

## WHEN ARE TWO PHENOLOGICAL PATTERNS DIFFERENT?

GEORGE F. ESTABROOK,\* JAMES A. WINSOR,\* ANDREW G. STEPHENSON,† AND HENRY F. HOWE‡

\*Department of Botany, University of Michigan, Ann Arbor, Michigan 48109; †Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802; and ‡Department of Zoology, University of Iowa, Iowa City, Iowa 52242

The importance of the timing of certain events, such as flowering, germination of seeds, bud break, and dispersal, in the lives of many species has long been recognized. In many cases, the phenological patterns of these events have evolved by natural selection. In some cases, we can formulate hypotheses of adaptation concerning phenological patterns. These hypotheses can be tested by observing phenologies only if there is some valid method for determining whether two phenological patterns are different. This paper describes such a method and presents three natural examples of its use.

### Introduction

In many species, the timing of various activities is important to survival and reproductive success and most probably has been determined, at least in part, by natural selection. Phenological events such as germination, bud break, flowering, fruit dehiscence, and leaf drop may be timed to coordinate with abiotic environmental conditions, pollinators, competitors, herbivores, dispersers, and others (HEITHAUS 1974; FRANKIE 1975; PULLIAM and BRAND 1975; OPLER, FRANKIE, and BAKER 1976; THOMPSON and WILLSON 1979; STILES 1980). The development in a few generations of horticultural varieties for early and late harvests attests to the speed with which phenological patterns can evolve in response to intense selection. Modern workers often collect data on the timing of phenological events. When such timing data have been collected for a species under two conditions or for two species under similar or different conditions, it is important to decide whether the timing is the same or different.

In this paper we describe a procedure to determine whether sufficient evidence exists to conclude that two timing patterns are different. This nonparametric statistical procedure is relatively free of assumptions and can be applied easily without a computer or calculator.

### Methods

To make the exposition concrete, we shall describe our method in terms of flowering, but consideration of the examples to follow will make it clear how to apply the procedure to other analogous situations. Consider two groups of plants: Do they flower at the same time? To answer this question, we must consider what data can be reasonably gathered to generate statistics, and we must make more precise,

in terms of probability, what is meant by the question.

To collect data, some flower buds in each group of plants are marked individually, before any flower buds have opened, or after only a very few have opened. Then on various dates during the flowering season, the marked buds are observed to see if they have opened yet. For each bud, the date on which it is first observed to have opened is noted. This is not necessarily the date on which it opened but may be later than that date. On any day of observation, all the previously unopened buds in a group are checked, although both groups need not be checked on the same day. These observations continue at intervals until all buds in both groups are opened. Aborted or dormant buds, which never open, should not be counted in the sample.

The data should be arranged in a table with a column for observation dates and a column for each group of plants. Opposite each observation date, record the number of flower buds from each group that are known to have opened by that time. Thus, opposite the last observation date should be entered the total number of buds that eventually flowered in each group. These are the data.

To conclude that the two groups have different timing, we must consider each flower bud to represent a single sample from a probability distribution that assigns to each date the probability that a bud will flower on that date. We suppose that all the flower buds within the same group of plants are governed by the same probability distribution. To state that two groups of plants have the same timing for flowering is to state that the probability that a bud will open on any given day is the same for all buds in both groups.

We use the timing data to test the null hypothesis that the probability that any flower bud will open on any given day is the same in both groups. First, divide each entry in the column for each group by the total number of marked buds to open in that group (i.e., divide every entry in a column by the last entry in that column) to express the data in

*Manuscript received October 1981; revised manuscript received March 1982.*

Address for correspondence and reprints: GEORGE F. ESTABROOK, Department of Botany, University of Michigan, Ann Arbor, Michigan 48109.

fractions of the total in each group and enter these fractions into two new columns of the data table. Then, for each date, determine the absolute value of the difference between the two fractions for the two groups and enter these differences into a new column of the data table. Choose the maximum absolute difference, and call it  $D$ . The larger is  $D$ , the stronger is the evidence that the two phenological patterns are different. We would like to determine the thresholds,  $T$ , that  $D$  will exceed with probability .05 or .01. For what value of  $T$  is it true that  $P(D > T) = .05$  (or = .01)? To determine these thresholds, we use the distributions derived by SMIRNOV (1939) and described by CONOVER (1971) or by SOKAL and ROHLF (1969). Accordingly:

$$P(D > 1.36 * \{[(m + n)/m*n]**(1/2)\}) = .05,$$

$$P(D > 1.63 * \{[(m + n)/m*n]**(1/2)\}) = .01.$$

In these statements,  $m$  is the number of buds that eventually flowered in one group, and  $n$  is the number of buds that eventually flowered in the other group. The thresholds provided above are the result of asymptotic convergences, and to ensure their accuracy, both  $m$  and  $n$  should be in excess of 20.

