

ARTICLE

Análisis Bayesiano para la Detección de Firmas Espectrales de Fusarium en Plantas de Banano.

Santiago Giraldo H.

Estudiante de maestría, Universidad Nacional de Colombia, Medellín

Author for correspondence: S.Giraldo, Email: sagiraldohe@unal.edu.co.

Abstract

La detección temprana de enfermedades en cultivos es fundamental para la seguridad alimentaria y la agricultura sostenible. Este estudio presenta un enfoque novedoso que combina análisis bayesiano espectral con modelos de aprendizaje automático para mejorar la detección temprana de *Fusarium oxysporum* f.sp. cubense Raza 1 (FOCR1) en plantas de banano Gros Michel. Mediante el desarrollo de modelos bayesianos independientes de regresión logística con selección de variables para cada día post-inoculación (1–15 días), incorporamos conocimiento previo de la literatura sobre las etapas de desarrollo de la enfermedad. Los resultados demuestran una caracterización espectral distintiva con picos de precisión en los días 3, 6 y 9, destacando la relevancia de las bandas del espectro Red Edge. El modelo XGBoost enriquecido con características bayesianas superó consistentemente al modelo basado únicamente en datos espectrales, logrando un aumento en la precisión.

Keywords: *Fusarium oxysporum*, análisis bayesiano, espectroscopía VIS/NIR, XGBoost

1. Introducción

El marchitamiento por *Fusarium* causado por *Fusarium oxysporum* f.sp. cubense representa una de las amenazas más devastadoras para la producción mundial de banano, con pérdidas económicas que pueden ser muy perjudiciales para las plantaciones afectadas. La detección temprana durante el período asintomático, que puede extenderse hasta 20 días post-inoculación, es crucial para implementar medidas de control efectivas antes de que los síntomas visibles se manifiesten.

Los métodos tradicionales de detección, incluyendo inspección visual y técnicas moleculares como PCR y ELISA, presentan limitaciones significativas en términos de costo, tiempo y aplicabilidad en campo. La espectroscopía de reflectancia VIS/NIR ha emergido como una alternativa prometedora, ofreciendo detección no destructiva y en tiempo real de cambios fisiológicos asociados con infecciones patógenas (Díaz, 2024).

Trabajos previos han demostrado el potencial de la espectroscopía para la detección de enfermedades en plantas, pero típicamente se basan en enfoques frecuentistas que no incorporan conocimiento previo sobre la progresión temporal de las enfermedades.

Este estudio presenta un enfoque bayesiano que busca mejorar la detección temprana de *Fusarium* en plantas de banano integrando información de la literatura sobre las etapas de desarrollo de *Fusarium* en modelos de selección de variables, logrando una caracterización espectral evolutiva de la enfermedad, identificando patrones temporales distintivos y ventanas críticas de detección durante el período asintomático. Posteriormente, se evalúa la calidad del enfoque por medio de la integración de los resultados con técnicas de aprendizaje automático avanzadas.

2. Metodología

2.1 Diseño Experimental y Adquisición de Datos

Este trabajo utiliza los datos recolectados en Díaz Herrera 2024 donde se utilizaron 240 plantas de banano Gros Michel cultivadas bajo condiciones controladas en el invernadero de AUGURA, Carepa, Antioquia. Las plantas fueron sometidas a ocho tratamientos diferentes: control, estrés hídrico, FOCR1, *Ralstonia solanacearum* Raza 2, y sus interacciones. Los datos espectrales se obtuvieron mediante un espectrofotómetro portátil ASD FieldSpec 4 Hi-Res NG en el rango de 350–2500 nm durante 15 días post-inoculación.

2.2 Análisis Bayesiano con Conocimiento Previo

Para cada día post-inoculación, se desarrolló un modelo bayesiano independiente de regresión logística con selección de variables:

$$\pi_j \sim \text{Beta}(\alpha_\pi, \beta_\pi) \quad (1)$$

$$z_j \sim \text{Bernoulli}(\pi_j) \quad (2)$$

$$B_j \sim \text{Normal}(0, \sigma_\beta) \quad (3)$$

$$B_0 \sim \text{Normal}(0, \sigma_a) \quad (4)$$

$$\theta_i = \text{Logit}(B_0 + \sum B_i \lambda_{ij} * z_j) \quad (5)$$

$$Y_i \sim \text{Bernoulli}(\theta_i) \quad (6)$$

Los parámetros α_π y β_π se configuraron basándose en conocimientos de la literatura, diferenciando entre las etapas de desarrollo de la enfermedad.

