DISSERTATION
SUBMITTED TO THE
COMBINED FACULTIES FOR THE NATURAL SCIENCES AND FOR MATHEMATICS
OF THE RUPERTO-CAROLA UNIVERSITY OF HEIDELBERG, GERMANY
FOR THE DEGREE OF
DOCTOR OF NATURAL SCIENCES

presented by
Diplom-Physiker Andreas Hartl
born in Menden

Oral examination: 5 February 2007

Tomographic Reconstruction of 2-D Atmospheric Trace Gas Distributions from Active DOAS Measurements

Referees: Prof. Dr. Ulrich Platt
Prof. Dr. Bernd Jähne

Zusammenfassung

Die Anwendung des tomographischen Prinzips auf die aktive DOAS Fernerkundung stellt eine neue Meßtechnik von atmosphärischen Spurengasverteilungen dar, für die analytische Standardmethoden zur Rekonstruktion von skalaren Feldern aus ihren Wegintegralen wegen der wenigen (10-50), unregelmäßig angeordneten Strahlen nicht anwendbar sind. Stattdessen wird die Verteilung durch eine begrenzte Zahl von lokalen, hier stückweise konstanten oder linearen, Basisfunktionen parametrisiert und das diskrete lineare inverse Problem durch einen Least Squares - Minimum Norm Ansatz gelöst.

Für räumlich stark begrenzte 2-D Konzentrationsspitzen wird in Abhängigkeit ihrer Ausdehnung gezeigt, wie die Rekonstruktion durch optimierte Wahl der Parametrisierung, d.h. Zahl und Art der Basisfunktionen, sowie des A Prioris erheblich verbessert werden kann. Die Regularisierung der Lösung spielt eine untergeordnete Rolle. Vorschläge zur Rekonstruktion von Konzentrationsspitzen durch Kombination verschiedener Parametrisierungen werden systematisch untersucht, wobei deren Erfolg stark von der am meisten interessierenden Eigenschaft der Verteilung abhängt. Vergleich verschiedener 2-D Strahlgeometrien ergibt, daß lineare Unabhängigkeit des resultierenden Systems entscheidend ist.
Eine detaillierte Analyse des Rekonstruktionsfehlers stellt Besonderheiten der Tomographie mit wenigen Strahlen heraus und argumentiert, daß dessen Abschätzung ohne A Priori nicht möglich ist. Unter dieser Voraussetzung wird ein numerisches Schema zur Berechnung des Rekonstruktionsfehlers vorgestellt.
Die Methoden werden auf ein Innenraumexperiment zur Simulation von Emissionsfahnen, sowie auf 2-D Modellverteilungen über einer Straßenschlucht angewendet, wobei für letztere gezeigt wird, wie Modellevaluation trotz weniger Lichtstrahlen möglich sein kann. Anders als für räumlich begrenzte Maxima besteht hier starke Abhängigkeit vom Grad der Regularisierung.

Abstract

Applying the tomographic principle to active DOAS remote sensing leads to a novel technique for the measurement of atmospheric trace gas distributions. Standard analytical methods for the reconstruction of a scalar field from its line integrals cannot be used due to low numbers of light paths (10-50) and their irregular arrangement, so that the concentration field is expanded into a limited number of local (piecewise constant or linear) basis functions instead. The resulting discrete linear inverse problem is solved by a least squares-minimum norm principle.

For sharp 2-D concentration peaks it is shown systematically with respect to their extension how the optimal choice of parametrisation (in terms of number and kind of basis functions) and a priori can tremendously improve the reconstruction. Regularisation plays a minor role. Proposals for retrieving peak distributions by combining different parametrisations are again examined systematically showing that their usefulness heavily depends on the features one is most interested in. Comparison of different 2-D light path geometries reveals that linear independency within the associated systems is pivotal.
A detailed analysis of the reconstruction error points out special issues of tomography with only few integration paths and argues that a complete error estimation is not possible without a priori assumptions. Based on this discussion a numerical scheme for calculating the reconstruction error is suggested.
The findings are applied to an indoor experiment simulating narrow emission puffs and 2-D model distributions above a street canyon, respectively. For the latter case it is demonstrated how model evaluation can be possible even with a relatively small number of light paths. Contrary to the reconstruction of peak distributions regularisation becomes crucial.

Contents

I. Basics 18

Symbols, Notation and Acronyms

$A = (\mathbf{a}_1, \dots, \mathbf{a}_m)^T$	system matrix of the discretised tomographic problem with column vectors $\mathbf{a}_i$	p. 54
$c_{(i)}(\mathbf{r}, t)$	atmospheric concentration field (of chemical species i)	p. 25
$\Delta\hat{c}(\mathbf{r}, t)$	reconstruction error field	p. 86
$\Delta\hat{c}_{l,u}(\mathbf{r}, t)$	lower/upper bounds of the reconstruction error field	p. 85
$\Delta\hat{c}_{\pm}(\mathbf{r}, t)$	positive/negative reconstruction error (under-/overestimation by the reconstruction)	p. 94
$\mathbf{d} = (d_1, \dots, d_m)^T$	vector of column densities ('data vector')	pp. 49, 57
$\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_m)^T$	measurement error	pp. 50, 90
$\mathcal{E}$	set (ensemble) of concentration fields	pp. 93, 107
i	*usually* index in discrete data space (light paths)	
I	index for concentration fields of a set $\mathcal{E}$ *or* index of subgrids for grid combination schemes	p. 103
j	*usually* index in discrete state space (parameters of the discretisation)	
J	index of grid knots for grid combination schemes	p. 103
k	iteration number	
l	*usually* length of reconstruction area	
L	*usually* light path length	
m	dimension of data space (number of light paths)	
M	number of subgrids for grid combination schemes	p. 103
n	dimension of state space (number of parameters of the discretisation)	
N	number of grid knots for grid combination schemes	p. 103
R	resolution matrix	p. 69
$S_{\mathbf{x},\boldsymbol{\epsilon},a}$	covariance matrix of the a posteriori, measurement error and apriori	pp. 65,72
S_j	column density from DOAS fit for chemical species j	p. 42
$\sigma_{x,y,z}$	plume dispersion coefficients	p. 28
$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_{\mathrm{rank}[A]}, 0, \dots)$ $\Sigma_r = \mathrm{diag}(\sigma_1, \dots, \sigma_{\mathrm{rank}[A]})$	matrix of singular values σ_i	p. 60
$U = (\mathbf{u}_1, \dots, \mathbf{u}_m)$	matrix of singular vectors $\mathbf{u}_i$ associated with the data space	p. 60
$V = (\mathbf{v}_1, \dots, \mathbf{v}_n)$	matrix of singular vectors $\mathbf{v}_j$ associated with the state space	p. 60
$\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$	vector of basis functions ϕ_j	pp. 54,57
$\mathbf{x} = (x_1, \dots, x_n)^T$	vector of discretisation parameters x_j ('state vector')	pp. 54,57
$\mathbf{x}_a$	a priori of the state vector	pp. 66,72
$\mathbf{X} = (X_1, \dots, X_J)^T$	vector of concentration values on knots for grid combination schemes	p. 103

$\langle\cdot\rangle$	average (of concentration fields)	
$\overline{(\cdot)}$	temporal or spatial mean	
	or implicit transformation by (S)ART, SIRT	p. 78
$\widehat{(\cdot)}$	reconstructed quantities	
$(\cdot)'$	turbulent fluctuations	p. 25
	or components in the system of the singular vectors	p. 79
$\widetilde{(\cdot)}$	transformation to diagonal covariance matrices	p. 73
BG	background	
LP	light path	
NMB	normalised mean bias (accuracy of reconstructed/modelled spatial mean concentrations)	p. 97
$NEARN$	nearness (specially normalised root mean square error of the reconstructed concentration field)	p. 96
PPA	peak prediction accuracy (accuracy of reconstructed/modelled peak maxima)	p. 97
$PIPA$	peak integral prediction accuracy (accuracy of reconstructed total emissions)	p. 97

1. Introduction

Trace gases form only a tiny part of ambient air – their volume fractions range from less than 10^{-9} to some 10^{-6} – but their actual amounts can be both vital and threatening to the life-forms of our ecosystem. On the one hand, it is their direct impact on the radiation budget of the earth which makes them so important, the so-called greenhouse effect, e.g. of water vapour, CO_2, O_3 or CH_4. On the other hand, it is their chemical reactivity which, besides making themselves or their reaction products a risk to creatures and plants, can change the composition of the atmosphere and thus indirectly affects its natural radiation balance. All these effects could or still can be observed since anthropogenic activities have started to increase trace gas levels locally or globally beyond their natural values: smog caused by harmful amounts of industrially emitted SO_2, photochemical smog as a noxious mixture of trace gases such as NO_x, O_3 and VOCs (volatile organic compounds), acid rain formed from emissions of SO_2 and NO_x, destruction of the ozone layer, e.g. by CFCs (chlorofluorocarbons), and finally – with deepest impact and on a global scale – the anthropogenic greenhouse effect. It has taken decades to understand some of these phenomena because of their complex chemistry. In view of the rapid industrialisation taking place in some densely populated parts of the world and with the most serious consequences of the anthropogenic greenhouse effect yet to emerge, it is beyond doubt that measurement and understanding of atmospheric trace gas concentrations is of paramount interest. In contrast to the principal constituents of the air, these concentrations depend on emission sources and the chemical environment, so that, in fact, we are speaking of trace gas concentration *distributions*, varying spatially and temporally.

Depending on the trace gas species and the part of the atmosphere one is concerned with, the concentration distributions are more or less well understood – or relevant, so that motives to deal with them may be, e.g., to understand their chemistry, to allocate emissions, for pure monitoring purposes or to evaluate models and so forth. These reasons have gained significance on a wide range of spatial scales. For example, global CO_2 distributions in the context of worldwide emission trading have to be monitored just as air pollution at different sites on a communal level. Global chemical transport models designed to ultimately predict future trends have to be validated by measurements just as disperion models allowed by EU regulations for monitoring air pollution on a local scale down to traffic hot spots [e.g. *Schatzmann et al.*, 2006; *Trukenmüller et al.*, 2004].

Measurements of trace gas concentrations are either *in situ*, i.e. local, or *remote sensing*, which here means using the interaction of molecules with electromagnetic waves to infer atmospheric parameters along the propagated path. A single measurement of the former kind provides the value of the concentration field at the measurement location.[1] A single measurement of the latter kind gives a line integral along the propagated path over a functional of the concentration field – ideally over the field itself. Either method bears more or less apparent drawbacks. Interpreting in situ measurements

---

[1] Measurements that restrict the wave propagation to a small volume by multiple reflections are regarded as point measurements in this context.

always raises the question of their representativeness, especially if the variability of the concentration field is unknown. Remote sensing methods give information only on line integrals involving the desired concentration distribution.

While getting spatial information from point measurements simply needs the experimental act of taking enough samples, in order to obtain the same from any number of remote sensing measurements additionally requires mathematical techniques to retrieve the concentration field from the integrals. The combination of taking the right remotely sensed samples and subsequent inversion techniques to unravel the integrals is commonly referred to as *tomography*. Naturally, it has become an extremely elaborate technique in disciplines where taking samples of the object of investigation is impossible or prohibitively expensive. Medical tomography benefits from small regions of interest and it is relatively easy to handle source-detector configurations, so that large numbers (of the order 10^5) of integrals can be measured. Huge numbers of data ($10^6 - 10^8$) for acoustic travel time tomography of the earth's interior are conveniently provided by earthquakes or explosions.

There is a variety of spectroscopic techniques applied to the remote sensing of trace gases in the atmosphere, differing either in the wavelength range or the physical principle they exploit for detection, with natural or artificial radiation sources (*passive* or *active* remote sensing). The most important ones are *infrared spectroscopy* (IR), the *light detection and ranging* (LIDAR) technique and the *differential optical absorption spectroscopy* (DOAS). Although these methods are able to detect various trace gases with high sensitivity for certain spatial ranges, in general, they have not been used for tomographic measurements of atmospheric trace gas distributions. *Byer and Shepp* [1979] put forward the idea to reconstruct 2-D atmospheric trace gas distributions near ground from tomographic absorption LIDAR measurements with a rather elaborate setup. About the same time *Fleming* [1982] proposed tomographic satellite measurements of IR and microwave emissions for 2-D reconstruction of temperature and trace gas distributions by increasing the number of viewing directions from one to five. While the latter proposal for passive measurement has been realised in a smiliar manner in aircraft DOAS measurements using sunlight [*Heue*, 2005], so far no active experiment has been performed in the atmosphere that comes near the number of integration paths considered by *Byer and Shepp* [1979]. Comparable optical setups for 2-D tomographic reconstruction of indoor air contaminants have been studied mostly in simulations – light path numbers in actual gas chamber experiments using the Fourier transform IR (FTIR) spectroscopy were between around 30 and 200.[2] Active DOAS tomographic measurements along 16 light paths were used by *Laepple et al.* [2004] to reconstruct the NO_2 distribution perpendicular to a motorway. *Belotti et al.* [2003] reconstruct volcanic CO_2 from horizontal open-path IR measurements along 15 paths.

On the one hand, active DOAS measurements allow *precise* detection of a variety of important trace gas species for light paths between a few and up to $10 - 20\,km$ *at the same time.* On the other hand, almost 30 years after first proposals of tomographic remote sensing, this technique still has not been used for genuine tomographic DOAS measurements – motivation enough to advance both experimental techniques and theoretical insight. As a theoretic contribution to these efforts, this thesis investigates the possibility to reconstruct 2-D distributions of tropospheric trace gas concentrations from active DOAS measurements along a moderate number of about 10 to 40 light paths. The mathematics of *computerised tomography* (CT) with regular geometries and large numbers of integration paths, as they appear, for example, in medical image processing, is well understood. The application at hand –

---

[2] See sec. 3.3 for references.

with only few light paths which are unlikely to be arranged in a regular fashion – requires consideration beyond the existing literature because

- analytic inversion methods that are standard in CT cannot be applied (due to the irregular geometry),
- estimation of the error of the reconstruction result is not standard and contributions that are negligible elsewhere become important,
- smooth atmospheric distributions may differ from absorption patterns in the human body or layer structures in the earth,
- disciplines like computerised tomography, geophysics and atmospheric profiling have developed powerful concepts for their inverse problems and sometimes quite different retrieval techniques. It is not always obvious why this is the case and which methods can be adopted. Existing studies with small numbers of integration paths are scarcely systematic.

The lack of a 'standard literature' that our methods could refer to is reflected by the structure of this work, being split into three parts. The *first part* reviews basic facts on trace gases, their remote sensing and the mathematics of tomography, while the *second part* contains the methodology – partly developed in this thesis – for the calculation of the reconstruction error, for the reconstruction itself and the planning of a tomographic experiment. It also contains all analytical results. The *third part* consists in numerical applications.

Not only is there no standard approach to atmospheric tomography, emissions, transport and chemical transformation make the situation also far too complex to admit a standard picture of atmospheric trace gas distributions. Therefore, after presenting some of the most important trace gases relevant for DOAS, *chapter* 2 considers two scenarios of trace gas distributions that will be relevant for the third part: dispersion from a point source and the complexity of urban trace gas distributions. *Chapter* 3 gives an introduction to the tomographic principle and the most important remote sensing methods that potentially could make use of it. The DOAS technique is discussed in more detail. Finally, previous tomographic measurements of atmospheric parameters are presented and put into the context of standard tomographic applications. *Chapter* 4 justifies the discretisation and reformulation of the tomographic reconstruction problem as a least-squares least-norm problem. Apart from the very common box discretisation, linear parametrisation is considered for the 2-D case. Special attention is paid to the question how known instabilities of this problem (so-called *ill-posedness*) affect different kinds of solutions. The optimal estimate that is commonly used to solve inverse problems in atmospheric science is presented both as a regularisation method (to stabilise the least-squares solution) and as a probabilistic approach. The iterative solution of the least-squares least-norm problem that has been adopted from computerised tomography is discussed in comparison with other iterative algorithms before the chapter closes with brief sketches of alternative reconstruction principles.

The detailed discussion of the reconstruction error in *chapter* 5 includes the error related to the finite representation of the solution that is usually not covered in the tomographic literature. We present a method for numerical estimation of the reconstruction error from test distributions that differs in several points from the one employed by *Laepple et al.* [2004]. The final section reviews and compares quality criteria used in image processing and atmospheric modelling. The discussion up to then will be irrespective of the concentration distributions or the dimensions of the reconstruction. The schemes proposed in *chapter* 6 use several grids for parametrisation and reconstruction and are especially intended for 2-D reconstruction of peak distributions as they arise from point sources. The discussion of a priori information and background concentrations is also specific for atmospheric trace gas distributions. *Chapter* 7 deals with experimental design issues like the signal to noise ratio, including additional point measurements and how tomographic measurements can be used to evaluate model results.

Concepts and methods of the first two parts are applied to the scenarios introduced in chapter 2: to 2-D Gaussians peaks representing sections through atmospheric emission puffs (*chapters* 8 & 9) and to model distributions of a highly polluted city street canyon (*chapter* 10). The simulations of *chapter* 8 investigate the 2-D reconstruction of concentration peaks systematically with respect to the parametrisation of the problem and the reconstruction scheme. Two iterative algorithms are compared to study the effect of regularisation. Finally, different kinds of path geometries are investigated in detail and the role of background concentrations is discussed. *Chapter* 9 applies the simulation results to data from a tomographic indoor experiment [*Mettendorf et al.*, 2006] and contains a detailed estimation of the complete reconstruction error. The final *chapter* 10 highlights the challenge of reconstructing trace gas concentration fields in a complex urban environment by means of high resolution model distributions and shows how model evaluation could be carried in a real experiment even with low spatial resolution of the reconstructed distributions.

Part I.

Basics

2. Trace Gas Distributions in the Atmosphere

This chapter starts with a brief motivation of the experimental efforts to measure the spatial distribution of trace gases. The discussion is not as general as the chapter title may suggest but, like the complete work, it is restricted to near ground trace gases detectable by DOAS and to spatial (and temporal) scales relevant to a special kind of this technique, the long-path DOAS, i.e. scales from a few $100\,m$ up to around $10\,km$. The remaining sections try to give an idea of how distributions actually arise on these scales from chemical and – maybe more important – transport processes. We concentrate on two special scenarios. Point emission (sec. 2.4) is important for practical applications like stack emission from industrial complexes or power plants or, in fact, for any distinct source. But because it is such a simple scenario, it is also ideal to study the nature of turbulent concentration fluctuations. Urban trace gas distributions, on the other hand, are of great public interest but far too complex to allow a universal picture. Therefore sec. 2.5 highlights some aspects of urban model development. The reconstruction of point source emissions will be the subject of simulations in chap. 8 and of a tomographic experiment in chap. 9, while results from the model system M-SYS designed for urban micro environments will be used in chap. 10.

2.1. The need for measured trace gas distributions

Apart from pure scientific curiosity directly measured trace gas concentration distributions are desirable in the following situations:

- Emission sources are unclear.
- Chemical transformation processes are not well known.
- For model evaluation and input.
- For air pollution monitoring.

The first point can mean that a source is suspected but not identified or characterised for all atmospheric conditions. Examples might be the yet not completely understood chemistry of halogen oxides in maritime areas [*Von Glasow et al.*, 2004; *Peters et al.*, 2005], the allocation of biogenic emissions [*Kesselmeier and Staudt*, 1999] or exploration of geological sources such as volcanoes or lava flows [*Bobrowski et al.*, 2003]. As emission into the atmosphere always implies transport, the easiest way to get a comprehensive picture of emission sources is to produce a concentration map for the region of interest. A second reason for unclear emissions might be that, although the source is known, it may be hard to measure directly under real world conditions. This is the case for traffic emissions on a motorway [e.g., *Corsmeier et al.*, 2005*b*], take-off and regular emissions of air planes are another example.

Certainly, the chemical transformation of reactive species is extremely complex and can be studied in the laboratory only to a limited extent, especially if precipitation or surfaces play a role. An important area of ongoing research is dealing with the details of the chemistry of photo smog as it occurs in heavily polluted urban areas and cities during summer. The situation concerning both the emissions and the effects of transport and turbulent mixing are too complicated to be simulated in a smog chamber. Complicated flow patterns in urban areas due to buildings make allocation of point-like sampled concentration values extremely difficult and time series of concentration fields combined with wind data would increase the information content tremendously.

The verification of models that – in their most complex form – are designed to simulate emission, transport and chemical evolution of trace gases and particles on different atmospheric scales is important. Even if emissions and reaction paths are modelled with sufficient precision there still remains the transport calculation. While describing dispersion of pollutants over even terrain and on mesoscales ($\sim \mathcal{O}(10^2)\,km$) might be a relatively easy task if one does not look into spatial details, complex terrain and flow patterns on microscales ($\sim \mathcal{O}(10^2)\,m$) are more of a problem both conceptually and computationally. Boundary conditions for pollution transport models are pre-processed by meteorological models. But while nobody will ask what the weather is like next street, concentration variations on a street scale are important, for example when it comes to accumulation of traffic emissions in a street canyon or whether the NO_x plume of a power plant meets a near surface VOC plume and locally forms ozone or not. In view of the fact that building effects can change concentration values by an order of magnitude [*Schatzmann and Leitl*, 2002], both meteorological and chemical transport models have to meet rather challenging requirements on urban scales (see sec. 2.5). The experimental proof that a model predicts concentrations with enough accuracy is thus indispensable and itself an intricate matter as discrepancies have to be traced back to either wrong emission data, incomplete chemical mechanisms, faulty transport calculations or insufficient parametrisation of atmospheric turbulence. Point measurements are in general not suited due to high concentration fluctuations both with time and location whereas concentration maps with sufficient temporal and spatial resolution give information on the location of emissions, their transport and concentration variability. The discussion of model evaluation will be resumed in sec. 7.4.

Addressing the last point, air pollution monitoring, it can be remarked that despite technical progress in reducing *emissions factors* – that is the average amount of a pollutant per amount of fuel or for a specific process – for example, by improved filtering of industrial discharge or by catalytic converters for cars, absolute amounts of emission can still be high in regions with extensive industry. They can be even growing if the absolute numbers of cars increase to a higher degree or if countries radically intensify their industrial production. Within the EU, air pollution and its monitoring on a national level is regulated by EU laws. As a matter of fact, the European Air Quality Guideline 96/62/EU and its daughter directives reach down to the municipal level by requiring concentration maps from members not only for the state in total and larger conurbations (more than 250 000 habitants) but also for heavily polluted micro environments [*Schatzmann et al.*, 2006]. If provided by point measurements, the EU directives regulate that measurements at urban background sites should be representative for an area of several km^2 and at traffic-orientated sites for at least $200\,m^2$ [*Trukenmüller et al.*, 2004].

pollutant	quality indicator	model quality objective
SO_2, NO_2, NO_x	hourly mean	50-60%
	daily mean	50%
	annual mean	30%
Ozone	$8\,h$ daily maximum	50%
	$1\,h$ mean	50%
Benzene	annual mean	50%
CO	$8\,h$ mean	50%

Table 2.1: Modelling quality objectives established by the EU directives 1999/30/EC, 2000/69/EC and 2002/3/EC. 'Quality objective' is defined as the maximum deviation of measured and calculated concentration values within the period given by the quality indicator (to which the limit values refer, too).

Applying an obstacle resolving chemical transport model with high spatial resolution (up to $1\,m$) to a heavily polluted street canyon, *Schlünzen et al.* [2003] show that point measurements at ground level barely fulfil these requirements even under stationary meteorological conditions. Effective experimental techniques to obtain concentration maps rather than point samples would be invaluable for air pollution monitoring. Legislation acknowledges the current state of measurement technique and the costs of monitoring stations by permitting modelled concentration maps for monitoring purposes (see also table 2.1) which brings us back to the evaluation of these models by measurements. Even if chemical dispersion models eventually reach a level allowing them to be used for operational air quality predictions, experimental concentration distributions become by no means obsolete. One reason is that models (at least in the short run) depend on the quality of the input, i.e. initial values and boundary conditions. Another reason of universal significance lies in the fact that models usually allow a large choice of parameters and different settings. For transport models this could be the grid size, grid spacing, turbulence closure scheme etc. and

> "*this gives the user additional degrees of freedom and makes it unlikely that two users applying the same complex numerical model to the same problem will produce the same results. This was demonstrated in the European project, Evaluation of Modelling Uncertainty (EMU) (Hall, 1997) in which four experienced user groups predicted the dispersion of dense gas releases around simply shaped buildings by using the same commercially available CFD [computational fluid dynamics] code. The variability between different modeller's results was shown to be substantial. Depending on the quantity under consideration, differences up to an order of magnitude were reported.*" [*Schatzmann and Leitl*, 2002]

2.2. Selected trace gases

Trace gases, occurring with number densities way below the main constituents of unpolluted air nevertheless play an important role because of their chemical reactivity, for the radiation budget of the earth (greenhouse effect) and for the human health. Following up the preceding discussion, here we are interested only in species that emerge in near ground air pollution and which can be detected with DOAS measurements relevant for this work. Some of these trace gases are listed in table 2.2

trace gas	major sources	major sinks	mixing ratios [*ppb*]		
			remote	rural-urban	polluted-urban
O_3	photochem. production, free radicals	deposition, hydrolysis, photolysis,	$30-40$	$80-150$	up to 300
NO_2	fossil fuel, soil emission, biomass burning, lightning	oxidation by OH or O_3 to HNO_3	$0.03-5$	$1-60$	$50-200$
SO_2	fossil fuel, volcanoes, sulphide oxidation	dry deposition, oxidation to H_2SO_4, wet deposition	$0.01-0.05$	$1-200$	up to a few 100
HONO	het. formation on surfaces	photolysis	$0.01-1$	$0.01-10$	$0.03-30$
CH_2O,	traffic, incomplete	photolysis (CH_2O),	$0.3-2$	$0.1-10$	$1-60$
Benzene,	combustion, photochem.	reaction with OH,	$0.008-0.2$	$0.1-0.6$	$0.9-26$
& *Toluene*	production (CH_2O)	dry/wet depos.	$0.01-0.25$	$0.05-0.8$	$2-39$

Table 2.2: Sources, sinks and approximate mixing ratios for selected trace gases occurring in near ground air pollution and that are detectable with long-path DOAS. Mixing ratios are taken from [Finlayson-Pitts and Pitts, 2000, sec. 11.A.4].

with typical concentrations[1] in polluted and unpolluted air.[2]

Starting to investigate the tomographic reconstruction of trace gas distributions from DOAS measurements, we concentrate only on the most important atmospheric processes responsible for the spatial patterns of the concentrations. Transport processes are addressed in the next section, very basic textbook facts on the chemical behaviour of important species are given now. Further details can be found, for example, in [*Seinfeld and Pandis*, 1998; *Finlayson-Pitts and Pitts*, 2000].

Nitrogen oxides and ozone

The largest source of NO and NO_2 is fossil fuel combustion in industrial processes, power generation and traffic. Relative amounts of NO and NO_2 in the exhaust depend on the details of the combustion process. For power plants the ratio of NO_2 to NO varies between less than 10% and 40%, with typical values between $5-10\%$ [*Bland et al.*, 2000]. For traffic emissions, the ratio takes values around $5-10\%$ [e.g. *Kohler et al.*, 2005; *Bäumer et al.*, 2005; *Stern and Yamartino*, 2001; *Berkowicz et al.*, 1997].

NO quickly reacts with ozone to NO_2, so that they are usually combined to NO_x by $[NO_x] = [NO] + [NO_2]$. NO_2 photodissociates for wavelengths $\leq 420\,nm$, leading to the formation of ozone. In total

$$NO + O_3 \longrightarrow NO_2 + O_2 \tag{R 2.1}$$

$$NO_2 \xrightarrow[\lambda \leq 420\,nm]{h\nu} NO + O(^3P) \tag{R 2.2}$$

$$O(^3P) + O_2 \xrightarrow{M} O_3 \;, \tag{R 2.3}$$

where the collision partner M enables the recombination. In the absence of any other chemical transformations, reactions (R 2.1)–(R 2.3) lead to an approximate steady state described by the so-

---

[1] To be correct, concentrations (amounts of trace gas per volume) should be distinguished from so-called *mixing ratios* which are amounts of trace gas per amount of air, both referring to volume, mole or mass and the corresponding units being *parts per million* (*ppm*) or *billion* (*ppb*) by volume, mass or mole. But we will make a distinction only if crucial and use the same symbol c for both quantities.

[2] Following the definition in [*Seinfeld and Pandis*, 1998, p. 49], a substance – benign or harmful – is regarded as pollutant, if normal ambient concentrations are exceeded due to anthropogenic activities with measurable effect on the environment or humans.

called *photostationary state relation*

$$\frac{[\mathrm{NO}]}{[\mathrm{NO_2}]} = \frac{j_{(\mathrm{R\,2.2})}}{k_{(\mathrm{R\,2.3})}\,[\mathrm{O_3}]}.$$

Here $k_{(\mathrm{R\,2.3})}$ is the rate constant of reaction (R 2.3) and $j_{(\mathrm{R\,2.2})}$ the photolysis frequency of NO_2. For typical O_3 mixing ratios of a few 10 *ppb*, at daytimes the ratio $[NO]/[NO_2]$ takes values of the order one.

The above system of reactions offers only an insufficient description if stronger oxidants play a role. In fact, these are usually present in the form of radicals, most importantly the hydroxyl (OH) and the hydroperoxyl (HO_2) radicals. They are photochemically produced, predominantly from ozone, HONO and formaldehyde (CH_2O).

On the one hand, the oxidation of NO by HO_2

$$\mathrm{NO} + \mathrm{HO_2} \longrightarrow \mathrm{NO_2} + \mathrm{OH}$$

instead of reaction (R 2.1) leads together with reactions (R 2.2) and (R 2.3) to a net production of O_3. On the other hand, there is a cycle between OH and HO_2 in which the radical HO_2 is formed by the reaction of OH with oxidisable gases like CO or CH_2O. The reaction of HO_2 back to OH via

$$\mathrm{O_3} + \mathrm{HO_2} \longrightarrow \mathrm{OH} + 2\,\mathrm{O_2}$$

plays a role for low NO_x concentrations and results in a net destruction of ozone. Whether the coupled NO_x and HO_x ($= OH + HO_2$) cycles eventually produce or destroy O_3 depends on the NO_x concentrations. For higher NO_x, levels O_3 production exceeds the destruction by radicals. Lower levels result in net destruction of O_3.

Apart from its role in the HO_x cycle, the highly reactive OH radical constitutes also a day time sink of NO_2 by the reaction

$$\mathrm{NO_2} + \mathrm{OH} \xrightarrow{M} \mathrm{HNO_3}\ ,$$

which is important for high NO_x levels, while at night the oxidation by O_3 becomes a NO_2 sink. In both cases the product is HNO_3, which is finally removed by precipitation and dry deposition.

Volatile organic compounds

The large class of *volatile organic compounds* (VOCs) comprises hydrocarbons such as alkanes, alkenes and alkines, aromatic hydrocarbons and oxidised compounds (aldehydes, ketones, alcohols, organic acids). Some occur naturally, like methane (CH_4) or terpenes which are produced by certain trees. The dominant anthropogenic source for VOCs like benzene (C_6H_6), toluene (C_7H_8) or formaldehyde (CH_2O) is (incomplete) combustion in traffic, domestic fuel or industry.

Some VOCs, like benzene, play prominant roles in air pollution because of their severe toxic character. But in any case, VOCs are important for the production of HO_2 and thus for the ozone production, as illustrated here for the group of alkanes (written as RH)

$$\begin{aligned} \mathrm{OH} + \mathrm{RH} &\longrightarrow \mathrm{R} + \mathrm{H_2O} \\ \mathrm{R} + \mathrm{O_2} &\xrightarrow{M} \mathrm{RO_2}\ . \end{aligned}$$

In polluted areas, the reaction of the peroxy radical RO_2 with NO is most likely so that

$$\begin{aligned} \mathrm{RO_2} + \mathrm{NO} &\longrightarrow \mathrm{RO} + \mathrm{NO_2} \\ \mathrm{RO} + \mathrm{O_2} &\longrightarrow \mathrm{R^{'}OHO} + \mathrm{HO_2}\ . \end{aligned}$$

Similarly, other VOCs drive the HO_x–NO_x system to produce ozone, while the VOCs themselves are reacted to aldehydes and finally to CO_2. The interplay between the HO_x and NO_x cycles results in an intricate dependency of the O_3 production on inital VOCs and NO_x concentrations (for details and the so-called VOC- and NO_x-limited ozone production see the references mentioned in the introduction).

Sulphur dioxide

Sulphur compounds enter the atmosphere both naturally (volcanoes, biological sources) and anthropogenically. In the latter case they are emitted predominantly as SO_2 when burning sulphur containing coal or fuel. In the context of air pollution, the reaction of SO_2 to sulphuric acid (H_2SO_4) and subsequent emergence of acid rain has been a major issue.
Concentrating on emissions in the form of SO_2, the most important gas phase reaction is oxidation by OH

$$\begin{aligned} \mathrm{SO_2} + \mathrm{OH} &\xrightarrow{M} \mathrm{HOSO_2} \\ \mathrm{HOSO_2} + \mathrm{O_2} &\longrightarrow \mathrm{SO_3} + \mathrm{HO_2} \\ \mathrm{SO_3} + \mathrm{H_2O} &\xrightarrow{M} \mathrm{H_2SO_4}\ , \end{aligned}$$

but SO_2 is also removed from the atmosphere by liquid phase oxidation and dry deposition. Depletion by the above oxidation process is quite moderate. For a relatively clean urban environment and a cloudless summer day, the $24\,h$ averaged rate of SO_2 oxidation amounts to $0.7\%\,/h$ assuming $[\mathrm{OH}] \sim 1.7 \cdot 10^6\, molecules \cdot cm^{-3}$ [*Seinfeld*, 1986, p.167].

In conclusion, it has to be emphasised that the picture of the reactions of trace gases just given does not even contain all basic mechanisms – for example, the NO_3 radical and the nitrous acid HONO were omitted, the sulphur compounds were reduced to SO_2 and so forth. But nevertheless, besides conveying insight into the complexity of air pollution, it explains important characteristic features of trace gases like diurnal cycles for some of them and gives an idea about when certain simplifying assumptions may be valid and when not. For example, on time scales of a few hours the evolution of a SO_2 plume may be approximated by wind transport and dispersion, neglecting its chemical transformation. In constrast, freshly emitted NO–NO_2 in the emissions, for instance, of a power plant undergo rapid transformation during their dispersion until they level off at some ratio more or less given by the photostationary state relationship. As there are no point sources for O_3 as such, spatial variations of the ozone concentration can only result from high local production rates, as in the so-

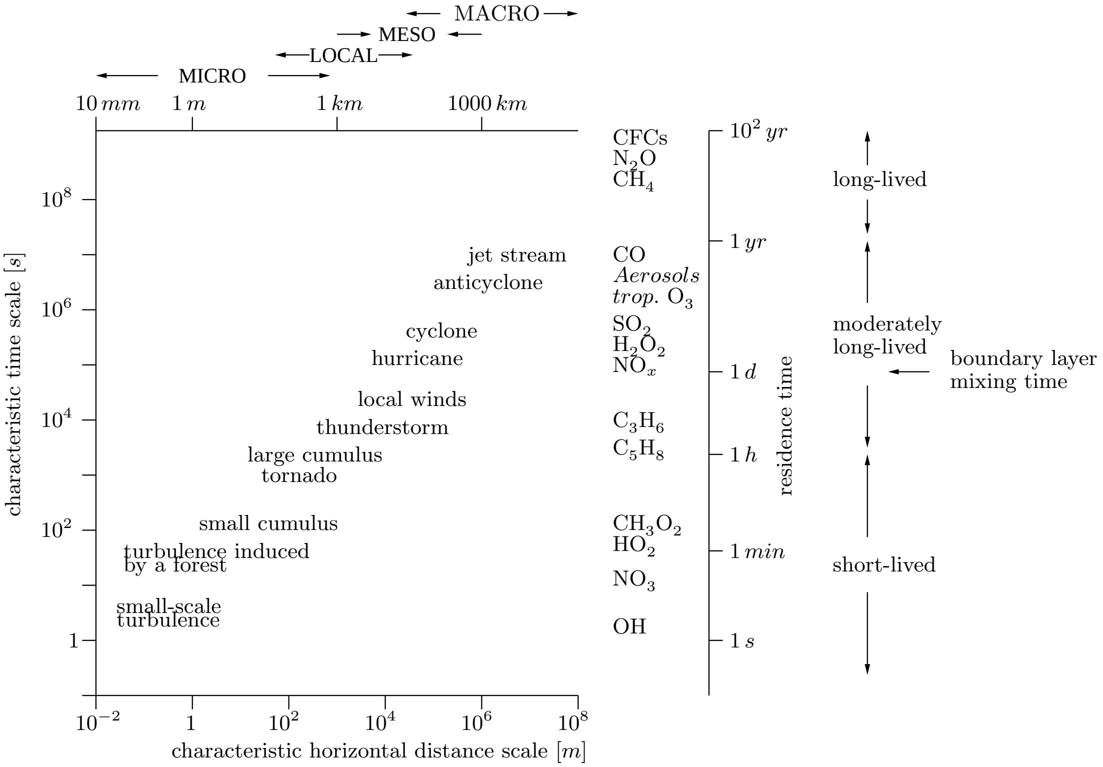


Figure 2.1: Temporal and spatial scales of turbulent atmospheric processes and lifetimes of some species in the atmosphere. Transport of the latter is characterised by spatial scales according to the turbulent scales. The precise definition of the terms micro-, meso- *and* macroscale *is not always the same in the literature but this of no concern here [combined from Stull, 1988; Wayne, 2000].*

called photo-smog, or depletion, for example by reaction with freshly emitted NO. Ozone distributions show little spatial variation at low concentration levels in unpolluted background air, but enhanced values can occur downwind of urban centres where the O_3, gradually formed in the NO_x and VOC enriched city plume, reaches its maximum.

The life time of a species in the atmosphere is of great interest for transport and dispersion processes. Figure 2.1 gives an idea about the time scales for some trace gases. But while these average values might be useful for species with only few reaction paths (as seen for SO_2), they become less meaningful for complex systems where not only emissions vary, but also factors like sunlight or cloud cover affect chemical transformations. For NO_2 *Neophytou et al.* [2005] specify diurnal variations of the transformation time scale between a few minutes and up to 10^4 minutes.

2.3. Dispersion by turbulent diffusion

The temporal evolution and spatial distribution of a scalar quantity like the concentration $c_i(\mathbf{r}, t)$ of a trace gas species i is governed by a continuity equation of the form

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (\mathbf{u}\, c_i) = D_i\, \Delta c_i + R_i(c_1, \ldots, T) + S_i(\mathbf{r}, t), \tag{2.1}$$

where $\mathbf{u}$ denotes the velocity field, D_i the molecular diffusion constant for species i, R_i its rate of generation or destruction by chemical reaction and S_i is a source term. Here and in the following it is assumed that this equation for c_i is independent of the equations for $\mathbf{u}$, the temperature T and all other meteorological quantities. But even in the most simple case of a chemically inert gas without source and sink (the molecular diffusion term is of the order 10^{-5}–10^{-6} smaller than the advection term $\nabla \cdot (\cdot)$), eq. (2.1) cannot be solved because $\mathbf{u}$ contains an unknown, stochastic component related to turbulences in the atmosphere. Turbulent eddies can be thermally generated by the heating of the atmosphere or mechanically caused when air moves past an obstruction. Turbulence occurs under all circumstances and on the whole time and length scale of atmospheric phenomena (fig. 2.1). The velocity field $\mathbf{u}$ is of course not totally erratic so that it makes sense to write it as a sum of the deterministic mean $\bar{\mathbf{u}}$ and a random part $\mathbf{u}'$

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'.$$

As a consequence the concentration field, too, consists of a deterministic and a random contribution in a sense illustrated in fig. 2.2a. The (hypothetical) average $\langle c_i(\mathbf{r}, t)\rangle$ for species i can be regarded as the ensemble mean of all measurable, real distributions $c_i(\mathbf{r}, t)$. It can be obtained by time averaging over measurements if the meteorological conditions for $\bar{\mathbf{u}}$ are stable and the chemical generation and destruction processes are more or less stationary compared to the turbulent time scales for the velocity field. Put in other words, these time scales give a clue to what degree the distribution $c_i(\mathbf{r}, t)$ is representative for the mean meteorological conditions. For this reason and without going into the details of the statistical representation of turbulence, I would like to point out the following with regards to the turbulent variations of the velocity and concentration fields.

- The time scale for the *velocity turbulence* can be defined by the time integral of the Lagrangian autocorrelation function of the velocity field, that is the autocorrelation in a system moving along with the velocity field. This function is hard to measure and an exact relation with the corresponding Eulerian function in usual fixed-point coordinates is not known. Often they are assumed to be of the form $\sim e^{-t/T}$ with different times T_L, T_E for the Lagrangian and Eulerian function. Based on the Taylor (frozen eddy) assumption one finds the estimates

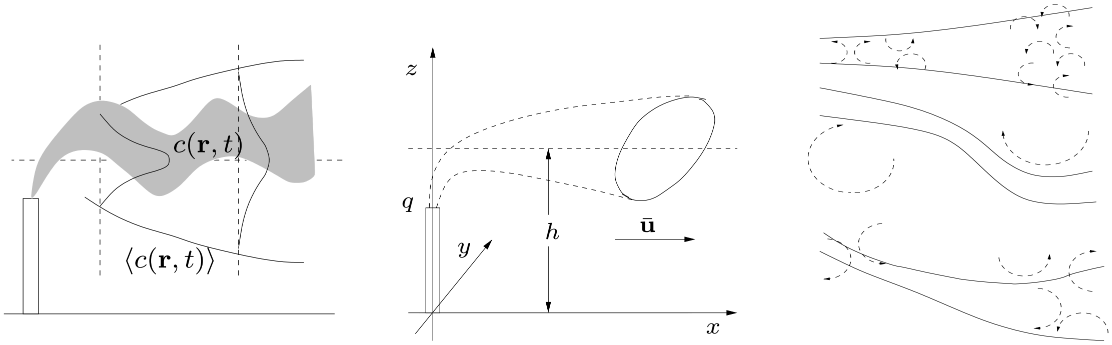
$$T_L \sim (\text{repetition rate around dominant eddies of size } l)^{-1} \sim \frac{l}{\sigma_v} \tag{2.2a}$$

$$T_E \sim (\text{repetition rate of dominant eddies})^{-1} \sim \frac{l}{\bar{u}} \tag{2.2b}$$

$$\frac{T_L}{T_E} = C\frac{\bar{u}}{\sigma_v} \tag{2.2c}$$

  with mean velocity $\bar{u}$, standard deviation σ_v of the vertical velocity component and $C \sim 0.7$ inferred from observations [e.g., *Blackadar*, 1997]. Empirical values for T_L and $\sigma_\alpha = \sigma_v/\bar{u}$ can be found in appendix A.

- Time scales for the *concentration turbulence* can be defined in a similar manner. But there is no reason why they should match the velocity time scales. On the contrary, for the concentration fluctuations in time series obtained from a number of smoke plume experiments, *Hanna and Insley* [1989] find that the time scale of the concentration spectra is $\sim 2 - 5$ times less than for the wind speed and $\sim 10 - 20$ times more than that of the vertical velocity. Turbulent patterns



(a) (b) (c)

Figure 2.2: (a) A concentration field $c(\mathbf{r},t)$ *under turbulent conditions as a member of a stochastic ensemble with mean* $\langle c(\mathbf{r},t)\rangle$ *which is assumed to be a Gaussian distribution here. (b) Emission from a point source at an effective height* h *with source strength* q $[kg\, s^{-1}]$*. (c) Evolution of a plume depending on the eddy size [adapted from Seinfeld and Pandis, 1998].*

in the time series translate into variances σ_c of the concentration around the ensemble mean (see next section).

As a deterministic approach to turbulence is not possible, the aim must be to describe the mean fields for the concentration $\langle c_i(\mathbf{r},t)\rangle$ as accurately as possible. Inserting the decompositions for $\mathbf{u}$ and the corresponding one

$$c_i(\mathbf{r},t) = \langle c_i(\mathbf{r},t)\rangle + c_i'(\mathbf{r},t) \tag{2.3}$$

into eq. (2.1) and averaging over the turbulent realisations leads to

$$\frac{\partial \langle c_i\rangle}{\partial t} + \nabla\cdot(\bar{\mathbf{u}}\,\langle c_i\rangle) + \nabla\cdot\langle \mathbf{u}'\, c_i'\rangle = D_i\,\Delta\langle c_i\rangle + R_i(\langle c_1\rangle + c_1', \ldots, T) + S_i(\mathbf{r},t) \tag{2.4}$$

with two unknown quantities. The fact that the unwanted flux $\langle \mathbf{u}' c_i'\rangle$ cannot be substituted without introducing combinations of higher order in u_j' and c_i' is known as the closure problem of turbulence.[3] There are ways of varying complexity to solve the closure problem leading to more or less involved numerical models. The most simple of them is to assume a diffusion like relationship for the turbulent flux and the gradient of the mean field, i.e.

$$\langle \mathbf{u}' c_i'\rangle = -K\,\nabla\langle c_i\rangle \quad \text{with eddy diffusity tensor } K\text{, here} \quad K_{jk} = K_{jj}\,\delta_{jk}.$$

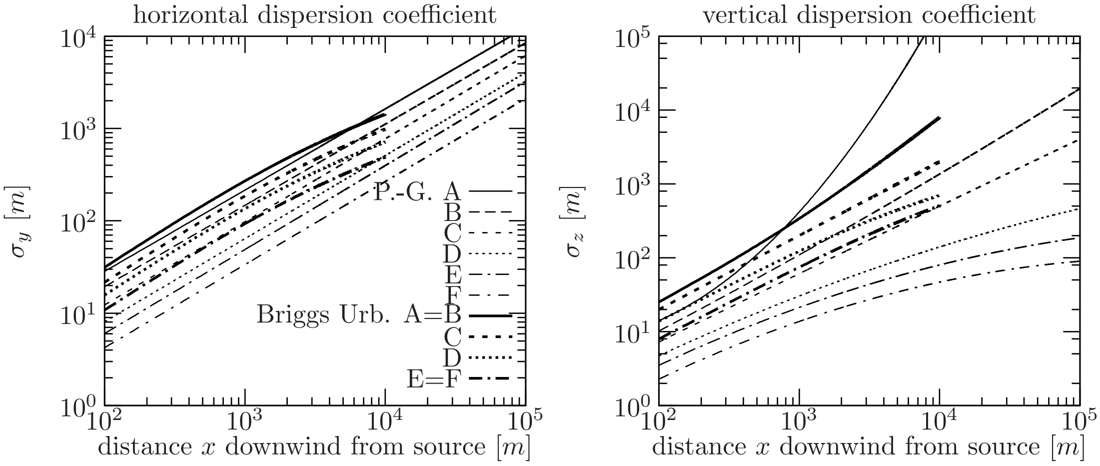
This approach, known as K-*theory*, *mixing-length theory* or *gradient transport theory*, results in equations that describe the turbulent dispersion of a pollutant as a diffusive process parametrised by the *eddy diffusities* K_{jj} (if molecular diffusion is again neglected and the mean velocity field is taken to be incompressible). They are referred to as atmospheric diffusion equations.

2.4. The Gaussian plume model

Assuming a constant mean velocity field $\bar{\mathbf{u}}$ and negligible chemical reaction, it is straightforward to show that the atmospheric diffusion equation has solutions in the form of Gaussian distributions. To

---

[3] More precisely the closure problem of the *Eulerian* approach as in eq. (2.1).



(a) *(b)*

Figure 2.3: (a) Horizontal dispersion coefficients ($\sigma_x = \sigma_y$) for the Pasquill-Gifford (thin lines) and the urban Briggs (thick lines) parametrisation. For the latter the dispersion coefficients for stability classes A and B and the ones for E and F are identical.
(b) The same as (a) for the vertical dispersion. The functional forms and further details can be found in appendix A.

get more specific, consider the situation depicted in fig. 2.2b, i.e. a point source at the origin, emitting at a rate q (in units $kg\ s^{-1}$) at an effective height h and a wind field $\mathbf{\bar{u}} = (\bar{u}, 0, 0)$. For instantaneous release of the total amount $Q = \int q\, dt$ one gets the *Gaussian puff* formula

$$\langle c(x,y,z,t)\rangle = \frac{Q}{8(\pi\, t)^{3/2}(K_{11}K_{22}K_{33})^{1/2}}\ \exp\Big(-\frac{(x-\bar{u}t)^2}{4K_{11}t} - \frac{y^2}{4K_{22}t}\Big) \\ \times\left[\exp\Big(-\frac{(z-h)^2}{4K_{33}t}\Big) + \exp\Big(-\frac{(z+h)^2}{4K_{33}t}\Big)\right], \quad 0 \le z \le \infty,$$

while the stationary concentration field resulting from continuous release is described by the *Gaussian plume* equation

$$\langle c(x,y,z)\rangle = \frac{q}{8\,\pi\,(K_{22}K_{33})^{1/2}x}\ \exp\Big(-\frac{\bar{u}}{4x}\frac{y^2}{K_{22}}\Big) \\ \times\left[\exp\Big(-\frac{\bar{u}}{4x}\frac{(z-h)^2}{K_{33}}\Big) + \exp\Big(-\frac{\bar{u}}{4x}\frac{(z+h)^2}{K_{33}}\Big)\right], \quad 0 \le z \le \infty, \quad (2.5a)$$

provided that the dispersion in the direction of the flow can be neglected, i.e. that the distance x a particle has travelled since its release is large compared to the spread at that time $\sqrt{K_{11}t}$ or $K_{11} \ll \bar{u}\,x$ (so called *slender plume approximation*). Both formulas are for boundary conditions corresponding to total reflection of the plume at ground ($z = 0$). Similar expressions can be derived for total or partial absorption at ground and additional reflection at an inversion layer above the plume [e.g., *Seinfeld and Pandis*, 1998].

Based on large numbers of observations (like the legendary prairie grass experiment [*Barad*, 1958] and the urban dispersion study in St. Louis [*McElroy and Pooler*, 1968]) several parametrisations of the plume dispersion coefficients $\sigma_x = \sigma_y = \sqrt{2K_{22}x/\bar{u}}$ and $\sigma_z = \sqrt{2K_{33}x/\bar{u}}$ have been proposed. They usually refer to a classification of atmospheric stability proposed by *Pasquill* [1961] and distinguish six (sometimes seven) classes ranging from *extremely unstable* (A) to *moderately stable* (F) (or *extremely stable* (G)). Appendix A contains details about this classification in terms of meteorological observables. For the widely used Pasquill-Gifford scheme [*Gifford*, 1961] fig. 2.3 shows σ_y, σ_z for the classes A – F as a function of the downwind distance x. *Briggs* [1973] gave different parametrisations for open country and urban conditions (fig. 2.3). Explicit formulas can be found in appendix A. Although these schemes have been introduced more than 40 years ago and new ones are being proposed [e.g., *Hanna et al.*, 2003] the original curves are still in use [*Barratt*, 2002]. When applying the Gaussian plume or puff model, apart from the steady state conditions assumed above, one should notice:

- The models are not fit for complex terrain.
- It is the effective emission height h that enters the formula (important if the gas is emitted with initial momentum, see [*Barratt*, 2002; *Seinfeld and Pandis*, 1998] for correction terms).
- The parametrisations are not reliable

  near the source [see also *Venkatram et al.*, 2004],
  too far away from the source ($\gtrsim 10\,km$), i.e. the uncertainty of σ_y, σ_z increases with x,
  for elevated sources higher $\sim 100\,m$ [*Barratt*, 2002].
- Strictly speaking, the σ_y, σ_z are valid only for the averaging times for which they have been derived ($3 - 10\,min$ and $1\,h$ for the parametrisations mentioned above, see appendix A for details).
- A single measurement over a limited period will not see a Gaussian concentration distribution, c.f. fig. 2.2a.

A few remarks on the last point which refers again to the context of fig. 2.2a and is especially critical for the meandering plume in fig. 2.2c. Loosely speaking this is the case if the effective eddy size is much larger than the size of the plume. Turbulent eddies of a size comparable to the plume change its size and disperse it diffusion-like while very small eddies mix the plume within. The size of the dominant eddies can be estimated with eqs. (2.2) by consulting empirical values for T_L and σ_α.

> For example, using Draxler's specification for T_L in appendix A one gets:
> For a surface source and stability class C (slightly unstable)
>
> $$T_L = \begin{cases} 60\ s & \text{lateral} \\ 10\ s & \text{vertical} \end{cases}, \quad \sigma_\alpha = 15^\circ \quad \overset{\text{eq. (2.2c)}}{\Rightarrow} \quad T_E = \begin{cases} 22\ s & \text{lateral} \\ 7.5\ s & \text{vertical} \end{cases}$$
>
> and for a wind speed of $\bar{u} = 5\ \frac{m}{s}$ thus roughly $l \sim 100\ m$.

If the measurement time T is shorter than necessary to reproduce the ensemble mean, the expected deviation of the mean concentration $\bar{c}_T$ during T at a specific location from the ensemble mean $\langle c \rangle$ at the same place is according to *Venkatram* [1979] given by

$$\frac{\mathrm{E}\big[(\bar{c}_T - \langle c \rangle)^2\big]}{\langle c \rangle^2} \sim \begin{cases} \frac{\mathrm{var}[\langle c \rangle^2]}{\langle c \rangle^2} & \frac{T}{T_E} \to 0 \\ 2(\Gamma - 1)\frac{T_E}{T} & \text{else} \end{cases}$$

where $\Gamma = \frac{c_{max}}{\langle c \rangle}$ with c_{max} being the maximum concentration detected at the location and T_E the Eulerian time scale for the concentration turbulence. For the moderate example $\Gamma = 10$, $T/T_E = 10$ one gets $\mathrm{E}\big[(\bar{c}_T - \langle c \rangle)^2\big] \sim 1.4 \langle c \rangle^2$, i.e. the deviation is of the same size as the ensemble mean value. For the special case of a Gaussian plume the author gives a further theoretical expression for the expected deviation at the centreline of the plume in [*Venkatram*, 1984]. *Ma et al.* [2005] give an estimate for the same quantity by analysing experimentally obtained crosswind sections through plumes. They find that the crosswind fluctuation $\sigma_c(y)/c(y)$ along y increases at the edges of the plume and gets smaller towards the centreline. For distances further away from the source and averaging times around $1\,h$, values at the edges of the plume amount to $\sigma_c(y)/c(y) \sim 10$ and at the centreline to $\sigma_c(0)/c(0) \sim 0.1$. Near the source the fluctuations increase dramatically. In fact, the authors propose a fit of the form $\sigma_c(0)/c(0) \sim a + \frac{b}{x}$ for the fluctuation at the centreline depending on the downwind distance to the source x.

In principle, the Gaussian plume model can be applied to line and area sources, too [e.g., *Barratt*, 2002; *Seinfeld and Pandis*, 1998]. Simple analytic corrections of the Gaussian solution concern the initial plume size, plume rise and exponential chemical decay of the emitted substance. More involved chemical transformations have to be implemented numerically [e.g. *Olcese and Toselli*, 2005; *Song et al.*, 2003; *Von Glasow et al.*, 2003]. The limitation to stable atmospheric conditions can be lifted by using a series of puffs instead of a continuous plume. The frequently employed numerical model SCIPUFF [*Sykes and Gabruk*, 1997] is based on this approach. Other numerical multi-plume dispersion models incorporating more or less directly the Gaussian plume model are mainly applied to impact assessment of regular or accidental emissions from industrial complexes, power plants etc. There is a large variety of these dispersion models employed by environmental institutions, national agencies and companies. The classic industrial source complex (ISC) model[4] was developed by the US Environmental Protection Agency (EPA) for source types from point to volume sources. It describes the dispersion by the Pasquill-Gifford parametrisation and accounts for deposition when concentration hits the ground. The further developed model AERMOD includes calculation of vertical profiles for meteorological parameters and treats dispersion under convective conditions more accurately. Apart from the input of emission and meteorological data it requires specification of the topography to correct flow patterns. Likewise other state of the art models like the often mentioned Atmospheric Dispersion Model System (ADMS) have modules for processing terrain data, meteorological parameters (e.g. the mesoscale meteorological model MM5), plume rise (e.g. PRIME) and so forth. The Lagrangian Atmospheric Dispersion Model (LADM) from the CSIRO Division of Atmospheric Research (Australia) is capable of modelling photochemical formation and dispersion of gases such as O_3 and NO_2, but in general model systems designed to predict atmospheric dispersion do not include reaction mechanisms. Short descriptions for some of the latest dispersion model systems can be found in [*Barratt*, 2002] which also contains a list of related internet sites.

2.5. Urban trace gas distributions

Here, we are interested in concentration distributions from a tomographic point of view, i.e. mainly in their spatial patterns, not so much in their chemistry. The latter aspect has been very briefly

---

[4] The model is available in the long term version ISCLT and the sort term version ISCST, the current version being ISCST3.

addressed above in sec. 2.2 and we restrict ourselves here to the following observations concerning emission source types.

- Traffic is an important primary source for NO_x, CO and VOCs (possibly SO_2). It can show strong temporal variability in terms of daytime or rush hours. The simplest way is to treat it as a line source, a more involved approach includes traffic induced turbulences (which according to *Tsai and Chen* [2004] can in street canyons be as important as wind induced turbulences for the mixing of the pollutants). The source strength does not only depend on the traffic frequency but also on the emission factors. These are far from well-known as pointed out by *Corsmeier et al.* [2005*b*] who report underestimations for the CO and NO_x emission factors used by models of around 20%.

- Domestic heating constitutes a seasonal source of SO_2, CO and CO_2 (with diurnal variations, see [*Rippel*, 2005] for measurements in Heidelberg) and can be regarded as an area source, at least in densely populated downtown and residential areas.

- Factories, industrial complexes as well as power plants or block power stations are sources for a variety of species like NO_x, SO_2, CO, VOCs and CH_2O. As gases are discharged via more or less high chimneys they represent the most typical case for the Gaussian plume model. Impact of the emission can be estimated once the effective stack height and the stability class have been specified. According to *Barratt* [2002] the stack height is usually chosen at least 2.5 times the height of any building within a radius of twice the stack height to make sure that downdraught of the emissions does not occur. The rule of thumb quoted in this context that the airflow is disturbed up to a height of $\sim$ 2.5 times the building height agrees with the observation that mechanically induced turbulences usually cause nearly neutral stabilities over urban areas up to heights around two times the average building height [*Hanna et al.*, 2003].

- Not only emissions from within the town and its suburbs make up the urban trace gas distributions. Polluted air or air with a special composition like ozone rich air imported from outside add to the concentrations and react in a way they might not do in unpolluted air.

Turning from the emissions to the immissions: What kind of measurements are there to give insight into the spatial distribution of urban concentration distributions? We would like to characterise these efforts in the following manner:

- Intensive measurements aiming at the concentration distributions on a mesoscale, e.g. the chemical composition for particular meteorological or topographical conditions or the chemical development of a city plume etc. This kind of campaigns typically comprises a variety of measurement techniques like point or mobile in situ sampling, remote sensing measurements and possibly airborne measurements and they usually address very specific questions. Examples are the BERLIOZ study that investigated the evolution of the Berlin plume [e.g., *Volz-Thomas et al.*, 2003], the FORMAT project focussing on formaldehyde in and around Milano in the heavily polluted Po valley [*Hak et al.*, 2005] or a series of measurements in the extremely polluted megacity of Mexico City to learn more about oxidation processes in the atmosphere [*Shirley et al.*, 2006].

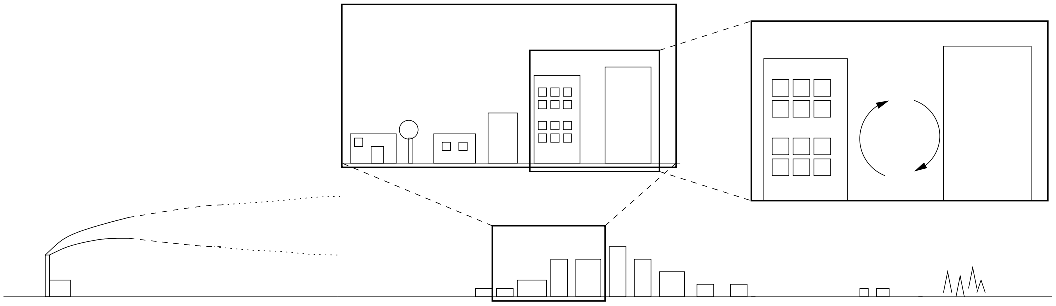


Figure 2.4: Different scales involved in urban modelling: from mesoscale models for background concentrations and meteorological boundary conditions to models on the scale of a street canyon (shown here with a typical rotor like flow pattern) [simplified from Fisher et al., 2006].

- Long-term in-situ measurements, especially operational air quality monitoring. They provide important statistical data for individual sites but detectors are usually not installed in a way that admits conclusions on the spatial distributions. The representativeness of these measurements, even for a local neighbourhood, are questioned by [*Schlünzen et al.*, 2003] on the grounds of model calculations with very high spatial resolution.

- Hot spot measurements on a microscale like the measurement of concentrations and airflow within a street canyon [e.g. *Schatzmann et al.*, 2006; *Vardoulakis et al.*, 2005] or around an industrial complex.

On its proper scale, each of these setups might give a comprehensive picture of the concentration fields, but very often they do not: Too few point measurements, too short time series of data, not all relevant species measured and so on. Not least because data interpretation might be impossible without model assumptions, some aspects of modelling trace gas distributions in an urban environment are now introduced.

The problem with urban distributions is that any model describing them satisfactorily has to be multiscale, see fig. 2.4. On the spatial microscale building effects become important (and, as already said, according to *Schatzmann and Leitl* [2002] can easily change concentrations by more than an order of magnitude). Mesoscale conditions drive transport of airmasses to the town and away from it. A result from long-term measurements at a heavily polluted street canyon in Hannover illustrates this: For a period of two years it was found that NO_x concentrations at roof level in the street canyon are to 30% due to regional transport [*Schäfer et al.*, 2005]. But mesoscale conditions can also be responsible for more complex patterns as shown in fig. 2.5. According to *Fisher et al.* [2006] the complexity of urban meteorology is not fully acknowledged by modellers and "[...] *there seems to be a reluctance from model developers to move away from familiar concepts of the boundary layer even if they are not appropriate to urban areas.*" Time scales relevant for the model are not just given by transport times but rather by chemical transformation times. These may range from $10^{-4}\,s$ for the OH radical to $10^{8}\,s$ for CH_4 (see fig. 2.1) and can vary for one and the same species with mixing ratios of other trace gases or physical parameters like radiation and temperature.

Among the different concepts for describing urban trace gas concentrations we are here interested only in methods that can – at least in principle – reproduce spatial (and temporal) variability on scales

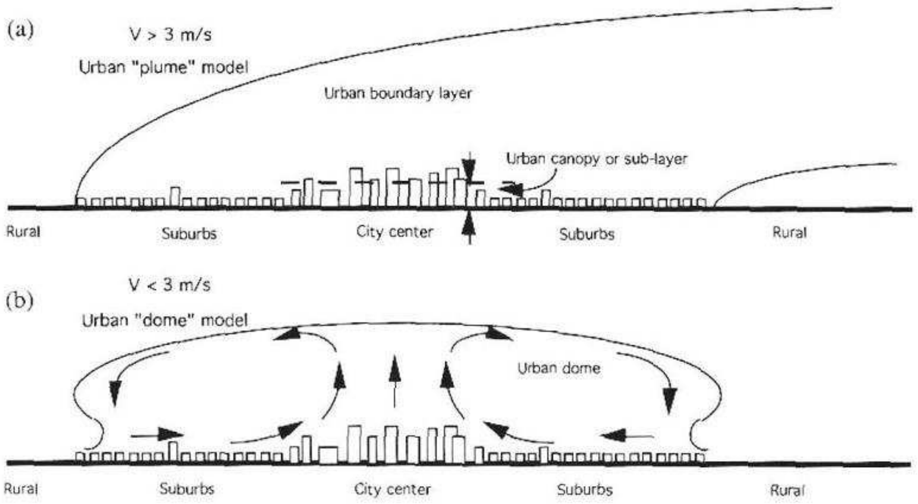


Figure 2.5: Example of how mesoscale meteorological conditions can show up in the urban airflow patterns. This illustrates both the complexity of urban meteorology and the necessity of applying models on different scales [adopted from Fisher et al., 2006].

relevant for long-path DOAS measurements. This excludes all kinds of statistical models, for example receptor models [e.g., *Seinfeld and Pandis*, 1998], or diagnostic models which use empirical functions to simulate flow around buildings and leaves us mainly with chemical transport models (CTM) that have a high enough resolution within the urban area. The problem of multiple scales can be tackled by either of the following two ways:

- A single (Eulerian) model nested at different scales [e.g., *Pielke et al.*, 1992].
- A system of models, each for a different scale [e.g., *Trukenmüller et al.*, 2004; *Soulhac et al.*, 2003; *Kessler et al.*, 2001].

Both approaches rely on further components that provide emissions, meteorological data and the topography within each model/nesting domain as input, as sketched in the cartoon fig. 2.6. Here the meteorological models provide flow fields (where the mesoscale model processes meteorological observations) that enter the (on- or offline) calculation of concentration fields by the chemical transport models. Any details depend on the specific model system. For example, the emission component can depend on the meteorological conditions or contain a more or less complex model to simulate traffic emissions. Measured background concentrations or values interpolated from monitoring stations could be used as immission constraints for the chemical models and so on. The scales *regional*, *urban* and *street* in fig. 2.6 are not universally defined.[5] Domain size and grid resolution at each scale (nesting) depend on the model and, of course, on the problem. A typical horizontal resolution for an area $\sim 200 \times 200\,km^2$ (regional scale) is around $2-5\,km$. To be more specific, model domains and and grid resolutions for the first two model systems mentioned above are as follows (the system in [*Kessler et al.*, 2001] lies somewhere between the former two). *Soulhac et al.* [2003] use a square area of $208 \times 208\,km^2$ with a grid spacing of $4\,km$ for the mesoscale model around Lyon and a resolution of $500\,m$ within an

---

[5] *Britter and Hanna* [2003] suggest the definition of *street scale* for less than $100-200\,m$, *neighbourhood* for scales up to $1-2\,km$, *city* up to $10-20\,km$ and *regional* up to $100-200\,km$.

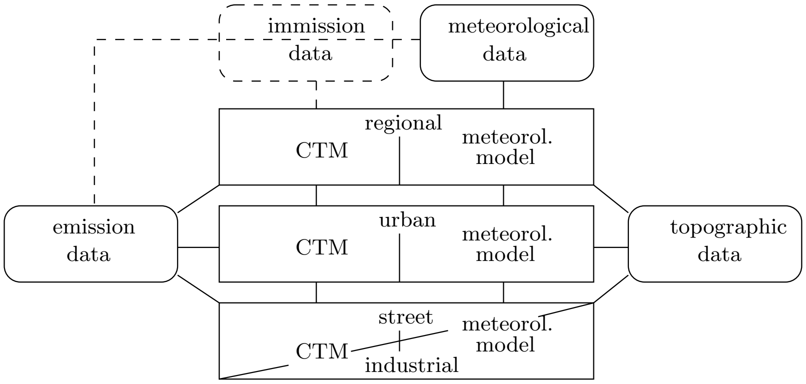


Figure 2.6: Simplified sketch of an urban model system. The scales street, urban *and* regional *do not have a fixed meaning. The* industrial *scale is given by the range of typical industrial emissions, e.g. up to* 10 *km. The corresponding transport model could be a Gaussian dispersion model. Model systems can have more or less scales than shown. Meteorological and chemical calculations can be done by the same models on different scales or by different models for different scales. Immission data could enter at any level of the system to nudge the CTM components towards observations. Emission data are supplied by inventories, models for point and line sources and possibly emission models for vegetation, traffic etc. Topographic data include surface elevation, buildings and any other obstacles.*

area of $32 \times 32\,km^2$ centred on the city. The vertical extension of the domains is $10\,km$ and $2.5\,km$, respectively, with 22 grid points in the first case and spacing between $20\,m$ (bottom) and $400\,m$ (top) in the second. The model system M-SYS presented in [*Trukenmüller et al.*, 2004] has been developed to provide concentration distributions on all scales relevant for EC air pollution regulation. For the model setup described in this publication all model domains are centred on a street canyon in the city of Hannover with the largest domain being $2000 \times 2000\,km^2$ with grid spacing $16\,km$ and a domain size of the innermost model of $1 \times 1\,km^2$. In this small area the grid is chosen non-uniformly such that the highest resolution ($1.5\,m$) is within the street canyon and the grid spacing at the boundary is $15\,m$. Input of meteorological and emission data as well as forcing of the lower scale models happens on an hourly basis in both model systems. The street canyon calculations in [*Trukenmüller et al.*, 2004] use a stationary flow field.

There are, of course, many other mesoscale transport models and applications on the regional scale and quite a few on the street scale, bur far less simulations on an urban scale as comprehensive and highly resolved as M-SYS. In particular, the Eulerian CTM CALGRID [*Yamartino et al.*, 1992] should be mentioned at this point for its popularity. Originally developed for regional and urban scales, it has been modified for use within the EU framework directives for applications on street scale, micro-CALGRID or MCG. For a first evaluation point measurements of O_3, CO, NO and NO_2 in a street canyon (Schildhornstrasse, Berlin) were compared with highly resolved calculations (horizontal domain of $1000 \times 300\,m^2$ with grid cells of $\sim 12 \times 3\,m^2$ and vertical spacing from $3\,m$ to $24\,m$ up to an height of $100\,m$) and showed good agreement [*Stern and Yamartino*, 2001]. It should be noted though that the model was adjusted to measured background concentration during the 5 day model run. Finally, for the sake of completeness, I would like to mention that concepts from computational fluid dynamics (CFD) have proven increasingly useful to describe flow around buildings and in street canyons [e.g., *Pullen et al.*, 2005; *Riddle et al.*, 2004; *Tsai and Chen*, 2004], in particular large eddy simulation models that are capable of describing at least larger turbulent structures. However, these models are computationally too expensive to be applied on urban scales [*Schatzmann and Leitl*, 2002].

3. Tomography and Remote Sensing of Atmospheric Trace Gases

This chapter describes the experimental foundations of this thesis. After a very basic introduction of the tomographic principle, common remote sensing methods are briefly reviewed with respect to their potential for tomographic measurement of atmospheric trace gases. Here it has to be emphasised that instrumental development is ongoing in all areas and we do not attempt an outlook. DOAS is covered in more detail – not least to make the origin of the detection limits and measurement errors clearer – although the theoretical and simulation results in the following basically apply to any remote sensing technique with well defined optical paths. A straightforward application of the tomographic principle to measurements with scattered sun light is not possible and a proper discussion unfortunately beyond the scope of this work. While the advantages of tomographic remote sensing are evident and have led to detailed proposals for atmospheric measurements (sec. 3.3), none of the existing applications can be called a well established technique. Finally, sec. 3.4 describes first tomographic DOAS experiments and sums up a comparison with other tomographic applications in table 3.2.

3.1. The principle of tomographic measurements

To start with an illustration consider the following

> *Example*
> An observer sends out two light beams A and B that travel on straight lines to two mirrors at different height above ground. From the spectra of the reflected light beams he can by some clever method deduce average concentrations of a trace gas along the light paths. What are the average concentrations in the two layers 1, 2 in fig. 3.1a?

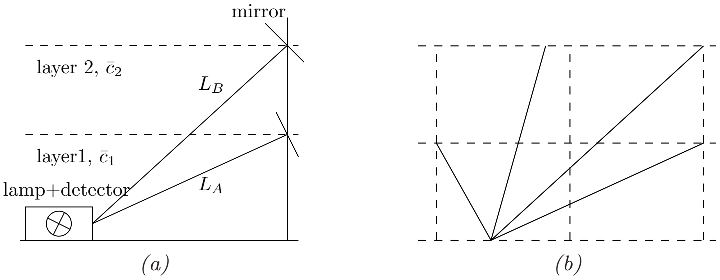
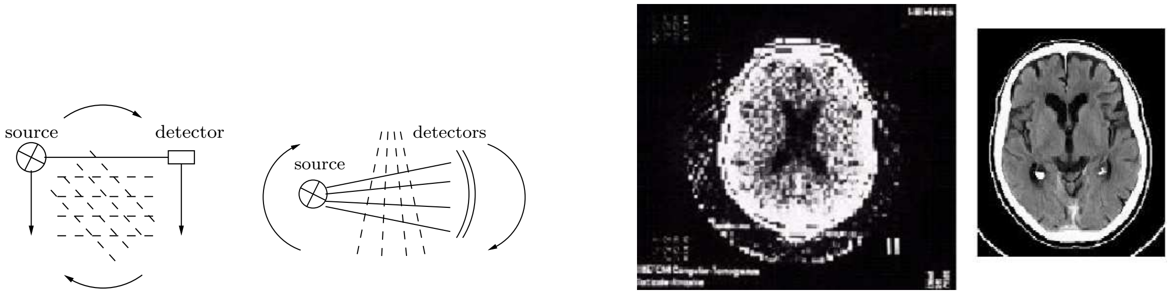


Figure 3.1: The principle of tomography. Local properties of a field are retrieved from path integrated quantities. (a) From average concentrations along the beams, the mean concentrations in two layers can be inferred. (b) In the same way, two dimensional properties can be reconstructed.



(a) Parallel scanning *(b) Fan-beam scanning* *(c) left:* 128×128 *image pixels, right:* 512×512

Figure 3.2: X-ray tomography. (a) and (b) show two kinds of scanning in transmission tomography [Adopted from Natterer, 2001]. (c) Axial CT scan of the brain from around 1975 and a state of the art scan using a scanning method as sketched in (b) [Siemens company]. Typical sizes of image pixels are around $0.1 - 1\,mm$. *Detector arrays consist of about* $500 - 1000$ *detectors of size* $\sim 1\,mm$. *A fan is made of about* 500 *rays and, depending on the application, the source-detector system is rotated for* $\mathcal{O}(100) - \mathcal{O}(1000)$ *projection angles.*

The average concentrations along the light paths are

$$\bar{c}_I = \frac{1}{l_I} \int_{L_I} ds\; c(s) \qquad I = A, B,$$

and if l_{Ai} and l_{Bi} are the lengths of the light paths in layer $i = 1, 2$, respectively, then one gets a simple system of equations for the average concentrations $\bar{c}_i$ in box i

$$\begin{aligned} \bar{c}_A &= \frac{l_{A1}}{l_A} \bar{c}_1 \\ \bar{c}_B &= \frac{l_{B1}}{l_B} \bar{c}_1 + \frac{l_{B2}}{l_B} \bar{c}_2. \end{aligned}$$

Despite its simplicity this example reveals the basic principle of a tomographic reconstruction. Extending it to two dimensions works obviously by adding integration or light paths in further directions and asking for the two dimensional dependency of the desired quantity as in fig. 3.1b. Three dimensional reconstructions can be obtained correspondingly or by adding slices of two dimensional reconstructions.[1]

THE PRINCIPLE OF TOMOGRAPHY
can be characterised by either of the two ways as:

- A non-invasive measurement technique that from the propagation of waves along a sufficient number of paths through a medium reconstructs local properties of this medium.
- The reconstruction of a function from a set of integral equations given by the line integrals of the function along a sufficient number of integration paths.

In practice, tomographic measurements are made using either acoustic waves or electromagnetic fields (mainly X-ray, microwaves, radiowaves). Depending on the phenomenon under investigation and the mechanical/electrical properties of the medium, the acoustic/electromagnetic waves can be emitted, absorbed or transmitted and possibly be reflected and diffracted. The associated tomographic techniques are called emission tomography, transmission tomography and so forth. The classic example of

---

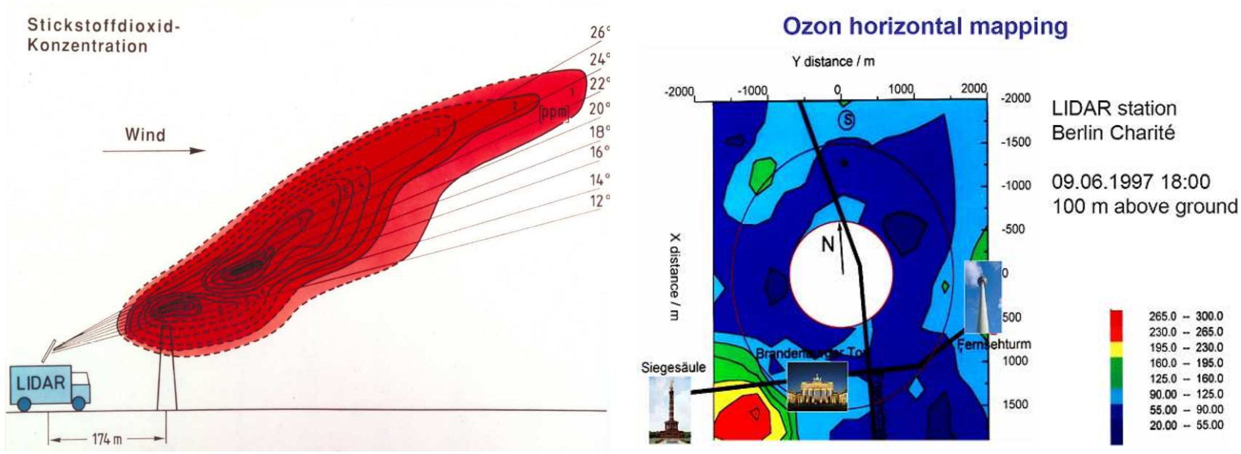
[1] This is where tomography got its name from: *tomos* meaning slice and *graphia* describing.

a tomographic measurement is the X-ray absorption tomography applied to the human body, fig.3.1. Passage of X-rays through the body can be described by straight lines and an intensity attenuation along the way x obeying Lambert-Beer's law $\Delta I/I = -f(x)\Delta x$, where the attenuation coefficient $f(x)$ depends on the tissue, bone structure etc. In its integrated form $\ln(I'/I_0) = \int_L dx\, f(x)$ this can readily be seen to be a line integral along a straight line L. Parallel scanning as shown in fig. 3.2a was introduced by A. Cormack in 1963 and also applied in the first commercial scanner. Later fan-beam scanners were used (fig. 3.2b). The field of X-ray tomography has coined the expression *computerized tomography* (also computed tomography) (CT) for computer assisted tomographic reconstruction in the widest sense and (together with later developments like the *single particle emission* CT (SPECT) or the *positron emission tomography* (PET)) it has produced numerous techniques and algorithms for tomographic reconstruction (see chapter 4, especially sec. 4.9). Another highly developed tomographic discipline is seismology that tries to infer the earth's interior from travel time differences of earth quake signals due to different material along their way (*travel time tomography*)[e.g., *Nolet*, 1987]. This technique is also pursued actively by acoustic or electromagnetic emitters in boreholes or on the surface as well as in oceans [e.g., *Munk et al.*, 1995] using either varying transmittance or reflectivity to learn something about the structures within the medium. There are numerous other experimental methods like the *nuclear magnetic resonance imaging* (NMRI) or the *electrical impedance tomography* (EIT) developed for medical use and tomographic applications in areas different from medicine and earth science, e.g. biology, material testing or plasma physics. Even a short survey of experimental techniques is beyond scope, therefore we will concentrate on atmospheric applications in sec. 3.3.

