

Leeghwaterstraat 44
2628 CA Delft
Postbus 6012
2600 JA Delft

www.tno.nl

T +31 88 866 22 00

TNO-rapport**TNO 2022 R10152****Onderzoek naar toepassing van AI voor
voorspelling van de opbrengst op basis van
het lokale kasklimaat in een Gerbera-kas**

Datum	1 februari 2022
Auteur(s)	A. Sapounas M.S. Molhoek R.A. Bezemer
Aantal pagina's	36
Aantal bijlagen	--
Opdrachtgever	Ministerie LNV Programma Kas als Energiebron
Projectnaam	Kansen AI voor precisie Klimaat
Projectnummer	060.41076

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, zonder voorafgaande toestemming van TNO.

Indien dit rapport in opdracht werd uitgebracht, wordt voor de rechten en verplichtingen van opdrachtgever en opdrachtnemer verwezen naar de Algemene Voorwaarden voor opdrachten aan TNO, dan wel de betreffende terzake tussen de partijen gesloten overeenkomst.

Het ter inzage geven van het TNO-rapport aan direct belanghebbenden is toegestaan.

© 2022 TNO

Samenvatting

De gewasontwikkeling en de oogstplanning in een kas zijn gebaat bij een homogeen kasklimaat. Echter, zelfs bij gebruik van geavanceerde systemen kan het klimaat nog behoorlijk inhomogeen zijn. Een inhomogeen kasklimaat en een daaraan gerelateerd niet uniform groeiend en producerend gewas zorgt voor een hoger energiegebruik. De teler zal te allen tijde doen wat mogelijk is om een gelijkmatiger klimaat te verkrijgen.

Dit project voorziet in een eerste stap om gedetailleerd inzicht te verkrijgen in de variatie van klimaatvariabelen, zowel ruimtelijk als over de tijd en wat deze variatie betekent voor de productie. Verschillende klimaatvariabelen zijn gedurende enkele maanden op goed bekende posities in de kas gemeten. Statistiek en kunstmatige intelligentie (AI) zijn ingezet om het aantal geoogste bloemen (gerbera's) te relateren aan tijd en plaats, en om handvatten te verkrijgen om deze opbrengst te verhogen.



Dit project is uitgevoerd in het programma Kas al Energiebron in de periode november 2019 t/m maart 2021. In dit project heeft TNO samengewerkt met Kwekerij Van Veen Gerbera uit De Lier (zie bovenstaande foto's) en Proeftuin Zwaagdijk (inmiddels Vertify) uit Naaldwijk.

Doelstellingen en conclusies van het onderzoek

Het onderzoek kende drie doelstellingen:

- A. In kaart brengen van verschillen in de temperatuur en de relatieve luchtvochtigheid in een commerciële kas en deze gegevens aanvullen met gegevens van de klimaatcomputer (zowel de meetwaarden als de ingezette control acties) en productiegegevens van het gewas.
- B. Relateren van de klimaatgegevens van de klimaatcomputer, weergegevens en het sensornetwerk aan productiegegevens met behulp van kunstmatige intelligentietechnieken.
- C. Op basis van B) in kaart brengen van het effect van klimaatverschillen op de gewasontwikkeling en aantonen dat kunstmatige intelligentie op basis van het klimaat de opbrengst (aantal geoogste bloemen) kan voorspellen.

De conclusies uit het onderzoek bij deze doelstellingen zijn de volgende:

- A. Er is een aanzienlijke ruimtelijke variatie van klimaatparameters binnen de kas qua temperatuur, relatieve luchtvochtigheid en PAR (Photosynthetic Active Radiation). Deze variatie van het klimaat, zoals blijkt uit de metingen van het draadloze sensornetwerk, is groter dan die gemeten door de geventileerde meetboxen van de klimaatcomputer. De statistische analyse van de gegevens geeft globaal inzicht in de relatie met de opbrengst (aantal geoogste bloemen).
- B. Op basis van de verschillende analyses kunnen in de donkere periode (oktober – maart) de volgende acties een positief effect hebben op de opbrengst:
 - Vermijden van lage verwarmingstemperaturen, zowel overdag als 's nachts, waarbij de normale dag-nacht schommelingen van de kastemperatuur blijven bestaan;
 - Vermijden van lage PAR in de tweede helft van de middag en de vroege avond, dus meer assimilatiebelichting, zodat meer fotosynthese kan plaatsvinden.
- C. Het is mogelijk gebleken om deze gegevens in te zetten in een AI-algoritme, waarmee een voorspelling gedaan kan worden van de toekomstige opbrengst. Hiermee kunnen complexe relaties tussen variabelen gerelateerd worden aan de opbrengst. Door het algoritme te laten leren (o.a. door het te voeden met de werkelijke opbrengst), kan het een redelijke opbrengstvoorspelling doen.

Opzet van het onderzoek: experimenteel en analyse van gegevens

In een proefperiode van ca. 10 weken (fase 1) en een praktijkexperiment van 5 maanden (fase 2) zijn resp. 10 en 12 draadloze sensoren geïnstalleerd die ieder temperatuur, relatieve luchtvochtigheid en PAR meten op bloemhoogte. Fase 1 diende om een goed beeld te krijgen van inhomogeniteit in de gehele kas, zodat in fase 2 een kleiner gebied nauwkeuriger en met wat meer sensoren onderzocht kon worden. In fase 2 is rond ieder van de 12 sensoren een veld afgezet, waarin het aantal geoogste bloemen werd geteld. Met de projectpartners is afgesproken dat de volgende gegevens worden verzameld:

- Oogstgegevens: aantal geoogste en aantal afgekeurde bloemen.
- Sensor-metgegevens: temperatuur, relatieve luchtvochtigheid, PAR.
- Gegevens klimaatcomputer buiten: temperatuur, windsnelheid en -richting.
- Gegevens klimaatcomputer binnen: kastemperatuur, relatieve luchtvochtigheid, raamopening aan de wind- en aan de luwe zijde, PAR instraling, verwarmingstemperatuur, %verduistering door scherm.

In fase 2 zijn de 12 velden totaal 36 maal geoogst, leidend tot 432 oogstwaarden. De waarden van de klimaatvariabelen zijn over een uur gemiddeld, zodat er per uur één waarde is. Om een relatie te kunnen leggen met de productie (de oogstwaarden), zijn de klimaatgegevens van een periode vóór de oogst van belang. Op aangeven van de kweker heeft het projectteam gekozen voor een periode van vier dagen voor de oogst. Daarom is van alle uurlijkse waarden van de klimaatgegevens het gemiddelde genomen over de laatste vier dagen, wat leidt tot 24 waarden per variabele, één voor ieder uur van de dag.

De verkregen gegevens zijn geanalyseerd op drie manieren:

- Statistische analyses, die inzicht geven in de variatie over de kas, het verloop van variabelen over de tijd, en de samenhang van temperatuur, relatieve luchtvochtigheid en PAR met het aantal geoogste bloemen (de opbrengst).
- Correlatieanalyses, die inzicht geven hoe sterk iedere variabele samenhangt met de opbrengst, uitgaande van lineaire afhankelijkheden; let op: dit impliceert nog geen *causale verbanden* van deze parameters met de opbrengst.

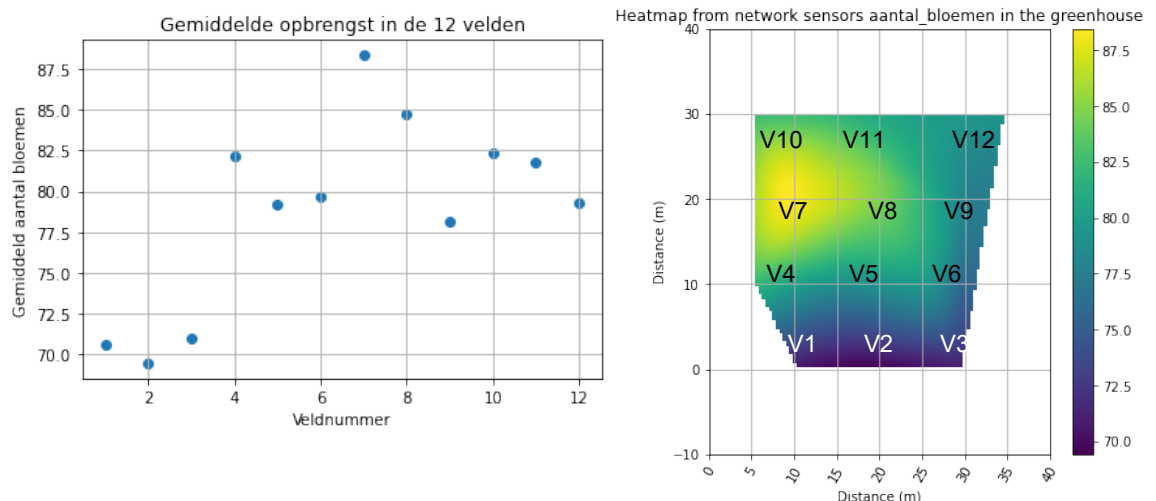
- Machine learning / kunstmatige intelligentie (AI), om inzicht te krijgen in complexe relaties tussen gecombineerde variabelen en de opbrengst. Daarnaast geeft AI informatie hoe de opbrengst beïnvloed kan worden (voorspellende functie).

Resultaten van de analyses

Uit de statistische analyse blijkt hoe de temperatuur, PAR en de relatieve luchtvochtigheid varieert over de onderzochte velden in de kas (mate van inhomogeniteit):

- Het verschil tussen de hoogste en laagste temperatuur loopt op tot 3 °C.
 - Het verschil in PAR varieert midden op de dag van ca. 300 tot 450 $\mu\text{mol}/\text{m}^2/\text{s}$.
 - Het verschil tussen de hoogste en laagste relatieve vochtigheid loopt op tot 7%.
- Dit zijn de maximale verschillen; deze treden midden op de dag op.

De analyse wijst verder uit dat de opbrengst over de 12 velden afhangt van de plaats in de kas. Onderstaande figuur toont de gemiddelde opbrengst per veld (V1 t/m V12) per oogst in een grafiek (links) en een 'heatmap' (rechts). De opbrengst (incl. afgekeurde bloemen) is gemiddeld 20% hoger aan de kant van het pad (velden 4, 7, 10, links in de heatmap) dan aan de zijkant van de kas (velden 1, 2 en 3, onderaan in de heatmap). Na correctie voor afkeur is het verschil nog steeds 17%.



Verder zijn de volgende afhankelijkheden gevonden:

- Een hogere temperatuur is gerelateerd aan een hogere opbrengst.
- Een hogere PAR gedurende de dag-uren is gerelateerd aan een hogere opbrengst. Dit is het duidelijkst tussen 11:00 en 16:00 uur.
- De relatieve luchtvochtigheid is niet duidelijk gecorreleerd aan de opbrengst.

Uit de correlatieanalyses blijkt dat een aantal parameters samenhangt met de opbrengst:

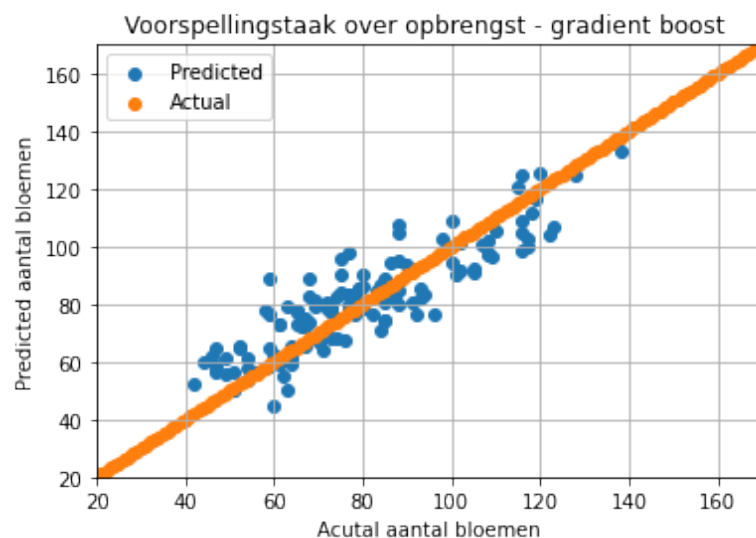
- Een hogere kastemperatuur, vooral aan het einde van de middag, hangt samen met een hogere opbrengst.
- De kasverwarmingstemperatuur, vooral 's morgens, hangt samen met een hogere opbrengst.
- Meer verduistering, vooral in het midden van de nacht, hangt samen met een hogere opbrengst.

- De raamopening luwe zijde, vooral tussen 0:00 en 6:00 uur en tussen 13:00 en 17:00 uur, hangt samen met een hogere opbrengst.

Dit betekent dat een kleine verandering van deze variabelen te relateren is aan een duidelijke verandering van de opbrengst. Opgemerkt wordt dat deze correlaties geen oorzakelijk verband aantonen; correlatie kan ontstaan doordat er een ander effect is met invloed op zowel de variabele als de opbrengst.

Uit de analyse op basis van *machine learning / AI* blijkt dat de opbrengst op basis van de gebruikte variabelen in behoorlijke mate voorspeld kan worden. Door de gegevens op verschillende manieren te analyseren, is gebleken dat deze voorspelling ook stabiel is, dus nauwelijks afhangt van hoe de gegevens verwerkt zijn.

Onderstaande figuur geeft de voorspelde opbrengst (verticaal) tegenover de werkelijke opbrengst (horizontaal). De blauwe punten zijn de voorspellingen; bij 100% overeenstemming zouden alle punten op de oranje lijn liggen. Gegeven het feit dat het algoritme alleen op basis van het kasklimaat is getraind en gewasvariabelen (zoals gewastemperatuur, water- en voedinggift en verdamping) nog niet zijn meegenomen, is de voorspellende waarde redelijk te noemen.



