J. Dairy Sci. 106:5501–5516
https://doi.org/10.3168/jds.2022-22980

Can meteorological data improve the short-term prediction of individual milk yield in dairy cows?

L. Gasser,[1] F. Perez Cruz,[1] and M. Cockburn[2]*
[1]Swiss Data Science Center, 8050 Zürich, Switzerland
[2]Digital Production, Agroscope, 8356 Tänikon, Switzerland

ABSTRACT

Many farms document daily milk yields of individual cows because these are a good indicator of cow well-being. It is established that extreme meteorological conditions influence the milk yields by causing heat and cold stress, whereas less is known about the effects of moderate changes in meteorological conditions. Thus, the aim of the present study was to evaluate whether individual daily milk yield predictions can be improved by considering such changes. We evaluated 8 years of milking and meteorological data from Eastern Switzerland with a total of 33,938 daily milkings from 145 Brown Swiss and 64 Swiss Fleckvieh cows. The cows were aged between 1.9 and 13.5 years at parturition. The data set was split into 7 periods according to the days in milk (DIM) and subsequently filtered into subsets by breed and parity. We applied Gaussian process regression to predict individual daily milk yield. We compared different models including DIM, lagged milk yield, and meteorological variables as features and found that models including the lagged milk yield performed best. Within the period of 5 to 90 DIM, we were able to predict individual next-day milk yield from the cow's last milkings with a root mean squared error (RMSE) of 2.1 kg. In contrast, without information on the previous milk yield, accuracy of milk yield prediction was lower, with an RMSE close to 8 kg. The models holding information about previous milk yields showed a substantial increase in performance. Within a more homogeneous data subset filtered by breed or parity or both, predictions were even better, with a relative RMSE of 4.3% for first-parity Fleckvieh cows. However, we found that including meteorological features, such as temperature, rainfall, wind speed, temperature humidity index, cooling degree, and barometric pressure, did not improve the predictions in any of the evaluated periods. This finding indicates that considering meteorological features in daily milk yield prediction models is not useful in moderate climates; considering lagged milk yield is sufficient. We hypothesize that this meteorological information, among other influences, is indirectly contained in the lagged milk yield.
Key words: milk production, short-term milk yield prediction, Gaussian process regression

INTRODUCTION

Increasing interest in smart farming has led to a wide range of digitalized processes, offering new opportunities for data exploitation in dairy farming, such as the evaluation of feeding strategies (Menardo et al., 2021) or data-driven decisions in livestock management (Ferris et al., 2020). Daily milk yield (**DMY**) data are the most commonly collected data on dairy farms, as 45% of Swiss dairy farmers collect this information during milking (Groher et al., 2020). To date, milk yield data are considered in a standalone manner, despite the circumstance that a few known variables can cause short- or long-term drops in milk yield.

Influence of Meteorological Conditions on Milk Production

Researchers have evaluated the effects of temperature and humidity on dairy farms with specific regard to heat stress, where they reported an interrelation of thermoregulation, productivity, and meteorological variables, particularly during summer (Tao et al., 2020). Based on the evidence of previous studies, Tao et al. (2020) describe that physiological heat stress responses in dairy cows develop rapidly after half a day of stress and include increases in rectal temperature and respiration rate as well as reduced milk production (Tao et al., 2020). Spiers et al. (2004) described increases in rectal temperature and respiration rate after 24 h and a drop in milk yield and feed intake after 48 h when cows were exposed to air temperatures between 19 and 29°C. Similarly, Bouraoui et al. (2002) found a negative correlation (r = −0.76) of temperature-humidity

Received November 4, 2022.
Accepted February 23, 2023.
*Corresponding author: marianne.cockburn@agroscope.admin.ch

index (**THI**), milk yield, and feed intake, where THI increased from 68 to 78 and milk yield dropped by 21%. Milk production started decreasing when temperatures exceeded 20°C (Tao et al., 2020) or, as reported in a more recent study in Southern Germany, at temperatures above 16°C or THI above 60 (Mbuthia et al., 2022). The degree of cooling is another combination of meteorological parameters that has been proposed to be relevant for determining physiological reactions to meteorological conditions in dairy cows (Baehr et al., 1984).

Prediction of Milk Yield

Predicting long-term milk yield allows farmers to monitor farm production and adapt management decisions accordingly (Grzesiak et al., 2006). Predicting and monitoring DMY additionally allows detection of physiological alterations of the cow and thus can be useful for the implementation of early warning systems. Using short-term predictions as an alert for physiological alterations requires high accuracy, so that the farmer does not miss actual physiological changes but also does not ignore alarms owing to a flood of false positive alerts.

Milk yield models have followed 2 paths: lactation curve models, which allow modeling of milk yields as well as protein and fat contents, while considering seasonal trends caused by feeding (Wood, 1976), or random regression test-day models, which have been applied to evaluate animal breeding strategies (Schaeffer, 2004). Lee and Wardrop (1984) predicted individual milk yield from test-day milk data of 49 Canadian herds and were able to predict 75 to 80% of DMY with a prediction error below 2 kg. This was extremely valuable and innovative in a time before the onset of digitalization in milking parlors, where daily cow individual milk recordings were hardly available. Since then, researchers have been developing models to improve predictions from lactation curve models, test-day milk yield, and later DMY (Zhang et al., 2018).

McParland et al. (2019) further predicted individual 24-h test-day milk yield from a single milking sample of the same day to evaluate whether the morning or evening measurement can be omitted using data from 237 farms in Ireland. They found that 24-h test-day milk yields can be predicted satisfactorily from the morning milking sample when also considering the milking interval.

Murphy et al. (2014) predicted daily herd milk yield in pasture-based Holstein-Friesian Irish dairy cows by using DIM and the number of milked cows as features and found nonlinear autoregressive models with exogenous input (**NARX** models) to be useful in that context. These models were able to predict daily herd milk yield more accurately than multiple linear regression and artificial neural networks (Murphy et al., 2014). Hereby, short-term predictions of the next 10 d achieved a lower relative root mean squared error (**RMSE**) of 5.8% compared with long-term predictions of a complete lactation of 305 d with a relative RMSE of 8.6%, whereas the conventional models reached relative RMSE values between 10.5 and 12.2% (Murphy et al., 2014). The prediction quality was improved by shortening the prediction horizon. The NARX models allowed for short- and long-term yield predictions with satisfactory accuracies.

Milk Yield Prediction Using Meteorological Data

Previous research considered whether meteorological features can improve milk yield prediction, and the outcome was not quite consistent throughout the different studies. Some studies reported no or very small effects of meteorological data on milk yield prediction. The NARX model published by Murphy et al. (2014) was refined in Zhang et al. (2020) by including sunshine hours, precipitation, and soil temperature to predict individual DMY and tested on lactation data from pasture-based Holstein-Friesian Irish dairy cows. Hereby, the authors found a small effect of the meteorological parameters, particularly sunshine hours, which they reported with the reservation that these parameters "did not have a substantial impact on [milk yield] forecast accuracy" (Zhang et al., 2020, p. 120). The authors provided individual RMSE for the 10-d prediction horizon of 39 cows in the test set, which we used to calculate the average RMSE. The model containing sunshine hours led to an average RMSE of 3.1 kg, an increase of 0.2 kg compared with the 3.3-kg average RMSE of the base model without meteorological features. Furthermore, a recent study, also from Ireland, aimed to predict the national weekly milk yield of grazing cows over 12 years by using a variety of models and prediction horizons of 1 up to 52 weeks ahead, and found that including meteorological variables did not improve their results (O'Leary and Lynch, 2022). However, Marumo et al. (2022) found a small effect of the minimum temperature lagged by 2 d when predicting individual DMY in indoor-housed Holstein-Friesian dairy cows on a Scottish farm, but only for primiparous cows.

Other studies, in contrast, reported effects of meteorological parameters. Mbuthia et al. (2022) found the herd monthly test-day milk yield to be decreasing for temperatures above 16°C and THI greater than 60 [calculated using NRC, 1971: THI = (1.8 × T + 32) −

(0.55 − 0.0055 × rH) × (1.8 × T − 26); temperature (T) in °C, relative humidity (rH) in %]. Additionally, a study on Lacaune ewes in Spain performed a correlation analysis and found significant yet small Pearson correlations between individual DMY and the meteorological variables mean temperature, relative humidity, THI, and solar radiation (Osorio-Avalos et al., 2022). Interestingly, Yano et al. (2014) applied a generalized additive model using multiple variables, such as DIM, parity, temperature, and daylight length, to predict the individual milk production traits of yield and fat content from test-day data; they found that particularly high-yielding dairy cows were more sensitive to heat.

Aim of the Study

The current study aimed to quantify the effect of moderate meteorological changes on milk production in housed dairy cows using Gaussian processes.

MATERIALS AND METHODS

The experiment was conducted under the animal welfare license 32348–TG02/20.

Description of Experimental Farm and Animals

The experimental dairy farm had 2 barns: the experimental barn and the research emissions barn. Only milking data from the experimental barn were used in this study. The experimental dairy barn housed 63 lactating cows in an open barn with elevated deep-littered (straw) cubicles (Supplemental Figures S1–S6; https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023). The barn had an Eternit roof with a ridge vent (Ondapress 57, Eternit Schweiz AG). Surfaces were plane-fixed inside the barn and slatted in the outdoor loafing area. The barn was separated into 2 feeding group areas, and feed was given according to the cows' stage of lactation (Supplemental Figures S2 and S3). The cows had access to a shared outdoor loafing area, which they could enter through electronic gates (Supplemental Figures S4 and S5). The cows were fed a TMR once per day, and the feed was pushed to the trough every 3 h by an automated feeding robot (Lely Juno). The TMR contained grass silage (47%) with maize silage (29%), hay (5%), sugar beet silage (14%), and concentrates (3.8%). Additionally, the cows were supplemented with concentrates on an individual level according to their DMY (range 0–8 kg of concentrates) via automated feeding stations placed within the outside loafing area (Supplemental Figure S5). The barn was open on 3 sides, allowing free airflow during summer, and could be closed with curtains during winter (Supplemental Figure S1). This barn represents the primary housing of the dairy cows. However, depending on experiments, some of the cows were moved to the research emissions barn, which was located 4 km away, for brief periods of approximately 3 to 5 mo. Therefore, the number of cows housed in the primary experimental barn varied, reaching 63 lactating cows at most. The cows were accustomed to these relocations. Because no data from the research emissions barn were considered in this study, the data set did not include complete lactation cycles for each cow and lactation period.

