Working Paper

Towards a simplified model to describe ozone formation in Europe

Chris Heyes and Wolfgang Schöpp

WP-95-34
April 1995

Towards a simplified model to describe ozone formation in Europe

Chris Heyes and Wolfgang Schöpp

WP-95-34

April 1995

Preface

Air pollution is a multi-faceted problem with a variety of pollutants released from a large number of different anthropogenic activities causing a multitude of environmental effects. Cost-effective strategies to reduce negative impacts of air pollution must take account of these complexities and consider the individual aspects of the air pollution problem simultaneously.

Integrated assessment models provide a consistent framework for a systematic analysis of alternative strategies. One of such models, the RAINS (Regional Acidification Information and Simulation) model developed at IIASA, has been used as a scientific support tool for the international discussions on further reductions of emissions of sulfur dioxide in Europe, which led in 1994 to an agreement on the 'Second Sulfur Protocol'.

The perceived success of this agreement motivated further work on extending model analysis to additional pollutants and effects. Elevated levels of tropospheric ozone are currently considered as one of the major air quality problems in Europe, calling for balanced reductions of emissions of nitrogen oxides and volatile organic compounds. A major obstacle for developing a practical integrated assessment model for ozone is the complexity and the size of most of the current models on atmospheric ozone formation. This paper tries to identify elements for potential model simplifications, which could contribute to the development of an operational European-scale ozone formation model. Such a simplified model would establish the core of an integrated assessment model for ozone formation, linking information on emissions and emission control costs with an assessment of their environmental impacts.

Acknowledgements

Contents   Page

1. INTRODUCTION

Concern over the environmental impact of ozone and other photochemically-derived oxidants in the troposphere goes back many years, with the first signs of problems being experienced in the Los Angeles basin. Investigations of the Los Angeles smogs during the 1950s showed that the formation of ozone involved chemical reactions between air pollutants, specifically nitrogen oxides (NO_x) and volatile organic compounds (VOCs), driven by solar radiation. Subsequent work has revealed that ozone formation occurs on a regional scale in many parts of the world. Within Europe interest has intensified in recent years, with increasing experimental evidence that ozone can have adverse effects on crops, trees, materials and human health.

The timescale of ozone production is such that ozone concentrations build up in polluted air over several days under suitable weather conditions, and this pollutant and its precursors can be transported considerable distances and across national boundaries. There are regional variations in the density of precursor emissions, in the potential for and the costs of emission control, and in the sensitivities of ecosystems to photochemical pollutants. Consequently, measures taken within Europe to control ozone require international cooperation if they are to be successful and cost-effective.

Within Europe, international agreements to protect the environment from the effects of air pollution have largely been negotiated under the auspices of the United Nations Economic Commission for Europe (UNECE), in particular within the Convention on Long-Range Transboundary Air Pollution (LRTAP). Several Protocols to this Convention are now in force. Initially, attention was focussed on the problem of acidification, leading to a Protocol to reduce sulphur emissions by at least 30% which was adopted in Helsinki in 1985. More recently, in 1994, a second sulphur Protocol aimed at further reducing sulphur emissions was adopted in Oslo. Nitrogen oxides also make a substantial contribution to the acidity of precipitation, and recognition of this led to a Protocol concerning the control of NO_x emissions, adopted in Sofia in 1988. The adverse effects of ground-level ozone have also been recognised within the LRTAP Convention, and in 1991, in Geneva, a Protocol concerning the control of VOC emissions was adopted. For the future, the highest priority is being given to the development of a strategy for the second step of the Protocol concerning the Control of Emissions of Nitrogen Oxides. In pursuing this, it is envisaged that, as far as possible, a multi-pollutant, multi-effect approach will be followed. Within this framework it is clear that the question of ground-level ozone will be a major consideration.

The integration of all the factors relevant to the efficient control of ozone is a rather complex task. An integrated assessment model which combines information on emission generation, emission control technologies and abatement costs, the long-range transport of pollutants and the environmental effects in different areas of Europe would be valuable in assessing alternative control strategies. Such models, including IIASA's Regional Acidification Information and Simulation (RAINS) model (Alcamo et al., 1991), have been developed to analyse strategies to reduce acidification. Recently, this approach has been used to derive guidelines for international reductions of SO_2 emissions in Europe.

Initial feasibility studies on the basic design of an integrated assessment model for ozone have been carried out, specifically in regard to the necessary VOC emission inventories (Olsthoorn, 1994) and the availability and costs of technologies to reduce VOC emissions (Caliandro, 1994). Central to the development of an integrated ozone model is the

requirement to provide a simplified description of ozone formation. This description may be divided into two parts. The first of these concerns the source-receptor relationships between precursor emissions and the concentrations of ozone in the atmospheric boundary layer, and forms the subject of this paper. The second part relates the boundary layer concentrations to ground-level ozone exposure, taking local factors into account. Analyses of European ozone measurement data, aimed at identifying the important features of this relationship, are reported separately (Kettunen et al., 1994; Baldi and Calori, 1994).

For application in an integrated assessment model for ozone, the source-receptor relationships need to be valid over a range of different spatial patterns of emission sources, and not restricted to the present-day situation alone. For this reason, attempts to define these relationships solely on the basis of recent ozone measurement data are likely to prove inadequate. Instead, the simplified ozone formation description needs to be based on a more complex mathematical model that has gained widespread international acceptance. To this end, the work is being carried out in collaboration with EMEP's Meteorological Synthesizing Centre - West, and the results of the EMEP ozone model provide the basis on which a simplified ozone formation module will be built.

It is possible to envisage several ways of condensing the results of the EMEP model in order to construct a means of representing source-receptor relationships, and investigations in this area have proceeded on parallel paths. This paper describes work carried out on one of the possible approaches, in which statistical techniques are used to investigate and describe the important relationships between ozone, its precursors and meteorological variables.

The second chapter of this paper provides background information on the factors affecting ozone concentrations in the troposphere, followed by a brief description of the EMEP ozone model, which is fundamental to the work described here, and some results obtained from it. An outline of the integrated assessment model concept is also included in order to place the present study in context. Progress with the development of a statistical model able to summarise the results of the EMEP model is presented in Chapter 3.

2.   TROPOSPHERIC OZONE

There are no emissions of ozone directly into the atmosphere; all the ozone found there has been formed by chemical reactions in the air. Most of the ozone is produced and resides in the stratosphere at altitudes between 12 and 40 km. Some stratospheric ozone is transported downwards through the troposphere into the atmospheric boundary layer and is finally removed at the earth's surface. This naturally occurring flux is supplemented by in-situ photochemical production of ozone in the troposphere, the most significant in-situ production occurring in the atmospheric boundary layer close to the surface. In this region, it is the reactions of primary pollutant gases, specifically nitrogen oxides and organic compounds, driven by the action of sunlight, which are responsible for enhanced ozone formation. Ozone generated in the lower atmosphere in this way is termed a secondary pollutant.

The following section provides an outline of the atmospheric chemistry and other factors affecting the formation of ozone in the atmospheric boundary layer.

2.1 Ozone Formation

The chemical production of ozone (O_3) in the atmosphere requires the recombination of atomic oxygen ($O(^3P)$, a ground state oxygen atom) and molecular oxygen (O_2). In the troposphere, photodissociation of nitrogen dioxide (NO_2) is the only significant source of oxygen atoms:

$$NO_2 + h\upsilon \rightarrow NO + O(^3P) \qquad (1)$$

Clearly, this process depends on sufficient solar energy (represented by $h\upsilon$, the product of Planck's constant, h, and the frequency, υ, of the solar radiation). The recombination of atomic and molecular oxygen:

$$O(^3P) + O_2 + M \rightarrow O_3 + M \qquad (2)$$

(where M can be any inert molecule such as nitrogen or oxygen) to produce ozone may be followed by the rapid reaction of ozone with nitric oxide (NO):

$$NO + O_3 \rightarrow NO_2 + O_2 \qquad (3)$$

to regenerate NO_2. In a relatively unpolluted atmosphere, a photochemical steady state is achieved during hours with sunlight, and, in the absence of other processes that convert NO to NO_2, the ozone concentration is linked to the NO_2 / NO concentration ratio:

$$[O_3] = J_1 [NO_2] / k_3 [NO] \qquad (4)$$

where J_1 is the NO_2 photolysis rate and k_3 is the rate constant for reaction 3.

Any processes, other than reaction with ozone, which convert NO to NO_2 will facilitate net production of ozone during daylight. Such processes are possible in polluted atmospheres containing volatile organic compounds (VOCs), originating primarily from pollutant emissions but also from natural sources like vegetation. The chemistry of the polluted troposphere is very complicated because of the presence of numerous VOCs of various classes (e.g. alkanes, alkenes, aldehydes, ketones and aromatic hydrocarbons) and the added complexities in the chemistry of these organic species. A full description of their role in atmospheric chemistry is clearly beyond the scope of this paper; the following outline of the important atmospheric reactions of the alkanes is intended merely to illustrate the salient features with regard to ozone production. More complete discussions may be found in several review articles (e.g. Atkinson, 1990; National Research Council, 1991).

The initial step in the important chain of reactions involving VOCs is attack by hydroxyl radicals ($OH^\bullet$). In a sunlit, polluted atmosphere the two main sources of this species are (i) the ultraviolet dissociation of ozone followed by rapid reaction with water vapour,

$$O_3 + h\upsilon \rightarrow O(^1D) + O_2 \qquad (5)$$

$$O(^1D) + H_2O \rightarrow OH^\bullet + OH^\bullet \qquad (6)$$

(where $O(^1D)$ is an electronically excited oxygen atom and H_2O is water) and (ii) the photochemical dissociation and subsequent reactions of aldehydes and ketones.

3

In the illustrative case of alkanes (represented by RH), reaction with the OH radical yields an alkyl radical (R$^\bullet$):

$$RH + OH^\bullet \rightarrow R^\bullet + H_2O \tag{7}$$

which is rapidly converted to an organic peroxy radical (RO$_2^\bullet$):

$$R^\bullet + O_2 \rightarrow RO_2^\bullet \tag{8}$$

The peroxy radicals are able to oxidise NO to NO$_2$:

$$RO_2^\bullet + NO \rightarrow RO^\bullet + NO_2 \tag{9}$$

and thus, in conjunction with reactions 1 and 2, lead to increased ozone production.

The alkoxy radical (RO$^\bullet$) produced in reaction 9 may react with oxygen to form an aldehyde or ketone, depending on the nature of R, and a hydroperoxy radical (HO$_2^\bullet$) :

$$RO^\bullet + O_2 \rightarrow carbonyl + HO_2^\bullet \tag{10}$$
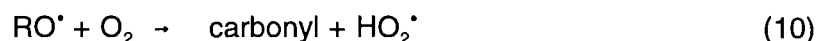
although they may also decompose or isomerise to other alkyl radicals and carbonyl compounds which can then react further. The hydroperoxy radical is another important trace component of the atmosphere which can also oxidise NO to NO$_2$ with the regeneration of a hydroxyl radical:

$$HO_2^\bullet + NO \rightarrow NO_2 + OH^\bullet \tag{11}$$

with the result that further oxidation of VOCs, and further ozone formation, can occur. Reaction of the hydroperoxy radical with ozone:

$$HO_2^\bullet + O_3 \rightarrow 2O_2 + OH^\bullet \tag{12}$$

is another mechanism by which the chain-propagating hydroxyl radical can be regenerated.

The main process which removes free radicals, and therefore opposes the ozone-producing reactions, is the combination of hydroxyl radicals with nitrogen dioxide to form gaseous nitric acid (HNO$_3$) :

$$OH^\bullet + NO_2 + M \rightarrow HNO_3 + M \tag{13}$$

Peroxy radicals also react with nitrogen dioxide to form peroxynitrates:

$$RO_2^\bullet + NO_2 + M \rightarrow RO_2NO_2 + M \tag{14}$$
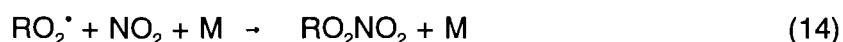
Many of the peroxynitrates are unstable and so do not effectively remove radicals. However, the more stable ones do act as radical sinks and can build up as significant secondary pollutants. One important example of such a compound is peroxyacetyl nitrate (PAN), formed in reaction 14 when the R fraction represents the acetyl group.

4

Reaction between peroxy radicals can also act as a radical loss process. One example is the self reaction of hydroperoxy radicals:

$$HO_2^{\bullet} + HO_2^{\bullet} \rightarrow H_2O_2 + O_2 \qquad (15)$$

which is the major route for the formation of atmospheric hydrogen peroxide (H_2O_2).

The preceding description of the photochemical production of ozone is applicable to the polluted atmospheric boundary layer. It is worth noting that some ozone formation can also take place in the absence of anthropogenic emissions of VOCs. In the clean, unpolluted troposphere, the atmospheric chemistry is largely determined by the reactions of methane (CH_4) and its degradation products, formaldehyde (HCHO) and carbon monoxide (CO). The hydroxyl radical reaction is again the important initial step (see, for example, National Research Council, 1991) but its rate of reaction with methane is much slower than with non-methane VOCs and, consequently, methane has a much longer tropospheric lifetime than the VOC species. In the lower troposphere, especially in polluted areas, the chemical reactions of the non-methane VOCs dominate those of methane and carbon monoxide.

A major sink for tropospheric ozone is removal by deposition to terrestrial surfaces, termed dry deposition. This process is regulated by stomatal uptake in vegetation canopies which depends on light, temperature and humidity. As a result, the dry deposition of ozone shows strong diurnal and seasonal cycles with daytime and summertime maxima.

The concentration of ozone in the atmosphere will clearly be influenced by the rates at which the chemical mechanisms proceed. Additionally, both large- and local-scale mixing processes occurring in the troposphere will affect ozone concentrations measured at ground level. During the day, any pollutants within the atmospheric boundary layer, having a depth of approximately 1 km or so, are generally well mixed. Following sunset and subsequent cooling at the surface, the depth of the boundary layer may contract to perhaps a few tens of metres above the surface. During night time, in the absence of significant ozone formation, the ozone concentration within the shallower boundary layer decreases due to deposition to the surface and, in areas subject to emissions of nitric oxide, through reaction with NO (reaction 3). At sunrise, convective processes break down the nocturnal stratification; ozone trapped aloft overnight is mixed down to the surface and its ground level concentration rises. As a consequence of these processes, a diurnal variation in ozone concentration is observed at many locations. Topographical features may influence the strength of this variation. Sites on elevated ground are exposed more frequently to free tropospheric air and may be less frequently capped by a nocturnal inversion layer. Similarly, land/sea breezes help to sustain mixing processes at coastal locations. In both these cases there may be a dampening of the diurnal cycle.

It should be emphasised that the preceding description of tropospheric ozone production is necessarily simplified and incomplete; for example, important night time reactions have been entirely omitted from the discussion. However, even this simple outline serves to highlight the most important factors involved.

Clearly, the enhanced production of ozone depends on emissions into the atmosphere of nitrogen oxides ($NO_x = NO + NO_2$) and volatile organic compounds. Emissions of NO_x arise mainly from the combustion of fossil fuels; the most important sources are motor vehicles and power plants. Major anthropogenic sources of VOCs in the atmosphere include emissions resulting from the use of solvents, motor vehicles (both exhaust and evaporative emissions)

and the chemical and petroleum industries. In addition, there is increasing recognition that biogenic, or natural, emissions of VOCs, mainly from trees and other vegetation, can also be important in some circumstances.

It is also evident from the preceding discussion that the ratio between NO_x and VOC concentrations in the atmosphere is an important factor, not merely the presence of these compounds. When the reactions of peroxy radicals with NO (reactions 9 and 11) proceed more rapidly than reactions between peroxy radicals (e.g. reaction 15) or the reaction of the hydroperoxy radical with ozone (reaction 12), the result is net ozone formation. The balance between these opposing tendencies is determined by the NO_x / VOC ratio. In very remote locations, mainly in parts of the southern hemisphere, the NO_x concentrations may be so low that the reactions of HO_2 with ozone and other peroxy radicals dominate, leading to net ozone removal. The NO_x concentrations in the atmospheric boundary layer over continental areas in the northern hemisphere, however, are generally high enough that the reactions of peroxy radicals with NO dominate, leading to ozone production.

