Correlation between *C. burnetii* Transmission Rates and Satellite Based Vegetation Indices

February, 2012

Author

Jaïrus Brandsma[1]

Jeroen van Leuken[2,3]

Peter Droogers[1]

Johannes Hunink[1]

Arno Swart[2]

Wim van der Hoek[2]

[1]*FutureWater*, [2]*RIVM*, [3]*UU/IRAS*

Report FutureWater: 109

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

Table of contents

Tables

Figures

1 Introduction

1.1 Background

Coxiella burnetii is a bacterium with a worldwide reservoir in many different animal species and may cause Q fever in humans. In some animals, especially small ruminants (sheep and goats), but also domestic cats, extremely large numbers of *C. burnetii* can be excreted with birth products like placenta and amniotic fluids (van der Hoek et al., 2011). Although direct exposure to parturient animals or their birthing products poses a high risk for infection with *C. burnetii*, the organism's ability to persist in the environment may result in a continued risk for infection weeks to months after the birthing event (Yanase et al., 1998). The potential for transmission is greatly enhanced by the low ID50 for *C. burnetii*. It is reported that a single organism is enough to cause infection (Tigertt et al., 1961). Q fever is a zoonotic disease with no evidence for human-to-human transmission.

Infection with *C. burnetii* is more likely to occur when a number of environmental conditions are met, such as warm weather with dry soil conditions and with wind speed and wind direction that allows airborne contaminated dust particles to be inhaled by people and animals that are at close distance to the contaminated soil. It has been suggested that under dry, dusty conditions infective aerosols can travel several kilometers down wind and large human outbreaks have been linked to wind dispersion from sites where infected animals are kept (Hawker et al., 1998; Tissot-Dupont et al., 1999).

Outbreaks of human Q fever of unprecedented size occurred in the Netherlands in 2007, 2008, and 2009 (Schimmer et al., 2008, 2009). The size of the community outbreak in the Netherlands suggests that transmission takes place through wide-scale environmental contamination or multiple point-source contamination sites, rather than from direct (occupational) contact with animals, consumption of contaminated unpasteurized milk, or from parturient pet animals. Ongoing studies in the south of the Netherlands suggest that infected animals in large-scale dairy goat farms are the major source of environmental contamination. The affected area has a very high density of dairy goat farms and dairy sheep farms and several farms experienced clinical signs with abortion waves that were confirmed as Q fever with immunohistochemistry.

An earlier study by Hunink et al. (2010) shows that environmental factors influence the transmission of the *C. burnetii* bacteria from the infected source to humans. The study indicated that the transmission is correlated with especially the vegetation density (NDVI, Normalized Difference Vegetation Index) and the average groundwater levels.

This study focuses on the vegetation density. We have investigated which vegetation indices are available and which of these has the highest correlation with the transmission rate from an infected source to humans.

1.2   The aerosol route

Aerosol transmission of *C. burnetii* occurs through the inhalation of contaminated particulate matter (PM). The origin of potentially contaminated PM will be either the infected farm itself or manure transported elsewhere. The transport and sedimentation of the contaminated PM is mainly a function of wind velocity and the near-surface turbulence. These factors are related to the surface roughness, vegetation density and land use characteristics.

The mobilization of fine particles requires the knowledge of the surface wind and the threshold velocity of wind erosion. The threshold velocity depends on the particle size and soil moisture. The mass flux of wind-transported dust particles depends on the excess of the wind friction speed over the threshold wind friction velocity for erosion.

Soil moisture increases the threshold friction velocity. Chepil (1956) considered that the influence of soil moisture on wind erosion rates depends on soil texture and can be explained by interparticle cohesion forces due to soil water retention processes. The soil clay content was found to be the main parameter that controls the minimal soil moisture required to induce an increase in the threshold friction velocity (Fécan et al., 1998).

Vegetation has a sheltering effect on erodible land surfaces (e.g. Lancaster and Baas, 1988 and Stockton and Gillette, 1990) and thus reduces dust emissions. Vegetation traps particles and extracts momentum from the air flow depending on the roughness structure of the vegetated area. A high density of roughness elements (e.g. trees) increases the threshold friction velocity and reduces the wind speed. This removes particles from the air flow and reduces dust concentrations (deposition). The degree to which dust emissions are controlled by vegetation cover and geomorphic setting was investigated using dust storm frequency data. Engelstaedter (2003) showed that dust storm frequency is inversely correlated with leaf area index (an index of vegetation density) and net primary productivity.

1.3   Q-fever transmission and environmental factors (previous studies)

A preliminary study was undertaken to explore the environmental factors which could contribute to the transmission of *C. burnetii* (Van der Hoek et al., 2011). This section summarizes the main findings of that particular study.

The distribution of human cases in the Netherlands did not show an even distribution around the farms contaminated with *C. burnetii.* Some highly infected farms did not have any human cases in the surroundings. This raised the question which factors or environmental conditions influence the transmission of *C. burnetii.* The study assessed the influence of the vegetation, land use, soil, and weather conditions, and accounted for other possible non-environmental factors, such as population and animal density.

Results of this preliminary study indicated that during the months of April through June the human infection of *C. burnetii* is highest (Figure 1). Therefore this research focused on the infections within these months only. A five kilometer buffer is taken around each farm, since previous research (Schimmer et al., 2010) showed that people living within

five km from a farm have a much higher risk of infection than people living further away. Farms were considered to be the only potential contamination source from the infection date onwards. Berri et al. (2007) described that *C. burnetii* bacteria and spores among animals are persistent, and will remain to be a potential contamination source for at least one year after the initial emission date.
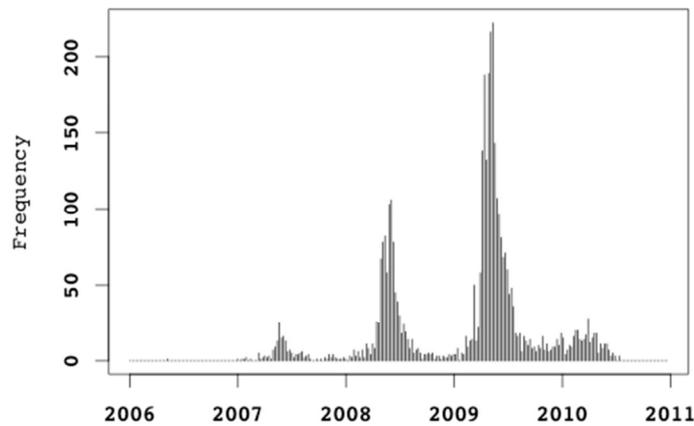
Figure 1: Human Q fever notifications in the Netherlands

The vegetation density was derived and calculated based on 250m resolution satellite images. The density was calculated with the Normalized Difference Vegetation Index (NDVI). The images used originate from the MODIS (Moderate Resolution Imaging Spectroradiometer). The times of production of these images do not coincide with the beginning and end of the months. It was therefore assumed that the image for which the 16-day range fell entirely within a month, represented the entire month.

The land use was assessed based on the Dutch land use database of 2004, on a 25 meter resolution. After extracting residential areas the following relevant classes were derived: (i) arable and cultivated land; (ii) pastures; (iii) open spaces with little or no vegetation (including heath land); and (iv) forest.

The soil map of The Netherlands 1:50,000 was reclassified into relevant classes which supplies information about the *soil texture* at surface level. This surface soil texture determines the wind-erosion threshold velocity.

Soil moisture was modeled, based on the groundwater depth classes. The temporal variation is mainly controlled by the water balance, i.e., evapotranspiration and precipitation which are included in the model.

For the *weather conditions* the wind velocity and temperature have been taken into account. Wind velocity is a contributing factor to erosion, and as such plays a role in spreading of *C. burnetii.* Temperature may influence the persistence of *C. burnetii* in the environment. These data have been obtained from the KNMI, and have been spatially interpolated to fit the buffer zones around each farm.

These five factors have been analyzed for each farm. Each human notification of Q fever within each 5 km buffer zone was assigned to the farm. When zones overlapped the case was assigned to the nearest farm. Buffer areas were considered contaminated when there were >1 notification per 10,000 population per month, with a minimum of two notifications. The contaminated buffer zones and the areas without transmission were

statistically compared concerning the five factors for the months April – June in 2008 and 2009 (see Table 1).

Table 1: Means and *p*-value (Student's *t*-test) for environmental variables for the complete dataset (n = 150 farm-months of observation). (van der Hoek et al., 2011)

	Transmission		
	No ($n = 95$)	Yes ($n = 55$)	P
NDVI*			
Median	0.70	0.67	<0.001
Minimum	0.27	0.24	0.14
Maximum	0.89	0.89	0.26
Land use (km^2)			
Arable lands	22.5	27.4	<0.001
Pasture lands	33.2	29.2	0.01
Forest cover	7.8	8.0	0.77
Bare cover and heath	0.3	0.2	0.10
Residential area	7.1	8.6	0.02
Soil texture (%)			
Peat covered	7.3	0.8	<0.001
Clay covered	24.2	26.4	0.67
Sand covered	59.5	62.6	0.58
Soil moisture			
Min groundwater table class	2.11	2.36	<0.001
Relative soil humidity (%)	0.62	0.66	0.19
Climate			
Mean wind velocity (0.1 m/s)	34.8	32.9	<0.001
Temperature (°C)	13.7	13.7	0.99
Global radiation (J/cm^2)	1,797	1,765	0.39

*NDVI, Normalized Difference Vegetation Index.

The results of the study, Table 1, show that contaminated farms have a lower median NDVI, a significantly lower groundwater table and when the soil is peat covered the probability of transmission is much lower. Contra intuitive, the wind velocity was found slightly lower in areas with transmission. The fact that wind-direction was not considered in the study could probably explain this contradiction.

The role of soil moisture and vegetation to prevent erosion and to remove dust particles from the air flow is also supported by the environmental science. The factors which highly correlate with the transmission do often coincide with each other. For example peat covered soil with a shallow ground water table. Or low mean NDVI with arable land and low soil moisture.

The study recommends investigating further on the spatial variation of vegetation density and soil moisture. This will result in a better insight on the transmission of *C. burnetii,* which could help decision makers, or to create a risk warning system.

1.4   Objectives

The overall objective for this study is to explore the correlation between 19 satellites based vegetation indices and *C. burnetii* transmission from contaminated farms to humans. The focus will be on selecting the most promising VI, which has the highest correlation with the source factor. Traditionally the NDVI (Normalized Difference Vegetation Index) is most used; this study however will involve 18 alternative VI´s. The indices will be calculated based on the MODIS satellite images, and linked on an annual basis to the data on the Q fever outbreaks in the period 2006 until 2009 in the Netherlands. In order to assess whether the distance to a contaminated farm is of any importance a spatial assessment was carried out as well.

The results of this study will contribute to a better understanding of the transmission of Q-fever and other air-borne zoonotic diseases. This is only one factor affecting the transmission of C. burnetii, and should therefore be combined with other factors to locate the vulnerable areas within the Netherlands.

