



Methaan Data Lake: Analyse en koppelen van externe data bronnen om trends en variatie van methaanemissie tussen melkveebedrijven en meetmethodes

Resultaten project F3 van de klimaat envelop 2020/2021

Esther D. Ellen, Carsten A. Schep, Michael Aldridge, Bert Klandermans, Claudia Kamphuis

Openbaar
Rapport 1401



WAGENINGEN
UNIVERSITY & RESEARCH

Methaan Data Lake: Analyse en koppelen van externe data bronnen om trends en variatie van methaanemissie tussen melkveebedrijven en meetmethodes

Resultaten project F3 van de klimaat envelop 2020/2021

Esther D. Ellen, Carsten A. Schep, Michael Aldridge, Bert Klandermans, Claudia Kamphuis

Wageningen Livestock Research

Dit onderzoek is uitgevoerd door Wageningen Livestock Research en gesubsidieerd door het ministerie van Landbouw, Natuur en Voedselkwaliteit, in het kader van het Beleidsondersteunend onderzoek thema 'Klimaat slim methaan emissie veehouderij' (projectnummer BO-53-003-047).

Wageningen Livestock Research
Wageningen, december 2022

Openbaar
Rapport 1401

Ellen, E.D., Schep, C.A., Aldridge, M., Klandermans, B., en Kamphuis, C., 2022. *Methaan Data Lake: Analyse en koppelen van externe data bronnen om trends en variatie van methaanemissie tussen melkveebedrijven en meetmethodes inzichtelijk te krijgen; Resultaten project F3 van de klimaat envelop 2020/2021*. Wageningen Livestock Research, Rapport 1401.

Samenvatting NL Reduceren van methaanemissie is complex en vereist inzicht in beschikbare data, en het begrijpen van variatie tussen bedrijven. Het doel van dit onderzoek was om een data architectuur te bouwen om verschillende methaan-datastromen bijeen te brengen, te koppelen, en daar externe databronnen aan toe te voegen om 1) trends en variatie tussen bedrijven zichtbaar te maken, 2) de invloed van (beschikbare) externe factoren op methaanemissie vast te stellen, en 3) bestaande rekenregels die methaan voorspellen te toetsen met wat er op bedrijven aan emissies gemeten wordt. In dit onderzoek hebben we laten zien dat het mogelijk is om verschillende datastromen en externe databronnen samen te brengen in de Cloud, in een Data Lake, en om in de Cloud analyses uit te voeren. Hiermee heeft dit onderzoek de basis gelegd van een modern methaan datawarehouse. Methaandata van drie verschillende sensoren (GreenFeed, Sniffer en Stal), gemeten op verschillende bedrijven in Nederland, hebben we samengebracht in dit Methaan Data Lake. Een vergelijking van de drie sensoren binnen een bedrijf liet zien dat er verschillen zijn in niveaus en patronen tussen de metingen van de drie sensoren. Er is geen directe verklaring van patronen uit externe meteorologische gegevens gevonden. Tot slot zijn de methaanmetingen vergeleken met voorspelde methaanemissie, waarbij we gebruik hebben gemaakt van een bestaand model, de kringloopwijzer (KLW). Voor korteretermijnvoorspelling is het KLW-model niet geschikt, echter voor langeretermijnvoorspellingen geeft KLW een goede schatting van de enterische methaanemissie.

Summary UK Reducing methane emissions is complex and requires insight into available data, as well as understanding variation between farms. The goal of this research was to build a data architecture to bring together, and merge, different methane data streams and to add external data sources, to 1) visualize trends and variation between farms, 2) establish the influence of (available) external factors on methane emissions, and 3) validate existing models predicting methane emissions with actual farm measurements. In this research, we have shown that it is possible to merge different data streams and external data sources in the Cloud, in a Data Lake, and to perform analyses in the Cloud. This research has thus laid the foundation for a modern Methane data warehouse. In this Methane Data Lake, we have brought together methane data from three different sensors (GreenFeed, Sniffer and Farm), recorded on different farms in the Netherlands. Comparison of the three sensors within a farm showed variation in the level and pattern of the methane measurements. No direct explanation of patterns from external meteorological data has been found. Lastly, we compared measured methane data to predicted methane emissions, which we obtained from an existing methane emission prediction model called Kringloopwijzer (KLW). Results showed that the KLW-model is less suitable for short-term predictions. However, for longer-term predictions (yearly basis) the KLW-model gives a good estimate of the enteric methane emission.

Dit rapport is gratis te downloaden op <https://doi.org/10.18174/583514> of op www.wur.nl/livestock-research (onder Wageningen Livestock Research publicaties).



Dit werk valt onder een Creative Commons Naamsvermelding-Niet Commercieel 4.0 Internationaal-licentie.

© Wageningen Livestock Research, onderdeel van Stichting Wageningen Research, 2022

De gebruiker mag het werk kopiëren, verspreiden en doorgeven en afgeleide werken maken. Materiaal van derden waarvan in het werk gebruik is gemaakt en waarop intellectuele eigendomsrechten berusten, mogen niet zonder voorafgaande toestemming van derden gebruikt worden. De gebruiker dient bij het werk de door de maker of de licentiegever aangegeven naam te vermelden, maar niet zodanig dat de indruk gewekt wordt dat zij daarmee instemmen met het werk van de gebruiker of het gebruik van het werk. De gebruiker mag het werk niet voor commerciële doeleinden gebruiken.

Wageningen Livestock Research aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Wageningen Livestock Research is NEN-EN-ISO 9001:2015 gecertificeerd.

Op al onze onderzoeksopdrachten zijn de Algemene Voorwaarden van de Animal Sciences Group van toepassing. Deze zijn gedeponeerd bij de Arrondissementsrechtbank Zwolle.

Openbaar Wageningen Livestock Research Rapport 1401.

Inhoud

Samenvatting	5
1 Inleiding	7
1.1 Aanleiding voor het onderzoek	7
1.2 De programmatische integrale aanpak	7
1.3 Doel van het onderzoek	8
2 Inleiding	10
2.1 Afstemming met andere projecten	10
3 Overzicht van methaanmetingen	11
3.1 Methaansensoren	11
3.1.1 GreenFeed	11
3.1.2 Sniffer	11
3.1.3 Stal	12
3.2 Dataverzameling	12
4 Data Lake	13
4.1 Data Lake	13
4.2 Materiaal en methode	14
4.2.1 GreenFeed	15
4.2.2 Sniffer	16
4.2.3 Stalmetingen	16
4.2.4 Akkerweb	16
4.2.5 Data Lake structuur en analyse	16
4.3 Resultaten en discussie	17
5 Data Lake	18
5.1 Materiaal en methode	18
5.1.1 Aanwezige variatie	18
5.1.2 Analyse van variatie en trends	20
5.1.3 Variatiecoëfficiënt	20
5.1.4 Vergelijking tussen bedrijven	21
5.1.5 Vergelijking tussen meetmethodes	22
5.2 Resultaten en discussie	22
5.2.1 Algemeen meetmethodes	22
5.2.2 Variatie	26
5.2.3 Tussen meetmethodes	28
5.2.4 Rangschikken van koeien en bedrijven	30
6 Koppelen met externe bronnen	32
6.1 Materiaal en methode	32
6.2 Resultaten en discussie	32
6.2.1 Jaarpatroon	32
6.2.2 Correlatie per bedrijf	34
7 Valideren bestaande modellen	35
7.1 Materiaal en methode	35
7.1.1 Modelleren met de KringloopWijzer	35

7.1.2	Vergelijking tussen de voorspelde emissies van de KringloopWijzer en daadwerkelijke methaanmetingen	36
7.2	Resultaten en discussie	37
7.2.1	Vergelijking tussen de KringloopWijzer en Stalmetingen	37
7.2.2	Vergelijking tussen de KringloopWijzer en GreenFeed	39
7.2.3	Vergelijking tussen de KringloopWijzer en Sniffers	40
7.2.4	Hoe kunnen de resultaten van de vergelijking tussen voorspeld methaan van de KringloopWijzer en gemeten methaan van de drie meetmethodes worden gebruikt in de praktijk?	41
	Literatuur	44

Samenvatting

Het reduceren van methaanemissie is een belangrijk onderdeel van de programmatische aanpak 'Integraal Aanpakken', als onderdeel van het LNV-Klimaatbeleid, gefinancierd door het Ministerie van LNV. Het overkoepelende onderzoeksdoel binnen deze programmatische aanpak is een reductie van 30% van methaan (**CH₄**) en ammoniak (**NH₃**), en een implementatiegraad van 80% in 2030. Het LNV onderzoeksprogramma Stal en Opslag heeft als doel om innovatieve technieken en maatregelen in stal, mest en verwerking voor emissiereductie te ontwikkelen en in de praktijk te implementeren om de reductieopgaven in 2030 te halen. Het reduceren van CH₄-emissie is echter complex en vereist inzicht in de beschikbare data, en het begrijpen van het ontstaan van variatie tussen bedrijven.

Het doel van dit onderzoek was om een data architectuur te bouwen om verschillende CH₄-datastromen bijeen te brengen, deze te koppelen, en daar externe databronnen aan toe te voegen om 1) trends en variatie tussen bedrijven zichtbaar te maken, 2) de invloed van (beschikbare) externe factoren op CH₄-emissie vast te stellen, en 3) bestaande rekenregels die CH₄ voorspellen te toetsen met wat er op bedrijven daadwerkelijk gemeten wordt.

In dit onderzoek hebben we laten zien dat het mogelijk is om verschillende datastromen samen te brengen in een Cloud-based Data Lake. Een Data Lake is een modern datawarehouse concept voor integratie, opslag en analyse van data. Werken met een Data Lake geeft een grote flexibiliteit in de dataverwerking en -analyse en er kan tegen relatief weinig kosten een grote dataset (schaalbaar tot enkele petabytes) in worden opgeslagen. Methaandata van drie verschillende meetmethodes (GreenFeed, Sniffer en Stal) gemeten op verschillende bedrijven in Nederland zijn samengebracht in het Methaan Data Lake. Daarnaast is er een koppeling gemaakt tussen dit Methaan Data Lake en een externe databron (Farmmaps) om de lokale meteorologische gegevens te verkrijgen. Hiervoor zijn verschillende Python-pipelines en Application Programming Interfaces (**APIs**) ontwikkeld. De meeste van deze pipelines zijn in dit onderzoek nog niet volledig geautomatiseerd, maar vergen nog een handmatige aanvraag. De ontwikkelde pipelines kunnen echter relatief eenvoudig geautomatiseerd worden. De scripts voor de data-analyse die ontwikkeld zijn in dit onderzoek zullen beschikbaar blijven zodat de analyses frequenter en herhaaldelijk uitgevoerd kunnen worden als meer data in het Methaan Data Lake geïntegreerd wordt.

In totaal zijn er op 27 Nederlandse melkveebedrijven CH₄-metingen uitgevoerd. Op twee bedrijven waren metingen verricht met alle drie de meetmethodes (GreenFeed, Sniffer en Stal), vijf bedrijven hadden GreenFeed- en Stalmetingen, vijf bedrijven hadden GreenFeed- en Sniffer-metingen, en negen bedrijven hadden Sniffer- en Stalmetingen. De data zijn tussen september 2018 en april 2021 verzameld om specifieke vraagstukken in andere onderzoeksprojecten te beantwoorden, en zijn in het Methaan Data Lake bijeengebracht. Dit leverde een wereldwijd unieke dataset op. Uit de data-analyse om trends en variatie te achterhalen blijkt dat de variatie tussen dieren op één bedrijf gemiddeld genomen nog groter is dan de variatie tussen bedrijven. De variatiecoëfficiënt (**VC**) is voor Sniffer 1.54 en 0.58, voor dieren en bedrijf respectievelijk; deze VC is voor GreenFeed 0.2 en 0.11, respectievelijk. Deze tussen-bedrijfsvariatie is vergelijkbaar met de variatie tussen dagen. Ook de variatie binnen dagen is groter dan de variatie tussen bedrijven, maar goed zichtbaar is het dagpatroon: aan het begin van de dag zien we lagere waarden voor CH₄ en CO₂ emissies of concentraties en in de avond zien we de hoogste waarden voor emissies of concentraties. In de variatie tussen dagen is een lange-termijn effect zichtbaar wat waarschijnlijk gerelateerd is aan de samenstelling van het rantsoen door het jaar heen en de hoeveelheid mest in de mestkelder (onder de roostervloeren). Stal- en Sniffermetingen resulteren in een hogere schatting van de CH₄-emissie (Stal) en -concentratie (Sniffer) gedurende de winter en een lagere schatting van de emissie en concentratie gedurende de zomer. Er is geen directe verklaring van meetpatronen uit de gekoppelde meteorologische gegevens gevonden. Variatie tussen bedrijven blijkt nog steeds het meest lastig om te verklaren. De drie meetmethodes meten CH₄ op verschillende plekken binnen een bedrijf en op verschillende manieren. Uit dit onderzoek blijkt dat de drie verschillende meetmethodes niet volledig substitueerbaar zijn als het gaat om het rangschikken van koeien (binnen een bedrijf) of van bedrijven op basis van de hoeveelheid CH₄-emissies.

Als laatste zijn de CH₄-metingen vergeleken met voorspelde CH₄-emissies, waarbij gebruik is gemaakt van het bestaande model Kringloopwijzer (**KLW**). Voor kortere termijnvoorspelling is het KLW-model niet geschikt, echter voor langere termijnvoorspellingen geeft KLW een goede schatting van de CH₄-emissie.

In dit onderzoek waren communicatie en samenwerking met onderzoekers die betrokken zijn bij andere onderzoeken die CH₄-data verzamelen belangrijke sleutelcomponenten voor succes. Naar de toekomst toe blijft het belangrijk om te communiceren over de mogelijkheden van het Methaan Data Lake en de infrastructuur van een Data Lake verder toe te passen. Door verschillende data stromen samen te voegen zal er meer inzicht verkregen worden in variaties tussen bedrijven en onderliggende factoren. In dit onderzoek is de focus gelegd op het bijeen brengen van de CH₄-data, die gemeten wordt door de drie verschillende meetmethodes voor specifieke onderzoeksdoeleinden. Daarnaast is er een focus gelegd op het binnenhalen van meteorologische gegevens van een externe databron. Het huidige Methaan Data Lake is dus nog relatief klein, en gevuld met een gelimiteerde hoeveelheid data. Hierdoor kon er in dit onderzoek nog weinig verklarend onderzoek verricht worden voor bijvoorbeeld de geobserveerde trends en variatie. Echter, dit onderzoek heeft de basis gelegd van een moderne methaan datawarehouse, waarbij CH₄-data centraal opgeslagen en bewerkt wordt. Door deze architectuur nu aan te leggen en de Methaan Data Lake vol te laten stromen met data die nu voor specifieke vraagstukken verzameld worden, kunnen reductiemaatregelen in de toekomst efficiënter en integraal ontwikkeld worden met de kennis en data die dan tot de beschikking staat.

1 Inleiding

1.1 Aanleiding voor het onderzoek

De Nederlandse melkveehouderij is een belangrijke sector met ongeveer 15.000 bedrijven en veel indirecte werkgelegenheid bij dienstverleners en in de verwerkende en toeleverende industrie. Daarnaast heeft de sector een belangrijke rol als beheerder van de groene ruimte. Het is daarom belangrijk om de Nederlandse melkveehouderij en het huidige productievolume te behouden. Echter, de melkveehouderij heeft een belangrijke bijdrage aan de uitstoot van broeikasgassen, waaronder de emissies van methaan (**CH₄**) en ammoniak (**NH₃**). In Nederland was de totale uitstoot van broeikasgassen 181,1 megaton koolstofdioxide-equivalenten in 2019 (RIVM, 2021). Ongeveer 10% hiervan is CH₄-emissie. Van de totale CH₄-emissie is ongeveer 70% afkomstig uit de agrarische sector. Daarvan is ongeveer 60% afkomstig uit de pens- en darmfermentatie van melkkoeien (Ruysenaars et al., 2020).

Het verminderen van CH₄-emissie is een belangrijk onderdeel geworden van een duurzame melkveehouderij. In het Nederlandse klimaatakkoord is vastgelegd dat de melkveehouderij voor 2030 de CH₄-emissie met 0.8 megaton CO₂-equivalenten moet reduceren ten opzichte van 2017, en dat deze reductie nog verder ingezet moet worden richting 2050. Daarnaast moet ook de NH₃-emissie in 2030 gereduceerd worden met 30% ten opzichte van 1990. Reductie van CH₄- en NH₃-emissie is echter complex, waarbij het belangrijk is om de bestaande kaders en doelen op het gebied van onder andere milieu, koeien in de wei, biodiversiteit, dierenwelzijn en diergezondheid goed mee te nemen. Vanuit de Klimaat Envelop wordt er gestuurd op een geïntegreerde aanpak van CH₄ en NH₃ via het voer- en dierspoor. Met als doel om de reductiedoelen in 2030 te behalen. Dit project tracht hier een bijdrage aan te leveren door metingen van drie verschillende methaan metingen samen te voegen en verschillen binnen en tussen bedrijven te kwantificeren, waarbij ook naar de impact van meteorologische kenmerken is gekeken, en door bestaande voorspellende rekenregels (Kringloopwijzer) te valideren met gemeten waardes.

1.2 De programmatische integrale aanpak

Dit onderzoek maakt deel uit van de programmatische aanpak 'Integraal Aanpakken' als onderdeel van het LNV-Klimaatbeleid, gefinancierd door het ministerie van LNV. Het overkoepelende onderzoeksdoel binnen deze programmatische aanpak is een reductie van 30% van CH₄ en NH₃ en een implementatiegraad van 80% in 2030. Voor de periode na 2030 zal een hogere reductie- en implementatiegraad tot meer emissiereductie kunnen leiden (de aanpak is een 'no regret policy' voor toekomstige klimaat- en stikstofdoelen).

Het LNV onderzoeksprogramma 'Integraal Aanpakken' heeft als doel om innovatieve technieken en maatregelen in stal, mest en verwerking voor emissiereductie te ontwikkelen en in de praktijk te implementeren om de reductieopgaven in 2030 te halen. Het onderzoeksprogramma Stal en Opslag bestaat uit 4 onderdelen: Meten en monitoring, onderzoek, pilots en demo's, en communicatie. Uit onderzoek moet blijken welke reductietechnieken werken en hoeveel emissiereductie deze opleveren. Naast onderzoek gericht op nieuwe stallen wordt ook gekeken naar kosteneffectieve aanpassingen aan bestaande stallen. De effecten van de reductieopties zullen in de praktijk worden getoetst en gedemonstreerd op pilot- en demobedrijven. Om te komen tot implementatie van reductiemaatregelen wordt kennis over nieuwe en bestaande maatregelen verspreid onder de doelgroep van veehouders en erfbetreders (adviseurs). De aanpak levert handelingsperspectief voor veehouders om emissies te reduceren en zetten aan tot toepassing.

1.3 Doel van het onderzoek

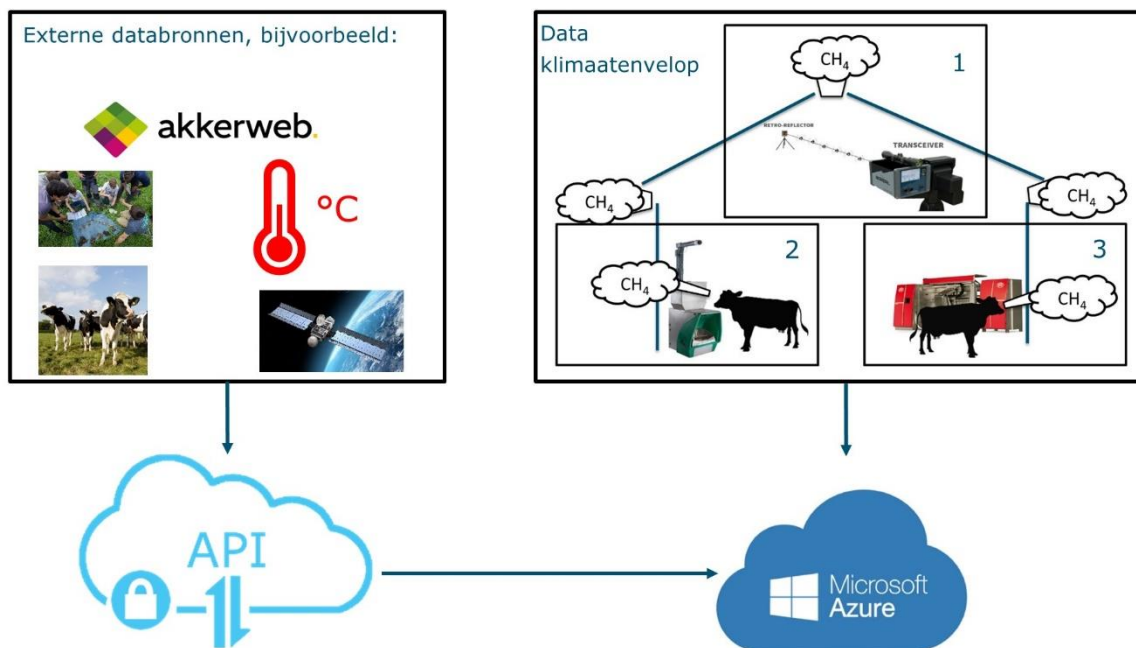
Het doel van dit onderzoek was om een moderne data architectuur te bouwen om verschillende CH₄-datastromen bijeen te brengen, te koppelen, en daar externe databronnen aan toe te voegen om 1) trends en variatie tussen bedrijven zichtbaar te maken, 2) de invloed van (beschikbare) externe factoren op CH₄-emissie vast te stellen, en 3) bestaande rekenregels die CH₄ voorspellen te toetsen met wat er op bedrijven daadwerkelijk gemeten wordt.

