UPTEC X 10 025
Examensarbete 30 hp
Oktober 2010

Prediction models for the emergence of mycotoxins in grain

Anders Sundström

UPPSALA UNIVERSITET

Bioinformatics Engineering Program

Uppsala University School of Engineering

UPTEC X 10 025	**Date of issue 2010-10**

Author

Anders Sundström

Title (English)

Prediction models for the emergence of mycotoxins in grain

Title (Swedish)

Abstract

The Swedish Meteorological and Hydrological Institute calculates an increase in temperature in Sweden by 4-6°C before the year 2100, with the possibility of increased incidence of mycological and mycotoxical material in grain intended for feed and food production. Four models for predicting emergence of mycotoxins in grain were developed using a regression analysis function in the statistical software Weka 3.6 along with data from SVA, SMHI, Lantmännen and the Swedish Museum of Natural History.

Keywords

Prediction models, *Alternaria*, mycotoxin, Tenuazonic acid

Supervisors

Gunnar Andersson
National Veterinary Institute

Scientific reviewer

Mats Gustafsson
Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **30**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 471 4687

Prediction models for the emergence of mycotoxins in grain

Anders Sundström

Populärvetenskaplig sammanfattning

Många tecken tyder på att Sveriges klimat kommer att bli både varmare och fuktigare i framtiden. SMHI förutspår att medeltemperaturen i Sverige kommer att öka med 4-6°C innan år 2100, och därmed närma sig dagens Mellaneuropeiska klimat. Det finns farhågor att denna varmare miljö kommer att leda till ökade problem med mögelgifter, så kallade mykotoxiner, i grödor som skall användas till livsmedel eller djurfoder. Vissa av dessa mykotoxiner har akuttoxiska effekter för människor och djur och därför blir förekomsten av dessa gifter viktig att förutsäga och övervaka. Bland mykotoxinproducerande svampar återfinns släktet *Alternaria* som är en frekvent mögelsvamp i spannmål och ett vanligt allergen bland människor. Det mest potenta Alternariatoxinet, Tenuazonic acid, påvisades i anmärkningsvärt höga koncentrationer vid en undersökning på spannmål i södra och mellersta Sverige 2006 och 2007.

Ett verktyg för att på ett tidigt stadium förutsäga risker för mykotoxinbildning i grödor är användandet av prognosmodeller. En prognosmodell i detta sammanhang beräknar förekomst av mykotoxiner som funktion av väderdata. Tanken är att man då skall kunna förutsäga bildandet av toxiner i god tid så att preventiva åtgärder skall kunna sättas in.

Det här examensarbetet är en del av ett MSB-finansierat samarbetsprojekt mellan Svenska Livsmedelsverket, Svenska Jordbruksverket, Statens Veterinärmedicinska anstalt, m.fl. med avsikten att i en pilotstudie beskriva förutsättningarna för att utveckla prognosmodeller för uppkomst av mykotoxiner i svensk spannmål. Syftet med arbetet var att redogöra för definitioner, variabler, kausala samband, etc. rörande framställning av en större prognosmodell på svenska klimatförhållanden, både samtida och framtida, samt att utveckla en enklare prognosmodell att utnyttja som beslutstöd vid kommande utredningar. Resultatet av den här undersökningen visade att det finns en koncensus i forskningen vad gäller indikatorer för mykotoxinproduktion i spannmål, och att dessa indikatorer bör utgöras av någon form av nederbörd, temperatur och fukt. Det visade sig också att tidsintervallet man applicerar dessa indikatorer på är av största betydelse och oftast utgår från spannmålets axgång.

Examensarbete 30hp

Civilingenjörsprogrammet Bioinformatik

Uppsala Universitet, Oktober 2010

Table of Contents

Abbreviations

CPL	Critical period length
CV	Cross validation
DON	Deoxynivalenol
LD_{50}	Dose required to kill half a test population after a given test duration
spp.	species, plural form
SVA	National Veterinary Institute
TeA	Tenuazonic acid

1. Introduction

Climate change is a hot topic in today's society. Swedish Meteorological and Hydrological Institute predict an increase in Sweden's mean temperature by 4-6°C until the year 2100 (SMHI). A consequence of the warmer climate is a possible increase of fungi and mycotoxin incidence in feed and food, with higher risk of intoxication as a result.

One way to handle this is by the utilization of prediction models. A prediction model can here be seen as a tool that forecasts emerging mycological risks based on meteorological factors. Appliance of such an instrument would give the farmer a head start, so preventive measures can be taken in due time. Also, a predictive model forecasting not only seasonal ups and downs but long term changes in incidence due to climate changes is interesting from a national point of view, calculating the upcoming agricultural conditions.

The fungi of interest is species of the genus *Alternaria*, which is an important mycotoxin producer and a decomposer of organic material. As such, it can cause both pre- and post-harvest decay as well as damage crops during growth (Andersen, 2001). It is estimated that *Alternaria* infection contributes in 20% -40% of the spoilage of agricultural output (Battilani, 2009).

There are three factors that are of interest to predict; occurrence of the fungi responsible for mycotoxin-production, the concentrations of toxin in the crop and the occurrence of a crop decease

caused by *Alternaria* called Black point. Black point is interesting from an economical point of view because infected grain originally intended for food production are downgraded and can only sell as animal feed. The idea behind the models is that the prediction should be based on meteorological and agricultural variables at some specific period pre- and/or post-harvest.

Previous research in this area can be divided into two categories; fungi and mycotoxin prediction models regarding non-*Alternaria* related fungi (Hooker, 2002), (Moschini, 1996), (Prandini, 2009), (Tarekegn, 2006) and *Alternaria* related prediction models based on climate that differ largely from Swedish climate (Iglesias, 2007), (Katial, 1997), (Languasco, 1994), (Moschini, 2006). These models are based on statistical analysis of relations between historical mycological data and meteorological and/or agricultural factors.

The non-*Alternaria* prediction models are more developed but they are not applicable to the *Alternaria* species, and the existing *Alternaria* models are not applicable on the Swedish climate. A model that is produced explicitly for *Alternaria* and Swedish climate would be of good use for containing the spread of mycotoxins in Sweden.

This thesis work is part of the Swedish Civil Contingency Agency's project *National collaboration on climate-related spread of molds and mycotoxins*, and the aim of the thesis is to start to describe the conditions for the development of *Alternaria* and *Alternaria* mycotoxin prediction models from metrological data in Sweden, and to develop a simple prediction model to use as decision support in future sampling.

2. Background

This section explains some biological aspects of the research area. The relevant mycological fundamentals are described along with the dominant cereal development scale used in agricultural science.

2.1. Alternaria

Alternaria is a diverse and omnipresent genus of the fungi. It is a common field fungus and therefore a frequent contaminator of grain. Its species includes both saprophytes, which means organisms that feed on decaying material, and plant pathogens, and therefore *Alternaria* plays a major role in crop damage and decay both before and after harvest (Logrieco, 1990).

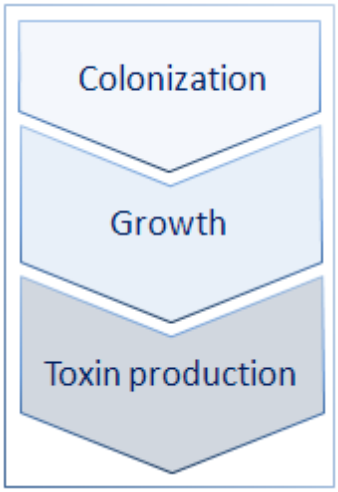

Figure 1: The infection procedure of *Alternaria*.

Alternaria colonies are white or gray at first, but because of their melanin production they gradually change color towards brown as times goes by. *Alternaria* survives the winter on infected debris and produces wind-carried spores that throughout the spring and summer will infect plants and seeds during moist and warm conditions (Battilani, 2009). Its peak spore count occurs in the late summer months (Katial, 1997).

Because of its saprophytic character, *Alternaria* incidence can

sometimes be seen as a symptom for an already ongoing damage that is caused by some other pathogen. This is the case with what is commonly called sooty molds (Saskatchewan, 2009). Sooty molds acts like a brown dust that is covering the head of a wheat straw, and is actually composed of *Alternaria* molds that is consuming decaying crops.

A lot of interest and research has been put into a crop decease called Black point, mainly from an economical point of view. The primary agent for causing Black point is *Alternaria alternata* and the decease turns out as a black to brown discoloration of the kernels of wheat and barley (Özer, 2005). Black point itself does not damage the nutritional quality of the crop but the discoloration it is responsible for makes the grain unattractive and difficult to market, which leads to a downgrading of the crop and economical loss for the farmer (Rees, 1984). This means that grain that wore intended to be sold as food is downgraded and can only sell as animal feed.

Studies have shown that toxins produced by *Alternaria* can disturb the development of the fetus of hamsters and mice, and have cytotoxic properties on bacterial and mammal cells (Battilani, 2009). *Alternaria* also has a direct affect on humans in form of allergy and airway diseases. Its spores are the main agent of indoor and outdoor fungal allergens and it is determined that *Alternaria* sensitization is the number one source of childhood asthma (Battilani, 2009).

2.2. Tenuazonic acid

Some fungi can produce organic compounds that are of toxic nature, mycotoxins, and we define these compounds as secondary metabolites since they are not involved in the growth or the reproductive system of the organism. The genus of *Alternaria* can produce 71 known toxic secondary metabolites but only a few are of toxicological significance (Battilani, 2009).

Tenuazonic acid, or TeA, is a colorless viscous oil produced by the species *A. longipes*, *A. tenuissima* and *A. alternata* and is probably the most toxic of all of the *Alternaria* mycotoxins with a Lethal-Dose$_{50}$–value of 81/168 mg/Kg female/male mice (Battilani, 2009). As an interesting side note and point of reference, caffeine is just slightly less toxic with a LD$_{50}$–value of 192 mg/Kg (University). As another point of reference TeA was given in the diet to chicken at concentration of 10mg/kg of feed, with decreased weight gain and lowered feed efficiency as a result (Giambrone, 1978).

TeA's toxicity originates from its inhibitory effect on the protein synthesis, or more precise, TeA suppresses the release of newly formed proteins from the ribosome (Battilani, 2009). The cytotoxicity has been established in test on mouse, hamster and human cells (Zhou, 2008). TeA also has antitumor, antibiotic and antiviral properties (Siegel, 2009) (Shephard, 1991).

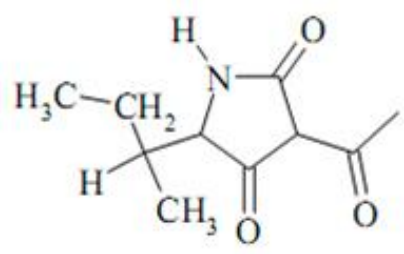

Figure 2: Molecular shape of Tenuazonic acid.

