



AP4 Abschlussbericht

Schwellenwerte

Der Gegenstand von Arbeitspaket 4 ist die Berechnung von Schwellenwerten für den Verkehr, um vorgegebene Schadstoff-Grenzwerte einzuhalten. Auf diese Weise kann ermittelt werden, wann das Luftschadstoff-Prognosemodell eine Überschreitung der Grenzwerte vorhersagt und wie stark der Verkehr reduziert werden müsste, um die Überschreitung zu verhindern. Darauf basierend können verkehrslenkende Maßnahmen ergriffen werden.

Generelles Vorgehen	3
Ausgangsbedingungen	4
Schadstoff-Grenzwerte	4
Messstationen	4
Mögliche Ansätze	5
Prognose der Verteilungen statt Punktschätzer	5
Trajektorien	6
Direkte Verwendung des normalen Schadstoffmodells	6
Selektion der Input-Variablen im normalen Schadstoffmodell	7
Gewicht des Verkehrs in der Modellschätzung erhöhen	8
Zweistufiges Modell	8
Umsetzung des zweistufigen Modells: Details	10
Variablen im ersten Teilmodell	10
Strategie zur Verkehrsreduktion	11
Tweak Values: Optimierung der Ergebnisse im entscheidenden Bereich	11
Prognosehorizont	12
Ergebnisse	12
Ergebnisformat	12
Beispiel: Mariendorfer Damm	13
Ermittelte Tweak Values	15
Überschreitungen in der Jahresübersicht	16



Genauigkeit: verpasste Überschreitungen vs. falsche Alarme	17
Speicherung und Bereitstellung der Schwellenwerte	18
Code und Code-Dokumentation	20
Ansätze zur Übertragung auf Straßen ohne Messcontainer	20



Generelles Vorgehen

Das Vorgehen von Arbeitspaket 4 gliedert sich in zwei Schritte auf:

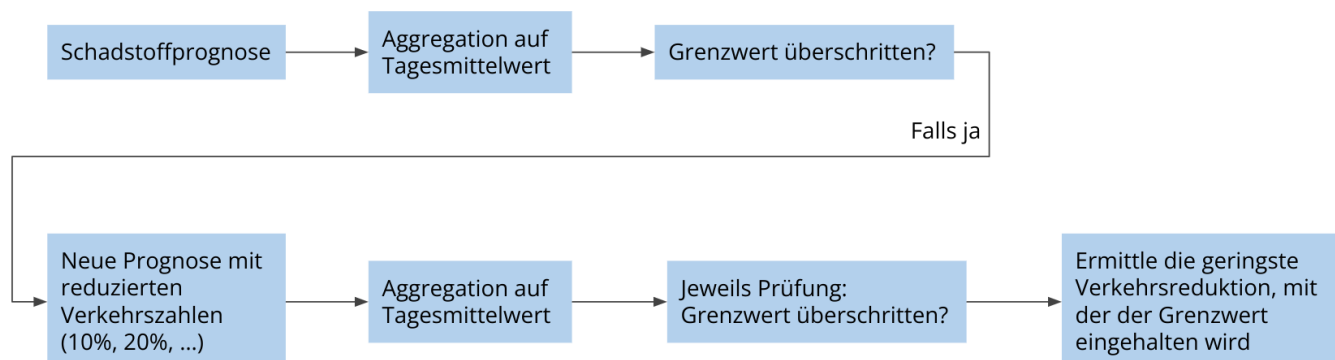
1. Prüfung, ob vorgegebene Schadstoff-Grenzwerte an einem bestimmten Tag überschritten werden,
2. An Tagen mit Überschreitung: Ermittlung der notwendigen Verkehrsreduktion, um eine Überschreitung des Grenzwerts zu verhindern.

Dazu wird zunächst die normale Modellprognose für die Luftschadstoffe erstellt, die in Arbeitspaket 2 entwickelt wurde. Das generelle Vorgehen sowie eine Evaluation sind im Endbericht zum Arbeitspaket 2 beschrieben. Da dieses Modell stündliche Prognosen liefert, die Grenzwerte aber als Tagesmittelwerte formuliert sind, erfolgt im nächsten Schritt die Aggregation der Prognosen auf Tagesmittelwerte. Dann wird geprüft, ob der Tagesmittelwert den erlaubten Grenzwert überschreitet. Falls ja, muss ermittelt werden, welche Verkehrsreduktion die Überschreitung verhindern würde.

Die Verkehrszahlen sind als Input-Variable im Schadstoffmodell enthalten. Daher kann man diese Variable beliebig verändern und die daraus resultierenden Prognosen berechnen. Deswegen werden insbesondere für Tage, für die eine Grenzwertüberschreitung vorhergesagt wird, neue Prognosen mit unterschiedlich stark reduzierten Verkehrszahlen berechnet. Diese Prognosen werden ebenfalls auf Tagesmittelwerte aggregiert. Auf diese Weise lässt sich ermitteln, ab welcher Reduktion der prognostizierte Tagesmittelwert nicht mehr über dem Grenzwert liegt. Der ermittelte maximale Kfz-Anteil ist der Schwellenwert für den Verkehr.

Dabei kann das Ergebnis auch sein, dass sich der Grenzwert an einem Tag voraussichtlich überhaupt nicht einhalten lässt, selbst wenn der Verkehr auf 0% reduziert werden würde.

Die folgende Abbildung fasst das Vorgehen zusammen.



Ausgangsbedingungen

Schadstoff-Grenzwerte

In Absprache mit der Senatsverwaltung erfolgt die Implementierung für die folgenden Schadstoffgrenzwerte (Tagesmittel):

- NO₂: 50 µg/m³
- PM₁₀: 45 µg/m³
- PM_{2,5}: 25 µg/m³

Die Werte sind angelehnt an aktuelle WHO-Empfehlungen¹. Eine nachträgliche Änderung ist in Absprache mit der SenUMVK möglich. Beispielsweise könnten die Werte jährlich angepasst werden, um sich kontinuierlich strengeren Vorgaben anzunähern.

Messstationen

Die Ermittlung für Schwellenwerte ist vor allem für Stationen in verkehrsreichen Gebieten von Interesse. Deswegen wurden die Berechnungen für die folgenden Stationen umgesetzt:

- Schildhornstraße (ID 117)

¹ World Health Organization. (2021). WHO global air quality guidelines: particulate matter (PM_{2,5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. <https://apps.who.int/iris/handle/10665/345329>. Lizenz: CC BY-NC-SA 3.0 IGO



- Mariendorfer Damm (ID 124)
- Silbersteinstraße (ID 143)
- Frankfurter Allee (ID 174)

Mögliche Ansätze

Im Laufe der Entwicklung wurden verschiedene Ansätze für die Simulation ausprobiert und zum Teil verworfen. Als erfolgreich erwies sich schließlich der Ansatz eines zweistufigen Modells. Dieser Ansatz und die zuvor verworfenen Ansätze werden im Folgenden erläutert.

Prognose der Verteilungen statt Punktschätzer

In Arbeitspaket 2 wurde ein Modell zur Prognose der Schadstoffwerte entwickelt. Es handelt sich um ein sogenanntes XGBoost-Modell (<https://xgboost.readthedocs.io/en/stable/>). Die Prognosen dieses Modells sind Punktschätzer, d.h. beispielsweise sagt das Modell für morgen 14 Uhr an einer bestimmten Station einen PM_{10} -Wert von $20 \mu\text{g}/\text{m}^3$ vorher. Der wahre Wert wird sehr wahrscheinlich nicht genau 20 sein, aber bei der 20 handelt es sich um den Erwartungswert.

