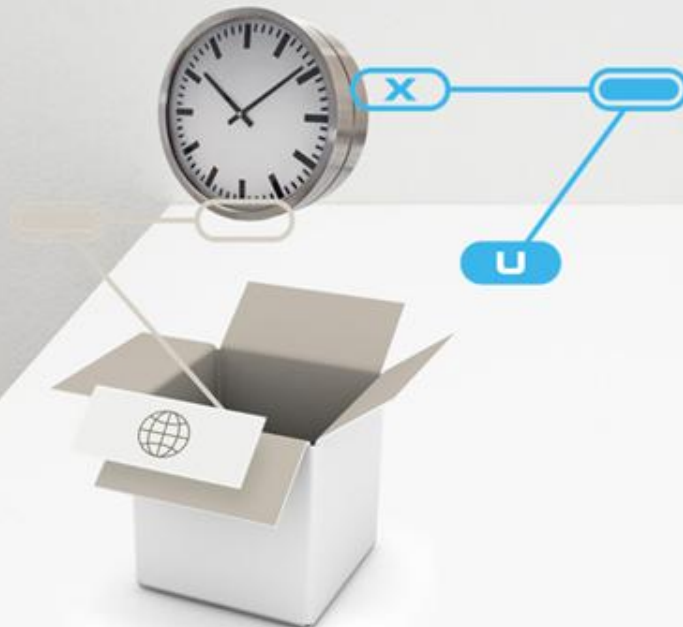




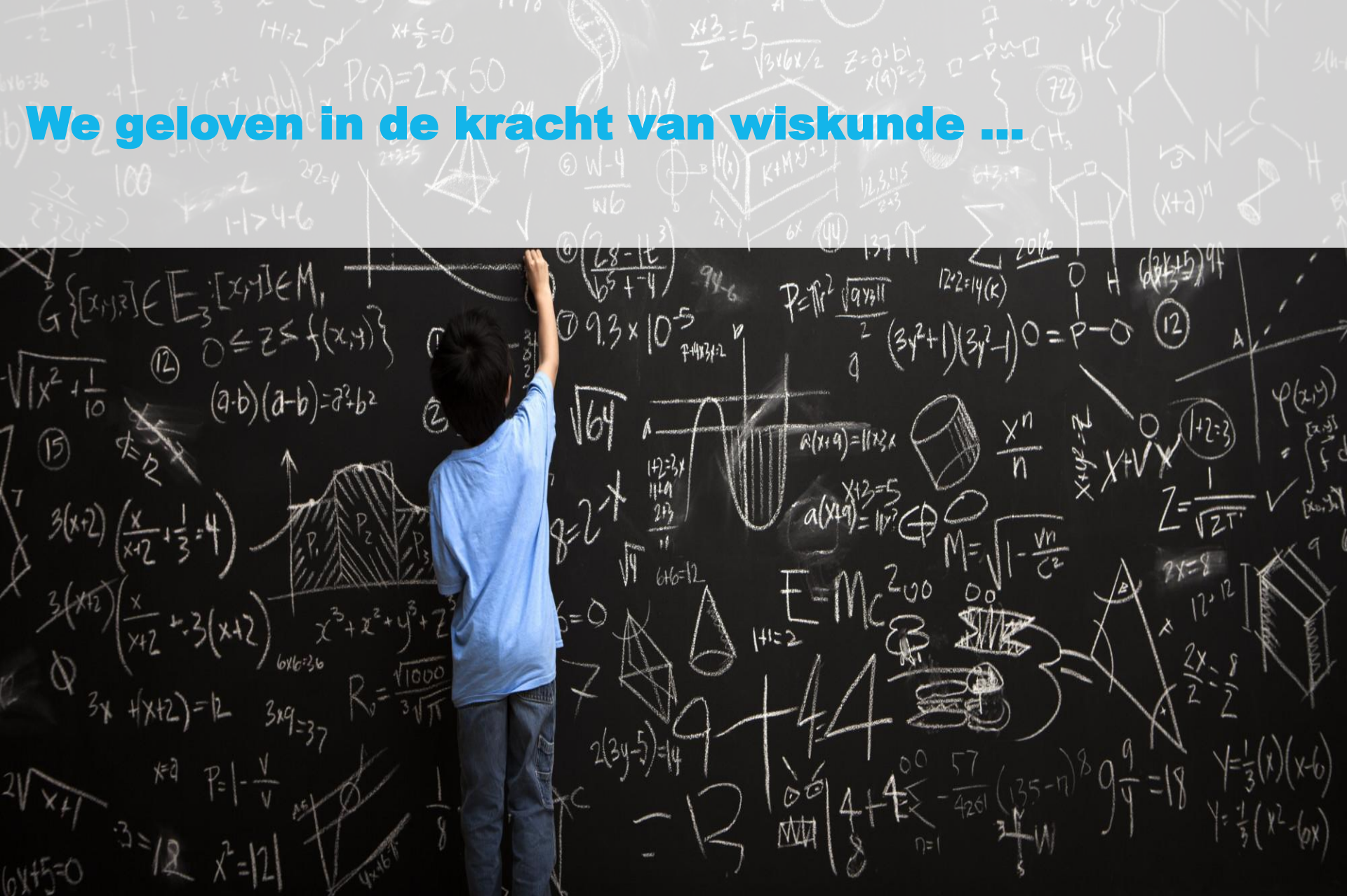
Consultants in Quantitative Methods

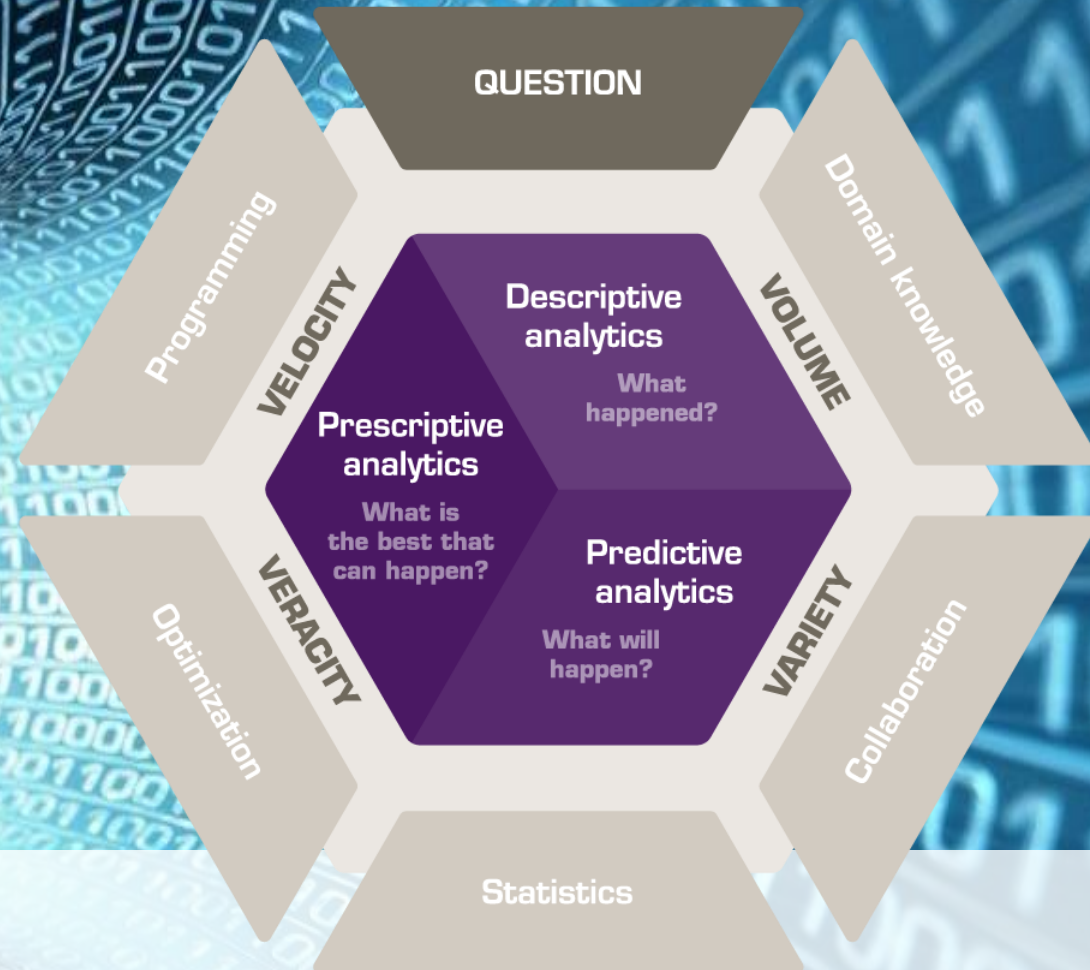
Data analytics in de praktijk

Arjen Vestjens – Managing Partner | 9 juni 2016 | vestjens@cqm.nl | 040 750 2318



We geloven in de kracht van wiskunde ...





