

Predicción de anomalías de producción agrícola con imágenes de detección remota

1 Contexto

En el marco de la generación del Sistema de Información Social Integral de la Secretaría de Desarrollo Social se ha planteado la necesidad de crear un Sistema de Alerta Temprana para prevención de hambrunas. En este avance se presenta un breve resumen de la dirección y el objetivo del sistema, así como resultados preliminares.

2 Diseño de solución

A partir de la décadas de los setenta y ochenta, en particular, a raíz de las hambrunas en Sudán y Etiopía de 1984-1985, se ha impulsado la creación de sistemas de alerta temprana para prevención de hambrunas. En este contexto, el Famine Early Warning System (FEWS) de US Agency for International Development ha sido uno de los principales actores, investigadores y desarrolladores de metodología para sistemas de alerta temprana.

El esquema que presentamos aquí está basado en varios estudios auspiciados por FEWS, a saber [Funk and Verdin, 2010], [Funk and Budde, 2009], [David López-Carr, 2016] y [Molly E. Brown, 2008], y se basa en análisis de imágenes de detección remota para crear un monitor de producción agrícola. A grandes rasgos, dichos estudios utilizan índices de vegetación obtenidos de imágenes satelitales para detectar anomalías en producción agrícola.

3 Marco Teórico: predicción fenológica de producción agrícola con índices de vegetación.

Entre los índices de vegetación más utilizados en monitores de vegetación se encuentra el Índice de Vegetación de Diferencia Normalizada, o NDVI por sus siglas en inglés, que sirve para estimar la calidad y cantidad de biomasa basado en la radiación que las plantas emiten. A grandes rasgos, la clorofila de las plantas absorbe gran parte de la luz visible, o en un rango 0.4 a 0.7 μm , mientras

que la estructura celular de las plantas refleja la luz casi-infrarroja en el rango 0.7 a 1.1 μm [Roderick, 1996]. Por tanto, las plantas se aprecian oscuras en un rango espectral, y brillantes en otro, y una gran diferencia entre ambas bandas indicaría la presencia de vegetación. Planteado así, el NDVI se define como

$$NDVI = \frac{NIR - VIS}{NIR + VIS} \quad (1)$$

Donde,

- NIR , es el espectro del infrarrojo cercano
- VIS , es el espectro de luz visible

Y la división pretende normalizar por diferencias de iluminación, ángulo y geometría. Por su planteamiento, el NDVI suele tomar valores entre -1 y 1, donde 1 representa abundante vegetación, valores cercanos a cero, tierra árida [Roderick, 1996].

3.1 Suavizamiento de NDVI

Existen cuantiosas fuentes de contaminación para el NDVI, entre las que destacan nubes, perturbaciones atmosféricas e iluminación variable, que disminuyen el valor de NDVI obtenido. Por esa razón, es recomendable utilizar alguna función de suavizamiento para series de tiempo de NDVI. [Swets, 1999] propone un método basado en mínimos cuadrados que pondera por la calidad del pixel y la altura del valor, por lo que observaciones anómalamente bajas tienen menor peso. Por otro lado, el Filtro de Savitzky-Golay, comúnmente utilizado en procesamiento de señales, calcula regresiones polinómicas locales a través de coeficientes de convolución. Dado un conjunto de coeficientes de convolución $\{C_i\}$, y observaciones de la forma $\{x_j, y_j\}$, para $j = 1, \dots, n$. El Filtro de Savitzky-Golay está dado por,

$$Y_j = \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \quad (2)$$

con $\frac{m+1}{2} \leq j \leq n - \frac{m-1}{2}$ y coeficientes de convolución C_i , m .

Los coeficientes de convolución se calculan según los parámetros m , h y p : el número de puntos adyacentes sobre el cual calcular un polinomio local, la distancia entre puntos y el grado del polinomio.

3.2 Predicción fenológica de producción agrícola

Se han planteado varios métodos para medir el desempeño de un cultivo dadas las condiciones climáticas. En particular, el Water Requirement Satisfaction Index (WRSI) tiene un enfoque teórico para analizar los requerimientos de agua de un cultivo en específico. En este sentido, crea un balance

entre agua disponible (a través de precipitación y temperatura) y agua requerida por un cultivo en un momento específico de su crecimiento; o bien, el requerimiento fenológico de un cultivo. Mientras el planteamiento del WRSI ha sido funcional, requiere de un error bajo en su insumo de datos para calcular evapotranspiración (ET), o la cantidad de agua que se regresa a la atmósfera por evaporación y transpiración vegetal. Por esa razón, su implementación a nivel nacional puede ser difícil.