2 Vegetation indices

2.1 Background

Vegetation is part of the friction surface which determines the wind velocities and erosion and deposition of particles near the ground. Forests and other vegetated areas are characteristically rough surfaces and thus have a relatively large contribution to the surface air turbulence. Especially leaf canopies are very effective in slowing down wind because of their large friction area and they enhance the deposition of wind-transported dust particles.

The size of the wind speed reduction by a vegetation patch depends on its internal structure, (e.g. density and height) and the local wind speed profile. The density of vegetation is strongly related to the total leaf area, a variable that differs strongly during the year. It is likely that the effect of vegetation on the transport of contaminated dust particles is highly correlated to the time of the year as well.

2.1.1 Remote sensing

The term "remote sensing" refers to obtaining and interpreting information from a distance, using sensors that are not in physical contact with the object being observed. The science of remote sensing in its broadest sense includes aerial, satellite, and spacecraft observations of the surfaces and atmospheres of the earth or even to other planets and stars. The term remote sensing is customarily restricted to methods that detect and measure electromagnetic energy, including visible light, that has interacted with surface materials and the atmosphere. Remote sensing of the Earth has many purposes, including making and updating maps, weather forecasting, and gathering military intelligence. In the even more strict use of the term remote sensing in environmental applications and studies is related to Earth's surface.

Remote sensors measure electromagnetic (EM) radiation that has interacted with the Earth's surface. Interactions with matter can change the direction, intensity, wavelength content, and polarization of EM radiation. The nature of these changes is dependent on the chemical composition and physical structure of the material. Changes in EM radiation resulting from its interactions with the Earth's surface therefore provide major clues to the characteristics of the surface materials. The fundamental interactions between EM radiation and matter are (Figure 2):

- Transmission: Electro- magnetic radiation that is transmitted passes through a material (or through the boundary between two materials) with little change in intensity.
- Absorption: Materials can also absorb EM radiation. Usually absorption is wavelength-specific: that is, more energy is absorbed at some wavelengths than at others. EM radiation that is absorbed is transformed into heat energy, which raises the material's temperature.
- Emission: Some of that heat energy may then be emitted as EM radiation at a wavelength dependent on the material's temperature. The lower the temperature, the longer the wavelength of the emitted radiation. As a result of

solar heating, the Earth's surface emits energy in the form of longer-wavelength infrared radiation. For this reason the portion of the infrared spectrum with wavelengths greater than 3 µm is commonly called the thermal infrared region.

Electromagnetic radiation encountering a boundary such as the Earth's surface can also be reflected:

- Specular: If the surface is smooth at a scale comparable to the wavelength of the incident energy, specular reflection occurs: most of the energy is reflected in a single direction, at an angle equal to the angle of incidence.
- Scattering: Rougher surfaces cause scattering, or diffuse reflection in all directions.

All remote sensing systems designed to monitor the Earth's surface rely on energy that is either diffusely reflected by or emitted from surface features. Current remote sensing systems fall into three categories on the basis of the source of the electromagnetic radiation and the relevant interactions of that energy with the surface: (i) reflection, (ii) thermal infrared, and (iii) radar.

Reflected solar radiation sensors. These sensor systems detect solar radiation that has been diffusely reflected (scattered) upward from surface features. The wavelength ranges that provide useful information include the ultraviolet, visible, near infrared and middle infrared ranges. Reflected solar sensing systems discriminate materials that have differing patterns of wavelength-specific absorption, which relate to the chemical make-up and physical structure of the material. Because they depend on sunlight as a source, these systems can only provide useful images during daylight hours, and changing atmospheric conditions and changes in illumination with time of day and season can pose interpretive problems. Reflected solar remote sensing systems are the most common type used to monitor Earth resources.

Thermal infrared sensors. Sensors that can detect the thermal infrared radiation emitted by surface features can reveal information about the thermal properties of these materials. Like reflected solar sensors, these are passive systems that rely on solar radiation as the ultimate energy source. Because the temperature of surface features changes during the day, thermal infrared sensing systems are sensitive to time of day at which the images are acquired.

Imaging radar sensors. Rather than relying on a natural source, these "active" systems "illuminate" the surface with broadcast micro- wave radiation, then measure the energy that is diffusely reflected back to the sensor. The returning energy provides information about the surface roughness and water content of surface materials and the shape of the land surface. Long-wavelength microwaves suffer little scattering in the atmosphere, even penetrating thick cloud cover. Imaging radar is therefore particularly useful in cloud-prone tropical regions.

Another important distinction in sensors is the resolution: spatial, spectral and temporal. **Spatial resolution** is a measure of the spatial detail in an image, which is a function of the design of the sensor and its operating altitude above the surface. Each of the detectors in a remote sensor measures energy received from a finite patch of the ground surface. The smaller these individual patches are, the more detailed will be the spatial

information that we can interpret from the image. For digital images, spatial resolution is most commonly expressed as the ground dimensions of an image cell (=pixel).

The **spectral resolution** of remote sensing system can be described as its ability to distinguish different parts of the range of measured wavelengths. In essence, this amounts to the number of wavelength intervals ("bands") that are measured, and how narrow each interval is. An "image" produced by a sensor system can consist of one very broad wavelength band, a few broad bands, or many narrow wavelength bands. The names usually used for these three image categories are panchromatic, multispectral, and hyperspectral, respectively.

Finally the **temporal resolution** of a sensor is an important characteristic .The surface environment of the Earth is dynamic, with change occurring on time scales ranging from seconds to decades or longer. The seasonal cycle of plant growth that affects both natural ecosystems and crops is an important example. Repeat imagery of the same area through the growing season adds to our ability to recognize and distinguish plant or crop types. A time-series of images can also be used to monitor changes in surface features due to other natural processes or human activity. The time-interval separating successive images in such a series can be considered to define the temporal resolution of the image sequence.
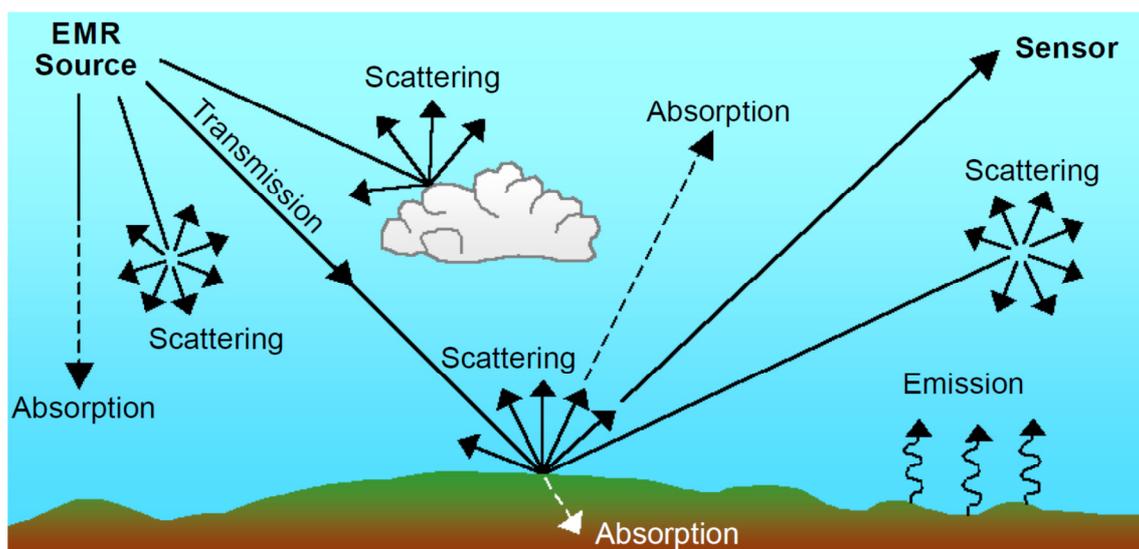
Figure 2. Typical electromagnetic (EM) radiation interactions in the atmosphere and at the Earth's surface.

2.1.2   Vegetation indices

Vegetation Indices (VI's) are combinations of surface reflectances at two or more wavelengths designed to highlight a particular property of vegetation. Each of the VI's is designed to accentuate a particular vegetation property. Analyzing vegetation using remotely sensed data requires knowledge of the structure and function of vegetation and its reflectance properties. This knowledge enables the linking of vegetative structures and their condition to their reflectance behavior in an ecological system of interest.

The solar-reflected optical spectrum spans a wavelength range of 400 nm to 3000 nm. Out of this range, the 400 nm to 2500 nm region is routinely measured using a variety of

optical sensors. Vegetation interacts with solar radiation differently from other natural materials, such as soils and water bodies. The absorption and reflection of solar radiation varies depending on the characteristics of the vegetation and the incident wavelength. Water, pigments, nitrogen, and carbon are each expressed in the reflected optical spectrum from 400 nm to 2500 nm, with often overlapping, but spectrally distinct, reflectance behaviors (See Figure 3).

Live green vegetation absorbs solar radiation, which is used in the process of photosynthesis. The ratio of absorption of red light wavelengths and the reflection of NIR, blue and green light wave lengths is an indicator to assess greenness, the relative density of vegetation or the health of vegetation (See Figure 4).
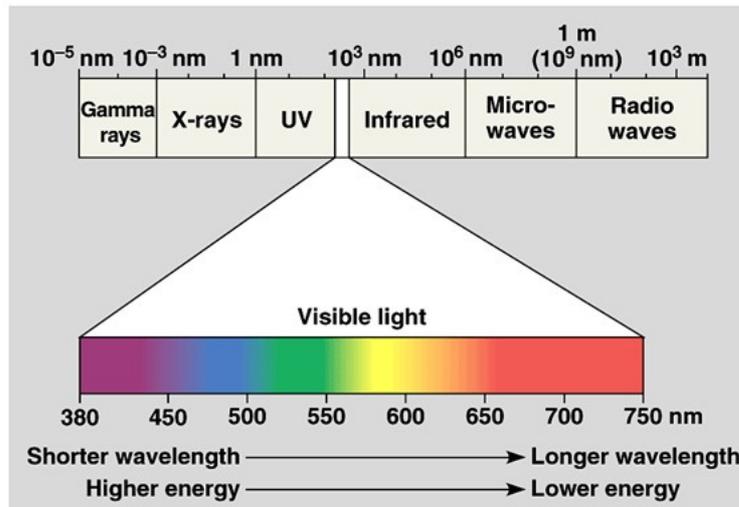
Figure 3: Light wavelengths
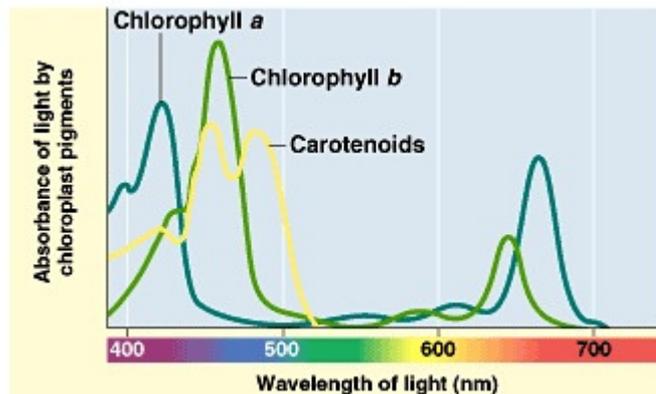
Figure 4: Absorption spectra

The different VI's all emphasize a different property of the vegetation, and some make an additional correction for the soil. The soil reflectance of light may disturb an accurate measurement, especially in sparse vegetated areas. The soil adjusted vegetation Indices, such as SAVI, MSAVI2 and OSAVI, tend to minimize the soil brightness (Panda et al., 2010).

