Internship Report:

Analysis of the variation of milk progesterone concentrations in dairy cows: relationship with milk production and reproductive performance

Wenting ZHANG

Toulouse School of Economics
MASTER IN STATISTICS AND ECONOMETRIC

Toulouse National Veterinary School
Reproduction Unit

Supervision:
Professor Sylvie CHASTANT-MAILLARD
Professor Nathalie VILLA-VIALANEIX
Dr Claire SABY-CHABAN
Mr. Rémi SERVIEN

August 2016

Contents

1 Basic professional context **1**

2 Background and Data Source **2**

2.1 Introduction 2

2.2 Data Description 7

2.2.1 Data source 7

2.2.2 Datasets 8

3 Content of Work **10**

3.1 Objectives of the work 10

3.2 Methodology 10

4 Input and insight **11**

4.1 Data Preparation 13

4.1.1 Imputation 13

4.1.2 Filters 15

4.1.3 LP and FP Estimation 18

4.1.4 Daily Progesterone Interpolation 24

4.1.5 Code Type of Cow 25

4.1.6 Final Data 27

4.2 Data Analysis 29

4.2.1 Univariate Description of Variables in LP Influential Factor 29

4.2.2 Influence of LP Cycle and Jours on lens 32

4.2.3 Other Variables' Influence on lens 34

4.2.4 Milk Production Analysis 35

5 Summary **36**

1 Basic professional context

ENVT (Ecole Nationale Vétérinaire de Toulouse) is one of four grand schools in France for the high training on French veterinarians. Each year, 500 vets graduate from this school.

In parallel of this teaching activity, ENVT develops research activities. L'UMR (Unité Mixte de Recherché 1225) of IHAP (Interaction hôtes-agents pathogènes) is a multidisciplinary research laboratory founded in 2003 in ENVT and under the joint supervision of INRA (Institut National de la Recherche Agronomique, one of the top agricultural research institute in the world). IHAP's works focus on emerging pathogens interaction with host and disease control in several species (bovine, small ruminants, horses, birds, dogs, pigs). IHAP develops various research approaches at the molecular to cellular level, and individual to population level. Their works contribute to a wide range of areas such as zoonosis, food safety, human and animal health, and are of great significance in both academics, industry and breeding.

Within IMM team (Régulation précoce des infections) in IHAP, Professor Sylvie CHASTANT-MAILLARD is involved in animal reproduction. She works on reproductive diseases in dogs, cats and cows with a special emphasis on the economic efficiency of dairy cows. In fact, the health of farm animals is one of the most important aspects to achieve economical benefits improvement. It not only increases the farm product quality but also accelerates the production procedure. She is assisted in this field by Dr Claire SABY, veterinarian specialized in bovine medicine and reproduction.

This 5-month internship (April-August 2016), on ENVT campus, is part of a large scale project aiming to describe and better understand global postpartum health status of dairy cows, combining data on metabolism, immunity, milk production and reproduction. The internship was focused specifically on the data related to reproduction, and more precisely on the cyclic variation of the ovarian function from data obtained on about 2510 cows.

To ensure the quality of data modelisation, besides the analysis of biological processes, the internship was co-supervised by Nathalie VILLA-VIALLANEIX (researcher, INRA, Mathématiques et Informatique Appliquées -Toulouse) and Rémi SERVIEN (researcher in biostatistics, INRA, UMR INRA-ENVT 1331 Toxalim). The basic biological knowledge (i.e. bovine reproduction, its impact on milk production economics, ovarian physiology) was taught me at the beginning of the internship. I also visited a dairy farm using the data recording system from which the dataset was built. Thereafter, we worked on the dataset : biological questions were defined, I found the appropriate methods of data analysis; every 2-3 weeeks, I presented my work and results to the four supervisors.

2 Background and Data Source

2.1 Introduction

The final objective of the dairyman is to maximize the total milk production obtained from each cow over her career. The basic requirement for a cow to produce milk is to have calved. The start of a milk production cycle (called « lactation») is marked by a calving event (first blue dot on figure 1). Milk production (red line) begins immediately after the birth of the calf, increases during the two first months of lactation, and then progressively decreases (-10% per month in mean). During this early postpartum period, the ovaries of the cow are first inactive (it is spoken of «postpartum anestrus»), the cow does not ovulate, thus cannot be inseminated. At a time variable between females, ovarian activity will resume: the cow expresses a specific behaviour during around 14 hours («estrus», also called heat), signal for the dairyman to perform artificial insemination. If the insemination is not successful (no pregnancy), the cow will come into estrus 19-25 days later and will be re-inseminated at each estrus until pregnancy is finally achieved. Pregnancy in the bovine lasts around 9 months, after which calving occurs (round point; figure 1). Two months before the expected time of calving, lactation is voluntary stopped by the dairyman in order to allow the cow to accumulate fat reserves for next lactation together with cell renewal within the mammary gland. The 2 months period, during which milk production is stopped, is called « dry-off » (blue inverted triangle; figure 1).

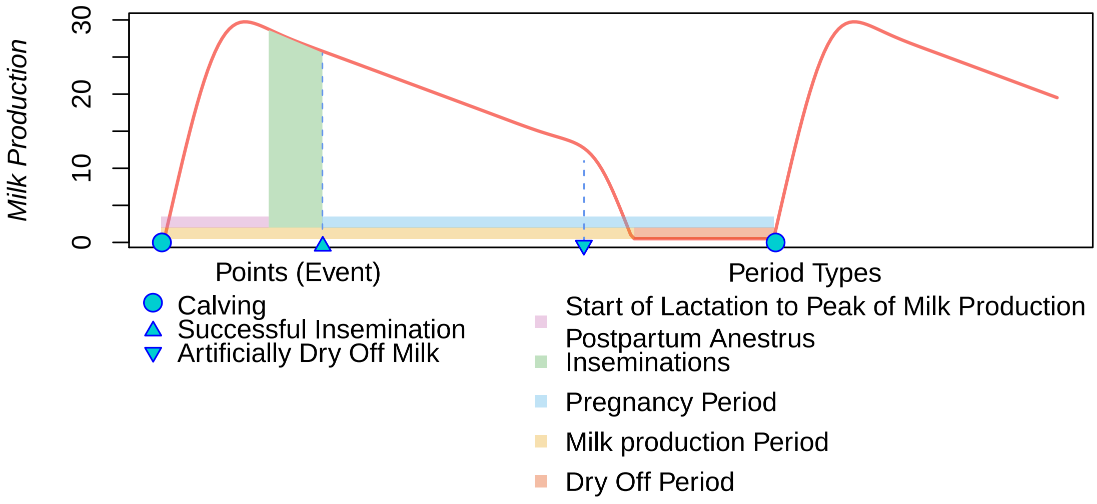


Figure 1: Life Periods of the Cow

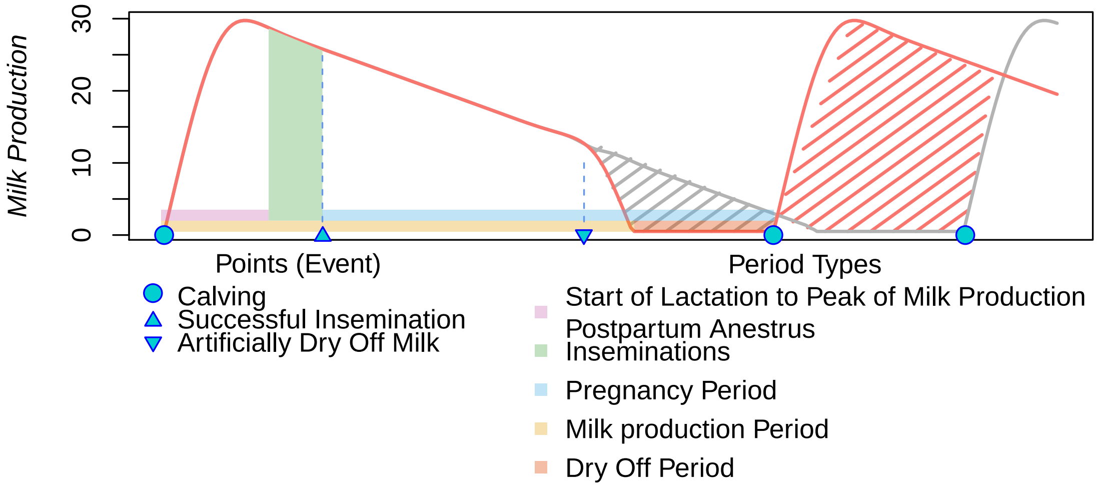


Figure 2: Comparison of Milk Production
Depending on Time of Successful Insemination

From an economical point of view, the more rapidly the cow can be get pregnant after calving, the more important the milk production over the cow career (figure 2 compares red lactation curve from a rapidly pregnant cow to the grey lactation curve from a cow that took more time to get pregnant). The optimum is considered to have a cow successfully inseminated between 45 and 90 days after calving, giving a 10 month lactation (Inchaisri et al, 2011).

It is thus crucial that:

1. The ovarian activity can resume as early as possible after calving.
2. Once resumed, the ovarian cyclicity has to be regular (one estrus every 19-25 days); otherwise, success rate of later inseminations will be decreased (Lamming and Darwash, 1998; Royal et al, 2000).
3. The estrus period has to be detected by the dairyman. Estrus detection, strictly required to inseminate cows, is currently one of the major challenges of the dairyman: specific estrus behavior is not only expressed during a short period of time (14 hours every 21 days) but also weakly expressed by most dairy cows. To obtain satisfactory estrus detection rates, 3 observation periods of 20 minutes each are necessary. Estrus detection is thus not only difficult (a mean of 50% of estruses only is detected) but also a time-consuming task (Diskin and Screenan, 2000).

The importance and the difficulties of estrus detection in dairy cows explain the development of some estrus assistance systems. One of the most potent one (HerdNavigator, Foss-Delaval) is based on repeated progesterone assay in milk. Progesterone is an hormone specifically synthetized by the ovary after ovulation (figure 3). During follicular period, before ovulation,

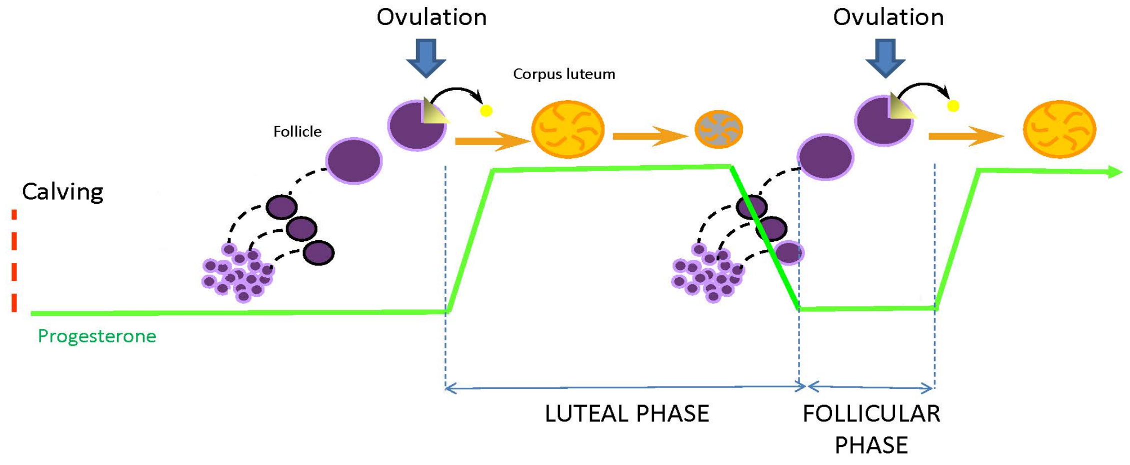
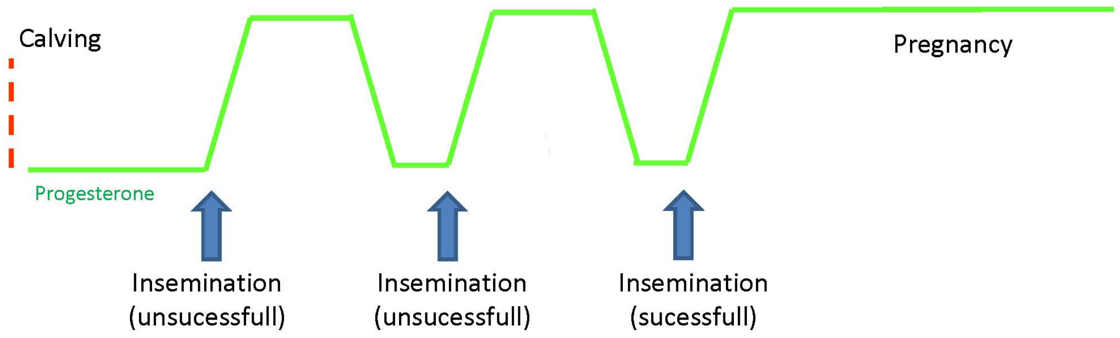
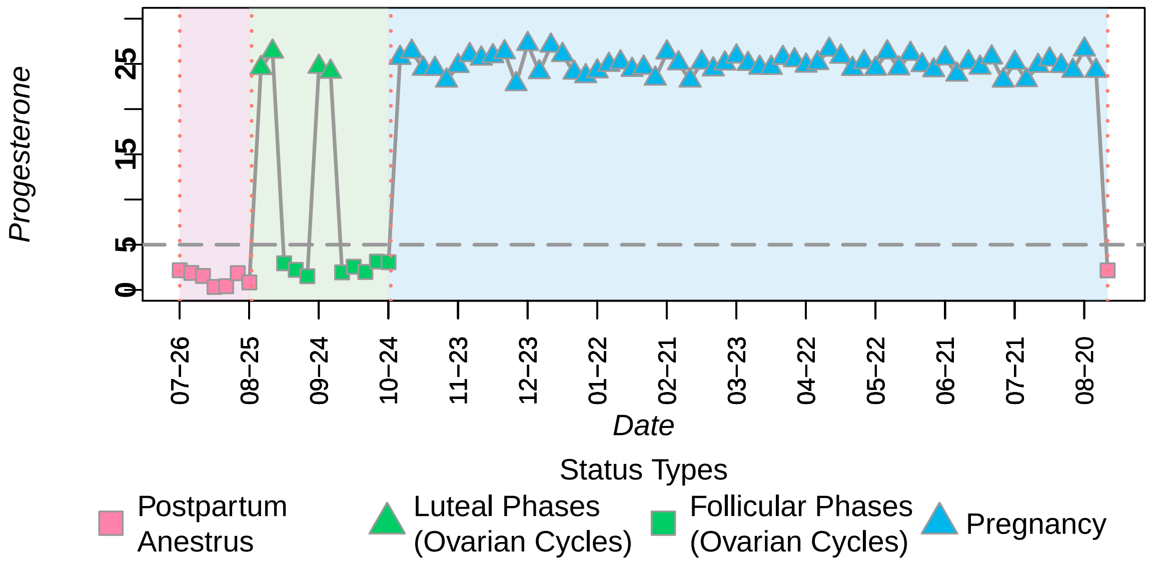


Figure 3: Reproductive Statuses of Cow


Calving
Pregnancy
Progesterone
Insemination (unsucessfull)
Insemination (unsucessfull)
Insemination (sucessfull)

(a) Schematic



(b) Real Curve

Figure 4: Progesterone pattern after calving

no progesterone is produced; after ovulation, progesterone is produced by a specific ovarian structure called « corpus luteum » during the « luteal phase ». In absence of fertilization, corpus luteum will spontaneously disappear (« luteolysis »), progesterone level will drop, the cow entering into a new « follicular phase », before a new ovulation. Repeated progesterone assays in milk allow detecting this drop in progesterone level, indicative of coming estrus within the following days (figure 3). If pregnant, the corpus luteum will persists until next calving, and progesterone thus remains high during 9 months (figure 4; table 1).