Data Collection

For this specific analysis, 2 data sources were exploited. First, we extracted historic milk yield data from the DairyPlan C21 herd management software (GEA Farm Technologies) for the time between December 2014 and June 2022 from the experimental dairy farm, located in Tänikon, Switzerland. The milk yields were recorded with the International Committee for Animal Recording-approved device Megatron P21 (GEA Farm Technologies). The milking parlor was professionally serviced yearly. Herd management data included morning and evening milk yield, parity, DIM, cow identity and name, breed, age, and date of calving. Until February 2019, these data were recovered from backup files. In spring 2019, 88 days of data were missing owing to technical issues (software update). From May 2019, daily output files were automatically saved.

Second, meteorological data, including temperature, rainfall, humidity, wind speed, solar radiation, and barometric pressure, were collected from the MeteoSwiss meteorological station located 300 m away from the dairy farm (Federal Office for Meteorology and Climatology, MeteoSwiss station: Tänikon; www.meteoschweiz.ch; Supplemental Figure S7; https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023). The temperature range of the data did not allow us to draw conclusions about heat stress, because mean daily temperatures above 20°C were only measured on 229 of the 3,108 d of the measurement period (7.4%).

We additionally calculated THI and degree of cooling (in W/m^2) by applying the following formulas to the meteorological data, where T is temperature (in °C), rH is relative humidity (in %), and v is wind speed (in m/s).

Temperature-humidity index:

$$THI = (0.8 \cdot T) + \left[\frac{rH}{100} \cdot (T - 14.4)\right] + 46.4$$

(Mader et al., 2006).

Cooling degree (CD):

$$CD = 36.5 - T \cdot \left(0.14 + 0.49 \cdot \sqrt{v}\right) \text{if } v > 1;$$

$$CD = 36.5 - T \cdot \left(0.20 + 0.40 \cdot \sqrt{v}\right) \text{if } v \leq 1$$

(Baehr et al., 1984).

Due to the use of historic milking data, we were not able to measure meteorological data directly at the farm for the entire study period. However, we collected temperature, humidity, and wind speed above the cubicle and outside in the loafing area, at 2 m height, during the last year of the study. These data on indoor and outdoor microclimates were collected via HOBO climate loggers (HOBO RX3000, HOBO Data Logging Solutions) and used to show the correlation and relevance of the data measured by the MeteoSwiss meteorological station.

Data Selection

We extracted milking data from December 2014 to June 2022 from the herd management system. The data were pre-processed by removing (1) zero milk yield and lactation number 0 entries, (2) duplicated entries with the same milk yield for several days (these correspond to manual entries where the cows were not milked in the experimental barn), and (3) duplicated entries per milking. If more than 2 entries per morning or evening milking existed, we used the last entry after plausibility checks. After preprocessing, 148,563 single milking entries remained, resulting in 550 lactation curves of any length, from 225 cows [157 Brown Swiss (**BS**) and 68 Swiss Fleckvieh (**FV**)]. Cows with clinical mastitis were milked into a bucket, and their yields were not recorded. This appears as missing data in the data set. After recovery, cows were milked normally and data were collected again.

Missing Data and Data Imputation. At the experimental barn, the cows were milked twice per day: in the morning (0600–0700 h) and in the evening (1630–1730 h). On 4 d, the data from either a morning or an evening milking were missing for all cows. Additionally, most cows lacked a few (mostly 1–2) milk yield entries. We considered only complete lactation periods with 2 milkings per day in a consecutive period (i.e., without any missing entry) for the prediction of milk yield.

To increase the number of complete lactation periods, we imputed missing values by linear interpolation of previous and subsequent milk yields of same milking periods; that is, morning or evening milking. If the previous or the subsequent milk yield was not available, no imputation was done. This means that gaps larger than 1 d were not filled, and affected lactation curves were excluded from the analysis. Less than 1% of milk yield entries were imputed for all data subsets (Table 1). For example, for the first 5 to 90 DIM, 81 lactation curves were available without imputation and 178 curves with imputation. Out of the 30,220 milkings, less than 0.5% were imputed, implying that most of the data were correct and only very few data points needed to be imputed (Table 1). For 89 lactation curves, 1 or 2 values were imputed. For 7 lactation curves, more than 2 milkings needed to be imputed. The other data subsets needed a similar amount of imputation ranging from 0.38 to 0.71% of the corresponding total number of entries (Table 1).

Selection of Periods. The moving of cows between the experimental and the research emissions barn led to an insufficient number of complete lactation curves with up to 305 DIM. The first 4 d after calving were excluded for all lactation curves. In the current study, we therefore evaluated five 30-d periods within the first 150 d of the cow's lactation, collected over the course of 8 years. These periods were established according to previous studies (Laevens et al., 1997). Additionally, we evaluated longer periods, namely, the first 90 DIM and the first 150 DIM. (See Table 2 for a detailed overview and Supplemental Table S1 (https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023) for the number of lactation curves per data subset.)

Summary of Selected Data. In total, 33,938 daily milkings (i.e., sum of morning and evening milkings) of 145 BS and 64 FV cows were considered in the evaluation. The cows were aged between 1.9 and 13.5 years (mean ± SD: 4.2 ± 2.0 yr) at calving. Modeled periods are presented in Table 2. We refer to a data subset as all

Table 1. Summary table of imputed values

	Period (DIM)						
Milk yield data	5–30	31–60	61–90	91–120	121–150	5–90	5–150
Entries (n)	12,056	17,254	13,784	11,790	8,974	30,220	19,630
Imputed (n)	57	66	80	57	64	147	117
Fraction (%)	0.47	0.38	0.58	0.48	0.71	0.49	0.60

Table 2. Overview of modeled periods based on the daily milk yield

		Parity[2]											
		1			2			3+			1+		
			Milk yield (kg)			Milk yield (kg)			Milk yield (kg)			Milk yield (kg)	
DIM	Breed[1]	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
5–30	Both	74	26.2	4.4	57	35.7	3.3	106 (75)	37.7	5.8	237 (149)	33.6	7.4
	BS	44	25.8	4.5	33	34.1	4.6	74 (51)	37.1	5.1	151 (99)	33.2	6.8
	FV	30	26.7	4.3	24	37.9	6.4	32 (24)	39.1	7.1	86 (50)	34.5	8.3
31–60	Both	88	27.0	4.9	64	36.3	6.2	135 (93)	39.7	5.8	287 (176)	35.1	7.9
	BS	64	25.7	4.5	38	34.0	4.8	104 (69)	38.4	5.4	206 (127)	33.7	7.5
	FV	24	30.8	3.8	26	39.6	6.6	31 (24)	44.1	5.1	81 (49)	38.7	7.6
61–90	Both	62	25.2	4.5	56	33.6	6.5	111 (82)	36.8	6.2	229 (157)	32.9	7.2
	BS	45	23.9	4.1	33	32.2	5.3	88 (62)	35.6	5.6	166 (112)	31.8	7.2
	FV	17	28.8	3.7	23	35.7	7.4	23 (20)	41.5	6.0	63 (45)	35.9	7.9
91–120	Both	49	24.4	4.7	50	30.2	6.6	97 (79)	33.3	6.2	196 (142)	30.3	7.0
	BS	34	23.0	4.4	27	28.5	5.5	70 (56)	37.6	5.6	131 (93)	28.8	6.4
	FV	15	27.5	3.6	23	32.1	7.3	27 (23)	37.6	5.6	65 (49)	33.3	7.1
121–150	Both	41	23.7	4.2	37	27.7	6.5	71 (61)	30.0	6.2	149 (115)	27.7	6.4
	BS	25	22.2	4.2	21	25.4	4.2	49 (42)	28.6	5.9	95 (74)	26.2	5.8
	FV	16	26.0	3.0	16	30.7	7.7	22 (19)	33.2	5.4	54 (41)	30.3	6.4
5–90	Both	51	25.8	4.6	41	35.2	5.7	86 (67)	38.6	5.8	178 (131)	34.2	7.7
	BS	37	24.8	4.4	23	33.3	4.3	66 (49)	37.6	5.4	126 (94)	33.1	7.4
	FV	14	28.5	4.0	18	37.6	6.3	20 (18)	41.9	5.9	52 (37)	36.8	7.8
5–150	Both	19	24.2	5.0	13	31.0	7.4	35 (31)	35.6	6.2	67 (60)	31.5	7.9
	BS	15	23.7	5.1	7	28.8	4.8	27 (23)	35.0	6.0	49 (42)	30.7	7.5
	FV	4	25.9	3.9	6	33.6	8.8	8 (8)	37.7	6.6	18 (18)	33.7	8.3

[1]BS = Brown Swiss; FV = Fleckvieh.

