

**Note:** Please find the English information about the air quality data on the next page.

## Ortsgebundene MOS- Luftqualitätsvorhersagen

**Kontakt für Rückfragen:** anke.kniffka@dwd.de

Vorhersagen der Stundenmittelwerte von Stickstoffdioxid (**NO<sub>2</sub>**), Ozon (**O<sub>3</sub>**) und Feinstaub (**PM<sub>10</sub>** und **PM<sub>2.5</sub>**) werden mit einer **Vorlaufzeit von 96 h** und einer zeitlichen Auflösung von 1h an etwa 450 Orten bereitgestellt. Zur Berechnung der ortsgebundenen Vorhersage wird der Ansatz des statistischen Postprocessings MOS = Model-Output-Statistics genutzt. Dabei wird die numerische CAMS-Luftqualitätsvorhersage (CAMS=Copernicus Atmospheric Monitoring Service) für einen Trainingszeitraum von mehreren Jahren mit ortsspezifischen Messungen und meteorologischen Vorhersagen korreliert, um Vorhersagegleichungen zu bestimmen. Die Vorhersagen werden stündlich neu berechnet und bereitgestellt, wobei in jeden Vorhersagelauf aktuelle Messwerte einfließen. Zusätzlich wird in der 10 UTC-Vorhersage die verwendete CAMS-Luftqualitätsvorhersage und in der 9 UTC sowie 21 UTC-Vorhersage die Vorhersagen für die meteorologischen Verhältnisse aktualisiert. Es werden zwei Ausgabeformate bereitgestellt: 1.) die 96-stündigen Zeitreihen und 2.) die zugehörigen Mittelwerte der nächsten drei Tage für alle Parameter und Stationen.

## Aufbau und Inhalt der Dateien

1.) Die Zeitreihen „**lq\_forecast\_YYYYMMDDHH.csv**“:

- Die Vorhersagen werden stündlich neu berechnet. Aus dem Dateinamen "lq\_forecast\_YYYYMMDDHH.csv" geht der Zeitpunkt des jeweiligen Vorhersagelaufes in UTC hervor. (UTC+1 ≡ MEZ; UTC+2 ≡ MESZ).
- Schadstoffkonzentrationen werden in µg/m<sup>3</sup> angegeben.
- Fehlwerte tragen den Wert „-999“
- Die Datentabelle ist wie folgt aufgebaut:
  - Erste Spalte - "Station": Kürzel der Station
  - Zweite Spalte - "Komponente": Luftschadstoff. Vorhergesagt werden die Stundenmittelwerte von Stickstoffdioxid, Ozon, PM<sub>10</sub>, PM<sub>2.5</sub> (PM2\_5) sowie das gleitende 24h-Mittel für PM<sub>10</sub> und PM<sub>2.5</sub>.
  - Dritte Spalte - "Datum": Zeitpunkt des aktuellen Vorhersagelaufes in UTC im Format YYYYMMDDHH
  - Vierte Spalte - "Lat.": Breitengrad des Vorhersageortes (in Dezimalgrad)
  - Fünfte Spalte - "Lon.": Längengrad des Vorhersageortes (in Dezimalgrad)
  - Sechste Spalte - "-01h": Letzter Messwert an dieser Station. Hierfür wird die Messung 1h vor dem Vorhersagelauf genutzt; z.B. bei einem Vorhersagelauf um 10 UTC stammt der letzte Messwert von 9 UTC (also von Stunde -01h).



- ab Spalte 7: Stündliche Vorhersagewerte in  $\mu\text{g}/\text{m}^3$  ab Stunde +01h bis zu 4 Tage (+96h) relative zum Zeitpunkt des Vorhersagelaufes.

