

PREDICCIÓN DE BACTERIAS MEDIANTE REDES BAYESIANAS EN LA BAHÍA DE BANDERAS, MÉXICO

BACTERIA PREDICTION BY MEANS OF BAYESIAN NETS IN BAHIA DE BANDERAS, MEXICO

Liza Kelly-Gutierrez¹, Luis Plata-Rosas³, Saúl Guerrero-Galván²

lizadke@pv.udg.mx / luispla@pv.udg.mx / guerrero@pv.udg.mx

RESUMEN. Para determinar las variables que permiten predecir la ocurrencia de bacterias (coliformes fecales y totales y estreptococos fecales) en la Bahía de Banderas mediante redes bayesianas se realizaron 21 campañas oceanográficas mensuales en las que se midieron 45 variables, en 21 estaciones de muestreo, de marzo de 2003 a diciembre de 2005. Se separaron los datos en secas y lluvias, se calcularon los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90 para cada variable y se crearon matrices con los datos que están dentro de cada uno de los siguientes intervalos: $x \leq P_{10}$, $P_{10} < x \leq P_{20}$, $P_{20} < x \leq P_{30}$, $P_{30} < x \leq P_{40}$, $P_{40} < x \leq P_{50}$, $P_{50} < x \leq P_{60}$, $P_{60} < x \leq P_{70}$, $P_{70} < x \leq P_{80}$, $P_{80} < x \leq P_{90}$ y $x > P_{90}$. Se calculó la probabilidad conjunta $[P(A \cap B)]$, la probabilidad condicional $[P(A / B)]$ y la independencia estadística con el teorema de Bayes $[P(A / B) = P(A \cap B) / P(B)]$, donde los eventos A, tanto en época de secas como de lluvias, son las matrices correspondientes a las bacterias, y los eventos B son todas las variables restantes. Para determinar en qué casos las probabilidades condicionales son significativas se hizo una prueba de bondad de ajuste con un nivel de significancia de 0.05. Con base en la periodicidad del muestreo, la duración total del estudio y la probabilidad de predicción, las mejores variables predictoras en la zona de mayor influencia continental en la Bahía de Banderas fueron: 1) para coliformes totales, en época de secas, la profundidad del disco de Secchi y la razón N/SiO₄ y, en época de lluvias, el fósforo total y la profundidad del disco de Secchi; 2) para coliformes fecales, tanto en época de secas como en época de lluvias, la salinidad, la profundidad del disco de Secchi y la razón N/SiO₄; 3) para estreptococos fecales, en época de secas, la temperatura, el oxígeno disuelto, la profundidad del disco de Secchi, demanda biológica de oxígeno, y las razones N/P, N/SiO₄ y MOP/SS y, en época de lluvias, la salinidad, la profundidad del disco de Secchi, el fósforo reactivo soluble, el fósforo total, las clorofilas C y A corregida por feopigmentos, y las razones N/P y N/SiO₄.

PALABRAS CLAVE: Coliformes fecales, coliformes totales, estreptococos fecales, redes bayesianas, Bahía de Banderas-México.

ABSTRACT. To determine the variables that allow predicting the occurrence of bacteria (fecal and total coliforms and fecal streptococcus) in the Banderas Bay by means of Bayesian networks from March 2003 to December 2005 were carried out 21 monthly oceanographic surveys in which 45 variables were measured in 21 sampling stations. Data were blocked by dry and rainy seasons, the percentils 10, 20, 30, 40, 50, 60, 70, 80 and 90 were calculated for each variable and matrices were made with data corresponding to the following intervals: $x \leq P_{10}$, $P_{10} < x \leq P_{20}$, $P_{20} < x \leq P_{30}$, $P_{30} < x \leq P_{40}$, $P_{40} < x \leq P_{50}$, $P_{50} < x \leq P_{60}$, $P_{60} < x \leq P_{70}$, $P_{70} < x \leq P_{80}$, $P_{80} < x \leq P_{90}$ y $x > P_{90}$. We calculated join $[P(A \cap B)]$ and conditional $[P(A | B)]$ probabilities, and we tested the statistical independence according to the Bayes theorem $[P(A | B) = P(A \cap B) / P(B)]$, where A events, both in dry and rainy seasons, are the matrices corresponding to bacteria and B events are the remaining variables. To establish in which cases the conditional probabilities are significant, a goodness of fit test, with a significant level of 0.05, was done. Considering the survey periodicity, the total duration of the study and the prediction probability, we proposed the following variables as bacteria predictors in the larger continental influence zone of the bay: 1) for total coliforms, during dry season, Secchi disc depth and the ratio N/SiO₄ and, during rainy season, total phosphorus and Secchi disc depth; 2) for fecal coliforms, both in dry and rainy seasons, salinity, Secchi disc depth and the ratio N/SiO₄; 3) for fecal streptococcus, during dry seasons, temperature, dissolved oxygen, Secchi disc depth, biological demand of oxygen, and the ratios N/P, N/SiO₄ and MOP/SS and, during rainy season, salinity, Secchi disc depth, soluble reactive phosphorus, total phosphorus, chlorophyll C and corrected chlorophyll A for feopigments, and the ratios N/P and N/SiO₄.

KEYWORDS: Total coliform, fecal coliform, fecal streptococcus, bayesian nets, Banderas Bay- México.

¹ Posgrado en Desarrollo Sustentable y Turismo, Centro Universitario de la Costa, Universidad de Guadalajara. Av. Universidad de Guadalajara No.203, Puerto Vallarta, 48280, Jalisco, México. www.cuc.udg.mx

² Departamento de Ciencias Exactas, Centro Universitario de la Costa, Universidad de Guadalajara. Av. Universidad de Guadalajara No.203, Puerto Vallarta, 48280, Jalisco, México. www.cuc.udg.mx

³ Departamento de Ciencias Biológicas y de la Salud, Centro Universitario de la Costa, Universidad de Guadalajara. Av. Universidad de Guadalajara No.203, Puerto Vallarta, 48280, Jalisco, México. www.cuc.udg.mx



Introducción

El desarrollo sustentable de la zona costera requiere de la evaluación del estado actual de los recursos marinos y de los procesos (físicos, químicos, geológicos y biológicos) que en ella tienen lugar. Estudios científicos que tienen por objeto generar este conocimiento son indispensables en el manejo integral costero para determinar la interdependencia entre las decisiones a tomar y la respuesta del medio [1]. Sin embargo, un diagnóstico realista que abarque todos los procesos relevantes en el análisis de un problema ambiental particular (e.g. la determinación de la calidad del agua de una bahía u otro cuerpo costero) es difícil debido a que la situación más usual es aquella en la que grupos de investigadores independientes estudian diferentes aspectos del sistema con base en sus propios intereses científicos [2].

Incluso en estos trabajos, la comprensión de cierto proceso es bastante limitada en las escalas de tiempo y espacio, lo que se traduce en diferencias o incertidumbres entre los valores actuales (desconocidos) y los valores medidos de los parámetros de procesos oceanográficos [3] debido a simplificaciones prácticas y a parametrizaciones inexactas. El uso de modelos de simulación para la cuantificación y predicción de estas incertidumbres es común actualmente. La mayoría de estos modelos mecanicistas intentan reproducir todos los procesos que ocurren en el sistema en cierta escala espacio-temporal predeterminada [4]. Sin embargo, debido a no hay una única escala en la que todos los procesos ocurran, esto conduce con frecuencia a modelos que están sobreparametrizados o subparametrizados con respecto a la escala de datos existentes [5].

Es también un problema inevitable durante la implementación de modelos de simulación de la zona costera el que la exactitud de una predicción en el comportamiento del ecosistema acuático requiere de una comprensión científica (i.e., entender las interacciones de los procesos oceanográficos que describen la variación espacio-temporal de los parámetros de interés), evidencia observacional (i.e., llevar a cabo mediciones en el área de estudio, el análisis de laboratorio de las muestras y el tratamiento de los datos medidos) y descripciones matemáticas (i.e., el empleo de modelos teóricos y ecuaciones) que, por lo general, son superiores a aquellas con las que se cuenta [6]. Además, la validación de un modelo involucra la selección de varios parámetros, provenientes de un registro histórico, que son “sintonizados” manualmente por el modelador o automáticamente a través de un algoritmo [7] de manera que las predicciones de la simulación coincidan con las mediciones, pero el principal interés de los responsables de las decisiones en el manejo costero radica en los resultados que el modelo predecirá con base en cualquier otro conjunto de datos actual o futuro, imposible de validar la mayoría de las veces.

Debido a las razones expuestas en el párrafo anterior, y a pesar del uso común de los modelos numéricos de simulación, una estrategia cada vez más empleada desde una perspectiva de toma de decisiones en ecología, es el uso de redes bayesianas de probabilidad [8-10]. Los avances ocurridos en el mundo de la informática durante los últimos años, especialmente en el aumento de la velocidad de procesamiento de datos, así como la reducción de los precios de los componentes informáticos, han permitido el desarrollo de programas específicos necesarios para la construcción y manejo de redes bayesianas complejas [11]. Las redes bayesianas representan en términos probabilísticos las relaciones entre las variables de un sistema que tienen relevancia directa en éste, y pueden ser usadas para expresar incertidumbre y asociaciones complejas (e.g. sistemas dinámicos no lineales) entre variables [12]. Desde fechas recientes las redes bayesianas se aplican en la resolución de problemas medioambientales, ya que estos presentan una estructura complicada, con altos grados de incertidumbre y con elevadas dificultades en la toma de datos [13]. Inclusive, algunos autores (e.g. [14]) consideran que el análisis bayesiano es el más apropiado en la toma de decisiones inherente al manejo ambiental.

Las redes bayesianas se están empezando a utilizar en medicina para descubrir patrones en los datos biomédicos y también como una base para disminuir la incertidumbre asociada a la toma de decisiones. Relacionadas con bacterias, se han aplicado redes bayesianas en los campos de biomedicina y microbiología de los alimentos. En biomedicina para predecir la eficacia del diagnóstico con base en estudios como rayos-X [15] o la sintomatología integrando información biomédica y clínica [16] y para seleccionar el tipo de antibióticos que se utilizara como tratamiento [17]. En microbiología de los alimentos para determinar el crecimiento o no crecimiento de *Escherichia coli* con base a variables ambientales como la temperatura entre otros [18] y en el estudio de Carlin *et. Al* en donde predicen la presencia de bacterias formadoras de esporas como *B. cereus* y *C. botulinum* en comida cocinada picante considerando temperatura, contenido de vegetales y las condiciones de almacenamiento [19].

Aunque existen numerosos trabajos en ecología que aplican el análisis bayesiano (e.g. evaluación de pesquerías [20, 21], reforestación [22] y restauración de hábitat [23]), son pocos los estudios que la empleen como parte del monitoreo y predicción de la calidad del agua en cuerpos como ríos [12], lagos [24] y estuarios [6].

En este estudio se usaron redes bayesianas para establecer la relación causal de las principales variables químicas del agua de mar en la Bahía de Banderas, Jalisco-Nayarit, México (Fig. 1), con base en mediciones mensuales realizadas en la zona costera de mayor influencia continental de marzo de 2003 a diciembre de 2005. Nuestros resultados son los primeros de una serie de trabajos sobre la oceanografía química de la bahía y permitirán hacer predicciones sobre la calidad del agua para uso recreativo (i.e. en contacto directo con el cuerpo humano), expresada en términos de la probabilidad de ocurrencia y concentración de bacterias patógenas (coliformes fecales, coliformes totales y estreptococos fecales). Debido a la importancia turística de esta región, la aplicación de nuestro modelo probabilístico tiene, además de un interés científico, una gran relevancia económica.

Materiales y métodos

De marzo de 2003 a diciembre de 2005 se realizaron 21 campañas mensuales en las estaciones de muestreo localizadas en la figura 1, a una distancia de entre 15 y 50 metros de la línea de costa, dependiendo de las condiciones del oleaje. Se establecieron tres estaciones en el centro de la bahía para obtener las muestras testigo de las condiciones marinas.