2.2   Data sets

To assess the vegetation indices for the 5 kilometers buffers around the contaminated farms MODIS (Moderate Resolution Imaging Spectroradiometer) images are used.

The MODIS instruments view the planet at a distance of 705 km above the earth's surface. The MODIS instruments are on board of the Terra and Aqua satellite, which were launched by NASA in 1999 and 2002 respectively. Together these satellites image the entire globe every 1-2 days.
MODIS gathers information by measuring and recording the light that is reflected by the various earth surfaces. Because the different surfaces reflect light in different patterns, detailed information is made available. The MODIS surface reflectance sensors measure the reflectance of the light in 36 bands. To assess the vegetation indices the following bands are used:

MYD09A1: 3(*459–479 nm*), 4(*545-565 nm*), 5(*1230–1250 nm*) on a 500 m resolution
MYD09Q1: 1(*620–670 nm*), 2(*841-876 nm*) on a 250 m resolution.

These bands are available on an 8 day interval, and show the clearest reflection values for each raster cell out of these 8 days. Taking this interval increases the quality; however prolonged cloudiness may reduce the quality or even make it impossible to measure, a specific area, within this 8-days period. Bands 3, 4 and 5 were resampled to the same resolution of bands 1 and 2 to obtain Vegetation Indices at spatial resolutions of 250 meter.

The reflected light waves that satellite sensors detect coming from vegetation on earth, can be affected by gasses, aerosols or thin clouds in the atmosphere as well as the angle at which the satellite views the ground. These 'noise' factors are corrected by the NASA (National Aeronautics and Space Administration), so that the values given in the bands 1-7 only show the values as it would be on surface level (Vermote and Vermeulen, 1999). For this analysis the remote sensing data is used from the Aqua satellite, as the orbital tracks are most suitable. The Aqua satellite passes directly over the Netherlands around mid-day every second day (See Figure 5).
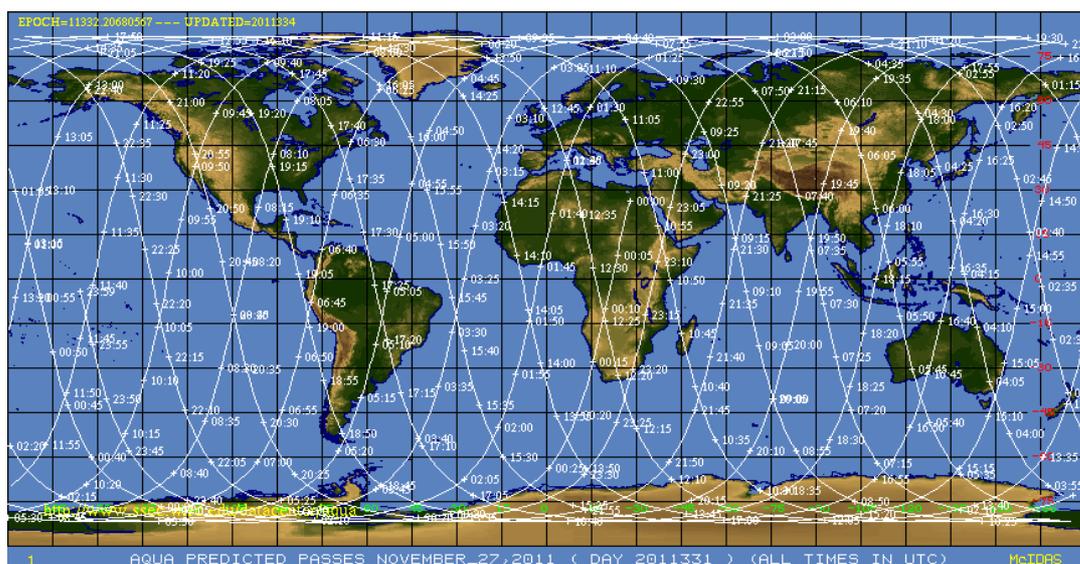

Figure 5: Sample Aqua Orbital Tracks 27-11-2011

2.3   Available indices

For this research 19 indices where used. These indices include the most usual and widespread indices, such as NDVI, EVI, SAVA, etc. These selected indices represent a wide range, among which, the simple ratio, the normalized difference vegetation index, soil adjusted vegetation index, atmospherically resistant vegetation index, modified SAVI, enhances vegetation index, optimized SAVI, and green NDVI. For the complete overview of the assessed indices and formulas see Table 3. A detailed description of these indicators is provided in the work of Solaimani et al. (2011).

The wavelengths in the formulas are one specific number, but since the MODIS reflectance images are measured in wavelength ranges the following ranges are used to approach this exact numbers and to calculate the VI's.

Table 2: Wavelength ranges per color.

Color	Band number	Wavelength range	Wavelength In formula
RED	1	620–670 nm	670
NIR	2	841-876 nm	800
BLUE	3	459–479 nm	450
GREEN	4	545-565 nm	550
Infrared	5	1230–1250 nm	1241

Table 3: Assessed indices

Abbreviation	Name	Formula
NDVI	Normalized Difference vegetation Index	$(\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670})$
RDVI	Renormalized Difference Vegetation Index	$(\rho_{800} - \rho_{670}) / \sqrt{(\rho_{800} \times \rho_{670})}$
MSR	Modified Simple Ratio	$(\rho_{800} / \rho_{670} - 1) / \sqrt{(\rho_{800} / \rho_{670} + 1)}$
SAVI	Soil-Adjusted Vegetation Index	$(1 + L)(\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670} + L)$, L=0.5
MSAVI2	Modified Soil Adjusted Vegetation Index	$(0.5)(2(\rho_{800} + 1) - \sqrt{((2 \times \rho_{800} + 1)^2 - 8(\rho_{800} - \rho_{670}))})$
TVI	Transformed Vegetation Index	$(NDVI + 0.5)^{0.5}$
MTVI1	Modified Triangular Vegetation Index	$1.2 (1.2 (\rho_{800} - \rho_{550}) - 2.5 (\rho_{670} - \rho_{550}))$
WDRVI	Wide Dynamic Range Vegetation Index	$(a \rho_{800} - \rho_{670}) / (a \rho_{800} + \rho_{670})$, a = 0.2
VARI	Visible Atmospherically Resistant Index	$(\rho_{550} - \rho_{670}) / (\rho_{550} + \rho_{670} - \rho_{450})$
NDWI	Normalized Difference Water Index	$(\rho_{800} - \rho_{1241}) / (\rho_{800} + \rho_{1241})$
IPVI	Infrared Percentage Vegetation Index	$\rho_{800} / (\rho_{800} + \rho_{670})$
EVI	Enhances Vegetation Index	$2.5 (\rho_{800} - \rho_{670}) / (1 + \rho_{800} + C1 \rho_{670} + C2 \rho_{450})$, C1 = 6, C2 = 7.5.
DVI	Difference Vegetation Index	$\rho_{800} - \rho_{670}$
RVI	Ratio Vegetation Index	ρ_{800} / ρ_{670}
GNDVI	Green Normalized Difference Vegetation Index	$(\rho_{800} - \rho_{550}) / (\rho_{800} + \rho_{550})$
MND	Modified Normalized Difference	$\rho_{800} - (1.2 \rho_{670}) / (\rho_{800} + \rho_{670})$
OSAVI	Optimized Soil-Adjusted Vegetation Index	$(1 + 0.16)(\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670} + 0.16)$

GI	Greenness Index	ρ_{550} / ρ_{677}
NLI	Non Linear Index	$(\rho_{800}^2 - \rho_{670}) / (\rho_{800}^2 + \rho_{670})$

3.1   Study area

The total number of registered small ruminant locations or farms in the Netherlands is approximately 52,000, of which 350 are professional dairy goat farms with more than 200 adult goats and 40 are professional dairy sheep farms. In principle the animals on these farms remain in the deep litter stable year-round and all practices are carried out indoors. Dairy goat farmers typically do not have land and manure is transported elsewhere.

Following the 2007 outbreak, an informal agreement was made that the veterinary and the public health sectors would exchange information on farms with newly diagnosed animal cases of Q fever to allow for an adequate response and control. In 2008, notification of *C. burnetii* abortions in ruminants became mandatory. In addition, a ban was imposed to spread manure during the three months following the detection of Q fever at the farm. In October 2009 a mandatory monitoring system of bulk tank milk (BTM) at dairy goat and dairy sheep farms with >50 animals was implemented. If DNA of the bacteria is detected in BTM, the farm is declared 'infected' but the environmental contamination from such farms is generally much lower than from farms where clinical signs of Q fever (abortion waves) have occurred.

For this research the farms in the Netherlands which have been infected from 2005 till 2010 are reviewed (See Figure 6). For the Vegetation indices assessment however, only the farms on which an abortion wave took place are considered as a potential contamination source. Farms which were found to be tank milk positive are not taken into account as a contamination source.
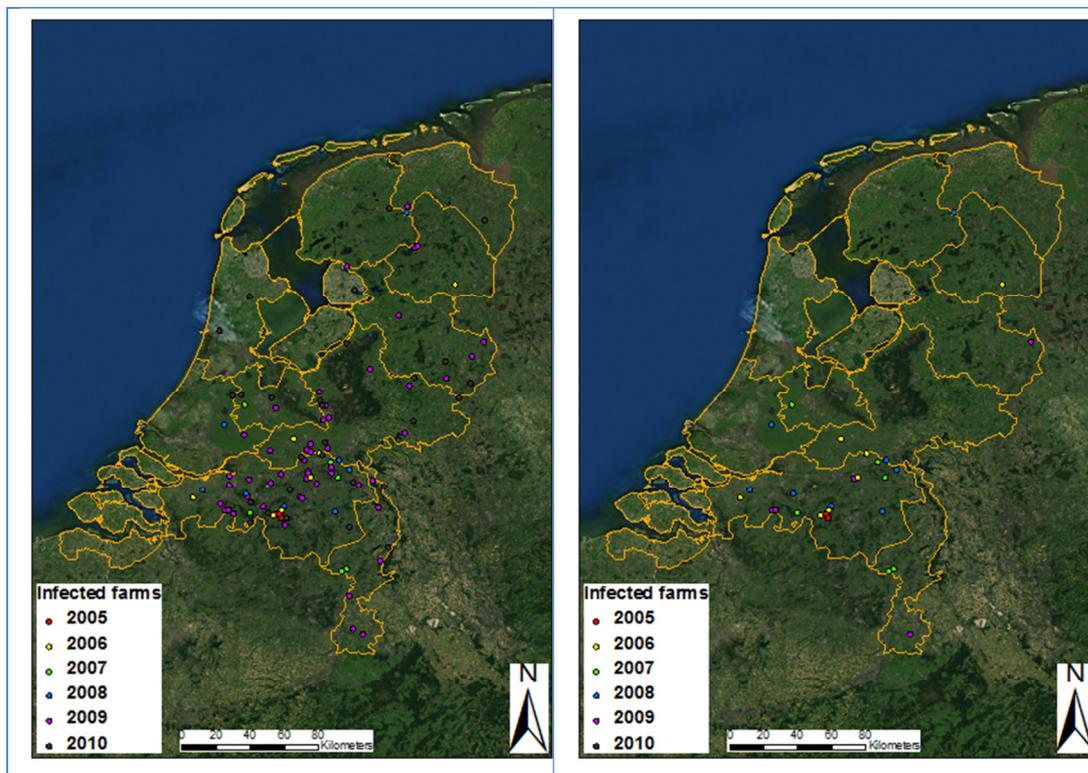

**Figure 6: Total infected farms (left) (n=106) and farms with abortion waves (right) (n=29)**

3.2 Data

The municipal health services provided the full postal code of the home address and date of onset of illness of all patients that had been notified with acute Q fever as separate anonymous files. Information on farms affected by clinical Q fever (abortion waves) was provided by the Animal Health Service. In October 2009 a system of mandatory bulk tank milk monitoring started. The addresses of bulk tank milk positive farms were publicly available on the website of the Food and Consumer Product Safety Authority. Locations of all farms with sheep and goats in the Netherlands were available from the Dutch Ministry of Economic Affairs, Agriculture and Innovation.