TeA has been detected in many human food sources such as olives, peppers, tomatoes, tangerines, melons, apples, sorghum, rice, wheat, sunflower seeds and in tobacco. In previous studies, the levels of TeA have been relatively modest, with few exceptions like India in 1978 (6mg/kg in sorghum) (Webley, 1998), (Battilani, 2009), but in 2006, a Swedish study found high levels, 4mg/kg, of TeA in whole grain and straw samples from Sweden (Häggblom, 2007).

2.3. Zadoks scale

In order to precisely measure a crop's development, independently of where it grows, the development process has been normalized with respect to the crop's visual stages of growth. This was done by the Dutch phytopathologist Jan C. Zadoks who proposed that the cereal growth should be divided into ten primary stages (Zadoks, 1974). Each primary stage is in turn divided into ten secondary stages, so the complete scale ranges from 00 to 99. So for example, Z39 means that development stage is in the stem elongation phase and more precisely that the flag leaf of the cereal just became visible. The scale thoroughly used in agricultural research and practice and it is commonly referred to as the Zadoks scale.

Table 1: ***Zadoks development stages***

Stage	Crop development
0	Germination
1	Seeding growth
2	Tillering
3	Stem elongation
4	Booting
5	Ear emergence
6	Flowering
7	Milk development
8	Dough development
9	Ripening

Each stage incorporates a more detailed sub-stage.

2.4. Aim

The aim of this thesis is to develop a rough prediction model for mycotoxins in grain, and to contribute a compilation of indicator selection research that could be of use in further development of *Alternaria* prediction models.

3. Materials and methods

This section starts with a presentation of the datasets used in this project and proceeds with a description of the pre-processing procedure of the aforementioned datasets. Then the statistical concepts involved in training and evaluation of the prediction models are explained. Further on, previous research in the field of prediction modeling are pointed out, introducing two major works in mycotoxin prediction modeling. And finally, the statistical modeling practice in this project is accounted for in the end of this section.

3.1. Datasets

Suitable data on this subject is scarce. The optimal dataset would consist of sampling from many locations during several seasons, but since Swedish *Alternaria* research is just starting to develop, unfortunately this optimal dataset does not yet exist. However, investigations lead to three datasets being located for the biological factor, one for each concept sought out to predict (incidence of fungi, toxin and plant disease). These dataset were also the only datasets to be found regarding *Alternaria* related data for Sweden. Data for the meteorological variables was collected using SMHI's databases. The basic structure of all datasets is a biological observation in conjunction with whether data and

since these datasets are the only ones regarding this research area, the purpose of all them is to be used for training and evaluation in the modeling process.

Weather data for a duration of approximately four months pre-harvest were obtained for all the datasets from SMHI's or LantMet's weather stations. The data consisted of precipitation, temperature, relative humidity and atmospheric pressure and was measured every three hours at SMHIs stations. The quality of the data is based on how closely the weather stations are situated in the area of study, which varies from a few kilometers up to 50km.

The first dataset corresponds to daily *Alternaria* spore counts per cubic meter air for a time period of 30 years. The sampling was made at Frescati in Djurgården in the north eastern part of central Stockholm.

The second dataset is based on incidence of black point infected kernels in oat. Incidence was measured as percent of infected kernels. The sampling was done by Lantmännen on 17 farms in the central and southern parts of Sweden during 3 consecutive seasons.

The third dataset comes from a study lead by professor Per Hägglund at SVA. The project set out to map the mycotoxical risks involved in the harvest of grain from rain damaged areas in Sweden in 2006, and also, to find out the toxical substances involved. The actual numbers here are concentrations of Tenuazonic acid. The dataset also includes crop species and all in all, 33 fields were sampled.

3.2. Pre-processing of data

Matching the biological data with the correct, in terms of date, meteorological data was an elaborate process since SMHI measure meteorological variables every three hours. When the time period gets sufficiently long, for example 30 years with the *Alternaria* spore count data, the result is a very large dataset. The size of the dataset would have made the concatenation of biological data and meteorological data a very time consuming task, and the process was therefore automated with a perl-script. The code can be viewed in the appendix.

The *Alternaria* dataset were reduced from 30 years to 18 years due to limitations in the weather data. More explicitly, the weather data had for some reason only precipitation measurements up until 1997.

As earlier mentioned, the Black point dataset were constituted of 17 sampled farms. However, nine of the farms were situated to close to be separated by the meteorological data, and thereby were given the same meteorological variables. Since these nine farms had varying Black point incidence, there exist a biological variation that cannot be explained by the available weather data, and the dataset was reduced to eight unique locations.

As with the black point dataset, some of the 33 fields in the toxin dataset were geographically to close to each other, and the resolution of the meteorological data reduced them to 12 unique locations.

Also, it is common practice in mycotoxin prediction modeling that the variables are attributed to the model during a certain period of time in the growth process when their predictive value are the

strongest (Hooker, 2002) (Moschini, 2006) (Franz, 2009). This window is called the *critical period length*, or CPL in short. Most of the variables consist of a sum of the number of days within this critical window that certain conditions is fulfilled. This kind of variables also reduces the size of the dataset so that the measure points of the final dataset used to train and validate the model consists of one sample of measured biological data (*Alternaria* spore counts, toxin concentration et.c.) and one value for each calculated dependent variable (number of days with rainfall et.c). So the size of dataset is the number of fields (and/or years) sampled times the number of weather variables. This means that the size, and in some sense quality, of the dataset largely depends on the number of sampled fields, so reducing the number of fields due to poor weather data resolution is a serious problem, both regarding to the resulting lack of training and validation data and the increased risk of overfitting the mode simply because of the tiny dataset.

Table 2: ***Dataset dimensionality***

Dimensionality	Alternaria	Black point	Toxin
Raw	27321 x 6	19760 x 6	39424 x 6
Pre-processed	3415 x 6	1377 x 6	1848 x 6
CPL-adjusted	306 x 6	153 x 6	204 x 6
With applied variables	18 x 6	9 x 6	12 x 6

The dimensionality of the datasets were severely reduced by pre-processing along with the limited geographical resolution of the weather data. Since the variables involved in the model are only applied to the data for a short period of time, the resulting dataset becomes even smaller. The nature of the variables themselves (summations over number of days when weather aspects fulfills a condition) reduces the dataset to their final size of *sampled fields x variables*.

3.3. Statistical concepts

3.3.1. Regression analysis

In statistical terminology, regression analysis is a collective term for the examination of the relationship between variables. The purpose is to determine the relationship of one or more variables, called independent variables, upon another variable called the dependent variable (Blom, 2005). For example, examining the effect of temperature and precipitation on *Alternaria* incidence in Swedish wheat fields.

The goal of regression analysis is to determine an equation that models the dependent variable as a function of the independent variables with the highest possible agreement between predictions and true (measured) values. This is accomplished by fitting a polynomial of some degree to the data, using for example a numerical method like Least squares.

Least squares is a method which seeks to minimize the difference between data and a polynomial by minimizing the function

Least squares is a method used to determine the model parameters which minimizes the mean sum of squared prediction errors. Assuming that the prediction model is a polynomial denoted f(x,μ), the objective function Q minimized is:

$$Q = \sum_{i=1}^{n} \left(y_i - f(x_i, \mu)\right)^2$$

where y_i is the data value of the point x_i and $f(x_i, \mu)$ is a function representing the polynomial. μ is a parameter vector. In linear regression for example, $f(x_i, \mu)$ represents a polynomial of the first degree where μ_0 is the intersect and μ_1 is the slope, resulting in $f(x_i, \mu) = \mu_0 + \mu_1 x$ (Blom, 2005).

The Least squares method has a famous disadvantage. It is sensitive to data values that strongly deviates from the rest of the data, commonly called outliers. There exist several ways to deal with this disadvantage, making the method more robust, for example switching the square to an absolute value. Another approach, which is used further on in this project, is to instead of minimizing a sum of squares, minimize the median square (Rousseeuw, 1984). The method is called Least median squares (LMS) and seeks the solution to the function:

$$Q = \min \underset{i}{\text{median}}(y_i - f(x_i, \mu))^2$$

The LMS method is much more rigid towards deviations in the data because of the nature of the median function, able to disregard any eventual outliers in a way that a sum cannot.

3.3.2. Correlation coefficient

Correlation in this context is a measure of agreement between the model predictions and the measured (true) values. The coefficient used here for quantification of this agreement is the classical (Pearson) correlation coefficient R defined as in the formula below:

$$R = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

This means that the correlation coefficient is the covariance of two variables divided by the product of their standard deviation. The correlation coefficient attains a value between -1 and 1, where the value 1 indicates complete positive correlation, 0 indicates no correlation and -1 complete negative correlation (Blom, 2005).

Sometimes the squared correlation coefficient, R^2, is used as a general measure of how well a model accounts for the variation in the data. This is commonly seen in prediction model articles as a performance estimator of the model (Hooker, 2002) (Franz, 2009) (Tarekegn, 2006) (Moschini, 2006), and is also used further on in this project when referring to previous work in the field of prediction modeling.

3.3.3. K-fold cross validation

Cross validation is a resampling method used to evaluate how well a model will generalize to an independent dataset. The dataset is divided into K subsets and one subset of the data is out held for validation purposes, while the rest of the data is used for training a model to evaluate on the validation set. This procedure is then repeated with a new validation subset and consequently, a new training set, until all samples have been used in both a validation set and a training set (Molinaro, 2005). A common usage is a K set to 10 (McLachlan, 2004). Cross validation is a good alternative when data is scarce and preferably used for modeling purposes rather than excluding some of the data for a validation dataset.

In this project, a special case of K-fold cross validation is sometimes used called Leave-one-out cross validation, which means that K is set to the number of instances in the dataset. Or in other words, every measure point is used both for training and for validation. This method is more time

consuming and computational heavy than other versions of K-fold cross validation, but this is not of concern in a situation where a small dataset is used, such as in this project. Leave-one-out cross validation was used when the number of samples was lower or equal than 10.

3.4. Prediction models

A prediction model in this context is a simplified mathematical model of a system of interest that maps system inputs into predicted system outputs. The model tries to summarize the system into elements that can be used to deduct relations between the elements or as a platform to propose hypotheses. In this case, a model should suggest incidence of spore concentration, black point prevalence or mycotoxin concentration as a function of whether variables. An important concept is the time factor involved.

