



Universitat de Girona

CONFORMAL PREDICTION OF AIR POLLUTION CONCENTRATIONS FOR THE BARCELONA METROPOLITAN REGION

Olga IVINA

Dipòsit legal: Gi. 473-2013

<http://hdl.handle.net/10803/108341>



Conformal prediction of air pollution concentrations for the Barcelona Metropolitan Region està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 No adaptada de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/)

© 2013, Olga Ivina



University of Girona

PhD Thesis

Conformal prediction of air pollution concentrations
for the Barcelona Metropolitan Region

Olga Ivina

2012



University of Girona

PhD Thesis

Conformal prediction of air pollution concentrations
for the Barcelona Metropolitan Region

Olga Ivina

Doctoral Programme in Experimental Sciences and Sustainability

Scientific advisor: Dr. Marc Saez Zafra

A thesis submitted for the degree of Philosophiæ Doctor (PhD) by
the University of Girona

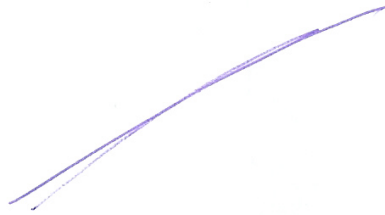
2012

Certificat de direcció de la tesi

El Dr. Marc Saez Zafra, professor del departament d'Economia de la Universitat de Girona, director del grup de recerca Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS),

CERTIFICO:

Que aquest treball, titulat "*Conformal prediction of air pollution concentrations for the Barcelona Metropolitan Region*", que presenta Olga Ivina per a l'obtenció del títol de doctora, ha estat realitzat sota la meua direcció i que compleix els requeriments per poder optar a Menció Europea.



Girona, 23 de març de 2012.

Acknowledgements

I am profoundly grateful to the University of Girona and, in particular, to its OITT unit, for awarding me the BR research fellowship that has financially supported me during all these years.

Moreover, I deeply thank the CIBER of Epidemiology and Public Health of Spain for endowing me with the scientific mobility fellowship to support my European doctorate research practice at the Helmholtz Zentrum München, Germany.

Besides, I cannot but acknowledge the enormous contribution of my scientific advisor, Prof. Marc Saez Zafra, PhD, to my work. I am thankful to him for his guidance, advice and a helping hand.

Also, I am very much obliged to Prof. Alex Gammerman, PhD, for his support and valuable counseling within my research internship at the Computer Learning Research Centre of the Royal Holloway, University of London, and afterwards.

Finally, I must thank my parents, my husband and my mother-in-law for being my towers of strength, for being so very patient with me and for constantly cheering me on.

Contents

Acknowledgements	i
Resum d'aquesta tesi	iv
Summary of the present dissertation	vii
1 Introduction	1
1.1 Air pollution and its effects	1
1.2 Air pollution exposure assessment	5
1.3 Conformal prediction for air pollution	7
2 Objectives	10
3 Methods and data	12
3.1 Kriging	12
3.1.1 General definitions	12
3.1.2 Ordinary kriging	14
3.2 Conformal predictors	15
3.2.1 General definitions	15
3.2.2 Ridge regression procedure	17
3.2.3 Ridge regression confidence machine	18
3.2.4 Dual form ridge regression confidence machine	20
3.2.5 Kernels	21
3.3 Computing	23
3.4 Data	24
3.4.1 Barcelona Metropolitan Region	24
3.4.2 The present data set	24
3.4.3 Previous research for BMR	29
4 Results	31
4.1 Ordinary kriging with exponential covariance	31
4.1.1 NO ₂	31
4.1.2 PM10	36
4.2 RRCM in iid setting	40

4.2.1	NO2	40
4.2.2	PM10	42
4.3	Comparison of both models in default settings	45
4.3.1	NO2	45
4.3.2	PM10	46
4.4	Ordinary kriging with Gaussian covariance	48
4.4.1	NO2	48
4.4.2	PM10	51
4.5	RRCM with RBF kernel	55
4.5.1	NO2	55
4.5.2	PM10	57
4.6	Comparison of both models in Gaussian setting	60
4.6.1	NO2	60
4.6.2	PM10	62
4.7	RRCM with polynomial kernel of second order	64
4.7.1	NO2	64
4.7.2	PM10	67
4.8	Comparison of kriging and ridge regression confidence machine with polynomial kernel	69
4.8.1	NO2	69
4.8.2	PM10	70
4.9	Comparison of ridge regression confidence machine models	72
4.9.1	NO2	72
4.9.2	PM10	79
5	Discussion	85
5.1	General notions	85
5.2	Fittng the models	85
5.3	Efficiency of predictions	87
5.4	Bottom line	88
6	Conclusion	90
6.1	The subject of this study	90
6.2	Conformal predictors and geostatistics	91
6.3	Key points and further development of this work	93
	Appendix A. R functions code	CD ROM
	Appendix B. Detailed results	CD ROM

Resum d'aquesta tesi

Aquest treball està destinat a introduir el nou mètode de les màquines d'aprenentatge, els predictors de conformació, per l'avaluació de la contaminació de l'aire. Per l'àrea d'estudis, que és la Regió Metropolitana de Barcelona (RMB), s'han desenvolupat uns models per la predicció conforme. Aquests models estan basats en l'especificació que es diu la màquina de confiança de la regressió cresta (RRCM) [1, 2]. La regressió cresta és una generalització del tradicional mètode dels mínims quadrats proposada pel A. E. Hoerl [3]. Aquesta tècnica té l'objectiu de fer front als problemes mal definits. Els predictors de conformació que s'han desenvolupat per les finalitats d'aquest estudi són uns models de regressió (cresta), que ofereixen prediccions vàlides, com faria qualsevol predictor de conformació. En lloc de les prediccions puntuals, un predictor de conformació genera un conjunt de predicció, que pot adoptar formes diferents - des d'un punt fins a tota la recta real. No obstant això, en la pràctica és gairebé sempre un interval. És naturalment desitjable que aquests conjunts serien tan petits com sigui possible, és a dir, que l'eficiència de la predicció seria òptima.

L'algorisme subjacent dels predictors de conformació derivats i discutits al llarg d'aquesta tesi és kriging ordinari [4]. És un dels mètodes de la família kriging anomenada de l'enginyer sud-africà D.G. Krige. Kriging implica l'interpolació espacial que s'utilitza per estimar el valor d'un factor en un lloc no observat sobre la base dels valors d'aquest factor en els llocs observats. L'avantatge d'aquest mètode de regressió és que proporciona l'estimació del valor del factor juntament amb l'estimació de la variància de l'error.

L'especificació del kriging ordinari permet considerar la dependència espacial entre els llocs d'observació amb l'ús dels funcions de variograma i de covariància [5]. El predictor de conformació basat en el kriging ordinari pot capturar la distribució i la dependència espacials mitjançant una tècnica que es diu "el truc del nucli" ("kernel trick"). Aquesta tècnica, introduïda pel V. Vapnik i els altres, està dirigida a tractar els problemes d'alta dimensió i s'ha utilitzat per primera vegada per a les màquines de vectors suport [1, 6]. En aquesta investigació, el nucli s'introdueix a un model de predicció de

conformació de forma anàloga com una funció de covariància s'introdueix a un model de kriging ordinari.

Des de la pràctica geoestadística se sap que el kriging té alguns inconvenients. La reclamació més greu sobre l'ús de kriging per a l'avaluació de la contaminació de l'aire és que proporciona les superfícies d'estimació molt suavitzades. Això passa perquè els models de kriging normalment fan la interpolació de la contaminació de l'aire sobre la base de dades de les xarxes de monitorització que solen ser escasses [7]. És per això que les variàncies de kriging, o les variàncies d'error, tendeixen a ser grans. Els models de predicció de conformació que es deriven en aquesta tesi no resolen el problema de l'excés de suavitzat. No obstant això, els intervals de predicció que proporcionen són vàlids, és a dir, el valor real d'un factor cau dins d'aquests intervals amb una probabilitat especialment donada. La grandària dels intervals de predicció, o amb altres paraules, l'eficiència de la predicció, pot ser alterada amb una selecció apropiada del nucli i amb un ajust adequat del paràmetre cresta.

En aquesta tesi, els models de kriging ordinari i els corresponents models de predicció de conformació s'han derivat per investigar les concentracions de contaminació de l'aire a la Regió Metropolitana de Barcelona. S'han estudiat dos compostos químics : el diòxid de nitrogen i les partícules. Les concentracions d'aquests contaminants tenien la forma de mitjanes anuals. La línia de temps de l'estudi s'estén des de 1998 fins a 2009 per al NO_2 , i de 2001 a 2009 per a les PM_{10} . En total, les dades han estat disponibles per 49 estacions de mesurament a tota la RMB. No obstant, la quantitat total dels valors mancants al conjunt de dades arribava gairebé a la meitat de tot el conjunt. Aquesta és la raó perquè les prediccions s'han fet de manera tòpica, és a dir, per a cada ubicació de l'estació i l'any i contaminant, on havien els valors mancants. Aquesta manera de predicció permet comparar les concentracions previstes als observats per altres anys, i la predicció pot ser estesa en una reixeta, si cal. La validació creuada, tant pels models de predicció de kriging com pels predictors de conformació, ha sigut realitzada per a cada any i contaminant. Diverses funcions han sigut utilitzades com a nuclis i/o funcions de covariància pels models pertinents: exponencial, gaussiana i polinomi.

Quant als resultats, dos enfocaments han donat resultats acceptables i comparables, amb una diferència important en el medi que consisteix en el fet que els resultats obtinguts pels predictors de conformació són sempre vàlids. Respecte a aquesta qualitat dels predictors de conformació, podria resultar beneficiós desenvolupar els models d'aquest tipus a la base dels altres algorismes estadístics de l'avaluació de contaminació de l'aire, com per exemple la regressió de l'ús del sòl.

Els resultats dels enfocaments de modelització proposats en aquest treball justifiquen el desenvolupament ulterior i l'aplicació d'aquests mètodes a les dades de contaminació de l'aire. Per tant dels contaminants considerats en aquest treball, l'obtenció precisa dels paràmetres de la distribució espacial és difícil i gairebé impossible a causa de la petita grandària dels conjunts de punts d'observació vàlids disponibles per a cada any. Malgrat això, sense previ anàlisi exhaustiva de la distribució de les dades, RRCM ha proporcionat conjunts eficaços de predicció. $35.32 \mu\text{g}/\text{m}^3$ és l'ample òptim d'interval de predicció per al NO_2 , i per al PM_{10} aquest valor és igual a $34.59 \mu\text{g}/\text{m}^3$. Aquests valors són mitjanes anuals. Una vegada que la forma de distribució de dades i els paràmetres de covariància poden ser estimats correctament, els models guanyaran força en eficiència.

Paraules clau: *predictors de conformació, kriging ordinari, regressió cresta, contaminació de l'aire, Barcelona.*

Summary of the present dissertation

The present dissertation is aimed to introduce the newly developed machine learning method, conformal predictors, for air pollution assessment. For the given area of study, that is Barcelona Metropolitan Region (BMR), several model for conformal prediction have been developed. These models have been based on the specification that is called ridge regression confidence machine (RRCM) [1, 2]. Ridge regression is a generalization of the traditional least squares method proposed by A. E. Hoerl [3] and aimed to deal with ill-posed problems. A conformal predictor that has been developed for the purposes of the present study is a (ridge) regression model, which provides valid predictions, as any conformal predictor would do. Instead of point predictions, a conformal predictor outputs a prediction set, which can take different forms - from a point to the whole real line. However, in practice it is almost always an interval. It is naturally desired that these sets would be as small as possible, i.e. that the efficiency of prediction would be optimal.

The underlying algorithm for the conformal predictors derived and discussed throughout this dissertation is ordinary kriging [4]. It is one of the so-called kriging methods named after the South African engineer D.G. Krige. Kriging methods imply spatial interpolation which is used to estimate the value of a factor at an unobserved spot on the basis of the values of this factor at the observed locations. The benefit of this regression method is that it provides an estimate of the value of the factor together with an estimate of the error variance.

The ordinary kriging specification of kriging allows to consider spatial dependence between observation locations with the use of variogram and covariance functions [5]. In the kriging-based conformal predictor, spatial distribution and dependence can be captured with the use of kernel trick. This technique, introduced by V. Vapnik and others, is aimed to deal with high-dimensional problems, and it has been first used for support vector machines [1, 6]. In this research, a kernel is introduced to a conformal predictor

model as an analogue of a covariance function in an ordinary kriging model.

In geostatistical practice, kriging is known to have some drawbacks. The most serious claim towards the use of kriging for air pollution assessment is that it provides over smoothed estimation surfaces. This happens because kriging models make their interpolation on the basis of air pollution monitoring data network which are usually sparse [7]. That is why the kriging variances, or the error variances, tend to be large. The conformal predictor models derived in this dissertation do not solve the problem of over smoothing. Nevertheless, the prediction interval they provide are valid, i.e. the real value of a factor falls within these interval with an ad hoc given probability. The size of the prediction intervals, i.e. the efficiency of the prediction can be altered with a proper kernel fit and with a choice of a ridge parameter.

In this dissertation, ordinary kriging and corresponding conformal predictor models have been derived to investigate air pollution concentrations at the Barcelona Metropolitan Region. Two substances have been taken up, and those are: nitrogen dioxide and particulate matter. The concentrations of these contaminants have been provided in a form of annual averages. The timeline of the study spreads from 1998 to 2009 for NO_2 , and from 2001 to 2009 for PM_{10} . In total, the data has been available for 49 measurement stations across the BMR. Nevertheless, the overall amount of missing values in the data has made up almost a half of the whole set. This is why the predictions have been made topically, i.e. for each station location and year and pollutant where the data was missing. This manner of prediction allows to compare the predicted concentrations to the observed ones for other years, and the prediction can be extended on a grid if needed. Cross-validation, both for kriging and conformal prediction models, has been performed for each year and pollutant. Several functions have been taken up as kernels and/or covariance functions in the relevant models: exponential, Gaussian and polynomial.

As far as results are concerned, on the bottom line, both approaches have provided decent and comparable results, with a major difference that consists of the fact that conformal prediction results are always valid. Regarding this quality of conformal predictors, it could be beneficial to derive models of such kind on top of other successfully applied air pollution assessment algorithms, such as land-use regression.

The results of the proposed modeling approaches justify further development and application of these methods to air pollution data. For both of the pollutants considered in this research work, the precise elicitation of spatial distribution parameters was difficult and almost impossible due to small sets of valid observation points available for each year. Despite this, with

no thorough preliminary analysis of data distribution, RRCM has provided effective prediction sets. $35.32 \mu\text{g}/\text{m}^3$ is the optimal width of prediction interval for NO_2 , and for PM_{10} this value is equal to $34.59 \mu\text{g}/\text{m}^3$. These values are annual averages. Once data distribution form and covariance parameters can be properly estimated, the models will sufficiently gain in efficiency.

Keywords: *conformal predictors, ordinary kriging, ridge regression, air pollution, Barcelona.*

Chapter 1

Introduction

1.1 Air pollution and its effects

Air pollutant is a problem of growing concern all over the world. By the present moment, there has been accumulated a great body of the scientific evidence on hazardous effect of air pollution on people's health and well-being, as well as on crops, plants and animals. Air pollution causes both acute and chronic effects in people's health: from minor upper respiratory irritation to chronic respiratory and heart disease, and lung cancer [8]. Also, exposure to contaminants has been connected with premature mortality and reduced life expectancy.

In people, air pollutants are found associated with adverse health outcomes both in adults and in children. Besides, it is proven that the least ones are more susceptible to contamination [9]. It was identified within the European Union's Ministerial Conference on Children's Environmental Health in 2004 that exposure to air pollution is among the major contributors to death and disabling of children in Europe [10]. It has been also reported that almost 90 per cent of the residents of urban environments, including children, were exposed to the levels of pollutants exceeding the World Health Organization's guideline values [11, 12]. Very young children, perhaps including the unborn ones, are particularly susceptible to pollution, and there is a great variety of evidence suggesting links between contaminants and health outcomes [11]. In particular, exposure to air pollution during pregnancy can affect fetal growth, as well as it is associated with low birth weight, and with preterm birth [13]. There is enough evidence that claims there exists the connection between air pollution and lung function development in children [14]. Also, air pollution is linked to childhood asthma aggravation, mostly due to particulate matter and ozone exposure, and to cough and bronchitis. Moreover, pollutants were connected to hay fever and allergic rhinitis. Air pollution is also associated with the excess risk for mortality

in young children, especially in those with respiratory illnesses and infants with lower respiratory infections [15].

In adult population, traffic-related air pollution has been linked with respiratory outcomes, such as COPD (Chronic obstructive pulmonary disease), asthma or chronic bronchitis [16]. Exposure to fine particles has been associated to all-cause, cardiopulmonary and lung cancer mortality, as well as sulfur oxide pollution has been associated with the elevated mortality risk [17]. Sulfur dioxide, oxides of nitrogen and ozone, were concatenated with mortality, too[15]. Older adults are especially susceptible to pollution due to the decrease in physiological processes [18]. There exist an evidence of the link of cardiovascular morbidity with the exposure to particles in elderly people. Moreover, the people of 75 years of age and more are more susceptible to nonaccidental mortality caused by short-term exposure to PM_{2.5} and PM₁₀, and by long-term exposure to ultrafine particles.

Not only the age or the health status of a person determine the influence of air pollution on people, but there are other factors as well. Two of the major agents, influencing the exposure to contamination in micro and macro levels respectively are the socioeconomic status (SES) and the geographic region of residence. Populations with a low socioeconomic status are most probably exposed to air pollution to a greater extent, as well as these disadvantaged populations are perhaps more sensitive to pollutants [19]. Low SES is linked to limited access to medical care facilities, as well as to fresh food that hampers the intake of vitamins and fatty acids [18]. Also, it is associated with the prevalence of preexisting diseases. All of these factors may contribute to the development of air pollution related health effects. When it comes to other characteristics that may modify susceptibility to contamination, the following can be named apart from age and SES: genetic polymorphisms, preceding cardiovascular and respiratory diseases, and some additional factors like obesity [18].

Person-related factors are not the only decisive ones, but the exposure generally varies from one geographic region to another. In the last twenty years, there have been conducted numerous studies considering the geographic region as a health determinant [19]. From an epidemiological point of view, air pollution related health effects appear mainly as a result of long-term exposure [20], and living in areas with filthy air is connected to emersion of these effects [21]. Persons residing in cities or in their immediate surroundings are more likely to be exposed to air pollution as a result of growing industrialization and increased combustion of fossil fuels [8].

Existing monitoring and exposure assessment information suggests that the levels of airborne particles are sometimes higher in major Asian cities

rather than in North American or Western European cities [15]. The European Space Agency reports on the global air pollution map, produced by its Envisat, the ten-instrument world's largest satellite for environmental monitoring launched in 2002 [22]. Its onboard Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) tool records the spectrum of sunlight, and then the obtained data is used to trace gases in the atmosphere. It has been revealed that high distribution of nitrogen dioxide were seen in major cities across North America, Europe and north-east China, as well as with Mexico City in South America, and with South Africa coal-fired power plants region. Figure 1.1 represents the global air pollution map created by Envisat's SCIAMACHY. This image has been produced by S. Beirle, U. Platt and T. Wagner of the University of Heidelberg's Institute for Environmental Physics.

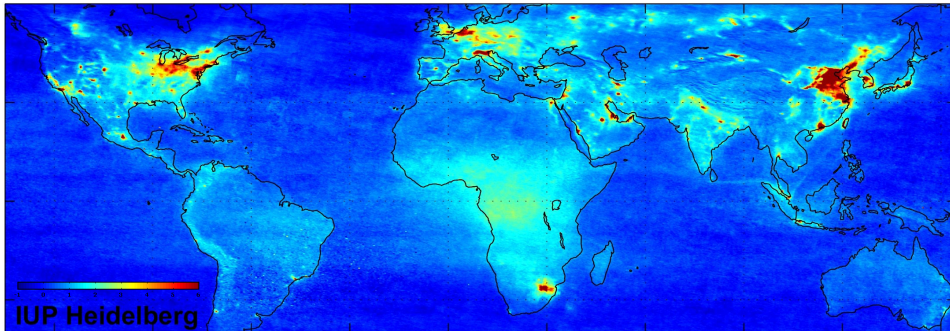


Figure 1.1: Global air pollution map produced by Envisat's SCIAMACHY

Air pollution is one of the major contributors to climate change. Black carbon is known for its capacity to absorb sunlight, heat the air and contribute to global warming. Generally, absorbing aerosols can affect regional climate [23], as well as lead to global effects [24]. For example, human-made aerosols can weaken the hydrological cycle, which can affect the availability and quality of fresh water.

In urban areas, the main source of air pollution is traffic [19]. About 50 per cent of nitrogen oxide emissions, including nitrogen oxide (NO) and nitrogen dioxide (NO₂), are produced by vehicle engine combustion [25]. Studies carried out in Europe conclude that traffic contributes up to 50 per cent of the average annual concentration of particulates with the aerodynamic diameter of less than 10 μm (PM₁₀) found in the ambient air [21]. This percentage is even higher for fine particles (those of the aerodynamic diameter less than 2.5 μm) and ultrafine particles (those of the aerodynamic diameter

less than $0.1 \mu\text{m}$). Sometimes, the major contaminants are referred to as “criteria” pollutants [26]. Those are: ground-level ozone, particles ($\text{PM}_{2.5}$ and PM_{10}), lead (Pb), nitrogen dioxide (NO_2), carbon monoxide (CO) and sulfur dioxide (SO_2). The present research takes up two contaminants that are most frequently monitored and assessed within epidemiological studies: nitrogen dioxide (NO_2) and particulate matter (PM_{10}).

NO_2 is a pollutant that has been associated with mortality and morbidity in people both in short-term and long-term exposure, as well as it contributes to forming of ground-level ozone, as causing the environmental and climate effects. Nitrogen dioxide is emitted to the atmosphere mainly from the traffic sources and power plants. World Health Organization points out multiple impact of NO_2 in their Air Quality Guidelines [12]. In short-term exposure, in high concentrations - those exceeding $200 \mu\text{g}/\text{m}^3$ - NO_2 is associated with adverse effects both in animals and in humans. Current research shows that short-term exposures, i.e. those in the range from 30 minutes up to 24 hours, are linked to adverse respiratory effects in healthy people, and with increase of respiratory symptoms in those with asthma [27]. Short-term exposures are also connected to augmented visits to emergency departments and hospital admissions for respiratory causes, in particular, asthma. In the long-term exposure, NO_2 is associated with the increase of the risk of coronary heart disease, particularly with fatal events [28]. Epidemiological studies indicate that the bronchitis symptoms increase in asthmatic children in association with annual concentrations of nitrogen dioxide. Furthermore, in the presence of the ultraviolet light and of hydrocarbons, nitrogen dioxide is transformed into ground-level ozone and of nitrate aerosols [12]. In order to reduce the burden of air pollution, WHO has established and published the guideline values for each of major pollutants, including NO_2 . Those values are based on expert evaluation of the up-to-date scientific knowledge, and they are intended mostly for policy-makers all over the world. Achieving the them for individual pollutants would yield public health benefits. For nitrogen dioxide, the guidelines are the following: $40 \mu\text{g}/\text{m}^3$ for the mean annual concentrations, and $200 \mu\text{g}/\text{m}^3$ for the hourly mean concentrations [12]. Those guideline values are the limit values for EU, that have entered into legal force January, 1, 2010 [29].

Particulate matter is a mixture of extremely small solid particles and liquid droplets [30]. The mixture is composed by acids, like sulfates and nitrates, metals, organic compounds and soil or dust particles. Particulate matter is classified by its size, i.e. aerodynamic radius, and those of special concern are with the aerodynamic diameter smaller than 10 and 2.5 micrometers respectively. The first ones are referred to as “coarse particles”, or PM_{10} , the least ones are named “fine particles”, or $\text{PM}_{2.5}$. The size of particles is directly connected with their potential health effects [30]. From

10 micrometers in diameter or smaller, they enter the lungs once inhaled, and they can also affect heart. The range of adverse health effects of airborne particulate matter is wide, but mainly it affects cardiovascular and respiratory systems [12]. The epidemiological studies indicate that there is an association with particles and diseases both in short- and long-term exposures. Short-term effects include aggravating the condition when suffering from cardiovascular or respiratory diseases, hospital admissions, or even premature deaths. Long-term exposure can lead to development of heart or lung disease, and they are also associated with premature mortality [26]. PM_{10} is the most widely reported measure of particles pollution, and also the indicator of relevance to the majority of the epidemiological data [12]. Nevertheless, the WHO air quality guidance is based on the studies that take up fine particles, i.e. $PM_{2.5}$. The values for the PM_{10} concentrations are then obtained by applying a $PM_{10}/PM_{2.5}$ ratio by 0.5. The guideline values for the PM_{10} are thus the following: $20 \mu\text{g}/\text{m}^3$ for the annual mean concentrations, and $50 \mu\text{g}/\text{m}^3$ for the 24-hour mean concentrations, which should not be achieved more than 35 times per year. [31]. Those guideline values are the limit values for EU, which have entered into legal force on January 1, 2005 [29].

1.2 Air pollution exposure assessment

By the present moment, there has been accumulated a great volume of studies that investigate exposure of the individuals to air pollution. Direct measurements of the concentrations of contaminants at the residence address of a given person are not always available. In order to predict the levels of air pollution concentration at a given geographic site for a given period of time, statistical modeling is used. There exist a large number of models suitable for various study cases and data available. Different model classes can be named: proximity models, geostatistical models, land use regression (LUR), dispersion models, integrated meteorological emission techniques, and hybrid methods [7].

Proximity-based models are based on the simple assumption that the exposure to air pollution is directly associated with the distance to the emission source, that, in case of traffic-related contaminants, is roads. Some studies that make use of this method establish a buffer around the source, and all those persons who live inside of the buffer are classified as exposed, whilst those who live outside are classified as not or less exposed.

Interpolation models are mostly based on geostatistical methods. The target pollutant is measured at a spatial point, and this measure is dis-

tributed over the study area. The objective of those methods is to generate the estimates of the pollutions in the locations other than the monitoring sites. The most common geostatistical interpolation technique is kriging [5, 7, 4, 32, 33, 34, 35]. It is a group of methods first introduced in the 1950s and formalized and developed since then [4]. Kriging is famous for its capacity to provide the best linear unbiased estimates (BLUE) for any point of a spatial region. The core hallmark of kriging, however, is that in each and every point it does not only provide the estimate, but also the error variance, otherwise known as the "kriging variance". As for other interpolation methods, they include splines, inverse distance weighting and similar techniques. Although the interpolation methods are relatively precise, they are purely mechanical, and thus might fall apart when there is a lack of data. The dependance on availability of monitoring data is the core disadvantage of kriging [7]. Geostatistical interpolation models require a dense network of sampling sites. The average number of measurement points involved in studies is between 10 and 100, and this number depends on many factors such as the scale of the analysis, characteristics of the study region, meteorological conditions etc. Poor availability of data leads to over-smoothed pollution surfaces and large errors in predictions, especially in parts where few observations are given. This problem can be aggravated for contaminants that are known to vary significantly over small scales, such as nitrogen dioxide. To tackle these problems, primary data collection is needed, but it might be costly.

Land-use regression models [7] are aimed to predict the levels of target pollutant at a spatial point based on the land use and traffic characteristics of the surrounding area. This technique involves the least squares regression modeling for predicting pollution surfaces on the basis of the exogenous independent variables. The main advantages of this method are: its structure which allows the adaptation of the method to various environments without additional data acquisition, and that it is relatively low-cost. The main drawback of LUR is that it works well for relatively homogeneous areas - in the sense of land use, meteorology and vehicle. Land-use regression has been recognized as a standard approach for predicting pollutant concentrations using concentration measures, GIS spatial parameters and site characteristics [36]. As being specific, these models have shown better capacity in capturing small-scale intra-urban variability of air pollution than kriging models, integrated meteorological-emission models, or dispersion models. Nonetheless, LUR models are hard to generalize, because the parameters of the model are derived for a given spatial region, and thus the estimates obtained for two different cities might not be comparable.

Dispersion models generally rely on Gaussian plume equations [37, 7]. Those models describe the dilution and transport of pollution from its sources as a stationary process, implying that the distribution is Gaussian. They

require pollution, meteorological and emission data. The advantage of such models is that they incorporate the spatiotemporal variation of air pollution without needing dense monitoring networks. Among the disadvantages, the following can be listed: those models are expensive, they might require extensive cross-validation, and the Gaussian approach is not always suitable in reality.

Integrated meteorological-emission techniques consist of two modules, which embrace the chemistry and the meteorology [7]. These techniques imply the simulation of pollution involving the meteorological data: it is provided to chemistry modules at every step of the simulation. The contrary is not necessary because chemical data might have a scant impact on meteorology. The prediction efficiency of these combined models depend on various factors like entry data, grid resolution etc. IME models are costly to implement, they require a high-end computer facilities and can be performed by high-skilled people only, and thus the complexity of their practical use is their main drawback. The advantage of the IME models is the high precision of the estimates, as these models are capable of simulating various scenarios, which could incorporate even secondary pollutants such as ozone.

Hybrid models incorporate personal, regional or some other type of monitoring with one or two of the previously named exposure models [7]. The main advantage of these models is the possibility of validation, while the drawbacks depend on the model specification being used.

Comparing all the listed techniques, it is obvious that the choice of a model mostly depends on the data available and the problem to investigate. It has been established in research that simple metrics like road buffers are associated with children respiratory health outcomes [7]. However, if the size of the effect matters, a more sophisticated model might be required. In general, while all the listed techniques make use of the GIS data, the interpolation models are low demanding but yet precise. This particular research makes use of the classical interpolation method, kriging. It is a well-developed approach, because it yields relatively precise estimates making use of relatively little data, and it provides the estimates together with the estimation variance.

1.3 Conformal prediction for air pollution

However, one of the problems of nowadays existing methods for air pollution exposure assessment is that nearly all of them yield point estimates that might lack confidence. Even in case of kriging, the estimation variance that

is computed together with the estimate can be quite broad, so the validity of the prediction can not be guaranteed. In order to tackle this problem, this research suggests making use of a newly developed approach that is conformal predictors [1].

A conformal predictor is determined by some nonconformity measure. Given a set of objects, or observations, when a new object comes along, the nonconformity measure indicates how different this new example is from the old ones in the set. The old objects in the set are somehow labeled, while the new object has no label, and it is to be predicted. A conformal predictor is a “confidence predictor”. Provided with a level of confidence, a conformal predictor outputs a set of possible labels for the new example that should contain the object’s actual label. The new object is labeled, assuming that it will conform with the old objects in the set. The confidence level for prediction is a complimentary term for the significance level which determines the required amount of conformity for prediction. In regression terms, the label for the new object is the value of the independent variable to estimate.

The main feature of conformal predictor is that its estimates are always valid, i.e. for a given level of confidence, in the long run, the probability of error of prediction does not exceed the value implied by this level of confidence. A desired confidence level is introduced ad hoc by a researcher to the model. For example, if the confidence level is set to 99 per cent, this means that there might be up to 1 per cent of errors in the output of the predictor. Thus it will be a highly confident prediction.

Another important point for conformal predictors is the efficiency of prediction. The output of a conformal predictor comes in a form of a prediction set which can be: an interval, a ray, a union of two rays, the whole real line, or empty. Also, it may be a union of such sets. In practice, the most common type of a prediction set is an interval. Efficiency of prediction implies that the obtained prediction sets (or intervals) should be as small as possible.

A valuable feature of a conformal predictor is its flexibility. This means that a conformal predictor can be build upon any algorithm used in statistics such as bootstrap, neural network, (ridge) regression, support vector machines, Bayesian algorithms etc. [1]. Those initial methods, or underlying algorithms, remain intact. A conformal predictor derived from a traditional algorithm inherits the prediction power of the underlying method, and in the same time, is always valid.

The classical interpolation approach used in this research is kriging. In order to build a conformal predictor on top of this method, a regression specification is used that is called Ridge Regression Confidence Machine (RRCM)

[1, 38]. Ridge regression is a sophistication of the classic least squares method proposed in the 1960s. It implies introducing a small ridge coefficient to the regression equation [39]. This is done in order to tackle possible problems encountering when the data is nonorthogonal, like the need to invert a matrix close to singular.

In this study the initial use is made of the iid model that makes no specific requirements on data distribution, apart from that they should be independent and identically distributed (iid) [38]. The linear iid model has been further advanced with the aim to boost the prediction efficiency. In order to do so, a special technique has been used, that is referred to as a “kernel trick”. This method has been applied by Boser, Guyon and Vapnik [6] for dealing with high-dimensional data. Kernel trick implies the employment of some kernel function instead of a scalar product in the regression equation. In case GIS data, which do not form a high-dimensional set, a non-linear kernel can be treated as a way to take into account spatial structure of the data and spatial covariances between observations at different sites.

RRCM can be treated as an alternative to kriging, that is based on the same regression principle, but provides valid prediction intervals with high confidence instead of point estimates. The kernel approach in the RRCM can be reciprocally handled as an analogue of a covariance function in kriging. The main aim of using kernels is to tighten up the prediction interval without a loss in confidence. Another way to boost the efficiency of an RRCM prediction is to adjust the ridge parameter in the regression equation.

Chapter 2

Objectives

This dissertation has two major objectives:

1. to demonstrate the capacity of conformal predictors as a method for spatial environmental modeling,
2. to provide valid estimates of nitrogen dioxide and fine particulate matter for Barcelona Metropolitan Region.

In our modern environment, especially in urban areas, there is a great need to evaluate the levels of personal exposure to air pollution. Numerous studies have shown that contaminants contribute to adverse health effects in individuals both in long-term and short term exposure [40, 17, 12, 14, 20, 26, 15]. Continuous growth of knowledge regarding air pollution and its impact on health causes continuous development of methodology for its estimation. Various approaches, including geostatistical, have been used in research to establish the actual levels pollution people are exposed to [7]. Some methods are more sophisticated than the other ones, they require more entry data and skilled staff and equipment, and therefore they can yield better estimates. All in all, the choice of method to implement depends on the data and resources available.

The present study evaluates the concentrations of nitrogen dioxide and airborne particles in the Barcelona Metropolitan Region. The research makes use of panel data, which includes mean annual levels of concentrations and geographical coordinates of the measurement stations for several years. The data for this study has been kindly provided by XVPCA of the Generalitat of Catalonia. Geostatistical approach is the suitable one to process such data. This study takes up a specific method named ordinary kriging for its purposes. However, small size of the given data set together with some limitations of the method may lead to lack of validity in the estimates. To cope with this problem, this research suggests to consider the recently

developed machine learning method called conformal predictors [1].

Conformal predictors always provide valid estimates, and their flexibility allows to derive a predictor on the basis of almost any underlying algorithm, e.g. kriging. Thus, this dissertation makes use of a conformal predictor similar to kriging that produces valid prediction intervals with a given level of confidence. In other words, for every chosen level of confidence, this predictor would come up with a set of values and guarantee that the actual value of the estimated concentration falls within this set with this chosen confidence. The objective, of course, is to make this set as small as possible, i.e. to make it effective. Ways to achieve efficiency in prediction are disclosed in this dissertation.

Unfortunately, the data set is small with a great share of missing values, and no land-use variables have been available for this study. However, for each year, it was desirable to come up with a prediction pollution surface that would optimally fit the actual data.

Chapter 3

Methods and data

3.1 Kriging

3.1.1 General definitions

Kriging is a spatial interpolation method. It converts the data into an estimate of the spatial random field together with the measure of uncertainty [4]. This measure of uncertainty is called *kriging variance* (or *error variance*). Kriging was first introduced by a South African engineer D. G. Krige in the work that has been devoted to estimation of a mineral ore body [32, 4]. The method has been first described in 1951 [41], but it has been further developed and formalized by a French mathematician G. Matheron [42]. In fact, “kriging” is a generic name for a whole group of spatial interpolation methods, and the particular cases include: simple kriging, ordinary kriging, co-kriging, Bayesian kriging etc. In its simplest form, a kriging estimate of the data at an unobserved location is a linear combination of the data at observed locations of the spatial field [4]. Coefficients of the kriging equation, and the estimation error depend on spatial configuration of the data, and on spatial correlation. The least stands for the degree to which data at one location can be a predictor for data at another location as a function of spatial separation.

In order to measure spatial dependence, there has been introduced a special term - *variogram*. A variogram (that is sometimes referred to as *semivariogram*) is the cornerstone of geostatistics. It is introduced as follows: suppose $Z(x) : x \in \mathbf{D}$ is a stochastic process on \mathbf{D} , where \mathbf{D} is a spatial domain. Then

$$2\gamma(h) = \text{var}(Z(x) - Z(x+h)) \quad \forall x, x+h \in \mathbf{D} \quad (3.1)$$

is the variogram, and it depends only on the distance between points, h [43]. A variogram is a common measure of spatial dependence. (3.1) can be

replaced with a stronger assumption:

$$\text{cov}(Z(x+h), Z(x)) = C(h) \quad \forall x, x+h \in \mathbf{D}, \quad (3.2)$$

where $C(h)$ denotes the *covariance function*. If the mean function is assumed constant:

$$E(Z(x)) = \mu \quad \forall x \in \mathbf{D} \quad (3.3)$$

then (3.2) and (3.3) define a class of second-order stationary processes on \mathbf{D} . Conditions (3.1) and (3.3) define intrinsically stationary processes, which form a wider class. A variogram and a covariance function are related as follows:

$$\gamma(h) = C(0) - C(h). \quad (3.4)$$

A covariance function is positive definite, and it is bounded by the value:

$$|C(h)| \leq C(0) = \text{var}(Z(x)). \quad (3.5)$$

A variogram is not always bounded, and that is why a variogram can be deduced from a covariance function, and the reverse is not always true [5].

There exist a variety of covariance functions that are used in practice [5]. They assume that the spatial process is isotropic (rotation invariant). The *nugget-effect function* serves to model a discontinuity at the origin of the variogram, and it is defined as follows:

$$C(h) = \begin{cases} c_0 & \text{for } |h| = 0 \\ 0 & \text{for } |h| > 0 \end{cases} \quad (3.6)$$

where $c_0 > 0$ is the *nugget effect*.

One of the most frequently used is the exponential covariance function:

$$C(h) = c_0 * \exp\left(-\frac{|h|}{\sigma}\right), \quad (3.7)$$

where σ is a range parameter, and h is a distance between two spatial points. Another one is the Gaussian covariance function:

$$C(h) = c_0 * \exp\left(-\left(\frac{|h|}{\sigma}\right)^2\right). \quad (3.8)$$

Both exponential and Gaussian covariance functions are special cases of Matérn family [32] model. It is defined as follows:

$$C(h) = C(0)\{2^{\kappa-1}\Gamma(\kappa)\}^{-1} \left(\frac{h}{\sigma}\right)^{\kappa} K_{\kappa} \left(\frac{h}{\sigma}\right) \quad (3.9)$$

where $\kappa > 0$ is a parameter called *order*, and it determines the smoothness of the underlying process $Z(x)$, K_{κ} denotes a modified Bessel function of order κ , and σ is a range parameter. For $\kappa = 0.5$, the Matérn function reduces to the exponential one, and with $\kappa \rightarrow \infty$ - to the Gaussian covariance function.

Both exponential and Gaussian covariances are used in the present research. A linear covariance function [44] has been also tested out for variogram modeling.