Sinds 2018 is CH₄-data verzameld op meerdere commerciële melkveebedrijven. Deze data worden gemeten op twee niveaus (individueel dier- en stalniveau) en middels verschillende meetmethodes (zie **Figuur 1.1**). Zo wordt CH₄-concentratie op stalniveau gemeten met een sensorsysteem en middels referentiemeetapparatuur om later gebruikt te worden voor het berekenen van CH₄-emissie op stalniveau. Daarnaast worden zogenoemde GreenFeeds en Sniffers ingezet om CH₄-data te meten op individueel dierniveau. In **Hoofdstuk 3** worden de drie verschillende meetmethodes verder toegelicht. Deze opzet, waarbij op hetzelfde moment met verschillende methodes en op diverse niveaus de CH₄-data wordt verzameld, is wereldwijd uniek. Ondanks deze unieke opzet, wordt er vooralsnog hoofdzakelijk eenzijdig gebruik gemaakt van de data, waarbij specifieke vragen worden beantwoord met slechts één component van deze dataset. Er kan meerwaarde uit deze dataset worden gehaald als de verschillende datastromen continu en langdurig aan elkaar gekoppeld zouden worden. Immers, door datastromen te koppelen, kunnen reductiemaatregelen sneller ontwikkeld worden met de kennis die dan tot de beschikking is, en hierdoor kunnen de effecten sneller bepaald worden.

Dit onderzoeksproject (F3) valt onder het onderdeel 'Onderzoek' van het onderzoeksprogramma Stal en Opslag, en richt zich op het verkrijgen van inzichten en het begrijpen van variatie van CH₄-emissies door gebruik te maken van beschikbare data. Om dit te doen, is in deze studie de basisarchitectuur aangelegd om CH₄-data van huidig onderzoek en van ouder onderzoek bijeen te brengen en te koppelen, en om toekomstige data eenvoudig toe te voegen. Daarnaast is de basisarchitectuur gebruikt om koppelingen te maken met externe databronnen (verzameld buiten het onderzoeksprogramma). Door deze verschillende CH₄-datastromen bijeen te brengen, te koppelen en daar externe databronnen aan toe te voegen (**Figuur 1.1**) is er de unieke kans om:

- (1) huidige trends en variatie tussen bedrijven zichtbaar te maken,
 - (2) de invloed van (beschikbare) externe factoren op CH₄-emissie vast te stellen,
- En
- (3) bestaande rekenregels die CH₄ voorspellen te toetsen met wat er op bedrijven gemeten wordt.

Inzicht in de variatie tussen bedrijven en over de tijd, en de invloed van externe factoren hierop, is belangrijk om effectieve reductiemaatregelen te ontwikkelen op dier- en stalniveau. Door de data architectuur te ontwikkelen en data integraal samen te brengen, kan in de toekomst deze data structuur ook gebruikt worden en kunnen analyses sneller en effectiever gebeuren. Immers, door datastromen bijeen te brengen, kunnen reductiemaatregelen sneller ontdekt en eventueel ontwikkeld worden met de kennis die dan tot de beschikking is. Hierdoor is het ook mogelijk om sneller effecten van de maatregelen te bepalen. Dit draagt bij aan een continue ontwikkeling van handelingsperspectief voor melkveehouders om ook in de toekomst die reductiemaatregel(en) te kiezen die het beste past bij de uitdaging en de bedrijfsvoering waar de melkveehouder op dat moment mee te maken heeft.



Figuur 1.1 Opslaan en combineren van verschillende methaan-datastromen (1: Stalmetingen, 2: GreenFeed, 3: Sniffer) verzameld op meerdere commerciële melkveebedrijven, en de koppeling daarvan met externe databronnen om trends en variaties van methaanemissie zichtbaar te maken tussen bedrijven en over de tijd.

In het huidige onderzoek zijn de gekoppelde CH₄-data gecombineerd geanalyseerd om trends en variatie in CH₄-emissie op stal- en dierniveau tussen de melkveebedrijven en de meetmethodes over de tijd en continu inzichtelijk te krijgen. Daarnaast zijn deze hoge frequentie aan meetwaardes gebruikt om bestaande rekenregels die CH₄ voorspellen te toetsen.

De unieke dataset is ook gebruikt om meetmethodes onderling te vergelijken. Dit heeft niet als doel gehad om de 'beste' meetmethode te achterhalen, maar om te kijken of de verschillende meetmethodes vergelijkbare trends opleveren en elkaar mogelijk kunnen versterken. Tenslotte zijn er externe databronnen toegevoegd om de invloed van deze externe factoren op CH₄-emissie vast te stellen. Omdat de infrastructuur de mogelijkheid biedt om direct data van deze externe factoren (in dit onderzoek lokale meteorologische gegevens) op te vragen, kunnen invloeden van bijvoorbeeld weersomslagen of extreme weersomstandigheden direct bekeken worden. De resultaten van deze analyses zijn op korte termijn al zinvol voor het onderzoek naar reductiemaatregelen, en daarmee voor alle veehouders. Daarnaast heeft de koppeling van de data meerwaarde voor de CH₄-reductie-vraagstukken van de toekomst.

2 Inleiding

Dit onderzoek borduurt voort op een data architectuur van een bestaand project (zie **Hoofdstuk 2.1** Afstemming met andere projecten), waarin de basis is gelegd om CH₄-data gemeten met de Sniffer (near real-time te verzamelen in de Cloud (Microsoft Azure) en waarbij een koppeling is gemaakt met temperatuur via een web-based API met Farmmaps. Farmmaps is een innovatief en advies platform dat verschillende databronnen die belangrijk zijn voor het telen van gewassen op één plaats heeft staan. Een beschikbare databron zijn meteorologische gegevens zoals temperatuur, luchtvochtigheid en windsnelheid. Deze gegevens zijn op basis van de coördinaten van de bedrijfslocatie en daarmee zeer lokaal. Dit is in tegenstelling tot de meteorologische gegevens die beschikbaar zijn vanuit het Koninklijke Nederlandse Meteorologisch Instituut (**KNMI**).

2.1 Afstemming met andere projecten

In dit onderzoek is er samengewerkt met andere lopende onderzoeken binnen de Klimaatenvolop Methaanemissie Veehouderij 2018-2020 die CH₄-data verzamelen op melkveebedrijven:

- Project 1 van KE2018: "Methaanemissie uit melkvee- en geitenstallen: monitoring en demonstratie emissies en bedrijfseffecten";
- Project 11 van KE2018: "Enterisch methaan: emissievariatie in de Nederlandse melkveestapel";
- Project 4 van KE2019: "Genetische mogelijkheden om enterische methaanemissie van melkvee te verlagen";
- Project E1 van KE2020: "Met sensoren en modellen tbv sturing op CH₄ en NH₃";

In het Kennis Basis project "Data Driven & High Tech" (KB DDHT – 04 Smart and privacy conserving infrastructures) is de basis gelegd voor een infrastructuur om CH₄-data gemeten met Sniffers (near) real-time te verzamelen, als ook de koppeling met Akkerweb voor temperatuurgegevens. Deze infrastructuur is in dit onderzoek als basis gebruikt en zal in **Hoofdstuk 4** verder toegelicht worden.

In dit onderzoek waren communicatie en samenwerking met onderzoekers die betrokken zijn bij andere onderzoeken die CH₄-data verzamelen essentiële sleutelcomponenten voor succes. Voor het gebruik van een nieuwe data-infrastructuur (Data Lake) moest er bewustwording van het werken met deze nieuwe infrastructuur gecreëerd worden. Voor die bewustwording zijn er het afgelopen jaar meerdere bijeenkomsten georganiseerd waarbij de infrastructuur is uitgelegd en uiteindelijke (toekomstige) impact van dit onderzoek besproken is. Het werken met een Methaan Data Lake, in de Cloud, geeft nieuwe uitdagingen, die in **Hoofdstuk 4** verder besproken worden.

3 Overzicht van methaanmetingen

Binnen het onderzoek werden CH₄-data bijeen gebracht die op drie verschillende manieren en op twee verschillende niveaus verzameld zijn; metingen met de GreenFeed en metingen met de Sniffer zijn beide metingen op dierniveau, en een derde meting op stalniveau (**Figuur 3.1**). De verschillende methodes zullen hieronder verder toegelicht worden. Het is belangrijk om in het achterhoofd te houden dat de meest precieze methode (de referentie of gouden standaard) het meten van CH₄ in een respiratiekamer is. Hierbij wordt één dier, in dit geval een melkkoe, voor een aantal uren in deze kamer geplaatst, waarbij de exacte uitstoot van CH₄ gemeten wordt. Het gebruik van een respiratiekamer heeft een aantal praktische beperkingen: ze zijn duur, hebben een lage doorvoer van dieren, zijn zeer arbeidsintensief, en het normale gedrag van een dier wordt tijdelijk, maar ernstig, beperkt (Hammond et al., 2015). Daarnaast zijn deze precieze metingen, onder zeer gecontroleerde omstandigheden, niet goed te vertalen naar een praktijksituatie. Immers, in de praktijk kunnen bekende en onbekende interacties wel een effect hebben op de CH₄-productie/emissie, maar deze worden door de gecontroleerde omstandigheden niet zichtbaar in een respiratiekamer. Doordat het gebruik van een respiratiekamer niet toe te passen is in praktijkgericht onderzoek, zijn er andere meetmethodes ontwikkeld die gericht zijn op het meten van CH₄ op stal- of dierniveau. Deze worden in **Hoofdstuk 3.1** toegelicht.

3.1 Methaansensoren

3.1.1 GreenFeed

GreenFeed-metingen worden uitgevoerd in een speciaal ontwikkeld krachtvoerstation (C-Lock Inc., Rapid City, SD, USA; Zimmerman, 2011). Dit station wordt in de stal geplaatst en meerdere keren per dag betreedt een koe dit station om te vreten. Gedurende dit vreetmoment wordt een ademanalyse uitgevoerd. De Greenfeed meet continu de CH₄- en CO₂-concentratie in de afgezogen luchtstroom (Koning et al., 2020). Waarbij de positie van de kop van de koe ook meegenomen wordt. Elke meting bij de Greenfeed duurt circa 3-4 minuten (een zogenaamd 'short-term breath measurement'-techniek) en er zijn meerdere metingen per koe per dag gewenst. Eerder onderzoek heeft laten zien dat er minimaal 20 metingen per koe verzameld moeten worden voordat de binnenkoevariatie stabiel is (Manafaziar et al., 2016; Arthur et al., 2017). Wanneer 30 koeien zes keer per dag in de Greenfeed staan voor een meting, is de bezetting van de Greenfeed 12 volle uren (50% van een etmaal). Het is daardoor raadzaam om niet meer dan 30 koeien tegelijk te bemeten. Omdat het meten met de GreenFeed een dure methode is, wordt er binnen de klimaatvelop-onderzoeken die gebruik maken van deze GreenFeed slechts vier weken op locatie gemeten (Koning et al., 2020). GreenFeed-metingen worden uitgedrukt in emissie (of productie; CH₄ kg/koe/dag en CO₂ kg/koe/dag). De ruwe data vanuit de GreenFeed wordt eerst bewerkt door C-Lock waarna de geaggregeerde data van een website gehaald kan worden voor verdere analyses (zie ook **Hoofdstuk 4.1.2**).

3.1.2 Sniffer

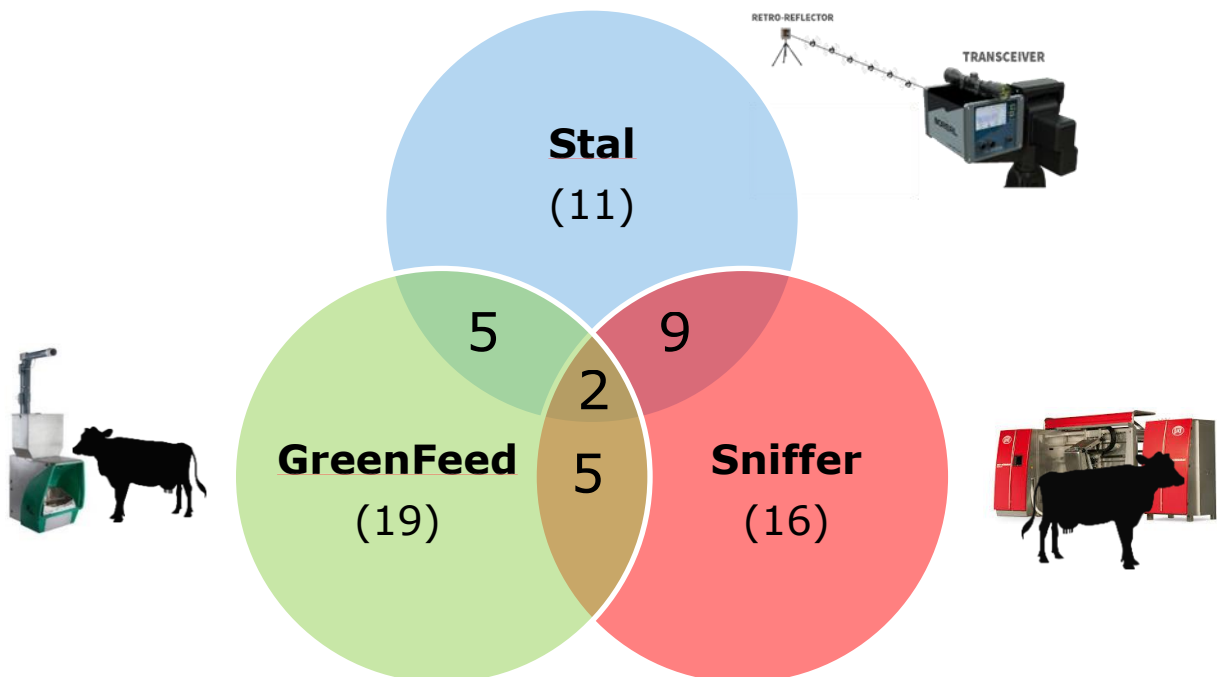
Sniffer-metingen worden uitgevoerd in de voerbak van de melkrobot. Hierbij wordt de CH₄-concentratie continu in de lucht gemeten (Lassen et al., 2012). Doordat elke koe meerdere keren per dag de robot bezoekt, zijn er meerdere meetmomenten per koe per dag. Daarnaast kan de Sniffer op een bedrijf blijven, waardoor er gedurende langere tijd een continue datastroom is van alle productieve melkkoeien die aanwezig zijn op een bedrijf. De nauwkeurigheid van Sniffer-metingen kan wat laag zijn, doordat de positie van de kop van de koe en de achtergrondconcentratie bepalend zijn voor de meting (de Haas et al., 2021). Daarom wordt de Sniffer met name ingezet om koeien binnen een bedrijf te rangschikken. In tegenstelling tot de GreenFeed worden Sniffer-metingen uitgedrukt in concentratie (CH₄ in parts per million (**ppm**) en CO₂ in ppm). In dit onderzoek zijn de ruwe Sniffer data gebruikt en verder bewerkt.

3.1.3 Stal

Emissies op stalniveau worden bepaald door een representatief luchtmonster te nemen uit de nok van de stal. Via een centrale leiding wordt deze lucht aangeboden aan verschillende sensoren die CH₄- en CO₂-concentratie meten. Via het CO₂ productiemodel (Pedersen et al., 2008) wordt het debiet bepaald en daarmee de CH₄-emissie in kg/dierplaats/dag. Ter standaardisatie wordt dit omgerekend naar kg CH₄/dierplaats/jaar. Continue metingen worden gecontroleerd en gecorrigeerd met behulp van referentiemetingen (Mosquera et al., 2021). In het huidige onderzoek zijn de gecontroleerde en gecorrigeerde waarden gebruikt, omdat er op de ruwe data complexe berekeningen uitgevoerd zijn om tot een betrouwbare stalemissie per dag te komen (Schep et al., 2022). Deze berekeningen zijn uitgevoerd door de Afdeling Veehouderij en Omgeving van Wageningen Livestock Research (**WLR**).

3.2 Dataverzameling

In totaal zijn er op 27 Nederlandse melkveebedrijven CH₄-metingen verzameld tussen september 2018 en april 2021 (**Tabel 3.1**). Over deze periode waren er twee bedrijven waarbij alle drie de metingen (GreenFeed, Sniffer en Stal) zijn uitgevoerd, vijf bedrijven met GreenFeed- en Stalmetingen, vijf bedrijven met GreenFeed- en Sniffer-metingen, en negen bedrijven met Sniffer- en Stalmetingen (**Figuur 3.1**; **Tabel 3.1**). Daarnaast zijn er een aantal bedrijven met maar één soort CH₄-meting. Het aantal overlappende dagen en het aantal koeien met metingen is weergegeven in **Tabel 3.1**. Zoals te zien in **Tabel 3.1** is de dataset gebruikt in dit onderzoek gelimiteerd, doordat er slechts een klein aantal bedrijven zijn waar 2 of 3 meetmethodes zijn toegepast.



Figuur 3.1 De drie verschillende meetmethodes en het aantal bedrijven met methaanmetingen.

Tabel 3.1 Overzicht van het aantal bedrijven, dagen en koeien met methaanmetingen voor periode september 2018 – april 2021.

	Aantal bedrijven	Aantal dagen	Aantal koeien
GreenFeed	19	1964	1894
GreenFeed-Sniffer	5	295	234
GreenFeed-Stal	5	455	
Sniffer	16	4540	2381
Sniffer-Stal	9	1784	
Stal	11	5377	
GreenFeed-Sniffer-Stal	2	95	

4 Data Lake

In dit hoofdstuk wordt een korte toelichting gegeven wat een Data Lake is, en hoe het concept is toegepast binnen dit onderzoek. Binnen dit onderzoek is er gewerkt aan een aantal data integraties van verschillende datastromen. Per datastroom worden de knelpunten en gekozen oplossingsrichtingen rondom deze opleveringen besproken.

4.1 Data Lake

Een Data Lake is een systeem of opslagplaats van data in haar natuurlijke of onbewerkte format¹. Deze opslagplaats bevindt zich vaak in de Cloud. Verschillende databronnen, gestructureerd en ongestructureerd, kunnen in een Data Lake bijeengebracht worden. Een Data Lake is meestal een centrale opslag van gegevens, inclusief onbewerkte kopieën van bronsysteemgegevens, sensorgegevens, persoonsgegevens en bewerkte gegevens die worden gebruikt voor taken zoals rapportage, visualisatie, geavanceerde analyses en machine learning.

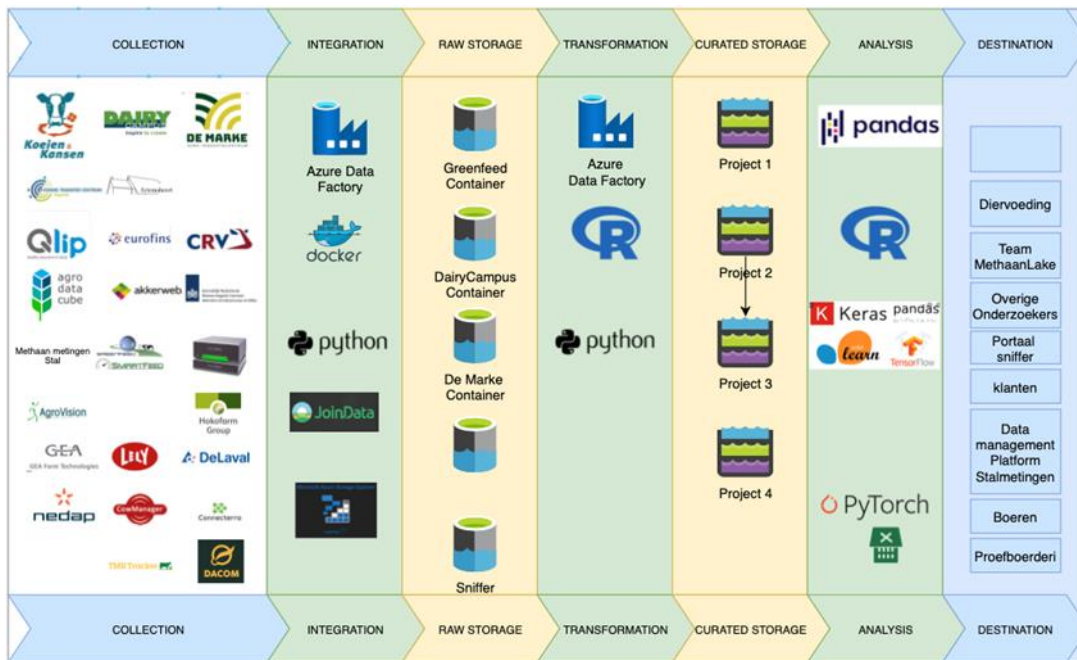
Het voornemen was om het Data Lake te bouwen op het fundament van het reeds in de praktijk draaiende Sniffer Data Lake. Na een inventarisatie van het Sniffer Data Lake bleek dat daar nog geen gebruik werd gemaakt van een Data Lake storage, maar enkel van een SQL server met een data-integratie. In dit onderzoek is ervoor gekozen om binnen de Microsoft Azure Cloud omgeving het Methaan Data Lake in te richten.

