Cathodic Protection Prediction

Jelle van Wezel: *Cathodic Protection Prediction,* Time-series predictions with GMLVQ,

Ohana means family.
Family means nobody gets left behind, or forgotten.

— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.

1939 – 2005

Foreword here ...

ABSTRACT

Short summary of the contents in English...a great guide by Kent Beck how to write good abstracts can be found here:

https://plg.uwaterloo.ca/~migod/research/beck00PSLA.html

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— **knuth:1974** [**knuth:1974**]

ACKNOWLEDGMENTS

Put your acknowledgments here.

Many thanks to everybody who already sent me a postcard!

Regarding the typography and other help, many thanks go to Marco Kuhlmann, Philipp Lehman, Lothar Schlesier, Jim Young, Lorenzo Pantieri and Enrico Gregorio[1], Jörg Sommer, Joachim Köstler, Daniel Gottschlag, Denis Aydin, Paride Legovini, Steffen Prochnow, Nicolas Repp, Hinrich Harms, Roland Winkler, Jörg Weber, Henri Menke, Claus Lahiri, Clemens Niederberger, Stefano Bragaglia, Jörn Hees, Scott Lowe, Dave Howcroft, José M. Alcaide, and the whole LaTeX-community for support, ideas and some great software.

Regarding LyX: The LyX port was intially done by *Nicholas Mariette* in March 2009 and continued by *Ivo Pletikosić* in 2011. Thank you very much for your work and for the contributions to the original style.

1 Members of GuIT (Gruppo Italiano Utilizzatori di TeX e LaTeX)

CONTENTS

1 INTRODUCTION 1
1.1 History of natural gas in The Netherlands 1
1.2 Current situation and future 1
1.3 Involved companies and institutions 3
1.3.1 University of Groningen 3
1.3.2 Cogas and Coteq 3
1.3.3 ValueA 3
1.3.4 Cathodic protection 3
1.4 Goals 4
1.5 Working Environment 5
1.6 Deliverables 5
1.7 Cogas 6
2 BACKGROUND 7
3 PIPELINE PROTECTION 9
3.1 Coating 9
3.2 Cathodic Protection 9
4 DATASET 11
4.1 Configuration 11
4.2 Missing Data 11
4.3 Interpolation 11
5 METHODS 13
5.1 Classification 13
5.1.1 Learning Vector Quantization 13
5.1.2 Generalized Learning Vector Quantization 13
5.1.3 Localized Generalized Matrix Learning Vector Quantization 15
6 RESULTS 17
7 DISCUSSION 19

BIBLIOGRAPHY 21

LIST OF FIGURES

Figure 1.1 The area Cogas has gas pipelines in. 6
Figure 4.1 The number of measurements per measuring point against the cumulative number of measuring points having that number of measurements. 12
Figure 4.2 The average leave one out cross validation over all areas plotted against the number of N coefficients used for the polynomial interpolation. 12

LIST OF TABLES

Table 1.1 Types of damages on gas pipelines in The Netherlands 3
Table 4.1 11
Table 4.2 11

LISTINGS

ACRONYMS

1

INTRODUCTION

HISTORY OF NATURAL GAS IN THE NETHERLANDS

In 1959 on of the worlds largest natural gas deposit was found in Slochteren, The Netherlands [2]. It was discovered by the "Nederlandse Aardolie maatschapij" (NAM) which translates to: the Dutch oil Company. This company was founded by Shell and Esso in 1947 [3] in order to find natural energy resources in the Dutch soil. The discovery of the natural gas deposit was kept with a low profile at first because Dutch law at that time [1] did not give ownership of an natural occurring resource to its discoverer but to the state. After a follow up exploratory drilling near Delfzijl showed the size of the natural gas deposit The NAM filed for the drilling rights with the Dutch Government for the Groningen area.