3.1.2 Ordinary kriging

The most commonly used kriging type is ordinary kriging [5]. This method has been also applied in the study of this dissertation. The ordinary kriging is specified in the following way: assume that $Z(x)$ is a realization of an intrinsic stochastic process, and it has a (semi)variogram $\gamma(h)$. Let there be n neighboring locations, x_1, \dots, x_n , and an unobserved location x_0 . Then, the kriging estimate of the value of $Z(x)$ at the point x_0 will be:

$$Z_{OK}^*(x_0) = \sum_{\alpha=1}^n \omega_{\alpha} Z(x_{\alpha}). \quad (3.10)$$

Here, the ω_{α} are the *kriging weights*. A kriging estimate is unbiased, furthermore, it is a *BLUE* (*best linear unbiased*) estimate. The unbiasedness of the ordinary kriging estimate is guaranteed by the following constraint on weights:

$$\sum_{\alpha=1}^n \omega_{\alpha} = 1. \quad (3.11)$$

As known, the main strength of kriging is that it provides an estimate together with a measure of uncertainty, that is estimation, or kriging, variance. The estimation variance for ordinary kriging is the variance:

$$\sigma_E^2 = \text{var}(Z^*(x_0) - Z(x_0)), \quad (3.12)$$

and a weight ω_0 is equal to -1 , so:

$$\sum_{\alpha=0}^n \omega_{\alpha} = 0. \quad (3.13)$$

The estimation variance is:

$$\sigma_E^2 = -\gamma(x_0 - x_0) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_\alpha \omega_\beta \gamma(x_\alpha - x_\beta) + 2 \sum_{\alpha=1}^n \omega_\alpha \gamma(x_\alpha - x_0). \quad (3.14)$$

The ordinary kriging system can be obtained by minimizing the estimation variance. It looks as follows:

$$\begin{cases} \sum_{\beta=0}^n \omega_\beta^{OK} \gamma(x_\alpha - x_\beta) + \lambda_{OK} = \gamma(x_\alpha - x_\beta) & \text{for } \alpha = 1 \dots n \\ \sum_{\beta=0}^n \omega_\beta^{OK} = 1. \end{cases} \quad (3.15)$$

Here, ω_β^{OK} are the ordinary kriging weights, and λ_{OK} is a Lagrange multiplier. Therefore, the ordinary kriging estimation variance takes form:

$$\sigma_{OK}^2 = \lambda_{OK} - \gamma(x_0 - x_0) + \sum_{\alpha=1}^n \omega_\alpha^{OK} \gamma(x_\alpha - x_0). \quad (3.16)$$

For ordinary kriging, confidence intervals can be obtained in a standard manner. Assuming that the data come from a Gaussian spatial process, a 95 % prediction intervals can be constructed as follows [44]:

$$A = \left(\widehat{Z}(x_0) - 1.96 * \sigma_{OK}(x_0), \widehat{Z}(x_0) + 1.96 * \sigma_{OK}(x_0) \right). \quad (3.17)$$

It is noteworthy that ordinary kriging is an exact interpolator. This means that if the unobserved location x_0 coincides with any of the data locations $x_\alpha : \alpha = \overline{1, n}$, then the value of the ordinary kriging estimate will coincide with the data value at that point:

$$Z_{OK}^*(x_0) = Z(x_\alpha), \quad \text{if } x_0 = x_\alpha, \alpha = \overline{1, n}. \quad (3.18)$$

3.2 Conformal predictors

3.2.1 General definitions

Conformal predictors is a name for an approach coming from the statistical learning theory [1]. It allows to obtain valid predictions with a given level of confidence. In other terms, conformal predictors are “*confidence predictors*”. In order to obtain a confidence prediction, one should specify a level of error probability for it, otherwise called a *significance level*. Assuming the error probability of 1, 5, or 10 %, the confidence level will respectively make up 99, 95 or 90 %. The of confidence would denote a highly confident prediction, a 90 per cent level would designate a confident prediction, and a, say, 20 per cent level would correspond to a casual prediction.

There are two important features of a confidence predictor:

1. It should be valid, which means that in the long run the frequency of error of prediction shall not exceed the chosen error probability;
2. It should be efficient, which means that the prediction set that is output is as small as possible.

As far as validity is concerned, conformal predictors are always valid, by definition. When it comes to efficiency, it must be said that in regression problems the a prediction set usually takes form of an interval. Thus, the efficiency of a conformal predictor in a regression setting would imply that the interval should have minimum length.

After all, the main feature of conformal predictors is their extreme flexibility. This implies that a conformal predictor can be built upon almost any machine learning (and statistical) algorithm: regression, support vector machines, bootstrap, decision trees, Bayesian methods etc. Such a predictor has the same predictive performance as its underlying algorithm, but, provided with an ad hoc confidence level, it is also a valid predictor.

Let's assume that there are given the pairs of observations of (x_i, y_i) where x_i will be called an *object* and y_i will be a *label*. To put it in a straightforward way, x_i can stand for an air pollution measurement station coordinates, and y_i - for a corresponding concentration for given pollutant and year. The space \mathbf{X} will then be named *object space*, and \mathbf{Y} - will be the *label space*. \mathbf{Z} defined as:

$$\mathbf{Z} := \mathbf{X} \times \mathbf{Y} \quad (3.19)$$

is then the *example space*. \mathbf{Z}^∞ is a measurable space, and the infinite data sequence $(x_1, y_1), (x_2, y_2), \dots$ belongs to it. Then, we believe that this infinite data sequence has a distribution P , which is *exchangeable*, i.e. for every positive integer n , every permutation π of $1, \dots, n$ and every measurable set $E \subseteq \mathbf{Z}^n$:

$$\begin{aligned} P(z_1, z_2, \dots) \in \mathbf{Z}^\infty : (z_1, \dots, z_n) \in E &= \\ = P(z_1, z_2, \dots) \in \mathbf{Z}^\infty : (z_{\pi(1)}, \dots, z_{\pi(n)}) \in E. & \end{aligned} \quad (3.20)$$

Suppose there is a sequence $(x_1, y_1), (x_2, y_2), \dots, (x_{n-1}, y_{n-1})$, and the aim is to predict the label y_n for the object x_n . For that purpose, the following function will serve:

$$D : \mathbf{Z}^* \times \mathbf{X} \rightarrow \mathbf{Y}. \quad (3.21)$$

This function, under the assumption that it is measurable, can be called a *simple predictor*. Here, $x_1, y_1, x_2, y_2, \dots; x_{n-1} \in \mathbf{Z}^*$, and the value of the

y_n will be derived as follows: $y_n = D(x_1, y_1, x_2, y_2, \dots; x_{n-1}), Y_n \in \mathbf{Y}$.

If we refer to air pollution stochastic interpolation, kriging can be classified as a simple predictor. It will output only one possible label y_n , i.e. it will be a point prediction.

However, there is a more sophisticated approach to prediction. Imagine we allow the predictor output subsets of \mathbf{Y} large enough to provide the confidence of the predictor. That will mean, that we will be confident in that the true value of the y_n falls within this subset. This approach would require an additional measure $\epsilon \in (0, 1)$ referred to as *significance level*, and a complementary measure $1 - \epsilon$ is called *confidence level*. For each significance level, the corresponding level of confidence of prediction is a desiderata, or that the probability of error in prediction is equal to ϵ .

A conformal predictor is a confidence predictor defined by some *nonconformity measure*. Given an object x_n and a significance level ϵ , the predictor provides a prediction set that should contain the actual value of y_n . A predictor obtains this set on the basis of the assumption that the observation (x_n, y_n) conforms with the rest of observations. The level of significance shows the degree of conformity.

3.2.2 Ridge regression procedure

The ridge regression method has been first suggested by A. E. Hoerl [45, 3]. This procedure is aimed to tackle the problem arising when the predictor vectors in the matrix X of the independent variables are far from being orthogonal. The correlation in predictors can cause the situation when the matrix $X'X$ is close to singular [45]: The least, in its turn, can lead to unstable parameter estimates. In order to cope with this problem, a small scalar coefficient introduced into the regression equation. This is done in the following way. Suppose Z is the matrix of predictors X , but scaled and centralized. Let the matrix X , and consequently, Z , be of size $n \times p$, so there are n observations and p parameters (predictor vectors). Y then would stand for the $1 \times n$ vector of dependent variables. Then, the ridge regression estimates are obtained as follows:

$$\omega(a) = (Z'Z + aI_p)^{-1}Z'Y, \quad (3.22)$$

where a is the *ridge parameter*, and I is the identity matrix. In applications, the values of a usually fall within the interval $(0, 1)$. By the present day, there is no analytical procedure known aimed to choose the optimal value of a . In order to do so, brute force method is normally employed, or method of successive approximations [46]. At a certain value of a the system

stabilizes and has the general characteristics of the orthogonal system [45, 3], and the regression coefficients will have proper signs, and the residual sum of squares will not be too large. When the ridge factor is equal to zero, the ridge regression estimate coincides with the least squares estimate: in other words, least squares is a special case of the ridge regression.

$$\omega(0) = (Z'Z)^{-1}Z'Y. \quad (3.23)$$

The ridge regression estimator can be expressed in terms of the least squares estimator [45]:

$$\omega(a) = (I_p + a(Z'Z)^{-1})^{-1}\omega(0) = Q\omega(0), \quad (3.24)$$

where Q is the sample covariance matrix. Thus, the ridge estimators are all linear combination of least regression estimator with coefficients given by the matrix:

$$(I_p + a(Z'Z)^{-1})^{-1}. \quad (3.25)$$

The plot of the components of the vector $\omega(a)$ against a is called *ridge trace*. With the increase of ridge parameter, the estimates become smaller in their absolute value, and they tend to zero as a tends to infinity. Ridge trace helps select the optimal value of ridge factor, because it provides visual representation of how changes in the values of a affect the estimates.

When the optimal value of the ridge factor, a^* , is *selected*, the ridge estimates $\omega(a^*)$ are obtained. Those estimates are not least squares, and they *are biased*, but they are more stable, they have the correct sign and reasonable values, and they provide a smaller mean square error. The bias is explained by the presence of the ridge parameter, a , in the equation.

3.2.3 Ridge regression confidence machine

Ridge regression confidence machine (RRCM) is a conformal predictor suggested by Nouretdinov [2], and it makes use of the ridge regression procedure as an underlying algorithm [1]. Suppose X_n is the $n \times p$ matrix of objects (or, independent variables), and Y_n is the vector of labels (dependent variables). The nonconformity score for this predictor is represented by the absolute value of the residuals: $e_i := y_i - \hat{y}_i$, where $i = \overline{1, n}$. Namely, the ridge regression estimate of the parameters ω will be written down as:

$$\omega = (X_n'X_n + aI_p)^{-1}X_n'Y_n. \quad (3.26)$$

It is noteworthy here, that normalization and centralization of the matrix X is not compulsory, yet desirable. It helps overcome the possible

multicollinearity problem, and it facilitates further computations to a great extent. Since the computing capacities are bounded, the problems may if the data is used in its initial representation. This may occur because of various reasons, for example, the difference of scaling, or just “huge” numbers composing a matrix, and making it impossible to be inverted.

The predictions \hat{y}_i for the objects x_i are given by the equation:

$$\widehat{Y}_n = (\hat{y}_1, \dots, \hat{y}_n)' = X_n(X_n'X_n + aI_p)^{-1}X_n'Y_n. \quad (3.27)$$

The matrix:

$$H_n = X_n(X_n'X_n + aI_p)^{-1}X_n' \quad (3.28)$$

is named the *hat matrix*, because it transforms y_i to \hat{y}_i . (3.28) is a symmetric and idempotent matrix when $a = 0$, together with the matrix $I_n - H_n$. Thus, the vector of regression residuals (or nonconformity scores) can be written as follows:

$$(|e_1|, \dots, |e_n|)' = |(I_n - H_n)Y_n|. \quad (3.29)$$

Now suppose we have the incomplete data sequence:

$$x_1, y_1, x_2, y_2, \dots, x_{n-1}, y_{n-1}, x_n, \quad (3.30)$$

and the aim is to predict the value of y_n . A significance level for prediction must be introduced, let it be equal to ϵ . Let y be a possible label for x_n , and then $Y := (y_1, \dots, y_{n-1}, y)$. The vector Y can be represented as splitter into two parts: $Y = (y_1, \dots, y_{n-1}, 0)' + (0, \dots, 0, y)'$, and the vector of the regression residuals (or, nonconformity scores) will take form $|A + By|$ where:

$$A = (I_n - H_n)(y_1, \dots, y_{n-1}, 0)'$$

and

$$B = (I_n - H_n)(0, \dots, 0, 1)'.$$

The p-value of the prediction, $p(y)$, changes sign only at points some points, this is why a set of points y can be calculated, such that the p-values, $p(y)$, exceed the given significance level, ϵ . For each $i = \overline{1, n}$, let:

$$S_i := \{y : \alpha_i(y) \geq \alpha_n(y)\} = \{y : |a_i + b_i y| \geq |a_n + b_n y|\},$$

where a_i and b_i are the components of the vectors A and B . Each set S_i can be either a real line, a ray, a union of two rays, an interval, a point, or empty. For example, if S_i is an interval, then, if $b_i \neq b_n$, $\alpha_i(y)$ and $\alpha_n(y)$ are equal in two points:

$$-\frac{a_i - a_n}{b_i - b_n} \quad \text{and} \quad -\frac{a_i + a_n}{b_i + b_n}. \quad (3.31)$$

Those points may coincide, and S_i may be a point. In order to calculate the p-value for y , it must be counted, how many S_i include y , and then divide it by n :

$$p(y) = \frac{|\{i = \overline{1, n} : y \in S_i\}|}{n}. \quad (3.32)$$

As y increases, the p-value $p(y)$ changes only at points (3.31), so for any significance level ϵ there can be found a union of a finite number of S_i . This will yield the final prediction set. The algorithm works as follows: the predictor arranges the points (3.31) into an increasing sequence $y_{(1)}, \dots, y_{(m)}$. Then it adds two points: $y_{(0)} := \infty$ and $y_{(\infty)} := \infty$ on the ends of the set. Then it computes $N(j)$, the number of i such that $(y_{(j)}, y_{(j+1)}) \subseteq S_i$ for $j = 0, \dots, m$, and $M(j)$, the number of i such that $y_{(j)} \in S_i$ for $j = 0, \dots, m$ [1]. A small set of the chosen significance levels, $\epsilon_k, k = 1, \dots, K$, is provided to the algorithm. The predictor outputs the corresponding nested family of prediction sets:

$$\Gamma_n^{\epsilon_k}(x_1, y_1, \dots, x_{n-1}, y_{n-1}, y) \quad (3.33)$$

for $k = 1, \dots, K$. For each k , $\Gamma_n^{\epsilon_k}$ is equal to the set of all labels $y \in Y$ such that:

$$\frac{|\{i = 1, \dots, n : \alpha_i \geq \alpha_n\}|}{n} > \epsilon. \quad (3.34)$$

3.2.4 Dual form ridge regression confidence machine

When the number of parameters is large, the ridge regression procedure can hardly handle the computation of the data, since it implies inverting the $p \times p$ matrix. In order to deal with high-dimensional data, Boser, Guyon and Vapnik have suggested a so-called “kernel trick” [6]. This method has been implemented to non-linear support vector machines, and here it can be used to introduce non-linear ridge regression procedure. The input space \mathbf{X} of objects (or, independent variables) X_n is mapped into another space \mathbf{H} , called *feature space* [1], or *linearization space* [6]. The substitution is done as follows: for $x^{(1)}, x^{(2)} \in \mathbf{X}$; $h^{(1)}, h^{(2)} \in \mathbf{H}$

$$k(x^{(1)}, x^{(2)}) = \langle h^{(1)}, h^{(2)} \rangle = F(x^{(1)}) \dots F(x^{(2)}), \quad (3.35)$$

where $k(\cdot, \cdot)$ is a kernel, and $\langle \cdot, \cdot \rangle$ denotes a scalar product. Thus, the dependency between the elements of the original (or, input) space is described through scalar product of the elements of the feature space.

In order to apply “kernel trick”, ridge regression confidence machine should undergo a specific transformation. The ridge regression procedure must be written in the “dual form”:

$$X_n(X_n'X_n + aI_p)^{-1} = (X_nX_n' + aI_n)^{-1}X_n, \quad (3.36)$$

or

$$(X_n'X_n + aI_p)^{-1}X_n' = X_n'(X_nX_n' + aI_n)^{-1}. \quad (3.37)$$

Thus, the ridge regression prediction (\hat{y}) for an object x based on examples $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ will be:

$$\hat{y} = Y_n'(X_nX_n' + aI_n)^{-1}X_n \quad (3.38)$$

The hallmark of this representation is that it depends of the objects x_1, \dots, x_n, x only via scalar products between them. If the object space \mathbf{X} is mapped into Euclidian feature space \mathbf{H} , $F : \mathbf{X} \rightarrow \mathbf{H}$, and *ridge regression is performed in the feature space*, the prediction (3.38) can be rewritten as:

$$\hat{y} = Y_n'(K_n + aI_n)^{-1}k_n, \quad (3.39)$$

where K_n is the matrix of the elements $(K_n)_{i,j} = k(x_i, x_j)$, k_n is the vector with the elements $(k_n)_i = k(x, x_i)$, and $k(\cdot, \cdot)$ is a kernel. The hat matrix (3.28) in the dual representation is written as:

$$H_n = X_n(X_n'X_n + aI_p)^{-1}X_n' = (X_nX_n' + aI_n)^{-1}X_nX_n'. \quad (3.40)$$

When ridge regression is performed in the feature space, it takes form:

$$H_n = (K_n + aI_n)^{-1}K_n, \quad (3.41)$$

and now, in order to perform the RRCM in a kernel form, or a non-linear RRCM, one only would need to substitute the hat matrix in the primary setting RRCM algorithm (and define a kernel).

3.2.5 Kernels

There is an infinite amount of kernels out there. Two most important features of a kernel is that it is symmetric:

$$k(x^{(1)}, x^{(2)}) = k(x^{(2)}, x^{(1)}), \quad \forall x^{(1)}, x^{(2)} \in \mathbf{X}, \quad (3.42)$$

and nonnegative definite:

$$\sum_{i=1}^m \sum_{j=1}^m K(x^{(i)}, x^{(j)}) a_i a_j \geq 0,$$

$$\forall x^{(1)}, \dots, x^{(m)} \in \mathbf{X}, \forall a_1, \dots, a_m \in \mathbb{R}.$$

It is also correct that any function $\mathcal{K} : \mathbf{X}^2 \rightarrow \mathbb{R}$ can be represented in the form of a kernel [1]:

$$\mathcal{K}(x^{(1)}, x^{(2)}) = F(x^{(1)}, x^{(2)}). \quad (3.43)$$

This research makes use of two popular kernels - apart from the scalar product in the linear model. These kernels are the *Gaussian radial basis function* (RBF kernel) and the *polynomial kernel* of the second order. These are positive definite kernels. The justification for use of kernels is the following: in literature, a covariance function is one of the terms used to describe positive definite kernels [6]. The RBF kernel has been chosen here in order to match the Gaussian covariance function in kriging prediction. The polynomial kernel of the second order has been chosen with no special justification. It has been guessed that the appropriate covariance function is probably not linear. A better kernel can be further found, for sure.

The RBF kernel is specified as follows [6]:

$$k(x^{(1)}, x^{(2)}) = \exp\left(\frac{\|x^{(1)} - x^{(2)}\|^2}{2\sigma^2}\right), \quad (3.44)$$

where $\sigma > 0$ is a scale parameter. As for polynomial kernels, there are distinguished two types of them: *homogeneous* polynomial kernels:

$$k(x^{(1)}, x^{(2)}) = \langle x^{(1)}, x^{(2)} \rangle^d, \quad (3.45)$$

where d is a degree (order), and *inhomogeneous* polynomial kernels:

$$k(x^{(1)}, x^{(2)}) = \left(\langle x^{(1)}, x^{(2)} \rangle^d + c\right)^d,$$

where $c > 0$ is a constant. This particular research makes use of an inhomogeneous polynomial kernel of the second degree, and it is specified as [1]:

$$k(x^{(1)}, x^{(2)}) = \left(1 + x^{(1)} \cdot x^{(2)}\right)^2.$$

3.3 Computing

All the computational work for the present dissertation has been made with the use of R statistical software [47]. The kriging calculations have been performed employing the **geoR** package [48, 32], which embraces all the essential tools for geostatistical analysis and modeling. As for RRCM modeling, the package **PredictiveRegression** [38] has been employed. This package has been developed to provide the computational support to the on-line predictive version of the common problem of linear regression. In this research, the function *iidpred* has been used in order to perform linear RRCM estimation for the data. Then, in order to consider a non-linear RRCM predictor, namely, to perform the “kernel trick”, the function *iidpred* has been rewritten by the author of the present dissertation so, that it would match the RRCM in the dual setting. This function has been further modified as involving kernels considered in this research, and two functions appeared as a result:

- function *iid_rbf* serves to compute ridge regression confidence machine estimates with the RBF kernel,
- function *polyn2* serves to compute the ridge regression confidence machine estimates with the polynomial kernel of the second degree.

For code of the functions, please see the **Appendix A** to the dissertation.

It is noteworthy that prior to application of both approaches, ordinary kriging and ridge regression, the data has been scaled and centralized, as ridge regression procedure suggests [45]. The scaling has been done in order to be able to run the ridge regression procedure correctly, since it involves matrix inversion, and also to overcome the difference in measurement. The independent variables of both regression procedures, kriging and ridge regression, are the geographical longitude and latitude, and not only they are given in quite a big numbers, but also there is a difference in scaling of one order (10 times) beet wen them. Both longitude and latitude have been transformed into unit scale in the following way:

$$\widehat{x^{(i)}} = \frac{x^{(i)} - \bar{x}^{(i)}}{x_{max}^{(i)} - x_{min}^{(i)}}, \quad i = 1, 2, \quad (3.46)$$

Data transformed in such a manner has been used used both for kriging and ridge regression - for accurate comparison of outputs of both models. Although, using the original units in kriging would not affect the estimation results.

3.4 Data

3.4.1 Barcelona Metropolitan Region

Barcelona Metropolitan Region (BMR) is situated on the north-east of Spain. Spread along the Mediterranean coast, it has a territory of about 3200 km^2 and accommodates over 5 million inhabitants [49]. In 1986, there has started a process of economic and geographic expansion, which has converted the region into one of the 10 biggest urban agglomerations of Europe [50]. In Spain, it is the second largest metropolitan area, after Madrid. The BMR comprised 90 municipalities back in 1986. It has grown up to 220 municipalities in 2011, and it has consequently seen the growth of the population: from 3.58 million inhabitants in 1986 to 5.4 million in 2011. The actual population of the BMR makes up 91 per cent of the whole population of the Province of Barcelona; the Region houses about 195.000 enterprises, which is a 91-percent share of the whole province as well. As far as the administrative subdivision is concerned, the BMR consists of: the city of Barcelona surrounded by the Barcelona Metropolitan Area (BAM), encompassing 35 municipalities, and the rest of the BMR, embracing 128 other municipalities. Sometimes, they refer to BMR simply as the Barcelona Metropolitan Area which might provoke confusions. The city of Barcelona is the economic centre of the BMR, as well as of the whole Catalonia. It is inhabited by 1.6 million people, and embodies 44 per cent of the whole volume of jobs of the province. Badalona and L'Hospitalet de Llobregat are other important cities within the region, together with Mataró, Granollers, Sabadell, Terrassa and Vilanova i la Geltrú, which have played the role of the industrial centers since the 19th century.

Every week, two third part of the whole amount of movements that take place in Catalonia, occur in BMR, which is proportional to the share of inhabitants and to the concentration of economic activities [49]. In BMR, there happen about 107 million displacements weekly, with averagely 16.8 moves during a work day and 11.2 dislocations during a day-off. 54.1 per cent of total weekly displacements happen by means of motorized transport: both public and private [51]. The share of private engine-powered vehicles alone is equal to 35.6 percent, or 38.8 million weekly moves.

3.4.2 The present data set

It is a well-known fact that traffic is the main contributor to air pollution in urban surroundings. The present research makes use of the data representing the concentrations of two air pollutants, nitrogen dioxide (NO_2) and particulate matter (PM_{10}), at the measurement cites spread over the Barcelona Metropolitan Regions. The data for this research has been kindly provided

by XVPCA (Network for Monitoring and Forecasting of Air Pollution) of the Generalitat of Catalonia [19]. The dataset comprises observations made in 27 municipalities of the BMR, that are: Barcelona, Badalona, Barberà de Vallès, Castellbisbal, Cornellà de Llobregat, El Papiol, El Prat de Llobregat, Esplugues, Gavà, Granollers, L'Hospitalet de Llobregat, Martorell, Molins de Rei, Mollet, Montcada i Reixac, Ripollet, Rubi, Sabadell, Sant Adrià de Besòs, Sant Andreu de la Barca, Sant Cugat de Vallès, Sant Feliu de Llobregat, Sant Vicenç dels Horts, Santa Coloma de Gramenet, Santa Perpètua de Mogoda, Terrassa and Vilanova i la Geltrú (see Table 2.1 and Figure 2.1)

Table 3.1: Distribution of measurement sites over BMR

municipality	stations	pollutants
Barcelona	13	NO ₂ , PM ₁₀
Badalona	1	NO ₂ , PM ₁₀
Barberà de Vallès	1	NO ₂ , PM ₁₀
Castellbisbal	2	PM ₁₀
Cornellà de Llobregat	1	NO ₂
El Papiol	1	PM ₁₀
El Prat de Llobregat	1	NO ₂ , PM ₁₀
Esplugues	1	NO ₂
Gavà	3	NO ₂ , PM ₁₀
Granollers	2	NO ₂ , PM ₁₀
L'Hospitalet de Llobregat	1	NO ₂ , PM ₁₀
Martorell	1	NO ₂ , PM ₁₀
Molins de Rei	1	PM ₁₀
Mollet	1	NO ₂ , PM ₁₀
Montcada i Reixac	1	NO ₂ , PM ₁₀
Ripollet	1	total suspended particles
Rubi	3	NO ₂ , PM ₁₀
Sabadell	3	NO ₂ , PM ₁₀
Sant Adrià de Besòs	1	
Sant Andreu de la Barca	1	NO ₂ , PM ₁₀
Sant Cugat de Vallès	1	NO ₂ , PM ₁₀
Sant Feliu de Llobregat	1	PM ₁₀
Sant Vicenç dels Horts	2	NO ₂ , PM ₁₀
Santa Coloma de Gramenet	1	NO ₂ , PM ₁₀
Santa Perpètua de Mogoda	1	NO ₂ , PM ₁₀
Terrassa	1	NO ₂ , PM ₁₀
Vilanova i la Geltrú	1	NO ₂ , PM ₁₀

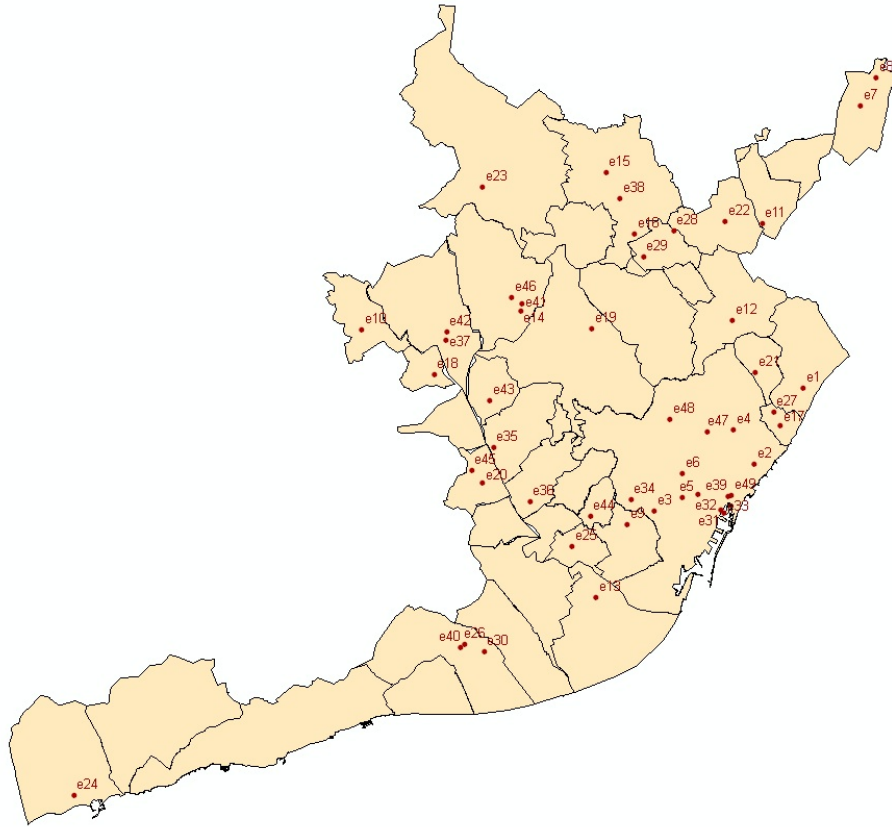


Figure 3.1: Distribution of stations over BMR

The time scope for the study is the following: for NO_2 observations for the years 1998-2009 were available - with the absence of observations for the year 2003, and for PM_{10} the timeline was from 2001 to 2009, again with no data for the year 2003. Both pollutants were measured on the 49 monitoring sites mentioned above on the daily basis. Then the mean annual concentration were obtained with a geographic information system. Thus, in the data set the pollutants are represented in the form of mean annual concentrations together with the geographic coordinates of the measurement sites, their labels, the names of municipalities where they are situated, and the year the observations correspond to. This is a pooled data set. Unfortunately, the data was not available for each year and station for neither of pollutants. With 49 measurement stations all together, there were approximately 24 valid observations per year for NO_2 , and 29 for PM_{10} . For each given year, the concentrations of NO_2 were provided at 22 to 25 stations (see Table 3.2).

Table 3.2: Data on mean annual nitrogen dioxide concentrations

Available observations for each year										
1998	1999	2000	2001	2002	2004	2005	2006	2007	2008	2009
24	25	25	25	25	24	22	24	25	25	24

Table 3.3 demonstrates how many observations of PM_{10} concentrations were available for each of the study years.

Table 3.3: Data on mean annual particulate matter concentrations

Available observations for each year							
2001	2002	2004	2005	2006	2007	2008	2009
22	24	28	28	29	30	33	36

Thus, there can be named two major drawbacks, or limiting factors, of the data set:

- **Size:** there was a small number of observations for each year and pollutant,
- **Distribution:** the measurement spots are situated quite far apart from one another, and they are distributed, or placed, unevenly over the geographic region.

Those two limitations sufficiently reduce the scope of the study, since it implies the interpolation of the given data in order to obtain a “good” estimate at any unobserved location over the study region.

Also, a few words should be said about the reliability of the data. Both nitrogen dioxide and particulate matter concentrations taken up in this study are mean annual concentrations. Initially, the data has been measured hourly, and these measurements have been further averaged. Only these average observed concentrations were available for this study, while the initial hourly concentrations were not. Mean annual measurements may be uncertain, because they do not allow to make unambiguous conclusions about the data distribution. Because of the lack of observations and their low frequency, the shape and the parameters of the data distribution cannot be established precisely, and thus the inference regarding the existence or absence, or character of spatial covariance between observations might be

misleading. Not only this may affect the modeling prediction capacity, but also it can reduce the possibility for validation of the models.

Table 3.4 shows the descriptive statistics for the observed nitrogen dioxide concentrations. Those concentrations are the annual averages. For each year of the study timeline, Table 3.4 depicts the following values: minimum, first quartile of the distribution, median, mean, third quartile of the distribution, maximum, and the percentage of missing values. If the data set for nitrogen dioxide would be complete, the number of observations for each year would be equal to 49, as this is the amount of measurement stations at the Barcelona Metropolitan Region where the nitrogen dioxide pollution has been recorded. However, the data set is not complete, and the Table 3.2 shows the number of valid observations available for each year for evaluation. The right column of the Table 3.4 reflect the percentage that the missing values for each year make up of the total of 49. It is seen that for each year, on the average, at half of the stations the concentrations were not observed.

Table 3.4: Annual nitrogen dioxide concentrations ($\mu\text{g}/\text{m}^3$). Descriptive statistics

year	min	1st q.	median	mean	3rd q.	max	missing
1998	11	35.75	47	46.50	61.00	68	51.02
1999	17	32.00	45	44.40	57.00	71	48.98
2000	13	33.00	42	41.20	48.00	65	48.98
2001	23	39.00	48	46.16	54.00	64	48.98
2002	25	38.00	45	45.24	53.00	69	48.98
2004	22	33.50	38	39.88	45.00	67	51.02
2005	21	36.50	44	44.23	48.00	83	55.10
2006	19	36.50	46	43.96	50.25	74	51.02
2007	21	40.00	44	43.40	50.00	66	48.98
2008	18	39.00	42	42.36	47.00	65	48.98
2009	11	41.00	44	43.12	50.00	63	48.98

Table 3.5 shows the descriptive statistics for the observed particular matter concentrations. Those concentrations are also annual averages, too. The right column of the Table 3.5 shows the percentage of the missing values in the whole data set for each year. The amount of the valid observations of the PM_{10} observations for each year can be seen in Table 3.3.

Table 3.5: Annual particulate matter concentrations ($\mu\text{g}/\text{m}^3$). Descriptive statistics

year	min	1st q.	median	mean	3rd q.	max	missing
2001	33	40.25	43	47.82	55.25	75	55.10
2002	28	43.75	48	49.29	55.25	67	51.02
2004	26	38.50	45	44.36	50.50	57	42.86
2005	31	43.75	47	48.61	54.25	70	42.86
2006	29	45.00	48	49.24	53.00	69	40.82
2007	33	40.25	44	45.93	48.75	89	38.78
2008	29	36.00	38	39.94	43.00	63	32.65
2009	25	33.00	36	36.19	40.00	48	26.53

3.4.3 Previous research for BMR

It is noteworthy that a similar research aimed to predict the spatial concentrations of nitrogen dioxide and particles (PM_{10}) has previously been carried out [33]. This study has considered the same geographic region, that is Barcelona Metropolitan Region. There, spatial distribution for both of the pollutants has been obtained for the study region. Also, the concentrations have been predicted in centre of the census track for every day of the entire study period. These daily values have been further averaged. Moreover, this was the first evidence of using kriging for predicting the concentrations of air contaminants for this region. A preliminary data analysis has been carried out in this study, aimed to establish the existence or absence of spatial tendencies in the distribution of the pollutants. It has been done by means of linear regression. As far as kriging is concerned, the four spatial dependence structures have been taken up: exponential, spherical, Gaussian and Gneiting, and Bayesian inference has been used in modeling. The calculus have been made employing R, especially its **geoR** library [48]. The timeline for the study is the following: the years 1994-2004 for NO_2 , and the years 2004-2004 for PM_{10} . For both of the pollutants, the observations have been made at up to 9 stations.

The difference between the above mentioned study and the one presented in the current dissertation is not only denoted by somewhat dissimilar time scope and number of the measurement stations involved. When it comes to methodology and results, there are differences as well. The study of Saez and Lertxundi is very accurate when it comes to preliminary data analysis. Here, this initial exploration aimed to elicit spatial tendencies is not done intentionally. Here, instead, two types of spatial dependency have been assumed, i.e. two different covariance functions in kriging have been taken up,

and then the corresponding kernels were employed in the RRCM modeling.

The main goal of the current study is to introduce a new method for air pollution assessment, that is, conformal prediction on the basis of kriging, and testing this method out. The main aim was to implement and “debug” the method, rather than find a perfect fit for this particular data. This is, however, a plan for future research. It is planned to extend the current data set for this purpose.

By now, in order to “test and debug” the method, cross-validation analysis has been carried out. Regarding the fact, that 25 observations without missing values have been available for each year and pollutant, leave-one-out cross-validation has been considered [52].

Chapter 4

Results

4.1 Ordinary kriging with exponential covariance

4.1.1 NO₂

Model for 1999

Below, it will be shown below how the computation of the ordinary kriging with exponential covariance function has been performed and which results it has yielded. Linear covariance function has also been considered, but, due to relatively small distances, the approximation with linear variomodel is practically the same as with the exponential one. Here, the computations and the results are shown for year 1999. For the ease of comparison of the results, this year has been taken up for all further models for nitrogen dioxide. The detailed results for each year in the data set can be found in **Appendix B**.

For ordinary kriging computation, the function *krige.conv* from the R [47] package **geoR** [48] has been used. The initial data (the regressors, i.e. coordinates) has been scaled and normalized. This has been done to compare the results of the ordinary kriging model with the results of the further demonstrated RRCM model. The scaling of the coordinates has been performed following the formula (3.46).

The computation of the model is shown below. First, a *geodata* [32] object should be created.

```
> geo_no2_1999 <- as.geodata(no2_1999[1:25, ],
+   coords.col = 1:2, data.col = 3)
> locs_no2_1999 <- no2_1999[26:49, 1:2]
```

An exponential variogram model has been fitted:

```
> aae <- variofit(variog(geo_no2_1999), cov.model = "ex")
```

```

variog: computing omnidirectional variogram
variofit: covariance model used is exponential
      sigmasq phi      tausq kappa
initial.value "327.08" "0.91" "163.54" "0.5"
status        "est"   "est"   "est"   "fix"
loss value: 2171172.0506172

```

A linear variomodel has been fitted, too. This has been done to ascertain, which variomodel would make a better fit. Also, the kriging modeling results are to be further compared to the predictions of ridge regression confidence machine, which is linear and makes use of no kernels in its standard iid setting.

```
> aal <- variofit(variog(geo_no2_1999), cov.model = "li")
```

```

variog: computing omnidirectional variogram
variofit: covariance model used is linear
      sigmasq phi      tausq kappa
initial.value "327.08" "0"    "163.54" "0.5"
status        "est"   "est" "est"   "fix"
loss value: 2015475.3850871

```

Figure 4.1 shows how the theoretical variogram models fits the data. This is a two-dimensional plot where the values of the empirical variogram (or semivariogram) are plotted against the values of the Euclidian distance between points. Moreover, two curves for the fitted theoretical variogram models are added. On this plot, the two fitted model curves coincide. Thus, the exponential function will be used as it is considered by default for conventional kriging in the **geoR** package.

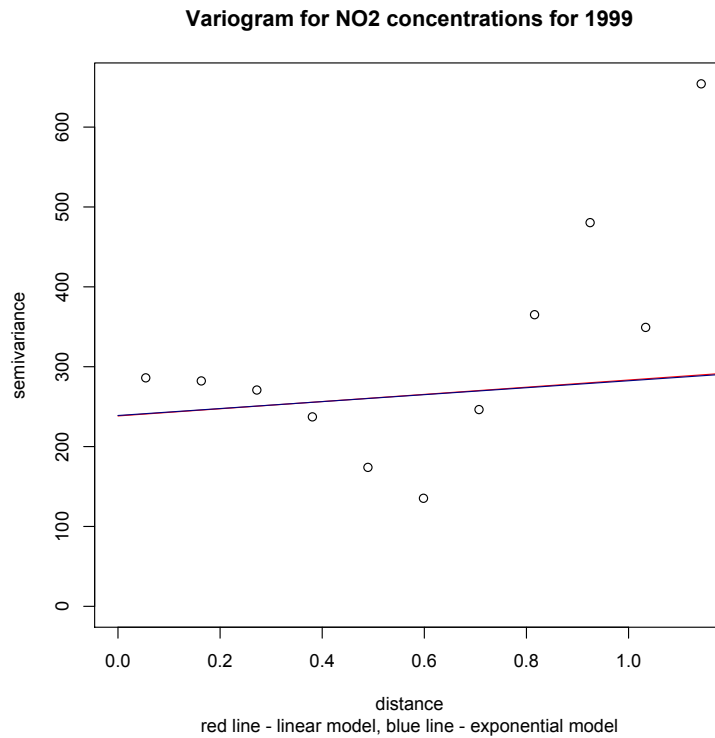


Figure 4.1: Empirical and modeled variograms

Running the ordinary kriging model with exponential covariance function:

```
> ok_no2_1999 <- krige.conv(geo_no2_1999,
+   locations = locs_no2_1999,
+   krige = krige.control(type.krige = "ok",
+   obj.model = aae))
```

Table 4.1 shows the results of the modeling. It is seen that the prediction turned out to be very smooth, as the predicted values vary from 42.57 ($\mu\text{g}/\text{m}^3$) to 45.30 ($\mu\text{g}/\text{m}^3$). Ascending order in the predicted values is a result of that the data has been sorted. Column 2 of the table shows the estimates for the kriging variances. They are very big, which is explained by the small number of the observed data.

