MODELO COMPUTACIONAL PARA LA ESTIMACIÓN DEL TAMAÑO DE FRUTAS

F. J. Mayorano[1], A. J. Rubiales[1], V. Herrero[1,2], F. Núñez[4], A. Clausse[1,3]
[1]*PLADEMA, Universidad Nacional del Centro*, 7000 Tandil, Argentina.
[2]También Ejército Argentino.
[3]También CNEA y CONICET.
[4]Expofrut, General Roca, Río Negro, Argentina.
{fmayoran, arubiale, vherrero, clausse} @exa.unicen.edu.ar ; Tel./Fax: 02293-442202
fnunez@expofrut.com.ar ; Tel./Fax: 02941-439600

Palabras Claves: Simulación Computacional, Procesos Estocásticos, Pronóstico de Fruta, Método de Montecarlo.

RESUMEN

Se presenta un simulador del crecimiento de fruta para soporte de decisiones comerciales y logísticas. El objetivo principal del simulador es estimar el tamaño que presentará la fruta en tiempo de cosecha, en base a pronósticos sobre muestreos estadísticos durante las etapas de crecimiento. Se desarrolló un modelo de procesos estocásticos, utilizando las correlaciones históricas entre el crecimiento de la fruta y la temperatura. Se incluye la implementación del simulador en peras y manzanas de la zona del valle de Río Negro.

1) Introducción

El conocimiento de la distribución del tamaño de fruta es de crucial importancia para la industria frutícola ya que conocer este dato anticipadamente aumenta los beneficios de la empresa. Existen inevitables *lead-times* cuando una compañía compra los materiales necesarios para el procesamiento de la fruta (por ejemplo, el material para empaque). Por esta razón, la materia prima debe ser adquirida con anticipación. Qué y cuánto comprar depende de las características de la fruta a cosechar. Comprar por demás implica pérdidas en inventario de materia prima obsoleta; mientras que comprar de menos produce rupturas de inventarios con sobrecostos y retrasos en el procesamiento de la fruta.

Además, la calidad de la cadena comercial esta afectada por la medida en la que la producción de fruta se ajusta el mercado estándar. Conocer anticipadamente las características de la producción ayuda a conducir los esfuerzos comerciales para adaptarse a la combinación de tamaños que el mercado demanda.

Existen en la actualidad algunos estudios realizados para el modelado del crecimiento de la fruta. Entre ellos se puede citar un enfoque en el cuál se utilizan ecuaciones diferenciales ordinarias, y otro en el cuál se obtienen resultados a través de modelos estocásticos.

En el primer enfoque la tasa de cambio de ciertas propiedades de la fruta se expresa como función del estado de la fruta y factores ambientales determinantes. Estos son construidos observando los datos y planteando ecuaciones para los promedios que aproximan los datos observados. Se suelen ajustar curvas sigmoides de crecimiento incluyendo parámetros dependientes de variables climáticas (*e.g.,* temperatura media) [1].

Por otra parte, si lo que interesa conocer es la distribución de tamaños, no basta con modelar los promedios. En estos casos se suele introducir el concepto de ‘ruido’, describiendo el crecimiento de la fruta con ecuaciones diferenciales estocásticas, cuya solución es la evolución temporal de la distribución de tamaños [2].

En este trabajo se presenta un nuevo enfoque para la simulación del crecimiento de frutas, basado en la teoría de procesos de Markov. El modelo presentado permite predecir la distribución final del tamaño de la fruta, basándose en muestreos estadísticos durante la etapa de crecimiento.

2) Descripción del modelo matemático

Las cadenas de Markov [3] son una clase de procesos estocásticos de tiempo discreto, representados por una variable aleatoria que toma valores desde un conjunto finito, llamados el espacio de estados del sistema. La estadística de una cadena de Markov está completamente determinada por la distribución de probabilidad inicial (por ejemplo, V_i es la probabilidad que el sistema este en el estado i) [5]. Dada la distribución de probabilidad al tiempo t, $\{V_1(t), V_2(t), \ldots, V_N(t)\}$, la distribución de probabilidad al tiempo $t + 1$ esta dada por:

$$V_i(t+1) = \sum_{i=1}^{N} p_{ij}(t) V_j(t) \qquad (1)$$

donde $p_{ij}(t)$ representa la probabilidad condicional de que el proceso se encuentre en el estado i en el tiempo $t+1$, dado que se encontraba en el estado j en el tiempo t.