[2]For single parities (i.e., 1 and 2), the number of periods coincides with the number of cows. For parity ranges (i.e., 3+ and 1+), the same cow can have more than one lactation curve (e.g., from the first and second parities); therefore, the number of cows is given in parentheses after the number of lactation curves.

lactation curves from 1 of the 7 periods. For example, from 5 to 90 DIM for a breed (i.e., both, BS, and FV) and a parity range [i.e., first parity (1), second parity (2), third parity and above (3+), and all parities (1+)].

Modeling

We applied Gaussian processes and linear models as follows.

Gaussian Process Regression. Gaussian processes (**GP**) are state-of-the-art Bayesian tools for discriminative machine learning; that is, regression (Williams and Rasmussen, 1995), classification (Kuss et al., 2005), and dimensionality reduction (Lawrence and Hyvärinen, 2005). They were first proposed in statistics by O'Hagan (1978) and are well known to the geostatistics community as kriging. However, owing to their high computational complexity (i.e., cubic in the number of samples), they did not become widely applied tools in machine learning until the early 21st century (Williams and Rasmussen, 2006). In its simplest form, a GP for linear regression uses a Gaussian prior over the weights of the regressor. It couples them with a least square error loss function (Gaussian likelihood), which allows for computing in closed form the best prediction for each input and its confidence interval. By relying on the kernel trick (Williams and Rasmussen, 2006), GP can also solve nonlinear regression problems in closed form. It is the main feature of GP to provide accurate predictions, which naturally come with confidence intervals. The entire Appendix describes further details on GP.

The analysis was performed in Python 3.10.4 by applying the GP package gpflow 2.5.2 (Matthews et al., 2017).

Models. For prediction, DMY was used as one of the 4 target variables. Alternatively, differenced milk yields were used, namely, the difference between the current and the previous milk yield, and the difference between the current and the average of the last 2 and 3 d, respectively.

For each of the 4 target variables, we applied 3 baseline models. The simplest model predicted a constant milk yield for all cows, which coincides with the average milk yield of the data subset the model was trained on. A slightly more complex model considered that the milk yield depends on the lactation progression by using DIM as the single feature. The third baseline model used the previous milk yield as the prediction of the current milk yield.

To determine the features yielding the best prediction, several models with varying features were fit, again for the different target variables. We evaluated DIM and the milk yield (in kg) of up to 3 previous lagged days. Meteorological features included temperature (in °C), rainfall (in mm/d), wind speed (in m/s), barometric pressure (in hPa), THI, and degree of cooling (in W/m^2). The provided hourly values were averaged over 24 h with a time range from 1800 h of the previous day to 1800 h of the current day, accounting for the fact that the current DMY cannot be affected by the weather after the evening milking. After averaging, the meteorological features were merged with each cow's individual milking data. We included the meteorological features of either the current day, the previous day, some of the previous days (1, 3, 5, 7), or all 10 previous days. For temperature and THI, we also used moving averages over the last 2, 3, 5, and 7 d, respectively.

In all but the baseline models, each feature was standardized (zero-mean, unit SD) based on the training data. The predictions for the differenced milk yields were calculated back to the original scale in kilograms. In the GP models, we used 2 parts of the kernel in Appendix Equation [13], a radial basis function with the same length scale for all dimensions, describing the relation between the data points as well as the noise variance.

Data Splitting and Model Evaluation. We ran the models on all data subsets from Table 2, except for the period of 5 to 150 DIM, for which the models were only fit to data subsets with both breeds because there were not enough data points for the 2 breeds to be evaluated separately. For each data subset, an 80:20 train test split was applied, followed by 5-fold cross-validation on the training data. All splits were applied on the cow level, stratified by breed, ensuring that no data leakage occurred between train and test set or between the cross-validation folds. We calculated the evaluation metrics RMSE, mean absolute error (**MAE**), and relative RMSE (divided by average milk yield of the respective period as in Murphy et al., 2014). The final models were retrained on all training data and evaluated on the test data to obtain test error.

Models Based on Linear Extrapolation. To expand on the baseline model that uses the last milk yield as the prediction of the current milk yield, we implemented a simple linear extrapolation. To one of the data subsets we fitted linear regression models on a varying number of lagged milk yields and used these models to predict the current DMY. We considered 1, 2, 3, 5, and 10 lagged days and used the scikit-learn (Pedregosa et al., 2011) implementation.

Linear Temperature Models. We also investigated linear relationships between milk yield and temperature by using scikit-learn (Pedregosa et al., 2011). To one of the data subsets, we applied bootstrapped simple linear regression on the training data, with DMY as independent variable and temperature as the single dependent variable. The bootstrap sample size was set to 1,000.

Limitations

The data used in the current study originated from only 1 farm. However, using these data has the advantage of well-documented processes with adequate housing conditions in terms of management and animal welfare, as well as a carefully managed herd of dairy cows with different breeds (FV and BS), a relatively long data collection period, and fully digitalized management processes. Additionally, Shock et al. (2016) recommended not using data from meteorological stations to predict microclimate, because the on-farm temperature was higher than that of the meteorological station. However, we controlled for this condition by comparing on-farm data to those from the meteorological station and found temperature in the barn to be roughly 2°C below that of the meteorological station for the entire time frame compared, supporting the good management of the farm (Supplemental Figure S8; https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023).

Furthermore, we selected shorter periods than the usual classifications of early, mid-, and late lactation for milk yield prediction. This was a compromise we made to include a larger number of complete data sets, but it also allowed for a detailed analysis of the prediction quality. Although the most critical period of a dairy cow's health during lactation is the transition phase (3 wk before to 3 wk after parturition; Sundrum, 2015), the period from 5 to 90 DIM reflects a vulnerable time of lactation, where cows are particularly affected by changes of meteorological conditions owing to their peaking milk production (Yano et al., 2014). At the same time, milk production strongly increases during this time, which makes it both particularly interesting and difficult to predict.

Additionally, we would point out that we did not consider feeding information in the study because these historic data were not available for the entire period. However, feeding strategies were consistent throughout the measured period.

RESULTS

We first present the results for the period from 5 to 90 DIM with cows from both breeds and all parities before we look at the breeds, parity ranges, and periods separately. Unless stated otherwise, cross-validation errors are given.

Table 3. Validation evaluation metrics for the 3 baseline models on all target variables, performed on the data subsets of the 5- to 90-DIM period for both breeds and all parities[1]

Baseline model	Daily milk yield			Difference from last milking			Difference from last 2 milkings			Difference from last 3 milkings		
	RMSE (kg)	RMSE (%)	MAE (kg)	RMSE (kg)	RMSE (%)	MAE (kg)	RMSE (kg)	RMSE (%)	MAE (kg)	RMSE (kg)	RMSE (%)	MAE (kg)
Last day	2.52	7.26	1.72	2.52	7.26	1.72	2.21	6.36	1.53	2.19	6.30	1.51
Constant	7.69	22.13	6.31	2.52	7.26	1.72	2.21	6.36	1.53	2.19	6.30	1.51
DIM	7.60	21.85	6.23	2.52	7.25	1.71	2.20	6.33	1.52	2.16	6.21	1.49

[1]MAE = mean absolute error; RMSE = root mean squared error.

Baseline Models

We were able to predict individual next-day milk yield from the cow's last milking(s) with an RMSE of 2.2 to 2.5 kg (relative RMSE: 6.2–7.3%; MAE: 1.5–1.7 kg). In contrast, without information on the previous milk yield (i.e., for 2 of the 12 combinations of target variables and baseline models), accuracy of milk yield prediction was lower, with an RMSE of 7.6 to 7.7 kg (relative RMSE: 22%; MAE: 6.2–6.3 kg; Table 3). Considering DIM as the single feature [i.e., the model *DMY* ∼*DIM* (presented as target variable ∼features)] led to an RMSE of 7.6 kg, which was slightly better than fitting a constant value across all cows, which resulted in an RMSE of 7.7 kg.

The baseline models holding information about previous milk yields, either in the target variable or when the current milk yield was predicted as the cow's milk yield from the last day, showed a substantial increase in performance. By including more previous milk yield knowledge in the independent variable, we noticed a saturation effect. The previous milk yield reduced the RMSE from almost 8 kg to 2.5 kg (i.e., prediction improved by 5 kg), whereas including the last 2 milk yields led to a small additional reduction of 0.3 kg, and including the last 3 milk yields hardly improved the prediction any further (<0.1 kg). Therefore, we did not consider the difference to the last 3 milkings any further in the other models.

Linear Extrapolation Models

The linear extrapolation models were found to perform at par or worse than the models including previous milk yield knowledge; that is, in the range of 2.4 kg and 9.0 kg RMSE for varying number of lagged milk yields used to base the extrapolation on. The highest RMSE, obtained by using 2 lagged milk yields to predict the difference in milk yield, is actually even worse than the RMSE of the models without previous milk yield knowledge. In the first 90 d of the lactation, the lactation curve is not linear but rather first increasing, then reaching a plateau at 60 to 80 DIM, and then decreasing again but with a smaller slope. This shape of the lactation curve cannot be captured well with linear extrapolations, in contrast to using the last milk yield as the prediction of the current milk yield. The latter, 1 of the 3 baseline models, coincides with the linear extrapolation model using 1 lagged milk yield and is able to capture the changing curvature of the lactation curve and sudden changes well.

The linear extrapolation models, which are simple but restricted, are outperformed by the GP models using lagged milk yields as features. The GP models are nonlinear and therefore more suited to learning nonlinear relationships between variables.

Models with Meteorological Variables

Including meteorological variables did not improve the prediction error, independently of considering DIM in the model (Table 4). For DMY as target variable, the RMSE stagnated close to 8 kg when including the meteorological variables temperature, wind speed, rainfall, and barometric pressure, and quantities derived from them, such as THI or cooling degree (Table 4). This was the case for numerous ways of adding meteorological information, such as from the current day, the previous day, the 10 previous days, and the moving averages of the 2, 3, 5, and 7 previous days (Supplemental Table S2; https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023). Adding temperature, wind speed, and rainfall together to the same model or adding temperature of selected days of the previous week did not affect the prediction (results not shown). For the differenced target variables, RMSE is substantially lower but not also affected by the different meteorological variables (Table 4).

