Development of a computing model for resistance screening of *Citrus limon* cultivars infected by the causal agent of "Mal secco" *Phoma tracheiphila*

K. Khanchouch[1,2]**, M.R. Hajlaoui**[2]**, and H. Kutucu**[3]
[1]Department of techniques, ISAJC University of Tunis, Tunis, Tunisia
[2]Laboratory of plant protection, National research agronomic institute, Tunis, Tunisia
[3]Department of Mathematics, Izmir institute of technology, Ural-Izmir, Turkey

Abstract - *The mathematical survey of the different studied cultivars, using a polynomial model, permits to describe the resistance state of the infected plants. The polynomial interpolation at 5 degrees appears to be the most adequate for this mathematical model. Comparison of R^2 values showed that the polynomial regression at the 5 degree gives the best results. The statistical analysis confirm those obtained by the polynomial model. This polynomial model have the advantage to give a strict evaluation of the state of the resistance of the cultivar tested and not a relative estimation as its in the case of the different mathematical and statistical others classic tools usually used to evaluate the state of the plant resistance. The computing model is able to distinguish between the three resistances levels of Citrus limon cultivars tested.*

Keywords: Biomath, Modelling, *Citrus limon*, *Phoma tracheiphila* and Bioinformatic

1 Introduction

Plant diseases are responsible of 14.1% of the world crop loses which represent $220 billion of dollars. These phytopathological damages imply several others problems in different sectors concerning human health, the environment and some social and economic aspects of our life [1]. In order to have an efficient solution to control the causal agents of these diseases it is very important to understand the mechanism of these illnesses very well [2, 3].

In phytopathology, the Mathematical tools used offer models describing the process of the infection [4]. These mathematical models allow to describe the pathological processes and therefore to foresee the most efficient control methods. The mainly mathematical tools used to model the plant diseases are: Disease progress curves, Linked Differential Equation (LDE) and Area Under disease Progress Curve (AUDPC). Statistical analyses are also employed in the studies of epidemiology of plant diseases. Each tool is utilized for an acute appropriate purpose to model some aspects of the disease development.

The specificity of the host-parasite relationships determine the variables and the adequate mathematical model to be used. On the basis of these chosen mathematical tools the most Known model developed in the phytopathological studies are: Monomolecular, Exponential, Gompertz and Logistic models. The logistic model which was proposed firstly by Veshulst in 1838 to represent human population growth was after developed by Van der Plank (1963) [5], to being more appropriate for most polycyclic diseases. This growth model is the most widely used for describing epidemics of plant disease [3,6].

Using the logistic model alone or combined with others tools many plant diseases were described. In the case of the Citrus disease ''Mal secco'', there is no reports referring to the development of a model allowing to test the resistance degree of the susceptible infected host plants. The causal agent of the ''Mal secco'', Phoma tracheiphila (Petri) [Kanc et Ghik.], is responsible of many important losses in the Citrus crop orchards and it's the most destructive fungal disease of lemon plantation worldwide [2]. As fungicides treatments showed non efficient results to control this pathogen, the research of resisting cultivars remains the most efficient solution to decrease the losses inflicted by the pathogen [7].

2 Material and methods

2.1 Biological Material

Plants belong to *Citrus limon* cultivars and a highly virulent pathogenic isolate of the causal agent of Mal secco were used. The green house inoculation method used is its described by Hajloui et al (2000) [8]. 120 inoculation points in total are assessed per plant. A scale of six classes is used to evaluate the reaction of tested plants. The classes are numbered from 0 to 5.

2.2 The Mathematical Model

2.2.1 Polynomial interpellation

The cumulative percentage frequency of each class is calculated for all the tested plants. The calculation of the cumulative frequency is determined as described below:

$$\mathbf{Yi= [\sum_0^i \mathit{frequency\ of\ xi}\ )\ /120]*100}$$

Yi= The cumulative frequency at the respective class, Xi.

Xi= class ‘i’’ varying from “0” to “5”

120, it’s the number of the inoculation points tested.

The representative curve of the cumulative percentage frequency for the different tested plants it’s a polynomial. The degree of the polynomial is fixed referring to the theorem of Lagrange. This is approved using the data of the inoculation test and the calculation of the coefficient of determination “R^2”.

$$f(x_i) = a\, x_i^{\,5}+b\, x_i^{\,4}+c\, x_i^{\,3}+d\, x_i^{\,2}+e\, x_i^{\,1}+f$$

