

Comparativa entre modelos estocásticos de crecimiento bacteriano a distintas escalas

Nerea Martínez-López, Carlos Vilas and Míriam R. García
Ingeniería de Procesos, IIM-CSIC, Vigo, España
nmartinez@iim.csic.es, carlosvf@iim.csic.es, miriamr@iim.csic.es

Resumen

En la industria alimentaria la calidad y la seguridad del alimento, así como los procesos implicados en su producción, envasado y transporte, se optimizan acudiendo a la microbiología predictiva. Esta disciplina modela el comportamiento de microorganismos, tales como bacterias, empleando modelos primarios para describir su crecimiento.

Uno de los mayores retos en este contexto es modelar y simular el crecimiento cuando el número inicial de bacterias (inóculo) es reducido, ya que la población se comporta estocásticamente. Existen distintas alternativas de modelado, dependiendo de (1) el grado de detalle de la solución requerida, (2) los tiempos de cálculo y (3) la facilidad de calibración de sus parámetros.

En este trabajo se comparan las soluciones y tiempos de cálculo, a la vez que se discute la identificabilidad estructural, cuando se aborda el problema desde una escala microscópica, con modelado basado en individuos, o desde una escala mesoscópica, mediante la derivación de la equivalente ecuación de Fokker-Planck modificada.

Palabras clave: Ecuación de Fokker-Planck modificada; procesos estocásticos; modelado multiescala; identificabilidad estructural.

1. INTRODUCCIÓN

La industria alimentaria está sufriendo una transformación gracias al uso de modelos matemáticos que permiten la optimización de los procesos productivos [12], los tratamientos de conservación [13], el envasado [14], así como el transporte y la distribución de alimentos [7]. En estos modelos, una parte fundamental es la descripción del crecimiento bacteriano. Las bacterias bajo estudio pueden ser tanto patógenas como deterioradoras del alimento (determinando la seguridad y calidad alimentaria), o bien resultar interesantes para los procesos de producción en la industria alimentaria, como aquellas que se encargan de iniciar procesos fermentativos o de producir bioconservantes.

Los modelos más conocidos y empleados, tanto

por la industria como por las agencias evaluadoras de riesgos (por ejemplo la EFSA [9]), están basados en estudios estadísticos puros o dinámicas deterministas. Estos últimos modelos reciben comúnmente el nombre de microbiología predictiva cuando se aplican en ingeniería de alimentos. Sin embargo, su uso está limitado cuando el inóculo es bajo, pues en este caso la población tiende a evolucionar estocásticamente.

Cuando el comportamiento de la población es estocástico, una opción son los modelos basados en individuos [3, 2, 8], que permiten la implementación de mecanismos, por ejemplo de división celular, de una forma sencilla e intuitiva. Sin embargo, su calibración y cómputo para tareas de optimización resulta, muchas veces, prohibitivo. Otra alternativa es acudir a los modelos mesoscópicos, que se centran en describir propiedades medias de la población, normalmente trabajando con distribuciones de probabilidad de las variables de interés.

En [5] se presenta un modelo de crecimiento estocástico de bacterias basado en individuos mediante la implementación de mecanismos de división bien estudiados (cada bacteria crece exponencialmente [4] y estocásticamente [1], y se divide al llegar a cierto tamaño una vez la bacteria está adaptada al medio [10]). Este modelo permite estimaciones de dinámica del número de bacterias para la fase exponencial y de una cota superior (conservadora) para el caso de la fase de latencia. En el mismo trabajo, se presenta la ecuación equivalente de Fokker-Planck modificada, comparando los resultados con datos obtenidos a partir de citometría de flujo para las distribuciones de probabilidad de tamaño celular en la población.

En el presente trabajo se analizan los resultados obtenidos por medio de dos enfoques de modelado (microscópico y mesoscópico) para la distribución de tamaños de una población de bacterias que evoluciona estocásticamente, a la vez que se compara la eficiencia computacional de ambos modelos. Se presenta, además, un breve análisis de la identificabilidad estructural, con el fin de detectar posibles dificultades en la calibración de estos modelos por medio de datos obtenidos con la técnica de citometría de flujo.

2. MODELADO BASADO EN INDIVIDUOS DEL CRECIMIENTO BACTERIANO

Supongamos que $\mathcal{P} = \{1, 2, \dots, N\}$ representa una población compuesta por $N \in \mathbb{N}$ bacterias. El modelo escogido trata de explicar el crecimiento exponencial de cada individuo $i \in \mathcal{P}$ como un proceso que verifica la Ecuación Diferencial Estocástica (EDE) [1]:

$$dX_t^i = \mu dt + \xi dW_t^i, \tag{1}$$

de forma que la variación que experimenta con el tiempo el volumen del i -ésimo individuo (V_t^i , o $X_t^i = \ln V_t^i$ en escala logarítmica) se explica como combinación de una fluctuación estocástica (conducida por un proceso de Wiener, W_t^i) de intensidad ξ , más una componente de desarrollo determinista, cuantificada por el ratio de crecimiento $\mu > 0$. Para $\xi = 0$ la incertidumbre es eliminada del proceso, y el volumen de cada individuo crece exponencialmente con el tiempo.