Eine Erweiterung von XGBoost (XGBoostLLSS, s. <https://github.com/StatMixedML/XGBoostLSS>) ermöglicht die Prognose von Verteilungen statt Punktschätzern. Dabei wird für eine gegebene Verteilung (z.B. Normalverteilung) zusätzlich zum Punktschätzer auch die Streuung prognostiziert, um die Unsicherheit der Prognose zu quantifizieren.

Die erste Idee für die Schwellenwerte war, mit dieser Erweiterung zu arbeiten und somit vorherzusagen, in welchem Bereich sich die wahren Werte mit welcher Wahrscheinlichkeit bewegen. Im nächsten Schritt könnte man die Berechnung an kritischen Tagen für verschiedene Verkehrszahlen wiederholen und ermitteln, bei welcher Verkehrsreduktion die Grenzwerte mit zufriedenstellend hoher Wahrscheinlichkeit eingehalten werden.

Dieser Ansatz wurde verworfen, weil die Prognose auf Stundenbasis erfolgt, sodass eine Aggregation auf Tagesmittelwerte notwendig ist. Verteilungen lassen sich aber - anders als Punktschätzer - nicht einfach mitteln oder aufsummieren, insb. weil die Verteilungen der einzelnen Stundenprognosen nicht unabhängig voneinander sind. Ein



Lösungsansatz für dieses Problem wäre ein komplett neues Modell, das nicht Stunden- sondern direkt Tageswerte vorhersagt. Dies wurde aufgrund des enormen Zusatzaufwands verworfen. Ein anderer Ansatz wird im folgenden Abschnitt erläutert.

Trajektorien

Um die im vorangegangenen Abschnitt erläuterten prognostizierten Verteilungen der Stundenwerte auf einen Tagesmittelwert zu aggregieren, kann man nicht einfach die Verteilungen "mitteln". Dies liegt unter anderem daran, dass die Werte nicht unabhängig voneinander sind. Ist der Wert einer Stunde sehr hoch, so ist auch der Wert der darauffolgenden Stunde mit großer Wahrscheinlichkeit erhöht, was im Modell an den Messstationen auch durch sogenannte Lags (zeitversetzte Variablen, z.B. der Schadstoffwert eine oder 24 Stunden zuvor) berücksichtigt wird.

Um diese Problematik zu umgehen, könnte man für jede Stunde die Verteilung vorhersagen, daraus einen Wert zufällig ziehen und darauf basierend die Verteilung der darauffolgenden Stunde vorhersagen und so weiter. Damit ergibt sich eine Art Pfad für die nächsten Stunden, eine sogenannte Trajektorie. Aus diesen Werten kann man die Tagesmittelwerte berechnen. Der Prozess wird in der Simulation sehr häufig wiederholt, sodass man im Endergebnis eine große Menge Trajektorien und damit eine große Menge Tagesmittelwerte pro Tag erreicht. Auf diese Weise wird die Verteilung der Tagesmittelwerte generiert und man kann berechnen, ob nach dieser Verteilung der Grenzwert wahrscheinlich überschritten wird oder nicht. Da dieser Ansatz sehr aufwändig und rechenintensiv ist, wurde er zunächst zugunsten anderer Ansätze zurückgestellt und letztendlich verworfen, nachdem sich ein anderer Ansatz als geeignet erwiesen hatte.

Direkte Verwendung des normalen Schadstoffmodells

Der erste Ansatz, der testweise implementiert wurde, basiert auf dem "normalen" Schadstoffmodell, welches in Arbeitspaket 2 entwickelt wurde.

Dazu wurden zunächst die Prognosen mit dem vorhandenen Modell erstellt und daraus Tagesmittelwerte berechnet. Für die Tage, deren Mittelwerte den Grenzwert überschreiten, wurden in einer Simulation erneut Prognosen erstellt, wobei die Input-Variable des Verkehrsaufkommens gezielt reduziert wurde. Dadurch lässt sich



berechnen, wie hoch der prognostizierte Schadstoffwert wäre, würde man den Verkehr entsprechend reduzieren.

Erste Ergebnisse zeigten jedoch, dass der rein statistisch ermittelte Einfluss der Verkehrsvariable auf die prognostizierte Schadstoffbelastung im Modell deutlich unter dem physikalisch erwartbaren Einfluss liegt. Damit ist der Einfluss zu schwach, um die prognostizierte Schadstoffbelastung nennenswert zu reduzieren. Dies ist kein Modellfehler, sondern liegt daran, dass im Modell weitere Variablen enthalten sind, die stark mit dem Verkehrsaufkommen zusammenhängen, wie zum Beispiel die Uhrzeit. Das statistische Modell kann an dieser Stelle nicht erkennen, ob eine erhöhte Schadstoffbelastung dem Verkehr oder der Uhrzeit zuzuschreiben ist. Dies ist für die generelle Prognosegüte irrelevant. Es hat aber zur Folge, dass eine Veränderung der Verkehrs-Input-Variablen einen unrealistisch schwachen Effekt auf die Prognose hat. Analog zur Uhrzeit gilt dies auch für die Lags: War der Schadstoffwert in der vorangegangenen Stunden hoch, so ist oft auch die aktuelle Verkehrsbelastung eher hoch (z.B. an Wochentagen um 18 Uhr). Im statistischen Modell haben die Lags einen hohen Erklärungsgehalt, der zum Teil aber genauso gut dem Verkehr zugeschrieben werden könnte.

Selektion der Input-Variablen im normalen Schadstoffmodell

Im vorangegangenen Abschnitt wurde die Problematik erklärt, dass im normalen Schadstoffmodell aus Arbeitspaket 2 dem Verkehr nicht genug Erklärungsgehalt zugeschrieben wurde, sodass Änderungen am Verkehrsaufkommen auch nur einen geringen Einfluss auf die Prognosen hatten. Deswegen wurden in einem ersten Versuch die Variablen, die mit dem Verkehrsaufkommen hoch korreliert sind und daher um Erklärungsgehalt konkurrieren, testweise aus dem Modell entfernt. Sowohl durch das Entfernen der Lags als auch durch das Entfernen der Uhrzeit wurde die Modellgüte jedoch deutlich schlechter, sodass diese Optionen verworfen wurden. In einem weiteren Schritt wurden die Lag-Variablen angepasst: Die direkten Lags wurden nur noch für 24 und 48 Stunden ins Modell aufgenommen. Kurzfristige Lags wurden entfernt und durch den Mittelwert der letzten fünf Stunden ersetzt. Dies verbesserte die Modellprognosen generell, löste aber nicht das Problem des zu geringen Verkehrseinflusses.



Gewicht des Verkehrs in der Modellschätzung erhöhen

Das XGBoost-Modell setzt sich aus vielen kleinen Entscheidungsbäumen zusammen. Jeder dieser Entscheidungsbäume hat isoliert betrachtet nur eine geringe Aussagekraft. Die Stärke des Modells begründet sich daher in der Kombination dieser Bäume.

