

Clasificación de Sistemas Productivos Preponderantes utilizando técnicas de agrupamiento para la estimación de emisiones de Gases de Efecto Invernadero.

Lozza, Anabella^{1,2}; Bellini Saibene, Yanina ¹; Lorda, Héctor ¹

¹ Estación Experimental Agropecuaria Anguil, "Ing. Agr. Guillermo Covas",
Ruta Nacional N° 5, Km 580, 6326, Anguil, La Pampa

² CONICET

{lozza.anabella,bellini.yanina,lorda.hector}@inta.gob.ar

Abstract. *This work uses k-means and k-medoids to process data of the 2014 Provincial Agricultural Register Survey from microregion 5 (La Pampa) to obtain preponderant production systems. K-medoids yielded the best results finding two groups with a stability index > 0.78. These groups represent a cow-calving rearing fattening bovine system on native grassland and a second one of cow-calving rearing fattening on native grassland and cultivated pastures. The emissions of both groups are similar, being bovine the greatest contribution with cows explaining 75% within this category. The detail achieved by each group will allow the calculation of emissions to be improved by incorporating the agricultural component in future works.*

Resumen. *Este trabajo utiliza k-means y k-medoids en los datos del Registro Provincial Agropecuario de la microrregión 5 (La Pampa) del año 2014 para obtener sistemas productivos preponderantes. K-medoids arrojó los mejores resultados encontrando dos grupos con un índice de estabilidad > 0.78. Estos grupos representan un sistema de cría- recría bovina sobre pastizal natural y otro de cría-recría bovina sobre pastizal natural y forrajeras cultivadas. Las emisiones de ambos grupos son similares, siendo bovinos el de mayor aporte con las vacas explicando el 75% dentro de esta categoría. El detalle logrado por grupo permitirá mejorar el cálculo de emisiones incorporando el componente agrícola en futuros trabajos.*

Keywords: *clustering, agro data mining, data minig, AgICT, models*

Palabras claves: *clustering, agro data mining, minería de datos, agrotics, modelos*

1. Introducción

La provincia de La Pampa presenta una multiplicidad de sistemas de producción, debido a la disponibilidad y calidad de los recursos naturales y a la diversidad de estructuras productivas, organizaciones de la producción y productores (Caviglia et al., 2010). Este contexto hace necesario estudiar, dentro de las posibilidades que ofrece la información disponible (Caviglia et al., 2010), cuales son los sistemas de producción preponderantes,

entendidos como distintas combinaciones de actividades agropecuarias que más se utilizan en una zona o región.

Conocer estos sistemas de producción, permite contar con información objetiva de la realidad regional, la cual se convierte en un insumo para a) acciones de gestión institucional (ej.: priorización de problemas, determinación de líneas de trabajo. (Caviglia et al., 2010)) b) acciones de transferencia y extensión (Ej.: elaboración de indicadores económicos y técnicos productivos a nivel de sistema y de actividades (Bellini Saibene et al., 2013)) y c) acciones de investigación y desarrollo: insumo para poder asignar un valor de emisión de gases de efecto invernadero (GEI) en distintas cadenas industriales del agro para resolver de manera efectiva problemas de cambio climático asociados al agro-negocio (Frank, et. al., 2014). La cría bovina explica gran parte de las emisiones GEI de la actividad ganadera y aparece como actividad prevalente en la microrregión 5 (Caviglia et al., (2010); Ministerio de la Producción. Gobierno de La Pampa, (2012)).

Existen trabajos previos de identificación y caracterización de sistemas productivos, tanto en La Pampa (Caviglia et al., (2010); Bellini Saibene et al., (2013)) como en otras provincias y regiones (Van den Bosch et. al., (2011); R. de Rodríguez, et. al., (2008); Marinissen, Josefina et al., (2009); Coronel Derenolfi & Ortuño Pérez, (2005); Zehnder, R. & Quaino, O., (2005); Calvi, Mariana et al., (2011); y Ghida Daza, Carlos Alberto, (2006)). La mayoría de los trabajos utiliza datos de los censos nacionales agropecuarios (CNA) y la metodología predominante es la sugerida por (Ghida Daza, Carlos Alberto, 2006). Para La Pampa, también se cuenta con información del Registro Provincial Agropecuario (REPAGRO), una declaración jurada, enfocada a relevar estadísticas agropecuarias, que deben realizar anualmente todos los productores agropecuarios pampeanos. Este relevamiento es coordinado desde la Dirección General de Estadística y Censos (DGEC) del Ministerio de la Producción de la provincia.

En los antecedentes se realiza una división en zonas agroecológicas homogéneas previo al análisis de los sistemas productivos, basados en el hecho que los recursos naturales disponibles condicionan el tipo de actividad que se puede llevar adelante. En La Pampa, se definieron microrregiones como estrategia central para el desarrollo territorial. El INTA también ha definido su estrategia institucional a nivel de los territorios que se definen a partir de las microrregiones provinciales (Bellini Saibene et al., 2013). Como estas microrregiones contemplan divisiones políticas además de agroecológicas, algunos estudios de sistemas productivos como (Caviglia et al., (2010); Bellini Saibene et al., (2013); Iglesias, et. al., (2016)) unifican estas microrregiones de acuerdo a criterios ambientales y agropecuarios definidos en otros trabajos como (Lorda, Héctor et. al., (2008); Campos Carlés, Silvina et al., (2010)).

El objetivo de este trabajo es identificar y describir sistemas de producción preponderantes en la microrregión 5 de la provincia de La Pampa (figura 1) utilizando técnicas de agrupamiento para calcular las emisiones GEI.