La división se modela añadiendo los nuevos individuos a la población (*células hijas*), y reajustando el tamaño del individuo que se ha dividido (*célula madre*) al tamaño de sus hijas. Se tiene que el número de individuos generados por una sola madre al término de cada división es $N_h = 2^g - 1$, donde g representa el número de planos en los que se divide la bacteria. Suponiendo que este proceso se activa para un individuo $i \in \mathcal{P}$ cuando su tamaño iguala (o excede) cierto valor límite (definido en función de los tamaños de las madres, $X_m \sim \mathcal{N}(x_m, \sigma^2)$, variable aleatoria normal de media $x_m = \ln(v_m)$ y varianza σ^2 , ambos parámetros del modelo) la división se esquematiza:

$$\begin{cases} \mathcal{P} &= \mathcal{P} \cup \{N + j, 1 \leq j \leq N_h\}, \\ X_t^{N+j} &= X_t^i - g \ln(2), 1 \leq j \leq N_h, \\ X_t^i &= X_t^i - g \ln(2), \end{cases} \tag{2}$$

donde N representa el número de individuos de la población \mathcal{P} antes de la división, y el volumen de la madre es $2g$ veces mayor que el de las hijas ($v_m = 2^g v_h$). En este estudio se considera que la división tiene lugar de forma determinista (las bacterias se dividen a un volumen prefijado v_m), por lo que se toma $\sigma = 0$, y así $X_m = x_m \sim \delta_{x_m}$.

Las simulaciones del modelo basado en individuos (1)–(2) se han llevado a cabo usando distintos esquemas numéricos para la ecuación del crecimiento. Se puede probar que la solución exacta de la EDE (1) es un proceso conocido como *browniano con deriva*:

$$X_t^i = X_{t_0}^i + \mu(t - t_0) + \xi(W_t^i - W_{t_0}^i), \quad t_0 \leq t < t_f,$$

donde t_0 es el instante en el que el individuo i se origina (o finaliza su anterior división), mientras que t_f es el instante en el que se divide nuevamente. Por simplicidad, usamos t_0 y t_f , pero conviene notar que $t_0 = t_0(i, j)$ y $t_f = t_f(i, j)$, con $j \geq 0$ el número de divisiones que ha experimentado la i -ésima bacteria hasta tiempo t_0 .

Discretizando el intervalo $[t_0, t_f]$ en $n \in \mathbb{N}$ puntos $\{t_k, 0 \leq k \leq n - 1\}$, el crecimiento del individuo $i \in \mathcal{P}$ puede implementarse, para $1 \leq k \leq n - 1$, empleando cualquiera de los esquemas numéricos:

Esquema exacto:

$$X_{t_k}^i = X_{t_0}^i + \mu(t_k - t_0) + \xi(W_{t_k}^i - W_{t_0}^i). \tag{3}$$

Euler–Maruyama:

$$X_{t_k}^i = X_{t_{k-1}}^i + \mu(t_k - t_{k-1}) + \xi \Delta W_k^i, \tag{4}$$

donde $\Delta W_k^i = W_{t_k}^i - W_{t_{k-1}}^i$. Ambos esquemas producen resultados idénticos.

Dado que $X_t^i = \ln V_t^i$, aplicando el Lema de Ito [6] se puede probar que V_t^i es un proceso que satisface la EDE (lineal y con ruido multiplicativo):

$$dV_t^i = (\mu + \frac{1}{2}\xi^2)V_t^i dt + \xi V_t^i dW_t^i, \quad t_0 \leq t < t_f,$$

cuya solución es el *browniano geométrico*:

$$V_t^i = V_0^i \exp(\mu(t - t_0) + \xi(W_t^i - W_{t_0}^i)).$$

Así, es posible implementar directamente el modelo de crecimiento en volumen para la partición $\{t_k, 0 \leq k \leq n - 1\}$ con alguno de los esquemas:

Esquema exacto:

$$V_{t_k}^i = V_0^i \exp(\mu(t_k - t_0) + \xi(W_{t_k}^i - W_{t_0}^i)). \tag{5}$$

Euler–Maruyama:

$$V_{t_k}^i = V_{t_{k-1}}^i (1 + \alpha(t_k - t_{k-1}) + \xi \Delta W_k^i). \tag{6}$$

Milstein–II.

$$\begin{aligned} V_{t_k}^i &= V_{t_{k-1}}^i (1 + \alpha(t_k - t_{k-1}) + \xi \Delta W_k^i) + \\ V_{t_{k-1}}^i &\alpha(t_k - t_{k-1}) (\xi \Delta W_k^i + \frac{\alpha}{2}(t_k - t_{k-1})) + \\ V_{t_{k-1}}^i &\frac{\xi^2}{2} ((\Delta W_k^i)^2 - (t_k - t_{k-1})), \end{aligned} \tag{7}$$

donde se ha tomado $\alpha = \mu + \xi^2/2$.

