Neural Modelling of the Spatial Distribution of Air Pollutants

A new method developed considering as example Cyprus

Von der Fakultät Geo- und Biowissenschaften der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Heiko Pfeiffer

aus Stuttgart

Hauptberichter:	Prof. Dr. W. D. Blümel
Mitberichter:	Prof. Dr. G. Baumbach
Tag der mündlichen Prüfung:	14.7.2005

Institut für Geographie der Universität Stuttgart

2006

Vorwort

Die vorliegende Dissertation entstand während meiner Tätigkeit am Institut für Verfahrenstechnik und Dampfkesselwesen der Universität Stuttgart in der Abteilung Reinhaltung der Luft. Dort hatte ich die Gelegenheit, an dem äußerst interessanten Projekt “Preliminary Assessment of Ambient Air Quality in Cyprus” mitzuwirken und dabei verschiedene Bereiche des weiten Gebietes der Luftreinhaltung kennenzulernen.

Mein Dank gebührt vor allem dem Leiter der Abteilung Reinhaltung der Luft, Herrn Prof. Dr.-Ing. G. Baumbach, der mich jederzeit unterstützte und wissenschaftlich orientierte. Seine freundliche und motivierende Art war und ist für mich eine persönliche Bereicherung.

Für die Übernahme des Hauptberichts danke ich Herrn Prof. Dr. rer. nat. W.-D. Blümel vom Institut für Geographie, dessen freundliche und konstruktive Art ich als überaus positiv empfand.

Weiterhin sei den Mitarbeitern des Projekts auf Zypern gedankt, die mit ihren Messungen die Grundlage für die Entwicklung meines Modells legten. Sie sind alle namentlich im Anhang genannt.

Meiner Familie und meinen Freunden danke ich dafür, dass sie immer für mich da waren.

Table of Contents

Abbreviation Index

UNOPS	United Nations Office for Project Services
NO_2	Nitrogen Dioxide
SO_2	Sulphur Dioxide
VOC	Volatile Organic Compounds
O_3	Ozone
NO_x	Nitrogen Oxides
$PM_{2,5}$	Particulate matter < 2,5 µm
PM_{10}	Particulate matter < 10 µm
NO	Nitrogen Oxide
MLP	Multilayer perceptron
CO	Carbon Monoxide
LST	Local sidereal time
UTC	Coordinated universal time
X-RFA	X-ray fluorescence
AAS	Atomic absorption spectrometry
GIS	Geo information system
hv	Radiation intensity
IWD	Inverse distance to a power interpolation
LCR	Learn control rate
UTM	Universal Transverse Mercator
GCC	Greek Cypriot Community
TCC	Turkish Cypriot Community
DEM	Digital elevation model
r^2	Coefficient of determination

Zusammenfassung

Das Ziel dieser Arbeit war die Entwicklung einer Methode zur Berechnung der flächenhaften Verteilung von Luftverunreinigungen. Atmosphärische Ausbreitungsmodelle modellieren die Verteilung von Gasen oder Partikel anhand mathematisch formulierbarer Wirkmechanismen. Messergebnisse von Luftverunreinigungen werden nur zur Validierung der berechneten Verteilung, bzw. zur Modellentwicklung eingesetzt. Jedoch ist die Verwendung vieler, gut verteilter Messdaten als Modelleingabe die einzige Möglichkeit, die tatsächlich existierende Verteilung der gesuchten Komponente zu berücksichtigen. Mit einem Netz von Passivsammlern kann eine solche Verteilung kostengünstig und zeitgleich erhoben werden. Im Rahmen des UNOPS-Projekts "Preliminary Assessment of Ambient Air Quality in Cyprus" wurde an 270 Punkten NO_2 Passivsammler in sechs Messkampagnen exponiert. Auf diese Weise entstand eine gute Datenbasis für die Entwicklung einer Methode, mit der man die tatsächlich gemessene Verteilung von Luftverunreinigungen berücksichtigen kann. Darüber hinaus sollte man bei einem realitätsnahen Modell die wichtigsten Einflüsse, wie etwa die Bevölkerungsdichte oder die Emissionsverteilung miteinbeziehen. Gegenwärtig erfüllen drei Verfahren diese Anforderungen: Regressionsmodelle, Interpolation und Künstliche Neuronale Netze. Regressionsanalysen sind zwar generell geeignet, kommen hier jedoch nicht in Frage, da es sich um eine starre Methode handelt mit sehr vielen theoretischen Bedingungen, die man in der Praxis schwer einhalten kann. Die Interpolation ist ein gut entwickeltes Standardverfahren, welches auch von der Europäischen Union empfohlen wird. Es wurde deshalb im Rahmen dieser Arbeit untersucht, inwieweit man mit dieser Methode eine realistische Verteilung von Luftverunreinigungen modellieren kann. Einfache Interpolationsalgorithmen eignen sich hierfür nicht, da das Ergebnis stark von der geographischen Lage der Messpunkte abhängt. Quellen und Senken zwischen zwei Stützpunkten werden vernachlässigt. Eine mögliche Lösung für dieses Problem sind Interpolationsalgorithmen wie Cokriging, mit denen man zusätzliche Variablen betrachten kann. Jedoch konnte auch hier keine Zufriedenstellende Ergebnisgüte erzielt werden. Dennoch ist die Interpolation für eine schnelle Visualisierung von Messungen durchaus geeignet.

Künstliche Neuronale Netze sind derzeit die einzige Möglichkeit, ein Modell unter Berücksichtigung der obengenannten Kriterien zu entwickeln. Der heute meistverwendete Netzwerktyp im Bereich Luftqualitätsmodellierung ist das sogenannte "Multilayer Perceptron", das auch im Rahmen dieser Arbeit zum Einsatz kam. Zwei prinzipielle Kriterien beeinflussten die Entwicklung des Modells: Die Netzwerktopologie, also die Anordnung der einzelnen Neuronen im Netzwerk, inklusive deren Eigenschaften und vor allem die Wahl der Eingangsvariablen. Im Trainingsmodus wurde ein sogenannter kontrollierter Lernalgorithmus verwendet, bei dem der Anwender das Netz mit Ein- und bekannten Ausgangsvariablen, also Luftverunreinigungsmessungen trainiert. Die Aufgabe des neuronalen Netzes ist das Erlernen der zumeist nichtlinearen Zusammenhänge zwischen Ein- und Ausgabe. Zu diesem Zweck wurde ein Analysegitter mit 1x1 km Kantenlänge über das Untersuchungsgebiet Zypern gelegt und jeder Gitterzelle wurden die entsprechenden Eingabevariablen, wie etwa die UTM-Koordinaten zugeordnet. Bei den Gitterzellen mit Passivsammlern konnten die Zusammenhänge zwischen Eingabe und Ausgabe ermittelt und anschließend auf alle anderen Zellen übertragen werden. Im Laufe der Entwicklung wurden alle verfügbaren Variablen in uni- und multivariaten Modellen getestet.

Mit den UTM-Koordinaten als Eingangsdatensatz konnte zunächst eine neuronale Interpolation erzielt werden. Das Ergebnis war eine vereinfachte Interpolationskarte mit NO_2 Konzentrationen von 30 bis 40 µg/m³ in den Städten und geringeren Konzentrationen in ländlichen Gebieten. Bemerkenswert sind hierbei unrealistische, gerade Streifen von NO_2

Konzentrationen um 20 µg/m³, welche die Städte Nicosia, Limassol und Larnaka miteinander verbinden.

Anschließend wurde dem Analysegitter ein digitales Höhenmodell angepasst und das Netzwerk mit den resultierenden Höhenwerten trainiert. Auch hier traten physikalisch unmögliche Werte auf, wie etwa hohe NO_2 Konzentrationen in Flusstälern. Dies verdeutlichte die Notwendigkeit, bei der Entwicklung nur physikalisch sinnvolle Variablen einzusetzen, trotz der Fähigkeit Neuronaler Netzen jegliche Art von Zusammenhängen, auch beispielsweise Trivialkorrelationen zu erkennen. In diesem Fall korrelierte das Netzwerk den Umstand, dass alle großen Städte Zyperns am Meer liegen (geringen Höhenlage) mit dem Auftreten hoher NO_2 Konzentrationen.

Ein sehr wichtiger Einfluss auf die Verteilung von Luftverunreinigungen sind die atmosphärischen Ausbreitungsbedingungen. Da alle meteorologischen Parameter ständig variieren ist es unmöglich ein mittleres Windfeld zu berechnen. Windstatistiken wären eine mögliche Lösung, können jedoch nicht direkt als Eingabe verwendet werden, da es sich im Falle der Windrichtung um eine Verteilung von mehreren Werten handelt. Um dennoch die Ausbreitungsbedingungen zu berücksichtigen, wurden Abgasfahnen der wichtigsten Emissionsquellen in Zypern berechnet. Methodisch kamen hierbei das Gauß-Modell P&K 3782 und statistische Analyseverfahren wie Regressionsanalysen zum Einsatz. Die Ergebnisse wurden mit einem neuen rechnerischen Ansatz auf das Analysegitter verteilt, wobei das Konzept der "Distributed Emissions" entwickelt wurde. Mit diesem Input für das Neuronale Netz konnte bereits eine sehr genaue NO_2 Immissionskarte berechnet werden, auf der die Lage der Quellen und deren Emissionsstärken gut wiedergegeben werden.

Nach den oben beschriebenen univariaten Modellen wurden multivariate Berechnungen durchgeführt, um auch noch die vorhandenen weiteren Einflussparametern miteinzubeziehen. Alle Modelle mit UTM-Koordinaten und Höhenwerten als Eingabe produzierten wiederum unrealistische Verteilungsmuster.

Das beste Ergebnis konnte mit einem Neuronalen Netzwerk erzielt werden, das mit "Verteilten Emissionen" und der Populationsdichte trainiert wurde. Diese Modellkonfiguration bewahrte die positiven Aspekte der univariaten Ansätze und machte außerdem noch weitere Quellen wie Dörfer deutlich sichtbar. Eine realistische, fein strukturierte Immissionskarte von Zypern ist das Resultat dieser Berechnungen.

Während der gesamten Entwicklungsphase wurden die Modellergebnisse einer ständigen statistischen und visuellen Prüfung unterzogen. Wichtigster Bestandteil war hierbei ein Testdatensatz aus 50 Passivsammlern zur Überwachung des Trainingsfortschritts und zum direkten Vergleich von berechneten und gemessenen NO2-Konzentrationen. Mit dem Modell konnte ein Pearson Korrelationskoeffizient von 0,75 erzielt werden.

1 Introduction / Aims / Questions

Strict laws and guidelines together with constant efforts of the scientific community in the development of new technologies helped to improve the air quality in western countries to date. This positive trend is expected to be partly compensated by the increase of pollution sources called forth by a parallel growth of demand for energy. Less developed countries rely strongly on poor quality fossil fuel and on technology with high pollutants emission rates. In addition, large parts of the world population are potential new energy consumers often neglecting environmental issues. A continuous monitoring of the air pollution level is therefore inalienable and the base for adequate actions. Air quality measurements however show only a small part of the whole situation and may miss hot spots since they are only valid for the site where the measurement units are located. The spatial distribution of the pollutants remains unknown since it depends on the actual dispersion conditions and the emission strength. A possibility to partly bypass this problem is to use mobile measurement stations that can be moved after a certain time [1] or to operate a network of several stations. This is however economically not always acceptable. A cost-effective alternative are diffusive samplers which can be placed at an almost unlimited number of sites, but the principal problem for all methods remain unsolved: Measurements are punctiform, with some minor exceptions for special problems. Air quality models can partly fill this gap and therefore are essential tools in the development of action plans for improving air quality. Thus, the EU Framework Directive 96/62 EC on air quality assessment refers to “the use of other techniques of estimation of ambient air quality besides direct measurements” [2]. It is stated, that these “other techniques” may be modelling techniques. In preliminary assessment member states will designate zones and design measurement networks in each zone according to modelling results. Rather than sheer numbers, these results should be presented in an intelligible way – in the form of pollutant maps [3,4]. This opens the possibility for a more cost-effective and goal-orientated assessment strategy. Measurement stations can be placed on sites where limit value exceedances are predicted by the model, which might save investments in expensive equipment.

In the Framework Directive and its Daughter Directives, no special modelling techniques are mentioned, whereas single countries have their own provisions on the usage of specific models, like “Austal2000” [5] in the German TA-Luft [6]. Equivalent guidelines for European right are in preparation. Most of the models are mainly suitable for case scenarios or they are limited in their flexibility and their possibility to consider the most important influencing factors. To calculate territory covering pollutant maps in high resolution, a method is needed that can establish correlations between the pollutant and it’s main influencing factors like the dispersion conditions, the strength and location of emission sources and the population density. But still measurements are the only way to provide a direct link to the reality. So the optimal model should consequently include results from air quality measurements. For most of the current approaches this is not case. The aim of the following work is to find such a method – a model that enables us to calculate pollutant concentration maps, covering a large area like Cyprus.

During the UNOPS project “Preliminary Assessment of Ambient Air Quality in Cyprus” in which the author was involved, measurements and surveys were carried out, providing an excellent database for the model development. Of special interest is the diffusive sampling programme, carried out at 270 sites over the course of one year.

Below the eligible established and new modelling approaches, the most promising ones are:

- regression analysis
- interpolation algorithms considering additional variables
- artificial neural networks

Artificial neural networks are able to quantify complex relationships of different variables without particular preconditions like normal distribution. Comparable to the human brain, neural networks are flexible and dynamic, with the ability to learn. A lot of limitations from conventional methods do not apply here. That’s why this work mainly investigates the possibility to use artificial neural networks for the calculation of distribution maps of air pollutants.

1.1 Background information - Project “Preliminary Assessment of Air Quality in Cyprus”

As a new member of the European Union, Cyprus has to meet European legislation. Within the field of environmental protection a preliminary assessment of air quality is required, based on the Council Directive 96/62/EC. The aim of this air quality policy is to provide a healthy environment for inhabitants of Cyprus and to prevent diseases, which can be caused by air pollutants especially in the respiratory system of humans. Another target is to offer a clean environment to the tourists who come to Cyprus for recreation. To realise this preliminary assessment, a project was financed by the US Agency for International Development (USAID) and the United Nations Development Programme (UNDP). It was executed by the United Nations Office for Project Services (UNOPS) and the Greek and Turkish Cypriot Communities. Contractor for carrying out the project was the University of Stuttgart together with Cypriot partners.

The main objectives of the project are as follows:

- Assessment of the spatial distribution and temporal variation of air pollutants over the whole island of Cyprus
- To assist Cyprus to optimise the ambient air quality monitoring network in order to comply with the relevant Directives of the European Union including the reporting to the commission
- To supply the necessary input for the formulation of air pollution management policies in Cyprus including preparation of plans on how to meet the EU limits and other EU requirements
- To increase public awareness on the issues of urban and rural air pollution

Since the air does not recognise borders, measurement programmes over the course of one year were carried out in both communities of Cyprus. The tasks were as follows:

1. Emissions Inventory: Daily emissions of all major line, point and area sources in Cyprus
2. Diffusive Sampling: Spatial distribution of the pollutants NO_2, SO_2, VOC (Volatile Organic Compounds) and O_3 at up to 270 categorised sites all over Cyprus
3. Continuous Monitoring: Temporal variation of the most important pollutants with fully equipped measurement vans/containers, NO_x/O_3 background monitoring stations, weather stations and VOC measurements at selected sites
4. Tethered Balloon Measurements: Vertical variation of pollutants and meteorological parameters, measured in Nicosia and Limassol
5. Particulate Matter (PM) Measurements: Spatial and temporal variation of PM, including receptor modelling with principal component analysis

6. Modelling: Spatial distribution of the pollutants based on diffusive sampling results
7. Air Quality Management: Recommendations on how to meet EU regulations and to set up the future monitoring network in Cyprus

All these tasks are summarized in Figure 1.1.

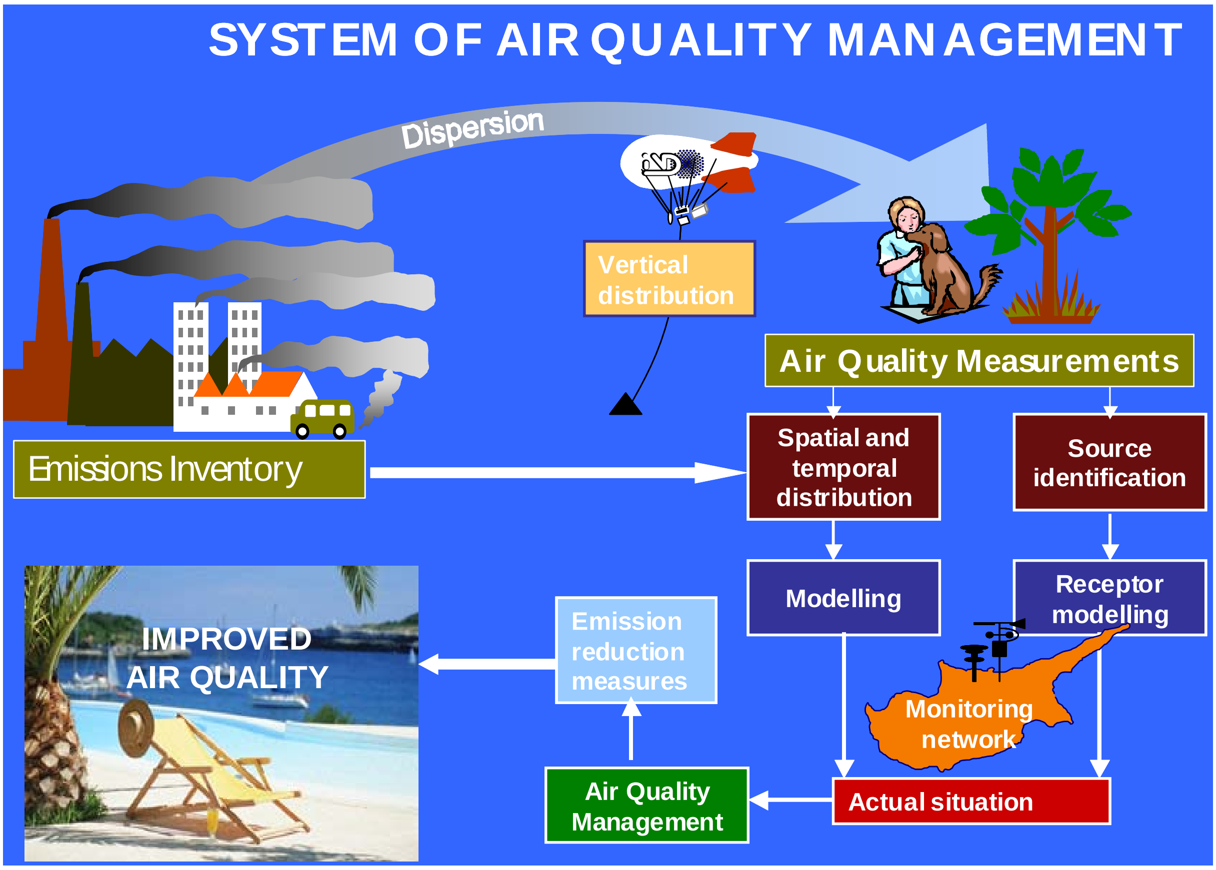


Figure 1.1. System of air quality assessment as applied during the UNOPS project "Preliminary Assessment of Air Quality in Cyprus" [34]

1.2 Methodical approach

The aim of this work was to develop a model, with which one is able to calculate the spatial distribution of pollutants. According to the project requirements, the model should have the following characteristics and abilities:

- Ability to handle diffusive sampling measurements as model input and not only for validation purposes
- High accuracy of the model results
- Ability to incorporate the most important parameters, that influence the distribution of air pollutants
- The model should be easy to handle
- Short computing time and no high hardware requirements
- Ability to update and enhance the model

The general approach to develop a method that meets these requirements is depicted in Figure 1.2.

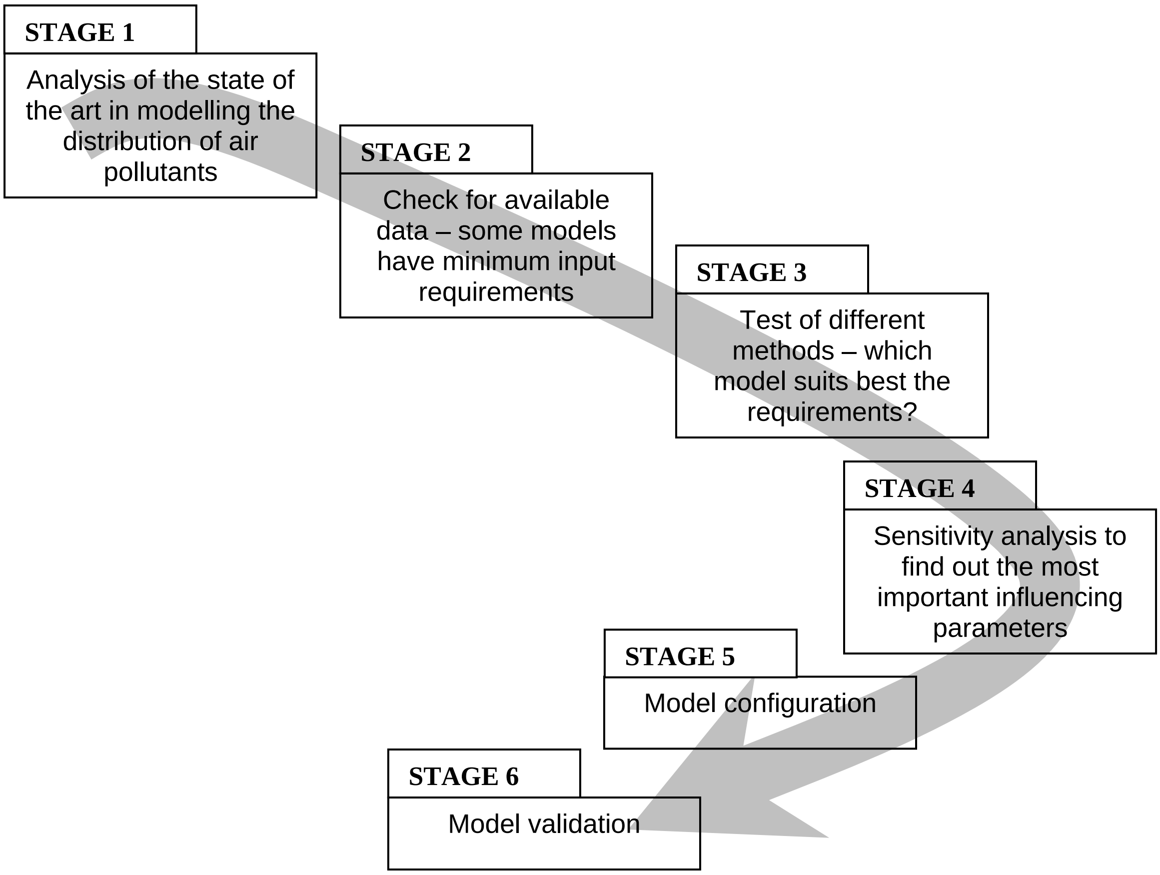


Figure 1.2. General approach to model the spatial distribution of air pollutants

2 State of the art in air quality modelling

As described above, the aim of this work was to develop a model that has certain quality characteristics. This presumes that there currently exists no available method that suits this demand. In the following state of the art analysis, available methods are checked with regard to the required criteria.

2.1 Dispersion modelling

With dispersion models the dispersion of pollutants starting from one or more emission sources that might be point, line or area sources is calculated. Depending on the requirements, dispersion models are able to consider the topography, the development, the velocity-, turbulence- and temperature-field, chemical and physical alterations and other parameters [7].

There are different possibilities to classify dispersion models:

- Mathematical and physical models [8]
- Mesoscale or microscale models [9]
- Models with a diagnostic or prognostic flow field pre-processor
- Simple models with homogenous terrain, models that consider the relief and land usage, models that consider the development [7]

Today’s models are usually combined with a pre-processor, which calculates the flow field – the end result depends strongly on the beforehand-calculated flow field. For local scale considerations, CFD (computational fluid dynamics) models are first choice today [10, 11].

Generally it should be emphasized, that the requirements for the input data are very high [12]. Their preparation and acquirement usually is very time consuming, as well as the final calculation run on the computer, which may take several days depending on the model and the regarded problem.

Table 2.1 gives a summary of the currently available dispersion models. The classification here follows the mathematical principle of the method [8, 62].

2.2 Interpolation methods

In many countries simple interpolation algorithms are officially applied, e.g. Kriging, Inverse Distance Weighting, Modified Shepard’s Method and Radial Basis Function. [15]

Input parameters are the geographical coordinates and the pollutant concentration values. The calculated concentrations at a certain site are a function of the distance to the measurement points [7]. According to the implemented approach, the number, direction and distance to the real concentration values can be considered. With Kriging, it is moreover possible to include the spatial variation of the measured concentrations by using variograms [16].

One can differentiate between statistical and non-statistical approaches or “exact” (the input value is preserved in the output) and “inexact” methods. Single cases are normally not considered, the adjusted interpolation parameters are valid for all cases. Therefore, the result depends strongly on the geographical location of the measurement sites, that shouldn’t be influenced by local emission sources [3].

Table 2.1. Classification and assessment of dispersion models

Model Type	Theoretical background	Advantages	Disadvantages	Example
Gaussian models	- Gaussian plume model: analytical solution of the steady-state advection-diffusion equation [13] - Gaussian puff model: analytical solution of the time-varying advection-diffusion equation [14]	- short computing time - easy to handle - input data requirements are low	- theoretical simplifications (homogeneous velocity and turbulence field) - not suitable for hourly values - mainly suitable for homogeneous terrain	Austal 86 P&K 3782 PROKAS
Eulerian grid models (k-models)	- numerical solution of the advection-diffusion using a finite difference technique	- flexibility to process flow and turbulence inhomogeneities over time and space - higher-order chemical transformation considered - variable time scale	- problems treating the advection (numerical diffusion, mass deficits, negative mass densities) - long computing time - input data requirements are high	EURAS FITNAH REM3 MITRAS
Particle models (Lagrange)	- the model tracks point-like particles representing a trace species on their path - the vector of the turbulent velocity is varied for each particle at each time step using a Markov process [8]	- natural phenomena involved in turbulent diffusion are largely reflected - no numerical diffusion - mass conserving - delivers non-negative mass densities - consideration of complex geometry - consideration of large areas - physical and chemical alterations considered - variable time scale	- sampling error associated with the particle count - long computing time - input data requirements are high	LASAT, AUSTAL2000

2.3 Interpolation methods considering additional variables

Beier and Doppelfeld developed a model for the spatial interpolation of air quality data using local weight functions normalised by a radius of influence. Outside a circle determined by the radius of influence of a measurement site, other monitoring sites are not taken into account for the interpolation [17].

Another method is called Cokriging, where the correlation of the concentration value with another variable is considered. Own tests have shown, that the resulting enhancement doesn't justify the increased effort.

Drüeke applied the interpolation method Kriging with external drift to model the ozone concentration. The additional variable here was the altitude that was used to calculate daily ozone averages in an area of 260 km² [18].

2.4 Regression methods

In general, regression methods are particularly suitable for the calculation of complete concentrations fields. The functional correlation of the pollutant with the most important influencing factors can be quantified in a linear or non-linear regression formula [46]. Prerequisite is, that all independent variables are available for the whole study area.

Thoma used multiple, linear regressions to formulate statistical models to regionalize NO-, NO_2- und O_3-measurements. The independent variables were the height above sea level, the number of days with minimum-temperature-inversions, the ventilation situation, the wind velocity, land usage, the number of days with heat stress and others. The calculations were carried out in a 250 m-raster, for summer and winter. For NO and NO_2, he found out correlations with the height above sea level and the total pollution value, resulting in a regression model with an r^2 of 0,3 to 0,5 [19].

Regression models make high demands to the data collective – statistical preconditions like the normal distribution have to be met and there is a high sensitivity to outliers. Figure 2.1 shows the steps to configure a regression model.

2.5 Applied methods in EU countries

Now follows a brief summary of methods for the calculation of complete concentrations fields officially applied in EU countries to fulfil the recommendation of the EU framework directive 96/62/EC [2, 15].

2.5.1 Belgium

The spatial extend of air pollutant concentrations are interpolated using an "Inverse Distance to a Power" algorithm in which a weighting factor inversely proportional to the distance from a monitoring station is applied. The results are calculated for 5x5 km squares of a grid that covers the complete Belgium territory.

2.5.2 United Kingdom

Rural concentration maps are produced by the interpolation of air pollution data from monitoring sites. At local scale, a box model is applied using relationships between measured concentrations and the values of local pollution statistics. Again the results (background

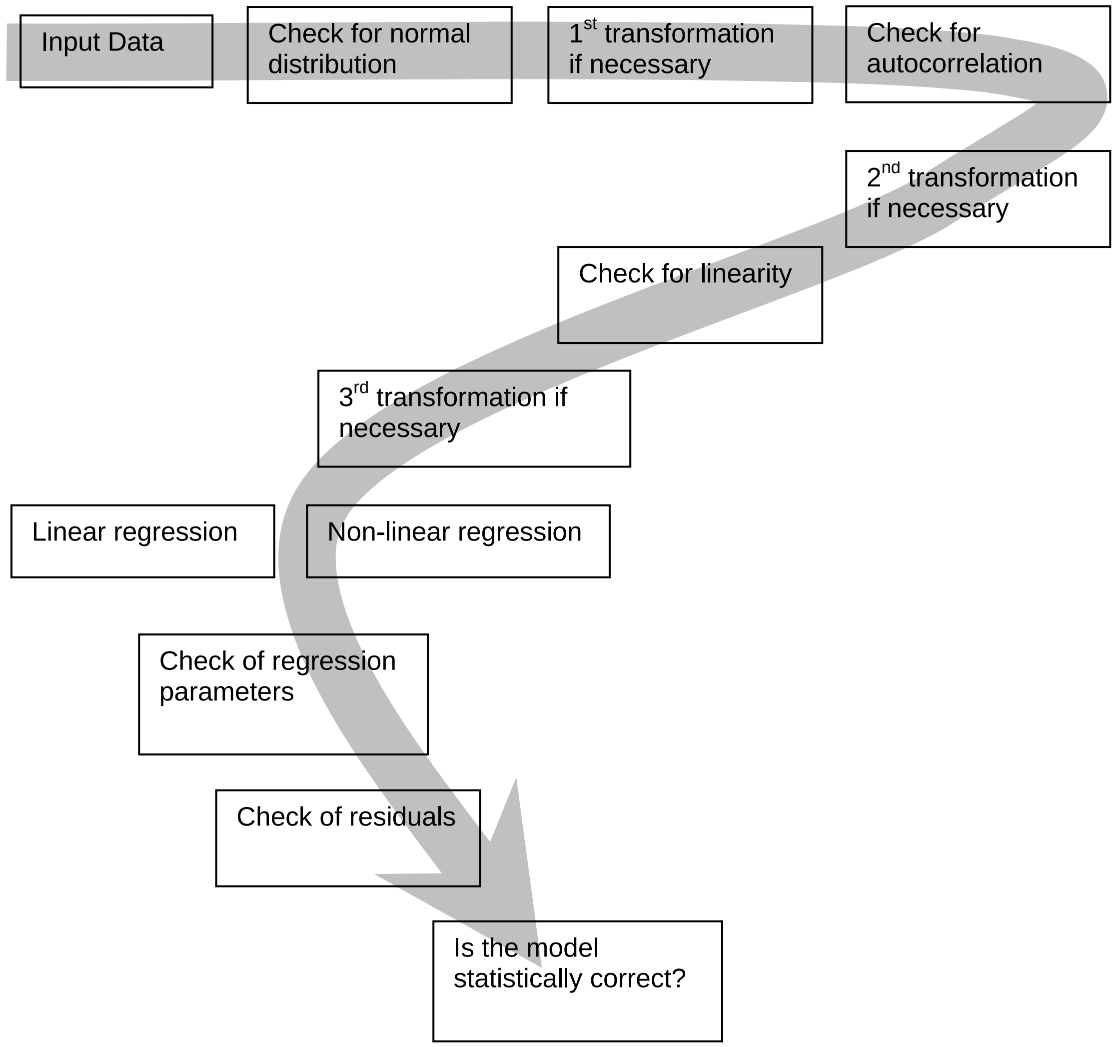


Figure 2.1. Steps to create a regression model

concentration levels) are calculated for a grid, in this case with a resolution of 1x1 km.

2.5.3 Germany

FLADIS is a modelling system that includes different modelling approaches, which are explained below. With Fladis a grided concentration field can be calculated on a time base that depends on the available input data from continuous measurements.

Modelling Approach 1: Interpolation
For every time step, an interpolation between all available continuous measurement sites is calculated.

Modelling Approach 2: Regression Model
A statistical regression analysis quantifies the dependence between elevation values and the pollutant concentration at the grid cell, where the measurement station is located. Since elevation values exist in every grid cell, the obtained dependency can be applied for cells where no measurement data are available. This method delivers especially for Ozone good values and also for NO_x at high concentration levels.

Modelling Approach 3: Gaussian Dispersion Model

First, FLADIS merges different emissions inventories into a summarized uniform grid. The emission inventories may include point, line and area sources. Second, a cluster analysis helps reducing the summarized emissions inventory. The reduction is necessary to ensure the calculation in an acceptable time. Third, a gaussian dispersion model is applied with which the dispersion of the emissions from the clustered sources is calculated. The input meteorological data are derived from all available measurements by averaging the wind vectors (including speed and direction). Also the topography is considered through a digital elevation model. For every method a coefficient of determination is calculated for every time step. Finally all results are combined according to this coefficient – the output pollution map is a weighted combination of all modelling approaches implemented.

2.5.4 Netherlands

Air quality data of the Dutch National Monitoring Network are used to estimate the total length of city roads where limits are exceeded. The method includes the dispersion model "CAR", a database with input information on roads and a statistical model to extrapolate the results for the cities in the database to all cities in the country.

With a long range Lagrangian transport/deposition model, territory covering maps of PM_{10} concentrations with a resolution of 5 km are produced. In a second step, the difference between measured concentrations from the existing monitoring network and the results from the dispersion model are calculated and then interpolated on a 5 km grid. Adding the difference map to the modelled map generates the final concentration field map.

2.5.5 Guidance report on preliminary assessment under EC air quality directives [3]

According to the above mentioned guidance report, the production of concentration maps is the central output of a preliminary assessment of ambient air quality under the framework directive 96/62/EC. On these maps, all areas of exceedances or near-exceedances should be clearly visible. In chapter 3.4 of the guidance, the use of the diffusive sampling technique is recommended to determine the pollutant distribution over a large area. Clear, stepwise instructions are given how to carry out the task of diffusive sampling. In this context step 8 is of special interest: "Calculate the distribution of the pollution levels by interpolation of the measurement made in each grid cell."

2.5.6 Guidance on Assessment under the EU Air Quality Directives [15]

The report recommends the interpolation of measurement results to obtain maps of the pollutant concentrations. According to the guidance these maps "can be used for the mapping of air pollutants over an area in particular for the following applications: Assessment of areas exceeding the limit value and of the population exposed; support for the definition of zones; classification of a territory in areas of homogeneous air quality; design and optimisation of monitoring networks and the control of the effectiveness of abatement measures".

2.5.7 Overview of Methods and Results of the Preliminary Assessment of Air Quality in Europe under Directives 96/62/EC [21]

The European commission has prepared a report where methods and results are summarized. Here it is stated, “the primary method for the preliminary assessment used in the countries was the analysis of the results of the existing monitoring network. Most respondents used in addition mathematical methods ranging from interpolation to computer modelling”.

2.6 Neural Networks

Neural networks in the field of air quality modelling have already been applied by some scientists and have shown promising results. Their functionality is described in chapter 5.2, but briefly they can be characterised as a flexible method to recognise complex patterns in a multivariate data environment. They fit best in the category of statistical methods.

M. Gerboles from the “Institute for Environment and Sustainability”, Ispra (Italy) has used a multilayer perceptron (MLP) to calculate the distribution of NO_2 in Bologna, Italy. His calculations were based on 15 continuous monitoring stations and 100 diffusive sampling sites, distributed over an area of 120 km². The input variables were: NO/NO_2 data from the monitoring stations mentioned above, the population density per km², the rectangular coordinates and the polar coordinates related to the position of each monitoring station. As output variables served the diffusive sampling measurement results – 800 cases for training and 400 for testing. Compared to pollutant maps obtained by Kriging interpolation of the diffusive sampling results, the NO_2 distribution calculated by the neural network shows the same patterns with some lack of peak areas [22].

A. Pellicioni et al. have applied a neural filter to the calculations of a virtual height dispersion model. In particular, the predicted concentrations levels were filtered by a multilayer perceptron to account for the systematic influence of important variables related with atmospheric processes. A comparison between the performances of the dispersion model alone and those of the coupled model showed a significant improvement when the neural filter was applied [23, 24].

The goal of another work was to forecast O_3 and NO_2 levels with a linear regression model and a multilayer perceptron. A comparison of the results from both methods showed, that the neural network was able to give better predictions [25].

H. Omasreiter realised pollutant forecasts with inputs from continuous monitoring stations in the area around Stuttgart, Germany. With the help of multilayer perceptrons and Elman networks, he established correlations between CO concentrations and other variables like the wind speed and traffic data. Furthermore he developed the theory of neural air pollutants modelling and included the already existing concept of virtual neural sensors [26].

L. Partanen et al. used multilayer perceptrons to calculate punctiform NO_2 and PM_{10} levels using meteorological, traffic and air quality data. A conclusion of this work was, that dispersion models are better applicable for the prediction of the spatial distribution of air pollutants than neural networks. [27]

Future (24 h later) daily ground level SO_2 concentrations in Istanbul were modelled by A. Saral and F. Ertürk. A multilayer perceptron was applied to correlate SO_2 concentrations with a detailed set of meteorological parameters including wind speed, wind direction, pressure, temperature, cloudiness, relative humidity, dominant wind direction, solar radiation and data on the mixing layer height. Some problems occurred here predicting peak values [28].

F. Liguori correlated CO, SO_2 and O_3 measurements with traffic and meteorological data using a neural network. This work aimed to assess air quality and to implement models for the monitoring of severe pollution episodes [29].

As an intermediate summary it can be stated that neural networks gain more and more importance in the field air quality modelling, especially for the calculation of pollutant concentrations at single places. Distribution calculations are rarely realized to date.

2.7 Discussion of the different methods – conclusions for own research

In the previous chapters it has been shown, that a great part of all available modelling methods do not include measured pollutant concentrations. They rely on theoretically and experimentally explored correlations between the pollutant and other variables like wind data or turbulence parameters. Here, measured concentrations serve only for validation purposes. The only approaches that base on actual measurements of the modelled species are statistical models, interpolation methods and neural networks. With these methods a direct link to the measured air pollutants is assured. So which of these three methods fulfils all the requirements laid down in chapter 1.2? Interpolation delivers results that look nice but neglects important additional parameters like sources and sinks between the measurement points, a problem, which only can be solved partly by quasi-multivariate interpolation methods. Considering Cokriging e.g., the additional variables only enhance the variogram and don’t deliver real correlations [16]. The end result is still mainly determined by the selected sampling sites – hot spots may be missed.

The regression analysis is a method that allows correlating different variables. With a good database, it reflects the real situation quite well, since it quantifies the functional correlation of the pollutant concentration with the major influence variables. However, as depicted in Figure 2.2, it makes high demands on the data quality in terms of basic theoretical assumptions. For multiple regression analysis, even a multivariate normal distribution has to exist [30, 31, 32]. To meet these preconditions, extensive data transformations and tests have to be carried out [33]. The result will be a rigid formula; the functional correlation of the variables might be blurred due to its complexity and non-linearity.

Finally, neural modelling is a flexible method, which enables one to recognise highly complex non-linear correlations. Statistical assumptions like normal distribution are not necessary, which makes them easy to handle in principle. The network can be trained with real measurement data and updated with new measurements, enhancing its quality and making it the ideal method for the purpose of this work. Since this is a relatively new method in the field of air quality modelling, basic research work still has to be done. This work here is a contribution to this.

In Figure 2.3 the possible methods and their capabilities regarding the demanded characteristics to calculate the spatial distribution of air pollutants are summarized.

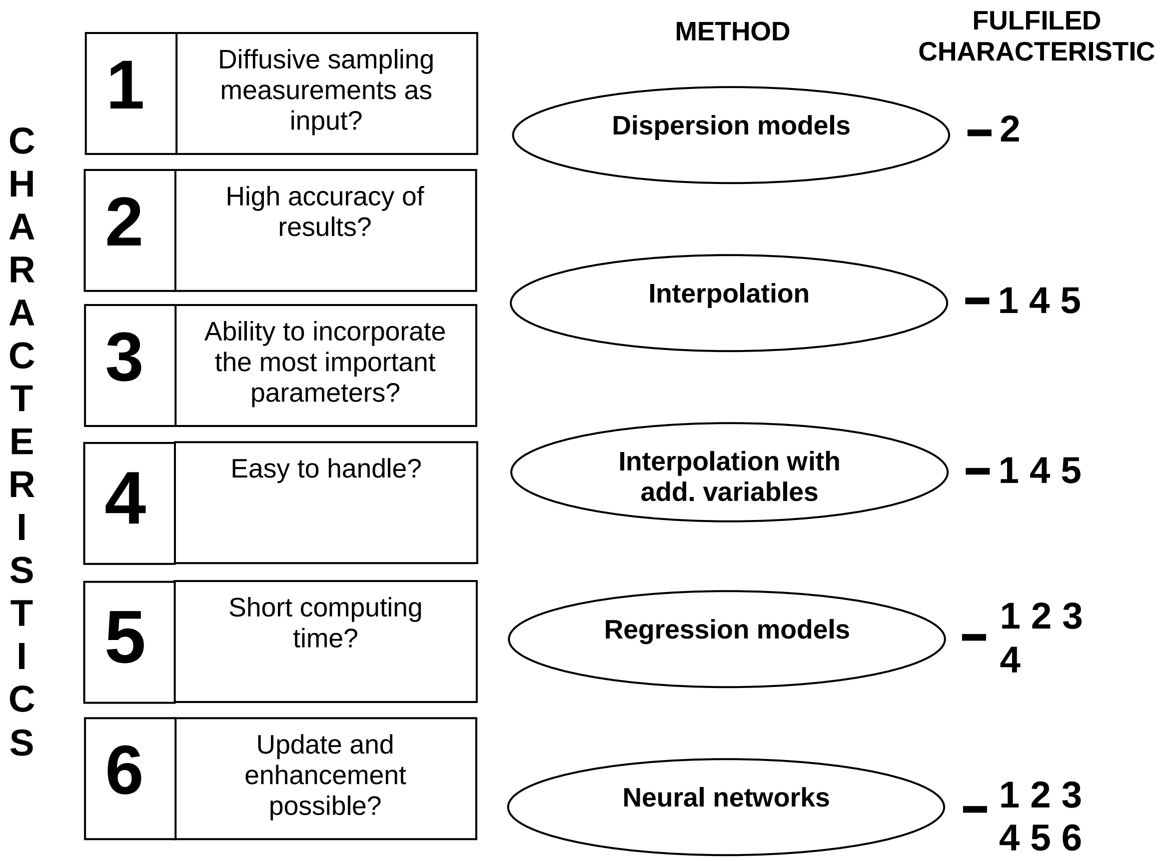


Figure 2.2. Demanded characteristics for a method to simulate the spatial distribution of air pollutants

Neural networks are the only tools that fulfil all characteristics as shown in chapter 1.2 and summarized in Figure 2.2.

3 Local conditions in Cyprus

3.1 Description of the study area Cyprus

The now following description of the study area Cyprus refers mainly to it's characteristics influencing the spatial distribution of air pollutants. As mentioned above, the model was developed for the UNOPS project "Preliminary Assessment of Air Quality in Cyprus".

3.1.1 General considerations

Cyprus is located in the east Mediterranean Sea, south of the Turkish coast, crossing the 35th longitude and the 33rd latitude. Its diverse geomorphology includes coastal areas, mountains of up to 2000 meters height (Troodos mountains, Kyrenia Range), a narrow promontory (Karpezia) and a wide plain (Messaoria). In Figure 3.1 the major topographic properties of Cyprus in a three dimensional view are shown.

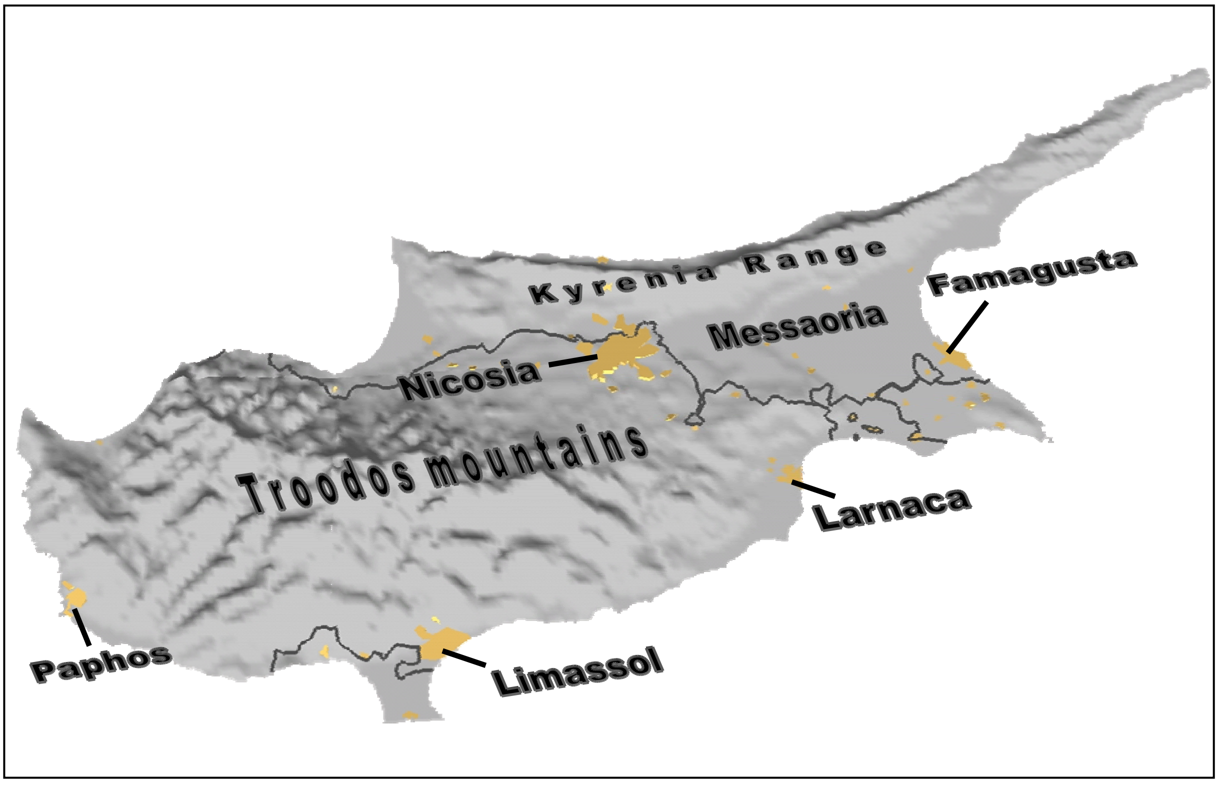


Figure 3.1. Three-dimensional view of Cyprus with the major topographic properties and cities – the raising factor is ten

The population concentrates in the major cities being Nicosia, Limassol, Larnaka, Paphos, Famagusta and Kyrenia. They are all located at the coast except Nicosia, which is situated in the centre of Cyprus on the plane of Messaoria. During summer time, the population grows strongly, because of the tourism, which is the most important economic factor.

Since 1974, Cyprus is divided in two parts by the so called “green line”, which runs from East to West in the plane of Messaoria, North of the Troodos mountains passing the old city centre of Nicosia. South of this line the Greek Cypriot Community (GCC) is living, northerly the Turkish Cypriot Community (TCC).

