



HAL
open science

Evaluation of a functional Bayesian method to analyse time series data in precision viticulture

C. Laurent, Meïli Baragatti, J. Taylor, Thibaut Scholasch, Aurélie Metay,
Bruno Tisseyre

► **To cite this version:**

C. Laurent, Meïli Baragatti, J. Taylor, Thibaut Scholasch, Aurélie Metay, et al.. Evaluation of a functional Bayesian method to analyse time series data in precision viticulture. 12th European Conference on Precision Agriculture, ECPA 2019, Jul 2019, Montpellier, France. pp.67-73, 10.3920/978-90-8686-888-9_7. hal-02609779

HAL Id: hal-02609779

<https://hal.inrae.fr/hal-02609779>

Submitted on 16 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation of a functional Bayesian method to analyse time series data in precision viticulture.

C. Laurent^{1,2,3}, M. Baragatti⁴, J. Taylor¹, T. Scholasch³, A. Metay², B. Tisseyre¹

¹ITAP, Univ. Montpellier, Montpellier SupAgro, Irstea, France

²SYSTEM, Univ. Montpellier, CIHEAM-IAMM, CIRAD, INRA, Montpellier SupAgro, France

³Fruition Sciences, Montpellier, France

⁴MISTEA, Univ. Montpellier, Montpellier SupAgro, INRA, France

cecile@fruitionsciences.com

Abstract

In precision agriculture, most studies focus on spatial crop variability whereas temporal variability and its role in decision-making is equally important. The classical methods for temporal analysis have limitations, potentially resulting in information loss. A novel method based on a Bayesian functional Linear regression with Sparse Steps functions (BLiSS method) is evaluated in this paper to investigate continuous influence analysis when working with time series data. The example of the influence of temperature on the number of clusters per vine during the year before harvest was considered as an example application. The evaluation of the BLiSS results was done by comparing identified critical time periods with traditional viticulture knowledge in the literature. It showed the relevance of the BLiSS method, highlighting already known results and identifying new critical time periods for yield elaboration.

Key words: BLiSS, temporal analysis, temperature, yield, number of clusters

Introduction

In precision agriculture, most studies have focused on spatial crop variability (Oliver et al., 2013) whereas temporal variability and its role in decision-making is equally important. Indeed, most indicators of production yield and quality are time-dependant. Understanding their evolution during the crop growth cycle and the factors that drive them is often an issue for better production management. This implies understanding the pattern of influencing factors over a whole production period, as well as the correlations between influences.

Historically, ‘temporal’ studies have focused on using climate variables at a few known key phenological stages or time steps. This has been done to simplify the analysis and permit the use of classical methods, such as least squares linear regression. However, these classical approaches have limitations: i) they depend on choices of climate variables and timing, ii) it is often necessary to suppress data or to analyse only parts of a time series, and iii) times series temporality (observations correlated over time) is not considered. Therefore, information about temporal crop physiological regulation is potentially missed.

To advance temporal analysis in the crop production domain, this paper evaluates the potentialities of a new approach to study quantitative time series data. It uses a Bayesian functional Linear regression with Sparse Steps functions (BLiSS method) (Grollemund et al., 2019). The principle of this method is to analyse the complete history of an

explanatory variable to identify the periods during which it has an impact on the explained variable. It takes into account the data correlation over time *i.e.* it allows the analysis of a period's impact according to the impact of others periods. By overcoming the classical approaches' limitations, the BLiSS method could result in the discovery of new periods of influence of an explanatory variable on an explained variable in many agricultural sectors.

The aim of this paper is therefore to test the potentialities of the BLiSS method on a study case that is well understood in the literature. To this end, the impact of temperature history in the year before harvest on the number of clusters per vine has been chosen as a study case. Indeed, grape yield elaboration is an interesting case study in temporal terms as grapevine is a perennial crop, meaning its yield determining process covers two growth cycles (Howell, 2001 ; Vasconcelos et al., 2009). These two years are noted $n-1$ and n , the year n being the year of harvest (Fig. 1). Thus, at least two growth cycles overlap and there is an interdependency between the different physiological processes leading to memory effects in vine physiology (Ravaz & Sicard, 1911 ; Keller et al., 2010).