3. ECUACIÓN DE FOKKER - PLANCK MODIFICADA

Sea $p(x, t)$ la función que describe la probabilidad de que las células tengan un tamaño x (en escala logarítmica) a tiempo t , es decir, la función de densidad de probabilidad (FDP). Una propiedad bien conocida de la FDP es:

$$\int_{-\infty}^{\infty} p(x, t) dx = 1. \tag{8}$$

Cuando no se considera división ni muerte celular, $p(t, x)$ se obtiene como la solución de la ecuación de Fokker–Planck:

$$\frac{\partial p(x, t)}{\partial t} = \underbrace{\frac{\xi^2}{2} \frac{\partial^2 p(x, t)}{\partial x^2} - \mu \frac{\partial p(x, t)}{\partial x}}_{\text{crecimiento celular con incertidumbre}}, \quad (9)$$

donde, al igual que en la sección anterior, μ representa la velocidad de crecimiento determinista y ξ la intensidad del ruido.

Sin embargo, estamos considerando que, al llegar a cierto volumen v_m , las bacterias se dividen en 2^g células hijas. La solución de la ecuación de Fokker–Planck clásica, teniendo en cuenta dicha división, no verificaría la igualdad (8), pues el valor de la integral aumentaría con el tiempo. Para evitar esto, se añade al modelo un término de normalización, resultando en la ecuación de Fokker–Planck modificada (EFPM) [5]:

$$\frac{\partial p(x, t)}{\partial t} = \frac{\xi^2}{2} \frac{\partial^2 p(x, t)}{\partial x^2} - \mu \frac{\partial p(x, t)}{\partial x} - \underbrace{p(J(x_h, t) - J(x_m, t))}_{\text{Normalización}}, \quad (10)$$

siendo $J(x, t)$ el flujo de la FDP [5]:

$$J(x, t) = \mu p(x, t) - \frac{\xi^2}{2} \frac{\partial p(x, t)}{\partial x}.$$

La derivación de la expresión (10) se puede consultar en el Apéndice A.

Dado que las células se dividen al alcanzar un tamaño $x_m = \ln(v_m)$, la probabilidad de encontrar células de tamaño mayor o igual que x_m es cero. Matemáticamente, esto se representa imponiendo condiciones frontera de *wand absorbente*, es decir:

$$p(x_m, t) = 0. \quad (11)$$

Además, el flujo de nuevas células que surgen de la división (es decir, $J(x_h, t)$) es 2^g veces el flujo de células que se dividen ($J(x_m, t)$), de forma que $J(x_h, t) = 2^g J(x_m, t)$. Por lo tanto, la condición frontera en x_h quedaría:

$$-\frac{\xi^2}{2} \frac{\partial p(x, t)}{\partial x} \Big|_{x=x_h} = \mu(2^g p(x_m, t) - p(x_h, t)) - 2^g \frac{\xi^2}{2} \frac{\partial p(x, t)}{\partial x} \Big|_{x=x_m}. \quad (12)$$

Inicialmente, las células tienen un tamaño x_h . Por tanto, la distribución de probabilidad inicial es:

$$p(x, 0) = \delta_{x_h}(x), \quad (13)$$

donde $\delta_{x_h}(x)$ es la delta de Dirac centrada en x_h .

La FDP respecto del volumen, $p(v(x), t)$, se obtiene mediante la expresión:

$$p(v(x), t) = \frac{p(x, t)}{dv(x)/dx} = \exp(-x)p(x, t). \quad (14)$$

4. RESULTADOS

Los valores de los parámetros empleados en las simulaciones (véase la Tabla 1) fueron obtenidos a partir de citometría de flujo para *Pediococcus acidilactici* [5]: una bacteria productora de pediocina que supone un antimicrobiano relevante para la industria alimentaria [12].

Tabla 1: Parámetros usados en las simulaciones.

Símb.	Significado	Valor	Uds.
v_m	Volumen de las madres	4.55740	μm^3
ξ	Intensidad del ruido	0.13439	$\text{h}^{-1/2}$
μ	Velocidad crecimiento	1.16190	h^{-1}
g	Planos de división	2.00000	-

4.1. MODELADO BASADO EN INDIVIDUOS

Las simulaciones arrancan en $t_0 = 0$ con una única bacteria de volumen $V_0^1 = v_h$, la cual evoluciona de acuerdo con la ecuación (1) y se divide según (2), hasta alcanzar un tiempo final de simulación. En la Figura 1 se muestra la evolución que experimenta el volumen de cada una de las bacterias de la población durante las primeras cinco horas de desarrollo, empleando el esquema exacto (5) y un paso de tiempo constante $\Delta t = 0.01$ h. El resultado que se obtiene es idéntico cuando se parte de una bacteria de tamaño $X_0^1 = \ln(x_h)$ y se aplica el método (3), realizando posteriormente la transformación $V_t^i = \exp(X_t^i)$, para cada $i \in \mathcal{P}$.

Dada la naturaleza estocástica del modelo, la trayectoria que se muestra en la Figura 1 para el volumen de cada bacteria $i \in \mathcal{P}$ varía al tomar distintas realizaciones del proceso de Wiener W_t^i en la expresión (5). Se tiene, no obstante, $E[V_t^i] = V_0^i \exp(\mu(t - t_0) + \xi^2(t - t_0)/2)$, propiedad que se aprecia intuitivamente en la figura. Análogamente, se verifica $E[X_t^i] = X_{t_0}^i + \mu(t - t_0)$.