There is a further complication when NO_x concentrations are particularly high. In this case, the hydroxyl radicals, which propagate VOC oxidation and conversion of NO to NO_2, are scavenged by the relatively high concentrations of NO_2 (reaction 13). The NO_2 effectively competes with the VOCs for the OH radical, reducing the production of RO_2 and HO_2 radicals, and inhibiting the formation of ozone. This feature of tropospheric chemistry has important implications for the development of ozone control strategies. If, in these circumstances, the NO_2 concentration is decreased, there will be a greater number of OH radicals available to react with the VOCs, leading to greater formation of ozone. Hence, a reduction in the atmospheric NO_x level may result in an increase in ozone concentration.

While the emissions of ozone precursors are obviously essential for ozone formation, meteorological factors also exert a crucial influence. Ozone production in an air mass containing NO_x and VOC compounds will occur only if the weather conditions are favourable. Firstly, it will be clear from the preceding description of ozone formation chemistry that the intensity of the incoming solar radiation must be high since this is the driving force behind the photochemistry of ozone production. Temperature is also important since it influences the rate of many of the chemical reactions involved. Other conditions favourable to ozone formation are low windspeeds and a persistent, well-defined boundary layer. These conditions allow sufficient build-up of precursor concentrations, contained within a limited volume, for long enough to generate ozone and other secondary pollutants before the polluted air is dispersed into the free troposphere above the boundary layer.

2.2 The EMEP Ozone Model

The formation of tropospheric ozone involves the complex interaction between precursor emissions, photochemistry, transport and deposition. The complete representation of these processes in a mathematical model is difficult, and the many models developed to study the ozone formation processes have been greatly simplified in one or more of these areas. Both Eulerian, or grid, and Lagrangian, or trajectory, models have been constructed, using different spatial and temporal scales, and employing chemical mechanisms of varying complexity, as deemed appropriate for each model's main purposes. A simplified description of ozone formation suitable for incorporation into an integrated assessment model needs to be based on a more complex model designed to calculate ozone concentrations over long periods and covering the whole of Europe. This requirement is satisfied by the EMEP MSC-W photo-oxidant model (Simpson, 1992a, 1993).

The EMEP ozone model is a single-layer Lagrangian trajectory model which calculates concentrations of photochemical oxidants every six hours for a set of up to 740 arrival points (on a 150 km x 150 km grid) covering the whole of Europe. Columns of air in the atmospheric boundary layer are followed along specified 96-hour trajectories, picking up emissions of NO_x, VOC, CO and SO_2 from the underlying grid. The height of the air column, the mixing height, containing the bulk of the polluted air is reset at 1200 GMT each day from radiosonde data. Along each trajectory the mass conservation equations are integrated, taking into account emission inputs, photolysis and chemical reactions, dry and wet removal, and the influence of meteorological parameters. These equations are solved numerically, currently using the quasi-steady state approximation method with a fixed time step of 15 minutes.

The six-hourly meteorological data required by the EMEP model are taken from the output of the Norwegian Numerical Weather Prediction model. Wind velocity data permit calculation of 96-hour back-trajectories to any point in the EMEP grid. The ozone model simulates the exchange of boundary layer air with free tropospheric air as a result of convective clouds. Photolysis rates are adjusted for cloud cover, and temperature data are used to calculate appropriate chemical reaction rates and in estimating both natural VOC emissions and the emissions of NO_x from soils. Other meteorological data are used in estimating deposition velocities, which are calculated as a function of atmospheric stability, latitude, time of year and time of day.

The anthropogenic emissions of NO_x, VOC and SO_2 used in the model are based, as far as possible, on data supplied officially at a national level. The emission inventory includes estimates of SO_2 and NO_x emissions from ship traffic in the Atlantic Ocean, North Sea and Baltic Sea. Emissions of CO are estimated by scaling the anthropogenic VOC emissions by a factor of 3.7. For use in the ozone model, the emissions are spatially disaggregated on to the 150 km x 150 km EMEP grid. The SO_2 emissions and emissions of NO_x from high-level sources are assumed to vary sinusoidally over the year with a maximum in January and a minimum in July. By contrast, anthropogenic VOC emissions are kept constant from one month to the next. Both low-level NO_x emissions and anthropogenic VOC emissions are assumed to follow a simple diurnal variation with daytime emissions three times greater than those during night.

Natural emissions of VOCs are represented in the EMEP ozone model by the emissions of isoprene from forests and agricultural crops (Simpson, 1994a). These are calculated at each time step using surface temperature data, land-use data for each grid square and published emission-temperature relationships. Estimates of NO_x emissions from soils are based on temperature and a simple categorisation of type of land cover.

The chemical mechanism formerly employed in the EMEP model has been shown (Derwent, 1993) to compare well in polluted situations with a range of 24 other mechanisms. This scheme has recently been updated (Simpson et al., 1993) and now comprises 136 reactions, including 25 photolysis reactions. The model calculates explicitly the concentrations of 61 chemical species. Six organic species - ethane, n-butane, ethene, propene, o-xylene and ethanol - are used to represent the emitted anthropogenic VOCs. Their proportions have been chosen to reflect the mix of species and chemical reactivities thought to be present in actual VOC emissions. In the version of the model used for this study, the same VOC speciation is used to represent man-made emissions from anywhere in Europe. Future versions of the EMEP model will allow a larger number of emitted VOC species and the use of different VOC species profiles for different source sectors in different countries. Methane is treated in the model as a tropospheric background species. Initial concentrations at the

7

start of each 96-hour trajectory are taken from previous model calculations, if available; otherwise a set of assumed background tropospheric concentrations, based on appropriate measured data or tropospheric model calculations, is used.

The EMEP ozone model has been applied to the summer periods of 1985, 1989 and 1990 (Simpson, 1992a, 1993; Labancz, 1993) and has been used to assess the effects of various scenarios of precursor emission reductions (Simpson, 1992b, 1993). Some of the uncertainties associated with the modelling process have been addressed by comparison with other photochemical oxidant models (Builtjes et al., 1991). The role played by biogenic VOCs and their influence on the predicted effects of reducing anthropogenic emissions have also been investigated (Simpson, 1994b).

2.3   An Integrated Assessment Model for Ozone

Stated briefly, the aim of an integrated assessment model for tropospheric ozone is to describe the relationship between ozone exposure and the emissions of ozone precursors, NO_x and VOCs, in such a way that the costs and effectiveness of emission reduction strategies within Europe can be quantified. To achieve this requires information from each of the following areas:

- the emissions of NO_x and VOCs, both man-made and natural.

- the abatement technologies available for NO_x and VOCs, and their costs.

- a concise description of the relationship between ozone concentrations and precursor emissions, taking account of meteorological influences on ozone formation.

- ozone measurement data from across Europe in order to take account of local variations, due to altitude for example, and to investigate the relationships between accumulated exposure to ozone and measured ozone concentration statistics.

- studies of the effects of ozone on agricultural crops, forests and human health, leading to the establishment of critical levels for ozone in different areas of Europe.

Data from each of these areas are necessary for the development of the integrated assessment model and all are critical to its ultimate success.

On the emissions side, the current RAINS integrated assessment model (Alcamo et al, 1991) already provides the necessary information on NO_x emissions. A design study (Olsthoorn, 1994) of the VOC emission inventories, scenarios and emission control options proposed using the CORINAIR90 inventory as the basis for the VOC part of the integrated model. The availability and costs of technologies to reduce VOC emissions have been reviewed (Caliandro, 1994).

The third item in the list above concerns the source-receptor relationships involved in ozone formation and is the subject of this paper. The EMEP ozone model is believed to give reliable estimates of ozone concentrations during the afternoon period when daily maximum values are most likely to occur. This model is designed to calculate ozone over extended periods - of several months - and throughout Europe and can, therefore, provide the basis for the ozone formation module of an integrated assessment model.

In this study, the formulation of a simplified description of ozone formation is dependent on the results of the EMEP ozone model. Because this model is designed to cover the whole of Europe, it uses a relatively large grid size of 150 km. However, an integrated assessment model for ozone needs to consider the effects of ozone exposure (above a specified threshold level) on ecosystems that may be much smaller than this. Consequently, rather than base the simplified description of ozone formation on the EMEP model's predictions of ground-level ozone concentrations, which represent an average over an area related to the model's grid size, a two-stage approach is adopted here. The first stage of the ozone formation module, the subject of this paper, would provide boundary layer average ozone concentrations; the development of this module is based on boundary layer average ozone concentrations predicted by the EMEP model. In the rest of this paper, reference to EMEP model results implies boundary layer average concentrations. The relationships between the EMEP model, and its inputs and outputs, and the simplified ozone formation module are shown schematically in Figure 1.
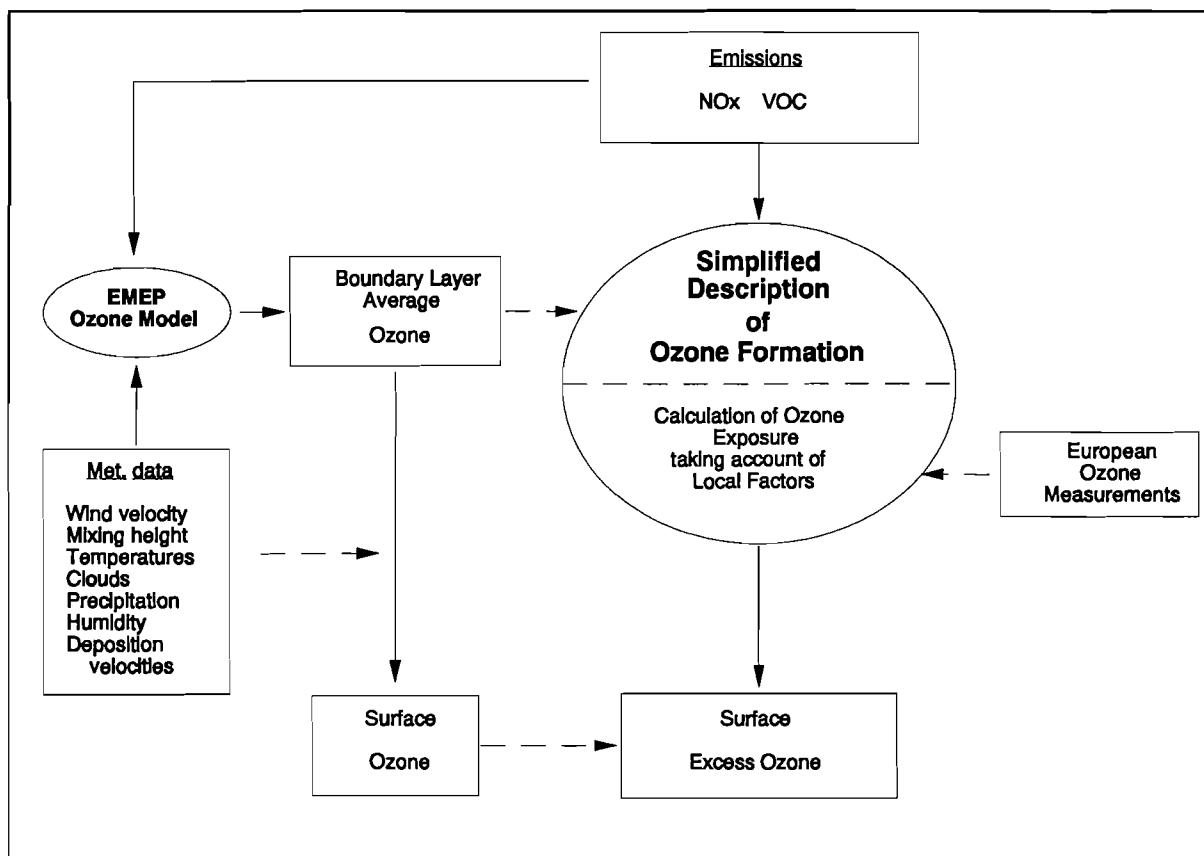

Figure 1 Relationships between Simplified Ozone Formation Module and EMEP Model

In a subsequent step, the ozone exposure at ground level at a particular location would be calculated from the boundary layer average ozone for the area, taking local factors into account as necessary. These factors might include features such as latitude, distance from the coast, altitude and local emissions of NO_x. Analyses of ozone measurement data, aimed at identifying the important features and investigating ways of incorporating them into an integrated assessment model, are reported separately (Kettunen et al, 1994; Baldi et al., 1994). Figure 1 provides a schematic overview of the two-part concept of the ozone formation module; the first stage depends on the results of the EMEP ozone model and the second is derived from analyses of European ozone measurement data.

9

Studies of the impacts of ozone have resulted in the apparent consensus that critical levels for ozone in order to protect agricultural crops and forests can best be established with long-term exposure measures, in particular, by the 'accumulated excess ozone' concept. Currently, a threshold concentration of 40 ppb is proposed for both crops and trees. This exposure index is referred to as AOT40, the accumulated exposure over a threshold of 40 ppb (Fuhrer and Achermann, 1994). For agricultural crops the accumulated exposure should be calculated for daylight hours, defined as those hours with a mean global radiation of 50 Wm^{-2} or greater, and for three months (May-July, except in areas where this period is inappropriate). The current proposal for forest trees recommends that the cumulative exposure be calculated for 24 hours a day during a six-month period. Furthermore, most current international air quality standards to protect human health are defined as short-term concentrations. An integrated assessment model needs to take these different measures into account.

2.4   EMEP Ozone Model Results

This section provides details of the calculations performed with the EMEP ozone model to provide the necessary basic set of data for the study. These results are then illustrated using a form of ozone isopleth diagram.

2.4.1 Model Calculations

Initial investigations of simple relationships between ozone and its precursors were based on model results for 25 sites for the period April - September 1989. Sixteen of these sites, shown in Figure 2, lie on a cross-section of the EMEP grid (EMEP x coordinate = 19), providing data for a range of different environments; the other nine are measurement sites. For the base case with no emission reductions, there were data available on emissions, meteorological parameters and the concentrations of some other chemical species, in addition to the final ozone concentrations. These parameters were provided as averages (and, for the emissions, also as totals) over the final 72 hours along each trajectory. Additionally, the model had been used to predict ozone concentrations for 36 scenarios corresponding to different combinations of NO_x and VOC reductions, shown in Figure 3, assuming in each case that the same fractional reduction had been applied to all emission sources in Europe.

2.4.2 Ozone Isopleth Diagrams

A convenient way to illustrate these results is by means of ozone isopleth diagrams. Such diagrams have been most commonly used, particularly in North America, to show how maximum ozone concentrations depend on the initial concentrations of NO_x and VOCs on a particular day at a specific location. Lines of constant value, or isopleths, of the maximum ozone concentrations are constructed by connecting points having the same ozone concentration but corresponding to various initial conditions. Ozone isopleth diagrams in this form provide a concise representation of the effect of reducing initial NO_x and VOC concentrations on peak ozone concentrations and, in the past, they have been used quantitatively to develop ozone control strategies as part of the U.S. EPA's empirical kinetic modelling approach (EKMA).