The major asset of this system is to be non-invasive (assay is performed on milk, moreover collected by robotized automated milking) and global (all cows can be assayed, since they are milked 2-3 times a day). It is thus a potent scientific tool to study ovarian activity and reproductive performances (delay from calving to insemination success for example) on large numbers of animals. The system simultaneously assays two other parameters: lactate dehydrogenase (LDH), enzyme activated by white blood cells and indicative of the inflammatory status of the animal (namely of udder health) — beta hydroxybutyrate (BHB), residue of the energetic metabolism indicative of a negative energy balance (table 2). These two parameters will not be considered during the internship.

Table 1: Reproductive Statuses in Milk Production Cycle Defined by Milk Progesterone Concentration[1]

Reproductive Status	*Definition*
Postpartum Anestrus	The first several days after calving when progesterone is lower than 5 ng/ml
Luteal Phase (LP)	The period when progesterone is higher than 5 ng/ml and lasts less than 9 month. Length of LP > 3 days
Follicular Phase (FP)	The period when progesterone is lower than 5 ng/ml and happens after a luteal phase. Length of FP > 2 days. Insemination should be performed within FP
Ovarian Cycle (OC)	A luteal phase plus the follicular phase next to it
Pregnancy Period	A 9 months period when progesterone stays higher than 5 ng/ml

[1] In each milk production cycle, there are only one postpartum anestrus and one pregnancy period, but can have more than one ovarian cycle, luteal phase and follicular phase.

Table 2: Parameters Checked by HerdNavigator

Name	*Definition and Usage*
Progesterone	Produced by the ovaries, placenta, and adrenal glands. It is the "gold standard" used to detect reproduction status such as heat, pregnancy.
LDH (Lactate dehydrogenase)	A specific measure of udder's health. LDH is an intracellular enzyme and inflammation episode in the udder will destroy the cells and release the LDH to milk.
BHB (Beta hydroxybutyrate)	The most common ketone body in dairy cow used by muscle and nervous tissue. Excessive BHB ketosis is indicative of a negative energy balance (more energy required for life and lactation than energy ingested).

2.2 Data Description

2.2.1 Data source

The data was collected in French dairy farms. Every farm owns an independent HerdNavigator system. Records were collected by HerdNavigator system and then the system management software stored the data on local PC in farms. Old data will be overwritten thus vet should visit farms regularly to export them. Raw data came from 2510 cows in 23 French farms all milking the most common dairy breed (Holstein). Table 3 shows the data collocation duration and number of observations in each farm. First date is the date when HerdNavigator started to run (first date will be the last date when data was collected, in new data in future) and last date is the date when Claire SABY exported the data. The number of cows varies a lot from farm to farm, ranging from 64 to 203.

Table 3: Summary of Information by Farm [1]

Farm	*First Date*	*Last Date*	*Num of Progesterone*	*Num of Cows*
1	2014-03-17	2015-04-22	46276	203
2	2014-03-11	2015-04-16	34517	128
3	2014-05-20	2015-05-11	23828	89
4	2014-04-13	2015-05-19	17397	75
5	2014-05-21	2015-05-18	14022	71
6	2014-05-20	2015-05-22	27544	120
7	2014-04-10	2015-05-10	14730	71
8	2014-05-15	2015-06-15	29013	116
9	2014-04-20	2015-05-25	18239	69
10	2014-11-06	2015-05-29	13653	85
11	2014-09-02	2015-06-01	29287	183
13	2014-04-27	2015-06-02	15890	69
14	2014-04-28	2015-06-03	27917	139
15	2014-04-28	2015-06-02	19894	84
16	2014-04-27	2015-06-01	29255	124
17	2014-04-29	2015-06-03	21600	106
18	2014-04-29	2015-06-04	24798	108
19	2014-12-03	2015-06-04	14205	119
20	2014-04-27	2015-06-01	39902	189
21	2014-05-06	2015-06-10	22607	93
22	2014-05-11	2015-06-15	27923	96
23	2014-05-11	2015-06-16	21030	109
24	2013-11-27	2015-06-16	20165	64
Total	**2013-11-27**	**2015-06-16**	**553692**	**2510**

[1] In each milk production cycle, there are only one postpartum anestrus and one pregnancy period, but can have more than one ovarian cycle, luteal phase and follicular phase.

2.2.2 Datasets

Two raw data sets, named as “cows” and “rapport production lait”, are in CSV format.

Data set “cows” covers the aspects of cow we needed, including value of three parameter measurements (progesterone, BHB, LDH), health statuses and reproductive statuses (estrus, pregnancy, etc). It has 553692 records and 26 variables: some records are in the same date because although a record stores information of one indicator, it is possible that HerdNavigator may check several indicators a day. Table 4 is the first three rows of data "cows".

Data set “rapport production lait” contains 24 data tables. Every data table is the daily milk production of a farm. The names and the number of variables in tables are not completely consistent. Since “rapport production lait” is used to update the milk production data, we will only use 8 variables (animal, lactation, Date, jours, prodJ, prod7J) in later step. There are 714826 records in data set “rapport production lait”. Table 5 is the first three rows of Data “rapport production lait”.

Table 6 shows the description of variables in data set "cows". "Error:512" means HerdNavigator breaks down and NA means not applicable. The definitions of variables in data set “rapport production lait” are exactly the same as in data set “cows”, thus we don’t need a new data description for data set “rapport production lait”.

Table 4: First three records in Data "cows"

id	elevage	IndexHN	Date	animal	jours	lactation	prodJ	prod7J	prog_raw	prog_smooth
1-6	1	17/03/2014-6	17/03/2014	6	194	2	37,06	52,93	0.00	0.00
1-14	1	17/03/2014-14	17/03/2014	14	154	2	53,86	49,03	0.00	0.00
1-15	1	17/03/2014-15	17/03/2014	15	139	2	66,05	56,83	10.59	10.49

HN_BHB_raw	HN_BHB_smooth	HN_LDH_raw	HN_LDH_smooth	HN_alarme	chaleur	Evenement_IA
0.00	0.00	19.81	18.78	Decochee		
0.00	0.00	20.19	12.86	Decochee		
0.00	0.00	0.00	0.00	Decochee		

gestation	IAF	prob_chaleur	IAnum	h_insemination	HN_Type_biometric	HN_risk	HN_diagnostic
					LDH	9	Mammites
					LDH	13	Mammites
					Progesterone	0	Kyste folliculaire

Table 5: First three records of Farm 1 in
Data “rapport production lait”

id_new	Date	animal	jours	lactation	prodJ	prod7J
1-4-2	17/03/2014	4	206	2	31.75	34.55
1-4-2	17/03/2014	4	206	2	31.75	34.55
1-6-2	17/03/2014	6	194	2	37.06	52.93

Table 6: Data Description of Additional Variables

	Variable	*Explanation*	*Value*	*Type*	*Source*
1	id	Cow's id (unique for every cow); There are 2510 different id	id = "elevage-animal"; 1-6, 1-14, 1-15 ...	Character	HerdNavigator
2	elevage	id of farm	1, 2 ... 11, 13 ... 24	Numeric	Claire SABY
3	IndexHN	Index of record given by HerdNavigator in each farm; There are 346358 different values, which means some cows have several records at the same date	IndexHN = "Date - animal". For example: 17/03/2014-6, 17/03/2014-14 ...	Character	Claire SABY
4	Date	Date when data was collected	From 27 Nov. 2013 to 16 June 2015	Character	HerdNavigator
5	animal	Index of cow	Integers less than 5. One exception: a cow coming from farm 24 has the animal index "54574"	Integer	HerdNavigator
6	jours	Number of days after calving. In later part, the date of calving is called as **day 0** and the date of N days after calving date is called as **day N**	From 0 to 642 and with 350 "Err:512"	Character	HerdNavigator
7	lactation	Milk production cycle	From 1 to 9 with 300 "Err:512" and 293 ""	Character	HerdNavigator
8	prodJ	Daily milk production; If the record is used to check an indicator that a cow has health problem or an AI, no milk production appears in this record	From 0.8 to 120.05 liters with 358 "Err:512" and 1 missing value	Character	HerdNavigator
9	prod7J	Average daily milk production within previous 7 days	From 0 to 75.4 liters with 358 "Err:512" and 1 missing value (due to the "Err:512" and "" in prodJ)	Character	HerdNavigator
10	prog_raw	Progesterone concentration (ng/ml) checked by HerdNavigator from milk samples. Both 0 and NA mean progesterone didn't be checked in these records	From 0 to 28 ng/ml with around 70% of prog_raw are 0 and 7349 are NA	Numeric	HerdNavigator
11	prog_smooth	Smoothed progesterone concentration (ng/ml) automatically calculated by HerdNavigator, way of smoothing because patented	From 0 to 27.8 ng/ml with 7349 NA (due to the NA in prog_raw)	Numeric	HerdNavigator
12	HN_BHB_raw	BHB concentration (nmol/L) assayed by HerdNavigator from milk samples. NA mean BHB was not assayed from milk samples taken	From 0 to 3.9 nmol/L with 7349 NA	Numeric	HerdNavigator
13	HN_BHB_smooth	BHB concentration (nmol/L) automatically smoothed by HerdNavigator, way of smoothing because patented	From 0 to 1 nmol/L with 7349 NA (due to the NA in HN_BHB_raw)	Numeric	HerdNavigator
14	HN_LDH_raw	LDH concentration (UI/L) assayed by HerdNavigator from milk samples. NA means LDH was not assayed from milk samples taken	From 0 to 427 UI/L with 7349 NA	Numeric	HerdNavigator
15	HN_LDH_smooth	LDH concentration (UI/L) automatically smoothed by HerdNavigator,way of smoothing because patented	From 0 to 348 UI/L with 7349 NA (due to the NA in HN_LDH_raw)	Numeric	HerdNavigator
16	HN_alarme	When cow is checked by HerdNavigator and it considered to be in heat, an heat alarm will be issued by HerdNavigator to remind dairyman performing AI	Different values have the same meaning due to the default setting of HerdNavigators. There are 5 values: (1)"Cochee" and "Oui" mean heat alarm issued; (2)"Decochee", NULL and " " mean unchecked or no heat;	Character	HerdNavigator
17	chaleur	Heat is confirmed manually by dairyman. There are 5363 confirmed heat	Two values: (1)"Oui" means be in heat (2)"" means not in heat	Character	Dairyman
18	Evenement_AI	Dummy variable whether AI is performed or not There are many reasons that dairyman gives up AI even with heat confirmed: maybe the cow is unhealthy, or too early after calving or programmed to be culled; There are 5117 AI	Two values: (1)"Oui" means insemination takes place; (2)"" means that insemination was not performed;	Character	Dairyman
19	gestation	Dummy variable indicating whether cow get pregnant or not There are 2017 "Positif", 328 "Negatif", 32 "Incertain", 551315 unchecked	Four values: (1)"Positif" means cow is checked by farmer and farmer thinks it get pregnant; (2)"Negatif" means cow is checked but doesn't get pregnant; (3)"Incertain" means cow is checked and HerdNavigator not sure of its pregnancy status; (4)"" means cow is unchecked;	Character	Dairyman
20	IAF	Dummy variable indicating the success of AI	Four values:"", "Oui", "Oui?" and "Non". Detail meanings of "" and "Oui?" are unknown	Character	Claire SABY
21	prob_chaleur	Probability (%) of being in heat	From 0 to 83, if no heat alarm during this day, prob_chaleur is NA.	Integer	HerdNavigator
22	IAnum	Number of times the cow has been artificially inseminated. NA means cow wasn't inseminated before.	From 1 to 13	Integer	Dairyman
23	h_insemination	unknown variable	/	Character	
24	HN_Type_biometric	Type of parameters. There are 153715 progesterone, 78362 BHB, 314266 LDH and 7349 NA.	4 types: "Progesterone", "LDH", "BHB", "";	Character	HerdNavigator
25	HN_risk	Risk of disease based on BHB or LDH (%)	From 0 to 100	Integer	HerdNavigator
26	HN_diagnostic	The disease the cow has or status of cow		Character	HerdNavigator

3 Content of Work

3.1 Objectives of the work

Based on HerdNavigator obtained from 553692 progesterone data of 2510 cows in 23 French herds, the objectives were:

1. To fill progesterone with interpolated values and compare methods.

2. To characterize ovarian cyclicity resumption in dairy cows in commercial (non experimental) conditions: timing of first ovulation (first progesterone rise), postpartum cyclicity profiles (LP and FP lengths regularity of progesterone cycles).