4.2. ECUACIÓN DE FOKKER – PLANCK MODIFICADA

Para implementar el modelo (10)–(14) se ha empleado el método de Diferencias Finitas (DF). Concretamente, se ha considerado un esquema de cinco puntos de DF centradas para aproximar la derivada segunda, y un esquema de cuatro puntos de DF progresivas para aproximar la derivada primera con respecto de x . Una explicación detallada de estos métodos se puede encontrar en [11]. Se han elegido 400 puntos de discretización en la variable x , dado que emplear un mayor número de puntos no altera significativamente los resulta-

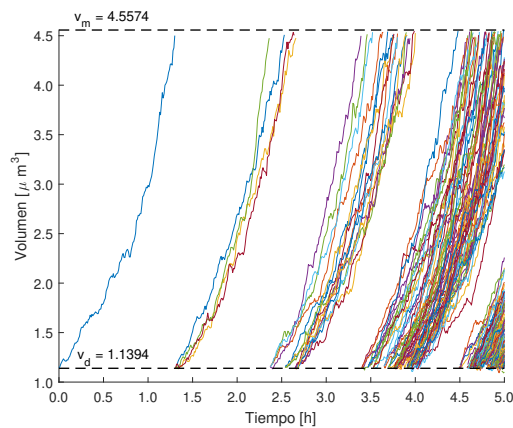


Figura 1: Evolución del volumen para el modelo de crecimiento y división (1)–(2). El número de bacterias para $t = 5$ horas es $N = 172$.

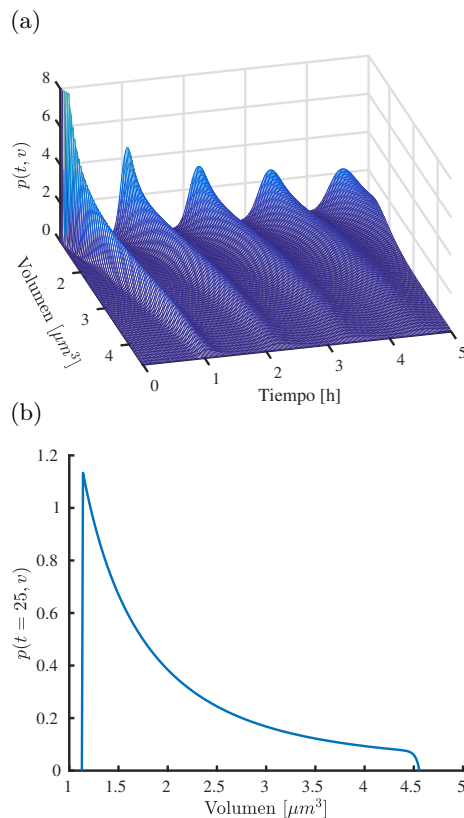


Figura 2: Solución numérica del sistema de ecuaciones (10)–(14). (a) Evolución en el tiempo hasta $t = 5$ h. (b) Estado estacionario ($t = 25$ h).

dos. La integración temporal se ha llevado a cabo usando el integrador *ode15s* en Matlab®.

La solución numérica obtenida con el proceso que se acaba de describir se muestra en la Figura 2. Para poder visualizar mejor la evolución de la distribución de probabilidad, se ha detenido la simulación en tiempo $t = 5$ h -Figura 2(a)-. Desde el ins-

Modelado, Simulación y Optimización
 Tabla 2: Comparación entre métodos de simulación. MBI = Modelo Basado en Individuos; EFPM = Ecuación de Fokker–Planck modificada; AN = Analítica; EM=Euler–Maruyama; MII=Milstein-II; DF=Diferencias finitas.

Modelo	Método	Tiempo comp. [s]	RMSE	
			t	t_f
MBI (X)	AN	813	-	-
MBI (X)	EM	744	0	0
EFPM (x)	DF	1	0.107	0.043
MBI (V)	AN	1365	-	-
MBI (V)	EM	685	0.031	0.017
MBI (V)	MII	876	0.019	0.010
EFPM (v)	DF	1	0.063	0.035

tante inicial, la probabilidad de encontrar células de mayor (resp. menor) tamaño, aumenta (resp. disminuye) con el tiempo. Como era de esperar, la probabilidad de obtener células de tamaño mayor que v_m es 0, ya que la división se produce al alcanzar dicho volumen máximo. Esto ocurre en la vecindad de los tiempos $t \in \{1, 2.1, 3.3, 4.5, \dots\}$. A medida que avanza el tiempo, también aumenta el tamaño de dicha vecindad, hasta generar un flujo continuo de células de volumen v_h . Aproximadamente a tiempo $t = 25$ h se alcanza el estado estacionario -Figura 2(b)-. El tamaño más probable (modo de la distribución) se corresponde con v_h , y va decreciendo de forma no lineal hasta alcanzar el valor máximo v_m .