Binnen het onderzoek bleek er vaak een spraakverwarring te bestaan tussen de term Data Lake en de term data warehouse. Bij een Data Lake is de opslag van data zonder vooraf vaste structuur, waarbij gegevens in ruwe en originele vorm centraal worden opgeslagen. Hierbij worden ook data opgeslagen die in eerste instantie niet verder gebruikt worden in verdere verwerkings- en analysestappen. Dit is in tegenstelling tot een data warehouse. Hier worden alleen data uit specifiek geselecteerde bronnen opgeslagen, waarbij de gegevens gestructureerd zijn met een vast format². In dit onderzoek is er gekozen voor een Data Lake, zodat alle verschillende CH₄-data centraal opgeslagen konden worden voor verdere analyses. Door dit in de ruwe originele vorm te doen, zonder vooraf bedachte vaste structuur, kunnen de data ook nog gebruikt worden voor toekomstig onderzoek waarbij eventueel een ander format of structuur vereist is. Echter, een Data Lake is alleen een opslagmethode van data. Binnen de bedoelingen van het Methaan Data Lake vallen zowel de opslag, verwerking en analyse van data samen. Wanneer deze elementen aan het systeem worden toegevoegd spreekt men van een "modern data warehouse"³ (Orenstein et al., 2017; How, 2020). Het geheel van systemen is dus een voorbeeld van een modern data warehouse, uiteraard zijn er meerdere uitvoeringen te bedenken van een modern data warehouse. **Figuur 4.1** laat de stappen in het huidige onderzoek zien, gestructureerd volgens de conventies van een 'modern data warehouse'. De logo's en merknamen zijn in dit figuur illustratief; ze zijn niet allemaal aanwezig in de Methaan Data Lake die binnen dit onderzoek ontwikkeld is.

¹ https://en.wikipedia.org/wiki/Data_lake

² <https://www.ictinformatiecentrum.nl/data-management/data-lake>

³ <https://docs.microsoft.com/en-us/azure/architecture/example-scenario/data-warehouse/dataops-mdw>



Figuur 4.1 Voorbeeld van het "modern data warehouse" model die gebruikt is binnen het onderzoek (Bert Klandermans, WLR, persoonlijke communicatie). Van links naar rechts zijn de verschillende stappen gevisualiseerd die samen een modern data warehouse vormen. De logo's en merkmaken zijn illustratief en niet allemaal aanwezig binnen dit onderzoek.

Zoals te zien in **Figuur 4.1** bestaat het concept van een modern data warehouse uit een aantal verschillende stappen. Het totaal aan deze stappen wordt een 'stack' genoemd. Het Methaan Data Lake dat in dit onderzoek is opgezet heeft al deze stappen aan boord:

1. Collection: deze stap richt zich op het verzamelen van data op de plek (in dit onderzoek een Nederlandse melkveehouderijbedrijf) waar de data worden gegenereerd;
2. Integration: deze stap richt zich op het ophalen van de data bij de bron in stap 1 'Collection', en het wegzetten van deze data in het Data Lake;
3. Raw storage: in deze stap worden de data, in haar originele format, ongewijzigd opgeslagen in het Data Lake;
4. Transformation: hier worden de data in haar originele format geconverteerd tot een uniform format en in het Data Lake gezet.
5. Curated storage: het Data Lake waar de geconverteerde uniforme data staat. Zodat deze bevraagd kan worden voor allerlei bewerkingen, zoals rapportages, visualisaties en (geavanceerde) analyses;
6. Analysis: deze stap bevat de tools die verbinding maken met het Data Lake om de analyses daadwerkelijk uit te voeren.

4.2 Materiaal en methode

Voor de technische uitwerking van dit onderzoek is gebruik gemaakt van een Agile werkwijze: door middel van korte ontwikkelopdrachten wordt er periodiek een prototype in de vorm van minimum viable product⁴ (**MVP**) opgeleverd. Deze periodieke MVPs worden ingezet om snel (agile) feedback te ontvangen of om de werking aan te tonen. Na evaluatie, feedback en aanbevelingen zijn er diverse revisies van de prototypes opgeleverd. De verschillende uitgewerkte prototypes bestonden in dit onderzoek voornamelijk uit integratie-pipelines; pipelines om data op te halen bij de bron (zie **Figuur 4.1**). Tijdens het onderzoek zijn er per databron diverse prototypes gemaakt. In dit onderdeel wordt er een korte samenvatting van de ontwikkelde prototypes, problemen gedurende de ontwikkeling, en de gekozen oplossingsrichtingen gegeven.

⁴ https://en.wikipedia.org/wiki/Minimum_viable_product

4.2.1 GreenFeed

Bij de GreenFeed is er sprake van drie datastromen (ruw, geaggregeerd door C-Lock, en verrijkt). Een korte uitleg van deze drie datastromen volgt hieronder. De drie datastromen zijn in het Methaan Data Lake opgenomen:

Ruwe data

Voor dit onderzoek is onderzocht of de ruwe data (dit zijn data zoals de GreenFeed ze vastlegt) periodiek opgehaald konden worden. Na overleg met C-Lock (<https://www.c-lockinc.com/>) zijn de data via een ssh-/sftp-verbinding beschikbaar gesteld. Binnen de SFTP server wordt onderscheid gemaakt tussen de data van de GreenFeed sensor en de geaggregeerde data van C-Lock.

Prototype 1 Azure data factory

In het eerste prototype zijn de data geïntegreerd middels Azure data factory (**ADF**). Een kenmerk van ADF is dat het telkens alle data verwerkt. Doordat alle data op de sftp-server van C-Lock bleef staan, werd niet alleen nieuwe data maar ook de al verwerkte data telkens door ADF verwerkt. Deze verwerking door ADF duurde dagen, en leidde tot zeer hoge rekeningen in Azure. ADF werd daardoor als ongeschikt bevonden voor dit onderzoek.

Prototype 2 Python pipeline

In het volgende prototype is gebruik gemaakt van een sftp-integratie met Python. Python is een scripttaal, ontwikkeld met het oog op een leesbare code, en kan uitgevoerd worden zonder de code te compileren. Door de toepassing van eigen scripts was er met enkele regels code meer controle over de dataverwerking dan met de ADF. Door gebruik te maken van een in-memory databasetechniek om de reeds verwerkte bestanden op te slaan, werd de tijd voor het binnenhalen van nieuwe ruwe data gereduceerd tot enkele minuten. Het Python script draait nu op aanvraag (dit betekent dat het script handmatig wordt aangezet om data binnen te halen). Het kan eenvoudig geïntegreerd worden in "Azure functions" om de data integratie pipeline te automatiseren.

Geaggregeerde data van C-Lock

Deze data is bewerkt door C-Lock en wordt van de website van C-Lock gehaald. Door de directe integratie met het sftp-systeem van C-Lock is het mogelijk om de geaggregeerde data van C-Lock automatisch op te halen en in een eigen database op te slaan.

Prototype 1 Azure Data Factory

Met ADF zouden de verschillende bestanden van de geaggregeerde data kunnen worden ingelezen en worden verwerkt. Echter, er ontstonden al snel problemen omdat C-Lock niet consequent blijkt te zijn in de bestandsindeling. Dit betekent dat bij iedere aanlevering van data er verschillen zijn ten opzichte van de vorige aanlevering. Een voorbeeld hiervan is dat kolommen bij iedere aanlevering verschillende namen hadden. Deze inconsistente manier van data aanleveren vanuit de leverancier levert grote uitdagingen op om in de toekomst geautomatiseerde pipelines te ontwikkelen.

Er is met C-Lock afgesproken om voortaan de data in het ICAR 'animal data-exchange' protocol aan te leveren. Dit is echter nog in ontwikkeling en valt verder buiten de scope van dit onderzoek.

Prototype 2 Python pipeline

De uitdaging van de inconsistente data leverantie vanuit C-Lock is verholpen door een Python-pipeline te ontwikkelen waarin een script van een Dictionary is gemaakt van de verschillende versies, uitzonderingen en conversies. Dit script draait dagelijks, waarbij er dagelijks data op de interne harde schijf vanuit C-Lock wordt ontvangen. Het script heeft wel met grote regelmaat onderhoud nodig, om de verschillende versies van C-Lock bestanden te kunnen blijven ontvangen. Het script is dus een tijdelijke oplossing, maar geen oplossing dat werkbaar is op de lange termijn.

Verrijkte data door WLR

Voor de verrijkte data worden de GreenFeed-databestanden verrijkt met koe- en boerderijnummer (Koning et al., 2020). Tijdens de duur van het onderzoek zijn een aantal keer deze verrijkte data handmatig aangeleverd. Deze verrijkte data zijn vervolgens handmatig in het Methaan Data Lake gezet.

Er zijn pipelines ontwikkeld om GreenFeed data (ruw, geaggregeerd en verrijkt) te transformeren naar curated storage in de Methaan Data Lake stack (stap 5, zie **Figuur 4.1**). Omdat de opmaak van de data in de verschillende handmatige data-aanleveringen verschildte en de structuur van de curated storage gedurende het onderzoek wijzigde, zijn er meerdere prototypes gemaakt. Omdat de data handmatig geleverd werden, was er nog geen noodzaak om de pipeline verder te automatiseren.

4.2.2 Sniffer

Bij de aanvang van het onderzoek werden de ruwe data van de Sniffers al in de Azure-omgeving verwerkt. De zogenaamde ruwe data waren beschikbaar in een SQL server. Via een handmatige download van de ruwe data, werden deze Sniffer data verrijkt met een koenummer. Deze verrijkte data moest ook in het Methaan Data Lake komen. Omdat de data handmatig verwerkt werden, was er geen noodzaak om een automatische pipeline te maken. Wel is er een pipeline gemaakt om de bewerkte data te transformeren naar curated storage (stap 5, **Figuur 4.1**).

4.2.3 Stalmetingen

Wat betreft stalmetingen, deze metingen worden gebruikt om complexe berekeningen uit te voeren om uiteindelijk tot een betrouwbare stalemissie per dag te komen (Mosquera et al., 2021; Schep et al., 2022). In dit onderzoek zijn deze gecontroleerde en gecorrigeerde waarden gebruikt. Deze zijn handmatig toegevoegd in het Methaan Data Lake. In de toekomst zouden de stalmetingen automatisch gekoppeld kunnen worden.

4.2.4 Akkerweb

Voor de integratie van meteorologische data is een integratie-pipeline ontwikkeld met Akkerweb. Er zijn verschillende prototypes gemaakt, getest en geoptimaliseerd zodat de data uit de pipeline het best te verwerken was in de analyses.

Prototype 1 Python pipeline for a single data set

Dit Python gescripte prototype haalde de data van de Akkerweb API op en zette deze data ieder dag in een apart bestand op het Methaan Data Lake. Dit bestand bevatte 24 waarnemingen per uur, voor alle 27 boerderijen met CH₄-metingen. Omdat alle data van alle bedrijven in één bestand geplaatst werden, nam de verwerking van deze data veel tijd in beslag.

Prototype 2 Python pipeline for farm-specific data sets

In het tweede prototype werd de data van Akkerweb per boerderij naar het Methaan Data Lake gezet, waarbij per boerderij alle weerdata in 1 groot bestand kwam te staan. Deze manier was veel efficiënter voor verdere verwerking en is daarom voor de rest van het onderzoek toegepast.

4.2.5 Data Lake structuur en analyse

Bij aanvang van het onderzoek was een Data Lake structuur ontworpen. In een aantal prototypes zijn de ontwerpen geëvalueerd en geoptimaliseerd. In het eerste concept was het idee om alle beschikbare data op te splitsen naar kleine bestandjes per tijdseenheid. Door de dataset op te delen in logische kleinere delen is het datamodel beter schaalbaar. Echter, zoals bij de Akkerweb data had dit een sterk vertragende factor bij de analyse. Er was verder ook geen noodzaak voor de kleine bestandjes. Optimalisaties die uitgevoerd zijn zijn met name gericht op partitionering (opsplitsen) van de data. Partities binnen Data Lakes kunnen helpen om data sneller te bevragen, teveel kleine partities daarentegen vertragen de analyse.

Een Data Lake is een oplossing die vaak wordt gebruikt voor Big Data toepassingen. Tijdens de analyse is er onderzocht hoe de data bevraagd kon worden met Spark & Databricks. Dit is gebruikt om de data te koppelen en te middelen op dagbasis, uurbasis en dierbasis. Deze verwerkte data zijn geschikt om met R-studio te verwerken. Deze offline ontwikkelde analyses zijn later toegepast in de Cloud.

4.3 Resultaten en discussie

Het doel van dit onderzoek was om een basis architectuur te maken waarin alle CH₄-datastromen centraal opgeslagen en bewerkt konden worden, en waar de mogelijkheid was om ook de verbinding te maken met externe databronnen. De keuze om een Methaan Data Lake te bouwen heeft te maken met de flexibiliteit die dit moderne data warehouse biedt met betrekking tot data opslag, -verwerking en -analyse. Een Data Lake is schema vrij, wat betekent dat nieuwe datastromen eenvoudig kunnen worden toegevoegd. Een Data Lake heeft een hoge beschikbaarheid (de tijd per jaar dat het systeem gegarandeerd beschikbaar is voor gebruik; 99,99%), eventueel wereldwijd. En een Data Lake is schaalbaar: de opslagcapaciteit kan tot petabytes oplopen, terwijl er relatief weinig kosten zijn om deze hoeveelheid data op te slaan en te verwerken.

Het automatiseren van de ontwikkelde pipelines was lastig, omdat de verschillende datastromen niet in alle gevallen direct bij de bron konden worden ontsloten. Daarnaast waren we afhankelijk van menselijke verwerking en aanlevering. Hierdoor was er geen noodzaak om de pipeline van A tot Z te automatiseren, maar worden ontwikkelde pipelines nog handmatig gestart indien nodig. Automatisering is echter wel mogelijk door gebruik te maken van Azure Functions mocht dit in de toekomst waardevol lijken te zijn. In Azure Functions kunnen de Python-pipelines, ontwikkelt in dit onderzoek, eenvoudig worden geïntegreerd.

Uiteindelijk bleek dat in dit onderzoek geen grote hoeveelheden aan data bij elkaar werden gebracht. Van Big Data hoeveelheden bleek geen sprake te zijn (qua volume), echter in dit onderzoek waren er veel verschillende formats aan data. Door de relatief kleine volumes aan data lijkt een Data Lake oplossing soms niet de meest efficiënte oplossing, echter de Data Lake oplossing zal wel efficiënter worden als in de toekomst langdurig en op veel meer bedrijven data wordt verzameld. In dit opzicht is in dit project ervaring opgedaan, en een infrastructuur opgezet, die wel toekomstbestendig is. De scripts die in dit onderzoek voor de data analyses zijn ontwikkeld, zullen beschikbaar blijven in de Cloud. Hierdoor kunnen de analyses frequenter en herhaaldelijk uitgevoerd worden als meer data in het Methaan Data Lake geïntegreerd wordt.

5 Data Lake

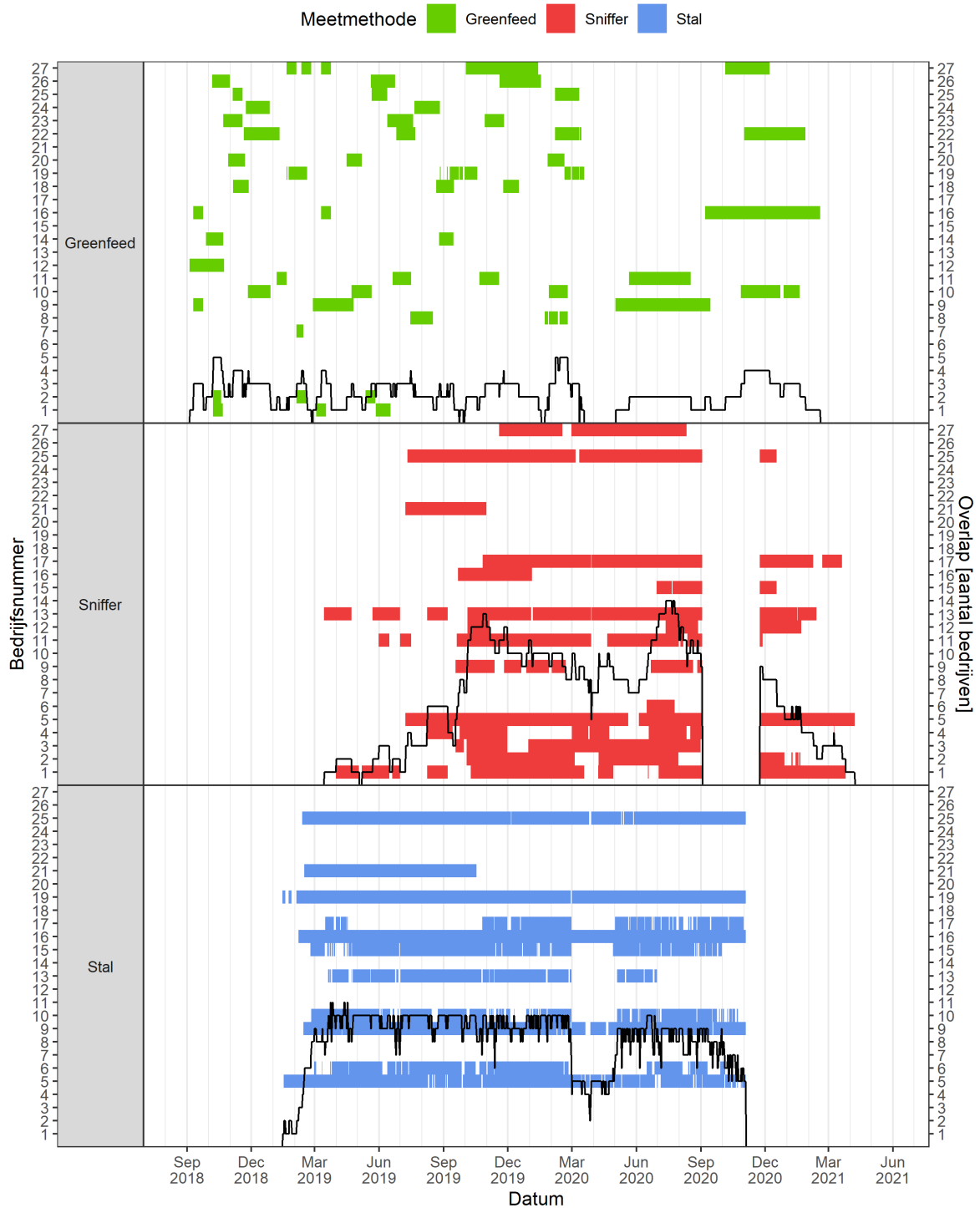
Eén van de doelen van dit onderzoek betreft het bestuderen van variatie en trends op basis van alle CH₄-data die beschikbaar zijn. Ten eerste zijn er CH₄-data van meerdere bedrijven beschikbaar, waardoor patronen door de tijd heen of interacties met verklarende variabelen beter onderbouwd worden. Ten tweede kan de consistentie van variatie, trends, of de ranking van bedrijven worden bepaald door dit te vergelijken tussen de drie verschillende meetmethodes. Zo worden patronen robuuster en leren we hoe meetmethodes elkaar kunnen versterken.

5.1 Materiaal en methode

Data van de GreenFeed-, Sniffer- en Stalmetingen tussen september 2018 en april 2021 zijn in het Methaan Data Lake bijeen gebracht. **Figuur 5.1** geeft een schematisch overzicht van de beschikbaarheid van data per meetmethode. Ook is het aantal bedrijven met overlap per meetmethode te zien. Ter voorbereiding van de analyses zijn de data op verschillende manieren opgeslagen: als daggemiddelden per bedrijf, maar ook uurgemiddelden per bedrijf en diergemiddelden per dag per bedrijf. Vanuit deze datasets zijn vervolgens de analyses uitgevoerd.

5.1.1 Aanwezige variatie

De onzekerheid die aanwezig is rondom een gemiddelde emissie of concentratie kent drie componenten (**Figuur 5.2**) en wordt uitgedrukt als variantie. De eerste component in de formule is de tussen-bedrijfsvariantie. Deze variatie tussen bedrijven wordt veroorzaakt door bijvoorbeeld management- of rantsoenverschillen, maar ook verschillen tussen stalsystemen en rassen spelen hierin een rol. De tweede variantiecomponent is de binnen-bedrijfsvariantie. Deze component bevat in deze studie de verschillen tussen en binnen dagen per bedrijf. Als derde variantiecomponent is er de variantie gerelateerd aan de metingen zelf. De meetonzekerheid in concentratie van het instrument bepaalt hoe groot deze derde variantiecomponent is. Voor emissies wordt deze component vergroot met de onzekerheid gerelateerd aan het ventilatiedebiet. Er wordt aangenomen dat de laatste variantiecomponent random is en dus de ene keer een overschatting en de andere keer een onderschatting is (Hofschreuder et al., 2008). Deze aanname maakt het mogelijk om dieren, bedrijven, trends over tijd en meetmethodes met elkaar te vergelijken.



