Employing high resolution big data for predictive modelling in precision dairy farming

G. Katz

Speaker: Gil Katz
Dairy farming in the emerging era of IOT

Gil Katz
Afimilk
CONVERGENCE OF MEGA TRENDS

The INTERNET OF THINGS
Automated Data Collection and Analysis

- **Objective**
- **Accurate**
- **Effortless**
- **Consistent**
- **Accessible**

3D data-base

**Analysis**
- **Herd & group level**
- **Time domain**
- **Cow domain**

**Sensor domain**

**Diagnosis and Response**
- **Cow/Herd**
  - Feed, Health status, lactation, Gynecological status, ...
- **Interface**
  - Parlor maintenance, Staff, Cow preparation, Washing system ...

**Validity of data**

- **S1**
- **S2**
- **S3**
- **S4**

**Devices**
- Calibration, Technical malfunction, ..
Manage and merge different Data types

Quantitative data (monotonic structure)
milk yield, milk components, milk flow, weight ....

Qualitative data (discreet structure)
gynecological status, health status ...

Behavioral data (pattern based)
activity pattern, grouping pattern, rest pattern, feed pattern ....

Milking stall sensors – milk yield, milk flow, milk conductivity, milk fat, protein, lactose, blood, coagulation potential

Cow sensors – activity, lying times, lying bouts
Big Data

Complex biological systems

Challenge:
construct data, collect data, mine data, Develop predictive models,
Validate models, construct comparative standards

Disciplines

- Biology, Chemistry, Physics
- Data science, Mathematics, computer science
- Health, Fertility, Feed, Genetics, Production

Challenge: Pattern recognition of subjective multi dimensional data
Descriptive: From highlighting irregularities to diagnostics
From Data Collection to Decision Making

Data → Information → Knowledge → Intelligence

Data integrity? Normalize and classify

Optimization
- Predictive Modeling
- Descriptive Modeling
- Analytical on-line Reports
- Processed Data
- Raw Data

What happened?
- What is going to happen?
- What is the best that could happen?
- Why did it happen?

Arkadi Slezberg, 2009
From retrospective to prospective prediction of production

Real time measurement of milk yield and composition
AfiLab concept

- Casein, un-saturated fatty acids, saturated fatty acids, mono & poli Unsaturated fatty acids, igG count in colostrum
Predictive: From diagnostics to prediction

Different heuristic approach

From classical statistics terminology:

- Mixed models
- Decision trees
- Bayesian models

To time dependent terminology:

- Dynamic modeling
- Markovian and non-Markovian processes
- Memory stamps

- **Scope:** >37,000 Holstein cows spanning over 2 years

- Finds agreement between Afimilk's inline milk lab real time analysis and between DHIA monthly tests.

- Selected for 'Editor's Choice' of JDSc
Objectives of the study

✓ Comparison of lactation yields between the traditional testing & Afilab
✓ Calculation & comparison of Predicted Transmitting Ability (PTA)
✓ Calculation of genetic & phenotypic correlations
✓ Establishing correction factors for Season, Age & Open Days
✓ Calculation of extended yield factors for cows with truncated data (partial records)
Heritabilities, genetic and environmental correlations among 7,866 first parity 305 d lactations computed from the ICBA and AfiLab records.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Heritabilities</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICBA</td>
<td>AfiLab</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>0.23</td>
<td>0.31</td>
</tr>
<tr>
<td>Protein</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>(kg)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% fat</td>
<td>0.48</td>
<td>0.57</td>
</tr>
<tr>
<td>% protein</td>
<td>0.55</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Heritabilities were higher for the AfiLab records for all traits, except for % protein.
Phenotypic correlations among complete and extended 1st parity lactations the last ICBA test day and the last two weeks of AfiLab records.