source

how big

Then in 1962 the Dutch government passed the natural gas bill . It created a partnership between the Dutch state, Shell and Esso. The Dutch state would get a share of 50%, Shell and Esso would both get a share of 25%. The bill also founded the 'Nederlandse Gasuni' (the Dutch union for natural gas). This union would be responsible for distributing the natural gas from the Slochteren field across the Netherlands. At that time there were local gas companies producing and distributing light gas distilled from coal. These companies would now distribute the Slochteren gas with their existing distribution network rather than produce their own.

cite

source

In less than ten years most Dutch households would be connected with the gas network. The Netherlands would cook and be warmed by the Natural gas from Slochteren for foreseeable future. The natural gas from Slochteren brought economical prosperity to The Netherlands in the second part of the 20th century. However there were also downsides. So much so that an economical term is named after it: 'Dutch disease'. This term is used when a nations products become expensive due to a strong currency, which is fueled by a newly discovered natural resource. Expensive exports causes a nations production to decrease and unemployment rates to rise.

cite journaal wiki

cite

CURRENT SITUATION AND FUTURE

80 percent of the natural gas is thought to be extracted and at the current rate of extraction it is predicted that the gas will last for atleast another ten years. The gas is providing the Netherlands with warm homes and meals however the 'mijnwet' gave ownership of the natural gas to the state. This meant that the revenues went the The Netherlands as a whole and not to directly to the regionals living di-

rectly on the gas field. Due to the extraction of the gas from the lower layers of the Earth, the upper layers start to shift with earthquakes as result. Because earthquakes are not common in The Netherlands, structures are not build to withstand their impact. This causes the houses of the residents directly above and near the gas field show signs of damages and sometimes become uninhabitable.

Without any direct benefit from the gas field but with their homes collapsing the people of Groningen started to oppose the extraction of the gas form the Slochteren field. Multiple protest groups have been formed over the years and not without success. The extraction of the natural gas will be limited from.. to .. a year. There have been made promises by the Dutch government to compromise the owners of damaged homes, however this is still to be implemented.

when

number of nations

The earthquakes are not the only incentive for The Netherlands to stop extracting the gas. The Dutch government signed the Paris agreement in . The Paris agreement is a climate accord in which made promises to reduce their carbon emissions in order to slow the global rising temperature. One of the promises The Netherlands made was to reduce the carbon emissions by ... percent. The Dutch government has stated that it wants to reduce the dependency on gas and start using other forms of energy instead.

find source

A third reason to reduce the dependency on gas is when the gas eventually runs out, the Netherlands will have to rely on importing gas from foreign nations such as Russia. Which can have undesirable consequences since Russia has been known to use their gas supplies as diplomatic leverage .

For these three reasons The Netherlands is moving away from fossil fuels and transitioning to sustainable energy sources like wind and solar power. The transition to 'green' energy will take time. The energy infrastructure needs to be able to handle a bigger dependency on electricity and facilities need to be build to produce the electricity. During this time the current gas infrastructure will still be in use.

source energy trends

The gas infrastructure or transition systems in The Netherlands are hold by operator companies (TSOs). Replacing broken pipe infrastructure is a large expenditure for the TSOs, in 2014 together the TSOs spend 679 billion euro in replacing and maintaining the current transition systems. Over 84 percent of this was spent on replacing pipeline infrastructure. Table 1.1 shows what the main causes are for pipeline malfunctions. The largest contributer is the aging of the pipelines.

is how long the current gas pipe network can be sustained without replacing large parts of it. In this thesis we will try to answer this question.

As of 2014 there are 7 transmission system operators in The Netherlands. Coteq is one of them holding transmission systems in the middle of The Netherlands, in and around the city of Almelo.

30	Aging	For example oxidation
21	Excavation damage	
16	Soil movement	
12	Construction fault	
11	Unknown	
5	Point load	Tree roots
5	Other	

Table 1.1: Types of damages on gas pipelines in The Netherlands

INVOLVED COMPANIES AND INSTITUTIONS

University of Groningen

FWN, Biehl

Cogas and Coteq

In The Netherlands the energy distributors and the companies holding the infrastructure are separated by law. Here Cogas is the energy company and Coteq is the company holding the infrastructure. The Cogas group is the umbrella organization for Cogas and Coteq.