Una manera práctica de visualizar los procesos de Markov es a través de grafos de transición.

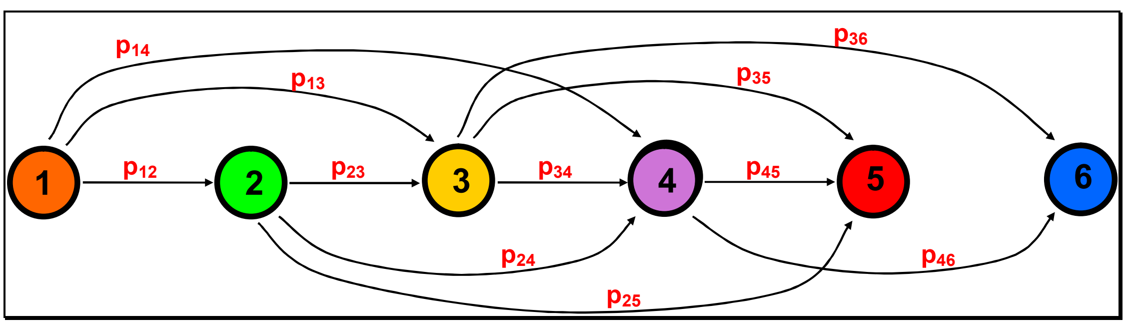

Figura 1: Grafo de transición de Estados

En la Fig. 1 se puede observar un grafo representando el crecimiento de la fruta durante una temporada. Cada círculo representa un rango de tamaño de frutas (por ejemplo, frutas cuyo diámetro se encuentra entre 70 y 74 mm). Las flechas entre círculos indican las transiciones en el crecimiento, las cuales tienen sus respectivas probabilidades, $p_{ij}(t)$, que dependen en principio de los factores climáticos. V_i en este caso representa la cantidad de fruta que se encuentra en el estado i. Otra manera de representar el grafo de la Fig. 1, es a través de una matriz de pasaje, en donde cada uno de los valores de la matriz, representa las diferentes flechas del grafo:

$$\begin{bmatrix} 0 & p_{12} & p_{13} & p_{14} & 0 & 0 \\ 0 & 0 & p_{23} & p_{24} & p_{25} & 0 \\ 0 & 0 & 0 & p_{34} & p_{35} & p_{36} \\ 0 & 0 & 0 & 0 & p_{45} & p_{46} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