... en innovatie gebaseerd op data



R&D



SUPPLY CHAIN
ANALYTICS



TRANSPORT PLANNING



RAIL



ENERGY



DATA SCIENCE





**OPPORTUNITY
AHEAD**

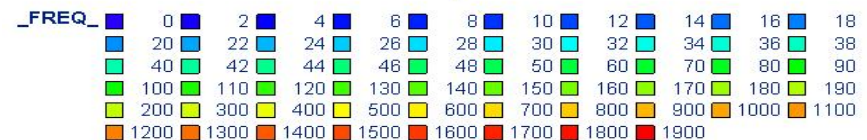
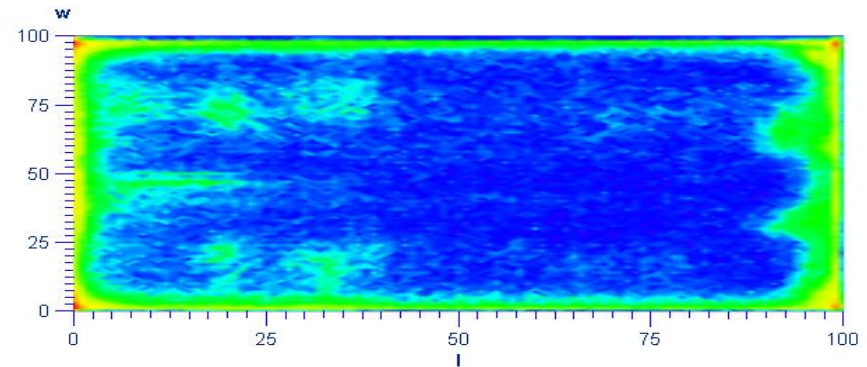
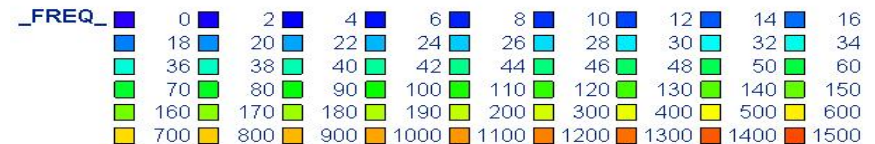
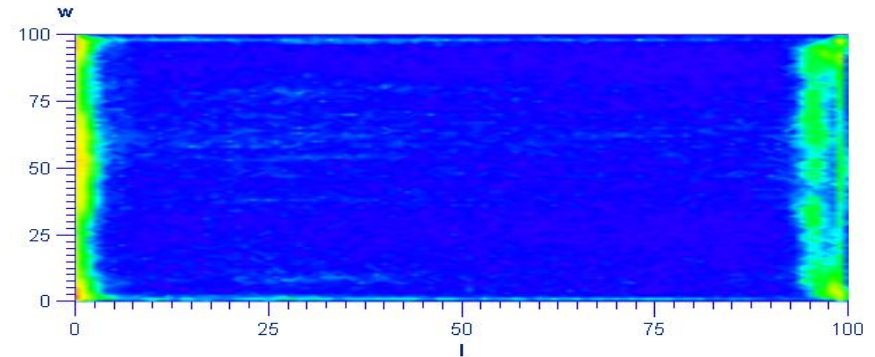