Sin embargo, [Funk and Verdin, 2010] y [Nagler P. L., 2005] han planteado que el NDVI puede ser un buen proxy para evapotranspiración si se ajusta de manera correcta con la métrica Σ_v . Bien así,

$$\Sigma_v \propto \sum_{t_0}^{t_0+LGP} ET_i \quad (3)$$

donde t_0 es el inicio de siembras y LGP el periodo de crecimiento del cultivo (*length of growth period*).

La métrica Σ_v tiene cuatro parámetros utilizados para ajustar el estadístico de forma fenológica. Estos son t_0 , o inicio de periodo (ya sea de siembras o de lluvias), lag , o periodo de espera entre siembra o lluvias y la respuesta en NDVI, y LAP o periodo de acumulación en el que el NDVI es informativo. Adicionalmente, al estadístico se resta el NDVI inicial para evitar influencia vegetación anterior. Siendo así,

$$\Sigma_v = \sum_{t_0+lag}^{t_0+lag+LAP} (NDVI_t - NDVI_{t_0}) \quad (4)$$

4 Orígenes y Fuentes de Datos

El monitor de producción agrícola, como presentado por [Funk and Budde, 2009] tiene tres insumos principales: imágenes satelitales, datos históricos de producción de maíz y datos de precipitación. Adicionalmente, es recomendable utilizar una máscara de uso de suelo para determinar zonas agrícolas en el país.

Para imágenes satelitales se utilizará MODIS, o Moderate-resolution Imaging Spectroradiometer, un instrumento científico de libre acceso presente en los satélites Terra (EOS AM) y Aqua (EOS PM) de la NASA, que desde el año 2000 provee lecturas de la Tierra con periodicidad de uno a dos días. MODIS ofrece 36 bandas espectrales a tres distintas resoluciones: 250 m^2 , 500 m^2 , y 1 km^2 . Adicionalmente, presenta compuestos de índices de vegetación con periodicidad de 16 días y un mes. Mientras es posible mantener el monitor en funcionamiento con la periodicidad de MODIS, los resultados actuales se realizaron con los índices de vegetación compuestos de MODIS con periodicidad de 16 días desde el año 2004.

Adicionalmente, se cuenta con datos de SAGARPA para avance agrícola mensual desagregados a nivel municipal desde 2004. El reporte mensual de SAGARPA, presentado a través del Servicio

de Información Agroalimentaria y Pesquera cuenta con información sobre superficie sembrada, superficie cosechada, superficie siniestrada, producción, ciclo agrícola, y modo de riego.

Por último, contamos con mapas de uso de suelo publicados por INEGI en los años 2003, 2007 y 2012.

4.1 Procesamiento y análisis de datos de detección remota

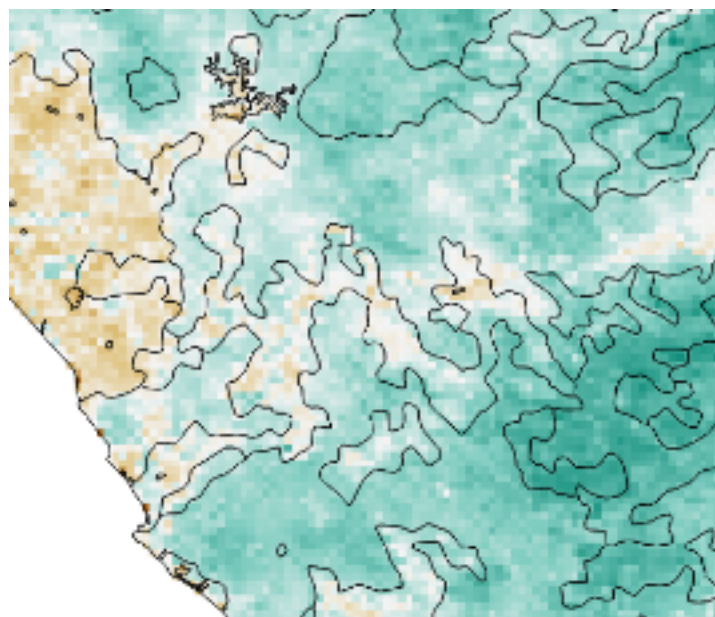
Utilizando el software Python y GDAL se obtuvo información de MODIS con periodicidad de 16 días y resolución de 1 km^2 a partir de 2004 para los *tiles* de información satelital que cubren el territorio mexicano. Utilizando un mapa vectorial del territorio nacional del INEGI, se ajustaron las imágenes satelitales para cubrir únicamente el territorio mexicano. Con R 3.3.1 se aplicó el filtro de Savitzky-Golay para obtener una serie de tiempo suavizada.