In addition, removing DIM from the models did not help (Table 4; Supplemental Table S2). For the raw target variable, the RMSE was slightly worse. For the differenced milk yields, we observed no change in prediction error (Table 4). Because DIM is the basic descriptor of lactation curves, we kept it for the remaining models.

Table 4. Validation root mean squared error (RMSE; kg) for models with different meteorological variables of the current day, not including lagged milk yields as features, performed on the data subsets of the 5- to 90-DIM period for both breeds and all parities[1]

	Daily milk yield		Difference from last milking		Difference from the last 2 milkings	
Meteorological variable	w/ DIM	w/o DIM	w/ DIM	w/o DIM	w/ DIM	w/o DIM
None	7.60	7.69	2.52	2.52	2.20	2.21
Temperature	7.68	7.77	2.52	2.52	2.20	2.21
Rainfall	7.60	7.69	2.52	2.52	2.20	2.21
Wind speed	7.60	7.70	2.52	2.52	2.20	2.21
THI	7.69	7.77	2.52	2.52	2.20	2.21
Cooling degree	7.67	7.75	2.52	2.52	2.20	2.21
Barometric pressure	7.63	7.71	2.52	2.52	2.20	2.21

[1]THI = temperature-humidity index; w/ = with; w/o = without.

Models with Previous Milk Yield Knowledge

By adding lagged milk yield features to the models that predict DMY, the RMSE dropped to 2.1 to 2.5 kg (Table 5), which coincided with the error range of the baseline models with previous milk yield knowledge. As in the models with meteorological variables, this outcome was independent of adding DIM or meteorological variables such as temperature, THI, and degree of cooling as features (Table 5). As in the baseline models, we observed a saturation effect with increasing knowledge of previous milk yields (Table 5).

Across Breeds, Parity Ranges, and Periods

The same patterns were evident when considering parity and breeds separately (still considering the period from 5 to 90 DIM). The prediction error depended on the data subset. More homogeneous groups, such as BS or FV cows in the first parity, tended to achieve a lower prediction error (Table 6). The RMSE for single parities (e.g., first and second parity) and parity ranges (e.g., third and later parities) was up to 3.5 kg lower than the RMSE for all parities. Modeling the breeds separately had a smaller effect of up to 1.9 kg and did not necessarily improve the prediction for both breeds (e.g., for the second parity). Across the breeds and considered parity ranges, adding lagged milk yield reduced the prediction error substantially, yet again we noticed a saturation effect with increasing previous knowledge.

The findings for the period from 5 to 90 DIM were consistent throughout all selected periods. It was not possible to improve the short-term milk yield prediction by any meteorological parameter, as shown in Figure 1 for both breeds. We plotted the results per breed (BS and FV cows) in Supplemental Figures S9 and S10, respectively (https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023). These figures summarize the results presented in the previous sections. Each subplot represents a period, and each line represents the results from a data subset (e.g., for a period, breed, and parity range). The prediction error of each model run is depicted with a symbol. Models without meteorological variables are shown with a circle and solid lines connecting them. The cross, plus, and diamond represent models with temperature, THI, and cooling degree, respectively. They mostly coincide with the model without weather variables. Including previous milk yield to predict the current yield provides a significant error reduction, illustrated in the plot when we move from 0 to 1 in the x-axis, representing previous milk yield knowledge. Adding the milk yields of 2 or 3 previous days has a marginal effect, leading to an additional reduction in RMSE of only up to 1 kg across

Table 5. Validation root mean squared error (RMSE; kg) for models on daily milk yield with knowledge of previous milk yields through lagged milk yield features, performed on the data subsets of the 5- to 90-DIM period for both breeds and all parities[1]

	Lagged milk yield variables							
	0		1		2		3	
Meteorological variable	w/ DIM	w/o DIM	w/ DIM	w/o DIM	w/ DIM	w/o DIM	w/ DIM	w/o DIM
None	7.60	7.69	2.47	2.48	2.16	2.17	2.11	2.13
Temperature	7.68	7.77	2.48	2.48	2.16	2.18	2.12	2.14
THI	7.69	7.77	2.48	2.48	2.16	2.18	2.12	2.14
Cooling degree	7.67	7.75	2.48	2.48	2.16	2.18	2.12	2.14

[1]THI = temperature humidity index; w/ = with; w/o = without.

Table 6. Validation root mean squared error (RMSE; kg) and relative RMSE (%) for models on daily milk yield with knowledge of previous milk yields through lagged milk yield features, with all models including DIM as a feature, performed on the data subsets of the 5- to 90-DIM period with the data split by breeds and parity ranges

		Lagged milk yield variables							
		0		1		2		3	
Breed[1]	Parity	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)
Both	1	4.53	17.11	2.03	7.67	1.76	6.67	1.74	6.56
	2	6.08	17.38	2.60	7.42	2.13	6.09	2.06	5.88
	3+	5.61	14.35	2.73	6.98	2.41	6.17	2.36	6.04
	1+	7.60	21.85	2.47	7.12	2.16	6.21	2.11	6.08
BS	1	4.30	16.83	2.14	8.37	1.85	7.23	1.82	7.14
	2	4.16	12.65	2.36	7.16	2.06	6.25	2.00	6.07
	3+	5.27	13.86	2.80	7.35	2.45	6.44	2.40	6.30
	1+	7.32	21.73	2.59	7.70	2.24	6.65	2.19	6.50
FV	1	4.19	14.48	1.69	5.83	1.46	5.04	1.41	4.87
	2	7.38	19.68	2.28	6.08	2.32	6.19	2.21	5.90
	3+	5.60	13.22	2.49	5.88	2.25	5.31	2.21	5.22
	1+	7.67	20.49	2.18	5.83	1.96	5.25	1.92	5.14

[1]BS = Brown Swiss; FV = Fleckvieh.

all data subsets. Once we use the previous milk yield the model becomes "personalized" for each cow, which explains the measured improvement.

In Figure 2, we show individual lactation curves and the prediction for 2 models for 23 cows from the period from 5 to 150 DIM. The first 3 rows correspond to first-, second-, and third-parity cows, respectively, whereas the fourth row shows cows with parity 4 and above. The first 3 columns contain BS cows, the last 3 rows FV cows. The actual lactation curve is depicted as a black line, which is overplotted by the prediction from one of the models. We compare the baseline model *DMY ∼DIM*, which has no "personalization" and yields the same prediction for all cows (purple lines), to the baseline model *DMY diff ∼DIM* (where *DMY diff* is the difference between current and previous milk yield), which contains information of the prior milk yield in the target variable. The latter model is "personalized" to each cow, meaning its predictions are different for each cow and coincide with the true lactation curve.

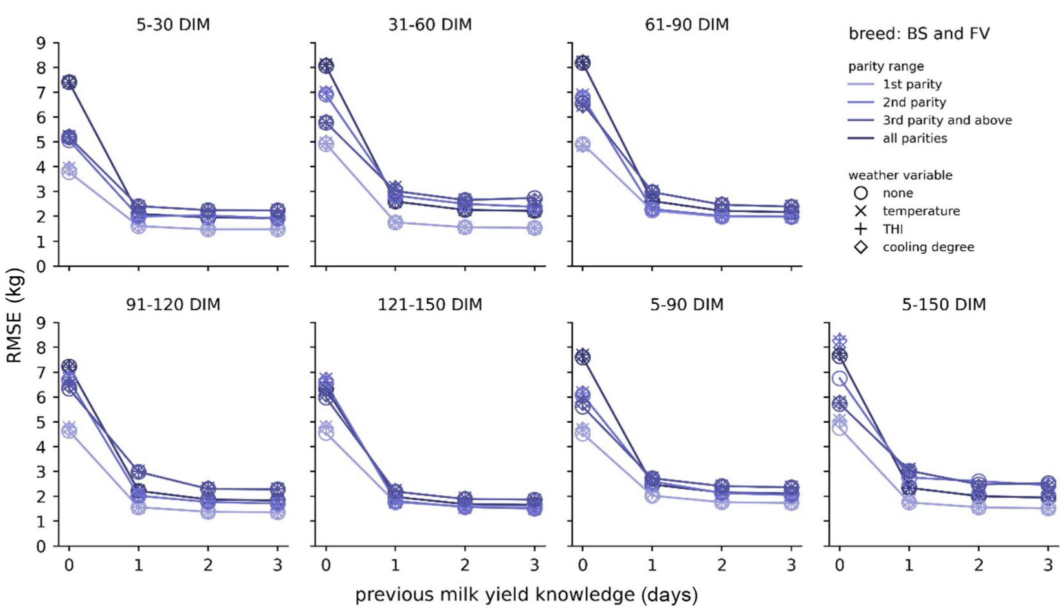

Figure 1. Validation root mean squared error (RMSE) depending on previous milk yield knowledge in days of lagged milk yield for different periods and parity ranges (for both breeds). Meteorological variables such as temperature, temperature-humidity index (THI) and cooling degree did not improve the prediction. BS = Brown Swiss; FV = Fleckvieh.