4.3. COMPARACIÓN ENTRE MODELOS Y MÉTODOS DE SIMULACIÓN

4.3.1. ERROR Y TIEMPO DE CÁLCULO

La Tabla 2 resume los resultados obtenidos para los modelos basados en individuos y para la EFPM. Los cálculos se han llevado a cabo en Matlab R2016b®, usando una estación de trabajo con 32 GB de RAM (2400 MHz) y CPU Intel® Core™ i7-7700K (4.20 GHz). La columna *Tiempo comp.* indica el tiempo empleado por cada método para obtener la solución. En el caso del modelo basado en individuos, la simulación se para cuando el proceso alcanza el tiempo $t_f = 17$ h. Para los distintos esquemas numéricos aplicados al modelo basado en individuos, el tiempo de computación aumenta exponencialmente con la duración del proceso - ver Figura 3-. Dado que el proceso es estocástico, la solución depende de la semilla que se use para generar los números aleatorios necesarios para construir trayectorias del browniano. En dicha figura se muestra la media de tiempos para

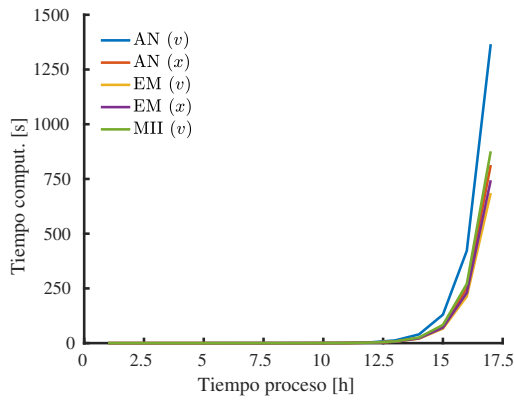


Figura 3: Evolución del tiempo de computación como función del tiempo de proceso para los distintos métodos de resolución del modelo basado en individuos.

250 simulaciones con distintas semillas. El método más eficiente es el esquema de Euler–Maruyama (con crecimiento en volumen) y el menos eficiente se obtiene implementando la solución analítica.

La primera columna del error (t) se refiere a la raíz del error cuadrático medio (RMSE), el cual se ha calculado empleando los perfiles obtenidos en todos los tiempos de proceso. La segunda columna del error (t_f) muestra el RMSE calculado usando los perfiles a tiempo final. El RMSE de la primera columna es mayor que el de la segunda, pues el número de células al principio del proceso es pequeño, de manera que se obtienen perfiles muy ruidosos -ver Figura 4-. A medida que el proceso transcurre, el número de células aumenta significativamente y los perfiles se suavizan, aproximándose cada vez más a la solución obtenida mediante la EFPM. Los perfiles que se muestran en la Figura 4 para el modelo basado en individuos se corresponden con el valor medio de 3000 simulaciones con distintas semillas de generación de números aleatorios. Se puede ver, tanto en la Tabla 2 como en la Figura 4(d)-(f), que (tal y como se ha mencionado en la Sección 2) las soluciones analítica y Euler–Maruyama coinciden en el dominio X , pero el esquema de Euler–Maruyama es más eficiente computacionalmente. Respecto a la EFPM, se puede ver que representa con bastante precisión el comportamiento medio de los modelos basados en individuos. Las diferencias entre ellos vienen dadas, principalmente, porque los modelos basados en individuos permiten tamaños menores que x_h , mientras que la solución de la EFPM está limitada al dominio $[x_h, x_m]$. Sin embargo, la EFPM es mucho más eficiente que los modelos basados en individuos, los cuales, a partir de un cierto número de células, se vuelven demasiado costosos y presentan problemas de memoria.

4.3.2. IDENTIFICABILIDAD ESTRUCTURAL

La medición directa de los parámetros necesarios para la implementación de los modelos que se presentan en este trabajo (Tabla 1) requiere experimentos complejos, donde las estimaciones no son fiables al referirse a propiedades individuales de cada célula (ver [5] para detalles). Una alternativa es el uso de distribuciones de probabilidad de tamaño celular, que se pueden medir rápidamente, y de forma relativamente barata, empleando la técnica de citometría de flujo. Sin embargo, las distribuciones habitualmente se miden solo en estado estacionario, por lo que son incapaces de proporcionar demasiada información útil para la estimación de estos parámetros.

De esta forma, se hace necesario establecer si, dada la información que proporciona la distribución de tamaños en el estado estacionario, podemos llegar a estimar los parámetros de la Tabla 1 de forma unívoca con uno u otro modelo. En tal caso, el modelo sería estructuralmente identificable.

Dos de los parámetros que intervienen en los modelos son estructuralmente identificables: el volumen de las madres v_m , y el número de planos de división g . Nótese que el valor del volumen de las madres corresponde con el final de la distribución de probabilidad, es decir, con el mayor tamaño celular cuya probabilidad es superior a cero. Los planos de división se pueden obtener mediante la relación entre los volúmenes de las madres y las hijas, correspondiendo estas últimas con el modo de la distribución.