Scratch free surface. A big data quest.
Herleiden klantfeedback naar root causes in
productieprocessen

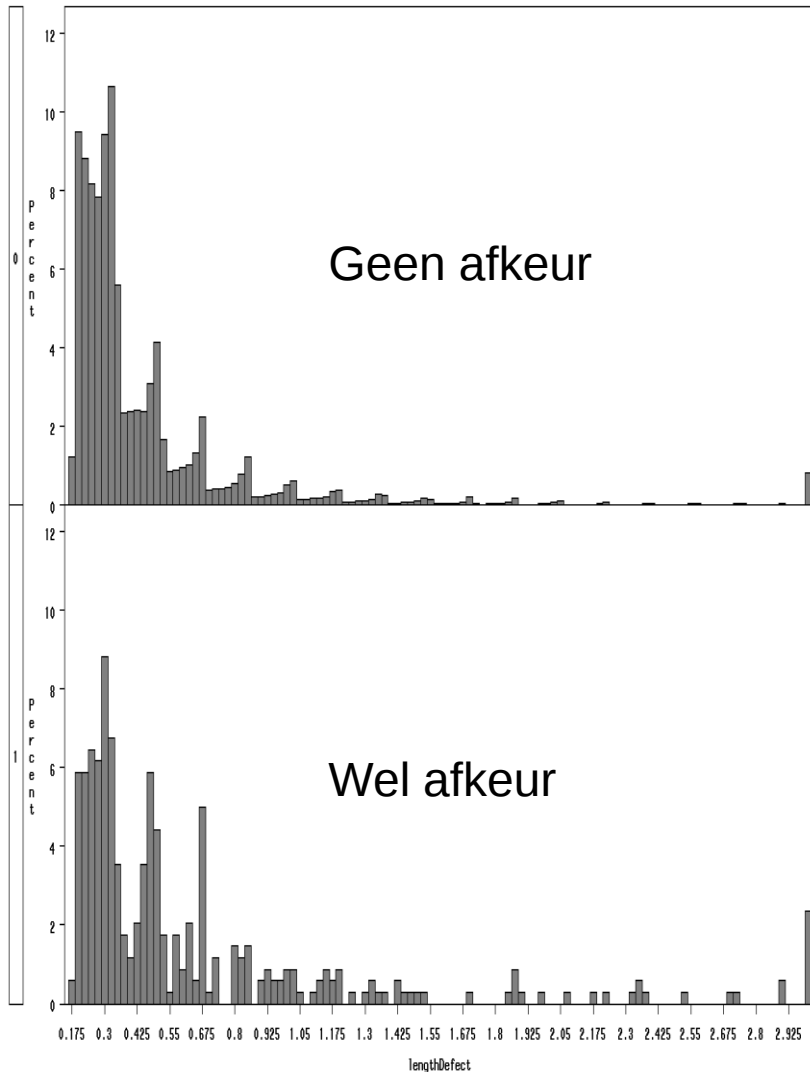
Reductie klantenklachten

- Het doel is om het aantal klantenklachten te reduceren
- Hiervoor is achterhaald welke klantenklachten het meest voorkomen en welke aangepakt gaan worden (top 10 van klantenklachten)
- De aanpak van de klantenklachten verloopt indirect met behulp van cameragegevens
- Hiervoor dient er een verband gelegd te worden tussen cameragegevens enerzijds en klantenklachten (of afkeur) en productiestappen anderzijds

Getoond:
locatie (grotere) krassen op rol,
onder en bovenkant

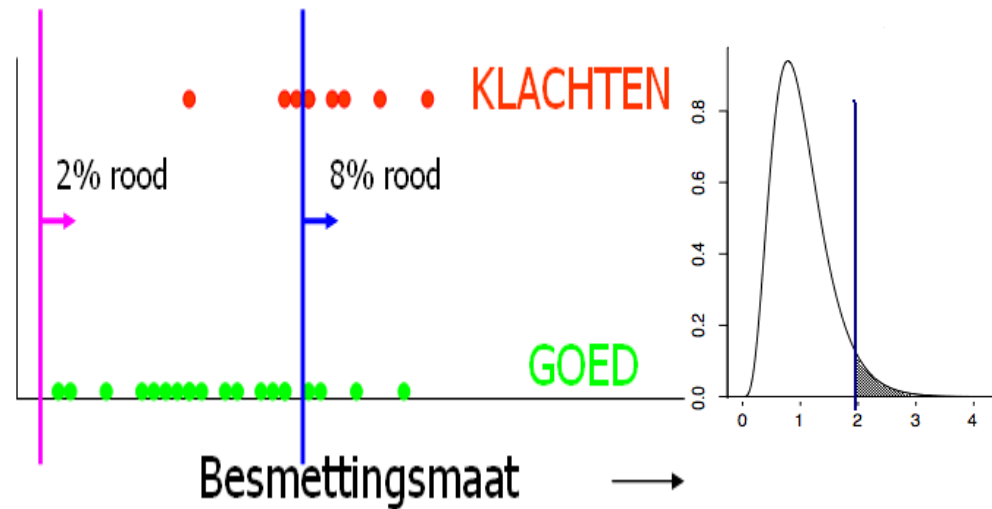


Meest discriminerende besmettingsmaat?