La figura 1 muestra el desempeño de un pixel aleatorio en territorio nacional para las series de tiempo obtenida de MODIS (*raw*), y la serie suavizada (*smooth*). El efecto de suavizamiento se puede apreciar en casos en los que el NDVI tiene una tendencia, una observación única más baja, y continúa la tendencia anterior. Tal es el caso de enero y agosto 2016; y, en menor medida, julio 2009. Una vez aplicado el filtro, mientras no se pierde la observación anómala de enero 2016, sí se atenúa su efecto.

Una vez aplicado el filtro para suavizamiento, se procedió de dos formas: utilizando una máscara de división municipal de INEGI, y una máscara de uso de suelo, también de INEGI, se calculó el NDVI promedio municipal, y el NDVI promedio municipal en zonas agrícolas. La figura 2 muestra las series de tiempo de NDVI promedio, por municipio, de Aguascalientes.

Se puede apreciar que los municipios siguen algunas tendencias: Calvillo es el municipio con el mayor índice de vegetación, mientras El Llano tiene consistentemente los periodos de menor vegetación. Por su parte, San José de Gracia suele tener un descenso en NDVI a la par con Calvillo, a pesar de tener valores más consistentemente más bajos en época de lluvias.

El uso de máscaras de uso de suelo puede consistentemente añadir precisión al modelo en algunos casos, particularmente en zonas de grandes productoras agrícolas. La figura 3 muestra un mapa por uso de suelo para una sección de Sinaloa. En color se muestran los pixeles de NDVI con resolución de 1 km^2 , siendo verde para valores de NDVI más altos, y café más bajos. Por uso de suelo, las zonas cafés corresponden a cuerpos de agua, ciudades y, en mayor medida, a zonas agrícolas. En la figura se muestra el NDVI en enero de 2012, época que corresponde a la posterior a la cosecha, lo que explica el valor de NDVI bajo de zonas agrícolas.



Elaboración propia en QGIS con datos de MODIS e INEGI

Figure 3: NDVI de 17 de enero 2012 por uso de suelo en Sinaloa

El uso de máscaras también puede alterar de manera sustancial el NDVI total. La figura 4 contiene series de tiempo para el NDVI total y de zonas agrícolas de Guasave, Sinaloa. Como se puede corroborar con datos de producción de SAGARPA en la figura 5, la totalidad de producción de maíz en Guasave se realiza en modo de riego, y la gran mayoría durante el ciclo agrícola Otoño-Invierno, ciclo que no corresponde con el periodo de lluvias propio de la entidad. Como resultado, el NDVI promedio del municipio tiene picos de vegetación que no corresponden al patrón que se había observado en otros municipios. Considerando únicamente el NDVI de zonas agrícolas, se puede apreciar un ciclo completamente distinto al ciclo del municipio total.

En la figura 5 se puede apreciar la producción por ciclo agrícola. La gran mayoría de la producción se realiza en el ciclo Otoño-Invierno (OI), lo cual afecta la serie de tiempo de NDVI total del municipio.

Mientras el mapa de uso de suelos puede revelar patrones interesantes en zonas de gran producción agrícola, existen municipios en zonas de autoconsumo como Guerrero, Oaxaca y Chiapas donde los campos agrícolas son demasiado pequeños como para poder utilizar mapas de INEGI. En dicho caso, el objetivo principal del monitor debe ser la agricultura de temporal, donde la cosecha de maíz se encuentre ligada a las condiciones climatológicas observables.

4.2 Relación Visual NDVI y producción de maíz

Por último, en la figura 6 muestra una relación visual de NDVI y producción de maíz para dos municipios. Se incluyen datos de porcentaje de superficie de temporal cosechada y NDVI para Milpa Alta, en Ciudad de México, y Pabellón de Arteaga, en Aguascalientes. En ambos casos se observan patrones: Milpa Alta tiene ciclos estables de NDVI, y su campo consistentemente cosecha la totalidad de la superficie sembrada; Pabellón de Arteaga ha tenido un NDVI y una producción irregular. En azul se incluye también el periodo de LAP en el cálculo de Σ_v .