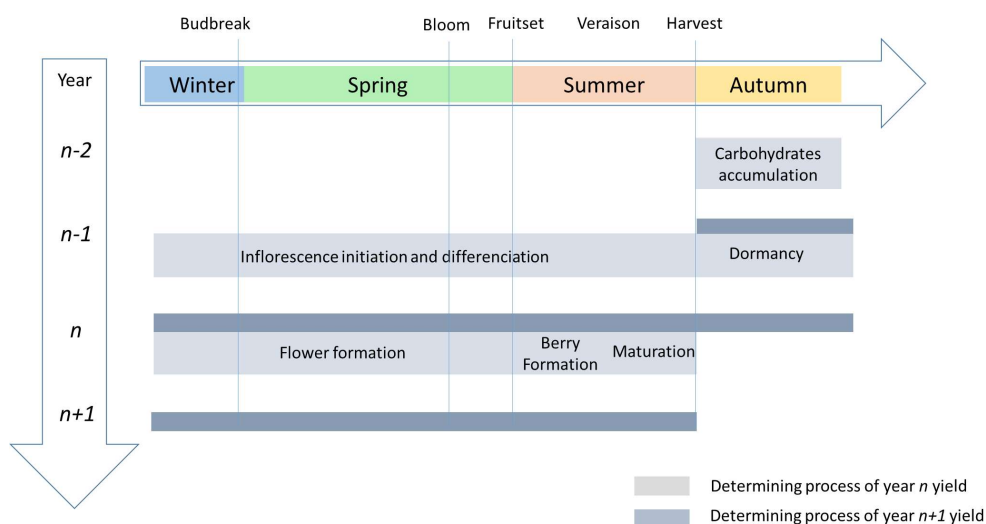


Figure 1. Key phenological stages and the temporal pattern of the grapevine yield determining process over two seasons (aligned for the northern hemisphere).

The number of clusters per vine (CN) is a major yield component, said to explain up to 60% of the final yield variability (Clingeffer, 2001). However, understanding the timing of its elaboration and its sensitivity to external factors such as temperature is complicated by its interdependency with other physiological processes over time. Therefore, there is a challenge to determine all the periods during which any explanatory variable, such as temperature, has an impact on CN taking into account that the effects of these periods are correlated to many others periods' multivariable impacts. To do this, classical approaches have been carried out using key phenological stages (Buttrose, 1974 ; Pouget, 1981) or time steps (Guilpart et al., 2014 ; Molitor & Keller, 2017). They have shown that temperature is one of the major factors driving grape fruitfulness (Buttrose, 1974 ; Srinivasan & Mullins, 1981 ; Petrie & Clingeffer, 2005). In the literature, periods of phenological development during year $n-1$ for one cycle are referred to by their concomitance with the phenological stages of year n . For example, the development of inflorescence primordia linked to the yield in year n starts

with budbreak in year $n-1$ (Vasconcelos et al., 2009). Following this notation, a few hours of elevated temperature around budbreak of year $n-1$ should be sufficient to positively impact inflorescences' or clusters' number per vine in year n (Buttrose, 1974 ; Srinivasan & Mullins, 1981 ; Pouget, 1981 ; Lobell et al., 2006). Critical stages have also been highlighted around bloom (Durquety et al. ; 1982 ; Molitor & Keller, 2017) of year $n-1$. Molitor and Keller (2017) found that high maximum temperatures around veraison and low average temperatures during maturation of year $n-1$ favoured high yield.

As a study case, this paper aims to assess whether using the BLiSS method to analyse the impact of the whole history of temperature in year $n-1$ on CN permits the identification of critical periods previously described and well understood within literature. It also tests the potential detection of previously unknown critical periods.