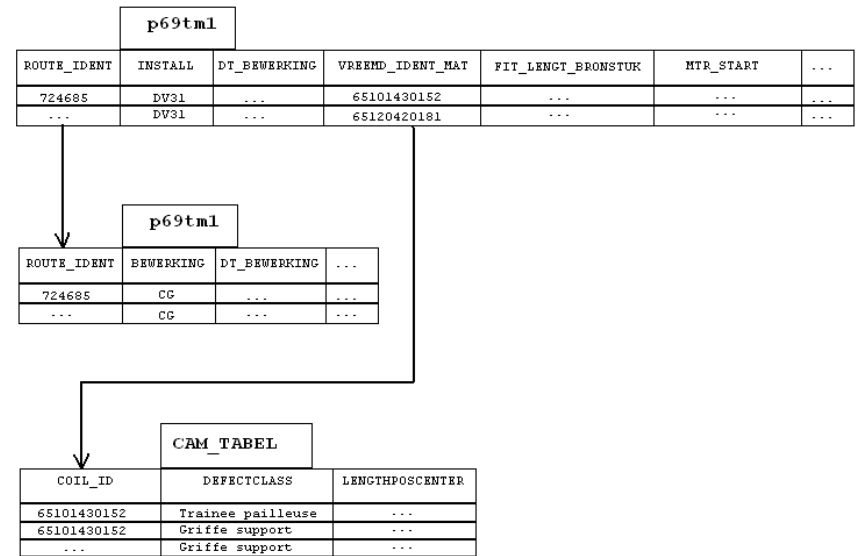
Mogelijke besmettingsmaten:

- Gemiddeld aantal krassen per m²;
- Gemiddelde foutlengte krassen per m²;
- Gemiddelde oppervlakte van krassen per m²;
- Grootste kras per rol;
- Som van foutlengtes groter dan x per rol, met verschillende filters;
- Maten gemiddeld per m² en per m.



Issues bij up-to-date houden overzicht en terugtransformeren

- Vervuilde data – misclassificaties
- Niet alle detectiesystemen worden op dezelfde manier opgeslagen
- Om een aggregatietabel consistent te houden moet rekening gehouden worden met het doorknippen van bronstukken
- Een rol kan uit een paar plakken van de staalfabriek bestaan
- Rekentijd updaten aggregatietabel
- Veel verschillende camerarecepten
- Verschillende oppervlakte\staal-kwaliteiten



ROUTE_IDENT	C_FOUT_ZIJDE	FOUT_ERNST	FOUT_CODE_CAMERA	AANTAL	SUM_OPP_DEFECT_M2	SUM LENGTE M
724685	T	Z	106	2	0,186218	0,699
724685	B	L	34	1	0,000042	0,007
724685	B	L	40	5	0,003033	0,285
724685	B	M	73	3	0,000112	0,028
724685	B	Z	88	2	0,043799	2,21
724685	B	Z	106	16	0,492564	0,765
...	...	...	...	...	...	...

Resultaat

- ⇒ Klachten gereduceerd door enkele structurele oplossingen in het productieproces
- ⇒ Continue monitoring van kwaliteit en inzicht plek in productieproces waar kwaliteit degradeert
- ⇒ Geeft mogelijkheid om vroegtijdig in te grijpen en problemen bij volgende rollen te voorkomen
- ⇒ Geeft mogelijkheid om klant direct de juiste kwaliteit te leveren – geen (dure) verrassingen achteraf





Voorspellen van aantal passagiers in de trein om beschikbare capaciteit te balanceren met verwachte reizigersstromen.