The isopleth diagrams used in this section are constructed rather differently, although there are obvious similarities in appearance. Firstly, the ozone statistic depicted by the isopleths is the mean, over the six-month summer period, of the daily maximum ozone concentrations calculated by the EMEP model. The daily maximum has been taken simply as the largest of the four values predicted by the model for each day. Secondly, in the version
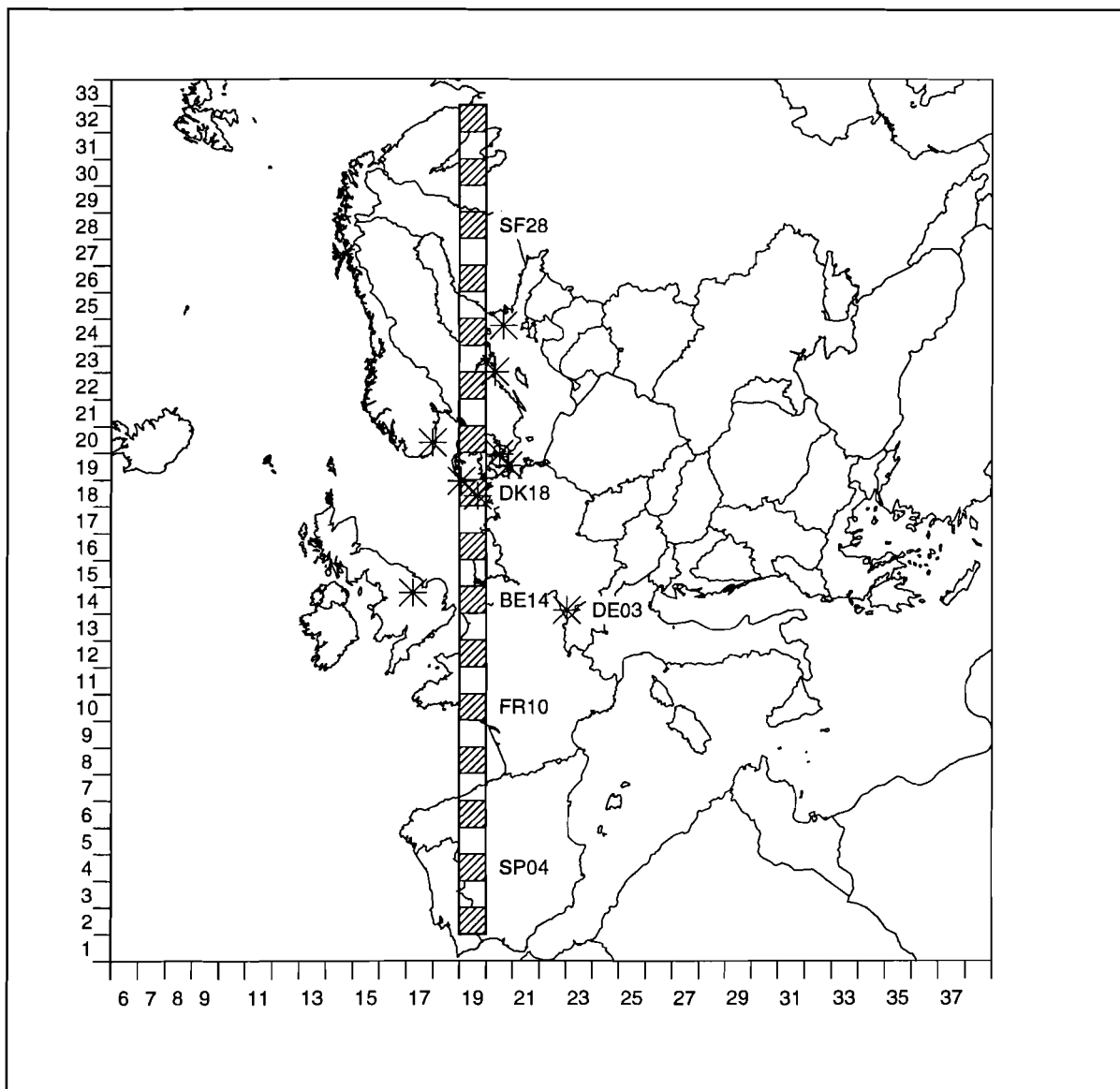
10

Figure 2 EMEP Cross-Section and Measurement Sites used as Model Receptors

used here, ozone is shown as a function of the percentage reduction in emissions of NO_x and VOCs. Thus, the top right-hand corner of each diagram represents the base case without any reduction in precursor emissions.

Figure 4 shows the mean daily maximum ozone at six receptor sites as a function of the percentage reduction in emissions of NO_x and VOCs implemented uniformly throughout Europe. Figures 4(a-e) are isopleth diagrams for grid squares along the EMEP cross-section shown in Figure 2. Figure 4(f) presents results for Schauinsland in Germany, a site which has been heavily used in the development of the statistical description of ozone formation described in later sections of this report.

In Figure 4, locations near the ends of the EMEP cross-section are represented by the grid squares labelled SP04 in Spain (shown in Fig. 4a) and SF28 in Finland (Fig. 4e). At these sites, it can be seen that the ozone concentration (six-month mean of daily maxima) is largely independent of VOC emission reductions. The importance of the NO_x / VOC ratio in
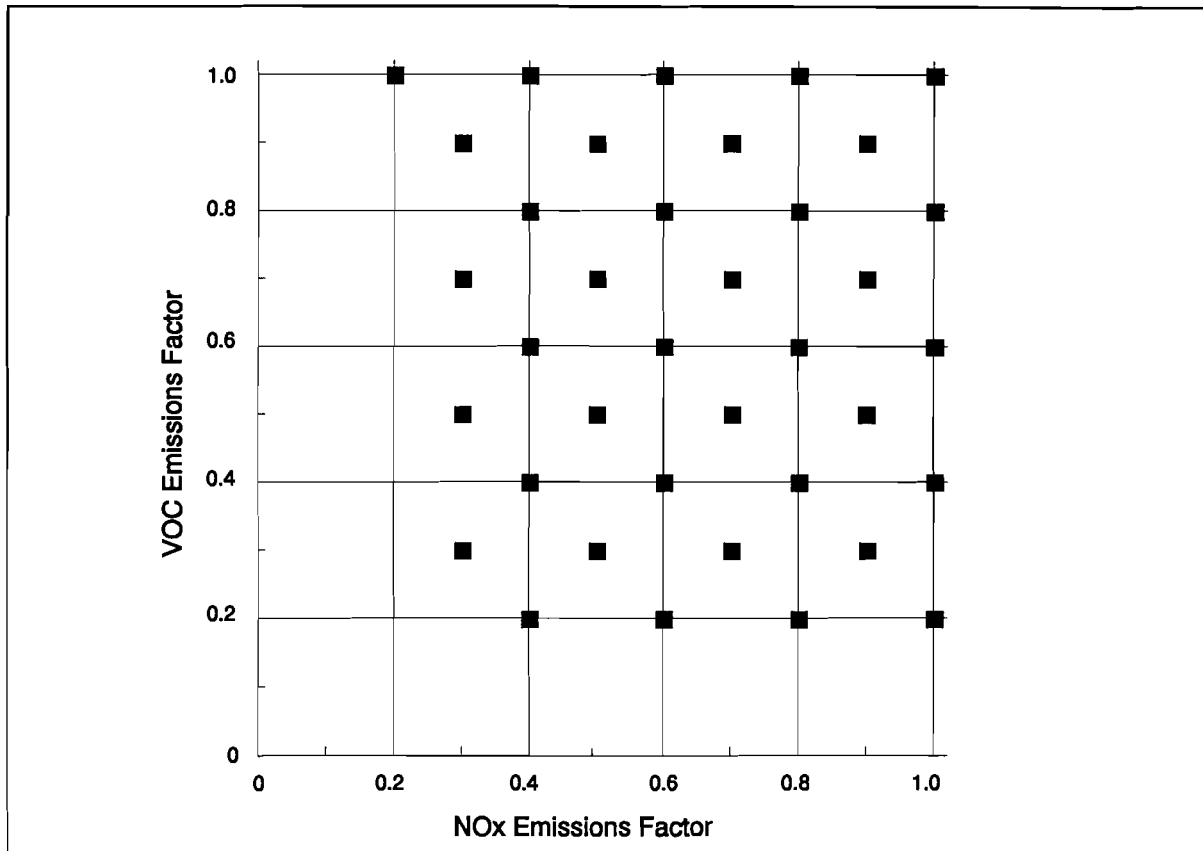
Figure 3 Combinations of NO_x and VOC Emission Reduction Factors

determining ozone production, and the underlying atmospheric chemistry, were discussed in Section 2.1. In regions exemplified by the SP04 and SF28 grid squares, the NO_x / VOC ratio is relatively low. There is an ample supply of peroxy radicals (RO_2 and HO_2) to convert NO to NO_2 and, thus, lead to ozone production. Decreasing the available NO_x leads directly to a decrease in ozone. In these circumstances, ozone formation is limited by the availability of NO_x, and the atmospheric chemistry system is said to be NO_x-limited. In such regions, then, reductions in emissions of NO_x are likely to be effective in reducing ozone concentrations, but ozone is insensitive to reductions of VOC at constant NO_x and is also insensitive to variations in VOC composition.

In the area of NW Europe typified by the BE14 grid square in the Belgium/Netherlands area, the isopleth diagram (Fig. 4c) shows a different picture. The isopleths form a ridge dividing the diagram into two areas. On the left of the ridge, corresponding to the greatest reductions in NO_x emissions, the system tends towards the NO_x-limited case discussed above. On the right side of the ridge, the NO_x / VOC ratio is relatively high and the NO_2 concentrations are sufficiently great that NO_2 competes with the VOCs for reaction with the OH radical, as discussed in Section 2.1. In this region of the diagram, reducing VOC emissions results in lower ozone concentrations; however, ozone concentrations may be increased, at least initially, by NO_x reductions in the absence of concurrent reductions in VOC emissions.

The ozone isopleth diagrams for the FR10 (Fig. 4b) and DK18 (Fig. 4d) grid squares, which lie in intermediate positions between the centre and ends of the EMEP cross-section, show intermediate ozone formation behaviour corresponding to their geographical locations.
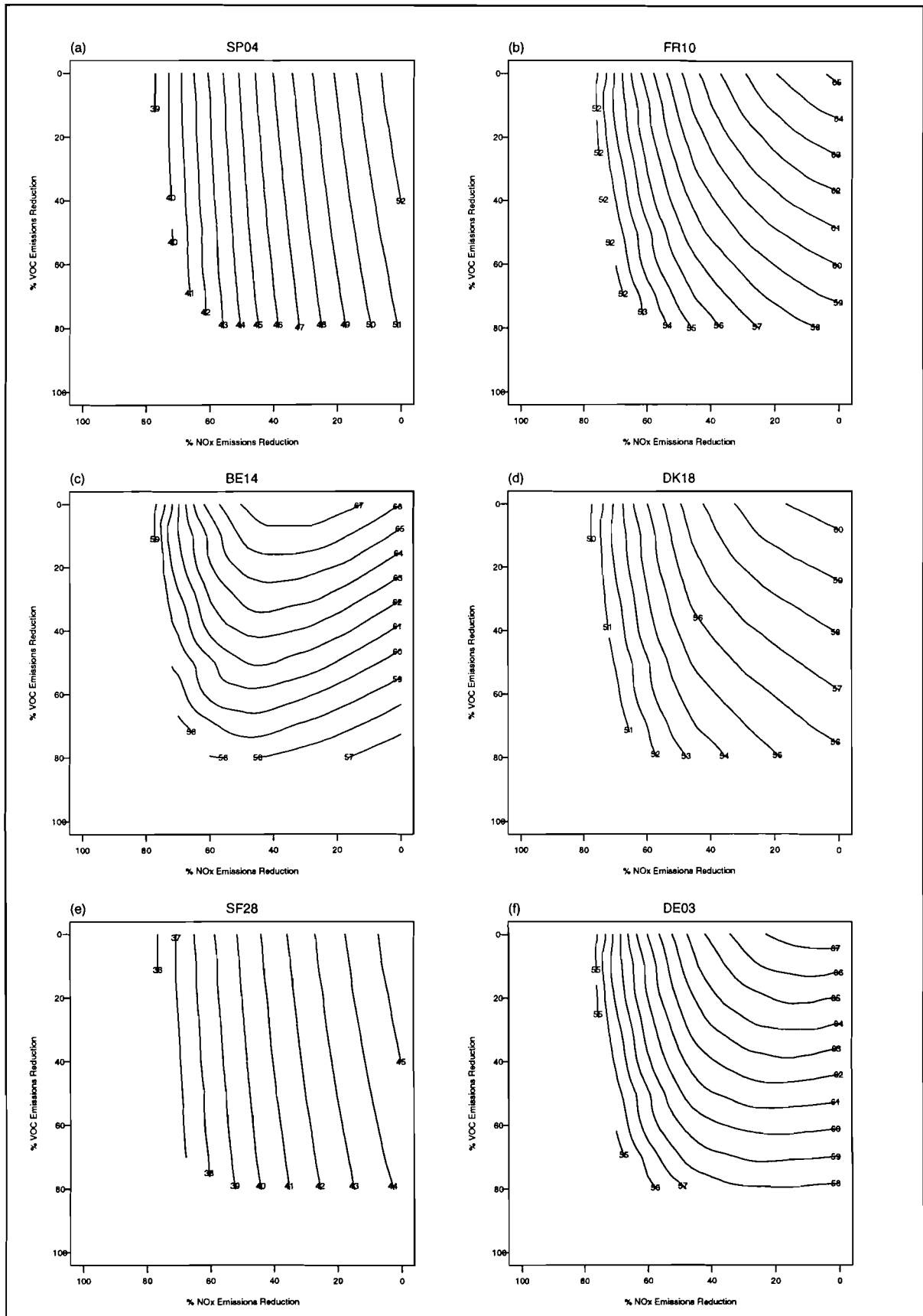
12

Figure 4 Modelled Ozone Concentrations (ppb) as a Function of NO$_x$ and VOC Reductions

A similar pattern can be observed in the isopleth diagram for the Schauinsland site shown in Figure 4(f). Here, a ridge can be discerned in the isopleths but it is less pronounced than that seen in the BE14 grid square plot.

Although instructive in providing a qualitative illustration of the relationship between ozone concentrations and NO_x and VOC emissions, such isopleth diagrams are not ideally applicable to an integrated assessment model. Firstly, they are based on the premise that the same fractional emission reductions take place throughout Europe whereas one of the main purposes of an integrated assessment is to consider the relative merits of regional differences in the application of abatement measures. Secondly, in the form presented here, they provide a picture of the overall effects of emission control over a six-month period. While this may be adequate in relation to the effects of ozone on forests, consideration of ozone exposure to crops requires greater temporal resolution because of differences in growing season between different crop species and between different locations. These factors indicate that a more detailed description of ozone formation is required than can be obtained from the simple ozone isopleth diagrams presented here.

3.   THE STATISTICAL APPROACH

As indicated in the Introduction, a number of ways of constructing a simplified description of ozone formation may be envisaged, ranging from those in which the emphasis is placed on the chemical processes to others that rely on statistical methods. The work described here took an approach of the latter type, with the aim of using statistics to summarise the results of a detailed photochemical oxidant model.

In essence, the aims of this approach are to identify the most important factors leading to ozone production, in other words the best predictor variables, and then find an adequate description of their relationship to ozone concentrations. The data required for this investigation are provided by the EMEP ozone model and, indeed, the present study is entirely dependent on the EMEP model. Consequently, the search for the best set of ozone predictor variables is restricted to information that can be obtained from the EMEP model. If successful, the outcome will be a condensation of the most important facets of the EMEP ozone model into a format that is appropriate to an integrated assessment model.

The ozone isopleth diagrams described in Section 2.4.2 summarise the mean formation of ozone from its precursors over a six-month period. In contrast, the development of this statistical approach is based on daily data. Thus, the ozone concentrations considered are either daily maximum values or the final concentrations associated with air mass trajectories arriving at 1200 GMT. Similarly, for each ozone datum, the predictor variables take their values from the trajectory leading to that particular ozone concentration.

The ozone formation process is very complex; the chemistry included in the EMEP ozone model is a considerable simplification of reality yet still consists of 136 reactions. Meteorological factors add further to the complexity and it would be impossible to derive a simple analytical form to describe all the interconnected processes. However, although in the EMEP model the emissions and meteorological data are changed in step form every two hours, the model chemistry uses only ordinary differential equations, and, with respect to changes in emissions, a smooth analytical form can be expected. Therefore, any statistical

model for this system must be capable of approximating smooth functions. This is what standard non-parametric models are designed for.