Figuur 5.1 Overzicht van de beschikbare data per meetmethode (GreenFeed, Sniffer, Stal) per bedrijf per dag. De zwarte lijn geeft op de rechter-as het aantal bedrijven weer met overlappende metingen.

$$\sigma_{\text{total}}^2 = \frac{\sigma_b^2}{k} + \frac{\sigma_w^2}{k \cdot l} + \frac{\sigma_m^2}{k \cdot l \cdot m}$$

Figuur 5.2 De onzekerheid rondom een gemiddelde emissie of concentratie bestaande uit drie componenten; (1) σ_b^2 : variantie tussen bedrijven; (2) σ_w^2 : variantie binnen bedrijven en (3) σ_m^2 : variantie tussen metingen met k het aantal bedrijven, l het aantal metingen per bedrijf en m het aantal herhalingen per meting (Hofschreuder et al., 2008).

5.1.2 Analyse van variatie en trends

Om te onderzoeken op welke manier de meetmethodes elkaar kunnen versterken en wat er te leren valt aan de hand van grote datasets, is er een vergelijking gemaakt tussen bedrijven (paragraaf 5.1.4) en tussen meetmethodes (Paragraaf 5.1.5). Daarnaast kan op basis van de variatiecoëfficiënt (**VC**; paragraaf 5.1.3) en de ranking een vergelijking gemaakt worden tussen dieren. Een vergelijking tussen dieren laat zien welke variatie in CH₄-productie er aanwezig is binnen de Nederlandse melkveestapel. Een vergelijking tussen bedrijven laat zien of gevonden trends consistent zijn wanneer deze van veel verschillende bedrijven bekeken worden. De vergelijking tussen meetmethodes laat zien of dezelfde conclusies getrokken kunnen worden met verschillende meetmethodes.

In de analyses zijn zoveel mogelijk bedrijven meegenomen. Omdat Sniffermetingen betrouwbaarder worden wanneer gemeten over een langere tijd, is de keuze gemaakt om bedrijven met meer dan 270 dagen aan data te selecteren voor de analyse.

5.1.3 Variatiecoëfficiënt

De VC is berekend om per meetmethode en per gas (CH₄ en CO₂) te kunnen beoordelen welke soort variatie (binnen dagen, tussen dagen, tussen bedrijven of tussen dieren) het grootst is. De VC drukt de standaarddeviatie uit als fractie van het gemiddelde (**Formule 1**). Om de VC van alle bedrijven samen te voegen is deze berekend in twee stappen.

$$VC_{ij} = \frac{SD_{ij}}{G_{ij}} \quad 1$$

Waar VC_{ij} de variatiecoëfficiënt is voor dag of dier i en bedrijf j . SD_{ij} is de standaarddeviatie en G_{ij} het gemiddelde. Onder meetperiode i worden alle data verstaan die per bedrijf beschikbaar waren. Om de VC per dag of per dier samen te voegen tot één getal op bedrijfsniveau is **Formule 2** toegepast.

$$VC_j = \sqrt{\frac{\sum_{i=1}^n (VC_{ij})^2}{n}} \quad 2$$

Hierbij worden de VC voor alle dagen of alle dieren per bedrijf gemiddeld tot VC_j ; de VC per bedrijf. Door nu met **Formule 3** alle variatiecoëfficiënten per bedrijf met elkaar te middelen, wordt de uiteindelijke VC per bedrijf verkregen.

$$VC = \sqrt{\frac{\sum_{j=1}^n (VC_{ij})^2}{n}} \quad 3$$

Met j als bedrijf 1 tot n (n is maximaal 27).

5.1.4 Vergelijking tussen bedrijven

Dagpatronen vergelijken

Om dagpatronen van verschillende bedrijven met elkaar te vergelijken, is het nodig de variatie tussen dagen, tussen bedrijven en tussen dieren buiten beschouwing te laten. Emissies en concentraties zijn daarom eerst per bedrijf op uurbasis gemiddeld (**Formule 4**). Dagpatronen kunnen vervolgens worden vergeleken tussen bedrijven op een relatieve schaal door **Formule 5** toe te passen. Dit betekent dat het daggemiddelde 100% is en uurgemiddelden worden uitgedrukt als percentage van het daggemiddelde.

$$\overline{E_{uur}} = \frac{1}{n_{dagen}} \sum_{dag=1}^{n_{dagen}} E_{uur} \quad 4$$

Met $\overline{E_{uur}}$ als emissie of concentratie per bedrijf per uur (uur variërend tussen 1 en 24) en n_{dagen} als het totaal aantal dagen per bedrijf. Dit betekent dat de emissie of concentratie per uur een gemiddelde is van alle beschikbare dagen van het desbetreffende bedrijf. In **Formule 5** worden de uurgemiddelden vervolgens omgezet van absolute waarden naar een fractie van het daggemiddelde.

$$\overline{Rel_{uur}} = \frac{\overline{E_{uur}}}{\frac{1}{24} \sum_{uur=1}^{24} \overline{E_{uur}}} \quad 5$$

Met $\overline{Rel_{uur}}$ als relatieve uurgemiddelde per bedrijf.

Jaarpatroon vergelijken

De trend in CH₄-concentratie en -emissie is bepaald aan de hand van Sniffer- en Stalmetingen. Deze meetmethodes zijn geselecteerd vanwege de langere doorlooptijd per bedrijf. GreenFeed-metingen zijn minder geschikt voor dit doeleinde, omdat deze slechts gedurende een korte periode op een bedrijf aanwezig waren (zie **paragraaf 3.1.1**), waardoor bedrijfsgerelateerde kenmerken een grotere rol spelen dan het jaarpatroon. Bedrijven met Sniffer-data zijn geselecteerd op minimaal 270 dagen aan data. De periode tussen september 2019 en september 2020 is gebruikt om het jaarpatroon te bepalen, omdat hier de meeste Snifferdata beschikbaar waren (**Figuur 5.1**).

Om de jaarpatronen van verschillende bedrijven te kunnen vergelijken, is de variatie binnen dagen en tussen bedrijven buiten beschouwing gelaten. Variatie binnen een dag is niet meegenomen door met daggemiddelde waarden te werken. Variatie tussen bedrijven is buiten beschouwing gelaten door de emissies of concentraties op dagbasis te delen door de jaargemiddelde emissie per bedrijf (**Formule 6**). Op deze manier wordt de emissie per dag uitgedrukt als fractie van de emissie op jaarbasis.

$$\overline{E_{dag;bedrijf}} = \frac{E_{dag;bedrijf}}{\overline{E_{bedrijf}}} \quad 6$$

Met $\overline{E_{dag}}$ als gemiddelde relatieve emissie of concentratie per bedrijf per dag in het jaar, $E_{dag;bedrijf}$ als daggemiddelde emissie of concentratie per bedrijf en $\overline{E_{bedrijf}}$ als jaargemiddelde emissie of concentratie per bedrijf.

5.1.5 Vergelijking tussen meetmethodes

Een vergelijking tussen meetmethodes is gemaakt om te onderzoeken of dezelfde conclusies kunnen worden getrokken op basis van verschillende meetmethodes.

GreenFeed – Stal

Het verschil tussen GreenFeed- en Stalmetingen kan een indicatie geven van de mestbijdrage aan de totale CH₄-emissie. De GreenFeed meet immers de CH₄ uit pensfermentatie en in de Stalmetingen wordt de totale CH₄-emissie gemeten. Hierbij moet echter wel rekening worden gehouden met het verschil in eenheid tussen GreenFeed (in kg/koe/dag) en Stal (in kg/dierplaats/jr). Beide metingen zijn omgerekend naar kg/koe/jr. GreenFeed-metingen zijn daarvoor vermenigvuldigd met 365. Dit resulteert waarschijnlijk in een overschatting van de jaarproductie op basis van GreenFeed, aangezien de GreenFeed enkel dieren meet die in lactatie zijn, terwijl koeien ook een deel van het jaar geen melk produceren. Tijdens de droogstand is de opname van voer lager, omdat er geen energie nodig is voor melkproductie en daarmee is ook de CH₄-emissie lager (Colucci et al., 1982; Bannink et al., 2011). Om de Stalmetingen om te rekenen naar kg/koe/jr zijn alleen de stallen geselecteerd waarin geen jongvee gehuisvest werd. Vervolgens zijn de Stalmetingen omgerekend naar een gelijke bezetting van de stal.

GreenFeed – Sniffer

In een vergelijking tussen GreenFeed- en Sniffer-metingen is onderzocht of er een verband aanwezig is tussen CH₄-concentratie uit de adem van de koe en de gemeten CH₄-emissie. De vergelijking is op twee manieren uitgevoerd: enerzijds kwantitatief door van beide meetmethodes weekgemiddelde waarden te berekenen en deze in een grafiek tegen elkaar uit te zetten, en anderzijds is een kwalitatieve vergelijking gemaakt op koe-niveau. Het aantal dieren voor de vergelijking is beperkt door de criteria te stellen dat er per koe minimaal 270 dagen aan Sniffer-data beschikbaar moesten zijn en 30 dagen aan GreenFeed-data. De vergelijking is kwalitatief gemaakt door dieren volgens beide meetmethodes te ranken van laag (lage concentratie/emissie, rank 1) naar hoog (hoge concentratie/emissie).

GreenFeed – Sniffer – Stal

De overlap tussen alle drie de meetmethodes is dat ze een uitspraak kunnen doen op bedrijfsniveau. Een vergelijking is gemaakt door bedrijven te ranken van laag naar hoog. Het criterium voor de selectie van Sniffer-data was dat minimaal 270 dagen aan data beschikbaar moesten zijn. Vervolgens zijn de bedrijven geselecteerd waarvoor er overlap was tussen een van de combinaties GreenFeed – Stal, Sniffer – Stal of GreenFeed – Sniffer.

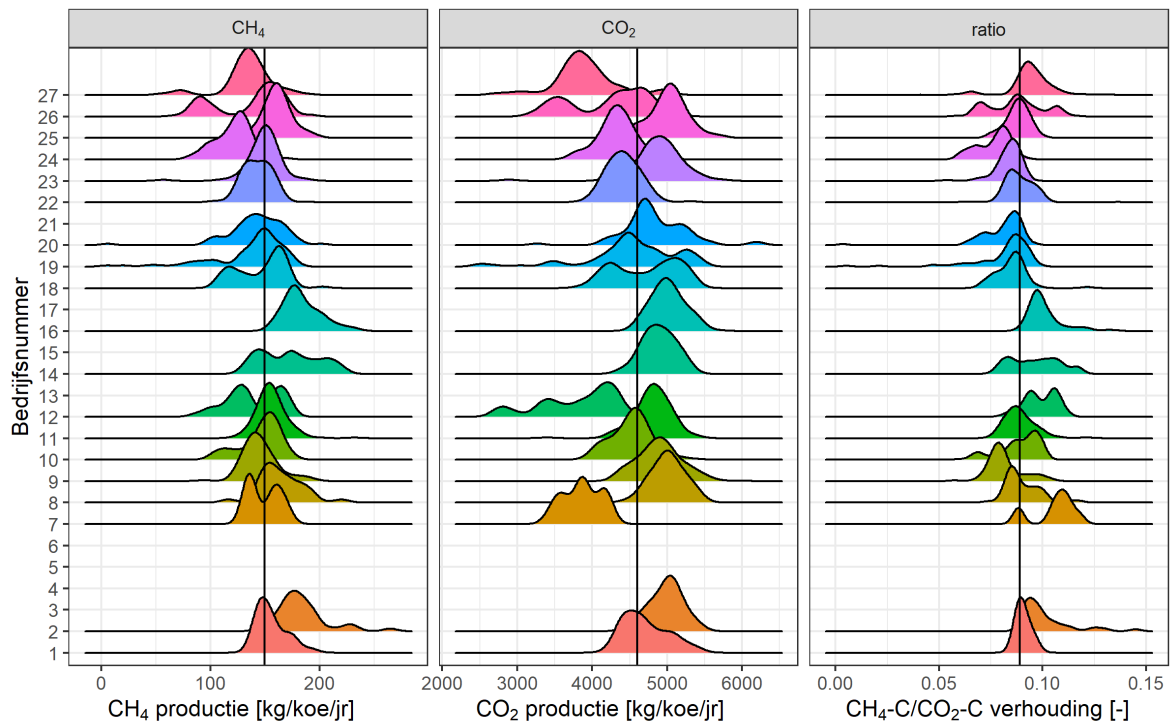
5.2 Resultaten en discussie

Hieronder zullen de verschillende resultaten gepresenteerd en bediscussieerd worden. Eerst worden de algemene gegevens voor de verschillende meetmethodes besproken. Daarna wordt ingegaan op de variatie binnen en tussen bedrijven, en de variatie binnen een jaar. Vervolgens worden de verschillende meetmethodes met elkaar vergeleken en als laatste wordt er gekeken naar de rangschikking van dieren binnen een bedrijf en de rangschikking van bedrijven.

5.2.1 Algemeen meetmethodes

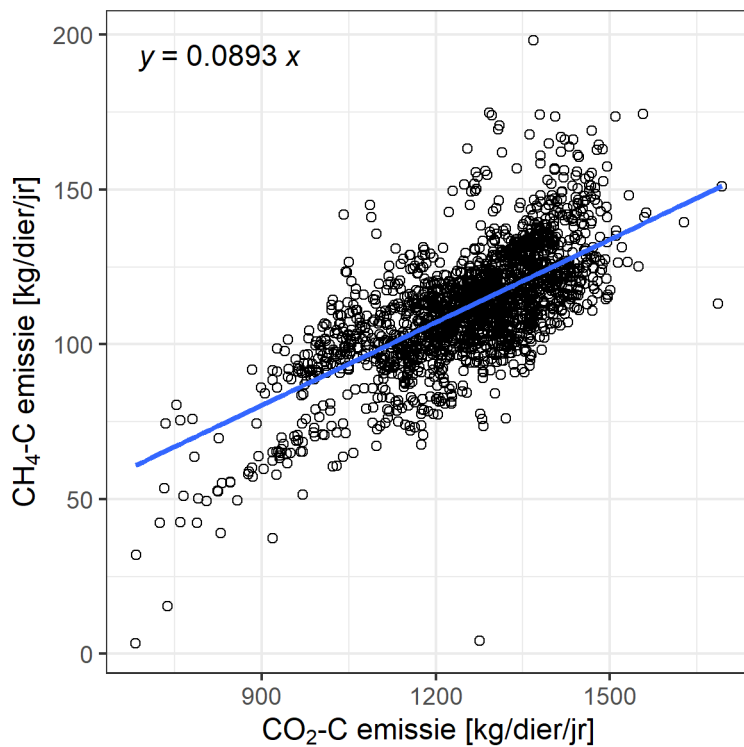
GreenFeed

Figuur 5.3 geeft een overzicht van de spreiding binnen een bedrijf en tussen bedrijven voor GreenFeed-metingen. De breedte van de curve laat zien wat de spreiding binnen een bedrijf is. Hieruit wordt zichtbaar dat er bandbreedtes zijn per bedrijf, maar dat deze wel vrijwel allemaal overlappen op de bedrijven waar gemeten is met de GreenFeed. Het valt op dat bedrijven met een gemiddeld lagere CO₂-productie (bedrijven 7, 12 en 27), een hogere CH₄-C/CO₂-C verhouding kennen. Dit zou gerelateerd kunnen zijn aan bijvoorbeeld voer of ras; wat op deze bedrijven wordt gehouden, dit wordt in andere onderzoeken nog onderzocht.



Figuur 5.3 Frequentiegrafiek van de CH₄-productie, CO₂-productie en de CH₄-C/CO₂-C verhouding per bedrijf (aangegeven in verschillende kleuren) gemeten met de GreenFeed. De hoogte van de curve geeft een indicatie van het aantal waarnemingen per mate van de productie. De gemiddelde CH₄- en CO₂-productie en CH₄-C/CO₂-C verhouding is aangegeven met een verticale lijn.

De productieverhouding tussen CH₄ en CO₂ op koolstofbasis (0.089 ± 0.011) zoals zichtbaar in **Figuur 5.3**, is ook zichtbaar in **Figuur 5.4**. **Figuur 5.4** maakt zichtbaar dat er een duidelijk verband is tussen de CH₄- en CO₂ productie van het dier. Dit verband komt overeen met eerdere bevindingen zoals door Madsen et al. (2010) met een CH₄/CO₂ verhouding tussen de 0.06 en 0.10 waarbinnen de gevonden productieverhouding valt.



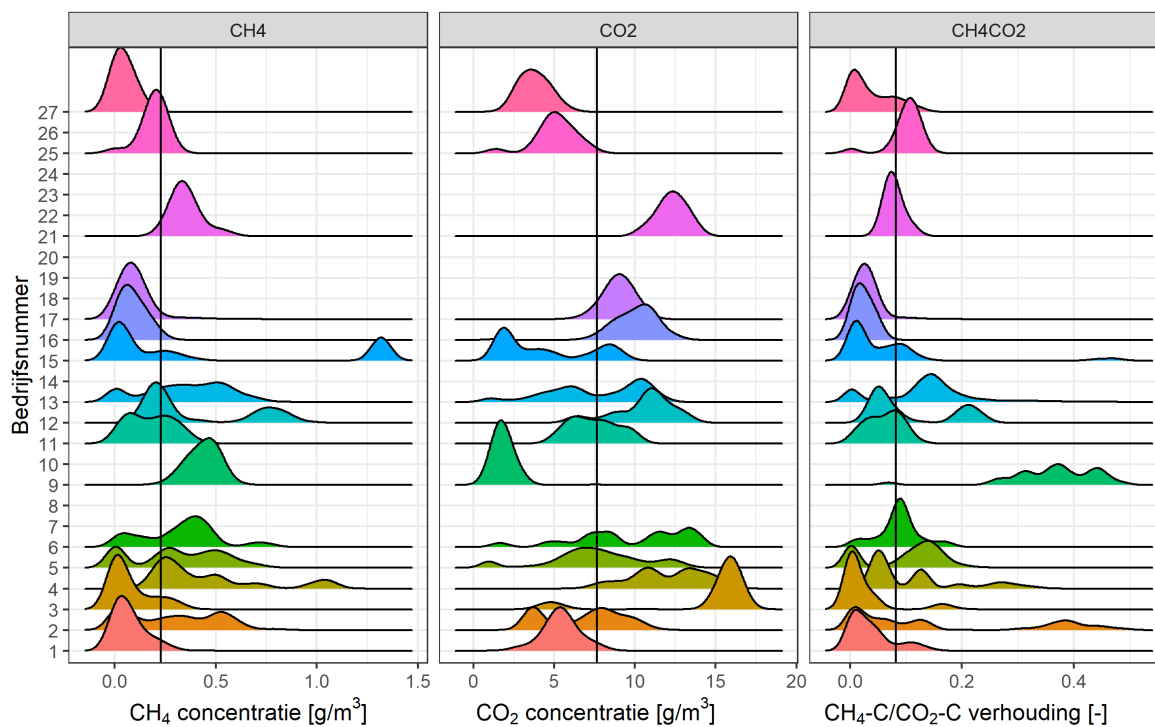
Figuur 5.4 Daggemiddelde CH₄-C en CO₂-C emissies in kg/dpl/jr gemeten met de GreenFeed.

Sniffer vs GreenFeed

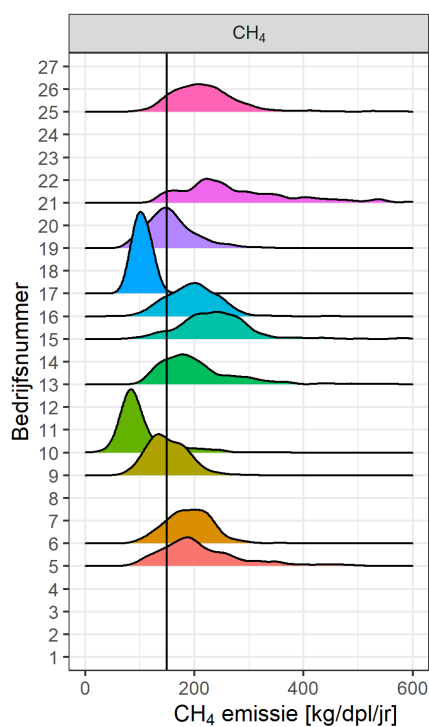
Sniffer-metingen zijn weergegeven in **Figuur 5.5**. Hierin is te zien dat de spreiding per bedrijf behoorlijk groot is en dat er meer verschillen tussen bedrijven worden waargenomen dan bij de GreenFeed. Dit wordt verklaard doordat de Sniffer een geheel andere meetmethode toepast dan de GreenFeed. De GreenFeed kan nauwkeurige metingen verrichten, maar is een erg dure techniek. De Sniffer daarentegen, is een veel goedkopere techniek en is daardoor makkelijker opschaalbaar in aantal bedrijven en in de duur van de metingen. De Sniffer wordt gebruikt om veel metingen te doen die geaggregeerd een goede schatting kunnen geven. Door de verschillen in meetmethodes tussen Sniffer en Greenfeed is het lastig om op basis van individuele waarnemingen, zoals weergegeven in **Figuur 5.5**, algemene uitspraken te kunnen doen. Geaggregeerde analyses zullen in de volgende hoofdstukken verder aan de orde komen.

Stal

Evenals de GreenFeed-metingen, zijn ook de Stalmetingen erop gericht een kleine meetonzekerheid te realiseren rondom een individuele meting. **Figuur 5.6** geeft een overzicht van de gemeten CH₄-emissie per bedrijf. Net als bij de GreenFeed-metingen, overlappen de emissies van alle bedrijven deels. Bedrijf 10 en 17 hebben een duidelijk lagere CH₄-emissie per dierplaats dan het gemiddelde. Hierbij moet in acht worden genomen dat deze emissie niet gecorrigeerd is voor bezettingsgraad in de stal en dat dit een oorzaak kan zijn van de lagere emissies.