To determine the coefficients of the polynomial function a linear system of six equations is used:

$$\begin{cases} Y_0 = a\, x_0^{\,5}+b\, x_0^{\,4}+c\, x_0^{3}+d\, x_0^{2}+e\, x_0^{1}+f \\ Y_1= a\, x_1^{\,5}+b\, x_1^{\,4}+c\, x_1^{3}+d\, x_1^{2}+e\, x_1^{1}+f \\ Y_2= a\, x_2^{\,5}+b\, x_2^{\,4}+c\, x_2^{3}+d\, x_2^{2}+e\, x_2^{1}+f \\ Y_3= a\, x_3^{\,5}+b\, x_3^{\,4}+c\, x_3^{3}+d\, x_3^{2}+e\, x_3^{1}+f \\ Y_4= a\, x_4^{5}+b\, x_4^{\,4}+c\, x_4^{3}+d\, x_4^{2}+e\, x_4^{1}+f \\ Y_5= a\, x_5^{5}+b\, x_5^{\,4}+c\, x_5^{3}+d\, x_5^{2}+e\, x_5^{1}+f \end{cases}$$

To calculate the coefficients a, b, c, d, e and f, we use Gaussian elimination method [9]. After the coefficients are determined, the polynomial is used to calculate the area under the curve of the infected plant by integrating f(x) from 0 to 5:

$$\int_0^5 \mathrm{f(x)} = \left[a\frac{1}{6}x^6 + \mathrm{b}\ \frac{1}{5}x^5 + \ \mathrm{c}\ \frac{1}{4}x^4 + \ d\frac{1}{3}x^3 + \mathrm{e}\ \frac{1}{2}x^2 + \mathrm{f}\, x^1\right]$$

2.2.2 Statistical analysis

For data analysis of the artificial inoculated plants, the biosoftware Statisica version 5.1 was used. Newman and keul-Keuls test at p=0,05 of Anova order 1 was performed.

2.2.3 Computing the plant resistance level

Mathematical description of the resistance level of the infected plants is based on the characteristics of their polynomial regression curve. Three types of polynomial curve can be described:

Type A: with an upper concave convection
Type B: with a mixed convection curve
Type C: with a lower concave convection

The parametric analysis of each polynomial curve by calculating its derivative near the convection points, allows distinguishing between the three types.

$$f'(x_i) = 5a\, x_i^{4}+4b\, x_i^{3}+3c\, x_i^{2}+2d\, x_i+e$$

The derivative calculation for each polynomial regression curve is performed from the point x_i =1 to the point x_i =4. A linear regression curve to fit the calculated values of the derivative is determined using these formulations:

$$y_i' = ax_i+b$$

$$a= \sum_1^i (\boldsymbol{xi} - \ \bar{\mathrm{x}})(\boldsymbol{y'i} - \overline{\boldsymbol{y'}}) / \ \sum_1^i (\boldsymbol{xi} - \ \bar{\mathrm{x}})^2$$

$$b= \overline{\boldsymbol{y'}} - a\bar{\mathrm{x}}$$

with :
$\bar{\mathrm{x}}$= means of xi
$\overline{\boldsymbol{y'}}$= means of y'_i

The adjustment of the fitted linear derivative regression curve is appreciated using its R'^2 value. Higher value of R'^2 characterizes both the type A and C while the type B is determined by its R'^2 as equal or less than 0,5. The coefficient ‘‘a’’ is positive for type A and negative for type C.

2.3 Algorithm building

In order to determine the coefficients a, b, c, d, e and f, we use Gaussian elimination method:

INPUT : A(n, n), b(n)
OUTPUT: x(n) as the solution

```
for k=1 to n-1
    for i = k+1 to n
        factor = A(i , k)/A(k , k)
        for j=k+1 to n
            A(i , j) = A(i , j) – factor*A(k ,j)
        end
        b(i)=b(i)-factor*b(k)
        end
end
x(n)=b(n)/A(n,n)
```

```
for i=n-1 to 1
    sum=0
  for j=i+1 to n
    sum=sum+a(i,j)*x(j)
  end
  x(i)=(b(i)-sum)/a(i,i)
end
```

After the coefficients are determined, the polynomial is used to calculate the area under the curve of the infected plant by integrating f(x), from 0 to 5. We use Simpson's method to compute the area under the regression polynomial curve:

```
INPUT : p(x), a (lower limit of integration), b (upper limit of
integration), n (the number of subintervals to divide interval
[a,b], n must be divisible by 3)
OUTPUT: Integral of p(x) from a to b

SECTIONS = n/3
h = (b-a) / n
APPROX = 0
for i=1 to SECTIONS:
  x0 = a + 3 * (i–1) * h
  x1 = x0 + h
  x2 = x1 + h
  x3 = x2 + h
  APPROX = APPROX + p(x0) + 3*p(x1) + 3*p(x2) + p(x3)
  end
INTEGRAL = 3 * h/8 * APPROX
```