Para el análisis de la identificabilidad estructural de la velocidad de crecimiento μ y la intensidad del ruido ξ , se hace un estudio adimensional de la EFPM (véase el Apéndice B para más detalles):

$$\frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{t}} = \frac{1}{Pe} \frac{\partial^2 p(\bar{x}, \bar{t})}{\partial \bar{x}^2} - \frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{x}} - pL(J(0, \bar{t}) - J(1, \bar{t}))Pe, \quad (15)$$

donde $Pe = 2\mu L/\xi^2$ y

$$J(\bar{x}, \bar{t}) = \frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{x}} - p(\bar{x}, \bar{t})Pe,$$

$$J(0, \bar{t}) = 2^g J(1, \bar{t}),$$

$$p(1, \bar{t}) = 0.$$

Nótese que existen infinitas combinaciones de los parámetros ξ y μ que resultan en el mismo valor de Pe y, por lo tanto, en soluciones idénticas de la ecuación (15). Como consecuencia, se deduce que los parámetros ξ y μ no son estructuralmente identificables a partir de las distribuciones de tamaño obtenidas con citometría de flujo, pudiéndose extraer solo una relación de dependencia entre ellos.

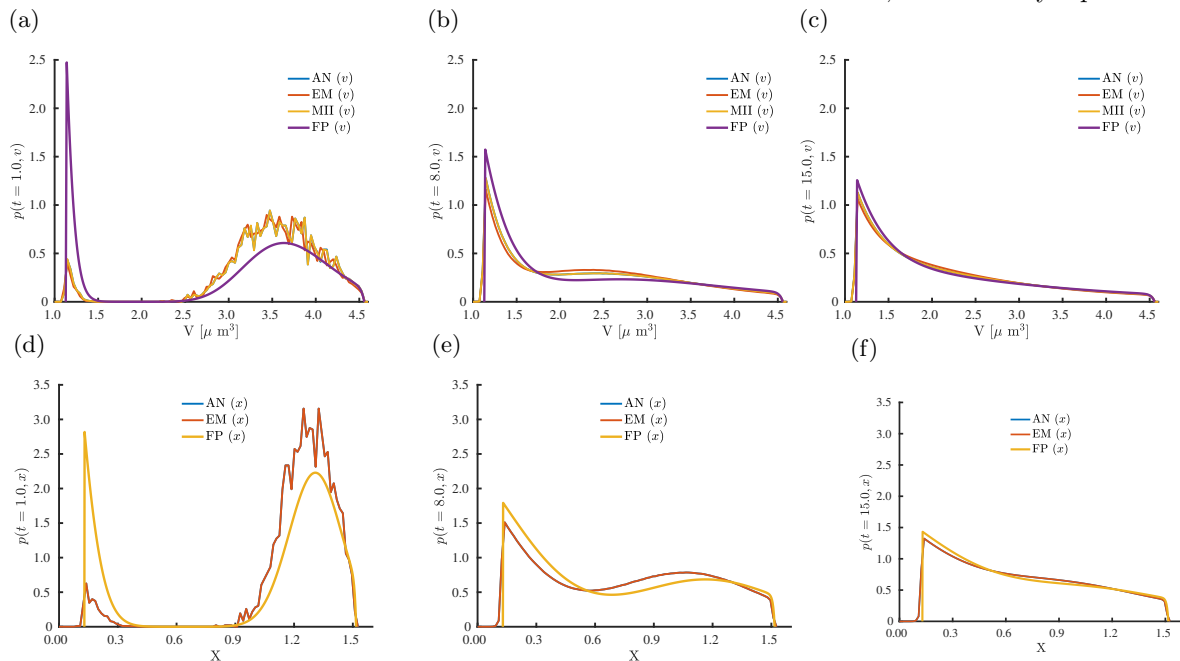


Figura 4: Comparación entre los resultados de simulación obtenidos con el modelo basado en individuos y la EFPM. Las figuras de arriba se corresponden con las soluciones obtenidas en el dominio v mientras que las de abajo fueron obtenidas en el dominio x . Los tiempos de proceso de las figuras son $t = 1$ h (a),(d). $t = 8$ h (b),(e). $t = 15$ h (c),(f).

Este análisis de identificabilidad es válido siempre que se use una distribución estacionaria de tamaños como única fuente de información, y afectaría tanto a la EFPM como al modelo basado en individuos. Para este último tipo de modelado, además, la obtención de las distribuciones es más costosa y requiere mayor tiempo de computación.

5. CONCLUSIONES

En este trabajo se han considerado distintas alternativas para modelar la evolución de poblaciones de bacterias cuando el crecimiento es estocástico. Por una parte, se ha hecho uso de modelos basados en individuos, para los cuales se han considerado varios esquemas: (i) solución exacta (el menos eficiente), (ii) Euler–Maruyama (el más eficiente), y (iii) Milstein–II. La raíz del error cuadrático medio (RMSE) entre las soluciones analíticas y los esquemas numéricos es menor de 0.04 en todos los casos, lo que demuestra su robustez. Sin embargo, el tiempo de cálculo necesario y la memoria requerida en estos esquemas se incrementa exponencialmente con el número de células, es decir, con la duración del proceso simulado. Por otra parte, se ha usado la ecuación de Fokker–Planck modificada (EFPM) cuya principal ventaja, en comparación con los modelos basados en individuos, es la eficiencia para un número de células medio-alto. Además, se ha mostrado que la distribución de probabilidad obtenida mediante la EFPM se

ajusta a la media de los resultados obtenidos con los modelos basados en individuos. La EFPM se ha usado para mostrar que la velocidad de crecimiento (μ) y la intensidad de ruido (ξ) no se pueden identificar simultáneamente usando datos de citometría de flujo en estado estacionario.

