

Construcción de tipologías, una forma de manejar la diversidad de las fincas

Directrices generales para Humidtropics



Stéphanie Alvarez
Wim Paas
Katrien Descheemaeker
Pablo Tiftonell
Jeroen Groot

Wageningen, Países Bajos
2014

Humidtropics, un Programa de Investigación del CGIAR liderado por el IITA busca transformar la vida de las poblaciones campesinas pobres de América, Asia y África tropical. Las organizaciones de investigación involucradas en la alianza central con Humidtropics son [AVRDC](#), [Bioversity International](#), [CIAT](#), [CIP](#), [FARA](#), [icipe](#), [ICRAF](#), [ILRI](#), [IITA](#), [IWMI](#), y [WUR](#). humidtropics.cgiar.org

Agradecimientos



Agradecemos por la financiación parcial de este trabajo a Humidtropics, un Programa de Investigación del CGIAR cuyo objetivo es desarrollar nuevas oportunidades de mejores medios de subsistencia en un ambiente sostenible.

Humidtropics es un Programa de Investigación del CGIAR liderado por el IITA (Instituto Internacional de Agricultura Tropical), que busca transformar la vida de las poblaciones campesinas pobres de América, Asia y África tropical y que utiliza la investigación de sistemas integrados y plataformas de asociación únicas para tener un impacto en la reducción de la pobreza y en la integridad de los ecosistemas. Las organizaciones de investigación involucradas en la alianza central con Humidtropics son AVRDC (Centro Mundial de Vegetales), Bioversity International, CIAT (Centro Internacional de Agricultura Tropical), CIP (Centro Internacional de la Papa), FARA (Foro para la Investigación Agrícola en África), icipe (Ciencia Africana en Insectos para la Salud y la Alimentación), ICRAF (Centro Mundial Agroforestal), IITA (Instituto Internacional de Agricultura Tropical), ILRI (Instituto Internacional de Investigación en Ganadería), IWMI (Instituto Internacional de Gestión del Agua), y WUR (Universidad y Centro de Investigación de Wageningen). humidtropics.cgiar.org

El CGIAR es una asociación de investigación agrícola mundial para un futuro alimentario seguro. Su ciencia se lleva a cabo a través de los 15 centros de investigación que integran el Consorcio CGIAR en colaboración con centenares de organizaciones socias. www.cgiar.org

Agradecemos por la financiación de la traducción de esta guía en español a la Fundación McKnight (Andes) cuyo objetivo es ayudar a los campesinos a alimentar a sus familias y comunidades. <http://www.ccrp.org/>



La misión de Wageningen UR (Universidad y Centro de Investigación) es “explorar el potencial de la naturaleza para mejorar la calidad de vida”. Dentro de Wageningen UR, nueve institutos de investigación especializados pertenecientes a la Fundación DLO (Servicio de Investigación Agronómica, DLO por sus siglas en holandés) han unido esfuerzos con la Universidad de Wageningen para ayudar a responder las preguntas más importantes en el ámbito de la alimentación y condiciones de vida saludables. Con aproximadamente 30 sedes, 6000 empleados y 9000 estudiantes, Wageningen UR es una de las organizaciones líderes del mundo en su campo. www.wageningenur.nl/en.htm

Referencia bibliográfica

Alvarez, S., Paas, W., Descheemaeker, K., Tiftonell, P., Groot, J.C.J., 2014. *Construcción de tipologías, una forma de manejar la diversidad de las fincas: directrices generales para Humidtropics*. Informe para el Programa de Investigación de CGIAR sobre Sistemas de los Trópicos Húmedos. Grupo de Ciencias de las Plantas, Universidad de Wageningen, Países Bajos.

Tabla de contenido

Resumen	4
1 Introducción	5
1.1 Propósitos	5
1.2 Métodos	5
2 Directrices sobre tipologías	7
2.1 Objetivos de las tipologías	7
2.2 Hipótesis sobre la estructura de las tipologías	7
2.3 Selección de variables clave para los métodos estadísticos	10
2.4 Muestreo	12
2.5 Estadística multivariante	13
2.6 Verificación de hipótesis	15
3 Clasificación rápida, dinámicas y distribución en el territorio	17
3.1 Derivación del árbol de decisiones para una clasificación rápida de las fincas	17
3.2 Dinámicas de las fincas	18
3.3 Sistemas de geoinformación	19
4 Inclusión de resultados del desarrollo en las tipologías	20
4.1 Seguridad alimentaria y nutrición	20
4.2 Género	20
Referencias bibliográficas	21
Apéndice A: precisión de la encuesta	24
Apéndice B: ejemplo de ACP y AJ	Error! Bookmark not defined.

Resumen

Para captar la variabilidad de los sistemas de explotación agropecuaria, las tipologías son una herramienta de uso frecuente. Las tipologías en investigación para proyectos de desarrollo se utilizan para obtener eficazmente los ajustes, mejores políticas e innovaciones que se requieren para el modelo de finca¹ óptimo, con el fin de lograr los objetivos de Humidtropics.

El objetivo de este documento es proporcionar directrices para la construcción de tipologías. El enfoque general de tipología que aquí se propone es combinar con estadística multivariante el conocimiento de expertos obtenido a través de un enfoque participativo.

Después de compartir brevemente alguna información previa sobre el proceso de formulación de tipologías, a continuación se describen los diferentes pasos para su construcción:

1. Establecer con precisión el objetivo de la tipología;
2. Formular una hipótesis sobre la diversidad de los sistemas de explotación agropecuaria (fincas);
3. Seleccionar las variables que caracterizan a los sistemas de explotación agropecuaria;
4. Diseñar un método de muestreo para la recolección de datos;
5. Agrupar los sistemas de explotación agropecuaria utilizando la estadística multivariante;
6. Comparar los resultados de la tipología con la hipótesis y validar la tipología con expertos locales.

En los apéndices se dan ejemplos detallados de los métodos estadísticos que pueden aplicarse para la construcción de tipologías.

¹ El término “finca” se emplea aquí para designar los sistemas de producción agropecuaria en general (incluye hacienda, rancho, granja, chacra, etc.).

1 Introducción

En los trópicos húmedos los sistemas de explotación agropecuaria se hallan presentes a través de una gran variedad de culturas y paisajes. Los impulsores biofísicos, institucionales, sociales y económicos difieren entre contextos, dando como resultado que las respuestas de los campesinos y comunidades sean distintas entre regiones y al interior de ellas. Además, las fincas se encuentran en diferentes etapas de desarrollo y los campesinos tienen diferentes habilidades y aspiraciones. Con el tiempo, estas diferencias en los impulsores y en las características de las fincas conducen a una variabilidad espacial y temporal entre los sistemas de explotación agropecuaria y al interior de ellos.

Es difícil entender totalmente la variabilidad que existe en los sistemas de explotación agropecuaria, lo que resulta en una representación parcial de la realidad. Se han desarrollado varias herramientas y métodos (p. ej. clasificación por patrimonio, tipologías de fincas, distribuciones) para entender y manejar la diversidad de sistemas de explotación agropecuaria (fincas). Cuando se utilizan estos métodos y herramientas, se sopesan los pros y contras entre la calidad de representar la realidad y el nivel de detalle requerido. Una metodología de uso frecuente para manejar la variabilidad y diversidad es la construcción de tipologías, es decir, el agrupamiento de fincas.

En este documento se proporcionan directrices para la construcción de tipologías. Nuestros objetivos son presentar un enfoque por pasos para estructurar este proceso, suministrar recomendaciones y sugerencias prácticas sobre técnicas útiles, y crear consciencia sobre las oportunidades y obstáculos que surgen durante la construcción de tipologías.

1.1 Propósitos

Las tipologías responden a preguntas de investigación que requieren **que se tome en cuenta la heterogeneidad agrícola dentro de una región** (p. ej. Alary y col., 2002; Righi y col., 2011; ver Figura 1).

Las cuatro principales razones para desarrollar una tipología son:

1. *Enfocarse*: la diferenciación entre los sistemas de explotación agropecuaria busca identificar las intervenciones apropiadas por tipo de sistema.
2. *Escalar horizontalmente*: las tipologías ayudan a entender cómo las intervenciones apropiadas pueden ser difundidas a gran escala.
3. *Seleccionar*: las tipologías sirven como base para la selección de fincas representativas o para definir fincas prototipo (promedio) para análisis detallados.
4. *Escalar verticalmente*: las tipologías sirven como base para la extrapolación de evaluaciones de impacto ex-ante a mayores escalas espaciales u organizacionales (Ewert y col., 2011).

1.2 Métodos

Las tipologías de fincas se pueden construir usando varios métodos:

- *Comparación paso a paso del funcionamiento de una finca* (Capillon, 1993; Landais, 1998): para un área delimitada, este método de clasificación se basa en una extensa información acerca del funcionamiento de la finca (familia, objetivos, historia, producciones, manejo, resultados tecnoeconómicos, restricciones biofísicas, etc.), la cual se puede obtener mediante encuestas realizadas en una muestra estratificada de explotaciones agropecuarias. El agrupamiento por tipos se hace utilizando la comparación "paso a paso" de explotaciones vecinas (ver Landais, 1998 para más detalles sobre el método).
- *Conocimiento de expertos*: la construcción de tipologías se basa en agrupar las fincas en clases definidas por expertos locales, informantes clave, o campesinos (Giller y col., 2011; Landais, 1998; Pacini y col., 2013). Este enfoque conduce al establecimiento de una base de referencia común

(Landais, 1998). Generalmente, la construcción de tipologías basada en conocimiento de expertos requiere de poco tiempo y dinero (Landais, 1998).

- *Clasificaciones participativas*: clasificación de las fincas, principalmente según su patrimonio (clasificación por patrimonio), realizada por expertos y/o por los mismos campesinos en un proceso participativo. Los bienes observables son importantes cuando la clasificación se basa en el patrimonio (Kebede, 2007).
- *Análisis multivariante que incluya métodos de ordenación y agrupamiento*: este método se puede ver como el equivalente cuantitativo del "enfoque de conocimiento de expertos". Los métodos estadísticos (p. ej. Análisis de Componentes Principales, Análisis de Correspondencias Múltiples, Análisis Factorial Múltiple, Escalamiento Multidimensional) se utilizan para clasificar los objetos (en este caso las fincas). Idealmente no se proyectan ninguna jerarquía o preconcepciones sobre los objetos (Alary y col., 2002, Giller y col., 2011). Esta clase de métodos también son llamados técnicas de "reducción de dimensión" o "reducción de datos" (Pacini y col., 2013) porque tienen la ventaja de capturar la complejidad de los sistemas de explotación agropecuaria al tener en cuenta, al mismo tiempo, numerosas dimensiones de las fincas y resaltar luego unas cuantas dimensiones que explican mejor la diversidad de las fincas (Alary y col., 2002).

En los proyectos se tiene que cumplir con al menos dos estándares importantes: (i) los estándares de la ciencia en los cuales son importantes la precisión, la objetividad y la reproducibilidad, y (ii) los estándares de los resultados del proyecto, los cuales dependen de las diferentes necesidades, percepciones, intereses, etc. de los diversos actores.

Los métodos de estadística multivariante con frecuencia se prefieren a los enfoques basados en conocimiento de expertos debido a la reproducibilidad inherente a sus fundamentos estadísticos (Pacini y col., 2013), lo cual contribuye principalmente al estándar (i). No obstante, para contribuir también al estándar (ii) y aumentar el éxito de los proyectos, las tipologías tienen que ser relevantes para los grupos de interés. Por lo tanto, se podrían emplear de manera complementaria los diferentes métodos de tipología, utilizando en este caso estadísticas multivariantes además de enfoques participativos (Alary y col., 2002; Pacini y col., 2013; Righi y col., 2011).

Utilice de forma complementaria la estadística multivariante y el conocimiento de expertos

2 Directrices sobre tipologías

En la Figura 1 se muestra la estructura del marco de trabajo para la construcción de tipologías. Esta comprende seis pasos para pasar de una población heterogénea de fincas a una agrupación entre tipos coherentes de fincas. Los seis pasos son:

1. Establecer con precisión el objetivo de la tipología;
2. Formular una hipótesis sobre la diversidad de los sistemas de explotación agropecuaria (fincas);
3. Seleccionar las variables que caracterizan a los sistemas agropecuarios;
4. Diseñar un método de muestreo para la recolección de datos;
5. Agrupar los sistemas agropecuarios utilizando la estadística multivariante;
6. Comparar los resultados de la tipología con la hipótesis y validar la tipología con expertos locales.

2.1 Objetivos de las tipologías

Una **tipología de fincas o sistemas agropecuarios depende de la pregunta de investigación** (p. ej. Köbrich y col., 2003). Se pueden construir tipologías para un objetivo de investigación específico para un área específica (p. ej. "para mejorar el suministro de forraje en las tierras altas de Madagascar") o para un objetivo general para una zona extensa (p. ej. "para mejorar la seguridad alimentaria en los trópicos húmedos"). En ambos casos, es importante tener presente el objetivo al momento de construir una tipología y especialmente durante la selección de las variables (Figura 1).

Las fincas son puntos de mira móviles (Giller y col., 2011), mientras que las **tipologías (basadas en mediciones de una sola vez) dan una fotografía instantánea de la situación de las fincas en un período de tiempo determinado** (Kostrowicki, 1977). Debido a la dinámica de las fincas, las tipologías podrían volverse obsoletas muy rápidamente, así que es preferible actualizarlas periódicamente (Landais, 1998; Valbuena y col., 2014). Por eso, las tipologías deben ser evaluadas y actualizadas continuamente.

Evalúe y actualice continuamente su tipología

Otro punto a tener en cuenta es que uno deba enfrentar la escasez de datos y las restricciones de tiempo. En ese caso una "clasificación simple" basada principalmente en dotación de recursos (conocida como tipología estructural), podría ser la mejor opción, después de todo (Giller y col., 2011). Este enfoque rápido proporciona un punto de partida para explorar más en detalle las restricciones e impulsores existentes.

2.2 Hipótesis sobre la estructura de las tipologías

Como punto de partida del desarrollo de tipologías, se recomienda **establecer una hipótesis sobre la diversidad de fincas o sistemas agropecuarios del área estudiada** (Tiftonell, 2014a). Se puede estructurar la hipótesis utilizando conocimiento de expertos, métodos participativos y/o estudios en el área u observación de campo realizados previamente. La hipótesis debe estar relacionada con el propósito de la construcción de la tipología y, preferiblemente, basada en el conocimiento y teorías sobre agricultura (Whatmore y col., 1987). La hipótesis puede tratar sobre el número de tipos de sistemas agropecuarios, sus principales características y su proporción en las áreas de estudio.

Formule una hipótesis sobre la diversidad de sistemas de explotación agropecuaria (fincas)

La eficacia del desarrollo de tipologías se podría mejorar con la participación de los grupos de interés locales en el proceso de construcción de hipótesis (Righi y

col., 2011). Proponemos utilizar enfoques participativos, en los cuales se incluya a los grupos de interés locales (investigadores, actores y campesinos locales), con el fin de formular las hipótesis conjuntamente, lo cual dará como resultado una **descripción ex-ante de diferentes tipos de fincas**. Las hipótesis sobre los tipos de fincas deben reflejar los criterios seleccionados por los grupos de interés locales para describir los sistemas locales de explotación agropecuaria (Alary y col., 2002; Pacini y col., 2013). Estos criterios deben formar parte del cuestionario utilizado para las encuestas en las fincas en el paso siguiente. Una ventaja adicional de incluir a los grupos de interés locales es que la comunicación y el grado de participación pueden aumentar.