   - Does LP's length differ between LP cycles?
   - Does number of days after calving have impact on LP's length? If it does, how?
   - Other possible influential factors?

3. To identify type of cyclicity profile before insemination.

4. To identify feature of LP length such as distribution of LP's length.

Also it is expected to use data exploration techniques to discover and move to the other interesting topics related to milk production and reproduction. The core responsibilities of internship are divided into two parts: Data preparation and data analysis.

3.2 Methodology

The data preparation and data analysis were on the basis of a R programming.

Coding scripts are saved in 3 folders: Data, Preparation and Analysis. Raw data is stored in "∼/Data/Input/". In data preparation folder, there are 23 scripts. Each script is a step and the step order can be found in script's name. The data created in each step is saved in sub folders of "∼/Data/Output/". There are 3 sub folders (OLS, Panel data, Test) in Analysis folder and scripts are grouped into sub folders by the questions answered and methods applied to.

From script "0. Install package and set environment", we can see all packages installed and loaded. This is a pre-preparation step automatically checking and installing packages. Customer-defined functions are also used and they are at the top of the script it referred.

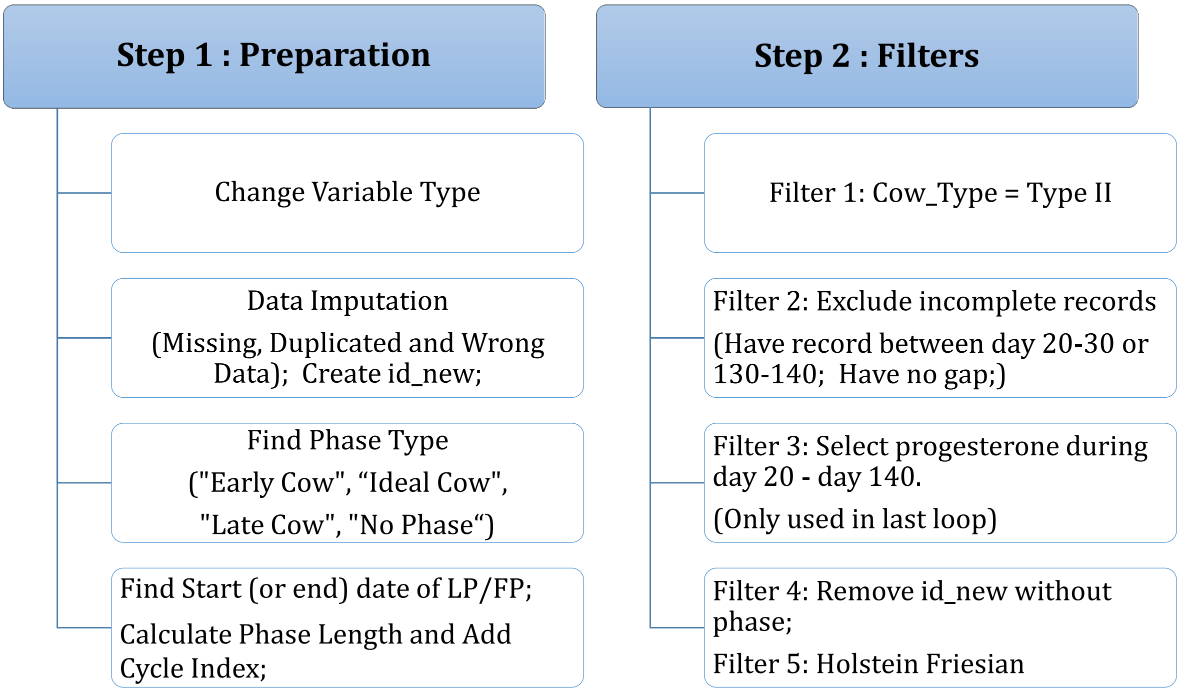
The main packages and functions used in data preparation section are: {data.table}, {plyr}, {dplyr}, {zoo} and as.POSIXct{lubridate}. And for data analysis, dunn.test{dunn.test}, kruskalmc{pgirmess} and are used to perform statistic tests, lm{stats} were used to solve linear model, plm{plm} was used in panel data model.

4 Input and insight

We first change the type of variables and set the uniform default value (for example: Err:512, NA and “” are missing values in different HerdNavigator). Then we delete the cows which have only one or two records, because a complete LP or FP should have at least three records and two records cannot tell us anything. In this case, cow “19-5528” and “7-2645” are deleted.

The main work of data preparation are: creating a new id (variable: id_new) and filtering the data by a series of biological criterion.

The flow chart summarizes the logic of coding in data preparation section:



Step 1 : Preparation

Change Variable Type

Data Imputation
(Missing, Duplicated and Wrong Data); Create id_new;

Find Phase Type
("Early Cow", “Ideal Cow",
"Late Cow", "No Phase“)

Find Start (or end) date of LP/FP;
Calculate Phase Length and Add Cycle Index;

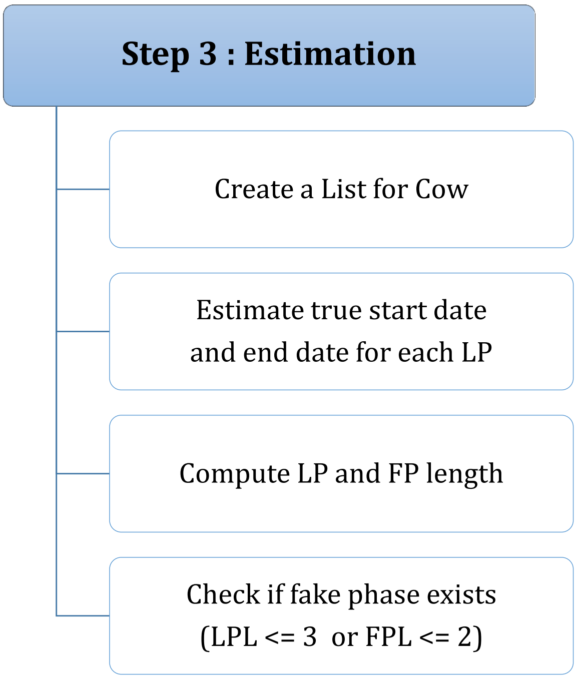
Step 2 : Filters

Filter 1: Cow_Type = Type II

Filter 2: Exclude incomplete records
(Have record between day 20-30 or 130-140; Have no gap;)

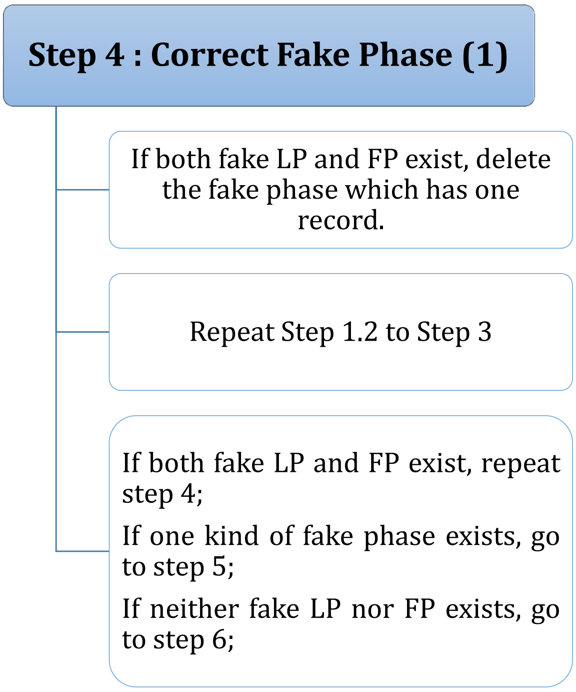
Filter 3: Select progesterone during day 20 - day 140.
(Only used in last loop)

Filter 4: Remove id_new without phase;
Filter 5: Holstein Friesian



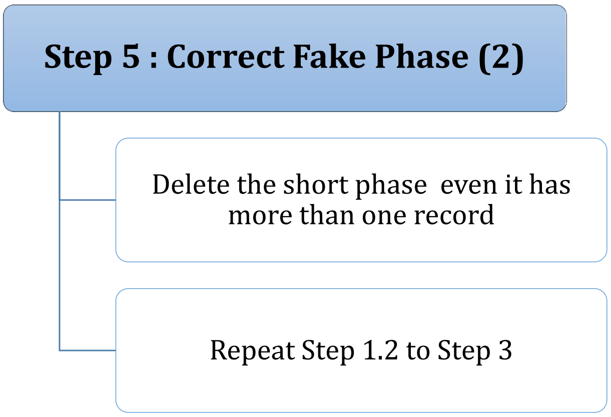
Step 3 : Estimation

- Create a List for Cow
- Estimate true start date and end date for each LP
- Compute LP and FP length
- Check if fake phase exists (LPL <= 3 or FPL <= 2)



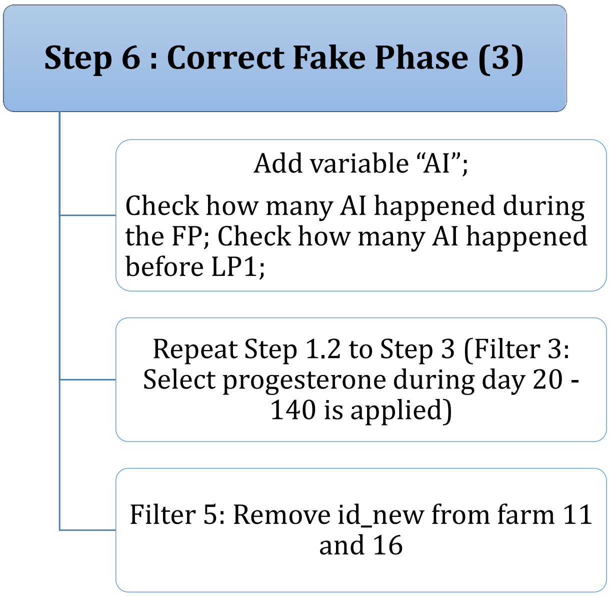
Step 4 : Correct Fake Phase (1)

- If both fake LP and FP exist, delete the fake phase which has one record.
- Repeat Step 1.2 to Step 3
- If both fake LP and FP exist, repeat step 4;
  If one kind of fake phase exists, go to step 5;
  If neither fake LP nor FP exists, go to step 6;



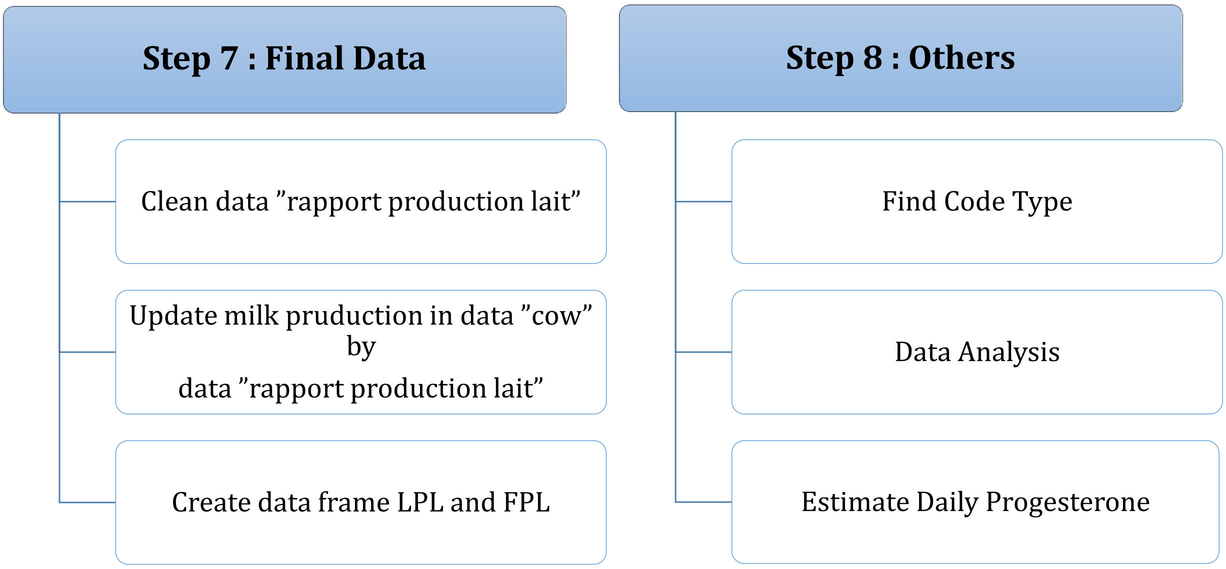
Step 5 : Correct Fake Phase (2)

- Delete the short phase even it has more than one record
- Repeat Step 1.2 to Step 3



Step 6 : Correct Fake Phase (3)

- Add variable "AI";
  Check how many AI happened during the FP; Check how many AI happened before LP1;
- Repeat Step 1.2 to Step 3 (Filter 3: Select progesterone during day 20 - 140 is applied)
- Filter 5: Remove id_new from farm 11 and 16


Step 7 : Final Data
Clean data ”rapport production lait”
Update milk pruduction in data ”cow” by data ”rapport production lait”
Create data frame LPL and FPL
Step 8 : Others
Find Code Type
Data Analysis
Estimate Daily Progesterone

4.1 Data Preparation

4.1.1 Imputation

(1) Data "cows.CSV"

Missing values. Because we are interested in progesterone concentration within every OC, thus individual in our analysis is no longer a cow but a cow in a specific milk production cycle. A new variable id_new which is in “elevage-animal-lactation” format should be created. As mentioned in table 6, there are 300 "Err:512" and 293 "" in 72 cows’ lactation value which are all missing values. Thus to create id_new, we should replace missing values first.