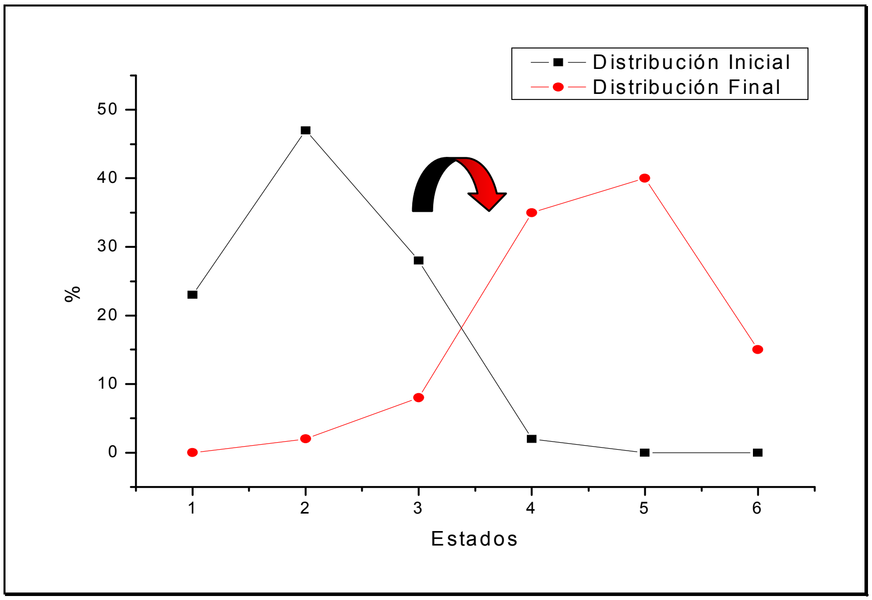

Figura 2: Evolución de las Distribuciones

En la Fig. 2 se muestra un ejemplo de el cambio de la distribución de tamaño de frutas de una cierta muestra medida en dos tiempos distintos. Los estados representan rangos crecientes de diámetro. En principio, la distribución final puede ser calculada aplicando la matriz de pasaje a la distribución inicial de acuerdo a la Ec. (1). Para simular el crecimiento de la fruta usando la Ec. (1), es necesario encontrar un conjunto apropiado de probabilidades de transición, que representen adecuadamente los procesos biológicos involucrados.

Entre los diferentes factores que afectan el crecimiento de la fruta, la temperatura es el más influyente. Las variaciones del tamaño de la fruta entre diferentes temporadas y lugares pueden ser comprendidas si definimos el mismo como función de la acumulada de la temperatura [1]. Este efecto esta relacionado con la activación de la multiplicación y crecimiento de las células debido a la energía térmica depositada en la fruta.

Otro factor que es importante en las regiones frías (como la del valle de Río Negro) son las heladas y periodos de bajas temperaturas. Por ejemplo, el crecimiento de la manzana se detiene a temperaturas menores que 7º C.

Los factores mencionados influyen de diferente manera a lo largo del proceso de crecimiento de la fruta. En experimentos dentro de ambientes controlados sobre manzanas, por ejemplo, bajas temperaturas en los primeros 40 días tienen un gran efecto en el tamaño y madurez de la fruta, mientras que condiciones similares entre los 40 y 70 días después de floración no tienen efecto importante sobre la manzana.

En base a estas consideraciones, es razonable plantear una dependencia de las probabilidades del crecimiento, p_{ij}, con la temperatura media total acumulada y con la acumulada menor a 7º C. De la observación de las tendencias históricas se concluyó que un modelo lineal era suficiente para obtener una buena descripción en manzanas y peras creciendo en el Valle del Río Negro, esto es:

$$p_{ij}(t) = a_{ij}(t)\overline{\Sigma}(t) + b_{ij}(t)\Sigma_7(t) + c_{ij}(t) \tag{2}$$

donde:

$$\overline{}(\boldsymbol{t}) \qquad \overset{\boldsymbol{t}}{\underset{\boldsymbol{t'}\ \ \boldsymbol{0}}{}} \boldsymbol{\overline{T}(t')} \tag{3}$$

es la temperatura media diaria acumulada, y:

$$\Sigma_7(t) = \sum_{t'=0}^{t} \min(0, T_{\min}(t') - 7^o C) \tag{4}$$

es la acumulada de los subenfriamientos diarios por debajo de los 7º C.

Los parámetros $a_{ij}(t)$, $b_{ij}(t)$ y $c_{ij}(t)$ dependerán en general de la localización geográfica, del suelo y de las prácticas de agricultura que se realicen sobre la fruta. Se considerará que pueden ser estimados utilizando las correlaciones históricas entre el crecimiento de la fruta y la temperatura. En principio, cuanto más información histórica se encuentre disponible, más confiable y más exacto será el modelo. Desde esta perspectiva, la Ec. (2) debe ser vista como una herramienta de soporte para estimar el tamaño de la fruta, la cuál debe ser complementada con un consistente monitoreo de la fruta y de las condiciones climáticas.