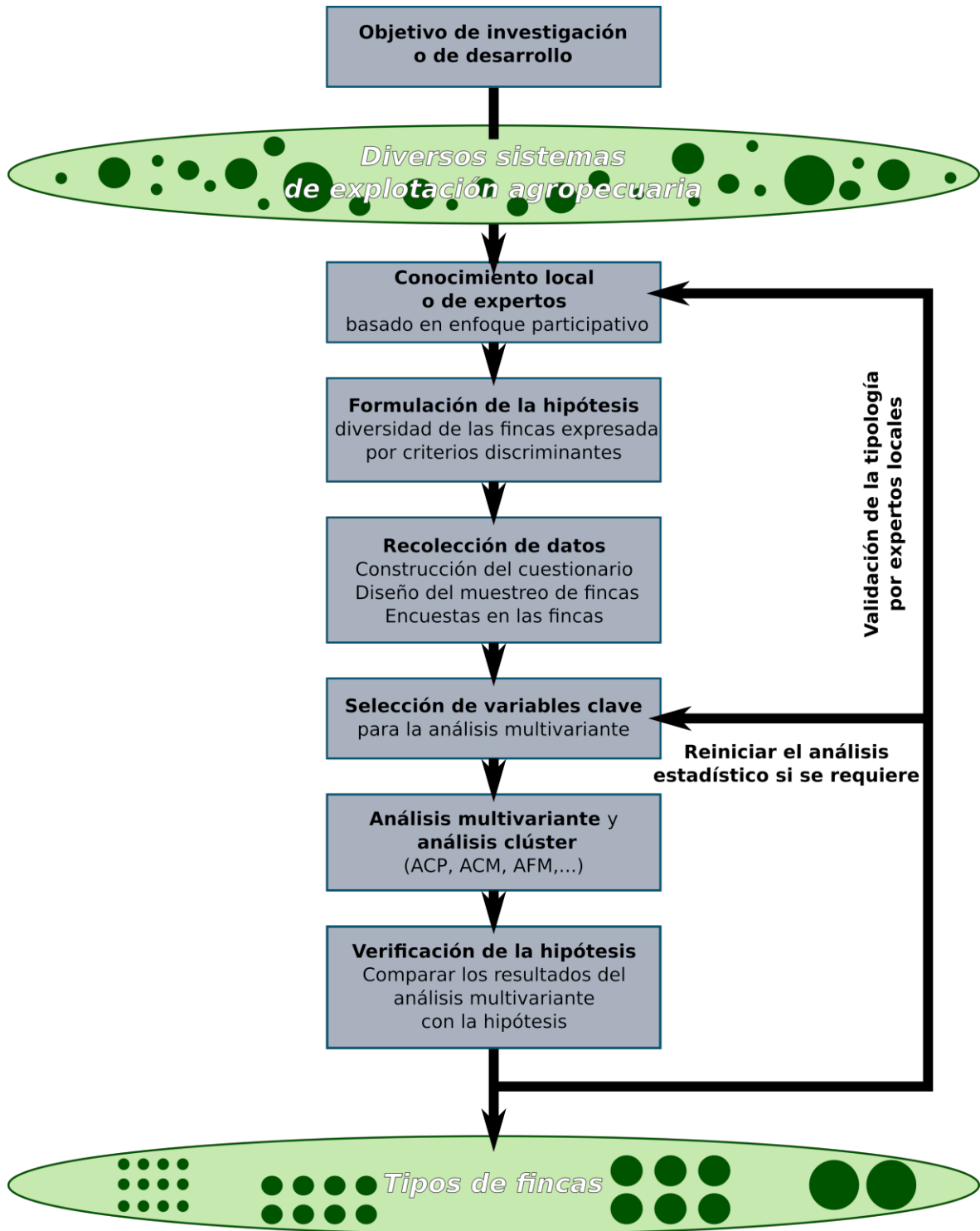


Figura 1: Marco de trabajo general del proceso de tipologías (ACP: Análisis de Componentes Principales; ACM: Análisis de Correspondencias Múltiples; AFM: Análisis Factorial Múltiple) (de acuerdo con Tiftonell, 2014a)

2.3 Selección de variables clave para los métodos estadísticos

Las tipologías se podrían agrupar en dos clases principales: (i) **tipologías estructurales** basadas principalmente en variables que describen los niveles de recursos y activos, y (ii) **tipologías funcionales** basadas en variables que describen las estrategias de subsistencia y las dinámicas de las fincas (Tittonell, 2014a). El objetivo de la tipología debe orientar el proceso de desarrollo de tipologías, y por lo tanto la **selección de variables**² (Pacini y col., 2013).

La recolección de datos provenientes de las fincas es un paso esencial en la construcción de tipologías. Se recomienda diseñar una encuesta para captar todo el sistema de explotación agropecuaria (Giller y col., 2011; Tittonell y col., 2010). Las variables utilizadas en los cuestionarios a las fincas podrían agruparse en categorías específicas, por ejemplo:

- Variables de características estructurales y variables del funcionamiento de las fincas (Sanago y col., 2010);
- Disponibilidad de los recursos de las fincas y Gestión (Pacini y col., 2013);
- Recursos biofísicos, Aspectos socioeconómicos, y Equipos (Righi y col., 2011);
- Cultivos, Insumos químicos, Cosecha, densidad y calidad de los frutos, Recursos Económicos, Naturaleza del sistema de explotación agropecuaria, Contexto físico, Aspiraciones personales, Social, Desempeño agronómico, Desempeño económico y Desempeño ambiental (Blazy y col., 2009).

En este caso, con el fin de garantizar un enfoque de sistemas, se recomienda considerar variables relacionadas con los principales componentes del sistema de explotación agropecuaria (es decir, hogar o familia, sistema de cultivo, sistema de cría de animales) y sus interacciones con el medio ambiente o con el exterior (p. ej. contexto ambiental, contexto económico, contexto sociocultural). Dado que los sistemas de explotación agrícola y ganadera son la columna vertebral de la agricultura de pequeños productores en los países en vías de desarrollo (Thornton y Herrero, 2001), presentamos aquí un ejemplo de un conjunto de variables que se utilizan comúnmente en el contexto de los sistemas de explotación de cultivos y ganadería (Tabla 1). Naturalmente, la selección de variables clave para el análisis multivariante debe adaptarse al propósito de la tipología, el área y el contexto del sistema de explotación agropecuaria, siguiendo el primer paso de intercambio con los grupos de interés locales y de formulación de hipótesis. Blazy y col. (2009), por ejemplo, estudiaron las innovaciones en gestión de cultivos por tipo de finca, y por lo tanto se enfocaron más en las variables que indican la naturaleza técnica del sistema de gestión de cultivos, y menos en las variables que son un indicador para los factores sociales.

Antes de la **selección de las variables clave para el análisis multivariante**, es importante evaluar la calidad requerida para los datos (cantidad, exactitud). Los datos de las encuestas en fincas en los países en vías de desarrollo son a menudo erróneos (Howe y McKay, 2007; en el Apéndice A se proporcionan ejemplos), lo cual perjudica la fiabilidad de las estadísticas. De ahí que sea bueno revisar los datos y escoger ciertas variables de manera acertada. En el Apéndice B se describe un paso de control de datos para el Análisis de Componentes Principales.

² La variable como un atributo de las fincas en la población de fincas que queremos estudiar

Tabla 1: Ejemplo de variables que describen los sistemas de explotación agrícola y ganadera. La columna “conjunto común” propone un conjunto de variables que podrían incluirse en la construcción de tipologías en todas las sedes de Humidtropics

Atr.	Categoría	Variable	Unidad	Conjunto común
R	Hogar	Tamaño de la familia	cápita	✓
R	Hogar	Edad de la persona jefe de familia	año	
R	Hogar	Mano de obra familiar en las actividades agropecuarias	cápita u hombre-día/año ^a	✓
R	Hogar	Mano de obra contratada	cápita u hombre-día/año ^a	✓
O	Hogar	Meses de autosuficiencia alimentaria	meses	
R	Hogar-Medio ambiente	Actividades fuera de la finca	clases ^b	✓
R	Hogar-Medio ambiente	Margen bruto total de la finca (ingresos - gastos)	moneda local	
R	Hogar-Medio ambiente	Ingresos por labores fuera de la finca (% de los ingresos del hogar)	%	✓
O	Hogar-Medio ambiente	Compra de alimentos (% de los gastos del hogar)	%	
R	Sistema de cultivo	Área de propiedad de la finca	ha	
R	Sistema de cultivo	Área trabajada por la familia ^c	ha	✓
R/O	Sistema de cultivo	Área con cultivos de alimentos	ha o % de área cultivada	
R/O	Sistema de cultivo	Área con cultivos de forrajes	ha o % de área cultivada	✓
R/O	Sistema de cultivo	Área con cultivos comerciales	ha o % de área cultivada	✓
O	Sistema de cultivo-Medio ambiente	Ventas de la producción de los cultivos (% de los ingresos)	%	
O	Sistema de cultivo-Medio ambiente	Compra de fertilizantes minerales u orgánicos (% de los gastos)	%	
O	Sistema de cultivo-Medio ambiente	Compra de pesticidas (% de los gastos)	%	
R	Sistema de ganadería	Número total de animales	UGT (Unidades Ganaderas Tropicales)	✓
R/O	Sistema de ganadería	Cantidad de ganado local	n ^o	
R/O	Sistema de ganadería	Cantidad de ganado de raza mejorada	n ^o	
R/O	Sistema de ganadería	Cantidad de rumiantes menores	n ^o	
R/O	Sistema de ganadería	Cantidad de animales menores (cerdos y/o aves)	n ^o o UGT (Unidades Ganaderas Tropicales)	
O	Sistema de ganadería	Producción de leche	l/año	
O	Sistema de ganadería-Medio ambiente	Total de ventas de productos de origen animal (% de los ingresos)	%	
O	Sistema de ganadería-Medio ambiente	Ventas de estiércol (% de los ingresos)	%	
O	Sistema de ganadería-Medio ambiente	Compras de concentrado o de forraje (% de los gastos)	%	
O	Hogar-Medio ambiente	Objetivos o estrategias de producción (desafiando las restricciones externas)	clases ^d	✓

Atr.: Atributo; O: Orientaciones; R: Recursos

a: hombre-día/año podría permitir considerar diferentes tipos de mano de obra (p. ej. persona de tiempo completo, niño, mujer)

b: clases por definir de acuerdo al conjunto de los resultados de las encuestas (p. ej. Sí/No, Agrícola/No agrícola, Agrícola/Urbano)

c: el área dedicada a la explotación agropecuaria podría incluir áreas cultivadas, de pastoreo o de plantación

d: clases por definir de acuerdo al conjunto de los resultados de las encuestas (p. ej. Aumentar/Mantener/Disminuir la producción, Aumentar/Mejorar/Diversificar/Cambiar la producción)

Se pueden utilizar las variables externas (p. ej. distancia a la carretera, disponibilidad de acceso a los mercados) **para explicar la diversidad** entre las fincas, y los factores que la impulsan. Algunos ejemplos en la literatura combinan en la construcción actual de tipologías las variables internas de la finca y las variables del medio ambiente externo (p. ej. Ansoms y McKay, 2010; Tittonell y col., 2010). La disponibilidad de variables externas ofrece una excelente oportunidad para probar las teorías sobre los impulsores de la diversidad. No obstante, podría ser preferible no usar esas variables externas para la identificación de los tipos de fincas (es decir, como variables clave para el análisis multivariante), con el fin de distinguir más claramente entre las variables que describen la diversidad de los sistemas de explotación agropecuaria (fincas) y las que explican esta diversidad.

Distinguir entre las variables que describen la diversidad de las fincas y los indicadores que explican la diversidad

Se recomienda utilizar un "número pequeño" de variables clave (Kostrowicki, 1977) y asegurarse de que el número de fincas encuestadas sea por lo menos cinco veces mayor que el número de variables clave usadas para el análisis multivariante (Hair y col., 2010). Así que el número de variables clave requeridas y seleccionadas para la estadística multivariante puede ser diferente al número de variables planteadas durante la encuesta. La Tabla 1 muestra algunas variables que podrían ser seleccionadas como variables clave para el análisis multivariante. Además, otras variables integradoras (p. ej. la relación Mano de obra/Terreno, la relación UGT/Área de forraje, la relación Área de cultivos alimenticios/Área cultivada) podrían calcularse a partir de las variables planteadas durante la encuesta.

Utilice la proporción de 5 unidades de observación (fincas) por 1 variable clave.

El número de variables clave utilizadas en estadística multivariante para fines de tipología es altamente cambiante: de 5 a 46 variables, con un promedio de aproximadamente 15 variables³. Una visión general de 21 estudios² mostró un promedio de 9 unidades de observación por variable de finca. Debe notarse que la selección de variables clave introduce un grado de subjetividad en el proceso de construcción de tipologías.

Es importante tener en cuenta que, como resultado del análisis multivariante, no todas las variables clave que se procesaron en este análisis se conservarán necesariamente como variables discriminantes. **Las variables discriminantes son las variables que el análisis multivariante muestra como las que mejor describen la variabilidad de la finca** (ver ejemplo en el Apéndice B). Es posible observar que pueden construirse tipologías intermedias por clase de variable, con el fin de obtener mayor percepción en cuanto a la diversidad por clase de variable (p. ej. Maton y col., 2005; Moreno-Pérez y col., 2011).

2.4 Muestreo

El muestreo de fincas debe cubrir la diversidad de fincas o sistemas de explotación agropecuaria del área estudiada (Pacini y col., 2013). Por lo tanto el muestreo debe ser elaborado con base en la hipótesis inicial, y especialmente en las proporciones esperadas de tipos de finca. Si el muestreo es completamente aleatorio, se requiere una muestra de gran tamaño. Para reducir el tamaño del muestreo, se pueden utilizar métodos que capturan la diversidad a lo largo de un gradiente (p. ej. el método en forma de "Y" descrito por Tittonell y col., 2010), o métodos basados en estratificación o por

³ Con base en 21 estudios de tipologías

transectos (p. ej. el “transecto que sigue a un gradiente de intensificación” utilizado por Pacini y col., 2013).

No se recomienda pedirle a “todos los campesinos” que vengan a un lugar de encuentro para que llenen la encuesta: la muestra de fincas podría sesgarse por la capacidad y/o motivación de los campesinos para venir al sitio de la reunión. Por otra parte, las visitas a las fincas permiten revisar otras cosas, por ejemplo el área de campo sembrada, los productos cultivados, las herramientas que se poseen y los animales que se manejan.

Usualmente se recomienda por razones estadísticas tener una muestra de por lo menos 50 fincas (Hair y col., 2010). En la práctica la muestra de fincas para estudios de tipología oscila entre 18 y 2646 fincas, con una mediana de 138 fincas encuestadas⁴. Además, Kumar (2014) da información adicional sobre metodología de muestreo. Es importante tener siempre en cuenta que el tamaño de la muestra y el método de muestreo pueden impactar la proporción de fincas pertenecientes a cada tipo de finca resultante. Por ejemplo, cuando una muestra contiene 100 fincas, un tipo de finca que en realidad combine cerca del 10% de los sistemas de explotación agropecuaria de la zona podría estar representado por sólo una o dos fincas encuestadas, debido al proceso de muestreo (Hair y col., 2010). Además, durante el proceso multivariante, estas dos fincas podrían ser consideradas como valores atípicos, o podrían estar combinadas en otros tipos de fincas.