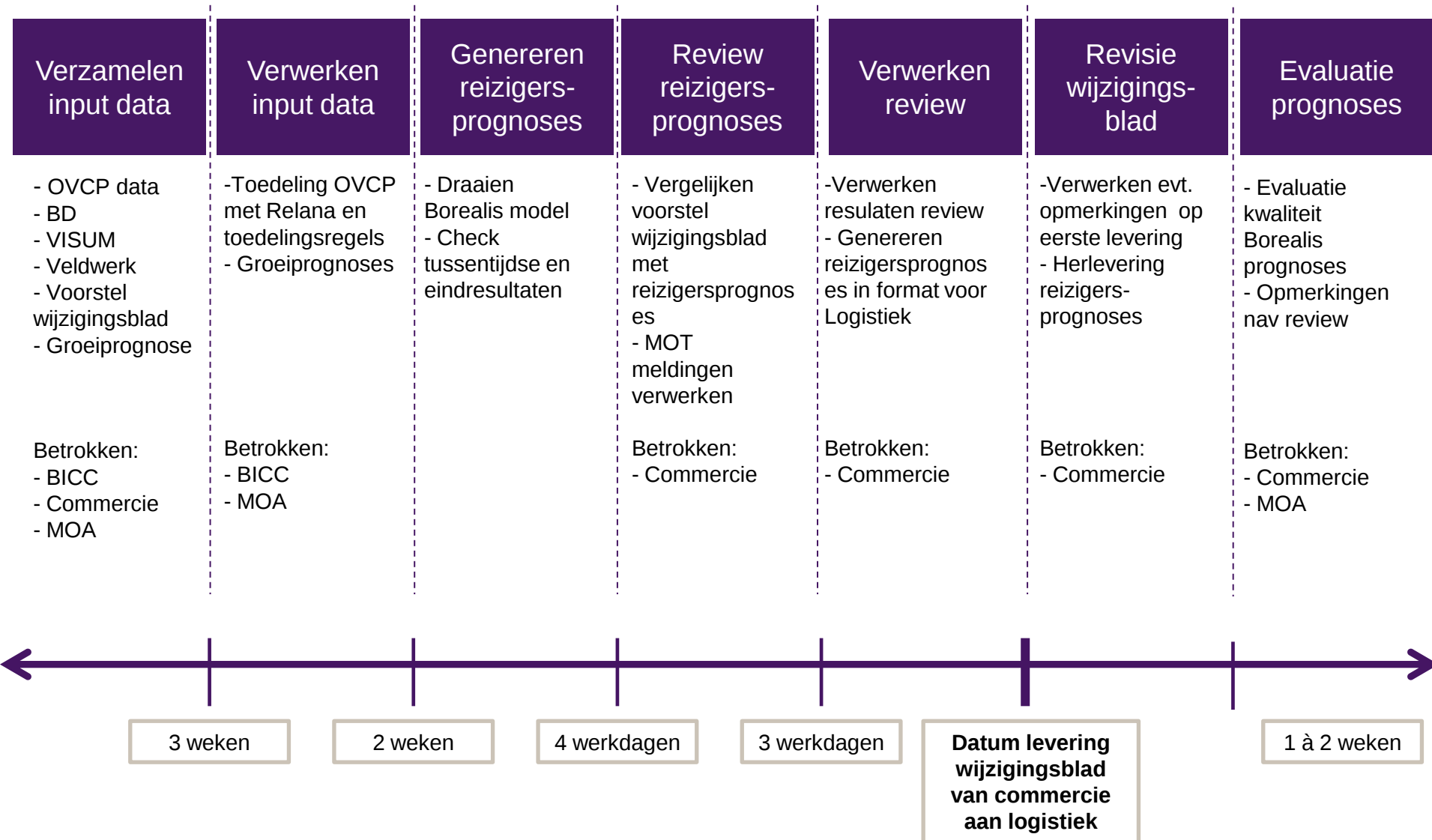
>280M records invoer en 150K records uitvoer.

Uitvoer

Trnr	Dvdw	Van	Naar	Vertrektijd	Prognose
527	1	Rtd	Gd	805	579
527	1	Gd	Ut	824	559
527	1	Ut	Amf	850	569
527	1	Amf	Zl	907	355
527	1	Zl	Asn	947	296
527	1	Asn	Gn	1027	296
527	2	Rtd	Gd	805	571
527	2	Gd	Ut	824	641
527	2	Ut	Amf	850	569
527	2	Amf	Zl	907	345
527	2	Zl	Asn	947	257
527	2	Asn	Gn	1027	257
527	3	Rtd	Gd	805	571
527	3	Gd	Ut	824	641
527	3	Ut	Amf	850	569
527	3	Amf	Zl	907	345
527	3	Zl	Asn	947	257
527	3	Asn	Gn	1027	257
527	4	Rtd	Gd	805	571
527	4	Gd	Ut	824	641

