

Corona uit de riooldata filteren

De coronapandemie is voorbij, maar het Nederlandse RIVM houdt nog steeds het aantal deeltjes coronavirus in ons rioolwater bij. Door natuurlijke variaties pieken die waarden soms. Kan de Studiegroep Wiskunde met de Industrie bepalen of dat toevallige meetafwijking zijn, of het teken van een nieuwe uitbraak?

Sinds 5 mei 2023 beschouwt de wereldgezondheidsorganisatie WHO COVID-19 niet meer als pandemie. Toch worden er nog steeds mensen geïnfecteerd, soms met de jarenlang slepende aandoening *long covid* als gevolg. Ook nieuwe varianten van het virus zijn de moeite om in de gaten te houden. Daarom houdt het Nederlandse RIVM (Rijksinstituut voor Volksgezondheid en Milieu) de vinger aan de pols van corona.

"Een van de manieren waarop je de verspreiding van het coronavirus kunt meten is via het aantal virusdeeltjes in het rioolwater", vertelt Wouter Hetebrij. Na zijn promotie in de wiskunde werd hij modelleur bij het RIVM. "Even omschakelen; je moet toch een hoop achtergrond opdoen. Maar daarna kon ik mijn vaardigheden prima toepassen."

Bijvoorbeeld op de analyse van coronavirusdeeltjes in het rioolwater. Op diezelfde manier wordt al gemonitord op de ziekte polio en naar bacteriën die ongevoelig zijn voor antibiotica. Het levert een interessant vraagstuk op voor de Studiegroep Wiskunde met de Industrie (SWI), waarvan op voorhand niet duidelijk is welke techniek de beste oplossing zal geven.

Metingen teruggeschroefd

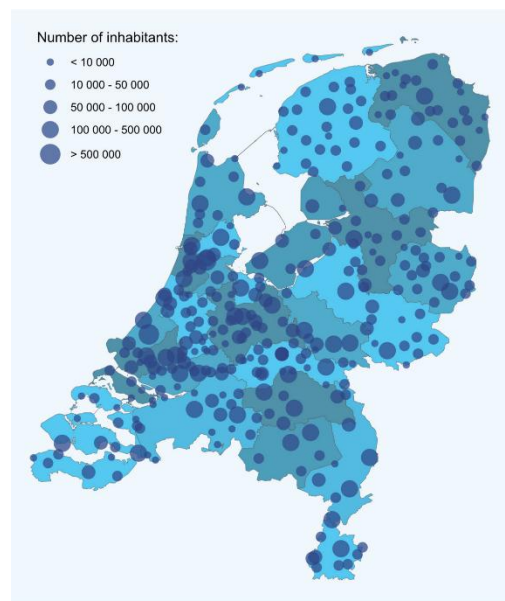
Op het hoogtepunt van de coronapandemie namen de Nederlandse waterschappen vier keer per week monsters bij de ruim 300 rioolwaterzuiveringsinstallaties (RWZI's) in ons land. Het RIVM analyseerde die monsters om per RWZI-gebied het aantal virusdeeltjes per 100.000 inwoners te achterhalen. Steeg het coronagehalte in het rioolwater ineens, dan was dat een aanwijzing voor een virusuitbraak.

Twee jaar na het eind van de pandemie is het aantal rioolwatermetingen teruggeschroefd. Niet alleen is de hoge nood voorbij, ook de kosten van het meten en analyseren tellen mee. Veertig procent van de RWZI's analyseert twee keer per week op SARS-CoV-2, de rest nog maar wekelijks.

Betrouwbaarheid

Met nog maar 1 à 2 metingen per week groeit de invloed van meetonzekerheden. Hetebrij: "Je kunt altijd een meting krijgen die eruit springt, waarbij het aantal virusdeeltjes ineens piekt of duikt." Daarbij is het lastig dat er geen data van eergisteren meer is om mee te vergelijken. Om te zorgen dat er geen onrealistische gegevens in officiële rapportages komen, doet het RIVM aan trendanalyse. Daarbij worden nieuwe metingen vergeleken met voorgaande waarden van dezelfde RWZI; een flinke rekenklus die enkele uren kost.