Of this class of statistical models, the most commonly used are local and polynomial regression models. The question arises: how well do these models approximate any smooth function in practice? The number of degrees of freedom of the regression surface is limited by the size of the sample. Furthermore, there are numerical limitations to the complexity of the functional form. These problems manifest themselves in different ways in the different types of models. Within these limitations, both types of model provide tests to assess the optimal number of degrees of freedom.

The technique selected for investigating and summarising the relationships between ozone and its predictors in this study was a local regression model although, in fact, the model actually used here contained some elements of polynomial fitting. The following section provides a brief overview of local regression models. Subsequent sections present the results of the investigation.

3.1 Local Regression Modelling

Local regression models provide methods for fitting regression functions, or regression surfaces, to data. If only one predictor is used, then the fitted function is a curve. With two predictors a surface is fitted and the result can be shown by a contour plot. Local regression models provide considerable flexibility in that no restrictions are placed on relationships between the predictors; the model is fitted as a single, smooth function of all the predictors.

One basic specification in a local regression model is that for a point x in the space of the predictors there is a local region containing x in which the regression surface can be well approximated by a function from a specific parametric class. The particular model used in this study allowed two classes - linear or quadratic polynomials. The specifications of local regression models lead to methods of fitting that consist of smoothing the response as a function of the predictors; thus the fitting methods are non-parametric regression procedures.

The local region, or neighbourhood, is determined by recursively splitting the data space in half. Starting with a rectangular cell containing the values of the predictor variables, the cell is divided at the median value of the predictor whose spread is the greatest. This procedure is applied recursively to each sub-cell, as demonstrated in Figure 5, until each cell contains fewer than a pre-determined number of points, this number depending on the amount of smoothing desired. Within each cell a parametric function is fitted to the data and then constrained to join smoothly with the functions of neighbouring cells.

In fitting a local regression model, assumptions are made about properties of the regression surface and the errors involved. This study made use of a method of fitting local regression models called *loess*, available within the S statistical software package (Chambers and Hastie, 1992). The specifications possible within this method are listed briefly below. The following choices about the properties of the errors are available:

• Gaussian or symmetric distribution of errors;

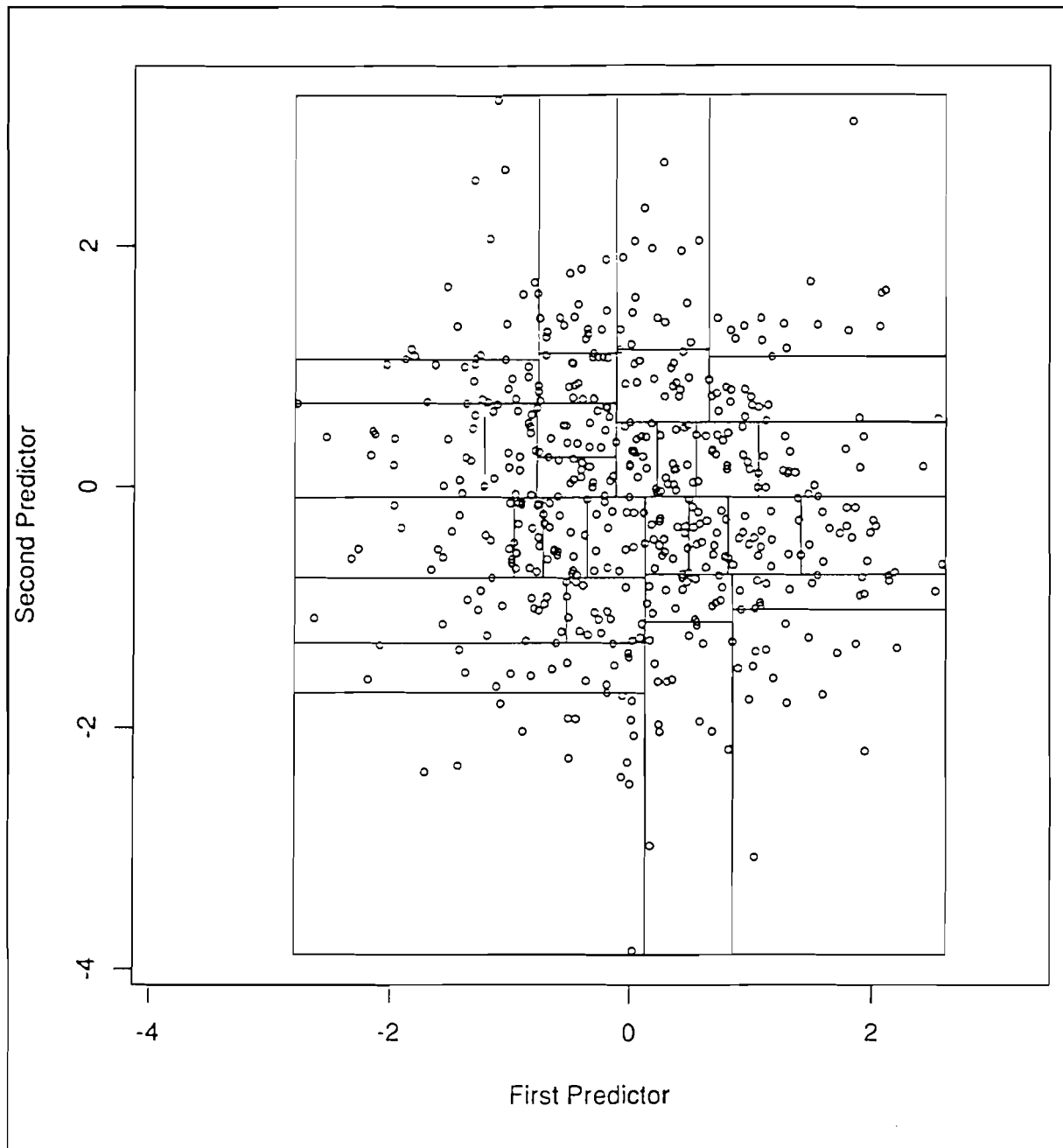• constant variance or *a priori* weights.

Figure 5 Division of Data Space in Local Regression Model with Two Predictors

In this study it was assumed that the errors had constant variance and followed a Gaussian distribution. The regression surface could be specified by selecting:

- locally linear or locally quadratic polynomials;

- the local neighbourhood size, by means of a smoothing parameter or span;

- normalisation of the scales of the predictors;

- when quadratic interpolation has been specified, the option of dropping the square term(s) of one or more predictors from the polynomial function;

16

- a conditionally parametric subset. If a particular predictor, u, is specified to be conditionally parametric, this has the effect of ensuring that, given values of the other predictors, the regression surface is a parametric function (either linear or quadratic) of u. If the given values are changed, the surface remains parametric in u, although the parameters might change. Specifying a regression surface to be conditionally parametric in a particular variable is sensible if examination of the data suggests that the surface is everywhere a very smooth function of that variable.

Further discussion of the theory and implementation of local regression models is beyond the scope of this report. Detailed treatments of local regression models and the computational methods crucial to their practical implementation (Cleveland and Grosse, 1991), and the fitting method used in the S language (Chambers and Hastie, 1992) are available in the literature.

3.2   Local Regression Model of Ozone Formation

In attempting to develop a regression model that would provide a relatively simple but reliable description of ozone formation at a particular site, the following general approach was adopted:

- investigation, typically by graphical means, of the data generated by the EMEP ozone model;

- construction of a regression model based on variables identified by the data exploration;

- graphical diagnosis of the regression model to assess goodness of fit and validity of the assumptions made in specifying the model;

- evaluation of the success of the regression model by comparison with the original EMEP model predictions of ozone concentrations and, where possible, with measured data.

Initially, this approach was applied to the EMEP model data obtained for 25 sites, as described in Section 2.4.1, to provide an idea of the differences to be expected in different parts of Europe. The most recent developments have required additional model runs of the EMEP model in order to obtain further data, particularly to investigate a greater number of potential predictor variables and to improve the homogeneity of the data with respect to NO_x and VOC emissions. To date, testing of the latest refinements has been largely restricted to one site, Schauinsland in SW Germany.

3.2.1 Data Exploration

Initial exploratory data analysis relied heavily on the conditioning plot, or *coplot*, graphical method, which provides a way of seeing how a response depends on one predictor variable given the values of other predictors. In the present application, the response is the ozone concentration predicted by the EMEP model at a particular site, and the predictor variables can, in principle, be any parameters, obtainable from the EMEP model, relating to the appropriate air mass trajectory arriving at that site. Since the purpose of the exercise is to develop an integrated assessment model, two of the predictor variables will, inevitably, be

parameters involving the emissions of NO_x and VOCs that are picked up by an air mass along the trajectory path.

Several emission-related parameters were assessed for their applicability as ozone predictors, together with a number of meteorological variables. Initially, the total anthropogenic emissions, of NO_x and VOCs, picked up along the trajectory during the final 72 hours before arrival at the site of interest were investigated as predictor variables. This simple first approach was improved upon by including natural VOC emissions in the VOC total. A further improvement was suggested by recent studies with the EMEP model (Simpson, 1994b) which included an investigation of how much of the emissions, ozone, NO_x, etc. survived from the earlier stages of a trajectory to have an impact on the final ozone concentration. At least at some sites, it seems that often rather little does, with the result that analysis of ozone in terms of trajectory-integrated emissions may be misleading. The EMEP model includes two processes by which boundary-layer air can be mixed with free tropospheric air, namely day-to-day increases in mixing height and the venting effect of cumulus clouds. To allow for these effects, emissions along the trajectory were weighted by the amount of dilution that subsequently takes place within the air mass (Simpson, 1994b). Details of the weighting procedure are provided in Appendix I. The dilution-weighted or "effective" NO_x and VOC emissions were then assessed as predictors of ozone in a regression model. Examples of regression models based on these two predictors alone are presented in section 3.2.2.

Further analysis of the errors associated with statistical models based solely on emission-related predictors indicated that the errors had a seasonal dependence. This could be removed, and the model improved considerably, by including as an additional predictor variable the mean NO_2 photolysis rate over the four days of the trajectory. These three trajectory-related variables, dilution-weighted or "effective" NO_x emissions, "effective" VOC emissions, and mean NO_2 photolysis rate, constitute the optimum set of predictors identified by the present study. No substantial improvements were achieved by the inclusion of additional meteorological parameters, such as temperature, for the sites and grid squares so far investigated.

Figure 6 shows an example *coplot* illustrating the relationships between ozone and the three predictor variables at the Schauinsland site. The figure consists of a 4 x 4 array of dependence panels, in each of which ozone concentration (ppb) is plotted against "effective" VOC emissions, in this example, for the data corresponding to the given ranges of "effective" NO_x emissions and mean NO_2 photolysis rate. The additional panels at the top and right of the figure indicate the ranges of these latter variables. These ranges have been chosen such that there are an equal number of points within each range, rather than simply dividing the total range equally, and, as can be seen, with some overlap between adjacent ranges. The bottom row of dependence panels shows only those data corresponding to the lowest quartile, approximately, of mean NO_2 photolysis rates. Similarly, the leftmost column of panels corresponds to low "effective" NO_x emissions. The panel in the top right-hand corner, for example, shows how ozone depends on "effective" VOC emissions when both the "effective" NO_x emissions and mean NO_2 photolysis rate are high. Moving from left to right through the array of dependence panels shows how the dependence of ozone on VOC emissions varies with increasing NO_x emissions for a given range of NO_2 photolysis rates. Moving from bottom to top indicates the influence of the NO_2 photolysis rate on the ozone:VOC relationship.

At Schauinsland, and many other sites, conditioning plots such as Figure 6 showed that, in general, ozone concentrations increase, approximately linearly, with increasing "effective" VOC emissions and with increasing NO_2 photolysis rates. The role of "effective"
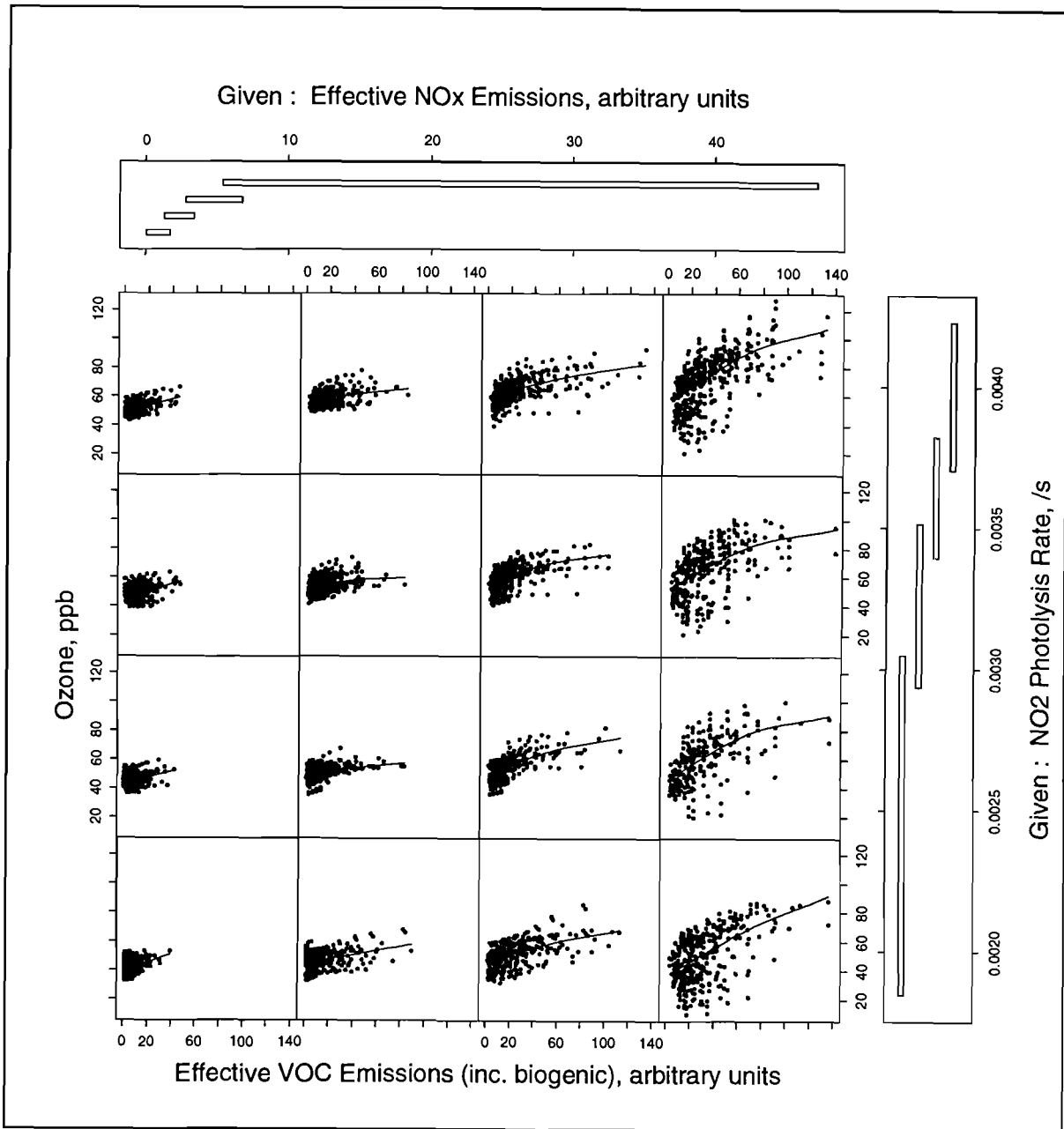
18

Figure 6 Schauinsland data - Coplot of O_3 against Effective VOC Emissions given Effective NO_x Emissions and Mean NO_2 Photolysis Rate

NO_x emissions is more complex, however, with the highest NO_x emissions leading to reductions in ozone unless both the "effective" VOC emissions and the NO_2 photolysis rates are sufficiently high. In all cases, there is scatter in the data, especially when the "effective" NO_x emissions are high. This scatter is not unexpected in view of the many other influences on ozone formation, discussed in section 2.1, which will affect the data but are not explicitly taken into account by the three predictor variables used in this analysis.