ValueA

ValueA is a company hired by Cogas to facilitate software solutions. They focus on Front-end-develpment , integration and expert support.

Cathodic protection

Steel gas-pipes are placed in the ground where they form corrosion. The amount of formed corrosion is thought to be depending on multiple environment factors. For example, the composition of the soil, the ground water level, and the depth the pipes are placed. In order to prevent the formation of corrosion on the pipes a method called cathodic protection is applied. Cathodic protection uses an electrical system where an anode (metal) is placed in the ground near the cathode (the steel pipes). When current is introduced to the system corrosion forms on the anode and not on the cathode. This method is not only applied in the gas industry but also in the shipping sector, where it is used to protect boats from corroding.

This method is applied on multiple pipe-lines by an energy company called 'Cogas' which is located in Almelo. They have collected the environment factors by hand and know how much current is needed per pipe section. Because of the corrosion buildup over time on the anodes and the pipes, at some point in time the amount of current needed exceeds a threshold. At this point it is no longer possible or economically viable to produce the amount of current needed to prevent the corrosion. When this happens more anodes need to be placed or the steel pipes have to be replaced. Cogas wants to be able to predict when this will happen based on their collected data.

The project will be done for ValueA. ValueA provides ICT solutions for utility companies and has Cogas as a client. One of the software solutions ValueA has provided to Cogas is an interactive dashboard where pipe sections with some of their properties are shown. This project will focus on predicting when a pipe section needs maintenance by replacing the pipes or placing more anodes. These predictions could be incorporated in this dashboard. It is unknown which environment factors (features) leads to the deterioration of the steel pipes. Methods like Supervised Variational Relevance Learning (SUVREL) and Global Metric Learning (GML) could give insight on which features are more important than others in this deterioration process. With the knowledge about the features predictions can be done on the dataset. For this project Learning Vector Quantization (LVQ) will be used to do the predictions. Its performance will be compared to the intuitive and widely used classifier K nearest neighbor (K-nn).

The cathodic protection predictions project is not the only project ValueA is involved with. Another classification problem involves Fraud detection in the energy grid. This problem forms the base of the graduation project of Sebastiaan van Loon. Since both the project of Sebastiaan van Loon and this project will be conducted at ValueA in the same period some collaboration will occur. This collaboration will consist of the development of an abstract module which includes an API that allows the existing software of ValueA to connect to a classifier implementing the abstract module.

This abstract module will be developed to ensure that the ValueA software can connect to a classification module independent of the methods the classification uses. This means that the development of the abstract module does not include the development of the actual classification algorithms and distance measures, which will be developed separately and are not part of the collaboration between Jelle van Wezel and Sebastiaan van Loon.

GOALS

- Research existing techniques to map LVQ and KNN on a regression problem
- Research existing methods for feature reduction/relevance
- Develop a classification module for ValueA which can predict the lifespan of steel gas pipes with LVQ, KNN and a or multiple preprocessing methods like GML and SUVREL.

- Offer a comparison of the results between the used methods
- Describe the research, used methods, developed software, and the results.

WORKING ENVIRONMENT

The student will be partly embedded both in the Intelligent Systems research group within the JBI, and with ValueA at the location of Cogas, a customer of ValueA located in Almelo. It is envisioned that the student will work, on average, two days a week at Cogas and the remaining days at the university and at home. The student will have a meeting at least every fortnight with the first supervisor and will have at least one stand up a week with the third supervisor and/or other employees of ValueA and will attend the monthly meetings of the Intelligent Systems research group.