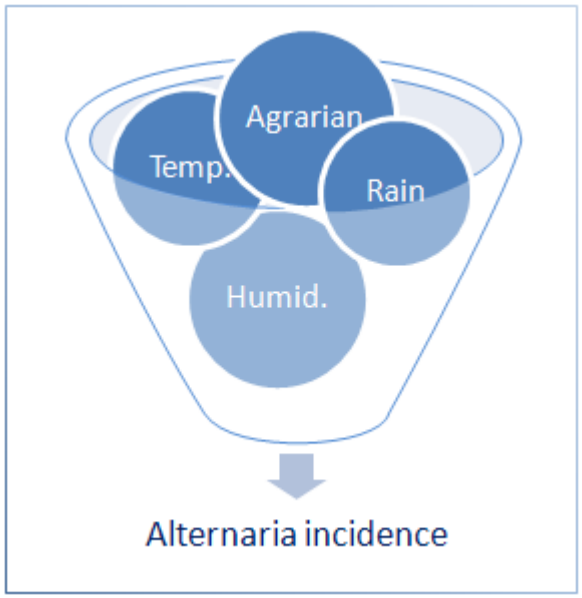

Figure 3: A schematic picture describing a model of *Alternaria* incidence as a function of the pre-harvest weather indicators temperature, precipitation and humidity as well as an agrarian indicator. This schematics is also applicable as generalizations for the other models in this project.

3.4.1. Previous research

Previously, there have been a number of attempts at modeling the presence of fungi or incidence of toxin in grain by identification of the most important weather variables (Tarekegn, 2006), (Hooker, 2002), (Moschini, 2006), (Prandini, 2009), but most of the efforts have been done in regards to the fungi genus *Fusarium* and related mycotoxins. However, there exist rough prediction models for *Alternaria* as well (Iglesias, 2007), (Katial, 1997), (Languasco, 1994), (Moschini, 2006). The two most thorough models are elaborated on in more detail in the following paragraphs. These models have to some extent been used as guidelines in this project, and their method and practice would have been followed even more closely if sufficient Swedish data would have been available. Also, calculating the critical period length in analogy with either model would have been interesting, but the adequate weather data were acquired at a late stage in the projects development process and was not able to be fitted in due to time constraint.

DONcast

DONcast is a commercialized prediction model for the mycotoxin Deoxynivalenol, or DON in short, in wheat developed by Hooker et al. (Hooker, 2002). The model consists of three equations, using temperature and rainfall from a period of seven days before heading to 10 days after heading (Zadoks 59), calculating the predictions.

During the model design phase, DON concentrations in grain samples from 399 wheat fields was measured during five years. The weather data were collected for the period of 48 days around wheat heading. To locate the CPL, the weather variables were given daily binary values, i.e. one or zero, depending on if they match a predetermined weather criterion (for example mean daily temperature exceeds 32°C). The criterions used were selected since expert opinion deemed them relevant regarding to either fungi development or strengthening fungi inoculation potential. Every day's binary values during the 48 day stint were then summed in four day windows according to the following procedure:
For each i between i = 1 to i = 45, calculate $WV_i = \sum_{d=1}^{4} B_{(d+i-1)}$. B in this procedure is the binary value for the weather variables applied to day d in the four day interval i, and the outcome of the whole procedure is a list of scored intervals to be used in CPL calculations. This means The procedure resulted in 45 proposals that in part can build up the model's CPL, and the best combination of intervals were then calculated with regression analysis, and defined as the CPL. They found the CPL to be the interval from 7 days before heading to 10 days after heading.

The actual modeling process is not described in satisfactory detail in the paper and Hooker et all suffice by merely saying that the concentration of DON were transformed by the natural logarithm (hence the exponential regression equations) and that the final models were derived from regression procedures.

Table 3: ***DONcast variables***

Variable	Explanation
RAINa	Number of days with rain > 5mm/day 4-7 days before heading.
RAINb	Number of days with rain > 3mm/day 3-6 days after heading.
RAINc	Number of days with rain > 3mm/day 7-10 days after heading.
Tmin	Number of days with temperature < 10°C 4-7 days before heading.
Tmax	Number of days with temperature >32°C 4-7 days before heading.

Hooker et all's results are the following three exponential equations where equation 1 utilizes weather data before wheat heading and equation 2 and 3 utilizes weather data from before and after heading. Equation 1 is used when RAINb > 0, and equation 2 when RAINb = 0.

(1) $DON = \exp[-0{,}30 + 1{,}84RAINa - 0{,}43(RAINa)^2 - 0{,}56Tmin] - 0{,}1$

(2) $DON = \exp[-2{,}15 + 2.21RAINa - 0{,}61(RAINa)^2 + 0{,}85RAINb + 0{,}52RAINc - 0{,}30Tmin - 1{,}10Tmax] - 0{,}1$

(3) $DON = \exp(-0{,}84 + 0{,}78RAINa + 0{,}40RAINc - 0{,}42Tmin) - 0{,}1$

When evaluated on 17 farm fields in 2000, equation 1 had a value of R^2=0,55, equation 2 had R^2=0,71 and equation 3 had R^2=0,56.

Moschini el al's black point model

As earlier mentioned, black point is a crop decease caused by the fungi *Alternaria,* and however not harmful to animals and humans, it discolor the grain and thereby makes the grain hard to sell with the consequence of economical loss for the farmers. Moschini et al developed a prediction model based on meteorological variables in the Argentinean Pampas region (Moschini, 2006). The model consists of an equation which accounts for 87% of the total variance in decease incidence (% of kernels discolored).

Black point incidence was recorded at 5 different locations for three consecutive years, 1995, 1996 and 1997 and meteorological data were collected from the same years during the period from wheat heading (Zadoks 59) to harvest. The meteorological variables used are listed in table 4.

Table 4: *The variables used in the Moschini et al model*

Variable	Explantation
Td	Daily mean temperature
DD	Degree-day. The accumulation of mean temperatures during a period of days
MTx[a]	Mean value of daily maximum temperature
MTn[a]	Mean value of daily minimum temperature
DDTx[a]	Accumulation of days with exceeding daily maximum temperatures over a threshold (28-32°C)
DDTn[a]	Accumulation of days with exceeding daily minimum temperatures under a threshold (9-13°C)
DDTd[a]	Accumulation of days with mean temperature over a threshold (7-17°C)
TPr	The sum of total daily precipitation
DPr[a]	Accumulation of days with precipitation
DRH[a]	Accumulation of days with relative humidity over a threshold (60-85%)
DPrDDTd[a]	A combination of DPr and DDTd, computed as the multiplication of the variables

[a]) **Calculated over proposed time periods.**

To calculate the CPL, a computer program analyzed several irregular time periods with varying length, from heading (Zadoks 59) to ripening (Zadoks 90), in search of the time period with the strongest associations between weather variables and decease data. The coefficient of determination, R^2, was used to rank the different time periods, and the result was that the CPL were defined as to start 543 DD from heading and to end 861 DD from heading.

Different combinations of variables were tested with regression analysis to compute equations with high predictive ability. The equation with the highest coefficient of determination (R^2) was:

PI% = -6,50 + 0,07 DPrDDTd + 0,23 DRH, where PI% is predicted decease incidence, and the equation has a value of R^2=0,870.

The model was validated with an independent validation set of samples from the same five locations used for training, although the validation used data from the next growth season, 1998.

3.5. Indicator selection

In selecting the appropriate weather variables that have good prediction ability for the biological factors involved, a literature study was made. A Dutch study published 2009 had a similar topic as this project, although regarding a different but related fungi (van der Fels-Klerx, 2009). They conducted a literary study in combination with expert opinion to identify the 12 foremost important indicators for emerging mycotoxins in wheat cultivation. These indicators were used as a guideline as of pointing in the right direction in this literature study.

Table 5: *The 12 most important indicators for identification of emerging mycotoxins according to the van der Fels-Klerx study* (van der Fels-Klerx, 2009)

Rank	Indicator
1	Relative humidity/Rainfall
2	Crop rotation (previously cultivated crop)
3	Temperature
4	Tillage practice
5	Water activity in the kernels
6	Crop variety
7	Harvest conditions
8	Changes in fungal populations
9	Fungicide use
10	Plant health
11	Regional infection pressure
12	Awareness of food safety

3.6. Design of predictive models

Analysis of correlation between weather variables and incidence of mycotoxin was performed using the software Weka 3.6 (http://www.cs.waikato.ac.nz/ml/weka/). Weka is a java based collection of pattern recognition algorithms developed at the University of Waikato in New Zealand (Frank, 2005). Tools for data pre-processing, regression analysis among others are included in the Weka software.

The built-in Weka function for classifier design called ”Least median squares” (LMS) was used for the design of predictive models in this project. The LMS design procedure tries to numerically minimize the median squared error between model predictions and observed values. This means that the LMS function generates regression equations in a least squared sense from random subsamples of the dataset, and the best equation is chosen in terms of the lowest median squared error.

The design procedure were evaluated with Leave-one-out cross validation when the number of samples was lower or equal than 10, and otherwise, K-fold cross validation was used with K set to 10. The usage of cross validation implicates that the design procedure was evaluated on all available data, which in consequence might cause overtly optimistic results.

During the cross validation procedure Weka calculates the correlation coefficient for each fold describing the agreement between observation and prediction. The actual correlation coefficient output is then the average generated from all the folds.

3.6.1. Alternaria model

The number of spores per day was sampled during a 30 year time span, from 1980 to 2010, but since precipitation was only registered up till 1997, the dataset was reduced to 17 years. The dependent variable spores/day is here the mean value of the total number of spores registered during the year. The model used the weather variables *total amount of rain* (Rtot) and number of days with *mean daily temperature over 15°C* (Tsum) as independent variables and the CPL was 4-6 weeks after flowering. The design procedure was evaluated using k-fold cross validation width k = 10.

3.6.2. Black point model

The Black point model were developed from the Lantmännen dataset. Black point incidence were sampled in 17 farms during three years. However, some of the farms were situated to close

geographically to be separated by the weather data, and was thereby excluded from the dataset, which then was reduced to nine measure points of Black point incidence.

The Black point incidence was linked with weather data from approximately three month pre-harvest and the CPL used were identical to Moschini et al's research, which means the interval of 540-860 degree days after heading. Heading date for wheat in the sampled area varies but was approximated to June 7st. The model used *total amount of rainfall* and *number of days with mean daily temperature exceeding 15°C* as independent variables. Because of the low number of instances, the design procedure was evaluated using Leave-one-out cross validation, which in this case means K-fold cross validation with K=9.

3.6.3. Toxin model

The toxin model were developed using Per Hägglunds data from the SVA research. It consisted of measurements of Tenuazonic acid concentrations at 33 farms in the central and southern part of Sweden in 2006. The weather data for modeling toxin concentration holds the same disadvantage as did the Black point dataset, and the low resolution of weather data reduces the dataset to 13 unique locations. But the low quality weather data should be given even more credit here since the Black point weather data, although with low resolution, had a dataset covering 3 years, while the toxin data did only cover one year, and thus the variety in the weather data becomes even smaller.

The variables used in the modeling was *total amount of rainfall (Rtot)* and *number of days with mean daily temperature exceeding 15°C* (Tsum) and the CPL was chosen to be identical to DONcast's CPL, id est spanning from 7 days prior to heading to 10 days after heading.

