

We thank the two anonymous referees for their careful reading of our manuscript and constructive and pertinent comments, which have undoubtedly helped to improve the quality of our work.

As a general note, we are aware that the original LF paper (Wind et al., 2020) presented the method as a way of tracking source contributions, and not as a sensitivity model. This can certainly be a cause of confusion. However, for linear species, we would argue that there is no real fundamental difference. The “generalization” presented here, instead concerns mostly the range of applicability. Compared to the first LF paper from 2020, the methodology essentially only adds the module that propagates the local fractions through the chemistry transformations. We also change units in the outputs. Where the original local fractions were dimensionless (being treated as fractions), we here multiply by the total concentration to get a concentration unit. The fundamental methodology and computer code (outside of the calculation of chemical derivatives in the chemistry module) is unchanged compared to the 2020 paper.

While not strictly necessary for the reply to reviewers, we would like to give some additional background information by elaborating on the original 2020 submission, and how this reflects on any possible confusion with regards to the LFs being a source apportionment or sensitivity methodology. In our first submission of the 2020 paper, we in fact tried to present the method as a sensitivity method, but this was not accepted by the referees (and eventually reformulated), which might be a reason for the final form to be somewhat misleading (with the original LFs being presented as a way to determine the source contributions, i.e. source apportionment, of different sources to primary particulate matter). Here is an example of one of our replies to reviewer comments from the original LF paper:

“In the future we plan to generalize the method to include chemical processes. The Local Fractions could then give information about sensitivities to changes in emissions without necessarily summing up to one hundred percent.”

Referee comments for the first paper also noted the following:

“It seems to me that the local fractions deliver information about contributions, not sensitivities. Page 16, line 14: Why wouldn’t the local fractions add up to 100%, and why isn’t this a problem? It seems that the final sentence of the manuscript creates all sorts of problems for the interested reader. The authors could consider simply deleting this sentence.”

For the current work, we also would like to stress that we aimed for this to be a technical paper. To this end, we want to document the mathematical and computational foundations of a new tool. This new tool has already been used in important applications (some of them briefly mentioned in the discussion), but each new application requires detailed discussions that we do not think fit in this methodological description paper, and which will instead be the subject of separate future publications. With this being said, we now proceed to address the comments from reviewer 1 and 2 step-by-step. Each of the comments is repeated, while our replies are highlighted in blue.

RC1: ‘Comment on egusphere-2024-3571’, Anonymous Referee #1, 31 Mar 2025

The authors are providing a description of how the original local fractions method has been extended making it suitable for application to pollutants influenced by non-linear processes. While the technique is quite complicated and requires some brain crunching to grasp its approach, the authors have done a good job at taking the reader by the hand through two illustrative examples, allowing easier understanding of the methodology.

The paper thereby can be considered as a good reference for a novel methodology, which provides opportunities to advance in the science of emission change sensitivities. The paper is well written and structured.

General comments:

- Since it seems the method is developed to be applied for source attribution applications (as illustrated in the applications mentioned in the conclusions some more attention should be paid to the applicability of the method for such applications. Explain a bit further how this method could be used to investigate potential impacts of 15% emission reductions, which is a common emissions reduction used for policy studies. The reader

50 is now referred to an EMEP report for the comparison with brute force at 15% emission reduction, but it would be good to include the conclusions of that work in this paper as well focusing on potential differences and possible implication for the interpretation of LF results.

55 We understood this comment to effectively come down to the question of how large the differences due to non-linearities are between the 1 % reductions employed by the LF methodology, and the 15 % reductions typically employed by the BF method. In part since the BF method is sometimes used to calculate (or estimate) source attribution, even though this is not the aim of the LF method as it is presented in the current work. To provide additional context, we now refer to earlier investigations of the impact of non-linearities by adding to Sect. 4:

60 "In particular, in EMEP Status Report 1/2024 (2024) country source-receptor matrices for yearly averages using the BF and LF methods have been compared for several pollutants and indicators, including O<sub>3</sub> and PM<sub>2.5</sub>. The differences are overall less than 10 %, and can be considered as small, since they are of the same order of magnitude as methodological differences (advection scheme and filtering) shown in Fig. 1 and Fig. 2 in the following. The differences due to non-linearities introduced by reducing the emissions by 15 % in the BF method, compared to, in principle, infinitesimally small reductions in the LF methodology, represent only a small fraction of methodological differences. We do stress that the 15 % emission reduction employed by the BF method is in principle arbitrary (EMEP Status Report 1/2004, chapter 4, 2004), and that the differences due to non-linearities with the LFs do not represent an actual source of methodological error."

70 In several places in the text it is mentioned that because of the small perturbations linear assumptions can be made, what does this mean for application to investigate the impact of potential policy measures.

Also the LF results are presented as contributions, i.e. transforming the sensitivity to a very small emission change to a 100% contribution, please explain what is the validity and limitations of this, for example are the contributions adding up to the total concentration? This is related to the interpretation of the results for policy makers.

80 In effect, the interpretation of the assumptions is that the LFs give accurate results when the (policy relevant) emission reductions are very small. For emission changes that are larger, the method as it is presented in the current work might not be entirely accurate, because then non-linear effects may become important. However, for such applications, multiple LF simulations could be performed to calculate the actual non-linear dependence on emission changes. But as alluded to in the conclusion section, such applications will be the subject of a future publication. Nevertheless, to add further interpretation to the LF methodology, we have added to the introduction section:

85 "The methodology is a sensitivity method. It does not directly attempt to determine the impact of any finite change of emissions. Neither does it assign the relative total contributions from different sources (in contrast to other tagging methods (e.g., Butler et al., 2018; Emmons et al., 2012; Dunker et al., 2002; Kwok et al., 2015; Grewe, 2013; Grewe et al., 2017; Wang et al., 1998)). Total contributions can only be directly inferred if the species are considered to have a linear dependency on emissions, which is not the case in general. Still, due to the lower computational cost, non-linear responses can be inferred by performing sensitivity analysis at several emission levels, thus providing indirectly the effect of non linear changes (EMEP Status Report 1/2024, 2024,  
90 "

and to Sect. 2.2:

95 "It is important to keep in mind that even if the results are presented in units of concentrations for a 100% change of emissions, they can not be interpreted as total contributions for non-linear species. The values must be interpreted as sensitivities to small emission changes. Those sensitivities can be both positive or negative, and will in general not sum up to total

concentrations.”

- 100 - Throughout the paper make sure to consistently use the word generalized in combination with local fractions wherever appropriate to distinguish between the original local fractions method and the newly developed generalized version.

105 As explained in the introduction to this author reply document, we do not consider that there is a fundamental difference: The “generalization” presented here, does essentially only add the module that propagates the local fractions through the chemistry transformations (and also a change in units in the outputs; where the original local fractions were dimensionless, we here multiply by the total concentration to get a concentration unit). To highlight that the generalized LFs and original LFs are fundamentally interchangeable, we have added “The generalized LF methodology (hereafter interchangeably referred to as “LF”)...” to the introduction section.

110 Specific comments:

- Introduction: explain why it is fundamental to relate the air pollution to emission sources (i.e. for policy), add some background on how this is normally done (current practice broader than BF) and how LF comes in.

115 To provide additional context, we have added to the introduction: “The information about the relative impact from different countries (and possibly emission sectors) can then be used to propose an optimized set of abatement measures, as is done for example using the GAINS (Greenhouse gas – Air pollution Interactions and Synergies) model (Amann et al., 2011)”

- 120 - Introduction page 1, mention also the need for tracking source sectors besides source regions.

125 The sources tracked using the LFs are effectively (a collection of) sectors from any particular source region, based on how emission input files are typically formatted. To highlight that we can track sectors from any specific source region, we have added “In this work the word “source” is defined as a set of emitted species from a predefined region and from any individual or combination of emission sector(s).”

- Introduction line 25-26, mention interaction between sources

130 We have added: “The species involved will also usually originate from different emission sources.”

- Introduction line 28 – chemical species, make it clear that this concerns non-inert chemical species

135 rewritten as “sensitivity (i.e., rate of change) of such active chemical species to source emission changes”

- Sections 1 and 2, as mentioned in the general comments put the method in the context of other methods, what is the range of applicability, explain additivity limitations or limitations with respect to the range of emission changes for which the method can be applied. What happens if there is a limited chemical regime and an emission change does not have any impact? This question relates mainly to applicability for single source sectors or pollutants.

145 To add depth to the interpretation of the calculations, we have added (as discussed above): “It is important to keep in mind that even if the results are presented in units of concentrations for a 100% change of emissions, they can not be interpreted as total contributions for non-linear species. The values must be interpreted as sensitivities to small emission changes. Those

150 sensitivities can be both positive or negative, and will in general not sum up to total concentrations.” If there is a limited chemical regime where an emission change will not have any impact, the calculated LFs will also simply be zero. In reply to other comments, we have added elsewhere in the text a note that the LFs can be either positive or negative (e.g., positive when reducing NO<sub>x</sub> emissions leads to an increase in O<sub>3</sub> due to reduced titration effects). We hope that this also further clarifies that the sensitivities may be zero in certain regimes.

155 - Page 3 line 66, what do you mean with second order effects, which ones? Again here you focus on small perturbations but how would you apply this method when the policy maker is interested in impact of large emission changes?

The method as is described in the current work does not directly address this (policy) question. However, the text added in reply to an above comment (repeated here), gives a more explicit explanation:

160 “The methodology is a sensitivity method. It does not directly attempt to determine the impact of any finite change of emissions. Neither does it assign the relative total contributions from different sources (in contrast to other tagging methods (e.g., Butler et al., 2018; Emmons et al., 2012; Dunker et al., 2002; Kwok et al., 2015; Grewe, 2013; Grewe et al., 2017; Wang et al., 1998)). Total contributions can only be directly inferred if the species are considered to have a linear dependency on emissions, which is not the case in general. Still, due to the lower computational cost, non-linear responses can be inferred by performing sensitivity analysis at several emission levels, thus providing indirectly the effect of non linear changes (EMEP Status Report 1/2024, 2024, ”

170 - Page 3 – lines 70-73, What is the difference with the linear adjoint model? What is new to this generalized LF, why not use the existing linear adjoint model?

To highlight the differences and to provide additional context, we have added: “To our knowledge, the usage of Tangent Linear Models (TLM) has primarily been to compute backward trajectories (emission inversions (Zheng et al., 2024)), while the adjoint of TLMs is also applied in the context of data assimilation (Shankar Rao, 2007). We will here show how a TLM can be used in a ‘forward mode’ to compute sensitivities to emission sources in the form of, for example, source-receptor matrices.”

175 - Page 3 – line 76, can it also be a source sector?

180 Indeed it can, and as described above, we have added “(from a certain sector), or in principle any other sub-class of a pollutant)” to the introduction text.

- Page 5 equation 8, explain a bit further, especially the minus sign

185 For clarity, in the revised manuscript the equations have been separated into two, one for the minus sign (new eq. 8) and one for calculation of deposition sensitivities (new eq. 9). An extra line of interpretation is also added.

- Page 5 line 145, ‘however the range of validity is now limited’, please explain what that means in practice.

190 As highlighted in the additions to the text described above, we now emphasize more that the method relates to the impact of very small emission changes, because then the calculated responses are effectively linear. For larger emission changes, the response will in general be non-linear, therefore being beyond the range of validity of the linear assumption. To further highlight this, we have added: “and the calculated deposition sensitivities may no longer be representative.”

- Page 6 line 147, ‘this may not be the case anymore’, what does that mean for the

195 applicability of the method?

As for the previous comment, this statement reflects that for larger perturbations the response may no longer be linear, thus making the method unsuited for calculating the impact of large emission changes for (highly) non-linear species. We hope to have addressed this comment also with the reply given to the previous comment.

200 - Page 6 line 154, general equations: what do you mean with general? do you mean the equations for the entire method, not only chemistry? Do you show the SIA chemical transformations as example for other chemistry?

205 We have added that ‘general equations’ refers to the sensitivity equations for all species and relevant model processes. In addition, a more descriptive explanation of the general equation has been added to Sect. 2.4.2: “Eq. 25 is a general equation that describes how to update the sensitivities for any transformation. For example in the (linear) deposition case, the Jacobian is simply equal to  $1 - v_i$ . In the advection case (below) the indices of the concentrations would refer to pollutants at different position in space and the Jacobian will be equal to the fluxes.”

210 - Page 6 line 175, ‘those are’ I guess you are referring to  $a_{ij}$  with the word those? Or to the environment, .....? So do you mean that for example the change in NO<sub>3</sub> because of a change in SO<sub>4</sub> is assumed constant?

215 Indeed this is what is meant. For added clarity, we have reformulated to “ $a_i^j$  are considered constants during the time step, being independent of the scenario  $k$ .”

- Page 7 lines 197-199, this is a difficult part, for me hard to follow, explain the small  $c$  versus capital  $C$

220 The lower case is now defined in Eq. 1 as “ $C_{i,k}$  represents the concentration of pollutant  $C_i$  from source  $k$  and we use the lower case  $c_{i,k}$  to define the “dimensionless concentration”, or the proportion of the pollutant that originate from source  $k$ ”, analogous to the dimensionless fractional contributions of the original LFs.

225 - Page 7 lines 190-199, it made it easier for me to follow when I filled in example species, maybe an extra sentence can be added giving the example of delta SO<sub>4</sub> as a result of delta NO<sub>3</sub>.....

230 While we agree that such an addition would be didactic, we also feel that it might distract from the overall flow of the text. In our view it works in the favor of the text to keep the methodological description overall analytical.

- Page 8 line 227-228 I do not understand this sentence

235 Reformulated as “Note that  $\epsilon$  in Eq. 4 is assumed small, however  $\Delta t$  or the changes in Eq. 24 are not assumed to be small. In the general case, and in the model code, the function  $f$  is not assumed to be linear, and Eq. 25 is valid also for non-infinitesimal  $\Delta t$ . For example, in the chemical solver  $\Delta t$  is divided into micro-iterations, each time step capturing the full non-linear chemistry.” We hope that this better supports the final sentence of the above quote, which was also meant to serve as a further explanation of the original (unclear) sentence referred to by the referee.