Se puede apreciar en el año 2009 de la figura (b) que el ciclo de NDVI del año tuvo un retraso de un par de meses, y la producción agrícola sufrió una fuerte pérdida, lo cual sugiere que el NDVI debe ser saludable en el periodo de crecimiento sobre el cual se calcula Σ_v . Los cuatro años que cosecharon la totalidad de su cosecha muestran un patrón similar de crecimiento de NDVI en el periodo LAP .

5 Implementación de Modelo

Para la implementación del modelo se propone un planteamiento Bayesiano para predicción de producción agrícola a nivel municipio, que constituye el mayor grado de desagregación del que se tiene información de producción agrícola. De esta forma, se pueden hacer uso de modelos jerárquicos según la variación que presenten las variables: regional o temporal. Para un municipio m y un año t , el modelo propuesto es de la forma

$$Y_{m,t} \sim Gamma\left(\phi, \frac{\phi}{\mu_{m,t}}\right) \quad (5)$$

Donde,

$$\log(\mu_{m,t}) = \alpha_m + \beta_m * \Sigma_{V_{m,t}} \quad (6)$$

Y $\Sigma_{V_{m,t}}$ se define por las variables latentes,

$$\Sigma_{V_{m,t}} = \sum_{i=T_{0t}+lag}^{T_{0t}+lag+LAP} (NDVI_i - NDVI_{T_{0t}}) \quad (7)$$

Por su parte, las variables T_{0t} , α_m y β_m se definen de forma jerárquica para temporalidad y municipios:

$$T_{0t} \sim Cat(p_{T_0}) \quad (8)$$

$$\alpha_m \sim Normal(\alpha, \tau_\alpha) \quad (9)$$

$$\beta_m \sim Normal(\beta, \tau_\beta) \quad (10)$$

Las variables latentes son de la forma:

$$LAP \sim Cat(p_{LAP}) \quad (11)$$

$$lag \sim Cat(p_{lag}) \quad (12)$$

Por último, las distribuciones a priori son de la forma:

$$p_R \sim Dirichlet(\alpha_R) \quad (13)$$

Y ϕ , τ_α , τ_β , α , β contienen distribuciones a priori no informativas. Por último, para cada α_R para LAP , lag y T_0 se puede tomar información conocida que incorpore información conocida sobre cada caso. Pata T_0 se puede ajustar a el inicio de simebra, y para LAP a la duración esperada que dura el crecimiento del maíz.

5.1 Resultados Ejemplo

Utilizando un esquema de muestreo MCMC a través de JAGS 4.1 y R 3.3.1 se obtuvieron estimaciones de las distribuciones posteriores para cada municipio. En la figura 7 se muestra la distribución predictiva de producción agrícola para nueve municipios contiguos en Oaxaca. En este caso, se utilizaron datos de producción agrícola de 2009 a 2014 para ajustar el modelo, y se predijo la distribución de 2015. En cada caso, las líneas en negro muestran la producción realmente obtenida en ese periodo; en todos los casos presentados la producción real se encuentra en zonas de alta probabilidad posterior.