3.2.2 Potential Models

When just two predictors are used in a regression model, it is possible to display the model as an isopleth diagram, or contour plot. This is done in Figure 7 for the same set of
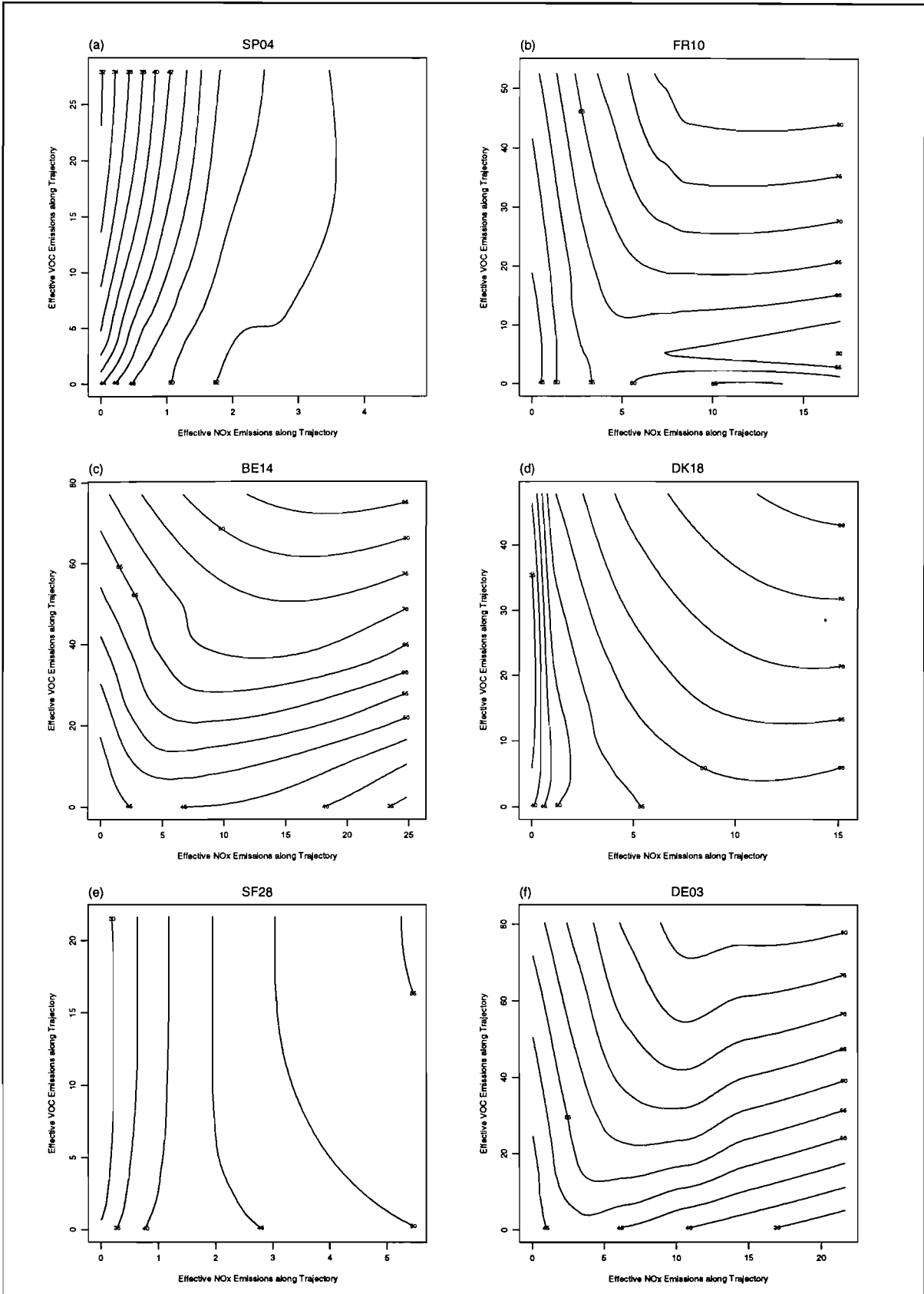
Figure 7 Isopleth Diagrams of Regression Model Ozone Predictions as a Function of Effective NO$_x$ and VOC Emissions

sites as used in the simple percentage emissions reduction isopleth diagrams shown in Figure 4. The two predictor variables used are the "effective" NO_x emissions and "effective" VOC emissions discussed in the previous section. For each site a local regression model of the 1200 GMT ozone concentration was constructed based on these predictors. In the examples presented here, the model was specified to be locally linear. The resulting models were used to predict ozone concentrations on a regular grid in the NO_x - VOC emissions plane and, thus, to produce isopleth diagrams as a function of emissions along the trajectory. Because of some spatial inhomogeneity in the set of data, leading to potential artefacts introduced by the surface-fitting and smoothing procedures where data are sparse, the diagrams shown in Figures 7(a-f) are restricted to an area bounded by the 98th-percentile values of the emissions parameters.

The first obvious information provided by such isopleth diagrams is the variation in the amounts of NO_x and VOC emissions picked up along trajectories arriving at different parts of Europe. Air masses reaching Belgium, the Netherlands, Germany, UK and Denmark receive more ozone precursor emissions than areas to the north or south of those countries.

For sites and grid squares in central areas of the EMEP cross-section, the ozone isopleths take the form of a ridge running from the origin towards higher NO_x and VOC emission values. For some sites, the greatest ozone concentration is predicted to occur in the area of maximum NO_x and maximum VOC emissions. At others, however, the maximum ozone concentrations occur at values of the NO_x emissions below the maximum. In these cases (e.g. Fig. 7c), particularly when the isopleth ridge is sharper, reductions in NO_x emissions may lead to an increase in ozone. This was most notably the case in the Belgium/Netherlands area and also in parts of Germany, Denmark and the UK.

As described in the previous section, it is possible to improve on the regression models shown in Figure 7 by including the mean NO_2 photolysis rate as a third predictor variable. A model constructed on this basis for the Schauinsland site is shown using a *coplot* representation in Figures 8(a-c). In this case, the model was specified to be locally quadratic, as is evident from the figure. Each dependence panel in Figure 8 shows the relationship between the 1200 GMT ozone concentrations and one predictor variable, for particular given values of the other two predictors. Here, the given values have been chosen at equally spaced intervals between the 5th and 95th percentile values of the relevant predictor. In contrast to the *coplot* shown in Figure 6, the Figure 8 *coplots* show the regression model predictions rather than the original set of data, and so these figures use particular values of the given variables instead of the ranges employed in Figure 6. In other respects, the Figure 8 *coplots* should be interpreted in the same way as the earlier figure.

Conditioning plots are also useful in examining residuals, the differences between the model predictions and original data. Figure 9 provides an example, again for the Schauinsland model shown in Figure 8, in which the residuals are plotted against the mean NO_2 photolysis rate for given ranges of the other two predictor variables. The residuals shown here do not indicate lack of fit in any particular area but they do highlight the increased errors that tend to occur when the "effective" NO_x emissions are high.
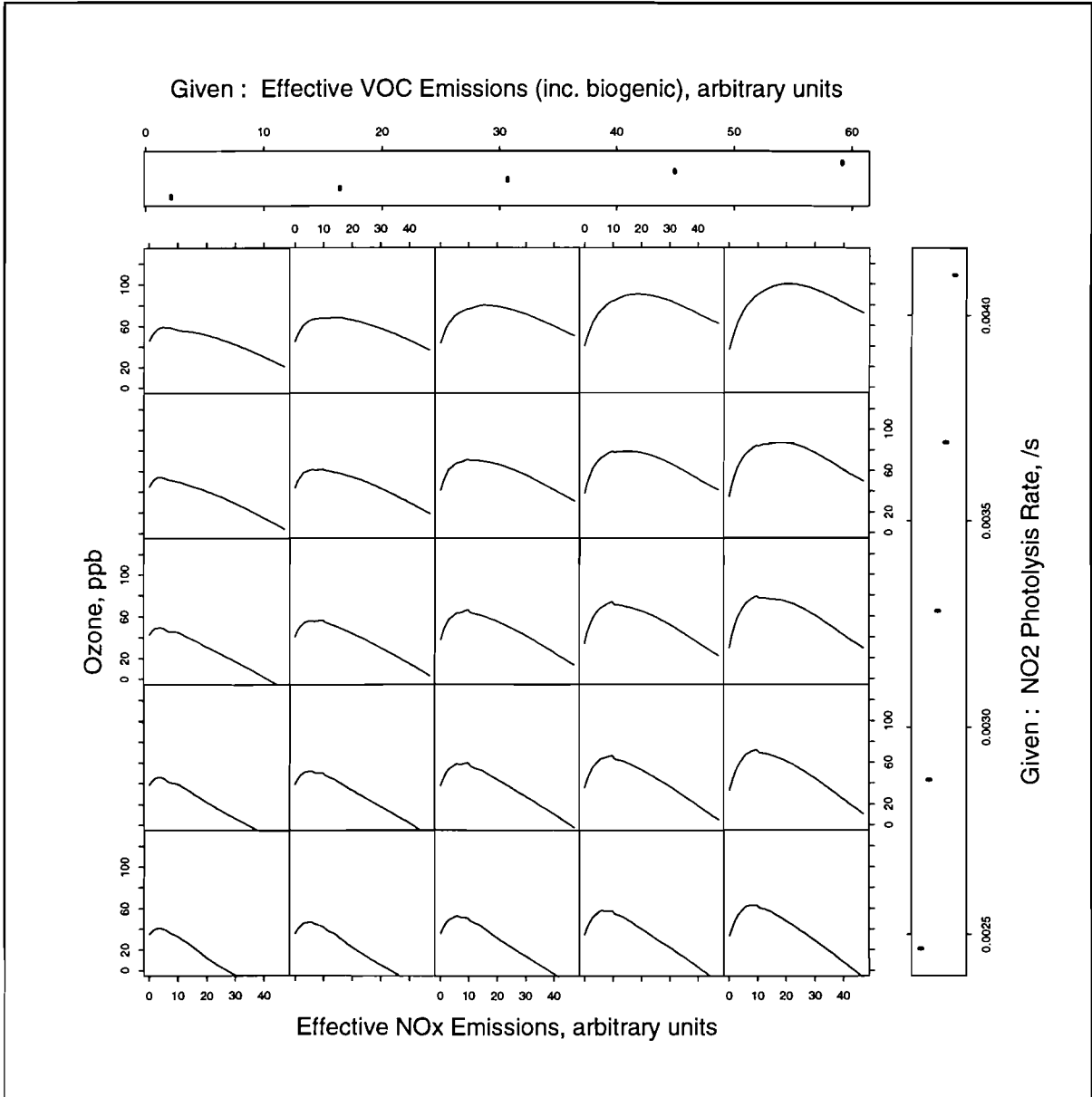
Figure 8(a)   Schauinsland Regression Model - Coplot of O_3 against Effective NO_x Emissions given Effective VOC Emissions and Mean NO_2 Photolysis Rate
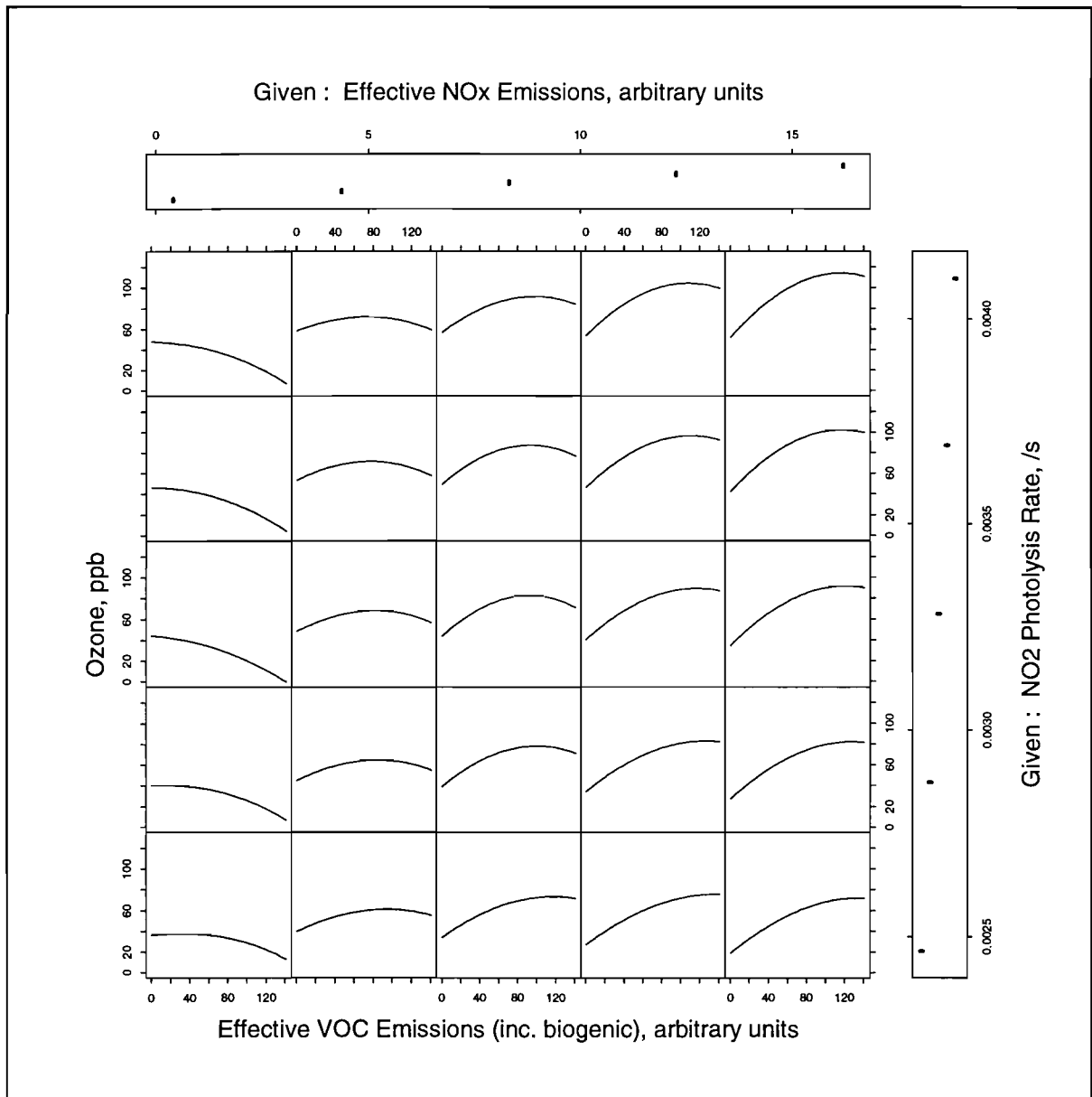
Figure 8(b) Schauinsland Regression Model - Coplot of O_3 against Effective VOC Emissions given Effective NO_x Emissions and Mean NO_2 Photolysis Rate
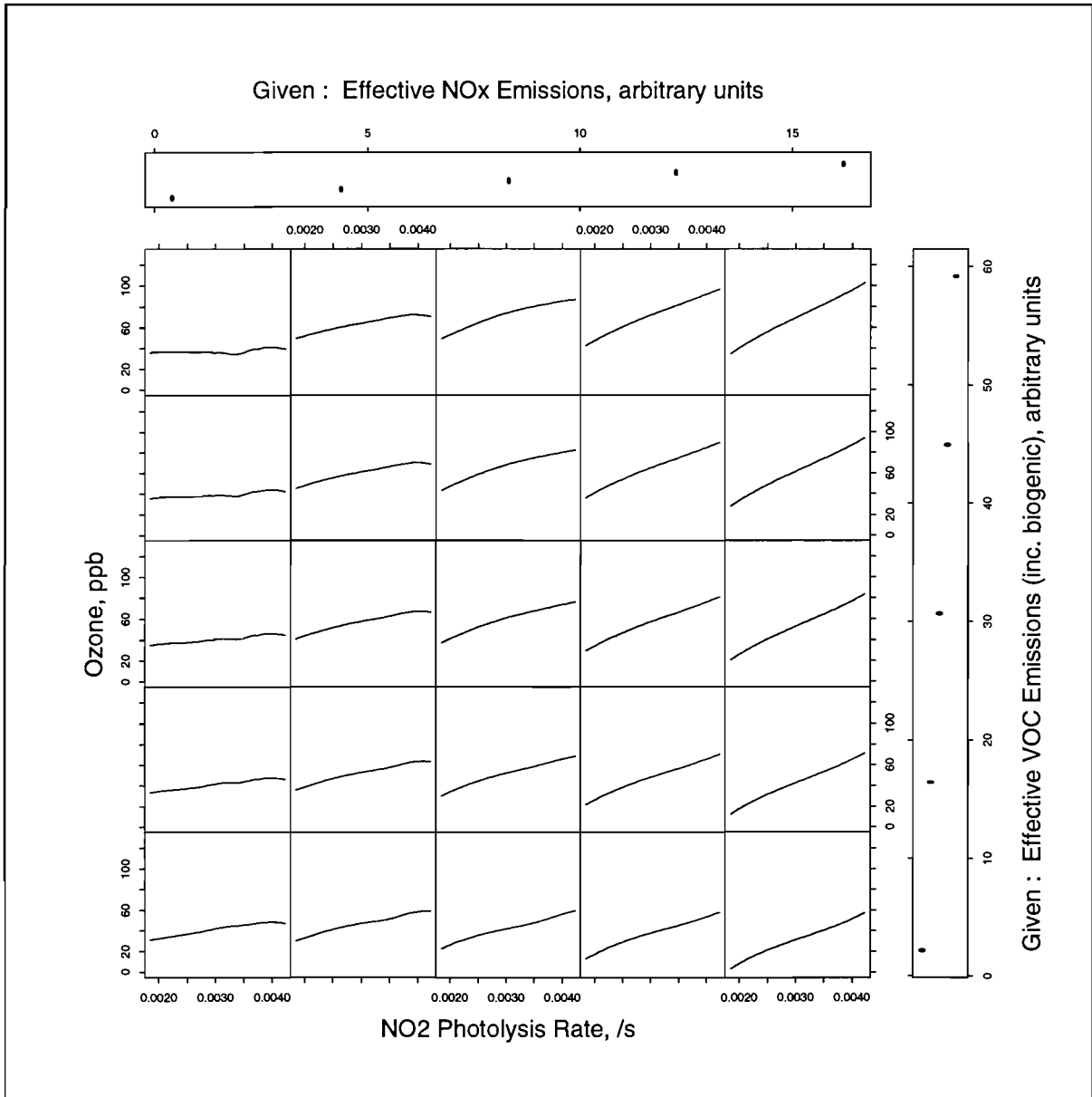
23