### Applications

#### 1. GERMINATION

The time a seedling emerges from the soil relative to its neighbors is a determinant of its success in competition (HARPER 1961; BLACK and WILKINSON 1963). A seedling that emerges early has the benefit of a longer growing season; in addition, the temporary absence of competition enables it to gain an advantage, which may be accentuated as the season progresses (ROSS and HARPER 1972).

*Impatiens capensis* Meerb. (Balsaminaceae) is an annual herb of wet ground that frequently occurs with its perennial competitor, *Urtica dioica* L. (Urticaceae). LECK (1979) reported that *I. capensis* seedlings emerge very early in the spring, but the emergence times of co-occurring species were not recorded. WINSOR (1981) studied a site where both species occur; *I. capensis* and *U. dioica* seeds were planted in the fall, and seedling emergence was observed in the spring. The number of newly emerged seedlings of each species was recorded on 15 dates over 30 days. The seedlings were marked so that they would be counted only once. The data and calculations to test for a difference in emergence times for these two competing species are presented in table 1. Note that the time intervals for observation need not be equal.

The maximum difference of 0.719 (tables 1–3, asterisk) far exceeds the critical value of 0.103 for the .05 confidence level, indicating that the times of emergence for the two species are indeed different.

TABLE 1  
DISTRIBUTIONS OF EMERGENCE TIMES FOR  
IMPATIENS CAPENSIS AND URTICA  
DIOICA SEEDLINGS

DATE	<i>I. capensis</i>		DIFFER- ENCE	<i>U. dioica</i>	
	Count	Fraction		Fraction	Count
April:					
16.....	2	.010	.010	.000	0
18.....	38	.185	.167	.018	20
19.....	64	.311	.286	.025	28
21.....	107	.519	.462	.057	63
23.....	137	.665	.589	.076	84
25.....	167	.811	.700	.111	122
29.....	181	.879	.719*	.160	177
30.....	200	.971	.709	.262	289
May:					
2.....	204	.990	.543	.447	493
5.....	206	1.000	.212	.788	870
7.....	206	1.000	.137	.863	953
9.....	206	1.000	.081	.919	1,014
12.....	206	1.000	.036	.964	1,064
14.....	206	1.000	.006	.994	1,097
16.....	206	1.000	.000	1.000	1,104

NOTE.— $m = 206$ ,  $n = 1104$ . Asterisk denotes the maximum difference.

$$5\% \text{ threshold} = 1.36 * \{[(206 + 1104)/206*1104]**(1/2)\} = .103.$$

Therefore, we can interpret the trend to conclude that *I. capensis* seedlings emerge earlier than *U. dioica* seedlings. If we move the distribution for *Impatiens* so that April 23 corresponds to May 5 for *Urtica*, the maximum difference is minimized and falls to 0.123. Thus, *Impatiens* emerges about 12 days earlier than *Urtica*. But even this minimum maximum difference is still significant at the 5% level, so *Urtica* and *Impatiens* probably have different “shaped” germination timing patterns as well; i.e., *Urtica* emerges over a longer period of time.

Such moving of distributions in time and retesting for shape differences is a post hoc procedure. As such, it can be effective in generating specific new phenological hypotheses about the shape of timing distributions, but to maximize the statistical validity of conclusions concerning any new hypotheses that might be generated in this way, a new experiment should be undertaken to generate new data with which to test it statistically.

#### 2. FRUIT ABORTION

*Catalpa speciosa* (Warder ex Barney) Englem. (Bignoniaceae) and many other species consistently produce mature fruits from only a small fraction of the fruits that initiate. In these species the number of mature fruits is limited by resources. Juvenile fruits abort when there is an imbalance between the resource needs of the developing fruits and the ability of the parent plant to provide the resources necessary for growth (STEPHENSON 1981). In *C. speciosa*, the proportion of juvenile fruits that mature depends on many factors, such as the number of

juvenile fruits competing for the limited resources and the total resources available for reproduction. These resources in turn are influenced by the intensity of herbivory and the quality of the growing season (STEPHENSON 1979, 1980). Consequently, the proportion of initiated fruits that mature varies from year to year.