Materials and Methods

Data description

Clusters Number per vine data

Data were collected from a commercial Cabernet-Sauvignon (*Vitis vinifera* cv. *Cabernet Sauvignon*) vineyard field (~1 ha) in the Bordeaux region, France. The vines are short-pruned, trellised, non-irrigated, partially grass covered in the inter-row and planted at a density of 8696 vines/ha. The CN was determined after fruitset on the same 100 vines in the vineyard from 2007 to 2018, with the exception of 2014. Note that depending on the seasonal conditions, cluster thinning is often performed in this vineyard, such that the final harvest differs from the amount of fruit set. The occurrence date of the principal phenological stages (budbreak, bloom, fruitset, veraison) was recorded by the vineyard manager according to the Gregorian calendar from 2006 to 2018.

Temperature data

Temperature data were collected by a local weather station (provided by DE.MET.E.R, Villenave d'Ornon, France) located in the wine estate from 2006 to 2018. Multiple measurements were observed, however, only daily minimum (T_{min}) and average (T_{avg}) temperature (°C) are considered in this analysis as their impact on grape yield is well documented in the literature.

The T_{min} and T_{avg} influences on CN have only been studied during the growing season of year $n-1$ (from 1st of March to 25th of October). The potential impact of winter or the growing season of year n was not investigated in this preliminary study, but it could be included following the same steps. To reduce processing time, temperature was aggregated into 10-day periods. The T_{min} therefore corresponded to the minimum daily temperature during each of these periods. The T_{avg} corresponded to the average of daily temperatures over the 10-day periods. It is expected that better software and increased computing power will remove the need for this pre-processing step in the near future.

Analysis by the BLiSS method

Theory

The BLiSS method (Grollemund et al., 2019) allows quantitative data (variable Y) to be explained by functional data (variable X) using a linear functional regression model.

$$Y = \alpha + \int_0^T \beta(t)X(t)dt + \varepsilon \quad (1)$$

Where Y is the explained variable, α a scalar, X the explanatory variable, $\beta(t)$ the parameter function to be estimated and ε a residual error supposed to follow a normal distribution $N(0, \sigma^2)$. In this paper, Y corresponds to CN and X corresponds to $Tmin$ or $Tavg$. BLiSS is a Bayesian method: it supposes that the parameter $\beta(t)$ is random, formulates an *a priori* estimation of the $\beta(t)$ distribution and updates it with the data to produce an *a posteriori* estimation of the $\beta(t)$ distribution. The BLiSS method investigates the estimated distribution support, *i.e.* all the t instants for which $\beta(t)$ differs from 0. The BLiSS method thus delivers two estimators: one of the $\beta(t)$ function and another of the $\beta(t)$ function support (here time). The latter allows periods during which X has an impact on Y to be identified.

The output of BLiSS is best represented in a graphical form (*e.g.* Fig. 2). By way of explanation, Fig. 2 features the *a posteriori* distribution of the BLiSS estimator of $\beta(t)$ as a line estimated over time. If the $\beta(t)$ estimator is null ($= 0$) during a certain period, it means that the variable X is not related to Y . If it has a positive value during a given period of time, X is positively correlated to Y . In contrast, if it has a negative value during a period of time, X is negatively correlated to Y . The duration of the detected periods corresponds to their time length (distance on the x-axis). The magnitude of the detected periods corresponds to their estimated $\beta(t)$ value (distance on the y-axis). The higher the magnitude, the greater the impact X has on Y . The BLiSS method also allows the correlation between X and Y to be estimated. Practically, it corresponds to the influencing period detection confidence. The colour code is a visual way to represent it. The darker the colour is (red on the extreme case), the more confident the correlation between X and Y during this period. In contrast, the lighter the colour is (white on the extreme case), the less confident the correlation.

Analysis

Each individual corresponds to a CN per year. The dataset was therefore composed of 5368 individuals to be analysed. The analysis was performed using the package *bliss* version 1.0.0 (Grollemund, 2019) in R 3.5.1. The $Tmin$ and $Tavg$ time series were analysed independently.

Results interpretation

Assuming that the ecophysiological results reported in the literature and reviewed in the introduction represent the diversity of the classical method applications, the BLiSS method results were evaluated by comparison with the identified critical time periods. The literature review focused on temperature impact on the number of inflorescences or clusters per vine and on production yield during the year $n-1$.

