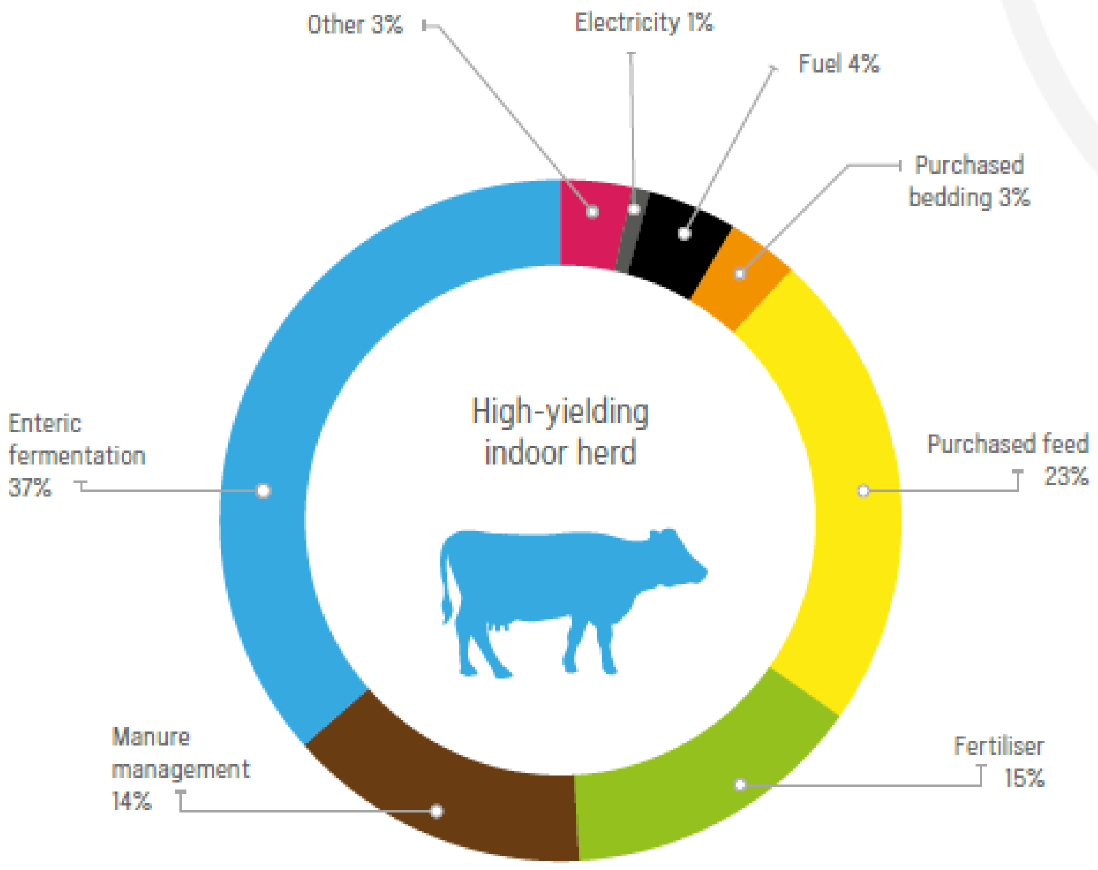
Modelling enteric methane emissions in UK dairy herds milking the farm data

Tom Chamberlain, MRCVS
MRes student

Results from preliminary data
(150 herds, 37k cows)

Methane – what's the problem?

- Methane is a powerful green house gas
  - COP26 (Glasgow, 2021)
  - Pledged 30% reduction by 2030
- Methane x86 more potent than CO2 (GWP*)
- Methane is 51% of carbon footprint on dairy unit
- Most (73%) methane from enteric fermentation



Contributions to carbon footprint for high yielding dairy herd (CIEL 2020)

Possible mitigation solutions

- Feed additives – NOP/Bovaer, Silvair, Rumitech
  - Available on international market, cost 1 – 2 ppl, antimicrobial
  - Also Seaweed, etc in development.
  - 10% to 30% reduction on emissions
- Genetics – 10 – 15 year to get national impact
- Others – fitted devices, boluses to time drug release, etc
- Generally
  - Limited on-farm testing, deployment, uptake – slow adoption
  - Who pays? private cost cf public good
  - These are NEW = Unforeseen problems (AMR, side effects, etc.)

Will improving efficiency help?

- More milk per cow per day of life
  - dilution of methane production impact
  - ? How big is this effect
- Look at lifetime production
  - Lifetime methane production / lifetime milk yield
- Look at variation between herds in UK national herd
  - Will show range that is achievable
  - Uses existing technologies
    - no development, implementation, uptake barriers and delays
  - Improving efficiency should improve farmer's gross margin (GM) and possibly profit.