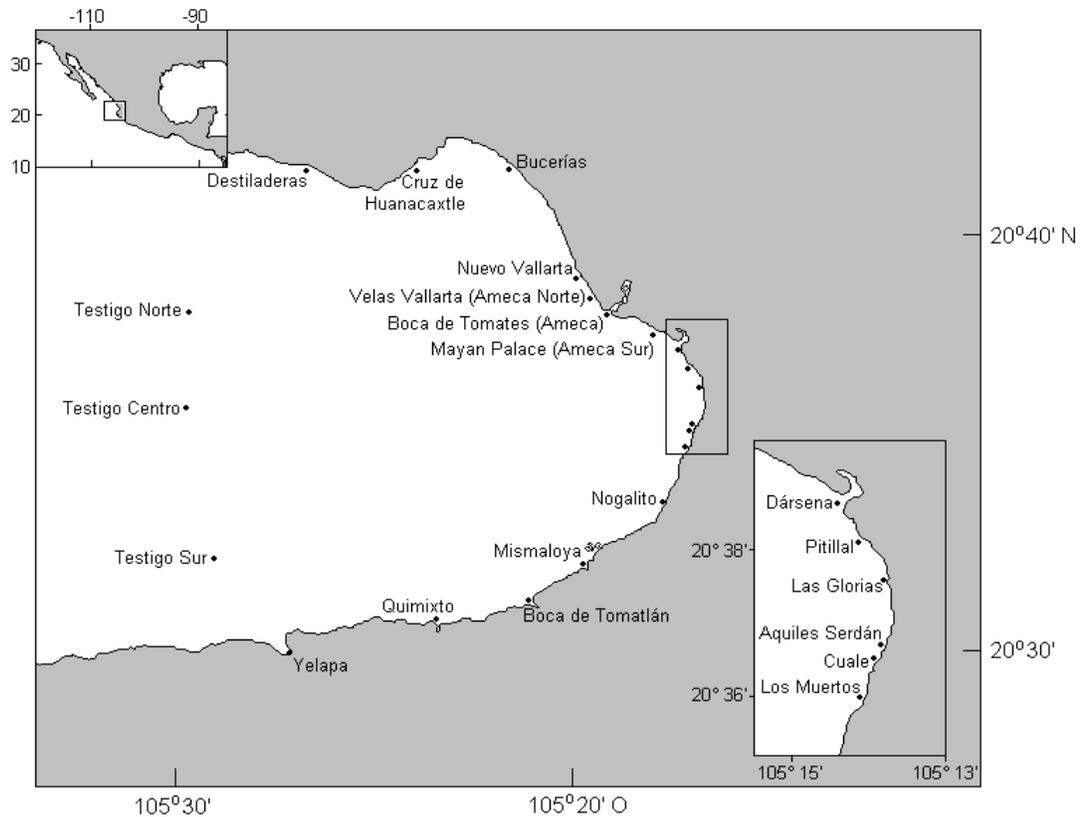


Figura 1. Estaciones de muestreo de las 21 campañas mensuales (marzo de 2003 a diciembre de 2005) en la Bahía de Banderas.

En cada campaña se obtuvieron 47 variables:

- i) Mediante mediciones *in situ*: temperatura, salinidad, oxígeno disuelto (OD) (sistema integrado YSI modelo 85) y profundidad del disco de Secchi.
- ii) Mediante toma de muestras de agua y su posterior análisis en laboratorio con los métodos descritos en la [Tabla 1](#): coliformes totales (CT), coliformes fecales (CF), estreptococos fecales (SF), sustancias reactivas al azul de metileno (SAAM), grasas y aceites (GYA), demanda bioquímica de oxígeno por autoconsumo en 5 días (DBO_5), sólidos suspendidos (SS), materia orgánica particulada (MOP), pigmentos fotosintéticos, nitritos, nitratos (NO_3), silicatos (SiO_4), fósforo reactivo soluble (PO_4), fosforo total (P) y nitrógeno amoniacal (NH_4).
- iii) Mediante cálculos: coeficiente de atenuación de la luz con la ecuación propuesta por Holmes ($C_{Atenuación} = 1.7/\text{profundidad del disco de Secchi}$) [25], nitrógeno total inorgánico (NTI) [26], fosforo residual (diferencia entre PO_4 y P), PO_4/P , NTI/P , NTI/SiO_4 , MOP/SS y los índices de calidad (ICA) de OD , DBO_5 , NH_4 , NO_3 , PO_4 , SAAM, CT, CF, GYA (metodología propuesta por EPA [27] y adaptada por la SEDUE [28]).

Tabla 1. Métodos analíticos para determinar las variables químicas y biológicas que influyen en la calidad del agua en la Bahía de Banderas.

<i>Análisis</i>	<i>Referencia</i>
Coliformes fecales, coliformes totales, estreptococos fecales, sustancias reactivas al azul de metileno y grasas y aceites.	APHA (1992) [29]
Demanda bioquímica de oxígeno por autoconsumo en 5 días. Sólidos suspendidos y materia orgánica particulada.	Rodier (1981) [30]
Pigmentos fotosintéticos.	Rodier (1981) [30], Holm-Hansen y Riemann (1978) [31]
Nitritos, nitratos y silicatos. Fósforo reactivo soluble y total. El fósforo residual se determina con la diferencia entre total y reactivo. Oxígeno disuelto.	Strickland y Parsons (1972) [32]
Nitrógeno amoniacal.	Solórzano (1969) [33]

Los datos se analizaron en escenarios de secas y lluvias, se calcularon los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90 (P_{10} a P_{90}) para cada variable y se crearon matrices con los datos incluidos en cada uno de los siguientes intervalos: ($x \leq P_{10}$), ($P_{10} < x \leq P_{20}$), ($P_{20} < x \leq P_{30}$), ($P_{30} < x \leq P_{40}$), ($P_{40} < x \leq P_{50}$), ($P_{50} < x \leq P_{60}$), ($P_{60} < x \leq P_{70}$), ($P_{70} < x \leq P_{80}$), ($P_{80} < x \leq P_{90}$) y ($x > P_{90}$). Se calculó la probabilidad conjunta [$P(A \cap B)$], la probabilidad marginal $P(B)$ y la probabilidad condicional [$P(A | B) = P(A \cap B) / P(B)$] y se probó la independencia estadística con el teorema de Bayes [$P(A | B) = P(A \cap B) / P(B)$], la sumatoria de la $p(x=1)$ para cada variable. Los eventos A, tanto en época de secas como de lluvias, fueron las matrices correspondientes a las variables coliformes totales, coliformes fecales y estreptococos fecales, y los eventos B, todas las variables restantes (Tabla 2).

Los datos de probabilidad de ocurrencia de bacterias según la fórmula de probabilidad condicional [$P(A | B) = P(A \cap B) / P(B)$] se filtraron mediante los siguientes criterios de selección: 1) Que el número de bacterias (coliformes totales, coliformes fecales o estreptococos totales) fuera mayor al P_{90} con el fin de determinar las variables y las probabilidades con las que se puede predecir la ocurrencia de que las bacterias sean altas; 2) que la probabilidad de que el número de bacterias sea mayor al P_{90} fuera igual o mayor a 0.5 y significativamente diferente a las probabilidades calculadas para los otros percentiles de las variables analizadas (para ello se usó una prueba de bondad de ajuste con un nivel de significancia de 0.05 y 9 grados de libertad); y 3) que los eventos A y B sean dependientes entre sí con base en el teorema de Bayes.

Tabla 2. Variables utilizadas en la construcción de redes bayesianas para la predicción de coliformes totales, coliformes totales y estreptococos fecales.

<i>Variable No.</i>	<i>Nombre</i>	<i>Variable No.</i>	<i>Nombre</i>
1	Temperatura	24	Estreptococos fecales (SF)
2	Salinidad	25	Grasas y aceites (GyA)
3	Oxígeno disuelto (OD)	26	Clorofila A
4	Potencial oxido-reducción (ORP)	27	Clorofila B
5	Profundidad del disco de Secchi	28	Clorofila C
6	Coefficiente de atenuación de la luz	29	Clorofila A corregida por feopigmentos
7	DBO salinidad	30	Feopigmentos
8	Demanda bioquímica de oxígeno (DBO ₅)	31	Sólidos suspendidos 0.45 µm (SS)
9	Amonio (NH ₄)	32	Sólidos suspendidos 1.2 µm (SS)
10	Nitritos (NO ₂)	33	Materia orgánica particulada (MOP)
11	Nitritos+nitratos	34	MOP/SS
12	Nitratos (NO ₃)	35	Utilización aparente de oxígeno (AOU)
13	Nitrógeno total inorgánico (NTI)	36	ICA OD
14	Fosforo reactivo soluble (PO ₄)	37	ICA DBO ₅
15	Fosforo total (P)	38	ICA NH ₄
16	Fosforo residual	39	ICA NO ₃
17	PO ₄ /P	40	ICA PO ₄
18	NTI/P	41	ICA SAAM
19	Silicatos (SiO ₄)	42	ICA C total
20	NTI/SiO ₄	43	ICA C fecal
21	Detergentes (SAAM)	44	ICA GyA
22	Coliformes totales (C total)	45	ISS
23	Coliformes fecales (C fecal)		

Resultados y discusiones

Los valores de los descriptores para cada variable: número de datos, intervalo (mínimo y máximo), promedio y desviación estándar, se muestran en la [Tabla 3](#).

Tabla 3. Descriptores para cada variable.

<i>Variable</i>	<i>Unidades</i>	<i>n</i>	<i>Intervalo</i>	<i>Promedio ± desviación estándar</i>
Temperatura	°C	428	20.73-32.58	27.81±2.75
Salinidad	ups	428	0.43-36.59	31.63±6.00
Oxígeno disuelto (OD)	mg/l	428	2.57-8.89	6.00±0.72
Potencial oxido-reducción (ORP)	V	393	(-87.00)-179.00	-15.32±57.45
Profundidad del disco de Secchi	m	325	0.07-30.00	3.76±4.45
Coefficiente de atenuación de la luz	m ⁻¹	325	0.05-24.28	1.54±2.89
Demanda bioquímica de oxígeno (DBO ₅)	mg/l	404	(-0.33)-6.80	1.21±0.93
Amonio (NH ₄)	µM	393	0.21-42.70	3.97±4.47
Nitritos (NO ₂)	µM	431	0.01-20.03	0.43±1.49
Nitratos (NO ₃)	µM	429	0-322.60	6.28±19.10
Nitrógeno total inorgánico (NTI)	µM	391	0.71-329.27	10.27±20.96
Fosforo reactivo soluble (PO ₄)	µM	430	0.05-18.43	1.18±2.15
Fosforo total (P)	µM	410	0.05-13.67	1.17±1.72
Fosforo residual	µM	404	0-12.18	0.37±1.01
Silicatos (SiO ₄)	µM	428	0.01-1167.306	83.85±150.54
Detergentes (SAAM)	mg/l	271	0-0.06	0.02±0.01
Coliformes totales (C total)	ufc/100 ml	359	1-105210.00	3174.55±9442.42
Coliformes fecales (C fecal)	ufc/100 ml	401	1-39078.00	1492.77±4592.98
Estreptococos fecales (SF)	ufc/100 ml	259	1-24640.00	246.53±1753.34
Grasas y aceites (GyA)	mg/l	353	0-660.68	6.24±36.29
Clorofila A	µg/l	277	0-37.38	2.99±3.77
Clorofila B	µg/l	277	0-51.59	1.80±4.18
Clorofila C	µg/l	277	0-140.27	4.49±11.81
Clorofila A corregida por feopigmentos	µg/l	278	0-120.15	5.44±10.10
Feopigmentos	µg/l	278	0-23.70	0.39±1.72
Sólidos suspendidos 0.45 µm (SS)	mg/l	372	0-655.00	22.08±50.09
Sólidos suspendidos 1.2 µm (SS)	mg/l	395	0.15-792.50	35.49±71.42
Materia orgánica particulada (MOP)	mg/l	372	0-135.00	4.23±8.98
Utilización aparente de oxígeno (AOU)	mg/l	429	(-2.41)-14.62	0.49±1.02

Debido a que muchas variables son diferentes entre dichas épocas, en todos los análisis que se realizaron en el presente trabajo se estratificaron los datos en secas y lluvias (Tablas 4 y 5).

Tabla 4. Descriptores para cada variable en época de secas.