Bei der Schätzung des XGBoost-Modells können für jede Variable sogenannte Feature Weights definiert werden. Dadurch wird die Wahrscheinlichkeit verändert, dass eine Variable in einem einzelnen Entscheidungsbaum verwendet wird. Allerdings hat die Variable trotzdem nur dann einen Einfluss, wenn sie auch tatsächlich einen Erklärungsgehalt hat.

Dem Verkehrsaufkommen wurde ein höheres Gewicht zugewiesen. Für NO_2 zeigte sich in etwa der gewünschte Effekt: Das Verkehrsaufkommen hatte durch das erhöhte Gewicht einen deutlich höheren Erklärungsgehalt, der zuvor anderen Features (Lags, Uhrzeit, ...) zugeschrieben worden war. Für PM_{10} und $\text{PM}_{2.5}$ zeigte sich jedoch nicht der gewünschte Effekt, sodass dieser Ansatz ebenfalls verworfen wurde.

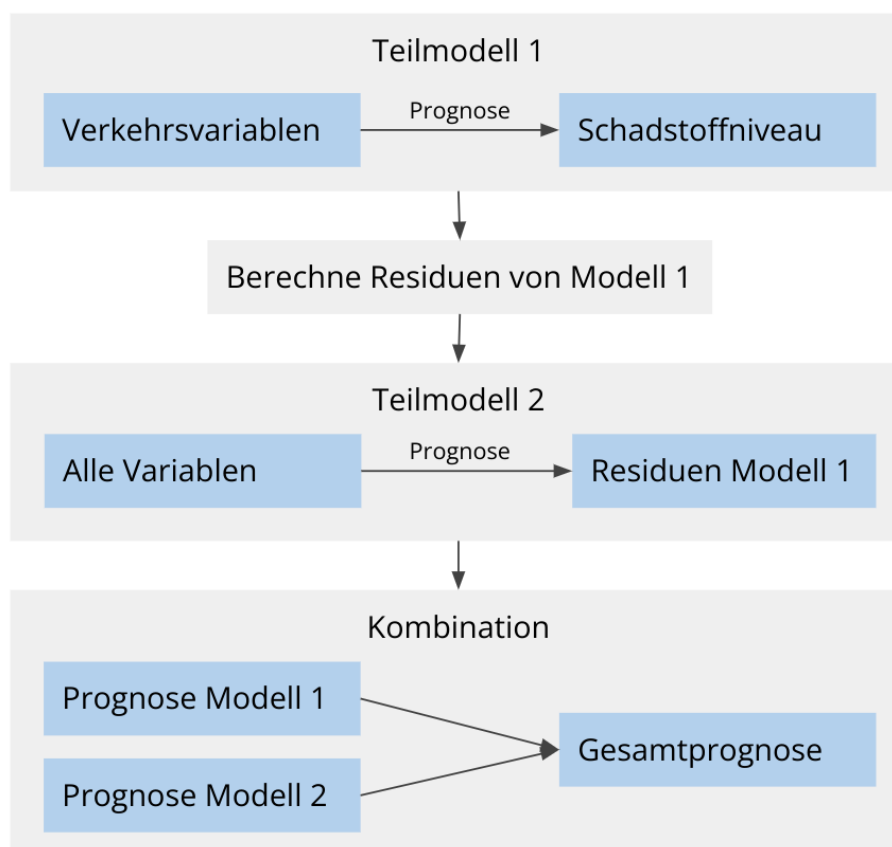
Zweistufiges Modell

Der Ansatz, der letzten Endes umgesetzt wurde und zum Ziel führte, war ein zweistufiges Modell.

In der ersten Stufe werden die Luftschadstoffe mit einem reduzierten Modell prognostiziert, welches nur Verkehrsvariablen, zeitlich konstante Variablen sowie Variablen, die mit dem Verkehr nicht stark korreliert sind, enthält. Dadurch wird sichergestellt, dass die gesamte Variation der Schadstoffbelastung, die durch den Verkehr erklärt werden kann, auch durch diesen erklärt wird. Der Erklärungsgehalt kann nicht der Uhrzeit oder den Schadstoff-Lags zugeordnet werden.

Die Modellgüte dieser ersten Stufe ist im Vergleich zum vollen Modell aus Arbeitspaket 2 natürlich deutlich reduziert. Dies spielt jedoch keine Rolle, denn im nächsten Schritt wird ein zweites Modell berechnet, das alle Variablen enthält und einen möglichst großen Anteil der noch verbleibenden Abweichungen erklärt. Im Detail: Basierend auf den Prognosen der ersten Stufe werden die Abweichungen von den wahren Werten berechnet. Diese Abweichungen nennt man Residuen. Die zweite Stufe hat nicht mehr den Luftschadstoff selbst, sondern die Residuen aus dem ersten Modell als Zielvariable. Die endgültige Prognose dieses zweistufigen Modells erhält man durch Kombination der Prognose aus der ersten und der zweiten Stufe.

Die folgende Abbildung stellt den Aufbau des zweistufigen Modellsystems schematisch dar:



Betrachtet man nur die endgültigen Prognosen, so ist die Modellgüte genau so gut wie bei einem einstufigen Modell. Allerdings wird die gesamte Varianz, die durch den Verkehr erklärt werden kann, auch dem Verkehr zugeschrieben. Simuliert man nun die Prognosen für unterschiedlich stark reduzierte Verkehrszahlen, zeigt sich ein Effekt auf die Schadstoffwerte, der im erwarteten Bereich liegt.

Um den Wartungsaufwand zu reduzieren, wird das zweistufige Modell auch für die normalen, stündlichen Schadstoffprognosen eingesetzt, denn die Modellgüte ist nicht schlechter als beim zuvor verwendeten einstufigen Modell aus Arbeitspaket 2.

Details zur Umsetzung werden im nächsten Abschnitt erläutert, die Ergebnisse im übernächsten.



Umsetzung des zweistufigen Modells: Details

Im vorangegangenen Abschnitt wurde beschrieben, dass nach der Überprüfung verschiedener Ansätze letzten Endes ein zweistufiges Modell verwendet wurde. Auch die generelle Funktionsweise wurde dort beschrieben. In diesem Abschnitt werden die Details der Umsetzung genauer beschrieben.

Variablen im ersten Teilmodell

Die folgenden Variablen wurden ins erste Teilmodell aufgenommen, um möglichst viel Erklärungsgehalt durch den Verkehr abzudecken:

- Fahrzeug-Anzahl
- Fahrzeug-Geschwindigkeit
- Verkehrsmenge (basierend auf dem Detailnetz und der Verkehrsmengenkarte)
- Koordinaten
- Landnutzung, z.B. Wohnnutzung oder Grünfläche
- Emissionen aus dem Emissionskataster
- Windrichtung inkl. Lags
- Windgeschwindigkeit
- Niederschlag inkl. Lags
- Bewölkung

Weitere Variablen außer der Verkehrsvariablen wurden aufgenommen, damit auch Interaktionen zwischen dem Verkehr und diesen Variablen bereits im ersten Teilmodell abgebildet werden können.

Basierend auf den Prognosen des ersten Teilmodells werden die Abweichungen von den wahren Werten, d.h. die Residuen, berechnet. Die zweite Stufe hat nicht mehr den Luftschadstoff selbst, sondern die Residuen aus dem ersten Modell als Zielvariable. Die endgültige Prognose dieses zweistufigen Modells erhält man durch Kombination der Prognose aus der ersten und der zweiten Stufe.