2.5 Estadística multivariante

Los análisis multivariantes y de agrupamientos se utilizan **para identificar variables explicativas (variables discriminantes) y para agrupar las fincas en tipos homogéneos**. La estadística multivariante permite reducir el número de variables y preservar el máximo de la variabilidad total de la muestra. De acuerdo con la naturaleza de las variables clave seleccionadas (cuantitativas y/o categóricas) se deben usar diferentes estadísticas multivariantes:

- *Análisis de Componentes Principales* (ACP) para variables cuantitativas (continuas o discretas) (p. ej. Bidogeza y col., 2009; Sanogo y col., 2010; Tittone y col., 2010) (ver ejemplo de ACP en el Apéndice B);
- *Análisis de Correspondencias Múltiples* (ACM) para variables categóricas (p. ej. Blazy y col., 2009);
- *Análisis Factorial Múltiple* (AFM) para variables categóricas organizadas en conjuntos de datos multitabla y multibloque (Alary y col., 2002);
- *Análisis de Hill-Smith* para variables cuantitativas y cualitativas mezcladas (p. ej. Rueff y col., 2012);
- *Escalamiento Multidimensional* para construir una configuración de clasificación en una dimensión específica (p. ej. Pacini y col., 2013; Righi y col., 2011).

Como se mencionó anteriormente, el primer paso del análisis multivariante concierne a la selección de las variables clave. Es necesario verificar que una determinada categoría de variables (es decir, un grupo de variables que describen un mismo aspecto del sistema; ver ejemplos de categorías en la sección 2.3) no esté sobrerrepresentada (es decir, cuando el número de variables en esta categoría es mucho mayor que para las otras categorías); de lo contrario eso podría dar más peso a esta categoría de variables y por lo tanto sesgar el análisis (Blazy y col., 2009; Kostrowicki, 1977).

Otra medida de precaución, específicamente para el ACP, es estandarizar todas las variables seleccionadas (cuantitativas), utilizando por ejemplo porcentajes, “para evitar la influencia de diferentes niveles de variación debido a la unidad de medida” (Pacini y col., 2013). Esta precaución mejora la comparación de variables con diferentes unidades, por ejemplo el área cultivada que oscila entre 0.5 y 2.0 ha, y los ingresos por agricultura que van de \$150/año a \$5000/año. Los métodos ACM y AFM son sensibles a bajos números de unidades de observación o a clases desbalanceadas; por lo tanto en este caso podría ser necesario combinar algunas clases.

⁴ Mediana observada en una muestra de 22 estudios de tipologías

Además, se requiere probar la independencia de las variables, por ejemplo con la Prueba de Chi-cuadrado de Pearson. De hecho, si dos variables estuvieran estrechamente correlacionadas eso haría que en el análisis multivariante la información dada por esas variables pesara dos veces más.

Por otra parte, el ACM podría ser más difícil de interpretar que los análisis ACP y AFM; la interpretabilidad de ACM es mayor cuando el número de variables seleccionadas es limitado (preferiblemente ≤ 20 variables; Hervé, 2011).

Los análisis multivariantes son bastante sensibles a los valores atípicos (posibles errores o unidades de observación "excepcionales"), por lo que se recomienda eliminar del análisis estos valores (Hair y col., 2010). Si se sacan del análisis multivariante las "fincas excepcionales", tal como se sugiere (Apéndice B), puede ser útil resaltarlas en un informe final que presente la diversidad de los sistemas de explotación agropecuaria (fincas).

El número de ejes (es decir, los componentes o factores principales) para los ACP, ACM y AFM se puede determinar de acuerdo con un criterio establecido antes del análisis (p. ej. el número de ejes que explican un mínimo de $x\%$ de la variabilidad – "usualmente 60% o más alto" (Hair y col., 2010)), o utilizando el Criterio de Kaiser para el ACP (es decir, que se escogen todos los ejes con un valor propio superior a 1; Hervé, 2011).

Seleccione los ejes con un valor propio superior a 1.0 y/o los ejes que expliquen más del 60% de la variabilidad

El Análisis Clúster (AC) tiene por objetivo agrupar las fincas en clases o tipos lo más "homogéneos" posible.

- *Agrupamiento no jerárquico*, es decir, una separación del espacio de las unidades de observación o fincas en grupos o tipos disjuntos donde el número de grupos (k) es fijo;
- *Agrupamiento jerárquico*, es decir, una agregación paso a paso del espacio de las unidades de observación o fincas en grupos o tipos disjuntos. Primero cada finca es un grupo por sí mismo, y luego en cada paso, los dos grupos más similares se fusionan hasta que solo queda un grupo con todas las fincas. El resultado visual de estos pasos (algoritmo) es un dendrograma o árbol de clasificación (Figura 2). La altura de las ramas del dendrograma representa la distancia (disimilitud) promedio entre las unidades de observación al interior de los grupos y entre los grupos mismos. Por lo tanto, el dendrograma proporciona una representación visual de la variabilidad de los datos y una herramienta útil para justificar la selección de una partición, es decir, el número de clústeres. La partición del dendrograma podría hacerse con base en: (i) la apariencia general del dendrograma, (ii) el número de clústeres y (iii) su interpretabilidad, y (iv) el examen del delta de alturas. Por consiguiente, partiendo de la parte superior del dendrograma (el nivel más elevado de altura o "nodos raíz"), la estructura del dendrograma sugiere una división entre un número n de clústeres cuando la reducción del nivel de disimilitud al pasar de un $(n-1)$ clústeres a n clústeres (es decir, Δ Altura $_{(n-1) \text{ a } n \text{ clústeres}}$) es mucho mayor que cuando se pasa de n clústeres a $(n+1)$ clústeres (Husson y col., 2011). Por último, es importante anotar que, a pesar del uso de criterios para respaldar la partición del dendrograma, tales como Δ Altura, sigue habiendo subjetividad en la selección de la partición.

El algoritmo de *Agrupamiento Aglomerativo Jerárquico* se utiliza a menudo en el proceso de construcción de tipologías (p. ej. Alary y col., 2002; Blazy y col., 2009; Pacini y col., 2013; Sanogo y col., 2010).

Los dos métodos de agrupamiento se pueden utilizar juntos para combinar sus respectivas fuerzas (Michielsens y col., 2002; Iraizoz y col., 2007). En la combinación, el agrupamiento jerárquico se usa para estimar el número de clústeres, mientras que el agrupamiento no jerárquico se usa para calcular los

centros de clústeres. El número de tipos de fincas normalmente oscila entre 3 y 7, con una mediana de 5 tipos de fincas⁵.

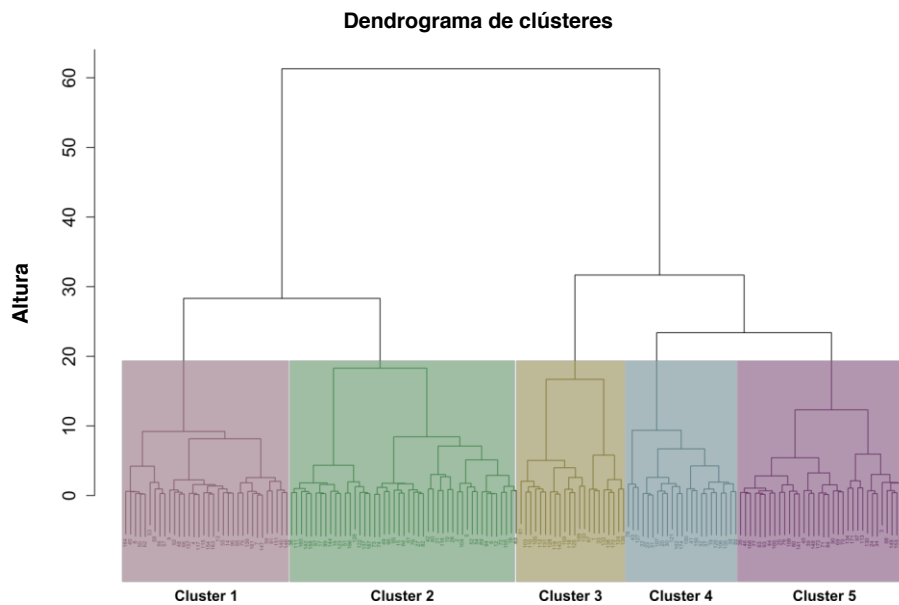


Figura 2: Ejemplo de un dendrograma correspondiente al Agrupamiento Aglomerativo Jerárquico de 161 fincas en Tanzania (Apéndice B) en el que se muestran cinco clústeres (tipos de fincas).

Los resultados del análisis multivariante y del análisis de agrupamiento se utilizan juntos **para interpretar el significado de cada clúster** (Figura 3). En el Apéndice B se detalla un ejemplo concreto de ACP y AC.

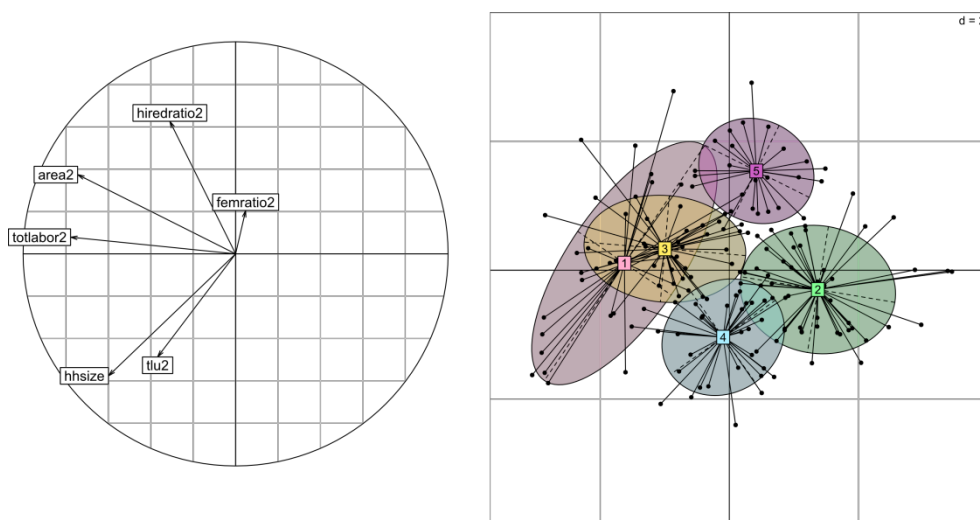


Figura 3: Ejemplo de resultados del Análisis de Componentes Principales y del Agrupamiento Jerárquico en 161 fincas en Tanzania (en el Apéndice B se proporcionan más detalles).

2.6 Verificación de hipótesis

Los tipos de fincas se deben seleccionar con base en su valor explicativo, es decir, tienen que ser significativos a nivel conceptual (Moreno-Pérez, 2011). Los tipos de fincas que resulten de los análisis

⁵ Datos observados en una muestra de 20 estudios de tipologías

multivariante y de agrupamientos se deben comparar con la hipótesis inicial (ver 2.2). Es necesario discutir y tratar de entender las diferencias entre la hipótesis y los resultados del análisis multivariante. En caso de que haya resultados inesperados, es posible que se deba repetir el análisis multivariante y de clústeres y/o que se deban reiniciar las sesiones de discusión y retroalimentación con los grupos de interés locales. Por último, es deseable contar con una validación de los resultados de la tipología por parte de expertos y/o campesinos locales (Figura 1).

Más aún, hacer una comparación entre los tipos de fincas resultantes y los hallazgos de otras investigaciones brinda una oportunidad para lograr una mejor comprensión del sector agrícola.

3 Clasificación rápida, dinámicas y distribución en el territorio

3.1 Derivación del árbol de decisiones para una clasificación rápida de las fincas

Se pueden visualizar los resultados de la tipología utilizando diagramas de caja. Los diagramas de caja pueden respaldar no solo la interpretación de los tipos de fincas sino también la identificación de variables con poder distintivo (Figura 4).

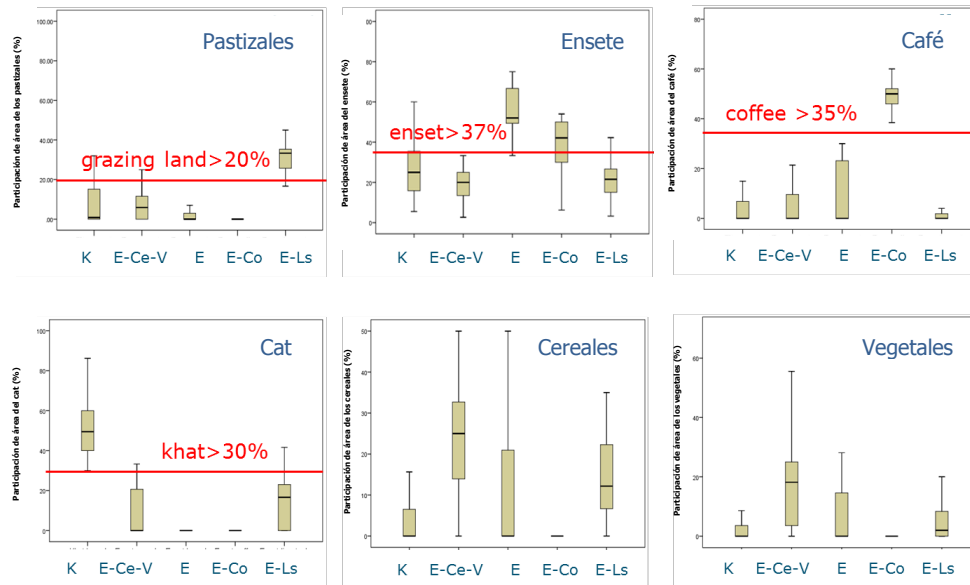


Figura 4: Diagramas de caja de indicadores distintivos para diferentes tipos de fincas con base en el uso de la tierra (%) para ciertos cultivos. Siguiendo el orden de selección es posible diferenciar entre tipos de fincas basándose solo en unos cuantos indicadores. Fuente: van de Ven y Melisse (2014). Las letras sobre el eje de las X representan los cultivos (K para cat (kath), E para ensete (enset), Ce para cereales, V para vegetales, Co para café (coffee)) y el ganado (Ls (Livestock))

Estos indicadores distintivos se podrían usar para crear un árbol de clasificación, proporcionando así una herramienta para una "clasificación rápida" de otras fincas con base en una cantidad reducida de variables (en estos 4, participación por área de pastizales, ensete, café y cat). Por ejemplo, en la Figura 4, más del 35% de la tierra utilizada para café es distintiva para el tipo de finca E-Co; en un siguiente paso, más del 30% de la tierra utilizada para cat es distintiva para el tipo de finca K, etc. (Figura 5).