Studying this dataset, it was obvious that crop species played an important role in toxin concentration. For instance, toxin concentration measured on oat were much higher than any other crop, and up to ten times higher than measured on wheat for samples in the same vicinity. Since the same weather data could result in such a great span of toxin concentration depending on crop species, two models were developed based on the two crop species which provided the largest quantity of samples in the dataset, namely barley and wheat. Each of the designs was evaluated with the Leave-one-out cross validation method.

4. Results

4.1. Indicator selection

The literature study on indicator selection resulted in the following indicators, sorted by category. Each indicator is then further elaborated on in the discussion section.

Table 6: *Indicators*

Category	Indicator	Explanation	Platform
Pre-harvest			
Precipitation			Alt,Bp,Tx
	Rtot, [1,3,4]	Total amount of rain during a period of time	
	Rd, [1, 3]	Days with rain	
	Rn, [1,2]	Number of days with rain over a threshold n	
	R7, [9]	*Current growth stage <Z65:Annual rainfall >=700mm	
		*Current growth stage >Z65:rain in the last 7 days >5mm	
Temperature			Alt,Bp,Tx
	DD, [1]	Degree days. Accumulation of the mean daily temperature during several days	
	DDTn, [1]	Accumulation of the exceeding temperatures under a threshold	
	DDTx, [1]	Accumulation of the exceeding temperatures over a threshold	
	Tsum, [1, 2]	Days with mean temperature over a threshold	
	Tmin, [1, 3, 5]	Minimum temperature registered during a day	
	Tmax, [1, 3, 5]	Maximum temperature registered during a day	
	Tmean, [1, 4, 5]	Mean value of the temperature registered during a day	
Humidity			Alt,Bp,Tx
	RHn, [1]	Number of days with relative humidity over a threshold n	
	RH, [3, 4]	Relative humidity registered during a day	
	aW, [7]	Water activity	
	LWD, [8]	Leaf wetness duration	
Combinations			Bp
	DPrDDTd, [1]	Days of precipitation * total degree-day accumulation of mean daily temperature greater than a given threshold.	
Agrarian			Alt,Bp,Tx
	Crop rotation, [6]	Previous crop cultivated on the field	
	Tillage practice, [6]	Method of tillage	
	Crop varieties, [6]	The specific crop species	
	Dom. Species, [6]	The dominating *Alternaria* species in the area	
	Crop stress, [6]	Crop stress such as weather damage, late harvest etc.	
Post-harvest			Tx
	aW in kernels, [6]	Water activity in kernels	
	RH in product, [6]	Relative humidity in the product	
	Ventilation, [6]	Ventilation during transport and storage	
	Temperature, [6]	Temperature during transport and storage	

Indicators with explanations along with a note on witch model they correspond to (*Alternaria*- (Alt), Black point- (Bp) or toxin-model (Tx)). *1.* (Moschini, 2006), *2.* (Hooker, 2002), *3.* (Tarekegn, 2006), *4.* (Katial, 1997), *5.* (Iglesias, 2007), *6.* (van der Fels-Klerx, 2009), *7.* (Magan, 1984), *8.* (Detrixhe, 2003), *9.* (Bailey, 2000).

4.2. Prediction models

Four equations were developed with the intention of predicting mycotoxical concepts and two more was adopted from an Argentinean study (Moschini, 2006) and evaluated on Swedish data. The scarce pool of data limits the models predictive value and they should only be viewed as very rough guidelines at this stage of development, but they could be somewhat meaningful as decision support for sampling.

All available data were used in building the models and the correlation coefficients derives from the build in cross validation feature in Weka (more details can be found in chapter 3.6.).

Table 7: *Prediction models*

Modeled aspect	Regression equation	Correlation
Alternaria	0,47 * Rtot – 0,83 * Tsum + 27,92	r=0,15
Toxin – Wheat	-18,02 * Rtot – 67,27 * Tsum + 881,93	r=-0,43
Toxin – Barley	-14,83 * Rtot + 376,33 * Tsum – 2879,50	r=-0,22
Black point – Moschini model 1	0,06 * DPr * DDTd – 3,94	r=-0,10
Black point – Moschini model 2	0,07 * DPr * DDTd + 0,23 * DRh – 6,50	r=-0,11
Black point – Swedish model	-0,04 * Rtot + 0,58 * Tsum – 4,78	r= 0,51

The developed regression equations in correspondence with their respective correlation coefficient resulting from the cross validation process.

5. Discussion

This section presents commented visualizations of sampled incidence versus predicted incidence from the cross validation process for all the prediction models along with a discussion on the results from the literature study regarding the different indicators. The visualization shows sampled values along with predicted values from each CV fold and predictions are calculated from the prediction model trained on all the other samples, that is, on all the other folds. The discussions are separated by respective modeled aspect and all of the post-harvest discussion is merged into one section.

5.1. Alternaria model

Looking at the table below, we can see that a large difference between sampled and predicted values usually occurs simultaneously with atypical values on Rsum (Total amount of rain), and most often in combination with low values on Tsum (number of days with mean temperature below 15°C), i.e. a cold, and wet or dry summer.

Table 8: *Data for Alternaria model*

Year	CV-fold	Sampled	Predicted	Deviation	Rsum	Tsum
1980	4	18,68	23,484	4,803	9,6	11
1981	5	17,699	27,219	9,52	6,4	8
1982	6	22,131	40,785	18,654	33,8	5
1983	8	26,316	10,522	-15,793	11,8	12
1984	2	27,679	24,707	-2,971	3,8	3
1985	4	21,983	20,051	-1,932	2,4	11
1986	8	19,39	21,987	2,597	6	15
1987	1	8,951	26,562	17,611	9,6	9
1988	5	13,639	22,028	8,389	12	16
1989	3	38,574	12,943	-25,631	0,2	15
1990	9	49,581	31,936	-17,646	64,2	14
1991	7	33,913	17,823	-16,09	0	11
1992	1	45,676	22,535	-23,141	9,9	13
1993	2	32,019	24,726	-7,292	11,5	9
1994	6	9,638	17,772	8,134	0,6	15
1995	10	26	18,378	-7,622	1,8	11
1996	3	35,795	47,926	12,131	28,9	7
1997	7	21,356	17,313	-4,043	3,2	13

Data obtained from the cross validation procedure along with the two weather variables used in the prediction model for *Alternaria* incidence. The type of cross validation used is 10-fold cross validation.

The *Alternaria* model fails to predict the major peaks and dips of the sampled spores/day in a moderately accurate manner. The model predicts at its best when the total amount of rain is low, but not too close to zero as in 1984-1986 and in 1997. It has a tendency to exaggerate the predicted spores/day when the summer is cold, as in 1982-1983, 1987 and 1996, but this is contradicted by the coldest summer in the dataset, which is accurately predicted namely 1984.

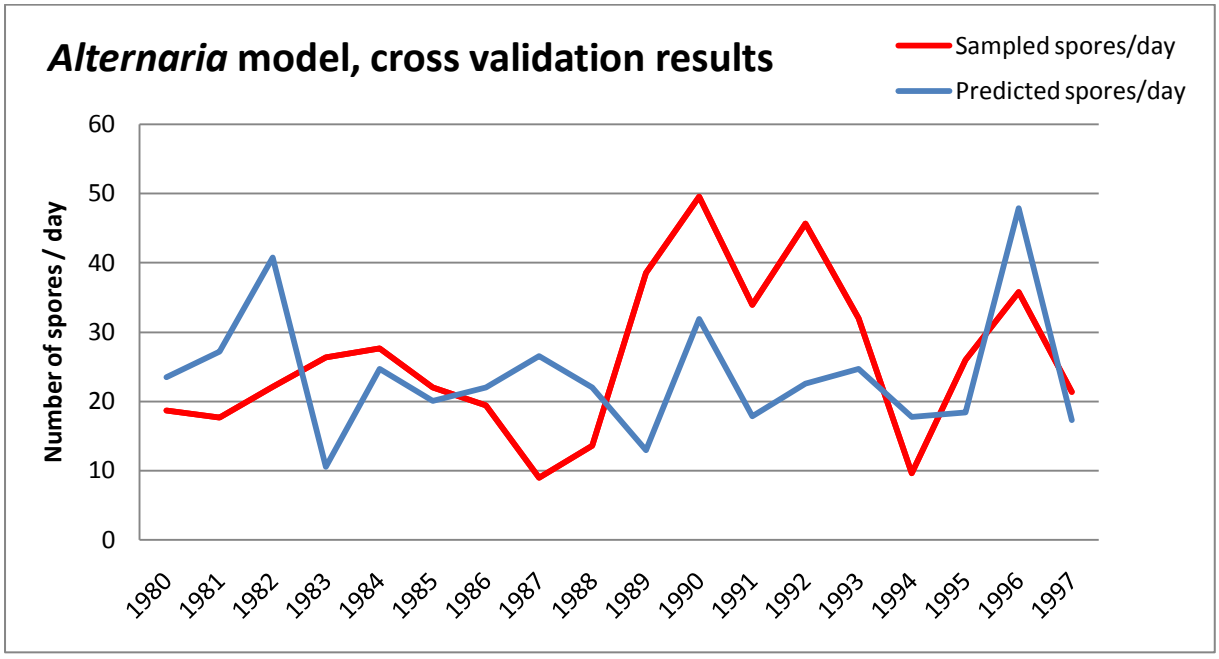

Figure 4: Results from the 10-fold cross validation procedure. The figure describes sampled spores per day and modeled spores per day. Cross validation with 10 folds and 18 samples means that eight folds consists of two years/samples and two folds have only one year/sample. The predicted spores/day for each year is predicted by the model trained using all other folds.

5.2. Alternaria indicators

5.2.1. Pre-harvest indicators

The literature agrees on defining the most important pre-harvest weather indicators for Alternaria as precipitation, temperature and humidity. Agrarian indicators are also an important factor in *Alternaria* prediction models and the topic is discussed below.

Precipitation

Several articles has indicated that rain at certain points in time in the crops maturity process is integral in the development of *Alternaria* infection on the crop (Iglesias, 2007), (Tarekegn, 2006). Tarekegn found correlations between incidence of *Alternaria* moulds and weather variables such as total amount of rainfall 4-6 weeks after flowering as well as 5-8 weeks after flowering, depending on the crop hybrid. The variable *R7* is related to a decision system for prediction of sooty mold decease which in turn is caused by *Alternaria* infection (Bailey, 2000). The reason for its low relevance here steams from its implicit nature.