3.1.2 Meteorological conditions

Besides the emission sources which are of course the base for the existence of any pollutant in the air, the meteorological conditions are the most important factor to influence concentration levels. Wind, for example, dilutes the gaseous pollutants, the higher the wind speed the better the grade of dilution. Inversion layers on the other hand, force their accumulation. The wind direction determines the dispersion direction from the source. The interaction of the different influences is very complex and can be reviewed in [33]. A detailed description of the general and local meteorological conditions of Cyprus is given in the chapters below.

3.1.1.1 General meteorological considerations

The mediterranean climate of Cyprus means constantly hot, dry summers from mid May to mid September and mild, rainy winters from November to mid March. The transitional seasons spring and autumn are characterized by diverse meteorological conditions.

This can be explained by the pressure system around Cyprus [34]:

- **Summer:** Cyprus is influenced by a continental depression over Southwest Asia, which extends to Cyprus in the form of a stable trough. Monotonous weather conditions with a maximum temperature of 40 °C and more are the result.
- **Winter:** This monotony ends in winter due to small depressions crossing the island from West to East between the anticyclone of Eurasia and the low-pressure belt of North Africa.
- **Spring:** The Siberian anticyclone collapses causing a continental depression over Southwest Asia, which extends westwards. A similar depression appears over Sahara.
- **Autumn:** The Siberian anticyclone starts to develop and the continental depression cantered over the southwest Asia starts growing.

Table 3.1 summarizes synoptic features influencing Cyprus and averages of temperature and humidity.

3.1.1.2 Local scale meteorology

The main synoptic features described in chapter 3.1.1.1 are modified by local processes, being:

- **Land-Sea breeze:** A phenomenon often observed at the coast. During daytime, the air flows from the sea to the land (sea breeze), during nighttime from the land to the sea (land breeze). The motor behind this lies in the different thermal properties of landmasses and water. Land warms up and cools quickly, while water needs more time to warm up, but is able to store this heat longer. This causes local pressure systems driving the winds.

- **Mountain-Valley breeze:** During daytime warm air from the valley rises along the mountain slope (valley breeze), during nighttime the cool “heavy” air slides down the slopes (mountain breeze).

Table 3.1. Synoptic features influencing Cyprus averaged over 10 years [34]

Parameters	**Winter**			**^Summer**			**Spring**		**Autumn**	
Synoptic features	Small depression crosses Cyprus from west to east			Trough over Cyprus extending from depression of southwest Asia			Influence of depression over southwest Asia		Influence of Siberian anticyclone	
Temperature (°C) Nicosia	Mean Max January	Mean Min January	Lowest Min January	Mean Max July	Mean Min July	Highest Max July	Mean Max April	Mean Max April	Mean Max October	Mean Max October
	15	5	- 3	37	21	43	24	10	28	15
Relative Humidity (%) Nicosia	0800 LST* January	1400 LST January		0800 LST July	1400 LST July		0800 LST April	1400 LST April	0800 LST October	1400 LST October
	78	52		52	31		63	40	68	44

*Local sidereal time

- **Inversion layers:** The vertical dispersion of pollutants is limited by atmospheric barrier layers. The so-called nocturnal surface inversion can lead to an enrichment of pollutants in the lowest layer where people are living. In the morning hours this layer is dissolved caused by warm air at the ground, which is heated by the sun radiation. An elevated inversion layer is limiting the mixing height for the pollutants over Cyprus. In summer this mixing height reaches up to 4000 m above ground whereas in fall and wintertime it stays below 2000 m. The view from Mount Olympus during nice weather in fall time (November 2002) illustrates the vertical extent of the mixing layer with its brown colour containing the pollutants emitted from ground level (Figure 3.2). Above this sharp barrier layer very clean air can be observed. During aircraft ascents and descents this barrier layer separating the polluted mixing layer and the clear free atmosphere above can be seen too very clearly. The warm climate and strong solar radiation in Cyprus induces convective movements and creates an average mixing layer height of at least 1700 meters, where the first inversion layer is located (Figure 3.3). In some cases, the second inversion layer is linked with the first, but usually it can be found several hundred meters above and may serve as the decisive barrier for mixing processes. This allows a good dispersion and dilution of pollutants. In winter however, lifted ground inversions can be observed sometimes. As expected, the general inversion layer height in summer increases and decreases in winter (Figure 3.3) [34, 35, 64].

- **Topography:** As described in chapter 3.1.1, Cyprus has a quite diverse topography, which directs the winds physically (Figure 3.4) and causes special meteorological features like the mountain-valley breeze. So the wind system in Cyprus is also very diverse with distinct local differences in a comparable small area.

Figure 3.2. Picture taken at the Troodos Mountains showing a brown mixing layer, were all the pollutants are released and a clear, unpolluted layer above which is separated by the inversion layer

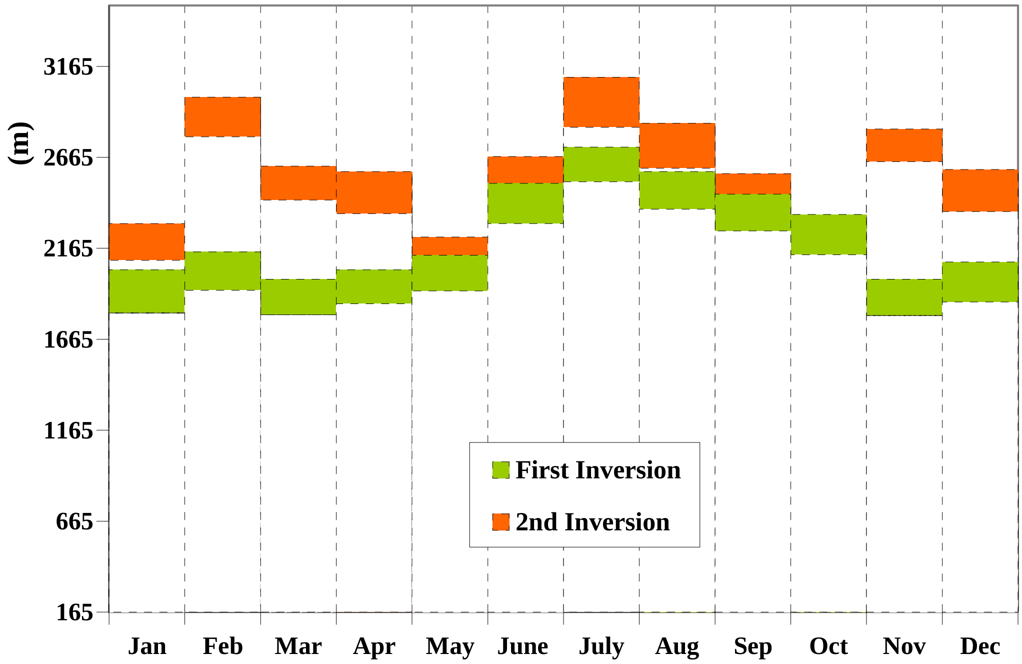


Figure 3.3. Mean annual profile of the inversions over Nicosia at 11 UTC as determined by an eight years statistical study of inversions by Meteorological Service Cyprus 1991 [34]

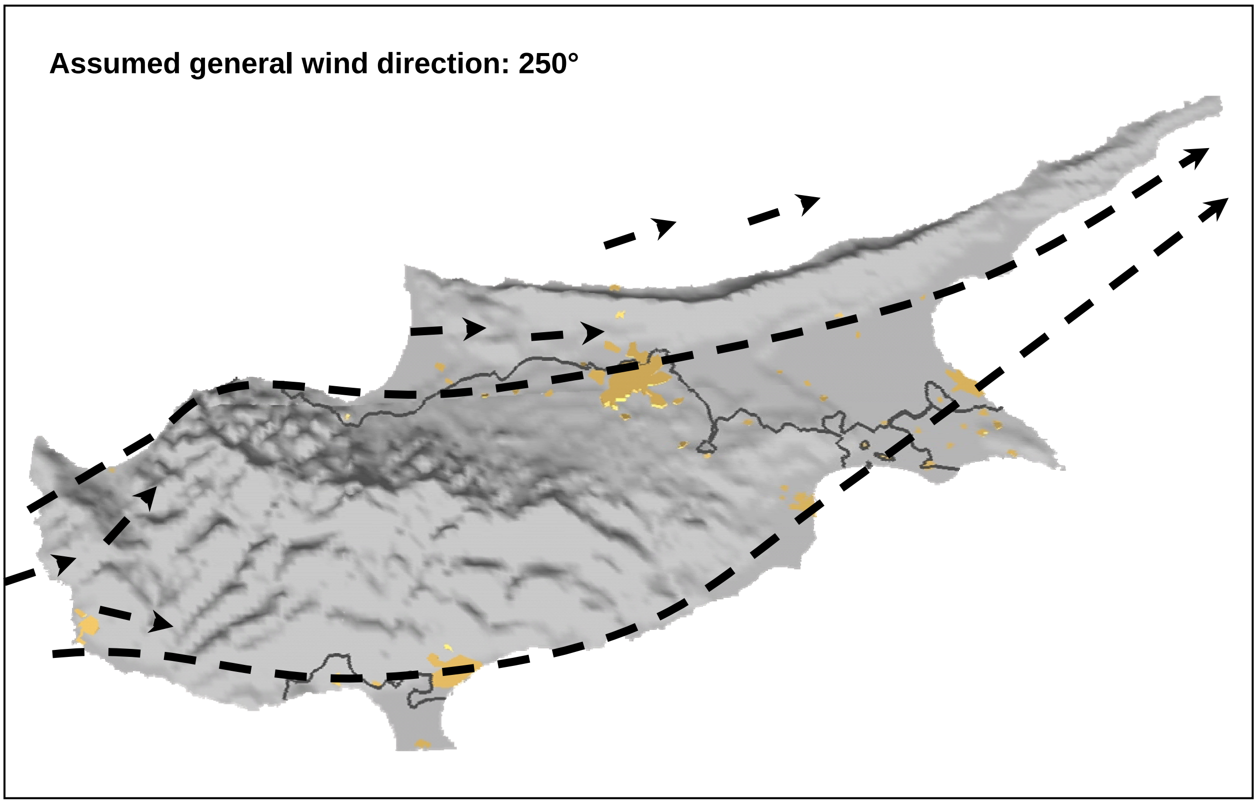


Figure 3.4. Theoretical wind field over Cyprus assuming a general wind direction of 250° - after [56]

3.2 Air quality measurements

Within the project "Preliminary Assessment of Ambient Air Quality in Cyprus" an extensive measurement programme and surveys were carried out (see chapter 1.1). Herewith, a good database for the model development was created. A more detailed description of the tasks that are of special interest for this work is given below.

3.2.1 Diffusive sampling

At 270 sites NO_2 diffusive samplers were exposed (see Annex A.1.) in six campaigns over the course of one year (July 2002 to July 2003). For the pollutants Ozone, VOC and SO_2 the site number was reduced since their distribution was expected to be less structured. The samplers were primary placed in the major cities, selecting categorized sampling sites to cover all major emission sources and characteristic land usage types. For the development of the neural net an important issue – it's generalizing ability increases the more categories are considered. Such a categorization is also demanded by the EU, but was refined for project purposes. In Table 3.2 the average annual values from all measured components at all site categories are shown, reproducing very well what can be expected: The highest values occur in the centre of the cities in commercial streets and traffic sites, the lowest in mountainous rural areas. Bold numbers indicate exceedances of the actual upper assessment threshold or of future EU limit values. For the secondary pollutant Ozone other formation mechanisms apply [33], which is why inverse observations can be established here. To verify the results of the diffusive sampling campaigns, a comparison of these results with continuous measurements at the same site was carried out. In Figure 3.5 this comparison is depicted and it shows a very good agreement between the two methods. Up to four diffusive samplers were exposed at the same time to assure the comparability of the results. No major deviations were observed, although the high temperatures in Cyprus might influence the relative uptake rate [36, 37]. So these values can be considered as reliable and therefore suitable for model development.

Table 3.2. Average diffusive sampling results for NO_2, divided into site categories [34]

Site Category	NO_2 average in µg/m³	Benzene average in µg/m³	SO_2 average in µg/m³	Ozone average in µg/m³
commercial (Municipality Market, Larnaka + Armenias Street + Ezekia Papaioannou Str., Nicosia)	**48,7**	**8,4**	16,1	-
urban background	**39,7**	**7,3**	11,4	60,9
traffic	**38,9**	**6,7**	13,2	
recreation	32,9	-	-	-
residential	23,2	2,8	7,5	74,4
industrial	22,7	3,5	6,6	92,7
touristic beaches	19,9	1,0	9,2	77,6
peripheral	16,8	1,7	4,2	-
airport	15,0	1,3	5,1	-
village>700	14,0	1,7	6,9	81,0
touristic	11,9	2,2	8,5	
sensitive area (Akrotiri – Salt Lake)	10,7		-	-
village<700	8,1	1,2	4,8	78,8
agricultural	7,0	1,6	-	73,4
mountainous, forests	2,6	0,5	3,2	95,5
mountainous, no forests	2,0	1,1	2,2	102,6

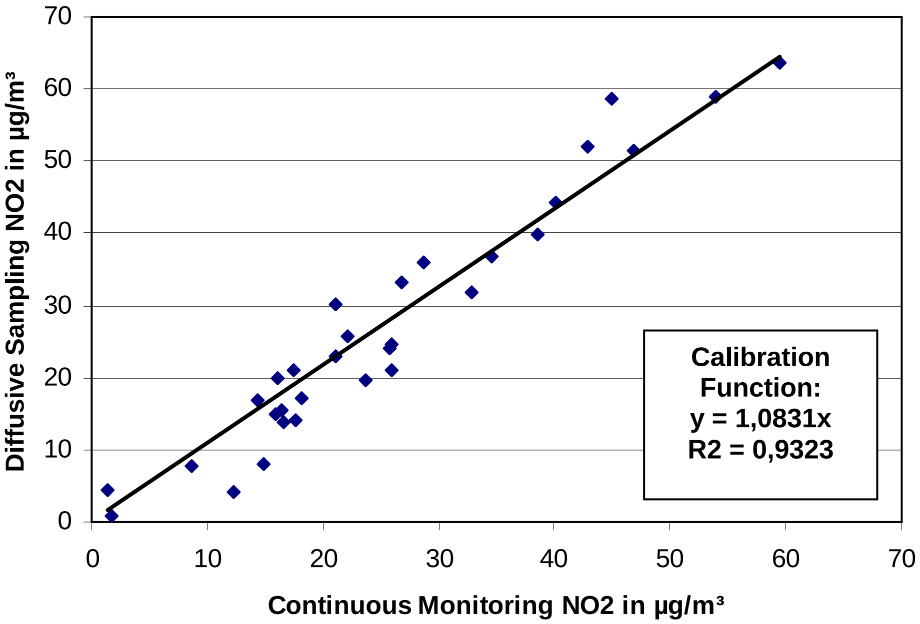


Figure 3.5. Quality assurance of NO_2 diffusive sampling - comparison of NO_2 diffusive sampling results and NO_2 continuous monitoring. The latter are considered to be correct.

Finally in Figure 3.6, the results of all NO_2 diffusive sampling campaigns are summarized, showing the annual NO_2 distribution over Cyprus. As expected, the highest concentrations are measured in the big cities. Hot spots are to be found in Nicosia and Limassol. Large rural areas like the Troodos mountains and most of the area of the Turkish Cypriot Community are more or less unaffected by NO_2 due to a lack of emission sources and good dispersion conditions. Alone with these results, a very good insight into the NO_2 distribution is already given.

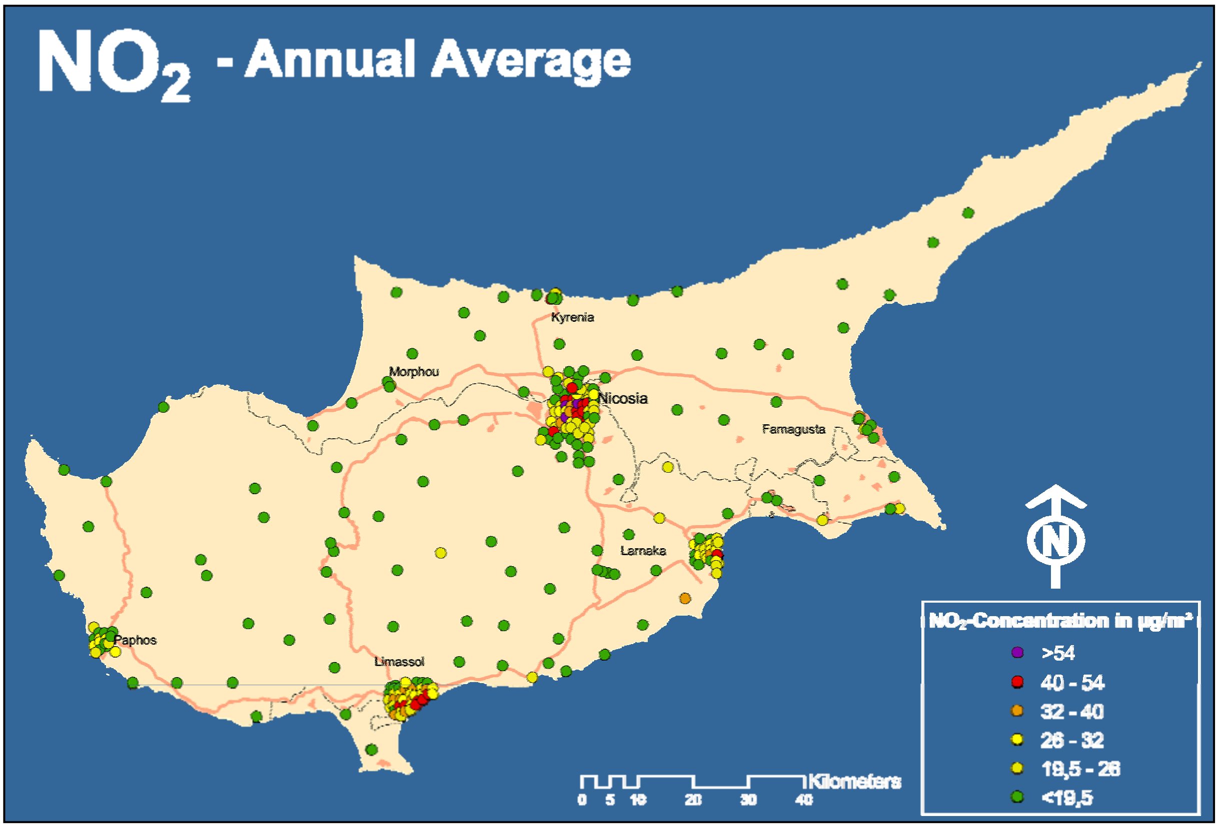


Figure 3.6. Mean annual diffusive sampling results at 270 sampling sites over Cyprus (summer 2002 to summer 2003)

3.2.2 Continuous monitoring

To observe the temporal variation of the pollutants and meteorological parameters and to capture worst cases of maximum concentrations, continuous monitoring stations were placed all over Cyprus. All stations were equipped with air quality monitoring instruments for recording the concentrations of ozone (O_3) and nitrogen oxides (NO, NO_2, NO_x). Further, instruments for measuring the concentrations of carbon monoxide (CO) and sulphur dioxide (SO_2) were installed in the multi-component stations.

Mini stations served for investigating and monitoring the distribution of ambient ozone in rural background areas. To exclude external influences on the ozone measurements, the mini stations were placed at sites where no emission sources e.g. road traffic, power plants or residential areas had been close or present at all. By parallel measurement of nitrogen oxides at each mini station, influences by emissions from local sources on the ozone concentration thus could be identified.

The multi-component stations were placed in urban areas with medium to high emission loads at traffic or residential sites. Exemplary results can be viewed in Annex A.1. As stated above, the results were in good agreement with the diffusive samplers.

Compared to other European cities, the air quality situation in Cyprus can be regarded as problematic as shown for the case of NO_2 in Annex A.1. In the countryside, the air quality is much better due to the good ventilation and sparse emission sources.

3.2.3 Other measurements

During the project additional measurements were necessary to complete the preliminary assessment in agreement with the European Framework Directive. For the model development these measurements couldn’t be implemented directly, but served as valuable background information. In detail, these measurements are: Tethered balloon measurements, particulate matter measurements and wind measurements.

The objectives of tethered balloon measurements were to determine the meteorological conditions, which are influencing the pollutants distribution within the cities of Cyprus, as an example in Nicosia and Limassol. This more specifically addresses the diurnal and nocturnal dynamics of mixing heights and the wind systems like the land-sea breeze, which are responsible for cleaning urban air and transporting pollutants to the rural areas. In combination with the ground level measurements, a description and explanation of the air quality situation of the selected cities Nicosia and Limassol was given as shown in Annex A.1. [34].

The complete tethered balloon measurement system was developed at the Institute for Process Engineering and Power Plant Technology at the University of Stuttgart for continuous vertical soundings of different meteorological parameters and air pollutants (NO_2 and O_3). VOC and particulate matter are measured discontinuously by collecting samples in three different height levels. The balloon can be elevated with a special rope up to a height of 1000 m above ground level.

Earlier in this work, the meteorological conditions in Cyprus according to past statistics are described. This allows a good insight in the general situation in Cyprus and can be considered as valid today. But to assess and quantify the influence of meteorological parameters on pollutant concentrations, the measurements have to be carried out at the same time. That’s why wind measurements were carried out parallel to continuous pollutant measurements.

Finally, particulate matter was measured, namely PM_{10} and $PM_{2,5}$. Due to long-range transport events (Annex A.1.), the usage of diesel fuel and the dryness of the land, particulate matter is a problematic pollutant in Cyprus. The measurement devices were placed at sites to cover different categories. Together with the results from element analyses using X-RFA and AAS devices, a statistical source apportionment was carried out. The principle of this method lies in the fact that most sources can be identified by combinations of certain elements or by single elements. Herewith, the most important emission sources of PM could be identified [34].

3.3 Emissions inventory

The emissions inventory was one of the main phases of the project and was aiming to the collection and processing of appropriate data for the estimation of air pollutants emissions from different sources. This was the first time that a systematic and coherent inventory was performed for the whole island of Cyprus. The methodology, the data collection and the calculated emissions for all sources are depicted in the final report of the project [34].

The air pollution sources being considered are treated as line, point and area sources and cover:

- Emissions due to road traffic
- Emissions due to the use of industrial boilers
- Emissions due to dry cleaners
- Emissions from the hotel industry
- Emissions due to domestic heating, heating in hospitals/other buildings
- Emissions due to agricultural activities
- Emissions due to petrol stations
- Emissions from airports

The air pollutants being considered here are:

- Oxides of nitrogen (NO_x)
- Sulphur dioxide (SO_2)
- Carbon monoxide (CO)
- Volatile Organic Compounds (VOC)
- Particulate Matter (PM)

The calculated emissions were edited for ArcGis, a geoinformation system (GIS), which allows visualizing and analysing data regarding their spatial reference. In Figure 3.7 the daily emissions of boilers in Cyprus are shown. Five violet spots attract attention. They indicate the most important single sources of Cyprus: The power plants of Kyrenia, Dhekelia, Vasilikos (Annex A.1.) and Moni and the cement factory of Vasilikos (Annex A.1.).

The traffic emissions were calculated according to the traffic strength and the relevant emission factors from the European database “CORINAIR” as mass emissions per km road length [34, 63]. The roads are coloured in maps according to their emission load. An example of the traffic sector is presented in Figure 3.8. Here the daily NO_x emissions of the traffic sector are shown for the city of Limassol. Parallel to the coast, a red line indicates high traffic loads in the city centre with commercial and touristic traffic. The thick red line that is directed East-West stems from the highway connecting the cities Larnaka, Limassol and Paphos.

For Nicosia, a new method was developed in a diploma thesis to improve the accuracy of the emissions inventory for the traffic sector (see Annex A.1.). The method is based on multichromatic high-resolution pictures from Quickbird satellite as shown in Annex A.1.. Concretely, the pictures were used to exactly determine the traffic density over the whole city without any time delay and without missing any roads [63].

It was attached importance to prepare the emissions inventory in accordance with the requirements of the neural network model, since the emissions inventory obviously is an important variable to declare the pollutants distribution.

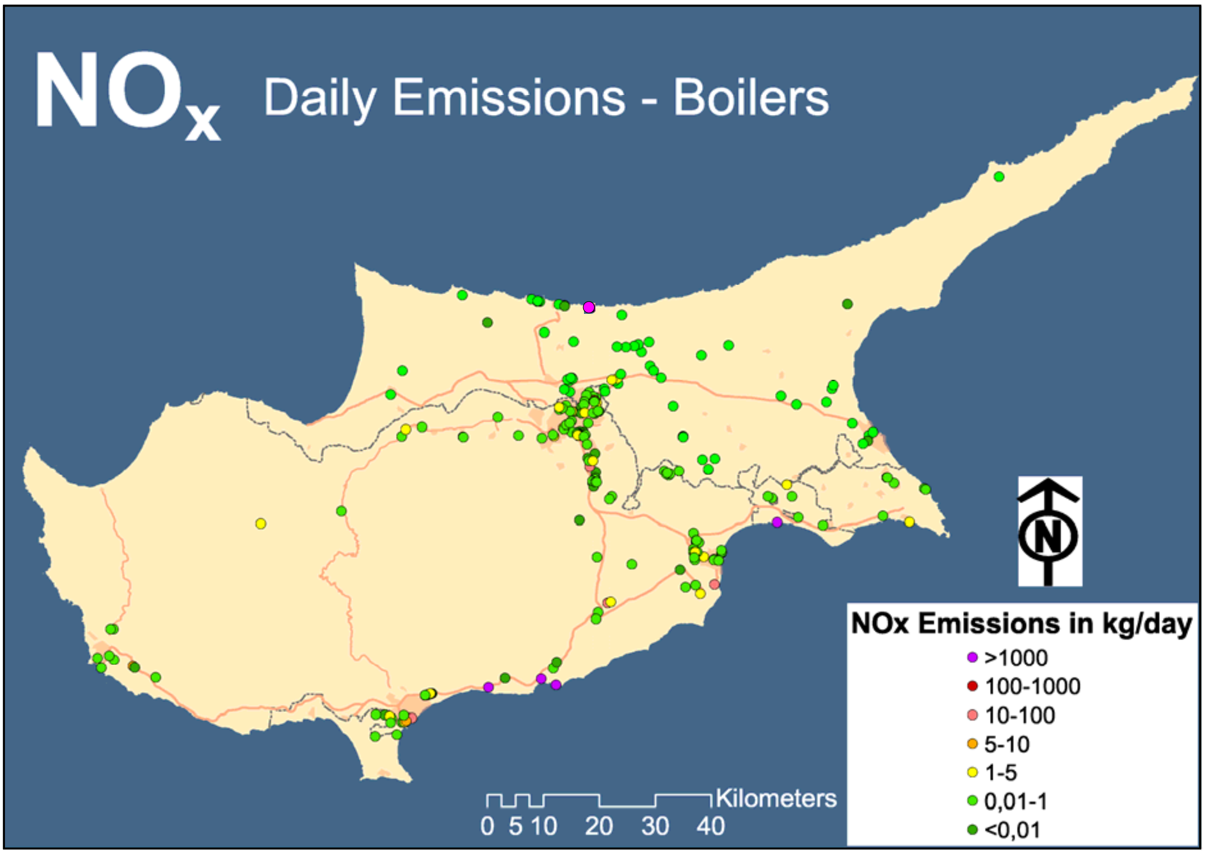


Figure 3.7. Daily NO_x emissions of boilers in Cyprus

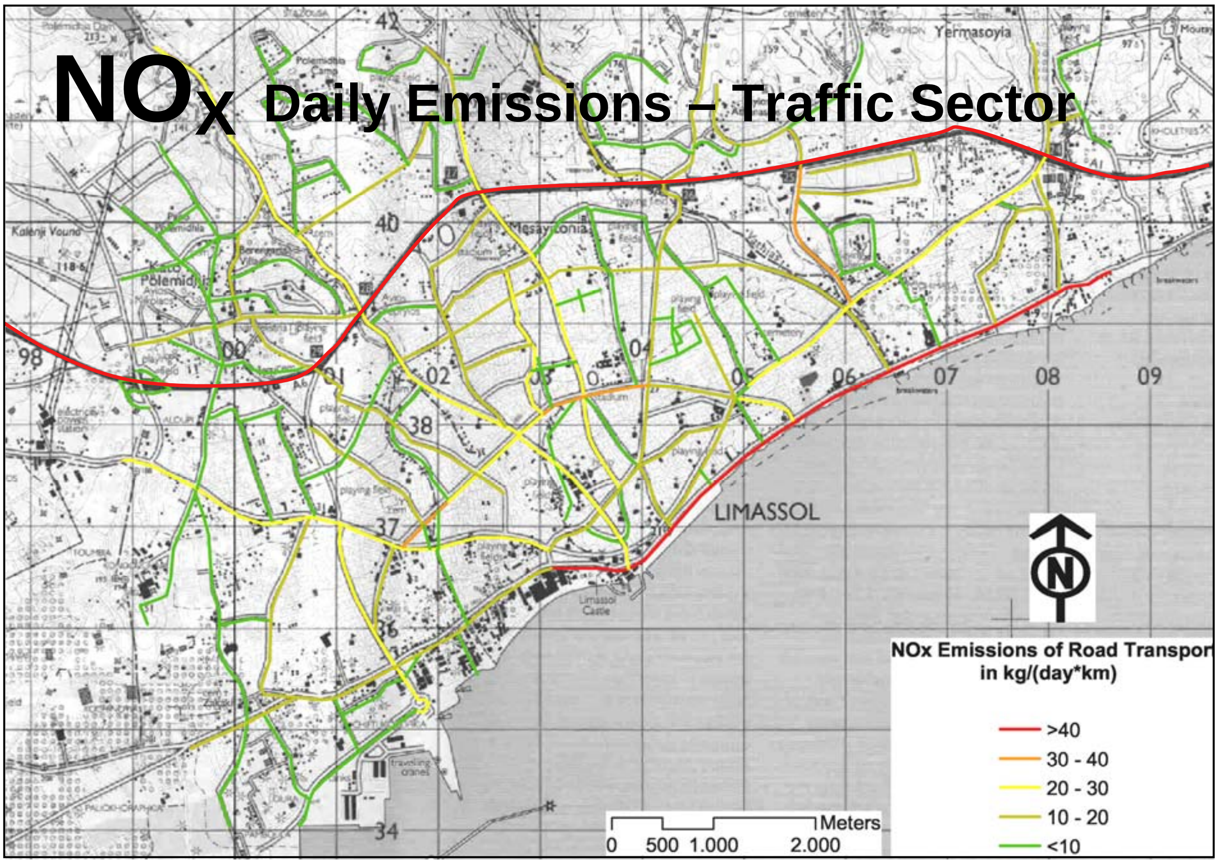


Figure 3.8. Daily NO_x emissions of road transport sector in Limassol

3.4 Other influences on the pollutant concentrations

Chapters 3.1 to 3.3 describe the most important parameters, that influence the spatial distribution of air pollutants. There are however also other influences that can be rather important depending on the regarded pollutant and on the time and spatial scale:

- For Ozone - the height above sea level
- For Volatile Organic Compounds - the vegetation
- For NO_x - forest fires
- For NO_2 - the Ozone concentration (oxidation of NO)
- etc.

Many of these additional variables cannot be quantified or measured – only assumptions can be made, so they are only conditionally usable.

It was finally decided to build the model development on the pollutant NO_2, since this is the only pollutant that was measured at all 270 diffusive sampling sites in Cyprus (see chapter 3.2). So, for NO_2 the most important influence parameters are summarized in Figure 3.9.

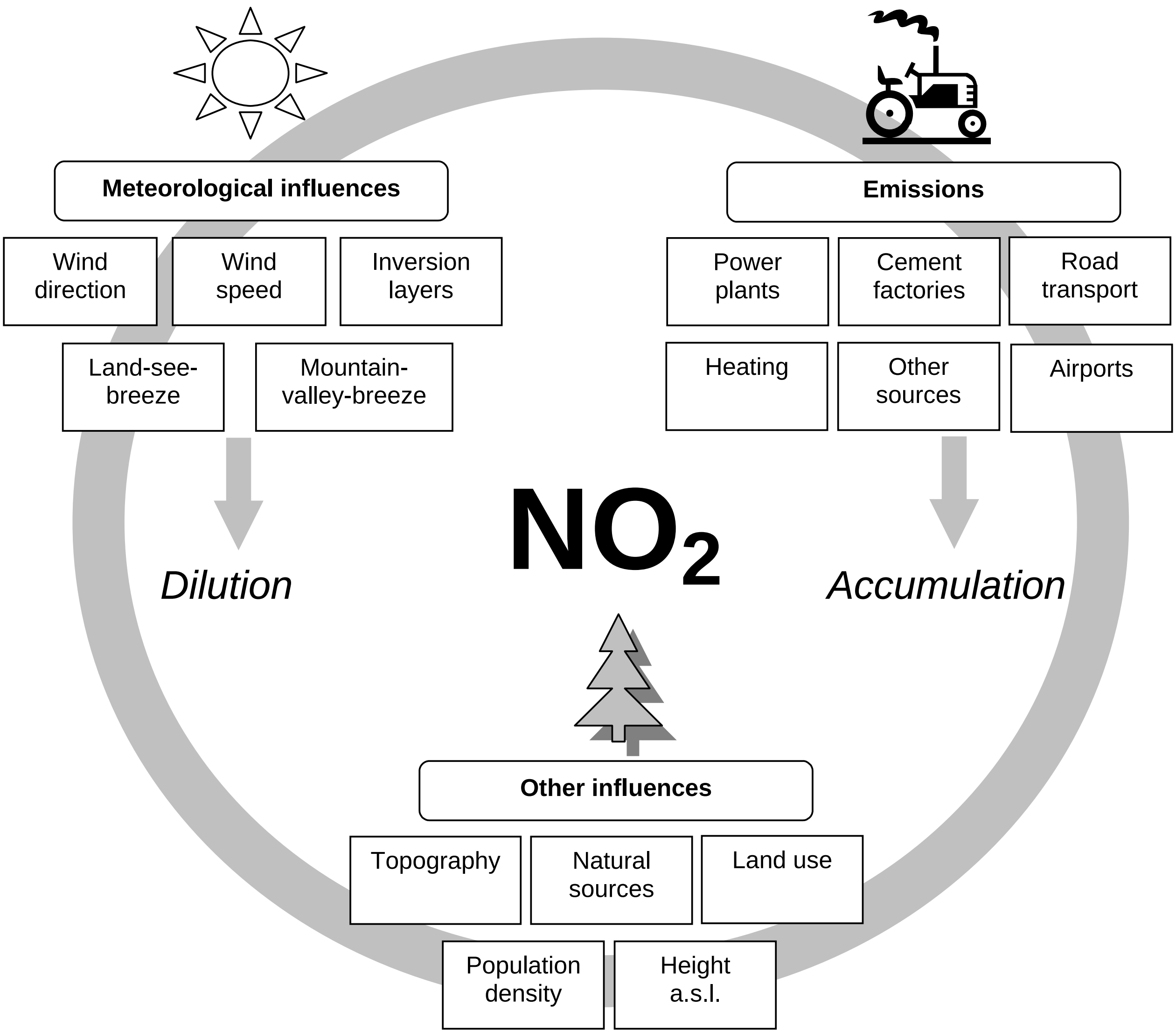


Figure 3.9. Factors that influence the spatial distribution of NO_2

Another important issue on NO_2 formation is the so called photostationary equilibrium between NO, NO_2, O_3 and O_2, expressed in the photostationary equilibrium:

$$NO_2 + O_2 \underset{}{\overset{hv}{\rightleftharpoons}} NO + O_3$$

$$c(O_3) = c(NO_2)/c(NO) * k_2(hv)/k_1$$

These equations show, that the ozone concentration depends on the NO_2/NO ratio and on the effective light intensity. Near a highly frequented road this ratio remains small and only a little amount of ozone is formed. An exception is a process that leads to the photochemical smog as first observed in Los Angeles. Here, intense sunlight forces the formation of OH radicals and together with hydrocarbons the formation of peroxy radicals that finally oxidize NO to NO_2. The oxidization of NO via O_2 is a very slow process and therefore negligible.

Ideally a model should incorporate all possible influence parameters, but in practice this is hardly possible. The developer has to fall back on available parameters from measurements or surveys and digital maps. But first of all, it is important to consider what makes physical sense! In this context, the possible input parameters for the model development are:

- Results of pollutant measurements
- Emissions inventory
- Population density
- Land usage
- Meteorological data, especially wind speed and wind direction
- A digital elevation model

On this database, the development of the neural network was carried out, described in the following chapter.

4 Interpolation maps

4.1 General modelling approach

As stated in chapter 2.7, there are only three modelling methods, that enable the user to process diffusive sampling measurements – regression analysis, interpolation and neural networks. Since interpolation is a widely applied method [15] and delivers fast results, interpolation maps were calculated firstly. The intention was to obtain a first insight on the pollutants distribution and to see how powerful interpolation algorithms are. In Figure 4.1 the general development steps of this work from the measurements to the final neural network model are illustrated.

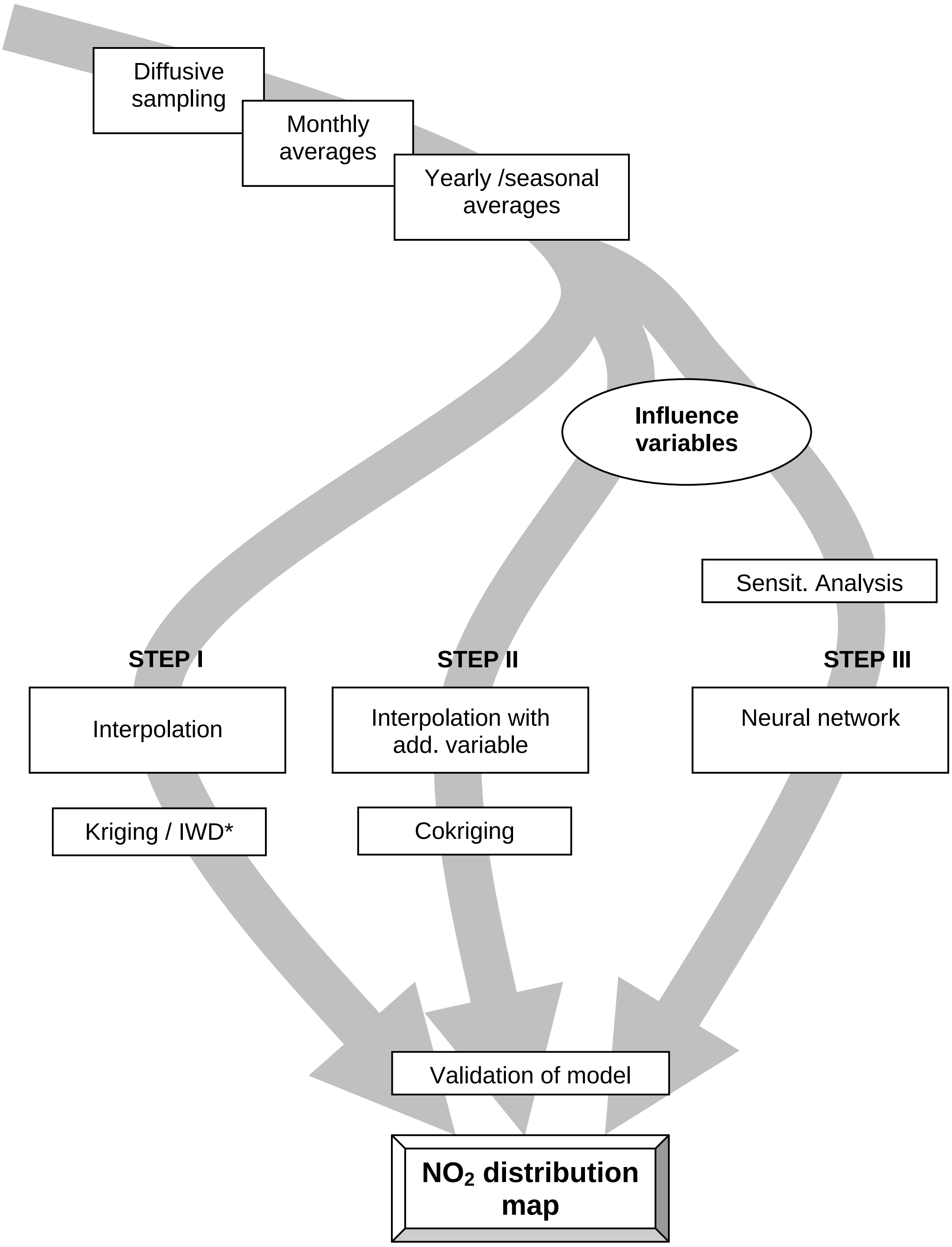


Figure 4.1. Steps carried out to model the spatial distribution of NO_2 in Cyprus

*"Inverse Distance to a Power" interpolation algorithm

Linear and non-linear regression analysis was only used to perform a statistical sensitivity analysis which included graphics like scatter plots to visualize bivariate dependencies or histograms to depict the distribution of a certain parameter.

Earlier it was decided not to use regression analysis as modelling tool due to its limited flexibility and strict theoretical assumptions that cannot be fulfilled practically [30, 31, 33].

4.2 Interpolation

Interpolation maps were calculated for all major cities, for the seasons and for whole Cyprus. Two interpolation results are presented in the following chapter. In Annexes B.1. to B.8 more of these maps are shown.

4.2.1 Simple interpolation

Several interpolation methods were tested using different software tools. The first maps were created with “Surfer”, which is very fast and incorporates the most important interpolation methods. However, another software was chosen, the geoinformation system “ArcGis”. Using the implemented geostatistical analyst, it is not only possible to adapt all necessary interpolation parameters, the great advantage is, that one can directly preview what effect has a modification like changing the lag size of the variogram. More specifically, one can see the percentage of influence of all measurement points at any site and also preview the interpolation result in a rough form. A problem here could be that the end result is very subjective, since one can produce a vast number of quite different maps using the same input data (Annex B.1.). This problem can be partly solved through the cross validation option of the geostatistical analyst and also by setting aside a validation data set, which can be done quite comfortably.

In Figure 4.2 the result of an “Inverse Distance to a Power” interpolation of NO_2 in Limassol is shown. Transition colours were avoided, for not to pretend exactness that doesn’t exist in reality. Peak concentrations can be observed in the city centre of Limassol near the coast with a heavy traffic load. A look at the NO_x emissions inventory of the traffic sector, shown in Figure 3.8, confirms this. There is a gradual circular decrease of NO_2 pollution, the lowest values are predicted north of Limassol - on the map they emerge as two green areas.

This result can be assessed as follows:

- The general trend is reproduced quite well, since the database is rather solid – the interpolation map of Limassol is based on 20 diffusive sampling sites.
- The spatial distribution of NO_2 depends strongly on the geographical location of the sampling sites, which can be concluded from the two green areas, which are two sampling points.
- Interpolation does not care if there are any sources between two interpolation points it simply “over-interpolates” such a source. So it does with the highway, the dark thick line travelling from the west margin of the map to the east.
- Interpolation does not regard any obstacles – it passes buildings.
- The IWD algorithm produces circular shapes, which would mean the wind blowing from all directions with the same speed.

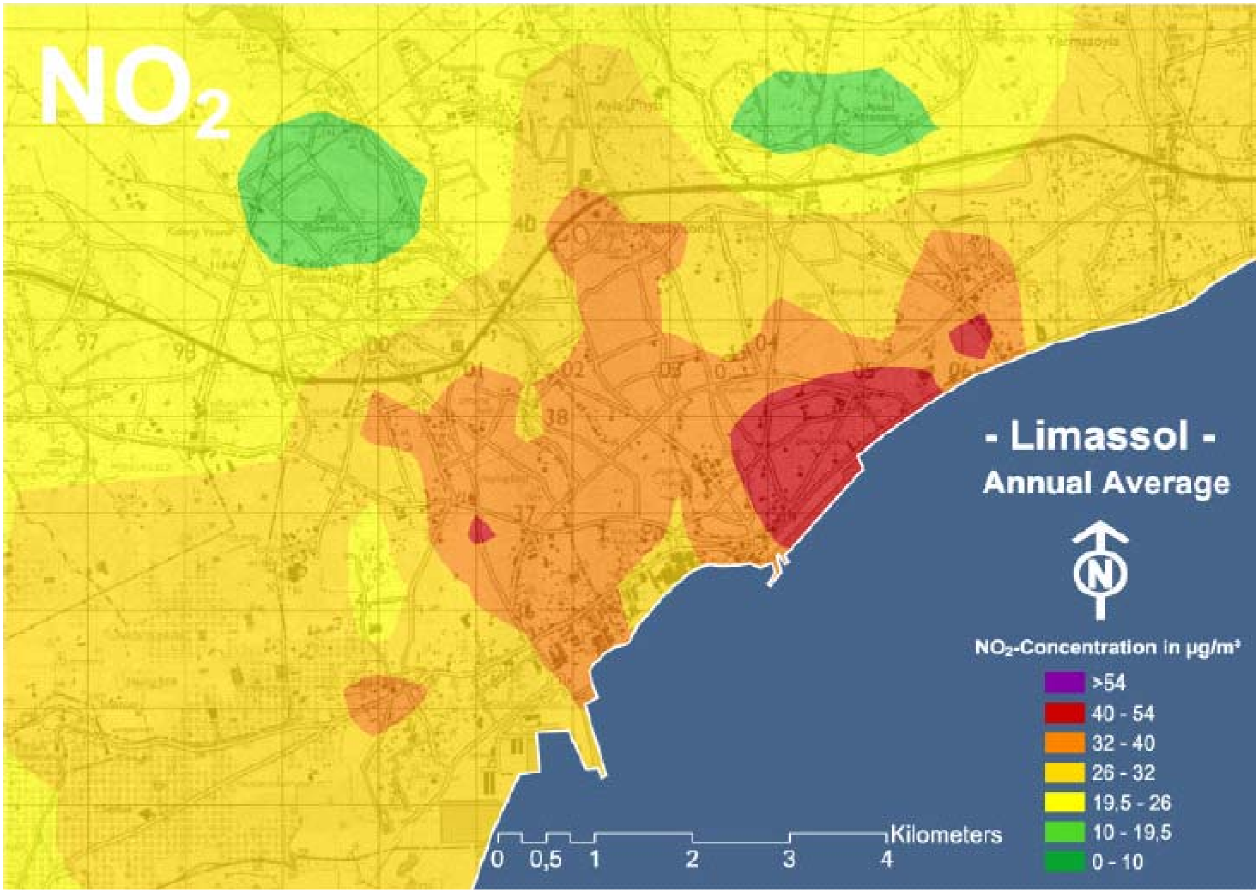


Figure 4.2. Mean annual interpolated NO_2 distribution in Limassol using an “Inverse Distance to a Power” (IWD) interpolation algorithm

In Figure 4.3 the annual distribution of NO_2 all over Cyprus is depicted. With this large scale, the specified disadvantages seem to be blurred, since all major cities emerge quite clearly and large parts of Cyprus, where no emissions are located, are in green. But of course the same problems discussed for the city interpolation do also occur here. Quite in the centre of the map, 25 km north of Limassol, there is a yellow dot with a light green circular surrounding caused by one single diffusive sampling point, which was placed near a highly frequented road. This shows again that the result depends on the location and in addition it clarifies that points, which are influenced by single sources, should not be included in the interpolation. This approach is also being demanded by the European Union in the “Guidance report on preliminary assessment under EC air quality directives” [3]. Here it is recommended to place the diffusive samplers in a way that they are representative for a large area.