2.2.1 Configuración de aprioris

Etapla temprana (días 1-5): Detección pre-sintomática

La especificación de distribuciones a priori para la etapa temprana se fundamenta en el principio de que los cambios bioquímicos preceden a las manifestaciones sintomáticas visibles. Se implementó un prior base restrictivo Beta(1,3) con probabilidad de inclusión esperada del 0.25, reflejando la naturaleza especializada de las respuestas espectrales tempranas y promoviendo parsimonia en la selección de variables.

La banda de clorofila (680-733 nm) recibió la máxima prioridad con Beta(4,1) [$E[\pi] = 0.80$], justificada por la degradación temprana de clorofila como indicador universal de estrés biótico antes de síntomas visibles (Mahlein et al., 2012). Las bandas de absorción de agua (1400-1450, 1900-1950 nm) fueron priorizadas con Beta(3,1) [$E[\pi] = 0.75$], reflejando que *Fusarium* causa estrés hídrico temprano por obstrucción vascular.

Los compuestos fenólicos (300-400 nm) recibieron priors moderadamente informativos Beta(2,1) [$E[\pi] = 0.67$], reconociendo su papel como biomarcadores de respuesta temprana de estrés y defensa. La región red-edge (700-750 nm) mantuvo la misma probabilidad, reflejando su sensibilidad a cambios estructurales tempranos (Mahlein et al., 2012).

Etapla media (días 6-10): progresión sintomática

Durante la etapa media, la progresión de la infección sistémica justifica una menor restricción en el prior base Beta(1,2) [$E[\pi] = 0.33$], permitiendo mayor diversidad espectral informativa.

La intensificación de síntomas se refleja en el incremento de probabilidades de inclusión para regiones establecidas: clorofila aumenta a Beta(5,1) [$E[\pi] = 0.83$], mientras que las bandas de agua alcanzan Beta(4,1) [$E[\pi] = 0.80$] por el estrés hídrico severo establecido (Bauriegel et al., 2011).

Etapla tardía (días 10-15): colapso sistémico

Esta etapa se caracteriza por el colapso fisiológico, justificando el prior base menos restrictivo Beta(1,1.5) [$E[\pi] = 0.40$]. El colapso sistémico amplifica las señales espectrales, alcanzando la máxima probabilidad de inclusión para clorosis y muerte foliar (560-733 nm) con Beta(6,1) [$E[\pi] = 0.86$], mientras que el colapso del sistema hídrico se intensifica Beta(5,1) [$E[\pi] = 0.83$] (Ploetz, 2015).

2.2.2 Implementación del modelo

El modelo fue implementado en Stan, tal como se detalla en el anexo A. Para la implementación se utilizaron los siguientes hiperparámetros:

- Cadenas: 4
- Iteraciones: 2000
- Warmup = 10000

Con esta configuración se espera una buena convergencia de los modelos.

3. Resultados

3.1 Evaluación del enfoque

Con el fin de evaluar la relevancia del análisis bayesiano para mejorar la capacidad de detección de la enfermedad en plantas de banano, se desarrollaron dos modelos XGBoost para comparación:

Modelo espectral básico: Utilizando únicamente datos espectrales e información temporal

Modelo con características informadas: Incorporando las siguientes características derivadas del análisis bayesiano:

- Probabilidades de inclusión (media, máximo, mediana, Q75)
- Coeficientes efectivos (media, máximo, desviación estándar)
- Métricas de importancia y selección de variables

4. Resultados

4.1 Diagnóstico de convergencia

La convergencia de los modelos bayesianos implementados en Stan fue evaluada mediante diagnósticos estándar para asegurar la validez de las inferencias realizadas.