Figure 8(c) Schauinsland Regression Model - Coplot of O_3 against Mean NO_2 Photolysis Rate given Effective NO_x Emissions and Effective VOC Emissions
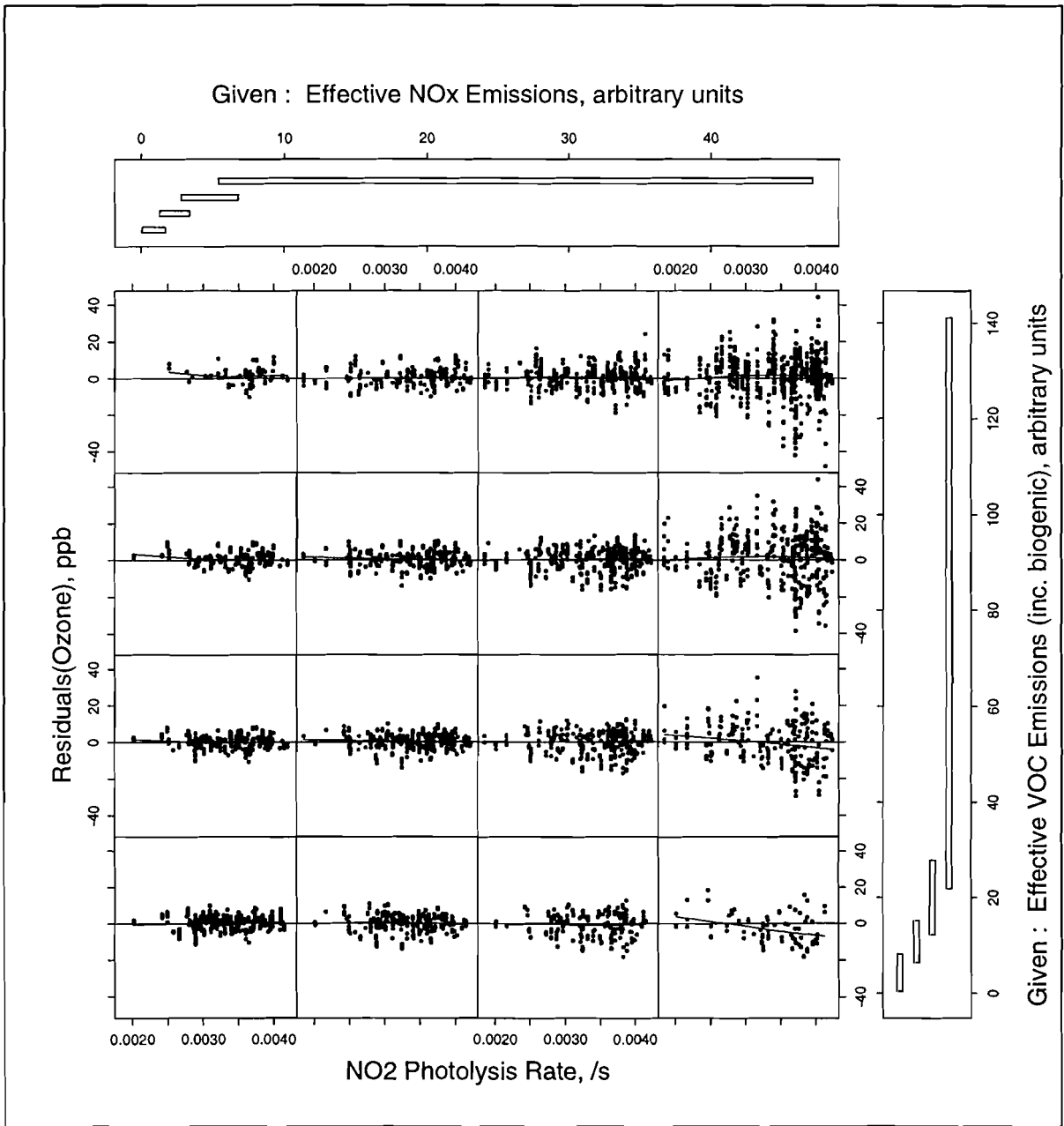
Figure 9 Schauinsland Regression Model - Coplot of Residuals

The regression surface illustrated in Figure 8 was fitted assuming a Gaussian distribution of the errors. To check the validity of this assumption, a Gaussian probability plot of the residuals was constructed, shown in Figure 10. The straight line added to this plot is drawn through the lower and upper quartiles of the data to assist in assessing the linearity of the points. The result shown here indicates that the tails of the error distribution are longer than a Gaussian distribution, suggesting that the Gaussian specification may not be completely justified.

Following graphical diagnostic tests of the sort exemplified above, the optimal specifications of a local regression model to represent ozone formation at Schauinsland appear to be locally quadratic interpolation with the "effective" VOC emissions designated as conditionally parametric. Figures 8-10 are based on such a model with the smoothing
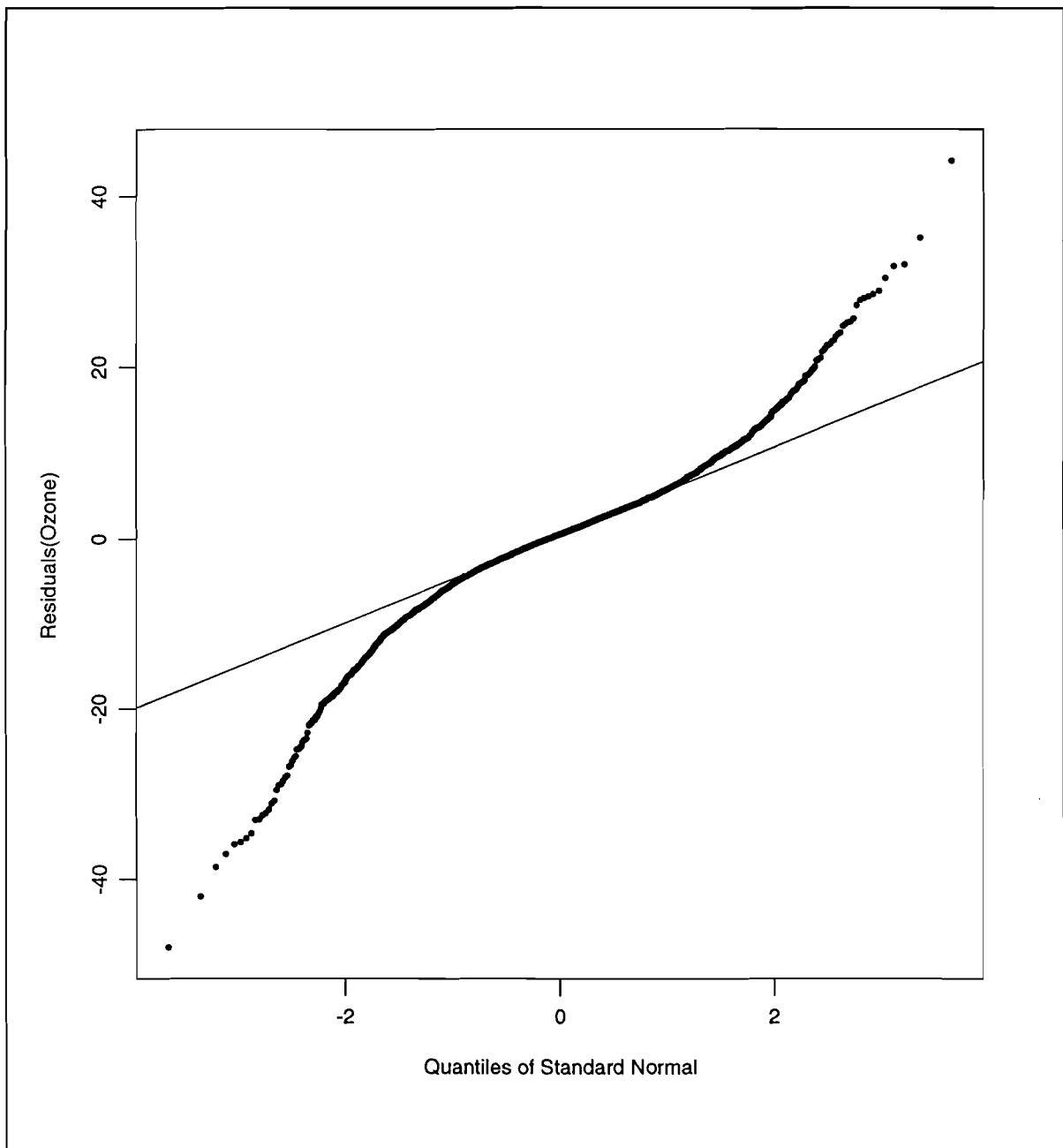

Figure 10 Schauinsland Regression Model - Gaussian Probability Plot of Residuals

parameter, related to the size of the local data neighbourhoods, set to a value of 0.5. In practice, this span value results in the data space being divided until the local data cells or neighbourhoods each contain less than 10% of the total number of data points. Figures 8-10 illustrate the form of the regression surface and provide an indication of the quality of fit achieved. For this model the residual variance is estimated to be 7.3 ppb.

Having developed a local regression model for Schauinsland that provides a reasonable description of ozone formation, it is useful to investigate how the quality of fit provided by this model is affected by the smoothing parameter, or span. Analysis of variance tests can provide a quantitative assessment of the change in quality of fit in such circumstances. Results of successive analysis of variance tests on a sequence of models in which the span value was varied from 0.2 to 1 are given in Table 1.

Table 1 Results of Analysis of Variance Tests

Model No	Span	Equivalent No. of Parameters	Residual Standard Error	Multiple R^2	Test	F Value
1	0.2	28.4	7.184	0.74	1 vs 2	0.54
2	0.3	19.3	7.178	0.74	2 vs 3	18.17
3	0.4	14.9	7.302	0.73	3 vs 4	1.20
4	0.5	12.3	7.302	0.73	4 vs 5	66.08
5	1	10.0	7.674	0.70		

The values of the F statistic presented in Table 1 indicate that increasing the span value above 0.3 leads to a distortion in a statistical sense. On the other hand, reducing the span to 0.2 does not improve the fit, suggesting that 0.3 may be close to the optimal value for this parameter. A span value of 0.5 also appears to be better than its immediate neighbours; an increase in span leads to distortion whereas decreasing the span to 0.4 fails to improve the quality of fit.

It is also of interest to consider to what extent the quality of fit deteriorates if a simpler parametric model specification is used. This may be investigated by setting the smoothing parameter such that no division of the data space into local neighbourhoods is carried out. This is achieved when the span value is set to 1 (model no. 5 in Table 1). Such a model is equivalent to a second-order polynomial regression. This polynomial regression has a constant and three independent variables, leading to ten parameters, as indicated in the Table. The result of an analysis of variance test of the comparison between such a model and the model with a span of 0.5 is shown in Table 1. Again, the F statistic indicates that the simpler model will be distorted.

In interpreting these analysis of variance tests, a number of factors need to be recognised. Firstly, because of the way the data set was generated from the EMEP ozone model, the distribution of data points is unevenly spread and it is most likely that there is a complex correlation between the error terms. Consequently, the estimate of variance is biased and all test statistics derived from this estimate will also be biased. Secondly, the

loess model is a rather complex package and there is no easy way to combine it with the usual extensions and tests of a simpler regression model. For example, it would be very complicated to go from a least-squares to a general least-squares approach with the *loess* package. Thirdly, the focus of this study is on the quality of fit rather than on parameter estimates. It can be seen from Table 1 that the parameters describing the quality of fit - the residual standard error and the multiple R^2 value - show minor differences between the non-parametric and simpler parametric model.

As a result, further investigation of this approach will also concentrate on polynomial regression models, which can overcome the problems identified above. The solutions provided by the polynomial regression technique can then be fed back into a local regression model to test again the effects of changing the span parameter.

3.2.3 Comparison of Statistical Model with Original Data

An example plot of the residuals found in fitting a regression surface to the entire Schauinsland data set was shown in Figure 9. In order to demonstrate how well such a model might perform in practice, this section presents comparisons between regression model predictions of ozone concentrations at Schauinsland for the base case, with no emission reductions, and the corresponding EMEP model results. For all the results given in this section the ozone concentrations are those appropriate to air mass trajectories arriving at Schauinsland at 1200 GMT. The regression model used here is the one described in the previous section based on three predictors: "effective" NO_x emissions, "effective" VOC emissions and mean NO_2 photolysis rate along the trajectory. The regression surface was specified to be locally quadratic, with the "effective" VOC emissions designated as conditionally parametric, and 0.5 was the value adopted for the smoothing parameter.

The cumulative frequency distributions of the base case results from the EMEP model and the regression model are compared in Figure 11. At the very lowest ozone concentrations the regression model has a tendency to overpredict but this is likely to be unimportant in the context of an integrated assessment model designed to assess the cumulative exposure to ozone above a threshold value. Throughout the rest of the data range, the EMEP and regression model distributions are seen to be remarkably similar. The mean value of the regression model results is 61.7 ppb, which exceeds the EMEP mean by 1.2 ppb. The regression model median value is 58.6 ppb compared to the EMEP model value of 56.2 ppb.

A scatter plot of the same data is shown in Figure 12(a). This figure also includes the 1:1 correspondence line and a least-squares linear regression fit to the data. Generally, the agreement is rather good but the plot also shows the existence of a number of outliers where the regression model predicts ozone concentrations of about 80 ppb for EMEP model values of 60 ppb or less.