El uso de estas técnicas se justifica porque: a) los datos del análisis son observacionales por lo que fueron recolectados para un fin diferente al del estudio a realizar, b) no se cuenta con un conjunto de datos previamente etiquetado, así que se debe buscar productores que se comporten de forma similar desde el punto de vista productivo a partir del análisis de la información disponible, c) es necesario obtener aquellos grupos

más importantes (ya sea por mayor cantidad, superficie ocupada o cabezas manejadas) y d) es necesario obtener una descripción de las características productivas de cada grupo.

Las técnicas de agrupamiento intentan dividir los casos en grupos naturales. Estos grupos, presumiblemente, reflejan algún patrón donde los objetos en un grupo son más similares entre sí y los objetos en grupos separados son menos similares entre sí. El conocimiento de los grupos permite una descripción sintética de un conjunto de datos multidimensional complejo, sustituyendo la descripción de todos los elementos del grupo por la de un representante característico del mismo. También posibilita conocer cuáles son las variables que marcan las diferencias entre los grupos y generar un modelo que clasifique nuevos casos (Bellini Saibene et al., (2013); Witten, Frank, & Hall M. A., (2011)).

La Huella de Carbono (HC) es una medida que permite estimar las emisiones de GEI asociadas a las actividades y bienes producidos, comercializados, y utilizados por los humanos. Siendo el cambio climático una preocupación mundial, este indicador adquiere especial importancia.

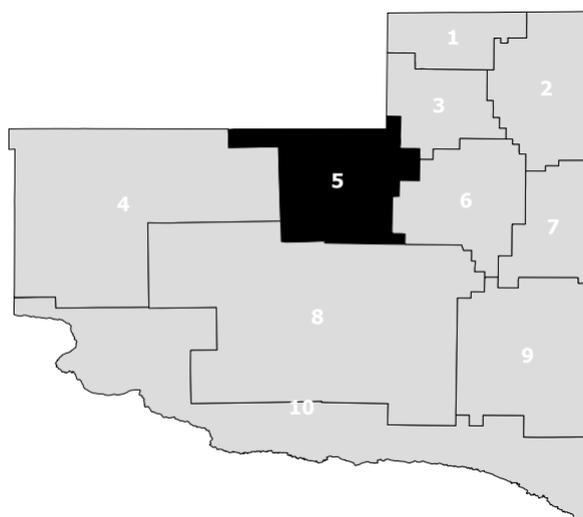


Figura 1. División por microrregiones de La Pampa (fuente: elaboración propia en base a (Dirección de Estadística y Censo de La Pampa, 2015)

La certificación ecológica de productos agrarios, permite a los consumidores decidir qué alimentos comprar en base a la contaminación generada como resultado de los procesos por los que ha pasado (Viglizzo et. al., 2006). Esta tendencia ya es una realidad en la Unión Europea, quienes a partir del 2016 aplican mayores restricciones relacionadas con la problemática ambiental; monitoreando la HC en diferentes cadenas. La HC genera agregado de valor de los productos, en estrategias de marketing internacional, posibilitando un mejor posicionamiento en los mercados.

El ganado, particularmente el vacuno por su importancia cuantitativa, contribuye a la emisión de GEI, de manera directa (a partir de la fermentación entérica y el aprovechamiento del estiércol) e indirecta (actividades de producción y conversión de bosques en pastizales) (Hristov et al., 2013).

El sector ganadero emite GEI de mayor potencial de calentamiento que el dióxido de carbono (CO₂). El proceso de fermentación ocurrido en la digestión entérica de los

rumiantes genera metano (CH₄), un gas que tiene un poder de calentamiento global 21 veces mayor al CO₂ y las heces bovinas aportan el óxido nitroso (N₂O), con un poder de calentamiento global 310 veces mayor que el CO₂ (Solomon, 2007). La actividad pecuaria emite el 37% y 65% del CH₄ y N₂O antropógeno mundial, respectivamente (Steinfeld et al., 2009). En Argentina, la producción ganadera representa el 17% de las emisiones totales, que sumadas a las emisiones relacionadas a la producción de forrajes, este porcentaje alcanza el 25%. La cadena de carne bovina en La Pampa tiene presencia en toda la geografía provincial, siendo el sector primario casi el 50% del Producto Bruto Geográfico del sector agropecuario (Iglesias, et. al., 2010). Esta actividad es un gran contribuidor de GEI a escala provincial y fuente potencial de acciones locales de mitigación. Anticiparse a futura imposición (competencia) de la trazabilidad de emisiones para la cadena de carne bovina, sería apropiado. Comenzando a medir el desempeño ambiental de las producciones locales, a través de una metodología de estimación de GEI a nivel predial.

2. Materiales y Métodos

2.1. Los datos

El presente estudio se realizó utilizando 62 variables registradas en el año 2014 por el REPAGRO (tabla 1). Esta información es de libre disponibilidad, anualmente se publica un informe técnico realizado por la DGEC, quien además pone a disposición de los interesados la base de datos correspondiente. Al ser de cumplimiento obligatorio y tener carácter de declaración jurada, se asegura la disponibilidad de la información en el tiempo. Estas características, permiten considerarlo como una herramienta estratégica para el desarrollo y posterior aplicaciones de los modelos resultantes.

Las variables seleccionadas se han registrado en todas las ediciones del REPAGRO, han sido utilizadas en trabajos previos de caracterización e identificación de sistemas productivos (Bellini Saibene et al., (2013); Caviglia et al., (2010); R. de Rodríguez, Mabel G. et al., (2008); Zehnder, R. & Quaino, O., (2005); Calvi, Mariana et al., (2011); Campos Carlés, Silvina et al., (2010)) e indagan sobre aspectos productivos (indican qué actividades hace el productor) a partir de contabilizar la cantidad de cabezas de cada categoría animal por especie y la cantidad de hectáreas por uso del suelo. Las categorías animales se definen dentro de una especie (ej.: bovinos, porcinos) según la edad, el sexo y el uso del animal, así un novillo se diferencia de un toro porque el primero fue castrado y tiene un fin de producción de carne, mientras que en el otro su funcionalidad es reproductiva. Los usos del suelo, hacen referencia a que tipo de ocupación tiene la superficie, como por ejemplo cultivos, médanos, construcciones, etc. Los cultivos se diferencian en aquellos que se realizan para cosecha y venta del grano (ej.: trigo para cosecha fina, girasol para cosecha gruesa) y aquellos que se realizan para alimentar al ganado (ej.: verdeos de invierno como el centeno, verdeos de verano como el maíz y pasturas como la alfalfa).