Temperature

Temperature is highly important regarding both *Alternaria* infection and its production of mycotoxins. A predictive model must test if the temperature variable is inside the boundaries of the favorable conditions for *Alternaria* development, and for how long it remains there. Optimum

temperature for *Alternaria* development lies in the interval 15°C < t <24°C (Battilani, 2009), (Magan, 1984). This can be described by degree days, which is defined as the sum of daily mean values over a period of time. *Tmin* and *Tmax* are also of significance since values on these two variables, outside of the temperature boundaries for *Alternaria* growth and mycotoxin production, indicate that no growth and production is in process. Because of the relatively cold summers in Sweden, *Tmin* might be more relevant than *Tmax*.

Humidity

Relative humidity and water activity is two closely related variables, and since one is build up by the other, their separation here is in need of a comment. The use of *RH* as a variable basically make a_W redundant but considering the fact that a_W is quite frequently used in the literature that focus on the biological aspects of this subject, it is of some importance. It is also of major significance in the post-harvest contamination process. A Belgian study proposes leaf wetness duration as an indicator and argues that it has a strong relationship with plant deceases since many pathogens needs a layer of water to move on the surface of the plant and to start their infection processes. The argument makes sense but the practical measurement would be too complicated to incorporate in a model of this type.

One way to implement humidity in a model is by using interrupted wet periods, or *IWP*. An *IWP* day is when relative humidity is higher than 95% for six consecutsive hours at night followed by relative humidity lower than 80% during the day, for six consecutive days. This indicator has been favorably used predicting infection from *Alternaria* species on potatoes in Spain (Iglesias, 2007)

Agrarian

There has been little research on quantifying the effect of agrarian indicators in prediction models, but a qualitative consensus exists. Crop rotation is the foremost important agrarian indicator of *Alternaria* development, followed by tillage practice and crop varieties (van der Fels-Klerx, 2009). Regarding crop stress, in a study by Hudec *Alternaria* incidence in barley at four different locations in Slovakia for two consecutive years were measured (Hudec, 2007). 32 Samples were taken at standard harvest time and at late harvest time. The result showed that a late harvest caused increased *Alternaria* incidence in only 15 of the 32 samples and in the other 17, the incidence had decreased. The result is in contrast to the general hypothesis that late harvest stimulates *Alternaria* incidence, and also serves as an example of the uncertainty of crop damage as an indicator for *Alternaria* development.

5.3. Black point model

In general, the deviation between sampled and predicted values is quite small with the exception of Nybble, Fransåker in 2002 and Kölbäck. There is no obvious pattern in these exceptions as they have very typical values on the temperature variable and widespread values on the precipitation variable.

Tabell 9: *Data for Black point model*

Location	CV-fold	Sampled	Predicted	Deviation	Rsum	Tsum
Vintrosa, 2001	1	0,867	1,666	0,8	63	14
Borgeby, 2002	2	4,96	5,31	0,35	48	18
Nybble, 2002	3	2,86	0,411	-2,449	123	17
Fransåker, 2002	4	2,84	4,775	1,935	69,9	18
Fransåker, 2003	5	3,9	3,472	-0,428	18,4	16
Kölbäck, 2003	6	8,7	4,428	-4,272	35	17
Borgeby, 2003	7	3,93	5,199	1,269	47	18
Brunnby, 2003	8	3,63	3,539	-0,091	21	16
Kampetorp, 2003	9	5,4	4,352	-1,048	43	18

Data obtained from the cross validation procedure along with the two weather variables used in the prediction model for Black point incidence. The type of cross validation used is Leave-one-out cross validation, hence one sample in each CV-fold.

Six of the nine predictions have a deviation of less than 1,3 percentage. The biggest deviation derives from Kölbäck where sampled and predicted value deviates 4,3 percentage. The weather in Kölbäck were not atypical.

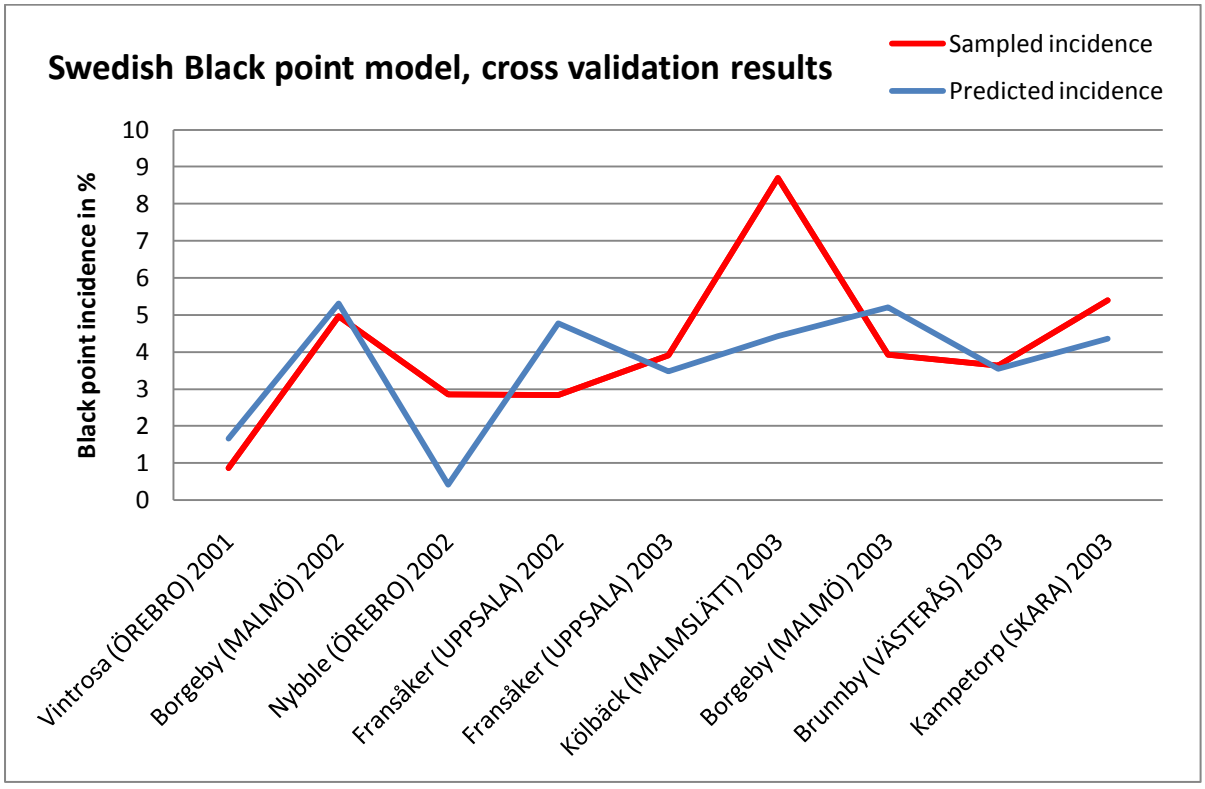

Figure 5: Predicted Black point incidence derived from the leave-one-out cross validation result along with sampled incidence. Each sample in the graph has been used for both training and evaluation, but not at the same time. The data originates from the Lantmännen dataset. The location inside the parenthesis is the closest available weather station.

An attempt using Moschini et al's Black point models on the Swedish dataset from Lantmännen resulted in correlation coefficients of 0,10 and 0,11 for model 1 and model 2 respectively. The poor result probably relates to the fact that the test was made with the same parameters that wore adjusted for Argentinean conditions, and tweaking these parameters with regards to Swedish environmental factors might give a better result. Also, adjusting the threshold for the variables themselves, or the CPL, might prove worthwhile.

Table 10: *Data for the Argentinean model*

Location	Sampled	Model 1	Model 2	Deviation (1st, 2nd)		Dpr	DDTd	DRH
Vintrosa, 2001	0,867	2,06	4,87	1,19	4,0	10	10	19
Borgeby, 2002	4,96	-0,04	2,19	-5	-2,77	5	13	18
Nybble, 2002	2,86	1,1	3,06	-1,76	0,2	7	12	16
Fransåker, 2002	2,84	4,16	7,09	1,32	4,25	9	15	18
Fransåker, 2003	3,9	-0,34	0,92	-4,24	-2,98	4	15	14
Kölbäck, 2003	8,7	0,86	3,01	-7,84	-5,69	5	16	17
Borgeby, 2003	3,93	0,14	2,4	-3,79	-1,53	4	17	18
Brunnby, 2003	3,63	1,46	3,48	-2,17	-0,15	6	15	16
Kampetorp, 2003	5,4	3,62	6,46	-1,78	1.06	9	14	18

The two Moschini et al's Black point models evaluated on the Lantmännen dataset. The deviation is listed as deviation of first model and deviation of second model.

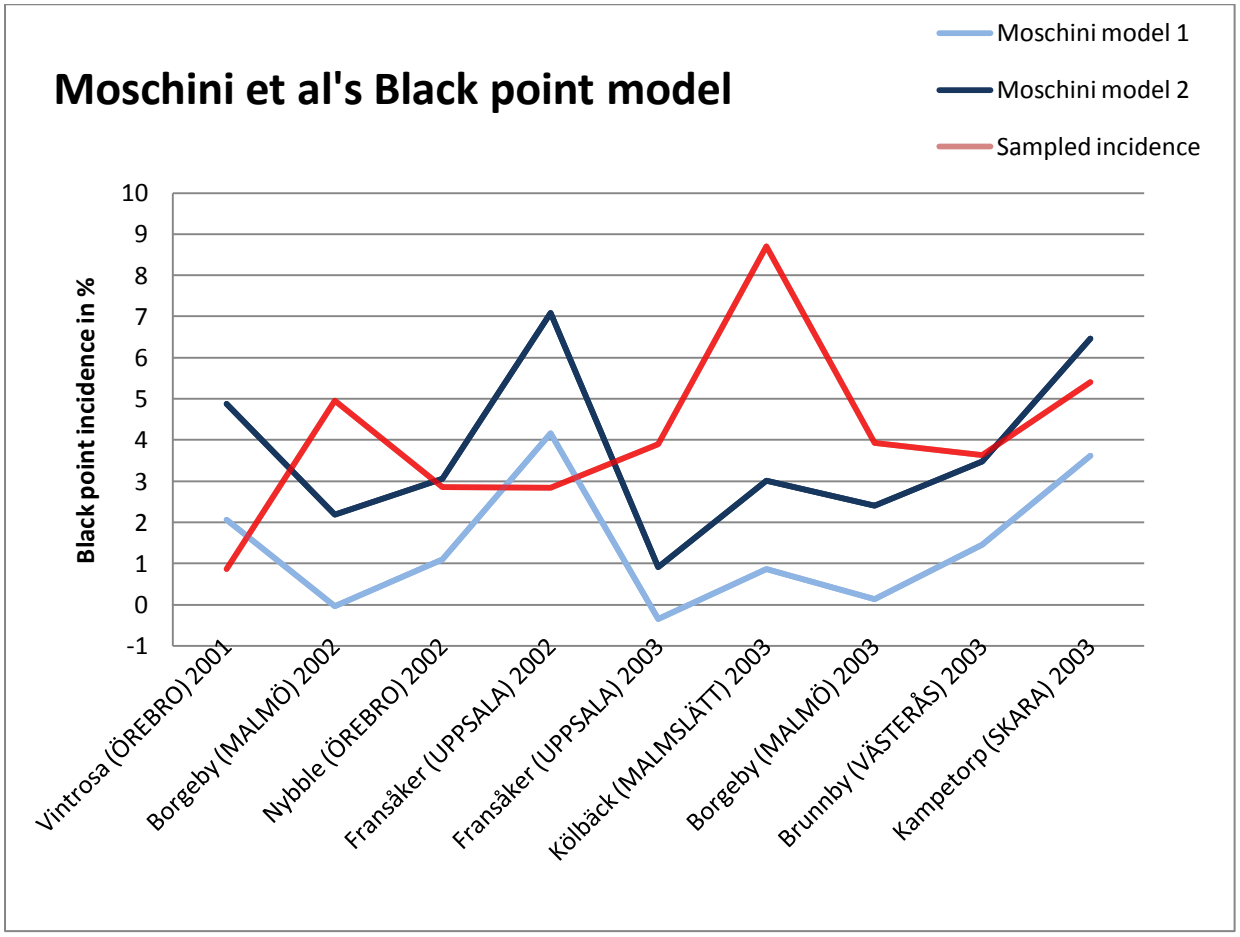

Figure 6: Predicted incidence using the two models proposed by Moschini et al 2006 versus measured incidence of black pointed kernels from the Lantmännen dataset. The location inside the parenthesis is the closest available weather station.