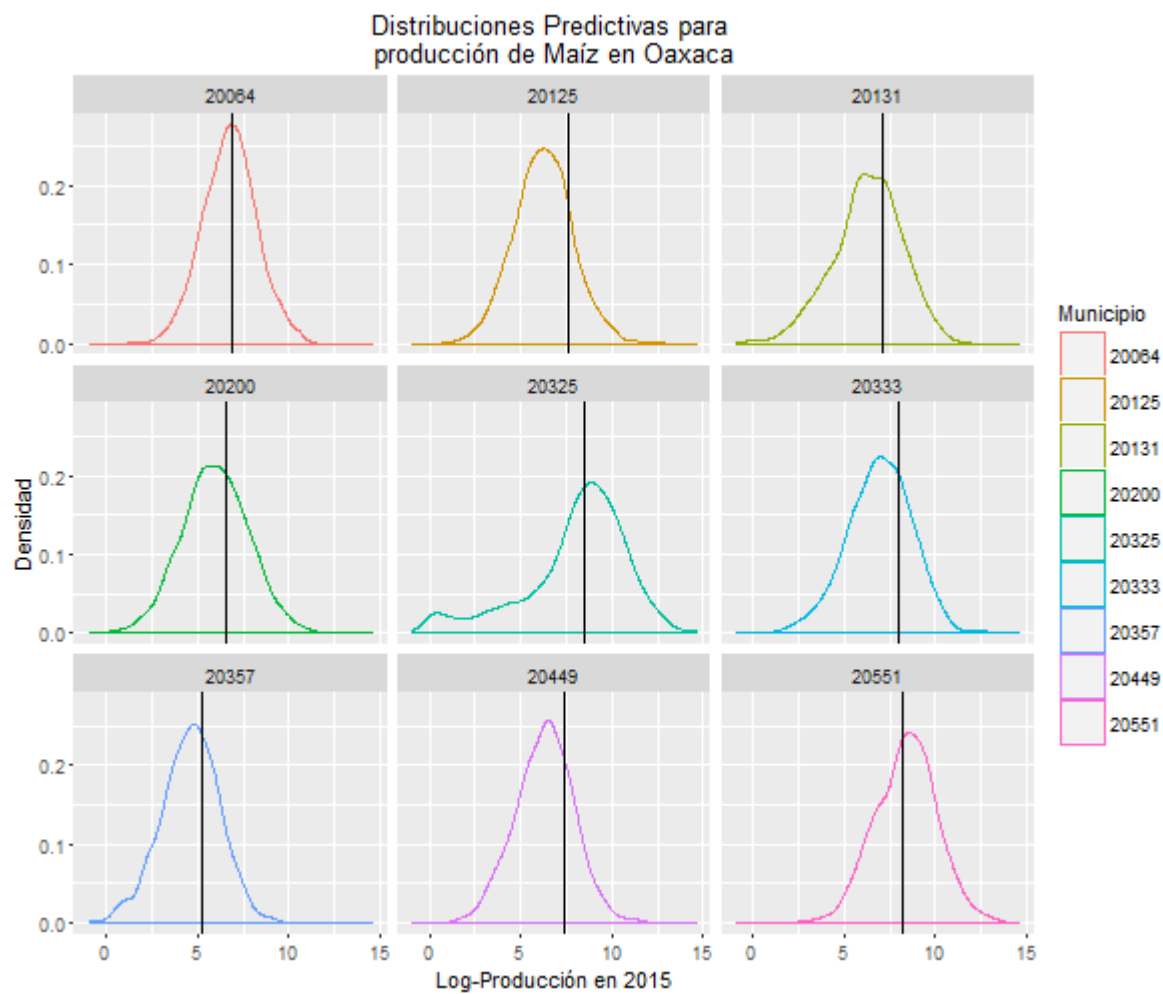


Figure 7: Distribuciones predictivas de 2015 para predicción de maíz para 9 municipios en una misma región en Oaxaca. Las líneas representan la producción real para cada municipio en 2015

References

- [David López-Carr, 2016] David López-Carr, e. a. (2016). Climate-related child undernutrition in the lake victoria basin: An integrated spatial analysis of health surveys, ndvi, and precipitation data.
- [Funk and Budde, 2009] Funk, C. C. and Budde, M. E. (2009). Phenologically-tuned modis ndvi-based production anomaly estimates for zimbabwe.
- [Funk and Verdin, 2010] Funk, C. C. and Verdin, J. P. (2010). Real-time decision support systems: the famine early warning system network.
- [Molly E. Brown, 2008] Molly E. Brown, e. a. (2008). Using satellite remote sensing data in a spatially explicit price model: Vegetation dynamics and millet prices.
- [Nagler P. L., 2005] Nagler P. L., e. a. (2005). Evapotranspiration on western u.s. rivers estimated using enhanced vegetation index from modis and data from eddy covariance and bowen ratio flux towers.
- [Roderick, 1996] Roderick, M, e. a. (1996). Calibrating long term avhrr- derived ndvi imagery.
- [Swets, 1999] Swets, e. a. (1999). A weighted least-squares approach to temporal smoothing of ndvi.