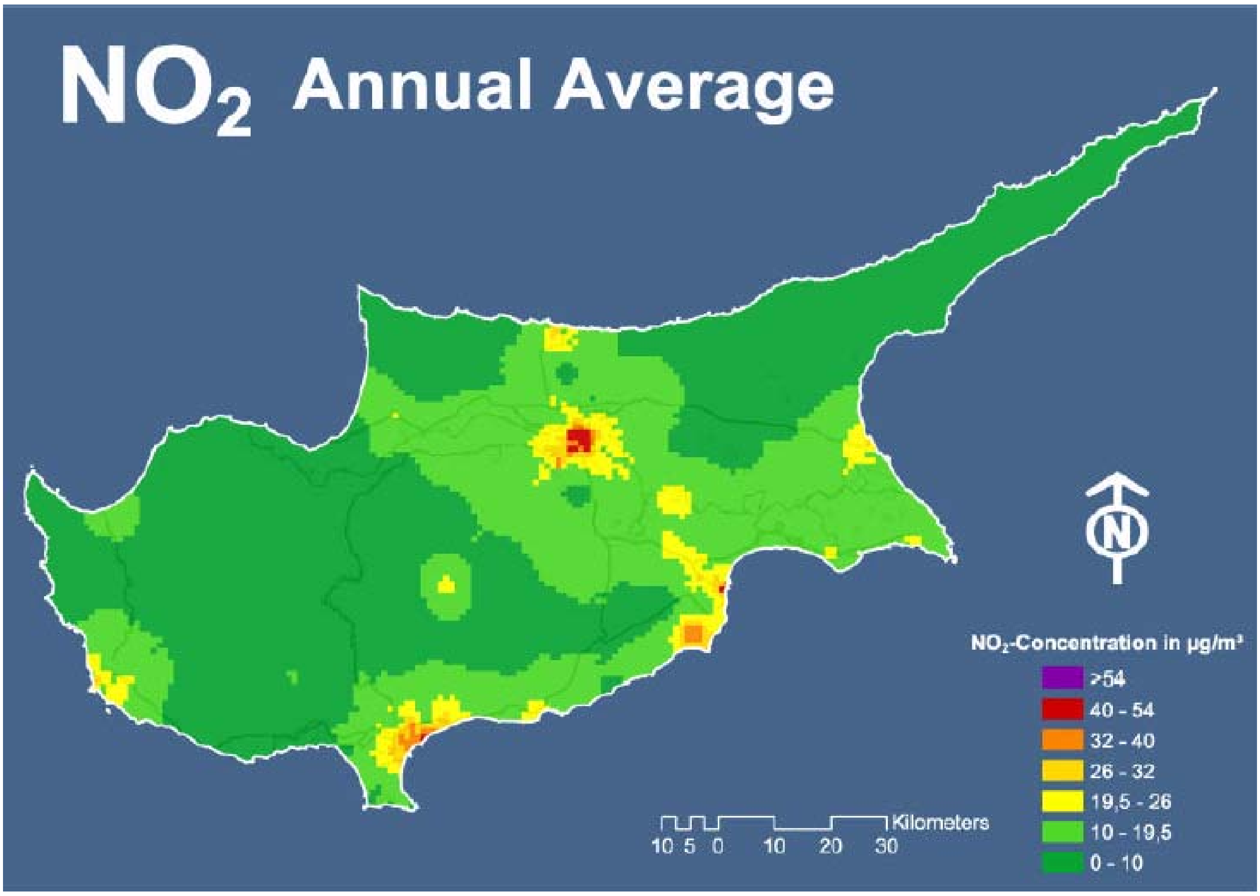


Figure 4.3. Mean annual interpolated NO_2 distribution over Cyprus (summer 2002 to summer 2003) using an Inverse distance to a power (IWD) interpolation algorithm

Nevertheless, the result reproduces quite well general tendencies on a larger spatial scale, as confirmed by the cross validation in Figure 4.4. The cross validation of ArcGis works by gradually leaving one data point out from the interpolation and estimating it's value with the remaining others. Higher measured NO_2 concentrations are underestimated; the points should ideally concentrate around the zero line, which indicates a zero prediction error.

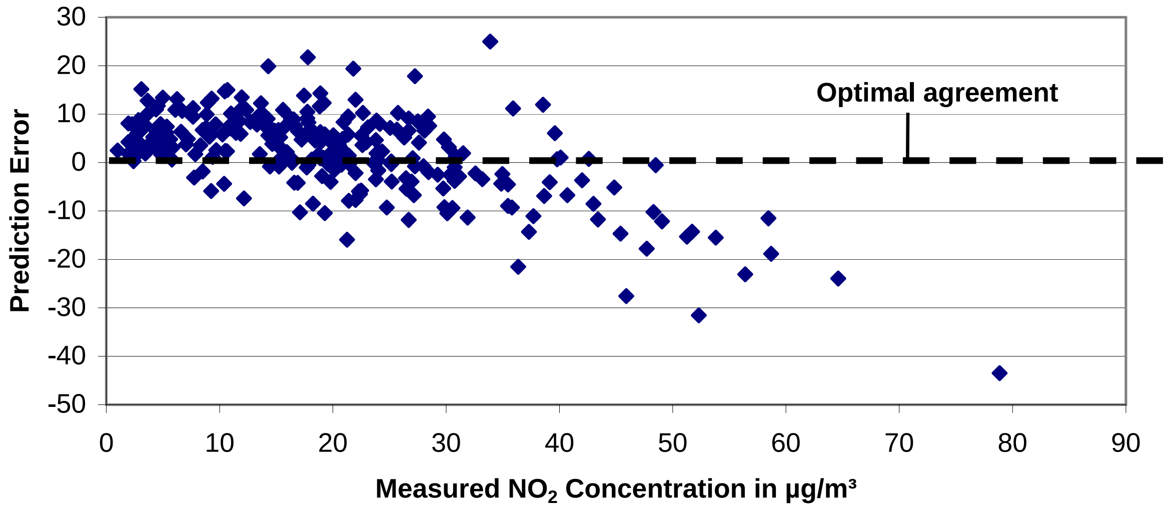


Figure 4.4. Cross validation of the annual NO_2 interpolation map of Cyprus

4.2.2 Interpolation considering additional variables

As a conclusion from the previous chapter, it can be stated, that simple interpolation is a strongly limited method and should be only used for quick visualization. The greatest limitation is, that it doesn't consider the space between two points, but only their spatial distance. In the case of Kriging the "spatial roughness" is also considered, but this doesn't really enhance the result, since it cannot identify additional sources or e.g. obstacles.

With Cokriging, we have on the first view a method that can consider such important influence parameters, but in practice it only uses this additional information to enhance the variogram [38]. The general limitations of interpolation remain, as proven by own tests. In combination with other methods like regression, interpolation can be an option. In the software FLADIS [15], interpolation is an important component and delivers acceptable results also in combination with a gaussian dispersion model, with which emission sources can be considered.

At the moment no satisfactory stand-alone method exists to calculate the distribution of air pollutants under consideration of the most important parameters. This work treats the development of such a method, based on artificial neural networks.

5 Neural network modelling

Artificial neural networks are a very simplified version of real neural networks. The human nervous system consists of 10^{11} to 10^{12} nerve cells and is able to carry out 10^{12} to 10^{13} “switching processes” - a complexity that cannot be rebuilt technically. Nevertheless, it is possible to understand the principles and to reconstruct a few cells that simulate the most important processes. In the year 1943, Warren McCulloch and Walter Pitts showed in their paper “A logical calculus of the ideas immanent in nervous activity” that even simple neural networks are able to calculate any arithmetic or logical function [39]. 1957, Frank Rosenblatt et al. developed the first successful neuro-computer, the so-called “Mark 1 perceptron”, which was able to recognize simple patterns. Neural networks on the base of backpropagation were developed in the early seventies and still are today the most popular networks [40].

5.1 Biological neural networks

To understand neural modelling, it is important to know basics about biological neurons. Unlike a personal computer, information processing in the brain is not directly influenced from outside, only input data are provided. The brain organises the processing of information on it’s own and has the ability to modify this process to enhance the results – in other words: it learns. The only information it receives from outside is a statement of the environment on the quality of the results. A neural network therefore can be described as a “black box” to which no interference takes place and whose concrete behaviour is invisible.

Biological neural networks are built by neurons, which consist in principal of four parts (Figure 5.1):

1. **Cell body:** Processes the information.
2. **Axon:** A fibre that conducts the output signal of the cell body to other neurons.
3. **Synapse:** The axon thickens at its end to form the synapse, which links the axon to the dendrites of other neurons.
4. **Dendrites:** Dendrites receive input signals of other neurons and direct them to the cell body.

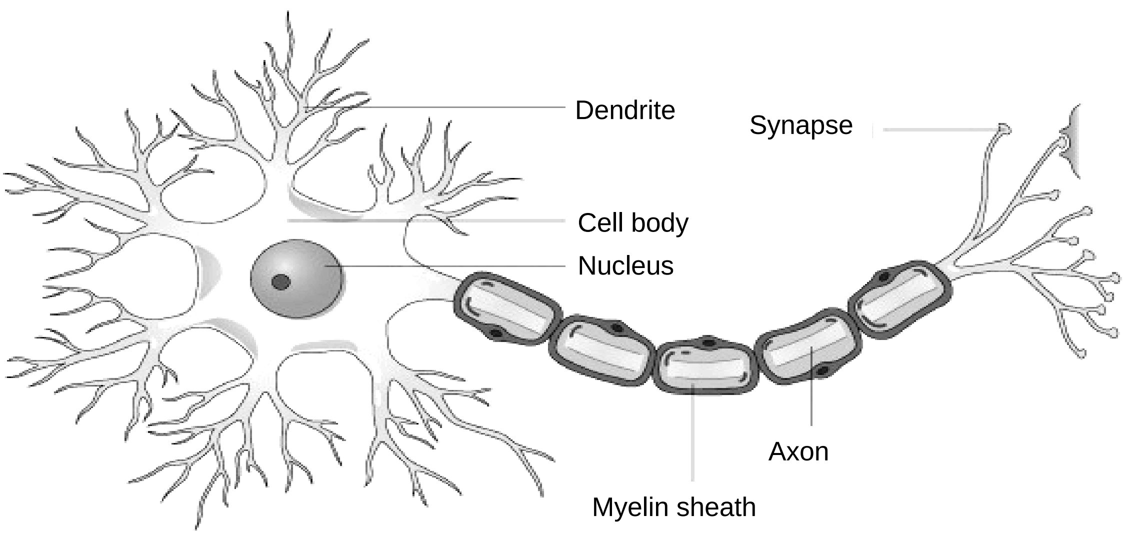


Figure 5.1. Biological neuron and its principal components [40].

When the impulse from another cell exceeds a specific threshold stimulus, the neuron is activated, processes the information and sends a modified impulse using the axon. This happens in the human nervous system up to 10^{13} times per second [40]. The network then stimulates e.g. muscles according to the processed information it received from the receptors (eye, skin…). Summarized, there are three principal tasks the network has to fulfil (Figure 5.2):

Figure 5.2. The three principal tasks of a neuron

By connecting many neurons, a biological neural network is formed. By changing the wiring diagram, the thresholds stimuli or the strength of impulses, the network can develop, respectively learn. Practically, the synapses are growing or shrinking to strengthen or weaken a connection [39, 41].

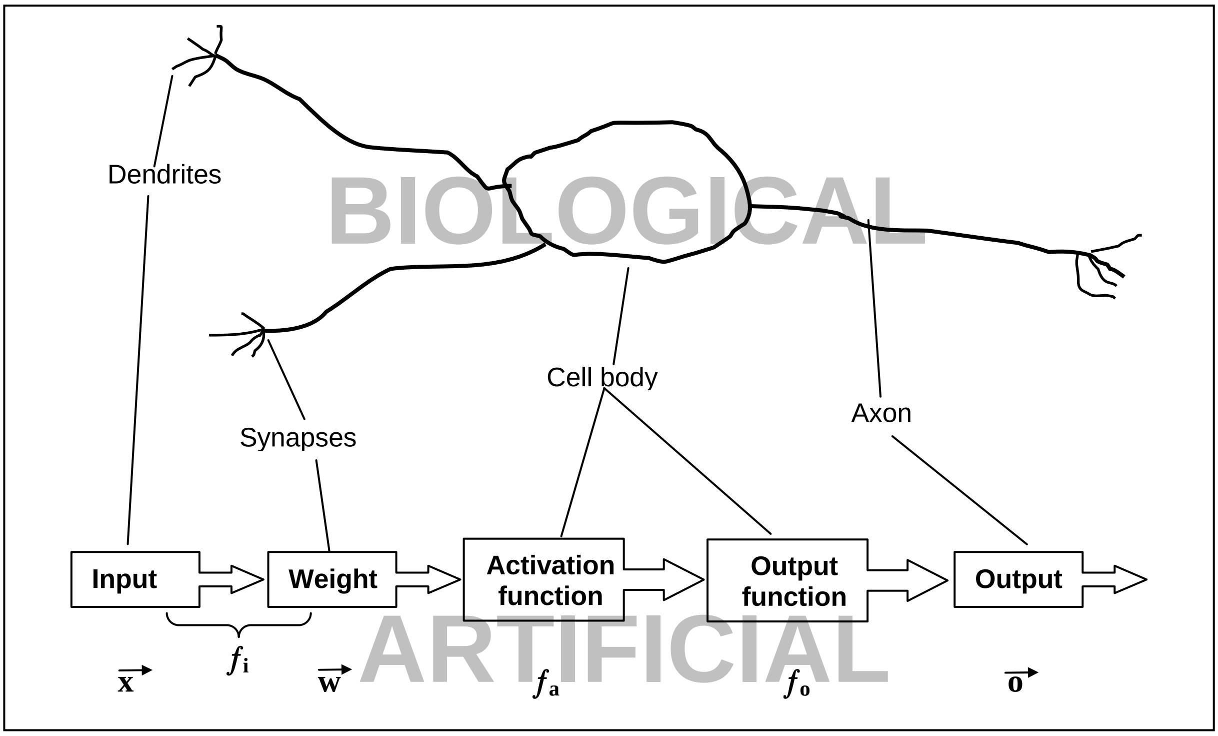


Figure 5.3. Principal biological components of neurons and their mathematical realization

5.2 Artificial neural networks

The components of an artificial neural network are strongly idealized neurons. Just like their biological example, they consist of four principal components: The cell body, the axon, the dendrites and the synapses. An artificial neuron has the same principal tasks as biological neurons. They are realized trough mathematical functions and vectors, which are connected with each other in a chain. In Figure 5.3 the principal functional components of biological neuron and their mathematical translation are shown.

One neuron alone doesn't build a network. A complete artificial neural network consists of layers of many single neurons that are connected with each other. The way how neurons are organised in layers and connected with each other defines the **network topology**. The most common topology is the so-called **multilayer perceptron** (Figure 5.4). It is very well developed and suitable for most problems, since it has theoretically the ability to approximate every continual function [39]. Most authors in the field of air quality modelling used a multilayer perceptron [e.g. 22, 23, 24, 26].

It consists of one input layer, one output layer and some intermediate layers. Every single neuron is connected with the previous layer where it receives its input from and with the following layer to which it passes the processed information. There are no feedback connections between the different layers – the information flows unidirectional (Figure 5.4).

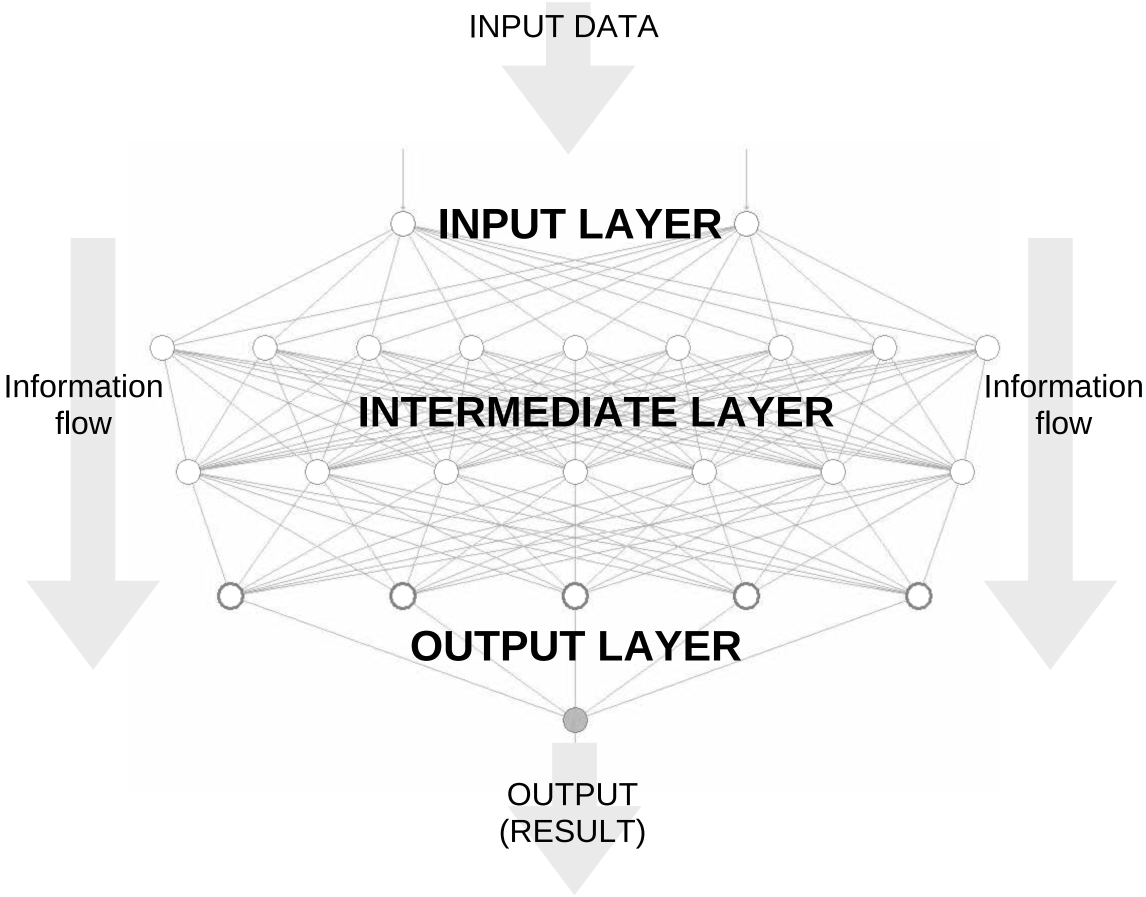


Figure 5.4. Topology of a multilayer perceptron with three principal layers

Of course there exist also many other well-developed network topologies, but they are mainly suitable for other purposes. Forecasts of time rows e.g. can be done with the so-called Elman-network [42] since it includes feedback connections, which memorizes previous values of the time row (Figure 5.5). Statistical methods use the comparable “auto regression” [43, 44, 45].

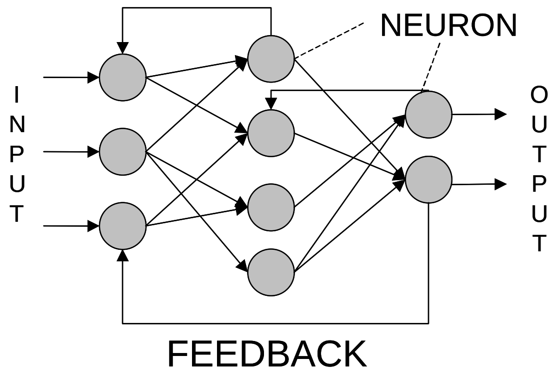


Figure 5.5. Elman network with feedback connections

Since this work here is based on a multilayer perceptron, the following comments will refer to this network-type.

In general there are two operation modes:

1. **Training mode:** The network learns the functional correlation of the provided input and output variables.
2. **Recall mode:** The learned correlations are applied to a dataset with unknown outputs.

During training, the network learns the dependencies of the different variables and finds the principles that connect them. In the case of **supervised learning**, applied in this work, an extern “teacher” provides both input and output - the most practicable and fastest method to calculate air pollutant concentrations, since the exact outputs are known. Other learning concepts require more extensive training and are suitable for other questions [39, 41].

In detail, the information processing of the applied network is organised in the following way [40]:

1. The user provides input data for the **input layer.**
2. The input data are normalized to a value range of 0 to 1.
3. Random starting weights of each connection in the whole network are being set.
4. Every neuron in the first intermediate layer multiplies each input value with the starting weight and finally summarizes all – the **input function** f_i.
5. The result from the input function is passed to the sigmoid activation function f_a.
6. The result from the activation function is passed to the output function f_o, which creates the output.
7. Now the steps 4 to 6 are repeated with the outputs of the previous layer being the input for the following layer.
8. When the output layer is reached, an **output vector** is created.
9. The calculated output is compared with the correct value using the squared distance. An **error function** is calculated.
10. According to this error function, the weights are adjusted in reverse direction towards the input layer.
11. Steps 1 to 10 are repeated, until a satisfactory error minimum is reached.

This training method is called **backpropagation** and using the correct terminology, steps 1 to 9 are called **forward-pass** and step 10 **backward-pass**. More concretely, the network optimisation is realized by finding the minimum of the error function, which can be depicted simplified as a curve with the weights on the x-axis and the errors on the y-axis. Starting the training, a random point on the curve is selected, and then the weights are set in a way to descent on the curve to find a minimum, ideally the global minimum. This method is called gradient descent (Figure 5.6) [40].

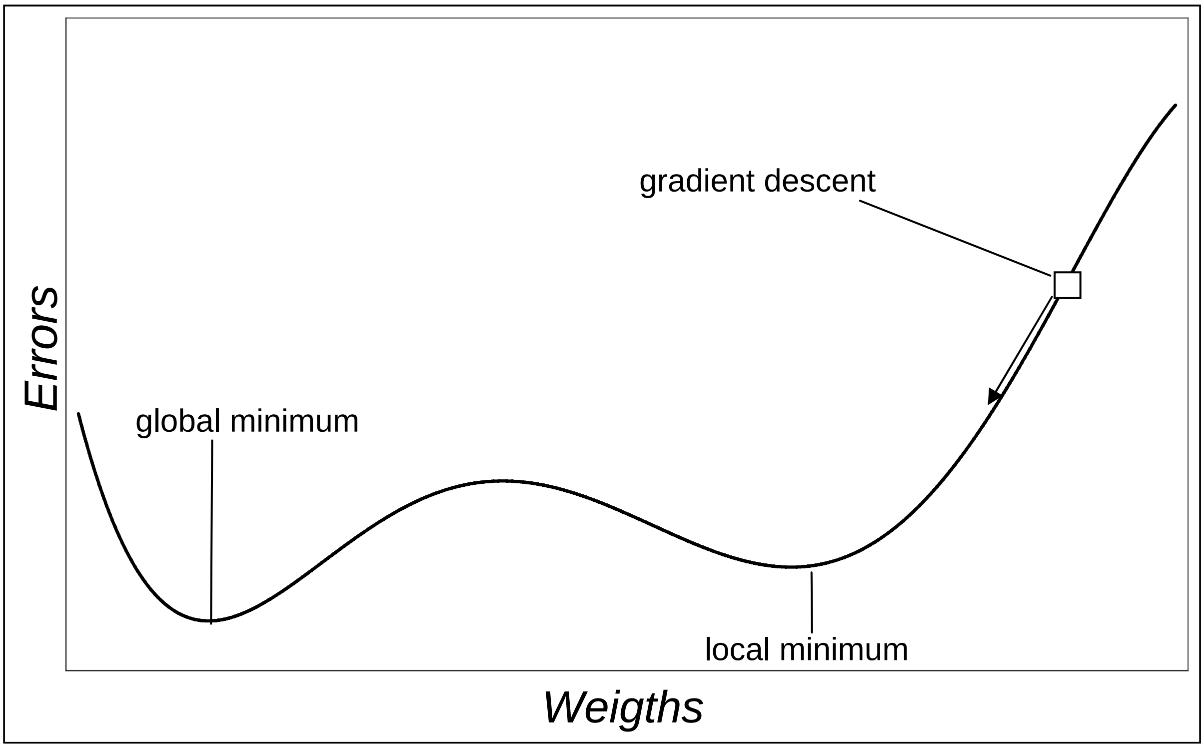


Figure 5.6. Training of a backpropagation neural network – gradient descent

Problems with backpropagation [40]

1. **Local minima:** When the dimension of the network is growing and more connections and neurons are set, the error surface is increasingly jagged. The training could therefore end in a local minimum.
2. **Symmetry breaking:** If the initialising weights are all the same, the following weight adjustments for the connections remain the same.
3. **Flat plateaus:** If there is a flat plateau on the error surface, the backpropagation stagnates. Many iteration steps are needed, which is the same behaviour when a minimum is reached. This could lead to the wrong assumption, that the training has been finished successfully.
4. **Oscillation:** If the gradient at the margin of a valley is too big, the training can jump, respectively oscillate between the two slopes of the valley.

5.3 Model development

As aforementioned, the development was carried out with a backpropagation neural network, a so-called **multilayer perceptron**. Developing a new model is an iterative process, where a systematic approach is inalienable. It is however not possible to document every single result and development step, this would go beyond the scope of this thesis. Only the most important results and findings are described below.

5.3.1 Systematic approach

First of all, the proposed problem was to calculate the annual average NO_2 distribution all over a large region (in this case Cyprus). As already discussed above, the NO_2 concentration field depends on certain influences that result from the three basic processes: **Emission – Transmission – Immission** (transition from the air to the receptor). These dependencies can be used to declare the searched immission, which leads to the applied approach: The study area Cyprus was overlayed by a one by one kilometre grid with each grid cell containing the influence parameters and some grid cells – the training cells – containing the influence parameters AND the NO_2 diffusive sampling results (Figure 5.7). With these training cells the functional dependencies are quantified (network training) and afterwards applied to the cells where the measured NO_2 concentrations are unknown (network recall).

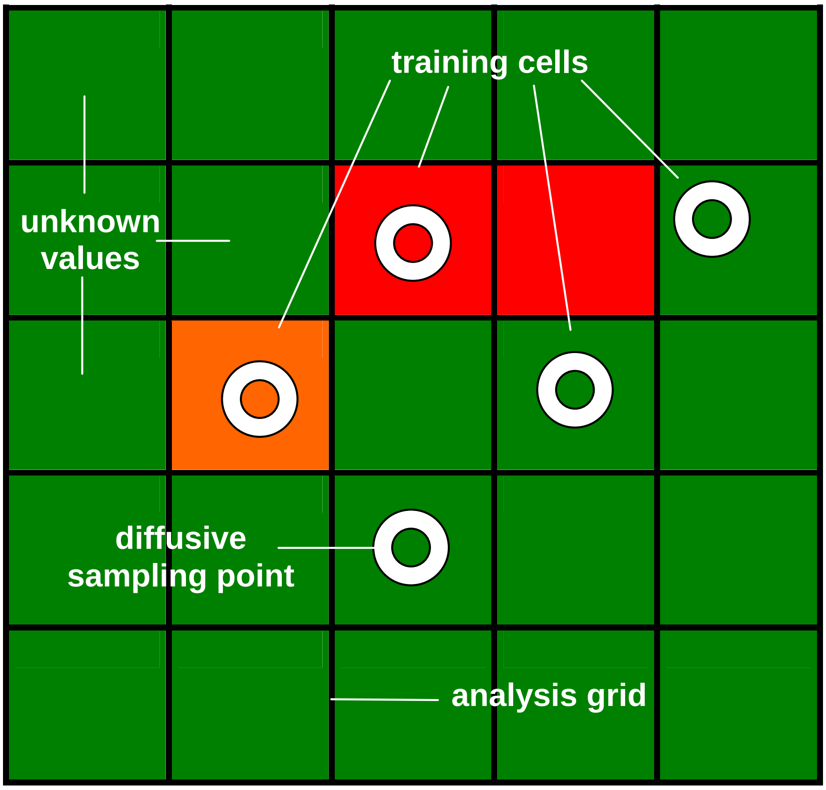


Figure 5.7. Training of a backpropagation neural network – gradient descent

5.3.2 Configuration of the neural network

The model configuration was carried out stepwise regarding different criteria:

1. **Neural network software:** As described before, the chosen software should include a model based on backpropagation. Furthermore it should have an interface to supervise the training progress, to evaluate the results with respect to the errors and to influence the most important training parameters. Since the development was carried out under the frame of a project with the duty to deliver a usable model, it should be also easy to handle. The finally chosen model was the Windows based application “Qnet2000” [48].

2. **Network topology:** Since a neural network can be considered as a black box and no real rules for the ideal configuration exist, except some experiences of other authors that can’t be transferred completely to other problems, one of the main tasks was to find the best possible network topology. This included the amount of intermediate layers, the number of neurons and the type of the activation functions. Since the software allows 8 hidden layers with unlimited neurons each, the possible combinations are huge.

3. **Training parameters:** There are some adjustable network parameters, influencing the training performance. The most important is the learn rate.

4. **Training variables:** A major task of the model development was to find the variables that deliver a significant contribution to model the NO_2 distribution. High demands were made on the data – they had to be available for the whole area of Cyprus and reliable, since incorrect values could badly influence the network performance, although a neural network is not as sensitive to such values like the regression analysis.

5. **Quality assessment:** An important issue is to assess the quality of the model outputs, whether they agree with real measured values. This finally defines the quality of the model, if it makes sense to use it or further develop it in future. Different assessment methods were used, being an internal assessment of the modelling software and an external assessment including the mapping of a result that passed the internal assessment**.**

Steps 2 to 4 were carried out quasi-simultaneously, since the results of each step influences the approach of the others. To find out how the output reacts on a change of the topology, the training parameters and the different variables, a systematic approach was chosen: For each training cycle only one characteristic was modified, the others remained the same. In addition it was also tested how the network behaves on combinations of parameter changes. With this method, general tendencies were found out and constantly checked by using internal assessment methods like the rout mean squared deviation of the results and the correlation coefficient. This approach is depicted in Figure 5.8.

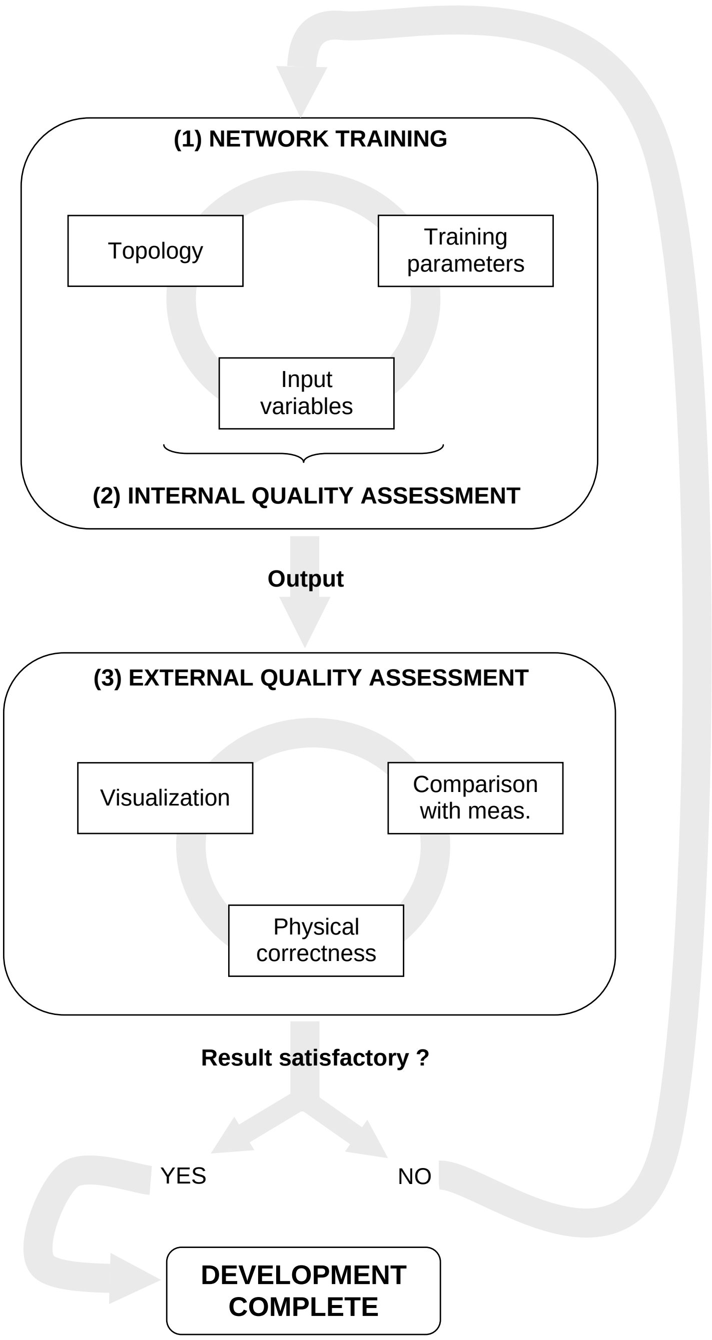


Figure 5.8. Steps in the development of the neural network

5.3.3 Generalization and memorization

A test set of 50 cases was set aside for every training run to monitor a possible **overtraining**. 50 cases are considered to be statistically significant [47]. Here, the real measured values of the test set are compared with the network output and statistical errors are calculated. If the error of the test set decreases, the model still **generalizes**, which means, that it learns the correlations, if the error of the test set increases, the network **memorizes** – it recalls the values of the training set. So a test set is very important to find the right point to terminate the training, since the error of the training set always decreases. In Figure 5.9 an exemplary error course during training is depicted. After approximately 10.000 iterations the root mean square error falls dramatically. This shows, that the network learns the major correlations relatively fast – a pattern that was found quite often. Afterwards, a slow but constant error decrease can be observed. This could lead to the conclusion that more iterations would improve the model performance.

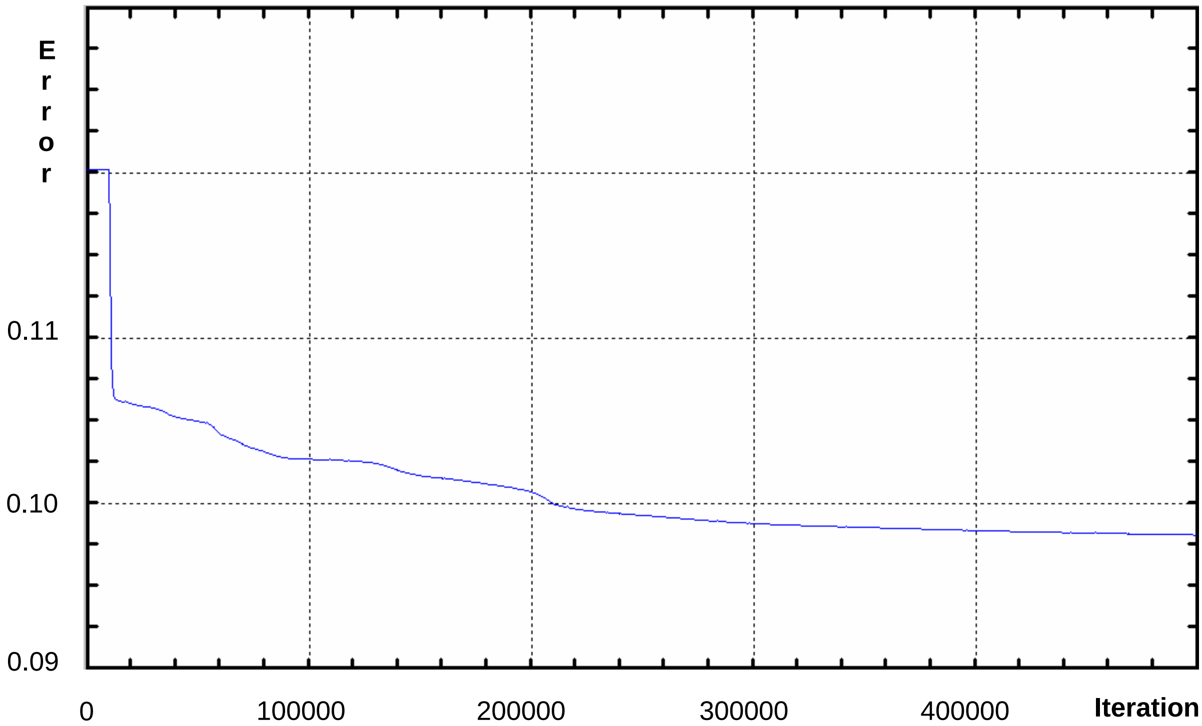


Figure 5.9. Root mean square error of the normalized model output (NO_2) as calculated by the neural network software using the training data set with 270 values

Figure 5.10 shows however a different picture: Here, the error course of the test data set is shown. The same abrupt error-decrease after 10.000 appears, followed by alternating ascents and descents to reach a local minimum after 80.000 iterations. Finally the error decreases more or less constantly - an indicator of overtraining and memorization. Patterns like these were observed throughout the whole model development, so that the iteration number was limited to 100000.

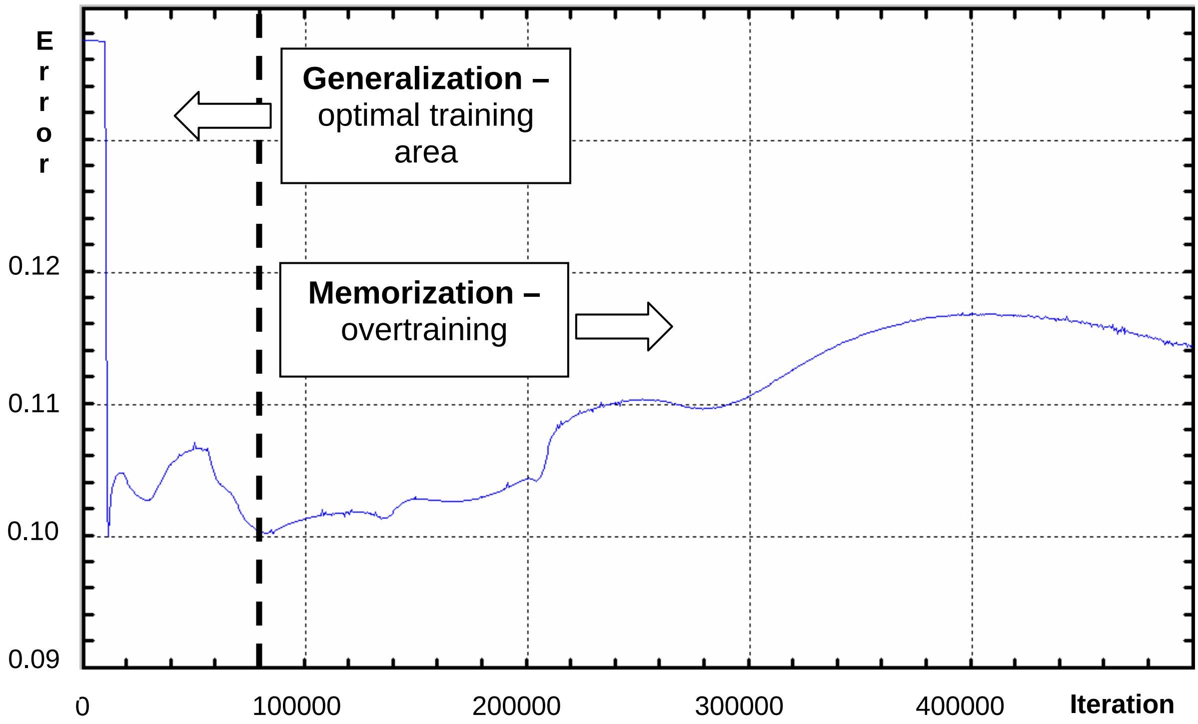


Figure 5.10. Root mean square error of the normalized model output (NO_2) as calculated by the neural network software using the test data set with 270 values

5.3.4 Optimisation of network topology

To find out the best network topology, three principal criteria have to be considered:

1. The **number of input neurons**: Choosing the best number and types of input neurons is the most time consuming part of the optimisation process and will be discussed later.
2. The **number of hidden layers and neurons**: Authors like Omasreiter [26] and Gerboles [22] used only one hidden layer with up to 12 neurons, but own experiences have shown that three hidden layers with 3 to 5 neurons each deliver the best results. Multi-hidden layer networks tend to grasp complex concepts more easily than networks with one layer. One reason for this is that the multi-hidden layer construction creates an increased cross-factoring of information and relationships [48]. With the so-called “hidden node analyser”, the contribution in percentage of every neuron to the next layer can be retrieved – an important tool to assess the best network topology. A low contribution of a neuron indicates an over dimensioned network and should lead to a reduction of neurons. In Table 5.1 an exemplary output of the hidden node analyser from a trained network with three hidden layers containing three neurons each is shown. Every neuron contributions at least around 30% to the next layer, which indicates a well configured network topology. But this alone doesn’t guarantee a good performance.

Table 5.1. Hidden node analyser of a neural network with three neurons in three hidden layers

Average Hidden Node Contribution to Next Layer
Network Name: training coord emi hi pop
Iterations: 100000

Hidden Layer	Node	Percent Contribution
1	1	31.99
1	2	39.09
1	3	28.92
2	1	30.60
2	2	29.21
2	3	40.20
3	1	28.80
3	2	32.77
3	3	38.43

To assess the performance, the error history of different topologies with the same frame conditions has to be regarded. A significant increase or decrease of neurons doesn't increase the network performance as in Figure 5.11 is shown: Configuration 1 with 3 to 5 neurons results in the low errors compared to configuration 3 with 9 neurons. Research results have shown that the same is valid for the number of hidden layers. Another disadvantage of using a larger number of neurons is the increase of computing time.

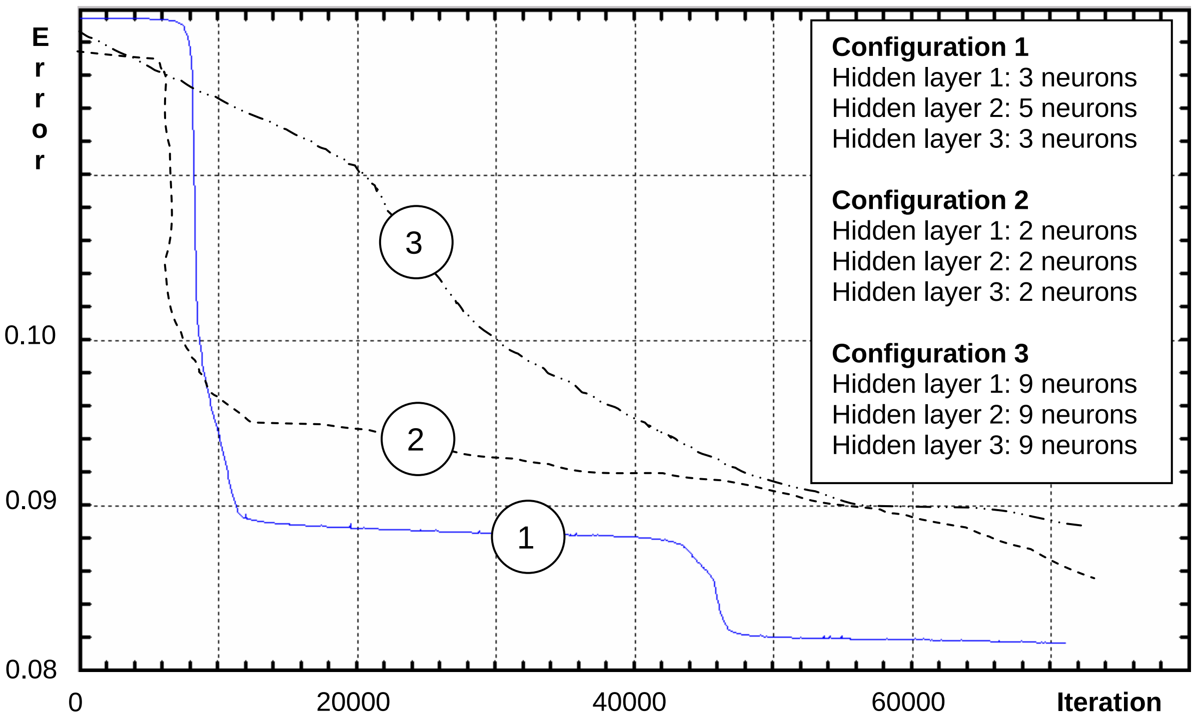


Figure 5.11. Different network topologies and their history during training – the network was trained with the summed NO_x emissions and the population

3. The **activation functions**, respectively the **output function**: As previously shown, the neuron's activation function serves for the purpose of controlling the output signal strength for the neuron, except for the input layer which uses the provided input. Four different function types are possible: Sigmoid, gaussian, hyperbolic tangent and hyperbolic secant. The sigmoid function is mostly applied and agrees best with biological neurons. Nevertheless, all of these functions have been tested and finally it was discovered that the **sigmoid activation function** performs best. It can be described by the mathematical relationship **1/(1+e-x)** and acts as an output gate that can be opened (value = 1) or closed (value = 0). Since the function is continuous, it is also possible for the gate to be partially opened.

5.3.5 Optimisation of training parameters

Two factors are used to control the training algorithm's adjustment of the weights. They are the **"learning rate coefficient"**, **eta**, and the **"momentum factor"**, **alpha**.

Eta determines the size of the node weights adjustments during training. If the learning rate is too fast (i.e., eta is too large), network training can become unstable (oscillation problem). If eta is too small, the network will learn at a very slow pace and flat plateaus on the error surface could cause a stagnation of learning. With the "Learn Rate Control" (LCR) feature of the software Qnet2000, eta can be altered automatically. LRC will drive eta higher or lower in a systematic fashion depending on the current learning activity. If the network appears to be learning at a relatively slow rate, eta is driven up quickly. Conversely, if the network is learning at a fast pace, eta will be held constant or even lower to avoid instabilities. Following the recommendations of the software tutorial, the training was started at a low eta without LCR. LCR was turned on after 10.000 iterations. Training without LCR didn't lead to an optimisation of the network performance.

The **momentum factor** has a smaller influence on learning speeds, but it can influence training stability and promote faster learning since it damps high frequency weight changes. During the development phase, different momentum factors were tried, but finally the default values were applied since they delivered the best results. Practically, the influence of alpha was negligible.

5.3.6 Reproducibility

A very important issue of a model is the question, if its results can be reproduced at any time. Since the initialisation of connection weights is always random, every model is unique. Therefore, the learned functional correlations of two models with the same frame conditions are also always different. The worst case would be symmetry breaking, where all weights are the same at start – a hardly probable case. Practically, it was found out that the deviations are negligible. Nevertheless, the models should be checked and rerun with a new weight initialisation. This was done two to three times for every model.

Generally speaking, it was found out that neural network modelling is a reproducible method if the same frame conditions are set (eta, alpha, topology).

5.4 Application of input variables

The most problematic part of the model development was to find the available variables that explain the searched spatial NO_2 distribution. These variables had to be formatted or produced for every cell of the 1x1km grid Cyprus was overlayed with. From the logical point of view it is clear what influences the average NO_2 concentration field (see chapter 3.4), but practically it is a large-scale task to finally have the data in the desired format or even to obtain them.

Table 5.2 lists the possible influence data and comments their availability and possible contribution to the model.

Table 5.2. Parameters that influence the spatial distribution of NO_2 and their possible contribution to the model performance

Input variable	Data type	Availability	Possible contribution to model
Coordinates	GIS map from cartographical service of Cyprus	Very good	Neural interpolation of NO_2 values – same problems like normal interpolation
Height above sea level (3d- model)	GIS map from cartographical service of Cyprus	Good - has to be generated from contours	Secondary influence of height through the vertical distance to the sources and to the inversion
NO_x Emissions	GIS map from emissions inventory of project	Good - has to be made	Very important, since these sources are also the origin for the secondary component NO_2
Population density	GIS map from cartographical service of Cyprus	Good - has to be altered and formatted	Very important, since the NO_2 concentrations depend largely on human activity
Wind direction and wind speed	Statistics, measurements during project	Problematic – measurements are only point values and cannot be taken as valid for a whole grid cell	Very important, since it is responsible for the dispersion of the pollutants
Land usage	Printed maps, partly GIS maps, Satellite data	Problematic – no good data base available, satellite data with a rough resolution	Generally very important, but it intersects with the variables emissions and population density

The central analysis tool for formatting, producing and visualizing the data was ArcGis 8.x.

5.4.1 UTM coordinates

With the UTM coordinates as input data the neural network can learn the spatial dependency of measured NO_2 concentrations. Theoretically, the produced output cannot be more than a simple interpolation map, since no additional information is provided to the net. In recall mode, the UTM coordinates from the centre of every 1x1km grid cell were entered. The recall-grid for all models was aligned to the UTM grid starting at 430000E 3820000N and ending at 650000E 396000N. As expected, the result (Figure 5.12) is very similar to a normal interpolation with the difference that the actual interpolation (measurement) points are blurred. The reason for this behaviour lies in the generalization, the network carries out: It learns that certain coordinate combinations result in high NO_2 values and others in low values. Since the peak levels are to be expected near NO_2 emission sources, the map shows red colours in and around the major cities and the highway at the south coast. Only the cities of Nicosia, Limassol, Larnaka and conditionally Paphos emerge clearly on the map. Kyrenia and Famagusta are blurred, since no high NO_2 concentrations were measured and a low number of measurement (training) points were placed. This lowers the network response to these sites. Again, the result depends largely on the chosen measurement sites, since the UTM coordinates do only indirectly indicate if there are emission sources in a very generalized way. That’s why the result map is only poorly structured, containing clear, simple shapes.