Strategie zur Verkehrsreduktion

Die Schwellenwerte sollen verwendet werden, um verkehrslenkende Maßnahmen zu ergreifen. Da diese Maßnahmen nur tagsüber stattfinden, wurde in der Simulation der Input "Verkehr" immer nur für die Uhrzeiten von 5:00 bis 21:59 Uhr verändert.

Dabei wurde die Reduktion in 10%-Schritten getestet, d.h. die Prognosen wurden für 100% des üblichen Verkehrs, für 90%, 80%, ..., 10% und 0% erstellt. Auch bei 0% Verkehr liegen die Prognosen der Luftschadstoffe selbstverständlich nicht bei 0, weil es weitere Schadstoffquellen und eine gewisse Hintergrundbelastung gibt.

In Abstimmung mit der SenUMVK fiel die Entscheidung für eine prozentuale Reduktion und gegen eine absolute Anzahl von Fahrzeugen. Wann am Tag und mit welchen Maßnahmen diese Reduktion erreicht wird, wird im Rahmen des eUVM-Projektes von SenUMVK entschieden.

Tweak Values: Optimierung der Ergebnisse im entscheidenden Bereich

Das Modell wird auf den kompletten Daten trainiert - sowohl auf sehr hohen Schadstoffwerten als auch auf sehr niedrigen, die weit unter den Grenzwerten liegen. Das Modell wird dadurch so optimiert, dass es insgesamt mit den vorhandenen Variablen möglichst genaue Prognosen generiert. In der Folge variiert die Qualität der Vorhersagen in unterschiedlichen Szenarien. Üblicherweise werden sehr hohe Werte unter- und sehr niedrige Werte überschätzt (sog. Regression zur Mitte), es sei denn, ein Modell liefert perfekte Prognosen.

Für die Schwellenwerte ist von Interesse, dass die Schadstoffprognosen im Bereich der Schadstoffgrenzwerte besonders genau sind. Das heißt, hier zählt ausschließlich, ob Über- oder Unterschreitungen korrekt vorhergesagt werden ("True Positive", bzw. "True Negative"). Ist der Grenzwert eher hoch, so ist zu erwarten, dass das Modell in diesem Bereich die echten Werte tendenziell eher unterschätzt. In diesem Fall wird es tatsächliche Überschreitungen zu selten vorhersagen (sog. "False Negatives"). Umgekehrt sähe es aus, wenn das Modell im Bereich der Grenzwerte eher zu hohe Werte vorhersagt: Dann würde es zu häufig Alarm schlagen (sog. "False Positives"). Um die Prognosen hierfür nochmal zu optimieren, wurde ein fester Wert (sog. "Tweak Value") zu allen Prognosen addiert. Der Tweak Value wurde in Abstimmung mit der



SenUMVK so gewählt, dass die False-Positive-Rate (Anzahl falscher Alarme dividiert durch Anzahl aller Unterschreitungen) und die False-Negative-Rate (Anzahl verpasster Überschreitungen dividiert durch die Anzahl aller Überschreitungen) ungefähr gleich hoch sind. Diese "getweakten" Prognosen wurden dann verwendet, um zu entscheiden, ob der Grenzwert überschritten wird. Es besteht somit ein Trade-Off zwischen der Anzahl an Fehlalarmen und der Anzahl an verpassten Alarmen.

Prognosehorizont

Die Prognosen werden für mindestens zwei Tage erstellt, damit ggf. frühzeitig Maßnahmen ergriffen werden können. Dabei wird jeder Tag separat behandelt, das heißt: Im Szenario für 50% Verkehrsreduktion am zweiten Tag werden diese Tag-2-Prognosen unter der Annahme erstellt, dass an Tag 1 keine Maßnahmen zur Verkehrsreduktion umgesetzt wurden.

Wird in der Praxis dann für Tag 1 bereits eine Maßnahme ergriffen, sollte sich der Effekt in den Luftschadstoffwerten an der Messstation zeigen. Wird schließlich nach Ablauf des Tages die Prognose und Schwellenwert-Simulation für den nächsten Tag (Tag 2) erneut berechnet, startet diese bereits bei einem geringeren Schadstoffniveau und sagt entweder keine Grenzwertüberschreitung mehr vorher oder identifiziert eine schwächere Verkehrsreduktion bereits als ausreichend.

Ergebnisse

Ergebnisformat

Das Ergebnis besteht aus einer Tabelle, die für jeden Prognosetag, für jeden Schadstoff und an jeder der vier ausgewählten Messstationen Vorhersagen mit entsprechend reduzierten Verkehrs-Werten für ein bis drei Tage in die Zukunft liefert. Die folgende Tabelle veranschaulicht das Ergebnisformat anhand fiktiver Beispieldaten:

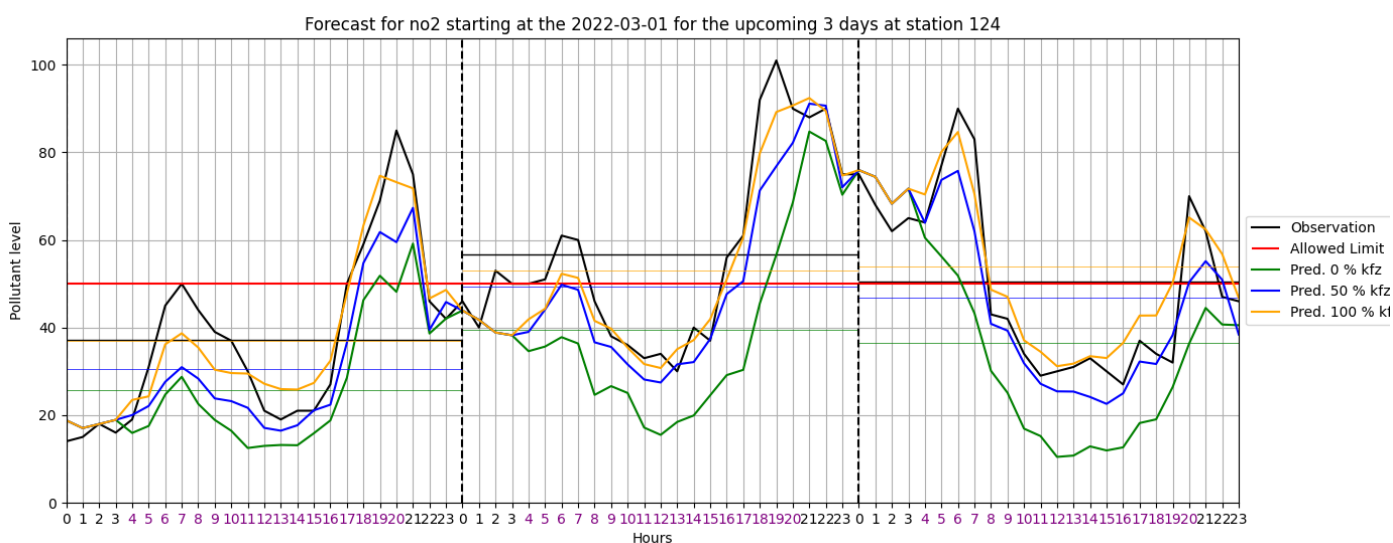


model	date time forecast	date	station_id	tweak value	kfz 100%	kfz 90%	...	kfz 0%
NO ₂	2022-09-01	2022-09-01	174	5.3	55.3	53.2	...	45.1
NO ₂	2022-09-01	2022-09-02	174	5.3	45.7	44.7	...	37.1

Beispiel: Mariendorfer Damm

Im Folgenden wird ein Beispiel für Anfang März am Mariendorfer Damm (ID 124) im Detail beleuchtet.