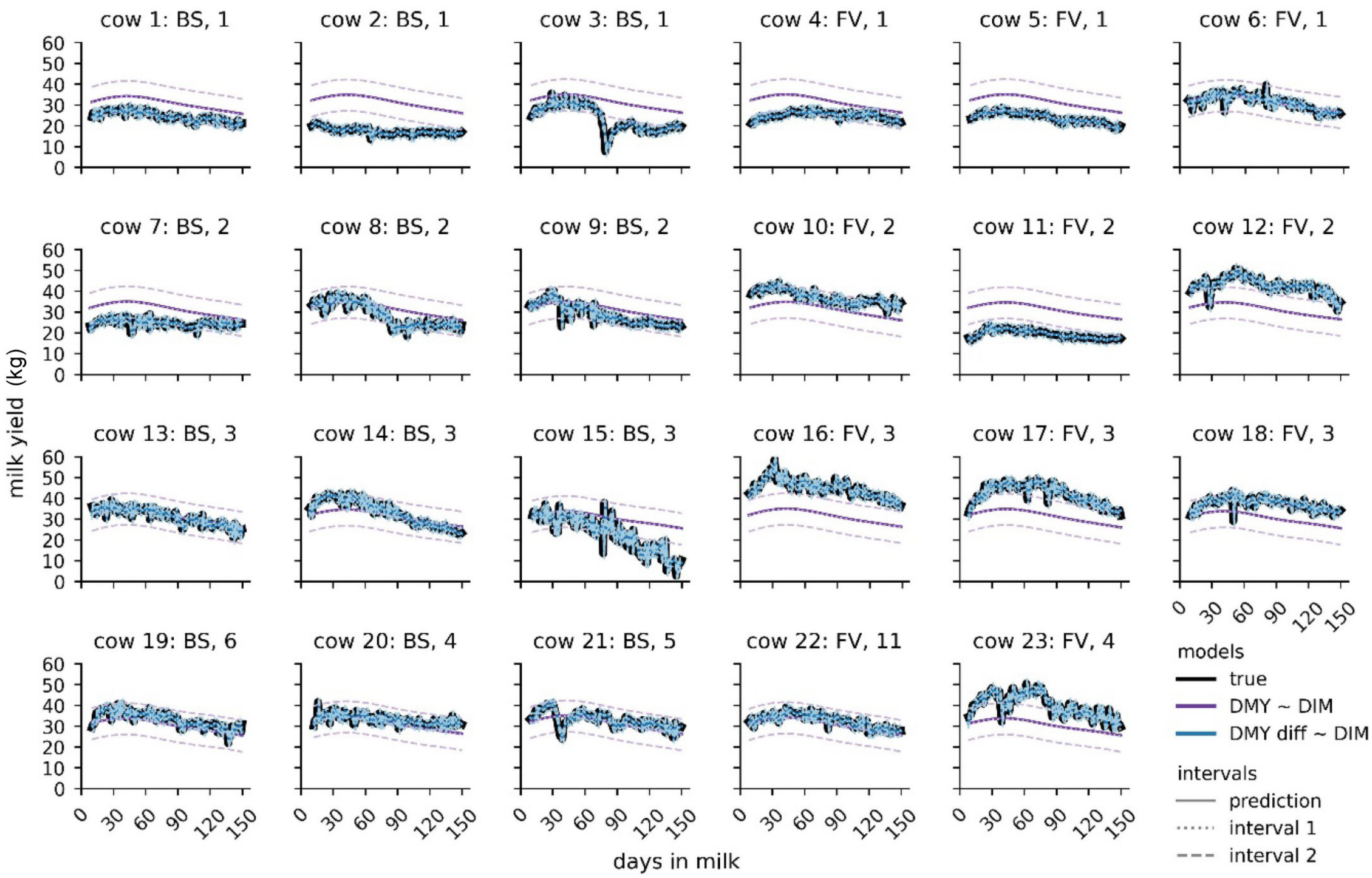

Figure 2. Examples for cow individual milk yield prediction (for the period from 5 to 150 DIM). The subplot titles state breed and parity of each cow. The dotted interval 1 corresponds to the prediction ± the SD of the radial basis function kernel, whereas the dashed interval 2 reflects the prediction ± the sum of the SD of the radial basis function kernel and the noise (Appendix Equation [12]). BS = Brown Swiss; FV = Fleckvieh; DMY = daily milk yield; DMY diff = difference from last milking.

Similar Outcomes From Test Data

For the models in Figure 1, the evaluation on the test data revealed the same behavior with slightly different error values (Table 7 and Supplemental Figure S11; https://doi.org/10.6084/m9.figshare.23309519; Gasser et al., 2023).

Improved Performance for More Homogeneous Data Subsets

To ensure the assumption that our model performs better for more homogeneous data subsets, we compared the prediction error with the spread of the milk yield and found that the RMSE correlated with the standard deviation for most models (Figure 3).

Daily Milk Yield or Differenced Target Variable

Using differenced milk yield instead of DMY as the target variable did not improve the prediction. This holds for all periods (Figure 4). Predictions for the DMY improved with every added milk yield (Figure 4, blue curve). In contrast, predictions for the differenced milk yields did not improve when adding lagged milk yields (Figure 4, green and yellow curves).

Linear Temperature Models

We applied bootstrapped simple linear regression to the 5- to 90-DIM period on both breeds and all parities and found that the regression coefficient for temperature was negligible (mean of the bootstrap samples: 0.01; SD of the bootstrap samples: 0.01) compared with the average milk yield of 34.8 kg. The predicted milk yields for the lowest and highest measured temperatures, respectively, differed by 0.4 kg, a range that is of minor importance compared with the measured milk yield range of 58.6 kg. The outcomes for the differenced independent variables were similar (data not shown).

Table 7. Test root mean squared error (RMSE; kg) and relative RMSE (%) for models on daily milk yield with knowledge of previous milk yields through lagged milk yield features, with all models including DIM as a feature, performed on the data subsets of the 5- to 90-DIM period with the data split by breeds and parity ranges

		Lagged milk yield variables							
		0		1		2		3	
Breed[1]	Parity	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)	RMSE (kg)	RMSE (%)
Both	1	4.98	20.41	1.80	7.49	1.40	5.74	1.36	5.58
	2	4.26	11.75	2.55	7.04	2.24	6.18	2.17	5.99
	3+	5.69	15.10	3.05	8.09	2.65	7.02	2.58	6.84
	1+	7.97	24.21	2.44	7.43	2.12	6.44	2.07	6.30
BS	1	4.52	19.88	1.92	8.43	1.47	6.48	1.43	6.28
	2	4.03	11.62	2.18	6.27	1.97	5.68	1.93	5.56
	3+	5.03	13.67	3.25	8.82	2.80	7.61	2.70	7.33
	1+	7.81	24.52	2.49	7.83	2.18	6.84	2.13	6.69
FV	1	4.34	15.44	1.42	5.03	1.24	4.41	1.21	4.29
	2	3.80	9.84	2.98	7.73	2.55	6.62	2.49	6.45
	3+	6.66	16.06	2.24	5.40	2.09	5.04	2.04	4.93
	1+	7.71	21.66	2.33	6.55	2.03	5.71	1.98	5.56

[1]BS = Brown Swiss; FV = Fleckvieh.

DISCUSSION

The exploitation of daily herd management data allowed for satisfactory predictions of milk yield. We were able to predict DMY with an accuracy of 1.4 to 2.4 kg (RMSE) for the different data subsets of the period from 5 to 90 DIM (Table 6, penultimate column). This is sufficient to detect physiological challenges in dairy cows that have been reported to cause a 20% drop in milk yield (Spiers et al., 2004). Hereby, the best prediction model in our study considered cow's individual 3-d lagged milk yield. The prediction models showed consistent results across all periods. Considering meteorological parameters in the model did not improve the prediction outcome. Because temperature is not the best indicator of thermal comfort, we tested a variety of meteorological parameters (temperature, wind speed, rainfall, THI, cooling degree, barometric pressure), which all resulted in similar prediction errors. This finding indicates that moderate meteorologi-

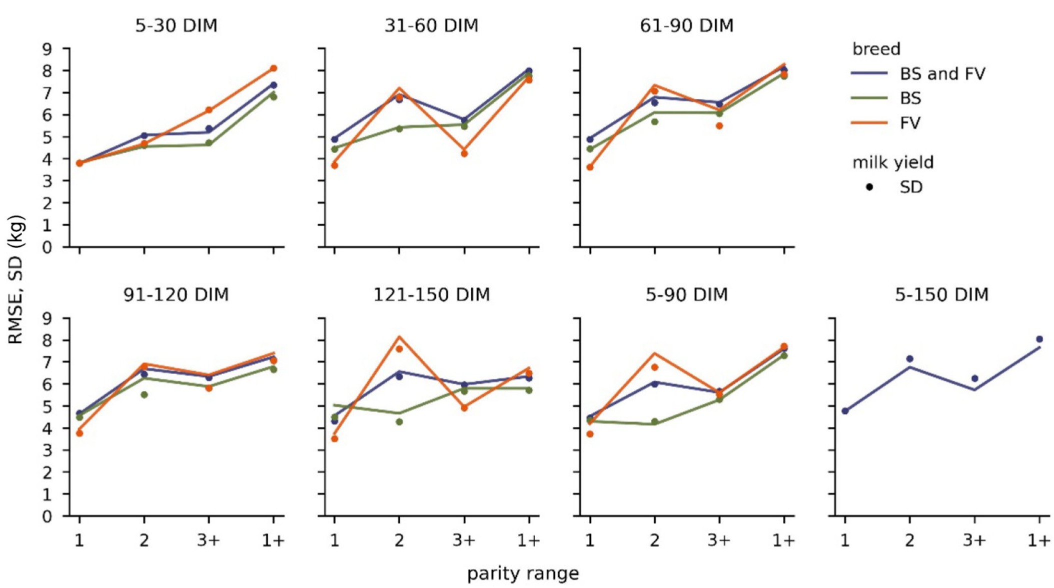

Figure 3. Comparison of validation root mean squared error (RMSE) of the baseline model $DMY \sim DIM$ and SD of daily milk yield (DMY) for all modeled data subsets. BS = Brown Swiss; FV = Fleckvieh.