Results

Figure 2A shows the influence of daily minimum temperature during the year $n-1$ on CN . Four periods of influence were detected: (i) from mid-March to late April, (ii) from

mid-July to early August, (iii) from early August to late September and (iv) from late September to mid-October. The detection of each of these periods had an adequate degree of confidence (dark colour). Three out of these four periods (i,ii and iv) were positively correlated to the *CN* i.e. elevated *Tmin* during these periods favoured a high *CN*. In contrast, high *Tmin* during the (iii) period favoured a lower *CN*. The duration of periods (i) and (iii) exceeded one month whereas the period (ii) and (iv) lasted between two and three weeks. The absolute magnitude for each period gave an indication of the intensity of the temperature impact on *CN*. It was 6 for periods (ii) and (iv) and respectively 1.5 and 3 for periods (i) and (iii).

Figure 2B shows the influence of daily average temperature on *CN* during year $n-1$. Three periods of influence were detected: (v) from late April to early June, (vi) from late June to late July and (vii) from late July to early September. The detection of each period had a lesser degree of confidence (lighter yellow) than periods (i) and (iii) (Fig. 1A). Two periods, (v) and (vi), were positively correlated to *CN*, meaning high *Tavg* during these periods favoured a higher *CN*. In contrast, high *Tavg* during period (vii) reduced *CN*. The duration of these three periods were between one month and a month and a half. The absolute magnitude of periods (v), (vi) and (vii) were 9, 2 and 11 respectively.

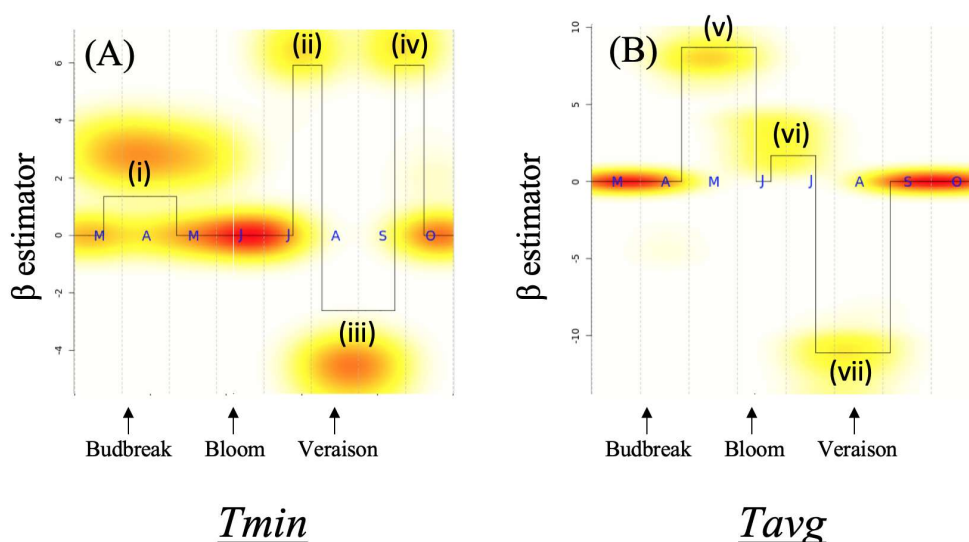


Figure 2. Analysis of possible periods of influence respectively of (A) daily minimal temperature (*Tmin*) and (B) average temperature (*Tavg*) on the number of clusters per vine from 1st of March to 25th of October of year $n-1$. The black line indicates the β estimator distribution. The colour gradient from white to red illustrates the density probability function of the β estimator distribution. The x axis labels *M,A,M,J,J,A,S,O* represent March, April, May, June, July, August, September, October respectively. Average dates of budbreak, bloom and veraison recorded on the commercial field from 2006 to 2018 vintages are indicated.