2.) Die Mittelwerte „**lq\_average\_allstats\_YYYYMMDDHH.csv**“:

- Die Mittelwerte werden auf Basis der stündlich erneuerten Vorhersage ebenfalls stündlich neu berechnet. Sie zeigen die Mittelwerte der nächsten drei Tage, jeweils von 01 UTC bis 24 UTC.
- Aus dem Dateinamen "lq\_average\_allstats\_YYYYMMDDHH.csv" geht der Zeitpunkt des jeweils genutzten Vorhersagelaufes in UTC hervor.
- Schadstoffkonzentrationen werden in  $\mu\text{g}/\text{m}^3$  angegeben.
- Fehlwerte tragen den Wert „-999“
- Die Datentabelle ist wie folgt aufgebaut:
  - Header in erster Zeile: Überschriften der Spalten
  - Erste Spalte – „Komponente“: Name der vorhergesagten Größe. Z.B. Stickstoffdioxid
  - Zweite Spalte – „Station“: Kürzel der Station
  - Dritte Spalte – „TimeStamp“: Datum und Stunde im Iso-format in UTC
  - Vierte bis sechste Spalte – „Mittel1,2 und 3“: Mittel von Tag eins, zwei und drei

## Location-based MOS air quality forecast

**Contact:** [anke.kniffka@dwd.de](mailto:anke.kniffka@dwd.de)

Predictions of the hourly mean concentration of nitrogen dioxide (**NO<sub>2</sub>**), ozone (**O<sub>3</sub>**) and particulate matter (**PM<sub>10</sub>** and **PM<sub>2.5</sub>**) are provided with a **lead time of 96 hours** and a temporal resolution of 1 hour at approx. 450 locations. The statistical post-processing approach MOS = Model-Output-Statistics is used to calculate the location-based forecast. In order to determine the location-based forecast equations, the numerical CAMS air quality forecast (CAMS = Copernicus Atmospheric Monitoring Service) is correlated with location-specific measurements and meteorological forecasts for a training period of several years. The forecasts are recalculated and provided on an hourly basis, with current air quality measurement being incorporated into each forecast run. In addition, the MOS system is updated every day in the 10 UTC forecast with latest CAMS air quality forecast; Likewise, the latest forecasts for the meteorological conditions are included in the 9 UTC and 21 UTC forecasts. The results are distributed in two formats: 1.) as 96h-forecast time-series for each parameter and station and 2.) as the averages for day one, two and three for all parameters and stations.



## Structure and content of the data files

### 1.) Time-series "lq\_forecast\_YYYYMMDDHH.csv":

- The forecast is calculated every hour, giving the date and time of the current run in the filename in UTC: "lq\_forecast\_YYYYMMDDHH.csv"
- The unit of pollutant concentrations is  $\mu\text{g}/\text{m}^3$ .
- Missing values are marked with "-999".
- The structure of the data file is:
  - First column – "Station": Station-ID
  - Second column – "Komponente": Pollutant. Provided are the predicted hourly mean values for NO<sub>2</sub> (Stickstoffdioxid), O<sub>3</sub> (Ozon), PM<sub>10</sub> (PM10), PM<sub>2.5</sub> (PM2\_5) and the 24h-mean value of PM<sub>10</sub> and PM<sub>2.5</sub>.
  - Third column – "Datum": Date of the forecast run in UTC in the format YYYYMMDDHH
  - Fourth column – "Lat.": Latitude of the forecast location (in decimal degree)
  - Fifth column – "Lon.": Longitude of the forecast location (in decimal degree)
  - Sixth column – "-01": Last measured concentration at this location. The most recent considered concentration in each forecast run is measured one hour before the run. E. g. for the 10UTC run, the last considered observation is from 9 UTC (means from hour -01h).
  - From column 7: forecast in  $\mu\text{g}/\text{m}^3$  for every hour from +01h to +96h (4 days) relative to the date of the forecast run (given in column 3).

### 2.) Averages of day one, two and three "lq\_average\_allstats\_YYYYMMDDHH.csv":

- Averages are computed once per hour based on the hourly refreshed forecasts. They contain the averages of the next three days, each from 01 UTC to 24 UTC.
- The filename shows the date of the forecast used to calculate the averages.
- The unit of pollutant concentrations is  $\mu\text{g}/\text{m}^3$ .
- Missing values are marked with "-999".
- The csv-files are structured as follows:
  - Header in first line contains column headers
  - First column "Komponente": Name of the pollutant
  - Second column "Station": Station-ID
  - Third column "TimeStamp": Time stamp in isoformat in UTC
  - Column four to six "Mittel1", "Mittel2", "Mittel3": average pollutant concentration of day one, two and three for each pollutant and station