Kan dat sneller en betrouwbaarder, door cijfers van nabijgelegen RWZI's met elkaar te vergelijken? Voor naburige grote steden zou je vergelijkbaar gedrag verwachten. Hetebrij, die tijdens zijn eigen



Nederlandse rioolwaterzuiveringsinstallaties in 2022. Bron: van Boven, Hetebrij, et al. Euro Surveill (2023)

promotie al eens deelnam aan de Studiegroep Wiskunde met de Industrie (SWI), legt het vraagstuk voor aan de organisatie.

Populaties

Voor Niek Mooij van de Universiteit Utrecht is de RIVM-case een natuurlijke keuze. Zijn promotie gaat over hoe de vorm van netwerken met meerdere locaties complexe processen zoals ecologie of economie beïnvloeden. Daar komen veel differentiaalvergelijkingen bij kijken. "Die techniek leek ons ook kansrijk om de coronadeeltjes in het rioolwater te beschrijven."

In differentiaalvergelijkingen hangt de snelheid waarmee een meetwaarde verandert af van zijn huidige waarde. Zo kun je bijvoorbeeld de populaties van roof- en prooidieren modelleren, maar ook de verspreiding van virussen. Tijdens de SWI-week bouwt een deel van de wiskundigen een model van het aantal coronainfecties in het gebied rond een RWZI.

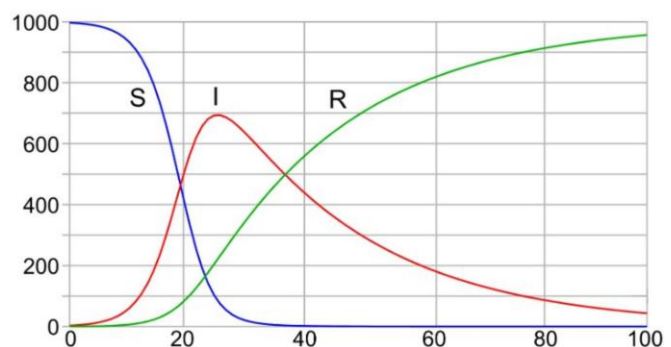


Het rekenmodel SIR. Bron: SWI

Het zogeheten SIR-rekenmodel (voor *Susceptible*, *Infectious* en *Recovered*) simuleert drie verschillende groepen mensen: mensen die nog *vatbaar* zijn voor infectie, mensen met de ziekte die *besmettelijk* zijn en dus virusdeeltjes verspreiden, en zij die *hersteld* zijn. Elke tijdstap van het rekenmodel kunnen mensen Corona oplopen of daarvan genezen.

Hoeveel mensen in het model besmet raken of genezen, hangt telkens af van de grootte van hun groep, vermenigvuldigd met een voorfactor: veel vatbaren betekent veel besmettingen. De uitdaging bij het SIR-model is om de juiste voorfactoren te vinden die de werkelijkheid goed beschrijven. Dat doet het team door hun formules te vergelijken met historische coronadata.

Met een betrouwbaarheidsinterval is te bepalen of een nieuw datapunt binnen de normale spreiding rond de functie ligt, of onwaarschijnlijk is. De SWI-week blijkt te kort om die analyse in te bouwen; ook is er geen tijd om de SIR-modellen van naburige steden met elkaar te vergelijken.



Een mogelijke oplossing van het SIR-model, met horizontaal de tijd en verticaal de grootte van de drie populaties. Bron: SWI

Voor Wouter Hetebrij is het SIR-model interessant, maar nog geen directe oplossing voor zijn vraagstuk. En ook Niek Mooij kiest uiteindelijk voor een andere aanpak: "Met differentiaalvergelijkingen moet je heel zeker zijn dat je het juiste gedrag modelleert. In het geval van corona heb je veel factoren die je hier niet mee kunt nemen." Herbesmettingen en nieuwe varianten bijvoorbeeld, maar ook vaccinaties. Mooij sluit zich daarom aan bij een team dat zich baseert op methodes uit de kansrekening.