240 - Page 9 line 233 – what are those sets? You mentioned before 3 sets for SIA, but what would be the others, can you give some additional examples?

245 To clarify that sets here meant to refer to species, this line has been rewritten as “In our present implementation, 60 species will have a direct or indirect effect on other chemical species in the chemistry module. This implies that the function  $f$  must be evaluated for a perturbation of each of those 60 species. In addition perturbations for emissions must be performed, but usually only a few sources will contribute for a given grid cell (typically one country, four species, and only for the lowest seven vertical levels).”

250 - Page 9 line 244 how are the fluxes dependent on the five nearest cells?

255 The exact dependence of the fluxes on concentrations in the five nearest grid cells is described in detail in the cited work of Bott (1989a, b). However, a more precise description of this dependence is difficult to explain without going more into the details of the scheme. We think such a description would distract from the current work, as the exact nature of the scheme is not necessary for the discussion in Sect. 2.5. In the context of the LFs and the equations described in this section, all that really matters is that a relationship exists, no matter the exact details of this relationship.

- Lines 246-247: these are nearly a duplicate of lines 244-245

260 Yes, removed!

- Lines 248-250, isn't a change in the advection rate also physical?

265 To better distinguish what we consider physical versus non-physical advection effects, we have added “These fluxes are the fraction of air masses transported to a neighboring grid cell and are in reality independent of the pollutant concentrations.”

- Line 276 ‘improve’ you have already put in italics, but isn't it more stable results?

Yes, perhaps, but since we remove a contribution that we know is wrong, we believe we can say that the result is improved.

270 - Line 281-283 how small/large do you expect this impact to be and in which situations can it be expected?

275 It is difficult to quantify as it must be disentangled from other effects, and the effect will vary largely with time, space and chemical regime. However the sum of all the differences is shown to be small in the result section, which we think suffices for making the point that these secondary effects are small.

- Line 292-293 again is expected impact small/large and in which situations can it be larger?

280 See answer to the above referee comment.

285 - Line 318-319 this is not the case for all tagging models, also in those methods they make use ‘smart’ computationally more efficient methods, such as calculating the loss and production rates for the total tracer concentration and then multiplying a followed by a (sparse) matrix multiplication to get the chemical conversions for each tag.

Thanks for this clarification, removed!

290 - Table 1 caption is fairly complicated to understand change title to relative run times generalized LF and speed up in comparison to BF. And move the sentence ‘the reference time....’ to the end by the time required using the LF method (...) where the reference

time is taken .....

295 Clarified by rephrasing to “Relative run times of LF simulations and their speed up in comparison to BF. The time unit is defined as the time to run a single BF scenario, amounting to 29.1 seconds ( $T_{\text{scenario}}$ ) for the model setup described in Sect. 4 (48 hours simulation on a  $400 \times 260 \times 20$  grid). 50-2, 50-4, and 50-5 stands for 50 countries with 2, 4 or 5 different sector contributions (each emitting five pollutants). The results in the table show the time to execute the LF code on eight compute nodes. The speed up is defined as the time it would take to compute all the scenarios if a direct method (BF) had been used, divided by the time required using the LF method ( $\frac{\text{number of scenarios} \cdot T_{\text{scenario}}}{\text{total LF time}}$ ).”

300 - Table 1 total time LF total time and add unit here (reference time) to clarify

Done.

305 - Lines 341-342 is this for generalized LF? Which spatial resolution

To clarify, we have added and rephrased to “In practice (EMEP Status Report 1/2024, 2024), a full SR matrix calculation for 55 countries on a latitude-longitude-altitude grid with dimensions  $400 \times 260 \times 20$  ( $0.2^\circ \times 0.3^\circ$  horizontal resolution), will for a full year LF simulation on 16 compute nodes (running 128 MPI processes on each compute node) take 18 wall time hours.”

310 - Lines 343-344 I think the detailed numbers could be omitted, to me it would be enough to state that in practice the LF run is 10 times faster than corresponding BF simulation, which then agrees with table 1.

315 Done.

- Section 3.4 add a short introduction what this is about, are you looking for options to simplify and obtain a speedup? Maybe change the title to optimization through approximating chemically active species. Implemented and possible simplifications seem to be mixed in this section, make this more clear. Are you trying to say that the method can be sped up if you would look at S, N, .. atoms instead of species such as NO<sub>2</sub>, NH<sub>3</sub>, .....?

325 To better clarify our intended purpose, we have added: “Each term of the Jacobian matrix,  $\frac{\partial f_i}{\partial C_j}$  in Eq. 25, is explicitly evaluated in our present implementation. This might not be necessary, and in this section we will briefly indicate the types of simplifications that might be implemented in the future.”

- Lines 383-388 which emissions are used?

330 Added information that the runs were “employing standard EMEP reported emissions”, but also that “The results can be reproduced using the code and data provided under the “Code and data availability” section.”

- Lines 394-397 the paper would benefit from including the conclusions of the 15% emission reduction investigations since these are relevant to understand the applicability (range) of the method

335 As noted in an earlier reply, to clarify this we have added “In particular, in EMEP Status Report 1/2024 (2024) country source-receptor matrices for yearly averages using the BF and LF methods have been compared for several pollutants and indicators, including O<sub>3</sub> and PM<sub>2.5</sub>. The differences are overall less than 10 %, and can be considered as small, since

340 they are of the same order of magnitude as methodological differences (advection scheme and filtering) shown in Fig. 1 and Fig. 2 in the following. The differences due to non-linearities introduced by reducing the emissions by 15 % in the BF method, compared to, in principle, infinitesimally small reductions in the LF methodology, represent only a small fraction of methodological differences. We do stress that the 15 % emission reduction employed by the BF method is in principle arbitrary (EMEP Status Report 1/2004, chapter 4, 2004), and that the differences due to non-linearities with the LFs do not represent an  
345 actual source of methodological error.”

- Line 430-432 and 437-438 do you mainly see them over sea? Did you check other days, while July seems an appropriate period for ozone chemistry, maybe for SIA another time period should be preferred. And are the differences so small they would not impact in a  
350 policy application?

The instabilities of the thermodynamics equilibrium module are a property of this module and not of the LF method. In the ISORROPIA-Lite equilibrium module (Kakavas et al., 2022) (also sometimes employed by the EMEP model), these instabilities can be an even more pronounced (see for example also Capps et al., 2012a; Miller et al., 2024). This is a problem  
355 also for the BF method, and its effects can be large especially when compared against the impact of small emission perturbations. The LF method gives an opportunity to address this problem through the use of a filter. To clarify this, we have changed the figures to demonstrate that it is the BF method that can give poor results, and that the LF filter can detect those. We do not have the ground truth though, so we cannot show the absolute “error” of the different methods, except for that we know the results to be occasionally poor because of numerical noise and not because of physical processes.

360 - Line 438, what do you mean by these are also present in BF simulations

See answer above.

365 - Lines 468-477 here you make the connection to practical source attribution applications, I believe it is important already earlier in the document to identify the needs for this, how the assumptions and simplifications made will impact this and how to deal with larger emission reductions.

370 As noted in an earlier comment, in the introduction section we have now added “The methodology is a sensitivity method. It does not directly attempt to determine the impact of any finite change of emissions. Neither does it assign the relative total contributions from different sources (in contrast to other tagging methods (e.g., Butler et al., 2018; Emmons et al., 2012; Dunker et al., 2007)). Total contributions can only be directly inferred if the species are considered to have a linear dependency on emissions, which is not the case in general. Still, due to the lower computational cost, non-linear responses can be inferred by performing  
375 sensitivity analysis at several emission levels, thus providing indirectly the effect of non linear changes (EMEP Status Report 1/2024, 2024).”

Technical corrections

- Page 3 line 83 and 85, page 5 line 118 (also in other places in the report), please add the  
380 units of e.g. Ci, E, vi.... Done.  
- Line 248 add an s after emission → emissions Done.  
- Line 262 same fluxes → same base fluxes Done.  
- Line 280 ‘a change of emissions will change’ → may change Done.  
- Line 287-288 move the word however : the impact of this limitation is however.... Done.  
385 - Line 297 a test may → a test is Done.  
- Line 299 procedure may → procedure is; iterations used may → iterations used depends Done.  
- Lines 301, 312, 317, 327, 379, LF approach → generalized LF approach



390 As explained in the first part of this reply, we do not consider that there are two fundamentally different LF approaches: the “generalization” rather reflects the range of applicability.

- Line 356 this represent → this would represent Done

- Line 363 emission → emissions (Line 248?) Done

RC2: 'Comment on egusphere-2024-3571', Anonymous Referee #2, 07 Apr 2025

395 Summary and Overall Evaluation

This manuscript presents an important and methodologically sound extension of the Local Fractions (LF) approach to account for chemical transformations in chemically active pollutants, enabling efficient estimation of source sensitivities in a single simulation. The generalized LF method is implemented in the EMEP MSC-W model and validated against the brute force (BF) method. This is a highly relevant contribution to the field of atmospheric chemistry and chemical transport modelling, particularly in the context of source-receptor relationships and emission control policy support.

400 The paper is well-written, technically solid, and clearly illustrates the method, its implementation, and validation. The generalization of LF to chemically active species represents a significant advancement with substantial computational benefits. I recommend that the paper be accepted with minor revisions.

Major Comments

405 The core methodology is presented in detail and with rigor. However, the paper could benefit from including a flowchart or schematic summarizing how the generalized LF sensitivities are calculated in comparison to the BF sensitivities.

410 To give a clearer picture of how the implementation of the LF calculations functionally works, we have added “The LF modules are additions to the original EMEP model code, but do not affect the original results (e.g., those used for BF calculations). For each module (advection, chemistry, emissions, depositions etc.) a corresponding LF module exists, which at each time step computes the updates to the local fractions, but there is no feedback from the LF modules into the regular concentrations.” to the introduction of Sect. 3.

415 The validation section is appropriate but could be strengthened. The comparison is limited to a 24-hour period and a single country (Germany). While this suffices for proof-of-concept, readers might benefit from including or summarizing more from the referenced EMEP Status Reports (e.g., provide direct comparisons of longer-term simulations or bias metrics).

420 Also in reply to comments from the first referee, we have added additional information referring to results discussed in the EMEP status reports “In particular, in EMEP Status Report 1/2024 (2024) country source-receptor matrices for yearly averages using the BF and LF methods have been compared for several pollutants and indicators, including O<sub>3</sub> and PM<sub>2.5</sub>. The differences are overall less than 10 %, and can be considered as small, since they are of the same order of magnitude as methodological differences (advection scheme and filtering) shown in Fig. 1 and Fig. 2 in the following. The differences due to non-linearities introduced by reducing the emissions by 15 % in the BF method, compared to, in principle, infinitesimally small reductions in the LF methodology, represent only a small fraction of methodological differences. We do stress that the 15 % emission reduction employed by the BF method is in principle arbitrary (EMEP Status Report 1/2004, chapter 4, 2004), and that the differences due to non-linearities with the LFs do not represent an actual source of methodological error.”

430 The handling of discrepancies arising from chemical regime discontinuities is important. The authors mention filtering but do not show filtered vs. unfiltered results. Including one such illustration would be helpful.

It is true that the handling of discontinuities is a major practical problem. In fact, since the submission of the original manuscript we already have developed another filtering method to give even more satisfactory results. In the revised manuscript, we take the opportunity to discuss this implementation in more details, while also including the effect of switching off the filter.

435 The citation of other relevant literature in the paper is done in a very narrow way. The authors tend to cite large amounts of their own work and work of their colleagues, and what seems like a bare minimum of other work from the broader literature. As a result of this, the paper lacks context about how the work fits in with other related studies. The narrow citation strategy

will also limit the discoverability of this new work through tools like citation databases. The authors should make an effort to discuss their work in relation to previous literature on adjoint / tangent linear modelling, other source attribution techniques such as tagging, and policy-relevant source/receptor modelling.

To add context, also in reply to a comment from the first referee, we have added “The methodology is a sensitivity method. It does not directly attempt to determine the impact of any finite change of emissions. Neither does it assign the relative total contributions from different sources (in contrast to other tagging methods (e.g., Butler et al., 2018; Emmons et al., 2012; Dunker et al., 2007)). Total contributions can only be directly inferred if the species are considered to have a linear dependency on emissions, which is not the case in general. Still, due to the lower computational cost, non-linear responses can be inferred by performing sensitivity analysis at several emission levels, thus providing indirectly the effect of non linear changes (EMEP Status Report 1/2024, 2024, ” to the introduction section.

In addition, we have added a paragraph stating “To our knowledge, the usage of Tangent Linear Models (TLM) has primarily been to compute backward trajectories for emission inversions (e.g., Zheng et al., 2024), while the adjoint of TLMs is also applied in the context of data assimilation (Shankar Rao, 2007). We will here show how a TLM can be used in a ‘forward mode’ to compute sensitivities to emission sources in the form of, for example, source-receptor matrices.” to Sect. 2.

Minor Comments and Suggestions

Line 59: how large is “large enough”? added: “larger than about ten (Sect. 3.3.2)”

Equation 1: explain why there is no subscript  $i$  on the LHS. Shouldn’t the local fractions be defined per chemical species?  
Line 85: shouldn’t the emissions  $E_k$  also have subscript  $i$ ?

We could have added a subscript  $i$ , however here it is assumed as part of source  $k$ . In the 2020 paper the source  $k$  is tracked, and therefore there is no real distinction between the source species and the pollutant species (e.g., total  $PM_{2.5}$  was tracked rather than a specific chemical subspecies contained within  $PM_{2.5}$ ). In the generalized LF case the tracked source-specific species can be different. For clarification, we have now added an example of what the indices mean in the original formulation: “For instance  $i$  could refer to primary particulate matter, and  $k$  could refer to primary particulate matter from the transport sector in Paris.”

