한국농림기상학회지, 제16권 제4호(2014) (pISSN 1229-5671, eISSN 2288-1859)
Korean Journal of Agricultural and Forest Meteorology, Vol. 16, No. 4, (2014), pp. 304~308
DOI: 10.5532/KJAFM.2014.16.4.304

An Improved Method for Phenology Model Parameterization Using Sequential Optimization

Kyungdahm Yun and Soo-Hyung Kim*
School of Environmental and Forest Sciences, College of the Environment,
University of Washington, Seattle, WA 98195, USA
(Received September 5, 2014; Revised November 4, 2014; Accepted November 4, 2014)

순차적인 최적화 기법에 의한 생물계절모형 모수추정 방식 개선

윤경담 · 김수형
미국워싱턴대학교 환경대학 환경산림과학부
(2014년 9월 5일 접수; 2014년 11월 4일 수정; 2014년 11월 4일 수락)

ABSTRACT

Accurate prediction of peak bloom dates (PBD) of flowering cherry trees is critical for organizing local cherry festivals and other associated cultural and economic activities. A two-step phenology model is commonly used for predicting flowering time depending on local temperatures as a result of two consecutive steps followed by chill and heat accumulations. However, an extensive computation requirement for parameter estimation has been a limitation for its practical use. We propose a sequential parameterization method by exploiting previously unused records of development stages. With an extra constraint formed by heat accumulation between two intervening stages, each parameter can then be solved sequentially in much shorter time than the brute-force method. The result was found to be almost identical to the previous solution known for cherry trees (*Prunus* × *yedoensis*) in the Tidal Basin, Washington D.C.

***Key words*: Parameterization, Calibration, Phenology, Cherry blossom, Peak bloom**

I. Introduction

Cherry blossoms are celebrated to signify the onset of spring in many countries and cultures around the world. Cities and towns organized cherry blossom festivals that are centered on peak bloom of their local cherry trees. Therefore, it is critical to predict peak bloom date of these trees accurately using local climate data. The phenology of flowering trees including dormancy release, bud break, floral development, and peak bloom is often described using thermal time accumulations. A two-step method was proposed to predict bud-burst after dormancy release, which was modeled with two consecutive steps of chill and heat unit accumulations (Cesaraccio *et al.*, 2004). The model was later modified to predict flowering time, with an assumption that the flowering was also a consequence of extended heat accumulation (Yun, 2006).

While this approach was able to estimate peak bloom dates of flowering cherry trees in the Tidal Basin, Washington D.C., with reasonable accuracy (Chung *et al.*, 2011), adapting the model for new location and variety required an arduous parameterization process through extensive grid search. This brute-force method has several shortcomings. Firstly, a large amount of computation time was required for the grid search

* Corresponding Author : Soo-Hyung Kim
(soohkim@uw.edu)

when finding an optimal parameter set. Secondly, the parameter space was indeed not unimodal, leading to multiple solutions with a low cost of error. Thirdly, available observation records were not fully exploited, but only peak bloom dates were used. Intermediate development stages are often reported together and they could possibly improve parameterization.

In this study, we propose a sequential parameter calibration algorithm to address these issues. Each parameter is optimized individually with an additional constraint on heat accumulation between two development stages. When parameterized with the existing dataset from Tidal Basin, the result from the new approach was almost identical to the known solution found by brute-force method but with smaller errors and considerably reduced computation times (i.e., less than a minute compared to days).

II. Materials and Methods

2.1. Peak bloom model

The phenology of flowering is modeled with consecutive accumulation of two types of thermal units: chill days and anti-chill days. The unit is a degree day above 0°C and partitioned with base temperature (T_c) and its relation to daily maximum and minimum temperatures (Cesaraccio *et al.*, 2004). Heating unit (C_a) is calculated first and its difference with total degree day becomes chilling unit (C_d). Once a rest period of dormancy is initiated in October 1, the chilling units need to be accumulated up to chill requirement (R_c). It is then considered the dormancy has released and moved onto a quiescence period. From this point, the heating units are accumulated in the opposite direction to reach heat requirement (R_h) until peak bloom occurs. Consequently, an estimated peak bloom date $\hat{PB}$ of year y can be obtained by a function f of three parameters, T_c, R_c, and R_h, as shown in Eq. (1). Fig. 1 illustrates an example run of the model with associated parameters and variables.