3.3 Spatial distribution

Based on ongoing studies (around Helmond) it was demonstrated that the risk for Q fever infection shows a monotonous decline with distance, with a significantly decreased risk for people living more than 5 kilometers from the infection source. For this reason, the analysis is done using 5 km buffers around the infected farms as the radius of exposure. To give a better view on the spatial distribution of the vegetation within these five km buffers a weighted analysis is carried out as well. This weighted average will give more importance to the vegetation near to the farm (Figure 7).
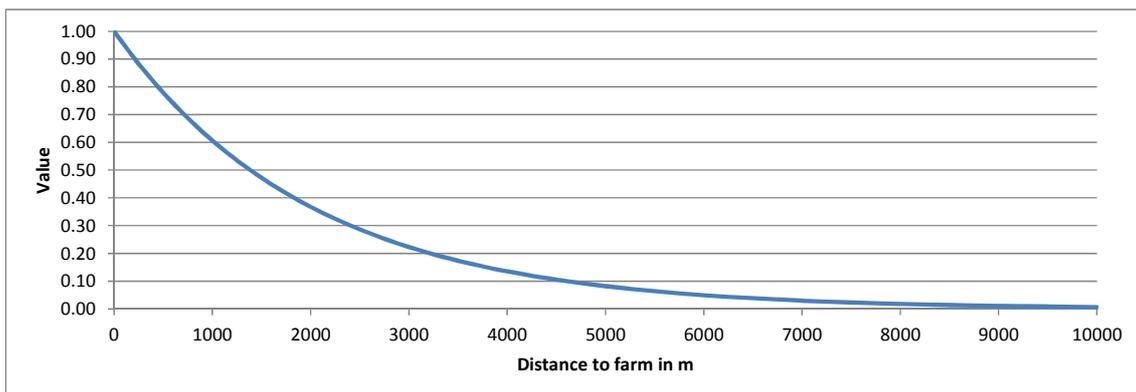
Figure 7: Weighted values

The spatial distribution of human cases in 2008 and 2009 shows outbreaks around farms where
Q fever was detected during previous years. Besides, the Q fever bacteria are known to persist for a long time in the environment. For this reason, it was assumed that a farm remained infected during the years following the first detection of Q fever. The information available confirms that at a few locations Q fever was detected twice, 1 to 3 years after the first detection.

3.4 Analysis

3.4.1   Vegetation indices

The remote sensing images are the basis for the VI analysis. The analysis is carried out for the years 2006 until 2010 and for each year the months March through June were assessed. For each 8-day interval image the 19 different indices were calculated based on the equations shown in Table 3. Finally a yearly average is created over the months March through June.

Around each contaminated farm a buffer with a 5 kilometer radius was considered. For the area within the buffer zone and for each year, the March to June mean value for each of the vegetation indices is calculated (Table 4).

Table 4: Typical example of mean NDVI values per year (March to June).

Farm	2006	2007	2008	2009	2010
A	0.53	0.57	0.54	0.57	0.51
B	0.54	0.61	0.56	0.61	0.55
C	0.49	0.59	0.54	0.57	0.46
D	0.54	0.67	0.56	0.61	0.51
E	0.53	0.58	0.54	0.59	0.51
F	0.54	0.57	0.58	0.61	0.55
G	0.51	0.54	0.52	0.55	0.49

A similar analysis is carried out to obtain a weighted-distance average, which gives more importance to vegetation immediately surrounding the farm. For this analysis a weight raster is created for each buffer zone surrounding farm, based on the following formula, which is plotted in Figure 7.

$$Value = e^{(-[distance\ raster]\ *\ 0.0005)}$$

This weighted raster is multiplied with the annually averaged indices raster. From this multiplied raster the spatial average is calculated, which results in a table as shown in Table 4. The example of the relation between the weighted-distance average and the unweighted average is shown in Figure 8. The three outliers in the Figure are farms where the 5 km zone extents to Belgium for which insufficient data were available.
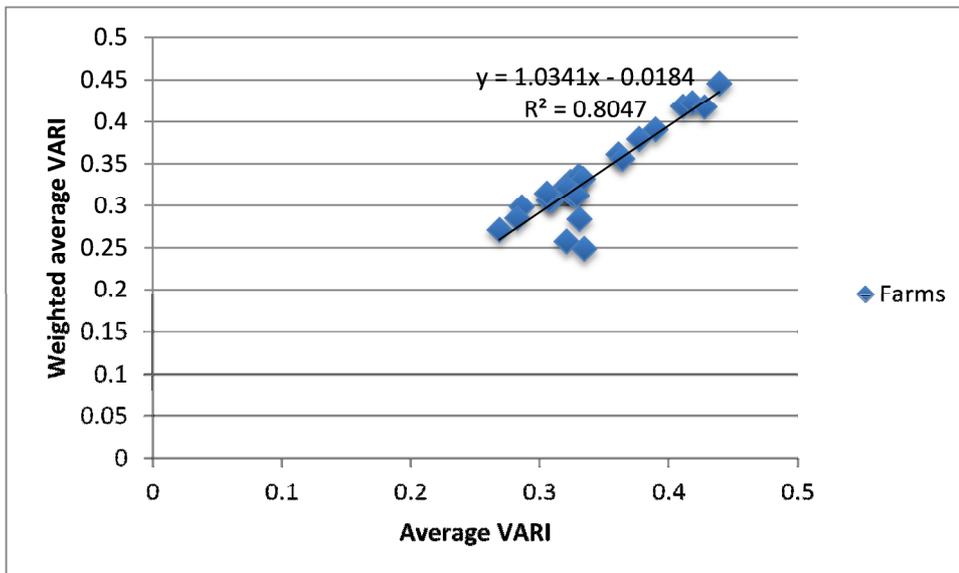
Figure 8: Example relation average VARI and weighted average VARI.

3.4.2 Transmission factor

Van Leuken et al. (2012) describe the method of calculating so-called transmission factors in order to determine the effect of an environmental factor (e.g., the mean vegetation index) on the incidence of Q fever as a function of the infected farms. These factors are described by three parameters:

1) The emission strength (by means of the number of goats and sheep, d);
2) The distance between a source and a patient (r);
3) The population density (y) of the 4 digit zip code area (PC4) of each patient.
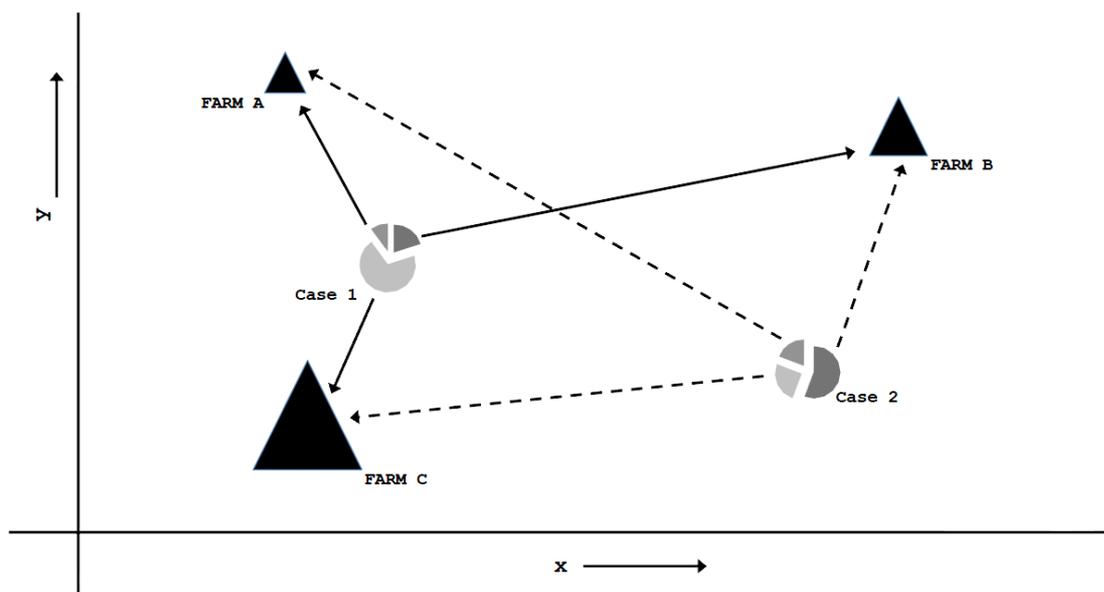
Figure 9: Example of the calculation of the transmission factors.

Patient 1 is divided over all farms on the basis of each farm's number of animals only. Thus, the farm with the largest number of animals receives the largest part of patient 1. Patient 2 is divided over all farms on the basis of the distance to each farm only, e.g., the largest fraction is assigned to farm B, which is closest to patient 2. (Van Leuken et al., 2012)

Since it is not possible to determine by which source a patient was infected, each patient was divided over all sources. Thus, each source has some potential contribution to the infection of a patient. This contribution depends on the three parameters. For example, a farm with many animals at a small distance to a certain patient will have a larger share of that patient than a farm with a few animals at a large distance (Figure 9)

Each source j receives a fraction of a patient i that is equal to:

$$\phi_{i,j} = \frac{r_{i,j}^{-1} \cdot d_j}{\sum_{k=1}^{K}[r_{i,k}^{-1} \cdot d_k]}$$

[1]

The denominator in [1] is the sum of the number of animals of all farms multiplied by the reciprocal distance between patients i and all farms. For example, if patient 1 could be infected by farms A, B, or C, each with 100, 200 and 700 animals respectively and located at 250, 600, and 300 m respectively from patient 1, then [1] becomes for the relation of farm A to patient 1:

$$\phi_{1,A} = \frac{(r_{1,A})^{-1} d_A}{\sum_{k=1}^{K}[(r_{1,k})^{-1} d_k]} = \frac{250^{-1} \cdot 100}{250^{-1} \cdot 100 + 600^{-1} \cdot 200 + 300^{-1} \cdot 700}$$

All values are part of the matrix Φ in $\mathbf{R}^2$ with I patients and K sources. Summing all fractions $\phi_{1,A}$ results in 1 by definition. One obtains the transmission factors τ by multiplication of each fraction ϕ with the population density y [km^{-2}] of each patient's PC4, and by subsequently summing over all patients:

$$\tau_j = \sum_{i=1}^{I}[y_i^{-1} \cdot \phi_{i,j}]$$

[2]

The annual averaged VI values are correlated with the transmission factors. The VI which shows the highest correlation with the transmission factor is advised to use in further research.

