



GASTCOLUMN ZORGTECHNOLOGIE

Gegevenswetenschap benutten in crisistijd

Het eerste Covid-19 sterftegeval in Nederland werd gerapporteerd op vrijdag 6 maart 2020. De crisis was toen al een paar maanden in China in volle gang en Italië had ook al serieuze problemen. In Nederland werden de gevolgen van het virus op dat moment nog steeds onderschat.

Als statisticus vond ik het nodig een bijdrage te leveren aan de werkzaamheden rond de crisis, omdat ik van mening ben dat we niet alleen moeten onderwijzen hoe we data kunnen modelleren of analyseren, maar dat we dit ook moeten laten zien, vooral in tijden wanneer het er echt toe doet. Mijn doelstelling was het dagelijks voorspellen van het aantal infecties en doden in verschillende landen en het publiek hierover te informeren.

Nu is er een enorme literatuur over modellen die beschrijven hoe de verspreiding van het virus zou kunnen verlopen – het is een wetenschap op zich. Een van de meest simpele van deze modellen leidt tot een Verhulst-groei-curve, een model dat halverwege de negentiende eeuw werd ontwikkeld voor het voorspellen van de totale populatieomvang in landen. Het model bevat drie parameters met een directe betekenis voor het simpele verspreidingsmodel. Een van de parameters geeft de gemiddelde groeisnelheid aan, waarmee in dit geval de populatie wordt besmet met het virus. De tweede parameter is het maximumaantal besmettingen dat zal plaatsvinden. En de derde parameter is het aantal dagen (na de start) waarop vijftig procent van de besmettingen heeft plaatsgevonden. Deze laatste parameter wordt ook wel het kantelpunt genoemd, omdat op het kantelpunt het maximale aantal nieuwe besmettingen bereikt is.

Met een klein team van onderzoekers startten we onze analyses met de data van China. Het Verhulst-model past heel goed op de Chinese data. Dit feit bewijst niet dat het simpele verspreidingsmodel ook goed past bij de onderliggende data van Covid-19. Dit lijkt misschien een tegenspraak, maar dat is het niet. Het simpele verspreidingsmodel legt voorwaarden op de relatie tussen de drie parameters die wij niet bij het *fit*ten van het Verhulst-model implementeren. Het Verhulst-model

kan dus prima functioneren zonder de onderliggende interpretatie. Ik heb Verhulst-groei-curve in het verleden ook toegepast op gegevens die helemaal niets met de verspreiding van virussen te maken hebben. Het gaat dus uiteindelijk om het adequaat analyseren van data, waarbij kan worden aangetoond dat het gekozen model de taak kan volbrengen. Het model staat niet voorop, maar de data, het doel, en de (onderbouwning van de) aanpak.

Onze voorspellingen voor China op basis van het eerste deel van de groei-curve lieten zien dat nieuwe gevallen en het maximum in de toekomst gemiddeld genomen worden onderschat. Simulaties die we hebben uitgevoerd bevestigen deze tekortkomingen. We passen daarom een kalibratie toe om dit effect te corrigeren. Vanaf 11 maart rapporteren we zo dagelijks onze resultaten om het publiek te informeren (zie link onderaan). Tevens hebben we onze aanpak ingezet om te helpen bij het voorspellen van ziekenhuis- en intensive care -opnamen.

Dit lijkt misschien een mooi succesverhaal voor de gegevenswetenschap, maar eigenlijk hebben we maar heel beperkt kunnen helpen en is er nog steeds veel onduidelijk. Gegevens worden helaas niet op wetenschappelijke principes verzameld. We weten bijvoorbeeld niet welk deel van de populatie nu werkelijk besmet is en of onze groei-curve daarmee een sterke relatie vertonen. Een veranderend testbeleid helpt ook niet bij het monitoren van de crisis. Verder geldt dat sommige gegevens niet precies worden verzameld. We weten niet goed hoeveel mensen er dagelijks beter worden. Niet alle gegevens worden gedeeld, ondanks de grote wens dat wetenschap publiek beschikbaar zou moeten zijn. Mijn grote wens is dat we in de (nabije) toekomst hiervan leren, zodat de gegevenswetenschap beter kan worden ingezet in een crisis(tijd).

Edwin van den Heuvel
hoogleraar Statistiek Technische Universiteit Eindhoven

<https://www.tue.nl/nieuws/nieuwsoverzicht/11-03-2020-datawetenschappers-tue-voorspellen-groei-coronabesmettingen-per-land/>