The data source– milk recording data

- Three companies in UK NMR (52%) CIS (44%) QMMS (4%)
  - Data + milk collected monthly by company (impartial)
  - Uniform structure, complete (much can be verified through biology)
  - All adhere to international standards (ICAR)
  - Overall high quality, very complete data set
  - Biased towards ‘better’, more proactive farms
  - Data is freely available
- Full data set size
  - ~365 herds, 89K cows, 609M litres/year
- Preliminary dataset
  - ~168 herds, 40K cows

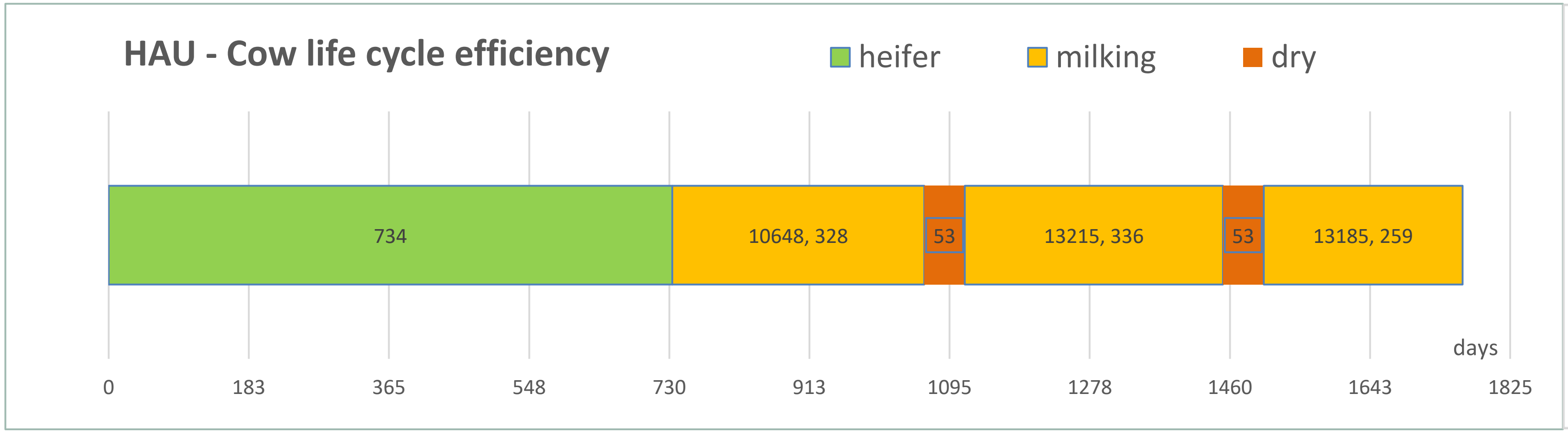
OEE - Overall Equipment Effectiveness measure of factory machine efficiency

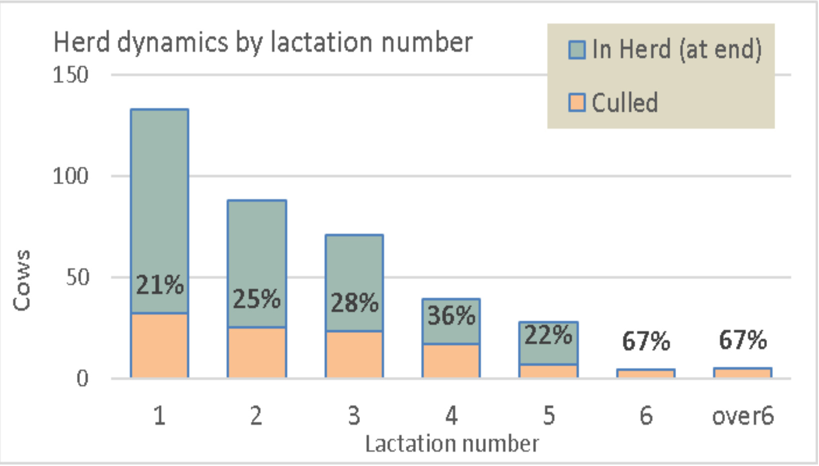
- Time + resources taken to build machine
- Level of production when working
  - Quality of product
  - Defective product
- Service intervals - down time
- Life span – production runs

- Age to first calving
- Lactation length and milk yield
  - Butterfat, protein, SCC
  - Milk discarded
- Length of dry period
- Lifespan – culling patterns

The data

- Births, deaths, calving, dry off dates
- Daily Milk yields + quality - recorded every month
- Can determine what a cow does every day of her life and when cows are culled
- OEE inspired approach
  - – Gold sections are productive, others 'non-productive'





Pre-analysis perceptions

- Methane output linked to
  1. Age at first calving (CEIL, 2020)- delays start of 'productive period'
  2. Milk yield – dilution of maintenance (and ration differences)
     1. Corrected for fat and protein
  3. Calving interval – dilutes time spend as non-productive dry cow
  4. Simple overall culling rate % of herd leaving each year
     1. Number of productive cycles
  5. Average age of cows at culling
     1. Better captures culling patterns
     2. Cows culled at L=1 more detrimental than culled at L=5