<i>Variable</i>	<i>Unidades</i>	<i>n</i>	<i>Intervalo</i>	<i>Promedio ± desviación estándar</i>
Temperatura	°C	220	20.73-29.90	25.67±1.92
Salinidad	ups	220	11.61-35.34	33.40±3.81
Oxígeno disuelto (OD)	mg/l	220	2.57-8.31	6.02±0.69
Potencial oxido-reducción (ORP)	V	205	(-87)-78	-27.62±43.99
Profundidad del disco de Secchi	m	173	0.5-30	4.13±4.35
Coefficiente de atenuación de la luz	m ⁻¹	173	0.05-3.4	0.77±0.53
Demanda bioquímica de oxígeno (DBO ₅)	mg/l	210	(-0.17)-6.8	1.34±0.97
Amonio (NH ₄)	µM	205	0.21-32.82	4.57±4.60
Nitritos (NO ₂)	µM	223	0.01-20.03	0.66±1.94
Nitratos (NO ₃)	µM	222	0-50.10	2.95±5.75
Nitrógeno total inorgánico (NTI)	µM	204	0.71-73.40	7.77±9.70
Fosforo reactivo soluble (PO ₄)	µM	223	0.05-17.85	1.08±2.01
Fosforo total (P)	µM	212	0.05-5.63	0.87±0.87
Fosforo residual	µM	210	0.00-2.41	0.23±0.41
Silicatos (SiO ₄)	µM	220	0.04-607.26	45.76±91.47
Detergentes (SAAM)	mg/l	147	0.00-0.05	0.01±0.00
Coliformes totales (C total)	ufc/100 ml	172	1-20094	979.90±2673.17
Coliformes fecales (C fecal)	ufc/100 ml	208	1-17340	541.41±1822.69
Estreptococos fecales (SF)	ufc/100 ml	158	1-792	32.41±96.83
Grasas y aceites (GyA)	mg/l	187	0-97.5	5.10±11.00
Clorofila A	µg/l	147	0.21-12.02	5.57±2.14
Clorofila B	µg/l	147	0-6.72	1.20±0.96
Clorofila C	µg/l	147	0-80.43	3.06±6.62
Clorofila A corregida por feopigmentos	µg/l	147	0-19.54	3.60±3.33
Feopigmentos	µg/l	147	0-6.64	0.18±0.78
Sólidos suspendidos 0.45 µm (SS)	mg/l	206	0-51.66	12.31±6.56
Sólidos suspendidos 1.2 µm (SS)	mg/l	206	0.51-87.14	20.10±9.62
Materia orgánica particulada (MOP)	mg/l	209	0-15.88	3.04±1.82
Utilización aparente de oxígeno (AOU)	mg/l	221	(-1.54)-14.62	0.63±1.22

Tabla 5. Descriptores para cada variable en época de lluvias.

<i>Variable</i>	<i>Unidades</i>	<i>n</i>	<i>Intervalo</i>	<i>Promedio ± desviación estándar</i>
Temperatura	°C	208	32.58-24.3	30.07±1.29
Salinidad	ups	208	0.43-36.59	29.75±7.20
Oxígeno disuelto (OD)	mg/l	208	3.1-8.89	5.97±0.74
Potencial oxido-reducción (ORP)	V	188	(-72)-179	-1.89±66.79
Profundidad del disco de Secchi	m	154	0-22	3.34±4.54
Coefficiente de atenuación de la luz	m ⁻¹	152	0.07-24.28	2.41±4.03
Demanda bioquímica de oxígeno (DBO ₅)	mg/l	194	(-0.33)-6.34	1.07±0.85
Amonio (NH ₄)	µM	188	0.44-42.69	3.29±4.23
Nitritos (NO ₂)	µM	208	0-7.74	0.18±0.65
Nitratos (NO ₃)	µM	207	0-322.6	9.83±26.41
Nitrógeno total inorgánico (NTI)	µM	187	0.73-329.27	13.00±28.36
Fosforo reactivo soluble (PO ₄)	µM	207	0.04-18.43	1.27±2.29
Fosforo total (P)	µM	198	0.16-13.67	1.48±2.26
Fosforo residual	µM	194	0-12.18	0.51±1.37
Silicatos (SiO ₄)	µM	208	0.01-1167.30	124.14±185.81
Detergentes (SAAM)	mg/l	124	0.002-0.05	0.02±0.008
Coliformes totales (C total)	ufc/100 ml	187	1-105210	5193.16±12509.48
Coliformes fecales (C fecal)	ufc/100 ml	193	1-39078	2518.06±6191.11
Estreptococos fecales (SF)	ufc/100 ml	101	1-24640	581.49±2780.45
Grasas y aceites (GyA)	mg/l	166	0-660.68	7.51±51.67
Clorofila A	µg/l	130	0-37.37	3.46±4.98
Clorofila B	µg/l	130	0-51.59	2.47±5.96
Clorofila C	µg/l	130	0-140.26	6.09±15.61
Clorofila A corregida por feopigmentos	µg/l	131	0-120.15	7.49±14.02
Feopigmentos	µg/l	131	0-23.69	0.62±2.35
Sólidos suspendidos 0.45 µm (SS)	mg/l	166	0.97-655	34.19±72.94
Sólidos suspendidos 1.2 µm (SS)	mg/l	189	0.14-792.5	52.27±100.33
Materia orgánica particulada (MOP)	mg/l	165	0-135	5.74±13.20
Utilización aparente de oxígeno (AOU)	mg/l	208	-2.41-4.14	0.34±0.72

Las bacterias CT, CF y SF son considerados en conjunto como buenos indicadores de la calidad del agua, sobre todo de contaminación por descargas de aguas negras [34] y las consecuencias que estas últimas pueden implicar: hay muchos microorganismos patógenos que transmiten enfermedades por estar en contacto con agua contaminada [35]. Una gran desventaja del uso de bacterias como indicadores es que su determinación en laboratorio requiere al menos 2 días, incluyendo las pruebas microbiológicas presuntivas y confirmativas. Así, cuando se determina en laboratorio una zona presenta contaminación en sus aguas, ya no hay mucho que hacer para preservar la vida acuática ni con el turista que estuvo en contacto con estas aguas. Una ventaja de utilizar el análisis de redes bayesianas es que se puede estimar mediante ciertas variables una alta ocurrencia de bacterias indicadoras de contaminación por bacterias en tiempo real.

Una de las ventajas de haber usado el análisis bayesiano en este trabajo es que, debido a que en ésta es posible analizar los datos estratificados en percentiles, donde cada variable es dividida en 10 conjuntos de datos ordenados, se pueden estimar valores de probabilidad de ocurrencia de bacterias mayores al P_{90} que son evidentes únicamente cuando las variables predictoras están dentro de cierto intervalo, o cuando son mayores o menores que un cierto valor (cualquier percentil en este caso) y esto, en este caso, es independiente de la distribución muestral y de la escala de los datos. Este comportamiento se enmascara si el análisis estadístico considera todos los datos en conjunto, como en el caso de las correlaciones múltiples entre las variables o parámetros como el promedio. Como ejemplo de esto tenemos el caso de la variable N/SiO₄: la probabilidad de que SF > P_{90} es 0.9166 si la razón N/SiO₄ es menor al P_{10} (32.0803), es decir el 91.66% de las veces que la razón N/SiO₄ fue menor a 32.0803 (P_{10}), los estreptococos fueron mayores a 390 ufc (P_{90}). Este resultado significa que, con base únicamente en el monitoreo de los valores de N/SiO₄ en el área de estudio, podemos predecir una ocurrencia alta de SF.

Una segunda ventaja de las redes bayesianas en estudios como el nuestro es que ellas posibilitan trabajar con un número ilimitado de variables, sin importar que tengamos un intervalo amplio en el orden de magnitud de los datos. No menos importante en términos de programación, es posible llevar a cabo el análisis estadístico sin importar que las matrices de datos contengan espacios vacíos lo cual representa una gran ventaja sobre otros análisis estadísticos como el análisis de regresión.

A pesar de lo dicho anteriormente, autores como Hajmeer y Basheer [36] sugieren, luego de comparar varios métodos de tratamiento de datos, que los mejores resultados al determinar las interrelaciones entre los parámetros se obtienen combinando técnicas tradicionales, como el análisis de regresión lineal y no lineal, con el análisis de redes bayesianas.

Debido a que los aportes del continente hacia el mar se incrementan en época de lluvias, los valores del P_{90} varían significativamente entre secas y lluvias. Los valores para el número de bacterias que corresponden al P_{90} en las dos épocas del año se presentan en la [Tabla 6](#).

Tabla 6. Valores de los percentiles 90 de bacterias estratificando los datos en época de secas y lluvias.

	P₉₀ en Secas (unidades formadoras de colonias/100 ml)	P₉₀ en Lluvias (unidades formadoras de colonias/ 100 ml)
CT	2741.2	16446.4
CF	1144	6792.8
SF	74.1	390

Coliformes totales

Las variables que permiten predecir si los CT serán mayores a 2741.2 ufc (percentil 90 en secas) o mayores a 16446.4 ufc (percentil 90 en lluvias) con una probabilidad mayor o igual a 0.5 y significativa en la prueba de bondad de ajuste ($\alpha = 0.05$ y g.l.=9) se presentan en las [Tablas 7](#) y [8](#) respectivamente.

Tabla 7. Variables que permiten predecir CT mayores a 2741.2 ufc/100 ml en época de secas.

Evento B:	Evento A: CT > P ₉₀ (2741.2 ufc)
Salinidad < P ₁₀ (30.661 ups)	P(A B) = 0.6666
Profundidad del disco de Secchi < P ₁₀ (1.2 m)	P(A B) = 0.8888
Fósforo residual ≤ P ₅₀ (0)	P(A B) = 0.5
N/ SiO ₄ < P ₁₀ (0.1712)	P(A B) = 0.6666
CF > P ₉₀ (1144 ufc)	P(A B) = 0.6315

En secas, las variables que permiten predecir la ocurrencia de CF mayor al percentil 90 (2741.2 ufc) son: salinidad, profundidad del disco de Secchi, razón de nitrógeno entre SiO₄, CF y el ICA de CF.

Tabla 8. Variables que permiten predecir CT mayores a 16446.4 ufc/100 ml en época de lluvias.

Evento B:	Evento A: CT > P ₉₀ (16446.4 ufc)
Salinidad < P ₁₀ (22.55 ups)	P(A B) = 0.5555
Profundidad del disco de Secchi < P ₁₀ (0.263 m)	P(A B) = 0.5333
PO ₄ (total) > P ₉₀ (3.4297 μM)	P(A B) = 0.5294
N/P < P ₁₀ (2.6012)	P(A B) = 0.6
CF > P ₉₀ (6792.8 ufc)	P(A B) = 0.6315

En lluvias, las variables que pueden usarse para predecir la probabilidad de que el número de CF sea mayor al P₉₀ son: salinidad, profundidad del disco de Secchi, PO₄(total), razón de N/P y CF.

Debe mencionarse que una bacteria indicadora de contaminación no necesariamente es patógena, aunque sí puede estar asociada a organismos patógenos. Algo similar sucede con todos los ICA y el índice de CT, que requiere necesariamente la cantidad de coliformes totales para su determinación.

Así, La probabilidad de que los CT sean mayores al percentil 90=2741.2 ufc en secas es: i) 0.6666, si la salinidad es menor a 30.661 (P₁₀); ii) 0.8888, si la profundidad del disco de Secchi es menor a 1.2 m (P₁₀); iii) 0.5, si el fósforo residual es igual a cero; iv) 0.6666 y si la razón N/ SiO₄ es menor a 0.1712 (P₁₀).

La probabilidad de que los CT sean mayores al percentil 90=16446.4 ufc en lluvias es: i) 0.5555, si la salinidad es menor a 22.55 (P₁₀); ii) 0.5333 si la profundidad del disco de Secchi es menor a 0.263 m (P₁₀); iii) 0.5294, si la concentración de PO₄(total) es mayor a 3.4297 μM (P₉₀); y iv) 0.6, si la razón de nitrógeno entre fósforo (N/P) es menor a 2.6012.

Otro aspecto a considerar al buscar una variable que permita predecir bacterias es el número de muestreos que abarca el estudio y la propuesta. Para analizar esto se realizaron redes bayesianas considerando: i) sólo el primer muestreo de secas, ii) tomando en cuenta el primer y segundo muestreo, iii) y así sucesivamente hasta que, con los muestreos acumulados, se completaron los 11 meses de secas y los 10 de lluvias medidos en este trabajo.