Equation 3: the absence of the subscript  $i$  here implies that the reduction factor applies equally to all species emitted by sector  $k$ . Is this correct, and if so, is this a limitation of the method?

$k$  refers also to an emitted species/pollutant (e.g.,  $NO_x$  from traffic, or benzene from traffic, etc.), and is now more generally defined as “The source  $k$  can refer to any specific class of pollutants (from a certain sector, or in principle any other sub-class of a pollutant) either from a given grid cell within the tracked region (termed the ‘local region’) or from a predefined fixed region, such as a country.”

For different chemical species emitted by a single source (in the sense of the source referring to sector such as road traffic), the calculation of the source sensitivities indeed applies the same reduction factor (perturbation) to all emitted species. But since these perturbations are taken to be infinitesimally small, this reduction factor is equally valid for all emitted species, and is not a limitation of the method.

Equation 4: Define “ $S$ ” here.

added “is then defined as the sensitivity  $S$ , as”.

Section 2.3.1: Are there any examples of this non-linear deposition in a real CTM simulation?

Yes, or at least it has been show that it is relevant. We have added reference to the paper of Fowler et al. (2001), describing a relationship between the deposition rate of SO<sub>2</sub> and NH<sub>3</sub> concentrations.

490 Section 2.4.1: Does the choice of a small delta limit the size of the perturbations (difference from the base case) that can be calculated using Local Fractions?

495 We do only attempt to evaluate the first order derivative. We have added “Mathematically a very small value can be chosen, however for numerical stability and robustness reasons in the SIA calculations, a somewhat larger value than a 1 % perturbation is preferred here.” The choice of using a small perturbation does in general limit the applicability of the calculated LFs to a regime where the response to emission perturbations is approximately linear.

Line 233: how are these 60 perturbations chosen?

500 We have added a more explicit explanation (also in reply to a comment from the first referee) that “In our present implementation, 60 species will have a direct or indirect effect on other chemical species in the chemistry module. This implies that the function  $f$  must be evaluated for a perturbation of each of those 60 species. In addition perturbations for emissions must be performed, but usually only a few sources will contribute for a given grid cell (typically one country, four species, and only for the lowest seven vertical levels). The evaluation of the function  $f$  for each of these perturbations represents the most time-consuming part of the entire code”

505

Line 396: This strikes me as a cop-out. The whole point of the Local Fractions method is to approximate the sensitivities of non-linear systems, so some discussion of the differences between the results obtained from Local Fractions and the results of 15% perturbations is definitely appropriate in this paper.

510 To elaborate on this (also in reply to comments from the first referee), we have expanded the text to include “In particular, in EMEP Status Report 1/2024 (2024) country source-receptor matrices for yearly averages using the BF and LF methods have been compared for several pollutants and indicators, including O<sub>3</sub> and PM<sub>2.5</sub>. The differences are overall less than 10 %, and can be considered as small, since they are of the same order of magnitude as methodological differences (advection scheme and filtering) shown in Fig. 1 and Fig. 2 in the following. The differences due to non-linearities introduced by reducing the emissions by 15 % in the BF method, compared to, in principle, infinitesimally small reductions in the LF methodology, represent only a small fraction of methodological differences. We do stress that the 15 % emission reduction employed by the BF method is in principle arbitrary (EMEP Status Report 1/2004, chapter 4, 2004), and that the differences due to non-linearities with the LFs do not represent an actual source of methodological error.”

515

520

## **Generalized local fractions – a method for the calculation of sensitivities to emissions from multiple sources for chemically active species, illustrated using the EMEP MSC-W model (rv5.5)**

Peter Wind<sup>1,2</sup> and Willem van Caspel<sup>1</sup>

<sup>1</sup>Climate Modelling and Air Pollution Division, Research and Development Department, Norwegian Meteorological Institute (MET Norway), PO 43 Blindern, 0313 Oslo, Norway

<sup>2</sup>Department of Chemistry, UiT - The Arctic University of Norway, N-9037 Tromsø, Norway

**Correspondence:** Peter Wind (peter.wind@met.no)

## Abstract.

This paper presents an extension of the original Local Fraction methodology, to allow the tracking of the sensitivity of chemically active air pollutants to emission sources. The generalized Local Fractions are defined as the linear sensitivities of chemical species to source emission changes, as propagated through the full set of non-linear chemical transformations. The method allows to track simultaneously sensitivities from hundreds of sources (typically countries or emission sectors) in a single simulation. The current work describes how the non-linear chemical transformations are taken into account in a rigorous manner, while validating the implementation of the method in the European Monitoring and Evaluation Programme (EMEP) Meteorological Synthesizing Centre – West (MSC-W) chemistry-transport model by examples. While effectively producing the same results as a direct ‘brute force’ method, where the impact of emission reductions of each source has to be computed in a separate scenario simulation, the generalized Local Fractions are an order of magnitude more computationally efficient when large numbers of scenarios are considered.

## 1 Introduction

Air pollution negatively impacts both human and ecosystem health (Manisalidis et al., 2020), while being the product of a complex interplay between chemistry, meteorology, and natural and anthropogenic emissions. One of the fundamental questions in air quality research is to relate the different emission sources with air pollution in certain receptor regions (see for example The Forum for Air Quality Modelling (FAIRMODE, Thunis et al. (2022)). To this end, the Local Fractions (LF) method was originally developed as a practical way to track the relative contributions to primary air pollution from local and regional sources (Wind et al., 2020). Within the LF framework, tracking a set of sources simultaneously greatly increased the numerical efficiency, allowing the tracking of thousands of (distant) source contributions within a single simulation. However, the method did not consider the effects of chemistry, making it suitable only for chemically inert species. In this work, the LF methodology is generalized to include chemistry, allowing also the description of source sensitivities for species that undergo complex chemical processes.

The original LF method worked by assuming linearity, in the sense that the fractions computed for each source or pollutant can be considered independently. That is, the effect of an increase of emission from two sources being equal to the sum of the effects obtained by increasing each source individually. However, many important pollutants have a non-linear dependency on emissions, while also being the product of chemical reactions rather than being emitted directly. The species involved will also usually originate from different emission sources. To track the source impacts to species such as ozone (O<sub>3</sub>) and Secondary Inorganic Aerosol (SIA), the LF method is generalized to instead track the ~~sensitivity of~~ sensitivity (i.e., rate of change) of such active chemical species to source emission changes. The approach of tracking these chemical sensitivities allows for a mathematically well-defined description of non-linear processes, while also allowing for the quantification of the impact of emission changes of those sources to the total air pollutant. In this work the word “source” is defined as a set of emitted species from a predefined region and from any individual or combination of emission sector(s).

In effect, the sensitivities obtained through the generalized LF method define the changes in pollutant concentrations that would result from a small change in emissions. By ‘small’, we mean that the method calculates pollutant perturbations arising from infinitesimally small emission perturbations, thereby tracking the tangent of concentrations expressed as a function of emission intensity. Since the emission of a species will affect an entire group of species through non-linear chemical reactions, the sensitivities of all species that are directly or indirectly involved in the chemical processes must be tracked separately for each source. However, despite this added complexity, the method still allows for the tracking of hundreds of sources in a single simulation, building on the efficient formulation of the original LF methodology.

One such applications arises in the calculations of the country-to-country Source-Receptor (SR) contributions to air pollution, which is a standard product of the European Monitoring and Evaluation Programme (EMEP) Chemistry-Transport Model (CTM) developed at the Meteorological Synthesising Centre - West (MSC-W). The CTM developed at MSC-W (hereafter “EMEP model”) is a three-dimensional Eulerian model for tropospheric chemistry, having a long history of policy application and research development on both the European and global scales (e.g., Ge et al., 2024; van Caspel et al., 2024; Jonson et al., 2018; Simpson et al., 2012). Traditionally, the country-to-country SR contributions, or “blame-matrices”, have been calculated using a large number of Brute-Force (BF) simulations (around 50 per pollutant species), where country emissions for primary particles, nitrogen oxides ( $\text{NO}_x = \text{NO} + \text{NO}_2$ ), Non-Methane Volatile Organic Compounds (NMVOC), sulfur oxides ( $\text{SO}_x = \text{SO}_2 + \text{SO}_4$ ) or ammonia ( $\text{NH}_3$ ) are individually reduced by 15 %. [The information about the relative impact from different countries \(and possibly emission sectors\) can then be used to propose an optimized set of abatement measures, as is done for example using the GAINS \(Greenhouse gas – Air pollution Interactions and Synergies\) model \(Amann et al., 2011\).](#) With the generalized LF methodology, the impact of emission reductions for all individual countries and precursor species can be calculated from a single simulation.

The [methodology is a sensitivity method. It does not directly attempt to determine the impact of any finite change of emissions. Neither does it assign the relative total contributions from different sources \(in contrast to other tagging methods \(e.g., Butler et al., 2018; Emmons et al., 2012; Dunker et al., 2002; Kwok et al., 2015; Grewe, 2013; Grewe et al., 2017; Wang et al., 1998\)\). Total contributions can only be directly inferred if the species are considered to have a linear dependency on emissions, which is not the case in general. Still, due to the lower computational cost, non-linear responses can be inferred by performing sensitivity analysis at several emission levels, thus providing indirectly the effect of non linear changes \(EMEP Status Report 1/2024, 2024,](#)

The generalized LF methodology (hereafter interchangeably referred to as “LF”) is introduced and described in more detail in Sect. 2. The latter includes an illustrative example of the core ideas behind the method through its implementation for the effects of dry deposition and SIA, followed by a more general mathematical description. Sect. 3 discusses the computational cost. Sect. 4 validates the methodology through comparison against the BF method for several model configurations, in terms of numerical accuracy. The results are concluded in Sect. 5, where also a number of future applications for the generalized LFs are discussed.

## 2 Generalized Local Fractions methodology

In a scenario (or BF) approach, an independent simulation is performed for each source of interest. Conceptually, the core of the Local Fractions method is to perform a single simulation for a set of scenarios (up to a few hundred), and perform the computations in such a way that some key variables can be reused for all scenarios instead of being computed independently for each simulation. The results from the generalized Local Fractions method are in principle the same as the results from a series of BF scenario runs with small perturbations in the emissions, but can be obtained at a lower computational total cost, if the number of scenarios is ~~large enough~~ larger than about ten (Sect. 3.3.2).

For example, some variables are independent of emission intensity. Meteorological data is an obvious example, since these fields are independent of the emission scenarios, but are nevertheless read in and processed for each individual simulation. In the Local Fractions method these fields have to be computed only once to cover all the scenarios.

However, the largest benefits arise from avoiding the repetition of computationally expensive processes such as advection and chemical transformations. In the case of advection, the flux for a pollutant is computed only once in the LF method, and can then be reused for all the tracked scenarios. If the different scenarios only differ by a small perturbation (for example, 1 % emission intensity), second order effects are negligible, and this can be taken advantage of to efficiently calculate the effect of complex transformations. To this end, we will show that computing the Jacobian of the chemical transformations allows for a more efficient treatment of the sensitivities to emission perturbations, avoiding the calculation of the explicit chemical transformations for each individual scenario.

The fundamental theory underlying the generalized Local Fractions is not new within the context of Numerical Weather Prediction (NWP) and chemistry-transport modelling. It is called the ~~linear-adjoint-tangent linear~~ model, already conceptually introduced by Lorenz (1965), relating output variables with small perturbations of the initial state. ~~We will however here~~ To our knowledge, the usage of Tangent Linear Models (TLM) has primarily been to compute backward trajectories (emission inversions (Zheng et al., 2024)), while the adjoint of TLMs is also applied in the context of data assimilation (Shankar Rao, 2007). We will here show how a TLM can be used in a ‘forward mode’ to compute sensitivities to emission sources in the form of, for example, source-receptor matrices. We will give a description from ~~a more an~~ applied point of view in the context of CTMs, utilizing the efficient formulation of the original Local Fractions.

### 2.1 The original Local Fractions

As described in Wind et al. (2020), the original LF methodology was designed to track the fraction of the total pollutant with an origin from a specific source  $k$ . The source  $k$  can refer to ~~a any~~ a specific class of pollutants (from a certain sector, or in principle any other sub-class of a pollutant) either from a given grid cell within the tracked region (termed the ‘local region’) or from a predefined fixed region, such as a country. In the following, we will use upper cases (Local Fractions or LFs) when referring to the methodology used in computing the local fractions, and lower cases (local fractions) for the physical quantity that gives the actual fraction of a pollutant (though this definition changes when considering the generalized Local Fractions, as will be discussed in the following section).

The original LFs defined the local fractions as the dimensionless quantity

$$620 \quad \text{lf}_k = \frac{C_{i,k}}{C_i} = \underline{c_{i,k}}, \quad (1)$$

where  $C_{i,k}$  represents the concentration of pollutant  $C_i$  from source  $k$  and we use the lower case  $c_{i,k}$  to define the “dimensionless concentration”, or the proportion of the pollutant that originate from source  $k$ . The source  $k$  in this notation, refers to a source of pollutant  $C_i$  (units of  $\mu\text{g m}^{-3}$  for example). For instance  $i$  could refer to primary particulate matter, and  $k$  could refer to primary particulate matter from the transport sector in Paris. Since the original LF method did not consider the effects of chemistry, Eq. 1 applies only to chemically inert species such as primary Particulate Matter (PM). For such species, considering the emissions ( $E$ ) of pollutant  $C_i$  by source  $k$  (written as  $E_k$ , unit  $\mu\text{g m}^{-3}\text{s}^{-1}$ ), the local fractions at time  $t + \Delta t$  can be written as

$$625 \quad \text{lf}_k(t + \Delta t) = \frac{C_{i,k}(t) + E_k(t)\Delta t}{C_i(t + \Delta t)}, \quad (2)$$

where  $C_i(t + \Delta t)$  is the total pollutant concentration calculated by the CTM based on all emission sources. In effect, Eq. 2 then also represents the fraction by which the total pollutant concentration would be reduced if source  $E_k$  is omitted (assuming linearity). One practical advantage of using the formulation of Eq. 2 is that physical processes such as wet and dry deposition affect the total pollutant concentrations, but not the fractional source contributions. Hence such processes do not have to be calculated for each individual scenario, or source, but only once.