Table 4.1: Ordinary kriging prediction results ($\mu\text{g}/\text{m}^3$)

NO2	kriging variance
42.57	260.75
42.60	260.86
43.19	272.10
43.25	271.72
43.57	261.27
43.40	271.61
43.14	259.15
43.57	259.81
43.17	258.36
43.21	257.77
44.12	260.17
44.45	260.94
44.56	258.78
44.60	258.73
44.86	257.04
44.78	255.15
44.80	256.88
45.57	256.87
45.11	255.23
45.63	258.53
45.63	258.85
45.68	257.95
45.69	257.98
45.30	256.52

Cross-validation of the model

Kriging is an exact interpolator, so the prediction has been obtained for the points where the data was not available, on the basis of the observed values. For 1999, the data was available at 25 points out of the total number of 49. So, for 24 spatial locations, the concentrations of nitrogen dioxide for 1999 were unavailable. They have been predicted with the use of ordinary kriging, as shown above. However, the measure of error, i.e. the kriging variance, is by itself an estimate, since the real observations are not given. This means there is no way to ascertain the goodness of fit of the model since one cannot compare the model output to reality. Nevertheless, the goodness of fit of the model can be tracked with the use of a cross-validation method. Regarding the number of observations available, the leave-one-out specification of cross-validation has been considered [53]. On each step, the value of

nitrogen dioxide concentrations at every location has been predicted on the basis of its values at the other observed locations. Table 4.2 shows that the absolute values of the errors of prediction are big, which is due to that the predictions are very smooth while the real data distribution is not so even. A wide range of the observed values of the nitrogen dioxide concentrations might be explained by general variability of the pollutant. A little range in the kriging estimates indicates that the exponential covariance model does not match the empirical variogram of the data very well. However, a better model can be barely derived for this data.

Table 4.2: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	predicted	absolute error
17.00	47.42	30.42
20.00	45.42	25.42
22.00	44.38	22.38
25.00	48.10	23.10
28.00	47.25	19.25
30.00	47.44	17.44
32.00	45.29	13.29
33.00	44.46	11.46
37.00	43.92	6.92
38.00	42.19	4.19
38.00	47.02	9.02
44.00	44.95	0.95
45.00	44.86	0.14
48.00	44.25	3.75
50.00	44.09	5.91
51.00	43.00	8.00
52.00	43.75	8.25
53.00	44.42	8.58
57.00	43.68	13.32
61.00	44.14	16.86
61.00	44.10	16.90
64.00	43.58	20.42
65.00	41.37	23.63
68.00	43.70	24.30
71.00	43.77	27.23

4.1.2 PM10

Model for 2009

Here, a similar computational process has been used for particulate matter data, and its results are shown below. The year that has been chosen for this example is 2009, simply because there are more observations available for this year than for any other year in the data set.

Computing a variogram, considering the linear covariance function:

```
> aal <- variofit(variogram(geo_pm10_2009), cov.model = "li")

variogram: computing omnidirectional variogram
variofit: covariance model used is linear
           sigmasq phi   tausq   kappa
initial.value "52.94" "0"   "17.65" "0.5"
status        "est"  "est" "est"   "fix"
loss value: 18785.3075752401
```

Computing a variogram, considering the exponential covariance function:

```
> aae <- variofit(variogram(geo_pm10_2009), cov.model = "ex")

variogram: computing omnidirectional variogram
variofit: covariance model used is exponential
           sigmasq phi   tausq   kappa
initial.value "35.29" "0.89" "17.65" "0.5"
status        "est"  "est" "est"   "fix"
loss value: 18208.9238162109
```

It is visible that the fitted variogram parameters are smaller for the particulate matter data, rather than for nitrogen dioxide data. The σ^2 parameter stands for the variance of the spatial process of which the data (signal) is assumed to be a realization, ϕ denotes the practice range, τ^2 is the measurement error variance, and κ is the smoothness parameter [32]. Such a difference in the parameter estimates for two contaminants might either be contributed by the difference in years (1999 and 2009), or generally indicate the tendency of nitrogen dioxide to have a greater variability than particulate matter.

Figure 4.2 shows how the chosen variogram would fit the data. It is seen that the fitted curves coincide:

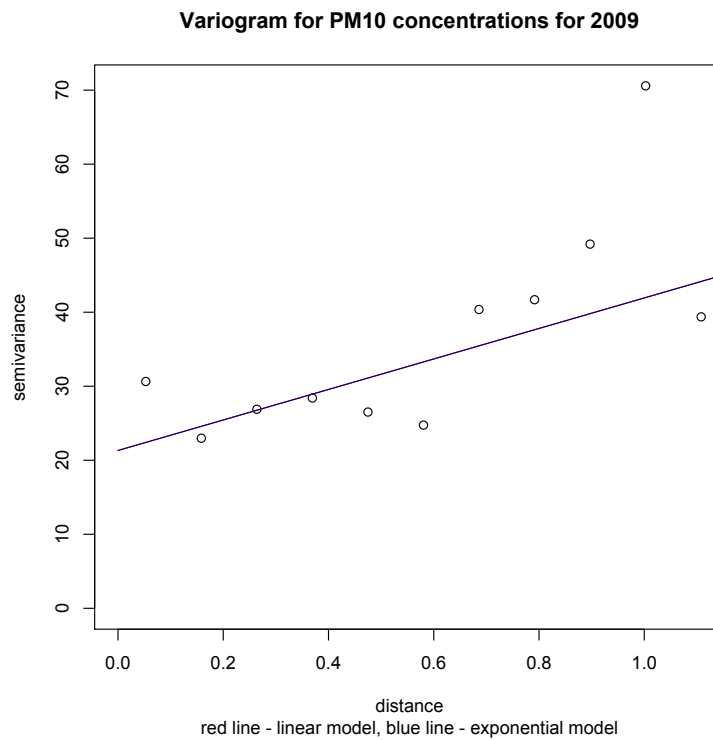


Figure 4.2: Empirical and modeled variogram

The next step is to perform ordinary kriging.

```
> ok_pm10_2009 <- krige.conv(geo_pm10_2009,  
+   locations = locs_pm10_2009,  
+   krige = krige.control(type.krige = "ok",  
+   obj.model = aae))
```

The results are shown in the Table 4.3.

Table 4.3: Ordinary kriging prediction results ($\mu\text{g}/\text{m}^3$)

PM10	kriging variance
35.61	24.79
32.75	27.15
33.09	27.39
36.04	24.37
35.62	24.71
37.07	24.81
37.22	24.95
37.03	24.14
37.15	24.24
37.21	24.01
37.00	24.26
36.49	24.57
37.35	30.64

It is noteworthy that the estimates of the kriging variance are roughly 10 times smaller than for nitrogen dioxide. The reasoning behind that is perhaps the same as behind the difference in covariance model parameter estimates.

Cross-validation of the model

Leave-one-out cross-validation procedure has been executed for particulate matter data, too. The results of the LOO cross-validation for PM₁₀ data for 2009 can be seen from the Table 4.4. Comparing these results to the same results for NO₂, it is seen that absolute errors of prediction for both pollutants are averagely in the same range. The mean absolute value of prediction errors for PM₁₀ is equal to 13.49 ($\mu\text{g}/\text{m}^3$).

Table 4.4: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	predicted	absolute error
25.00	42.77	17.77
25.00	61.00	36.00
28.00	20.00	8.00
29.00	42.79	13.79
31.00	65.00	34.00
31.00	45.42	14.42
32.00	45.46	13.46
33.00	43.23	10.23
33.00	44.06	11.06
33.00	46.09	13.09
33.00	45.00	12.00
34.00	38.00	4.00
34.00	44.00	10.00
34.00	44.25	10.25
34.00	44.50	10.50
34.00	45.32	11.32
35.00	43.50	8.50
36.00	51.00	15.00
36.00	33.00	3.00
37.00	64.00	27.00
38.00	43.00	5.00
38.00	38.00	0.00
39.00	53.00	14.00
39.00	30.00	9.00
40.00	22.00	18.00
40.00	43.57	3.57
40.00	61.00	21.00
40.00	45.57	5.57
40.00	17.00	23.00
41.00	50.00	9.00
41.00	32.00	9.00
41.00	68.00	27.00
42.00	43.17	1.17
43.00	52.00	9.00
46.00	71.00	25.00
48.00	25.00	23.00

4.2 RRCM in iid setting

4.2.1 NO2

Model for 1999

Ridge regression confidence machine prediction has been performed for all the data in the data set, in order to see how this machine learning method works with geostatistical data. At first, the basic linear iid setting [38] of the RRCM has been considered. Then, the model has been advanced to use a non-linear approach. For nitrogen dioxide, the same 1999 year has been used for demonstration of the application of the RRCM method - in order to better show its capacity in comparison to classical kriging. The data that has been used here has been scaled and normalized, as said before. The RRCM prediction has been carried out with the use of the **PredictiveRegression** package [38]. In this ridge regression procedure, the ridge factor has been set to 0.01, and the confidence level has been chosen equal to 95 percent - due to the size of the data set. By its specification, iid RRCM model has the following limitation: for a significance level ϵ , a data set must count at least $1/\epsilon$ observations to impart valid predictions [38]. Therefore, a 95 % confidence is the highest feasible level. The observed data has been taken up as a training set, and the unobserved part, for which the prediction has been made, is the test set.

First, a training and a test set must be created:

```
> train <- no2_1999[1:25, ]
> test <- no2_1999[26:49, 1:2]
```

For the correct performance of the model, training and test sets should be declared of type *matrix*:

```
>train <- as.matrix(train)
>test <- as.matrix(test)
```

Then, the model is to be performed:

```
> iid_no2 <- iidpred(train, test, 0.05, 0.01)
```

The results of the application of the method can be seen from the Table 4.5. There, the first column denotes lower bounds of the prediction interval, the second column denotes their upper bounds, and the third column indicates width of the intervals, namely, the difference between the upper and the lower bounds. As seen from the results, the mean width of the prediction interval for this (test) data set is equal to 66.31 ($\mu\text{g}/\text{m}^3$). The mean value of the lower bound of the prediction interval is 66.31 ($\mu\text{g}/\text{m}^3$), while the mean value of upper bound of the prediction interval makes up 77.76 $\mu\text{g}/\text{m}^3$.

Table 4.5: RRCM iid for NO2 for 1999 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
5.60	72.64	67.05
5.45	72.72	67.27
6.04	76.22	70.18
6.26	76.28	70.01
8.13	73.07	64.94
6.80	77.15	70.36
8.64	73.13	64.49
9.08	73.40	64.32
8.28	74.04	65.76
8.93	74.16	65.23
10.49	75.04	64.55
11.49	77.41	65.92
12.41	77.40	64.98
11.22	77.62	66.39
13.78	78.72	64.94
15.40	78.89	63.49
14.12	78.64	64.51
15.41	81.65	66.24
16.25	80.74	64.49
15.61	83.28	67.67
15.61	83.54	67.93
15.95	83.28	67.32
16.03	83.41	67.38
17.74	83.74	66.00

Cross-validation of the model

Cross-validation of this method is useful because of two reasons:

1. the data set (as well as the data sets for other years) is small,
2. the modeling is performed to predict the values which are not observed, i.e. there is no possibility of comparison of the observed and the predicted.

The results of cross-validation can be seen from the Table 4.6. There is one error in the prediction, i.e. for the first point, the observed value, 17 ($\mu\text{g}/\text{m}^3$), falls below the prediction interval $[22.94, 73.60]$ ($\mu\text{g}/\text{m}^3$). As up to 5 per cent of errors is allowed, the method generally works well.

Table 4.6: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	lower	upper	is in
17.00	22.94	73.60	no
20.00	-29.97	104.26	yes
22.00	6.88	74.96	yes
25.00	15.09	83.56	yes
28.00	17.30	85.39	yes
30.00	18.45	88.72	yes
32.00	13.66	79.09	yes
33.00	12.53	76.66	yes
37.00	9.04	74.80	yes
38.00	-1.63	73.04	yes
38.00	18.07	92.43	yes
44.00	13.26	79.69	yes
45.00	17.61	82.90	yes
48.00	16.72	91.42	yes
50.00	8.78	77.14	yes
51.00	-0.13	74.06	yes
52.00	10.00	79.11	yes
53.00	15.57	79.88	yes
57.00	12.09	76.06	yes
61.00	14.78	78.66	yes
61.00	17.84	83.30	yes
64.00	16.55	79.77	yes
65.00	7.01	71.09	yes
68.00	13.61	77.98	yes
71.00	15.45	80.05	yes

4.2.2 PM10

Model for 2009

The same RRCM model in the default iid setting has been performed for the other contaminant, particulate matter, as well. Here below it is shown, how the method has been applied stepwise, and which results it has yielded. For the better comparison to kriging technique, the same 2009 year has been used for demonstration of the conformal predictor as well. Training and test sets for PM₁₀ data for 2009 should be created first, and then the model can be executed:

```
> train <- pm10_2009[1:36, ]
> test <- pm10_2009[37:49, 1:2]
```

```
> iid_pm10_2009 <- iidpred(train, test, 0.05, 0.01)
```

The results can be seen from the Table 4.7. The mean value of the lower bound of prediction intervals makes up 22.80 ($\mu\text{g}/\text{m}^3$), while the mean value of the upper bound is 51.22 ($\mu\text{g}/\text{m}^3$). The mean width of prediction intervals is equal to 28.41 ($\mu\text{g}/\text{m}^3$).

Table 4.7: RRCM iid for PM10 for 2009 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
20.71	48.92	28.22
19.72	48.58	28.86
19.90	48.84	28.94
22.02	49.61	27.59
22.02	49.53	27.52
23.55	51.41	27.86
24.05	51.91	27.87
24.04	51.49	27.44
23.19	51.97	28.77
23.40	51.98	28.58
24.13	51.97	27.84
24.35	52.75	28.39
25.35	56.91	31.56

Cross-validation of the model

The results of leave-one-out cross-validation for the linear RRCM model in iid setting for PM₁₀ for 2009 can be seen from the Table 4.8. Out of 36 points in the set, there is 1 error: one observed value (25 $\mu\text{g}/\text{m}^3$) falls below the prediction interval [26.42, 50.60] ($\mu\text{g}/\text{m}^3$), but nonetheless, regarding that 5 per cent of errors is allowed, the prediction is still considered as valid.

Table 4.8: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	lower	upper	is in
25.00	20.48	49.96	yes
25.00	26.42	50.60	no
28.00	4.77	56.75	yes
29.00	22.45	50.16	yes
31.00	21.35	48.69	yes
31.00	22.51	50.17	yes
32.00	23.52	52.38	yes
33.00	20.84	49.05	yes
33.00	21.57	48.76	yes
33.00	23.89	52.92	yes
33.00	24.73	53.20	yes
34.00	18.50	48.79	yes
34.00	22.80	50.58	yes
34.00	23.02	50.56	yes
34.00	23.83	51.17	yes
34.00	23.53	52.57	yes
35.00	21.79	48.99	yes
36.00	20.25	50.95	yes
36.00	23.13	50.22	yes
37.00	24.98	53.71	yes
38.00	21.36	48.69	yes
38.00	25.32	57.66	yes
39.00	24.54	52.74	yes
39.00	23.88	52.85	yes
40.00	19.88	48.30	yes
40.00	20.75	48.17	yes
40.00	23.11	50.77	yes
40.00	22.81	51.05	yes
40.00	24.67	51.98	yes
41.00	22.22	51.45	yes
41.00	23.59	51.01	yes
41.00	22.62	50.66	yes
42.00	21.27	49.30	yes
43.00	20.18	49.15	yes
46.00	22.31	51.06	yes
48.00	23.09	49.37	yes

4.3 Comparison of both models in default settings

4.3.1 NO₂

After introducing both models, the classical ordinary kriging and the newly developed ridge regression confidence machine, it is interesting to compare their predictive performance. All the data pools composed of the observations for all the 11 years, are taken up for the comparison. Figure 4.3 shows the results. The upper and the lower dark green dashed lines depict the mean values of upper and lower bounds of RRCM iid prediction intervals for each year. The purple line delineates the mean values of kriging predicted (point) concentrations. It is noteworthy to mention that there have occurred *no errors* in prediction of the RRCM model - in the sense that all the kriging predicted values have fallen within the RRCM intervals. For the general comparison of the models, the ridge factor has been taken up equal to 0.01, and the confidence level - to 0.95, which has been dictated by the average size of the data sets available for each year.

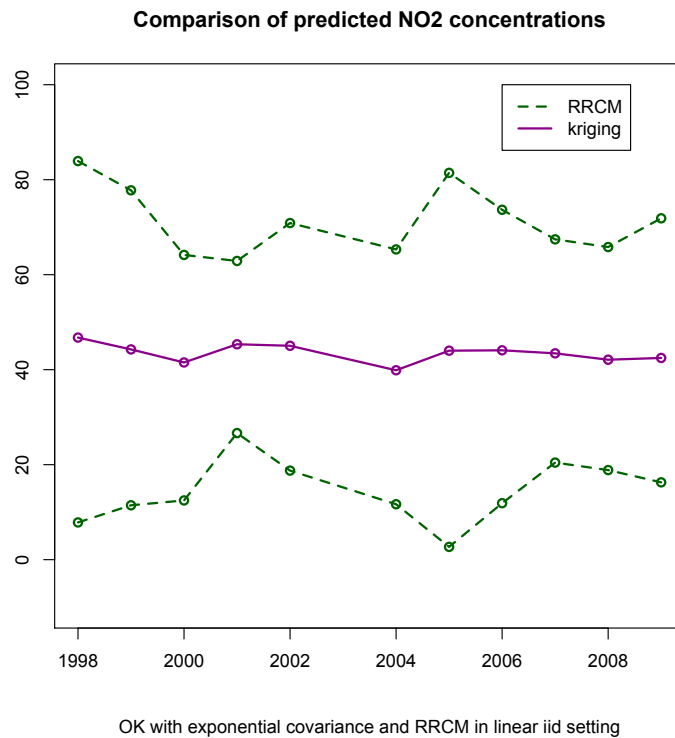


Figure 4.3: Ordinary kriging and ridge regression confidence machine predictions for nitrogen dioxide

It is noteworthy to bring up again that all the predictions have been

performed for the locations where the actual observations are not available. Thus there is no way to compare the modeling results to real data and so to find the model error. Ordinary kriging allows to estimate prediction errors, and prediction intervals for kriging can be derived following the formula (3.17). Kriging prediction intervals are built under the assumption that the confidence level is equal to 95 %. The histogram shown on the Figure 4.4 opposes the kriging intervals to mean width of the RRCM prediction intervals. On this histogram, the light purple bars denote mean kriging intervals, and the shaded purple bars stand for mean RRCM intervals. With no loss of generality, here and below, kriging intervals have been derived assuming that the spatial process is Gaussian. As Figure 4.4 shows, they are generally smaller than those of RRCM, but RRCM intervals are valid and the model does not require any prior assumption on data distribution.

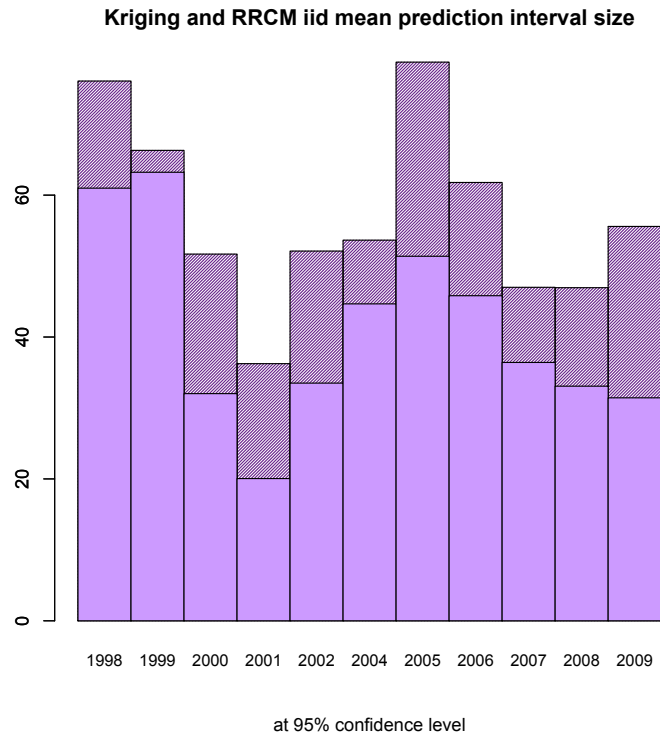


Figure 4.4: Estimated variance for kriging and RRCM prediction

4.3.2 PM10

For particulate matter, the same comparison of the models has been made. This contrasting has been also carried out for the whole data available: from 2001 to 2009, with the exception of 2003. Figure 4.5 depicts the results of the

comparison. As in the nitrogen dioxide case, there have occurred no errors in RRCM prediction - in the same sense as before, i.e. that all the kriging predicted values have fallen within the corresponding RRCM intervals. For RRCM modeling, the ridge factor has been set to 0.01, and the confidence level has been chosen equal to 0.95. Here, the dashed teal lines stand for the mean values of the upper and the lower bounds of RRCM prediction intervals, and the dark orange line denotes the mean kriging predictions.

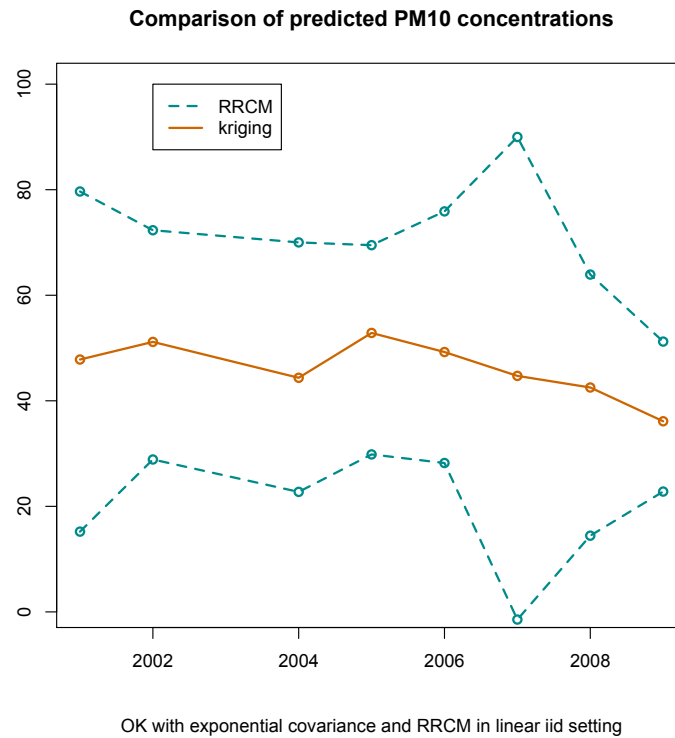


Figure 4.5: Ordinary kriging and ridge regression confidence machine predictions for particulate matter

For this data set, no real observations are available at the modeling locations, and the comparison of the model variance has been made. In order to evaluate the estimated variance in kriging prediction, kriging prediction (or confidence) intervals have been derived. Figure 4.6 shows a histogram where kriging intervals are opposed to RRCM prediction intervals. On this histogram, the blue bars reflect the estimated kriging intervals, while the shaded bars stand for RRCM prediction intervals. These are mean prediction intervals for confidence level of 95 per cent. Kriging variance estimates for PM_{10} are generally smaller than for that NO_2 .

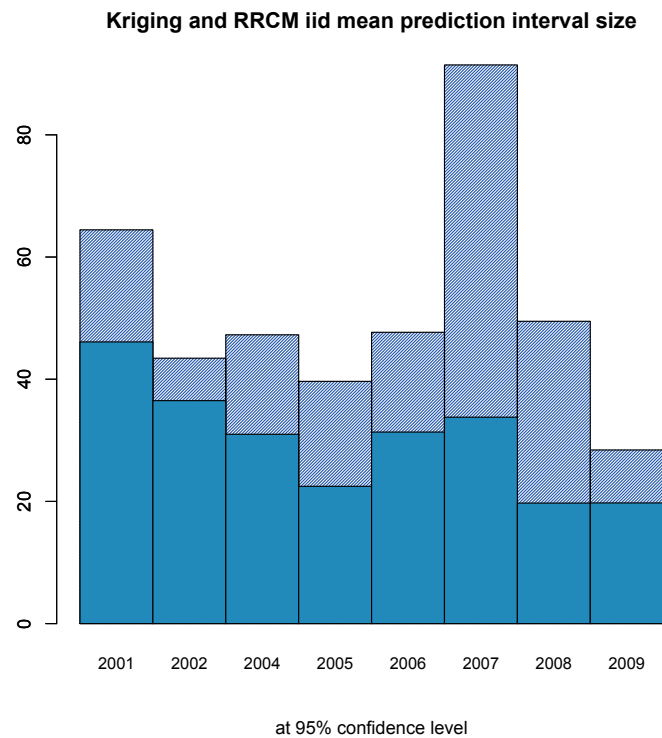


Figure 4.6: Estimated variance for kriging and RRCM prediction

4.4 Ordinary kriging with Gaussian covariance

4.4.1 NO₂

Here another specification of the ordinary kriging model is presented: the one that makes use of the Gaussian covariance function. As before, the detailed computation and results are provided for 1999. All the detailed results for every year are listed in **Appendix B**.

For practical realization of the model, the **geoR** package for R has been employed as well. The Gaussian variomodel has been specified, and ordinary kriging procedure has been applied to the NO data for 1999.

```
> aag <- variofit(variog(geo_no2_1999), cov.model = "ga")
```

```
variog: computing omnidirectional variogram
variofit: covariance model used is gaussian
          sigmasq phi   tausq kappa
initial.value "327.08" "0.91" "163.54" "0.5"
```

```
status      "est"  "est"  "est"  "fix"  
loss value: 2212678.3434232
```

Figure 4.7 shows how the chosen variomodel fits the actual data.

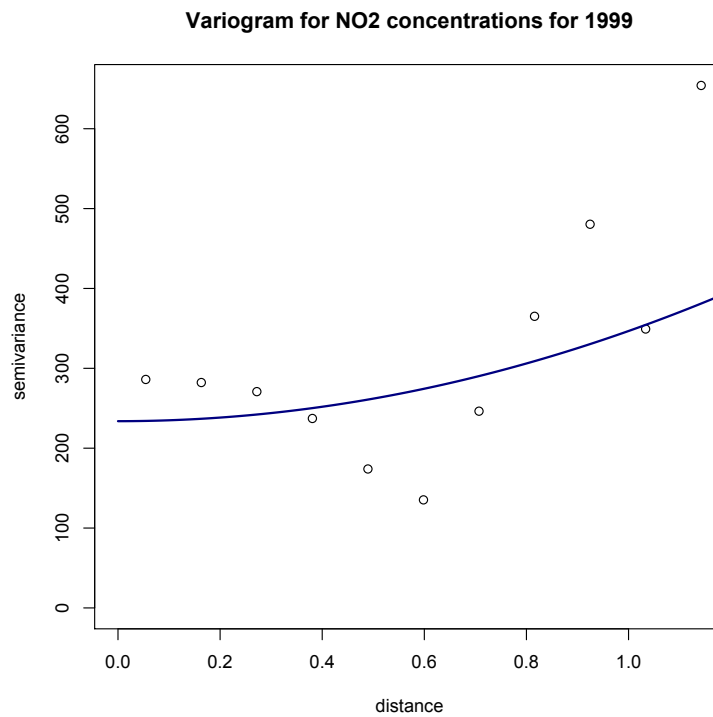


Figure 4.7: Empirical and modeled variogram

Running the ordinary kriging model with Gaussian covariance function:

```
> ok_no2_1999 <- krige.conv(geo_no2_1999,  
+   locations = locs_no2_1999,  
+   krige = krige.control(type.krige = "ok",  
+   obj.model = aag))
```

Table 4.9 shows the results of this prediction.

Table 4.9: Ordinary kriging prediction results ($\mu\text{g}/\text{m}^3$)

NO2	kriging variance
41.78	250.12
41.80	250.37
41.39	260.62
41.46	260.19
41.89	248.55
41.72	260.42
42.28	246.67
42.25	246.99
42.79	247.67
42.93	246.89
42.66	247.67
43.17	249.81
43.49	247.73
44.52	249.34
44.10	247.03
44.80	244.24
45.22	246.69
45.05	248.38
45.30	245.62
45.35	250.39
45.39	250.76
45.48	249.78
45.53	249.84
46.29	247.86

The estimated kriging variances are slightly smaller for this model specification than those for the one making use of the exponential covariance function. This decrease might be indicating that this variomodel is a better fit for the data (which is also seen, comparing Figure 4.1 and Figure 4.7). Nevertheless, the variances are still very big.

Cross-validation of the model

Cross-validation of this model has been performed as for the previous model. Its results can be seen in Table 4.10. A quick view on the absolute errors of predictions reveal that for this particular data set, a Gaussian covariance function does not provide any significant improvement of efficiency of prediction. The errors in both kriging models for this data are within the same

range of values.

Table 4.10: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	predicted	absolute error
17.00	47.81	30.81
20.00	45.42	25.42
22.00	42.80	20.80
25.00	45.87	20.87
28.00	46.97	18.97
30.00	47.83	17.83
32.00	45.41	13.41
33.00	44.35	11.35
37.00	43.30	6.30
38.00	40.83	2.83
38.00	49.53	11.53
44.00	43.99	0.01
45.00	44.38	0.62
48.00	44.25	3.75
50.00	43.99	6.01
51.00	41.90	9.10
52.00	42.57	9.43
53.00	45.42	7.58
57.00	43.98	13.02
61.00	43.89	17.11
61.00	45.52	15.48
64.00	43.58	20.42
65.00	40.37	24.63
68.00	43.33	24.67
71.00	44.20	26.80

4.4.2 PM10

For PM₁₀ data, an ordinary kriging model with Gaussian covariance function has also been derived. Here below the detailed computation and results are demonstrated for the data set corresponding to 2009. For the rest of the years, the results can be seen in **Appendix B**.

A Gaussian covariance model has been fitted as follows:

```
> aag <- variofit(variog(geo_pm10_2009), cov.model = "ga")
```



```

variog: computing omnidirectional variogram
variofit: covariance model used is gaussian
           sigmasq phi   tausq  kappa
initial.value "35.29" "0.71" "17.65" "0.5"
status        "est"  "est"  "est"  "fix"
loss value: 21882.6068581801

```

Figure 4.8 depicts how the chosen variomodel fits the given data.

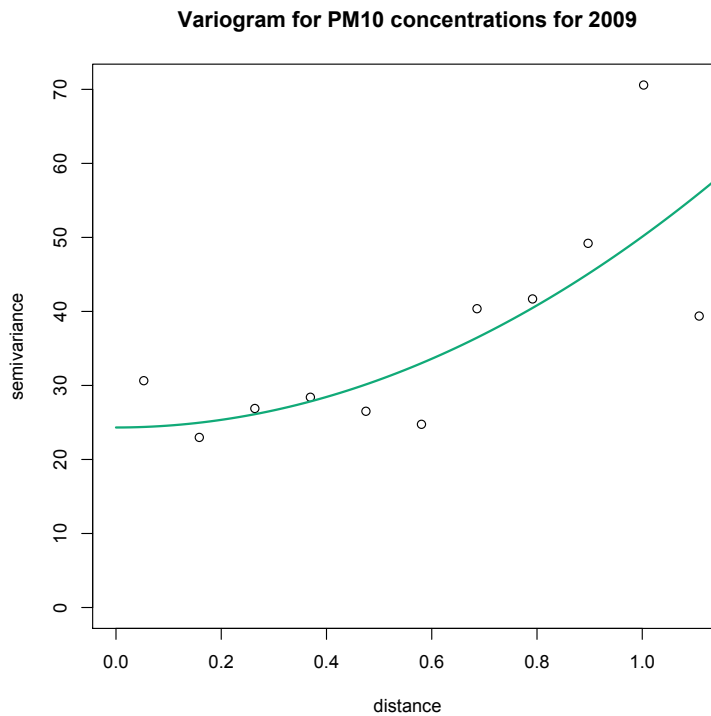


Figure 4.8: Empirical and modeled variogram

Then an ordinary kriging procedure with the fitted Gaussian variomodel has been executed:

```

> ok_pm10_2009 <- krige.conv(geo_pm10_2009,
+   locations = locs_pm10_2009,
+   krige = krige.control(type.krige = "ok",
+   obj.model = aag))

```

The results of the modeling can be seen from Table 4.11. It is noteworthy that the estimated kriging variance is very low: the maximum value of it

(28.46 $\mu\text{g}/\text{m}^3$) is smaller than the smallest predicted concentration (34.41 $\mu\text{g}/\text{m}^3$).

Table 4.11: Ordinary kriging prediction results ($\mu\text{g}/\text{m}^3$)

NO2	kriging variance
35.45	25.87
34.41	26.54
34.51	26.57
36.01	25.50
35.47	25.50
37.04	25.74
37.31	25.77
36.78	25.31
36.59	25.84
36.68	25.75
36.96	25.47
37.29	25.71
39.20	28.46

Cross-validation of the model

Table 4.12 shows the results of cross-validation for ordinary kriging model that makes use of the Gaussian covariance function for PM_{10} data for 2009. This kriging model provides a significantly better fit for this data compared to the ordinary kriging model that makes use of the exponential (or linear) covariance function. The absolute values of the errors of prediction are small.

Table 4.12: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	predicted	absolute error
25.00	35.35	10.35
25.00	38.67	13.67
28.00	36.43	8.43
29.00	36.42	7.42
31.00	35.30	4.30
31.00	35.90	4.90
32.00	37.50	5.50
33.00	35.59	2.59
33.00	35.43	2.43
33.00	36.95	3.95
33.00	37.55	4.55
34.00	35.01	1.01
34.00	36.01	2.01
34.00	36.12	2.12
34.00	36.67	2.67
34.00	36.75	2.75
35.00	35.39	0.39
36.00	36.26	0.26
36.00	36.42	0.42
37.00	37.96	0.96
38.00	35.39	2.61
38.00	39.26	1.26
39.00	37.56	1.44
39.00	37.15	1.85
40.00	34.93	5.07
40.00	34.88	5.12
40.00	36.30	3.70
40.00	36.30	3.70
40.00	37.20	2.80
41.00	36.71	4.29
41.00	36.79	4.21
41.00	36.14	4.86
42.00	35.68	6.32
43.00	34.87	8.13
46.00	36.40	9.60
48.00	35.56	12.44

4.5 RRCM with RBF kernel

4.5.1 NO₂

Model for 1999

In order to switch to a nonlinear setting, ridge regression confidence machine makes use of the so-called “kernel trick” [6]. With the use of this technique any kernel can be considered. Hereby it is demonstrated, how the Gaussian RBF kernel has been employed for the nitrogen dioxide data. For the sake of precise comparison, the same 1999 year has been taken up for the example below. The R [47] function for executing the RRCM procedure with an RBF kernel has been derived on the basis of the function *iidpred* from the **PredictiveRegression** [38] package. The detailed listing of the code can be found in the **Appendix A**. The new function has been named *iid_rbf*. The confidence level has been set to 0.95, and the ridge factor has been chosen equal to 0.01, as in the other models presented in this dissertation. The R function has been called as follows:

```
> iid_no2 <- iid_rbf(train, test, 0.05, 0.01)
```

The results of the application of the method can be seen in the Table 4.13. It should be pointed out that this prediction has seen no errors - in the sense that all the prediction intervals contained the relevant kriging predictions.

Table 4.13: RRCM RBF for NO₂ for 1999 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
6.56	73.23	66.67
6.60	73.29	66.69
0.30	88.02	87.72
0.83	87.84	87.01
6.70	75.12	68.42
0.75	90.83	90.09
9.19	74.32	65.13
8.71	75.15	66.44
10.97	74.62	63.65
11.51	74.77	63.26
11.34	77.43	66.09
13.27	80.65	67.38
14.83	79.37	64.54
14.05	79.18	65.13
16.79	79.60	62.81
17.46	78.46	60.99
15.94	78.54	62.60
18.36	81.35	63.00
18.27	79.59	61.32
17.99	83.38	65.40
17.85	83.78	65.93
18.20	82.87	64.68
18.17	82.96	64.79
17.39	81.96	64.57

Cross-validation of the model

Cross-validation of the RRCM model with the RBF kernel has been performed as well. As before, the leave-one-out cross-validation has been chosen, which has been dictated by the size of the data set available. This method helps assess the efficiency of the method's specification as opposing its prediction to the observed data. The results of the application of the leave-one-out cross-validation to the NO₂ data for 1999 can be seen in the Table 4.14. The cross-validation has revealed one error in prediction: for the first observation the actual value lies below the RRCM predicted interval. Nevertheless, the assumed level of confidence allows up to 5 per cent of error, so with 1 error the prediction is still valid. It is clear that the prediction interval for the second observation is very ineffective, since its size is huge, but the actual value lies within the interval, so this cannot be classified as

an error.

Table 4.14: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	lower	upper	is in
17.00	22.25	75.20	no
20.00	-8477.56	579.32	yes
22.00	7.82	79.18	yes
25.00	19.58	85.12	yes
28.00	18.73	84.78	yes
30.00	18.33	89.36	yes
32.00	15.88	80.71	yes
33.00	14.59	78.72	yes
37.00	11.56	76.20	yes
38.00	-11.42	79.65	yes
38.00	9.89	95.32	yes
44.00	16.36	82.19	yes
45.00	17.38	81.38	yes
48.00	-2.53	97.45	yes
50.00	10.26	81.32	yes
51.00	-3.49	81.31	yes
52.00	5.46	93.48	yes
53.00	16.14	78.22	yes
57.00	14.26	76.32	yes
61.00	17.43	78.30	yes
61.00	7.28	79.96	yes
64.00	15.76	75.92	yes
65.00	2.35	71.19	yes
68.00	14.94	78.31	yes
71.00	15.67	78.19	yes

4.5.2 PM10

Model for 2009

This modeling approach has also been used for particulate matter data. The RRCM procedure with the Gaussian RBF kernel has been called as follows:

```
> iid_pm10 <- iid_rbf(train, test, 0.05, 0.01)
```

The results of the application of this method can be seen from the Table 4.15. As in the other models, this prediction has seen no errors - in the sense that all the kriging predicted values fall within the RRCM predictive

intervals. On the average, the predictive intervals obtained with this RRCM specification are of slightly bigger size than the corresponding estimated kriging variance. Nevertheless, those intervals are valid, i.e. the real concentrations fall within those intervals with a given level of confidence, that is, 95 per cent. Considering that the mean width of the prediction intervals is equal to 29.15, it is a good result.

Table 4.15: RRCM RBF for PM10 for 2009 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
21.88	49.01	27.14
16.99	49.04	32.06
16.78	49.85	33.07
23.32	49.48	26.16
22.34	48.87	26.53
24.16	50.85	26.68
24.44	51.19	26.75
24.70	51.17	26.47
23.49	51.63	28.15
23.80	51.63	27.83
24.56	51.71	27.15
24.17	52.79	28.62
16.07	58.50	42.43

Cross-validation of the model

Cross-validation has been performed for this model and data. The result of the leave-one-out cross-validation can be seen from the Table 4.16. There has occurred one error in the prediction: in the second observation, the lower bound of the prediction interval is bigger than the actual value. With the assumed confidence of 95 per cent, up to 1.8 errors are allowed, so generally the prediction is good. Nonetheless, it is seen from the table that in the third observation the prediction set is the whole real line, which means that the prediction at this point is completely inefficient.

Table 4.16: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	lower	upper	is in
25.00	19.53	51.23	yes
25.00	26.06	51.81	no
28.00	-Inf	Inf	yes
29.00	23.95	50.17	yes
31.00	22.05	48.62	yes
31.00	23.54	49.56	yes
32.00	24.07	52.06	yes
33.00	22.02	49.42	yes
33.00	22.40	48.82	yes
33.00	24.35	52.97	yes
33.00	25.09	53.76	yes
34.00	17.47	51.82	yes
34.00	23.72	49.94	yes
34.00	24.12	50.02	yes
34.00	25.01	51.13	yes
34.00	23.96	52.48	yes
35.00	22.33	48.63	yes
36.00	18.69	52.73	yes
36.00	24.24	50.29	yes
37.00	24.40	52.81	yes
38.00	22.21	48.70	yes
38.00	10.95	65.22	yes
39.00	24.31	51.72	yes
39.00	23.26	53.03	yes
40.00	20.56	48.27	yes
40.00	20.81	47.90	yes
40.00	23.94	50.31	yes
40.00	23.47	50.55	yes
40.00	24.57	51.58	yes
41.00	20.73	51.19	yes
41.00	24.14	50.56	yes
41.00	23.35	50.14	yes
42.00	22.47	49.12	yes
43.00	17.93	48.74	yes
46.00	22.05	50.23	yes
48.00	23.72	48.95	yes

There are two ways to tackle the efficiency problem. First is to consider a lower confidence level, and second is to boost the ridge factor. Also, it

should be assured that covariance function (or, kernel) parameters values are specified well. Table 4.17 shows the results obtained for this “trouble” point using both of the proposed solutions.