Er zijn veel variabelen die gerelateerd zijn aan de opbrengst, variërend over verschillende tijden op de dag, vooral: raamopening luwe zijde, windsnelheid, verduistering, verwarmingstemperatuur, gemeten PAR ('s middags) en relatieve luchtvochtigheid (RV, 's morgens). Daarbij geldt:

- Een lagere PAR is gerelateerd aan lagere opbrengst, maar een hogere PAR is niet sterk gerelateerd aan een hogere opbrengst; dit is zoals verwacht, omdat dan andere factoren beperkend worden.
- Een hogere RV is gerelateerd aan een hogere opbrengst, maar een lagere RV niet sterk aan een lagere opbrengst.

Aanbevelingen

Algemeen wordt opgemerkt dat de aanbevelingen voortkomen uit het onderzoek in de kas van Kwekerij Van Veen Gerbera uit De Lier en niet zonder meer te generaliseren zijn naar andere (gerbera)kassen. Wel bieden de resultaten handvatten aan kwekers voor onderzoek naar verbetering van het kasklimaat.

De periode waaruit gegevens gebruikt worden om te relateren aan de opbrengst kan nauwkeuriger worden afgestemd op het proces van bloei-inductie. Dat zal leiden tot een analyseperiode van meer dan de nu gehanteerde vier dagen tot aan de oogst.

Het is een meerwaarde om een camera (voor temperatuurmeting van het gewas) en een netto radiometer (voor inkomende en uitgaande straling) in te zetten op zowel blad als bloem. Daarmee kan beter inzicht verkregen worden in klimaatongelijkheid en in mogelijkheden voor een AI-algoritme. Daarmee kan bovendien de relatie tussen verduistering en opbrengst nader worden onderzocht.

Het verdient aanbeveling om meetwaarden van de irrigatiemomenten, van het drainpercentage en van de gift en het verbruik van nutriënten mee te nemen, zodat het AI-algoritme betere voorspellingen kan doen. Hetzelfde geldt voor ziektedruk en de kwaliteit van de bloem (bijvoorbeeld via de verhouding afkeur/goedkeur).

Dit onderzoek heeft uitgewezen dat een dichter netwerk van sensoren de voorkeur verdient boven een ruimer netwerk. De sensoren zouden nóg dichter bij elkaar kunnen worden geplaatst om de sterkste correlaties te vinden tussen gewasprestaties en klimatologische omstandigheden. Dan kan ook nauwkeuriger onderzocht worden wat de beïnvloedende factoren zijn voor afkeur.

Bij de gegevensanalyse is het waardevol om perioden met verschillende externe klimatologische omstandigheden (zoals zomer / winter) mee te nemen. Daarmee worden de conclusies die zijn gerelateerd aan de opbrengst robuuster.

De data-analyse kan versterkt worden door de verschillende variabelen te groeperen in twee categorieën: variabelen die direct samenhangen met de teelt en dus de opbrengst en variabelen die primair het kasklimaat beïnvloeden.

Om een kwantitatieve indruk te geven van de homogeniteit van het kasklimaat kunnen in toekomstige analyses de relaties tussen setpoints, externe klimatologische omstandigheden en de variaties daarin worden onderzocht. Dat kan leiden tot een 'klimaat uniformiteitsindex' van de kasomgeving.

De homogeniteit van het kasklimaat kan gekwantificeerd worden door een 'klimaat uniformiteitsindex' te definiëren, die zich baseert op de verschillen tussen posities in de kas en de variaties over de tijd. Daarvoor is van belang dat de kas voldoende meetpunten (sensoren) bevat. Onderzocht kan dan worden hoe de uniformiteit samenhangt met de controlstrategie en de setpoints, en wat daarop de invloed van het weer en de seizoenen is.

Uit de analyse van de data bleek dat het waardevol kan zijn om de data afkomstig van de klimaatcomputer, het draadloze sensornetwerk en het productie-oogstregistratiesysteem te integreren. Door deze integratie zou de statistische analyse continu kunnen worden uitgevoerd, in plaats van na een bepaalde periode. Daarvoor is het nodig om tools te ontwikkelen voor data-integratie, data-analyse met behulp van AI-technieken en voor het trekken van conclusies. Dit kan de teler helpen zijn teeltpraktijk aan te passen, of zelfs de kasklimaatstrategie te verbeteren.

Inhoudsopgave

	Samenvatting	2
1	Inleiding	8
1.1	Aanleidingen	8
1.2	Onderzoeksvragen en doelstelling van het project	9
1.3	Algemene projectgegevens en samenwerking.....	10
2	Opzet praktijkexperimenten	11
2.1	Fase 1: Proefperiode	11
2.2	Fase 2: Volledige experiment	14
3	Methode data-analyse	17
3.1	Databewerking	17
3.2	Statistische analyse	19
3.3	Correlatieanalyse en AI	19
4	Resultaten en bespreking	20
4.1	Fase 1: Proefperiode	20
4.2	Fase 2: Statistische analyse	22
4.3	Fase 2: Correlatieanalyse	26
4.4	Fase 2: AI en voorspellende variabelen voor opbrengst	29
4.5	Bespreking van de resultaten	32
5	Conclusies en aanbevelingen	34
5.1	Conclusies	34
5.2	Aanbevelingen	34
6	Ondertekening	36

1 Inleiding

De glastuinbouw maakt momenteel grote stappen op het gebied van energiebesparing. Toepassing van Het Nieuwe Telen (HNT) is in de gerberateelt gedurende de afgelopen 10 jaar gemeengoed geworden. Energiezuinige teeltconcepten zijn ontwikkeld in combinatie met veel aandacht voor de behoeften van de plant. Tegelijk leidt de energietransitie van minder (geen) fossiel en meer aardwarmte en elektrificeren ertoe dat kassen en klimaatsystemen complexer worden. Dit is een belangrijke aanleiding voor veel onderzoek naar de inzet van kunstmatige intelligentie in de glastuinbouw, met als doel efficiënter met energie om te gaan en om stappen te zetten in de richting van autonoom telen.

1.1 Aanleidingen

Dataverzameling, -analyse en -visualisatie is in toenemende mate een sleutel voor het telen in moderne kassen. Het is bekend dat het klimaat binnen de kas (in het bijzonder temperatuur en relatieve luchtvochtigheid) niet homogeen is, zelfs bij gebruik van geavanceerde klimaatbeheersingssystemen. Voor HNT is homogeniteit van het kasklimaat van groot belang. Door het klimaat op meerdere goed bekende posities in de kas te meten, ontstaat gedetailleerd inzicht in de inhomogeniteit qua tijd en plaats. Door het buitenklimaat erbij te betrekken, kan de relatie tussen binnen- en buitenklimaat zichtbaar worden.

Als eerste is het belangrijk inzicht te verkrijgen in deze 'big data'. Het is belangrijk om te onderzoeken of er bepaalde patronen zijn en of ruimtelijke klimaatverschillen samenhangen met andere variabelen. Om vervolgens de homogeniteit van het kasklimaat te kunnen verbeteren, moet rekening worden gehouden met deze klimaatverschillen. Kunstmatige intelligentie biedt hierbij kansen om complexe patronen en variabelen in beeld te brengen. Dit project voorziet in een eerste stap op dit gebied.

Klimaatverschillen

De belangrijkste oorzaken voor klimaatverschillen in de kas zijn:

- De klimaatbeheersing is gebaseerd op metingen op een enkele, specifieke locatie in de kas, vaak op één positie per een aantal hectare. Daardoor wordt de ruimtelijke variabiliteit van het kasklimaat niet in redelijk detail waargenomen.
- Klimaatsystemen (bijvoorbeeld verwarming, scherming, verneveling, luchtkanalen, enz.) in combinatie met externe weersomstandigheden creëren klimaatverschillen. Voorbeelden zijn bijvoorbeeld het afschot van de kas, waardoor de koude lucht boven het scherm naar de laagste zijde "rolt" en de kieren in het scherm waardoor er ongewenste 'luchtgaten' boven het gewas ontstaan.

Gegevens en aansturing

De inzet van sensoren gaat steeds vaker gepaard met het visualiseren van gegevens en een vorm van verwerking ervan, zoals 2D-kaarten voor temperatuur en luchtvochtigheid. Dergelijke dashboards presenteren de gemeten gegevens, maar door de veelheid aan variabelen (incl. buitenklimaat) is het voor een mens steeds lastiger om alles te overzien. Inzet van kunstmatige intelligentie kan uitkomst

bieden, in eerste instantie om inzicht en overzicht te verkrijgen, in tweede instantie ook om (meer decentraal) de kas te kunnen aansturen.

Productie en energie

Kleine verschillen in klimaatparameters leiden niet noodzakelijkerwijs tot verschillen in plantgroei en -productie. Het is wel van belang om te weten of bijvoorbeeld koude plekken in de kas zich vaak op dezelfde positie bevinden. Het is bekend dat planten hun eigen mechanismen hebben om zich aan de klimaatomstandigheden aan te passen. Zij reageren niet alleen op het klimaat van het moment, maar ook op dagelijkse of zelfs langer dan dagelijkse veranderingen. Het is wel algemeen bekend dat een gelijkmatiger klimaat zorgt voor een lager energieverbruik doordat onder andere de bandbreedte van de relatieve luchtvochtigheid kleiner kan worden gehouden.

1.2 Onderzoeksvragen en doelstelling van het project

Klimaatvariaties zijn momenteel niet zodanig gekwantificeerd dat een klimaatbeheersingssysteem de uniformiteit van de omgeving kan verbeteren. De doelstellingen zoals oorspronkelijk in het onderzoeksvoorstel opgenomen, zijn daarom aan het begin van dit project geherformuleerd:

- A. In kaart brengen van verschillen in de temperatuur en de relatieve luchtvochtigheid in een commerciële kas en deze gegevens aanvullen met gegevens van de klimaatcomputer (zowel de meetwaarden als de ingezette control acties) en productiegegevens van het gewas.
- B. Relateren van de klimaatgegevens van de klimaatcomputer, weergegevens en het sensornetwerk aan productiegegevens met behulp van kunstmatige intelligentietechnieken.
- C. Op basis van B) in kaart brengen van het effect van klimaatverschillen op de gewasontwikkeling en aantonen dat kunstmatige intelligentie op basis van het klimaat de opbrengst (aantal geoogste bloemen) kan voorspellen.

Oorspronkelijk was er ook een doelstelling D (gerelateerd aan onderzoeksvraag 3): “Op basis van B) in kaart brengen van het effect van de verschillende controle mogelijkheden van de klimaatcomputer op het wegnemen van de klimaatongelijkheid. Achterliggend doel is dat de tuinder scherpere setpoints kan aanhouden als de klimaatcomputer voor kleinere ongelijkheden kan zorgen.” Echter, dit kon niet onderzocht worden, omdat de computer en actuatoren in de kas niet ingericht zijn op klimaatsturing die nauwkeuriger plaatsgebonden is dan met de huidige één sensorpositie per een aantal hectare.

Ook werd in de oorspronkelijke doelstelling (C) de relatie tussen klimaatverschillen en ziektedruk benoemd. Omdat er ten tijde van het onderzoek geen ziektesymptomen zijn waargenomen, kon deze relatie niet worden aangetoond.

A en B kunnen resulteren in het definiëren van een 'klimaat uniformiteitsindex' als indicatie van de klimaatuniformiteit, bruikbaar voor een klimaatbeheersysteem.

1.3 Algemene projectgegevens en samenwerking

Titel: Kansen AI voor precisie Klimaat, "KAIK"
Looptijd: november 2019 t/m maart 2021
Samenwerking met: Kwekerij Van Veen Gerbera, De Lier
Demokwekerij Westland (inmiddels Vertify), Naaldwijk

2 Opzet praktijkexperimenten

Het project bestond uit twee fasen: een proefperiode waarin 8 weken 'proef' gemeten is en het volledige experiment, dat 20 weken in beslag heeft genomen. Het experimentgebied van zowel fase 1 als 2 bevat hetzelfde Gerbera-ras.

2.1 Fase 1: Proefperiode

In de proefperiode (fase 1) heeft de kick-off van het project plaatsgevonden, zijn de eerste experimenten uitgevoerd, en zijn de resultaten geanalyseerd ter voorbereiding op het volledige experiment (fase 2).

Sensortypes en -locaties in de kas zijn gedefinieerd. De locaties van de sensoren zijn gekozen om gegevens uit het hele kasgebied te krijgen en door deze gegevens te analyseren om de locatie met hoge inhomogeniteit te identificeren voor de tweede fase van het project. In de tweede fase werden de klimatologische omstandigheden gecorreleerd met de opbrengst (aantal snijbloemen) behaald in een bepaalde tijdsperiode.

Besloten is om in deze fase 10 draadloze geventileerde Sigrow-sensoren (zie Figuur 1) verspreid over de kas te plaatsen, die temperatuur, relatieve luchtvochtigheid en PAR meten (totaal dus feitelijk 30 sensoren). De selectie en posities van de sensoren zijn met de teler besproken, waarbij de sensoren op plekken in de kas zijn geïnstalleerd waarvan werd aangenomen dat deze verschilden in klimaat. Een draadloze zender is geïnstalleerd, die de meetgegevens via een 4G-verbinding naar een server stuurt, waarvan de gegevens kunnen worden gedownload.



Figuur 1. Geventileerde Sigrow-sensor met ingebouwde sensoren voor T, RV en PAR.

