

Modelación numérica de eco y socioecosistemas.
Diversos enfoques con lenguaje R.

Vega, E. (compilador)

2024-02-28

Contents

1	Presentación	5
1.1	Agradecimientos	6
1.2	Referencias	6
2	Optimización uni y multiobjetivo	9
2.1	Presentación.	9
2.2	Sistemas de ecuaciones lineales	9
2.3	Aspectos básicos de los modelos de equilibrio parcial	16
2.4	Generalidades de la modelación multiobjetivo (MMO)	31
2.5	Referencias	60
3	Análisis de conectividad del paisaje.	61
3.1	Introducción	61
3.2	Modelación de nicho ecológico y distribución de especies	62
3.3	Referencias	75
4	Modelación de datos globales con NETCDF	77
4.1	Introducción.	77
4.2	El origen de las mediciones globales	78
4.3	El formato NetCDF	79
4.4	Instalación del formato NetCDF en cualquier computadora y librerías correspondientes en R.	80
4.5	Procesamiento de datos globales en formato NetCDF en R.	81
4.6	Ejemplos y sitios web relevantes para encontrar datos.	92
4.7	Conclusiones.	94

5	Modelación de procesos biológicos con ecuaciones diferenciales ordinarias	95
5.1	Introducción	95
5.2	Nociones básicas de modelos dinámicos.	97
5.3	Algunas propiedades de los sistemas dinámicos.	100
5.4	Ecuaciones diferenciales ordinarias.	102
5.5	Ecuaciones diferenciales ordinarias: introducción, definiciones y soluciones.	106
5.6	Construcción de un modelo de EDOs	115
5.7	Construcción modelo exponencial	117
5.8	Construcción modelo logístico	121
5.9	Construcción del modelo depredador-presa	125
5.10	Sección 6: Modelado de Abundancia y Riqueza (MAR)	126
5.11	Modelo de nicho	130
5.12	5.2. El modelo MAR	132
5.13	Conclusiones/Discusión, posibles cursos lecturas posteriores	136
5.14	Bibliografía	137
6	Modelos de ecuaciones estructurales (SEM)	141
6.1	Presentación	141
6.2	Preliminares	143
6.3	Modelo del “PIB Mediador”	146
6.4	Bibliografía	154

Chapter 1

Presentación

Estamos en una etapa importante en la investigación en las ciencias del ambiente. Por un lado, el conocimiento generado hasta ahora sugiere la necesidad de incorporar a la Ecología conceptos nuevos (la variación espacial explícita, los estados estables alternos, etc.) y nuevas maneras de medirlos. Por otro lado, el intensivo uso del medio por parte de las sociedades industrializadas ha puesto en riesgo los ciclos de regeneración y de homeostasis de los múltiples subsistemas del planeta, por lo que es urgente realizar caracterizaciones precisas y reales acerca del funcionamiento de los sistemas ecológicos y de su interacción con los sociales.

Se necesita información sólida acerca del funcionamiento de los sistemas ecológico-ambientales, para tomar decisiones de gestión y uso que no sean tan desacertadas. Una porción de esa información puede generarse como consecuencia de concebir a los sistemas socioecológicos como complejos, queriendo decir con esto lo siguiente: *sistemas con elementos interconectados que pueden mostrar “comportamientos” espaciales y temporales no codificados en sus componentes*. Este tipo de sistemas presentan “conductas sorprendidas”, es decir, pueden presentar estados estables alternos o incluso desaparecer, en términos funcionales (Scheffer y Carpenter, 2003).

Entender cómo se vinculan las sociedades con la dinámica del ambiente es quizás el reto más importante y urgente para implementar estrategias sustentables de consumo de recursos naturales. Si bien se han desarrollado avances en la comprensión de la dinámica de los sistemas naturales y de los sociales por separado, la modelación de la interacción entre ambos sistemas ofrece un enorme reto, tanto para las disciplinas básicas como las aplicadas. Sin embargo, los sistemas sociales y naturales, a pesar de ser cualitativamente diferentes, poseen propiedades compartidas, pues ambos son complejos (Watts y Strogatz 1998, Gordon et al. 2008, Mitchell 2009, Deffuant y Gilbert 2011).

Este documento pretende ofrecer un conjunto de métodos de análisis muy dis-

tintos, pero que poseen una “visión compartida”: *el sistema de estudio puede representarse con cajitas y flechitas que conectan a las cajitas*. Esta representación (pareciera trivial) es la base para el desarrollo de las estrategias de modelación ofrecidas en este documento.

Este modelo básico aparece claramente en los capítulos de modelación de sistemas dinámicos con ecuaciones diferenciales, de ecuaciones estructurales y de conectividad del paisaje. Y con un poquito más de esfuerzo podrán encontrarlo en los demás capítulos.

Aquí verán un conjunto de tópicos aparentemente inconexos y que tratan de contestar una o algunas de estas preguntas generales acerca de los sistemas Socioecológicos:

- * Cómo estudiar los cambios temporales de un sistema?
- * Cómo evaluar las relaciones de causa y efecto en un conjunto de variables?
- * Cómo encontrar la mejor manera para que un sistema funcione?
- * Cómo analizar el efecto del clima en un sistema determinado?
- * Cómo evaluar el grado de conexión entre dos o más partes de un sistema?

1.1 Agradecimientos

A la UNAM y al PAPIME que financió este proyecto (PAPIME PE200623 “INTRODUCCION A LA MODELACION DE SISTEMAS SOCIOECOLOGICOS, CON METODOS DE SISTEMAS COMPLEJOS”).

A mis colegas “coperpetradores” de este documento: Alejandro, César, Eugenio, Fernando, Germán, Memo, Norma, Pacho, Rafa. Sin ellos este ejercicio no se hubiera podido hacer.

1.2 Referencias

- Deffuant, G., Gilbert, N. 2011. Variability and Resilience of Complex Systems. Springer.
- Gordon, L.J., Peterson, G.D. and E.M. Bennet. 2008. Agricultural modifications of hydrological flows create ecological surprises. Trends in Ecology and Evolution, 23: 211-219.
- Mitchell, M. 2009. Complexity. A Guided Tour. Oxford University Press.
- Scheffer, M., and S.R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in Ecology and Evolution, 18: 648-656.

- Watts, D. and Strogatz, S.H. 1998. Collective dynamics of “small-world” networks. *Nature*, 393(4):440-442.

Chapter 2

Óptimización uni y multiobjetivo

Dr. German Wies (IIES, UNAM Campus Morelia)
Dr. Ernesto Vega (IIES, UNAM Campus Morelia)

2.1 Presentación.

El manejo de socioecosistemas implica la conciliación de objetivos productivos y ambientales, que frecuentemente pueden ser antagónicos. Por ello es pertinente usar herramientas que permitan identificar de qué forma pueden coexistir en una misma región, tanto intereses económicos como de conservación del ambiente (Wies et al. 2023).

En este capítulo revisaremos los aspectos básicos de dos de estas herramientas: los *modelos de equilibrio parcial* y la *modelación multiobjetivo*. La primera de ellas optimiza un sólo objetivo, mientras que la segunda lo hace con dos o más. Ambos enfoques ofrecen perspectivas complementarias para entender mejor la gestión de sistemas socioambientales.

2.2 Sistemas de ecuaciones lineales

La base de estas estrategias de modelación es la línea recta:

$$y = mx + b \tag{1}$$

Tanto x como y se denominan variables, mientras que m se denomina coeficiente y la b es una constante.

Este modelo puede representarse de otra forma:

$$mx - y = b \quad (2)$$

Y se puede generalizar un poco más:

$$ax + by = c \quad (3)$$

Los coeficientes a y b no pueden valer cero simultáneamente.

Este es un ejemplo con dos variables (i.e. dos dimensiones), que puede modificarse para representar sistemas con muchas dimensiones. Este es un ejemplo que involucra tres dimensiones:

$$ax + by + cz = d \quad (4)$$

Nuevamente, la terna (a, b, c) no pueden ser cero simultáneamente.

2.2.1 Concepto de linealidad

Los anteriores se conocen como modelos lineales, porque las variables aparecen a la primera potencia (no son raíces cuadradas o productos) y no están como argumentos de funciones logarítmicas, exponenciales ni trigonométricas.

Estos son ejemplos de ecuaciones no lineales:

$$\begin{aligned} \sin(x) + y &= 0 \\ 3x + 2y - xy &= 0 \end{aligned}$$

2.2.2 Sistemas de ecuaciones lineales.

Ahora imaginemos **un conjunto de ecuaciones lineales** que comparten las mismas variables:

$$\begin{aligned} 5x + y &= 3 \\ 2x - y &= 4 \end{aligned} \tag{5}$$

De manera general este sistema se representa así:

$$\begin{aligned} a_{11}x_{11} + a_{12}x_{12} + a_{13}x_{13} + \cdots + a_{1n}x_{1n} &= b_1 \\ a_{21}x_{21} + a_{22}x_{22} + a_{23}x_{23} + \cdots + a_{2n}x_{2n} &= b_2 \\ &\dots \\ a_{m1}x_{m1} + a_{m2}x_{m2} + a_{m3}x_{m3} + \cdots + a_{mn}x_{mn} &= b_m \end{aligned} \tag{6}$$

Este sistema tiene m ecuaciones lineales (filas) y n variables o incógnitas (columnas). El primer subíndice de los coeficientes en a_{xy} corresponde a la ecuación, mientras que el segundo corresponde a la variable. Un sistema “cuadrado” es en el que $m = n$ (hay tantas ecuaciones como variables).

El reto fundamental que ofrece un sistema de ecuaciones es **encontrar aquellos valores de las variables que permitan que se cumplan todas las ecuaciones (cumplir significa que la igualdad expresada en la ecuación sea verdadera)**. A este conjunto de valores de x se le conoce como *solución del sistema de ecuaciones*.

2.2.3 Notación matricial de un sistema de ecuaciones lineales.

El sistema de ecuaciones mostrado en (6) tiene una “representación simplificada”:

$$A * x = b \tag{7}$$

en la que \mathbf{A} es una matriz de coeficientes a_{xy} :

$$\mathbf{A} = \left\langle \begin{array}{cccc} a_{11}, a_{12}, a_{13}, \dots, a_{1n} \\ a_{21}, a_{22}, a_{23}, \dots, a_{2n} \\ \dots \\ a_{m1}, a_{m2}, a_{m3}, \dots, a_{mn} \end{array} \right\rangle$$

El vector x corresponde a las variables:

$$x = (x_1, x_2, x_3, \dots, x_n)$$

Y b es el vector de constantes b :

$$b = (b_1, b_2, b_3, \dots, b_m)$$

El problema básico de los sistemas de ecuaciones lineales ahora consistiría en encontrar el vector x que permite que la igualdad $A * x = b$ se cumpla.

2.2.4 Cómo calcular las soluciones de un sistema de ecuaciones lineales.

En este módulo nos enfocaremos solamente en el uso de las capacidades del lenguaje R para resolver sistemas de ecuaciones. Las estrategias básicas de solución pueden encontrarse en los primeros capítulos del texto de referencia.

Consideremos el sistema de ecuaciones (5):

$$\begin{aligned} 5x + y &= 3 \\ 2x - y &= 4 \end{aligned} \tag{5}$$

Recordemos el problema: deseamos encontrar los valores de x y y que permiten que se cumplan las igualdades del sistema de ecuaciones.

La representación matricial de los coeficientes es:

```
##      [,1] [,2]
## [1,]    5    1
## [2,]    2   -1
```

Esto es, un **renglón por ecuación** y una **columna para cada variable**.

El vector de constantes:

```
## [1] 3 4
```

Ahora podemos usar la función “solve” (parte de la librería “base”):

```
## [1] 1 -2
```

Entonces, el par (1,-2) corresponde a los valores de x y y que hacen que se cumpla la igualdad del sistema (6).

Veamos un sistema con tres incógnitas y dos ecuaciones:

$$\begin{aligned}4x_1 - x_2 + 3x_3 &= -1 \\3x_1 + x_2 + 9x_3 &= -4\end{aligned}$$

La “traducción” de este sistema al lenguaje R puede ser así:

```
##      [,1] [,2] [,3]
## [1,]    4  -1    3
## [2,]    3   1    9
```

```
## [1] -1 -4
```

y la solución se implementaría con la función “qr.solve” (también de la librería “base”):

```
## [1] -0.7142857 -1.8571429 0.0000000
```

La tercia ($x_1 = -0.714$, $x_2 = -1.857$, $x_3 = 0.000$) son los valores que permiten que se cumpla la igualdad.

2.2.5 Excepciones para considerar

No siempre es posible encontrar soluciones para todos los sistemas de ecuaciones.

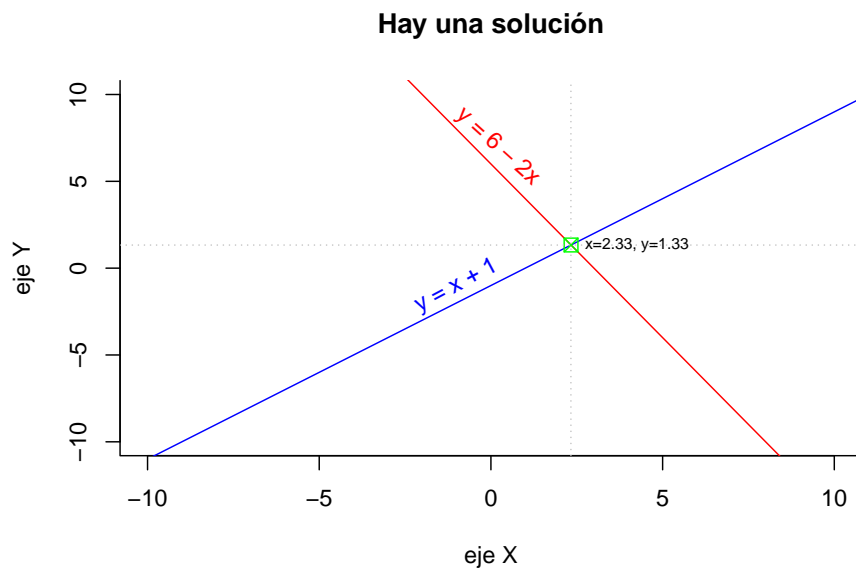
El siguiente es un sistema “bien portado”, con una única solución:

$$\begin{aligned}x - y &= 1 \\2x + y &= 6\end{aligned}$$

```
##      [,1] [,2]
## [1,]    1  -1
## [2,]    2   1
```

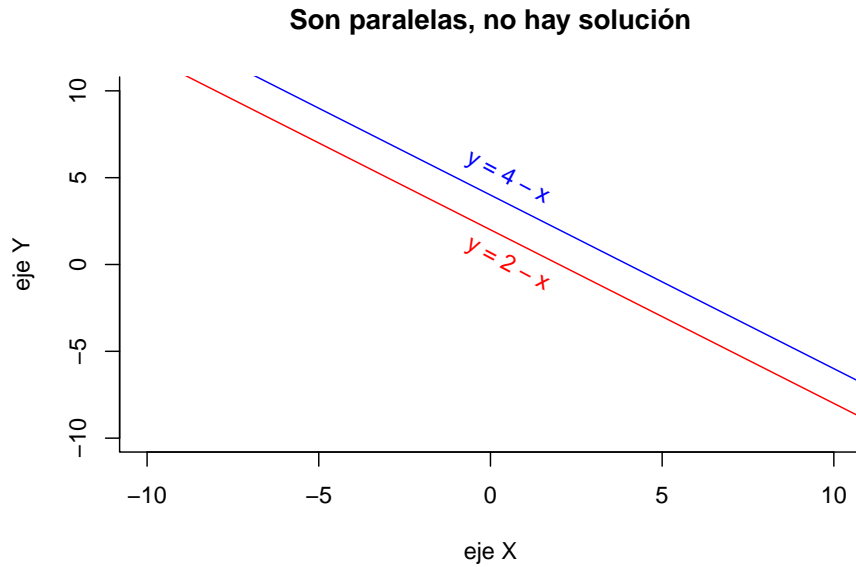
[1] 2.333333 1.333333

Este sistema de ecuaciones se puede representar en un plano cartesiano, en el que cada ecuación representa una línea recta. La solución del sistema corresponde a la intersección de dichas líneas:



En cambio, este es un sistema sin solución, porque son rectas *paralelas*, que *nunca se intersectan*:

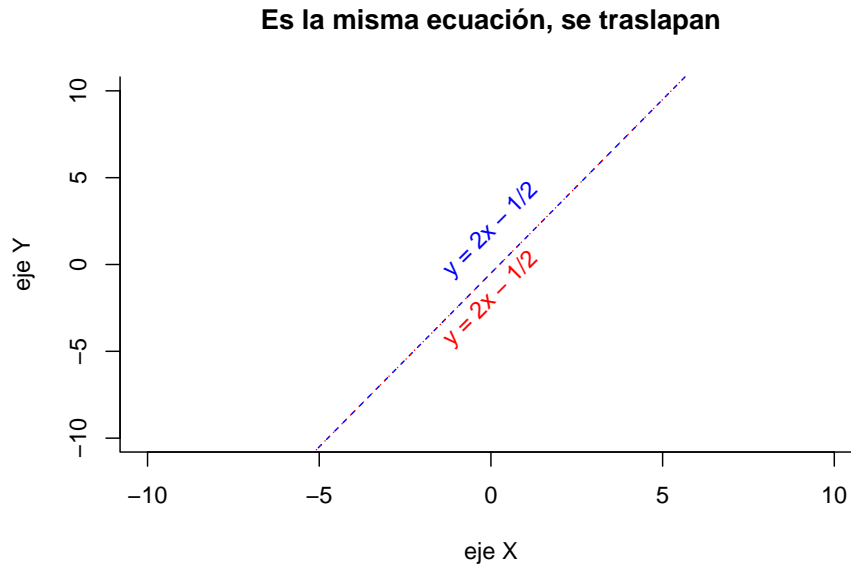
$$\begin{aligned}x + y &= 4 \\3x + 3y &= 6\end{aligned}$$



Si se usara la función “solve” aparecería un mensaje de error, indicando que el sistema es singular, y por tanto no se puede resolver.

Finalmente, en este sistema existe un número *infinito* de soluciones, porque se trata de la misma ecuación, se traslapan, por lo que hay un *número infinito* puntos que cumplen las igualdades del sistema:

$$\begin{aligned}4x - 2y &= 1 \\16x - 8y &= 4\end{aligned}$$



Aquí también la función “solve” indicaría que se trata de sistemas “singulares”, por lo que no tienen solución.

Para concluir, hay que comentar que la gran mayoría de los sistemas de ecuaciones “aplicados”, i.e. que representan sistemas naturales, económicos, o socioambientales no son “cuadrados” (que tienen el mismo número de ecuaciones y de variables), sino **rectangulares** (los números de ecuaciones y de variables no son iguales). Esta situación hace que la solución de estos sistemas se vuelva un reto matemático importante, que ha llevado al desarrollo de múltiples métodos y enfoques para resolver dicho tipo de problemas.

2.3 Aspectos básicos de los modelos de equilibrio parcial

2.3.1 Consideraciones generales

Los conceptos y métodos vistos en la sección previa son la base para el desarrollo de modelos “aplicados”. Para implementar un modelo con estas herramientas conviene considerar lo siguiente:

- 1) El conjunto de ecuaciones debería representar los principales procesos “**antagónicos**” presentes en el sistema.

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL¹⁷

- 2) Las variables (columnas) pueden representar “**atributos**” o “**propiedades**” del sistema cuyos roles o funciones dependen del “proceso o función” que se considere.
- 3) Los renglones (ecuaciones) deberían representar los “**procesos**” o “**actividades**” de interés.

Los modelos de equilibrio parcial (MEP) son sistemas de ecuaciones lineales, como los vistos líneas arriba. Esta herramienta se utiliza mucho en el campo de la Economía, la Economía agrícola, la Administración.

El nombre de “equilibrio parcial” proviene de las ciencias Económicas, y hace referencia al supuesto de que el resto de procesos económicos que influyen en el modelo (y no están incluidos en él) *no cambian*. Se utiliza para caracterizar sistemas en los que:

- 1) existe un proceso de interés que se desea *optimizar* (maximizar o minimizar) que se denomina *función objetivo*;
- 2) existe al menos otro proceso que *impone limitaciones* a la función objetivo. Se les conoce como *ecuaciones de restricción*.

2.3.2 Ejemplo simple de MEP, con representación gráfica

Los MEP son *sistemas de desigualdades*, es decir que las ecuaciones no tienen signo de igual, en su lugar hay \leq o \geq (desigualdades). El análisis gráfico de sistemas con dos variables permite comprender de manera cualitativa el funcionamiento de los componentes del MEP.

2.3.2.1 La función objetivo

Pongamos como ejemplo un paisaje en el que hay dos usos de suelo, agrícola y agroforestal con superficies x_1 y x_2 respectivamente (ejemplo modificado de Budnick 2007, cap. 10).

Por cada unidad de superficie de uso de suelo se pueden obtener ganancias de 5 y 6 unidades por unidad de superficie respectivamente. La ganancia total T obtenida del paisaje es:

$$T = 5x_1 + 6x_2$$

Esta es la *función objetivo*, la actividad o proceso que se desea hacer lo más grande o pequeño posible. Como esta función representa ganancias, entonces la

finalidad de este modelo es encontrar la combinación de superficies agrícola y agroforestal que maximice la ganancia.

Si consideramos que la máxima cantidad de superficie que se puede asignar a una actividad es 70, entonces estas serían las ganancias para cada combinación de superficies x_1 y x_2 . Las columnas muestran el aumento de x_1 y los de los renglones los de x_2 :

```
# para calcular la ganancia en
# distintas combinaciones de x1 y x2
xx<-as.matrix(rev(seq(10,70,10)))
fux<-function(xx){
  xx2<-seq(10,70,10)
  ttx<-5*xx2 + 6*xx
  ttx
}
qq<-apply(X=xx, MARGIN=1,FUN=fux)
qq<-t(qq)
qq
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] 470 520 570 620 670 720 770
## [2,] 410 460 510 560 610 660 710
## [3,] 350 400 450 500 550 600 650
## [4,] 290 340 390 440 490 540 590
## [5,] 230 280 330 380 430 480 530
## [6,] 170 220 270 320 370 420 470
## [7,] 110 160 210 260 310 360 410
```

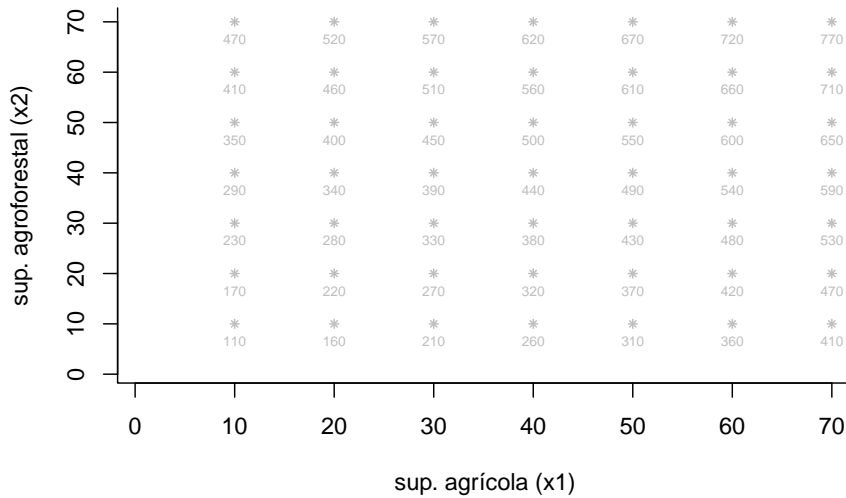
Y así se verían estos valores en un plano (x_1, x_2):

```
plot(1:70,1:70,type="n",bty="l",main=" ",
     xlab="sup. agrícola (x1) ",
     ylab="sup. agroforestal (x2)")

xx2<-rev(seq(10,70,10))
xx<-as.matrix(seq(10,70,10))

for (rx in 1:length(xx)) {
  #rx<-1
  zz<-length(xx)-(rx-1)
  points(y=rep(xx[rx],times=length(xx2)),x=xx,
         pch=8,col="gray",cex=0.6)
  text(y=rep(xx[rx],times=length(xx2)),x=xx,pos=1,
       labels=qq[zz,],col="gray",cex=0.6)
}
```

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL¹⁹



Podemos ver cómo las distintas combinaciones de $x_{\{1\}}$ y $x_{\{2\}}$ producen valores de ganancia diferentes. Pensando de manera simple, entre más superficies de x_1 y x_2 haya, mayor debería ser la ganancia. Por ello, la máxima ganancia se obtendría eligiendo las superficies máximas de x_1 y x_2 (la ganancia sería de 770). Sin embargo, **existen dos restricciones**, que impiden la elección de esta combinación. Por lo tanto, lo que se desea es *identificar la combinación de x_1 y x_2 que permitan la mayor ganancia, considerando las restricciones*.

2.3.2.2 Las funciones de restricción

El manejo de este paisaje está condicionado por dos aspectos (o restricciones). Uno es el **tiempo total invertido** en el manejo de cada tipo de uso de suelo. Para una unidad de superficie agrícola se requieren *tres* unidades de tiempo, mientras que para la superficie agroforestal se necesitan *dos* unidades de tiempo. Pero los usuarios del paisaje *unicamente pueden dedicar como máximo 120 unidades de tiempo*. Por tanto, el aspecto de manejo se puede representar mediante la desigualdad:

$$3x_1 + 2x_2 \leq 120 \quad (\text{er } 1)$$

Esta es la **primera ecuación de restricción**.

El otro aspecto es el **presupuesto** que se puede asignar a los usos de suelo para que sean productivos. En este presupuesto pueden ir salarios, fertilizantes,

maquinaria, etc. Como buen presupuesto, está restringido a un monto máximo. Para la superficie x_1 se pueden asignar *cuatro* unidades, mientras que para la superficie x_2 se asignan *seis* unidades. Esta segunda limitación tiene la siguiente representación algebraica:

$$4x_1 + 6x_2 \leq 260 \quad (\text{er } 2)$$

Y esta es la **segunda ecuación de restricción**.

El planteamiento del problema sería así:

Deseo saber cuánta superficie agrícola y agroforestal debería haber en el paisaje, para que la **ganancia sea lo más grande posible**, considerando **dos restricciones en el manejo**.

De manera sucinta este sistema se representa así:

Maximizar:

$$T = 5x_1 + 6x_2$$

sujeo a:

$$3x_1 + 2x_2 \leq 120$$

$$4x_1 + 6x_2 \leq 260$$

considerando que x_1 y x_2 sean mayores a cero.

2.3.2.3 Representación y análisis gráfico

El análisis de este sistema empieza con la representación gráfica de las ecuaciones de restricción.

La ecuación de la primera restricción “*tiene*” *dos partes*. **La primera es la igualdad** $3x_1 + 2x_2 = 120$, que puede representarse como:

$$x_1 = 40 - (2/3)x_2$$

Esta última línea muestra las combinaciones de superficies agrícola y agroforestal (al conjunto de pares (x_1, x_2)) que *requieren el máximo tiempo disponible (120 horas)*.

La segunda parte corresponde a la desigualdad $3x_1 + 2x_2 < 120$. Esta línea *divide el plano en dos secciones que tienen significados diferentes*. Si, p.ej. elegimos el punto $(x_1 = 20, x_2 = 20)$ y lo sustituimos en $3x_1 + 2x_2 \leq 120$, queda que $60 + 40 \leq 120$, lo cual es *cierto*. Por tanto, **todos los puntos que están abajo de la línea corresponden a combinaciones de x_1 y x_2 que cumplen la restricción**.

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL21

Pero si elegimos el par $(x_1 = 50, x_2 = 50)$, la sustitución produce $150 + 50 = < 120$, lo que es *falso*. Por ello es que **todos los puntos por arriba de la línea son combinaciones de x_1 y x_2 que NO cumplen la restricción.**

La segunda restricción $4x_1 + 6x_2 = < 260$ puede analizarse de la misma manera. Si la reescribimos en términos de x_1 queda que:

$$x_1 = 65 - 1.5x_2$$

Esta es la línea que corresponde las combinaciones de superficies agrícola y agroforestal que *cumplen con el valor máximo de la restricción.*

Nuevamente, si elegimos el punto $(x_1 = 20, x_2 = 20)$ y lo sustituimos en $4x_1 + 6x_2 = < 260$, queda que $80 + 120 = < 260$, lo cual es cierto. Por tanto, todos los puntos que están *abajo* de la línea también corresponden a **combinaciones de x_1 y x_2 que cumplen con la restricción.**

Pero si elegimos el par $(x_1 = 50, x_2 = 50)$, la sustitución produce $200 + 300 = < 260$, lo que es *falso*. Por ello es que todos los puntos **arriba de la línea nuevamente son combinaciones de x_1 y x_2 que NO cumplen con la restricción.**

En el panel izquierdo se muestra la representación gráfica del análisis de la primera restricción, mientras que en el panel derecho está la segunda restricción.

```
par(mfrow=c(1,2))

#####
## primera restricción

plot(1:70,1:70,type="n",bty="l",
     main="restricción 3x1 +2x2 =<120 ",
     xlab="sup. agrícola (x1)",ylab="sup. agroforestal (x2)")
# restricción 1
#lines(x=c(0,65),y=c(43.3,0),col="blue")
abline(a=43.3,b=-2/3,col="blue")
text(x=49,y=17,labels="restricción 1", col="blue",srt=-39)

# zona de soluciones viables
points(20,20,pch=8,col="green")
text(20,20,labels="zona factible",pos=1,col="green")

# zona de soluciones INviables
points(50,50,pch=8,col="red")
text(50,50,labels="zona inviable",pos=1,col="red")

#####
## segunda restricción

plot(1:70,1:70,type="n",bty="l",
```

```

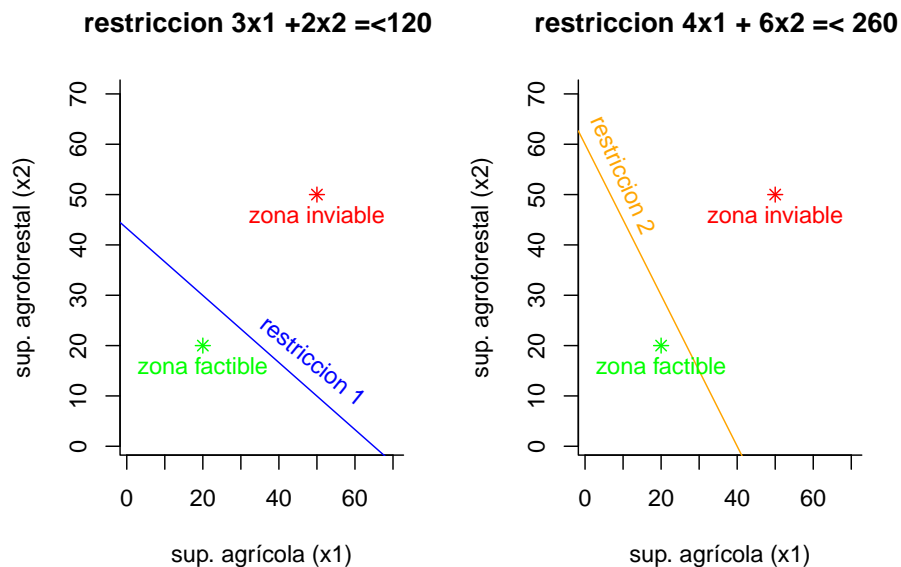
main="restriccion 4x1 + 6x2 =< 260 ",
xlab="sup. agricola (x1)",
ylab="sup. agroforestal (x2)")

# restriccion 2
#lines(x=c(0,40),y=c(60,0),col=" orange")
abline(a=60,b=-3/2,col="orange")
text(x=10,y=54,labels="restriccion 2", col="orange",srt=-66)

# zona de soluciones viables
points(20,20,pch=8,col="green")
text(20,20,labels="zona factible",pos=1,col="green")

# zona de soluciones INviables
points(50,50,pch=8,col="red")
text(50,50,labels="zona inviable",pos=1,col="red")

```



Si juntamos las dos lineas en un mismo plano tendremos el análisis completo. Ahora podemos distinguir cuatro regiones:

- * a) La región en la que se cumplen ambas restricciones;
- * b) la región en la que no se cumplen ninguna;
- * c) las de cumplimiento parcial (una restriccion se cumple y la otra no);
- * d) La solución del sistema, que es la intersección de las dos líneas.