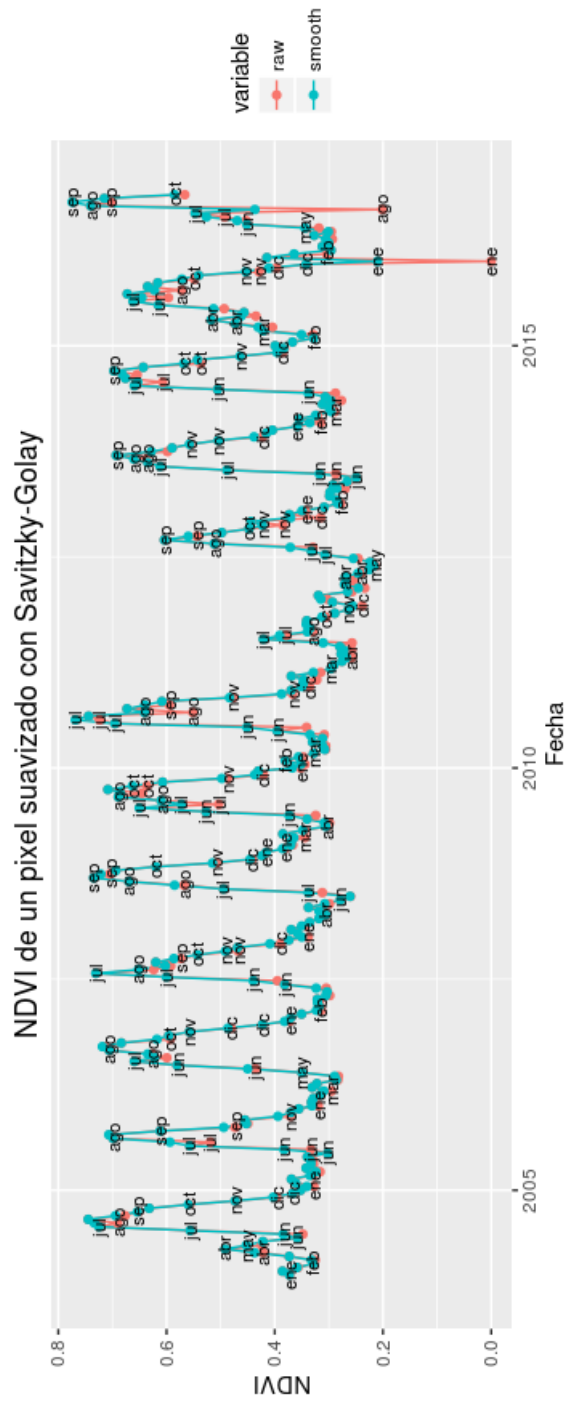
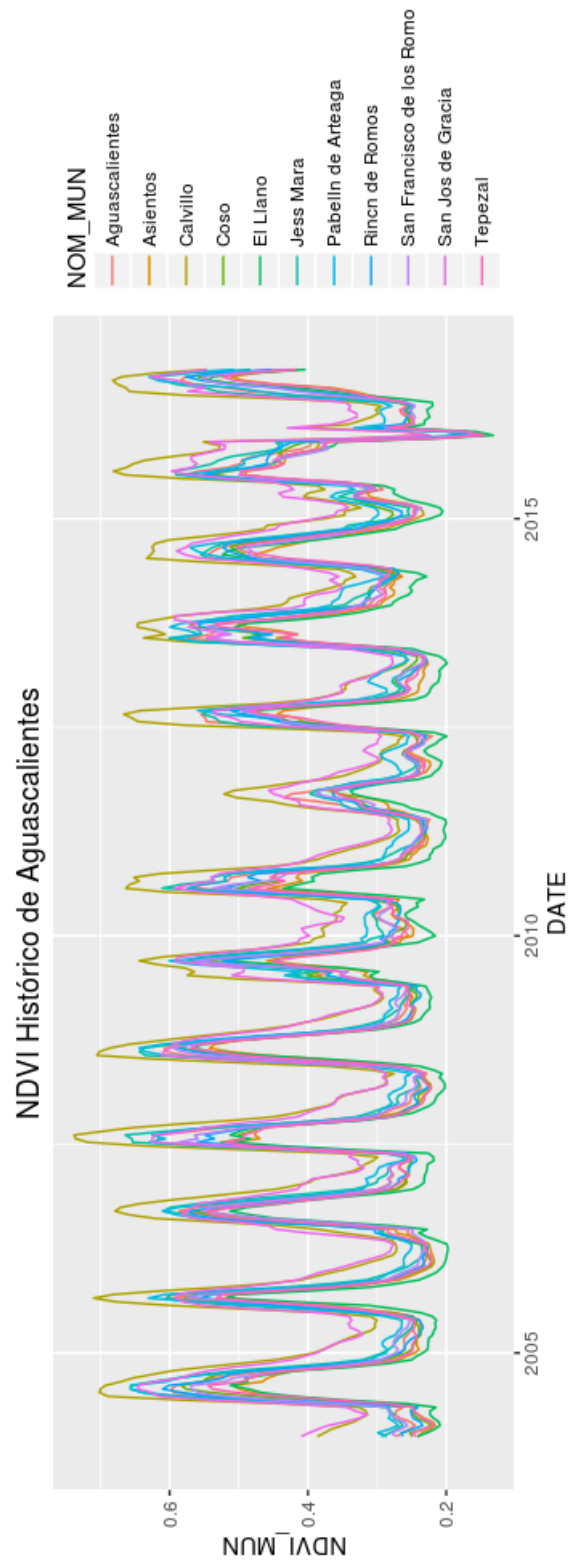
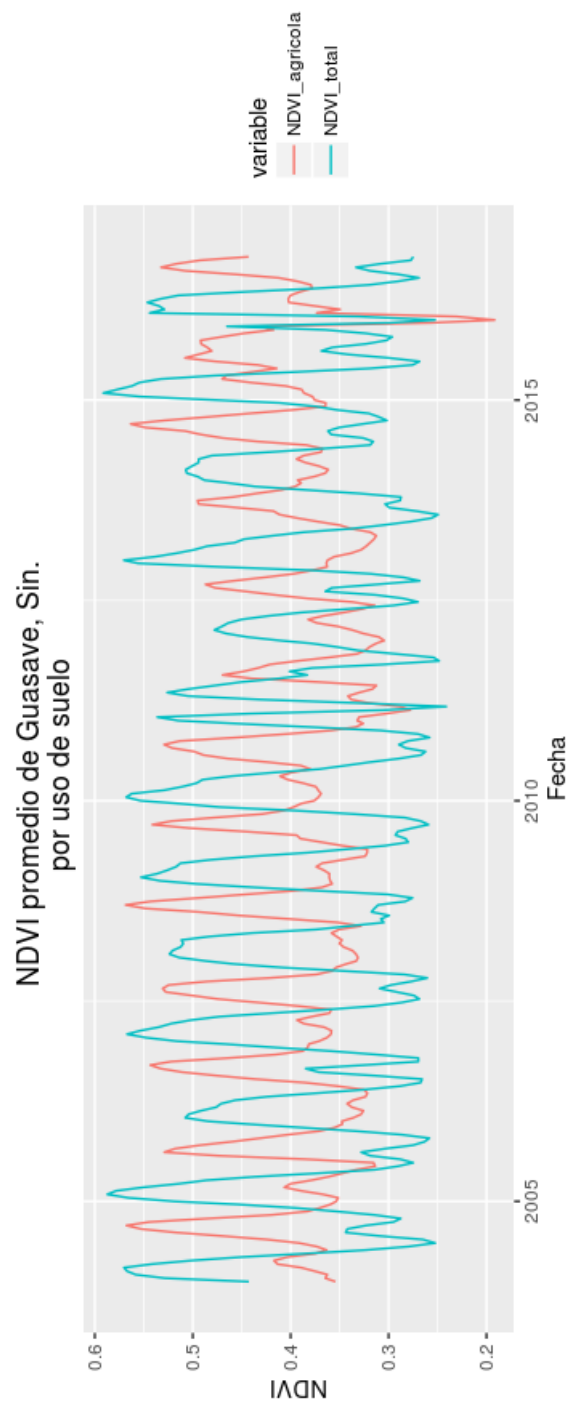