## 2.2 Sensitivity to emission changes

The conceptual idea behind the generalized Local Fractions is introduced here. We can rewrite  $E_k$  using a scalar multiplicative factor  $e_k$ , which can be used to define a uniform reduction factor on  $E_k$ , which itself is fully space and time dependent. This can be written as

$$\tilde{E}_k(x, y, z, t, e_k) = e_k E_k(x, y, z, t), \quad (3)$$

where the “base case” with full emissions corresponds to  $e_k = 1$ . In calculating the sensitivities to emission changes, the  $e_k$  values are perturbed to calculate the resulting impacts on the full set of chemical species included in the CTM. The index  $k$  thus describes a particular scenario where the emissions from source  $k$  are modified. Written in mathematical form, the definition of the generalized Local Fractions is then defined as the sensitivity  $S$ , as

$$S(C_i, e_k E_k) = \frac{\partial C_i}{\partial e_k} = \lim_{\epsilon \rightarrow 0} \frac{C_i(e_k E_k) - C_i((e_k - \epsilon) E_k)}{\epsilon}. \quad (4)$$

In a BF approach,  $\epsilon$  is a fixed fraction (typically 15%), and both terms in the numerator are computed using two independent simulations. In the generalized LF framework, the derivative in the case of a linear system instead becomes

$$645 \quad \lim_{\epsilon \rightarrow 0} \frac{C_i(e_k E_k) - C_i((e_k - \epsilon) E_k)}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{C_i(e_k E_k) - C_i(e_k E_k) + \epsilon C_i(E_k)}{\epsilon} = C_i(E_k), \quad (5)$$



where  $C_i(E_k)$  represents the concentration of pollutant  $C_i$  resulting from emissions from source  $k$ . The For linear processes, the definition of Eq. 4 is thus equivalent to the original LF definition of Eq. 1, multiplied by the total pollutant concentration:-

$$\frac{\partial C_i}{\partial e_k} = c_{i,k} C_i(t). \quad (6)$$

This shows that the normalization with respect to  $\epsilon$  is such that the sensitivities of Eq. 1 give the concentration change extrapolated linearly to a 100% change of emissions. However, the original interpretation as the fraction of pollutant originating from a specific source cannot be used anymore if non-linear (chemical) transformations are involved, even though the calculated sensitivities themselves reflect linearly extrapolated perturbation impacts (i.e., calculated from the tangent from an otherwise complex set of non-linear transformations). The equations and code valid for the original LFs can still be used, however, as will be discussed in the following.

It is important to keep in mind that even if the results are presented in units of concentrations for a 100% change of emissions, they can not be interpreted as total contributions for non-linear species. The values must be interpreted as sensitivities to small emission changes. Those sensitivities can be both positive or negative, and will in general not sum up to total concentrations.

### 2.3 Illustrative example: Dry deposition

Before describing the mathematical formulation of the general case in more detail, we will show in a simpler case how the LF method fundamentally differs from a mere parallel run of scenarios. To that end, we will describe the procedure for computing the contributions of different sources to dry deposition.

The deposition process of species  $C_i$  can be written as the product of the concentration of species  $C_i$  and an effective deposition velocity  $v_i$ , as

$$\text{Dep}_i = v_i C_i. \quad (7)$$

Here the effective deposition velocity is taken as a dimensionless number between 0 and 1 (with  $v_i = 1$  meaning that all of species  $C_i$  is lost to deposition):

$$C_i(t + \Delta t) = C_i(t) - v_i C_i(t). \quad (8)$$

With this notation,  $\text{Dep}_i$  has units of  $\mu\text{g m}^{-3}$  per timestep, and must be multiplied by the grid cell volume to get the weight of pollutants deposited during the time step. The effective deposition velocity will depend on many parameters, such as fractions of land use, leaf area index, meteorological parameters, etc. The computation of  $v_i$  can therefore be computationally demanding. In a parallel scenario approach, each scenario would recompute the value of  $v_i$ , and the computation time for those terms is therefore proportional to the number of scenarios. If we instead look at the sensitivity of  $\text{Dep}_i$  for species  $C_i$  to emission changes from source  $k$ :

$$\frac{\partial \text{Dep}_i}{\partial e_k} = \frac{\partial v_i C_i}{\partial e_k} = v_i \frac{\partial C_i}{\partial e_k}, \quad (9)$$



675 the  $\frac{\partial C_i}{\partial e_k}$  term is the sensitivity calculated using the generalized Local Fractions, where we have assumed that the deposition velocity  $v_i$  is independent of the concentrations for small concentration perturbations. This shows that the deposition velocity has to be computed only once, and the sensitivity to emission changes of the deposition, can be computed with a simple multiplication for each scenario  $S_k(C_i)$ . The latter being a shorthand notation of the sensitivity of species  $C_i$  to source  $k$  given by the definition of Eq. 4.

680 Another aspect of the original LFs was that dry deposition does not change the values of the local fractions itself. That is, deposition changes the total modeled concentrations, but not the fractional contributions of different emission sources tracked by the LFs. Using the dimensionless notation (see Eq. 1), for the deposition process:

$$\underline{c_{i,k}(t + \Delta t) = c_{i,k}(t)}. \quad (10)$$

This property is maintained in the generalized LFs framework, since the generalized LFs are equivalent to the original LFs but multiplied by the total pollutant concentration, Eq. 6. Therefore, while the process of dry deposition affects the total pollutant concentration, the value of the sensitivities (or generalized local fractions) remains unchanged except for the renormalization:

$$\frac{\partial C_i}{\partial e_k}(t + \Delta t) = \frac{\partial C_i}{\partial e_k}(t) - \underline{v_i \frac{\partial C_i}{\partial e_k}(t)} = \frac{\partial C_i}{\partial e_k}(t) \frac{C_i(t + \Delta t)}{C_i(t)}. \quad (11)$$

While Eq. 11 reflects dry deposition during time step  $\Delta t$ , the handling of other processes during  $\Delta t$  (such as chemical transformations) will be discussed later on.

### 690 2.3.1 Non-linear deposition

So far the results are simply proportional to the magnitude of the linear emission changes. It has been shown Fowler et al. (2001) that for example SO<sub>2</sub> deposition is affected by NH<sub>3</sub> concentration. If we want to describe ~~a situation~~ such situations where the deposition velocity depends on the concentration of other species  $C_j$ , where  $j$  can represent any number of species (also  $j = i$ ), we need to add this dependency as an additional transformation in Eq. 9, as

$$695 \quad \frac{\partial \text{Dep}_i}{\partial e_k} = \frac{\partial v_i C_i}{\partial e_k} = \sum_j \frac{\partial v_i}{\partial C_j} \frac{\partial C_j}{\partial e_k} + v_i \frac{\partial C_i}{\partial e_k}. \quad (12)$$

However, the range of validity is now limited, because Eq. 12 assumes that  $\frac{\partial v_i}{\partial C_j}$  is independent of the scenario, which is valid only in a first order approximation (i.e., that the scenarios only differ slightly). For larger deviations from the base concentrations, this may not be the case anymore, and the calculated deposition sensitivities may no longer be representative. Compared to a regular simulation, the additional computational cost will now include the calculation of the derivatives  $\frac{\partial v_i}{\partial C_j}$ .

700 Note however that this additional cost is still independent of the number of scenarios (indexed by  $k$ ) considered in the LF simulation, such that including additional scenarios comes at practically no additional computational cost.

## 2.4 Chemistry

In Wind et al. (2020), we showed how the local fractions are transformed during advection and other linear processes, largely analogous to the dry deposition example from the previous section. In this section we will show how they are transformed

705 through non-linear chemical transformations. But before presenting the general equations [for all species and relevant model process sensitivities](#), we will show in more details how the chemical transformations of SIA are treated within the generalized LF framework. This will illustrate in a simpler context the way the generalized LFs are transformed when chemical transformation are taking place.

### 2.4.1 Generalized Local Fractions approach for SIA

710 In the thermodynamic equilibrium chemistry modules of CTMs, the concentrations of the species  $\text{HNO}_3$ ,  $\text{NO}_3$ ,  $\text{NH}_4$ ,  $\text{NH}_3$  and  $\text{SO}_4$  are partitioned into the gas and particulate phases, with a dependence on each other as well as on physical environment variables such as temperature and humidity.

We consider one grid cell, and the chemical transformation during one time step. In a direct scenario approach, a set of concentrations is considered for each scenario. For each scenario, the following transformation is applied:

$$715 \quad C_i^k(t + \Delta t) = f_i(\{C_j^k(t)\}), \quad (13)$$

where  $C_i^k$  are the concentrations,  $k$  is the scenario index and  $i$  and  $j$  run over the five species  $\text{HNO}_3$ ,  $\text{NO}_3$ ,  $\text{NH}_4$ ,  $\text{NH}_3$  and  $\text{SO}_4$ .  $f_i$  represents the non-linear thermodynamic equilibrium transformation. In this formulation the equilibrium module is applied for each scenario, and the computational cost will then be proportional to the number of scenarios.

We now assume that all the scenarios differ by only small amounts compared to a “base case”  $C_i$ :

$$720 \quad C_i^k(t) = C_i(t) + \delta C_i^k(t). \quad (14)$$

In a linearized picture, where we assume that  $\delta C_i^k$  are small, the resulting concentrations at time step  $t + \Delta t$  can be approximated to first-order using a Taylor-series expansion as

$$C_i^k(t + \Delta t) = C_i(t + \Delta t) + \sum_j^5 a_i^j \delta C_j^k(t), \quad (15)$$

725 where  $C_i$  is the concentration of one of the five species ( $i = 1, \dots, 5$ ) and  $a_i^j$  are the transformation parameters, which describe the transformation of the concentrations of species  $C_i$  as a function of species  $C_j$  during time step  $\Delta t$ .  $a_i^j$  will depend on the environment, the size of the time step, but also on the base case concentrations; however in a linearized picture, ~~those  $a_i^j$~~  are considered constants [, during the time step, being](#) independent of the scenario  $k$ .

The clue here is that the matrix  $a_i^j$  is of a fixed size ( $5 \times 5$ ), independent of the number of scenarios  $k$ . The number of operations required in Eq. 15 is still proportional to the number of scenarios. However, the number of operations for each  
730 scenario is very small (five multiplications, and six additions). By comparison, the application of the full equilibrium module typically requires thousands of operations.

To compute the updated values of the generalized local fractions at time  $t + \Delta t$ , the  $\delta C_j^k(t)$  terms in Eq. 15 can be chosen to be proportional to the sensitivities:

$$\delta C_j^k(t) = \delta \frac{\partial C_j}{\partial e_k}. \quad (16)$$

735 Then Eq. 15 can be written as

$$C_i^k(t + \Delta t) = C_i(t + \Delta t) + \delta \frac{\partial C_i}{\partial e_k}(t + \Delta t) = C_i(t + \Delta t) + \sum_j^5 a_i^j \delta \frac{\partial C_j}{\partial e_k}(t), \quad (17)$$

and thus

$$\frac{\partial C_i}{\partial e_k}(t + \Delta t) = \sum_j^5 a_i^j \frac{\partial C_j}{\partial e_k}(t). \quad (18)$$

This shows that each updated local fraction can be computed directly using only the matrix  $a_i^j$ , without having to explicitly  
740 apply the full transformation of Eq. 13 for each source  $k$ .

We still need to compute the matrix  $a_i^j$ . This can be done in practice by evaluating Eq. 13 for five additional sets of input values, where in each set only one of the pollutants is perturbed:

$$C_j^j(t) = C_j(t) + \delta C_j(t) \quad (19)$$

$$C_i^j(t) = C_i(t) \quad i \neq j. \quad (20)$$

745 We get then one column of the matrix for each iteration by inverting the equation:

$$a_i^j = \frac{C_i^j(t + \Delta t) - C_i(t + \Delta t)}{\delta C_j(t)}, \quad (21)$$

where  $C_i^j(t + \Delta t)$  is the result for species  $i = 1, \dots, 5$  from the calculation perturbing species  $C_j$  and  $C_i(t + \Delta t)$  is the result from the unperturbed (base) case.

Alternatively one can use dimensionless local fractions  $\frac{\partial c_j}{\partial e_k} = \frac{\partial C_j}{\partial e_k} \frac{1}{C_j}$ , as is done in our code implementation for the EMEP  
750 model:

$$b_i^j = a_i^j \frac{C_j(t)}{C_i(t + \Delta t)} = \frac{C_i^j(t + \Delta t) - C_i(t + \Delta t)}{\delta C_i(t + \Delta t)}. \quad (22)$$

The updated values can then computed in the same way:

$$\frac{\partial c_i}{\partial e_k}(t + \Delta t) = \sum_j^5 b_i^j \frac{\partial c_j}{\partial e_k}(t). \quad (23)$$

The reason for using dimensionless local fractions in the code itself stems from historical reasons with simulations of only  
755 primary PM.

We note that in the EMEP model, the resulting equilibrium solutions as calculated with the MARS thermodynamic equilibrium module (as used in the current work) depends only on the sums of  $\text{HNO}_3 + \text{NO}_3$  and  $\text{NH}_4 + \text{NH}_3$ , and on the  $\text{SO}_4$  concentrations at time  $t$ . Therefore only three supplementary evaluations of Eq. 13 are actually required to determine the entire  $a_i^j$  matrix.