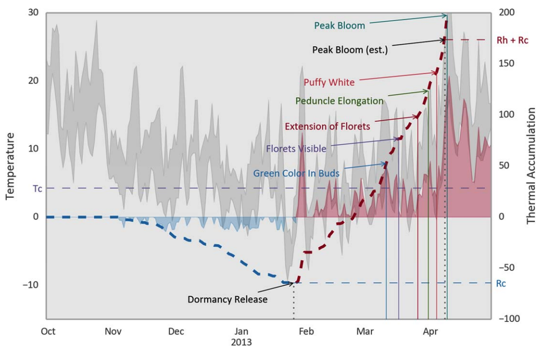

Fig. 1. An illustration of the peak bloom model ran for the year of 2013 in the Tidal Basin, Washington D.C. The parameter set (T_c = 4.28, R_c = -64.38, R_h = 238.13) was found by the proposed algorithm. Chilling units were accumulated until dormancy release at January 26, and then heating units started growing in the other direction. The estimated peak bloom date was April 9 instead of actual peak bloom date, April 10. Blue and red shades are daily units of chilling (C_d) and heating (C_a). A gray shade represents a daily variation of temperature observed at Reagan National Airport located 4 km away from the site.

$$\hat{PB}_y = f_{T_c,R_c,R_h}(y) \tag{1}$$

2.2. Phenological stages

The U.S. National Park Service monitors flower development of cherry trees (e.g., *Prunus* × *yedoensis*) in the Tidal Basin, Washington D.C. in regards with five distinct stages: green color in buds (GC), florets visible (FV), extension of florets (EF), peduncle elongation (PE), puffy white (PW), and peak bloom (PB). The first stage is usually detected in mid February to early March, and then followed by next stages until peak bloom, which usually happens around late March or early April. As visible changes should be accompanied by bud burst after dormancy release (DR), noticing these stages presumably mean that it is in the second step of the model, where only heat accumulation is counted. Eq. (2) states that the heat requirement δ for year y between any two stages, α and β, is simply a sum of heat accumulation C_a in the corresponding period p. T_c is required for driving C_a, but R_c is not, because only the relative difference is needed in this case.

$$\delta_{T_c,\alpha,\beta}(y) = \sum_{p=\alpha_y}^{\beta_y} c_a(p, T_c) \tag{2}$$

2.3. Sequential optimization

Calibrating three parameters, T_c, R_c, and R_h, is an optimization problem of minimizing sum of squared differences between the estimated peak bloom dates ($\hat{PB}$) and the actual dates (PB) observed in the site. Instead of a time-consuming grid search fitting all parameters simultaneously (Chung *et al.*, 2011), a sequential algorithm is applied for fitting each parameter separately, using other parameters discovered in the

Table 1. Comparison of two solutions in various metrics: root mean square error (RMSE), mean absolute error (MAE), maximum absolute error (XE), and total error (TE). Error units are in days and time unit is in seconds. The running time of brute-force algorithm depends on the number of grid points N, which may scale up to millions.

Algorithm	T_c	R_c	R_h	R_h-R_c	RMSE	MAE	XAE	TE	Time
Brute-force	4.3	-78.9	221.1	300.0	3.69	2.7	9	51	0.25N
Sequential	4.28	-64.38	238.13	302.51	3.18	2.6	7	50	33

previous steps. Sequential parameter estimation is often preferred over simultaneous procedure when phenology parameters have no known interdependencies (Wallach *et al.*, 2014).