De specificaties van de draadloze sensoren zijn als volgt:

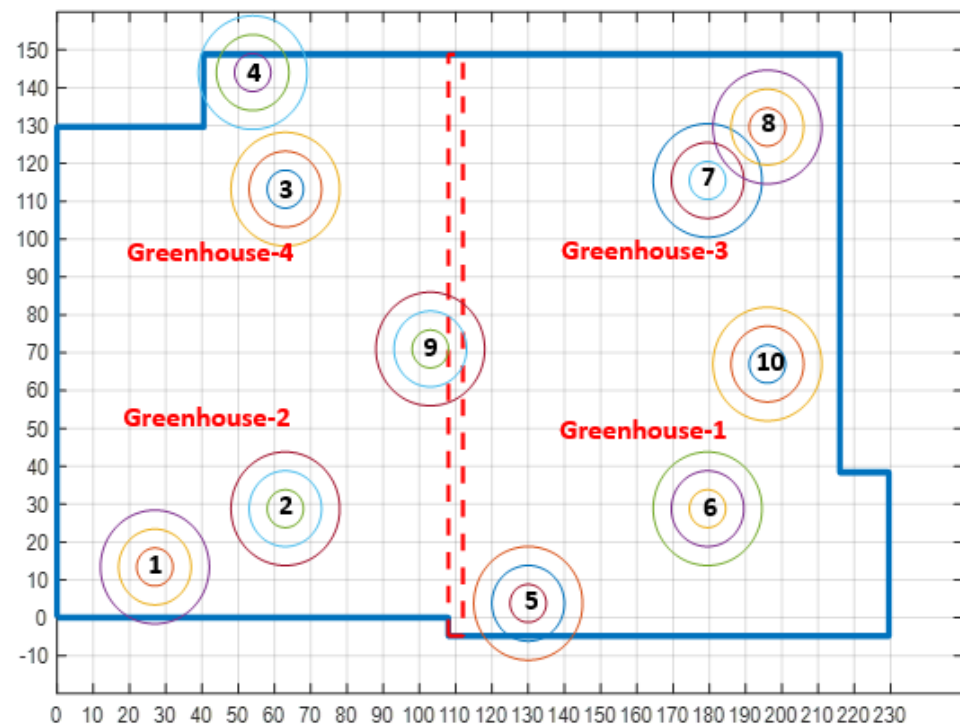
- Luchttemperatuur: sensor Sensirion SHT21:
 - resolutie 0,01°C
 - range 0-50°C
 - nauwkeurigheid 1% met max 0,3°C
- Relatieve luchtvochtigheid: sensor Sensirion SHT21:
 - resolutie 0,01%
 - range 0-100%
 - nauwkeurigheid 3% (RV<80%), 4% (80%<RV<90%) of 5% (90%<RV<100%)
- PAR: sensor van Sigrow:
 - resolutie: 1 $\mu\text{mol}/\text{m}^2/\text{s}$
 - range 0-2000 $\mu\text{mol}/\text{m}^2/\text{s}$
 - nauwkeurigheid 5%.

Besloten is om het CO₂-niveau niet te meten, omdat dat – zeker in de winter – naar verwachting nauwelijks varieert over de kas.

Het oorspronkelijke plan om luchtstromen te meten is losgelaten. De reden is dat de betrouwbaarheid van deze metingen onvoldoende geacht werd om zinvolle relaties te kunnen leggen met inhomogeniteit en productievariabelen.

De frequentie van gegevensverzameling is bepaald: meetgegevens van de sensoren, setpoints en meetgegevens van de klimaatcomputer en oogstgegevens die door Van Veen worden geregistreerd. In de analyses zijn ook de actuele weergegevens meegenomen.

Begin januari 2020 is de apparatuur geïnstalleerd. Het eerste praktijkexperiment is uitgevoerd gedurende de proefperiode van iets meer dan de vooraf geplande 8 weken, van begin januari tot medio maart 2020. Vervolgens zijn de ontvangen gegevens geanalyseerd, waarmee een goed beeld verkregen werd van gebieden in de kas waar inhomogeniteit een grote rol kan spelen. De plaatsing van de sensoren is weergegeven in Figuur 2.



Locatie:	Sensornummer: (in te vullen door TNO)	Padnummer	Pootnummer:	Camera Gearbox	Locatie binnen de kap
1	14808	111-112	18	Ja	bed nummer 5
2	18022	121-122	11	Nee	bed nummer 3
3	27929	175-176	11	Nee	bed nummer 3
4	47910	193-194	13	Nee	bed nummer 3
5	54496	*5-6	4	Ja	bed nummer 5
6	23505	22-21	14	Nee	bed nummer 3
7	6436	75-76	14	Nee	bed nummer 3
8	39250	86-85	21	Ja	bed nummer 5
9	20436	147-148	2	Ja	bed nummer 5
10	11700	46-45	21	Nee	bed nummer 3

Figuur 2. Plaatsing van de sensoren ten opzichte van de kas, fase 1. Afstanden in de plattegrond zijn een benadering, aangegeven in meters. "Greenhouse-1" t/m "Greenhouse-4" is de aanduiding die de kweker gebruikt voor gebieden in de kas, elk van hen wordt onafhankelijk aangestuurd door de klimaatcomputer.

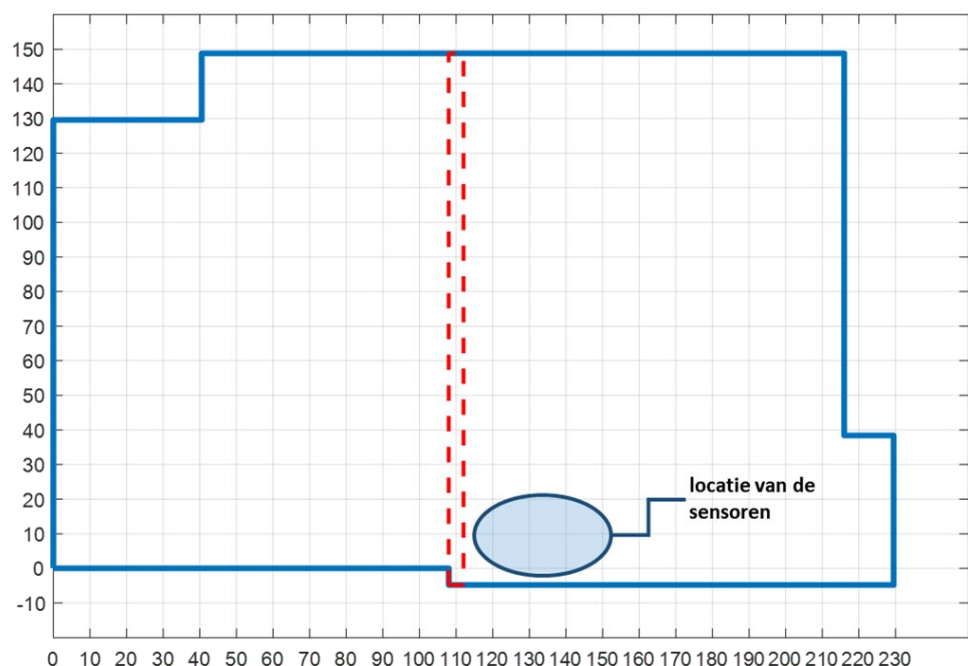
Op basis van deze resultaten zijn de plaatsing van de sensoren, de opzet van de data-analyse en de inzet van AI (kunstmatige intelligentie) voor het volledige experiment (fase 2) bepaald. Tevens is bepaald welke productieve variabele het best meetbaar was: opbrengst (oogstgegevens) of echte meeldauwdruk. Omdat echte meeldauw erg weinig voorkwam, is besloten om het aantal geoogste bloemen per oppervlakte, in een vastgesteld deel van een pad, als productieve variabele te hanteren. Deze variabele is direct gerelateerd aan de ontwikkelingssnelheid die de bloem heeft vertoond in de voorafgaande periode.

In deze gehele proefperiode is de kweker intensief betrokken geweest en is de kwaliteit van het verdere onderzoek zeker gesteld.

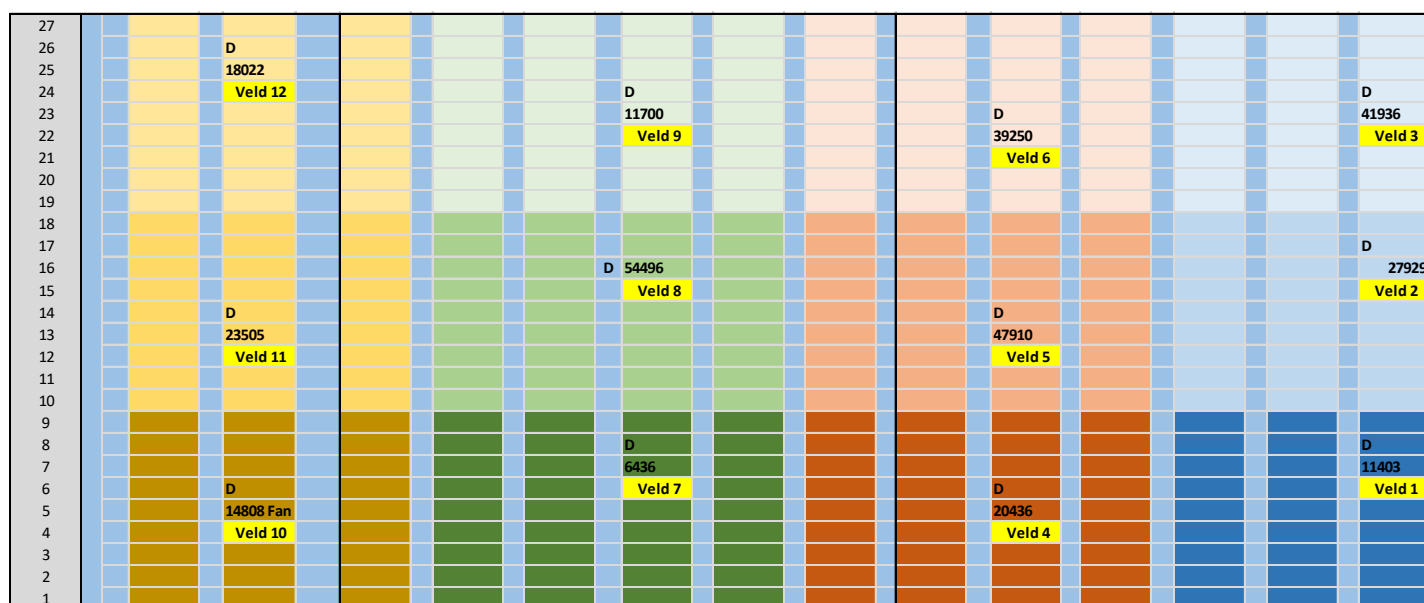
2.2 Fase 2: Volledige experiment

Na fase 1 is besloten om fase 2 uit te voeren met 12 sensoren geplaatst in één afdeling bij een hoek van de kas, waar in fase 1 duidelijke inhomogeniteit en variatie van het binnenklimaat zijn gesignaleerd; de kweker merkt ook verschillen op in het geoogste aantal bloemen. Besloten is om de experimenten niet uit te voeren in de oorspronkelijk geplande drie perioden van 4-5 weken, maar in twee perioden van 10 weken, wat ook beter past bij de tijd die nodig is voor de vorming van gerberabloemen. De reden is dat voor de data-analyse voldoende gegevens nodig zijn per configuratie. Een periode van 5 of zelfs 7 weken zou daarvoor te kort zijn. De twee perioden liepen van medio oktober tot eind december 2020 en van begin januari tot medio maart 2021.

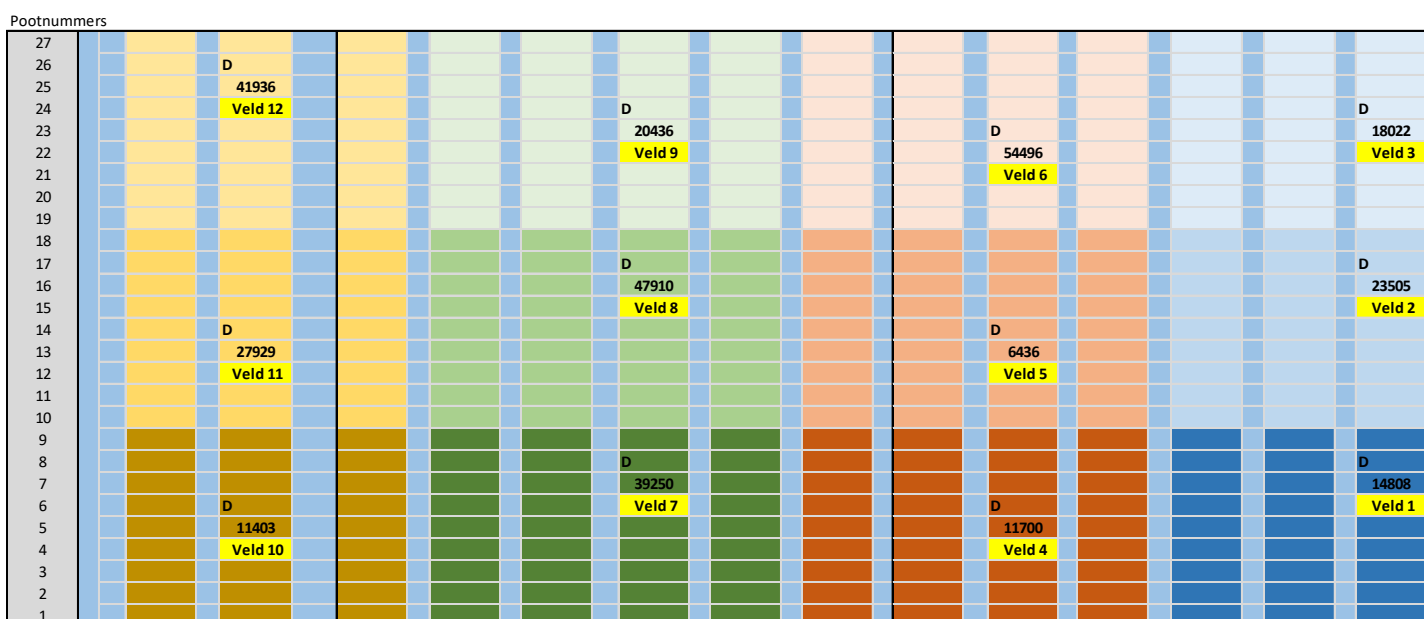
De plaatsing van de sensoren is weergegeven in Figuur 3, Figuur 4 en Figuur 5.



Figuur 3. Plaatsing van de sensoren 1 t/m 12 ten opzichte van de kas, fase 2. Afstanden in de plattegrond zijn een benadering, aangegeven in meters.