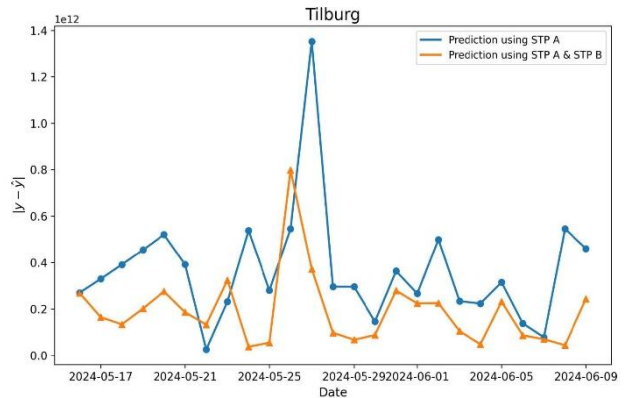
Autoregressie

In plaats van de werkelijkheid na te bootsen in een model, kiezen de statistici ervoor om wiskundige patronen af te leiden uit de werkelijkheid. Ze gebruiken de techniek van *autoregressie*, die populair is

om patronen te vinden in tijdreeksen van meetdata. Ook de relatie tussen verschillende rioolwaterzuiveringsinstallaties die het RIVM meet, is er goed in te vangen.

"Bij autoregressie bouw je een functie op basis van de historische data", legt Mooij uit. "Als je honderd datapunten, hebt, voer je een computer de eerste vijf en vraag je om een functie die het zesde datapunt voorspelt. Die verfijn je door met datapunt twee tot en met zes de zevende waarde te voorspellen. En zo door tot je bij het heden bent."

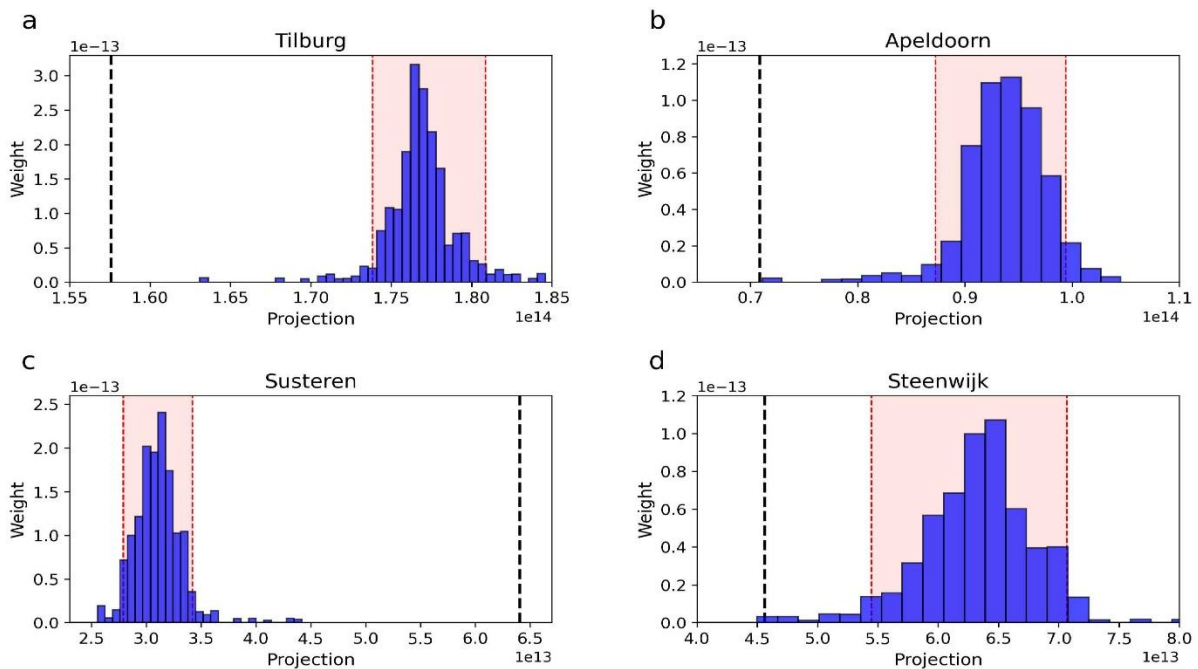
De uitkomst is serie coëfficiënten: per tijdverschil een coëfficiënt die het precieze gedrag bepaalt. En een bijbehorende onzekerheid in de voorspelling, want de data bevatten immers meetruis. De cruciale truc, vertelt Mooij: "vervolgens doe je de hele exercitie nog een keer met de data van een naburige RWZI erbij. Als dat een betere voorspelling levert, weten we hoe sterk die twee verbonden zijn."