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL23

```
plot(1:70,1:70,type="n",bty="l",
     xlab="sup. agrícola (x1)",
     ylab="sup. agroforestal (x2)")

# restricción 1
#lines(x=c(0,65),y=c(43.3,0),col="blue")
abline(a=43.3,b=-2/3,col="blue")
text(x=45,y=17,labels="restriccion 1", col="blue",srt=-24)

# restricción 2
#lines(x=c(0,40),y=c(60,0),col="orange")
abline(a=60,b=-3/2,col="orange")
text(x=6,y=54,labels="restriccion 2", col="orange",srt=-42)

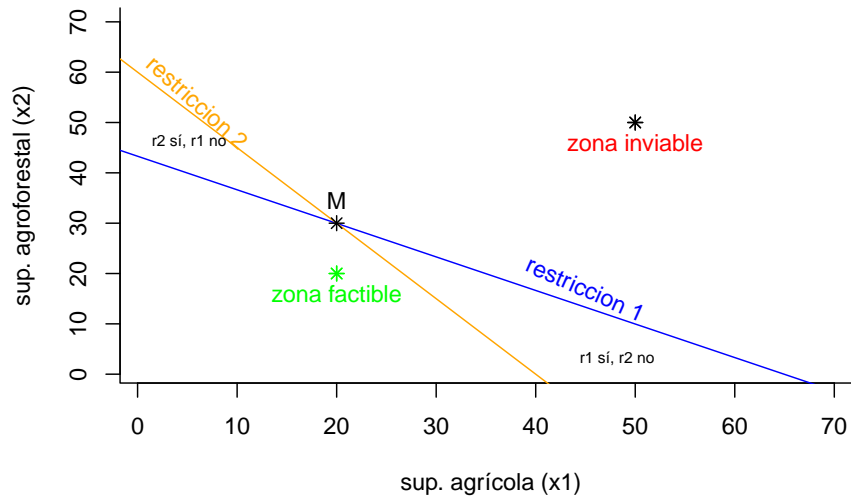
# zona de soluciones viables
points(20,20,pch=8,col="green")
text(20,20,labels="zona factible",pos=1,col="green")

# zona de soluciones INViables
points(50,50,pch=8,col="black")
text(50,50,labels="zona inviable",pos=1,col="red")

# zona de soluciones SEMI viables
points(20,20,pch=8,col="green")
text(0,46,labels="r2 sí, r1 no",pos=4,col="black",cex=0.7)

points(50,50,pch=8,col="black")
text(43,3,labels="r1 sí, r2 no",pos=4,col="black",cex=0.7)

### intersección
points(20,30,pch=8)
text(20,30,labels="M",pos=3)
```



La intersección M corresponde a la **combinación de superficies x_1 y x_2 que hacen que la ganancia sea máxima y cumple con las dos restricciones**. Si juntamos este análisis gráfico con la gráfica en la que se ven las ganancias como función de x_1 y x_2 podemos ver cuál es el valor de ganancia que corresponde a la intersección y los correspondientes valores de x_1 y x_2 :

```
#####
## calcula ganancias en
# distintas combinaciones de x1 y x2
xx<-as.matrix(rev(seq(10,70,10)))
fux<-function(xx){
  xx2<-seq(10,70,10)
  ttx<-5*xx2 + 6*xx
  ttx
}
qq<-apply(X=xx, MARGIN=1,FUN=fux)
qq<-t(qq)

#####
## dibuja plano x1 vx x2 y zonas
plot(1:70,1:70,type="n",bty="l",
     main=" ",
     xlab="sup. agrícola (x1)",ylab="sup. agroforestal (x2)")
```


2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBIO PARCIAL²⁵

```
# restricción 1
#lines(x=c(0,65),y=c(43.3,0),col="blue")
abline(a=43.3,b=-2/3,col="blue")
text(x=45,y=17,labels="restriccion 1", col="blue",srt=-24)

# restricción 2
#lines(x=c(0,40),y=c(60,0),col=" orange")
abline(a=60,b=-3/2,col="orange")
text(x=6,y=54,labels="restriccion 2", col="orange",srt=-42)

# zona de soluciones viables
points(20,20,pch=8,col="green")
text(20,20,labels="zona factible",pos=1,col="green")

# zona de soluciones INviables
points(50,50,pch=8,col="black")
text(50,50,labels="zona inviable",pos=1,col="red")

# zona de soluciones SEMI viables
points(20,20,pch=8,col="green")
text(0,46,labels="r2 sí, r1 no",pos=4,col="black",cex=0.7)

points(50,50,pch=8,col="black")
text(43,3,labels="r1 sí, r2 no",pos=4,col="black",cex=0.7)

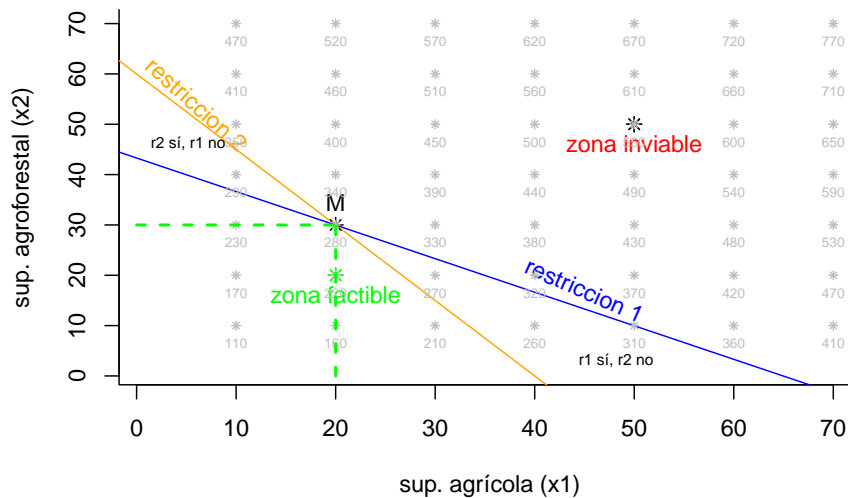
### intersección
points(20,30,pch=8)
text(20,30,labels="M",pos=3)

#####
## combinaciones de x1 y x2 y su ganancia
xx2<-rev(seq(10,70,10))
xx<-as.matrix(seq(10,70,10))

for (rx in 1:length(xx)) {
  #rx<-1
  zz<-length(xx)-(rx-1)
  points(y=rep(xx[rx],times=length(xx2)),x=xx,
         pch=8,col="gray",cex=0.6)
  text(y=rep(xx[rx],times=length(xx2)),x=xx,pos=1,
       labels=qq[zz,],col="gray",cex=0.6)
}

#####
```

```
## valores de x1 y x2
lines(x=c(20,20),y=c(0,30),lty="dashed",col="green",lwd=2)
lines(x=c(0,20),y=c(30,30),lty="dashed",col="green",lwd=2)
```



La combinación de $x_1 = 20$ y $x_2 = 30$ produce una ganancia de 280, que es la *máxima posible* considerando las restricciones.

2.3.3 Análisis de modelos de equilibrio parcial con software

Es poco frecuente encontrar sistemas tan simples, con dos dimensiones (x_1, x_2) y dos ecuaciones de restricción. Lo más esperado es trabajar con sistemas de muchas dimensiones y con varias restricciones. Por eso el uso de software especializado es indispensable.

La solución de este sistema usando la librería “lpSolve” del R es:

```
## Success: the objective function is 280
```

```
## [1] 20 30
```

Esta función tiene dos resultados principales.

El primero es el valor de la función objetivo, i.e. el valor máximo de ganancia que se puede obtener de este sistema, considerando las restricciones. Esta ganancia

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL 27

máxima es $T_{max} = 280$.

El segundo resultado es la combinación de valores de x_1 y x_2 que producen la ganancia máxima. Estos valores son $x_1 = 20$ y $x_2 = 30$.

2.3.3.1 Análisis de sensibilidad de los modelos de equilibrio parcial

Una vez hecho el análisis de optimización, un aspecto relevante es identificar *qué tanto cambiarían los resultados al cambiar atributos del modelo*. Si, por ejemplo, las mediciones fueran realizadas con otros criterios de precisión ¿cambiarían las soluciones? ¿Cuál parte del modelo es la que, al modificarse, cambiaría más los resultados?

Estas preguntas forman parte del llamado **análisis de sensibilidad**. Veremos dos casos básicos: a) el precio sombra; b) el análisis de sensibilidad propiamente dicho; .

- Se pueden revisar las sensibilidades de los coeficientes de la función objetivo;
- Se pueden revisar las sensibilidades de los coeficientes de las ecuaciones de restricción.

Sin entrar en muchos detalles se puede decir que los análisis de sensibilidad están basados en derivadas parciales, en las que se trata de vincular las tasas de cambio de un parámetro con respecto a las tasas de cambio de las soluciones del sistema.

2.3.3.1.1 El “precio sombra” Esta es la magnitud en la que el valor óptimo de la función objetivo cambiaría si el valor limitante de una restricción aumenta en una unidad.

Recordemos el ejemplo 1, que consiste en maximizar:

Maximizar:

$$T = 5x_1 + 6x_2$$

sujeto a:

$$3x_1 + 2x_2 \leq 120$$

$$4x_1 + 6x_2 \leq 260$$

El análisis indicó que el óptimo para la función objetivo es 280. En este caso, *el precio sombra indica en cuanto cambiaría el 280 si las restricciones (120 o 260) cambiaran en una unidad*:

```
## Success: the objective function is 280
```

```
## [1] 20 30
```

```
## [1] 0.6 0.8 0.0 0.0
```

Los precios sombra de las restricciones son 0.6 y 0.8. Esto indica tres cosas:

El cambio de *una* unidad en el *tiempo invertido* en el sistema (120) produciría un cambio de 0.6 en la ganancia máxima del sistema (280)

El cambio de *una* unidad en el *presupuesto disponible* (260) produciría un cambio de 0.8 en la ganancia máxima del sistema (280).

Por esto el valor óptimo de la función objetivo es más sensible a la restricción 2 ($4x_1 + 6x_2 \leq 260$), que es la del presupuesto.

Los precios sombra de las variables son cero.

Los intervalos inferior y superior indican el rango de variación en el que el análisis es válido:

```
## [1] 8.666667e+01 1.600000e+02 -1.000000e+30 -1.000000e+30
```

```
## [1] 1.95e+02 3.60e+02 1.00e+30 1.00e+30
```

2.3.3.1.2 Análisis de sensibilidad Este análisis pretende saber qué tanto pueden cambiar los resultados del modelo si se alteran los coeficientes de las ecuaciones. Particularmente, el análisis de sensibilidad desea *saber si un cambio en un coeficiente afecta mucho o no al valor óptimo de la ecuación objetivo*.

Este análisis es relevante, porque en la estimación de los coeficientes de las ecuaciones siempre hay errores de medición, o procesos o considerados, o supuestos que no son válidos en todos los casos. La sensibilidad permite identificar aquellas partes del modelo que pueden alterar mucho a los resultados.

En este contexto, al valor óptimo se le conoce como **base**. Un coeficiente **insensible** es el que al cambiar afecta muy poco a la base. Pero si su cambio ocasiona un gran variación en la base, entonces se le denomina **sensible**.

2.3.3.1.3 Sensibilidad de la función objetivo Este análisis identifica qué tanto puede cambiar un coeficiente de la función objetivo sin que el valor óptimo varíe mucho.

```
## Success: the objective function is 280
```

2.3. ASPECTOS BÁSICOS DE LOS MODELOS DE EQUILIBRIO PARCIAL29

Acá está la implementación del análisis de sensibilidad:

```
## [1] 4.000000 3.333333
```

```
## [1] 9.0 7.5
```

```
##   lim. Inf coef. original lim.Sup
## 1 4.000000           5      9.0
## 2 3.333333           6      7.5
```

El rango del primer coeficiente (5) va de 4 a 9, y representa **el rango de variación del coeficiente en el que la base no cambia significativamente**. Si (por algún motivo) el coeficiente quedara fuera de estos rangos, afectaría mucho a la base.

El rango del segundo coeficiente (6) va de 3.3. a 7.5 y es de **menor amplitud que el primero**. Por ello es que el *segundo coeficiente es más sensible que el primero*.

2.3.3.1.4 Sensibilidad de las funciones de restricción El mismo razonamiento puede aplicarse para los coeficientes de las ecuaciones de restricción:

```
## Success: the objective function is 280
```

Y esta es la sensibilidad de las restricciones:

```
##   lim. Inf. vals. orig. lim. Sup
## 1 86.66667         120      195
## 2 160.00000        260      360
```

Estos son los *rangos de variación de los valores de restricción en los que no hay un efecto isgnificativo en el óptimo de la función objetivo (base)* .

2.3.3.1.5 Ejercicio 1 MEP

- Construya un esbozo de MEP con su sistema de estudio (o uno de su interés);
- En lo posible, hágalo tan simple como el primer ejemplo (una ecuación objetivo, dos ecuaciones de restricción, y dos variables);

- Lo fundamental es “practicar la traducción”: representar un proceso o modelo verbal-conceptual en términos numéricos;
- Enfatizar en la utilidad o propósitos del modelo: qué deseo contestar, para qué me sirve, qué espero entender con el modelo;
- Enfatizar en los supuestos y limitaciones del modelo;
- Primero hacer el ejercicio de manera individual (10 min);
- Luego comparar los ejercicios;
- LLenar un cuadro con tres columnas: sistema de interés, proceso para optimizar, procesos restrictivos.

2.3.3.2 Cambio de la función objetivo

Tal vez ya sea evidente que el principal atributo del modelo (la función objetivo) sea quizás su limitación conceptual más notoria.

Esta herramienta está concebida para optimizar un único proceso, aunque los sistemas reales no siempre “cumplan” este supuesto.

En una gran variedad de sistemas ocurre (o se necesita) **la optimización multiobjetivo (MMO), que hace referencia a estrategias de modelación en las que se optimizan simultáneamente dos o más procesos.**

La modelación multiobjetivo (MMO) es un enfoque que conceptualmente es más afín a la condición de los sistemas naturales y sociales, en los que existen dos o más procesos que se optimizan simultáneamente.

Una forma “rápida” de implementar la modelación multiobjetivo con los modelos de equilibrio parcial es mediante el **cambio de la función objetivo**. Consiste simplemente en el intercambio de roles entre funciones: una función de restricción se vuelve objetivo y una función objetivo se vuelve de restricción.

2.3.3.2.1 Primer cambio de objetivo. Reconsideremos el primer ejercicio e intercambiemos la función objetivo inicial (la ganancia obtenida del sistema, $T = 5x_1 + 6x_2$) con la primera restricción (el tiempo dedicado al sistema, $3x_1 + 2x_2 \leq 120$).

El primer cambio del modelo tal vez sea “conceptual”. Ahora se trata de encontrar las mayores cantidades de tiempo que se pueden dedicar al sistema, considerando las restricciones de ganancia e inversión.

La implementación no es muy complicada:

```
## Success: the objective function is 168
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTIOBJETIVO (MMO)31

```
## [1] 56 0
```

Este ejercicio muestra que la ganancia máxima (168) es menor a la original y sugiere que toda la superficie se dedique a agricultura.

2.3.3.2.2 Segundo cambio de objetivo Ahora intercambiamos la función objetivo inicial (la ganancia obtenida del sistema, $T = 5x_1 + 6x_2$) con la segunda restricción (presupuesto que se puede asignar a los usos de suelo, $4x_1 + 6x_2 \leq 260$).

Este modelo modificado pretende encontrar las mayores inversiones que se pueden realizar en el sistema, considerando las restricciones de ganancia y tiempo.

Este sería el análisis usando la segunda restricción:

```
## Success: the objective function is 280
```

```
## [1] 0.00000 46.66667
```

Ahora, la ganancia se mantiene igual (280), pero toda la superficie debería dedicarse a la agroforestería.

2.4 Generalidades de la modelación multiobjetivo (MMO)

Al implementar los MEP surge una inquietud básica: los problemas/procesos son más complicados, porque se necesita encontrar soluciones (i.e. optimizar) que **involucren varios procesos simultáneamente**.

Los MEP incluyen la modelación simultánea de varios procesos, pero sólo uno de ellos es de interés, los demás procesos son importantes sólo porque que limitan o restringen al proceso de interés.

En este nuevo escenario, ahora todos los procesos son “igualmente importantes”, por lo que se necesita optimizar (maximizar o minimizar) a **todos los procesos al mismo tiempo**.

Esta actividad se conoce como Modelación Multi Objetivo (MMO).

2.4.1 Soluciones de compromiso, disyuntivas o trade-offs

En el caso de la modelación con MEP hay dos resultados principales. Uno es el *valor óptimo* (máximo o mínimo) que puede alcanzar la función objetivo. El otro son los *valores de las variables* que permiten a la función objetivo alcanzar su valor óptimo, considerando a las funciones de restricción.

En el caso de la modelación MMO, el resultado principal es un **conjunto de soluciones que definen las mejores *disyuntivas* (tradeoffs, soluciones de compromiso) entre los procesos de interés.**

La disyuntiva entre dos procesos A y B significa que mientras A aumenta B disminuye y viceversa.

2.4.2 Dominancia

Para encontrar una solución que sea mejor que otra en un problema con un solo objetivo únicamente hay que comparar las soluciones e identificar la que sea mejor.

Este proceso se complica un poco más con la MMO, por lo que se requiere de un concepto nuevo, *la dominancia*.

Una solución x_1 domina a la solución x_2 si:

1. x_1 no es peor que x_2 en todos los objetivos;
2. x_1 es estrictamente mejor que x_2 en al menos un objetivo.

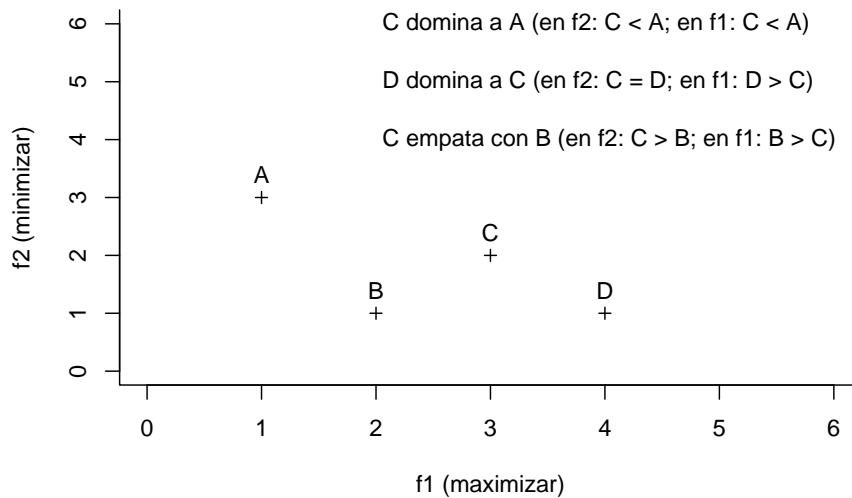
Veamos esto en la siguiente figura:

```
xx<-c(1,2,3,4)
yy<-c(3,1,2,1)
tt<-c("A","B","C","D")
plot(xx,yy,xlim=c(0,6),ylim=c(0,6),bty="l",
      pch=3,cex=0.8,
      xlab="f1 (maximizar)",ylab="f2 (minimizar)")
text(xx,yy,labels=tt,pos=3)

#relaciones de dominancia
CdA<-" C domina a A (en f2: C < A; en f1: C < A)"
DdC<-" D domina a C (en f2: C = D; en f1: D > C)"
CeB<-" C empata con B (en f2: C > B; en f1: B > C)"

text(2,6,labels=CdA,adj=0)
text(2,5,labels=DdC,adj=0)
text(2,4,labels=CeB,adj=0)
```


2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO)33



En la figura podemos ver algunos ejemplos de dominancia entre cuatro posibles soluciones. en este caso, una solución es una *combinación* de valores de $f1$ y $f2$. La solución C domina a la solución A porque en el eje $f2$ C “minimiza más” que A , mientras que en el eje $f1$ C “maximiza más” que A . Pero la solución C no es mejor que la solución B : C es mejor en el eje $f2$, pero B es mejor en el eje $f1$. La solución D dominaría a todas, ya que es la que mejor “maximiza” en $f1$ y la que mejor “minimiza” en $f2$.

Veamos con más detalle este concepto. En la gráfica siguiente se supondrá que las dos funciones $f1$ y $f2$ deben maximizarse.

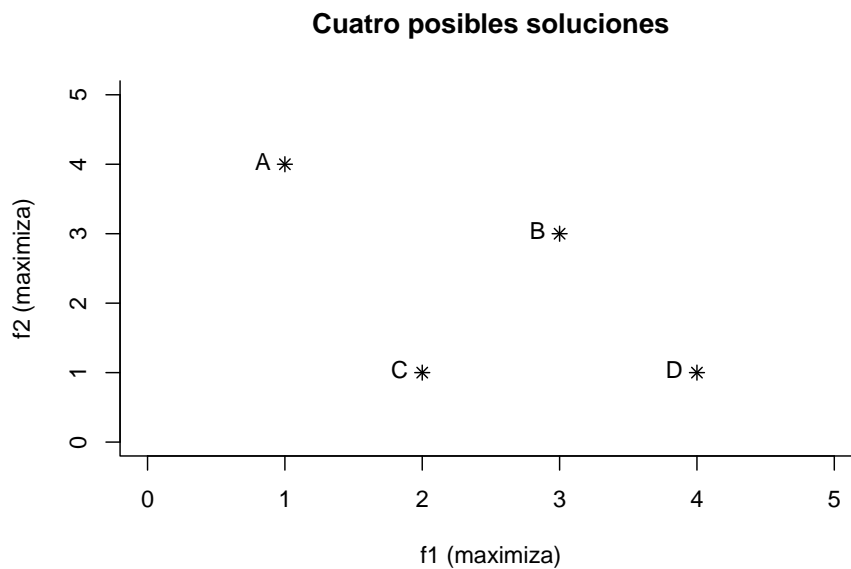
Para empezar, establezcamos las coordenadas de las soluciones (cuatro posibles en este ejemplo). Primero hacemos una pequeña base de datos:

```
x<-c(1,2,3,4)
y<-c(4,1,3,1)
nn<-c("A","C","B","D")
df<-cbind(x,y)
rownames(df)<-nn
df
```

```
##  x y
## A 1 4
## C 2 1
## B 3 3
## D 4 1
```

Y luego la graficamos:

```
plot(df,ylim=c(0,5),xlim=c(0,5),type="n",bty="l",
     main="Cuatro posibles soluciones",
     xlab="f1 (maximiza)",ylab="f2 (maximiza)")
points(df,pch=8)
text(df,labels=nn,pos=2)
```



Si comparamos la solución B , encontramos que ella domina unicamente a la solución C , (B es mayor que C en los dos ejes).

La función “dominated” de la librería {caRamel} permite averiguar si una solución (combinación de valores de $f1$ y $f2$) domina o no a las demás soluciones. Este análisis se implementa así.

```
## Warning: package 'caRamel' was built under R version 4.3.2
```

```
## Warning: package 'geometry' was built under R version 4.3.2
```

```
## x y
```

```
## 3 3
```

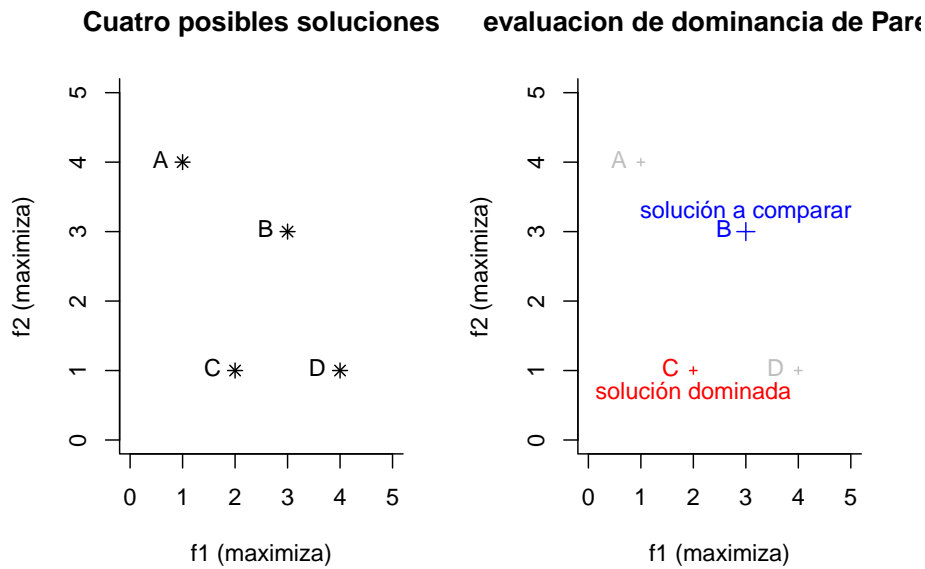
```
##      A      C      B      D
```

```
## FALSE TRUE FALSE FALSE
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO)35

```
##      A      C      B      D
## TRUE FALSE  TRUE  TRUE
```

El resultado de la función “dominated” (*FALSE, TRUE, FALSE, FALSE*) muestra un valor *TRUE* en la solución *C*, lo que indica que la solución *B* domina a *C* nada más. Para verlo en colores:



2.4.3 ¿Pareto?

Los problemas de optimización de muchos objetivos se empezaron a estudiar a finales del siglo XIX. El economista-sociólogo Vilfredo Pareto (1848~1923) es quizá el principal antecedente en este campo de la modelación matemática. Los conceptos “frente de Pareto”, “soluciones de Pareto”, “índice de Pareto,” “distribución de Pareto” y otros, se nombran en su memoria.

2.4.4 Conjunto de soluciones no dominadas de Pareto

Por el hecho de que en la modelación multiobjetivo se busca la optimización simultánea de varios procesos, no hay una solución, sino *un conjunto* de posibles soluciones que cumplen los requerimientos del sistema de interés.

Un conjunto de soluciones no dominadas de Pareto es un conjunto de soluciones que no están dominadas por ningún miembro del conjunto.

El conjunto óptimo de Pareto es el subconjunto no dominado del conjunto posible de soluciones.

Cuando se grafican las posibles soluciones en un plano f_1 vs. f_2 se puede distinguir el *Frente Óptimo de Pareto*, que es el borde que separa al conjunto óptimo de soluciones de las otras soluciones.

La MMO tiene dos propósitos:

- 1) Encontrar el conjunto de soluciones más cercano al frente óptimo de Pareto;
- 2) Que ese conjunto de soluciones sea lo más diverso posible.

2.4.5 Técnicas para la MMO

Hay muchas técnicas para encontrar las soluciones de un problema multiobjetivo:

- Suma ponderadas (weighted sums)
- Restricción epsilon (-constraint)
- métricas ponderadas (weighted metrics)
- programación evolutiva y algoritmos genéticos

En la actualidad, el enfoque de programación evolutiva se ha desarrollado mucho. Los ejercicios de este módulo se implementaron con una variante de esta técnica.

2.4.5.1 Algoritmos genéticos, o la naturaleza como inspiración

Los algoritmos de optimización computacionales enfrentan algunos problemas de gran complejidad teórica y práctica:

- Encontrar soluciones entre un enorme número de posibilidades;
- Analizar procesos de manera simultánea;
- Elegir las secuencias que se van a revisar;
- Adaptarse, i.e. seguir funcionando bien aunque cambie el entorno.

A mediados de los años 60 del siglo XX, John Henry Holland (físico, ingeniero, psicólogo y computólogo!) y sus colegas iniciaron los estudios que dieron origen a los algoritmos genéticos. La fuente básica de inspiración de Holland fue la evolución biológica. Este proceso, visto desde la perspectiva computacional, resuelve los problemas antes mencionados. En consecuencia, los sistemas biológicos (seres vivos) pueden ser considerados como propuestas de solución a un problema fundamental: dejar descendencia viable. Otro aspecto interesante es

2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO)³⁷

que la enorme diversidad biológica es resultado de un único proceso (la evolución) que ha generado múltiples soluciones particulares (i.e. las poblaciones biológicas).

El proceso evolutivo, en su representación más simple está compuesto de:

1. Una población de soluciones candidatas para resolver un problema (inicialmente se denominan cromosomas);
2. Una función de adaptación (criterio o regla que permite identificar y comparar la calidad de las soluciones candidatas);
3. La selección de las mejores soluciones de acuerdo a los criterios de la función de adaptación.
4. La recombinación de las soluciones elegidas (combinar partes de una solución con partes de otra, para producir soluciones nuevas)
5. La mutación (variación aleatoria de las soluciones nuevas).

La búsqueda de soluciones consiste en:

1. Crear una población inicial de soluciones $N(0)$;
2. Evaluar los desempeños de cada elemento de la población usando la función de adaptación;
3. Quedarse con las mejores soluciones y descartar las demás;
4. Recombinar las soluciones ganadoras para producir una nueva población de descendientes $f(0)$, que sean una combinación de las soluciones con mejor desempeño;
5. Variar aleatoriamente a los miembros de $f(0)$;
6. La población resultante se convierte en $N(1)$ y el ciclo de evaluación comienza de nuevo.

A partir de este algoritmo simple se han desarrollado muchos métodos de búsqueda de soluciones en sistemas multiobjetivo.

2.4.6 Ejemplo 1 recalculado con MMO

Recordemos el sistema del ejemplo 1 analizado con MEP:

```
## Success: the objective function is 280
```

```
## [1] 20 30
```

2.4.6.1 Análisis MMO

Al igual que con MEP, hay muchos softwares para hacer estas cuentas. En esta implementación con la librería “caRamel” existe la posibilidad de indicar qué hacer con cada función (maximizarla o minimizarla).

Como esta rutina de optimización está basado en algoritmos genéticos, se pueden modificar el tamaño de la población de soluciones y las tasas de mutación.

Aspectos para notar en esta función:

- Se debe indicar *qué se desea hacer* con cada función (maximizar o minimizar, por medio del argumento “minmax”);
- Hay que indicar *rangos de variación* de las variables, para que los algoritmos de búsqueda no se pierdan (para esto es el argumento “bounds”);
- Se puede elegir la *cantidad de soluciones* para conservar, con el argumento “popsize”.

```
## Beginning of caRamel optimization <-- Wed Feb 28 09:59:04 2024
```

```
## Number of variables : 2
```

```
## Number of functions : 3
```

```
## |
```

```
## Done in 0.0083010196685791 secs --> Wed Feb 28 09:59:04 2024
```

```
## Size of the Pareto front : 20
```

```
## Number of calls : 20
```

2.4.6.1.1 Resultados del MMO: La frontera de Pareto En lugar de encontrar una sola combinación de valores de las variables (como en los MEP), ahora aparecerán *conjuntos de soluciones: combinaciones de valores de las variables que tienen mejor o peor desempeño*. En este ejemplo se obtienen 20 soluciones (argumento popsize=20).