Table 4.17: Solutions comparison

solution	observed	lower	upper
<i>conf</i> = 0.90	28	-7.70	53.22
<i>ridge</i> = 1	28	5.30	39.35

It must be said that shifting the ridge factor has resulted more beneficial than decreasing the confidence level. It has not only improved the efficiency of the prediction, but also has yielded 0 errors in it, i.e. all the observed values have appeared to be within the prediction intervals. The mean prediction interval size for the RRCM model with the ridge factor set to 1 is $26.15 \text{ } (\mu\text{g}/\text{m}^3)$. For a cross-validation result, i.e. the average for 36 different models, it is a good efficiency. It is clear that the choice of the ridge factor influences the efficiency of the RRCM prediction to a great extent. This point is discussed below. All of the models are introduced and demonstrated assuming the confidence level equal to 0.95 and the ridge factor equal to 0.01.

4.6 Comparison of both models in Gaussian setting

4.6.1 NO₂

The results of the comparison of the kriging and RRCM models in the Gaussian setting are presented here. Figure 4.9 opposes the ordinary kriging with Gaussian covariance model to the ridge regression confidence machine predictor with Gaussian RBF kernel. The dark red upper and lower dashed lines reflect the mean values of upper and lower bounds of prediction intervals of RRCM, and the blue line between them shows the mean kriging predictions. It is noteworthy that there have appeared no errors in the predictions of these nonlinear models as well - in the sense that all of the RRCM prediction intervals withhold the kriging predicted values for all the years throughout the given timeline.

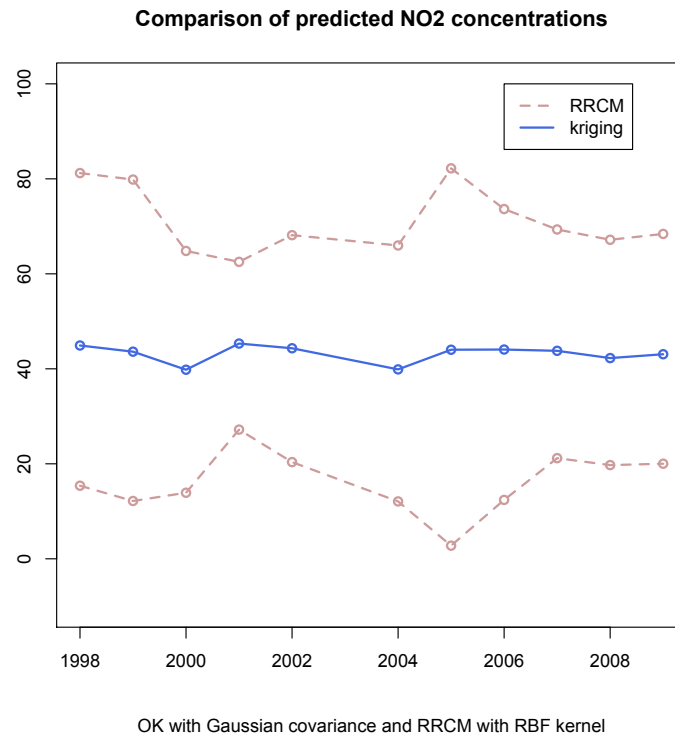


Figure 4.9: Ordinary kriging and ridge regression confidence machine predictions for nitrogen dioxide

With this model, there is no possibility to evaluate the “true” error of the estimation, since the observations are not available for the points where the prediction has been made. Mead kriging prediction intervals, derived under assumption of Gaussianity for 95 per cent confidence, have been opposed to mean RRCM intervals. Figure 4.10 reflects this comparison: the dark orange bars denote the estimated kriging intervals, while the shaded ones stand for the mean size of RRCM prediction intervals.

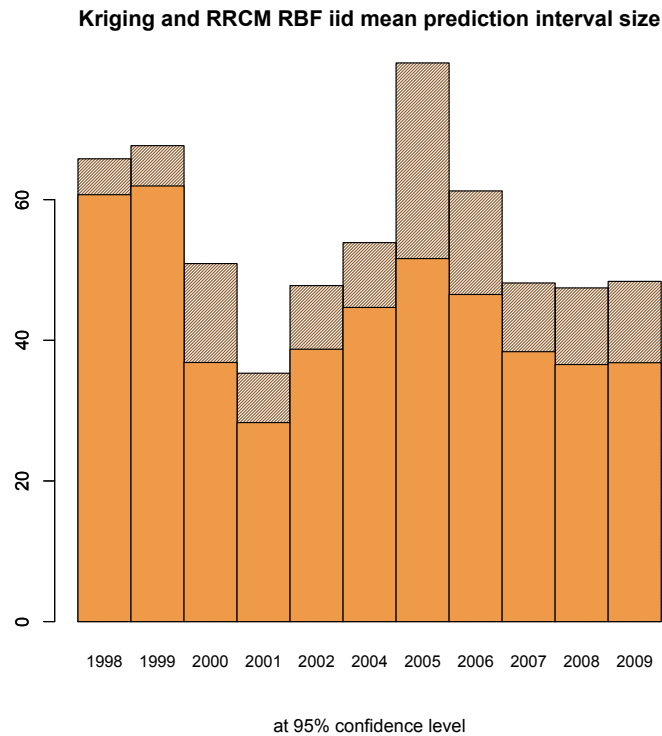


Figure 4.10: Estimated variance for kriging and RRCM prediction

4.6.2 PM10

The same analysis has been carried out for the particulate matter data. The results of comparison of the ordinary kriging model with Gaussian covariance to the ridge regression confidence machine with RBF kernel prediction has been done, and the results of it are shown on the Figure 4.11. The dashed grey lines depict the upper and the lower bounds of the RRCM prediction intervals, while the center red line shows the mean kriging prediction. Like in the other models, there has been no errors in the prediction - in the same sense as above, i.e. all the RRCM intervals have contain the relevant kriging prediction.

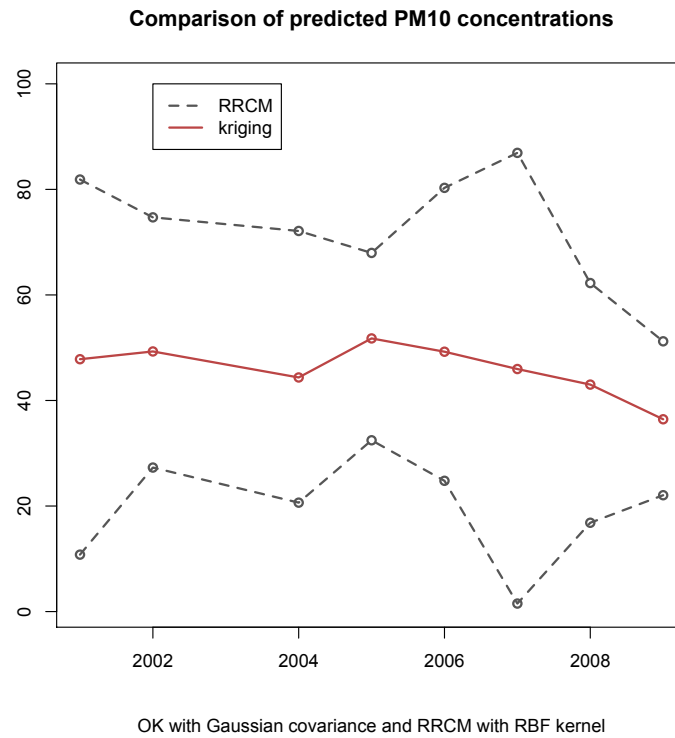


Figure 4.11: Ordinary kriging and ridge regression confidence machine predictions for particulate matter

As for NO_2 data, no observed values for the points where the prediction has been carried out are available. So, mean kriging prediction intervals have been constructed and compared to RRCM prediction intervals. The results of this comparison can be seen on the Figure 4.12. Here, the dark red bars denote the mean kriging intervals, and the shaded bars stand for the mean RRCM interval size.

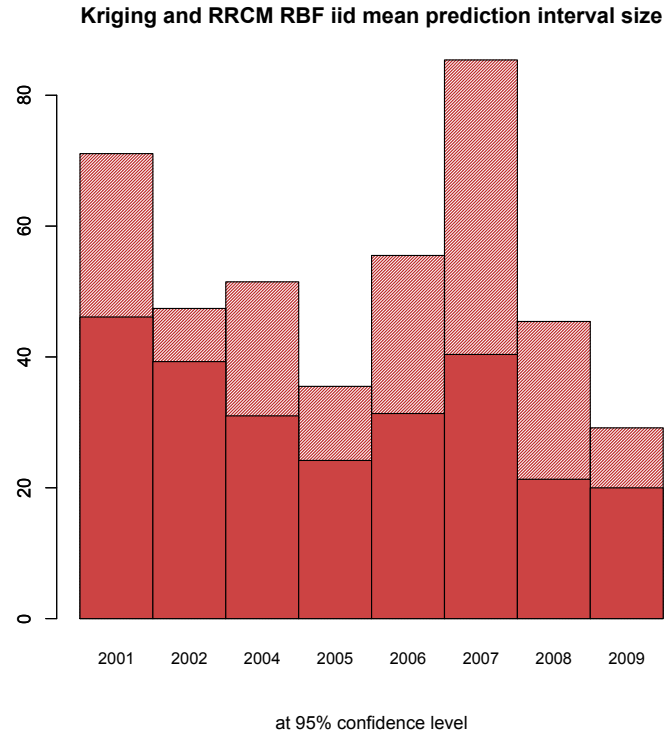


Figure 4.12: Estimated variance for kriging and RRCM prediction

4.7 RRCM with polynomial kernel of second order

4.7.1 NO₂

Model for 1999

The last model to be introduced is the RRCM model with the inhomogeneous polynomial kernel of the second order [1]. This kernel has been applied to the data as a result of a guess that since it makes use of the square of longitude and latitude, it might be a fit for the data. The data pools for each year are rather small to come up with a precise distribution for the data, so in this dissertation, as a matter of introduction of the methods and their testing, various approaches are used, regardless of how well do they actually sort with the data. Polynomial kernel of the second order might not be the best fit, but its implementation demonstrates how any kernel can be employed, once it suits the data. As said before, ridge regression confidence machine can be treated as an additional method or even an alternative to kriging. In some sense it is a similar method, so every covariance function that can be used in kriging can be also used as a kernel for RRCM.

In order to actually employ the RRCM with a polynomial kernel of the second order for the nitrogen dioxide (and particulate matter) data, an R function has been created: *iid_polyn_2*. In fact, this function is a modification of the *iidpred* function from the **PredictiveRegression** package, and it makes use of the inhomogeneous polynomial kernel of the second order. The function is called as follows:

```
> iid_no2 <- iid_polyn_2(train, test, 0.05, 0.01)
```

The results of modeling can be seen in Table 4.18. Again, there have occurred no errors in the prediction - in the sense that all the prediction intervals included the corresponding kriging predicted values. Mean prediction interval size is 65.31 ($\mu\text{g}/\text{m}^3$).

Table 4.18: RRCM polynomial for NO2 for 1999 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
4.28	69.67	65.39
4.28	69.80	65.52
-1.16	70.84	72.01
-0.95	70.76	71.70
3.91	68.57	64.66
-0.82	70.87	71.70
5.12	68.71	63.58
4.73	68.49	63.75
5.72	70.04	64.31
5.96	69.88	63.92
4.53	68.57	64.03
4.11	69.10	64.99
5.07	68.91	63.84
6.33	72.82	66.49
5.77	69.16	63.39
7.21	69.51	62.30
7.15	72.35	65.20
6.06	70.05	63.99
7.09	70.05	62.96
5.67	70.59	64.92
5.59	70.67	65.09
5.94	70.62	64.68
5.96	70.68	64.72
7.01	71.45	64.44

Cross-validation of the model

Cross-validation has been also made for this model as before. The results of it can be seen in the Table 4.19. There has occurred one error in the prediction: for the last observation in the set, the actual value is bigger than the upper bound of the correspondent prediction interval. Nevertheless, 1 error makes up less than 5 per cent of errors, which is allowed under 95 per cent confidence level, and thus the prediction is correct. The average size of the prediction intervals in the cross-validation is $69.35 \text{ } (\mu\text{g}/\text{m}^3)$.

Table 4.19: Leave-one-out cross-validation results ($\mu\text{g}/\text{m}^3$)

observed	lower	upper	is in
17.00	9.80	71.47	yes
20.00	-29.03	97.00	yes
22.00	5.11	70.09	yes
25.00	7.11	69.56	yes
28.00	8.25	70.50	yes
30.00	7.92	71.59	yes
32.00	7.79	71.53	yes
33.00	7.48	69.66	yes
37.00	5.95	69.85	yes
38.00	1.14	71.15	yes
38.00	3.46	81.70	yes
44.00	4.36	69.29	yes
45.00	6.90	71.27	yes
48.00	0.41	82.88	yes
50.00	4.64	73.22	yes
51.00	2.31	72.52	yes
52.00	0.49	69.76	yes
53.00	5.82	73.16	yes
57.00	5.46	71.48	yes
61.00	3.94	69.72	yes
61.00	4.32	72.25	yes
64.00	4.51	73.57	yes
65.00	0.96	67.89	yes
68.00	2.41	69.80	yes
71.00	4.62	69.18	no

4.7.2 PM10

Model for 2009

The RRCM model with the second order polynomial kernel has been also applied to the particulate matter data. The function has been called as follows:

```
> iid_pm10 <- iid_polyn_2(train, test, 0.05, 0.01)
```

The results of the implementing of this model can be seen from the Table 4.20. Mean prediction interval size is 32.26 ($\mu\text{g}/\text{m}^3$), and the prediction is correct - in the sense that all the prediction intervals contain the kriging predicted values. Here, the kriging model with the Gaussian variomodel is meant.

Table 4.20: RRCM polynomial for PM10 for 2009 ($\mu\text{g}/\text{m}^3$)

lower bound	upper bound	interval width
17.03	48.55	31.53
14.56	48.69	34.12
14.60	48.71	34.12
17.67	48.47	30.80
16.77	47.86	31.09
17.91	49.47	31.56
17.93	49.68	31.75
17.93	48.30	30.37
17.24	48.49	31.25
17.39	48.48	31.09
17.87	48.52	30.65
17.79	48.95	31.16
15.03	54.98	39.95

Cross-validation of the model

Leave-one-out cross-validation has been performed for this model and data. The results are in the Table 4.21. Cross-validation has revealed one error in the last observation: the actual value is higher than the upper bound of the correspondent prediction interval. Nevertheless, since the confidence level for the prediction is chosen equal to 95 per cent, the prediction is still correct.

Table 4.21: Leave-one-out cross-validation

observed	lower	upper	is in
25.00	15.23	49.01	yes
25.00	18.43	49.53	yes
28.00	1.18	59.55	yes
29.00	17.89	48.65	yes
31.00	16.90	47.79	yes
31.00	17.29	47.78	yes
32.00	17.71	50.11	yes
33.00	16.98	48.50	yes
33.00	17.12	47.76	yes
33.00	17.39	48.51	yes
33.00	18.04	48.93	yes
34.00	15.72	49.23	yes
34.00	17.07	47.92	yes
34.00	17.35	47.87	yes
34.00	17.98	48.08	yes
34.00	17.17	48.46	yes
35.00	16.82	47.69	yes
36.00	16.41	50.23	yes
36.00	17.92	48.27	yes
37.00	17.38	50.39	yes
38.00	17.01	47.85	yes
38.00	14.94	53.87	yes
39.00	17.37	50.01	yes
39.00	17.18	48.94	yes
40.00	16.33	48.11	yes
40.00	16.27	47.72	yes
40.00	17.16	48.10	yes
40.00	16.90	48.23	yes
40.00	17.69	48.99	yes
41.00	16.76	50.06	yes
41.00	17.57	49.02	yes
41.00	16.81	48.13	yes
42.00	17.02	48.40	yes
43.00	15.04	48.12	yes
46.00	16.44	48.60	yes
48.00	18.92	45.44	no

4.8 Comparison of kriging and ridge regression confidence machine with polynomial kernel

4.8.1 NO₂

In order to see what polynomial kernel can yield in relation to nitrogen dioxide data, the corresponding mode has been run on the data for all the given years. The results of this model have been opposed to those of ordinary kriging with Gaussian covariance. No other covariance function has been taken up here, because both Gaussian covariance function and polynomial kernel are nonlinear of the same order, i.e. of the second degree. A finer analytical form of a covariance function to contrast with a polynomial kernel of second order can be derived, of course. Figure 4.13 shows the results of the comparison of approaches. The upper and the lower lines stand for mean values of RRCM prediction intervals for each year, and the central line depicts mean kriging predictions. There have appeared no errors in the prediction - in the sense that all the kriging predictions have fallen within the correspondent RRCM prediction intervals.

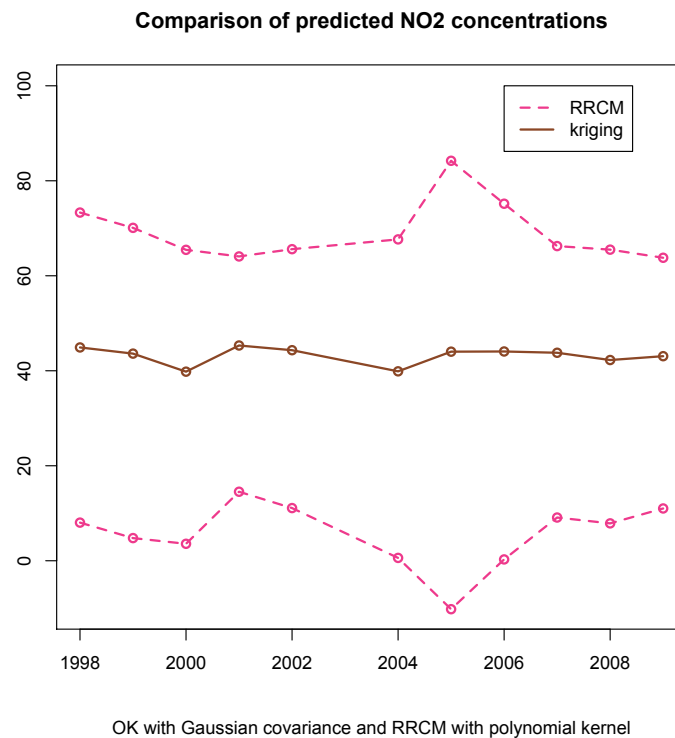


Figure 4.13: Ordinary kriging and ridge regression confidence machine predictions for nitrogen dioxide

The measure of uncertainty of prediction, the kriging variance, has been opposed to mean width of prediction intervals for RRCM with polynomial kernel of the second order. To do so, as before, kriging intervals have been constructed, and their mean size has been compared to mean size of RRCM intervals. The results of the comparison can be seen on Figure 4.14. The light purple bars show the kriging intervals for each year, and the shaded bars stand for mean size of RRCM prediction intervals.

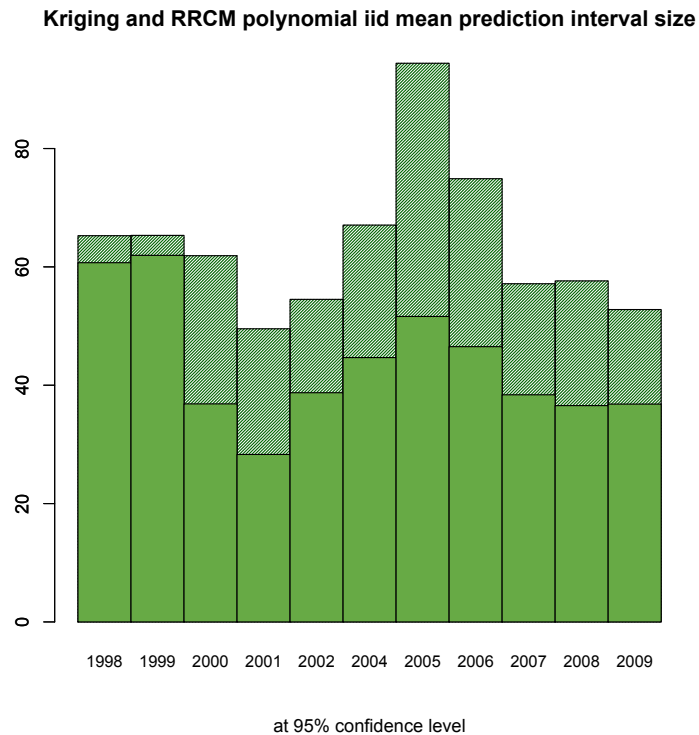


Figure 4.14: Estimated variance for kriging and RRCM prediction

4.8.2 PM10

Models of the same type have been used for particulate matter data. Figure 4.15 shows the comparison of the ordinary kriging and the RRCM predictions. Here, there also have appeared no errors in prediction - in the sense that the RRCM intervals for all of the years contain the kriging predicted values.

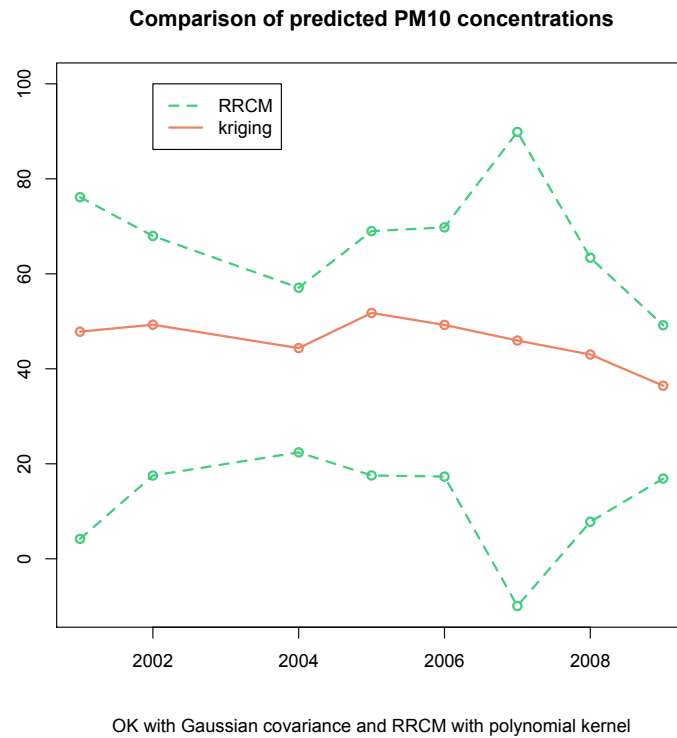


Figure 4.15: Ordinary kriging and ridge regression confidence machine predictions for particulate matter

The estimated error variance of kriging model in for of prediction intervals is opposed to the RRCM intervals, and the results of this contrasting for each year are shown on the Figure 4.16. The yellow bars stand for kriging intervals, and shaded bars depict the mean interval size of RRCM with polynomial kernel of the second order.

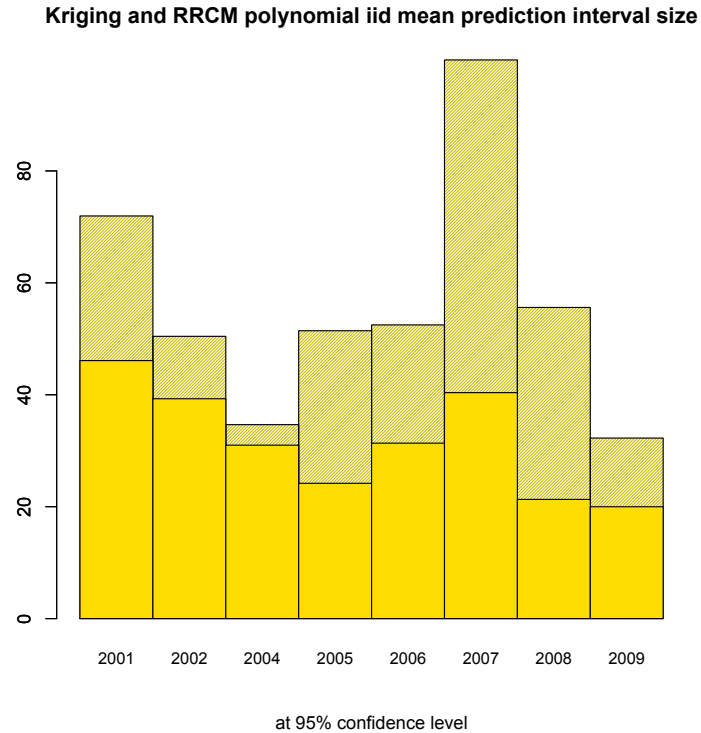


Figure 4.16: Estimated variance for kriging and RRCM prediction

4.9 Comparison of ridge regression confidence machine models

4.9.1 NO₂

The models

It has been demonstrated that ridge regression confidence machine is a suitable approach for assessment of air pollution data. It can be treated as an add-on to kriging, because it adds validity to kriging prediction. It can be also treated as an independent prediction technique, but the setting of the method must be carefully selected. This dissertation takes up both linear and nonlinear approaches to RRCM, and two nonlinear kernels have been taken up: the Gaussian radial basis function and the polynomial kernel of the second order. The RBF kernel has been used because it is a direct analogue of the Gaussian covariance function in kriging, the the polynomial kernel has been employed on the basis of a guess that it might be a good fit for spatial data. However, the implementation of the polynomial kernel is not aimed to match the actual data distribution in the best possible way

rather than to demonstrate that kernels of any kind can be used in RRCM. Those kernels should, of course, sort with the data.

First, performance of the basic linear RRCM model in the iid setting is investigated. Table 4.22 shows the description of the results of modeling, which could help assess how the model suits the given data. Minimum and maximum values of the lower and the upper bounds of prediction intervals are shown in order to track to which extremes do they get (if they contain negative or very big values). Minimum, mean and maximum values of prediction intervals are also provided for each year with the aim to elicit the variation in the estimates. Numbers in bold depict the minimum and the maximum values of the predictions for correspondent columns. It is seen that the “best” prediction in terms of efficiency has been achieved for 2001, and the “worst” - for 2005.

Table 4.22: RRCM iid model for NO₂ ($\mu\text{g}/\text{m}^3$)

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
1998	-0.55	14.73	72.89	76.08	80.73	79.91	90.20
1999	5.45	17.74	63.49	66.31	70.36	72.64	83.74
2000	3.61	19.11	49.68	51.69	54.13	57.75	69.43
2001	20.21	31.84	34.28	36.25	38.15	57.23	69.74
2002	9.15	25.04	50.06	52.12	54.12	63.26	76.48
2004	2.19	16.49	50.71	53.65	61.59	61.16	72.38
2005	-10.09	8.50	73.99	78.75	90.92	76.51	90.56
2006	6.50	15.26	59.10	61.79	69.98	69.57	84.59
2007	14.63	25.29	44.53	47.01	51.91	61.78	73.04
2008	13.92	23.37	44.60	46.96	52.12	60.57	73.68
2009	10.97	21.20	53.08	55.59	62.47	67.34	82.24

Table 4.23 shows the same parameters for the prediction obtained with the ridge regression confidence machine that makes use of the Gaussian RBF kernel. Naturally, this model is also iid. It is seen that the “best” prediction in the sense of efficiency has been obtained for 2001, and the “worst” one has been obtained for 2005, as with the plain iid model. However, RRCM with the RBF kernel acts “worse” as its plain linear counterpart, because the lowest of the lower bounds of the prediction intervals over all the years reaches -31.35 ($\mu\text{g}/\text{m}^3$) for 2005, while for the linear model it is equal to -10.09 ($\mu\text{g}/\text{m}^3$). Generally, there are more negative values amongst the minimums of the lower bounds of prediction intervals for RRCM Gaussian

RBF model. Nevertheless, the mean prediction interval size is smaller for the RBF iid specification ($56.92 \mu\text{g}/\text{m}^3$) than for the plain iid one ($55.09 \mu\text{g}/\text{m}^3$).

Table 4.23: RRCM RBF model for NO₂ ($\mu\text{g}/\text{m}^3$)

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
1998	4.89	20.88	59.41	65.81	85.35	76.53	90.24
1999	0.30	18.36	60.99	67.68	90.09	73.23	90.83
2000	-1.77	20.49	47.09	50.91	58.32	56.55	72.64
2001	18.24	32.05	32.94	35.32	40.67	57.96	67.90
2002	10.63	27.78	44.52	47.78	53.73	61.83	74.83
2004	-7.08	18.17	48.99	53.89	83.14	62.90	76.06
2005	-31.35	11.56	71.38	79.44	124.73	78.26	93.39
2006	-11.00	18.20	56.24	61.24	94.33	69.09	83.33
2007	1.40	27.66	43.21	48.15	71.33	63.97	80.98
2008	6.16	24.94	42.28	47.45	68.97	62.02	76.11
2009	8.05	26.18	43.65	48.38	72.05	63.53	80.10

Finally, Table 4.24 presents the same characteristics for the ridge regression confidence machine model with the polynomial kernel of the second order. This model is iid, too. It is seen that the “best” prediction in terms of efficiency has been achieved for 2001, as with the other models, and the “worst” for 2005. Compared to other models, the RRCM with polynomial kernel of the second order is “worse”: for 2005, the maximum lower bound of the prediction intervals is negative, but it must be taken into consideration that 2005 is clearly an “outlying” year for the whole data set. The mean prediction interval width is equal to $63.68 (\mu\text{g}/\text{m}^3)$, which is higher than for both linear iid and Gaussian RBF iid model settings. This might indicate that the polynomial kernel of the second order is not suitable for this particular nitrogen dioxide data.

Table 4.24: RRCM polynomial model for NO₂ ($\mu\text{g}/\text{m}^3$)

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
1998	1.18	11.83	62.58	65.27	72.11	71.51	77.42
1999	-1.16	7.21	62.30	65.32	72.01	68.49	72.82
2000	-4.27	7.19	59.36	61.89	68.17	63.28	69.89
2001	8.31	17.36	47.52	49.54	54.55	62.06	67.37
2002	4.33	14.29	52.36	54.51	59.91	63.57	69.72
2004	-7.28	2.95	63.49	67.06	84.93	65.58	77.65
2005	-22.78	-6.78	88.99	94.41	119.77	81.53	96.99
2006	-7.04	2.67	71.00	74.90	94.59	72.90	87.55
2007	3.60	11.10	54.24	57.15	71.84	64.39	75.44
2008	3.05	9.93	54.69	57.63	72.44	63.36	76.44
2009	5.09	13.28	50.02	52.79	66.28	61.16	76.03

Comparison to observed data

In order to complete the comparison of predictive powers of the models, hereby the minimum, the mean and the maximum observed concentrations are provided. It is clear that at the points where RRCM, and kriging, modeling has been performed, the real observations are not available. However, those descriptives can serve as indicators for the assessment of the performance of the used models. Table 4.25 shows the descriptive statistics for the observed nitrogen dioxide concentrations for each given year.

Table 4.25: Observed NO₂ concentrations ($\mu\text{g}/\text{m}^3$)

	min	mean	max
1998	11.00	46.50	68.00
1999	17.00	44.40	71.00
2000	13.00	41.20	65.00
2001	23.00	46.16	64.00
2002	25.00	45.24	69.00
2004	22.00	39.88	67.00
2005	21.00	44.23	83.00
2006	19.00	43.96	74.00
2007	21.00	43.40	66.00
2008	18.00	42.36	65.00
2009	11.00	43.12	63.00

Choice of the ridge factor

The efficiency of ridge regression confidence machine predictor to a great extent depends of the choice of the ridge factor. There is no analytical procedure known by now that is used for that purpose, so the brute force method is recommended. In other words, it is recommended to test out various ridge factors until the optimal modeling results will be achieved. The ridge regression procedure suggests plotting the obtained regression coefficients against the ridge factor values, and such a plot is called *ridge trace* [45]. In this research, for each contaminant, year and model a similar plotting method has been used: mean prediction intervals obtained by each model have been plotted against the correspondent ridge factor values. Machine has repeatedly evaluated the mean values of RRCM prediction interval size for each ridge factor value within the interval from 0 to 2 with a step of 0.01. Resulting predictions have been plotted against the ridge factor values. Such plots help elicit the optimal value of the ridge factor that for every particular model yields the most effective prediction, i.e. the smallest mean prediction set. For each contaminant, average annual concentrations are not the same throughout the years, so there is no common value of the ridge factor that would suit all of the models of a kind. In this chapter, all of the RRCM models are demonstrated in one common setting: with the ridge factor equal to 0.01 and the confidence level equal to 95 per cent. Considering the variation in years and data distribution, it is questionable whether this setting is the optimal one for every year and pollutant. While the choice of the confidence parameter value is explained above, and this value is hardly to improve, the choice of the ridge factor can alter the efficiency of prediction. Ridge factor evaluating plots for every year and contaminant can be found in the **Appendix B**.

Two plots are provided below. Figure 4.17 shows the dependence of the mean width of prediction intervals from the ridge factor for the nitrogen dioxide 1999 data. It is visible from the graph that the optimal prediction is obtained when the ridge factor is equal to 1.4.

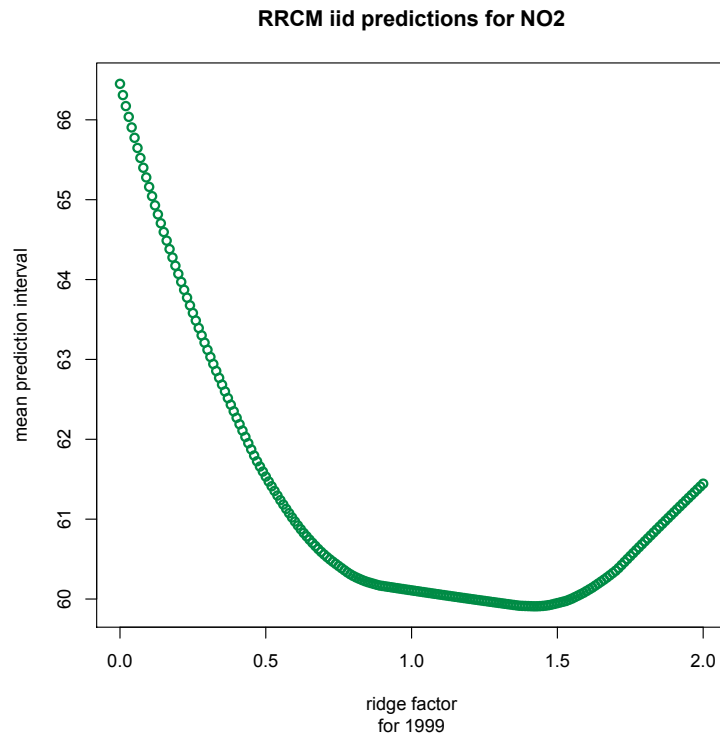


Figure 4.17: Mean prediction interval size and ridge factor

Figure 4.18 depicts how the mean size of the prediction interval changes with the ridge factor value for the 2009 data. It is clear from the plot that the optimal value of the ridge factor is 2 (or may be, a bigger value). Also, it is visible that the character of the dependence is different for both of the years: for the 2009 data set, the mean width of prediction intervals is clearly monotonically decreasing with growth of the ridge factor, while for the 1999 data, the function seems to decrease first, and the further increase, starting from the 1.4 point at the ridge factor scale.

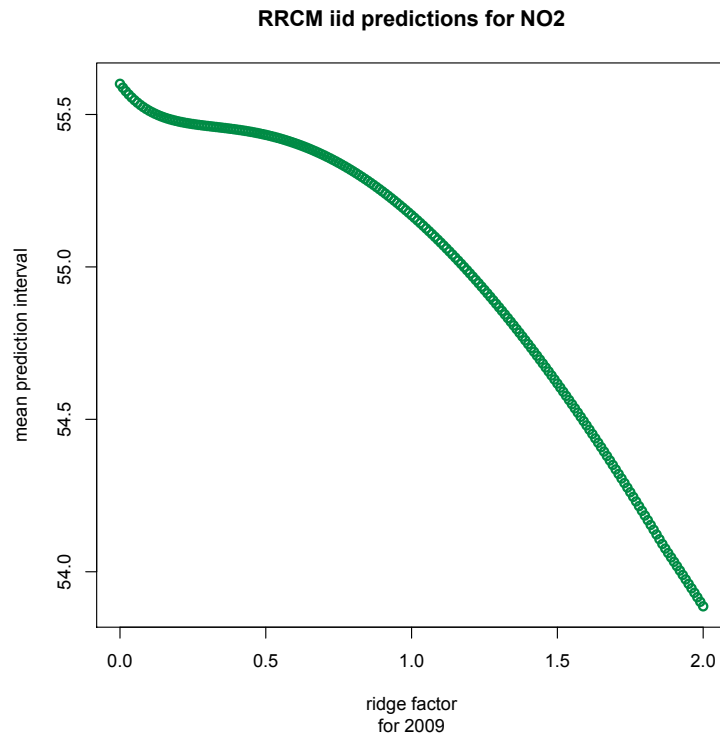


Figure 4.18: Mean prediction interval size and ridge factor

In order to provide some quick visual comparison, Table 4.26 provides the values of the size of the mean prediction intervals for each of the models considered here and a set of ridge factors. This set has been dictated by the common sense, and it is noteworthy that the prediction of the RRCM models with both polynomial and RBF kernels are not computable for the ridge factor equal to 0. In other words, for these specifications it is not possible to obtain least squares estimates. This is why its value of 0.01 has been used instead. For each model specification, the optimal prediction is marked in bold.

Table 4.26: Comparison of models for different ridge factors ($\mu\text{g}/\text{m}^3$)

ridge	linear iid			RBF			polynomial		
	0.01	1	2	0.01	1	2	0.01	1	2
1998	76.08	72.33	68.27	65.81	72.37	68.37	65.27	64.71	65.99
1999	66.31	60.11	61.44	67.68	60.57	60.39	65.32	68.20	70.87
2000	51.69	55.27	57.89	50.91	52.90	55.63	61.89	64.19	66.38
2001	36.25	41.30	44.90	35.32	38.65	42.36	49.54	52.34	54.95
2002	52.12	46.57	49.51	47.78	51.44	57.38	54.51	56.99	59.37
2004	53.65	59.11	62.46	53.89	56.95	60.41	67.06	69.36	71.60
2005	78.75	84.77	88.57	79.44	82.18	86.14	94.41	96.94	99.43
2006	61.79	66.39	69.78	61.24	63.82	67.38	74.90	77.36	79.76
2007	47.01	49.35	53.13	48.15	47.11	51.04	57.15	59.91	62.48
2008	46.96	50.15	53.58	47.45	48.04	51.55	57.63	60.21	62.63
2009	55.59	55.17	53.89	48.38	54.35	52.68	52.79	55.19	57.57

4.9.2 PM10

The models

As for nitrogen dioxide data, for particulate matter the predictive capacities of the models have also been compared. Table 4.27 demonstrates the results of RRCM modeling in plain linear iid setting. For PM_{10} data, the minimum and the maximum values in the relevant tables are marked in bold - for this particular model, and for the two subsequent models. Here, the “best” prediction in terms of efficiency is obtained for 2009, and the “worst” corresponds to 2007. The minimum prediction interval’s lower bound for 2007 is negative. Also, the highest maximum upper bound and the highest maximum and thus the mean size of prediction intervals have correspond to this year, which might indicate that this is sort of an “outlying” year for the data. In terms of mean prediction, the average prediction interval for this model has the width of 51.48 ($\mu\text{g}/\text{m}^3$).

Table 4.27: RRCM iid model for PM10 ($\mu\text{g}/\text{m}^3$)

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
2001	9.18	18.53	61.42	64.46	67.06	74.46	84.93
2002	22.65	32.76	41.41	43.43	50.10	67.82	82.86
2004	18.07	26.59	44.94	47.26	54.87	65.17	79.37
2005	25.37	33.26	37.61	39.65	46.34	63.92	73.47
2006	24.99	30.33	44.68	47.68	53.02	70.85	78.58
2007	-6.67	1.39	87.01	91.43	102.21	87.08	100.25
2008	9.77	17.58	47.05	49.48	58.41	60.60	72.47
2009	19.72	25.35	27.44	28.42	31.56	48.58	56.91

The results of modeling with the use of ridge regression confidence machine with a Gaussian RBF kernel are shown in Table 4.28. The mean prediction interval size that this model yields for this data is $52.62 (\mu\text{g}/\text{m}^3)$, which is slightly bigger than for the model in the linear iid specification.