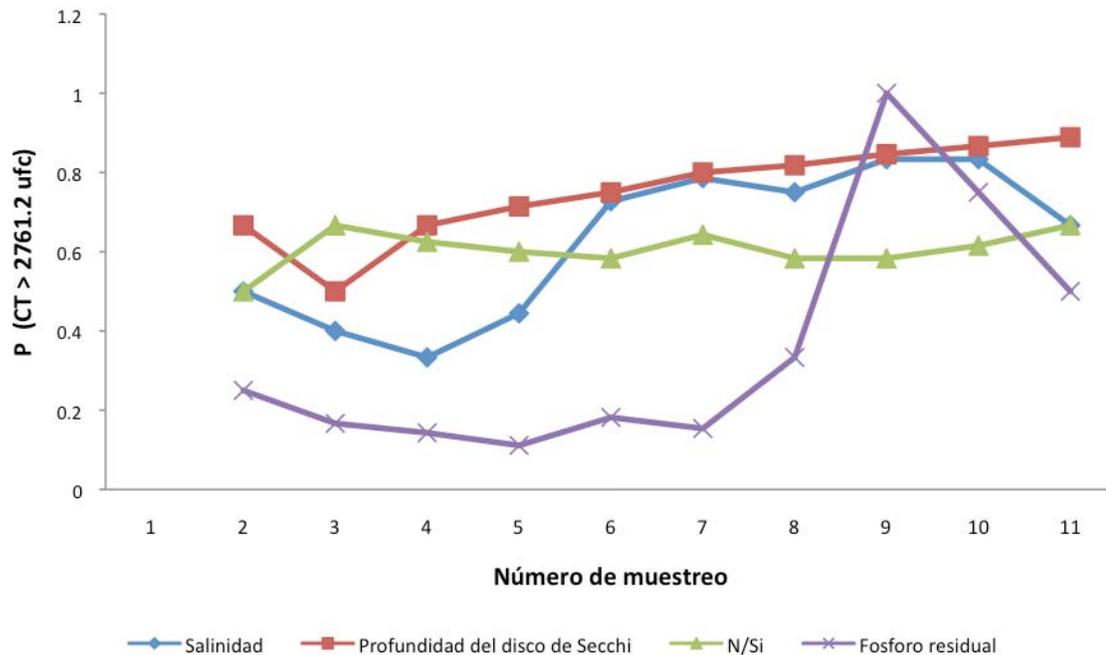


Figura 2. Distribución de probabilidad condicional $P(A|B)$ acumulada en época de secas del evento A: $CT > P_{90}$ (2761.2 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, la razón N/SiO_4 , y fósforo residual.

En las figuras 2 y 3 se muestra la variación en las predicciones condicionales de CT en época de secas y lluvias. Se realizaron pruebas de bondad de ajuste para determinar si las probabilidades de ocurrencia de bacterias variaban significativamente con el número de muestreos. En época de secas la profundidad del disco de Secchi y la razón N/SiO_4 fueron iguales significativamente desde el primer muestreo. La salinidad presentó mayor variación en la estimación de la probabilidad de ocurrencia de CT, pero se estabilizó (igualdad significativa) a partir del sexto muestreo.

La probabilidad de que las bacterias CT fueran mayores a 2741.2 (P_{90}) en secas, estimada a partir de fósforo residual, fue diferente significativamente en los muestreos 10 y 11, por lo que este último se descartó como predictor. La probabilidad de que las bacterias CT fueran mayores al P_{90} (2741.2) en secas, estimada a partir de sulfuros, fue siempre la misma ($\alpha=0.05$) a partir del octavo muestreo. Se presentó una gran variación en los valores de probabilidad condicional para la predicción de $CT > P_{90}$ con respecto al número de muestreo acumulado.

Si el estudio de predicción de CT se hubiera hecho con seis meses, en vez de 11, se hubieran obtenido los mismos resultados con salinidad, disco de Secchi y la razón N/SiO_4 . Así, estas variables fueron seleccionadas posibles predictores.

En lluvias, el PO_4 (total) predijo la ocurrencia de CT con igual valor de probabilidad significativa desde el primer muestreo hasta el décimo. Salinidad, profundidad del disco de Secchi y N/P presentaron diferencias significativas con respecto al número de muestreos; en la figura 3 se muestra la variación en los valores de probabilidad condicional.

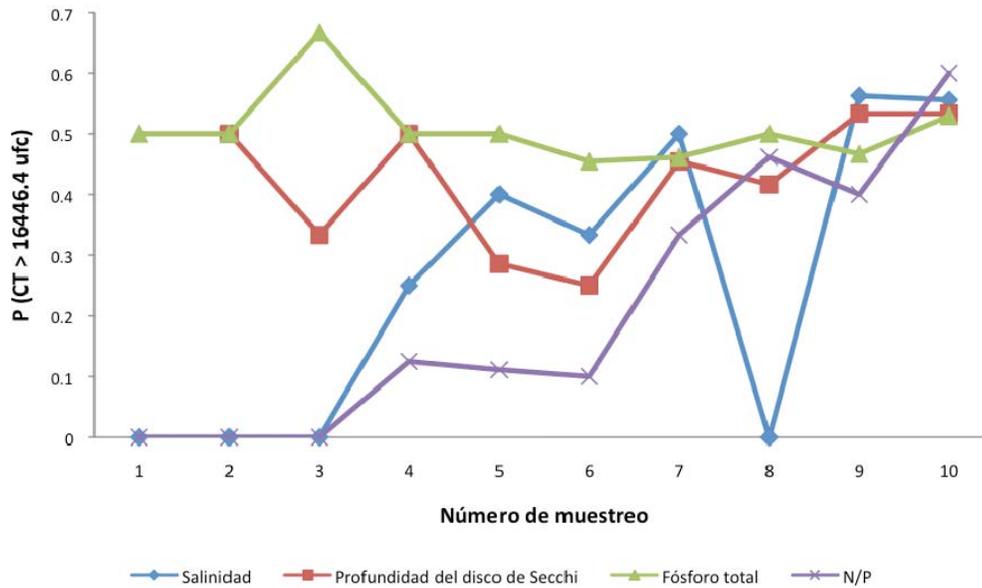


Figura 3. Distribución de probabilidad condicional $P(A|B)$ acumulada en época de lluvias del evento A: $CT > P_{90}$ (16446.4 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, PO_4 (total) y la razón N/P.

La salinidad se descartó como predictora de CT en lluvias debido a que no fue sino en el noveno muestreo que presentó igualdad significativa con respecto al décimo, en el que se consideraron todos los datos para esta estación. La profundidad del disco de Secchi desde el séptimo muestreo presenta igualdad significativa con respecto al décimo; la razón N/P, desde el octavo. Como la curva de N/P no presentó estabilidad y sí un aumento en la probabilidad condicional con respecto al muestreo, fue descartada como predictor.

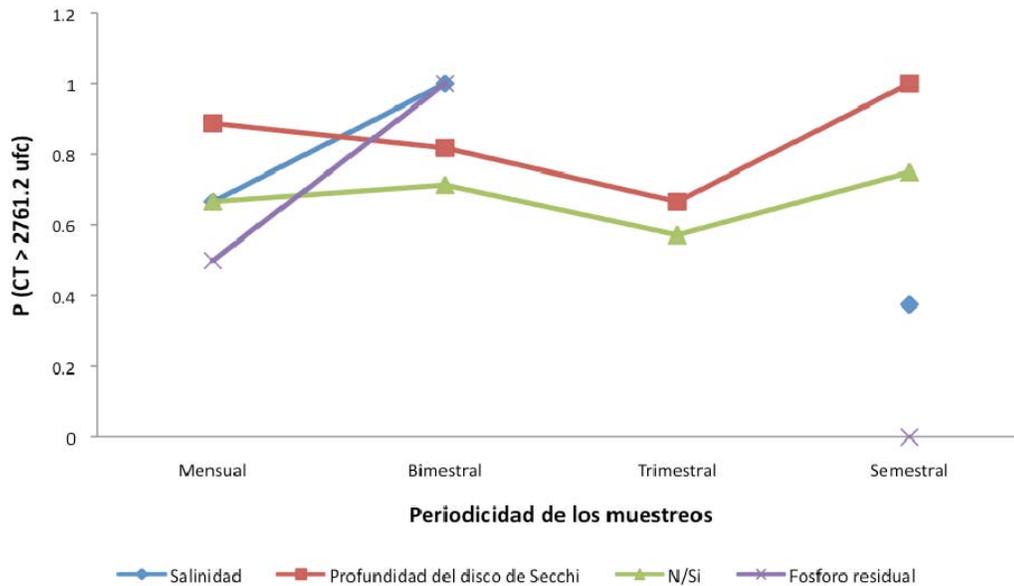
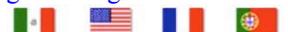


Figura 4. Distribución de probabilidad condicional $P(A|B)$ con diferente periodicidad en los muestreos del evento A: $CT > P_{90}$ (2761.2 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, la razón N/ SiO_4 y fósforo residual en época de secas.



Debido a los costos y al tiempo que se requiere para un monitoreo de calidad del agua, se calcularon las redes bayesianas considerando que además de muestrear mensualmente, como se hizo en este trabajo, se hubiera muestreado bimestral, trimestral y semestralmente. Este experimento se hizo también para determinar la validez de haber muestreado mensualmente. Se realizaron pruebas de bondad de ajuste con un nivel de significancia de 0.05 para determinar la periodicidad de muestreo con la que se obtendrían resultados similares a los correspondientes a muestreos mensuales.

En las figuras 4 y 5 se observa que el tiempo entre muestreos es directamente proporcional a la dispersión con respecto al resultado más exacto que se obtuvo (muestreos mensuales). En secas algunas variables (e.g. la profundidad del disco de Secchi y la razón N/ SiO₄), dan los mismos resultados si se muestrea mensual, bimestral, trimestral o semestralmente; esto no sucede con la salinidad, el fósforo residual o los sulfuros, que ni siquiera bimestralmente presentan los mismos resultados que en un muestreo mensual. Esto redujo las variables predictoras para CT > 2761.2 ufc (P90) a la profundidad del disco de Secchi y la razón N/ SiO₄.

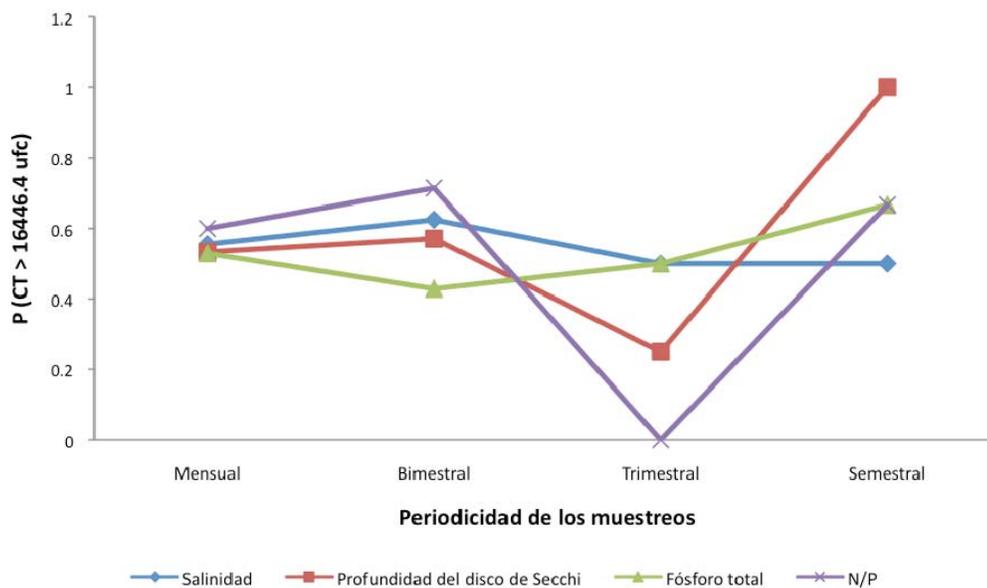


Figura 5. Distribución de probabilidad condicional $P(A|B)$ con diferente periodicidad en los muestreos del evento A: CT > P₉₀ (16446.4 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, PO₄ (total) y la razón N/P en época de lluvias.

En lluvias, las variables con resultados iguales sin importar la periodicidad de los muestreos fueron salinidad y PO₄ (total). La profundidad del disco de Secchi y la razón N/P hubieran podido medirse bimestralmente con los mismos resultados que en el muestreo mensual.

En conclusión, las mejores variables para predecir CT en la zona de mayor influencia continental en la Bahía de Banderas, en época de secas, fueron la profundidad del disco de Secchi y la razón N/ SiO₄ y, en época de lluvias, la profundidad del disco de Secchi y el PO₄ (total).