Figure 1: Comparación de NDVI con y sin suavizamiento



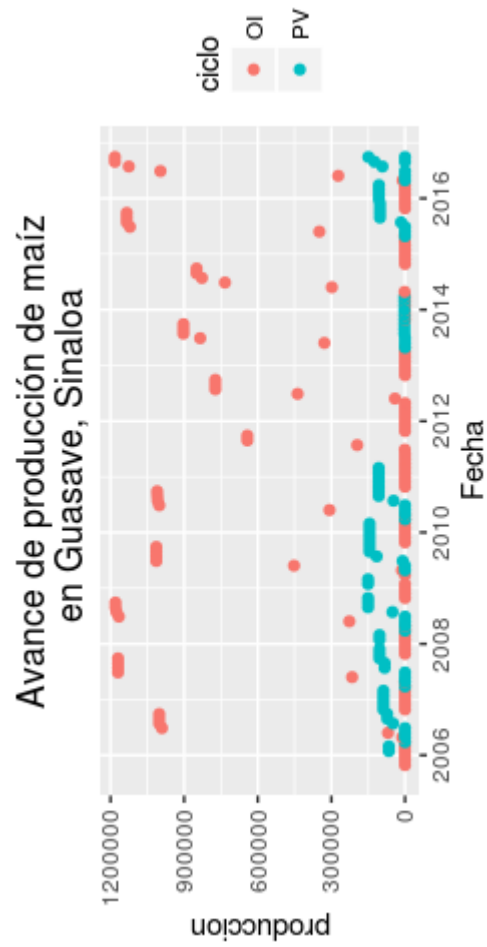
Elaboración propia con datos de MODIS e INEGI

Figure 2: NDVI promedio por municipio en Aguascalientes



Elaboración propia con datos de MODIS e INEGI

Figure 4: NDVI por uso de suelo en Guasave, Sinaloa



Elaboración propia con datos de SAGARPA

Figure 5: Producción de maíz por ciclo agrícola en Guasave, Sinaloa. La totalidad de producción es de riego. OV indica ciclo Otoño-Invierno, y PV, Primavera-Verano.