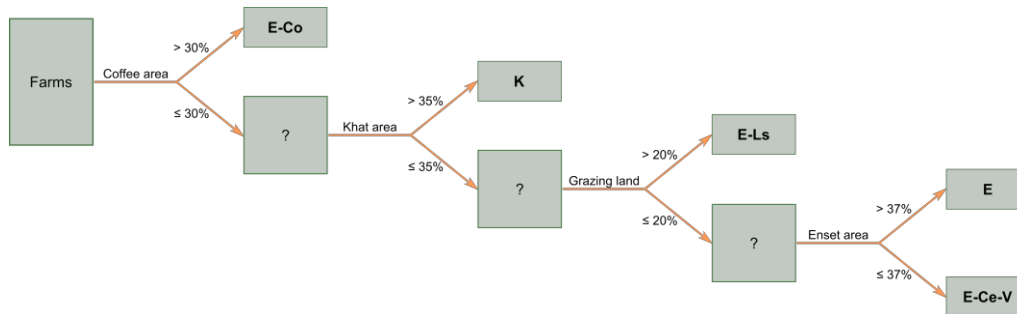


Figura 5: Ejemplo de una “clasificación rápida” utilizando un árbol de clasificación. El árbol de clasificación fue creado con base en diagramas de caja de los resultados de las tipologías y en los indicadores distintivos que se identificaron (K=cat, E=ensete, Ce=cereales, V=vegetales, Co= café and Ls=ganado)

3.2 Dinámicas de las fincas

Como se mencionó anteriormente, las fincas son reconocidas como puntos de mira móviles (Giller y col., 2011). Por lo tanto, el estudio en una finca y la tipología que resulte de él son fotografías instantáneas en un momento dado (Kostrowicki, 1977; Laurent y col., 1999). De ahí que sea necesario emplear las tipologías con cautela. Es bueno conocer el contexto temporal dentro del cual se realizó la encuesta, con el fin de reflejar la representatividad de ciertas variables.

En la literatura hay algunos ejemplos que toman en consideración la dinámica:

- Los movimientos de fincas individuales se pueden proyectar en trayectorias de tipos de fincas, de los cuales existe la hipótesis de que son predecibles, es decir, que siguen tendencias generales (Laurent y col., 1999; Tittonell, 2014b). Sin embargo, hay que tener presente que las fincas individuales podrían optar por seguir patrones diferentes o verse forzadas a ello (Valbuena y col., 2014).
- Cuando se puede acceder a datos de múltiples años, el método de Falconnier y col. (presentado) es interesante. Falconnier y col. (enviado) realizaron un análisis multivariante para la situación actual. Después analizaron el uso de la tierra para los cultivos y construyeron un árbol de decisiones para una clasificación fácil de las fincas, como hicieron van de Ven y Melisse (2014; Figuras 5 y 6). Con base en el árbol de decisiones y en unas cuantas variables de datos de la finca obtenidos en años anteriores, fue posible clasificar las fincas y compararlas con la clasificación actual.
- Incluir el conocimiento de los grupos de interés o de expertos ofrece una oportunidad para captar algunas de las dinámicas de las fincas, ya que ellos pueden evaluar los tipos de fincas a la luz de tendencias de largo plazo. Su experticia puede ayudar a rastrear la historia o evolución de las fincas, lo cual conduce a su estructura real y de esta manera a los tipos de fincas resultantes.
- Para establecer prioridades al interior de las sedes de acción y/o las sedes de campo, y para realizar comparaciones entre áreas, se pueden identificar los tipos de fincas que están por debajo de la línea de la pobreza (Davis y col., 1997; Howe y McKay, 2007; Tittonell y col., 2010). Además, identificar los tipos que se encuentran ubicados en una situación recurrente de pobreza revela algunas de las posibilidades o imposibilidades de los hogares para escapar de la pobreza (Howe y McKay, 2007).
- Los flujos de recursos (p. ej. alimentos, mano de obra, dinero, conocimiento) y la interdependencia entre los tipos de fincas en una comunidad son otro indicador del potencial para las dinámicas. Laurent y col. (1999) presentan un diagrama de flujo de recursos entre los tipos de fincas, el cual sirve de apoyo al desarrollo de una comprensión de la situación.

3.3 Sistemas de goinformación

La explotación agropecuaria es una actividad relacionada con un territorio. Algunos artículos abogan por tipologías de fincas vinculadas a nivel territorial, es decir, que conectan los tipos de fincas con su posición en el paisaje, especialmente para planificación del uso de la tierra (Landais, 1998; Carmona y col., 2010; Madry y col., 2013). No obstante, llevar a cabo un análisis territorial requiere un contexto rico en datos, lo cual no es algo que se dé con frecuencia en los países en vías de desarrollo (Carmona y col., 2010). El uso de las coordenadas GPS de las fincas para mapear los tipos de fincas permite evaluar las relaciones entre los tipos y los elementos del paisaje, por ejemplo las carreteras. La Figura 6 presenta un mapa en el que los tipos de fincas están representados por puntos de diferentes colores, lo cual ayuda a visualizar la diversidad de las fincas, es decir, la organización de los diferentes tipos de fincas dentro del territorio.

A: Descripción de los tipos de fincas

<i>Tipos de fincas</i>	<i>Principales características</i>
1	Área mediana de tierra cultivable, la vivienda y hato ganadero más grandes + amplias actividades fuera de la finca
2	La mayor área de tierra cultivable, alta proporción de legumbres, gran hato ganadero + orientada al mercado
3	Gran área de tierra cultivable, la más alta proporción de legumbres y rumiantes menores + la mayoría de mano de obra en labores de la finca
4	Área de tierra mediana, pequeño hato de animales, alta proporción de legumbres, la mayoría de la mano de obra contratada + orientada al mercado
5	La menor área de tierra cultivable, la más alta proporción de maíz, la más baja proporción de legumbres + los más bajos ingresos por actividades fuera de la finca
6	Pequeña área cultivable, los más pequeños hatos ganaderos + más que todo ventas de ganado

B: Organización de los tipos de fincas en el territorio

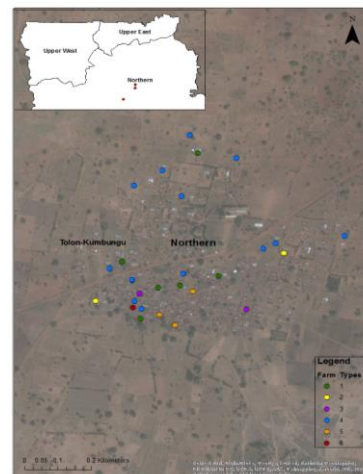


Figura 6: Distribución territorial de los tipos de fincas en una aldea de la región norte de Ghana.

Fuente: Kuivanen (2014).

4 Inclusión de resultados del desarrollo en las tipologías

4.1 Seguridad alimentaria y nutrición

Dado que una nutrición mejorada es un importante resultado de programa de la investigación para proyectos de desarrollo, es fundamental que en la construcción de tipologías se incluyan indicadores de nutrición para la seguridad alimentaria, especialmente cuando la pregunta de investigación aborda específicamente la seguridad alimentaria y la condición nutricional de los hogares. La desnutrición y el hambre oculta son dos formas de malnutrición.

Algunos estudios incluyen la desnutrición indicando los meses de seguridad alimentaria (p. ej. Tabla 1). Esto podría especificarse aún más si se pregunta en cuáles meses la familia come alimentos provenientes de la finca (p. ej. Tiftonell y col., 2010) y de otras fuentes (p. ej. Ebanyat y col., 2010). También se podrían indicar los meses en los cuales las familias luchan por conseguir el alimento. Es útil tener indicadores que especifiquen las "otras fuentes", las cuales podrían ser ayuda alimentaria o alimentos comprados con salario obtenido por labores fuera de la finca.

El hambre oculta se refiere a las deficiencias en la dieta, cuando la ingesta de calorías y/o proteínas es suficiente. Entre las deficiencias importantes en la dieta en los países en vías de desarrollo se incluyen las deficiencias de hierro, yodo, vitamina A, zinc y folato (Kennedy y col., 2003; Muthayya y col., 2013; ambos artículos presentan mapas de deficiencias de nutrientes en el mundo). Las deficiencias se deben en parte a las dietas monótonas, de ahí que la diversidad de la dieta sea un indicador importante. Para evaluar la diversidad de la dieta podrían usarse el Puntaje de Diversidad Alimentaria (es decir, el número de ciertos grupos de alimentos consumidos por un individuo o por el núcleo familiar (Kennedy y col., 2011)) o la Diversidad Funcional (Remans y col., 2011). Es posible evaluar qué tan adecuada es la dieta por el equilibrio entre ingestas y requerimientos; se recomienda el método Recordatorio de 24 horas (24h-recall) para cuantificar las ingestas humanas. El contenido nutricional de la mayoría de los cultivos tropicales puede encontrarse consultando las páginas web de la FAO (1968).

4.2 Género

No logramos encontrar en literatura muchos ejemplos donde se incluyeran explícitamente indicadores de género en los estudios de tipologías. De Lima Vidal (2013) desarrolló una tipología con solo mujeres cabeza de familia en un área de Brasil. Molua (2011) presenta un estudio donde el único determinante para la tipología fue el género de la cabeza del hogar. Djurfeldt y col. (2008) encontraron que los asuntos de género en relación con la productividad de las fincas son determinados específicamente por el contexto y no siempre dependen de la hostilidad del ambiente o de la abundancia en una región. Por lo tanto, aunque simple, la distinción de Molua (2011) podría ser una primera evaluación cuantitativa rápida de la importancia relativa de los asuntos de género en una región.

En la actual encuesta ImpactLite del ILRI, se ha prestado atención a los aspectos de género en los hogares. Los indicadores obtenidos ofrecen oportunidades para evaluar la influencia del género en los tipos de fincas. Sería interesante ver si los indicadores específicos de género dentro de un hogar están asociados con diferentes estructuras y niveles de productividad de las fincas. En particular, con respecto al enfoque en intervenciones técnicas que mejoren las condiciones de trabajo y la eficiencia laboral de las mujeres, podría ser deseable diferenciar grupos de fincas con base en los aspectos de género. Algunas posibles variables podrían ser: la proporción de mujeres y niños en los hogares; el total de mano de obra masculina y femenina por finca; el ganado, tierra y otros activos que pertenecen a hombres y a mujeres.

Referencias bibliográficas

- Ansoms, A., McKay, A., 2010. A quantitative analysis of poverty and livelihood profiles: The case of rural Rwanda. *Food Policy* 35, 584-598.
- Alary V., Messad S., Taché C., Tillard E., 2002. Approach to the diversity of dairy farm systems in Reunion. *Revue Elev Méd Vét Pays Trop* 55, 285-297.
- Bidogeza, J. C., Berentsen, P. B. M., Graaff, J., Oude Lansink, A. G. J. M., 2009. A typology of farm households for the Umutara Province in Rwanda. 1(3): 321-335.
- Blazy, J.-M., Ozier-Lafontaine, H., Doré, T., Thomas, A., Wery, J., 2009. A methodological framework that accounts for farm diversity in the prototyping of crop management systems. Application to banana-based systems in Guadeloupe. *Agricultural Systems* 101, 30-41.
- Capillon, A., 1993. Typologie des exploitations agricoles, contribution à l'étude régionale des problèmes techniques. Doctoral thesis, INA P-G, Paris.
- Carmona, A., Nahuelhual, L., Echeverría, C., Báez, A. 2010. Linking farming systems to landscape change: An empirical and spatially explicit study in southern Chile. *Agriculture, Ecosystems and Environment* 139: 40-50.
- Davis, J., Mack, N., Kirke, A., 1997. New perspectives on farm household incomes. *Journal of Rural Studies* 13, 57-64.
- De Lima Vidal, D., 2013. Work division in family farm production units: Feminine responsibilities typology in a semi-arid region of Brazil. *Journal of Arid Environments* 97, 242-252.
- Djurfeldt, G., Larsson, R., Holmquist, B., Jirström, M. and Andersson, A., 2008. African farm dynamics and the sub- continental food crisis – the case of maize. *Food Economics, Acta Agriculturae Scandinavica C* 5, 75–91.
- Ebanyat, P., De Ridder, N., De Jager, A., Delve, R.J., Bekunda, M.A., Giller, K.E., 2010. Drivers of land use change and household determinants of sustainability in smallholder farming systems of Eastern Uganda. *Population & Environment* 31, 474-506.
- Ewert, F., Van Ittersum, M.K., Heckeley, T., Therond, O., Bezlepina, I. and Andersen, E., (2011) Scale changes and model linking methods for integrated assessment of agri-environmental systems. "Agriculture, Ecosystems and Environment" 142, 6–17.
- Falconnier, G., Descheemaeker, K., van Mourik, T.A., Sanogo, O.M., Giller, K.E., *submitted*. Two decades of change in southern Mali: understanding farm trajectories and development pathways. Submitted to *Agricultural Systems*.
- FAO, 1968. Food Composition for use in Africa. <http://www.fao.org/docrep/003/X6877E/X6877E00.htm>
- Giller, K.E., Tittonell, P., Rufino, M.C., van Wijk, M.T., Zingore, S., Mapfumo, P., Adjei-Nsiah, S., Herrero, M., Chikowo, R., Corbeels, M., Rowe, E.C., Baijukya, F., Mwijage, A., Smith, J., Yeboah, E., van der Burg, W.J., Sanogo, O.M., Misiko, M., de Ridder, N., Karanja, S., Kaizzi, C., K'ungu, J., Mwale, M., Nwaga, D., Pacini, C., Vanlauwe, B., 2011. Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. *Agricultural Systems* 104, 191-203.
- Hair, J. F., Black, W.C, Babin, B.J, Anderson,R.E., 2010. *Multivariate Data Analysis: A Global Perspective*, Seventh Edition, Pearson.
- Hervé M., 2011. Aide-mémoire de statistique appliquée à la biologie – Construire son étude et analyser les résultats à l'aide du logiciel R, 2ème version.