Missing values arise for many reasons (for example table 7). The default lactation value for those cows who have never lactated before is not 0 but a missing value. Besides, some missing values are due to duplicated records. In some situations, data is collected more than once per day, it is possible that only one record has lactation value. Also the break down of HerdNavigator is a possible source.

To fill missing values in lactation, first, we found milk production cycles by creating a new variable Calving_Date. Calving date is the start of a milk production, and as the definition previously mentioned, calving date (day 0) is the date of a record when its value of variable "jours" equals to 0. In the second step, we filled the missing values based on the idea that within a milk production cycle, lactation value should be unique. Third, since step 2 can only

Table 7: Three Sources of Missing Values in lactation

(a) No Pregnancy Before

id	*Date*	*lactation*
24-6246	2015-06-15	1
24-6247	2013-11-30	NA
24-6247	2014-01-16	NA
24-6247	2014-02-04	NA
24-6247	2014-11-24	1

(b) Duplicated Data

id	*Date*	*lactation*
09-3760	2015-02-20	1
09-3760	2015-02-20	1
09-3760	2015-02-22	NA
09-3760	2015-02-22	NA
09-3760	2015-02-23	1

(c) Break Down

id	*Date*	*lactation*
21-6613	2014-07-09	1
21-6613	2014-07-10	1
21-6613	2014-07-15	NA
21-6613	2014-07-16	NA
21-6613	2014-07-19	1

replace missing values in case 7b and 7c, thus we checked and ensured that the rest of 287 missing values were in case 7a which could be deleted directly. Finally, there were 553392 records remain and new variable id_new was created. In the rest of report, the individual is no longer id but id_new. Each id_new can be treated as a new cow and now every cow has one milk production cycle at most.

As mentioned before, jours is needed in later analysis but there are 350 missing values in variable jours. We filled the missing values by computing it as the difference between variable Date and variable Calving_date.

The missing values in variables prodJ and prod7J can be replaced by milk production data in “rapport production lait”.

Duplicated data: Because several measurements are made at the same day or progesterone may be checked more than once per day, we need to select the progesterone records and then combine duplicated progesterone records to make sure that there is only one record each day. The way of merging the duplicated records depends on the type and meaning of the variable. There are 153714 progesterone records, we used variables "id_new” and "Date" to identify the duplicated records. For these duplications, we took average for numeric variables, merged the data if the variable type was character and reassembled the logic variable by if there was at least one TRUE in duplication the logic was TRUE otherwise it was FALSE. Finally we kept 3142 cows with 134002 records.

Wrong data: There is an error in data: variable jours is not consistent with variable Date. More precisely, lactation should have one and only one unit of increase in the next milk production cycle. If grouping the records by id_new and sorting the data in date order, it is surprised that not all of jours are in non-decreasing order. Actually, this is due to the reason that dairyman had entered a wrong calving date at first and then he realized the mistake and entered at least one new calving date. Thus variable jours should be recompute. In this case, the corresponding lactation, id_new were corrected at the same time. Note that we need to look into the data case by case (see table 8) and choose an appropriate way to correct. Thus the code in this part is not a general solution.

Table 8: The Cow Having Wrong Data

id_new	*Way to correct error*
01-3010-1	Delete first 2 lines
01-3015-1	Delete first 2 lines
06-9967-4	Delete first four lines and correct calving date and jours in line 5-8
09-1886-2	Correct calving date and jours in lines with error
15-0551-2	Delete last lines
16-8673-1	Delete first four lines and correct lactation(=2) and id_new in last line
19-5390-2	Correct calving date and jours in error lines
20-0691-2	Correct lactation and id_new in last line

(2) Data "rapport production lait.CSV"

Data “rapport production lait” is more easy to deal with. We made the duplicated rows unique and then found that there were 363 errors (different milk production in the same date due to one unit less input of date) in variable Date and 154 errors (jours is calculated wrong thus it is not inconsistent with Date) in variable jours. After correcting the errors, we computed the total amount of milk quantity during day 0 and day 140 for each cow and average the milk quantity within 7 days before each LP.

4.1.2 Filters

In this section, we filtered the records of data "cows" with a series of biological criterion:

Criteria 1: Cow_Type = Type II.
Criteria 1 restricts the duration of records to ensure the comparability of cow. There are three exclusive types (figure 5). First date and last date have the same definitions as in table 3. Type I cow starts to produce milk before First date, thus HerdNavigator has no information about when it began to lactate thus data is left truncated. Type II cow is well recorded by HerdNavigator, the data collection begun before its day 0 and ended after its day 140. Type III cow starts to produce milk very late (several days before we stop to collect data), thus data is right truncated. We only want type II cows.

The reason why we choose day N = 140 as a threshold in definition is that: If there is no upper bound for N, few of cows will remain, while if, in turn, there is very small lower bound, we will cut the majority of data and loss many LPs. N = 140 was chosen because the mean interval between calving and first artificial insemination in French Holstein cows is 144 days (Le Mezec et al, 2014). After criteria 1, 1354 cows with 54608 records remain.

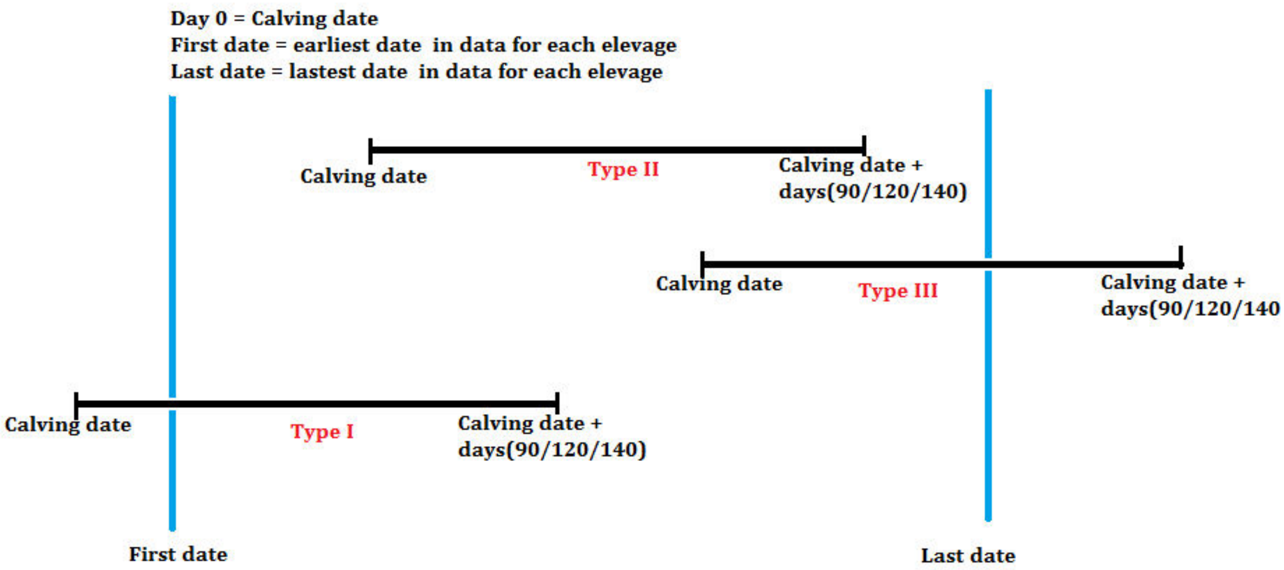

Day 0 = Calving date
First date = earliest date in data for each elevage
Last date = lastest date in data for each elevage
Calving date
Type II
Calving date + days(90/120/140)
Calving date
Type III
Calving date + days(90/120/140
Calving date
Type I
Calving date + days(90/120/140)
First date
Last date

Figure 5: Type of Cow

Criteria 2: Exclude incomplete records.

Criteria 2.1: Keep the cow who has at least 1 record during day 20 and day 30 (day 20 and 30 included) and during day 130 and day 140 (day 130 and 140 included).
Because, for example, if a cow has a health problem such as broken leg at day 90, it will be moved out of herd and it is possible that it won't come back before day 140. Thus the records during day 120 to day 140 are lost. Similarly, an incomplete data at day 20 may occur. We choose day 20 because usually 1st LP starts from day 20. In this step, 1064 cows with 41596 records remain.

Theoretically, according to biological point of view and working mechanism of HerdNavigator, it is impossible that a type II cow has no record during day 20 and day 30 or during day 130 and day 140. However we observed more than 100 cows in such a case. It might be due to the break down of machine or the mistake made by dairyman. In fact, it is a kind of missing data but cannot be filled.

Criteria 2.2: Delete the cow if it has at least one gap.
Gap is defined as a time interval between a record with the next record which is larger than 10 days. Gap means during those days cow no data were recorded for the cow by HerdNavigator, as a result, HerdNavigator doesn't have the records of the cow during this period (figure 6.a). Because the minimum length of OC is 5, therefore computing LP length

without excluding the gap leads to merge two short OC as a long OC, which will artificially lengthen the OC and miscount the number of LP. For instance, in figure 6, (b) and (c) are two possible true phase of (a). In this case, cow should be deleted if it has at least one gap. Note that during day 0 and day 20 any time interval longer than 10 days is not a gap since during which no complete LP or FP exists. We kept 826 cows with 36598 cows after this step.

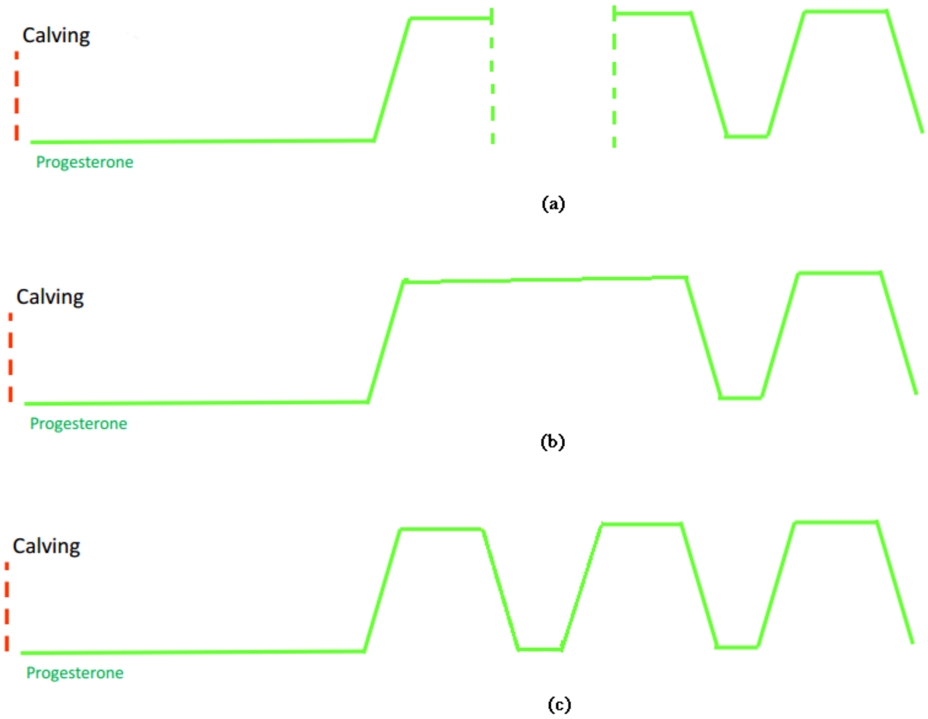


Figure 6: Gap

Criteria 3: Select progesterone during day 20 to day 140.
Usually, ± 20% of the cows resume before Day 20. There are 184 assays between Day 0 to Day 20, they are caused by dairyman and HerdNavigator should assay only after Day 20. As noted earlier, we need progesterone to detect LP and FP, and 5 ng/ml is the threshold value. If progesterone data between day 0 and day 19 is included, it might cause fake LP or FP and therefore enlarge the count of LP and FP. After applying this criteria, 36414 records remain but the number of cow is still 826.

Criteria 4: Delete Phase_Type = No phase.
Depending on the time of first progesterone rise, Phase_Type is defined as in table 9. Detect Phase Type is an important step without which we cannot count LP and FP.

Table 9: Phase Types

Phase_Type	*Definition*
Early Cow	Has incomplete first luteal phase since its first luteal phase starts before day 20, thus its first record has prog > 5
Ideal Cow	First luteal phase begins between day 20 and day 60
Late Cow	First luteal phase begins after day 60
No Phase	If progesterone is always <= 5 ,or always >= 5, or when number of records < 3, we cannot observe any phase

Figure 7 is an example of four phase types. Cow “01-0907-2” is an early cow, we don’t know how long its 1st luteal phase has began; Cow “01-0004-3” is an ideal cow which has 4 complete LPs and 3 complete FPs; Cow “01-1328-1” is a late cow, its progesterone has a slight increase at day 56 but fails to reach 5ng/ml, its 1st LP finally starts around day 90; Cow “17-0928-1” is the only one special cow without any phases during its day 20 and day 140 and we will remove it in later the analysis.

Criteria 5: Choose Holstein Friesian.
We want to study cows of the same breed. All the cow are Holstein friesian, except the ones in elevage 11 and 16 (Normande or Pie Rouge des Plaines). There are 754 Holstein friesian in data.

4.1.3 LP and FP Estimation

In this part, we will use the data we have already obtained through previous section to find the true LP and FP, then update milk production to new estimated data and add AI information to each LP.

