorption capacity of the sediments. It may be caused by other, purely geographical factors as it is very strongly (positively correlated) to the distance to faults.

The negative correlation between I and distance to the coast line and the major geological faults suggests that high I groundwater is associated with these two features or their position. It should be remarked that the samples included in this model (n = 46) are situated relatively close to sea or fjord coast lines (1.28 - 12.40 km, mean 5.28 km) and to faults (0.35-19.90 km, mean 6.04 km).

According to the model results, I is positively correlated to Ba and HCO₃, which suggests that similar processes or sources are influencing their concentration in the groundwater samples. Additionally, the negative correlation to pH points to the possibility that the equilibrium between carbonate minerals and the groundwater is one of the important processes responsible for the I variation. According to Mokrik et al. (2009), Ba content in groundwater is mainly caused by dissolution of Ba-holding minerals, e.g. barite (BaSO₄) or wetherite (BaCO₃). Saturation indexes show that all samples (n = 46) are oversaturated with respect to BaCO₃, when SI with respect to BaSO₄ ranges from -3.22 to 1.37 (mean 0.03, median 0.45). This means that some of the
groundwater samples are oversaturated and barite may precipitate, while others are undersaturated and barite may dissolve. The SI distribution is similar for some other carbonates (CaCO$_3$, MgCO$_3$, and SrCO$_3$) – both undersaturated and oversaturated samples are present. However, all samples are oversaturated with respect to dolomite and undersaturated with respect to FeCO$_3$ (see Figure 8 from Supplementary materials).

For aquifers where groundwater recharge is mixing with seawater, Santos et al. (2011) discuss that no single process can explain Ba distribution; they describe four processes to be of major interest: mixing of sea- and freshwater, ion exchange driven sorption/desorption to/from sediments, marine organic matter mineralisation, and Fe and Mn oxide cycling. Santos et al. (2011) argued that marine organic matter mineralisation together with Fe and Mn oxide cycling are the prime processes responsible for the Ba distribution on their study site, located at coastal plain in the north-eastern Gulf of Mexico (Florida, USA). Organic matter decay and digenesis also tend to release Br$^-$ (Davis et al. 1998) and possibly I. De-halogenation of organic matter is expected to proceed in the order I, Br, Cl and the proportions of released Br and I are controlled by the amount of released DOC and the concentration of I and Br in the peat (Biester et al. 2006). Hence, positive correlation in our data set between I and Ba, Br, and HCO$_3^-$ may be due to desorption from organic matter and/or break-down of organic matter during diagenesis. Unfortunately, NVOC was not included in the dataset due to the insufficient number of co-joint analyses present; in order to confirm this interpretation more detailed studies are needed.

Figure 7.

Score Plot relationships

In addition to the model interpretations based on LWp (Figure 6d), the following interpretation is based on score plots (SP) in Figure 7. Groundwater samples (sample clusters) can be characterised with respect to different factors by using annotated versions of the basic SP (PC1-PC2) relationships. The factors used in our study are: geology type; screen depth; redox type; Cl/Br mass ratio; Na/Cl molar ratio. Samples with similar characteristics are either connected to each other so they form a polygon, or enclosed in common shaded areas on the SP.

There is groundwater from seven Danish municipalities in the 46 sample data set: Gladsaxe, Copenhagen, Frederiksberg, Ishøj, Sønderborg, Jammerbugt, and Lejre (Figure 7a and Figure 5). The polygons of Gladsaxe, Copenhagen, Frederiksberg, and Ishøj are overlying each other, showing a high degree of similarity in the groundwater composition. The similarity is not surprising due to the geographical proximity of these four municipalities, Figure 5.

When using the six different geological settings (see Methods and Materials), the grouping of the samples on the SP is not substantial (not presented here). The samples delineated with respect to the depth of the intake’s bottom did neither show any discernible pattern (Figure 7b). When the samples are annotated as originating from Quaternary or Pre-Quaternary settings, the picture is somewhat clearer – the Pre-Quaternary samples are more or less grouped together at the centre of SP (PC1-PC2). However, the Quaternary samples are spread all over the SP, Figure 7c. The largest I variation (4-194 µg/L, mean 51 µg/l) is found in the Quaternary samples (n=10), whereas the I concentration of the Pre-Quaternary samples (n=27) ranges from 10 µg/l to 172 µg/l (mean 42 µg/l). Iodine concentrations in groundwater from these two types do not differ significantly as low as well as high I concentrations are found in both types. In Quaternary sediments above marine carbonate deposits I concentrations are generally elevated according to Fuge (2005). Additional to these two categories, almost the same ranges are seen for the unknown geology samples (n=9): 4-160 µg/l, mean 55 µg/l. These results and the way the samples group on the SP implies that the I rich groundwaters included in this model are not governed only by the geology at depth of sampling, but most likely by other processes too.
The overall picture is that the location (geographical position, geology type, depth of the screen) alone do not govern the grouping of the samples (the three clusters) significantly in the SP and therefore cannot explain the correlations between the different X-variables and the I concentration. Other geochemical factors should preferentially be included in the analysis.