- Howe, G., McKay, A., 2007. Combining quantitative and qualitative methods in assessing chronic poverty: The case of Rwanda. *World Development* 35, 197-211.
- Husson F., Lê S., Pagès J., 2011. *Exploratory Multivariate Analysis by Example Using R*. Chapman & Hall/CRC.
- Iraizoz, B., Gorton, M., Davidova, S., 2007. Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain. *Agricultural Systems* 93(1-3): 143-169.
- Kebede, B., 2007. Community wealth-ranking and household surveys : An integrative approach. Q-Squared Working paper No. 38, Centre for International Studies, University of Toronto, Canada.
- Kennedy, G., Ballard, T., & Dop, M., 2011. Guidelines for measuring household and individual dietary diversity. Rome, Italy: FAO.
- Kennedy, G., Nantel, G., Shetty, P., 2003. The scourge of 'hidden hunger': global dimensions of micronutrient deficiencies. *Food, Nutrition and Agriculture* 32, 8-16.
- Köbrich, C., Rehman, T., Khan, M., 2003. Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in Chile and Pakistan." *Agricultural Systems* 76(1): 141-157.
- Kostrowicki, J., 1977. Agricultural typology concept and method. *Agricultural Systems* 2, 33-45.
- Kuivanen, K., 2014. Patterns of persistence – characterizing smallholder farming system diversity in Ghana's Northern Region. MSc-thesis report, Farming Systems Ecology Group, Wageningen University, Wageningen.
- Kumar, R., 2014. Research methodology: a step-by-step guide for beginners – fourth edition. Sage, Los Angeles. <http://www.uk.sagepub.com/kumar4e/>
- Landais, E., 1998. Modelling farm diversity: new approaches to typology building in France. *Agricultural Systems* 58, 505-527.
- Laurent, C., van Rooyen, C.J., Madikizela, P., Bonnal, P., Carstens, J., 1999. Household typology for relating social diversity and technical change, *Agrekon: Agricultural Economics Research, Policy and Practice in Southern Africa*, 38: S1, 190-208.
- Madry, W., Mena, Y., Roskowska-Madra, B., Gozdowski, D., Hryniewski, R., Castel, J.M., 2013. An overview of farming system typology methodologies and its use in the study of pasture-based farming system: a review. *Spanish Journal of Agricultural Research* 11, 316-326.
- Maton, L., Leenhardt, D., Goalard, M., Bergez, J.E., 2005. Assessing the irrigation strategies over a wide geographical area from structural data about farming systems. *Agricultural Systems* 86, 293-311.
- Michielsens, C.G.J., Lorenzen, K., Phillips, M.J., Gauthier, R., 2002. Asian carp farming systems: towards a typology and increased resource use efficiency. *Aquaculture Research* 33, 403-413.
- Moreno-Pérez, O., Arnalte-Alegre, E., Ortiz-Miranda, D., 2011. Breaking down the growth of family farms: A case study of an intensive Mediterranean agriculture. *Agricultural Systems* 104, 500-511.
- Molua, E., 2011. Farm income, gender differentials and climate risk in Cameroon: typology of male and female adaptation options across agroecologies. *Sustainability Science* 6(1): 21-35.
- Muthayya, S., Rah, J.H., Sugimoto, J.D., Roos, F.F., Kraemer, K., Black, R.E., 2013. The global hidden hunger indices and maps: An advocacy tool for action. *PLoS ONE* 8(6): e67860.
- Pacini, G.C., Colucci, D., Baudron, F., Righi, E., Corbeels, M., Tittonell, P., Stefanini, F.M., 2013. Combining multi-dimensional scaling and cluster analysis to describe the diversity of the rural households. *Experimental Agriculture FirstView*, 1-22.

Remans, R., Flynn, D.F.B., DeClerck, F., Diru, W., Fanzo, J., 2011 Assessing Nutritional Diversity of Cropping Systems in African Villages. *PLoS ONE* 6(6):e21235.

Righi, E., Dogliotti, S., Stefanini, F.M., Pacini, G.C., 2011. Capturing farm diversity at regional level to up-scale farm level impact assessment of sustainable development options. *Agriculture, Ecosystems & Environment* 142, 63-74.

Rueff, C., Choisis, J.P., Balent, G., Gibon, A., 2012. A Preliminary Assessment of the Local Diversity of Family Farms Change Trajectories Since 1950 in a Pyrenees Mountains Area, *Journal of Sustainable Agriculture*

Sanogo, O.M., de Ridder, N., van Keulen, H., 2010. Diversité et dynamique des exploitations agricoles mixtes agriculture-élevage sud du Mali. *Cahier Agriculture*. 19, 185-193.

Thornton, P.K., Herrero, M., 2001. Integrated crop-livestock simulation models for scenario analysis and impact assessment. *Agricultural Systems* 70, 581-602.

Tittonell, P., Muriuki, A., Shepherd, K.D., Mugendi, D., Kaizzi, K.C., Okeyo, J., Verchot, L., Coe, R., Vanlauwe, B., 2010. The diversity of rural livelihoods and their influence on soil fertility in agricultural systems of East Africa – A typology of smallholder farms. *Agricultural Systems* 103, 83-97.

Tittonell, P., 2014, Categorising diversity of smallholder farming systems Household typologies. FSE staff seminar, 30 Jan 2014, Wageningen.

Tittonell, P., 2014. Livelihood strategies, resilience and transformability in African agroecosystems. *Agricultural Systems* 126(0): 3-14

Valbuena D., Groot, J., Mukalama, J., Gérard, B., Tittonell, P., 2014, Improving rural livelihoods as a “moving target”: trajectories of change in smallholder farming systems of Western Kenya. *Regional Environmental Change*, 1-13.

Van de Ven, G., Mellisse, B.T. (2014). Historic analysis of home garden agroforestry system changes in Ethiopia. Plant Production Systems Group, Wageningen University. [presentation]

Whatmore, S., Munton, R., Little, J.O., Marsden, T., 1987. Towards a typology of farm businesses in contemporary British Agriculture. *Sociologia Ruralis* 27, 21-37.

Apéndice A: precisión de la encuesta

Tomado de un informe realizado por Lotte Klapwijk (Septiembre 2013) sobre recolección de información detallada llevada a cabo en Tanzania para el proyecto AfricaRISING:

"... algunas veces hubo grandes diferencias entre los datos reportados en las encuestas que entregaron los encuestadores y los datos encontrados durante la recolección de información detallada. Por ejemplo, para varios campesinos, el número total de días laborales fue sorprendentemente bajo, llegando hasta 40 días por año. Después de expresar sorpresa y reformular la pregunta, las respuestas fueron diferentes. Esto podía ser el efecto de varias cosas, tales como una forma diferente de preguntar; preguntas hechas por una persona diferente (factor = encuestador), o una respuesta diferente simplemente porque es un día diferente (factor = campesino). Incluso ahora, dentro del período de la recolección de datos detallados, se encontraron algunas diferencias al comparar las notas del traductor con las notas del autor de este reporte (por ejemplo 60k por bolsa de maíz, contra 80k), lo cual muestra la dificultad de obtener claridad en la información. Otro problema es el hecho de que la definición de ciertas palabras o términos no siempre fue clara, o acordada de antemano. Por ejemplo, 'el año pasado' o 'la temporada pasada' significaban diferentes cosas para diferentes personas.

Según datos de la encuesta, un campesino poseía 39 acres de tierra, mientras que durante la segunda visita reportó un total de 22 acres. Sólo después de minutos de formular la pregunta de diferentes maneras, se logró aclarar la situación: hacía años el campesino había distribuido 5 acres de tierra a cada uno de sus 4 hijos mayores. Durante el resto de la entrevista no fue fácil separar la historia del campesino de la de esos cuatro hijos mayores; como en muchas familias africanas, estaban profundamente entrelazadas. Por ejemplo, las cosechas de los hijos se almacenaban en pilas separadas en una especie de kraal, pero cuando una pila se acababa, el padre se aseguraba de que habría todavía comida para alimentarse. (N.T: Kraal: asentamiento de chozas esparcido en forma de círculo, con un espacio en el centro para encerrar ganado. Fuente: Wikipedia)

Por otra parte, la base de recolección de datos es cuestionable; se pide a los campesinos, sin previo aviso, que recuerden números y cifras acerca de casi todo lo que acontece en sus vidas, cuando la mayoría de ellos no mantiene registro de ningún número. En relación con esto, algunas preguntas podrían ser casi imposibles de responder; la gente parecía tener gran dificultad para pensar en porcentajes. Tampoco es descabellado pensar que las personas algunas veces dan respuestas por razones sociales, o para no parecer tontos, o porque es más fácil (por ejemplo, un campesino aseguraba que sus animales necesitaban 12 horas de trabajo por día, mientras que durante la entrevista, al mediodía, estaban todavía en el kraal, y otro afirmaba que gastaba 12 horas por día con 7 cerdos). Como resultado, los datos pasaron de 100% de consumo doméstico de un producto de cierto cultivo a sólo el 25%. Para otro campesino, se pasó de dos campos de 1.5 acres a 1 campo de 2 acres, y otro más pasó de 2 campos de 7 y 8 acres a 3 campos de 5, 10 y 11 acres. Es necesario decir que otros cambios son tristes, pero reales y, sobre todo, muestran cómo es de variable y vulnerable la vida de los pequeños productores; uno perdió 39 de sus 45 aves de corral por la Enfermedad de Newcastle..."

Apéndice B: ejemplo de ACP y AJ

Este apéndice proporciona un ejemplo de los pasos para llevar a cabo un Análisis de Componentes Principales (ACP) y un Agrupamiento Jerárquico (AJ) para la construcción de tipologías usando lenguaje de programación R (software R), con un estudio de caso de un conjunto de datos de fincas en Tanzania.

El software R es una fuente de información abierta y disponible en Internet: <http://cran.r-project.org/>. En este ejemplo, el ACP y el AJ funcionan en R (versión 3.0.3) utilizando el [Paquete R ade4](#).

Preparación del conjunto de datos

Con el fin de ejecutar el ACP en R, el conjunto de datos debe organizarse en forma de tabla con las unidades de observación (fincas) en las filas y las variables en las columnas (Tabla A2). Los valores faltantes se podrían expresar con "NA" o como celdas vacías y se eliminarán en el primer paso del control de datos del ACP.

Tabla A1: Conjunto de datos de fincas tanzanas (n=174) utilizado para el ejemplo de ACP y AJ; el conjunto de datos es llamado "tanza".

obs	region	hhsz	area	totlabor	femratio	hiredratio	tlu	ncrop
1	1	13	6.27	3202	0.46	0.2	17.8	2
2	1	5	1.01	976	0.38	0.15	2.7	3
3	1	3	7.69	2455	0.29	0.83	19.3	3
4	1	7	1.42	1225	0.54	0.09	4.5	3
5	1	3	0.81	954	0.32	0.2	0.73	2
...								
174	2	5	4.96	2208	0.5	0.5	0.1	4

Se recomienda escoger nombres cortos para las variables (Tabla A3), sin espacio, y destacarlos para facilitar el trabajo en R.

Tabla A2: Variables del conjunto de datos *tanza*

	Código	Variable
1	obs	número de la finca
2	region	región en Tanzania
3	hhsz	tamaño del hogar (número de integrantes del hogar)
4	area	área del terreno (ha)
5	totlabor	total de mano de obra (horas/año ⁻¹)
6	femratio	proporción de mano de obra femenina (mano de obra femenina/mano de obra total)
7	hiredratio	proporción de mano de obra contratada (mano de obra contratada / mano de obra total)
8	tlu	unidad ganadera tropical (tlu: tropical livestock unit)
9	ncrop	número de cultivos

Control de datos para el ACP y el AJ

Uno de los primeros pasos para el funcionamiento del ACP es verificar los datos, es decir, encontrar valores faltantes, posibles errores, valores atípicos, correlaciones "fuertes" y controlar la distribución de las variables.

Para eliminar todos los valores faltantes en el conjunto de datos *tanza*:

```
tanza <- na.omit(tanza)
```

Los posibles errores, valores atípicos, correlaciones "fuertes" de *tanza* pueden detectarse gráficamente usando gráficos X-Y (diagramas) o gráficos de distribución (Figura A1 y Figura A2).

Para crear una matriz de diagramas X-Y para las variables (Figura A1) del conjunto de datos *tanza* (excepto la variable 1 (*obs*) y la variable 2 (*region*)):

```
pairs(tanza[,-c(1,2)], panel=panel.smooth)
```

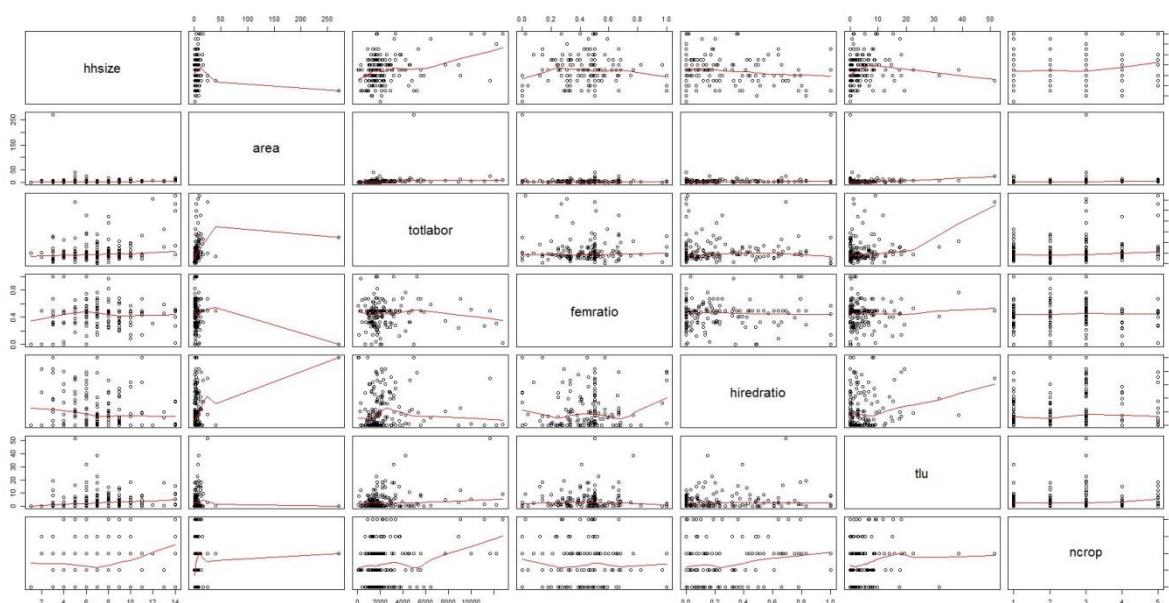


Figura A1: Matriz de diagramas X-Y para las variables de *tanza*

Para crear un gráfico con todas las distribuciones de las variables (Figura A2) de *tanza* (excepto las variables 1 y 2, *obs* y *region* respectivamente):

```
hist(tanza[,-c(1,2)])
```

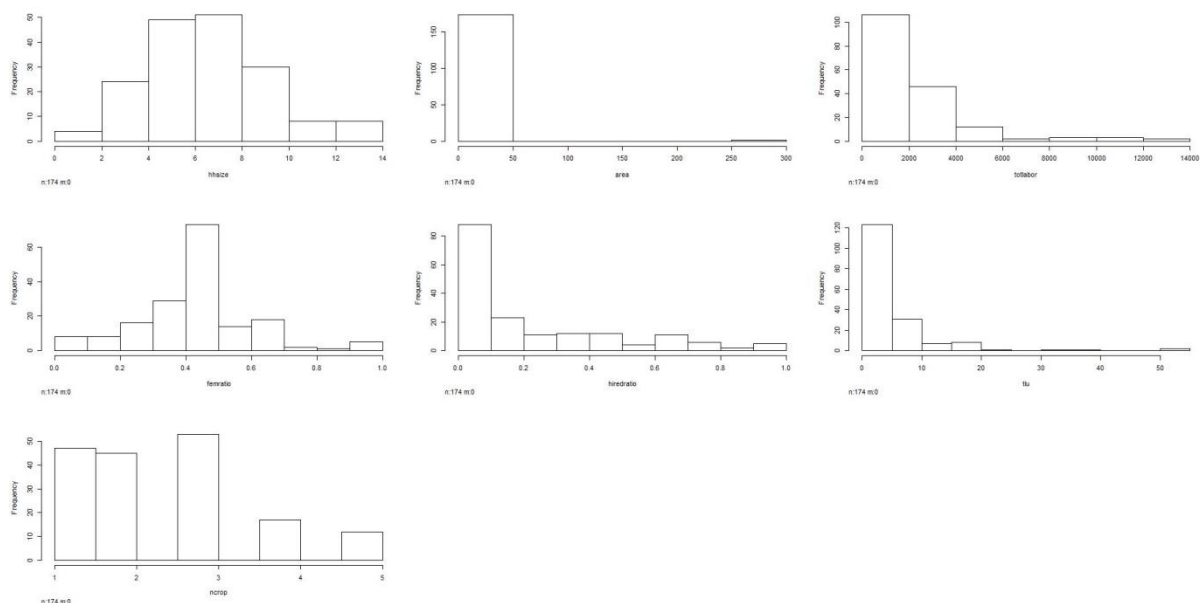


Figura A1: Distribuciones de las variables del conjunto de datos *tanza*

En las Figuras A1 y A2, observamos la existencia de valores atípicos en términos de área de campo (*area* > 50 ha) y en términos de ganado (*tlu* > 30). Estos valores atípicos podrían ser posibles errores o podrían ser casos existentes en el área. En ambas situaciones se aconseja quitarlos del ACP debido a su fuerte impacto sobre los resultados. Sin embargo, si estos valores atípicos son fincas existentes, podrían mencionarse en el resultado general de la tipología como fincas que son la excepción.