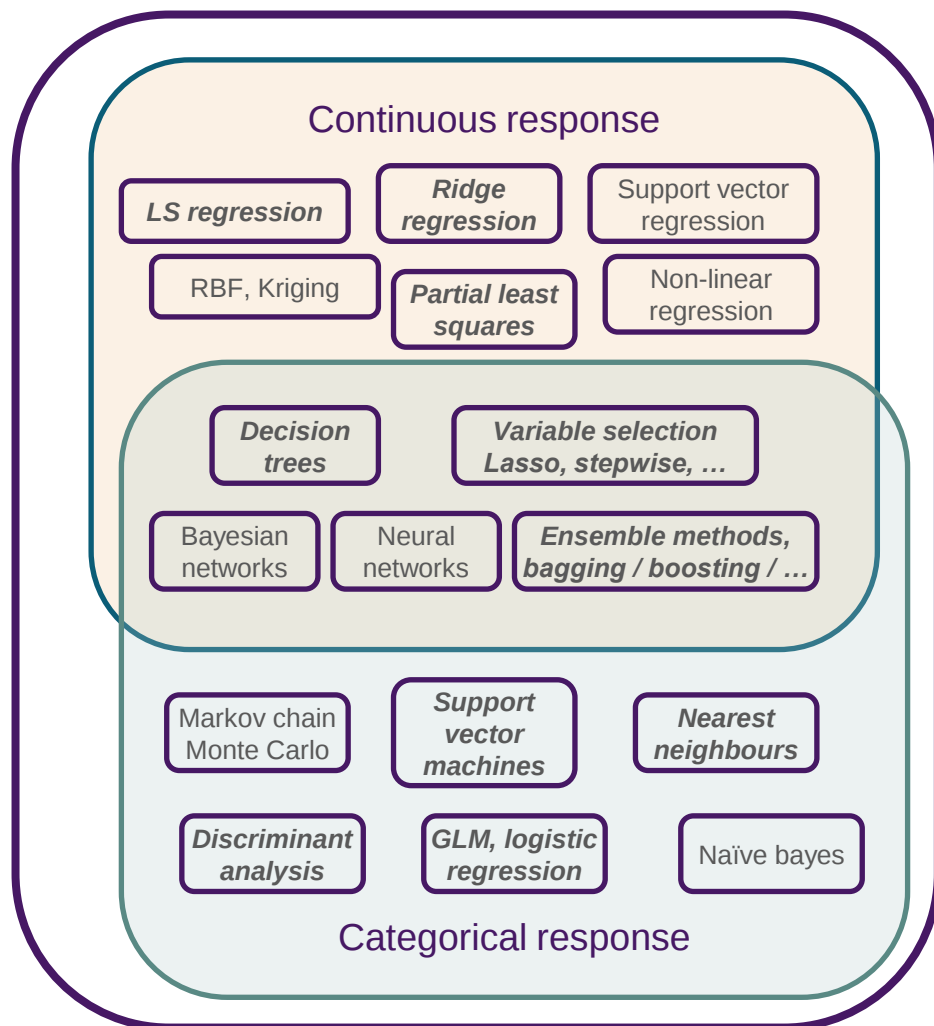


Proces van voorspellen

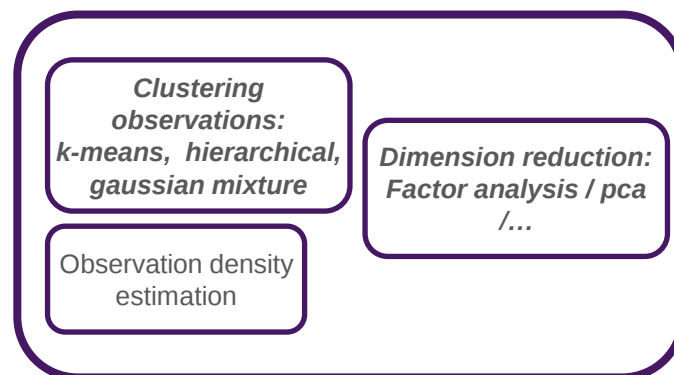


Overzicht statistiek & machine learning technieken

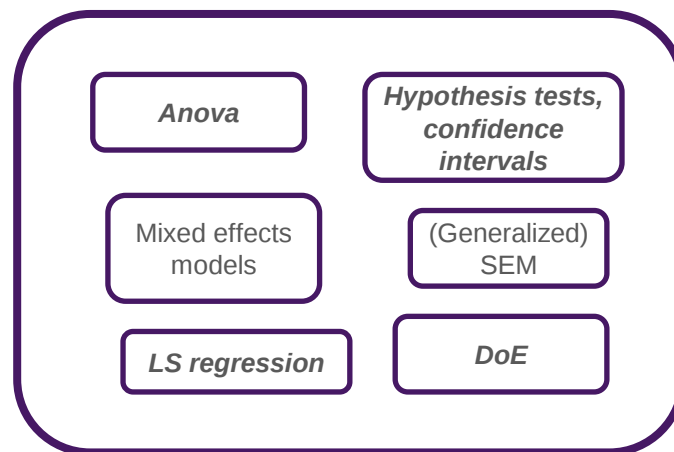
Predictive model building

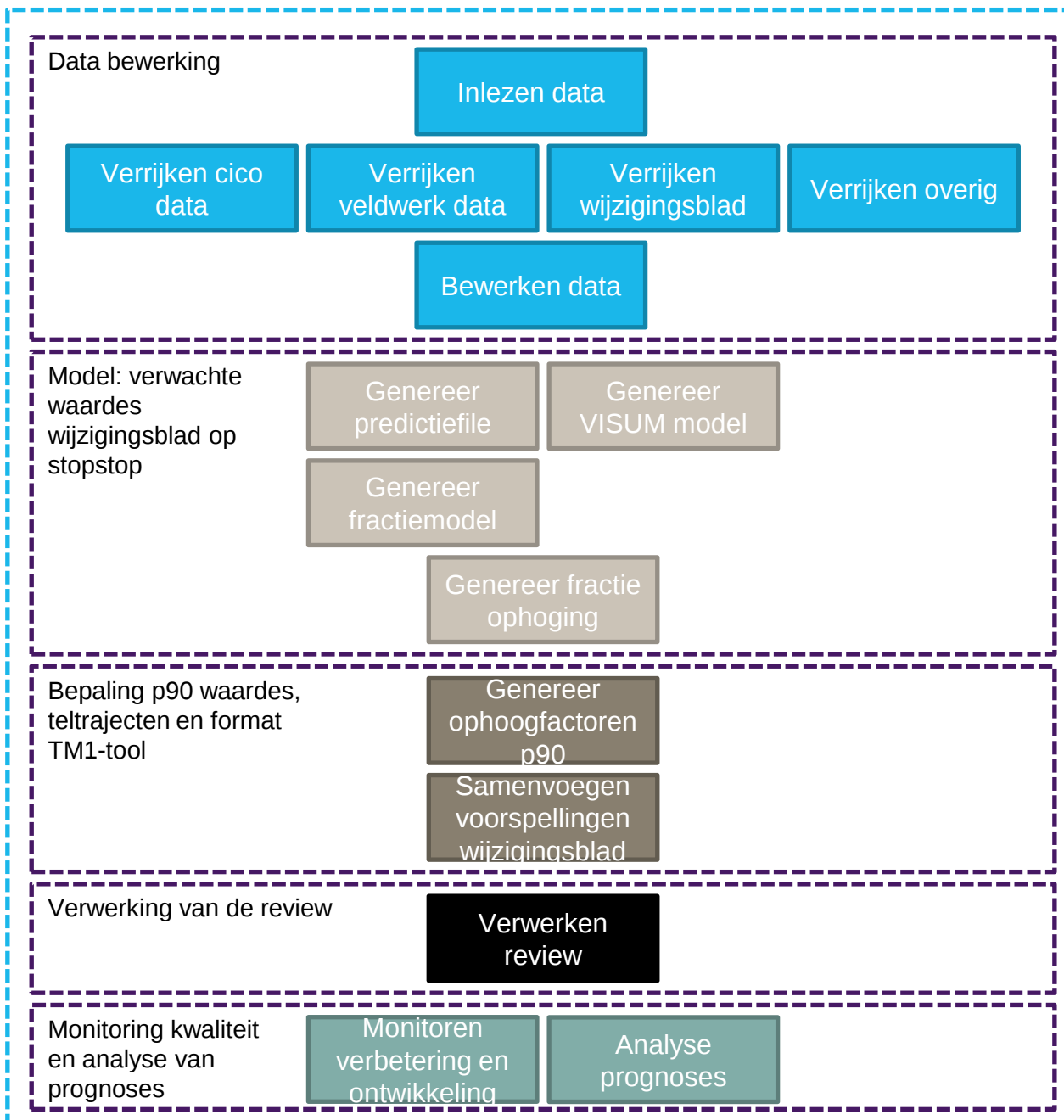


Discovering structure in observed data