Agradecimientos

Este trabajo ha sido financiado por los proyectos PIE 20213AT001 (NM), PGC2018-099312-B-C33 (CVF) y RTI2018-093560-J-I00 (MCIU/AEI/FEDER, UE), RYC2019-028006-I/AIE/10.13039/501100011033 (MRG).

Los autores quieren agradecer y dedicar este trabajo al Prof. Antonio A. Alonso, uno de los investigadores que iniciaron originalmente este trabajo. Agradecemos su apoyo incansable y su continuo asesoramiento.

English summary

COMPARING STOCHASTIC MODELS OF BACTERIAL GROWTH AT DIFFERENT SCALES

Abstract

Quality and safety in the food industry, as well as the processes involved in the production, packaging, and transport of food,

may be optimised using predictive microbiology. This discipline models the behaviour of microorganisms and uses the so-called primary models to describe bacterial growth.

One of the main challenges consists of explaining and simulating the growth when the number of initial bacteria (inoculum) is reduced because of the stochastic behaviour. There exist different alternatives to modelling such stochastic processes depending on: (1) the required level of detail for the solution; (2) the computational time; and (3) the possibility of estimating parameters from fast and inexpensive measurements.

In this work, we compare the solutions, computational times and structural identifiability obtained using different alternatives: (1) microscopic scale considering individual-based models; and (2) mesoscopic scale considering the modified Fokker-Planck equation.

Keywords: Modified Fokker-Planck equation; Stochastic processes; Multiscale modelling; structural identifiability.

Referencias

- [1] A. A. Alonso, I. Molina, and C. Theodoropoulos. Modeling Bacterial Population Growth from Stochastic Single Cell Dynamics. *Applied and environmental microbiology*, 80(17):5241–5253, 2014.
- [2] J. C. Augustin, R. Ferrier, B. Hezard, A. Lintz, and V. Stahl. Comparison of individual-based modeling and population approaches for prediction of foodborne pathogens growth. *Food Microbiology*, 45:205–215, 2015.
- [3] J. Ferrer, C. Prats, D. López, and J. Vives-Rego. Mathematical modelling methodologies in predictive food microbiology: A SWOT analysis. *International Journal of Food Microbiology*, 134(1-2):2–8, 2009.
- [4] I. Fishov, A. Zaritsky, and N. B. Grover. On microbial states of growth. *Molecular microbiology*, 15(5):789–94, mar 1995.
- [5] M. R. García, J. A. Vázquez, I. G. Teixeira, and A. A. Alonso (2018). Stochastic individual-based modeling of bacterial growth and division using flow cytometry. *Frontiers in Microbiology*, 8:2626.
- [6] P. E. Kloeden and E. Platen. Numeric Solution of Stochastic Differential Equations. *Springer*, 1992.
- [7] K. Koutsoumanis, A. Allende, A. Alvarez-Ordóñez, D. Bolton, M. Chemaly, R. Davies, A. De Cesare, L. Herman, F. Hilbert, R. Lindqvist, M. Nauta, L. Peixe, G. Ru, M. Simmons, P. Skandamis, E. Suffredini, S. Arason, K. Bekaert, M. R. García, M. Georgiadis, W. Messens, O. Mosbach-Schulz, and S. Bover-Cid. The use of the so-called ‘tubs’ for transporting and storing fresh fishery products. *EFSA Journal*, 18(4):1–123, 2020.
- [8] K. P. Koutsoumanis and Z. Aspidou. Individual cell heterogeneity in Predictive Food Microbiology: Challenges in predicting a “noisy” world. *International Journal of Food Microbiology*, 240:3–10, jan 2017.
- [9] A. Ricci, A. Allende, D. Bolton, M. Chemaly, R. Davies, P. S. Fernández Escámez, R. Girones, L. Herman, K. Koutsoumanis, R. Lindqvist, L. Robertson, G. Ru, M. Sanaa, M. Simmons, P. Skandamis, E. Snary, N. Speybroeck, B. Ter Kuile, J. Threlfall, H. Wahlström, J. K. Andersen, M. Uyttendaele, A. Valero, M. T. Da Silva Felício, W. Messens, and B. Nørrung. Guidance on the requirements for the development of microbiological criteria. *EFSA Journal*, 15(11):5052, nov 2017.
- [10] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. Cell-Size Control and Homeostasis in Bacteria. *Current Biology*, 25(3):385–391, feb 2015.
- [11] A. Vande Wouwer, P. Saucez, and C. Vilas. *Simulation of ODE/PDE Models with MATLAB®, OCTAVE and SCILAB. Scientific and Engineering Applications*. Springer International Publishing, Cham, 2014.
- [12] J. A. Vázquez, M. P. González, and M. Murado. Preliminary tests on nisin and pediocin production using waste protein sources: factorial and kinetic studies. *Bioresource Technology*, 97(4):605–613, 2006.
- [13] C. Vilas, M. García, J. Banga, and A. Alonso. Desarrollo de una librería de componentes

en ecosimpro para la operación de plantas de procesamiento térmico de alimentos. *RIAI - Revista Iberoamericana de Automática e Informática Industrial*, 5(1):51–65, 2008.