> *Remark on the effect of measurement errors in the example on page 35*
> Trivial as they may be, the example reveals another characteristic feature of tomographic measurements. Suppose that the observer intends to make a very precise measurement for a very thin layer 2, i.e. $l_{B2}/l_{B1} \ll 1$. If the relative measurement errors for the average concentrations $\bar{c}_A$ and $\bar{c}_B$ are $\tilde{\epsilon} \sim \Delta\bar{c}_A/\bar{c}_A \sim \Delta\bar{c}_B/\bar{c}_B$ then straightforward error propagation yields the relative errors for the reconstructed concentrations
>
> $$\frac{\Delta c_1}{c_1} = \tilde{\epsilon}$$
> $$\frac{\Delta c_2}{c_2} \sim \frac{l}{\Delta l}\sqrt{2}\,\tilde{\epsilon} \qquad \text{with} \qquad l = l_A \sim l_B,$$
>
> i.e. the relative error of the reconstructed concentration for the thin layer gets magnified by $l/\Delta l \gg 1$. Anticipating the discussion in sec. 4.4, we remark here that the instable behaviour occurs also in two and three dimensional problems and can in general not be avoided by special choices of the integration paths.

3.2. Atmospheric remote sensing

Among the many experimental techniques to analyse the chemical composition of the atmosphere some methods like spectroscopic methods (gas spectroscopy, mass spectroscopy, optical spectroscopy) are fairly universal while others can be only applied to a group of species or just one species. Following up the preceding section, the most important remote sensing methods are now reviewed, having in mind



(a) Vertical scanning of a NO_2 plume. *(b) Horizontal scan of O_3 over Berlin at 100 m height.*

Figure 3.3: Examples for the range resolving capacity of LIDAR measurements (Courtesy of K. Stelmaszczyk, FU Berlin).

their potential to provide maps of concentration distributions either directly as a range resolving technique or by tomographic measurements. The role of in-situ, i.e. point like measurements for examining concentration distributions will be briefly addressed in sec. 7.4. All remote sensing methods to detect atmospheric trace gases use optical spectroscopy differing either in the wavelength range or the physical phenomenon they exploit. The method is denoted as *active* if it provides an artificial radiation source, *passive* if it uses a natural one. Current experiments use wavelengths from the infrared (IR) to the ultraviolet (UV). Active microwave sensing does not play a role for the detection of trace gases (at least up to now).

3.2.1. IR spectroscopy

Infrared spectroscopy was originally used to detect CO_2. The development of the Fourier transform IR spectroscopy (FTIR) allows the measurement of path averaged concentrations for trace gases like NO, CH_4, HNO_3, CH_2O, HCOOH and H_2O_2. This technique was used recently by *Belotti et al.* [2003] with spatially separated source and detectors to make a tomographic measurement of volcanic CO_2 emissions. The maximum length of the optical paths for this technique is roughly $\sim 1\,km$. Another development is the use of tunable diode laser spectrometers (TDLS) for measurements of NO, NO_2, HNO_3, CH_2O and H_2O_2. Both FTIR and TDLS have been used in a variety of tomographic chamber experiments to study indoor gas dispersion by means of tracer gases such as SF_6 and CH_4 [*Fischer et al.*, 2001; *Samanta and Todd*, 1999; *Drescher et al.*, 1997].

3.2.2. LIDAR

The *light detection and ranging* (LIDAR) technique works similar to the probably better known RADAR (radiowave detection and ranging). By sending out short pulses of (laser) light into the atmosphere and comparing them to the back-scattered signals, in principal, the spatial distribution of aerosols and trace gases along the light path can be retrieved. Because the back-scattered signal is too weak for confident detection of trace gases, measurement at two wavelengths is necessary: at

		SO_2		O_3		NO		NO_2	
λ_{on}	[nm]	300.05		291.40		226.8		448.10	
λ_{off}	[nm]	299.40		300.55		226.83		446.50	
lower (upper) DL	[ppm] ([ppb])	4	(270)	4	(250)	2	(65)	10	(800)
accuracy	[ppm] ([ppb])	±10	(±10)	±10	(±10)	±5	(±5)	±15	(±15)
max. spatial resolution	[m]	3	(500)	3	(500)	3	(250)	3	(1000)
best range	[km]	1	(10)	1	(10)	0.5	(1.2)	3	(19)
mean range	[km]	3	(-)	3	(-)	0.5	(-)	5	(-)

Table 3.1: Specifications for the mobile LIDAR given in [Kölsch, 1990]. Values in brackets (·) refer to measurements with topographical targets. DL stands for detection limit*, see sec. 3.2.3. Wavelengths are listed for completeness only.*

λ_{on} for which the species absorbs and at λ_{off} for which it does not (so-called differential absorption lidar (DIAL)). If mirrors or topographic targets are used to reflect the laser signal, the sensitivity of the method is greatly enhanced but the range resolving power gets lost. The pure back-scattering DIAL can be used for measurement of strong emissions (with a limited range) or vertical immissions. Horizontal and vertical sections of SO_2 plumes from power plants with high spatial resolution have been obtained with a mobile instrument in [*Beniston et al.*, 1990; *Kölsch*, 1990]. Simultaneous measurement of NO and NO_2 from an artificial source (maximum values in the plume around 10 *ppm* for both species) was reported in [*Kölsch et al.*, 1989], see fig. 3.3a. The same system was used to measure horizontal path averaged concentrations of SO_2, NO and NO_2 in various European cities and horizontal distributions of O_3 at 100 *m* above ground over the city centre of Berlin, see fig. 3.3b. Table 3.1 gives an idea about ranges and resolutions for the instrument used. Vertical profiles of O_3 have been obtained in similar experiments [e.g., *Duclaux et al.*, 2002]. In conclusion it can be said that the current state of LIDAR technology allows measurements with

- high spatial resolution in the metre range for distances up to $\sim 1\,km$ only for high trace gas concentrations of the order *ppm*,
- long range (up to $10\,km$) and poor spatial resolution for concentrations on the *ppb* level.

3.2.3. DOAS

The *differential optical absorption spectroscopy* (DOAS) [e.g., *Platt*, 1994] is based on Lambert-Beer's law retrieving integrated concentrations of trace gases along a light path from their narrow band absorption structures. In principle, it can be applied if these have a smaller width than $\sim 10\,nm$ which is the case for many trace gases like O_3, SO_2, NO_2, HONO, CH_2O and BrO. The technique is used both with direct or scattered sunlight and with artificial light sources in the UV, visible and near IR wavelength range. The focus will be on the active DOAS method as the experimental results for this thesis were obtained by this technique.

Method

The Lambert-Beer's law applied to the case of a single gaseous absorber with homogeneous concentration c describes the attenuation of radiation at wavelength λ after passing the distance L by

$$I(\lambda, L) = I_0(\lambda)\, e^{-L\,\sigma(\lambda)\,c} \tag{3.1a}$$

where $I_0(\lambda)$ is the initial intensity and $\sigma(\lambda)$ the wavelength-dependent absorption cross section of the trace gas. Knowledge of the distance, the cross section and measurement of I and I_0 allows to determine the concentration as

$$c = \frac{\tau}{\sigma(\lambda)\,L}, \qquad \text{where} \quad \tau = \ln\frac{I_0(\lambda)}{I(\lambda)} \tag{3.1b}$$

is the so-called *optical density* of the layer of thickness L. For the wavelengths of interest here – the UV and visible – trace gases in the atmosphere exhihibit absorption bands due to rotational, vibrational and electronic excitations, so that a variety of species contribute to absorption at a given λ. Furthermore the light is attenuated also by scattering at the air molecules (essentially through Rayleigh scattering with wavelength dependency $\sim \lambda^{-4}$) and at aerosol particles (usually described by Mie scattering $\sim \lambda^{-n}$ with n often taken around ~ 1.3). Finally atmospheric turbulences lead to an intensity loss by widening the beam. Taking into account spatial dependencies, Lambert Beer's law in the atmosphere thus takes the form

$$I(\lambda, L) = I_0(\lambda)\,A(\lambda)\,\exp\big[-\int_0^L ds \sum_j \sigma_j(\lambda, p, T) c_j(\mathbf{r}(s)) + \epsilon_R(\lambda, \mathbf{r}(s)) + \epsilon_M(\lambda, \mathbf{r}(s))\big] \tag{3.2}$$

where $I(\lambda, L)$ is the intensity at the detector, $c_j(\mathbf{r}(s))$ the number density of species j at position $\mathbf{r}(s)$ along the light path and $\sigma_j(\lambda, p, T)$ its absorption cross section depending also on air pressure p and temperature T. $\epsilon_R(\lambda, \mathbf{r})$ and $\epsilon_M(\lambda, \mathbf{r})$ are the extinction coefficients for Rayleigh and Mie scattering, respectively and the factor $A(\lambda)$ stands for intensity loss due to turbulences. The Lambert-Beer's law describes the statistical behaviour of photons along a straight line from the source to the detector and does not apply straightforwardly to photons that reach the detector via scattering processes, as is the case in scattered sun light experiments. In any case, to infer the c_j from eq. (3.2) in addition to the absorption cross sections at least I_0 and the extinction coefficients have to be known, which is extremely difficult to achieve. Aerosol concentrations are usually not known. I_0 for passive measurements is not known exactly, for artificial light sources it might fluctuate in time, and so forth. The basic principle common to all applications of DOAS relies on the observation that the unknown or unwanted quantities vary only slowly with λ while the absorption cross sections have characteristic narrow band structures *unique* for each species, see fig. 3.4. Dividing the absorption cross sections into broad band and narrow band contributions σ_{0_j}, σ'_j such that for the exponentials in eq. (3.2) $\exp[-\ldots\sigma_j] = \exp[-\ldots\sigma_{0_j}]\,\exp[-\ldots\sigma'_j]$, one gets

$$I(\lambda, L) = \underbrace{I_0(\lambda)\,A(\lambda)\exp\big[-\int_0^L ds \sum_j \sigma_{0_j}(\lambda, p, T) c_j(\mathbf{r}(s)) + \epsilon_R(\lambda, \mathbf{r}(s)) + \epsilon(\lambda, \mathbf{r}(s))\big]}_{I'_0(\lambda, L, p, T)} \times$$
$$\exp\big[-\int_0^L ds \sum_j \sigma'_j(\lambda, p, T) c_j(\mathbf{r}(s))\big].$$

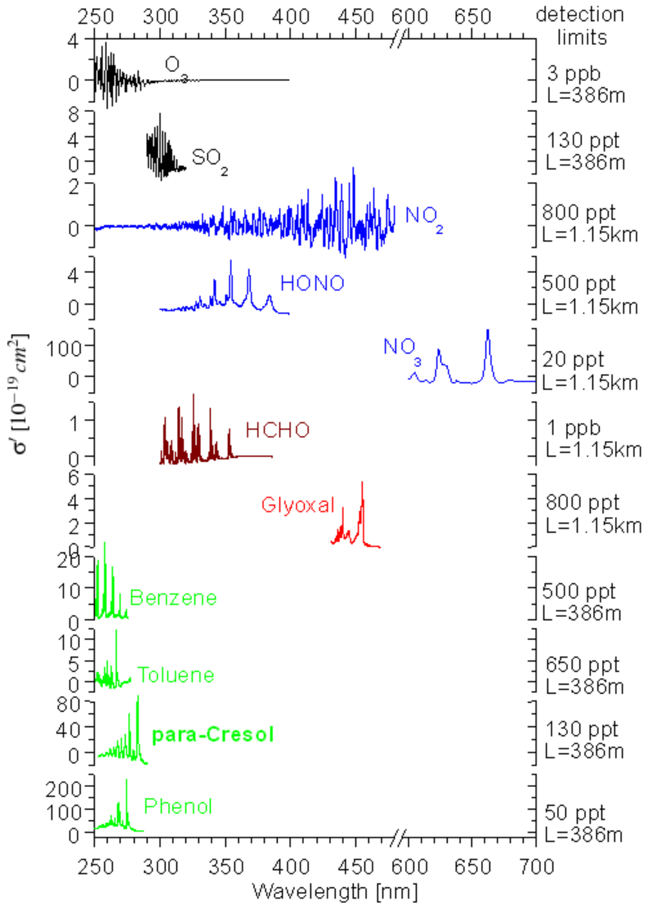


Figure 3.4: Illustration of the characteristic structures of the differential absorption cross sections σ' for selected trace gases that can be measured by DOAS. Detection limits of the mixing ratios for the given light paths length L are only examples (as the detection limits, eq. (3.6)*, depend amongst others on the instrumental details). [Adopted from Volkamer, 2001].*

The last exponential in this equation represents intensity loss by narrow band absorption only, so-called *differential* absorption, while $I'_0(\lambda, L, p, T)$ is the intensity as it would be without differential absorption. Assuming that the absorption cross sections in the lower troposphere, where measurements take place, do not vary with pressure and temperature (which is not vital but simplifies things) leads thus to

$$I(\lambda, L) = I'_0(\lambda, L)\ \exp\big[-\sum_j \sigma'_j(\lambda)\,\bar{c}_j\,L\big] \qquad \text{where} \quad \bar{c}_j = \frac{1}{L}\int_0^L ds\ c_j(\mathbf{r}(s)). \tag{3.3}$$

Inference of the path-integrated concentrations $\bar{c}_j$ as in eqs. (3.1) is still not possible because $I'_0(\lambda, L)$ is not available and the species are not separated. But if the wavelength dependency of $I(\lambda, L)$ is measured within the range where the trace gases of interest show absorption structures, the distinct functional form of the differential cross sections $\sigma'_j(\lambda)$, see fig. 3.4, allows to find the concentrations $\bar{c}_j$ by fitting the $\sigma'_j(\lambda)$ for the relevant trace gases simultaneously to $I(\lambda, L)$. However, to carry out this fit the impact of the measurement system on $I(\lambda, L)$ – that up to now refers to the intensity as it would be *before* entering the detector – has to be taken into account. First, the wavelength analysis, typically by a grating spectrograph, has only limited spectral resolution. This can be expressed mathematically by convolution with a so-called instrument function $H(\lambda)$, i.e. $I(\lambda, L) \otimes H(\lambda) = \int_{-\infty}^{\infty} d\lambda\ I(\lambda - \lambda', L)\,H(\lambda')$. Second, the continuous wavelengths get mapped to pixels or channels i of the recording system. The further procedure for finding $\bar{c}_j$ from the recorded spectrum $I'(i, L)$ depends largely on the following factors: The size of the so-called *differential optical density* $\tau' = L \sum_j \sigma_j'(\lambda)\, \bar{c}_j$ deciding whether the exponential in eq. (3.3) can be Taylor expanded, the spectral resolution of the detector compared to the width of the trace gas absorption bands and finally, structures in the spectrum of the light source $I_0'(\lambda, L)$ compared to the resolution of the instrument.

In the case of a weak absorber, i.e. small optical densities, the exponential can be expanded and, assuming a smooth spectrum of the light source, one gets (after a number of steps omitted here)

$$J(i) = \ln I'(i) = \sum S_j\, \sigma'_{Hj}(i) \ + \text{broad band variations} + \text{noise} + \text{pixel sensitivity} \tag{3.4}$$

with

$$S_j = L\, \bar{c}_j \tag{3.5}$$

and $\sigma'_{H_j}(i)$ being the differential cross section convoluted with the instrument function and mapped to the channels. Broad band variations mean contributions from the lamp, from Rayleigh and Mie scattering etc. Pixel sensitivity refers to variations of the spectral sensitivity of the detector within the pixels or channels. The S_j are called *column densities* (more precisely column number densities or column mass densities).

The fit problem in the special case of weak absorbers is now linear in the S_j and could be solved, for example, by a least squares principle. In practice, the fit is complicated by a number of uncertainties. We mention just two of them. The broad band variations in eq. (3.4) can be taken care of by an additional polynomial in a way that preserves the linearity of the problem. Misalignment of the wavelengths on both sides of eq. (3.4) is a more awkward matter as even a mere shift between the zero positions of the wavelengths enters nonlinearly. It can be caused by inaccurate positioning of the grating, unwanted dispersion or wrong calibration of the absorption cross sections and has to be corrected. In the end, the fit problem is linear with respect to the S_j and the parameters of the broad band correction polynomial, but nonlinear in the parameters for the wavelength alignment. Correspondingly, the fit algorithm consists of a least squares fit for the linear parameters and a nonlinear fit for the wavelength alignment, where for the latter the Levenberg-Marquardt iteration has proven useful.

The case of weak absorption and smooth spectrum of the light source described above is the one most relevant within this work. A general discussion can be found in [*Platt*, 1994], but as the tomographic reconstruction only depends on the column densities as end products, it is of minor importance here. We would just like to point out the rigorous formulation of DOAS in terms of operators (for the convolution with the instrument function, the discretisation of the spectrum etc) in [*Wenig et al.*, 2005] which becomes useful to keep track of various approximations, especially for passive measurements.

Applications

DOAS is applied on various platforms ranging from ground based experiments over airplane and balloon measurements to satellite observations. Measurements on ground employ both artificial light sources and the sun, while air and space borne applications use either direct or scattered sun light (we neglect sources as the moon or stars).

Among the passive techniques notably the so-called *multi-axis* (MAX) measurements have found application on ground [e.g., *Heckel et al.*, 2005; *Hönninger et al.*, 2004*a*,*c*,*b*; *Wagner et al.*, 2004] and on airplanes [e.g., *Heue et al.*, 2005; *Petritoli et al.*, 2002]. Here, scattered sun light is detected in several lines of sight and allows thus conclusions on the vertical distribution of trace gas, in particular the distinction of tropospheric and stratospheric contributions, or on concentration and extension of emission plumes [e.g., *Bobrowski et al.*, 2003].

Active DOAS measurements relevant in our context typically consist of a telescope emitting light from a broad band light source (e.g., a Xenon arc lamp) and retro-reflectors that send the light beam back to the instrument where the receiving part of the optics transmits the light to the detector. As the detection of trace gases like O_3, NO_2 or SO_2 typically requires light paths of the order of a few hundred m up to $10 - 20\,km$ this kind of technique is referred to as *long-path* (LP) DOAS [e.g., *Stutz and Platt*, 1997]. Among other things, it has been very important for the detection of radicals like OH and NO_3 [*Perner et al.*, 1976; *Platt et al.*, 1980; *Geyer et al.*, 2001], the important point being that it is a non-contact measurement.

Active measurements have the advantage of being conceptionally simple as the light travels along a defined path but they have the disadvantage of of being logistically expensive and being restricted by the setting up of the very sensitive telescope and the arrangement of the retro-reflectors. Passive measurements like MAX-DOAS, on the other hand, are easily set up with the instruments getting smaller and cheaper. But the technique itself is less sensitive than the active method and data interpretation is in general not possible without employing numerical models that simulate the transport of solar radiation through the atmosphere, so-called radiative transport models (RTMs).

Measurement errors and detection limits

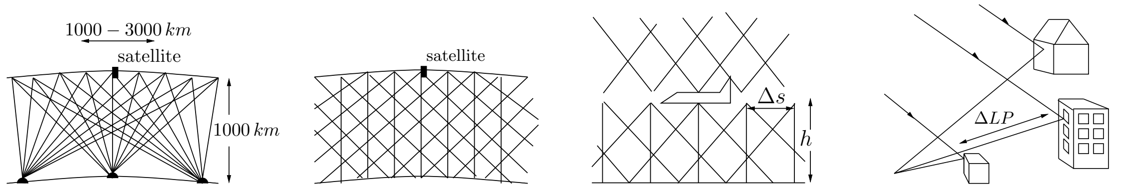
The smallest amount of trace gas detectable with a given method, the *detection limit*, is a key value for the sensitivity of this method. As only the combination $\sigma'_j\,\bar{c}_j\,L$, i.e. the optical density τ', enters the argument of the exponential in eq. (3.4) there is no universal concentration limit, but only detection limits for specific species and lengths of the light path. A proper estimate of the detection limit has to be based on actual measurement errors, which are either of statistic or systematic nature. Random noise can be caused by the photon statistics or the electronic components of the instrument etc. While it is trivial to calculate its propagation from the recorded spectrum to the column densities in the case of the linear absorption least squares fit for the linear case eq. (3.4) (namely as in eq. (4.40)), the effect of the nonlinear wavelength alignment has to be found from simulations. The following estimate for the detection limit of the retrieved $S_j = \bar{c}_j\,L$ takes into account the linear fit only and defines that a species is detectable if for the relative error $\Delta S_j/S_j = \Delta\bar{c}_j/\bar{c}_j > 0.5$ [*Stutz and Platt*, 1996]. The detection limit δS_j for the S_j is then

$$\delta S_j \sim \frac{6}{\sqrt{m-1}}\,\sigma_j \tag{3.6}$$

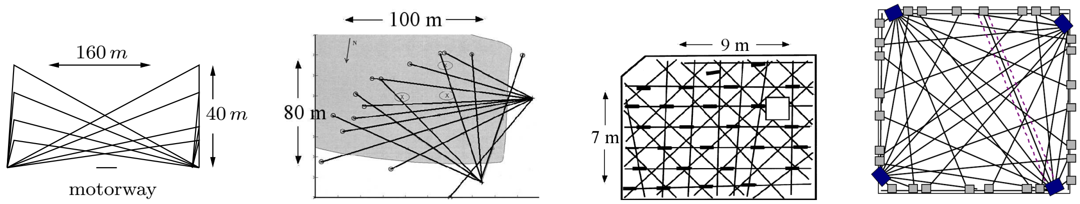
where $\sigma_j^2 \sim \mathrm{var}[S_j]$ and m is the number of degrees of freedom for the measurement, i.e. the number of channels or pixels of the detector. The order of the detection limit for some of the trace gases important for DOAS are contained in fig. 3.4.

The impact of systematic errors is more difficult to estimate in a general fashion. Systematic errors can be introduced by faulty reference cross sections, but often their origin remains unclear. Common sources are stray light in the spectrograph, an offset of the measured intensities etc. Systematic errors will not be considered anywhere in the following.

3.3. Tomographic applications in the atmosphere



(a) Ionospheric tomography *(b) Proposal for satellite emission tomography* *(c) Tomographic AMAX--DOAS* *(d) Proposal for tomographic MAX-DOAS*



(e) Vertical LP-DOAS tomography of motorway emissions *(f) Horizontal IR tomography of volcanic* CO_2 *(g) Indoor gas tomography* *(h) Indoor gas tomographic study*

Figure 3.5: Geometry of optical paths for some tomographic applications in the atmosphere. (a) Ionospheric tomography, three ground receivers are shown. (b) Proposal for IR or microwave emission tomography from space by [Fleming, 1982], here with three lines of sight. (c) Multi-axis airplane measurements [Heue, 2005], shown three downwards and two upwards looking telescopes with arbitrary lines of sight. For Δs, h *see text. (d) Principle of tomographic measurements by (passive) multi-axis DOAS with defined light paths* ΔLP *as suggested in [Frins et al., 2006]. (e) Vertical 2-D projection of the geometry with two telescopes and eight reflectors used for long-path DOAS measurements at a motorway [Laepple et al., 2004]. (f) Tomographic FTIR measurements of volcanic* CO_2 *[Belotti et al., 2003]. (g) Measurement geometry for a gas chamber FTIR experiment [Fischer et al., 2001] and (h) a theoretical study [Verkruysse and Todd, 2004].*

Having briefly sketched the principle of tomography and presented the most important atmospheric remote sensing methods, we turn now to the combination of both techniques. The reconstruction of atmospheric parameters like temperature, pressure or concentrations from tomographic remote sensing instead of conducting, for example, a corresponding number of point measurements can be motivated by various reasons.

- The advantages of the remote sensing method can be used. This could be contactless measurement of reactive gases or the simultaneous measurement of species for the special case of DOAS etc.
- Point measurements providing the same spatial resolution and/or quality are too expensive. The best example here are probably global maps of trace gas columns from satellite observations. Air pollution monitoring on a microscale has been addressed in sec. 2.1.
- In-situ sampling is difficult or impossible. Volcanic emissions have been mentioned, wood fire or noxious gas are further examples.

- Although not primarily intended, tomography in the atmosphere profits from the fact that stochastic variations of the concentration field due to turbulent fluctuations are reduced by integrating, i.e. averaging, over the optical path (see also sec. 7.4).

Some measurements and proposals using tomography with different remote sensing techniques are now being presented in the order of the spatial scale involved. DOAS tomographic applications will be introduced in the next section.

An application on the largest scale is the ionospheric tomography that tries to reconstruct the spatial distribution of the electron number density in the ionosphere, i.e. the atmospheric layer roughly between 80–1000 km, by its influence on the propagation of radio waves through changes in the diffraction index [e.g., *Kunitsyn et al.*, 1995]. Frequency dispersion, Doppler shifting, angle of arrival or signal attenuation can be used to measure the total electron content, that is the integral of the electron density along the line between emitter and receiver as in the example in fig. 3.5a [*Fehmers*, 1996]. Here, an orbiting satellite sends out radio signals to ground receivers and the total electron content along all lines of sight derived by the differential Doppler technique allows reconstruction of two dimensional vertical distributions.
While the first proposal for ionospheric satellite measurements was made in 1986 [*Austen et al.*, 1986], *Fleming* [1982] simulates (in a to my opinion extraodinarily intelligible paper) the two dimensional reconstruction of temperature distributions from IR (or microwave) radiation measured by satellite and suggests similar measurements to retrieve concentration fields, see fig. 3.5b. He compares temperature profiles reconstructed from hypothetical measurements along three and five viewing angles (zero degree zenith, $\pm 45°$ and additionally $\pm 60°$) to conventionally retrieved profiles and finds improvements of the accuracy up to 34%. It should be noted that he assumes measurements at five different frequencies (channels) for each line of sight. To my knowledge, satellite observations of this kind have not been realised so far.

Somewhat earlier than these travel time and emission tomographic satellite measurements *Byer and Shepp* [1979] proposed absorption LIDAR for (horizontal) two dimensional tomographic measurement of air pollution. The reflector-LIDAR system assumed for the simulations in [*Wolfe and Byer*, 1982] consists of a laser source and mirrors that reflect fan beams[2] through the area of interest ($\varnothing \sim 3\,km$) to a detector array. With around 100 virtual sources and 70 detectors, the authors find that a spatial resolution of less than 200 m becomes realistic (an adjective that might not apply to the setup – as far as I know, this setup has not been realised with anything near the number of optical sources and detectors assumed here).

Horizontal temperature distributions on an area of about $200 \times 200\,m^2$ obtained by acoustic travel time measurements are compared in [*Weinbrecht et al.*, 2004] with highly resolved results from large eddy simulations. The measurements used a moderate number of six acoustic sources and five receivers at a height of 2 m above ground.

The tomographic IR measurements in an active volcanic region have already been mentioned [*Belotti et al.*, 2003]. They were performed on an area of $\sim 100 \times 100\,m^2$ measuring CO_2 absorptions successively along 15 optical paths given by the laser transmitter/receiver system and a retro-reflector (fig. 3.5f). The detection limit for the column density is stated as $1000\,ppm \cdot m$ and the sampling time per path was chosen between 5–10 min. Reconstructed CO_2 varies between $\sim 350\,ppm$ (background)

---

[2] Buildings, trees or topographic objects could also be used at the expense of intensity.

and 3 700 *ppm*.

Finally, the indoor gas dispersion studies already mentioned try to reconstruct gas distributions on a metre scale from IR measurements. Although not really being atmospheric applications they are among all studies listed the most relevant from a geometrical point of view as some of their light path configurations are very similar to what can be expected from tomographic LP-DOAS experiments (figs. 3.5g, 3.5h). For the same reason their approach to reconstruction of concentration fields is worth looking at.
Todd and Leith [1990] start ambitiously with a theoretical study to detect contaminants like ammonia, benzene, tri- and perchloroethylene by FTIR spectroscopy for parallel projection like beam geometries with as many as 400 paths. This boils down to 136 paths actually realised in an experiment [*Samanta and Todd*, 1999] with four spectrometers by rotating the sources successively towards different mirrors and eventually down to 40 rays in a series of theoretical studies on path configurations and reconstruction algorithms [*Todd and Leith*, 1990; *Todd and Ramachandran*, 1994*a*,*b*; *Todd and Bhattacharyya*, 1997]. For similar experiments *Drescher et al.* [1996] present an approach especially suited for the reconstruction of point emissions (SBFM, see sec. 4.9.3) and find from a detailed time series analysis that the evolution of an indoor gas emission shows far more variability than expected [*Drescher et al.*, 1997]. The method is adopted in [*Philip*, 1999; *Hashmonay et al.*, 1998] to investigate the possibility of reconstructing point sources by radial scanning with one rotating optical source only. SBFM is given up in [*Fischer et al.*, 2001] for the reconstruction of CH_4 tracer distributions released in a chamber experiment for being too slow (see footnote 3, p. 101).

3.4. First tomographic DOAS measurements

The idea to infer information on the spatial distribution of trace gases from DOAS remote sensing along more than one light path has been realised in profile measurements with retro-reflectors on mountain sites [*Platt*, 1978] or mounted on towers [e.g., *Stutz et al.*, 2004], c.f. fig. 3.1a on p. 35, and also in balloon experiments [*Veitel et al.*, 2002].

A genuine tomographic experiment aimed at the 2-D vertical reconstruction of the exhaust plume perpendicular to a motorway from LP-DOAS measurements [*Pundt et al.*, 2005]. The experiment was part of a joint campaign including in-situ measurements and model studies [*Corsmeier et al.*, 2005*a*]. It utilised two LP telescopes each sending successively one light beam along the motorway to one of eight retro-reflector arrays mounted on two towers next to the carriage way (fig. 3.5e). Scanning all reflectors to achieve the complete geometry of 16 light paths took about 45 *min*. With optical path lengths around 800 *m*, NO_2 concentrations could be measured most reliably and were thus considered for tomographic reconstruction in [*Laepple et al.*, 2004]. For a time averaging interval of 3 *h*, the pure DOAS error on the column densities was about $2-5\%$, while the error due to the successive scanning was estimated from time series of the column densities about to be 2%. Maximum reconstructed NO_2 concentrations in the plume lie around 20 *ppb* and the maximum of the mean reconstruction error was estimated to be roughly 5 *ppb* which is about as much as the maximum deviation from model predictions.

The motorway experiment pointed out a major problem of active DOAS tomography: To generate as many light paths as possible within an acceptable time one has to either employ as many light sources (telescopes) as possible or direct the telescopes successively at the reflectors (‘*scanning*’) or a combination of both. The first is (at least at the moment) prohibitively expensive, the second not easy to accomplish as conventional LP instruments are rather heavy and have to be installed stable. This led to the construction of a telescope that emits up to four light beams at once [*Mettendorf*, 2005; *Pundt and Mettendorf*, 2005]. Three instruments have been tested in a tomographic indoor experiment [*Mettendorf et al.*, 2006] that used one or two cylindric containers filled with NO_2 to mimic puff emissions. A geometry corresponding to 36 integration paths was realised in three steps à 4 beams per telescope. While spatial dimensions can be scaled up to atmospheric experiments this is not true for the time scales of the measurement. As the reconstruction from the column densities of this experiment was part of this thesis, further details will be presented in chap. 9. The instruments are currently used for 2-D horizontal measurements over the centre of Heidelberg.

The tomographic principle can be applied straightforwardly to DOAS measurements of sun light if scattering can be neglected, i.e. for observations of direct or directly reflected sun light, which restricts possibilities somehow, unless some kind of mirrors are used. For tomographic measurements of scattered sun light – depending on the degree of homogeneity assumed – 1-D, 2-D or 3-D simulations have to model the light paths. These radiative transport models are predominantly 1-D as calculations in two and three dimensions are quite involved. The Monte-Carlo algorithm TRACY-II [*Wagner et al.*, 2006] is one of the very few RTM capable of both 2-D and 3-D simulations.

First passive tomographic measurements have been attempted by installing multi-axis DOAS instruments onto an airplane (AMAX-DOAS) with three telescopes looking upwards and seven looking downwards [*Heue*, 2005]. Thus sun light from above the airplane and reflected from ground can be used for tomographic reconstruction, see fig. 3.5c. Flying at heights h between 500–2 000 m above ground at a speed of about 200 $\frac{km}{h}$ together with integration times around 30 s for the visible wavelength range and 11 s in the UV results in measurement points every $\Delta s \sim 1.7\,km$ (VIS) and $\sim 600\,m$ (UV). The method was used to examine the emission plume of a power plant in the Po valley. Preliminary 2-D vertical recontructions of the NO_2 plume were obtained from 2-D RTM calculations.

A very interesting idea put forward in [*Frins et al.*, 2006] amounts to using passive MAX-DOAS in a way similar to LP-DOAS, thus combining the benefits of the former (simple setup, no reflector arrays necessary) and the latter (simple concept, especially for tomographic measurements). As for LP-DOAS, the objective is to measure near-surface concentrations and this is achieved by pointing the telescope at a target in a way that the light reflected from it passes the region of interest. This part of the light path is well defined and all one has to do is to subtract the remaing part from the top of the atmosphere to the target (fig. 3.5d), which can be done almost routinely. With several (relatively cheap) mini MAX-DOAS instruments and suitable targets tomographic experiments identical to LP-DOAS measuremets can be set up. Drawbacks coming along with the passive technique are that measurements depend, of course, on the sun light, they might not be as sensitive as LP-DOAS measurements and if scattering between target and instrument cannot be neglected, radiative transport modelling gets involved.

	integration paths m	model params. n	m/n	area $[km^2]$	time $[min]$	reconstruction method[a]
Axial X-ray	$\mathcal{O}(10^5)$	$\mathcal{O}(10^5)$	$m > n$	$\mathcal{O}(10^2)\,cm^2$		transf. meth.
typical	~ 500 rays$\times$ $(10^2 - 10^3)$ proj.	$\sim 512 \times 512$ pixels	~ 4			FBP
Seismology	$\sim (10^6 - 10^8)$	$\sim (10^5 - 10^6)$	$m \gg n$			(non)lin. LS (LSQR, CG)
2-D Satellite tomography of ionosphere[b]	~ 500	100×20	$1/4$	~ 2000 $\times$ 600	10 ... 20	constrained LS, constr. opt.
AMAX-DOAS[c]	60			20×2	~ 6	
LP-DOAS	$\mathcal{O}(10)$	$\mathcal{O}(10)$		$(0.5 \ldots 10)\times$ $(0.5 \ldots 10)$	$\mathcal{O}(1)$ $\ldots \mathcal{O}(10)$	
motorway[d]	16	5×4	0.8	$160 \times 40\,m^2$	~ 45	SIRT
Indoor FTIR[e]	28	63	~ 0.4	$7 \times 9\,m^2$	$7\,s$	general. LS
[f]	56	$6 - 18$	$m > n$	$\sim 6 \times 7\,m^2$	6	SBFM

---

[a] See chap. 4.
[b] [*Fehmers*, 1996]
[c] For the flight around a power plant reported in [*Heue*, 2005] and a single scattering approximation.
[d] [*Laepple et al.*, 2004]
[e] Simultaneous measurement with 28 instruments by *Fischer et al.* [2001].
[f] Sequential measurement with one source and fix mirrors [*Drescher et al.*, 1996].

Table 3.2: Dimensions of some tomographic applications.

4. Tomography and Discrete, Linear Inverse Problems

The previous chapter has clearly shown that tomographic DOAS measurements lie beyond classical applications of tomography. Consequently, there is no standard method to gain the concentration distribution from the measured line integrals. Indeed, it is not exaggerated to say that there are as many suggestions for reconstruction methods as there are potential or actual applications. It will become clear in parts II and III that the choice of the reconstruction procedure for a limited set of measurement data is a delicate matter and we prefer to discuss it on the background of conventional methods. This implies jargon, concepts and algorithms from computerised tomography, image processing, inverse theory and atmospheric profiling and, as I am not aware of a suitable synoptic reference, I have tried to treat the matter as concise as possible in a common, uniform context. Still the coverage is by no means complete and especially the discussion of regularisation only scratches the surface.

First we relate the tomographic problem to other inverse problems given by integral equations and, by considering continuous tomographic inversion methods, are then led to a discretisation of the problem. The reconstruction method chosen here within the discrete approach is essentially least squares minimisation which on the one hand has the advantage of leading to linear systems of equations. On the other hand, the least squares formalism arises naturally if the underlying statistics are assumed to be normally distributed. We briefly review solution strategies for the least squares problem and elaborate on a family of iterative algorithms well known in computerised tomography and image reconstruction, but not common in atmospheric sciences. If not self evident or stated explicitely in this chapter, justification of the approach will be given in sec. 6.1. The chapter concludes by brief descriptions of alternative approaches for later reference.

4.1. Forward model and inverse problem

The formulation and solution of the tomographic reconstruction problem is carried out for the concentration distribution of each species seperately and we write

$$d \equiv S_j, \qquad \text{see eq. (3.5)},$$

for the column densities obtained by the DOAS analysis for the species under consideration (d standing for the input *data* of the inverse problem). As starting point we assume that an ideal long-path (LP) DOAS measurement along a light path LP with parametrisation $LP : \mathbf{r} = \mathbf{r}(p) \in \mathbb{R}^3$, $p \in [a, b]$, provides a column density d measured at time t which simply is the line integral of the concentration distribution c at the same time. That is

$$d(t) = \int_a^b dp \, |\tfrac{d\mathbf{r}(p)}{dp}| \, c\,(\mathbf{r}(p), t) \tag{4.1a}$$

$$= \int_{LP} ds \, c(\mathbf{r}, t), \tag{4.1b}$$

where in the shorthand notation eq. (4.1b) ds denotes the line element along LP and $\mathbf{r} = \mathbf{r}(s)$.

The model eq. (4.1) represents a measurable cause-effect relationship between a given concentration field and the column densities that result from it – in contrast to the pure mathematical backward inference of the unknown concentration field from the measured integrals – so that it is commonly referred to as a *forward model*. It relies on the assumptions:

- The light path is known.
- The beam diameter is negligible.
- There is no significant time delay between absorption process and the recording of the absorption spectrum.

In this work we do not take into consideration any modifications that arise from relaxing the above assumptions except for errors that ultimately can be expressed as random noise on the column densities:

$$d_\epsilon = d + \epsilon. \tag{4.2}$$

Here d_ϵ is the error afflicted measured column density, d the ideal one and ϵ an (unbiased) random quantity. Time averaging does not cause trouble as long as it acts consistently on both sides of eq. (4.1)

$$\overline{d(t)}_{\mathcal{T}} = \int_{LP} ds \, \overline{c(\mathbf{r}, t)}_{\mathcal{T}}, \tag{4.3}$$

$\mathcal{T}$ indicating the time averaging interval $\mathcal{T} = [t; t + \Delta t]$. This is the case, for example, when averaging over subsequent spectra to reduce stochastic noise. If data is taken by the same instrument (e.g. the multi-beam instrument, see sec. 3.4, that emits several light beams simultaneuously) the question of correlated measurement errors has to be addressed. The (idealised) tomographic inverse problem can now be stated as follows.

LP-DOAS TOMOGRAPHIC RECONSTRUCTION PROBLEM

Given the column densities d_i of LP-DOAS measurements along m light paths LP_i and a compact set $\Omega \subset \mathbb{R}^3$ containing the LP_i, find a function $c(\mathbf{r})$ over Ω satisfying the forward model

$$d_i = \int_{LP_i} ds \, c(\mathbf{r}), \quad i = 1, \ldots, m, \tag{4.4}$$

where the column densities refer to a particular time: $d_i = d_i(t)$, or time averaging interval: $d_i = \overline{d_i(t)}_{\mathcal{T}_i}$ that is the same for all measurements ($\mathcal{T}_i \equiv \mathcal{T}$).

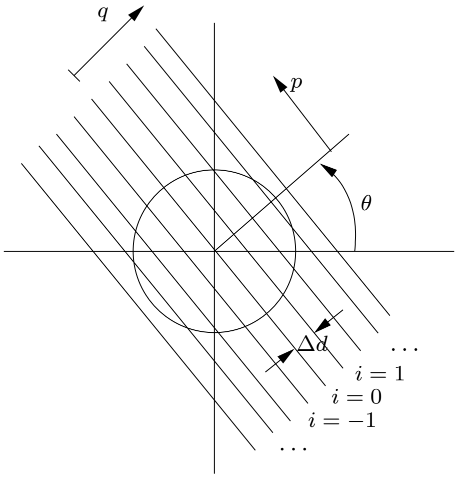


Figure 4.1: Parallel beams in the plane for the projection angle θ_k. The circle indicates the region of interest.

4.2. Continuous inversion methods

The most elegant way to solve problem eq. (4.4) would be analytic inversion of the integral equations. This is indeed possible under certain conditions. But before addressing these transform methods, it is instructive to compare the tomographic inverse problem with another large class of inverse problems given by the integral equation

$$g(q) = \int_0^1 dp\, k(q,p)\, f(p), \quad 0 \le p \le 1. \tag{4.5}$$

Here the integral value is a function of the parameter q of the kernel function k (see [*Groetsch*, 1993] for an elementary introduction). These so called Fredholm equations of the first kind appear in numerous applications. A prominent example in atmospheric research is the integral equation given by the solution of the atmospheric radiative transfer equation and the two textbooks [*Twomey*, 1997; *Rodgers*, 2000] practically exclusively deal with the inversion of this equation. If the discrete light path index i in eq. (4.1) can be transformed into a continuous parameter q, the tomographic problem becomes a Fredholm equation. Consider the 2-D example of parallel light paths in fig. 4.1. Their parameter representation for the projection angle θ is

$$\mathbf{r}_i(p) = i\Delta d\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} + p\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}, \quad i = 0, \pm1, \pm2, \ldots, \quad p \in \mathbb{R}. \tag{4.6}$$

If the distance Δd of the beams is small against all relevant scales (i.e. the length of the area of interest, the length of typical structures within the area etc.), $q = i\Delta d$ effectively becomes a continuous parameter and the line integrals are now

$$d_\theta(q) = \int_{-\infty}^{\infty} dp\; c(x,y), \qquad \text{with} \qquad \begin{pmatrix} x(p,q) \\ y(p,q) \end{pmatrix} = \begin{pmatrix} q\cos\theta - p\sin\theta \\ q\sin\theta + p\cos\theta \end{pmatrix}. \tag{4.7}$$

Apart from the fact that the unknown function depends on two parameters instead of one as in eq. (4.5) the kernel is trivial so that inference from the one dimensional $d_\theta(q)$ to the two dimensional $c(x,y)$ is impossible. Providing further projections, it turns out that the kernel is both problem and cure as its special form allows the use of very efficient transform methods. Discrete methods treat $c(x, y)$ as a one dimensional object and the resulting discrete system is formally identical to the system obtained from discretising a Fredholm equation – but only formally, as will be discussed in sec. 4.4.

4.2.1. Transform methods

Transform methods rely on a quasi-continuous representation of the measurement results while the parametrisation itself depends crucially on the configuration of the integration paths, i.e. the kind of projection. We resume the example of the parallel projection, fig. 4.1, for its simplicity, although this kind of projection is not realistic for a DOAS tomographic measurement. Other cases like fan beam projections (fig. 3.2b) and incomplete projections can be found in the application-oriented introduction by *Kak and Slaney* [2001] and in the extensive, mathematical coverage [*Natterer*, 2001].

The transform from the 2-D space for the (concentration) field to the space of projections of c with its two dimensions θ and q in eq. (4.7) is called *Radon transform.* If the projections along the q-axis are measured for sufficiently many angles θ, inversion of the Radon transform becomes feasible. Straightforward Fourier transformation does the job, but does not take into account the geometry of the problem to simplify the solution. There is a number of formulations of the inverse Radon transform that do exploit the nature of the projection, but they are numerically not very stable. By far the most widespread way to invert the Radon transform in practical and commercial applications is the so called *filtered backprojection* (FBP). It was also used in the theoretical study on the reconstruction of ambient trace gas distributions from tomographic LIDAR measurements [*Wolfe and Byer*, 1982] that was mentioned in sec. 3.3.

The filtered backprojection can be motivated by considering the 2-D Fourier transform

$$\tilde{c}(k_x, k_y) = \iint\limits_{-\infty}^{+\infty} dx\, dy\; e^{-i2\pi(k_x x + k_y y)}\; c(x, y)$$

which in the case $k_y = 0$, i.e. no dependence in y, takes the form

$$\tilde{c}(k_x, 0) = \int\limits_{-\infty}^{+\infty} dx\, e^{-i2\pi k_x x} \Big[\int\limits_{-\infty}^{+\infty} dy \;\; c(x, y) \Big] .$$

The term in brackets is the line integral along the y-axis $d_{\theta=0}(q)$. Thus one gets the relation $\tilde{c}(k_x, 0) = \widetilde{d_{\theta=0}}(k_x)$ for the Fourier transforms. Obviously this result cannot depend on the choice of the coordinate system and by rotating the axis one gets more generally the so called *Fourier slice theorem*

$$\tilde{c}(k\, \cos\theta, k\, \sin\theta) = \widetilde{d_\theta}(k)\, .$$

In other words, the 2-D transform of c along the projection is given by the 1-D Fourier transform of the projection. This reduces the number of integrations in the inverse Fourier transform for $c(x, y)$ to only two:

$$c(x, y) = \int_0^\pi d\theta \int_{-\infty}^{\infty} dk\; |k|\; e^{i\, 2\pi\, q\, k}\; \widetilde{d_\theta}(k) = \int_0^\pi d\theta\; D(q). \tag{4.8}$$

where as above $q = x\cos\theta + y\sin\theta$. D is up to the so called filter factor, here $|k|$, the inverse Fourier transform of $\tilde{d}$ and hence called filtered projection. For a given x, y in the image plane there is a unique tuple θ, q, but eq. (4.8) shows that in fact all points on the line with equal distance q to the origin contribute the same for a given θ. The reconstruction seems to project each filtered projection $D(q)$ back into the image plane along the line of constant q. This is why the transform is called filtered backprojection. In practice, the theoretical filter $|k|$ does not work and has to be modified. One reason is the inherent instability of the inverse problem discussed in sec. 4.4.

4.2.2. Remark on existence and uniqueness

The questions of existence and uniqueness of the solution for the tomographic inverse problem eq. (4.4) are mathematically rigorously discussed by *Natterer* [2001, chap. 2]. He shows irrespectively of the inversion method that for an arbitrary function c and a compact region there is always another function c' for which any finite number of line integrals have the same values on this arbitrary small region. As this holds even with both functions being $C^{(\infty)}$ this sounds rather discouraging. However, c' is a highly oscillating, artificial function and if its variation is restricted in terms of some norm the agreement of c and c' cannot always be achieved and the solution becomes unique. In the framework of the transform method just sketched the question of uniqueness is addressed by sampling theorems similar to the well-known Nyquist theorem. The conditions for the norm of the function of c correspond now to bounds on the frequency range of c, i.e. it has to be band-limited [*Natterer*, 2001; *Kak and Slaney*, 2001, chap. 2]. Further discussion is beyond the scope of this work, as it will deal only with the discretised version of the reconstruction problem.

4.3. The discrete, linear inverse problem

It is clear from the preceding discussion that we are not addressing the kind of discretisation that is eventually necessary for the computational solution by any numerical algorithm, but the reformulation of the problem in terms of a limited number of parameters. For either way of looking at it, there are essentially two ways of discretising integral equations [*Hansen*, 1998].

4.3.1. Discretisation by quadrature

Applying some quadrature rule like the midpoint rule, Simpson's rule etc. leads to the following representation

$$d_i = \int\limits_{LP_i} ds\, c(\mathbf{r}) = \sum_{j=1}^{N_i} w_{ij}\, c_{ij} + \Delta J_i(c), \tag{4.9}$$

where the weights w_{ij} depend on the details of the quadrature (line segments Δs in the simplest case) and $c_{ij} = c(\mathbf{r}_i(s_j))$ are the concentration values on the interpolation nodes s_j along path i. The functional ΔJ_i stands for the error due to the discretised evaluation of the integral. To make use of the tomographic character of the measurement concentration, values for different rays have to be set to equal values so that the individual equations (4.9) get coupled. This can be done by identifying concentrations in certain regions, e.g. cubes in three dimensions, and adjusting the interpolation nodes to them. For the most simple case that approximates the integral as $\int ds\, c(\mathbf{r}(s)) \approx \sum_{j=1}^{N} \Delta s_j\, c(\mathbf{r}(s_j))$

and uses n cubes in which concentrations are identical one gets:

$$d_i = \sum_{j=1}^{n} A_{ij}\, c_j + \Delta_{disc} J_i(c).$$

Here the sum is over the cubes and the weight A_{ij} is the length of ray i in cube j. The number $\Delta_{disc} J_i(c)$ represents the error that comes with the approximations of the integral and the function c. Obviously, for a given level of discretisation, i.e. number of cubes, applying a very sophisticated quadrature is somewhat pointless, if concentration values over a large region are forced to equal values. Refined schemes involving 2-D or 3-D interpolation would be necessary, however, we do not pursue this issue any further as there is a more elegant and consistent way of discretising the problem.

4.3.2. Finite element discretisation - local basis functions

Instead of approximating the integral one can discretise the function c first by representing it on a finite set of so called *basis functions*[1] $\{\phi_j\}_{j=1}^{n}$:

$$c(\mathbf{r}) = \sum_{j=1}^{n} x_j\, \phi_j(\mathbf{r}) + \Delta_{disc}(c(\mathbf{r})), \tag{4.10}$$

where $x_j \in \mathbb{R}$ are the n parameters of the representation and Δ_{disc} stands for the error related with it. Inserting this into the forward model eq. (4.4) yields

$$d_i = \sum_{j=1}^{n} A_{ij}\, x_j + I_i$$

with

$$A_{ij} = \int\limits_{LP_i} ds\, \phi_j(\mathbf{r}) \quad \text{and} \quad I_i = \int\limits_{LP_i} ds\; \Delta_{disc}(c(\mathbf{r})). \tag{4.11}$$

Depending on the literature consulted this approach is mostly referred to as Galerkin's method or finite element method (FEM). The latter term implies that the basis functions have compact support.

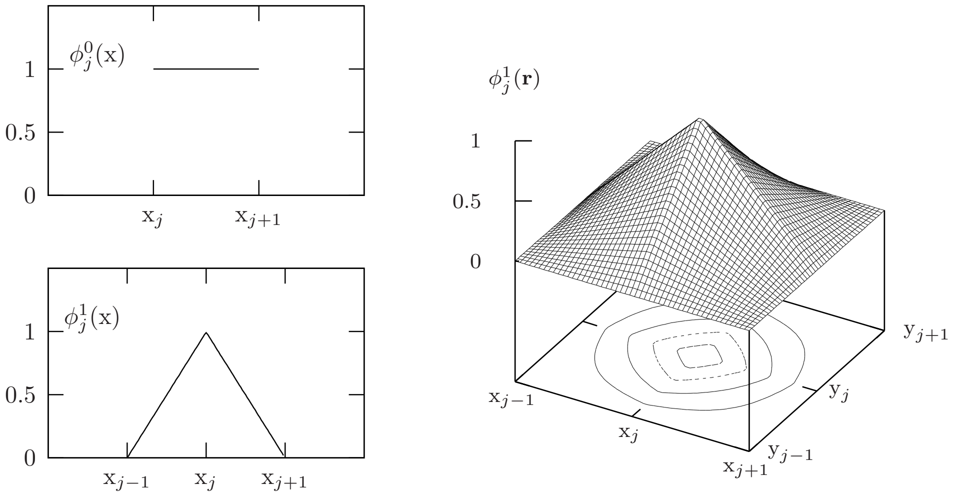
Several aspects are relevant for the choice of the basis functions:

- continuity, smoothness etc.
- computational expense ($m \times n$ integrals A_{ij} have to be calculated)
- regularisation behaviour (this will be discussed later in sections 4.6.3 and 8.2.2)
- other mathematical reasons (e.g., do derivatives appear in the problem ?)

For large problems computational economy will dominate over the aim to minimise discretisation errors. In fact, at one end of the scale of this trade-off there are applications of computerised tomography for which use of any other than piecewise constant basis functions is too expensive [*Kak and Slaney*, 2001]. For moderate numbers of light paths, DOAS tomography finds itself at the other end

---

[1] This term is slightly misleading as, in general, the ϕ_j do *not* form a basis of the function space of the concentration fields c.



(a) *(b)*

Figure 4.2: (a) The piecewise constant and linear basis function, or splines of degree 0 and 1, respectively. The support (finite element) is defined by two and three nodes. (b) The (2-D) bilinear basis function given by eq. (4.14) with contour lines in the bottom graph.

of this scale where computation time does not play any role – at least not for calculating the integrals eq. (4.11) – and more expensive basis functions may lead to a better representation and solution of the problem.

Turning to their explicit form, the basis functions are divided into two categories: A set of basis functions $\{\phi_j\}$ is called *local* if the ϕ_j have compact support $\Omega_j \subset \Omega$. It is called *global* if $\Omega \subseteq \Omega_j$. Legendre and Chebyshev functions are examples of global polynomial basis functions in one dimension, another example that relates closely to the tomographic reconstruction problem will be met in sec. 4.9.3. A most simple example of a set of local basis functions is the aforementioned piecewise constant basis (over the disjoint union of the Ω_j)

$$\phi_j(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in \Omega_j \\ 0 & \text{else} \end{cases}, \quad \text{with}\, \Omega_j \cap \Omega_k = \emptyset \;\text{ for}\, j \neq k \;\text{ and }\; \cup_j \Omega_j = \Omega.$$

A widely used class of local basis functions consists in functions built from piecewise polynomials, motivated by the observation that smooth functions can be approximated to arbitrary accuracy by polynomials of the right degree. In the widest sense, piecewise polynomial functions used for interpolation and/or smoothing are called *splines*. Fig. 4.2a shows basis functions constructed from the most simple splines of degree 0 and 1 for the case of one spatial dimension. The functional form $\phi_j^0(\mathrm{x})$ for the spline of degree 0 reads[2]

$$\phi_j^0(\mathrm{x}) = \begin{cases} 1 & \text{if } \mathrm{x} \in [\mathrm{x}_j, \mathrm{x}_{j+1}[\\ 0 & \text{else.} \end{cases} \tag{4.12}$$

The two parameters in $b\,\mathrm{x} + a$ for the representation of $\phi_j^1(\mathrm{x})$ are given by demanding $\phi_j^1(\mathrm{x}_j) = 1$ and

---

[2] The nodes x_j here must not be confused with the parameters x_j in eq. (4.10). There will be no cause for confusion in the following.

$\phi_j^1(\mathrm{x}_{j-1}) = 0 = \phi_j^1(\mathrm{x}_{j+1})$:

$$\phi_j^1(\mathrm{x}) = \begin{cases} \dfrac{\mathrm{x}-\mathrm{x}_{j-1}}{\mathrm{x}_j-\mathrm{x}_{j-1}} & \text{if } \mathrm{x} \in [\mathrm{x}_{j-1}, \mathrm{x}_j[\\ \dfrac{\mathrm{x}_{j+1}-\mathrm{x}}{\mathrm{x}_{j+1}-\mathrm{x}_j} & \text{if } \mathrm{x} \in [\mathrm{x}_j, \mathrm{x}_{j+1}[\\ 0 & \text{else.} \end{cases} \tag{4.13}$$

Evidently, representation by the piecewise linear basis functions (also called *hat* or *tent* functions) is equivalent to linear interpolation between two nodes. Higher order polynomials can be constructed by including nodes other than the neighbouring ones or by providing conditions on the derivatives. The next higher order basis function emerging naturally in this context is the cubic basis function, basically because the four parameters in $d\,\mathrm{x}^3 + c\,\mathrm{x}^2 + b\,\mathrm{x} + a$ can be determined by the function values plus slopes on two neighbouring nodes. An alternative construction leading to the so called *natural cubic spline* with continuous first and second derivatives can be found in [*Press et al.*, 1992].

A simple way to build higher dimensional basis functions is to form tensor products of lower dimensional ones. In two dimensions this leads for the above cases to the piecewise constant

$$\phi_j^0(\mathbf{r}) = \phi_j^0(\mathrm{x})\phi_j^0(\mathrm{y})$$

and bilinear basis functions (fig.4.2b)

$$\phi_j^1(\mathbf{r}) = \phi_j^1(\mathrm{x})\phi_j^1(\mathrm{y}), \tag{4.14}$$

with support $[\mathrm{x}_j, \mathrm{x}_{j+1}[\times[\mathrm{y}_j, \mathrm{y}_{j+1}[$ and $[\mathrm{x}_{j-1}, \mathrm{x}_{j+1}[\times[\mathrm{y}_{j-1}, \mathrm{y}_{j+1}[$, respectively, i.e. rectangular areas for both functions. As in the 1-D case, the piecewise constant basis is orthogonal (supports do not overlap) whereas a piecewise linear basis function does overlap with its nearest neighbours. Representation by this bilinear basis corresponds to 2-D interpolation (on a square grid). Piecewise constant and trilinear basis functions in three dimensions with their supports being cubes can be constructed similarly.

Of course, there is a far larger variety of basis functions, e.g. piecewise Hermite polynomials or B-splines (a generalisation of the Bézier curve) – just to mention polynomials. And disciplines like digital image processing or the finite element method for solving differential equations use basis functions with support chosen to optimally suit the problem at hand, e.g. in two dimensions triangular and quadrilateral elements instead of the rectangular above. On the other hand, many of the examples of tomographic reconstruction cited in chapter 3 that take a discrete approach do not use particularly elaborate discretisation schemes. One reason might be the computational expense mentioned before, especially if the spatial resolution given by the size of the meshes of the discretisation grid is high anyway. For applications with low spatial resolution like the LP-DOAS tomography a different argument becomes important: Contrary to a problem of best function approximation where the number of parameters is limited by computer memory or calculation time, for a discrete inverse problem every free parameter coming along with the discretisation has to be determined from the limited amount of data. And it is by no means clear that established discretisation schemes work here as well as for their original application. Fig. 4.3 illustrates a very simple case. To parametrise a concentration field on 2-D finite elements given by rectangles as shown in the figure, one parameter per element is required for the piecewise constant basis functions whereas the four nodes of the bilinear basis carry four parameters. In three dimensions, a cubic finite element still only needs one parameter, the trilinear basis

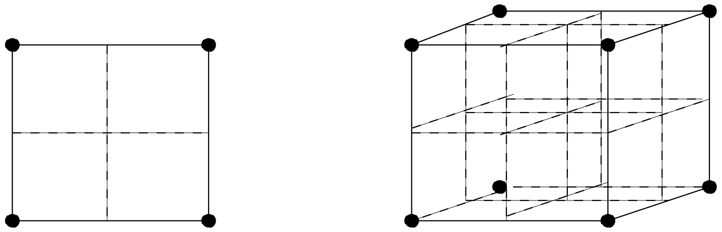


Figure 4.3: Left: For the 2-D parametrisation on a rectangle the bilinear representation needs four concentration values on the nodes (•) as parameters. Taking the same rectangle as support for the piecewise constant basis, only one paramater occurs. With four parameters, i.e. four finite elements (− −), gradients within the rectangle can be described by the piecewise constant basis, too. Right: The same for three dimensions, where the representation by trilinear basis functions on a cube requires eight parameters.

eight. Of course, the bi-or trilinear representation, in principle, allows a far better approximation of smooth fields – but only if the parameters can be reconstructed correctly. The remainder of this thesis will almost exclusively deal with 2-D tomographic reconstruction and we will systematically compare the piecewise constant and the bilinear basis functions, having in mind the points just mentioned, to tackle the question whether more involved discretisation schemes are worth considering.

Using the discretisation by local basis functions and combining into vectors like[3]

$$\begin{aligned} \mathbf{d} &= (d_1, \dots, d_m)^T \in \mathbb{R}^m, \\ \mathbf{x} &= (x_1, \dots, x_n)^T \in \mathbb{R}^n, \\ \boldsymbol{\phi} &= (\phi_1, \dots, \phi_n)^T \in \mathbb{R}^n \end{aligned} \tag{4.15}$$

the discrete problem is now stated explicitly in two (or three) dimensions.

DISCRETISED 2-D(3-D) LP-DOAS TOMOGRAPHIC RECONSTRUCTION PROBLEM
Given a vector of column densities $\mathbf{d} \in \mathbb{R}^m$ of LP-DOAS measurements along m light paths LP_i with time dependency as in the continuous problem eq. (4.4), a compact set $\Omega \subset \mathbb{R}^2\,(\mathbb{R}^3)$ containing the LP_i and a set of n basis functions $\{\phi_j(\mathbf{r})\}$ with supports $\Omega_j \subset \mathbb{R}^2\,(\mathbb{R}^3)$ so that $\cup_j \Omega_j = \Omega$, then the reconstructed concentration field represented by these basis functions is *defined* as

$$\hat{c}(\mathbf{r}) = [\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x}, \tag{4.16a}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a solution of

$$\mathbf{d} = A\,\mathbf{x} \tag{4.16b}$$

and $A \in \mathbb{R}^{m \times n}$ is given by

$$A_{ij} = \int\limits_{LP_i} ds\, \phi_j(\mathbf{r}).$$

$\mathbf{x}$ and $\boldsymbol{\phi}$ are defined in eq.(4.15).

(*i*) *Existence and uniqueness of the solution*: In contrast to the conditions for the solution of the

---

[3] The – in this case discrete – spaces of all vectors **d**, **x** are usually referred to as *data* and *model space*, respectively. The vector **x** is called *state vector*.

original reconstruction problem eq. (4.4), existence and uniqueness for the discretised problem are now clearly defined by the linear system of equations (4.16b):

$$\begin{aligned} \text{solution exists} \quad &\Leftrightarrow \quad \operatorname{rank}[A] = \operatorname{rank}[(A\ \mathbf{d})] \\ \textit{unique}\ \text{solution exists} \quad &\Leftrightarrow \quad n = \operatorname{rank}[(A\ \mathbf{d})] \end{aligned}$$

where $(A\ \mathbf{d})$ denotes the augmented matrix formed by adding $\mathbf{d}$ as a column to A. In practice, classifying the system as

even-determined	$m = n$	no solution or infinitely many or exactly one
over-determined	$m > n$	no solution (or exactly one)
under-determined	$m < n$	infinitely many solutions (or none)

is more helpful. While a solution of the discretised problem will in general not be the same as a solution of the original reconstruction problem eq. (4.4) due to the discretisation error related with the finite representation (c.f. eq. (4.10)), the column densities can be reproduced *exactly*, if only eq. (4.16b) is satisfied .

(*ii*) *Interpretation of A*: A is sometimes called *weighting* (function) matrix or *kernel* (function) matrix. Its element A_{ij} is the contribution of basis function with index j to the line integral along LP_i and in the case of the piecewise constant basis functions this is just the length of LP_i within Ω_j. For the construction of the 2-D(3-D) basis functions used here, Ω_j is a rectangle (cube), in the following called *box* in both cases. The expressions *pixel* (picture element) and *voxel* (volume element) adopted from image processing are also commonly used for the 2-D and 3-D cells. The piecewise constant basis functions will also be referred to as *box basis functions*.

4.4. The question of ill-posedness

In the preceding section it was not assumed that the number of free parameters x_j equals the number of light paths, i.e. $m = n$. But even if one cared to design a discretisation in a way that assures a square matrix A whose inverse exists, the solution $A^{-1}\mathbf{d}$ might be useless because it is numerically unstable. This means that very small changes in the data (measurement errors or even numerical inaccuracies) have large effects on the reconstruction result. Before tracing its origin this instability is illustrated by an

Example from [*Twomey*, 1997]. Let

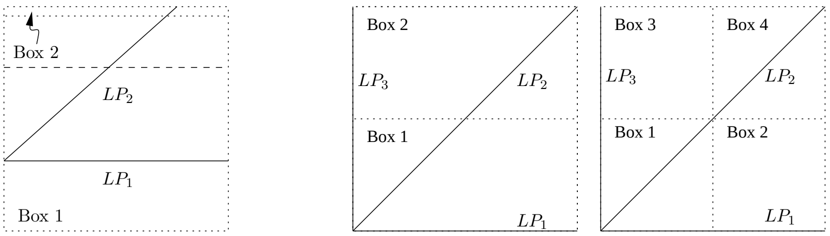
$$\begin{aligned} d_1 &= x_1 \\ d_2 &= A_{21}\, x_1 + A_{22}\, x_2 \end{aligned} \qquad \text{with} \qquad \begin{aligned} &A_{21} = 0.99995,\ A_{22} = 0.01 \\ &d_1 = 1,\ d_2 = 1.00995, \end{aligned} \tag{4.17}$$

where all units are suppressed. The exact solution is $x_1 = 1$, $x_2 = 1$. Now assume an error on the data around one percent:

$$\begin{aligned} d_1 &\mapsto d_1 - 0.01 \\ d_2 &\mapsto d_2 + 0.01. \end{aligned}$$

Then the resulting change of x_2 amounts to 1.99995 or almost 200 %.

The instability in this example can be understood by the almost linear dependency of the row vectors in the above system $(1,\ 0)$ and $(0.99995,\ 0.01)$. Applying Cramer's rule to solve the system a determinant



(a) *(b)*

Figure 4.4: (a) Two light paths and boxes (···) as in the example on p. 58 (but with different proportions to make the figure clearer). Dividing into two boxes as indicated by the dashed line (--) makes the problem perfectly well behaved. (b) Finer discretisation leads to under-determined systems of equations but with linear independent rows.

appears in the denominator which gets small for almost linear dependent vectors (In two dimensions this determinant is the $(\text{area})^2$ of the parallelogram the two vectors span, in general basically the $(\text{volume})^2$ of the corresponding parallelopiped). Small perturbations get thus amplified by a large factor. A nice geometrical explanation of this behaviour can be found in [*Twomey*, 1997, section 5.3]. An immediate question is whether this example with its near-singular system matrix A is at all relevant for any inverse problem. We first show that it is and afterwards turn to the case of tomographic inverse problems.

Assume for the moment that the system of equations was derived from discretising (by whatever method described above) a problem like the Fredholm equation (4.5) with a smooth kernel function k. For example, processes involving radiation attenuation in the atmosphere, i.e. processes that can be described by the radiative transfer equation, would lead to an exponential kernel – something very smooth. Unless the discretisation is very coarse, the row vectors of the system matrix A will show a high degree of interdependency. In fact, the paradoxical situation arises that the finer the discretisation, in other words the more accurate you try to be, the more unstable your solution gets. Of course this is only a very superficial, yet plausible, argument and one could try to work around the problem of instability by a more clever discretisation or choice of basis functions or by orthogonalising the kernel functions etc. (see [*Twomey*, 1997] for a more elaborate, but still not mathematically rigorous discussion). Nevertheless some inverse problems remain inherently unstable. These are called *ill-posed*, a term going back to *Hadamard* [1902], who defined a problem as *well-posed* if 1.) A solution exists, 2.) The solution is unique and 3.) The solution depends continuously on the data. There are more and less clever ways to approach an ill-posed problem but in the end usually a reformulation – also known as regularisation (c.f. section 4.6.3) – of the problem is necessary. Here, we do not deal with general criteria for well- or ill-posedness of continuous and the related discrete inverse problems (As one might expect they are closely related to the mathematical properties of the kernel: compactness, its smoothing behaviour etc. [*Groetsch*, 1993; *Hansen*, 1998]). Instead, we focus on the tomographic inverse problem. Take again the numerical example from above, this time understood as a tomographic measurement (fig. 4.4a). The almost linear dependent rows in the matrix A are due to a very inappropriate discretisation:

$$A = \begin{pmatrix} \text{length } LP_1 \text{ in box 1} & \text{length } LP_1 \text{ in box 2} \\ \text{length } LP_2 \text{ in box 1} & \text{length } LP_2 \text{ in box 2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.99995. & 0.01 \end{pmatrix}$$

Dividing the area differently into two boxes yields a well behaved system. Choosing only one box leads to an over-detemined system with exactly linear dependent rows. This is further illustrated in fig. 4.4b. If the square area has a length of 2 in arbitrary units, one gets

$$A_{2\text{ boxes}} = \begin{pmatrix} 2 & 0 \\ 2\sqrt{2} & 2\sqrt{2} \\ 1 & 1 \end{pmatrix}, \quad A_{4\text{ boxes}} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 2\sqrt{2} & 0 & 0 & 2\sqrt{2} \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Coarse discretisation leads to over-determined systems with unstable solutions as the rows of A become linear dependent. Finer discretisation results in under-determined systems of equations which are ill-posed in the sense that the solution is not unique, but unless light paths and discretisation do not suit each other and the rows in A become linear dependent there is no reason why the resulting system *has* to be unstable. What if both the area coverage by the light paths and the discretisation become very fine ? As suggested by the continuous approach in section 4.2, the kernel now turns into a smooth function and the argument of interdependency becomes effective again. Indeed, it turns out that tomographic measurements which provide such a high ray density lead to ill-posed inverse problems. To quote *Natterer* [2001, p. 85]: "*All problems in Computerised Tomography are ill-posed, even if to a very different degree*". Assuming reasonable smoothness of the unknown function and further properties of the operator (or matrix) norm of A, he shows that compared to other ill-posed inverse problems tomographic problems are only mildly to modestly ill-posed (The exact definition of these categories is beyond the scope of this work).

To sum up, the tomographic reconstruction problem is in principle ill-posed. Parametrisations for measurements with an abundance of data that make use of high spatial resolution will lead to unstable solutions. For measurement with only few data, like the LP-DOAS tomography, ill-posedness in the sense of instability is not given *per se*. Instability crucially depends on the details of the discretisation and is less likely to occur for under-determined systems. The next section introduces a tool that allows to quantify the degree of instability of a discrete inverse problem.

4.5. The singular value decomposition

The Singular Value Decomposition (SVD) is a generalisation of the ordinary expansion of a square symmetric matrix in terms of orthogonal eigenvectors. In fact, one way to motivate it is to look at the eigensystems of the square symmetric matrices AA^T and A^TA. We do not derive it here (see [*Golub and van Loan*, 1996] for a standard reference) and just mention its importance both theoretically and practically for numerous areas such as signal and image processing, time series analysis, pattern recognition etc. The usefulness of the SVD for the discrete tomographic problem will become evident from frequent applications in the remainder of this thesis. It states as follows.

THE SINGULAR VALUE DECOMPOSITION
Let $A \in \mathbb{R}^{m\times n}$ be a matrix of rank $r = \text{rank}[A]$. Then there exist orthonormal matrices U, V, $UU^T = U^TU = \mathbb{1}_m$ and $VV^T = V^TV = \mathbb{1}_n$, such that

$$A = U\Sigma V^T, \quad \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}, \tag{4.18}$$

where $\Sigma_r = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. The σ_i are called *singular values* of A.

Without proof, we list some properties of the SVD for later use.

(i) *Singular vectors*: Writing $U = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)$, $V = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ the orthonormal column vectors $\mathbf{u}_i$ and $\mathbf{v}_j$ are associated with σ_i, $i = 1, \dots, r$, via

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \ . \tag{4.19}$$

The SVD can be written as

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \ . \tag{4.20}$$

(ii) *Related eigenvalue decomposition*: In general, the singular values for square A do not agree with the (ordinary) eigenvalues. This is the case only for symmetric A. However, the singular values are eigenvalues of the following eigenvalue problems

$$AA^T\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i \ , \tag{4.21a}$$

$$A^T A\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i \ . \tag{4.21b}$$

(iii) *Uniqueness*: The σ_i are unique. Two singular vectors $\mathbf{u}_i$, $\mathbf{v}_i$ are unique up to a common sign, except for vectors associated with multiple singular values. Here only the spaces spanned by these vectors are unique. Thus U and V are unique up to linear combination of the corresponding column vectors.

(iv) *Fundamental subspaces*: All column vectors of U and V form a basis of the data and model space, respectively. Defining the *range* (or column space) of A as

$$\mathcal{R}(A) = \{\mathbf{y} = A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \tag{4.22}$$

and the *nullspace* (or *kernel*) as

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}, \tag{4.23}$$

the column vectors of V associated with the zero singular values are a basis of $\mathcal{N}(A)$. More completely, for the four fundamental subspaces of A :

$$\begin{aligned} \mathcal{N}(A) &= \mathrm{span}[\mathbf{v}_{r+1}, \dots, \mathbf{v}_n], \quad \mathcal{R}(A) = \mathrm{span}[\mathbf{u}_1, \dots, \mathbf{u}_r], \\ \mathcal{N}(A^T) &= \mathrm{span}[\mathbf{v}_1, \dots, \mathbf{v}_r], \quad \mathcal{R}(A^T) = \mathrm{span}[\mathbf{u}_{r+1}, \dots, \mathbf{u}_m]. \end{aligned} \tag{4.24}$$

The importance of these seemingly abstract spaces lies in the fact that components of atmospheric states that belong to the nullspace of the matrix A describing the measurement go unnoticed by the experiment. Similarly, components of the data lying outside the range of A cannot be reproduced exactly by any retrieval $\mathbf{x}$.

(v) *Characteristic features*: The following two features are very common for discrete ill-posed problems [*Hansen*, 1998] and will also appear later in the tomographic applications of this work.

- The singular values decay gradually to zero without particular gaps in between. Increasing the dimensions of A will increase the number of small singular values.

- The singular vectors have more sign changes in their components with increasing index i, i.e. decreasing σ_i. If AA^T and A^TA are totally positive, then $\mathbf{u}_i$ and $\mathbf{v}_i$ have exactly $i-1$ sign changes.

Taking eq. (4.20) and (v) together allows the interpretation that high frequency components of the state get damped by small singular values so that they add little information to the measurement. On the other hand, the inversion will be dominated by these oscillating components (in the same way the solution of a square system is unstable because of a small system determinant). Small singular values lie at the heart of the ill-posed problem making it effectively rank-deficient. Obviously the absolute size of the σ_i is irrelevant – otherwise rescaling of A would solve the problem of ill-posedness. Instead one expects some relative expression for the instability. It can be derived by considering the perturbed system

$$A(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{d}+\boldsymbol{\epsilon} \tag{4.25}$$

and assuming for the moment that A^{-1} exists. Then (choosing the induced norm of the vector norm) from $\|\Delta\mathbf{x}\| \leq \|A^{-1}\|\|\boldsymbol{\epsilon}\|$ and $\|\mathbf{d}\| \leq \|A\|\|\mathbf{x}\|$ it follows for the relative error of the solution

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\|\;\|A\|\;\frac{\|\boldsymbol{\epsilon}\|}{\|\mathbf{d}\|}. \tag{4.26}$$

The factor $\|A^{-1}\|\;\|A\|$ measures the relative sensitivity of the solution against perturbations and is called *condition number*. Using the 2-norm one gets

$$\mathrm{cond}[A] = \|A^{-1}\|_2\;\|A\|_2 = \frac{\sigma_1}{\sigma_r} \tag{4.27}$$

where the last expression defines the condition number for any matrix A. The matrix in section 4.4 has singular values $\sigma_1 = 1.4142$, $\sigma_2 = 0.0071$, thus a condition number of about 200, meaning that simple inversion magnifies the relative error on the data at worst by a factor 200. Typical condition numbers for applications in CT are around 10^6!

4.6. The least squares problem

As pointed out in the preceding section, the discrete tomographic problem is ill-posed in the literal sense: The *formulation* in the form $A\mathbf{x} = \mathbf{d}$ causes the trouble, problems with the solution are only secondary. *Reformulation* depends on the origin of the ill-posedness – instability or lack of information – and the physical nature of the underlying problem in terms of hard, empirical or statistical constraints for admissible solutions. The solution method, i.e. the algorithm, depends on the mathematical and numerical properties of the problem (linear/nonlinear, small/large systems of equations, sparse/dense systems etc).[4] The least squares approach is a mathematically well covered formalism which allows simple formulation and solution of basic ill-posed problems irrespective of the cause of ill-posedness.

An extensive reference for a variety of least squares problems with focus on their numerical treatment is [*Björck*, 1996]. Least square criteria in a wider conceptional context of inverse theory can be found

---

[4] And the working field. It seems that in geophysics, atmospheric profiling and image reconstruction people traditionally employ different algorithms for the same problems, sometimes mathematically very similar ones, yet with different names. The mathematical literature offers a more comprehensive treatment.

in [*Tarantola*, 1987, 2005]. We start our discussion with the most simple cases.

4.6.1. Least squares and least norm solutions

There are certainly many ways to motivate the use of least squares methods, like the top-down derivation from statistical properties of the model, here we take a rather pragmatic point of view and distinguish two cases:

(i) *Least squares minimisation* ($m > n$):
Having more equations than unknowns cannot lead to unique solutions if measurement errors are present. Instead, minimum discrepancy between data and forward calculated data in terms of a vector norm is demanded. Choosing the 2-norm (see the discussion in section 4.6.5) gives the following reformulation of the discrete inverse problem. For given $\mathbf{d} \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$, find $\mathbf{x} \in \mathbb{R}^n$ such that

$$\min_{\mathbf{x}} \|\mathbf{d} - A\mathbf{x}\|_2 \quad \text{or} \quad \min_{\mathbf{x}} \left(\mathbf{d} - A\mathbf{x}\right)^T \left(\mathbf{d} - A\mathbf{x}\right). \tag{4.28}$$

Derivation leads to the normal equations

$$A^T A\mathbf{x} = A^T \mathbf{d} \tag{4.29}$$

and the solution

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T \mathbf{d}, \tag{4.30}$$

provided the inverse of the $m \times m$ matrix $A^T A$ exists. This is the case if and only if $\text{rank}[A] = n$ which can easily be shown using the SVD.

(ii) *Least norm minimisation* ($m < n$):
In the case where there are more parameters to determine than measurements available, conditions have to be added to the measurement equations in a way that gives rise to a unique solution. Picking the vector of smallest norm from all admissible solutions makes sense from an economic point of view (unless you expect the solution to be very fluctuating). Also this least norm solution has attractive mathematical properties (see below). But nevertheless it remains arbitrary and if it can be replaced by physical constraints or statistical information, all the better. The mathematical implementation is such that the equations $A\mathbf{x} = \mathbf{d}$ are regarded as constraints for the norm minimisation problem, i.e.

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2\,, \quad \mathbf{d} = A\mathbf{x}. \tag{4.31}$$

With a Lagrange multiplier λ_j for each of the n equations and writing $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$, one gets the normal equations of the second kind

$$AA^T \boldsymbol{\lambda}/2 = \mathbf{d}, \quad \mathbf{x} = A^T \boldsymbol{\lambda}/2 \tag{4.32}$$

with unique solution

$$\mathbf{x} = A^T \left(AA^T\right)^{-1} \mathbf{d} \tag{4.33}$$

if $\text{rank}[A] = m$. One might be familiar with the characteristics of the least square method as it is *the* tool for any kind of regression, but the behaviour of the least norm solution may not be equally clear.

Therefore, we would like to present a simple

Example based on fig. 4.4b, p. 59.
Assume a constant concentration in all four square boxes of length 1 (again we suppress all units for the moment):

$$c = 1.$$

Neglecting any measurement errors yields the column densities

$$d = (2, 2\sqrt{2}, 2)^T$$

and together with

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \sqrt{2} & 0 & 0 & \sqrt{2} \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

this gives the least norm solution

$$\mathbf{x}_{LN} = (1.5, 0.5, 0.5, 0.5)^T,$$

i.e. overestimation in the box with three light paths, underestimation in the others. Why not a uniform solution $(1, 1, 1, 1)^T$? Because $\mathbf{x}_{LN}$ is compatible with the measurements *and* has a smaller norm:

$$\|\mathbf{x}_{LN}\|_2 = \sqrt{3} < 2 = \|(1, 1, 1, 1)^T\|_2.$$

In the following case, a concentration peak is located in one of the boxes which are crossed by one light beam, here the upper left one in fig. 4.4b:

$$c = (1, 1, 10, 1)^T \Rightarrow d = (2, 2\sqrt{2}, 11)^T \quad \text{and} \quad \mathbf{x}_{LN} = (3.75, -1.75, 7.25, -1.75)^T,$$

i.e. underestimation of the peak. In this case negative concentrations compatible with the data lead to smallest norm.