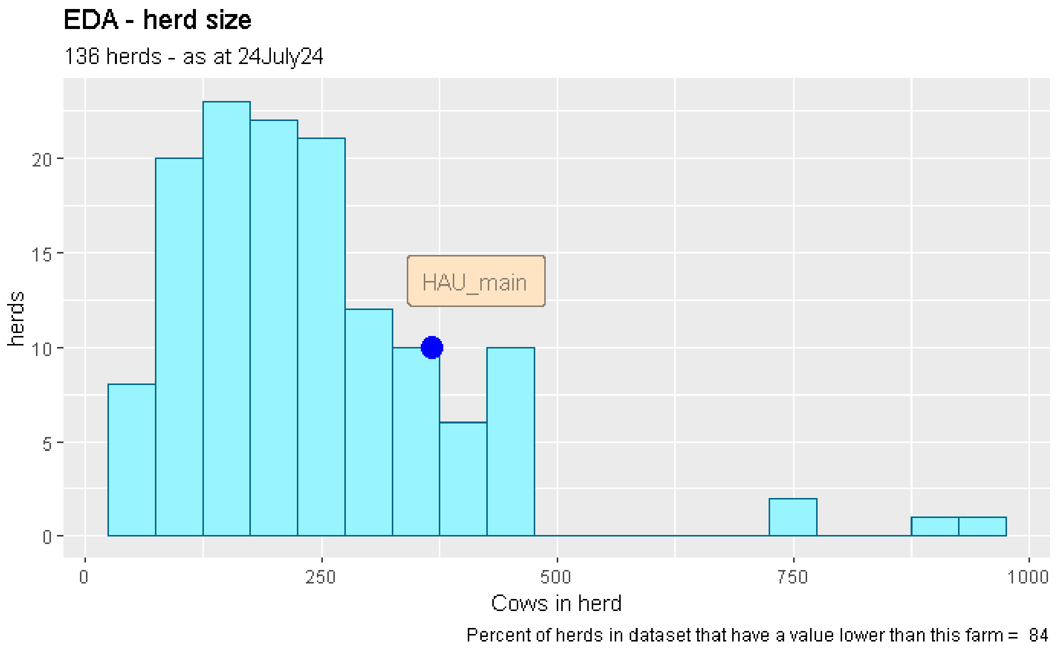
This is the base (vanilla, nvars=5) model

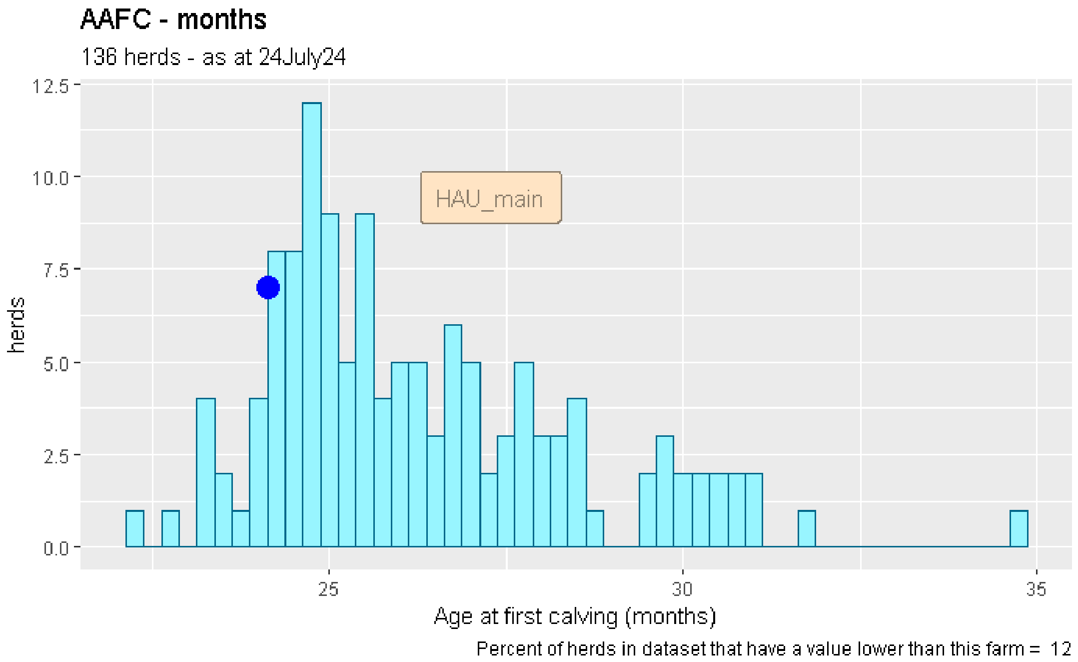
Modelling assumptions

- Model Domain = weaning (8wks) to leaving farm at culling
- Work at cow-life time scale
- Assume (over lifetime when include DLWG)
  - Energy requirements = energy supplied
- Energy (ME) modelled using AFRC 1990 (TCORN)
  - Covers all production stages – later models don't eg FiM
  - Can use a factorial approach (maintenance, DLWG, milk, pregnancy)
  - Calculate ME/day required for every day of life
- Predict methane from ME intake (Ellis, 2007)
  - CH_4 (MJ day) = 4.12 + (0.0901 x ME.intake) [MJ/day] for all cattle *RMSPE =28.2%*

Results - EDA

- Mean herd size 250 cows
  - Slightly larger than other data sources
- Age at first calving
  - Target = 24 months
  - many herds exceed target

