CRV, leading in health and efficiency

A Non-parametric approach to assessing cow performance

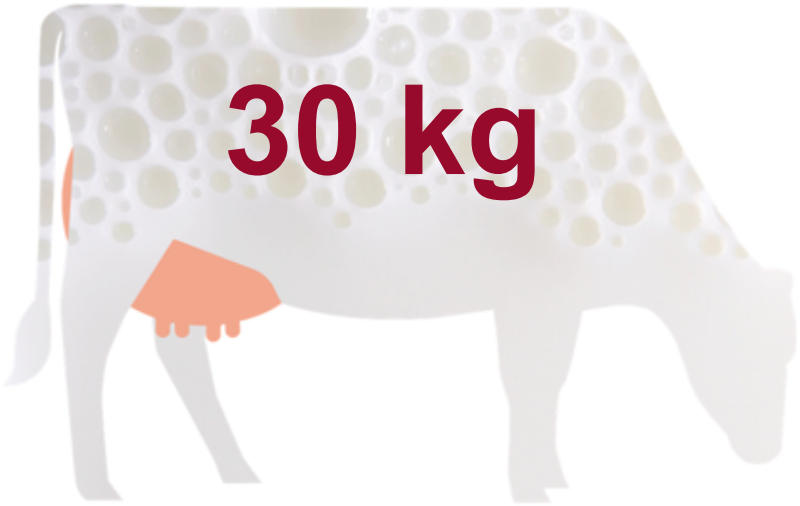
Ivo Lagarde | June 4 | 14:00 – 14:15 | CRV BV

CRV
BETTER COWS > BETTER LIFE

TS12: From Guidelines to the Barn: Implementing section 2 of ICAR's Milk Recording Guidelines – Updates and Lessons from the Field

What is the performance of this cow?

Test day milk yield:



What is her lactation stage?

200 days-in-milk (DIM)

What month did she calve?

December

What was her age at calving?

25 months

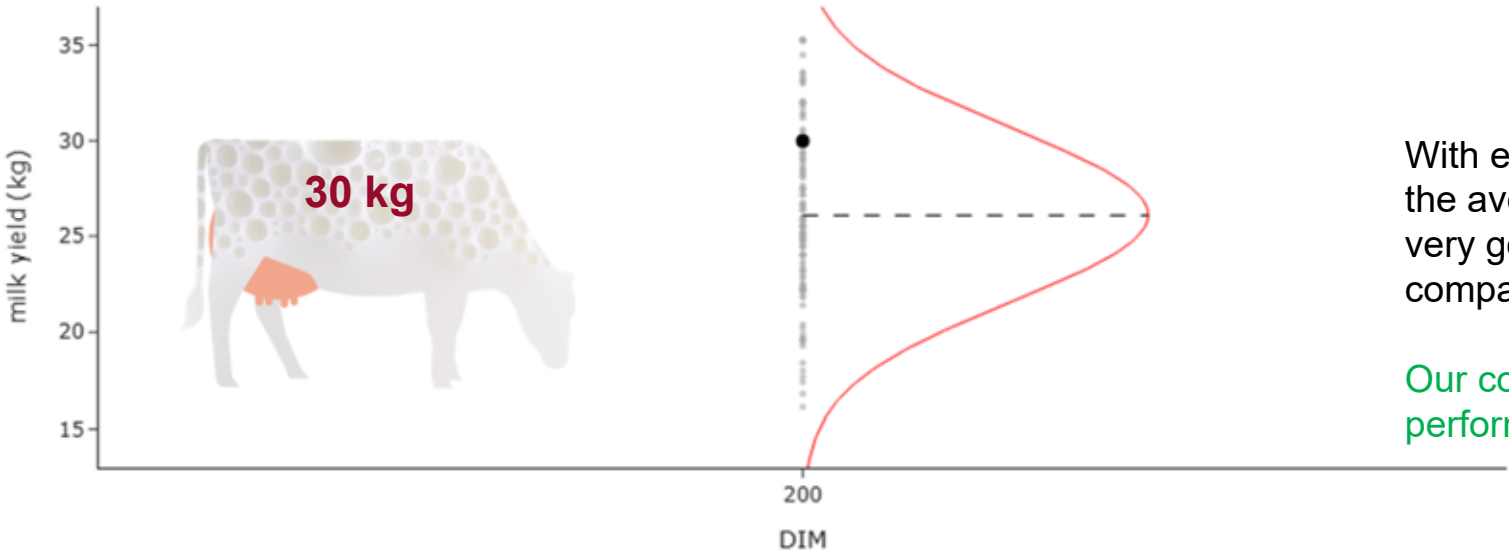
What is the overall level of the herd?