(1) True luteal phase

STEP 1: Estimate true start date and end date for each LP.
Because HerdNavigator doesn’t make real-time record, we don’t know the exact time when progesterone becomes higher (or lower) than 5ng/ml. MD Royal et al (2000) defined the start date of LP as the date when progesterone becomes higher than 5 ng/ml and the end date of LP as the last date before the date when progesterone becomes lower than 5 ng/ml (figure 8). But we are not in the same case since they used daily data. If we follow their method, the largest deviation due to the inaccuracy can reach ± 9 days, so we should find the true start date and end date by estimation.

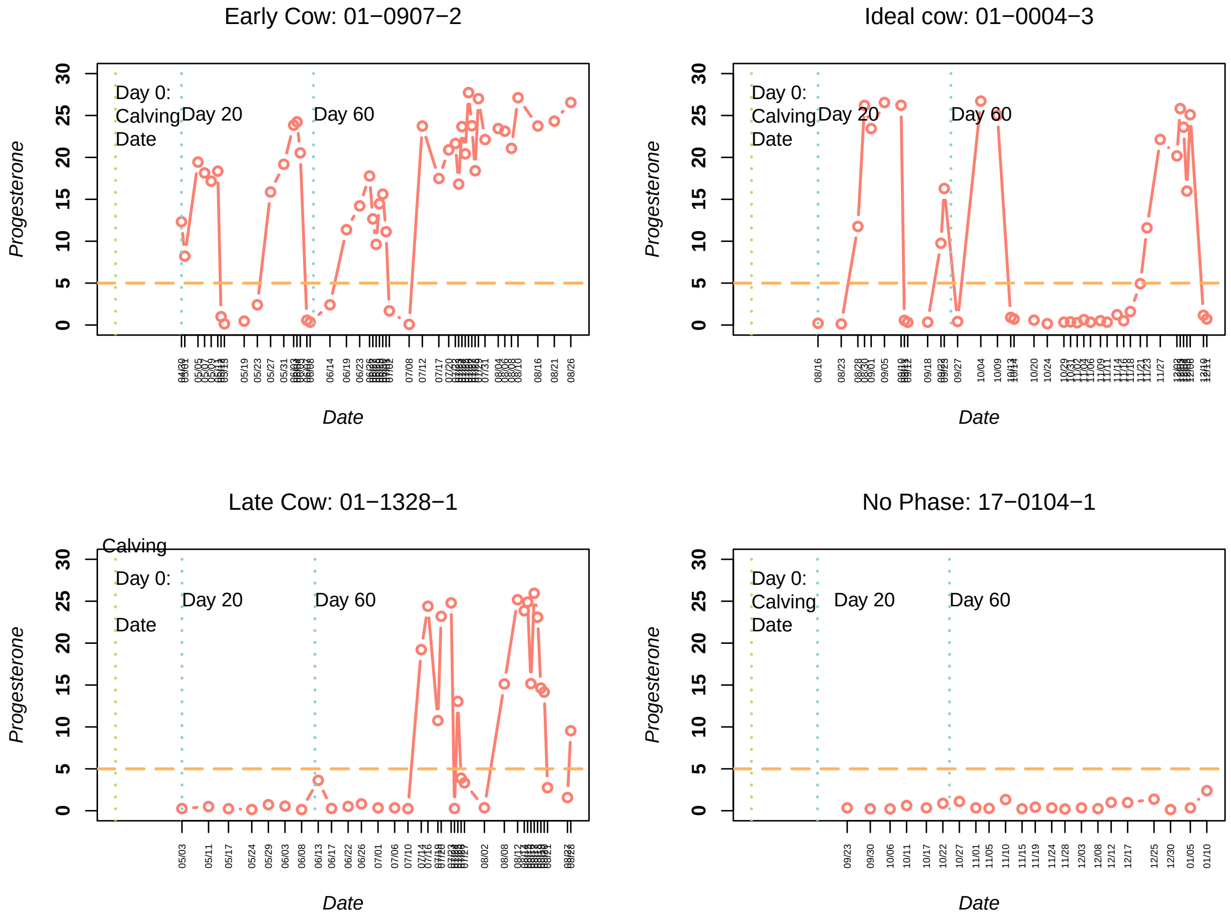


Figure 7: Record Type

We use linear model to estimate true start date and end date of phase. In the example of figure 8, the yellow point is start date of 4th LP (also the end date of 3rd FP), it is estimated by the date of record at which progesterone will become higher than 5 (point A) and the first forward date of which progesterone is higher than 5 (point B), similarly, the blue point - the end date of 4th LP (also is the start date of 4th FP), can be estimated by records at points B and C.

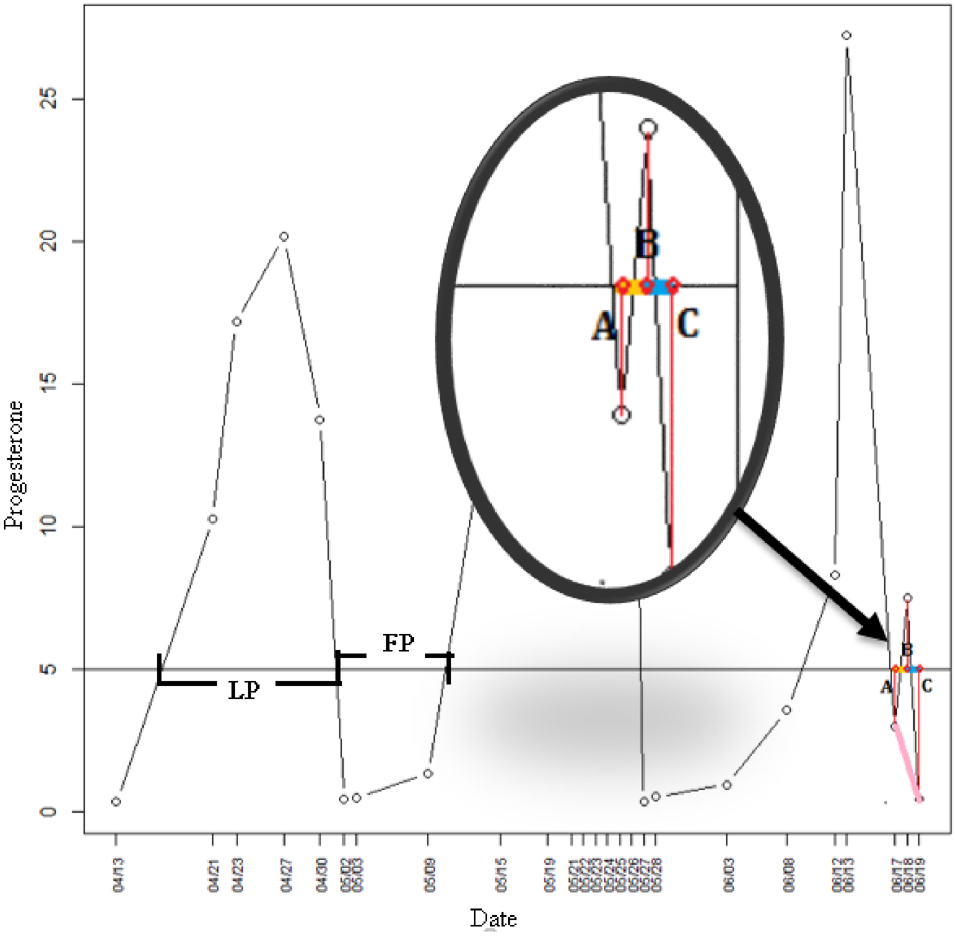


Figure 8: Record Type

STEP 2: Compute LP and FP length.

Using formulas below, we can calculate the length of LP and FP (figure 8) by corresponding start dates and end dates found in step 1:

LP length = end date of LP – start date of LP

FP length = end date of FP – start date of FP

STEP 3: Delete fake phase.

Remind of table 1, we defined fake LP (or FP) as a phase whose progesterone is higher (or lower) than 5 ng/ml and with phase length <= 3 (or <= 2). In other words, to index the true phase cycle, any LP (or FP) no longer than 3 days (or 2 days) should be ignored.

In this case, the idea of deleting fake phases is:

(1) Delete the short phase length (LP length <= 3 or FP length <= 2) which has only one record;

(2) Repeat previous step until no such one-record short phase exists. Only short LPs or FPs

remain after this step;

(3) Delete short LPs or FPs no matter how many records within them;

Figure 9 is an example. Note that although we don't know the correct values of progesterone for those wrong points and we deleted them, it is not equivalent with they are missing values. As a matter of factor, we still know that their true progesterone values are higher (or lower) than 5 ng/ml. Thus, even after correction some time intervals between records become larger than 10 days, they are not gaps.

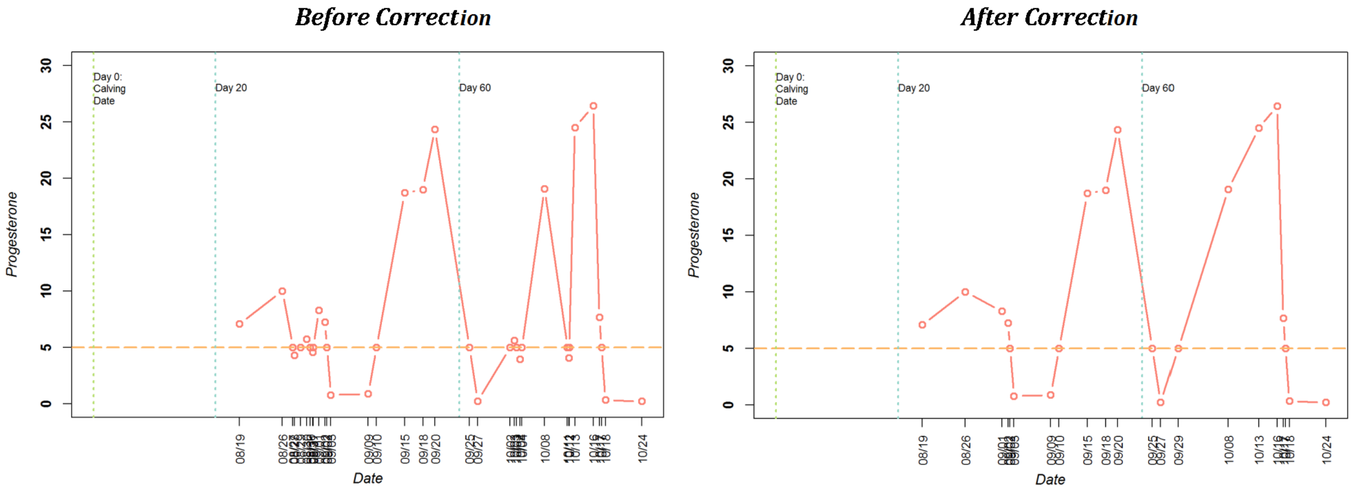


Figure 9: An Example Result of Step 1 to 3

STEP 4: Count phase cycle. For each id_new, after computing the phase length in each cycle and removing fake phases, we count the phases and index the LP (or FP) as 1st LP, 2nd LP (or 1st FP, 2nd FP), etc. Figure 10 and figure 11 summarizes the number of cows in each phase cycle and phase type.

712 cows have the 2nd LP and 80.48% of them have the 3rd LP, this percentage keep decreasing during later cycles, around 64.40% of 3rd LP cow has the next LP and only 33.71% 4th LP cow has 5th LP, this is because as time goes by, more cows have gotten pregnancy thus their LP stopped. The proportion of ideal type and late type decrease with phase cycle, while the proportion of early type increases. In other words, early cow tends to have more LPs before pregnancy. Among the first 5 phase cycle, more than half of cows are in ideal type.