Coliformes fecales

Los valores de probabilidad de que el número de CF sea mayor al percentil 90 (i.e. 1144 ufc) y las variables con las que podría predecirse esta ocurrencia de bacterias se muestran en las tablas 9 y 10

Con base en las razones expuestas en la sección anterior para CT, para predecir la alta ocurrencia (mayor al P90=1144 ufc) de CF, en secas, se propuso OD, profundidad del disco de Secchi y la razón N/ SiO₄; y, en lluvias, salinidad, profundidad del disco de Secchi y las razones N/P y N/ SiO₄.

Tabla 9. Variables que permiten predecir CF mayores a 1144 ufc/100 ml en época de secas.

Evento B:	Evento A: CF > P ₉₀ (1144 ufc)
Salinidad < P ₁₀ (30.661 ups)	P(A B) = 0.75
OD < P ₁₀ (5.279 mg/l)	P(A B) = 0.5909
Profundidad del disco de Secchi < P ₁₀ (1.2 m)	P(A B) = 0.9444
N/ SiO ₄ < P ₁₀ (0.1712)	P(A B) = 0.7619
CT>P90 (2741.2 ufc)	P(A B)=0.7058
SF>P90 (74.1 ufc)	P(A B)=0.6

Tabla 10. Variables que permiten predecir CF mayores a 6792.8 ufc/100 ml en época de lluvias.

Evento B:	Evento A: CF > P ₉₀ (6792.8 ufc)
Salinidad < P ₁₀ (22.55 ups)	P(A B) = 0.6315
Profundidad del disco de Secchi < P ₁₀ (0.263 m)	P(A B) = 0.8666
N/P < P ₁₀ (2.6012)	P(A B) = 0.5294
N/SiO ₄ < P ₁₀ (32.0803)	P(A B) = 0.7894
CT>P90 (16446.4 ufc)	P(A B)=0.6315

Las variables propuestas como predictores de CF se estabilizaron a partir del cuarto muestreo (Fig. 6). Esto implica que, si se hubieran realizado únicamente cuatro muestreos, se habrían obtenido los mismos resultados que con los 11 muestreos efectuados en secas.

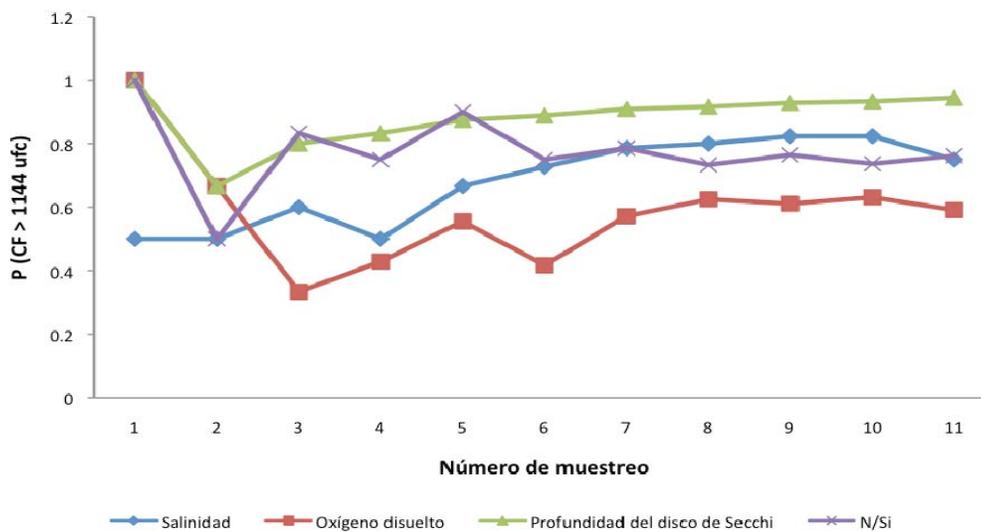


Figura 6. Distribución de probabilidad condicional P (A|B) acumulada en época de secas del evento A: CF > P₉₀ (1144 ufc/100 ml) y los eventos B: salinidad, OD, profundidad del disco de Secchi y la razón N/ SiO₄.



En el caso de lluvias se descartó N/P como predictor debido a que no fue sino hasta el octavo muestreo que presentó una igualdad significativa con el resultado obtenido en el total de los muestreos ($P = 0.5294$). Su comportamiento indica que, si se hubieran realizado mas campañas de muestreo, con N/P menor a 2.6012 (P_{10}) la probabilidad de ocurrencia de coliformes >6792.8 (P_{90}) sería mayor. Por consiguiente, se requiere hacer más muestreos para poder incluir a N/P como predictor de CF.

En lluvias los resultados de salinidad, profundidad del disco de Secchi y N/ SiO₄ fueron significativamente iguales a los calculados desde el segundo muestreo para el total de los muestreos (Fig. 7). Esto no ocurrió con la razón N/P, cuyas probabilidades de ocurrencia de CF $> P_{90}$ fueron significativamente iguales a los datos totales (10 muestreos) desde el octavo muestreo.

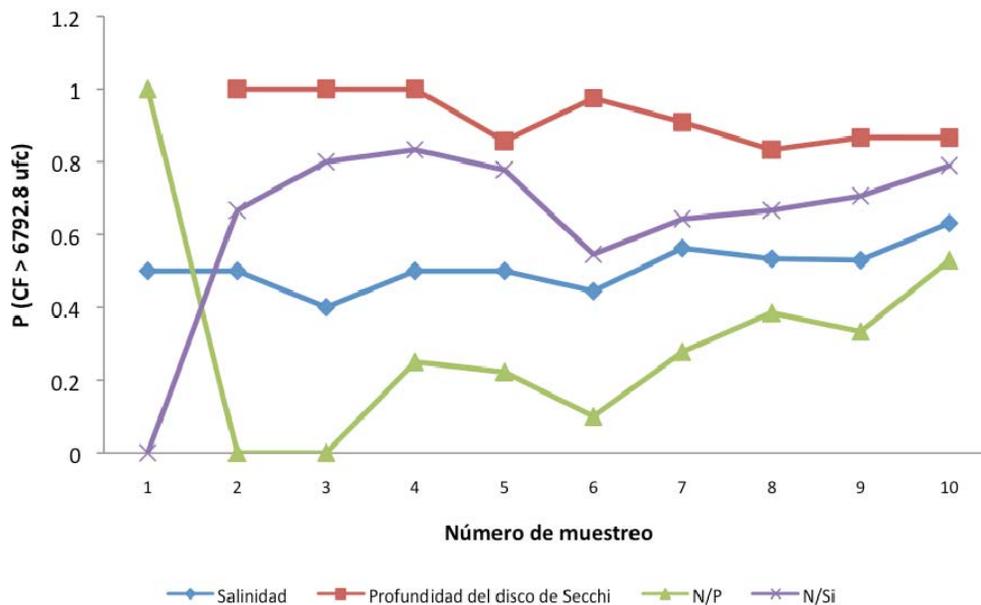


Figura 7. Distribución de probabilidad condicional $P(A|B)$ acumulada en época de lluvias del evento A: CF $> P_{90}$ (6792.8 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, la razón N/P y la razón N/ SiO₄.



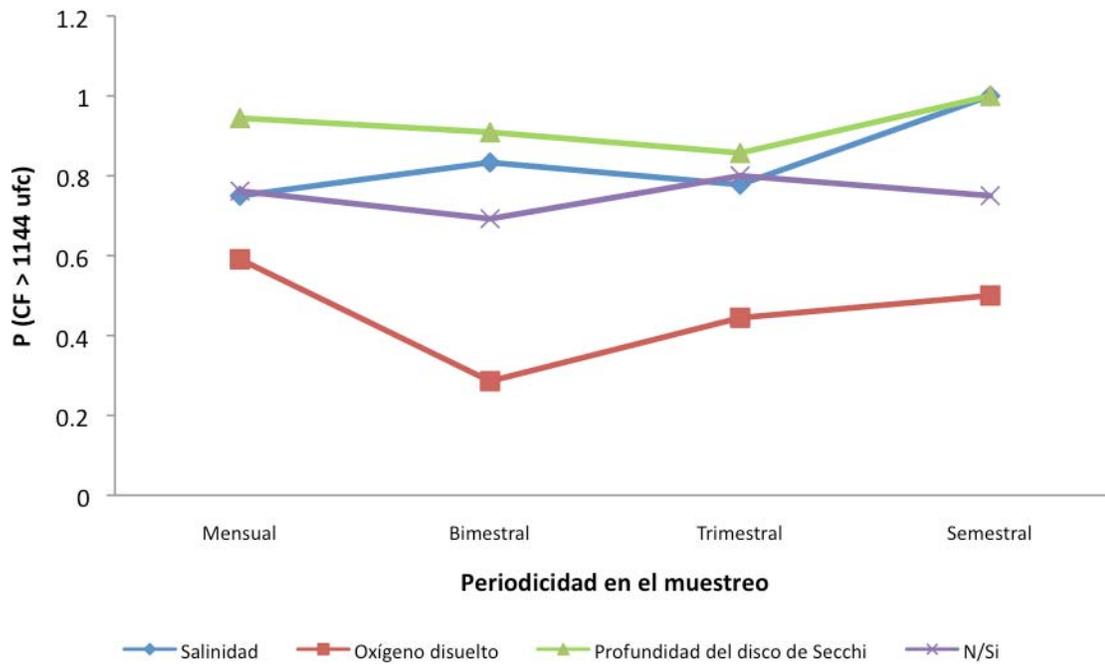


Figura 8. Distribución de probabilidad condicional $P(A|B)$ con diferente periodicidad en los muestreos del evento A: $CF > P_{90}$ (1144 ufc/100 ml) y los eventos B; salinidad, OD, profundidad del disco de Secchi y la razón N/SiO_4 en época de secas.

Con respecto a la periodicidad del muestreo, en secas las variables salinidad, profundidad del disco de Secchi y N/SiO_4 pudieron ser medidas bimestral, trimestral e incluso semestralmente, y obtenerse las mismas probabilidades para predecir que los coliformes fueran mayores a 1144 ufc (P_{90}). Este no fue el caso del OD, en el que la probabilidad de ocurrencia de CF mayor a P_{90} se subestima si no se realizan los muestreos mensualmente (Fig. 8) y, además, hay diferencias significativas entre muestrear mensual y bimestralmente.

En la figura 9 se muestra la probabilidad de que el número de CF sean mayores a 6792.8 (P_{90}) en lluvias con respecto a la periodicidad del muestreo. Los resultados de profundidad del disco de Secchi y N/SiO_4 fueron iguales sin importar la periodicidad del muestreo. Las estimaciones con salinidad y N/P se pudieron haber realizado bimestralmente con los mismos resultados significativos ($\alpha = 0.05$).

En conclusión, las variables propuestas para predecir la probabilidad de que los CF sean mayores al percentil 90 (1144 ufc) en época de secas fueron: salinidad, profundidad del disco de Secchi y la razón N/SiO_4 , con una probabilidad de: i) 0.75, si la salinidad es menor a 30.661 (P_{10}); ii) 0.9444, si la profundidad del disco de Secchi es menor a 1.2 m (P_{10}); y iii) 0.7619, si la razón N/SiO_4 es menor a 0.1712 (P_{10}).