Average

CRV

BETTER COWS > BETTER LIFE

Lets visualise all test day milk yields for cows with the same systematic variables of the past 3 years:

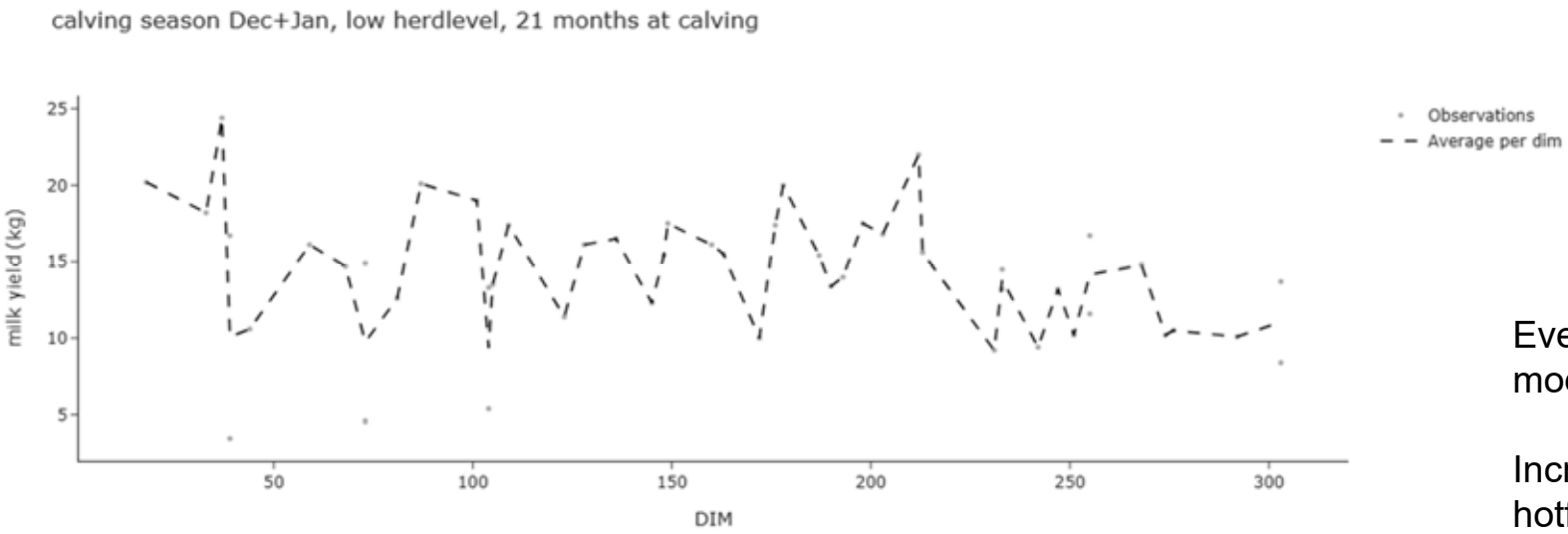


With enough observations the average is actually a very good benchmark to compare against.

Our cow with 30kg is performing well.

BETTER COWS > BETTER LIFE

This approach works well with a lot of data, but breaks down when observations become sparse.



Even parametric models will struggle

Increasing bins is a hotfix not a robust solution

BETTER COWS > BETTER LIFE

Looking back at past 3 years we at CRV
have over 16 million test day observations.

But spread over:

- 20 herdlevel classes
- 6 calving seasons
- 18 calving age classes
- 5 to 305 days-in-milk

20*6*18*300 = **648000 possible stratifications**

→ **Many will be (nearly) empty**

CRV

BETTER COWS > BETTER LIFE

Goal

We want a model that:

- Is robust even in regions of low data availability
- Is scalable to millions of observations
- Can easily be extended with other systematic variables