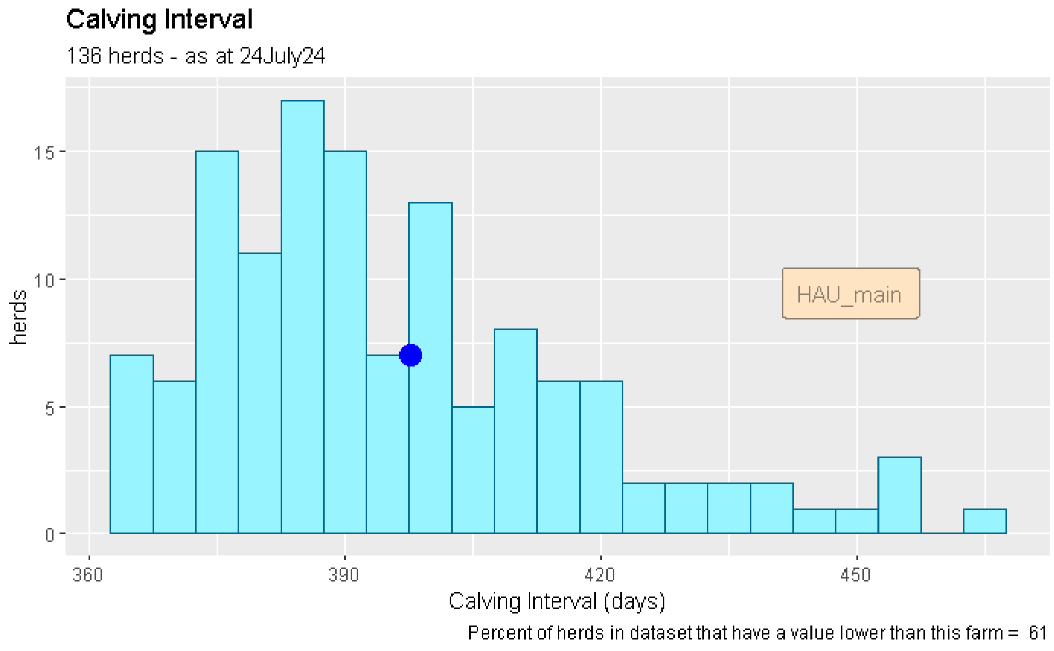


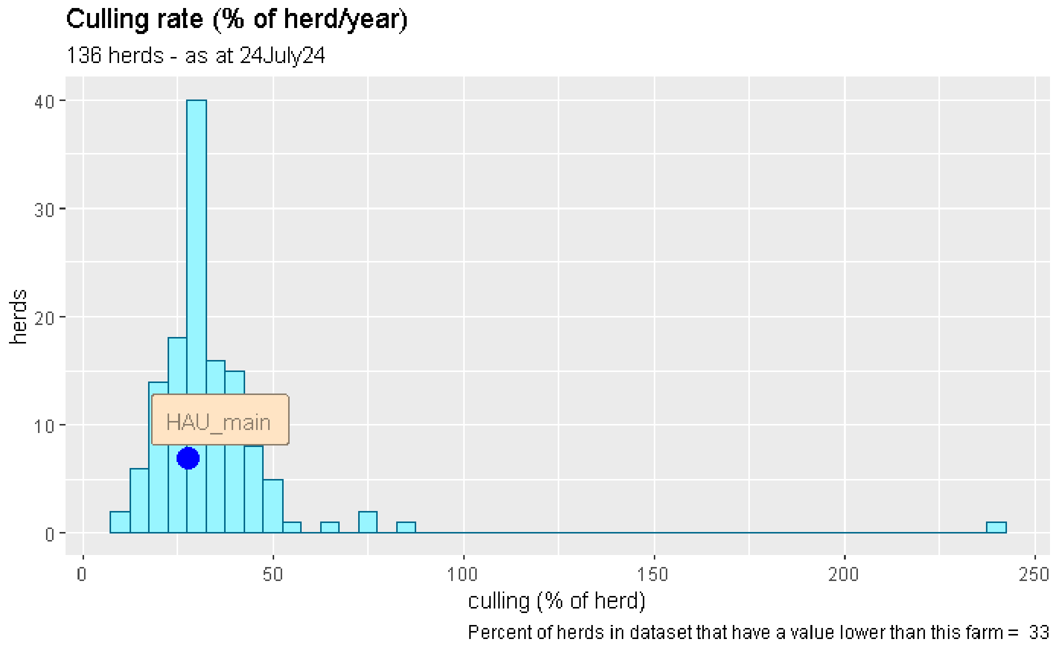


Results - EDA

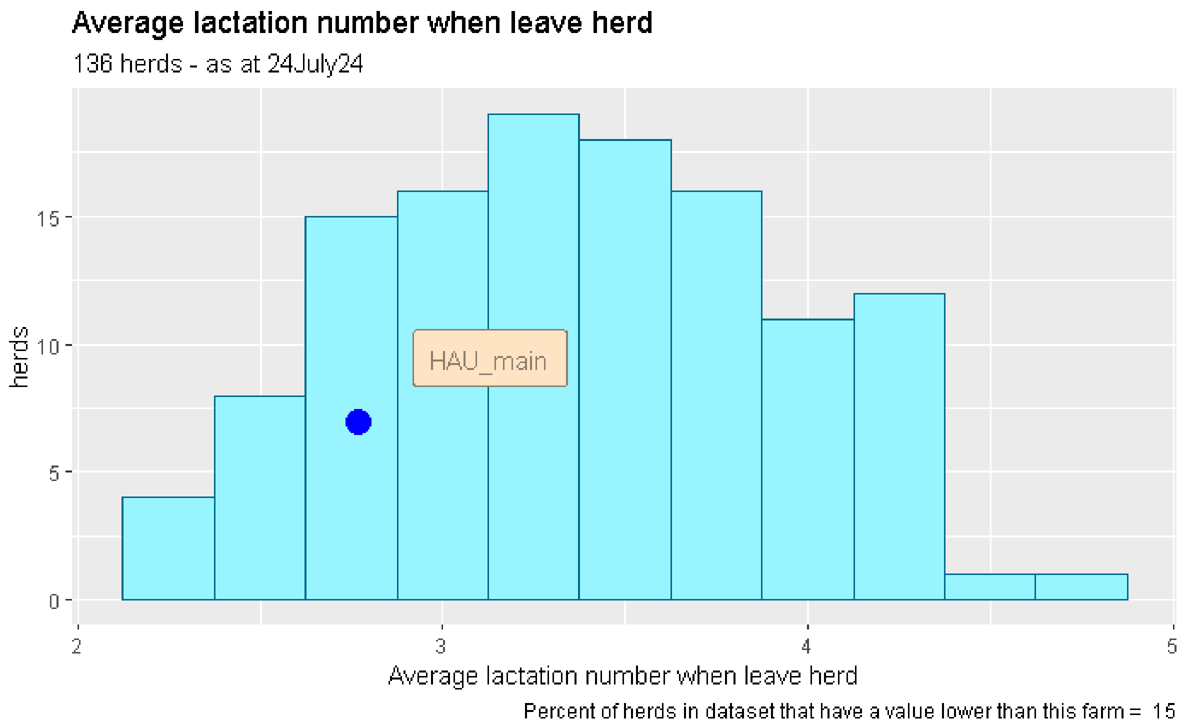
- Calving Interval
  - Generally higher than target (365 days)
  - Many tools available to control

- Culling rate
  - Some VERY high
    - ? TB, retirement, etc.
    - Pruned from data base
  - High butter fat also pruned
    - Probably Channel Island herds