3) Pasos a seguir para la instanciación del modelo

Llamamos instanciación del modelo a la determinación de las matrices de parámetros, $a_{ij}(t)$, $b_{ij}(t)$ y $c_{ij}(t)$, en base a mediciones de tamaño de frutas. El proceso esta dividido en los siguientes pasos:

a) Determinar fechas de medición y estados

Las fechas de medición se definen en base a cuándo es importante tener una estimación de las distribuciones del calibre o diámetro de la fruta en tiempo de cosecha, las cuales están asociadas a los planes comerciales. Por otra parte, los estados (rangos de tamaño de fruta) deben definirse de manera tal que no queden todas las mediciones en un mismo estado; ni tampoco demasiados estados con pocas mediciones cada uno. Los estados que estén mayormente asociados a mediciones en tiempo de cosecha se definen en base a los tamaños comerciales. Se incrementa el detalle (por lo tanto, se define mayor cantidad de estados con rangos mas pequeños) en aquellos estados donde se encuentre concentrada la mayor parte de la demanda.

b) Generar distribuciones de acuerdo a los estados definidos en el punto anterior

Tomando los datos que se poseen sobre los diferentes calibres, se calculan las distribuciones de probabilidad correspondientes a cada momento de la temporada en los que se realizan mediciones. Este trabajo se debe realizar no solo para las diferentes etapas de crecimiento de la fruta, sino también en tiempo de cosecha.

c) Obtener $\overline{\Sigma}(t)$ y $\Sigma_7(t)$

De acuerdo a las temperaturas medias y mínimas registradas se calculan los parámetros $\overline{\Sigma}(t)$ y $\Sigma_7(t)$. Para el cálculo de ambos parámetros se toman los valores de las temperaturas desde la fecha de floración hasta el momento en que se realiza la medición.

d) Pasos a seguir para la estimación de una temporada:

Una vez completados los pasos anteriores se está en condición de realizar una estimación para una especie, variedad, región y temporada en particular. Los pasos a seguir son:

I. Seleccionar temporadas para realizar la estimación

Es importante que los datos que se utilicen para definir las matrices de crecimiento sean de temporadas similares a la temporada que se quiere estimar. Para poder definir que dos temporadas son similares debemos definir un criterio que nos permita establecer esa similitud. Se definen como características principales o relevantes de cada una de las temporadas los siguientes 3 aspectos:

- $\overline{\Sigma}(t)$, ya definida en la ecuación (3).
- $\Sigma_7(t)$, ya definida en la ecuación (4).
- La distribución a partir de la cual se quiere estimar.

Con las temporadas de las cuales se posean datos se debe realizar una selección, y en base a los criterios antes mencionados, elegir las temporadas que más se adaptan a la temporada que se desea estimar.

II. Generar coeficientes

Se ajustan las matrices de coeficientes para que estos, al aplicarlos a la función de transición antes definida (2), generen valores lo más similares posibles a los que se obtuvieron en las temporadas seleccionadas. El algoritmo utilizado para tal objetivo es el método de Monte Carlo, el cuál realiza una búsqueda aleatoria modificando los elementos de las matrices hasta reducir el error cuadrático entre lo medido y lo calculado [4].

Pseudo-Código del algoritmo:

```
Generacion_Coeficientes (error, #temporadas, #estados, mA, mB, mC,
                         mDDF,  mCosecha, T1, T2,  delta);

begin
```

```
   error_Actual = ∞;                                                             (a)
   while (error_Actual > error) do                                               (b)
   begin
      matriz_Seleccionada = Seleccionar_Matriz(0, 2);                            (c)
      indice_i = Seleccionar_Indice(0, #estados -1);                             (d)
      indice_j = Seleccionar_Indice(0, #estados -1);                             (e)
      incremento = Seleccionar_Incremento(0,1);                                  (f)
      if incremento = 0 then                                                     (g)
         signo = -1;
      else
         signo = 1;
      matriz_Seleccionada[indice_i, indice_j] =
          matriz_Seleccionada [indice_i, indice_j] + delta*signo;                (h)
      error_Anterior = error_Actual;
      mPasaje = Normalizar_Matrices(mA, mB, mC, T1, T2);                         (i)
      error_Actual =
          Calcular_Error(#temporadas, #estados, mDDF, mCosecha, mPasaje);        (j)
      if error_Actual  > error_Anterior then
         matriz_Seleccionada[indice_i, indice_j] =
          matriz_Seleccionada [indice_i, indice_j] + delta*signo;                (k)
   end;
end;
```

Los parámetros de entrada del algoritmo son:

- Error : El error máximo que se espera para la matriz de transición.
- #temporadas : Cantidad de temporadas utilizadas en la estimación.
- #estados : Cantidad de estados o rangos entre los cuales se separo la distribución de tamaño de la fruta.
- mA : Matriz de coeficientes que en la ecuación acompaña al termino $\overline{\Sigma}(t)$, asociado a la temperatura media.
- mB : Matriz de coeficientes que en la ecuación acompaña al termino $\Sigma_7(t)$, asociado a las temperaturas mínimas.
- mC : Matriz de coeficientes que en la ecuación representa al término independiente.
- mDDF : Matriz en donde se encuentran definidas los distribuciones en cierto cantidad de días despues de floracion para las temporadas seleccionadas.
- mCosecha : Matriz en donde se encuentran definidas los distribuciones en cosecha reales para las temporadas seleccionadas.
- T1 : Arreglo que contiene todos los valores de $\overline{\Sigma}(t)$ correspondientes a las temporadas utilizadas en el algoritmo.
- T2 : Arreglo que contiene todos los valores de $\Sigma_7(t)$ correspondientes a las temporadas utilizadas en el algoritmo.
- Delta : Valor absoluto del incremento o decremento que se aplica en cada iteración.
- mPasaje : Matriz de tres dimensiones en la que se almacena la matriz de pasaje para cada una de las temporadas seleccionadas.

Y las referencias que se muestran en el Pseudo-Código anterior significan lo siguiente:

(a) Se define el error actual con un valor inicial máximo, el cual se va a ir disminuyendo sucesivamente con el correr de las iteraciones.
(b) Define que el algoritmo iterará hasta que el error calculado para la función de transición definida sea menor a un error definido externamente.
(c) Elije aleatoriamente una matriz de coeficientes (a, b, o c) sobre la cuál se modificará uno de sus valores.
(d) Define la fila de la matriz de coeficientes en la cual se encuentra el valor a modificar.
(e) Define la columna de la matriz de coeficientes en la cual se encuentra el valor a modificar.
(f) Define si el valor del coeficiente se incrementara o se decrementará.
(g) En base a lo definido en el punto anterior, genera el valor que se le sumará al coeficiente.
(h) A la matriz de coeficientes, y en la posición seleccionada aleatoriamente se le suma el incremento definido en el punto anterior.
(i) Se calcula la matriz de pasaje de acuerdo a las matrices de coeficientes y los parámetros $\overline{\Sigma}(t)$ y $\Sigma_7(t)$, para cada una de las temporadas seleccionadas. Además, normaliza cada matriz de manera tal que la suma de probabilidades de las columnas sea igual a 1. Esta normalización se debe realizar para que el sistema cumpla la propiedad de conservación.
(j) Se calcula el nuevo valor para el error teniendo en cuenta las matrices de pasajes generadas para cada temporada.
(k) La matriz de coeficientes seleccionada, vuelve a tomar los valores que tenía antes de ser modificada. Cabe destacar que esta acción se realiza solo en el caso de que el error general no sea mejorado.

El valor del Error se corresponde con la suma de los errores cuadráticos existentes entre las distribución a cosecha dada por los coeficientes definidos (y calculadas con los valores de $\overline{\Sigma}(t)$ y $\Sigma_7(t)$ correspondiente a cada una de las temporadas) y la verdadera distribución de tamaños de esa temporada que se genera en base a los datos recolectados para la misma.

Es decir:

$$= \sum_k \sum_i \left\{ r_{k,i} - \sum_j \left[ddf_{k,j} * (\overline{\Sigma}_k . a_{i,j} + \Sigma_{7k} . b_{i,j} + c_{i,j}) \right]^2 \right\} \quad (5)$$

donde a, b y c son matrices de coeficientes; $\overline{\Sigma}_k$ y Σ_{7k} son los valores de $\overline{\Sigma}(t)$ y $\Sigma_7(t)$ correspondientes a la temporada k, $r_{k,i}$ es el valor real correspondiente a la probabilidad del estado i en la distribución de cosecha para la temporada k, y $ddf_{k,j}$ es el valor correspondiente a la probabilidad del estado j en la distribución para una cantidad dada de días después de floración para la temporada k.