BETTER COWS > BETTER LIFE

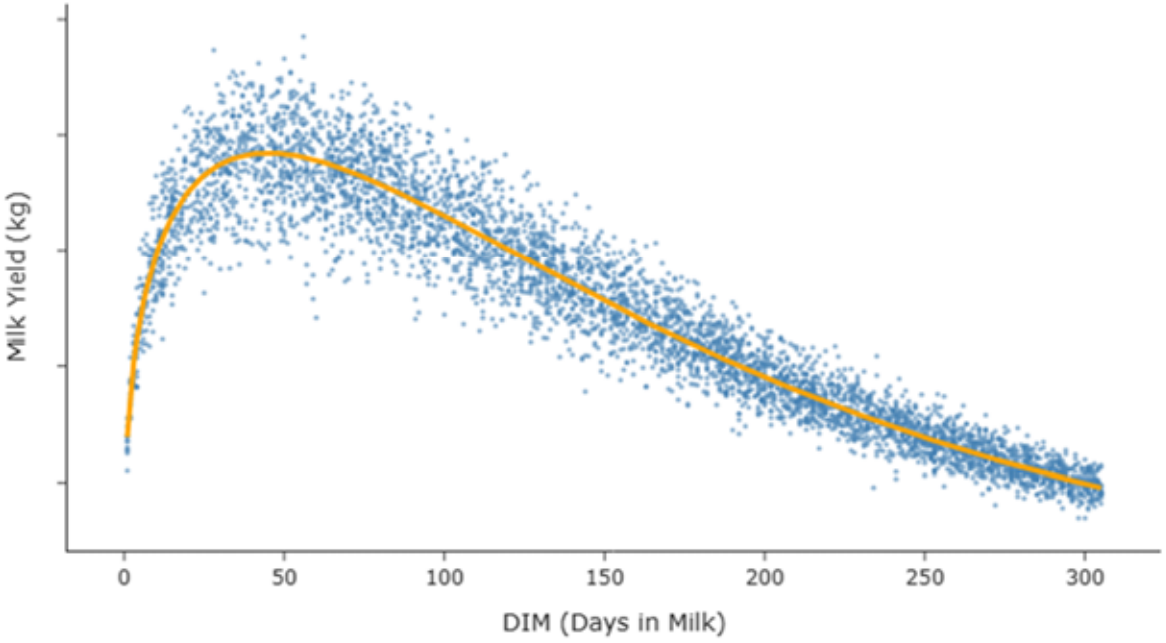
Can you draw a lactation curve?

You inherently already know a rough shape

Now we observe some data

We nudge our prior shape a bit upward.

Even more data comes in. Don’t care about our prior belief, simply fit the data.



CRV
BETTER COWS > BETTER LIFE

Bayesian concept

Balancing between prior belief (structure of a curve) and availability of data

Elegant way to still have robust predictions in regions of sparse data

CRV
BETTER COWS > BETTER LIFE

Our solution:

Stochastic Variational Gaussian Process (SVGP)

SVGP is basically an elegant mathematical representation of this trade-off

- **Learns smooth shared lactation-curve structure from nearby data**
  - DIM, calving season, age at calving, herdlevel
- **Automatically balances between learned structure and data fit**
- **Scalable to millions of observations**

CRV
BETTER COWS > BETTER LIFE

Approach

16 million test day milk yield observations:

80% used for training our model
10% used for validating our model
10% used for testing our model

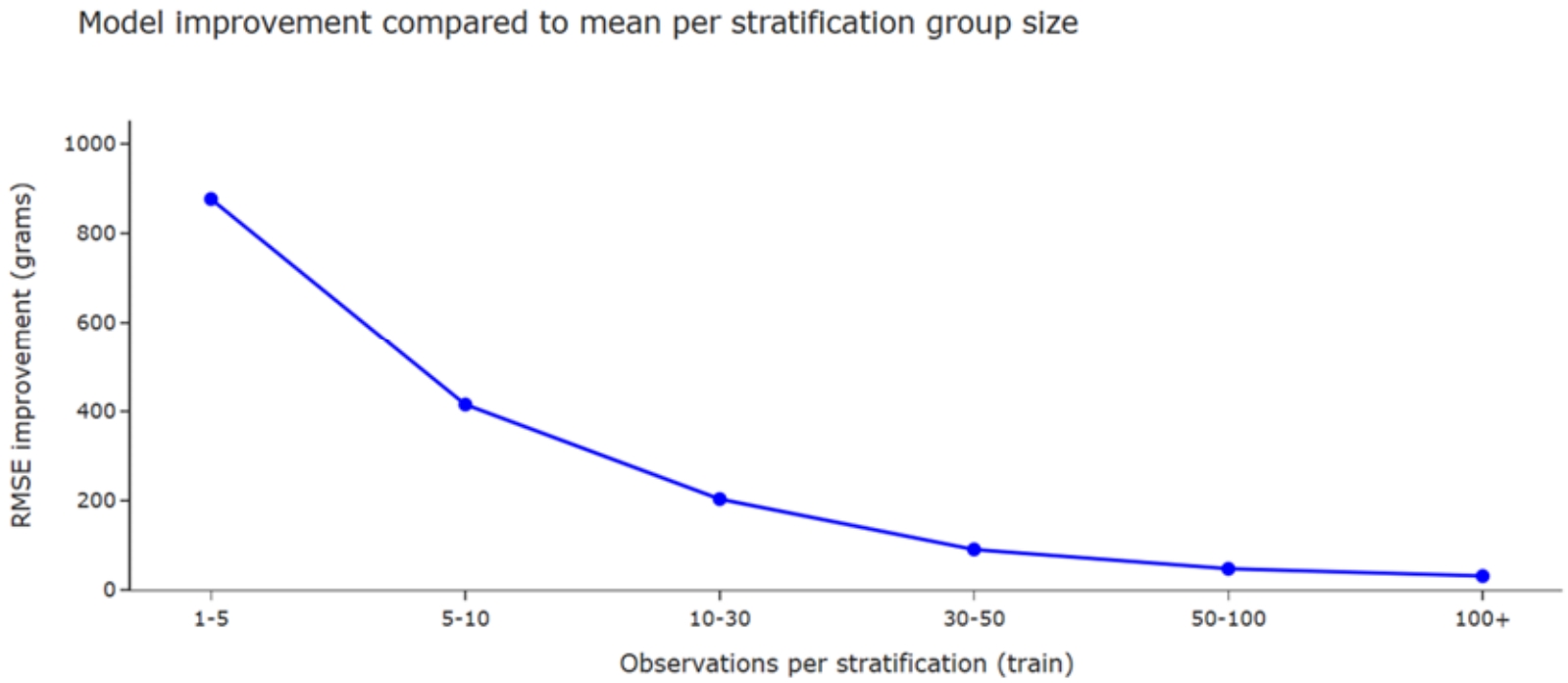
Calculated the Root mean squared error (RMSE) on test set for every combination of the stratifications

