

## Empalme por Interpolación Estructural de las series de los principales cultivos estimados por la Encuesta de Superficie y Producción Agropecuaria Continua 2002 – 2014

Salomé Borja Parreño\*

Javier Núñez Salgado\*<sup>1</sup>

### RESUMEN

En el presente documento se detalla el método aplicado para el Empalme por Interpolación Estructural de las series de superficie plantada y cosechada de los principales cultivos estimados mediante la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC). En primer lugar, se expone la motivación para realizar este ejercicio considerando los diferentes agravantes tales como la desactualización de los marcos de muestreo, cambios estructurales en los dominios de estudio y cambios metodológicos introducidos en la ronda 2014. Luego, se detalla el método a utilizar explicando su desarrollo e información necesaria a utilizar. Para finalizar, se presenta el resultado obtenido para la serie de superficie cosechada de Banano.

**Palabras clave:** Empalme por Interpolación Estructural, ESPAC, Test de sesgo

<sup>1</sup> Los autores son investigadores del Instituto Nacional de Estadística y Censos.

## Introducción

Durante el periodo 2002-2013, el Instituto Nacional de Estadística y Censos (INEC) produjo indicadores para el sector agropecuario utilizando como fuente de información la Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC), la cual generaba estimaciones a partir de un Muestreo en Dos Fases (Doble) implementado a partir del Censo Nacional Agropecuario realizado en el año 2000 (CNA 2000). Debido a su diseño muestral, la muestra usada para la ESPAC era una submuestra seleccionada a partir de las unidades de observación que participaron en el CNA 2000.

El CNA 2000 no fue un ejercicio censal por enumeración completa, puesto que se ejecutó a partir de un Muestreo de Marco Múltiple (MMM), compuesto por un Marco de Muestreo de Áreas (MMA) y un Marco de Muestreo de Lista (MML), donde:

- El diseño del MMA fue probabilístico, estratificado y replicado de áreas (conglomerados) con dos etapas de selección mediante el cual, se seleccionaron sistemáticamente las Unidades Primarias de Muestreo (UPM's) y en la segunda etapa, se escogió aleatoriamente al menos un Segmento de Muestreo (SM) en cada UPM incluida en la muestra de primera etapa de selección.
- EL MML contemplaba las Unidades de Producción Agropecuarias (UPA's) importantes (100 ha. o más), sectoriales (menos de 50 ha. con cultivos principales focalizados) y especiales (flores, maracuyá, mango, avícolas, porcinos, camaronerías, entre otras).

En la etapa de campo del CNA 2000 se encuestaron 12,277 SM del MMA y 15.238 UPA's correspondientes al MML, de los cuales, 2000 SM y 4000 UPA's fueron seleccionados para la primera ronda de ESPAC en el año 2002.

Debido al Muestreo en dos Fases (Doble), la eficacia estadística de los estimadores resultantes de la ESPAC se basan en la posibilidad de correlacionar la información obtenida de la muestra de Primera Fase (muestra CNA 2000) con la información recolectada en la muestra de Segunda Fase (muestra ESPAC), pero esta ha venido perdiendo consistencia en el transcurso del tiempo debido principalmente a factores como:

- Desactualización de los marcos de muestreo,
- Tasas de no respuesta, mayores al 5%,
- Utilización de la misma muestra para todas las rondas entre 2002 y 2013,
- Marco de Muestreo de Lista incompleto.

Considerando estos factores en el marco de la cooperación INEC-FAO entre el segundo y tercer trimestre del año 2014, el INEC realizó un proceso de cambio metodológico y actualización de la ESPAC, el cual contó con el desarrollo de un nuevo Marco de Muestreo de Áreas y una actualización completa del Marco de Lista, dando como resultado una ronda 2014 de ESPAC con sustanciales mejoras, entre las cuales se encuentran:

- El diseño muestral es estratificado, constando con 4 estratos definidos por la intensidad de uso de suelo agropecuario que posee cada Unidad Mínima de Estratificación. Además, el tamaño de cada SM es inversamente proporcional a la intensidad de uso de suelo, siendo los más pequeños los SM con una intensidad de uso de suelo más elevada.
- Los segmentos muestrales del nuevo Marco de Muestreo tienen límites geométricos, puesto que fueron construidos en función a los mapas de uso y cobertura de suelo generados por un convenio entre MAGAP y MAE,
- La unidad de observación son todos los terrenos que se encuentran parcialmente o totalmente contenidos en los segmentos seleccionados para la etapa de campo, a excepción de la información referente a cantidad de cabezas de ganado y empleo,
- El método de estimación se fundamenta en la metodología del segmento ponderado, sin embargo, la superficie dentro del SM se obtiene de planimetrar los terrenos investigados,
- El Marco de Lista a utilizar fue reestructurado y actualizado, considerando nuevamente los cultivos a incluirse en él.