4.1 Results C. burnetii transmission

The transmission factors are calculated for the farms that experienced abortion waves during the period 2006 – 2009. A clear progressive trend can be observed over the years in (i) number of contaminated farms, (ii) infected people (Figure 1), and (iii) higher transmission factors. (Table 5)

**Table 5: Transmission factors per year 2006-2009**

ID	#ANIMALS	TRANSMISSION2006	TRANSMISSION2007	TRANSMISSION2008	TRANSMISSION2009
5	1519	NA	NA	0.028	0.031
19	1349	NA	NA	NA	0.020
24	797	0.000	0.001	0.002	0.005
25	1196	NA	0.002	0.006	0.015
27	2228	NA	0.018	0.038	0.055
29	604	0.000	0.002	0.010	0.016
34	1497	NA	NA	0.017	0.026
40	756	NA	NA	0.007	0.016
43	2194	NA	NA	0.013	0.031
47	1213	NA	NA	NA	0.030
52	573	NA	NA	NA	0.005
53	860	NA	NA	NA	0.007
54	933	NA	NA	NA	0.004
62	343	NA	NA	0.001	0.005
73	705	NA	NA	0.002	0.006
95	1728	0.000	0.002	0.009	0.021
96	730	0.000	0.000	0.001	0.003
97	800	0.000	0.001	0.004	0.011
98	805	0.000	0.001	0.005	0.011
99	1700	0.000	0.005	0.027	0.045
100	902	0.000	0.001	0.005	0.011
101	830	0.000	0.001	0.007	0.014
102	560	NA	0.000	0.002	0.005
103	4146	NA	0.006	0.018	0.042
104	2679	NA	0.002	0.009	0.022
105	1484	NA	0.003	0.019	0.034
106	1354	NA	0.001	0.004	0.011
107	986	NA	NA	0.005	0.011
108	185	NA	NA	0.001	0.001
Average		0.000	0.003	0.010	0.018

4.2 Correlation between Vegetation Indices

A rigorous discussion on the differences between the various VI's is beyond the scope of this report. However, this section will provide some typical examples of the various VI's and their spatial and temporal distribution and correlations.

VI's show a large variation over the year. Within the study period from March through June the vegetation density increases significantly. A typical example is shown in Figure 10 indicating spring and summer vegetation, here expressed as the Normalized Difference Vegetation Index (NDVI).
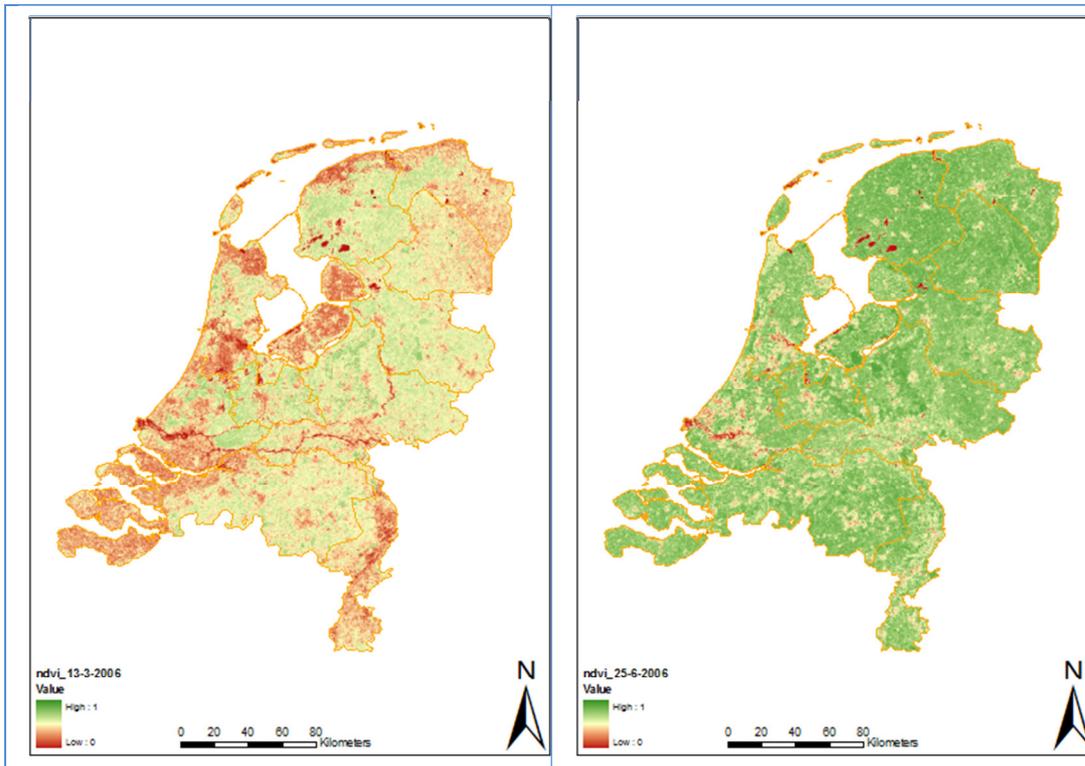
Figure 10: NDVI values for 13-3-2006 (left) and 25-6-2006 (right) for the Netherlands.

Correlation between the various VI's exists especially since some of the VI's are based on exactly the same satellite bands. A typical example is shown in Figure 11. Although absolute values differ substantially because of the nature of the equation, correlation is quite high.
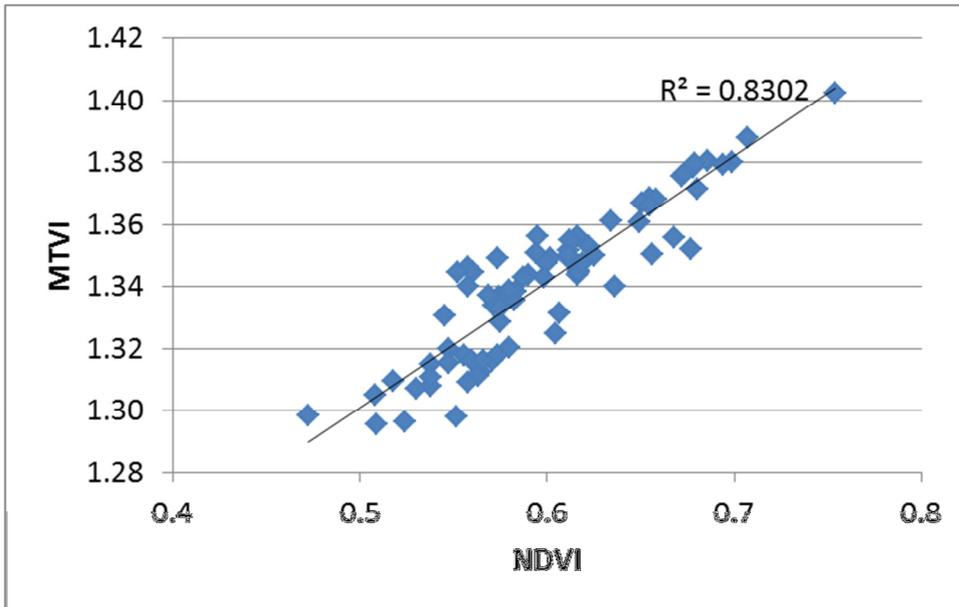
Figure 11: Typical example of correlation between two different Vegetation Indices. Plotted are 78 values of contaminated farms in years (2006-2009).
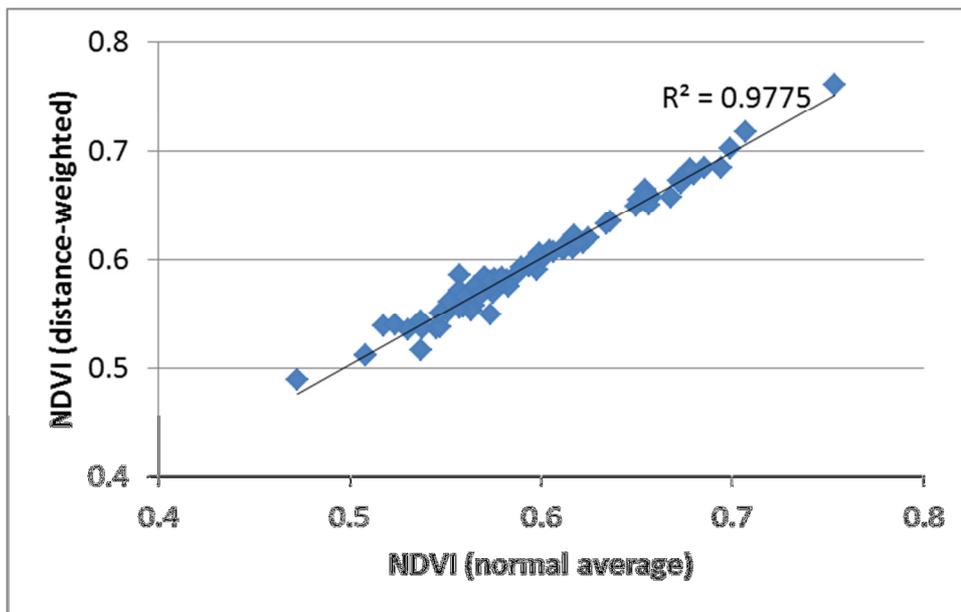

Figure 12: Correlation between distance-weighted and equal-weighted MTVI1. Plotted are 78 values of contaminated farms in years (2006-2009).

4.3 Correlation between Vegetation Indices and transmission factors

The correlation (r^2) between transmission factors and VI's is presented in Table 6. It is clear that the correlation is relatively low. Highest correlation can be found for the simplest VI, the Difference Vegetation Index (DVI). Difference between weighted distances and equal-weight is minor. The value of the slope of the regression line is, as expected, negative. This indicates that more vegetation correlates to lower transmission rates, according to what was expected. Only for one value this correlation was reverse (MSR). The weighted-distance slope was in almost all cases steeper than the non-weighted distance one, indicating that the weighted-distance approach might be a more favorable approach.