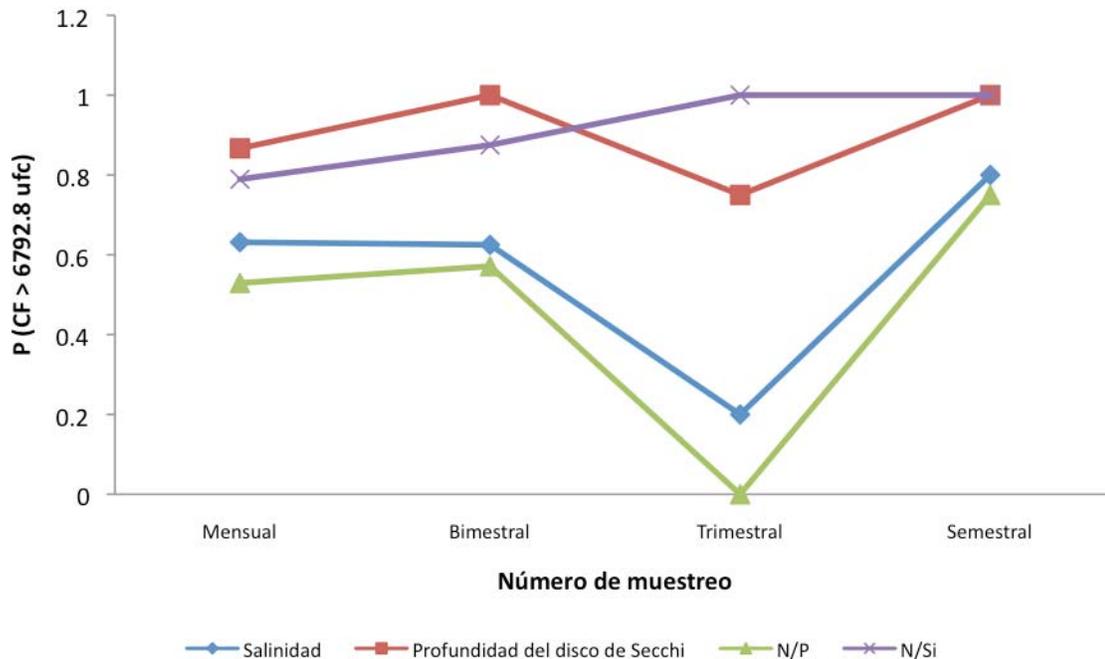


Figura 9. Distribución de probabilidad condicional P (A|B) con diferente periodicidad en los muestreos del evento A: CF > P₉₀ (6792.8 ufc/100 ml) y los eventos B: salinidad, profundidad del disco de Secchi, la razón N/P y la razón N/ SiO₄ en época de lluvias.

Las variables propuestas para predecir la probabilidad de ocurrencia de un número de CF mayores al percentil 90 (6792.8 ufc) en época de lluvias fueron: salinidad, profundidad del disco de Secchi y la razón N/ SiO₄. La probabilidad de que el número de CF fuera mayor a 6792.8 ufc en lluvias es: i) 0.6315, si la salinidad es menor a 22.55 (P₁₀); ii) 0.8666, si la profundidad del disco de Secchi es menor a 0.263 m (P₁₀); y iii) 0.7894, si la razón N/ SiO₄ es menor a 32.0803 (P₁₀).

Estreptococos fecales

Las variables que permiten predecir la ocurrencia de SF > P₉₀ (74.1 ufc) en secas y mayores al P₉₀ (390 ufc) en lluvias con una probabilidad mayor o igual a 0.5 se muestran en las tablas 11 y 12.

Tabla 11. Variables que permiten predecir SF mayores a 74.1 ufc/100 ml en época de secas.

Evento B:	Evento A: SF > P ₉₀ (74.1 ufc)
Temperatura < P ₁₀ (23.114 °C)	P(A B) = 0.6923
P ₈₀ (35.06 ups) < Salinidad ≤ P ₉₀ (35.18 ups)	P(A B) = 0.5
OD < P ₁₀ (5.279 mg/l)	P(A B) = 0.6153
Profundidad del disco de Secchi < P ₁₀ (1.2 m)	P(A B) = 0.9411
DBO < P ₁₀ (0.19 mg/l)	P(A B) = 0.6153
N/P < P ₁₀ (3.7575)	P(A B) = 0.5798
N/SiO ₄ < P ₁₀ (0.1712)	P(A B) = 0.9285
A < P ₁₀ (0.6890 µg/l)	P(A B) = 0.5714
B < P ₁₀ (0.2906 µg/l)	P(A B) = 0.5714
MOP/SS < P ₁₀ (0.0977)	P(A B)=0.6428



Se analizó la probabilidad de que el número de SF fuera mayor al P_{90} , tanto en época de secas como de lluvias, con base en la periodicidad del muestreo, y se utilizó como criterio para eliminar algunas variables el hecho de que las probabilidades obtenidas al haber muestreado mensualmente fueran diferentes haberlo hecho bimestralmente. Aplicando este criterio se eliminan, como predictoras, las variables clorofila A, clorofila B y salinidad en secas y SAAM y MOP/SS en lluvias.

Tabla 12. Variables que permiten predecir SF mayores a 390 ufc/100 ml en época de lluvias.

<i>Evento B:</i>	<i>Evento A: SF > P90 (390 ufc)</i>
Salinidad < P_{10} (22.55 ups)	$P(A B) = 0.7894$
Profundidad del disco de Secchi < P_{10} (0.263 m)	$P(A B) = 0.875$
$NO_2 > P_{90}$ (0.2884 μ M)	$P(A B) = 0.75$
$PO_4 > P_{90}$ (3.2883 μ M)	$P(A B) = 0.8888$
$P > P_{90}$ (3.4297 μ M)	$P(A B) = 0.8888$
Fósforo residual > P_{90} (1.0681 μ M)	$P(A B) = 0.6$
$N/P < P_{10}$ (2.6012)	$P(A B) = 0.7058$
$N/SiO_4 < P_{10}$ (32.0803)	$P(A B) = 0.9166$
P_{70} (0.026mg/l) < SAAM $\leq P_{80}$ (0.03mg/l)	$P(A B) = 1$
$GYA \leq P_{20}$ (0)	$P(A B) = 0.6666$
$C > P_{90}$ (9.959 μ g/l)	$P(A B) = 1$
A corregida > P_{90} (17.8 μ g/l)	$P(A B) = 1$
$MOP/SS < P_{10}$ (0.25 mg/l)	$P(A B) = 0.8333$

Si los muestreos se hubieran hecho semestralmente no se habría obtenido ningún resultado ni se tendrían variables predictoras para $SF > P90$ (74.1 ufc) en la época de secas (Fig. 10). Esto no sucede para la época de lluvias, en la que sí se obtienen probabilidades de ocurrencia de coliformes mayor a P_{90} (390 ufc) para la mayoría de las variables que pasaron los primeros criterios de filtrado.

En la figura 11 se observa la diferencia significativa entre las predicciones para la ocurrencia de SF mayores al percentil 90 en lluvias (390 ufc) a partir de GYA si el muestreo hubiera sido bimestralmente ($P(A/B) = 0$) con respecto a un muestreo mensual ($P(A/B) = 0.6666$). Dado que, como en casos anteriores, este es un criterio para no considerar a una variable como predictora, se eliminó GYA para predecir la ocurrencia de un número de SF mayor al percentil noventa en época de lluvias.

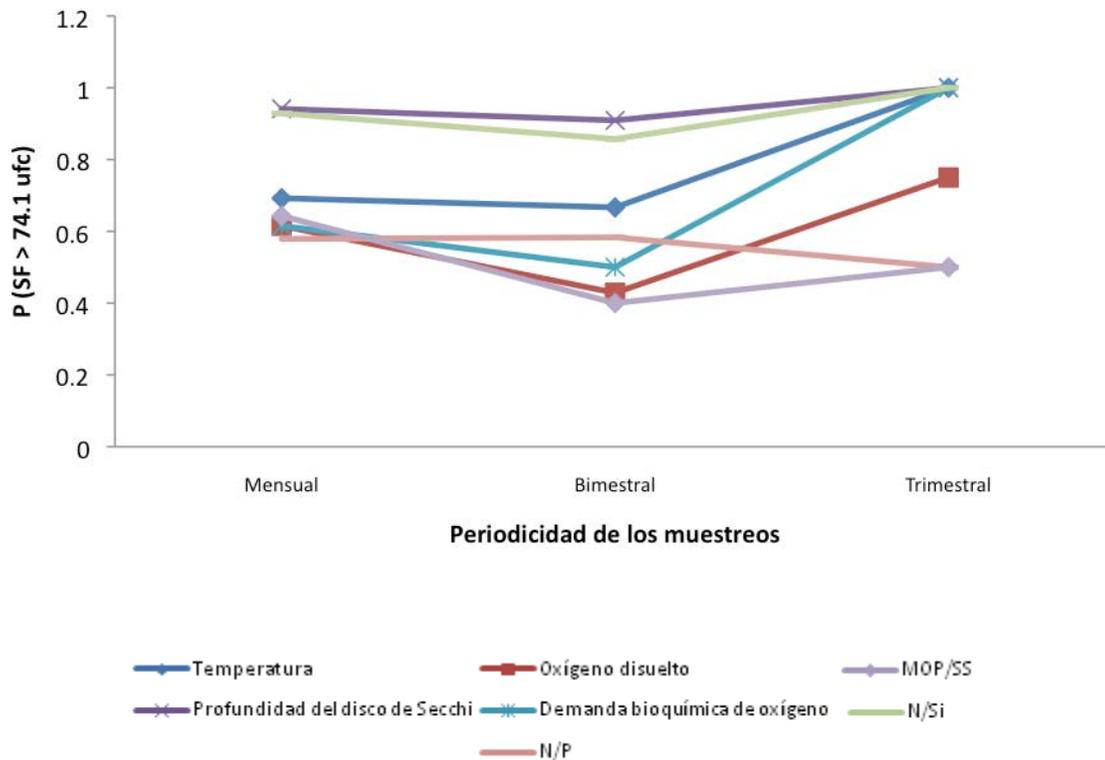


Figura 10. Distribución de probabilidad condicional P (A|B) con diferente periodicidad en los muestreos del evento A: SF > P₉₀ (74.1 ufc/100 ml) y los eventos B: temperatura, profundidad del disco de Secchi, la razón MOP/SS, OD, DBO₅, la razón N/P y la razón N/SiO₄ en época de secas.

La **Tabla 13** muestra los valores de probabilidad de que SF > P₉₀=74.1 ufc a partir de la probabilidad condicional de las variables temperatura, OD, profundidad del disco de Secchi, DBO, N/P, N/SiO₄ y MOP/SS tomando en cuenta desde únicamente un muestreo hasta todos los datos de secas.

Algunas variables (e.g. profundidad del disco de Secchi y N/SiO₄) desde el primer muestreo tuvieron una probabilidad de ocurrencia de SF > P₉₀ significativamente igual al último muestreo acumulado de secas (n=11). Su estabilidad en la predicción con respecto al tiempo, las hace considerarse como predictoras.

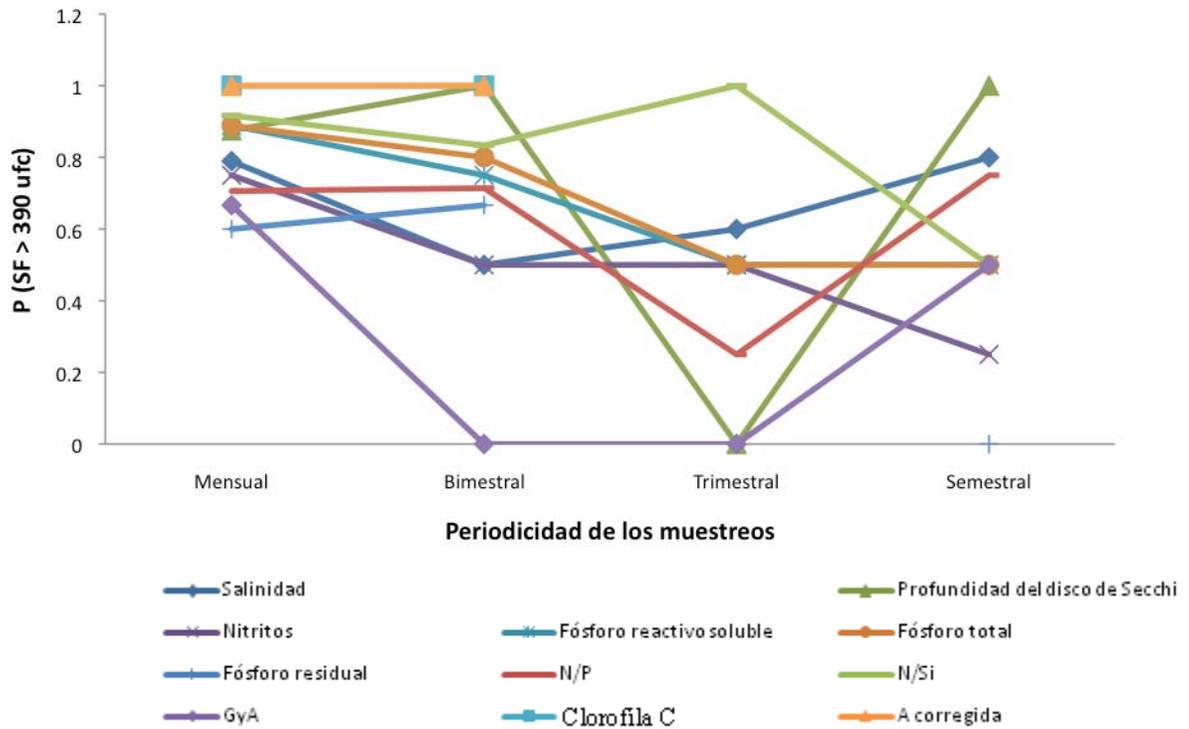
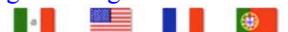


Figura 11. Distribución de probabilidad condicional P(A|B) con diferente periodicidad en los muestreos del evento A: SF > P₉₀ (390 ufc/100 ml) y los eventos B; salinidad, NO₂, fósforo residual, GYA, PO_{4(s)}, la razón N/P, clorofila C, profundidad del disco de Secchi, PO_{4(total)}, la razón N/SiO₄ y clorofila A corregida por feopigmentos en época de lluvias.