We finally address the cases of rank deficiency, i.e. rank$[A] < \min(m, n)$. To prepare the formal solution, notice that the inverse of a square matrix A with rank$[A] = m = n$ can easily be expressed in terms of the SVD eq. (4.18)

$$A^{-1} = V\Sigma_r^{-1}U^T. \tag{4.34}$$

A similar expression holds for the full-rank over-determined case eq. (4.30)

$$\left(A^T A\right)^{-1} A^T = V \begin{pmatrix} \Sigma_r^{-1} \\ 0 \end{pmatrix} U^T, \qquad \text{rank}[A] = m \tag{4.35}$$

and the full-rank under-determined case eq. (4.33)

$$A^T \left(AA^T\right)^{-1} = V \begin{pmatrix} \Sigma_r^{-1} & 0 \end{pmatrix} U^T, \qquad \text{rank}[A] = n. \tag{4.36}$$

It is tempting to consider the expression $V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T$ as a generalised inverse in the rank deficient case. Indeed, the following theorem holds:

LEAST SQUARES-MINIMUM NORM SOLUTION
Let $A \in \mathbb{R}^{m \times n}$ be of rank $r \leq \min(m, n)$ and the SVD of A be given by eq. (4.18) then the general

least squares problem

$$\min_{\mathbf{x}\in S}\|x\|_2, \quad S=\{\mathbf{x}\in\mathbb{R}^n \mid \|A\mathbf{x}-\mathbf{d}\|_2=\min\} \tag{4.37a}$$

has the unique solution

$$\mathbf{x}=A^{\dagger}\mathbf{d}, \tag{4.37b}$$

where the matrix

$$A^{\dagger}=V\begin{pmatrix}\Sigma_r^{-1} & 0\\ 0 & 0\end{pmatrix}U^T \tag{4.37c}$$

is called the *pseudoinverse* of A.

See [*Björck*, 1996] for a proof. Depending on the dimensions m, n the pseudoinverse (also called *generalised inverse* or *Moore-Penrose inverse*) automatically generates a least squares, least norm or the exact solution if A is full rank. Otherwise the result is a mixture referred to as *least squares-minimum norm solution*.

4.6.2. Weighted least squares and weighted least norm with a priori

The least squares and least norm principles of the preceding section might need modification to incorporate the underlying physics. To make the point clearer we assume $\operatorname{rank}[A]=\min(m,n)$.

(i) *Data weighting*:
Imagine that the uncertainty differs substantially within the measured data d_i (something not too hard to imagine). This means, for example, that a large data value which is large because of a large measurement error enters the least squares minimisation problem eq. (4.28) in the same way as a large value with a small measurement error. Therefore, it is common practice to weight the data according to their errors, usually by the inverse of the error variances $\sigma_{\boldsymbol{\epsilon}\,i}$ or more general, the inverse of the error covariance matrix $S_{\boldsymbol{\epsilon}}=\mathrm{E}[(\mathbf{d}_{\boldsymbol{\epsilon}}-\mathbf{d})(\mathbf{d}_{\boldsymbol{\epsilon}}-\mathbf{d})^T]$. Here $\mathrm{E}[\cdot]$ denotes the expected values and $\mathbf{d}=\mathrm{E}[\mathbf{d}_{\boldsymbol{\epsilon}}]$ according to eq. (4.2) with unbiased errors. The *weighted* least squares minimisation replaces eq. (4.28) by[5]

$$\min_{\mathbf{x}}\left(\mathbf{d}-A\mathbf{x}\right)^T S_{\boldsymbol{\epsilon}}^{-1}\left(\mathbf{d}-A\mathbf{x}\right). \tag{4.38}$$

$S_{\boldsymbol{\epsilon}}$ is positive semidefinite. If none of the $\sigma_{\boldsymbol{\epsilon}\,i}$ equals zero it is positive definite and the inverse exists. It is straightforward to obtain the solution

$$\mathbf{x}=\left(A^T S_{\boldsymbol{\epsilon}}^{-1}A\right)^{-1}A^T S_{\boldsymbol{\epsilon}}^{-1}\mathbf{d} \tag{4.39}$$

and, employing the relation $S_{\mathbf{x}}=BS_{\mathbf{y}}B^T$ for the covariances that holds for any linear relation $\mathbf{x}=B\mathbf{y}$, the covariance of $\mathbf{x}$ is

$$S_{\mathbf{x}}=\left(A^T S_{\boldsymbol{\epsilon}}^{-1}A\right)^{-1}. \tag{4.40}$$

(ii) *A priori*:
Data weighting has no effect on the least norm solution (apart from unintended numerical implications) as the equations $A\mathbf{x}=\mathbf{d}$ are fulfilled exactly. But the minimisation principle for $\|\mathbf{x}\|_2$ might not be a good choice for cases like the following. Assume that the concentration field shows local variability

---

[5] Another way to look at it is that the 2-norm gets replaced by a norm defined as $\|\cdot\|^2_{S_{\boldsymbol{\epsilon}}^{-1}}=(\cdot)S_{\boldsymbol{\epsilon}}^{-1}(\cdot)^T$.

on an otherwise constant but high background that is more or less known. Instead of demanding that the norm is small, it would make far more sense in this case to demand that the deviation from this background is minimal, i.e. $\min\|\mathbf{x}-\mathbf{x}_a\|_2$, where $\mathbf{x}_a$ stands for this known, estimated or assumed concentration. It is usually referred to as *a priori* because whatever information is used for its construction, in general, it has to be known beforehand and cannot be gained as a product or byproduct of the reconstruction process. Inserting a weighting matrix as in eq. (4.38) the minimisation principle with a priori reads now

$$\min_{\mathbf{x}} \left(\mathbf{x}-\mathbf{x}_a\right)^T S_a^{-1} \left(\mathbf{x}-\mathbf{x}_a\right) \ , \quad \mathbf{d} = A\mathbf{x}. \tag{4.41}$$

There is another drawback of the least norm principle eq. (4.31) which becomes apparent from eq. (4.36). Whenever a component of the retrieval lies in the nullspace of A it is set to zero. One could argue that naturally the principle cannot produce values that are not given by the data, but zero is as arbitrary as any other guess. Solving eq. (4.41) makes this clearer:

$$\mathbf{x} = \mathbf{x}_a + S_a^{-1} A^T \left(A S_a^{-1} A^T\right)^{-1} \left(\mathbf{d} - A\mathbf{x}_a\right). \tag{4.42}$$

Obviously, eq. (4.33) is a special case of this with $\mathbf{x}_a = \mathbf{0}$, i.e. it contains an a priori, too.

In principle, a solution similar to eq. (4.37b) could be given for the modified least squares - minimum norm problem

$$\min_{\mathbf{x}\in S} \left(\mathbf{x}-\mathbf{x}_a\right)^T S_a^{-1} \left(\mathbf{x}-\mathbf{x}_a\right), \quad S = \{\mathbf{x}\in\mathbb{R}^n \mid \left(\mathbf{d}-A\mathbf{x}\right)^T S_{\boldsymbol{\epsilon}}^{-1} \left(\mathbf{d}-A\mathbf{x}\right) = \min\}, \tag{4.43}$$

with a priori $\mathbf{x}_a \in \mathbb{R}^n$, measurement error covariance $S_{\boldsymbol{\epsilon}}$ and an a priori weighting matrix S_a, but for the remainder only the explicit form of the solutions eqs. (4.39,4.41) for the full-rank case is important.

4.6.3. Regularisation and constrained least squares problems

The least squares-minimum norm principle solves the problems of non-uniqueness and non-existence but does not make any difference for the instability of the solution because of small singular values. This is very clear for the solutions written in the form of eqs. (4.34-4.36) (see also [*Twomey*, 1997, chap. 6.2]). Although ill-posedness due to instability is not the main issue for the under-determined reconstruction problems focused on in this work, several strategies to *regularise* (as this stabilising is called) the least squares-least norm solution are briefly sketched in the following for two reasons. Firstly these methods are widely used, indispensable reference and starting points for any discussion and future development. Secondly, while some regularisation methods are purely mathematical, others aim to involve physical constraints related to the actual measurement to generate a stable and unique solution. These latter approaches are attractive alternatives to methods that contain arbitrary ad hoc parameters and are especially interesting for under-determined problems – provided that there is enough information to formulate these physical constraints.

(*i*) *Tikhonov regularisation*:
Notice that the instability of the least squares and least norm solution lies in the expressions $\left(A^T A\right)^{-1}$

and $\left(AA^T\right)^{-1}$, respectively. Making the substitution

$$\left(A^TA\right)^{-1} \longmapsto \left(A^TA + \lambda^2 \mathbb{1}_n\right)^{-1} \tag{4.44a}$$

$$\text{or} \quad \left(AA^T\right)^{-1} \longmapsto \left(AA^T + \lambda^2 \mathbb{1}_m\right)^{-1}, \tag{4.44b}$$

i.e. adding a term proportional to the unit matrix, changes the singular values in the decompositions eqs. (4.35,4.36) to

$$\sigma_i^{-1} \longmapsto \frac{\sigma_i}{\sigma_i^2 + \lambda^2}. \tag{4.45}$$

For the right choice of λ small singular values are suppressed as desired. Following Tikhonov a choice $\alpha = \alpha(\lambda)$ is said to result in a *regular* solution $\mathbf{x}_\lambda$ of $A\mathbf{x} = \mathbf{d}$ if

$$\lambda \to 0 \Rightarrow \alpha(\lambda) \to 0, \quad \mathbf{x}_\lambda \to A^\dagger \mathbf{d}$$

with $A^\dagger$ as in eqs. (4.37). There is a minimisation problem that leads to the same result as the substitutions eqs. (4.44):

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 \tag{4.46}$$

(Mind that λ is *not* a Lagrange-multiplier here). The reasoning for this ansatz in the over-determined case is that minimisation of the discrepancy $\|A\mathbf{x} - \mathbf{d}\|_2$ alone leads to oscillating solutions while the addition of $\|\mathbf{x}\|_2$ favours smooth solutions. Allowing additionally weighting and an a priori as in section 4.6.2 one has the more general *regularised* (also called *damped* or *generalised*[6]) least squares principle

$$\min_{\mathbf{x}} \ (A\mathbf{x} - \mathbf{d})^T S_{\boldsymbol{\epsilon}}^{-1} (A\mathbf{x} - \mathbf{d}) \ + \ \lambda^2 \ (\mathbf{x} - \mathbf{x}_a)^T D^T D \, (\mathbf{x} - \mathbf{x}_a) \tag{4.47}$$

with the straightforward solution

$$\mathbf{x} = \mathbf{x}_a + \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + \lambda^2 D^T D\right)^{-1} A^T S_{\boldsymbol{\epsilon}}^{-1} (\mathbf{d} - A\mathbf{x}_a) \tag{4.48a}$$

$$= \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + \lambda^2 D^T D\right)^{-1} \left(A^T S_{\boldsymbol{\epsilon}}^{-1} \mathbf{d} + \lambda^2 D^T D \, \mathbf{x}_a\right) \tag{4.48b}$$

$$= \mathbf{x}_a + \left(D^T D\right)^{-1} A^T \left(A (D^T D)^{-1} A^T + \lambda^2 S_{\boldsymbol{\epsilon}}\right)^{-1} (\mathbf{d} - A\mathbf{x}_a) \, . \tag{4.48c}$$

Here $D \in \mathbb{R}^{n \times n}$ can be the identity matrix, a diagonal weighting matrix or a discrete approximation of a derivative operator to influence the smoothness of $\mathbf{x}$ etc. In the last step it was assumed that $D^T D$ is regular. Following from the normal equations, the solution is unique if $\mathcal{N}(S_{\boldsymbol{\epsilon}}^{-1/2} A) \cap \mathcal{N}(D) = \{0\}$. Calculation of the covariance $S_{\mathbf{x}}$ can be carried out as in eq. (4.40). The generalised least squares solution formally contains the least squares and least norm solution in the limits

$$\text{generalised least squares} \quad \begin{array}{l} \xrightarrow{\lambda \to 0} \quad \text{least squares} \\ \xrightarrow{\lambda^2 S_{\boldsymbol{\epsilon}} \to 0} \quad \text{least norm} \end{array} . \tag{4.49}$$

This method is generally attributed to *Tikhonov* [1963] but sometimes also referred to as Twomey-Tikhonov approach [*Twomey*, 1963]. While there exist estimations for the effect of the regularisation, i.e. bounds on how much the solution changes compared to the unregularised (see sec. 5.2.3), there

---

[6] The term *constrained* least squares used by *Twomey* [1997] is not common in this context. The nomenclature is far from uniform.

is nothing telling what exact value the regularisation parameter λ ought to take. If λ is chosen too small there is no regularisation, if it is too large the solution is oversmoothed. There is a variety of suggestions for the parameter choice, some of them are very intuitive like the *discrepancy principle* that chooses λ such that the perturbation bounds are smaller than the residual norm $\|A\mathbf{x} - \mathbf{d}\|$. This is not pursued any further, instead we refer to [*Hansen*, 1998, chap. 7] and quote *Linz* [1984]:

> "*There are no general criteria by which different algorithms can be compared. Consequently, many methods are proposed which, on the evidence of a few special cases, are claimed to be effective*".

(*ii*) *Constrained least-squares*:
The following least squares problems with quadratic constraints are dual to each other in that – for the right choice of parameters $\delta, \epsilon \in \mathbb{R}$ – their solutions are identical to the Tikhonov-regularised solution

$$\min_{\mathbf{x}\in S}\|A\mathbf{x} - \mathbf{d}\|_2, \quad S = \{\mathbf{x} \in \mathbb{R}^n \mid \|D(\mathbf{x} - \mathbf{x}_a)\|_2 \leq \delta\} \tag{4.50a}$$

$$\min_{\mathbf{x}\in S}\|D(\mathbf{x} - \mathbf{x}_a)\|_2, \quad S = \{\mathbf{x} \in \mathbb{R}^n \mid \|A\mathbf{x} - \mathbf{d}\|_2 \leq \epsilon\}. \tag{4.50b}$$

This time the constraints can be taken into account by Lagrange multipliers as the inequalities are almost always fulfilled as equalities for *quadratic* constraints. The advantage of this formulation lies in the fact that it allows to express the regularisation in terms of bounds on physical quantities like gradients in the case of eq. (4.50a) or measurement errors for eq. (4.50b). For some inverse problems, like the tomographic problem, a further inequality constraint is necessary to assure nonnegativity of the solution:

$$\min_{\mathbf{x}\in S}\|D(\mathbf{x} - \mathbf{x}_a)\|_2, \quad S = \{\mathbf{x} \in \mathbb{R}^n \mid \|A\mathbf{x} - \mathbf{d}\|_2 \leq \epsilon \wedge \mathbf{x} \geq \mathbf{0}\}. \tag{4.51}$$

While there are standard methods to solve least squares problems with linear inequality constraints or quadratic constraints like eqs. (4.50a,4.50b) [e.g., *Björck*, 1996, chaps.5.2,3], solving the optimisation problem eq. (4.51) is rather more involved [e.g., *Fehmers*, 1998].

(*iii*) *Truncated singular value decomposition and iterative regularisation*:
Both methods rely on the SVD. The truncated SVD exploits the fact that the generalised inverse eq. (4.37c) has an expansion similar to A eq. (4.20):

$$A^{\dagger} = \sum_{i=1}^{r} \sigma_i^{-1}\mathbf{u}_i\mathbf{v}_i^T \tag{4.52}$$

with $r = \text{rank}[A]$ and increasing contributions as i increases. So cutting off the sum at some $i_c < r$ will stabilise the inverse. Basically it amounts to replacing the ill-conditioned, full-rank matrix A by a well-conditioned, rank-deficient matrix. Because $\mathbf{u}_i$, $\mathbf{v}_j$ are orthonormal, the solution will also have a smaller norm than the unregularised. Again there is no canonical choice for the cut-off parameter. It could be given by an upper bound f_ϵ for the relative error magnification eq. (4.27), i.e. $\sigma_1/\sigma_{i_c} < f_\epsilon$ or $\sigma_{i_c} > \sigma_1/f_\epsilon$.

Regularisation by means of iterative algorithms that are – for reasons to be discussed later – used to generate least squares solutions works very similar to the cut-off of small singular values. If the algorithm initially picks up components in the SVD that belong to large singular values or converges faster for these, then stopping the iteration prematurely will suppress small singular values. The role of the regularisation parameter is played by the iteration number. Relevant details will be discussed in chap. 4.8.

(*iv*) *Parametrisation of the continuous problem*:
There are other factors that influence the stability of the problem although it might be hard to formulate them according to the mathematical definition of regularisation given by Tikhonov. For example, in section 4.4 it was argued that the number of basis functions influences the conditioning of the system matrix. The same holds for the kind of basis functions. We would like to mention a thorough study by *Doicu et al.* [2004] who theoretically investigated the reconstruction of 1-D temperature and concentration profiles from far infrared airborne observations and infrared limb emission spectra, respectively. Apart from using a regularised algorithm, the authors show that different basis functions (Chebyshev, Hermite polynomials and B-splines in this case) clearly have different regularising properties. Another approach totally ignored here is to regularise *before* discretising. In fact, this approach is related to the choice of basis functions [see *Groetsch*, 1993, section 5.2]. And one could still think of other reconstruction schemes involving a finite number of paramaters that circumvent the problem of instability and lead to possibly not ideal, yet stable solutions.

4.6.4. Resolution matrix and averaging kernels

Assume that the retrieval is given by the linear map A^{inv}, so that $\mathbf{x} = A^{inv}\mathbf{d}$. If all errors in the discrete forward model are neglected, the column densities can be calculated from the true state of the system $\mathbf{x}_{true}$ as $\mathbf{d} = A\mathbf{x}_{true}$. Inserting this into the previous equation yields

$$\mathbf{x} = A^{inv}A\,\mathbf{x}_{true} = R\,\mathbf{x}_{true}. \tag{4.53}$$

R is called *resolution matrix* as it quantifies how well a state is resolved by the measurement plus retrieval process. Its column vectors describe to what degree the components x_{true_i} are averaged over, i.e. for the tomographic reconstruction: how the original box concentrations are spread over the retrieval grid and finally give rise to the concentration x_i in box i. The rows of R are called *averaging kernels*. R takes a simple form for the unweighted least norm solution (i.e. $m < n$) eq. (4.33):

$$R = A^T(AA^T)^{-1}A = \sum_{j=1}^{r=\text{rank}[A]} \mathbf{v}_j\mathbf{v}_j^T. \tag{4.54}$$

This is just the projection matrix onto the orthogonal complement of the nullspace of A, that is the subspace where the retrieval is completely given by the measurement data. The matrix

$$P = \mathbb{1}_n - R = \mathbb{1}_n - A^T(AA^T)^{-1}A \tag{4.55}$$

projects onto the nullspace of A.

In the same way, inserting $\mathbf{d} = A\mathbf{x}_{true}$ into the generalised least squares solution eq. (4.48a) leads to

$$\mathbf{x} = \mathbf{x}_a + \underbrace{\left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + \lambda^2 D^T D\right)^{-1} A^T S_{\boldsymbol{\epsilon}}^{-1} A}_{R} (\mathbf{x} - \mathbf{x}_a) \tag{4.56a}$$

$$= R\,\mathbf{x}_{true} + (\mathbb{1}_n - R)\,\mathbf{x}_a \tag{4.56b}$$

$$= A^{inv}\mathbf{d} + (\mathbb{1}_n - R)\,\mathbf{x}_a \tag{4.56c}$$

with similar expressions for eq. (4.48c). This formulation shows nicely that the information from the measurement is mapped to the retrieval $\mathbf{x}$ by R while anything else $(\mathbb{1}_n - R)$ has to come from the a priori.

For the over-determined case $(m > n)$ the projector in eq. (4.54) becomes $R = \mathbb{1}_n$ if $r = \mathrm{rank}[A] \geq n$. In this case definition of a corresponding data resolution matrix would be appropriate. But as we do not make use of it, we leave the matter to it.

4.6.5. Remark on the norm

Finally, a justification for our choice of norm is mandatory, even if only very briefly. In principle, any norm could be used in the above minimisation problems, e.g. any member of the l_p norm family defined by

$$\|\mathbf{x}\|_{l_p} = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p}, \quad p \geq 1.$$

But the choice of a norm very different from the standard ones requires firm knowledge of the analytic and algebraic implications. For example, on the one hand the problem

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{d}\|_{l_p}, \quad \mathrm{rank}[A] = n$$

has a unique solution for $p > 1$ only, while on the other hand solutions with norms near $p = 1$ have more stable numerical properties [*Scales and Gersztenkorn*, 1988]. Furthermore, the norm is closely related to the statistics of the problem (see [*Tarantola*, 1987] for an extensive discussion). Here, we assume all relevant quantities to be Gaussian distributed which naturally leads to the 2-norm. In other words, the choice is dictated by the statistical nature of the measurement, not by the details of the retrieval method. There might be norms which are better suited for the inversion of certain problems or concentration fields, even if the statistics behind is Gaussian. Then all mathematical tools that are based on the 2-norm, e.g. the SVD, would have to be modified. The choice of the norm is a fundamental matter under several aspects and with no indication against the conventional 2-norm at hand, a reconsideration is beyond the scope of this work.

4.7. Statistical approach to the least squares problem

The statistical approach to the inverse problem describes all quantities involved by probability densities, i.e. $P(\mathbf{x})$ and $P(\mathbf{d})$ for $\mathbf{x}$ and $\mathbf{d}$ in our notation (see [*Tarantola and Valette*, 1982] for a concise yet general formulation of the approach). The forward model and thus ideally the measurement is expressed as conditional probability $P(\mathbf{d}|\mathbf{x})$, the reconstruction by the a posteriori conditional prob-

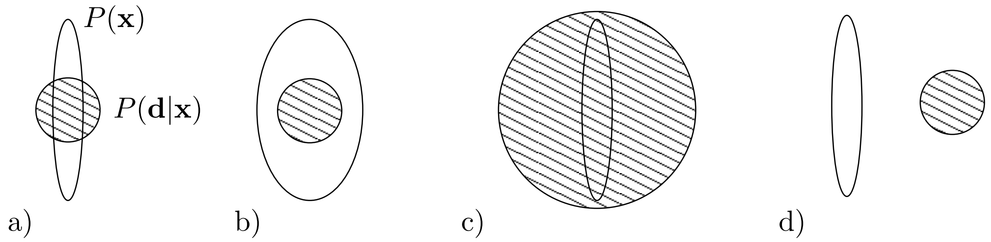


Figure 4.5: Schematic picture of how the measurement can update the a priori information about the system. (a) The measurement $P(\mathbf{d}|\mathbf{x})$ increases the knowledge about the a priori $P(\mathbf{x})$ but does not confine the a posteriori estimate completely. (b) The measurement constrains the a priori completely. (c) The measurement does not provide any new information and the retrieval distribution is basically given by the a priori estimate. (d) A priori and measurement are inconsistent in some user-defined sense.

ability $P(\mathbf{x}|\mathbf{d})$. As both are related via Bayes' theorem

$$P(\mathbf{x}|\mathbf{d})P(\mathbf{d}) = P(\mathbf{d}|\mathbf{x})P(\mathbf{x}), \tag{4.57}$$

the statistical approach is also referred to as *Bayesian* approach. To infer the a posteriori conditional probability $P(\mathbf{x}|\mathbf{d})$ from eq. (4.57) both $P(\mathbf{d}|\mathbf{x})$ and $P(\mathbf{x})$ have to be known. $P(\mathbf{d}|\mathbf{x})$ is given by the forward model, the measurement data and its statistics. The a priori probability density is a more intricate matter as it asks for (statistical) knowledge about the system before the measurement. If the Bayesian approach is intended to increase the information on the actual state of the system after the measurement, the a priori has to be based on either theoretical insight or empirical ground that is firm enough to allow for a statistical description of it. This could be certain knowledge like "concentrations are always positive" or a climatology, for example confining the temperature profile of the atmosphere from past measurements etc. Whether the a priori is useful for the reconstruction is a different matter, as sketched in the cartoon in fig. 4.5. Another way to use the Bayesian method is the following: An *assumption* about the state model is formulated as a probabilistic statement which serves as a priori probability for the Bayesian inversion. Here the point is to constrain the a priori model assumption more tightly by the a posteriori estimate (see also fig. 4.5). The measurement data can also help to construct the a priori but then the uncertainties have to be corrected. This method is known as *empirical* Bayes approach [e.g., *Carlin and Louis*, 1998]. As we will see, yet another way is to (mis)use the statistical inversion as a regularisation method.

An attractive feature of the statistical inversion is that by its nature it provides an uncertainty for the retrieval, i.e something like an error estimate. The Bayesian method is widely used for any of the questions just mentioned in geophysics [e.g., *Scales and Snieder*, 1997; *Tarantola*, 2005] and atmospheric research, here mostly in the very popular form of the *optimal estimate* for Gaussian probability densities [*Rodgers*, 2000].

4.7.1. The optimal estimate

To get a specific retrieval one has to choose a state from the a posteriori distribution. The optimal estimate is the solution that maximises the a posteriori probability. It is therefore more correctly called *maximum a posteriori* (MAP) or less correctly *maximum likelihood* (ML) solution. Assuming a Gaussian distribution

$$P(\mathbf{d}|\mathbf{x}) \sim e^{(\mathbf{d}-A\mathbf{x})^T S_{\boldsymbol{\epsilon}}^{-1}(\mathbf{d}-A\mathbf{x})}$$

for the measurement is plausible (here $S_{\boldsymbol{\epsilon}}$ denotes again the measurement error covariance). The statistics of the a priori is less evident. If it is obtained from some kind of climatology, then a Gaussian distribution makes sense. But if, for example, it consists in an educated guess based on model results, choice of Gaussian statistics is more than questionable. Nevertheless we assume a Gaussian probability density

$$P(\mathbf{x}) \sim e^{(\mathbf{x}-\mathbf{x}_a)^T S_a^{-1}(\mathbf{x}-\mathbf{x}_a)}$$

where $\mathbf{x}_a$ is again the a priori (state vector) and S_a its covariance. The MAP solution coincides with the expected mean E$[P(\mathbf{x}|\mathbf{d})]$ and takes the following forms [*Rodgers*, 2000; *Tarantola and Valette*, 1982]:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_a + \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + S_a^{-1}\right)^{-1} A^T S_{\boldsymbol{\epsilon}}^{-1} \left(\mathbf{d} - A\mathbf{x}_a\right) && \text{(4.58a)} \\ &= \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + S_a^{-1}\right)^{-1} \left(A^T S_{\boldsymbol{\epsilon}}^{-1}\mathbf{d} + S_a^{-1}\mathbf{x}_a\right) && \text{(4.58b)} \\ &= \mathbf{x}_a + S_a A^T \left(A S_a A^T + S_{\boldsymbol{\epsilon}}\right)^{-1} \left(\mathbf{d} - A\mathbf{x}_a\right). && \text{(4.58c)} \end{aligned}$$

with covariance

$$\begin{aligned} S_{\mathbf{x}} &= \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + S_a^{-1}\right)^{-1} && \text{(4.59a)} \\ &= S_a - S_a A^T \left(A S_a A^T + S_{\boldsymbol{\epsilon}}\right)^{-1} A S_a. && \text{(4.59b)} \end{aligned}$$

The solution is formally identical with the generalised least squares solution eq. (4.48) if one sets $S_a^{-1} = \lambda^2 D^T D$. And this is exactly how the optimal estimate regularises or can be used to regularise the least squares solution, as becomes especially clear for diagonal $S_a = \sigma_a^2 \mathbb{1}_n$. σ_a^{-1} plays now the role of the regularisation parameter λ. For completely non-committal a priori $\sigma_a \to \infty$ (see fig. 4.5b) or infinitely precise measurement, the optimal estimate turns into the unregularised least squares solution. If the a priori state of the system is known exactly $\sigma_a \to 0$ (fig. 4.5c) or the data error is extremely large, the solution is just the a priori $\mathbf{x}_a$ (oversmoothing). For the standard case (fig. 4.5a) the optimal estimate can be seen as the weighted mean of what is known from the measurement and the a priori. In fact, treating the a priori as an additional measurement by building the augmented measurement vector, system matrix

$$\boldsymbol{d} = \begin{pmatrix} \mathbf{d} \\ \mathbf{x}_a \end{pmatrix} \in \mathbb{R}^{m+n}, \quad \mathcal{A} = \begin{pmatrix} A \\ \mathbb{1}_n \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)} \tag{4.60a}$$

and covariance

$$\mathcal{S}^{-1} = \begin{pmatrix} S_{\boldsymbol{\epsilon}}^{-1} & 0 \\ 0 & S_a^{-1} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)} \tag{4.60b}$$

eqs. (4.58) can be obtained from the weighted least squares problem

$$\min_x \left(\boldsymbol{d} - \mathcal{A}\mathbf{x}\right)^T \mathcal{S}^{-1} \left(\boldsymbol{d} - \mathcal{A}\mathbf{x}\right).$$

Conditions for the uniqueness of the solution are as for eqs. (4.48).

4.7.2. Degrees of freedom and information content

A further benefit of the Bayesian approach is that it allows to adopt concepts from any other context that uses a stochastic description like statistical physics, information theory etc. Two quantities are introduced here for later use without deriving them or putting them into proper context [c.f. *Rodgers*, 2000, chap. 2.4].

The optimal estimate minimises the expression $(\mathbf{x}-\mathbf{x}_a)^T S_a^{-1}(\mathbf{x}-\mathbf{x}_a)+\boldsymbol{\epsilon}^T S_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} = \mathbf{d}-A\mathbf{x}$. The first term provides information about $\mathbf{x}$, the second contribution is just noise. Their actual average values are interpreted as *degrees of freedom for the signal* d_s and *noise* d_n, respectively. For the optimal estimate one gets

$$d_s = \mathrm{tr}[R] = \sum_{i=1}^{\mathrm{rank}[\tilde{A}]} \tilde{\sigma}_i^2/(1+\tilde{\sigma}_i^2), \tag{4.61a}$$

$$m = d_s + d_n, \tag{4.61b}$$

where $\tilde{A} = S_{\boldsymbol{\epsilon}}^{-1/2} A S_a^{1/2}$ and $\tilde{\sigma}_i$ are the singular values of $\tilde{A}$. The resolution matrix R for the optimal estimate is given as in eq. (4.56), again with the substitution $\lambda^2 D^T D \rightarrow S_a^{-1}$:

$$R = \left(A^T S_{\boldsymbol{\epsilon}}^{-1} A + S_a^{-1}\right)^{-1} A^T S_{\boldsymbol{\epsilon}}^{-1} A. \tag{4.62}$$

The d_s and d_n add up to the degrees of freedom of the measurement, as expected. Identifying $d_{s_i} = \tilde{\sigma}_i^2/(1+\tilde{\sigma}_i^2)$ as the degree of freedom associated with the ith singular value, one would say that it increases the information about the system if $d_{s_i} > d_{n_i}$ or $\tilde{\sigma}_i > 1$. Otherwise its contribution to the measurement cannot be distinguished from noise. Consequently, these quantities can be used to evaluate the measurement setup and reconstruction for different levels of noise. Expressing the resolution matrix with the a posteriori covariance $S_{\mathbf{x}}$ (eq. 4.59) gives

$$R = \mathbb{1}_n - S_{\mathbf{x}} S_a^{-1} \tag{4.63a}$$

$$\text{or} \qquad n = d_s + \mathrm{tr}(S_{\mathbf{x}} S_a^{-1}), \tag{4.63b}$$

where the trace term can be interpreted as the number of parameters resolved by the a priori.

Up to now, expressions like 'increase of information' or 'knowledge' have been used only in a very loose sense. They can now be given a specific meaning in terms of entropy S. In thermodynamics S is given by the logarithm of the number of microstates compatible with a given macrostate.[7] So the difference between the entropy before and after the measurement agrees perfectly with what intuitively would be described as the effect of the measurement: Exclude states of the system that, in principle, are possible. The *information content* of the measurement is thus defined as

$$H = S\left(P(\mathbf{x})\right) - S\left(P(\mathbf{x}|\mathbf{d})\right).$$

The probability densities replace probabilities in the formal definition of the entropy and for Gaussian

---

[7] Provided that all microstates are equally probable.

density functions this yields

$$H = -\tfrac{1}{2} \ln |\mathbb{1}_n - R| = \tfrac{1}{2} \sum_{i=1}^{\mathrm{rank}[\tilde{A}]} \ln \left(1 + \tilde{\sigma}_i^2\right). \tag{4.64}$$

which agrees with a calculation from the difference $S\left(P(\mathbf{d})\right) - S\left(P(\mathbf{d}|\mathbf{x})\right)$.

d_s and H represent two different ways to calculate a single number with concrete physical meaning from the rather unhandy resolution matrix or averaging kernels.

4.8. Iterative solution of the least squares problem

After formulating a number of least squares problems and presenting the formal solution for the unconstrained cases, we turn now to the actual computation of these solutions.

4.8.1. Iterative versus direct methods

Numerical algorithms for least squares problems are based on either type of the normal equations (4.29,4.32), not on matrix inversion as in $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{d}$. (The SVD for which calculation of the inverse is trivial, may be regarded as an exception). They are either *direct* methods, i.e. they generate a solution in a finite number of steps, or *iterative*, i.e. converging towards the solution in a sequence $\{\mathbf{x}^{(k)}\}$, $k = 1, \ldots, \infty$. A direct method could be any method for solving systems of equations that leads to stable solutions of the normal equations, like the QR or Cholesky decomposition or the Householder transformation. For ill-conditioned problems the system with a regularised matrix has to be solved. Iterative algorithms are chosen mainly for two reasons. First, it usually suffices to know the action of the matrix during the iteration, one does not have to store the whole matrix itself. This is particularly attractive if the matrix is very large or its action easily generated. The second application concerns cases where the initially favourable form of the matrix is not utilised but instead worsened by a direct method, e.g. fill-in of a sparse matrix. This is the case especially for reconstruction problems where A is randomly sparse as in tomographic problems [*Björck*, 1996, §7.1.1]. While for small systems there is still no urgent need to resort to iterative solvers for the normal equations, the situation may change if constraints have to be taken into account. The least squares problem becomes now a constrained optimisation problem

$$\min_{\mathbf{x} \in S} f(\mathbf{x}),$$

where S is the set of feasible solutions (e.g. positive solutions etc.) and the so called *objective* or *cost* function is quadratic here. As mentioned in sec. 4.6.3, there are direct solutions for simple inequality constraints, but their implementation is much easier for a certain class of iterative algorithms that has been extensively studied in image reconstruction and signal processing, the so called *projections onto convex sets* (POCS) [see *Byrne*, 2004, and references therein]. This is now illustrated for the under-determined case, where the least squares-least norm solution satisfies $A\mathbf{x} = \mathbf{d}$. Each of the m equations can be regarded as a set S_i (hyperplane in this case) in $\mathbb{R}^n$. If their intersection $S = \bigcap_i S_i$ is nonempty then the sequence P^k of the product of the projection operators P_i onto the sets $P = \prod_i P_i$ will converge towards a member of the feasible set S, fig. 4.6a. The details of convergence have to be

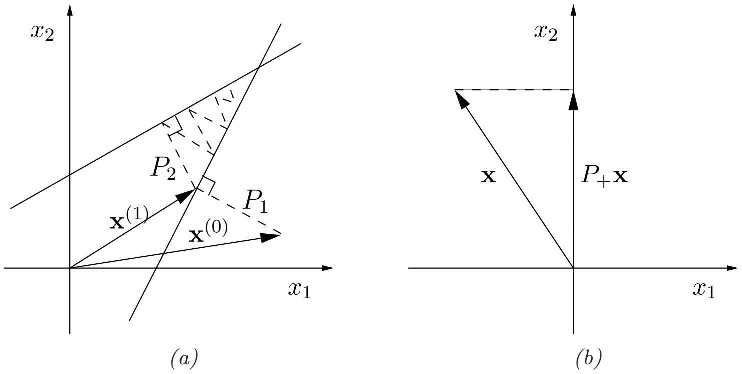


Figure 4.6: (a) The projection onto convex sets *(POCS) for two planes $d_i = \sum_j A_{ij}x_j$, $i = 1, 2$ in the case of a unique solution. The cycle of orthogonal projections $P_1, P_2, P_1, \ldots$ shown here is equivalent to the* algebraic reconstruction technique *(ART). Notice that the smaller the angle between the planes (small singular values) the slower the convergence will be. (b) The projection P_+ onto $\{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} \geq \mathbf{0}\}$ acts by setting negative components to zero.*

specified, of course, also this does not hold for arbitrary sets (it does for convex sets). Furthermore, the operator P does not have to be a projection operator [for details *Byrne*, 2004] but the preceding example is general enough for our purpose. The important point in favour of these iterative algorithms is now that the implementation of inequality constraints like nonnegativity becomes trivial by just inserting the corresponding operator into the above sequence of projections (e.g., fig. 4.6b for the positivity constraint).

4.8.2. Typical convergence behaviour

Before comparing some important iterative methods used to solve least square problems, we mention some characteristics common to their convergence behaviour when applied to ill-conditioned systems.

(i) *Data residuum*: For the iterative solution $\mathbf{x}^{(k)}$ the residual

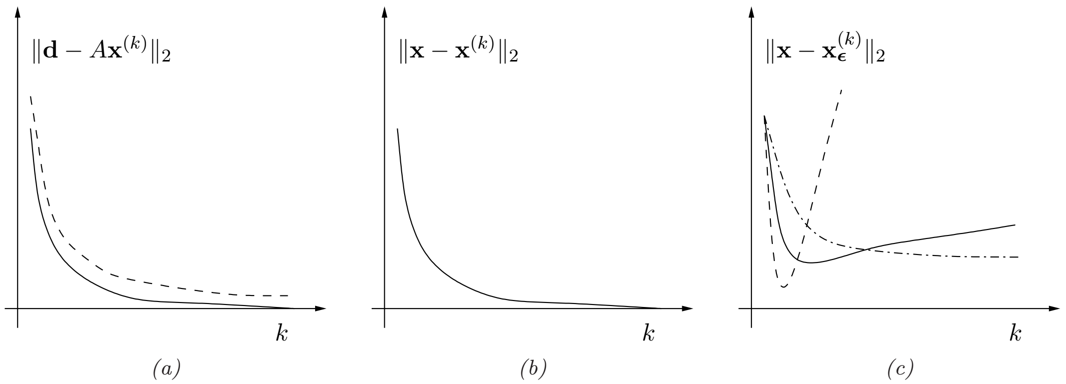
$$\|\Delta\mathbf{d}^{(k)}\|_2 = \|\mathbf{d} - A\mathbf{x}^{(k)}\|_2 \tag{4.65}$$

will typically be monotonically decreasing during the iteration because the iterative algorithms considered here are *constructed* to reduce $\|\Delta\mathbf{d}^{(k)}\|_2$, see fig. 4.7a. In the consistent under-determined case $\|\Delta\mathbf{d}^{(k)}\|_2$ can reach zero.

(ii) *Convergence rates*: Without any errors $\mathbf{x}^{(k)}$ will converge to the exact least-squares solution (or least squares-minimum norm to be more precise) $\mathbf{x}$ as $k \to \infty$, fig. (4.7b). Convergence rates are conveniently analysed with the help of the SVD. Similar to the decomposition of the generalised inverse $A^\dagger = \sum \sigma_i^{-1}\mathbf{u}_i\mathbf{v}_i^T$, eq. (4.34), one writes

$$\mathbf{x}^{(k)} = \sum_i f^{(k)}(\sigma_i, \mathbf{d})\, \sigma_i^{-1}\, \mathbf{u}_i\mathbf{v}_i^T\mathbf{d} \tag{4.66}$$

where the *filter factors* or *response functions* $f^{(k)} \to 1$ as $k \to \infty$. For many iteration schemes they do not depend on $\mathbf{d}$. The filter factors quantify the behaviour necessary for iterative regularisation as suggested in 4.6.3(iii), that is, they give information on the convergence rates of the frequency modes



(a) *(b)* *(c)*

*Figure 4.7: (a) Data residual versus iteration number k for the consistent (–) and inconsistent (--) case. (b) Without errors the iterative algorithm converges to the exact (least squares) solution. (c) Typical behaviour of iterative algorithms if errors are present (*semiconvergence*). The solid line (–) corresponds to the ART algorithm, (--) to the CG method and (-··-) to the Landweber iteration in the example [Hansen, 1998, p. 165].*

of the solution.

(*iii*) *Regularisation*: In the presence of errors the convergence of ill-conditioned equations turns into a behaviour denoted as *semi-convergence*. In the beginning $\mathbf{x}^{(k)}$ approaches the exact solution but deteriorates again for larger iteration numbers, see fig. 4.7c. The solution is regularised by stopping the iteration before this happens – ideally at the optimal iteration number.

4.8.3. Brief comparison of some iteration methods

Dealing with iterative algorithms can be slightly confusing. For example, the very simple *Landweber* iteration discussed extensively by *Twomey* [1997] is known under (at least) five different names in the literature [*Hansen*, 1998]: *Landweber*, *Richardson*, *Fridman*, *Picard* and *Cimino* iteration. Although limited to a minimum, a short, qualitative overview seems in order. The classical methods for least squares problems like the mentioned *Landweber*, the *Jacobi* and *Gauß-Seidel* algorithms are all based on different splittings $A^T A = M - N$, M non-singular, leading to the sequences $M\mathbf{x}^{(k+1)} = N\mathbf{x}^{(k)} + \mathbf{d}$. Only one component of $\mathbf{x}$ is updated in one step which makes it easy to introduce constraints, e.g. in the form of a projection P as discussed above. As their convergence is rather slow, the Gauß-Seidel algorithm has been endowed with a so-called *relaxation* parameter, yielding the *successive over-relaxation* (SOR) scheme. Thanks to its simplicity and flexibility it is being employed in applications for which this property is essential [e.g., *Sauer and Bouman*, 1993; *Fessler*, 1994]. Another classical group of algorithms is the class of *projection* methods (not to be confused with the projections onto convex sets from above. The projections here have different meaning and target space). Widespread projection methods are the gradient-based methods like the *steepest descent* and the *conjugate gradients* (CG) algorithms. They can be viewed as optimisation schemes based on a local search where the search directions are not random but depend on the local properties of the supposed result, like its gradients. In the geophysical literature they appear to be the standard tool to solve inverse problems, but they are not very stable for ill-posed problems, especially not for higher iteration numbers. This led to the mathematically equivalent but more stable LSQR method [see *Björck*, 1996]. Furthermore, these methods "do not easily accommodate inequality constraints" [*Fessler*, 1994], see also [*Calvetti et al.*, 2004] for a short discussion. Finally, there are the *ART like* methods, named after the first algorithm the *algebraic reconstruction technique*, which was visualised in fig. 4.6a. These algorithms belong to the POCS and, depending on whether the projections are performed successively as in $P \sim \prod_i P_i$ or at once $P \sim \sum_i P_i$, they are further classified as *sequential* or *simultaneous*. Because their computation is very cheap in terms of computer memory space, these algorithms have been widely used in image reconstruction. In passing, we note that the original ART is equivalent to the Gauß-Seidel algorithm for the normal equations of the second kind, while the simultaneous version *SART* is a special case of the Landweber iteration. Without giving proofs, we add some relevant facts concerning the convergence and filtering to get the following picture

- SOR allows simple implementation of constraints.
  Small singular values (high spatial frequencies) converge faster than large singular values [*Sauer and Bouman*, 1993; *Fessler*, 1994].

- CG cannot easily be combined with inequality constraints.
  Large eigenvalues converge faster.
  It converges quickly with small residuals, but it is not apt for ill-posed problems. In the case of semi-convergence it converges fast to the optimum, but deteriorates quickly past this point, too [*Björck*, 1996].

- ART like methods can easily be augmented with constraints.
  The filtering behaviour depends on the specific algorithm. ART favours smaller singular values, the simultaneous methods large singular values.

The convergence behaviour of SOR is suspicious, in the sense that the faster converging small singular values are more susceptible to noise. Indeed, it will be shown that the similarly behaving ART performs worse than simultaneous methods. Requesting simplicity and flexibility for the implementation of constraints rules out CG. Thus the next chapter will have a closer look at the ART like methods.

4.8.4. ART, SART and SIRT

The original method of these *row acting methods*, as they are also called because they act only on one row of the matrix A at the same time, goes back to *Kaczmarz* [1937]. It was rediscovered and generalised by *Tanabe* [1971], applied to image reconstruction by *Herman et al.* [1973] where it has been widely used and further developed. A simultaneous version of ART, the *simultaneous image reconstruction technique, SIRT,* was proposed by *Gilbert* [1972]. *SART* [*Andersen and Kak*, 1984], a slightly different simultaneous analogue of ART, improved reconstructed images in medical applications significantly. Numerous modifications of these algorithms concern the acceleration of convergence speed, the application to constrained least squares problems and their formulation as block-iterative (i.e. only parts of the matrix act simultaneously, the rest sequentially) and other aspects. We refer to [*Censor and Herman*, 1987] for a brief review. Although being much slower than direct methods like the filtered back projection (sec. 4.2.1), it appears that with increasing computer power these iterative algorithms regain interest thanks to their flexibility [*Leahy and Byrne*, 2001], especially for cases with inconsistent or incomplete data [e.g., *Mueller and Yagel*, 2000].

Iteration step:
Introducing the row vectors of the $m \times n$ matrix $A = (\mathbf{a}_1, \dots, \mathbf{a}_m)^T$, ART in its simplest form can be written as

$$\mathbf{x}^{(k+1)} = P_i \mathbf{x}^{(k)} = \mathbf{x}^{(k)} + \omega \, \frac{\mathbf{a}_i}{\mathbf{a}_i^T \mathbf{a}_i} (\mathbf{d} - A\mathbf{x}^{(k)})_i, \quad \begin{aligned} k &= 0, 1, 2, \dots \\ i &= k \bmod m + 1, \\ \omega &> 0, \end{aligned} \tag{4.67a}$$

where ω is a relaxation parameter and $\mathbf{x}^{(0)}$ a user defined starting vector. P_i corresponds to the projection onto the ith hyperplane, for $\omega = 1$ illustrated in fig. 4.6a. More explicitly, the iteration cycle looks like this:

$$\begin{aligned} \mathbf{x}^{(1)} &= P_1 \mathbf{x}^{(0)} & \qquad \mathbf{x}^{(m+1)} &= P_1 \mathbf{x}^{(m)} \\ \mathbf{x}^{(2)} &= P_2 \mathbf{x}^{(1)} & &\vdots \\ &\vdots & & \\ \mathbf{x}^{(m)} &= P_m \mathbf{x}^{(m-1)} & & \end{aligned} \tag{4.67b}$$

SIRT reads as

$$\mathbf{x}^{(k+1)} = P\mathbf{x}^{(k)} = \mathbf{x}^{(k)} + \omega \sum_i \frac{\mathbf{a}_i}{\mathbf{a}_i^T \mathbf{a}_i} (\mathbf{d} - A\mathbf{x}^{(k)})_i. \tag{4.68}$$

SART was in [*Andersen and Kak*, 1984] experimentally motivated in order to reduce the impact of noise and derived in the following form

$$\mathbf{x}_j^{(k+1)} = P\mathbf{x}^{(k)} = \mathbf{x}_j^{(k)} + \omega \, \frac{1}{\sum_{i_1} A_{i_1 j}} \sum_i \frac{A_{ij}}{\sum_{j_2} A_{ij_2}} (\mathbf{d} - A\mathbf{x}^{(k)})_i. \tag{4.69}$$

In both cases the correction to the update is a weighted mean of the projections onto the individual hyperplanes. The positivity constraint can be incorporated by inserting the projection P_+ (fig. 4.6b) for every iteration step in ART and SIRT/SART by substituting

$$P_i \rightarrow P_+ P_i \quad \text{with} \quad P_+ x_j = \max(0, x_j)\,, \tag{4.70a}$$

$$P \rightarrow P_+ P\,. \tag{4.70b}$$

Convergence:
None of the three algorithms actually solves $A\mathbf{x} = \mathbf{d}$ or the related normal equations but *rescaled* versions thereof. This can be seen [*Van der Sluis and van der Vorst*, 1987; *Trampert and Lévêque*, 1990] by introducing the weighting matrices

$$L_\phi = \text{diag}[\sum_{i=1}^{m} (A_{ij})^\alpha] \qquad \in \mathbb{R}^{n \times n}, \tag{4.71a}$$

$$L = \text{diag}[\sum_{j=1}^{n} (A_{ij})^{2-\alpha}] \qquad \in \mathbb{R}^{m \times m} \tag{4.71b}$$

with nonnegative parameter α and performing the transformation

$$\bar{A} = L^{-1/2}\, A\, L_\phi^{-1/2}\,, \tag{4.71c}$$

$$\bar{\mathbf{x}} = L_\phi^{1/2}\, \mathbf{x}\,, \tag{4.71d}$$

$$\bar{\mathbf{d}} = L^{-1/2}\, \mathbf{d}\,. \tag{4.71e}$$

The iteration steps now take the form:
ART ($\alpha = 0$)

$$\bar{\mathbf{x}}^{(k+1)} = \bar{\mathbf{x}}^{(k)} + \omega\, \bar{A}^T (\bar{\mathbf{d}} - \bar{A}\bar{\mathbf{x}}^{(k)})_i\, \mathbf{e}_i\,, \quad \text{with} \quad (\mathbf{e}_i)_j = \delta_{ij} \tag{4.72a}$$

SIRT ($\alpha = 0$) and SART ($\alpha = 1$)

$$\bar{\mathbf{x}}^{(k+1)} = \bar{\mathbf{x}}^{(k)} + \omega\, \bar{A}^T (\bar{\mathbf{d}} - \bar{A}\bar{\mathbf{x}}^{(k)})\,. \tag{4.72b}$$

Eqs. (4.72) show that a solution $\mathbf{x}^{(k)} \to \mathbf{x}$ satisfies the rescaled normal equations

$$\bar{A}^T \bar{A}\bar{\mathbf{x}} = \bar{A}^T \bar{\mathbf{d}} \quad \text{or} \quad A^T L^{-1} A\mathbf{x} = A L^{-1}\mathbf{d} \tag{4.73}$$

that correspond to a weighted least squares problem, c.f. eq. (4.39). A correct analysis using the SVD yields:

CONVERGENCE OF SIRT AND SART

The one parameter family of simultaneous algorithms eqs. (4.72b) with a nonnegativity constraint as in eq. (4.70b) converges for $0 \le \alpha \le 2$, $0 < \omega < 2$ to the solution of the least square-minimum norm problem

$$\min_{\mathbf{x}\in S} (\mathbf{x} - \mathbf{x}^{(0)})^T L_\phi (\mathbf{x} - \mathbf{x}^{(0)}), \quad S = \{\mathbf{x} \in \mathbb{R}^n \mid\ (\mathbf{d} - A\mathbf{x})^T\, L^{-1}\, (\mathbf{d} - A\mathbf{x}) = \min, \quad \mathbf{x} \ge 0\}. \tag{4.74}$$

This means

- The iteration start $\mathbf{x}^{(0)}$ is the a priori of the least squares problem.
- *Implicit* weighting of the a priori by
  SIRT ($\alpha = 0$): $L_\phi = m\ \mathbb{1}_n$, i.e. no weighting
  SART ($\alpha = 1$): $L_\phi = \text{diag}(\sum_i A_{ij})$, "sum of all light path lengths in box j, i.e. sampling of box j".
- *Implicit* weighting of the data by
  SIRT ($\alpha = 0$): $L^{-1} = \text{diag}(\sum_j A_{ij}^2)^{-1}$, "inverse sum of (box LP lengths)^2 for LP_i"
  SART ($\alpha = 1$): $L^{-1} = \text{diag}(\sum_j A_{ij})^{-1}$, inverse length of LP_i,

where the simple geometrical interpretation holds only for the box basis. Finally, the filter factors, see eq. (4.66), of the nonzero singular values for the transformed system are

$$f^{(k)}(\bar{\sigma}_i) = \left(1 - (1 - \omega\ \bar{\sigma}_i^2)^k\right), \quad i \le \text{rank}[\bar{A}], \tag{4.75}$$

where for the singular values of $\bar{A}$ holds $0 \le \bar{\sigma}_i \le 1$. For the components of $\bar{\mathbf{x}}$ in the system of the singular vectors, i.e. of $\mathbf{x}' = V^T\bar{\mathbf{x}}$, the convergence rate is given by

$$x'^{(k)}_j - x'^{(\infty)}_j = (1 - \omega\ \bar{\sigma}_j^2)^k (x'^{(0)}_j - x'^{(\infty)}_j), \quad j \le \text{rank}[\bar{A}]. \tag{4.76}$$

As the convergence is faster for larger values of $\omega\,\bar{\sigma}_i^2$, the relaxation parameter ω can be used to speed up the iteration process. Eq. (4.76) establishes the above statement for the frequency filtering of the simultaneous ART methods.

For ART the simple SVD approach does not work because of the series of operator products. Basically the method converges similarly to the simultaneous versions except for the inconsistent case. Here the algorithm may not converge to a single fix point but to a cyclic subsequence. Comparing eqs. (4.67a), (4.68) it follows an implicit weighting of ART like SIRT. For details see [*Kak and Slaney*, 2001] where a geometrical illustration is given and [*Jiang and Wang*, 2003] for further mathematical references.

4.9. Other reconstruction methods

One could think of various other approaches to the tomographic inverse problem eq. (4.4) like genetic algorithms, neural networks or Monte Carlo methods. While some of these approaches indeed occur in the literatue and might be well justified, they are nevertheless black box methods that may not bear great potential for further insight. This chapter concludes by sketching some important methods that are also employed within the context of tomographic or discrete inverse problems.

4.9.1. Maximum entropy and maximum likelihood

The concept of entropy was applied in sec. 4.7.2 to calculate the information content of a measurement under the assumption that all quantities involved can be described by Gaussian probability density functions. But entropy can also be used constructively as a reconstruction principle based on the assumption that the most probable reconstruction among all that reproduce the data (and possibly further constraints) is the one with maximum entropy. This principle has the important property that it does not introduce any new correlation in the reconstruction result that goes beyond the data [*Gull and Daniell*, 1978]. Strictly speaking, the *maxium entropy method* MEM (or MaxEnt) requires probability density functions but usually the configurational entropy $S(\mathbf{x}) = \sum_j x_j \ln x_j$ is computed directly in state space. The most simple formulation of the reconstruction problem then reads

$$\max_{\mathbf{x}\in S'} S(\mathbf{x}), \quad S' = \{\mathbf{x} \in \mathbb{R}^n \mid \ (\mathbf{d} - A\mathbf{x})^T (\mathbf{d} - A\mathbf{x}) = \min,\ \mathbf{x} \geq 0\}. \tag{4.77}$$

The maximum entropy method has found numerous applications, sometimes rigorously founded, but in many cases on a mere empirical basis [see *Censor and Herman*, 1987, for references]. As for the quadratic objective function several iterative algorithms have been proposed to solve eq. (4.77) for the consistent, i.e. equality constrained case which is especially important in image reconstruction. The *multiplicative algebraic reconstruction technique* MART was introduced by *Gordon et al.* [1970] and has found a number of modifications including its simultaneous version SMART (see [*Byrne*, 2004; *Reis and Roberty*, 1992] and references therein). The iterative step is as simple as for ART like methods but the correction is applied multiplicatively, not additatively. Another maximum entropy method sometimes used is MENT [*Minerbo*, 1979]. The entropy objective function in image reconstruction has been especially successful for noisy images.

The *maximum likelihood* ML method applied in this context tries to estimate parameters $\mathbf{x}$ from an incomplete set of data **d** by maximising the *likelihood* $L(\mathbf{x}) = P(\mathbf{d}|\mathbf{x})$, i.e. the probability that a given **x** leads to the observed data **d**. Using the log-likelihood its most simple form states thus

$$\max_{\mathbf{x}\in\mathbb{R}^n} \ln P(\mathbf{d}|\mathbf{x})$$

A universal iterative algorithm converging towards the ML estimate, the *expectation maximisation* EM, was developed by *Dempster et al.* [1977]. It was applied to the specific case of emission tomography under the assumption that the measurement can be described as a Poisson process in [*Shepp and Vardi*, 1982; *Vardi et al.*, 1985]. The resulting iterative algorithm, usually referred to as EMML but also as MLEM or EM, is frequently used in image reconstruction. It is also a very simple algorithm but it seems to be quite sensitive to data noise [*Censor and Herman*, 1987; *Byrne*, 2001].

Although developed independently, the entropy maximising SMART and the likelihood maximising EMML (and in fact further algorithms) can be obtained by minimising a functional of the form $KL(\mathbf{x},\mathbf{y}) = \sum_j x_j \ln(x_j/y_j) + y_j - x_j$ [*Byrne*, 2001]. The Kullback-Leibler distance KL is closely related to the cross entropy which is used in information theory to measure the overall difference between two probability distributions.

4.9.2. Backus-Gilbert method

The Backus-Gilbert method (see [*Backus and Gilbert*, 1970] for the original paper) is often quoted in geophysical and atmospheric science, as it represents a diagnostic formulation of the inverse problem that also leads to a least squares solution. *Hansen* [1998] refers to in the context of mollifier methods. Basically, it attempts to *construct* the inverse from a limited amount of data in a way that leads to optimal resolution. Usually the presentation is for a forward model in the form of ordinary integral equations $d_i = \int dr\ K_i(r)\ c(r)$, here it is the discrete system $A\mathbf{x} = \mathbf{d}$. First, it is assumed that the reconstructed or estimated concentration $\hat{c}$ at some $\mathbf{r}_0$ is given by a linear combination of the observed data

$$\hat{c}(\mathbf{r}_0) = \sum_i a_i^{inv}(\mathbf{r}_0)\ d_i. \tag{4.78}$$

Second, assume that there is a function, the *averaging* or *resolving kernel* R, relating the estimated to the true concentration field c

$$\hat{c}(\mathbf{r}_0) = \int d\mathcal{V}\ R(\mathbf{r},\mathbf{r}_0)\ c(\mathbf{r})\,. \tag{4.79}$$

Ideally, R would be the δ-functional, in practice it is some smoothing or averaging functional (c.f. sec. 4.6.4). The key idea is now to choose R as narrow as possible while still reproducing the measurement data. To relate eqs. (4.78, 4.79) recall that $c(\mathbf{r}) \sim \sum_j x_j\ b_j(\mathbf{r})$ where from now on – merely for simplicity – the basis functions are taken to be orthonormal. Then also

$$R(\mathbf{r},\mathbf{r}_0) = \sum_j r_j(\mathbf{r}_0)\ b_j(\mathbf{r}) \tag{4.80}$$

and replacing d_i in eq. (4.78) by $(A\mathbf{x})_i$ it follows

$$r_j(\mathbf{r}_0) = \sum_j A^T_{ji}\ a_i^{inv}(\mathbf{r}_0)\,. \tag{4.81}$$

The third assumption concerns the functional that measures the width of the averaging kernel $R(\mathbf{r}, \mathbf{r}_0)$. Its generic form is to some extent arbitrary but typically something like

$$J(\mathbf{r}_0) = \alpha \int dV \;\, R(\mathbf{r}, \mathbf{r}_0)^2 |\mathbf{r} - \mathbf{r}_0|^\beta \, .$$

Inserting eq. (4.80) leads with eq. (4.81) to a quadratic functional of $\mathbf{a}^{inv} = (a_1^{inv}, \ldots, a_m^{inv})^T$ and thus to the least squares problem

$$\min_{\mathbf{a}^{inv}} J \ .$$

The solution of this problem is simply a particular solution of the under-determined inverse problem [see also *Tarantola*, 2005, pp.191–194]. It might have to be modified (regularised) not only if it is unstable, but also for the following reason. The variance of the estimated concentration $\hat{c}$ due to measurement errors can with eq. (4.78) easily be seen to equal $\mathbf{a}^{inv}(\mathbf{r}_0)^T S_{\boldsymbol{\epsilon}} \mathbf{a}^{inv}(\mathbf{r}_0)$, i.e. it also depends on the a_i^{inv}. In order not to risk too large variances while tightening up the resolution, it might be necessary to minimise a weighted mean of J and the variance. This amounts to a Tikhonov regularisation.

4.9.3. Example of global optimisation – fitting Gaussian exponentials

So far the fit problem

$$\min_{x_1, x_2, \ldots} \; \| d_i - \int_i ds \; c(\mathbf{r}; x_1, x_2, \ldots) \|_2$$

with model parameters x_j has been considered only for parametrisations that lead to discrete quadratic minimisation problems which in turn yield linear equations. In cases where the functional form of the concentration field is known to some extent it can make more sense to exploit this knowledge rather than to use a completely unspecific parametrisation. If the resulting equations cannot be solved by standard methods, global optimisation algorithms have to be employed. The disadvantage that a black box algorithm is totally ignorant to the tomographic origin of the problem is not as critical as the convergence behaviour and speed of the general purpose optimisation algorithm.

An example for this approach can be found in [*Drescher et al.*, 1996] for the 2-D reconstruction of indoor gas concentrations. Assuming that the indoor dispersion can be approximated by a Gaussian diffusion equation, the authors make the following ansatz for the resulting concentration field in the case of a limited number of point sources

$$\begin{aligned} &\hat{c}(x, y; C_1, x_1, y_1, \sigma_{x_1}, \sigma_{y_1}, \phi_1, C_2, \ldots) \\ &\qquad = \sum_k C_k \; e^{-\frac{1}{2}\left((\cos\phi_k \;\, (x - x_k) + \sin\phi_k \;\, (x - x_k))^2 / \sigma_{x_k}^2 + (\, x \leftrightarrow y \,) \right)} \, , \end{aligned} \tag{4.82}$$

i.e. for each Gaussian six parameters for peak height, location, orientation and variances. The optimisation problem arising is highly nonlinear and solved using the *Amebsa* routine from [*Press et al.*, 1992], a combination of the simplex and the simulated annealing method. A standard search for one reconstruction takes about $2\,h$ according to the authors. As another way to look at the method is that the discrete (local) basis functions have been replaced by smooth (global) basis functions, the approach was named the “smooth basis function minimization” (SBFM) in the original work.

An approach based on the fitting of symmetric Gaussians is also proposed by *Giuli et al.* [1999] and applied to the reconstruction of volcanic CO_2 distributions in [*Belotti et al.*, 2003], c.f. sec. 3.3.

Part II.

Methodology

5. The Error of the Reconstructed Distribution

This chapter develops a closed methodology for the complete reconstruction error of the continuous inverse problem eq. 4.4, p. 50. For reasons that will become clear from the first section, there is hardly a standard treatment of the *full* error of the reconstruction result. While errors related to the discrete system $A\mathbf{x} = \mathbf{d}$ are commonly covered (especially if the optimal estimate is employed, otherwise very often merely in form of sensitivity studies) the error associated with the discretisation is rarely addressed. I am not aware of my approach to it in sec. 5.2.1 for local basis functions appearing in the literature on tomography with small numbers of light paths. The theoretical discussion of the different error contributions in sec. 5.2 is neither restricted to tomographic inverse problems nor does it depend on any specific inversion method. Sec. 5.3 treats the purely numerical estimation of the total reconstruction error as it was more or less used in [*Laepple et al.*, 2004]. Finally, the last section on error norms (or overall errors) connects quality evaluation measures from image reconstruction and atmospheric modelling.

5.1. The problem in defining the reconstruction error

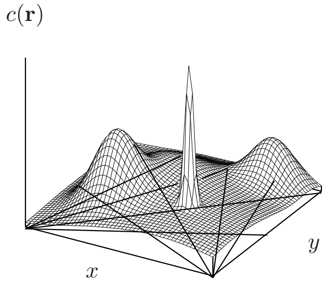


Figure 5.1: The problem of quantifying the reconstruction error, illustrated here for the case where the 2-D section through $c(\mathbf{r})$ shows structures that cannot be resolved by a measurement with the path geometry shown. If the variability of $c(\mathbf{r})$ is not confined, the reconstruction error can become arbitrary large. Using the definitions of sec. 5.2 it is the sum of discretisation *and* inversion error *that is unbounded if $c(\mathbf{r})$ is not constrained.*

The aim of an error calculation is to derive bounds for the deviation $\Delta\hat{c}(\mathbf{r})$ of the reconstructed concentration field $\hat{c}(\mathbf{r})$ from the true state $c(\mathbf{r})$:

$$\Delta_l\hat{c}(\mathbf{r}) \;\leq\; \underbrace{c(\mathbf{r}) - \hat{c}(\mathbf{r})}_{\Delta\hat{c}(\mathbf{r})} \;\leq\; \Delta_u\hat{c}(\mathbf{r}). \tag{5.1}$$

These bounds can be hard constraints or probabilistic statements. An example for the former is the remainder term of the Taylor expansion, an example for the latter the 1σ declaration of a measurement result. Neglecting all measurement errors for the moment, ideally, it should be possible to calculate these bounds solely from the measured data $\mathbf{d}$ and the details of light path geometry and inversion algorithm. It was pointed out in remark 4.2.2 that a *finite* number of measurements cannot pin down the continuous function $c(\mathbf{r})$. Unfortunetaly, the proof in [*Natterer*, 2001] is not constructive. That is, it does not contain a recipe for the functions that give rise to exactly the same column densities and I am not aware of any other work that does. Evidently the ambiguity of the solution is aggravated for decreasing number of light paths and coarser discretisations (see fig. 5.1). On the other hand, it was noticed in the same remark that the space of solutions can be reduced by further assumptions on the properties of $c(\mathbf{r})$. Thus, we find

- The reconstruction error cannot be estimated (at least not within this work) from the measurement data $\mathbf{d}$ and reconstruction procedure only. Further information about the true concentration $c(\mathbf{r})$ is necessary or assumptions have to made. In the case of statistical inversion this information consists in the a priori probability density function.

- In other words, the declaration of a reconstruction error as in eq. (5.1) only makes sense in combination with the assumptions on the true solution $c(\mathbf{r})$ it is based on.

- It is important to realise that this ambiguity is not an effect of the inverse problem's instability. This is taken care of by regularisation and the resulting change of the solution can be calculated and expressed in terms of perturbation bounds (sec. 5.2.3).

A remedy of this dissatisifying situation, where not only the solution of the ill-posed problem involves more or less arbitrary assumptions but also the estimate of its uncertainty, is beyond the scope of this thesis. But not all contributions to the total reconstruction error $\Delta\hat{c}(\mathbf{r})$ imply hidden assumptions as shown in the following.

5.2. Composition of the total error

Going through the logical steps of the reconstruction again, there appears first the approximation of the continuous $c(\mathbf{r})$ by a finite set of basis functions, c.f. eq. (4.10). We call this discrepancy the *discretisation error*. Even if $c(\mathbf{r})$ can be expressed exactly as a linear combination of the basis functions, the inversion of $A\mathbf{x} = \mathbf{d}$ is in general not possible in a unique and/or stable way. This *inversion error* is defined in $\mathbb{R}^n$. Finally, the *measurement error* $\boldsymbol{\epsilon}$ will affect the fit result $\mathbf{x}_{\boldsymbol{\epsilon}}$ and propagate to the reconstructed field. The effect of the measurement error is controlled by regularising the solution $\mathbf{x} \mapsto \mathbf{x}_\lambda$ with regularisation parameter λ, giving rise to the *regularisation error*. The deviation of the perturbed from the error free regularised solution is called *perturbation error*. In formulas

$$\Delta\hat{c}(\mathbf{r}) = \underbrace{c(\mathbf{r}) - [\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x}_{id}}_{\Delta c_{disc}(\mathbf{r})} + \underbrace{[\boldsymbol{\phi}(\mathbf{r})]^T (\overbrace{\mathbf{x}_{id} - \mathbf{x}}^{\Delta\mathbf{x}_{inv}})}_{\Delta\hat{c}_{inv}(\mathbf{r})} + \underbrace{[\boldsymbol{\phi}(\mathbf{r})]^T (\overbrace{\mathbf{x} - \mathbf{x}_\lambda}^{\Delta\mathbf{x}_{reg}})}_{\Delta\hat{c}_{reg}(\mathbf{r})} + \underbrace{[\boldsymbol{\phi}(\mathbf{r})]^T (\overbrace{\mathbf{x}_\lambda - \mathbf{x}_{\boldsymbol{\epsilon}}}^{\Delta\mathbf{x}_{pert}})}_{\Delta\hat{c}_{pert}(\mathbf{r})}, \tag{5.2}$$

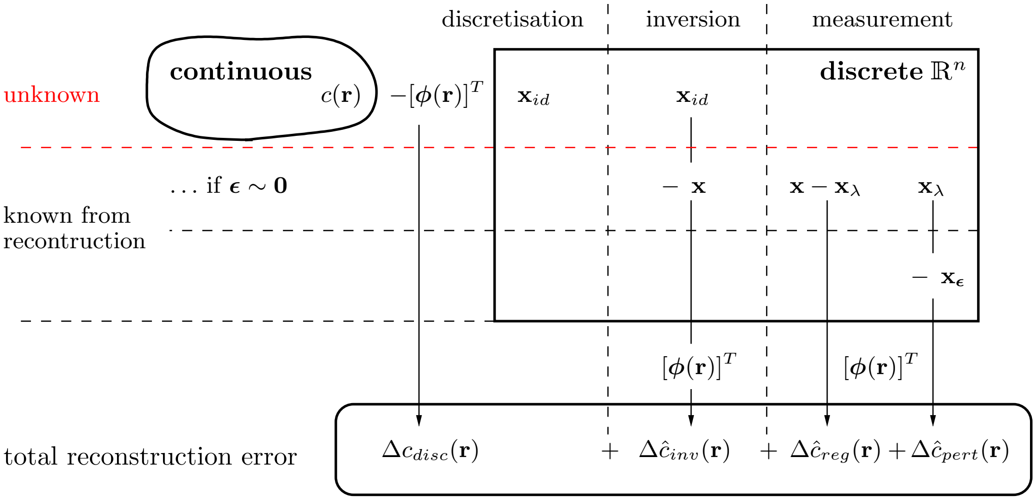


Figure 5.2: The composition of the total reconstruction error $\Delta\hat{c}(\mathbf{r})$.

see also fig. 5.2. Here $\mathbf{x}_{id}$ is the ideal representation of the continuous $c(\mathbf{r})$ for a given basis, i.e. the $\mathbf{x}$ that leads to the smallest discretisation error in terms that yet have to be defined. $[\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x}$ is the reconstruction result as it would be without measurement errors and regularisation (e.g. a least-squares or least-squares least-norm solution). $[\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x}_\lambda$ is the regularised solution without errors, $[\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x}_{\boldsymbol{\epsilon}}$ with. Before discussing the individual contributions in turn, we observe that:

- The different errors are usually related to each other. For example, choosing a coarser discretisation that makes the problem over-determined will lead to small or vanishing inversion errors, but to larger discretisation errors and higher sensitivity to noise. Fine discretisation that makes the problem under-determined implies larger inversion errors but the regularisation and perturbation error may become less important (see sec. 4.4).

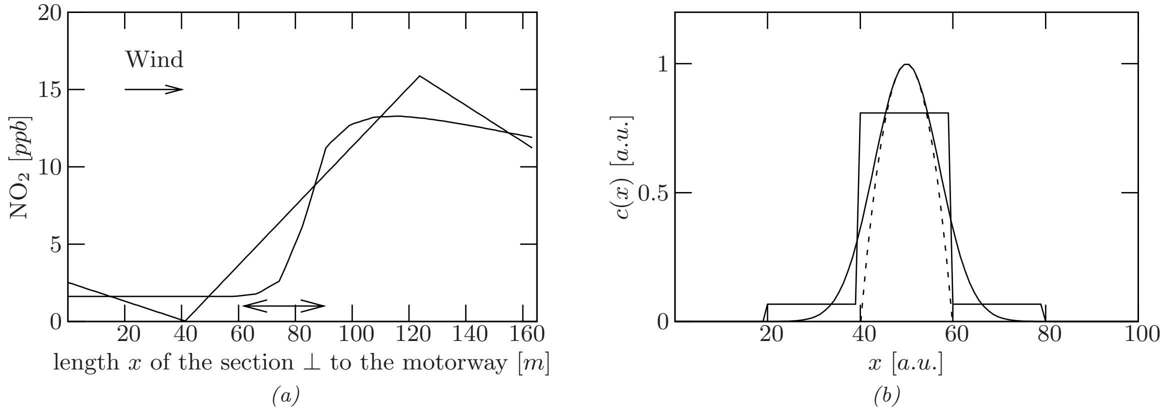
- The significance of each contribution to the total error largely depends on the application. For tomographic measurements with large numbers of integration paths like X-ray tomography the discretisation error is usually neglected (or not discussed). Likewise the error analysis for atmospheric inversion problems is predominantly in $\mathbb{R}^n$ for the $\Delta\mathbf{x}_{inv}$, $\Delta\mathbf{x}_{reg}$ and $\Delta\mathbf{x}_{pert}$ only.

- Statistical inversion in $\mathbb{R}^n$ (this is how the optimal estimate, sec. 4.7.1, is usually used) provides an estimate of $\Delta\mathbf{x}_{inv}+\Delta\mathbf{x}_{reg}+\Delta\mathbf{x}_{pert}$ in terms of the a posteriori probability density function (or the a posteriori covariance in the case of the optimal estimate where this function is a Gaussian distribution) but it has to be specified, whether the discretisation error is contained in the a priori uncertainty or neglected.

5.2.1. The ideal discretisation and the discretisation error

The ideal parameters are given as minimum of some objective function measuring the misfit between continuous and parametrised field. Following remark 4.6.5, again the 2-norm is chosen so that $\mathbf{x}_{id}$ is given by

$$\min_{\mathbf{x}} \| c(\mathbf{r}) - [\boldsymbol{\phi}(\mathbf{r})]^T \mathbf{x} \|_2, \quad \mathbf{x} \geq \mathbf{0} \tag{5.3}$$



(a) *(b)*

Figure 5.3: (a) Horizontal profiles at 5 m above ground of the NO$_2$ concentration field perpendicular to a motorway [Bäumer et al., 2005] and its bilinear discretisation on grid of 4×3 boxes as used for reconstruction in [Laepple et al., 2004]. ↔ indicates the position of the motorway. The maximum discretisation error is ∼ 3 ppb. (b) Box discretisation of a Gaussian function with $\sigma = 7$ a.u.. The parabola shows the quadratic Taylor expansion around the peak maximum used in the text to estimate the width of a general peak.

with

$$\|\cdot\|_2^2 = \int_\Omega d\mathcal{V}\ (\cdot)^2 \tag{5.4}$$

and Ω being the volume or area for two and three dimensions, respectively. The normal equations can be written as

$$\Phi\,\mathbf{x}_{id} = \mathbf{c} \tag{5.5a}$$

where

$$\Phi = \int_\Omega d\mathcal{V}\ \boldsymbol{\phi}(\mathbf{r})[\boldsymbol{\phi}(\mathbf{r})]^T, \qquad \mathbf{c} = \int_\Omega d\mathcal{V}\ \boldsymbol{\phi}(\mathbf{r})c(\mathbf{r}). \tag{5.5b}$$

For the orthonormal box basis calculation of these quantities and inversion of the normal equation is trivial:

$$\Phi_{ij} = \mathcal{V}_i\ \delta_{ij}, \qquad c_i = \mathcal{V}_i\ \overline{c(\mathbf{r})}_i \tag{5.6a}$$

where $\mathcal{V}_i$ is the area or volume of box i and $\overline{(\cdot)}_i$ denotes the spatial mean in it. Thus

$$x_{id_i} = \overline{c(\mathbf{r})}_i, \tag{5.6b}$$

i.e. the ideal discretisation parameters $\mathbf{x}_{id}$ for the box basis are the box averages of $c(\mathbf{r})$. For the box basis it is easy to see that if $A\mathbf{x} = \mathbf{d}$ has a unique solution, it is $\mathbf{x}_{id}$. This is not obvious neither for a more general local basis nor for the special case of bilinear basis functions. Also the inversion of the normal equations for the bilinear basis is not so simple because Φ is not diagonal. Fig. 5.3a is the result of a numerically calculated ideal bilinear discretisation for the NO$_2$ distribution perpendicular to a motorway modelled by *Bäumer et al.* [2005]. It shows horizontal profiles of the model concentration and the ideally parametrised $c_{id}(\mathbf{r}) = [\boldsymbol{\phi}(\mathbf{r})]^T\,\mathbf{x}_{id}$ for a discretisation grid as it was used by *Laepple et al.* [2004] to reconstruct the motorway plume from measurement data.

Before trying to estimate the discretisation error, we illustrate which properties of $c(\mathbf{r})$ might be important by considering one dimension for simplicity. The gradient of c is a critical property for the box basis functions. Only if the relative change $\Delta c/c$ within a box is much smaller than 1, we would say

> that the box basis is a good approximation of c. Thus for a box length Δx: $\Delta c/c \sim |\frac{\partial c}{\partial x}|\ \Delta x/c \ll 1$ or $\Delta x \ll c\ |\frac{\partial c}{\partial x}|^{-1}$. Linear basis functions can easily model gradients but not changes of the slope. The change of the slope, essentially the curvature, characterises peaks (or valleys). Assuming that the peak can locally be approximated by a quadratic Taylor expansion (c.f. fig. 5.3b), the peak width Δ at the baseline is $\Delta \sim 2\sqrt{2}\sqrt{c\,|\frac{\partial^2 c}{\partial x^2}|^{-1}}$, where the function values refer to the maximum. The local basis functions can mimic the peak appropriately only if the box length is small compared to the peak width, thus $\Delta x \ll \sqrt{c\,|\frac{\partial^2 c}{\partial x^2}|^{-1}}$.

As the above definition eq. (5.3) of the ideal discretisation involves collocation not interpolation, simple estimates for interpolation errors cannot be used. Instead, we take the definition of the discretisation error (ignoring the nonnegativity constraint for the moment)

$$\Delta c_{disc}(\mathbf{r}) = c(\mathbf{r}) - [\boldsymbol{\phi}(\mathbf{r})]^T \Phi^{-1} \mathbf{c} \tag{5.7}$$

and approximate $c(\mathbf{r})$ in $\mathbf{c}$ by a Taylor expansion to second order. For the 2-D box basis functions it is shown in appendix C that

$$\Delta c_{disc}(\mathbf{r}) \sim c(\mathbf{r}) - \sum_{i=1}^{n} \left(c(\mathbf{r}_i) + \tfrac{1}{24}\big(\Delta x_i^2 \frac{\partial^2 c}{\partial x^2} + \Delta y_i^2 \frac{\partial^2 c}{\partial y^2}\big)(\mathbf{r}_i) \right) \phi_i^0(\mathbf{r}) \tag{5.8a}$$

where

$\mathbf{r}_i$: centre of box i,
$\mathbf{r}$: chosen such that Taylor expansion around $\mathbf{r}_i$ is valid,
Δx_i : length of box i in x direction, Δy_i the same for y

and especially in the centre of the box

$$\Delta c_{disc}(\mathbf{r}_i) \sim -\tfrac{1}{24}\big(\Delta x_i^2 \frac{\partial^2 c}{\partial x^2} + \Delta y_i^2 \frac{\partial^2 c}{\partial y^2}\big)(\mathbf{r}_i). \tag{5.8b}$$

> Fig. 5.3b shows the example of a Gaussian peak with $\sigma = 7$ *a.u.* for the piecewise constant discretisation with 5 boxes. At the maximum the second derivative of a Gaussian peak is given by $-\sigma^{-2}$. Inserting this and the box length of 20 *a.u.* into eq. (5.8b) yields for one dimension $\Delta c_{disc} \sim 0.3$ *a.u.* which roughly argrees with the actual values.

A similar estimate for the bilinear basis is complicated by the calculation of the inverse Φ^{-1} and not pursued further[1].

---

[1] Trying to estimate discretisation errors from the textbook error bounds for interpolating quadrature in box i, again for one dimension, in the form

$$\frac{1}{\Delta x_i} \Big| \int_{box\ i} dx\ c - I_0(c) \Big| \le \max_{box\ i} \Big| \frac{\partial c}{\partial x} \Big|$$

$$\frac{1}{\Delta x_i} \Big| \int_{box\ i} dx\ c - I_1(c) \Big| \le \tfrac{1}{2}\Delta x_i \max_{box\ i} \Big| \frac{\partial^2 c}{\partial x^2} \Big|$$

(with I_0, I_1 the integrals over the interpolation polynomials of degree 0 and 1, respectively) did not prove very useful in that these bounds are too high.