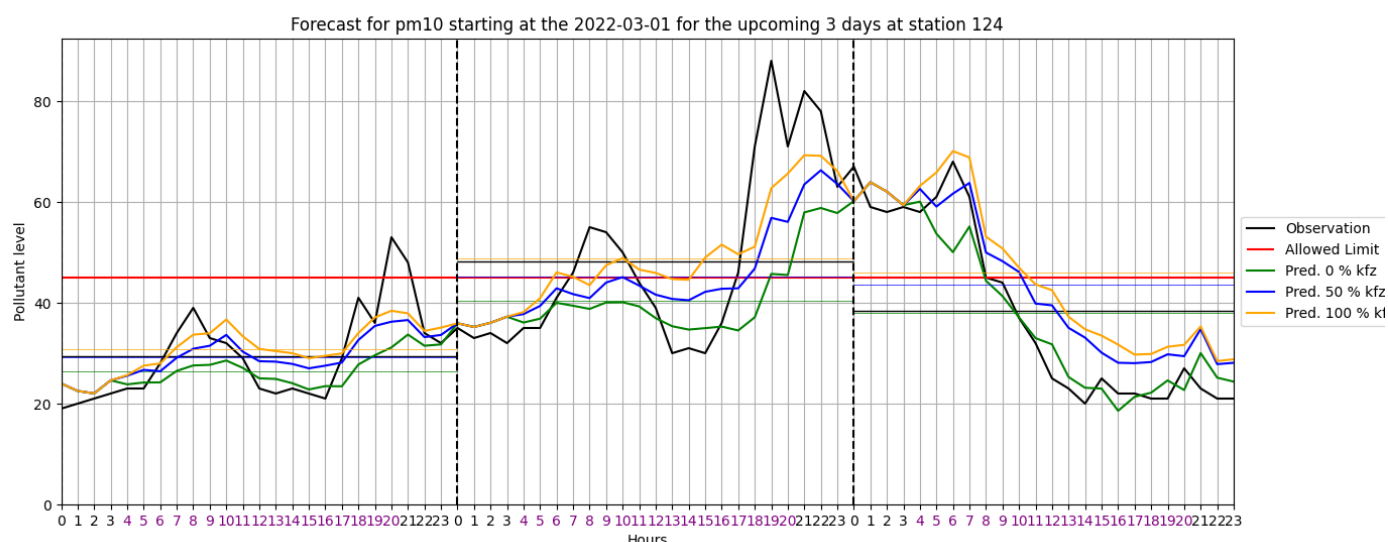
Die folgende Grafik zeigt den Verlauf der (beobachteten und prognostizierten) NO₂-Belastung für drei Tage. Die Vorhersage wurde ausgehend vom 1. März erstellt. Abgebildet sind der 1., 2. und 3. März. Die x-Achse zeigt die Uhrzeit an, wobei Reduktionen lediglich in den lila markierten Stunden vorgenommen werden. Die y-Achse zeigt den NO₂-Wert in µg/m³. Die schwarze Linie ist der beobachtete Verlauf ("Observation"). In Gelb, Blau und Grün sind die Prognosen für 100%, 50% bzw. 0% Verkehr abgebildet. Die weiteren 10%-Zwischenschritte sind der Übersichtlichkeit halber nicht in der Grafik abgebildet. Der Grenzwert ist als waagerechte rote Linie eingezeichnet ("Allowed Limit"). Außerdem sind pro Tag die Mittelwerte (beobachtet, 100% Kfz, 50% Kfz sowie 0% Kfz) als waagerechte Linien eingezeichnet.



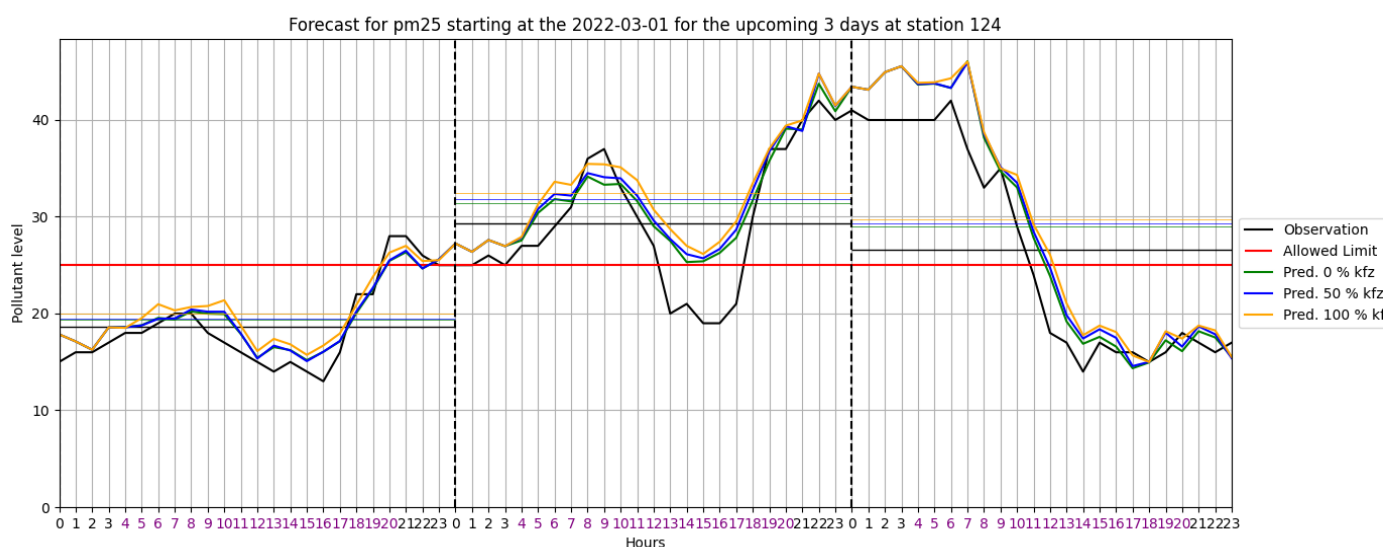
Es ist erkennbar, dass die beobachteten Tagesmittelwerte (schwarz) und die Prognose für 100% Verkehr (gelb) recht dicht beieinander liegen.

Außerdem sieht man für NO₂ eine vorhergesagte und beobachtete Überschreitung des Tagesgrenzwertes von 50 µg/m³ für den 2. März (mittlerer abgebildeter Tag). Durch die Reduktion des Verkehrs auf 50% wird eine Unterschreitung des Grenzwerts vorhergesagt (die waagerechte blaue Linie liegt knapp unter der roten).

Ähnlich sieht es für PM₁₀ aus - auch hier wird für den 2. März eine Überschreitung vorhergesagt, die durch eine Reduktion des Verkehrs auf 50% laut Prognose verhindert werden könnte:



Im Gegensatz dazu gibt es bei PM_{2.5} zwar eine Überschreitung des Grenzwertes von 25 µg/m³, jedoch wird durch die Verkehrsreduktion nur ein marginaler Effekt erzielt. Die gewünschte Verbesserung der Luftqualität in Bezug auf PM_{2.5} kann nicht herbeigeführt werden (d.h. der Tagesmittelwert kann durch eine Verkehrsmaßnahme nicht unter den Grenzwert gebracht werden). Dies ist nicht unerwartet, weil der Verkehr erfahrungsgemäß nur einen kleinen Teil zur PM_{2.5}-Belastung beiträgt.