Pseudo-Código del algoritmo para el cálculo del error:

```
Calcular_Error(#temporadas, #estados, mDDF, mCosecha, mPasaje);

begin
   error_Acumulado = 0;                                                        (a)
   for indice_Temporada = 0 to #temporadas - 1 do                              (b)
      for indice_i = 0 to #estados - 1 do                                      (c)
      begin
         prob = 0;                                                             (d)
         for indice_j = 0 to #estados - 1 do                                   (e)
         begin
              prob = prob +  (mPasaje[indice_Temporada,indice_i,indice_j]*
                             mDDF[indice_Temporada, indice_j]);                (f)
         end;
         error_Acumulado= error_Acumulado +
                (prob - mReal[indice_Temporada, indice_i])² ;                  (g)
      end;
   return error_Acumulado;                                                     (h)
end;
```

Donde los parámetros de entrada del algoritmo se definen de la misma manera que los del algoritmo anterior.

Y las referencias que se muestran en el Pseudo-Código anterior significan lo siguiente:

(a) Se define el acumulador del error, el cual se va incrementando a medida que voy comparando el resultado calculado con el real para cada temporada.
(b) Define que se debe calcular dicho error para cada una de las temporadas.
(c) Define un índice para iterar sobre las filas de las matrices.
(d) Define un acumulador que indica cual es la distribución calculada para el estado seleccionado.
(e) Define un índice para iterar sobre las columnas de las matrices.
(f) En base a la matriz de pasaje y la distribución en cierta cantidad de días después de floración, se calcula la distribución en cosecha.
(g) Incrementa el acumulador del error de acuerdo a lo especificado en la ecuación (5).
(h) Regresa el error de la comparación entre las distribuciones reales y las calculadas.

III. Obtener matriz de pasaje

Una vez que se obtienen los coeficientes, y los parámetros $\overline{\Sigma}(t)$ y $\Sigma_7(t)$ para la temporada que se desea estimar, se obtiene las probabilidades $p_{ij}(t)$ aplicando la ecuación (2).

IV. Obtener la estimación de los tamaños de la fruta en tiempo de cosecha

Partiendo de las matrices (probabilidades de transición) calculadas en el paso anterior se aplica la Ec. (1) a la distribución que se posee para esa temporada y se obtiene la distribución en cosecha. Cabe destacar que el tiempo al que pertenece la distribución debe ser el mismo que el utilizado para el cálculo de los coeficientes.

4) Aplicación del modelo en el Valle de Río Negro

El modelo antes descrito se aplicó a un caso real, en la zona del Valle de Río Negro, Argentina. Se usaron mediciones del tamaño de la fruta a 40 y 80 días después de floración y en cosecha, para la variedad Granny Smith (manzanas) en la zona del Alto Valle. Cabe destacar que a medida que la temporada avanza, el nivel de certidumbre de los datos inferidos es mayor. Por lo tanto, la distribución que se genera utilizando los datos en 80 días después de floración, tiene una precisión mayor que cuando se utilizan los datos en 40 días después de floración.

