
Abstract

Air pollution poses a significant threat to human health and the environment. Although efforts are being made to develop emission inventories, and characterise and quantify emission rates, the absence of direct measurements introduces uncertainty into model-based estimations, leading to uncertain air pollution predictions. Chemistry transport models, such as the EUROpean Air pollution Dispersion-Inverse Model (EURAD-IM), can combine physical, chemical, and meteorological information to simulate the state of the atmosphere. A common methodology for incorporating observational data into chemistry transport models for further analysis is data assimilation. The EURAD-IM uses four-dimensional variational data assimilation, which allows for the correction of emission inventories. However, the accuracy of these corrections depend on the information provided by the observations. It is therefore a fundamental requirement of the observational data to be representative of the state of the atmosphere.

In order to improve the representativity of the model analysis, a methodology to generate a representative split of ground-based observation data is developed in this work. A clustering algorithm is used to determine observation stations with common statistical features e.g. mean and variance of the annually averaged diurnal cycle. By sampling each cluster individually, the resulting assimilation and validation data sets are representative of the available measurement data. The representativity is evaluated as the difference in the root mean square error of the assimilation data to the validation data after the model analysis. It is shown that the representativity of the new data sets is improved compared to an operational configuration.

To assess the influence of observation configurations on the model result, a series of observing system simulation experiments is performed. Hereby, artificial observations are extracted from a model simulation, the nature run, which serves as ground truth to gauge the accuracy of subsequent model simulations. The nature run is generated with a known emission perturbation for CO (carbon monoxide), NO_x (nitric oxide and nitrogen dioxide), NMVOC (non-methane volatile organic compounds), and SO_x (sulfur oxide) from EURAD-IM simulations. This allows to examine the influence of the observation network on the emission rate correction directly. A systematic benchmark study addressing the relation between spatial coverage of the observational network and the quality of the model analysis is performed. The ability of the model to reproduce the emission rate perturbation of the nature run is evaluated for different observation networks with varying coverage. It is shown that the best performance is achieved by an observation configuration covering approximately 11 % of the model grid. Increasing the observation density further reduces the accuracy of the emission rate corrections.

In addition, a surrogate model for the evaluation of the impact of individual observations is developed, focusing on the definition of an optimal observation network for emission rate

optimization. It combines an adjoint sensitivity field of the simulated day with the emission inventory data and the root-mean square error (RMSE) of the model concentrations of a reference simulation (without data assimilation) with respect to the observation. An algorithm for the generation of an observation configuration based on the observation impact is developed and the resulting observation configuration is tested in the framework established by the benchmark study. The generated observation network outperforms the best benchmarking configurations in terms of RMSE reduction and convergency rate with a lower number of observations. The algorithm is extended to supplement satellite-based column density observations to ground observation networks. It is shown that additional satellite observations to an optimal ground-based observation configuration are detrimental to the emission rate optimization. The emission rate RMSE reductions are decreased by 7 to 16 %. This is due to the column density observations affecting the entire model column in the assimilation, while mainly the ground layers are relevant for the emission rate optimization. Additional satellite observations to an operational configuration show an improvement in the emission rate RMSE reductions (17 to 29 %). This is attributed to the improved capturing of emission sources not adequately captured by the operational configuration. The studies in this thesis enhances the understanding of the impact of observation configurations for the correction of emission rates, which has not been presented in the literature to date. They lay the foundation for future investigations on the optimal combination of different measurement instruments to enable more accurate emission rate assessments and improved air quality predictions.

Kurzzusammenfassung

Luftverschmutzung stellt eine signifikante Bedrohung für die menschliche Gesundheit und die Umwelt dar. Während Anstrengungen unternommen werden, um Emissionsinventare zu entwickeln und Emissionsraten zu charakterisieren und zu quantifizieren, führt die Abwesenheit direkter Messungen zu Unsicherheiten in modellbasierten Emissionsabschätzungen, welche wiederum zu unsicheren Vorhersagen der Luftverschmutzung führen. Chemie-Transportmodelle, zum Beispiel das EUROpean Air pollution Dispersion-Inverse Model (EURAD-IM), sind in der Lage physikalische, chemische und meteorologische Informationen zu kombinieren, um den Zustand der Atmosphäre zu simulieren. Eine häufig genutzte Methode zur Einbindung von Beobachtungsdaten in Chemie-Transportmodelle für weitere Analysen ist die Datenassimilation. Das EURAD-IM nutzt vier-dimensionale variationelle Datenassimilation, welche unter anderem die Korrektur von Emissionsraten erlaubt. Die Genauigkeit dieser Korrekturen hängt jedoch von den Informationen ab, die durch die Beobachtungen bereitgestellt werden. Es ist daher ein grundsätzlicher Anspruch an die Beobachtungsdaten, dass sie repräsentativ für den Zustand der Atmosphäre sind.