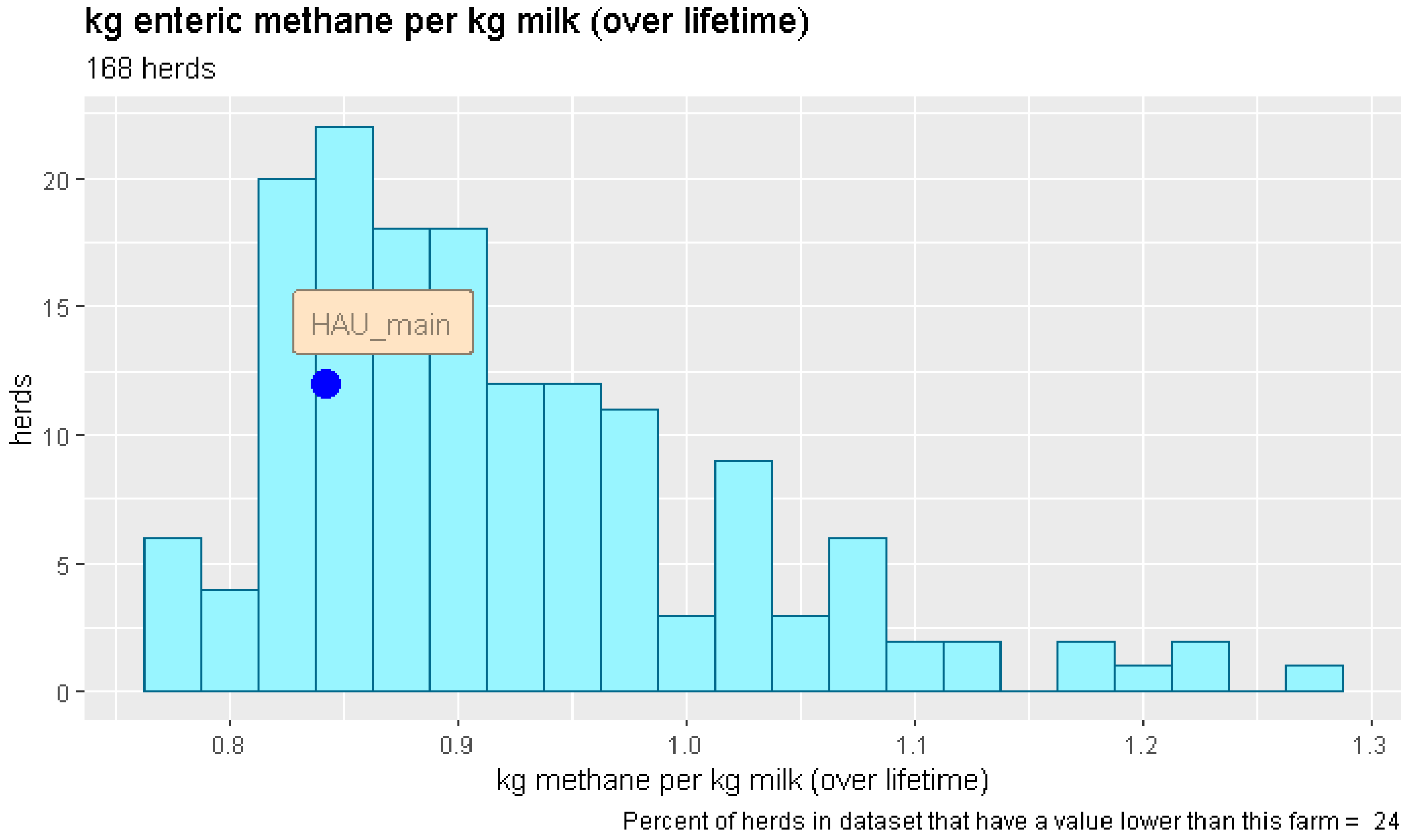
Results - EDA



Average lactation number when leave herd
136 herds - as at 24July24
herds
15
10
5
0
HAU_main
2
3
4
5
Average lactation number when leave herd
Percent of herds in dataset that have a value lower than this farm = 15

Modelled methane emissions

- Typical long left tail.
- IQR = 0.12 kg
  – 13% reduction
- IDR = 0.24 kg
  – 23% reduction
- Comparisons
  – Bovaer ~30%
  – Silvair ~ 10%
  – Rumitech ~10-15%
- Efficiency gains are in same ‘ball park’ and should be synergistic



Quantile	10%	25%	50%	75%	90%
kgCH4/kg FPCorr Milk	0.82	0.85	0.90	0.97	1.06

Predicting methane emissions
Vanilla model

- Simple regression [lm]
- Single run
- No validation(see later)
- Base line RMSE = 0.033
  - ? Can this be bettered
  - ?effect of noise in data
- Milk yield dominates