Figuur 4. Plaatsing van de sensoren ten opzichte van de kas, fase 2 (eerste periode). De rijen zijn in deze figuur verticaal weergegeven; de rechterzijde komt overeen met de onderzijde in Figuur 3.



Figuur 5. Plaatsing van de sensoren ten opzichte van de kas, fase 2 (tweede periode). De rijen zijn in deze figuur verticaal weergegeven; de rechterzijde komt overeen met de onderzijde in Figuur 3.

In de twee perioden van 10 weken zijn gegevens verzameld:

- Oogstgegevens: aantal geoogste en aantal afgekeurde bloemen.
- Sensor-meetgegevens: temperatuur, relatieve luchtvochtigheid, PAR.
- Gegevens uit de klimaatcomputer voor buiten: buitentemperatuur, windsnelheid en windrichting.
- Gegevens uit de klimaatcomputer voor binnen: kastemperatuur, relatieve luchtvochtigheid, raamopening aan de wind- en aan de luwe zijde, PAR instraling, verwarmingstemperatuur, %verduistering door scherm.

Voor wat betreft oogstgegevens is besloten geen ziektedruk te meten, omdat de verwachting van de kweker was dat er geen sprake zou zijn van ziektes/plagen. Achteraf is dit inderdaad het geval geweest. Daarom is alleen het totaal aantal geoogste bloemen en het aantal afgekeurde bloemen geteld. Afkeur vindt plaats op basis van de visueel waargenomen kwaliteit door degene die plukt.

De gegevens zijn geanalyseerd op drie manieren:

- Statistische analyses: basisanalyses leidend tot 'heatmaps' van de waarden van de variabelen verspreid over de kas, inclusief het verloop van de waarden van de variabelen over de tijd. Dit is gerelateerd aan de opbrengst (oogstgegevens).
- Correlatieanalyses: 1-op-1 correlatie tussen variabelen met de opbrengst.
- Machine learning: om complexe relaties tussen multivariabele en opbrengst te vatten inclusief uitlegbaarheid van variabelen op de opbrengst.

De resultaten van de analyses en de consequenties voor de aansturing in de kas zijn met de partners besproken in drie bijeenkomsten in maart en mei 2021.

3 Methode data-analyse

De gegevens zijn verwerkt tot een grote dataset, waarop analyses mogelijk zijn. De analyses zijn onderverdeeld in een statistische analyse, een correlatieanalyse en als laatste een machine learning analyse. Deze analysemethoden kunnen in toenemende mate complexe relaties tussen de variabelen blootleggen met als doel om te bepalen hoe de kas optimaal kan functioneren.

3.1 Databewerking

Er is voor gekozen om alle beschikbare variabelen mee te nemen in de analyse. Dat biedt de mogelijkheid om ook minder voor de hand liggende relaties tussen de opbrengst en variabelen op het spoor te komen. Er zijn variabelen gemeten door twee verschillende bronnen: het netwerk van 10 sensoren (fase 1) en later 12 sensoren (fase 2) en de sensoren van de klimaatcomputer. In Tabel 1 staan de afkortingen per gemeten variabele en door welke bron die gemeten is.

Tabel 1. Overzicht van gemeten variabelen door de klimaatcomputer (KC) en door het sensornetwerk (SN) en de gebruikte afkortingen per klimaatparameter. De temperatuur wordt gemeten door zowel de KC (één meetbox, meet KT) als het SN (10 of 12 sensoren, meten T), echter op verschillende posities.

Afkorting	Uitleg variabele	Afkorting	Uitleg variabele
KC	Klimaatcomputer	KLZ	Raamopening aan luwe zijde (%) door KC
SN	Sensor Network	KRV	RV Klimaat (%) door KC
T	Temperatuur (°C) door SN	KVT	Verwarmingstemperatuur (°C) door KC
P	Photosynthetic Active Radiation (PAR) ($\mu\text{mol}/\text{m}^2/\text{s}$) door SN	KST	Instraling PAR ($\mu\text{mol}/\text{m}^2/\text{s}$) door KC
H	Relatieve luchtvochtigheid (%) door SN	KWS	Windsnelheid (m/s) door KC
KT	Kas temperatuur (°C) door KC	KWR	Windrichting (graden) door KC
KV	Kas verduistering (%) door KC	KBT	Buitentemperatuur (°C) door KC
KWZ	Raamopening aan windzijde (%) door KC		

De meetwaarden (elke vijf minuten) van de variabelen zijn per uur gemiddeld, zodat er per uur één waarde is. De oogstgegevens zullen gecorreleerd worden aan het gemiddelde van de klimaatgegevens gedurende vier dagen voorafgaand aan de oogst. Daarom is van alle uurlijkse waarden het gemiddelde genomen over de laatste vier dagen.

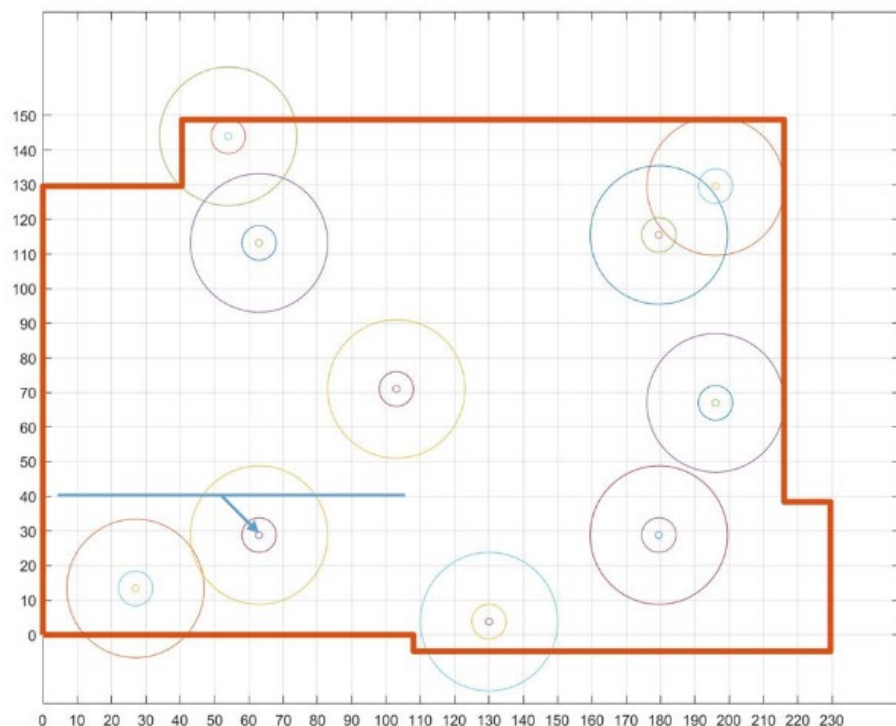
Deze variabelen worden als volgt weergegeven (voorbeeld):

- T13: gemiddelde temperatuur tussen 13:00 en 14:00 over de afgelopen 4 dagen tot oogstmoment.
- KWR21: gemiddelde windrichting tussen 21:00 – 22:00 over de afgelopen 4 dagen tot oogstmoment.

3.1.1 *Fase 1 van het project: matchen van de opbrengst met gegevens van de klimaatcomputer en de sensoren*

In fase 1 is er met 10 sensoren gemeten over de gehele kas. Er zijn 189 rijen met productie. Dit levert 1443 unieke datapunten voor de opbrengst op doordat er 7 à 8 keer geoogst is over de gemeten periode. Iedere rij heeft een dichtstbijzijnde sensor. Van de helft van de rij die zich het dichtst bij die sensor bevindt, is dagelijks de opbrengst geteld. Die is in verband gebracht met de meetwaarden van de sensor, om verbanden tussen de opbrengst en het binnenklimaat te onderzoeken.

Een voorbeeld van de matching tussen een rij en een sensor is weergegeven in Figuur 6. Hier wordt de blauwe lijn (een halve rij) gelinkt aan de sensor met de pijl. Omwille van de simpliciteit is voor deze methode van matching gekozen. Voor een vervolg kan een geavanceerdere methode gekozen worden waarbij sensoren geïnterpoleerd worden over de rijen.



Figuur 6. Blauwe lijn: halve rij waarvan de opbrengst wordt gerelateerd aan de meetwaarden van de dichtstbijzijnde sensor (blauwe pijl).

Aan het einde van fase 1 zijn de resultaten geanalyseerd door de opbrengsten te relateren aan de sensormetingen. Vervolgens is besloten om in fase 2 een andere set-up te kiezen. Daarbij zijn de sensoren dichter bij elkaar in een kleiner deel van de kas geplaatst om de afstand tussen het gebied van de opbrengst en de sensor te verkleinen. Zo kan nauwkeuriger de relatie tussen metingen en opbrengst worden onderzocht.

3.1.2 *Fase 2 van het project: matchen van de opbrengst met gegevens van de klimaatcomputer en de sensoren*

Het klimaat van het gebied in de kas waar de sensoren stonden opgesteld, wordt geregeld door klimaatcomputer 1. Gedurende 18 weken zijn alle variabelen gemeten, van 23-10-2020 tot en met 02-03-2021. Er is twee keer per week geoogst, eenmaal is er 3 keer per week geoogst. De 36 oogsten van de 12 velden hebben dus geleid tot 432 datapunten.

Voor dezelfde aanpak als in fase 1 is gekozen: sensorgegevens worden gekoppeld aan oogstgegevens. De sensoren zijn geplaatst in een strakker grid in een kleiner gebied, en halverwege de meetperiode zijn de sensoren iets verplaatst, zie Figuur 4 en Figuur 5. Het doel hiervan was te onderzoeken of dat leidt tot andere uitkomsten van de analyses.

3.2 Statistische analyse

Er is gestart met een statistische analyse. Hierbij zijn alle getelde opbrengsten in drie groepen met gelijke hoeveelheid datapunten verdeeld: lage opbrengst, hoge opbrengst en een middengroep. Zo is onderzocht of een verandering van een variabele (bijvoorbeeld de gemiddelde temperatuur op een bepaalde tijd van de dag) een relatie heeft met de hoogte van de opbrengst.

Daarnaast zijn er visualisaties als [heatmaps](#) gemaakt over de verdeling van de variabelen over de plaats in de kas. Zo zijn verschillen in de kas direct inzichtelijk.

3.3 Correlatieanalyse en AI

Na de statistische analyse volgt een correlatieanalyse. Hierbij is onderzocht per variabele hoe die zich verhoudt tot de opbrengst. Een hoge positieve of negatieve correlatie betekent dat een hogere waarde van de variabele samengaat met een hogere of lagere opbrengst. Zo kan onderscheid gemaakt worden tussen variabelen die sterk of juist weinig samenhangen met de opbrengst. Hierbij is het belangrijk de volgende eigenschappen van correlatieanalyse te duiden: 1) correlaties tonen puur een relatie aan, en gaan niet over oorzaak en gevolg (dat is causatie). 2) Standaard correlatieanalyse is vooral geschikt voor lineaire relaties, en achterhaalt geen complexe relaties zoals curves of meerdere externe variabelen die enkel samen een invloed hebben.

Vervolgens zijn verschillende AI (machine learning, ML) modellen geïmplementeerd, waarbij [Gradient Boost](#) de beste resultaten behaalde gekeken naar de Mean Squared Error (gemiddelde afwijking van werkelijkheid) waarop geoptimaliseerd is. Vervolgens is er met [SHAP](#)-techniek (SHapley Additive exPlanations) gekeken naar de samenhang van een variabele met de uitkomsten. De SHAP-waarde geeft ook aan of een hoge of juist een lage waarde van de variabele positief of negatief gecorreleerd is met de uitkomst (opbrengst).

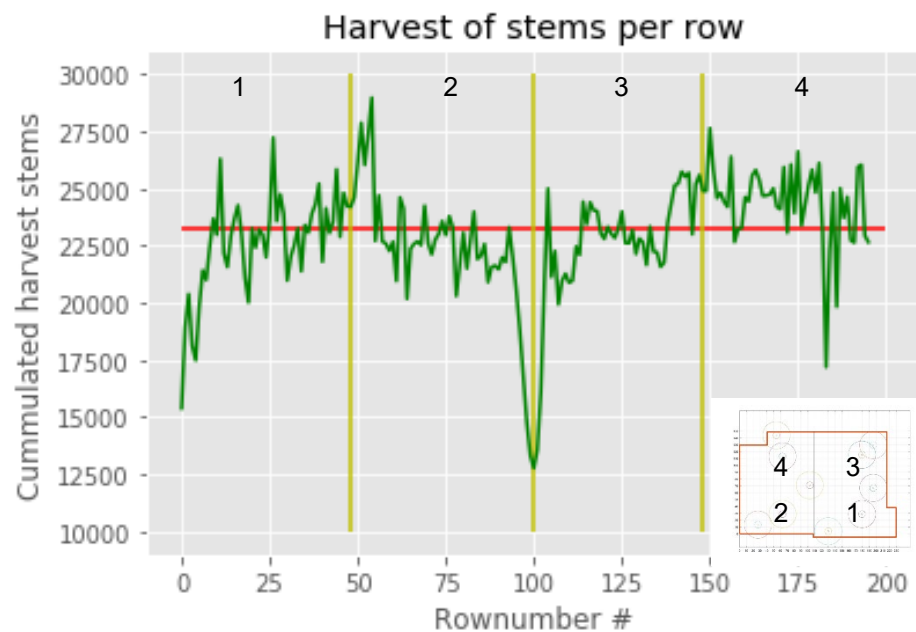
In combinatie met de heatmaps geeft deze AI-analyse inzicht in de complexe relaties tussen verschillende waarden van de variabele en de opbrengst, verspreid over het onderzochte gebied in de kas.