Tabla 13. Muestreo acumulado en el cual se obtiene una igualdad significativa (prueba de bondad de ajuste con α=0.05) con respecto al total de los muestreos realizados en secas (n=11).

Variable	
Temperatura	6
OD	5
DBO	5
DQO	5
DBO/DQO	5
N/P	4
MOP/SS	6

Con todas las otras variables se hubieran obtenido los mismos resultados de probabilidad de que SF > P₉₀ en secas si el muestreo hubiera durado seis meses de secas, por lo que se consideraron suficientemente estables como para proponerlas como predictoras.



La probabilidad de que los estreptococos fueran mayor al P_{90} (74.1 ufc) en época de secas estimada es de: i) 0.6923 si la temperatura es menor o igual que 23.114°C (P_{10}), ii) 0.6153 si el oxígeno disuelto es menor o igual que 5.279 mg/l (P_{10}), iii) 0.9411 si la profundidad del disco de Secchi es menor o igual que 1.2 m (P_{10}), iv) 0.6153 si la DBO es menor o igual que 0.19 mg/l (P_{10}), v) 0.5798 si N/P es menor o igual que 3.7575 (P_{10}), vi) 0.9285 si N/SiO₄ es menor o igual que 0.1712 (P_{10}) y vi) 0.6428 si MOP/SS es menor o igual que 0.0977 (P_{10}) (Fig. 12).

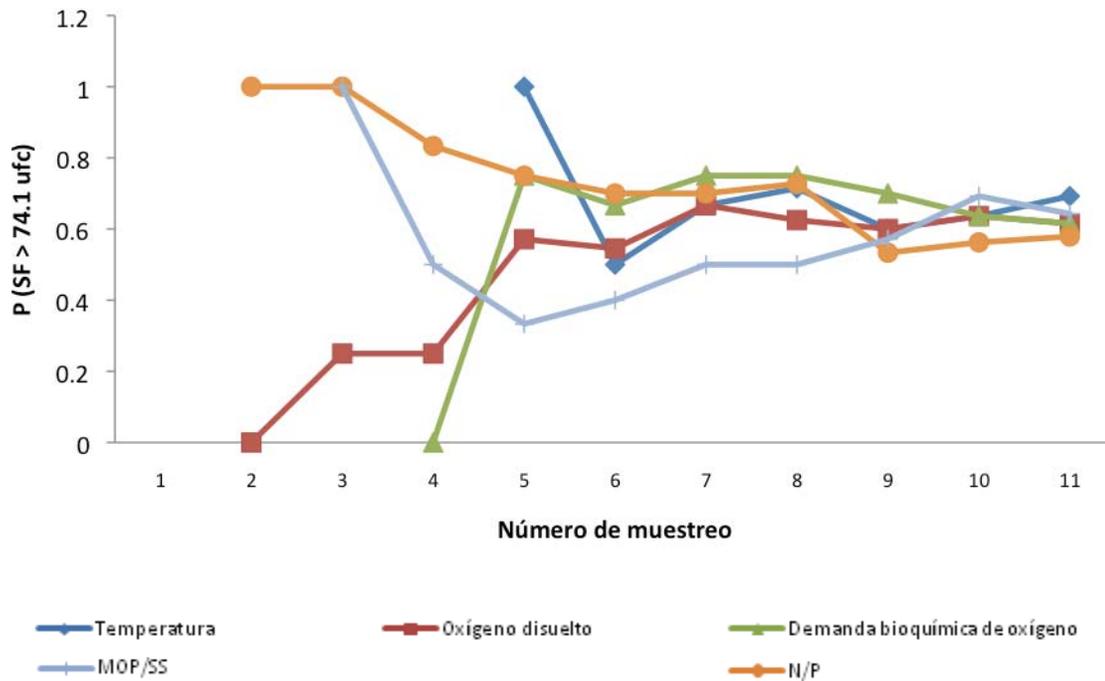
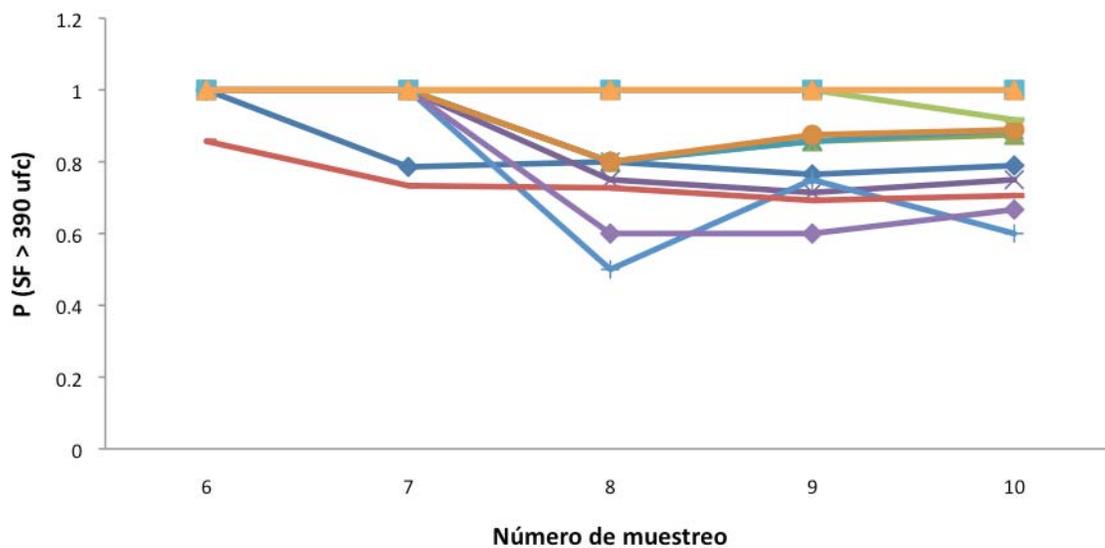


Figura 12. Distribución de probabilidad condicional $P(A|B)$ acumulada en época de secas del evento A: $SF > P_{90}$ (74.1 ufc/100 ml) y los eventos B: temperatura, la razón MOP/SS, OD, DBO₅ y la razón N/P.



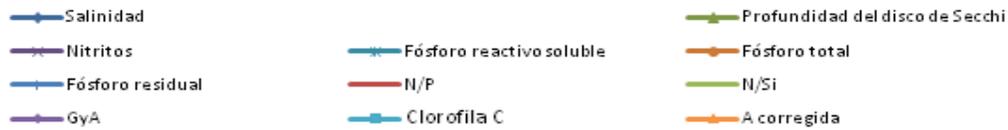


Figura 13. Distribución de probabilidad condicional P(A|B) acumulada en época de lluvias del evento A: SF > P₉₀ (390 ufc/100 ml) y los eventos B: salinidad, NO₂, fósforo residual, GYA, PO_{4(s)}, la razón N/P, clorofila C, profundidad del disco de Secchi, PO_{4(total)}, la razón N/SiO₄ y clorofila A corregida por feopigmentos.

El análisis de las probabilidades condicionales de que SF >P₉₀ (390 ufc) en lluvias se hizo de manera acumulada considerando un muestreo, luego dos y así sucesivamente hasta considerar la totalidad de los 10 muestreos correspondientes a la época de lluvias (Fig. 13).

Se eliminaron como predictoras las variables fósforo residual, NO₂ y GYA debido a que los resultados de probabilidad condicional para SF > P₉₀ son iguales significativamente al total de los datos hasta los muestreos 8, 7 y 8 respectivamente.

La probabilidad de que los SF sean mayor al P₉₀ (390 ufc) en época de lluvias es de: i) 0.7894 si la salinidad es menor o igual que 22.55 (P₁₀), ii) 0.875 si la profundidad del disco de Secchi es menor o igual que 0.263 m (P₁₀), iii) 0.8888 si PO_{4(s)} es mayor a 3.2883 μM (P₉₀), iv) 0.8888 si el PO_{4(total)} es mayor a 3.4297 μM (P₉₀), v) 0.7058 si la razón N/P es menor o igual que 26012 (P₁₀), vi) 0.9166 si la razón N/ SiO₄ es menor o igual que 320803 (P₁₀), vii) 1 si la clorofila C es mayor a 9.959 μg/l (P₉₀) y viii) 1 si la clorofila corregida por feopigmentos es mayor a 17.8 μg/l (P₉₀).

Tabla 14. Resumen de las variables predictoras que permiten predecir la ocurrencia de coliformes totales, coliformes fecales y estreptococos fecales en época de secas y lluvias.

Época	Evento A:	Evento B: Variables predictoras	[P(A B)=P(A∩B)/P(B)]
Secas	CT > P ₉₀ (2741.2 ufc/100ml)	Profundidad del disco de Secchi < P ₁₀ (1.2 m)	0.8888
		N/ SiO ₄ < P ₁₀ (0.1712)	0.6666
	CF > P ₉₀ (1144 ufc/100ml)	Salinidad < P ₁₀ (30.661 ups)	0.75
		Profundidad del disco de Secchi < P ₁₀ (1.2 m)	0.9444
	SF > P ₉₀ (74.1 ufc)	N/ SiO ₄ < P ₁₀ (0.1712)	0.7619
		Temperatura < P ₁₀ (23.114 °C)	0.6923
		OD < P ₁₀ (5.279 mg/l)	0.6153
		Profundidad del disco de Secchi < P ₁₀ (1.2 m)	0.9411
		DBO < P ₁₀ (0.19 mg/l)	0.6153
		N/P < P ₁₀ (3.7575)	0.5798
Lluvias	CT > P ₉₀ (16446.4 ufc)	N/SiO ₄ < P ₁₀ (0.1712)	0.9285
		MOP/SS < P ₁₀ (0.0977)	0.6428
	CF > P ₉₀ (6792.8 ufc)	Profundidad del disco de Secchi < P ₁₀ (0.263 m)	0.5333
		PO _{4(total)} > P ₉₀ (3.4297 μM)	0.5294
	SF > P ₉₀ (390 ufc)	Salinidad < P ₁₀ (22.55 ups)	0.6315
		Profundidad del disco de Secchi < P ₁₀ (0.263 m)	0.8666
		N/SiO ₄ < P ₁₀ (32.0803)	0.7894
		Salinidad < P ₁₀ (22.55 ups)	0.7894
		Profundidad del disco de Secchi < P ₁₀ (0.263 m)	0.875
		PO ₄ > P ₉₀ (3.2883 μM)	0.8888
		P > P ₉₀ (3.4297 μM)	0.8888
		N/P < P ₁₀ (2.6012)	0.7058
		N/SiO ₄ < P ₁₀ (32.0803)	0.9166
		C > P ₉₀ (9.959 μg/l)	1
A corregida > P ₉₀ (17.8 μg/l)		1	



Aunque en este trabajo se utilizaron los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90, se pueden emplear en lugar de ellos valores de relevancia económica, como son los límites permisibles establecidos por la legislación mexicana. En el caso de los coliformes totales, por ejemplo, donde se tiene un valor de 200 ufc como número más probable (NMP) por cada 100 ml de agua para propósitos de nado y recreacionales [37]), mediante redes bayesianas se pueden determinar aquellas variables que nos permitan establecer la probabilidad de que estas bacterias no excedan dicho valor. Aunque esto se puede hacer con otras distribuciones, el análisis bayesiano tiene la ventaja de que: i) no importa la distribución de las variables, ii) no importa la escala en la que están las variables y iii) permite usar matrices con huecos.