Tabla 1. Listado de variables originales utilizadas en el procesamiento.

| Tema | Detalle |
|-------------------------|---|
| Existencias bovinas (9) | total de cabezas por categoría (terneras menor 1 año, terneros menor 1 año, vaquillonas 1 a 2 años, vaquillonas |

| | |
|----------------------------|---|
| | mayor 2 años, vacas, novillos 1 a 2 años, novillos mayor 2 años, toritos 1 a 2 años, toros mayor 2 años) |
| Existencias ovinas (7) | Total de cabezas por categoría (ovejas, corderos, corderas, borregos, borregas, capones y carneros) |
| Existencias caprinos (3) | Total de cabezas por categoría (madres, chivitos, castrones) |
| Existencias porcinas (5) | Total de cabezas por categoría (cerdas, lechones, cachorros, capones, padrillos) |
| Existencias equinas (2) | Total de cabezas por categoría (machos y hembras) |
| Cultivos de cosecha (13) | total de hectáreas por cultivo (trigo, avena para grano, cebada, centeno, otros fina, girasol, maíz, sorgo granífero, soja, maní, otros gruesa, rastrojo, barbecho) |
| Cultivos para forraje (21) | total de hectáreas por cultivo (maíz verdeo, maíz diferido, mijo, sorgo verdeo, sorgo diferido, avena, cebada, centeno, alfalfa pura, alfalfa consociada, otras leguminosas puras, otras leguminosas consociadas, pasto llorón, digitaria, panicum, gramíneas puras, monte natural, campo natural sin monte, otros verdeos de invierno, otros verdeos de verano, otros cultivos diferidos). |
| Total de hectáreas (1) | |
| Número identificador* | Se conserva para identificar los casos de cada grupo |
| Microrregión | Se usa para filtrar los casos |

2.2. Análisis de los datos

El REPAGRO 2014 contiene 7.766 casos para toda la provincia, los cuales se clasifican según su actividad, distribuyéndose: 1- Ganadería: 3.331, 2- Agricultura: 316, 3- Tambo: 71,4- Mixto: 2.478, 5-Mixto con tambo: 40, 6- Pastoreo o contratista: 1.156 y, 7 y 8- Futuras operaciones /Sin actividad: 374. Los casos categorizados como “7 u 8” no presentan datos porque no han realizado actividades productivas en el período del registro o no son establecimientos activos, por lo que se descartan del estudio. Los casos con actividad “6 - Pastoreo o contratistas” son productores que realizan sus actividades en un campo ajeno, el cual referencian en el formulario.

La microrregión 5 tiene 422 casos totales (4,8% de la provincia), 33 casos con actividad 7 y 8 y 90 casos con actividad 6, lo que resulta en 299 casos a procesar. Los casos con actividad 6 se analizaron particularmente para chequear la consistencia entre los datos declarados por el dueño de la tierra y los datos declarados por el productor que realiza la actividad. De los 90 casos, 41 presentaron inconsistencias (45,5%), las cuales pudieron ser subsanadas con ayuda de la DGEC.

2.2.1. Análisis estadístico

Se calcularon los estadísticos descriptivos y se realizaron los diagramas de caja de las 62 variables seleccionadas, utilizando R. No hay valores faltantes. La desviación estándar caracteriza una muestra con gran dispersión en todas las variables analizadas. La mayoría de las variables presentan valores atípicos, especialmente porque las magnitudes utilizadas (cabezas y hectáreas) tienen que ver con la disponibilidad de

superficie para trabajar y hay mucha diferencia entre las explotaciones agropecuarias de menor tamaño y los de mayor superficie (rango > 28.000 ha). El análisis de la variable superficie total, mostró 41 explotaciones de tamaño muy reducido en hectáreas (< 100 ha); las cuales en trabajos previos (Bellini Saibene et al., (2013); Caviglia et al., (2010)) son descartadas de estos análisis por ser menores a la unidad económica de la microrregión y encontrarse cercanas a las ciudades, por lo cual entran en otra lógica de producción, más urbana y periurbana, que rural. Para este análisis se descartaron del set de datos, previa revisión geográfica de la ubicación catastral coincidente con influencia por cercanía a los municipios, quedando en total 264 casos para analizar.

2.2.2. Variables Calculadas

Basado en las características observadas en los datos, sumada a la experiencia de trabajos anteriores (Bellini Saibene et al., 2013) se realizaron cálculos de proporciones para las actividades. Como se buscan sistemas productivos, es necesario determinar qué hace el productor y qué importancia tiene esa actividad en su sistema. El cálculo de proporciones es el más adecuado a la hora de representar esa decisión, antes que las hectáreas o la cantidad de cabezas. La tabla 2 presenta las proporciones e índices y el cálculo correspondiente.