4 Resultaten en bespreking

Omdat fase 1 de proefperiode was, zijn die gegevens alleen beoordeeld op correlaties tussen de gemeten variabelen en de opbrengst (paragraaf 4.1). Op basis van de uitkomsten is fase 2 ingericht. De analyseresultaten van fase 2 zijn opgesplitst in statistische, correlatie- en machine learning resultaten (resp. paragraaf 4.2, 4.3 en 4.4).

4.1 Fase 1: Proefperiode

De opbrengst in fase 1 is weergegeven in Figuur 7. Per rij is de totale opbrengst over de gehele periode omgerekend naar een opbrengst naar een gelijke lengte van 108 m (de mediaan van rijlengte), zodat de opbrengsten in alle rijen vergeleken kunnen worden. De rode lijn geeft de gemiddelde opbrengst per rij weer. De gele lijnen verdelen de kas in dezelfde vier gebieden als weergegeven in Figuur 2: gebied 1 is links weergegeven (vanaf rij 1); gebied 4 is rechts weergegeven (tot rij 189).



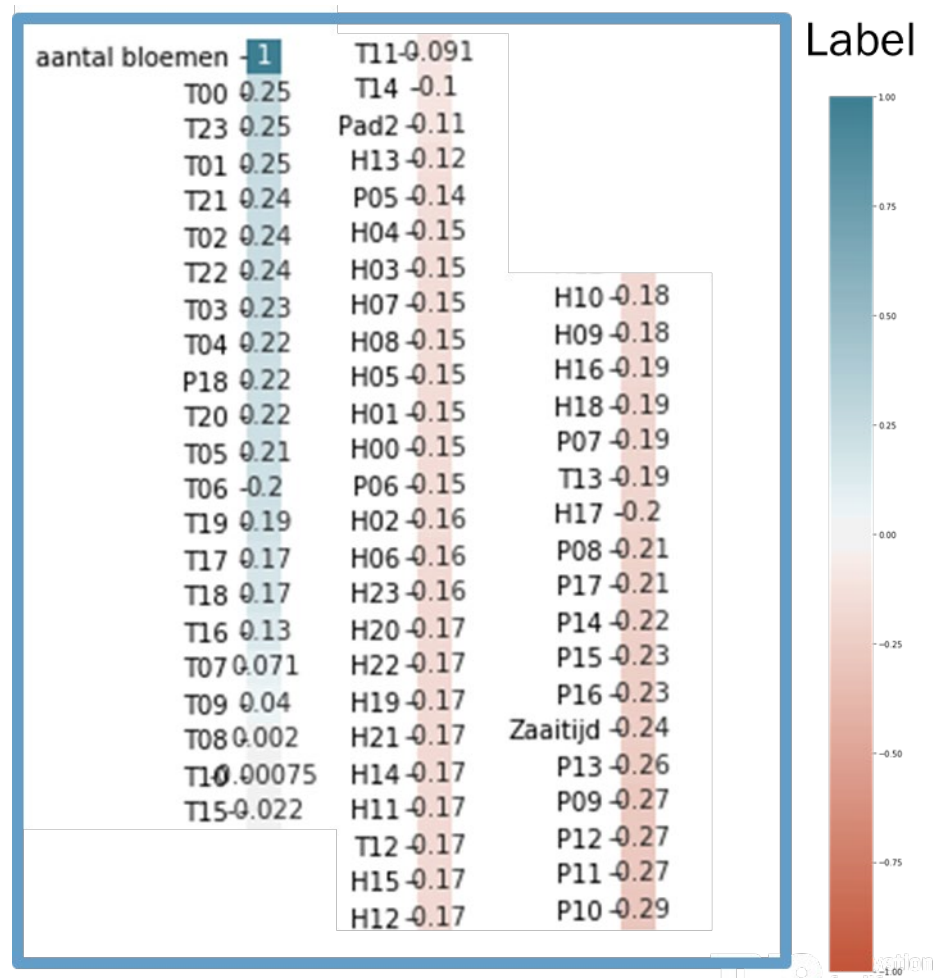
Figuur 7. Opbrengst gedurende heel fase 1 voor de 189 rijen.

Er is duidelijke fluctuatie te zien in opbrengst over de kas. Er zijn gebieden die dicht bij elkaar liggen en toch een heel verschillende opbrengst hebben. De eerste rijen in de kas (links in de grafiek) hebben een beduidend lagere opbrengst (er is genormaliseerd voor rij-lengte).

De opbrengsten zijn gecorreleerd met de uurlijkse gemiddelden van de gemeten variabelen over de voorgaande vier dagen. In de correlatieplot (Figuur 8) zijn lage correlaties te zien tussen de opbrengst (aantal bloemen) en de verschillende variabelen van het sensorennetwerk. Blauw betekent een hoge positieve correlatie, bijvoorbeeld: bij een hogere T00 (temperatuur tussen 0:00 – 1:00) hoort een hogere

opbrengst. Rood geeft een hoge negatieve correlatie aan, bijvoorbeeld: bij een lagere P10 (PAR tussen 10:00 - 11:00) hoort een hogere opbrengst. Lichtere kleuren geven een lagere correlatie weer; bijvoorbeeld: T08 (temperatuur tussen 8:00 – 9:00) hangt nauwelijks samen met de opbrengst.

Er zijn twee extra variabelen te zien Pad2 = padnummer, en Zaaityd = tijd tussen vorige oogst en deze oogst, die in fase 1 afhangt van het vorige oogstmoment. Met de telers is gesproken over een betere manier om de variabelen te relateren aan de opbrengst. Er is toen voor gekozen om in fase 2 de variabelen van de laatste 4 dagen te beschouwen.



Figuur 8. Correlaties tussen opbrengst en gemeten variabelen. Opbrengst heeft de sterkste correlatie met temperatuur (T) en PAR (P); in mindere mate ook met relatieve luchtvochtigheid (H).

Vanwege lage correlaties en slechte voorspellingen is er onderzocht of de afstand tot een sensor van invloed is op correlaties en voorspellingen. Dat is gedaan voor alle sensoren door correlaties te bepalen van sensormetingen met opbrengsten in een deel van de rij die zich dicht bij de sensor bevinden. Die afstand blijkt inderdaad invloed te hebben: bij afstanden tot een maximum van 20 meter van een rij tot een sensor werden de correlaties beter naarmate de afstand kleiner werd.

Voorbeelden:

- De correlatie van T03 is 0.35 bij een afstand onder 20m.
- De correlatie van P18 is 0.37 bij een afstand onder 20m.

Bij afstanden tot 10 meter tot een sensor verbeterden de correlaties nog verder, echter slechts 10% van de opbrengstmetingen valt binnen die afstand.

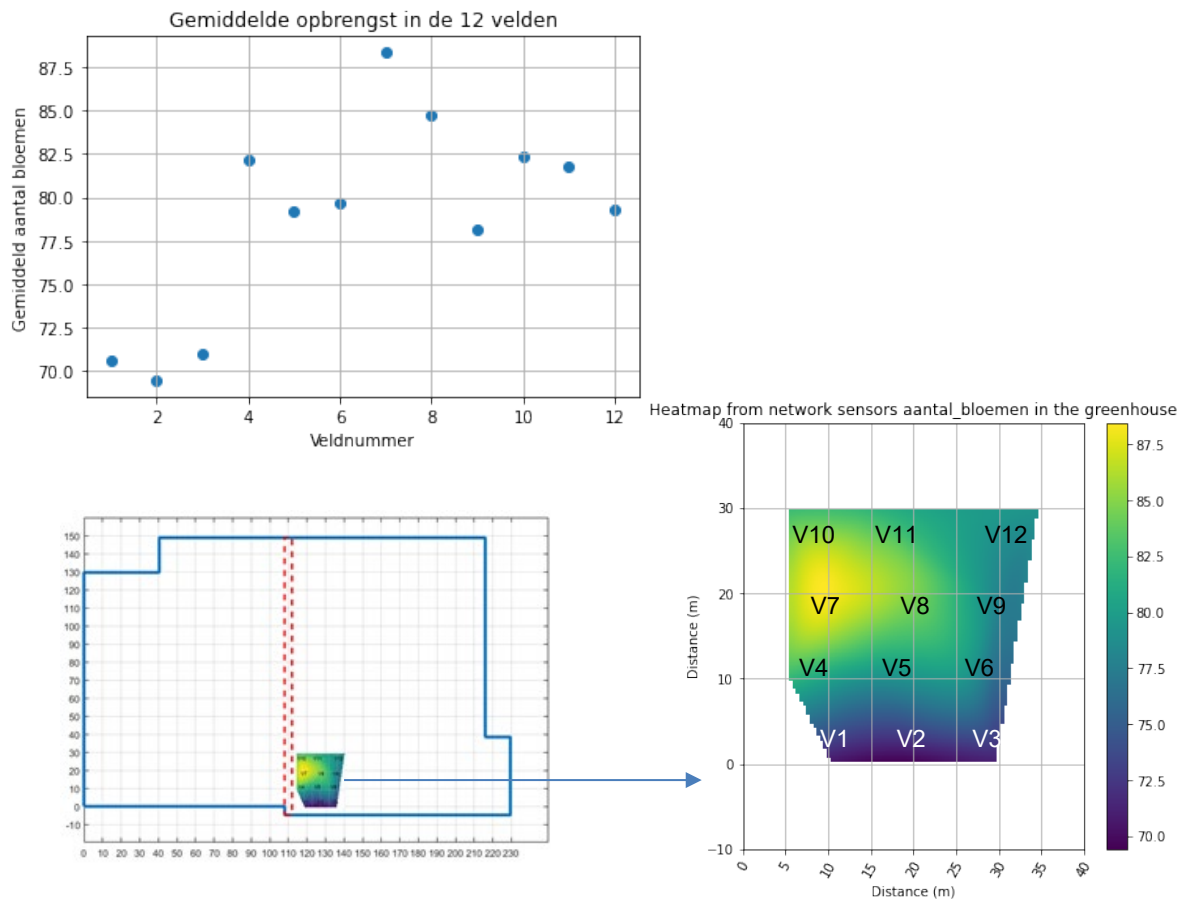
Voorbeelden:

- De correlatie van T03 is 0.43 bij een afstand onder 10m.
- De correlatie van P18 is 0.41 bij een afstand onder 10m.

Om deze reden is besloten een nieuwe set-up in fase 2 te hanteren, waarbij de sensoren dichter bij een beperkt aantal rijen geplaatst worden. Zo kunnen er betere conclusies getrokken worden over de samenhang van variabelen met de opbrengst.

4.2 Fase 2: Statistische analyse

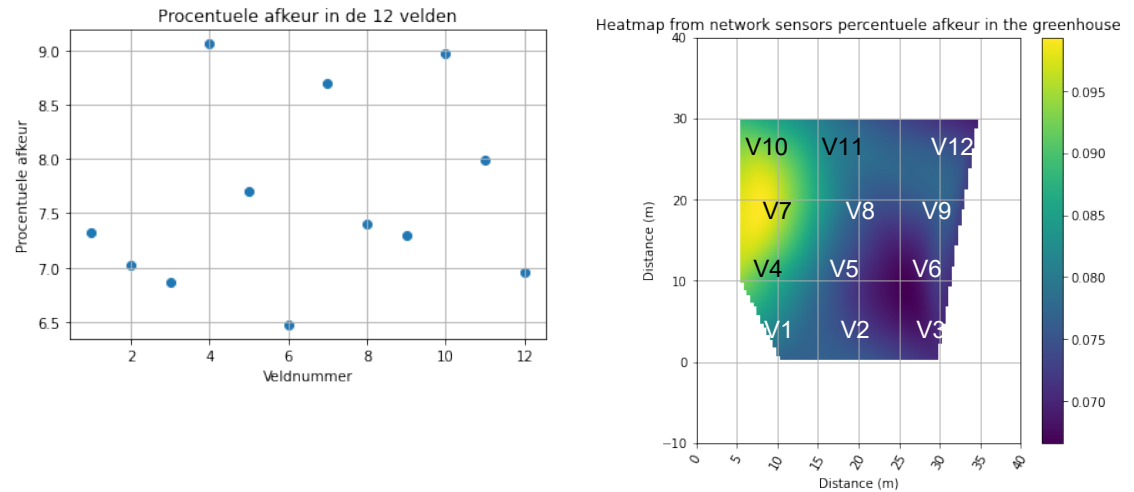
De opbrengst in fase 2 gemeten over de 12 velden is weergegeven in Figuur 9. De linker figuur toont de gemiddelde opbrengst (inclusief afkeur) per oogst voor elk van de 12 velden. Aan de rechterkant staat de heatmap van de gemiddelde opbrengst verspreid over het onderzochte deel van de kas.



Figuur 9. Opbrengst per veld (links) en als 'foto van boven' van het onderzochte gebied (rechts, heatmap). Naarmate het gebied geler kleur, is de opbrengst hoger. Velden 1, 2 en 3 liggen onderaan van links naar rechts; veld 12 is rechtsboven. Het raster correspondeert met afstanden in meters, niet met de velden. De schuine zijden in de heatmap zijn het gevolg van interpolatie van de resultaten.

De opbrengst (inc. afgekeurde bloemen) is gemiddeld 20% hoger aan de kant van het pad (velden 4, 7, 10, links in de heatmap) dan aan de zijkant van de kas (velden 1, 2 en 3, onderaan in de heatmap).

Figuur 10 toont de relatieve afkeur, het deel van de opbrengst dat is afgekeurd. Na correctie voor afkeur is de opbrengst aan de kant van het pad nog steeds 17% hoger dan aan de zijkant van de kas.