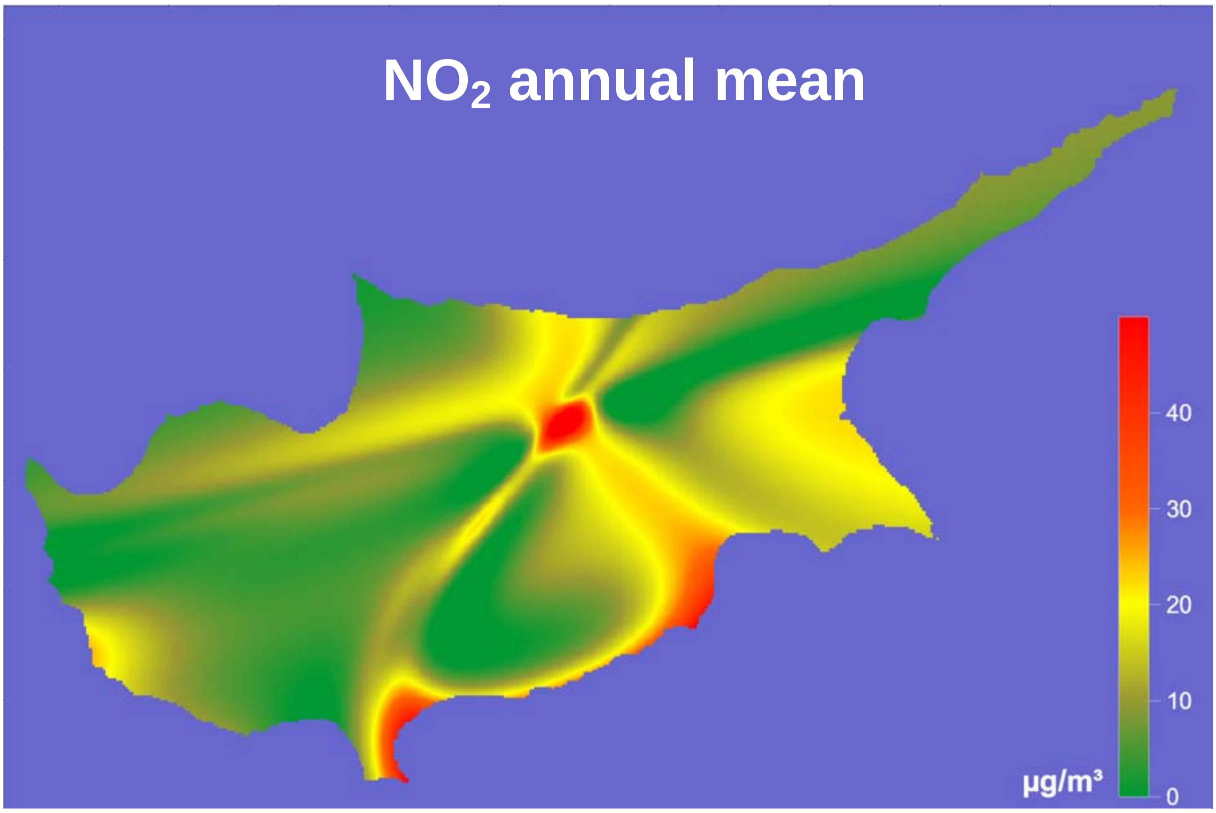


Figure 5.12. Spatial NO_2 distribution as predicted by a neural network, trained only with UTM coordinates

5.4.2 Height above sea level

Theoretically, the spatial NO_2 distribution corresponds indirectly with the height above sea level:

1. **Photochemistry:** As described above, the so-called photostationary equilibrium between NO, NO_2, O_3 and O_2 influences the NO_2 distribution. O_2 and above all O_3 oxidises NO to NO_2 – Ozone therefore promotes the formation of NO_2 depending of course on the offered amount of NO. It is well known, that the ozone concentration directly correlates with the height above sea level [49] and consequently so does indirectly the NO_2 concentration. Nevertheless this is questionable since there is always Ozone in the air and other components like peroxy radicals Volatile Organic

Compounds are also part in this reaction chain. Finally, solar radiation splits NO_2 in NO and O and positively correlates with the height above sea level.

2. **Mixing layer:** Mountainous areas usually are located above the nocturnal ground inversion layer. Most of the emission sources are located below this layer; the transport of polluted air masses to higher elevations breaks down during night. The average concentrations are therefore lower in mountainous areas.
3. **Emission activities:** A so called trivial correlation [50, 51] is the fact that on the one hand the emission activity of NO_x correlates with the land usage and on the other hand, the land usage with the height above sea level. Of course, this cannot be generalized, but in the case of Cyprus, the major cities are located at the coast, except Nicosia, which is situated on the flat plain of Messaoria on a low elevation level. All power plants in GCC and the cement factory of Vasilikos are located at the south coast, the power plant of TCC at the North coast near Kyrenia. Also the strongly frequented highway in GCC is mainly situated at a low height above sea level.

Considering the facts mentioned above, a good correlation between the height and the NO_2 concentrations (received through diffusive sampling measurements) can be expected. Sensitivity analyses confirm this conjecture: Figure 5.13 depicts the dependence of the two variables in a scatter plot where a clear non-linear correlation emerges.

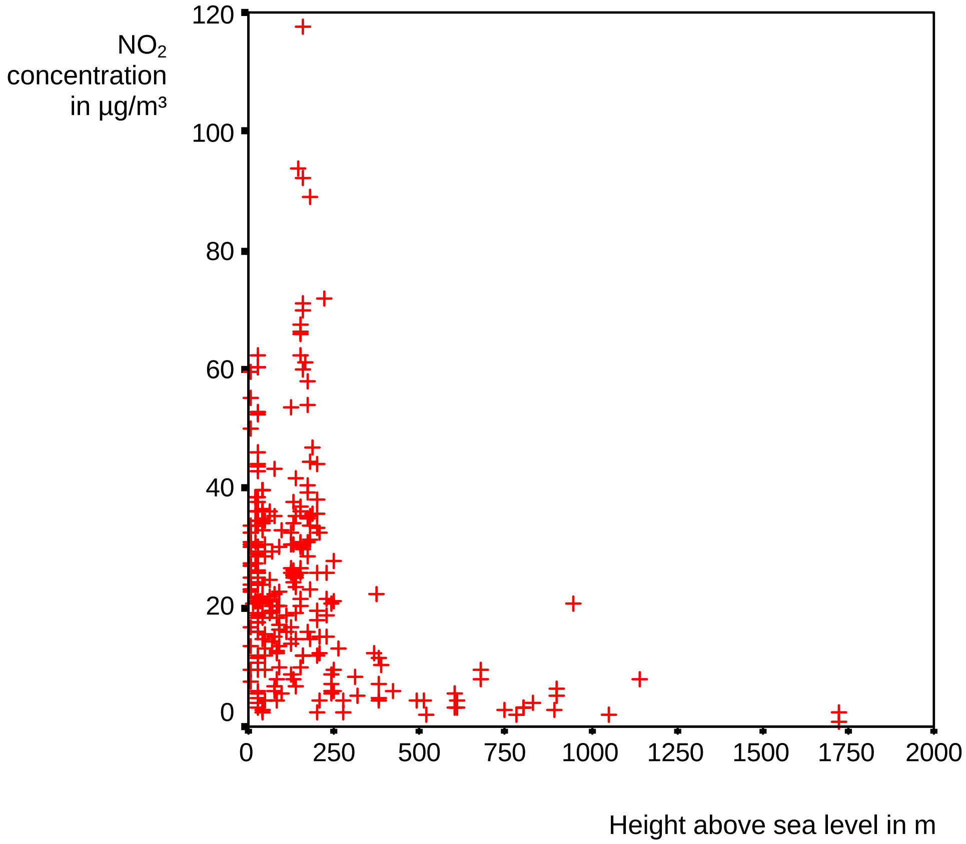


Figure 5.13. Scatter plot of NO_2 concentrations measured by diffusive sampling and the height above sea level at the measurement sites

The height values from the above sensitivity analysis were directly measured on site with a GPS (Global Positioning System) device. The measured altitude values were also used for network training, but for the network recall they had to be calculated for every grid cell. To achieve this, a digital elevation model (DEM) was calculated by the interpolation of the contour information of the cartographic service of Cyprus. For every grid cell an average height value was created and then applied to the network. To assure the reliability of the digital elevation model, the measured values from the GPS were compared with the interpolated results. Since the vertical resolution was chosen as 50m, the comparison resulted in a very good agreement. A higher resolution is not practicable, because the heights were averaged for every 1x1 km grid cell and the created digital elevation model was based on the digital 1:250.000 map of Cyprus. With this resolution, the network could be trained with 40 height classes – from the low coastal regions and plains to the highest point of Cyprus, Mt. Olympus with 1952m above sea level.

As expected, with the network training it was possible to reproduce the non-linear correlation as depicted in Figure 5.13. Agreement statistics calculated by the neural network software were accordingly good. The network recall lead to the result map presented in Figure 5.14. On the first view it seems to be well structured, but looking more precisely, one can see that the actual locations of the major cities are shifted. Some special features of this map reveal which correlations the network learned: Paphos at the southwest coast emerges quite clearly in red reproducing quite well it's actual geographical location, but the north and northeast edge of Paphos shows three red finger-like structures. This is a place were no emission sources are located and no NO_2 measurements were taken, they match perfectly with three river valleys coming from the Troodos mountains. Obviously the network applied here the above-described correlation that the major emission sources are situated at a low altitude, the same problem appears at the north-western coast near Lefka.

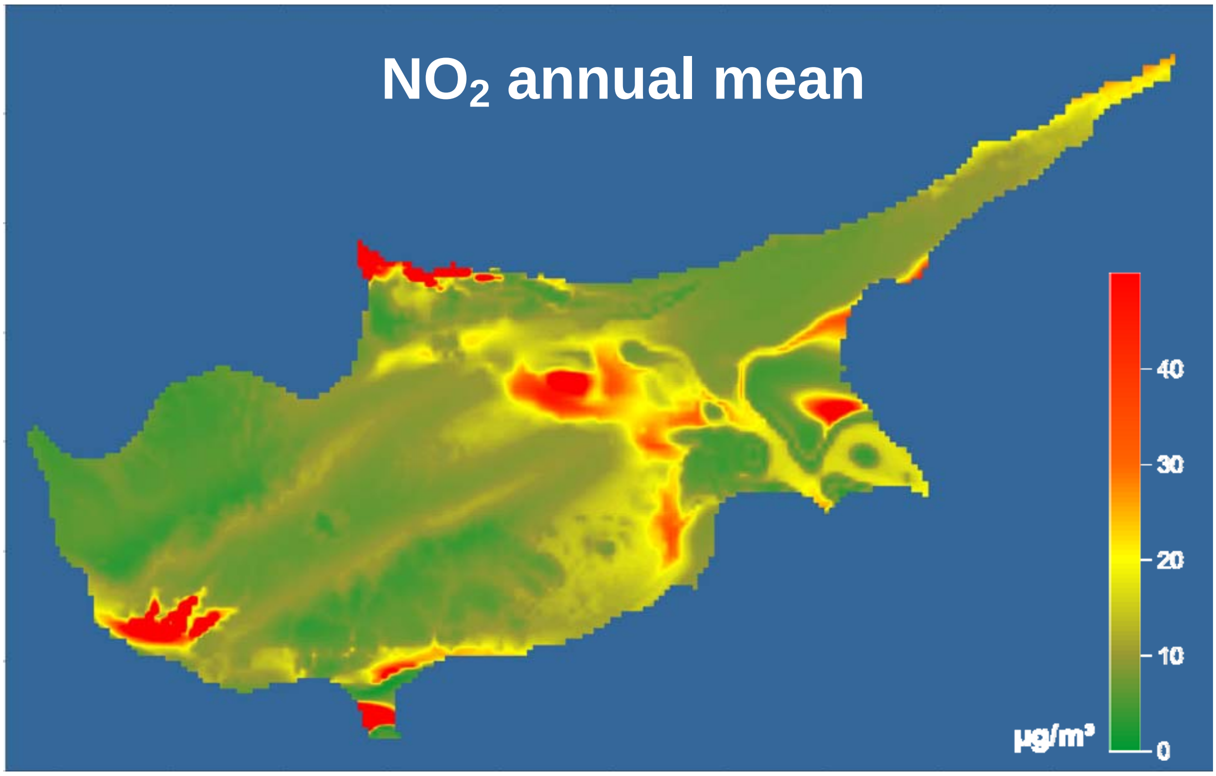


Figure 5.14. Spatial NO2 distribution as predicted by a neural network, trained with UTM coordinates and the height above sea level

Other possible correlations as mentioned above seem to be generally suppressed by this major feature. From the physical point of view it can be concluded that there is no justification to use the altitude as an input variable for the neural network, although it might optimise the statistical error. Later attempts in combination with other variables that were expected to compensate the problem showed the same structures.

5.4.3 Wind direction and wind speed

The dispersion conditions are determined by meteorological parameters like wind speed, wind direction and atmospheric stability. Their values directly influence the dilution of air pollutants (see also chapter 0). An assessment of air quality therefore always has to include the consideration of the local meteorological conditions. Furthermore, seasonal weather changes influence human emission activities and cause for example an increased heating during winter or additional touristic traffic in summer. These two principals have to be considered separately since they underlie different mechanisms.

- **Dispersion conditions:** All meteorological parameters are highly dynamic and fluctuate permanently. The most important ones wind speed and wind direction can be described with meteorological models. Since annual averages are considered, statistics for every grid cell should be calculated. For the wind direction, this doesn't make sense since statistics would include distribution parameters (multiple values of one characteristic) that cannot be implemented in the neural network as developed in this work. For the wind speed, an average value would be practicable, but couldn't be calculated during this thesis.
- **Emission activities**: The consideration of seasonal changes could be carried out by calculating two different models – one for winter and one for summer [53]. This requires strictly speaking the existence of seasonal input data, which is unfortunately not the case. Nevertheless two seasonal models were calculated by training the network with annual input data and seasonal diffusive sampling results. This resulted in a satisfactory overall network performance. It was correctly calculated, that the NO_2 concentrations are higher in winter. Only course tendencies could be reproduced here, a complete data set with seasonal values can be expected to improve the results significantly.

Another possibility to include the dispersion conditions in the model is to implement them in the emissions inventory for the most important sources in Cyprus as described in the following chapter.

5.4.4 NO_x emissions

Special attention was directed to the emissions inventory – its potential contribution to the network performance is very high. All serious modelling methods to calculate a spatial pollutant distribution like dispersion models or FLADIS include emissions as input. It is obvious why: Without emissions there are no air pollutants, so there must exist a correlation between the NO_x emissions and the searched NO_2 concentrations. During the project the first coherent emissions inventory for Cyprus was prepared and could be used for this work. Its characteristics determine the modelling approach:

- **Quality:** Are all important sources included? If sources are omitted, the model performance may suffer. Some sources are actually not included in the emissions inventory due to difficulties in getting numbers, especially for the northern part of Cyprus. Domestic heating and waste burning activities at several unofficial sites (see Annex A.1.) could be only partly registered or described qualitatively. So, for some sources rough estimations had to be made, since they would have required extensive surveys and additional personnel or project time.

- **Time resolution:** Daily or hourly values, weekdays, seasons or months distinguished? The time resolution of the model is mainly determined by the emissions inventory and partly by the other variables. Here, average daily values valid for the whole year are available, representing an average weekday.
- **Spatial resolution:** Are the emissions summed in a grid with a certain spatial resolution or are line, point and area sources considered? Are the coordinates known for each source? Fortunately, the emissions inventory is divided in the three basic source geometries and the spatial locations are known.

Since the author was involved in the preparation including the realization in a geoinformation system, the emissions inventory was prepared to be suitable for the model development.

As mentioned above, the spatial resolution of the model was chosen to be a 1x1 km grid, so the emissions were prepared to match the same grid. Practically, all emission sources that fall in a particular grid cell were summed. The area and line sources were proportionately converted and assigned according to their length or area. This method assumes however that the emissions of all sources directly pass over to ambient air pollutant concentrations in the considered grid cell. This is for the majority of sources with relatively low emissions at a low height level an acceptable approximation, but for the major sources a strong distortion of the real conditions. In reality, the air pollutant leaves the outlet of the source and disperses vertically and horizontally according to the present wind and turbulence field. This process is called **transmission** and depends on the source height, the source strength and the dispersion conditions. In Figure 5.15 this context is depicted.

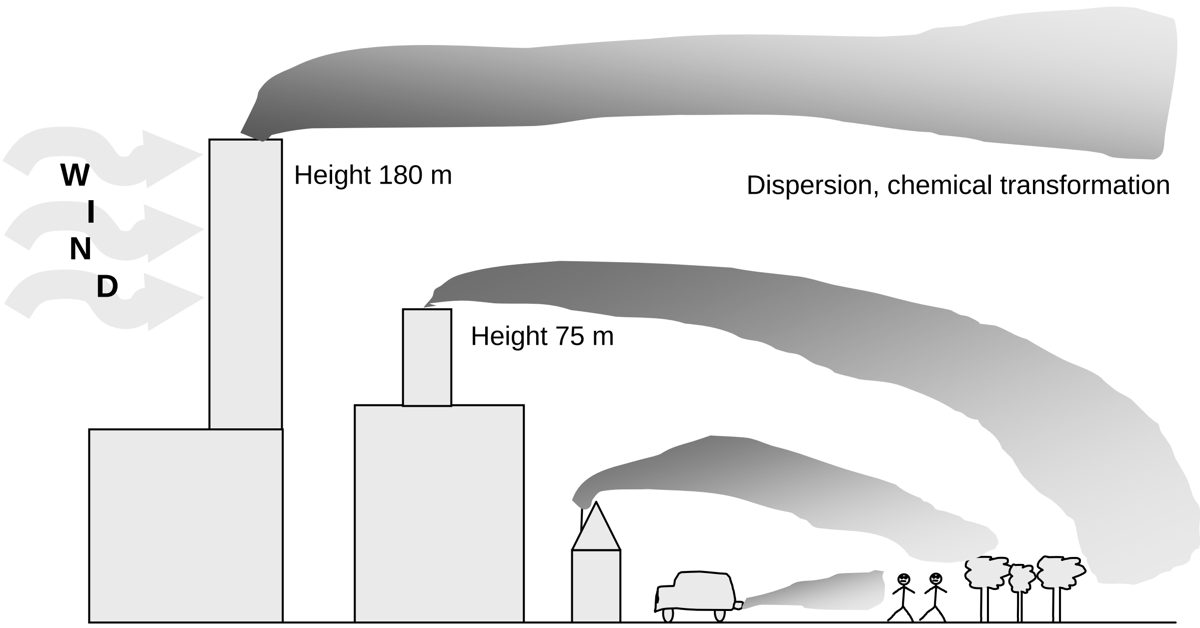


Figure 5.15. Paths of air pollutants from points of emission to the location of influencing ambient air quality

In a high source altitude the air pollutants are distributed over a large area and do not affect the direct neighbourhood. To count these emission sources to the grid cell of their actual geographical position would result in an overestimation of air pollutant concentrations in this grid cell.

For the correlation of ambient air concentrations with emissions, the actual location where the emissions touch the ground have to be considered. Usually a dispersion model with a wind field processor is the only possibility to calculate the pollutants distribution around a source [54]. It is not possible to replace a dispersion model by a neural network, especially when the regarded area is topographically structured with different elevations [52]. Moreover, the shape of the dispersion plume has to be explicitly given by a mathematical formula. Different approaches were used to solve this problem for the major sources in Cyprus. They are described below.

5.4.4.1 Methods to determine the plumes of the major sources in Cyprus

Point sources:
According to the emissions inventory, the major single point sources of NO_x are: The cement factory near Vassilikos and the power plants of Moni, Dhekelia, Vassilikos and Kyrenia. All important point sources of Cyprus are located at the south coast between Limassol and Agia Napa, except the power plant of Kyrenia, which is situated at the north coast. A Gaussian dispersion model (“P&K 3782”) was used to calculate the NO_x distribution around these sources. Since this work serves to find a method based on a neural network it was decided not to use a more sophisticated model which would have stretched the development time significantly. Another reason is that Gaussian models deliver good results in a flat modelling area [13], which is the case here, since the coastal regions are flat in the first approximation. Figure 5.16 shows a three dimensional view of Cyprus with the most important point sources.

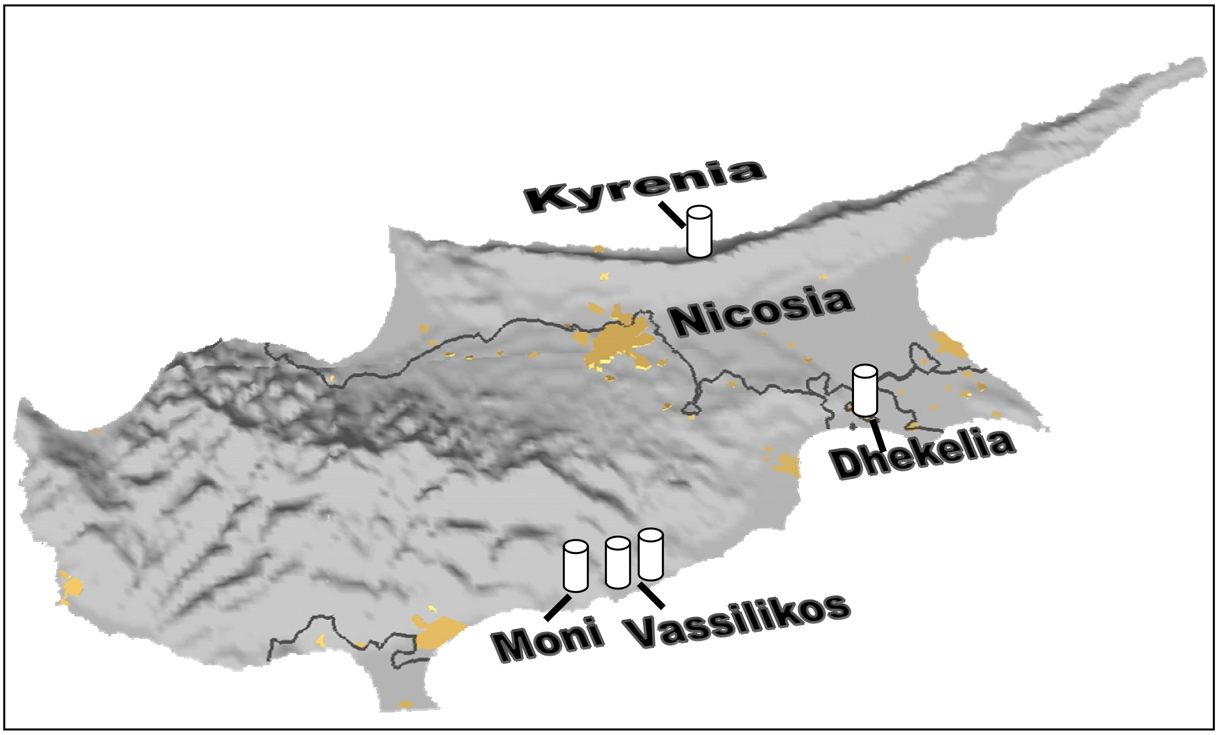


Figure 5.16. Three-dimensional view of Cyprus with the major point sources – raising factor is ten; the stack heights in the map do not reflect their actual heights

In addition it is not so important to match the exact shape of the plume since the calculations are added after some conversions to the emissions inventory, which has a model input resolution of 1x1 km.

Some of the input parameters for the Gaussian dispersion model had to be calculated or estimated due to a lack of information. For all sources the total daily emission of NO_x was known. For the Vassilikos power plant, the most important parameters were provided by the Electric Authority of Cyprus. They are listed in Table 5.3. Similar information was provided for the power plant near Kyrenia. Information concerning the other sources could be found in the internet but most of the parameters had to be estimated and derived from the available ones using known technical specifications [55].

Table 5.3. Emission parameters of the Vasilikos power plant

Parameter	Value
Stack height	120 m
Stack diameter	2,9 m
Gas temperature	125-130 °C
Pollutant volume flow	83,3 m/s
Fuel consumption	8 l/s
SO_2 emission rate	146,1 g/s
SO_2 concentration	1753,3 mg/m³
Gas flow velocity	12,6 m/s

Besides the operational specifications of the source, wind statistics are a necessary input for the dispersion model. For each source a wind statistic was created according to own measurements, measurements from the project partners, measurements from the Electricity Authority of Cyprus and wind statistics from the Meteorological Service [56]. Wind roses for the used measurement stations are shown in Annex B.1.. With the dispersion software the concentration field were calculated according to this statistic, so that the result should be reflected by the wind rose.

In Figure 5.17 the mean annual dispersion of NO_x around the cement factory near Vassilikos as calculated with the modelling software P&K3782 is depicted. In Annex B.1. to B.1. the same for the power plants of Dhekelia, Moni, Vasilikos and Kyrenia is shown.

The highest concentrations occur at the southern part of the plume that crosses the coast to the mediterranean sea. Close to the cement factory near Vassilikos, the power plant of Vasilikos is situated for which the SO_2 distribution was modelled by Joao Ferreira in his master thesis with a more sophisticated model - the Particle model Austal 2000 [57]. Here the same wind data were used and also a digital elevation model. If the topography would have a great influence on the dispersion, the plumes of the two modelling approaches would differ significantly, but in fact they match very well (compare Figure 5.17 and Annex B.15).

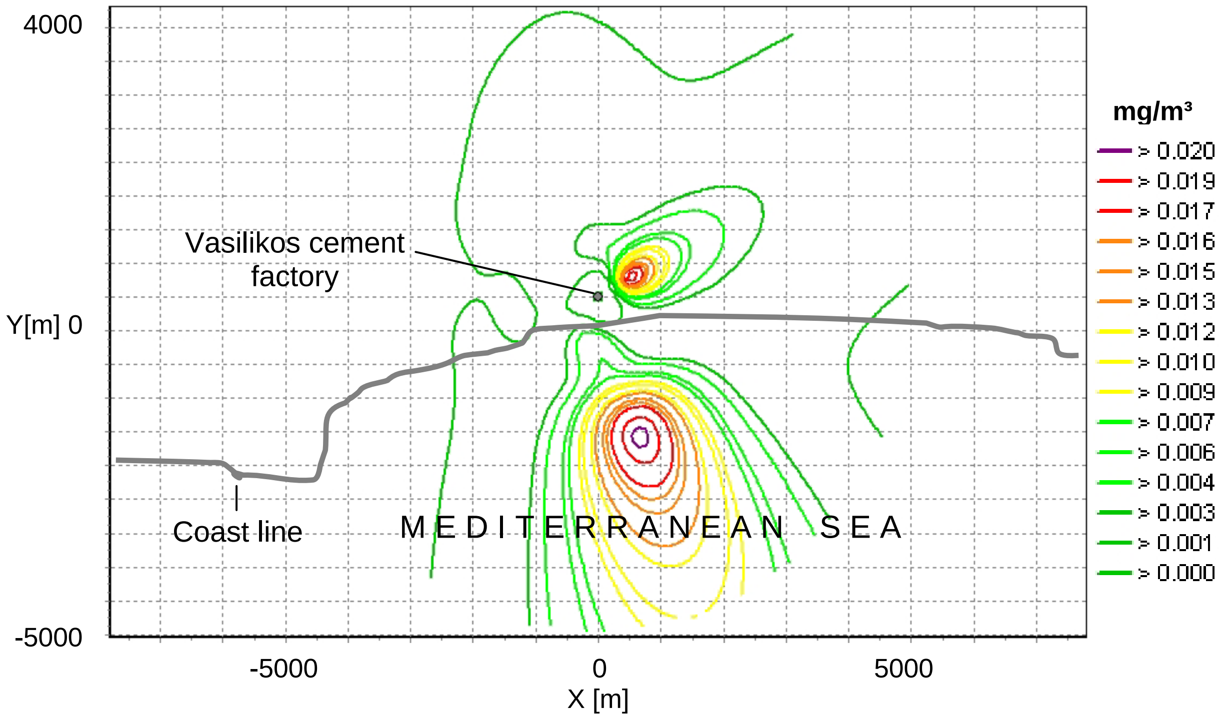


Figure 5.17. Dispersion modelling of the mean annual NO_x concentration for the cement factory near Vasilikos. Output of Gaussian dispersion model P&K 3782

The dispersion modelling results had to be converted afterwards to fit into the emissions inventory. For this purpose, all dispersion modelling results were digitised and transferred into ArcGis and then converted according to the formula:

$$C_E=(C_I*k*t)/A_{St}$$

where

C_E …	Daily emissions in kg/km²
C_I…	Pollutant distribution according to dispersion model in µg/m³
k…	Factor for the conversion of m to km → 10^6
t…	Factor for the conversion of s to day → 3600*24
A_{St}…	Stack area in m² → 6,6 m^2

It has to be emphasized, that this formula is purely a conversion calculation to perform the transition of pollutant concentrations in µg/m³ to **"distributed emissions"** in kg/km². The results were cross-validated with the known total daily emissions of the stack by summarizing the values of the distributed emissions according to the covered area. A good agreement could be observed here. This method can be considered as a way to weigh the emissions according to their source height, which was already tried by other authors, but in a more simplified way with statistical methods [58]. With the described approach, a normal emissions inventory can be enhanced significantly. In Figure 5.18 the distributed emissions of the Vassilikos cement factory are shown. All distributed emissions were finally intersected with the 1x1km grid and then included in the emissions inventory.

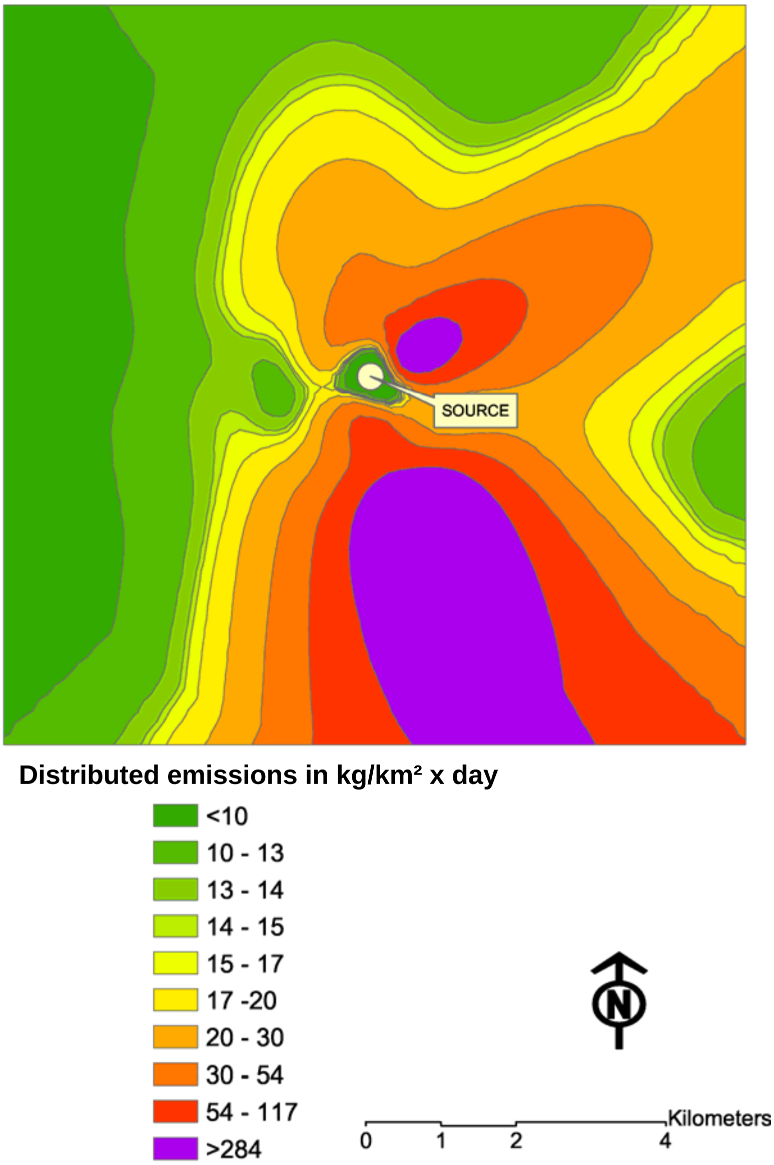


Figure 5.18. Dispersion plume of the cement factory near Vasilikos and its distributed NO_x emissions – mean annual distribution

Highways:
The major line sources in Cyprus are the highways in Cyprus. They connect the cities Nicosia, Larnaka, Limassol, Paphos, Famagusta and Kyrenia and are an important NO_x source since the main emission source is the traffic sector.
To assess the influence of the highways on their neighbourhood, an NO_2 abatement curve could be created by calculating the closest distance of each diffusive sampling point to a highway and inserting it into a scatter plot together with the measured NO_2 concentrations. A clear correlation was found out as shown in Figure 5.19.
When this curve is intersected with the background value line of 7 µg/m³, which was obtained from the site category analysis (Table 3.2), the influence distance of the highway can be determined to be 500 m in maximum. Here it is assumed that the diffusive sampling points are not influenced by other sources and that to both sides of the roads wind directions and speeds are equal. This is an acceptable approximation, since the measurements cover a whole year resulting in a good distribution of wind speeds and directions.

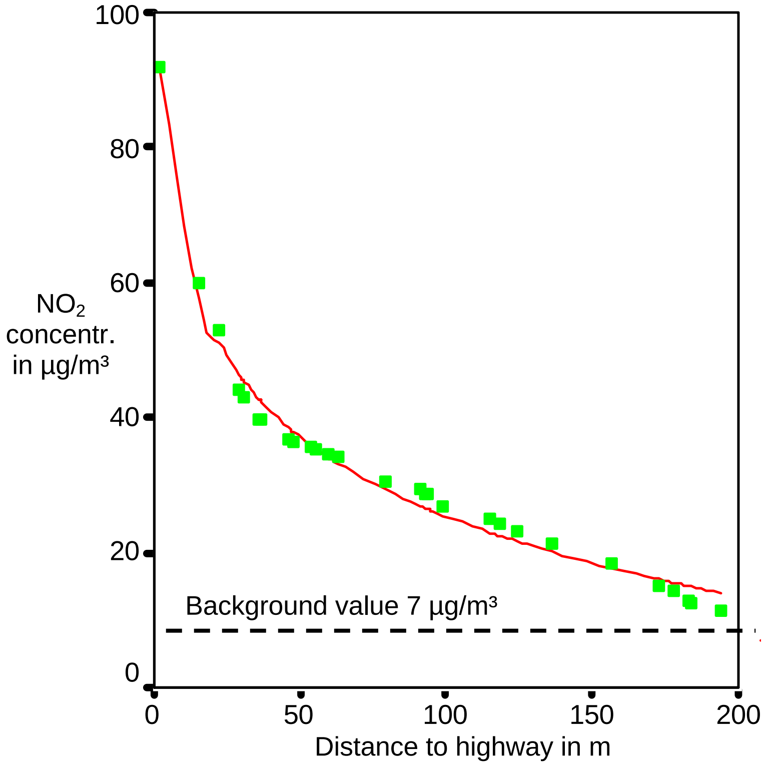


Figure 5.19. Abatement of NO_2 with increasing distance to highway

Therefore a buffer of 500 m was applied to each segment of the highways and intersected with the 1x1km grid. After this, the area of each intersected buffer was calculated and proportionally the emissions were calculated (Figure 5.20) and included as distributed emissions to the final emissions inventory.

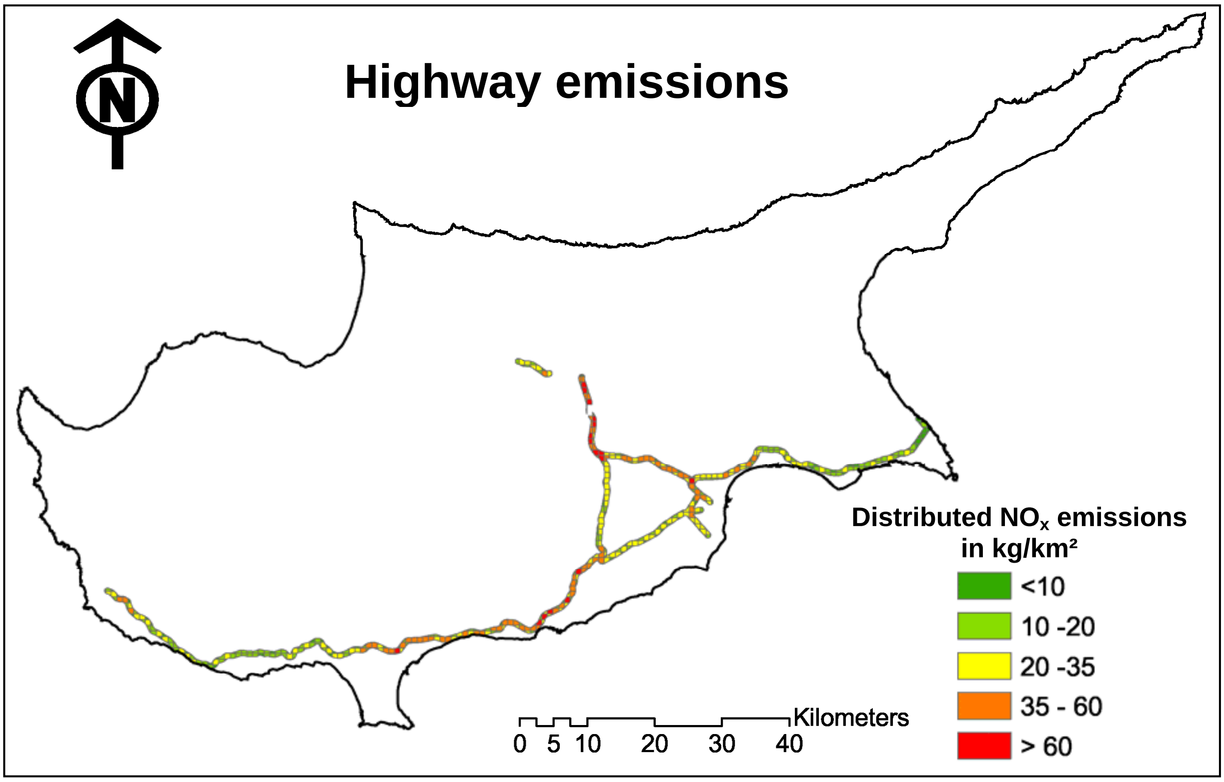


Figure 5.20. Distributed NO_x emissions of the highly frequented highways in Cyprus including their influence on the environment

Cities:

On the chosen model scale, 1x1km, cities can be treated as area sources, since all emission sources are concentrated densely at a definable area. Most of the sources in a city are negligible, but all together they form a very important source that influence greatly the vicinity. Depending on the dispersion conditions, a large plume of air pollutants leaves the city and affects the suburbs or even a larger area. With the two balloon measurement campaigns of the project, the plumes of Nicosia and Limassol could be detected [34]. To determine the actual shape of a city plume, the use of a dispersion model is usually the appropriate tool. The dense network of diffusive samplers that were exposed during the project however allows another approach. It was investigated, whether the distances of the diffusive samplers to the cities correlate with the measured NO_2 concentration. After the exclusion of locally influenced samplers, a regression analysis was carried out. The functional dependency of the two variables could be explained with a logarithmic approach (Figure 5.21). The dashed line indicates the average diffusive sampling value of agricultural points which can be considered as the background value in non urban places with a low population density: 7,0 µg/m³ NO_2. The value for mountainous areas cannot be taken, since this does not reflect the average value close to a city. The intersection between the logarithmic curve and the dashed line can be considered as the influence distance of the city: It is 12 km.

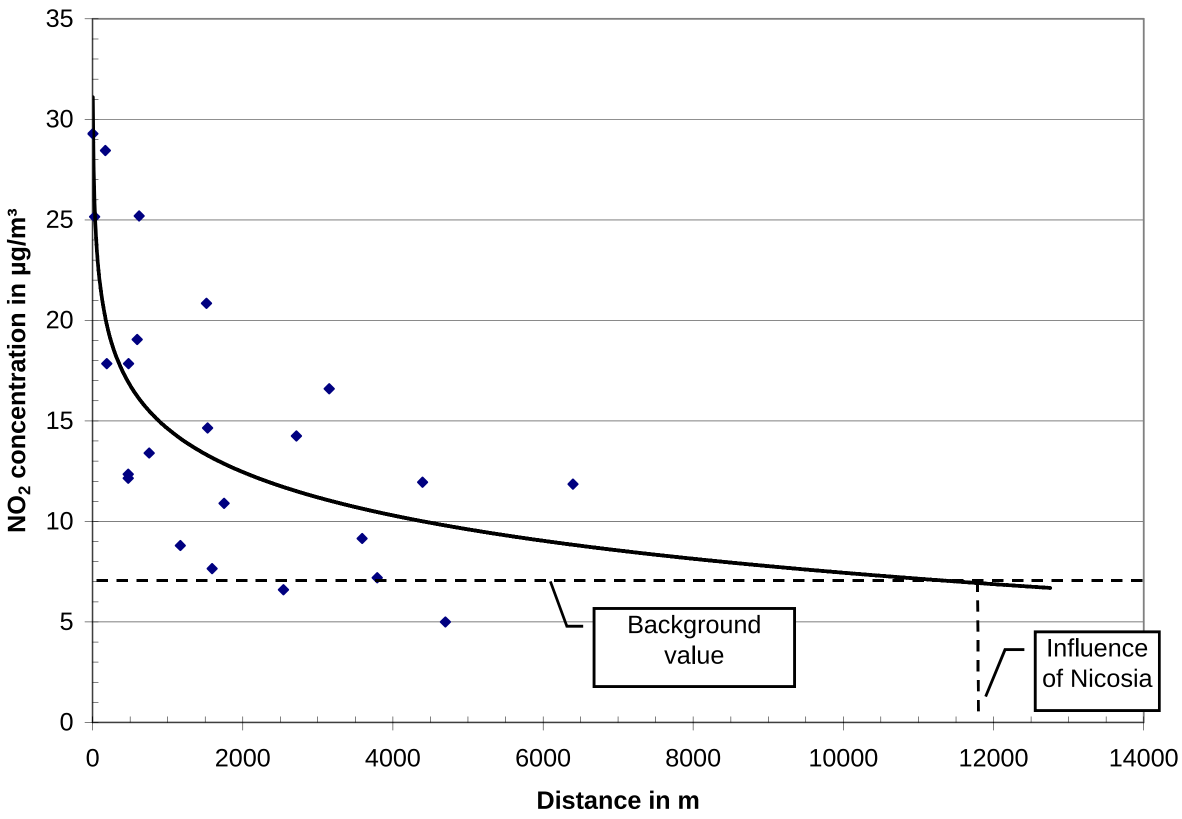


Figure 5.21. NO_2 abatement curve: Influence of Nicosia on the NO_2 concentrations in its vicinity considering all directions

To calculate the shape of the air pollutant plume, wind statistics from Athalassa were divided into 16 wind direction classes. For each class, the average wind speed was calculated and then weighted with the frequency of its occurrence. This value then was multiplied with the maximal plume length of 12 km, which finally resulted in the plume length for the regarded wind direction sector. Therefore, the calculated plume resembles the wind rose from the used wind statistics. This is an acceptable idealisation, since the result is added as “distributed emissions” to the coarse 1x1km modelling grid. Basically, the following assumptions were made:

- The shape of the plume is determined by the frequency of wind directions
- The plume length in each wind direction sector correlates with the wind speed
- The total plume length correlates also with the total emissions of the city, since less emissions are diluted faster

For the gradual abatement of the distributed emissions, the before determined logarithmic regression formula of the curve in Figure 5.21 was converted to calculate “distributed NO_x emissions” instead of NO_2 concentrations. As conversion factor, the ratio of the average NO_x emissions in kg/km² of Nicosia to the maximum NO_2 concentration was used.

In Figure 5.22, the calculated plume of distributed NO_x emissions from Nicosia is shown. Black framed squares indicate the analysis grid, black diagonal lines the frontiers of 16 wind sectors. In the city, the grid cells are without a frame, they indicate the emissions according to the emissions inventory subtracted by the distributed emissions outside Nicosia.

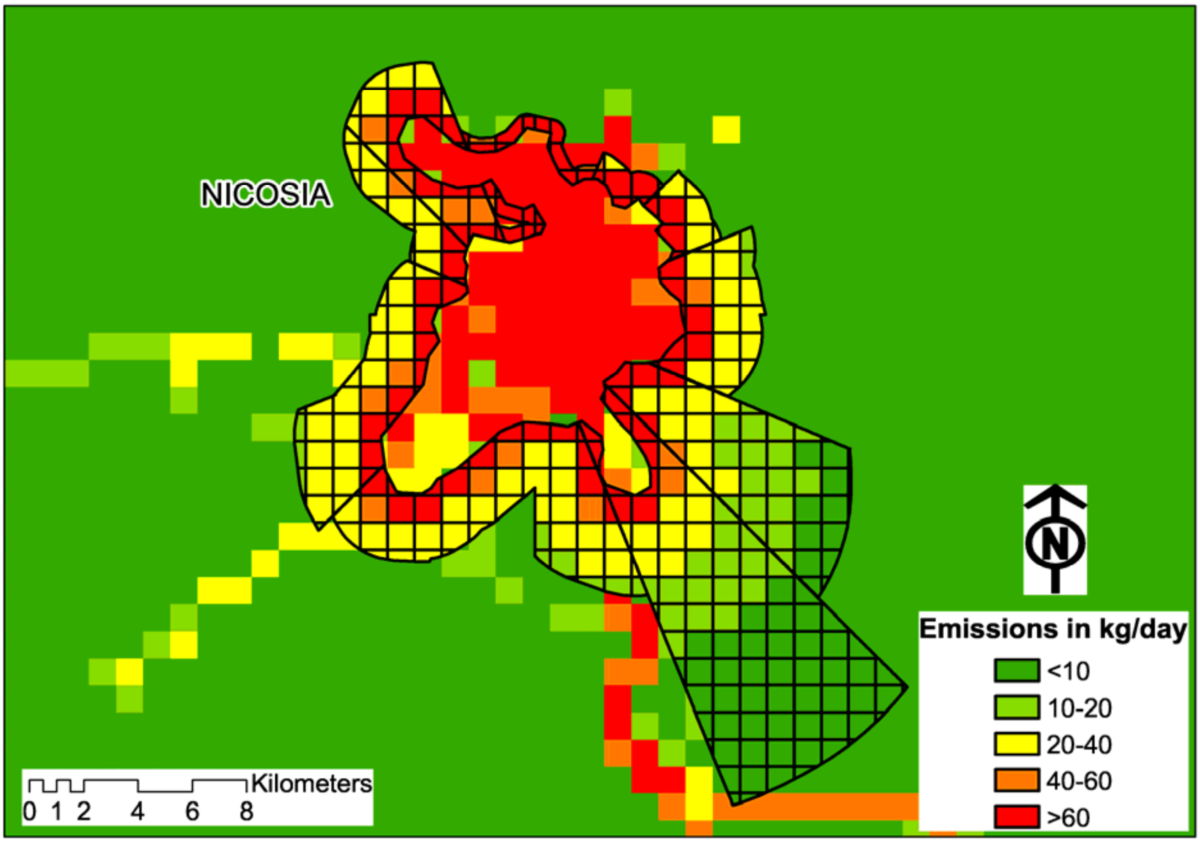


Figure 5.22. Dispersion plume of Nicosia with real NOx emissions from emissions inventory and the calculated distributed emissions of the plume

Following the above-mentioned assumptions, the other cities were treated equally (see Annex B.1. to view Limassol), with the difference that no own regression curve could be determined, since not enough diffusive samplers were available to make a secure statistical statement. Instead of this the formula from Nicosia had been fitted. The total plume length was calculated using the ratio of the summed emissions of the regarded city to Nicosia. In Figure 5.23 all calculated city plumes are summarized.