Ermittelte Tweak Values

Aus dem im Abschnitt "Tweak Values" beschriebenen Verfahren ergeben sich folgende Werte, die als "Sicherheitsabstand" zum prognostizierten Tagesmittelwert addiert werden, um zu entscheiden, ob dieser oberhalb eines Grenzwerts liegt oder nicht.

- NO_2 : $5.3 \mu\text{g}/\text{m}^3$
- $\text{PM}_{2.5}$: $2.7 \mu\text{g}/\text{m}^3$
- PM_{10} : $5.0^2 \mu\text{g}/\text{m}^3$

Diese Werte werden im Rahmen eines regelmäßigen Modell-Reviews (Arbeitspaket 8) überprüft und ggf. angepasst.

Anmerkung: Die Werte werden nicht generell zu den Prognosen addiert, sondern lediglich für die Berechnungen der Tagesmittelwerte mit unterschiedlichen Verkehrs-Reduktionen verwendet.

² Das Verfahren ergab einen Wert von 10.3, der in Absprache mit der SenUMVK auf einen weniger extremen Wert von 5.0 gesetzt wurde.



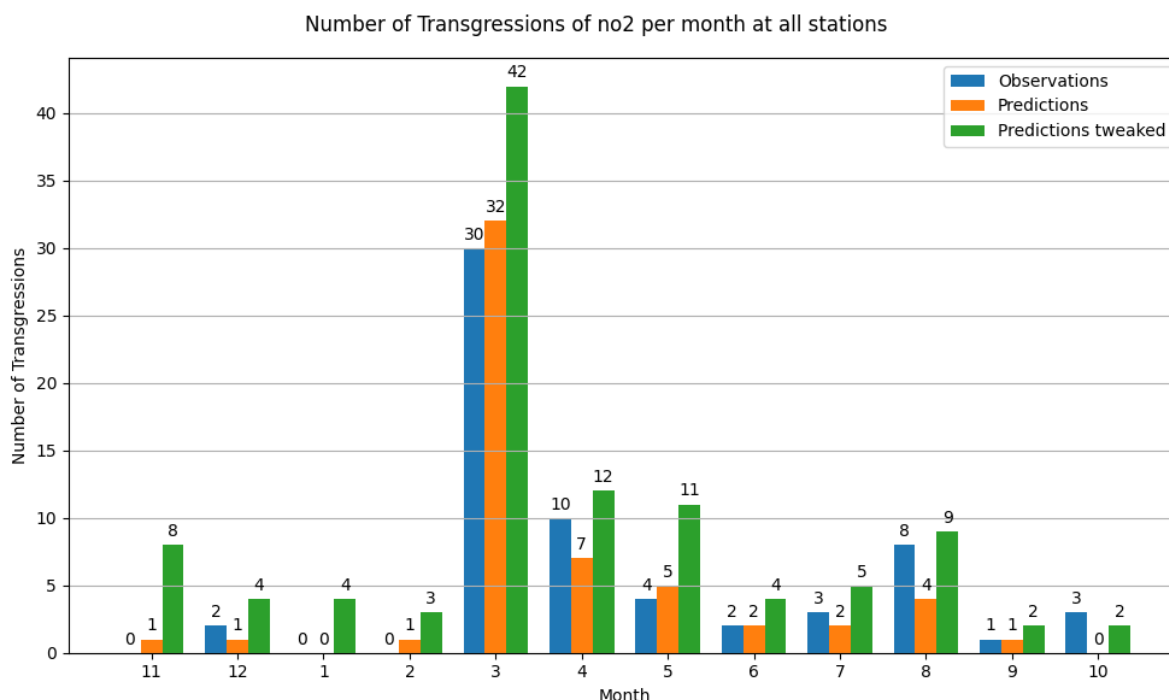
Überschreitungen in der Jahresübersicht

Um die Plausibilität der Prognosen zu prüfen, wurde für das vergangene Jahr (November 2021 bis Oktober 2022) rückwirkend berechnet, für welche Tage Überschreitungen vorhergesagt sowie beobachtet wurden.

Am Beispiel von NO₂ ist im Folgenden ein Überblick über die Anzahl der Grenzwertüberschreitungen pro Monat abgebildet, die im letzten Jahr an den selektierten Stationen

- beobachtet ("Observations", blau),
- vorhergesagt ("Predictions", orange),
- und unter Berücksichtigung der Tweak Values vorhergesagt ("Predictions tweaked", grün)

wurden.



Man erkennt sowohl bei den Vorhersagen als auch bei den Beobachtungen mehr Überschreitungen zwischen März und August, wobei besonders der März hervorsticht. Ähnliche Muster ergeben sich für PM_{2.5} und PM₁₀.

Generell ähnelt die Monatsverteilung der prognostizierten Überschreitungen der Monatsverteilung der tatsächlich beobachteten Überschreitungen.



Außerdem ist erkennbar, dass das generelle Niveau der Anzahl von Überschreitungen bei der Berücksichtigung der Tweak Values (grün) wie erwartet höher ist als bei den unveränderten Vorhersagen (orange). Auch ist es höher als das Niveau der beobachteten Vorhersagen, was erwünscht ist, um möglichst wenige Überschreitungen zu verpassen. Details dazu werden im nachfolgenden Abschnitt erläutert.

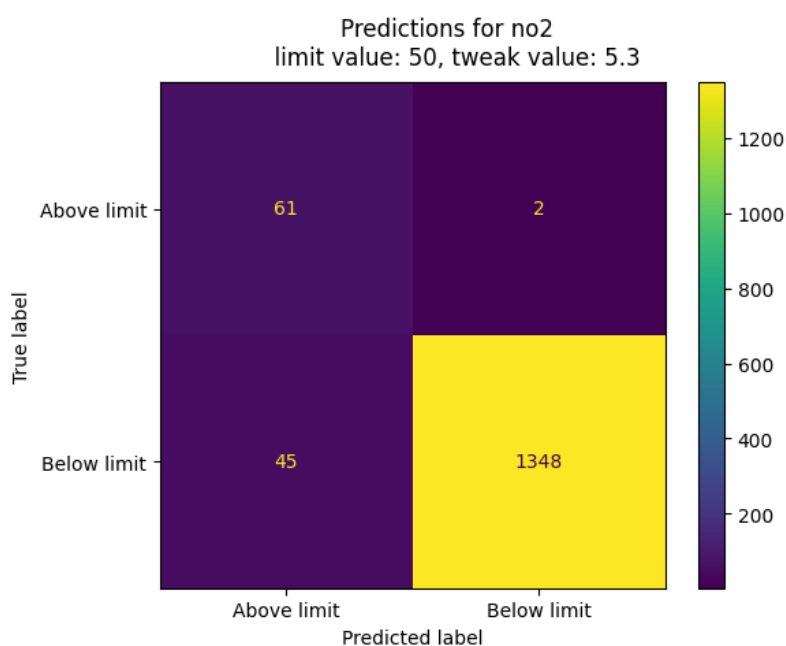
Genauigkeit: verpasste Überschreitungen vs. falsche Alarmer

Im Folgenden wird am Beispiel von NO₂ untersucht, wie häufig Vorhersagen über- oder unterhalb des Grenzwertes lagen und ob diese Vorhersagen korrekt waren.