Figure 12(b) presents these data in the form of a quantile - quantile plot, in which the sets of data are first ordered from lowest to highest before being plotted against each other. The closeness of the points to the 1:1 correspondence line provides further evidence that the regression model predictions and EMEP model results belong to the same distribution.
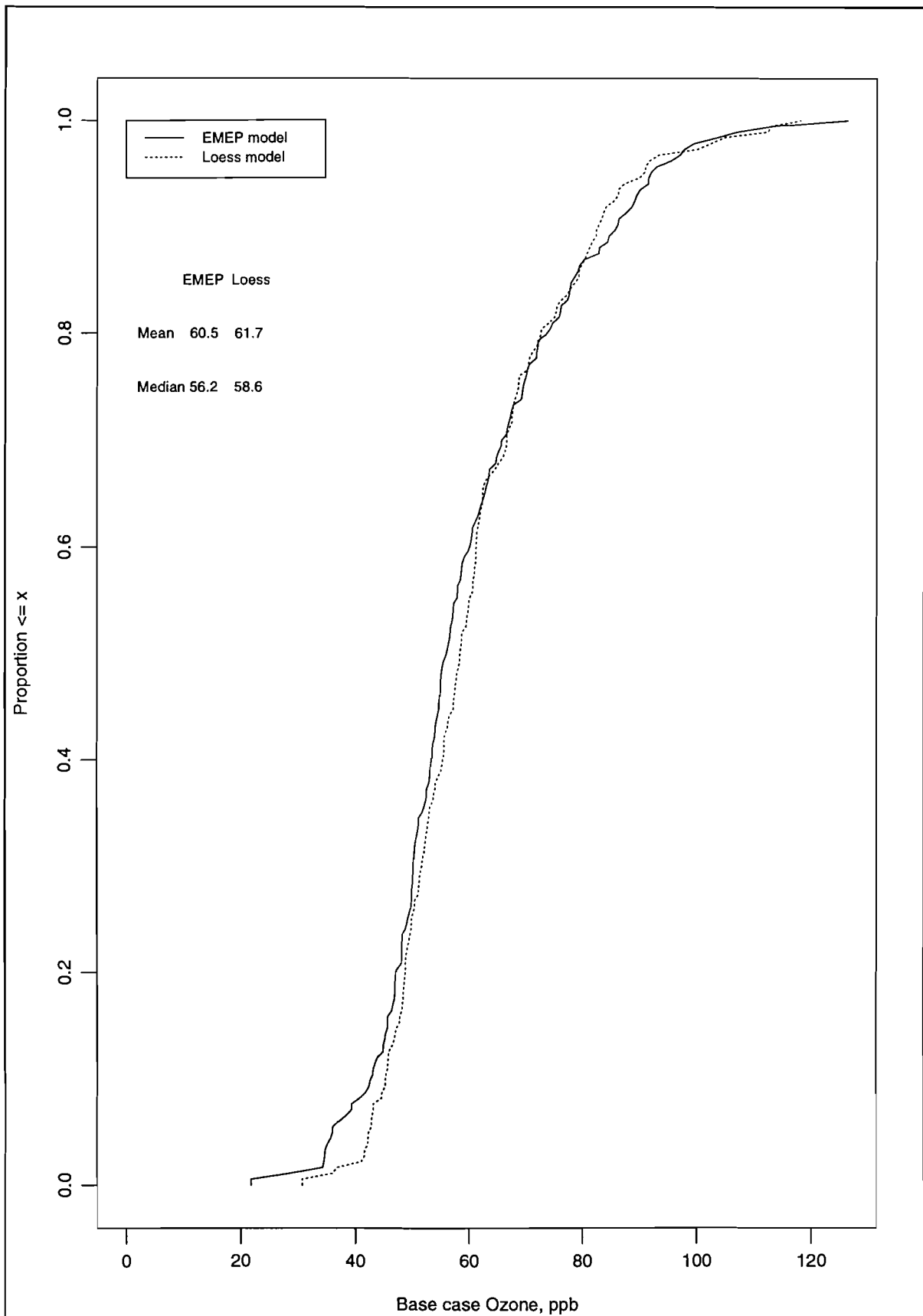
Figure 11 Comparison of EMEP Model and Regression Model Results at Schauinsland for 1989 Base Case - Cumulative Frequency Distributions

(a) Scatter Plot

Loess base case Ozone, ppb (y-axis)
EMEP base case Ozone, ppb (x-axis)

(b) Quantile-quantile Plot

Loess base case Ozone, ppb (y-axis)
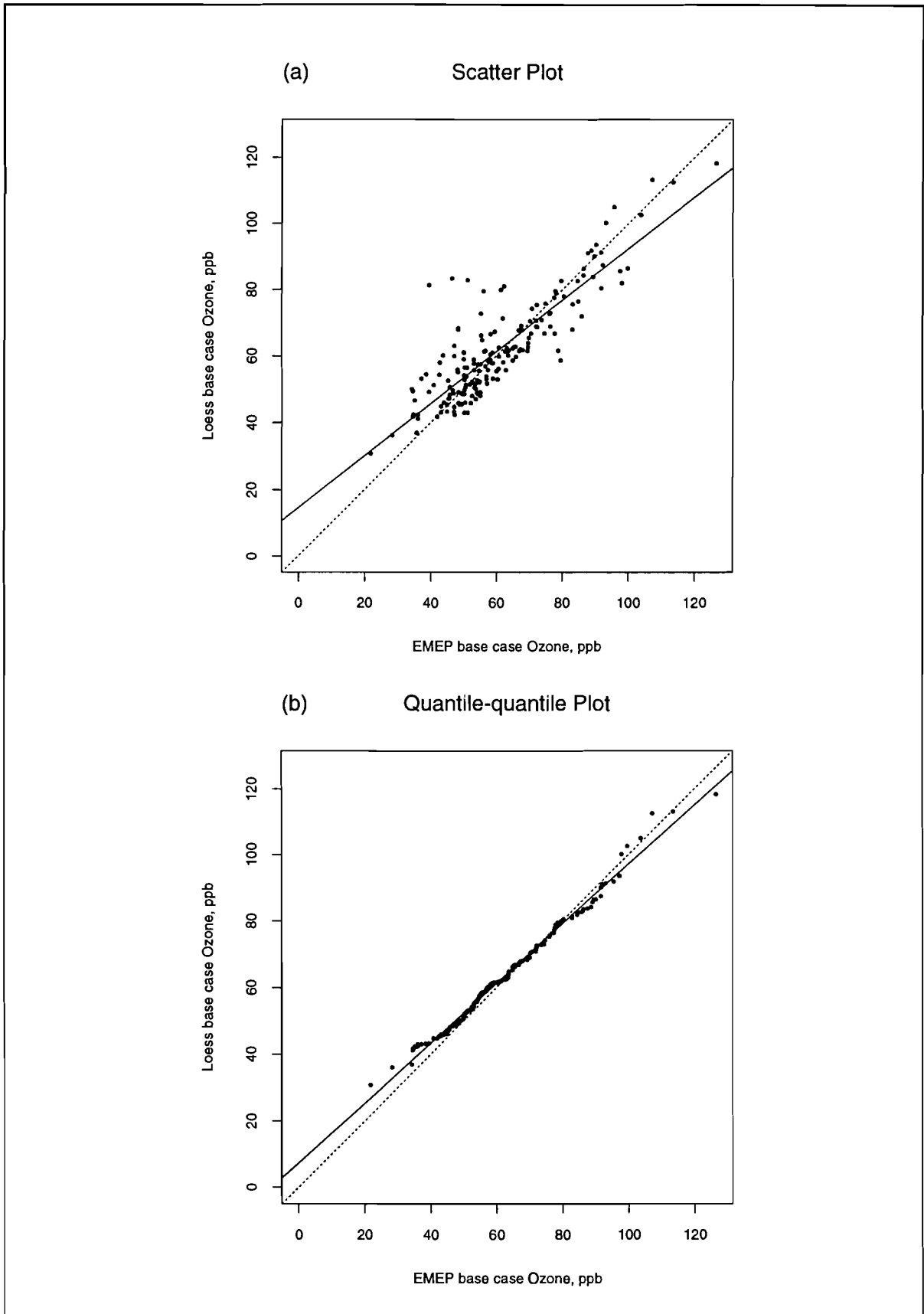EMEP base case Ozone, ppb (x-axis)

Figure 12   Comparison of EMEP Model and Regression Model Results at Schauinsland for 1989 Base Case - Scatter Plot and Quantile-Quantile Plot

The differences between the EMEP model results and the regression model predictions are shown day by day in Figure 13(a). In this way it is possible to identify when the largest regression model overpredictions, evident in Figure 12(a), actually occurred. It can be seen from Figure 13(a) that there are several groups of consecutive days during which the EMEP-loess differences lay in the same direction, suggesting that the differences may arise from particular wind directions or other meteorological conditions which remain similar over a number of days. This feature has not yet been fully investigated but it is evident that several, although not all, of the largest regression model overpredictions are associated with trajectories that have passed over the grid square having the greatest NO_x emission density within the EMEP domain.

It is also of considerable interest to compare the regression model predictions with measurements of ozone concentration made at the Schauinsland site. In making this comparison it must be remembered that the regression model is derived from EMEP model results that are calculated to be representative of the atmospheric boundary layer as a whole, whereas the measurements are made at a height of, typically, 2m. This comparison depends upon the validity of the assumption that, at the relevant time of day, the boundary layer is sufficiently well-mixed that ground-level ozone concentrations are not significantly different from the boundary-layer average. At a relatively high altitude site such as Schauinsland, this is expected to be a reasonable assumption, especially during the afternoon period.

The comparison between Schauinsland measurements of ozone during the summer of 1989 and the regression model predictions is shown in Figure 13(b). It should be noted that gaps in the figure, or apparent differences of zero, are generally due to missing measurement data rather than perfect agreement. Clearly, there are larger differences between the regression model and the measured data than were apparent between the regression model and the original EMEP calculations.

The time series of ozone measurements, EMEP model calculations and regression model predictions for Schauinsland during the 1989 summer period are presented together in Figure 14. The general pattern of measured ozone is reproduced reasonably well by the models, although some larger differences occur on individual days, as also shown in Figure 13(b) for the regression model. The differences between the model predictions and the measurements are summarised in Table 2.

Table 2 Summary of Differences between Ozone Measurements and Model Calculations

Model	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum	Variance	Median Absolute Difference
EMEP	-43.4	-11.8	-2.9	-2.6	5.9	43.9	244.8	9.6
Loess	-39.0	-11.5	-0.1	-0.9	8.4	50.9	260.6	9.6

It is evident that the differences between the regression model predictions and ozone measurements are generally of similar magnitude to the differences between EMEP model calculations and measured ozone concentrations at Schauinsland.
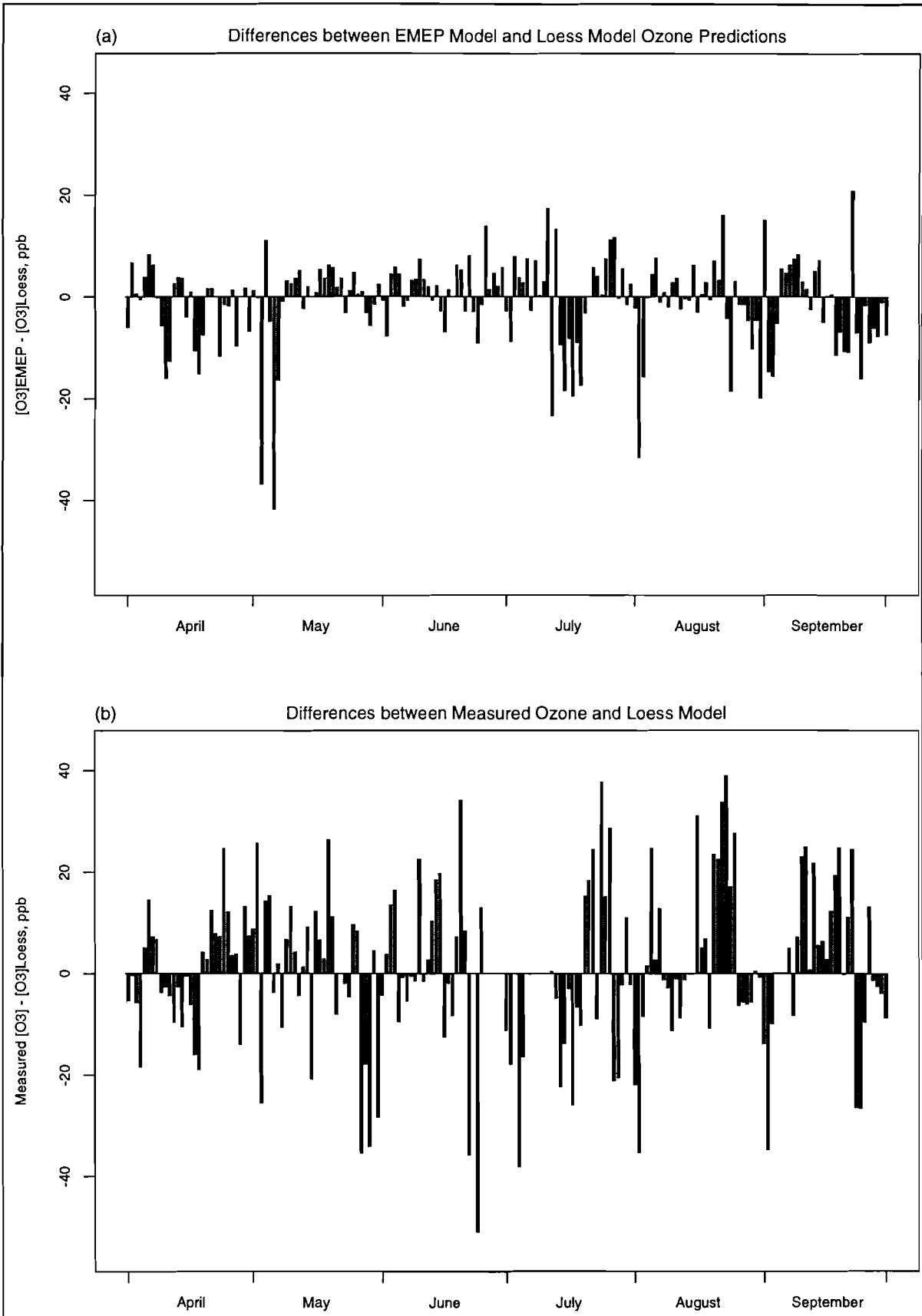
31