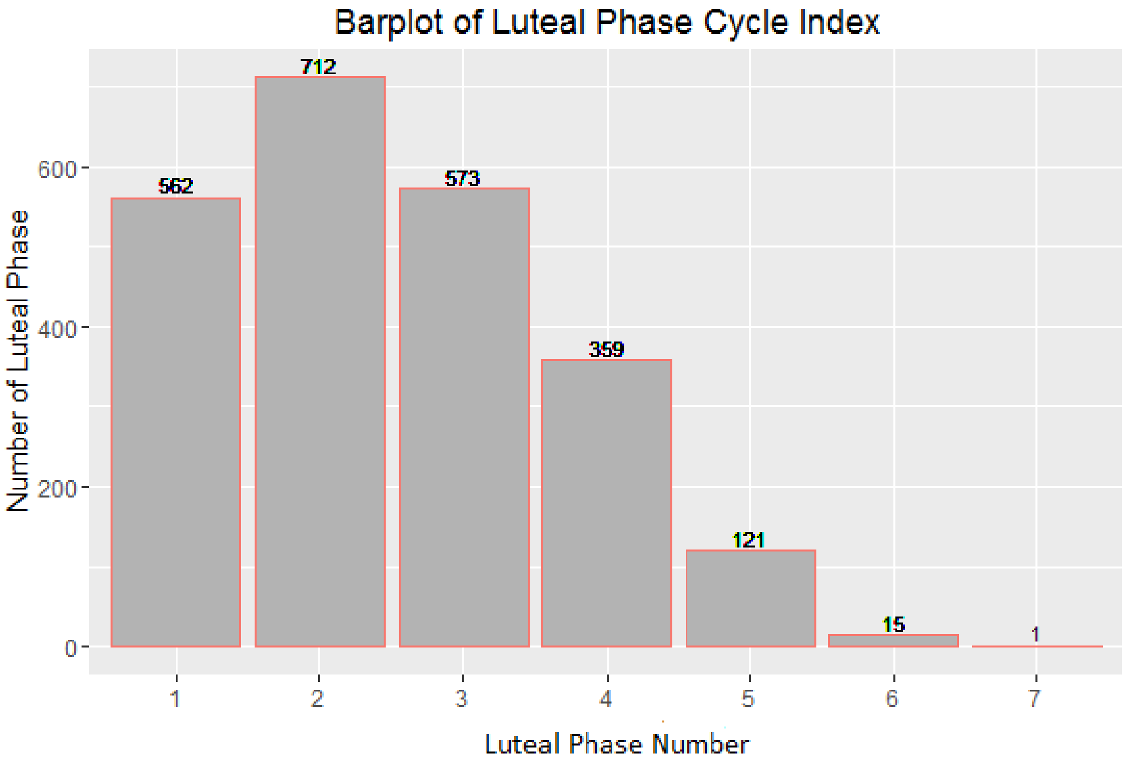


Figure 10: Number of id_new in each phase cycle

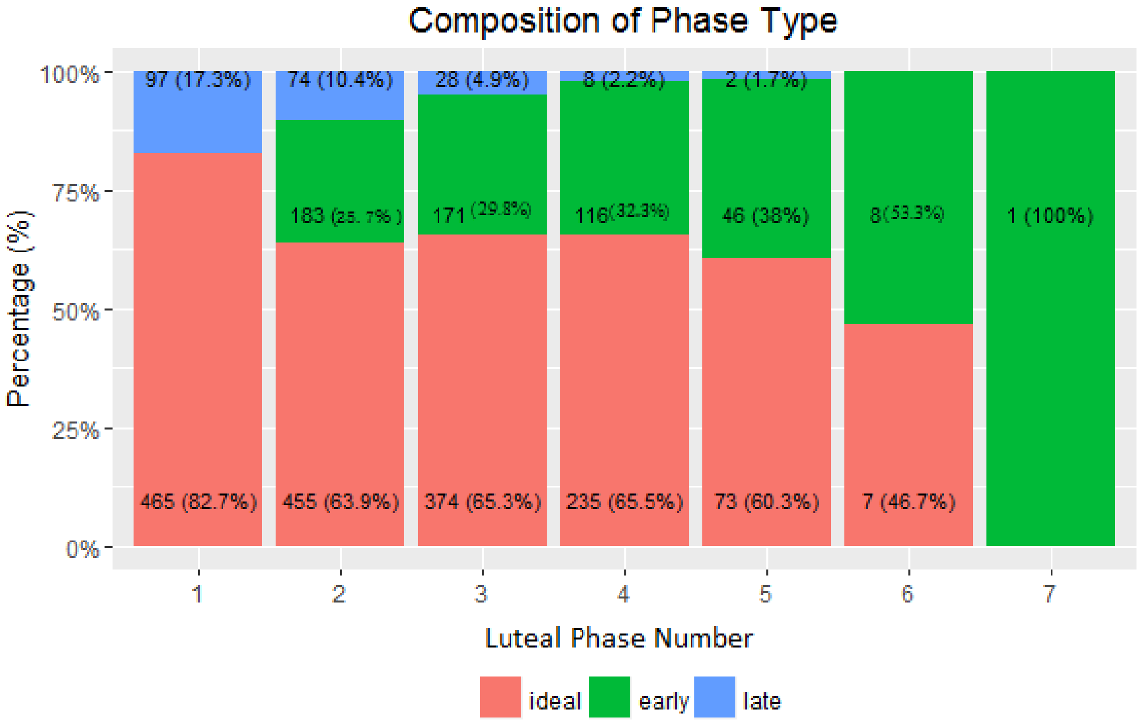


Figure 11: Proportion of phase type in each phase cycle

STEP 5: Merge data "cows" with cleaned data "rapport production lait".

For the estimated date, we replace the missing milk production value in variable "prodJ" (daily milk production) and "prod7J" (average milk production within 7 days forward) by the daily milk production in data "rapport production lait".

STEP 6: Detect artificial insemination activities.
Since an AI, whether is successful or not, can lengthen the LP, it is necessary to distinguish the LP with AI with those without AI. To do this, we created a new variable which indicates whether there is an AI during the latest FP forward.

Four cows ("02-0604-4", "08-0037-1", "18-8567-2" and "20-1733-2") were inseminated before 1 st LP due to the non-standard operations of dairyman. In later analysis, those cows cannot be taken into account in some questions. Also, there are 24 cows which were inseminated during LP. In this case, AI has effect on LP.

The proportion of LP with AI increases with phase cycle index (figure 12). This increase is considerable during first 4 phases but slows down and stays around (60%) among 4th to 6th LP. As explained in Introduction, for economic purposes, cows have to be inseminated as early as possible after 45 days postpartum.

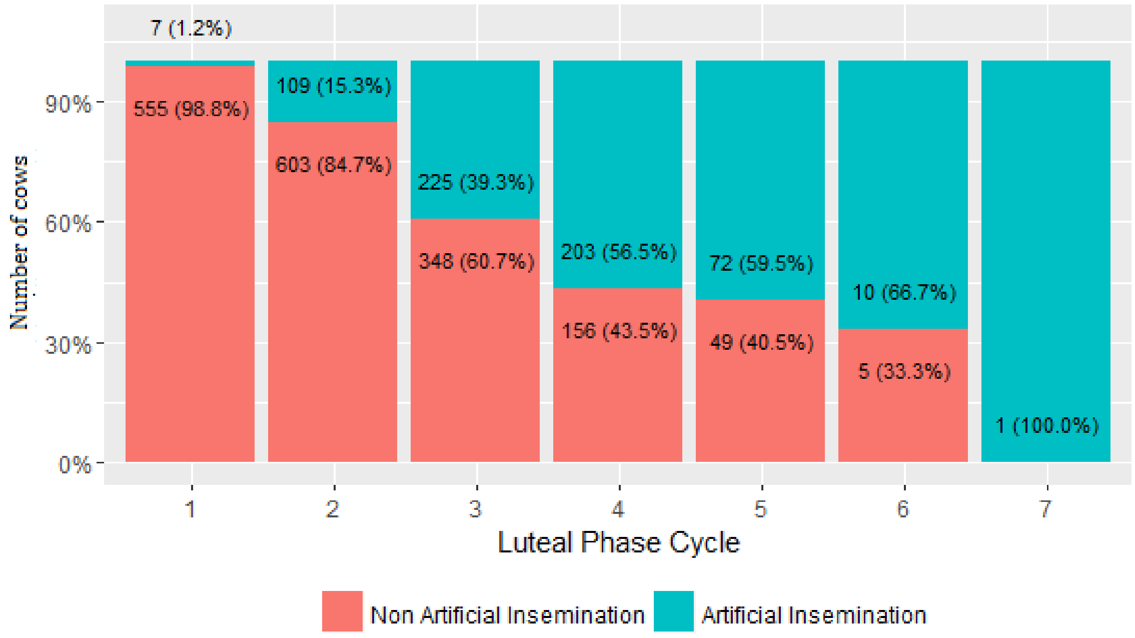
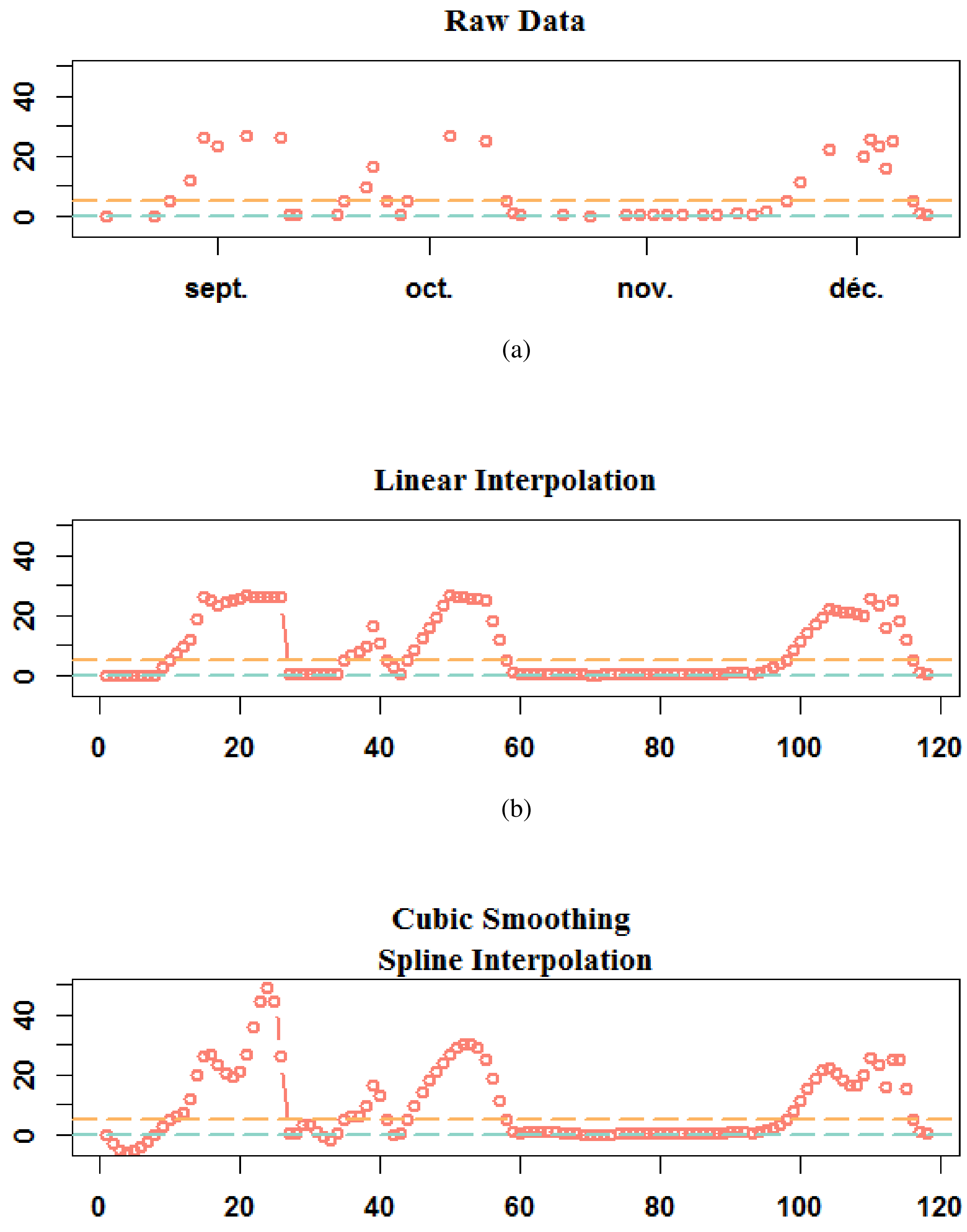


Figure 12: Proportion of artificial inseminations
depending on luteal phase number

4.1.4 Daily Progesterone Interpolation

To prepare for the content of fertility analysis, we want to know the daily progesterone concentration (objective 1). In this case, we use two interpolation methods (linear interpolation and cubic spline interpolation) to estimate. I finally chose linear instead of cubic spline interpolation because the latter will lead to negative progesterone value which is unrealistic (figure 13).



(a)

(b)

(c)

Figure 13: Comparison of Interpolation Methods

4.1.5 Code Type of Cow

Table 10 are the definitions we applied to find the code type of cow and table 11 summarises how many cows in each type based on different rules and answer objective 3. Rule 1 and rule 2 have different the threshold values in definition.

Table 10: Definitions of Code Types

AI	*Code Type*	*Name*	*Definition*
All	Normal Type	I	At least 1 progesterone higher than 5 ng/ml before 45 days (included) after calving and no other problems
All	Delayed Ovulation Type I	II	Progesterone <= 5 ng/ml during the first 45 days (included)
All	Delayed Ovulation Type II	III	Progesterone <= 5 ng/ml during N1 days (included) between two LP <=> FP length >= N1
No	Persistent CL Type I	IV	Progesterone higher than 5 ng/ml during N2 days in 1st LP <=> length of 1st LP >= N2
No	Persistent CL Type II	V	LP length >= N2 (excluding 1st LP) without IA event during previous FP
All	Short Luteal Phase	VII	LP length < 10
Yes	Late Embryo Mortality	VI	LP length >= N2 with IA event during previous FP

Note: N1 and N2 refer to table 11

Table 11: Results of Code Types

	Rule 1 (N1 = 12, N2 = 19)	*Rule 2* (N1 = 14, N2 = 20)
I	145	191
II	161	161
III	325	212
IV	118	106
V	124	106
VI	110	93
VII	262	262

We also answered the questions below:

1) For Code II: What proportion of id_new has an ovulation later than 45 days within 140 days postpartum?
In both Rule 1 and Rule 2, there are 562 (75.4%) of cows whose 1st LP are complete. 161 cows had their first ovulation later than 45 days after postpartum. This number accounts for 21.6% among 745 cows and 28.6% among 562 cows.

2) For Code III: Among 745 id_new, what percentage suffer at least once between day 0 and day 140 from a FP length $\geq$12 or FP length $\geq$14?
   Because that cow "15-0414-3“, "15-0616-1“, "18-8513-3“, “18-8535-3“, "22-7771-2“ and "24-0891-3“ have only one LP and non complete FP, thus the total number of cows having FP is 739. In Rule 1,325 cows have at least one FP $\geq$12, they account for 44.0% of 739 cows which have FP and 43.6% of all 745 cows; In Rule 2, 212 cows have at least one FP $\geq$14, they make up 28.7% of 739 cows which have FP and 28.5% of all 745 cows.

3) Code IV: What proportion of id_new whose 1st LP is known has length of 1st LP $\geq$19 or 20? (Excluding the id_new whose 1st LP has an AI)
   555 cows have 1st LP which without AI. More precisely, 118 cows (21.2% of 555 cows and 15.8% of 745 cows) have long 1st LP according to Rule 1; and 106 cows (19.1% of 555 cows and 14.2% of 745 cows) have long 1st LP according to Rule 2.

4) Code V: Among the id_new having at least a LP (excluding 1st LP and without AI), how many of them has at least a length of LP $\geq$19 or 20 (excluding 1st LP)?
   Among 745 cows, 627 cows have n^{th} LP(n $\geq$2) without AI. Moreover, in Rule 1, 124 cows (19.8% in 627 cows and 16.6% in 745 cows) has length of n^{th} LP $\geq$ 19; while in Rule 2, 106 cows (16.9% in 627 cows and 14.2% in 745 cows) have n^{th} LP $\geq$20.