Datenbasis sind auch hierfür die rückwirkenden Prognosen an den vier ausgewählten Stationen im Zeitraum November 2021 bis Oktober 2022. Ein Tag an einer Station ist eine Beobachtungseinheit.

Die folgende Abbildung zeigt:

- die Anzahl der korrekt vorhergesagten Grenzwertüberschreitungen (oben links)
- die Anzahl der verpassten Grenzwertüberschreitungen (oben rechts)
- die Anzahl der falschen Alarmer (unten links)
- die Anzahl der korrekt vorhergesagten eingehaltenen Grenzwerte (unten rechts)





Von besonderem Interesse ist hier die Minimierung der "False Negatives" (oberes rechtes Feld), d.h. der verpassten Grenzwertüberschreitungen. Die Ergebnisse basieren auf den Berechnungen mit dem zuvor bestimmten Tweak Value von 5,3 für NO₂. Damit gibt es im ganzen Jahr an allen vier Stationen nur noch zwei Überschreitungen, die nicht prognostiziert werden. Diese Genauigkeit "erkauft" man sich mit einer erhöhten Zahl von "False Positives" (untere linke Ecke, falscher Alarm) von 45. Diese falschen Alarme sind in Absprache mit der SenUMVK hinnehmbar. Die Auswertung wurde für die vier untersuchten Standorte (IDs 117, 124, 143 und 174) durchgeführt, sodass insgesamt im ganzen untersuchten Jahr ca. 11 verpasste Überschreitungen ("False Negatives") zustande kamen, was wiederum nahe am einzuhaltenden Grenzwert lag.

Eine detaillierte Betrachtung der beobachteten und prognostizierten Tagesmittelwerte sowie der theoretisch benötigten Verkehrsreduktion zeigte unter anderem:

- Bei falschen Alarmen liegt der tatsächliche Tagesmittelwert in der Regel nur knapp unter dem erlaubten Grenzwert, sodass ggf. umgesetzte verkehrslenkende Maßnahmen trotzdem einen gesundheitlichen Nutzen erwarten lassen.
- Überschreitungen des NO₂-Tagesmittelwertes lassen sich in der Regel durch Verkehrsreduktionen theoretisch abwenden (auch wenn diese teilweise sehr groß sein müssten). Bei PM₁₀ ist dies nur teilweise der Fall. Bei PM_{2,5} lassen sich die Überschreitungen aufgrund der begrenzten Relevanz des Verkehrs üblicherweise laut Prognose nicht durch eine Verkehrsreduktion abwenden.

Speicherung und Bereitstellung der Schwellenwerte

Täglich wird berechnet, ob der Tagesgrenzwert an den kommenden vier Tagen überschritten wird. Falls ja, werden die Schwellenwerte berechnet (d.h. die notwendige Verkehrsreduktion). Alle Ergebnisse werden in eine interne Datenbank geschrieben. Die Umsetzung dieser Automatisierung erfolgt im Rahmen von Arbeitspaket 5.

Die Datenbanken liegen auf Servern, die vom Cloud-Anbieter Hetzner gehostet werden. Es handelt sich um Clickhouse-Datenbanken, einem SQL-basierten Datenbanksystem, das besonders gut für große Datenmengen geeignet ist. Die SQL-Statements zum initialen Aufsetzen der Datenbanken werden gesammelt und mit dem Code hinterlegt.



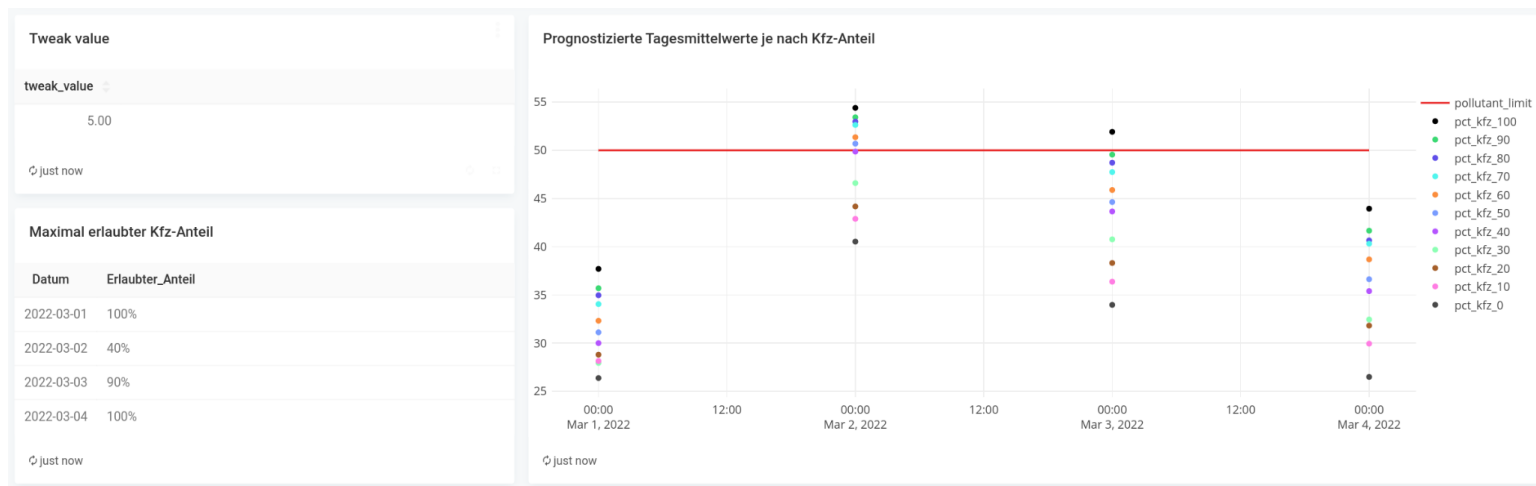
Die Datenbank ist in drei sogenannte Schemata unterteilt:

- fairq-raw: Rohdaten, die nur wenigen Verarbeitungsschritten unterzogen wurden
- fairq-features: stärker aufbereitete Daten, die in die Schadstoff-Modellierung eingehen ("Features") bzw. Zwischenschritte, um die Daten für die Modellierung zu berechnen
- fairq-output: Modellergebnisse, die später nach außen zur Verfügung gestellt werden

Im Rahmen von Arbeitspaket 4 sind ausschließlich die Schemata fairq-features und fairq-output relevant. Aus fairq-features stammt der Modell-Input; die Ergebnisse werden in fairq-output geschrieben.

Ein direkter Zugriff auf die Datenbank von außen, z.B. durch die SenUMVK oder weitere Dienstleister, ist nicht möglich und unter anderem aus Sicherheitsgründen auch nicht empfehlenswert. Um von außen auf die Schwellenwerte zugreifen zu können, werden diese im Rahmen von Arbeitspaket 7 (Schnittstellen) als weiterer Endpunkt bereitgestellt. Außerdem wird der SenUMVK im Rahmen von Arbeitspaket 7 Zugriff auf ein Dashboard gewährt, das die neuesten Schwellenwert-Prognosen anzeigt.