The BLiSS results are in agreement with literature

Period (i) (Fig. 2A) was associated with the budbreak period in the Bordeaux region. It is in agreement with Pouget's (1981) results on the number of inflorescences per vine under experimental conditions and the findings of Lobbell et al. (2006) when studying

night temperature around budbreak in California. Period (v) corresponded to the bloom period in the Bordeaux region. This agrees with the results of Durquety et al. (1982) on *CN* for Petit Manseng and the Molitor and Keller (2017) study on the impact of minimal temperature on production yield for Müller-Thurgau and Riesling in Alsace (North-East of France). In the latter case, the correlation between minimum temperature around bloom and yield was tested and found positive, as observed in period (v) (Figure 2B). Period (ii) coincided with veraison in the Bordeaux region. Only Molitor and Keller (2017) reported this period studying daily *Tmax* effects on yield for Müller-Thurgau. The (iii) and (vii) periods are associated with the maturation period in the Bordeaux region. Molitor and Keller (2017) also reported a negative impact of minimum temperature on future yield potential during this period.

The BLiSS results allow further exploration of the data

Periods (iv) and (vi) have not been previously cited in the literature but they had a good degree of certainty according to the BLiSS method. They respectively correspond to a post-harvest and a fruitset period. Since the method correctly identified known periods of influence (previous section), the hypothesis is that these periods are also influential on *CN*. This shows the potential ability of BLiSS to highlight unknown periods of influence of temperature on *CN*. Unlike the classical method results, for which the duration of the resulting periods is determined by the pre-analysis choices, the periods detected by the BLiSS method were characterized by different durations. It is also interesting to note that the temperature influence during summer was characterized by several periods in Fig. 2, whereas Molitor and Keller (2017) only reported one period of long duration. The BLiSS method had also refined the duration of known key influencing periods. The proposed alterations in duration and the identification of new periods of influence now need to be properly validated in a viticulture context. If true, these results have increased the knowledge and the temporal resolution of environmental effects on yield and consequent potential management.

Discussion

The BLiSS method : a novel and relevant method to analyse time series data

The BLiSS method allowed the detection of already confirmed periods of daily minimal and average temperature on *CN* and potential vineyard yield. This point validated the use of this method to analyse time series of agricultural data. Moreover, the method allowed the detection of previously unknown periods of influence or the refinement of the duration of known key periods. This demonstrated the advantages of a linear functional model in a Bayesian framework for the temporal analysis of agricultural data.

Comment on the interpretation of the BLiSS results

Each period that was detected with the BLiSS method was characterized by a magnitude, a duration and a degree of confidence. There are therefore three criteria that can be used to compare the different periods. However, classifying the periods by decreasing order of influence required an integration of the information linked to both the degree of confidence and the magnitude. This may not be simple in every case. For example, a high magnitude but medium degree of confidence period cannot be

prioritized in comparison with a high degree of confidence but medium magnitude period. Thus, the interpretation of the BLiSS results should be decomposed into the three criteria first and an expert knowledge of the field of study should then balance the interpretation. Clearer interpretation of the BLiSS outputs is an area for further development.

Potentialities for various study field analysis using the BLiSS

The BLiSS method allows further exploration of many types of datasets. Any variable characterized by time series data, which is a common case in agriculture, can be analysed with the BLiSS method. The method could also be applied to other variables evolving over different continuous supports such as wavelength or spatial coordinates.

While two univariate cases were presented here, several explanatory variables could be studied in a multivariable analysis to identify their potential combined effects as well as their individual (univariate) effects. For example, *Tmin* and *Tavg* have simultaneously been considered in a bivariate BLiSS analysis to distinguish their proper effects on *CN* (data not shown).

However, there are some limitations to the current BLiSS method. Firstly, the computation time has to be taken into account when designing the analysis. For example, the results presented in Fig. 2 (derived from 5368 individuals and using 27 10-day climatic periods) took approximately 4 hours on a machine of 4 CPU and 5 Gb memory. Secondly, it is not possible to know the portion of the *Y* variability explained by *X* at the moment. It is also not yet possible to include qualitative explanatory variables but this is planned to be implemented soon. In addition, it is important to remember that the BLiSS method is subject to the same limitations that any other linear analysis, such as disturbance due to the explanatory variables correlation.

Finally, the BLiSS method presents another option which has not been tested in this paper. Prior knowledge can be taken into account by the Bayesian-based method. In practice, this can help the method to detect relevant periods using *a priori* literature results or field experience, in the case of a small number of data for example.