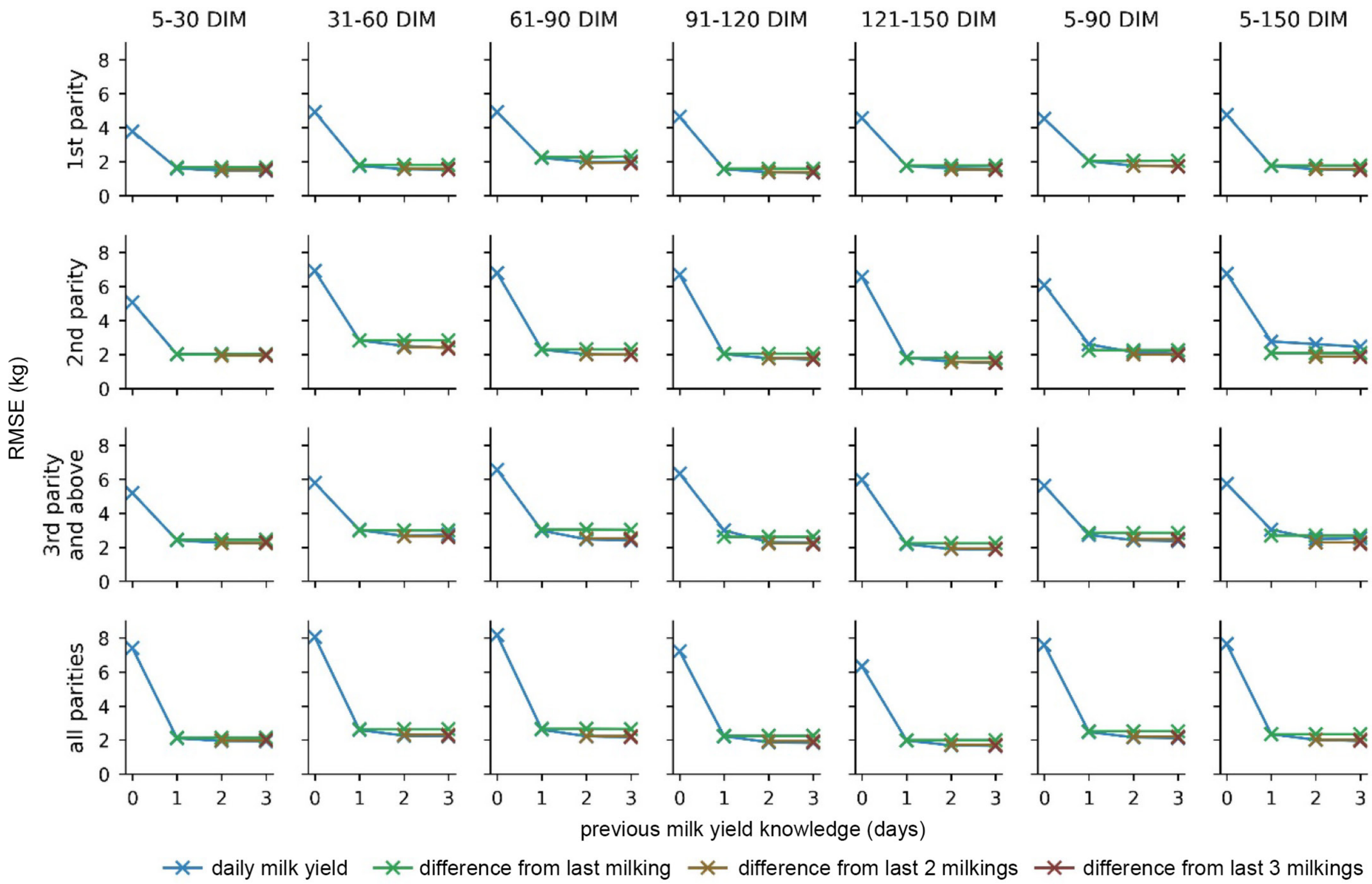

Figure 4. Validation root mean squared error (RMSE) vs. previous knowledge in number of days of lagged daily milk yield (for both breeds).

cal changes do not influence milk yield, and including this parameter in prediction models does not improve their outcome.

Prediction of Milk Yield

Murphy et al. (2014) predicted DMY at herd level for pasture-based cows with a relative RMSE of 6% for a 10-d prediction horizon. In the current study, we achieved the same accuracy for daily cow individual yield predictions but with a much shorter prediction horizon of 1 d and for barn-housed cows. Due to the different prediction horizons and farming types, a direct comparison of our study to that of Murphy et al. (2014) is difficult.

Furthermore, we found that considering lagged milk yield in the model improved prediction quality, which is in line with the findings of McParland et al. (2019), who found the last milking to be the best predictor of 24-h milk yield if the milking interval is also included in the model. Lee and Wardrop (1984) were able to predict milk yield from test-day data, where the models predicted 75 to 80% of DMY with a prediction error below 2 kg, whereas the current study achieved a prediction error of 1.4 to 2.4 kg for the different data subsets for individual DMY predictions. For these 2 studies, a direct comparison to our study is also difficult, as they present results for test-day milk yield prediction of 237 and 49 herds, respectively.

Influence of Climate

In the current study, we evaluated whether we could improve DMY predictions by including meteorological parameters in the prediction models to detect moderate influences of these parameters. However, including the THI, the cooling degree, wind speed, temperature, rainfall, or barometric pressure did not improve the accuracy of the model. These results are in line with a recent Swiss study reporting no effect of heat periods on annual milk revenues, veterinary expenses, or feed purchases (Bucheli et al., 2022). These results do not mean that meteorological parameters have no effect on milk yield prediction. Our line of discussion is two-fold. First, our calculations cover moderate temperature ranges, which present the focus of this study, but do not cover heat periods because the daily mean temperature exceeded 20°C in only 7% of the measurement period. Therefore, the data do not allow for evaluating predictions related to heat stress. Second, the milk yield considered in the model indirectly already incorporates the information on the meteorological variables by the cow herself, which indicates that meteorological data do not add a lot of new information to the model.

Yano et al. (2014) addressed the "non-linear nature of milk production and the individual differences in dairy cows" (p. 1) as the biggest challenges in creating models that are flexible enough to process this information, and they developed an extension of existing models to account for temperature effects. The authors modeled such effects of temperature on milk production in Holstein-Friesian cows and found that individual cows react differently to heat stress (Yano et al., 2014). They reported that susceptibility to heat stress increases with the cow's daily milk production (Yano et al., 2014). However, the authors also found that not all cows followed this profile, because individual cows' relative milk production increased with rising temperatures (Yano et al., 2014). In contrast, in our study, meteorological conditions did not affect higher-yielding cows more than lower-yielding cows.

Zhang et al. (2020) reported the individual RMSE for all the models in their study. Their best-performing model reached a mean RMSE of 3.13 kg (±0.74 kg SD; minimum: 2.12 kg; maximum: 5.32 kg) on 39 cows in the fourth parity for a 10-d prediction horizon. With our best-performing model, we reached a mean test RMSE of 1.94 kg (±0.84 kg SD; minimum: 0.68 kg; maximum: 4.11 kg) in the 5- to 90-DIM period for both breeds and all parities for a prediction horizon of 1 d. As stated earlier, the different prediction horizons and farming types make a comparison of our study to that of Zhang et al. (2020) difficult.

Practical Implications

Pontiggia et al. (2021) reported that stabling cows offered an effective way to reduce heat stress in dairy cows, compared with cows exposed to heat during grazing. In addition to the moderate temperature ranges covered in our study, the professional barn management provided the cows with adequate conditions, allowing them to cope with uncomfortable environments without developing physiological reactions. This situation indicates that the conditions at the research farm compensated for challenging meteorological conditions and further supports the hypothesis that moderate temperatures do not affect milk production. Bucheli et al. (2022) described Swiss dairy farms to be "robust to current heat exposure" (p. 304). For practical purposes, this means that providing adequate thermal comfort during heat periods allows farms to maintain good production levels.

Additionally, our research implies that the last day's milk yield is the most suitable parameter to predict milk yield. This means that farmers can take advantage of these consistent predictions without being required to install additional sensors.

Data Availability

The data are available on Zenodo (https://doi.org/10.5281/zenodo.7924864).

CONCLUSIONS

Our results suggest that a cow's individual previous day's milk yield allows prediction of milk yield with satisfactory accuracy. Including meteorological variables in the prediction model did not improve the prediction outcome for the underlying data set. This finding indicates that considering meteorological features in daily yield prediction models is not useful in moderate climates. We hypothesize that this information is indirectly contained in the lagged milk yield.

ACKNOWLEDGMENTS

This project was funded by the Swiss Federal Office for Agriculture (Bern, Switzerland) within the project "CowData" under the ICT-AGRI ERA-NET Call "Farm Data Integration–Key to Cattle Management Success," project ID 37729. We thank Markus Keller (Agroscope, Tänikon, Switzerland) for data collection and scraping and Christoph Bühler (Swiss Future Farm, Tänikon, Switzerland) and Thomas Anken (Agroscope. Tänikon, Switzerland) for fruitful discussions and input. The authors have not stated any conflicts of interest.

REFERENCES

Baehr, J., H. Schulte-Coerne, K. Pabst, and H. Gravert. 1984. Verhalten von Milchkuhen in Laufstallen. Zuchtungskunde. 56:127–138.

Bouraoui, R., M. Lahmar, A. Majdoub, M. Djemali, and R. Belyea. 2002. The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate. Anim. Res. 51:479–491. https://doi.org/10.1051/animres:2002036.