The first and second supervisor will provide necessary expertise regarding the classification algorithms and dissimilarity measures used in the project. The third supervisor will provide necessary expertise about both the ValueA software solution and the energy utility market, provide data sets for the training and testing of the classifiers, and evaluate and offer feedback on the software developed.

The student will work on his own laptop and PC, and in case of failure work on a university managed system. Deployment of software will be done on a cloud environment provided by ValueA. Code revisions will be stored in a GIT repository managed by ValueA. Data which ownership belongs to ValueA will remain in the ValueA cloud environment and will not be copied to the student's laptops unless explicit permission is granted from ValueA.

DELIVERABLES

As requested by ValueA the student will develop the software in Python.

- End of December. Classification module description. A document describing the classification module, its main features and its integration with the existing ValueA software.
- Start of January. Literature survey. A document summarizing the main findings of a literature review and the relevant references regarding Cathodic protection predictions, LVQ, KNN, Metric learning.
- End of January. Abstract classification module. An abstract module facilitating the connection of a classifier with the existing software and data of ValueA. (In collaboration with Sebastiaan van Loon.)
- Mid of February. KNN classification module. A KNN classification module for predictions of gas pipe lifespan.

- End of March. LVQ classification module. A LVQ module for predictions of gas pipe lifespan with relevance learning as initialization.
- Mid of April. Test results.
- Start of May. Thesis draft.
- Mid of May. Presentation draft.
- End of May. Presentation to the board of Cogas.
- End of May. Final presentation, code, and thesis.

COGAS

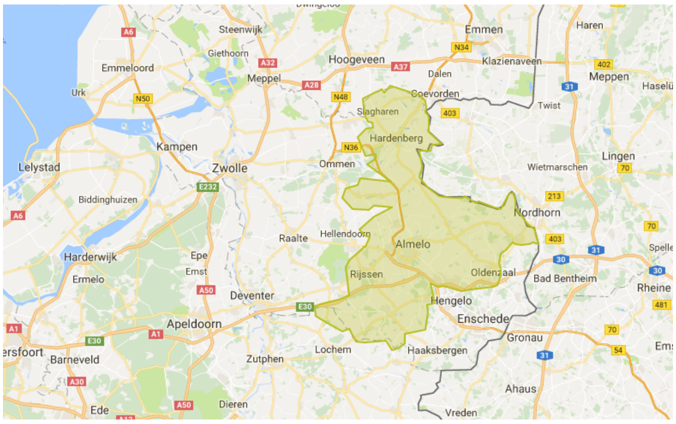

Figure 1.1: The area Cogas has gas pipelines in.

2 BACKGROUND

3 PIPELINE PROTECTION

COATING

CATHODIC PROTECTION

4 DATASET

CONFIGURATION

The data consisted of 24 CP areas. Each of these areas have one anode point where the current is pushed into the ground. The current is then propagated through the ground and led back to the anode through the steel gas pipes completing the electric system.

The pipes spread from the anode and branch off from each-other, creating a tree like structure. The pipes are divided into segments, Usually where a pipe branches into two pipelines, two new segments form. A pipe can only be part of one segment but a segment can consist of multiple pipes. In between the segments measuring points are placed to measure the current and difference in voltages in the segment. The number of segments an measuring points differ per area and is shown in table

Table 4.1

right	center	left
right	center	left

Table 4.2

right	center	left
right	center	left

MISSING DATA

Usually the measurements are conducted once per year at every measuring point. The measurements started as early as 1987 up-until 2016. This means their should be 39 measurements per measuring point. However this is not the case as is shown in figure

Figure [REF] shows a fast drop in measurement points after the 20 measurements mark. At 26 measurements the drop stagnates and the maximum number of measurements is reached at 37 measurements. This means that none of the measuring points have the maximum number of 39 measurements.

INTERPOLATION

The method we proposed to use to predict the voltages of the pipelines with LVQ [REF] depends on the data being consistently spaced in time.