Tabla 2 Variables de proporciones e índices.

| Variable y Fuente | Cálculo |
|---|--|
| Porcentaje de cultivos de cosecha fina (Bellini Saibene et al., 2013) | $(\text{Trigo} + \text{Avena} + \text{Cebada} + \text{Centeno} + \text{Otros}) / \text{Total de Hectáreas}$ |
| Porcentaje de cultivos de cosecha gruesa (Bellini Saibene et al., 2013) | $(\text{Girasol} + \text{Maíz} + \text{Sorgo Granífero} + \text{Soja} + \text{Maní} + \text{Otros}) / \text{Total de Hectáreas}$ |
| Porcentaje de forrajeras anuales (Bellini Saibene et al., 2013) | $(\text{Maíz} + \text{Mijo} + \text{Sorgo} + \text{Avena} + \text{Cebada} + \text{Centeno} + \text{otros verdes de invierno} + \text{otros verdes de verano}) / \text{Total de Hectáreas}$ |
| Porcentaje de forrajeras perennes (Bellini Saibene et al., 2013) | $(\text{Alfalfa Pura} + \text{Alfalfa Consociada} + \text{Otras Leguminosas Puras} + \text{Otras Leguminosas Consociadas} + \text{Pasto Llorón} + \text{Digitaria} + \text{Panicum} + \text{Gramíneas Puras}) / \text{Total de Hectáreas}$ |
| Porcentaje de Caldenal (Bellini Saibene et al., 2013) | $(\text{Caldenal} + \text{Jarillar} + \text{Renoval}) / \text{Total de Hectáreas}$ |
| IndiceOrientacionGanadera | $\text{Total Bovinos} / (\text{Total bovinos} + \text{Total ovinos} + \text{Total porcinos} + \text{Total caprinos} + \text{Total equinos})$ $\text{Total ovinos} / (\text{Total bovino s} + \text{Total ovinos} + \text{Total porcinos} + \text{Total caprinos} + \text{Total equinos})$ $\text{Total porcinos} / ((\text{Total bovinos} + \text{Total ovinos} + \text{Total porcinos} + \text{Total caprinos} + \text{Total equinos}))$ $\text{Total caprinos} / (\text{Total bovinos} + \text{Total ovinos} + \text{Total porcinos} + \text{Total caprinos} + \text{Total equinos})$ $\text{Total equinos} / ((\text{Total bovinos} + \text{Total ovinos} + \text{Total porcinos} + \text{Total caprinos} + \text{Total equinos}))$ |
| Índice Novillo (Fernández, Julio César et al., 2010) | $(\text{Total Novillo de 1 a 2 años} + \text{Total Novillos} > 2 \text{ años}) / \text{Total Cabezas}$ |

$$\text{Índice de Cría (Bellini Saibene et al., (Total Vacas + Total Vaquillonas de 1 a 2 años + Total 2013) Vaquillonas >2años) + Total Toros + Total Toritos)/Total Cabezas}$$

La tabla 3 presenta la media, mediana, y desviación estándar de estas variables. Se aprecia que se representa correctamente los usos del suelo, siendo los de cosecha los de menor presencia (casi nula) y las forrajeras perennes y el bosque los de mayor presencia. En los índices también se aprecia la importancia de los bovinos sobre el resto de las especies en el aspecto ganadero. Los índices asociados a la cría son los más importantes; lo que es coherente con las posibilidades productivas de la zona bajo estudio. Los valores presentes en la desviación estándar muestran dispersiones importantes en las variables asociadas a la ganadería. Con excepción de PorBosque el resto presenta valores atípicos.

Tabla 3. Estadísticos descriptivos variables de proporciones e índices

| Variables | Media | Mediana | Desviación estándar |
|----------------------------------|--------|---------|---------------------|
| PorFina | 0.0005 | 0.0000 | 0.0061 |
| PorGruesa | 0.0003 | 0.0000 | 0.0055 |
| PorForrajeraAnual | 0.0320 | 0.0000 | 0.0881 |
| PorForrajeraPeremne | 0.0532 | 0.0000 | 0.1374 |
| PorBosque | 0.6881 | 0.8133 | 0.3259 |
| IndiceOrientacionGanaderaOvina | 0.0656 | 0.0000 | 0.1564 |
| IndiceOrientacionGanaderaPorcina | 0.0040 | 0.0000 | 0.0320 |
| IndiceOrientacionGanaderaCaprina | 0.0036 | 0.0000 | 0.0210 |
| IndiceOrientacionGanaderaEquina | 0.0211 | 0.0079 | 0.0694 |
| IndiceOrientacionGanaderaBovina | 0.8979 | 0.9823 | 0.1997 |
| IndiceCria | 0.6526 | 0.6670 | 0.2563 |
| IndiceNovillo | 0.0537 | 0.0000 | 0.1471 |

2.3. Clasificación

Para correr los algoritmos se generaron tres set de datos, todos con la misma cantidad de casos, pero con diferentes combinaciones de variables: 12 variables en el primero (Tabla 1), en el segundo se quita la variable PorGruesa y en el tercero PorGruesa y PorFina, debido a la presencia casi nula de estas dos actividades. Las variables fueron estandarizadas para que sean la media cero y la varianza uno.

Se utilizaron los algoritmos k-means, ya utilizado en Bellini Saibene et al., (2013) y k-medoids porque es más robusto al ruido y a los valores extremos, implementados en R. Ambos algoritmos necesitan como dato de entrada la cantidad de grupos a buscar. Debido a las características productivas de la zona y las variables disponibles, se presupone que pueden existir a lo sumo dos sistemas productivos (Caviglia et al., (2010); Lorda, Héctor et al., (2008)). Para contar con más elementos para determinar la cantidad de grupos, se realizó por cada dataset un gráfico de la suma de cuadrados dentro de cada grupo generado con kmeans desde 2 a 10 grupos, buscando en la gráfica una pequeña curva ascendente, luego de un marcado descenso de los valores, el número de grupos posibles está antes de esa pequeña curva (Hennig, 2010). De acuerdo a los 3 gráficos generados la cantidad es de tres grupos o menos. También se utilizó la función de R pamk (que implementa kmedoids) (Hennig, 2010), que estima la cantidad de grupos, dando como resultado un sólo grupo para los tres dataset. Ante esta situación se decidió correr ambos algoritmos con 2 y 3 grupos.