760 In these calculations  $\delta$  is chosen to be a fixed small number, normally  $\delta = 0.05$ . [Mathematically a very small value can be chosen, however for numerical stability and robustness reasons in the SIA calculations, a somewhat larger value than a 1 % perturbation is preferred here.](#)

## 2.4.2 General case

Given the generalized local fractions in a given grid cell at time  $t$ , we want to compute the new values after that the concentrations have been updated through the Chemical module of the host CTM. We will assume (as is the case in our model code), that the emissions are included as additional terms in the Chemistry module.

The concentrations at time  $t + \Delta t$  can be expressed as a general function of all the input concentrations and emissions at time  $t$ . We will only write explicitly the parameters that are affected by the emission sources (e.g.,  $\Delta t$ , temperature, etc., are not modified by a change in ~~emission~~emissions in our model, and are therefore not included). If some parameters are emission dependent (for example, solar radiation may depend on aerosol concentrations, and hence on emissions), they can be included in a similar way as discussed in the following.

Writing the concentrations at time  $t + \Delta t$  gives

$$C_i(t + \Delta t) = f_i(C_1, C_2, \dots, C_{n_C}, E_1, E_2, E_3, \dots, E_{n_E})(t), \quad (24)$$

where all concentrations on the right-hand side are at time  $t$  and  $E_j$  are the emissions sources. If we derive the equation with respect to  $e_k$  (assuming that  $f_i$  and its partial derivatives are continuous functions) we get:

$$\frac{\partial C_i}{\partial e_k}(t + \Delta t) = \sum_j^{n_E} \frac{\partial f_i}{\partial E_j} \frac{\partial E_j}{\partial e_k} + \sum_j^{n_C} \frac{\partial f_i}{\partial C_j} \frac{\partial C_j}{\partial e_k}(t) + \underbrace{\sum_j^{n_E} \frac{\partial f_i}{\partial E_j} \frac{\partial E_j}{\partial e_k}}_{\text{wavy line}}, \quad (25)$$

where  $\frac{\partial C_j}{\partial e_k}(t)$  and  $\frac{\partial C_j}{\partial e_k}(t + \Delta t)$  are the values of the sensitivities at time  $t$  and  $t + \Delta t$ . In Eq. 25, the  $\frac{\partial f_i}{\partial E_j}$  and  $\frac{\partial f_i}{\partial C_j}$  terms define the Jacobian of the transformation,  $\frac{\partial f_i}{\partial E_j}$  represents the average rate of change in the concentration  $C_i$  due to the emissions  $E_j$  during  $\Delta t$ .  $\frac{\partial E_j}{\partial e_k}$  shows the dependence of the emissions of species  $j$  to source  $k$ ; typically this could be 1 within a country referenced by  $k$  that emits species  $j$ , and zero outside of the country; it could also be some other fraction if one looks at more complex situations such as sector specific emissions.

Note that  $\epsilon$  in Eq. 4 is assumed small, however  $\Delta t$  or the changes in Eq. 24 are not ~~necessarily~~ assumed to be small. In the general case, and in the model code, the function  $f$  is ~~also~~ not assumed to be linear, ~~even within the~~ and Eq. 25 is valid also for non-infinitesimal  $\Delta t$  time-scale. For example, in the chemical solver  $\Delta t$  is divided into micro-iterations, each time step capturing the full non-linear chemistry.

If in a chemical scheme the chemical transformations can be assumed linear within the time step, the Jacobian might be directly accessible as an analytical function of the input concentrations. Otherwise, the computation of the Jacobian matrix must be computed for example by a method similar as shown for SIA in Sect. 2.4.1.

In our present implementation ~~this involves the evaluation of~~, 60 species will have a direct or indirect effect on other chemical species in the chemistry module. This implies that the function  $f$  ~~for approximately 60 different sets of input values (each set being must be evaluated for a perturbation of the base input values), and is each of those 60 species.~~ In addition perturbations for emissions must be performed, but usually only a few sources will contribute for a given grid cell (typically one country, four species, and only for the lowest seven vertical levels). The evaluation of the function  $f$  for each of these perturbations represents the most time-consuming part of the entire code (see Sect. 3.3).

Eq. 25 is a general equation that describes how to update the sensitivities for any transformation. For example in the (linear) deposition case, the Jacobian is simply equal to  $1 - v_i$ . In the advection case (below) the indices of the concentrations would refer to pollutants at different position in space and the Jacobian will be equal to the fluxes.

## 2.5 Advection

Mathematically the advection can be treated similarly in the Local Fraction method. Now the sensitivity updates can depend on the values at different positions in space, but there is no mixing between different species caused by advection.

In a one-dimensional advection time step, some pollutants are transferred from one cell to a neighboring cell, which can be written as

$$C_i(x_0, t + \Delta t) = C_i(x_0, t) + F_i(x_{-\frac{1}{2}})C_i(x_{-1}, t) - F_i(x_{+\frac{1}{2}})C_i(x_{+1}, t). \quad (26)$$

Here  $F_i(x_{-\frac{1}{2}})$  and  $F_i(x_{+\frac{1}{2}})$  are unitless fluxes through the cell boundaries; they represent the fraction of pollutants transferred between two neighboring cells. These fluxes are the fraction of air masses transported to a neighboring grid cell and are in reality independent of the pollutant concentrations. In a simplified zero order scheme the fluxes would simply be proportional to the wind speed  $\frac{v\Delta t}{\Delta x}$ , with  $v$  wind speed and  $\Delta x$  the size of the cell. However, in the EMEP model the one-dimensional Bott's fourth order scheme (Bott, 1989a, b) is used. ~~In this scheme, the fluxes are dependent on the concentrations in the five nearest cells, not only the wind speed.~~ The main purpose of the scheme is to minimize the so-called numerical diffusion. In effect, using the Bott scheme means that the fraction of pollutant that is transferred between neighboring grid cells ~~depends~~ will depend not only on the wind, but also on concentrations ~~in those five of five surrounding~~ grid cells and thus indirectly on emissions. This has consequences when a brute force method is applied: the changes in concentrations observed when ~~emission~~ emissions are modified, are not only a consequence of physical effects. The fluxes calculated in two scenarios will be different and this will affect the transport patterns, because the advection scheme will be based on different concentration distributions and this will have an indirect, non-physical effect. One visible adverse consequence of this effect is that in a BF approach, it can sometimes be observed that an increase of emissions in a grid cell, can lead to a decrease of concentrations in an upwind grid cell.

Using the formalism from the preceding section (Eq. 25), Eq. 26 becomes:

$$\begin{aligned} \frac{\partial C_i}{\partial e_k}(x_0, t + \Delta t) = & \frac{\partial C_i}{\partial e_k}(x_0, t) + \\ & \sum_n \frac{\partial F_i(x_{-\frac{1}{2}})}{\partial C_i(x_n)} \frac{\partial C_i(x_n)}{\partial e_k}(t) C_i(x_{-1}, t) + F_i(x_{-\frac{1}{2}}) \frac{\partial C_i(x_{-1})}{\partial e_k}(t) + \\ & \sum_n \frac{\partial F_i(x_{+\frac{1}{2}})}{\partial C_i(x_n)} \frac{\partial C_i(x_n)}{\partial e_k}(t) C_i(x_{+1}, t) + F_i(x_{+\frac{1}{2}}) \frac{\partial C_i(x_{+1})}{\partial e_k}(t), \end{aligned} \quad (27)$$

where  $n$  runs over the five neighboring cells and  $\frac{\partial C_i(x_n)}{\partial e_k}$  is the sensitivity of species  $C_i$  to source  $k$  in cell  $x_n$ . The  $\frac{\partial F_i}{\partial C_i}$  terms reflect the dependence of the fluxes on the concentrations of neighboring cells. If we would use a simpler scheme where the

fluxes do depend only on wind speed, and not on the concentrations,  $\frac{\partial F_i}{\partial C_i} = 0$  and the corresponding terms in Eq. 27 do not contribute (this is the case for the zero order advection approximation used for comparisons in Sect. 4).

825 In principle we could use Eq. 27 in the generalized Local Fractions framework, and reproduce also the results arising from advection changes from BF in the case of small ~~emission~~-emissions changes. Instead, for the horizontal advection of the sensitivities, or local fractions, we will set  $\frac{\partial F_i}{\partial C_i} = 0$ . That is, use the same base fluxes  $F_i$  for all the scenarios  $k$ , and not take into account the changes in the fluxes that appear when concentrations change. Eq. 27 can then be written as

$$\frac{\partial C_i}{\partial e_k}(x_0, t + \Delta t) = \frac{\partial C_i}{\partial e_k}(x_0, t) + F_i(x_{-\frac{1}{2}}) \frac{\partial C_i(x_{-1})}{\partial e_k}(t) + F_i(x_{+\frac{1}{2}}) \frac{\partial C_i(x_{+1})}{\partial e_k}(t), \quad (28)$$

830 which gives a more stable transport pattern and ensures that for primary species an increase of emissions can never give a decrease in concentrations in the advection process. The  $F_i$  fluxes in Eq. 28 are then simply the same as those calculated by the CTM for the regular advection of species.

For the vertical advection, the EMEP model uses a second order scheme. In that case we have still chosen to include all terms of Eq. 27. This is because  $O_3$  has high values at high altitude, and this can have a strong effect on the vertical transport patterns. To keep a better compatibility with the BF method, we have found it preferable here to not use only the base fluxes, as in the horizontal advection case.

### 3 Computational aspects

The LF modules are additions to the original EMEP model code, but do not affect the original results (e.g., those used for BF calculations). For each module (advection, chemistry, emissions, depositions etc.) a corresponding LF module exists, which at each time step computes the updates to the local fractions, but there is no feedback from the LF modules into the regular concentrations.

#### 3.1 Differences with BF approach

In theory, the generalized Local Fractions method can give results identical to a direct method when two runs which differ only by a small change of emissions are compared. In practice some differences are still present. The largest differences are due to the differences in the treatment of advection. As discussed in the preceding section, this is expected to slightly *improve* the quality of the results rather than deteriorate them. For testing purposes, it is still possible to use a simplified advection scheme (“zero order” Bott scheme), for which the results from the advection module using both methods will be identical.

Some transformations (i.e., partial derivatives) are not yet implemented. For example some reaction rates depend on the surface of particulate matter present in a grid cell. A change of emissions ~~will change~~-may affect the size of the particulate matter, and thereby give a change of those reaction rates. ~~Those~~-These secondary effects are at present not taken into account in the local fraction calculations.

Another limitation stems from the photolysis rate ( $J$ -value) calculations made using the Cloud- $J$  module, which is the default photolysis rate scheme used by the EMEP model from version 4.47 onward (van Caspel et al., 2023). Since Cloud- $J$

takes into account the instantaneous modeled abundance of  $O_3$  and a number of aerosol species throughout the atmospheric column, the chemistry inside a single grid cell is no longer completely local, depending also on the radiative impact of chemical concentrations (and perturbations) in the above grid cells. The impact of this limitation is however expected to be comparatively small, ~~however~~, with the majority of over-head absorption of radiation relevant to active chemistry occurring above the EMEP model top (100 hPa), for which (UV-absorbing)  $O_3$  concentrations are specified based on observations.

There are other differences due to the details of the numerical schemes. One example, is the scheme for chemical transformations: the chemical scheme uses a fixed number of iterations. The starting guess will depend on the concentrations from the previous time step. For the local fractions, the starting guess is also taken from the corresponding scenario at the previous iteration, however this is not completely equivalent.

### 3.2 Filtering

Some processes may present discontinuities, where an infinitesimal change in an input value can produce a non-infinitesimal change in the output values. This is not uncommon in for example two situations:

- A test ~~may is~~ be performed on the value of a concentration (chemical regime), and the code can branch into one transformation in one case and into another for the other case.
- An iterative procedure ~~may is~~ be used, and the number of iterations used ~~may depend~~ depends on some concentration-dependent criterion. The number of iterations may then vary slightly between otherwise almost identical cases.

In a LF approach, if numerical derivation is used, the chance of being just at the two sides of a branching point is small, but the effect will also be large, since the derivative value is obtained by dividing by  $\epsilon$  in Eq. 4. In a BF approach these discontinuities may happen more often, but their effect is also smaller. Such discontinuities have been observed in thermodynamic equilibrium chemistry modules (e.g., Capps et al., 2012b), as will also be discussed in Sect. 4.3.

One advantage of the LF approach is that it is possible to filter out such effects, if they are not too numerous: since in those cases the calculated derivatives will be much larger than can reasonably be expected, they can be detected and some action taken (simply keep the values of the local fractions unchanged for this particular point and time step for example). Filtering the results is much more difficult in a BF approach, since it is difficult to recognize the problematic situations: since the BF base run and scenario run are independent, it is not possible to detect those special situations. It is also not trivial to define what to do if one would detect a problematic chemical regime.

### 3.3 Computational cost

Since the main advantage of the LF method is its computational cost, we will present in some details how the cost compares to direct scenario runs.

3.3.1 Cost of chemistry

The computation of the chemical transformations is the computationally most expensive part of the EMEP model; this is probably the case for most CTMs. The calculation of the Jacobian matrix will increase this cost substantially and is the computationally most demanding part of the LF calculation. This cost is however in theory independent of the number of pollutant sources that are traced. This is a fundamental difference compared to direct ~~or tagging~~ methods. For those methods, the number of components that undergo the full chemistry scheme increase proportionally with the number of scenarios~~or tags~~.

The number of operations performed in Eq. 25 will still increase with the number of traced pollutants, but it has the form of a matrix multiplication. Matrix multiplications can be done extremely ~~effectively~~ efficiently on most computers, and will in practice have a negligible computational cost.

The computation of the Jacobian matrix is a fully local process (i.e., local to each grid cell), and with our code it will scale perfectly with the number of Message Passing Interface (MPI) processes in a multi-core parallel run. This means that the computation time can in practice be reduced by increasing the number of processors used.