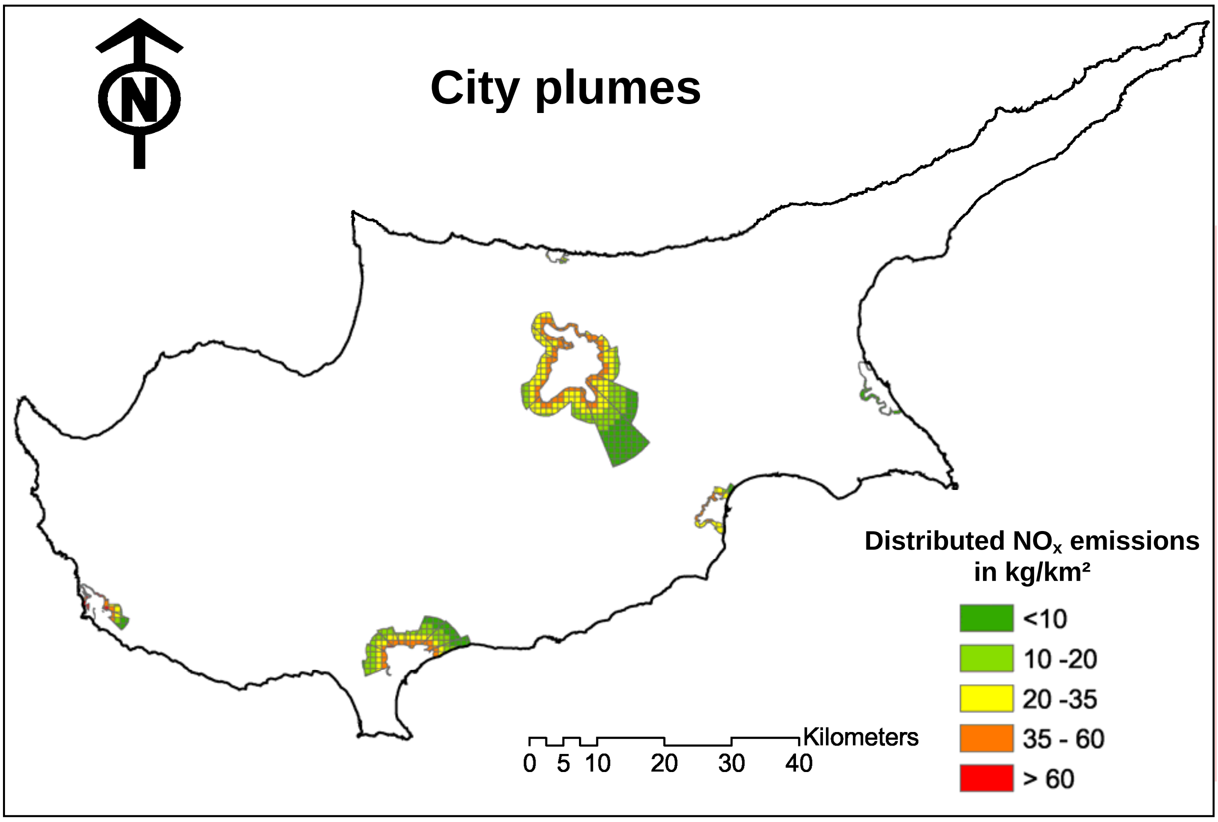


Figure 5.23. Dispersion Plumes of the large cities in Cyprus containing distributed NO_x emissions

Total NO_x emissions of Cyprus

Finally all distributed emissions were added to the accordingly reduced "normal" emissions inventory and formatted for the training with the neural network. The result is shown in Figure 5.24.

5.4.4.2 Sensitivity analysis of NO_x emissions

To assess the possible contribution of NO_x emissions to the model performance, the diffusive sampling results were compared with the NO_x emissions of the regarded grid cell. As expected, a good correlation was discovered, shown in Figure 5.25. A power curve with a coefficient of determination (r^2) of 0,62 could be fitted to the bivariate data distribution. Some outliers occur due to locally influenced diffusive samplers, which were not omitted, since the neural network should also be able to learn such cases. The origin of these outliers can be viewed in Annex B.1., where the maximum value range of diffusive samplers in grid cells with more than one sampler is shown. In some cases deviations of around 20 to 50 µg/m³ can be observed here. This creates a positive or negative deviation compared to the emission value of the regarded grid cell.

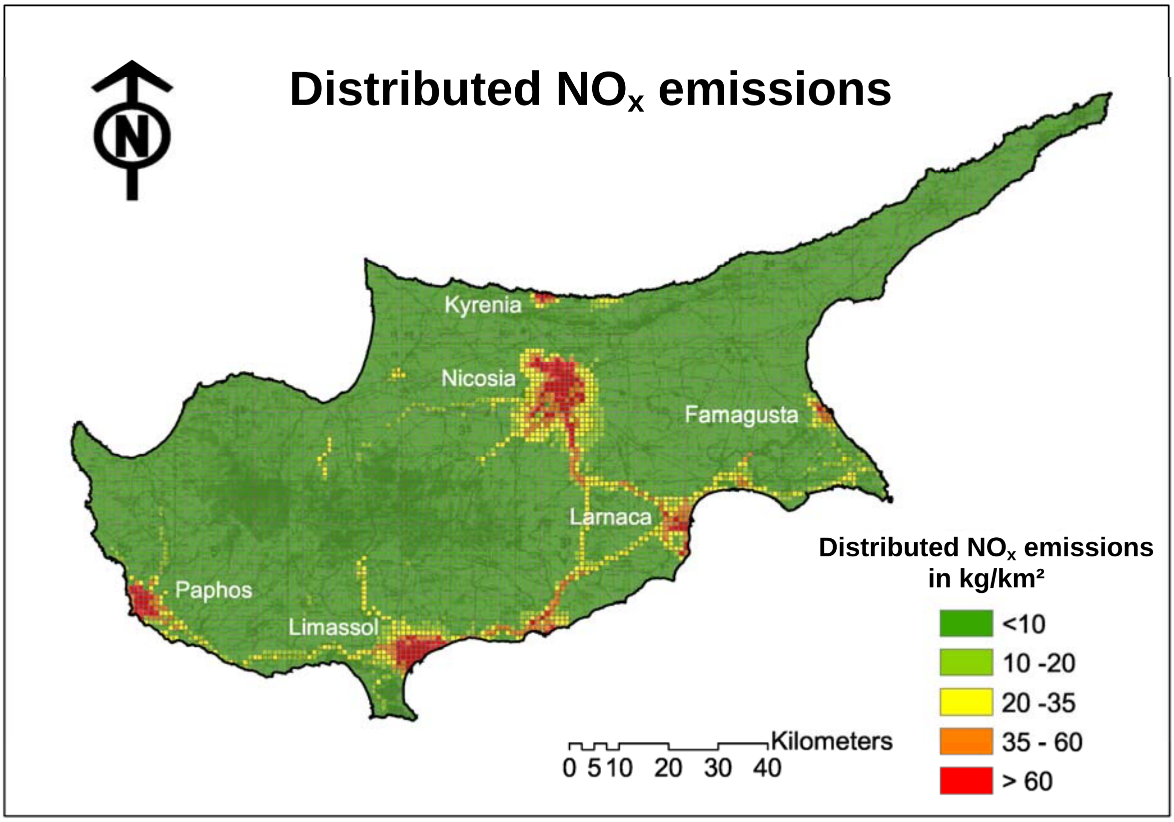


Figure 5.24. All summed NO_x emissions in Cyprus including the plumes of the major sources

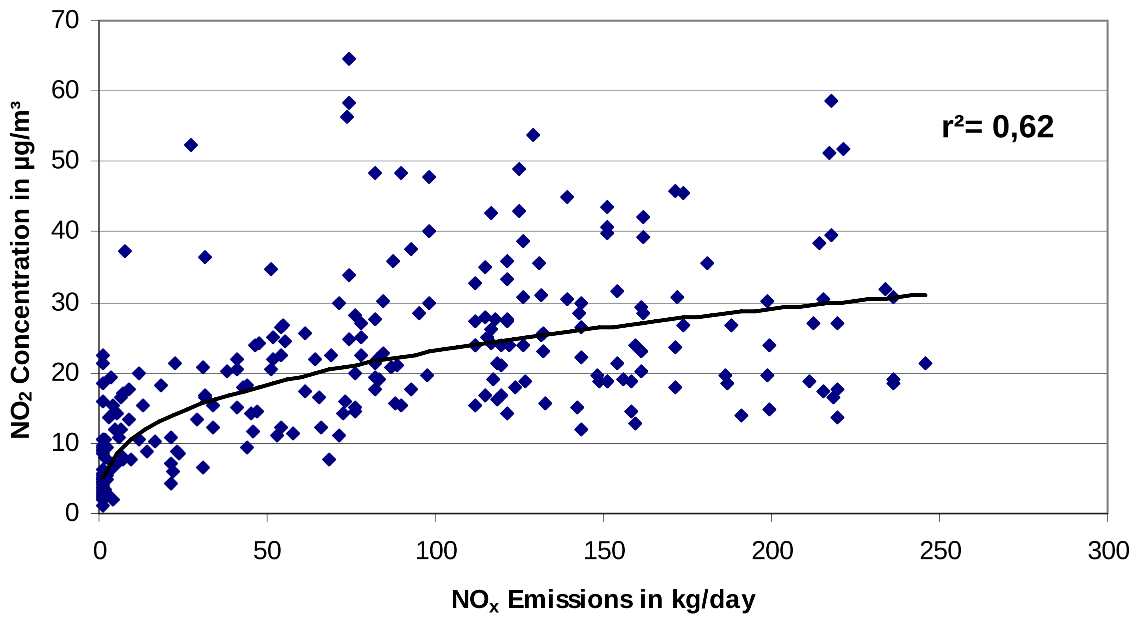


Figure 5.25. Correlation of NO_x emissions and annual NO_2 concentrations as measured by diffusive sampling

Training of the neural network with NO_x emissions

After several test runs, the optimal network topology was found to be three hidden layers with 5-3-3 neurons per layer. Figure 5.26 shows the result after recalling the trained network. All major sources and their plumes are reproduced quite well; minor sources however are hardly visible since they are not always represented in the emissions inventory. The most important streets and highways emerge clearly, except the road from Nicosia to Kyrenia, which already appears only weakly in the emissions inventory – perhaps an underestimation. Nevertheless the author did no additional adjustments since this would have only been a subjective rough estimation. East of Limassol, the cement factory and the power plant of Vasilikos cause distinct air pollution with concentrations of up to 54 µg/m³.

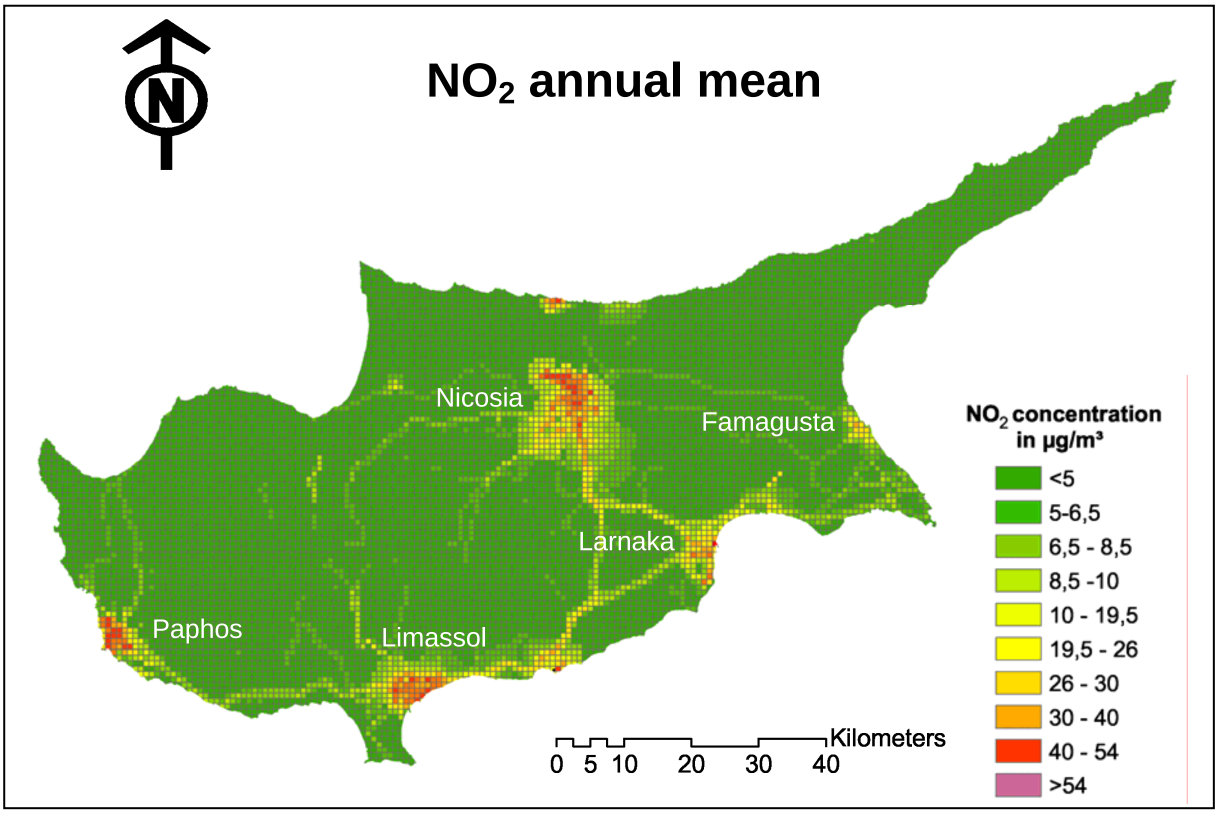


Figure 5.26. NO_2 concentrations distribution calculated by a neural network that was trained with NO_x emissions

It should be finally pointed out, that the correlation of NO_x and NO_2 theoretically cannot be perfect, since NO_x (nitrogen oxides) contains the two species NO and NO_2. But, since regarded time scale is one year, it can be assumed that all NO is ultimatively transformed into NO_2. For some single diffusive samplers near roads, there might be also a significant amount of NO in the air, which might lead to an underestimation of NO_2 by the model.

5.4.5 Population density

The training data on population density was based on the digital GIS maps from the cartographical service of Cyprus. Either centre points of all villages or polygons of the district boundaries were available. Both geometries would however blur the actual physical distribution of inhabitants, which is determined by the settlement and therefore worsen the network performance or non-existent correlations would be established. It was necessary to improve the digital population density map by digitising the actual extent of all cities and villages and assigning the population numbers to these polygons. The digitisation was based on a 1:250000 paper map from the year 2000 which is not ideal since there are many built-up areas in Cyprus, but a good approximation, since all values are again intersected with the 1x1km grid. In reality, the extent of the cities and villages should be generally larger. Actual satellite photos covering the whole island would be a solution for this problem. Figure 5.27 shows the enhanced population density map assigned to the 1x1km analysis mask.

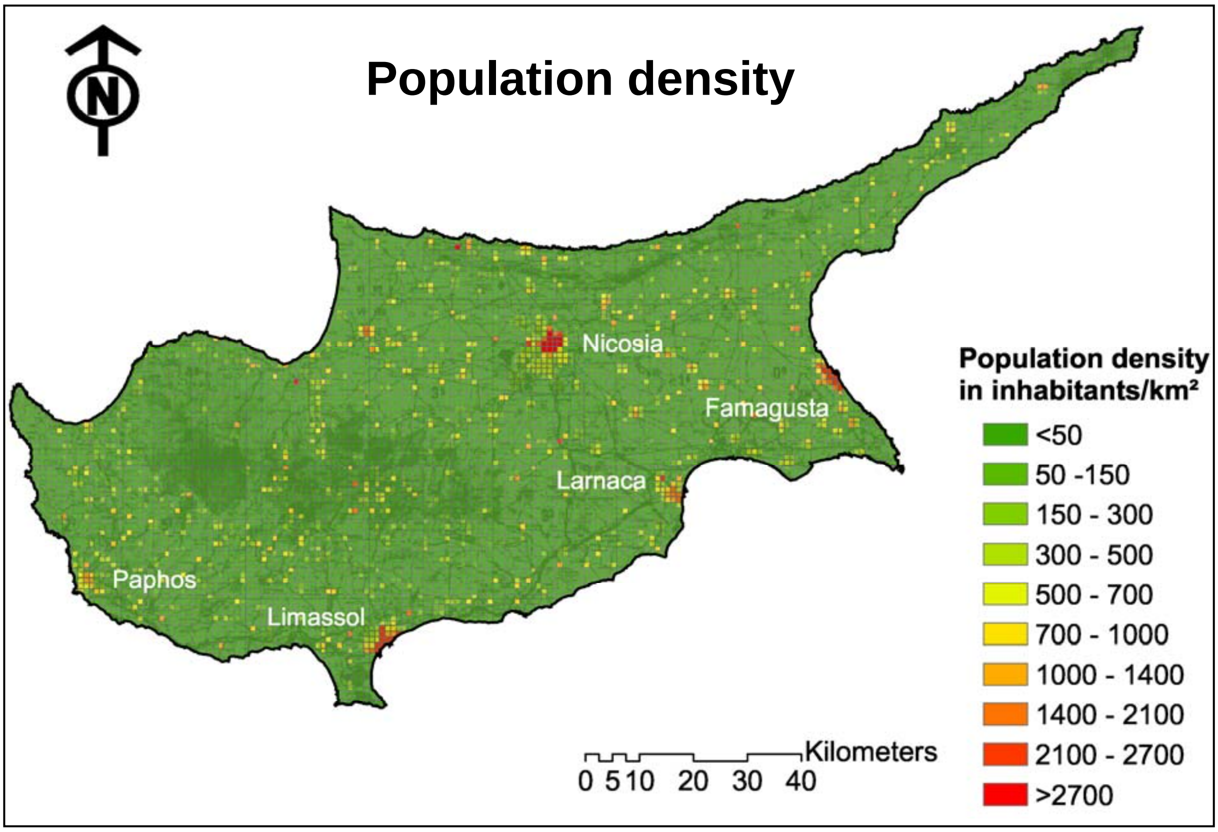


Figure 5.27. Improved population density map assigned to the 1x1km analysis grid

The population density is an indicator for human activity, which is the main cause for NO_x emissions. Unlike other air pollutants like Particulate Matter the natural sources for NO_2 are negligible, only special phenomena and events like volcanic activity or forest fires are taking place under high temperature conditions, which are necessary for NO emission. High NO_2 concentrations are therefore usually caused by humans. A high population density occurs in cities and entails an infrastructure of transport and energy whose base are combustion processes. So on the one hand, the population density correlates with the emissions inventory, but on the other hand it might reveal less tangible sources and always is an indicator for basic land usage characteristics – rural or urban, respectively anthropogenic or natural. Its potential contribution to the model performance can be regarded as a supplement and an enhancement to the emissions inventory. An assumption that is confirmed by the sensitivity analysis of this variable, depicted in Figure 5.28: The scatter plot with the diffusive sampling results on the y-axis and the population density on the x-axis shows on the first view a weak correlation with a low gradient of the logarithmic regression curve. On a closer view however it becomes clear that the general tendencies are reproduced quite well: NO_2 concentrations smaller than 10 µg/m³ only occur at a population density below approx. 1300 inhabitants/km². A reverse conclusion cannot be drawn as the highest NO_2 concentration is measured at a less crowded place of about 600 inhabitants/km². The reasons for this are local emission sources that influence this diffusive sampling point, which cannot be reflected by the population density but might be compensated by the emissions inventory. A univariate network training here wasn't necessary, since the previous development steps allowed a forecast what would have been the result for this – a map where at populated places the predicted NO_2 concentrations ascent proportionally to the population density, but due to the distribution pattern in the sensitivity analysis, the model would not be able to reproduce high concentrations. A significant model enhancement only can be expected in a multivariate environment.

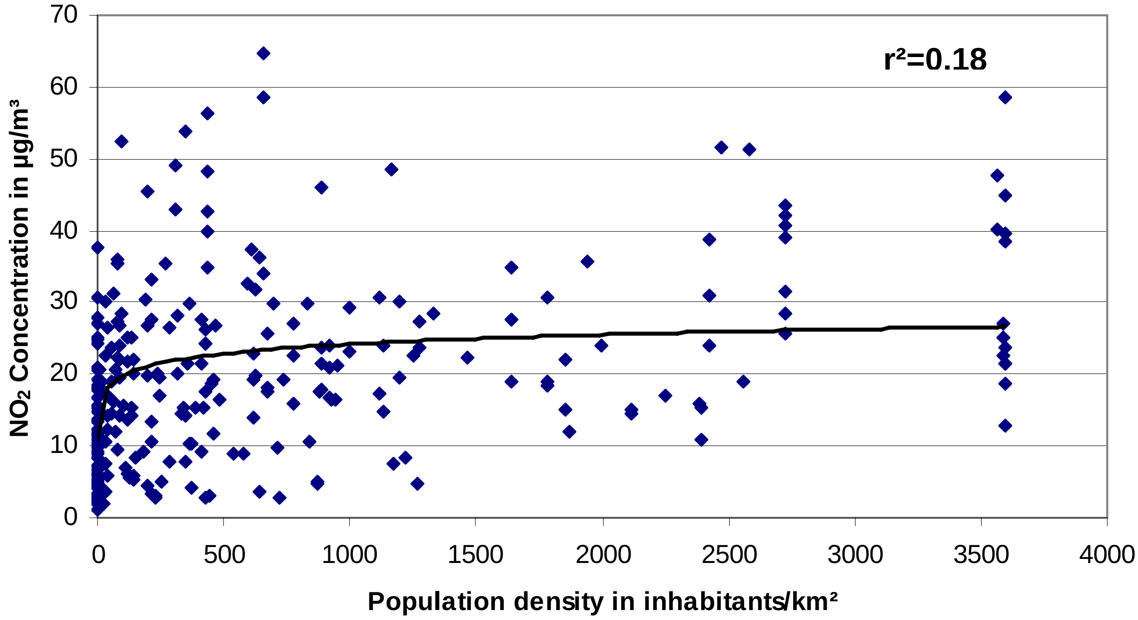


Figure 5.28. Correlation of population density and annual NO_2 concentrations as measured by diffusive sampling

5.4.6 Land use

Land usage is an ordinal variable that cannot be quantified. It can be considered as a qualitative description of a certain feature. As all diffusive sampling sites were assigned to different site categories, a useful step would be to assign these categories to a land usage map. Unfortunately, the only available maps were derived from Landsat satellite data, which turned out to be too coarse, and its classification was useless for modelling purposes. To prepare an own land usage map would have gone beyond the scope of this work. Nevertheless it was tried to apply land usage in a simplified form as so called dummy variables "urban/rural" or by providing the actual average NO_2 value of urban and rural diffusive sampling sites. A level shift of the model output was expected but did not occur. Therefore, land use was not directly included in the model.

5.4.7 Combination of input variables (multivariate modelling)

A univariate model alone cannot precisely predict the NO_2 distribution, as there is more than one influencing factor (see Figure 3.9). A good model performance could however already be established with the emissions inventory as input. This positive result can be explained by two reasons:

- It was not only a "normal" emissions inventory, since the dispersion of the most important sources in Cyprus was calculated. The model therefore can be considered as multivariate with the additional variables wind speed, wind direction and atmospheric stability
- The emissions are the source for the distribution of NO_2 and are therefore the ideal training variable

The weak point of this model is the underestimation of peak values and the neglect of many small sources that are hard to cover with a survey. The problem can be only compensated by considering additional information. In the previous chapters, the potential contribution to the model by the different available input variables has been discussed. Each of them correlates directly or indirectly with NO_2 concentrations and should for this reason improve the quality of the calculations. On the other hand it was also found out that an existing correlation does not necessarily create a realistic result, which is the case for the height above sea level. In addition it is not clear, how the network responds to the combination of different variables.

A lot of effort was therefore put into this development step. In the following, different variable combinations for the training of the neural network are presented.

5.4.7.1 Emissions inventory and UTM coordinates

A neural network with three hidden layers was trained with the UTM coordinated and the emissions inventory, before it's improvement with the dispersion plumes of the major emission sources. The visualisation of the result reveals the problems of this model configuration: an unrealistic ellipsoid shaped NO_2 plume at about 7 µg/m³ coming from the Southeast of Cyprus (Figure 5.29). Similar results always occurred, when the UTM variables were included. A possible explanation for this effect is that the neural network recognises high pollutant loads at the south coast of Cyprus, especially between Limassol and Dhekelia, where the major single sources and a part of the highways are located. In numbers, this can be expressed with UTM Northing values below 3870000 and Easting values between 490000 and 590000. The ellipsoid shape of the plume results most possibly from a generalized neural interpolation. In other words, the network generalizes the correlation of high NO_2 concentrations in the above-mentioned coordinate range and at the same time it interpolates to the lower polluted Northern part of Cyprus and the Troodos mountains. The emissions of Nicosia are obviously not included in this generalisation, as the sum of the emissions at the south coast is higher and weighted stronger – the neural interpolation only touches the Southern part of Nicosia.

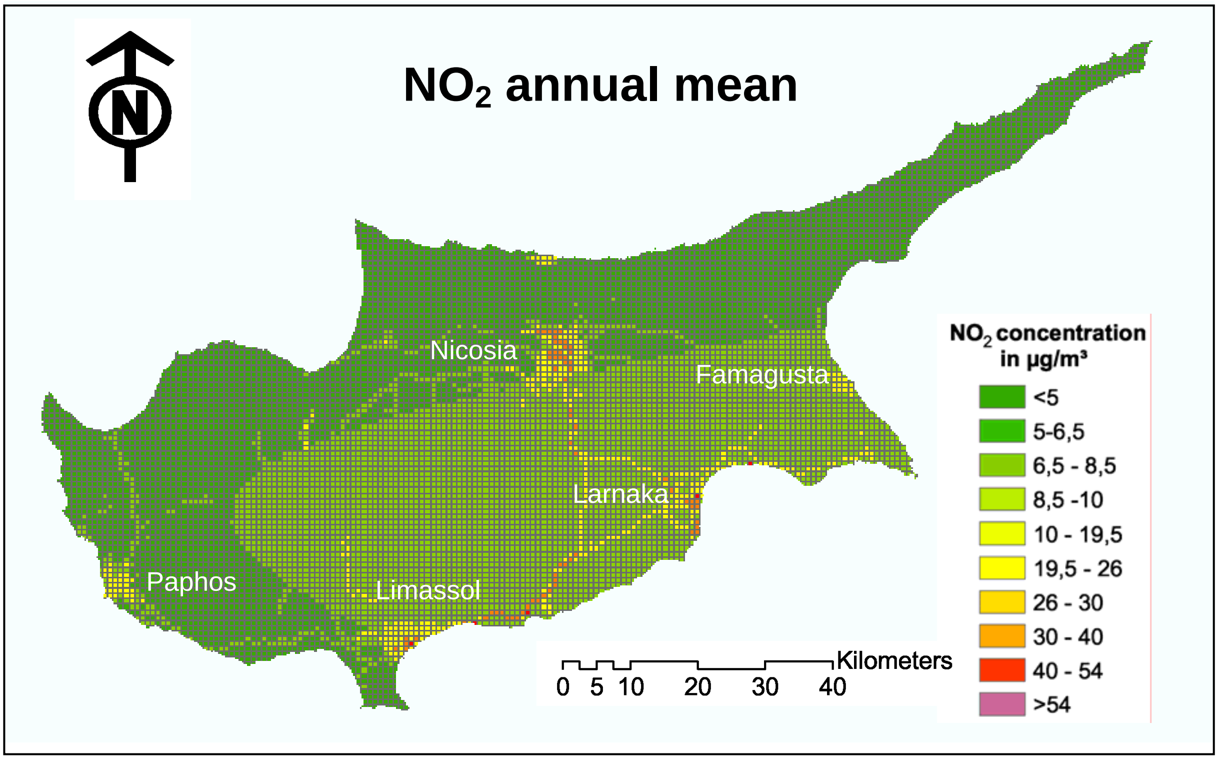


Figure 5.29. NO_2 distribution as calculated by neural network trained with NO_x emissions and UTM coordinates

5.4.7.2 Emissions inventory, UTM coordinates and population density

Adding the population density to the before described model, the output becomes more realistic although the plume at the Southwest still is there (Figure 5.30). All attempts to include the UTM coordinates in a multivariate model created similar unrealistic plume shapes. A fine structured neural interpolation seems to be only possible with a univariate model. It is remarkable that the major cities and the single sources are reproduced quite well. Generally, the network respond to the emissions inventory was in all tested models very good.

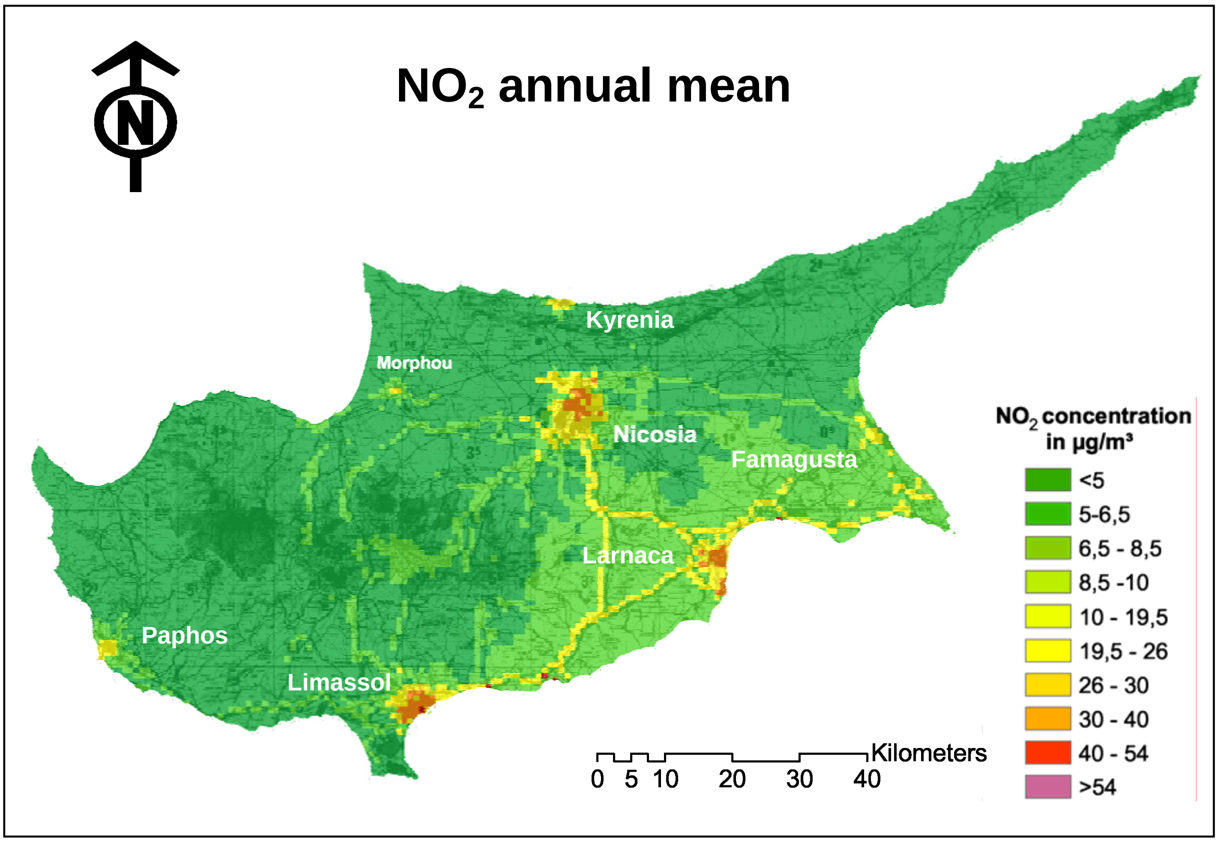


Figure 5.30. NO_2 distribution as calculated by a neural network trained with coordinates, population density and NO_x emissions as input variables

5.4.7.3 Emissions inventory, population density and height above sea level

The univariate training with the height above sea level produced physically impossible results as clearly shown in Figure 5.14. This is however not necessarily also true for a multivariate model since the neural network may learn the fact that high NO_2 concentrations do not automatically occur at a low altitude. The possibility is given here to consider additional information like the emissions. Yet, the recall of the trained neural network with three hidden layers at three neurons each still calculated an unrealistic map as shown in Figure 5.31. Clear structures cannot be found, the streets disappear and the city of Limassol is completely underestimated, even Famagusta seems to be more polluted. Other attempts to apply the altitude were also unsuccessful.

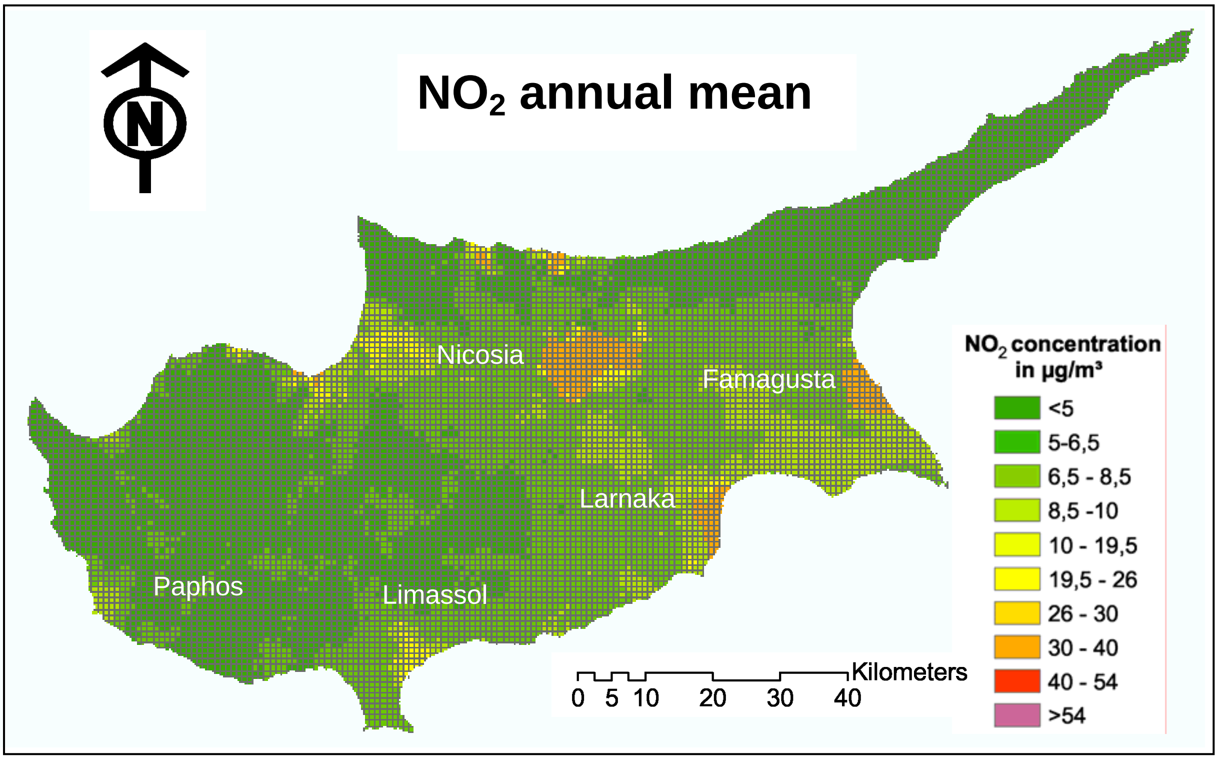


Figure 5.31. NO_2 distribution as calculated by a neural network, with height above sea level, population density and NO_x emissions as input variables

5.4.7.4 Emissions inventory and population density

From the above-described modelling progress the following conclusions can be drawn:

1. A univariate model cannot fully explain the spatial distribution of NO_2, since there are more than one influencing factors.
2. The improved emissions inventory is potentially the most important input variable.
3. The population density might be a supplementation to the emissions inventory.
4. Neural interpolation with the UTM coordinates only leads in a univariate environment to useful results.
5. With the height above sea level as input variable, the neural network calculates physically incorrect NO_2 distribution maps.

Consequently, the next step is to train a neural network with the improved versions of the emissions inventory and the population density map.

After several tests runs to find the best network topology, a model with three hidden layers and 3 – 5 – 3 neurons per layer was finally chosen. Further training parameters were:

- 500000 iterations in total
- During the first 10000 iterations, the learn rate control was switched off and eta was set to 0,01
- The momentum factor (learning coefficient) was set to 0,8
- The weight update mode was set to process all training patterns per cycle (offline training)

In Figure 5.32 the error progress of the test set during training as provided by the neural network software is depicted. Since the input data are normalised prior to training and the user has no direct influence on the internal error assessment tools of the software, the normalized average root mean square error is shown. During the first training iterations the error decreases only very slowly due to the fact, that the learn rate control was switched off and the learning rate coefficient eta was chosen quite low at 0,01. After 10.000 iterations, the learn rate control was turned on and the network learned at a much faster pace driving the error to a first local minimum. A second local minimum occurs at 35.000 iterations and the error rises again afterwards which could lead to the assumption that the training is finished, but again the curve descents to reach the third minimum at 80.000 iterations. 300.000 iterations later, a light error decline seems to occur, but this could not be reproduced with a new run and with learn rate control feature of the software turned on it is unlikely that the error could reach an equal minimum. During the whole development process, overtraining always started latest after 100.000 iterations.

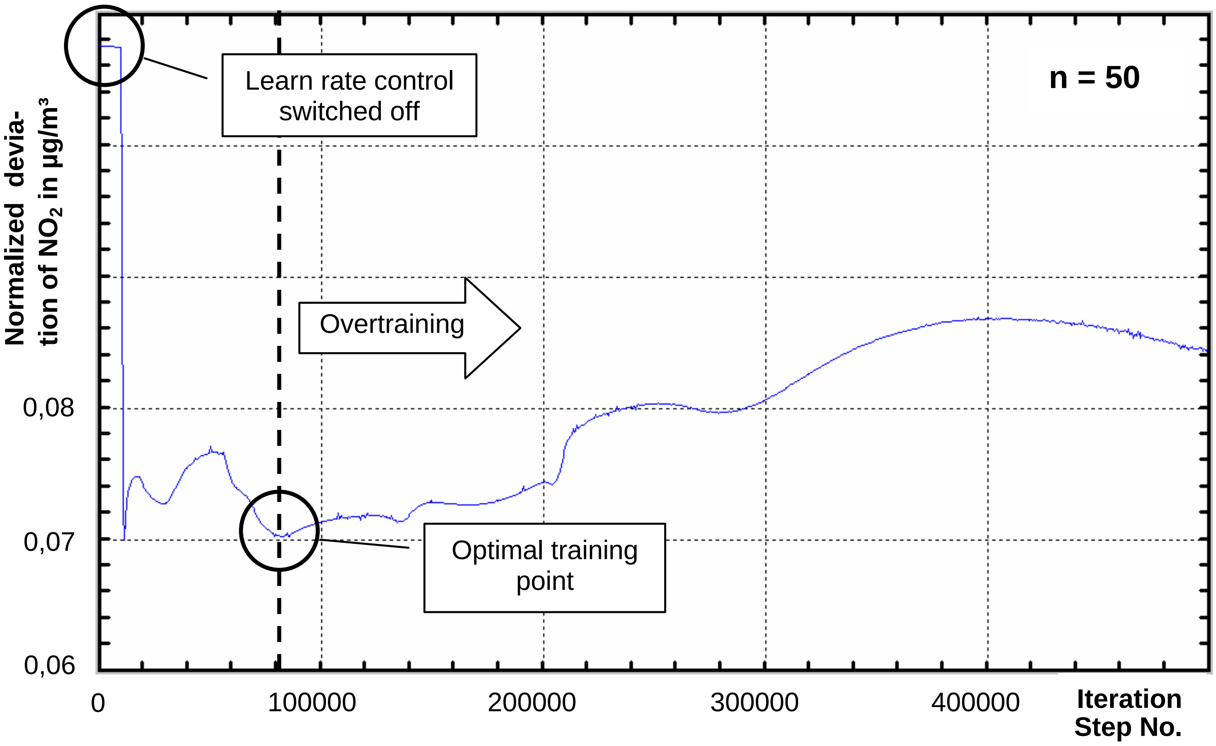


Figure 5.32. Progress of the network output deviation (normalized NO_2 values) related to the test set during training with NO_x emissions and population density as input variables

The error progress of the training set (Figure 5.33) shows a similar behaviour: After 210000 iterations, the error declines gradually and progresses almost parallel to the x-axis. A continuation of the training would bring no significant benefit. Considering the fact, that after 100000 iterations the network tends to memorize the input values, the training should be halted exactly at this point.

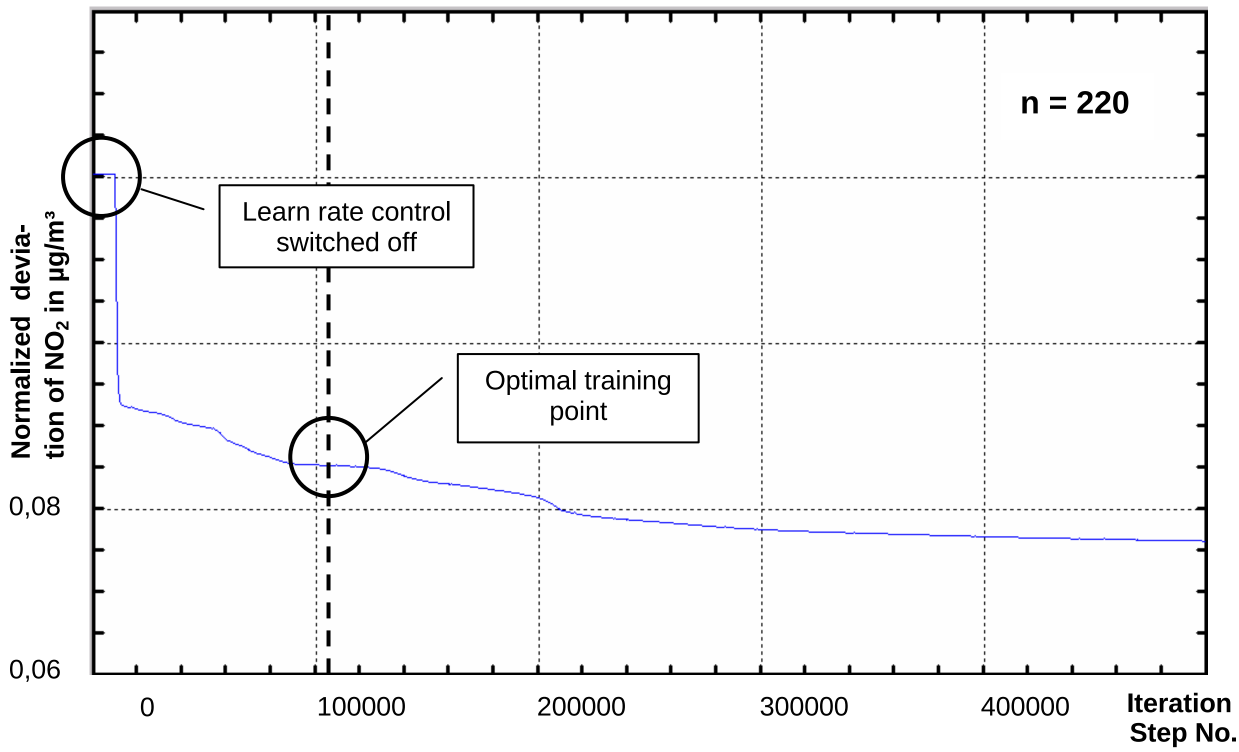


Figure 5.33. Progress of the network output deviation (normalized NO_2 values) related to the training set during training with NO_x emissions and population density as input variables

To assess the topology of the neural network, the hidden node analyser of the model is depicted in Table 5.4. The hidden node analyser quantifies the average contribution of each neuron to the next layer in other words it describes the activity of a neuron.

Table 5.4. Hidden node analyser after training the neural network with NO_x emissions and population density

Average Hidden Node Contribution to Next Layer
Network Name: training coord emi hi pop
Iterations: 100000

H.Layer	Node	Percent Contribution
1	1	35.81
1	2	28.78
1	3	35.41
2	1	29.60
2	2	8.50
2	3	11.39
2	4	27.42
2	5	23.09
3	1	15.00
3	2	31.92
3	3	53.08

None of the nodes contributes less than 8,5% to the layer's output signals over all training patterns. This indicates a well-configured network topology - the lowest contribution values belong to the second hidden layer with 5 neurons, which means a theoretical ideal contribution of 20%.

The relative importance of the two input nodes to the network response could be determined by the input node interrogator of the software. As expected, the emissions contributed with 66,93% most to the networks output, the population density 33,07% - values that justify the implementation of both variables in the model.

Comparing the calculated network outputs with the measured outputs excluding the test data set leads to Figure 5.34. The theoretical optimal agreement of the model is indicated with a dashed line. Below a measured NO_2 concentration of 50 µg/m³, the model calculations are very close to reality, the dots scatter continuously along the dashed line with a light tendency to overestimation. Two thick black lines indicate the EU annual limit value of 54 µg/m³ as laid down in the daughter guideline 1999/30/EC of the framework directive 96/62/EC [59]. Other authors experienced similar problems with their models [19]. Above measured 54 µg/m³, the model significantly underestimates three of four values marked grey. Again the reason for this can be found in exceptional influences of single sources that cannot be reproduced by the network. All other calculations agree very well with the measurements resulting in a Pearson correlation coefficient of 0,75.

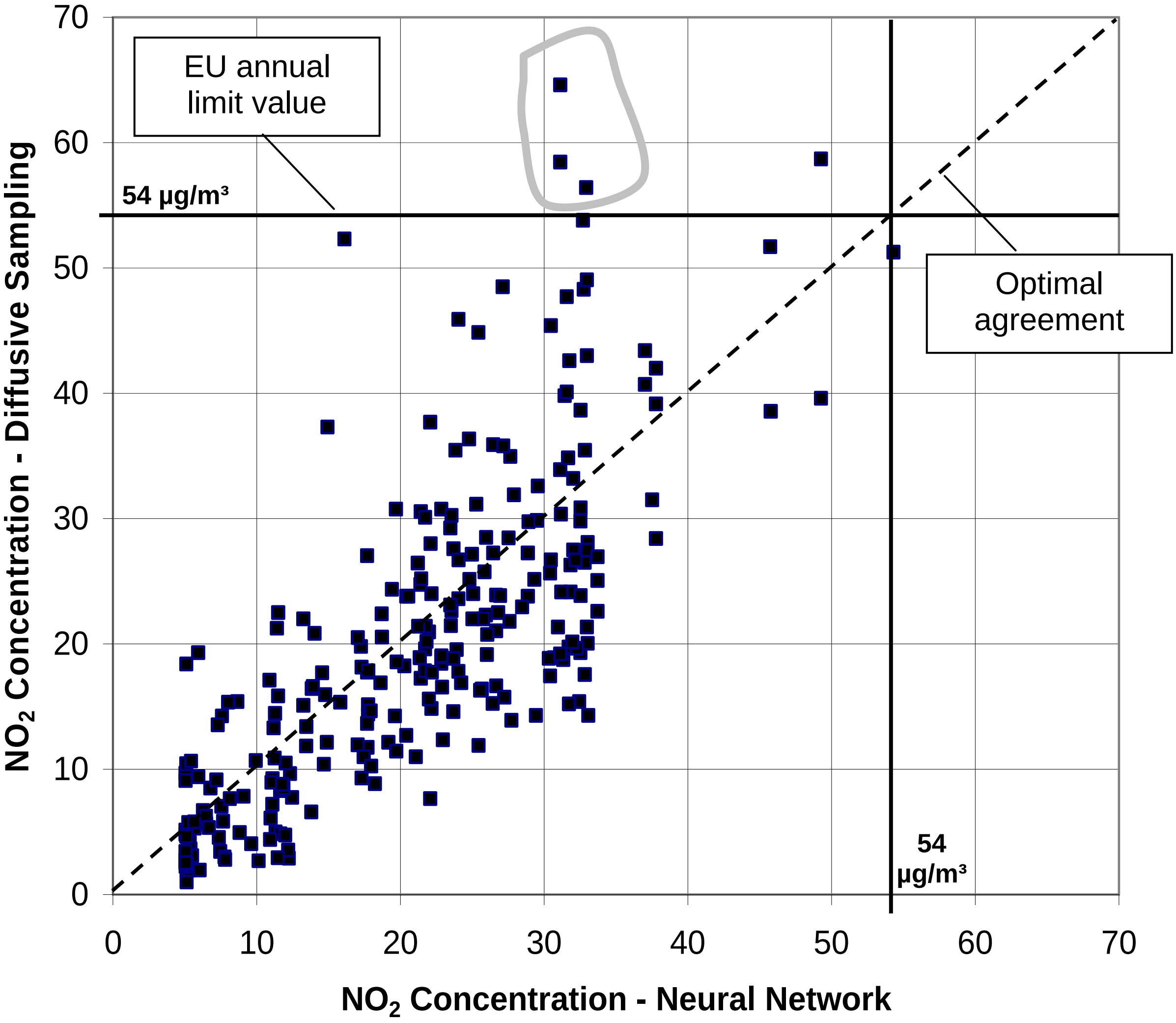


Figure 5.34. Quality assurance of the final model – comparison of NO_2 diffusive sampling results from training data set and modelled NO_2 values with n =220

To confirm this good result and exclude memorization, the same procedure was carried out for the test data set, the diffusive sampler results that were not used for network training. These test values were selected randomly, which is common practice in statistics [e.g. 32, 47]. The scatter plot in Figure 5.35 shows similar results as the comparison of the training data set: A slight scatter around the dashed line of optimal agreement and two outliers stemming from exceptional values from locally influenced diffusive sampling sites. The correlation coefficient could be determined as 0,62. These observations are in good agreement with the ones in Figure 5.34 – the model can therefore be regarded as statistically robust.