5.2.2. The inversion error

The inversion error was defined as $\Delta\hat{c}_{inv}(\mathbf{r}) = [\boldsymbol{\phi}(\mathbf{r})]^T(\mathbf{x}_{id} - \mathbf{x})$ for an error free retrieval $\mathbf{x}$ and can only be estimated for appropriate $\mathbf{x}_{id}$. If the reconstruction is given by a linear map and we regard $\mathbf{x}_{id}$ as the true vector state of the atmosphere for a given discretisation, then by recalling the definition of the resolution matrix eq. (4.56)

$$\mathbf{x} - \mathbf{x}_a = R\,(\mathbf{x}_{id} - \mathbf{x}_a)$$

one has

$$\Delta\mathbf{x}_{inv} = \mathbf{x}_{id} - \mathbf{x} = (\mathbb{1}_n - R\,)(\mathbf{x}_{id} - \mathbf{x}_a). \tag{5.9}$$

and

$$\begin{aligned} \|\Delta\mathbf{x}_{inv}\|_2 &\le \|\mathbb{1}_n - R\|_2\,\|\mathbf{x}_{id} - \mathbf{x}_a\|_2 \\ &= \sqrt{n - \operatorname{rank}[A]}\,\|\mathbf{x}_{id} - \mathbf{x}_a\|_2 \quad \text{for the generalised inverse.} \end{aligned} \tag{5.10}$$

The difference $-\Delta\mathbf{x}_{inv}$ is called *smoothing error* in [*Rodgers*, 2000, sec. 3.1.2] for reasons discussed in sec. 4.6.4. Mind that the inversion error is 'proportional' to the discrepancy between R and the ideal resolution matrix $\mathbb{1}_n$ *and* the difference of the guess $\mathbf{x}_a$ and the true state. If $\mathbf{x}_a$ happens to be the true $\mathbf{x}_{id}$, the inversion error becomes zero despite a resolution matrix deviating from the ideal one. If the reconstruction method is nonlinear $\Delta\hat{c}_{inv}(\mathbf{r})$ can only be inferred numerically.

5.2.3. The measurement error

Random noise

Only unbiased random noise is considered here. It appears on the right hand side of $A\mathbf{x}_{\boldsymbol{\epsilon}} = \mathbf{d}_{\boldsymbol{\epsilon}}$ and propagates through the inversion to the solution $\mathbf{x}_{\boldsymbol{\epsilon}}$ and further to $\hat{c}(\mathbf{r})$. Regularisation controls the effect of errors, e.g. by some parameter λ as in the Tikhonov method, $\lambda = \sigma_a^{-1}$ in the optimal estimate or the iteration number k in the case of iterative regularisation.

We consider the case where the inversion is performed by a linear map A_λ^{inv}. Then

$$\mathbf{x} - \mathbf{x}_{\boldsymbol{\epsilon}} = A_\lambda^{inv}(\mathbf{d} - \mathbf{d}_{\boldsymbol{\epsilon}}) = A_\lambda^{inv}\boldsymbol{\epsilon} \tag{5.11a}$$

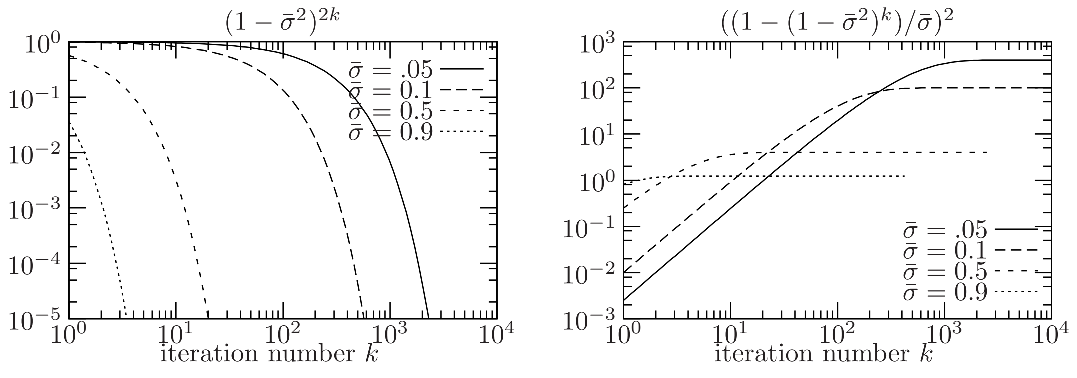
with covariance

$$S_{\mathbf{x}} = \mathrm{E}[(\mathbf{x} - \mathbf{x}_{\boldsymbol{\epsilon}})(\mathbf{x} - \mathbf{x}_{\boldsymbol{\epsilon}})^T] = A_\lambda^{inv} S_{\boldsymbol{\epsilon}} A_\lambda^{inv\,T} \tag{5.11b}$$

(mind that for Bayesian least squares, e.g. the optimal estimate, eq. (4.59) holds instead) and variance for the concentration field

$$\sigma_{\hat{c}}^2(\mathbf{r}) = \operatorname{var}[\hat{c}(\mathbf{r})] = [\boldsymbol{\phi}(\mathbf{r})]^T S_{\mathbf{x}} \boldsymbol{\phi}(\mathbf{r}). \tag{5.11c}$$

Thus in principle, the impact of random noise can be estimated without any assumptions on the unknown concentration $c(\mathbf{r})$. But in practice the regularisation depends on the concentration field at hand. For iterative solution the optimal iteration number, for example, takes quite different values depending on the dimensions of the system $A\mathbf{x} = \mathbf{d}$ and the type of distribution, e.g. peaks or smooth background etc. We look at the individual contributions $\Delta\mathbf{x}_{reg}$ and $\Delta\mathbf{x}_{pert}$ in more detail. Assuming uncorrelated measurement errors $S_{\boldsymbol{\epsilon}} = \sigma_{\boldsymbol{\epsilon}}^2\mathbb{1}_m$ to simplify the formulas, a straightforward calculation



(a) *(b)*

Figure 5.4: (a) Contribution of the singular value $\bar\sigma$ to the regularisation error $\|\Delta\bar{\mathbf{x}}_{reg}\|_2^2$, eq. (5.13a), as a function of the iteration number for four selected values of $\bar\sigma$. (b) The same for $\|\Delta\bar{\mathbf{x}}_{pert}\|_2^2$ in eq. (5.13b).

by means of the SVD yields in general [*Van der Sluis and van der Vorst*, 1987]

$$\|\Delta\mathbf{x}_{reg}\|_2^2 = \sum_i {x'_i}^2 (f(\sigma_i, \mathbf{d}) - 1)^2 \tag{5.12a}$$

$$\mathrm{E}\big[\|\Delta\mathbf{x}_{pert}\|_2^2\big] = \sigma_{\boldsymbol\epsilon}^2 \sum_i \Big(\frac{f(\sigma_i,\mathbf{d})}{\sigma_i}\Big)^2 \tag{5.12b}$$

$$\mathrm{E}\big[\|\mathbf{x}-\mathbf{x}_{\boldsymbol\epsilon}\|_2^2\big] = \|\Delta\mathbf{x}_{reg}\|_2^2 + \mathrm{E}\big[\|\Delta\mathbf{x}_{pert}\|_2^2\big], \tag{5.12c}$$

where $f(\sigma_i, \mathbf{d})$ are the filter factors of the regularised solution, c.f. eq. (4.66), and $\mathbf{x}' = V^T\mathbf{x}$ the vector $\mathbf{x}$ in the system of singular vectors. Inserting the filter factors eq. (4.75) for SIRT like algorithms eq. (4.72b) gives for the *transformed* system, eq. (4.71), $\bar A = L^{-1/2}\, A\, L_\phi^{-1/2}$, $\bar{\mathbf{x}} = L_\phi^{1/2}\,\mathbf{x}$, $\bar{\mathbf{d}} = L^{-1/2}\,\mathbf{d}$ the bounds

$$\|\Delta\bar{\mathbf{x}}_{reg}\|_2^2 = \sum_i {x'_i}^2 (1-\omega\,\bar\sigma_i^2)^{2k} \tag{5.13a}$$

$$\mathrm{E}\big[\|\Delta\bar{\mathbf{x}}_{pert}\|_2^2\big] = \bar\sigma_{\boldsymbol\epsilon}^2 \sum_i \Big(\frac{1-(1-\omega\,\bar\sigma_i^2)^k}{\bar\sigma_i}\Big)^2 \tag{5.13b}$$

with x'_i now as in eq. (4.76).

Their dependence on the iteration number k is sketched in fig. 5.4. Clearly, small singular values have more impact on both errors than larger singular values (remember that $0 \le \bar\sigma \le 1$). But while the contribution to the regularisation error decreases in the course of the iteration (it has to vanish for $k \to \infty$, i.e. no regularisation), the perturbation error is monotonically increasing. The perturbation error depends only on $\bar\sigma$, i.e. the measuring system, while calculation of the regularisation error requires the reconstructed state or the unperturbed column densities. Eqs. (5.12,5.13) are the quantitative explanation for the semiconvergence of iterative algorithms in the presence of noise that was mentioned in sec. 4.8.2.

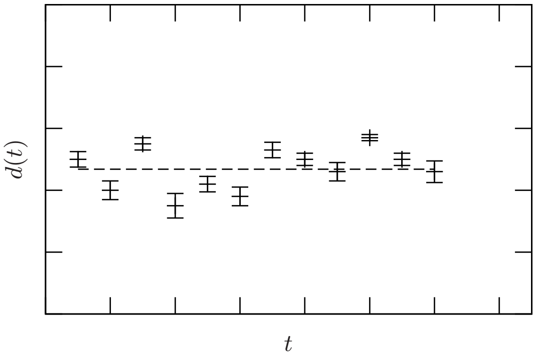


Figure 5.5: Time series of a column density showing stochastic variations that cannot be explained by the measurement error. The deviation around the mean can be used as an estimate for the error Δd associated with the source of this error.

Errors if measurements do not refer to the same point or period of time

If not all light paths finally used for reconstruction measure at the same time, this will introduce an error if the concentration distribution changes during the measurement interval. Examples of such non-simultaneous measurements are setups where the complete geometry is obtained by scanning the retro-reflectors in several steps (see secs. 3.4, 9.1) or aircraft and satellite measurements where the geometry arises through the motion of the detector.
Temporal changes of the concentration field can be due to

1. varying sources, sinks or chemical transformation,
2. turbulent fluctuations,
3. changing meteorological conditions,
4. transport.

Concentrating on LP-DOAS measurements using multibeam instruments (see sec. 3.4), the measurement cycle is of the order of minutes [*Mettendorf et al.*, 2006] – details depending, of course, on the number of steps necessary to generate the complete geometry, integration times etc. Variations in 1.-4. that are slow compared to this time scale can be neglected. If the variations are of stochastic nature, like turbulences or traffic emissions in an urban environment, they will partly cancel when averaging over the path. Otherwise, the error on the column densities related to them can be estimated from column density time series, if all other factors remain stable (fig. 5.5). Errors due to changing meteorology are best avoided by selecting stable periods.

For the tomographic reconstruction of concentration puffs that will be studied in part III changes of the concentration field caused by wind transport are especially troublesome. The spatial scales, given by $\Delta s = \bar{\mathbf{u}}\,\Delta t$ and thus of the order of several hundred metres for wind speeds $\bar{u}$ of a few metres per second and measurement times Δt of a couple of minutes, are comparable to the dispersion coefficients for travel distances of a few hundred metres (c.f. fig 2.3). To quantify the error on the column densities, we consider two light paths which, if measured at the same time, would yield the column densities

$$d_1(t) = \int\limits_{LP_1} ds\, c(\mathbf{r}, t),$$
$$d_2(t) = \int\limits_{LP_2} ds\, c(\mathbf{r}, t).$$

If the actual measurement for LP_2 takes place at $t + \Delta t$, its column density is two first order in Δt given by

$$d_2(t + \Delta t) \sim \int\limits_{LP_2} ds\, c(\mathbf{r}, t) + \Delta t \int\limits_{LP_2} ds\, \frac{\partial c}{\partial t}.$$

Assuming that the dominant contribution to $\frac{\partial c}{\partial t}$ is given by advection with constant wind vector, eq. (2.4) gives $\frac{\partial c}{\partial t} = -\bar{\mathbf{u}}\nabla c = -\bar{u}\nabla_{\bar{\mathbf{u}}} c$, where $\nabla_{\bar{\mathbf{u}}}$ is the directional derivative along the wind direction. Writing the integrals as average quantities along the light path, one gets with $\Delta s = \bar{u}\Delta t$

$$d_2(t + \Delta t) \sim L\,\bar{c} - L\,\Delta s\,\overline{\nabla_{\bar{\mathbf{u}}} c}.$$

The second term can be treated as small error Δd, if

$$\Delta s \frac{\overline{\nabla_{\bar{\mathbf{u}}} c}}{\bar{c}} \ll 1,$$

or, in other words, if the relative average change of c along $\bar{\mathbf{u}}$ over the distance Δs is much smaller than one.

5.3. Numerical estimation of the reconstruction error

Even if the discretisation error can be neglected, an estimation of $\Delta\hat{c}(\mathbf{r})$ according to eq. (5.2) involves assumptions about the unknown true state $\mathbf{x}_{id} = \mathbf{x}_{true}$. While the statistical impact of noise can easily be estimated either analytically by eqs. (5.11) or by Monte-Carlo simulations, the calculation of the inversion error is often reduced to a sensitivity analysis of the form: If my concentration profile is like this, what does my retrieval look like?

Here, we refer to a statistical estimate of $\Delta\hat{c}(\mathbf{r})$ – including the discretisation error – based on a sufficient number of simulated reconstructions for concentration fields $c_I(\mathbf{r})$, $I = 1, 2, \ldots, N$, from a suitable set (*ensemble*)

$$\mathcal{E} = \{\text{admissible } c_I(\mathbf{r}),\ I = 1, 2, \ldots, N\}. \tag{5.14a}$$

If the c_I are very similar, the sign of the individual reconstruction errors $\Delta\hat{c}_I(\mathbf{r}) = c_I(\mathbf{r}) - \hat{c}_I(\mathbf{r})$ will behave similar within the reconstruction area, too, and it makes sense to take the ensemble mean $\langle\Delta\hat{c}_I(\mathbf{r})\rangle_{\mathcal{E}}$. Otherwise one could use a positive quantity like $\langle\sqrt{(\Delta\hat{c}_I(\mathbf{r}))^2}\rangle_{\mathcal{E}}$, but then all information about systematic over- or underestimation within the area gets lost. Instead of the mean reconstruction error used by *Laepple et al.* [2004], we consider the error fields $\Delta_+\hat{c}(\mathbf{r})$ and $\Delta_-\hat{c}(\mathbf{r})$ as measures of under- and overestimation, respectively, defined by

$$\Delta\hat{c}(\mathbf{r}) = \Delta_+\hat{c}(\mathbf{r}) + \Delta_-\hat{c}(\mathbf{r}), \quad \text{where } \Delta_\pm\hat{c}(\mathbf{r}) = \begin{cases} \Delta\hat{c}(\mathbf{r}) & \text{if } \Delta\hat{c}(\mathbf{r}) \gtrless 0, \\ 0 & \text{else} \end{cases}. \tag{5.14b}$$

Without measurement errors one gets thus the scheme:

$$\begin{array}{llllr}
c_1(\mathbf{r}) \in \mathcal{E} \xrightarrow[\mathbf{d}_i = \int_i ds\, c(\mathbf{r})]{\text{forw. model}} & \mathbf{d}_1 \xrightarrow{\text{reconstruct}} & \hat{c}_1(\mathbf{r}) \Rightarrow & & \Delta\hat{c}_1(\mathbf{r}) = c_1(\mathbf{r}) - \hat{c}_1(\mathbf{r}), \\
c_2(\mathbf{r}) \in \mathcal{E} \longrightarrow & \mathbf{d}_2 \longrightarrow & \hat{c}_2(\mathbf{r}) \Rightarrow & & \Delta\hat{c}_2(\mathbf{r}) = c_2(\mathbf{r}) - \hat{c}_2(\mathbf{r}), \\
\vdots & \vdots & \vdots & & \vdots \\
\hline
 & & \Rightarrow & & \Delta\hat{c}(\mathbf{r}) \sim \langle \Delta\hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}}, \\
 & & \text{or} & & \Delta\hat{c}_{u/l}(\mathbf{r}) \sim \langle \Delta_\pm\hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}},
\end{array} \tag{5.14c}$$

where $\hat{c}_{l,u}(\mathbf{r})$ are the lower and upper bound of the reconstruction error as in eq. (5.1). In principle, an ensemble of smooth, very similar random distributions can give rise to the same mean reconstruction errors as an ensemble of highly fluctuating distributions (a less dramatic example will occur in chap. 9, sec. 9.3). But the standard deviation will be much higher in the latter case. Therefore, we define the lower and upper bounds of the reconstruction error as

$$\Delta\hat{c}_{u/l}(\mathbf{r}) = \langle \Delta_\pm\hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}} \pm \mathrm{std}_{\mathcal{E}}[\Delta_\pm\hat{c}_I(\mathbf{r})]. \tag{5.14d}$$

For simulated measurements, where the error free column densities $\mathbf{d}$ are known, the statistic contribution of an unbiased measurement error to the reconstruction error fields $\Delta_\pm\hat{c}$ is given by the standard deviation $\mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}(\mathbf{r})]$, with $\hat{c}$ retrieved from $\mathbf{d}$. For linear reconstruction the standard deviation is given by eqs. (5.11). The error bounds eq. (5.14d) take now the form

$$\Delta\hat{c}_{u/l} = \langle \Delta_\pm\hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}} \pm \mathrm{std}_{\mathcal{E}}[\Delta_\pm\hat{c}_I(\mathbf{r})] \pm \mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}(\mathbf{r})]. \tag{5.14e}$$

For actual measurements, where only the error afflicted column densities $\mathbf{d}_{\boldsymbol{\epsilon}}$ are known, one way to get an estimate for the impact on the unknown column densities $\mathbf{d}$ is to calculate the mean impact within the ensemble $\mathcal{E}$ by replacing the above simulation step with

$$\begin{array}{rl}
c_I(\mathbf{r}) \in \mathcal{E} \longrightarrow \mathbf{d}_I \xrightarrow{+\boldsymbol{\epsilon}} \hat{c}_I(\mathbf{r}) \overset{\overline{(\cdot)}_{\boldsymbol{\epsilon}}}{\Longrightarrow} & \overline{\Delta\hat{c}_I(\mathbf{r})}_{\boldsymbol{\epsilon}} \pm \mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}_I(\mathbf{r})] \\
\text{or} & \overline{\Delta_\pm\hat{c}_I(\mathbf{r})}_{\boldsymbol{\epsilon}} \pm \mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}_I(\mathbf{r})]
\end{array} \tag{5.14f}$$

to get the average

$$\Delta\hat{c}_\pm(\mathbf{r}) \sim \langle \overline{\Delta\hat{c}_{\pm I}(\mathbf{r})}_{\boldsymbol{\epsilon}} \rangle_{\mathcal{E}} \pm \langle \mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}_I(\mathbf{r})] \rangle_{\mathcal{E}}. \tag{5.14g}$$

This approach was followed by *Laepple et al.* [2004]. But unless it can be shown that the synthetic column densities obtained from the c_I are a better estimate of $\mathbf{d}$, there seems little reason to prefer them to the measured column densities $\mathbf{d}_{\boldsymbol{\epsilon}}$ for the estimation, especially for small errors. Besides, the ensemble estimate is computationally much more expensive if the standard deviations cannot be calculated analytically. In the following, the approach resulting in the bounds eq. (5.14e) will be taken.

Strictly speaking only distributions consistent with the actually measured column densities $\mathbf{d} = (d_1, \ldots, d_m)^T$ should be taken into account, i.e.

$$\mathcal{E} = \{\text{admissible } c_I(\mathbf{r}) \,\Big|\, |d_i - \int_{LP_i} ds\, c_I| \leq \sigma_{\epsilon i}, \quad i = 1, \ldots, m\}, \tag{5.15}$$

but this is rather involved in practice. Instead it is assumed in the following that the ensemble means $\langle \Delta_{(\pm)} \hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}}$ for suitably chosen $\mathcal{E}$ are good estimates for the reconstruction error of the concentration field $\hat{c}$ retrieved from experiment, without demanding consistency of the column densities. The test concentrations c_I can be provided by model calculations, be based on the assumption that the reconstruction is correct within certain bounds (see appendix B and also sec. 7.4) etc. One should notice that, in general, the ensemble will be biased in the sense that

$$\langle \hat{c}_I(\mathbf{r}) \rangle_{\mathcal{E}} \neq \hat{c}(\mathbf{r}). \tag{5.16}$$

5.4. Overall reconstruction errors and quality criteria

The spatial dependency of the reconstruction error is not always of interest. A typical situation where the error field is an inconvenient quantity is the following. You want to find optimal parameters of the reconstruction procedure for a possibly large number of test concentration fields which hopefully come close to the unknown real one. These parameters are then used for the actual reconstruction from the measurement data (see sec. 6.4). For the processing of larger amounts of data you therefore want a simple criterion, best a number Q, telling that if for two sets of parameters p_1 and p_2 $Q_1 > Q_2$ then reconstruction with p_1 is better (or worse) than with p_2. Denoting the true (simulated) quantities as c, $\mathbf{d}$, the reconstructed ones as $\hat{c}$, $\hat{\mathbf{x}}$ and defining $\hat{\mathbf{d}} = A\hat{\mathbf{x}}$, then two cases can be distinguished.

(*i*) *Quality criteria based on the measurements, i.e.* $Q = Q(\mathbf{d}, \hat{\mathbf{d}})$:
The most simple number constructed from $\mathbf{d}$, $\hat{\mathbf{d}}$ is the *data residual*, c.f. eq. (4.65), i.e. the overall agreement between measured and reconstructed data

$$\|\Delta \mathbf{d}\|_2 = \|\mathbf{d} - \hat{\mathbf{d}}\|_2 \tag{5.17}$$

or the relative data residuum if it is normed by the column densities. Another quantity that has the advantage of being normalised is the correlation coefficient between the d_i and $\hat{d}_i$ (see below). However from the discussion so far it follows:

- In the under-determined case of an ill-posed problem, quality criteria based on the agreement between measured (simulated) and reconstructed path integrals are not decisive as the measured data can be reproduced by (in general) infinitely many solutions. A regularised solution with higher residual can easily be better than an unregularised solution with vanishing residual. In fact, for a measurement error $\boldsymbol{\epsilon}$ it does not really make sense to strive for a residuum smaller than $\|\boldsymbol{\epsilon}\|_2$.
- If the solution of the over-determined problem involves $\|\mathbf{d} - A\hat{\mathbf{x}}\|_2 = \|\mathbf{d} - \hat{\mathbf{d}}\|_2$ in the cost

> function, as it does in the least squares fit, the residual can be regarded as a measure for the quality of the solution – provided that the measurement errors are not too large. If they are, small residual means fitting to noise. For the regularised solution the data residual is in general not a useful indicator. We refer here to the discussion of the choice of the regularisation parameter in [*Twomey*, 1997] and [*Hansen*, 1998, chap. 7]. In any case, forcing the data residuum to be smaller than $\|\boldsymbol{\epsilon}\|_2$ is questionable.

To sum up, quality criteria based on data agreement possess the attractive feature that they do not rely on the unknown concentration field. But for the (linear) ill-posed inverse problem they are of no use in the under-determined case. In the over-determined case their significance depends on kind and degree of regularisation.

(ii) Quality criteria based on unknown concentrations, i.e. $Q = Q(c, \hat{c})$ *or* $Q(\mathbf{x}, \hat{\mathbf{x}})$:
We consider criteria for the continuous $c(\mathbf{r})$, $\hat{c}(\mathbf{r})$ first. The average reconstruction error $\mathcal{V}^{-1} \int_\Omega d\mathcal{V} \Delta\hat{c}(\mathbf{r})$ is not suitable because over- and underestimation cancel each other within the area or volume Ω. Instead one could use $\int_\Omega d\mathcal{V} |\Delta\hat{c}(\mathbf{r})|$ or $\int_\Omega d\mathcal{V} [\Delta\hat{c}(\mathbf{r})]^2 = \|\Delta\hat{c}(\mathbf{r})\|_2^2$. The integrated square of the reconstruction error is mathematically more appealing. Especially

$$\mathrm{E}\left[\|\Delta\hat{c}(\mathbf{r})\|_2^2\right] = \|\Delta c_{disc}(\mathbf{r})\|_2^2 + \|\Delta\hat{c}_{inv}(\mathbf{r})\|_2^2 + \|\Delta\hat{c}_{reg}(\mathbf{r})\|_2^2 + \mathrm{E}\left[\|\Delta\hat{c}_{pert}(\mathbf{r})\|_2^2\right]. \tag{5.18}$$

This holds *only* if the discretisation error is defined as in eqs. (5.5). We note in passing that if $S_\mathbf{x}$ as in eq. (5.11b) is known, it follows immediately that

$$\mathrm{E}\left[\|\Delta\hat{c}_{\boldsymbol{\epsilon}}(\mathbf{r})\|_2^2\right] = \mathrm{tr}\left[S_\mathbf{x}\, \Phi\right].$$

These equations imply that the different contributions to the overall error defined by the 2-norm can be calculated individually and added up later. And it means that one does not have to vary all parameters at once, but can vary e.g. parameters for the discretisation first, then vary measurement errors and so forth. The quasi normalised key figure *nearness*, built from $\|\Delta\hat{c}(\mathbf{r})\|_2$, was introduced by [*Herman et al.*, 1973; *Herman and Rowland*, 1973] into tomographic image reconstruction:

$$NEARN = \|c(\mathbf{r}) - \bar{c}\|_2^{-1}\, \|\Delta\hat{c}(\mathbf{r})\|_2, \quad \text{with} \quad \bar{c} = \mathcal{V}^{-1} \int_\Omega d\mathcal{V}\, c(\mathbf{r}). \tag{5.19}$$

If the reconstruction $\hat{c}(\mathbf{r}) = [\boldsymbol{\phi}(\mathbf{r})]^T \hat{\mathbf{x}}$ equals the spatial mean $\bar{c}$ of the unknown concentration field (i.e. $\hat{\mathbf{x}} = \mathbf{const}$) then $NEARN = 1$. Vice versa, if $NEARN = 1$ then the retrieved field has the same overall reconstruction error in terms of nearness as the spatial mean of $c(\mathbf{r})$ would have.

If the real and reconstructed concentrations are compared not on the entire area or volume $\mathcal{V}$ but on a subset of points $\{\mathbf{r}_i\}_{i=1}^N$, then evaluating the agreement between the real concentrations $c_i = c(\mathbf{r}_i)$ and the reconstructed $\hat{c}_i = \hat{c}(\mathbf{r}_i)$ corresponds to the problem of atmospheric model evaluation from N measurements or from simulations for a fixed time. An exceeding number of measures (also called *indices* or *metrics*) is used for this purpose, depending on whether the spatial evaluation concerns meteorological parameters, chemical transport or air pollution models. In the context of dispersion modelling a set of mostly statistical and relative[2] measures suggested by *Hanna* [1989] frequently

---

[2] Air quality assessment usually uses relative numbers to evaluate model performance while for meteorological models absolute numbers are prefered. The reason being that the significance of a relative error can be very different for

metric	Q	analytical or statistical expression	range	best
overall error				
nearness	$NEARN$	$= \left(\sigma_c^{-2}\ \overline{(\Delta\hat{c})^2}\right)^{1/2}$	$0 \leq NEARN$	0
normalised mean square error	$NMSE$	$= (\bar{c}\,\bar{\hat{c}})^{-1}\ \overline{(\Delta\hat{c})^2}$	$0 \leq NMSE$	0
index of agreement	IOA	$= 1 - \overline{(\lvert c-\bar{c}\rvert + \lvert\hat{c}-\bar{\hat{c}}\rvert)^2}^{\,-1}\ \overline{(\Delta\hat{c})^2}$	$0 \leq IOA \leq 1$	0
normalised mean absolute error	$NMAE$	$= \bar{c}^{\,-1}\ \overline{\lvert\Delta\hat{c}\rvert}$	$0 \leq NMAE$	0
bias				
normalised mean bias	NMB	$= \bar{c}^{\,-1}\ \overline{\Delta\hat{c}}$	$0 \leq NMB$	0
fractional bias	FB	$= \frac{1}{2}(\bar{\hat{c}} + \bar{c})^{-1}\ \overline{\Delta\hat{c}}$	$-2 \leq FB \leq 2$	0
correlation	r	$= (\sigma_c\,\sigma_{\hat{c}})^{-1}\ \overline{(c-\bar{c})(\hat{c}-\bar{\hat{c}})}$	$-1 \leq r \leq 1$	1
other				
fraction of factor 2	$FA2$	= %-age of data with $\frac{1}{2} \leq \frac{\hat{c}}{c} \leq 2$	$0 \leq FA2 \leq 100\%$	100%
peak prediction accuracy	PPA	$= c_{max}^{-1}\ \hat{c}_{max}$	$-\infty \leq PPA \leq \infty$	1
peak integral prediction accuracy	$PIPA$	$= (\int_{\mathcal{P}eak} c)^{-1}\ \int_{\mathcal{P}eak} \hat{c}$	$-\infty \leq PIPA \leq \infty$	1

Table 5.1: Quality measures considered in this thesis. For evaluation of the tomographic reconstruction c is the true, i.e. simulated field, $\hat{c}$ the reconstructed. For model evaluation c refers to measured concentrations and $\hat{c}$ to the modelled ones. The mean $\overline{(\cdot)}$ can be understood as $\overline{(\cdot)} = \mathcal{V}^{-1} \int_\Omega d\mathcal{V}\ (\cdot)(\mathbf{r})$ or $N^{-1} \sum_i (\cdot)_i$, both being identical for $N \to \infty$. In the same sense $\sigma^2_{(\cdot)} = \overline{\left((\cdot) - \overline{(\cdot)}\right)^2}$. c_{max}, $\hat{c}_{max}$ are the absolute maximum values of $c(\mathbf{r})$, $\hat{c}(\mathbf{r})$ within Ω. $\int_{\mathcal{P}eak}$ is the integral within the area of a peak (eq. (5.20)) so that $PIPA$ measures the precision of reconstructed total emissions (see text).

appears in the literature, some of which are listed in table 5.1 and will be used later. The *normalised mean square* and *absolute error* $NMSE$ and $NMAE$ measure the overall discrepancy between the measured or simulated concentrations c_i and the values $\hat{c}_i$ predicted by the model. *Normalised mean* and *fractional bias* NMB and FB indicate how much the model average concentration value differs from the observed mean. r is the common (Pearson) *correlation coefficient* so that a value of 1 means perfect correlation, -1 perfect anticorrelation and a value of 0 no correlation. $FA2$ is self-explanatory. The *index of agreement* IOA was defined by *Willmott* [1981] to include the variances between model and measurement and between model, measurements and the model mean in one quantity. In contrast to r it is sensitive to model and observed means. Values of 0 mean perfect agreement, 1 poor agreement. The non-statistic *peak prediction accuracy* PPA refers to the (unpaired) absolute maxima of original and reconstructed distributions within Ω.

For the 2-D simulations in part III one further index is introduced. Thinking of concentration peaks as emission plumes or puffs, a measure for the precision of reconstructed total emissions suggests itself. The *peak integral prediction accuracy* $PIPA$

$$PIPA = \frac{\int_{\mathcal{P}eak} dA\ \hat{c}(\mathbf{r})}{\int_{\mathcal{P}eak} dA\ c(\mathbf{r})} \tag{5.20}$$

---

different meteorological parameters. Just think of a 10% error on a wind speed of $10\,m/s$ ($1\,m/s$) versus the same relative error on a temperature of $300\,K$ (namely $30\,K$!).

compares the concentration integrated over the area of the reconstructed peak with the one of the original. For the evaluation of the integrals in the case of Gaussian peaks, $3\times\max(\sigma_x, \sigma_y)$ is taken as the radius of the original peak distribution (at $3\,\sigma$ the peak value has fallen to 1%). For the reconstructed peak, one pixel length of the reconstruction grid is added to this radius to take account of the spreading within the pixels for the box basis and the interpolation between neighbouring grid points for the linear basis functions, respectively. This, especially in the case of several large distributions with overlap and for larger pixel sizes, will only be a rough estimate. If the concentration peak originates from point source emission, then by writing the emission rate $\dot{q}$ (in units $kg \cdot s^{-1}$) like

$$\dot{q} = \oint d\mathbf{A}\, \mathbf{u}\, c(\mathbf{r}) \sim u_\perp \int_{\mathcal{P}lume} dA\; c(\mathbf{r}),$$

it can be seen that $PIPA = \dot{\hat{q}}/\dot{q} = \hat{q}/q$ provides a measure for the precision of emission rates or total emissions (if the wind speed has a component perpendicular to the reconstruction area). For horizontal cuts through the atmosphere this is the case, e.g., if the area contains a source with vertical transport. For horizontal transport of a plume this number can be related to total emissions only with further assumptions on the vertical dispersion since its release.

Finally, the case $Q = Q(\mathbf{x}, \hat{\mathbf{x}})$ can be obtained by choosing the sample points $\mathbf{r}_i$ such that they coincide with the boxes for the piecewise constant discretisation or the grid nodes for the linear discretisation, i.e. $c_i = x_i$, $\hat{c}_i = \hat{x}_i$. We conclude this chapter by some remarks

- Evidently, one measure is not enough to grasp the continuous field $\Delta\hat{c}(\mathbf{r})$ completely. Therefore, depending on the application a variety of more or less specific criteria exists apart from the ones in table 5.1. For higher resolved fields far more refined evaluation methods can be applied, especially for time dependent analysis.

- Quality criteria can differ largely in their degree of sensitivity to error patterns. For example, the correlation coefficient is insensitive to an overall offset between measured and modelled/reconstructed values, no matter how large this offset is, whereas it would give a large contribution to, e.g., $NMSE$.

- Above all, the significance of most quality measures is a relative one, because its rough size largely depends on the concentration fields. For example, the best possible modelling or reconstruction of a sharp concentration peak will in general have larger r values than a poorly modelled or reconstructed smooth background concentration.

6. Reconstruction Procedure

The focus of this work with respect to the reconstruction method lies on improved parametrisation of the discrete inverse problem as well as on the choice and implementation of a priori assumptions as they might typically occur in atmospheric applications and not on algorithm development for the inversion of the discrete system of equations itself. Therefore, after a brief summary of the reconstruction method as far as it was described in chap. 4 and has been applied before in the theoretic study [*Fleming*, 1982] for satellite tomography of trace gases, by *Todd and Ramachandran* [1994*a*] for indoor gas concentrations[1] and by *Laepple et al.* [2004] for the reconstruction of a motorway emission plume in sec. 6.1, sec. 6.2 presents generalisations of the so-called grid translation method suggested by *Verkruysse and Todd* [2004] that relies on reconstructions from several grids. The subsequent sec. 6.3 introduces a priori constructions for the (under-determined) least-squares problem based on the kind of concentration distributions discussed in chap. 2. After a discussion of additional linear (in)equality constraints, the grid translation scheme is revisited in sec. 6.3.4 for peak concentrations to introduce the maximum reconstructed value as constraint. Finally, sec. 6.4 reviews all variables of the reconstruction method as it will be used in part III and how they are optimised through simulations.

6.1. Reconstruction principle and inversion algorithm

It was argued in sec. 4.2 that long-path DOAS tomography is not fit for the application of transform methods like the filtered back projection (sec. 4.2.1). The alternative formulation in terms of a finite number of parameters can lead to a linear or nonlinear problem. The former allows powerful concepts from linear inverse theory like the resolution matrix (averaging kernels) and simpler discussion of reconstruction errors (see secs. 5.2.2, 5.2.3), the latter in general requires global optimisation algorithms, except for special cases where linearisation is possible or where particular algorithms exist (c.f. sec. 4.9.1). For these reasons I have chosen the parametrisation by local basis functions described in sec. 4.3.2 for this first systematic study of the possibility to reconstruct various forms of atmospheric trace gas distributions from realistic long-path DOAS tomographic measurements. This does not mean that for certain applications special schemes like the fit of Gaussian distributions (SBFM, sec. 4.9.3) may not lead to better results, but we want to keep the discussion general and among other drawbacks, the last example introduces unavoidable functional a priori that is not always appropriate (for example not in the case of the motor way emission plume in [*Laepple et al.*, 2004][2], see also fig. 5.3a, p. 88).

The resulting linear discrete ill-posed problem is replaced by a least-squares least-norm principle as in secs. 4.6.1, 4.6.2. Demanding a minimum norm of the solution is mathematically appealing. Alternatively, often the weighted sum of the residual and the norm of the solution or discrete approximations

---

[1] Both only with piecewise constant basis functions.

[2] Private communication with T. Laepple.

of its gradients or higher derivatives serves to regularise (Tikhonov regularisation, sec. 4.6.3). It can thus be used as a selection criterion to smooth the solution for the right choice of the ad hoc weight. For the applications in part III neither of these reasons appeared to be critical so that this approach was not considered here.[3]

The least-squares least-norm problem plus nonnegativity constraint is solved iteratively as discussed in secs. 4.8.1, 4.8.3 by row acting methods (sec. 4.8.4), where only ART and SIRT will be considered as representatives for the sequential and simultaneous algorithms. The simultaneous SART was found to be less appropriate for the reconstruction of Gaussian peaks in [*Todd and Ramachandran*, 1994*a*]. The implicit weighting of these algorithms according to eq. (4.74) will not be corrected and ignored in the subsequent discussion because for ART and SIRT it only acts on the norm of $\mathbf{d} - A\mathbf{x}$ and in the under-determined case, which is the most relevant in the following, this kind of weighting is not or less effective. Furthermore, the relaxation parameter ω, usually used to speed up the convergence for large systems of equations, will not be varied and set to one as the systems in the following are rather small. Regularisation, if necessary, is carried out iteratively, as discussed in sec. 5.2.3.

6.2. Reconstruction grid

The development of the reconstruction procedure in chap. 4 tacitly assumed that the parameters of the discretisation are the same as the ones entering the least-squares fit, in other words, that the discretisation grid coincides with the reconstruction grid. For the over-determined case it is easy to think of schemes for which this is not the case, especially if the discretisation is carried out by quadrature as sketched in sec. 4.3.1 (see [*Doicu et al.*, 2004] for a 1-D example). In this work they are identical.

Piecewise constant and linear basis functions are constructed by tensor products of one dimensional functions (sec. 4.3.2) so that the supports of the piecewise basis functions are rectangular (*boxes*). For discretisation that is fine relative to the whole area this is not a problem, but for very coarse discretisation this may become critical, e.g., if the region defined by the light paths is triangular. *Reconstruction grid* means in the following the set of points consisting in the corner points of the boxes (which are the discretisation nodes in the case of the bi(tri)linear basis functions). The number of parameters, i.e. the number of boxes for the piecewise constant or the number of grid nodes in the case of the linear basis functions, is refered to as *grid dimension*. The region of interest, the *reconstruction area* or volume, is always such that it is defined by the outermost grid points.

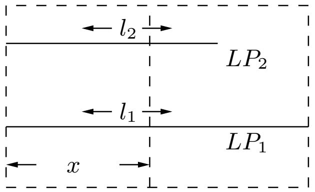


Figure 6.1: The ideal reconstruction grid. How has the box length x of the lower box to be chosen so that for the two light paths shown the singular values of A, and therefore their information content, become as equal as possible ?

6.2.1. The ideal reconstruction grid

We assume for the moment that no additional information enters the problem, i.e. there are as many integration paths as parameters to be determined: $m = n$. The ideal parametrisation of $c(\mathbf{r})$ by the basis functions would be such that A becomes diagonal with entries of the same size. Then these coincide with the singular values which are equal to the roots of the eigenvalues of A^TA or AA^T, see eqs. (4.21), and thus carry all the same information content according to eq. (4.61). For typical light path geometries as in fig. 3.5, p. 44, and basis functions with rectangular support diagonal A is in general not feasible, but one can try to choose the grid such that singular values do not become too different. To see how this can be accomplished Gerschgorin's theorem[4] is applied to the eigenvalue system of A^TA to give the following bounds on any singular value σ

$$\sum_i a_{ij}^2 \le \sigma^2 \le \sum_i a_{ij}^2 + \sum_{j' \neq j} \sum_i a_{ij} a_{ij'} \tag{6.1a}$$

$$\vee \quad \sum_i a_{ij}^2 - \sum_{j' \neq j} \sum_i a_{ij} a_{ij'} \le \sigma^2 \le \sum_i a_{ij}^2 \tag{6.1b}$$

for some j. For the box basis $\sum_i a_{ij}^2$ measures the total square length of all light paths added up in box j. $\sum_{j' \neq j} \sum_i a_{ij} a_{ij'}$ is basically the overlap of the light paths. As $\mathrm{tr}[AA^T] = \sum_{i,j} a_{ij}^2 = \sum_j \sigma_j^2$ eqs. (6.1) imply that the grid has to be chosen such that the lengths of the light paths within the boxes have to be similar and that the overlap has to be small compared to it. Conversely, a singular value can become small if a light path's length goes to zero (eq. (6.1a)) or, for arbitrary length, if the grid is chosen in a way that results in large overlap with other light paths (eq. (6.1b)). To study the question of the right grid quantitatively, we consider the following simple

Example: $A = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ with box basis functions and light paths of lengths l_1, l_2 as in fig. 6.1, i.e. $a_{11} = x$, $a_{12} = l_1 - x$ and $a_{21} = x$, $a_{22} = l_2 - x$ if $x \le l_2$, $a_{21} = l_2$, $a_{22} = 0$ otherwise. What value x should the box length have to minimise the difference between the two singular values σ_1, σ_2? The singular values are

$$\sigma_{1/2}^2 = \tfrac{1}{2}\mathrm{tr}[AA^T] \pm \tfrac{1}{2}\sqrt{\mathrm{tr}[AA^T]^2 - 4\det[A]^2}. \tag{6.2}$$

The general expression for x minimising the root is quite lengthy, so I give three examples instead for $l_1 = 2\,a.u.$ and the following values for l_2:

---

[3] *Price et al.* [2001] have applied the Tikhonov approach minimising the discrete approximation of the third derivative to the reconstruction of indoor gas concentrations. The method hardly needs the nonnegativity constraint and is therefore linear. Unfortunately, this study is not systematic with respect to the shape of the distribution and does not compare its method to other solutions like the minimum norm solution.

[4] Gerschgorin's theorem sais that for each eigenvalue λ of a matrix M there is an i such that

$$|\lambda - m_{ii}| \le \sum_{i' \neq i} |m_{ii'}|.$$

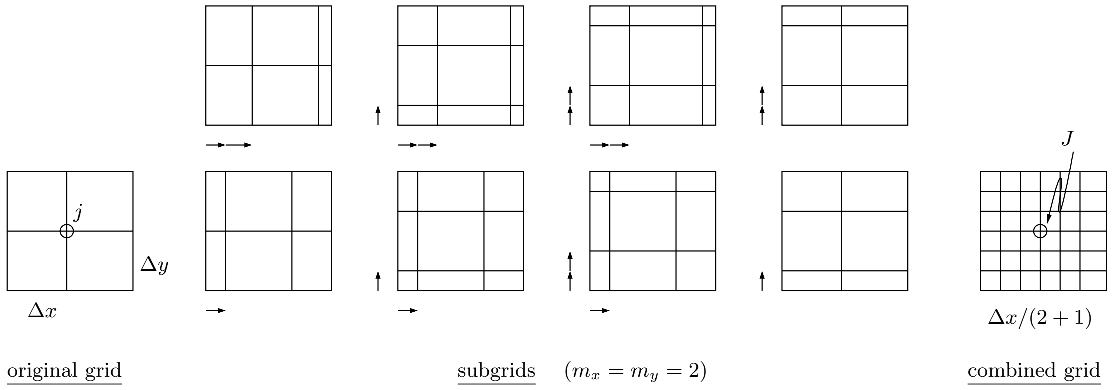


Figure 6.2: 2-D grid translation scheme proposed by Verkruysse and Todd [2004] for a rectangular reconstruction area and grid, depicted here for a square area with identical spacing in x and y of the grid which is shifted twice in both directions. The dimension of the final composite grid is $N = [3{\cdot}2{+}1]^2 = 49$. Mind that the subgrids have different dimensions. Subgrids with small boxes near the boundaries cause smaller singular values according to sec. 6.2.1.

l_2 $[a.u.]$	x $[a.u.]$	σ_2 $[a.u.]$	σ_2 $[a.u.]$
0.5	0.877	1.460	0.385
1	0.909	1.589	0.572
1.5	0.911	1.590	0.573

This may be not what one expects and illustrates the intricate interplay between the box-lengths of the light paths (the trace term in eq. (6.2)) and their overlap (which leads to small determinants in eq. (6.2)). For complex irregular light path geometries the design of an ideal grid from the singular values' point of view will become quite complicated and a constructive method needs further thought. But one should be aware of the fact that the reconstruction method finally will not use the exact singular values of A, if a priori (or constraints etc.) enters the problem.

6.2.2. Combining grids for the reconstruction

While the type of basis functions ϕ_j and the location of the grid nodes may be only of minor importance for very fine discretisations, the shape of the reconstruction is mainly given by the $\phi_j(\mathbf{r})$ if $c(\mathbf{r})$ is represented only by a few basis functions. In this case, especially if the light path geometry is irregular, the exact position of the grid nodes is to a large extent arbitrary, which suggests to take into account representations based on different grids (provided that these are equally sensible).

Motivated by the fact that a reconstruction underestimates the maximum value of a sharp peak if the position of the original peak lies unfavourable relative to the boxes of the reconstruction grid, *Verkruysse and Todd* [2004] suggested to use several grids shifted against each other in the plane as shown in fig. 6.2. The reconstruction result is then defined on the grid that arises from putting these subgrids on top of each other and attributing to a box of this higher resolved grid the concentration that corresponds to the 'nearest' of the boxes among all subgrids (nearest is defined by the grid lines and if there are several equally near, the average concentration is taken). This 'grid translation' scheme was defined for box basis functions and a regular original grid that was shifted m_x times in x-direction and m_y times in y-direction like this: The first grid is generated by shifting the original nodes by $\Delta x/(m_x+1)$ in x-direction, where Δx is the x-distance between two nodes of the original grid. This is performed m_x times, then the initial grid is shifted by $\Delta y/(m_y+1)$ in y and the procedure is repeated, and so forth.

The idea was adopted in [*Hartl et al.*, 2006] for bilinear basis functions. The reconstruction area is assumed to be rectangular, as the applications in part III are all on rectangular areas. The shape of the reconstruction area is not essential but, especially for the bilinear basis, the design of regular shifted grids for irregular areas becomes difficult or impossible unless the number of basis functions is increased significantly. As before, the final, higher resolved grid is given by the union of the nodes from all subgrids and the parameter vector $\mathbf{X} = (X_1, \ldots, X_N)^T$ for this grid is defined in two different ways:
(i) In a *composite* scheme the component X_J for grid node $\mathbf{r}_J$ is taken from the Ith subgrid that has a coinciding node there:

$$X_J = \hat{c}^I(\mathbf{r}_J). \tag{6.3}$$

The boundary grid lines do not change and here the average of coinciding nodes is taken. This scheme corresponds to the suggestion by *Verkruysse and Todd* [2004] for the box basis.
(ii) In the *avaraging* scheme X_J is taken as the average of all subgrid reconstructions evaluated on grid node $\mathbf{r}_J$:

$$X_J = \frac{1}{M}\sum_{I=1}^{M} \hat{c}^I(\mathbf{r}_J), \tag{6.4}$$

where M is the number of subgrids. This scheme is completely linear.
If the subgrids are generated by the translation of an original grid as sketched above and shown in fig. 6.2, this results in M reconstructions for the subgrids and a total of N nodes for the composite final grid given by

$$M = m_x m_y + m_x + m_y + 1 \tag{6.5a}$$

$$N = [(m_x+1)(n_x-1)+1][(m_y+1)(n_y-1)+1], \tag{6.5b}$$

where n_x and n_y are the numbers of nodes in x- and y-direction. Both schemes will be discussed in detail in 8.2.3.

6.3. A priori and constraints

A priori knowledge is any information available prior to the reconstruction, in terms of probabilities, equalities, inequalities etc. Here – as often done in the literature – we refer to a priori as an initial guess $\mathbf{x}_a$ of the parameter vector $\mathbf{x}$ with or without additional information about its uncertainty. Such an a priori is always present for an under-determined least-squares problem and corresponds to the iteration start $\mathbf{x}^{(0)}$ for the iterative row acting algorithms discussed in sec. 4.8.4. We are now going to specify it for the reconstruction of atmospheric concentration distributions in part III.

6.3.1. Choice of the a priori

According to eqs. (4.56) the contribution of the a priori to the least-squares least-norm solution is given by $(\mathbb{1}_n - R)\mathbf{x}_a$, where R is the resolution matrix of the linear inversion.[5] We assume throughout the following an a priori that is completely noncommittal with respect to the spatial distribution, i.e.

$$\mathbf{x}_a = c_a(1,\dots,1)^T \quad \Rightarrow \quad [(\mathbb{1}_n - R)\mathbf{x}_a]_j = c_a(1 - \sum_{j'} R_{jj'}). \tag{6.6}$$

Following the discussion in chapter 2, two scenarios for the concentration c_a are distinguished.

(i) *Locally enhanced (or reduced) concentrations* (e.g. emission puffs or plumes) where the concentration peaks or sinks are on confined areas within the total reconstructed region. A typical example for a “negative peak” would be local ozone depletion due to strong emission of NO. c_a is set to zero if the background concentration of the trace gas of interest is negligible. The case of substantial background concentrations is addressed in the next section.

(ii) *Smooth concentration distributions* without significant peaks. In this case an average concentration is taken as a priori that is obtained from the measurements as

$$c_a = \frac{1}{m}\sum_{i=1}^{m} \bar{c}_i = \frac{1}{m}\sum_{i=1}^{m} \frac{d_i}{L_i}, \tag{6.7}$$

where L_i is the length of LP_i.

Using the definitions of chap. 5, an incorrect a priori is related with the inversion error according to eq. (5.10). An estimated error Δc_a of c_a leads to a contribution

$$(\Delta\mathbf{x})_j = (1 - \sum_{j'} R_{jj'})\,\Delta c_a \tag{6.8}$$

to the inversion error.

6.3.2. Background concentrations

We refer to the peak scenario in sec. 6.3.1(i). Background is understood in a graphical sense as what remains of the distribution after subtracting the peak(s) and it does not have to agree with the natural background or any other atmospheric offset concentration. Taking into account the limited resolution of the tomographic experiments here, only constant background is considered

$$c_{BG}(\mathbf{r}) \equiv c_{BG}, \quad \text{i.e.} \quad \mathbf{x}_{BG} = c_{BG}(1,\dots,1)^T. \tag{6.9}$$

(i) *Fitting the background:*
Inserting

$$c(\mathbf{r}) = c'(\mathbf{r}) + c_{BG} \tag{6.10a}$$

---

[5] And in the unweighted case $P = (\mathbb{1}_n - R)$ is the projector onto the nullspace of A, see eq. (4.55).

into the forward model $d_i = \int_i ds\, c(\mathbf{r})$ leads to the discrete system

$$\begin{aligned} \mathbf{d} = B\,\mathbf{y}, \qquad \text{with} \qquad B &= (A \quad \boldsymbol{l}) && (6.10b) \\ \mathbf{y} &= (\mathbf{x}', \quad c_{BG})^T && (6.10c) \end{aligned}$$

and $\boldsymbol{l} = (L_1, \ldots, L_m)^T$ is the vector of the light path lengths.
In the over-determined case a least-squares fit with the additional parameter c_{BG} leads to (see appendix C)

$$\begin{aligned} \mathbf{x}' &= (A^T L\, A)^{-1} A^T L\, \mathbf{d} = A_L^{\dagger} \mathbf{d} \qquad (6.11) \\ c_{BG} &= \frac{\boldsymbol{l}^T}{\boldsymbol{l}^T \boldsymbol{l}} (\mathbb{1}_m - R_{\mathbf{L}})\, \mathbf{d}, \end{aligned}$$

where $L = \boldsymbol{l}^T \boldsymbol{l}\, \mathbb{1}_m - \boldsymbol{l}\boldsymbol{l}^T$ and $R_L = A\, A_L^{\dagger}$ is the resolution matrix of the weighted inverse in eq. (6.11). In the under-determined case the least-norm principle can be applied to the variation $\mathbf{x}'$ alone, i.e. to $\|\mathbf{x}'\|_2$. Minimisation of both variation and background can be obtained by $\|\mathbf{y}\|_2$, while $\sum_j (x'_j - c_{BG})^2$ minimises the deviations from the background. All three cases can be written as $\mathbf{y}^T H \mathbf{y}$ with suitable matrix H, and the solution of the corresponding least-norm problem

$$\min_{\mathbf{y}} \mathbf{y}^T H \mathbf{y}, \qquad \mathbf{d} = B\,\mathbf{y}$$

can be expressed as weighted least-norm solution for $\mathbf{x}'$ with a priori $\mathbf{x}'_a$ (c.f. appendix C). For the three special choices one gets

$$H = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} : \qquad \begin{aligned} \mathbf{x}' &= A^T (AA^T + \boldsymbol{l}\boldsymbol{l}^T)^{-1} \mathbf{d} \\ c_{BG} &= \boldsymbol{l}^T (AA^T + \boldsymbol{l}\boldsymbol{l}^T)^{-1} \mathbf{d}, \end{aligned} \qquad (6.12a)$$

$$H = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{pmatrix} : \qquad \begin{aligned} \mathbf{x}' &= A^T \left(\mathbb{1}_m - \tfrac{(AA^T)^{-1} \boldsymbol{l}\boldsymbol{l}^T}{\boldsymbol{l}^T (AA^T)^{-1} \boldsymbol{l}} \right) (AA^T)^{-1} \mathbf{d} \\ c_{BG} &= \tfrac{\boldsymbol{l}^T}{\boldsymbol{l}^T (AA^T)^{-1} \boldsymbol{l}} (AA^T)^{-1} \mathbf{d}, \end{aligned} \qquad (6.12b)$$

$$H = \begin{pmatrix} 1 & & & -1 \\ & \ddots & & \vdots \\ & & 1 & -1 \\ -1 & \ldots & 1 & n \end{pmatrix} : \qquad \begin{aligned} \mathbf{x}' &= \left(A^T - \tfrac{A^T (AA^T)^{-1} \boldsymbol{l}\boldsymbol{l}^T - \frac{1}{2} \boldsymbol{l}\boldsymbol{l}^T}{\boldsymbol{l}^T (AA^T)^{-1} \boldsymbol{l}} \right) (AA^T)^{-1} \mathbf{d} \\ c_{BG} &= \tfrac{1}{2} \tfrac{\boldsymbol{l}^T}{\boldsymbol{l}^T (AA^T)^{-1} \boldsymbol{l}} (AA^T)^{-1} \mathbf{d}. \end{aligned} \qquad (6.12c)$$

The sum $\mathbf{x} = \mathbf{x}' + c_{BG}(1, \ldots, 1)^T$ is the same for the last two cases. The background in eq. (6.12b) can be written as

$$c_{BG} = \frac{\boldsymbol{l}^T}{\boldsymbol{l}^T (AA^T)^{-1} \boldsymbol{l}} (AA^T)^{-1} \mathbf{d} = \frac{1}{\operatorname{rank}[A]} \sum_j x_j^{\dagger},$$

with $\mathbf{x}^{\dagger} = A^{\dagger} \mathbf{d}$ and for more or less constant $c(\mathbf{r}) \sim c$ this is roughly the same as c_a in eq. (6.7).[6]

---

[6] In fact, c_{BG} in eq. (6.12b) can also be written as $c_{BG} = \operatorname{rank}[A]^{-1} \sum_i \frac{d'_i}{l'_i} {e'_i}^2$ where $\mathbf{d}'$, $\boldsymbol{l}'$ and $\mathbf{e}'$ are the vectors $\mathbf{d}$, $\boldsymbol{l}$ and $\mathbf{e} = (1, \ldots, 1)^T$ in the system of singular vectors. As the transformation to this system is orthonormal, the

(*ii*) *Subtracting the background:*

The idea is to subtract the background as an offset on both sides of eq. (6.10b) and to solve the reduced system

$$\mathbf{d}' = \mathbf{d} - c_{BG}\boldsymbol{l} = A\mathbf{x}'. \tag{6.13}$$

The rationale behind subtracting c_{BG} rather than using it as a priori is that the reduced system eq. (6.13) with zero a priori and positivity constraint allows finer discretisation as will be shown in sec. 8.5. In principle, c_{BG} can be taken from a fit as in (i), but the choice

$$c_{BG} = \min(\tfrac{d_1}{L_1}, \cdots, \tfrac{d_m}{L_m}) \tag{6.14}$$

assures positive d_i'. The case of a local sink can be treated accordingly, if after subtraction of the background the reduced system is written as $\mathbf{d}'' = -\mathbf{d}' = A\mathbf{x}'$.

6.3.3. Additional constraints

Additional constraints to the system $A\mathbf{x} = \mathbf{d}$ can arise from further measurements (c.f. sec. 7.3), from model estimates or they may occur in simulations. Linear inequality constraints can be implemented in the same way as the positivity constraint through iterative projection (secs. 4.8.1, 4.8.4). Linear equalities of the form

$$F\mathbf{x} = \mathbf{c}, \quad F \in \mathbb{R}^{f\times n} \tag{6.15}$$

lead to the augmented system

$$\underbrace{\begin{pmatrix} A \\ F \end{pmatrix}}_{\mathcal{A}\,\in\,\mathbb{R}^{(m+f)\times n}} \mathbf{x} = \underbrace{\begin{pmatrix} \mathbf{d} \\ \mathbf{c} \end{pmatrix}}_{\boldsymbol{d}\,\in\,\mathbb{R}^{(m+f)}}, \tag{6.16}$$

where the rows should have similar weights, i.e. be multiplied with suitable factors if necessary. The associated least-squares and least-norm problems can be solved iteratively for the system $\mathcal{A}\mathbf{x} = \boldsymbol{d}$ in the nonlinear constrained case or explicitly in the linear case as shown in appendix C:

$$\mathbf{x} = (A^TA + F^TF)^{-1}(A^T\mathbf{d} + F^T\mathbf{c}) \qquad \text{if} \quad m+f \geq n, \tag{6.17}$$

see also eqs. (4.60), and

$$\begin{aligned} \mathbf{x} &= \mathcal{A}^T(\mathcal{A}\mathcal{A}^T)^{-1}\boldsymbol{d} + P_{\mathcal{A}}\mathbf{x}_a \\ &= \mathbf{x}_a + P_F A^T\big(AP_FA^T\big)^{-1}(\mathbf{d} - A\mathbf{x}_a) \\ &\qquad + P_A F^T\big(FP_AF^T\big)^{-1}(\mathbf{c} - F\mathbf{x}_a) \qquad \text{if} \quad m+f \leq n, \end{aligned} \tag{6.18}$$

with P as in eq. (4.55).

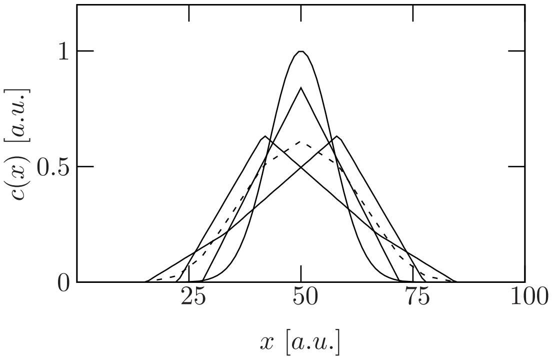


Figure 6.3: Example showing how the reconstructed peak height gets reduced by averaging over shifted grids. The initial regular grid carries five linear interpolation nodes and is shifted twice in x-direction. The Gaussian peak is as in fig. 5.3b. The solid lines show the ideal discretisation (sec. 5.2.1) for the shifted grids, the dashed line is their average.

6.3.4. Fitting peak maxima

The grid averaging scheme (sec. 6.2.2) will inevitably smear narrow peaks and reduce the maximum value, c.f. fig. 6.3. Looking at the individual subgrid reconstructions of Gaussian peaks (chap. 8) it was observed that the grid node of the absolute maximum of all subgrids correlates well with the grid node next to the actual peak maximum. Furthermore, the grid nodes and the values of the reconstructed absolute maximum among the subgrids agrees well with what can be expected from an ideal parametrisation of the Gaussian peak (sec. 5.2.1). It thus makes sense to regard this absolute maximum $c_{max} = \hat{c}^I(\mathbf{r}_{J_{max}})$ belonging to subgrid I and the node J_{max} of the composite grid as a better representation of the true maximum than the averaged value $X_{J_{max}}$ according to eq. (6.3). Therefore, the grid averaging scheme is run a second time with the additional constraint

$$x^I_{J_{max}} = c_{max} \tag{6.19}$$

for all subgrids $I = 1, \ldots, M$ as described in the preceding section. Results for this procedure will be presented in sec. 8.2.3. In principle, several local maxima could be treated in the same manner.

6.4. Finding optimal settings from simulations

The variables for the reconstruction procedure sketched so far are summarised in fig. 6.4. A specific choice of these parameters leads to a reconstruction result and an error estimate, e.g. by the numerical scheme eqs. (5.14), related to this particular setting. The aim is, of course, to find the optimal setting in the sense that it leads to smallest possible reconstruction errors. Similar to the numerical error estimation this can only be done by simulated reconstruction for suitable test distributions. As now the reconstruction error averaged over the ensemble has to be compared for possibly many realisations of the parameter set $\mathcal{P}$, scalar quality measures Q as discussed in sec. 5.4 are evaluated rather than the error field $\Delta\hat{c}(\mathbf{r})$. For a choice of parameters $p \in \mathcal{P}$ a simulation step for the test distribution c_I from the ensemble $\mathcal{E}$ takes the form

$$c_I(\mathbf{r}) \in \mathcal{E} \xrightarrow[\mathbf{d}_i = \int_i ds\, c(\mathbf{r})]{\text{forward model}} \mathbf{d}_I \xrightarrow[p]{\text{reconstruct}} \hat{c}(\mathbf{r}) \Rightarrow \quad Q(c_I, \hat{c}_I, p). \tag{6.20}$$

---

latter expression should be comparable to c_a in eq. (6.7).

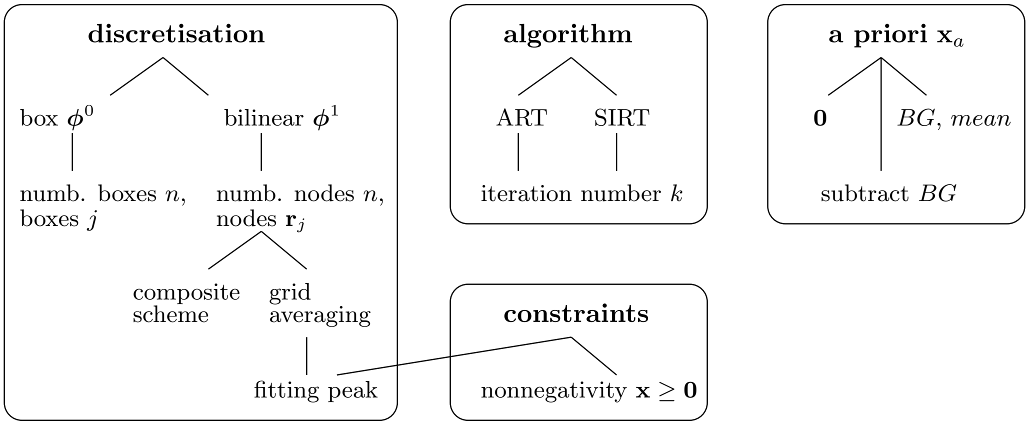


Figure 6.4: Choices for elements of the reconstruction procedure applied in part III.

Taking the ensemble average and repeating the procedure for all parameters $p_1, p_2, \ldots \in \mathcal{P}$ yields the optimal parameters as those for which $\langle Q(c_I, \hat{c}_I, p)\rangle_{\mathcal{E}}$ shows the optimum.

In general, measurement errors should be taken into account – not only if the parameter varied is the regularisation parameter (iteration number) – as in eqs. (5.14), so that finally the expression

$$\langle \overline{Q(c_I, \hat{c}_I, p)}_{\boldsymbol{\epsilon}} \rangle_{\mathcal{E}} \tag{6.21}$$

has to be optimised with respect to the parameters p.

7. Designing a Tomographic Experiment

This chapter covers various aspects relevant for planning a tomographic measurement and its targets but not affecting the reconstruction procedure as such. Secs. 7.1, 7.2 deal with the setup of the path geometry and related questions like: *How many light paths are minimally necessary*? or *How much information will be lost due to noise*? etc. The inclusion of additional (point) measurements is discussed in sec. 7.3 depending on whether these are treated equally to the path measurements or regarded as a priori that might be overridden by the latter. The last section examines how a tomographic reconstruction can be used for model evaluation (statistically, see sec. 5.4) or even verification (deterministically, e.g. sec. 5.3), inevitably involving both the reconstruction error and the model uncertainty. It briefly addresses the problem of relating concentration values from point measurements to modelled values thus following up the discussion in secs. 2.3, 2.4.

7.1. Requirements for the setup – number of light paths

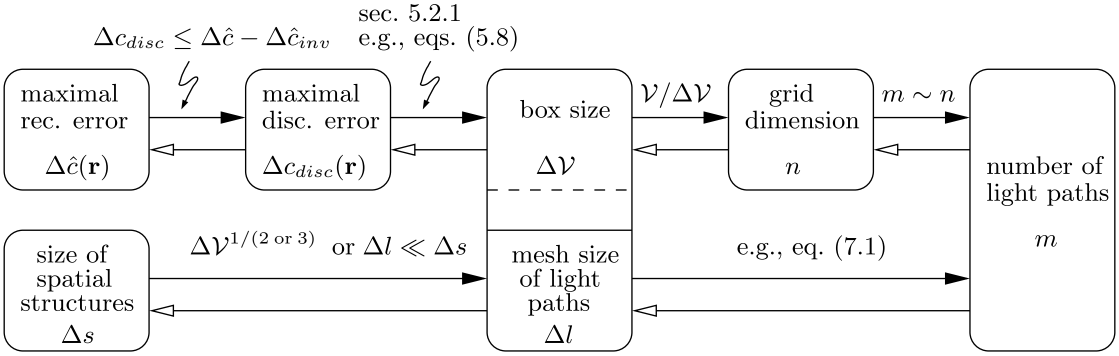


Figure 7.1: Estimation of the number of paths necessary for a maximally acceptable reconstruction error or minimally resolvable spatial structures and vice versa, neglecting the measurement error and assuming the even-determined case.

Evident questions for a tomographic experiment are of the sort "Given a hypothesis or model assumption on the spatial distribution of a certain trace gas, what experimental setup, i.e. essentially what number of light paths is needed for its verification or falsification?". This translates into a maximally acceptable error of the reconstructed distribution or a minimum resolving power for spatial structures. Fig. 7.1 illustrates the conclusion to the number of light paths necessary to meet the requirements under the assumptions that no further (a priori) information is included and that the measurement error plays a minor role. The mesh size of the light paths in fig. 7.1 is understood as the typical, smallest area enclosed by the crossing rays and, obviously, it can vary significantly within the reconstruction area or for different directions. For DOAS measurements the complete geometry

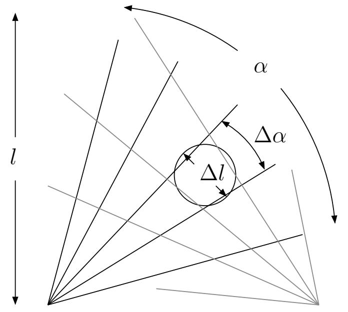


Figure 7.2: Illustration of the mesh size Δl for a typical fan beam geometry generated by static emitting or receiving instruments.

is generated by telescopes emitting (or receiving) light for different angles α_i (elevations in the case of vertical measurements).

> For example, assuming a number of n_t static instruments and $\Delta\alpha_i = \alpha_{i+1} - \alpha_i \sim \Delta\alpha$ to be more or less the same for all instruments, then the size Δl of the light path meshes in the centre of the reconstruction area of size $l \times l$ and the number m of the light paths according to fig. 7.2 are related roughly via
>
> $$\Delta l \sim \frac{1}{\sqrt{2}}\,\Delta\alpha\, l \sim \mathcal{O}(1)\,\alpha\, n_t\, l\, m^{-1}, \tag{7.1}$$
>
> where the same number of rays and angle of beam spread α for all telescopes was assumed.

7.2. Light path geometry – degrees of freedom and influence of the a priori

The preceding section related the light path geometry to the discretisation. Unless there is a continuous a priori $c_a(\mathbf{r})$ the role of measurement error and a priori for the geometry can be discussed in the discrete vector space in terms of linear algebra.

(i) Arguments based on the discrete forward model only:
Without measurement error the arrangement of the light paths for a given number of discretisation parameters n is as good as it can be if

$$\mathrm{rank}[A] = \min(m, n).$$

Components of $\mathbf{x}$ in the $n - \mathrm{rank}[A]$-dimensional nullspace of A are given by information other than the measurement data. According to sec. 4.5, $\mathbf{x}$ can be represented in the basis of singular vectors as $\mathbf{x} = \sum_j x'_j \mathbf{v}_j$, where the $\mathbf{v}_j$ with $j = \mathrm{rank}[A]+1, \ldots, n$ span the nullspace. Consequently, one gets for the matrix V of the column vectors $\mathbf{v}_j$:

$$\text{If} \quad \sum_{j'>\mathrm{rank}[A]} V_{jj'}^2 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}, \quad \text{then } x_j \text{ is fixed completely by the} \begin{Bmatrix} \text{data} \\ \text{a priori or constraints} \end{Bmatrix}. \tag{7.2}$$

How this picture changes in the presence of measurement errors has been implicitly addressed in sec. 4.4 on the ill-posedness of the tomographic reconstruction and in sec. 6.2.1 on the ideal reconstruction grid. The former discussion suggests that the light paths should be as linear independent as possible. The latter discussion used the eigenvalues of $A^T A$ but it can be led in the same way for AA^T, showing that the geometry should be such that the individual paths have similar length and as little overlap in the boxes as possible to ensure singular values of similar size. To quantify how much interdependency of the light paths is acceptable, it is helpful to look at the singular value decomposition of A:

$$\mathbf{d} + \boldsymbol{\epsilon} = A\mathbf{x} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T \mathbf{x} + \underbrace{\cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \mathbf{x}}_{\gtrless |\boldsymbol{\epsilon}|\ ?}, \quad r = \text{rank}[A]. \tag{7.3}$$

As indicated, only those modes of the decomposition add information which give a contribution above the noise level. The singular vectors are normalised to one so that this can be seen to be the case if

$$\sigma_j \gtrsim \epsilon \ / \ |\mathbf{v}_j^T \mathbf{x}| \sim \epsilon \ / \ |\mathbf{x}|, \tag{7.4}$$

where ϵ is the root mean square error, i.e. $|\boldsymbol{\epsilon}|/\sqrt{m}$.[1]

(*ii*) *Arguments based on the inversion*:
Only the case of a linear inversion by a matrix A^{inv} is considered. In this case the resolution matrix $R = A^{inv}A$, sec. 4.6.4, can be used to study how for a given light path geometry the original concentration values $x_{true\,j}$ are mapped to the whole reconstruction grid by $x_j = \sum_{j'} R_{jj'} x_{true\,j'}$. We note in passing that $(\mathbb{1}_n - R)_{jj} = \sum_{j' > \text{rank}[A]} V_{jj'}^2$ for the generalised inverse, eqs. (4.37), so that for the ideal reconstruction ($R_{jj} = 1$) the sum over the nullspace indeed equals zero.

The Bayesian approach in the special form of the optimal estimate gives bounds similar to eq. (7.4) (see sec. 4.7.2): A mode j of the SVD of A increases the information content of the measurement, if for the singular value $\tilde{\sigma}_j$ of $\tilde{A} = S_{\boldsymbol{\epsilon}}^{-1/2} A S_a^{1/2}$ holds

$$\tilde{\sigma}_j \gtrsim 1. \tag{7.5}$$

It was remarked at the beginning of sec. 4.7 that, even if not feasible as a reconstruction method, the Bayesian method can be used as a diagnostic tool for *what-if* scenarios, testing the outcome of an experiment for an assumed a priori state of the system. In this sense the optimal estimate will be used in sec. 8.4.2 to study the impact of measurement noise for different light path geometries.

7.3. Including point measurements and profile information

Planning a remote sensing experiment can involve point measurements in two ways

- The location of the point measurements is fixed – how should the light paths be chosen ?
- The light paths are fixed – where should the point measurements be performed ?

A point measurement at the location $\mathbf{r}_0$ with result c_0 takes in the discrete framework the form

$$c(\mathbf{r}_0) = [\boldsymbol{\phi}(\mathbf{r}_0)]^T \mathbf{x} = c_0 \quad \text{(with measurement error } \sigma_{c_0}). \tag{7.6}$$

---

[1] A quite lengthy calculation without using the SVD for continuous kernel functions (and the over-determined case) in [*Twomey*, 1997, sec. 8.4] can be adopted to the discrete case and gives the same threshold.

If there are p such measurements at locations $\mathbf{r}_{i_p}$, $i_p = 1, \dots p$, their results can be joined in a vector $\mathbf{c} \in \mathbb{R}^p$ and a matrix $F \in \mathbb{R}^{p \times n}$ as

$$F\mathbf{x} = \mathbf{c} \qquad \text{with} \qquad \begin{array}{l} \mathbf{c} = (c_1, \dots, c_p)^T \\ F_{i_p j} = \phi_j(\mathbf{r}_{i_p}) \end{array} \quad . \tag{7.7}$$

For the box basis or if the location of the point measurement coincides with the grid nodes in the case of the bi-(tri)linear basis, the rows in the system become

$$x_{i_p} = c_{i_p} .$$

Concentration measurements along a profile $\Pi : \ \mathbf{r} = \mathbf{r}(s) \in \mathbb{R}^3, \ s \in \mathbb{R}$, can be considered as a series of point measurements $c(\mathbf{r}(s_{i_p})) = c_\Pi(s_{i_p})$, where the $\mathbf{r}(s_{i_p})$ lie along the profile, and can thus in principle also be expressed in the form of eq. (7.7).

There are principally two ways to incorporate information from point or profile measurements into the tomographic reconstruction: (i) as additional constraints like in sec. 6.3.3 or (ii) as (additional) a priori.

(i) *Point measurements as additional constraints*:
The augmented system $\mathcal{A}\mathbf{x} = \boldsymbol{d}$ defined in eq. (4.60a) can again be solved explicitly. Assuming the under-determined case $m+p < n$ with a priori $\mathbf{x}_a$ and no further inequality constraints, the least-norm solution is given by eq. (6.18).

(ii) *Point measurements as additional a priori*:
The point measurement(s) will in general not provide an a priori for all components of $\mathbf{x}$, so that further a priori $\mathbf{x}_a$ is still needed. The augmented a priori

$$\boldsymbol{x}_a = F^T(FF^T)^{-1}\mathbf{c} + P_F\,\mathbf{x}_a \tag{7.8}$$

draws rank$[F]$ degrees of freedom from the point measurement(s) $F\mathbf{x} = \mathbf{c}$ and the remaining n-rank$[F]$ from $\mathbf{x}_a$ (appendix C). P_F is the projector onto the nullspace of F. The least-squares least-norm solution of $A\mathbf{x} = \mathbf{d}$ without further constraints reads then as

$$\mathbf{x} = A^T(AA^T)^{-1}\mathbf{d} + P_A\big(F^T(FF^T)^{-1}\mathbf{c} + P_F\mathbf{x}_a\big). \tag{7.9}$$

Case (i) treats remote sensing and point measurement in the same way. The (ad hoc) a priori $\mathbf{x}_a$ contributes only in the reduced n-rank$[\mathcal{A}]$ dimensional nullspace $\mathcal{N}(\mathcal{A})$. In (ii) the a priori $\mathbf{x}_a$ adds components in the intersection of $\mathcal{N}(A)$ and $\mathcal{N}(F)$: $P_A P_F\,\mathbf{x}_a$.[2]

---

[2] The difference is best illustrated by an *example*: Let there be five boxes and one light path through two boxes, a second one through the other three such that $A = \begin{pmatrix} 1\,1\,0\,0\,0 \\ 0\,0\,1\,1\,1 \end{pmatrix}$. Let there further be a point measurement in the first box, i.e. $F = (1\,0\,0\,0\,0)$, $\mathbf{c} = (c)^T$. Then for the augmented system

$$\mathbf{x}_{(i)} = \begin{pmatrix} c \\ d_1 - c \\ \frac{1}{3}(d_2 + 2x_{a_3} - x_{a_4} - x_{a_5}) \\ \frac{1}{3}(d_2 - x_{a_3} + 2x_{a_4} - x_{a_5}) \\ \frac{1}{3}(d_2 - x_{a_3} - x_{a_4} + 2x_{a_5}) \end{pmatrix}, \quad \text{while} \quad \mathbf{x}_{(ii)} = \begin{pmatrix} \frac{1}{2}(d_1 + c - x_{a_2}) \\ \frac{1}{2}(d_1 - c + x_{a_2}) \\ \text{as for } \mathbf{x}_{(i)} \\ \shortparallel \\ \shortparallel \end{pmatrix}$$

with augmented a priori $\boldsymbol{x}_a = (c,\ x_{a_2},\ x_{a_3},\ x_{a_4},\ x_{a_5})^T$.