Autoregressie van de Coronawaardes in het rioolwater van Tilburg. In de oranje uitkomst is ook een naburige RWZI meegenomen, in de blauwe alleen Tilburg. Bron: SWI

Door alle 300 ruim RWZI's met elkaar te vergelijken, krijgen de wiskundigen een netwerk van sterk en minder sterk op elkaar lijkende RWZI's. Na het doorploegen van de historische data gaat de berekening snel: elke nieuwe week hoeft je per RWZI maar één a twee nieuwe rekenstappen te doen. Rekening: anderhalve minuut.

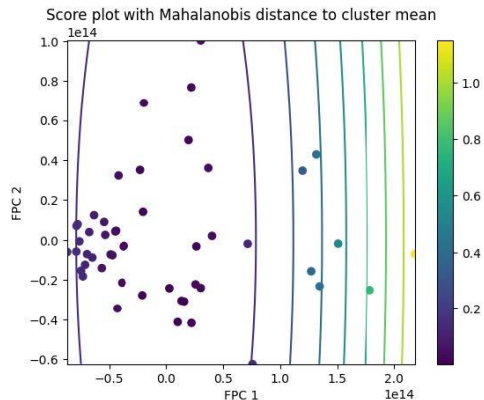
"Valt een nieuwe meetwaarde dan buiten de bulk van wat het netwerk zou verwachten, dan weet je dat die écht onverwacht is", schetst Mooij. Omgekeerd: als meerdere sterk verbonden gebieden tegelijkertijd pieken, is de uitkomst betrouwbaarder.



Betrouwbaarheidsintervallen in de autoregressieaanpak. Bron: SWI

Onderliggende patronen

Ook een derde groep onderzoekers zoekt naar onderliggende patronen in de coronadata. Maar waar autoregressie per RWZI één beschrijvende functie oplevert, maakt *Functional Principal Component Analysis* (FPCA) er zoveel je maar wil. Het achterliggende idee is om de gemeten data te ontbinden in een serie basisfuncties die samen het waargenomen gedrag voorspellen.



De meeste coronawaardes bij de RWZI Holten blijken te voorspellen met vergelijkbare combinaties van FPCA-functies. Bron: SWI

Alsof je losse tonen bij elkaar optelt tot je de juiste klank krijgt: de mix van de componenten is allesbepalend en de precieze verhoudingen werken zelfs als een vingerafdruk. Als een nieuw meetpunt een heel andere mix vraagt om goed te voorspellen, is dat weer een teken dat de meting buiten de verwachtingswaarde valt.

Ook voor de FPCA-team blijkt een week te kort om het RIVM-vraagstuk volledig op te lossen. Hun data-analyse werkt, maar naburige meetlocaties vergelijken is voor de toekomst. Dat kan bijvoorbeeld door grafieken van rioolwaterzuiveringspunten die op verschillende momenten beginnen te pieken te verschuiven tot ze zoveel mogelijk over elkaar vallen.

De kracht van wiskunde

Ook al ligt er geen kant-en-klare oplossing voor zijn vraagstuk, opdrachtgever Wouter Hetebrij is enthousiast over de opbrengst van SWI: "Het is mooi hoeveel verschillende methodes er zijn geprobeerd. Toch de kracht van wiskunde: je hebt uiteenlopende gereedschappen zoals modelleren of statistiek die allemaal toepasbaar zijn. Vooral de autoregressie lijkt goed bruikbaar. We gaan nu bespreken welke van de drie methodes van SWI we een tijd uittesten naast onze standaardaanpak, zodat we ze over een tijdje kunnen vergelijken."

Ook promovendus Niek Mooij is tevreden over de SWI-week: "In een team werken aan een probleem uit de praktijk was heel interessant, erg anders dan ieder aan zijn eigen project. Het was extra leuk om samen te werken met mensen uit andere gebieden van de wiskunde, die je normaal gesproken niet 1-2-3 tegenkomt. Samen konden we heel wat meters maken!"