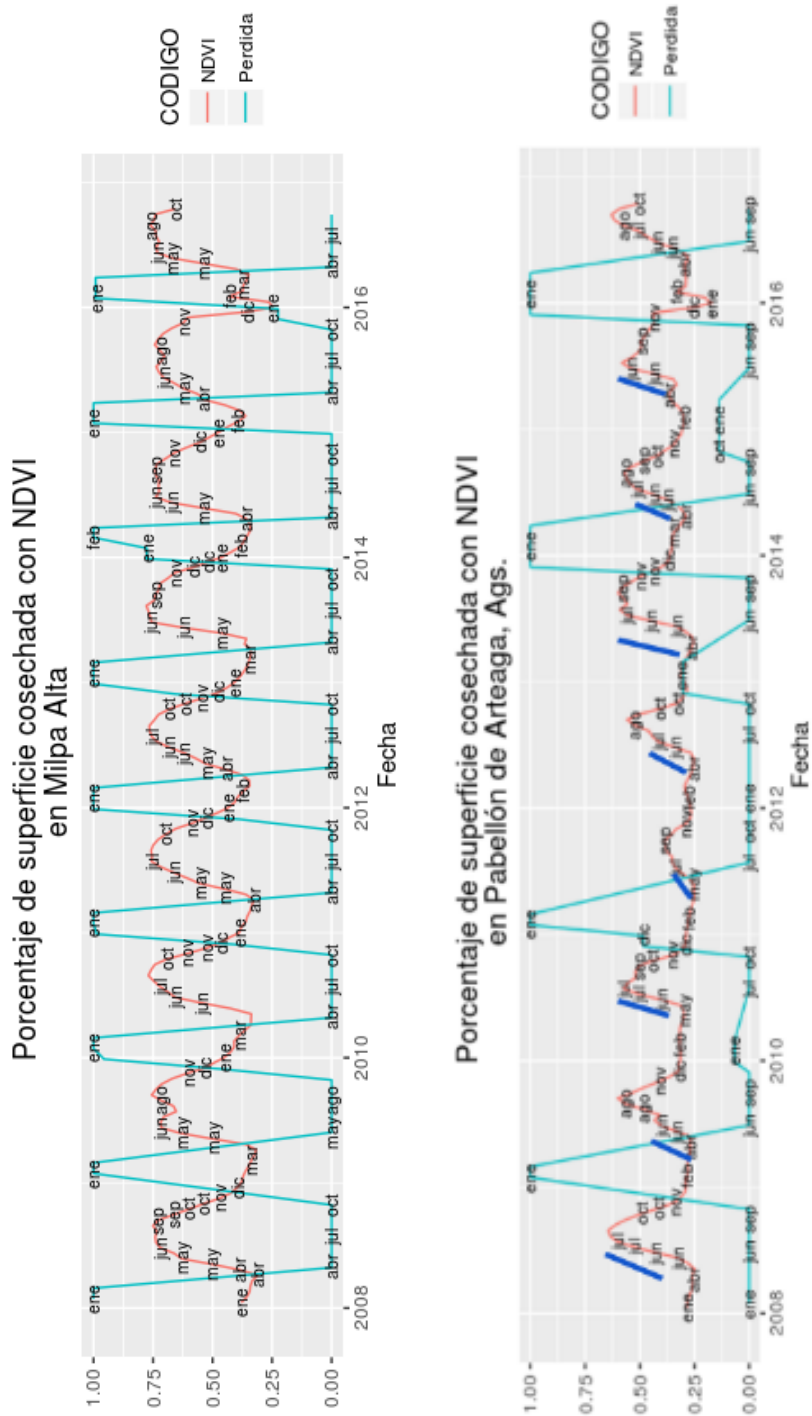


Figure 6: En (a) Milpa Alta, y (b) Pabellón de Arteaga. En (b) se incluye también una línea azul que indica el periodo sobre el cual se calcula Σv , que comienza durante o poco después de mayo, y termina a principios de agosto o finales julio, según el año.