5.4. Black point indicators

Literature on black point prediction models is scarce and only two reports on the subject, one from Argentina and one from Italy (Moschini, 2006), (Languasco, 1994), where found.

5.4.1. Pre-harvest indicators

The Argentinean study found the susceptible time interval for the effect of weather variables to occur during the milk and mealy dough wheat stages, which in the Zadoks scale translates to Z71 - Z87 (Moschini, 2006). In other terms, the period coincides with 550 degree days after wheat heading

to 850 degree days after heading. This is defined as the *CPL* or critical period length. The Italian study found significant correlations between weather variables and black point incidence the first decade after heading, which is a *CPL* that does not overlap with the *CPL* in the Argentinean report.

Precipitation

There are three common ways to quantify precipitation; number of days with rainfall, number of consecutive days with rainfall and amount of rain fallen. Moschini uses days with rain (Moschini, 2006), while Languasco finds significant correlations (r=0,84; P≤0,01) between amount of rain in mm during the first decade after heading with regards to the Zadoks scale (Languasco, 1994).

Temperature

Moschini utilizes degree days both for locating the *CPL* and in a predictive fashion as a variable in the model. He proposes five different variables based on temperature, of which three are new to this report. The new variables are *DDTx*, *DDTn* and *DDTd. DDTx* and *DDTn* are both accumulation of exceeding temperatures outside a threshold of maximum and minimum respectively. *DDTd* is the accumulation of positive degree days over threshold.

Combined

The study by Moschini combined two different weather variables by multiplying their effect and then made a stepwise regression analysis to find the relation between the variable and black point decease. The suggested variable was called *DPrDDTd* and consists of *DPr* which means number of days with precipitation and *DDTd* using a threshold of 17°C. After defining the parameters, this variable explained 84% of the variance in black point incidence in wheat during three consecutive growing seasons at different locations in the Argentinean Pampas region. In combination with a variable based on days with relative humidity over 62%, the explained variance increased to 87%.

Agrarian

Reports show that the sensitivity for black point infection in different varieties of wheat varies significantly. There exist resistant varieties like Benito, Glenlea and Park, as well as varieties with intermediate sensitivity, Leader and Sadash, and susceptible ones like all durum and soft white spring wheat (Saskatchewan). If there would be an agronomical category included in the prediction model for black point decease, inclusion of crop varieties could play a fundamental part.

5.5. Toxin model

The toxin model for barley predicted TeA concentrations rather accurate in two out of five instances, but is very inaccurate in the fifth instance.

Table 11: ***Data for the Toxin model on barley***

Location	CV-fold	Sampled	Predicted	Deviation	Rtot	Tsum
Klippan	1	101,5	-3,419	-104,919	2	8
Kristianstad	2	374	0,179	-373,821	9	9
Svalöv	3	57	80,5	23,5	5	8
Uppsala	4	94,5	70,243	-24,257	3	8
Vintrosa	5	111	858,643	747,643	2	10

Data obtained from the cross validation procedure along with the two weather variables used in the prediction model for Toxin incidence in barley. The type of cross validation used is Leave-one-out cross validation, hence one sample in each CV-fold.

The models inaccuracy at Vintrosa, a very warm and dry location at the time of measurement, is probably due to the model's high sensitivity for temperature. This sensitivity might result from the lack of a sufficient quantity of training data, which does not only apply to the toxin dataset but were a dilemma throughout the project.

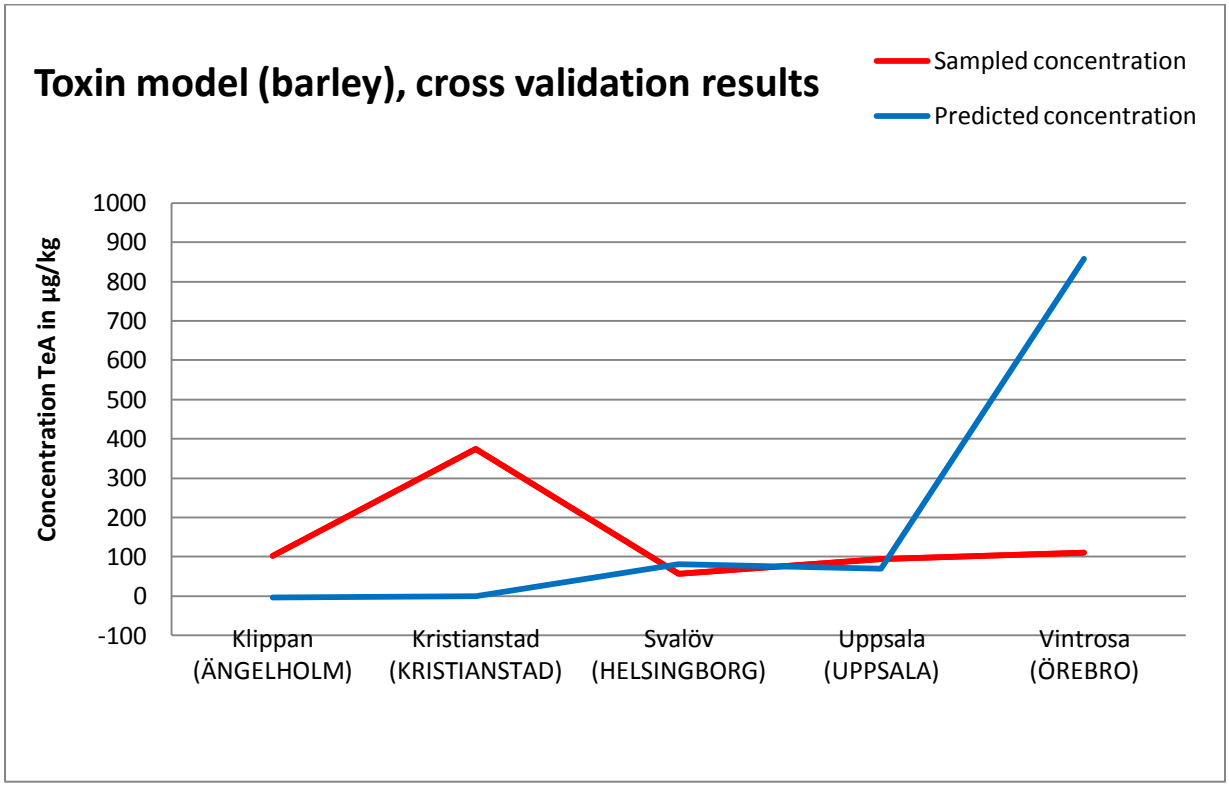

Figure 7: Sampled concentration of Tenuazonic acid in barley along with predicted concentration. The sampled concentration is derived from the SVA dataset and the predicted concentration was modeled from all available data.

In general, the toxin model for wheat is quite consistent in its predictions, with deviations of more or less the same magnitude across the board. The sample at Söderköping could almost be seen as an outlier with roughly three times the toxin concentration of the second highest sample.

Tabell 12: *Data for the Toxin model on wheat*

Location	CV-fold	Sampled	Predicted	Deviation	Rtot	Tsum
Gamleby	1	59	201,167	142,167	6	10
Klippan	2	73	307,697	234,697	2	8
Kristianstad	3	70	202,371	132,371	9	9
Linköping	4	182,5	39,837	-142,663	6	10
Skara	5	224	122,968	-101,032	1	9
Svalöv	6	293	60,703	-232,297	5	8
Söderköping	7	813	111,918	-701,082	7	10

Data obtained from the cross validation procedure along with the two weather variables used in the prediction model for Toxin incidence in wheat. The type of cross validation used is Leave-one-out cross validation, hence one sample in each CV-fold..