The administration of contaminated farms and infected patients only started properly in 2008. Therefore the correlation between the 19 Vegetation Indices and the equal-distance and weighted-distance averaging, for the most reliable data (2008 and 2009) are shown in Table 7. This stratification has significantly increased the correlation values (See Table 6). Moreover, slopes of the regression lines are much steeper as well indicating a stronger relationship between transmission and vegetation.

Somewhat surprisingly, the performance of the various VI's to explain transmission factors, expressed as r^2, is not consistent between 2008 and 2009. For some years one VI performs better, while for another year/period the same VI does not perform very well. Important is to realize that the number of cases included in the statistical analysis is relatively low.

Table 6: Correlation, expressed as r^2, between VI's and transmission factors for all infected farms and all years.

2006-2009	Normal r2	Weighted average r2
RDVI	0.026	0.025
NDVI	0.022	0.036
DVI	**0.046**	**0.053**
EVI	0.024	0.031
GI	0.027	0.036
IPVI	0.022	0.036
MND	0.040	0.051
MSR	0.001	0.000
NDWI	0.028	0.030
NLI	0.033	0.044
RVI	0.030	0.038
SAVI	0.036	0.046
TVI	0.022	0.036
VARI	0.007	0.007
MSAV	0.042	0.050
MTVI	0.012	0.020
OSAV	0.029	0.041
GNDV	0.018	0.028
WDRV	0.024	0.037

**Table 7: Correlation, expressed as r², between Vegetation Indices and transmission factors.**

Indici 2008	Normal r2	Weighted average r2	Indici 2009	Normal r2	Weighted average r2
RDVI	0.090	0.084	RDVI	0.172	0.119
NDVI	0.070	0.107	NDVI	0.182	0.218
DVI	0.059	0.069	DVI	0.181	0.198
EVI	0.062	0.080	EVI	0.190	0.201
GI	0.098	0.095	GI	0.076	0.088
IPVI	0.070	0.107	IPVI	0.183	0.218
MND	0.060	0.078	MND	0.154	0.175
MSR	0.014	0.028	MSR	0.004	0.002
NDWI	0.078	0.085	NDWI	0.094	0.098
NLI	0.064	0.087	NLI	0.160	0.184
RVI	0.103	0.123	RVI	0.167	0.183
SAVI	0.060	0.079	SAVI	0.190	0.212
TVI	0.070	0.105	TVI	0.182	0.217
VARI	**0.151**	**0.156**	VARI	0.051	0.051
MSAV	0.059	0.072	MSAV	0.186	0.204
MTVI	0.075	0.097	MTVI	**0.287**	**0.273**
OSAV	0.062	0.088	OSAV	0.191	0.218
GNDV	0.066	0.093	GNDV	0.211	0.214
WDRV	0.083	0.117	WDRV	0.169	0.201

Some typical VI's scatter plots showing the correlation between the VI and the transmission factor are shown in Figure 13 for the year 2009. These plots indicate that the correlation between transmission and VI is not a clear linear one. The overall pattern is that at high values of VI transmission factors are very low, and at low or intermediate VI level no correlation with transmission exists. In a more practical way: it is clear that dense vegetation will reduce the risk of transmission of *C. burnetii* substantially. However at low or intermediate vegetation densities, other factors determine the transmission of Q fever. Interesting is that farms having high values of VI and low transmission rates (the five dots rights in the plots in Figure 13) are all located outside the main Q fever areas in the Provinces of Brabant and Limburg (Figure 14).

If we expand the analysis to all years (2006-2009: 78 cases) the same pattern occurs (Figure 15). If NDVI is below about 0.67, correlation between transmission factor and NDVI is low. However at NDVI values higher than 0.67 transmission factor is always very low.

A detailed exploration for all 19 vegetation indices revealed that all of them show this threshold based correlation. The only exception is the Normalized Difference Water Index (NDWI) as shown bottom-right in (Figure 15). Interesting is that this VI is the only one that uses the infrared band (band 5) information of the MODIS sensor.
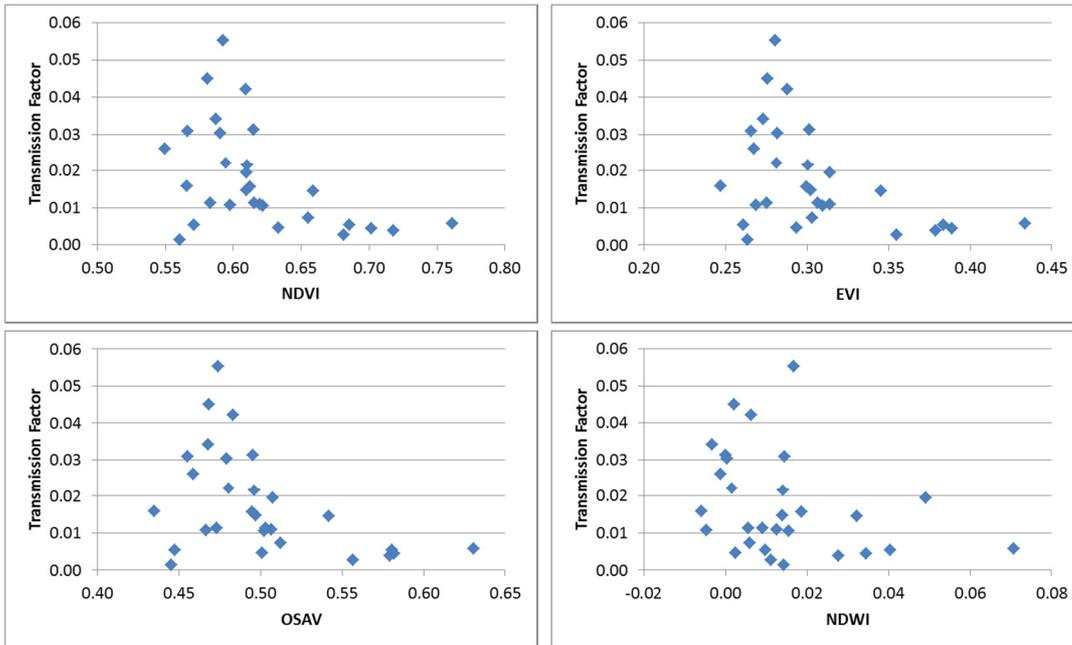
Figure 13: Correlation between transmission factor and vegetation index for four typical indices (year 2009).
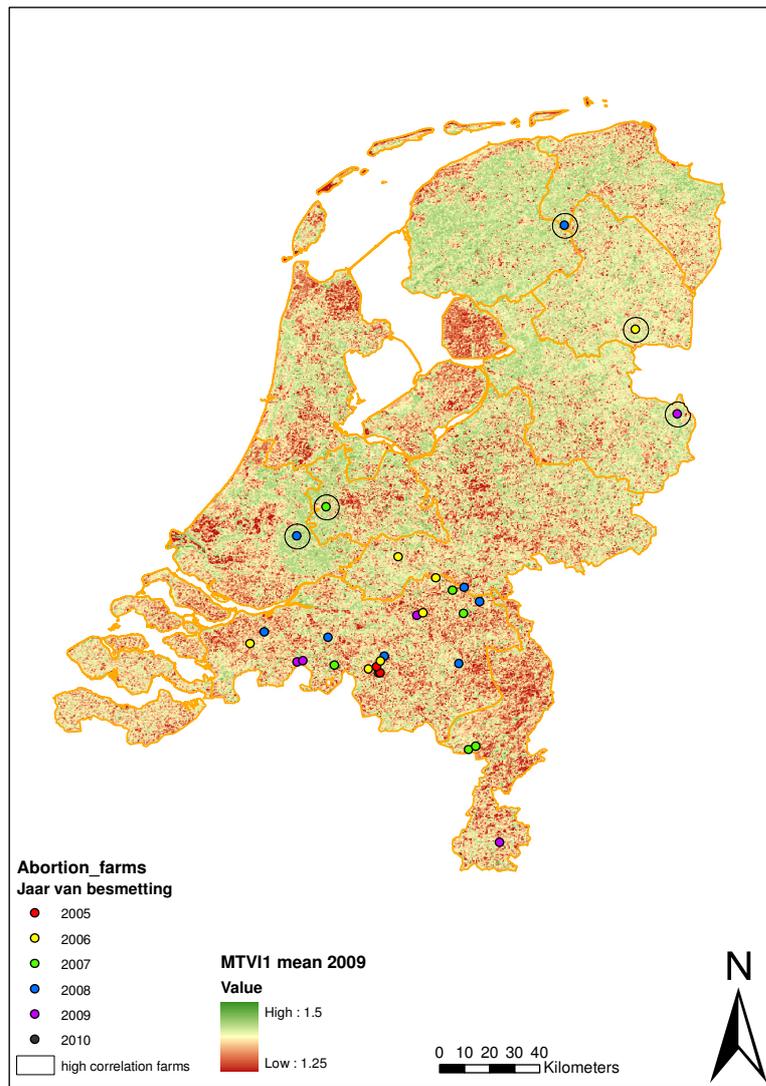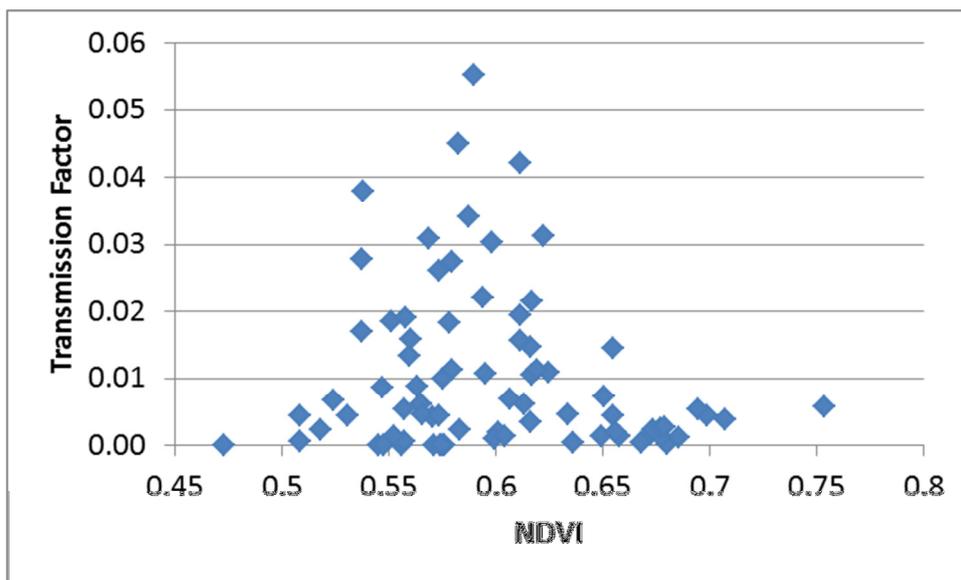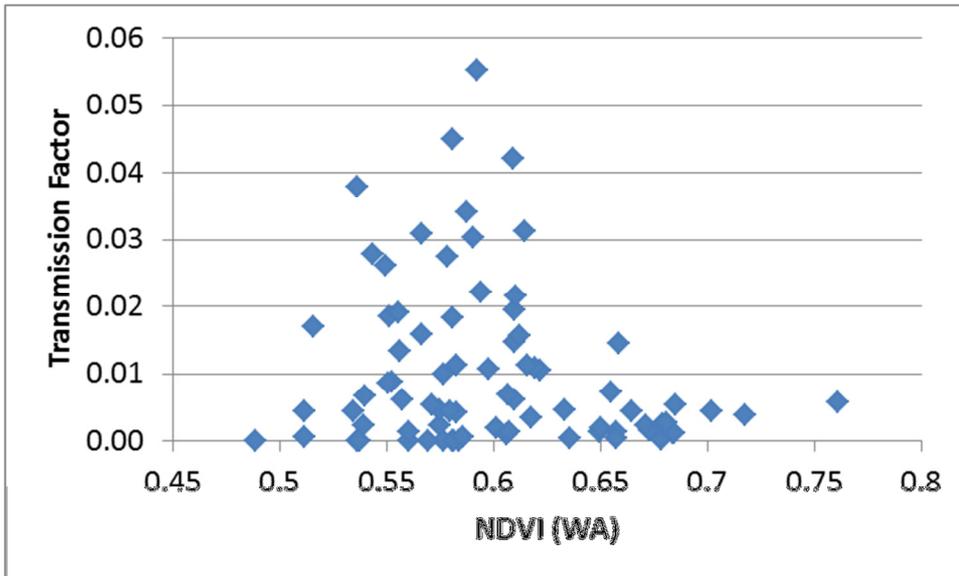
Figure 14: Location of the five farms having high VI and low transmission rates.