Además, se evaluó la estabilidad de los clusters generados con la función clusterboot() de R. Este algoritmo utiliza el coeficiente Jaccard, una medida de similitud entre conjuntos. Los grupos con un coeficiente Jaccard inferior a 0,6 deben considerarse inestables. Valores entre 0,6 y 0,75 indican que el grupo está midiendo un patrón en los datos, pero no hay mucha certeza acerca de qué casos deben agruparse. Los clusters con valores por encima de 0,85 pueden considerarse altamente estables (es probable que sean clusters reales). Se realizaron 100 corridas con cada dataset, para cada algoritmo (kmeansCBI y pamkCBI) (Hennig, (2010); Hennig, (2007); Hennig, (2008)). K-means utiliza el algoritmo de Hartigan and Wong por defecto, con la distancia Euclídea.

Los resultados obtenidos para cada grupo se analizaron con expertos del negocio zonales como está descrito en Caviglia et al., (2010); Bellini Saibene et al., (2013); Calvi, Mariana et al., (2011); Campos Carlés, Silvina et al., (2010); Houghton, J. T. et. al., (1997). Esto permitió validar que las características del grupo correspondieran a prácticas factibles en la microrregión estudiada y que las relaciones de superficie y actividades son coherentes.

2.4. Estimación de emisiones GEI

Para la estimación de las emisiones de gases CH₄ y N₂O producidas por el ganado de los sistemas productivos, se utilizó la metodología de IPCC (Change, 2006). Se empleó el nivel 2 para Bovinos y el nivel 1 para el resto de las especies ganaderas.

El Nivel 1 requiere datos de stock y regionalización con factores de emisión por defecto (Change, 2006). El Nivel 2 minimiza la utilización de datos por defecto, para lo cual requiere una caracterización detallada de la población bovina, generando un buen nivel de desagregación a efectos de obtener factores de emisión ajustados a las circunstancias locales. Para este estudio, se tomaron los factores de emisión más ajustados a la microrregión 5, correspondientes a la región Semiárida según los Sistemas Modales definidos en (Secretaría de Ambiente y Desarrollo Sustentable, 2012). Asimismo, y atento a que no existen valores de tasas de conversión de CH₄ validadas a nivel local, se tomaron los valores recomendados por (Change, 2006).

Se estimaron las emisiones de CH₄ por fermentación entérica y emisiones de CH₄ y N₂O por manejo de estiércol. También se estimaron las emisiones directas e indirectas

en suelos, que se generan por la deposición de animales en pastoreo, por volatilización de N₂O. Las emisiones indirectas por lixiviación de N₂O no se aplican ya que en la región estudiada no se excede la capacidad de retención de agua del suelo debido a precipitaciones y/o a irrigación (Change, 2006). Los resultados obtenidos fueron convertidos a CO₂ equivalente (CO₂Eq). Esto permite comparar distintas emisiones de GEI en base a su potencial de calentamiento global a 100 años (Houghton et al., 1997). En esta instancia, no se estimaron los gases producidos por la agricultura y otras labores generadas por el productor.

3. Resultados y Discusión

3.1. Modelos

La tabla 4 presenta los resultados de los agrupamientos realizados. K-means logra la estabilidad más alta con el set de datos sin las variables de cosecha y buscando dos grupos, aunque los grupos logrados no son balanceados. K-medoids obtiene la estabilidad más alta con dos grupos, set de datos sin la variable de cosecha gruesa y con grupos balanceados. Para ambos algoritmos la disolución de los grupos es mínima. Los valores de estabilidad de 0.85 y 0.78 en el mejor resultado de K-medoids exhiben dos grupos muy estables. Los valores de silueta de todos los grupos generados no superan el valor 0,59 (valores entre 0,50 y 0,70 indica que una estructura razonable se ha encontrado), presentando alguno de ellos valores negativos (no se encontró estructura en esos datos).

En los informes provinciales (Dirección de Estadística y Censo de La Pampa, (2015); Ministerio de la Producción. Gobierno de La Pampa, (2012)) y estudios anteriores de la zona, (Caviglia et al., (2010); Calvi, Mariana et al., (2015); Roberto, Zinda Edith et al., (1998); Campos Carlés, Silvina et al., (2010)) se presenta un solo sistema productivo preponderante: ganadero de cría, con pequeñas variaciones en la cantidad de cabezas disponibles, los recursos forrajeros para su alimentación y la superficie de las explotaciones. Estos antecedentes, refuerzan la hipótesis que es más probable que existan uno o dos sistemas productivos y no un mayor número de agrupamientos.

Tabla 4. Resumen de resultados de los agrupamientos.

| Algoritmo | DataSet | Grupo | Jaccard | Disoluciones | Casos |
|-----------|---------|-------|----------|--------------|-------|
| K-means 2 | TV | 1 | 0.076227 | 99 | 10 |
| | TV | 2 | 0.829200 | 0 | 254 |
| | TV | 1 | 0.495480 | 58 | 68 |
| K-means 3 | TV | 2 | 0.796465 | 0 | 166 |
| | TV | 3 | 0.661954 | 27 | 30 |
| K-means 2 | SG | 1 | 0.845245 | 0 | 255 |

