

Smart emission: O₃

Pieter Marsman

May 12, 2016

Dit document beschrijft welke stappen er moeten worden genomen om vanuit de metingen van Jose de O₃ meting van het RIVM te voorspellen.

Stap 1: variabelen berekenen en selecteren.

In twee stappen is een set van variabelen verkregen die iets zegt over de O₃ metingen van het RIVM.

Allereerst zijn er op basis van de weerstandswaarden van Jose nieuwe variabelen berekent. Voor elke weerstandswaarde is de logaritmische en wortel ervan toegevoegd aan de data. Daarna zijn alle producten van (op) weerstandswaarden (gebaseerde variabelen) toegevoegd. Hierdoor kunnen ook interacties tussen verschillend gasstoffen meegenomen worden.

Ten tweede is er een selectie gemaakt van de variabelen. Allereerst zijn alleen de sensor waarden van Jose gebruikt en niet de status waarden. Daarnaast zijn tijdsmaten, gps, licht en regen metingen verwijderd. Op de overige variabelen is een forward selectie met een linear model toegepast. Hierbij wordt begonnen met een klein linear model en wordt stapsgewijs de variabele toegevoegd die het residu het meest verkleint. Het residu is het verschil tussen de voorspelling en de daadwerkelijke meting van het RIVM voor O₃.

Een aantal verschillende variabelen bleek interessant:

- $s.coresistance * \log(s.o3resistance)$
- $s.light.sensor.bottom$
- $s.coresistance * s.temperature.ambient$
- $s.temperature.ambient * \log(s.coresistance)$
- $\log(s.coresistance) * \sqrt{s.coresistance}$
- $s.coresistance * s.no2resistance$
- $s.no2resistance * \log(s.o3resistance)$
- $s.barometer * s.coresistance$
- $s.humidity$
- $s.temperature.unit * \sqrt{s.no2resistance}$

Stap 2: variabelen filteren

In de weerstandswaarden van O₃, CO en NO₂ zit veel ruis. Dit betekent dat opeenvolgende metingen erg anders kunnen zijn. Door de ruis te verminderen kan de voorspelling verbeterd worden.

De ruis kan verminderd worden door een voortschrijdend gemiddelde te berekenen. Een voortschrijdend gemiddelde is het gemiddelde van de laatste x aantal metingen. Om x te bepalen werden verschillende waarden geprobeerd. Hieruit blijkt dat voor alle gassensoren 350 een goede waarde is. Dit komt overeen met een gemiddelde over 6 uur.

Stap 3: model leren

Er is een linear model geleerd dat vanuit de 11 onafhankelijke variabelen (inclusief intercept) de afhankelijke variabele O₃ gemeten door het RIVM voorspeld. Alle 11 variabelen voorspellen een significant deel.

Uit het lineaire model komt de volgende rekensom:

$$\begin{aligned} O_3 &= 89.1177 \\ &+ 0.03420626 * s.coresistance * \log(s.o_3resistance) \\ &- 0.008836714 * s.light.sensor.bottom \\ &- 0.02934928 * s.coresistance * s.temperature.ambient \\ &- 1.439367 * s.temperature.ambient * \log(s.coresistance) \\ &+ 1.26521 * \log(s.coresistance) * \sqrt{s.coresistance} \\ &- 0.000343098 * s.coresistance * s.no_2resistance \\ &+ 0.02761877 * s.no_2resistance * \log(s.o_3resistance) \\ &- 0.0002260495 * s.barometer * s.coresistance \\ &+ 0.0699428 * s.humidity \\ &+ 0.008435412 * s.temperature.unit * \sqrt{s.no_2resistance} \end{aligned}$$

Stap 4: betrouwbaarheid beoordelen

De betrouwbaarheid wordt beoordeeld aan de hand van de Root Mean Squared Error (RMSE). De RMSE is de standaard deviatie van de error. Een lage RMSE is dus wenselijk. De RMSE van deze voorspelling is 6.1 microgram / m³.

De O₃ waarde gemeten door het RIVM varieert tussen de 0 en 75 microgram / m³. De error is dus klein in vergelijking met de spreiding in de data.

De luchtkwaliteitsindex van het RIVM wordt uitgedrukt in hoofdklassen en subklassen. Voor Ozon bestaan de hoofdklassen uit groepen van minstens 40 microgram / m³ en de subklassen uit groepen van minstens 10 microgram / m³. De hoofdklassen zijn dus met zekerheid vast te stellen en de subklassen met enige zekerheid.