Figuur 5.5 Frequentiegrafiek van de CH₄-concentratie, CO₂-concentratie en de CH₄-C/CO₂-C concentratieverhouding per bedrijf (aangegeven in verschillende kleuren) gemeten met Sniffers. De hoogte van de curve geeft een indicatie van het aantal waarnemingen per concentratieniveau.



Figuur 5.6 Frequentiegrafiek van de CH₄-emissie per bedrijf (aangegeven in verschillende kleuren), op basis van Stalmetingen. De hoogte van de curve geeft een indicatie van het aantal waarnemingen per hoeveelheid emissie.

5.2.2 Variatie

Tabel 5.1 geeft een overzicht van de gevonden variatiecoëfficiënten (**VC**) per meetmethode en per gastype (CH₄ of CO₂). Vanwege de grote variatie tussen individuele Sniffer-metingen (**Figuur 5.5**) ten opzichte van de GreenFeed-metingen (**Figuur 5.3**), is ook de VC van de Sniffer groter. Het valt op dat de variatie tussen dieren per bedrijf het grootst is, op basis van zowel de GreenFeed als de Sniffer. Op basis van de Sniffer is de variatie tussen dagen groter dan de variatie binnen dagen. Dit kan verklaard worden door drift van de meetmethode over tijd. De GreenFeed-metingen resulteren in een grotere variatie binnen dagen ten opzichte van de variatie tussen dagen. Dit kan verklaard worden door een dagpatroon in activiteit van de dieren of metabolisme van de dieren wat afhankelijk kan zijn van factoren zoals temperatuur (Ngwabie et al., 2011).

Op basis van CH₄ is de variatie tussen bedrijven kleiner dan de variatie tussen dagen, terwijl voor CO₂ het omgekeerde het geval is. Dit heeft er mee te maken dat de CO₂-productie gerelateerd is aan de metabolische energiebehoefte. Deze metabolische energiebehoefte is meer stabiel over de tijd dan de CH₄-productie. Deze laatste is sterk afhankelijk van voeropname, rantsoensamenstelling en momenten van voeren (Jentsch et al., 2009; Hegarty, 2013; Knapp et al., 2014).

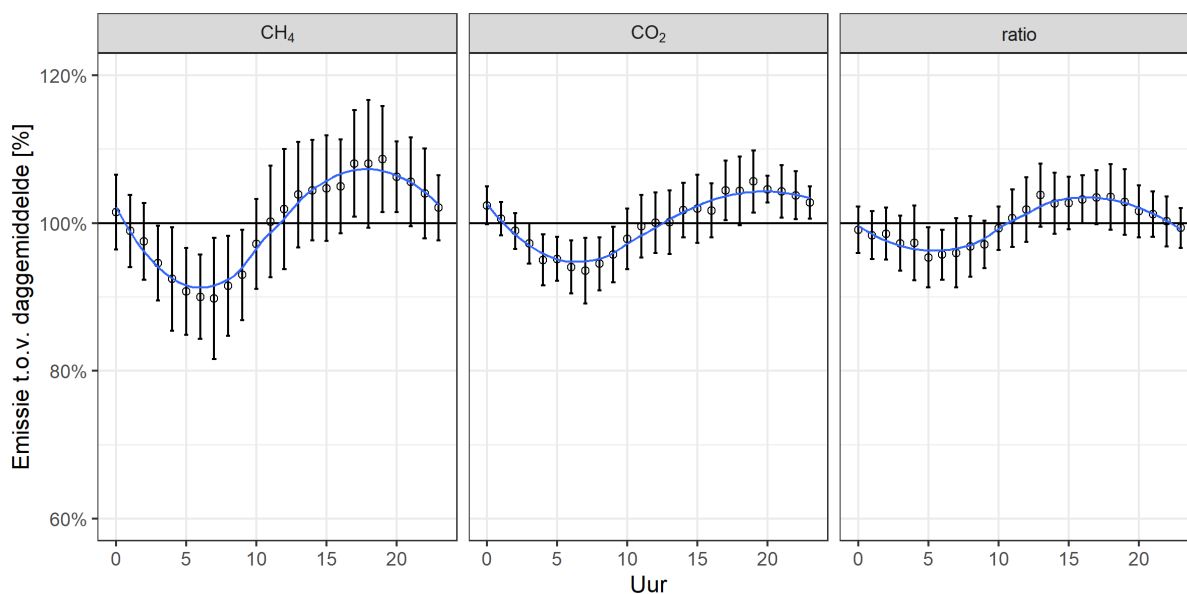
Tabel 5.1 Variatiecoëfficiënten per meetmethode van CH₄ en CO₂ binnen dagen, tussen dagen, tussen bedrijven en tussen dieren. De kleurverdeling per rij geeft de hoogte van de variatiecoëfficiënt weer en loopt van laag (wit) naar hoog (oranje).

Gas	Meetmethode	Binnen dagen per bedrijf	Tussen dagen per bedrijf	Tussen bedrijven	Tussen dieren per bedrijf
CH ₄	GreenFeed	0.19	0.14	0.11	0.2
CH ₄	Sniffer	0.61	0.71	0.58	1.54
CH ₄	Stal	-	0.31	0.3	-
CO ₂	GreenFeed	0.11	0.07	0.09	0.13
CO ₂	Sniffer	0.21	0.27	0.43	0.88
CO ₂	Stal	-	-	-	-

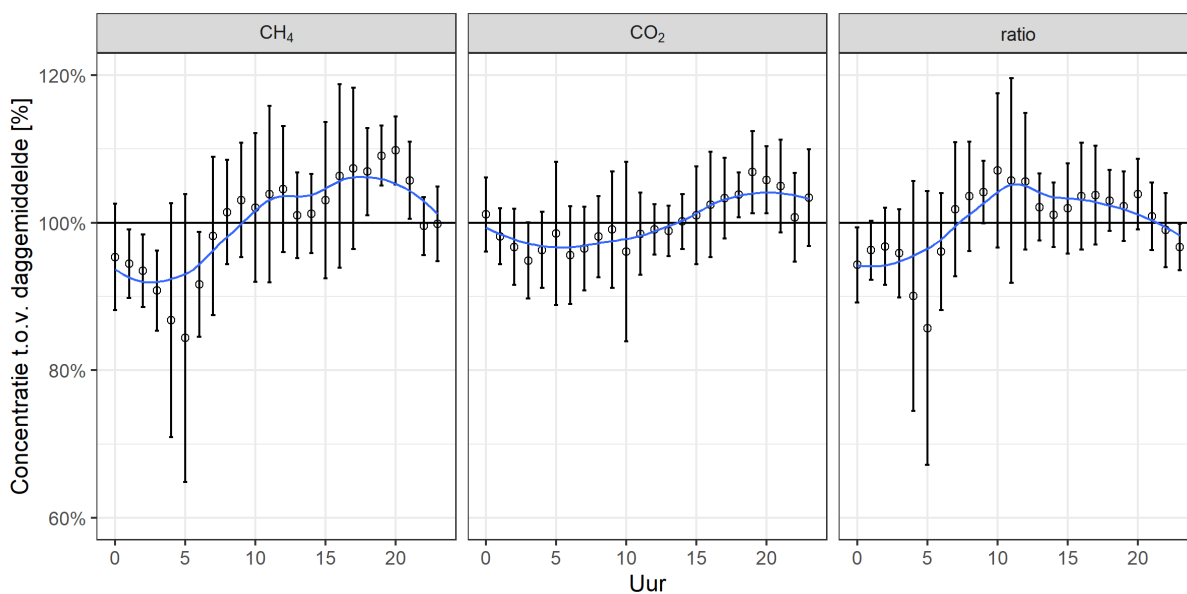
Tussen bedrijven – GreenFeed en Sniffer

Dagpatronen van de bedrijven waar gemeten is met de GreenFeed en/of de Sniffer zijn gevisualiseerd in respectievelijk **Figuur 5.7** en **Figuur 5.8**. Beide figuren laten zien dat er een duidelijk patroon in emissie (GreenFeed) en concentratie (Sniffer) aanwezig is door de dag heen. Aan het begin van de dag zien we lagere CH₄ en CO₂ emissies en in de avond zien we de hoogste emissies. De spreiding door de dag heen is ≤10% voor CH₄ en ≤5% voor CO₂. Dit is vergelijkbaar met de variatie die gevonden is in Manafiazar et al. (2016) en Bell et al. (2018). Door het verschil in spreiding, is er ook een dagpatroon zichtbaar in de CH₄-C/CO₂-C verhouding. In de ochtend wordt er per eenheid CO₂ minder CH₄ (enterisch) uitgedemd dan in de middag. Dit patroon wordt eveneens verklaard door het metabolisme wat een meer constant proces is dan de CH₄-productie (Manafiazar et al., 2016).

Het valt op dat bij de Sniffer de CO₂-concentratie later op de dag boven de 100% komt dan de CH₄-concentratie (**Figuur 5.8**). Hier is verder geen literatuur over bekend.



Figuur 5.7 Dagpatroon van de CH₄- en CO₂-emissie en de CH₄-CO₂ ratio op basis van 19 bedrijven waar gemeten is met de GreenFeed. Standaarddeviatie rondom de uurgemiddelden geeft de spreiding tussen bedrijven weer.

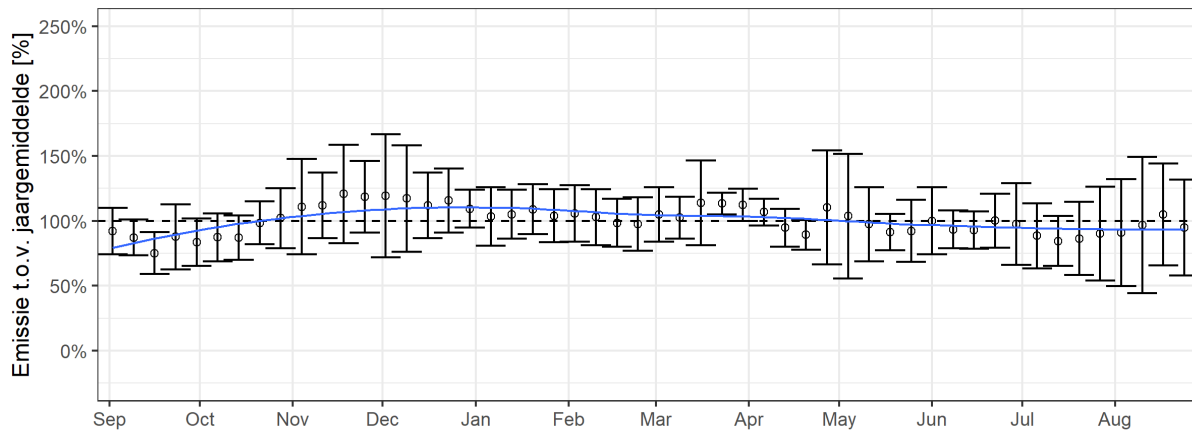


Figuur 5.8 Dagpatroon van de CH₄- en CO₂-concentratie en de CH₄-CO₂ ratio op basis van 11 bedrijven waar langer dan 9 maanden gemeten is met de Sniffer. De standaarddeviatie rondom de uurgemiddelden geeft de spreiding tussen bedrijven weer.

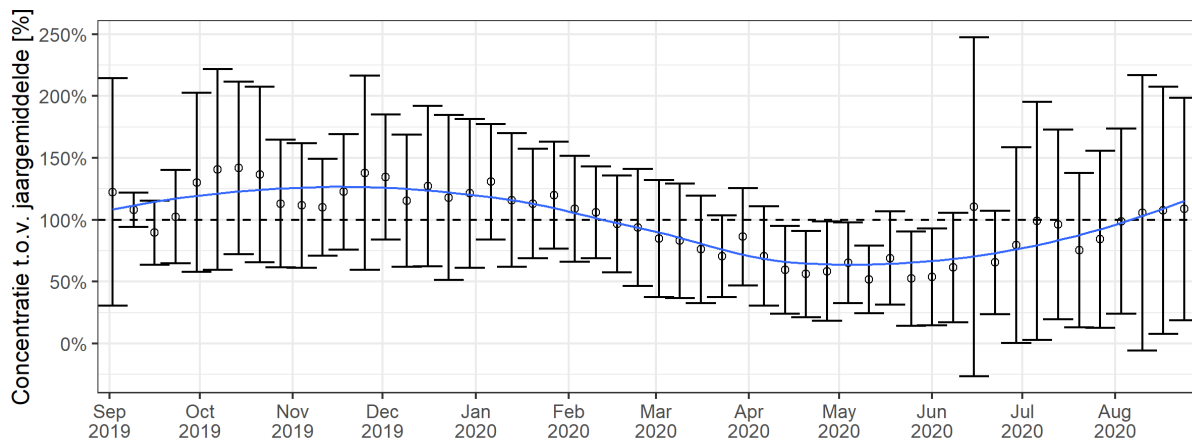
Jaarpatroon Sniffer en Stal

De jaarpatronen volgend uit de emissie- of concentratiemetingen van CH₄ van respectievelijk Stal en Sniffer, zijn zichtbaar in **Figuur 5.9** en **Figuur 5.10**. Het verschil in spreiding binnen het jaar valt op. Stalmetingen variëren gemiddeld 20% rondom het jaargemiddelde en Sniffer-metingen 50%. Wel laten beide meetmethodes een zelfde soort patroon zien door het jaar heen. Beide meetmethodes resulteren in een hogere CH₄-emissie (Stal) en -concentratie (Sniffer) gedurende de winter en een lagere emissie en concentratie gedurende de zomer. Verwacht wordt dat bij de stalmetingen gedurende de winter het effect van mestopslag zichtbaar wordt. Mest mag niet worden uitgereden tussen 1 september en 14 februari. Daarnaast is het bekend dat de CH₄-productie afhankelijk is van de temperatuur en de 'leeftijd' van de mest; bij een hogere temperatuur en 'oudere' mest is de CH₄-productie hoger (Zeeman, 1991). Dit patroon is, los van de variatie, zichtbaar in **Figuur 5.9**. Gedurende de winterperiode (september – februari) is de meeste

mest aanwezig in de stal en is ook de CH₄-emissie het hoogst. Ngwabie et al. (2011) heeft echter aangetoond dat ook de enterische CH₄-productie hoger is bij lagere temperaturen. Bij lagere temperaturen is de activiteit van de dieren hoger en verliest de koe meer warmte waardoor er een hoger metabolisme is en daarmee ook een hogere CH₄-emissie. Dit zien we ook terug in **Figuur 5.10** waar het jaarpatroon van de Sniffer concentratie ook hogere waarden geeft gedurende de winterperiode. Tevens kan ook ruwvoersamenstelling een effect hebben op het jaarpatroon van de CH₄-emissie (Van Dijk et al., 2020).



Figuur 5.9 Jaarpatroon op basis van de wekelijkse gemiddelde CH₄-emissie van 11 bedrijven waar Stalmetingen gedaan zijn in de periode van september 2019 tot september 2020. De standaarddeviatie rondom de uurgemiddelden geeft de spreiding tussen bedrijven weer.



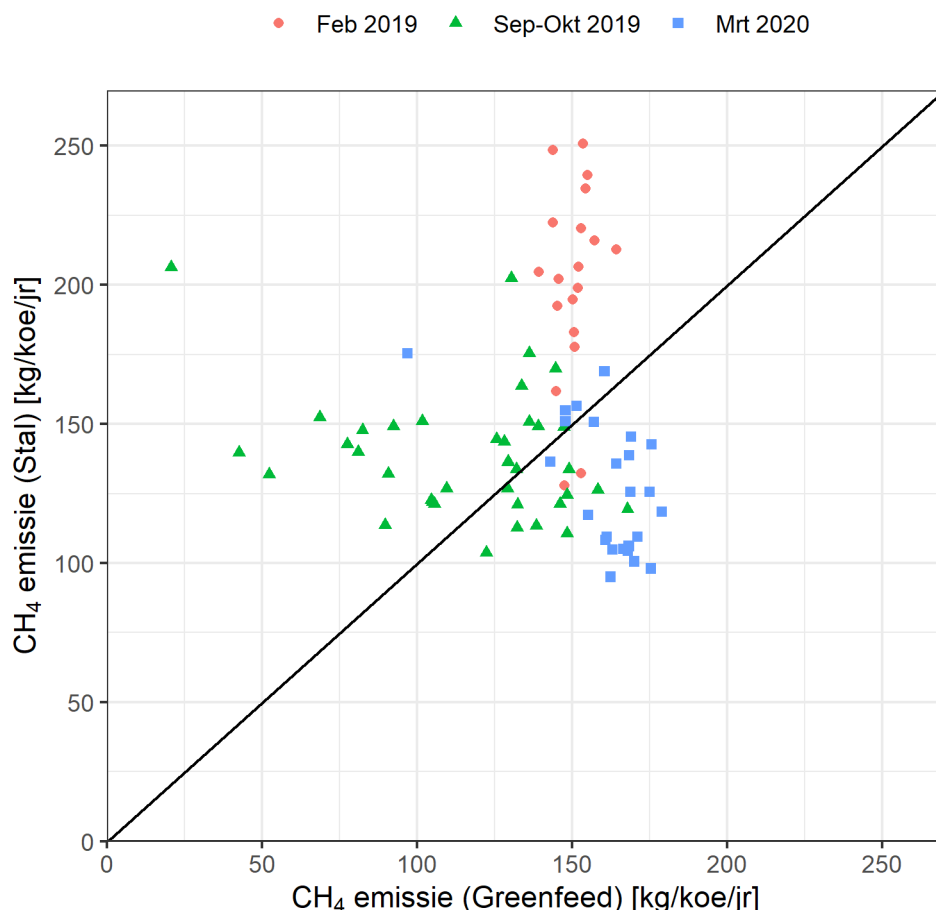
Figuur 5.10 Jaarpatroon van de wekelijkse gemiddelde CH₄-concentratie op basis van 11 bedrijven waar Sniffer metingen gedaan zijn in de periode van september 2019 tot september 2020. De standaarddeviatie rondom de uurgemiddelden geeft de spreiding tussen bedrijven weer.

5.2.3 Tussen meetmethodes

GreenFeed – Stal

Op één bedrijf waar zowel Stal- als GreenFeed-metingen zijn uitgevoerd, was er gemeten in een stal met enkel melkvee. Jongvee en droogstaande koeien waren hier afwezig. Hierdoor kon de stalemissie in kg/dierplaats/jr omgerekend worden naar kg/koe/jr en vergeleken worden met de GreenFeed-metingen. Een onderzoeksvraag was of de mestbijdrage aan de CH₄-productie berekend kan worden door de GreenFeed-emissie in mindering te brengen op de Stalemissie. **Figuur 5.11** laat zien dat de stalemissie niet altijd op een hoger niveau ligt dan de emissie gemeten met de GreenFeed, wat resulteert in een negatieve mestbijdrage. Dit is fysisch onmogelijk. Een mogelijke oorzaak voor een GreenFeed emissie die hoger is dan de Stalemissie ligt in de selectie van de dieren. Op basis van de Stalmetingen wordt er een gemiddelde CH₄-emissie bepaald op basis van alle dieren. De GreenFeed geeft een schatting op basis van een selectie van dieren. Wanneer de steekproef niet representatief is, kan hierdoor een verschil ontstaan. Een andere reden voor de negatieve mestbijdrage is de meetonzekerheid rondom de meetmethodes. Dit wordt ook zichtbaar in **Tabel 5.2** waar de onzekerheid rondom de mestbijdrage is aangegeven (grote standaardafwijking).

In totaal is er in drie perioden van drie weken tegelijkertijd gemeten met GreenFeed en Stalmetingen (zie ook **Figuur 5.11**). Het gemiddelde per periode, is opgenomen in **Tabel 5.2**. De hoogste berekende mestbijdrage wordt gevonden in februari. De laagste mestbijdrage is geschat in maart 2020. September 2019 zit hier precies tussen wat betreft berekende mestbijdrage. Deze mestbijdrage komt overeen met de verwachte momenten van mest uitrijden zoals besproken bij **Figuur 5.9**, omdat de periode van het uitrijden van mest start op 14 februari en er daardoor in maart een lagere mestbijdrage is gevonden. Hierbij nemen we nog steeds aan dat de GreenFeed hogere emissies schat.