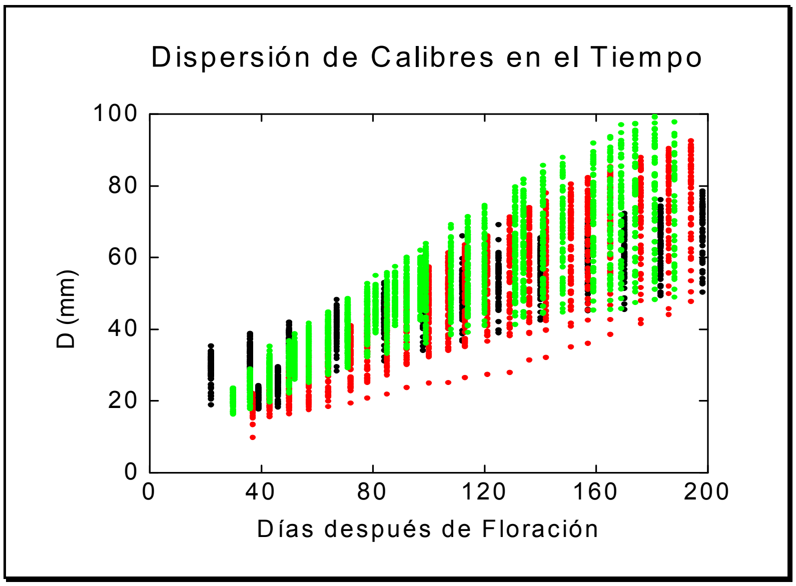

Figura 3: Dispersión de los calibres en el tiempo

Los rangos de tamaño se seleccionaron teniendo en cuenta los estándares de envases y consideraciones estadísticas.

Estados	Diámetro (mm)
1	menor que 35
2	35-45
3	45-55
4	55-60
5	60-65
6	65-70
7	70-73
8	73-76
9	76-80
10	Mayor que 80

Tabla 1: Estados definidos

La dispersión de los datos obtenidos del tamaño de la fruta durante las diferentes temporadas se puede visualizar en la Fig. 3, mientras que los estados que se consideraron necesarios para el caso particular se encuentran en la Tabla 1.

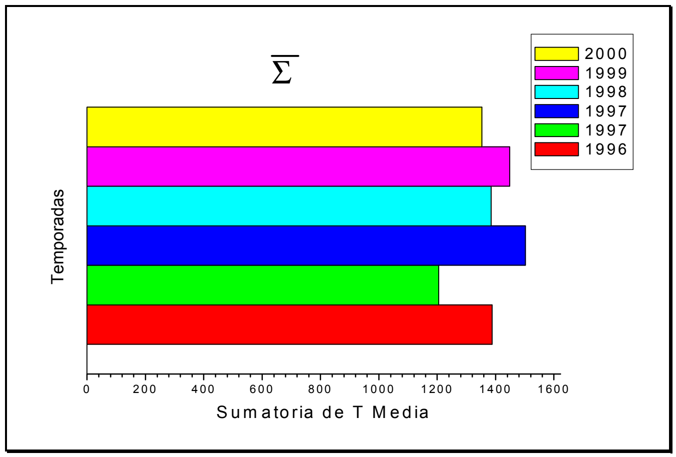

Figura 4: Cálculo del $\overline{\Sigma}(80\ ddf)$ para las diferentes temporadas

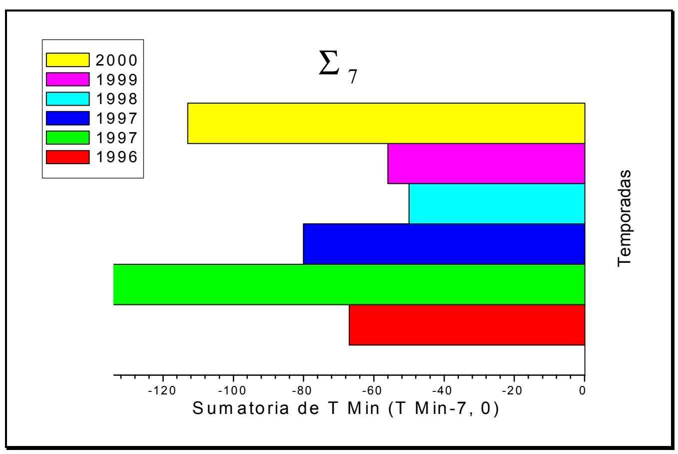

Figura 5: Cálculo del $\Sigma_7(80\ ddf)$ para las diferentes temporadas

Los valores de $\overline{\Sigma}(80\ ddf)$ y $\Sigma_7(80\ ddf)$ obtenidos para las diferentes temporadas son presentados en la Figs. 4 y 5 respectivamente

Para el caso de estudio, se estimó la temporada 1996. Al observar los parámetros de entrada, se concluye que las temporadas más significativas para realizar la generación de coeficientes son la temporada 1997, 1998 y la 1999, por lo tanto, dichas temporadas son las que se consideraron para generar las matrices de coeficientes. En la Fig. 6 se muestran los resultados obtenidos.