T_c is the first parameter in a series as it controls the amount of thermal units, thus affecting how much degree R_c and R_h are being filled up. Assuming an equivalent heat requirement for each year y between the first and last stages, GC and PB, $\hat{T}_c$ is estimated by minimizing sum of annual variances as shown in Eq. (3).

$$\hat{T}_c = \underset{T_c}{\mathrm{argmin}} \sum_y (\delta_{T_c,\alpha,\beta}(y) - \overline{\delta_{T_c,\alpha,\beta}})^2 \Big|_{\alpha=\mathrm{GC}}^{\beta=\mathrm{PB}} \tag{3}$$

R_c can be found in a similar manner. With a chilling unit (C_d) for the range determined by a given $\hat{T}_c$, a dormancy release date (DR) now becomes a function of R_c. $\hat{R}_c$ is then estimated by minimizing sum of variances of the entire heat requirement ($\delta_{T_c,DR,PB} = R_h - R_c$) as shown in Eq. (4).

(4)

After two parameters revealed, R_h is solved by an original scheme shown in Eq. (5) that minimizes sum of differences between estimated and actual peak bloom dates, $\hat{PB}$ and PB, respectively.

$$\hat{T}_c = \underset{R_h}{\mathrm{argmin}} \sum_y (f_{\hat{T}_c,\hat{R}_c,R_h}(y) - \mathrm{PB}_y)^2 \tag{5}$$

III. Results and Discussion

3.1. Calibration

Phenology observation records for cherry trees (*Prunus × yedoensis*) in the Tidal Basin, Washington D.C. were used for parameterization. The starting year was set to 1992 because the detailed records have been only available since then. The end year was chosen to be 2010 for comparison with previous research (Chung *et al.*, 2011).

Our approach resulted in T_c estimate of 4.28°C, which is almost identical to the solution found in Chung *et al.* (2011) using a grid search method (Table 1). On the other hand, our R_c and R_h were slightly larger by 14.52 and 17.03 respectively than those from Chung *et al.* (2011). Interestingly, the entire heat requirements (R_h-R_c) were still very close to each other. It would suggest that our solution successfully captured a right amount of heat accumulation required for the phenology development.

As the performance of the previous algorithm was dependent on the grid size, it could easily become very slow. A fine-grained search whose T_c spanning from 0.0°C to 10.0°C by 0.1°C, R_c from -200 to 0 by 1, and R_h from 100 to 200 by 1, would generate a grid of 100·200·200=4,000,000 points, which may take up to several days to run. On the other hand, our algorithm had a constant running time, less than a minute, which would be almost negligible on regular circumstances.

Various metrics indicated that our solution was marginally better in terms of error, at least during the cali-

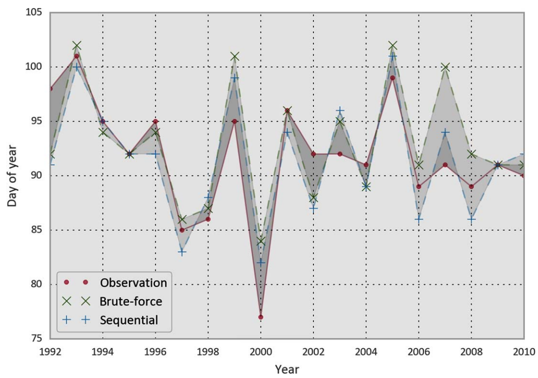

Fig. 2. Evaluation of calibrated parameters. The actual peak bloom dates are plotted with estimated dates from two different parameter sets. A brute-force method was parameterized with data of the Tidal Basin from 1991 to 2010. Our solution was calibrated with the same data, but from 1992, due to missing phenology observation.

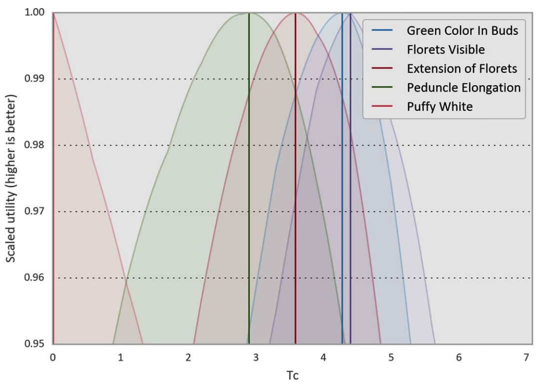

Fig. 3. T_c estimated with different heat constraints: 4.28 for GB, 4.40 for FV, 3.59 for EF, 2.90 for PE, and 0.00 for PW. If the distance to PB becomes smaller, its corresponding T_c also tends to decrease and finally converge to zero.

bration period. Fig. 2 shows a plot of peak bloom dates and corresponding errors from each parameter set.