Figure 13   (a) Comparison of EMEP and Regression Model Results at Schauinsland
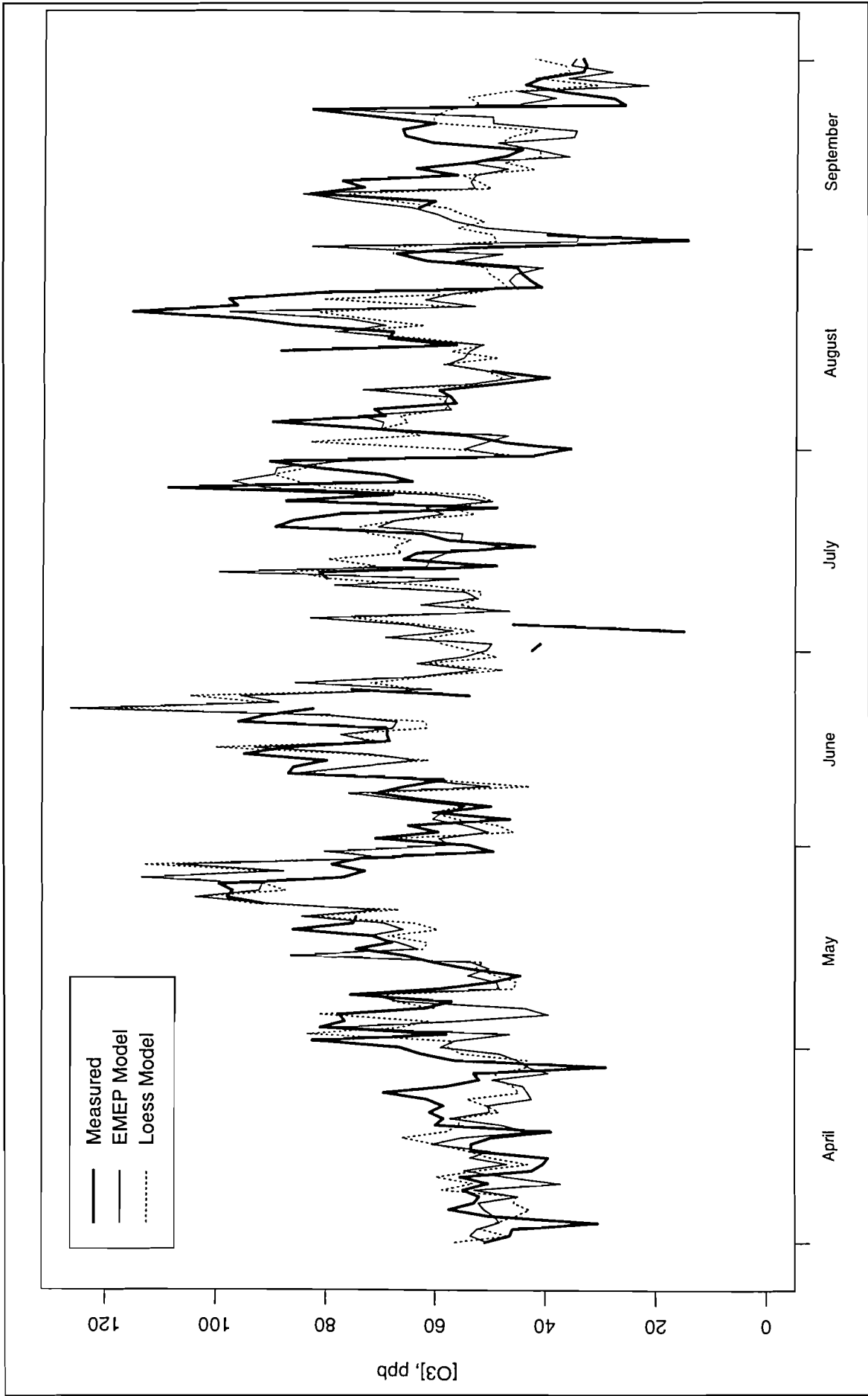   (b) Comparison of Ozone Measurements and Regression Model Results

32

Figure 14   Time Series of Measured and Modelled Ozone Concentrations, Schauinsland 1989

3.2.4 Potential Application in an Integrated Assessment Model

The results obtained with a regression model for the Schauinsland site are sufficiently promising that the application of such an approach within an integrated assessment model seems to have the potential for success. Of course, the assessment of this approach has so far been restricted largely to one site, with the regression surface derived from data relating to the six-month summer period from just one year. More extensive testing of the method, at sites covering a wide variety of ozone climates and with data from several years, will obviously be necessary before practical application of the regression surface approach can be fully justified.

Other, as yet unresolved, difficulties with the present approach also need to be acknowledged. Firstly, the regression models investigated to date have all been derived from EMEP ozone model calculations for emission control scenarios which, although covering a wide range of possible percentage reductions, have all assumed spatially uniform emission changes. It is not yet clear how well the regression surface approach will deal with a situation in which changes in one country's ozone precursor emissions are substantially different from those in a neighbouring country. In the context of an integrated assessment model, this is a potentially serious drawback which remains to be fully evaluated.

Secondly, the regression method, as currently formulated, takes little account of the age of the emissions at the arrival point, in other words, where along a trajectory the emissions occur. The regression model accounts for dilution of the precursor emissions due to mixing with the free troposphere above the boundary layer but does not explicitly consider the time available for the emissions to undergo chemical and photolytic reactions before the air mass reaches its destination. This may be an important factor as high NO_x emissions might depress ozone formation locally but, nevertheless, lead to enhanced ozone production further downwind. Further work is required to assess the significance of this factor.

Thirdly, because the regression approach uses an aggregated, not speciated, VOC emissions parameter and because it takes no account of the age of the emissions, the method is insensitive to potential modifications of the VOC species profile. Preliminary investigations of the effect of changes in the VOC speciation on the results of the EMEP ozone model have suggested that this effect is probably of minor importance. However, the lack of sensitivity to the VOC species profile must still be considered a potential disadvantage of the regression surface approach in comparison to any alternative method able to take account of such changes.

Clearly, the work of developing a regression surface to describe ozone formation cannot be regarded as complete. Nevertheless, it is worthwhile to consider how such a regression surface might be used in practice in an integrated assessment model.

Within the integrated model, the first part of the ozone formation module is required to relate emissions of NO_x and VOC, specified by country of origin, to ozone concentrations in one or more grid squares. The regression model, as formulated here, uses daily, trajectory-based data to do this and, therefore, needs a database of relevant trajectory information, which remains unaffected by emission control scenarios, as well as the ozone precursor emissions from each country. For each trajectory, this approach requires information on the countries traversed by the air parcel, their NO_x and VOC emissions, emissions of NO_x from soils and natural emissions of VOCs along the trajectory, and the mean NO_2 photolysis rate. Table 3 summarises the data requirements, differentiating between the factors for which data

are required per country and those for which a single trajectory-integrated value per trajectory (or day) will suffice.

Table 3 Data Required by Regression Surface Approach

For each country traversed by the trajectory	Trajectory-integrated
Countries contributing to emissions Effective man-made NO_x emissions by country Effective man-made VOC emissions by country	Effective soil NO_x emissions Effective biogenic VOC emissions Mean NO_2 photolysis rate

All the above data are obtainable from calculations using the EMEP ozone model for the base case without emission controls. For each receptor grid, there will be 183 trajectories to deal with for each summer period considered. Thus, for the trajectory-integrated parameters, 183 values will need to be stored for each grid-year combination. The EMEP model uses 96-hour back trajectories, with locations calculated, and pollutant emissions identified, at two-hour intervals. The number of countries contributing pollutant emissions to the air mass during the 96 hours will vary from one trajectory to the next. For a trajectory collecting emissions from n different countries, a 3 x n matrix of data will need to be stored. There will be 183 such matrices for each combination of receptor grid and year. In effect, this set of matrices constitutes a daily, country-to-grid, transfer matrix system.

The practical application of this approach is envisaged as follows. Any particular emission control scenario to be examined using the integrated assessment model would generate a new set of NO_x and VOC emissions from the countries of Europe. The ratios between these new emissions and the original, base-case values for each country would be used to calculate revised "effective" emission inputs for each country along relevant trajectories. This enables "effective", dilution-weighted, anthropogenic NO_x and VOC emissions to be derived for each of the required trajectories by summation of the individual country contributions. Combining the dilution-weighted, anthropogenic NO_x emissions with the dilution-weighted soil NO_x emissions parameter gives the "effective" NO_x emissions for the trajectory, as required by the regression model. The second predictor is obtained similarly from the sum of the "effective", anthropogenic and biogenic VOC emissions. The third predictor variable, the trajectory-mean NO_2 photolysis rate, is unaffected by emission abatement strategies and can be used directly. From these three predictors the regression model will provide a prediction of the final, boundary-layer average ozone concentration for each trajectory. In this way, for the receptors and time periods of interest, a distribution of boundary-layer average ozone concentrations may be obtained.

This ozone concentration distribution would be passed, at least conceptually, to the second stage of the ozone formation module (see Fig. 1) so that the accumulated, ground-level exposure to ozone concentrations above a given threshold value may be estimated. The subsequent comparison with critical exposure levels for ozone and assessment of the costs and benefits of the emission abatement strategy, which are important topics in themselves, are beyond the scope of this report.

4. CONCLUSIONS

The development of an integrated assessment model for tropospheric ozone requires a simplified yet reliable description of ozone formation in order to represent the source-receptor relationships involved. This paper has demonstrated how such an ozone formation module can be constructed using a regression surface to summarise, in effect, the results obtained from numerous calculations performed by the EMEP ozone model.

In this statistical approach, the ozone concentration at a receptor is estimated on a daily basis from the values of a number of predictor variables calculated over the four days prior to the arrival of the relevant air mass at the receptor. Three predictors were found to provide an adequate estimate of ozone; these are the mean emissions of each of the ozone precursors, NO_x and VOCs, and the mean photolysis rate of NO_2. Both the emissions parameters are weighted to allow for the effects of dilution due to mixing processes with the free troposphere.

In formulating the local regression model, the most convincing fit was obtained when locally quadratic interpolation was specified and the relationship between ozone and VOC emissions was forced to remain parametric throughout the data range. Such a model has been assessed most thoroughly at the Schauinsland site in SW Germany, for which the residual variance of the model was estimated to be 7.3 ppb. Ozone concentrations predicted by the regression model for the current situation without further emission controls compared well with the original EMEP model results. The level of agreement between the regression model results and measured ozone concentrations at this site was similar to that between the EMEP ozone model and measured data.

This statistical description of tropospheric ozone production takes into account many of the important factors involved - emissions of precursor species, solar radiation and the effects of changes to the atmospheric boundary layer - but does not explicitly include other parameters, such as temperature, known to influence ozone formation. This approach has the advantages of relative simplicity and, therefore, speed of operation, essential requirements in the context of an integrated model designed to assess many different scenarios. In addition, because the method described here provides daily ozone concentrations, it has sufficient flexibility to cover the necessary range of ozone exposure assessment periods resulting from different growing seasons across Europe.

Potential disadvantages of the method are its lack of sensitivity to changes in the VOC species profile and that it fails to recognise the potential importance of the age of the emissions at the arrival point. This latter factor may be particularly important in cases where the NO_x emissions are high, when the NO_x might depress ozone formation locally but, nevertheless, lead to enhanced ozone production further downwind.

Clearly, further work is required before a validated ozone formation module will be sufficiently developed for the integrated assessment model. It will be necessary to assess the significance of the potential disadvantages of the statistical approach identified above. More extensive testing of the method, at sites covering a wide variety of ozone climates and with data from several years, is also required. Finally, this statistical approach is just one potential method of formulating a simplified description of ozone formation. Other methods might be even simpler, if based, for example, on the sort of ozone isopleth diagrams presented in Figure 4, or they might be based more on a chemical description of ozone formation. The

advantages and disadvantages of these alternatives should be compared with those of the statistical approach when fully developed.

REFERENCES

Alcamo, J., Shaw. R. and Hordijk, L., Eds. (1991) *The RAINS Model of Acidification.* Kluwer, Amsterdam.

Atkinson, R. (1990) *Gas-phase tropospheric chemistry of organic compounds: A review.* Atmos. Environ., 24A, 1-41.

Baldi, M. and Calori, G. (1994) *Tropospheric Ozone: A data base for Europe and statistical analysis.* WP-94-xxx. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Builtjes, P.J.H., Simpson, D., Zlatev, Z., Christensen, J. and Frohn, T.P. (1991) *Comparison of three models for long term photochemical oxidants in Europe.* ed. by T. Iversen. EMEP MSC-W Report 3/91. MSC-W, Norwegian Meteorological Institute, Oslo, Norway.

Caliandro, B. (1994) *Control Technology Options and Costs for Reducing Volatile Organic Compounds.* WP-94-80. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Chambers, J.M. and Hastie, T.J., eds. (1992) *Statistical Models in S.* Wadsworth & Brooks/Cole, Pacific Grove, California.

Cleveland, W.S. and Grosse, E. (1991) *Computational Methods for Local Regression.* Statistics and Computing, 1.

Derwent, R.G. (1993) *Evaluation of the chemical mechanism employed in the EMEP photochemical oxidant model.* Atmos. Environ., 27A, 277-279.

Fuhrer, J. and Achermann, B., eds. (1994) *Critical Levels for Ozone, a UN-ECE workshop report.* Swiss Federal Research Station for Agricultural Chemistry and Environmental Hygiene, Liebefeld-Bern, Switzerland.

Kettunen A., Schöpp, W. and Klimont, Z. (1994) *Statistical Analysis of Tropospheric Ozone Concentrations.* WP-94-88. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Labancz, K. (1993) *Evaluation of the EMEP MSC-W oxidant model: comparison with measurements.* In Anttila, P. (ed.), *EMEP workshop on the control of photochemical oxidants in Europe.* Finnish Meteorological Institute, Helsinki, Finland.

National Research Council (1991) *Rethinking the Ozone Problem in Urban and Regional Air Pollution.* National Academy Press, Washington, DC, U.S.A.

Olsthoorn X. (1994) *Towards an Integrated Assessment Model for Tropospheric Ozone. Emission Inventories, Scenarios and Emission Control Options.* WP-94-27. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Simpson, D. (1992a) *Long period modelling of photochemical oxidants in Europe. Calculations for July 1985.* Atmos. Environ., 26A, 1609-1634.

Simpson D. (1992b) *Long Period Modelling of Photochemical Oxidants in Europe: A) Hydrocarbon reactivity and ozone formation in Europe; B) On the linearity of country-to-country ozone calculations in Europe.* EMEP MSC-W Note 1/92. MSC-W, Norwegian Meteorological Institute, Oslo, Norway.

Simpson, D. (1993) *Photochemical model calculations over Europe for two extended summer periods: 1985 and 1989. Model results and comparisons with observations.* Atmos. Environ., 27A, 921-943.

Simpson, D. (1994a) *Biogenic VOC Emissions in Europe. Part 1: Emissions and Uncertainties.* EMEP MSC-W Note 5/94, Norwegian Meteorological Institute, Oslo, Norway.

Simpson D. (1994b) *Biogenic VOC in Europe. Part II: Implications for Ozone Control Strategies.* Submitted to J. Geophys. Res.

Simpson, D., Andersson-Sköld, Y. and Jenkin, M. E. (1993) *Updating the chemical scheme for the EMEP MSC-W oxidant model: current status.* EMEP MSC-W Note 2/93, Norwegian Meteorological Institute, Oslo, Norway.

APPENDIX I. Calculation of Dilution-Weighted Emissions

The EMEP ozone model simulates the production of ozone from its NO_x and VOC precursors along an air parcel trajectory over a four-day period. It has been shown recently (Simpson, 1994b) that, at least for some receptor grids, dilution of the air parcel with free tropospheric air may mean that rather little of the precursor emissions and ozone from the early stages of the air parcel's trajectory will survive to reach the trajectory destination. In such a case, the early precursor emissions may not have much influence on the final ozone concentration and analysis of ozone in terms of trajectory-integrated emissions may be misleading. Simpson (1994b) has proposed a method, adopted in this study, of weighting trajectory-related variables to take account of the dilution processes. The following description of the procedure is based upon Appendix B of Simpson (1994b).

In the EMEP model, the emissions and meteorological input data are revised at two-hour intervals, so that there are 49 time steps during the four-day trajectory. Two processes are included in the model which lead to mixing of boundary layer air parcels with free-tropospheric air: the venting effect of cumulus clouds and day-to-day increases in mixing height. These exchange mechanisms operate at two-hour intervals, with chemical reactions calculated within each 2-hour time step.

If the emissions of an ozone precursor during time step i are denoted by E_i, and the exchange processes result in a dilution of boundary layer air by a factor f_i $(0 < f \le 1)$, the contribution from time step i to the trajectory-integrated value of the precursor emissions, E, at time step $(i + 1)$ is given simply by:

$$E_i \times f_i$$

Subsequent mixing events would further reduce the contribution of E_i, so that the contribution of time step i to the final trajectory-integrated value would be:

$$E_i \times (f_i \times f_{i+1} \times f_{i+2} \times f_{i+3} \times \ldots \times f_{49})$$

The integrated contributions from all 49 time steps, denoted by <E>, are given by:

$$\langle E \rangle = \sum_{i=1}^{i=49} E_i \times \prod_{j=i}^{j=49} f_j$$

Such quantities were calculated for both NO_x and VOC emissions along each trajectory and were investigated as predictor variables in regression models. In the main text of this report, they are referred to as dilution-weighted or "effective" emissions.