3.3.2 Scaling of computation time with number of scenarios

Table 1 compares the time required to run the EMEP model using the LF method, for different numbers of scenarios. We assume that each source region (country) is analyzed for five different ~~emission~~ emissions reductions: SO<sub>x</sub>, NO<sub>x</sub>, NH<sub>3</sub>, VOC and primary PM, and therefore count five scenarios for each country in addition to the baseline scenario. Here we note that some more specific details about the EMEP model itself will be discussed in Sect 4.

**Table 1.** Relative run times ~~comparing BF to of~~ LF simulations and their speed up in comparison to BF. The time unit is defined as the time to run a single BF scenario, amounting to 29.1 seconds (without LF  $T_{\text{scenario}}$ ) for the model setup described in Sect. 4 (48 hours simulation on a  $400 \times 260 \times 20$  grid). 50-2, 50-4, and 50-5 stands for 50 countries with 2, 4 or 5 different ~~sectors~~ sector contributions ~~One run simulates a 48 hours scenario in a  $400 \times 260 \times 20$  grid. The reference time is taken 29.1 seconds, which is half of the time used to run without LF on four compute nodes (each emitting five pollutants).~~ The results in the table show the time to execute the LF code on eight compute nodes. The speed up is defined as the time it would take to compute all the scenarios if a direct method (BF) had been used, divided by the time required using the LF method ( $\frac{\text{number of scenarios}}{\text{total time}} \frac{\text{number of scenarios} \cdot T_{\text{scenario}}}{\text{total LF time}}$ ).

<del>number</del> <u>Number</u> of countries	2	6	12	25	50	50-2	50-4	50-5
<del>number</del> <u>Number</u> of scenarios	11	31	61	126	251	501	1001	1251
<del>total time</del> <u>Total LF time / <math>T_{\text{BF scenario}}</math></u>	10.5	11.0	11.0	13.5	18.3	47.9	109.0	137.1
<del>speed</del> <u>Speed</u> up	1.1	2.8	5.6	9.3	13.7	10.5	9.2	9.1

When a small number of scenarios are evaluated, the total time for a LF run is almost independent of the number of scenarios. This is because most of the time is spent computing the Jacobian matrix of the chemistry module, and this time does not depend on the number of scenarios. For the cases with large number of scenarios, the computation time is dominated by the



local fractions updates required for each scenario (mainly in the advection module), and the run time increases linearly with the number of scenario. This is reflected by a constant time per scenarios, or speed up.

905 The speed up is better for around 250 scenarios than it is for higher number of scenarios. That means that it is more efficient to run for instance 250 scenarios twice, than to run 500 scenarios in a single run. This can be explained, because even if the number of operations is formally smaller when doing all the scenarios together, in practice the data arrays become very large; if the number of array elements that are looped over becomes too large, the CPU will run out of memory cache, thereby causing a drop in hardware efficiency. In the future, such limitations should be avoidable by improved code design.

910 In practice (EMEP Status Report 1/2024, 2024), ~~for a full SR matrix using calculation for 55 countries ,on a latitude-longitude-altitude grid with dimensions  $400 \times 260 \times 20$  ( $0.2^\circ \times 0.3^\circ$  horizontal resolution), will for a full year LF simulation on 16 compute nodes (running 128 MPI processes on each compute node) will take 18 wall time hours.~~ This can be compared to a single run on four compute nodes which requires 2 hour and 40 minutes. A direct BF method would require  $(55 \times 5 + 1)$  single runs. The computational resources required would then be approximately 10  ~~$(=(55 \times 5 + 1) \times 2.7 \times 4 / (18 \times 16))$~~ -times higher compared to the LF run.

Because all species involved in the transformations have to be tracked independently, tracking 55 countries for four chemically active species ( $\text{SO}_x$ ,  $\text{NO}_x$ ,  $\text{NH}_3$ , VOC) and primary inert particles ( $\text{PPM}_{2.5}$  and  $\text{PPM}_{\text{co}}$ ) represent more than 15000 individual local fractions to track.

### 3.4 ~~Approximating Optimization by approximating~~ chemically active species

920 ~~Each term of the Jacobian matrix,  $\frac{\partial f_i}{\partial c_j}$  in Eq. 25, is explicitly evaluated in our present implementation. This might not be necessary, and in this section we will briefly indicate the types of simplifications that might be implemented in the future.~~

The number of Sulfur (S) or Nitrogen (N) atoms are conserved during chemical transformations. We can in principle track the atoms from different sources in a physically meaningful way, and the sum of contributions from different sources will be equal to the total contribution. The difficulty arises because the atoms are part of different type of molecules. In a given point in time and space, the relative amount of the different molecules ( $\text{SO}_2/\text{SO}_4$ ,  $\text{NO}_2/\text{NO}$ , and  $\text{NH}_3/\text{NH}_4$  for example) will be different for the atoms from different sources, because they have a different history.

In a first approximation one can assume than those relative amounts are source independent. For the Local Fraction method, this ~~would~~ represent a great simplification, as the  $\text{SO}_x$ ,  $\text{NO}_x$ , or  $\text{NH}_x$  molecules can then be treated as primary (as show and discussed in Wind et al. (2020)). Additional effects can be taken into account if the different molecules are also tracked separately for the different sources. However this will still not be exact, since all chemical species involved in the transformations (such as OH,  $\text{O}_3$ ) should all be tracked for completeness.

One could try to further generalize this approach by grouping species into families where the total number of members of a family is conserved during chemical transformations. This can then be combined with a simpler chemical scheme for the computation of the Jacobian matrix.

935 As explained for SIA, even if the  $a$  matrix (Eq. 18) is a  $5 \times 5$  matrix, only three additional evaluation of the SIA operator are necessary to compute it, not five as in a general case. For the full chemistry module similar simplifications can be found.

In our code the species involved in the ozone chemistry can produce SOA (Secondary Organic Aerosols), but the SOA species do not influence  $O_3$ , not even indirectly. Meaning that a block of the Jacobian matrix is known to be zero and does not need to be computed. This is taken advantage of in the generalized LF code. The oxidized and reduced nitrogen atoms are conserved during the chemical process, which means that some simplifications could be obtained as for SIA. This is however not implemented yet.

From a more mathematical view point, one can consider the chemical transformations (in one grid cell, during one time step) as a matrix relating small changes in input to change in the output concentrations. This matrix can be diagonalized, and the eigenvectors with eigenvalues with value zero reflect a conserved quantity. The rank of the matrix is then smaller than its size. Building on this, one could keep only the eigenvectors with largest eigenvalues, and neglect the one below a threshold.

There are many more paths to explore that could improve the efficiency of our code, ranging from very simple (updating the Jacobian matrix every second time step only), to purely computational (using a GPU accelerator for the evaluation of the Jacobian matrix).

## 4 Validation

In order to verify that the code actually gives the expected values for the emission sensitivities, the sensitivities calculated using the LF method are compared to the brute force method for 1 % emission perturbations. The latter are sufficiently small that differences between the two methodologies due to non-linear chemistry are avoided, thereby isolating only the methodological differences. Indeed, if the ~~emission~~-emissions differences are small enough, the calculated derivative obtained by finite differences (BF) should be equal to the sensitivities calculated with the LFs.

The main EMEP model settings are essentially standard, employing standard EMEP reported emissions, except that some natural emissions are omitted (soil  $NO_x$ , ocean Dimethyl sulfide (DMS), lightning, forest fires, dust, aircrafts) for simplicity. The model is run on a  $0.3^\circ \times 0.2^\circ$  horizontal grid, employing 20 vertical levels up to a model top height of 100 hPa. The meteorology is based on 3-hourly data derived from the ECMWF Integrated Forecasting System (IFS) cycle 40r1 model (ECMWF, 2014). While the EMEP model uses its default EmChem19 mechanism (Bergström et al., 2022), employing a simplified set of lumped VOC species (Ge et al., 2024). As noted before, the following setup also employs the MARS equilibrium chemistry module. The results can be reproduced using the code and data provided under the “Code and data availability” section.

The length of the simulation is only 24 hours, in order to have a light weight setup. The purpose here is only to validate the principle of the methodology, not to quantify the differences in all possible situations (climate, species, time scales, emissions, etc.). However, more extensive and realistic comparisons can be found in ~~EMEP Status Report 1/2024 (2024)~~ EMEP Status Report 1/2024 (2024) Ch. 5 and ~~EMEP Status Report 1/2023 (2023)~~ EMEP Status Report 1/2023 (2023) Ch. 5. Furthermore, while the current work focuses on the differences in results arising from methodological differences, in practice BF calculations are often performed using 15 % rather than 1 % emission reductions. The impact on the differences between the BF and LF methodologies for 15 % emission reductions are also investigated in more detail in EMEP Status Report 1/2023 (2023) Ch. 5 and ~~EMEP Status Report 1/2023 (2023)~~ EMEP Status Report 1/2024 (2024) Ch. 5, noting that the results are qual-

970 itatively similar to those discussed in the following. ~~However, since for 15 % perturbations differences arise due to non-linear chemical effects and not from methodological differences in itself, these are not discussed in more detail here~~

In particular, in EMEP Status Report 1/2024 (2024) country source-receptor matrices for yearly averages using the BF and LF methods have been compared for several pollutants and indicators, including O<sub>3</sub> and PM<sub>2.5</sub>. The differences are overall less than 10 %, and can be considered as small, since they are of the same order of magnitude as methodological differences (advection scheme and filtering) shown in Fig. 1 and Fig. 2 in the following. The differences due to non-linearities introduced by reducing the emissions by 15 % in the BF method, compared to, in principle, infinitesimally small reductions in the LF methodology, represent only a small fraction of methodological differences. We do stress that the 15 % emission reduction employed by the BF method is in principle arbitrary (EMEP Status Report 1/2004, chapter 4, 2004), and that the differences due to non-linearities with the LFs do not represent an actual source of methodological error.

975

## 980 4.1 Country example

In this section, the net impact of 1 % perturbations in NO<sub>x</sub>, SO<sub>x</sub>, NH<sub>3</sub> and VOC emissions are investigated for Germany (DE). Here DE is taken as being a representative country for the comparison of the LF and BF methods, featuring considerable geographic differences in emissions and chemical regimes. The output species included in the comparison are daily mean O<sub>3</sub> and SIA, both involving highly active chemical transformations. However, the LF outputs are in general also available for a range of other species and derived O<sub>3</sub>-indicators, such as Peak Season (April-September) average Maximum Daily 8-hour Average (MDA8) O<sub>3</sub>, and reactive nitrogen and sulfur deposition. Here we note that, in order to calculate the impact of a 1 % emission reduction using the LF outputs, the LF outputs are multiplied by a factor of 0.01, since the LF outputs represent the impact of emission perturbations linearly extrapolated to a 100 % change of emissions.

985

As discussed in Sect. 2.5, one source of discrepancies between the BF and generalized LF methods is the (intended) difference arising from the advection scheme used. In order to distinguish differences due to other causes, a set of test runs have also been performed with a simplified advection scheme (zero order advection). In this simplified scheme, there are no differences due to difference in the advectons treatment, at the cost of some additional numerical diffusion in both results.

990

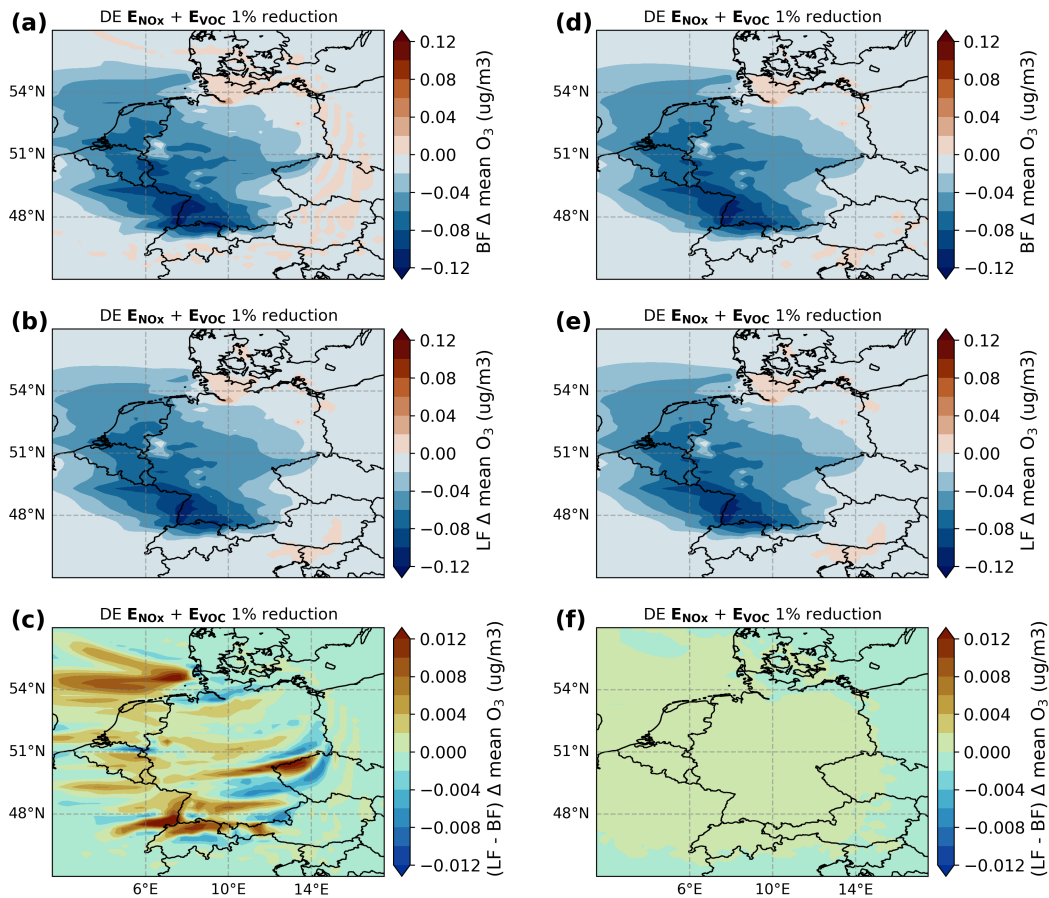
## 4.2 Surface O<sub>3</sub>

The comparison between the BF and LF methods in the regular setup (fourth order advection) as well as for the zero order advection setup is illustrated in Fig. 1, here shown for the change in surface O<sub>3</sub> resulting from the combined perturbations in NO<sub>x</sub> and VOC emissions (being those precursor species to which O<sub>3</sub> is most sensitive).