Before the interpolation method was applied the data was checked for outliers by taking the mean and two times the standart deveation. Then removing the data points that did not fall in the range of $x > 2 \times \sigma - \mu$ and $x < 2 \times \sigma + \mu$.

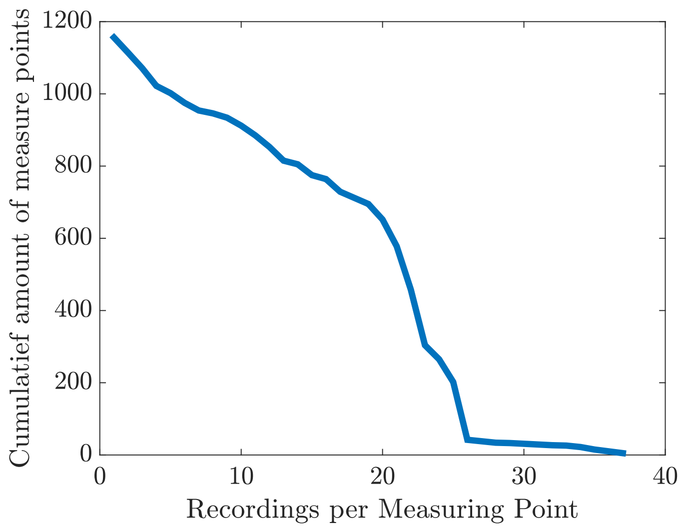

Figure 4.1: The number of measurements per measuring point against the cumulative number of measuring points having that number of measurements.

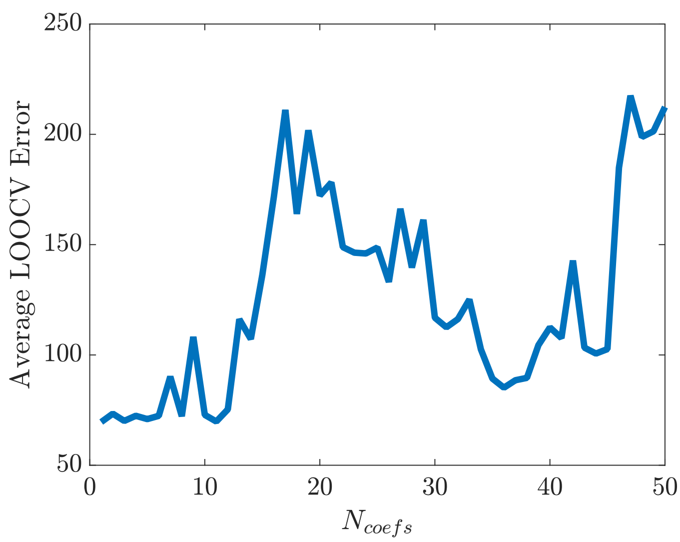

Figure 4.2: The average leave one out cross validation over all areas plotted against the number of N coefficients used for the polynomial interpolation.

5 METHODS

CLASSIFICATION

Learning Vector Quantization

Learning vector quantization (LVQ) was introduced in the '90s by Kohonen [**kohonen**] as a precursor to Self Organizing Maps (SOM). It is a prototype-based supervised classification algorithm. It uses a winner takes it all supervised Hebbian learning based approach. LVQ is defined by its prototypes often denoted as $W = (w_i, ...w_n)$. These prototypes are defined in the feature space of the dataset. Each epoch the algorithm iterates over the dataset and for each data-point it determines the closest prototype. The prototypes have labels from the same classes as the data-points. If the label of the prototype is the same as the label of the data-point, the prototype is updated by moving it in the direction of the data-point. If the labels differ, the prototype is moved away from the data-point. The distance a prototype travels is influenced by its location and a predetermined learning rate. The original heuristic has seen many variations since it was invented. The following sections go through the variations we used for classification on the datasets in this paper.