- [14] C. Vilas, M. Mauricio-Iglesias, and M. R. García. Model-based design of smart active packaging systems with antimicrobial activity. *Food Packaging and Shelf Life*, 24, jun 2020.



© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).

APÉNDICES

A. EFPM: OBTENCIÓN

Sea $n(x, t)$ el número de células de tamaño x (en escala logarítmica) a tiempo t . Sea, además, $J_n(x, t)$ el flujo de células. La evolución del número de células se puede representar de la siguiente forma:

$$\frac{\partial n(t, x)}{\partial t} = -\frac{\partial J_n(t, x)}{\partial x}. \tag{16}$$

Al llegar a cierto volumen $v_m = \exp(x_m)$, las células se dividen en 2^g células hijas de tamaño $v_h = v_m/2^g$. Este fenómeno se puede representar usando una condición de muro absorbente en la frontera x_m :

$$n(x_m, t) = 0, \tag{17}$$

de forma que la condición frontera en $x = x_h$ se define mediante los flujos:

$$J_n(x_h, t) = 2^g J_n(x_m, t). \tag{18}$$

Debido a esta división y, dado que la muerte de las células es despreciable en fase exponencial [5], el número de células $n(x, t)$ será creciente. Esto se puede evitar usando la forma fraccional $p(x, t)$, donde:

$$p(x, t) = \frac{n(x, t)}{N(t)}, \text{ siendo } N(t) = \int_{x_h}^{x_m} n(x, t) dx.$$

Derivando respecto del tiempo¹:

$$\frac{\partial p}{\partial t} = \frac{1}{N} \frac{\partial n}{\partial t} - \frac{p}{N} \frac{dN}{dt}. \tag{19}$$

El primer término del lado derecho de la ecuación (19) se puede expresar, usando la ecuación (16), de la siguiente forma:

$$\frac{1}{N} \frac{\partial n}{\partial t} = -\frac{1}{N} \frac{\partial J_n}{\partial x} = -\frac{\partial J}{\partial x},$$

¹A partir de aquí se omitirán los argumentos de las funciones, excepto para los términos en la frontera, por motivos de claridad

donde $J = J_n/N$. Para el segundo término se tiene que

$$-\frac{p}{N} \frac{dN}{dt} = -\frac{p}{N} \int_{x_h}^{x_m} \frac{\partial n}{\partial t} dx.$$

Usando la ecuación (16) y reordenando términos:

$$-\frac{p}{N} \frac{dN}{dt} = \frac{p}{N} \int_{x_h}^{x_m} \frac{\partial J_n}{\partial x} dx = -p(J(x_h, t) - J(x_m, t)).$$

Por lo tanto, la ecuación (19) se puede escribir de la siguiente forma:

$$\frac{\partial p}{\partial t} = -\frac{\partial J}{\partial x} - p(J(x_h, t) - J(x_m, t)), \tag{20}$$

que se corresponde con la expresión presentada en la Sección 3. Además, dividiendo las condiciones frontera (17)-(18) entre el número de células (N), se tiene que:

$$p(x_m, t) = 0, \tag{21}$$

$$J(x_h, t) = 2^g J(x_m, t). \tag{22}$$

B. FORMA ADIMENSIONAL DE LA EFPM

Para derivar una forma adimensional de la EFPM, utilizaremos la longitud característica $L = x_m - x_h$ y la velocidad de crecimiento μ . De esta forma, se puede definir:

$$\bar{x} = \frac{x - x_h}{L}, \quad \bar{t} = \frac{L}{\mu} t,$$

donde $x \in [0, 1]$. Las derivadas respecto de x y t quedarían, en el nuevo sistema de coordenadas adimensionales \bar{x} y \bar{t} , de la siguiente forma:

$$\frac{\partial}{\partial x} = \frac{1}{L} \frac{\partial}{\partial \bar{x}}, \quad \frac{\partial^2}{\partial x^2} = \frac{1}{L^2} \frac{\partial^2}{\partial \bar{x}^2}, \quad \frac{\partial}{\partial t} = \frac{\mu}{L} \frac{\partial}{\partial \bar{t}}.$$

Sustituyendo estas expresiones en la ecuación (10), reordenando los términos, y usando un número adimensional similar al Péclet, $Pe = 2\mu L/\xi^2$, se tiene:

$$\frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{t}} = \frac{1}{Pe} \frac{\partial^2 p(\bar{x}, \bar{t})}{\partial \bar{x}^2} - \frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{x}} - pL(J(0, \bar{t}) - J(1, \bar{t})) Pe, \tag{23}$$

donde

$$J(\bar{x}, \bar{t}) = \frac{\partial p(\bar{x}, \bar{t})}{\partial \bar{x}} - p(\bar{x}, \bar{t}) Pe.$$

Aplicando la misma transformación a las condiciones frontera (11)-(12), se tiene:

$$p(1, \bar{t}) = 0, \\ J(0, \bar{t}) = 2^g J(1, \bar{t}).$$