  - Can we look at residuals to model constant milk yields.
  - Cf ‘Residual feed intake’

Regression	F value	F df	P	Adj $R2$	RMSE
Annual FPCMY	449.4	1,152	<0.001	0.746	0.050
AAFC (months)	35.0	1,152	< 0.001	0.182	0.090
CI	1.07	1,152	0.303	0.000	0.099
Culling	3.22	1,152	0.075	0.014	0.099
avLact OnLeaving	1.00	1,152	0.317	0.000	0.099
All five	**235**	**5,148**	**<0.001**	**0.884**	**0.033**

Validating vanilla model

- Test:train [50:50]
  - 77 obs in each set
- Train set
  - $F = 121.8$ (df 5,71), $p<0.0001$, adj $R2 = 0.8882$, RMSE = 0.033
- Predict for test set (n=77)

Train:test validation

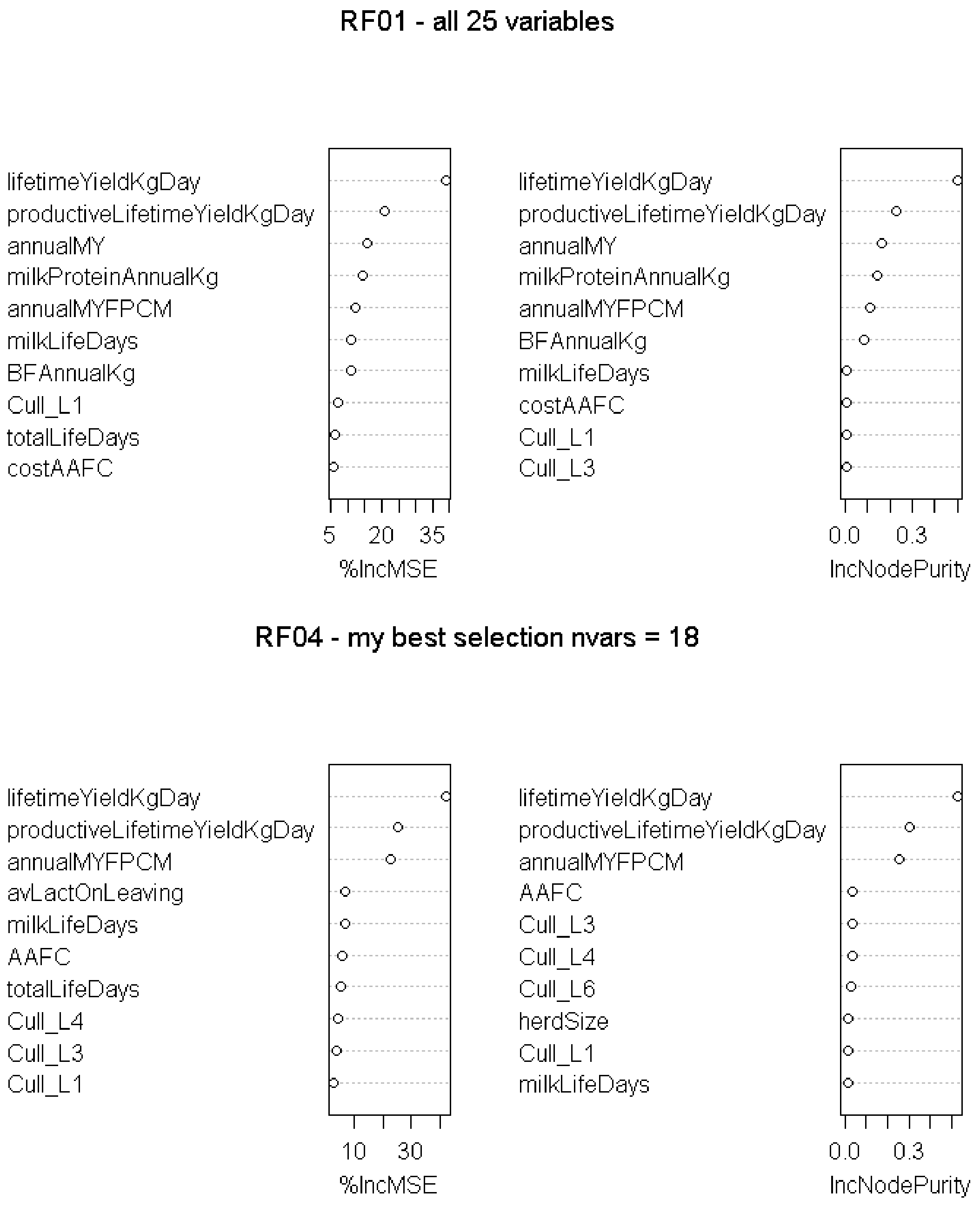
	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Mean
RMSE train	0.0333	0.0333	0.0333	0.0333	0.0333	0.0333	0.0333
RMSE test	0.0381	0.0334	0.0359	0.0302	0.0383	0.0343	0.0350
						Deterioration	5.2%