Returning to the question of how both kinds of measurements should be set up, one finds that if they are meant to be complementary, ideally the light paths or the location of the point measurement(s) should be set up such that the nullspaces of A and F according to eq. (7.7) have as little in common as possible, e.g. by choosing the row vectors of F as a linear combination of any p singular vectors that span the nullspace of A. In practice, such a construction will be hard to realise and might be corrupted by noise for the reasons discussed in the preceding section. Therefore, an estimate of the degrees of freedom (or the information content) of different configurations of remote sensing and point measurement for an expected level of noise as outlined in section 7.2 can be helpful. For the augmented system of case (i) the measurement error and a priori covariances are given by $\mathcal{S}_{\boldsymbol{\epsilon}} = \left(\begin{smallmatrix} S_\epsilon & 0 \\ 0 & L^2 S_c \end{smallmatrix}\right)$ and S_a, where the system $F\mathbf{x} = \mathbf{c}$ was scaled with a typical length L of the light paths. The augmented a priori covariance for the case (ii) is given as (appendix C)

$$\mathcal{S}_a = F^T(FF^T)^{-1}S_c(FF^T)^{-1}F + P_F S_a P_F. \tag{7.10}$$

7.4. Aspects of model evaluation

Motivation for evaluating the performance of dispersion or chemical transport models on very small spatial scales was given in chap. 2 in the context of urban air pollution monitoring. *Evaluation* is understood here as the comparison of modelled concentrations with corresponding experimental values.[3] According to sec. 2.3 this is more complicated than it sounds as both refer to fundamentally different quantities: For given meteorological conditions models describe the mean field given by eq. (2.3) plus, possibly, a more or less accurate approximation of the turbulent part c':

$$c_{mod}(\mathbf{r}) \sim \langle c(\mathbf{r},t)\rangle + c'_{approx}(\mathbf{r},t) \quad \text{for } \bar{\mathbf{u}}, T, p, \text{ etc. at time } t,$$

whereas a point measurement averaging over the period $\mathcal{T} = [t; t+\Delta t]$ yields the value

$$c_{exp}(\mathbf{r},\mathcal{T}) \sim \langle c(\mathbf{r},t)\rangle|_{\mathcal{T}} + \overline{c'(\mathbf{r},t)}_{\mathcal{T}}, \tag{7.11}$$

where $\langle c(\mathbf{r},t)\rangle|_{\mathcal{T}}$ denotes the value of $\langle c(\mathbf{r},t)\rangle$ (assumed to be constant) during $\mathcal{T}$. On the one hand, c_{mod} and c_{exp} will differ substantially for short averaging times

$$c_{mod}(\mathbf{r}) \neq c_{exp}(\mathbf{r},\mathcal{T}) \quad \text{for } \Delta t \to 0.$$

On the other hand, for large enough Δt

$$c_{mod}(\mathbf{r}) \approx c_{exp}(\mathbf{r},\mathcal{T}) \quad \text{for } \Delta t \to \infty \text{ and stationary } \bar{\mathbf{u}}, T, p, \text{ etc.}$$

Because meteorological conditions are never stationary, a compromise must be found between averaging times long enough to sufficiently reduce the randomness of the measured values and short enough to guarantee the same constant atmospheric parameters to be comparable to the model prediction. Averaging times of around 30 *min* are common but according to *Schatzmann and Leitl* [2002], who consider tracer measurements for a wind tunnel model of a street canyon for different averaging times,

---

[3] Whereas *validation* would mean a consistency check of the underlying physical concepts and their translation into computer code.

not always sufficient to obtain representative values. Scaling the averaging times corresponding to the real dimensions of the street canyon, the conclusion on ambient measurements can easily be drawn because in a wind tunnel stationary boundary conditions for $\Delta t \to \infty$ can actually be realised. An even more pessimistic estimation is inferred from long term time series of NO_x concentrations measured in the original street canyon where concentrations referring to the same boundary conditions scatter largely. It should be added in this context that it is common practice to convert a mean concentration c_1 for averaging time Δt_1 to another (usually shorter) Δt_2 by a relation $(\Delta t_2)^p\, c_2 = (\Delta t_1)^p\, c_1$ with p typically between 0.17 and 0.2 [*Barratt*, 2002]. This procedure is entirely empirical, giving no information on the turbulent part for different times and it is furthermore controversial [*Venkatram*, 2002].
The discussion so far concerned the representativeness of measured concentration values at a fixed place with respect to time. Interpretation of measurement results at different places $\mathbf{r}_1$, $\mathbf{r}_2$ and the same time interval $\mathcal{T}$ is similarly difficult, as the random and mean contributions in eq. (7.11) cannot be discriminated.

One way to deal with the stochastic uncertainty in eq. (7.11) in order to compare it with the corresponding model value is to use empirical or theoretical assumptions to relate the measured value $c_{exp}(\mathbf{r}, \mathcal{T})$ to the unknown mean $\langle c(\mathbf{r}, t)\rangle$ or to derive bounds for the size of the random contribution $\overline{c'(\mathbf{r}, t)}_{\mathcal{T}}$. For example, for urban pollution where not only the dispersion, but also the emissions follow more or less stochastic patterns, the experimental concentrations are often assumed to be distributed log-normally around the (model) mean (e.g., [*Venkatram et al.*, 2005], see also [*Seinfeld*, 1986, chap. 17]). The estimation of turbulent fluctuations has been shortly addressed in sec. 2.4 in terms of variances of the turbulent concentration field within a plume. A more comprehensive approach is a decription by probability density functions [e.g., *Yee and Chan*, 1996; *Ma et al.*, 2005]. While these value-by-value comparisons are necessary when it comes to judging model capability for individual concentrations or special features of the distribution, for an evaluation of the overall model performance the stochastic character of the measured values can be reduced by considering statistic quality measures like the ones introduced in sec. 5.4.
It is a bit difficult to specify the size of typical model errors or values of statistic quality measures because both heavily depend on the kind of application, the averaging times, atmospheric stability, the kind of species, the number of measurements and where they took place etc. Also the way and degree in which meteorological observations or immission measurements enter the model prediction can be quite different (c.f. fig. 2.6 on page 34). In the context of roadway dispersion modelling *Held et al.* [2003] regard an agreement between a dispersion model and measurement within 30% as perfect. For extensive calculations with a CFD model, see sec. 2.5, *Tsai and Chen* [2004] find discrepancies between modelled and measured concentrations over an averaging period of $2\frac{1}{2}\,h$ about $5 - 10\%$ for CO, $\sim 20 - 50\%$ for NO_x and $\sim 10 - 15\%$ for SO_2. Statistical evaluation of dispersion models on urban scales, e.g. by means of concentration data from SF_6 dispersion experiments like those mentioned in sec. 2.4, shows roughly the ranges listed in table 7.1 for statistical indices from table 5.1. The values should be taken with a grain of salt. Especially $NMSE$ and FB depend on the type of distribution. They are generally higher for small scale plumes, higher for larger scale distributions (see also sec. 8.2.1).

Turning from the point measurements implicitely referred to till now to tomographic remote sensing measurements, one expects the reconstructed field to depend less on the turbulent field $c'(\mathbf{r}, t)$. For

Table 7.1: Ranges for some metrics from table 5.1 as found for dispersion and chemical transport models on urban scales [Seinfeld and Pandis, 1998, sec. 23.8], [Hurley et al., 2005; Brandt et al., 2001; Moreira et al., 2005; Venkatram et al., 2004, 2005] (In most cases for 1 h averages).

$NMSE$	$\gtrsim$	0.2 %
IOA	$\sim$	0.5 - 0.7
$\vert FB\vert$	$\sim$	5 - 15%
r	$\sim$	0.3 - 0.8
$FA2$	$\sim$	40 - 80 %
PPA	$\sim$	15 - 20%

the individual column densities, eq. (4.3),

$$\overline{d(t)_{i}}_{\mathcal{T}} = \int\limits_{LP_i} ds\, \langle c(\mathbf{r},t)\rangle|_{\mathcal{T}} + \int\limits_{LP_i} ds\, \overline{c'(\mathbf{r},t)}_{\mathcal{T}}, \quad i = 1, \ldots, m, \tag{7.12}$$

the random part disappears for uniform fluctuations along the light path LP_i if their length scale is small against the length of LP_i (c.f. sec. 2.4, especially the example on p. 29). A numerical justification of this presumption could be carried out with a Monte-Carlo model as the one provided by *Blackadar* [1997]. Having mentioned this agreeable property of path integrating measurements, scenarios for model evaluation like the following ask for an estimation of the total reconstruction error.

(i) Given a modelled field $c_{mod}(\mathbf{r})$ and a field reconstructed from experiment $\hat{c}(\mathbf{r})$, do both agree within the reconstruction error, i.e.

$$|c_{mod}(\mathbf{r}) - \hat{c}(\mathbf{r})| \overset{?}{<} \Delta\hat{c}(\mathbf{r}) \tag{7.13}$$

(ii) Given a presumed uncertainty of the model, i.e.

$$c_{min}(\mathbf{r}) \lesssim c(\mathbf{r}) \lesssim c_{max}(\mathbf{r}), \tag{7.14a}$$

where $c(\mathbf{r})$ is the unknown true distribution, could a tomographic experiment further constrain the model (or verify the assumed uncertainty) ? That is

$$\begin{aligned} & c_{min}(\mathbf{r}) \overset{?}{<} \hat{c}(\mathbf{r}) + \Delta_l\hat{c}(\mathbf{r}) \\ \text{and} \qquad & \hat{c}(\mathbf{r}) + \Delta_u\hat{c}(\mathbf{r}) \overset{?}{<} c_{max}(\mathbf{r}) \end{aligned} \tag{7.14b}$$

in the sense of eq. (5.1).

While problem (i) requires estimation of $\Delta\hat{c}(\mathbf{r})$ independently of the model distribution, it is a consistent approach to (ii) to use the model uncertainty for numerical error estimation as described in sec. 5.3, based on the ensemble

$$\mathcal{E} = \{c_I(\mathbf{r}) \;\Big|\; c_{min}(\mathbf{r}) \lesssim c_I(\mathbf{r}) \lesssim c_{max}(\mathbf{r}) \wedge \ldots, \quad I = 1, \ldots, N\}, \tag{7.15}$$

where the dots refer to further specifications for $c_I(\mathbf{r})$ as described in appendix B. The bounds c_{min},

c_{max} could be given by roughly known relative errors x of the model, i.e.

$$c_{min/max}(\mathbf{r}) = (1 \pm x)\, c_{mod}(\mathbf{r}). \tag{7.16}$$

Alternatively, they could be based on empirical experience (comparison with other models etc) or be derived from statistical indices like $FA2$. If, for example, $FA2 \sim 100\%$, then one gets the bounds

$$c_{min}(\mathbf{r}) = 0.5\, c_{mod}(\mathbf{r}), \qquad c_{max}(\mathbf{r}) = 2\, c_{mod}(\mathbf{r}). \tag{7.17}$$

Part III.

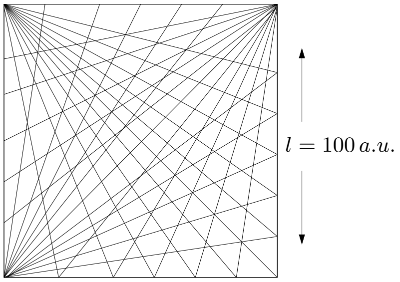
Numerical Results

8. 2-D Simulation Results for Gaussian Peaks

This chapter presents simulation results for the 2-D reconstruction of trace gas distributions emerging from a point source by turbulent diffusion, which can be described by the Gaussian plume or puff model (sec. 2.4). The case of emission puffs will be studied in the tomographic experiment of the next chapter. To be precise, it is assumed that the 2-D concentration distribution on the section through the plume or puff is Gaussian. The orientation of this section – that is, the reconstruction area – is arbitrary, but for LP-DOAS experiments it is likely to be horizontal. Superposition of several Gaussians can be used to mimic more complex concentration fields. It is assumed throughout that the trace gas species occurs only as concentration puff, i.e. there is no natural or anthropogenic background. This might be unrealistic for most trace gases relevant for DOAS measurements, but at the very end of the chapter in sec. 8.5, it will be shown that the preceding results hold for a smooth background, too, if subtracted before reconstruction.

The simulation results fall into three parts on the parametrisation of the problem (sec. 8.2), the inversion algorithm (sec. 8.3) and the light path geometry (sec. 8.4). All three parts employ a set of test ensembles defined in sec. 8.1. The first two parts use a fixed test geometry which is more or less the one realised in the aforementioned experiment.
The influence of the parametrisation of the reconstruction result in terms of grid dimension and grid combination schemes on different aspects of the reconstruction quality is examined systematically with respect to the extension of the puffs. The focus is on the smaller puffs of the test ensembles, for which box and bilinear parametrisation are compared in detail with respect to discretisation and inversion errors. Finally, sample reconstructions are shown. Anticipating the results of the second part (sec. 8.3) and in agreement with the literature, which generally finds SIRT superior to ART, all reconstructions in sec. 8.2 use SIRT.
The comparison of ART and SIRT in sec. 8.3 thus serves mainly to compare the influence of the choice of the algorithm to the influence of the parametrisation. Measurement noise and regularisation are discussed only for a special case that will be needed for the analysis of the experiment in chap. 9.
The third part of the simulation results starts with a short section (8.4.1) – more of an outlook – illustrating the SVD as a diagnostic tool. The main part consists of an analysis of different geometries with the same light path number by explicit numerical simulation for Gaussian distributions on the one hand (sec. 8.4.2(*i*)), and algebraic and Bayesian concepts on the other (sec. 8.4.2(*ii*)). The latter is used in sec. 8.4.3 in an example to optimally include a point measurement in the sense of sec. 7.3. Finally, a simple recipe is given to transfer simulation results for different numbers of light paths (sec. 8.4.4), thus increasing the use of the special cases studied here tremendously.

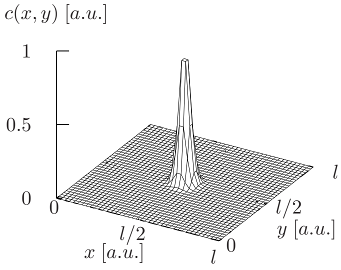
8.1. Test distributions and light path geometry



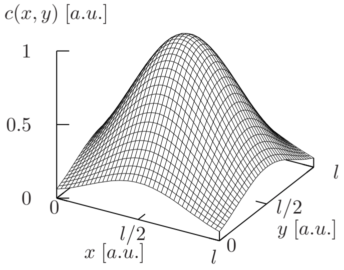
Light path geometry with three 90°-fans and 36 paths in total.

ensemble	$\sigma_{x,y}/l$	$\varnothing_{2\sigma}/l$	$\varnothing_{2\sigma}/\Delta l$
$\mathcal{E}_1$	$0.03\ldots0.05$	$0.1\ldots0.2$	$1\ldots2$
$\mathcal{E}_2$	$0.05\ldots0.1$	$0.2\ldots0.4$	$2\ldots4$
$\mathcal{E}_3$	$0.1\ldots0.2$	$0.4\ldots0.8$	$4\ldots8$
$\mathcal{E}_4$	$0.2\ldots0.3$	$0.8\ldots1.2$	$8\ldots12$

Variances and $2\,\sigma$-diameters of the peaks within the four ensembles. At $2\,\sigma$ the exponential has fallen to $\sim 14\%$ of its peak value. The area integral within $2\,\sigma$ around the peak centre accounts for 95% of the total integral. The last column presents the ratio of the peak extension $\varnothing_{2\sigma}$ to the mesh size Δl at the center of the geometry given by eq. (8.1).



Most narrow peak from $\mathcal{E}_1$ with $\sigma_x = \sigma_y = 0.03\,l$



Broadest peak from $\mathcal{E}_4$ with $\sigma_x = \sigma_y = 0.3\,l$

Table 8.1: Light path geometry used for the simulations in sections 8.2 and 8.3 and Gaussian concentration fields.

The reconstruction area is chosen as a square of side length $l = 100$ in arbitrary units (*a.u.*). The complete geometry of 36 light paths is generated by three 90°-fans with 12 beams each, emitted by three telescopes sitting in the corners of the square (see table. 8.1). This number of optical paths corresponds to the one of the indoor experiment and it is high enough to guarantee a certain degree of regularity required for a meaningful comparison of different types of path geometries later (sec. 8.4.2). The mesh size in the centre of the reconstruction area is according to eq. (7.1)

$$\Delta l \sim \tfrac{1}{\sqrt{2}}\,\alpha\,n_t\,l\,m^{-1} = \tfrac{1}{\sqrt{2}}\,\tfrac{\pi}{2}\,3\,\tfrac{l}{12} \sim 0.1\,l. \tag{8.1}$$

The Gaussian test concentration distributions are located randomly within the reconstruction area. According to their peak width, they are divided into four ensembles $\mathcal{E}_\mathcal{I}$, $\mathcal{I} = 1, \ldots, 4$:

$$\mathcal{E}_\mathcal{I} = \Big\{ c_I(x,y),\ I = 1,\ldots,N \;\Big|\; c_I(x,y) = C_0\ e^{-\frac{1}{2}\left((x-x_0)^2/\sigma_x^2 + (y-y_0)^2/\sigma_y^2\right)},$$
$$\text{with random } 0 \le x_0, y_0 \le l, \quad 0.1\,C_{max} \le C_0 \le C_{max}, \quad \sigma_{x,y} \text{ as in table 8.1} \Big\}, \tag{8.2}$$

where $C_{max} = 1\, a.u.$ All random quantities are uniformly distributed and N is the number of random samples.

Taking the radius of the peak as $2\,\sigma$, ensemble $\mathcal{E}_1$ in table 8.1, with a smallest peak diameter similar to the mesh size given by eq. (8.1), represents the lower limit of narrow peaks that are still detectable for the coverage with light paths at hand. Here the aim is rather to locate the peaks and to reconstruct total amounts of concentrations, i.e., total emissions, whereas for the broad peaks of ensembles $\mathcal{E}_3$ and $\mathcal{E}_4$ reconstruction of the actual distribution should become feasible (see the lower panels in table 8.1). If the test concentration field consists of more than one peak, maximum values for all Gaussians vary randomly between $0.1\,C_{max} \ldots C_{max}$. Maximally four peaks are considered at the same time, as reconstruction of any further, especially small peak becomes highly unreliable for the given number of light paths.

8.2. Parametrisation

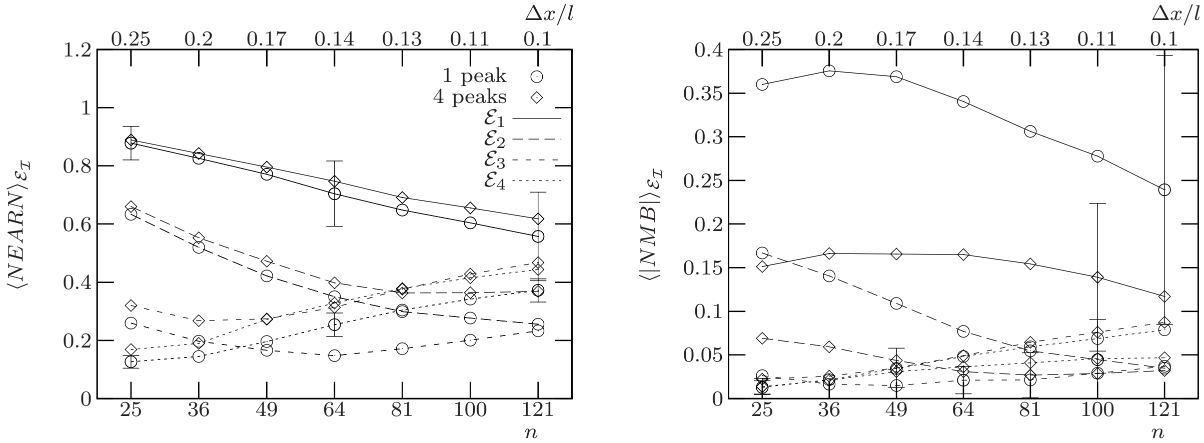
Referring to fig. 6.4 (p. 108), this section presents results for different choices within the discretisation procedure. As said, background concentration distributions in the sense of sec. 6.3.2 are assumed to be negligible or to have been subtracted as proposed in sec. 6.3.2(*ii*).

All results in this section are for optimised iteration number k in the sense of sec. 6.4. The dependence on k will be addressed in sec. 8.3. The nonnegativity constraint is always implemented.

8.2.1. Grid dimension

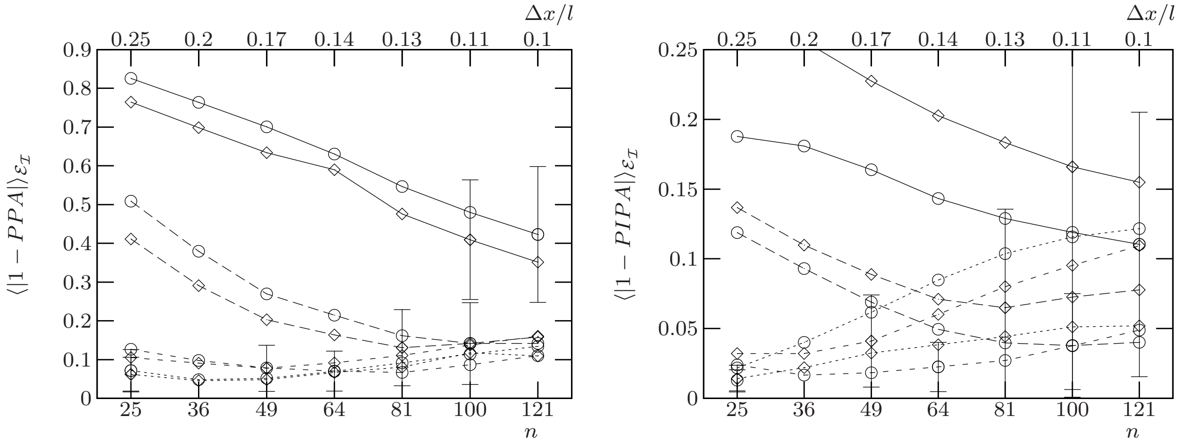
Several quality criteria from table 5.1 are calculated within each ensemble depending on the grid dimension n. To concentrate on the important point, only bilinear basis functions and the simultaneous algorithm SIRT will be considered. Furthermore, the grids are regular with uniform spacing, so that a grid of dimension n has $n_x = n_y = \sqrt{n}$ nodes in each direction and grid spacing $\Delta x = \Delta y = l/(\sqrt{n}-1)$. Nearness $NEARN$, representing the overall reconstruction error, the quality of the reconstructed mean concentration in the form of the normalised mean bias NMB and the accuracy of reconstructed maxima of the highest peak PPA and of peak integrals/total emissions $PIPA$ basically cover the most relevant features one would be interested in when trying to retrieve peak distributions. They are shown for one and four peaks of each ensemble in fig. 8.1. For ensembles $\mathcal{E}_3$ and $\mathcal{E}_4$ the mean a priori $\mathbf{x}_a$ according to eq. (6.7) was chosen as iteration start $\mathbf{x}^{(0)}$, otherwise the a priori is zero. The most important points emerging from these curves are

- For *narrow* peaks ($\mathcal{E}_1$ and $\mathcal{E}_2$) all error measures in fig. 8.1 can be tremendously reduced relative to the even-determined case $n \equiv m = 36$ by choosing discretisation grids that lead to highly under-determined systems of equations. For absolute maxima and total emissions the improvement amounts to more than 50%.
- This holds also for more (four) peaks. Here the ambiguity of individual quality measures becomes apparent. While the overall reconstruction error for four peaks is always smaller than for one peak (the reconstruction is less accurate, as one might expect), the relative error of the mean concentration NMB behaves inversely. In the same way, the peak prediction accuracy is higher



(a) Nearness

(b) Normalised mean bias



(c) Error of the maximum peak prediction accuracy

(d) Error of the peak integral prediction accuracy

Figure 8.1: Ensemble averages of different quality criteria for one and four random peaks depending on the grid dimension. The number of light paths is 36, the geometry is as in tab. 8.1. The bilinear grid is regular, so that dimension n means $\sqrt{n}-1$ *pixels in each direction. Also shown is the grid spacing* Δx*. The a priori is zero for* $\mathcal{E}_1$*,* $\mathcal{E}_2$ *and the mean eq. (6.7) for* $\mathcal{E}_3$*,* $\mathcal{E}_4$*. SIRT was used (see sec. 8.3 for iteration numbers). The number of random samples is* $N = 300$*. For the sake of clearness only some standard deviations* $\mathrm{std}_{\mathcal{E}_\mathcal{I}}(\cdot)$ *are shown. (a) Overall reconstruction error* $NEARN$*. (b) Absolute error of the mean concentration* $|NMB|$*. A value of 0.25 means that the reconstructed spatial mean deviates by* 25% *from the real one. (c) Absolute deviation of the reconstructed global maximum from the real one (*$|1 - PPA|$*). A value of 0.5 means deviation by* 50%*. (d) The same for the peak integral, i.e. a value of 0.1 means average over- or underestimation of total emissions by* 10%*.*

for four peaks, while the error of the reconstructed peak integral is lower than for one peak only (suggesting that, with several peaks, the largest 'accumulates' concentration at the expense of the others).

- While the errors decrease with increasing n, the standard deviations $\mathrm{std}_{\mathcal{E}_\mathcal{I}}(\cdot)$ get larger, as one might expect because the degrees of freedom of the picture increase with growing n (visible in the form of artefacts, like bumps, in the reconstructed distribution).
- For *broad* peaks ($\mathcal{E}_3$ and $\mathcal{E}_4$) under- or even-determined discretisation grids lead to best results. For the broadest peaks from $\mathcal{E}_4$ the overall reconstruction quality rapidly deteriorates for higher dimensional grids (resulting in large artefacts).

- The optimal grid dimension is not always the same, but similar for all quality criteria.

Solutions discretised by 10 × 10 pixels imply about $121/36 \sim 3$ more unknown parameters than measured integrals! How can the resulting continuous distributions for narrow peaks still be better approximations of the real peaks? The reasons are

1. For narrow peaks most of the grid nodes carry the concentration zero, and this is exactly the value provided by the a priori (in the linear case through eqs. (4.56)).

2. The discretisation error as defined in sec. 5.2.1 decreases with increasing grid dimension. This readily explains the decrease in the overall reconstrcution error ($NEARN$) and also the improved estimate of maximum concentration values (PPA), as the peak is spread over smaller pixels (mind that for 10 × 10 pixels the grid spacing Δx becomes comparable to the peak diameter $\varnothing_{2\sigma}$). For the total and peak area integrals ($NMB, PIPA$) it should be noticed that these may increase if the original concentration peak is spread over more pixels by the reconstruction while the column densities are kept constant, as illustrated by the figure below.

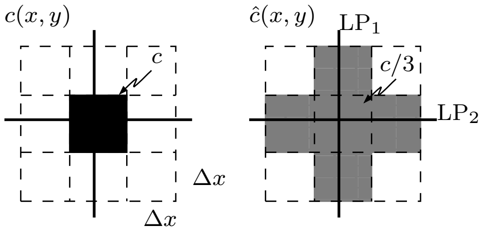


*Figure 8.2: For the original distribution (*left*) the column densities are $d_1 = d_2 = c\Delta x$ and the integral I over the complete reconstruction area is $I = \int dx\, dy\, c(x,y) = (\Delta x)^2 c$. For the reconstruction (*right*) $d_1 = d_2$ remain unchanged, but the area integral is now $\frac{5}{3} I$.*

3. The nonnegativity constraint becomes active around zero so that the nullspace is reduced by the negative solutions.

Putting points 1. and 2. together means that, when increasing the grid dimension, a possible increase of the inversion error –with a bad a priori being part of it– remains smaller than the decrease of the discretisation error. This will be further investigated in the next section. An illustration of point 3. will be given in sec. 8.4.1.

How can these findings be implemented to improve the reconstruction from real experiments, i.e. when actual plume extensions are *not* known? First, the experimentalist is not as ignorant as a set of equations and there might well be information that does not enter the formulation of the tomographic reconstruction problem. For example, the dimension of an emission puff driven through the setup of the light paths can roughly be inferred from the increase and decrease of column densities combined with wind speed. Or empirical dispersion coefficients might be useful (sec. 2.4). A way to obtain information about the true peak extension within the reconstruction itself is contained in tab. 8.3b, p. 129, which shows sample reconstructions for each ensemble and various grid dimensions. The peak in tab. 8.3b, reconstructed from a Gaussian belonging to $\mathcal{E}_2$, increases its height going from a grid with 8×8 pixels to 10×10 pixels, but decreases again for larger grid dimension while artefacts appear. Both features together might be interpreted as the optimal grid dimension being exceeded.

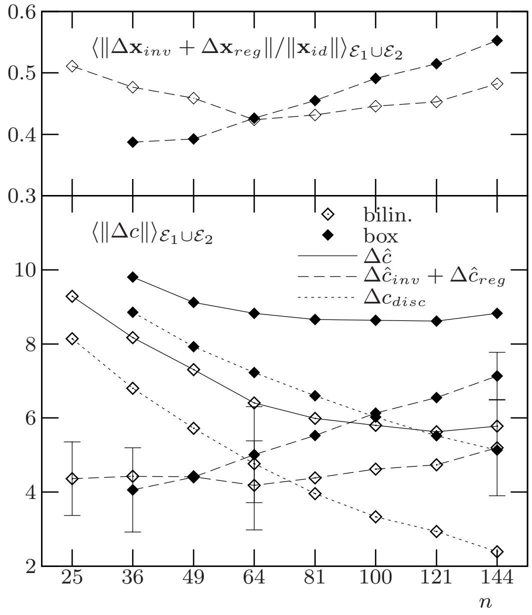


Figure 8.3: Lower graph*: Overall discretisation, inversion plus regularisation and total reconstruction errors for piecewise constant (box) and bilinear parametrisation of four (narrow) concentration peaks from $\mathcal{E}_1 \cup \mathcal{E}_2$ versus grid dimension. The grids are regular and dimension n means n pixels for the box and $(\sqrt{n}-1) \times (\sqrt{n}-1)$ pixels for the bilinear basis. The under-determined solutions are practically unregularised, so that $\Delta c_{reg}(\mathbf{r}) \sim 0$. The a priori is zero as in fig. 8.1, again for SIRT and $N = 300$.*
Upper graph*: The same for the relative discrete inversion error.*

8.2.2. Basis functions

The piecewise constant and bilinear basis functions are compared by explicitly calculating the discretisation error $\Delta c_{disc}(\mathbf{r})$ as proposed in sec. 5.2.1. The inversion error $\Delta\hat{c}_{inv}(\mathbf{r})$ immediately follows as $[\boldsymbol{\phi}(\mathbf{r})]^T (\mathbf{x}_{id} - \mathbf{x})$ and if the solution is regularised we consider $\Delta\hat{c}_{inv} + \Delta\hat{c}_{reg} = [\boldsymbol{\phi}(\mathbf{r})]^T (\mathbf{x}_{id} - \mathbf{x}_\lambda)$ (It turns out that, especially in the over-determined region, regularisation can improve the solution even if there are no measurement errors. I postpone the discussion to sec. 8.3). Fig. 8.3 contains in the lower graph the mean overall discretisation error $\|\Delta c_{disc}(\mathbf{r})\|_2$ and the inversion error of the regularised solution $\|(\Delta\hat{c}_{inv} + \Delta\hat{c}_{reg})(\mathbf{r})\|_2$ for narrow peaks from ensembles $\mathcal{E}_1$ plus $\mathcal{E}_2$ for both types of basis functions. In fact, here only the solutions with $m \gtrsim n$ are regularised and the regularisation error hardly contributes for $n \gtrsim m$.

As expected, the overall discretisation error, and therefore the total reconstruction error, is far larger for the piecewise constant than it is for the piecewise bilinear parametrisation. Furthermore, fig. 8.3 shows that:

- The squares of the overall discretisation and inversion error do indeed add up to the reconstruction error according to eq. (5.18).
- For the concentration distributions considered in fig. 8.3 the discretisation and inversion error are of similar size for the optimal grid dimension.
- The inversion error increases with grid dimension n. The decreasing total reconstruction error is finally solely due to the faster decreasing discretisation error.
- In the under-determined regime the inversion error for the box basis is larger and faster growing than for the bilinear basis.

Concentrating on the under-determined cases $n > m$ for the narrow peaks (the over-determined case applies to the broad peaks which are not part of these simulations), the last point is most remarkable because the discrete systems for both parametrisations are combined with the same a priori $\mathbf{x}_a = 0$

relative error		box	bilinear
mean concentration	$\vert 1-NMB\vert$	4%	5%
total emission	$\vert 1-PIPA\vert$	15%	10%
maximum concentration	$\vert 1-PPA\vert$	40%	20%

Table 8.2: Relative errors of concentration integrals and maxima for a bilinear grid of dimension $n = 81$. Distributions and all parameters are as in fig. 8.3.

and it is not obvious why $\Delta\hat{c}_{inv}(\mathbf{r})$ should differ significantly. One has to be careful though when carrying over conclusions from continuous solutions to $\mathbb{R}^n$ and vice versa. For the norm this step takes the form $\|\hat{c}\|_2^2 = \mathbf{x}^T\Phi\mathbf{x}$ with Φ as in eqs. (5.5) yielding $\|\hat{c}\|_2 = \sqrt{\Delta A}\,\|\mathbf{x}\|_2$ for the box basis and regular discretisation grids, where ΔA is the area of a grid cell. The upper graph of fig. 8.3 shows that the relative norm of the discrete inversion error for the bilinear basis is still smaller for highly under-determined systems of equations, but less distinctive. The fact that the unregularised least norm solution for the bilinear parametrisation is better in terms of inversion errors has to be attributed to the a priori and/or constraints. For example, the bilinear basis functions have overlapping supports so that the matrix A has less zero entries than the corresponding matrix for the box basis and the nonnegativity constraint might become more effective.

So far error fields $\Delta c(\mathbf{r})$ have been considered, but with the piecewise constant basis representing box average concentrations (see eqs. (5.6)) one might ask whether the box representation enables better reconstruction of spatial mean concentrations. In general it does not, as illustrated by table 8.2 which gives the relative error of reconstructed mean concentrations (NMB) and total emissions ($PIPA$) for an under-determined parametrisation by 8×8 pixels. The standard deviations of these errors within the test ensembles are of the size of the errors themselves, so that the difference in the mean concentration is not really significant. Nevertheless the reconstruction errors of both mean concentration and total emissions are persistently larger for the box than for linear parametrisation for all under-determined grids. Recalling that for the calculation of the peak integrals the pixel size is taken into account (see explanation to eq. (5.20)), the fact that – compared to the bilinear basis – the error of the peak integrals for the box basis is relatively larger than the error of the mean concentration suggests the interpretation that the concentration is distributed on a far larger area than the peak or elsewhere within the reconstruction area (artefacts). In both cases the inversion error would become larger, agreeing with the curves in fig. 8.3 which show a larger inversion error for the box parametrisation in the under-determined case.

8.2.3. Grid translation

As in the previous sections the effect of combining grids on the reconstruction is examined statistically for quality indices measuring the overall square error (nearness), the error of the spatial mean concentration (normalised mean bias NMB), the error of absolute maximum concentrations (peak prediction accuracy PPA) and the error of total emissions $PIPA$). Fig. 8.5 shows averages within ensembles $\mathcal{E}_1$ to $\mathcal{E}_4$ for the composite scheme (sec. 6.2.2(i)), the averaging scheme (sec. 6.2.2(ii)) and the averaging scheme that uses the absolute maximum as additional constraint (sec. 6.3.4) compared to the case of a single grid. The original regular grid ($\Delta x = \Delta y$) is shifted four times in each direction,

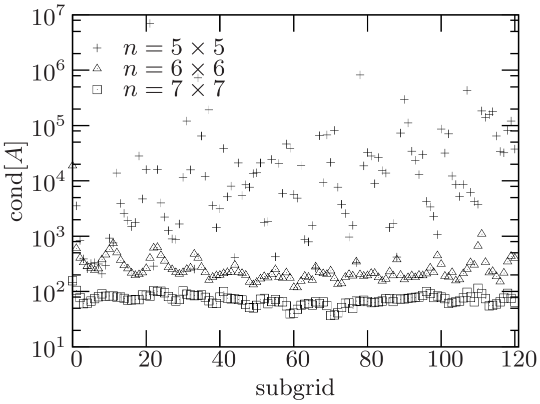
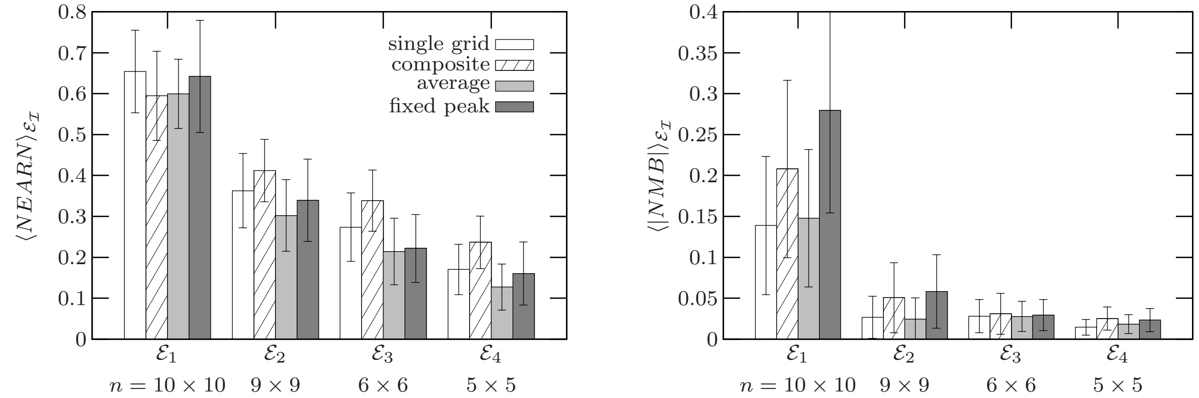


Figure 8.4: Condition number cond[A] = $\sigma_1/\sigma_{rank[A]}$, see eq. (4.27), for subgrids generated by shifting the original bilinear grid of dimension n ten times in x and y (c.f. fig. 6.2). Obviously, the variation of cond[A] within the subgrids is less distinct than the variation with n.

i.e. $m = m_x = m_y = 4$, resulting in $\sqrt{N} = (5(\sqrt{n} - 1) + 1)$ grid nodes in each direction for the final composite grid of higher resolution, see eq. (6.5b). The number of shifts is not critical in that choosing $m = 3$ does not drastically deteriorate the reconstruction results, while shifting five or six times does not improve them significantly – but increases calculation time considerably (according to eq. (6.5a) with $\sim m^2$). Strictly speaking, different conditioning of the matrix A for each subgrid should be taken into account by adjusting the iteration number accordingly as illustrated in fig. 8.4, but at least for the averaging and composite scheme it affects the subgrid reconstructions in the same way and should not matter for a comparison. Therefore the same iteration number for all subgrids is chosen.
First of all, general conclusions emerging from figs. 8.5a to 8.5d are:

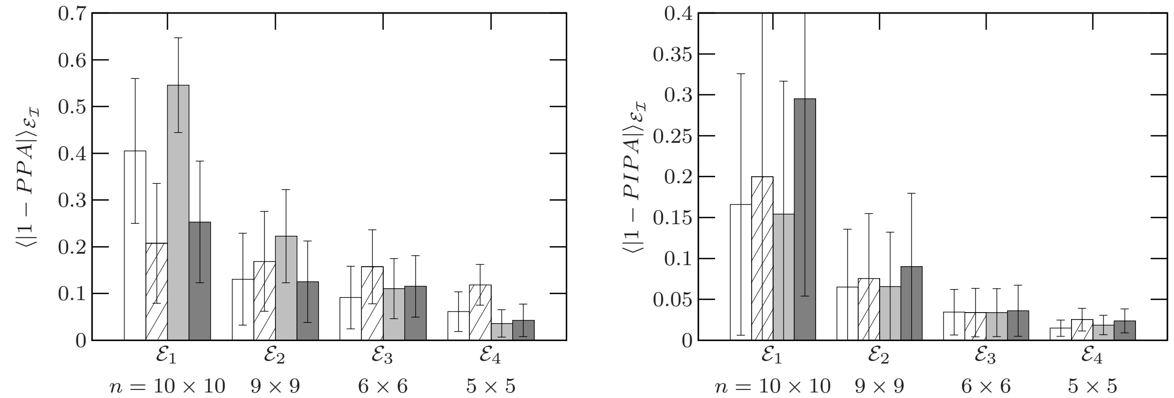
- The performance of each reconstruction scheme heavily depends on both the kind of concentration distribution (ensemble) and the quality criterion used for evaluation.
- The relative amount of change for different schemes (relative to the single grid reconstruction) varies strongly within the quality criteria.

The composite scheme reduces the overall root mean square error for narrow peaks ($\mathcal{E}_1$) by $\sim 8\%$ while the error of the reconstructed maximum concentration decreases by around 50% ! The fact that, at the same time the accuracy of mean concentrations and emissions decreases by 30% and 25%, respectively, can be understood as overestimation following from the construction of the scheme that combines maximum node values from different grids. It cannot be expected to conserve concentration integrals. Originally proposed by *Verkruysse and Todd* [2004] for better reconstruction of maximum concentration values in the context of indoor air contamination, the composite scheme thus works well for this purpose – as long as one is not interested in total amounts of emitted gas. The restriction to narrow peaks is not addressed in a systematic manner in the aforementioned study. The authors use piecewise constant parametrisation combined with the MLEM approach (c.f. sec.4.9.1) and consider Gaussian peaks with $\sigma/l \sim 0.07 \ldots 0.18$, roughly corresponding to peaks from $\mathcal{E}_2$ and $\mathcal{E}_3$. For the geometry with 40 integration paths shown in fig. 3.5h and four peaks they obtain mean values for the overall error $NEARN$ around 0.7 (single grid), 0.5 (translated grids) and for the peak prediction error $|1 - PIPA|$ around 0.35 (single grid) and 0.2 (translated grid). Comparing these to the values for $\mathcal{E}_2$ (to take the narrower peaks) in fig. 8.5a and 8.5c, it follows that already the reconstruction with a *single bilinear* grid leads to better results than the grid translation method for the box basis, underlining again the importance of optimal parametrisation. The composite scheme does not work reliably for peaks other than the narrow ones from ensemble $\mathcal{E}_1$.



(a) Nearness

(b) Normalised mean bias



(c) Error of the maximum peak prediction accuracy

(d) Error of the peak integral prediction accuracy

Figure 8.5: Ensemble averaged quality indices for different methods of combining reconstructions from several bilinear grids, again for four peaks within each ensemble and all settings as in fig. 8.1. For each ensemble the optimal grid in terms of $NEARN$ was chosen. The grid dimension n refers either to the actual grid dimension (when reconstructing from a single grid) or to the dimension of the original grid used to generate the shifted grids. In the latter cases the grid was shifted four times in each direction, see fig. 6.2.

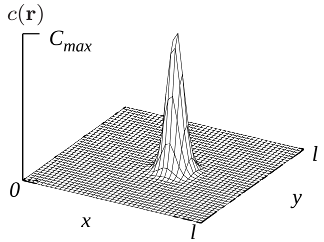
The averaging method reduces the mean square error for all peaks and makes hardly any difference for the concentration integrals – as to be expected for a linear method. It does underestimate the peak maximum though as anticipated in sec. 6.3.4, fig. 6.3, leading to lower peak prediction accuracy for the first three ensembles. For narrow peaks ($\mathcal{E}_1$) the averaging scheme increases the peak error by $\sim 50\%$, taking us right to the modified approach where the maximum grid node value among all subgrids is implemented as a constraint for each subgrid reconstruction. This scheme does indeed give better estimates for the absolute maximum concentration for peaks from $\mathcal{E}_1$ and $\mathcal{E}_2$ but also worse values for mean concentrations and total emissions, especially for very narrow peaks. This is not surprising because for these an enhanced peak concentration has relatively more impact on area integrals (see also fig. 8.2).

To conclude the statistic evaluation, it can be said that similar to the choice of the best reconstruction grid discussed before in sec. 8.2.1, selection of the optimal reconstruction scheme requires some idea about the extension of the peaks. The averaging scheme is safe in that it generally reduces the overall error, but at the expense of the peak reconstruction accuracy. For concentration peaks that lead to under-determined optimal parametrisations the composite scheme and the averaging method with peak constraint offer alternatives for better maximum values, where the composite scheme should be preferred for very narrow peaks. Both schemes should not be used if one is interested in total emissions. In fact, none of the multiple grid schemes discussed here improves the quality of reconstructed integration integrals.

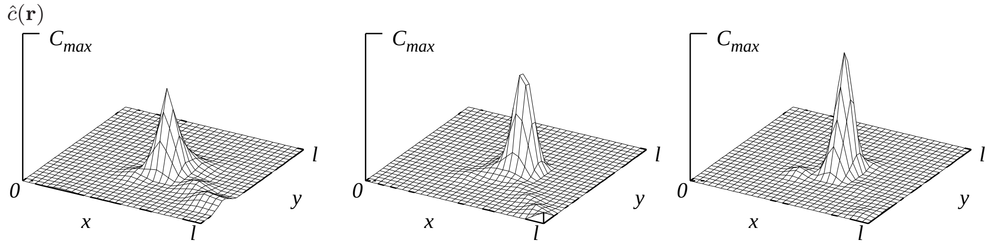
Finally, tables 8.3 and 8.4 illustrate the effect of the reconstruction schemes discussed.[1] Some characteristic features visible in all reconstructions like the peak reduction of the averaging scheme or the noisiness of the composite scheme follow directly from the construction of the method. Despite obvious shortcomings of individual schemes, it is hardly possible to make out an optimal method among all sample reconstructions and it appears that the best qualitative picture arises by combining the information from different approaches. Table 8.4a is a good example where use of a single grid (the 9 by 9 nodes grid in this case) can give a rather misleading picture. Another important aspect is contained in table 8.3b as mentioned earlier. *All* four reconstruction schemes show a decrease of the maximum and a broadening of the peak for the very fine $n = 13 \times 13$ grid which can only be interpreted as the a priori not working any more and the grid dimension being too large. This behaviour might thus be considered as a criterion for the choice of the parametrisation.

---

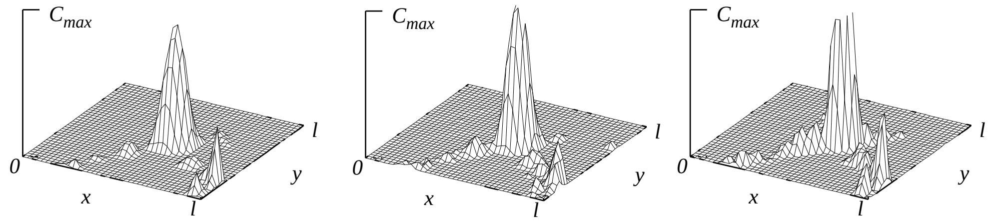
[1] 3-D surface graphs instead of, for example, 2-D contour graphs are used for this purpose simply because they visualise structures more clearly.



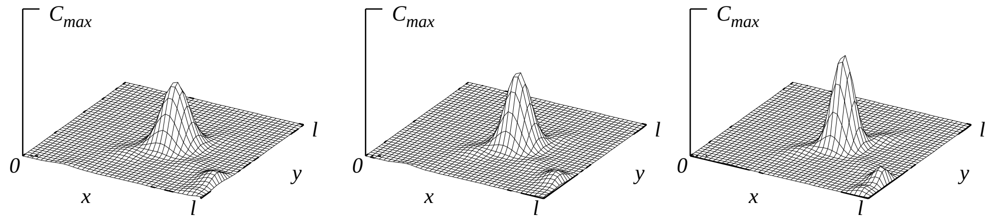
original ($\in \mathcal{E}_1$)



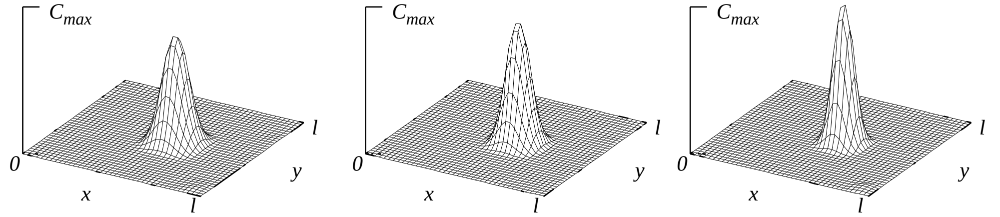
single grid: $n = 10 \times 10$ $n = 11 \times 11$ $n = 13 \times 13$



composite scheme: $n = 10 \times 10$ $n = 11 \times 11$ $n = 13 \times 13$



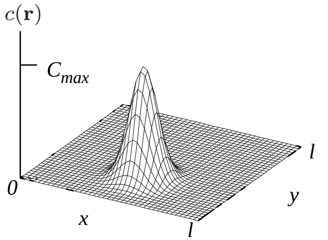
averaging scheme: $n = 10 \times 10$ $n = 11 \times 11$ $n = 13 \times 13$



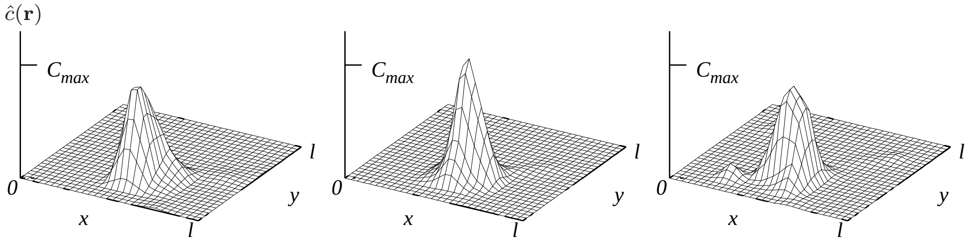
fixed peak scheme: $n = 10 \times 10$ $n = 11 \times 11$ $n = 13 \times 13$

(a) $c(\mathbf{r})$ from $\mathcal{E}_1$ with $\sigma_x/l = \sigma_y/l = 0.04$

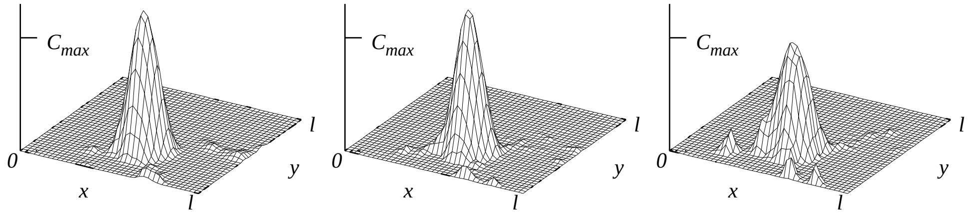
Table 8.3: Sample reconstructions using a single grid or one of the translation schemes, respectively, for one Gaussian from each ensemble and various grid dimensions. All reconstructions use bilinear parametrisation and SIRT. The grids are translated four times in each direction ($m_x = m_y = 4$).



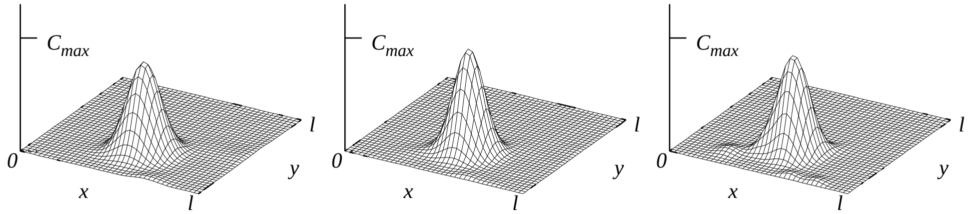
original ($\in \mathcal{E}_2$)



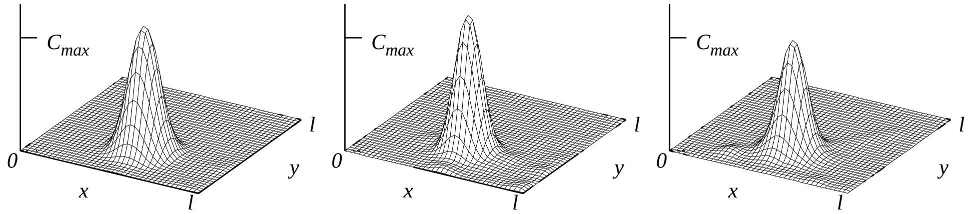
single grid: $n = 9 \times 9$ $n = 11 \times 11$ $n = 13 \times 13$



composite scheme: $n = 9 \times 9$ $n = 11 \times 11$ $n = 13 \times 13$



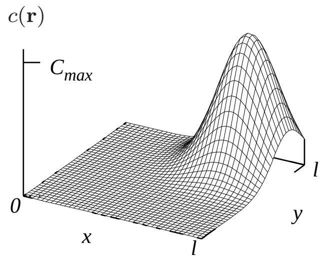
averaging scheme: $n = 9 \times 9$ $n = 11 \times 11$ $n = 13 \times 13$



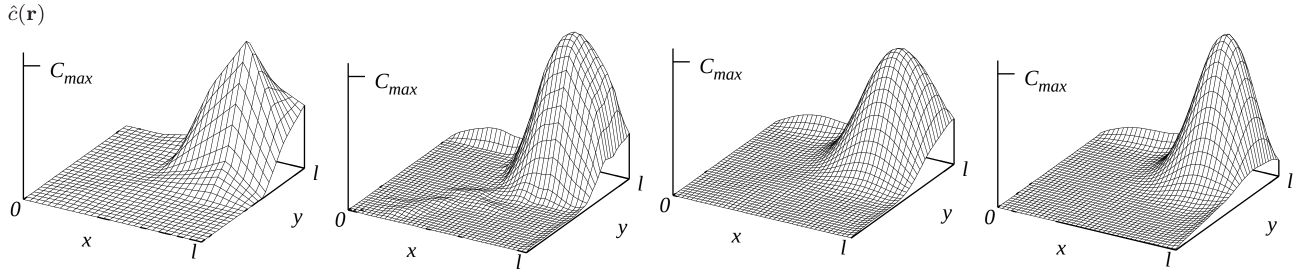
fixed peak scheme: $n = 9 \times 9$ $n = 11 \times 11$ $n = 13 \times 13$

(b) $c(\mathbf{r})$ *from* $\mathcal{E}_2$ *with* $\sigma_x/l = \sigma_y/l = 0.07$

Table 8.3: Sample reconstructions (continued).



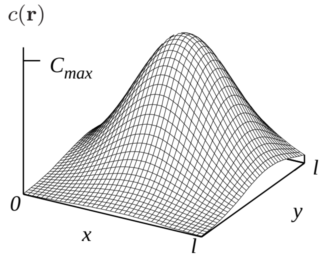
original ($\in \mathcal{E}_3$)



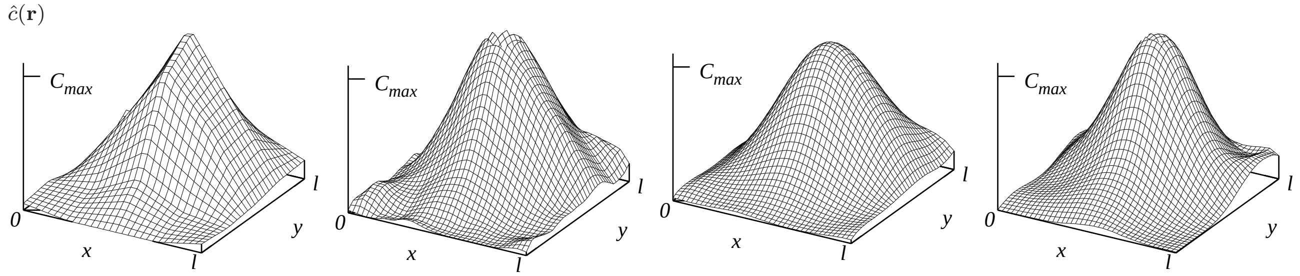
single grid, $n = 6 \times 6$ comp. scheme, $n = 6 \times 6$ averag. scheme, $n = 6 \times 6$ fix. peak scheme, $n = 6 \times 6$

(c) $c(\mathbf{r})$ from $\mathcal{E}_3$ with $\sigma_x/l = \sigma_y/l = 0.15$

Table 8.3: Sample reconstructions (continued). Only the even-determined grid is shown.



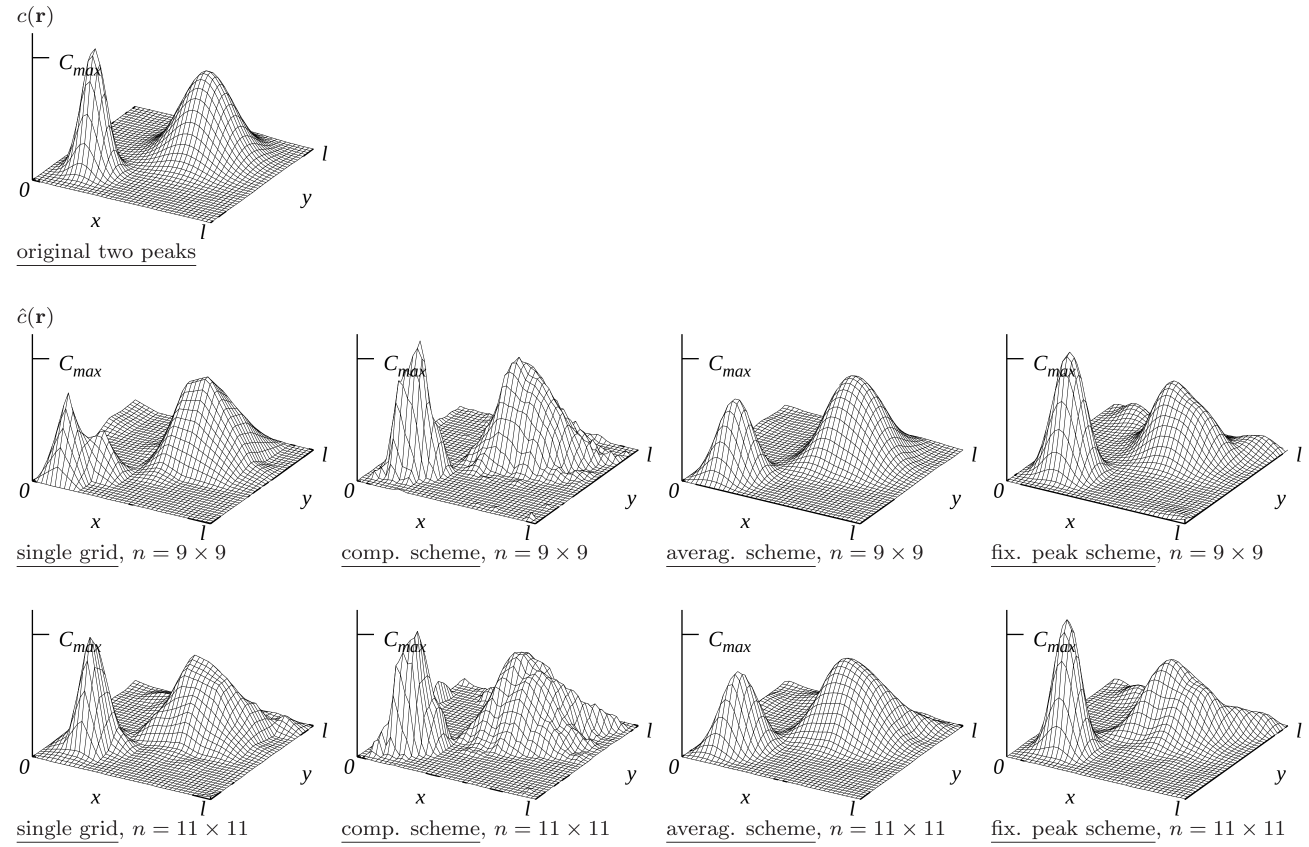
original ($\in \mathcal{E}_4$)



single grid, $n = 5 \times 5$ comp. scheme, $n = 5 \times 5$ averag. scheme, $n = 5 \times 5$ fix. peak scheme, $n = 5 \times 5$

(d) $c(\mathbf{r})$ from $\mathcal{E}_4$ with $\sigma_x/l = \sigma_y/l = 0.25$

Table 8.3: Sample reconstructions (continued). Only the under-detemined grid is shown (according to fig. 8.1a).

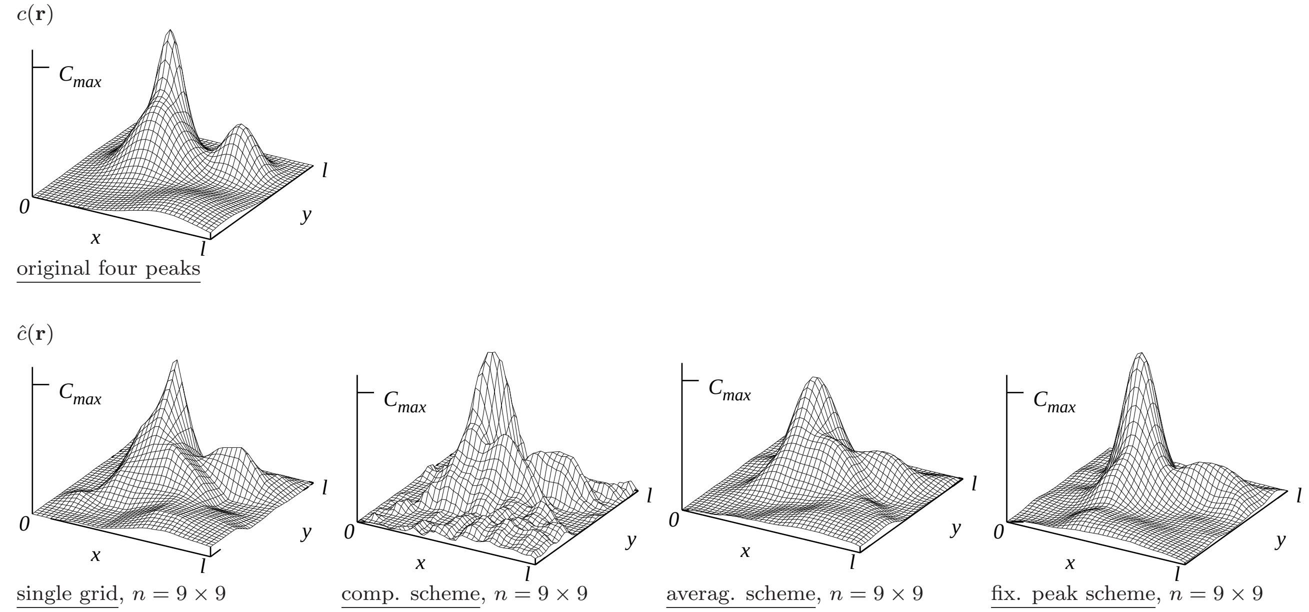


original two peaks

single grid, $n = 9 \times 9$ comp. scheme, $n = 9 \times 9$ averag. scheme, $n = 9 \times 9$ fix. peak scheme, $n = 9 \times 9$

single grid, $n = 11 \times 11$ comp. scheme, $n = 11 \times 11$ averag. scheme, $n = 11 \times 11$ fix. peak scheme, $n = 11 \times 11$

(a) Two Gaussians, one from $\mathcal{E}_2$ with $\sigma/l = \sigma_x/l = \sigma_y/l = 0.06$, $C_{max} = 1\,a.u.$, the other from $\mathcal{E}_3$ with $\sigma/l = 0.13$, $C_{max} = 0.7\,a.u.$, respectively.



original four peaks

single grid, $n = 9 \times 9$ comp. scheme, $n = 9 \times 9$ averag. scheme, $n = 9 \times 9$ fix. peak scheme, $n = 9 \times 9$

(b) Four Gaussians from $\mathcal{E}_2$ to $\mathcal{E}_4$ with $\sigma/l = 0.06$, 0.08, 0.13, 0.2 and $C_{max} = 1$, 0.5, 0.7, $0.2\,a.u.$.

Table 8.4: Sample reconstructions for several peaks. Settings of the reconstruction as in table 8.3.

8.3. Algorithms: ART versus SIRT

Solution of large over-determined least squares problems by row acting methods like ART and SIRT has been (and is still) subject of extensive investigation in computerised tomography and there is little need to go into much detail here. They have also been applied to small under-determined problems like the one at hand in indoor gas tomography, but some of the respective studies [*Drescher et al.*, 1996; *Todd and Ramachandran*, 1994*a*; *Laepple et al.*, 2004] lack systematics and mathematical background necessary for a general conclusion. Therefore, some results comparing the performance of ART and SIRT along the lines of the previous sections are presented, at the same time specifying details of the iteration. At last, the sensitivity to measurement noise is discussed.

8.3.1. Optimal iteration number and reconstruction quality

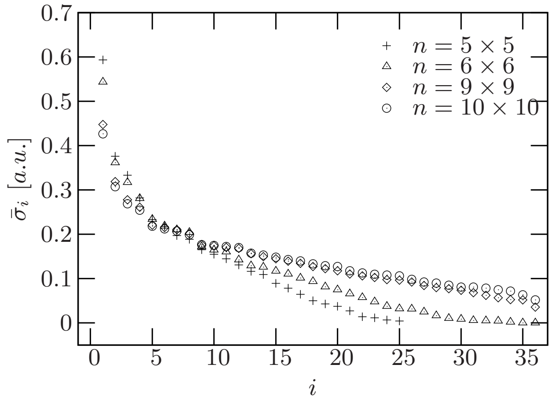


Figure 8.6: Singular values $\bar{\sigma}_i$ of the matrix $\bar{A}$ belonging to the rescaled system eqs. (4.71) *solved by ART and SIRT for regular bilinear grids of dimension n.*

The behaviour of the iteration numbers remains largely obscure in the aforementioned studies. According to sec. 4.8.2 we expect convergence like $\mathbf{x}^{(k)} \to \mathbf{x}^{\dagger}$, where $\mathbf{x}^{\dagger}$ is the (rescaled) least-squares minimum-norm solution, for error free data. The rate of convergence for the state vector components in the system of the singular vectors is given by eq. (4.76)

$$\frac{{x'}_j^{(k)} - {x'}_j^{(\dagger)}}{{x'}_j^{(0)} - {x'}_j^{(\dagger)}} = (1 - \omega\ \bar{\sigma}_j^2)^k, \quad j \leq \text{rank}[\bar{A}].$$

Rescaled singular values $\bar{\sigma}_i$ for various grid dimensions are shown in fig. 8.6, while the square of the right hand side was estimated in fig. 5.4a on page 91. Taking $\mathbf{x'}^{(0)} = \mathbf{0}$, iteration numbers around 10^3 are necessary for singular values of the order 0.01 to achieve relative deviations of the iterative solution from the ideal $\mathbf{x}^{\dagger}$ of about 10^{-2}. For narrow peaks the discrepancy between a priori and ideal solution will in general be smaller than for broader peaks, so that smaller iteration numbers suffice. Increasing the grid dimension will both lower the large singular values and enhance the smaller ones, where the former leads to higher, the latter to smaller iteration numbers. It turns out, that for the peak distributions considered here the iteration number increases with n. This is shown in tab. 8.5 which also confirms the predictions made before. The size of the iteration numbers for ART can immediately be estimated from the construction of the iteration cycles of ART and SIRT eqs. (4.67,4.68): One projection step for SIRT corresponds to m (here 36) projections for ART so that one expects the iteration numbers of ART to be by a factor of the order $1/m$ smaller than those of SIRT.

n	$\mathcal{E}_1$		$\mathcal{E}_2$		$\mathcal{E}_3$		$\mathcal{E}_4$	
	ART	SIRT	ART	SIRT	ART	SIRT	ART	SIRT
5 × 5	5-7^a	400 (400)	10-15^a	400 (400)	15	500 (500)	$\underline{30}$	$\underline{1\,500}$ ($\underline{5\,000}$)
6 × 6	15^a	500 (500)	15-20^a	500 (700)	20	500 ($\underline{800}$)	60	2 000 (6 000)
7 × 7	20-25^a	600 (600)	30	700 (900)	$\underline{40}$	$\underline{1\,000}$ (2 000)	60	3 000 (3 000)
8 × 8	35	1 000 (900)	35	1 000 (1500)	$\underline{60}$	$\underline{\text{-"-}}$ (3 000)	40	4 000 (1 000)
9 × 9	40	1400 (1500)	40	-"- ($\underline{2\,000}$)	80	2 000 (-"-)	-"-	-"-
10 × 10	$\underline{50}$	$\underline{1\,400}$ ($\underline{1\,800}$)	$\underline{50}$	$\underline{1\,200}$ ($\underline{2\,500}$)	120	3 000 (-"-)	50	5 000 (500)
11 × 11	$\underline{60}$	$\underline{1\,600}$ ($\underline{2\,200}$)	$\underline{80}$	$\underline{2\,000}$ (-"-)	125	4 000 (-"-)	60	-"-

---

[a] Semi-convergent case

Table 8.5: Approximate iteration numbers for which NEARN reaches convergence. The values are ensemble means for one Gaussian peak, in the case of SIRT numbers in brackets $(\cdot)$ refer to four peaks. Underlined figures indicate optimal grids in terms of nearness.

The case of convergence does not always apply to the error free case. Semi-convergence as discussed in sec. 4.8.2 can occur even without substantial noise (a fact that has not been acknowledged in the studies mentioned in the introduction). For the peak distributions considered so far, it occurs only for ART and over-determined systems – as one would expect – but with additional background or smooth distributions extending over the whole reconstruction area, semi-convergence is dominant for small dimensional grids. With increasing grid dimension the iteration behaves convergent again.

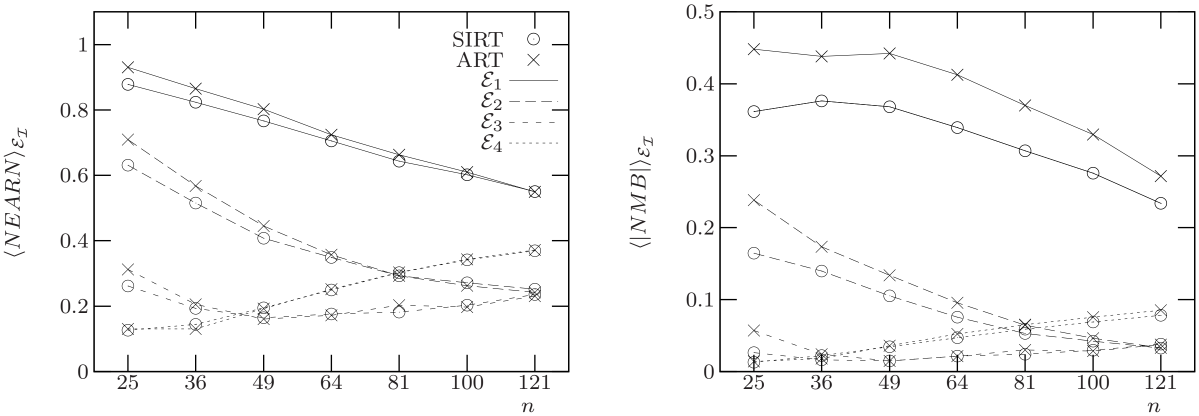
As far as calculation time is concerned, ART is preferable to SIRT. But, as realised also by *Todd and Ramachandran* [1994*a*] and *Laepple et al.* [2004], ART produces higher overall errors in terms of nearness, visible in more noisy concentration maps with a tendency to artefacts. Here, I just want to point out two further aspects following from evaluating the set of quality indices used before for ART and SIRT, respectively. Fig. 8.7 is analogous to fig. 8.5, here for one peak only, but the following conclusions remain valid for more peaks, too.

1. Combining complementary quality measures adds information. For example, from the fact that ART shows a significantly higher mean bias than SIRT (fig. 8.7b) while maxima and peak integrals are similar (figs. 8.7c,8.7d), it can be deduced that concentration is reconstructed outside the region of the peak, where originally there has not been any, i.e. aardvarks. This reliably confirms what has been guessed from inspecting a limited set of reconstructions.

2. Again, the systematic approach reveals that the performance for a specific setting of the reconstruction – in this case the algorithm – depends on the distributions. For smooth distributions ($\mathcal{E}_4$) ART and SIRT become very similar and, indeed, there are studies that report better results for ART compared to SIRT.
   More important for the applications of this work is the observation that the improvement by the optimal parametrisation (type of basis function, grid dimension) outweighs by far what can be gained by choosing one row acting method instead of another. This also holds for the MLEM algorithm as the quantitative discussion on page 125 has shown.

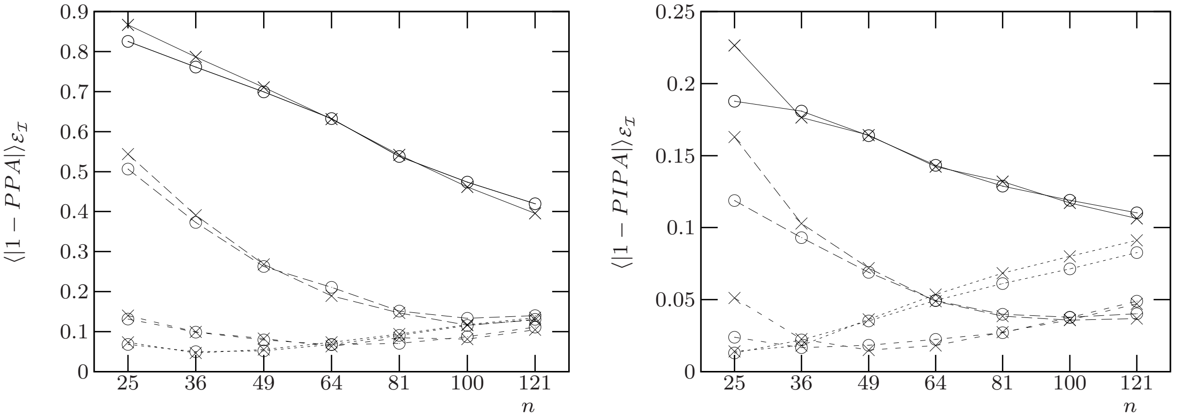
8.3.2. Sensitivity to noise

As found in the preceding section, there is no clear cut distinction between the error free convergent case and semi-convergence in the presence of noise. Following from the same discussion one expects



(a) Nearness

(b) Normalised mean bias



(c) Error of the maximum peak prediction accuracy

(d) Error of the peak integral prediction accuracy

Figure 8.7: Ensemble averages of different quality criteria for one random peak depending on the grid dimension for ART and SIRT. All settings as in fig. 8.1. Standard deviations are similar for both algorithms and have been omitted for clarity.

the under-determined case to be less susceptible to noise than the over-determined one which is usually dealt with in the literature. Here, I concentrate on the strongly under-determined case relevant for the measurements presented in the next chapter.

Fig. 8.8 illustrates the mean impact of unbiased random noise on ART and SIRT for a sharp concentration peak with $\varnothing_{2\sigma}/l = 0.05$ located randomly within the reconstruction area. Clearly, random noise affects ART more than SIRT, which is somewhat plausible because for the simultaneous iterative update corrections from each light path are added up, thus cancelling stochastic errors to some degree. While this observation remains generally true for other kinds of distributions, the very moderate effect of even large errors is special for the sharp peak distribution. Semi-convergence mildly occurs for errors of about 20% and only for ART. Nearness values increase for all levels of noise but the choice of the iteration number, i.e. the regularisation parameter, is not critical and its values can more or less be taken like in the error free case. Fig. 8.9 shows how noise deteriorates the reconstruction of plume properties like maximum concentrations and emissions for the same peak and SIRT. For random noise up to 10% – a realistic size for measurement noise – the impact stays quite moderate. The contribution of noise to the plume integral error can be estimated by means of eqs. (5.13b) as

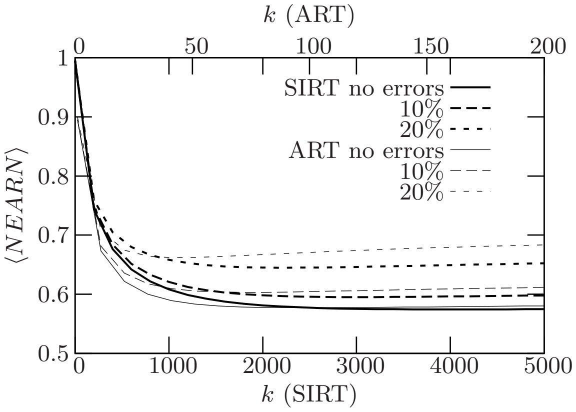


Figure 8.8: Sensitivity of SIRT and ART to random noise versus iteration number for a narrow concentration peak ($\varnothing_{2\sigma}/l \sim 0.05$) located randomly in the reconstruction area. For integration paths with zero concentration an absolute *error with variance of the size of the one with largest column density was assumed.*

follows (for the same iteration number the regularisation part in eq. (5.13a) is the same as in the error free case). Assuming that the plume extends over n_p grid nodes with the same perturbation $\Delta\bar{x}$ for each component, the left hand side of eq. (5.13b) takes the form

$$\mathrm{E}\big[\|\Delta\bar{\mathbf{x}}_{pert}\|_2^2\big] \sim n_p\, m\, (\Delta x)^2,$$

where the transformation eqs. (4.71) was taken into account. The scaling factor L for $\bar{\sigma}_\epsilon = \sigma_\epsilon/\sqrt{L}$ on the right hand side can according to sec. 4.8.4 be written as l^2/n_ϕ, with l being the characteristic length of a light path and n_ϕ the number of basis functions (boxes) contributing to it. Assuming furthermore a constant concentration level x within the plume of diameter $\varnothing$, one can rewrite the absolute error with the help of the relative data error r_ϵ as $\sigma_\epsilon = r_\epsilon\, d = r_\epsilon \varnothing\, x$. Putting the above LHS of eq. (5.13b) and the RHS in the form

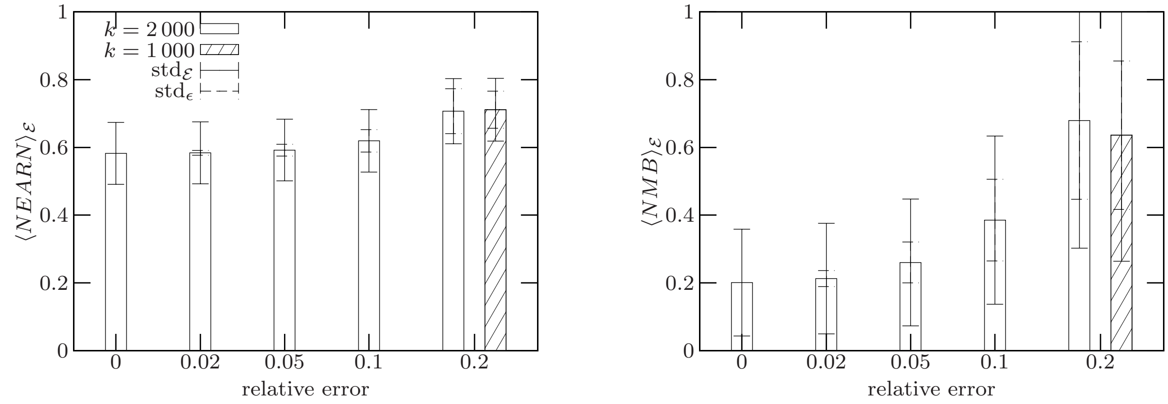
$$\bar{\sigma}_{\boldsymbol{\epsilon}}^2 \sum_i \Big(\frac{1-(1-\omega\,\bar{\sigma}_i^2)^k}{\bar{\sigma}_i}\Big)^2 \sim (r_\epsilon\,\frac{\varnothing}{l}\,x)^2\, n_\phi\, \sum_i \Big(\frac{1-(1-\omega\,\bar{\sigma}_i^2)^k}{\bar{\sigma}_i}\Big)^2$$

together leads to the following estimate for the error perturbation of the plume concentration

$$\frac{\Delta x}{x} \sim r_\epsilon \sqrt{\frac{n_\phi}{n_p\, m}}\,\frac{\varnothing}{l}\left(\sum_i \Big(\frac{1-(1-\omega\,\bar{\sigma}_i^2)^k}{\bar{\sigma}_i}\Big)^2\right)^{1/2}.$$

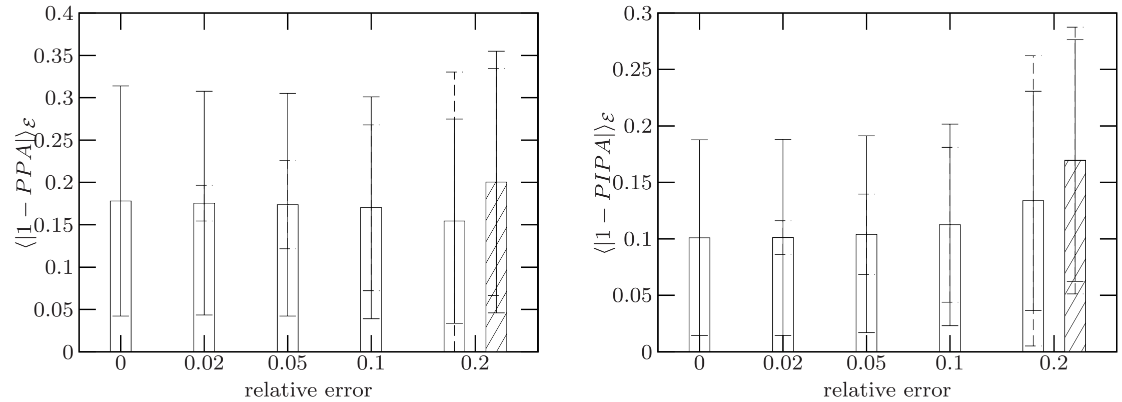
The sum involving the rescaled singular values takes for all discretisation grids shown in fig. 8.6 a value around $3\cdot 10^3$. The number of grid points covered by a plume with $\varnothing/l \sim 0.2$ on an 12×12 grid is about 6. Combining this with $n_\phi \sim 10$ and $m = 36$ results in a relative perturbation around $\Delta x/x \sim 0.2$ for data noise of 10%, agreeing nicely with the number in fig. 8.9d.