| | | | | | |
|-------------|----|---|------------------|----|-----|
| | SG | 2 | 0.079358 | 97 | 9 |
| | SG | 1 | 0.755686 | 13 | 30 |
| K-means 3 | SG | 2 | 0.503836 | 58 | 57 |
| | SG | 3 | 0.837895 | 0 | 177 |
| | SF | 1 | 0.915774* | 0 | 233 |
| K-means 2 | SF | 2 | 0.646721 | 31 | 31 |
| | SF | 1 | 0.705572 | 22 | 31 |
| K-means 3 | SF | 2 | 0.332079 | 88 | 22 |
| | SF | 3 | 0.816590 | 1 | 211 |
| | TV | 1 | 0.784063 | 15 | 111 |
| K-medoids 2 | TV | 2 | 0.852501 | 4 | 153 |
| | TV | 1 | 0.698114 | 13 | 83 |
| K-medoids 3 | TV | 2 | 0.822169 | 7 | 153 |
| | TV | 3 | 0.633103 | 35 | 28 |
| | SG | 1 | 0.783270 | 14 | 111 |
| K-medoids 2 | SG | 2 | 0.856370* | 2 | 153 |
| | SG | 1 | 0.693481 | 18 | 83 |
| K-medoids 3 | SG | 2 | 0.815127 | 13 | 153 |
| | SG | 3 | 0.602551 | 41 | 28 |
| | SF | 1 | 0.754939 | 17 | 111 |
| K-medoids 2 | SF | 2 | 0.843958 | 2 | 153 |
| | SF | 1 | 0.523399 | 61 | 133 |
| K-medoids 3 | SF | 2 | 0.523399 | 48 | 64 |
| | SF | 3 | 0.517583 | 72 | 67 |

* Mejores valores. TV: Todas las Variables. SG: sin PorGruesa, SF: Sin PorGruesa y PorFina.

3.2. Grupos finales

La tabla 5 presenta la media de las características productivas de ambos grupos. El Grupo 1 (G1) presenta un sistema de **cría- recría bovina sobre pastizal natural** y el Grupo 2 (G2) un sistema también de **cría-recría bovina pero sobre pastizal natural y forrajeras cultivadas**. Las diferencias más marcadas son el tamaño de las explotaciones (más de 800 ha de diferencia), la cantidad de superficie y variedad del área cultivada

(7% de la superficie versus 40%), el monte natural como recurso (48% del natural versus 92%) y la presencia de ovinos como segunda especie ganadera en importancia.

Tabla 5. Características de los dos grupos seleccionados y estimaciones de EGEI.

| Variable | Media G1 casos 111 | EGEI (tnCO2Eq) | Media G2 casos 153 | EGEI (tnCO2Eq) |
|----------------------------|-----------------------|--------------------|-----------------------|--------------------|
| Superficie (ha) | 1873 | | 2682 | |
| Área Cultivada (ha) | 909 | | 198 | |
| Cosecha Fina (ha) | 2,25 | | 0 | |
| Verdeos de Invierno | 44 | | 12 | |
| Verdeos de Verano | 16,5 | | 2 | |
| Cultivos Diferidos | 14,7 | | 2 | |
| Pasturas Perennes | 364 | | 72,6 | |
| Pasto Llorón | 269 | | 53,6 | |
| Gramíneas Puras | 61 | | 16 | |
| Alfalfa | 30 | | 3 | |
| Otras leguminosas | 4 | | 0 | |
| Barbechos y rastrojos | 59,5 | | 25 | |
| Natural (ha) | 1351 | | 2548 | |
| Monte Natural | 652 | | 2344 | |
| Pastizal Natural | 699 | | 204 | |
| % Área Cultivada | 40% | | 7% | |
| % Forraje Natural | 60% | | 93% | |
| Ganadería (cabezas) | | | | |
| <i>Bovinos (cabezas)</i> | 419 | 396,47 (95,35%) | 545 | 501,63 (98,25%) |
| Terneritas < 1 año | 53 | 0 | 68 | 0,00 |
| Vaquillonas 1 a 2 años | 24 | 18,86 | 36 | 27,81 |
| Vaquillonas >2 años | 21 | 16,62 | 27 | 21,10 |
| Vacas | 240 | 296,93 | 316 | 390,85 |
| Novillos 1 a 2 años | 20 | 27,04 | 20 | 27,04 |
| Novillos > 2 años | 8 | 11,02 | 4 | 5,68 |
| Toritos 1 a 2 años | 3 | 5,14 | 3 | 5,14 |
| Toros > 2 años | 13 | 20,85 | 15 | 24,00 |
| Terneros < 1 año | 37 | 0,00 | 56 | 0,00 |
| Ovinos (cabezas) | 46 | 10,14 (2,44%) | 9 | 1,98 (0,39%) |
| Porcinos (cabezas) | 4 | 0,17 (0,04%) | 1 | 0,04 (0,01%) |
| Caprinos (cabezas) | 2 | 0,45 (0,11%) | 1 | 0,23 (0,04%) |

| | | | | |
|-------------------|------------|--------------|------------|--------------|
| Equinos (cabezas) | 9 | 8,57 (2,06%) | 7 | 6,67 (1,31%) |
| % Bovinos | 87% | | 97% | |

El análisis de informantes calificados (técnicos de INTA), planteó la necesidad de revisar si el promedio es la medida adecuada para caracterizar cada grupo, por aquellas actividades que aparecen con muy pocas hectáreas (ej.: cosecha con 2 ha). Recomendaron agregar una variable que contabilice la cantidad de actividades totales (ej.: cuántas forrajeras realiza) y desagregar las proporciones de las variables *PorForrajeraAnual* y *PorForrajeraPerenne*. Con respecto de las variables bovinas y de forrajes opinan que se comportan de la manera esperada y se ajustan a sistemas productivos existentes en la zona. Finalmente, coinciden con los estudios previos que la cantidad y variedad de grupos posibles no puede exceder los dos sistemas productivos preponderantes para esta microrregión.

La estimación de emisiones en la categoría bovinos, se realizó con el nivel de detalle recomendado por IPCC y se usaron factores de emisión estimados específicamente para Argentina, previa regionalización según zonas agroecológicas. Las emisiones de la ganadería son explicadas casi en su totalidad por los bovinos (más del 95%). Ese porcentaje está distribuido principalmente en la categoría vacas, representando el 74,89% y el 77,92% del G1 y G2 respectivamente.