2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO) 39

2.4.6.1.2 Valores de las variables y de las funciones Un resultado de este análisis es el conjunto de valores de variables para cada conjunto de soluciones:

```
##           [,1]      [,2]
## [1,] 48.79463 60.53239
## [2,] 24.88826 62.87782
## [3,] 58.47874 39.25066
```

La columna 1 corresponde a la variable x_1 (superficie agrícola) y la segunda a x_2 (superficie agroforestal). Cada combinación de valores de variables produce un valor para cada una de las funciones objetivos:

```
##           [,1]      [,2]      [,3]
## [1,] 607.1675 267.4487 558.3729
## [2,] 501.7082 200.4204 476.8199
## [3,] 527.8977 253.9375 469.4189
```

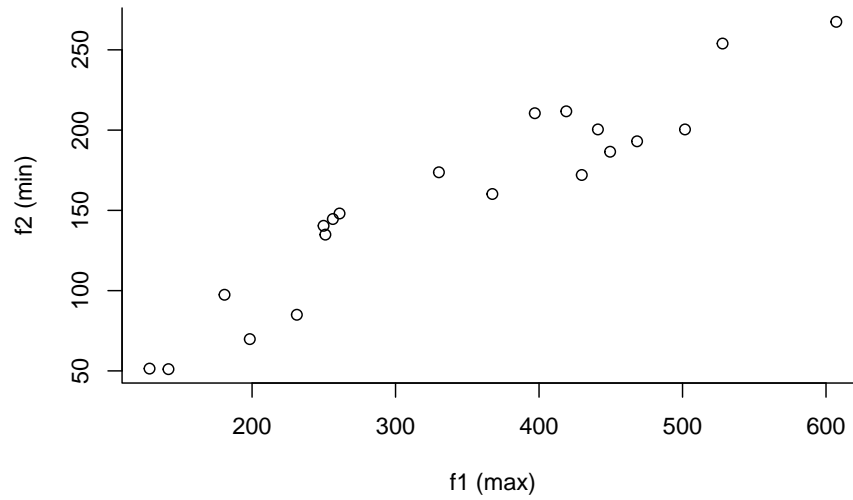
Sólo por ocioso, si reemplazamos $x_1 = 48.7946306$ y $x_2 = 60.5323931$ (primer renglón de valores) en $f_1 = 5x_1 + 6x_2$ (la primera función objetivo), vemos que $f_1 = 607.1675118$:

```
## [1] 607.1675
```

En el MEP se graficaban los *valores de las variables* (x_1 y x_2) para encontrar el valor óptimo de la función.

En la MMO se grafican los **valores de las funciones** (f_1, f_2, f_3), para **detectar disyuntivas o antagonismos (e.g. tradeoffs) entre funciones**. En este ejemplo, el análisis se hace mediante la graficación de pares de columnas (f_1 vs. f_2 , f_2 vs. f_3 , f_1 vs. f_3).

Veamos un par de ellas (f_1 vs. f_2), la primera función se debe maximizar mientras que la segunda se debe minimizar:

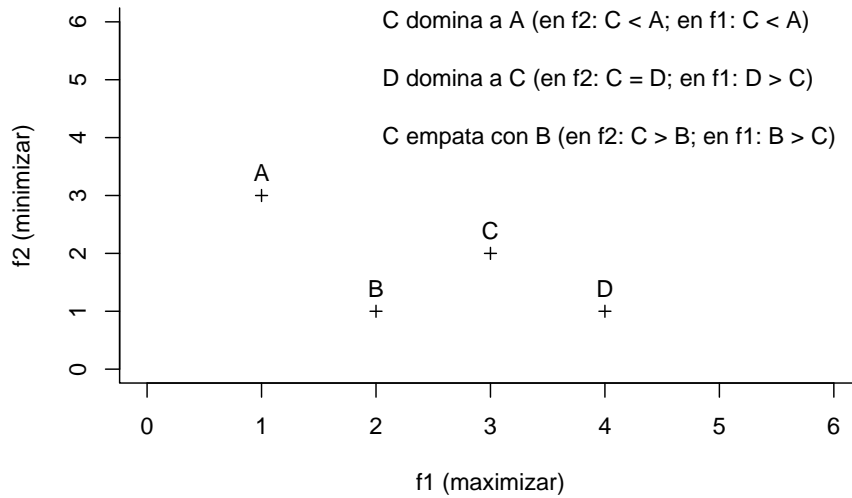


A cada punto en este plano (combinación de valores de f_1 y f_2) se le puede llamar *solución*. Recordemos que una solución A domina a la solución B si:

- A no es peor que B en todos los objetivos;
- A es estrictamente mejor que B en al menos un objetivo.

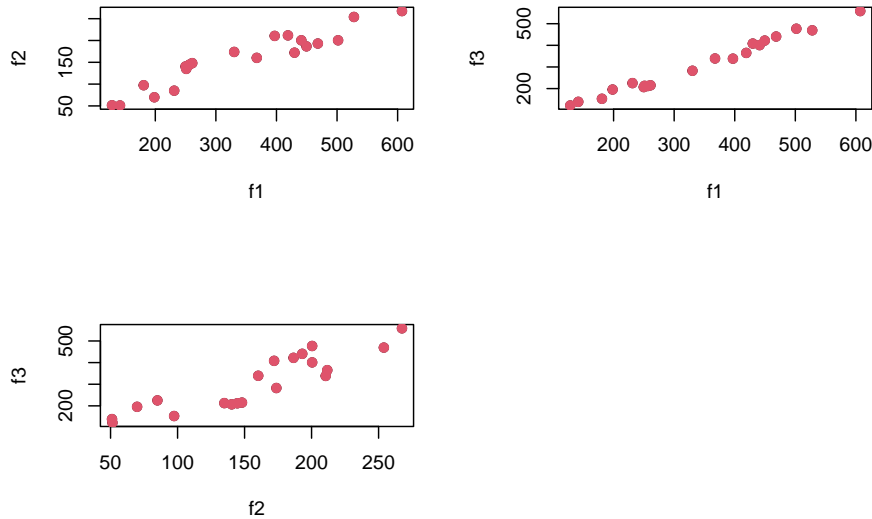
Veamos nuevamente la figura de dominancia de la sección anterior:

2.4. GENERALIDADES DE LA MODELACIÓN MULTIOBJETIVO (MMO)41



Un conjunto de soluciones no dominadas de Pareto es un conjunto de soluciones que no están dominadas por ningún miembro del conjunto. **El conjunto óptimo de Pareto es el subconjunto no dominado del conjunto posible de soluciones.** Se trata de combinaciones de valores de las funciones que, al menos, son igualmente buenas. En cualquier otra combinación (solución) no se cumplirán las condiciones de dominancia, por lo que se trata de soluciones “subóptimas”.

Con la función “plot_pareto” es posible identificar los conjuntos de Pareto para cada par de funciones. Notar que uno de los argumentos hace referencia a si la función se debía maximizar o no (argumento “maximized”):



Los puntos rojos muestran soluciones de Pareto. Cada una de ellas corresponde a una combinación de valores de las funciones con un desempeño (maximizan o minimizan) por lo menos equivalente.

Los patrones de puntos de estos planos coordenados pueden darnos información acerca de cómo interactúan las funciones entre sí. Por ejemplo, si la nube de puntos fuera *descendente* entre dos funciones, entonces podemos pensar en *influencias antagónicas* entre esas dos funciones.

En este ejemplo con pocas soluciones ($n=20$) podemos ver que no hay disyuntivas entre las funciones objetivo, ya que las nubes de puntos sugieren correlaciones directas entre ellas.

Otra posible situación sería aquella con pocos puntos “rojos” (soluciones de Pareto) y que estuvieran localizados en una pequeña región del plano. Eso indicaría que la zona de variación de las soluciones es restringida (como en el ejemplo numérico).

2.4.6.1.3 Ejercicio

- Cambiar el “destino” de las funciones objetivo (minimizar o maximizar) y comparar los frentes de Pareto;
- Cambiar de signo las funciones objetivo y comparar los frentes de Pareto.

2.4.7 Ejemplo de MMO: el caso de la región de Marqués de Comillas (Chiapas, México).

El ejemplo de esta sección se basa en el trabajo de Wies et al. (2023). Imaginemos un paisaje con cuatro coberturas/ usos de suelo:

- agrícola (A);
- cultivos forestales (F);
- pastoreo (P);
- áreas de conservación (C)

Recordemos que es el modelador quien decide qué procesos o aspectos incluir y decidir cuáles de ellos se maximizan y cuáles se minimizan. La elección de ellos se realiza primeramente en función de los *intereses del modelador* y de la *información disponible* para calcular los coeficientes de las ecuaciones.

Acá hay un cuadro con cuatro procesos que se desean representar y el “tratamiento” que se les desea aplicar (maximizar o minimizar).

proceso o atributo	“tratamiento”
(G_T) Ganancia de la producción en todo el paisaje	maximizar
(Q_T) Uso de agroquímicos	minimizar
(W_T) Trabajo invertido	minimizar
(B_T) Biodiversidad	maximizar

2.4.7.1 Definición de las ecuaciones.

Una vez identificadas los procesos o atributos que se desean optimizar, hay que elaborar las ecuaciones lineales correspondientes. Esto significa que hay que calcular los coeficientes de cada ecuación. Los datos para calcularlos se estandarizarán para poder interpretar los valores de los coeficientes de manera directa y simple.

2.4.7.1.1 Ganancia total (G_T) (gross margin) de la producción en todas los usos de suelo. Este es un atributo que no puede ignorarse a la hora de pensar en sistemas agroalimentarios y usos de paisaje.

De manera general, esta ecuación se construye calculando primero cuál es la ganancia que se obtiene por cada unidad de superficie de cada tipo de cobertura o uso del suelo ($g_{cobertura}$). A continuación se multiplican dichos valores por las superficies correspondientes de cada cobertura ($a_{cobertura}$) y se representan

como sumandos en la ecuación que calcula la ganancia total (G_T) que se genera al usar el paisaje. La forma general de la función de ganancia sería algo como esto:

$$G_T = a_A g_A + a_F g_F + a_P g_P + a_C g_C \quad (\text{Ganancia})$$

Hay aspectos que se deben considerar, como el intervalo de tiempo en el que se generó la ganancia, la unidad monetaria que se utiliza, entre otros. En la medida que se consideren el modelo será mejor.

2.4.7.1.2 Uso de agroquímicos (Q_T). La estrategia de construcción de esta ecuación es similar a la anterior. Hay que estimar cuántos agroquímicos se usan en cada tipo de uso de suelo ($q_{cobertura}$) y multiplicarlos por las coberturas correspondientes. Su expresión general sería:

$$Q_T = q_A g_A + q_F g_F + q_P g_P + q_C g_C \quad (\text{AgroQuímicos})$$

En este caso el uso de agroquímicos puede medirse en unidades monetarias, en peso, o en volumen. También es importante considerar el intervalo de tiempo para el que se calculan las cantidades de uso.

2.4.7.1.3 Cantidad de trabajo invertido (W_T). Las consideraciones básicas de este aspecto son las ya descritas previamente. Aquí la forma general:

$$W_T = w_A g_A + w_F g_F + w_P g_P + w_C g_C \quad (\text{Trabajo})$$

En este caso, el trabajo invertido (w) puede medirse como número de personas, o unidades de tiempo (horas, días, ect.) por persona, o número total de tiempo invertido.

2.4.7.1.4 Biodiversidad en el paisaje (B_T). La unidades de medición pueden ser riqueza de especies, o algún descriptor como el índice de Shannon.

$$B_T = b_A g_A + b_F g_F + b_P g_P + b_C g_C \quad (\text{Biodiv})$$

2.4.7.2 Por fin, el modelo completo.

Finalmente, el modelo completo se ve así:

$$\begin{aligned}
 G_T &= 5018g_A + 1538g_F + 6216g_P + 100g_C && \text{(modelo Completo)} \\
 Q_T &= 1.109g_A + 0g_F + 0.6P + 0g_C \\
 W_T &= 2.04g_A + 0g_F + 0.722g_P + 0g_C \\
 B_T &= 0g_A + 0g_F + 0g_P + 4.38g_C
 \end{aligned}$$

Cada coeficiente muestra la “importancia relativa” de una unidad de superficie de cada uso de suelo en cada proceso (renglón) que se pretende optimizar. Los coeficientes de valor cero (0) indican que esa cobertura no tiene aporte al proceso.

Y acá la implementación en la función “caRamel”:

- La ganancia y la conservación de la biodiversidad deberían *maximizarse*, mientras que el uso de agroquímicos y el trabajominvertido deberían *minimizarse*, por lo que el argumento “minmax” recibe el vector (TRUE,FALSE,FALSE,TRUE);
- Los *rangos de variación* de las coberturas de uso de suelo se obtienen de los valores máximos y mínimos encontrados en el sistema;
- Se eligieron 1000 soluciones (argumento “popsize”=1000).

```

# Caramel optimization
library(caRamel)

fpaisaje <- function(i) {
  GT<-5018*x[i,1]+1537.6*x[i,2]+6216*x[i,3]+100*x[i,4]
  AgT<-1.109*x[i,1]+0*x[i,2]+0.6*x[i,3]+0*x[i,4]
  WkT<-2.04*x[i,1]+0*x[i,2]+0.722*x[i,3]+0*x[i,4]
  BdT<-0*x[i,1]+0*x[i,2]+0*x[i,3]+4.38*x[i,4]
  return(c(GT,AgT,WkT,BdT))
}

# Número de funciones objetivo
nobj <- 4
# Número de variables
nvar <- 4

```

```

# número de posibles soluciones
nsol<-1000

# minimizar (FALSE)
#maximizar (TRUE)
# un valor por cada función objetivo
minmax <- c(TRUE,FALSE,FALSE,TRUE)

# ncol siempre es 2
# si las variables son superficies de territorio,
# los minimos pueden ser ceros, y los maximos
# son el total de area considerada, pej.
minsup<-c(0.25,3,2.5,1.1)
maxsup<-c(8,136,40,51)

#Uso      Minimo  Maximo
#Maiz    0.25    8
#Ganaderia 3    136
#Cultivo forestal (Palma) 2.5 40
#Bosque en conservacion 1.1 51

bounds <- matrix(data = 1, nrow = nvar, ncol = 2)
bounds[, 1] <- minsup * bounds[, 1]
bounds[, 2] <- maxsup * bounds[, 2]

resum <- caRamel(nobj = nobj,
                nvar = nvar,
                minmax = minmax,
                bounds = bounds,
                func = fpaisaje,
                popsize = nsol,
                archsize = nsol,
                maxrun = 400,
                prec = matrix(1.e-3, nrow = 1, ncol = nobj),
                carallel = FALSE,graph=FALSE,
                sensitivity=TRUE)

```

```
## |
```

Todos los resultados principales están juntos en la “sección” “total_pop” del objeto “resum”.

```
## [1] 1000 8
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTIOBJETIVO (MMO)47

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,] 1.2628401 125.1369 13.365400 3.051928 282131.9 9.419730 12.22601 13.36744
## [2,] 0.8959703 127.1357 31.716136 4.151449 397542.4 20.023312 24.72683 18.18335
## [3,] 1.6378339 133.6679 9.337478 5.031072 272291.3 7.418845 10.08284 22.03609
## [4,] 0.6283007 125.0060 29.718582 5.367748 380629.5 18.527935 22.73855 23.51074
## [5,] 0.8557017 114.7746 14.076124 7.000907 268968.7 9.394647 11.90859 30.66397
## [6,] 2.2900837 128.2957 17.946707 8.866565 321202.5 13.307727 17.62929 38.83556
```

Este objeto tiene 1000 filas, que corresponden al número de soluciones que se indicaron en la función. Las primeras columnas de este objeto corresponden a las *coberturas*, y las siguientes son los *coeficientes de las funciones objetivo*.

Veamos la primera fila:

```
## [1]      1.263      125.137      13.365      3.052 282131.880      9.420      12.226
## [8]      13.367
```

Ella muestra el siguiente sistema de ecuaciones:

$$2.8213188 \times 10^5 = (1.263 \times 5018) + (125.137 \times 1538) + (13.365 \times 6216) + (3.052 \times 100)$$

$$9.42 = (1.263 \times 1.109) + (125.137 \times 0) + (13.365 \times 0.6) + (3.052 \times 0)$$

$$12.226 = (1.263 \times 2.04) + (125.137 \times 0) + (13.365 \times 0.722) + (3.052 \times 0)$$

$$13.367 = (1.263 \times 0) + (125.137 \times 0) + (13.365 \times 4.38) + (3.052 \times 0)$$

Los primeros cuatro elementos (del 1 al 4) corresponden a los valores de superficie de cada tipo de cobertura (se repiten en cada ecuación). Los siguientes elementos (del 5 a 8) corresponden a los lados izquierdos de cada ecuación.

Esta es una solución de Pareto. Representa una manera de “organizar” el paisaje, con una combinación de coberturas que optimiza las ecuaciones de acuerdo a los criterios establecidos en el modelo: maximizar la ganancia y la conservación de la biodiversidad, y minimizar el uso de agroquímicos y el trabajo (TRUE, FALSE, FALSE, TRUE).

2.4.7.3 El frente de Pareto, o ¿cuántas soluciones hay?

De este total de soluciones (1000), *sólo una fracción son soluciones de Pareto*. Ellas están en la sección “parameters” del objeto “resum”:

```
## [1] 81 4
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.2628401 125.1369 13.365400 3.051928
## [2,] 0.8959703 127.1357 31.716136 4.151449
## [3,] 1.6378339 133.6679  9.337478 5.031072
## [4,] 0.6283007 125.0060 29.718582 5.367748
## [5,] 0.8557017 114.7746 14.076124 7.000907
## [6,] 2.2900837 128.2957 17.946707 8.866565
```

Si comparamos estos valores con las primeras cuatro columnas de la sección “total_pop” veremos que son iguales:

```
## [1] 1000      8

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,] 1.2628401 125.1369 13.365400 3.051928 282131.9  9.419730 12.22601 13.36744
## [2,] 0.8959703 127.1357 31.716136 4.151449 397542.4 20.023312 24.72683 18.18335
## [3,] 1.6378339 133.6679  9.337478 5.031072 272291.3  7.418845 10.08284 22.03609
## [4,] 0.6283007 125.0060 29.718582 5.367748 380629.5 18.527935 22.73855 23.51074
## [5,] 0.8557017 114.7746 14.076124 7.000907 268968.7  9.394647 11.90859 30.66397
## [6,] 2.2900837 128.2957 17.946707 8.866565 321202.5 13.307727 17.62929 38.83556
```

De ellos, **los primeros 81 corresponden a las soluciones de Pareto**. El resto (919) NO son soluciones de Pareto.

Recordemos que en este ejercicio se pidió al algoritmo que calcule 1000 soluciones para el sistema (con el argumento “archsize”). Sin embargo, el análisis encontró 81 soluciones de Pareto.

2.4.7.4 Coberturas de uso de suelo

Acá podemos ver las primeras combinaciones de coberturas que cumplen con los criterios de optimización de Pareto.

```
colnames(resum$parameters)<-c("Agric","Forestal","Ganaderia",
                             "Conservacion")
head(resum$parameters, n=3)
```

```
##           Agric Forestal Ganaderia Conservacion
## [1,] 1.2628401 125.1369 13.365400      3.051928
## [2,] 0.8959703 127.1357 31.716136      4.151449
## [3,] 1.6378339 133.6679  9.337478      5.031072
```

Recordemos que cada línea es **una combinación de superficies que puede optimizar el uso del paisaje**.

2.4.7.5 Funciones objetivo

Y acá vemos los valores de cada ecuación del sistema para cada combinación de coberturas.

```
# VALORES DE LAS FUNCIONES OBJETIVO
# Los tres primeros valores de cada función

colnames(resum$objectives)<-c("GT", "QT", "WT", "BT")
head(resum$objectives, n=3)
```

```
##           GT           QT           WT           BT
## [1,] 282131.9  9.419730 12.22601 13.36744
## [2,] 397542.4 20.023312 24.72683 18.18335
## [3,] 272291.3  7.418845 10.08284 22.03609
```

La primera línea de coberturas da como resultado la primera línea de valores de la función objetivo.

2.4.7.6 Sinergias y antagonismos en el frente de Pareto.

Ahora es posible buscar antagonismos o sinergias entre cada proceso modelado.

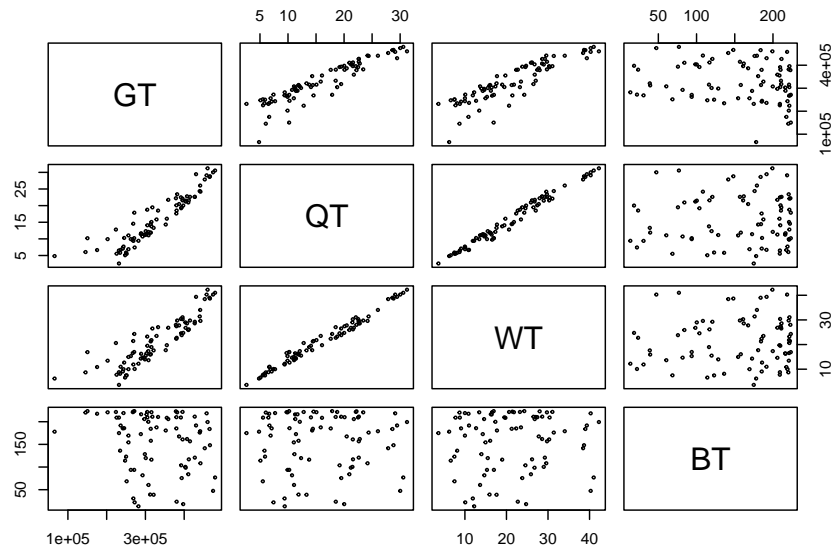
Esto se logra graficando las combinaciones de todos los procesos.

Las interacciones entre G_T , Q_T y W_T son sinérgicas, pues las gráficas sugieren correlaciones positivas entre ellas.

Las nubes de puntos con G_T en el eje X muestran una especie de “quiebre” (i.e cambio de pendiente) a la mitad del eje. Esto sugiere que hay dos grandes conjuntos de soluciones cuando se considera este proceso.

La biodiversidad (B_T) no parece estar asociada a alguno de los otros procesos, pues no se distinguen patrones en las nubes de puntos que la incluyen.

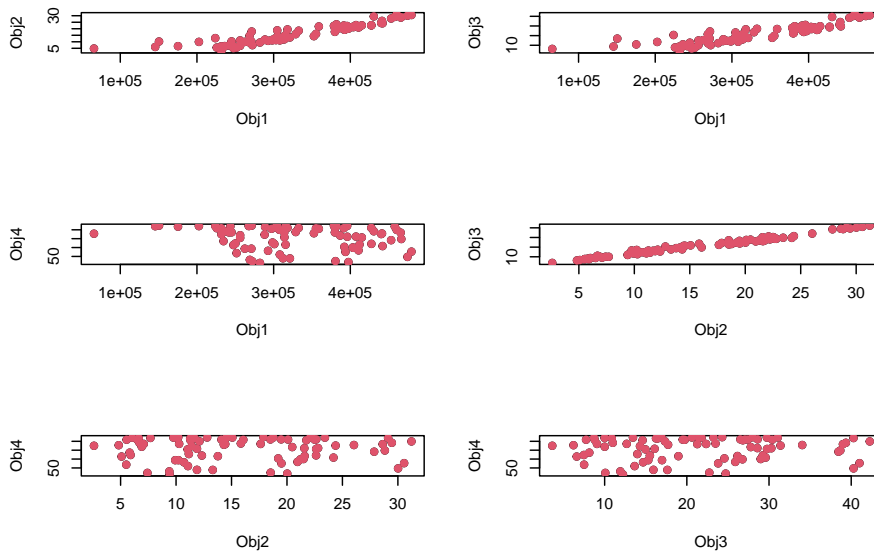
```
pairs(resum$objectives, cex=0.4)
```



Acá el mismo resultado, pero con las funciones diseñadas para ello.

```
# objetivos del frente de pareto
# valores de cada función del modelo
 #(matrix, nrow = archsize, ncol = nobj+nadditional)
plot_pareto(resum$objectives,maximized=minmax)
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTIOBJETIVO (MMO) 51



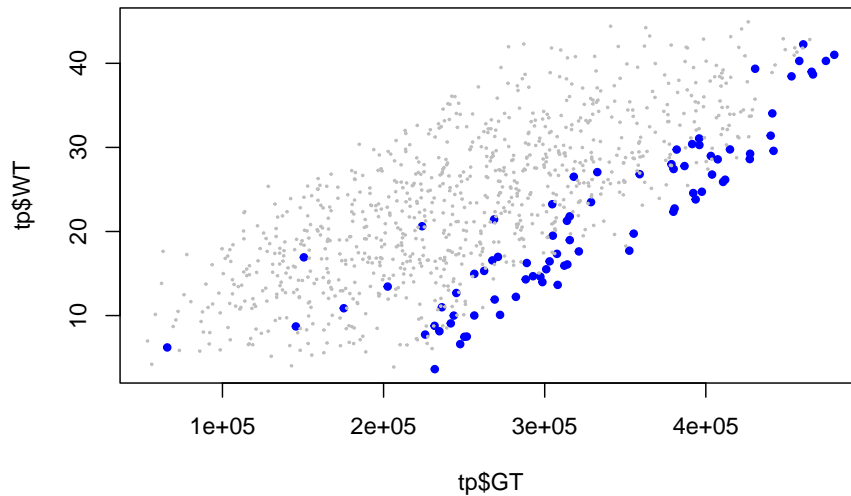
Para tener una representación más clara del frente de Pareto, grafiquemos todas las soluciones de un par de objetivos y resaltemos las soluciones de Pareto.

```
# nombres a las columnas
tp<-as.data.frame(resum$total_pop)
colnames(tp)<-c("Agric", "Past", "Fores", "Consr", "GT",
               "QT", "WT", "BT")

dd<-dim(resum$objectives)

# para distinguir quien es Pareto
# y quien no
cx<-ifelse((1:nsol)<dd[1], "blue", "gray")
tx<-ifelse((1:nsol)<dd[1], 1, 0.3)

plot(x=tp$GT,
     y=tp$WT, pch=20, col=cx,
     cex=tx)
```



Las soluciones de Pareto están agrupadas a lo largo de un solo lado del conjunto de soluciones, por lo que *parece que forman un “frente”*.

2.4.7.7 Caracterización de las soluciones de Pareto

El resultado principal de este modelo son los frentes de Pareto. Sin embargo, estos resultados pueden analizarse de manera complementaria con un PCA. Este análisis permite identificar **si hay grupos de soluciones más similares y las influencias de cada función objetivo sobre las soluciones.**

```
# PCA de las soluciones
```

```
library(MASS)
```

```
colnames(resum$objectives)<-c("GT", "QT", "WT", "BT")
```

```
pca0<-princomp(x=resum$objectives, cor=TRUE)
```

```
head(pca0$scores)
```

```
##          Comp.1  Comp.2  Comp.3  Comp.4
## [1,] -1.2729115 2.345365 -0.16707652 0.009565011
## [2,]  1.0725776 2.206776 -0.11962514 -0.075893239
## [3,] -1.6287298 2.238067 -0.04533936 0.053701034
## [4,]  0.7212575 2.133667 -0.10232915 -0.080881535
## [5,] -1.3884626 2.056006 -0.21891625 -0.028778310
## [6,] -0.3960863 1.906642 -0.14597772 0.025662100
```

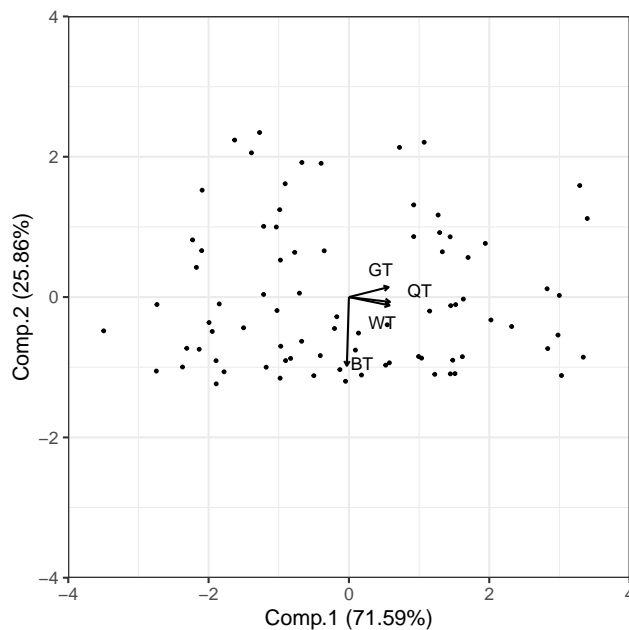
2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO) 53

El biplot del PCA confirma lo que se observó en los gráficos de Pareto, que hay una correlación positiva muy fuerte entre G_T , Q_T y W_T , mientras que B_T se comporta de una manera “casi independiente” de las otras.

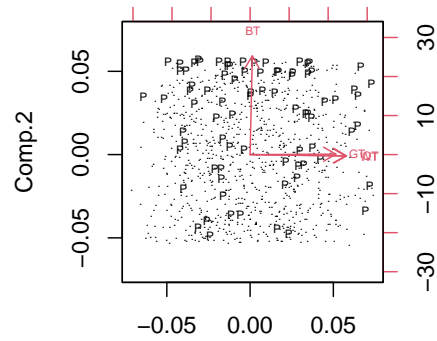
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
library(ggord)  
ggord(ord_in=pca0,size=0.5,  
      xlims=c(-4,4),ylim=c(-4,4),  
      arrow=0.1,repel=TRUE,txt=3)
```



2.4.7.7.0.1 **Cómo se ve el frente de Pareto en el PCA de todas las soluciones?** En este caso el frente de Pareto está “disperso” en la nube de puntos multivariada.

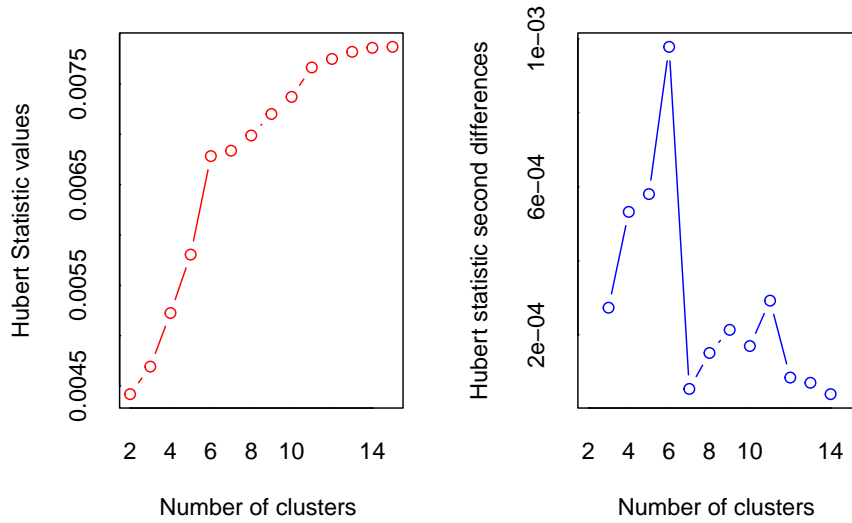


2.4.7.7.0.2 Métodos para identificar grupos de soluciones en la nube del PCA Una vez calculados los scores del PCA, es posible analizar la nube del PCA para identificar grupos (función “NbClust”, librería {NbClust}).