Random forest assessments (ntrees=1000)

1. The ‘kitchen sink’ data set (nvar = 25)
2. Prune out annual MY and some non-relevant vars (n=20)
3. Remove all yield variables (n = 15)
4. My best selection (n = 18)
5. Extended Vanilla data set (incl cull by lactation number) (n=12)
6. Vanilla set (n = 5)

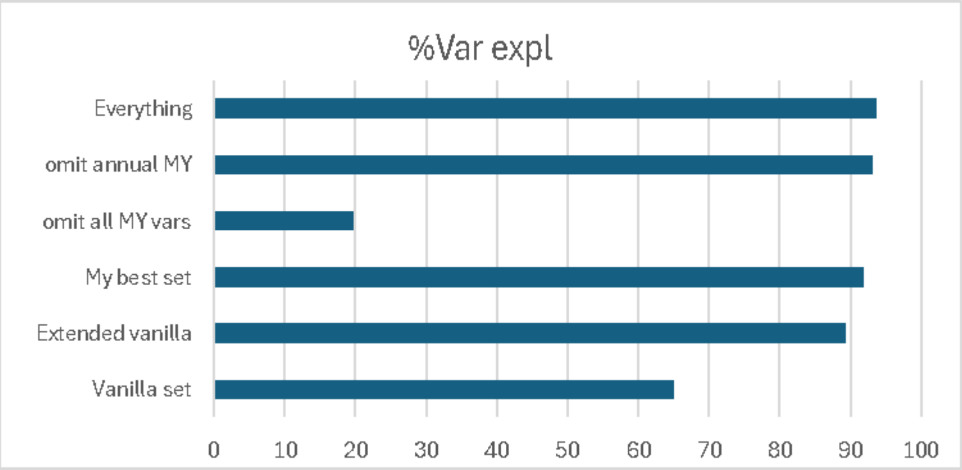
Assessing RF models

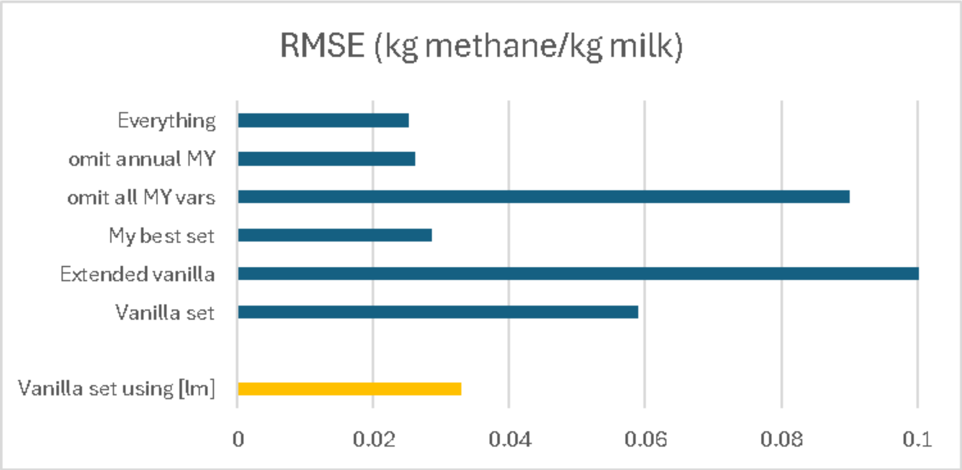
- Yield related variables dominate in both models
- 'best selection'
  - Removed some 'derived' variables
  - AAFC coming through
  - Detailed culling
- How to develop these into a stand alone model?