4. Conclusiones y trabajos futuros

Los grupos analizados tienen características coherentes para la zona de estudio y se condice con trabajos anteriores que han estudiado los sistemas de producción en esa región y las potencialidades de la misma. La descripción sintética de ambos grupos permite iniciar el armado de un inventario para la estimación de emisiones GEI para el seguimiento de la HC.

Las técnicas de agrupamiento, aportan nuevos métodos para realizar las caracterizaciones de los sistemas productivos preponderantes. Los criterios de clasificación obtenidos responden a los datos provistos por los productores agropecuarios y no sobre generalizaciones realizadas desde marcos teóricos económicos-productivos. Esta información puede retroalimentar estas metodologías más empíricas, a partir del análisis de la importancia de las variables para los modelos que pueden ser validados en el territorio al tener la posibilidad de entrevistar a una serie de productores que se agruparon en cada grupo generado.

Se plantea la posibilidad de realizar un análisis de agrupamiento de la provincia de La Pampa en su conjunto, sin tener en cuenta la división por microrregión debido a las características agroecológicas similares de varias microrregiones, apoyado además en los valores de silueta encontrados, que marcan una estructura razonable para algunos grupos pero agrupamientos sin estructura definida en otros. El algoritmo K-medoids se presenta como una alternativa interesante para realizar esta tarea, a la luz de los resultados obtenidos en este trabajo.

La estimación de emisiones de la ganadería, no representa grandes diferencias entre ambos planteos productivos. La información de la encuesta resulta insuficiente para estimar factores de emisión propios para cada grupo. Para lograrlo, sería necesario profundizar con entrevistas a los productores e información local que detalle los datos necesarios para la estimación de Nivel 2 de IPCC para cada categoría animal (condición corporal, ganancia diaria de peso, coeficientes reproductivos, proteína y digestibilidad de la dieta, etc.). Ambos grupos poseen una discordancia notable en la distribución del uso de tierras (cultivado/natural). En este estudio, no se pudieron evaluar las diferencias de stock de carbono contenido en el suelo y las variaciones de posibles entre áreas cultivadas y áreas naturales. No se cuenta, hasta el momento, con un factor de emisión oficial sensible, que permita evaluar las diferencias de emisiones por hectárea de los cultivos. Lograr asignarles un factor de emisión, permitiría obtener resultados que reflejen las diferencias entre los sistemas de forma íntegra.

Finalmente, con los grupos resultantes, se solicitará a la DGEC los datos de contacto para una muestra de los casos de cada grupo con el objeto de entrevistar a los productores. La finalidad de ello, es chequear el ajuste de esos casos a la descripción de los grupos obtenidos y del detalle necesario para mejorar la estimación de emisiones realizada.

Referencias

- Bellini Saibene, Y., Iturrioz, G., & Lorda, H. (2013). Identificación de Sistemas Productivos Preponderantes en una zona de la provincia de La Pampa utilizando técnicas de DataMining. *Anales del CAI 2013. Congreso Argentino de AgroInformática*, 46-60.
- Calvi, Mariana, Cendon, María Laura, Frade, María Virginia, Iturrioz, Gabriela Marina, Nigro, Hugo Marcelo, Pavan, Enrique, Rodríguez Vázquez, Gustavo. (2015). *Caracterización de la producción bovina | INTA: Instituto Nacional de Tecnología Agropecuaria*. Buenos Aires. Recuperado a partir de <http://inta.gob.ar/documentos/caracterizacion-de-la-produccion-bovina>
- Calvi, Mariana, Rodríguez, M., Pueyo, J. M., Sampredo, D., Dupleich, J., & Pizzio, R. (2011). Caracterización productiva y económica del sistema de cría vacuna predominante en la región centro sur de Corrientes y centro norte de Entre Ríos. *INTA*, 477.
- Campos Carlés, Silvina, Correa, Macarena, Covelli, María Paula, Del Río, Juan, Eleicegui, Gastón, Garré, Paula, Walter, Martina. (2010). *Provincia de La Pampa. Análisis del sector agropecuario en relación a la economía provincial*. (p. 61). Movimiento CREA. Recuperado a partir de <http://www.aaglp.org.ar/novedades/analisis-sector-agropecuario-la-pampa-junio2010.pdf>
- Caviglia, J., Lorda, H., & Lemes, J. D. (2010). Caracterización de las unidades de producción agropecuarias en la provincia de La Pampa. *Boletín de divulgación técnica*, (99), 10–15.