Figure 15: Correlation between transmission factor and vegetation index for NDVI (years: 2006 to 2009). Top: normal average, bottom: distance-weighted average.

Finally, four typical examples of farms, vegetation indices and transmission factors are shown in the following Figures:

- Figure 16 (id 96) is located North of Coevorden and represents a farm with a relatively high VI and a very low transmission rate.
- Figure 17 (id 73) is located close to Gouda and has also low transmission rate and a very high VI.
- Figure 18 (id 108) is located North of Roosendaal and has a very low transmission rate but also a low VI.
- Figure 19 (id 99) is located East of 's-Hertogenbosch and has a high transmission rate and a low VI.
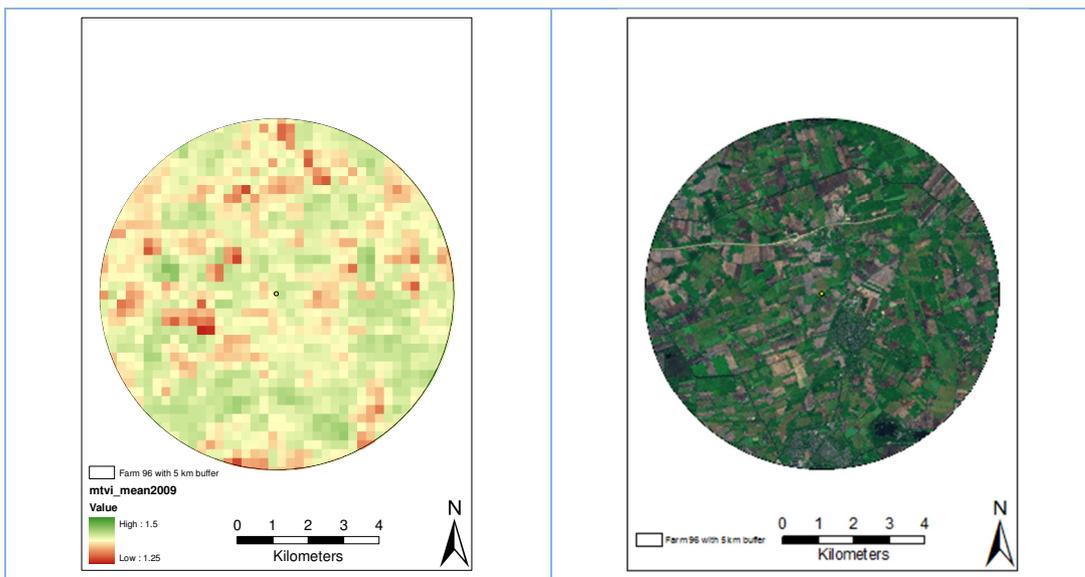
Figure 16: Typical example of vegetation index for farm ID 96, year 2009. Left MTVI 2009 and right visible image (unknown) date.
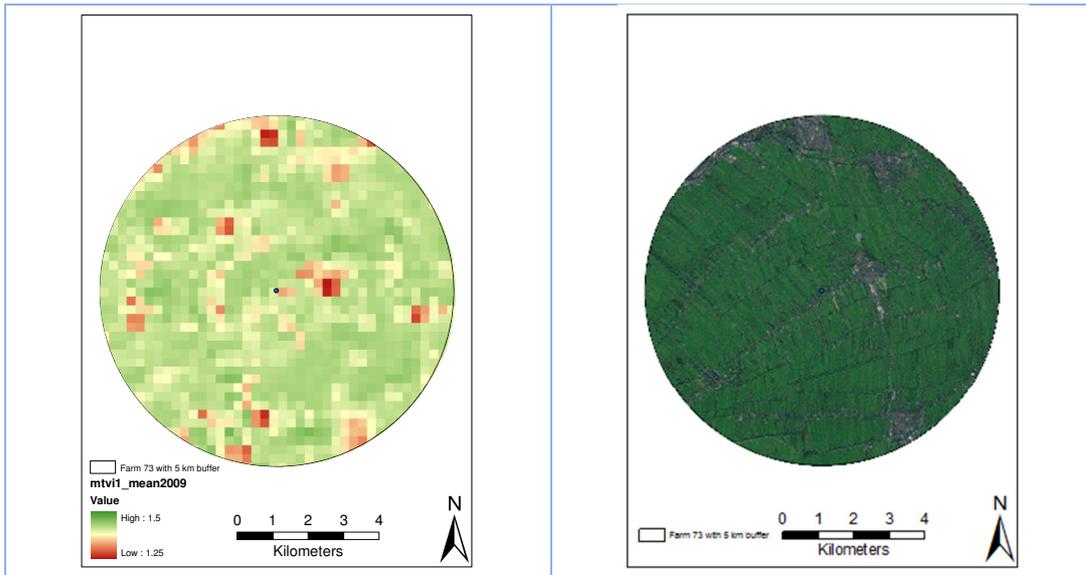
Figure 17: Typical example of vegetation index for farm ID 73, year 2009. Left MTVI 2009 and right visible image (unknown) date.

Figure 18: Typical example of vegetation index for farm ID 108, year 2009. Left MTVI 2009 and right visible image (unknown) date.

Figure 19: Typical example of vegetation index for farm ID 99, year 2009. Left MTVI 2009 and right visible image (unknown) date.

5 Conclusion and recommendations

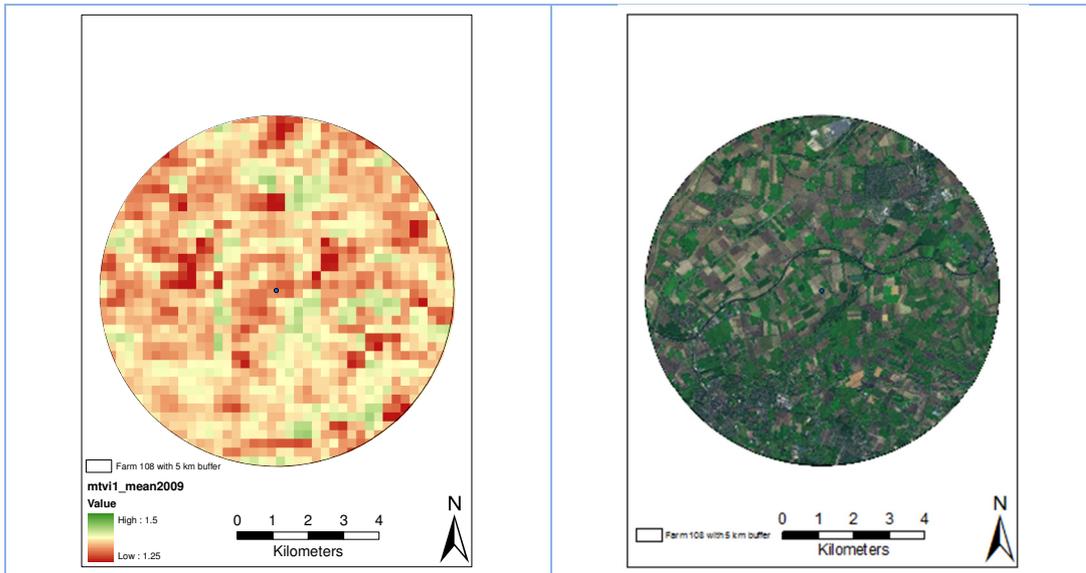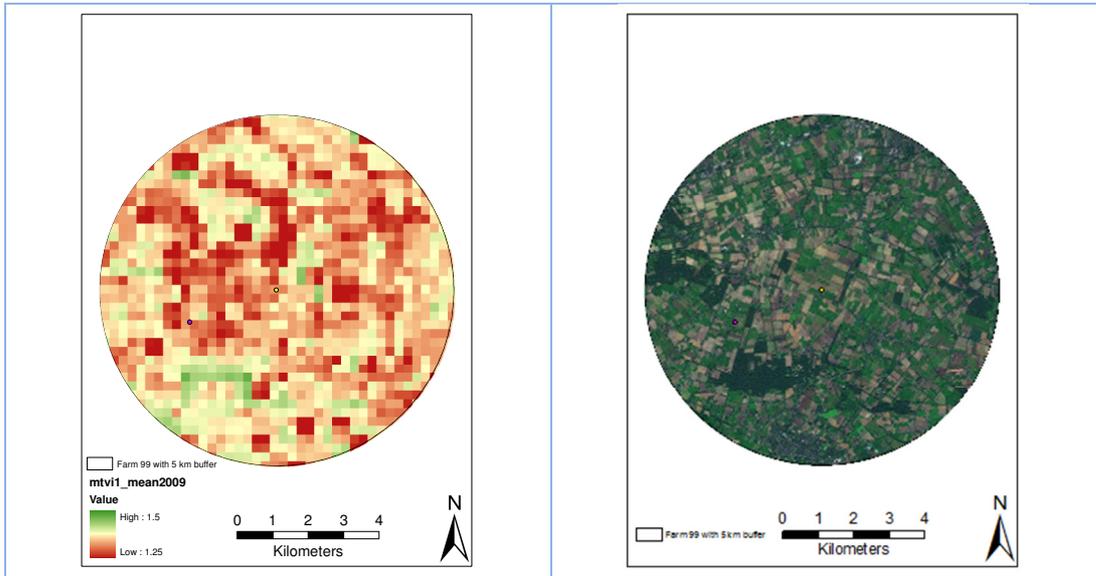Severe Q fever outbreaks in the Netherlands occurred in 2007, 2008 and 2009. Q fever is caused by infection with *C. burnetii*. Transmission of *C. burnetii*. is air-borne and is therefore more likely to occur under specific environmental conditions. However, so far most research on Q fever is clinical and/or epidemiological oriented. A first explorative study concluded that environmental factors might play an important role on the transmission of *C. burnetii* (Hunink et al., 2010; Van der Hoek et al., 2011). This explorative study concluded that vegetation around infected farms might be an important environmental factor.