Two other aspects are worth mentioning. The first concerns the disproportionate sensitivity of the total area integrated concentration (NMB) in fig. 8.9b. This is because the simulation puts an error on all zero column densities, too. Due to the nonnegativity constraint, this can only lead to rising concentration values outside the plume while the data residuals increase. Secondly, quality measures and plume properties incorporated by them depend quite differently on the regularisation parameter. While the overall error is hardly affected by the iteration number, the other properties show ambivalent behaviour as indicated for the relative error of 20%. Stronger regularisation, i.e. smaller iteration numbers, will smooth the picture (see the discussion in sec. 4.6.3). This means smearing the peak and reducing the peak maximum, thus giving rise to higher errors of $PIPA$ and PPA.



(a) Nearness

(b) Normalised mean bias



(c) Error of the maximum peak prediction accuracy

(d) Error of the peak integral prediction accuracy

Figure 8.9: Impact of random noise on the overall reconstruction error, on the reconstrction mean concentration, maximum concentrations and total emissions for the sharp concentration peak from fig. 8.8. The standard deviation std$_{\mathcal{E}}$ refers to the variance within the reconstruction area, std$_{\epsilon}$ to the ensemble averaged variance from the random error.

8.4. Light path geometry

The discussion in the following is for bilinear parametrisation and a single reconstruction grid. Grid combination schemes are disregarded because a) for the larger class of peaks considered shortly they would not lead to significantly better results and b) for computational ease. Application of a grid combination scheme should not change the essential points.

8.4.1. Singular value decomposition of the geometry

Scalar quantities like the degrees of freedom (eq. (4.61)) and the information content (eq. (4.64)) of the measurement or the perturbation norm of the state vector (eqs. (5.12)) are governed by the singular *values* σ_i only. But it can be instructive to look at the singular *vectors* $\mathbf{u}_i$ (spanning the data space of the $\mathbf{d}$) and $\mathbf{v}_i$ (spanning the state space of the $\mathbf{x}$), relating the eigenmodes to the light paths and the reconstruction area, respectively, too – especially, for very small systems, see table 8.6.

The singular vectors belonging to the largest and smallest singular values for the geometry used in

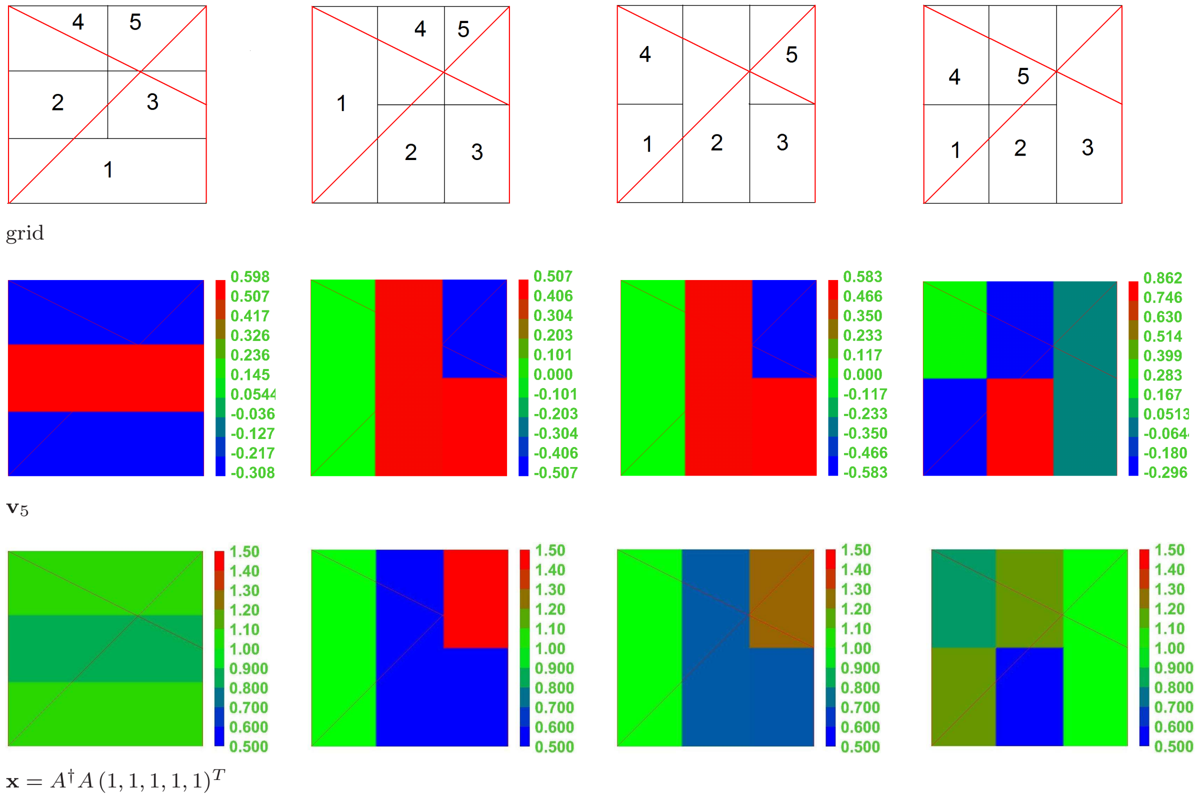


Table 8.6: Under-determined problem with four light (in red) paths and different parametrisation by five box basis functions (upper row). While the parametrisations are equally good in that they all lead to full rank systems with similar singular values, the – in this case 1-D – nullspaces look quite different (middle row) and leads to quite distinct retrievals (last row, for the generalised inverse and a homogeneous field). Mind that the nullspace does not contribute to box 1 for the 2^{nd} grid, to boxes 1 and 4 for the 3^{rd} and to box 3 for the 4^{th} grid, i.e. the retrieval in these boxes is exact.

the previous sections and a regular bilinear grid of dimension $n = 6 \times 6$ are shown in table 8.7. They show the oscillating behaviour of the 'small' singular vectors described as a general feature in sec. 4.5(v) and illustrate how regularisation smoothes the retrieval by cutting off small singular values and their vectors. The table also contains an example of how the map from the singular vectors $\mathbf{u}_i$ to the light paths can be used to identify certain modes, e.g. with small singular values. In this particular example the regular grid with $n = 6 \times 6$ shows an exceptionally high condition number (see fig. 8.4 in which it corresponds to the subgrid with number 0). The small singular value of the order 10^{-4} that is responsible for it (see the figure for $\mathbf{u}_{36}$) can clearly be attributed to some light paths whose arrangement for the given grid leads to an accidental linear dependency in the system matrix A. Different choice of the boxes containing these light paths can improve the situation.

Finally, table 8.8 indicates how the singular value decomposition can help with the construction of the grid for a given arrangement of light paths, especially if these cover the reconstruction area irregularly. Following the discussion in sec. 6.2.1, the choice between different grids for two geometries shown can be made by opting for the one which gives more balanced singular values. In both cases this choice give rise to a nullspace, more precisely the sum $\sum_{j'>\mathrm{rank}[A]} V^2_{jj'}$ according to sec. 7.2(i), that is more evenly distributed within the reconstruction area. This is indeed desireable if the a priori is noncommittal. More quantitative arguments need specification of the a priori and the noise level and will be presented

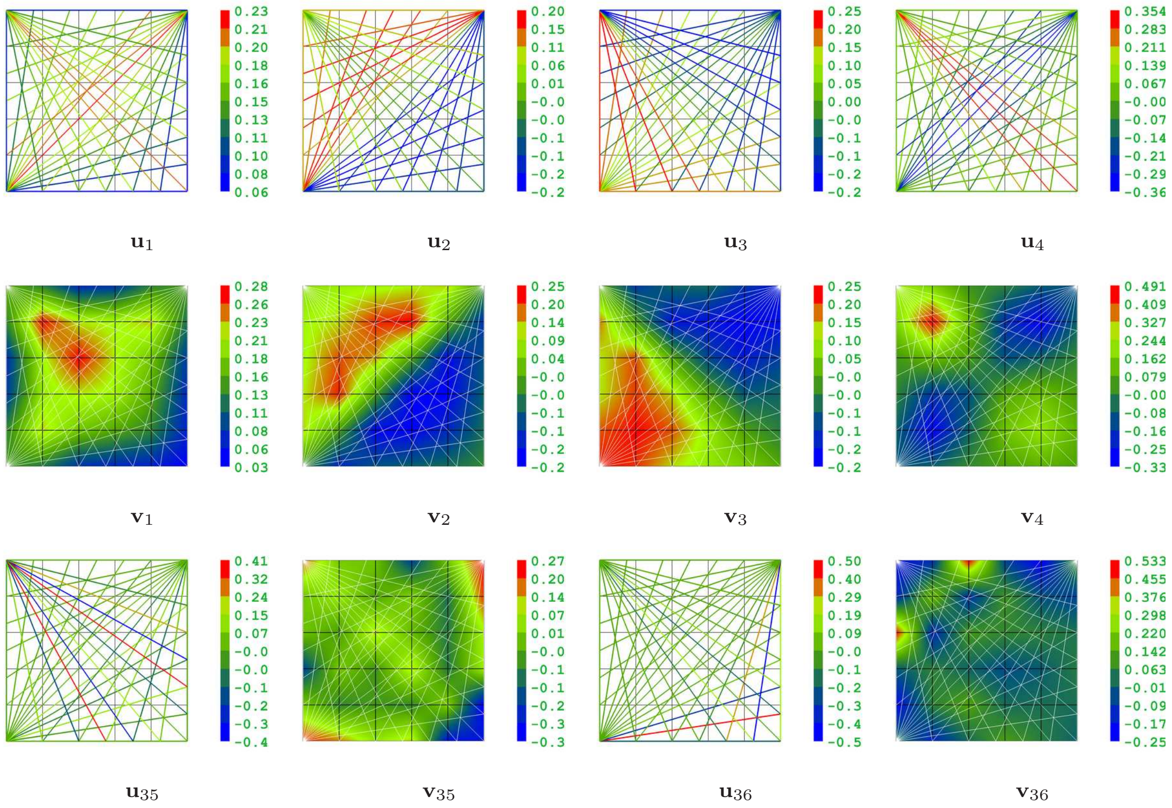
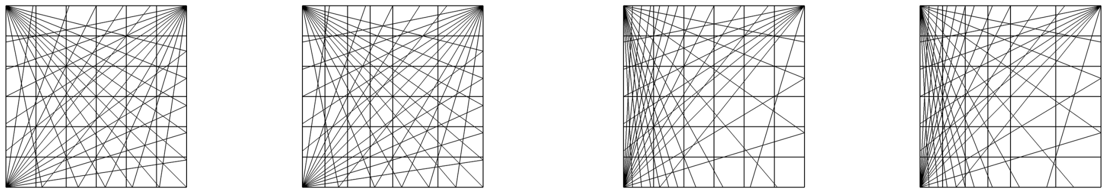
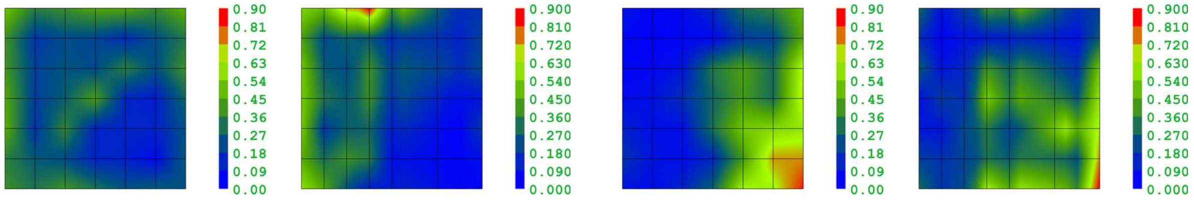


Table 8.7: Singular vectors $\mathbf{u}_i$ *and* $\mathbf{v}_i$ *spanning the space of column densities and state vectors, respectively, for the largest four and smallest two singular values of a regular bilinear grid of dimension* $n = 6 \times 6$ *(see also fig. 8.6). The components of* $\mathbf{u}_i$ *are mapped to the light paths while those of* $\mathbf{v}_i$ *are mapped to the grid nodes. Values in between the nodes are linearly interpolated.*

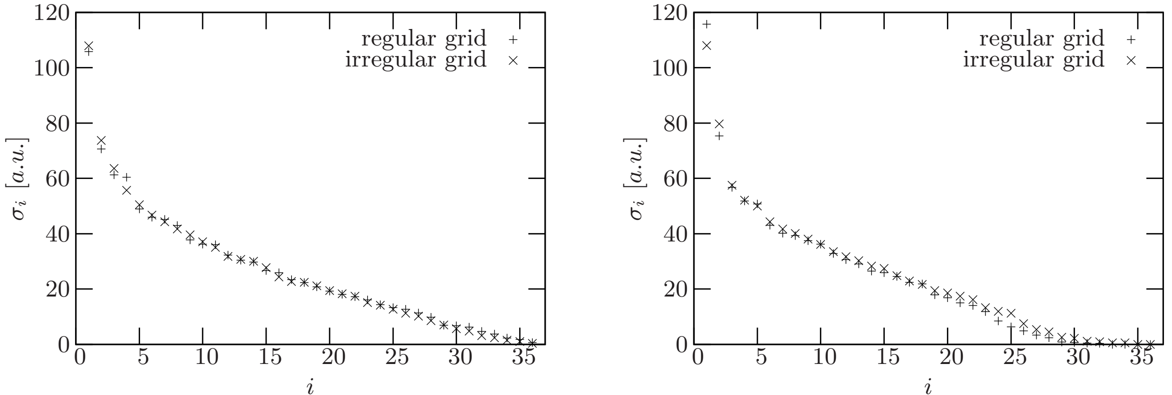
in the subsequent section.



'regular' geometry *irregular geometry*



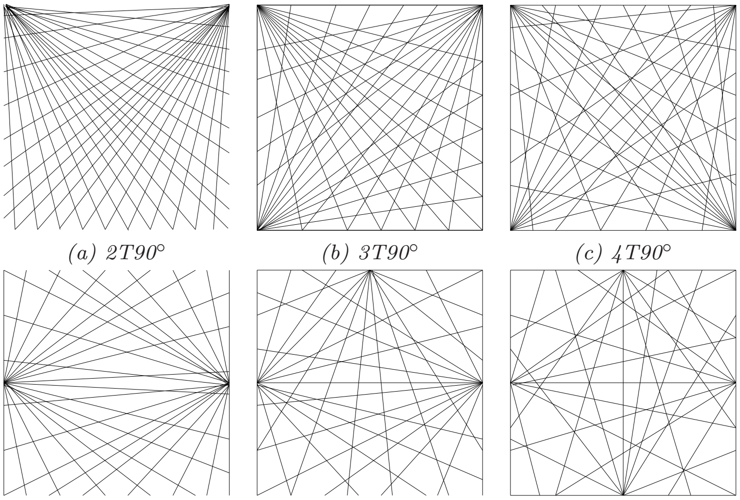
$\sum_{j'>\mathrm{rank}[A]} V_{jj'}^2$



singular values

Table 8.8: Nullspaces and singular values for two geometries and bilinear regular or irregular grids, respectively, of dimension $n = 7 \times 7$. *The vector* $\sum_{j'>rank[A]} V_{jj'}^2$ *is defined only on the grid nodes and values in between are merely interpolated. The singular values differ only slightly and the better balanced sets of singular values correspond to better balanced nullspaces.*

8.4.2. Comparison of different geometries



(a) $2T90°$ (b) $3T90°$ (c) $4T90°$

(d) $2T180°$ (e) $3T180°$ (f) $4T180°$

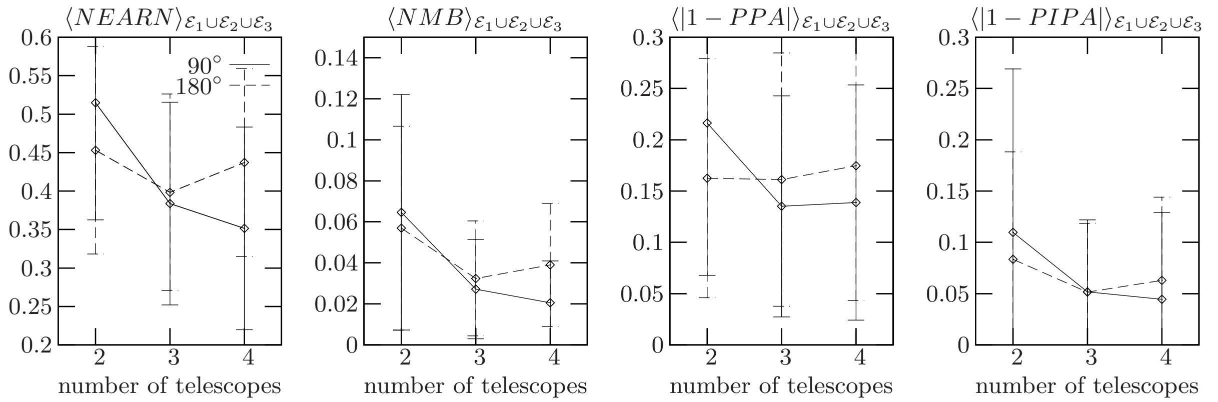
Figure 8.10: Geometries with two (a,d), three (b,e) and four (c,f) telescopes and light emitted in 90°(upper row)- and 180° (lower row)- fan beams. As before, the square area has a length of $l = 100\,a.u.$ *In all cases* $m = 36$. *The geometry labelled 3T90° is the one considered so far.*

From a simplified theoretical point of view the question for the optimal arrangement of a given number of integration paths has now been addressed on several occasions, especially in sec 7.2, and the answer basically is: Diagonalise the matrix A as much as possible, i.e. make the integration paths as linear independent as possible. Light paths get more independent if emitted (or received) by a larger number of instruments or in wider angles. In practice, both strategies will increase the mesh size according to eq. (7.1) and make the geometry more irregular. For peak distributions the resolution in terms of the mesh size becomes crucial and it is by no means obvious that linear independency outweighs a larger mesh size. Then again the answer to this question might depend on the level of noise, the impact of which is given largely by the singular values, i.e. by the linear independency. The problem was investigated by means of the geometries depicted in fig. 8.10, all generated by $m = 36$ light paths emitted (or received) in either 90°- or 180°-fan beams by between 2 and 4 telescopes. The light paths were chosen to cover the reconstruction area more or less regularly without claiming to present the ideal solution (for example, with respect to the number of retro-reflectors that would be needed to realise the light paths in an actual LP-DOAS measurement).

(i) First, the same statistical approach used earlier for the evaluation of different parametrisations and algorithms was chosen, this time for different geometries and random peaks taken from the union of the ensembles $\mathcal{E}_1$ to $\mathcal{E}_3$. These test distributions were chosen such that the mesh size should play a role and the same a priori (zero) could be used for all peaks. Results for four random peaks reconstructed on a single regular bilinear grid of dimension $n = 7 \times 7$ are shown in fig. 8.11. Differences between the geometries are substantial (the relative standard deviation lies between 15% for the overall error $NEARN$ and 50% for the error of the mean concentration), with a persistent trend showing that

1. increasing the number of emitting (or receiving) instruments generally enhances the reconstruction quality, provided that the coverage with light paths and their regularity can be maintained.

2. For the limited number of light paths considered here, large 180°-fans lead to worse results.



(a) Nearness *(b) Normalised mean bias* *(c) Error of the maximum peak prediction accuracy* *(d) Error of the peak integral prediction accuracy*

Figure 8.11: Ensemble mean values of the overall error (a) and the errors of mean concentration (b), maximum concentrations (c) and total emissions (d) for the geometries in fig. 8.10 and four peaks from ensembles $\mathcal{E}_1$ to $\mathcal{E}_3$. The reconstruction used a single bilinear grid with $n = 7 \times 7$.

3. The performance of the configuration with two telescopes and 90°-fans ($2T90°$) is exceptionally poor.

All in all the curves agree remarkably well with what one would expect from just looking at the geometries – except maybe for the geometry called $2T90°$. The light path arrangement $3T90°$ used in the preceding sections represents a good average of the geometries examined here.

To get an idea how the reconstruction quality varies within the reconstruction area for the individual geometries the same quality criteria as before were evaluated for one peak on a fine grid of peak positions (x_0, y_0) and mapped to the reconstruction area. The results are shown in the first three columns of table 8.9 on page 142. Apart from the pattern of smaller nearness values if peak positions coincide with grid nodes, there are spots of larger reconstruction errors where the coverage with light paths is coarse. The maps of the mean concentration and total emissions show this pattern, too. Both get underestimated in gaps between light paths and overestimated around the telescopes where mostly similar light paths from the same fan contribute. Taking the plots of mean concentration NMB and total emissions $PIPA$ together gives information about where the reconstructed trace gas is located: If the error of the mean concentration is distinctly larger than for the concentration within the peak, the concentration is spread over a larger region (as is the case near the telescopes) or artefacts appear somewhere else (which are here, for one peak, less pronounced).

The impact of random noise on the geometries was investigated for the very same peak distribution and a noise level of 10%. Similarly to the findings in sec. 8.3.2, the impact on the reconstruction quality for this special peak distributions is quite moderate. Furthermore, it turns out that the average aggravation is comparable for all geometries. This is shown for the overall error $NEARN$ in fig. 8.12, other quality measures behave similarly. For individual geometries, the sensitivity to noise does vary though within the reconstruction area, in particular for the special cases $2T90°$, $2T180°$ and $4T180°$, see tab. 8.10. While the geometries $2T90°$ and $2T180°$ (not shown) are less sensitive to noise in parts of the reconstruction area with poor coverage, the arrangement $4T180°$ shows higher sensitivity for these regions. An explanation for this opposed behaviour is that for the former two geometries the corresponding light paths are highly linear dependent (almost parallel at the top of

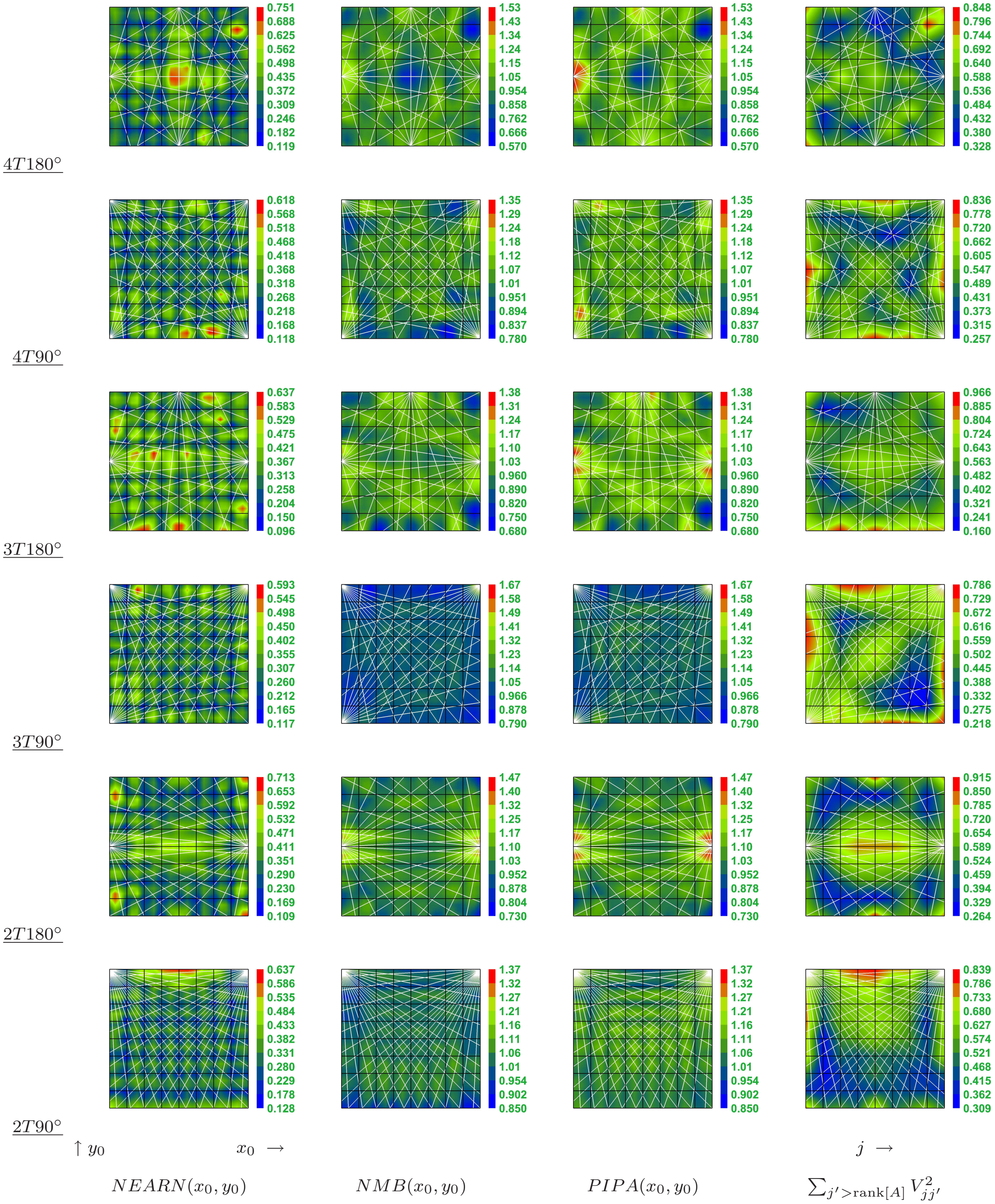


Table 8.9: First to third column: Overall error ($NEARN$), reconstructed mean concentration (NMB) and total emissions ($PIPA$) as function of peak position for the geometries in fig. 8.10. The peak from ensemble $\mathcal{E}_2$ with $\sigma_x/l = \sigma_y/l = 7$ was reconstructed for 30×30 positions using a single bilinear grid with $n = 9 \times 9$. Fourth column: Components j of the sum $V_{jj'}^2$ over the nullspace of A mapped to the nodes of the $n = 9 \times 9$ grid. Values in between grid nodes are merely interpolated

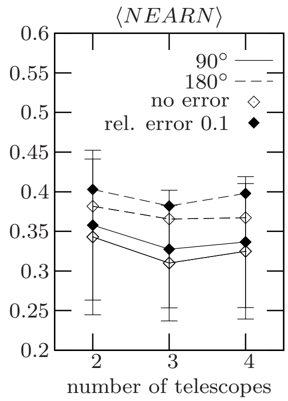
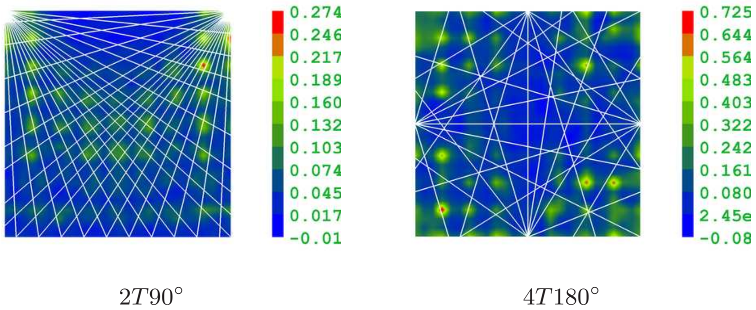


Figure 8.12: Mean nearness for reconstruction of the peak as in table 8.9 without and with 10% noise.

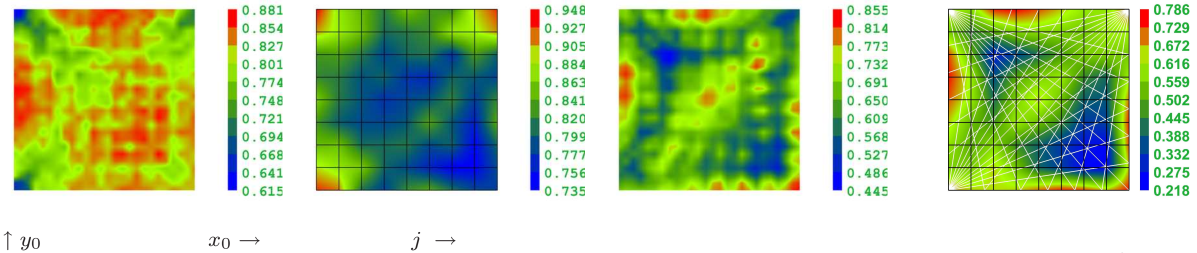


$2T90°$ $4T180°$

Table 8.10: Noise sensitivity of the nearness, defined as $(NEARN_\epsilon - NEARN)NEARN^{-1}$ as a function of the peak position (x_0, y_0) for reconstruction of the peak as in table 8.9 and selected geometries.

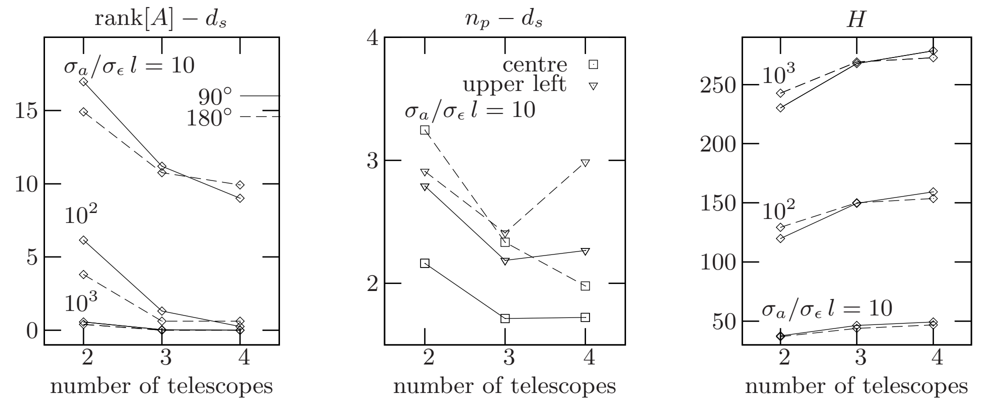
$2T90°$ and in the centre of $2T180°$) whereas in the latter case of $4T180°$ they are totally independent. In the first case the parts of the reconstruction area with poor coverage of light paths are related with small singular values and are thus particularly affected by regularisation, in the other case they are not. This is exactly what can be observed.

(*ii*) The results for the comparison of various light path arrangements have been obtained by explicit simulated reconstruction. Especially the investigation of the reconstruction quality within the area was restricted to a specific distribution. We are now going to show how the – mostly algebraic – arguments presented in sec. 7.2 can be used to arrive at similar conclusions, at least qualitatively, and start with an inspection of the nullspace according to eq. (7.2) like before in table 8.8. Results, given in the fourth column of table 8.9 for a bilinear grid of dimension $n = 9 \times 9$, show indeed a correlation between large values of $\sum_{j'>\mathrm{rank}[A]} V_{jj'}^2$, i.e. large nullspace, and high overall reconstruction error (mind again that the sum equals $1 - R_{jj}$, where R is the resolution matrix of the generalised inverse). The correlation is not very distinct and in the case of the geometry $3T90°$ hardly visible. But one has to bear in mind that the first three columns of the table refer to reconstruction which is a) nonlinear due to the nonnegativity constraint and b) for a specific distribution. The significance of both is nicely



(a) activity(x_0, y_0) *(b) average activity(j)* *(c) $\|\mathbf{x} - \mathbf{x}^\dagger\|_2 / \|\mathbf{x}\|_2 (x_0, y_0)$* *(d) $\sum_{j'>rank[A]} V_{jj'}^2$*

Table 8.11: Activity of the nonnegativity constraint for reconstruction of the peak distribution as in table 8.9 and the geometry $3T90°$. (a) As function of the peak position (x_0, y_0). A value of, e.g., 0.9 at (x_0, y_0) means that the constraint is active for 90% of the projections when the peak is at (x_0, y_0). (b) Activity on node j averaged over all reconstructions. (c) Difference norm of state vectors $\mathbf{x}$, $\mathbf{x}^\dagger$ for reconstruction of the peak at (x_0, y_0) with and without constraint. (d) Nullspace as in tab. 8.9.



(a) Uniform a priori variance

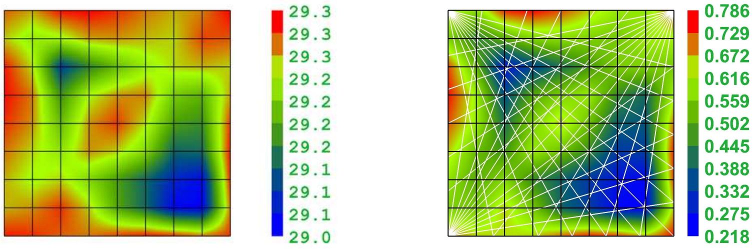
(b) Nonuniform a priori variance with values σ_a on $n_p = 3 \times 3$ nodes, zero else

(c) Uniform a priori variance

Figure 8.13: Degrees of freedom d_s and information H for the geometries of fig. 8.10 and different levels of noise. The bilinear reconstruction grid has dimension $n = 7 \times 7$. All covariance matrices are diagonal. The area covered by the n_p grid nodes in example (b) corresponds to about $0.3 \times 0.3\, l^{-2}$ and is located either in the centre or the upper left corner of the reconstruction area.

illustrated by table 8.11 showing the activity of the nonnegativity constraint during the reconstruction of the peak in table 8.9 for the geometry $3T90°$. It is evident from the figure that the constraint is active almost always and on all grid nodes with a tendency to 'fill up the nullspace'. That is, the larger the nullspace the more active the constraint becomes, thus explaining at least partly the difference in the patterns of overall reconstruction error and the nullspace. The strong nonlinearity of the inversion process following from table 8.11 is the reason why *quantitative* statements should rely on explicit numerical calculations.

Nevertheless, for a qualitative comparison of the geometries, we proceed with the linear approach and employ the Bayesian method as reasoned in sec. 7.2(*ii*) to see how individual geometries perform for a given level of noise and a supposed uncertainty of the true state. Using the degrees of freedom of the signal d_s, defined by eq. (4.61a) in sec. 4.7.2, as measures of performance means looking at the diagonal elements of the resolution matrix (or averaging kernel) R. We consider first a uniform a priori covariance $S_a = \sigma_a^2\, \mathbb{1}_n$ – corresponding to the case where the location of the peak within the reconstruction area is unknown and everywhere equally probably – and uncorrelated measurement errors of the same size σ_ϵ. In this case only the ratio σ_a/σ_ϵ enters. The limit $\sigma_a/\sigma_\epsilon \to \infty$ corresponds to the generalised inverse, for which $d_s = \mathrm{trace}[R] = \mathrm{rank}[A]$. Fig. 8.13a shows how d_s differ from the maximally possible 36 for different levels of noise – or better, for different degrees of ignorance about the true state vector. The parametrisation is the same as in fig. 8.11, i.e. a single bilinear grid of dimension $n = 7 \times 7$. Writing the variances in terms of relative errors r like $\sigma_a \sim r_a\, \overline{c_p}$, $\sigma_\epsilon \sim r_\epsilon\, \overline{c_p}\, \varnothing$, with a representative peak concentration $\overline{c_p}$ and peak diameter $\varnothing$, gives the estimate $\sigma_a/\sigma_\epsilon \sim r_a/r_\epsilon\, \varnothing^{-1}$. For the peaks considered in this section the ratio is thus of the order $(\mathcal{O}(0.1) \ldots \mathcal{O}(1)) \times r_a/r_\epsilon\ l^{-1}$ (see table 8.1). The curves show substantial variation of d_s for measurement with different geometries and a ranking very similar to the one found from simulated reconstructions in fig. 8.11a. Subareas of the reconstruction area can be compared by choosing the a priori variance correspondingly. To examine the degrees of freedom of a plume measurement, e.g. in the centre, one would assume a large a priori variance within this subarea and a small one outside. Fig. 8.13b illustrates this for a plume



(a) d_s *(b)* $\sum_{j'>rank[A]} V_{jj'}^2$

Table 8.12: (a) Degrees of freedom d_s *as function of the position of the point measurement for geometry* $3T90^\circ$ *and the special case that it lies on one of the nodes of the* $n = 9 \times 9$ *bilinear recontruction grid. All variances are diagonal and* $\sigma_a/\sigma_\epsilon\, l = 10$ *for all measurements. (b) Nullspace as in tab. 8.9. Values between the grid nodes are merely interpolations.*

that extends over $n_p = 9$ grid nodes either in the centre or the upper left corner of the reconstruction area. The a priori covariance is again diagonal with equal entries σ_a^2 for the nodes of interest and zero else. Differences of up to one degree of freedom (out of nine) for the given level of noise occur both with geometry and location of the plume. The figure suggests distinctly better reconstruction results for the centre areas of the 90°-fan geometries, in agreement with fig. 8.9. Finally, fig. 8.13c is the counterpart of fig. 8.13a for the information content as defined by eq. (4.64). Both agree qualitatively (mind that the real meaning of the information content lies in the number of states distinguishable by the measurement and given by 2^H), the same holds for the example with nonuniform a priori covariance (not shown).

The important point following from comparing the geometries by means of (i) the overall reconstruction error on the one hand and (ii) the resolution matrix (averaging kernel) on the other is not that they differ in details, but they agree so well. After all, evaluation of the former in continuous state space takes fully into account the shape of the distribution and its discretisation, while the latter takes place in discrete space, irrespective of the kind of distribution. This suggests that, in the end, it is the property of the linear independence that matters most, at least for the average of peaks taken into consideration here.

8.4.3. Point measurements

The conclusion in favour of the information concept at the very end of the preceding section justifies to use the degrees of freedom (or the information content) to analyse the optimal place of one or more additional point measurements in a given geometry of remote sensing paths by just comparing their values for different positions. It was argued in sec. 7.3 that the measurement location should be such that the associated equation(s) lie in the nullspace of A. Considering one measurement and treating it as additional constraint (case (ii) in sec. 7.3) gives values of d_s as shown in fig. 8.12 for reconstruction on a $n = 9 \times 9$ grid for the geometry labelled $3T90^\circ$. All covariance matrices were again assumed to be diagonal with equal entries and relative errors of the point and remote sensing measurement were taken to be of the same size. The measurement points were put on the nodes of the reconstruction grid for simplicity, but this should not restrict the conclusion that emerges when comparing fig. 8.12 to the nullspace of $3T90°$ in tab. 8.9. The Bayesian approach indeed predicts a maximum increase of d_s, if the point measurement takes place where the nullspace associated with the remote sensing measurements is largest. According to the last section, for noncommittal a priori this should hold equally for the reconstruction quality.

8.4.4. Scaling with light path number

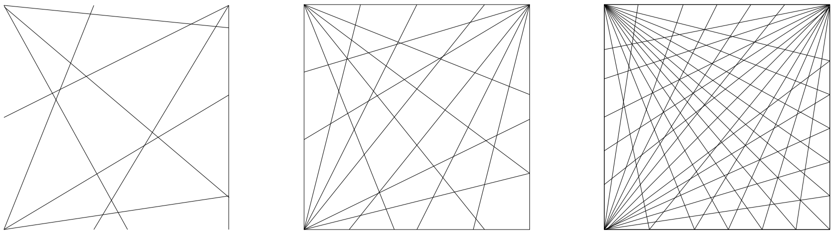

(a) 3T90° with $m = 9$ integration paths
(b) 3T90° with $m = 18$ integration paths
(c) 3T90° with $m = 36$ integration paths

Figure 8.14: The same generic type of geometry with light path numbers scaled by a factor $f = 2$.

Numerical results from the beginning of this chapter referred to a specific geometry ($3T90°$) with a fixed number of integration paths. The problem of transferring these numbers to arbitrary light path configurations can be stated like this: If, for a certain geometry g with m light paths, simulations involving test distributions c yield a value q for a quality Q, like the peak maximum precision, what will the value q' be for another geometry g' consisting of m' paths? I.e.,

$$Q\left(g(m), c\right) = q \quad \Rightarrow \quad Q\left(g'(m'), c\right) = q' = ?$$

Such a conclusion is easy, if the quality Q obeys simple scaling laws, for example with the meshsize Δl. But sec. 8.4.2, which dealt with $Q\left(g(m), c\right)$ for different geometries and fixed m, has shown that at least for the meshsize this is not the case. It was found that the generic type of the geometry, i.e. the linear independency of the light paths, is more important than the meshsize for the same light path number. Possible scaling with quantities related to the singular value decomposition of the geometry was not investigated. Instead, I show how quality measures scale with the light path number m for the *same* type of geometry.

The assumption goes like this: Scaling m with a factor f, so that $m' = f\,m$, gives for a more or less regular geometry a centre meshsize $\Delta l' = f^{-1}\,\Delta l$ (see eq. (7.1)). If the spatial dimensions of the concentration distributions are scaled by the same factor f^{-1}, the quantity Q should stay invariant. For Gaussian distributions with variances σ this takes the form:

$$Q\left(g(f\,m), c(f^{-1}\,\sigma)\right) = Q\left(g(m), c(\sigma)\right). \tag{8.3}$$

It was assumed here that Q has an appropriate normalisation, which is the case for all relative quality measures in table 5.1 on page 97, except for the normalised mean square error $NMSE$. Taking both test distribution c and retrieval $\hat{c}$ to be Gaussians, it is easy to see that the transformation $\sigma \mapsto f\,\sigma$ implies $NMSE \mapsto f^{-1}\,NMSE$.

$\varnothing_{2\sigma}/\Delta l$	m	n/m	$NEARN$	$NMSE$	IOA	$\|1-NMB\|$	r	$FA2$	$\|1-PIPA\|$
One peak									
	9	2.8	0.39	2.9	0.94	0.07	0.84	0.11	0.09
$1\dots2$	18	3.6	0.39	0.8	0.94	0.07	0.84	0.31	0.08
($\mathcal{E}_1$)	36	> 3.5	0.55	-	-	0.36	-	-	0.1
	9	1.8	0.22	0.08	0.99	0.03	0.95	0.70	0.03
$2\dots4$	18	2	0.24	0.32	0.98	0.04	0.95	0.36	0.05
($\mathcal{E}_2$)	36	2.3	0.25	-	-	0.03	-	-	0.04
	9	1.8	0.23	0.035	0.98	0.03	0.94	0.91	0.03
$4\dots8$	18	1.4	0.15	0.036	0.99	0.02	0.98	0.75	0.02
($\mathcal{E}_3$)	36	1.8	0.15	-	-	0.02	-	-	0.02
$8\dots12$	18	0.9	0.15	0.037	0.99	0.02	0.98	0.75	0.02
($\mathcal{E}_4$)	36	0.7	0.13	-	-	0.013	-	-	0.013
Four peaks									
($\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3$)	9	1.8	0.38	0.09	0.95	0.04	0.84	0.85	0.04
	18	1.4	0.37	0.14	0.95	0.03	0.85	0.75	0.04
	36	2.3	0.39	-	-	0.04	-	-	0.07

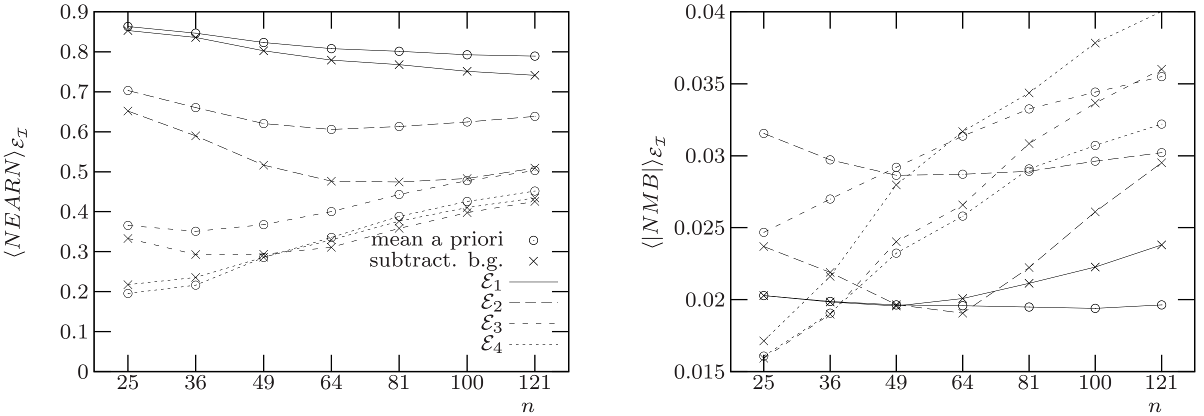
Table 8.13: Invariance of quality indices Q, if light path number and peak distribution are scaled according to eq. (8.3). The $\mathcal{E}_1$, etc. refer to the ensembles from table 8.1 for $m = 36$. All values are ensemble averages for one or four random peaks and reconstruction on a single, regular bilinear grid of dimension n. Some indices were taken into account later for $m = 9, 18$, but simulations were not rerun for $m = 36$. These values are missing.

To test the above scaling relation, we take again the geometry with three telescopes and 90°-fans $3T90^\circ$ and a scaling factor $f = 2$ as shown in fig. 8.14. The test distributions arise from the original ensembles $\mathcal{E}_1$ to $\mathcal{E}_4$ by scaling the variances with f, so that the ratio $\varnothing_{2\sigma}/\Delta l$ is the same within each ensemble for the geometries with 9, 18 and 36 light paths, respectively. In other words, σ/l ranges in $\mathcal{E}_1$ from 0.03 to 0.05 for $m = 36$ (see table 8.1), from 0.06 to 0.1 for $m = 18$ and so forth. Simulation results for reconstruction of one and four peaks on a single bilinear grid are given in table 8.13. For comparison with typical values of quality indices achieved by dispersion models that have been presented in table 7.1, it contains some indices beyond the ones used so far. Apart from confirming the expected scaling behaviour for all measures except $NMSE$, the table exhibits large variations in the sensitivity of different indices (for example, the index of agreement IOA seems to be rather insensitive). Whether the quality of the reconstructed peak distribution can keep up with dispersion models depends on the extension of the peaks relative to the meshsize (and the quality index), as the comparison with table 7.1 shows.

The invariance eq. (8.3) was found for other geometries, too, so that, in conclusion, the recipe for applying simulation results for peak distributions from one geometry to a similar geometry with a different number of light paths simply states:
Suitably normalised quality indices for peak distributions with the same ratio $\varnothing_{2\sigma}/\Delta l$ have equal values.

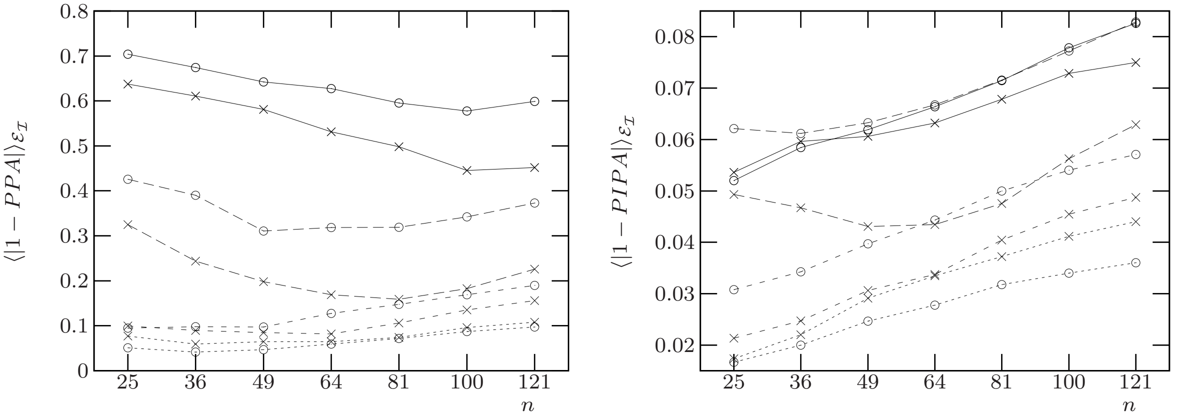
8.5. Background concentrations

The simulations in this chapter have assumed that the concentration values go down to zero (or arbitrary small values) far away from the peaks, i.e. that there is no background concentration in the sense of sec. 6.3.2. This precondition is of little significance for broad peaks, as long as the background is not too large. But for narrow peaks, the a priori $\mathbf{x}_a = \mathbf{0}$, supported by the nonnegativity constraint, has been seen to be essential. Instead of incorporating the background concentration into the a priori



(a) Nearness

(b) Normalised mean bias



(c) Error of the maximum peak prediction accuracy

(d) Error of the peak integral prediction accuracy

Figure 8.15: Same as fig. 8.1 for four random peaks and additional random background concentration (see text), reconstructed either using a mean priori or by subtracting the background.

$\mathbf{x}_a$, it was therefore suggested in sec. 6.3.2(*ii*) to use it as a nonnegativity constraint after subtracting the associated path integrals from the measurement data.

Both methods are compared again for random peaks from each ensemble and an additional non-uniform background distribution (see appendix B) with values varying randomly in the reconstruction area between 5% and 25% of the maximum peak value C_{max} and maximum gradients of about C_{max}/l.[2] Fig. 8.15 shows the results for the special case where the a priori is given by the mean concentration of all light paths eq. (6.7), while the value being subtracted is the minimum mean concentration eq. (6.14). Because the iteration number partly serves as regularisation parameter, it is given for completeness in table 8.14. Except for the broadest peaks from ensemble $\mathcal{E}_4$, the reconstruction errors obtained by subtracting the background are in almost all cases smaller than those using the mean a priori. Figs. 8.15a, 8.15b and 8.15d cannot be directly compared with the corresponding subfigures in fig. 8.1 for the case without background because, here, norm and integrals contain also the background. Details on how much the reconstruction of peak properties is deteriorated by the presence of a background concentration field will depend on its variability and thus the quality of the a priori (e.g., eqs. (6.10)). This has not been investigated systematically, but, at least for the situation

---

[2] For narrow peaks the contributions to the total column density from the peak and the background are thus of similar size. For broad peaks the contribution from the peak can be *maximally* 15 times that of the background.

n	$\mathcal{E}_1$		$\mathcal{E}_2$		$\mathcal{E}_3$		$\mathcal{E}_4$	
	subtract. b.g.	mean a priori	subtr.	mean	subtr.	mean	subtr.	mean
5×5	~ 200	~ 40	$\sim 400 - 500$	120	400	400	$\sim 2\,000$	2 000
6×6	-”-	~ 60	~ 800	150	800	600	$\sim 3\,000$	3 000
7×7	-”-	~ 80	-”-	300	2 000	-”-	$\sim 2\,000$	2 000
8×8	~ 400	~ 100	1 600	~ 400	-”-	500	~ 500	600
9×9	~ 500	~ 150	2 000	600	-”-	400	~ 400	-”-
10×10	-”-	-”-	-”-	-”-	-”-	-”-	-”-	400
11×11	-”-	150	-”-	400	-”-	-”-	-”-	-”-

Table 8.14: Iteration numbers for fig. 8.15. The '$\sim$' indicates optimal values of $NEARN$ for semi-convergence. Values without '$\sim$' mean that $NEARN$ has approximately reached convergence for this iteration number.

considered here, the peak maximum (fig. 8.15c) can be reconstructed with similar accuracy as in the case without background (fig. 8.1c). With the background distribution modelled here being rather unspecific, it seems thus fair to say that

1. Subtracting a moderately variable background field as formulated in eqs. (6.13), (6.14) rather than using a mean a priori in the least-squares minimum-norm scheme leads to significantly better results for narrow peaks (from ensembles $\mathcal{E}_1$ to $\mathcal{E}_3$), provided the solution is regularised appropriately.
2. For the 'differential' system that remains after subtraction of the background, the findings of the preceding sections apply, at least qualitatively. Details depend on the actual background variability.

8.6. Discussion

Some aspects concerning the comparison with existing studies are discussed here in brief, while a further, concluding discussion is held in chap. 11. Unlike for routine applications of computerised tomography (CT), there is not a standard problem and no common way of evaluating reconstruction methods in atmospheric science. For applications of CT in indoor gas dispersion, which come closest to our kind of reconstruction problem, there are no publications by different authors among all referred to in this thesis that use the same evaluation method (Despite the fact that there *is* a large set of standard metrics used in atmospheric modelling, see sec. 5.4).

Furthermore, none of these indoor studies examines the parametrisation systematically, neither in terms of the number of parameters, nor in terms of basis functions (Although most of them deal with Gaussian peaks as well, where these issues matter). The possibility of highly under-determined parametrisation is special to narrow peaks on a negligible or, as I have suggested, sufficiently smooth or known background. The fact that for this special kind of distribution the representation by bilinear basis functions not only leads to smaller discretisation errors, but also inversion errors, does not seem to be a regularisation effect because, in general, the conditioning of A for bilinear functions is not better than for piecewise constant ones. On the other hand, for over-determined discrete inverse problems, e.g. in atmospheric profiling, the impact of the basis functions on the quality of the retrieval has been attributed to different regularisation behaviour (e.g., [*Doicu et al.*, 2004] and references therein). For either case, the choice of parametrisation of the discrete inverse tomographic problem certainly requires further attention.

Most of the tomographic studies dealing with problems similar to ours are concerned with algorithms, i.e. either with inversion algorithms for the same reconstruction principle (like ART and SIRT) or with different principles altogether (like the least-squares and the maximum entropy principles). In many cases, the comparison with other algorithms remains insufficient or unsystematic with respect to the kind of distribution (However, it was argued in sec. 8.3.1 that the kind of distribution does play a role, explaining some seeming contradictions in the literature). For example, *Drescher et al.* [1996] finds better results for her method SBFM (see sec. 4.9.3) compared to ART on the grounds of less noisy concentration maps, but does not compare it to SIRT which is known to give less noisy results. *Price et al.* [2001] introduce a method, basically the Tikhonov approach with the third differences used for regularisation, but compare it only to SBFM using the data residual. The first method being over-determined and their discrete parametrisation being under-determined, there is hardly a point in comparing the data residual as argued in sec. 5.4. Furthermore, the authors do not show any sample reconstruction with original distribution. Doing this for Gaussian peaks suggests that the method is not only smoothing, as intended by the authors, but oversmoothing in terms of sec. 4.6.3[3]. *Samanta and Todd* [1999] have found that the MLEM principle (sec. 4.9.1) gives better results for the reconstruction of Gaussian peaks than the least-squares least-norm principle. This algorithm combined with box basis functions and a grid translation scheme was quantitatively compared to our approach on p. 125, in favour our method.
This does obviously *not* mean that the pure least-squares least-norm reconstruction principle is the ultimate choice. In fact, for over-determined problems, experience from 1-D profile reconstruction suggests that it is not. But according to our findings, any (new) principle or inversion algorithm should be compared to others systematically with respect to the kind of distribution and parametrisation. The same holds for the regularisation approach.

Light path geometries have been investigated by *Todd and Bhattacharyya* [1997]; *Todd and Ramachandran* [1994*b*] for indoor gas reconstruction, but – apparently inspired by X-ray tomographic setups – they use an excessive number of $120 - 400$ light paths, combining fan beams and projection paths by additional mirrors. By and large, carrying out numerical simulations similar to sec. 8.4.2(i), the authors come to the same conclusions, except for our geometries that are especially irregular.

---

[3] Cehlin, M., Computed tomography for gas sensing indoors using a modified low third derivative method. Submitted to *Atmospheric Environment*.

9. 2-D Reconstruction of NO_2 Peaks from an Indoor Test Experiment

The indoor experiment referred to in this chapter was carried out to study the performance of the multibeam instrument, especially developed for tomographic measurements [*Mettendorf*, 2005; *Pundt and Mettendorf*, 2005]. The experimental conditions are highly artificial (see [*Mettendorf et al.*, 2006; *Mettendorf*, 2005] for details of the experiment and its complications), therefore it is not used here to discuss atmospheric measurements of trace gas plumes, but rather to illustrate the reconstruction procedure for an experimental situation *without* further knowledge about the true concentration distribution. Only one of the distributions measured and reconstructed is considered in sec. 9.2, as variations basically are contained in the simulations of the previous chapter (see [*Mettendorf et al.*, 2006; *Mettendorf*, 2005] for a discussion of all results). In particular, the reconstruction error is estimated and re-estimated in sec. 9.3 in a consistent way from the reconstructed distributions and the simulations of chap. 8. Finally, sec. 9.4 examines the detectability of trace gas plumes in general and for the largest point sources near Heidelberg for illustrative purposes and as a case study for tomographic measurements currently taking place in Heidelberg.

9.1. The experiment

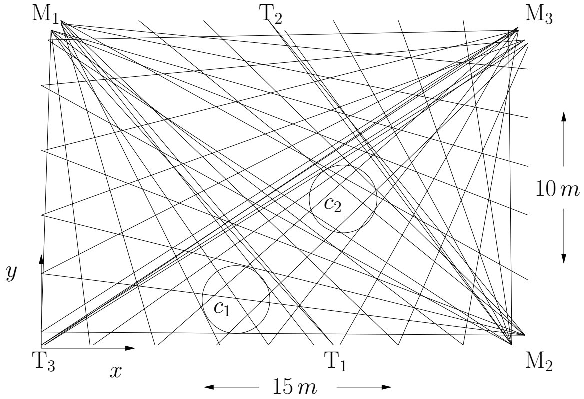


Figure 9.1: Light path geometry of the indoor experiment with 39 rays in total. Beams, simultaneously emitted by the telescopes (T), are redirected by mirrors (M) towards individual retro-reflectors. The test distributions consist of polycarbonate cylinders with radius $1\,m$ *and* NO_2 *concentrations* $c_1 = (1.95 \pm 0.15)\,10^{14} molec/cm^3$, $c_2 = (3.35 \pm 0.1)\,10^{14} molec/cm^3$.
(This arrangement of the cylinders was labelled distribution no. 16 in [Mettendorf et al., 2006; Mettendorf, 2005].)

The experiment was carried out on a factory floor on an area of $15 \times 10\,m^2$ with three of the multibeam instruments. Emission puffs were simulated by placing one or two polycarbonate cylinders, each filled with NO_2 and a radius of $1\,m$, at various positions in the area (fig. 9.1).

The design of the light path geometry was mainly dictated by instrumental factors. Light beams leave the multibeam telescope basically in the same direction and have to be redirected by external mirrors (one per beam) to the final destination. To appear separately on the mirrors the rays have to pass a

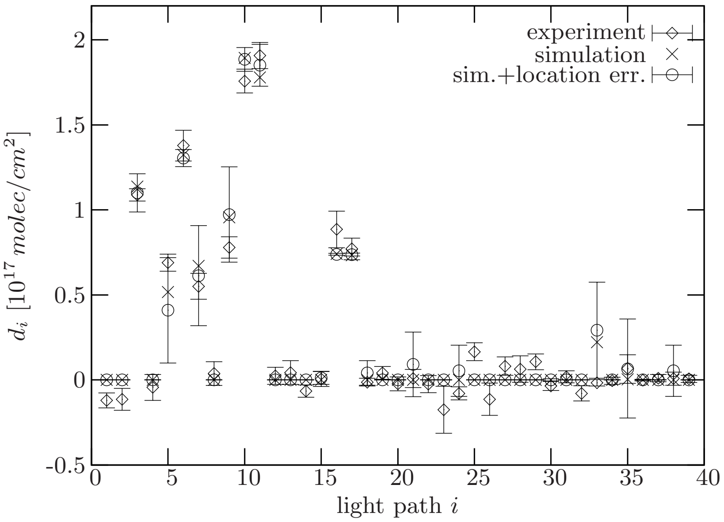
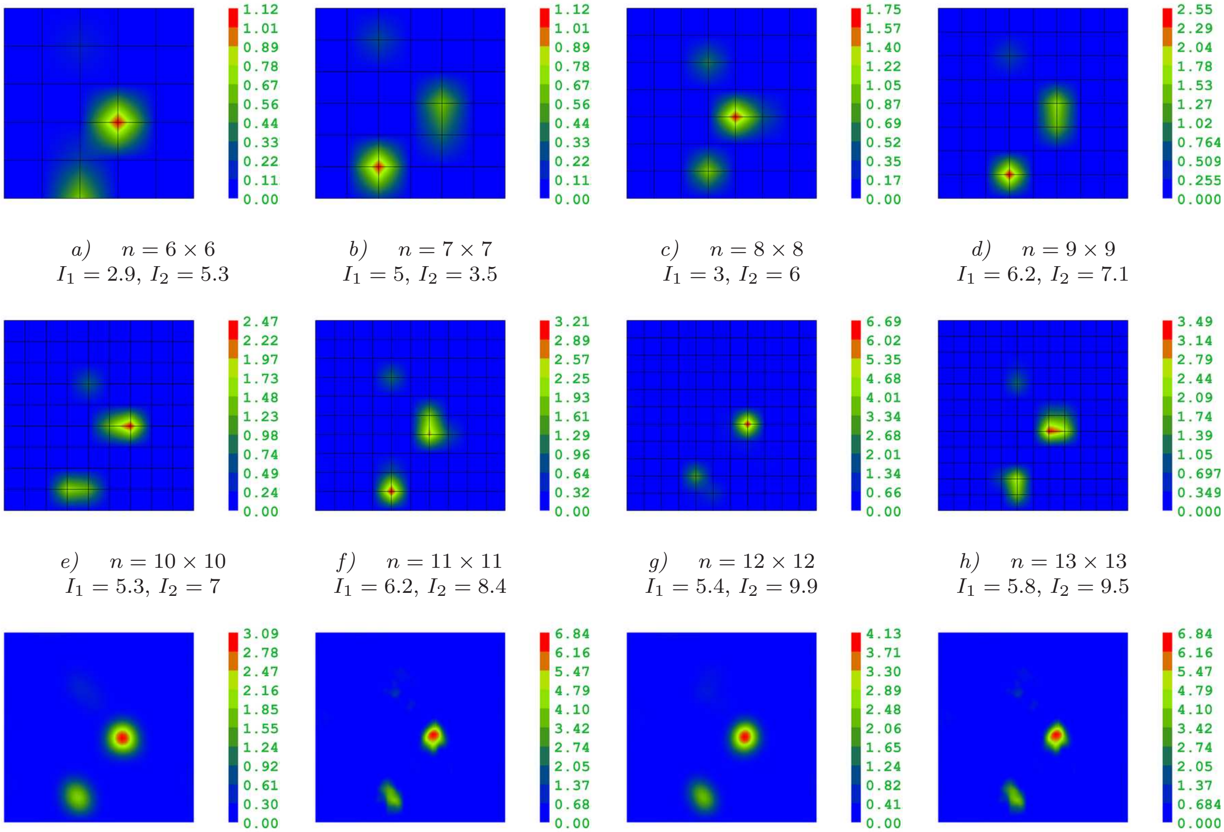


Figure 9.2: NO_2 column densities obtained from the experiment [Mettendorf et al., 2006; Mettendorf, 2005] and from integration over the known test distribution of fig. 9.1. The optical and integration path are twice the way from the telescope to the reflector. The location error in the simulation takes into account an uncertainty of the cylinder position of about $5\,cm$.

distance of several metres. A further restriction is given by the fact that each mirror can be turned horizontally by maximally ±30°, resulting in fans of maximally 120°. These requirements led to the arrangements of telescopes and mirrors as shown in fig. 9.1, i.e. three 90°-fans sitting in the corners of the test area. The number of light paths that can be realised at the same time is given by the number of rays emitted by each telescope (four in this case). To increase the number of light paths in atmospheric measurements, the external mirrors are successively directed to different targets, so that the final geometry used for reconstruction consists of more integration paths, yet not all referring to the same time (see the discussion in sec. 5.2.3). This scanning was carried out in the indoor experiment in three steps, each with 3×12 light paths, so that the complete geometry comprises 36 paths. All beams emitted by the same telescope share the distance TM within the reconstruction area, which would not be the case in real atmospheric measurements – or more precisely, this distance would be completely negligible compared to the remaining distance to the reflector. Besides, it is unfavourable for the tomographic reconstruction. Therefore, one optical path from the telescope to a retro-reflector at M has been added for each telescope, thus removing the ambiguity introduced by the common path TM. The final composite geometry with 39 light paths is essentially identical to the one labelled $3T90°$ in chap. 8. The three intermediate geometries with 12 light paths (see [*Mettendorf et al.*, 2006; *Mettendorf*, 2005] for details) were set up under static conditions, i.e. the concentration distribution did not change. Contrary to the geometry with 12 light paths in fig. 8.14, they were not intended to serve for reconstruction individually, so only the complete geometry of fig. 9.1 will be considered in the following.

Due to the short distances and the polycarbonate in the optical path, the DOAS measurement and its analysis differ considerably from an atmospheric measurement (see again [*Mettendorf et al.*, 2006; *Mettendorf*, 2005] for details in the following). In general, column densities obtained from the DOAS analysis agree well with expectations, as illustrated by fig. 9.2 for the concentration distribution with two puffs shown in fig. 9.1. In the simulations the finite beam diameter of around $6-30\,cm$ is neglected and the error of the experimental column densities is given by eq. (3.6), without taking into account any systematic errors, e.g. like the one coming along with an incorrect absorption cross section of NO_2. The average measurement error for all test distributions is about $0.05 \cdot 10^{17}\,molec/cm^2$ for beams passing through a cylinder and $0.025 \cdot 10^{17}\,molec/cm^2$ for beams that do not. For 'nonzero'

$\hat{c}(x,y)$ $[10^{14} molec/cm^3]$



a) $n = 6 \times 6$
$I_1 = 2.9$, $I_2 = 5.3$

b) $n = 7 \times 7$
$I_1 = 5$, $I_2 = 3.5$

c) $n = 8 \times 8$
$I_1 = 3$, $I_2 = 6$

d) $n = 9 \times 9$
$I_1 = 6.2$, $I_2 = 7.1$

e) $n = 10 \times 10$
$I_1 = 5.3$, $I_2 = 7$

f) $n = 11 \times 11$
$I_1 = 6.2$, $I_2 = 8.4$

g) $n = 12 \times 12$
$I_1 = 5.4$, $I_2 = 9.9$

h) $n = 13 \times 13$
$I_1 = 5.8$, $I_2 = 9.5$

i) $n = 11 \times 11$, aver.
$I_1 = 5.6$, $I_2 = 8.9$

j) $n = 11 \times 11$, comp.
$I_1 = 5.6$, $I_2 = 9.8$

k) $n = 13 \times 13$, aver.
$I_1 = 5.8$, $I_2 = 9.0$

l) $n = 13 \times 13$, comp.
$I_1 = 5.6$, $I_2 = 9.8$

Table 9.1: Reconstruction of the column densities of fig. 9.2 using SIRT, $\mathbf{x}^{(0)} = \mathbf{0}$*,* $k = 2\,000$*, a single bilinear grid (a-h) and the averaging (i,k) or composite (j,l) grid translation scheme.*
I_1 and I_2 are given by the integrals over peaks 1 and 2: $I_i = (10^{18} molec/cm)^{-1} \int_{\mathcal{P}eak_i} dA\, \hat{c}(\mathbf{r})$.

column densities in fig. 9.1 this amounts to relative errors between 3% and 10%.

9.2. Sample reconstruction

The reconstruction procedure is illustrated for the measurement shown in fig. 9.2, using the simulation results of the previous chapter *but no information other than what is provided by the measured column densities themselves.*

The fact that most of the column densities have zero (or negligible values) suggests the a priori $\mathbf{x}_a = \mathbf{0}$ as iteration start $\mathbf{x}^{(0)}$. Starting the reconstruction with a single regular bilinear grid of dimension $n = 6 \times 6$ – i.e. the over-determined case – leads to the concentration map shown in tab. 9.1a[1]. According to table 8.5 an iteration number of several thousand steps should be chosen to assure convergence. Assuming the convergent and not the semi-convergent case which needs regularisation is justified by the following.

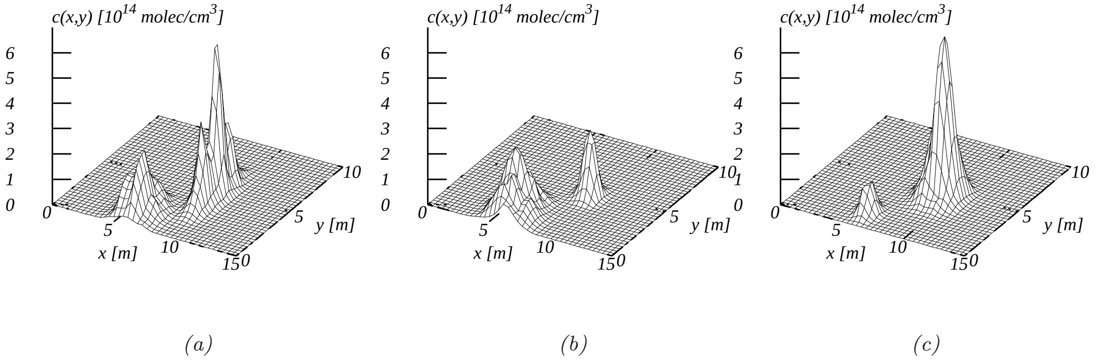
---

[1] Colour maps are a bit clearer than 2-D contour plots for the narrow peaks here.

The reconstructed map clearly shows two distinct peaks (and a hardly visible, weak accumulation in the upper half) which, on the grounds of the results of chap. 8, encourages to increase the grid dimension n. Even up to $n = 13 \times 13$ the reconstruction shows the same structure of two peaks and a very flat peak present in all maps except g, raising the question whether it genuinely belongs to the true distribution or whether it is an artefact of the reconstruction. The fact that no further artefacts appear when the grid dimension gets larger is an indicator for the a priori still working and the peaks indeed being very narrow. But while the area integrals over the individual peaks for higher dimensional grids give a fairly consistent picture, reconstruction of the peak maximum values – ranging from 2.5 to $6.7 \cdot 10^{14} molec/cm^3$ for the higher peak – is not conclusive.

To get a clearer picture of shape and maximum values, we note that in terms of the previous chapter the peaks belong to $\mathcal{E}_1$, so that according to fig. 8.5 the averaging grid translation scheme is likely to lead to a smaller overall error, while the composite scheme should give a more precise value for the peak maxima. The corresponding concentration maps are shown in tab. 9.1i-l. Again the peak integrals, especially of the lower peak, agree fairly well, but the maxima differ considerably. Trusting that the results of fig. 8.5 apply here, too, the maxima of the composite scheme ought to be more accurate. Further evidence is provided by an explicit estimation of the reconstruction error as suggested in sec. 5.3.

9.3. Discussion of the reconstruction error



(a) (b) (c)

Figure 9.3: Examples of random fields $c_I(x, y,) \in \mathcal{E}$ created by convoluting Gaussian peaks with totally random fields as described in the text.

The numerical estimation of the reconstruction error is based on the following assumptions

I. The true distribution consists of two peaks with negligible background concentration.

II. Taking the reconstructed peak maxima of peak 1 and 2, $\hat{C}_{0,1} \sim 2 \cdot 10^{14} molec/cm^3$, $\hat{C}_{0,2} \sim (3-6) \cdot 10^{14} molec/cm^3$ and the statistical scattering from fig. 8.5c, the true peak maxima are assumed to lie between

$$1 \cdot 10^{14} molec/cm^3 \leq C_{0,1} \leq 4 \cdot 10^{14} molec/cm^3,$$
$$1.5 \cdot 10^{14} molec/cm^3 \leq C_{0,2} \leq 8 \cdot 10^{14} molec/cm^3. \tag{9.1}$$

III. From the peak extensions reconstructed with the averaging and the composite grid translation scheme and from the meshsize of the geometry bounds for minimum and maximum 3σ-extension of the real peaks, respectively, are estimated as

$$100\,cm \leq 3\sigma_{x,y} \leq 400\,cm.$$

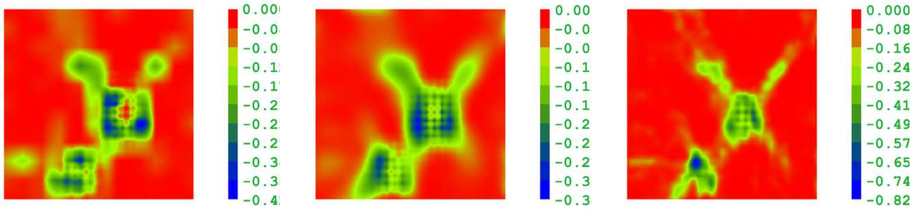
IV. The location (x_0, y_0) of the peak centre reconstructed with the averaging and composite scheme is to $\pm 50\,cm$ correct.

For the case of the real atmosphere, it is reasonable to generate the ensemble $\mathcal{E}$ for the error estimation from Gaussians. Here the random peaks are created by convoluting Gaussians with random fields as described in appendix B such that the resulting field $c_I(x, y) \in \mathcal{E}$ fulfils requirements I-IV and has at every point a maximum absolute gradient of $|\nabla c_I| \leq 10 \cdot |\nabla c_{Gaussian}|$. The latter bound is not physically motivated but intended to impose some variability on the otherwise smooth Gaussian. Samples of this procedure are shown in fig. 9.3.

Tab. 9.2 shows the mean error fields $\Delta_-\hat{c}(x, y,)$ (overestimation) and $\Delta_+\hat{c}(x, y,)$ (underestimation) as defined by eq. (5.14b) for reconstruction on a single 10×10 pixel grid and the averaging and composite scheme, respectively. The maps look very similar, with overestimation smeared around the peaks and underestimation concentrated mainly on the maximum of the higher peak. According to sec. 5.3 the contribution from the measurement error is calculated by adding random errors with variances given by the meaurement errors to the actually measured column densities and taking the standard deviation $\text{std}_\epsilon[\hat{c}]$ around the reconstructed mean..
We notice first that the very weak concentration peak in the reconstructions from experimental data found in the previous section is very likely an artefact, caused either by the imperfect inversion (tab. 9.2a-c) or by errors of the measurement data (tab. 9.2i-k). Second, calculation of the area integrals of $\Delta_\pm\hat{c}$, i.e. the mean errors, shows smaller values for the averaging scheme, followed by the single grid reconstruction. The difference between averaging and composite scheme amounts to $\sim 7\%$. Reconstruction of the peaks on a single grid is significantly more susceptible to measurement noise. Finally, mean errors for reconstruction on a 12×12 pixel grid are throughout larger than the corresponding ones for the 10×10 pixel grid, without changing the ranking of the methods.
Disregarding the results of the composite scheme on grounds of larger reconstruction errors fits neatly into the picture obtained from the reconstructed maps in tab. 9.1, where maximum values of the single grid and the averaging scheme generally are in much better agreement compared to the composite method.

To get a clearer picture of the reconstructed distribution and its error, fig. 9.4 shows concentration profiles along the x-axis through the centre of the true peaks. For the cross section considered here, the reconstruction errors $\Delta_\pm\hat{c}$ (long dashed line) are largest at the peak maxima, which for the Gaussian error ensemble from above tend to be underestimated, especially for single grid reconstruction of

$\Delta\hat{c}(x,y)\ [10^{14} molec/cm^3]$

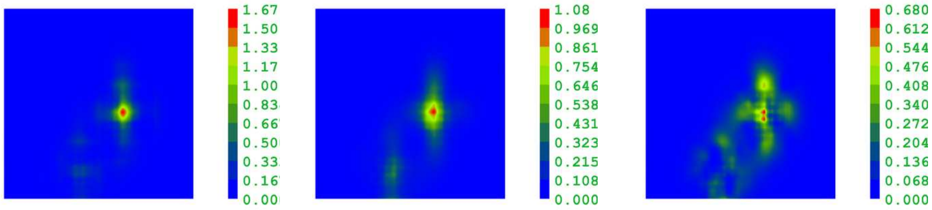


$\langle \Delta_- \hat{c}(x,y) \rangle_{\mathcal{E}}$

(a) single grid $I = -7.1\ (-7.5)$

(b) averaging scheme $I = -6.8\ (-7.3)$

(c) composite scheme $I = -7.3\ (-8.0)$

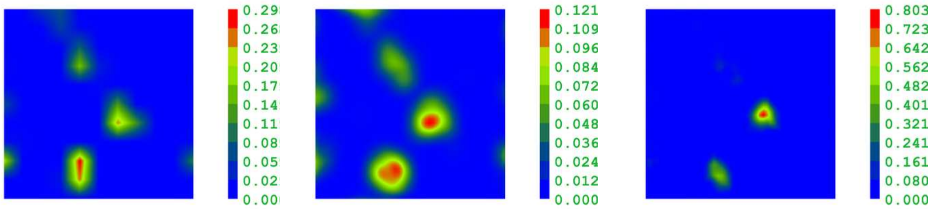


$\langle \Delta_+ \hat{c}(x,y) \rangle_{\mathcal{E}}$

(e) single grid $I = 5.6\ (5.9)$

(f) averaging scheme $I = 5.4\ (5.7)$

(g) composite scheme $I = 5.8\ (6.3)$



$\mathrm{std}_{\boldsymbol{\epsilon}}[\hat{c}(x,y)]$

(i) single grid $I = 2.6\ (3.1)$

(j) averaging scheme $I = 1.8\ (2.3)$

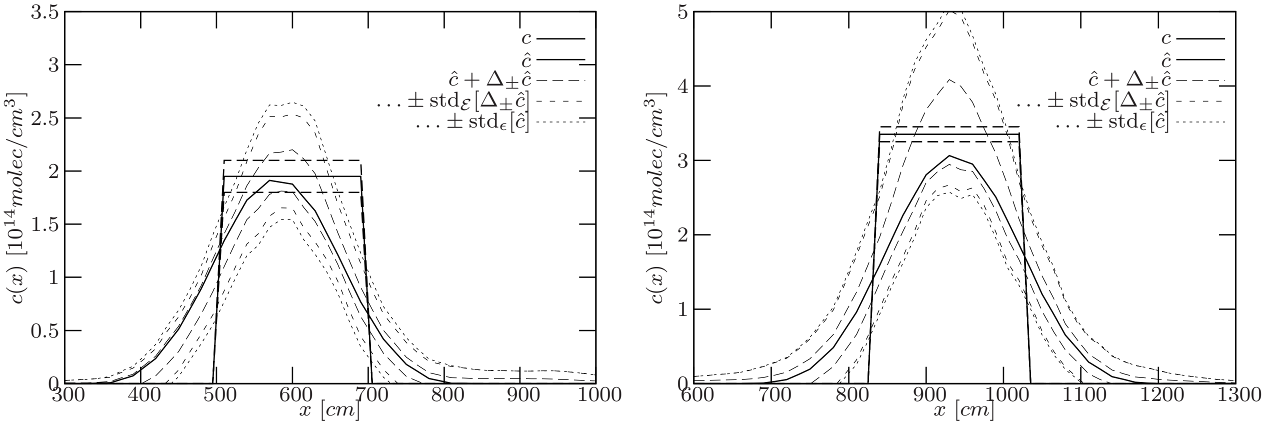
(k) composite scheme $I = 1.9\ (2.0)$

Table 9.2: Mean error fields $\Delta_-\hat{c}$, indicating overestimation, (a-c) and $\Delta_+\hat{c}$, indicating underestimation, (d-f) for reconstruction of the random distributions from the ensemble $\mathcal{E}$ described in the text and illustrated in fig. 9.3 ($N = 200$ samples) on a 10×10 pixel grid. The impact of the random measurement error (i-k) is calculated for the experimental column densities using the estimated measurement error.
I is given by the integral over the entire reconstruction area $\mathcal{A}$: $I = (10^{18} molec/cm)^{-1} \int_{\mathcal{A}} dA\ \hat{c}(\mathbf{r})$. Values in brackets $(\cdot)$ refer to a 12×12 pixel grid.

the peak 2 (fig. 9.4d). While the impact of the measurement error (dotted line) is indeed small, the standard deviations of $\Delta_\pm\hat{c}$ are of similar size as the errors themselves. Taking the standard deviations into account as they suggested in sec. 5.3 gives in the case of the averaging scheme for the peaks with

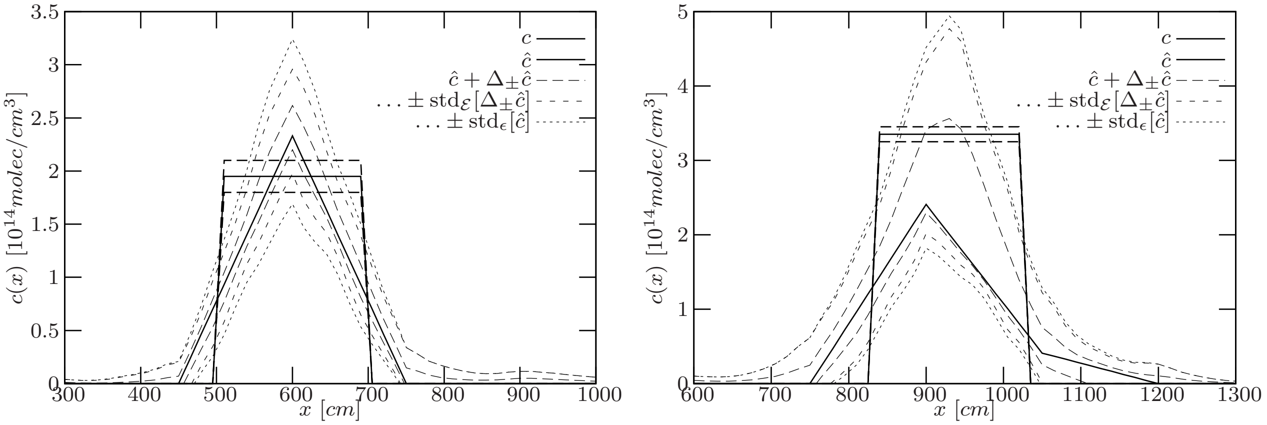
$$\hat{C}_{0,1/2} = \begin{Bmatrix} 1.9 \\ 3.1 \end{Bmatrix} \cdot 10^{14} molec/cm^3,$$

$$\int_{\mathcal{P}eak,1/2} dA\ \hat{c} = \begin{Bmatrix} 5.6 \\ 8.9 \end{Bmatrix} \cdot 10^{18} molec/cm, \tag{9.2}$$



(a) peak 1, averaging scheme

(b) peak 2, averaging scheme



(c) peak 1, single grid

(d) peak 2, single grid

Figure 9.4: Profiles along the x-axis through the centres of the peaks in tab. 9.1 reconstructed on a 10×10 grid and successive addition of the reconstruction errors $\Delta_\pm \hat{c}$, their standard deviations $\mathrm{std}_\mathcal{E}[\Delta_\pm \hat{c}]$ and the propagated measurement error $\mathrm{std}_\epsilon[\hat{c}]$. Also shown the true cylinder distribution (the uncertainty of about $\pm 5\,cm$ in the cylinder position is not taken into account).