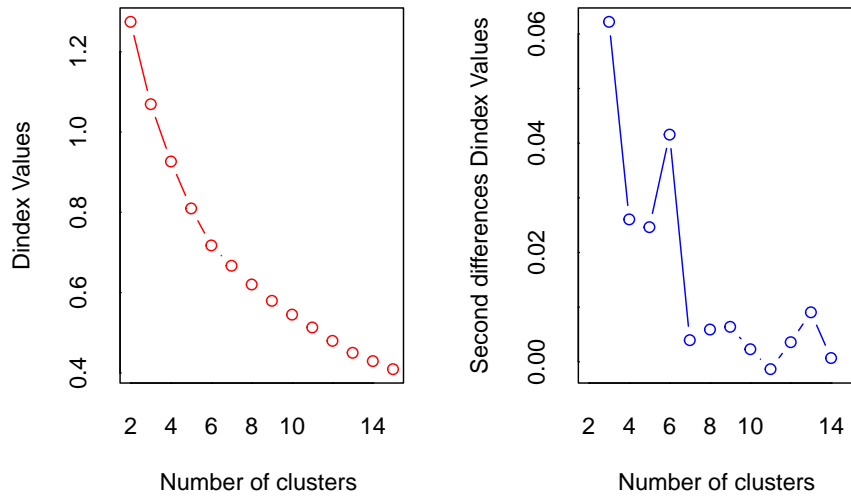
```
# Identifica cuantos grupos de soluciones hay
library(NbClust)

#resum$objectives
Ng<-NbClust(data=pca0$scores,method="ward.D")
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTIOBJETIVO (MMO)55



```
## *** : The Hubert index is a graphical method of determining the number of clusters.  
##           In the plot of Hubert index, we seek a significant knee that corresponds to a  
##           significant increase of the value of the measure i.e the significant peak in H  
##           index second differences plot.  
##
```



```
## *** : The D index is a graphical method of determining the number of clusters.
##           In the plot of D index, we seek a significant knee (the significant
##           second differences plot) that corresponds to a significant increase
##           the measure.
##
```

```
## *****
```

```
## * Among all indices:
```

```
## * 7 proposed 2 as the best number of clusters
## * 4 proposed 3 as the best number of clusters
## * 1 proposed 5 as the best number of clusters
## * 4 proposed 6 as the best number of clusters
## * 1 proposed 8 as the best number of clusters
## * 2 proposed 13 as the best number of clusters
## * 4 proposed 15 as the best number of clusters
```

```
##
```

```
##           ***** Conclusion *****
```

```
##
```

```
## * According to the majority rule, the best number of clusters is 2
```

```
##
```

```
##
```

```
## *****
```

```
# Aquí se identifica a cual grupo
# pertenece cada solución
```


2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO) 57

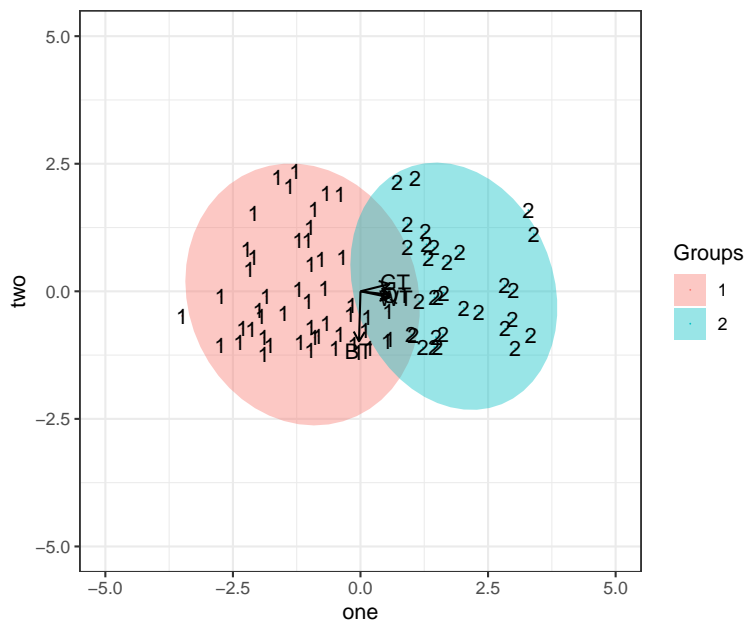
```
pg<-Ng$Best.partition
```

Ahora podemos ver con más claridad los grupos en la nube del pca.

```
# separo los primeros dos ejes  
dfc<-pca0$scores[,1:2]  
dfc$grupos<-Ng$Best.partition
```

```
## Warning in dfc$grupos <- Ng$Best.partition: Realizando coercion de LHD a una  
## lista
```

```
bp<-ggord(ord_in=pca0,size=0.1,  
         grp_in=factor(Ng$Best.partition),arrow=0.2)+  
  ylim(-5,5)+xlim(-5,5)+  
  geom_text(label=Ng$Best.partition)  
bp
```



2.4.7.7.1 Diferencias significativas entre grupos? Finalmente, es posible averiguar si esos grupos son diferentes (función “anosim”, librería {vegan}).

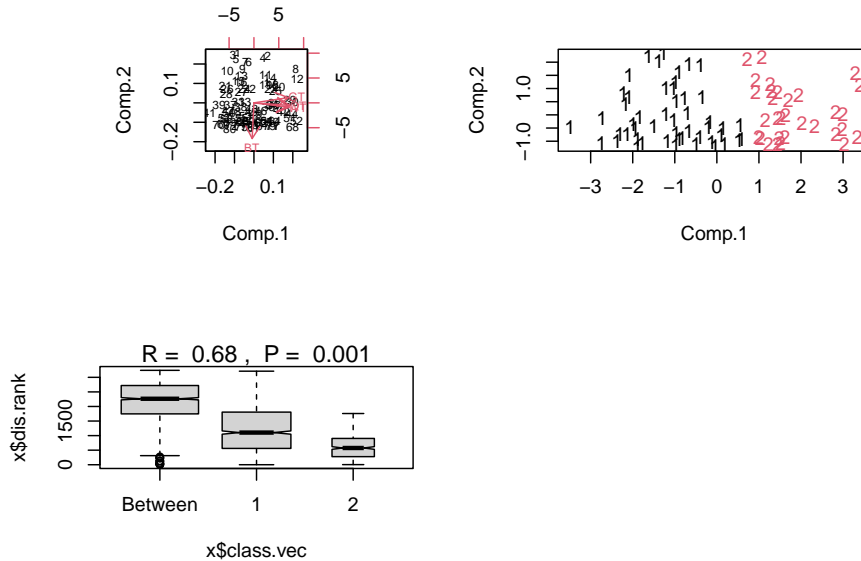
```
# Averigua si los grupos son diferentes  
# Se utiliza la matriz de  
# valores de las funciones objetivo  
library(vegan)  
dif0<-anosim(x=resum$objectives,grouping=pg)  
dif0
```

```
##  
## Call:  
## anosim(x = resum$objectives, grouping = pg)  
## Dissimilarity: bray  
##  
## ANOSIM statistic R: 0.6797  
##      Significance: 0.001  
##  
## Permutation: free  
## Number of permutations: 999
```

Veamos todo el ejercicio junto:

```
# dibuja todo junto  
par(mfrow=c(2,2))  
biplot(pca0,cex=0.6)  
plot(pca0$scores[,1:2],type="n")  
text(pca0$scores[,1:2],labels=pg,col=pg)  
plot(dif0)
```

2.4. GENERALIDADES DE LA MODELACIÓN MULTI OBJETIVO (MMO) 59



2.4.7.7.1.1 Son diferentes en las coberturas? Si usamos la partición de grupos del Nbclust podemos explorar si los grupos tienen distintos valores de cobertura.

Recordemos que al especificar el modelo, *se declararon rangos de las superficies de cada uso de suelo*. En este ejemplo cada cobertura puede variar entre 0.25, 3, 2.5, 1.1 y 8, 136, 40, 51 unidades de superficie.

Podemos ver que las soluciones sugieren paisajes con proporciones muy diferentes de cada cobertura.

Además, todos los valores de superficie son menores al umbral superior (8, 136, 40, 51), lo que puede interpretarse como si hubiera “superficie libre”, ya sea para otras actividades o para explorar otros escenarios de uso.

```
ddg<-as.data.frame(resum$parameters)
ddg$grupos<-Ng$Best.partition

## detalles de cada grupo
promSUP<-aggregate(ddg[,1:(ncol(ddg)-1)],
                    by=list(ddg$grupos),mean)
promSUP
```

```
## Group.1 Agric Forestal Ganaderia Conservacion
## 1 1 2.746907 111.9262 13.04897 37.05326
## 2 2 3.569745 124.2618 33.06063 33.59103
```

2.4.7.7.1.2 Qué tanto difieren los grupos de soluciones en cada proceso? Tan solo observando la columna de ganancia (G_T) podemos ver que cada grupos de estrategias ofrece ganancias diferentes, lo que sugiere que cada grupo tiene “intensidades de funcionamiento” contrastantes.

```
ddg<-as.data.frame(resum$objectives)
ddg$grupos<-Ng$Best.partition

## detalles de cada grupo
promPROC<-aggregate(ddg[,1:(ncol(ddg)-1)],
                    by=list(ddg$grupos),mean)
promPROC
```

```
##   Group.1      GT      QT      WT      BT
## 1      1 270699.4 10.87570 15.02504 162.2933
## 2      2 417841.9 23.79522 31.15205 147.1287
```

2.4.7.8 Ejercicio

- Modifique los coeficientes de las funciones objetivo y evalúe comparativamente los resultados de las simulaciones.

2.5 Referencias

Anton, H., & Rorres, C. 2014. Elementary Linear Algebra. Wiley.

Budnick, F.S. 2007. Matemáticas aplicadas para administración, economía y ciencias sociales. McGrawHill, www.FreeLibros.me.

Hazel, P., Norton, D. 1986. Mathematical programming for economic analysis in agriculture. McMillan.

Jeroen C.J. Groot, Walter A.H. Rossing, Andrew Jellema, Derk Jan Stobbe-laar, Henk Renting, Martin K. Van Ittersum. 2007. Exploring multi-scale trade-offs between nature conservation, agricultural profits and landscape quality- A methodology to support discussions on land-use perspectives. Agriculture, Ecosystems and Environment, 120: 58-69.

Mitchell, M. 1999. An Introduction to Genetic Algorithms. The MIT Press.

Wies, G., Groot, J.C.J., Martinez-Ramos, M., 2023. In highly-biodiverse tropical landscapes, multiple-objective optimization reveals opportunities for increasing both conservation and agricultural production. Ecol. Model. 483, 110435. <https://doi.org/10.1016/j.ecolmodel.2023.110435>

Chapter 3

Análisis de conectividad del paisaje.

M en C. Alejandro López-Mendoza (Programa de Maestría Institucional en Ciencias Biológicas, UMSNH)

Dr. Fernando Pineda-García (ENES, UNAM, Unidad Morelia)

Dr. Rafael Aguilar-Romero (ENES, UNAM, Unidad Morelia)

3.1 Introducción

La conectividad del hábitat es un atributo que permite el movimiento de organismos en el paisaje esto a través de los remanentes o fragmentos de bosques (Taylor et al. 1993; Correa et al. 2006; 2014; Saura, 2013). La conectividad puede clasificarse como **estructural** o **funcional** (Saura, 2013). En general, la conectividad está determinada por *la configuración espacial y tamaño de los fragmentos (estructural)*, así como por *la capacidad de dispersión de los organismos (funcional)*, por ejemplo, la capacidad de movimiento de un animal, la dispersión de semillas o de polen entre parches (Saura, 2013; Correa-Ayram et al. 2014).

Analizar la conectividad ayuda a identificar el nivel de accesibilidad que tiene un organismo a los distintos fragmentos de bosque que conforman el paisaje y el grado de asilamiento que pudieran tener algunos remanentes de bosque. De la misma forma, la conectividad permite mantener procesos ecológicos, como mantenimiento de la diversidad genética a través del intercambio de individuos, recolonización de parches que han sufrido diferentes alteraciones o extinciones locales y una mayor resiliencia de los ecosistemas (Bennett, 1999; Damschen et al. 2006; Saura, 2013; Oyama et al. 2017; Hawn et al. 2018).

Actualmente, se ha considerado la conectividad como una estrategia para la conservación al mitigar los impactos de la fragmentación y facilitar la adaptación de las especies a procesos como el cambio climático y pérdida de la conectividad (Saura, 2013; Godínez-Gómez et al. 2019).

Dada la importancia de este atributo del paisaje, en este capítulo abordaremos un ejercicio práctico para determinar el grado de conectividad estructural que tienen los remanentes de bosque. Para los fines prácticos se utilizarán bases de datos previamente recabadas y publicadas donde se analizó el grado de conectividad que tienen los fragmentos de bosques de encinos en la Cuenca de Cuitzeo, Michoacán.

3.2 Modelación de nicho ecológico y distribución de especies

Para estimar la conectividad de los remanentes de bosques como primer paso es necesario modelar el nicho ecológico (climático) de las especies y su distribución en el paisaje. Por lo tanto, en el caso de este ejercicio se utilizarán los datos disponibles recabados por Aguilar-Romero et al., (2016). En la siguiente liga se encuentran las bases de datos que reúnen los registros de presencias de las especies de encinos (*Quercus*) en el área de estudio Cuenca de Cuitzeo, Michoacán, México: (https://drive.google.com/drive/folders/1OGBvXysLwHAHbUe1djwvkW1dt9C5O649?usp=drive_link).

Además, encontrarán un base de datos que contienen registros de ausencias confirmadas para cada especie y para cada punto de registro se cuenta con los datos climáticos. En caso de que quieran seguir este ejercicio con otros datos, lo podrán hacer con datos de registro de campo de la especie de interés o realizar una búsqueda en publicaciones y/o colecciones científicas o en bases de datos públicas de biodiversidad como el GBIF (<https://www.gbif.org/es/>). Generalmente, siempre encontrarán datos de presencia, aunque para el modelo se pueden utilizar datos de presencia y/o ausencia o abundancia. Cabe señalar que para la construcción de los modelos de distribución se requiere información ambiental asociada a los puntos de registros de la especie de interés, ya que la información ambiental es fundamental para la construcción de los modelos. En el caso de que deseen acceder a datos (capas) climáticos precisos a continuación está la siguiente liga: <https://github.com/AngelaCrow/variables-bioclimatica> (Cuervo-Robayo et al. 2014). Las capas climáticas son archivos ráster que contienen matrices conformadas por celdas o píxeles, que cuentan con un valor correspondiente a cada área o superficie (i.e. valores de temperatura, precipitación).

3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES 63

3.2.1 Modelación de nicho con *MaxEnt*.

Para modelar el nicho se utilizará MaxEnt, este algoritmo *predice la distribución de las especies en un área basada en datos climáticos de sus registros*. Estos resultados se proyectan nuevamente en el espacio geográfico, generando mapas de una determinada superficie con valores que reflejan la similitud ambiental a la de las zonas donde se registra la presencia de la especie. Generalmente, de los puntos de presencia se suelen utilizar el 80% de los datos para construir los modelos y el 20% restantes se utilizan para evaluar el nivel de certidumbre del modelo (Phillips et al. 2006; Maher et al. 2014; Hallfors et al. 2016). Se recomienda que del 80 % de los datos se realicen re-muestreos para generar la base de datos que se utilizará en MaxEnt. Estos se desarrollarán mediante el siguiente script en el software R. Una vez generadas las bases de datos para cada especie con los puntos de presencia y las capas climáticas se desarrollarán los modelos en el algoritmo de MaxEnt.

- 1) Remuestrear datos para construir los sets y/o conjuntos de datos tomando como ejemplo los datos de presencia de *Quercus calophylla* para finalmente modelar la distribución por especie. A continuación, se muestra el script.
- 2) Cargar la base de datos de los registros de presencia de *Quercus calophylla*. La base de datos de esta especie cuenta con nueve registros. A continuación, se muestran las instrucciones para leer la base de datos en la consola de R.

En esta parte se pone la ruta en la que estén los datos que se usarán en este ejercicio. Hay que reemplazar esta ruta por la que corresponda en su propia computadora.

```
ruta<-"C:/Users/vegpe/OneDrive/Documentos/papime2023/feb2024_orasiLibro/n12/"
```

Y acá se lee el archivo de datos.

```
# se agrega el nombre de la base de datos
arch<-paste(ruta,"Q_calophylla_presencias.csv",sep="")

#Esta instrucción es para leer la base de
# datos cuyo nombre es calophylla_presencia.csv

o1<- read.csv (file=arch, header=TRUE)
#Esta función permitirá observar el nombre de las columnas
print(names(o1))
```

```
## [1] "especie" "Longitud" "Latitud"
```

```
## [1] "especie" "Longitud" "Latitud"

# Esta función te permite observar la base de datos original
# con nueve registros
print(o1)
```

```
##      especie Longitud  Latitud
## 1 Qcalophylla -101.1758 19.57040
## 2 Qcalophylla -101.5260 19.47754
## 3 Qcalophylla -101.5195 19.47820
## 4 Qcalophylla -101.5244 19.47200
## 5 Qcalophylla -101.4730 19.48353
## 6 Qcalophylla -101.0779 19.65067
## 7 Qcalophylla -101.2254 19.60002
## 8 Qcalophylla -101.3736 19.64041
## 9 Qcalophylla -101.2048 19.61211
```

- 3) Generar muestreo aleatorio. Para este ejercicio la especie cuenta con nueve registros, así que el seleccionaremos el 80 % de los nueve registros, lo cual equivale a siete registros de manera aleatoria. Este procedimiento lo haremos siguiendo las siguientes instrucciones del script de abajo:

```
# Esta función permite definir el número de registros
# a seleccionar
tsu<-7
#Esta función sample permite realizar el muestreo aleatorio
mues<-o1[sample(1:length(o1$especie),tsu,replace=F),]
# Esta función permite visualizar los siete registros
# seleccionados de manera aleatoria
print(mues)
```

```
##      especie Longitud  Latitud
## 5 Qcalophylla -101.4730 19.48353
## 7 Qcalophylla -101.2254 19.60002
## 3 Qcalophylla -101.5195 19.47820
## 2 Qcalophylla -101.5260 19.47754
## 8 Qcalophylla -101.3736 19.64041
## 4 Qcalophylla -101.5244 19.47200
## 1 Qcalophylla -101.1758 19.57040
```

Con la siguiente instrucción se guarda el archivo en formato csv, desde la consola de R. Este archivo despues será cargado en la plataforma de MaxEnt.

3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES 65

```
write.csv(mues, "Q.calophylla_aleatorio.csv")
```

- 4) En este ejercicio utilizaremos el algoritmo de MaxEnt. A continuación, se detallarán los pasos a seguir para la descarga y uso del programa.
- 4.1) Para llevar a cabo la descarga de MaxEnt se debe ingresar a la siguiente liga: https://biodiversityinformatics.amnh.org/open_source/maxent/. Al finalizar la descarga se obtiene una carpeta que contiene el ejecutable de MaxEnt.
- 4.2) Dentro de la carpeta de MaxEnt se encuentran las siguientes carpetas:
- Presencias*: contiene los archivos de las especies del área de estudio que se utilizarán para llevar a cabo la modelación en maxEnt, (Q.calophylla_aleatorio.csv, Quercus_castanea.csv, Q_deserticola.csv, Q_laeta.csv, Q_magnoliifolia.csv, Q_crassipes.csv y Q_rugosa.csv). MaxEnt utiliza los archivos en formato “.csv”, para especificar los sitios de presencia. Para examinar cualquiera de los archivos lo pueden hacer mediante Excel o bloc de notas, en cada caso encontrarán en la primera columna el nombre del archivo y el set al que pertenecen, en la segunda columna la longitud y en la tercera la latitud es recomendable que estas últimas se encuentren en grados decimales.
 - Parámetros ambientales* (capas climáticas): contiene los archivos en formato ráster ASCII (.asc), el cual utiliza MaxEnt; cada capa climática describe un variable ambiental distinta (e.g. Precipitación anual). Todos los archivos en esta carpeta deberán tener la misma resolución y extensión en términos de la superficie o del área de estudio nuestro ejemplo Cuenca del Lago de Cuitzeo.
- 4.3) Para abrir el programa simplemente hay que hacer clic sobre el archivo “maxent.bat”. En el recuadro Samples, se debe cargar el archivo que tiene los registros del muestreo mencionado arriba “Q.calophylla_aleatorio.csv”. En el recuadro Environmental Layers se deben cargar las capas climáticas o variables ambientales; en este caso seleccionamos toda la carpeta Parámetros ambientales y deberían aparecer todas listadas en el recuadro. En este paso se pueden seleccionar las variables a utilizar y se debe indicar si son categóricas o continuas; nosotros tildaremos todas y marcaremos todas como continuas. También debemos seleccionar la carpeta Output en el recuadro Output directory.
- 4.4) Ingresando en la opción Settings, se definirán los parámetros por default de 1000 iteraciones, un límite de convergencia de 0.00001, y un valor regularizado de 1. Esto creará un mapa binario de probabilidad de MaxEnt, considerando una probabilidad máxima y mínima predicha por el modelo para cada pixel.
- 4.5) Seleccionaremos las opciones Create response curves y Do jackknife test to measure variable importance, esta opción nos ayuda a conocer

cuáles de las variables ambientales utilizadas contribuyen en mayor proporción la predicción del modelo. Las curvas de respuesta muestran cómo la idoneidad ambiental predicha por el modelo cambia a medida que varía el conjunto de variables utilizadas en la construcción de los modelos. El parámetro de test Jackknife considera las variables que más aumentan la predicción del modelo cuando son incluidas y puede mostrar la baja predicción del modelo cuando son excluidas.

4.6) Una vez que han sido definidos los parámetros presionamos el botón Run y esperamos unos minutos. Todas las salidas del programa se encontrarán en la carpeta Output; el archivo “Q._calophylla.html” y “Q._calophylla.tif” muestra la salida del modelo y sus principales resultados que se pueden abrir con cualquier navegador.

4.7) Para la evaluación de los respectivos modelos se pueden usar dos métodos de evaluación; 1) la validación de los modelos se realizó mediante el índice de Kappa (K), el cual evalúa la exactitud del modelo basado en presencias y ausencias confirmadas. Este índice evalúa la cantidad de registros de presencias y ausencias que caen dentro del modelo de distribución (suitable área), y fuera del modelo (unsuitable área), este índice va de 0 a 1, un valor cercano a uno indica un buen modelo; y 2) la otra evaluación es mediante el área bajo la curva AUC.

4.8) Finalmente, en sistemas de información geográfica (SIG), se manipularán los modelos con menor incertidumbre por especie y se interceptaron con la capa de vegetación (e.g. “Quercus_calophylla.html”) de Mendoza et al (2011). Para generar la capa vectorial que comprende 510 de fragmentos que representan a la comunidad de encinos en el área de estudio contendrá el nombre de “encinos_conectividad.shp”. en formato shape.

3.2.2 Análisis de conectividad

A continuación, se evaluará la conectividad del paisaje utilizando el estudio de caso de López-Mendoza et al. (2022). En este ejemplo práctico se evaluará el índice de Probabilidad de Conectividad (PC), evalúa el grado de conectividad del paisaje, incluyendo las características de cada fragmento, calidad del fragmento, posición topológica de manera general. Además, se evaluaron las **fracciones relativas de la probabilidad de conectividad**: i) *Parches de flujo* ($dPCfluxk$), ii) *Parches conectores* ($dPCconnectork$) y iii) *Parches con conectividad intra* ($dPCintrak$) (Saura y Torné, 2012).

Para llevar a cabo el ejemplo será necesario implementar la paquetería (Makurhini) desarrollada por Godínez-Gómez y Correa Ayram (2020). Para realizar las evaluaciones de los diferentes índices se utilizarán como insumos las capas vectoriales de 510 fragmentos de bosques de la comunidad de encinos en la Cuenca del lago de Cuitzeo. Para evaluar la conectividad funcional se utilizarán las especies focales de *Melanerpes formicivorus* (pájaro carpintero) y *Aphelocoma ultramarina* (Chara azul), como especies potenciales en la

3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES 67

dispersión de bellotas utilizando una distancia máxima de 10 km medida correspondiente a su rango hogareño de las especies (Koenig et al., 1996; Sutherland et al., 2000; Brown, 2000; López-Barrera et al 2006; Pasendorfer et al 2016; Ramos-Palacios et al 2014).

En este ejercicio utilizaremos el software R studio. A continuación, se detallarán los pasos a seguir para implementar el script y el análisis de la conectividad.

- 1) Para abrir el programa simplemente hay que hacer clic sobre el icono de Rstudio.
- 2) En la consola se usarán las siguientes librerías:
- 3) *Devtools*: desarrolla paquetes para facilitar y simplificar tareas comunes en R: cargar, probar, actualizar, reinstalar y construir paqueterías (Wickham et al., 2022).
- 4) *Remotes*: paquete que nos facilitará la descarga e instalación de repositorios ‘Github’, ‘GitLab’, ‘Bitbucket’, ‘Bioconductor’, u otras subversiones y repositorios (Csárdi et al., 2018).
- 5) *sf*: paquetería enfocada a modelos de atributos simples, utilizado para cargar, manipular, crear y proyectar capas vectoriales. Algunos atributos considerados simples son: puntos, conjunto de puntos, líneas, conjunto de líneas, polígonos, conjunto de polígonos o un conjunto de geometrías antes señalados (Pebesma et al., 2023).
- 6) *Makurhini*: paquete desarrollado para calcular índices de fragmentación y conectividad del paisaje usados en la planeación de la conservación. El paquete provee información útil para identificar los fragmentos del paisaje que se consideran relevantes para mantener la conectividad y modelar escenarios de cambio en la conectividad (Godínez-Gómez y Correa-Ayram, 2020).

Estas instrucciones se deben usar *únicamente si no se han instalado las bibliotecas*.

```
# PARA INSTALAR LAS BIBLIOTECAS POR PRIMERA VEZ

#install.packages("devtools") #Esta instrucciones es para llamar
#paquetes
#install.packages("remotes") #Esta instrucciones es para llamar
#paquetes
```

```
#install.packages("sf") #Esta instrucciones es para llamar paquetes
# Este se instala desde GitHub
#install_github("connectscape/Makurhini", dependencies = TRUE,
#upgrade = "never")
```

Y si ya están instaladas en la compu, estas se deben ejecutar, para activar las bibliotecas.

```
## Warning: package 'devtools' was built under R version 4.3.2
## Warning: package 'usethis' was built under R version 4.3.2
## Warning: package 'remotes' was built under R version 4.3.2
## Warning: package 'cppRouting' was built under R version 4.3.2
## Warning: package 'sf' was built under R version 4.3.2
```

3.2.2.1 Ver los fragmentos

Una vez instalados todas paqueterías en la consola de R se carga el archivo que tienen la información de las capas vectoriales de los 510 fragmentos generados arriba.

A continuación, se muestra el script para leer las capas vectoriales de los 510 fragmentos.

OJO: Hay que definir el **directorio de trabajo** en el que están los datos. Para eso se utiliza la función `setwd()` y la dirección en la que se encuentra la información base. Hay que reemplazar la ruta mostrada en el ejemplo con la ruta de la computadora propia.

```
#setwd ("C:/conectividad")
# Se utilizará la función st_read( ) la cual permite cargar capas
#espaciales y conocer atributos como: Tipo de geometría: Polygon,
#Dimensión : XY, Límites (Bounding Box) Xmax. Ymax, Xmin,
#Ymin.Proyección: WGS 84 / UTM zone 14N
nomenc<-paste(ruta,"encinos_conectividad.shp",sep="")
encinos <- st_read((nomenc),quiet= TRUE)
```

La función `dim` del archivo permite visualizar el número de renglones y columnas de la base de datos.

3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES69

```
dim(encinos)
```

```
## [1] 510 4
```

```
## [1] 510 4
```

Esta función muestra el nombre de las columnas de la base de datos

```
names(encinos)
```

```
## [1] "OBJECTID" "area" "Node" "geometry"
```

```
## [1] "OBJECTID" "area" "Node" "geometry"  
#data("encinos", package = "Makurhini")
```

Con la función `plot()` se visualizará la capa `encinos` con la información asociada al área de cada parche, la clasificación y asignación de color va estar dada por el método de clasificación de natural breaks `jenks breaks = "jenks"`.

```
plot(encinos["area"], breaks = "jenks")
```

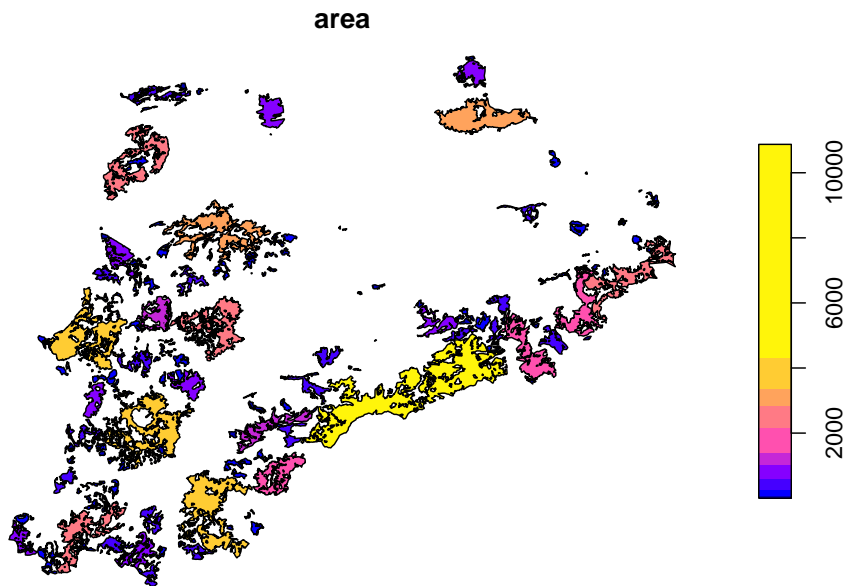


FIGURA 1.

3.2.2.2 Cálculo de la Probabilidad de conectividad (dPC)

En las siguientes instrucciones se evaluará la probabilidad de la conectividad (dPC).

La función `MK_dPCIIC` se utiliza para evaluar la probabilidad de la conectividad y se definen los parámetros por default de una probabilidad de 0.05 y el umbral de distancia de 10,000 m correspondiente al rango hogareño de las especies focales. Se utiliza al final de la instrucción `TRUE` para evaluar la conectividad general.

```
PC <- MK_dPCIIC(nodes = encinos, attribute = NULL,
               distance = list(type = "centroid"),
               metric = "PC", probability = 0.05,
               distance_thresholds = 10000,
               overall =FALSE)
# resumen de la conectividad general
summary(PC)
```

```
##      OBJECTID      area      Node      dPC
## Min.   : 1.0      Min.   : 0.208  Min.   : 1.0  Min.   : 0.00005
## 1st Qu.:128.2    1st Qu.: 3.152  1st Qu.:128.2 1st Qu.: 0.00261
## Median :255.5    Median : 7.483  Median :255.5 Median : 0.00749
## Mean   :255.5    Mean   :142.574 Mean   :255.5 Mean   : 0.22646
## 3rd Qu.:382.8    3rd Qu.: 28.480 3rd Qu.:382.8 3rd Qu.: 0.02369
## Max.   :510.0    Max.   :10872.240 Max.   :510.0 Max.   :42.28164
##      dPCintra      dPCflux      dPCconnector      geometry
## Min.   : 0.00000    Min.   :0.0000491  Min.   :0.000e+00  POLYGON      :510
## 1st Qu.: 0.00000    1st Qu.:0.0026055  1st Qu.:1.565e-15  epsg:32614   : 0
## Median : 0.00002    Median :0.0074751  Median :5.260e-15  +proj=utm ...: 0
## Mean   : 0.16570    Mean   :0.0607616  Mean   :5.303e-15
## 3rd Qu.: 0.00028    3rd Qu.:0.0233603  3rd Qu.:8.692e-15
## Max.   :41.30055    Max.   :1.8798825  Max.   :1.232e-14
```

```
# muestra el resultado de la conectividad general
print (PC)
```

```
## Simple feature collection with 510 features and 7 fields
## Geometry type: POLYGON
## Dimension: XY
## Bounding box: xmin: 233137.8 ymin: 2148570 xmax: 325116.8 ymax: 2221723
## Projected CRS: WGS 84 / UTM zone 14N
## First 10 features:
## OBJECTID area Node dPC dPCintra dPCflux dPCconnector
## 1 1 3.986289 1 0.0025455 0.0000056 0.0025400 1.000025e-14
```

3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES 71

```
## 2      2  0.614712      2 0.0004330 0.0000001 0.0004329 0.000000e+00
## 3      3 55.519574      3 0.0330731 0.0010770 0.0319961 7.979700e-16
## 4      4 23.274051      4 0.0171903 0.0001893 0.0170010 2.740900e-16
## 5      5  6.024369      5 0.0045834 0.0000127 0.0045707 2.337540e-15
## 6      6  1.844975      6 0.0014349 0.0000012 0.0014337 1.083790e-14
## 7      7 293.966858      7 0.1853195 0.0301936 0.1551260 0.000000e+00
## 8      8  3.975813      8 0.0033965 0.0000055 0.0033909 5.538100e-16
## 9      9  6.834431      9 0.0061356 0.0000163 0.0061193 0.000000e+00
## 10     10 13.266170     10 0.0113392 0.0000615 0.0112778 0.000000e+00
##
##                                geometry
## 1 POLYGON ((253313 2149514, 2...
## 2 POLYGON ((251181.8 2150373,...
## 3 POLYGON ((237843.4 2150910,...
## 4 POLYGON ((251181.8 2150373,...
## 5 POLYGON ((242420.1 2151391,...
## 6 POLYGON ((250107 2151580, 2...
## 7 POLYGON ((252309.4 2149742,...
## 8 POLYGON ((250441.5 2151946,...
## 9 POLYGON ((250460.6 2152360,...
## 10 POLYGON ((238430.6 2151968,...
```

3.2.2.3 Visualización de los índices de Probabilidad de conectividad

Esta instrucción genera el mapa de la conectividad general. Con la función `plot()` se visualizará la capa `encinos` con la información asociada al área de cada parche, la clasificación y asignación de color va estar dada por el método de clasificación de natural breaks `jenks` `breaks = "jenks"`.