Para estudiar más detalladamente los valores atípicos, es posible crear un diagrama de caja por variable (por ejemplo para *area* y *tlu*, Figura A3):

```
boxplot(tanza$area)
boxplot(tanza$tlu)
```

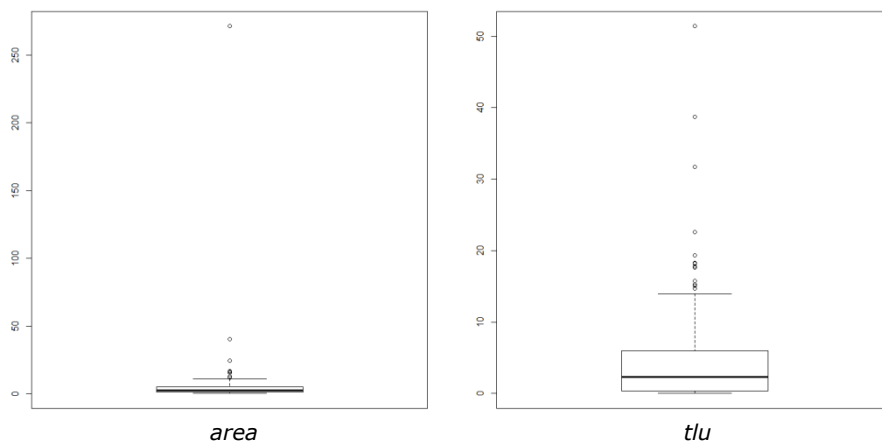


Figura A2: Diagramas de caja para las variables *area* y *tlu* del conjunto de datos *tanza*

Para eliminar los valores atípicos del conjunto de datos *tanza*:

```
tanza<- tanza[tanza$area < 50,]
tanza<- tanza[tanza$tlu < 30,]
```

Los nuevos diagramas de caja que resultan de la reducción de *tanza* están representados en la Figura A4: este tipo de figura puede ayudar a identificar con más precisión otros valores atípicos que tienen que ser excluidos para la ejecución del ACP.

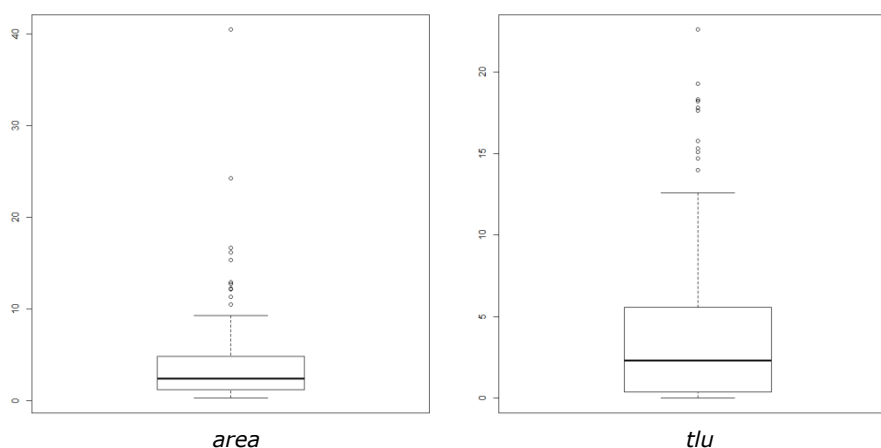


Figura A3: Nuevos diagramas de caja para las variables *area* y *tlu* del conjunto de datos *tanza*

Con base en lo que se observa en los nuevos diagramas de caja, elegimos realizar el ACP con un subconjunto *tanza* con *area* < 20 ha y *tlu* < 20:

```
tanza<- tanza[tanza$area < 20,]
tanza<- tanza[tanza$tlu < 20,]
```

Comentarios: aquí hay todavía valores atípicos, pero consideramos que quedan lo suficientemente agrupados como para formar un tipo de finca.

Ya que para el ACP se recomienda utilizar variables cuantitativas con distribución normal o al menos simétrica, algunas variables de *tanza* necesitan ser transformadas (Figura A5). Por lo tanto se crearon nuevas variables (p. ej. *area2*, *totlabor2*, *femratio2*) aplicando logaritmo (*log10*), raíz cuadrada (*sqrt*) u otras funciones a las variables originales:

```
tanza$area2 <- log10(tanza$area)
tanza$totlabor2 <- log10(1+tanza$totlabor)
tanza$femratio2 <- log10(1+tanza$femratio)
tanza$hiredratio2 <- sqrt(tanza$hiredratio)
tanza$tlu2 <- sqrt(tanza$tlu)
tanza$ncrop2 <- sqrt(tanza$ncrop)
```

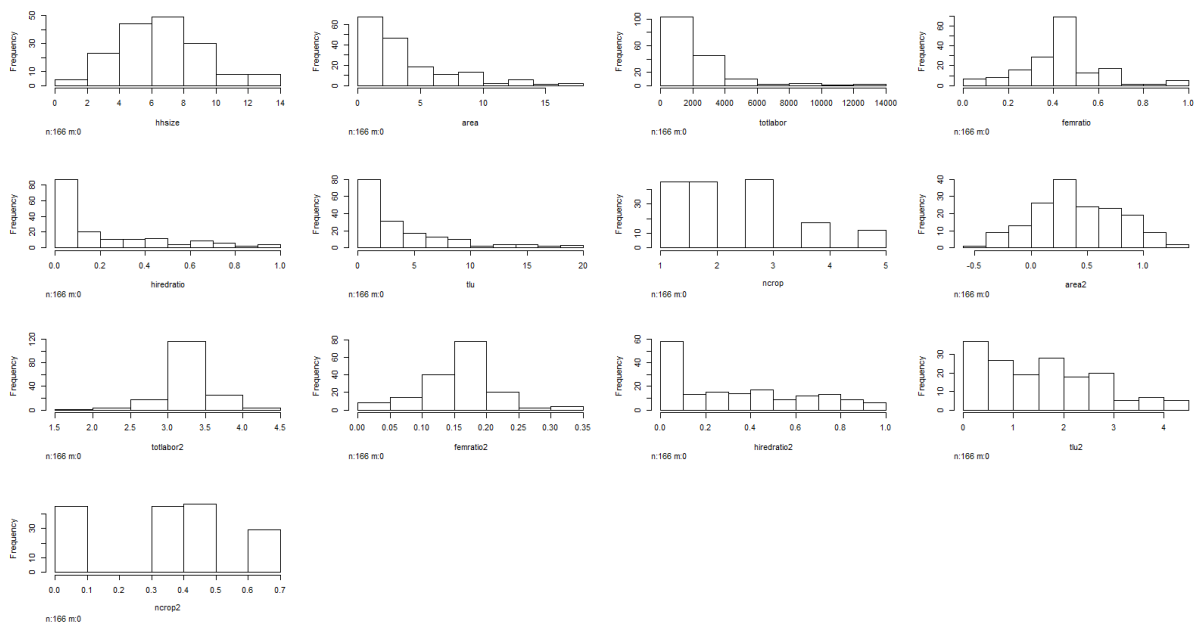


Figura A4: Distribuciones de las variables del conjunto de datos *tanza* con las variables transformadas.

Comentarios: incluso si se mejora la distribución de las variables, alcanzar una “buena” simetría puede ser difícil para algunas variables, tales como *hiredratio2* and *tlu2* (Figura A5).

ACP sobre los datos transformados

El paquete R utilizado para el análisis ACP es “ade4”; por lo tanto debe ser instalado y cargado:

```
install.packages("ade4")
library(ade4)
```

ACP No. 1

La función `dudi.pca()` se utiliza para operar el ACP sobre el conjunto de datos *tanzaT*, el cual contiene las variables transformadas:

```
tanzaT <- tanza[,match(c("hysize", "area2", "totlabor2", "femratio2",
"hiredratio2", "tlu2", "ncrop2"), dimnames(tanza)[[2]])]
tanza.pca <- dudi.pca(tanzaT, center=T, scale=T, scannf=T, nf=5)
> Select the number of axes: 3
```

La función `dudi.pca()` despliega automáticamente el diagrama de barras de los valores propios (Figura A6) con el fin de ayudar a seleccionar el número de componentes principales (CP) que se van a guardar para el análisis posterior. El uso del Criterio Kaiser promueve la selección de todos los ejes que tengan valores propios superiores a 1:

```
tanza.pca$eig
> 1.8996519 1.1906887 1.0549658 0.9962744 0.7951804 0.6305626 0.4326761
```

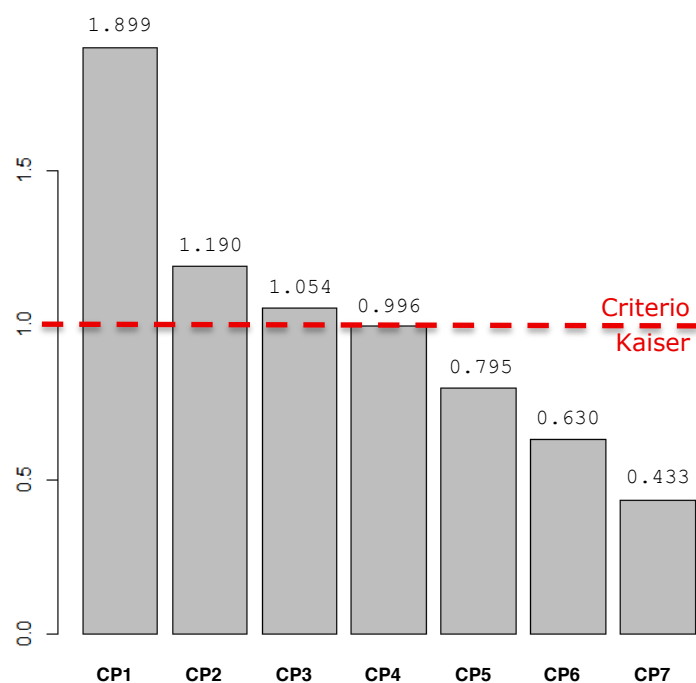


Figura A5: Diagrama de barras de los valores propios por componente principal (CP)

La selección del número de CP también podría ser determinada por el porcentaje de variabilidad explicada que se quiera representar. Para calcular el porcentaje acumulado de variabilidad explicada por el CP:

```
cumsum(tanza.pca$eig) / sum(tanza.pca$eig)
> 0.2713788 0.4414772 0.5921866 0.7345116 0.8481088 0.9381891 1.0000000
```

En este caso, eso significa que con CP1 y CP2 se explica alrededor del 44.1% de la variabilidad de las fincas y con CP1, CP2 y CP3 se explica cerca del 59.2% de la variabilidad de las fincas. En ese paso elegimos conservar 3 ejes (CP1, CP2 y CP3)

Debe notarse que con más CP la interpretación de los resultados finales del ACP y el AJ se torna más difícil.

La interpretación de los CP está basada en los círculos de correlación (Figura A7) y en la matriz de correlación (Tabla A4)

Se utiliza la función `s.corcircle()` para crear los círculos de correlación:

```
s.corcircle(tanza.pca$co, xax=1, yax=2 )
s.corcircle(tanza.pca$co, xax=1, yax=3 )
```

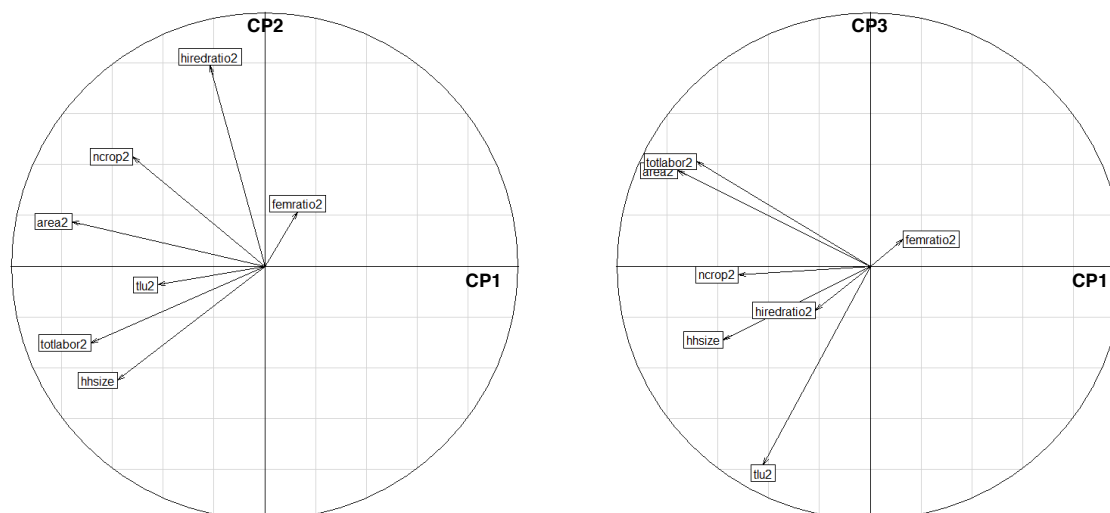


Figura A6: Círculos de correlación para los componentes principales CP1-CP2 y CP1-CP3.

Los coeficientes de correlación entre los CP y las variables están contenidos en el objeto `tanza.pca$co`.

Tabla A3: Matriz de correlación entre los componentes principales (CP) y las variables del conjunto de datos `tanzaT`.

Variables	CP1	CP2	CP3
hhsiz2	-0.581	-0.450	-0.291
area2	-0.761	0.176	0.377
totlabor2	-0.686	-0.303	0.412
femratio2	0.130	0.214	0.106
hiredratio2	-0.216	0.793	-0.175
tlu2	-0.422	-0.074	-0.785
ncrop2	-0.520	0.431	-0.034

Aquí las variables `area2`, `totlabor2` y `hhsiz2` se correlacionan (negativamente) con el CP1, `hiredratio2` se correlaciona con el CP2 y `tlu2` se correlaciona con el CP3 (Figura A7 y Tabla A4). En resumen, el CP1 expresa información general acerca del tamaño de la finca (en términos de tierra y mano de obra), el CP2 expresa la importancia relativa de la mano de obra contratada y el CP3 expresa el tamaño de los hatos ganaderos.