Debido a la obsolescencia de las estimaciones motivada por la desactualización de los marcos de muestreo, cambios estructurales en los dominios de estudio de ESPAC<sup>2</sup> y cambios metodológicos introducidos debido a las mejoras implementadas, el INEC presenta el siguiente documento en el cual se detalla el proceso desarrollado para el empalme por interpolación estructural de las series de superficie plantada y cosechada de los principales productos a nivel provincial y sus series agregadas a nivel nacional para el periodo 2002-2014.

<sup>2</sup> Desagregación de los dominios de estudio de la Amazonía en sus respectivas provincias. Creación de las provincias de Santa Elena y Santo Domingo de los Tsáchilas en el año 2007.

## 1. Empalme por interpolación estructural de las series ESPAC 2002 - 2014

El método de empalme por interpolación estructural permite mantener el comportamiento de la serie nacional, considerando como años base a los puntos iniciales (año 2002) y finales (año 2014) de cada una de las series<sup>3</sup>. El modelo a utilizar se desarrolla en el documento "Alternativas metodológicas para el empalme estadístico de serie temporales: caso Venezuela 1950-2005" donde éste se aplica para el empalme por interpolación estructural de las series del PIB de Venezuela para los años 1950 – 2005.

En el mismo se expone que el objetivo del empalme por interpolación estructural es recuperar los niveles y estructura proporcional del PIB venezolano a lo largo de los años 1950-2005, en los cuales existen más de un año base bajo los cuales se calculan los indicadores para determinados periodos.

Este método es compatible con las estimaciones resultantes de la ESPAC, puesto que es una encuesta que brinda información estructural de uso y cobertura de suelo para sus dominios de estudio, es decir, la ESPAC brinda información sobre qué cultivos se siembran, los dominios de estudio dónde se siembran y sus niveles de participación proporcionales.

La fórmula de cálculo del empalme por interpolación estructural de la ESPAC expresa:

$$\omega_{ij,t} = \left( \frac{N_{ij,t^1}}{T_{j,t^1}} - \frac{N_{ij,t^0}}{T_{j,t^0}} \right) \cdot \left( \frac{t - t^0}{t^0 - t^1} \right)^{-1/\ln(\alpha)} + \frac{N_{ij,t^0}}{T_{j,t^0}}$$

Dónde:

$\omega_{ij,t}$  = Disposición estructural de la superficie del cultivo  $j$  en el  $i$ -ésimo dominio en el periodo de tiempo  $t$ .

$N_{ij,t^1}$  = Superficie del cultivo  $j$  en el  $i$ -ésimo dominio en el año base  $t^1$ .

$T_{j,t^1}$  = Superficie del cultivo  $j$  a nivel nacional en el año base  $t^1$ .

$N_{ij,t^0}$  = Superficie del cultivo  $j$  en el  $i$ -ésimo dominio en el año base  $t^0$ .

$T_{j,t^0}$  = Superficie del cultivo  $j$  a nivel nacional en el año base  $t^0$ .

<sup>3</sup> A excepción de los cambios en los dominios de estudio de las provincias de Santa Elena y Santo Domingo de los Tsáchilas, para las cuales el año base 0 es 2009.

$j$  = Índice que recorre los cultivo (permanentes o transitorios) a ser empalmados.

$i$  = Índice que recorre los dominio de estudio por cultivos a ser tratados.

$t^0$  = Año base 0 (2002).

$t^1$  = Año base 1 (2014).

$t$  = Años comprendidos en el periodo 2002-2014.

$\alpha$  = Parámetro entre 0 y 1.

A diferencia del planteamiento original, en vez de determinar  $\alpha$  a través de una fórmula que maximice la correlación entre las variables de la serie expresada en el año base oficial para cada periodo de tiempo, se optó por utilizar el complemento de la probabilidad de existencia de sesgo de obsolescencia entre las estimaciones.

La probabilidad de existencia de sesgo de obsolescencia es obtenida a partir de una prueba de hipótesis que contrasta la proporción de cada cultivo por dominio de estudio entre 2002 y 2014. Esta prueba se realiza para medir si las diferencias entre las proporciones de cada cultivo por dominio de estudio son significativas. Gracias a esta prueba, obtenemos una aproximación del impacto de la desactualización del marco sobre las estimaciones obtenidas a través de la ESPAC.