Compared SVGP with benchmark: average per stratification in train set applied to corresponding test set stratifications.

Results

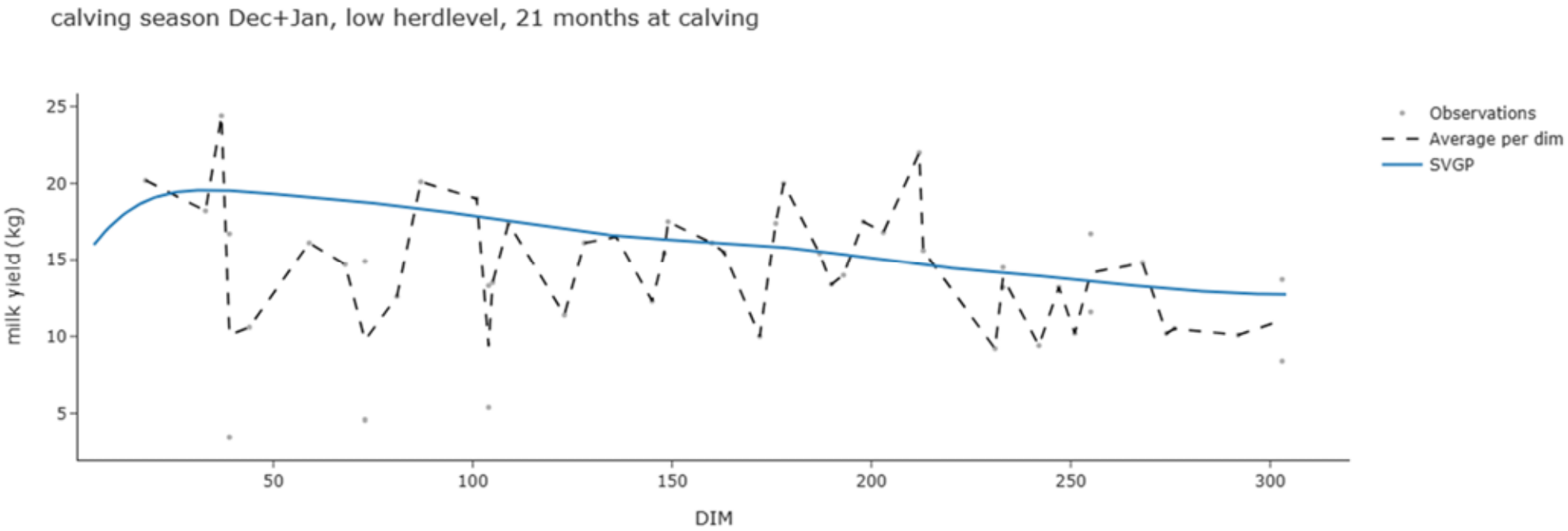
High availability data means taking average is pretty good → we want our model to perform on par.

Low availability data means taking average is not good → we want our model to perform a lot better



CRV
BETTER COWS > BETTER LIFE

Visual inspection of learned curves



Conclusion

- **The SVGP is a model that is able to estimate systematic effects even under low availability of data.**
- **This enables CRV to create robust performance metrics for any cow**

What's next?

- **Inspect other systematic variables (e.g. parity, pregnancy, 305+ DIM)**

BETTER COWS > BETTER LIFE

CRV

BETTER COWS > BETTER LIFE