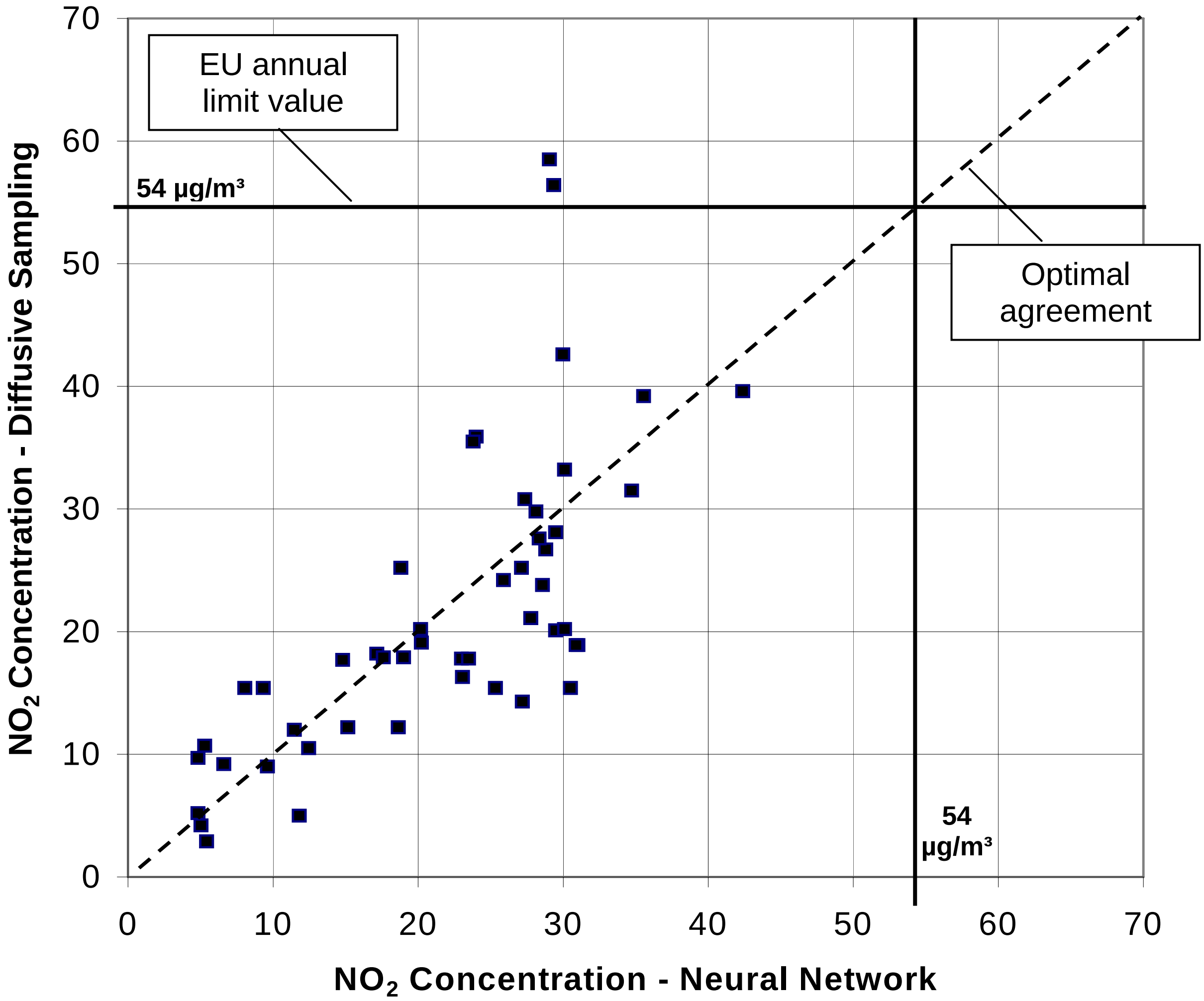


Figure 5.35. Quality assurance of the final model – comparison of NO_2 diffusive sampling result from the test data set and modelled NO_2 values with n = 50

Nevertheless, the goal of air quality assessment is to react on limit value exceedances and a model should enable the user to correctly predict high NO_2 concentrations for this purpose. It was therefore tried to increase the network response on such high values by providing weighting factors for the training process. With these factors, the modeller can advise the network to emphasize on certain training cases. The disadvantages of this method are obvious: The choice of a factor is subjective, time-consuming and hardly reproducible for other users. All attempts to increase the accuracy through factors failed, as Figure 5.36 makes clear.

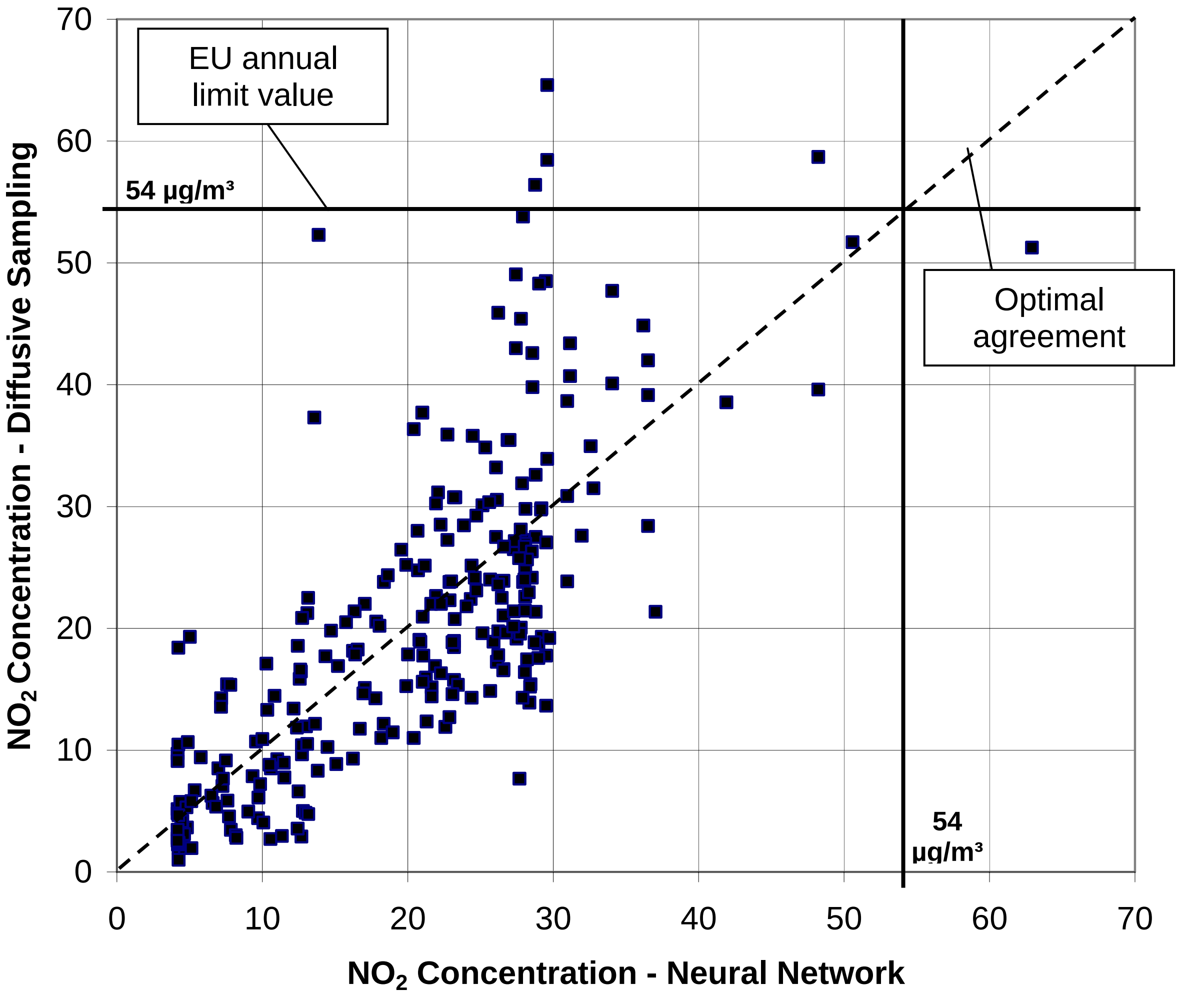


Figure 5.36. Quality assurance of the final model, enhanced with case weightings – comparison of NO_2 diffusive sampling result and modelled NO_2 values with n = 220

In spite of the implementation of weighting factors during training, measured values above 54 µg/m³ remain mostly underestimated. Only a slight improvement of the root mean square error and a raise of the correlation coefficient to 0,76 could be reached as shown in Figure 5.36. This doesn't justify the extra work and the acceptance of the other, above-mentioned disadvantages.

The recall of the trained network with the variables distributed emissions and population density finally resulted in the NO_2 distribution map, shown in Figure 5.37. A remarkable feature of this map is that the major emission sources and their plumes are accurately reproduced in terms of their shape, geographical position and strength. Thus the cities of Nicosia, Limassol, Paphos, Larnaka, Kyrenia and Famagusta appear very clear. The highest NO_2 concentrations above 40 µg/m³ (the future EU limit value valid in 2010) occur in Nicosia and Limassol. Famagusta is the least polluted city in this row.

Black circles in Figure 5.37 indicate the major point sources of Cyprus - four power plants and one cement factory. Each of them is depicted exactly according to their geographic location, their strength and their calculated dispersion plumes. Above all, the cement factory of Vasilikos and the power plants of Moni and Vasilikos, marked by an ellipsoid, have a significant influence on their vicinity that can be compared thoroughly with a city like Famagusta. Significant NO_2 concentrations around 30 µg/m³ are caused by these few sources.

Another important feature in the map are the highways in the southern part of Cyprus. In Nicosia, the highway starts as a thick yellow diffusive line, blurred by other sources and proceeds southwards as a clear line. About 10 km outside Nicosia it splits into two parts, one that leads directly to Larnaka city and the airport and another one to Limassol. From Larnaka, another highway proceeds to Limassol unifying with the other one, after which the two power plants of Vasilikos and Moni and the cement factory of Vasilikos blur the highway. The last part of the highway is the connection of Limassol with Paphos that is less frequented and causes lower NO_2 concentrations. More light green occurs here, whereas the average concentration around the other highways is about 26 µg/m³ up to 30 µg/m³. The influence of the highways might seem to be overestimated as the abatement curve (Figure 5.19) is running out after several hundred meters. This blur effect stems from the coarse analysis grid of 1x1km. Other important line sources that can be found on the map are the connections to the holiday resorts, Polis and Aya Napa and also the most important roads to the Troodos mountains West of Nicosia and North of Limassol. In the North of Cyprus, the line sources are hardly recognisable, although there are indeed strongly frequented roads – a problem that already occurred in the emissions inventory and therefore can only be eliminated by correcting the emission data. Nevertheless, the connections of Nicosia with the cities of Famagusta and Kyrenia can be seen as weak, light green lines.

Single or agglomerations of light green and yellow dots are distributed all over the air pollution map. They reflect smaller and bigger villages and match the population density map for which the author digitised the actual geographical extent of all villages. So, these dots indicate the location and the NO_2 pollution level of these villages ranging from 5 to 26 µg/m³. Large gaps with wide green areas of low NO_2 pollution can be found in the Troodos Mountains at hardly accessible areas with loose surface roads. There are no emissions and the pollutants of the cities hardly reach these areas. So, there are only background concentrations.

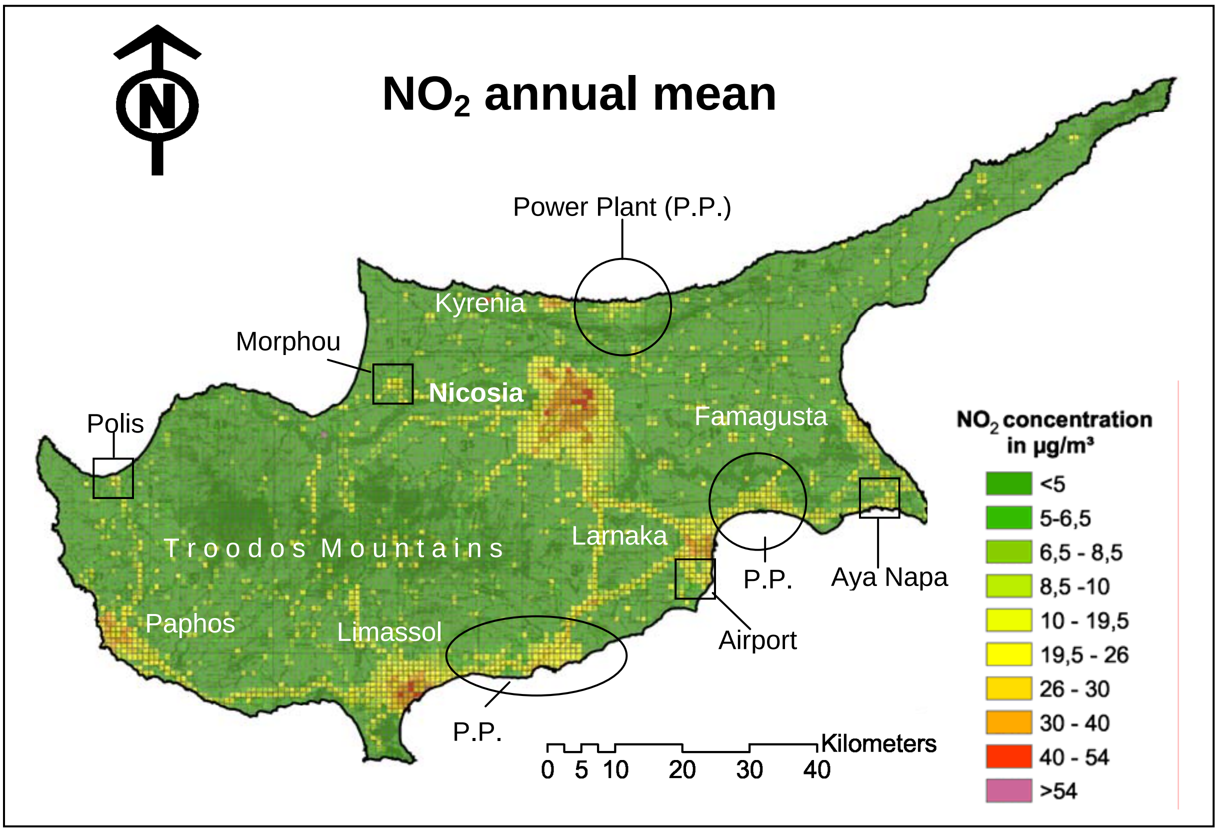


Figure 5.37. Mean annual NO_2 distribution in Cyprus as calculated by a neural network that was trained with NO_x emissions and population density

Other sources worth mentioning are the airport of Larnaka and the village of Morphou.

The three dimensional view of the result map is depicted in Figure 5.38. It underlines the observations described above: A distinct accumulation of high NO_2 pollution at the south coast and in the plane of Messaoria with Nicosia in its centre. Green areas can be found in the Troodos Mountains and in the Northern part of Cyprus. Due to technical reasons, the grid lines of the analysis cells have been omitted and contrast and brightness have been changed to better recognise the three dimensional view. That is why the line features appear more clearly as in the same result map shown in Figure 5.37 and the colours seem to be brighter.

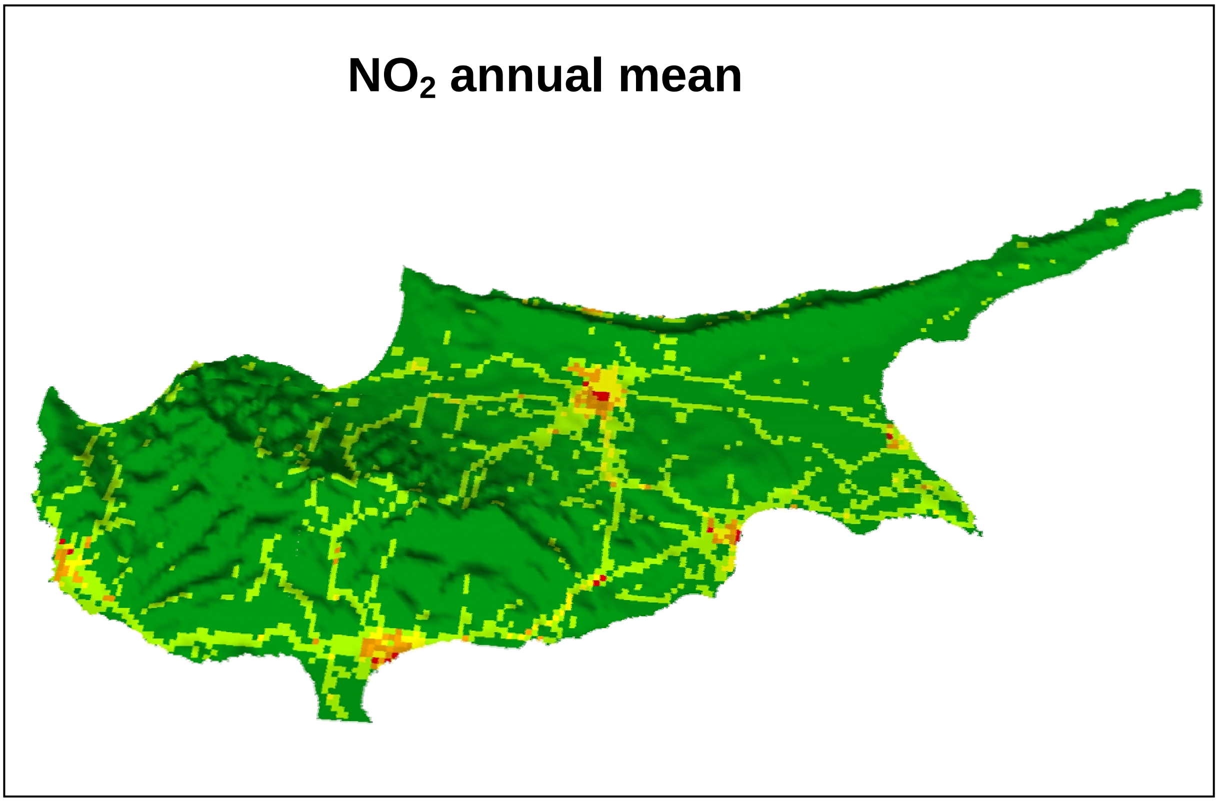


Figure 5.38. Three dimensional depiction of the final modelling result of the annual NO_2 distribution in Cyprus

Based on these maps and the statistical analysis described above, it can be concluded that with the neural network model a very realistic map of the NO_2 distribution over Cyprus can be produced, where the sources and their spatial influences are reflected precisely.

This observation is valid for the regarded regional scale. It is also of great interest how the model performs on a smaller, local scale. A closer view of the result map is therefore presented in Figure 5.39. Here, the modelling result for Nicosia is laid over the topographic map of Cyprus. Dark shapes in the centre of the map indicate settlement; white and black lines mainly symbolize roads. Noticeable are orange to red colours that cover almost the whole area of Nicosia, indicating NO_2 concentrations higher than 30 µg/m³. Top-level NO_2 concentrations above 40 µg/m³ can be observed in the commercial area. The reason for this can be found in high traffic loads and dense settlement towards the city centre of Nicosia. Thus, towards the vicinity the colours abate slowly to yellow and green and therefore to lower concentrations. An important factor for high concentrations is the dense settlement that causes unfavourable dispersion conditions with bad ventilation. In the old city centre, marked with a dashed circle, no traffic is allowed on the side of the Greek Cypriot Community, which is why the NO_2 concentrations remain below 40 µg/m³. Thus, the direct affect of pedestrian areas on the pollution level can be demonstrated.

In the south east of the city, the recreation place Athalassa Park is located with no emission sources, but a still significant pollution level of 26 µg/m³, which agrees with the average value for residential areas. Close to Athalassa, the industrial area of Latsia and some important roads are located that cause together NO_2 concentrations of 30 µg/m³ and more.

Another industrial area with even a higher loading (red colour) can be found in the North. A feeder road that connects Nicosia with the Troodos mountains and the suburbs comes from the western corner of the map and emerges clearly with concentrations of up to 19,5 µg/m³, whereas the highway to Famagusta appears only weakly due to the low values of emission input data. The whole vicinity of Nicosia is overlayed with the calculated average air pollution plume of Nicosia which has is maximum extension towards the southeast in agreement with the wind statistics. Single sources are blurred as they melt with this plume.

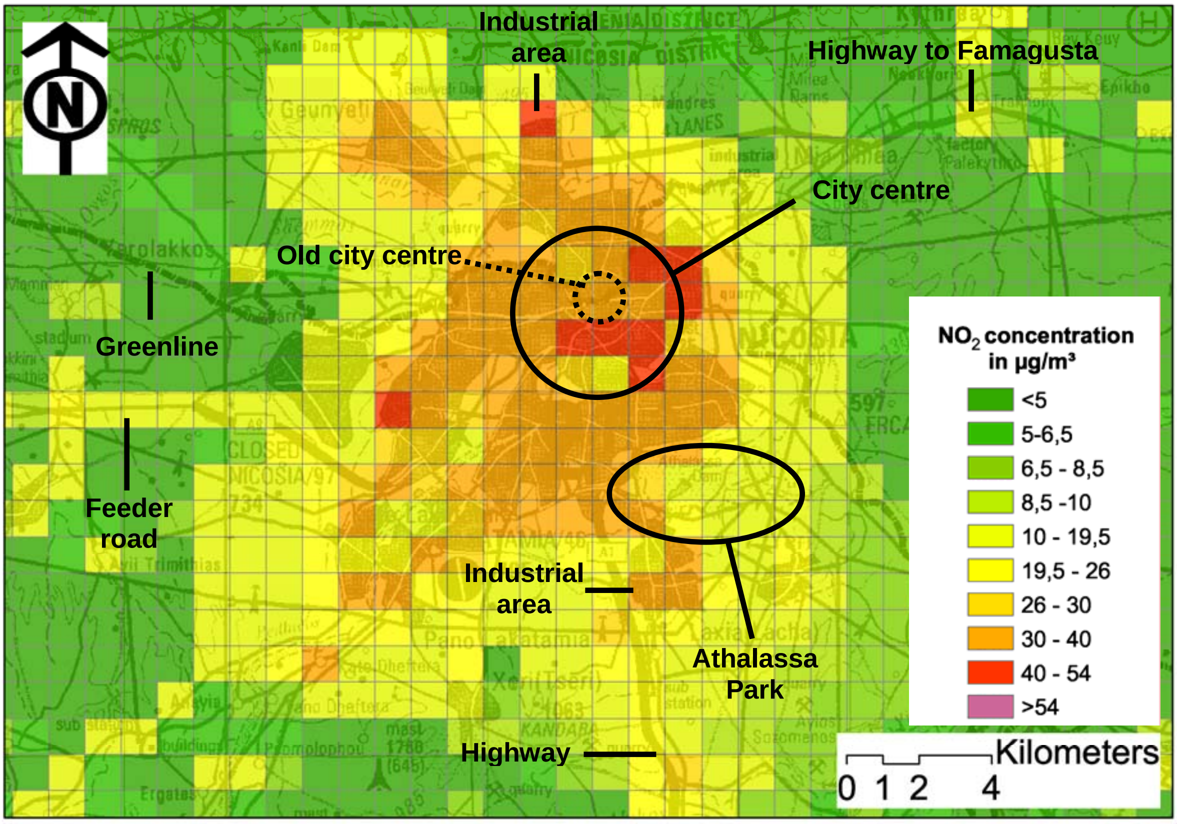


Figure 5.39. Annual average NO_2 distribution in Nicosia as calculated by the neural network model

The analysis above shows the suitability of the model for air quality assessment on a local scale. To increase the resolution significantly would not be advisable, since phenomena on a smaller scale become more and more important which would not be registered in detail in the emissions inventory - the accuracy of the model would suffer. For a higher resolution the wind field and its interaction with the development and the topography has to be regarded.

5.4.8 Quality assessment of models and input variables

A major task of this work was to find out the most suitable variables with which one is able to functionally explain the NO_x distribution over Cyprus. The internal assessment option of the neural modelling software allows observing directly the behaviour of the input neurons. This has been described above for the final model. In Table 5.5 the percentage of contribution of the different variables to the model performance of some selected models is shown. Contrary to what a model is able to declare in a univariate configuration, a multivariate environment causes different patterns. In Table 5.5 this can be observed considering the population density: Its contribution to the overall network performance in models 3 to 5 equals or even exceeds the contribution of the emissions. In models 6 and 7 however the emissions contribute about 2/3 to the performance. An explanation can be found in the dynamic nature of neural networks – a new variable creates different interactions of the weight adjustments during learning. High contribution values do not necessarily indicate the physical correctness of values as model 3 clarifies with a high contribution value of the altitude. The model quality can be assessed considering the standard deviation values and correlation coefficients listed in Table 5.5. As described above, statistical numbers can be considered only as a part of the assessment methods, which have to include the visualization of the results and physical considerations.

Table 5.5. Contribution of input variables to the model performance

No.	Model from chapter No.	Contribution of Input Variable to Model Performance in %					Standard Deviation	Correlation Coefficient
		Longitude	Latitude	Population Density	Emissions	Altitude		
1	5.4.1	63,78	36,22	-	-	-	15,74	0,48
2	5.4.4	-	-	-	100	-	11,3	0,62
3	5.4.7.2	29,74	17,82	25,14	27,29	-	10,01	0,67
4	-	13,58	16,08	20,76	14,98	34,6	9,79	0,66
5	5.4.7.3	-	-	32,21	37,42	32,21	14,97	0,51
6	5.4.7.4	-	-	33,07	66,93	-	8,27	0,75
7	5.4.7.4	-	-	38,3	61,7	-	8,1	0,76

5.5 Practical application of the model

As already stated in the chapters before, this work served to find a method to calculate maps of the distribution of air pollutants. But what are the benefits of having such maps and how can the model be used in practice? First of all, the model was designed to provide a tool that enables policy makers:

- To react on exceedances of existing limit values – the model allows to locate zones with the heaviest average pollution load and to quantify the average exposure of humans and nature to different air pollutants
- To establish locations where emissions reduction measures are necessary
- To formulate strategies to meet future limit values – trends of NO_2 pollution can be calculated by providing new emissions inventory data as an update or as forecasts for future scenarios
- To further optimise the monitoring network – continuous monitoring stations should be placed according to the calculated air pollution map
- To assess the effect of measures on air pollution prevention like traffic restrictions or the introduction of new fuel quality specifications for power plants or the traffic sector
- To increase public awareness

Furthermore, the model fulfils so far the requirements of the European Union: In combination with measurements, modelling techniques are recommended by the EU to assess the compliance of assessment thresholds in Directive 2000/69/EC [60], to observe zones with values one level lower than the limit values in Framework Directive 96/62/EC [2] and to provide essential information for the management of the air quality.

6 Critical discussion and outlook

6.1 Critical discussion

The model development to the final, realistic result was an iterative process where several failures had to be accepted and new directions had to be taken. This can be explained only partly by the fact that it was a completely new development. The reason for the problems stem also from the nature of neural networks as “black boxes” where the user has no direct view and influence to what happens during training. At the beginning, frame conditions like the network topology and the most important parameters are being set. When the training starts, the influence of the user ends with some minor exceptions. This own dynamic a neural network develops is partly it’s strength since it is the precondition for learning and partly it’s weakness, which can be only compensated by experience of the user.

Apart from this, the model can be critically evaluated as follows:

- **Quality of the model output and capability of the model:**

  With the neural network model realistic pollutant distribution maps on an annual base can be calculated. All assessment methods consisting of statistical analysis, internal criteria of the neural network software and the physical-visual control proved the high quality of the model output. All important emission sources are considered through the emissions inventory and the population density which helped to identify small villages that were omitted in the survey but certainly contribute to the composition of the air. With a 1x1 km resolution, the result map is very fine structured and all sources appear clearly on the map – area sources (cities), line sources (highways) and point sources (power plants and a cement factory). To achieve the highest possible agreement with the reality, the emission sources were distributed prior to the model run with a gaussian dispersion model and a statistical approach using the annual wind rose. This resulted in a new concept: The so-called “distributed emissions”.

  The model principle depends on the availability of diffusive sampling measurements that have to be on-hand highly resoluted. Other measurement methods could also be used of course, but they are not affordable with the required spacing. For the model presented here, 270 diffusive samplers were placed all over Cyprus with a higher density in the cities – for example 78 sampling sites in Nicosia (GCC and TCC). This opened the possibility for a reliable depiction of the urban air quality at the chosen resolution of 1x1 km. In rural areas less, widely spaced samplers are sufficient, since there are only a few sources, except some villages and roads. Generally speaking, the model calculated correctly the highest annual NO_2 concentrations in the cities of Nicosia, Limassol, Larnaka, Paphos, Kyrenia and Famagusta followed by some important single sources like power plants, highways and the airport near Larnaka. A few peak values were underestimated as the statistical evaluation showed. The result map also demonstrates the model sensitivity related to weak emission sources: Widely unpolluted areas like the Troodos mountains are located correctly as well as little villages at a comparable low emission rate. The output quality of the model depends mainly on the provided input data. The more influencing variables are provided, the higher the quality of the dispersion calculation, the more diffusive samplers are exposed – the better the model output. Artificial neural networks have the ability to approximate any function, linear or non-linear, multivariate or univariate. So if there are any correlations of the implemented variables that would help to functionally describe the NO_2 concentration, the neural network will most likely find them.

- **All kinds of numerical values can be implemented in the neural network model:**

  Since a neural network only calculates mathematical correlations between the provided characteristics, there are no limitations in the use of input variables. This opens vast possibilities in the development of neural air quality models. Subject to the availability of measurement data, a neural network model can be extended without limitations in terms of the consideration of quantifiable influences. Different model types are thinkable to cover the requirements of particular questions, e.g. to model chemical alterations of highly reactive species.

  To calculate the distribution of air pollutants like it was carried out in this work, high demands are however made to the input data: They have to be existent for every grid cell and based on the same time level unless they are constant like the height above sea level. Prior to the actual implementation in the model, big efforts for data preparation are necessary to obtain a data set that covers the whole research area (the same is true for conventional statistical models or interpolation methods using additional variables). But at the same time, the model principle to calculate pollutant concentrations for every grid cell based on the realization of correlated characteristics is the precondition for the major advantage of the model: The possibility to implement measurement data of the searched output, in this case the NO_2 concentration as measured with diffusive samplers. This is actually the only possible approach for a direct access to the real air quality situation in the regarded grid cell.

  A restriction of the model, which may appear at first sight contradictory, lies in the difficulty selecting the appropriate input data. Since a neural network has the ability to correlate any kind of variables (an improvement of the correlation coefficient occurs almost every time when a new variable is implemented), there is a danger of using all available data for network training. This may cause an improvement of the statistical quality of the model, but also it may cause a physically not justifiable pollutant distribution. During development, this occurred with the UTM coordinates and the altitude above sea level. Both variables are actually partly justifiable, since there is a physical reason for their usage. But this physical justification cannot be applied globally. This means in the case of the altitude above sea level that a low altitude value does not always directly correlate with a high pollutant concentration. It is true that the major emission sources are located at the coastline, but is also true that river valleys are located at a low altitude value. The neural network should be able to learn that high pollutant concentrations only occur if there are emission sources, which is of course not the case for river valleys. Nevertheless, the model predicted high NO_2 concentrations in river valleys, a problem that might be compensated with more training data or supplemental information. A careful check-up of the used input variables is inalienable to assure a physical correct model.

- **No statistical preconditions for the input data of the neural network model have to be fulfilled:**

  The method as developed in this work can be generally considered as a statistical model. Conventional statistical models were developed on the base of theoretical assumptions like the normal distribution or the exclusion of autocorrelation. Practically, tests and extensive data conversions are necessary to approximate the compliance with these preconditions prior to their implementation in the model. The same applies to the results that also have to bear up some statistical tests (e.g. test for heteroscedasticity). Normally the conditions cannot be perfectly met – compromises have to be accepted, whereas neural networks are very modest in their requirements to the input data. The only precondition is that the input data have to be numerical values – a great advantage compared to statistical methods like the regression analysis, since a great part of the data preparation prior to the actual model run can be left out. This saves time and makes neural networks easier to use than conventional methods.

- **Neural networks versus conventional regression analysis:**

  As mentioned above, there are statistical methods that theoretically could replace the neural network model, since it “only” calculates correlations of different variables. Above all, the regression analysis would be the method of first choice to develop such a model based on conventional statistical methods. Due to the theoretical assumptions that have to be kept, conventional statistical methods require a very laborious data preparation. But on the other hand, these are established methods and easy to use with software programmes like SPSS. Strictly speaking this applies however only for “simple” univariate/multiple linear regressions and univariate non-linear regressions.

  Comparable to neural networks are multiple non-linear regressions. They are much more complex than the more simple linear and univariate regression approaches and require a pre-formulation of a formula which is the starting point for the iterative approximation of the best fitting model. Different iteration approaches like Levenberg-Marquart are possible. Neural networks however are able to approximate any function, linear or non-linear, without any greater adjustments. Another advantage is that neural networks are much more dynamic compared to the regression analysis. An outlier caused by a measurement error or a wrong data entry by the user could dramatically worsen the quality of the goodness in a regression model. In a neural network model, this would not have such a big effect. According to discussions with other scientists and own experiences [53], it can be concluded that neural networks are indeed the most flexible “statistical” method, but a well-configured non-linear model based on regression analysis could also deliver results that come close to it.

- **Neural interpolation versus conventional interpolation:**

  In chapter 5.4.1 it was found out that with the UTM coordinates as input data for the neural network it is possible to create a neural interpolation of the NO_2 diffusive sampling measurements. The obtained NO_2 distribution reflected very well the most important emission sources, namely the cities, but in a generalized way. It also created unrealistic linear and circular/ellipsoid structures. The reason for this behaviour lies in the generalization the network carried out: It learned that certain coordinate combinations result in high NO_2 values and others in low values. All sampling sites influence each other; there is no possibility to isolate single sites or groups of sites. This is however necessary to reflect the physical dispersion process – in the case at hand, sampling sites at the coast are influencing sites far away in the mountains and vice versa. Using only the UTM coordinates as input is not the solution for a good neural interpolation.

  Conventional interpolation methods require also the coordinates as input. But here, the actual geographic location is used in the algorithm to calculate the distance between the different sampling sites. A sphere of influence can be specified that doesn’t have to be necessarily circular. In the case of Kriging it is also possible to consider the “spatial roughness” of the data set by fitting a variogram.

To develop an adequate interpolation method for neural networks, it would be necessary to translate the conventional interpolation parameters in a way that they can be used in a neural network. Above all these are the distances of the different points to each other, or to important emission sources. For this work it was not necessary to develop such a tool since the distances to the most important sources are contained in the distributed emissions and the final model principle differs from interpolation.

Using a neural network only as an interpolation tool, it must be stated that conventional interpolation methods are more powerful and easier to use at the moment. Software programmes like Surfer and ArcGis allow a total control over the interpolation process. There are many adjustable parameters that highly influence the final result.

- **Neural networks versus dispersion models:**

Dispersion models in combination with a wind field pre-processor simulate how gaseous air pollutants or particles leave an emission source and subsequently move towards the receptor. Certain important aspects of the dispersion process are parameterised and implemented in the model, trying to reproduce the physical reality of the mixing layer. So the needed input for this model-type is limited to physical values like wind speed, wind direction, emission mass flow rate and topography. The treatment of other important influences like the population density is only realizable with big efforts or in combination with other methods – a restriction that doesn't apply to neural networks. Dispersion models are very useful to model case studies, how a single or a few emission sources affect a defined area. To calculate the complete spatial distribution of air pollutants in a large area like Cyprus at a 1x1 km resolution under consideration of all emission sources is hardly realizable. It would require carrying out 9.251 dispersion calculations since the island of Cyprus covers an area of 9.251 km².

With neural networks however, the mathematical dependency of any numerical variable can be calculated. Practically, the measured concentrations of NO_2 are correlated with area-wide available secondary information and assigned to a 1x1 km analysis grid. The possible size of the research area only depends on the availability of input data, whereas the physical process of dispersion cannot be parameterised and implemented. At the moment, it is impossible to replace a dispersion model and a wind field model by a neural network.

Artificial neural networks and dispersion models are suitable for different questions, but can supplement each other:

- A neural network can be used as a filter to improve the performance of a dispersion model as carried out by Pellicioni et al. [23].
- A dispersion model was used in this work to calculate the spatial influence of the most important emission sources in Cyprus.

- **Grid spacing – representativeness of diffusive sampling measurements:**

For the development of the neural network model, Cyprus was overlayed with a 1x1 km analysis grid – a fine resolution since the research area amounts to 9.251 km². Anyhow, a problem might be especially in cities that the concentration values in a grid cell may fluctuate highly, depending on the distribution of emissions and the dispersion conditions. A larger cell spacing increases the probability of irregular distributed emission sources with a wide range of emission strengths. So the measured concentration used for training may not mirror the average value, which is why the sampling sites should not be directly influenced by a single emission source. An outlier could worsen the model performance, although neural networks are less sensitive to such a case than conventional statistical models. The statistical evaluation of the model output of this work actually showed a few outliers that stem from locally influenced diffusive samplers.

To reduce the size of the grid cells would not be the appropriate solution, respectively it would solve the problem only partly since pollutant concentrations would still fluctuate significantly. Considering the microscale range (meter-range), special meteorological phenomena like eddies at the lee of a building gain more and more importance. Microscale dispersion models still are the only possible approach to model such cases.

- **Neural networks are black boxes, no direct interference during training is possible with minor exceptions:**

  As described in chapter 5.3 the user has only little influence on the weight changes of the connections between the neurons. These weight changes simulate the learning process of a biological network and represent the shrinking or growing of real synapses. If a satisfactory result is reached depends amongst others on how the starting weights are initialised and how the learn rate is set. Practically a gradient descent on the error function is carried out aiming to reach the global minimum. Local minima, flat plateaus, error oscillation and symmetry breaking are possible problems that could occur during training. It requires however experience to distinguish during model training if the run indeed has reached it’s error minimum or if overtraining, respectively one of the problems mentioned above, occurs. This circumstance and the fact that the influence of the user ends after setting up the network topology and the training parameters turn artificial neural networks into black boxes. In most of the cases however, after some experience in the model usage and the distribution of the input data, it is possible to glean the general response of a network and the possible correlation of the input data. It is recommended to analyse scatter plots prior to the model configuration to reveal obvious correlations among the data sample. Knowledge on the physical context of the used input helps to understand the results and prevents following a wrong path. In this way it is possible to illuminate the black box neural network for the most part.

- **The neural network model can’t be used as a standalone tool:**

  Earlier it was set fourth that a crucial influencing factor can’t be calculated with a neural network – the dispersion conditions. Although meteorological parameters gain more importance when a small time or spatial scale is chosen, they also can’t be neglected for annual averages as calculated with the model developed in this work. The only exception are stable weather conditions with inversion layers and very low wind speeds. Normally such barrier layers are built up during nighttime and gradually disappear after sunrise, but in some cases, especially in winter, they may continue several days. So a more or less static mixing layer is a single phenomenon with a short duration. Considering an annual time base and a 1x1 km grid, it is not necessary to include diurnal weather phenomena and microscale processes in the calculation of a spatial pollutant distribution. Simplified approaches are sufficient to calculate annual statistics.

  At the moment is not possibility to model a wind field with a neural network since too many variables have to be considered. Some models like FLADIS simply interpolate the wind vectors of available measurement stations to create a wind field. It would be possible to carry out a similar neural interpolation, but this would lead to a complete disregard of the physical mass exchange processes. Neural or conventional interpolation of wind data would only lead to acceptable results in a flat terrain situation and a very dense network of wind measurements, which is quite rarely the case. So alternative approaches were utilized to provide the necessary input data: For the point sources in Cyprus (mainly power plants) a gaussian dispersion model, for the area sources (cities) and line sources (highways) a statistical approach, derived from wind roses. The used statistical approach could also be partly solved with a neural network. Moreover, all calculated pollutant concentrations had to be converted to the newly conceived “distributed emissions” prior to the model run.

Due to the described problems it is at the moment not possible to present a standalone neural network model without the support of other methods or pre-processors that however only serve to provide the necessary input data. The final and respectively the actual calculation of the pollutants distribution is in fact to be carried out with the neural network model. It would be possible furthermore to include all methods in one software package, like it is often realized in the case of dispersion models.

- **Neural networks are very fast:**

  Neural network models are principally developed in two steps: First of all they are trained with a set of the searched output variables (in the case of this work NO_2 concentrations), afterwards the learned correlations are assigned to the area-wide known distributed characteristics like emissions and population density where no measurements of the NO_2 concentration exists. This second step is referred to as the “recall mode”. Most of the running time consumes the first step since the training procedure is an iterative process. When training is initiated, starting weights of the neural connections are randomly set and step per step optimised to approximate the searched function. This process may take several minutes depending on the network architecture (more input variables, more hidden layers and more neurons per layer entail more training time) and the hardware specification of the used computer – the power of the CPU is the determinant element here, neither large temporary files are created nor the RAM is notably affected. In the case of the present work a 1,8 GHz Pentium 4 system was used and the maximum training time was about 5 minutes. For the recall mode only a few seconds are required.

  Compared with statistical methods it can be stated that the multiple linear regression analysis is much faster, but since a linear correlation is improbable in most cases, a data transformation is necessary prior and after the actual regression calculation. A non-linear multiple regression analysis takes approximately the same time as neural training, since it is also a method based on iteration. Dispersion models are much slower and take at least several minutes (Gaussian models) up to several days (Lagrangian models or Eulerian models). Restrictively it must be added that the data preparation for the neural network model presented here or a comparable regression model require very much time for data preparation. This partly compensates the headstart of these methods, but since dispersion models are intrinsically useful for case studies, it would be very time consuming to consider as many emission sources as it was carried out with the neural network model in this work.

- **The neural network model can be updated (trained) successively with new measurement data and improve it’s calculation:**

  It is easily possible to add new influencing variables to the neural network model as well as it is possible to feed it with new diffusive sampling results. Normally a neural network improves its performance when more training cases are available. On the other hand it is recommended to constantly update the database of influencing factors, namely emissions inventory and population density. Especially new sources have to be considered in the emissions inventory, otherwise the pollutant concentrations of the grid cells in the vicinity of the new source would be underestimated since the network has no other information.

New, actual pollutant distributions can be calculated by a new training of the neural network with diffusive sampling measurement results.

- **Reproducibility:**

  A very important issue on the possible application of the model in practice is its reproducibility. It must me assured that the model always delivers reliable results. Theoretically, a neural network model is never 100% reproducible. Since the initialisation of connection weights is always random, every model is unique. Therefore, the learned functional correlations of two models with the same frame conditions are also always different. Training has to be carried out very carefully to avoid the possible traps: Local minima, flat plateaus, error oscillation and symmetry breaking. With the used software Qnet2000 the LCR-feature (learn control rate) helps to avoid such cases. During the development in the context of this work reproducibility never was a problem.

  Restrictions on the model reproducibility may however appear in the case of further developments when new variables are implemented. A so far powerful model may suffer dramatically when a new variable is added – training may reconfigure the connection weights significantly resulting in a completely different model with a poor performance. What would at the worst only have no effect at all on a regression model (no significant change of model formula, only marginal improvement of the coefficient of determination) could highly lessen the quality of a neural network model.

  Other methods suffer comparable problems:

  - Lagrangian dispersion models: The number of simulation particles will usually be of many orders of magnitude smaller than the actual number of the trace species particles (atoms, molecules, aerosol particles) to be modelled. When repeating the same simulation with different random numbers, it can normally be observed that the number of simulation particle in a sampling volume and hence, also the concentration value determined for the respective grid cell, are subject to random fluctuations. Consequently, the result of the dispersion calculation is always associated with a certain degree of uncertainty, the sampling error.
  - Multiple non-linear regression analysis: The choice of starting values for the parameters and the chosen algorithm highly influence the final model. Convergence is not always assured.

Finally it can be concluded, that a neural network is an excellent tool with great advantages concerning the ability to learn correlations of different variables. But at the same time it still is new and mainly applied by scientists besides its implementation in some fields like satellite picture analysis, speech recognition or optical character recognition (OCR). For inexperienced users, there is a lack of standardized methods at the moment, but this will change most probably in the near future.

The model presented here enables us to calculate the distribution of NO_2 over a large area. According to the statistical and qualitative assessment of the result, the model seems to reproduce the reality very well. A great advantage is the possibility to feed the model with a lot of measurement data; in other methods they are not only used for validation. Unlike dispersion models, this characteristic enables the neural network to learn from actual values of the investigated area providing a direct link to the reality. But on the other hand, with a neural network the shape of a dispersion plume from a power plant cannot be predicted unless there exists a dense net of measurement results around such a source. A neural network still cannot replace an atmospheric dispersion model, which is why a Gaussian Plume model was used to support the calculations. Another limitation is the difficulty to find the best network topology and training parameters. The latter can be partly compensated with the Learn Rate Control feature of the used software, whereas the best topology can be only found through tests. Another problem encountered was the underestimation of peak values that stem from local influences. Some of those miscalculations can be explained by diffusive samplers that do not represent the average air pollution of the whole grid cell, since they are influenced by local sources. So, the calculations might come closer to the actual average pollution of a grid cell as the measurements - an assumption that cannot be proven. This underlines the importance to expose diffusive samplers at representative sites uninfluenced by a local source. Exceedances of limit values still should be monitored with continuous measurements.

6.2 Outlook

Although the neural network delivered very good and realistic results, there are still many possibilities to expand and enhance it in the future.