Random forest comparisons

- Big RF's did better
- Yield is important!
- Smaller RF's not as good as simple LM
- BUT: is LM overfitted
  - Se later
- How to deploy RF outside 'R'?
  - Need to create a deliverable





How to find the best model

- Want to apply model outside R
- Linear model – 5 variables. Train:test [50:50]

Train:test validation

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Mean
RMSE train	0.0333	0.0333	0.0333	0.0333	0.0333	0.0333	0.0333
RMSE test	0.0381	0.0334	0.0359	0.0302	0.0383	0.0343	0.0350
					Deterioration		5.2%

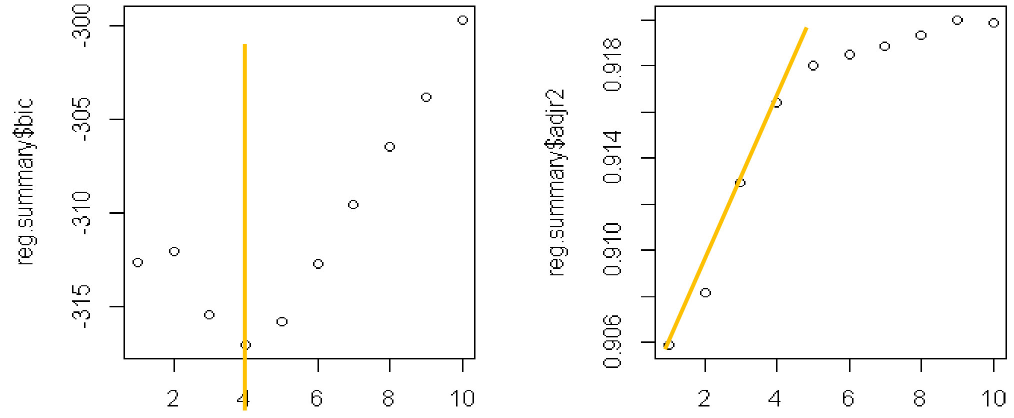
- Best subset – in progress

- Other ideas please

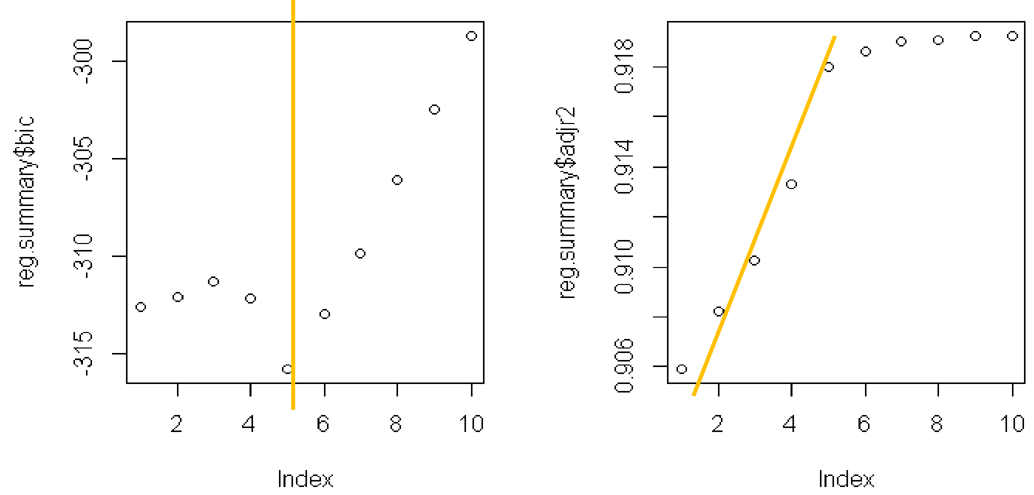
Best subsets

- Used 'best selection' data set
- BIC minimal and elbow at 4 or 5 variables
- Backward
  - annualMYFPCM
  - CI
  - avLactOnLeaving
  - lifetimeYieldKgDay
- Forward
  - annualMYFPCM
  - CI
  - avLactOnLeaving
  - lifetimeYieldKgDay
  - Cull_L3
- Cross validation [50:50] of 5 component model
  - Full data set
  - F=332 (df 5,148), adj R2=0.916, p<0.0001
  - RSME = 0.0284
  - Validation deterioration 11.5%

Backward



Forward



Train:test validation

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	**Mean**
RMSE train	0.0296	0.0264	0.0273	0.0280	0.0297	0.0256	**0.0278**
RMSE test	0.0277	0.0310	0.0310	0.0310	0.0289	0.0361	**0.0310**

Deterioration 11.5%

Further work

- Use full data base ~350 – 400 herds
- Other regression models, etc. but needs to be simple to apply
- Can put costs (GM ish) to changes in major KPI's
  - AAFC, CI, MY, culling rate
  - Relate GM cost savings to fall in methane production (kg/kg)
  - ? Get a win:win , profit positive situation.
- Develop farmer-friendly interface
  - *'improve this KPI and your CH4 footprint falls this much'*
  - Will this be seen (by farmers) as another 'farmer bashing' tool?
- Look at predicted eMethane at the cow level
  - Could have 40k cull records (11k so far from 100 herds)
  - Does this relate to genomic test results?
  - Can we selectively breed for cows with a low eCH4/kg milk score?
  - Develop a specific SNP-key for longevity, low methane potential?
  - ??Innovate UK consortium project