Generalized Learning Vector Quantization

The LVQ heuristic by Kohonen [**LVQ**] needed proof of convergence, this was done by Sato and Yamada with Generalized Learning Vector Quantization (GLVQ) [**GLVQ**]. They introduced an energy function S based on the LVQ 2.1 scheme. Here there are two winning prototypes, one closest to the data-point with the same label as the data-point (denoted as $\vec{w}_J$) and the closest data-point with a different label, denoted by $\vec{w}_K$. Here $\vec{w}_J$ is moved towards the data-point and $\vec{w}_K$ is moved away from the data-point. They show that by optimizing the energy function through the prototypes this generalized version of LVQ always converges. The energy function is defined as

$$E = \sum_{i=1}^{P} \Phi(\mu(\vec{x}_i)) \text{ , with } \mu(\vec{x}) = \frac{d_J - d_K}{d_J + d_K}. \tag{5.1}$$

The function, $\Phi(\mu)$, is monotonically increasing and d_J and d_K the squared distances from a data-point to the winning prototypes, defined by

$$d_i = (\vec{x} - \vec{w}_i)^\top (\vec{x} - \vec{w}_i) \text{ , with } i = J, K, \tag{5.2}$$

with $\vec{x}$ the data-point and $\vec{w}_i$ one of the winning prototypes. The general update rule for the prototypes are defined by

$$\Delta \vec{w}_i = -\alpha \frac{\partial E}{\partial \vec{w}_i} \text{ , with } i = J, K, \tag{5.3}$$

with α as the learning rate. Then based on equation 5.3 the update rules for the prototypes are derived to be

$$\Delta \vec{w}_J = \alpha \Phi'(\mu(\vec{x})) \frac{d_K}{(d_J + d_K)^2} (\vec{x} - \vec{w}_J), \tag{5.4}$$

$$\Delta \vec{w}_K = -\alpha \Phi'(\mu(\vec{x})) \frac{d_J}{(d_J + d_K)^2} (\vec{x} - \vec{w}_K). \tag{5.5}$$

In equations 5.4 and 5.5 we can see prototype $\vec{w}_J$ is updated towards the data-point and prototype $\vec{w}_K$ is updated away from the data-point minimizing the energy function. The following LVQ methods (RGLVQ, GMLVQ, and LGMLVQ) are all based on this generalized variant of the LVQ 2.1 algorithm. In the experiments we only used these generalized versions of the LVQ algorithm.

Because GLVQ uses a gradient descent (a stochastic gradient decent was used) to find an optimal location for its prototypes we used a fixed maximum iterations set to 2500 for all variants of the algorithm. Furthermore the learning rate used was quadratically declining spread out over the number of iterations.

Generalized Matrix Learning Vector Quantization

Generalized Matrix Learning Vector Quantization (GMLVQ) as proposed by Schneider et al. [**GMLVQ**] uses a matrix Λ instead of a vector (as in RGLVQ) to map the relevances and the combinations of the relevances between features. In order to gain a meaningful distance measure, Λ has to be semi positive definite. This can be obtained by searching for the matrix Ω such that $\Lambda = \Omega^\top \Omega$. The method described here is based on the work by Schneider in her Ph.D. thesis [**Schneider**]. The energy function for GMLVQ is written as

$$E = \sum_{i=1}^{P} \Phi(\mu^\Lambda(\vec{x}_i)) \text{ , with } \mu^\Lambda(\vec{x}) = \frac{d_J^\Lambda - d_K^\Lambda}{d_J^\Lambda + d_K^\Lambda}. \tag{5.6}$$

Here the distance is influenced by the relevance matrix Λ and defined as

$$d_i^\Lambda = (\vec{x} - \vec{w}_i)^\top \Lambda (\vec{x} - \vec{w}_i), \text{ with } i = J, K. \tag{5.7}$$