The quality and abilities of the model depend strongly on the provided input data. A first step for a further development would be to improve the database, in the case of Cyprus especially the emissions inventory of the Turkish Cypriot Community. Moreover, a higher time resolution of emissions would open the possibility to calculate seasonal or monthly pollutant maps. Daily or hourly calculations are not possible with this method, since the diffusive samplers are exposed for about one month. Together with wind field calculations and input data from continuous monitoring stations however, such a model would be conceivable. Here, the diffusive samplers would serve as pollutant level indicators. Another enhancement of the database would be land use maps that can be expected to further differentiate the network outputs. Their resolution should be at least 1x1 km. It was stated that the meteorology is a very important influence on the distribution of air pollutants. Instead of using statistics for the model, which was found to be impossible, the average ventilation conditions, characterised by the average wind speed of a grid cell could be calculated using a diagnostic wind field model. For models with a higher time resolution the application of stability classes as measured by balloon soundings or derived from cloud coverage and ground measurements would be a further option. Due to regular weather phenomena like the land-see-breeze or mountain-valley-breeze a division into a nighttime and daytime model is also thinkable. Of course the model is also expandable to other species that can be measured by diffusive samplers. For some pollutants like Benzene NO_2 is a good indicator, their correlation could be calculated with a neural network. This would allow reducing the necessary number of measurement sites. But also the influence of other variables, like the altitude in the case of Ozone, should be checked here.

For the future in Cyprus, a net of 270 diffusive samplers is economically not justifiable. With the NO_2 distribution map it is now possible to strategically place and reduce the number of diffusive samplers without loosing the information of the trained network. New training cases will help the network to better understand the correlations. It is strongly recommended to select sites with no direct influences of single sources. A grid nesting with a higher resolution in the cities would also be a meaningful approach, but only to a limited extent: The pollutants distribution in the street network of a city can be only calculated with a micro scale dispersion model since the buildings have to be considered [61]. A neural network could be used here as a filter to process the result data as described by Pellocioni et al. [23, 24].

To sum it up it can be said that the potential of artificial neural networks in the field of air quality modelling still isn’t exhausted, but for many problems they will always at best serve as a support.

7 Summary

The objective of this work was to develop a method to calculate realistic air pollutants distribution maps. Most common state of the art dispersion models are limited in using air pollutant measurement data. They are only used for validation of the results - a fact that does not limit the importance of this model type which is still first choice for many questions. Yet, the usage of many, well distributed measurement data as a model input is the only possibility to provide a direct link to the actual air pollution distribution. So the first demand to the model was the ability to consider such data in its algorithm. Diffusive samplers are a predestined method to assess the spatial distribution of air pollutants, since they are the only economic possibility to carry out simultaneous measurements at many places [37]. During the UNOPS project "Preliminary Assessment of Ambient Air Quality in Cyprus" NO_2 diffusive samplers were exposed at 270 sites in six campaigns throughout one year. In this way, an excellent database for the model development was created. A second requirement for the model was the ability to consider the most important influence parameters, like the emissions or the population density – most available methods are limited in this respect. Finally it was important that the model could be updated with new data to increase the quality of its calculations.

Three methods were found to generally fulfil these requirements: Regression models, interpolation algorithms and artificial neural networks. First, the regression analysis was excluded since it is too rigid, too limited to reproduce non-linear correlations and based on too many theoretical statistical assumptions that are hard to be kept. Interpolation is a well-developed standard method that is also recommended by the European Union. It was therefore tried to investigate the possibilities that lay in this approach. Simple interpolation failed to reproduce a realistic distribution of air pollutants, although exact interpolation algorithms preserve the measured values. The result maps depend strongly on the location of the measurement site; sources in between are neglected. A possible solution to this limitation are interpolation methods that consider additional variables like Cokriging. But still no significant improvement could be achieved here. For a quick visualisation, interpolation maps can be considered as a good tool. Finally, artificial neural networks were found out to be the best solution.

In air pollution modelling and comparable fields of neural network application, it is common sense that multilayer perceptrons with backpropagation deliver the best results. Such a network is implemented in the Windows based software "Qnet 2000" that was finally chosen as the development tool. Two principal criteria influenced the configuration of the neural network model, which where the network topology and above all the consideration of the most important influencing factors – the input variables. In training mode, a so-called "controlled learning" was applied, where the user provides the input data together with a set of known outputs. The task of the network is to learn the correlation between inputs and outputs. For this purpose, an analysis grid of 1x1 km was laid over the research area of Cyprus and the most important influencing variables were intersected with this grid. For all grid cells with diffusive samplers, the correlations of the annual NO_2 concentration with the input variables were calculated.

First of all a neural interpolation of the diffusive sampler measurements could be established by using the UTM coordinates as input. The result was a simplified interpolation map with NO_2 concentrations around 30 to 40 µg/m³ in the cities and lower concentrations in rural areas. Remarkable here are unrealistic straight stripes of medium NO_2 concentrations of about 20 µg/m³ connecting the cities of Nicosia, Limassol and Larnaka where the highest concentrations were measured.

A digital elevation model then was fit to the analysis grid and used as input. Again some improbable values like high NO_2 concentrations in river valleys were predicted. This clarified the necessity to consider physical aspects instead of using any available input variable. In this case, the network correlated the fact that all major cities in Cyprus are located at a low altitude with the occurrence of high NO_2 in these cities.

A very important influence on the distribution of air pollutants are the dispersion conditions. Since all meteorological parameters are constantly varying, it is impossible to calculate an average wind field [54]. Statistics would be the only solution, but cannot be used directly as input. It is true that an attempt to calculate two different models for winter and summer season correctly resulted in higher concentrations in winter, but the overall performance here was not satisfactory. The only realizable possibility to include meteorological data was to consider case studies for the major emission sources in Cyprus. So the dispersion plumes for the most important sources were calculated using annual wind statistics: For the largest cities, the power plants, one cement factory and the highways. Applied methods for this purpose were the Gaussian dispersion model P&K 3782 and statistical evaluations including regression analysis. Finally the plumes were transferred into so-called distributed emissions. Trained with this input data, the network reproduced the actual location and strength of all the sources that were used as input.

After using univariate models, multivariate calculations were carried out to depict the complex correlations that influence the distribution of air pollutants. All models with UTM coordinates again created unrealistic large and regular plumes coming from the south-eastern bottom of the map, which lead to their exclusion. The same is true for the altitude. For both maps the population density was added with no positive effect.

Finally, the best result could be established with a model that uses the enhanced emissions inventory (the plumes of the most important sources were calculated using wind statistics) and the population density as inputs. This configuration preserved the positive aspects of the univariate model with emissions and added additional sources like villages, which were omitted in the emissions inventory. A realistic, fine structured map was the result.

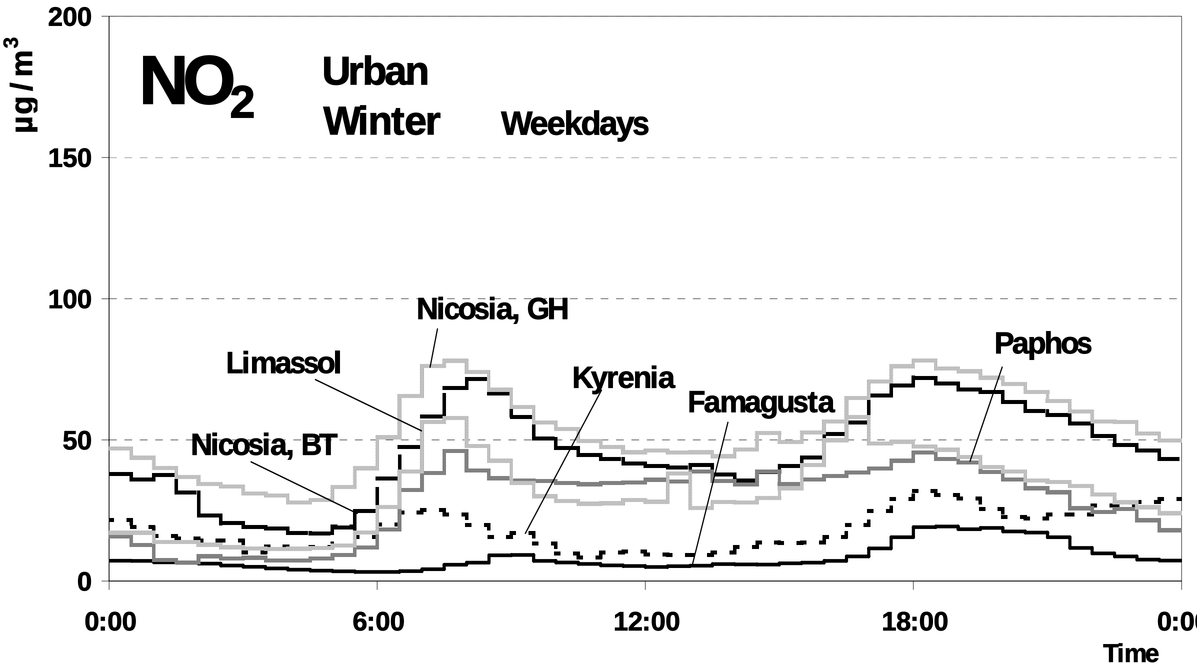
Internal and external assessments of the results were carried out during the whole development phase such as the monitoring of the prediction error from a test set of input data (observation of overtraining), the comparison of the prediction results with diffusive sampling measurements and the visualization of the results as GIS maps. The final model lead to a Pearson correlation coefficient of 0,75.

Annex A

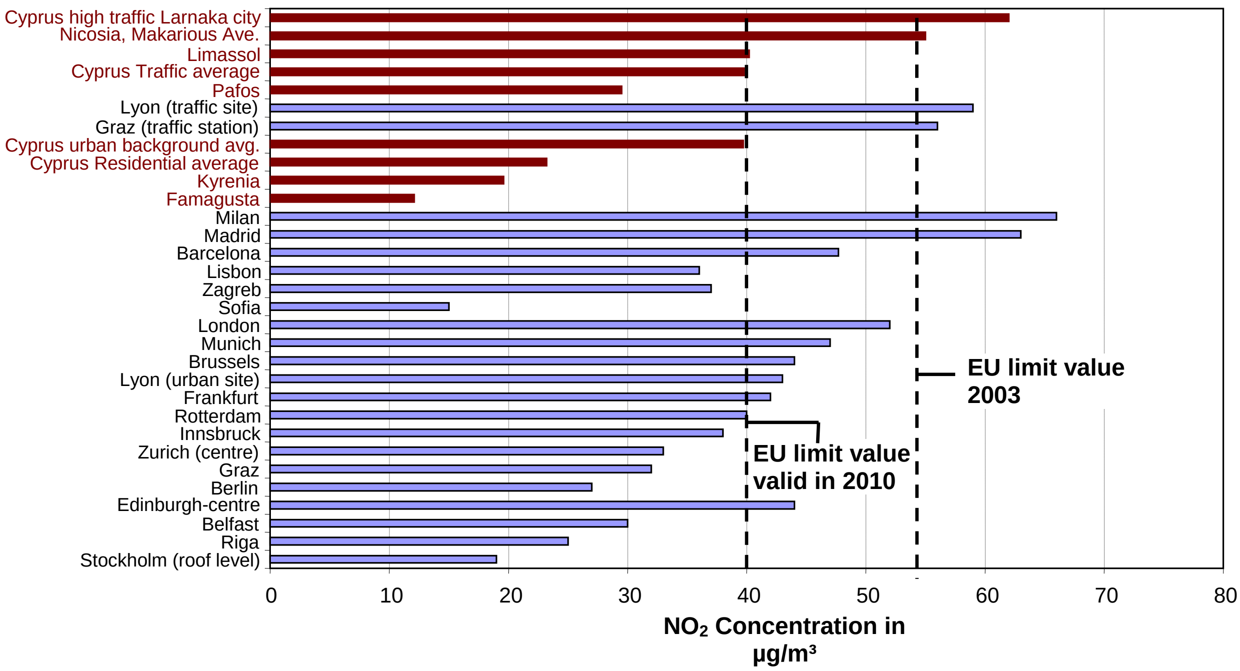
Local conditions in Cyprus

Exemplary measurement results and observations from the project “Preliminary Assessment of Ambient Air Quality in Cyprus” are shown in this part of the Annex.

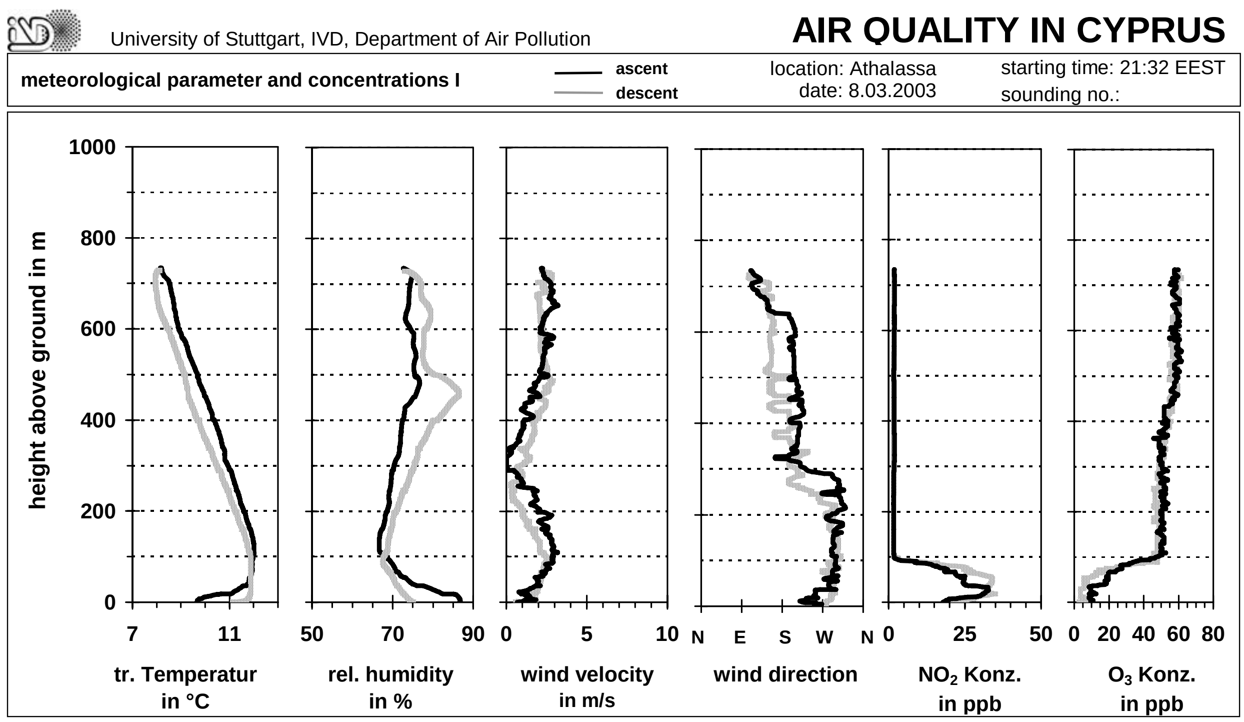
Annex A.1. Shelter with diffusive sampler used in the field



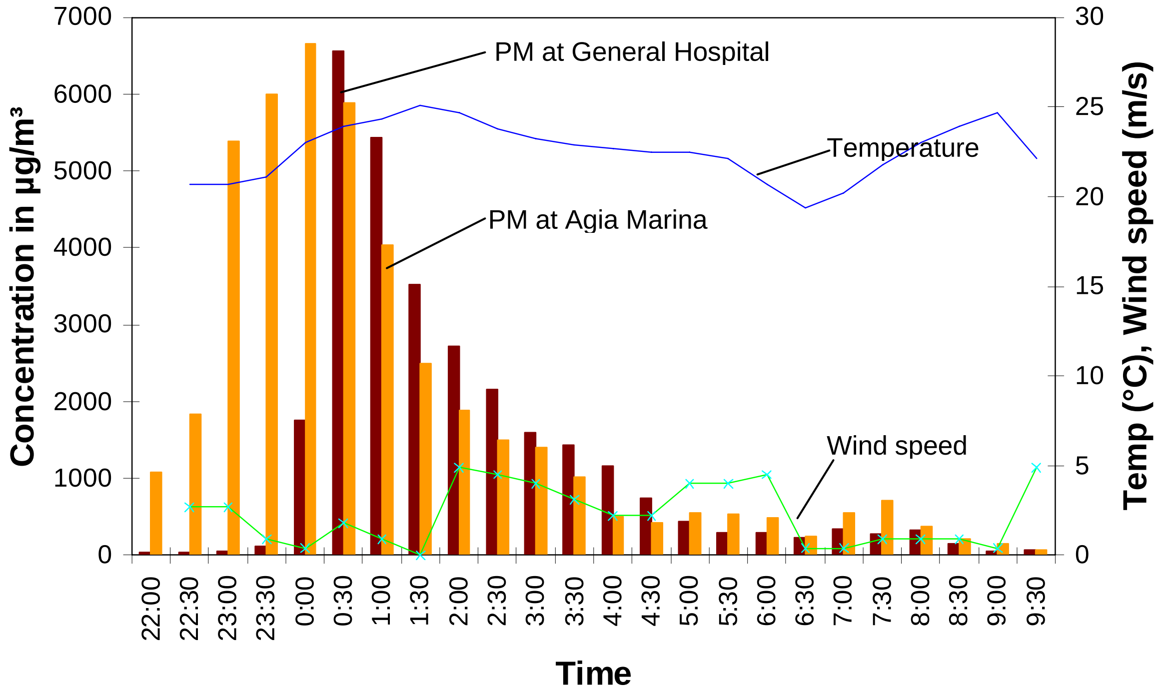
Annex A.2. Continuous measurements of NO_2 at urban sites – average diurnal courses on weekdays



Annex A.3. Comparison of the Air Quality of cities in Europe - Annual mean values (Cyprus: 2002/2003, European cities: 2001)



Annex A.4. Balloon sounding results in Nicosia on 8.3.2003 at 21:32 – a lifted ground inversion occurs at approx. 50m, which can be clearly seen in the vertical temperature and NO_2 profiles [34]

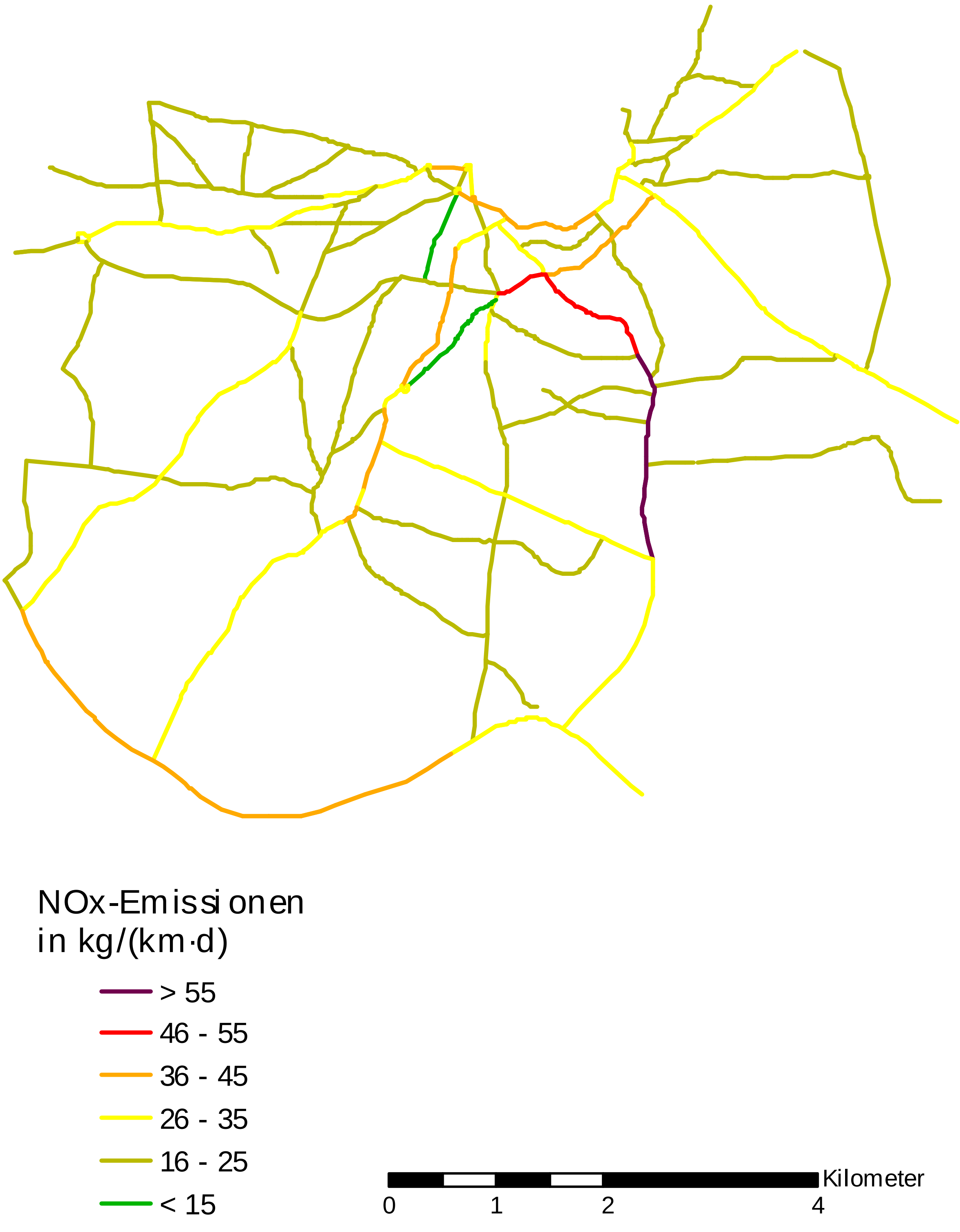


Annex A.5. Sahara dust event on 29./30.5.2003. PM_{10} measurements show a significant peak both at a traffic influenced measurement station (General Hospital in Nicosia) and a background station (Agia Marina, South of Nicosia and therefore with a one hour shift) [34]

Annex A.6. Power plant near Vasilikos

Annex A.7. Cement factory near Vasilikos

Annex A.8. Uncontrolled waste burning in Cyprus – a large plume can be seen in the middle of the picture, travelling from left to right



Annex A.9. Daily emissions of the traffic sector in Nicosia – improved with traffic data from quickbird satellite pictures [63]

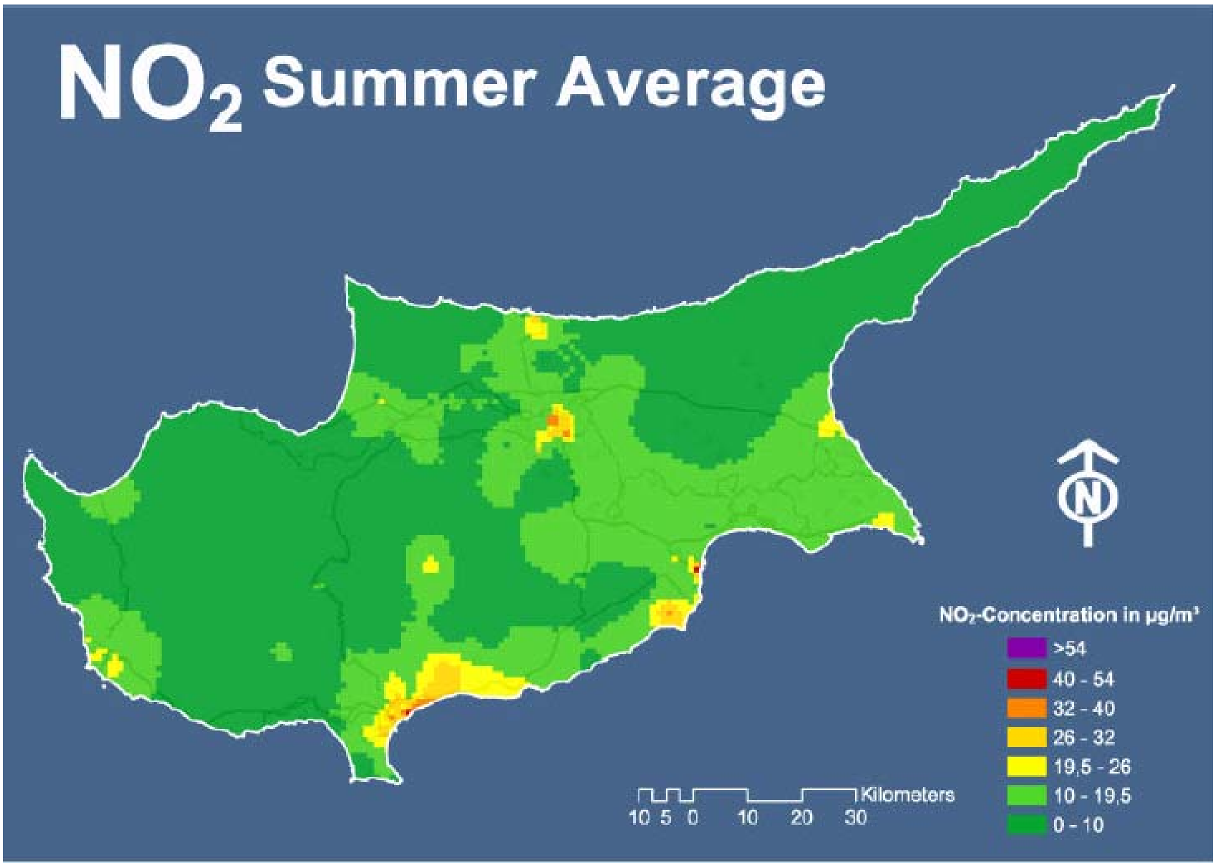
Annex A.10. Quickbird satellite picture of Nicosia for car countings [63]

Annex B

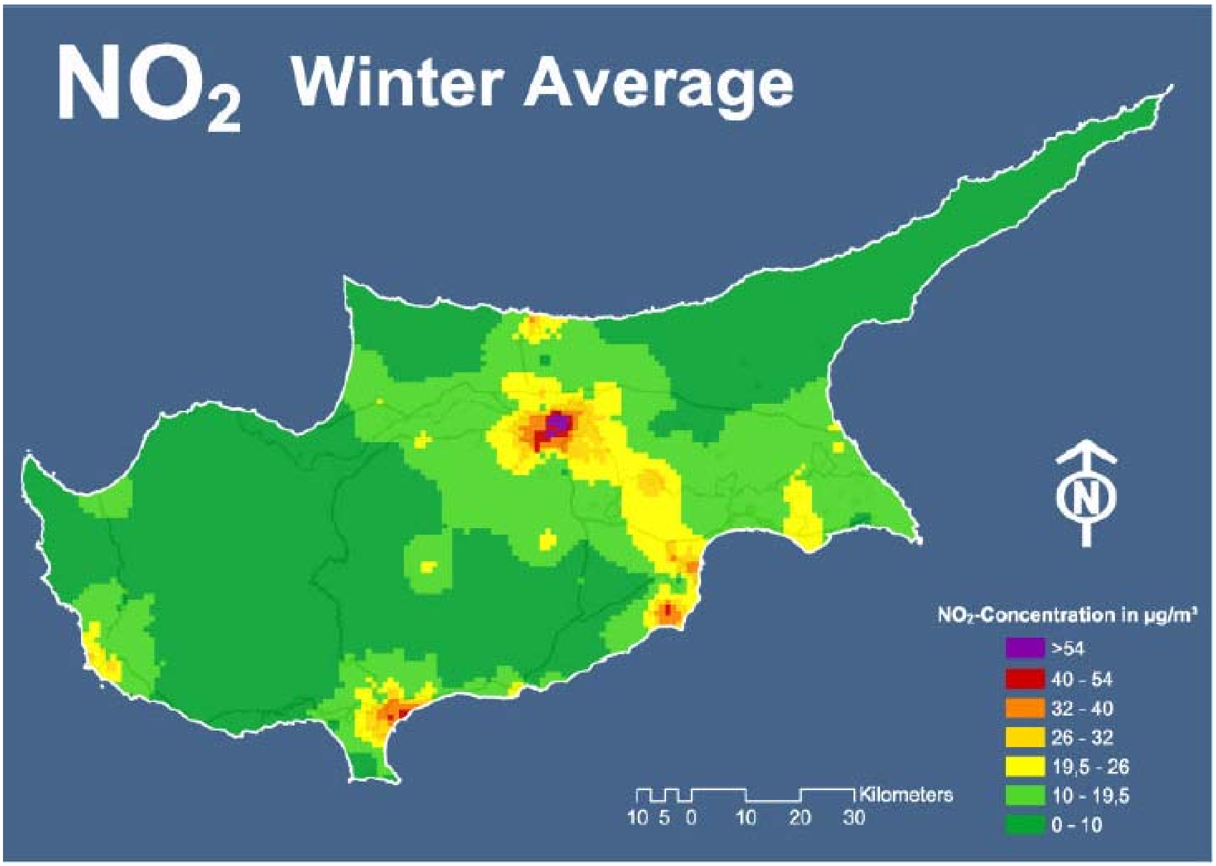
Modelling

In this part of the Annex, detailed interpolation maps of the largest cities in Cyprus and for the whole island are shown as well as supplementary information on the method interpolation. Summer and winter are differentiated for Nicosia and the whole island, which gives a good impression on the seasonal dependencies of the pollutant distribution due to meteorological changes and changes in emissions activity.

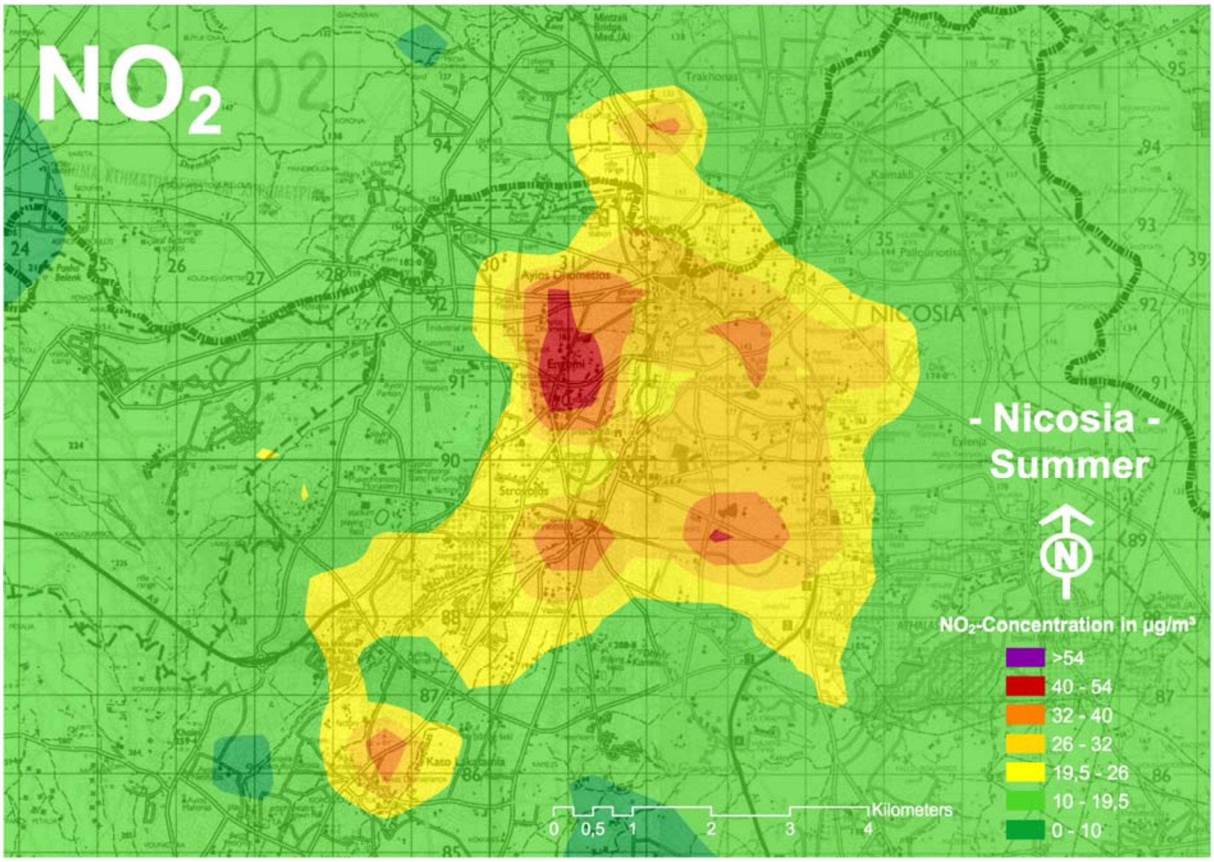
For the emission part of the model, a gaussian dispersion calculation of the 5 major sources in Cyprus was carried out. The results and other information related to the development of the neural network model are also shown here.



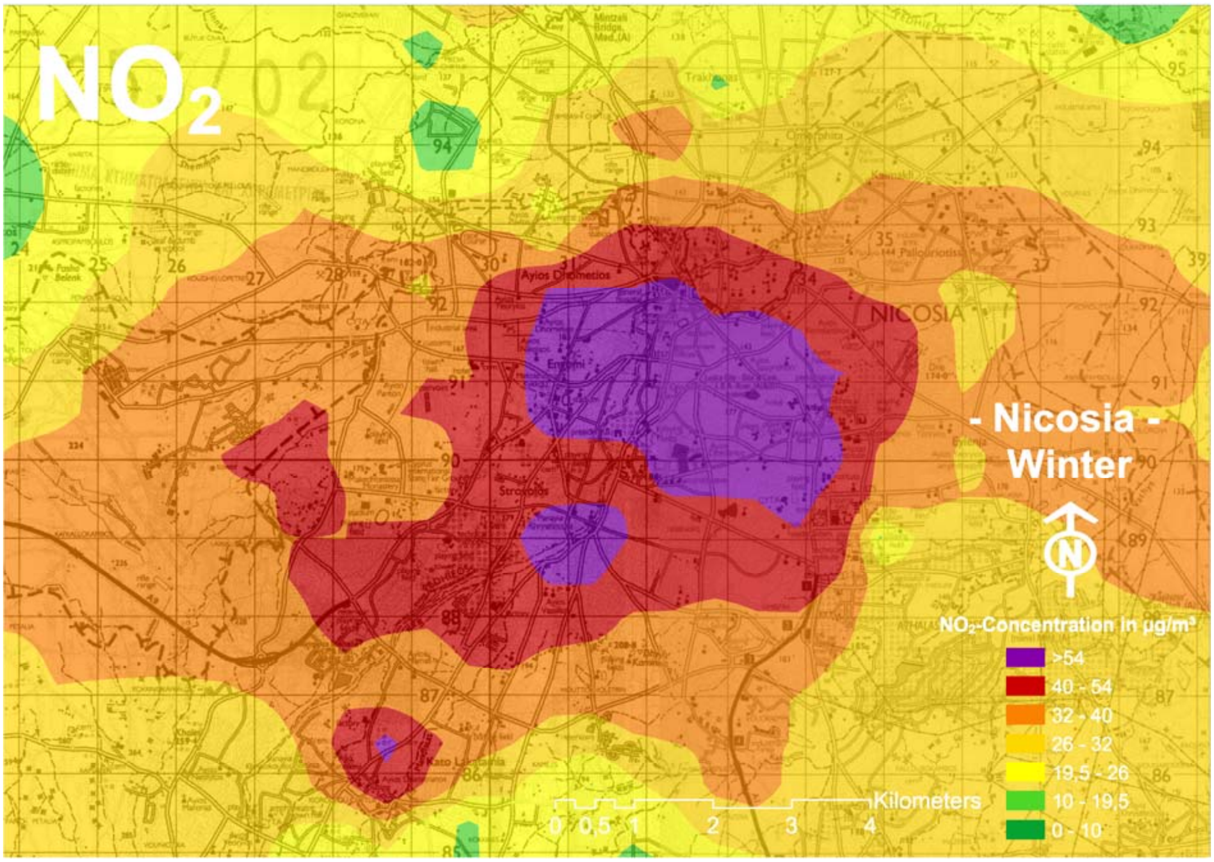
Annex B.1. Interpolated NO_2 distribution in Cyprus – July to September 2002 and April to June 2003



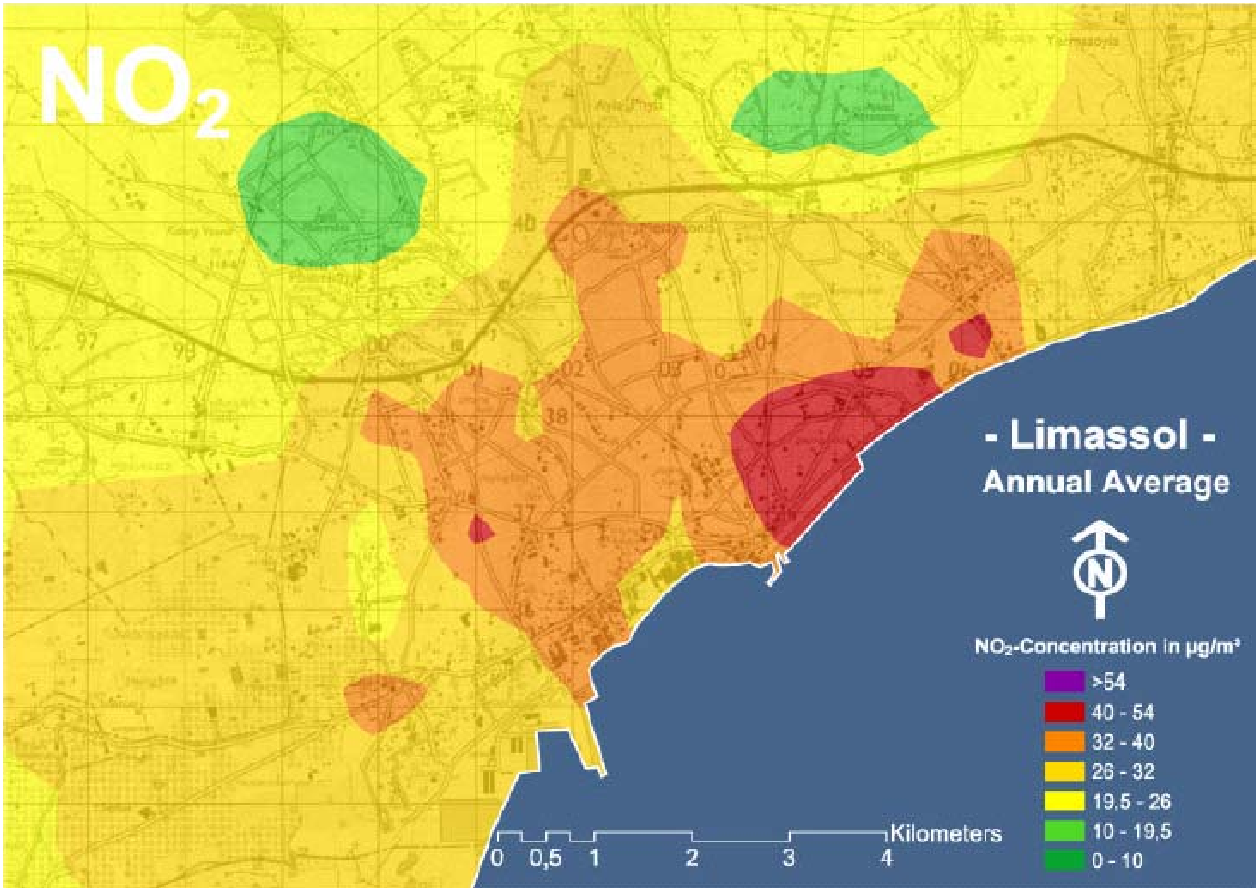
Annex B.2. Interpolated NO_2 distribution in Cyprus – winter 2002/2003



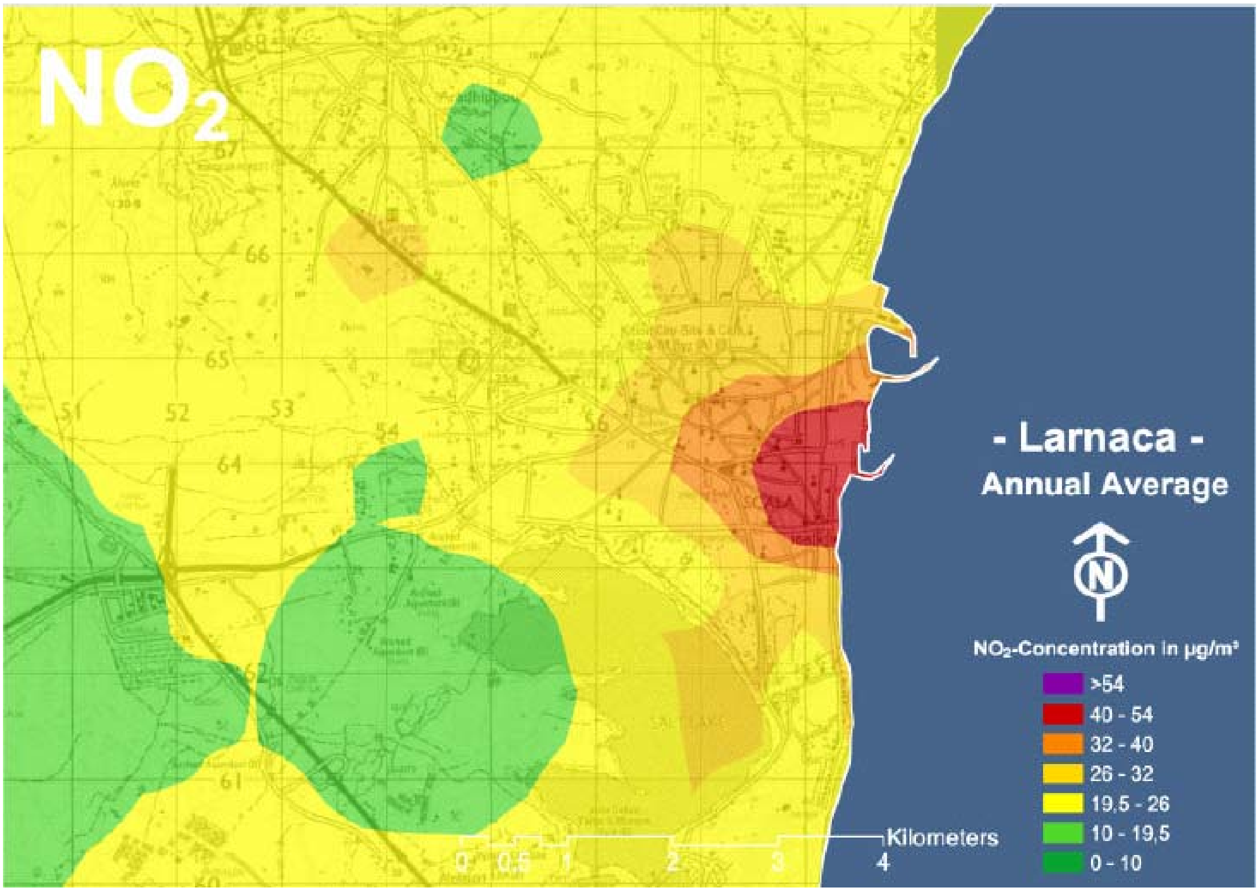
Annex B.1. Interpolated NO_2 distribution in Nicosia – July to September 2002 and April to June 2003



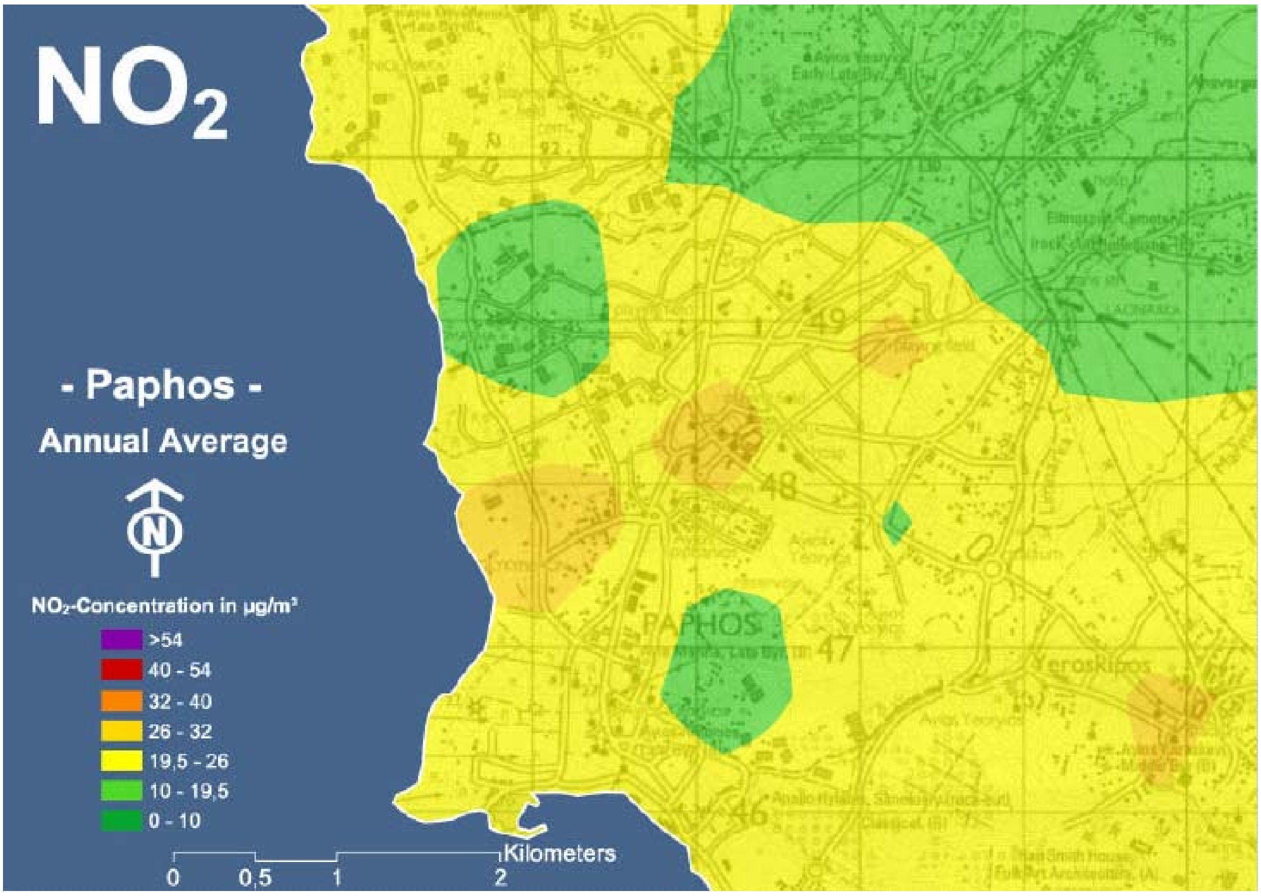
Annex B.4. Interpolated NO_2 distribution in Nicosia – winter 2002/2003



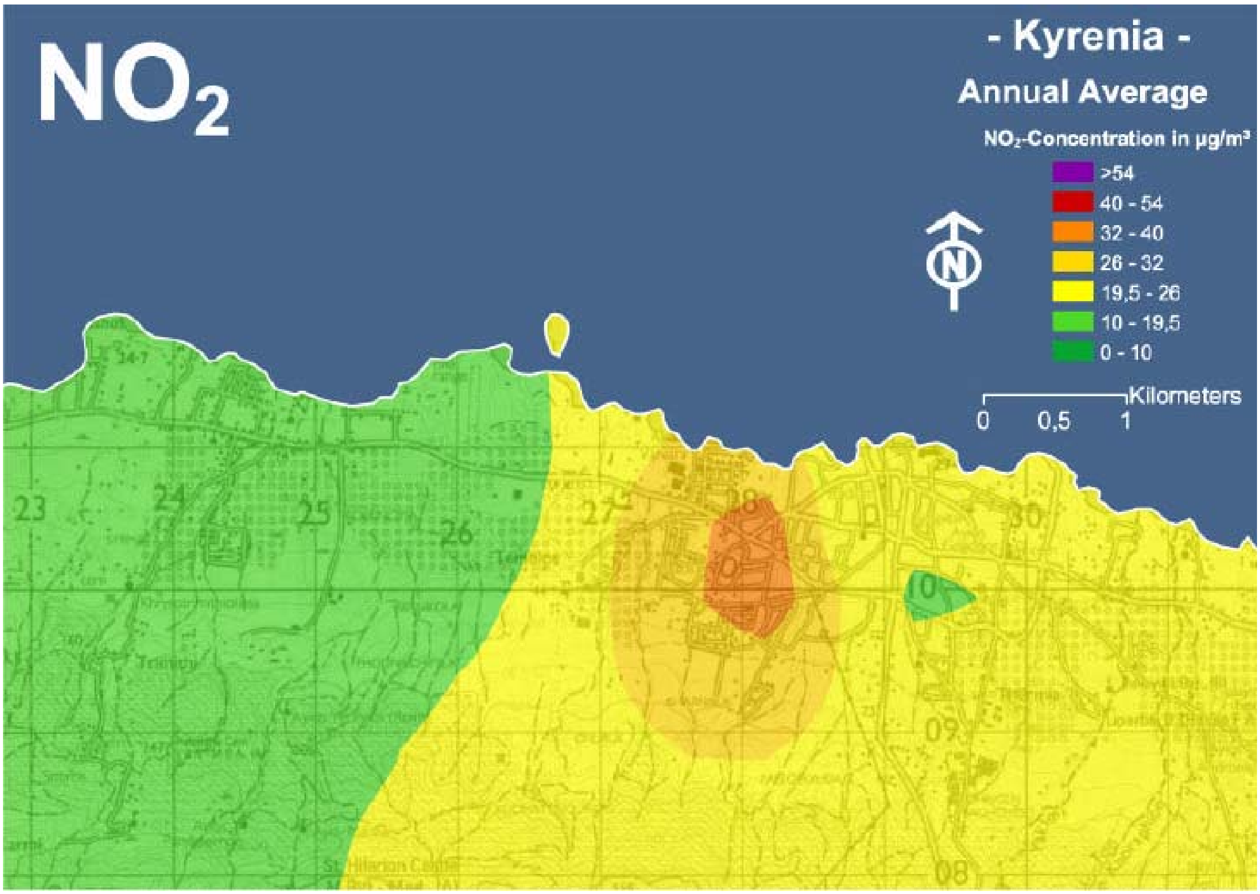
Annex B.5.. Mean annual interpolated NO_2 distribution in Limassol