Table 4.28: RRCM RBF model for PM10 ($\mu\text{g}/\text{m}^3$)

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
2001	4.79	16.75	66.14	71.08	80.05	75.34	88.61
2002	13.78	35.57	42.12	47.41	70.32	68.45	98.79
2004	9.32	25.80	45.33	51.48	77.39	65.19	93.08
2005	28.02	35.89	30.58	35.50	58.10	59.12	86.80
2006	15.61	29.30	48.30	55.51	80.11	73.41	95.73
2007	-15.46	7.35	74.01	85.40	122.28	81.36	113.64
2008	9.52	22.12	38.50	45.42	77.07	56.20	86.81
2009	16.07	24.70	26.16	29.16	42.43	48.87	58.50

RRCM model with the polynomial kernel of the second order has been employed, too, and the results of the modeling are shown in Table 4.29. This specification is denoted by the lowest minimum lower bound for all of the prediction intervals across all the RRCM models ($-18.75 \mu\text{g}/\text{m}^3$). To top it up, this model setting is the only one for which there has appeared a negative value for the maximum lower bound of the prediction intervals ($-6.61 \mu\text{g}/\text{m}^3$). This model generally performs less effectively for the given data than both iid and Gaussian RBF iid specifications.

Table 4.29: RRCM polynomial model for PM10

	min lower bound	max lower bound	min interval width	mean interval width	max interval width	min upper bound	max upper bound
2001	-1.81	5.82	69.22	71.95	78.69	74.16	78.88
2002	13.06	19.38	47.55	50.44	63.89	65.70	79.30
2004	18.56	23.83	32.16	34.66	45.49	54.94	64.04
2005	11.15	19.28	48.15	51.44	65.22	66.54	76.37
2006	13.50	19.00	49.15	52.48	65.49	68.15	78.99
2007	-18.75	-6.61	93.46	99.83	124.23	86.85	105.48
2008	3.86	9.82	52.64	55.60	68.66	61.12	74.19
2009	14.56	17.93	30.37	32.26	39.95	47.86	54.98

Comparison to observed data

As in case of the nitrogen dioxide data, the modeled results unfortunately cannot be directly compared to the observed values at the same locations for particulate matter data. Nevertheless, Table 4.30 shows descriptive statistics for observed concentrations of each year. This information can help assess the comparative performance of the above presented models for this data.

Table 4.30: Observed PM10 concentrations ($\mu\text{g}/\text{m}^3$)

	min	mean	max
2001	33.00	47.82	75.00
2002	28.00	49.29	67.00
2004	26.00	44.36	57.00
2005	31.00	48.61	70.00
2006	29.00	49.24	69.00
2007	33.00	45.93	89.00
2008	29.00	39.94	63.00
2009	25.00	36.19	48.00

Choice of the ridge factor

Similar to the nitrogen dioxide data, the analysis of the optimal value of the ridge factor for the particulate matter data has been made for each year. All of the graphs can be found in the **Appendix B**, and here it will be shortly mentioned that there is no common value of the ridge factor for each of the employed RRCM models that would be optimal for all of the years. That

means that the distribution of the data from year to year is not the same. In order to demonstrate this, two graphs are provided below. Figure 4.19 shows how of the average size of the prediction interval changes dependent on the ridge factor for the RRCM iid model for 2009 data.

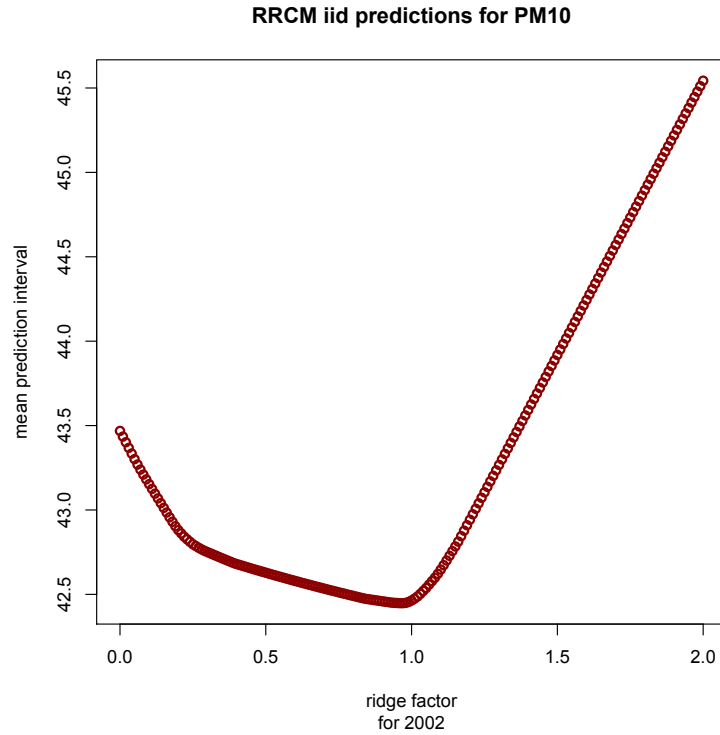


Figure 4.19: Mean prediction interval size and ridge factor

Figure 4.20, in its turn, demonstrates the same for the same model specification, but for 2004 data. It is clear that the character of the dependence is different for both years.

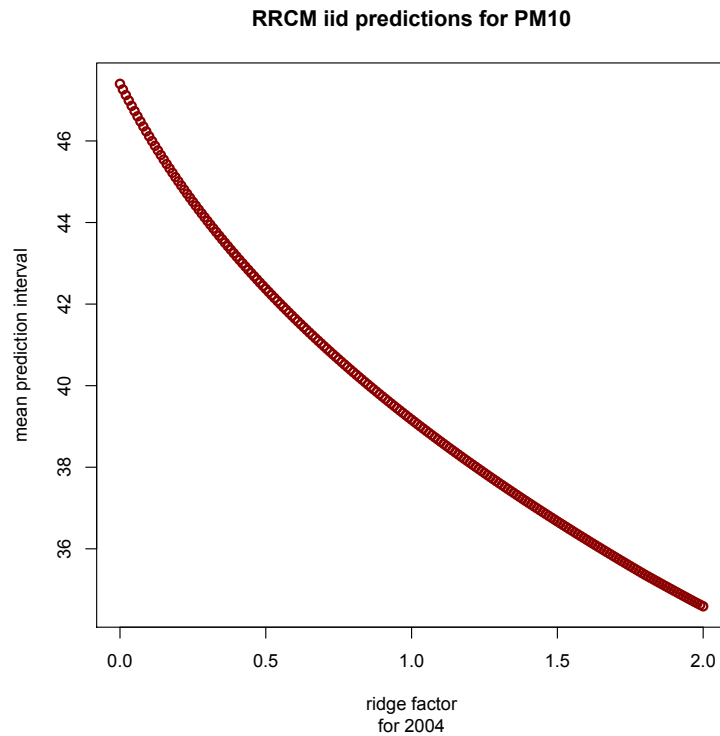


Figure 4.20: Mean prediction interval size and ridge factor

Those graphs have been derived assuming the ridge factor taking every value within the range $[0, 2]$ with a step of 0.01. For the 2002 model, the optimal value of ridge factor is about 0.95, while the optimal value for the parameter for the 2004 model is 2 (or, perhaps, smaller). Table 4.31 presents the differences in modeling results for each of the three RRCM models employed in this research for a set of ridge factors (dictated by common sense): 0.01, 1, 2. It should be mentioned that the RBF and the polynomial specifications of the ridge regression confidence machine are not computable with the ridge factor equal to 0, so the value of 0.01 has been chosen as a “start point”. The optimal predictions for each year and model specification and ridge factor are marked in bold.

Table 4.31: Comparison of models for different ridge factors ($\mu\text{g}/\text{m}^3$)

ridge	linear iid			RBF			polynomial		
	0.01	1	2	0.01	1	2	0.01	1	2
2001	64.46	64.44	67.13	71.08	63.11	66.06	71.95	74.63	77.24
2002	43.43	42.46	45.54	47.41	42.91	45.05	50.44	53.17	55.82
2004	47.26	39.17	34.59	51.48	39.29	35.19	34.66	37.00	39.51
2005	39.65	45.14	49.28	35.50	47.60	51.91	51.44	54.76	57.76
2006	47.68	45.40	48.63	55.51	46.09	48.86	52.48	55.27	57.86
2007	91.43	94.02	96.45	85.40	94.09	96.65	99.83	102.11	104.29
2008	49.48	50.90	52.58	45.42	55.27	58.21	55.60	57.26	58.91
2009	28.42	27.32	29.01	29.16	26.11	27.79	32.26	33.67	35.09

Chapter 5

Discussion

5.1 General notions

On the bottom line, the results have shown the following: conformal predictors can be successfully used for air pollution modeling. In this dissertation, an application of a kriging-based conformal predictor is discussed, and it has provided good predictive capacity. The ultimate conclusion is that this predictor, or its variations, can be treated as an upgrade of kriging. As bringing validity into kriging prediction, it can be employed as an additional validation method. Also, it can perform as an independent prediction technique always when interval output is convenient.

The results have revealed two major points of discussion. First is the goodness of fit of the chosen models. Second is the comparative efficiency of predictions imparted by kriging and ridge regression confidence machine. Based on the results, the predictive capacity of the models is still to be explained prior to advocating the use of these models for other data sets.

5.2 Fitting the models

First of all, goodness of fit should be explained. It can be described in common terms for both methods used in this work since these models are based on the same geostatistical paradigm. This study uses two major approaches: ordinary kriging and ridge regression confidence machine. Ordinary kriging is the specification of the kriging methods that is to be used to handle raw data with unknown mean. The present data set is represented by mean annual concentrations of two pollutants, nitrogen dioxide and fine particles, within several years. Considering a great share of missing values for each year and pollutants, and a scarce number of spatial points where data is observed, it can be concluded that the means are unknown. Moreover, it

is complicated to approach data distributions for each year with theoretical variograms. As it is often done in practice [5], variograms have been chosen without a particular justification. Three variograms have been used: linear, exponential and Gaussian. Graphical representations of variogram fits divulge that linear approach generally performs equally or worse as an exponential function. Also, exponential model is the default one for conventional kriging prediction in the **geoR** package [48]. Ordinary kriging with exponential function has been performed. Overall, it has yielded smooth predictions with huge variance estimates for nitrogen dioxide and smooth predictions with small variance estimates for particulate matter. Smoothness of prediction is a common kriging problem, since it does not take into account local pollution sources. As for the big difference in variance estimates for two pollutants, nitrogen dioxide is prone to strong spatial variability, whereas for particulate matter the variability is small [12]. Approximations with other variomodels can be done and have been done for Barcelona Metropolitan Region before [33]. Nonetheless, this study makes use of very little data, and these are annual average concentrations with up to 50 per cent of missing values, so it is doubtful that a model other than the suggested ones could provide a remarkably better approach.

Ridge regression predictor in its basic iid setting is aimed to be opposed to kriging models with linear and exponential covariance functions. Instead of point predictions, it imparts intervals. This predictor considers variomodels with the use of “kernel trick”. The standard iid model uses dot product as a kernel, and it can be seen as an analogue of a linear variomodel. An exponential kernel could be practically implemented, too, but it would not significantly boost the efficiency of prediction with this data. On almost all variogram plots, exponential curves are flat as straight lines. As for non-linear approximation, Gaussian functions serve this purpose. Ordinary kriging with Gaussian covariance function has been performed. Ridge regression predictor with the Gaussian RBF kernel is the counterpart approach of the kriging models. Gaussian covariance function fits the data generally better than linear and exponential functions do, but, nevertheless, spatial distribution can vary for each year and pollutant. Also, an inhomogenous polynomial kernel of the second order has been employed in RRCM. In the present data set, the concentrations of pollutants are annual averages, so it is hard to establish the initial data distribution for every year.

Table 5.1 summarizes the main features of ordinary kriging and ridge regression confidence machine.

Table 5.1: Comparison of ordinary kriging and ridge regression confidence machine

OK	RRCM
point predictions	prediction sets (usually intervals)
regression algorithm	regression algorithm
Gaussianity assumption	iid assumption
estimates error variance	-
uses variogram and covariance function to approach it	uses any appropriate kernel
-	ridge factor
may lack confidence	confidence level is chosen and guaranteed

All in all, when using both of the approaches, ordinary kriging and ridge regression confidence machine, for spatial prediction, in case of full data availability, preliminary data analysis cannot be neglected. It can help establish a proper variomodel which would yield better predictions. In case of this particular data set, however, finding a proper variomodel is challenging.

5.3 Efficiency of predictions

Second issue to clarify is the comparative efficiency of ridge regression and ordinary kriging models. Kriging predictions are smooth and predicted values vary to a small extent. Error estimates, however, are huge in case of nitrogen dioxide, and small in case of airborne particles. These estimates help come up with 95 % confidence intervals for the predicted values. Results generally suggest that RRCM prediction intervals are of comparable size or larger than these kriging confidence intervals. Nevertheless, two major subjects are to dispute. First of all, kriging confidence intervals can be derived only under the assumption that the data is a realization of a Gaussian spatial process. This is a common assumption in geostatistics [32, 54], but it is not always correct. The specification of a ridge regression conformal predictor, the RRCM iid model, does not make any assumption regarding data distribution, apart from the iid property. This is an important benefit of ridge regression conformal predictors in comparison to kriging. Second is that kriging confidence intervals, left alone the Gaussianity assumption, are based on kriging variances, which are, in their turn, estimates themselves. No real observations are available for the points where kriging is performed, thus real errors of prediction cannot be established, too. Ordinary kriging

capacity would be dependent on variomodel fit only, and thus validity of prediction cannot be guaranteed. Cross-validation can be made, and is made in this research, but the size of the data set is insufficient to provide solid validation capacities for kriging models. When it comes to ridge regression confidence machine prediction intervals, they are always valid, by definition of conformal predictors.

Although RRCM intervals are valid predictions, they are relatively big, as the results claim. Nevertheless, ridge regression confidence machine can make use of two adjustment tools to boost efficiency of its predictions, besides the variomodel fit, or kernel. These tools are: ridge factor and confidence level. RRCM allows to choose the value of these parameters. When it comes to confidence, a researcher is free to choose almost any convenient confidence level. Naturally, the highest confidence is preferred. However, if the iid setting of RRCM is taken up, it is important to mention that it puts the following restriction on the value of the confidence level: for a chosen confidence level $1 - \epsilon$, where ϵ is the significance level, or the maximum error probability, the data set must count at least $1/\epsilon$ observations to deliver valid predictions [38]. In other words, a data set composed of 20 observations is sufficient to provide 95% confidence. In this research, the confidence of prediction has been set to 95 per cent. This choice has been dictated by the average size of the datasets available for every year.

As for ridge parameter, considering that the data has been scaled and normalized (as the ridge regression procedure suggests), the optimal ridge factor has been picked from the interval from 0 to 2 by brute force method. For each year and pollutant, for each year, plots similar to ridge traces [45] have been derived to elicit the optimal values of the ridge factor.

5.4 Bottom line

In the main, with validity and efficiency of RRCM predictions explained, regarding the limitations of the study data set, optimal models for each year and pollutant can be produced. The best result for nitrogen dioxide, in terms of mean prediction interval size, is $36 \text{ } (\mu\text{g}/\text{m}^3)$. Compared to the overall distribution of nitrogen dioxide at the BMR for 11 years, this interval is equal to the first quartile of the distribution, and it is smaller than the mean observed NO_2 concentration ($43.67 \text{ } \mu\text{g}/\text{m}^3$). On the average, the size of optimal prediction intervals for NO_2 for each year is from 58 to 68 $(\mu\text{g}/\text{m}^3)$. It is smaller than the observed maximum concentration ($83 \text{ } \mu\text{g}/\text{m}^3$). Compared to the guideline values recommended by WHO for annual concentrations of nitrogen dioxide ($40 \text{ } \mu\text{g}/\text{m}^3$) [12], this interval could

house this value 1.5 times. For PM_{10} , the size of the best average prediction interval was equal to $26.5 \text{ } (\mu\text{g}/\text{m}^3)$. Compared to the overall distribution of particulate matter, this interval is slightly bigger than the minimum observed concentration ($25 \text{ } \mu\text{g}/\text{m}^3$), but it is smaller than the mean observed concentration ($44.6 \text{ } \mu\text{g}/\text{m}^3$). For all the years, the optimal prediction set width varies between 53 and 63 ($\mu\text{g}/\text{m}^3$). Compared to the guideline annual concentrations recommended by WHO ($20 \text{ } \mu\text{g}/\text{m}^3$), this average interval is three times bigger.

On the basis of the results, it can be concluded that RRCM iid model and its derivations with kernels are a suitable approach to assess air pollution exposure data. The only desideratum for further development of this research and therefore the models is: more data. A denser monitoring network and, perhaps, observations with higher frequency than annual means are needed for better variomodel fits.

Chapter 6

Conclusion

6.1 The subject of this study

The present dissertation uses the newly developed approach of conformal predictors to evaluate concentrations of two air pollutants for the Barcelona Metropolitan Region for several years. Also, it suggest that conformal predictors can be seen as a good method for spatial prediction. This work compares the application of a specification of conformal predictors, called ridge regression confidence machine (RRCM) [1], to a classical geostatistical method, ordinary kriging. After employing both methods, a conclusion is made that RRCM can be seen as an additional and/or complementary method to ordinary kriging.

In this research, conformal predictors have been used to predict the pollutants concentrations for the Barcelona Metropolitan Region (BMR). This geographic region is situated in the UTM31 time zone, and the study dataset is composed of the GIS data. Those are the observations of two pollutants, nitrogen dioxide (NO_2) and particulate matter with the aerodynamic diameter of 10 micrometers (PM_{10}), and they are provided together with the geographical coordinates of the measurement stations. For NO_2 , the timeline of the study embraces the period from 1998 to 2009, with the exception of 2003. For PM_{10} , the time period is from 2001 to 2009, also with the exception of 2003. The concentrations of pollutants in the dataset are the mean annuals, and there are 49 monitoring sites across the region. However, not for every station and year and pollutant the observations are available. The average number of the available observations for nitrogen dioxide is 24, while for particulate matter it is 22. For every year, the concentrations of pollution have been predicted for the unobserved measurement spots. Those prediction can be expanded to a grid over the whole Barcelona Metropolitan Region, of course. The predictions have been made in such a way because thus they can be compared with the observed values for the other years

(when there are such). The data for the study has been kindly provided by the XVPCA of the Generalitat of Catalonia.

Taking into consideration the limitations imposed by the data, conformal predictors have performed as a good prediction method to tackle the given problem. For future investigation, the main goal is, of course, to increase the efficiency of prediction. In order to do so, a bigger data set must be available. Then, other kernels and covariance functions can be implemented if they would work well for the data.

6.2 Conformal predictors and geostatistics

Ridge regression confidence machine method is based on the ridge regression procedure [45], but instead of providing a point estimate of the value of factor of interest, it outputs its prediction in the form of a prediction set, which, in a specific case, can be a point, too. A prediction set can be a ray, a union of two rays, the whole real line, a point, or an interval. Practically, it is almost always an interval. It is desired to obtain the smallest possible intervals.

An outstanding hallmark of a conformal predictor is its flexibility: it can be build upon almost any machine learning (or, statistical) algorithm. Provided with a confidence level, a newly developed predictor inherits all the predictive power of its underlying algorithm, but it is also always valid.

In this work, ridge regression confidence machine is contrasted with the classical geostatistical approach that is kriging [4, 5]. Kriging is a well-developed procedure, which is famous for its capacity to provide the measure of uncertainty together with the prediction [4]. This measure is called kriging error of kriging variance. However, some drawbacks of this approach are known. When it comes to air pollution assessment, the main criticism towards kriging is that it ignores the local pollution sources such as traffic arteries. Kriging comes up with a smooth prediction surface, and the prediction at each given point may be incorrect. The algorithm takes up observed values at a set of points where measurement stations are situated, and it performs an interpolation procedure to output the predicted value for an unobserved point. The predictions can be spread to a grid of any size and density. As its adjustment tool, it uses a variogram model. It allows to introduce spatial variability into the estimation equation, but it is fitted on the basis of the observed points only. Thus, the local variability of air pollution is not mentioned. Nevertheless, the unavailability of monitoring data is a general problem for geostatistical interpolation methods [7].

Scientific evidence infers that the average number of monitoring sites for a geographic study region is between 10 to 100 stations. Poor network of available sampling sites may provide over-smoothed pollution surfaces and large estimation errors.

Furthermore, a definite letdown of the ordinary kriging method is that it lacks validity in its estimates. Kriging is an exact interpolator which implies that it outputs the observed values of the factor of interest at the observed locations, while the measure of uncertainty for the unobserved locations, i.e. the kriging variance, is itself an estimate [4].

Conformal predictors help overcome the lack of validity problem. RRCM, like kriging, is a regression method, and it can be applied in a similar way. It can take the observed locations as a set of its objects, where the concentration of a given pollutant at a given spot is the label of the corresponding object. Unlike kriging, that provides point estimates, RRCM yields valid intervals as its prediction. These intervals are valid, as they contain the real value with a given level of confidence.

Prediction intervals can be, however, quite wide, and it is naturally preferred to come up with the smallest possible intervals. In order to achieve this goal, RRCM suggest two three expedients. First of all, the confidence level can be abated, but this is perhaps not desired. Secondly, a ridge parameter can be adjusted, which is usually done by a brute force method. Another solution is the use of kernels. A non-linear setting can be considered with the help of so-called “kernel trick” [6]. It has been proposed by Boser, Guyon and Vapnik for non-linear support vector machines, and it is aimed to deal with high-dimensional problems [1]. The object space is mapped to a so-called feature space by means of a kernel function, and the ridge regression procedure is performed in the feature space.

Kernel trick can be applied to RRCM in a similar way a a covariance function is applied in kriging, i.e. with the intention to consider spatial distribution and spatial covariance between observations. For modeling, the empirical variograms are replaced by theoretical ones, and this fit is done by eye [5]. There are several classes of covariance functions that are implemented on practice: the Matérn family, spherical, “nugget effect” etc. The main point is that they are positive definite, since it is the necessary and sufficient condition for a parametric family of functions to be legitimate to define a class of covariance functions [32]. This condition allows using covariance functions as kernels in ridge regression procedure and vice versa.

6.3 Key points and further development of this work

Conformal prediction is a newly developed approach. Coming from machine learning, this paradigm can be successfully applied to epidemiological studies. As providing confidence predictions, conformal predictors can be of great use for clinical research, because it is dealing with health and well-being of people. Air pollution has been associated with adverse effects in people [14, 12, 18, 15], and it is known that the grade and the duration of exposure to contamination determines the occurrence of certain outcomes. Unlike the common methods, conformal predictors are capable of providing exposure estimates with confidence, and the levels of confidence can be very high.

This research shows that conformal predictors can be used for geostatistical modeling, and that it can handle GIS data well. Nearly any good statistical air pollution assessment algorithm can be transformed to a conformal predictor, and it will both inherit the predictive power of the underlying technique and yield estimates with confidence. Thus, popular air pollution assessment models, such as land use regression [36], can be turned into conformal predictors. Those methods would be feasible subject to availability of relevant data, such as, for example, land use data. Bayesian models, such as trans-Gaussian kriging [32], can be also altered to follow conformal prediction paradigm [1].

Overall, with practically any regression model allotted for spatial prediction, the aim of converting it to a conformal predictor is to obtain not only valid, but also efficient prediction intervals. The efficiency of prediction mostly depends on the correct specification of the underlying algorithm and on the data set size. On the bottom line, there are many ways to obtain efficient predictions. Therefore, as being a valid, efficient and extremely flexible method, conformal predictors are concluded to be a beneficial method for air pollution assessment and, generally, for spatial prediction.

Popular algorithms that are successfully applied in air pollution investigation practice, such as land use regression (LUR) or dispersion models [7], can be transformed to conformal predictions, provided with a significance level for prediction. LUR is a linear regression model that makes use of land use data, e.g. traffic density. A regression conformal predictor can be easily derived on the basis of a LUR algorithm, and regression residuals will serve as a nonconformity measure. Moreover, when the independent variables are not heavily correlated, ridge factor can be omitted, and a plain least squares predictor would work out. Dispersion models are based on Gaussian plume equations that represent air pollution concentrations via complex exponen-

tial function of meteorological parameters and height above the sea level. A transformation of a dispersion model into a informal predictor is not as straightforward as for LUR. It is certainly doable, but it requires some specific focused investigation aimed to determine a nonconformity measure.

Derivation of conformal predictors on the basis of well-developed air pollution modeling algorithms is one of important directions for future research. Another important thing to do is to further develop kriging-based predictors. This would imply two major issues. First of all, various covariance functions should be used as kernels for kriging-based conformal predictors. Secondly, Bayesian approach to geostatistical modeling should be taken up, and a Bayesian kriging-based conformal predictor should be developed. Finally, all the models should be tested on various data sets. The Barcelona data set is to be amplified soon. It will help derive new models for Barcelona's data and also adjust the covariance function parameters for the models that have already been developed, and therefore yield more efficient predictions. Nevertheless, data for one city is not enough to validate the models, and they should be tested on data sets for other cities, too.

Bibliography

- [1] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer, 2005. ISBN: 9780387001524.
- [2] I. Nourtdinov, T. Melliush, and V. Vovk. “Ridge regression confidence machine”. In: *In Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufman, 2001, pp. 385–392.
- [3] A. Hoerl and R. Kennard. “Ridge regression: biased estimation for nonorthogonal problems”. In: *Technometrics* 12.1 (1970), pp. 55–67.
- [4] P. Switzer. “Kriging”. In: *Encyclopedia of Environmetrics*. John Wiley & Sons Ltd, 2006. ISBN: 9780470057339. DOI: 10.1002/9780470057339.vak003.
- [5] H. Wackernagel. *Multivariate geostatistics: an introduction with applications*. Springer, 2003. ISBN: 9783540441427.
- [6] B. Schölkopf and A.J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, 2002. ISBN: 9780262194754.
- [7] M. Jerrett et al. “A review and evaluation of intraurban air pollution exposure models”. In: *Journal of Exposure Analysis and Environmental Epidemiology* 15 (2005), pp. 185–204.
- [8] M. Kampa and E. Castanas. “Human health effects of air pollution”. In: *Environmental Pollution* 151.2 (2008), pp. 362–367. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2007.06.012.
- [9] D. van Leeuwen et al. “Genomic analysis suggests higher susceptibility of children to air pollution.” In: *Carcinogenesis* 29.5 (2008), pp. 977–83. ISSN: 1460-2180.
- [10] M. Lloyd-Smith and B. Sheffield-Brotherton. “Children’s environmental health: intergenerational equity in action - a civil society perspective”. In: *Annals of the New York Academy of Sciences* 1140.1 (2008), pp. 190–200. ISSN: 1749-6632.
- [11] D. Dalbokova et al. *Children’s health and the environment in Europe: a baseline assessment*. World Health Organization, 2007. ISBN: 9789289072977.

- [12] World Health Organization. *Air quality guidelines: global update 2005 : particulate matter, ozone, nitrogen dioxide and sulfur dioxide*. EURO Nonserial Publication. World Health Organization, 2006. ISBN: 9789289021920.
- [13] I. Aguilera et al. “Prenatal exposure to traffic-related air pollution and ultrasound measures of fetal growth in the INMA sabadell cohort”. In: *Environmental Health Perspectives* 118.5 (Jan. 2010). DOI: 10.1289/ehp.0901228.
- [14] World Health Organization, World Health Organization. European Centre for Environment, and Bonn Office Health. *Effects of air pollution on children’s health and development: a review of the evidence*. World Health Organization, 2005.
- [15] N. Vichit-Vadakan, N. Vajanapoom, and B. Ostro. “Part 3. Estimating the effects of air pollution on mortality in Bangkok, Thailand.” In: *Res Rep Health Eff Inst* 154 (2010), pp. 231–68. ISSN: 1041-5505.
- [16] A. Lindgren et al. “Traffic-related air pollution associated with prevalence of asthma and COPD/chronic bronchitis. A cross-sectional study in Southern Sweden”. In: *International Journal of Health Geographics* 8.1 (2009), p. 2. ISSN: 1476-072X. DOI: 10.1186/1476-072X-8-2.
- [17] C. Pope et al. “Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution”. In: *JAMA: The Journal of the American Medical Association* 287.9 (2002), pp. 1132–1141. DOI: 10.1001/jama.287.9.1132.
- [18] J. Sacks et al. “Particulate matter-induced health effects: who is susceptible?” In: *Environmental Health Perspectives* 119.4 (Oct. 2010), pp. 446–54. DOI: 10.1289/ehp.1002255.
- [19] Barceló M.A., Saez M., and Saurina C. “Spatial variability in mortality inequalities, socioeconomic deprivation, and air pollution in small areas of the Barcelona Metropolitan Region, Spain”. In: *Science of The Total Environment* 407.21 (2009), pp. 5501–5523. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2009.07.028.
- [20] N. Künzli et al. “Public-health impact of outdoor and traffic-related air pollution: a European assessment”. In: *The Lancet* 356.9232 (2000), pp. 795–801. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(00)02653-2.
- [21] O. Ivina, M. Saez, and M.A. Barceló. “Do socioeconomic factors modify the effects of traffic-related air pollution on health at within-city scale?” In: *Lomonosov-2010*. 2010. ISBN: 9785317031978.
- [22] European Space Agency. *Global air pollution map produced by Envisat’s SCIAMACHY*. 2004. URL: http://www.esa.int/esaCP/SEM340NKPZD_UnitedKingdom_0.html.

- [23] S. Menon et al. “Climate effects of black carbon aerosols in China and India”. In: *Science* 297.5590 (2002), pp. 2250–2253. DOI: 10.1126/science.1075159.
- [24] V. Ramanathan et al. “Aerosols, climate, and the hydrological cycle”. In: *Science* 294.5549 (2001), pp. 2119–2124. DOI: 10.1126/science.1064034.
- [25] U.S. Environmental Protection Agency. *NO_x. How nitrogen oxides affect the way we live and breathe*. 1998. URL: <http://www.epa.gov/airquality/nitrogenoxides/health.html>.
- [26] D. Doll, United States. Environmental Protection Agency. Office of Air Quality Planning, and Standards. *Our nation’s air: status and trends through 2008*. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, 2010.
- [27] U.S. Environmental Protection Agency. *Health-effects of NO₂ air pollution*. 2011. URL: <http://www.epa.gov/airquality/nitrogenoxides/health.html>.
- [28] M. Rosenlund et al. “Traffic-related air pollution in relation to incidence and prognosis of coronary heart disease”. In: *Epidemiology* 19.1 (2008), pp. 121–8. ISSN: 1044-3983.
- [29] European Commission. *Air Quality Standards*. 2012. URL: <http://ec.europa.eu/environment/air/quality/standards.htm>.
- [30] U.S: Environmental Protection Agency. *Particulate Matter*. 2011. URL: <http://www.epa.gov/pm/>.
- [31] European Environmental Agency. *Particulate matter (PM₁₀) 2009. Annual limit value for the protection of human health*. 2011. URL: <http://http://www.eea.europa.eu/data-and-maps/figures/particulate-matter-pm10-2007-annual-limit-value-for-the-protection-of-human-health-3>.
- [32] P. Diggle and P. Ribeiro Jr. *Model-based geostatistics*. Springer, 2007. ISBN: 9780387329079.
- [33] A. Lertxundi-Manterola and M. Saez. “Modelling of nitrogen dioxide (NO₂) and fine particulate matter (PM₁₀) air pollution in the metropolitan areas of Barcelona and Bilbao, Spain”. In: *Environmetrics* 20.5 (2009), pp. 477–493. ISSN: 1099-095X. DOI: 10.1002/env.939.
- [34] R. Webster and M.A. Oliver. *Geostatistics for environmental scientists*. Wiley, 2007. ISBN: 9780470028582.
- [35] J.P. Chilès and P. Delfiner. *Geostatistics: modeling spatial uncertainty*. Wiley series in probability and statistics. Wiley, 1999. ISBN: 9780471083153.

- [36] J. Clougherty et al. “Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants”. In: *Environmental Health* 7.1 (2008), p. 17.
- [37] J. Cyrus et al. “GIS-based estimation of exposure to particulate matter and NO₂ in an urban area: stochastic versus dispersion modeling”. In: *Environmental Health Perspectives* 113.8 (2005), pp. 987–92. ISSN: 0091-6765.
- [38] V. Vovk, I. Nouretdinov, and A. Gammerman. “On-line predictive linear regression”. In: *Annals of Statistics* 37.3 (2009), pp. 1566–1590.
- [39] A.E. Hoerl and R. W. Kennard. “Ridge regression: applications to nonorthogonal problems”. In: *Technometrics* 12.1 (1970), p. 69.
- [40] M. Saez et al. “A combined analysis of the short-term effects of photochemical air pollutants on mortality within the EMECAM project”. In: *Environmental Health Perspectives* 110.3 (2002), pp. 221–228. ISSN: 0091-6765.
- [41] D.G. Krige. “A statistical approach to some basic mine valuation problems on the witwatersrand”. In: *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52.6 (1951), pp. 119–139.
- [42] N. Cressie. “The origins of kriging”. In: *Mathematical Geology* 22 (3 1990). 10.1007/BF00889887, pp. 239–252. ISSN: 0882-8121.
- [43] N. Cressie. “Geostatistics”. In: *The American Statistician* 43.4 (1989), pp. 197–202. ISSN: 00031305.
- [44] N. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993. ISBN: 9780471002550.
- [45] N.R. Draper and H. Smith. *Applied regression analysis*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, 1981. ISBN: 9780471029953.
- [46] L. Kantorovich. “The method of successive approximations for functional equations”. In: *Acta Mathematica* 71 (1 1939), pp. 63–97. ISSN: 0001-5962.
- [47] R Development Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing, 2011. URL: <http://www.R-project.org>.
- [48] P. Ribeiro Jr. and P. Diggle. “geoR: a package for geostatistical analysis”. In: *R-NEWS* 1.2 (2001). ISSN: 1609-3631. URL: <http://cran.r-project.org/doc/Rnews>.

- [49] A. Cebollada and C. Miralles-Guasch. “La movilidad en la región metropolitana de Barcelona: entre los nuevos retos y las viejas prácticas”. In: *Finisterra : Revista Portuguesa de Geografia XLV* (90 2010), pp. 33–47. ISSN: 04305027.
- [50] J. Trullén, R. Boix, and V. Galletto. “La metròpoli de Barcelona : territori”. In: *Papers: Regió Metropolitana de Barcelona: Territori, estratègies, planejament* 54 (2011), pp. 52–60. ISSN: 1888-3621.
- [51] *Enquesta de la Mobilitat Quotidiana de Catalunya 2006. Regió Metropolitana de Barcelona*. 2006.
- [52] D.M. Dziuda. *Data mining for genomics and proteomics: analysis of gene and protein expression data*. Wiley Series on Methods and Applications in Data Mining. Wiley, 2010. ISBN: 9780470163733.
- [53] G. Cawley and N. Talbot. *Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers*. 2003.
- [54] J. Pilz and G. Spöck. “Why do we need and how should we implement Bayesian kriging methods”. In: *Stoch. Env. Res. Risk A*. 22.5 (Aug. 2008), pp. 621–632.

Appendix A

R functions code

iid_dual function

iid_dual is an R [3] function that has been developed on the basis of the *iidpred* function from the **PredictiveRegression** package [5, 1]. This is an auxiliary function aimed to present the ridge regression equation (2.26) in the *dual form* (2.32 - 2.34) [4]. The duality approach is needed to further implement the “*kernel trick*”. To create this function, only a part of the initial function *iidpred* code has been modified. The initial unmodified code is marked in grey.

```
> function(train, test, epsilons = c(0.05, 0.01), ridge = 0) {
+   N <- dim(train)[1]
+   K <- dim(train)[2] - 1
+   N2 <- dim(test)[1]
+   K2 <- dim(test)[2]
+   flag <- 0
+   up <- array(Inf, c(N2, length(epsilons)))
+   low <- array(-Inf, c(N2, length(epsilons)))
+   if (K2 != K) {
+     flag <- 1
+     return(list(low, up, flag))
+   }
+   ZZ <- train[, 1:K]
+   dim(ZZ) <- c(N, K)
+   ZZ <- cbind(1, ZZ)
+   dim(ZZ) <- c(N, K + 1)
+   yy <- train[, K + 1]
+   dim(yy) <- c(N, 1)
+   ZZ2 <- cbind(1, test)
+   dim(ZZ2) <- c(N2, K + 1)
```

```

+   A <- array(0, c(N + 1, 1))
+   B <- array(0, c(N + 1, 1))
+   toinvert <- array(0, c(K + 1, K + 1))
+   inverseZZ <- array(0, c(K + 1, N + 1))
+   P <- array(0, c(2 * N + 2, 1))
+   NM <- array(0, c(2 * N + 2, 1))
+   for (n2 in 1:N2) {
+     P[1] <- -Inf
+     sizeP <- 1

+     ZZ_ext <- rbind(ZZ, ZZ2[n2, ])
+     yy0 <- rbind(yy, 0)
+     OO1 <- rbind(array(0, c(N, 1)), 1)
+     KO <- ZZ_ext %*% t(ZZ_ext)
+     inverseZZ <- solve(KO + ridge * diag(N + 1))
+     H <- inverseZZ %*% KO
+     A <- yy0 - H %*% yy0
+     B <- OO1 - H[, N + 1]

+     A[B < 0] <- -A[B < 0]
+     B[B < 0] <- -B[B < 0]
+     L <- 0
+     R <- 0
+     for (n in 1:N) {
+       if (B[n] != B[N + 1]) {
+         point1 <- (A[n] - A[N + 1]) / (B[N + 1] - B[n])
+         point2 <- -(A[n] + A[N + 1]) / (B[N + 1] + B[n])
+         P[sizeP + 1] <- min(point1, point2)
+         P[sizeP + 2] <- max(point1, point2)
+         if (B[n] < B[N + 1]) {
+           NM[sizeP + 1] <- 1
+           NM[sizeP + 2] <- -1
+         }
+         else {
+           NM[sizeP + 1] <- -1
+           NM[sizeP + 2] <- 1
+           L <- L + 1
+           R <- R + 1
+           if (point1 == point2) {
+             sizeP <- sizeP - 2
+           }
+         }
+       }
+     }
+     sizeP <- sizeP + 2

```

```

+         }
+     else {
+         if (A[n] == A[N + 1]) {
+             L <- L + 1
+             R <- R + 1
+         }
+         else {
+             if (B[N + 1] != 0) {
+                 point1 <- -(A[n] + A[N + 1]) / (B[N + 1] + B[n])
+                 if (A[n] > A[N + 1]) {
+                     NM[sizeP + 1] <- 1
+                     R <- R + 1
+                 }
+                 else {
+                     NM[sizeP + 1] <- -1
+                     L <- L + 1
+                 }
+                 sizeP <- sizeP + 1
+             }
+             else {
+                 if (A[n] > A[N + 1]) {
+                     L <- L + 1
+                     R <- R + 1
+                 }
+             }
+         }
+     }
+ }
+
+ P[sizeP + 1] <- Inf
+ sizeP <- sizeP + 1
+ NM[1] <- L + 1
+ NM[sizeP] <- -R - 1
+ P_order <- order(P, -NM)
+ P <- P[P_order]
+ NM <- NM[P_order]
+ eps_order <- order(epsilons)
+ p <- 0
+ P_reached <- 1
+ for (eps_index in 1:length(epsilons)) {
+     eps_real_index <- eps_order[eps_index]
+     epsilon <- epsilons[eps_real_index]
+     found <- FALSE
+     for (P_index in P_reached:sizeP) {
+         p <- p + NM[P_index] / (N + 1)

```

```

+         if (p > epsilon) {
+             found <- TRUE
+             low[n2, eps_real_index] <- P[P_index]
+             P_reached <- P_index
+             p <- p - NM[P_index]/(N + 1)
+             break
+         }
+     }
+     if (!found) {
+         for (eps_ind in eps_index:length(epsilons)) {
+             eps_real_ind <- eps_order[eps_ind]
+             low[n2, eps_real_ind] <- Inf
+             up[n2, eps_real_ind] <- -Inf
+         }
+         break
+     }
+ }
+ p <- 0
+ P_reached <- sizeP
+ for (eps_index in 1:length(epsilons)) {
+     eps_real_index <- eps_order[eps_index]
+     epsilon <- epsilons[eps_real_index]
+     found <- FALSE
+     for (P_index in P_reached:1) {
+         p <- p - NM[P_index]/(N + 1)
+         if (p > epsilon) {
+             found <- TRUE
+             up[n2, eps_real_index] <- P[P_index]
+             P_reached <- P_index
+             p <- p + NM[P_index]/(N + 1)
+             break
+         }
+     }
+     if (!found) {
+         cat("You have found a bug in the program.\n")
+         cat("Please contact the package's maintainer.\n")
+     }
+ }
+ }
+ max_eps <- max(epsilons)
+ if (N + 1 < 1/max_eps)
+     flag <- 2
+ list(low, up, flag)
+ }

```

***iid_rbf* function**

This function has been created in the same manner as the *iid_dual*. It allows to calculate the ridge regression equation, implementing the "kernel trick" and considering the Gaussian RBF kernel [2]. Only the modified part of the code is presented below.

```
> ZZ_ext <- rbind(ZZ, ZZ2[n2, ])
> yy0 <- rbind(yy, 0)
> OO1 <- rbind(array(0, c(N, 1)), 1)
> K0 <- ZZ_ext %*% t(ZZ_ext)
> K1 <- matrix(diag(K0), dim(K0)[1], dim(K0)[2])
> D <- K1 + t(K1) - (K0 + K0)
> RBF <- exp(-D/2)
> inverseZZ <- solve(RBF + ridge * diag(N + 1))
> H <- inverseZZ %*% RBF
> A <- yy0 - H %*% yy0
> B <- OO1 - H[, N + 1]
```

***iid_polyn_2* function**

iid_polyn_2 function has been developed similarly, as substituting the same part of the initial code with a new chunk. This function allows to consider the inhomogenous of the second order [4] in the regression equation.

```
> ZZ_ext <- rbind(ZZ, ZZ2[n2, ])
> yy0 <- rbind(yy, 0)
> OO1 <- rbind(array(0, c(N, 1)), 1)
> NA
> K0 <- (diag(N + 1) + ZZ_ext %*% t(ZZ_ext))^2
> inverseZZ <- solve(K0 + ridge * diag(N + 1))
> H <- inverseZZ %*% K0
> A <- yy0 - H %*% yy0
> B <- OO1 - H[, N + 1]
```

Bibliography

- [1] I. Nourtdinov, T. Melliush, and V. Vovk. “Ridge Regression Confidence Machine”. In: *In Proceedings of the Eighteenth International Conference on Machine Learning*. Morgan Kaufman, 2001, pp. 385–392.
- [2] B. Schölkopf and A.J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, 2002. ISBN: 9780262194754. URL: <http://books.google.com/books?id=y8ORL3DWt4sC>.
- [3] R Development Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing, 2011. URL: <http://www.R-project.org>.
- [4] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer, 2005. ISBN: 9780387001524. URL: <http://books.google.com/books?id=pEXc4C1ymakC>.
- [5] V. Vovk, I. Nourtdinov, and A. Gammerman. “On-Line Predictive Linear Regression”. In: *Annals of Statistics* 37.3 (2009), pp. 1566–1590.