The fruits of *C. speciosa* have a characteristic growth pattern. Flowering terminates in late June, and during July the fruits grow very slowly. By August 1, the healthy, developing fruits contain less than 10% of their mature dry weight and total protein. During August, the fruits grow rapidly to their mature size (STEPHENSON 1980). We hypothesize that the timing of fruit abortion in *C. speciosa* should not vary annually, as does the proportion of fruits that abort, but rather it should be fixed by natural selection to precede the period of maximal resource investment by the maternal parent.

To test this hypothesis, the timing of fruit abortion on the same *C. speciosa* tree was monitored in both 1976 and 1977. In 1976, 207 fruits on 75 infructescences and, in 1977, 299 fruits on 100 infructescences were marked on July 1. At ca. 2-wk intervals, the number of aborted fruits was recorded. The data and computations to test for differences in fruit abortion timing in 1976 and in 1977 are shown in table 2. The greatest difference was 0.101 on July 20. This value is less than the 5% threshold of 0.143; thus, the timing distributions for the 2 yr are not shown to be different. Note that the dates of observation are not always the same for both years. In 1976, no observation was made on August 12. The correct count for August 12 is actually between 149 and 159. To calculate the difference for August 12, we chose the extreme that is most inimical to our hypothesis, in this case 159. Even this inflated value is not large enough to conclude that the timing in the 2 yr is different.

In 1976 and 1977, respectively, 81% and 91% of the abortion activity took place during the unobserved interval between July 7 and July 20. Differences in excess of the 5% threshold might

have occurred in this interval. It would have been a stronger conclusion if observations had been made during this time. However, if the time scale of interest to the biological phenomenon is 2 wk, then this interval is the appropriate one for making observations. If differences existed in the timing patterns but the timing patterns were displaced temporally by an amount too small to be of interest biologically, the statistical conclusion would have no relevance to the biological argument. The choice of time interval should be considered in advance of taking data: The measurement frequency should be often enough to measure rates at a biologically relevant accuracy, yet not so frequent as to waste research effort by taking unnecessary data. For *C. speciosa*, it is clear that the period of fruit abortion precedes the period of maximal resource investment by the maternal parent, whether 78% of the crop is aborted as in 1976 or 67% of the crop is aborted as in 1977.

### 3. DISPERSAL

HOWE (1977), HOWE and ESTABROOK (1977), and HOWE and VANDE KERCKHOVE (1979) argued that the timing of the availability to dispersers of fruit of some tropical trees may have evolved under the control of natural selection, with the tree manipulating the disperser to its advantage. H. F. HOWE gathered data in 1976 on the timing of fruit availability in trees of *Casearia corymbosa* HBK (Flacourtiaceae) in the seasonal dry forest in Parque Nacional Santa Rosa, Provincia de Guanacaste, Costa Rica. We present some of them here to test the hypothesis that two trees offer fruits to dispersers in the same temporal pattern. To illustrate the flexibility of the test, we compare a tree offering many fruits with a tree offering few, so that  $m$  and  $n$  are of strikingly different magnitudes. Since fruits ripen overnight and are removed by the dispersers during the morning, it is easy to observe the timing of the availability of fruits by counting the brilliantly colored fruits on the trees early each morning before the dispersers arrive. Table 3 presents the data and calculations to test the hypothesis that these two trees present fruits to dispersers in different timing patterns.

The largest difference, 0.285, which occurred on October 3, is nearly twice the 5% threshold of 0.140, providing evidence that the two timing patterns are not the same. The tree with more fruit is making them available somewhat sooner than the tree with fewer fruits. The difference in the distributions lies between September 30 and October 5, during which both trees offer about 50% of their total crop. But the tree with more fruits makes its largest offerings about 3 days earlier than the tree with fewer fruits. Although the differences are there, it remains to be argued whether 3 days are enough to be biologically meaningful. It was not possible to take the data at

TABLE 2  
DISTRIBUTION OF TIME OF FRUIT ABORTION  
BY CATALPA SPECIOSA IN 1976 AND 1977

DATE	1976		DIFFER- ENCE	1977	
	Count	Fraction		Fraction	Count
July 7 . . . . .	0	.000	.000	.000	0
July 20 . . . . .	131	.809	.101*	.910	182
Aug 1 . . . . .	149	.920	.075	.995	199
Aug 12 . . . . .	149	.920	.080	1.000	200
Aug 17 . . . . .	159	.981	.019	1.000	200
Sept 1 . . . . .	162	1.000	.000	1.000	200

NOTE.— $m = 162$ ,  $n = 200$ . Asterisk denotes the maximum difference.

5% threshold =  $1.36 * \{[(162 + 200) / (162 * 200)]^{1/2}\} = .143$ .