```
plot(PC[c("dPC")], breaks = "jenks")
```

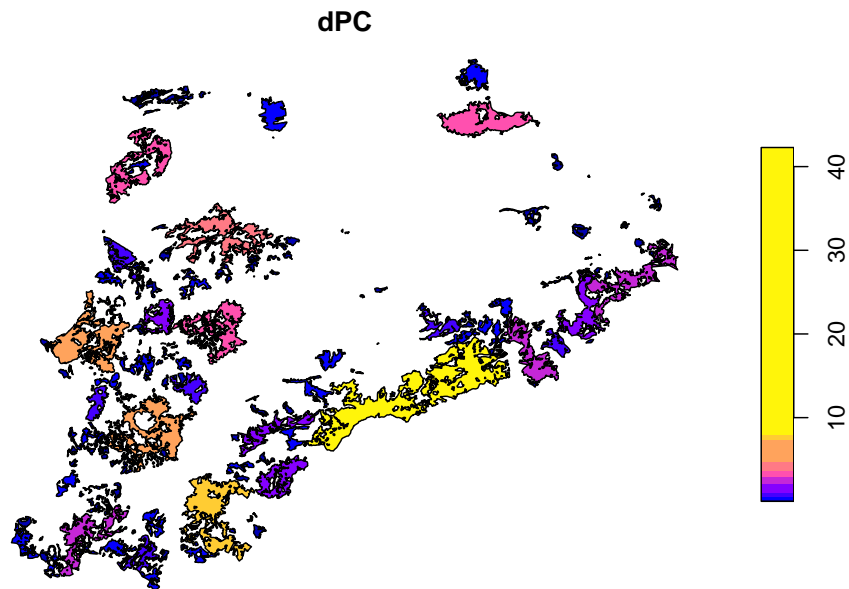


FIGURA 2

3.2.2.4 Cálculo de las diferentes fracciones del (dPCk) y sus respectivas métricas (dPCintra, dPCflux, dPCconnector).

La función `MK_dPCIIC` se utiliza para evaluar la probabilidad de la conectividad y se definen las tres fracciones mencionadas líneas arriba (*dPCintra*, *dPCflux*, *dPCconnector*), con los mismos parámetros por default de una probabilidad de 0.05 y el umbral de distancia de 10,000m correspondiente al rango hogareño de las especies focales. Al final de la función se utiliza el argumento `overall=FALSE`, para evaluar las fracciones de las métricas que se muestran abajo para los primeros seis fragmentos.

```
PC <- MK_dPCIIC(nodes = encinos, attribute = NULL,
               distance = list(type = "centroid"),
               metric = "PC", probability = 0.05,
               distance_thresholds = 10000,
               overall = FALSE)
```

PC

```
## Simple feature collection with 510 features and 7 fields
## Geometry type: POLYGON
## Dimension: XY
```


3.2. MODELACIÓN DE NICHOS ECOLÓGICOS Y DISTRIBUCIÓN DE ESPECIES73

```
## Bounding box: xmin: 233137.8 ymin: 2148570 xmax: 325116.8 ymax: 2221723
## Projected CRS: WGS 84 / UTM zone 14N
## First 10 features:
##   OBJECTID      area Node      dPC dPCintra dPCflux dPCconnector
## 1         1    3.986289    1 0.0025455 0.0000056 0.0025400 1.000025e-14
## 2         2    0.614712    2 0.0004330 0.0000001 0.0004329 0.000000e+00
## 3         3   55.519574    3 0.0330731 0.0010770 0.0319961 7.979700e-16
## 4         4   23.274051    4 0.0171903 0.0001893 0.0170010 2.740900e-16
## 5         5    6.024369    5 0.0045834 0.0000127 0.0045707 2.337540e-15
## 6         6    1.844975    6 0.0014349 0.0000012 0.0014337 1.083790e-14
## 7         7  293.966858    7 0.1853195 0.0301936 0.1551260 0.000000e+00
## 8         8    3.975813    8 0.0033965 0.0000055 0.0033909 5.538100e-16
## 9         9    6.834431    9 0.0061356 0.0000163 0.0061193 0.000000e+00
## 10        10   13.266170   10 0.0113392 0.0000615 0.0112778 0.000000e+00
##                               geometry
## 1 POLYGON ((253313 2149514, 2...
## 2 POLYGON ((251181.8 2150373,...
## 3 POLYGON ((237843.4 2150910,...
## 4 POLYGON ((251181.8 2150373,...
## 5 POLYGON ((242420.1 2151391,...
## 6 POLYGON ((250107 2151580, 2...
## 7 POLYGON ((252309.4 2149742,...
## 8 POLYGON ((250441.5 2151946,...
## 9 POLYGON ((250460.6 2152360,...
## 10 POLYGON ((238430.6 2151968,...
```

Con la función `plot()` se visualizará la capa encinos con la información asociada al área de cada parche. La clasificación y asignación de color se define por el método de clasificación de natural breaks `jenks` (argumento `breaks = "jenks"`).

```
plot(PC[c("dPCintra", "dPCflux", "dPCconnector")], breaks = "jenks")
```

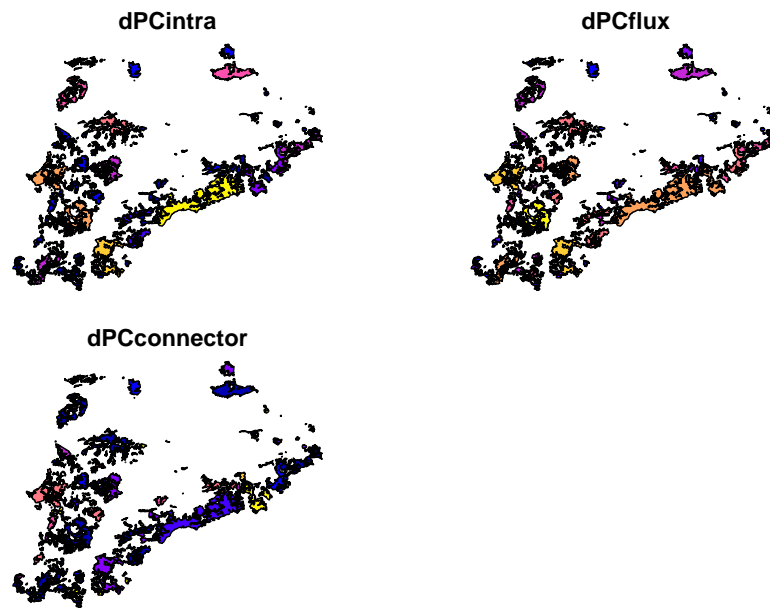


FIGURA 3

3.2.3 Conclusiones

El presente ejercicio brinda un acercamiento a diferentes metodologías y herramientas para la modelación de nichos ecológicos y análisis de conectividad del paisaje utilizando datos de presencias, ausencias, variables ambientales y atributos intrínsecos de la estructura del paisaje p.e. el área de los fragmentos de bosques. Se clasificaron los fragmentos de bosque por área (Figura 1), el análisis de conectividad evaluado por la probabilidad de conectividad (dPC) presenta valores bajos a nivel de parches, con un valor promedio de 0.22 y un valor máximo de 42.28, El análisis de la probabilidad de la conectividad general permite visualizar los fragmentos de mayor importancia para mantener la conectividad general (Figura 2). Por otro lado, las fracciones relativas de la probabilidad de la conectividad nos permite definir cuales fragmentos tienen mayor importancia en función de la métrica que se analice (dPCintra, dPCflux, dPCconnector), mostrando la importancia que presentan otros parches a partir del cálculo de los tres elementos de conectividad, ya sea como parches de flujo, peldaños o parches conectores y los fragmentos con alta conectividad intra. (Figura 3). Este análisis podría traducirse en la baja capacidad de movimiento de las especies focales consideradas como dispersoras de bellotas resultado de la pérdida de hábitats. Los análisis de conectividad tienen la finalidad de conocer el estado actual de los bosques de encinos, identificar los fragmentos más importantes para mantener la conectividad del paisaje y generar información útil para definir y

priorizar acciones de conservación y restauración manteniendo la conectividad de los bosques a largo plazo.

3.3 Referencias

Aguilar-Romero R, García-Oliva F, Pineda-García F, Torres I, Peña-Vega E, Ghilardi A, Oyama K. 2016. Patterns of distribution of nine *Quercus* species along an environmental gradient a fragmented landscape in central Mexico. *Botanical Science* 94 (3): 471-482 <https://doi.org/10.17129/botsci.620>

Brown, J.L. (2001). The Mexican Jay a Model System for the Study of Large Group Size and Its Social Correlates in a Territorial Bird. In *Model Systems in Behavioral Ecology Integrating Conceptual, Theoretical, and Empirical Approaches*; Dugatkin, L.A., Ed.; Princeton University Press: Oxford, UK, 2001; Volume 70, pp. 338–358. ISBN 0-691-00652-0.

Csárdi G, Hester J, Wickham H, Chang W, Morgan M, Tenenbaum D (2023). remotes: R Package Installation from Remote Repositories, Including ‘GitHub’. R package version 2.4.2, <https://github.com/r-lib/remotes#readme>, <https://remotes.r-lib.org>.

Godínez-Gómez, O. and Correa Ayram C.A. (2020). Makurhini: Analyzing landscape connectivity. <https://doi.org/10.5281/zenodo.3771605>

Hallfors, M.H.; Liao, J.; Dzurisin, J.; Grundel, R.; Hyvarinen, M.; Towle, K.; Wu, G.C.; Hellmann, J.J. (2016). Addressing potential local adaptation in species distribution models: Implications for conservation under climate change. *Ecol. Appl.* 2016, 26, 1154–1169.

Koenig, W.D.; Vuren, D.V.; Hooge, P.N. (1996). Detectability, philopatry, and the distribution of dispersal distance in vertebrates. *Trends in Ecology & Evolution*, 11, 514–517.

López-Barrera, F.; Manson, R.H.; González-Espinoza, M.; Newton, A.C. (2006). Effects of the type of montane forest edge on oak seedling establishment along forest-edge-exterior gradients. *Forest Ecology and Management*, 225, 234–244.

López-Mendoza, A., Oyama, K., Pineda-García, F., y Aguilar-Romero, R. (2022). Defining Conservation Priorities for Oak Forests in Central Mexico Based on Networks of Connectivity. *Forests*,13(7), 1085.

Maher, S.P.; Randin, C.F.; Guisan, A.; Drake, H.J. (2014). Pattern-recognition ecological niche models fit to presence-only and presence-absence data. *Methods Ecol. Evol.* 5, 761–770.

Mendoza, M.E.; López-Granados, E.; Geneletti, D.; Pérez-Salicrup, D.R.; Salinas, V. (2011). Analyzing land cover and land use change processes at watershed level: A multitemporal study in the Lake Cuitzeo Watershed, Mexico (1975–2003). *Applied Geography*, 31, 237–250

- Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With Applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>
- Pesendorfer, M.B.; Sillet, T.S.; Koenig, W.D.; Morrison, A.A. (2016). Scatterhoarding corvids as seed dispersers for oaks and pines: A review of widely distributed mutualism and its utility to habitat restoration. *Ornithological Application*. 118, 215–237.
- Phillips, S.J.; Anderson, R.P.; Schapire, R.E (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Modell*, 190, 231–259
- Ramos-Palacios, C.R.; Badano, E.I.; Flores, J.; Flores-Cano, J.A.; Flores-Flores, J.L. (2014). Distribution patterns of acorns after primary dispersión in a fragmented oak forest and their consequences on predators and dispersers. *European Journal of Forest Research*, 133, 391–404.
- Saura, S., y Torné, J. (2012). *Conefor 2.6 User Manual* (May 2012). Universidad Politécnica de Madrid.
- Sutherland, G.D.; Harestad, A.S.; Price, K.; Lertzman, K.P. (2000) Roedores Scaling of Natal Dispersal Distances in Terrestrial Birds and Mammals. *Conservation Ecology*, 4, 16.
- Sutherland, G.D.; Harestad, A.S.; Price, K.; Lertzman, K.P. (2000). Roedores Scaling of Natal Dispersal Distances in Terrestrial Birds and Mammals. *Conservation Ecology*, 4, 16.
- Wickham H, Hester J, Chang W, Bryan J (2022). *devtools: Tools to Make Developing R Packages Easier*. <https://devtools.r-lib.org/>, <https://github.com/r-lib/devtools>.

Chapter 4

Modelación de datos globales con NETCDF

Dr. Guillermo Murray (IIES UNAM Morelia)

4.1 Introducción.

Durante la mayor parte de nuestra humanidad los seres humanos vivimos desconectados unos de otros. Vivíamos en aldeas y pueblos remotos, con pocas personas a nuestro alrededor, con la que compartíamos un sinnúmero de similitudes culturales. No obstante, conforme nuestra tecnología y sociedad fueron evolucionando, las distancias se hicieron cada vez más cortas y el mundo cada vez más pequeño.

En el mundo moderno disponemos de fotografías casi instantáneas de cualquier parte del mundo, de satélites capaces de monitorear toda clase de condiciones terrestres (incendios, gases de efecto invernadero, la tasa fotosintética, por mencionar algunos) y modelos matemáticos computacionales que pueden simular el clima de todo el mundo de los últimos miles de años o los próximos cientos.

Si bien algunas de estas mediciones globales se procesan en forma de imagen, la mayoría ha transitado a un formato común, conocido como NetCDF, que facilita el manejo de las mismas. En este capítulo te explico cómo se procesan y usan los datos NETCDF en R para poder interpretarlos, graficarlos y generarlos.

4.2 El origen de las mediciones globales

Los seres humanos siempre hemos estado interesados en medir el mundo que nos rodea. Debido a lo anterior, decir cuando comenzamos a medir todo el Globo Terráqueo es complicado. La primera circunnavegación mundial, que contaba con un geógrafo especializado en generar mapas fue en 1519 con Fernando de Magallanes y el primer mapamundi completo fue publicado en 1527 por el portugués Diego Ribero. Después de eso comenzamos a tener mapas de yacimientos de minerales preciosos, vegetación, población y casi todo lo que se nos fue ocurriendo que podíamos representar en dos dimensiones. No obstante, el mundo seguía estando bastante desconectado. La globalización, y con ella la sistematización de las mediciones planetarias, comenzó realmente con el fin de la segunda guerra mundial en 1950 y el proceso se intensificaría aún más con el lanzamiento de satélites interplanetarios, capaces de tomar fotografías con diferentes sensores, de toda la superficie terrestre durante los 1960s.

En estos últimos 60 años nuestra capacidad de medir el planeta aumentó de forma exponencial. Las primeras fotografías eran del espectro visible (como las que tomamos nosotros con nuestras cámaras), pero rápidamente se comenzaron a usar sensores infrarrojos, ultravioletas y refinados para detectar con mucho detalle ciertas bandas del espectro electromagnético. Con ello, el día de hoy, somos capaces de medir desde el espacio: la concentración de gases diversos gases en la atmósfera (p.ej. dióxido de carbono), el contenido de agua y biomasa de la vegetación (conocido como VOD, por la técnica en inglés vertical optical depth), la humedad del suelo (a través de microondas capaces de penetrar hasta 20cm en el mismo), la fotosíntesis terrestre y oceánica (como fracción de la luz absorbida y por fluorescencia), incendios (como puntos de temperatura extrema), por mencionar algunos.

A la par de la evolución de las mediciones satelitales, también se dio un incremento en el poder y capacidad computacional, y junto con ello de nuestros modelos matemáticos del sistema Tierra. Los primeros modelos de rejilla globales se crearon en la 1960 y sólo incluían componentes atmosféricos y oceánicos. Sesenta años después cuenta con módulos acoplados para hielo, ciclos biogeoquímicos, el ciclo del carbono, vegetación, aerosoles atmosféricos, la química de la atmósfera, resolución vertical del ambiente marino, incendios, por mencionar algunos. En la actualidad, este tipo de modelos reciben el nombre de ESMs (Earth System Models o Modelos del Sistema Tierra) por su capacidad para simular y predecir el estado de una amplia diversidad de factores físicos, químicos y bióticos de nuestro planeta.

La conjunción de datos satelitales y de modelación, sumado a los esfuerzos sistemáticos de mediciones de campo globales, han hecho posible conocer el estado de nuestro planeta como nunca antes y poder generar predicciones para un futuro próximo y lejano. El *Panel Intergubernamental de Cambio Climático (IPCC por sus siglas en inglés)* es el principal órgano encargado de reunir y sintetizar dicha información. No obstante, debido a la magnitud de los datos, la

complejidad y diversidad de procesos, mucha de la información existente no ha sido analizada a profundidad. Lo anterior es particularmente cierto cuando la atención se centra en un país o una región, escala a la que existen pocos análisis usando los datos disponibles y en la que se puede tener una alta influencia para la toma de decisiones y la generación de políticas públicas.

Una de las razones por las cuales no hay procesamiento de datos a escala regional, es por la dificultad que implica usar y entender este tipo de datos. El utilizar imágenes globales de alta resolución necesita de mucho poder computacional, ocupa mucho espacio de almacenamiento y tiene una inherente dificultad técnica elevada. Con ello en mente, y para tratar de facilitar el uso de datos globales, se creó un formato específico conocido como **NetCDF (Network Common Data Form)**. El objetivo de este capítulo es mostrar cómo se utilizan este tipo de datos, sus ventajas y el tipo de información que pueden generar. Así mismo, se mostrarán algunos repositorios clave donde se pueden descargar de manera gratuita y de acceso abierto, diversas variables provenientes de mediciones globales, satélites y modelos matemáticos.

4.3 El formato NetCDF

El formato NetCDF (formulario de datos comunes en red) es una estructura de archivos generada para almacenar variables multidimensionales. Si bien, teóricamente podría ser utilizado para guardar cualquier dato en forma matricial, en la práctica se utilizan para guardar mapas que además incluyen la variable temporal. En otras palabras, los datos almacenados como NetCDF suele tener latitud, longitud y tiempo, para una o más variables (p.ej. temperatura, precipitación, radiación) (Figura 1). El formato para almacenar los datos de esta manera fue creado por la Corporación Universitaria para la Investigación Atmosférica (UCAR por sus siglas en inglés) y es gratuito y de acceso abierto (<https://www.unidata.ucar.edu/software/netcdf/>).

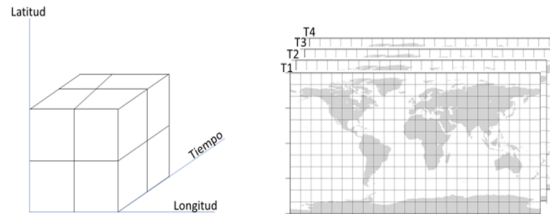


Figura 1. Ejemplo conceptual de una matriz tridimensional geográfica regular y en rejilla. Este tipo de datos son los más comunes de encontrar y usar en formato NetCDF.

Algunas de las funciones más importantes de los archivos NetCDF son: su portabilidad e interdependencia (se pueden acceder desde cualquier computadora y se encuentran disponibles en Windows, Mac y Linux), compatibilidad entre versiones (i.e. aunque se actualicen los formatos, se pueden procesar cualquier versión de los mismos), son compartibles y complementables (i.e. cualquiera puede agregar datos extras a una base existente o modificar los que recibe). Adicionalmente a lo anterior, tienen la propiedad de comprimir las bases de datos de manera excepcional, logrando reducir el espacio computacional necesario hasta en 90%.

Debido a lo anterior la mayoría de las bases de datos geográficas mundiales se encuentran en este formato. Algunos ejemplos son todos los datos de modelos ESM utilizados por el IPCC, datos climáticos de pasado provenientes de universidades y centro de investigación (p.ej. CRU), datos de humedad de suelo provenientes de satélites, entre otros. Cabe mencionar que la idea central del formato, fue facilitar el compartir datos entre instituciones de manera sencilla, debido a lo anterior, casi todas las iniciativas globales que emplean NetCDF suelen ofrecerlos de manera gratuita. En este caso particular, la limitante no es la existencia de datos, pero la formación de personas capaces de utilizarlos, graficarlos y utilizarlos, para generar información útil para sus contextos.

4.4 Instalación del formato NetCDF en cualquier computadora y librerías correspondientes en R.

Para poder utilizar estos datos en cualquier computadora, *lo primero que se necesita es instalar los ejecutables para poder manejarlos*. Para descargar los

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.81

archivos, se debe ingresar a la siguiente dirección: <https://downloads.unidata.ucar.edu/netcdf/>. Una vez dentro del sitio web se debe buscar la última versión disponible, correspondiente al sistema operativo que se desee utilizar. Por ejemplo, en el caso de una computadora Windows se descargaría e instalaría el archivo “netCDF4.9.2-NC3-64.exe” (actualizado para 2023). *Una vez instalados los archivos, se puede proceder a instalar las librerías en R. Estos pasos no pueden ser ejecutados a la inversa (primero las librerías en R y luego instalar NetCDF en la computadora), dado que las librerías de R necesitan que exista el formato instalado en el sistema.*

Una vez instalado el formato NetCDF en la computadora, en R se tienen dos posibles opciones de librería. La librería más antigua es “netcdf”, que fue más funcional en su momento, pero la mayoría de sus funciones ya no se encuentran actualizadas. Debido a lo anterior, se recomienda utilizar la librería “ncdf4”, misma que se puede instalar directamente desde la terminal. La instrucción en la terminal de R se verá así:

```
#install.packages("ncdf4").  
#library(ncdf4)
```

4.5 Procesamiento de datos globales en formato NetCDF en R.

La primera parte para procesar datos NetCDF en R es decidir con cuales se quiere trabajar. Para los ejemplos vertidos en este capítulo se utilizará la precipitación global histórica, proveniente de la Climate Research Unite (CRU) de la Universidad de East Anglia en Reino Unido. CRU tiene una serie amplia de datos globales climáticos, procesados de estaciones de monitoreo internacionales, desde 1901 hasta el año en curso (en este caso 2022). Los datos se pueden consultar en: <https://crudata.uea.ac.uk/cru/data/hrg/>. El sitio web incluye versiones promediadas por país, en formato de texto y los datos en rejilla, en formato NetCDF. Para los ejemplos de este capítulo se descargaron los datos de: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/cruts.2304141047.v4.07/pre/, que corresponden a precipitación global en la versión 4.07. Utilizaremos los datos de 1901-2022.

El archivo de datos se llama “cru_ts4.07.1901.2022.pre.dat.nc”,

OJO: en la pagina están en formato “gz”, que está comprimido. Primero hay que descargar el archivo comprimido a lacompu local, luego descomprimirlo y luego se podrá leer.

4.5.1 Abrir un archivo NetCDF en R.

Y ahora cargamos el archivo:

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.83

La segunda instrucción importante es “dim”, misma que nos ayudará a comprobar las dimensiones que tiene la variable extraída (en este caso lluvia). Si escribimos en la terminal:

```
#dim(lluvia)
```

el resultado debe ser 720,360,1464. Correspondiente a una matriz mundial de medio grado por medio grado (360° x 180° cada 0.5°) y 1464 pasos temporales (número de meses en 112 años).

4.5.2 Mapeo sencillo de datos NetCDF.

Existen diversas maneras de representar un mapa gráficamente, para este primer ejemplo utilizaremos dos de las opciones más sencillas. La primera es simplemente la función “image” parte de las librerías básicas de R; mientras que la segunda, es la función “image.plot” parte de la librería “fields” (hay que instalarla con antelación al igual que hicimos con “ncdf4”, mediante la función “install.packages”). Ambas funciones nos permiten graficar una matriz de dos dimensiones, por lo que, para este primer ejercicio, debemos elegir únicamente un paso temporal. La forma más sencilla de hacer lo anterior es con corchetes, donde las primeras dos dimensiones se dejarán vacías, indicando que se seleccionen completas y la tercera tendrá un único punto temporal. Para nuestro ejemplo se verá así:

```
#mapa1<-lluvia[, ,1]  
#image(mapa1)  
  
#library(fields)  
#image.plot(mapa1)
```

Lo anterior debe resultar en dos mapas similares a los del ejemplo de la Figura 3.

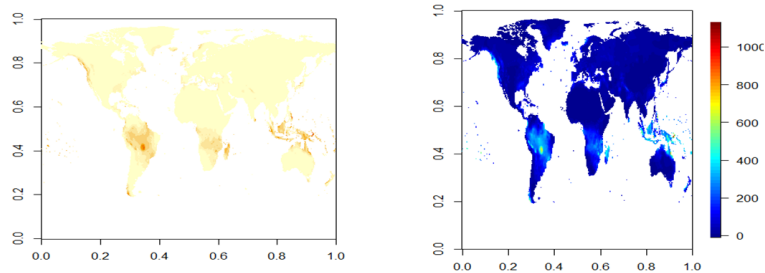


Figura 3. Dos ejemplos para la precipitación mundial durante enero de 1901. A la izquierda usando la función “image” y a la derecha “image.plot”.

Partiendo de los ejemplos anteriores, se puede probar hacer otros ejercicios. Por ejemplo, se puede seleccionar otro mes (en lugar de 1, se puede poner 1464 para tener diciembre de 2022, o 1458 para tener junio de 2022, etc.). También podemos utilizar la misma lógica para seccionar partes del mapa, por ejemplo, si deseamos visualizar una región más pequeña (rectangular). Lo anterior se hace nuevamente mediante el uso de corchetes, pero esta vez eligiendo un rango para la latitud y la longitud (usando dos puntos). Por ejemplo, para seleccionar un rectángulo que contenga únicamente a México, podríamos usar la instrucción:

```
#Mapa2<-lluvia[123:188,210:246,1]
```

4.5.3 Transformación a datos anuales.

Una vez que logramos los primeros pasos, lo siguiente es comenzar a manejar los datos para calcular patrones espaciotemporales de interés para contestar preguntas. Para ello lo primero es transformar los datos mensuales en datos anuales. Esto requiere de dos instrucciones, la primera es transformar la matriz inicial a un arreglo de cuatro dimensiones en que se separen mes y año. La segunda es una de las funciones más importantes en el manejo matricial, la función “apply”, que permite aplicar una función a través de una dimensión. Entonces nuestro script se verá así:

```
#lluvia2<- array(lluvia,dim=c(720,360,12,122))
#lluvia3<-apply(lluvia2,c(1,2,4),sum)
```

Esta última instrucción mantiene las dimensiones 1 (longitud), 2 (latitud) y 4 (años), sumando los datos mensuales. La nueva configuración se puede com-

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.85

probar con la función “dim” como se mencionó anteriormente y tiene tres dimensiones: 720,360,122 (longitud,latitud, años comenzando desde 1901). Otro punto importante, es que en este caso usamos la instrucción “sum” pues la precipitación mensual se debe sumar para dar la anual; no obstante, en el caso de la temperatura, se tiene que usar la instrucción “mean” pues la temperatura mensual se promedia para dar la anual. En este momento vale a pena parar para experimentar con la instrucción apply, pues su uso no sólo se aplica en mapas globales, sino en el manejo de cualquier tipo de matriz multidimensional. Aquí también se pueden concatenar instrucciones de las secciones pasadas, por ejemplo, para obtener la lluvia anual únicamente para México (usamos el “apply” sobre Mapa2) o si queremos visualizar la lluvia de un año en particular (por ejemplo, para 2011, un año muy seco en México), mediante el uso de corchetes. Un ejemplo se vería así:

```
#Mex<-apply(array(Mapa2,c(66,37,12,144)),c(1,2,4),sum)
#image.plot(Mex[, ,111])
```

Lo anterior resulta en un mapa que se ve como la figura 4.

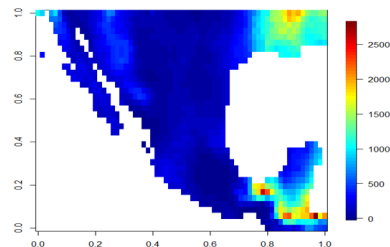


Figura 4. Ejemplo de concatenación de instrucciones para visualizar la lluvia anual de México durante 2011.

4.5.4 Series de tiempo.

La siguiente parte para el manejo de datos geográficos matriciales provenientes de NetCDF es generar series de tiempo. Esto nos servirá para contestar preguntas como: ¿Cómo ha cambiado la precipitación promedio global de 1901 a 2022? Por supuesto, estas nuevas herramientas se pueden concatenar con lo explicado anteriormente, para recortar secciones rectangulares del mundo (p.ej. México) y elegir grupos de años de interés (p.ej. 2000-2022). Pero antes de proceder se necesita de un paso intermedio en que se calcule el área de los mapas. Cada

matriz que se usa para representar al mundo, tiene una proyección propia, en que la Tierra tridimensional se convierte a dos dimensiones. Con ello, aunque las celdas sean regulares (p.ej. un grado o medio grado), su tamaño no lo es. Para cada proyección se debe hacer una corrección por el tamaño de celda. En el caso particular de los datos de CRU que usamos para este ejemplo, así como la mayoría de los usados por el IPCC, se utiliza una proyección cilíndrica. La corrección para el área está dada por la siguiente fórmula:

```
lat=seq(-89.75,89.75,by=0.5)
da=abs((55103)/(55103))cos(lat*pi/180)
dax<-t(replicate(720,da))
DA<-1.29*1014
```

Donde la primera instrucción se utiliza para crear un arreglo de la latitud según la resolución que se tenga (en este caso 0.5°) y la segunda es el tamaño de una celda en el Ecuador (55km de largo y 55km de ancho) multiplicada por la corrección latitudinal, caracterizada por la función coseno. La última instrucción es ajustar la matriz a las mismas dimensiones que la original (720*360) para poder multiplicarla. Finalmente, la cuarta instrucción (DA), es simplemente el área total de la superficie terrestre sin contar el océano. v

Una vez que tenemos el área por celda, se puede proceder al cálculo de las series de tiempo, utilizando la función “apply” sobre los datos anuales, manteniendo únicamente la dimensión tercera (tiempo). En un primer paso se multiplica la lluvia por el área global y en un segundo se hace el “apply”. Las instrucciones se ven así:

```
#lluvia4<-lluvia3*replicate(122,dax)
#lluvia_anual<-apply(lluvia4,c(3),sum,na.rm=T)/DA
```

En la primera instancia se multiplica la lluvia por una reproducción del área global, según el número de pasos temporales que se tengan (años), en este caso trabajamos con 1901-2022, así que son 122 años. En la segunda instrucción se utiliza la función “apply” para sumar la lluvia de todo el planeta para cada año (manteniendo únicamente la dimensión tres), se agrega la instrucción na.rm=T, para que ignore los datos vacíos sobre el océano. Finalmente se divide entre el área total de la superficie terrestre (sin océanos), para obtener el valor de precipitación promedio anual del mundo. Si se desea la precipitación total (en mm) se puede eliminar esa última división. El resultado es un vector con 122 valores de largo, mismo que se puede graficar simplemente con la instrucción plot. Un ejemplo de la instrucción se ve así:

```
#plot(1901:2022,lluvia_anual,type="l",ylab="(mm)",
#      main="Lluvia promedio anual global",
#      las=1,xlab="")
```

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.87

El resultado se puede ver en la figura 5, que representa una serie de tiempo de la precipitación global durante los últimos 122 años (1901-2022).

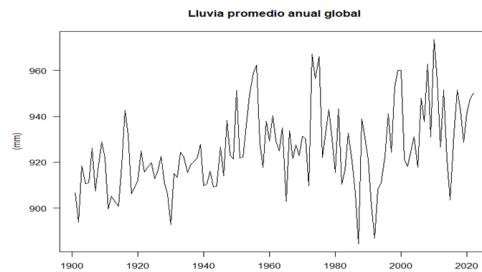


Figura 5. Ejemplo de una serie de tiempo para datos globales. En este caso se graficó la lluvia global promedio anual, para 1901-2022.

Un buen ejercicio en este momento es tratar de concatenar lo que se ha visto hasta ahora, para tratar de generar la misma serie de tiempo, pero únicamente para México o una región seleccionada (dado que se está trabajando con rectángulos, no será exactamente un país, pero se puede tener una aproximación a modo de ejercicio). Para ello se tienen que modificar las últimas tres instrucciones descritas en esta sección.

4.5.5 Mapas de medias (gridded means) y anomalías.

De manera similar a cómo se procesan los datos para calcular series de tiempo, se puede usar la función “apply” para generar mapas de promedios en el espacio (p.ej. la lluvia promedio por celda durante todo el periodo), o para calcular anomalías espacialmente explícitas. Una anomalía es simplemente un mapa en que se muestra que tanto difieren los valores de un año dado (o mes, o periodo) con respecto al promedio de la serie. Las herramientas para generar ambos productos (mapas de promedios o de anomalías) son las mismas y se describen en esta sección. La primera instrucción necesaria es promediar los datos manteniendo únicamente la latitud y la longitud (promediar en el tiempo por rejilla), nuevamente con la función “apply”. En este caso no es necesario hacer la corrección por área, dado que los datos se presentan para cada rejilla. El script entonces se verá así:

```
#Mapa3<-apply(lluvia3,c(1,2),mean)
#image.plot(Mapa3)
```

Lo anterior resulta en un mapa para la precipitación promedio mundial por celda de 1901-2022, que se puede ver en la figura 6.

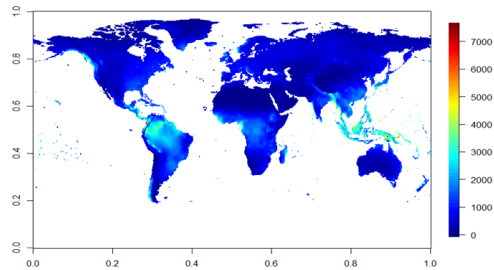


Figura 6. Mapa de la precipitación promedio mundial por celda, de 1901-2022.

Ahora, para hacer un mapa de una anomalía, se tiene que elegir un año en particular (mediante corchetes) y restarle el promedio calculado en el mapa anterior. El resultado nos permitirá ver, por rejilla, que tanto se diferencia un año del promedio de la serie. Tomemos por ejemplo el 2011, un año muy seco para México y Europa. El script se verá así:

```
#Mapa4<-lluvia3[, , 111]-Mapa3
#clr<-colorRampPalette(c("red", "grey90", "blue"))(1000)
#Mapa5<-ifelse(Mapa4>=500, 500, Mapa4)
#Mapa5<-ifelse(Mapa5<=-500, -500, Mapa5)
#image.plot(Mapa5, zlim=c(-500, 500), col=clr)
```

En este caso agregamos instrucciones adicionales para controlar el color del mapa y los valores máximos y mínimos que se grafican, para mejorar la visualización. La primera instrucción es simplemente el cálculo de la anomalía espacial (usamos los corchetes para elegir el año 2011 y le restamos el promedio de la serie, calculado con antelación).

En la segunda instrucción creamos una paleta de colores, en este caso usamos un degradado sencillo del rojo al azul, para indicar extremos negativos y positivos. Esta instrucción es sumamente útil para crear paletas de colores degradadas, y se puede combinar con todas las formas de selección de color en R. Adicionalmente uno puede elegir el número de intervalos que desea (en el ejemplo se eligieron mil).

La cuarta y quinta instrucciones las usamos para seleccionar los límites máximos de la variable, e impedir que el mapa tenga extremos que impidan la

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.89

visualización. Estos límites se determinan en función de la anomalía esperada y cambiarán según el periodo y la variable elegida. Si no se incluyen el mapa aparecerá con espacios en blanco, como se muestra en la figura 7. Finalmente, la última instrucción combina todos los elementos para generar el mapa, definiendo un límite de visualización igual al establecido con antelación.

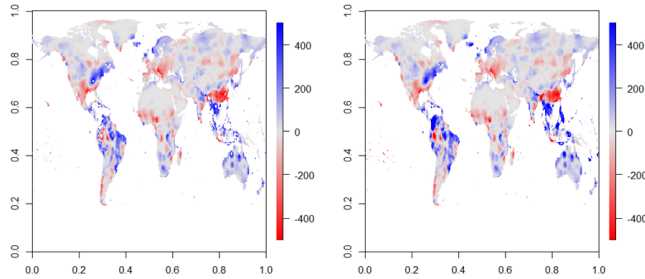


Figura 7. Ejemplo de graficación de una anomalía temporal. A la izquierda se muestra sin incluir límites de visualización y se pueden apreciar espacios en blanco. A la derecha el mapa correcto con los límites aplicados. En este caso particular se muestra la anomalía de lluvias durante 2011 comparada con el promedio de 1901-2022, en rojo se aprecian lugares con lluvia por debajo del promedio y en azul por encima. Se puede observar que México y Europa sufrieron una fuerte sequía como consecuencia del fenómeno “La Niña” durante ese año.

Nuevamente este ejercicio se puede concatenar con lo anterior, seleccionando únicamente una región de interés. Adicionalmente, la anomalía se puede mostrar como porcentaje, si dividimos el mapa entre el promedio de la serie, esto se vería así:

```
#Mapa6<-(lluvia3[, ,111]-Mapa3)/Mapa3*100
```

Para este ejemplo se tendrían nuevamente que definir límites de visualización (-100 y 100) mediante la función “ifelse”, igual que en el procedimiento anterior y nuevamente definir los colores que se deseen.

4.5.6 Mapas de tendencias (gridded linear trends).

La última técnica de manejo de datos de este tipo, es la visualización del cambio promedio en el tiempo por celda. Esto es conocido como una tendencia lineal espacial (spatial trend) e implica el cálculo de la pendiente de regresión para cada celda del mapa. Para ello se tiene que definir un horizonte temporal de interés, en este ejemplo utilizaremos los últimos 30 años. La tendencia lineal se crea a partir de una función que calcule la tendencia lineal y extraiga únicamente la pendiente. Esto se ve así:

```
#period<-91:120
#trend<-apply(lluvia3[, ,period],c(1,2),function(x){
#if(any(is.na(x))){return(NA)} else{
#fit<-lm(x~period)
#return(fit$coef[2]) }})
```

Donde la primera instrucción se utiliza para definir el periodo de interés y debe ser modificada de acuerdo a la estructura de los datos (p.ej. en este caso hay 122 años, pero si sólo tuviéramos 80, entonces elegiríamos 51:80, si queremos los últimos 30). La siguiente instrucción es la creación de una función que genera una tendencia lineal por celda y extrae la pendiente (el segundo coeficiente del modelo lineal), utiliza la función “apply” para indicar que el cálculo se hará en la tercera dimensión, manteniendo las primeras dos (longitud y latitud). Se debe seleccionar la variable de interés, en este caso la lluvia ya transformada a años; el resultado es un mapa de pendientes. Para este ejemplo en particular, el resultado se puede observar en la figura 8. Para este ejemplo (figura 8) se hizo nuevamente uso de las instrucciones usadas con antelación, donde se definen los límites de visualización y los colores del mapa.

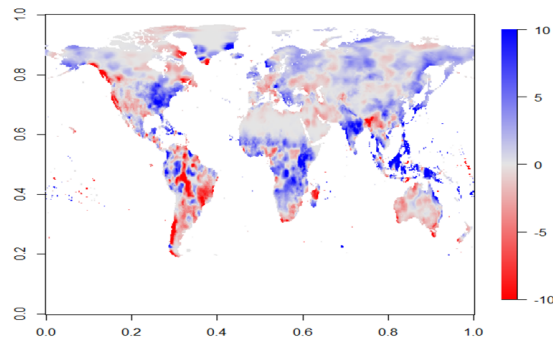


Figura 8. Mapa de tendencias lineales para la precipitación global de 1991-2020. En rojo se observan lugares con una disminución en la lluvia y en azul con un aumento.

4.5.7 Herramientas para pulir los mapas.

Las herramientas que se han explicado hasta este momento son sobre todo exploratorias. Sirven para visualizar los datos geográficos de diferentes maneras y encontrar patrones espaciotemporales de interés. Para ello es fundamental siempre partir de una hipótesis o pregunta con antelación, pues de lo contrario se corre el riesgo de encontrar patrones que pueden no tener una significancia real. No obstante, una vez encontrado el patrón o la respuesta a la pregunta,

4.5. PROCESAMIENTO DE DATOS GLOBALES EN FORMATO NETCDF EN R.91

los mapas que hemos hecho hasta ahora no tienen la calidad suficiente para ser presentados en una publicación (de cualquier tipo). En esta sección se muestra una forma de pulir los mapas, para que tengan un mejor acabado final. Resalto, que existen diversas maneras de hacer lo anterior y que la que muestro a continuación es una de las más sencillas. En general esta estructura funciona únicamente para mapas globales, aunque si se cuenta con una máscara (land-mask) adecuada para seleccionar un país o región, también se pueden usar.

El procedimiento consiste en tres partes. La primera es transformar los datos de matriz a coordenadas. La segunda es generar los detalles del mapa (color, título, leyenda, etc.) y finalmente incluir el contorno de los continentes o los países (según se desee). Estos procedimientos se pueden hacer con diversas librerías, en este caso se dará el ejemplo usando las librerías “sp”, “sf” y “fields”.

Para transformar los datos de una matriz regular a una de coordenadas, primero se deben crear tres vectores: latitud, longitud y la variable que se desea graficar. Posteriormente se deben pegar en un mismo arreglo y después transformarlos a coordenadas (en este caso regulares, siguiendo la misma estructura de celdas de los datos originales). Para el ejemplo dado, para una matriz de medio grado esto se verá algo así:

```
#library(sp)
#library(sf)
#lat<- as.vector(replicate(720,seq(-89.75,89.75,0.5)))
#lon<- as.vector(t(replicate(360,seq(-179.75,179.75,0.5))))
#lim<-as.vector(t(lluvia2[, ,120]))
#rdf<-data.frame(cbind(lon,lat,lim))
#names(rdf) <- c("lon", "lat", "V3")
#points_df <- rdf
#coordinates(points_df) = c("lon", "lat")
#pixel_df <- points_df
#gridded(pixel_df) <- TRUE
#rdf_sp = as(pixel_df, "SpatialGridDataFrame")
```

Lo anterior deriva en un mapa que se ve como la figura 9.

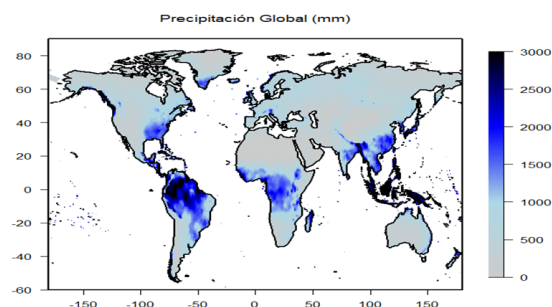


Figura 9. Ejemplo de un mapa detallado para poder ser presentado. El mapa incluye el contorno continental, las coordenadas geográficas (absolutas), la variable, sus unidades y la escala de visualización. En este caso se muestra la precipitación promedio mundial durante 2020.

4.6 Ejemplos y sitios web relevantes para encontrar datos.

Existen un sinnúmero de publicaciones académicas y de divulgación que emplean las técnicas mostradas aquí. Enumerarlas todas sería prácticamente imposible, así que se han elegido ejemplos con base al criterio personal del autor. En primer lugar, hay que destacar todos los informes y reportes del Panel Intergubernamental para el Cambio Climático (IPCC por sus siglas en inglés). Un ejemplo del tipo de mapas generados por el IPCC se puede ver en la figura 10. El IPCC utiliza datos de grandes grupos de modelación internacionales y todos ellos son de acceso abierto al público. Los de la última convocatoria se pueden encontrar en: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=overview>.

4.6. EJEMPLOS Y SITIOS WEB RELEVANTES PARA ENCONTRAR DATOS.93

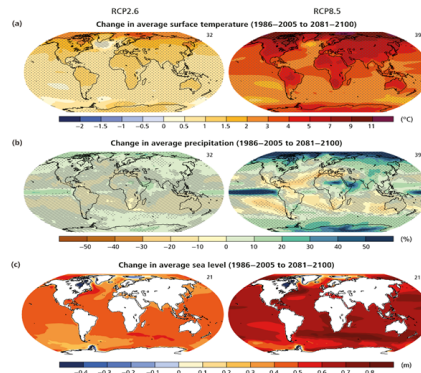


Figura 10. Figura 2.2. del reporte especial sobre cambio climático a futuro del IPCC (IPCC, 2014). IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Otros importantes ejemplos se pueden encontrar en agencias internacionales espaciales, particularmente la NASA, como lo muestra la figura 11. La NASA tiene un enorme repositorio de datos abiertos que se pueden consultar en: <https://gpm.nasa.gov/data>.

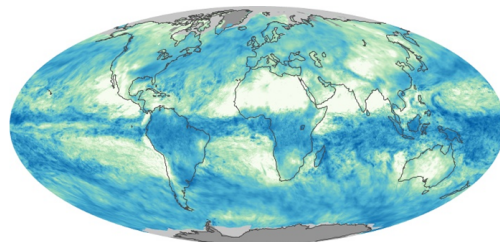


Figura 11. Precipitación promedio mensual de toda la Tierra registrada por el observatorio terrestre de NASA (satélite). Aunque se utiliza otro tipo de proyección para la visualización, los datos originales se manejan con las técnicas empleadas en este capítulo. Los datos se pueden consultar en https://earthobservatory.nasa.gov/global-maps/GPM_3IMERGM.

En la actualidad existen revistas especializadas en la publicación de datos globales, donde se muestran diferentes productos y sus aplicaciones. Un requisito de estas revistas es que los datos publicados sean públicos y de acceso libre. Posiblemente la más importante en el ramo en la actualidad sea Earth System Science Data, que se puede consultar en: <https://www.earth-system-science-data.net/>.

En cada artículo particular, se da una dirección a los datos que se emplean, la mayoría de los cuales están en formato NetCDF. Finalmente, los datos utilizados

como ejemplo durante este capítulo, provienen del Climate Research Unit, de la Universidad de East Anglia en Reino Unido. Los mismos representan el esfuerzo más grande por sintetizar datos de estaciones de monitoreo climáticas a nivel mundial. Todos sus resultados son de libre acceso, en formato NetCDF y se pueden encontrar en: <https://crudata.uea.ac.uk/cru/data/hrg/>.

4.7 Conclusiones.

En este capítulo se muestran algunas técnicas básicas para leer y procesar datos NetCDF en R. Como se mencionó en la introducción de este capítulo esto es útil para procesar datos globales, sumamente utilizados en la ciencia global del sistema de la Tierra. Este tipo de datos puede incluir variables climáticas, de la biósfera, hidrósfera, litósfera e inclusive demográficas. Cuentan además con la ventaja de que la mayoría son datos abiertos y que existe una plétora de información por investigar y estudiar. El ejemplo vertido a lo largo del capítulo es para procesar datos históricos de precipitación global, pero puede ser fácilmente adaptado a patrones regionales o nacionales; así como replicado para otro tipo de variables climáticas.

Hay que hacer notar que el objetivo de este capítulo es presentar las técnicas básicas de procesamiento y mapeo de los datos, pero para un manejo más avanzado -en especial de las técnicas de mapeo- existen muchas herramientas y scripts disponibles en línea. En particular en este trabajo se mostró un solo tipo de proyección y únicamente se mapearon datos terrestres (y no oceánicos). Así mismo se mostró únicamente una forma de representación gráfica (mediante la función “`image.plot`”). Pero es importante hacer notar que esta no es la única manera de visualización. Pese a lo anterior, el uso de los scripts presentados aquí es suficiente para presentar figuras en publicaciones científicas, de divulgación, clases o conferencias.

Chapter 5

Modelación de procesos biológicos con ecuaciones diferenciales ordinarias

Dr. Eugenio Azpeitia (CCM, UNAM Morelia)

Dra. Norma Leticia Abrica Jacinto (CCM, UNAM Morelia)

5.1 Introducción

Si bien la palabra *modelo* tiene múltiples significados y connotaciones, aquí, en un sentido general, diremos que un modelo es una abstracción de la realidad. En este sentido, los modelos, matemáticos y no matemáticos son diversos. Por dar algunos ejemplos de modelos podemos mencionar nuestra comprensión del mundo, por medio de la formulación de ideas, conceptos y teorías; los modelos experimentales, como lo son la rata de laboratorio, *Arabidopsis thaliana* en las plantas o *Escherichia coli* en las bacterias y desde luego, los modelos matemáticos, los cuales serán el tema de este capítulo. A diferencia de otros modelos, los modelos matemáticos tienen la característica de describir los procesos de interés en términos de fórmulas, ecuaciones o reglas que definen y contienen las propiedades del sistema bajo estudio. Como veremos, el expresar un proceso en términos matemáticos tiene fortalezas y limitaciones.

Una de las críticas más comunes a los modelos matemáticos es su naturaleza simplificadora. Vale la pena resaltar que esta es una propiedad de todo modelo, matemático o no. Todo modelo es una abstracción y por lo mismo una simplificación. El uso de un marco conceptual, una teoría o un modelo experimental requiere seleccionar las propiedades que consideramos relevantes de

la realidad para entenderla, interpretarla o estudiarla. En otras palabras, de forma consciente o no, todos estamos haciendo constantemente uso de simplificaciones para interpretar el mundo que nos rodea a través de modelos. El estudio de cualquier proceso por medio de experimentos o trabajo en campo, también requiere enfocarnos y medir u observar un conjunto de propiedades de interés mientras ignoramos otras. Aún más, es opinión de los autores, que si bien las simplificaciones restringen la capacidad explicativa de los modelos, al mismo tiempo, estas son indispensables para integrar y ordenar el conjunto de información que se considera necesaria, conocer sus alcances y para poder comprender e interpretar los análisis y resultados que se obtienen (Alon 2020; Ingalls 2013; Allman 2004). En otras palabras, las simplificaciones son necesarias, y al limitar los modelos también determinan sus alcances y su utilidad.

Los modelos matemáticos tienen la cualidad de que la mayoría de las simplificaciones son explícitas o es posible encontrarlas analizando las fórmulas o ecuaciones del modelo. Por otra parte, en los modelos no matemáticos muchas veces las simplificaciones son menos precisas o implícitas, lo que ocasiona que sus alcances puedan ser difíciles de evaluar. Por ejemplo, un modelo teórico conceptual suele ser impreciso, entre otras cosas, porque muchas veces no detalla magnitudes o cantidades, y usa conceptos que no tienen una definición unívoca. En el caso de un modelo experimental, éste tendrá variación entre los individuos experimentales. Parte de la variación suele ser imposible de controlar y su efecto o importancia es muchas veces difícil de evaluar. Los modelos matemáticos solventan varias de estas complicaciones. Al formalizar un proceso en un conjunto de reglas matemáticas es posible incluir de forma cualitativa y cuantitativa las propiedades de interés. También es posible definir de forma consistente al sistema, lo que permite reproducir siempre los mismos resultados (aunque la interpretación de los resultados pueda variar de individuo a individuo). Finalmente, en los modelos matemáticos también es posible controlar la variabilidad del mismo o “desaparecerla” si fuera necesario. Por otra parte, la formalización en reglas también ayuda a clarificar los aspectos y las suposiciones que se están considerando. En otras palabras la formalización matemática de un fenómeno, de forma implícita “obliga” a evidenciar las suposiciones e hipótesis que sustentan al modelo. Todo esto, hace que los alcances de los modelos matemáticos sean, en general, mucho más claros comparados a otro tipo de modelos (ver Broadlan 2015; Miramontes 2004 para una discusión más amplia de las ventajas de los modelos matemáticos).

Existen diferentes tipos de modelos matemáticos. En este capítulo, nos enfocaremos en modelos matemáticos *mecanísticos*, los cuales distinguiremos de los modelos matemáticos *estadísticos* algunos de los cuales se ven en otra parte del libro/curso. Mientras que los modelos estadísticos buscan encontrar relaciones entre un conjunto de variables, los modelos mecanísticos tratan de encontrar las variables que pueden explicar un fenómeno. Por ejemplo, podemos generar un modelo estadístico que predice las condiciones ambientales bajo las cuales crece o decrece una población sin entender cómo interactúa dicha población con las condiciones del ambiente. Por otro lado, un modelo mecanístico tratará de

encontrar cómo es que las condiciones del ambiente podrían modificar el comportamiento de la población y ocasionar que crezca o decrezca. Es importante notar que la distinción aquí hecha es meramente pedagógica, ya que en la realidad, tanto modelos estadísticos como mecanísticos tienen múltiples aplicaciones y en muchas ocasiones se utilizan en conjunto.

Los modelos que veremos también tendrán la particularidad de ser *dinámicos*. Es decir, serán modelos que se enfocan en los mecanismos que ocasionan los cambios en un sistema. En sistemas biológicos estos cambios normalmente se dan en el tiempo o en el tiempo y en el espacio. En este capítulo nos enfocaremos únicamente en los cambios temporales (ver Iglesias 2010). Notemos que todo fenómeno biológico es el producto de un conjunto de procesos. Por dar algunos ejemplos, los rasgos fenotípicos son el producto de la evolución, el tamaño de una población depende del crecimiento y decrecimiento de la misma, el estado celular depende de cambios físico, químicos y genéticos. Como todo proceso está dado por una progresión de eventos, todo fenómeno biológico es dinámico. Por este motivo, las herramientas que veremos aquí, aunque se enfocarán en su uso en procesos de dinámica poblacional, pueden ajustarse para estudiar cualquier proceso biológico (Murray 2002; Allman 2004; Otto 2007; Ingalls 2013; Alon 2020).

La finalidad de este capítulo es dar una primera introducción al uso de modelos dinámicos para estudiar el comportamiento de una población o comunidad biológica para estudiantes con una formación en un área que no sea físico-matemática. En particular nos enfocaremos en modelos de ecuaciones diferenciales (ED), las cuales son una de las herramientas más empleadas para la construcción de modelos dinámicos. Con esto en mente, hemos organizado el capítulo de la siguiente manera. Comenzaremos dando un panorama general del tipo de modelos dinámicos existen (sección 2) y qué propiedades podemos estudiar con ellos (sección 3). Posteriormente, en la sección 4 se da una introducción a las ecuaciones diferenciales, qué son, cómo se interpretan y cómo se analizan. En la sección 5 mostraremos cómo construir un modelo dinámico básico de crecimiento poblacional usando ED y cómo analizar algunas de las propiedades fundamentales de estos modelos. En la sección 6, veremos cómo podemos construir modelos más complejos que reproduzcan propiedades conocidas de los sistemas biológicos. Finalmente, en las conclusiones discutimos brevemente los alcances de lo visto en este capítulo y proveemos más referencias para el lector interesado.

5.2 Nociones básicas de modelos dinámicos.

Para describir un sistema dinámico que evoluciona en el tiempo requerimos tres elementos mínimos. Una representación del *tiempo*; una representación del *estado del sistema*; y las *reglas de transformación* que describen cómo evoluciona el sistema en el tiempo. Dependiendo de cómo definamos cada uno de estos

elementos tendremos diferentes tipos de modelo.

El tiempo puede ser discreto o continuo. Una representación discreta del tiempo refleja cambios donde las unidades pueden ir desde fracciones de segundos, como en reacciones químicas, hasta años, como podría ser para una población de árboles que se reproduce en un momento específico del año. El tiempo continuo, por otra parte, nos permitiría representar cambios en cualquier instante. Por ejemplo, cambios en la cantidad de biomasa de una población y la concentración de una molécula sufren cambios de forma constante.

Por otra parte, los estados del sistema suelen ser un vector⁽¹⁾ que contiene el valor de las *variables* que conforman al sistema. Las variables del sistema, representan a las entidades que cambian en el tiempo que nos interesa estudiar. Por ejemplo, en un modelo del crecimiento de una población, tendríamos una variable que podría representar al número de individuos o la densidad de la población, mientras que en un modelo del decaimiento de alguna proteína, la variable representaría el número de proteínas o su concentración. Un modelo puede contener tantas variables como requiera. Para un modelo del crecimiento de una población bastaría con una variable, mientras que un modelo de la interacción entre varias poblaciones (i.e., de una comunidad), se requerirían tantas variables como poblaciones haya. Por este motivo, el vector que representa al estado del sistema contiene un valor por cada variable que hay en el modelo. Las variables también pueden ser discretas o continuas. Por ejemplo, una variable discreta nos hablaría del número de individuos en una población, o el número de moléculas de algún tipo en una célula. Una variable continua la podríamos usar para describir la biomasa de una población o la concentración de alguna molécula en una célula.

¹ Se utiliza el término *vector* para indicar una cantidad que tiene magnitud y dirección. Vea la Sección 8.2 del libro de *Biocalculus Calculus for the Life Sciences* de James Stewart y Day Troy, Cengage Learning (2015) para una definición matemática de vector y algunos ejemplos biológicos.

Finalmente, hay diferentes tipos de reglas de transformación. Aquí las dividiremos en deterministas y estocásticas. Las reglas deterministas producen el mismo resultado al aplicarse a un mismo estado del sistema. En otras palabras, al aplicar una regla determinista a un estado del sistema, esto siempre resultaría en el mismo resultado. Por otra parte, las reglas estocásticas se refiere a transformaciones que pueden producir un resultado diferente a partir del mismo estado del sistema. Para diferenciar entre ambos, podríamos pensar en un modelo donde los individuos de una población se reproducen. Un modelo determinista nos podría decir la proporción de individuos que se reproducen a cada paso de tiempo (e.g., 50% de la población se reproduce cada mes), mientras que el modelo estocástico nos podría mencionar la probabilidad que tiene un individuo de reproducirse (e.g., un individuo se reproduce con 50% de probabilidad cada mes).

Una vez definido si el tiempo y la variable es discreta o continua y si el modelo es determinista o estocástico, suele ser simple definir el tipo de reglas de trans-

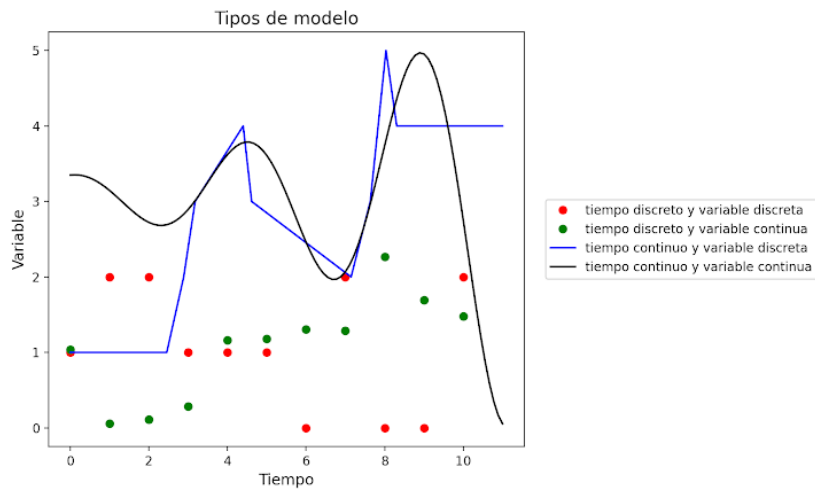


Figure 5.1: Esquema del comportamiento que se observa en modelo producidos combinando tiempo discreto o tiempo continuo con variable discreta o variable continua. En puntos vemos los modelos en tiempo discreto. En ellos podemos observar que todos los cambios se dan en unidades enteras de tiempo (i.e., 0, 1, 2, etc; ver eje vertical). En líneas continuas, vemos modelos en tiempo continuo. En estos modelos podemos observar que los cambios se dan en cualquier momento. Los modelos representados por puntos rojos y por la línea azul, muestran modelos de variable discreta. En ellos vemos que la variable toma valores enteros. El modelo en puntos verdes y línea negra, muestran modelos de variable continua. En ellos la variable toma cualquier valor real.

formación. En el cuadro 1 se observan algunas de las opciones para las reglas de transformación que se pueden emplear para construir un modelo matemático dependiendo del tipo de variable y tipo de tiempo empleado. El tipo de dinámica que produciría cada combinación de tipo de variable y tipo de tiempo está esquematizado en la Figura 1.

Table 5.1: Ejemplos de los tipos de reglas de transformación que se pueden emplear para construir un modelo al definir el tipo de variable y tipo de tiempo de sistema dinámico. * muestra reglas de transformación que cuentan con una versión determinista y una estocástica.

Tiempo	Variable Discreta	Variable Continua
discreto	Modelos Booleanos, <i>multivaluados</i> , cadenas de Markov (estocástico)	Modelos de diferencias*, cadenas de Markov (estocástico).
continuo	Piecewise linear*, Ecuación maestra (estocástico), Modelo de Gillespie (estocástico).	Modelos de ecuaciones diferenciales*, cadenas de Markov (estocástico).

Qué tipo de modelo usar depende de la naturaleza del problema, la pregunta y objetivo del modelo, así como la información disponible. Sin embargo, uno de los modelos matemáticos más empleados y que cuenta con más herramientas de análisis, son los modelos de ED. Es por este motivo que en este capítulo nos enfocaremos en ellos. Para el lector interesado, existen varios libros donde describen y enseñan otras opciones de modelos, incluidas las que se muestran en la Tabla 1 (Murray 2002; Goel y Richter-Dyn 2003; Allman 2004; Otto 2007; Ingalls 2013; Alon 2020).

5.3 Algunas propiedades de los sistemas dinámicos.

Independientemente del tipo de modelo dinámico que se elija, una de las grandes fortalezas es que nos permiten explorar o estudiar propiedades de relevancia para el sistema. Por ejemplo: *trayectorias* de estados, *puntos de equilibrio* (i.e., estados que no cambian al aplicarles las reglas de transformación), *estabilidad* (i.e., puntos de equilibrio a los que se regresa al ser perturbados), entre muchos más. Es importante resaltar, que ha sido posible estudiar muchas de éstas y otras propiedades gracias a los modelos. De hecho, muchas de las propiedades de los sistemas biológicos son difíciles de estudiar sin el uso de modelos ya que aparecen debido a la naturaleza intrínsecamente no lineal de la mayoría de los sistemas biológicos. De forma intuitiva (ver sección 4 para una definición un

poco más precisa de no linealidad), un sistema dinámico lineal es un sistema cuyos cambios son proporcionales a un estímulo. Debido a que los cambios son proporcionales, su estudio es más simple y es más fácil hacer predicciones sobre ellos. En los modelos no lineales el cambio no es proporcional al estímulo. Esto complica su estudio y a su vez ocasiona la aparición de propiedades muy interesantes que no existen en los sistemas lineales. Algunos ejemplos son:

Aparición de patrones: La presencia de una organización en todos los sistemas biológicos. Ejemplos de esta organización van desde ecosistemas hasta la célula. Esta organización, al distinguirse de un acomodo aleatorio se le conoce como patrones. El estudio de los sistemas dinámicos no lineales ha generado algunas de las primeras propuestas de cómo es que estos patrones aparecen. Uno de los ejemplos más famosos es el trabajo de Turing (1952), en el que mostró que patrones similares a los encontrados en la coloración de los animales se puede producir por la interacción y movimiento de dos químicos que interactúan de forma no lineal entre sí. Desde Turing, se ha visto, por una parte, que su modelo tiene la capacidad de explicar otro tipo de patrones (Vandermeer y Perfecto 2017), y a su vez, que otro tipo de modelos con interacciones no lineales pueden producir otro tipo de patrones (e.g., Yonekura 2019, Godin et al., 2020; Jaeger y Reinitz 2006; Kondo y Miura 2010).

Auto organización: Algunos sistemas no lineales, tanto físicos como biológicos, son capaces de auto organizarse. Es decir, de producir alguna estructura o patrón sin necesidad de alguna influencia externa o de un regulador central, sino como resultado de las interacciones propias entre los elementos que conforman el sistema. La auto organización implica que comportamientos globales del sistema están basados en interacciones dadas en niveles inferiores y que las reglas que especifican dichas interacciones requieren únicamente información local. Los ejemplos de auto organización en los sistemas biológicos son bastos, desde agregaciones de bacterias, patrones de coloración de animales, comportamientos de grupos de animales, procesos del desarrollo, entre muchos más (ver Camazine et al., 2001).

Caos: El caos se refiere a sistemas con una alta sensibilidad a las condiciones iniciales. En otras palabras, los sistemas caóticos son sistemas que al iniciar desde dos condiciones tan similares (nunca iguales) como se quiera, terminan produciendo respuestas diferentes. Algo interesante de los sistemas caóticos, es que son deterministas. Es decir, siempre producen los mismos resultados desde la misma condición inicial. Sin embargo, al ser tan sensibles a las condiciones iniciales siguen siendo impredecibles a mediano o largo plazo. Ejemplos de sistemas caóticos hay muchos, siendo probablemente los más conocidos el clima y la economía. Sin embargo, desde los trabajos iniciales de May (1976) se ha discutido mucho sobre la posible presencia del caos en la biología y en los ecosistemas. Actualmente, si bien aún hay debate, ha quedado establecido que varios sistemas biológicos presentan comportamientos que podrían ser caóticos (Vandermeer y Perfecto 2017).

Aquí solo damos algunos ejemplos de propiedades de los sistemas no lineales.

Una revisión más amplia de las propiedades que caracterizan a los sistemas no lineales se encuentra en (Vandermeer y Perfecto 2017; Mitchell 2011).

Los modelos matemáticos son especialmente útiles para el estudio de sistemas no lineales, debido a que la mayoría de sus propiedades no se pueden ver de forma intuitiva. Por este motivo, y dado que este es el objetivo de este capítulo, ahora nos enfocaremos en ED deterministas, las cuales son un formalismo de tiempo y variable continua que permite estudiar a los sistemas no lineales.

5.4 Ecuaciones diferenciales ordinarias.

Las ED están presentes en muchas áreas de las ciencias y la tecnología, y por ello diversos científicos buscan resolverlas. Una ecuación diferencial es una regla que explica/describe cómo cambia una cantidad en función de su propia tasa de cambio. La tasa de cambio se refiere a cuánto cambia una cantidad en relación con otra cantidad. En este caso, como estamos con sistemas dinámicos, la tasa de cambio se refiere a cómo cambia una cantidad (variable) respecto al tiempo (t). Por ejemplo, para una ecuación que describe cómo cambia la temperatura en función del tiempo, la tasa de cambio es entonces la velocidad a la que la temperatura cambia en cada instante. En biología, por ejemplo, las ED se utilizan para modelar el crecimiento de poblaciones, la difusión de sustancias químicas en un tejido, la dinámica de reacciones químicas en sistemas biológicos, entre otros. Como vemos, el concepto de tasa de cambio es importante y este se relaciona con el concepto de derivada, siendo el primero más amplio ya que puede considerarse otras variables diferentes al tiempo.

Para entender los modelos dinámicos que se presentan en este libro, iniciamos con un breve resumen de los conceptos y soluciones básicas de las ED, para más detalles puede consultar Hernández et al. (2008), Brannan y Boyce (2015), Boyce et al. (2021), por mencionar algunas referencias. También se presentarán dos ejemplos muy conocidos en biología matemática, y un ejemplo un poco más sofisticado para el estudio y análisis de la abundancia y riqueza de especies en una comunidad.

5.4.1 La derivada: conceptos e interpretaciones

Geoméricamente, el concepto de derivada está relacionado con la recta tangente a una curva. La *recta tangente* es aquella recta que toca en *un sólo punto* a la curva de una función. Supongamos que $x(t)$ es una función derivable definida en el plano t, x (Purcell et al, 2007; Stewart, 2012). La gráfica de la función es el conjunto de puntos (pares ordenados) de la forma $(t, x(t))$, denotados también como (t, x) con $x = x(t)$. La Figura 2 muestra la recta tangente a la gráfica de la función $x(t)$ en el punto $t = a$.

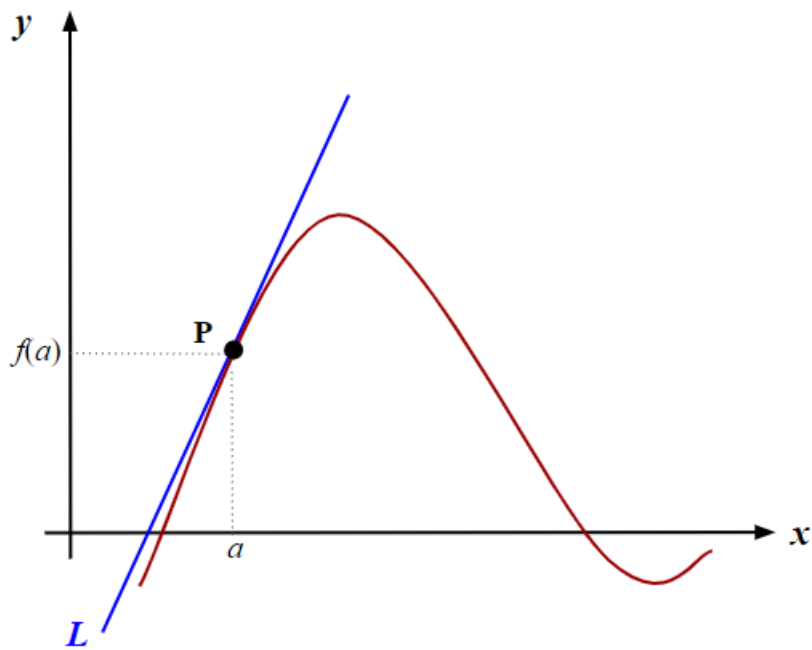


Figure 5.2: En azul la recta tangente L a la gráfica de una función $x(t)$ en el punto $P(a, x(a))$.

La pregunta ahora es, ¿cuál es la ecuación de esa recta? Para dar respuesta es necesario recordar la fórmula de la *ecuación recta-pendiente* para una recta (Purcell et al, 2007; Stewart, 2012). Esta fórmula nos dice que si conocemos un punto (t_0, x_0) y la pendiente m de la recta, entonces la ecuación recta-pendiente es:

$$x - x_0 = m(t - t_0), \quad (1)$$

con (t, x) cualquier punto de la recta distinto de (t_0, x_0) . Por definición, la pendiente m de la recta tangente en el punto (t_0, x_0) se conoce como la derivada de la función en el punto (t_0, x_0) . Por otro lado, si no conocemos el valor de la pendiente, basta con tener las coordenadas del punto (t_0, x_0) y cualquier otro punto de la recta. En general, para aproximar el valor m de la pendiente de la recta L , consideremos las pendientes de las rectas secantes. Una recta secante es aquella recta que corta a una curva en dos puntos. Supongamos que una recta secante pasa por el punto $P(a, x(a))$ y cualquier otro punto $Q(a + h, x(a + h))$ con $h > 0$ como se muestra en la Figura 3.

Una fórmula para encontrar la pendiente de una recta secante m_{sec} es:

$$m_{sec} = (\text{Cambio en el Eje Vertical}) / (\text{Cambio en el Eje Horizontal}). \quad (2)$$

De acuerdo con la Figura 3, la pendiente de la recta secante m_{sec} está en función del cambio en el eje vertical $x(a + h) - x(a)$, y el cambio en el eje horizontal por $(a + h) - a$. Al sustituir tenemos:

$$m_{sec} = (x(a + h) - x(a)) / ((a + h) - a) = (x(a + h) - x(a)) / h. \quad (3)$$

Notemos que si el valor de h se hace cada vez más pequeño, el punto $a + h$ se acercará al punto a . Esto, en términos formales se dice que “cuando h tiende a cero, entonces $a + h$ tiende a a ”, y en símbolos “si $h \rightarrow 0$, entonces $a + h \rightarrow a$ ”. De modo que el punto Q se acercará cada vez más al punto P (Figura 3). Y aquí entra un concepto fundamental del Cálculo diferencial, el **límite de una función** (lim), el cual se relaciona con saber “a qué valor se aproxima la función conforme los valores de entrada se acercan más a cierto número” (Purcell et al, 2007; Stewart, 2012). Entonces, la pendiente de la recta secante se acercará cada vez más a la pendiente de la recta tangente m_{tan} , conforme el valor de h tiende a 0, es decir,

$$m_{tan} = \lim_{h \rightarrow 0} m_{sec} \quad (4)$$

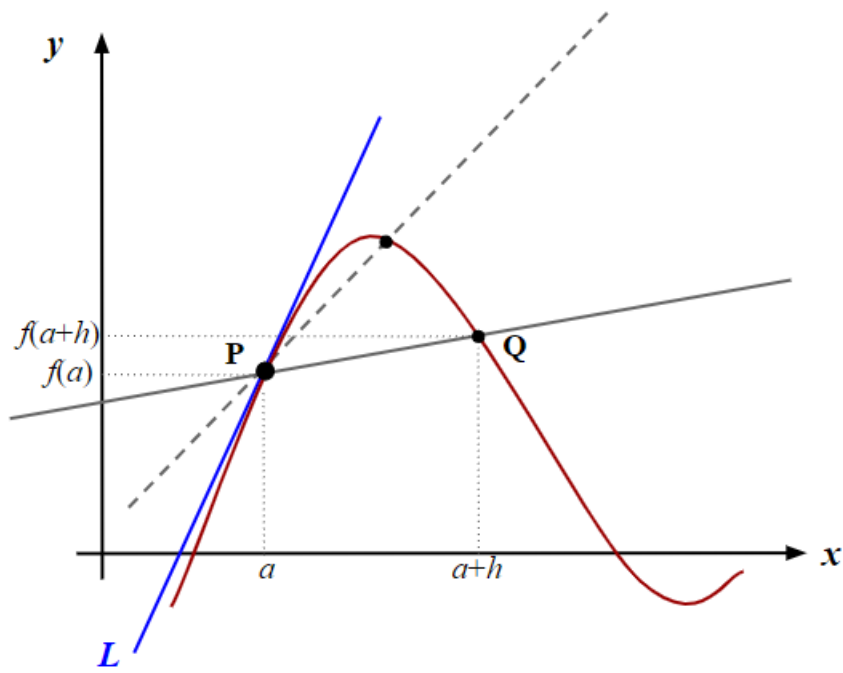


Figure 5.3: Gráfica de una función $x(t)$. La recta azul es la recta tangente en el punto P . Las rectas grises punteadas y continuas son rectas secantes que pasan por P y otro punto de la gráfica de $x(t)$. Notemos que al mover el punto Q hacia P , las rectas secantes (grises) se aproximan a la recta tangente L (azul).

En general la derivada de una función $x = x(t)$ en t está dada por

$$x'(t) = \frac{d}{dt}x(t) = \lim_{h \rightarrow 0} \frac{(x(t+h) - x(t))}{h} \quad (5)$$

siempre que el límite exista. ¡Sí! Habrá ocasiones en que el límite de la función no pueda ser ni calculado u obtengamos valores imprecisos, pero este no es el tema en esta sección. Para más detalle al respecto, el lector puede revisar cualquier libro de cálculo diferencial (Purcell et al, 2007; Stewart, 2012).

Considerando ahora funciones que cambian con el tiempo, es decir, funciones de la forma $f(t)$, la deriva es una *medida de cambio instantáneo de la función en relación con el tiempo*. Es decir, nos va diciendo cómo cambia el valor de la función en un momento específico, lo que se conoce también como *tasa de cambio*. Como la derivada de una función $f'(t)$ es una función, es posible derivar esta función teniendo entonces lo que se conoce como *derivadas de orden superior*:

primer derivada $\rightarrow f'(t) = \frac{d}{dt}f(t)$;

segunda derivada $\rightarrow f''(t) = \frac{d}{dt}f'(t) = \frac{d^2}{dt^2}f(t)$;

tercer derivada $\rightarrow f'''(t) = \frac{d}{dt}f''(t) = \frac{d^3}{dt^3}f(t)$;

...

n -ésima derivada (“ n -ésima derivada”) $\rightarrow f^{(n)}(t) = \frac{d}{dt}f^{(n-1)}(t) = \frac{d^n}{dt^n}f(t)$

Existen reglas para encontrar las derivadas de cualquier función, dejamos al lector revisar las reglas de derivación las cuales puede consultar en Purcell et al. (2007) y Stewart (2012), o en cualquier libro de cálculo diferencial.

Ahora, para retomar el interés de este capítulo exploremos ¿Qué nos dice la derivada en términos biológicos? Es decir, ¿cuál es la relevancia de la derivada para sistemas/fenómenos biológicos? Simplemente nos dice “qué tan rápido” se mueve o cambia la variable del sistema que se está describiendo con respecto del tiempo. Los ejemplos de variables que cambian en el tiempo en biología, y que por tanto son susceptibles de ser descritos con derivadas, son innumerables: los ciclos de vida hacen que los organismos vayan pasando por diferentes etapas del desarrollo a los largo del tiempo, el ritmo circadiano nos mantiene en diferentes estados fisiológicos durante un día, la migración animal modifica la localización de un individuo o de una población a lo largo del año, el crecimiento de una población modifica el número de individuos, por mencionar algunos. Sin embargo, aún nos queda una pregunta por resolver ¿Qué tienen que ver las derivadas con las ED?

5.5 Ecuaciones diferenciales ordinarias: introducción, definiciones y soluciones.

Las ED son ecuaciones que describen relaciones entre la función y sus tasas de cambio. Estas tasas de cambio, como hemos mencionado, son las derivadas.

Es decir, las ED son expresiones que describen el cambio de las variables por medio de una o más derivadas de una función. Por este motivo, son fundamentales para modelar fenómenos en la física, la ingeniería, la biología, las ciencias sociales, y muchas otras disciplinas. De esta forma, los procesos biológicos que se mencionaron al final de la sección anterior, suelen modelarse por medio de derivadas usando ED. Por ejemplo, si queremos conocer cómo se desarrolla una epidemia a través del tiempo, podemos usar las ED para predecir cómo aumenta o disminuye en el tiempo, usando las derivadas que describen cómo cambia el número de individuos sanos, infectados y recuperados de la enfermedad. El modelo SIR de W. O. Kermack y A. G. McKendrick (Kermack y McKendrick, 1927) es uno de los más conocidos, siendo la epidemiología matemática el área de investigación (Fowler, 1997; Diekmann y Heesterbeek, 2000). Otro ejemplo es en el campo de la dinámica de poblaciones, donde los modelos de ED propuestos describen la evolución de tamaño de las poblaciones en términos de las derivadas que describen el crecimiento (y decrecimiento) de las poblaciones. Algunos de los modelos más conocidos es el Lotka-Volterra, donde el cambio en el tamaño de las poblaciones está dado por la interacción entre presas y depredadores lo que afecta la reproducción y mortalidad de los individuos (Lotka 1925; Volterra 1927). Más adelante se describe cómo construir modelos de EDO presentado tres ejemplos conocidos en dinámica de poblaciones: el modelo exponencial, el modelo logístico y el modelo depredador-presa.

Como se ha mencionado, estamos considerando sistemas dinámicos, por lo que las derivadas son de la forma $x'(t) = \frac{d}{dt}x(t)$, para x una función cualquiera, es decir, son cambios en función del tiempo. Existen varias notaciones para referirse a una ecuación diferencial, aquí usaremos la notación $d/dtx(t)$, $(dx(t))/dt$ o dx/dt . En general, las ED se clasifican de acuerdo a su tipo, orden y grado. De acuerdo al tipo se clasifican en ordinarias y parciales. Las *ecuaciones diferenciales ordinarias* (EDO) son ecuaciones que contienen las derivadas de la función con respecto a *una sola variable*, en el caso de este capítulo, el tiempo. Las *ecuaciones diferenciales parciales* (EDP) son ecuaciones que contienen derivadas parciales de la función, es decir, derivadas con respecto a *más de una variable*, por ejemplo tiempo t , temperatura T . Las EDOs son las que estudiaremos en este capítulo. El orden de una ecuación diferencial es la derivada de mayor orden que aparezca en la ecuación. Por ejemplo,

$$\text{EDO de 1er. orden} \rightarrow \frac{d}{dt}x(t) = t + 1$$

$$\text{EDO de 2do. orden} \rightarrow \frac{d^2}{dt^2}x(t) + 5x(t) = \sin(t)$$

$$\text{EDO de 3er. orden} \rightarrow \frac{d^3}{dt^3}x(t) + 5\frac{d^2}{dt^2}x(t) - 3\frac{d}{dt}x(t) = 0$$

$$\text{EDO de 4to. orden} \rightarrow \frac{d^4}{dt^4}x(t) + 5\frac{d^3}{dt^3}x(t) + x(t) = e^t$$

Los términos que multiplican cada elemento de la ecuación diferencial, son llamados *coeficientes de la ecuación diferencial*, en el ejemplo anterior los coeficientes son numéricos, llamados también constantes: 1, 5, -5 y -3 . Finalmente, el **grado** de una ED se refiere al mayor exponente que tenga la derivada de mayor orden, y en este caso se clasifican en lineales y no lineales. Las ED *lineales* son

aquellas en las que la función y cualquiera de sus derivadas aparecen elevadas a la potencia 1; se distinguen porque los coeficientes de la función y de sus derivadas dependen sólo del tiempo t , o bien son constantes. La forma general de una *EDO lineal* es:

$$a_n(t) \frac{d^n x(t)}{dt^n} + a_{n-1}(t) \frac{d^{n-1} x(t)}{dt^{n-1}} + \dots + a_2(t) \frac{d^2 x(t)}{dt^2} + a_1(t) \frac{dx(t)}{dt} + a_0(t)x(t) = g(t) \quad (6)$$

donde los términos $a_n(t) \dots a_0(t), g(t)$ son funciones de la variable t . Las EDNL son todas aquellas que no son lineales. Por ejemplo,

$$\text{EDO no lineal} \rightarrow \left[\frac{d^2}{dt^2} x(t) \right]^3 - \frac{d}{dt} x(t) = 3t$$

$$\text{EDO lineal} \rightarrow \frac{d^2}{dt^2} x(t) + 2 \frac{d}{dt} x(t) + 3x^2(t) = 0$$

Una vez que tenemos un sistema de ED que describen el cambio en las variables de interés, lo que se busca es resolverlas. Pero, ¿qué significa resolver una ecuación diferencial? En términos matemáticos, es encontrar una ecuación que satisfaga la ecuación diferencial. En otras palabras, es encontrar una ecuación cuya derivada produce la ecuación diferencial. Es decir, una ecuación que al sustituirla por su derivada en la ecuación diferencial se cumple la igualdad. Por ejemplo, la ecuación diferencial $dx(t)/dt = t+1$, tiene como solución la ecuación $x(t) = t^2/2 + t$, ya que su derivada $x'(t) = t+1$ por lo que al derivar la función encontrada $x(t)$ se cumple la ecuación diferencial:

$$\frac{d}{dt} x(t) = \frac{d}{dt} \left(\frac{t^2}{2} + t \right) = \frac{d}{dt} \left(\frac{t^2}{2} \right) + \frac{d}{dt} (t) = \frac{2t}{2} + 1 = t + 1 \quad (7)$$

Pero, ¿qué significa resolver una ecuación diferencial en términos biológicos? En términos biológicos, una ecuación diferencial nos dice cómo cambia una variable, y al resolverla, su solución nos dice cuál es su valor a lo largo del tiempo. Por ejemplo, una ecuación diferencial para describir el crecimiento poblacional, digamos $dx(t)/dt$, nos diría cuánto cambia el tamaño de una población debido a la muerte y nacimiento de nuevos individuos, mientras que su solución $x(t)$ nos diría cuál es el tamaño de la población en tiempo. Imaginemos que tenemos una población con cinco individuos, donde uno muere y dos nacen, la ED nos diría que el cambio en la población sería de un individuo, ya que nació uno más de los que murieron, mientras que su solución nos mostraría que su tamaño pasaría de cinco a seis individuos. La solución ayudaría a predecir, por ejemplo, cuándo la población alcanzará su capacidad máxima (sí es que así sucede), cómo se comportará en diferentes momentos de tiempo o si en algún momento la población desaparece. En el caso de la propagación de una enfermedad, la solución de la ED puede mostrar cómo el número de individuos infectados cambia con el tiempo, lo cual es fundamental para tomar

decisiones para la intervención y control de la enfermedad. Como vemos, la interpretación de la solución de una ED dependerá del contexto biológico y del fenómeno que se está modelando. La solución proporciona información valiosa sobre el cambio de las variables biológicas a lo largo del tiempo, lo cual conlleva a un mejor entendimiento y predicción de los procesos biológicos.

Entonces, ¿cómo encontramos la solución de una ED? En realidad, existen una infinidad de ED y hasta ahora, no hay una única forma de resolver todas las ecuaciones imaginables. Por ello, los primeros cursos de ED estudian métodos de solución de casos especiales (ecuaciones separables, homogéneas, exactas, lineales, etc.) los cuales, a través de conceptos y herramientas básicas del Cálculo es posible determinar tales soluciones (Boyce 1990; Zill 2002; Nagle et al, 2005; Boyce et al, 2021). Los métodos de solución de ED también se pueden clasificar en soluciones *analíticas* y soluciones *numéricas*.

Una *solución analítica de una ED* es definir de forma explícita la ecuación que satisfaga la ecuación diferencial, es decir, encontrar la función que cumpla con la ecuación diferencial. En el ejemplo anterior, vimos que $x(t) = t^2/2 + t$ es solución de la ED $dx(t)/dt = t + 1$. Dejamos al lector el ejercicio de comprobar que $f(t) = t^2/2 + t + 5$, $g(t) = t^2/2 + t - 3$ y $h(t) = t^2/2 + t + c$ son también soluciones de la ecuación diferencial, con c cualquier número real. La solución de la forma $x_{g(t)} = t^2/2 + t + c$ es llamada *solución general de la ecuación diferencial*.

Son muchas las ED que no tienen una solución analítica, en particular, las ED con solución analítica son “simples” en su forma. Los cursos de ED se dedican especialmente a encontrar las soluciones a estas ecuaciones. Sin embargo, sabemos que la naturaleza no se comporta de forma simple, de modo que si deseamos representar un sistema biológico lo más realista posible por medio de EDs, es casi seguro que esta ED no tendrá una solución analítica, ¿y qué hacemos entonces? Recurrir a un método numérico para encontrar la solución numérica de la ED.

La solución numérica de ED se basa en métodos numéricos. Un método numérico es una técnica para aproximar la solución de ED mediante cálculos y algoritmos. Con métodos numéricos obtenemos un conjunto de valores que representa el cambio en los valores de la variable, mientras que con la solución analítica se obtiene una expresión algebraica representando combinaciones de números, variables y operaciones matemáticas (una ecuación).

Uno de los métodos más conocidos y simples, utilizados en biología y otras áreas de la ciencia, es el método de *Euler* y el cual será el empleado en este capítulo. El método de Euler se basa en la siguiente aproximación (Ingalls 2013): supongamos que tenemos una ecuación diferencial de la forma $(dx(t))/dt = f(x(t))$; la derivada $(dx(t))/dt$ se puede aproximar por una diferencia cocientes (ver sección 4.1.):

$$\frac{d}{dt}x(t) \approx \frac{x(t+h) - x(t)}{h} \quad (9)$$

para h muy pequeño. Sustituyendo con $\frac{d}{dt}x(t) = f(x(t))$, tenemos

$$\frac{x(t+h) - x(t)}{h} \approx \frac{d}{dt}x(t) = f(x(t)) \quad (10)$$

Considerando la igualdad, obtenemos una fórmula que se puede usar para calcular (aproximar) el valor de $x(t+h)$ dado el valor de $x(t)$:

$$\frac{x(t+h) - x(t)}{h} \approx \frac{d}{dt}x(t) = f(x(t)) \quad (11)$$

El método de Euler consiste en aplicar repetidamente esta fórmula de actualización. Podemos usar el método de Euler de dos maneras: implementado (haciendo) el código correspondiente o utilizando las funciones del software para cálculo científico, por ejemplo R, python, Matlab, Octave, Mathematica. En este texto utilizaremos R.

Para implementar el método de Euler, primero se elige un tamaño de paso h . Con esto, generamos una colección de puntos en la variable tiempo t , de la siguiente manera: supongamos que el tiempo toma valores en un intervalo de la recta, es decir, $t \in [a, b]$. Entonces podemos considerar varios puntos dentro de este intervalo de la siguiente forma:

$$\begin{aligned} t_0 &= a \\ t_1 &= a + h \\ t_2 &= a + 2h \\ &\dots \\ t_n &= a + nh = b \end{aligned} \quad (12)$$

para un número fijo n de pasos. Dado el valor inicial $x(t_0)$, se usa la fórmula 10 para aproximar el valor de $x(t_1), x(t_2), \dots, x(t_n)$:

$$\begin{aligned}
 x(t_0) &= x(a) & (13) \\
 x(t_1) &= x(a+h) = x(a) + h * f(x(a)) \\
 x(t_2) &= x(a+2h) = x(a+h) + h * f(x(a+h)) \\
 x(t_3) &= x(a+3h) = x(a+2h) + h * f(x(a+2h)) \\
 &\dots \\
 x(t_n) &= x(a+(n-1)h) = x(a+(n-1)h) + h * f(x(a+(n-1)h)) = x(b)
 \end{aligned}$$

A continuación describimos el algoritmo y el código computacional para el método de Euler.

5.5.1 Algoritmo del método de Euler

Consideremos una EDO de la forma $\frac{d}{dt}x(t) = f(x(t))$, entonces para $t \in [a, b]$, con x_0 como condición inicial y n número de pasos, el algoritmo de Euler es (modificación del presentado en Mora 2015):

Datos de entrada: $f(x, t), a, b, x_0, n$

Salida: Imprime las aproximaciones (t_i, x_i) , para $i = 0, 1, \dots, n$

- definir el ancho de paso $\rightarrow h = \frac{b-a}{n}$

- asignar $t_0 \rightarrow t_0 = a$

```

for i = 1 to n do
   $x_1 = x_0 + h * f(t_0, x_0)$ 
   $t_0 = a + i * h$ 
   $x_0 = x_i$ 
  print( $(t_0, x_0)$ )
end for

```

Esta es la implementación en lenguaje R.

5.5.1.1 Ejemplo 1.

Consideremos la ecuación diferencial $\frac{d}{dt}x(t) = 1 + x^2 + t^3$ en el intervalo $[1, 2]$, con $x(1) = -4$ y $n = 100$ pasos, hacemos lo siguiente:

1. Definir la función, llamemosla *ejem01.R* y escribimos el script:

```
ejem01 <- function(t,x) {
  f <- 1 + x^2 + t^3$
  return(f)
}
```

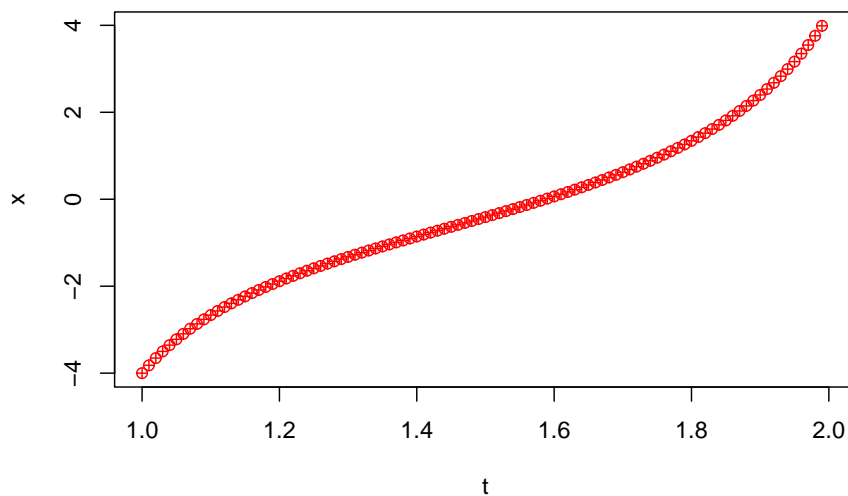
2. LLamar a la función: `source("~/ejem01.R")`
3. Usando el código del método de Euler (`euler.R`) y las condiciones, ejecutar la línea `euler(ejem01, 1, 2, -4, 100)`.

La Figura 4 muestra el resultado de la simulación.