To discriminate between the three types of the regression polynomial curves determined by the mathematical model we elaborated this Algorithm:

```
INPUT : p(x)= $ax^5+bx^4+cx^3+dx^2+ex+f$
OUTPUT: Fitted Line (y= $a'x+b'$), $R^2$ and Decision

Step 1: for i=1 to 4
     $x_i$=i
  end

Step 3: Find the derivative of p(x)
        $p'(x) = 5ax^4+4bx^3+3cx^2+2dx+e$

Step 4: for i=1 to 4
     $y_i = p'(i)$
  end

Step 5: Find $\bar{x}$ (the average of $x_i$)
          for i=1 to 4
        sum=sum+$x_i$
          end

          $\bar{x}$ =sum/4

Step 6: Find $\bar{y}$ (the average of $y_i$)
      for i=1 to 4
        sum=sum+$y_i$
          end

      $\bar{y}$ =sum/4
Step 7: Find $a'$ (the coefficient of x for the fitted line)
     for i=1 to 4

          sum1=sum1+$(x_i-\bar{x})(y_i-\bar{y})$

              sum2=sum2+$(x_i-\bar{x})^2$
            end
     $a'$=sum1/sum2

Step 8: Find $b'$ (the constant term of the fitted line)
  $b' = \bar{y} - a\bar{x}$


Step 9: Find $\hat{y}_i$ for all i=1,…,4
            for i=1 to 4
          $\hat{y}_i = a' \cdot i + b'$
           end

Step 10: Find $R^2$
     for i=1 to 4
          sum1=sum1+$(\hat{y}_i-\bar{y})^2$

              sum2=sum2+$(y_i-\bar{y})^2$
            end
     $R^2$=sum1/sum2

Step 11: If ($R^2$ <= 0.5) Then
                 the tested plant is classified as tolerant

    Else If (a>0) Then
      the tested plant is classified as sensible

         Else If (a<0) Then
                 the tested plant is classified as resistance

         End If
```

2.4 Results and discussion

The greenhouse artificial inoculation of the tested lemon cultivars was surveyed on the basis of severity of the disease symptoms following foliar inoculation. Three Citrus cultivars; Eureka, Interdonato and Monachello were used in this test.

All the tested plants developed the pathological symptoms caused by the parasite after the incubation period. Cultivars Eureka and Monachello expressed, relatively, the highest and the reduced degree of the disease index. However, Interdonato, showed an intermediate index of the disease severity. Statistical analysis indicate a significant differences between the three inoculated cultivars at p=0,05. According to this analysis, the tested plants were ranked in three different groups of resistance (Table 1).

Table 1: Infection severity rating of lemon cultivars.

Cultivars	Severity of the disease	Ranked group
Eureka	4,216	I
Interdonato	2,225	II
Monachello	0,8	III

I= sensible, II= Tolerant and III= Resistant

Calculation of the coefficients of the polynomial regression curves by the biomathematical model for the three resistance groups is resumed in the table 2.

Table 2: Parameters of the fitted polynomial curves

Cultivars	Regression curve	area
Eureka	$y = 0{,}187x^5 - 1{,}736x^4 + 6{,}701x^3 - 12{,}43x^2 + 13{,}944x + 2{,}5$	119,17
Interdonato	$y = 0{,}263x^5 - 2{,}534x^4 + 7{,}638x^3 - 5{,}798x^2 + 0{,}430x + 47{,}5$	297,85
Monachello	$y = 0{,}208x^5 - 3{,}159x^4 + 18{,}611x^3 - 54{,}340x^2 + 82{,}847x + 41{,}666$	455,43

The polynomial's coefficients determined by the bio-mathematical model were in concordance with its obtained by Mathematica uses Hermite interpolation technique to find fitting curves to a given sets of data [10].

Elevated value of R^2, determined by the model for all the groups of resistance, indicates the strength of fit between the polynomial regression curve and the percentage of the cumulative frequency. The area under the curve (AUC) calculated for the group I was the lowest. The group II and III were characterized but their respective increased value of the AUC.

Drawing the representative fitted curve, it was also found that the type A of the polynomial is correlated with the group I of resistance. While the type B was attributed to the group II of resistance. The type C is associated with the group III of resistant plant (Figure 1).

The algorithm build recognize the different studied groups and attribute to the tested plant its level of resistance. The tests performed on *citrus limon* cultivars using the biomathematical model are in perfect concordance with those obtained by the usually statistical tools.

The mathematical model is able to evaluate the resistance of the infected plant without using comparative methods. The model describes the repartition of the different classes of the disease to evaluate more precisely the reaction of the host infected by the parasite.