Appendix B

Detailed results

NO₂

1998

Fitting linear and exponential variomodels:

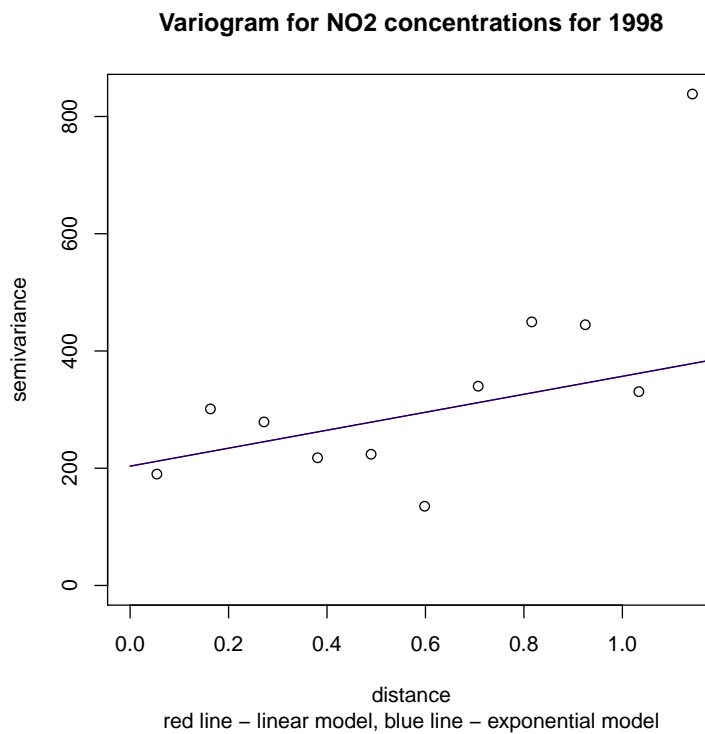


Figure 1: Empirical and modeled variograms

In Table 1, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 1: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
48.70	242.90	8.32	85.38	77.06
48.64	243.32	8.38	85.72	77.34
44.14	275.31	-0.55	79.91	80.46
44.26	274.23	-0.35	79.94	80.28
48.60	244.66	6.76	81.17	74.41
44.24	273.98	-0.53	80.20	80.73
48.39	241.07	8.87	82.82	73.95
48.49	242.27	7.89	81.59	73.71
47.69	238.44	10.77	86.33	75.56
47.35	236.96	11.04	85.96	74.92
47.67	242.75	6.79	80.75	73.97
46.53	243.13	5.30	80.96	75.65
46.74	237.94	6.75	81.32	74.57
50.14	239.11	13.88	90.20	76.33
46.75	232.66	7.55	82.13	74.58
48.06	234.77	14.06	87.58	73.51
45.53	232.05	10.67	83.56	72.89
49.06	235.60	14.73	88.79	74.07
46.94	231.70	7.56	83.81	76.25
44.91	229.93	10.13	84.26	74.13
46.16	237.02	6.62	84.65	78.03
46.06	238.05	6.44	84.78	78.34
45.98	235.43	7.21	84.82	77.61
45.87	235.57	7.25	84.93	77.68
42.09	231.64	10.41	86.44	76.03

Fitted Gaussian variomodel:

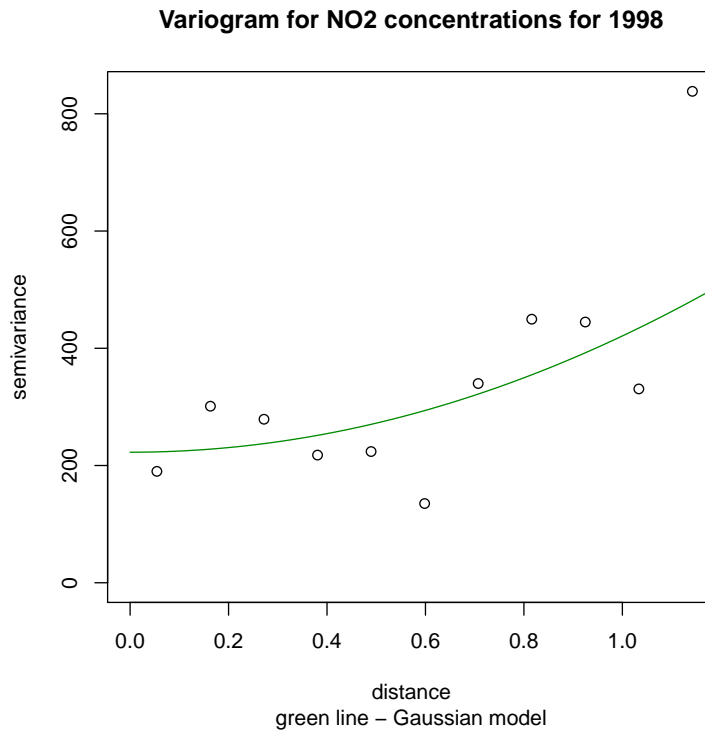


Figure 2: Empirical and modeled variograms

Table 2 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 2: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
241.18	242.90	17.17	82.19	65.02	9.56	73.90	64.34
241.58	243.32	17.28	82.33	65.06	9.64	74.08	64.44
251.90	275.31	5.20	89.06	83.86	1.18	73.29	72.11
251.41	274.23	5.68	88.89	83.20	1.38	73.20	71.82
238.04	244.66	15.80	81.43	65.63	7.69	71.90	64.20
251.83	273.98	4.89	90.24	85.35	1.29	73.23	71.94
236.24	241.07	17.58	80.97	63.40	9.42	72.40	62.98
236.30	242.27	16.89	80.98	64.09	8.55	71.88	63.33
238.23	238.44	19.68	81.92	62.24	10.89	74.32	63.43
237.16	236.96	19.75	81.60	61.85	10.97	74.08	63.11
236.92	242.75	16.81	80.98	64.16	7.54	71.51	63.97
239.63	243.13	16.13	81.69	65.56	6.39	71.67	65.28
237.21	237.94	17.34	80.12	62.77	7.47	71.58	64.11
240.32	239.11	20.88	84.82	63.94	11.75	77.42	65.68
236.56	232.66	17.64	78.74	61.10	8.01	71.80	63.79
235.17	234.77	20.33	81.17	60.84	11.83	75.48	63.65
233.24	232.05	17.82	77.23	59.41	9.91	72.49	62.58
236.55	235.60	20.44	81.65	61.21	11.79	76.58	64.79
238.57	231.70	16.00	77.63	61.64	7.86	72.54	64.68
235.06	229.93	16.61	76.53	59.92	9.40	72.86	63.46
241.25	237.02	13.34	78.51	65.17	7.17	72.96	65.79
241.74	238.05	12.87	78.76	65.89	7.04	73.03	65.98
240.50	235.43	13.38	77.92	64.54	7.52	73.05	65.53
240.58	235.57	13.13	77.91	64.77	7.53	73.11	65.58
238.00	231.64	12.17	76.88	64.71	9.06	74.26	65.19

2000

Fitting linear and exponential variomodels:

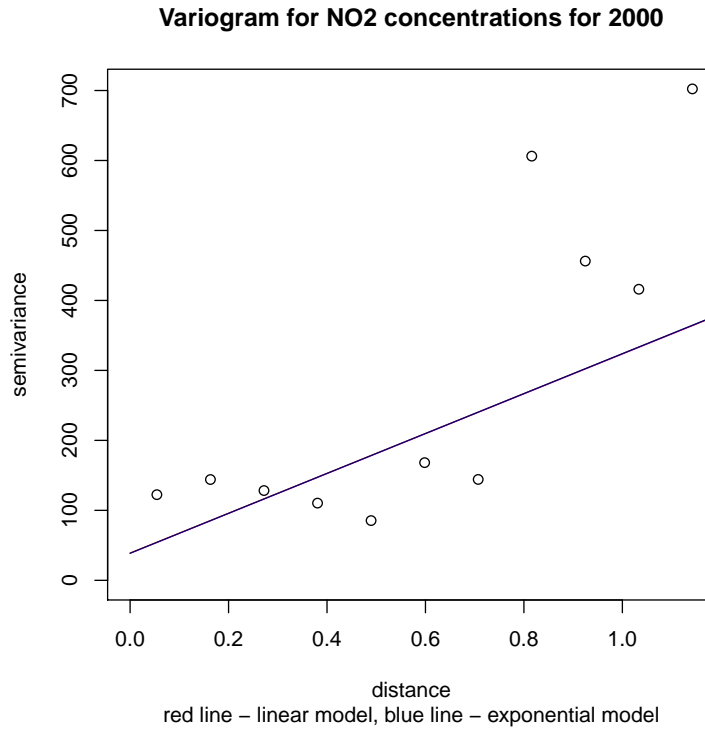


Figure 3: Empirical and modeled variograms

In Table 3, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 3: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
38.96	68.44	10.41	62.72	52.31
38.98	69.72	10.48	62.99	52.51
27.13	76.34	3.61	57.75	54.13
27.27	74.35	3.84	57.86	54.03
34.44	67.70	9.39	59.80	50.41
37.44	73.18	11.33	61.51	50.18
36.03	71.97	10.60	60.59	49.98
37.79	63.94	13.34	64.63	51.29
37.11	60.46	13.68	64.54	50.86
35.02	72.97	10.33	60.48	50.15
33.98	69.98	9.84	61.15	51.31
38.19	67.29	11.25	61.90	50.65
45.13	63.00	17.74	69.43	51.69
44.33	59.37	12.52	63.28	50.75
44.10	69.04	15.61	65.30	49.68
44.43	61.67	19.11	69.35	50.24
53.76	57.26	13.52	65.57	52.05
45.77	60.92	15.76	66.37	50.61
51.79	67.76	13.07	66.39	53.32
51.45	69.61	12.95	66.50	53.55
52.32	65.37	13.66	66.71	53.06
52.23	65.60	13.74	66.85	53.11
43.54	61.55	17.00	69.03	52.03
45.33	64.12	16.58	69.19	52.61

Fitted Gaussian variomodel:

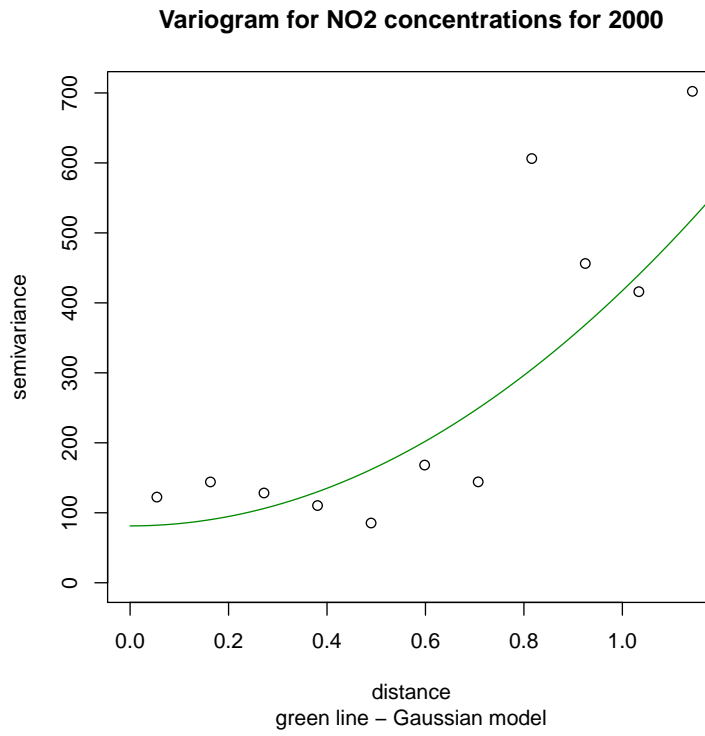


Figure 4: Empirical and modeled variograms

Table 4 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 4: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
89.55	68.44	12.79	64.03	51.23	3.16	65.41	62.24
89.81	69.72	13.05	64.30	51.26	3.22	65.59	62.37
92.39	76.34	-1.77	56.55	58.32	-4.27	63.90	68.17
92.21	74.35	-1.33	56.67	58.00	-4.03	63.86	67.89
86.88	67.70	9.05	60.10	51.06	1.93	63.37	61.44
86.47	73.18	12.48	62.29	49.81	3.64	64.13	60.50
86.19	71.97	10.93	61.02	50.09	2.93	63.55	60.62
88.02	63.94	16.34	65.22	48.88	4.99	66.25	61.26
87.39	60.46	16.50	65.07	48.57	5.20	66.08	60.88
86.32	72.97	10.57	60.03	49.46	2.43	63.28	60.86
87.80	69.98	10.42	60.21	49.79	1.79	63.54	61.75
86.81	67.29	12.67	61.20	48.54	2.98	63.68	60.70
88.74	63.00	19.14	69.16	50.01	6.51	69.89	63.38
86.84	59.37	14.90	62.81	47.91	3.87	64.19	60.32
85.37	69.04	18.31	65.40	47.09	5.92	65.28	59.36
86.64	61.67	20.49	68.38	47.89	7.19	69.36	62.17
88.47	57.26	16.72	66.42	49.70	4.35	65.29	60.95
86.58	60.92	18.88	67.05	48.17	5.80	65.81	60.01
90.13	67.76	15.98	68.50	52.52	3.92	65.76	61.84
90.43	69.61	15.76	68.83	53.07	3.83	65.83	62.00
89.78	65.37	16.72	68.91	52.18	4.30	65.92	61.62
89.85	65.60	16.78	69.18	52.40	4.34	66.00	61.66
88.52	61.55	19.51	71.80	52.29	6.04	67.51	61.48
89.28	64.12	18.99	72.64	53.65	5.75	67.56	61.81

2001

Fitting linear and exponential variomodels:

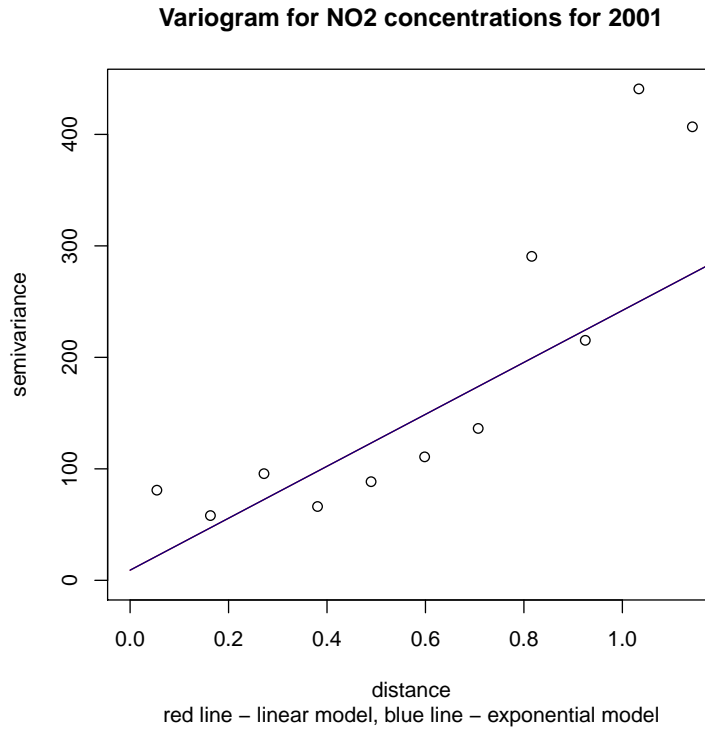


Figure 5: Empirical and modeled variograms

In Table 5, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 5: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
40.59	27.48	22.16	58.36	36.21
39.96	28.84	22.03	58.49	36.46
35.75	28.48	20.21	57.23	37.02
35.98	26.81	20.40	57.35	36.95
43.98	23.86	23.06	57.58	34.52
42.87	32.23	24.15	58.56	34.42
44.25	29.85	24.02	58.31	34.28
36.72	22.69	24.97	60.50	35.53
35.49	18.94	25.53	60.68	35.16
45.41	31.59	24.69	59.28	34.59
46.35	29.87	25.23	60.83	35.61
47.43	27.95	26.22	61.50	35.29
49.28	23.12	28.67	65.40	36.73
49.60	21.50	27.52	63.15	35.63
48.08	30.64	29.57	64.79	35.21
49.98	23.54	30.72	66.60	35.88
53.44	19.38	29.04	65.99	36.95
46.87	22.93	30.25	66.33	36.09
50.41	28.33	29.16	67.13	37.97
50.13	29.84	29.14	67.29	38.15
49.75	26.20	29.56	67.41	37.85
49.34	26.26	29.64	67.56	37.92
48.81	22.79	31.84	69.41	37.56
48.00	24.65	31.74	69.74	38.00

Fitting Gaussian variomodel:

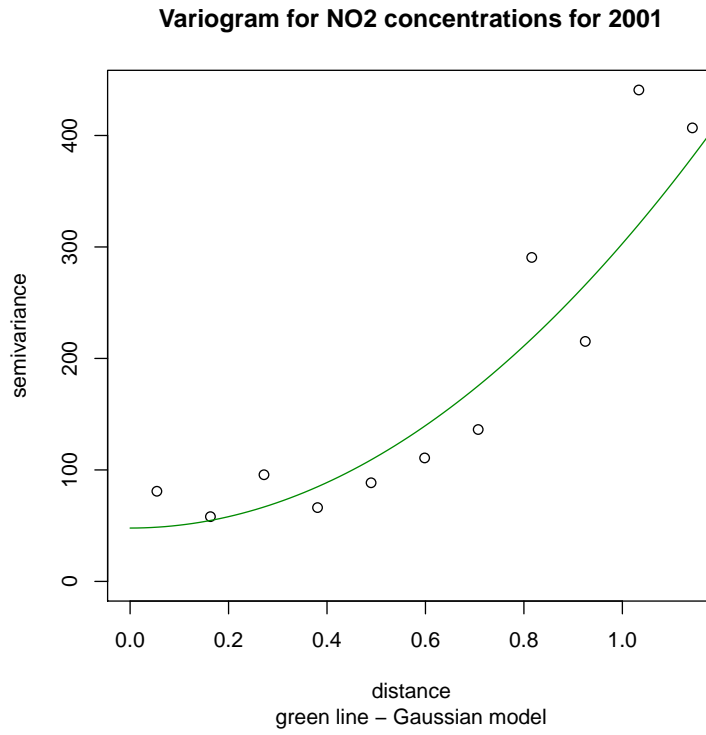


Figure 6: Empirical and modeled variograms

Table 6 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 6: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
52.87	27.48	22.75	58.57	35.82	13.56	63.40	49.84
53.04	28.84	22.87	58.71	35.84	13.60	63.54	49.94
54.50	28.48	18.24	58.92	40.67	8.31	62.87	54.55
54.40	26.81	18.55	59.00	40.45	8.51	62.84	54.33
51.21	23.86	22.24	57.96	35.72	12.87	62.06	49.19
50.98	32.23	24.10	58.94	34.85	14.16	62.60	48.44
50.80	29.85	23.66	58.71	35.04	13.69	62.22	48.53
51.94	22.69	26.56	60.74	34.18	15.16	64.21	49.05
51.56	18.94	26.92	60.88	33.96	15.36	64.12	48.75
50.87	31.59	24.86	59.44	34.59	13.49	62.21	48.72
51.77	29.87	26.05	60.80	34.75	13.19	62.61	49.42
51.18	27.95	27.30	61.20	33.90	14.12	62.71	48.59
52.35	23.12	30.57	65.53	34.97	16.62	67.37	50.75
51.20	21.50	28.97	62.43	33.46	14.91	63.20	48.29
50.31	30.64	30.79	63.73	32.94	16.53	64.05	47.52
51.07	23.54	32.05	65.54	33.49	17.36	67.14	49.78
52.21	19.38	30.47	64.70	34.23	15.47	64.25	48.78
51.05	22.93	31.43	64.79	33.35	16.56	64.60	48.04
53.23	28.33	30.15	66.01	35.86	15.22	64.72	49.49
53.41	29.84	30.03	66.26	36.24	15.16	64.78	49.62
53.01	26.20	30.51	66.13	35.62	15.52	64.85	49.33
53.06	26.26	30.53	66.29	35.76	15.56	64.91	49.36
52.24	22.79	31.76	67.29	35.52	16.92	66.13	49.22
52.71	24.65	31.45	67.90	36.45	16.73	66.21	49.48

2002

Fitting linear and exponential variomodels:

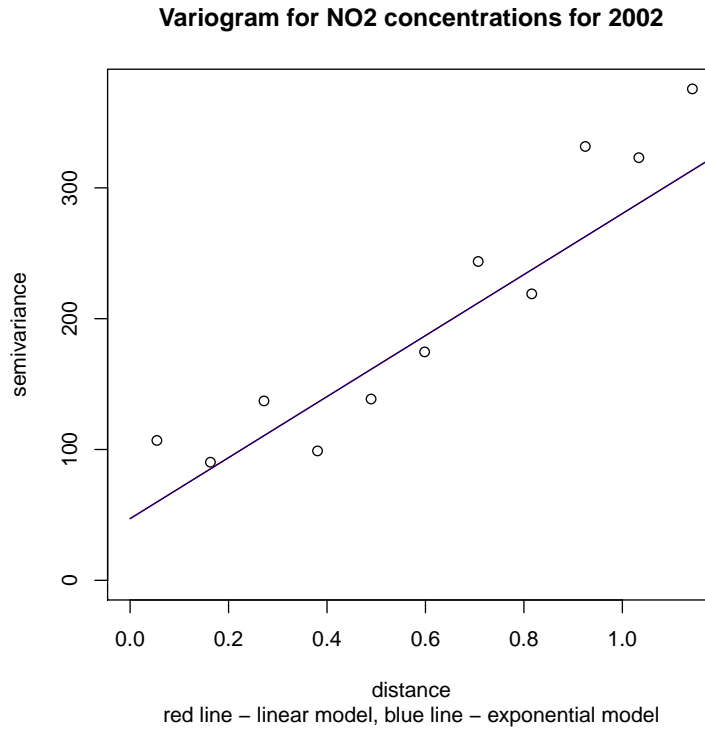


Figure 7: Empirical and modeled variograms

In Table 7, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 7: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
41.27	74.70	15.84	68.70	52.86
41.28	75.64	15.74	68.98	53.24
34.07	83.24	9.15	63.26	54.12
34.22	81.59	9.39	63.40	54.01
39.93	74.49	14.92	65.64	50.72
40.67	77.85	16.94	67.54	50.59
40.56	77.00	16.22	66.53	50.31
40.74	71.09	18.73	70.94	52.21
39.73	68.53	19.32	70.87	51.54
40.77	74.97	16.10	66.44	50.34
42.63	68.96	17.29	68.04	50.75
50.58	70.04	22.45	76.44	53.99
45.41	65.72	18.72	69.60	50.88
47.23	75.23	21.85	71.92	50.06
49.17	68.28	25.04	76.48	51.44
51.55	65.31	20.04	72.22	52.18
49.60	71.55	22.18	73.15	50.97
50.84	73.81	19.80	73.17	53.38
50.70	75.31	19.71	73.30	53.59
51.37	72.27	20.38	73.55	53.17
51.38	72.56	20.47	73.71	53.23
49.62	70.84	22.67	74.28	51.62
48.39	71.01	23.74	76.26	52.52
49.01	73.16	23.40	76.45	53.05

Fitting Gaussian variomodel:

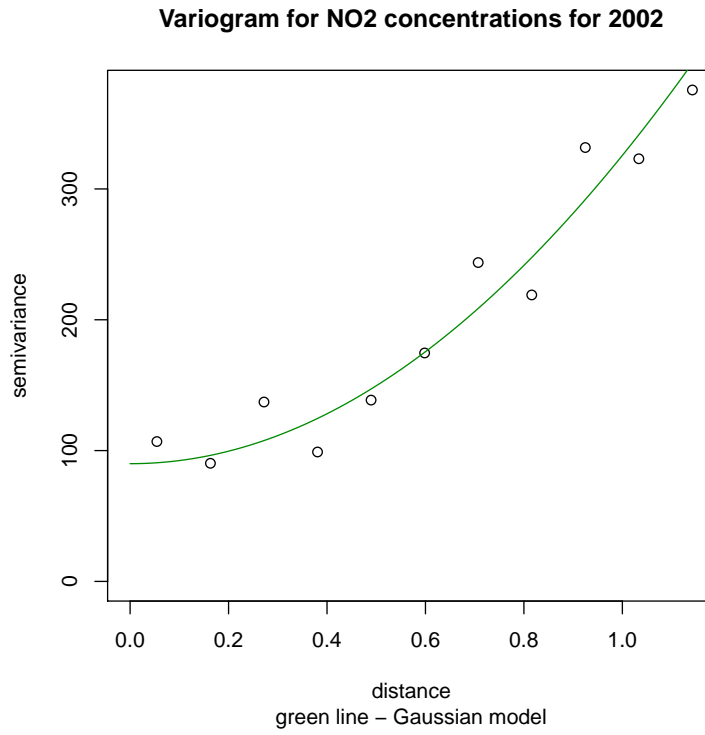


Figure 8: Empirical and modeled variograms

Table 8 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 8: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
98.94	74.70	16.10	64.66	48.56	10.81	65.66	54.86
99.22	75.64	16.41	65.00	48.59	10.86	65.84	54.98
101.57	83.24	10.63	64.36	53.73	4.33	64.24	59.91
101.39	81.59	10.95	64.38	53.43	4.52	64.19	59.66
96.13	74.49	13.83	61.83	48.00	9.63	63.73	54.10
95.77	77.85	16.25	63.31	47.07	11.12	64.43	53.31
95.45	77.00	15.38	62.54	47.16	10.48	63.88	53.40
97.33	71.09	20.72	67.02	46.30	12.39	66.41	54.02
96.69	68.53	20.87	66.87	45.99	12.55	66.23	53.69
95.50	74.97	16.68	62.92	46.24	9.98	63.57	53.59
95.97	68.96	19.61	64.87	45.27	10.38	63.84	53.47
98.00	70.04	27.31	74.83	47.53	13.80	69.72	55.93
96.03	65.72	21.49	66.33	44.84	11.10	64.25	53.16
94.62	75.23	23.38	67.89	44.52	12.89	65.25	52.36
95.93	68.28	27.78	73.37	45.59	14.29	69.16	54.87
97.74	65.31	23.32	69.62	46.30	11.43	65.16	53.74
95.87	71.55	24.13	69.45	45.31	12.73	65.67	52.94
99.46	73.81	22.93	71.73	48.80	11.02	65.54	54.53
99.76	75.31	22.81	72.10	49.29	10.93	65.60	54.67
99.12	72.27	23.19	71.77	48.58	11.34	65.69	54.35
99.20	72.56	23.18	71.97	48.79	11.37	65.76	54.39
96.76	70.84	24.22	70.90	46.68	12.74	66.16	53.42
98.00	71.01	23.96	73.46	49.50	12.88	67.15	54.27
98.78	73.16	23.54	74.17	50.63	12.61	67.17	54.57

2004

Fitting linear and exponential variomodels:

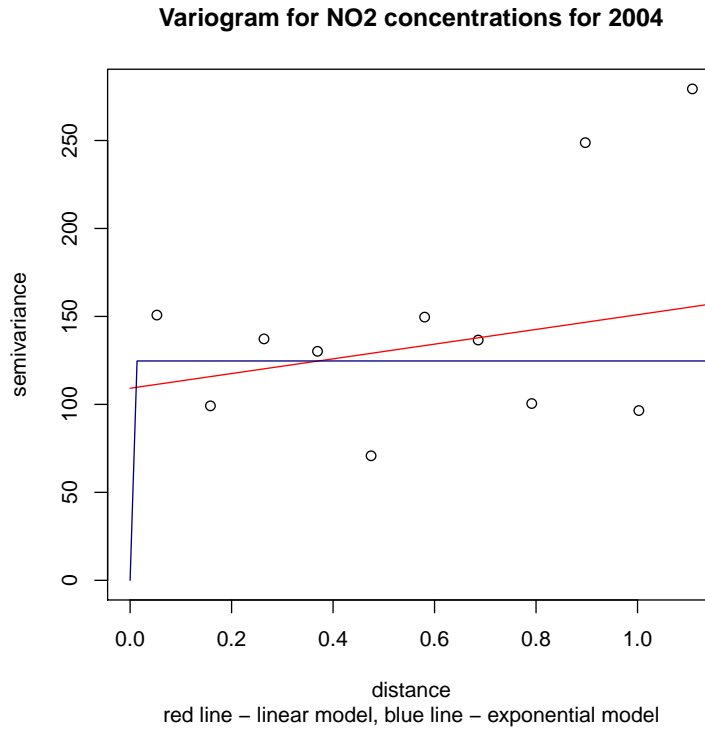


Figure 9: Empirical and modeled variograms

In Table 9, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 9: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
39.88	129.87	5.87	61.16	55.29
39.88	129.87	5.63	61.24	55.60
39.88	129.87	10.09	64.25	54.17
39.88	129.87	9.70	61.41	51.71
39.88	129.87	2.19	63.79	61.59
39.88	129.87	9.23	61.36	52.13
39.88	129.87	10.20	61.60	51.40
39.88	129.87	7.57	62.03	54.46
39.88	129.87	8.32	62.01	53.69
39.88	129.87	12.01	62.92	50.91
39.88	129.87	13.68	64.78	51.10
39.88	129.87	8.72	64.54	55.81
39.88	129.87	14.55	65.76	51.21
39.88	129.87	10.41	64.36	53.95
39.88	129.87	14.98	65.69	50.71
39.88	129.87	11.61	65.18	53.58
39.88	129.87	15.62	68.04	52.42
39.88	129.87	15.72	67.17	51.44
39.88	129.87	15.79	69.35	53.55
39.88	129.87	15.80	69.55	53.75
39.88	129.87	15.98	69.41	53.43
39.88	129.87	16.09	68.14	52.04
39.88	129.87	13.76	67.58	53.82
39.88	129.87	16.49	69.43	52.94
39.88	129.87	11.74	72.38	60.64

Fitting Gaussian variomodel:

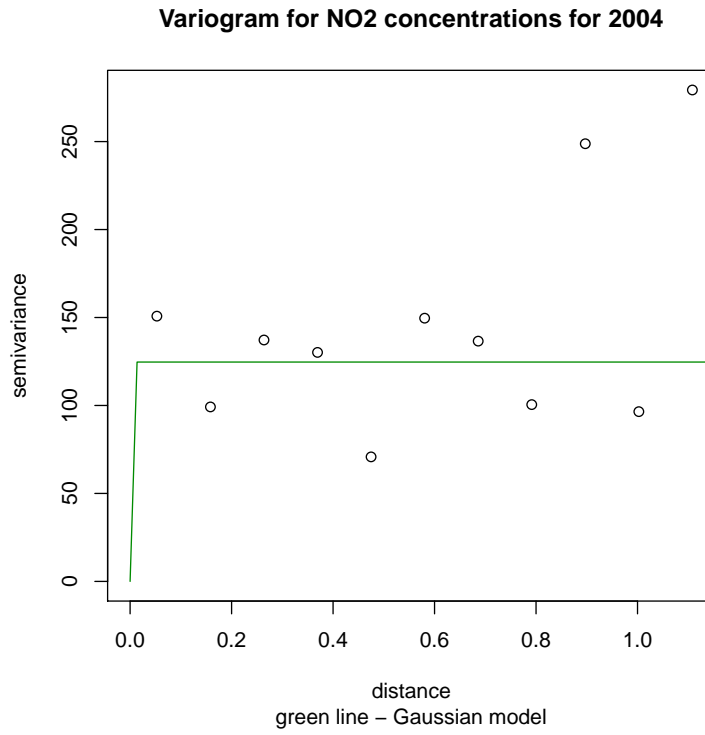


Figure 10: Empirical and modeled variograms

Table 10 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 10: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
129.87	129.87	8.37	63.23	54.87	-0.79	66.17	66.96
129.87	129.87	8.18	63.29	55.11	-0.89	66.24	67.14
129.87	129.87	7.63	63.33	55.70	-2.55	69.23	71.78
129.87	129.87	10.62	63.21	52.59	0.38	65.92	65.54
129.87	129.87	-0.06	68.91	68.97	-3.33	68.66	71.99
129.87	129.87	11.06	63.19	52.13	0.77	65.58	64.81
129.87	129.87	11.49	63.37	51.88	0.93	65.67	64.74
129.87	129.87	10.05	62.92	52.87	0.09	66.18	66.09
129.87	129.87	10.77	62.90	52.13	0.48	65.99	65.52
129.87	129.87	13.20	63.72	50.52	1.32	66.09	64.77
129.87	129.87	15.47	64.85	49.38	1.97	66.51	64.54
129.87	129.87	8.15	64.11	55.97	-0.72	68.21	68.93
129.87	129.87	16.79	65.78	48.99	2.46	66.65	64.19
129.87	129.87	10.92	63.50	52.58	0.38	67.51	67.12
129.87	129.87	17.06	66.06	49.00	2.95	66.44	63.49
129.87	129.87	11.65	63.75	52.11	0.50	67.97	67.46
129.87	129.87	18.07	67.95	49.89	2.64	67.51	64.87
129.87	129.87	17.96	67.24	49.27	2.95	67.08	64.13
129.87	129.87	17.82	69.51	51.70	2.41	68.17	65.76
129.87	129.87	17.71	69.79	52.09	2.36	68.28	65.92
129.87	129.87	18.05	69.59	51.54	2.52	68.16	65.64
129.87	129.87	18.17	68.15	49.98	2.86	67.57	64.71
129.87	129.87	12.10	65.14	53.04	0.09	69.65	69.56
129.87	129.87	17.75	69.57	51.81	2.57	68.38	65.81
129.87	129.87	-7.08	76.06	83.14	-7.28	77.65	84.93

2005

Fitting linear and exponential variomodels:

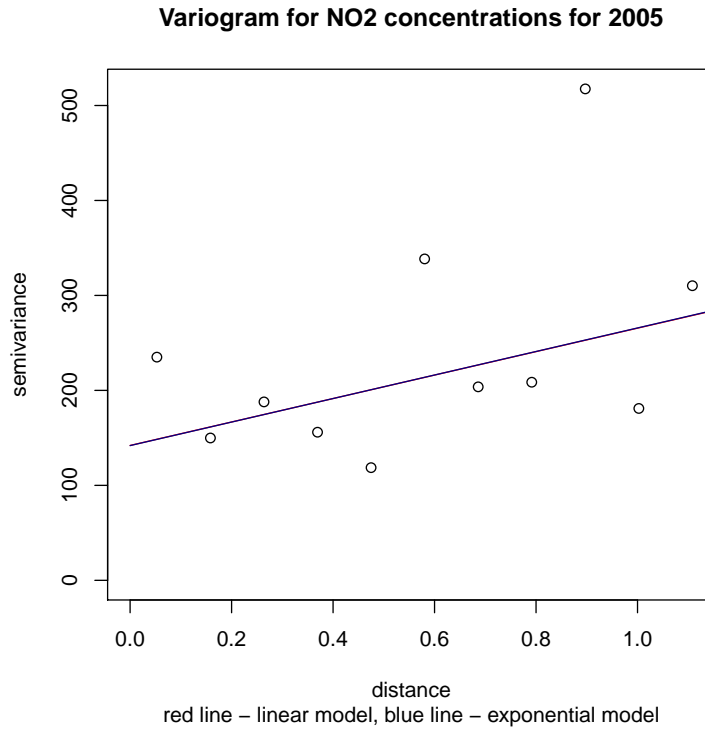


Figure 11: Empirical and modeled variograms

In Table 11, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 11: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
41.03	173.19	-4.05	76.71	80.75
40.85	173.82	-4.40	76.84	81.23
38.70	180.09	1.19	81.25	80.06
42.99	170.75	1.25	76.73	75.47
39.32	190.49	-10.09	80.83	90.92
39.47	181.27	1.64	82.11	80.47
42.70	170.34	0.48	76.51	76.03
43.46	169.90	1.82	76.81	74.99
40.19	170.76	-2.21	77.52	79.74
40.45	169.06	-1.17	77.38	78.55
44.61	167.91	4.03	78.43	74.40
46.59	164.05	5.87	80.61	74.74
40.49	176.25	-1.78	80.64	82.42
48.50	161.75	6.81	81.66	74.85
40.92	172.00	0.66	80.08	79.42
47.98	164.18	7.22	81.22	73.99
41.33	173.63	2.04	80.94	78.90
50.80	161.08	7.78	84.38	76.60
48.89	162.78	7.96	83.03	75.06
50.01	164.33	7.78	86.09	78.31
49.90	165.11	7.75	86.36	78.61
41.63	178.16	3.29	82.47	79.17
49.87	163.24	7.99	86.07	78.08
48.49	162.88	8.26	84.20	75.94
41.91	180.73	4.35	83.67	79.33
47.52	164.27	8.50	85.74	77.24
39.32	207.30	-0.33	90.56	90.89

Fitting Gaussian variomodel:

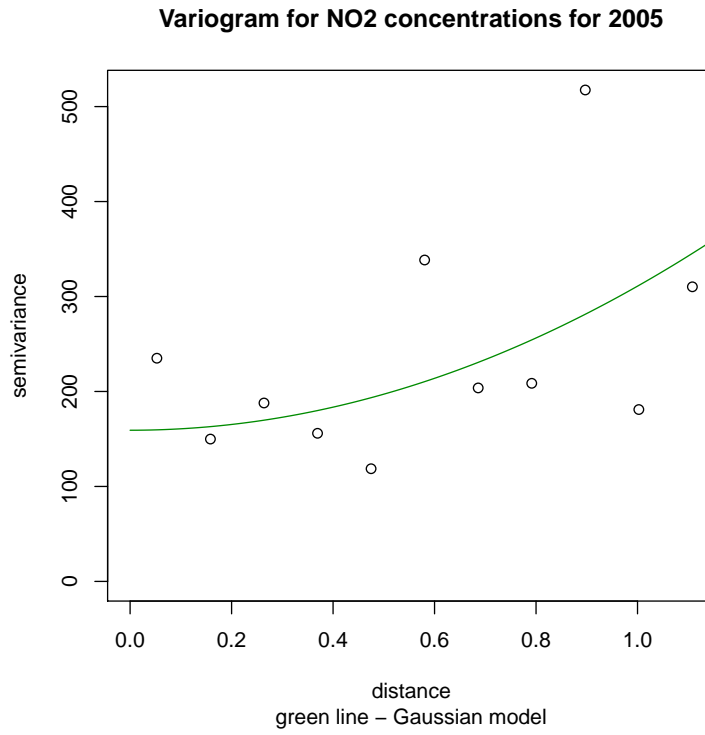


Figure 12: Empirical and modeled variograms

Table 12 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 12: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
174.93	173.19	0.83	80.60	79.78	-11.39	82.44	93.82
175.44	173.82	0.54	80.69	80.15	-11.54	82.54	94.08
176.77	180.09	-4.69	81.19	85.88	-14.07	86.89	100.96
169.52	170.75	3.30	80.09	76.79	-9.79	82.11	91.90
186.65	190.49	-12.78	88.12	100.90	-15.27	85.72	100.99
177.00	181.27	-5.11	82.43	87.54	-13.86	87.06	100.91
169.65	170.34	4.05	79.90	75.84	-9.30	81.53	90.83
168.65	169.90	4.45	80.06	75.61	-9.09	81.68	90.77
173.73	170.76	2.06	79.25	77.19	-10.40	82.24	92.65
172.38	169.06	3.01	79.10	76.09	-9.87	81.96	91.83
168.03	167.91	6.03	80.17	74.14	-8.64	82.21	90.84
168.03	164.05	8.56	81.13	72.56	-7.89	82.63	90.52
178.17	176.25	-3.62	79.42	83.05	-12.10	84.69	96.78
167.88	161.75	10.25	81.91	71.66	-7.32	82.68	90.00
174.50	172.00	0.54	78.41	77.87	-10.53	83.69	94.21
167.03	164.18	10.36	81.74	71.38	-6.78	82.21	88.99
174.34	173.63	0.85	78.26	77.41	-10.51	84.21	94.72
169.63	161.08	11.56	84.24	72.68	-7.25	83.68	90.93
168.21	162.78	11.29	83.01	71.71	-6.88	83.00	89.88
171.45	164.33	10.95	86.28	75.33	-7.64	84.54	92.18
171.77	165.11	10.75	86.67	75.92	-7.72	84.68	92.40
175.43	178.16	0.21	78.43	78.22	-11.11	85.41	96.52
171.24	163.24	11.25	86.24	74.99	-7.52	84.48	92.00
169.26	162.88	11.29	84.03	72.74	-7.10	83.60	90.70
176.09	180.73	-0.01	79.14	79.15	-11.46	86.27	97.73
171.05	164.27	10.13	85.65	75.52	-7.65	84.59	92.24
195.24	207.30	-31.35	93.39	124.73	-22.78	96.99	119.77

2006

Fitting linear and exponential variomodels:

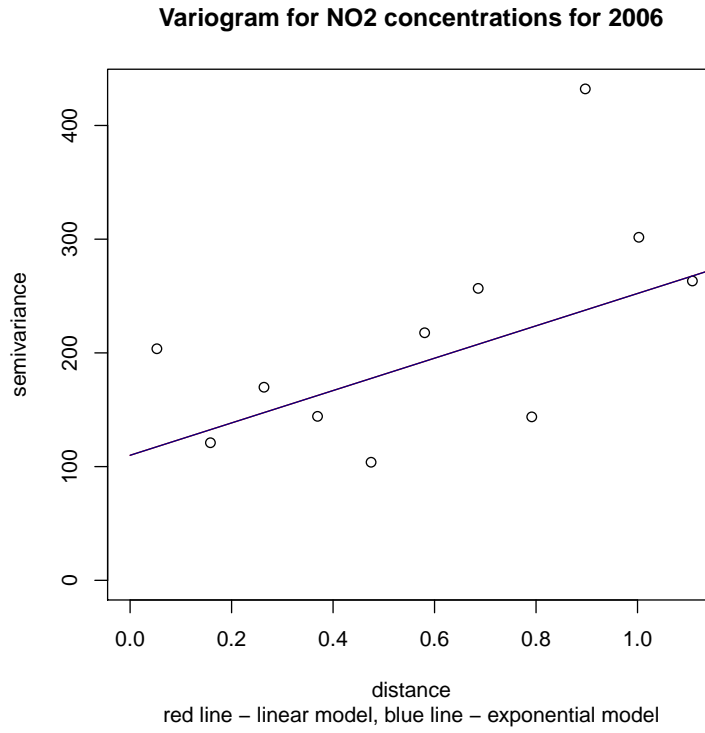


Figure 13: Empirical and modeled variograms

In Table 13, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 13: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
41.65	138.60	8.14	71.00	62.86
41.64	139.04	8.05	71.17	63.12
35.38	146.33	6.50	70.47	63.96
41.37	138.12	9.45	69.57	60.12
36.28	148.06	6.72	70.94	64.22
41.94	137.87	10.15	70.26	60.11
42.06	137.69	10.25	69.91	59.67
40.98	135.32	10.14	72.11	61.97
40.74	134.21	10.60	71.88	61.28
42.71	135.72	10.84	70.38	59.54
44.67	131.44	11.89	71.72	59.84
43.99	138.75	12.11	75.55	63.44
46.49	128.97	12.80	72.68	59.89
43.29	135.82	13.11	74.82	61.71
47.52	132.43	14.37	73.47	59.10
43.70	136.98	14.04	75.55	61.50
49.86	127.84	13.53	74.67	61.14
48.73	130.39	14.64	74.59	59.95
49.04	131.34	13.28	75.66	62.38
48.92	132.23	13.21	75.81	62.59
44.16	138.95	14.97	76.85	61.88
49.20	129.97	13.66	75.87	62.21
48.61	130.06	14.86	75.47	60.61
47.88	130.66	15.26	76.83	61.57
41.27	169.25	14.62	84.59	69.98

Fitting Gaussian variomodel:

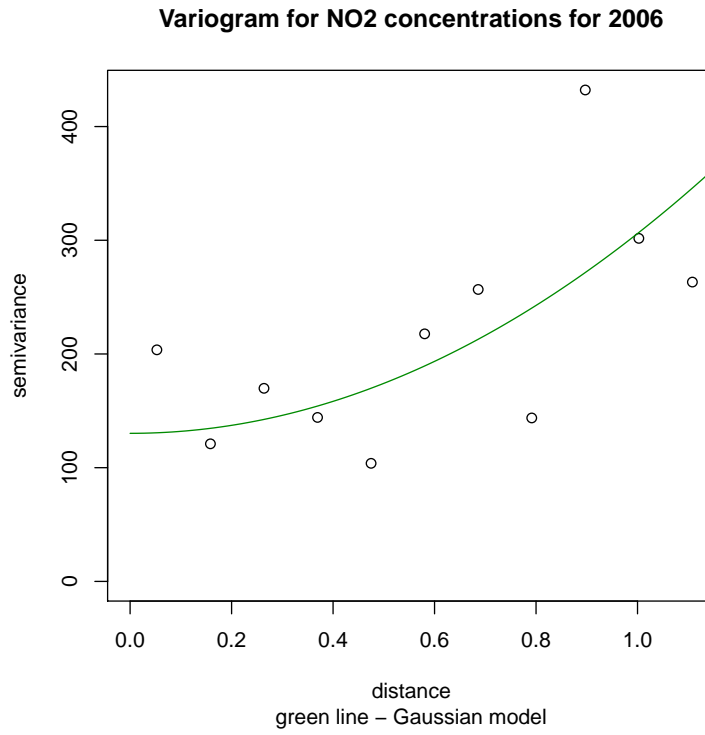


Figure 14: Empirical and modeled variograms

Table 14 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 14: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
142.12	138.60	11.45	73.07	61.63	-0.53	74.08	74.61
142.49	139.04	11.62	73.30	61.68	-0.59	74.20	74.79
145.30	146.33	1.35	69.09	67.75	-4.92	75.84	80.76
138.43	138.12	9.23	69.90	60.67	-0.28	73.08	73.36
145.54	148.06	1.29	70.93	69.64	-4.77	75.96	80.73
138.10	137.87	11.55	71.28	59.73	0.66	73.05	72.39
137.57	137.69	10.71	70.44	59.73	0.47	72.90	72.42
140.76	135.32	13.85	72.96	59.11	0.63	74.23	73.59
139.74	134.21	13.89	72.61	58.72	0.95	73.94	72.99
137.39	135.72	11.35	69.91	58.56	0.46	73.04	72.58
137.54	131.44	13.87	71.18	57.31	1.02	73.36	72.34
143.66	138.75	12.66	74.45	61.79	0.20	76.84	76.65
137.42	128.97	15.82	72.41	56.59	1.63	73.54	71.91
140.94	135.82	13.97	73.24	59.26	1.14	75.86	74.72
136.28	132.43	17.00	73.24	56.24	2.67	73.67	71.00
140.85	136.98	13.78	72.99	59.21	1.20	76.33	75.13
139.03	127.84	17.72	75.13	57.40	1.79	74.44	72.65
137.46	130.39	17.94	74.43	56.49	2.53	74.26	71.73
140.73	131.34	17.56	77.08	59.52	1.40	75.06	73.66
141.03	132.23	17.45	77.43	59.98	1.32	75.16	73.84
141.79	138.95	12.97	72.87	59.91	0.84	77.39	76.55
140.50	129.97	17.94	77.17	59.23	1.61	75.11	73.50
138.42	130.06	18.20	75.47	57.27	2.40	74.78	72.38
139.93	130.66	17.76	77.08	59.32	2.13	75.71	73.58
157.31	169.25	-11.00	83.33	94.33	-7.04	87.55	94.59

2007

Fitting linear and exponential variomodels:

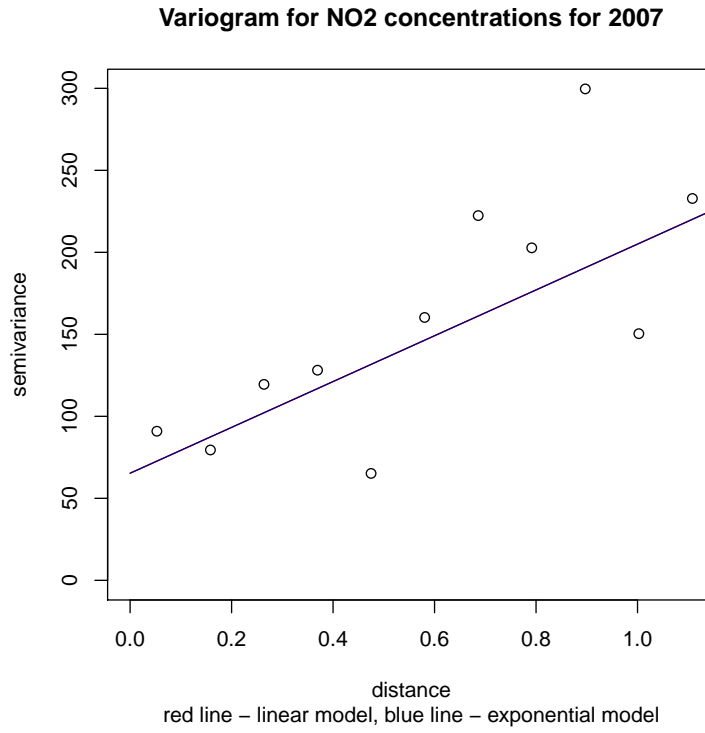


Figure 15: Empirical and modeled variograms

In Table 15, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 15: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
39.55	88.10	14.88	61.86	46.97
39.43	88.56	14.63	61.78	47.15
35.63	93.08	18.44	69.66	51.22
41.14	88.05	18.71	64.33	45.62
37.03	95.29	18.87	70.45	51.58
40.86	88.65	18.27	63.29	45.02
41.75	88.38	19.25	64.24	44.99
38.29	85.24	16.18	62.38	46.20
38.22	84.25	16.94	62.64	45.71
43.93	86.65	20.61	66.36	45.75
47.06	82.64	22.11	68.41	46.31
42.23	87.01	16.62	63.74	47.12
48.66	80.41	22.99	69.19	46.19
41.12	84.80	18.41	64.27	45.86
46.84	84.72	23.66	68.19	44.53
42.24	84.68	19.52	65.21	45.69
50.74	78.94	24.13	71.36	47.22
47.66	82.28	24.37	69.80	45.44
49.56	82.14	24.36	72.81	48.45
49.44	83.00	24.37	73.04	48.67
48.90	80.60	24.57	72.69	48.12
46.94	81.52	24.77	70.73	45.96
45.49	81.14	25.29	71.80	46.51
39.94	111.12	18.54	70.45	51.91

Fitting Gaussian variomodel:

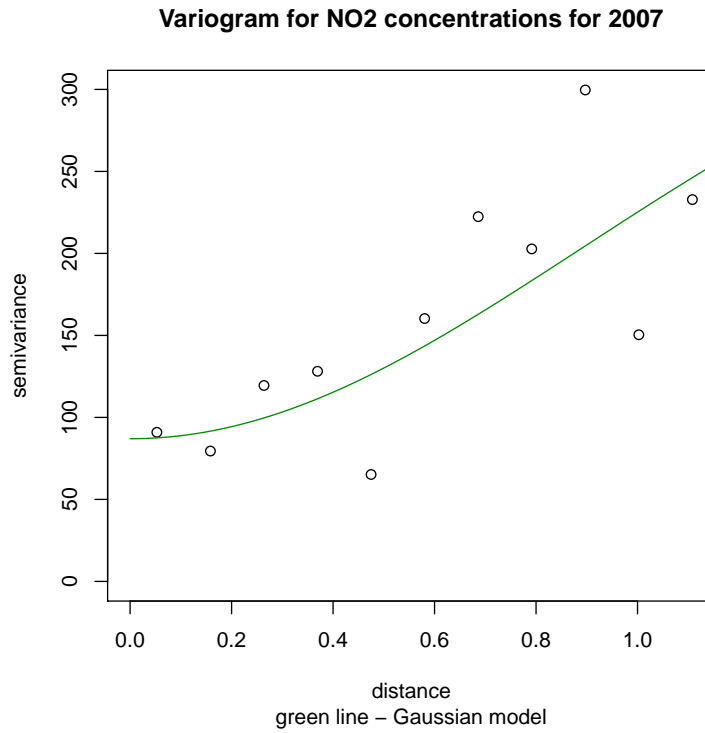


Figure 16: Empirical and modeled variograms

Table 16 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 16: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
97.34	88.10	16.41	64.03	47.63	7.89	64.90	57.00
97.55	88.56	16.30	63.97	47.67	7.83	64.97	57.13
100.35	93.08	17.06	78.79	61.73	5.40	67.20	61.80
94.89	88.05	19.85	67.18	47.32	8.48	64.58	56.10
100.60	95.29	17.34	80.98	63.64	5.58	67.36	61.78
94.38	88.65	19.39	65.60	46.22	9.07	64.39	55.32
94.07	88.38	20.48	66.89	46.41	9.05	64.42	55.37
95.41	85.24	18.63	64.20	45.57	8.86	65.04	56.19
94.70	84.25	19.14	64.40	45.27	9.15	64.88	55.73
93.61	86.65	22.76	69.33	46.57	9.28	64.78	55.50
93.20	82.64	25.17	71.00	45.84	9.88	65.20	55.32
96.54	87.01	18.08	65.22	47.14	8.60	67.00	58.41
92.75	80.41	26.45	70.90	44.45	10.40	65.38	54.98
94.51	84.80	19.47	64.73	45.26	9.42	66.38	56.96
91.57	84.72	25.17	68.38	43.21	11.09	65.32	54.24
94.48	84.68	19.53	64.64	45.11	9.57	66.82	57.25
93.58	78.94	27.66	71.96	44.31	10.66	66.20	55.54
92.29	82.28	25.88	69.26	43.38	11.10	65.89	54.80
94.90	82.14	27.58	73.67	46.09	10.45	66.77	56.31
95.14	83.00	27.51	74.01	46.50	10.41	66.86	56.45
94.68	80.60	27.49	72.97	45.47	10.60	66.78	56.18
92.98	81.52	25.64	69.61	43.97	11.06	66.34	55.29
94.29	81.14	24.10	69.72	45.62	10.91	67.09	56.18
117.12	111.12	1.40	72.73	71.33	3.60	75.44	71.84

2008

Fitting linear and exponential variomodels:

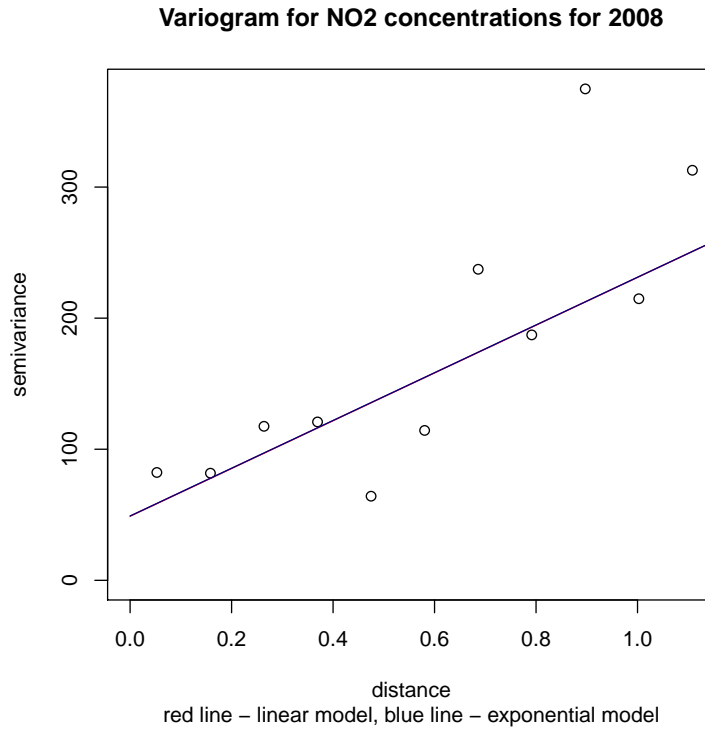


Figure 17: Empirical and modeled variograms

In Table 17, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 17: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
37.92	73.11	14.58	60.57	45.99
37.95	73.79	14.51	60.57	46.06
29.76	75.94	13.92	65.48	51.56
37.79	72.96	15.92	61.86	45.94
31.49	79.62	14.30	66.22	51.93
38.24	75.10	16.56	61.52	44.96
38.82	74.47	16.72	62.03	45.32
36.74	69.62	16.55	61.60	45.05
36.30	68.30	17.16	61.75	44.60
41.53	72.60	17.60	63.68	46.09
45.82	67.74	19.02	65.68	46.66
43.90	71.83	18.09	64.71	46.62
47.87	65.04	20.11	66.66	46.56
41.31	69.75	19.54	64.41	44.87
46.00	71.77	21.60	66.49	44.89
42.36	69.75	20.73	65.35	44.61
50.92	62.74	21.35	68.96	47.61
47.18	68.18	22.20	68.02	45.82
48.87	66.60	21.43	70.28	48.85
48.69	67.72	21.41	70.49	49.07
48.12	64.28	21.80	70.32	48.52
46.28	66.71	22.65	69.00	46.35
44.24	65.01	23.37	70.28	46.91
42.33	95.79	21.56	73.68	52.12

Fitting Gaussian variomodel:

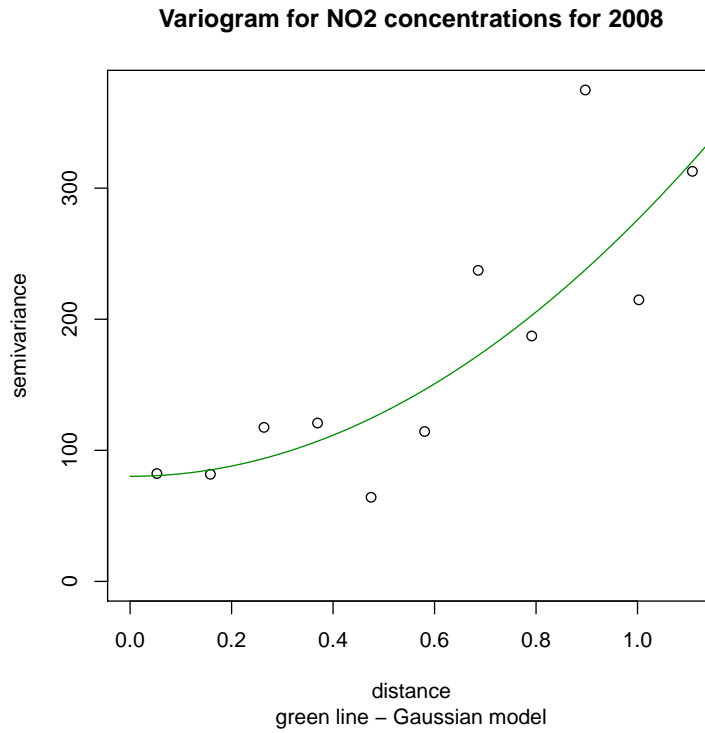


Figure 18: Empirical and modeled variograms

Table 18 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 18: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
88.14	73.11	15.97	62.02	46.05	6.72	64.20	57.48
88.41	73.79	15.97	62.06	46.09	6.69	64.31	57.61
90.04	75.94	12.13	73.90	61.77	3.05	65.37	62.32
85.43	72.96	16.32	63.67	47.34	6.84	63.40	56.57
90.24	79.62	12.43	76.11	63.68	3.22	65.52	62.30
85.23	75.10	17.87	62.85	44.97	7.72	63.50	55.79
84.84	74.47	17.57	63.64	46.07	7.53	63.36	55.83
87.03	69.62	18.71	62.77	44.06	7.94	64.60	56.66
86.29	68.30	19.03	62.79	43.76	8.20	64.39	56.20
84.69	72.60	18.91	65.72	46.81	7.59	63.56	55.97
84.85	67.74	21.48	67.58	46.10	8.22	64.00	55.79
88.58	71.83	19.74	65.32	45.58	8.21	67.11	58.90
84.78	65.04	23.09	67.84	44.74	8.86	64.30	55.44
86.72	69.75	20.60	64.35	43.75	8.92	66.36	57.44
83.85	71.77	23.84	66.13	42.28	9.92	64.61	54.69
86.55	69.75	20.81	64.43	43.61	9.13	66.86	57.73
85.98	62.74	24.74	69.40	44.66	9.23	65.23	56.00
84.72	68.18	24.69	67.15	42.46	9.92	65.18	55.26
87.24	66.60	24.73	71.20	46.47	8.99	65.78	56.79
87.47	67.72	24.67	71.55	46.88	8.94	65.86	56.93
87.06	64.28	24.93	70.66	45.72	9.21	65.86	56.65
85.40	66.71	24.94	67.71	42.77	9.93	65.68	55.75
86.43	65.01	24.13	68.24	44.11	9.89	66.55	56.65
97.02	95.79	6.16	75.12	68.97	3.99	76.44	72.44

2009

Fitting linear and exponential variomodels:

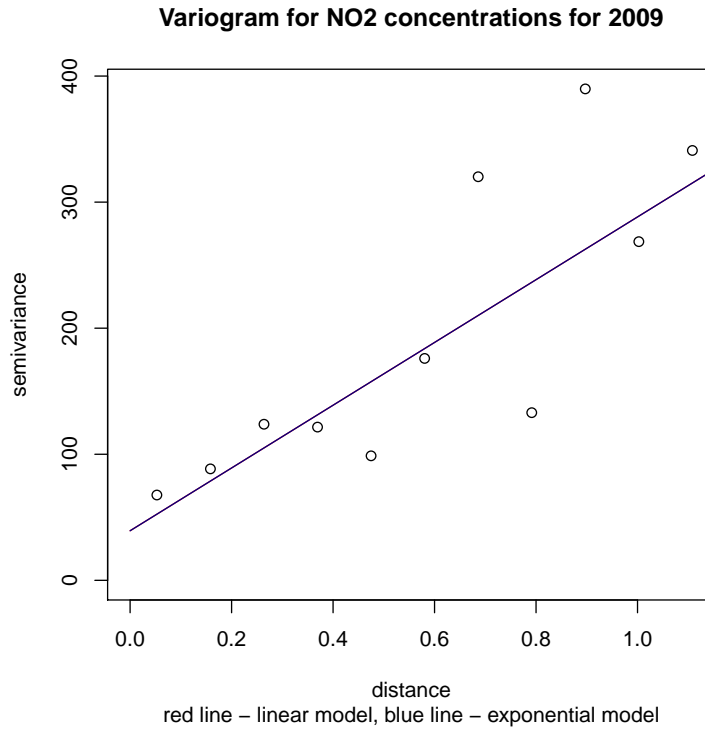


Figure 19: Empirical and modeled variograms

In Table 19, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 19: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
40.29	66.76	11.08	67.61	56.53
40.24	67.83	10.97	67.71	56.74
21.64	66.11	11.19	69.60	58.41
37.98	65.71	12.96	67.34	54.38
23.64	72.22	11.65	70.28	58.63
39.57	70.57	13.52	67.69	54.17
39.30	69.17	13.80	67.69	53.88
37.52	61.77	13.36	68.95	55.60
36.67	59.79	13.91	68.91	55.00
39.45	67.18	14.90	68.76	53.86
42.93	61.05	16.55	70.60	54.06
46.58	64.97	15.77	72.47	56.70
46.43	57.57	17.74	71.74	54.00
42.04	62.89	16.99	72.18	55.19
47.42	67.85	19.21	72.29	53.08
43.24	63.21	18.22	73.20	54.98
52.29	54.11	19.17	74.16	54.99
49.66	62.66	19.93	73.74	53.81
50.05	58.97	19.35	75.42	56.07
49.81	60.50	19.35	75.61	56.26
50.06	55.28	19.72	75.59	55.87
49.39	60.23	20.44	74.78	54.34
48.20	56.55	21.20	76.30	55.11
44.78	89.72	19.77	82.24	62.47

Fitting Gaussian variomodel:

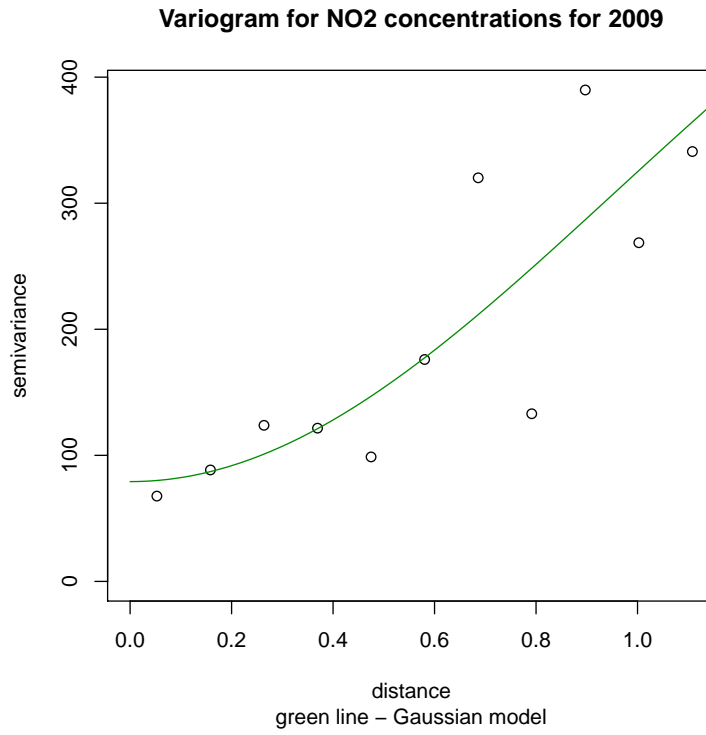


Figure 20: Empirical and modeled variograms

Table 20 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 20: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
89.88	66.76	16.50	64.61	48.11	9.90	62.48	52.58
90.10	67.83	16.63	64.78	48.15	9.92	62.62	52.70
92.61	66.11	8.61	69.45	60.85	5.09	62.94	57.86
87.15	65.71	15.73	63.53	47.80	9.42	61.16	51.74
93.01	72.22	8.77	71.40	62.63	5.26	63.13	57.87
86.75	70.57	17.40	64.08	46.68	10.62	61.64	51.03
86.36	69.17	17.03	63.91	46.88	10.22	61.29	51.07
87.70	61.77	19.44	65.47	46.03	11.38	63.21	51.83
86.97	59.79	19.57	65.29	45.72	11.57	62.97	51.41
85.71	67.18	18.35	64.69	46.33	10.09	61.28	51.19
85.20	61.05	21.14	66.45	45.32	10.75	61.77	51.02
88.66	64.97	20.97	68.59	47.62	12.38	66.26	53.88
84.73	57.57	23.18	67.48	44.30	11.53	62.23	50.70
86.48	62.89	21.85	67.56	45.71	12.89	65.43	52.54
83.54	67.85	24.17	67.82	43.65	12.98	63.01	50.02
86.40	63.21	22.31	67.88	45.56	13.19	66.01	52.82
85.60	54.11	25.54	70.16	44.62	12.06	63.28	51.21
84.26	62.66	25.40	69.21	43.81	13.03	63.57	50.54
87.07	58.97	25.92	72.13	46.21	11.84	63.76	51.93
87.35	60.50	25.92	72.48	46.56	11.78	63.84	52.06
86.83	55.28	26.18	72.11	45.93	12.14	63.95	51.81
84.99	60.23	25.90	70.32	44.41	13.11	64.10	50.99
86.45	56.55	25.89	71.97	46.08	13.28	65.10	51.82
112.77	89.72	8.05	80.10	72.05	9.75	76.03	66.28

PM₁₀

2001

Fitting linear and exponential variomodels:

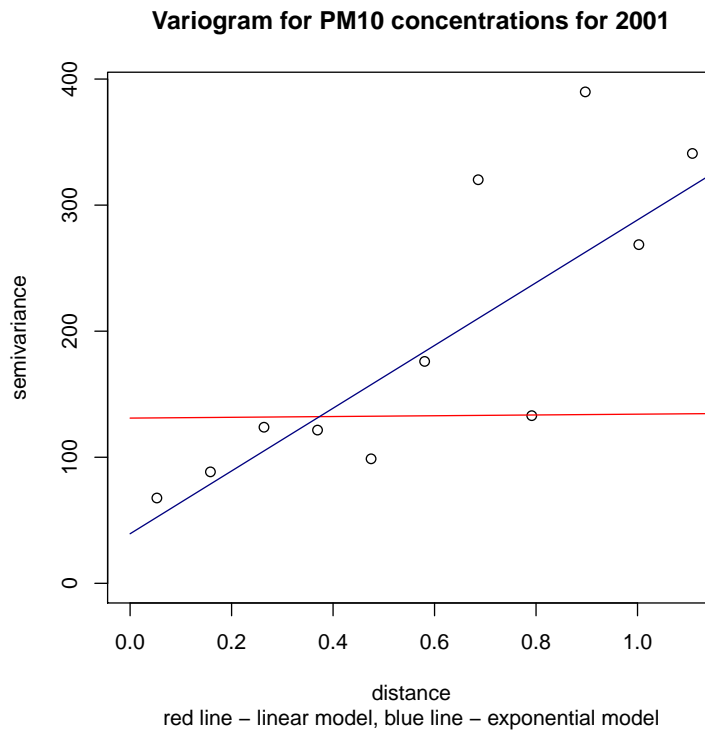


Figure 21: Empirical and modeled variograms

In Table 21, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 21: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
45.94	57.20	10.50	75.65	65.15
55.42	74.09	9.18	75.59	66.41
55.46	72.34	9.35	75.65	66.30
44.58	59.78	12.25	74.46	62.22
47.66	65.22	12.95	75.14	62.18
45.61	62.82	12.87	76.86	63.98
45.98	60.91	13.42	76.83	63.41
42.75	64.57	13.71	77.24	63.54
41.01	62.18	14.61	77.44	62.83
47.34	75.36	16.05	77.47	61.42
50.87	68.29	15.36	80.24	64.88
40.72	58.31	15.48	78.68	63.21
51.78	72.70	16.44	79.71	63.27
51.88	73.95	16.99	79.48	62.49
44.29	68.90	17.23	79.35	62.12
54.70	69.49	17.51	80.57	63.06
43.78	57.36	15.92	80.72	64.80
45.01	58.23	16.07	81.34	65.28
44.44	62.58	17.37	80.89	63.52
48.12	58.81	15.72	82.78	67.06
47.50	61.81	16.08	82.85	66.77
47.55	62.78	16.11	82.99	66.87
46.92	69.03	16.86	83.53	66.67
49.49	72.76	18.53	82.49	63.96
48.10	75.15	17.95	83.73	65.79
48.09	77.12	17.68	84.27	66.58
50.43	85.65	18.36	84.93	66.57

Fitting Gaussian variomodel:

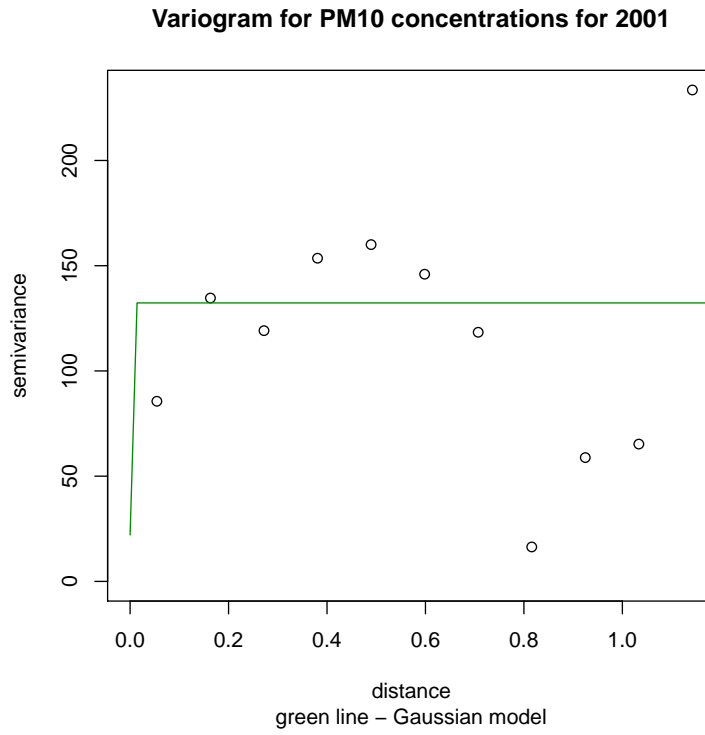


Figure 22: Empirical and modeled variograms

Table 22 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 22: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
138.34	57.20	4.79	75.34	70.55	3.02	75.35	72.34
138.34	74.09	5.66	85.16	79.50	-1.81	76.89	78.69
138.34	72.34	5.91	84.91	79.00	-1.61	76.77	78.38
138.34	59.78	7.20	76.32	69.12	2.95	74.16	71.21
138.34	65.22	7.98	76.19	68.22	4.01	74.27	70.26
138.34	62.82	9.46	77.78	68.32	4.37	75.72	71.36
138.34	60.91	10.07	77.97	67.90	4.62	75.55	70.92
138.34	64.57	11.45	78.98	67.53	3.01	74.60	71.59
138.34	62.18	12.17	78.39	66.22	3.88	74.35	70.46
138.34	75.36	12.95	79.76	66.81	5.75	75.18	69.42
138.34	68.29	14.19	85.61	71.42	4.74	78.88	74.14
138.34	58.31	12.75	78.89	66.14	4.45	74.57	70.13
138.34	72.70	15.27	83.95	68.67	5.44	77.81	72.38
138.34	73.95	15.48	83.25	67.77	5.77	77.22	71.44
138.34	68.90	13.59	80.25	66.66	5.76	74.98	69.22
138.34	69.49	16.75	85.03	68.28	5.61	78.38	72.77
138.34	57.36	12.09	80.48	68.38	4.50	75.26	70.77
138.34	58.23	11.65	81.14	69.49	4.52	75.51	70.99
138.34	62.58	12.89	81.42	68.52	5.53	75.55	70.02
138.34	58.81	9.74	83.05	73.31	4.05	76.09	72.05
138.34	61.81	9.90	83.12	73.22	4.30	76.14	71.84
138.34	62.78	9.73	83.34	73.60	4.31	76.20	71.90
138.34	69.03	9.41	84.49	75.08	4.71	76.59	71.87
138.34	72.76	12.77	84.67	71.90	5.82	77.01	71.19
138.34	75.15	10.06	85.81	75.74	5.27	77.17	71.90
138.34	77.12	8.86	86.52	77.65	5.02	77.30	72.28
138.34	85.65	8.56	88.61	80.05	5.15	78.17	73.02

2002

Fitting linear and exponential variomodels:

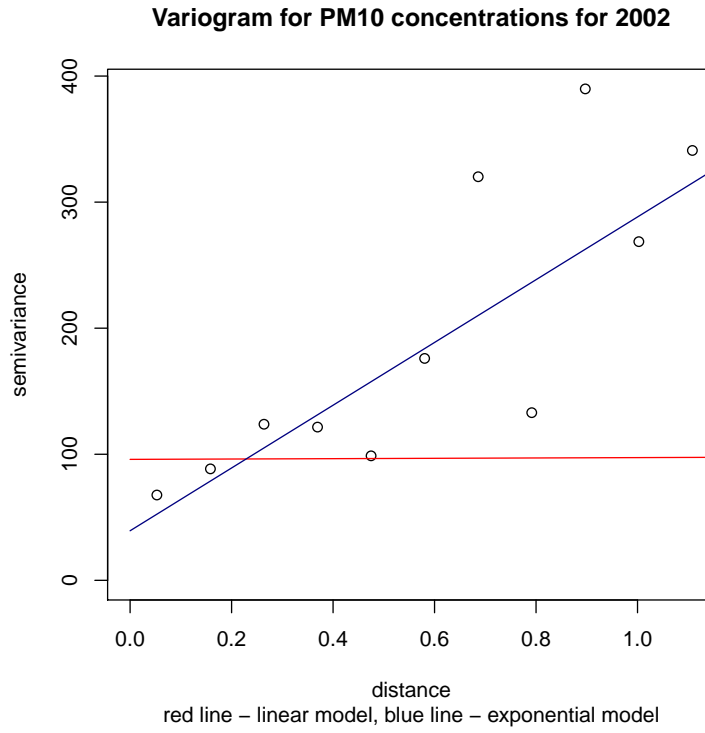


Figure 23: Empirical and modeled variograms

In Table 23, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 23: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
49.44	58.10	28.64	71.84	43.21
48.25	110.71	22.65	67.82	45.17
48.25	109.13	22.80	67.84	45.04
50.10	60.01	27.31	69.03	41.72
48.27	108.30	22.80	67.99	45.19
52.41	64.41	28.80	70.21	41.41
51.68	61.58	30.18	72.56	42.38
52.40	59.05	30.30	72.38	42.08
45.09	65.11	26.61	68.95	42.34
42.59	60.52	27.49	69.33	41.84
52.17	65.98	32.02	74.67	42.65
53.69	68.76	31.95	74.10	42.15
49.80	69.98	29.84	71.32	41.48
55.60	66.76	32.31	75.07	42.76
50.79	69.24	29.53	71.82	42.29
49.62	56.62	27.88	71.86	43.98
49.87	58.17	27.90	71.94	44.04
52.32	73.91	29.49	72.38	42.89
51.93	73.02	28.58	72.61	44.03
56.30	78.86	30.59	73.48	42.88
61.37	72.43	32.58	76.46	43.88
54.89	84.72	29.68	73.45	43.77
54.54	86.08	29.34	73.51	44.18
56.84	92.70	29.99	74.41	44.42
62.36	90.66	32.76	82.86	50.10