Um die Repräsentativität der Modellanalyse zu verbessern, wird in dieser Arbeit eine Methodik zur Erzeugung einer repräsentativen Aufteilung von boden-gestützten in situ Beobachtungen entwickelt. Ein Clustering-Algorithmus wird genutzt um Beobachtungsstationen mit gemeinsamen statischen Merkmalen zu identifizieren, z.B. Mittelwert und Varianz des jährlich gemittelten Tagesgangs. Durch die individuelle Aufteilung jedes Clusters sind die resultierenden Assimilations- und Validierungsdatensätze repräsentativ für die verfügbaren Messdaten. Die Repräsentativität wird durch die Differenz der Wurzel des quadratischen Mittelwertfehlers der Assimilationsdaten zu den Validierungsdaten nach der Modellanalyse bewertet. Es zeigt sich, dass die Repräsentativität der neuen Datensätze im Vergleich zu einer operativen Konfiguration verbessert ist.

Um den Einfluss der Beobachtungskonfigurationen auf das Modellergebnis abzuschätzen, wird eine Reihe von Experimenten mit simulierten Beobachtungssystemen durchgeführt. Dabei werden künstliche Beobachtungen aus dem sogenannten ‘nature run’ extrahiert, dieser dient als Vergleichsgrundlage, um die Genauigkeit weiterer Modellsimulationen zu bewerten. Der nature run wird mit einer bekannten Störung der CO (Kohlenstoffmonoxid), NO_x (Stickoxide), NMVOC (flüchtige organische Verbindungen außer Methan), und SO_x (Schwefeloxide) Emissionen aus EURAD-IM Simulationen generiert. Dadurch lässt sich der Einfluss des Beobachtungsnetzes auf die Korrektur der Emissionsraten durch das Assimilationssystem direkt evaluieren. Es wird eine systematische Benchmark-Studie durchgeführt, die sich mit dem Zusammenhang zwischen der räumlichen Abdeckung des Beobachtungsnetzes und der Qualität der Modellanalyse befasst. Die Fähigkeit des Modells, die Störungen der Emissionsraten des nature runs zu reproduzieren, wird für verschiedene Beobachtungsnetze mit variierenden Abdeckungen evaluiert. Es zeigt sich, dass die bes-

te Leistung mit einer Beobachtungskonfiguration erreicht wird, die näherungsweise 11 % des Modellgitters abdeckt. Eine weitere Erhöhung der Beobachtungsdichte reduziert die Genauigkeit der Emissionsratenkorrektur.

Darüber hinaus wird ein Surrogatmodell zur Bewertung der Auswirkung einzelner Beobachtungen entwickelt, wobei der Schwerpunkt auf der Definition eines optimalen Beobachtungsnetzes zur Optimierung der Emissionsraten liegt. Es kombiniert ein adjungiertes Sensitivitätsfeld des simulierten Tages mit den Emissionsinventardaten und der Wurzel des quadratischen Mittelwertfehlers (RMSE) der Modellkonzentrationen aus einem Referenzlauf (ohne Datenassimilation) in Bezug auf die Beobachtung. Es wird ein Algorithmus zur Generierung einer Beobachtungskonfiguration auf der Grundlage der Auswirkung einzelner Beobachtungen entwickelt, und die resultierende Beobachtungskonfiguration wird im Rahmen der Benchmark-Studie getestet. Die generierte Beobachtungskonfiguration übertrifft die Güte der besten Benchmark-Konfiguration hinsichtlich der RMSE Reduktion und der Konvergenzrate mit einer geringeren Anzahl an Beobachtungen. Der Algorithmus wird erweitert, um satellitengestützte Beobachtungen von Säulendichten zu Bodenbeobachtungsnetzen zu ergänzen. Es zeigt sich, dass zusätzliche Satellitenbeobachtungen zu einer optimalen bodengestützten Beobachtungskonfiguration sich nachteilig auf die Optimierung der Emissionsraten auswirken. Die Reduktion des RMSE der Emissionsraten wird um 7 bis 16 % vermindert. Das liegt daran, dass die Beobachtungen der Säulendichten die gesamte Modellsäule in der Assimilation beeinflussen, während für die Optimierung der Emissionsraten hauptsächlich die Bodenschichten relevant sind. Zusätzliche Satellitenbeobachtungen zu einer operativ genutzten Konfiguration zeigen eine Verbesserung der Reduktion des RMSE der Emissionsraten (17 bis 29 %). Dies ist auf die verbesserte Erfassung von Emissionsquellen zurückzuführen, welche von der operativ genutzten Konfiguration nicht hinreichend erfasst werden. Die Untersuchungen in dieser Arbeit verbessern das Verständnis des Einflusses von Beobachtungskonfigurationen auf die Korrektur von Emissionsraten, was bisher in der Literatur noch nicht behandelt wurde. Sie legen den Grundstein für zukünftige Studien zur optimalen Kombination verschiedener Messinstrumente, um genauere Emissionsratenbewertungen und verbesserte Luftqualitätsvorhersagen zu ermöglichen.