TABLE 3  
DISTRIBUTION OF TIME OF AVAILABILITY OF FRUITS  
IN TWO TREES OF CASEARIA CORYMBOSA

DATE	MANY FRUITS		DIFFER- ENCE	FEW FRUITS	
	Count	Fraction		Fraction	Count
Sept:					
20.....	0	.000	.010	.010	1
21.....	0	.000	.019	.019	2
22.....	3	.003	.016	.019	2
23.....	4	.004	.015	.019	2
24.....	22	.024	.005	.019	2
25.....	28	.031	.012	.019	2
26.....	29	.032	.013	.019	2
27.....	30	.033	.014	.019	2
28.....	34	.037	.018	.019	2
29.....	117	.128	.099	.029	3
30.....	244	.067	.219	.048	5
Oct.:					
1.....	299	.327	.251	.076	8
2.....	353	.389	.218	.171	18
3.....	531	.580	.285*	.295	31
4.....	594	.649	.230	.419	44
5.....	672	.735	.163	.572	60
6.....	699	.765	.060	.705	74
7.....	788	.862	.052	.810	85
8.....	824	.901	.015	.886	93
9.....	850	.928	.023	.905	95
10.....	871	.951	.018	.933	98
11.....	895	.977	.003	.974	102
12.....	915	1.000	.000	1.000	105

NOTE.— $m = 915$ ,  $n = 105$ . Asterisk denotes the maximum difference.  
5% threshold =  $1.36 * \{[(915 + 105)/915 * 105]^{.5} * (1/2)\} = .140$ .

weekly intervals, because the dispersers visit and remove fruit from the trees on a daily basis; but if they had been taken only on September 22, September 29, October 6, and October 13, then no significant differences would have been observed. The rains began on September 26, after which the bulk of the fruit was offered to the dispersers, following a lag of a few days, at a high constant rate until much of it was gone.

The two trees in this example differ in size and the total amount of fruit offered, as well as in many other respects. The hypothesis that the timing pattern in these two trees is the same has been tested with large sample sizes and rejected with confidence. However, the hypothesis that this difference in timing is causally related to tree size or size of total fruit crop cannot be confidently tested with a single tree of each kind. Nonetheless, determining as we have that the timing patterns are different is a necessary precursor to advancing hypotheses to explain the differences shown to exist.

### Discussion

We have presented here a method to determine whether the timing of a given activity in two groups of organisms is the same or different, and we have illustrated it with three natural examples. The ability of the method to accommodate the exigencies of data collecting is one of its strengths, but there are

costs, usually small, for this flexibility. Dates on which data are recorded need not be evenly spaced or numerous, but the number of entities (flower buds, fruits, seeds) representing single samples of the timing probability distributions should be large. A minimum of 20 events in each group is required to ensure the accuracy of the calculation with this asymptotic formula. Increasing the number of observed events will greatly add to the power of the test. For example, for  $m = n = 20$ , a maximum difference of at least 0.431 would be required to detect a difference at the 5% confidence level, but with  $m = n = 200$ , a maximum difference of 0.136 would be sufficient.

Because the test statistic is simply the maximum observed difference in cumulative frequency, increasing the number of dates when observations are made can only increase the value of this maximum and thus can only increase the chances that the two patterns will appear statistically distinct (see application 3). On the other hand, if two timing patterns have been shown to be distinct on the basis of few dates of observation, increasing the number of dates of observation will *never* reverse that conclusion, only strengthen it. This is a logical consequence of the way in which the test statistic is determined as the maximum difference between the two empirical distributions. More observation dates may reveal a different date at which the empirical distributions are even more different than they formerly maximally were, but more observation dates cannot reduce the magnitude of the largest difference formerly found.

With this in mind, if the biological theory to be tested predicts that two timing patterns will be the same, then it is best to make observations on many dates to provide ample opportunity for differences to be revealed and so disprove the theory. If the biological theory to be tested predicts that two timing patterns will be different, then only as many dates as are necessary to show this difference are required to confirm it. For biological purposes, "the same time" means "within the same time interval" where the length of the time interval has been chosen to reflect the biologically appropriate time scale. Temporal differences closer than this time interval could not be interpreted even if they were statistically different (see application 2).

When circumstances require observations on dates more widely spaced than is biologically desirable, the possibility exists that the data will fail to reveal a real difference. Statistical guidance for this situation would be a welcome contribution to the methods for studying phenology.