Mención aparte merece el caso de la variable profundidad del disco de Secchi, donde es el valor de ésta, mas no el coeficiente de atenuación de la luz que se calcula a partir de ella, el que sirve para predecir CT en secas, CF en secas y lluvias y SF en secas. La profundidad del disco de Secchi se obtiene en campo de una manera muy rápida y la inversión económica es la misma independientemente del número de lecturas realizadas. Por la cantidad de turistas que reciben anualmente, 175 millones, la costa de Estados Unidos desde Santa Barbara a San Diego es una de las más monitoreadas y se invierten anualmente más de 3 millones de dólares para evaluar la calidad microbiológica del agua [38]. En países que no tienen un presupuesto tan grande para monitorear sus playas, la herramienta de redes bayesianas puede ser aplicada para encontrar variables que permitan predecir una ocurrencia alta de bacterias (CT, CF y SF). Una ventaja no menor a la económica en el uso de profundidad de disco de Secchi es que la predicción se realiza en tiempo real, lo que permite restringir el uso de playas contaminadas cuando la contaminación está presente. Esto es imposible con el procedimiento actual recomendado por la APHA [29] que requiere hasta 48 horas después de tomada la muestra para la prueba confirmativa de coliformes fecales.

Aunque en este estudio se determinaron únicamente las bacterias CF, CT y SF, son muchos los organismos patógenos asociados a aguas negras que, de no recibir éstas el tratamiento necesario, podrían convertirse en un problema potencial de salud pública, ya que muchos sobreviven durante un tiempo largo en el medio ambiente marino [39], por ejemplo: las bacterias: *Salmonella* spp., *Shigella* spp., *Vibrio* spp., *Mycobacterium* spp., *Bacillus anthracis*, *Clostridium perfringens*, *Yersinia* spp., *Campylobacter* spp., *Pseudomonas* spp., *Leptospira* spp., *Listeria monocytogenes*, *Escherichia coli* y *Clostridium botulinum*); los virus: poliovirus, coxsackie A y B, echovirus, adenovirus, reovirus, parvovirus, rotavirus, hepatitis A, norwalk y virus gastroentéricos relacionados; los protozoarios: *Entamoeba histolytica*, *Acanthamoeba* spp. y *Giardia* spp.; y los helmintos: *Echinococcus granulosus*, *Hymenolepis nana*, *Taenia saginata*, *Fasciola hepatica*, *Ascaris lumbricoides*, *Enterobius vermicularis*, *Strongyloides* spp., *Trichuris trichiura*, *Toxocara canis* y *Trichostrongylus* spp. [40].

Si bien De Santa Olalla y colaboradores [11] utilizaron las redes bayesianas como herramienta de gestión hídrica de una gran superficie (por ejemplo un acuífero, una cuenca hidrográfica o una parte de ella), en oceanografía costera existe solamente un estudio relacionado con el desarrollo de un modelo de redes bayesianas para aguas superficiales contaminadas con bacterias [41]; sus autores exponen las consideraciones que deben hacerse al crear modelos de este tipo, pero no realizan mediciones de ningún tipo. A la fecha, nuestro trabajo sobre predicción de CT, CF y SF en la zona de mayor influencia continental es único en su tipo.

Conclusiones

El análisis bayesiano tiene la ventaja, sobre otros análisis estadísticos, de que: i) no importa la distribución de las variables, ii) no importa la escala en la que están las variables y iii) permite usar matrices con espacios vacíos (i.e., “huecos”).

Con base en la periodicidad del muestreo, la duración total del estudio y la probabilidad de predicción, las mejores variables para predecir coliformes totales en la zona de mayor influencia continental en la Bahía de Banderas, en época de secas, fueron la profundidad del disco de Secchi y la razón N/SiO₄ y, en época de lluvias el fósforo total y la profundidad del disco de Secchi. Para predecir coliformes fecales tanto en época de secas como en época de lluvias fueron la salinidad, la profundidad del disco de Secchi y la razón N/SiO₄. Temperatura, oxígeno disuelto, profundidad del disco de Secchi, demanda biológica de oxígeno, N/P, N/SiO₄ y MOP/SS fueron elegidas como variables predictoras de estreptococos fecales en época de secas, y salinidad, profundidad del disco de Secchi, fosforo reactivo soluble, fosforo total, la razón N/P, la razón N/SiO₄, clorofila C y clorofila A corregida por feopigmentos en época de lluvias.

Agradecimientos

A los integrantes del proyecto “Calidad del agua en la Bahía de Banderas, Jalisco” y que contribuyeron a la obtención de los datos de laboratorio y campo. A Mario Dueñas por su apoyo durante la programación de los algoritmos en MATLAB.

Literatura Citada

1. Pielke, R. A. 2001. The role of models in prediction for decision. *in* Proceedings of the Cary Conference IX.
2. Norton, B. G. 1998. Improving ecological communication: the role of ecologists in environmental policy formation. *Ecological Applications* 8:350-364.
3. Lermusiaux, P.F.J. 2006. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J. of Computational Phys.* 217, 176-199.
4. Fitz, H. C., E. B. DeBellevue, R. Costanza, R. Boumans, T. Maxwell, L. Wainger, and F. H. Sklar. 1996. Development of a general ecosystem model for a range of scales and ecosystems. *Ecological Modelling* 88:263-295.
5. Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* 73:1943-1967.
6. Borsuk, Mark E., Craig A. Stow, and Kenneth H. Reckhow. 2002. An integrated approach to TMDL development for the Neuse River estuary using a Bayesian probability network model (Neu-BERN). *Journal Water Resources Planning and Management*.
7. Sooroshian, S. and Dracup, J.A. 1980. Stochastic parameter estimation procedures for hydrologic-rainfall-runoff models. Correlated and heteroscedastic error cases. *Water Resources Research* 16(2), 430-442.
8. Berger, J.O. 1985. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, Berlin, Germany.
9. Lindley, D.V. 1985. *Making decisions*. Second edition. Wiley, New York, USA.
10. Pearl, J. 1999. Graphs, Structural Models, and Causality. Pages 95-138 *in* C. Glymour and G.F. Cooper, editors. *Computation, Causation & Discovery*. AAAI Press, Menlo Park, CA.

11. F Martin De Santa Olalla Manas, Francisco José Martín de Santa Olalla, Prudencio López Fuster, Alfonso Calera Belmonte. 2005. *Agua y Agronomía*. Mundi-Prensa Libros. 606 pp.
12. Reckhow, K.H. 1999. Water quality prediction and probability network models. *Can. J. Fish. Aquat. Sci.* 56(7): 1150–1158.
13. Simonovic, S. P. and Fahmy, H. (1999). A new modeling approach for water resources policy analysis. *Water Resources Research*, 35(1), 295-304.
14. Wolfson, L.J., J.B. Kadane and J.S. Mitchell. 1996. Bayesian environmental policy decisions: two case studies. *Ecological Applications* 6(4), 1056-1066.
15. Marcelo Fiszman, Wendy W. Chapman, Dominik Aronsky, R. Scott Evans, and Peter J. Haug. 2000. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *Journal of the American Medical Informatics Association* 7:593-604.
16. S. Andreassen, C. Riekehr, B. Kristensen, H. Schønheyder y L. Leibovici. 2009. Using probabilistic and decision–theoretic methods in treatment and prognosis modeling. *Artificial Intelligence in Medicine*, Volume 15, Issue 2, Pages 121-134.
17. Lucas Peter. 2004. Bayesian analysis, paterna analysis, and data mining in health care. *Current Opinion in Critical Care*: October 2000, Volume 10, Issue 5, pp 399-403.
18. M. N. Hajmeer and I. A. Basheer. 2003. A hybrid Bayesian–neural network approach for probabilistic modeling of bacterial growth/no-growth interface. *International Journal of Food Microbiology*. Volume 82, Issue 3, 15 Mayo 2003, Pages 233-243.
19. Frédéric Carlin, Hélène Girardin, Michael W. Peck, Sandra C. Stringer, Gary C. Barker, Antonio Martínez, Aurea Fernandez, Pablo Fernandez, William M. Waites, Sara Movahedi, Frans van Leusden, Maarten Nauta, Roy Moezelaar, Manuela Del Torre and Sonia Litman. 2000. Research on factors allowing a risk assessment of spore-forming pathogenic bacteria in cooked chilled foods containing vegetables: a FAIR collaborative Project. *International Journal of food microbiology*. Volume 60, issues 2-3, 25 september 2000, pages 117-135
20. Lee, D. C., and B. E. Rieman. 1997. Population viability assessment of salmonids by using probabilistic networks. *North American Journal of Fisheries Management* 17:1144-1157.
21. Kuikka, S., M. Hilden, H. Gislason, S. Hansson, H. Sparholt, and O. Varis. 1999. Modeling environmentally driven uncertainties in Baltic cod (*Gadus morhua*) management by Bayesian influence diagrams. *Canadian Journal of Fisheries and Aquatic Sciences* 56:629-641.
22. Haas, T. C., H. T. Mowrer, and W. D. Shepperd. 1994. Modeling aspen stand growth with a temporal Bayes network. *AI Applications* 8:15-28.
23. Rieman, B. E., J. T. Peterson, J. Clayton, P. Howell, R. Thurow, W. Thompson, and D.C. Lee. 2001. Evaluation of potential effects of federal land management alternatives on trends of salmonids and their habitats in the interior Columbia River basin. *Forest Ecology and Management* 5501:1-20.
24. Chapra, S. C. 1980. Application of the phosphorus loading concept to the Great Lakes. *in* R.C. Loehr, editor. *Phosphorus Management Strategies for Lakes*. Ann Arbor Science, Ann Arbor, MI.
25. Raymont J. E. 1980. *Plankton and Productivity in the oceans*. 2nd Ed. Pergamon Press.
26. Valderrama J. C. 1981. Determinación simultánea de Nitrógeno y Fósforo Totales. *Mar. Chem.* 10:109-116.
27. EPA. 1983 *Technical Support Manual: Waterbody Surveys and Assessments for Conducting Use Attainability Analyses*. Office of Water. USA pp 620
28. Secretaría de Desarrollo Urbano y Ecología (SEDUE). 1986. Lineamientos y criterios para la selección de índices e indicadores medio ambientales. Serie Ordenamiento Ambiental No. 2 México.
29. APHA/AWWA/WEF. 1992. *Standard Methods for water and wastewater analysis*. American Public Health Association Washington DC USA.
30. Rodier J. 1981. *Análisis de las aguas*. Ediciones Omega S.A. Barcelona 659 pp.

31. Holm Hansen, O. y B. Riemann. 1978. Chlorophyll a determination: improvements in methodology. *OIKOS*. 30:438-447.
32. Strickland, J.H.D. y T. R. Parsons. 1972. A Manual of Sea Water Analysis Fish. Res. Board Can. Bull. Vol. 167.
33. Solórzano, L. 1969. Determination of ammonia in natural waters byphenol-hypoclorite method. *Limnol. Oceanogr.* 14:799-801.
34. Noble R.T., D.F. Moore, M.K. Leecaster, C.D. McGee, S.B. Weisberg. 2003. Comparison of total coliform, fecal coliform, and enterococcus bacterial indicator response for ocean recreational water quality testing. *Water research* 37 (2003) 1637-1643.
35. Kennish Michael J. 1997. Practical handbook of Estuarine and Marine Pollution. CRC Pres Inc. Marine Science Series. Florida USA. 524 pp.
36. Hajmeer M. N. and I. A. Basheer. 2003. A hybrid Bayesian–neural network approach for probabilistic modeling of bacterial growth/no-growth interface. *International Journal of Food Microbiology*. Volume 82, Issue 3, 15 Mayo 2003, Pages 233-243.
37. Diario Oficial de la Federación: 1989, Criterios ecológicos de la calidad del agua CE-CCA-001/89, Tomo CDXXXV (9), 7-22.
38. Schiff, K.C., Dorsey J.H. y Weisberg S.B. 2001. Microbiological monitoring of marine recreational waters in southern california. *Environ manage*; 27:149-57.
39. San Millan María Dolores y Miguel Ibañez. 1979. Ultimos avances en el estudio de la polución del mar I: Metales pesados, bacterias y virus. *Lurralde; inv. Espac. N.2*, p. 81-101. ISSN 1697-3070
40. Michael J. Kennish. 2000. Practical handbook of Estuarine and Marine Pollution. CRC Pres Inc. Marine Science Series. Florida USA. 524 pp.
41. A.D Gronewold y K.H. Reckhow. 2007. Developing a Bayesian Network Model for Bacteriologically Impaired Surface Waters. *Proceedings of the 7th International (IWA) Symposium on Systems Analysis and Integrated Assessment in Water Management*. Washington DC.