Aquí la variable `femratio2` no está bien representada en ninguno de los planos ACP1-ACP2 y ACP1-ACP3 (flecha corta en la Figura A7). La variable `ncrop2` no parece aportar información adicional para el ACP: está correlacionada con `area2` y podría ser redundante con el ACP1, el cual ya está bien definido por `area2` and `totlabor2`. Por otra parte, por la estructura de la variable, `ncrop2` parece ser una variable categórica (Figura A1). De ahí que podríamos considerar eliminar una de esas dos variables o ambas para el siguiente intento de ACP. Para este ejemplo se escoge excluir la variable `ncrop2` en la próxima ejecución del ACP con el fin de probar si aumenta el porcentaje de la variabilidad explicada.

La función `s.label()` se utiliza para representar las unidades de observación (fincas) en el plano CP1-CP2 o CP1-CP3 (Figura A8):

```
s.label(tanza.pca$li, xax=1, yax=2)
s.label(tanza.pca$li, xax=1, yax=3)
```

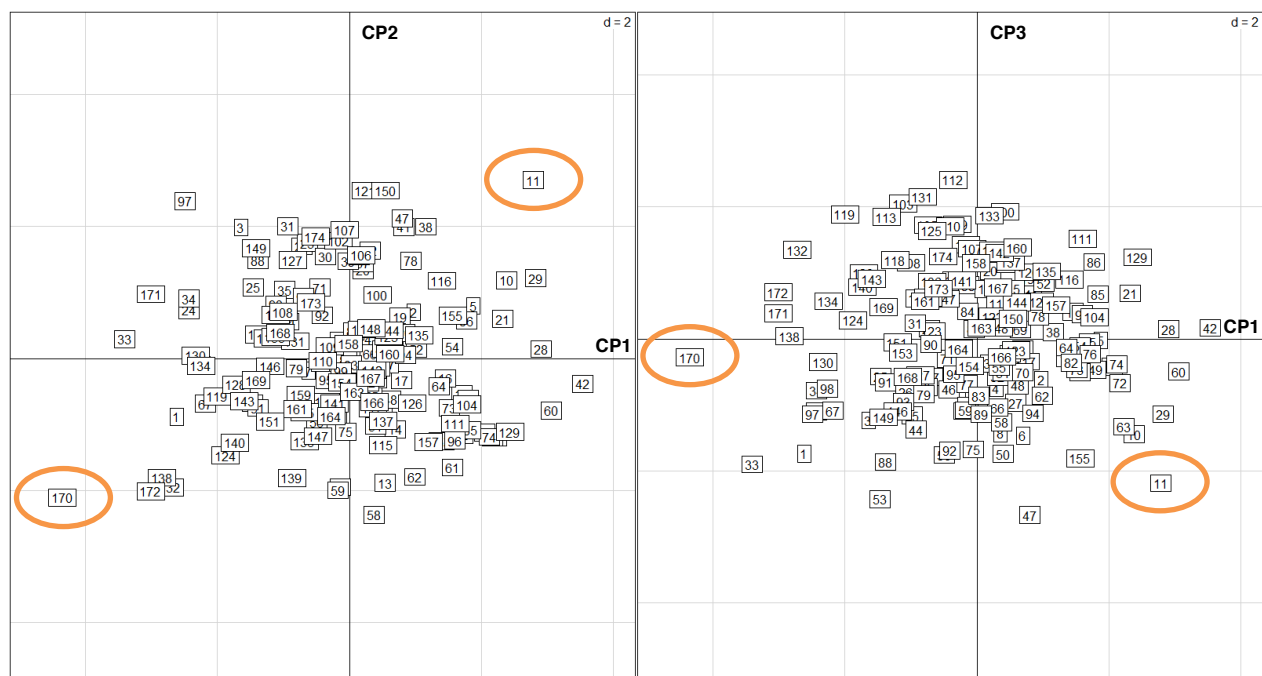



Figura A7: Localización de las fincas en los planos de componentes principales CP1-CP2 y CP1-CP3.

La Figura A8 ayuda a identificar posibles valores atípicos; aquí por ejemplo las fincas 11 y 170 están muy aisladas de las demás fincas. Los valores atípicos tienen un fuerte impacto sobre los resultados del ACP debido al procedimiento del ACP, el cual correlaciona linealmente los CP y las variables del conjunto de datos; estos valores atípicos, por lo tanto, afectan seriamente la inclinación de la línea de regresión lineal (Figura A9). En la próxima ejecución del ACP se podrían dejar por fuera estas dos fincas.

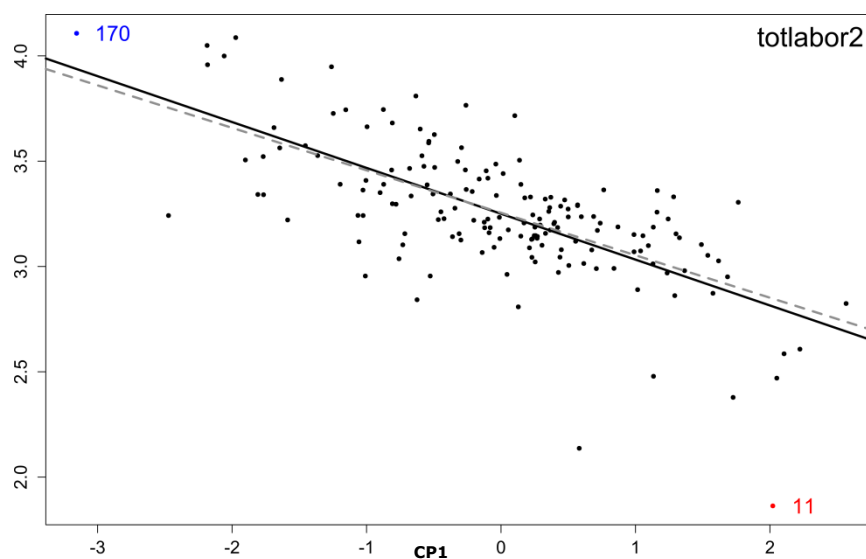


Figura A8: Regresión lineal entre la variable *totlabor2* de *tanzaT* y el Componente Principal CP1; la línea negra es la regresión lineal incluyendo las fincas 11 y 170, y la línea punteada representa el cambio en la inclinación al dejar por fuera las fincas 11 y 170.

ACP No. 2

En este ejemplo, la segunda ejecución del ACP se lleva a cabo sin las fincas 11 y 170 (valores atípicos según la Figura A8) y sin la variable *ncrop2*.

Para excluir las fincas 11 y 170 del conjunto de datos *tanza* (una por una):

```
tanza <- tanza[tanza$obs!=11,]
tanza <- tanza[tanza$obs!=170,]
```

Para crear un nuevo *tanzaT* sin la variable *ncrop* y ejecutar de nuevo el ACP:

```
tanzaT <- tanza[,match(c("hhsz", "area2", "totlabor2", "femratio2",
"hiredratio2", "tlu2" ), dimnames(tanza)[[2]])]
tanza.pca <- dudi.pca(tanzaT, center=T, scale=T, scannf=T, nf=5)
> Select the number of axes: 4
```

Para verificar los valores propios:

```
tanza.pca$eig
> 1.7208478 1.1131360 1.0632992 1.0017315 0.6261152 0.4748704
```

Para confirmar el porcentaje de la variabilidad explicada por los CP:

```
cumsum(tanza.pca$eig) / sum(tanza.pca$eig)
> 0.2868080 0.4723306 0.6495472 0.8165024 0.9208549 1.0000000
```

Ahora el CP1 y el CP2 juntos explican el 47.2% de la variabilidad de las fincas, el CP1, CP2 y CP3 juntos explican el 64.9% de la variabilidad y el CP1, CP2, CP3 y CP4 juntos explican el 81.6% de la variabilidad.

Aquí el Criterio Kaiser recomienda utilizar para el análisis los cuatro CP, pero también es posible que se decida usar únicamente tres CP, con lo cual se puede explicar cerca del 65% de la variabilidad.

Utilizamos el círculo de correlación (Figura A10) y la matriz de correlación (Tabla A5) para interpretar el significado de CP1, CP2, CP3 y CP4:

```
s.corcircle(tanza.pca$co, xax=1, yax=2)
s.corcircle(tanza.pca$co, xax=1, yax=3)
s.corcircle(tanza.pca$co, xax=1, yax=4)
```

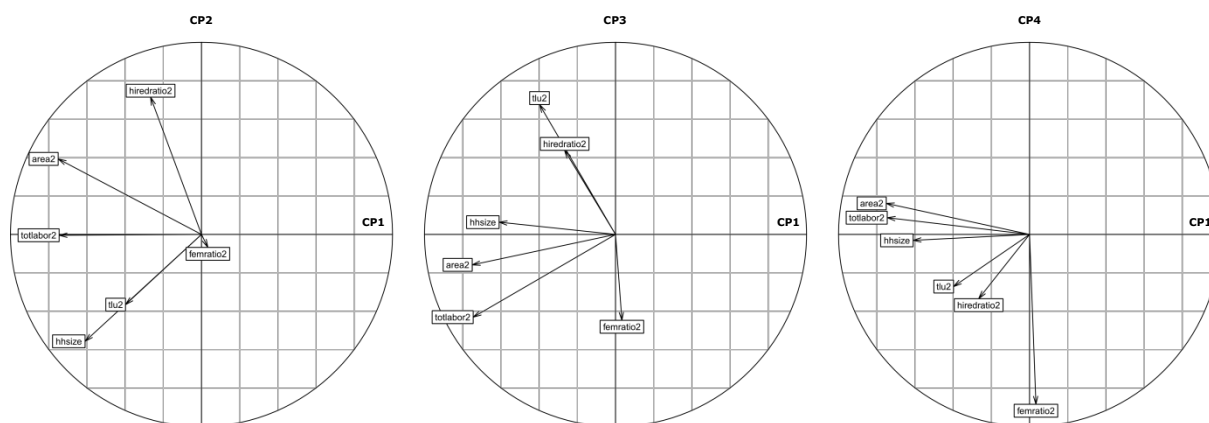


Figura A9: Círculos de correlación para los componentes principales CP1-CP2, CP1-CP3 y CP1-CP4.

Tabla A4: Matriz de correlación entre los componentes principales (CP) y las variables del conjunto de datos *tanzaT*.

Variables	CP1	CP2	CP3	CP4
hhsiz	-0.610	-0.556	0.064	-0.031
area2	-0.750	0.393	-0.158	0.162
totlabor2	-0.745	-0.004	-0.431	0.087
femratio2	0.033	-0.067	-0.447	-0.884
hiredratio2	-0.267	0.715	0.438	-0.335
tlu2	-0.398	-0.365	0.676	-0.271

Es conveniente revisar nuevamente los posibles valores atípicos (Figura A11):

```
s.label(tanza.pca$li, xax=1, yax=2)
s.label(tanza.pca$li, xax=1, yax=3)
s.label(tanza.pca$li, xax=1, yax=4)
```

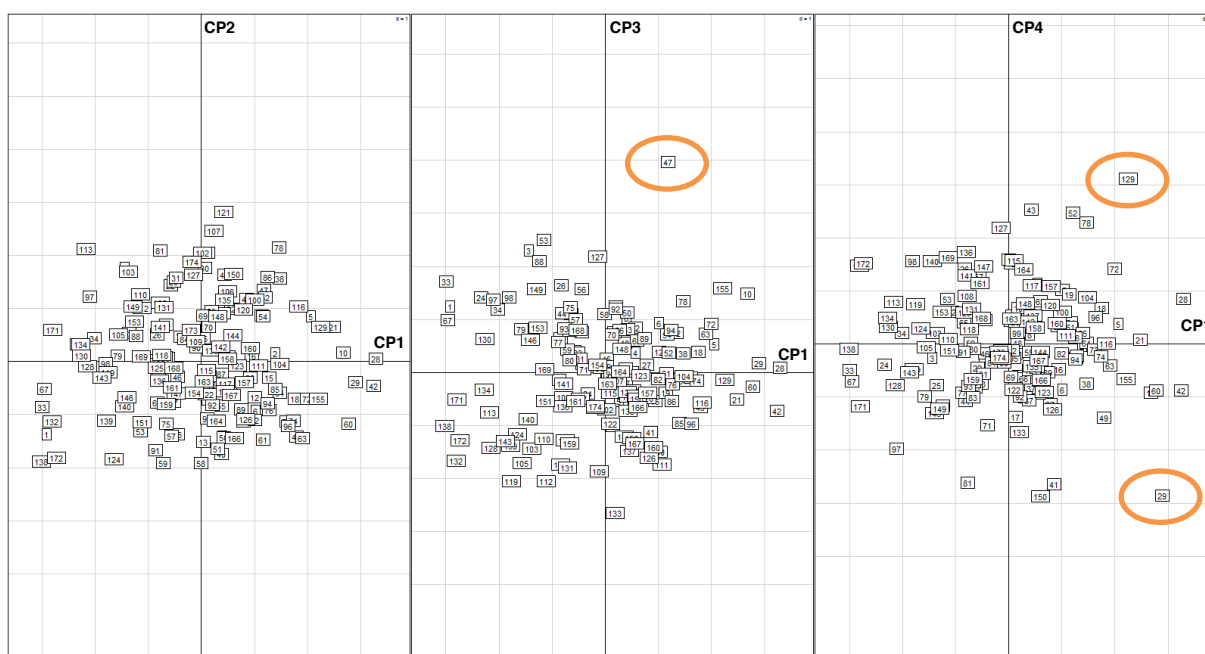


Figura A10: Campesinos representados gráficamente en los planos de componentes principales CP1-CP2, CP1-CP3 y CP1-CP4.

La Figura A11 ayuda a identificar nuevos valores atípicos plausibles; en la siguiente ejecución del ACP se escogió dejar por fuera la finca 47, aislada en el plano CP1-CP3 y las fincas 129 y 29, aisladas en el plano CP1-CP4.

ACP No. 3

La tercera ejecución del ACP se realiza sin las fincas 47, 129 y 29:

```
tanza <- tanza[tanza$obs!=47,]
tanza <- tanza[tanza$obs!=129,]
tanza <- tanza[tanza$obs!=29,]

tanzaT <- tanza[,match(c("hsize" ,"area2",
"totlabor2","femratio2","hiredratio2","tlu2"), dimnames(tanza)[[2]])]
tanza.pca <- dudi.pca(tanzaT, center=T, scale=T, scannf=T, nf=5)
> Select the number of axes: 3

tanza.pca$eig
> 1.7695090 1.1235104 1.0241344 0.9887622 0.6217878 0.4722962

cumsum(tanza.pca$eig) / sum(tanza.pca$eig)
> 0.2923875 0.4841733 0.6562492 0.8196921 0.9197690 1.0000000
```

Para mostrar los nuevos círculos de correlación (Figura A12) y la matriz de correlación (Tabla 6):

```
s.corcircle(tanza.pca$co, xax=1, yax=2)
s.corcircle(tanza.pca$co, xax=1, yax=3)

tanza.pca$co
```

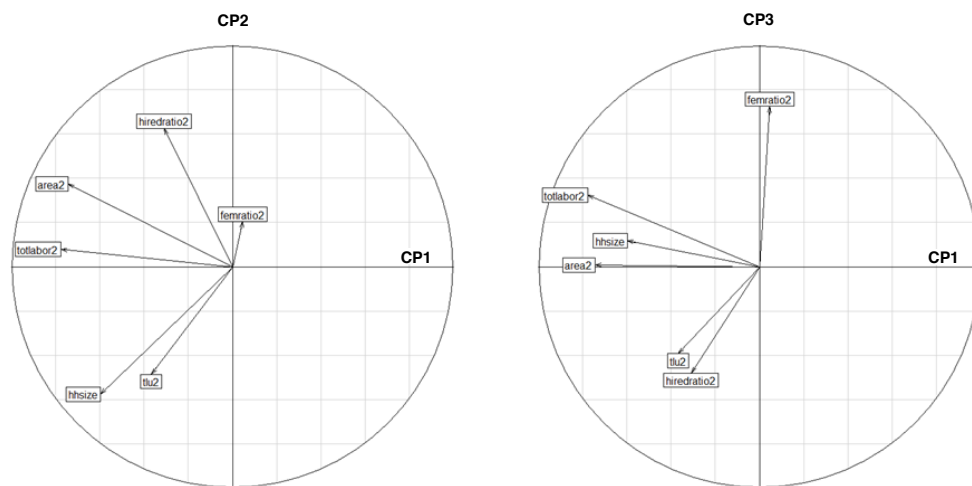


Figura A11: Círculos de correlación para los componentes principales CP1-CP2 y CP1-CP3.