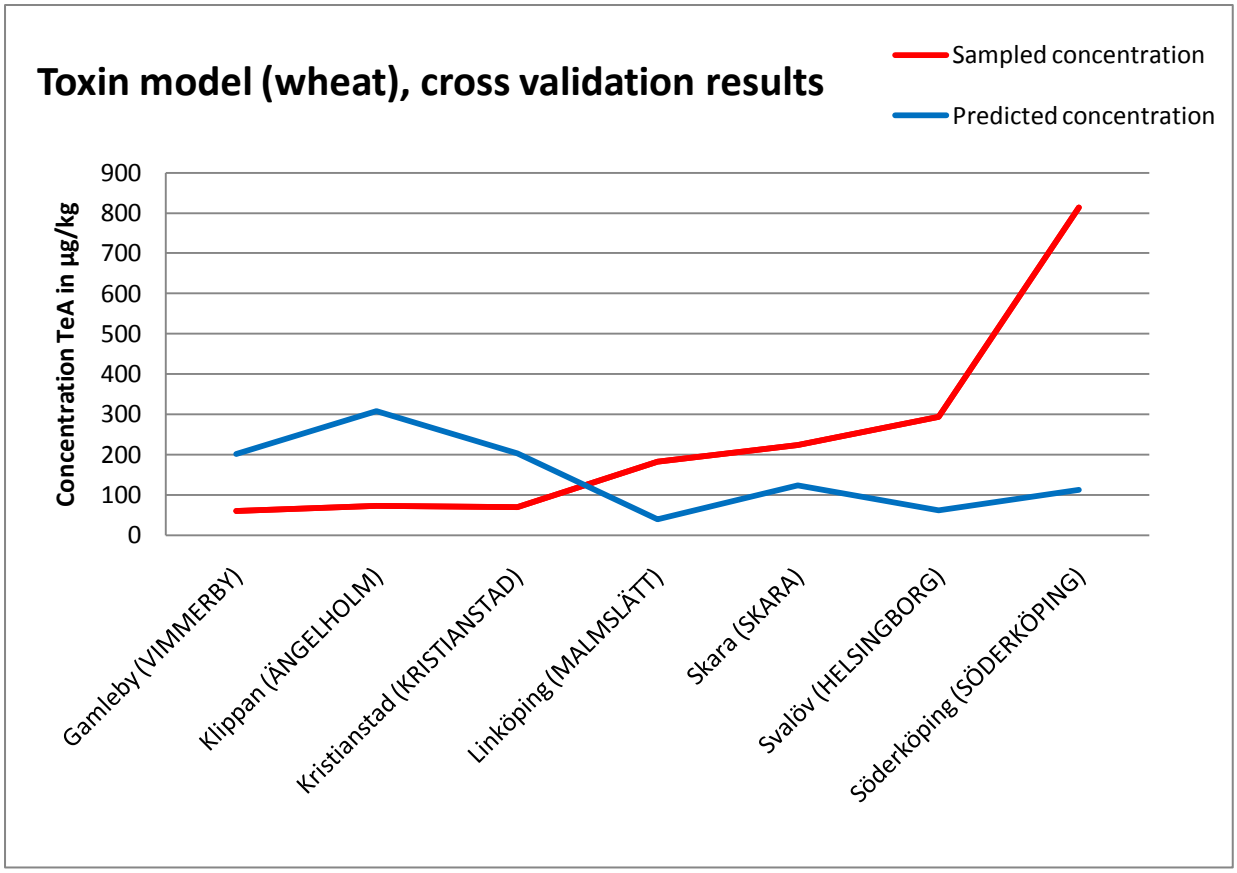

Figure 8: Sampled concentration and predicted concentration of Tenuazonic acid in wheat. The sampled concentration is derived from the SVA dataset.

5.6. Toxin indicators

5.6.1. Pre-harvest indicators

There exists several prediction models for mycotoxins but none is specialized on predicting TeA. The Canadian prediction model DONcast is arguably the most renowned and is already operating on a commercial level forecasting the mycotoxin Deoxynivalenol (Hooker, 2002). Although, not directly related to TeA, DONcast can be seen as a point of aim in the development of a prediction model for TeA.

Precipitation

DONcast utilizes four different variables to incorporate rain in their model, of which three are number of days with rain over a threshold during different periods and the last one is the square of one of the aforementioned. In more detail, *RAINA* equals number of days with rainfall >5mm in the 4-day period from 4-7 days before wheat head emergence (Zadoks 59), $RAINA^2$ is the square of that number, *RAINB* equals number of days with rainfall >3mm in the 4-day period from 3-6 days after wheat head emergence and *RAINC* equals number of days with rainfall >3mm in the 4-day period from 7-10 days after wheat head emergence.

Temperature

DONcast uses two variables to measure temperature. TMIN equals number of days with mean daily temperature <10°C during the 4-day period before wheat head emergence and TMAX equals mean daily temperature >32°C in the 4-day period 3-6 days after wheat head emergence. Optimal growth

temperature for Deoxynivalenol is 25°C (Ramireza, 2006), while Tenuazonic acid production by *A. tenuissima* peaks at 20°C (Magan, 1984), (Young, 1980), (Battilani, 2009), which indicates that if DONcast would be used as a point of reference, a fine-tune of the temperature thresholds might be necessary.

Humidity

Optimal TeA production occurs at 1,0 a_w (Young, 1980).

Agrarian

Dominating *Alternaria* species is a factor both in black point incidence and mycotoxin prevalence. It is well established that *A. tenuissima* is a potent TeA producer, as well as *A. alternata* though in smaller scale (Patriarca, 2007), but when *A. infectoria* is the dominating species the incidence of TeA is very low, (Webley, 1997). An Australian study by Webley in 1997 has an interesting implication regarding *A. infectoria* and *A. alternata* infection when he took the fact that *A. alternata* is a TeA producer in concern plus the fact that black point is associated with *A. alternata* infection (Webley, 1997), (Özer, 2005). By analyzing weather damaged and black point infected wheat along with healthy samples of wheat with different dominating *Alternaria* species, Webley found that there were only mycotoxins in the samples if the dominating *Alternaria* species was *A. alternata*. This meant that even if the sample were weather damaged and had moderate black point decease; there would be no measureable levels of mycotoxins, even in the presence of *A. alternata,* if the dominating species in the region were *A. infectoria*. The interesting aspect here is that while the implication black point -> *Alternaria* may hold, there is no necessary connection between black point -> *Alternaria* -> mycotoxin. This means that black point might be an indicator for mycotoxins in a prediction model, but only so in combination with dominating species. Webley's results also showed that the relative infection of *A. alternata* species in a field needs to be more than 40% to find moderate incidence of TeA.

Regarding dominating species, a Norwegian study focused on pointing out the dominating *Alternaria* spp. in Norway shows that the most common species in Norway is *A. infectoria,* closely followed by *A.tenuissima* (Kosiak, 2004). A similar study in Denmark on barley shows that the dominating species in Denmark is *A. infectoria* (Andersen, 1996). A study by Gannibal investigated the intraspecies distribution of *A. tenuissima* isolates in Russia (Gannibal, 2007). He found that *A. tenuissima* is the dominating *Alternaria* species in remotely geographical locations in the northern parts of Russia and that there is no host-specificity within the species. In a study on dominating species in Mediterranean countries, the dominating species was *A. alternata and A. triticina* while there were no sign of *A. tenuissima* (Logrieco, 1990). There is also a difference in optimal temperature for mycotoxin production by *A. alternata* and *A. tenuissima*. *A. alternata* has production optima at 25°C while *A. tenuissima* has an optima at 20°C (Paterson, 2009). All things considered, this information makes an insinuation that *A. tenuissima* could be a species of *Alternaria* that favors a colder climate, and may therefore be a dominating species in some parts of Sweden, but further research on the subject is needed to form and validate a more detailed hypothesis.

5.7. Post-harvest indicators

The post-harvest domain is essentially a black box since very little is known about the quantitive effect of post-harvest variables on mycotoxins, with the exception of the consequence of certain preventive measures. Cooling prolongs the shell life of fruits, dipping in hot water dramatically

inhibits disease caused by *Alternaria* spp., Gamma-irradiation has a preventive effect on TeA production by *A. alternata* and particular essential oils and fungi can be used with success as biocontrollers for certain *Alternaria spp*..

Despite the large uncertainty regarding the post-harvest process and the impact of post-harvest variables, the predictive problem is somewhat reduced since it is uncommon for new types of fungi to add to the equation while in transportation and storage condition. This means that the condition of the indicator variables are stable, there will be no change in dominating fungi, etc..

In a study on the effect of temperature and moisture on TeA production for *A. tenuissima* on cotton seed, the optimum production occurred at 20°C and 1.0 a_W (Young, 1980). The production optima of *A. alternata* occur at 25°C, which is slightly higher than its growth optima at 23°C (Paterson, 2009).

6. Conclusions

To conclude, I have here proposed four rough prediction models regarding different aspects of mycotoxins in grain along with important factors to include in a more thorough prediction model on a national level. These important factors can be divided into weather indicators and agrarian indicators, where the most essential weather indicators would be precipitation, temperature and humidity, either in combination or by themselves. Key agrarian indicators are suggested to be crop rotation, crop varieties and dominating *Alternaria* species. The mycotoxin and black point models presently existing in other countries are not immediately applicable on the colder Swedish climate, but with some adjustment they could represent guidelines for further studies. Furthermore, defining the CPL is central for the accuracy of a prediction model of this kind, but since there is no research on this subject, the actual CPL for Swedish weather conditions is regrettably unknown.

6.1. Future challenges

The predictive value of a model is depending largely on the quality of the indicators involved, the level of detail of sampled weather variables etc. The optimal conditions for producing a predictive model on mycotoxins in grain would include a substantial amount of data. A subject matter on which there is a lot more work to do. The way to develop a CPL for Swedish conditions is in my opinion rather straight forward. One have to test all indicators on all time periods during several seasons to find out which variables and time periods that have the greatest discriminatory ability. For this to work, again, more data is needed. During my brief modeling stint, one of my biggest concerns turned up to be getting hold of the correct weather data. Partly, this resulted from the difficulty of locating the closest weather station for the related data, and then merging and trimming it in an orderly fashion. A task much more complicated than the sound of it. Also, more research on the quantitative effects of post-harvest factors is necessary for including such variables in a prediction model. All in all, suggested future research areas are regarding the CPL for *Alternaria* mycotoxin production, quantitative post-harvest effects, pathways to streamline the acquisition of data for the model and foremost, more data needs to be sampled, and preferably from different growth seasons, to achieve higher predictive value.

6.2. Acknowledgements

I would like to thank my supervisor Gunnar Andersson for giving me the opportunity to conduct my M. S. thesis project at SVA and the Department of Animal Feed. Thanks for the invaluable guidance

and input. I would also like to thank Mats Gustafsson for accepting the role of scientific reviewer in this project. Furthermore, I want to thank Mats Lindblad at The Swedish National Food Administration, Tomas Börjesson at Lantmännen and Agneta Ekebom at The Swedish Museum of Natural History for their help in providing me with data. Finally, I would like to thank my family and friends for their support throughout my project.

References

Andersen, B., Kröger, E. and Roberts, R.G. 2001. Chemical and morphological segregation of Alternaria alternata, A.gaisen and A. longipes. *Mycological research.* Vol. 105:291-299.

Andersen, B., Thrane, U., Svendsen, A. and Rasmussen, I.A. 1996. Associated field mycobiota on malt barley. *Canadian Journal of Botany.* Vol. 74:854-858.

Bailey, B. 2000. *Decision Support System for Arable Crops (DESSAC).* Silsoe : Silsoe Research Institute, p40.

Battilani, P., Costa, L.G., Dossena, A., Gullino, M.L., Marchelli, R., Galaverna, G., Pietri, A., Dall'Asta, C., Giorni, P., Spadaro, D. and Gualla, A. 2009. *Scientific information on mycotoxins and natural plant toxicants.* s.l. : European Food Safety Authority, p126-193.

Blom, G., Enger, J., Englund, G., Grandell, J., Holst, L. 2005. *Sannorlikhetsteori och statistikteori med tillämpningar.* Lund : Studentliteratur, 91-44-02442-8, p358-371.