Fitting Gaussian variogram:

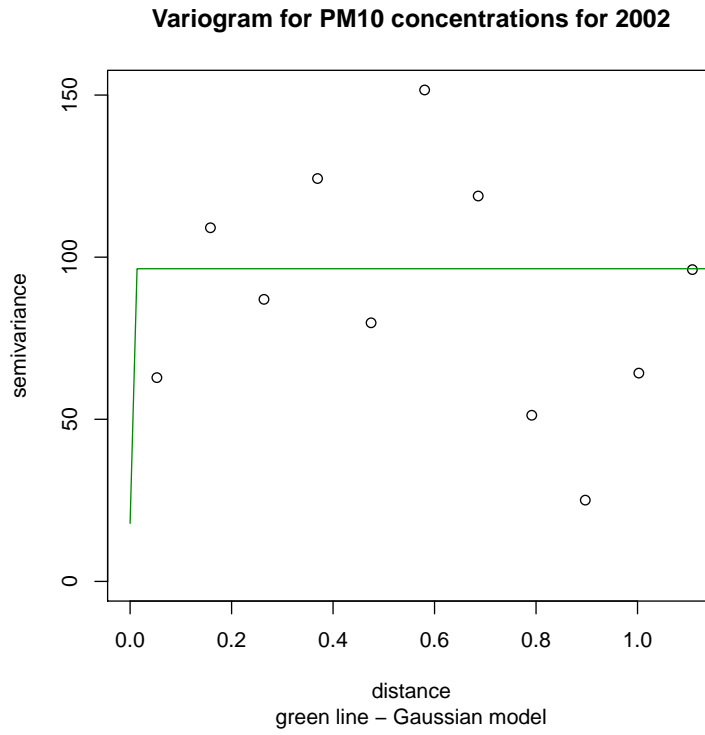


Figure 24: Empirical and modeled variograms

Table 24 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 24: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
100.45	58.10	25.35	71.05	45.71	17.61	67.47	49.86
100.45	110.71	14.52	72.85	58.33	13.06	67.31	54.25
100.45	109.13	14.81	72.62	57.81	13.20	67.23	54.02
100.45	60.01	24.98	70.73	45.75	17.02	66.00	48.98
100.45	108.30	13.78	72.83	59.04	13.21	67.21	54.00
100.45	64.41	26.83	71.25	44.42	17.99	66.29	48.30
100.45	61.58	28.36	71.69	43.33	18.55	67.65	49.10
100.45	59.05	28.80	71.89	43.09	18.69	67.49	48.80
100.45	65.11	23.83	68.45	44.62	16.65	65.80	49.15
100.45	60.52	25.82	68.76	42.94	17.34	65.70	48.36
100.45	65.98	32.91	75.53	42.62	19.23	69.23	50.00
100.45	68.76	33.34	75.64	42.30	19.38	68.71	49.33
100.45	69.98	30.48	72.60	42.12	18.85	66.40	47.55
100.45	66.76	34.04	77.07	43.03	19.27	69.62	50.35
100.45	69.24	30.14	72.94	42.80	18.60	66.69	48.09
100.45	56.62	26.69	70.74	44.05	17.53	66.78	49.24
100.45	58.17	26.70	70.93	44.23	17.54	66.82	49.28
100.45	73.91	29.98	74.03	44.05	18.49	67.05	48.57
100.45	73.02	27.74	73.40	45.66	17.88	67.20	49.32
100.45	78.86	31.81	77.70	45.89	18.89	67.92	49.03
100.45	72.43	35.57	81.63	46.06	18.90	71.03	52.13
100.45	84.72	29.56	77.09	47.52	18.38	67.82	49.44
100.45	86.08	28.59	76.94	48.35	18.17	67.84	49.68
100.45	92.70	29.23	80.40	51.17	18.31	68.61	50.30
100.45	90.66	28.48	98.79	70.32	15.41	79.30	63.89

2004

Fitting linear and exponential variomodels:

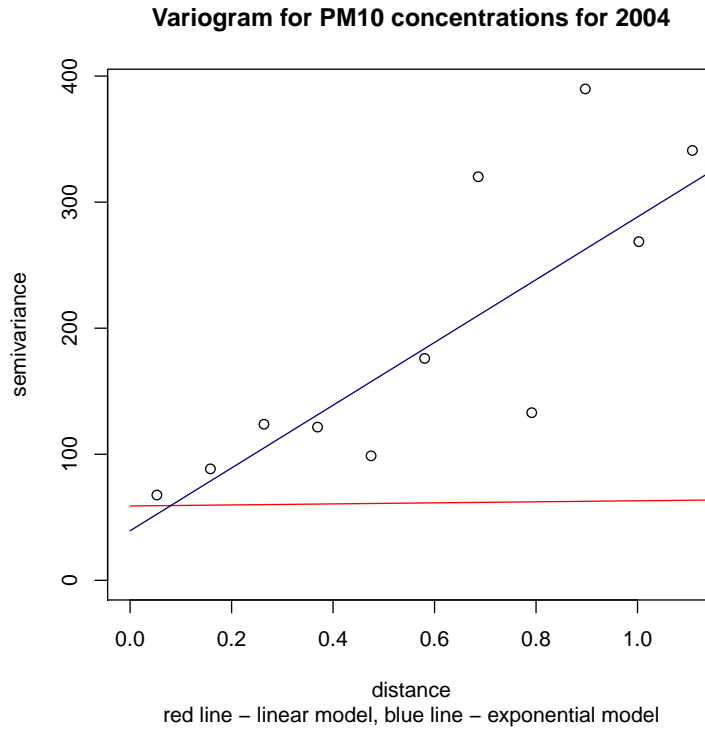


Figure 25: Empirical and modeled variograms

In Table 25, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 25: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
43.25	110.78	18.07	66.77	48.70
43.29	109.19	18.23	66.81	48.58
44.99	59.72	19.99	65.21	45.22
43.41	108.31	18.64	67.40	48.75
42.44	63.80	20.04	65.17	45.13
40.27	59.72	19.80	66.09	46.29
40.21	58.00	20.25	66.12	45.87
42.94	62.42	22.02	67.88	45.86
44.19	71.41	20.61	69.52	48.90
46.24	60.95	22.50	69.22	46.72
43.67	67.40	24.61	69.55	44.94
48.40	63.04	23.42	70.24	46.82
45.08	63.80	25.22	70.86	45.64
49.42	55.18	25.18	72.32	47.14
49.36	56.36	25.24	72.43	47.19
45.35	63.07	25.63	71.80	46.17
48.11	63.82	25.74	72.91	47.18
50.68	71.47	25.13	73.00	47.87
45.05	60.94	26.20	73.20	47.00
44.34	70.83	26.59	74.23	47.64
52.02	90.78	24.50	79.37	54.87

Fitting Gaussian variomodel:

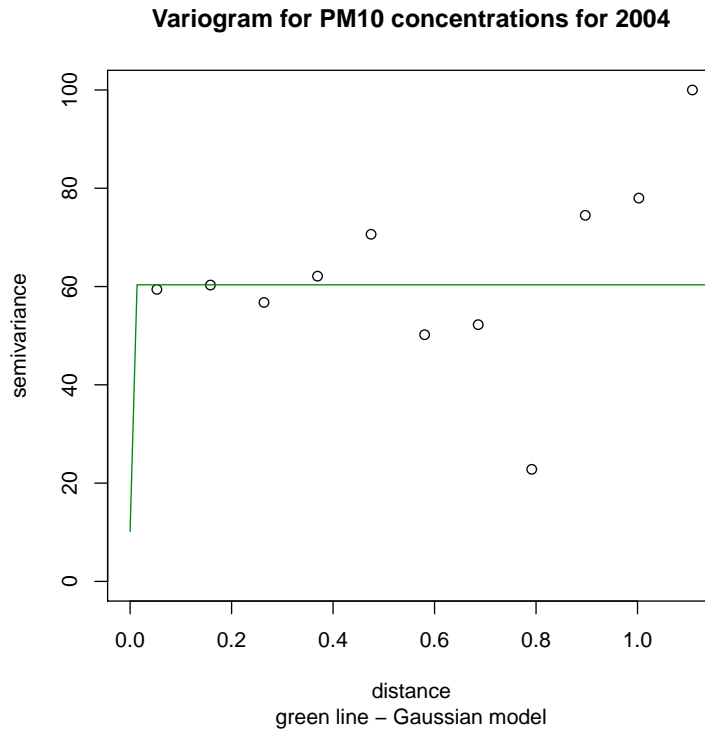


Figure 26: Empirical and modeled variograms

Table 26 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 26: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
62.51	110.78	9.54	70.59	61.05	20.31	57.30	36.98
62.51	109.19	9.95	70.47	60.52	20.41	57.23	36.82
62.51	59.72	18.34	65.58	47.24	22.45	55.22	32.78
62.51	108.31	9.32	71.43	62.11	20.48	57.33	36.85
62.51	63.80	19.06	65.19	46.13	22.67	54.94	32.26
62.51	59.72	19.30	65.57	46.26	22.37	55.46	33.08
62.51	58.00	19.79	65.72	45.93	22.56	55.37	32.81
62.51	62.42	20.16	67.67	47.51	22.63	55.84	33.21
62.51	71.41	18.29	71.24	52.95	21.53	57.55	36.01
62.51	60.95	22.31	70.08	47.76	22.56	56.83	34.27
62.51	67.40	24.71	70.04	45.33	23.81	55.97	32.16
62.51	63.04	23.57	71.59	48.02	22.63	57.27	34.64
62.51	63.80	25.45	71.54	46.08	23.83	56.42	32.58
62.51	55.18	24.90	72.53	47.63	23.40	56.90	33.50
62.51	56.36	24.93	72.74	47.82	23.40	56.94	33.54
62.51	63.07	25.75	73.09	47.34	23.81	56.78	32.98
62.51	63.82	25.43	74.35	48.93	23.57	57.15	33.58
62.51	71.47	25.62	76.13	50.50	22.43	58.60	36.17
62.51	60.94	25.80	76.33	50.53	23.70	57.41	33.72
62.51	70.83	25.54	79.48	53.94	23.52	57.97	34.45
62.51	90.78	15.69	93.08	77.39	18.56	64.04	45.49

2005

Fitting linear and exponential variomodels:

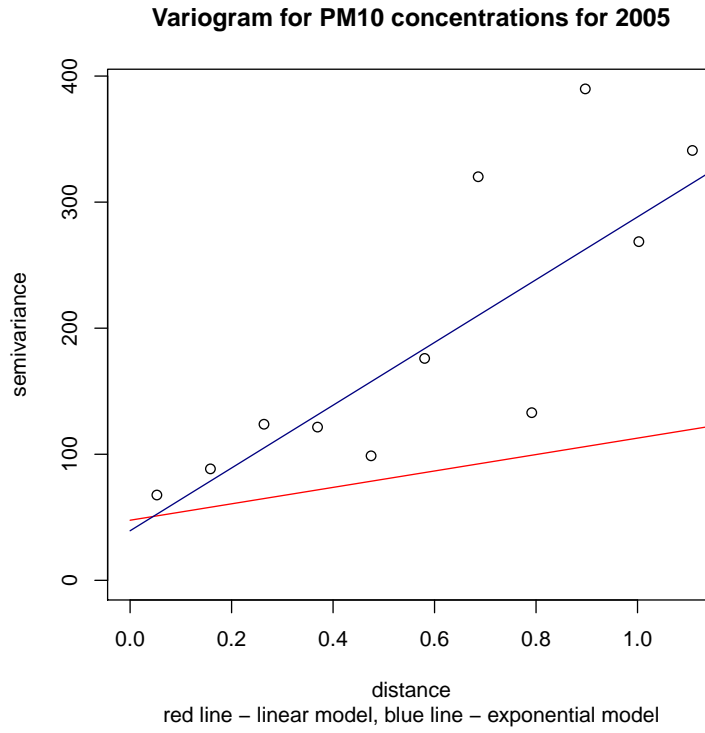


Figure 27: Empirical and modeled variograms

In Table 27, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 27: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
51.51	110.81	27.66	68.57	40.91
51.56	109.22	27.78	68.59	40.81
49.53	59.72	27.52	65.28	37.76
51.94	108.36	28.28	69.27	40.99
45.66	63.51	26.67	64.28	37.61
41.87	56.10	25.37	63.92	38.55
49.73	62.43	30.31	68.78	38.47
49.74	65.75	26.00	65.98	39.97
48.73	63.80	27.37	66.39	39.02
48.01	67.25	31.50	69.15	37.65
52.55	66.68	28.26	67.41	39.15
50.28	63.82	32.29	70.57	38.28
53.85	57.34	32.99	72.87	39.89
58.47	67.62	29.12	68.84	39.72
53.36	58.43	32.99	72.62	39.63
53.48	59.67	33.03	72.71	39.68
52.13	63.35	32.69	71.45	38.76
54.03	66.61	33.19	72.84	39.65
55.56	64.17	33.26	73.05	39.79
55.77	70.54	33.19	73.23	40.04
60.81	90.08	27.13	73.47	46.34

Fitting Gaussian variomodel:

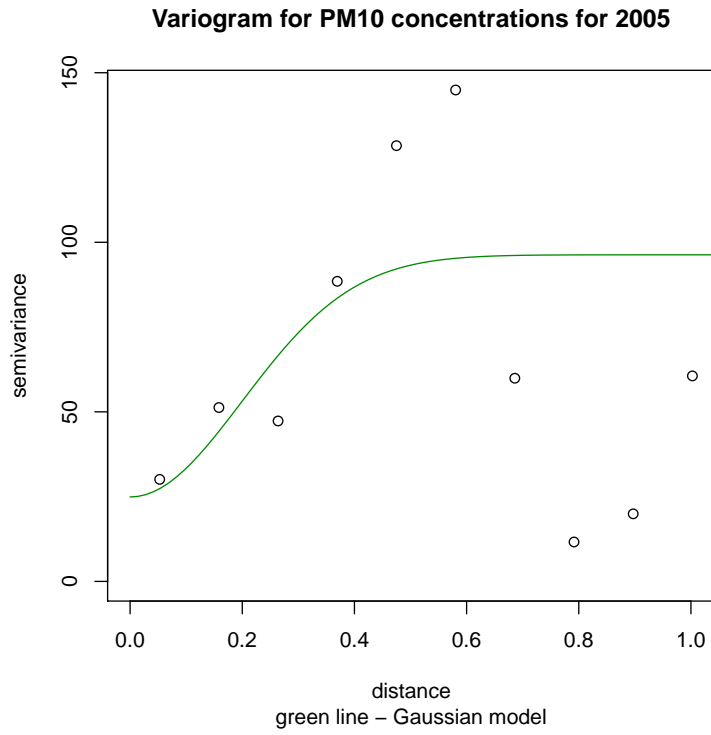


Figure 28: Empirical and modeled variograms

Table 28 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 28: Results($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
70.23	110.81	28.28	69.80	41.52	15.50	70.44	54.94
68.41	109.22	28.50	69.62	41.12	15.62	70.33	54.71
32.40	59.72	30.70	62.58	31.88	17.65	67.09	49.44
66.21	108.36	28.16	70.44	42.29	15.75	70.46	54.71
30.07	63.51	29.56	60.63	31.06	17.75	66.54	48.79
30.00	56.10	28.02	59.12	31.09	17.20	66.80	49.60
30.96	62.43	33.06	65.26	32.21	18.23	68.00	49.76
34.27	65.75	29.21	63.07	33.87	16.52	68.48	51.96
32.50	63.80	31.02	63.46	32.45	17.35	68.02	50.66
31.37	67.25	34.84	65.42	30.58	19.28	67.42	48.15
33.14	66.68	32.52	65.73	33.20	17.43	68.47	51.04
30.04	63.82	35.62	66.98	31.36	19.27	67.97	48.70
31.69	57.34	35.16	68.14	32.98	18.87	69.01	50.14
34.50	67.62	34.25	69.17	34.92	17.21	69.36	52.16
30.86	58.43	35.44	68.03	32.60	18.98	68.86	49.88
31.11	59.67	35.45	68.16	32.72	18.97	68.90	49.92
30.41	63.35	35.89	68.55	32.67	19.20	68.39	49.19
32.10	66.61	35.70	69.55	33.86	19.03	68.98	49.95
34.24	64.17	35.77	72.04	36.27	18.93	69.25	50.32
37.19	70.54	35.72	74.50	38.78	18.68	69.63	50.96
47.15	90.08	28.70	86.80	58.10	11.15	76.37	65.22

2006

Fitting linear and exponential variomodels:

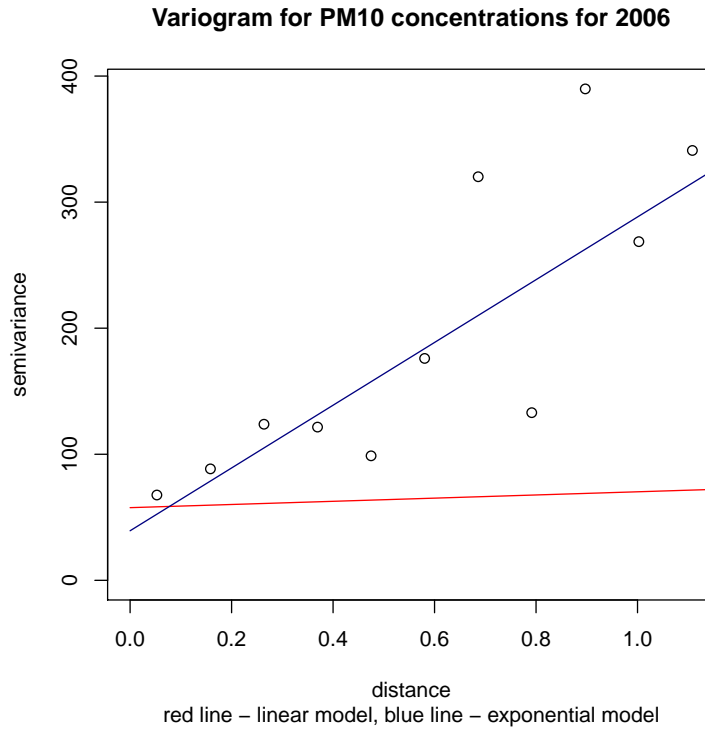


Figure 29: Empirical and modeled variograms

In Table 29, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 29: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
55.67	110.86	25.71	75.91	50.20
55.78	109.28	25.83	75.90	50.07
52.29	59.72	26.88	72.00	45.12
56.40	108.45	25.90	76.51	50.62
48.61	63.45	26.17	70.85	44.68
51.49	62.45	28.64	75.13	46.50
46.76	60.25	26.21	71.92	45.71
47.74	67.41	29.54	74.63	45.09
48.35	64.46	26.83	72.69	45.85
48.66	64.26	29.98	75.95	45.97
53.55	64.84	29.70	78.39	48.69
53.68	66.82	29.64	78.58	48.94
51.19	66.76	27.35	73.83	46.48
52.71	64.53	29.91	78.24	48.33
52.60	65.49	29.93	78.31	48.38
48.76	63.86	30.16	76.74	46.58
50.70	67.54	30.33	78.25	47.92
48.43	61.40	30.28	77.70	47.42
49.15	71.01	30.24	78.24	48.00
58.53	89.99	24.99	78.00	53.02

Fitting Gaussian variomodel:

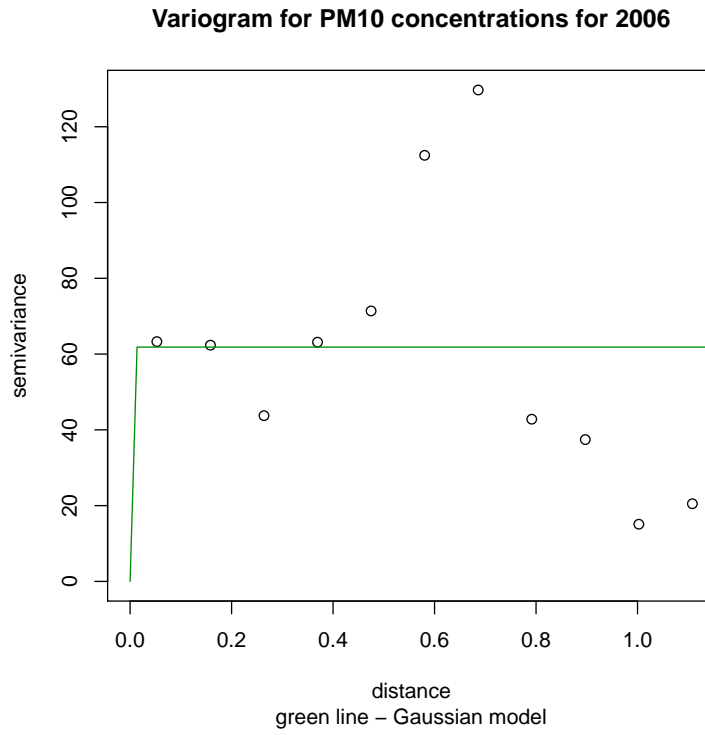


Figure 30: Empirical and modeled variograms

Table 30 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 30: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
63.97	110.86	24.10	92.20	68.10	13.98	70.38	56.40
63.97	109.28	24.41	91.84	67.44	14.12	70.29	56.16
63.97	59.72	26.45	76.87	50.42	17.71	68.22	50.51
63.97	108.45	23.10	93.56	70.46	14.10	70.29	56.19
63.97	63.45	24.56	73.41	48.85	18.47	68.19	49.72
63.97	62.45	29.30	80.80	51.50	17.33	68.33	51.00
63.97	60.25	23.90	73.48	49.58	18.78	70.10	51.33
63.97	67.41	26.67	74.97	48.30	19.00	68.15	49.15
63.97	64.46	24.27	74.20	49.93	18.69	70.40	51.70
63.97	64.26	26.42	75.76	49.33	18.71	68.46	49.75
63.97	64.84	26.91	79.59	52.67	17.61	68.86	51.25
63.97	66.82	26.76	79.92	53.15	17.53	68.92	51.39
63.97	66.76	24.39	75.48	51.09	18.33	71.13	52.80
63.97	64.53	26.56	78.94	52.38	17.76	68.86	51.09
63.97	65.49	26.39	78.96	52.57	17.76	68.89	51.13
63.97	63.86	25.61	76.30	50.69	18.52	68.76	50.25
63.97	67.54	25.04	78.21	53.17	17.96	69.06	51.10
63.97	61.40	23.56	77.28	53.72	18.24	69.37	51.13
63.97	71.01	21.57	78.26	56.70	18.00	69.97	51.96
63.97	89.99	15.61	95.73	80.11	13.50	78.99	65.49

2007

Fitting linear and exponential variomodels:

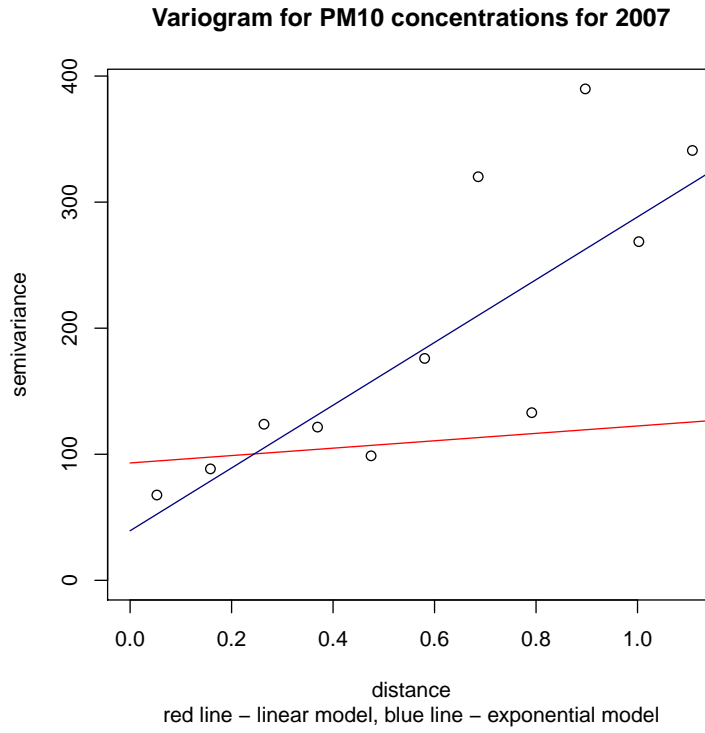


Figure 31: Empirical and modeled variograms

In Table 31, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 31: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
42.74	58.16	-1.89	89.50	91.39
59.59	110.79	-6.67	88.86	95.53
59.79	109.21	-6.48	88.81	95.29
48.39	59.72	-1.31	87.08	88.39
61.19	108.34	-6.62	89.04	95.66
44.26	64.16	-0.41	87.56	87.97
41.54	60.38	-0.38	89.44	89.83
41.32	58.52	-0.02	88.99	89.01
53.92	62.43	-1.83	87.81	89.64
42.82	59.72	1.05	90.74	89.69
40.35	63.47	1.24	88.25	87.01
44.30	61.55	1.39	90.99	89.60
46.90	56.02	-1.83	90.27	92.10
46.20	55.74	-1.31	90.13	91.44
46.34	56.46	-1.31	90.20	91.52
42.81	57.05	0.48	89.64	89.16
44.80	58.88	0.23	90.67	90.44
46.07	69.55	0.14	91.50	91.36
54.63	89.76	-1.96	100.25	102.21

Fitting Gaussian variomodel:

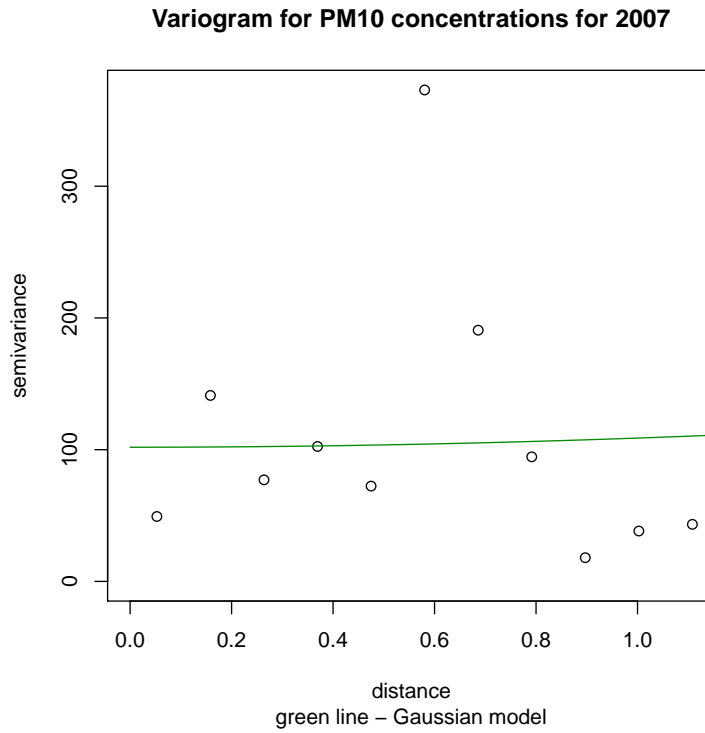


Figure 32: Empirical and modeled variograms

Table 32 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 32: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
105.87	58.16	3.52	82.98	79.46	-9.23	88.93	98.16
107.06	110.79	-13.13	94.92	108.05	-15.58	91.87	107.45
107.01	109.21	-12.45	94.58	107.03	-15.31	91.68	106.99
105.71	59.72	4.55	83.39	78.84	-9.07	87.48	96.55
107.00	108.34	-15.46	95.86	111.32	-15.29	91.67	106.96
105.52	64.16	5.14	82.12	76.99	-7.95	87.24	95.19
105.68	60.38	5.36	82.07	76.72	-8.04	88.52	96.56
105.59	58.52	5.64	81.85	76.21	-7.70	88.10	95.80
105.80	62.43	3.90	84.71	80.81	-9.38	87.67	97.05
105.88	59.72	6.85	83.87	77.02	-7.99	90.04	98.04
105.34	63.47	7.35	81.36	74.01	-6.61	86.85	93.46
105.95	61.55	7.31	84.42	77.11	-8.12	90.49	98.61
105.86	56.02	4.44	84.07	79.63	-8.92	88.44	97.36
105.79	55.74	5.17	83.47	78.30	-8.54	88.27	96.81
105.80	56.46	5.07	83.56	78.49	-8.55	88.32	96.88
105.59	57.05	6.41	82.58	76.17	-7.41	87.91	95.32
105.79	58.88	4.78	84.74	79.96	-7.97	88.91	96.88
105.97	69.55	3.16	87.32	84.16	-8.51	89.90	98.41
108.82	89.76	-8.64	113.64	122.28	-18.75	105.48	124.23

2008

Fitting linear and exponential variomodels:

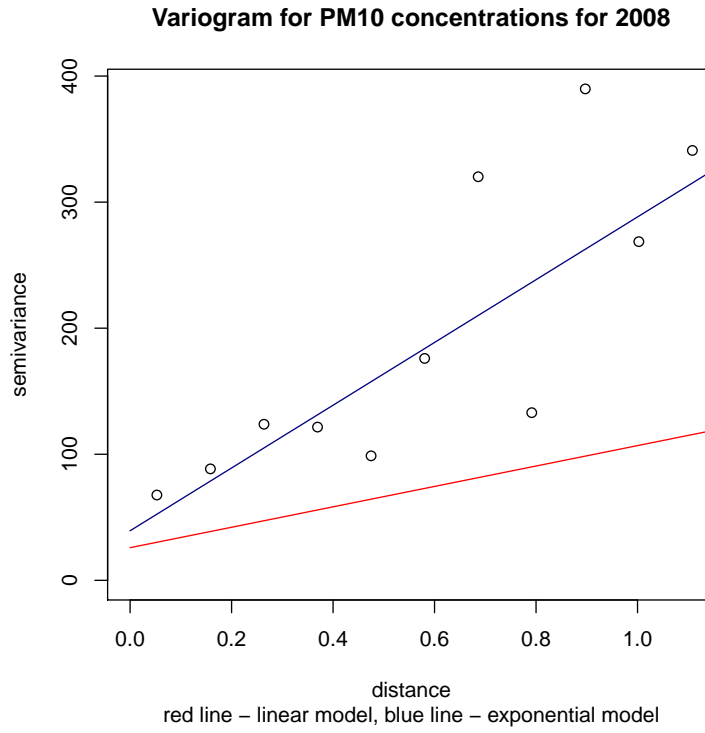


Figure 33: Empirical and modeled variograms

In Table 33, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 33: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
35.90	57.17	13.78	62.16	48.38
47.16	110.76	9.77	62.00	52.23
47.27	109.17	9.92	61.95	52.03
42.33	59.48	13.55	60.60	47.05
47.98	108.30	9.95	62.23	52.28
35.28	54.52	15.37	62.61	47.24
43.60	62.43	13.63	61.42	47.79
38.78	59.64	17.11	64.79	47.69
40.41	61.53	17.58	65.27	47.69
40.93	56.01	14.96	64.01	49.05
40.55	55.73	15.34	64.05	48.71
40.61	56.45	15.37	64.12	48.75
37.34	56.51	16.61	64.15	47.53
39.07	58.80	16.87	65.07	48.20
40.73	69.54	17.15	65.84	48.69
57.40	89.75	14.06	72.47	58.41

Fitting Gaussian variomodel:

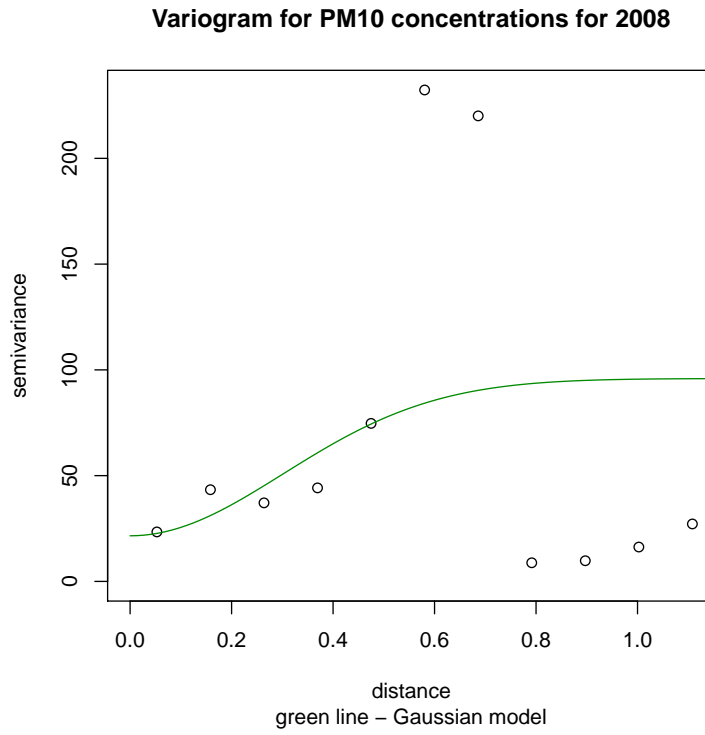


Figure 34: Empirical and modeled variograms

Table 34 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 34: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
24.87	57.17	16.20	56.20	40.01	8.12	62.21	54.09
44.75	110.76	10.73	66.28	55.55	3.86	63.17	59.31
43.81	109.17	11.03	66.00	54.97	4.02	63.08	59.06
25.91	59.48	18.04	57.96	39.92	7.86	61.12	53.26
43.34	108.30	9.52	67.16	57.64	4.06	63.10	59.05
24.42	54.52	18.05	56.55	38.50	9.25	62.06	52.82
25.39	62.43	18.18	59.44	41.26	7.75	61.32	53.57
24.73	59.64	21.03	60.23	39.20	9.74	63.84	54.09
24.85	61.53	22.12	61.42	39.30	9.82	64.25	54.43
25.03	56.01	18.36	59.02	40.66	8.51	62.28	53.77
24.56	55.73	18.80	58.77	39.97	8.79	62.26	53.47
24.64	56.45	18.76	58.83	40.06	8.79	62.30	53.51
24.01	56.51	19.88	58.73	38.85	9.65	62.29	52.64
25.59	58.80	19.55	60.33	40.78	9.54	63.05	53.51
27.70	69.54	19.36	62.28	42.91	9.43	63.79	54.36
38.93	89.75	9.75	86.81	77.07	5.53	74.19	68.66

2009

Fitting linear and exponential variomodels:

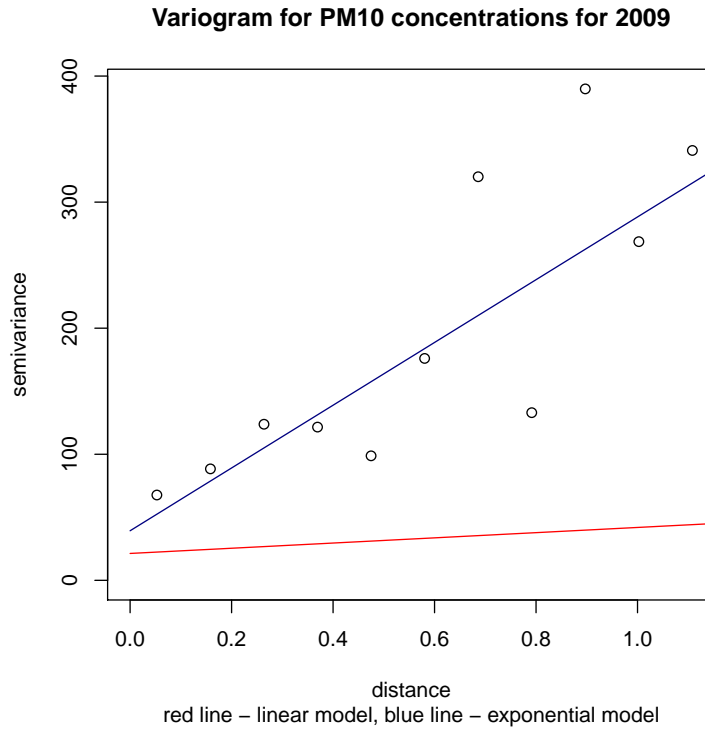


Figure 35: Empirical and modeled variograms

In Table 35, columns 1 and 2 correspond to ordinary kriging model with exponential covariance, and columns 2-5 correspond to linear RRCM iid model. Ridge factor equal to 0.01, confidence level equal to 0.95.

Table 35: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval
35.41	57.17	20.71	48.92	28.22
29.26	66.09	19.72	48.58	28.86
30.20	72.35	19.90	48.84	28.94
35.47	54.52	22.02	49.61	27.59
35.35	61.37	22.02	49.53	27.52
37.61	59.64	23.55	51.41	27.86
37.98	61.53	24.05	51.91	27.87
37.33	58.84	24.04	51.49	27.44
37.20	53.81	23.19	51.97	28.77
37.25	51.43	23.40	51.98	28.58
37.41	58.63	24.13	51.97	27.84
35.60	56.30	24.35	52.75	28.39
37.64	89.65	25.35	56.91	31.56

Fitting Gaussian model:

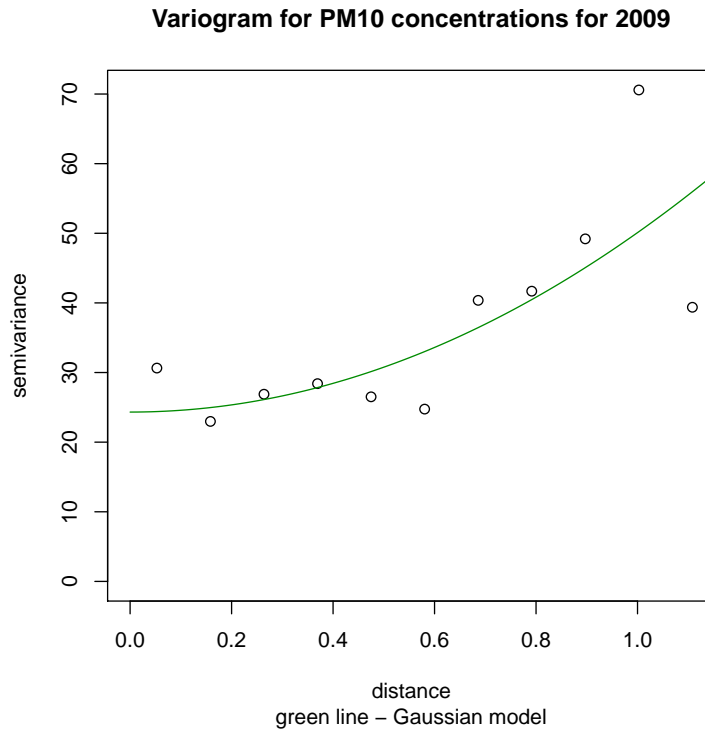


Figure 36: Empirical and modeled variograms

Table 36 shows the following results: columns 1 and 2 correspond to kriging prediction and variance, columns 3-5 correspond to RRCM model with Gaussian kernel, columns 6-8 correspond to RRCM model with polynomial kernel.

Table 36: Results ($\mu\text{g}/\text{m}^3$)

kriging	variance	lower	upper	interval	lower	upper	interval
25.87	57.17	21.88	49.01	27.14	17.03	48.55	31.53
26.54	66.09	16.99	49.04	32.06	14.56	48.69	34.12
26.57	72.35	16.78	49.85	33.07	14.60	48.71	34.12
25.50	54.52	23.32	49.48	26.16	17.67	48.47	30.80
25.50	61.37	22.34	48.87	26.53	16.77	47.86	31.09
25.74	59.64	24.16	50.85	26.68	17.91	49.47	31.56
25.77	61.53	24.44	51.19	26.75	17.93	49.68	31.75
25.31	58.84	24.70	51.17	26.47	17.93	48.30	30.37
25.84	53.81	23.49	51.63	28.15	17.24	48.49	31.25
25.75	51.43	23.80	51.63	27.83	17.39	48.48	31.09
25.47	58.63	24.56	51.71	27.15	17.87	48.52	30.65
25.71	56.30	24.17	52.79	28.62	17.79	48.95	31.16
28.46	89.65	16.07	58.50	42.43	15.03	54.98	39.95