Bajo esta consideración,  $\alpha$  se define como:

$$\alpha = 2 * (1 - \lambda)$$

dónde  $\lambda$  es la probabilidad de que la serie presente sesgo de obsolescencia. El test de sesgo de obsolescencia se realizó únicamente sobre el MA generado para ESPAC debido a que la unidad de observación del MA son terrenos y del ML son las UPA's.

Dado que la participación de todos los cultivos permanentes y transitorios no es equitativa, para este ejercicio se han seleccionado los cultivos que cumplan ciertos requerimientos tales como:

- Aporte desde el punto de vista de superficie plantada y cosechada, producción y ventas,
- Mayor presencia en los dominios de estudio, que se traduce en un número mínimo de observaciones muestrales<sup>4</sup>. Esta restricción es importante puesto que para un  $n$  muy pequeño se incurre en errores sistemáticos en el cálculo

<sup>4</sup> Tomamos como mínimo 30 observaciones muestrales para proporciones que se encuentran entre 0,2 y 0,8 (Kish, 1965).

de la varianza de los estimadores, afectando el resultado del sesgo de obsolescencia,

- Existencia de sesgo de obsolescencia en ese cultivo a nivel nacional o en los dominios de estudio considerado para el mismo.

Luego de este análisis, los cultivos considerados son los siguientes:

**Tabla 1.** Lista de cultivos considerados

Cultivos permanentes	Cultivos transitorios
Banano	Arroz
Cacao	Maíz Duro Seco
Café	Papa
Caña de Azúcar	
Palma	

Fuente: ESPAC  
Elaboración propia de los autores

Cabe mencionar que los estimadores de superficie con uso de suelo declarado como pasto cultivado, ganado vacuno y ganado porcino no se consideraron para el cálculo de sesgo puesto que estas 3 estimaciones tienen una gran participación del ML de la ESPAC. Además, puesto que cada cultivo precisa diferentes características para su producción, se consideraron dominios de investigación por producto en función a su volumen de producción. Estos dominios de estudio fueron considerados a nivel provincial o Resto país y quedan distribuidos de la siguiente manera:

**Tabla 2.** Dominios de estudio por cultivo a ser considerados

Tipo de cultivo	Cultivo	Dominios de estudio
Permanentes	Banano	El Oro, Guayas, Los Ríos, Manabí, Resto país.
	Cacao	Esmeraldas, Guayas, Los Ríos, Manabí, Resto país.
	Café	Loja, Manabí, Resto país.
	Caña de Azúcar	Guayas, Resto país.
Transitorios	Palma	Esmeraldas, Los Ríos, Santo Domingo, Resto país.
	Arroz	Guayas, Los Ríos, Manabí, Resto país.
	Maíz Duro Seco	Guayas, Loja, Los Ríos, Manabí, Resto país.
	Papa	Bolívar, Cañar, Carchi, Chimborazo, Cotopaxi, Pichincha, Tungurahua, Resto país.
Total	8 cultivos	28 series a empalmar.

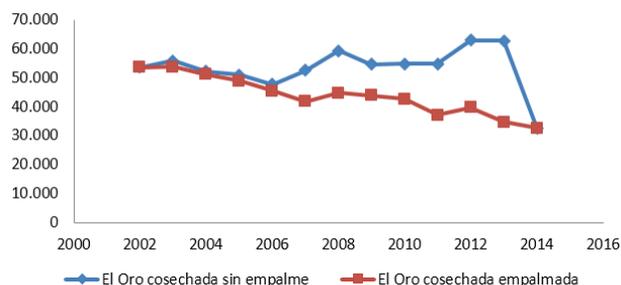
Fuente: ESPAC  
Elaboración propia de los autores

## 2. Aplicación del método de empalme por interpolación estructural.

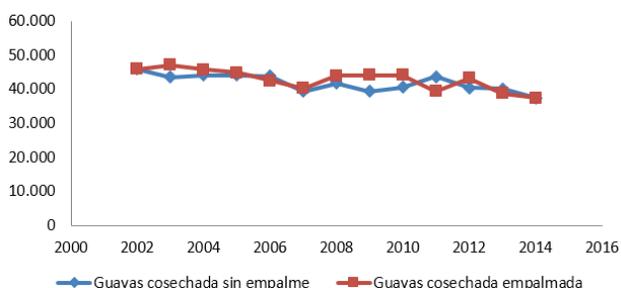
A continuación se presenta un ejemplo desarrollado para la serie superficie cosechada del cultivo permanente Banano y sus dominios de estudio, los cuales son las provincias de El Oro, Guayas, Los Ríos, Manabí y Resto país. En cada gráfico, se expresa el número de hectáreas sembradas de Banano para la serie 2002-2014.