Proving and quantifying effects

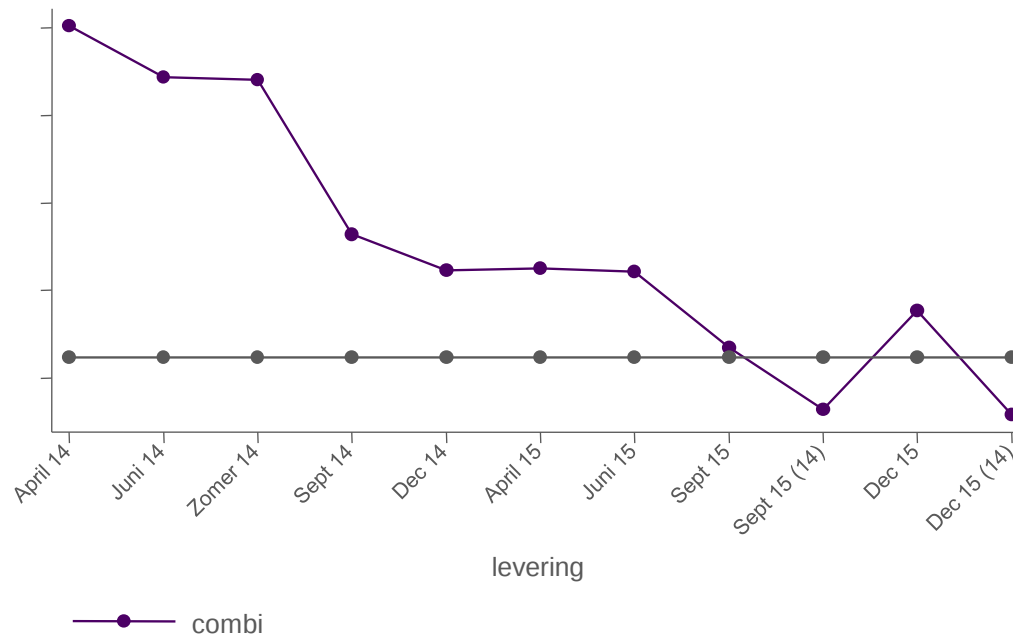




Resultaat

- Transparante voorspelmethode, volledig in beheer van NS, die voor ~97% van de trajecten, zonder tussenkomst van een expert, voldoende goede voorspelling geeft voor het capaciteitsplanningsproces.