Bousquet, O., and D. Herrmann. 2002. On the complexity of learning the kernel matrix. Adv. Neural Inf. Process. Syst. 15:415–422.

Bucheli, J., M. Uldry, and R. Finger. 2022. Heat risks in Swiss milk production. J. Agric. Appl. Econ. Assoc. 1:304–319. https://doi.org/10.1002/jaa2.24.

Ferris, M. C., A. Christensen, and S. R. Wangen. 2020. Symposium review: Dairy Brain—Informing decisions on dairy farms using data analytics. J. Dairy Sci. 103:3874–3881. https://doi.org/10.3168/jds.2019-17199.

Gasser, L., F. Perez Cruz, and M. Cockburn. 2023. Supplemental File for "Can meteorological data improve the short-term prediction of

individual milk yield in dairy cows?" figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.23309519.v1.
Groher, T., K. Heitkämper, and C. Umstätter. 2020. Digital technology adoption in livestock production with a special focus on ruminant farming. Animal 14:2404–2413. https://doi.org/10.1017/S1751731120001391.
Grzesiak, W., P. Błaszczyk, and R. Lacroix. 2006. Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). Comput. Electron. Agric. 54:69–83. https://doi.org/10.1016/j.compag.2006.08.004.
Kimeldorf, G., and G. Wahba. 1971. Some results on Tchebycheffian spline functions. J. Math. Anal. Appl. 33:82–95. https://doi.org/10.1016/0022-247X(71)90184-3.
Kuss, M., C. E. Rasmussen, and R. Herbrich. 2005. Assessing approximate inference for binary Gaussian process classification. J. Mach. Learn. Res. 6:1679–1704.
Laevens, H., H. Deluyker, Y. H. Schukken, L. De Meulemeester, R. Vandermeersch, E. De Muêlenaere, and A. De Kruif. 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80:3219–3226. https://doi.org/10.3168/jds.S0022-0302(97)76295-7.
Lawrence, N., and A. Hyvärinen. 2005. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. J. Mach. Learn. Res. 6:1783–1816.
Lee, A. J., and J. Wardrop. 1984. Predicting daily milk yield, fat percent, and protein percent from morning or afternoon tests. J. Dairy Sci. 67:351–360. https://doi.org/10.3168/jds.S0022-0302(84)81308-9.
MacKay, D. J. C. 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press.
Mader, T. L., M. S. Davis, and T. Brown-Brandl. 2006. Environmental factors influencing heat stress in feedlot cattle. J. Anim. Sci. 84:712–719. https://doi.org/10.2527/2006.843712x.
Marumo, J., D. Lusseau, J. Speakman, M. Mackie, and C. Hambly. 2022. Influence of environmental factors and parity on milk yield dynamics in barn-housed dairy cattle. J. Dairy Sci. 105:1225–1241. https://doi.org/10.3168/jds.2021-20698.
Matthews, A. G. G., M. Van Der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman. 2017. GPflow: A Gaussian process library using TensorFlow. J. Mach. Learn. Res. 18:1–6.
Mbuthia, J. M., A. Eggert, and N. Reinsch. 2022. Comparison of high resolution observational and grid-interpolated weather data and application to thermal stress on herd average milk production traits in a temperate environment. Agric. For. Meteorol. 322:108923. https://doi.org/10.1016/j.agrformet.2022.108923.
McParland, S., B. Coughlan, B. Enright, M. O'Keeffe, R. O'Connor, L. Feeney, and D. P. Berry. 2019. Prediction of 24-hour milk yield and composition in dairy cows from a single part-day yield and sample. Ir. J. Agric. Food Res. 58:66–70. https://doi.org/10.2478/ijafr-2019-0007.
Menardo, S., G. Lanza, and W. Berg. 2021. The effect of diet and farm management on N_2O emissions from dairy farms estimated from farm data. Agriculture 11:654. https://doi.org/10.3390/agriculture11070654.
Murphy, M. D., M. J. O'Mahony, L. Shalloo, P. French, and J. Upton. 2014. Comparison of modelling techniques for milk-production forecasting. J. Dairy Sci. 97:3352–3363. https://doi.org/10.3168/jds.2013-7451.
NRC. 1971. A Guide to Environmental Research on Animals. National Academy of Sciences.
O'Hagan, A. 1978. Curve fitting and optimal design for prediction. J. R. Stat. Soc. B 40:1–24. https://doi.org/10.1111/j.2517-6161.1978.tb01643.x.
O'Leary, C., and C. Lynch. 2022. An evaluation of machine learning approaches for milk volume prediction in Ireland. Pages 1–8 in Proc. 2022 33rd Irish Signals and Systems Conference (ISSC). IEEE.
Osorio-Avalos, J., D. M. Garza-Camargo, L. E. Robles-Jimenez, J. Plaza, J. A. Abecia, C. Palacios, and M. Gonzalez-Ronquillo. 2022. Influence of Mediterranean climate and lunar calendar on milk production in Lacaune breed ewes. Int. J. Biometeorol. 66:1191. https://doi.org/10.1007/s00484-022-02268-4.
Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12:2825–2830.
Pontiggia, A., A. Münger, M. Holinger, S. Ammer, R. Bruckmaier, F. Dohme-Meier, and N. M. Keil. 2021. Auswirkung der Stallhaltung während der heissesten Tageszeit auf die Hitzestressbelastung von weidenden Milchkühen. (Effects of housing during the hottest part of the day on heat stress in grazing dairy cows.) Pages 107–115 in Current Work on Species-appropriate Animal Husbandry 2021. German Veterinary Society, Board of Trustees for Technology and Construction in Agriculture.
Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35–45. https://doi.org/10.1016/S0301-6226(03)00151-9.
Shock, D. A., S. J. LeBlanc, K. E. Leslie, K. Hand, M. A. Godkin, J. B. Coe, and D. F. Kelton. 2016. Studying the relationship between on-farm environmental conditions and local meteorological station data during the summer. J. Dairy Sci. 99:2169–2179. https://doi.org/10.3168/jds.2015-9795.
Spiers, D. E., J. N. Spain, J. D. Sampson, and R. P. Rhoads. 2004. Use of physiological parameters to predict milk yield and feed intake in heat-stressed dairy cows. J. Therm. Biol. 29:759–764. https://doi.org/10.1016/j.jtherbio.2004.08.051.
Sundrum, A. 2015. Metabolic disorders in the transition period indicate that the dairy cows' ability to adapt is overstressed. Animals (Basel) 5:978–1020. https://doi.org/10.3390/ani5040395.
Tao, S., R. M. Orellana Rivas, T. N. Marins, Y.-C. Chen, J. Gao, and J. K. Bernard. 2020. Impact of heat stress on lactational performance of dairy cows. Theriogenology 150:437–444. https://doi.org/10.1016/j.theriogenology.2020.02.048.
Williams, C., and C. Rasmussen. 1995. Gaussian processes for regression. Adv. Neural Inf. Process. Syst. 8:514–520.
Williams, C. K., and C. E. Rasmussen. 2006. Gaussian Processes for Machine Learning. Vol. 2. No. 3. MIT Press.
Wood, P. D. P. 1976. Algebraic models of the lactation curves for milk, fat and protein production, with estimates of seasonal variation. Anim. Sci. 22:35–40. https://doi.org/10.1017/S000335610003539X.
Yano, M., H. Shimadzu, and T. Endo. 2014. Modelling temperature effects on milk production: a study on Holstein cows at a Japanese farm. Springerplus 3:129. https://doi.org/10.1186/2193-1801-3-129.
Zhang, F., P. Shine, J. Upton, L. Shaloo, and M. D. Murphy. 2018. A review of milk production forecasting models: Past and future methods. Pages 14–61 in Dairy Farming Operations Management, Animal Welfare and Milk Production. Nova Science Publishers.
Zhang, F., J. Upton, L. Shalloo, P. Shine, and M. D. Murphy. 2020. Effect of introducing weather parameters on the accuracy of milk production forecast models. Inf. Process. Agric. 7:120–138. https://doi.org/10.1016/j.inpa.2019.04.004.

APPENDIX

Gaussian Processes

Gaussian processes (GP) for regression model the output $y \in \mathbb{R}$ as a nonlinear function of the inputs $\boldsymbol{x} \in \mathbb{R}^d$:

$$y = f(\mathbf{x}) + v, \quad [1]$$

where v is a Gaussian random variable. Gaussian process regression, given a labeled data set $\mathcal{D}_n = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$, not only estimates y for a given input $\boldsymbol{x}$ but also provides a full statistical description, namely:

$$p\left(y|\boldsymbol{x},\mathcal{D}_n\right). \quad [2]$$

Gaussian processes can be presented as a nonlinear regressor that expresses the input–output relation in Equation [1] by assuming that a real-valued function $f(\boldsymbol{x})$, known as a latent function, underlies the regression problem and that this function follows a GP. Before the labels are revealed, we assume this latent function has been drawn from a GP prior. Gaussian processes are characterized by their mean and covariance functions, denoted by $\mu(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x'})$, respectively. Even though nonzero-mean priors might be of use, working with zero-mean priors typically represents a reasonable assumption and simplifies the notation. The covariance function explains the correlation between each pair of points in the input space and characterizes the functions that can be described by the GP. For example, $k(\boldsymbol{x}, \boldsymbol{x'}) = \boldsymbol{x}^T\boldsymbol{x'}$ only yields linear latent functions and is used to solve Bayesian linear regression problems.