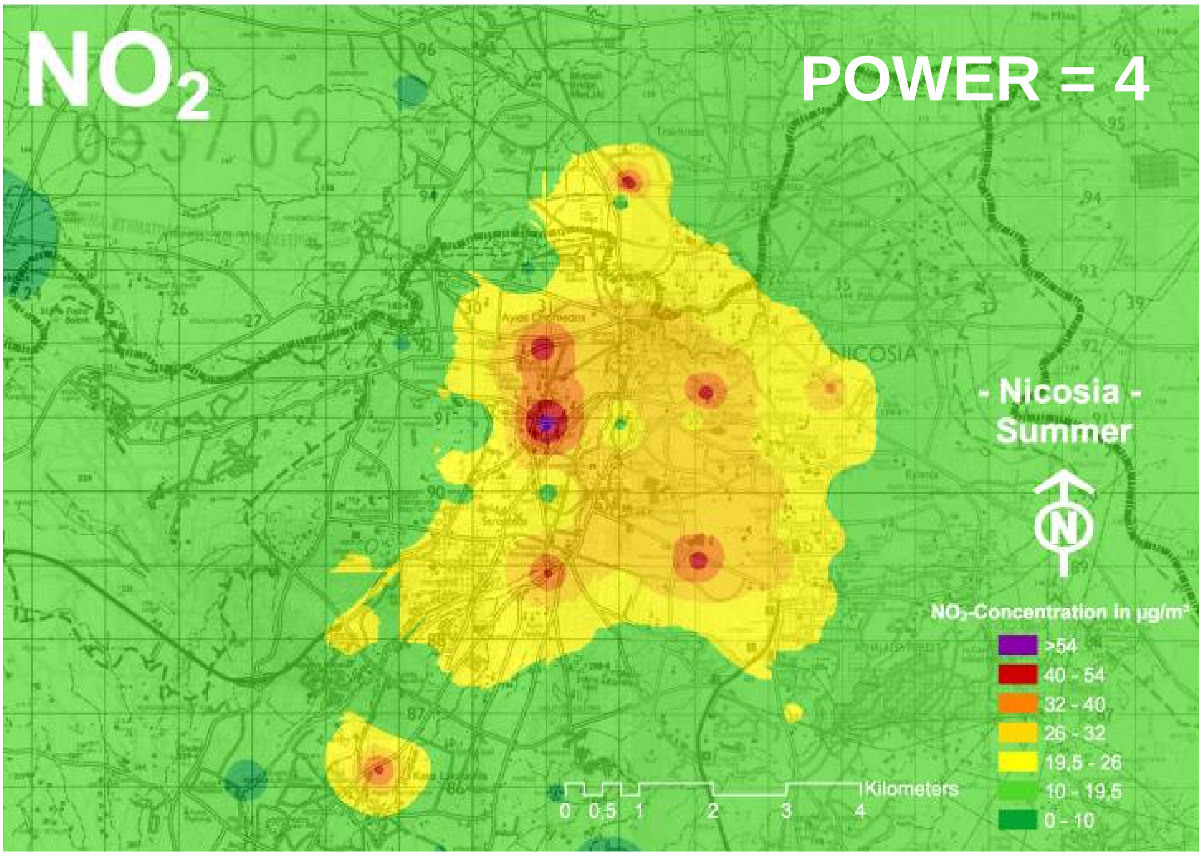
Annex B.6. Mean annual interpolated NO_2 distribution in Larnaca

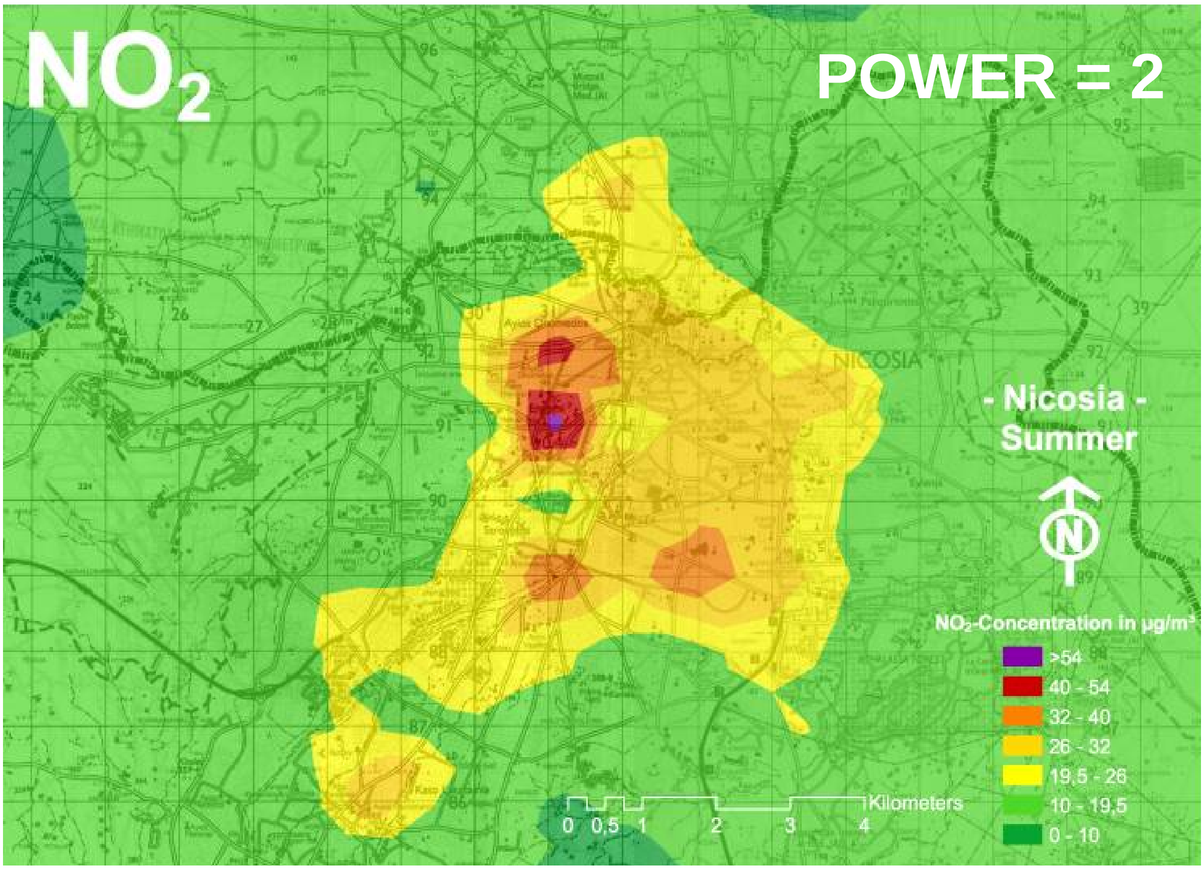


Annex B.7. Mean annual interpolated NO_2 distribution in Paphos

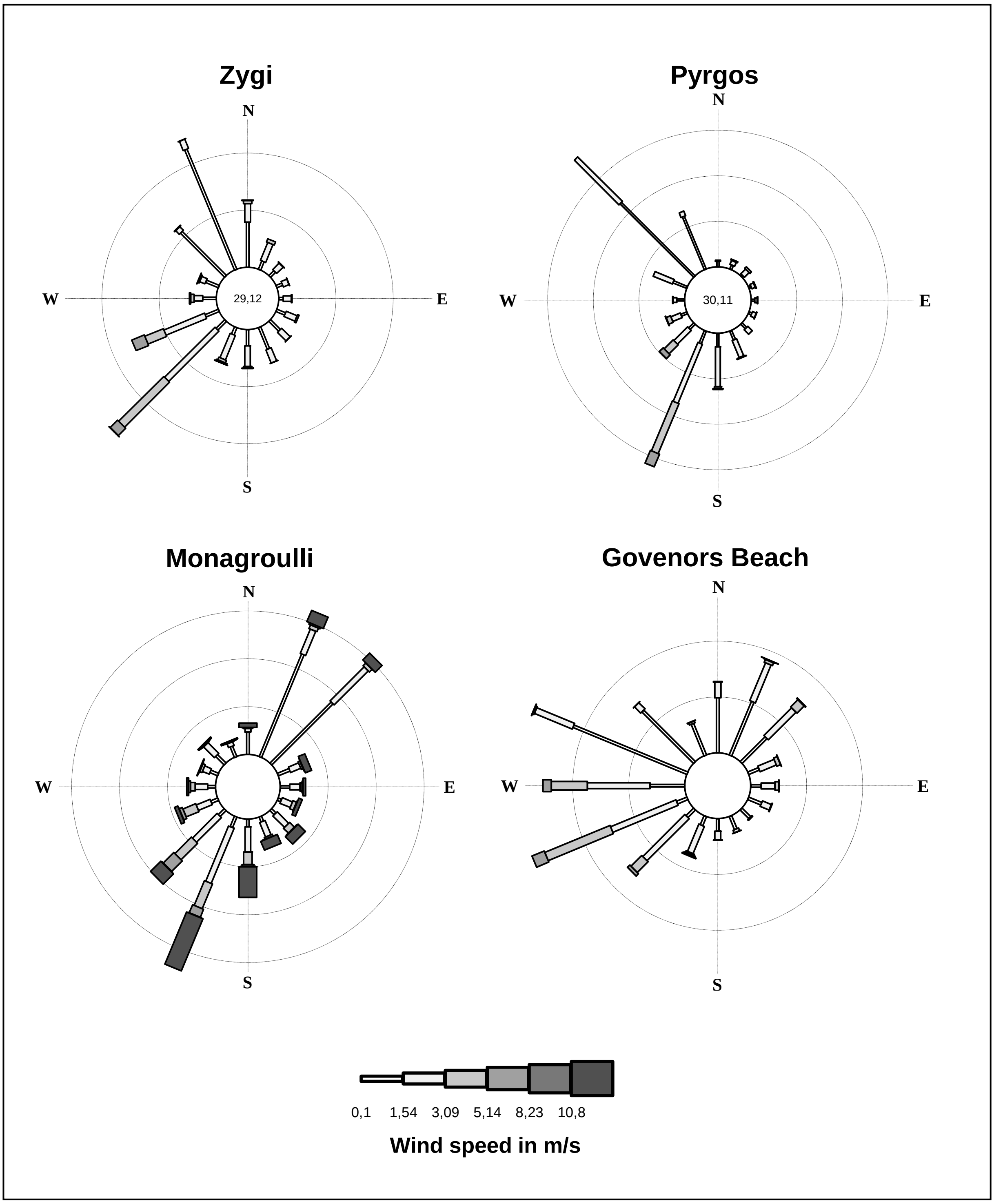


Annex B.8. Mean annual interpolated NO_2 distribution in Kyrenia

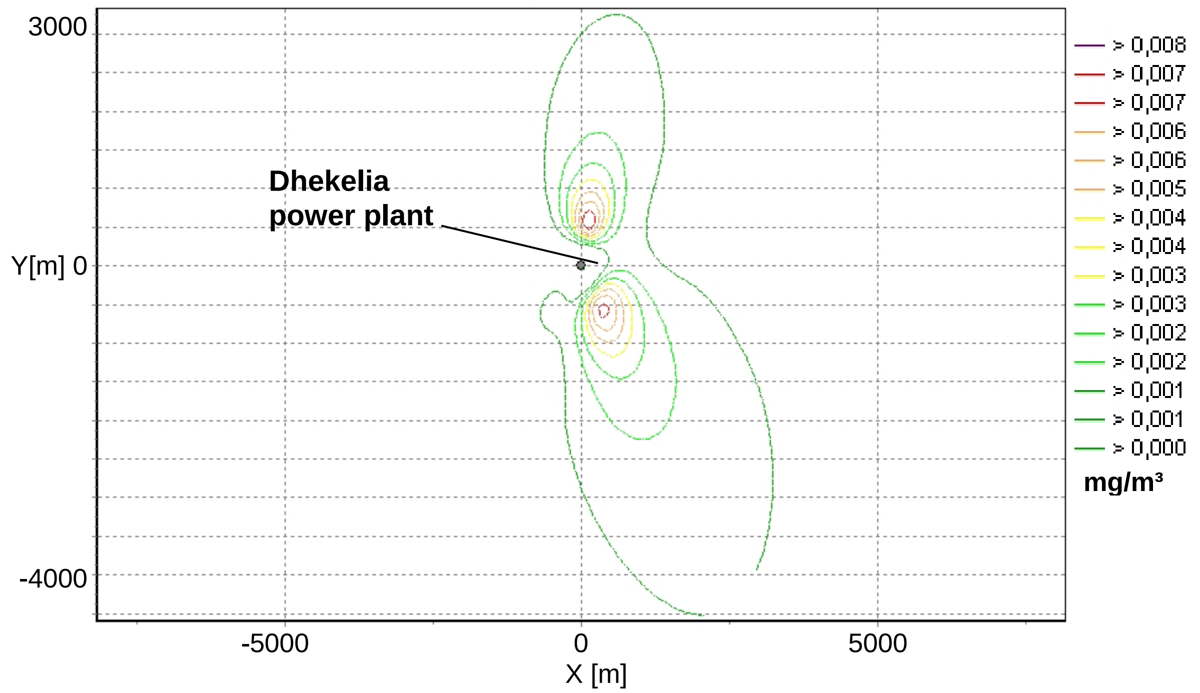




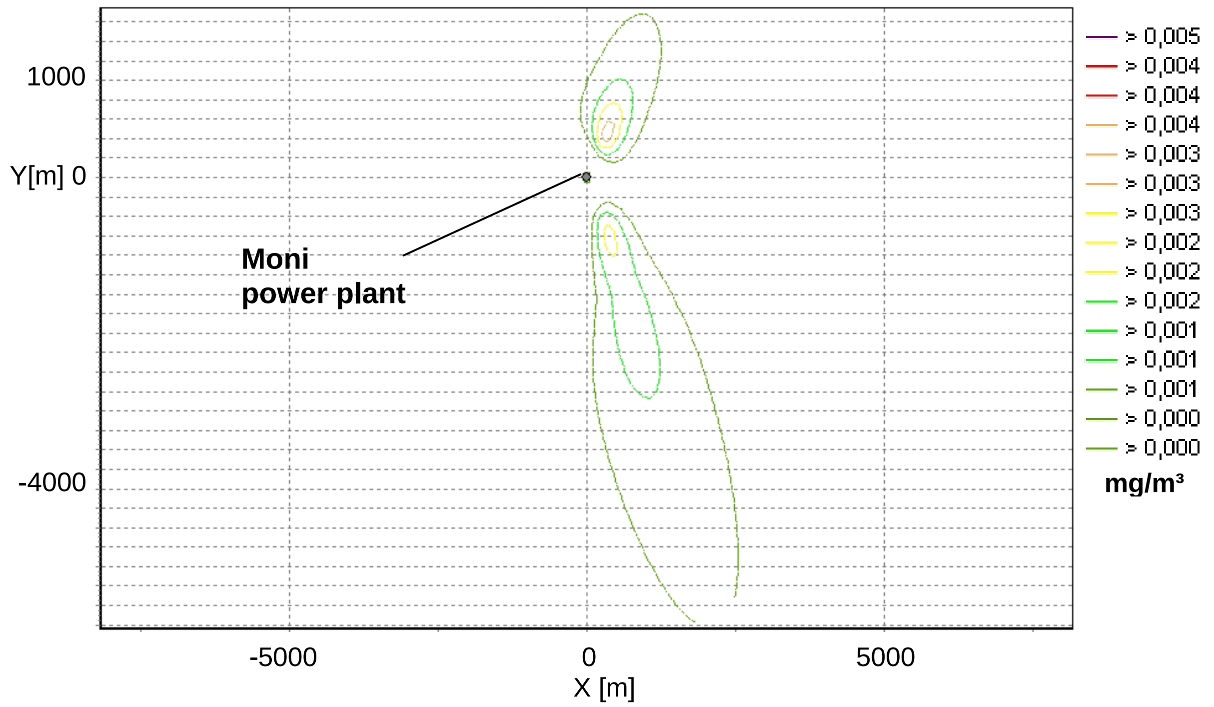
Annex B.9. Inverse distance to a power interpolation of NO_2 diffusive sampling results in Nicosia (summer average) using two different power values



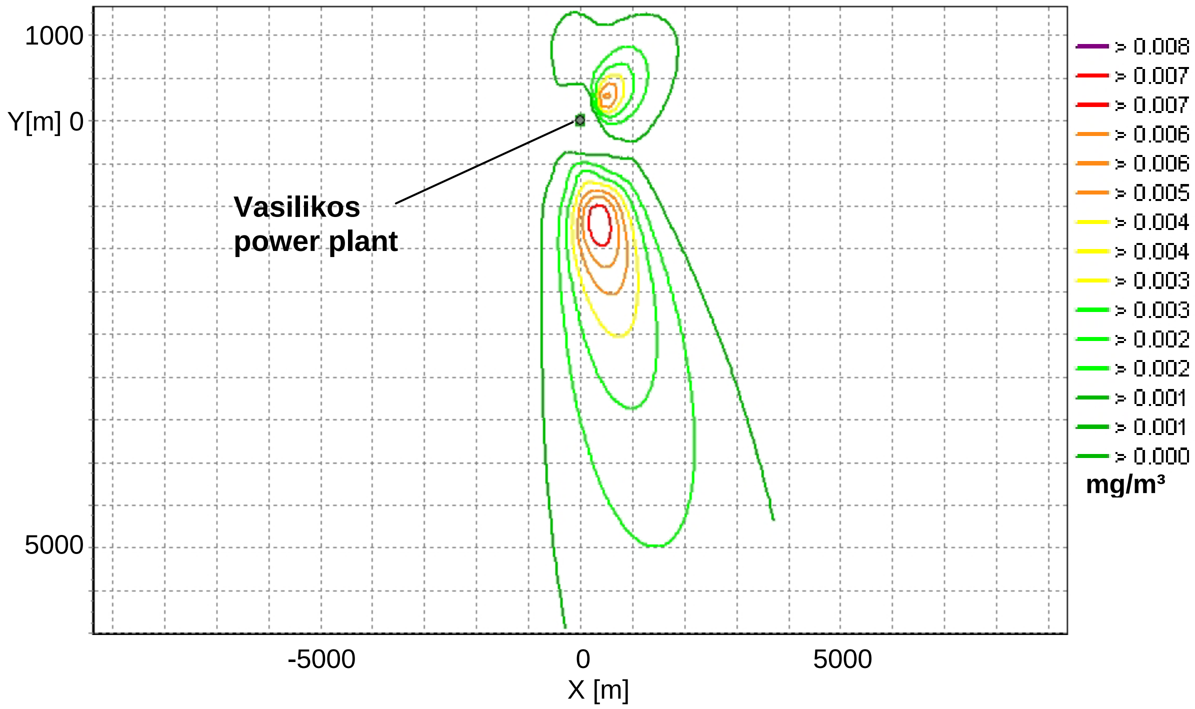
Annex B.10. Wind roses from selected sites in Cyprus at the south coast near important single sources - based on measurements of at least one year in 2001 to 2003



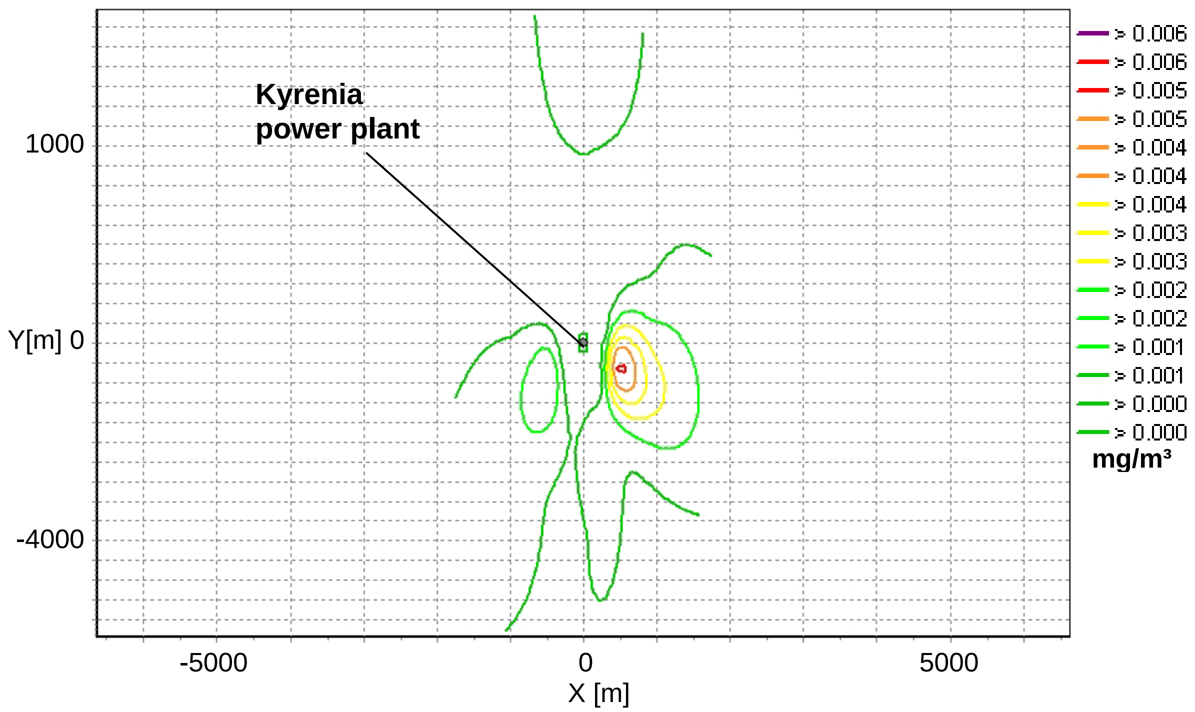
Annex B.11.. Dispersion modelling of the mean annual NO_x concentration for the power plant Dhekelia. Output of Gaussian dispersion model P&K 3782



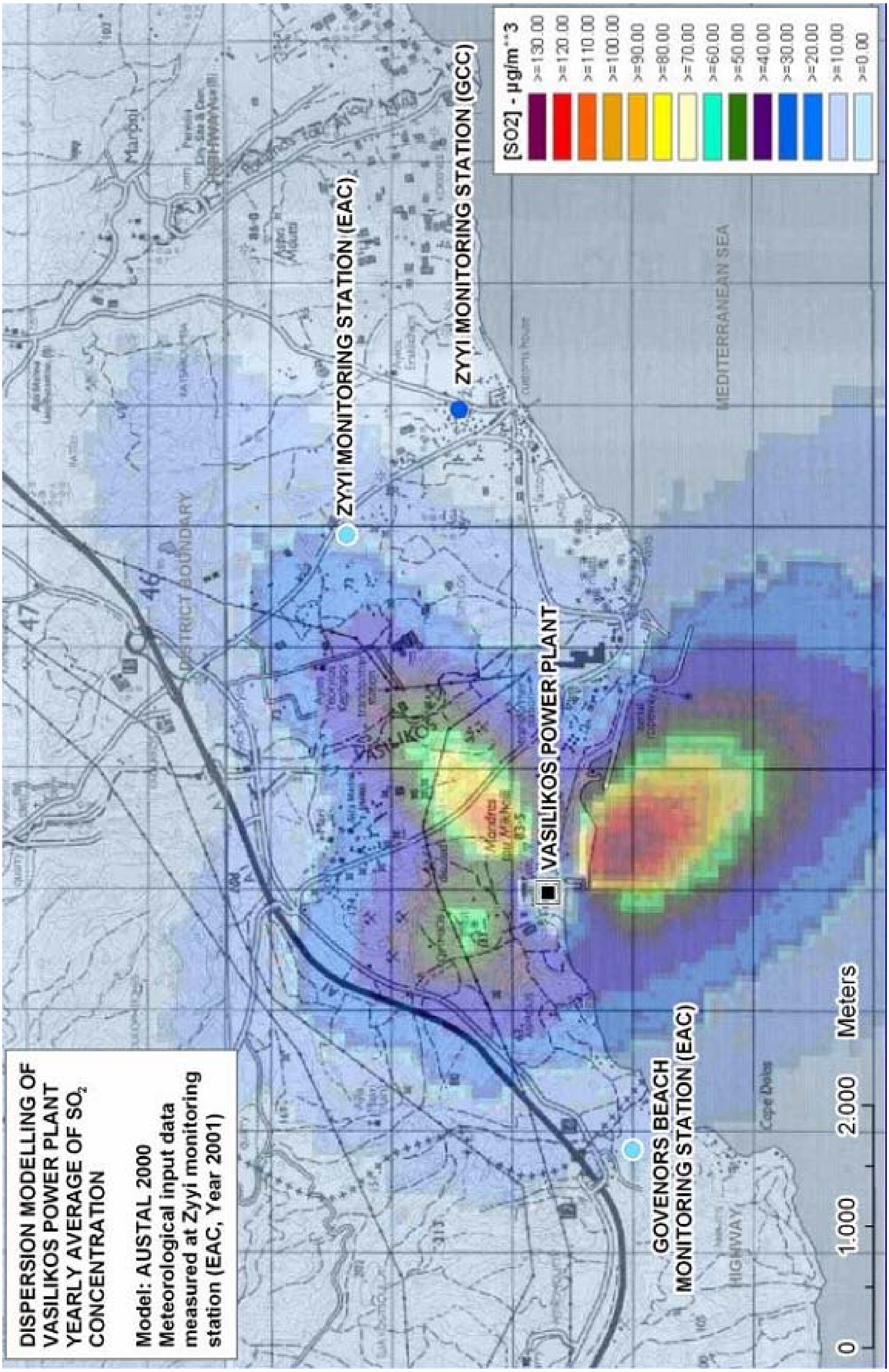
Annex B.12.. Dispersion modelling of the mean annual NO_x concentration for the power plant “Moni”. Output of Gaussian dispersion model P&K 3782



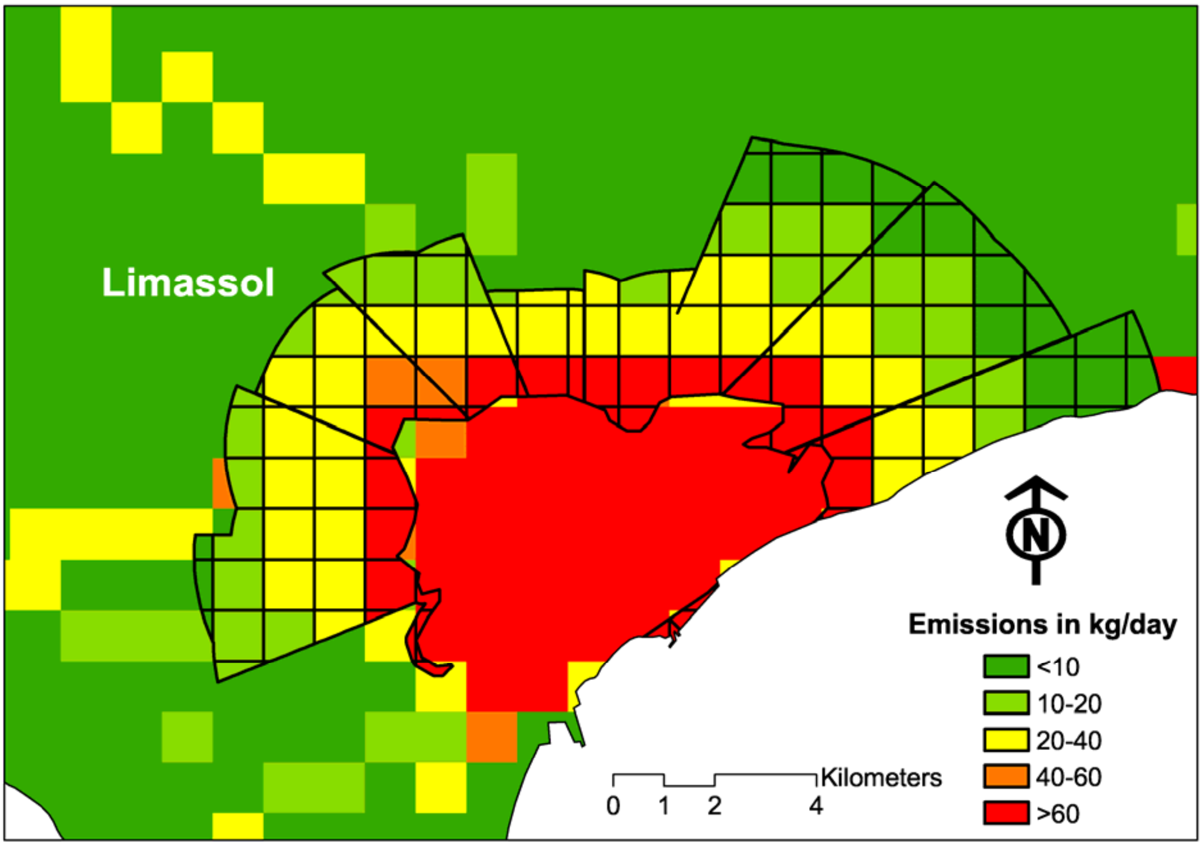
Annex B.13. Dispersion modelling of the mean annual NO_x concentration for the power plant "Vasilikos". Output of Gaussian dispersion model P&K 3782



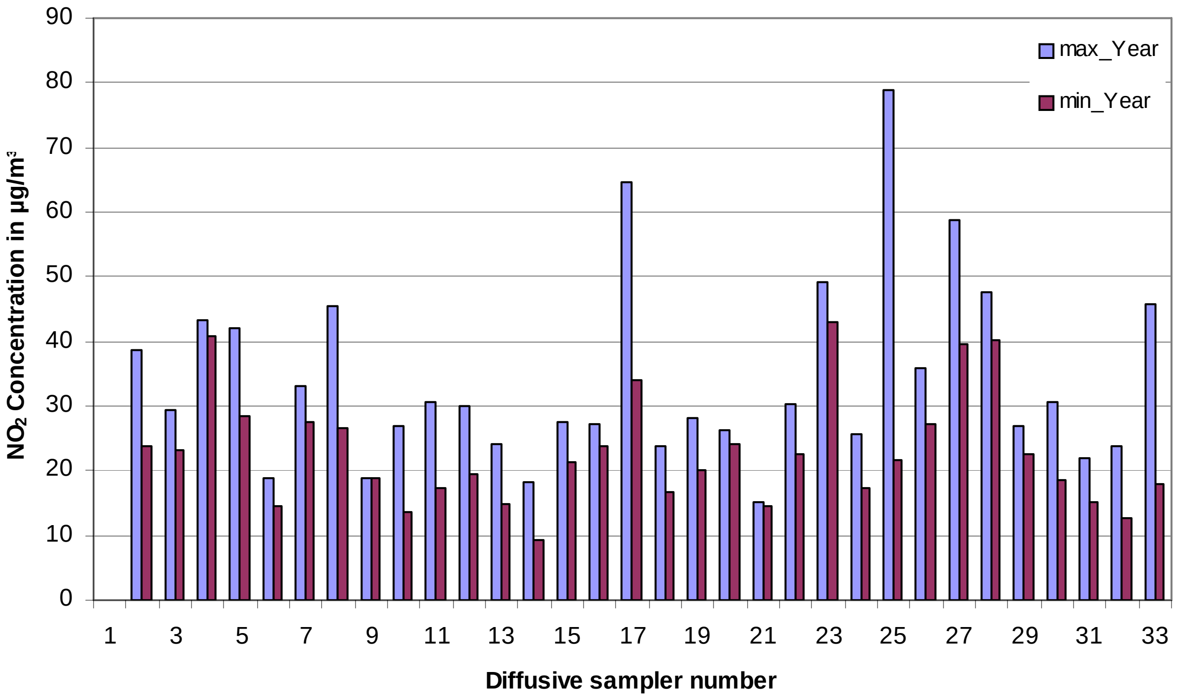
Annex B.14. Dispersion modelling of the mean annual NO_x concentration for the power plant "Kyrenia". Output of Gaussian dispersion model P&K 3782



Annex B.15. SO_2 dispersion modelling of the power plant near Vasilikos with the particle model Austal 2000 – annual average



Annex B.16. Dispersion plume of Limassol with real NOx emissions from emissions inventory and the calculated virtual emissions of the plume



Annex B.17. Value range of diffusive samplers that are placed in the same grid cell – base annual average of NO_2 concentration

8 References

[1] Seakins, P., Lansley, D., Hodgson, P. & Milne, P.: Evaluation of mobile NOx and Ozone measurements over domains ranging from 100m to 100km. Proceedings of the 4th international conference on urban air quality, Prague, University of Hertfordshire, University of Hertfordshire, 2003

[2] The council of the European Union: Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management. Official Journal of the European Union L 296, 1996

[3] Van Aalst, R. et al.: Guidance report on preliminary assessment under EC air quality directives. European Environment Agency, Technical report No. 11, Copenhagen, 1998

[4] VDI-Guideline 3787 Pt. 1: Climate and air pollution maps for cities and regions. VDI-Handbuch Reinhaltung der Luft, Beuth Verlag, Berlin, 1997

[5] Hartmann, U.: Erste Erfahrungen mit dem neuen Ausbreitungsmodell der TA Luft – Austal 2000. Immissionsschutz, 8. Jahrgang, Heft Nr. 3, Berlin, Bielefeld, München, 2003

[6] TA Luft (Technische Anleitung zur Reinhaltung der Luft): Erste Allgemeine Verwaltungsvorschrift zum Bundes-Immissionsschutzgesetz, Gemeinsames Ministerialblatt Heft 25 – 29, 2002

[7] Nagel, T., Flassak, T., Bächlin, W. & Lohmeyer, A.: Optimierung des Luftmessnetzes von Baden-Württemberg – Teil A Verfahren für Flächenhafte Immissionsdarstellung und Immissionsbezogene Klassifizierung. UMEG, Karlsruhe, 2002

[8] VDI-Guideline 3945 Pt. 3: Environmental meteorology. Atmospheric dispersion models. Particle model. VDI-Handbuch Reinhaltung der Luft, Beuth Verlag, Berlin, 1996

[9] Schädler, G., Lohmeyer, A., Bächlin, W. & Van Wees, T.: Vergleich und Bewertung derzeit verfügbarer mikroskaliger Ausbreitungsmodelle. Forschungszentrum Karlsruhe - Berichte Umweltforschung Baden-Württemberg, Karlsruhe, 1996

[10] Moussiopoulos, N.: Recent advances in urban air pollution research. Proceedings of the 8th International Conference on Environmental Science and Technology, University of the Aegean and Global Nest, Myrina, Lemnos island, Greece, 2003

[11] Louka P. & Moussiopoulos, N.: Optimisation of CFD modelling methods for traffic pollution in streets within TRAPOS research network. Proceedings of the 4th International Exhibition and Conference on Environmental Technology, Athens, Greece, 2003

[12] Bahmann, W., Schmonsees, N. & Oestereicher, R.: TA Luft 2002 – Neue Anforderungen an meteorologische Daten für Ausbreitungsrechnungen. Immissionsschutz, 8. Jahrgang, Heft Nr. 1, Berlin, Bielefeld, München, 2003

[13] VDI-Guideline 3782 Pt. 1: Environmental meteorology. Atmospheric dispersion models. Gaussian plume model. VDI-Handbuch Reinhaltung der Luft, Beuth Verlag, Berlin, 2001

[14] VDI-Guideline 3945 Pt. 1: Environmental meteorology. Atmospheric dispersion models. Gaussian puff model. VDI-Handbuch Reinhaltung der Luft, Beuth Verlag, Berlin, 1996

[15] Hout, D. et al.: Guidance on Assessment under the EU Air Quality Directives. European Environment Agency, draft, Copenhagen, 2000

[16] Barnes, R.: Variogram tutorial. Golden software Inc., Internet: www.goldensoftware.com, 2003

[17] Beier, R. & Doppelfeld, A.: Spatial Interpolation and Representativeness of Air Quality Data: An Intuitive Approach. International Conference Air Quality in Europe: Challenges for the 2000s, Stampe Violato, Padova, Italy, 1999

[18] Drüeke, M.: Die räumliche Modellierung bodennaher Ozonimmissionen im Mesoscale. Dissertation, Universität Trier, 1995

[19] Thoma, B.: Ermittlung der räumlichen Verteilung punktuell gemessener Immissionen in Baden-Württemberg - Ein Beitrag zur Regionalisierung von Luftschadstoffbelastungen. Diplomarbeit, Institut für Geographie, Universität Stuttgart, 1999

[20] Flassak, T., Nagel, T., Bächlin, W. & Lohmeyer, A.: Optimierung des Luftmessnetzes von Baden-Württemberg – Teil B Anwendung eines Hybridverfahrens Messung-Modellsimulation für das Untersuchungsgebiet Karlsruhe/Rastatt und Bewertung der Ergebnisse. UMEG, Karlsruhe, 2002

[21]European Commission: Overview of Methods and Results of the Preliminary Assessment of Air Quality in Europe under Directives 96/62/EC. DG Environment Unit C1 "Air and Noise", Internet: europa.eu.int/comm/environment/ air/pdf/reportpreliminaryasses.pdf, 2002

[22] Gerboles, M.: Application of artificial neural networks to the determination of the spatial distribution of ambient NO_2. International conference measuring air pollutants by diffusive sampling – Proceedings, 273-277, Montpellier, 2001

[23] Pellicioni, A., Tirabassi, T. & Gariazzo, C.: Improving of a dispersion model by neural network filter. Proceedings of the 4^{th} international conference on urban air quality, Prague, University of Hertfordshire, 2003

[24] Pellocioni, A., Tirabassi, T. & Gariazzo, C.: Coupling of neural network and dispersion models: A Novel methodology for air pollution models. 8^{th} international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes – Proceedings, National Institute of Meteorology and Hydrology, Sofia, 2002

[25] Agirre-Basurko, G., Ibarra-Barestegi, G. & Madariaga, I.: Regression and multilayer perceptron-based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. Proceedings of the 4^{th} international conference on urban air quality, Prague, University of Hertfordshire, 2003

[26] Omasreiter, H.: Neuronale Luftschadstoffmodellierung als Grundlage für ein immissionsgestütztes Verkehrsmanagementsystem. Doktorarbeit, Institut für Straßen- und Verkehrswesen, Universität Stuttgart, 2002

[27] Partanen, L. et al.: Inter-comparison of neural network, statistical and deterministic models for predicting the concentration of NO_2 and PM_{10} in urban air. Proceedings of the 4^{th} international conference on urban air quality, Prague, University of Hertfordshire, 2003

[28] Saral, A. & Ertürk, F.: Prediction of ground level SO_2 concentration using artificial neural networks. Water, Air and Soil Pollution - Focus 3, Netherlands, 2003

[29] Liguori, L.: Urban air pollution: Data interpretation and nowcasting by neural networks. Doctoral thesis, Instituto Veneto di Scienze, Venice, 2003

[30] Sachs, L.: Angewandte Statistik. Anwendung statistischer Methoden. 9^{th} edit., Springer Verlag, Berlin, Heidelberg, New York,1999

[31] Sachs, L.: Statistische Methoden. Planung und Auswertung. 7^{th} edit, Springer Verlag, Berlin, Heidelberg, New York, 1993

[32] Backhaus, K., Erichson, B., Plinke, W. & Weiber, R.: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. 9th edit., Springer Verlag, Berlin, Heidelberg, New York, 2000

[33] Baumbach, G.: Air Quality Control. , 3rd edit.; Springer Verlag, Berlin, Heidelberg, New York, 1996

[34] Baumbach, G. & Pfeiffer, H.: Preliminary Assessment of Ambient Air Quality in Cyprus. Final report. Department of Air Quality Control in the Institute of Process Engineering and Power Plant Technology, Universität Stuttgart, 2004

[35] Hadjigeorghiou, E.: Midday inversions over Athalassa. Summary of observations 1982-1989 at 11 U.T.C.. Meteorological Service – Meteorological note series No. 11, Nicosia, 1991

[36] Hangartner, M.: Influence of meteorological factors on the performance of diffusive samplers. International conference measuring air pollutants by diffusive sampling – Proceedings, Montpellier, 2001

[37] Deutsche Norm DIN EN 13528-3: Außenluftqualität – Passivsammler zur Bestimmung der Konzentration von Gasen und Dämpfen – Teil 3: Anleitung zur Auswahl, Anwendung und Handhabung, Kommission Reinhaltung der Luft im VDI und DIN – Normenausschuss, VDI/DIN-Handbuch Reinhaltung der Luft, Band 5, Beuth Verlag, Berlin, 2004

[38] Roth, C. & Bournel-Bosson, C.: Mapping diffusive sampling results: Including uncertainty and indirect information. International conference measuring air pollutants by diffusive sampling – Proceedings, Montpellier, 2001

[39] Rojas, R.: Theorie der neuronalen Netze. Eine systematische Einführung. Heidelberg, Springer Verlag, 1993

[40] Lippe, W.: Einführung in SoftComputing. Skript zur Vorlesung, Institut für Informatik, Westfälische Wilhelms Universität, Münster 2004

[41] Schneider, G.: Lebensdauerprognosen von SCR-Katalysatoren. Doktorarbeit, Institute of Process Engineering an Power Plant Technology, Universität Stuttgart, 1998

[42] Zell, A.: Simulation neuronaler Netze. Oldenburg Verlag, Oldenburg, 1997

[43] Box, G.E.P. & Jenkins, G.M.: Time series analysis. Forecasting and control. 1st edit., Holden Day, San Francisco, London, Sydney,1976

[44] Brosius, F.: SPSS 8.0. Professionelle Statistik unter Windows. 1st edit., MITP-Verlag, Bonn, 1998

[45] Schlittgen, R. & Streitber, B.: Zeitreihenanalyse. 7th edit., Oldenbourg Verlag, München, 2001

[46] Bäder, J.: Zusammenhänge zwischen Luftverunreinigungen und meteorologischen Parametern in der Region Stuttgart, Immissionsschutz, 6. Jahrgang, Heft Nr. 4, Berlin, Bielefeld, München, 2001

[47] Hartung, J.: Statistik. Lehr- und Handbuch der angewandten Statistik. 9th edit., Oldenbourg Verlag, München, Wien, 1993

[48] Qnet: Qnet2000. Manual of the software, Vesta Services 2004

[49] Mazzali, E., Angelino, E., Gerosa, G., Ballarin Denti, A.: Ozone Exposure (AOT 40) in a Complex morphology area with diffusive sampler data. International conference measuring air pollutants by diffusive sampling – Proceedings, Montpellier, 2001

[50] Eckstein, P.: Angewandte Statistik mit SPSS. Praktische Einführung für Wirtschaftswissenschaftler. 2nd edit., Gabler Verlag, Wiesbaden, 1999

[51] Johnson, R.A. & Wichern, D.W.: Applied Multivariate Statistical Analysis., 3rd edit., Prentice Hall, London, Sydney, Toronto, 1993

[52] Kukkonen, J. et al.: Evaluation of neural network, statistical and deterministic models against the measured concentration of NO_2, PM_{10} and $PM_{2,5}$ in an urban area. 8th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes – Proceedings, National Institute of Meteorology and Hydrology, Sofia, 2001

[53] Pfeiffer, H.: Methode zur Prognose von Luftverunreinigungen im Bereich verkehrsreicher Straßen. Diplomarbeit, Institut für Geologie und Paläontologie, Universität Stuttgart, 2001

[54] VDI-Guideline 3783 Pt. 6: Simulation of the wind field I. Regional dispersion of pollutants above complex terrain. VDI-Handbuch Reinhaltung der Luft, Beuth Verlag, Berlin, 1992

[55] Max Weishaupt GmbH: Weishaupt-Arbeitsmappe „Technik". Schwendi, 1970

[56] Pasiardi, S.: Statistical analysis of meteorological measurements in Cyprus. Meteorological Service – Meteorological note series No. 14, Nicosia, 1995

[57] Ferreira, J.: Modellierung der Ausbreitung von Abgasen industrieller Quellen in Zypern. Diplomarbeit, Institut für Verfahrenstechnik und Dampfkesselwesen, Universität Stuttgart, 2002

[58] Baumbach, G., Baumüller, J., Dröscher, F. & Reuter, U.: Lufthygenisches Gutachten für die Stadt Heilbronn. Amt für Straßenverkehr und Umwelt der Stadt Heilbronn, 1986

[59] The council of the European Union: Council Directive 1999/30/EC of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and nitrogen oxides, particulate matter and lead in ambient air. Official Journal of the European Union L 163/41.

[60]The council of the European Union: Council Directive 2000/69/EC of 16 November 2000 relating to limit values for benzene and carbon monoxide in ambient air. Official Journal of the European Union L 313/12.

[61] Nagel, T., Bächlin, W. & Lohmeyer, A.: Prognose der verkehrsbedingten Schadstoffbelastung im Zusammenhang mit der Planung „Stuttgart 21". Amt für Umweltschutz Abteilung Statdklimatologie - Untersuchungen zur Umwelt Heft 9, Stuttgart, 1996

[62] Zenger, A.: Atmosphärische Ausbreitungsmodellierung - Grundlagen und Praxis. 1st Edit.; Springer Verlag, Heidelberg, Berlin, New York, 1998

[63] Lutz, T.: Methode zur Erstellung eines Emissionskatasters für den städtischen Straßenverkehr in Zypern. Diplomarbeit, Institut für Verfahrenstechnik und Dampfkesselwesen, Universität Stuttgart, 2003

[64] Stull, R. B.: An introduction to boundary layer meteorology. Kluwer Academic Publisher, Dordrecht, Boston, London, 1988

Acknowledgement

Participants of the project “Preliminary Assessment of Ambient Air Quality in Cyprus”

UNOPS Coordinator:
Amelia de Coster

Project Partners:
GCC:
Savvas Kleanthous
TCC:
Ersever Beyaz
Aytekin Bagcier
Tolga Baki

Subcontractors / Consultants:
Medisell: George Orthodoxou
Ioannis Diakos
Lefteris Demetriou
Dragos Palanceann
Rasvan Flad
Michalis Spirou, Student
and other employees
Dr. Olga Poulida
Prof. Jürgen Baumüller
Dr. Reinhold Beier

IVD of University of Stuttgart
Prof. Günter Baumbach
Leire Sarachaga-Ruiz
Martin Winter
Patrick Mönckert
Ulrich Vogt
Dieter Straub
Anja Dreiseidler
Silke Drautz
Tina Lutz
Siegmar Parakennings
Lupin Hu
Mansoor Ahmad
Aynul Bari
Stefan Kiening
Yahefu Palida
Joao Ferreira
Andrea Krusch
Heike Grüner
Bernd Janisch
Heiko Pfeiffer
and others