The statistical test discussed here serves *only* to indicate whether data are sufficient to support the claim that two timing patterns are different. When differences between two temporal patterns are indicated, it is desirable to interpret them. Although

these test procedures are almost certain to become invalid with temporal shifts, one way to interpret differences is to move in time one empirical distribution relative to the other until the maximum difference is minimized. If this new minimum maximum is not "significant," then it is reasonable to hypothesize that one timing pattern precedes the other timing pattern by the amount moved. If the new minimum maximum is still "significant," it is reasonable to hypothesize that the two patterns have different shapes (see applications 1 and 3). In either case, it is scientifically desirable to gather new data to test new hypotheses. DOKSUM and SIEVERS (1976)

and DOKSUM (1977) described rigorous statistical procedures that would be appropriate for testing the hypothesis that one phenological pattern is the same as another except for a constant temporal shift, although their work is not presented in terms of phenological pattern.

The general applicability of this method to the comparison of any two temporal patterns is clear. Needed are: (1) two groups to be compared, (2) from each group, at least 20 independent "events" that occur in time, and (3) a means of making observations to determine when these events took place.

## LITERATURE CITED

- BLACK, J. N., and G. N. WILKINSON. 1963. The role of time of emergence in determining the growth of individual plants in swards of subterranean clover. *Australian J. Agr. Res.* **14**:628-638.
- CONOVER, W. J. 1971. *Practical nonparametric statistics*. Wiley, New York. 462 + x pp.
- DOKSUM, K. A. 1977. Some graphical methods in statistics: a review and some extensions. *Statist. Neerlandica* **31**: 53-68.
- DOKSUM, K. A., and G. L. SIEVERS. 1976. Plotting with confidence: graphical comparisons of two populations. *Biometrika* **63**:421-434.
- FRANKIE, G. W. 1975. Tropical forest phenology and pollinator plant coevolution. Pages 192-209 in L. E. GILBERT and P. H. RAVEN, eds. *Coevolution of animals and plants*. University of Texas Press, Austin.
- HARPER, J. L. 1961. Approaches to the study of plant competition. *Soc. Exp. Biol. Symp.* **15**:1-39.
- HEITHAUS, E. R. 1974. The role of plant-pollinator interactions in determining community structure. *Ann. Missouri Bot. Garden* **61**:675-691.
- HOWE, H. F. 1977. Bird activity and seed dispersal of a tropical wet forest tree. *Ecology* **58**:539-550.
- HOWE, H. F., and G. F. ESTABROOK. 1977. On intraspecific competition for dispersers among tropical trees. *Amer. Natur.* **111**:817-832.
- HOWE, H. F., and G. A. VANDE KERCKHOVE. 1979. Fecundity and seed dispersal of a tropical tree. *Ecology* **60**:180-189.
- LECK, M. A. 1979. Germination behavior of *Impatiens capensis* Meerb. (Balsaminaceae). *Bartonia* **46**:1-14.
- OPLER, P. A., G. W. FRANKIE, and H. G. BAKER. 1976. Rainfall as a factor in the release, timing, and synchronization of anthesis by tropical trees and shrubs. *J. Biogeogr.* **3**:231-236.
- PULLIAM, H. R., and M. R. BRAND. 1975. The production and utilization of seeds in plains grassland of southeastern Arizona. *Ecology* **56**:1158-1166.
- ROSS, M. A., and J. L. HARPER. 1972. Occupation of biological space during seedling establishment. *J. Ecol.* **60**:77-88.
- SMIRNOV, N. V. 1939. Estimate of deviation between empirical distribution functions in two independent samples [in Russian]. *Bull. Moscow Univ.* **2**:3-16.
- SOKAL, R. R., and F. J. ROHLF. 1969. *Biometry*. W. H. Freeman, San Francisco.
- STEPHENSON, A. G. 1979. An evolutionary explanation of the floral display of *Catalpa speciosa* (Bignoniaceae). *Evolution* **33**:1200-1209.
- . 1980. Fruit set, herbivory, fruit reduction, and the fruiting strategy of *Catalpa speciosa* (Bignoniaceae). *Ecology* **61**:57-64.
- . 1981. Flower and fruit abortion: proximate causes and ultimate functions. *Annu. Rev. Ecol. Syst.* **12**:253-280.
- STILES, E. W. 1980. Patterns of fruit presentation and seed dispersal in bird disseminated woody plants in the eastern deciduous forest. *Amer. Natur.* **116**:670-688.
- THOMPSON, J. N., and M. F. WILLSON. 1979. Evolution of temperature fruit/bird interactions: phenological strategies. *Evolution* **33**:973-982.
- WINSOR, J. A. 1981. Germination and establishment in *Impatiens capensis* (Balsaminaceae): adaptations to interspecific competition. Ph.D. diss. University of Michigan, Ann Arbor.