995

In Fig. 1a it can be seen that a reduction in NO<sub>x</sub> and VOC emissions leads to decreases in surface O<sub>3</sub> almost everywhere inside and in the vicinity of Germany. Nevertheless, due to titration effects parts of Northern Germany and Berlin see an increase in O<sub>3</sub> concentrations. Fig. 1b illustrates that these effects are well captured by the LF methodology, while the corresponding zero order advection runs (panels d and e) show very similar results. The difference between the BF and LF results for the fourth order advection setup (panel c) shows a pattern of positive and negative variations, having a magnitude of around 10 % of that of the individual BF and LF results. However, the lack of difference between BF and LF for the zero order advection scheme

1000



**Figure 1.** Impacts of 1 %  $\text{NO}_x$  and VOC emission reductions in Germany (DE) on daily mean surface  $\text{O}_3$  concentrations, calculated for the 1st on July 2018 using the EMEP model with a default fourth order advection (left-hand panels) and zero order advection (right-hand panels) configuration. Top panels show the impacts calculated using the BF method, middle panels with the LF method, and the bottom panels the difference between the two. The scaling of panels (c) and (f) is a factor of ten smaller than that of the other panels.

(panel f) demonstrates that the differences between the fourth order runs are almost entirely due to the choice of advection scheme. The remaining discrepancies in the zero order advection setup are very small (of the order of 1% in this particular test).

### 4.3 SIA

Fig. 2 shows tests of the impacts on surface SIA concentrations calculated for the combined effects of 1 % reductions in  $\text{NO}_x$ ,  $\text{SO}_x$ , and  $\text{NH}_3$  emissions. ~~In this comparison, the difference between the fourth order advection runs shows a pattern that is in many ways qualitatively similar to that found for the response in Fig. 1c. However, while the BF and LF using the~~ zero order advection setup. While the BF (panel (a)) and LF (panel (d)) calculations show generally agreeable results, ~~the~~

~~difference between the zero order advection runs~~ their difference (panel (b)) shows a comparatively large difference northwest of the Netherlands. Further diagnostic simulations finds that ~~this arises~~ these differences arise due to gas-aerosol partitioning calculations taking place in the MARS thermodynamics equilibrium module. While the differences between the BF and LF results are still comparatively small, they do point towards the general complication with the use of complex numerical models to calculate the impact of small (emission) perturbations, as discussed in Sect. 3.2.

For example, in the MARS module small perturbations to its input parameters can ~~easily~~ lead to a change in the number of iterations applied to certain solver routines. Furthermore, certain physical mechanisms, such as aerosol water uptake (in turn affecting the equilibrium solution), can show step-like behavior near certain threshold values. While we have modified the MARS module to smooth the solution in certain parts of its code, variations in the calculated SIA such as those shown in Fig. 2f-b nevertheless persist. These do, however, not have a large impact on the total simulation results, also for simulations performed over longer time periods. ~~The differences between SIA calculations are furthermore also present in BF simulations, and are therefore not necessarily a shortcoming of the LF method.~~

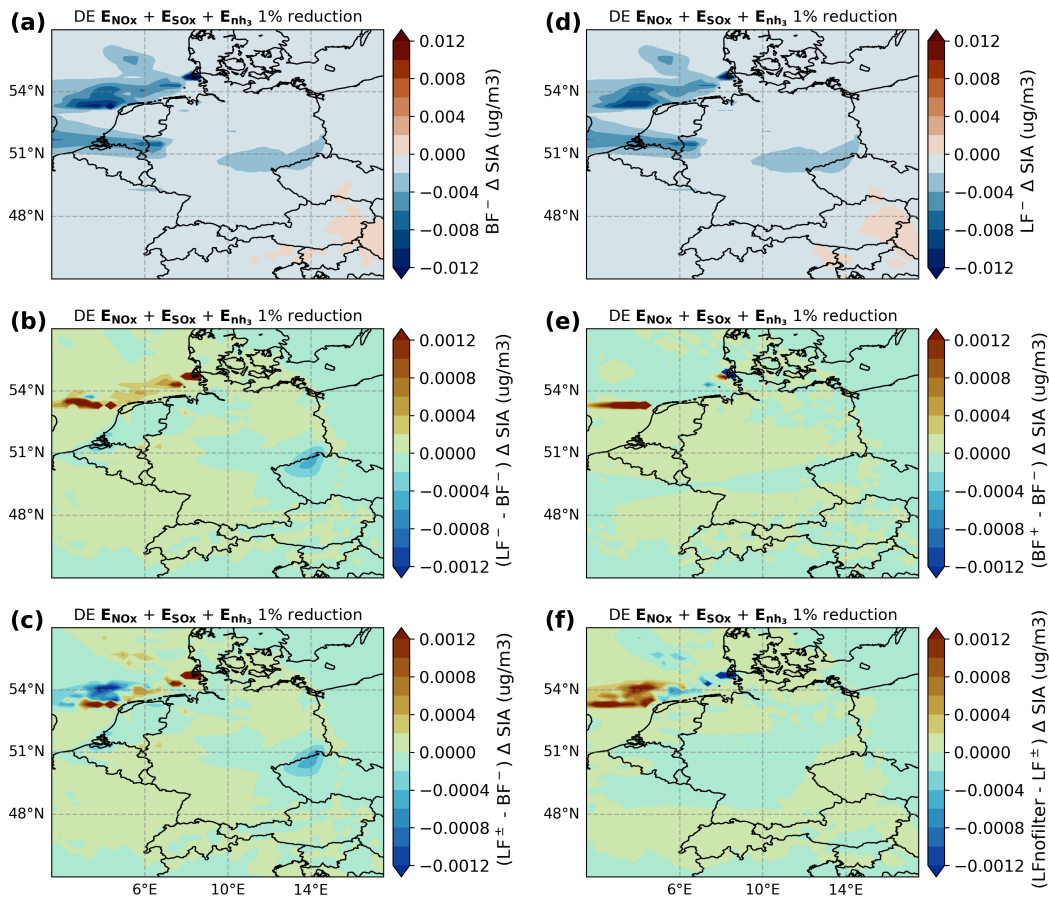
To further illustrate the effects of numerical instabilities, Fig. 2e shows the difference between BF calculations employing a 1 % emission reduction ( $BF^-$ ) and a 1 % emission increase ( $BF^+$ ). Noting that here the  $BF^+$  simulation has been used to likewise calculate the impact of a 1 % emission reduction by changing the sign of its results. One would expect the results between the regular  $BF^-$  and  $BF^+$  calculations to be almost identical, but in practice they are not, due to numerical effects arising from the complex thermodynamics calculations. This also demonstrates that the problems with discontinuities are present in the BF method too and are not specific to the LF method.

The principle of the SIA filter in the EMEP model is to reject results where more than four molecules in the outputs are created or destroyed for each additional (perturbation) molecule in the input. In Fig. 2c we also show the result for the LF method using an alternative filtering technique. In this filtering method each numerical derivative (Eq. 22) is performed twice, now for one with a positive  $\delta$  and also for one with a negative  $\delta$ . If the ratio between the positive and negative input perturbations does not fall within a factor of three, they are rejected, meaning that the local fractions are kept unchanged in the equilibrium module for this chemistry timestep. Furthermore, when the derivatives calculated with the positive and negative perturbations are found to be in agreement with each other, their geometric mean value is used as the final sensitivity.

Fig. 2f shows specifically the effect of this new filtering technique ( $LF^\pm$ ) by comparing against a LF run without any filtering at all ( $LF_{\text{nofilter}}$ ). As expected, the regions that show discontinuities in the BF method (panel (e)) are also regions where the filter is activated. However, there are also regions where the filter has a small effect, without the BF results showing any problem. We nevertheless find that the alternative SIA filter ( $LF^\pm$ ) overall produces the most numerically stable results.

## 5 Conclusion

The current work describes the theory and implementation of the generalized Local Fractions, which can be used to efficiently track the linear sensitivity to emission changes of air pollutants subject to complex non-linear chemical transformations. Building upon the efficient formulation of the original Local Fractions, the generalized formulation allows the tracking of the



**Figure 2.** As Fig. 1, but Different tests for the impacts of 1 % NO<sub>x</sub>, SO<sub>x</sub> and NH<sub>3</sub> emission reductions changes on daily mean surface SIA concentrations calculated for the 1st on July 2018 using the EMEP model with zero order advection. Panel (a): using the brute force method with an emission reduction (BF<sup>-</sup>). Panel (d): using LF and filtering of number of molecules created in response to a negative input perturbation (LF<sup>-</sup>). Panel (b): difference between panel (d) (LF<sup>-</sup>) and panel (a) (BF<sup>-</sup>). Panel (e): Difference of BF method using a reduction of 1 % (BF<sup>-</sup>) and an increase of 1 % (BF<sup>+</sup>). Panel (c) same as (b), but using a different filtering method (LF<sup>±</sup>, see text). Panel (f): Difference using LF with or without filtering.

sensitivities to hundreds of sources in a single simulation, increasing the computational efficiency by a factor of 10 over the  
1045 standard SR “blame-matrix” computations performed annually by MSC-W using the EMEP model. While differences between the emission reduction impacts calculated using the BF and LF methodologies exist, these can largely be understood, arising predominantly as an adverse side effect from the choice of advection scheme in the BF simulations.

The use of the original Local Fractions method has already proven fruitful in several applications considering pollutants as inert particles.

- 1050 – The uEMEP scheme is a down-scaling scheme (Denby et al., 2020), allowing to describe air pollution at fine resolution (down to 25 m), but still taking into account the effect of long range transport. In this scheme the local fractions gives the fraction of the pollutants which have a local origin, and those can then be replaced by more accurate, fine resolution values (Denby et al., 2024a, b).
- 1055 – For the GAINS (Greenhouse gas – Air pollution Interactions and Synergies) model (Amann et al., 2011), a full analysis of the SR relationships (from any part to any grid cell) over large regions have been produced using the local fractions (also called “transfer coefficients” in this context). Applications in Europe (Klimont et al., 2022) and South East Asia (World Bank Group, 2023) exist. A report on the methodology used in such applications of GAINS is under preparation.
- Using hourly time tagged emission sources, it is possible to use the LFs for inverse modelling, i.e., trying to reconstruct the emission sources based on observations. Such developments are presently underway.

1060 The generalized Local Fraction calculations with the full chemistry are not efficient enough to give results as detailed as for inert particles (where tens thousand of sources can be tracked simultaneously). Still, when a large number of scenarios are to be simulated, the new approach is much more efficient than previously available methods, opening up new fields of applications which are presently being investigated:

- 1065 – Due to computational limitations, the country-to-country blame-matrices calculated by MSC-W are normally performed on a reduced resolution  $0.3^\circ \times 0.2^\circ$  horizontal grid spanning  $30^\circ\text{N}$ - $82^\circ\text{N}$  to  $30^\circ\text{W}$ - $90^\circ\text{E}$ . However, with the considerably more efficient general LF method, future blame-matrix calculations could be performed on the regular  $0.1^\circ \times 0.1^\circ$  grid without loss of numerical accuracy.
- 1070 – The sensitivities to emission source perturbations calculated using the generalized LF method can be used as SR relationship coefficients in the calculation of cost-effective emission control strategies with the GAINS model, also for chemically active species such as  $\text{O}_3$ . Such calculations could furthermore benefit from the use of sensitivities calculated from simulations with different background emission levels, to include a description of the non-linear response to emission reductions when the reductions are comparatively large.
- 1075 – A complete picture of the SR relationships at different background levels, allows to describe the *accumulated* contributions from each country using the path integral method (Dunker, 2015). By integrating the sensitivities over a given emission change pathway, it is possible to determine the source contributions differences between two emission scenarios (see, e.g., EMEP Status Report 1/2024, 2024, Ch. 6). This allows to relate the source sensitivities with source apportionment (Clappier et al., 2017).

1080 The computer code is still under development. In the near future we intend to include more secondary processes, such as the dependency of some reaction rates on the surface area of aerosols and a more complete description of SOA, and to develop the user interface. There is also room for significant improvements in computational efficiency, although the present version already has proven an order of magnitude faster than BF methods in some relevant situations.

*Code and data availability.* A user friendly setup for testing and reproducing the results shown in this article is available at <https://doi.org/10.5281/zenodo.14162688> (Wind and Caspel, 2024). This includes a full copy of the EMEP MSC-W model code and a set of input data that can be used to produce the examples presented in this paper.

1085     For air pollution modelling purposes, we recommend to use the official version of the full EMEP MSC-W model code and main input data available through a GitHub repository under a GNU General Public License v3.0 through <https://zenodo.org/records/14507729> (EMEP MSC-W, 2024) (last access May 2025). The routines related to the Local Fractions are part of the standard model.

*Author contributions.* All authors contributed to the discussion and development of the main ideas, the applications, and the preparation of the paper. PW wrote the corresponding Fortran90 code.

1090     *Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* IT infrastructure in general was available through the Norwegian Meteorological Institute (MET Norway). Some computations were performed ~~on~~using resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway (grant NN2890k and NS9005k). The CPU time made available by ECMWF has been critical for the generation of meteorology used as input for the EMEP MSC-W model and the calculations presented in the current work.



- Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Hoeglund-Isaksson, L., Klimont, Z., Nguyen, B., Posch, M., Rafaj, P., Sandler, R., Schoepp, W., Wagner, F., and Winiwarter, W.: Cost-effective control of air quality and greenhouse gases in Europe: Modeling and policy applications, *Environmental Modelling & Software*, 26, 1489–1501, <https://doi.org/10.1016/j.envsoft.2011.07.012>, 2011.
- Bergström, R., Hayman, G. D., Jenkin, M. E., and Simpson, D.: Update and comparison of atmospheric chemistry mechanisms for the EMEP MSC-W model system, Tech. rep., [https://emep.int/publ/reports/2022/MSCW\\_technical\\_1\\_2022.pdf](https://emep.int/publ/reports/2022/MSCW_technical_1_2022.pdf), last access: November 2024, 2022.
- 1100 Bott, A.: A Positive Definite Advection Scheme Obtained by Nonlinear Renormalization of the Advective Fluxes, *Monthly Weather Review*, 117, 1006 – 1016, [https://doi.org/10.1175/1520-0493\(1989\)117<1006:APDASO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1006:APDASO>2.0.CO;2), 1989a.
- Bott, A.: Reply, *Mon. Weather Rev*, 117, 2633–2636, 1989b.
- Butler, T., Lupascu, A., Coates, J., and Zhu, S.: TOAST 1.0: Tropospheric Ozone Attribution of Sources with Tagging for CESM 1.2.2, *Geoscientific Model Development*, 11, 2825–2840, <https://doi.org/10.5194/gmd-11-2825-2018>, 2018.
- 1105 Capps, S. L., Henze, D. K., Hakami, A., Russell, A. G., and Nenes, A.: ANISORROPIA: the adjoint of the aerosol thermodynamic model ISORROPIA, *Atmospheric Chemistry and Physics*, 12, 527–543, <https://doi.org/10.5194/acp-12-527-2012>, 2012a.
- Capps, S. L., Henze, D. K., Hakami, A., Russell, A. G., and Nenes, A.: ANISORROPIA: the adjoint of the aerosol thermodynamic model ISORROPIA, *Atmospheric Chemistry and Physics*, 12, 527–543, <https://doi.org/10.5194/acp-12-527-2012>, 2012b.
- 1110 Clappier, A., Belis, C. A., Pernigotti, D., and Thunis, P.: Source apportionment and sensitivity analysis: two methodologies with two different purposes, *Geoscientific Model Development*, 10, 4245–4256, <https://doi.org/10.5194/gmd-10-4245-2017>, 2017.
- Denby, B. R., Gauss, M., Wind, P., Mu, Q., Grøtting Wærsted, E., Fagerli, H., Valdebenito, A., and Klein, H.: Description of the uEMEP\_v5 downscaling approach for the EMEP MSC-W chemistry transport model, *Geoscientific Model Dev.*, 13, 6303–6323, <https://doi.org/10.5194/gmd-13-6303-2020>, 2020.
- 1115 Denby, B. R., Kieseewetter, G., Nyiri, A., Klimont, Z., Fagerli, H., Wærsted, E. G., and Wind, P.: Sub-grid Variability and its Impact on Exposure in Regional Scale Air Quality and Integrated Assessment Models: Application of the uEMEP Downscaling Model, *Atmospheric Environment*, p. 120586, <https://doi.org/10.1016/j.atmosenv.2024.120586>, 2024a.
- Denby, B. R., Klimont, Z., Nyiri, A., Kieseewetter, G., Heyes, C., and Fagerli, H.: Future scenarios for air quality in Europe, the Western Balkans and EECCA countries: An assessment for the Gothenburg protocol review, *Atmospheric Environment*, 333, 120602, <https://doi.org/10.1016/j.atmosenv.2024.120602>, 2024b.
- 1120 Dunker, A. M.: Path-integral method for the source apportionment of photochemical pollutants, *Geoscientific Model Development*, 8, 1763–1773, <https://doi.org/10.5194/gmd-8-1763-2015>, 2015.
- Dunker, A. M., Yarwood, G., Ortmann, J. P., and Wilson, G. M.: Comparison of Source Apportionment and Source Sensitivity of Ozone in a Three-Dimensional Air Quality Model, *Environ. Sci. Tech.*, 36, <https://doi.org/10.1021/es011418f>, 2002.
- 1125 ECMWF: IFS Documentation CY40R1 - Part IV: Physical Processes, Tech. Rep. 4, ECMWF, <https://doi.org/10.21957/f56vvey1x>, 2014.
- EMEP MSC-W: OpenSource v5.5 (202412), <https://doi.org/10.5281/zenodo.14507729>, 2024.
- EMEP Status Report 1/2004, chapter 4: Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, EMEP MSC-W & CCE & CCC & ICP & JRC, [https://emep.int/publ/reports/2004/Status\\_report\\_int\\_del2.pdf](https://emep.int/publ/reports/2004/Status_report_int_del2.pdf), last access: 2025, Norwegian Meteorological Institute (EMEP/MS-C-W), Oslo, Norway, 2004.

- 1130 EMEP Status Report 1/2023: Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, EMEP MSC-W & CCC & CEIP, [https://emep.int/publ/reports/2023/EMEP\\_Status\\_Report\\_1\\_2023.pdf](https://emep.int/publ/reports/2023/EMEP_Status_Report_1_2023.pdf), last access: November 2024, Norwegian Meteorological Institute (EMEP/MSC-W), Oslo, Norway, 2023.
- EMEP Status Report 1/2024: Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components, EMEP MSC-W & CCC & CEIP, [https://emep.int/publ/reports/2024/EMEP\\_Status\\_Report\\_1\\_2024.pdf](https://emep.int/publ/reports/2024/EMEP_Status_Report_1_2024.pdf), last access: November 2024, Norwegian Meteorological Institute (EMEP/MSC-W), Oslo, Norway, 2024.
- 1135 Emmons, L. K., Hess, P. G., Lamarque, J.-F., and Pfister, G. G.: Tagged ozone mechanism for MOZART-4, CAM-chem and other chemical transport models, *Geosci. Model Dev.*, 5, doi:10.5194/gmd-5-1531-2012, 2012.
- Fowler, D., Sutton, M., and Flechard, C. e. a.: The Control of SO<sub>2</sub> Dry Deposition on to Natural Surfaces by NH<sub>3</sub> and its Effects on Regional Deposition, Water, Air, & Soil Pollution: Focus, 1, <https://doi.org/10.1023/A:1013161912231>, 2001.
- 1140 Ge, Y., Solberg, S., Heal, M. R., Reimann, S., van Caspel, W., Hellack, B., Salameh, T., and Simpson, D.: Evaluation of modelled versus observed non-methane volatile organic compounds at European Monitoring and Evaluation Programme sites in Europe, *Atmospheric Chemistry and Physics*, 24, 7699–7729, <https://doi.org/10.5194/acp-24-7699-2024>, 2024.
- Grewe, V.: A generalized tagging method, *Geosci. Model Dev.*, 6, doi:10.5194/gmd-6-247-2013, 2013.
- Grewe, V., Tsati, E., Mertens, M., Frömming, C., and Jöckel, P.: Contribution of emissions to concentrations: the TAGGING 1.0 submodel based on the Modular Earth Submodel System (MESSy 2.52), *Geosci. Model Dev.*, 10, <https://doi.org/10.5194/gmd-10-2615-2017>, 2017.
- 1145 Jonson, J. E., Schulz, M., Emmons, L., Flemming, J., Henze, D., Sudo, K., Tronstad Lund, M., Lin, M., Benedictow, A., Koffi, B., Dentener, F., Keating, T., Kivi, R., and Davila, Y.: The effects of intercontinental emission sources on European air pollution levels, *Atmospheric Chemistry and Physics*, 18, 13 655–13 672, <https://doi.org/10.5194/acp-18-13655-2018>, 2018.
- Kakavas, S., Pandis, S. N., and Nenes, A.: ISORROPIA-lite: A Comprehensive Atmospheric Aerosol Thermodynamics Module for Earth System Models, *Tellus*, 74, 1, 2022.
- 1150 Klimont, Z., Kiesewetter, G., Kaltenegger, K., Wagner, F., Kim, Y., Rafaj, P., Schindlbacher, S., Heyes, C., Denby, B., Holland, M., Borken-Kleefeld, J., Purohit, P., Fagerli, H., Vandyck, T., Warnecke, L., Nyiri, A., Simpson, D., Gomez-Sanabria, A., Maas, R., Winiwarter, W., Ntziachristos, L., Georgakaki, M., Bleeker, A., Wind, P., Höglund-Isaksson, L., Sander, R., Nguyen, B., Poupa, S., and Anderl, M.: Support to the development of the third Clean Air Outlook. Final Report, International Institute for Applied Systems Analysis, Laxenburg, Austria, [https://environment.ec.europa.eu/publications/third-clean-air-outlook\\_en](https://environment.ec.europa.eu/publications/third-clean-air-outlook_en), 2022.
- 1155 Kwok, R. H. F., Baker, K. R., Napelenok, S. L., and Tonnesen, G. S.: Photochemical grid model implementation and application of VOC, NO<sub>x</sub>, and O<sub>3</sub> source apportionment, *Geosci. Model Dev.*, 8, <https://doi.org/10.5194/gmd-8-99-2015>, 2015.
- Lorenz, E. N.: A study of the predictability of a 28-variable atmospheric model, *Tellus*, 17, 321–333, <https://doi.org/10.3402/tellusa.v17i3.9076>, 1965.
- 1160 Manisalidis, I., Stavropoulou, E., Stavropoulos, A., and Bezirtzoglou, E.: Environmental and health impacts of air pollution: a review, *Frontiers in public health*, 8, 505 570, <https://doi.org/10.3389/fpubh.2020.00014>, 2020.
- Miller, S. J., Makar, P. A., and Lee, C. J.: HETerogeneous vectorized or Parallel (HETPv1.0): an updated inorganic heterogeneous chemistry solver for the metastable-state NH<sub>4</sub><sup>+</sup>–Na<sup>+</sup>–Ca<sup>2+</sup>–K<sup>+</sup>–Mg<sup>2+</sup>–SO<sub>4</sub><sup>2-</sup>–NO<sub>3</sub><sup>-</sup>–Cl<sup>-</sup>–H<sub>2</sub>O system based on ISORROPIA II, *Geoscientific Model Development*, 17, 2197–2219, <https://doi.org/10.5194/gmd-17-2197-2024>, 2024.
- 1165 Shankar Rao, K.: Source estimation methods for atmospheric dispersion, *Atmospheric Environment*, 41, 6964–6973, <https://doi.org/10.1016/j.atmosenv.2007.04.064>, 2007.

- Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, A., and Wind, P.: The EMEP MSC-W chemical transport model – technical description, *Atmospheric Chemistry and Physics*, 12, 7825–7865, <https://doi.org/10.5194/acp-12-7825-2012>, 2012.
- Thunis, P., Janssen, S., Wesseling, J., Piersanti, A., Pirovano, G., Tarrasón, L., Guevara, M., Lopez-Aparicio, S., Monteiro, A., Martín, F., Bessagnet, B., Clappier, A., Pisoni, E., Guerreiro, C., and González Ortiz, A.: Recommendations for the revision of the ambient air quality directives (AAQDs) regarding modelling applications, <https://doi.org/10.2760/761078>, 2022.
- van Caspel, W. E., Simpson, D., Jonson, J. E., Benedictow, A. M. K., Ge, Y., di Sarra, A., Pace, G., Vieno, M., Walker, H. L., and Heal, M. R.: Implementation and evaluation of updated photolysis rates in the EMEP MSC-W chemistry-transport model using Cloud-J v7.3e, *Geoscientific Model Development*, 16, 7433–7459, <https://doi.org/10.5194/gmd-16-7433-2023>, 2023.
- van Caspel, W. E., Klimont, Z., Heyes, C., and Fagerli, H.: Impact of methane and other precursor emission reductions on surface ozone in Europe: scenario analysis using the European Monitoring and Evaluation Programme (EMEP) Meteorological Synthesizing Centre – West (MSC-W) model, *Atmospheric Chemistry and Physics*, 24, 11 545–11 563, <https://doi.org/10.5194/acp-24-11545-2024>, 2024.
- Wang, Y., Jacob, D. J., and Logan, J. A.: Global simulation of tropospheric O<sub>3</sub>-NO<sub>x</sub>-hydrocarbon chemistry: 3. Origin of tropospheric ozone and effects of nonmethane Hydrocarbons, *J. Geophys. Res.*, 103, 1998.
- Wind, P. and Caspel, W. E.: EMEP/MSC-W model version rv5.5 with sample of input data, <https://doi.org/10.5281/zenodo.14162688>, 2024.
- Wind, P., Rolstad Denby, B., and Gauss, M.: Local fractions – a method for the calculation of local source contributions to air pollution, illustrated by examples using the EMEP MSC-W model (rv4\_33), *Geoscientific Model Development*, 13, 1623–1634, <https://doi.org/10.5194/gmd-13-1623-2020>, 2020.
- World Bank Group: Striving for Clean Air: Air Pollution and Public Health in South Asia, Washington DC: World Bank, <https://doi.org/10.1596/978-1-4648-1831-8>, 2023.
- Wu, Q. Z., Z. F. Wang, A., Gbaguidi, C., Gao, L. N. L., and Wang, W.: A numerical study of contributions to air pollution in Beijing during CAREBeijing-2006, *Atmos. Chem. Phys.*, 11, doi:10.5194/acp-11-5997-2011, 2011.
- Zheng, T., Feng, S., Steward, J., Tian, X., Baker, D., and Baxter, M.: Development of the tangent linear and adjoint models of the global online chemical transport model MPAS-CO<sub>2</sub> v7.3, *Geoscientific Model Development*, 17, 1543–1562, <https://doi.org/10.5194/gmd-17-1543-2024>, 2024.