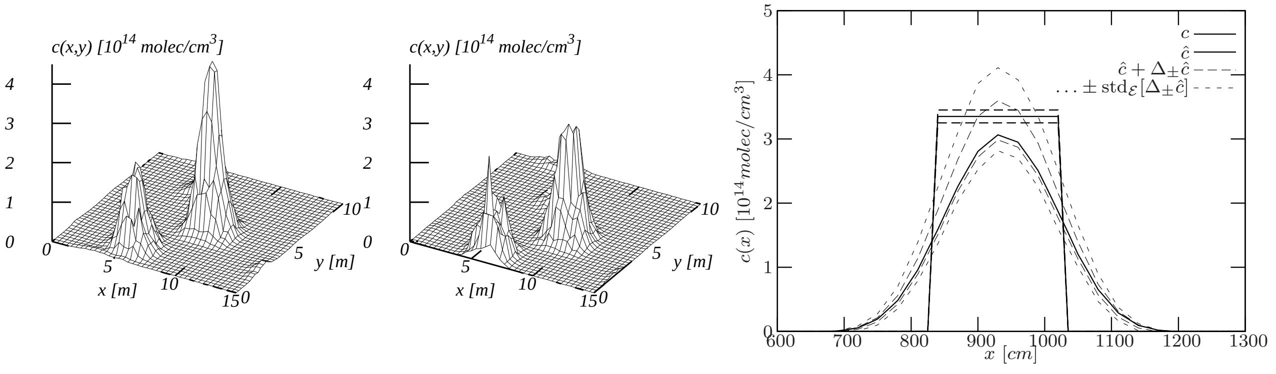
the following bounds on the peak maximum values and the peak integrals (total emissions):

$$\begin{aligned} \left\{\begin{matrix} 1.5\,(1.8) \\ 2.5\,(2.9) \end{matrix}\right\} \cdot 10^{14} molec/cm^3 \le C_{0,1/2} \le \left\{\begin{matrix} 2.5\,(2.2) \\ 5.0\,(4.0) \end{matrix}\right\} \cdot 10^{14} molec/cm^3, \\ \left\{\begin{matrix} 3.1\,(4.3) \\ 5.3\,(7.1) \end{matrix}\right\} \cdot 10^{18} molec/cm \le \int_{\mathcal{P}eak,1/2} dA\; c \le \left\{\begin{matrix} 8.8\,(6.8) \\ 16.3\,(11.9) \end{matrix}\right\} \cdot 10^{18} molec/cm, \end{aligned} \tag{9.3}$$

where the numbers $(\cdot)$ in brackets refer to $\hat{c} \pm \Delta_\pm \hat{c}$ without standard deviations $\mathrm{std}_\mathcal{E}[\Delta_\pm \hat{c}]$.

The wide ranges, particularly for the second peak with lower and upper bounds for the maxima and integrals differing by factors two and three, respectively, are quite dissatisfying. They are to a certain extent caused by the very pessimistic bounds on the random ensemble eqs. (9.1). The ranges in eqs. (9.1) can be narrowed down, for example, by assuming instead of I-IV on page 154 that

I′. In agreement with the error bounds calculated from I-IV, the distribution reconstructed on a 10×10 pixel grid by the averaging scheme (tab. 9.1i and fig. 9.4a,b) is within ±50% of its values correct.



(a) random distributions *(b) profile along the x-axis through the centre of peak 2*

Figure 9.5: Examples of random concentration fields generated from the field reconstructed on a 10×10 *pixel grid using the averaging scheme under the assumptions I′ and II′, p. 157, (a) and reconstruction errors for peak 2 (b). The measurement error is not included in the figure.*

II′. The peak centre is reconstructed with an accuracy of $\pm 50\,cm$ as before.

Generating again random distributions agreeing with I′ and II′ (see fig. 9.5a) leads to mean reconstruction errors $\Delta_\pm \hat{c}$ and standard deviations shown for peak 2 in the profile fig. 9.5b. Notably the bound on the underestimation $\Delta_+ \hat{c}$ and – as expected – its standard deviation $\mathrm{std}_{\mathcal{E}}[\Delta_+ \hat{c}]$ are considerably reduced. The new bounds on the reconstructed peaks are now (including the measurement error as in eq. (9.3))

$$\begin{Bmatrix} 1.5\ (1.8) \\ 2.6\ (2.8) \end{Bmatrix} \cdot 10^{14} molec/cm^3 \le C_{0,1/2} \le \begin{Bmatrix} 2.2\ (2.1) \\ 4.2\ (3.6) \end{Bmatrix} \cdot 10^{14} molec/cm^3,$$
$$\begin{Bmatrix} 3.7\ (4.8) \\ 5.9\ (8.1) \end{Bmatrix} \cdot 10^{18} molec/cm \le \int_{\mathcal{P}eak,1/2} dA\ c \le \begin{Bmatrix} 7.5\ (6.4) \\ 11.6\ (10.5) \end{Bmatrix} \cdot 10^{18} molec/cm \tag{9.4}$$

with values in brackets again for the case without the standard deviation of $\Delta_\pm \hat{c}$. Under the assumptions made, the estimated error of the peak maxima in eqs. (9.2) are thus maximally 20%(10%) for peak 1 and 35%(15%) for peak 2, whereas those of the integrals are 30%(15%) at most in both cases.

Finally, comparing the reconstructed fields and their error estimates to the known cylinder distributions with

$$C_{0,1/2} = \begin{Bmatrix} 1.95 \pm 0.15 \\ 3.35 \pm 0.1 \end{Bmatrix} \cdot 10^{14} molec/cm^3,$$
$$\int_{\mathcal{P}eak,1/2} dA\ c = \begin{Bmatrix} 6.1 \pm 0.5 \\ 10.5 \pm 0.3 \end{Bmatrix} \cdot 10^{18} molec/cm,$$

one finds that

- On the one hand, the true peak maxima and total emissions lie not only within the larger error bounds that take the scattering of the reconstruction errors $\Delta_\pm \hat{c}$ into account, but also within the ranges based on the mean errors $\langle \Delta_\pm \hat{c} \rangle_{\mathcal{E}}$ (and the measurement error) only.

- On the other hand, the profiles in figs. 9.4 and 9.5 clearly show that the reconstructed *distributions* do not agree with the true fields within their error bounds.

Both observations reflect fundamental problems of the error estimation for the inverse problem already discussed qualitatively in the introductory section 5.1.

The scattering of the error fields $\Delta_{\pm}\hat{c}(\mathbf{r})$, i.e. the standard deviations $\mathrm{std}_{\mathcal{E}}([\Delta_{\pm}\hat{c}(\mathbf{r})]$, is to a large extent given by the variability of the test distributions used for the statistical estimate. As pointed out in sec. 5.3, an ensemble of smooth concentration fields and one containing highly fluctuating fields can, in principle, give rise to the same mean error fields $\langle\Delta_{\pm}\hat{c}(\mathbf{r})\rangle_{\mathcal{E}}$, but their standard deviations $\mathrm{std}_{\mathcal{E}}[\Delta_{\pm}\hat{c}(\mathbf{r})]$ will be quite different. The variability of the random distributions above is mainly controlled by the choice of the maximum gradient. The fact that the lower and upper error bounds seem to be overestimated means that the arbitrary set variability of the ensemble is larger than the variability of the true concentration field.

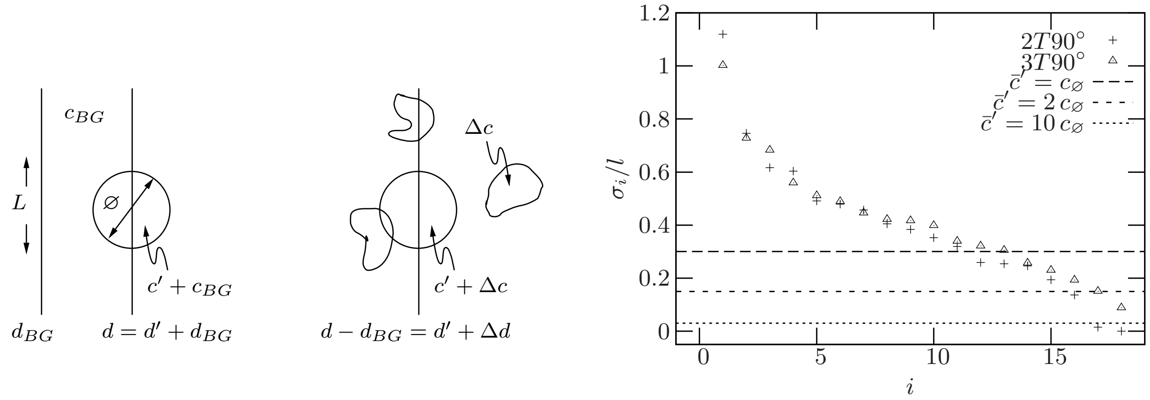
The inconsistency of the spatial distributions seen in the peak profiles is *not* due to an ensemble of test functions that is chosen too small in terms of lower, upper bounds and variability, but it has to be attributed to the generic form of the ensemble: continuous Gaussian peaks. Choosing discrete step functions as generating distributions would have increased the reconstruction error such that the real (cylinder) and the reconstructed distributions are compatible. However, in the atmosphere there would be little evidence for such an artificial choice if turbulent dispersion plays a role.

Although the two issues are related, the problem of the unknown variability is likely to play a less important role the longer the averaging time of the measurement is and thus the smoother the atmospheric distribution becomes.

9.4. Some aspects of atmospheric measurements of emission plumes

The results from the indoor experiment cannot be transferred to atmospheric measurements because – apart from the quite different experimental conditions (see [*Mettendorf et al.*, 2006; *Mettendorf*, 2005] for details) – the experiment assumed static puffs during the measurement cycle, which is rather unrealistic even for short times (see sec. 5.2.3), and the ratio of background to plume NO_2 concentrations would not correspond to atmospheric measurements. Therefore, a realistic setup, similar to measurements currently taking place over the centre of Heidelberg [*Pöhler*, 2006] is considered in the following instead. The area is assumed to be a square of size $l \times l = 2 \times 2\,km^2$ and the number of light paths generated by three multibeam instruments is taken to be 18, corresponding to the geometry in fig. 8.14b, p. 146, so that the simulation results of the previous chapter can – at least approximately – be adopted by downscaling as described in sec. 8.4.4. In particular, the smallest peak that can still be reconstructed corresponds to the lower limit of ensemble $\mathcal{E}_2$ in tab. 8.1, p. 119, with a 2σ diameter given by $\varnothing_{2\sigma}/l = 0.2$, i.e. here, $\varnothing_{min} = 400\,m$.

From a remote sensing point of view, the question whether locally enhanced concentrations can be measured purely depends on the detection limit of the trace gas species. The emission puff with mean concentration $\overline{c}'$ on top of the background c_{BG} depicted in fig. 9.6a can be detected, if the puff's



(a) Plume with concentration c' on a background c_{BG}.

(b) Plume concentration c' and residual Δc after subtracting the background.

(c) Singular values for the geometries $2T90°$, $3T90°$ with $m = 18$ light paths (figs.8.10, 8.14) scaled by the length l of the square reconstruction area. The thresholds indicate the singular values above the detection level for a plume concentration $\bar{c}'$ according to eq. (9.5).

Figure 9.6: Detection of trace gas plumes.

contribution to the total column density is above the detection limit δ for the column densities

$$d - d_{BG} > \delta \quad \text{or} \quad \bar{c}' > \frac{\delta}{2\varnothing},$$

where $\varnothing$ is the diameter of the puff. Detection limits of the multibeam instrument for a light path length of $5\,km$ are listed in tab. 9.3 for some trace gases (see also tab. 2.2, p. 22, for typical mixing ratios of these trace gases.)

From the inversion point of view, actual values of detection limits and measurement noise become important for the information content of the experiment and for the regularisation procedure. While the latter has been argued in sec. 4.4 and demonstrated in secs. 8.3.2, 9.3 to be less critical for the under-determined reconstruction of peaks than it would be for large over-determined problems, it is instructive to see how many degrees of freedom of the measurement system do in fact contribute for a given level of noise. According to sec. 7.2(i) it is given by the singular values with $\sigma_j \geq \epsilon/\|\mathbf{x}\|_2$, where ϵ is the root mean square error of the measurements. Writing – similarly to the approximations made on page 135 – $\|\mathbf{x}\|_2 \sim \sqrt{n_p}\,\bar{c}'$ for a plume with mean concentration enhancement $\bar{c}'$ extending over n_p grid nodes and taking ϵ as the detection limit δ, the inequality becomes

$$\sigma_j/l \gtrsim \frac{2}{\sqrt{n_p}}\,\frac{\delta}{d'} \quad \text{or} \quad \sigma_j/l \gtrsim \frac{2}{\sqrt{n_p}}\,\frac{c_\varnothing}{\bar{c}'}\,\varnothing/l. \tag{9.5}$$

Here, d' is the column density of the plume like in fig 9.6a and $c_\varnothing$ the concentration minimally detectable for a light path length of $2\varnothing$. For a narrow peak, e.g. $\varnothing_{2\sigma}/l = 0.3$ or $\varnothing_{2\sigma} = 600\,m$,

	O_3	NO_2	SO_2	HONO	CH_2O
detection limit [ppb] for $5\,km$	2.2	0.2	0.07	0.15	0.3

Table 9.3: Detection limits of the multibeam instrument for a light path of length $2L = 5\,km$ [Mettendorf, 2005].

tab. 8.13 suggests $n/m = 2$, corresponding to a 5 by 5 pixel grid and thus n_p something like three or four. If the mean concentration $\bar{c}'$ of the plume is around the detection limit $c_\varnothing$, only singular values $\sigma_j/l \gtrsim 0.3$ give contributions above the detection limit (fig. 9.6c, dashed line), for $\bar{c}' \sim 2c_\varnothing$ the threshold is 0.15 (dotted line) and so forth. The detection limit of NO_2, SO_2 and CH_2O for the light path length $2\varnothing = 2 \times 600\,m$ is $c_\varnothing = 0.8$, 0.3 and $1.25\,ppb$, respectively, so that for realistic pollution levels (see tab. 2.2) practically all modes contribute (in the case of geometry $3T90°$).

If the total measurement error, including noise, systematic errors and errors arising from changes of the concentration field during the measurement, is moderate, the decisive factors for the reconstruction of trace gas peaks are the discretisation and inversion errors. For negligible background concentration the relative discretisation and inversion errors do not depend on absolute peak concentrations and the detectability of a plume can be inferred from the simulations in the previous chapter.

The situation, where a plume with mean concentration enhancement $\bar{c}'$ is reconstructed by subtracting the supposed column densities of a moderately variable background concentration $c_{BG}(\mathbf{r})$ is shown in fig. 9.6b. The column densities after subtraction are up to a residual concentration field $\Delta c(\mathbf{r})$ given by those of the plume

$$\begin{aligned} d - d_{BG} &= d' + \Delta d \\ &= 2\varnothing\bar{c}' + 2L\overline{\Delta c}, \end{aligned}$$

where $\overline{\Delta c}$ is the average concentration of the residual field on the whole reconstruction area. On the one hand, this procedure strictly speaking demands

$$\Delta d \ll d'.$$

On the other hand, the presumably erratic character of the residual suggests to treat its column densities as noise, and the simulations in sec. 8.3.2 have shown that reconstruction of a (single) plume still makes sense for error levels below $\sim 10\%$. Taking into account that there are further measurement errors, one gets the rough bound

$$\frac{\Delta d}{d'} \ll 10\%, \tag{9.6a}$$

or

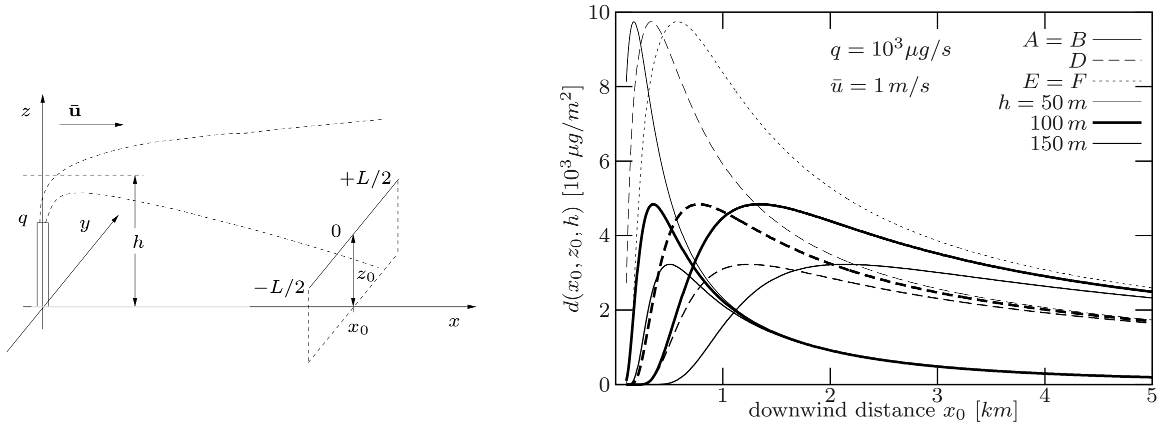
$$\frac{\overline{\Delta c}}{\bar{c}'} \ll 0.1\,\varnothing/L \sim 0.1\,\varnothing/l. \tag{9.6b}$$

For example, for a local enhancement of NO_2 around $10\,ppb$ with $\varnothing/l = 0.3$ as above and $c_{BG} \sim 1\,ppb$, the inequality reads

$$\overline{\Delta c} \ll 0.3\,ppb \sim \frac{1}{3}\,c_{BG},$$

which seems to be a realistic requirement.

Finally, the up to now hypothetical dimensions of the plume ought to be put in context to realistic dispersion of trace gas in the atmosphere. To this end, we consider a point source at distance x_0 away from the optical path, which is perpendicular to the wind direction for simplicity (see fig. 9.7a). The wind direction is assumed to be constant. The horizontal dimensions of the plume are given by the horizontal dispersion coefficients σ_y. For the Pasquill-Gifford parametrisation (appendix A) and the most frequent neutral stability class D, the 2σ-diameter at downwind distance, e.g., $5\,km$ is $\varnothing_{2\sigma} \sim 1\,080\,m$. Assuming a wind speed of $\bar{u} = 5\,m/s$ for this class, the travel time for this distance



(a) Special geometry with the optical path of length $2L$ at height z_0 parallel to the y-axis.

(b) Column density $d(x_0, z_0 = 20\,m, h)$ for an inert tracer according to eq. (9.7) for different stack heights and the urban Briggs parametrisation of the dispersion coefficients. The orientation of the optical path of length 4 km is as in figure a.

Figure 9.7: Measurement of a column density in a continuous plume emitted by a point source.

is $\sim 16\,min$. For stability classes A (extremely instable) and F (extremely stable) the corresponding values are $\varnothing_{2\sigma} \sim 1\,880\,m$ and $540\,m$, but at lower and higher wind speeds, respectively. While the plume passes the reconstruction area of length $l = 2\,km$ (in $\sim 7\,min$ at $\bar{u} = 5\,m/s$), the extension $\varnothing_{2\sigma}$ increases to $2\,530\,m$ (A), $1\,460\,m$ (D) and $730\,m$ (F). In terms of the ensembles of chap. 8, the peaks would thus belong to $\mathcal{E}_2$ to $\mathcal{E}_4$.

Turning to concentration levels in the plume, we take the most simple case of a continuous point source and, neglecting reflection at the mixing layer, the column density d measured along the light path for the geometry in fig. 9.7a can simply be calculated by integrating eq. (2.5a) along y

$$d(x_0, z_0, h) = \sqrt{\frac{2}{\pi}}\,\frac{q}{\bar{u}\,\sigma_z}\left[\exp\Big(-\frac{(z-h)^2}{2\sigma_z^2}\Big) + \exp\Big(-\frac{(z+h)^2}{2\sigma_z^2}\Big)\right]\operatorname{erf}\Big(\frac{L}{2\sqrt{2}\sigma_y}\Big), \qquad (9.7)$$

with σ_y, σ_z evaluated at x_0 and erf being the error function. Just as an example, we consider a cement works which happens to be located at about $5\,km$ distance south-east of Heidelberg and study the impact on column densities measured in Heidelberg for the scenario of fig. 9.7a.[2] The European pollutant emission register (EPER)[3] states mean annual emissions of

$$\bar{q}_{\mathrm{SO}_x} \sim 3.75 \cdot 10^5\,kg/a = 12\,g/s,$$
$$\bar{q}_{\mathrm{NO}_x} \sim 5 \cdot 10^5\,kg/a = 16\,g/s$$

for this plant. The actual direct emission of SO_2 can be written as

$$q_{\mathrm{SO}_2} = \underbrace{\frac{EF_{\mathrm{SO}_2}}{EF_{\mathrm{SO}_x}}}_{r}\, f_{\Delta t}\,\bar{q}_{\mathrm{SO}_x},$$

where EF denotes emission factors and r is the ratio of directly emitted SO_2 to SO_x. The factor $f_{\Delta t}$ is a

---

[2] This wind vector does not represent the predominant wind direction.

[3] http://www.eper.cec.eu.int/

dimensionless factor accounting for deviations from the annual mean and the corresponding expression holds for NO_2. Furthermore, the plume undergoes chemical transformation which in the case of NO_2 can be characterised by the downwind NO_x concentration and the Leighton ratio $L = [NO]/[NO_2]$ (c.f. sec. 2.2) and for SO_2 takes the simple form of an exponential decay by $\exp(-kx_0/\bar{u})$ with a value of $k = 4,81 \cdot 10^{-6} s^{-1}$ suggested in [*ISC-3*, 1995] for urban environment. For $x_0 = 5\,km$ and a wind speed of $2\,m/s$, the exponential takes a value of 0.99, so that the SO_2 plume can be treated as inert. Including the downwind chemical decay by the factor $D(x_0)$, we write

$$d_D(x_0, z_0, h) = D(x_0)\, d(x_0, z_0, h) = D(x_0)\, r\, f_{\Delta t}\, d_0(x_0, z_0, h). \tag{9.8}$$

Except for the most instable conditions (A), the column density at $x_0 = 5\,km$ does not depend significantly on the effective stack height h, as shown in fig. 9.7b, where $d(x_0, z_0, h)$ is calculated for different stack heights for a length $2L = 4\,km$ of the light path at height $z_0 = 20\,m$, using the Briggs urban parametrisation of the dispersion coefficients.[4]

Values of the column densities $d_0(x_0 = 5\,km, z_0 = 20\,m, h = 100\,m)$ without the correction factors D, r and f in eq. (9.8) and corresponding average concentrations $\bar{c}'_0$ along the light path amount to

SO_x			NO_x				
d_0 $[10^3 \mu g/m^2]$	$\bar{c}'_0$ $[\mu g/m^2]$	$[ppb]$	d_0 $[10^3 \mu g/m^2]$	$\bar{c}'_0$ $[\mu g/m^2]$	$[ppb]$	stab. class	$\bar{u}$ $[m/s]$
6	1.5	0.6	8	2	1	A/B	2
7.5	2	0.7	10	2.5	1.3	D	5
30	7	2.6	40	10	5	E/F	2

Taking the factors D and r into account will reduce these concentrations further, while $f_{\Delta t}$ might well be much larger than one. Average concentration values along a light path above the centre of Heidelberg obtained by *Rippel* [2005] during measurements in December 2004/January 2005 range from around $0.5\,ppb$ to $14\,ppb$ for SO_2 and from $\sim 2\,ppb$ to maximally $45\,ppb$ for NO_2. Whether the plume could be detected as such depends – apart from the meteorological conditions, the activity of the cement works and the chemical development of the plume – on the actual emission factors and on the concentration levels in Heidelberg itself. For a plume with $\varnothing/l \sim 1\,000\,m/2\,km = 0.2$ (stability class D, c.f. p. 161) eqs. (9.6) require $\overline{\Delta c}/\bar{c}' \ll 0.05$. Taking the corrected mean concentration $\bar{c}'$ of SO_2, for example, as a tenth of $\bar{c}'_0$ above, i.e. $\bar{c}' \sim 0.2\,ppb$, this means $\overline{\Delta c} \ll 10^{-3}$ – something unachievable.

The same analysis can be carried out for a large coal fired power station, the GKM, Mannheim, located $\sim 15\,km$ north-west of Heidelberg. The EPER specifies

$$\bar{q}_{SO_x} \sim 3 \cdot 10^9\, kg/a = 70\, g/s,$$
$$\bar{q}_{NO_x} \sim 3.8 \cdot 10^9\, kg/a = 120\, g/s.$$

The ratio $r = EF_{NO_2}/EF_{NO_x}$ takes values below 10% (see sec. 2.2), while according to the UK emission database[5] $EF_{SO2}/EF_{NO2} \sim 2.7$ for coal. Using the Pasquill-Gifford parametrisation (for rural environment) and ignoring again the mixing layer, the uncorrected column densities d_0 and average concentrations $\bar{c}'_0$ are

---

[4] We use the units $[\mu g/m^2]$ common in air pollution monitoring.

[5] http://www.naei.org.uk/

SO_x			NO_x				
d_0 $[10^3 \mu g/m^2]$	$\bar{c}'_0$ $[\mu g/m^2]$	$[ppb]$	d_0 $[10^3 \mu g/m^2]$	$\bar{c}'_0$ $[\mu g/m^2]$	$[ppb]$	stab. class	$\bar{u}$ $[m/s]$
35	9	3	60	15	8	B	2
56-84	14-21	5-8	96-144	24-36	13-20	D	5
70-315	18-80	7-30	120-540	30-135	16-71	F	2

where first values refer to an effective stack height of $200\,m$, the second to $100\,m$.
With the more realistic higher effective stack height and having the values of the factors r in mind, these concentrations can for highly above-average activity of the power station be a significant contribution to Heidelberg's pollution level, but in view of the plume extension $\varnothing_{2\sigma} \sim 800\,m\,(F) - 6\,000\,m\,(A)$, it would rather appear as an enhanced background than a distinct plume.

The two examples discussed represent the most important point sources for NO_2 and SO_2 in the vicinity of Heidelberg. Therefore, it can be concluded that (narrow) plumes with concentration maxima distinctly above the city level have to originate from within the area of Heidelberg.

10. 2-D Reconstruction of Model Trace Gas Distributions above a Street Canyon

This last chapter of numerical results focuses on tomographic reconstruction of trace gas distributions in an urban environment. For lack of experimental data results from an elaborate model system, ultimately designed to calculate concentrations within and in the vicinity of a highly polluted city street canyon, will be employed to simulate measurements with a moderate number of light paths. The reason for choosing these model distributions is simply that they cover a horizontal area large enough and with sufficient spatial resolution to represent a realistic state of the atmosphere on the scale of our tomographic measurement.

We are neither concerned with the particularities of the model, nor with the specific meteorological situation or the emission scenario, but only with the spatial patterns of the distributions. The description of the model situation in sec. 10.1 is therefore kept rather brief, before turning to the exemplary reconstruction of NO_2 in sec. 10.2. The reconstruction procedure turns out to be far less flexible than the one for peak distributions, instead regularisation becomes crucial. This becomes even more evident in the error calculations of the subsequent sec. 10.3, which examines the possibility to verify the model NO_2 distribution by means of the tomographic setup. The evaluation procedure follows the suggestion of sec. 7.4 on model evaluation.

10.1. Model system, set-up and results

Model results used for the simulations of this chapter were obtained by the model system M-SYS, which was briefly referred to in sec 2.5 as a tool for the evaluation of ambient air quality according to the EU framework directive 96/62/EC (see sec. 2.1). The trace gas distributions presented shortly were calculated by *D. Grawe*[1] for the street canyon Göttinger Straße in Hanover in the framework of the VALIUM project, which was carried out especially for the development and validation of these tools [*Schatzmann et al.*, 2006]. Motivation for the development of M-SYS was the assumption that contributions from all relevant scales have to be modelled properly to predict air pollutant concentrations on the street scale. Although only the innermost model domain will be considered here, some characteristics of the models should be mentioned for the sake of completeness. More details and references can be found in [*Trukenmüller et al.*, 2004].

Two models are used to calculate transport and chemistry of pollutants on the mesoscale: the non-hydrostatic Mesoscale Transport and Stream Model METRAS [*Schlünzen et al.*, 1996] and the

---

[1] Grawe, D., personal communication.

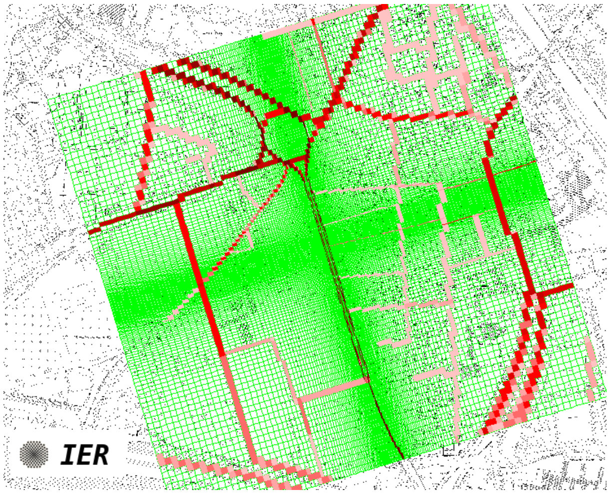


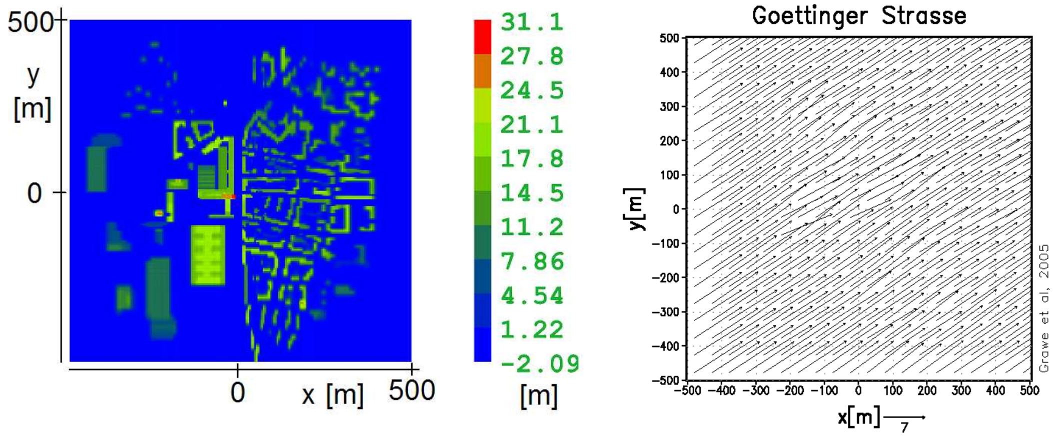
Figure 10.1: Non-uniform model grid for the microscale street canyon calculations. The area is $\sim 1050 \times 1060\,m^2$, resolved by $15\,m$ near the boundaries and $1.5\,m$ at the centre, where Göttinger Straße heads towards the roundabout in the North. Also shown the traffic emissions for the case study used here. They are highest for Göttinger Straße and the roundabout (The picture is misleading because of the uneven grid) [IER, University of Stuttgart. Courtesy of D. Grawe, now with the Division of Environmental Health and Risk Management, University of Birmingham].

Mesoscale Chemistry Transport Model MECTM [*Lenz et al.*, 2000]. For the case at hand they are applied on three levels with horizontal domains of $2000 \times 2000\,km^2$ (resolution $16\,km$), $352 \times 356\,km^2$ ($4\,km$) and $112 \times 200\,km^2$ ($1\,km$), respectively, to assess air quality on a regional scale and provide boundary conditions for one way nesting of the models. All domains are centred on Hanover-Brunswick, where the street canyon of Göttinger Straße is located. Boundary conditions for the outermost model are given by interpolated observations in the case of meteorological parameters and are partly based on climatologies in the case of background concentrations (see [*Trukenmüller et al.*, 2004] for details). The innermost mesoscale model includes major point and area sources (such as motorways and cities etc.) that contribute to the background concentrations of the microscale environment of the street canyon. The latter is defined horizontally by a $\sim 1 \times 1\,km^2$ area with the street canyon at its centre and vertically up to a height of about $400\,m$. Flow fields are calculated on a non-uniform grid using the obstacle-resolving microscale model MITRAS [*Schlünzen et al.*, 2003]. The grid spacing varies horizontally from $15\,m$ at the lateral boundaries to $1.5\,m$ at the centre of the street canyon, see fig. 10.1. The vertical grid consists of 50 layers with spacing $1.5\,m$ at the bottom and $30\,m$ at the top. Concentrations within the street canyon environment are computed by the microscale modification of MECTM, called MICTM[2]. Both models are based on the gas phase chemical mechanism RADM2 [*Stockwell et al.*, 1990] and calculate concentrations of 59 species. Benzene (an important pollutant subject to EU regulation, see also tab. 2.1, p. 21) is not modelled individually because emission inventories are lacking.

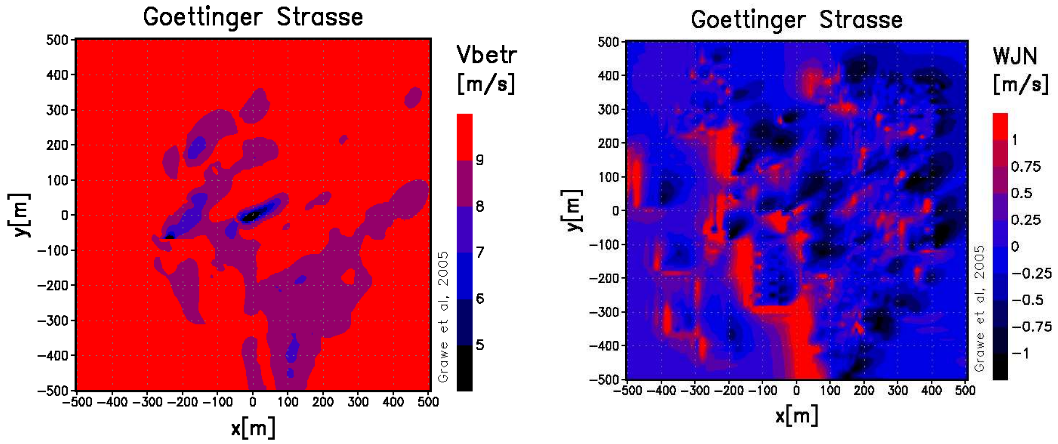
The obstacle-resolving model MITRAS was evaluated using wind tunnel data. The comparison shows adequate agreement for the flow fields in general, but near the edges of buildings and at roof top results depend on the turbulence closure scheme chosen [*Schlünzen et al.*, 2003]. The system M-SYS was compared to experimental observations of background concentrations of SO_2, NO_2 and O_3 on a regional scale with very good results for SO_2 and NO_2, but less accurate predictions for O_3 [*Trukenmüller et al.*, 2004]. The microscale performance of the model system was evaluated by intensive field measurements within the street canyon during the VALIUM project and included point and remote sensing measurements of important traces gases and the tracer SF_6 [*Schäfer et al.*, 2005]. Comparing $30\,min$ averages of point measurements taken at $2\,m$ above ground and the predictions by

---

[2] Grawe, D.: Verknüpfung von Modellen und Messungen zur Konzentrationsvorhersage, PhD-thesis, Fachbereich Geowissenschaften, University of Hamburg (in German), in preparation, 2006.



(a) Building height (street canyon at $x \sim 0$) *(b) Horizontal wind vector $30\,m$ above ground*



(c) Horizontal wind velocity $30\,m$ above ground *(d) Vertical wind velocity $30\,m$ above ground*

Figure 10.2: Buildings and wind field at $30\,m$ above ground for the microscale models. The street canyon is at $x \sim 0$ [Courtesy of D. Grawe, now with the Division of Environmental Health and Risk Management, University of Birmingham].

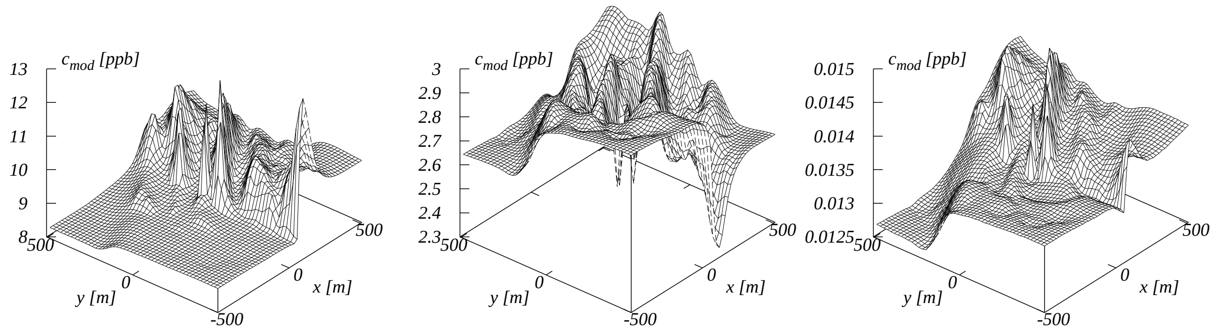
M-SYS shows mutual agreement well within a factor of 2.[3]

The trace gas distributions considered in the following were obtained from M-SYS under the assumption that thermal influences can be neglected on the microscale, which leads to a stationary solution for the street canyon. For the $30\,min$ run of the model system with $1\,min$ time steps it was furthermore assumed that background emissions in the mesoscale models do not change and that the only emissions relevant for the microscale models originate from traffic.

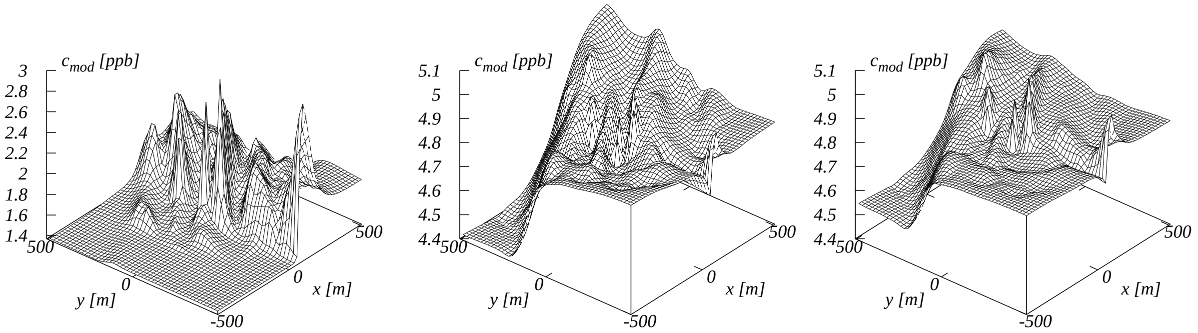
As the highest buildings within the microscale model domain reach $30\,m$, see fig. 10.2a, the optical paths of the tomographic remote sensing measurement are taken to lie in a plane at around $30\,m$ above ground. The associated wind field is shown in figs. 10.2b-d. Measuring column densities at $30\,m$ is admittedly somewhat high (for example, concentrations of NO_2 are about ten times lower than at ground). But there are several reasons why a tomographic measurement at this height yet is interesting, especially from a modeller's point of view.

---

[3] Grawe, D., personal communication.



(a) NO_2 at $t = 0$ (b) O_3 at $t = 0$ (c) HONO at $t = 0$



(d) CH_2O at $t = 0$ (e) SO_2 at $t = 0$ and $t = 30\,min$

Figure 10.3: M-SYS model distributions of selected trace gases on the section at $30\,m$ above ground. The street canyon is at $x \sim 0$. Time steps of the model run are $1\,min$.

- First, it gives information about the background contribution to the street canyon pollution level.

- Second, as mentioned before, at least for the microscale transport model MITRAS employed here, vertical momentum fluxes way above roof top can depend significantly on the turbulence closure [*Schlünzen et al.*, 2003, especially fig. 2].

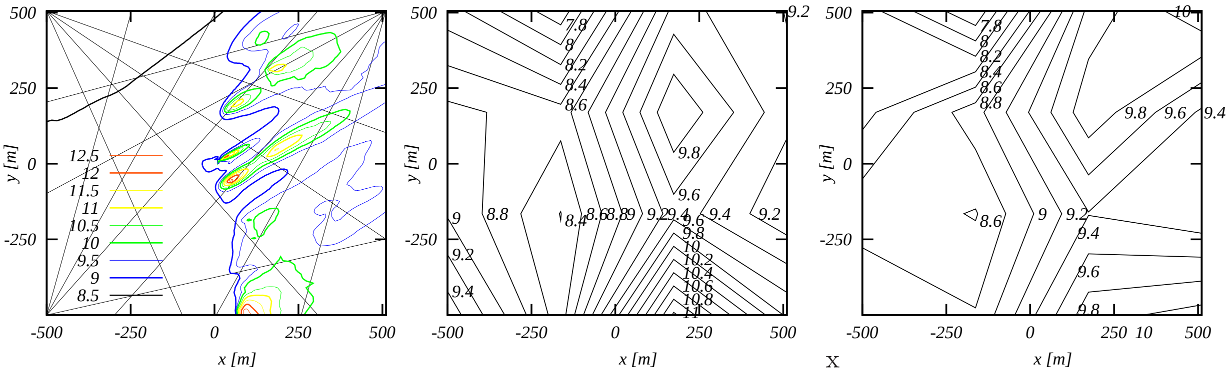
- And third, related to this topic, it has been observed in the field measurements mentioned above that tracers were uplifted from the street canyon, over the roofs and down into the backyards. This effect occurs in model simulations, too [*Schlünzen et al.*, 2003] and it would be interesting to test the quantitative agreement between measurement and model.

Finally, fig. 10.3 shows examples of 2-D model distributions for the height $\sim 30\,m$.[4] All species show concentration peaks (or sinks in the case of O_3) rising at the street canyon and extending plume like along the wind direction to the right hand side of the street canyon. They are caused by the building structure[5] and not very high at $30\,m$, for example maximally $3\,ppb$ for NO_2. For some trace gases, like NO_2, the background concentration field on the left hand side of the street canyon is very smooth

---

[4] Again 3-D contour plots are chosen to make the spatial variability clearer.

[5] Grawe, D., personal communication.



(a) $c_{mod}(x, y)$ [ppb] (NO_2 at $t = 0$) *(b) $c_{id}(x, y)$ [ppb]* *(c) $\hat{c}(x, y)$ [ppb]*

Figure 10.4: 2-D NO_2 distributions 30 m above ground. (a) M-SYS model as in fig. 10.3a. (b) Ideal parametrisation on a regular bilinear grid of dimension $n = 4 \times 4$. (c) Distribution reconstructed on a grid with $n = 4 \times 4$ and mean a priori. The iteration number is $k = 500$ and values of the quality indices are: $NEARN = 0.57$, $NMSE = 1.6 \cdot 10^{-3}$, $IOA = 0.89$, $NMB = 2 \cdot 10^{-3}$, $FB = -4 \cdot 10^{-4}$, $r^2 = 0.67$ and $FA2 = 100\%$.

compared to the peak structure, others like O_3 show more background variability. The concentration fields shown in fig. 10.3 change only slightly during the 30 min interval. The most significant change occurs for SO_2 due to a change in the background.

10.2. Sample reconstruction for NO_2

We consider the reconstruction of NO_2 from measurements with three telescopes and a moderate number of 18 optical paths, arranged in 90°-fans as in fig. 8.14b on page 146.

The meshsize Δl in the centre is about $\sim 200 - 250\, m$, the smallest peak structures have a size of about $\Delta s \sim 30\, m$. It is thus pure chance, if such a peak is hit by a light path like the one in the centre of fig. 10.4a. But even then a peak with $\varnothing \sim 30\, m$ and concentration enhancement $\bar{c}' \sim 3\, ppb$ would give rise to a column density $\sim 180\, ppb$, which is far below the detection limit of the column density $\delta = 1\,000\, ppb \cdot m$ according to tab. 9.3 (The minimal peak extension for an enhanced NO_2 concentration of $3\, ppb$ would be $\sim 170\, m$).

The NO_2 distribution with a smooth background to the left of the street canyon offers a good opportunity to illustrate different estimates of the background concentration c_{BG} according to sec. 6.3.2(i). Following eq. (6.11), the over-determined case of a single, regular 3 by 3 pixel grid with $n = 4 \times 4$ leads to

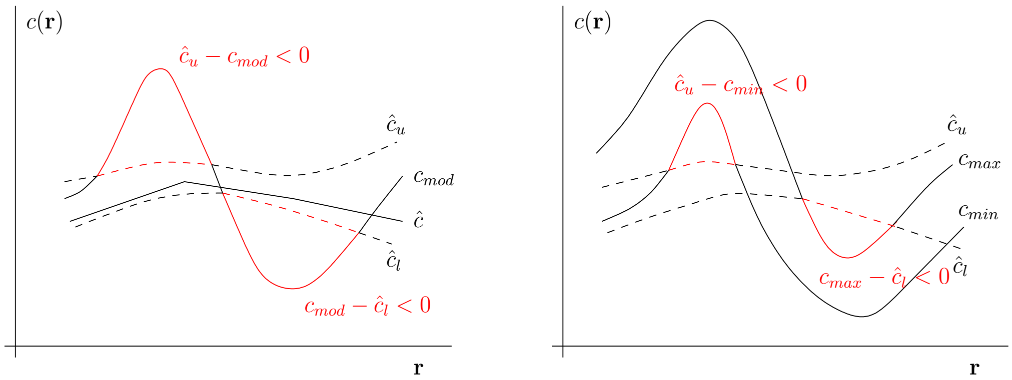
$$c_{BG} = 8.64\, ppb \qquad \text{for} \quad \text{eq. (6.11)},$$

coming very close to the values $8.3 - 8.5\, ppb$ in figs. 10.3a, 10.4a.

Values for the under-determined cases eqs. (6.12) and a regular grid with $n = 5 \times 5$ are

$$c_{BG} = 8.77,\ 9.12,\ 4.56\, ppb \qquad \text{for} \quad \text{eqs. (6.12a), (6.12b), (6.12c)}.$$

The last case does not reproduce the physical background, the first two lead to almost identical reconstructions $\hat{c}$.



(a) Evaluation of the model *(b) Evaluation of the model bounds*

Figure 10.5: Consistency of the experimental retrieval $\hat{c}$ with the model c_{mod} and its bounds $c_{min/max}$, understood as 1σ check. Upper and lower bounds of the reconstructed distribution are given by $\hat{c}_{l/u} = \hat{c} + \Delta_{l/u}\hat{c}$. (a) Case (i). (b) Case (ii).

Simulating error free measurements of the model NO_2 distribution with the above optical geometry of 18 light paths and reconstruction using SIRT with different grid dimension, a priori and reconstruction schemes reveals that

1. The over-determined parametrisation gives smaller overall errors and better statistical indices than the under-determined one.
2. Optimal iteration numbers, especially in the *under*-determined case, lie long before convergence.
3. The optimal choice for the a priori is the mean eq. (6.7).
4. Neither any of the the grid-shifting schemes, nor subtracting or fitting the background leads to better results than reconstruction with a single $n = 4 \times 4$ grid and mean a priori.
5. Except when subtracting the background, the nonnegativity constraint is not active.

Measurement errors have been neglected everywhere for the moment.

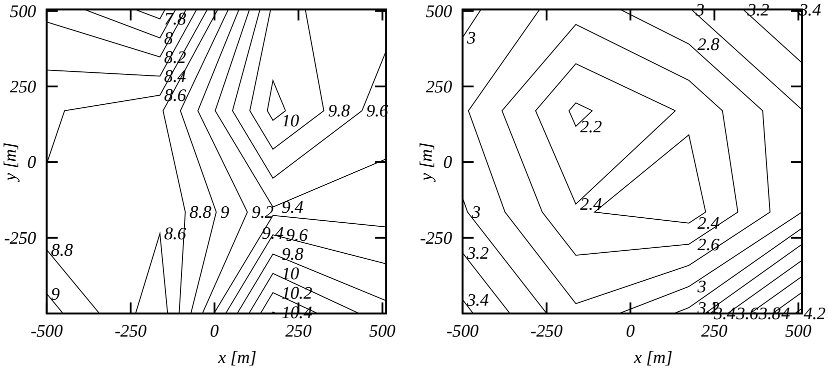
Point 4. is not too surprising as, even after ideal removal of the backgrund, the distribution does not represent a genuine peak distribution like in chap. 8. Points 3. and 5. are plausible. However, the second observation seems to be a sign of a crucial difference to the reconstruction of peaks. The fact that the *under*-determined solutions have to be strongly regularised suggests that the least-squares least-norm solution and thus the least-norm principle might not longer be the optimal selection criterion for this kind of distribution.

The reconstruction result on an over-determined $n = 4 \times 4$ grid shown in fig. 10.4c does not bear much resemblance to the original distribution above the street canyon, but comparison with the ideal parametrisation of the model distribution by the same grid (according to sec. 5.2.1) in fig. 10.4b makes clear that a field of this kind has to be expected from the given spatial resolution .

10.3. Discussion of model evaluation

We examine the two questions raised in the context of model evaluation in sec. 7.4:

Figure 10.6: Optimal estimate using $\mathbf{x}_{id}$ from fig. 10.4b as a priori with relative variances of 50% and 10% noise on the column densities obtained from the model distribution in fig. 10.4a. (a) The state vector. (b) A posteriori variances. Values in between the grid nodes are merely interpolated.



(a) $\hat{\mathbf{x}}$ [ppb] (b) $\sigma_{\hat{\mathbf{x}}i}$ [ppb]

(i) Given a distribution $\hat{c}$ reconstructed from a tomographic experiment and a model prediction c_{mod} for the same situation, do they agree within the recontruction error ?

(ii) If not, do they agree within the uncertainty of the model ?

The first question is illustrated by fig. 10.5a with the reconstruction error given by its lower and upper bounds $\Delta\hat{c}_l$ and $\Delta\hat{c}_u$, the second by fig. 10.5b with the model uncertainty given by the bounds c_{min} and c_{max}.

We take the NO_2 distribution from figs. 10.3a,10.4a as model prediction c_{mod} and assume its error according to the specification for the street canyon calculations given above on page 167 to be a factor of two. That is, *we take it for granted* in the following that the values of the *true* concentration field c lie between

$$c_{min}(\mathbf{r}) = 0.5 \cdot c_{mod}(\mathbf{r}) \leq c(\mathbf{r}) \leq 2 \cdot c_{mod}(\mathbf{r}) = c_{max}(\mathbf{r}), \tag{10.1}$$

(Especially the upper bound might seem pessimistic, but we recall that building effects can change concentration values to a high degree (see sec. 2.5) and that the model treatment of flows around buildings is far from trivial. See the discussion above.)

Because we do not have experimental column densities, measurements will be *simulated* for 'true' concentration fields c that are obtained by randomly varying the model within its uncertainty. Case studies for such simulated measurements will be presented in sec. 10.3.3. Before, the reconstruction error has to be quantified. The next section 10.3.1 will show that an estimation based on eq. (10.1) alone leads to reconstruction errors that are too large to allow any real evaluation of the model. Therefore, further assumptions will be added to the above inequality bounds in sec. 10.3.2.

10.3.1. Estimation of the reconstruction error using the optimal estimate

To get a first idea about the predictive power of the tomographic measurement under the given circumstances, we employ the optimal estimate, which makes sense because the nonnegativity constraint is not active (5. above). The ideal discretisation in fig. 10.4b is used as a priori $\mathbf{x}_a$ and the data error covariance is assumed to be of the form $S_\epsilon = \sigma_\epsilon^2 \mathbb{1}_n$ with σ_ϵ given by a relative measurement error of 10%. The asymmetric bounds in eq. (10.1) cannot be represented by a Gaussian a priori probability density. Instead, we use the covariance $S_a = \mathrm{diag}(\sigma_{a_i}^2)$ with $\sigma_{a_i} = 0.5 \cdot x_{a_i}$, that is a $\pm 50\%$ 1σ error. The state vector $\hat{\mathbf{x}}$ and the diagonal elements of its covariance matrix $S_{\hat{\mathbf{x}}}$ are shown in fig. 10.6.

Clearly, the 1σ uncertainty of the a posteriori lies within the bounds of eq. (10.1). But – also evident – its minimum value of about $2\,ppb$ is almost as high as the peak structures, and in fact all structues, of the original model distribution. And the optimal estimate does not even include the discretisation error arising from the finite representation of the continuous field. To be more precise, the optimal estimate provides a measure for the reconstruction error using an ensemble of distributions that are Gaussian distributed around the a priori, with a spatial variability given by the grid spacing. However, the variability of the model distribution is much higher.

10.3.2. Estimation of the reconstruction error using a random ensemble

We now narrow down the true distribution further by assuming additionally to eq. (10.1) that not only its concentration values c but also its gradients ∇c are confined by the model.

Generating random concentration fields as described in appendix B, it turns out that it is possible to constrain ∇c such that the background behaviour to the upwind left hand side of the street canyon, the increase of concentration downwind to its right and the appearance of the original peaks can be reproduced with a variability of the concentration values that sufficiently exploits eq. (10.1). Peak centres appear at the original sites. To account for model uncertainties of exact positions an arbitrary shift of the whole distribution in the plane by maximally 10% ($\pm 100\,m$) is allowed. To be precise, we now *assume* that the true distribution c for the atmospheric state modelled by c_{mod} lies within the bounds:

I. $0.5 \cdot c_{mod}(x,y) \leq c(x,y) \leq 2 \cdot c_{mod}(x,y)$.

II. $\nabla c(x,y)$ is maximally $f'_{max} \cdot \nabla c_{mod}(x,y)$ in the sense of eqs. (B.3), (B.4) with $f'_{max} = 5$.

III. $c(x,y) \mapsto c(x+\Delta_x, y+\Delta_y) \quad \text{with} \quad 0 \leq |\Delta_x|, |\Delta_y| \leq 100\,m$.

The exact value $f'_{max} = 5$ for the factor constraining the gradient is arbitrary. For example $f'_{max} = 10$ results in similar, yet very noisy concentration fields. Furthermore, as discussed at the end of sec. 9.3, the choice of maximum gradients affects the bounds of the reconstruction error. Here, f'_{max} is chosen such that the variability of the model distribution is reproduced and that the bounds given by I are exploited.

The ensemble $\mathcal{E}$ of distributions c_I, $I = 1, \ldots N$, for estimation of the reconstruction error is given by random fields satisfying I-III. Samples are shown in fig. 10.7a to 10.7f, while figure 10.7c represents a distribution that was obtained by a shift of more than $100\,m$ in the direction of negative y. (The 2-D contour plots in bottom panel will be referred to later.) The standard deviation around the ensemble mean $\langle c_I(x,y,)\rangle_{\mathcal{E}}$ is shown in fig. 10.8a. Indeed, it exhibits higher variabilty in the peak region to the right of the street canyon and the ensemble created from I-III appears to be a sensible random variation of c_{mod}, reproducing a smooth background to the left of the street canyon and decreasing peak concentrations towards the boundary right from it.[6]

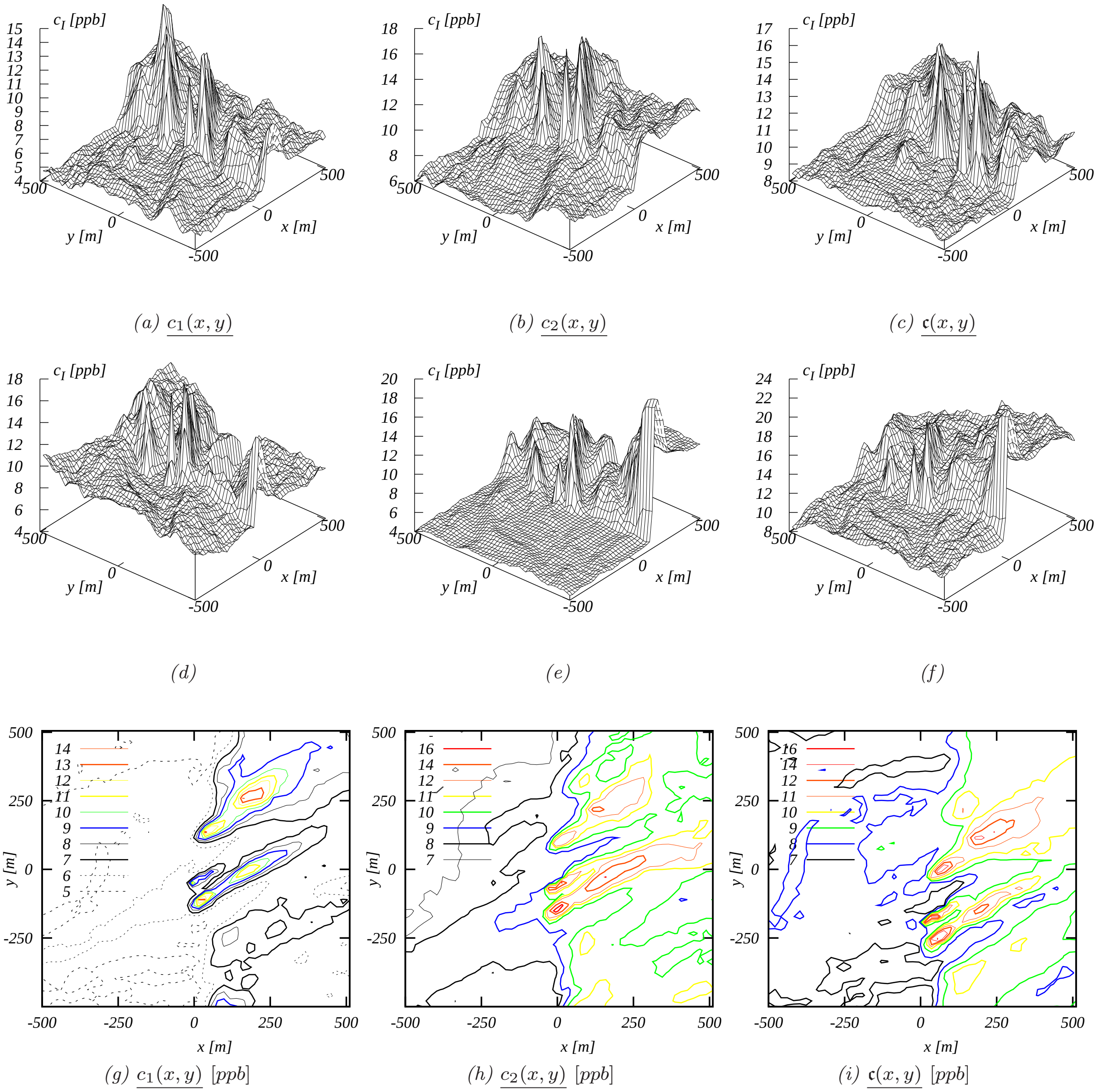
---

[6] Mind that $\langle c_I(x,y,)\rangle_{\mathcal{E}} \neq c_{mod}(x,y)$. In general, for a random number c uniformly distributed between c_{max} and c_{min}

$$\langle c\rangle = \frac{1}{2}(c_{max} + c_{min})$$

$$\mathrm{std}[c] = \frac{1}{2\sqrt{3}}(c_{max} - c_{min}).$$

In particular for $c_{max} = 2c_{mod}$, $c_{min} = \frac{1}{2}c_{mod}$: $\langle c\rangle = \frac{5}{4}c_{mod}$ and $\mathrm{std}[c] = \frac{\sqrt{3}}{4}c_{mod}$.



(a) $\underline{c_1(x,y)}$ (b) $\underline{c_2(x,y)}$ (c) $\underline{\mathfrak{c}(x,y)}$

(d) (e) (f)

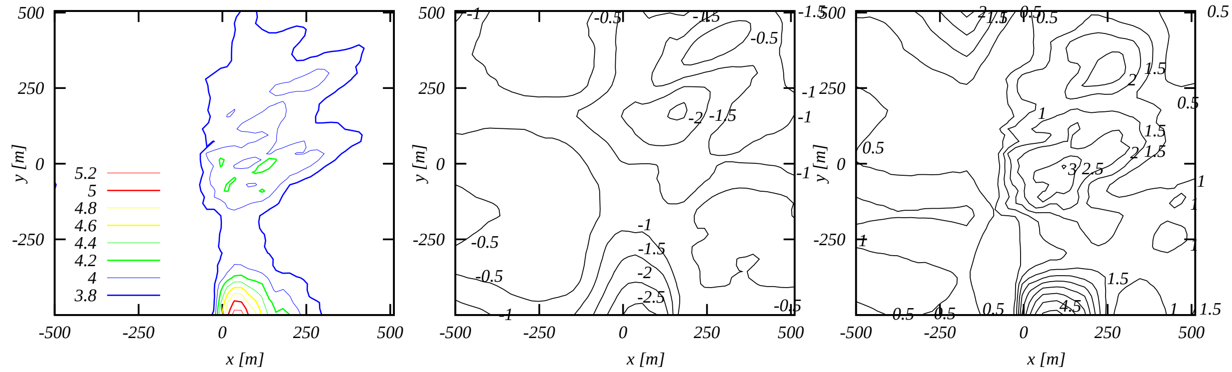
(g) $\underline{c_1(x,y)}$ [ppb] (h) $\underline{c_2(x,y)}$ [ppb] (i) $\underline{\mathfrak{c}(x,y)}$ [ppb]

Figure 10.7: Random concentration fields generated from the model NO$_2$ distribution (figs. 10.3a,10.4a) according to points I-III, except $\mathfrak{c}(x,y)$ in fig. 10.7c, which violates III in that $\Delta_y > 100\,m$. The bottom panel shows 2-D contour plots of c_1, c_2 and $\mathfrak{c}$ for later reference.

Figures 10.8b and 10.8c show lower and upper bounds of the reconstruction error field $\Delta\hat{c}(x,y)$ calculated for $\mathcal{E}$ according to eq. (5.14d)

$$\Delta_{u/l}\hat{c}(x,y) = \langle\Delta_{\pm}\hat{c}(x,y)\rangle_{\mathcal{E}} \pm \mathrm{std}_{\mathcal{E}}[\Delta_{\pm}\hat{c}(x,y)]. \tag{10.2}$$

Measurement noise was not taken into account. It follows that



(a) $std_{\mathcal{E}}[c_I(x,y)]$ $[ppb]$ (b) $\Delta_l\hat{c}(x,y)$ $[ppb]$ *without noise* (c) $\Delta_u\hat{c}(x,y)$ $[ppb]$ *without noise*

Figure 10.8: Random ensemble $\mathcal{E}$ *generated from I-III (a) Standard deviation around the mean of* $\mathcal{E}$*. (b) Lower and (c) upper bounds of the reconstruction error without noise, i.e.* $\Delta\hat{c} = \Delta c_{disc} + \Delta\hat{c}_{inv} + \Delta\hat{c}_{reg}$ *and* $\Delta_{l/u}$ *understood as* $\Delta_{u/l} = \langle\Delta_\pm\rangle_{\mathcal{E}} \pm std_{\mathcal{E}}[\Delta_\pm]$.

- If the true background shows a variability similar to the model (it is in fact higher for the distributions from $\mathcal{E}$, see fig. 10.7), it can be reconstructed with a precision of about $\pm 0.5\,ppb$.
- Reconstruction of the peak structure is hardly possible, especially not for the peak at $y \sim -500\,m$. If the location of the peaks within the reconstruction area is completely unknown, the resulting reconstruction error of at least $\pm 2\,ppb$ will exceed the structures in c_{mod}. Simulations where III is released to a shift over the entire area yield bounds $\Delta_{l/u}\hat{c}$ between $\sim \pm 1.5$ and $\pm 4\,ppb$. Even without measurement error, this is already of the order of what was obtained from the optimal estimate, see fig. 10.6b.

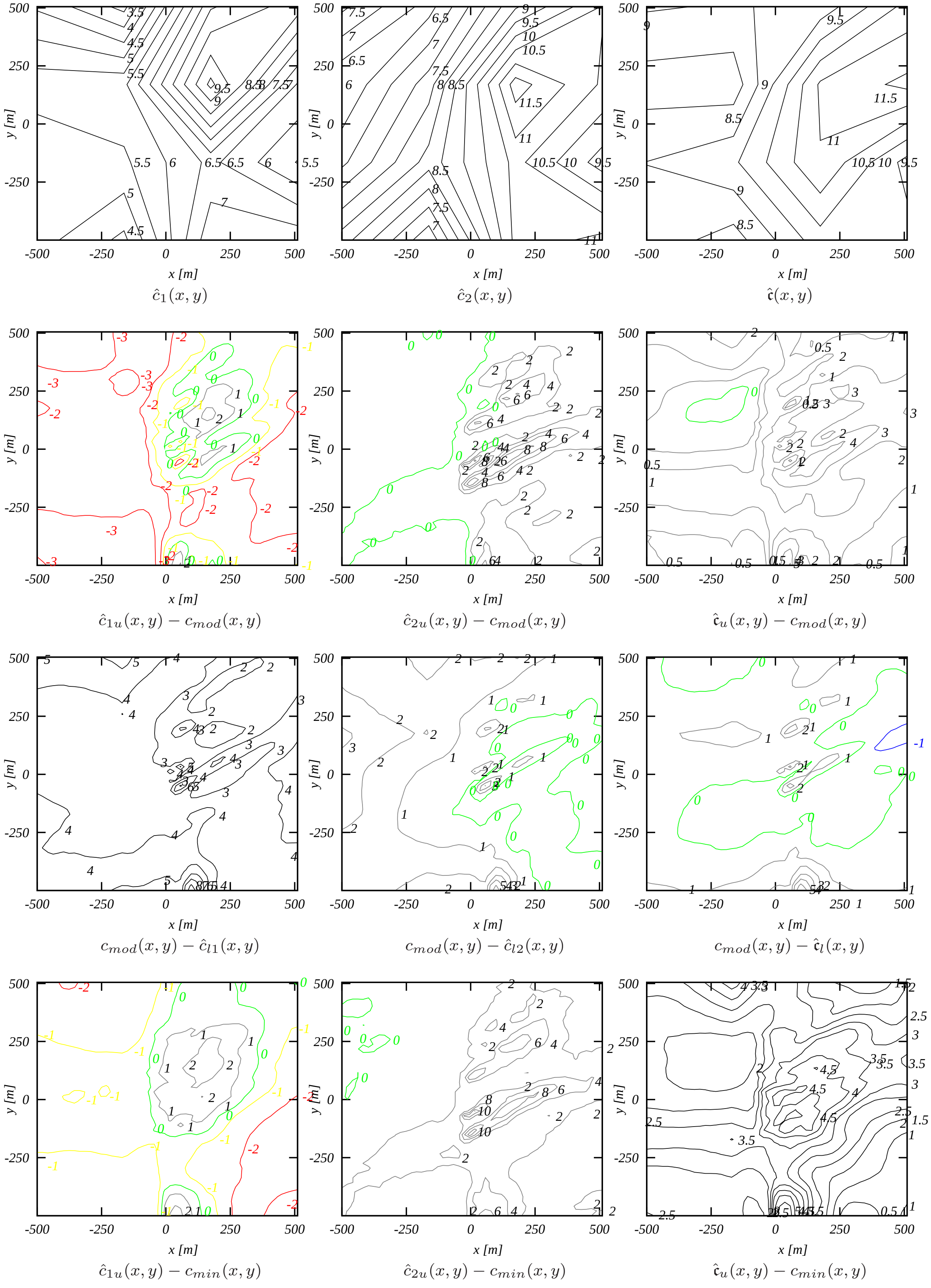
Additional information like II or III *has* to be provided to enable discrimination of the structures of the unkwown true distribution within the reconstruction error of the above tomographic setup.

10.3.3. Case studies

Simulated measurements are now studied for the 'true' 2-D NO_2 concentration being c_1, c_2 or $\mathfrak{c}$ from fig. 10.7 and the model prediction always being c_{mod} from figs. 10.3a,10.4a. The fields c_1 and c_2 represent the case of true distributions consistent with the model bounds. The field $\mathfrak{c}$ in fig. 10.7c serves as an example of the inconsistent case. (As discussed in sec. 7.4, to be correct this inconsistent case requires an independent estimate of the reconstruction error. With no such estimate available, we will use the above estimate of the reconstruction error.)
Reconstructions with and without measurement errors are treated seperately in the following.

Measurement noise can be neglected

If the experimental error plays only a minor role in the tomographic measurement, the parameters of the reconstruction are like in fig. 10.4c. 2-D maps reconstructed from simulated measurements of c_1, c_2 and $\mathfrak{c}$ are shown in the top panel of tab. 10.1 on page 175.
Ad (i): To find out whether these reconstructed fields $\hat{c}_1$, $\hat{c}_2$ and $\hat{\mathfrak{c}}$ are consistent with the model distribution c_{mod}, differences according to fig. 10.5a are formed, indicating inconsistency whenever



$\hat{c}_1(x,y)$ $\hat{c}_2(x,y)$ $\hat{\mathfrak{c}}(x,y)$

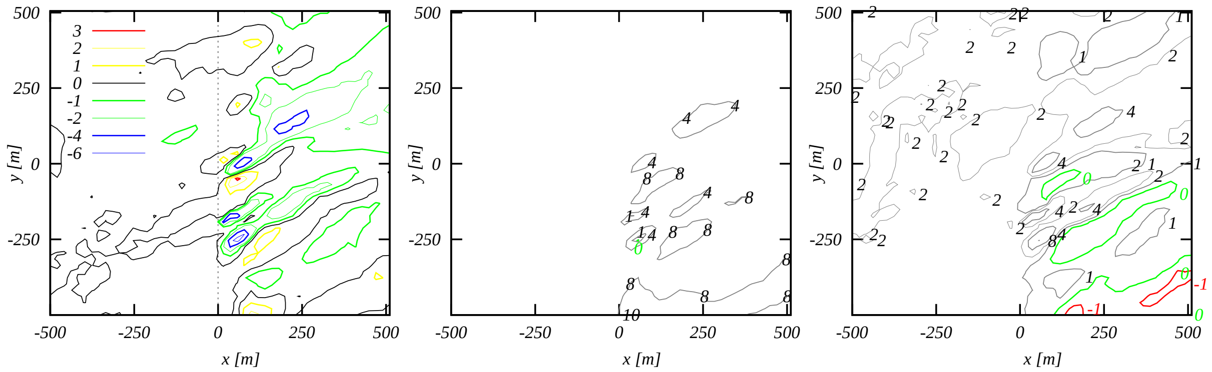
$\hat{c}_{1u}(x,y) - c_{mod}(x,y)$ $\hat{c}_{2u}(x,y) - c_{mod}(x,y)$ $\hat{\mathfrak{c}}_u(x,y) - c_{mod}(x,y)$

$c_{mod}(x,y) - \hat{c}_{l1}(x,y)$ $c_{mod}(x,y) - \hat{c}_{l2}(x,y)$ $c_{mod}(x,y) - \hat{\mathfrak{c}}_l(x,y)$

$\hat{c}_{1u}(x,y) - c_{min}(x,y)$ $\hat{c}_{2u}(x,y) - c_{min}(x,y)$ $\hat{\mathfrak{c}}_u(x,y) - c_{min}(x,y)$

(a) 'true' distribution c_1 [*ppb*] *(b)* 'true' distribution c_2 [*ppb*] *(c)* 'true' distribution $\mathfrak{c}$ [*ppb*]

Table 10.1: Reconstruction of a simulated, error free measurement (1^{st} panel from the top, parameters as in fig. 10.4c), consistency with the model c_{mod} (2^{nd} & 3^{rd} panel) and with the model bounds (bottom panel) according to fig. 10.5 for the distributions c_1, c_2 and $\mathfrak{c}$ from fig. 10.7. Red and yellow (and green) indicate that the model overestimates, blue (and green) that it underestimates the 'true' distribution.



(a) $c_{mod}(x, y) - \mathfrak{c}(x, y)$ [ppb] (b) $c_{max}(x, y) - \mathfrak{c}(x, y)$ [ppb] (c) $\mathfrak{c}(x, y) - c_{min}(x, y)$ [ppb]

Figure 10.9: Evaluation of (a) the model c_{mod} and (b,c) the model 1σ bounds by infinitely many ideal point measurements of $\mathfrak{c}$ covering the whole reconstruction area. In (a) these measurements would find a pattern of over-and underestimations above the street canyon (dashed line) and in the vicinity from the right.

they become negative. Negative values of the field $\hat{c}_u - c_{mod}$ in the second panel from the top show where model concentrations are too high to agree with the measurements. In the same way, $c_{mod} - \hat{c}_l < 0$ means that model values are too low.

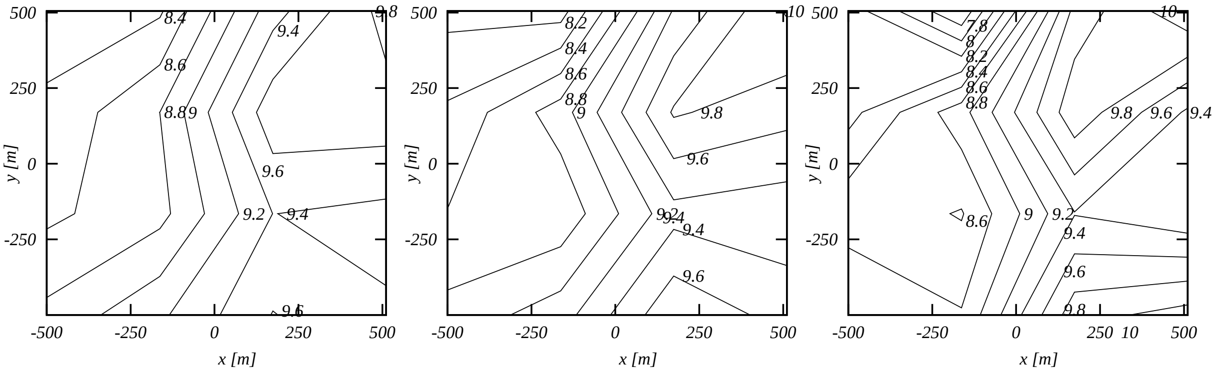
Within the 1σ bounds of the reconstruction error defined by eq. (10.2) the model quite correctly would agree with none of the 'true' distributions estimated from the measurement. The discrepancies identified from the error analysis agree well with what one would expect from comparing the original distributions in figs. 10.7g-10.7i with c_{mod} in fig. 10.4a. But the absolute numerical values of the differences below 1 *ppb* are rather small, except for the case of c_1 where the background concentration is considerably overestimated by the model c_{mod}.

Ideal point measurements of the real distribution immediately allow to compare true and model concentrations and thus model verification at the locations of the measurement. But a sufficient number of these samples at the right places is crucial in order to get a consistent picture of the model performance, as illustrated by fig. 10.9a where the concentration field $\mathfrak{c}$ representing the 'true' distribution is compared to the model prediction. While point samples in the left half of the picture and near the right boundary would be representative of a larger area round the place where they were taken, the interesting region above the street canyon requires quite a few measurements to get the pattern of over- and underestimation. This is exactly the problem of the representativeness of point measurements in an area with complex buildings addressed by *Schlünzen et al.* [2003] (c.f. sec. 2.1).

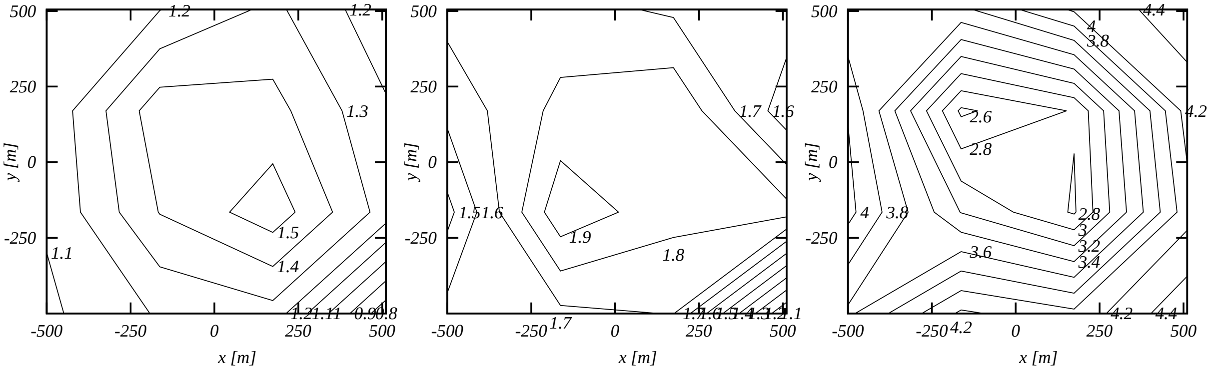
Ad (*ii*): We define the lower and upper model bounds c_{min}, c_{max} of the ensemble defined by I-III through the 1σ bounds

$$c_{max/min} = \langle c_I(x, y,) \rangle_{\mathcal{E}} \pm \mathrm{std}_{\mathcal{E}}[c_I(x, y,)] \tag{10.3}$$

with $\mathrm{std}_{\mathcal{E}}[c_I]$ as in fig. 10.8a. It turns out that the distributions c_1 and c_2, which are in $\mathcal{E}$, violate the 1σ bounds, while the field $\mathfrak{c} \notin \mathcal{E}$ is consistent with them (tab. 10.1, bottom panel). The seeming inconsistency of the former, especially c_1, in the region without peaks is due to the fact that their background concentrations take values at the lower end of admissible values (see also the footnote on page 172). The differences between the fields c_{mod} and $\mathfrak{c}$ appear mainly in the area of the peaks and cannot be resolved by the tomographic setup, see also fig. 10.9.