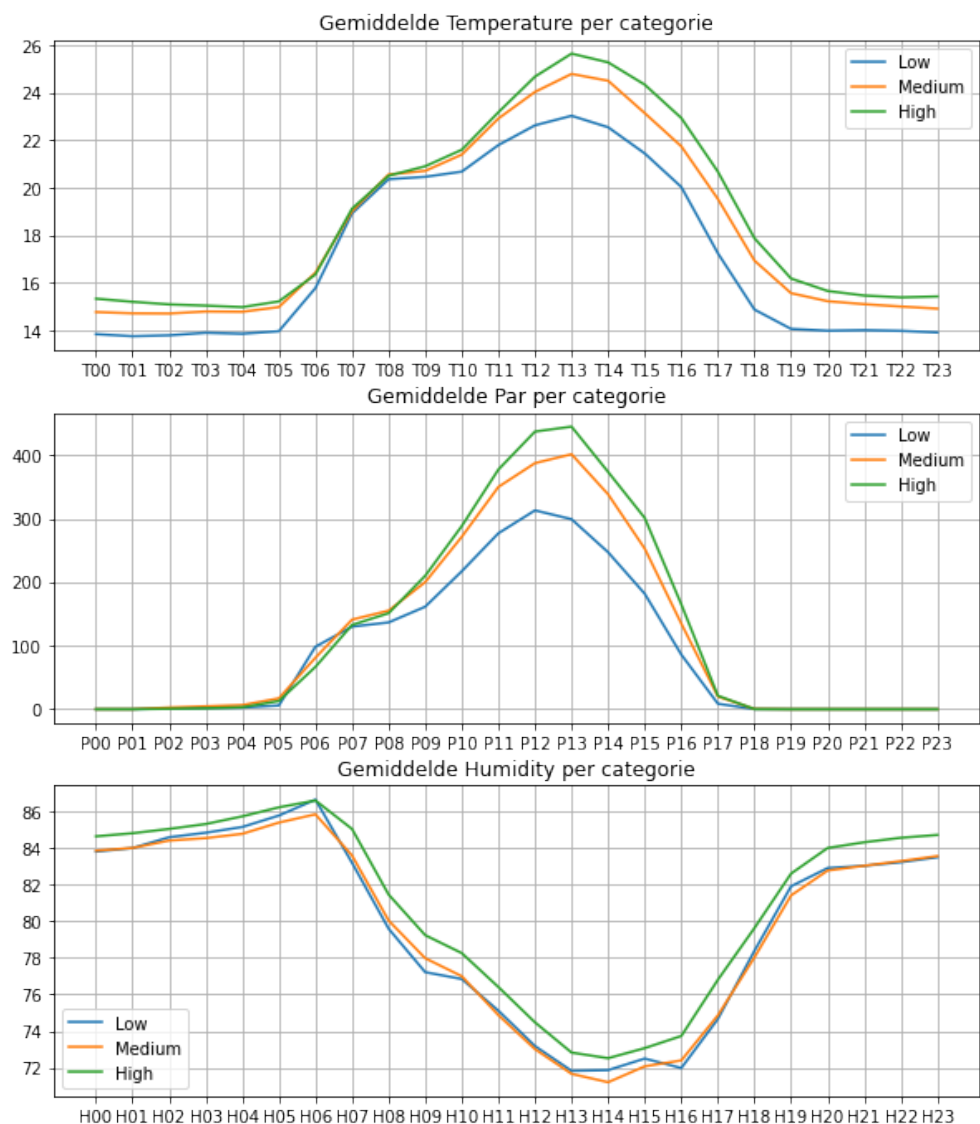


Figuur 10. Relatieve afkeur per veld (links) en als 'foto van boven' van het onderzochte gebied (rechts, heatmap). Naarmate het gebied geler kleurt, is de afkeurfractie hoger. Velden 1, 2 en 3 liggen onderaan van links naar rechts; veld 12 is rechtsboven. Het raster correspondeert met afstanden in meters, niet met de velden. De schuine zijden in de heatmap zijn het gevolg van interpolatie van de resultaten.

De gemeten opbrengsten zijn verdeeld in drie categorieën, zodanig dat iedere categorie evenveel meetpunten heeft: een lage (blauwe), gemiddelde (oranje) en hoge (groene) opbrengst. Er is voor niet meer dan drie categorieën gekozen, omdat het aantal meetpunten per categorie dan te laag wordt, waardoor de correlatieanalyse minder betrouwbaar zou worden.

Daarna zijn voor iedere categorie de waarden van de bijbehorende gemeten variabelen bepaald, gemiddeld over de gehele meetperiode per uur van de dag. De gemiddelde waarden binnen die categorieën van de temperatuur, de relatieve luchtvochtigheid en PAR van het sensornetwerk zijn weergegeven in Figuur 11.

Opgemerkt wordt dat in Figuur 11 de getoonde verschillen in temperatuur, PAR en relatieve vochtigheid niet alleen veroorzaakt worden door de lokale verschillen (inhomogeniteit) in het kasklimaat. Doordat de gegevens zijn verzameld in een periode van 20 weken, speelt ook de variatie van de opbrengst over de maanden een rol.



Figuur 11. Temperatuur, PAR en relatieve luchtvochtigheid per uur van de dag, ingedeeld in drie categorieën: de laagste 1/3 van de metingen (blauw), de middelste (oranje) en de hoogste (groen).

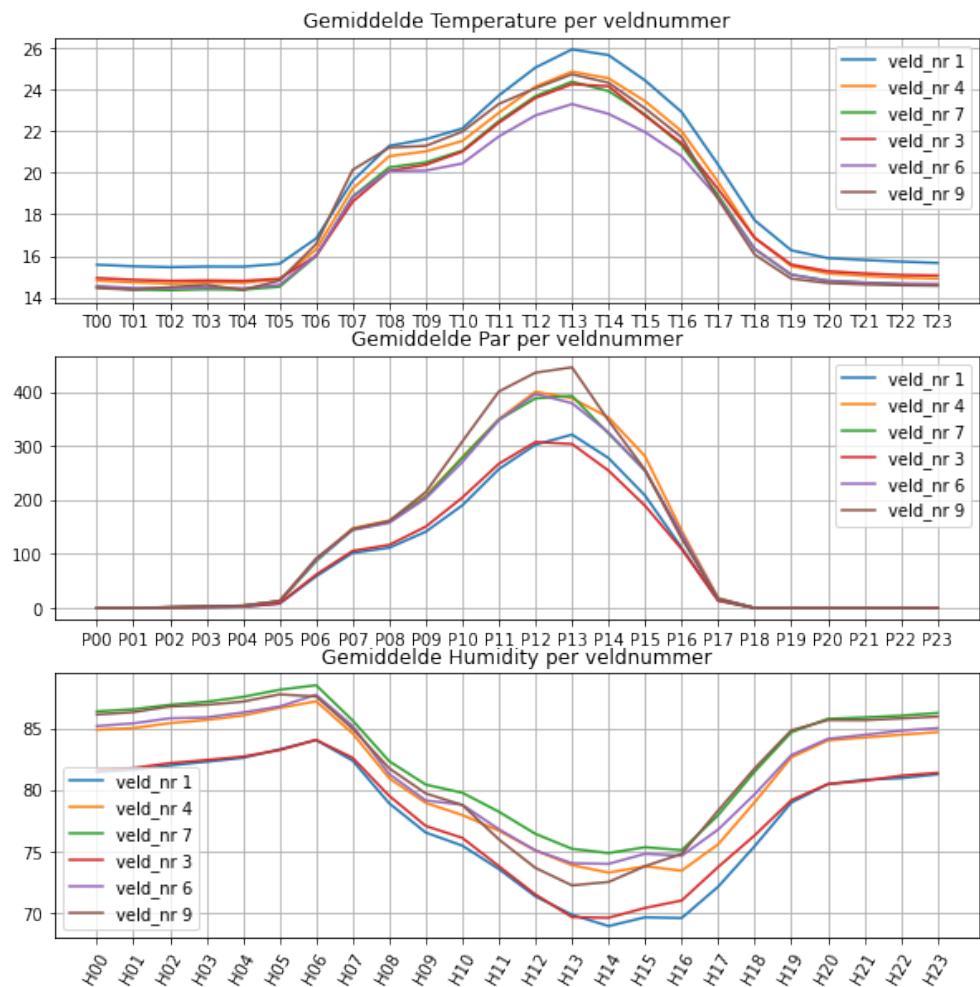
Bij de temperatuurverdeling is een hogere temperatuur (groene lijn) gerelateerd aan een hogere opbrengst.

Bij de PAR is een hogere PAR gedurende de dag-uren gecorreleerd aan een hogere opbrengst. Dit verschil is het grootst in de periode 11:00 – 16:00. Er is één uur op de dag waarop deze relatie omdraait, tussen 6:00 – 7:00.

Bij de relatieve luchtvochtigheid is er weinig variatie tussen een hoge, lage en gemiddelde opbrengst. Er is een kleine verhoging te zien van de lijn die de hogere opbrengst representeert (groen).

Vervolgens zijn temperatuur, PAR en relatieve luchtvochtigheid uitgesplitst per sensor. De drie grafieken staan in Figuur 12.

Variabelen per sensor

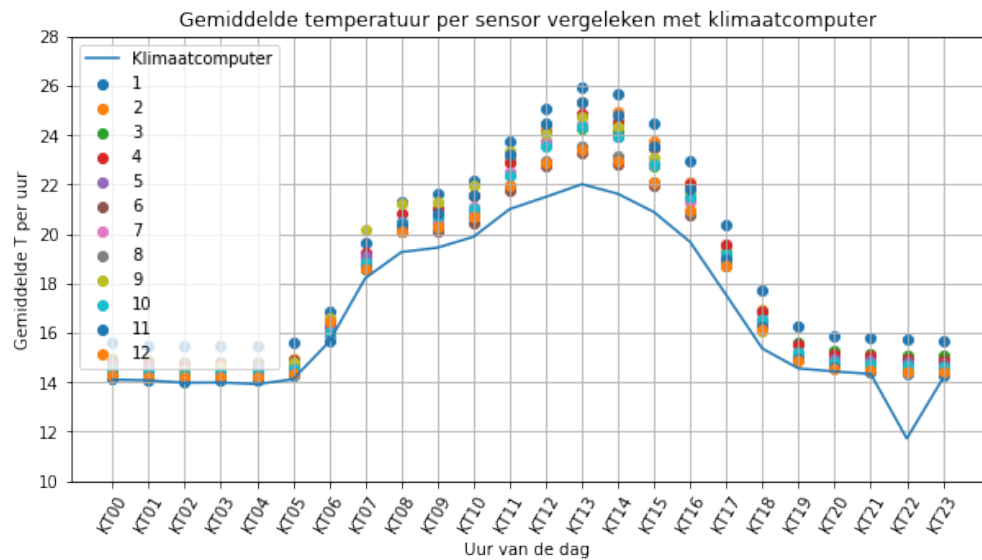


Figuur 12. Temperatuur, PAR en relatieve luchtvochtigheid gemiddeld over de dag, per sensor/veld.

Figuur 12 toont dat er vooral overdag flinke klimaatverschillen zijn tussen de velden:

- Het verschil tussen de hoogste en laagste temperatuur loopt op tot 3 °C.
- Het verschil in PAR varieert midden op de dag van ca. 300 tot 450 $\mu\text{mol}/\text{m}^2/\text{s}$.
- Het verschil tussen de hoogste en laagste relatieve vochtigheid loopt op tot 7%.

De temperatuur zoals gemeten door de klimaatcomputer is vergeleken met de sensormetingen op de verschillende velden in Figuur 13. De ononderbroken lijn is de klimaatcomputer, de stippen geven de verschillende sensoren op velden 1 t/m 12 weer.



Figuur 13. Gemiddelde temperatuur voor ieder uur van de dag gemeten door de klimaatcomputer (blauwe lijn) en door de 12 sensoren (gekleurde stippen).

Het blijkt dat de klimaatcomputer over het algemeen een lagere temperatuur meet dan de sensoren. Dit is een bekend verschijnsel bij draadloze sensoren. De belangrijkste oorzaak daarvoor is dat de kunststof behuizing van de sensoren een deel van de straling absorbeert en daardoor opwarmt. Dat leidt tot een iets hogere temperatuur in de sensorbehuizing. De sensoren zijn weliswaar geventileerd, maar niet continu, zodat deze temperatuurverhoging toch kan plaatsvinden. Deze verklaring is in lijn met de waarneming dat het verschil bij daglicht groter is dan wanneer het donker is.

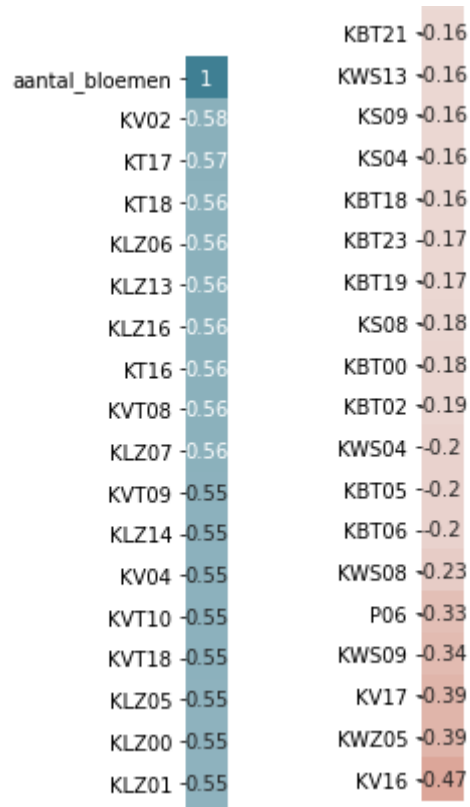
Er geen correctie uitgevoerd voor dit effect, om te voorkomen dat de verschillen in de temperaturen ter plaatse van de sensoren deels 'weggecorrigeerd' zouden worden.

Tussen 22:00 – 23:00 zijn er afwijkende temperaturen gemeten. Achteraf is nagegaan dat dit te wijten is aan een meetfout; dit meetpunt is niet meegenomen in de verdere analyses.

4.3 Fase 2: Correlatieanalyse

Correlaties tussen de 312 variabelen en de opbrengst zijn met een correlatieplot weergegeven in Figuur 14. Hierbij geeft wederom de blauwe kleur aan dat er positieve correlatie is met de opbrengst. De rode kleur geeft aan dat er een negatieve correlatie is met de opbrengst. In de figuur zijn slechts de variabelen weergegeven die de sterkste correlaties met de opbrengst hebben.