Based on the energy function and this distance measure the new update rules are derived to

$$\Delta \vec{w}_J = \alpha_1\, \Phi'(\mu^\Lambda(\vec{x}))\, \mu_J^\Lambda(\vec{x})\, \Lambda\, (\vec{x} - \vec{w}_J), \tag{5.8}$$

$$\Delta \vec{w}_K = -\alpha_1\, \Phi'(\mu^\Lambda(\vec{x}))\, \mu_K^\Lambda(\vec{x})\, \Lambda\, (\vec{x} - \vec{w}_K). \tag{5.9}$$

In these update rules α_1 is the learning rate for the prototypes, $\mu_J^\Lambda(\vec{x})$ and $\mu_K^\Lambda(\vec{x})$ are defined as

$$\mu_J^\Lambda(\vec{x}) = \frac{4 d_K^\Lambda}{(d_J^\Lambda + d_K^\Lambda)^2}, \tag{5.10}$$

$$\mu_K^\Lambda(\vec{x}) = \frac{4 d_J^\Lambda}{(d_J^\Lambda + d_K^\Lambda)^2}. \tag{5.11}$$

From the update rules in equations 5.8 and 5.9, the update rule for the matrix Ω is derived to

$$\begin{aligned} \Delta \Omega_{lm} = &- \alpha_2 \Phi'(\mu^\Lambda(\vec{x})) \\ &\Big(\mu_J^\Lambda(\vec{x}) \Big((x_m - w_{J,m}) [\Omega(\vec{x} - \vec{w}_J)]_l \Big) \\ &- \mu_K^\Lambda(\vec{x}) \Big((x_m - w_{K,m}) [\Omega(\vec{x} - \vec{w}_K)]_l \Big) \Big), \end{aligned} \tag{5.12}$$

where α_2 is an independent learning rate from the learning rate α_1 used in the update rules for the prototypes. After each update Λ needs to be normalized (as with λ in RGLVQ) to prevent GMLVQ from degeneration and this can be done by dividing all elements of Λ by $\sum_m \Lambda_{mm}$ and thus enforcing $\sum_m \Lambda_{mm} = 1$. This is a generalization for a simple diagonal metric of the normalization in GRLVQ where $\sum_m \lambda_m = 1$ was enforced.

Localized Generalized Matrix Learning Vector Quantization

The Localized GMLVQ (LGMLVQ) works with a more complex model using local matrices either attached to each prototype or in a class-wise manner. The update rules for the closest prototypes with the same and different labels are shown in equations 5.13 and 5.14.

$$\begin{aligned} \Delta \Omega_{J,lm} = &- \alpha_2 \Phi'(\mu^\Lambda(\vec{x})) \\ &\mu_J^\Lambda(\vec{x}) \Big((x_m - w_{J,m}) [\Omega_J(\vec{x} - \vec{w}_J)]_l \Big), \end{aligned} \tag{5.13}$$

$$\begin{aligned} \Delta \Omega_{K,lm} = &+ \alpha_2 \Phi'(\mu^\Lambda(\vec{x})) \\ &\mu_K^\Lambda(\vec{x}) \Big((x_m - w_{K,m}) [\Omega_K(\vec{x} - \vec{w}_K)]_l \Big). \end{aligned} \tag{5.14}$$

6 RESULTS

7 DISCUSSION

BIBLIOGRAPHY

[1] Dutch Government. *Mijnwet 1903*. Wet. 1903. URL: http://wetten.overheid.nl/BWBR0001870/1999-02-17.

[2] NAM. *Gasveld Groningen*. 2014. URL: https://www.nam.nl/algemeen/mediatheek-en-downloads/brochers/_jcr_content/par/textimage.stream/1474410748285/70797b9b4c4fe21a9e1e5b22d329e7ca170dd1a2c4e7abe8abd1b696a593b779/flyer-namg50ned.pdf.

[3] *Our History*. https://www.nam.nl/over-ons/onze-historie.html. Accessed: 25-04-2018.

DECLARATION

Put your declaration here.

Groningen, May 2018

Jelle van Wezel

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Thank you very much for your feedback and contribution.