(a) $\hat{c}(x,y)$ [ppb], $k=25$ (b) $\hat{c}(x,y)$ [ppb], $k=50$ (c) $\hat{c}(x,y)$ [ppb], $k=500$



(d) $std_{\epsilon}[\hat{c}(x,y)]$ [ppb], $k=25$ (e) $std_{\epsilon}[\hat{c}(x,y)]$ [ppb], $k=50$ (f) $std_{\epsilon}[\hat{c}(x,y)]$ [ppb], $k=500$

Figure 10.10: Regularised solutions for a simulated measurement of c_{mod} (a-c) and their standard deviation for a relative error of 10% on the synthetic column densities (d-f) for different values of the iteration number. All other parameters are as in fig. 10.4. Subfigure (c) is the same as fig. 10.4c.

Measurement noise has to be taken into account

The impact of random noise on the reconstruction, i.e. the perturbation error in terms of sec. 5.2.3, can be estimated along the lines of sec. 8.3.2. Equation (5.13b) takes again the form

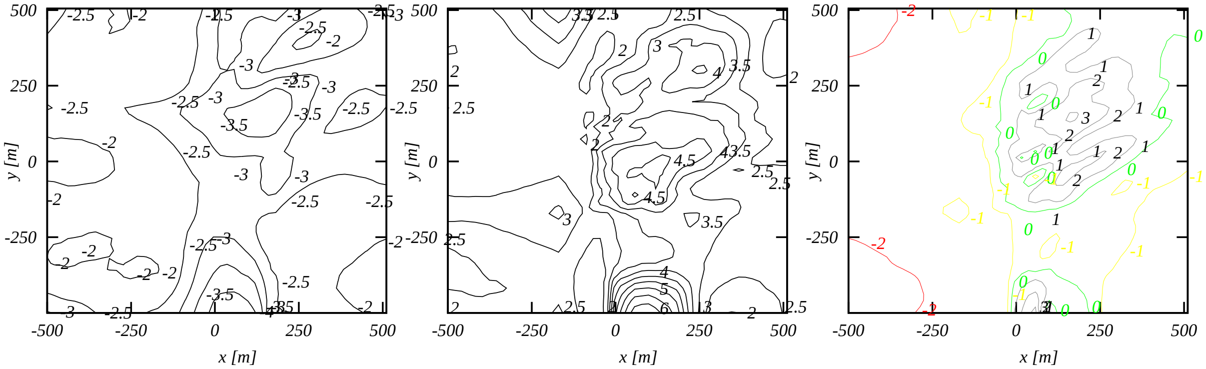
$$\mathrm{E}\big[\|\Delta\mathbf{x}_{pert}\|_2^2\big] \sim n\,m\,(\Delta x)^2 = \sum_i \Big(\frac{1-(1-\bar{\sigma}_i^2)^k}{\bar{\sigma}_i}\Big)^2,$$

where Δx represents the mean perturbation error of the state vector components, k is the iteration number and, as usual, n, m are the number of grid points and light paths, respectively. Introducing the mean relative error r_ϵ of the column densities, the spatial mean concentration $\bar{c}$ and a typical number n_ϕ of basis functions contributing to a light path ($\sim 3-5$ in this case) gives

$$|\Delta x| \sim r_\epsilon\,\bar{c}\,\sqrt{\frac{n_\phi}{n\,m}}\,\sqrt{\sum_i \Big(\frac{1-(1-\bar{\sigma}_i^2)^k}{\bar{\sigma}_i}\Big)^2}.$$

For relative errors of about 10%, $\bar{c} \sim 9\,ppb$ and $n_\phi = 4$ one gets

$$|\Delta x| \sim 1,\ 1.4,\ 3.2\,ppb \qquad \text{for} \quad k = 25,\ 50,\ 500.$$



(a) $\Delta\hat{c}_{mod\,l}(x,y)$ [ppb] (b) $\Delta\hat{c}_{mod\,u}(x,y)$ [ppb] (c) $\hat{c}_{1u}(x,y) - c_{mod}(x,y)$ [ppb]

Figure 10.11: Total reconstruction error including noise. (a,b) Upper and lower bounds $\Delta\hat{c}_{u/l} = \langle\Delta_\pm\hat{c}\rangle_\mathcal{E} \pm std_\mathcal{E}[\Delta_\pm\hat{c}] \pm std_\epsilon(\hat{c})$ *of the reconstruction error for the distribution reconstructed from a simulated measurement of* c_{mod}*. (c) Consistency of* c_{mod} *and the reconstruction from a measurement of* c_1 *(same as in tab. 10.1(a), but now with noise). The iteration number is* $k = 50$.

One finds that *the mean perturbation is substantial, its numerical value similar to the noiseless part of the reconstruction error (fig. 10.8) and it strongly depends on the choice of the regularisation parameter* k.

For the model distribution c_{mod} this is further pointed out in fig. 10.10, which shows regularised solutions for a simulated measurement (a-c) and their standard deviations for 10% noise (d-e) for different degrees of regularisation. Clearly, the iteration number in this case has more impact on the perturbation than on the least-squares solution. Not only absolute values of the standard deviation, but also the spatial patterns vary with k (for example, maximum values shift from the interior to the boundaries of the reconstruction area with growing iteration number). The weaker regularised case with $k = 500$ is very similar to the stochastic regularisation by the optimal estimate above (see fig. 10.6, p. 171), agreeing with the fact that the a priori is not very well constrained. Adding the perturbation error for the column densities from c_{mod} to the reconstruction error without noise (c.f. fig. 10.8) according to eq. (5.14e) leads to the bounds of the total reconstruction error shown in fig. 10.11a and 10.11b. These now have become so large that model evaluation is hardly possible within the bounds given by I-III, and among the distributions considered in the noisefree case only distribution c_1 with very low background concentrations can safely be distinguished from the model c_{mod} (fig. 10.11c).

Although evaluation of the specific model distribution here with low absolute concentration variations in the reconstruction area was found to be hard, if not impossible for realistic measurement errors, it appears fair to conclude that estimating the bounds of the reconstruction error from an ensemble of suitable random disributions is a consistent numerical scheme not only helpful to narrow down the uncertainty of the tomographic retrieval, but also a precise tool for the evaluation of models.

11. Conclusion and Outlook

11.1. Conclusions

This thesis theoretically investigates the possibility to retrieve 2-D distributions of trace gas concentrations from active DOAS measurements along a number of m light paths that is realistic at the current experimental state ($m \sim 10 - 40$).

The conventional aspects of the retrieval method can be summarised as follows: The approach discretises the inverse problem of finding the concentration field for measured column densities by parametrising the unknown distribution by a limited number of n local basis functions. The resulting linear discrete inverse problem is replaced by a least-squares least-norm problem for the discrete state vector, which is solved iteratively by the simultaneous iterative or algebraic reconstruction technique (SIRT and ART), commonly used in computerised tomography. Reasons for this approach were given in sections 4.2 and 6.1. The choice of the iterative solution was justified in section 4.8.1 by its flexibility with respect to additional (in)equality constraints. It also implies a specific regularisation behaviour depending on the iteration number.

The important novel contributions to tomography with a low number of integration paths are: A *systematic* investigation of the parametrisation with respect to the shape of the concentration field, including the number of parameters, the kind of basis function (piecewise constant and linear) and new schemes that take into account several reconstruction grids. While the focus is usually put on the inversion algorithm, it was shown here for peak distributions, which represent an important class of atmospheric concentration fields, that the parametrisation plays an equally, if not more important role. A second aspect that becomes vital for tomography with low spatial resolution, but which is mostly ignored, is the treatment of the *complete* reconstruction error, including the discretisation. The detailed approach of this thesis has revealed that the very common piecewise constant (box) functions should be avoided whenever possible – even for the reconstruction of spatial mean values. A systematic analysis of light path geometries using *alternative* arguments has pointed out not only their tremendous impact on the quality of the reconstruction, but also given insight into the reasons, including the role of measurement errors. Furthermore, it was suggested how numerical results can, to a certain degree, be generalised for arbitrary numbers of light paths.
Besides, it was attempted throughout to present similar and alternative concepts from different disciplines in a common context in order to put them into perspective (e.g., sections 4.2, 4.8.1, 4.8.3, 4.9) and to keep the discussion open for future developments (e.g., sections 4.6.3 and 4.7).

In more detail, the findings of this thesis can be stated in the following way.

The systematic approach to simulated tomographic measurements of 2-D Gaussian distributions (**‘peak distributions’**) with respect to their extension and parametrisation in chapter 8 shows that

1. for narrow emission peaks the reconstruction quality in terms of root mean square error, bias and total emissions can be tremendously improved by choosing parametrisations that lead to under-determined least-squares problems ($m < n$), provided that underlying background concentrations are known, negligible or smooth enough to be subtracted as proposed in section 6.3.2 (a quantification of 'smooth enough' was given in section 9.4) – in other words, that a zero concentration is a good a priori (after subtraction).
   At the same time, this means that the least-norm selection criterion and the nonnegativity constraint becomes more, the perturbation error less important (sections 4.4, 8.3.2 and tab. 8.11).

2. for the under- or even-determined solutions of 1. the bilinear parametrisation not only gives rise to smaller discretisation errors (which is evident) but also to smaller inversion errors (which is *not* evident).
   This holds also for bias and total emissions, i.e. for mean values of the concentration field.

3. the least-norm solution appears to be appropriate, at least no evidence against it was found.

4. the effectiveness of any of the schemes combining shifted grids by taking either the average, the maximum node values or the average while keeping the absolute maximum fixed strongly depends on the distribution and the feature of the reconstruction one is most interested in. The averaging scheme is 'safe' in that it generally reduces the root mean square error of the reconstruction.

5. from an experimental point of view, the temporal resolution of the measurement is as important as the spatial resolution if the concentration peaks represent emission puffs subject to wind transport (sec. 5.2.3).

The discussion of the fan beam geometries in section 8.4 reveals that

6. increasing the number of emitting and/or receiving systems while keeping the number of light paths fixed generally leads to better reconstruction results (provided that the geometry does not get too irregular, producing large gaps).

7. for the small numbers of light paths considered in this thesis, making the fans wider in general deteriorates the reconstruction quality.

8. for the peak distributions according to 1. the sensitivity to measurement errors does not vary much with the geometry, but can vary within the recontruction area.

9. numerical results of simulations for a certain geometry can be carried forward to geometries with different numbers of light paths, but of a similar type, by *scaling* the distributions with the mesh size of the geometry (section 8.4.4).

   Furthermore, the discussion presents the singular value decomposition as a useful diagnostic tool for the algebraic properties of the measurement system, independently of the distributions (see also section 9.4, fig. 9.6), and the optimal estimate as a simple method to analyse different physical scenarios without having to perform lengthy simulations (which in this case leads to very similar results, c.f. section 8.4.2).

Taking the indoor experiment (chapter 9) as an example of a 'real world' measurement of emission puffs where the true concentration distribution is not known proves that

ad 1.&3. for peak distributions it is possible to optimise the discretisation grid in a consistent way *without* knowing the true width of the peaks.
Comparing reconstructions from different grid combination approaches and from a single grid can lead to a better picture of the true concentration field (because the schemes are sensitive to different features of the peak) and thus to the a posteriori preference of a certain scheme.

10. the numerical approach to estimate the complete reconstruction error field from reconstructions of admissible (random) distributions leads to sensible upper and lower bounds of this field and allows to identify artefacts in the original reconstruction.
    It is important to get the variability of these admissible distributions right in order to neither over- nor underestimate the reconstruction error.

Finally, applying the same reconstruction schemes to highly resolved model distributions (**'smooth distributions'**[1]) above a city street canyon (chapter 10) suggests that

11. if the model distributions reflect the true variability of trace gas concentrations in the environment of complex buildings, their tomographic reconstruction is a rather challenging task.

12. contrary to the peak distributions, for the – on a larger scale – smooth model distribution the over-determined solution achieves minimal overall reconstruction errors and neither grid combination schemes nor subtraction of the background improves the reconstruction.
    The severe semi-convergence of the *under*-determined solution suggests that the least-norm selection criterion might not be appropriate for this kind of distribution, but comparison with alternative reconstruction principles for realistic atmospheric distributions is necessary to clear this point.

13. data errors have strong impact now and, therefore, regularisation becomes crucial. The influence of the basis functions on individual contributions to the reconstruction error in the ill-posed over-determined case has not been examined, but it is known, for example from atmospheric profile retrieving [e.g., *Doicu et al.*, 2004], that the choice of the basis functions affects regularisation (c.f. 4.6.3(*iv*)).
    Therefore, from an experimental point of view the ill-posedness of the problem demands *minimal* data errors.

14. contrary to a point measurement that – within its errors – can *directly* be compared to a model prediction at the measurement site, the interpretation of a tomographically reconstructed field is *not* possible without information on the true variability of the atmospheric trace gas.
    The numerical scheme proposed here, essentially parametrising the a priori concentrations by functional bounds and bounds for the gradients, leads to distinct error patterns consistent with the true distributions. Tomographic measurements with *low* spatial resolution can be used to verify specific model predictions if only the uncertainty of the model can be narrowed down sufficiently.

---

[1] Which here means they cannot be represented by narrow peaks on a moderately smooth background

11.2. Outlook

The results of this work will be applied to the measurements presently taking place in Heidelberg that were briefly addressed in sec. 9.4. Moreover, current and future technical development of active DOAS instruments, making these smaller, cheaper and easier to handle, will reduce the expense to set up tomographic experiments with eventually increasing numbers of light paths. With similar progress of the passive technique and all results of this thesis equally holding for any remote sensing method with well defined light paths, future tomographic MAX-DOAS measurements proposed by *Frins et al.* [2006] (c.f. sec. 3.4) would represent an extremely versatile application. Tomographic DOAS measurements from satellite similar to the suggestion by *Fleming* [1982] (sec. 3.3) are a further example that allows straightforward use of our results. In principle, everything said holds for IR and LIDAR remote sensing as well.

Given this perspective, certain aspects of the reconstruction procedure should be reconsidered, where we only refer to the discrete approach using basis functions with local support. For applications with regular light path geometries like satellite measurements – possibly air craft measurements – c.f. fig. 3.5, transform methods might become feasible (sec. 4.2.1). Methods like the explicit fit of Gaussian peaks to the measurement data (SBFM, c.f. 4.9.3) require further investigation, especially with respect to the stabilty of their solution (but distributions like those in fig. 10.3 do not look very promising for this special approach).

Basis functions

While it was explicitly shown in this thesis that linear parametrisation is superior to piecewise constant parametrisation in every respect for a *regular* reconstruction area, the implementation of the bilinear discretisation given by eq. (4.13) is not very flexible when it comes to parametrising concentration fields on an irregular area with only few basis functions. Different schemes, like triangular discretisation, might be more appropriate.

Another approach worth looking at is the parametrisation by higher order polynoms. This does make the problem nonlinear, but B-splines, for example (c.f. sec. 4.3.2), have excellent approximation and, as reported by [e.g., *Doicu et al.*, 2004], regularisation properties. Furthermore, they are very flexible with respect to their knots. [*Baussard et al.*, 2004] use a *small* number of *adaptively* chosen B-splines to parametrise their inverse problem, giving rise to the expectation of small discretisation and inversion errors.

Reconstruction principle

In image reconstruction the least-norm principle is frequently replaced by a maximum-entropy principle, or altogether by a maximum likelihood approach (sec. 4.9.1). Whether these would give better results for atmospheric reconstruction problems remains to be shown. By the same token, the iterative regularisation of the least-norm solution adopted here from image reconstruction should be compared with alternative methods commonly used for atmospheric inverse problems. This would be, in particular, the Tikhonov method, which if augmented with a nonnegativity constraint is essentially the constrained optimisation principle eq. (4.51). This formulation of the reconstruction problem would become especially attractive if the regularisation parameter δ in eq. (4.51) for a given regularisation matrix D could be *quantitatively* derived from atmospheric parameters. For example, if D is given by the first differences from maximum gradients of the concentration field. *Fehmers* [1996] employed this method for tomographic reconstruction of the ionosphere (see also fig. 3.5a, p. 44).[2]

Estimation of the reconstruction error

The Monte Carlo estimation of the reconstruction error proposed in this thesis is simple and flexible, but also time consuming and has the disadvantage of taking into account also distributions that do not agree with the measurements. A scheme that is consistent in this respect would be highly desirable. Using the a posteriori covariance of the discrete optimal estimate is only possible if the a priori can be formulated in terms of Gaussian probability densities *and* the discretisation error can be neglected (provided that the least-squares approach has been chosen as reconstruction principle).

---

[2] Computer code for an algorithm solving this and more general constrained optimisation problems is available, e.g., from the Numerical Algorithms Group [http://www.nag.co.uk/]

Part IV.

Appendices

A. Atmospheric Stability Classes and Dispersion Coefficients

Stability classes

wind speed @10 m [m/s]	day incom. solar radiation (insolation)			night	
	strong	moderate	slight	thinly overcast or $\geq 4/8$ low cloud cover	$\leq 3/8$ cloud cover
< 2	A	$A-B$	B	F	F
$2-3$	$A-B$	B	C	E	F
$3-5$	B	$B-C$	C	D	E
$5-6$	C	$C-D$	D	D	D
> 6	C	D	D	D	D

Table A.1: Atmospheric stability categories based on wind speed and insolation (see text).

On the basis of experimental observations (the Nebraska Prairie Grass Project) *Pasquill* [1961] proposed to describe different states of the atmosphere by means of six categories A to F derived from readily available meteorological observables. Table A.1 defines them in terms of wind speed and insolation [e.g., *Barratt*, 2002]. Estimation of the insolation from solar elevation and sky cover can be found in [*Barratt*, 2002, table 3.9]. Table A.2 relates the classes to some meteorological parameters.

Pasquill stability	classification	phenomena	freq. of occur. [%]	mean wind speed [m/s]	boundary layer depth [m]	σ_α
A	extremely unstable	strong thermal instability, bright sun	0.125	0.625	1 300	25
$A-B$			1.25	1.25	1 080	
B	moderately unstable	transitional period moderate mixing	3.8	2.0	920	20
$B-C$			2.6	3.37	500	
C	slightly unstable	transitional periods, slight mixing	15	4.12	840	15
$C-D$			-	5.0	600	
D	neutral	strong winds, overcast day/night transitions	62.4	4.12	500	10
E	slightly stable	transitional periods, night-time mod. winds	6.7	3.4	400	5
F	mod. stable	clear night time skies, very lim. vert. mixing	-	2.0	150	2.5
$F-G$			8.4	1.2	50	
G	extremely stable					

Table A.2: Selected meteorological parameters related to stability categories [Barratt, 2002, tables 3.7,13]. The parameter σ_α was defined by eqs. (2.2) as $\sigma_\alpha = \sigma_v/\bar{u}$. Values are taken from [Blackadar, 1997, table 10.1].

Estimates of T_L for various atmospheric stabilities have been given by several authors. Those proposed by

Draxler are given in the following table (see *Blackadar* [1997] and references therein):

T_L $[s]$	surface source		elevated source	
	stable	unstable	stable	unstable
lateral	60	60	200	200
vertical	10	20	20	100

Dispersion coefficients

There are numerous parametrisations of the plume dispersion coefficients defined in sec. 2.4 [*Seinfeld and Pandis*, 1998]. Tables A.3, A.4 give functionals forms known as Pasquill-Gifford and Briggs curves, respectively [*Barratt*, 2002]. Sampling times are around $10\,min$.

Stability class	$x\,[m]$	$\sigma_x(x) = \sigma_y(x)\,[m]$	$\sigma_z(x)\,[m]$
A	$100 - 300$	$0.493\,x^{0.88}$	$0.087\,x^{1.10}$
	$300 - 3\,000$	-”-	$\log_{10}\sigma_z = -1.67 + 0.902\log_{10}x + 0.181\,(\log_{10}x)^2$
B	$100 - 500$	$0.337\,x^{0.88}$	$0.135\,x^{0.95}$
	$500 - 20\,000$	-”-	$\log_{10}\sigma_z = -1.25 + 1.09\log_{10}x + 0.0018\,(\log_{10}x)^2$
C	$100 - 100\,000$	$0.112\,x^{0.91}$	$0.112\,x^{0.91}$
D	$100 - 500$	$0.128\,x^{0.90}$	$0.093\,x^{0.85}$
	$500 - 100\,000$	-”-	$\log_{10}\sigma_z = -1.22 + 1.08\log_{10}x + 0.061\,(\log_{10}x)^2$
E	$100 - 500$	$0.091\,x^{0.91}$	$0.082\,x^{0.82}$
	$500 - 100\,000$	-”-	$\log_{10}\sigma_z = -1.19 + 1.04\log_{10}x + 0.070\,(\log_{10}x)^2$
F	$100 - 500$	$0.067\,x^{0.90}$	$0.057\,x^{0.80}$
	$500 - 100\,000$	-”-	$\log_{10}\sigma_z = -1.91 + 1.37\log_{10}x + 0.119\,(\log_{10}x)^2$

Table A.3: Pasquill-Gifford parametrisation of the dispersion coefficients.

Stability class	$\sigma_x(x) = \sigma_y(x)\,[m]$	$\sigma_z(x)\,[m]$
Open country		
A	$0.22\,x\,(1 + 10^{-4}\,x)^{-0.5}$	$0.20\,x$
B	$0.16\,x$ (-”-)$^{-0.5}$	$0.12\,x$
C	$0.11\,x$ (-”-)$^{-0.5}$	$0.08\,x(1 + 2\cdot 10^{-4}\,x)^{-0.5}$
D	$0.08\,x$ (-”-)$^{-0.5}$	$0.06\,x(1 + 1.5\cdot 10^{-3}\,x)^{-0.5}$
E	$0.06\,x$ (-”-)$^{-0.5}$	$0.03\,x(1 + 3\cdot 10^{-4}\,x)^{-1}$
F	$0.04\,x$ (-”-)$^{-0.5}$	$0.016\,x(1 + 3\cdot 10^{-4}\,x)^{-1}$
Urban		
$A - B$	$0.32\,x\,(1 + 4\cdot 10^{-4}\,x)^{-0.5}$	$0.24\,x(1 + 10^{-3}\,x)^{+0.5}$
C	$0.22\,x$ (-”-)$^{-0.5}$	$0.20\,x$
D	$0.16\,x$ (-”-)$^{-0.5}$	$0.14\,x(1 + 3\cdot 10^{-4}\,x)^{-0.5}$
$E - F$	$0.11\,x$ (-”-)$^{-0.5}$	$0.08\,x(1 + 1.5\cdot 10^{-4}\,x)^{-0.5}$

Table A.4: Briggs parametrisation of the dispersion coefficients for rural and urban conditions.

B. Generation of Random Test Distributions

We consider the case that the 2-D random scalar fields $c_I(\mathbf{r})$, $I = 1, \ldots, N$, are generated from a given function $c(\mathbf{r})$ in the following way.

The $c_I(\mathbf{r})$ are defined on a finite number of grid nodes $\mathbf{r}_J$, $J = 1, \ldots, n_x \times n_y$ such that

1. for the function values on the nodes

$$c_{min}[c(\mathbf{r}_J)] \le c_I(\mathbf{r}_J) \le c_{max}[c(\mathbf{r}_J)], \tag{B.1}$$

   i.e. lower and upper bounds on each node are functions of c's value on it.
   For example,

$$f_{min}\, c(\mathbf{r}_J) \le c_I(\mathbf{r}_J) \le f_{max}\, c(\mathbf{r}_J), \tag{B.2a}$$

   like a factor of 2 uncertainty, or

$$c(\mathbf{r}_J) - \Delta[c(\mathbf{r}_J)] \le c_I(\mathbf{r}_J) \le c(\mathbf{r}_J) + \Delta[c(\mathbf{r}_J)], \tag{B.2b}$$

   like a ±50% uncertainty.

Depending on the spacing of the nodes $\mathbf{r}_J$, this constraint on its own can result in much larger gradients than the ones of the original field c, and for very fine grids it leads to unphysically fluctuating values. Therefore, it is necessary to restrict the gradients and possibly higher derivatives of c_I. Here, we restrict ourselves to the first differences and a constraint of the special form

2. for the first differences

$$f'_{min/max} \frac{\partial c(\mathbf{r}_J)}{\partial x} \le \frac{\Delta c_I(\mathbf{r}_J)}{\Delta x} \le f'_{max/min} \frac{\partial c(\mathbf{r}_J)}{\partial x} \qquad \text{for} \quad \frac{\partial c(\mathbf{r}_J)}{\partial x} \gtrless 0 \tag{B.3}$$

   and the same for y. To generate variability where the original field is very smooth, e.g. $|\nabla c| \sim 0$, this sign preserving relation is relaxed to

$$f'_{min} \Big| \frac{\partial c(\mathbf{r}_J)}{\partial x} \Big| \le \Big| \frac{\Delta c_I(\mathbf{r}_J)}{\Delta x} \Big| \le f'_{max} \Big| \frac{\partial c(\mathbf{r}_J)}{\partial x} \Big| \qquad \text{if} \quad \Big| \frac{\partial c(\mathbf{r}_J)}{\partial x} \Big| \le \text{some treshold}, \tag{B.4}$$

   and the same for y.

We further consider two different implementations of the constraints.

(*i*) Direct Monte-Carlo generation of the $c_I(\mathbf{r}_J)$ such that one of the equations (B.2) and eq,. (B.3), (B.4) are satisfied, starting on a certain grid node. The gradients of c have to be computed beforehand.

(*ii*) 'Convolution'

$$c_I(\mathbf{r}) = r(\mathbf{r}) \cdot c(\mathbf{r})$$

with a random field r constrained by

$$r_{min} \leq r(\mathbf{r}_J) \leq r_{max}, \tag{B.5}$$

$$\left|\frac{\Delta r(\mathbf{r}_J)}{\Delta x}\right| \leq r'_{max} \qquad \text{and the same for } y, \tag{B.6}$$

such that eq. (B.2a) with $f'_{min} = f'_{max}$ and eq. (B.3) are fulfilled. This implies

$$\begin{aligned} r_{min/max} &= f_{min/max}, \\ r'_{max} &= \frac{f'_{max} - f_{max}}{c(\mathbf{r}_J)} \max(\left|\frac{\Delta r(\mathbf{r}_J)}{\Delta x}\right|, \left|\frac{\Delta r(\mathbf{r}_J)}{\Delta y}\right|). \end{aligned}$$

The field r is produced as in (*i*).

Both methods have been tested and, for the right choice of parameters, lead to similar results. In any case, the mean and standard deviation of all c_I should be checked for bias and actual variability within the above bounds.

Unless stated differently, all random numbers are *uniformly* distributed.

C. Auxiliary Calculations

Proof of equation (5.8a): Estimate of the discretisation error

It is

$$\Delta c_{disc}(\mathbf{r}) = c(\mathbf{r}) - [\boldsymbol{\phi}(\mathbf{r})]^T \Phi^{-1} \mathbf{c}$$

with

$$\begin{aligned} c_j &= \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r}) c(\mathbf{r}) \\ \Phi_{jj'} &= \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r}) [\phi_{j'}(\mathbf{r})]^T \end{aligned}$$

and Ω_j the support of basis function ϕ_j. Expand c_j around a point $\boldsymbol{\rho} = (x_\rho, y_\rho)$ in Ω_j:

$$\begin{aligned} c_j \sim c(\boldsymbol{\rho}) & \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r}) \\ & + \frac{\partial c}{\partial x}(\boldsymbol{\rho}) \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r})(x - x_\rho) \quad + \quad (x \leftrightarrow y) \\ & + \frac{1}{2} \frac{\partial^2 c}{\partial x^2}(\boldsymbol{\rho}) \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r})(x - x_\rho)^2 \quad + \quad (x \leftrightarrow y) \\ & + \frac{\partial^2 c}{\partial x \partial y}(\boldsymbol{\rho}) \int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r})(x - x_\rho)(y - y_\rho). \end{aligned}$$

Taking $\boldsymbol{\rho}$ as the grid node $\mathbf{r}_j$ in the case of the bilinear basis and as the box centre in the case of the box basis, linear terms in x and y vanish for pixels entirely in the reconstruction area. The integral in the first contribution is

$$\int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r}) = \Delta x_i \Delta y_i = \Delta A_i,$$

where ΔA_i is the area of the rectangle formed by four grid nodes and in the case of the bilinear basis the grid is assumed to be regular in both x-and y-direction merely for simplicity. Similarly, for the quadratic term

$$\int_{\Omega_j} d\mathcal{V}\ \phi_j(\mathbf{r})(x - x_j)^2 = \begin{cases} \frac{1}{12}\ \Delta A_j (\Delta x_j)^2 & \text{box} \\ \frac{1}{6}\ \Delta A_j (\Delta x_j)^2 & \text{regular bilinear} \end{cases},$$

and the same for y. The vector $\mathbf{c}$ gets thus

$$c_j \sim \begin{cases} \Delta A_j \left(c(\mathbf{r}_j) + \frac{1}{24} \Big((\Delta x_j)^2 \frac{\partial^2 c}{\partial x^2} + (\Delta y_j)^2 \frac{\partial^2 c}{\partial y^2} \Big)(\mathbf{r}_j) \right) & \text{box} \\ \Delta A_j \left(c(\mathbf{r}_j) + \frac{1}{6} \Big((\Delta x_j)^2 \frac{\partial^2 c}{\partial x^2} + (\Delta y_j)^2 \frac{\partial^2 c}{\partial y^2} \Big)(\mathbf{r}_j) \right) & \text{regular bilinear} \end{cases},$$

which allows to estimate $\Delta c_{disc}(\mathbf{r})$ if Φ^{-1} is known explicitely. For the box basis $\Phi_{ij}^{-1} = \Delta A_i^{-1} \delta_{ij}$ gives the desired expression. The calculation of Φ^{-1} in the bilinear case is more involved and not considered here.

Proof of equations (6.11) & (6.12): Fitting the background

In the *over-determined* case the least-squares problem of eq. (6.10b)

$$\min \|\mathbf{d} - B\mathbf{y}\|_2, \qquad \text{with} \qquad \begin{aligned} B &= (A \quad \boldsymbol{l}) \\ \mathbf{y} &= (\mathbf{x}', \quad c_{BG})^T \end{aligned}$$

is given by

$$\begin{aligned} \mathbf{y} &= (B^T B)^{-1} B^T \mathbf{d} \\ &= \begin{pmatrix} A^T A & A^T \boldsymbol{l} \\ (A^T \boldsymbol{l})^T & \boldsymbol{l}^T \boldsymbol{l} \end{pmatrix}^{-1} \begin{pmatrix} A^T \mathbf{d} \\ \boldsymbol{l}^T \mathbf{d} \end{pmatrix}. \end{aligned}$$

The inverse $I = \left(\begin{smallmatrix} J & \mathbf{j} \\ \mathbf{j}^T & j \end{smallmatrix}\right)$ in the last expression can be computed from

$$\begin{pmatrix} A^T A & A^T \boldsymbol{l} \\ (A^T \boldsymbol{l})^T & \boldsymbol{l}^T \boldsymbol{l} \end{pmatrix} I = \mathbb{1},$$

giving

$$\begin{aligned} \mathbf{j} &= -\big(A^T(\boldsymbol{l}^T\boldsymbol{l} - \boldsymbol{l}\boldsymbol{l}^T)A\big)^{-1} A^T \boldsymbol{l}, \\ J &= \boldsymbol{l}^T\boldsymbol{l}\big(A^T(\boldsymbol{l}^T\boldsymbol{l} - \boldsymbol{l}\boldsymbol{l}^T)A\big)^{-1}, \\ j &= (\boldsymbol{l}^T\boldsymbol{l})^{-1}\Big(1 + \boldsymbol{l}^T A\big(A^T(\boldsymbol{l}^T\boldsymbol{l} - \boldsymbol{l}\boldsymbol{l}^T)A\big)A^T\boldsymbol{l}\Big). \end{aligned}$$

Inserting these expressions into

$$\begin{aligned} \mathbf{x}' &= J A^T \mathbf{d} + (\boldsymbol{l}^T \mathbf{d})\mathbf{j}, \\ c_{BG} &= \mathbf{j}^T A^T \mathbf{d} + (\boldsymbol{l}^T \mathbf{d}) j \end{aligned}$$

yields equation (6.11).

The *under-determined* case

$$\min_{\mathbf{y}} \mathbf{y}^T H \mathbf{y}, \qquad \mathbf{d} = B\,\mathbf{y}$$

contains the norms $\|\mathbf{x}'\|_2^2$, $\|\mathbf{y}\|_2^2$ and $\sum_j (x'_j - c_{BG})^2$ by the choices

$$H = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}, H = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 1 & & & -1 \\ & \ddots & & \vdots \\ & & 1 & -1 \\ -1 & \dots & 1 & n \end{pmatrix},$$

respectively. The normal equations of the second kind for the above minimum-norm principle

$$H\mathbf{y} = \frac{1}{2} B^T \boldsymbol{\gamma}, \quad \mathbf{d} = B\mathbf{y}$$

cannot be solved straightforwardly if H is singular, which is the case for the last two matrices H. Writing

$$H = \begin{pmatrix} \mathcal{H} & \mathbf{h} \\ \mathbf{h}^T & h \end{pmatrix}$$

with non-singular $\mathcal{H}$, the normal equations take the form

$$\begin{aligned}\mathcal{H}\mathbf{x}' + \mathbf{h}c_{BG} &= \frac{1}{2}A^T\boldsymbol{\gamma},\\ \mathbf{h}^T\mathbf{x}' + hc_{BG} &= \frac{1}{2}\boldsymbol{l}^T\boldsymbol{\gamma},\\ A\mathbf{x}' + \boldsymbol{l}c_{BG} &= \mathbf{d}\end{aligned}$$

with solution

$$\begin{aligned}c_{BG} &= N^{-1}\,\mathbf{v}^T(A\mathcal{H}^{-1}A^T)^{-1}\mathbf{d},\\ \mathbf{x}' &= \Big\{\mathcal{H}^{-1}A^T - N^{-1}\Big(\mathcal{H}^{-1}A^T(A\mathcal{H}^{-1}A^T)^{-1}\mathbf{v}\mathbf{v}^T - \mathcal{H}^{-1}\mathbf{h}\mathbf{h}^T\mathcal{H}^{-1}A^T + \mathcal{H}^{-1}\mathbf{h}\boldsymbol{l}^T\Big)\Big\}(A\mathcal{H}^{-1}A^T)^{-1}\mathbf{d},\end{aligned}$$

where

$$\begin{aligned}N &= \mathbf{v}^T(A\mathcal{H}^{-1}A^T)^{-1}\mathbf{v} - \mathbf{h}^T\mathcal{H}^{-1}\mathbf{h} + h,\\ \mathbf{v} &= \boldsymbol{l} - A\mathcal{H}^{-1}\mathbf{h}\end{aligned}$$

and $\boldsymbol{\gamma} = 2(A\mathcal{H}^{-1}A^T)^{-1}\big((A\mathcal{H}^{-1}\mathbf{h} - \boldsymbol{l})\,c_{BG} + \mathbf{d}\big)$. Equations (6.12) follow for the special choices of $\mathcal{H}$, $\mathbf{h}$ and h above.

The sum $\mathbf{x}' + c_{BG}(1,\dots,1)^T = \mathbf{x}' + c_{BG}\,\mathbf{e}$ can be written as

$$\mathbf{x}' + c_{BG}\,\mathbf{e} = \mathcal{H}^{-1}A^T(A\mathcal{H}^{-1}A^T)^{-1}\mathbf{d} + \Big(\mathbb{1} - \mathcal{H}^{-1}A^T(A\mathcal{H}^{-1}A^T)^{-1}A\Big)\mathbf{x}_a$$

where $A\,\mathbf{e} = \boldsymbol{l}$ was used and $\mathbf{x}_a = c_{BG}\,\mathcal{H}^{-1}(\mathcal{H}\mathbf{e} - \mathbf{h})$ can be interpreted as a priori (c.f., e.g. eq. (4.42)). Particularly, $\mathbf{x}_a = c_{BG}\,\mathbf{e}$ for the matrices H above.

Proof of equations (6.17) & (6.18): Additional constraints

In the *over-determined* case ($m + f \geq n$), the least-squares solution of $\mathcal{A}\mathbf{x} = \boldsymbol{d}$, with $\mathcal{A} = \begin{pmatrix} A \\ F \end{pmatrix}$ and $\boldsymbol{d} = \begin{pmatrix} \mathbf{d} \\ \mathbf{c} \end{pmatrix}$, in the form of eq. (6.17) immediately follows from

$$\mathbf{x} = (\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T\boldsymbol{d}$$

by the matrix multiplication in blocks:

$$\mathcal{A}^T\mathcal{A} = \begin{pmatrix} A^T & F^T \end{pmatrix}\begin{pmatrix} A \\ F \end{pmatrix} = A^TA + F^TF.$$

The *under-determined* case with a priori $\mathbf{x}_a$ has the solution

$$\mathbf{x} = \mathbf{x}_a + \mathcal{A}^T(\mathcal{A}\mathcal{A}^T)^{-1}(\boldsymbol{d} - \mathcal{A}\mathbf{x}_a),$$

where the calculation of the inverse $(\mathcal{A}\mathcal{A}^T)^{-1} = I = \left(\begin{smallmatrix} I_1 & I_3 \\ I_3^T & I_2 \end{smallmatrix}\right)$ is carried out as on page 190 and yields

$$\begin{aligned}
I_1 &= (AP_F A^T)^{-1}, \\
I_2 &= (FP_A F^T)^{-1}, \\
I_3 &= -(AP_F A^T)^{-1} AF^T (FF^T)^T = -(AA^T)^{-1} AF^T (FP_A F^T)^{-1},
\end{aligned}$$

with P_A defined by $P_A = \mathbb{1} - A^T(AA^T)^{-1}A$, see eq. (4.55), and the same for P_F.
For $\mathcal{A}^T(\mathcal{A}\mathcal{A}^T)^{-1} = \begin{pmatrix} A^T I_1 + F^T I_3^T & A^T I_3 + F^T I_2 \end{pmatrix}$ one gets thus

$$\begin{aligned}
A^T I_1 + F^T I_3^T &= P_F A^T (AP_F A^T)^{-1}, \\
A^T I_3 + F^T I_2 &= P_A F^T (FP_A F^T)^{-1},
\end{aligned}$$

which completes the proof.

Proof of equations (7.8)-(7.10): Point measurements as additional a priori

The augmented a priori $\boldsymbol{x}_a$ is constructed such that $f = \operatorname{rank}[F]$ of its degrees of freedom are determined by the point measurements, expressed as $F\mathbf{x} = \mathbf{c}$, and the remaining $n - f$ degrees from the ad hoc a priori $\mathbf{x}_a$. This is in fact a pure least-norm problem with solution according to eq. (4.56c)

$$\boldsymbol{x}_a = F^T(FF^T)^{-1}\mathbf{c} + P_F \mathbf{x}_a. \tag{C.1}$$

As it provides an instructive approach to the least-norm problem, we derive eq. (C.1) explicitly by using the singular value decomposition $F = U\Sigma V^T$ and write

$$x_{a_j} = \sum_{j'=1}^{f} V_{jj'} x'_{a_{j'}} + \sum_{j'=f+1}^{n} V_{jj'} x'_{a_{j'}}. \tag{C.2}$$

Inserting this into $F\boldsymbol{x}_a = U\Sigma\boldsymbol{x}'_a = \mathbf{c}$ gives the unique solution

$$x'_{a_j} = \Sigma_j^{-1}(U^T\mathbf{c})_j \quad \text{for} \quad j = 1, \ldots, f$$

by projecting onto the orthogonal complement of the nullspace. The remaining components are set to the components of $\mathbf{x}_a$ in the system of the singular vectors $\mathbf{v}$

$$x'_{a_j} = (V^T\mathbf{x}_a)_j \quad \text{for} \quad j = p+1, \ldots, n.$$

Inserting these equations back into eq. (C.1) gives

$$x_{a_j} = (\Sigma^\dagger U^T \mathbf{c})_j + \Sigma_{j'=f+1}^{n} V_{jj'} V_{j'k}^T x_{a_k},$$

where the matrix product in the second sum is the projector onto the nullspace of F and $\Sigma^\dagger = \begin{pmatrix} \Sigma_r^{-1} & 0 \end{pmatrix}$ as in eq. (4.36). Using the singular value decomposition, the first expression can be seen to be $(F^T(FF^T)^{-1}\mathbf{c})_j$, thus establishing eq. (7.8). Equation (7.9) follows immediately from eq. (4.36).

The augmented covariance matrix $\mathcal{S}_a$ can – like any matrix in $\mathbb{R}^{n\times n}$ – be decomposed in the following way

$$\mathcal{S}_a = P_F^\perp \mathcal{S}_a P_F^\perp + P_F^\perp \mathcal{S}_a P_F + P_F \mathcal{S}_a P_F^\perp + P_F \mathcal{S}_a P_F,$$

where P_F is again the projector onto the nullspace of F and $P_F^{\perp}$ the projector onto its orthogonal complement. The form eq. (7.10) follows from

1. $\mathbf{c} = F\boldsymbol{x}_a \quad \Rightarrow \quad P_F^{\perp}\mathcal{S}_a P_F^{\perp} = F^T(FF^T)^{-1}S_c(FF^T)^{-1}F$, with $S_c = \mathrm{E}\big[(\mathbf{c}-\bar{\mathbf{c}})(\mathbf{c}-\bar{\mathbf{c}})^T\big]$.
2. No correlation between the physical and unphysical subspaces related to the point measurements and the ad hoc a priori.
3. $P_F\mathcal{S}_a P_F = P_F S_a P_F$.

D. Software & Numerics

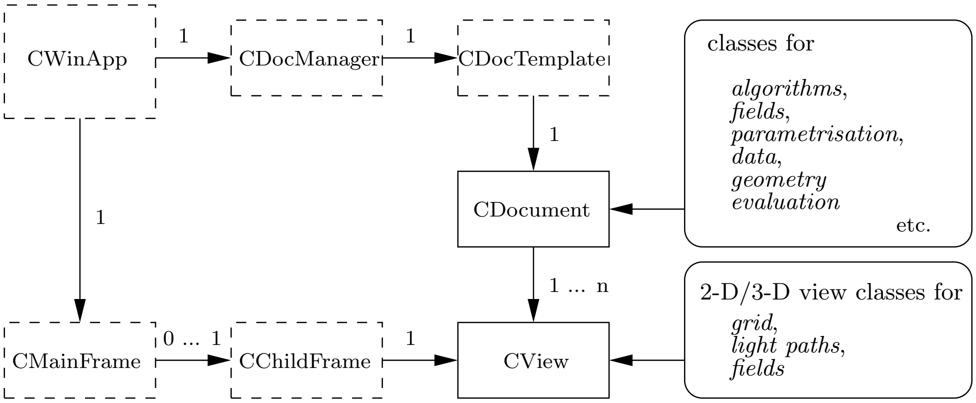


Figure D.1: The MFC Single Document/View architecture (simplified). Dashed boxes indicate code that remains largely untouched, round boxes contain the actual C++ classes for calculations and display.

Most of the calculations were carried out within a programme based on code originally developed by T. Laepple for reconstruction and display of experimental or simulated data [*Laepple et al.*, 2004]. The code is written in C++ and based on the *Microsoft Foundation Classes* (MFC), using the *Single Document/View* architecture. MFC is a Microsoft C++ library providing, e.g., classes and methods for windows as they are known from the operating software with the same name. The doc/view architecture is a special way to manage the storage of data (in the document), its display in one or multiple windows (views), the user interaction and the coordination and update of the views. Single doc/view means that there can be only one document in the application, but still multiple views of it (In fact, this architecture has been declared obsolete by now). The virtual algorithms represent only part of the complete code that is automatically created by Visual Studio's (the Microsoft programming interface) application wizard. Actual graphic display of the data uses the freeware *OpenGL* classes.

While the particular architecture allows convenient visualisation, the whole code seems heavily overloaded for purely scientific purposes and is sometimes hard to modify without some knowledge about MFC. Furthermore, it relies on Visual Studio, which is only commercially available.

Practically all original classes have been modified for the simulations of this thesis. The algorithm for the singular value decomposition (as well as vector and matrix classes) were adopted from the *TNT/JAMA* library.[1] Any other special routines have been taken from [*Press et al.*, 1992].

All colour plots in this thesis have been generated by a programme developed by B.C. Song which essentially does the same job as the earlier version by Laepple, but without using Microsoft's or any other commercial library. Contour plots have been created using the freeware *gnuplot*.

---

[1] Pozo, R., Mathematical and Computational Sciences Division, National Institute of Standards and Technology, http://math.nist.gov/tnt/

References

Andersen, A. H., and A. C. Kak, Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm, *Ultrason. Imag.*, *6*, 81 – 94, 1984.

Austen, J. R., S. J. Franke, C. H. Liu, and K. C. Yeh, Application of computerized tomography techniques to ionospheric research, *Proceedings of the URSI and COSPAR International Beacon Satellite Symposium on Radio Beacon Contribution to the Study of Ionization and Dynamics of the Ionosphere and to Corrections to Geodesy and Technical Workshop, part 1*, 23 – 35, 1986.

Backus, G. E., and J. F. Gilbert, Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. R. Soc. Lond.*, *266*, 123 – 192, 1970.

Barad, M. L. (Ed.), *Project Prairie Grass, A Field Program in Diffusion*, vol. I and II of *Geophysical Research Papers, No. 59*, Air Force Cambridge Research Center, 1958, Report AFCRC-TR-58-235.

Barratt, R., *Atmospheric dispersion modelling. An introduction to practical applications*, Business and Environment Practitioner Series, Earthscan Publications Ltd., London (UK), 2002.

Bäumer, D., B. Vogel, and F. Fiedler, A new parameterisation of motorway-induced turbulence and its application in a numerical model, *Atmos. Environ.*, *39*(31), 5750 – 5759, 2005.

Baussard, A., E. L. Miller, and D. Prémel, Adaptive B-spline scheme for solving an inverse scattering problem, *Inverse Problems*, *20*(2), 347 – 365, 2004.

Belotti, C., F. Cuccoli, L. Facheris, and O. Vaselli, An application of tomographic reconstruction of atmospheric CO_2 over a volcanic site based on open-path IR laser measurements, *IEEE Trans. Geosci. Remote Sensing*, *41*(11), 2629 – 2637, 2003.

Beniston, J. P., M. Wolf, M. Beniston-Rebetez, H. J. Kölsch, P. Rairoux, and L. Wöste, Use of Lidar measurements and numerical models in air pollution research, *J. Geophys. Res.*, *95*(D7), 9879 – 9894, 1990.

Berkowicz, R., O. Hertel, N. N. Sørensen, and J. A. Michelsen, in *Modeling Air Pollution from Traffic in Urban Areas. Flow and Dispersion Through Obstacles*, edited by R. J. Perkins and S. E. Belcher, pp. 121 – 142, 1997, Clarendon Press, Oxford, 1997.

Björck, Å., *Numerical methods for least squares problems*, SIAM, Philadelphia, 1996.

Blackadar, A. K., *Turbulence and diffusion in the atmosphere*, Springer, 1997.

Bland, V. V., J. P. Guarco, and T. V. Elderedge, Observations of NO_2 formation in two large natural gas fired boilers, *Proceedings of the 2000 International Joint Power Generation Conference*, 2000.

Bobrowski, N., G. Hönninger, B. Galle, and U. Platt, Detection of bromine monoxide in a volcanic plume, *Nature*, *423*, 273 – 276, 2003.

Brandt, J., J. H. Christensen, L. M. Frohn, F. Palmgren, R. Berkowicz, and Z. Zlatev, Operational air pollution forecasts from European to local scale, *Atmos. Environ.*, *35*(Suppl. 1), S91 – S98, 2001.

Briggs, G. A., Diffusion estimation for small emissions, *ARL Report ATDL–106*, NOAA, 1973.

Britter, R. E., and S. R. Hanna, Flow and dispersion in urban areas, *Ann. Rev. Fluid Mech.*, *35*, 469 – 496, 2003.

Byer, R. L., and L. A. Shepp, Two-dimensional remote air pollution monitoring via tomography, *Opt. Lett.*, *4*, 75 – 77, 1979.

Byrne, C., Likelihood maximization for list-mode emission tomographic image reconstruction, *IEEE Trans. Medical Imaging*, *20*(10), 103 – 120, 2001.

Byrne, C., A unified treatment of some iterative algorithms in signal processing and image reconstruction, *Inverse Problems*, *20*, 103 – 120, 2004.

Calvetti, D., G. Landi, L. Reichel, and F. Sgallari, Nonnegativity and iterative methods for ill-posed problems, *Inverse Problems*, *20*, 1747 – 1758, 2004.

Carlin, B. P., and T. A. Louis, *Bayes and empirical Bayes methods for data analysis*, Chapman & Hall, Roca Raton, 1998.

Censor, Y., and G. T. Herman, On some optimization techniques in image reconstruction from projections, *Appl. Numer. Math.*, *3*, 365 – 391, 1987.

Corsmeier, U., M. Kohler, B. Vogel, H. Vogel, and F. Fiedler, BAB II: A project to evaluate the accuracy of real-world traffic emissions for a motorway, *Atmos. Environ.*, *39*, 5627 – 5641, 2005*a*.

Corsmeier, U., D. Imhof, M. Kohler, J. Kühlwein, R. Kurtenbach, M. Petrea, E. Rosenbohm, B. Vogel, and U. Vogt, Comparison of measured and model-calculated real-world traffic emissions, *Atmos. Environ.*, *39*, 5760 – 5775, 2005*b*.

Dempster, A., N. Laird, and D. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.*, *39*, 1 – 38, 1977.

Doicu, A., F. Schreier, and M. Hess, An iterative regularization method with B-spline approximation for atmospheric temperature and concentration retrievals., *Environmental Modelling and Software*, *20*, 1101 – 1109, 2004.

Drescher, A. C., A. J. Gadgil, P. N. Price, and W. W. Nazaroff, Novel approach for tomographic reconstruction of gas concentration distributions in air: Use of smooth basis functions and simulated annealing, *Atmos. Environ.*, *30*(6), 929 – 940, 1996.

Drescher, A. C., D. Y. Park, A. J. Gadgil, S. P. Levine, and W. W. Nazaroff, Stationary and time-dependent tracer gas concentration profiles using open path FTIR remote sensing and SBFM computed tomography, *Atmos. Environ.*, *31*, 727 – 740, 1997.

Duclaux, O, E. Frejafon, H. Schmidt, A. Thomasson, D. Mondelain, J. Yu, C. Guillaumond, C. Puel, F. Savoie, P. Ritter, J. P. Boch, and J. P. Wolf, 3D-air quality model evaluation using the Lidar technique, *Atmos. Environ.*, *36*, 5081 – 5095, 2002.

Fehmers, G. C., *Tomography of the ionosphere*, Ph.D. dissertation, Technische Universiteit Eindhoven, The Netherlands, 1996.

Fehmers, G. C., An algorithm for quadratic optimization with one quadratic constraint and bounds on the variables, *Inverse Problems*, *14*, 893 – 901, 1998.

Fessler, J. A., Penalized weighted least-squares image reconstruction for positron emission tomography, *IEEE Trans. Medical Imaging*, *13*(2), 290 – 300, 1994.

Finlayson-Pitts, B., and J. N. Pitts, *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications*, Academic Press, New York, 2000.

Fischer, M. L., P. N. Price, T. L. Thatcher, C. A. Schwalbe, M. J. Craig, E. E. Wood, R. G. Sextro, and A. J. Gadgil, Rapid measurements and mapping of tracer gas concentrations in a large indoor space, *Atmos. Environ.*, *35*, 2837 – 2844, 2001.

Fisher, B., J. Kukkonen, M. Piringer, M. W. Rotach, and M. Schatzmann, Meteorology applied to urban air pollution problems: Concepts from COST 715, *Atmos. Chem. Phys.*, *6*, 555 – 564, 2006.

Fleming, H. E., Satellite remote sensing by the technique of computed tomography, *J. Appl. Meteorol.*, *21*, 1538 – 1549, 1982.

Frins, E., N. Bobrowski, U. Platt, and T. Wagner, Tomographic multiaxis-differential optical absorption spectroscopy observations of Sun-illuminated targets: a technique providing well-defined absorption paths in the boundary layer, *Appl. Opt.*, *45*, 6227 – 6240, 2006, accepted.

Geyer, A., B. Alicke, S. Konrad, J. Stutz, and U. Platt, Chemistry and oxidation capacity of the nitrate radical in the continental boundary layer near Berlin, *J. Geophys. Res.*, *106*(D8), 8013 – 8025, 2001.

Gifford, F. A., Use of routine meteorological observations for estimating atmospheric dispersion, *Nucl. Saf.*, *2*, 47 – 51, 1961.

Gilbert, P. F. C., Iterative methods for the three-dimensional reconstruction of an object from projections, *J. Theoret. Biol.*, *36*, 105 – 117, 1972.

Giuli, D., L. Facheris, and S. Tanelli, Microwave tomographic inversion technique based on stochastic approach for rainfall fields monitoring, *IEEE Trans. Geosci. Remote Sensing*, *37*(5), 2536 – 2555, 1999.

Golub, C. H., and C. F. van Loan, *Matrix computations*, John Hopkins University Press, Baltimore, 1996.

Gordon, R., R. Bender, and G. T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional microscopy and X-ray photography, *J. Theoret. Biol.*, *29*, 471 – 481, 1970.

Groetsch, C. W., *Inverse problems in the mathematical sciences*, Vieweg, Braunschweig, 1993.

Gull, S. F., and G. J. Daniell, Image reconstruction from incomplete and noisy data, *Nature*, *272*, 686 – 690, 1978.

Hadamard, J., Sur les problèmes aux dérivées partielles et leur signification physique, pp. 49–52, 1902, Princeton University Bulletin, 1902.

Hak, C., I. Pundt, S. Trick, C. Kern, U. Platt, J. Dommen, C. Ordóñez, A. S. H. Prévôt, W. Junkermann, C. Astorga-Lloréns, B. R. Larsen, J. Mellqvist, A. Strandberg, Y. Yu, B. Galle, J. Kleffmann, J. Lörzer, G. O. Braathen, and R. Volkamer, Intercomparison of four different in-situ techniques for ambient formaldehyde measurements in urban air, *Atmos. Chem. Phys.*, *5*, 2881 – 2900, 2005.

Hanna, S. R., Confidence limit for air quality models as estimated by bootstrap and jacknife resampling methods, *Atmos. Environ.*, *23*, 1385 – 1395, 1989.

Hanna, S. R., and E. M. Insley, Time series analyses of concentration and wind fluctuations, *Boundary-Layer Meteorol.*, *47*, 131 – 147, 1989.

Hanna, S. R., R. Britter, and P. Franzese, A baseline urban dispersion model evaluated with Salt Lake City and Los Angeles tracer data, *Atmos. Environ.*, *37*, 5069 – 5082, 2003.

Hansen, P. C., *Rank-deficient and discrete ill-posed problems*, Siam, Philadelphia, 1998.

Hartl, A., B.-C. Song, and I. Pundt, Reconstructing 2d concentration peaks from Long Path DOAS-tomographic measurements: Parametrisation and geometry within a discrete approach, *Atmos. Chem. Phys.*, *6*, 847 – 861, 2006.

Hashmonay, R. A., M. G. Yost, and C.-F. Wu, Computed tomography of air pollutants using radial scanning path-integrated optical remote sensing, *Atmos. Environ.*, *33*, 267 – 274, 1998.

Heckel, A., A. Richter, T. Tarsu, F. Wittrock, C. Hak, I. Pundt, W. Junkermann, and J. P. Burrows, MAX-DOAS measurements of formaldehyde in the Po-valley, *Atmos. Chem. Phys.*, *5*, 909 – 918, 2005.

Held, T., D. P. Y. Chang, and D. A. Niemeier, UCD 2001: an improved model to simulate pollutant dispersion from roadways, *Atmos. Environ.*, *37*, 5325 – 5336, 2003.

Herman, A., and S. W. Rowland, Three methods for reconstructing pictures from X-rays: A comparative study, *Comput. Graph. Image Proc.*, *1*, 151 – 178, 1973.

Herman, G. T., A. Lent, and S. W. Rowland, ART: Mathematics and applications. A report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques, *J. Theor. Biol.*, *42*, 1 – 32, 1973.

Heue, K.-P., *Airborne Multi Axis DOAS instrument and measurements of two dimensional tropospheric trace gas distributions*, Ph.D. dissertation, Institut für Umweltphysik, Universität Heidelberg, Germany, 2005.

Heue, K.-P., A. Richter, M. Bruns, J. P. Burrows, C. v. Friedeburg, U. Platt, I. Pundt, P. Wang, and T. Wagner, Validation of SCIAMACHY tropospheric NO_2-columns with AMAXDOAS measurements, *Atmos. Chem. Phys.*, *5*, 1039 – 1051, 2005.

Hönninger, G., C. v. Friedeburg, and U. Platt, Multi Axis Differential Absorption spectroscopy (max-doas), *Atmos. Chem. Phys.*, *4*, 231 – 254, 2004*a*.

Hönninger, G., H. Leser, O. Sebastián, and U. Platt, Ground-based measurements of halogen oxides at the Hudson Bay by Active Long Path Doas and Passive MAX-DOAS, *Geophys. Res. Lett.*, *31*(L04111), 2004*b*.

Hönninger, G., N. Bobrowski, E. R. Palenque, R. Torrez, and U. Platt, Reactive bromine and sulfur emissions at Salar de Uyuni, Bolivia, *J. Geophys. Res.*, *31*(L04101), 2004*c*.

Hurley, P., W. L. Physick, A. K. Luhar, and M. Edwards, The air pollution model (TAPM) version 3. part 2: Summary of some verification studies., *CSIRO Atmospheric Research Technical Papers No. 72*, CSIRO, 2005.

ISC-3, User's guide for Industrial Source Complex (ISC 3) Dispersion Models, vol. 2 (Description of model algorithms), *EPA–454/B–95–003a*, USEPA, 1995.

Jiang, M., and G. Wang, Convergence studies on iterative algorithms for image reconstruction, *IEEE Trans. Med. Imaging*, *22*(5), 569 – 579, 2003.

Kaczmarz, K., Angenäherte Auflösung von Systemen linearer Gleichungen, *Bull. Acad. Polon. Sci.*, *A35*, 355 – 357, 1937.

Kak, A. C., and M. Slaney, *Principles of computerized tomographic imaging*, Siam, Philadelphia, 2001.

Kesselmeier, J. K., and M. Staudt, Biogenic volatile organic compounds (VOC): An overview on emission, physiology and ecology, *J. Atmos. Chem.*, *33*(1), 23 – 88, 1999.

Kessler, Ch., W. Brücher, M. Memmesheimer, M. Kerschgens, and A. Ebel, Simulation of air pollution with nested models in North Rhine-Westphalia, *Atmos. Environ.*, *35*(Suppl. 1), S3 – S12, 2001.

Kohler, M., U. Corsmeier, U. Vogt, and B. Vogel, Estimation of gaseous real-world traffic emissions downstream a motorway, *Atmos. Environ.*, *39*, 5665 – 5684, 2005.

Kölsch, H. J., *Probing the Atmosphere: Air Pollution Studies by LIDAR*, Ph.D. dissertation, Freie Universität Berlin, Germany, 1990.

Kölsch, H. J., P. Rairoux, J. P. Wolf, and L. Wöste, Simultaneous NO and NO_2 DIAL measurements using BBO crystals, *Appl. Opt.*, *28*(11), 2052 – 2056, 1989.

Kunitsyn, V. E., E. S. Andreeva, A. Y. Popov, and O. G. Razinkov, Methods and algorithms of ray radiotomography for ionospheric research, *Annales Geophysicae*, *13*(12), 1263 – 1276, 1995.

Laepple, T., V. Knab, K. U. Mettendorf, and I. Pundt, Longpath DOAS tomography on a motorway exhaust plume: Numerical studies and application to data from the BAB II campaign, *Atmos. Chem. Phys.*, *4*, 1323 – 1342, 2004.

Leahy, R. M., and C. L. Byrne, Recent developments in iterative image reconstruction for PET and SPECT., *IEEE Trans. Med. Imaging*, *19*, 257 – 260, 2001.

Lenz, C.-J., F. Müller, and K. H. Schlünzen, The sensitivity of mesoscale chemistry transport model results to boundary values, *Env. Monitoring and Assessment*, *65*, 287 – 298, 2000.

Linz, P., Uncertainty in the solution of linear operator equations, *BIT*, *24*, 92 – 101, 1984.

Ma, Y., Z. Boybeyi, S. Hanna, and K. Chayantrakom, Plume dispersion analysis on surface concentration and its fluctuations from MVP field experiment, *Atmos. Environ.*, *39*, 3039 – 3054, 2005.

McElroy, J. L., and F. Pooler, The St. Louis dispersion study, *Report AP–53*, U.S. Public Health Service, National Air Pollution Control Administration, 1968.

Mettendorf, K. U., *Aufbau und Einsatz eines Multibeam Instrumentes zur DOAS-tomographischen Messung zweidimensionaler Konzentrationsverteilungen*, Ph.D. dissertation, Institut für Umweltphysik, Universität Heidelberg, Germany, 2005.

Mettendorf, K. U., A. Hartl, and I. Pundt, An indoor test campaign of the tomography long path differential absorption spectroscopy, *J. Environ. Monit.*, *8*, 279 – 287, 2006.

Minerbo, G., MENT: A maximum entropy algorithm for reconstructing a source from projection data, *Comput. Graphics Image Process.*, *10*, 48 – 68, 1979.

Moreira, D. M., J. C. Carvalho, and T. Tirabassi, Plume dispersion simulation in low wind conditions in the stable and convective boundary layers, *Atmos. Environ.*, *39*, 3643 – 3650, 2005.

Mueller, K., and R. Yagel, Rapid 3D cone-beam reconstruction with the Simultaneous Algebraic Reconstruction Technique (SART) using 2D texture mapping hardware, *IEEE Trans. Med. Imag.*, *19*, 279 – 287, 2000.

Munk, W., P. Worcester, and C. Wunsch, *Ocean acoustic tomography*, Cambridge Univ. Pr., Cambridge UK, 1995.

Natterer, F., *The mathematics of computerized tomography*, Siam, Philadelphia, 2001.

Neophytou, M. K., D. A. Goussis, E. Mastorakos, and R. E. Britter, The conceptual development of a simple scale-adaptive reacting pollutant dispersion model, *Atmos. Environ.*, *39*, 2787–2794, 2005.

Nolet, G. (Ed.), *Seismic Tomography with applications in global seismology and exploration geophysics*, D. Reidel Publishing Company, Dordrecht, 1987.

Olcese, L. E., and B. M. Toselli, Development of a model for reactive emissions from industrial stacks, *Environmental Modelling & Software*, *20*, 1239 –1250, 2005.

Pasquill, F., The estimation of the dispersion of windborne material, *Meterol. Mag.*, *90*, 33 –49, 1961.

Perner, D., D. H. Ehhalt, H. W. Pätz, U. Platt, E. P. Röth, and A. Volz, OH-radicals in the lower troposphere, *Geophys. Res. Lett.*, *3*, 466 – 468, 1976.

Peters, C., S. Pechtl, J. Stutz, K. Hebestreit, G. Hönninger, K. G. Heumann, A. Schwarz, J. Winterlik, and U Platt, Reactive and organic halogen species in three different European coastal environments, *Atmos. Chem. Phys.*, *5*, 3357 – 3375, 2005.

Petritoli, A., R. Ravegnani, G. Giovanelli, D. Bortoli, U. Bonaf, I. Kostadinov, and A. Oulanovsky, Off-axis measurements of atmospheric trace gases by use of an airborne ultraviolet-visible spectrometer, *Appl. Opt.*, *41*(27), 5593 –5599, 2002.

Philip, N. P., Pollutant tomography using integrated concentration data from non-intersecting optical paths, *Atmos. Environ.*, *33*, 275 –280, 1999.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, W. A. Lyons, L. D. Grasso, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland, A comprehensive meteorological modeling system – RAMS, *Meteor. Atmos. Phys.*, *49*, 69 –91, 1992.

Platt, U., Dry deposition of SO_2, *Atmos. Environ.*, *12*, 363 – 367, 1978.

Platt, U., Air monitoring by Differential Optical Absorption Spectroscopy (DOAS), in *Air Monitoring by Spectroscopic Techniques*, vol. 127 of *Chemical Analysis Series*, edited by M. W. Sigrist, 1994, John Wiley & Sons, Inc., New York, 1994.

Platt, U., D. Perner, G. W. Harris, A. M. Winer, and J. N. Pitts, Detection of NO_3 in the polluted troposphere by differential optical absorption, *Geophys. Res. Lett.*, *7*, 89 – 92, 1980.

Pöhler, D., Tomographic two-dimensional Long Path DOAS measurements of trace gas distributions above the city of Heidelberg, *Phd progress report*, University of Heidelberg, 2006.

Press, W. H., B. P. Flannery, and S. A. Teukolsky, *Numerical recipes in C: The art of scientific computing*, Cambridge Univ. Pr., Cambridge UK, 1992.

Price, P. N., M. L. Fischer, A. J. Gadgil, and R. G Sextro, An algorithm for real-time tomography of gas concentrations using prior information about spatial derivatives, *Atmos. Environ.*, *35*, 2827 –2835, 2001.

Pullen, J., J. P. Boris, T. Young, G. Patnaik, and J. Iselin, A comparison of contaminant plume statistics from a Gaussian puff and urban CFD model for two large cities, *Atmos. Environ.*, *39*, 1049 – 1068, 2005.

Pundt, I., and K. U. Mettendorf, Multibeam long-path differential optical absorption spectroscopy instrument: A device for simultaneous measurements along multiple light paths, *Appl. Opt.*, *44*(23), 4985 – 4994, 2005.

Pundt, I., K. U. Mettendorf, T. Laepple, V. Knab, P. Xie, J. Lösch, C. von Friedeburg, U. Platt, and T. Wagner, Measurements of trace gas distributions using Long-path DOAS-Tomography during the motorway campaign BAB II: experimental setup and results for NO_2, *Atmos. Environ.*, *39*(5), 967 – 975, 2005.

Reis, M. L., and N. C. Roberty, Maximum entropy algorithms for image reconstruction from projections, *Inverse Problems*, *8*, 623 – 644, 1992.

Riddle, A., D. Carruthers, A. Sharpe, C. McHugh, and J. Stocker, Comparisons between FLUENT and ADMS for atmospheric dispersion modeling, *Atmos. Environ.*, *38*, 1029 – 1038, 2004.

Rippel, B., Vorarbeiten für Langpfad DOAS Tomographische Messungen über der Stadt Heidelberg, Master's thesis, Institut für Umweltphysik, Universität Heidelberg, Germany, 2005.

Rodgers, C. D., *Inverse methods for atmospheric sounding: Theory and practice*, World Scientific, London, 2000.

Samanta, A., and L. A. Todd, Mapping chemicals in air using an environmental CAT scanning system: Evaluation of algorithms, *Atmos. Environ.*, *32*, 699 – 709, 1999.

Sauer, K., and C. Bouman, A local update strategy for iterative reconstruction from projections, *IEEE Trans. on Signal Processing*, *41*(2), 534 – 548, 1993.

Scales, J. A., and A. Gersztenkorn, Robust methods in inverse theory, *Inverse Problems*, *4*, 1071 – 1091, 1988.

Scales, J. A., and R. Snieder, To Bayes or not to Bayes, *Geophysics*, *62*, 1045 – 1046, 1997.

Schatzmann, M., and B. Leitl, Validation and application of obstacle-resolving urban dispersion models, *Atmos. Environ.*, *36*, 4811 – 4821, 2002.

Schatzmann, M., W. Bächlin, S. Emeis, J. Kühlwein, B. Leitl, W. J. Müller, K. Schäfer, and H. Schlünzen, Development and validation of tools for the implementation of European air quality policy in Germany (Project VALIUM), *Atmos. Chem. Phys.*, *6*, 3077 – 3083, 2006.

Schlünzen, K. H., K. Bigalke, C. Lüpkes, U. Niemeier, and K. v. Salzen, Concept and realization of the mesoscale transport- and fluid-model 'METRAS', *METRAS Techn. Rep. 5*, Meteorological Institute, University of Hamburg, 1996.

Schlünzen, K. H., D. Hinneburg, O. Knoth, M. Lambrecht, B. Leitl, S. Lopez, C. Lüpkes, H. Panskus, E. Renner, M. Schatzmann, T. Schoenemeyer, S. Trepte, and R. Wolke, Flow and transport in the obstacle layer – First results of the microscale model MITRAS, *J. Atmos. Chem.*, *44*, 113 – 130, 2003.

Schäfer, K., S. Emeis, H. Hoffmann, C. Jahn, W. J. Müller, B. Heits, D. Haase, W.-D. Drunkenmölle, W. Bächlin, H. Schlünzen, B. Leitl, F. Pascheke, and M. Schatzmann, Field measurements within a quarter of a city including a street canyon to produce a validation data set, *Int. J. Environ. Poll.*, *25*, 201 – 216, 2005.

Seinfeld, J. H., *Atmospheric chemistry and physics of air pollution*, Wiley and Sons, 1986.

Seinfeld, J. H., and S. N. Pandis, *Atmospheric chemistry and physics*, Wiley and Sons, 1998.

Shepp, L. S., and Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging*, *MI–1*, 113 – 122, 1982.

Shirley, T. R., W. H. Brune, X. Ren, J. Mao, R. Lesher, B. Cardenas, R. Volkamer, L. T. Molina, M. J. Molina, B. Lamb, E. Velasco, T. Jobson, and M. Alexander, Atmospheric oxidation in the Mexico City Metropolitan Area (MCMA) during April 2003, *Atmos. Chem. Phys.*, *6*, 2753 – 2765, 2006.

Song, C.H., G. Chen, S.R. Hanna, J. Crawford, and J.J. Davis, Dispersion and chemical evolution of ship plumes in the marine boundary layer: Investigation of $O_3/NO_y/HO_x$ chemistry, *J. Geophys. Res.*, *108*(D4), 4143, 2003.

Soulhac, L., C. Puel, O. Duclaux, and R. J. Perkins, Simulations of atmospheric pollution in Greater Lyon an example of the use of nested models, *Atmos. Environ.*, *37*, 5147 – 5156, 2003.

Stern, R., and R.J. Yamartino, Development and first evaluation of micro-calgrid: A 3-D urban-canopy-scale photochemical model, *Atmos. Environ.*, *35*(Suppl. 1), S149 – S165, 2001.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, *95*(D10), 16343 – 16367, 1990.

Stull, R. B., *An Introduction to Boundary Layer Meteorology*, Kluwer Acad. Publ., Dordrecht, Boston, London, 1988.

Stutz, J., and U. Platt, Numerical analysis and error estimation of differential optical absorption spectroscopy measurements with least squares methods, *Appl. Opt.*, *35*, 6041 – 6053, 1996.

Stutz, J., and U. Platt, A new generation of DOAS instruments, in *Instrument Development for Atmospheric Research and Monitoring*, edited by J. Bösenberg *et al.*, pp. 370 – 378, 1997, Springer, Berlin, Heidelberg, 1997.

Stutz, J., B. Alicke, R. Ackermann, A. Geyer, A. White, and E. Williams, Vertical profiles of NO_3, N_2O_5, O_3, and NO_x in the nocturnal boundary layer: 1. observations during the Texas Air Quality Study 2000, *J. Geophys. Res.*, *109*(D12306), 1 – 14, 2004.

Sykes, R. I., and R. S. Gabruk, A second-order closure model for the effect of averaging time on turbulent plume dispersion, *J. Appl. Met.*, *36*, 165 – 184, 1997.

Tanabe, K., Projection method for solving a singular system of linear equations and its applications, *Numer. Math.*, *17*, 203 – 214, 1971.

Tarantola, A., *Inverse Problem Theory*, Elsevier, New York, 1987.

Tarantola, A., *Inverse problem theory and model parameter estimation*, SIAM, Philadelphia, 2005.

Tarantola, A., and B. Valette, Inverse problems: Quest for information, *J. Geophys.*, *50*, 159 – 170, 1982.

Tikhonov, A. N., Solution of incorrectly formulated problems and the regularization method, *Soviet Math. Dokl.*, *4*, 1035 – 1038, 1963.

Todd, L., and R. Bhattacharyya, Tomographic reconstruction of air pollutants: Evaluation of measurement geometries, *Appl. Opt.*, *36*(30), 7678 – 7688, 1997.

Todd, L., and D. Leith, Remote sensing and computed tomography in industrial hygiene, *Am. Ind. Hyg. Assoc. J.*, *51*(4), 224 – 233, 1990.

Todd, L., and G. Ramachandran, Evaluation of algorithms for tomographic reconstruction of chemical concentrations in indoor air, *Am. Ind. Hyg. Assoc. J.*, *55*(5), 403 – 417, 1994*a*.

Todd, L., and G. Ramachandran, Evaluation of optical source-detector configurations for tomographic reconstruction of chemical concentrations in indoor air, *Am. Ind. Hyg. Assoc. J.*, *55*(12), 1133 – 1143, 1994*b*.

Trampert, J., and J. J. Lévêque, Simultaneous reconstruction technique: Physical interpretation based on the generalized least squares solution, *J. Geophys. Res.*, *95*(B8), 12553 – 12559, 1990.

Trukenmüller, A., D. Grawe, and K. H. Schlünzen, A model system for the assessment of ambient air quality conforming to EC directives, *Meteorol. Zeitschrift*, *13*(5), 387 – 3949, 2004.

Tsai, M. Y., and K. S. Chen, Measurements and three-dimensional modeling of air pollutant dispersion in an urban street canyon, *Atmos. Environ.*, *38*, 5911 – 5924, 2004.

Twomey, S., On the numerical solution of Fredholm equations of the first kind by the inversion of the linear system produced by quadrature, *J. Assoc. Comput. Mach.*, *10*, 97 – 101, 1963.

Twomey, S., *Introduction to the mathematics of inversion in remote sensing*, Dover Publication Inc., New York, 1997.

Van der Sluis, A., and H. A. van der Vorst, Numerical solution of large sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography*, edited by G. Nolet, chapter 3, pp. 49–83, Reidel, Hingham, Mass., 1987.

Vardi, Y., L. A. Shepp, and L. Kaufman, A statistical model for positron emission tomography, *J. Amer. Statist. Assoc.*, *80*, 8 – 37, 1985.

Vardoulakis, S., N. Gonzalez-Flesca, B. E. A. Fisher, and K. Pericleous, Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris, *Atmos. Environ.*, *39*, 2725 – 2736, 2005.

Veitel, H., B. Kromer, M. Mößner, and U. Platt, New techniques for measurements of atmospheric vertical trace gas profiles using DOAS, *Environmental Science and Pollution Research*, *special issue 4*, 17 – 26, 2002.

Venkatram, A., The expected deviation of observed concentrations from predicted ensemble means, *Atmos. Environ.*, *13*, 1547 – 1549, 1979.

Venkatram, A., The uncertainty in estimating dispersion in the convective boundary layer, *Atmos. Environ.*, *13*, 307 – 310, 1984.

Venkatram, A., Accounting for averaging times in air pollution modeling, *Atmos. Environ.*, *36*, 2165 – 2170, 2002.

Venkatram, A., V. Isakov, J. Yuan, and D. Pankratz, Modeling dispersion at distances of meters from urban sources, *Atmos. Environ.*, *38*, 4633 – 4641, 2004.

Venkatram, A., V. Isakov, and J. Pankratz D.and Yuan, Relating plume spread to meteorology in urban areas, *Atmos. Environ.*, *39*, 371 – 380, 2005.

Verkruysse, W., and L. A. Todd, Improved method "grid translation" for mapping environmental pollutants using a two-dimensional CAT scanning system, *Atmos. Environ.*, *38*, 1801 – 1809, 2004.

Volkamer, R., *A DOAS study on the oxidation mechanism of aromatic hydrocarbons under simulated atmospheric conditions*, Ph.D. dissertation, Institut für Umweltphysik, Universität Heidelberg, Germany, 2001.

Volz-Thomas, A., H. Geiss, A. Hofzumahaus, and K. H. Becker, Introduction to special section: Photochemistry experiment in BERLIOZ, *J. Geophys. Res.*, *108*(D4), 8252, 2003.

Von Glasow, R., M. G. Lawrence, R. Sander, and P. J. Crutzen, Modeling the chemical effects of ship exhaust in the cloud-free marine boundary layer, *Atmos. Chem. Phys.*, *3*, 233 – 250, 2003.

Von Glasow, R., R. Von Kuhlmann, M. G. Lawrence, U. Platt, and P. J. Crutzen, Impact of reactive bromine chemistry in the troposphere, *Atmos. Chem. Phys.*, *4*, 2481 – 2497, 2004.

Wagner, T., B. Dix, C. v. Friedeburg, U. Frieß, S. Sanghavi, R. Sinreich, and U. Platt, MAX-DOAS O_4 measurements: A new technique to derive information on atmospheric aerosols – Principles and information content, *J. Geophys. Res.*, *109*(D2205), 8252, 2004.

Wagner, T., J. Burrows, T. Deutschmann, B. Dix, F. Hendrick, C. von Friedeburg, U. Frieß, K.-P. Heue, H. Irie, H. Iwabuchi, J. Keller, C. McLinden, H. Oetjen, E. Palazzi, A. Petritoli, U. Platt, P. Postylyakov, J. Pukite, A. Richter, M. van Roozendael, A. Rozanov, V. Rozanov, R. Sinreich, S. Sanghavi, and F. Wittrock, Comparison of box-air-mass-factors and radiances for MAX-DOAS-geometries calculated from different UV/visible radiative transfer models, *Atmos. Chem. Phys. Discuss.*, *6*, 9823 – 9876, 2006.

Wayne, R. P., *Chemistry of Atmospheres: An Introduction to the Chemistry of the Atmospheres of Earth, the Planets, and their Satellites*, Oxford University Press, 3rd edition, 2000.

Weinbrecht, S., S. Raasch, A. Ziemann, K. Arnold, and A. Raabe, Comparison of large-eddy simulation data with spatially averaged measurements obtained by acoustic tomography – presuppositions and first results, *Bound. Layer Met.*, *111*, 441 – 465, 2004.

Wenig, M., B. Jähne, and U. Platt, Operator representation as a new differential absorption spectroscopy formalism, *Appl. Opt.*, *44*(16), 3246 – 3253, 2005.

Willmott, C. J., On the validation of models, *Phys. Geog.*, *2*, 184 – 194, 1981.

Wolfe, D. C., and R. L. Byer, Model studies of laser absorption computed tomography for remote air pollution measurement, *Appl. Opt.*, *21*(7), 1165 – 1177, 1982.

Yamartino, R.J., J.S. Scire, G.R. Carmichael, and Y.S. Chang, The CALGRID mesoscale photochemical grid model – I: Model formulation, *Atmos. Environ.*, *26A*, 1493 – 1512, 1992.

Yee, E., and R. Chan, A simple model for the probability density function of concentration fluctuations in atmospheric plumes, *Atmos. Environ.*, *31*(7), 991 – 1002, 1996.

Bedanken möchte ich mich herzlich bei Herrn Prof. Platt für die abschließende Betreuung dieser Arbeit, Herrn Prof. Jähne für die zweite Begutachtung und insbesondere bei Claudia Hak, Klaus-Peter Heue und Denis Pöhler für die sorgfältige Korrektur ihrer ersten Versionen.

Mein besonderer Dank geht auch an David Grawe für die freundliche Bereitstellung seiner Modelldaten, alle Kollegen des IUP, von denen ich in fachlichen Diskussionen lernen konnte, sowie an alle, die für Spaß und Kurzweil gesorgt habe und nicht zuletzt an Frau Clos für ihre stete, unerschütterliche Hilfsbereitschaft.

Above all I would like to thank Bing-Chao Song who not only taught me things about computers I did not dream of, gave me insight into a Chinese way of thinking (both with incredible patience) and made my time at the IUP a lot more fun, but who is also one of the finest persons I have ever met.