<table>
<thead>
<tr>
<th>Trait</th>
<th>FAT (kg)</th>
<th>Mean days in milk at truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>ICBA</td>
<td>0.67</td>
<td>0.75</td>
</tr>
<tr>
<td>Afilab</td>
<td>0.77</td>
<td>0.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trait</th>
<th>PROTEIN (kg)</th>
<th>Mean days in milk at truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>ICBA</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Afilab</td>
<td>0.72</td>
<td>0.83</td>
</tr>
</tbody>
</table>

11th April 2017  Oded Nir
The genetic values for 1st lactation cows were higher by Afilab except for % protein.

The prediction coefficients for 305 days Kgs milk, fat & protein were higher for Afilab.

The genetic & phenotypic correlations to 305 days lactation in 30 DIM are 0.75 and gradually rising to 0.98 in 240 DIM.

Prediction of complete lactation yields from partial data were more effective in Afilab.
Prediction of complete lactations in Afifarm

- Our objective: To adapt the large scale retrospective study’s method to a prospective prediction of complete (305_days) lactations in individual herds
  - For selection
  - For production planning (quota, summer/winter)

- The operational need: To enable farmers to get the decision as early as possible, but before breeding
## Waiting Periods

<table>
<thead>
<tr>
<th>Herds</th>
<th>Cows/herd</th>
<th>Voluntary waiting period (days)</th>
<th>Days to 1\textsuperscript{st} AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,885</td>
<td>158.4 ± 325 SD</td>
<td>58.4 ± 5.6 SD</td>
<td>95.2 ± 26.9 SD</td>
</tr>
</tbody>
</table>

Ferguson J.D. & Skidmore A. (2013). JDS 96 (2) 1269 -1289

<table>
<thead>
<tr>
<th>Days to 1\textsuperscript{st} AI</th>
<th>50</th>
<th>51 - 80</th>
<th>81 - 110</th>
<th>111 - 150</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\textsuperscript{st} lactation</td>
<td>0.4%</td>
<td>41.4%</td>
<td>45.2%</td>
<td>13.0%</td>
</tr>
<tr>
<td>2\textsuperscript{nd} lactation</td>
<td>9.7%</td>
<td>58.4%</td>
<td>26.9%</td>
<td>5.1%</td>
</tr>
</tbody>
</table>


Our objective is to be able to make the decision at 60 DIM
Early prediction of total lactation performance
Prediction calculated from 2014 data (new) compared to 2015 data (old)

- Calibration of models from cows calving in 2014 (26/01-31/12)
- Validation of models applied cows calving in 2015
- 6 herds of Israeli Holsteins with 371 to 1046 annual calving events and 11,840 Kg to 13,635 annual milk
Criteria for Success

- $R^2 = \text{RSquare of the summary of fit}$
- $r = \text{Correlations to actual production}$
- 75% & 90% tiles of the differences between the predicted & actual estimates of the various traits (for planning & selection)
- Predictive Values & accuracy for selection decisions
  - PPR (positive predicting value) = The probability that a cow defined by test as a “low yielder” is truly so
  - NPR (negative predicting value) = The probability that a cow defined by test as a “high yielder” is truly so

Oded Nir (Markusfeld)
Afimilk; Herd #3

<table>
<thead>
<tr>
<th></th>
<th>Milk, kg/305 days</th>
<th>Fat, kg/305 days</th>
<th>Protein, Kg.305 days</th>
<th>ECM, kg 305 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>34</td>
<td>54</td>
<td>84</td>
<td>34</td>
</tr>
<tr>
<td>RSquare</td>
<td>0.683</td>
<td>0.726</td>
<td>0.786</td>
<td>0.704</td>
</tr>
<tr>
<td>Correlations</td>
<td>0.930</td>
<td>0.949</td>
<td>0.968</td>
<td>0.926</td>
</tr>
<tr>
<td>+tive PV</td>
<td>65.0%</td>
<td>72.2%</td>
<td>84.6%</td>
<td>47.5%</td>
</tr>
<tr>
<td>-tive PV</td>
<td>78.6%</td>
<td>79.3%</td>
<td>79.0%</td>
<td>86.1%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>75.0%</td>
<td>77.6%</td>
<td>80.0%</td>
<td>65.8%</td>
</tr>
<tr>
<td>10%tile to 90%tile</td>
<td>-10.1%</td>
<td>-7.5% to -4.7%</td>
<td>-11.4%</td>
<td>-9.5% to -11.4%</td>
</tr>
</tbody>
</table>