RMSE ontwikkeling t.o.v. tellingen

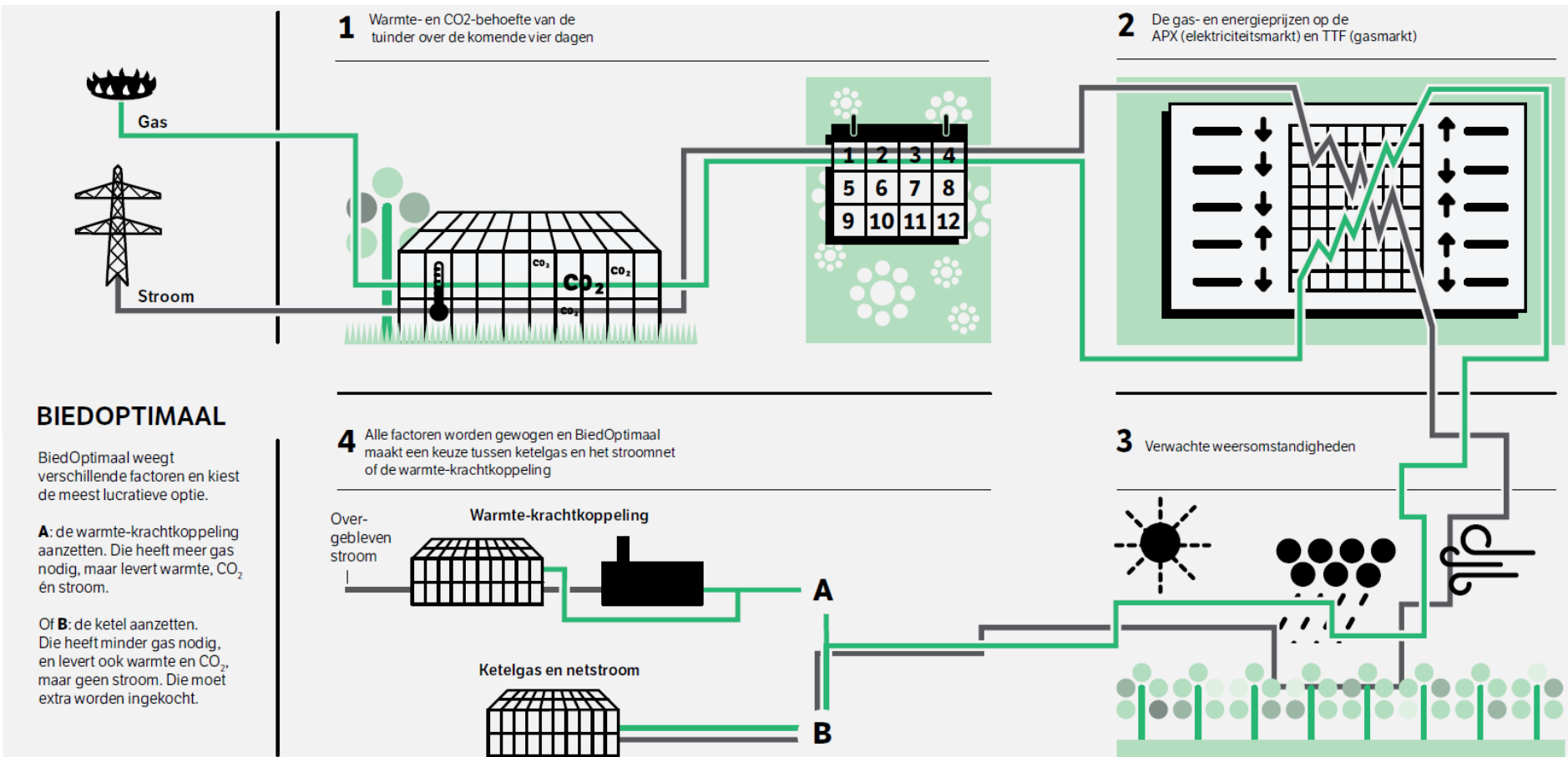


Tuinders worden energie leveranciers

Voorspel warmte- en energieconsumptie en gas- en elektriciteitsprijzen, en optimaliseer om de beste bieding op de energiemarkt te doen.



Hoe werkt het?



Databronnen

‘Intern’

- ⇒ Installatie(groep) kenmerken (e.g., min/max capaciteit en warmte-efficiëntie)
- ⇒ Onderhoudsperiodes per installatie per tijdseenheid
- ⇒ Historische waarden status (aan/uit op) per tijdstip
- ⇒ Historische waarden bufferstand per tijdstip
- ⇒ Historische waarden per installatiegroep per tijdseenheid (warmte en CO2)

‘Extern’

- ⇒ Historische prijsinformatie per tijdseenheid (APX en TTF)
- ⇒ Weersvoorspellingen per regio per tijdseenheid (temperatuur, zonnestraling, windrichting, windkracht, luchtvochtigheid)

Voorspellen

➤ Voorspellen warmte en CO2 behoefte

- Fysieke model van de kas bleek te moeilijk om te gebruiken in de praktijk
- Lineair regressiemodel gebaseerd op historische data werkt goed:
 $a * \text{tijd op de dag} + b * \text{verwachte temp} - c * \text{luchtvochtigheid} + d * \text{windrichting} * \text{windkracht} + \dots$



➤ Voorspellen gas en elektriciteitsprijzen

- Gas – niet voorspelbaar: prijs van morgen = prijs van vandaag
- Elektriciteit – lineair regressiemodel gebaseerd op historische data:
tijdstip op dag, dag in week, ...
Lastig: feestdagen, vakanties, zomer/wintertijd



Optimaliseren

☞ Te maken keuzes door tuinder:

- Warmte van te voren produceren en opslaan in buffer
- Warmte just-in-time produceren
- Gebruik ketel voor productie warmte (en inkoop elektriciteit op het net)
- Gebruik WKK voor productie warmte en electriciteit (en lever rest-electriciteit aan het net)

☞ Drie verschillende optimalisaties geïmplementeerd

- BufferRun Bepalen bufferstand-verwachting (input BiedRun)
- BiedRun Bepalen van de optimale bieding voor de tuinder
- KetelRun Bepalen optimale ketelinzet (na APX settlement)

☞ Bovengenoemde optimalisaties werken alle drie op hetzelfde wiskundig optimalisatiemodel (MILP)

☞ Solvers aangeroepen vanuit VAA software, die ook alle data klaar zet en de koppelingen met de andere systemen (APX, WKK) afhandelt



Resultaat

➤ Resultaat

- Online portal voor tuinders om biedingen mee te doen op energiemarkt
- Inclusief een tailor-made model voor optimale biedingen
- Inclusief optimale aansturing van WKK en ketel

➤ Samenwerking

- AgroEnergy: de visie, het budget en expertkennis van de markt
- VAA ICT Consultancy: de web-based portal, data opslag en koppeling met andere systemen
- CQM: voorspel- en optimalisatiemodules



Data analytics in de praktijk

Leerpunten en randvoorwaarden voor succes

- ⇒ Modelontwikkeling (voorspellen en optimaliseren) is een iteratiefproces
 - Een prototype (door)ontwikkelversie is nodig (naast een productieversie)
 - Modellen dienen altijd (eerst) gevalideerd te worden in de praktijk
 - Bij volautomaat moet er altijd een back-up scenario zijn voor geval er geen (realistische) uitkomst is
- ⇒ Combinatie van modellen en expert kennis geeft het snelste direct bruikbare resultaten
- ⇒ Modellen en methoden moeten robuust zijn tegen vervuilde of missende data
- ⇒ Voorspelmodellen moeten transparant zijn om er vertrouwen in te kunnen krijgen
- ⇒ Na anonimiseren is gebruik van gevoelige klantdata / persoonsgegevens mogelijk

➔ Data analytics werkt in de praktijk!



Bedankt voor uw aandacht!