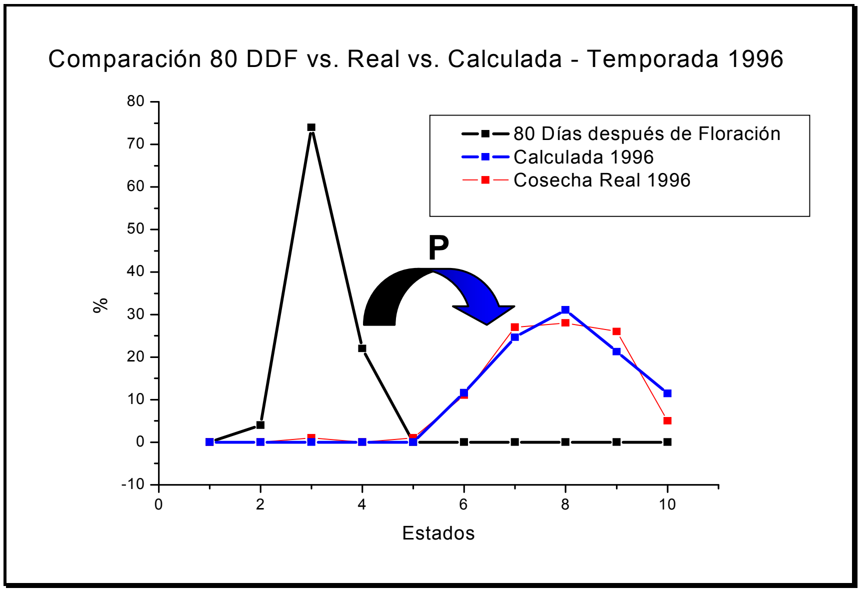

Figura 6: Resultado obtenido de aplicar el Modelo de Estimación

Se puede observar la distribución de tamaño a 80 Días después de Floración para la temporada 1996. Al aplicarle la matriz de pasaje P, ajustada en base a los parámetros $\overline{\Sigma}(80\ ddf)$ y $\Sigma_7(80\ ddf)$ para la temporada 1996, se obtiene una estimación de la distribución en cosecha para esa temporada, que, como se puede observar, es muy similar a la que se registro realmente ese año.

5) Conclusiones

Se desarrolló un modelo matemático para simular el comportamiento de la fruta de pepita durante su etapa de crecimiento. El mismo fue validado a través de un gran número de pruebas y se obtuvieron resultados aceptables. El principal aporte que realiza este modelo es la posibilidad de usar como entrada datos obtenidos a través de muestreo. El modelo está pensado para la aplicación en frutas de pepita, pero puede ser extensible para su aplicación a otro tipo de frutas u hortalizas.

Se implementaron casos particulares del modelo para la pera y manzana aplicada a la zona del valle de Río Negro. Para este caso se desarrollo un sistema que permite ingresar los parámetros de entrada al modelo y realizar las estimaciones para cada caso en particular. Este hecho permitió validar los resultados contra datos de mediciones reales. El mismo esta siendo utilizado actualmente por una empresa de la zona (Expofrut).

Los resultados del modelo de cierta forma deberían influir en la estrategia comercial y logística de la industria frutícola. Integrar los tamaños de fruta estimados en un sistema comercial permitiría optimizar diferentes aspectos en la compra y distribución de las materias primas a utilizar en cada empresa.

6) Bibliografía

[1] *Deterministic Models for Fruit Growth*

P. W. Gandar, A. J. Hall and H. N. de Silva
The Horticulture and Food Research Institute of New Zealand

[2] *Stochastic Models for Fruit Gorwth*

A. J. Hall and P. W. Gandar
Batchelar Research Centre – New Zeland
1996

[3] *Introduction to Probability Models – Seventh Edition*

Sheldon M. Ross
Department od Industrial Engineering and Operation Research
University of California, Berkeley, California
2000

[4] *Monte Carlo Methods*

J. M. Hammersley – Oxford University of Economics and Statistics
D. C. Handscomb – Oxford University Computing Laboratory
Chapman and Hall
1964

[5] *Probability, random variables, and stochastic processes - Second Edition*

Papoulis, A.
McGraw-Hill Book Co., New York.
1984