Herd #3: n for 12/14-11/15=717 (34 DIM); 1,195 (54 DIM); 1,912 (84 DIM); n for 12/14-02/16=76

- Prediction of all the production variables examined improved with time from calving
- The smaller herd behaved similar to the larger one

Oded Nir (Markusfeld)
Afilab &lt;=34 DIM vs. 1st ICBA milk test &lt;=34 DIM (All lactations combined)

<table>
<thead>
<tr>
<th>Herd #1</th>
<th>Milk, kg/305 d</th>
<th>Fat, kg/305 d</th>
<th>Protein, Kg.305 d</th>
<th>ECM, kg 305 d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Afi</td>
<td>ICBA</td>
<td>Afi</td>
<td>ICBA</td>
</tr>
<tr>
<td>RSquare</td>
<td>0.568</td>
<td>0.554</td>
<td>0.523</td>
<td>0.388</td>
</tr>
<tr>
<td>Correlations</td>
<td>0.858</td>
<td>0.800</td>
<td>0.866</td>
<td>0.727</td>
</tr>
<tr>
<td>+ve PV</td>
<td>75.0%</td>
<td>54.2%</td>
<td>60.6%</td>
<td>40.9%</td>
</tr>
<tr>
<td>-ve PV</td>
<td>83.1%</td>
<td>79.1%</td>
<td>87.0%</td>
<td>71.1%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>81.0%</td>
<td>70.1%</td>
<td>75.9%</td>
<td>61.2%</td>
</tr>
<tr>
<td>10%tile to 90%tile</td>
<td>-9.3% to 10.3%</td>
<td>-10.4% to 10.7%</td>
<td>-10.8% to 6.8%</td>
<td>-14.3% to 9.8%</td>
</tr>
</tbody>
</table>

Prediction for milk & fat, proved superior to that of ICBA (truncation at 34 DIM)
Afimilk; Afilab + Predicted Transmitting Ability (PTA)  
All lactations combined. Herd #3

<table>
<thead>
<tr>
<th></th>
<th>Milk, kg/305 days</th>
<th>Fat, kg/305 days</th>
<th>Protein, Kg.305 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIM34 +PTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.683</td>
<td>0.704</td>
<td>0.653</td>
</tr>
<tr>
<td>Correlations</td>
<td>0.930</td>
<td>0.926</td>
<td>0.918</td>
</tr>
<tr>
<td>+tive PV</td>
<td>65.0%</td>
<td>47.5%</td>
<td>65.0%</td>
</tr>
<tr>
<td>-tive PV</td>
<td>78.6%</td>
<td>86.1%</td>
<td>78.6%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>75.0%</td>
<td>65.8%</td>
<td>75.0%</td>
</tr>
<tr>
<td>10%tile to 90%tile</td>
<td>-10.1% to 8.4%</td>
<td>-11.4% to 7.0%</td>
<td>-8.7% to 9.8%</td>
</tr>
</tbody>
</table>

Adding PTA to the 34 DIM models in Herd #3 proved contributed more than in the 54 DIM models. Results were not different in Herd #1

Oded Nir (Markusfeld)
Prospective prediction of complete lactations in individual herds yielded similar results to Weller & Ezra’s large retrospective study.

Predictions using Afimilk in 34 DIM proved superior to those using the first Milk Test.

Though prediction improves with time in lactation, the present results allow for “safe” selection, culling & production planning at 54 DIM, and even earlier in lactation.

Results for small & large sized herds were similar.

Current average production planning error based on ICBA data is 20%-25% using daily afilab data the error drops down 5%-7%.

Adding PTA to the models slightly improved prediction of milk & protein in early lactation.
Take home message:
Not using available Daily data is a drawback to the industry.
Data reduction by averaging it is loss of information and knowledge.
Thank you