For any finite set of inputs $\mathcal{D}_n$, a GP becomes a multidimensional Gaussian distribution defined by its mean (zero in our case) and covariance matrix, $\left(\boldsymbol{K_n}\right)_{ij} = k\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right), \forall \boldsymbol{x}_i,\boldsymbol{x}_j \in \mathcal{D}_n$. The GP prior becomes

$$p\left(\boldsymbol{f}_n|\boldsymbol{X}_n\right) = \mathcal{N}\left(\boldsymbol{O},\boldsymbol{K}_n\right), \quad [3]$$

where $\boldsymbol{f}_n = [f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)]^T$ and $\boldsymbol{X}_n = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$. We want to compute the estimate for a general input $\boldsymbol{x}$ when the labels for the n training examples, denoted by $\boldsymbol{y}_n = [y_1, y_2, \ldots, y_n]^T$, are known. We can analytically compute Equation [2] by using the standard tools of Bayesian statistics: Bayes' rule, marginalization, and conditioning.

We first apply Bayes' rule to obtain the posterior density for the latent function:

$$p\left(f\left(\boldsymbol{x}\right),\boldsymbol{f}_n|\boldsymbol{x},\mathcal{D}_n\right) = \frac{p\left(\boldsymbol{y}_n|\boldsymbol{f}_n\right)p\left(f\left(\boldsymbol{x}\right),\boldsymbol{f}_n \mid \boldsymbol{x},\boldsymbol{X}_n\right)}{p\left(\boldsymbol{y}_n \mid \boldsymbol{X}_n\right)}, \quad [4]$$

where $p(f(\boldsymbol{x}), \boldsymbol{f}_n \mid \boldsymbol{x}, \boldsymbol{X}_n)$ is the GP prior in Equation [3] extended with a general input $\boldsymbol{x}$, $p(\boldsymbol{y}_n \mid \boldsymbol{f}_n)$ is the likelihood for the latent function at the training set, in which $\boldsymbol{y}_n$ is independent of $\boldsymbol{X}_n$ given the latent function $\boldsymbol{f}_n$, and $p(\boldsymbol{y}_n \mid \boldsymbol{X}_n)$ is the marginal likelihood or evidence of the model.

The likelihood function is given by a factorized model:

$$p\left(\boldsymbol{y}_n|\boldsymbol{f}_n\right) = \prod_{i=1}^{n} p\left(y_i \mid f\left(\boldsymbol{x}_i\right)\right), \quad [5]$$

because the samples in $\mathcal{D}_n$ are independent and identically distributed. In turn, for each pair $(f(\boldsymbol{x}_i), y_i)$, the likelihood is given by Equation [1]; therefore,

$$p\left(y_i|f\left(\boldsymbol{x}_i\right)\right) \sim \mathcal{N}\left(f\left(\boldsymbol{x}_i\right),\sigma_v^2\right). \quad [6]$$

A Gaussian likelihood function is conjugate to the GP prior, and hence the posterior in Equation [4] is also a multidimensional Gaussian distribution, which simplifies the computations to obtain Equation [2].

Finally, we can obtain the posterior density in Equation [2] for a general input $\boldsymbol{x}$ by conditioning on the training set and $\boldsymbol{x}$, and by marginalizing the latent function:

$$p\left(y|\boldsymbol{x},\mathcal{D}_n\right) = \int p\left(y|f\left(\boldsymbol{x}\right)\right)p\left(f\left(\boldsymbol{x}\right)|\boldsymbol{x},\mathcal{D}_n\right)df\left(\boldsymbol{x}\right), \quad [7]$$

where

$$p\left(f\left(\boldsymbol{x}\right)|\mathcal{D}_n,\boldsymbol{x}\right) = \int p\left(f\left(\boldsymbol{x}\right),\boldsymbol{f}_n|\boldsymbol{x},\mathcal{D}_n\right)d\boldsymbol{f}_n, \quad [8]$$

given the training data set, $\boldsymbol{f}_n$ takes values in $\mathbb{R}^n$ as it is a vector of n samples of a GP.

We have divided the marginalization into 2 separate equations to show the marginalization of the latent function over the training set in Equation [8] and the marginalization of the latent function at a general input $\boldsymbol{x}$ in Equation [7]. As mentioned earlier, the likelihood and the prior are Gaussians distributions, and therefore the marginalization in Equations [7] and [8] only involves Gaussian distributions. Thereby, we can analytically compute Equations [7] and [8] by using Gaussian conditioning and marginalization properties, leading to the following Gaussian density for the output:

$$p\left(f\left(\boldsymbol{x}\right)|\boldsymbol{x},\mathcal{D}_n\right) \sim \mathcal{N}\left(\mu_{f(\boldsymbol{x})},\sigma^2_{f(\boldsymbol{x})}\right), \quad [9]$$

where

$$\mu_{f(\boldsymbol{x})} = \boldsymbol{k}^T\boldsymbol{C}_n^{-1}\boldsymbol{y}_n \quad [10]$$

and

$$\sigma^2_{f(\boldsymbol{x})} = k\left(\boldsymbol{x},\boldsymbol{x}\right) - \boldsymbol{k}^T\boldsymbol{C}_n^{-1}\boldsymbol{k}, \quad [11]$$

with

$$\boldsymbol{k} = (k(\boldsymbol{x}_1, \boldsymbol{x}), k(\boldsymbol{x}_2, \boldsymbol{x}), \ldots, k(\boldsymbol{x}_n, \boldsymbol{x}))^T$$

and

$$C_n = K_n + \sigma_\nu^2 I_n.$$

The mean for $p(y \mid x, \mathcal{D}_n)$ is also given by Equation [10] (i.e., $\mu_y = \mu_{f(x)}$), and its variance is

$$\sigma_y^2 = \sigma_{f(\boldsymbol{x})}^2 + \sigma_\nu^2, \quad [12]$$

which, as expected, also accounts for the noise in the observation model.

Covariance Functions

In the previous section, we have assumed that the covariance functions $k(\boldsymbol{x}, \boldsymbol{x'})$ are known, which is not typically the case. In fact, the design of a good covariance function is crucial for a GP to provide accurate nonlinear solutions. The covariance function describes the relation between the inputs, and its form determines the possible solutions of the GP regression. It controls how quickly the function can change, or how the samples in one part of the input space affect the latent function everywhere else. For most problems, we can specify a parametric kernel function that captures any available information about the problem at hand. As already discussed, unlike kernel methods, a GP can infer these parameters, the so-called hyper-parameters, from the samples in $\mathcal{D}_n$ by using the Bayesian framework, instead of relying on computationally intensive procedures as cross-validation (Kimeldorf and Wahba, 1971) or learning the kernel matrix (Bousquet and Herrmann, 2002), as kernel methods need to.

The covariance function must be positive semidefinite because it represents the covariance matrix of a multidimensional Gaussian distribution. The covariance can be built by adding simpler covariance matrices, weighted by a positive hyper-parameter, or by multiplying them together, because the addition and multiplication of positive definite matrices yields a positive definite matrix. In general, the design of the kernel should rely on the information that we have for each estimation problem and should be designed to get the most accurate solution with the least number of samples. Nevertheless, the following kernel in Equation [13] often works well in signal processing applications:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \alpha_1 \exp\left(-\sum_{l=1}^{d} \gamma_d \left\| x_{il} - x_{jl} \right\|^2\right) + \alpha_2 \boldsymbol{x}_i^T \boldsymbol{x}_j + \alpha_3 \delta_{ij}, \quad [13]$$

where $\boldsymbol{\theta} = [\alpha_1, \gamma_1, \gamma_2, \ldots, \gamma_d, \alpha_2, \alpha_3]^T$ are the hyper-parameters.

The first term is a radial basis kernel, also denoted as RBF or Gaussian, with a different length-scale for each input dimension. This term is universal and allows construction of a generic nonlinear regressor. If we have symmetries in our problem, we can use the same length-scale for all dimensions: $\gamma_1 = \gamma$ for $l = 1, \ldots, d$. The second term is the linear covariance function. The last term represents the noise variance $\alpha_3 = \sigma_\nu^2$, which can be treated as an additional hyper-parameter to be learned from the data. We can add other terms or other covariance functions that allow for faster transitions, such as the Matérn kernel among others (Williams and Rasmussen, 2006).

If the hyper-parameters, $\boldsymbol{\theta}$, are unknown, the likelihood in Equation [5] and the prior in Equation [3] can be expressed as $p(\boldsymbol{y} \mid \boldsymbol{f}, \boldsymbol{\theta})$ and $p(\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{\theta})$, respectively, and we can proceed to integrate out $\boldsymbol{\theta}$. We have dropped the subindex n, because it is inconsequential and unnecessarily clutters the notation. First, we compute the marginal likelihood of the hyper-parameters of the kernel given the training data set

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{X}, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{f}. \quad [14]$$

Second, we can define a prior for the hyper-parameters, $p(\boldsymbol{\theta})$, that can be used to construct its posterior. Third, we integrate out the hyper-parameters to obtain the predictions. However, in this case, the marginal likelihood does not have a conjugate prior and the posterior cannot be obtained in closed form. Hence, the integration must be performed either by sampling or by approximations. Although this approach is well principled, it is computationally intensive, and it may not be feasible for some applications. For example, Markov-chain Monte Carlo methods require several hundred to several thousand samples from the posterior of $\boldsymbol{\theta}$ to integrate it out. Interested readers can find further details in Williams and Rasmussen (2006).

Alternatively, we can maximize the marginal likelihood in Equation [14] to obtain its optimal setting (Williams and Rasmussen, 1995). Although setting the hyper-parameters by maximum likelihood is not a purely Bayesian solution, it is fairly standard in the community, and it allows use of Bayesian solutions in time-sensitive applications. This optimization is nonconvex (MacKay, 2003), but, as we increase the number of training samples, the likelihood becomes a unimodal distribution around the maximum likelihood hyper-parameters, and the solution can be found using gradient ascent techniques. See Williams and Rasmussen (2006) for further details.