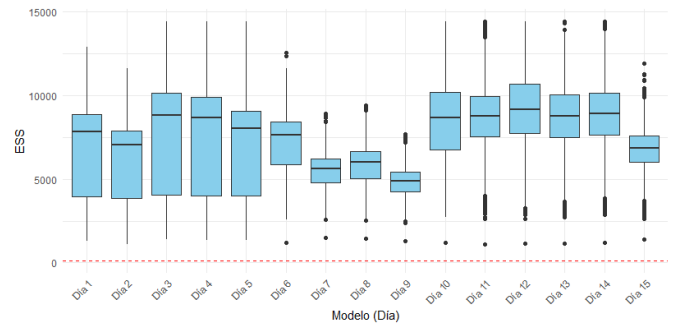


Figure 1. Tamaños efectivos de la muestra para cada modelo

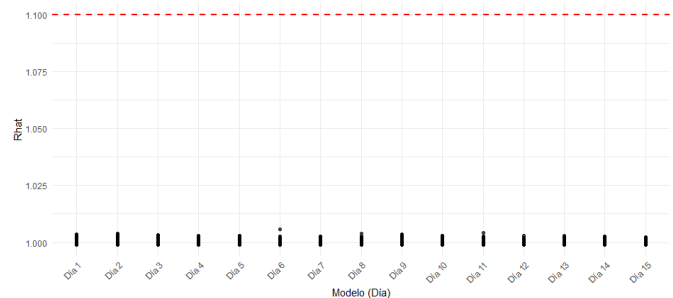


Figure 2. Estadístico (R-hat)

El estadístico (R-hat) de Gelman-Rubin mostró valores consistentemente menores a 1.1 para todos los parámetros de interés en los 15 modelos desarrollados (uno por cada día post-inoculación), indicando convergencia satisfactoria entre las cadenas MCMC (Figura 2). Los tamaños efectivos de muestra (N-eff) fueron superiores a 1000 para los parámetros, con valores típicos arriba de 5000, garantizando estimaciones precisas de las distribuciones posteriores (Figura 1).

4.2 Caracterización espectral temporal

El análisis reveló una caracterización espectral distintiva con patrones específicos de precisión a lo largo del tiempo. Se identificaron picos de precisión significativos en los días 3, 6 y 9, sugiriendo ventanas temporales críticas donde la discriminación entre plantas sanas e infectadas es óptima. Las bandas espectrales más relevantes se concentraron predominantemente en la región Red Edge (700–750 nm), confirmando la importancia de esta región para la detección de estrés vegetal. Esta región espectral es particularmente sensible a cambios en la estructura foliar y contenido de clorofila, indicadores tempranos de infección por Fusarium.

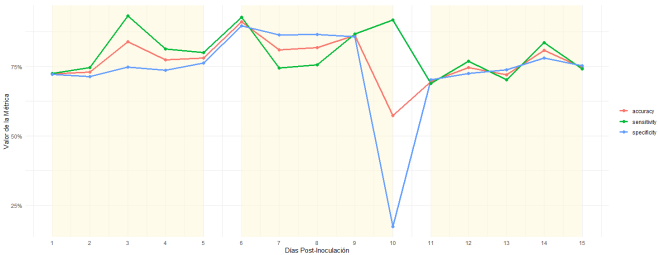


Figure 3. Métricas de precisión por día

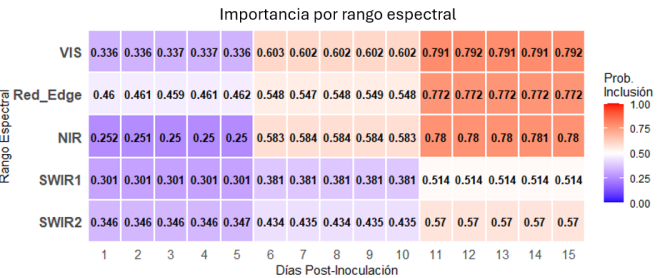


Figure 4. Importancia de rangos espectrales por día

4.3 Evolución Temporal de Bandas Relevantes

Se observaron cambios significativos en la relevancia de bandas espectrales en los días 5 y 10, marcando transiciones entre las etapas de desarrollo de la enfermedad. Particularmente notable fue un fenómeno emergente importante en el día 10, caracterizado por un cambio dramático en los patrones espectrales, posiblemente asociado con la transición de la fase. Esto se puede ver en el gráfico de estabilidad temporal (Figura 5), donde se muestra la correlación entre las bandas espectrales elegidas en cada día con las elegidas en el día siguiente.