Redox values were not included as variable in the model; however, the redox state of the groundwater samples was estimated separately as explained in Methods and Materials. From Figure 7d can be seen that most of the samples are representing slightly (C type) or strongly (D type) reduced water. Only two samples are representing oxic or anoxic (A or B type) groundwater and eleven mixed groundwater (X type). Redox conditions (as well as pH) affect the sorption of I in soils; in reduced conditions I is strongly desorbed from the soil (Fuge 2005). Changes in the redox state in soils (from oxidised to reduced) results in shift in the I speciation, but also induces the release of both I and IO$_3^-$ (Ashworth 2009). Our model (even though totally based on groundwater and not soil-water samples) is in agreement with this general finding, since we have mostly reduced conditions in all the studied I rich groundwater. Additionally, on the LWp I showed itself to be somewhat correlated to the Fe concentration in the groundwater sample, which indicates that I concentrations may be directly, or indirectly, influenced by redox conditions.

One distinctive trend that can be observed, is revealed when annotated samples are mapped with respect to their Na/Cl ratio on the SP (Figure 7e). Most of the samples with ratio below 0.7 are plotted in the lower part of the plot (with exception of two samples), which implies that there is saltwater infiltration to the Pre-Quaternary chalk and some of the Quaternary sand aquifers. Opposite, all samples with ratio above 1.25, indicating refreshening of marine sediments, are located in the upper part of the SP.

Also the Cl/Br ratio has been used as proxy for groundwater origin (Davis et al. 1998) and thereby the total I concentration as Br and I are found to correlate. Only three samples (from late glacial melt water sand) display a ratio below 175, which are believed to reflect Cl/Br ratios in the precipitation. Three samples are with ratios above 1600 might reflect water affected by halite dissolution or even road salt for the densely populated part of Denmark (Kristiansen et al. 2011). The rest of the samples are divided in two groups. In the first group are samples with ratios from 200 to 350, interpreted to reflect sea water or connate water influences. The second group consists of samples with ratios from 400 to 700. Davis et al. (1998) suggests that such a high ratios reflect possible sewage influence; however, sewage leaking to the groundwater in Denmark is very unlikely. Two other possible explanations include: 1) mixing of groundwater with high and low ratios or 2) brake-down of organic matter or high organic matter content in the groundwater and subsequent release of Br and/or I to the groundwater (notice, that NVOC is not included in the dataset and further exploration is not possible). An interesting trend can be observed along PC2: the samples with ratios above 1600 are in the bottom part of the plot, whereas the ones with very low ratios are predominantly in the upper part of the plot. Along PC1 can be seen that almost all saltwater (seawater or connate water) influenced samples are located in the left part or close to the centre of the plot, when the middle range (400 to 700) are mostly located in the right part. Still, these trends are not all that clear - the different groups of samples are grading and overlapping with one-another.

Bonnesen et al. (2009) suggested that the present composition of the pore-water in Danish Chalk formations cannot be explained simply by an origin as trapped sea water. The Danish sedimentary basin on Zealand (from where most of the samples originate) can be divided in three sections: shallow freshwater section; underlying saline section with near hydrostatic pressure; and deep saline geo-pressured zone (Bonnesen et al. 2009). They argue that deep saline water is flowing upwards and toward the basin margin in the east throughout the network of near vertical and horizontal fractures in the chalk formation. Both the refreshening of the chalk aquifer and the deep saline water could influence I concentrations in the groundwater as indicated by both the Na/Cl
and Cl/Br ratios. Unfortunately, based on the present limited dataset we are not in a position to state which process is dominating.

To summarise, Figure 8 gives an overview on the factors interpreted to be influencing the total I variability, based on the SP and the LWp of the PLS-R model of sub-set 2.