Detrixhe, P., Chandelier, A., Cavelier, M., Buffet, D. and Oger, R. 2003. Development of an agro-meteorological model integrating leaf wetness duration estimation to asess the risk of head blight infection in wheat. *Applied Biology.* Vol. 68:199-204.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B. and Witten, I.H. 2005. Weka - A Machine Learning Workbench for Data Mining. *Data Mining and Knowledge Discovery Handbook.* p1-10.

Franz, E., Booij, K., van der Fels-Klerx, I. 2009. Prediction of Deoxynivalenol content in Dutch winter wheat. *Journal of food protection.* Vol. 72:2170-2177.

Gannibal, P.B., Klemsdal, S.S and Levitin, M.M. 2007. AFLP analysis of Russian Alternaria tenuissima populations. *European Journal of Plant Pathology.* Vol. 119:175-182.

Giambrone, J.J., Davies, N.D. and Diener, U.L. 1978. Effect on tenuazonic acid on young chickens. *Poultry Science.* Vol. 57:1554-1558.

Hooker, D.C. and Schaafsma, A.W. 2002. The DONcast model: using wether variables pre- and post-heading to predict deoxynivalenol content in winter wheat. *Applied Biology.* Vol. 68:611-619.

Hudec, K. 2007. Influence of harvest date and geographical location on kernel symptoms fungal infestation and embryo viability of malting barley. *International Journal of Food Microbiology.* Vol. 113:125-132.

Häggblom, P., Stepinska, A. and Solyakov, A. 2007. Alternaria mycotoxins in Swedish feed grain. *Gesellschaft für Mykotoxin Forschung.* p1420-1422.

Iglesias, I., Rodríguez-Rajo, F.J. and Méndez, J. 2007. Evaluation of the different Alternaria prediction models on a potato crop in A Limia (NW of Spain). *Aerobiologiga.* Vol. 23:27-34.

Katial, R.K., Zhang, Y., Jones, R.H. and Dyer, P.D. 1997. Atmospheric mold spore couns in relation to meteorological parameters. *International Journal of Biometeorol.* Vol. 41:17-22.

Kosiak, B., Torp, M., Skjerve, E. and Andersen, B. 2004. Alternaria and Fusarium in Noweigian grains of reduced quality - a matched pair sample study. *International Journal of Foof Microbiology.* Vol. 93:51-62.

Languasco, L., Orsi, C. and Rossi, V. 1993. Forecasting black point of wheat using meteorological and fungal isolation data. *Istituto di Patologia Vegetale Università Cattolica del S. Cuore.* Vol. 7:203-209.

Logrieco, A., Bottalico, A., Solfrizzo, M. and Mule, G. 1990. Incidence of Alternaria species in grains from Mediterraniean countries and their ability to produce mycotoxins. *Mycologia.* Vol. 82(4):501-505.

Magan, N., Cayley, G.R. and Lacey, J. 1984. Effect of Water Activity and Temperature on Mycotoxin Production by Alternaria alternata in Culture and on Wheat Grain. *Applied and Environmental Microbiology.* Vol. 47:1113-1117.

McLachlan, G.J., Do, K.-A., Ambroise, C. 2004. *Analyzing microarray gene expression data,* 9780471226161.

Molinaro, A.M., Simon, R. and Pfeiffer, R.M. 2005. Prediction error estimation: a comparison of resampling. *Bioinformatics.* Vol. 21:3301-3307.

Moschini, R.C. and Fortugno, C. 1996. Predicting wheat head blight incidence using models based on meteorological factors in Pergamino, Argentina. *European Journal of Plant Pathology.* Vol. 102:211-218.

Moschini, R.C., Sisterna, M.N. and Carmona, M.A. 2006. Modelling of wheat black point incidence based on meterological variables in the southern Argentiniean Pampas region. *Australian Journa of Agricultural Research.* Vol. 57:1151-1156.

Paterson, R.R.M. and Lima, N. 2009. How will climate change affect mycotoxins in food. *Food Research International.* p1-13.

Patriarca, A., Azcarete, M.P., Terminiello, L. and Fernández Pinto, V. 2007. Mycotoxin production by Alternaria strains isolated from Argentinean wheat. *International Journal of Food Microbiology.* Vol. 119(3):219-222.

Prandini, A., Sigolo, S., Filippi, L., Battilani, P. and Piva, G. 2009. Review of predictive models for Fusarium head blight and related mycotoxin contamination in wheat. *Food and CHemical Toxicology.* Vol. 47:927-931.

Ramireza, M.L., Chulzeb, S. and Magan, N. 2006. Temperature and water activity effects on growth and temporal deoxynivalenol production by two Argentinean strains of Fusarium graminearum on irradiated wheat grain. *International Journal of Food Microbiology.* Vol. 106(3):291-296.

Rees, R.G., Martin, D.J. and Law, D.P. 1984. Black point in bread wheat:effects on quality and germination, and fungal associations. *Australian Journal of Experimental Agriculture and Animal Husbandry.* Vol. 24:601-605.

Rousseeuw, P.J. 1984. Least median of squares regression. *Journal of the American statistical association.* Vol. 79:871-880.

Saskatchewan, Agriculture Knowledge Center of. 2009. Sooty Moulds of Cereals at Harvest. *The Saskatchewan Agricultural website.* [Online] Government of Saskatchewan [Cited: 02 01, 2010.]

Saskatchewan, Government of. Blackpoint and Smudge of Wheat. *Government of Saskatchewan.* [Online] [Cited: 02 12, 2010.] http://www.agriculture.gov.sk.ca/Default.aspx?DN=c716c0a7-ed89-4ea8-82cf-775604bddaf4.

Shephard, G.S., Thiel, P.G.,Sydenham, E.W., Vlegaar, R. and Marasas, W.F.O. 1991. Reverced-phase high-performance liquid chromatography of tenuazonic acid and related tetramic acids. *Journal of chromatography.* Vol. 566:195-205.

Siegel, D., Rasenkoa, T., Kocha, M. and Nehls, I. 2009. Determination of the Alternaria mycotoxin tenuazonic acid in cereals by high-performance liquid chromatography–electrospray ionization ion-trap multistage mass spectrometry after derivatization with 2,4-dinitrophenylhydrazine. *Journal of chromatography.* Vol. 1216(21):4582-4588.

SMHI. Sveriges läns framtida klimat. *SMHI.* [Online] [Cited: 01 26, 2010.] http://www.smhi.se/klimatdata/klimatscenarier/klimatanalyser/Sveriges-lans-framtida-klimat-1.8256.

Tarekegn, G., McLaren, N.V. and Swart, J. 2006. Effects of wather variables on grain mould of sorghum in South Africa. *Plant Pathology.* Vol. 55:238-245.

University, Safety Officer in Physical Chemistry at Oxford. Safty (MSDS) data for caffein. [Online] [Cited: 01 28, 2010.] http://msds.chem.ox.ac.uk/CA/caffeine.html.

van der Fels-Klerx, H.J., Kandhai, M.C., Brynestad, S., Dreyer, M., Börjesson, T., Martins, H.M., Uiterwijk, M., Morrison, E. and Booij, C.J.H. 2009. Development of a European system for identification of emerging mycotoxins in wheat supply chains. *World Mycotoin Journal.* Vol. 2:119-127.

Webley, D.J. and Jackson, K.L. 1998. Mycotoxins in cereal - a comparison between North America, Europe and Australia. p63-66.

Webley, D.J., Jackson, K.L., Mullins, J.D., Hocking, A.D. and Pitt, J.I. 1997. Alternaria toxins in weather-damaged wheat and sorghum in the 1995-1996 Australian harvest. *Australian Journal of Agricultural Research.* Vol. 48:1249-1255.

Young, A.B., Davies, N.D. and Diener, U.L. 1980. The effect of temperature and moisture on tenuazonic acid production by Alternaria tenuissima. *Phytopathology.* Vol. 70:607-609.

Zadoks, J.C., Chang, T.T. and Konzak, C.F. 1974. A Decimal Code for the Growth Stages of Cereals. *Weed Research.* Vol. 14:415-421.

Zhou, B. and Qiang, S. 2008. Enviromental, genetic and cellular toxicity of tenuazonic acid isoltated from Alternaria alternata. *African Journal of Biotechnology.* Vol. 7:1151-1156.

Özer, N. 2005. Determination of the fungi responsible for black point in bred wheat and effects of the disease on emergence and seedling vigour. *Trakya University Journal of Science.* Vol. 6:35-40.

Appendix, Perl code

This is a Perl script written to concatenate selected parts of two files based on date. In specific, it matches the number of spores per cubic meter air in Stockholm with the correct SMHI weather data.

```
use 5.010;
use warnings;
#Defining variables to my files
my $infile_1 = 'sporer.txt';
my $infile_2 = 'vader.txt';
my $outfile = 'vader_sporer_1980-2009.txt';
#Defining variables for trimming
my $head = "D";
my $juni = '06';
my $juli = '07';
my $augusti = '08';
my $september = '09';
#Opening the first file, neatly correcting each line and putting it into the newly created array
open FIL, $infile_1 or die "Cant open 'sporer1.txt': $!";
my @array_sporer;
while(<FIL>){
		if (substr($_,5,2) eq $juni or substr($_,5,2) eq $juli or substr($_,5,2) eq $augusti or
substr($_,5,2) eq $september or substr($_,0,1) eq $head){
				chomp $_;
				push @array_sporer, $_;
		}
}
close FIL;
#Opening and storing the next file as previously explained
open FIL, $infile_2 or die "Cant open 'vader2.txt': $!";
my @array_vader;
while(<FIL>){
		if (substr($_,5,2) eq $juni or substr($_,5,2) eq $juli or substr($_,5,2) eq $augusti or
substr($_,5,2) eq $september or substr($_,0,1) eq $head){
				chomp $_;
				push @array_vader, $_;
		}
}
close FIL;
#Creating a file to write to
open UT, ">", $outfile or die "Cant open 'outfile': $!";
#Defining the header and writing it to the outfile
my $header = $array_sporer[0] . "\t" . substr($array_vader[0], 6) ;
say UT $header;

#Concatenates lines and writes them to the outfile IF the dates from the two files matches
```

```
my $count = 1;
my $arraylenght = @array_sporer;
foreach (@array_sporer) {
		$line = $_;
		say "$count of tot:$arraylenght";
		$count ++;
		foreach (@array_vader){
			if (substr ($line, 0, 10) eq substr ($_, 0, 10)){
				say UT $line, "\t", substr($_, 11);
			}
		}
}
```

This is a small script for merging all files in a directory. It was used for creating a single file of several files of weather data.

```
use 5.010;
use warnings;
#Loading a directory into an array, opening an outfile to write to and then adding the content of all the files in the aforementioned directory to the outfile.
@files = <./data/*>;
open UT, ">", 'TeA.txt' or die "Cant open: $!";
foreach (@files){
		open IN, $_ or die "cant open $_: $!";
		while(<IN>){
			print UT $_;
		}
		close IN;
}
```