5) Code VII: how many id_new has LP $\leq$10?
   295 cows have at least one short LP (433 short LP in total). They account for 39.6% of all cows. Among these 295 cows, 262 (88.8%) id_new have at least one short LP without AI (354 short LP in total) while 76 id_new have at least one short LP with AI (79 short LP in total).

6) Code VI: How many id_new with at least one LP with AI had at least one LP with AI and length $\geq$19? How many LP with AI had are longer than 19?
   620 LP have AI if excluding the AI happened before or during 1st LP. In Rule 1, 113 (18.2%) of them are long LP; 95(15.3%) of them are long LP. Considering id_new, 429 id_new have at least one LP with AI. In Rule 1, 110 (25.6%) of them have at least one long LP with AI and in Rule 2 this number is 93 (21.7%).

7) Code I: Normal Code Cow
   145 (19.5%) id_new are in normal code type in Rule 1 and 191 (25.6%) in Rule 2.

4.1.6 Final Data

There are two final data which answer objective 2. To measure performance of progesterone, the row of final data "Cows" represents a cow at specific date while the row in final data "LPLs" is a LP at the date it begins.

Final data "Cows" has 745 id_new. Table 12 is new variable descriptions and it is a additional table of table 6.

Table 12: Data Description of Additional Variables

	Variable	*Explanation*	*Value*	*Type*
1	id_new	New id (unique for cow in each of milk production cycle); There are 745 different id	id = "elevage-animal-lactation" 01-0004-3, 02-0514-2 ...	Character
2	Calving_Date	Day 0 in milk production cycle	From 2013-11-29 UTC to 2015-01-26 UTC	POSIXct
3	first_date	Date when HerdNavigator started to rund.	See table 3	POSIXct
4	last_date	Date when vet exported the data	See table 3	POSIXct
5	begin_LP	Date when LP begins		POSIXct
6	end_LP	Date when LP ends		POSIXct
7	begin_FP	Date when FP begins	begin_FP = previous end_LP	POSIXct
8	end_FP	Date when FP ends	end_FP = next begin_LP	POSIXct
9	Phase_Type	See definition in table 9	183 "Early Cow", 465 “Ideal Cow", 97 "Late Cow" and 0 "No Phase“	POSIXct
10	LP_index	Day 0 in lactation	From 1 to 7	Integer
11	FP_index	Day 0 in lactation	From 1 to 7	Integer
12	Added	If this record means perform AI	0 or 1	Factor
13	LP_AI	Whether there is an AI before LP	0 or 1	Factor
14	Num_AI	Number of AI performed during a FP	0, 1, 2	Integer
15	index_AI	How many previous LP has AI within former FP	From 0 to 4	Integer

Final data “LPLs” (table 13) is constructed as a data containing information of LP.

Table 13: First Three Rows in Data "LPLs"

	id_new	*elevage*	*lactation*	*animal*	*jours*	*Date*
1	01-0004-3	01	3	0004	29.09	2014-08-25 02:14:57
2	01-0004-3	01	3	0004	54.98	2014-09-19 23:29:25
3	01-0004-3	01	3	0004	63.22	2014-09-28 05:15:22

	Calving_Date	*Phase_index*	*LP_AI*	*Num_AI*	*index_AI*	*lens*
1	2014-07-27 00:00:00	1	0	0	0	16.73
2	2014-07-27 00:00:00	2	0	0	0	5.87
3	2014-07-27 00:00:00	3	0	0	0	14.10

To measure feature of LP, I use a new variable **"lens"** as the length of LP. In addition, I change variable in POSIXct format into Date format. This gives a strong advantage of extract season which could be an important step in data analysis. There are 2343 LP including the 1st LP with AI, among which 627 LP have AI during previous FP period while 1716 LP do not.

4.2 Data Analysis

4.2.1 Univariate Description of Variables in LP Influential Factor

LP. Table 14 is the summary of LP. Figure 14 is the distribution of LP length, which indicates that there are more short LP in non AI group. Thus our final data is in consistent with the reality: AI can lengthen the luteal phase length.

Table 14: Summary of LP length

	LP with AI	*LP without AI*
Min	3.3	3.0
Max	53.9	106.2
Mean	15.8	14.8
sd	7.1	8.2

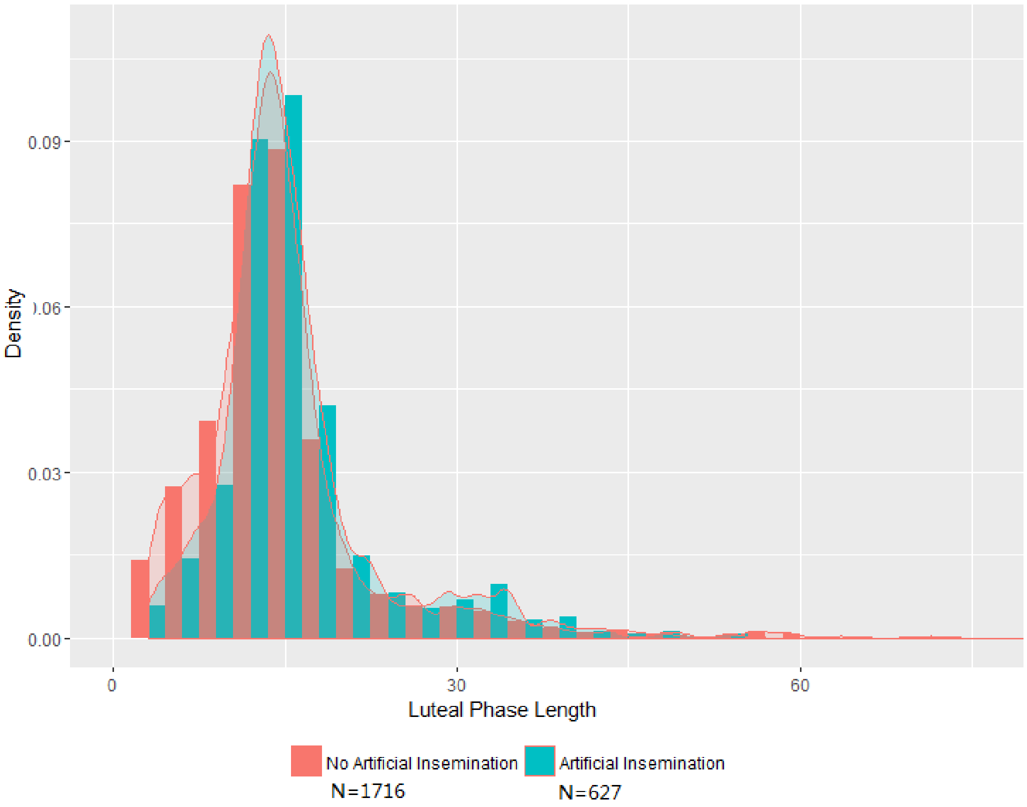


Figure 14: Distribution of LP Length

In figure 15, among those LP which have AI, 99% have only one AI performed. Moreover, in each phase cycle, the proportion of luteal phases which have twice AI is always around 0% to 2%. For the cow which has at least one AI, 3% of them have more than two AIs in one cycle. The second AI is performed based on the own feeling of dairyman, 12 to 24 hours after the first LP when the signs of heat persist.

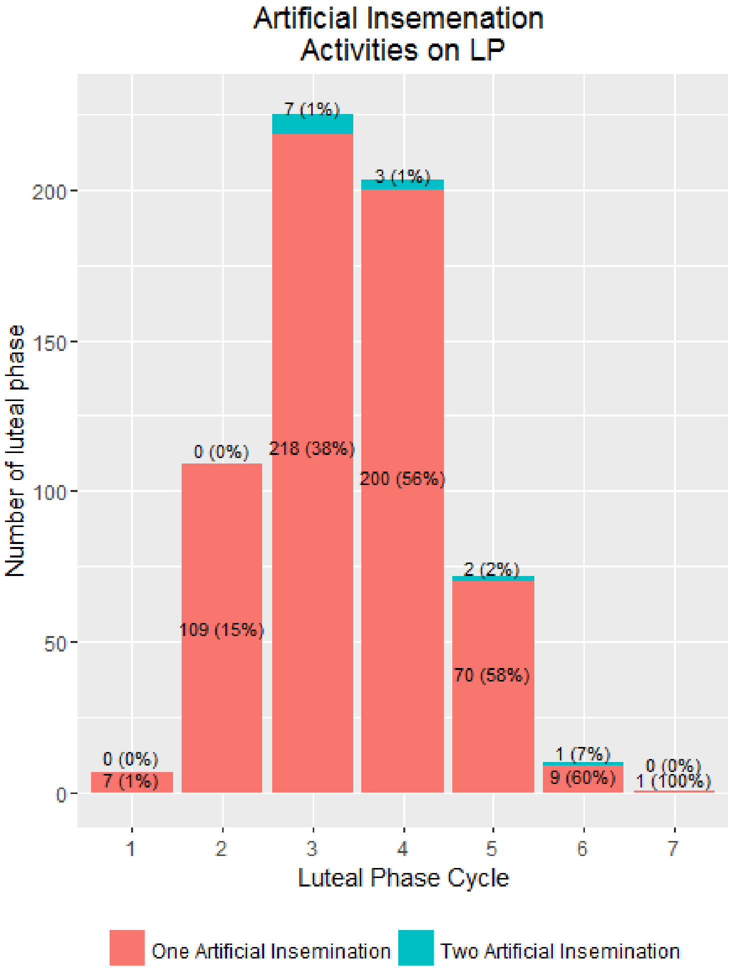


Figure 15: Artificial Insemination Activities

Jours. Figure 16 is the standard scatter plot used to check the relationship between quantitative variables: jours and lens. We cannot observe any obvious tendency between them. In fact, the data of LP length is not normal but right skewed distributed (figure 14), thus a non-parametric correlation test is applied and it says that the jours significantly has a negative but very weak relation with lens.

In other words, an increase in jours will lead to a slightly decrease in lens. Figure 16 implies that it may be caused by the fact that we choose to cut data at day 140, thus for those cows who have very long 1st or 2nd LP, we cannot avoid losing the information of their later LP. Thus in 3 rd or 4 th LP, only short lens remain. This problem can be improved when we will have more data in longer duration and do not needn to cut data anymore. A new data collection in 23 dairy herds is scheduled.

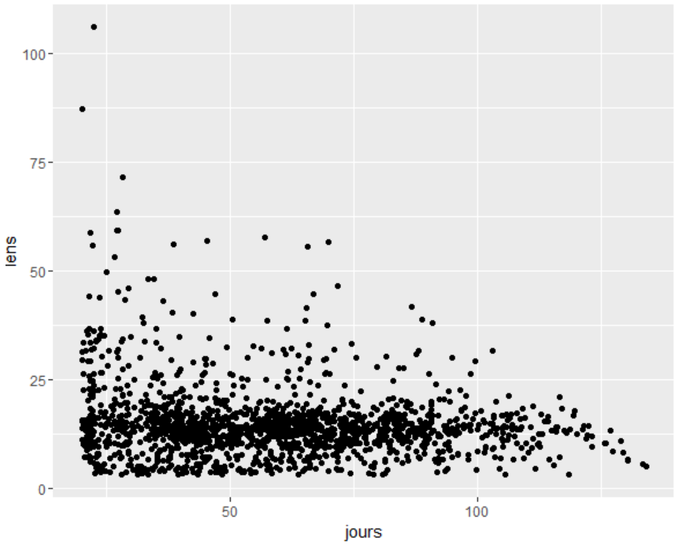


Figure 16: Relationship between time elapsed after calving
at the beginning of LP (variable "jours") and LP length (variable lens)
n = 1662

Phase_index. Figure 17 shows no significant difference in lens among LP cycles. As a matter of fact, it is not just a conclusion from eyes. Again for the reason of non-normality, as well as unbalance, and heteroscedasticity of the LPLs data in each phase (table 15), we use Kruskal-Wallis test to check no difference between LPL.

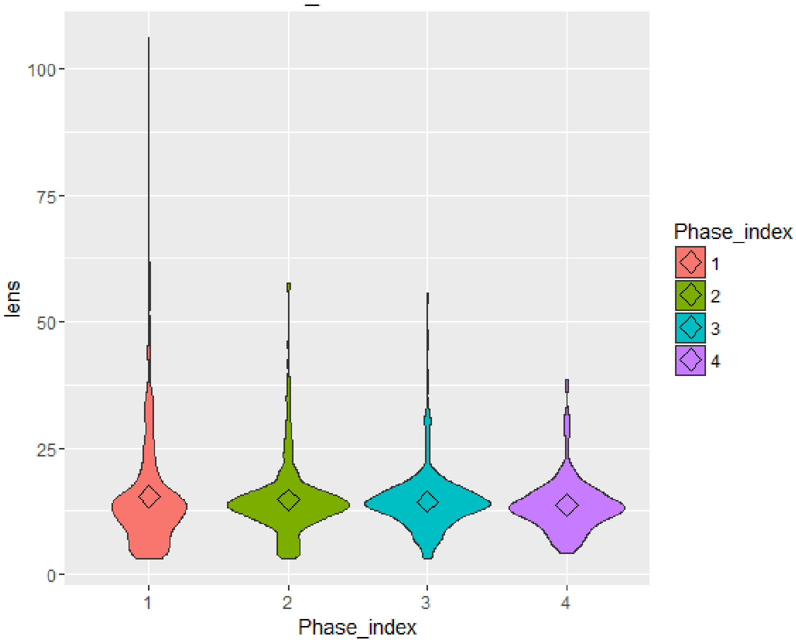


Figure 17: Relationship between Phase Cycle Index and LP length
n=1662

Table 15: Descriptive Statistics of LP Length (days)

LP Cycle Index	n	mean	sd	median	min	max	rqnge	skew	se
1	555	15.46	11.19	13.28	3.05	106.17	103.12	2.85	0.47
2	603	14.73	6.96	13.83	3.05	57.64	54.59	2.20	0.28
3	348	14.41	5.80	13.98	3.01	55.68	52.66	2.28	0.31
4	156	13.74	5.46	13.34	4.26	38.72	34.46	1.66	0.44

4.2.2 Influence of LP Cycle and Jours on lens

In this part, two regression models are used to answer the questions. All the analysis in this part is based on the data from the first four LP without AI because the last two cycles do not have enough observations (49 observations in 5th LP and 5 observations in 6th LP).