Based on the historical Jupiter data base it is possible to derive a first overview of Danish groundwater origin(s) and the processes governing its composition by a combination of MVDA and standard hydro-geochemical concepts and tools. It is emphasised that the quality of the input data (discussed in Supplementary materials) is a crucial success factor for the data analysis and for the reliability of the possible interpretations. Gathering historical hydro-geochemical data in a public data base is relatively straightforward and easy (although very laborious), but the individual co-joint data possibilities are of highly variable individual validity, and far from all of sufficient quality for multivariate analyse. Comprehensive scrutiny of the Jupiter data base resulted in substantial reductions of the dataset when questioned from the point of view of multivariate data accessibility. Moreover, the missing speciation data for I in the database is clearly an obstacle for a complete geochemical study. Complementary speciation studies are under way and will be presented elsewhere.

In order to study the geochemistry of I in Denmark, including its geo-medical relationships as comprehensively as possible, the present study forms only a natural starting point, serving to facilitate subsequent, more detailed and more localised studies.

Conclusions

The main two study objectives, as formulated in the Introduction, were fulfilled and the conclusions are presented below:

The key factors governing the total I concentration in sub-set 2 of r-MDS were revealed (1st study objective) by interpreting the positive correlation of I to Ba, HCO$_3^-$, Br, Na and partially Fe, and the negative correlation of I to distance to the coast line, major faults, and pH. Four major factors appear to influence the total I concentrations: wiz. distance to I source, carbonate equilibrium, and redox processes (PC1), and to a lesser extent ionic exchange relationships (PC2), as summarised in Figure 8.

The 2nd objective was to elucidate the mechanisms of influence. Based on the above findings we argued that desorption from the sediment, sea water intrusion, old saline water diffusion, and organic matter breakdown/digeneration are the likely key processes for high I concentrations in the available groundwater data set. However, specifying one dominating source of I is not consistent with the general complexity revealed by the MVDA. It is therefore reasonable to conclude that even in one single location there may be more than one source or process contributing to the total I concentration in the groundwater. From the general Danish hydro-geological understandings this is not a surprising result, but not before elucidated via I relationships.

Based on the extensive Danish historical groundwater data base, this pilot study gave a useful first impression of the I variability throughout the country, as well as its likely complex origin(s) and highlighted the spatial complexity of I possible sources. This study demonstrated that MVDA is a useful tool for interpretation of a complex dataset in respect to I in groundwater. However, improvements on the input data would undoubtedly be most useful, e.g. preforming new sampling campaigns in local Danish areas with different geological settings; using speciation I data: I$, IO_3^-$, organic bound I, and total I, could substantially improve the possibility for gathering further knowledge on the I sources and the processes affecting the I concentrations. Last but not least, stable isotope determination ($^{127}$I, $^{129}$I) and isotope ratios is another approach which could lead to
better understanding of the nature of I pathway(s) in selected areas of interest. The present MVDA first foray serves well as a foundation upon which to design such more advanced follow-up studies.

Acknowledgements

This paper is part of the Ph.D. study of the first author; the Ph.D. project was funded by GEOCENTER Denmark. We thus gratefully acknowledge the financial support by the Geological Survey of Denmark and Greenland (GEUS) and Aarhus University.

References (see p.19 of the progress report)

Figures (not in order)

**Figure 2** Iodine in groundwater in Denmark based on data from the national database Jupiter for the period 1933-2011 (display order of the I categories (symbol levels) from top to bottom: 0-5 µg/l, 101-1220 µg/l, 6-15 µg/l, 16-50 µg/l, 51-100 µg/l)
**Figure 1** Conceptual model of the terrestrial I geochemical cycle

**Figure 5** Location of the final 46 co-joint samples of sub-set 2, used for PLS-R analysis; the samples are located within the municipalities: Jammerbugt, Sønderborg, Lejre, Gladsaxe, Frederiksberg, København, and Ishøj.
Figure 3 Number of samples analysed for I and 25 co-joint elements and parameters in Jupiter (excluded/included number of parameters in master data set, see Dataset Preparation)

Figure 4 Flow path of multivariate data analysis, performed as a 3-step approach, specifying the datasets used, their specific number of variables and objects, and the multivariate data analysis method applied.

Figure 8 Factors interpreted as the most influential for the total I variance along the first two PCs
Figure 6 Multivariate data analysis results (a-c shows PCA of r-DMS, d-f shows PLS-R of sub-set 2): (a) PC1-PC2 score plot (SP), (b) PC2-PC3 score plot (SP), and (c) Residual calibration variance plot (RCVp); (d) Loading weights plot (LWp); (e) Predicted vs. Reference plot; (f) Residual variance plot (RVp);
Figure 7 Score plot from the partial least square regression (PLS-R) model of sub-set 2 with samples grouped and notated to reflect: (a) municipality; (b) screen depth; (c) geology type; (d) redox groundwater type; (e) ion-exchange ratio (Na/Cl); and (f) Cl/Br ratios.