Tabla A5: Matriz de correlación entre los componentes principales (CP) y las variables del conjunto de datos *tanzaT*.

	CP1	CP2	CP3
CP1	CP2	CP3	0.117
area2	-0.745	0.374	0.008
totlabor2	-0.778	0.079	0.324
femratio2	0.047	0.205	0.726
hiredratio2	-0.310	0.627	-0.481
tlu2	-0.368	-0.487	-0.394

De acuerdo con los resultados del ACP (Figura A12 y Tabla A6), aquí las variables discriminantes (las que mejor segregan las fincas) son *area2*, *totlabor2*, *hiredratio2* y *femratio2*.

En resumen, el CP1 expresa la información general sobre el tamaño de la finca, el CP2 la importancia relativa de la mano de obra contratada y el CP3 la importancia relativa de la mano de obra femenina.

Comentarios: el hecho de dejar por fuera las tres fincas con valores atípicos en el ACP anterior (Figura A11) produjo un fuerte impacto sobre los resultados del nuevo ACP (*tlu2* ya no está estrechamente correlacionado con los principales CP).

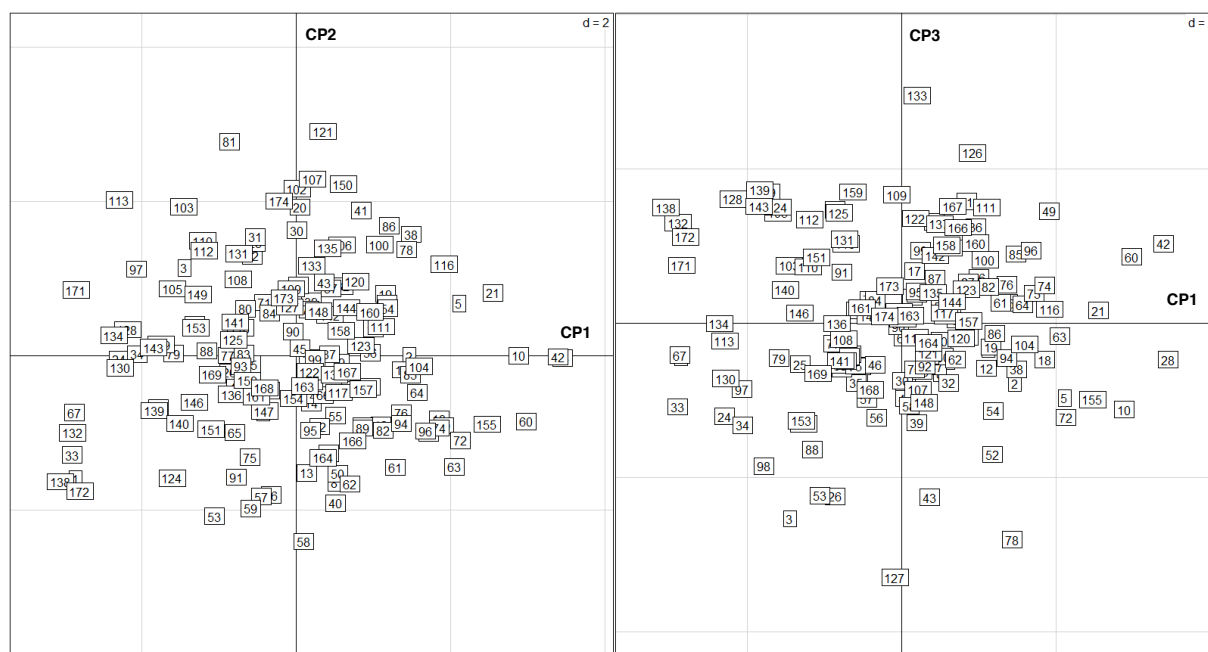


Figura A1: Campesinos representados gráficamente en los planos de componentes principales CP1-CP2 y CP1-CP3.

Para superponer el círculo de correlación y la unidad de observación en el plano CP1-CP2 (Figura A13):

```
scatter.dudi(tanza.pca)
```

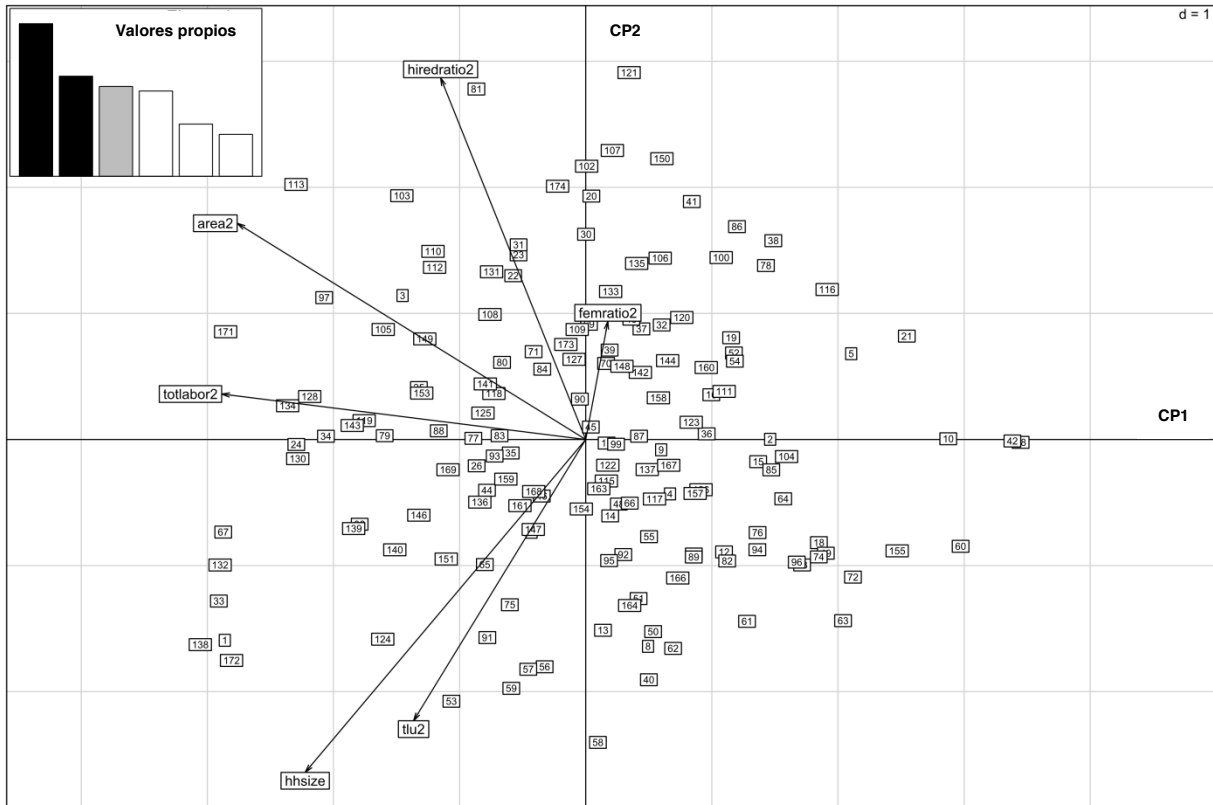


Figura A2: Campesinos y variables representados gráficamente en el plano de componentes principales CP1-CP2.

Análisis clúster sobre los resultados del ACP

Aquí se aplicó el Agrupamiento Jerárquico (AJ) a los resultados del ACP con la función `hclust()` utilizando el método Ward:

```
tanza.cah <- hclust(dist(tanza.pca$li), method="ward")
```

El dendrograma y el diagrama de barras de las alturas (Height) ayudan a escoger el número de clústeres (o tipos) que se van a utilizar (Figura A15):

```
barplot(tanza.cah$height)
```

```
plot(tanza.cah)
```

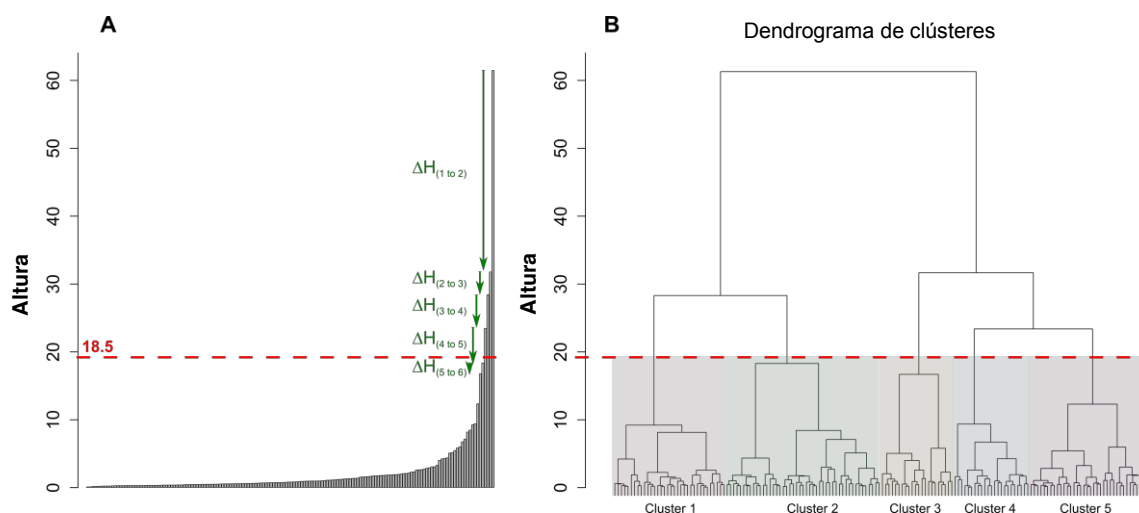


Figura A3: Diagrama de barras de la altura (A) y dendrograma (B). $\Delta H_{(n \text{ a } n+1)}$ corresponde al delta de Altura (disminuye) pasando n clústeres a $n+1$ clústeres.

En la Figura A15, la Altura es un indicador de la disimilitud dentro de los clústeres en relación con el número de clústeres. La barra del extremo derecho muestra la disimilitud máxima, es decir, la disimilitud cuando todas las fincas están agrupadas en un clúster. La revisión de las alturas disminuye de derecha a izquierda en la figura del diagrama de barras ($\Delta H_{(4 \text{ a } 5)} > \Delta H_{(5 \text{ a } 6)}$) y la estructura general del dendrograma sugiere hacer la partición del dendrograma ("cortar el árbol") a una Altura de aproximadamente 18.5 (línea punteada en la Figura A15), lo cual lleva a dividir el dendrograma en 5 clústeres.

Para separar las unidades de observación (fincas) en cinco clústeres (k es el número de clústeres deseados):

```
tanza.type <- cutree(tanza.cah, k=5)
```

Por último, para visualizar e interpretar los clústeres en los planos CP1-CP2 y CP1-CP3 (Figura A16):

```
s.corcircle(tanza.pca$co, xax=1, yax=2)
s.class(tanza.pca$li, fac = as.factor(tanza.type))
s.corcircle(tanza.pca$co, xax=1, yax=3)
s.class(tanza.pca$li, xax=1, yax=3, fac = as.factor(tanza.type))
```

Para agregar al conjunto de datos *tanza* una columna (*tipo*) que contenga los números de tipo de cada finca:

```
tanza$tipo <- tanza.type
```

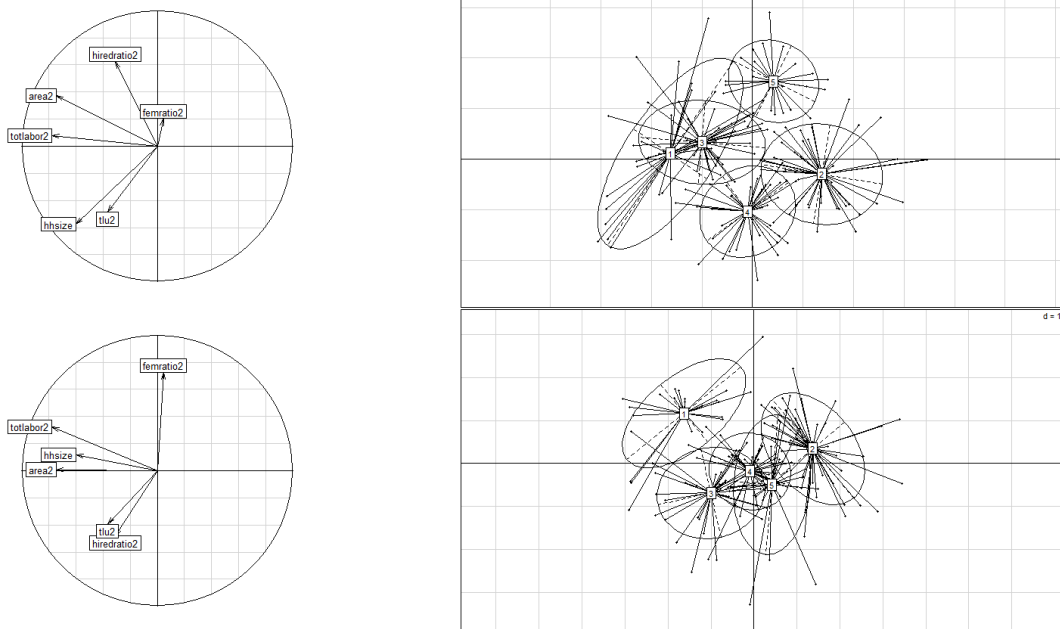


Figura A4: Resultado del Análisis de Componentes Principales y el Agrupamiento Jerárquico.

El próximo paso para la tipología

El paso siguiente es interpretar cada clúster (o tipo) utilizando los gráficos o resultados del ACP (Figura A16) y los cálculos estadísticos para cada clúster (p. ej. medias, rangos, etc.).

Luego, es necesario comparar el significado de cada tipo (es decir, los resultados de la interpretación de los clústeres) con el conocimiento acerca del área (Hipótesis). Las fincas que fueron excluidas se podrían mencionar como fincas atípicas que existen en la zona. Si los clústeres no tienen ningún significado según el conocimiento local (p. ej. que el papel que juega la variable *femratio* sea demasiado "fuerte"), se debería reiniciar el ACP.

Por último, es deseable que expertos locales validen los resultados de la tipología.