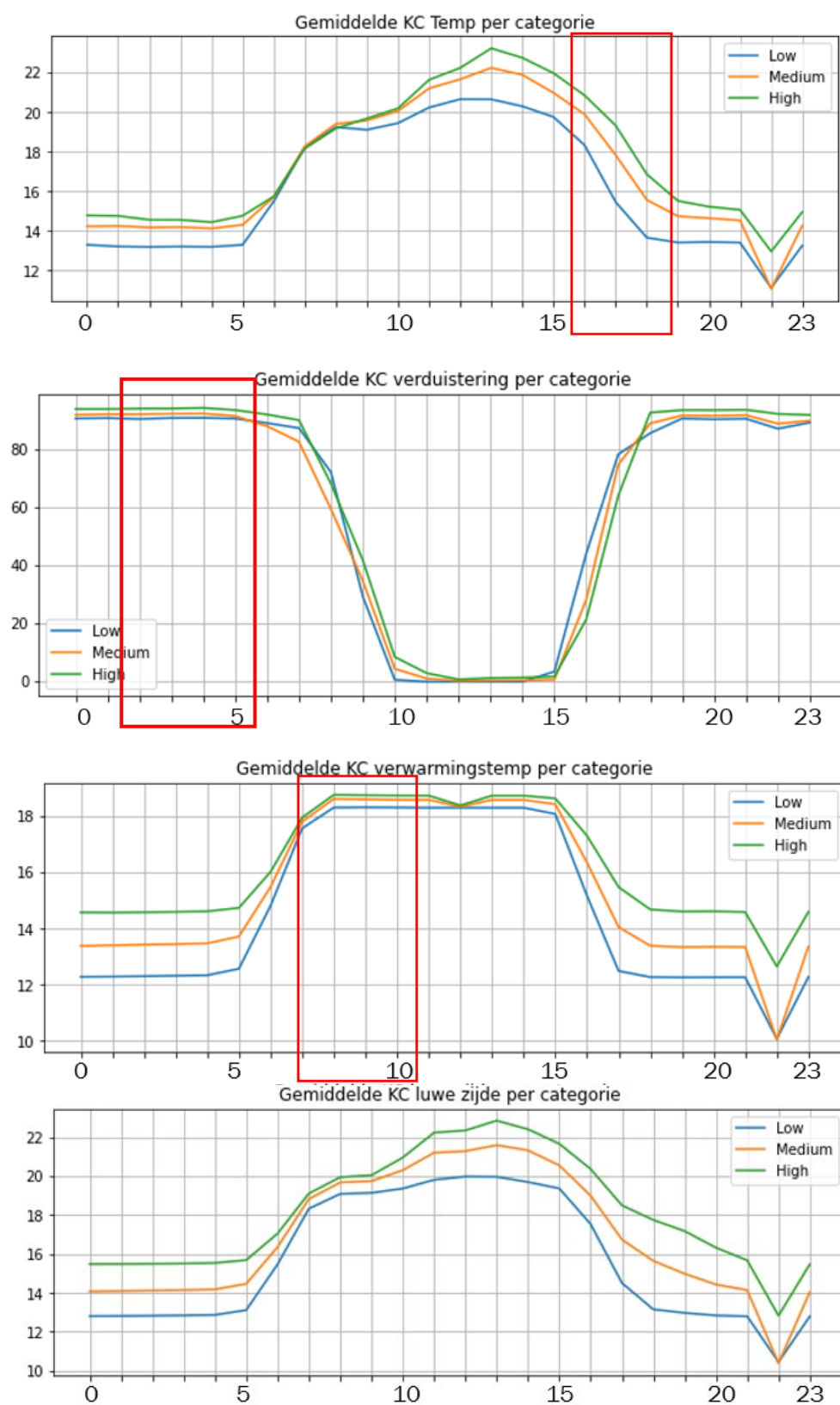
Opgemerkt wordt dat sterke correlaties niet impliceren dat er sprake is van een oorzakelijk verband. Er kan een ander onderliggend effect op beide variabelen zijn waardoor een samenhang ontstaat. Desondanks kan deze analyse nuttig zijn, omdat nieuwe inzichten kunnen ontstaan door naar diepere oorzaken achter de gevonden correlaties te zoeken. Een dergelijke diepere analyse valt buiten de scope van dit project.



Figuur 14. Sterkste correlaties van variabelen met de opbrengst. Blauw: positieve correlatie; rood: negatieve correlatie.

Er zijn beduidend meer variabelen die positief gecorreleerd zijn met de opbrengst dan negatief. De verduistering KV, de verwarmingstemperatuur KVT, de gemeten temperatuur KT en de raamopening aan de luwe zijde KLZ laten de sterkste positieve correlatie zien. De KV in het midden van de nacht (2:00 – 5:00) heeft een hoge correlatie met de opbrengst, echter is het niet duidelijk wat daarvan de oorzaak is, aangezien de verduistering geopend of gesloten is. De KVT heeft vooral 's morgens (8:00 – 11:00) een positieve correlatie met de opbrengst. Voor de KT geldt dat vooral de hoge temperaturen aan het einde van de middag een positieve correlatie hebben met de opbrengst. Er is geen duidelijk dagdeel te onderscheiden waarop de raamopening luwe zijde meer of juist minder gecorreleerd is aan de opbrengst. Zowel de nachtelijke uren (0:00 – 6:00) als de uren overdag (13:00 – 17:00) is deze sterk gecorreleerd. Er valt verder op dat de temperatuurwaardes (de enige meetwaarde die zowel uit het sensornetwerk als uit de klimaatcomputer komt) van de klimaatcomputer over het algemeen redelijk vergelijkbaar correleren met de opbrengst ten opzichte van de gemeten waarden van de sensoren.

In de analyse in Figuur 15 zijn deze vier variabelen van de klimaatcomputer (KT, KV, KVT, KLZ) weergegeven. De variabelen zijn verdeeld over de 3 categorieën opbrengsten (laag, gemiddeld, hoog). In rode rechthoeken zijn de tijden aangegeven wanneer de variabelen de hoogste correlaties laten zien.



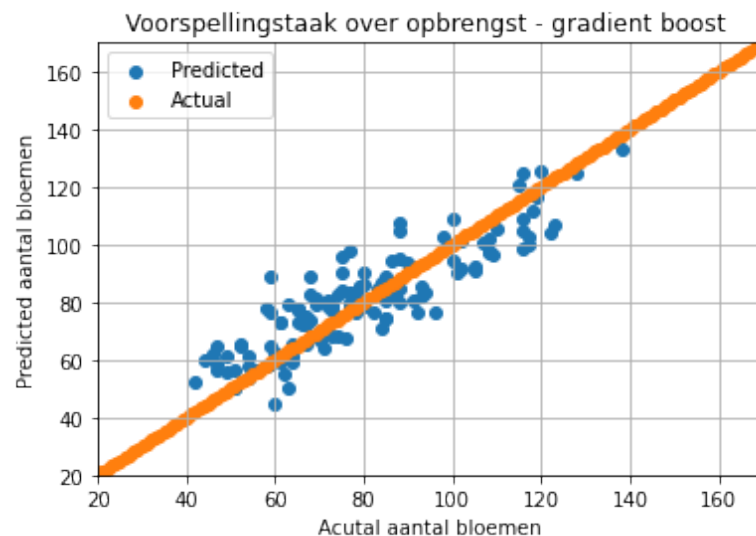
Figuur 15. Gemiddelde variatie over de dag van de variabelen die het sterkst correleren met de opbrengst, opgesplitst naar de bijbehorende opbrengstcategorie (laag – midden – hoog).

In de grafieken is zichtbaar dat de temperatuur het duidelijkst varieert met de opbrengst. De verduistering en de verwarmingstemperatuur variëren minder duidelijk met de opbrengst. Dat betekent dat een kleine verandering van deze variabelen te relateren is aan een duidelijke verandering van de opbrengst (bijv. van categorie 'laag' naar categorie 'hoog').

4.4 Fase 2: AI en voorspellende variabelen voor opbrengst

In de machine learning (ML) analyses zijn verschillende ML modellen vergeleken zoals de [random forest](#), [linear regression](#) en [Gradient Boost](#). Het gradient boost model laat de beste resultaten zien; daar is verder mee gewerkt. Het doel van het ML model is een uitkomst (in dit geval aantal geoogste bloemen) te voorspellen op basis van inputgegevens (in dit geval metingen door de sensoren en metingen of setpoints van de klimaatcomputer). Door het model te voeden met de werkelijke oogst, kan het ML model leren, de eigen voorspelling verbeteren, maar ook adviseren welke setpoints tot een optimale oogst leiden. Daarmee wordt toegewerkt naar meer autonoom telen. In dit project is de eerste stap gezet: het voorspellen van de oogst op basis van inputgegevens en vergelijken met de werkelijke oogst, om de betrouwbaarheid van het model te beoordelen.

De resultaten van de voorspellingen staan in Figuur 16. Op de X-as staat de werkelijke opbrengst (aantal bloemen). Op de Y-as staat de voorspelde opbrengst, op basis van de waarden van de variabelen. De blauwe punten zijn de voorspellingen van het ML-model. Als de voorspelling 100% zou overeenkomen met de werkelijkheid, zouden alle punten op de oranje lijn liggen.

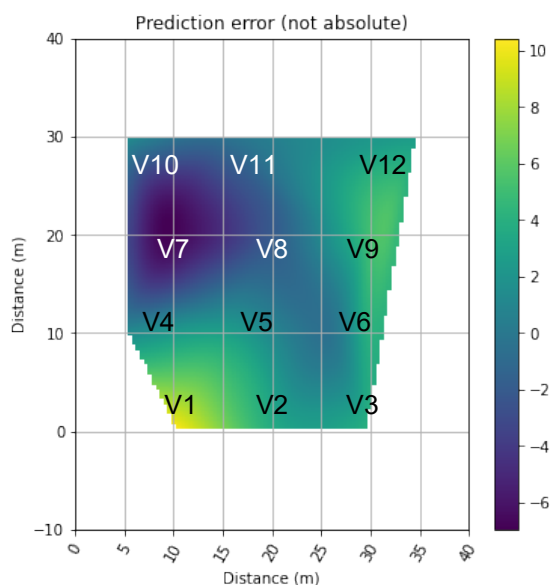


Figuur 16. Voorspelde opbrengst per veld (verticaal) op basis van de waarden van de variabelen ten opzichte van de werkelijke opbrengst.

Figuur 16 laat zien dat de voorspellingen in behoorlijke mate de oranje lijn volgen. Er zijn niet veel uitschieters naar de gebieden rechtsonder en linksboven. De Mean Absolute Error (MAE, de gemiddelde absolute afwijking) van deze voorspelling is 8,3. Met de crossvalidatie methode (analyses na het op andere manieren opsplitsen van de dataset) wordt gecontroleerd of de behaalde resultaten stabiel zijn. De

gemiddelde MAE wordt dan 9,4. Dit model geeft dus stabiele voorspellingen, wat het potentieel bruikbaar maakt in de praktijk. Daarvoor is het wel nodig dat de nauwkeurigheid van de voorspelling verder toeneemt. Dat zou kunnen door parameters aan het model toe te voegen die nu nog niet meegenomen zijn, zoals gewastemperatuur, water- en voedinggift en verdamping.

In de heatmap in Figuur 17 staat aangegeven waar in de kas de voorspellingen het verst afwijken van de werkelijke opbrengst. Het gebied rond veld 1 (linksonder) heeft de grootste positieve afwijking, gevolgd door velden 9 en 12. Hier is de opbrengst dus hoger voorspeld dan die daadwerkelijk is. Voor het gebied rond veld 7 wordt een te lage opbrengst voorspeld in relatie tot de werkelijkheid.



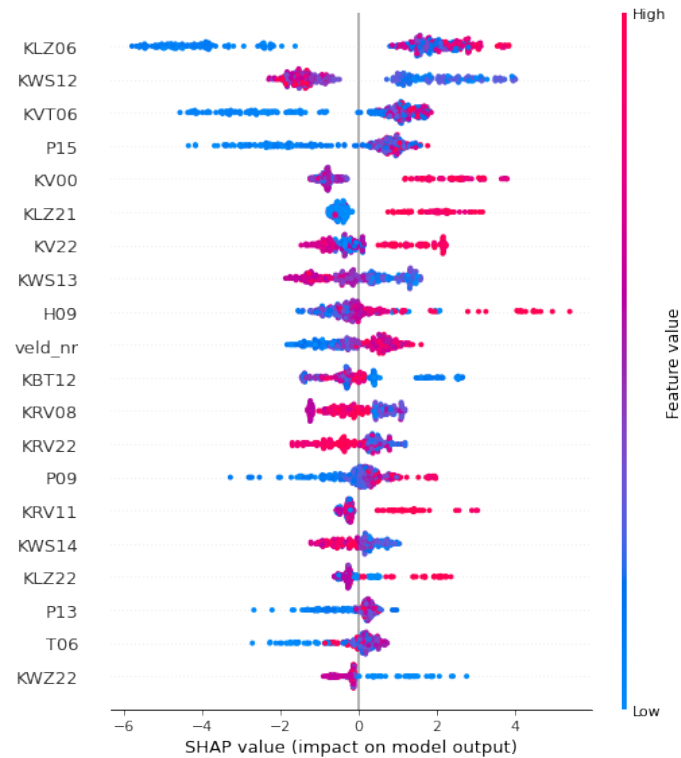
Figuur 17. Heatmap waarin zichtbaar is waar in het onderzochte gebied de voorspelde opbrengsten te hoog (geel) en te laag (blauw) zijn ten opzichte van de werkelijkheid. Velden 1, 2 en 3 liggen onderaan van links naar rechts; veld 12 is rechtsboven.