Luego de reconstruir las series en cada dominio de estudio, reconstruimos la serie nacional a partir de la agregación de los valores por dominio, obteniéndose:

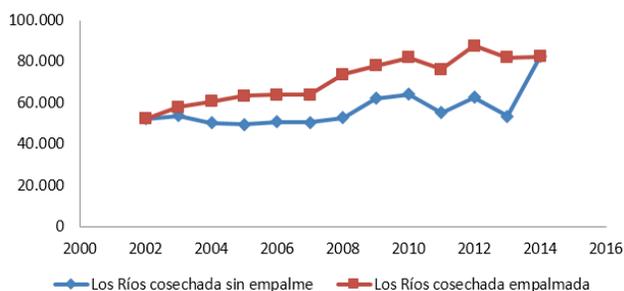
**Empalme Superficie Cosechada de Banano El Oro**



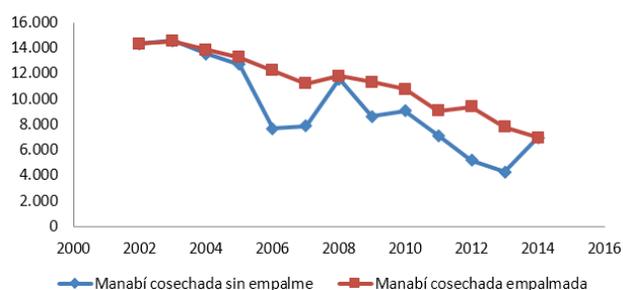
**Empalme Superficie Cosechada de Banano Guayas**



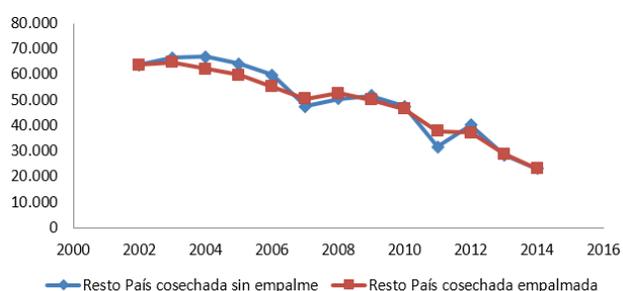
**Empalme Superficie Cosechada de Banano Los Ríos**



**Empalme Superficie Cosechada de Banano Manabí**



**Empalme Superficie Cosechada de Banano Resto País**



Para el empalme por interpolación estructural de las series se utilizaron los siguientes valores para  $\alpha$ :

**Tabla 3.** Valores  $\alpha$  usados en el ejercicio de empalme

Dominio de estudio	Valor $\alpha$
El Oro	0,39
Guayas	0,17
Los Ríos	0,36
Manabí	0,54
Resto País	0,63

Fuente: ESPAC  
Elaboración propia de los autores

El test de sesgo de obsolescencia fue calculado de manera independiente tanto para la superficie plantada como para la superficie cosechada para cada uno de los cultivos en los dominios a trabajar, por lo tanto, los puntajes  $\alpha$  son independiente para cada serie.

### 3. Conclusiones

El empalme por interpolación estructural expuesto es mejorable en los próximos años, puesto que se dispondrán de más datos en la serie y se podrán utilizar modelos más óptimos.

Existen ciertos cultivos donde en algunos dominios la superficie plantada es menor a la superficie cosechada. Esto se debe principalmente a características de cultivo que hacen que las diferencias entre estas superficies sean más o menos significativas al momento de calcular su sesgo de obsolescencia. El listado completo de las series que presentan este problema es el siguiente:

**Tabla 4.** Cultivos con diferencia negativa entre su superficie plantada y cosechada

Producto	Provincia	Años con inconvenientes
Banano	Guayas	1
Banano	Los Ríos	9
Cacao	-	-
Café	Loja	5
Caña de Azúcar	-	-
Palma Africana	Esmeraldas	1
Palma Africana	Resto país	1
Arroz	-	-
Maíz duro seco	-	-
Papa	Cañar	1
Papa	Carchi	1
Papa	Chimborazo	1
Papa	Cotopaxi	1

Fuente: ESPAC  
Elaboración propia de los autores

Considerando lo expuesto, el INEC realizará el ejercicio de reconstrucción en forma continua de las series de superficie sembrada y cosechada de los productos seleccionados con la nueva información disponible de la ESPAC en forma anual, realizando un análisis de las diferencias presentadas hasta que estas se consideren como estadísticamente no significativas.

Puesto que el empalme por interpolación estructural de las series de ESPAC considera dos periodos de referencia de la información (2002-2013 y 2014 en adelante), el ejercicio se focalizará en los puntos comprendidos entre 2002-2013 manteniendo las estimaciones de 2014 en adelante acordes a nuestras publicaciones anuales.