Der folgende Screenshot zeigt einen Ausschnitt des Dashboards mit dem eingangs verwendeten Datenbeispiel der NO₂-Werte am Mariendorfer Damm Anfang März 2022. Der Grenzwert ist in der Abbildung als rote Linie eingezeichnet. Die prognostizierten Tagesmittelwerte (für 4 Tage) für verschiedene Kfz-Anteile sind als bunte Punkte eingetragen. Im Dashboard ist die Grafik interaktiv, sodass bei Mouseover die exakten Werte zu jedem Punkt eingeblendet werden. Links neben der Grafik sind in zwei Tabellen der zugrundeliegende Tweak Value und die ermittelten maximalen Kfz-Anteile angegeben.



Code und Code-Dokumentation

Sämtlicher Code ist von Anfang an auf der Plattform Github (Repository: fairq-model) hinterlegt, welche auch für die Versionskontrolle des Codes und Code-Reviews verwendet wird. Alle Funktionen sind dokumentiert, um anderen Entwickler*innen die Wartung und Weiterentwicklung zu erleichtern. Außerdem sind automatisierte Tests auf korrekte Funktionsweise sowie automatisierte Tests des Code-Styles (z.B. Formatierung, Einrückungen etc.) eingerichtet, um eine hohe Qualität sicherzustellen.

Im Laufe des Gesamtprojekts werden die Repositories mit Readmes ausgestattet, um Entwickler*innen den Einstieg zu erleichtern. Außerdem ist geplant, alle Repositories open source zur Verfügung zu stellen.

Ansätze zur Übertragung auf Straßen ohne Messcontainer

Im Rahmen von Arbeitspaket 4 wurde die Schwellenwert-Berechnung nur auf vier ausgewählte Messcontainer an Hauptverkehrsstraßen angewendet. Die Schwellenwerte sind aber auch für Hauptverkehrsstraßen von Interesse, an denen keine Messcontainer stehen, damit auch dort verkehrslenkende Maßnahmen umgesetzt werden können.



Die bisher umgesetzte Logik lässt sich auch auf die Prognosen auf dem 50x50m²-Grid und somit auch auf die darauf aufbauenden Prognosen für die Straßen übertragen. Das Vorgehen ist dann folgendermaßen:

- Prognosen auf dem Grid erstellen (mit Hilfe eines zweistufigen Modells)
- Darauf basierend die Prognosen für die relevanten Straßen berechnen
- Jeweils prüfen, ob der Schadstoff-Grenzwert überschritten wurde
- Falls ja: Für alle Grid-Zellen, die in die Prognose der Straße einfließen, die hypothetischen Prognosen mit reduziertem Verkehrsanteil berechnen
- Darauf basierend die Prognosen für die Straßen mit reduziertem Verkehrsanteil

Eine Herausforderung dabei ist es, den Effekt zwischenzeitlich umgesetzter verkehrslenkender Maßnahmen miteinzubeziehen. Konkret geht es um folgendes Szenario:

- Für Tag 2 wird eine Grenzwertüberschreitung vorhergesagt
- An Tag 1 wird bereits der Verkehr reduziert

Im Modell an den Messstationen werden - wie weiter oben beschrieben - die Schadstoff-Lags berücksichtigt, sodass eine Verkehrsreduktion an Tag 1 automatisch die Prognose von Tag 2 beeinflusst. Sofern die Maßnahme an Tag 1 erfolgreich die Schadstoffbelastung gesenkt hat, wird somit auch die vorhergesagte Belastung an Tag 2 niedriger ausfallen. Das Modell auf dem Grid, auf dem auch die Prognosen für die Straßen basieren, enthält jedoch keine Schadstoff-Lags, weil diese mangels Messstationen nicht bekannt sind. Es ist allerdings wünschenswert, dass eine Verkehrsreduktion an Tag 1 auch hier einen Effekt auf die Vorhersage für Tag 2 hat. Dazu sind verschiedene Ansätze denkbar, zum Beispiel:

- *Reduktion der Prognosen aufgrund von Erfahrungswerten (pauschal):* Basierend auf den Daten an den Messstationen kann simuliert werden, wie stark eine Verkehrsreduktion um x% an Tag 1 die Prognosen an Tag 2 reduziert. Die auf diese Weise ermittelten Faktoren könnten als Korrekturfaktor für die Tag-2-Prognosen des Tagesmittelwerts auf dem Grid und an den Straßen verwendet werden.
- *Reduktion der Prognosen aufgrund von Erfahrungswerten (detailliert):* Statt für eine gegebene Verkehrsreduktion an Tag 1 immer denselben konstanten Korrekturfaktor für Tag 2 zu verwenden, kann dieser auch mit Hilfe eines Modells vorhergesagt werden. Es ist beispielsweise denkbar, dass der Zusammenhang



zwischen den beiden Tagen schwächer ausfällt, wenn an Tag 2 starker Wind weht, der die Schadstoffe ohnehin zu großen Teilen wegweht.

- *Aufnahme von Verkehrs-Lags:* Diese stehen im ganzen Berliner Stadtgebiet für beliebige Zeitpunkte zur Verfügung und können das Fehlen der Schadstoff-Lags teilweise kompensieren. So könnte berücksichtigt werden, wie sich eine Reduktion des Verkehrs um x% an Tag 1 auf die Prognose für Tag 2 auswirkt.
- *Übertragung der Schadstoff-Lags auf das Grid:* In einem ersten Schritt könnten die initialen Lags für das Grid berechnet werden, zum Beispiel mit dem bisher verwendeten Modell für die Prognosen auf dem Grid. Das Modell würde dann im laufenden Betrieb stets alle notwendigen Lags liefern, z.B. bis 48 Stunden. So könnte in einem zweiten Schritt das Modell mit Lags, das bisher an den Stationen verwendet wurde, auch auf beliebige Gridzellen ohne Messstation angewendet werden. Darauf basierend werden dann die Prognosen für die Straßen berechnet. Der Rest der Schwellenwertberechnung würde dann wie an den Messstationen erfolgen.

Diese Ansätze müssen - analog zu den oben im Detail beschriebenen unterschiedlichen Ansätzen für die bisherige Schwellenwertberechnung - zunächst in einem agilen Prozess getestet werden, um zu evaluieren, ob sie den gewünschten Effekt zeigen.

So muss beispielsweise erst untersucht werden, ob die Verkehrs-Lags im Modell eine ausreichend hohe Relevanz haben, um die Effekte verkehrslenkender Maßnahmen abbilden zu können. Im ursprünglich verwendeten ersten Modell zeigte sich eine sehr geringe Relevanz der Verkehrs-Lags; daher muss geprüft werden, ob sich die Relevanz durch die Zweiteilung ausreichend erhöht hat.

Bei der Übertragung der Schadstoff-Lags auf das Grid besteht die Möglichkeit, dass die Nutzung solcher hypothetischer Lags nicht zu einer tatsächlichen Verbesserung der Modellgüte führt. An den Stationen haben die Lags den Effekt, dass die Vorhersage mit Hilfe der tatsächlich beobachteten Werte noch einmal verbessert wird. Insbesondere kurzfristige Veränderungen im Schadstoffniveau, die das Modell nicht vorhergesehen hat, werden dadurch - mit leichter Verzögerung - doch noch abgebildet. Das kann im Grid nicht garantiert werden, weil es sich bei den Lags nicht um echte Beobachtungen handelt.

Die Priorisierung und Umsetzung erfolgt in Abstimmung mit der SenUMVK. Zur Abstimmung wird ein erstes Treffen Anfang Januar 2023 angestrebt.