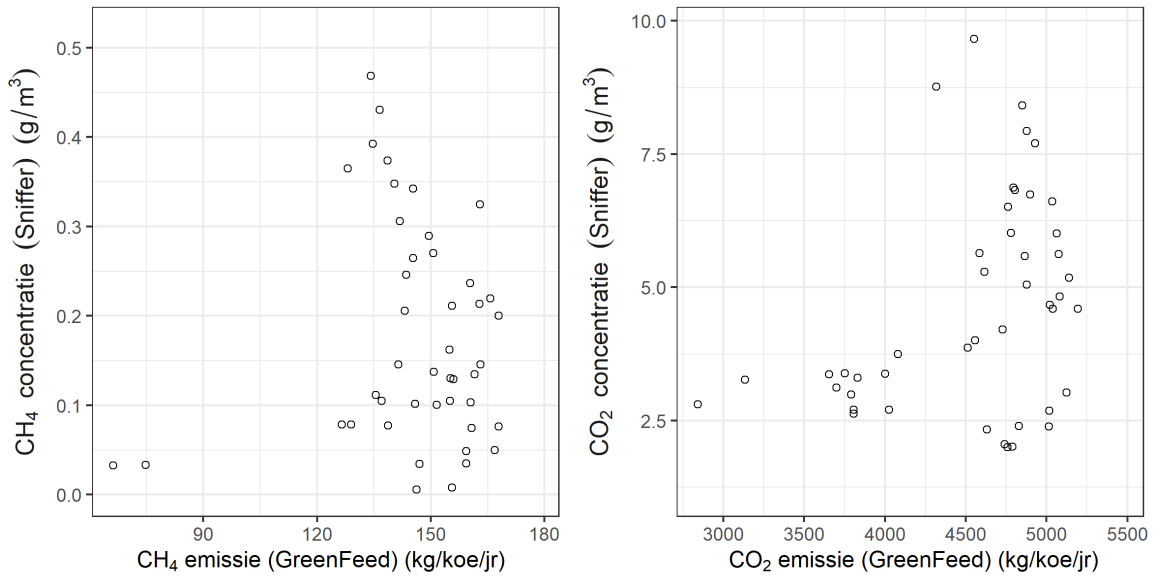
Figuur 5.11 CH_4 emissie gemeten op stalniveau [kg/koe/jr] tegen de enterische CH_4 emissie gemeten met de GreenFeed [kg/koe/jr] op bedrijf 19. De periodes van drie weken waarop GreenFeed en Stal tegelijk hebben gemeten, zijn onderscheiden op basis van kleur.

Tabel 5.2 Gemiddelde stal- en GreenFeed-emissie van bedrijf 19 gedurende drie perioden waarin beide meetmethodes zijn toegepast. De berekende mestbijdrage is verkregen uit het verschil tussen gemeten Stal- en GreenFeed-emissie.

Periode	Stal [Kg/Koe/jr]	GreenFeed [Kg/Koe/jr]	Berekende mestbijdrage [Kg/Koe/jr]
P1 februari 2019	201 ± 34	150 ± 6	51 ± 34
P2 september 2019	139 ± 23	116 ± 16	23 ± 45
P2 maart 2020	127 ± 24	161 ± 33	-34 ± 36

Sniffer – GreenFeed

Zoals bepaald uit eerdere resultaten (paragraaf 5.2.1), zijn Sniffer-metingen betrouwbaarder wanneer de waarden worden gemiddeld over een langere periode. **Figuur 5.12** laat de weekgemiddelde waarden zien voor die periodes waar de Sniffer en GreenFeed met elkaar vergeleken kunnen worden. Er is geen consistent patroon te vinden tussen beide meetmethodes op weekbasis en absoluut niveau.



Figuur 5.12 Weekgemiddelde CH₄ (links) en CO₂ (rechts) emissies (GreenFeed) en concentraties (Sniffer).

5.2.4 Rangschikken van koeien en bedrijven

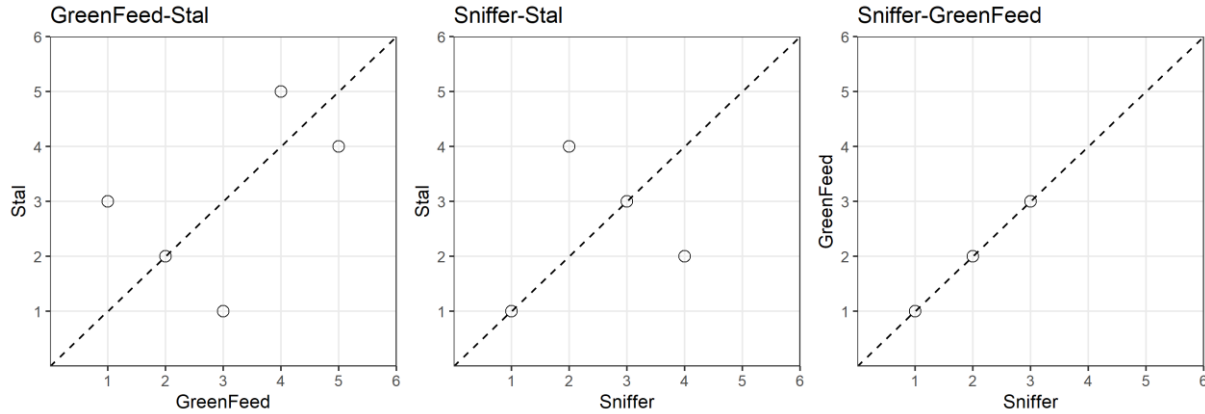
Figuur 5.13 laat de rangschikking van koeien zien op basis van GreenFeed of Sniffer, voor drie bedrijven. Hierbij is de rangschikking gedaan waarbij alleen dagen zijn genomen waar zowel GreenFeed als Sniffer metingen zijn. De rangschikking op basis van Sniffer of GreenFeed lijkt niet gecorreleerd te zijn. Dit kan betekenen dat emissie niet gerelateerd is aan concentratie. De inconsistente rangschikking kan ook het resultaat zijn van het feit dat de metingen niet op hetzelfde moment van de dag plaatsvinden, of dat de meetmethodes niet met elkaar overeenstemmen.



Figuur 5.13 Rangschikking van koeien op basis van GreenFeed (>30 dagen aan data; op de y-as) en Sniffer (>270 dagen aan data; op de x-as), per bedrijf met een lineaire regressielijn inclusief standaard error. Het totale aantal koeien verschilt per bedrijf.

Sniffer – GreenFeed – Stal

De vergelijking van de drie meetmethodes op basis van het rangschikken van bedrijven is weergegeven in **Figuur 5.14**. Het is te zien dat de rangschikking verschilt per meetmethode, maar de beperkte overlap tussen meetmethodes op dezelfde bedrijven maakt het niet mogelijk om hier een algemene uitspraak over te doen. Een mogelijk verschil in rangschikking kan worden veroorzaakt door het aantal dagen waarop gemeten is: er is per bedrijf meer Staldata beschikbaar dan GreenFeed-data. Desondanks komen, op basis van deze beperkte dataset, Sniffer en GreenFeed het best met elkaar overeen in het rangschikken van bedrijven.



Figuur 5.14 Rangschikking van bedrijven m.b.v. de drie meetmethodes. Links: Stal vs. GreenFeed op 5 bedrijven, midden: Stal vs. Sniffer op 4 bedrijven en rechts: GreenFeed vs Sniffer op 3 bedrijven. De onderbroken lijn geeft weer waar de rangschikking door beide meetmethodes gelijk zou zijn.

6 Koppelen met externe bronnen

6.1 Materiaal en methode

In het Methaan Data Lake zijn ook meteorologische data opgenomen. Deze data zijn verkregen via Akkerweb en omvatten luchtvochtigheid, temperatuur, windsnelheid, zonnestraling (stralingsintensiteit van de zon). Akkerweb verzameld deze gegevens. Door de coördinaten van de bedrijfslocatie in te voeren, zijn er locatie-specifieke meteorologische data verkregen. Deze data zijn vervolgens gekoppeld aan de emissie- en concentratiedata. Omdat meteorologische data veel fluctuaties kennen binnen een dag (temperatuur, luchtvochtigheid) en tussen dagen (windsnelheid), zijn de analyses uitgevoerd op basis van weekgemiddelden. Evenals voor de Sniffer- en Stalmetingen, is er een jaarpatroon bepaald op basis van weekgemiddelden tussen september 2019 en september 2020.

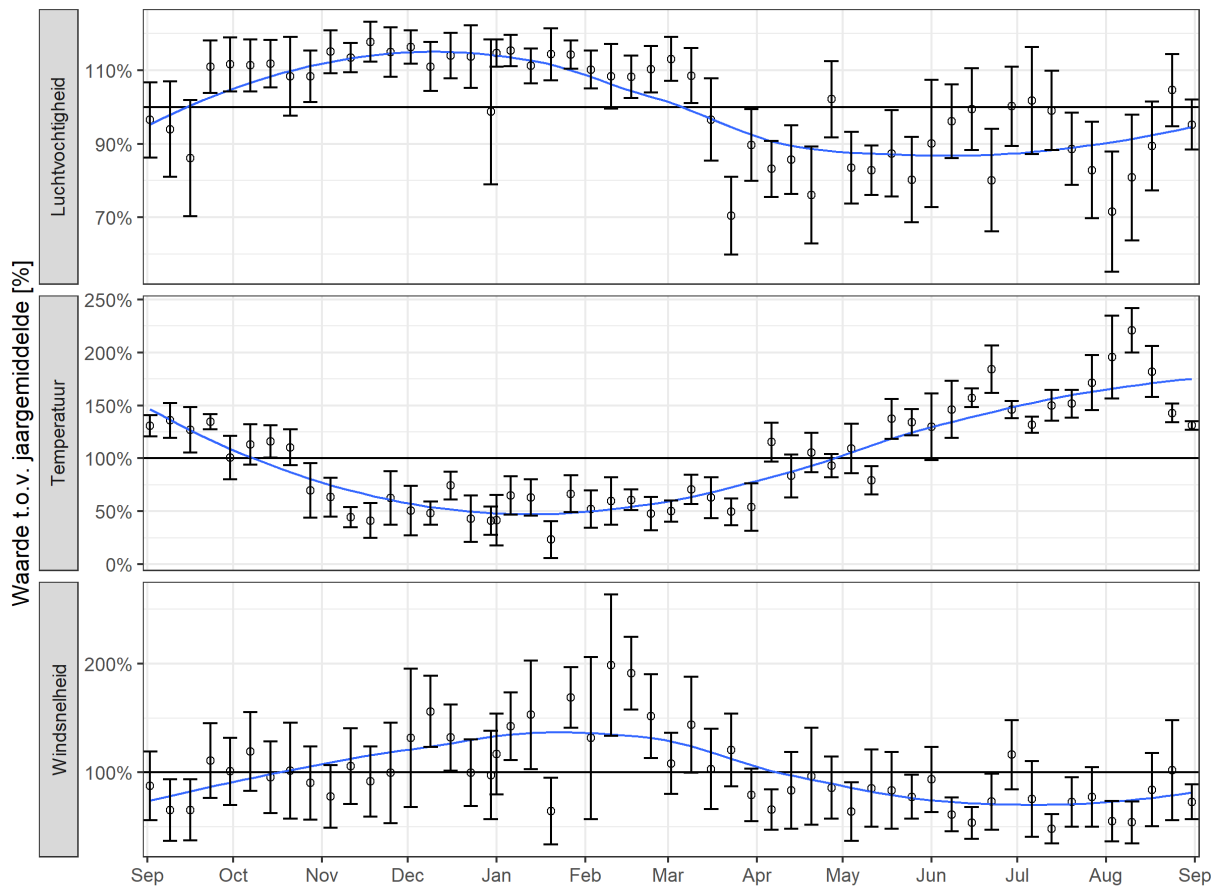
6.2 Resultaten en discussie

6.2.1 Jaarpatroon

Het jaarpatroon gebaseerd op weekgemiddelden van luchtvochtigheid, temperatuur en windsnelheid in de periode van september 2019 tot september 2020 is gevisualiseerd in **Figuur 6.1**. **Tabel 6.1** laat zien welke correlaties er gevonden zijn op basis van meteorologische weekgemiddelden en CH₄-emissies (Stal) en CH₄-concentraties (Sniffer). Voor het jaarpatroon op basis van de GreenFeed zijn geen significante correlaties gevonden. Dit komt voornamelijk doordat de GreenFeed gedurende drie weken op een bedrijf staat en daarmee weinig inzicht geeft in het jaarpatroon. Een significante correlatie met het jaarpatroon van de Sniffer data is gevonden voor luchtvochtigheid en deze correlatie is in dezelfde orde grootte als voor de stalmetingen. Voor Stal zijn alle correlaties met luchtvochtigheid, temperatuur en windsnelheid significant. Een correlatie geeft echter nog geen causaal verband aan en daardoor kunnen hier nog geen nieuwe inzichten aan worden ontleend, daar is nader onderzoek voor nodig. Wel is de negatieve correlatie tussen temperatuur en CH₄-emissie ook aangetoond door Ngwabie et al (2011). In dat onderzoek is ook aangetoond dat de activiteit van de dieren vermindert bij een hogere temperatuur. Een hogere emissie gedurende de winter zou daarom ook het effect kunnen zijn van toegenomen dierlijke activiteit.

Tabel 6.1 Correlatiecoëfficiënten tussen de jaarpatronen in luchtvochtigheid, temperatuur, windsnelheid en de jaarpatronen van de meetmethodes GreenFeed, Sniffer en Stal. Zonnestraling is niet significant.

	GreenFeed	Sniffer	Stal
Luchtvochtigheid	n.s.	0.51	0.53
Temperatuur	n.s.	n.s.	-0.56
Windsnelheid	n.s.	n.s.	0.66

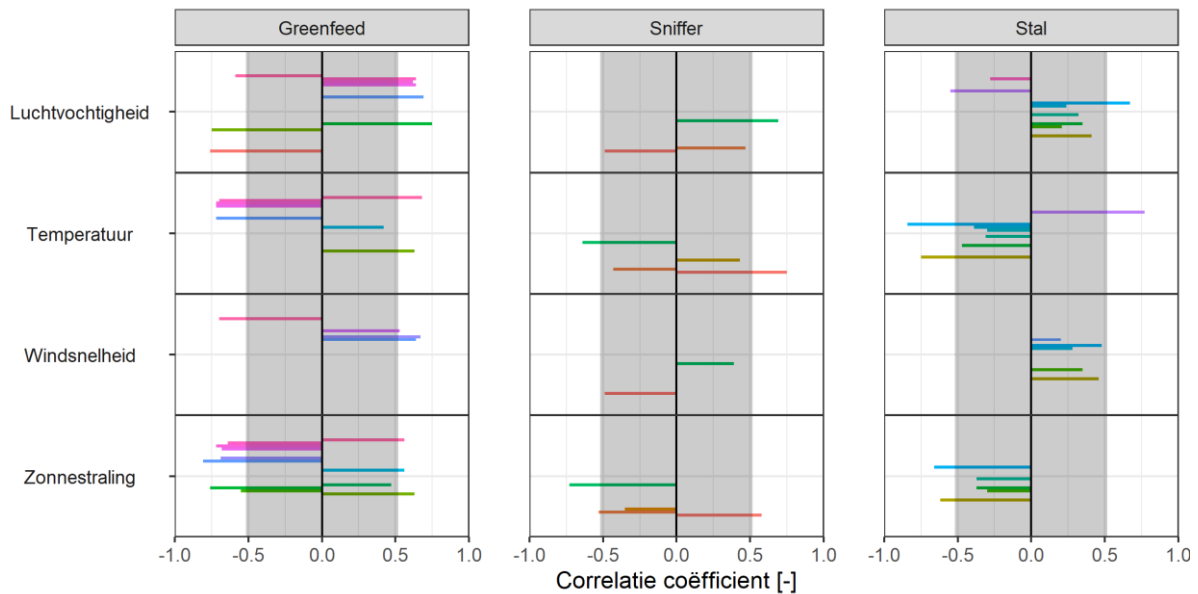


Figuur 6.1 Jaarpatroon op basis van de weegemiddelde luchtvochtigheid, temperatuur en windsnelheid in de periode van september 2019 tot september 2020. De standaarddeviatie rondom de uurgemiddelden geeft de spreiding binnen een week weer. Zonnestraling is niet significant.

6.2.2 Correlatie per bedrijf

Correlaties in jaarpatronen per bedrijf tussen de CH₄-emissie (GreenFeed, Stal), CH₄-concentratie (Sniffer) en luchtvochtigheid, temperatuur, windsnelheid en zonnestraling zijn weergegeven in **Figuur 6.2**. Het valt op dat de correlatiecoëfficiënten van GreenFeed en Sniffer meer verschillen tussen bedrijven dan de correlatiecoëfficiënten van de Stalmetingen. Een mogelijke verklaring ligt in de beschikbaarheid van de data. GreenFeed-data zijn in perioden van drie weken gemeten op een bedrijf, waardoor de correlaties zijn bepaald op korte tijdsreeksen en er verschillende beelden kunnen ontstaan. Van de Sniffer- en Stalmetingen zijn wel langere reeksen beschikbaar, van minimaal 270 dagen.

Hoewel de correlaties tussen Stalmetingen en weergegevens veelal marginaal zijn (<0.5), is de richting van de correlatiecoëfficiënten meer constant. Verschillen tussen bedrijven kunnen voortkomen uit de uitvoering van het stalsysteem. Voorbeelden van oorzaken zijn de aanwezigheid van ventilatoren of de grootte van luchtinlaatopeningen. Hier is echter geen informatie over beschikbaar in het Methaan Data Lake.



Figuur 6.2 Correlatiecoëfficiënten per bedrijf (kleur) voor vier weeggemiddelde weerfactoren en weeggemiddelde CH₄-emissie (GreenFeed; links), weeggemiddelde CH₄-concentratie (Sniffer; midden) en weeggemiddelde CH₄-emissie (Stal; rechts). Het grijsgekleurde oppervlak geeft de spreiding aan waarbinnen een correlatiecoëfficiënt kleiner is dan 0.5.

7 Valideren bestaande modellen

Het hebben van de mogelijkheid om CH₄ te modelleren op bedrijfsniveau is waardevol, omdat het kan helpen bij besluitvorming omtrent management door boeren, het opstellen van beleid en het beantwoorden van onderzoeksvragen. De grootste beperking van het modelleren is dat het model slechts een schatting van de werkelijkheid is, omdat het niet met alle factoren rekening houdt. De KringloopWijzer (**KLW**; van Dijk et al., 2020) gebruikt gegevens van de melkproductie, voer en de veestapel (bijvoorbeeld aantal melkkoeien) om de enterische- en mest- CH₄-emissies van een bedrijf te voorspellen. Binnen dit huidige onderzoek zijn CH₄-metingen van de drie verschillende methodes (zie **Hoofdstuk 3**) gebruikt om de voorspellingen van de KLW te vergelijken met gemeten waardes.

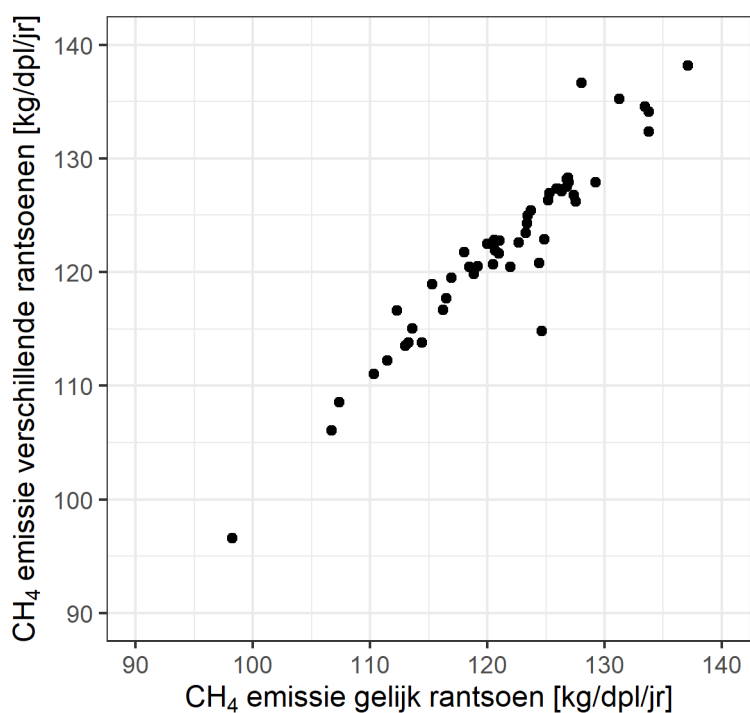
7.1 Materiaal en methode

7.1.1 Modelleren met de KringloopWijzer

In de KLW worden de CH₄-emissies op jaarbasis voor een melkveebedrijf voorspeld. Boeren leveren informatie aan over hun bedrijf met betrekking tot de voersamenstelling, de samenstelling van de veestapel (aantal koeien, pinken en kalveren), de begin- en eindvoorraad van het voer, de melkproductie, vet- en eiwitpercentages van de melk en het aantal dagen dat de koeien in de stal of het weiland hebben gestaan. Deze informatie wordt normaliter als totaalwaarden per jaar ingevoerd, maar op bedrijf 9 was al deze informatie op weekbasis beschikbaar voor het jaar 2019. In dit onderzoek hebben we de berekeningen van de KLW gebruikt om de CH₄-emissie op *wekelijkse* basis te voorspellen. In de KLW op jaarbasis wordt een voorspelling gedaan van de CH₄-emissie van zowel pens- en darmfermentatie (ongeveer 80% van de totale CH₄-emissie) alsook de CH₄-emissie uit de mest (20%). In de rekenmodule die gebruikt is binnen deze studie is enkel de CH₄ uit pens- en darmfermentatie berekend.