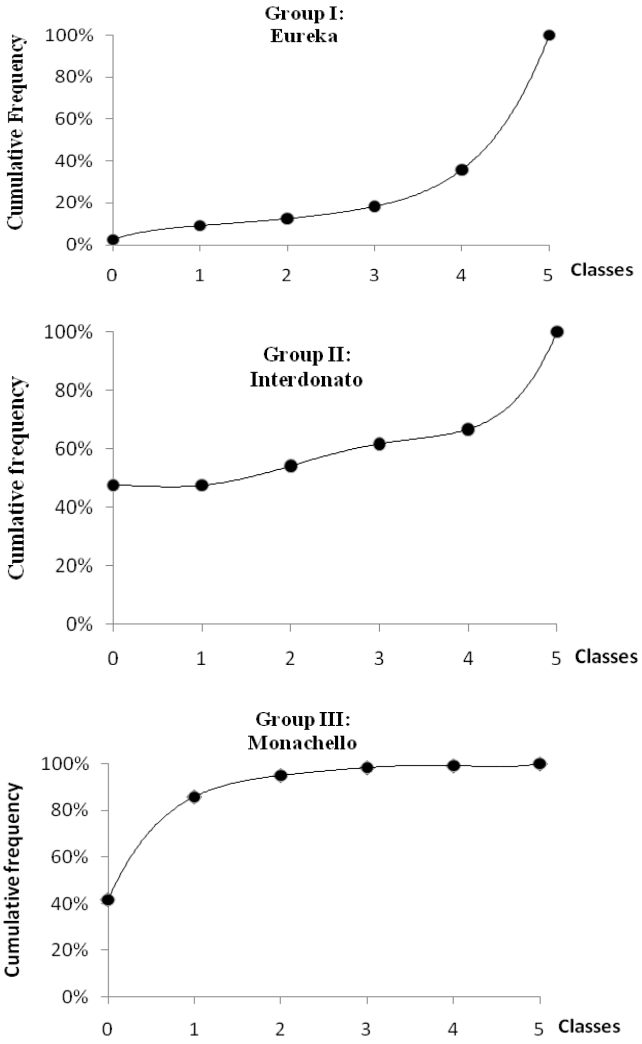

Figure 1: The polynomial regression fitted curves

The response of the studied cultivars, as it's recorded and analyzed by the mathematical model reflects the natural behavior of the tested plant in the orchards.

Taking into account the obtained results, the model proved to be an efficient new method for the resistance screening in the host parasite biological interaction system of *citrus limon* cultivars and their pathogenic fungi *Phoma tracheiphila*.

3 Conclusions

The factors most likely to influence the results of the phyopathological tests are mainly the techniques of the inoculation and the methods of the analysis used. As the procedures vary between the laboratories the results also differ even for the same tested biological sample. In order to overpass these difficulties and standardize the protocols of the disease assessment, the biomathematical model offer an appropriate solution.

Extending the results of this study to others groups of resistance in Citrus plants its needed to cover the whole specter of the host plants of the parasite, from the very susceptible one to the highly resistance.

4 References

[1] Agrios GN (2005). Plant Pathology. Fifth Edition, Elsevier Academic Press, London, UK. Royle DJ, Ostry ME (1995). Disease and pest control in the bioenergy crops poplar and willow. Biomass Bioenergy, 9: 69-79.

[2] Gulsen O., Uzun A., Pala, E H.. Canihos and Kafa G.. 2007. Development of seedless and Mal Secco tolerant mutant lemons through bud wood irradiation. Scientia Horticulturae Volume 112, Issue 2, Pages 184-190.

[3] Jeger MJ (2004). Analysis of disease progress as a basis for evaluating disease management practices. Annu. Rev. Phytopatol. 42: 61-82.

[4] Van Maanen A, Xu XM (2003). Modelling plant disease epidemics. European J. Plant Pathol. 109: 669-682.

[5] Van der Plank JE (1963). Plant disease: epidemics and control. Academic Press, NY., USA.eslie Lamport. “LaTeX: A Document Preparation System”. Addison-Wesley Publishing Company, 1986.

[6] Segarra J, Jeger MJ, Van den Bosch F (2001). Epidemic dynamics and patterns of plant diseases. Phytopathology, 91: 1001-1010.

[7] Slolel, Z., and Salerno, M. 1988. Mal secco. Pages 18-20 in: compendium of citrus Diseases. J.O. Whiteside, S. M. Garnsey, and L.W. Timmer, eds. American Phyotpathological Society, St. Paul, MN.

[8] Hajlaoui M.R., Kanchouch K., Guermech A. et Cherif M, 2000 a. Recherche sur l’agent du mal secco des Citrus, Phoma tracheiphila : I- Etude de quelques caractéristiques biologiques du parasite. Annales de l’Institut Nationale de la Recherche Agronomique de Tunisie. Vol 73: 163-182.

[9] Gareth Williams, Linear Algebra with Applications, Seventh Edition, Jones & Bartlett Publishers, 2009, 554 pages.

[10] Stephen Wolfram, The MATHEMATICA ® Book, Version 4, Cambridge University Press, 1999, 1469 pages.