Conclusion

The BLiSS method combines the advantages of both functional and Bayesian models to perform advanced temporal variability analysis. Functional analysis allowed a more complete and objective analysis of a data set, taking into account the explanatory variables histories and allowing detection of periods of influence in the time series. Bayesian analysis allowed a better uncertainty management using conditional probability. This approach can be used for further data exploration in many different fields, in agriculture or others. It is an interesting method to better study the effect of high temporal resolution variables, such as those generated in precision agriculture. The ability to correctly identify and manage key phenological stages in agriculture is an area that has often been overlooked in precision agriculture but is needed to improve the temporal-resolution of decision-making on farm.

Acknowledgements

The authors fully acknowledge the commercial estate that provided the data as well as the authors of the BLiSS method who spent time explaining its application.

References

- Buttrose M. S., 1974, Climatic Factors and Fruitfulness in Grapevines, *Horticultural Abstracts*, 44, 319–326
- Durquety P.M., Naude E. and Blanchard P., 1982, La Prévision de Récolte Sur Petit Manseng (*Vitis Vinifera* L.) Basée Sur Les Courbes-Niveaux de Fertilité et Les Températures Durant Une Période Critique (Yield prevision on Petit Manseng cultivar (*Vitis Vinifera* L.) based on fertility and temperature curves during a critical period), *Progrès Agricole et Viticole*, 99, 440–469
- Grollemund P., Abraham C., Baragatti M. and Pudlo P., 2019, Bayesian Functional Linear Regression with Sparse Step Functions, *Bayesian Analysis*, 14, 111-135
- Guilpart, N., Metay A., and Gary C., 2014, Grapevine Bud Fertility and Number of Berries per Bunch Are Determined by Water and Nitrogen Stress around Flowering in the Previous Year, *European Journal of Agronomy*, 54, 9–20
- Howell G.S., 2001, Sustainable Grape Productivity and the Growth-Yield Relationship: A Review, *American Journal of Enology and Viticulture*, 52, 165–174
- Keller M., Tarara J.M. and Mills L.J., 2010, Spring Temperatures Alter Reproductive Development in Grapevines. *Australian Journal of Grape and Wine Research*, 16, 445–544
- Lobell D.B., Field C.B., Cahill K.N. and Bonfils C., 2006, Impacts of Future Climate Change on California Perennial Crop Yields: Model Projections with Climate and Crop Uncertainties, *Agricultural and Forest Meteorology*, 141, 208–218
- Molitor D. and Keller M., 2017, Yield of Müller-Thurgau and Riesling Grapevines Is Altered by Meteorological Conditions in the Current and Previous Growing Seasons, *OENO One*, 50, 245–258
- Oliver M., 2013, An overview of precision agriculture. In : *Precision Agriculture for Sustainability and Environmental Protection*. Abindon, UK: Routledge. pp 21-37.
- Petrie, P. and Clingeffer P., 2005, Effects of Temperature and Light (before and after Budburst) on Inflorescence Morphology and Flower Number of Chardonnay Grapevines (*Vitis Vinifera* L.), *Australian Journal of Grape and Wine Research*, 11, 59–65
- Pouget R., 1981, Action de La Température Sur La Différenciation Des Inflorescences et Des Fleurs Durant Les Phases de Pré-Débourrement et Post-Débourrement Des Bourgeons Latents de La Vigne. (Impact of temperature on inflorescences and flowers differentiation during pre- and post- budbreak stages of the grapevine latent buds.), *Connaissance de La Vigne et Du Vin*, 15, 65–79
- Ravaz L. and Sicard L., 1911, L'effeuillage de La Vigne, (Leaf thinning for Grapevine), *Annales de l'École Nationale d'agriculture de Montpellier*, 11, 216-244
- Srinivasan C. and Mullins M.G., 1981, Physiology of Flowering in the Grapevine - A Review, *American Journal of Enology and Viticulture*, 32, 47–63
- Vasconcelos M.C., Greven M., Winefield C.S., Trought M.C.T. and Raw V., 2009, The Flowering Process of *Vitis Vinifera*: A Review, *American Journal of Enology and Viticulture*, 60, 411–434