Eén van de beperkingen van de formules in de KLW is dat het niet mogelijk is om verschillende rantsoenen toe te kennen aan de verschillende diercategorieën (koeien, pinken en kalveren). In de berekening op jaarbasis wordt aangenomen dat de pinken en kalveren hetzelfde rantsoen krijgen als de melkgevende dieren, maar in een lagere hoeveelheid. In werkelijkheid is er meer variatie in de rantsoensamenstellingen tussen diergroepen: melk producerende koeien krijgen bijvoorbeeld ander krachtvoer en in andere hoeveelheden dan het jongvee. Op jaarbasis maakt dit niet veel verschil, zeker wanneer de melkgevende koeien de overgrote meerderheid van de CH₄ produceren en pinken en kalveren in aanzienlijke kleinere aantallen aanwezig zijn. Echter, op weekbasis kan dit een effect hebben op de vergelijking van voorspelde en gemeten CH₄ en daarom was het model drie keer toegepast om de afzonderlijke rantsoenen goed mee te nemen; één keer voor iedere diergroep (model 1 koeien, model 2 pinken en model 3 kalveren) met het specifieke bijbehorende rantsoen voor die diergroep. De som van de CH₄-emissies per diergroep geeft dan een voorspelling van de CH₄-emissie op stalniveau. Voor droogstaande dieren is bij de vergelijking met CH₄-emissies op stalniveau aangenomen dat deze hetzelfde rantsoen hebben als de melkgevende dieren.

Er was een sterke correlatie tussen het modelleren van een enkel rantsoen en het modelleren van de drie afzonderlijke rantsoenen voor koeien, pinken en kalveren (**Figuur 7.1**), maar er waren ook enkele verschillen, met name in weken met vers gras. De schattingen per week voor het jaar 2019 zijn geüpload naar het Methaan Data Lake.



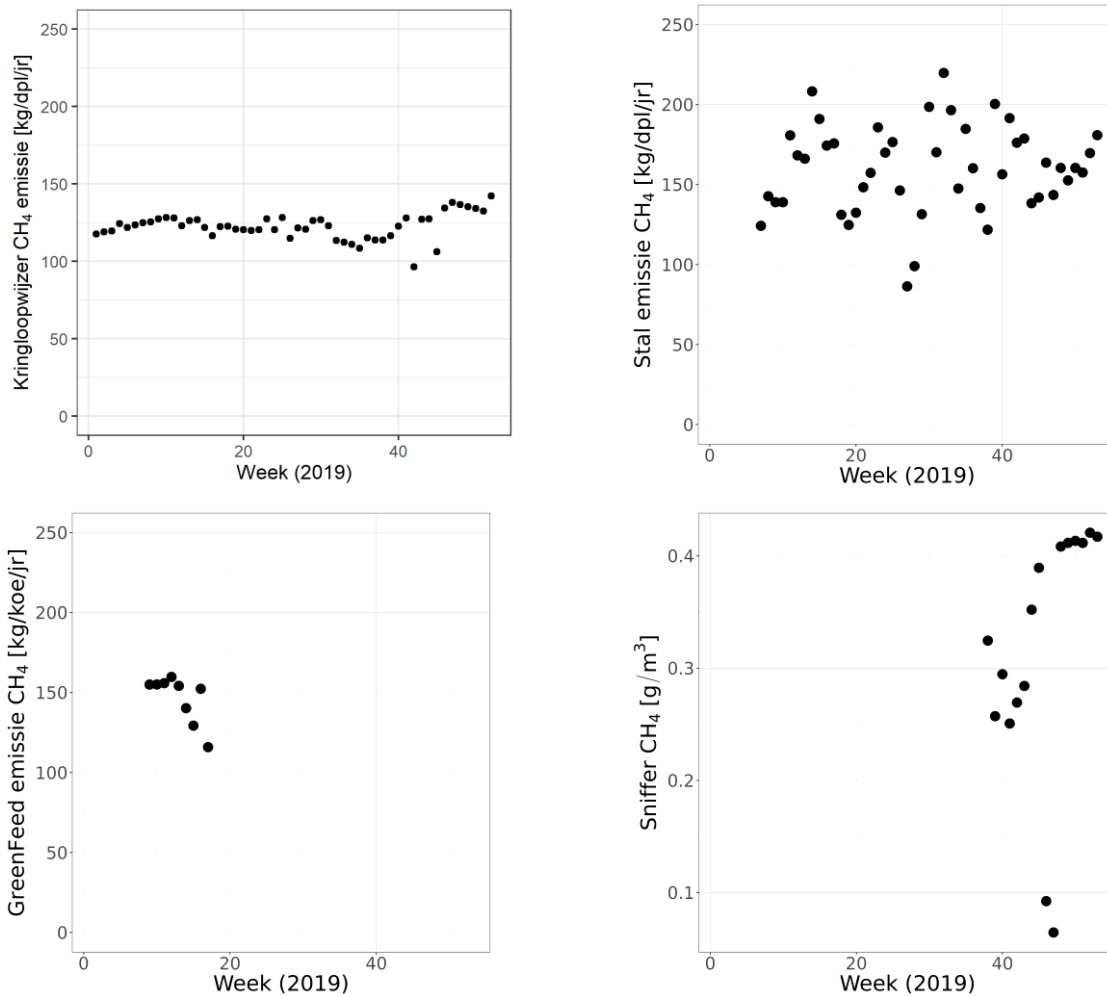
Figuur 7.1 Methaanemissies wanneer gemodelleerd met een enkel gemiddeld rantsoen (x-as) of met driediergroep-specifieke rantsoenen (y-as).

7.1.2 Vergelijking tussen de voorspelde emissies van de KringloopWijzer en daadwerkelijke methaanmetingen

Voor bedrijf 9 zijn de K LW-schattingen op wekelijkse basis gemaakt over het jaar 2019. In datzelfde jaar waren er ook metingen van de Stal, de GreenFeed en de Sniffer (**Figuur 5.1**). Al deze gegevens zijn beschikbaar in het Methaan Data Lake, zoals eerder beschreven (**Hoofdstuk 4**). Er waren twee weken waarin er inconsistenties waren in de voermetingen met betrekking tot vers gras en grazen, met mogelijke onderschattingen en dubbele tellingen, dus deze twee weken zijn verwijderd en corresponderen met de twee laagste metingen uit de K LW (**Figuur 7.2**, linksboven). Van de Stalmetingen waren er 47 weken beschikbaar die konden worden vergeleken met de K LW. Om dit te bewerkstelligen zijn de voorspelde K LW CH₄-emissies omgezet naar kilogram CH₄ uitgestoten per dierplaats (kg/dierplaats/jr), zodat de eenheden hierdoor direct vergelijkbaar waren tussen beide methoden (**Figuur 7.2**, rechtsboven). De GreenFeed heeft in totaal negen weken gemeten op bedrijf 9 (**Figuur 7.2**, linksonder) en heeft alleen CH₄-emissies van een deel van de melkgevende koeien gemeten (ongeveer 50% van de totale melkgevende groep op het bedrijf). Om een 'eerlijke' vergelijking tussen GreenFeed en K LW te kunnen maken, is de analyse alleen met melkgevende koeien uitgevoerd. Daarmee zijn de GreenFeed-metingen op koe niveau vergeleken met voorspelde K LW-emissies op dierniveau. Er waren vier weken met lagere CH₄-emissies van de GreenFeed, corresponderend met weken waarin de koeien toegang tot de weide hadden. Daarom is er een vergelijking gemaakt mét en zonder weidegangweken in de dataset. De vergelijking tussen de Sniffer en K LW was slechts beperkt mogelijk, omdat de Sniffer alleen de CH₄-concentraties heeft gemeten en deze gegevens niet kunnen worden omgerekend naar emissies. Hierdoor waren de meeteenheden tussen de K LW en de Sniffer niet direct vergelijkbaar. Daarnaast heeft de Sniffer gedurende zestien weken data verzameld (**Figuur 7.2**, rechtsonder). Dit is meer dan de GreenFeed, maar nog steeds beperkt. Twee van deze weken bevatten uitschieters waar de CH₄-meting lager was dan het verwachte achtergrondniveau van de omgeving.

Om voorspelde en gemeten CH₄-niveaus te kunnen vergelijken, zijn visualisaties gemaakt en correlaties berekend. Voor elke meetmethode is, ter visualisatie, de gemeten CH₄-emissie of -concentratie uitgezet tegen voorspelde K LW CH₄-emissie. De correlatie van de ruwe data geeft een indicatie van de relatie tussen de methoden. Er is een lineaire regressie van gemeten CH₄ op voorspeld CH₄ uitgevoerd en de r² is gebruikt

om de 'goodness-of-fit' te bepalen. Tot slot zijn de gemiddelden \pm standaarddeviaties vergeleken als indicatie voor hoe de langeretermijnmetingen zich verhouden tot gemodelleerde KLW CH₄-emissies.



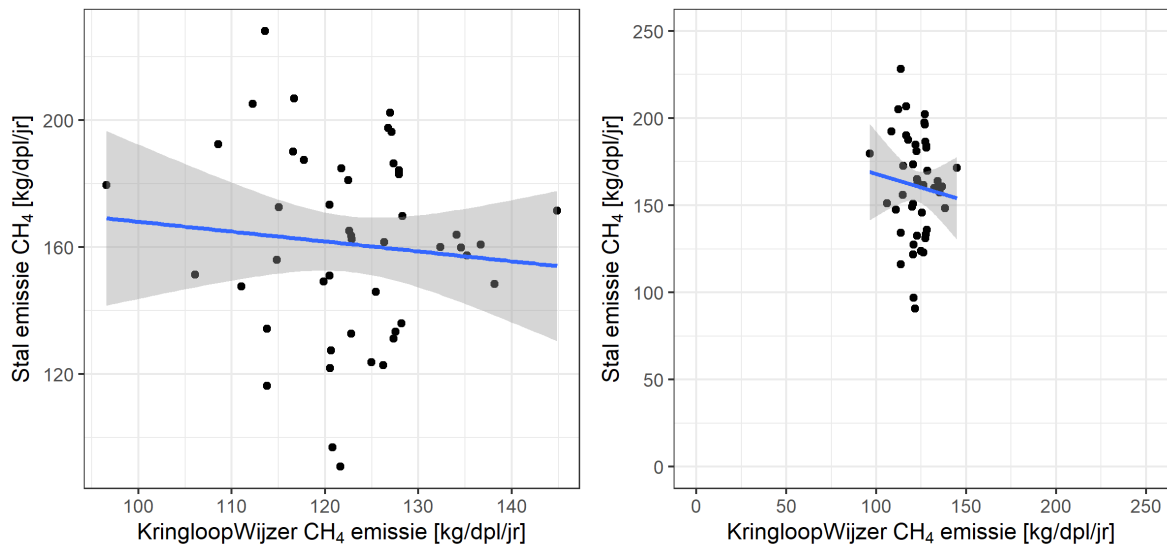
Figuur 7.2 Methaanemissies wanneer gemodelleerd met de KLW (linksboven), of wanneer gemeten met het Stalmeetsysteem (rechtsboven), de GreenFeed (linksonder) of Sniffers (rechtsonder).

7.2 Resultaten en discussie

Het eerste opvallende resultaat is de minimale variatie in voorspelde KLW CH₄-emissies over een jaar (**Figuur 7.2**, linksboven). De minimale variatie in voorspelde emissies is het gevolg van een minimale variatie in de inputvariabelen. Het rantsoen van de dieren was consistent over de weken in 2019 met als grootste verschil het aandeel vers gras in het rantsoen. De momenten waarop meer variatie wordt geobserveerd in de voorspelde KLW CH₄-emissie, vallen samen met de weken waarin de dieren vers gras gevoerd kregen of toegang hadden tot de weide. Naast rantsoen, zijn ook dieraantallen belangrijke inputvariabelen in het model en ook deze variabele is nagenoeg constant. Deze beperkte variatie betekent ook dat het lastig is om een sterke correlatie te vinden met de meer variabele gemeten emissies. Voor huidige toepassing van de KLW is de minimale variatie echter geen probleem aangezien het gaat om schattingen op jaarniveau.

7.2.1 Vergelijking tussen de KringloopWijzer en Stalmetingen

Er is een zwakke relatie tussen voorspelde KLW CH₄-emissie en gemeten CH₄-emissie van de Stalmetingen gevonden (**Figuur 7.3**, links). Deze zwakke relatie komt waarschijnlijk door de beperkte variatie tussen weken voor de KLW in vergelijking met het Stalmeetsysteem (**Figuur 7.3**, rechts): de KLW heeft metingen van 97-142 kg/dierplaats/jr (over de weken die overlappen met de Stalmetingen) en de Stalmetingen variëren tussen 91-228 kg/dierplaats/jr.



Figuur 7.3 *Vergelijking van methaanemissies voorspeld met KLW of gemeten met het Stalmeetsysteem, met de schaal passend gemaakt (links) en met gelijke schalen (rechts) en een lineaire regressie voor beide (in blauw).*

Met de zwakke relatie tussen voorspelde KLW CH₄-emissie en de Stalmetingen is het te verwachten dat de Pearsoncorrelatie (-0.10 ± 0.15) en r^2 waarde (nul) ook laag zijn (**Tabel 7.1**). Dit geeft aan dat het model op weekniveau geen goede voorspeller is van de Stalmetingen doordat het gemeten variatie niet verklaart. Dit door onder andere de beperkte variatie in de modelinvoer-parameters.

Tabel 7.1 *Indicatoren van overeenkomst tussen de KLW en Stal.*

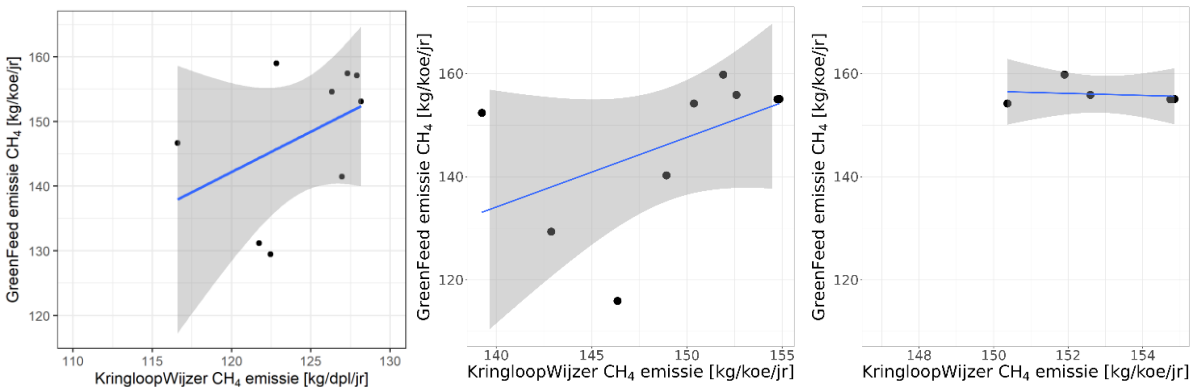
Indicator	Waarde
Pearsoncorrelatie	-0.10 ± 0.15
r^2	Nul
Gemiddelde Stal [kg/dpl/jr]	161 ± 29
Gemiddelde Kringloopwijzer [kg/dpl/jr]	123 ± 8

Tabel 7.1 laat ook de gemiddelde CH₄-emissies zien zoals voorspeld met de KLW (123 ± 8 kg/dpl/jr) en gemeten in de stal (161 ± 29 kg/dpl/jr). Zichtbaar wordt dat op jaarbasis de KLW in de stal hoger is dan voorspeld. Een belangrijke verklaring hiervoor is dat de KLW enkel een voorspelling maakt van enterisch CH₄ en daarmee de mestbijdrage buiten beschouwing laat terwijl via de Stalmetingen zowel de enterische CH₄-emissie alsook de CH₄-emissie gerelateerd aan mestopslag gemeten wordt. De verwachting is echter dat de CH₄-emissie als gevolg van mestopslag relatief klein is op bedrijf 9 aangezien hier bijna alle mest direct wordt verwerkt via vergisting. Wel bestaat er een vermoeden dat de mestvergister CH₄ lekt wat ertoe kan leiden dat de Stalmetingen de CH₄-emissie te hoog weergeven. Dit wordt onderzocht in een ander project.

Het verschil tussen gemeten stalemissie en voorspelde emissie kan ook betekenen dat 1) er ondanks de snelle afvoer naar de vergister toch een mestbijdrage aanwezig is en 2) dat het model de CH₄-emissie onderschat. Om deze twee hypothesen te toetsen is vervolgonderzoek vereist met bij voorkeur ook meerdere bedrijven om een aanname te toetsen.

7.2.2 Vergelijking tussen de KringloopWijzer en GreenFeed

Aangezien de GreenFeed alleen meet aan melkgevende dieren, zijn de GreenFeed metingen vergeleken met een K LW-voorspelling van alle diercategorieën (koeien, pinken, kalveren) én een K LW-voorspelling van alleen de melkgevende koeien. Tussen de GreenFeed metingen en de K LW voorspelling op basis van alle diercategorieën is een zwakke relatie gevonden (**Figuur 7.4**, links). Het vergelijken van de GreenFeed metingen met de K LW voorspelling van alleen de melkgevende dieren resulteerde in een iets sterkere relatie (**Figuur 7.4**, midden). In de weken dat de koeien in de weide liepen, zijn er lagere CH₄-emissies gemeten met de GreenFeed en was er meer variatie in CH₄-emissie. Wanneer deze weken met weidegang niet meegenomen werden in de vergelijking, was er geen relatie tussen de GreenFeed en K LW meer aanwezig (**Figuur 7.4**, rechts en $r^2 = 0.03$; **Tabel 7.2**).



Figuur 7.4 Vergelijking van methaanemissies wanneer voorspeld met de K LW of gemeten met GreenFeed, met alle dieren in de K LW (links), alleen melkgevende koeien in de K LW (midden) en observaties met weidegang verwijderd (rechts), met een lineaire regressie toegepast (in blauw).

Er is een positieve correlatie tussen de voorspelde K LW-emissies en de gemeten GreenFeed-emissies (**Tabel 7.2**), wat aangeeft dat er een relatie is tussen beide methoden. Wanneer alle dieren worden meegenomen in de K LW is deze correlatie zwak (0.42 ± 0.34). Wanneer er een vergelijking met alleen melkgevende koeien wordt gemaakt, wat eerlijker is omdat ook de GreenFeed alleen metingen doet bij melkgevende dieren, is deze correlatie iets hoger (0.49 ± 0.33). Eenzelfde zien we wanneer we kijken naar de 'goodness-of-fit': deze is lager wanneer alle dieren meegenomen worden ($r^2 = 0.18$), dan wanneer alleen melkgevende koeien worden meegenomen ($r^2 = 0.24$). Echter, deze waarden zijn nog steeds laag. Dit geeft aan dat niet veel van de variatie in CH₄-metingen wordt verklaard door voorspelde K LW CH₄-emissie.

De geschatte enterische emissie met de K LW komt echter sterk overeen met de gemeten enterische emissie met de GreenFeed wanneer er gemiddeld wordt over alle weken waarin gemeten is. Dit geeft aan dat er goede overeenstemming is tussen beide methoden voor emissiebepaling over een langere periode. In de analyses waar de metingen voor koeien in de weide achterwege zijn gelaten, was er een afname in correlatie, r^2 en een afname in de overeenstemming tussen de methoden. Dit geeft aan dat het verwijderen van deze metingen waarschijnlijk niet gepast is. Er zijn slechts negen weken met GreenFeed-metingen en het verwijderen van deze (weide) metingen betekent dat de helft van alle data wordt genegeerd. Daarnaast worden hierdoor juist de weken waarin er de meeste variatie is, in zowel de K LW- als de GreenFeed-resultaten, achterwege gelaten.

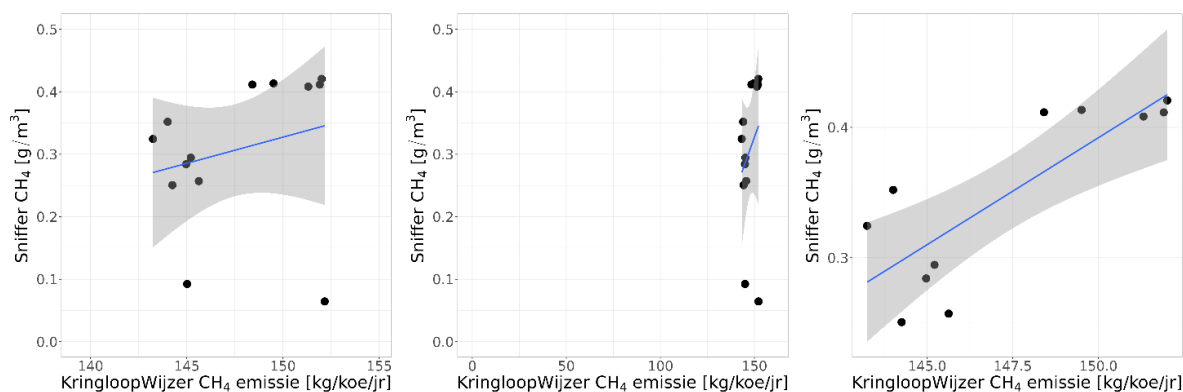
Tabel 7.2 Indicatoren van overeenkomst tussen de K LW en GreenFeed.

Indicator	Waarde		
	Alle dieren voorspeld met K LW vs koeien GreenFeed	Koeien voorspeld met K LW vs koeien GreenFeed	Koeien voorspeld met K LW vs koeien GreenFeed (zonder weidegang)
Pearsoncorrelatie	0.42 ± 0.34	0.49 ± 0.33	-0.16 ± 0.57
r^2	0.18	0.24	0.03
Gemiddelde GreenFeed	146 ± 15 kg/koe/jr	146 ± 15 kg/koe/jr	156 ± 2 kg/koe/jr
Gemiddelde Kringloopwijzer	124 ± 4 kg/dpl/jr	146 ± 8 kg/koe/jr	152 ± 3 kg/koe/jr

Op basis van bovenstaande resultaten kan er geconcludeerd worden dat de K LW niet geschikt is voor het modelleren van korte tijdsperioden en de K LW niet goed in staat is de variatie tussen weken te voorspellen. Verder onderzoek naar hoe CH₄-emissie wordt voorspeld wanneer koeien in de weide zijn kan bevorderlijk zijn, gezien dit onderzoek de indicatie laat zien dat deze resultaten momenteel wordt overschat door de K LW.