3.2. Selective constraints

The most important part of our sequential optimization algorithm is lying on the use of additional constraints formed by observing multiple phenological stages. Not only GB used in the proposed method, different stages could have been plugged into α for estimating T_c in Eq. (3). However, the profiling result shown in Fig. 3 suggested that other constraints might not be sufficient to construct a stable constraint. Only GB and FV had approached a previously known value of T_c, 4.3°C, with 4.28°C and 4.40°C respectively, while others had kept decreasing like until PW reached 0.0001°C.

The heat variance is basically controlled by T_c in the middle balancing out an amount of heating and chilling units. Once T_c becomes smaller than minimum temperature, it can no longer help minimizing the variance. T_c of 0°C is actually a boundary case by definition where no chilling unit can be generated at all. Therefore, later stages from much warm environments are apparently more susceptible to degeneracy.

3.3. Strengths and limitations of new approach

A sequential algorithm could estimate a decent parameter set that almost matches up with one found from a brute-force method. The entire process took only a fraction of time compared to the previous method, while calibration error was even smaller. However, its own drawbacks were also revealed. Observation of other development stage is mandatory and requires extra attention on its selection. Temperature range between two stages should span around base temperature, T_c, to ensure valid thermal accumulation.

IV. Conclusions

We propose a sequential parameter calibration algorithm for a cherry blossom phenology model. While shown its own strength in terms of speed and accuracy, it may not be always applicable because of the extra requirement. Additional constraints should be carefully chosen from intervening development stages that can balance out thermal unit generation. To apply this method successfully, it is advised to expand observations to include earlier phenological developments during heat accumulation step for avoiding the degeneracy issue on optimization.

적 요

벚꽃의 만개일은 관련 행사일정을 결정하는 중요한 요소로써 생육기간 중 기온에 따른 변화의 폭이 크다. 이를 예측하기 위한 방법으로는 벚꽃의 발달을 휴면기와 생장기의 2단계로 구분하여 저온(chill)과 고온(heat) 요구에 대한 온도시간(thermal time) 누적을 기술하는 모형이 개발되어 있다. 하지만 모수 추정시 모수공간내 일정 간격의 격자 전체를 계산하여 많은 시간을 소모한다는 단점이 있었다. 본 연구에서는 기존 모형이 고려하지 않던 벚꽃 발달의 중간단계 관측자료를 활용하여 고온요구에 대한 새로운 조건을 추가하고, 이를 기반으로 각 모수를 순차적으로 추정하여 최적화 시간을 단축하는 새로운 방법을 제안한다. 미국 워싱턴 DC 지역의 벚꽃개화 관측 자료를 기준으로 검증한 결과, 기존 모형에서 제안된 모수와 근사한 값을 단축된 시간 내에 계산해내는 것을 확인하였다.

Acknowledgements

This study was carried out with the support of "Cooperative Research Program for Agricultural Science & Technology Development (Project No. PJ01000707)", Rural Development Administration, Republic of Korea.

REFERENCES

Cesaraccio, C., D. Spano, R. L. Snyder, and P. Duce, 2004: Chilling and forcing model to predict bud-burst of crop and forest species. *Agricultural and Forest Meteorology*

126(1-2), 1-13.

Chung, U., L. Mack, J. I. Yun, and S.-H. Kim, 2011: Predicting the timing of cherry blossoms in Washington, DC and Mid-Atlantic states in response to climate change. *PloS one* **6**(11), e27439.

Wallach, D., D. Makowski, J. W. Jones, and F. Brun, 2013: *Working with Dynamic Crop Models (2nd ed.)*. Academic Press, 247-248.

Yun, J. I., 2006: Climate change impact on the flowering season of Japanese cherry (*prunus serrulata* var. *spontanea*) in Korea during 1941-2100. *Korean Journal of Agricultural and Forestry Meteorology* **8**(2), 68-76.