The study presented in this report focused on a more detailed analysis of satellite derived vegetation indices (VI's). The most important conclusions from this study can be summarized as:
- Satellite derived vegetation indices show a strong correlation between transmission of Q fever and the vegetation surrounding infected farms.
- Correlation between VI and transmission is not linear, but threshold based.
- High vegetation is strongly correlated with low transmission rates. Low and intermediate vegetation does not correlate with transmission rates.
- All VI's, with the exception of the NDWI one, perform approximately equally well.

Important recommendations for further studies are:
- Expand results of this study in a more inclusive analysis using soil moisture, amongst others, as other environmental factor.
- Expand analysis to longer time periods (e.g. 2010) and other areas with known outbreaks of Q fever.
- A more detailed analysis of timing of transmission and the VI at that particular moment could be undertaken.
- A clear distinction between bulk tank milk infection and abortion waves as contamination source.
- Research if other than the Pearson correlation will show stronger correlations.

For a policy oriented recommendation it seems evident that very dense vegetation can reduce the risk of transmission substantially.

6 References

Berri M, Rousset E, Champion JL, Russo P, Rodolakis A. 2007. Goats may experience reproductive failures and shed Coxiella burnetii at two successive parturitions after a Q fever infection. Res Vet Sci. 83:47–52.

Chepil, W. S., 1956. Influence of moisture on erodibility of soil by wind. Soil Science Society of American Proceedings, 20, 288-292.

Engelstaedter, S., K. E. Kohfeld, I. Tegen, and S. P. Harrison, 2003. Controls of dust emissions by vegetation and topographic depressions: An evaluation using dust storm frequency data Geophyiscal Research Letters, Vol. 30, 6, 1294, doi:10.1029/2002GL016471

Fécan, F., B. Marticorena and G. Bergametti, 1998. Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil moisture for arid and semi-arid areas. Annales Geophysicae 17, Nr. 1: 149-157. doi:10.1007/s00585-999-0149-7, .

Glenn, Edward P., Alfredo R. Huete, Pamela L. Nagler and Stephen G. Nelson, 2008. Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape. Sensors 2008, Nr. 8, doi: 2136-2160

Hawker J.I., Ayres, J.G., Blair, I., 1998. A large outbreak of Q fever in the West Midlands: windborne spread into a metropolitan area? Commun Dis Public Health 1998; 1: 180–87.

van der Hoek, W., J. Hunink, P. Vellema, P. Droogers. 2011. Q fever in The Netherlands: the role of local environmental conditions. International Journal of Environmental Health Research. 2011: 1–11.

Hunink, J.E., T. Veenstra, W. van der Hoek, P. Droogers, 2010. Q fever transmission to humans and local environmental conditions. FutureWater report 90, www.futurewater.nl/downloads/2010_Hunink_FW90.pdf

Kaufman et al., 1998 Y.J. Kaufman, D.D. Herring, K.J. Ranson and G.J. Collatz, Earth observing system M1 mission to earth. *IEEE Transactions on Geoscience and Remote Sensing*, 36 4 (1998), pp. 1045–1055.

Lancaster N., A Baas, 1988. Influence of vegetation cover on sand transport by wind: field studies at Owens Lake. Earth Surface Processes and Landforms, 1998 - aeoliangeomorphology.org

Van Leuken, J., Swart, A., Hunink, J., Havelaar, A., Van der Hoek, W. (2012) The role of environmental factors on the transmission of Coxiella burnetii to humans, National Institute of Public Health and the Environment, Bilthoven, the Netherlands, and Utrecht University, Utrecht, the Netherlands, in preparation.

Panda, Sudhanshu Sekhar; Ames, Daniel P.; Panigrahi, Suranjan. 2010. "Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques." *Remote Sens.* 2, no. 3: 673-696.

Schimmer B, Morroy G, Dijkstra F, Schneeberger P.M., Weers-Pothoff G, Timen A, Wijkmans C, van der Hoek W., 2008. Large ongoing Q fever outbreak in the south of The Netherlands. Euro Surveillance. 2008;13(31):pii=18939.

Schimmer B, Ter Schegget R, Wegdam M, Zuchner L, de Bruin A, Schneeberger PM, Veenstra T, Vellema P, van der Hoek W. 2010. The use of a geographic information system to identify

a dairy goat farm as the most likely source of an urban Q fever outbreak. BMC Infect Dis. 10:69.

Stockton, P.H., D. A. Gillette, 1990. Field measurement of the sheltering effect of vegetation on erodible land surfaces Land Degradation & Development Volume 2 Issue 2, Pages 77 – 85

Sokolik I.N., and Toon O.B., 1996. Direct radiative forcing by anthropogenic airborne mineral aerosols. Nature 381, 681 - 683; doi:10.1038/381681a0

Solaimani K. , Shokrian F., Tamartash R. and Banihashemi M., 2011. Landsat ETM+ Based Assessment of Vegetation Indices in Highland Environment. Journal of Advances in Developmental Research 2 (1) 2011 : 5-13.

Tegen, I., and I. Fung, 1994. Modeling of mineral dust in the atmosphere: Sources, transport, and optical thickness, Journal of Geophysical Research, 99(D11), 22,897–22,914.

Tigertt WD, Benenson AS, Gochenour WS., 1961. Airborne Q fever. Bacteriol Rev 25:285–293.

Tissot-Dupont, H., Amadei M., Nezri M. and Raoult D., 2004. Wind in November, Q fever in December. Emerging infectious diseases 10, Nr. 7: 1264-9.

Vermote, E. F., & Vermeulen, A. (1999). Atmospheric correction algorithm: Spectral reflectances (MOD09). University of Maryland, Dept of Geography. Technical   background document. Algorithm Theoretical background Document available online: http://modarch.gsfc.nasa.gov/data/ATBD/atbd_mod08.pdf.

Yanase T., Muramatsu Y., Inouye I., Okabayashi T., Ueno H., Morita C., 1998. Detection of *Coxiella burnetii* from dust in a barn housing dairy cattle, Microbiol. Immunol. 42, 51–53.

Appendix 1: R-script

Script to calculate the a, b, and r^2 values for the unweighted VI averages from 2006 through 2009.

```
1
2  # SYSTEM
3  #install.packages("gdata")
4  library(gdata)
5
6  # CONFIGURATIONS
7  years                          = c(2006,2007,2008,2009)
8  stat                           = c("a","b","r2")
9  ind                            = c("RDVI","NDVI","DVI","EVI","GI","IPVI","MND","MSR",
10                                    "NDWI","NLI","RVI","SAVI","TVI","VARI","MSAV","MTVI",
11                                    "OSAV","GNDV","WDRV")
12 cor                            = matrix(data = NA, nrow = length(ind),
13                                     ncol = length(years)*3)
14 colnames.cor                   = NULL        # Initialisation of columnames of matrix 'cor'
15 years.stat                     = seq(1,length(years)*length(stat),3) # Column name of each year in 'cor' for stat 'a'
16
17 # COLUMN AND ROW NAMES OF 'COR'
18 rownames(cor) = ind
19 for (i in years) {
20   for (j in stat) {
21     colnames.cor               = rbind(colnames.cor, paste(i,j,sep=""))
22   }
23 }
24 colnames(cor)                  = colnames.cor
25
26
27 # READ DATA
28 data2                          = read.xls(xls="c:/active/Q-koort/dp1_transmissionfactors-fin.xlsx",
29                                       sheet="1",perl="c:/perl/perl/bin/perl.exe")
30
31
32 # CREATE CORRELATION MATRIX
33 for (vi in 1:length(ind)) {
34 #for (vi in (1:1)) {
35   data1                        = read.table(file=paste("C:/Active/Q-koort/working/_29_",ind[vi],".txt",sep=""),
36                                       header=TRUE,sep=",")
37   for (y in 1:length(years.stat)) {
38     model                      = lm(data2[,y + 5] ~ data1[,y + 1])
39     summary                    = summary(model)
40     coefficients               = summary$coefficients
41     a                          = coefficients[2,1] # helling
42     b                          = coefficients[1,1] # intercept
43     r2                         = summary$r.squared
44     cor[vi,years.stat[y]:(years.stat[y]+2)] = c(a,b,r2)
45   }
46 }
47 write.table(x=cor,file="c:/Active/Q-koort/working/correlation.txt",sep=",",
48             row.names=T,col.names=T)
49
50
```

Script to calculate the weighted and unweighted r^2 values for all indicis in the year 2009.

```
1  |
2  # SYSTEM
3  #install.packages("gdata")
4  library(gdata)
5
6  # CONFIGURATIONS
7  years                          = c(2006,2007,2008,2009)
8  stat                           = c("a","b","r2","pf")
9  ind                            = c("RDVI","NDVI","DVI","EVI","GI","IPVI","MND","MSR",
10                                    "NDWI","NLI","RVI","SAVI","TVI","VARI","MSAV","MTVI",
11                                    "OSAV","GNDV","WDRV")
12 cor                            = matrix(data = NA, nrow = 5,
13                                     ncol = dim(data1)[2]-2)
14 colnames.cor                   = NULL        # Initialisation of columnames of matrix 'cor'
15 years.stat                     = seq(1,length(years)*length(stat),3) # Column name of each year in 'cor' for stat 'a'
16
17 # COLUMN AND ROW NAMES OF 'COR'
18 #rownames(cor) = ind
19 #for (i in years) {
20  # for (j in stat) {
21    # colnames.cor              = rbind(colnames.cor, paste(i,j,sep=""))
22  #}
23 #}
24 #colnames(cor)                 = colnames.cor
25
26
27 # READ DATA
28 data1                          = read.xls(xls="c:/active/Q-koort/working/all-excl-NA-2009.xlsx",
29                                       sheet="1",perl="c:/perl/perl/bin/perl.exe")
30
31
32 # CREATE CORRELATION MATRIX
33
34 for (i in 3:dim(data1)[2]) {
35   model = lm(data1[,i] ~ data1[,2])
36   summary                      = summary(model)
37   coefficients                 = summary$coefficients
38   a                            = coefficients[2,1] # helling
39   b                            = coefficients[1,1] # intercept
40   r2                           = summary$r.squared
41   P                            = pf(summary$fstatistic[1], summary$fstatistic[2],summary$fstatistic[3],lower=F)
42   cor[1:4,i-2]                 = c(a,b,r2,P)
43 }
44 write.table(x=cor,file="c:/Active/Q-koort/working/all-excl-NA-2009.txt",sep=",",
45             row.names=T,col.names=T)
46
47 #for (vi in (1:1)) {
48   #data1                       = read.table(file=paste("C:/Active/Q-koort/working/_29_",ind[vi],".txt",sep=""),
49   #                                  header=TRUE,sep=",")
50
51
```