7.2.3 Vergelijking tussen de KringloopWijzer en Sniffers

De twee belangrijkste resultaten uit de vergelijking tussen de K LW en Sniffer zijn: 1) het is belangrijk om de Snifferdata correct te verwerken (zie hieronder) en 2) er lijkt een sterke relatie tussen de twee methodes te zijn. Deze relatie moet echter met enige nuances worden geïnterpreteerd. De Sniffer meet namelijk enkel concentraties en geen emissies, hoewel de verwachting wel is dat een hogere CH₄-concentratie resulteert in een hogere CH₄-emissie (Garnsworthy *et al.* 2019). In dit onderzoek is er een zwakke tot gematigde relatie tussen de K LW en Sniffer gevonden (**Figuur 7.5**, links). Echter, **Figuur 7.5** laat ook zien dat de variatie in gemeten CH₄ groter is dan in voorspeld CH₄ (**Figuur 7.5**, midden). Na het verwijderen van de veronderstelde uitschieters (twee metingen die lager waren dan het verwachtte achtergrondniveau), is er een ogenschijnlijk sterke relatie tussen de K LW en Sniffer te zien (**Figuur 7.5**, rechts). Dit lijkt echter hoofdzakelijk te komen door clustering aan het hoge en lage uiteinde van de data. Samenvattend lijkt het erop dat weken met lage voorspelde K LW CH₄-emissies overeenkomen met weken met lage gemeten CH₄-concentraties van de Sniffer, en hetzelfde geldt voor weken met hoge CH₄-emissies of -concentraties. Deze uitkomst moet echter met enige voorzichtigheid worden geïnterpreteerd, omdat de Sniffer metingen een langere periode nodig heeft voor een betrouwbare meting. Het suggereert wel de potentie voor verder onderzoek.



Figuur 7.5 Vergelijking van methaanemissies wanneer voorspeld met de K LW of gemeten met Sniffers, met de schaal passend gemaakt (links), met gelijke schalen (midden) en met uitschieters verwijderd (rechts), met een lineaire regressie toegepast (in blauw).

De analyse waarin uitschieters van de Sniffer meegenomen waren, had een lagere Pearsoncorrelatie (0.24 ± 0.29) en een lagere r^2 waarde (0.06) dan de Pearsoncorrelatie (0.82 ± 0.19) en r^2 waarde (0.66) uit de analyse waarin de uitschieters niet waren meegenomen (**Tabel 7.3**). Dit laat het belang van het correct verwerken van de Snifferdata duidelijk zien. Een directe vergelijking tussen de Sniffer en de K LW kan niet worden gemaakt, vanwege de verschillende eenheden.

Tabel 7.3 Indicatoren van overeenkomst tussen de KringloopWijzer en Sniffer voor alle data en de data zonder sniffer uitschieters.

Indicator	Waarde	
	Alle data	Zonder Sniffer uitschieters
Pearsoncorrelatie	0.24 ± 0.29	0.82 ± 0.19
r^2	0.06	0.66
Gemiddelde Sniffer (g/m ³)	434 ± 169	493 ± 97
Gemiddelde Kringloopwijzer (kg/koe/jr)	147 ± 4	147 ± 3

7.2.4 Hoe kunnen de resultaten van de vergelijking tussen voorspeld methaan van de KringloopWijzer en gemeten methaan van de drie meetmethodes worden gebruikt in de praktijk?

Het belangrijkste resultaat van dit onderzoek was niet specifiek de relaties tussen K LW en de verschillende meetmethodes zelf die zijn geobserveerd, maar het feit dat de vergelijking überhaupt gemaakt kon worden. Als de data uit verschillende projecten binnen de klimaatenvelop niet centraal verzameld zou zijn geweest in het Methaan Data Lake, had dit onderzoek niet uitgevoerd kunnen worden. Voordat er conclusies getrokken kunnen worden over de vergelijking tussen voorspelde K LW CH₄-emissie en gemeten CH₄-emissie of -concentratie, is het belangrijk om de beperkingen van elke methode toe te lichten. De K LW is ontwikkeld voor jaarlijkse CH₄-emissies en niet om de variatie tussen weken weer te geven. In de berekeningen omvatte het model CH₄ uit pens- en darmfermentatie. Het grootste deel van de variatie in CH₄-emissies in de K LW is afkomstig van veranderingen in het rantsoen, tussen weken met hoge verhoudingen krachtvoer en weken met weidegang. Een andere variabele die een effect heeft op de CH₄-emissie is het aantal dieren. Echter, in dit onderzoek was deze variabele relatief constant. De Stalmetingen lieten een grote variatie in CH₄-emissies zien, maar 1) er kan nog niet bepaald worden wat deze variatie veroorzaakt doordat het Methaan Data Lake nog geen andere data samenbrengt die de variatie kan verklaren, en 2) de Stalmetingen omvatten niet de CH₄ die wordt geproduceerd in de weide; hiervan wordt alleen een schatting gemaakt. De beperking van de GreenFeed is dat deze slechts voor een korte periode, en voor een klein aantal melkgevende koeien gemeten heeft. Ook de Sniffer heeft in dit geval een beperking wat betreft de korte periode dat deze op bedrijf 9 heeft gemeten. Daarnaast zijn de metingen van de Sniffer uitgedrukt in concentraties. Er zijn andere manieren verkend om voorspeld en gemeten CH₄ te vergelijken, maar door het gebrek aan variatie in voorspelde K LW CH₄-emissie zijn er geen andere passende methoden om te testen zonder aanvullende data.

Zelfs met deze beperkingen is er een aantal duidelijke aanbevelingen voor het vergelijken van voorspeld versus gemeten CH₄. Als de voorspelde K LW CH₄-emissie vergeleken zal worden met echte metingen in de praktijk, is het belangrijk om de juiste diercategorieën te vergelijken. Als er variatie is die wordt veroorzaakt door een bekende factor (zoals weidegang), verwijder deze data dan niet, omdat deze waardevol is. Het regelmatig maken van vergelijkingen is nuttig voor het identificeren van factoren die kunnen bijdragen aan verbetering van de K LW, zoals betere voorspellingen voor weidegang. Het zou ook kunnen worden gebruikt om punten te identificeren waarop de meetmethodes (GreenFeed, Sniffer of Stal) zouden kunnen worden verbeterd. Informatie uit de K LW kan gebruikt worden om de datamodellen van de meetmethodes te verbeteren, waarbij bijvoorbeeld input uit de K LW als systematische effecten of als correcties gebruikt kunnen worden. Hierdoor zouden data (van bijvoorbeeld de Sniffer) van tevoren bewerkt kunnen worden om zo de data-analyse te optimaliseren. Met de analyses die mogelijk waren met de beperkte hoeveelheid data, is er enige overeenstemming tussen voorspelde K LW CH₄-emissie en de Sniffer en GreenFeed gevonden. Dit geeft enig vertrouwen dat de modellen en metingen beide naar behoren werken, maar er is nog ruimte voor verbetering voor elk van de methoden. Met name de vraag over de mestbijdrage blijft staan. Alleen tussen de K LW schatting en de Stalmeting is namelijk geen overeenstemming gevonden. De mestbijdrage is hiervoor de meest waarschijnlijke verklaring aangezien zowel de toegepaste als de GreenFeed en Sniffer enterisch CH₄ bepalen.

8 Algemene discussie en conclusie

Het doel van dit onderzoek was om een data architectuur te bouwen om verschillende CH₄-datastromen bijeen te brengen, te koppelen, en daar externe databronnen aan toe te voegen om 1) trends en variatie tussen bedrijven zichtbaar te maken, 2) de invloed van (beschikbare) externe factoren op CH₄-emissie vast te stellen, en 3) bestaande rekenregels die CH₄ voorspellen te toetsen met wat er op bedrijven daadwerkelijk gemeten wordt.

In dit onderzoek hebben we laten zien dat het mogelijk is om verschillende datastromen samen te brengen in een Cloud-based Data Lake, een modern datawarehouse concept voor integratie, opslag en analyse van data. Werken met een Data Lake geeft een grote flexibiliteit in de dataverwerking en -analyse en kan tegen relatief weinig kosten een grote dataset (schaalbaar tot enkele petabytes) worden opgeslagen. Methaandata van drie verschillende sensoren (GreenFeed, Sniffer en Stal) gemeten op verschillende melkveebedrijven in Nederland zijn samengebracht in het Methaan Data Lake. Daarnaast is er een koppeling gemaakt met een externe databron, Farmmaps, om zeer lokale meteorologische gegevens aan het Methaan Data Lake toe te voegen.

Om de CH₄-data bijeen te brengen in het **Methaan Data Lake** zijn verschillende pipelines en APIs ontwikkeld. De meeste van deze pipelines werken nog niet automatisch, maar vergen een handmatige aanvraag (moeten handmatig áangezet worden). Echter, de ontwikkelde pipelines zijn relatief eenvoudig te automatiseren in de toekomst door gebruik te maken van Azure Functions. Hier kunnen de Python-pipelines, ontwikkelt in dit onderzoek, eenvoudig in worden geïntegreerd. De verbinding met een externe databron via API is zeer geschikt om andere data, niet specifiek ontwikkelt om een onderzoeksvraag te beantwoorden, te verkrijgen. De scripts voor de data analyse die ontwikkeld zijn in dit onderzoek zullen beschikbaar blijven zodat de analyses frequenter en herhaaldelijk uitgevoerd kunnen worden als meer data in het Methaan Data Lake geïntegreerd wordt. **Hiermee heeft dit onderzoek de basis gelegd van een modern methaan datawarehouse.** Het huidige Methaan Data Lake is dus nog relatief klein, en gevuld met een gelimiteerde hoeveelheid data. Hierdoor kon er in dit onderzoek weinig verklarend onderzoek verricht worden om zo inzicht te krijgen in de geobserveerde trends en variatie. Naar de toekomst toe is het van belang om deze Methaan Data Lake verder toe te passen en verder te vullen met meer én andersoortige data. Naar de toekomst toe kan deze centrale data opslag plek resulteren in efficiënte inzet van data en kennis om integraal reductiemaatregelen te ontwikkelen.

Uit de data-analyse blijkt dat de variatie tussen dieren op één bedrijf gemiddeld genomen nog groter is dan de variatie tussen bedrijven. Deze tussen-bedrijfsvariatie is vergelijkbaar met de variatie tussen dagen. Ook de variatie binnen dagen is groter dan de variatie tussen bedrijven, maar goed zichtbaar in het dagpatroon. In de variatie tussen dagen is een lange-termijn effect zichtbaar wat waarschijnlijk gerelateerd is aan de samenstelling van het rantsoen door het jaar heen en de hoeveelheid mest in de mestkelder (onder de roostervloeren). Er is geen directe verklaring van patronen uit meteorologische gegevens gevonden. Variatie tussen bedrijven blijkt nog steeds het meest lastig om te verklaren. De drie meetmethodes meten CH₄ op verschillende plekken en verschillende manieren. Uit dit onderzoek blijkt dat de drie verschillende meetmethodes niet volledig substitueerbaar zijn als het gaat om het rangschikken van koeien of van bedrijven op basis van de hoeveelheid CH₄-emissies. Op basis van dit onderzoek kan niet een methode als meest geschikt aangewezen worden, Dat was ook niet het doel van dit onderzoek, en zal hoogstwaarschijnlijk afhangen van de onderzoeksvraag. Echter, dit onderzoek heeft wel laten zien dat meetmethodes niet volledig substitueerbaar zijn. Dit schept ook mogelijkheden om te onderzoeken hoe ze elkaar in toekomstig onderzoek kunnen versterken. Als laatste zijn de CH₄-metingen vergeleken met voorspelde CH₄-emissie, waarbij we gebruik hebben gemaakt van een bestaand model, de kringloopwijzer (KLW). **Voor korteretermijnvoorspelling is het KLW-model niet geschikt, echter voor langeretermijnmetingen geeft KLW een goede voorspelling van de enterische CH₄-emissie.**

9 Dankwoord

We willen Lianne Koning, Julio Mosquera Losada, Anouk van Breukelen, Nico Ogink, Arie Klop, Gerjan Hilhorst, Hendrik Jan van Dooren en Michel de Haan bedanken voor het **delen** van de CH₄-data en KLV-model en de waardevolle discussies. Yvette de Haas, Dirkjan Schokker, Leon Šebek, Sifra Bol en Karin Groenestein willen we bedanken voor de waardevolle discussies. Daarnaast willen we Dirkjan Schokker bedanken voor het meehelpen met het opzetten van het Methaan Data Lake. Malou van der Sluis willen we bedanken voor met name het vertalen van hoofdstuk 7.

Literatuur

- Arthur, P.F., Barchia, I.M., Weber, C., Bird-Gardiner, T., Donoghue, K.A., Herd, R.M., & Hegarty, R.S. (2017). Optimizing test procedures for estimating daily methane and carbon dioxide emissions in cattle using short-term breath measures. *Journal of Animal Science*, 95(2), 645– 656. DOI: 10.2527/jas.2016.0700
- Bannink, A., Van Schijndel, M. W., & Dijkstra, J. (2011). A model of enteric fermentation in dairy cows to estimate methane emission for the Dutch National Inventory Report using the IPCC Tier 3 approach. *Animal feed science and technology*, 166, 603-618. DOI: [10.1016/j.anifeedsci.2011.04.043](https://doi.org/10.1016/j.anifeedsci.2011.04.043)
- Bell, M. J., Craigon, J., Saunders, N., Goodman, J. R., & Garnsworthy, P. C. (2018). Does the diurnal pattern of enteric methane emissions from dairy cows change over time? *Animal*, 12(10), 2065-2070. DOI: 10.1017/S1751731118000228
- Colucci, P. E., Chase, L. E., & Van Soest, P. J. (1982). Feed intake, apparent diet digestibility, and rate of particulate passage in dairy cattle. *Journal of Dairy Science*, 65(8), 1445-1456. DOI: 10.3168/jds.S0022-0302(82)82367-9
- De Haas, Y., Aldridge, M., & van Breukelen, A. (2021). Genetics of enteric methane emissions of Dutch dairy cows: Climate Envelop project 2019. (No. 1318). Wageningen Livestock Research. DOI: 10.18174/546572
- Garnsworthy, P.C., Difford, G.F., Bell, M.J., Bayat, A.R., Huhtanen, P., Kuhla, B., Lassen, J., Peiren, N., Pszczola, M., Sorg, D., Visker, M.H.P.W., & Yan, T. (2019) Comparison of methods to measure methane for use in genetic evaluation of dairy cattle. *Animals*, 9(10), p.837. DOI: 10.3390/ani9100837
- Hammond, K.J., Humphries, D.J., Crompton, L.A., Green, C., & Reynolds, C.K. (2015). Methane emissions from cattle: Estimates from short-term measurements using a GreenFeed system compared with measurements obtained using respiration chambers or sulphur hexafluoride tracer. *Animal feed science and technology*, 203, 41-52. DOI: 10.1016/j.anifeedsci.2015.02.008
- Hegarty, R. S. (2013). Applicability of short-term emission measurements for on-farm quantification of enteric methane. *Animal*, 7, 401-408. DOI: 10.1017/S1751731113000839
- Hofschreuder, P., Zhao, Y. Y., Aarnink, A. J. A., & Ogink, N. W. M. (2008). Measurement protocol for emissions of fine dust from animal houses: considerations, draft protocol and validation (No. 134). Animal Sciences Group. <https://edepot.wur.nl/778>
- How, M. (2020). The modern data warehouse in Azure: Building with speed and agility on Microsoft's Cloud platform. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-5823-1
- Jentsch, W., Piatkowski, B., & Derno, M. (2009). Relationship between carbon dioxide production and performance in cattle and pigs. *Archives Animal Breeding*, 52(5), 485-496. DOI: 10.5194/aab-52-485-2009
- Knapp, J. R., Laur, G. L., Vadas, P. A., Weiss, W. P., & Tricarico, J. M. (2014). Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *Journal of dairy science*, 97(6), 3231-3261. DOI: 10.3168/jds.2013-7234
- Koning, L., Van Riel, J. & Šebek, L. (2020). Enteric methane emission of the Dutch dairy herd; Average and variation of enteric methane emission among the Dutch dairy herd. (No. 1267). Wageningen Livestock Research. DOI: 10.18174/531257
- Lassen, J., Løvendahl, P., & Madsen, J. (2012). Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. *Journal of dairy science*, 95 (2), 890-898. DOI: 10.3168/jds.2011-4544
- Madsen, J., Bjerg, B. S., Hvelplund, T., Weisbjerg, M. R., & Lund, P. (2010). Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. *Livestock Science*, 129(1), 223-227. doi:<https://doi.org/10.1016/j.livsci.2010.01.001>
- Manafiazar, G., Zimmerman, S., & Basarab, J. A. (2016). Repeatability and variability of short-term spot measurement of methane and carbon dioxide emissions from beef cattle using GreenFeed emissions monitoring system. *Canadian Journal of Animal Science*, 97(1), 118-126. DOI: 10.1139/cjas-2015-0190

-
- Mosquera, J., van Dooren, H.J.C., Ogink, N.W.M., van Well, E.A.P., & Monteny, G.J. (2021). Monitoring van methaan-, ammoniak-, en lachgasemissies uit melkveestallen: Praktijkmetingen in de periode oktober 2018-februari 2020. (No. 1286). Wageningen Livestock Research. DOI: 10.18174/536752
- Ngwabie, N. M., Jeppsson, K. H., Gustafsson, G., & Nimmermark, S. (2011). Effects of animal activity and air temperature on methane and ammonia emissions from a naturally ventilated building for dairy cows. *Atmospheric Environment*, 45(37), 6760-6768. DOI: 10.1016/j.atmosenv.2011.08.027
- Orenstein, G., Doherty, C., Boyarski, M., & Boutin, E. (2017). Data warehousing in the age of artificial intelligence. O'Reilly Media, Inc. ISBN 9781491997956
- Pedersen, S., Blanes-Vidal, V., Heetkamp, M.J.W., & Aarnink, A.J.A. (2008). Carbon dioxide production in animal houses: A literature review. *Agricultural engineering international*, X(BC 08 008). <https://edepot.wur.nl/22415>
- RIVM. (2021). <https://www.rivm.nl/nieuws/definitieve-emissiecijfers-over-2019-bekend>. Bezocht op 6/12/2021.
- Ruysenaars, P.G., Coenen, P.W.H.G., Rienstra, J.D., Zijlema, P.J., Arets, E.J.M.M., Baas, K., Dröge, R., Geilenkirchen, G., 't Hoen, M., Honig, E., van Huet, B., van Huis, E.P., Koch, W.W.R., Lagerwerf, L.A., te Molder, R.M., Montfoort, J.A., Vonk, J., & van Zanten, M.C. (2020). Greenhouse gas emissions in the Netherlands 1990-2018: National Inventory Report 2020. National Institute for Public Health and the Environment (RIVM). DOI: 10.21945/RIVM-2020-0031
- Schep, C.A., Dooren, H.J.C. van, Mosquera, J., Well, E.A.P. van, Keuskamp, J.A., & Ogink, N.W.M. (2022). Monitoring van methaan-, ammoniak- en lachgasemissies uit melkveestallen: Praktijkmetingen in de periode oktober 2018 - oktober 2020. (No. 1388). Wageningen Livestock Research. <https://doi.org/10.18174/576706>
- Van Dijk, W., de Boer, J.A., de Haan, M.H.A., Mostert, P., Oenema, J., & Verloop, J. (2020). Rekenregels van de KringloopWijzer 2020; Achtergronden van BEX, BEA, BEN, BEP en BEC: actualisatie van de 2019-versie. (No. WPR-1023). Stichting Wageningen Research, Wageningen Plant Research, Business Unit Agrosystems Research. DOI: 10.18174/533882
- Zeeman, G. (1991). *Mesophilic and psychrophilic digestion of liquid manure*. Wageningen University and Research.
- Zimmerman, P. (2011). Method and system for monitoring and reducing ruminant methane production. US Pat. No. 7966971B2

To explore
the potential
of nature to
improve the
quality of life



Wageningen Livestock Research
Postbus 338
6700 AH Wageningen
T 0317 48 39 53
E info.livestockresearch@wur.nl
www.wur.nl/livestock-research

Wageningen Livestock Research ontwikkelt kennis voor een zorgvuldige en renderende veehouderij, vertaalt deze naar praktijkgerichte oplossingen en innovaties, en zorgt voor doorstroming van deze kennis. Onze wetenschappelijke kennis op het gebied van veehouderijsystemen en van voeding, genetica, welzijn en milieu-impact van landbouwhuisdieren integreren we, samen met onze klanten, tot veehouderijconcepten voor de 21e eeuw.

De missie van Wageningen University & Research is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen University & Research bundelen 9 gespecialiseerde onderzoeksinstituten van Stichting Wageningen Research en Wageningen University hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 30 vestigingen, 6.500 medewerkers en 10.000 studenten behoort Wageningen University & Research wereldwijd tot de aansprekende kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen verschillende disciplines vormen het hart van de unieke Wageningen aanpak.