Figure 18 gives the idea that in a specific phase cycle how LP length will change when jours increase. There is a very weak negative tendency.

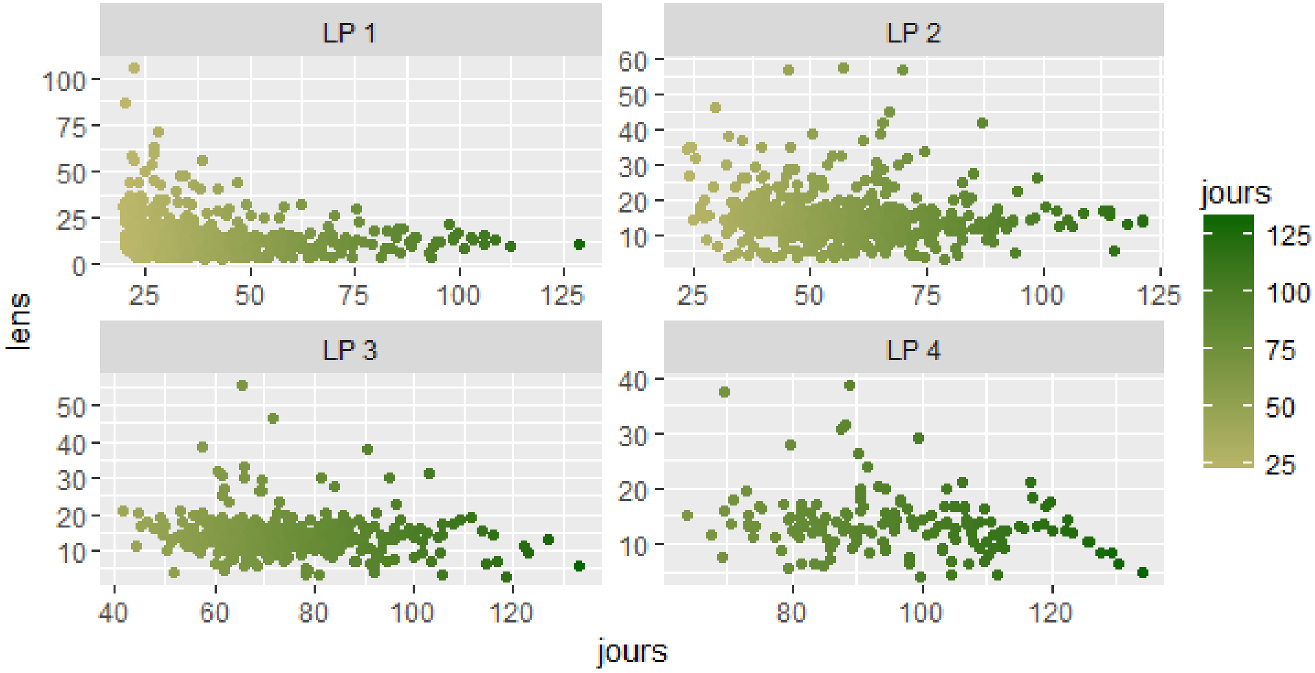


Figure 18: Relationship Between time elapsed postpartum (variable "jours") and LP length Grouped by Phase Cycle

In fact, we used a linear model as below to check the impact of jours and Phase_index on lens.

$$lens_i = \alpha_i + \beta_1 jours_i + \beta_2 Phase\_index_i + \varepsilon_i \tag{1}$$

Left column in table 16 is the result of OLS regression of log transformation of lens on jours and Phase_index. All the variables are statistically significant at the 0.001 level. We can say that for a one-unit increase in jours, we expect to see about a 0.5% decrease in lens, since

$exp(-0.005) = 0.995$. The impact of phase cycle on lens differs with the Phase_index. We can say that:

- lens in LP2 will be 13.2% longer than LP1;
- lens in LP3 will be 23.6% longer than LP1;
- lens in LP4 will be 30.2% longer than LP1.

Table 16: Regression Results

	Dependent variable:	
	log(lens)	
	OLS	*Panel Linear*
	(1)	(2)
jours	-0.005***	-0.027***
	(0.001)	(0.004)
Phase_index2	0.124***	0.575***
	(0.031)	(0.126)
Phase_index3	0.212***	1.293***
	(0.041)	(0.207)
Phase_index4	0.264***	2.003***
	(0.059)	(0.301)
Constant	2.724***	
	(0.034)	
Observations	1,662	184
R^2	0.028	0.270
Adjusted R^2	0.026	0.197
Residual Std. Error	0.491 (df = 1657)	
F Statistic	11.993*** (df = 4; 1657)	12.405*** (df = 4; 134)
Note:		*p<0.1; **p<0.05; ***p<0.01

We wouldn't use this model to predict lens because R-squares is very small which means the independent variables can only explain a small part of lens value. In fact, this model is not perfect. Since we are dealing with 733 id_new observed at least 1 times (1662 records in total), clearly not all records are independently distributed.

For this reason, it is preferable to take explicitly into account an individual specific effect as well as time effect in panel data models.

$$lens_{it} = \alpha_i + \beta_1 jours_{it} + \beta_2 Phase\_index_i + \varepsilon_{it} \tag{2}$$

where

t = LP Cycle
i = id_new
α_i = Individual effect. In fixed effect model, it is assumed to a constant in each id_new; In random effect model, it is assumed to follow a specific distribution in each id_new

After model diagnosis (Hausman test for panel Mmodels), the model 2 is a fixed effect model controlling for heteroskedasticity. From a biology point of view, the more phase cycle we keep in panel data, the higher possibility the cow has some health problems (sickness or broken leg).From the third row in table 16, we can say that for a one-unit increase in jours, we expect to see about a 2.664% decrease in lens, and again the impact of phase cycle on lens differs with the Phase_index.

- lens in LP2 will be 77.7% longer than LP1;
- lens in LP3 will be 264.4% longer than LP1;
- lens in LP4 will be 641.1% longer than LP1;

4.2.3 Other Variables' Influence on lens

Table 17: Result of Test:
Factors influencing LP length

Variable	Test	Influence	Details
jours	Yes(1)	Spearman Correlation Test	(1) p = 0.000081 < 0.05, thus correlation significantly differs from 0; Test statistics rho= -0.096493, this negative association is small. (2) The significant difference among LPL across farms only exists between farm 10 (in Middle of France) and farm 20 (in South of France). (3) LP length within 1st lactation significantly differs from the ones within 2nd (with p-value = 0.0068) and 3rd lactation (with p-value = 0.0054). The mean of LPL in first three lacatation is 15.189, 14.369 and 14.131 days. No other difference between any pair of LPL data.
Phase cycle index	No	Kruskal-Wallis Tests and Tukey Post-hoc	
elevage	Yes(2)		
lactation	Yes(3)		
Season of Calving Date	No		
Season when LP begins	No		
Number of FP (with AI) before	No		

Although the analysis is mainly focus on phase_index and jours, we still have a brief look at other variables. Results of other potential influence factors that were tested can be seen in table 17.

The lens differs among farms: LP length in Farm 10 (mean = 9.8) siginificantly differs from farm 1, 3, 20 and 21 (means > 14). In fact, the gene selection procedure in farm 10 is more active than other farms thus LP length is shorter.

Significant difference also exist between different lactation (milk production cycle). LP length in 1st lactation (mean = 15.16) is significantly larger than the ones in 2st (mean = 14.32) and 3rd lactations (mean = 14.06);

Section 4.2.1 to 4.2.3 answer the second objective.

4.2.4 Milk Production Analysis

Spearman test indicates that progesterone has weak negative correlation with daily milk quantity ($p < 0.01$ and τ = -0.091943), the correlation is more weak with average milk production within 7 days before. ($p < 0.01$ and τ = -0.086221). But this negative correlation becomes strong ($p < 0.01$ and τ = -0.23823) when using accumulated milk production data.

Total milk production during day 0 and day 140 (called as Milk.Total)

Difference between Milk.Total among different code type of cow is tested. Milk.Total has significantly higher value in Code III (mean = 5094.9 L) cows rather than other cows (mean = 4825.1 L).In fact, only Code_III and non Code_III cows have difference in milk production, no other significant difference exists with other code types.

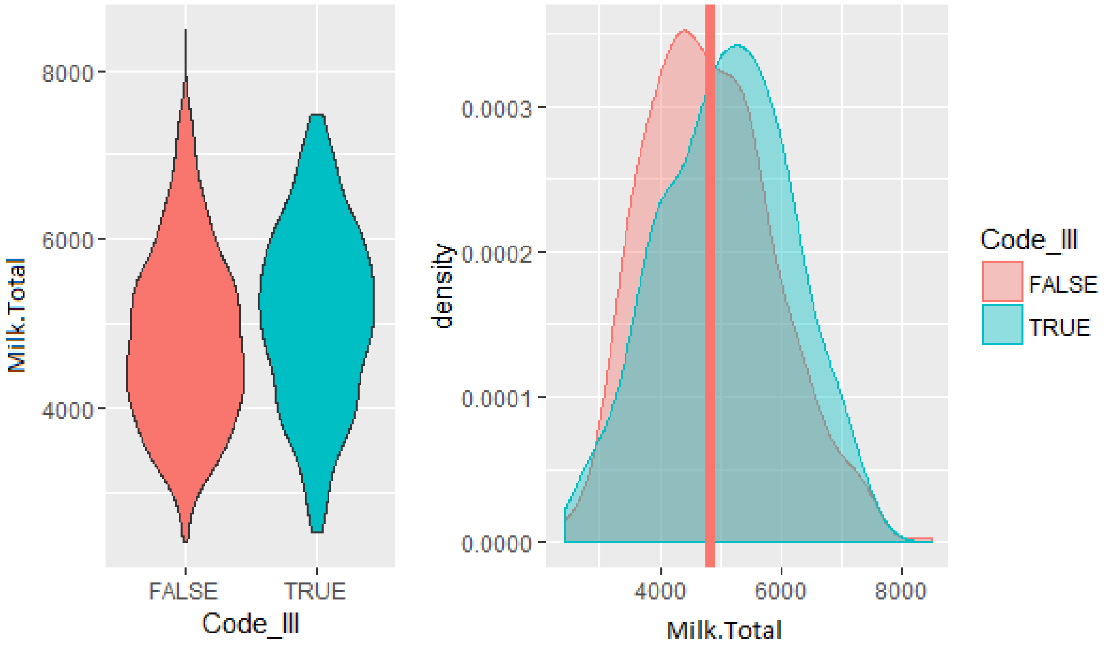


Figure 19: Comparison of Total Milk Production Between Code III Cow and Others

5 Summary

Based on HerdNavigator obtained from 553692 progesterone data of 2510 cows in 23 French herds, we extract 2343 complete luteal phase.

The main objectives in our work were: To fill progesterone with interpolated values and compare methods; To characterize ovarian cyclicity resumption in dairy cows in commercial (non experimental) conditions; To identify type of cyclicity profile before insemination; To identify feature of LP length such as distribution of LP's length.

Section 4.1.1 to 4.1.3 characterizes ovarian cyclicity resumption. The present work shows that, among 2510 cows from 23 farms, there are 2343 luteal phases among 744 complete and well recorded milk production cycles (id_new) in 744 Holstein Friesian. 46.7% of id_new have at least three luteal phases before pregnancy.

The distribution of luteal phase length is right skewed and left truncated. Although luteal phase length doesn't differ among luteal phase cycles, it has a weak negative correlation with number of days after calving. Also luteal phase length is not affected by season, but differs among farms and lactations. Although through test the mean value of luteal phase length decreases with its index, the regression result shows that this is a combined effect of luteal phase cycle and number of days after calving (they have opposite effect on luteal phase length): luteal phase length increases when cow goes into the next luteal phase cycle, but decreases as time goes by.

We described the profile of ovarian resumption elapsed from calving increases in our population. Depending of the criteria used to define normality. Only 19.5% to 25.6% of cows have normal progesterone file The last objective is in section 4.1.4: we use linear interpolation to find out daily progesterone concentration. It is a preparation step for insemination analysis.

Further work is needed to find influential factor of success rate of AI (especially the ovarian resumption profile) and try to find a prediction model based on the data we have already built. The dataset will also be implemented with new data from the same dairy herds, increasing the number of cows and of lactations.

References

[1] Michael G Diskin and Joseph M Sreenan. "Expression and detection of oestrus in cattle". In: *Reproduction Nutrition Development* 40.5 (2000), pp. 481–491.

[2] NC Friggens et al. "Improved detection of reproductive status in dairy cows using milk progesterone measurements". In: *Reproduction in Domestic Animals* 43.s2 (2008), pp. 113–121.

[3] C Inchaisri et al. "Analysis of the economically optimal voluntary waiting period for first insemination". In: *Journal of dairy science* 94.8 (2011), pp. 3811–3823.

[4] GE Lamming and AO Darwash. "The use of milk progesterone profiles to characterise components of subfertility in milked dairy cows". In: *Animal reproduction science* 52.3 (1998), pp. 175–190.

[5] Geert Opsomer et al. "Risk factors for post partum ovarian dysfunction in high producing dairy cows in Belgium: a field study". In: *Theriogenology* 53.4 (2000), pp. 841–857.

[6] MD Royal et al. "Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility." In: *Animal science* 70.3 (2000), pp. 487–501.