4.4 Comparación de resultados con la literatura científica

La correlación con la literatura científica resultó débil (figura 7), sugiriendo que los datos experimentales específicos predominan sobre el conocimiento previo generalizado para la clasificación en este contexto particular. Este hallazgo subraya la importancia de enfoques adaptativos que permitan que los datos informen la selección de características.

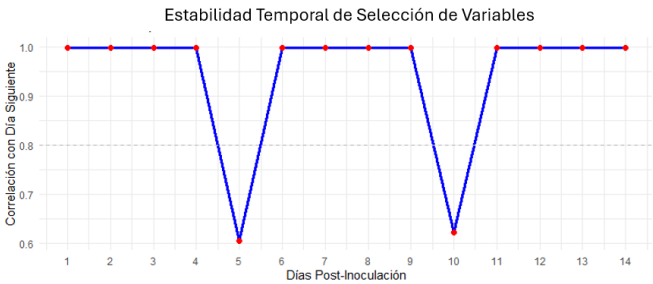


Figure 5. Gráfico de estabilidad temporal

4.5 Rendimiento de Modelos

El enfoque bayesiano demostró mejoras consistentes sobre el trabajo original, particularmente en los días 6 y 9 donde se lograron precisiones superiores (Figura 3). Notablemente, se obtuvo precisión consistente a lo largo de todos los días evaluados, manteniendo robustez a pesar de contar con un número limitado de muestras por día individual.

4.6 Evaluación de XGBoost con características Bayesianas

La comparación entre el modelo XGBoost básico y el enriquecido con características bayesianas mostró los siguientes resultados:

=== COMPARACIÓN DE RESULTADOS ===			
	Metric	Spectral_Only	Spectral_Bayesian Improvement
1	Accuracy	0.9844	0.9900 0.0056
2	Sensitivity	1.0000	1.0000 0.0000
3	Specificity	0.9683	0.9796 0.0113
4	Precision	0.9701	0.9806 0.0105
5	F1-Score	0.9848	0.9902 0.0054

Figure 6. Comparación entre modelos XGBoost

Se encontró que el modelo informado por el análisis Bayesiano mejoró hasta en un 1% las métricas de clasificación, alcanzando un notable 99.02% en el F1-score.

5. Conclusiones

Los resultados obtenidos demuestran que el enfoque bayesiano integrado con aprendizaje automático representa un avance para la detección temprana de Fusarium oxysporum en banana. La identificación de picos de precisión en los días 3, 6 y 9 revela la existencia de ventanas temporales críticas que coinciden con eventos fisiológicos clave: respuestas de defensa iniciales, establecimiento de la infección vascular e intensificación pre-sintomática, respectivamente. La predominancia de bandas en la región Red Edge confirma la sensibilidad de esta zona espectral para capturar cambios estructurales y metabólicos tempranos, validando su importancia como biomarcador espectral.

El fenómeno emergente observado en el día 10 marca una transición crucial hacia la manifestación sintomática, proporcionando una señal de alerta temprana invaluable para intervenciones preventivas.

La superioridad del modelo XGBoost enriquecido con características bayesianas sobre el enfoque espectral básico demues-