- Change, I. P. O. (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Recuperado a partir de https://www.ipcc-nggip.iges.or.jp/meeting/pdffiles/Washington_Report.pdf
- Coronel Derenolfi, M., & Ortuño Pérez, S. F. (2005). Tipificación de los sistemas productivos agropecuarios en el área de riego de Santiago del Estero, Argentina. *Problemas del desarrollo. Revista latinoamericana de economía*, 36(140). Recuperado a partir de <http://revistas.unam.mx/index.php/pde/article/view/7572>
- Dirección de Estadística y Censo de La Pampa. (2015). *Anuario Estadístico de La Pampa. 2015*. (Gobierno de La Provincia de La Pampa). Santa Rosa.
- Fernández, Julio César, Pinciroli, H, Girotti, C., Pordomingo, Aníbal, Caldera, Juan, & Juan, Néstor. (2010). Evolución de las existencias del ganado bovino en la regional La Pampa - San Luis en el período 2005-2009. En *Vinculación y articulación institucional. Información Agropecuaria para la gestión y toma de decisiones*. (Vol. 1, p. 44). INTA.
- Frank, F., Montero, G., Ricard, F., & Viglizzo, E. (2014). *La huella de carbono en la agroindustria* (INTA).
- Ghida Daza, Carlos Alberto. (2006). *Identificación y selección de sistemas productivos agropecuarios en el ámbito de la provincia de Córdoba*. (Interno) (p. 9). Manfredi, Córdoba: INTA.
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1), 258–271.
- Hennig, C. (2008). Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. *Journal of multivariate analysis*, 99(6), 1154–1176.
- Hennig, C. (2010). fpc: Flexible procedures for clustering. *R package version*, 2(2), 0–3.
- Houghton, J. T., Meira Filho, L. G., Lim, B., Treanton, K., Mamaty, I., & others. (1997). Revised 1996 IPCC guidelines for national greenhouse gas inventories. v. 1: Greenhouse gas inventory reporting instructions.-v. 2: Greenhouse gas inventory workbook.-v. 3: Greenhouse gas inventory reference manual. Recuperado a partir de <http://agris.fao.org/agris-search/search.do?recordID=XF2015041434>
- Hristov, A. N., Oh, J., Lee, C., Meinen, R., Montes, F., Ott, T., others. (2013). Mitigation of greenhouse gas emissions in livestock production: A review of technical options for non-CO2 emissions. Recuperado a partir de <https://www.ars.usda.gov/research/publications/publication/?seqNo115=280717>
- Iglesias, D., Ghezan, G., & others. (2010). Análisis de la Cadena de la Carne Bovina en Argentina. *Estudios Socioeconómicos de los Sistemas Agroalimentarios y Agroindustriales*, (5). Recuperado a partir de <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=book2.xis&method=post&formato=2&cantidad=1&expression=mfn=002464>
- Iglesias, D., Iturrioz, G., Lorda, H., Torrado Porto, Roberto, & Fernández, M. D. (2016, junio). Márgenes brutos de los principales productos agropecuarios de la provincia de La Pampa. *Boletín Económico*, 29.

- Lorda, Héctor, Roberto, Zinda Edith, Bellini Saibene, Yanina Noemí, Sipowicz, Andrés Horacio, & Belmonte, María Laura. (2008). Descripción de las zonas y subzonas agroecológicas RIAP. Área de influencia de la EEA Anguil. *Boletín de divulgación técnica*, 96, 40.
- Marinissen, Josefina, Castoldi, Federico José, Larreguy, Vicente, Perlo, Alberto Mario, Agamennoni, Raúl José, Pérez Pizarro, Julián, García, Carlos Daniel. (2009, Enero). Sistemas agropecuarios representativos de Villarino y Patagones. Análisis y propuestas. *Ediciones INTA*, 93.
- Ministerio de la Producción. Gobierno de La Pampa. (2012). Micro-Región 5 [Oficial]. Recuperado 11 de diciembre de 2016, a partir de <http://www.produccion.lapampa.gov.ar/micro-region-5.html>
- R. de Rodríguez, Mabel G., Ferrer, José Luis, Engler, Patricia L., & Cancio, Ricardo A. (2008). Identificación y caracterización de los sistemas de producción relevantes (SPR) en Entre Ríos según datos del censo nacional agropecuario 2002 - INDEC. *Serie Extensión -INTA Paraná*, 48, 87.
- Roberto, Zinda Edith, Viglizzo, Ernesto, Buschiazzo, D., Golberg, A., Pordomingo, A., & Frank, Oscar. (1998). Mapa de uso potencial de la tierra en La Pampa. Santa Rosa: INTA.
- Secretaría de Ambiente y Desarrollo Sustentable. (2012). *Tercera Comunicación Nacional de la República Argentina a la Convención Marco de las Naciones Unidas sobre Cambio Climático*. Recuperado a partir de <http://ambiente.gob.ar/tercera-comunicacion-nacional/>
- Solomon, S. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge University Press. Recuperado a partir de <https://books.google.com/books?hl=es&lr=&id=8-m8nXB8GB4C&oi=fnd&pg=PA339&dq=Chapter+2+IPCC+2007+the+physical&ots=hznpx4pgJ-&sig=9BqDKJjd2Ia0SRkioiOi8nmKQGc>
- Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., & de Haan, C. (2009). *La larga sombra del ganado problemas ambientales y opciones. Organización de las Naciones Unidas para la Agricultura y la Alimentación. Roma. 431 p. United Nations Population Fund. UNFPA*. www.unfpa.org (Consulta: junio 2013).
- Van den Bosch, María Eugenia, Silva, Jorge Horacio, & Salvarredi, Guillermo. (2011). Modelización de sistemas productivos agropecuarios relevantes y priorizados. *Ediciones INTA*, 93.
- Viglizzo, E. F., Frank, F., Bernardos, J., Buschiazzo, D. E., & Cabo, S. (2006). A rapid method for assessing the environmental performance of commercial farms in the Pampas of Argentina. *Environmental Monitoring and Assessment*, 117(1), 109–134.
- Witten, I. H., Frank, E., & Hall M. A. (2011). *Data Mining: Practical machine learning tools and techniques*. (Third Edition). Morgan Kaufmann. Recuperado a partir de <https://books.google.com/books?hl=es&lr=&id=QTnOcZJzIUoC&oi=fnd&pg=PR17&dq=Data+Mining.+Practical+Machine+Learning+Tools+and+Techniques.+Third+Edition.&ots=3hnEfsVjMe&sig=nBeKsjRLUveXZUSWf-ljH-xjgCA>

Zehnder, R., & Quaino, O. (2005). Identificación y Caracterización de los Sistemas Relevantes de Producción en Santa Fe según Información del Censo Nacional Agropecuario 2002-INDEC. Primera parte. Datos de estructura, tamaño, actividades y régimen de tenencia. *INTA Centro Regional Santa Fe. EEA Rafaela.*, 18.