Deze uitkomsten laten zien dat toepassing van het algoritme met de onderzochte variabelen in een aantal velden tot een andere opbrengstvoorspelling leidt dan de werkelijkheid. Dat is te verklaren uit het gegeven dat meer variabelen van invloed zijn, die niet in dit onderzoek zijn meegenomen. Denk daarbij aan gewasvariabelen, effecten van straling op de sensoren (zie het eind van paragraaf 4.2) en invloed van menselijk handelen voor of tijdens het oogsten.

Vervolgens is geanalyseerd welke variabelen een sterke correlatie hebben met de voorspelde opbrengst.

In Figuur 18 staat de hoogte van de variabele gekleurd van roze (hoog) naar blauw (laag). De variabelen zijn weergegeven op volgorde van de sterkte van de correlatie met de opbrengst. De variabele bovenaan heeft dus de sterkste correlatie met de voorspelde opbrengst. De eerste 20 variabelen van de 312 zijn weergegeven. Op de x-as is weergegeven wat de relatieve samenhang is met de opbrengst. Een negatieve samenhang betekent dat de opbrengst negatief beïnvloed wordt. Een positieve samenhang betekent dat de opbrengst positief beïnvloed wordt en dus omhoog gaat. Twee voorbeelden:

- Nr 1: KLZ06 (klimaatcomputer raamopening luwe zijde tussen 6:00 – 7:00) correleert sterk met de opbrengst bij lage waarden en bijna even sterk bij hogere waarden.
- Nr 4: P15 (PAR tussen 15:00 – 16:00) correleert sterk met de opbrengst als de PAR laag is (blauw), en minder sterk als de PAR hoog is (roze).



Figuur 18. Variabelen met de sterkste correlatie met de opbrengst, bovenaan is het sterkst. Blauw geeft een lage waarde van de variabele weer; rood een hoge waarde. Zichtbaar is bij welke waarden de variabelen goede voorspellers zijn voor de opbrengst. Bijvoorbeeld bij een lage KLZ21 (raamopening luwe zijde tussen 21 en 22u) is deze nauwelijks gecorreleerd met de opbrengst, maar bij een hoge KLZ21 wel.

Er zijn veel variabelen die correleren met de opbrengst, variërend over verschillende tijden op de dag. Wat de klimaatcomputer betreft zijn vooral lagere raamopening luwe zijde (KLZ), hogere windsnelheden (KWS), verduistering (KV) en lagere verwarmingstemperatuur (KVT) voorspellers voor een lagere opbrengst. Vanuit de sensoren gemeten zijn de PAR in de middag en relatieve luchtvochtigheid (H) in de ochtend goede voorspellers. Daarbij correleren voor PAR vooral de lage waarden sterk met opbrengst (lage PAR ↔ flink lagere opbrengst); de positieve waarden correleren minder sterk positief met opbrengst (hoge PAR ↔ iets hogere opbrengst). Bij relatieve luchtvochtigheid correleren juist hogere waarden sterk met opbrengst (hogere H ↔ flink hogere opbrengst), terwijl de lagere waarden slechts zwak correleren met opbrengst (lagere H ↔ weinig lagere opbrengst).

Er zijn twee belangrijke aandachtspunten m.b.t. de statistiek:

- Het ML-model haalt er één parameter uit die een tijdsgroep vertegenwoordigt. Bijvoorbeeld: KLZ06 betekent dat er in de (vroeg) ochtend een sterke correlatie is met de raamopening aan de luwe zijde.
- De SHAP-analyse houdt geen rekening met oorzaak-gevolg relaties, alleen met samenhang tussen verschillende variabelen. Het ligt in redelijkheid weliswaar

voor de hand dat er een aantal oorzaak-gevolg-relaties zijn, maar de analyses doen daar geen uitspraak over.

De statistische, correlatie- en machine learning resultaten laten duidelijke correlaties zien tussen de waarden van een aantal variabelen en de opbrengst.

4.5 Bespreking van de resultaten

Het onderzoek is uitgevoerd in de kas van Kwekerij Van Veen Gerbera uit De Lier. De resultaten van de analyses zijn niet zonder meer te generaliseren naar andere (gerbera)kassen. Wel bieden de resultaten handvatten voor kwekers om mogelijkheden te onderzoeken voor verbetering van de beheersing en de homogeniteit van het kasklimaat.

In dit project zijn buiten- en binnenklimaatgegevens en setpoints van de klimaatcomputer meegenomen in de analyses, evenals de opbrengst (aantal geoogste bloemen). Op deze gegevens zijn analyses uitgevoerd om correlaties te onderzoeken en de mogelijkheid te onderzoeken om de opbrengst te voorspellen op basis van de beschikbare gegevens. In dit onderzoek zijn – behalve opbrengst – geen andere variabelen gemeten, zoals voedingsgift, irrigatiemomenten, ontwikkelsnelheid, planttemperatuur, ontwikkeling van schadelijke en gunstige mijten en insecten en gedrag van stomata. De reden daarvoor is dat het onderzoek zich in eerste instantie richt op gedetailleerder inzicht in de variatie van klimaatvariabelen en hoe deze variatie samenhangt met de productie. Plantgegevens zijn minder gemakkelijk als inputgegevens te behandelen, maar voedingsgift, irrigatie en de ontwikkeling van schadelijke en gunstige mijten en insecten zouden – bijvoorbeeld in een vervolgonderzoek – aan deze inputgegevens toegevoegd kunnen worden om zo een nog betere voorspelling te krijgen door het AI-model.

Uit de analyse van de gegevens van de 2^e fase bleek dat er vooral aan de zijkanten van de kas minder opbrengst werd gemeten ten opzichte van de velden het dichtst bij het hoofdpad. Dit fenomeen moet verband houden met PAR. Er zijn schaduweffecten aan de zijkanten van de kas, wat PAR vermindert. De reden is dat dit deel van de kas grenst aan de serviceruimte. De wand van deze ruimte bestaat uit panelen in plaats van een transparant materiaal zoals glas.

Het percentage afgekeurde bloemen is hoger in velden 4, 7 en 10 en lager in velden 3, 6 en 12. Ook hebben de velden met de hoogste afkeur een hogere opbrengst. Dit zou verklaard kunnen worden als een snellere groei gepaard gaat met het vaker vóórkomen van bloemafstoting. In dat geval kan mogelijk meer duidelijkheid verschaft worden door analyse van andere parameters, voornamelijk gerelateerd aan gewassturing via voeding en irrigatiestrategie.

Uit de analyse van de gegevens blijkt dat de gegevens (metingen) van de klimaatcomputer een betere correlatie laten zien met de opbrengst dan de correlatie die blijkt uit de waarden van de draadloze sensoren. Dit is over het algemeen te verklaren uit het feit dat de sensoren die de klimaatcomputer gebruikt nauwkeuriger zijn dan de draadloze sensoren. De meteo-box in de kas met een natte en droge bol wordt constant geventileerd, dit in tegenstelling tot de draadloze sensoren en de PAR-sensor. Dit betekent niet dat het gebruik van draadloze sensoren afgeraden

moet worden; de hogere sensordichtheid leidt tot waardevolle informatie over de homogeniteit en correlaties van het klimaat met de opbrengst. Wel dienen het trekken van conclusies en in het bijzonder het nemen van beslissingen over de beheersing van het kasklimaat met de nodige voorzichtigheid te gebeuren. Daarnaast is het belangrijk dat sensoren regelmatig worden gekalibreerd om afwijkingen in meetwaarden te voorkomen.

De positieve correlatie van temperatuur en PAR met de opbrengst is conform de verwachting, maar relevant is dat dit het sterkst is tussen 11:00 en 16:00 uur. Daarbij wordt opgemerkt dat dit resultaat gebaseerd is op de metingen in de proefperiode, van oktober tot maart, waarin de daglichtperiode kort is en veel gebruik gemaakt wordt van assimilatiebelichting. In de zomerperiode kan deze relatie anders zijn.

Opvallend is daarnaast de positieve correlatie van de luchttemperatuur tegen het einde van de middag met de opbrengst.

De positieve correlatie van de verduisteringsperiode met de opbrengst is interessant, omdat deze verband houdt met het beheer van straling over het gewas, maar ook met de uitwisseling van warmte door thermische straling. De rol van verduisteringsschermen kan verder worden onderzocht door het gebruik van een camera om de temperatuur van het gewas te meten en door het gebruik van een netto radiometer.

De statistische en AI-methoden die in de huidige studie worden gebruikt zijn bevredigend, aangezien ze op duidelijke wijze de correlatie tussen de verschillende variabelen en de opbrengst beschrijven. De verwachting is dat het toepassen van deze methoden, in het bijzonder de AI (Gradient Boost) methode, tot betere opbrengstvoorspellingen leiden naarmate enkele belangrijke parameters worden toegevoegd, zoals gewastemperatuur, insectengevoeligheid en voeding- en watergift.

5 Conclusies en aanbevelingen

5.1 Conclusies

De conclusies uit dit onderzoek, aansluitend bij de doelstellingen in paragraaf 1.2, zijn de volgende:

- A. Zoals verwacht is er een aanzienlijke ruimtelijke variatie van klimaatparameters binnen de kas in termen van temperatuur, relatieve luchtvochtigheid en PAR (Photosynthetic Active Radiation). Deze ruimtelijke variatie van het klimaat, zoals blijkt uit de metingen van het draadloze sensornetwerk, is groter dan die gemeten door de geventileerde meetboxen van de klimaatcomputer. De statistische analyse van deze gegevens leidt tot globaal inzicht in de relatie met de opbrengst (aantal geoogste bloemen).
- B. Op basis van de verschillende analyses zouden in de donkere periode (oktober – maart) de volgende acties een positief effect kunnen hebben op de opbrengst:
 - Vermijden van lage verwarmingstemperaturen, zowel overdag als 's nachts, waarbij de normale dag-nacht schommelingen van de kastemperatuur blijven bestaan;
 - Vermijden van lage PAR in de tweede helft van de middag en de vroege avond, dus meer assimilatiebelichting, zodat meer fotosynthese kan plaatsvinden.
- C. Het is mogelijk gebleken om deze gegevens in te zetten in een AI-algoritme, waarmee een voorspelling gedaan kan worden van de toekomstige opbrengst. Hiermee kunnen complexe relaties tussen variabelen gerelateerd worden aan de opbrengst. Door het algoritme te laten leren (o.a. door het te voeden met de werkelijke opbrengst), kan het een redelijke opbrengstvoorspelling doen.

5.2 Aanbevelingen

Algemeen wordt opgemerkt dat de aanbevelingen voortkomen uit het onderzoek in de kas van Kwekerij Van Veen Gerbera uit De Lier en niet zonder meer te generaliseren zijn naar andere (gerbera)kassen. Wel bieden de resultaten handvatten aan kwekers voor onderzoek naar verbetering van het kasklimaat.

De periode waaruit gegevens gebruikt worden om te relateren aan de opbrengst kan nauwkeuriger worden afgestemd op het proces van bloei-inductie. Dat zal leiden tot een analyseperiode van meer dan de nu gehanteerde vier dagen tot aan de oogst.

Het is een meerwaarde om een camera (voor temperatuurmeting van het gewas) en een netto radiometer (voor inkomende en uitgaande straling) in te zetten op zowel blad als bloem. Daarmee kan beter inzicht verkregen worden in klimaatongelijkheid en in mogelijkheden voor een AI-algoritme. Daarmee kan bovendien de relatie tussen verduistering en opbrengst nader worden onderzocht.

Het verdient aanbeveling om meetwaarden van de irrigatiemomenten, van het drainpercentage en van de gift en het verbruik van nutriënten mee te nemen, zodat het AI-algoritme betere voorspellingen kan doen. Hetzelfde geldt voor ziektedruk en

de kwaliteit van de bloem (bijvoorbeeld via de verhouding afkeur/goedkeur).

Dit onderzoek heeft uitgewezen dat een dichter netwerk van sensoren de voorkeur verdient boven een ruimer netwerk. De sensoren zouden nóg dichter bij elkaar kunnen worden geplaatst om de sterkste correlaties te vinden tussen gewasprestaties en klimatologische omstandigheden. Dan kan ook nauwkeuriger onderzocht worden wat de beïnvloedende factoren zijn voor afkeur.

Bij de gegevensanalyse is het waardevol om perioden met verschillende externe klimatologische omstandigheden (zoals zomer / winter) mee te nemen. Daarmee worden de conclusies die zijn gerelateerd aan de opbrengst robuuster.

De data-analyse kan versterkt worden door de verschillende variabelen te groeperen in twee categorieën: variabelen die direct samenhangen met de teelt en dus de opbrengst en variabelen die primair het kasklimaat beïnvloeden.

Om een kwantitatieve indruk te geven van de homogeniteit van het kasklimaat kunnen in toekomstige analyses de relaties tussen setpoints, externe klimatologische omstandigheden en de variaties daarin worden onderzocht. Dat kan leiden tot een 'klimaat uniformiteitsindex' van de kasomgeving.

De homogeniteit van het kasklimaat kan gekwantificeerd worden door een 'klimaat uniformiteitsindex' te definiëren, die zich baseert op de verschillen tussen posities in de kas en de variaties over de tijd. Daarvoor is van belang dat de kas voldoende meetpunten (sensoren) bevat. Onderzocht kan dan worden hoe de uniformiteit samenhangt met de controlstrategie en de setpoints, en wat daarop de invloed van het weer en de seizoenen is.

Uit de analyse van de data bleek dat het waardevol kan zijn om de data afkomstig van de klimaatcomputer, het draadloze sensornetwerk en het productie-oogstregistratiesysteem te integreren. Door deze integratie zou de statistische analyse continu kunnen worden uitgevoerd, in plaats van na een bepaalde periode. Daarvoor is het nodig om tools te ontwikkelen voor data-integratie, data-analyse met behulp van AI-technieken en voor het trekken van conclusies. Dit kan de teler helpen zijn teeltpraktijk aan te passen, of zelfs de kasklimaatstrategie te verbeteren.

6 Ondertekening

Delft, 1 februari 2022

TNO

Ir. ing. M. Steins
Research Manager

Dr. A. Sapounas
Auteur