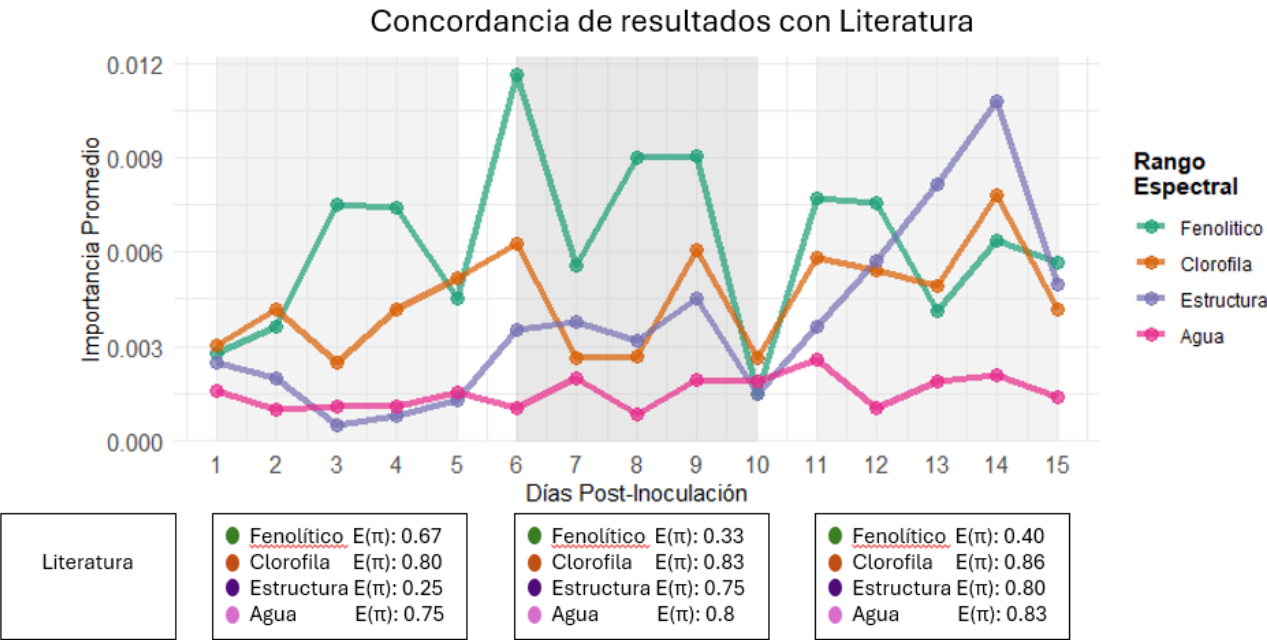


Figure 7. Gráfico de comparación de literatura con los resultados obtenidos

tra el valor agregado de incorporar conocimiento estructurado en sistemas de aprendizaje automático.

La consistencia temporal del rendimiento, mantenida a pesar del tamaño limitado de muestra por día, evidencia la robustez del marco metodológico desarrollado y su potencial para la implementación práctica en sistemas de monitoreo agrícola automatizado.

References

Bauriegel, Elke, Andrea Giebel, Martin Geyer, Uwe Schmidt, and Werner B. Herppich. 2011. Early detection of fusarium infection in wheat using hyper-spectral imaging. Research article on hyperspectral detection of Fusarium in wheat, *Computers and Electronics in Agriculture* 75 (2): 304–312. issn: 0168-1699. <https://doi.org/10.1016/j.compag.2010.12.006>.

Díaz Herrera, Cristian Camilo. 2024. Detección temprana de estrés biótico y abiótico usando modelos de clasificación de datos de espectroscopía de reflectancia vis/nir: aplicación en plantas de banano gros michel. Trabajo final de grado presentado como requisito parcial para optar al título de Maestría en Estadística. Tesis de Maestría en Estadística, Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de Estadística.

Mahlein, Anne-Katrin, Ulrike Steiner, Christian Hillnhütter, Heinz-Wilhelm Dehne, and Erich-Christian Oerke. 2012. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. Open Access article distributed under the terms of the Creative Commons Attribution License, *Plant Methods* 8 (1): 3. issn: 1746-4811. <https://doi.org/10.1186/1746-4811-8-3>.

Ploetz, Randy C. 2015. Fusarium wilt of banana. Review article on Fusarium wilt disease of banana, *Phytopathology* 105 (12): 1512–1521. issn: 0031-949X. <https://doi.org/10.1094/PHYTO-04-15-0101-RVW>.

```
data {
  int<lower=1> N;
  int<lower=1> P;
  matrix[N, P] X;
  int<lower=0,upper=1> y[N];
  vector<lower=0>[P] alpha_pi;
  vector<lower=0>[P] beta_pi;
  real<lower=0> sigma_beta;
  real<lower=0> sigma_alpha;
}

parameters {
  real B_0;
  vector[P] B_j;
  vector<lower=0,upper=1>[P] pi_j;
}

transformed parameters {
  // Coeficientes
  vector[P] effective_beta = B_j .* pi_j;
}

model {
  // Priors
  B_0 ~ normal(0, sigma_alpha);
  B_j ~ normal(0, sigma_beta);
  pi_j ~ beta(alpha_pi, beta_pi);

  // Likelihood
  y ~ bernoulli_logit(B_0 + x * effective_beta);
}
```

Figure 8. Modelo en Stan

Appendix 1. ANEXO A: Modelo implementado en Stan