```
euler = function(f, a, b, x0, n) {
#  paso h:
  h = (b-a)/n
# Datos igualmente espaciados: a=t0, a+h, ... a+nh=b
  t0 = a
  t = seq(t0, t0+n*h, by = h) # n+1 datos
  x = rep(NA, times=(n+1)) # n+1 datos
  x[1]=x0
  for(i in 2:n) x[i]= x[i-1]+h*f(t[i-1], x[i-1])
  #print(cbind(t,x)) # print
  plot(t,x, pch=10, col="red") # gráfica
}

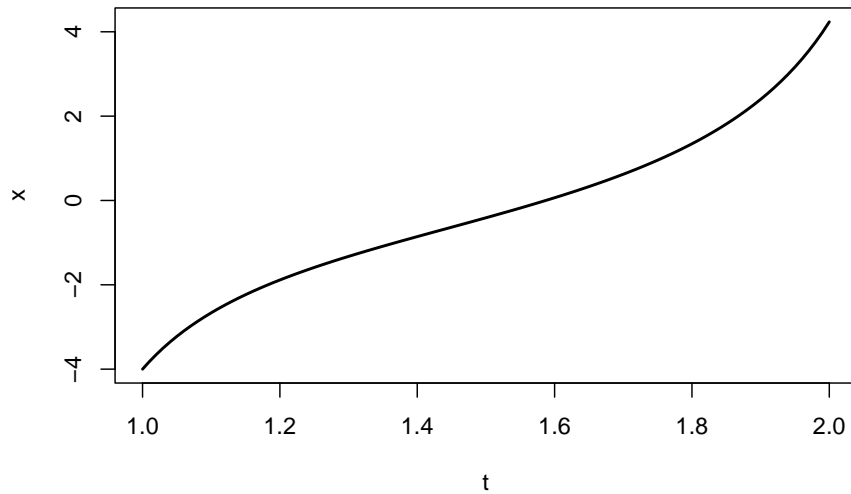
ejem01 <- function(t,x) {
  f <- 1 + x^2 + t^3
  return(f)
}

euler(f=ejem01, a=1, b=2, x0=-4, n=100)
```

Otra forma de encontrar la solución numérica es usando los métodos o funciones de R. Para este mismo ejemplo, ejecutamos las siguiente líneas de comando (ver Figura 5):

```
ejem01ode <- function(t,x,parms) {
  f <- 1 + x^2 + t^3
  list(c(f))
}
times <- seq(from = 1, to = 2, by = 0.01)
Xini <- -4
out=ode(times=times, y=Xini, func=ejem01ode, parms=NULL,
        method=rkMethod("euler"))
matplot(x = out[,1], y = out[,-1], type = "l", lwd = 2, xlab = "t",
        ylab = "x")
```



Elaborar los códigos computacionales o usar las funciones predefinidas de los lenguajes de programación (R, Matlab, Python, etc.), dependerá de la expertis del científico. Usar los lenguajes de programación supone el conocimiento, uso y manejo de las librerías y funciones predefinidas respectivas. De modo que los usuarios simplemente llaman a estas funciones con los parámetros adecuados, para resolver problemas específicos. Por otro lado, elaborar los códigos, supone entender el método (cómo funciona, cuándo es apropiado utilizarlo, limitaciones, requisitos) y la implementación computacional del método, es decir, traducir el algoritmo matemático del método a código de programación. La Figura 6 muestra las soluciones numéricas de la ecuación diferencial $dx(t)/dt = 1 + x^2 + t^3$ de ambos métodos, las cuales como vemos, coinciden.

Una vez que contamos con un algoritmo para resolver numéricamente nuestro sistema de ED diferenciales, podemos comenzar a estudiar algunas propiedades del sistema, como los que se mencionan al inicio de la sección 3. Por ejemplo, tal como se muestra en la Figura 5 y en el código anterior, podríamos estudiar el comportamiento del sistema por una determinada cantidad de tiempo, es decir, ver la *trayectoria* del sistema a partir de un estado inicial. También podríamos usar tiempos largos de simulación para ver si el sistema se estabiliza, es decir, si llega a un *punto de equilibrio*. Finalmente, estudiar la *estabilidad* de un punto de equilibrio simulando trayectorias desde puntos cercanos al punto de equilibrio. Si las trayectorias regresan al punto de equilibrio diríamos que el punto de equilibrio es *estable*. Si las trayectorias divergen del punto de equilibrio diríamos que el punto de equilibrio es *inestable*. En las siguientes secciones iremos viendo cómo podemos usar algunos de estos análisis para entender mejor a nuestro sistema. Sin embargo, nos falta cubrir un último punto, ¿cómo es

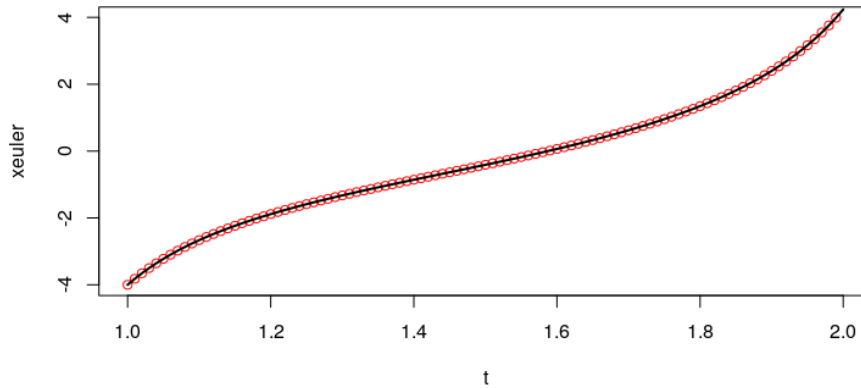


Figure 5.4: Comparación entre la solución numérica por el método de Euler (círculos) y la función ode (línea) de R.

posible obtener las ED para describir al sistema biológico? A continuación, a modo de ejemplo, presentamos una manera de construir modelos de EDOs que modelan el crecimiento poblacional en un ecosistema.

5.6 Construcción de un modelo de EDOs

En la sección anterior mostramos las bases para analizar e interpretar modelos de ED en un contexto biológico. Sin embargo, hasta el momento no hemos entrado en el proceso de construcción del modelo. Es decir, como pasar de la información disponible sobre algún proceso biológico a un sistema de ecuaciones. El proceso de construcción es un paso indispensable, del cual depende el buen funcionamiento del modelo, entendiéndose por esto que contenga de forma apropiada la información deseada para tener el alcance y capacidad para responder a las preguntas planteadas, lo que permite hacer una correcta interpretación de los resultados que se obtengan. Es por esto que realizar una construcción correcta es fundamental. Debido a que no existe una receta de construcción infalible, en esta sección veremos algunos ejemplos de construcción, haciendo especial énfasis en algunos aspectos que suelen estar presentes en la construcción de la gran mayoría de los modelos. Si bien los aspectos aquí descritos pueden servir de guía para construir un modelo, cabe mencionar que durante la construcción de un modelo pueden encontrarse particularidades de cada fenómeno, y muchos de los pasos aquí descritos tampoco requieren seguir siempre el mismo orden. Por estos motivos, la construcción apropiada de un modelo se consigue con la práctica.

Esta sección está organizada de la siguiente forma. Comenzaremos mostrando los elementos básicos de un modelo y cómo podemos identificarlos definiendo la pregunta o el problema bajo estudio. Después mostraremos cómo construir las ecuaciones de un par de modelos a partir de la información básica. Una vez hecho esto, discutiremos cuáles son las suposiciones de cada modelo. Finalmente usaremos los aspectos descritos para construir algunos de los modelos más estudiados en dinámica de poblaciones.

Elementos básicos de un modelo

Los primeros dos elementos básicos de un modelo que veremos aquí son las *variables* y los *parámetros*. Como dijimos anteriormente, las variables del modelo representan a las entidades que cambian en el tiempo que nos interesa estudiar. Por ejemplo, el número de individuos en una población, la cantidad de recursos en un ecosistema, la masa de un individuo. Notemos que a partir de aquí entenderemos por las variables a las variables dependientes, ya que el tiempo será la variable independiente en todos los casos que veremos, por lo que no es necesario especificarla. Los parámetros son valores constantes dentro de un modelo que afecta la dinámica de las variables. Por ejemplo, la tasa de mortalidad, la tasa de migración o la cantidad de recursos disponibles. Si bien, en un sistema real, el valor de los parámetros podría ser variable, en los modelos se suelen considerar constante para simplificar el análisis del modelo concentrándose en el efecto que tienen valores específicos de los parámetros.

En general, las variables suelen ser representadas con letras minúsculas como x , y o z , mientras que los parámetros suelen representarse con letras griegas, como α , β o γ . En muchos casos, cuando tenemos variables similares se usan subíndices x_1, x_2, \dots, x_n . Por ejemplo, para en una comunidad conformada por varias poblaciones, cada población podría estar identificada por un subíndice. Lo mismo ocurre con parámetros similares, donde los podemos distinguir usando subíndices.

Es importante notar, que desde el principio de la construcción de un modelo generalmente contamos con información sobre las variables y los parámetros, la cual es importante analizar. Por ejemplo, dado que en este capítulo trabajamos con ED, sabemos que tanto el tiempo como las variables serán continuas. Por este motivo, cualquier variable x se encuentra en los reales (i.e., $x \in \mathbb{R}$). Muchas veces, por motivos biológicos, tanto las variables como los parámetros solo pueden tomar ciertos valores, por lo que están acotados. Es importante verificar que se respeten dichas cotas. Supongamos, por ejemplo, que una variable representa el tamaño de una población. En este caso, no tendría sentido que la variable tomara valores menores a cero, ya que una población con un número negativo de individuos no existe. De forma similar, no tendría sentido que un parámetro que representa la tasa de mortalidad de una población tomara valores negativos, los cuales representarían una mortalidad negativa y hasta el momento la resurrección no forma parte de los comportamientos de las poblaciones reales.

Una vez que conocemos las variables del modelo, así como sus valores posibles, podemos definir cuáles son los estados del sistema. Un estado del sistema es un vector que contiene el valor de las variables del modelo en un tiempo dado. Por ejemplo, en un modelo del crecimiento de una población el estado del sistema estaría representado por un vector con un único valor representado el tamaño de la población. En un modelo de dos poblaciones, el estado del sistema sería un vector con dos valores, uno para cada población. Los estados del sistema nos permitirán estudiar propiedades del modelo, como a las trayectorias, las cuales son secuencias de estados del sistema que se obtienen al aplicar las *reglas de transformación*. Las reglas de transformación son el último elemento básico que veremos. En este capítulo las reglas de transformación serán las EDs, las cuales, como vimos, describirán el cambio en el valor de las variables. Notemos aquí, que los elementos que hemos definido hasta ahora son los mismos que se requieren para un sistema dinámico (ver Sección 2). Es decir, se requiere una representación del *tiempo*; una representación del *estado del sistema* (donde entran las variables); y las *reglas de transformación* (donde están los parámetros y las ED).

La cuestión es, cómo definimos cuales son las variables, cuáles son los parámetros y cuál es la forma de nuestras ecuaciones. Para responder a estas preguntas notemos que los modelos están hechos con la finalidad de responder alguna pregunta. Es decir, rara vez, un modelo es construido por el simple hecho de poderse construir. La pregunta nos ayudará a definir quiénes son las variables, quiénes son los parámetros y cómo son las ecuaciones. Debido a que en este capítulo nos enfocaremos en la dinámica poblacional, las preguntas suelen ser del tipo: ¿Qué ocasiona que el tamaño de una población aumente o disminuya? ¿Qué elementos podrían limitar el crecimiento de una población? ¿Qué efecto tiene la migración en el tamaño de una población? ¿Qué relevancia tienen las interacciones ecológicas en la dinámica de una población? Veamos ahora algunos ejemplos para ver cómo construir las ED.

5.7 Construcción modelo exponencial

Quizás la pregunta más simple en un modelo de dinámica poblacional es, ¿qué ocasiona que el tamaño de una población aumente o disminuya? Aun cuando la pregunta pudiera parecer simple, es una pregunta compleja que ha motivado una enorme cantidad de estudios. Para cualquier especie, el número de individuos cambia con el tiempo como respuesta a la disponibilidad de recursos, la interacción con otras especies, como la competencia y la depredación, también cambia debido a enfermedades y otras circunstancias no previstas, la migración, entre muchas otras cosas. Sin embargo, en sus términos más simples, una población decrece debido a la pérdida de individuos y aumenta debido a la aparición de nuevos individuos. Si no consideramos procesos como la migración, la principal razón de pérdida de individuos sería la muerte, mientras que el aumento

sería por el nacimiento de individuos. Es con esto que comenzaremos. Es decir, comenzaremos con un modelo que tendrá una variable que representará el cambio en el tamaño de una sola población debido a la muerte y nacimiento de individuos, donde el tamaño de la población podría representar el número de individuos (abundancia) o la biomasa de la población, entre otras cosas. Debido a que tenemos una sola variable, el cambio en el tamaño de la población estará dado por una ED, la cual constará de dos términos, la muerte de individuos y el nacimiento de individuos. Es decir, tendremos una ED de la forma:

$$\frac{d}{dt}x(t) = \text{nacimientos} - \text{muertes} \quad (14)$$

En esta ecuación $x(t)$ representa el tamaño de la población en el tiempo t , por lo que la ecuación diferencial representa el cambio en el tamaño de la población, el cual estará dado por la muerte y los nacimientos. Esto lo podemos expresar de la siguiente manera:

- a) En todo momento nace una cierta cantidad de individuos, la cual es proporcional al tamaño de la población $x(t)$. Esta cantidad es $bx(t)$ con $b > 0$, donde b es la tasa de nacimiento. La tasa de nacimientos debe ser multiplicada por el tamaño de la población ya que no tendría sentido que nacieran individuos si la población está extinta (aunque podrían aparecer nuevos individuos si incluyamos procesos como la migración), y esperaríamos que en una población más grande más individuos se reprodujeran. }
- b) En todo momento también muere una cierta cantidad de individuos, la cual también es proporcional al tamaño de la población $x(t)$. Esta cantidad es $dx(t)$ con $d > 0$, donde d es la tasa de muerte. Por motivos similares al caso anterior, la tasa de muerte es multiplicada por el tamaño de la población.

Una vez definida la forma de expresar la muerte y nacimientos, la ecuación diferencial que representa el cambio en el tamaño de la población es:

$$\frac{d}{dt}x(t) = bx(t) - dx(t) \quad (15)$$

la cual, factorizando se transforma en:

$$\frac{d}{dt}x(t) = (b - d)x(t) \quad (16)$$

Finalmente, haciendo $r = b - d$:

$$\frac{d}{dt}x(t) = rx(t) \quad (17)$$

donde r es llamada *tasa de crecimiento intrínseco* porque es la tasa a la que se esperaría que creciera la población si no hubiera ninguna limitante (ver abajo suposiciones del modelo). Por lo tanto, el modelo de crecimiento exponencial está dado por una ED lineal en la cual la razón de cambio en el tamaño de la población es *proporcional* al valor actual de su población.

Aun sin entrar en demasiado detalle, es posible hacer algunos análisis simples del modelo. Podemos notar que si r es negativa (respect. positiva), el valor de la ED será negativo (respect. positivo) (ya que $x \geq 0$), lo que significa que el tamaño de la población decae (respect. aumenta). Esto tiene sentido si notamos que r es la tasa de nacimiento menos la tasa de muerte. Si la tasa de nacimiento es mayor que la tasa de muerte, la población crece, y viceversa.

Por ejemplo, consideremos una población con $r = 1.07$ y la ecuación diferencial de valor inicial y $x(0) = 500$, de la forma $\frac{d}{dt}x(t) = 1.07x(t)$. Usando la función ode de R, y considerando que la condición inicial en el intervalo $t \in [0, 5]$ en tamaño de paso de 0.01, la Figura 7 muestra los valores obtenidos para el tiempo entre 0 y 1. Notemos que en el tiempo inicial (0) se tienen 500 individuos, y en el tiempo 0.9 se tienen 1303.0885 individuos. La Figura 8 muestra el comportamiento de la población en el intervalo de tiempo $[0, 5]$, es decir, la solución de la ecuación diferencial.

time	1	time	1	time	1	time	1
[1,] 0.00	500.0000	[21,] 0.20	618.6077	[61,] 0.60	946.9040	[81,] 0.80	1171.5242
[2,] 0.01	505.3500	[22,] 0.21	625.2268	[62,] 0.61	957.0359	[82,] 0.81	1184.0595
[3,] 0.02	510.7572	[23,] 0.22	631.9167	[63,] 0.62	967.2761	[83,] 0.82	1196.7289
[4,] 0.03	516.2223	[24,] 0.23	638.6782	[64,] 0.63	977.6260	[84,] 0.83	1209.5339
[5,] 0.04	521.7459	[25,] 0.24	645.5121	[65,] 0.64	988.0866	[85,] 0.84	1222.4759
[6,] 0.05	527.3286	[26,] 0.25	652.4191	[66,] 0.65	998.6591	[86,] 0.85	1235.5564
[7,] 0.06	532.9710	[27,] 0.26	659.4000	[67,] 0.66	1009.3448	[87,] 0.86	1248.7769
[8,] 0.07	538.6738	[28,] 0.27	666.4555	[68,] 0.67	1020.1448	[88,] 0.87	1262.1388
[9,] 0.08	544.4376	[29,] 0.28	673.5866	[69,] 0.68	1031.0603	[89,] 0.88	1275.6437
[10,] 0.09	550.2631	[30,] 0.29	680.7940	[70,] 0.69	1042.0927	[90,] 0.89	1289.2931
[11,] 0.10	556.1509	[31,] 0.30	688.0785	[71,] 0.70	1053.2431	[91,] 0.90	1303.0885
[12,] 0.11	562.1017	[32,] 0.31	695.4409	[72,] 0.71	1064.5128	[92,] 0.91	1317.0316
[13,] 0.12	568.1162	[33,] 0.32	702.8821	[73,] 0.72	1075.9030	[93,] 0.92	1331.1238
[14,] 0.13	574.1951	[34,] 0.33	710.4030	[74,] 0.73	1087.4152	[94,] 0.93	1345.3668
[15,] 0.14	580.3390	[35,] 0.34	718.0043	[75,] 0.74	1099.0505	[95,] 0.94	1359.7623
[16,] 0.15	586.5486	[36,] 0.35	725.6869	[76,] 0.75	1110.8104	[96,] 0.95	1374.3117
[17,] 0.16	592.8247	[37,] 0.36	733.4518	[77,] 0.76	1122.6961	[97,] 0.96	1389.0168
[18,] 0.17	599.1679	[38,] 0.37	741.2997	[78,] 0.77	1134.7089	[98,] 0.97	1403.8793
[19,] 0.18	605.5790	[39,] 0.38	749.2316	[79,] 0.78	1146.8503	[99,] 0.98	1418.9008
[20,] 0.19	612.0587	[40,] 0.39	757.2484	[80,] 0.79	1159.1216	[100,] 0.99	1434.0831
						[101,] 1.00	1449.4278

Figure 5.5: Datos de los puntos obtenidos con el método ode de R., para la ecuación diferencial $dx(t)/dt = 1.07x(t)$, con valor inicial $x(0) = 500$, en el intervalo $(0, 1)$

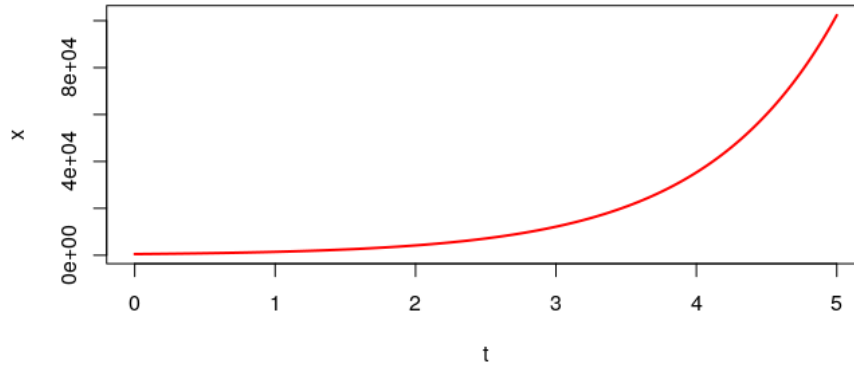


Figure 5.6: Solución de la ED $dx(t)/dt = 1.07x(t)$ con valor inicial $x(0) = 500$, en el intervalo de 0, 5. Este comportamiento es llamado *crecimiento exponencial*.

Quizás llame la atención que la población pueda tomar valores no enteros. Esto tiene sentido si consideramos que el tamaño de la población podría representar densidades poblacionales, cantidad de biomasa, o que el modelo es solo una aproximación.

Por último, vale la pena notar que dado que r es constante y $x \geq 0$, el signo de la ED nunca cambiará. Es decir, que una población que crece, seguirá creciendo hasta el infinito, mientras que una población que decrece inevitablemente se extinguirá. Esto se debe a que este modelo no considera ninguna restricción en la cantidad de recursos (e.g., espacio y alimento). Es decir, con una tasa de crecimiento positiva ($r > 0$) la población crecerá indefinidamente, como se ilustra en el ejemplo de la Figura 8. ¿Pero es esto realista desde el punto de vista biológico y ecológico? La respuesta es *no*. Siempre habrá eventos o factores que interfieren en el crecimiento o disminución de una población, como variables ambientales o físicas, entre otras. En general, el modelo de crecimiento exponencial tiene algunas limitaciones y desventajas en un contexto realista, por ejemplo:

- a) El supuesto de *recursos infinitos* (i.e., sin restricción en la cantidad de recursos). En la realidad, los recursos son limitados y con el tiempo, la población puede superar la capacidad de su entorno para sustentarla ocasionando, posiblemente, la competencia por recursos.
- b) *No considera factores externos*. Esto es claro ya que este modelo ignora factores naturales externos tales como desastres naturales, cambios climáticos o enfermedades.

- c) *No considera interacciones entre los individuos de la población.* Por ejemplo, en los ecosistemas naturales, las poblaciones interactúan de diferentes formas, como la depredación y la competencia. Este modelo no considera este tipo de interacciones.
- d) *No considera la densidad poblacional.* En condiciones reales, conforme aumenta el tamaño de una población, la competencia por los recursos es mayor lo que puede ocasionar una disminución en la tasa de crecimiento, y esto no lo considera el modelo (recordemos que para este modelo r es constante).
- e) *No predice el comportamiento a largo plazo.* Es decir, para tiempos “cortos” el modelo puede ser una buena aproximación, pero para tiempos mayores, el tamaño de la población crece desmesuradamente (en el caso de $r > 0$), como en el ejemplo de la Figura 8 en el tiempo $t = 5$ la población tiene un tamaño aproximado de 102,353 individuos cuando sólo empezó con 500 individuos.
- f) *Es determinista.* El comportamiento de la población está completamente determinado por las condiciones iniciales específicas, es decir, siempre se tendrá el mismo comportamiento bajo las mismas condiciones iniciales.
- g) *Sensibilidad a las condiciones iniciales.* Esto significa que pequeñas variaciones en los valores iniciales conducen a comportamientos (resultados) significativamente diferentes a medida que evoluciona el modelo.

Aun con estas limitaciones, el modelo exponencial ayuda a comprender la manera en cómo crece (decrece) una población y las implicaciones de los supuestos. Ahora veamos cómo podríamos incluir un poco más de realismo en el modelo.

5.8 Construcción modelo logístico

Una de las primeras cosas que se pueden modificar en el modelo para evitar crecimientos ilimitados es asumir que la cantidad de recursos en el medio son limitados. Si los recursos son limitados, esperaríamos que el crecimiento de la población se redujera conforme su tamaño aumenta ya que habría mayor competencia por los recursos disponibles. Aun más, si los recursos fueran insuficientes para la población, esperaríamos que esta decayera. Es decir, buscamos modificar la ecuación $dx(t)/dt$ de forma tal, que el crecimiento deje de ser proporcional a la población y en su lugar varíe con respecto al tamaño de la población. Una forma de hacer esto, es considerando la capacidad de carga K , la cual representa el tamaño de la población que el sistema puede mantener y modificando la ecuación del modelo exponencial a:

$$\frac{d}{dt}x(t) = \frac{r(1 - (x(t))/K)}{K} \quad (18)$$

donde r continúa siendo la tasa de crecimiento intrínseco dada por la diferencia entre la tasa de nacimiento y la tasa de mortalidad. A este modelo se le conoce como el *modelo logístico*, el cual ha sido sumamente empleado en el estudio de la dinámica de poblaciones. Detalles matemáticos de esta ecuación, así como de sus bases biológicas pueden ser consultados en otros libros (Allman y Rhodes, 2004; Otto y Day, 2007). Aquí solamente verificaremos que la ecuación cumple con el comportamiento deseado y explicaremos el por qué lo hace.

Para verificar que la ecuación cumple con el comportamiento esperado, notemos que la ecuación está formada por una multiplicación entre los términos $rx(t)$ y $1 - (x(t))/K$. El primer término es igual al modelo exponencial descrito en la sección anterior, mientras que el segundo término es la modificación a la ecuación que contiene la capacidad de carga. Para este análisis asumiremos que $r > 0$, lo que significa que la tasa de nacimiento es mayor a la de muerte y en principio, la población tenderá a crecer. Debido que $x(t) \geq 0$, también $rx(t) \geq 0$, por lo que basta conocer el signo de $1 - (x(t))/K$ para conocer el signo de la ecuación también. Por este motivo bastará con verificar que ocurre con los valores de $1 - (x(t))/K$ para poder describir cualitativamente el comportamiento de la ecuación.

Cuando $x(t)$ es muy pequeño (i.e., $x(t)$ es cercano a 0), el valor de la ecuación diferencial se aproxima a $dx(t)/dt \approx rx(t)$ debido a que $[1 - (x(t))/K] \approx 1$ (Figura 9). En otras palabras, cuando la población es pequeña no hay competencia por los recursos disponibles por lo que la cantidad de recursos no afectan de forma importante su crecimiento y la población crece a la misma tasa que crecía la población en el modelo exponencial, donde los recursos son ilimitados. Mientras el valor de $x(t)$ aumenta, si el tamaño de la población se mantiene por debajo de la capacidad de carga (i.e., $x(t) < K$), sabemos que la ecuación tendrá un valor positivo ya que tanto $rx(t)$ como $[1 - (x(t))/K] > 0$. Sin embargo, el valor va disminuyendo hasta que $x(t) = K$, donde $[1 - (x(t))/K] = 0$ (Figura 9). Cuando $x(t) = K$, tanto la ecuación diferencial como $1 - (x(t))/K$ son igual a cero, lo que significa que la población no cambia de tamaño. Finalmente, cuando $x(t) > K$, el valor de la ecuación diferencial se vuelve negativo ya que $[1 - (x(t))/K] < 0$ (Figura 9), lo que indicaría que el tamaño de la población disminuye. Es decir, cuando la población supera la capacidad de carga del sistema, su tamaño disminuye. De esta forma, podemos observar que el modelo tendrá el comportamiento esperado, con un crecimiento elevado cuando no hay competencia por recursos en poblaciones pequeñas, pero conforme el tamaño de la población aumente, el crecimiento irá disminuyendo hasta superar la capacidad de carga del sistema, donde la población tenderá a disminuir su tamaño.

Al igual que en el caso anterior, aún sin muchas herramientas se pueden estudiar algunas propiedades del modelo. Por ejemplo, la ecuación diferencial valdrá cero

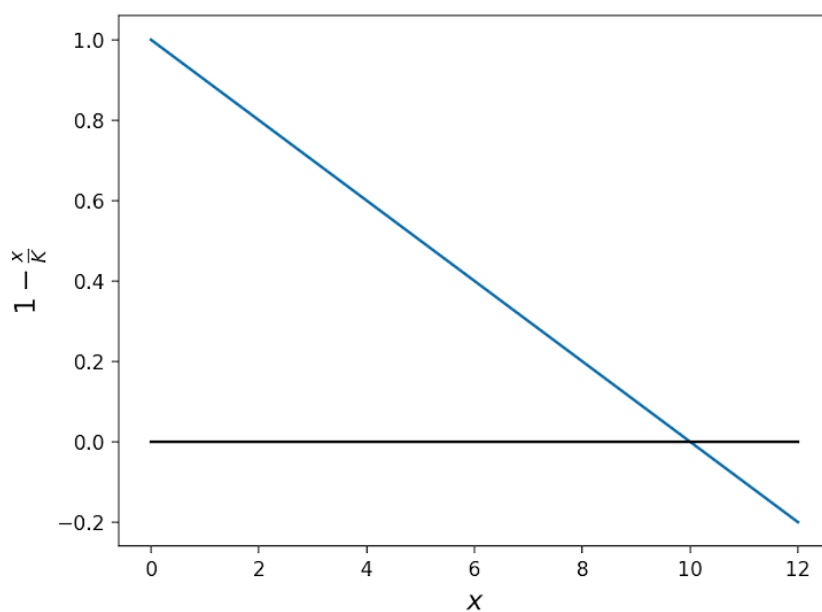
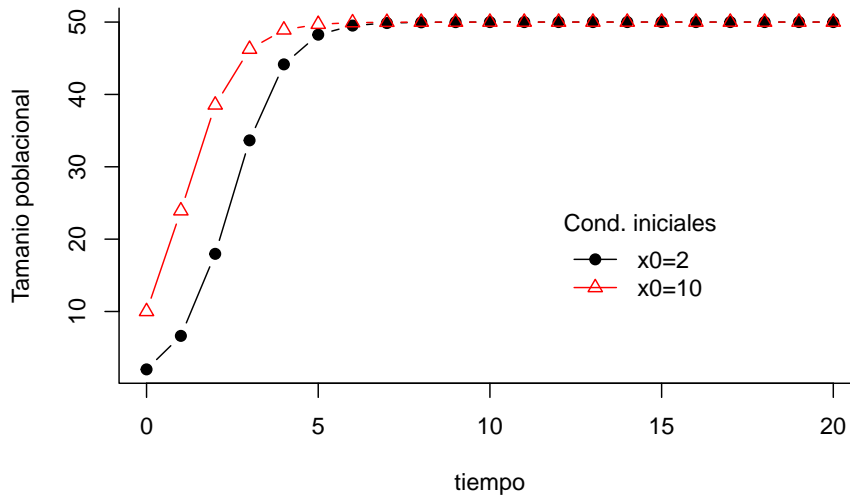


Figure 5.7: Evaluación del valor de $1 - x/K$ con $K = 10$. Podemos observar que conforme x tiende a 0 el valor de $1 - x/K$ se acerca a 1, por lo que la ecuación $rx(1 - x/K)$ tendería a rx . Cuando x tiende a 10, $1 - x/K$ tiende a 0, por lo que $rx(1 - x/K)$ tendería a 0. Finalmente, cuando x es mayor a 10, $1 - x/K$ es negativo, haciendo que $rx(1 - x/K)$ sea negativo también.

cuando cualquiera de los dos términos que se multiplican es igual a cero (i.e., $dx(t)/dt = 0$ si $rx(t) = 0$ o cuando $(1 - (x(t))/K) = 0$). Esto ocurre cuando $x(t) = 0$ y cuando $x(t) = K$. Debido a que el tamaño de la población no cambia cuando la ecuación diferencial es cero, este resultado significa que el tamaño de la población será estable cuando la población está extinta, o cuando alcanza la capacidad de carga del sistema.

De hecho, podemos verificar estos resultados por medio de simulaciones. Por ejemplo, el modelo logístico con $r = 1.3, K = 50$ desde $t = 0$ hasta $t = 20$ con la condición inicial $x_0 = 2.5$ y $x_1 = 10$ podemos ver que el sistema se estabiliza en el valor de la capacidad de carga (valor 50 en el eje vertical). Esto nos indica que este es un punto de equilibrio, el cual es estable (Figura 10). Le recomendamos al lector realizar otras simulaciones para otros valores de la condición inicial, sin comenzar con una población igual a 0.



Si bien este modelo elimina la suposición de recursos ilimitados, lo que hace que las poblaciones ya no crezcan de forma indeterminada, aún contiene muchas otras suposiciones. Por ejemplo,

- Linealidad en la capacidad de carga.* Este modelo asume que la capacidad de carga es constante y lineal (K). En situaciones reales (biológicas/ecológicas) esta capacidad puede variar debido a condiciones ambientales.
- Homogeneidad del entorno (ambiente).* Este modelo supone un entorno (ambiente) sin variaciones espaciales en los recursos o en la capacidad de

carga. En los ecosistemas naturales los recursos son de diversos tipos, es decir, heterogéneos; y la capacidad de carga puede cambiar según el área física del ecosistema.

- c) *No considera factores externos.* Como en el modelo exponencial, este modelo ignora factores naturales externos tales como desastres naturales, cambios climáticos o enfermedades.
- d) *Es determinista.* Similar al modelo exponencial, siempre se tendrá el mismo comportamiento bajo las mismas condiciones iniciales.
- e) *No predice el comportamiento a largo plazo.* De igual manera, este modelo no considera la aleatoriedad (y complejidad) de los sistemas naturales.
- f) *Sensibilidad a las condiciones iniciales.* Similar al modelo exponencial, pequeñas variaciones en los valores iniciales conducen a comportamientos diferentes.

Aun con estas limitaciones, es posible modificar el modelo para darle mayor realismo. Aquí veremos cómo incluir un último detalle, como podemos describir algunas interacciones entre poblaciones distintas. Sin embargo, para el lector interesado, como incluir otros detalles como migración, movimiento y varias más se pueden encontrar en Allman y Rhodes (2004), Otto y Day (2007), Haefner (2005, cap. 13).

5.9 Construcción del modelo depredador-presa

Este modelo, a diferencia del anterior, considera la interacción entre especies, siendo un sistema de tipo Lotka-Volterra (Lotka 1925; Volterra 1927), es decir, la interacción es de tipo depredación. Supongamos que $x(t)$ denota el tamaño de la población de la especie presa y $y(t)$ el tamaño de la población de la especie depredadora. Este modelo tiene varios supuestos:

- 1) En ausencia de depredadores, la población de presas crece. Es decir, la ausencia de depredadores significa que $y(t) = 0$, y por lo tanto la población de presas crece, es decir, $x(t) > 0$.
- 2) En ausencia de presas, la población de depredadores desaparece. Es decir, la ausencia de presas significa que $x(t) = 0$, y por lo tanto la población de depredadores desaparece debido a que no tienen recursos que consumir, es decir, $y(t) < 0$.
- 3) Cuando ambas especies interactúan implica que la presa es o puede ser devorada por el depredador, y el depredador entonces puede o no consumir

alguna especie presa. De modo, que en la interacción de ambas $(x(t) \cdot y(t))$, para la presa es perjudicial $(x(t) \cdot y(t) < 0)$ mientras que para el depredador es benéfico $(x(t) \cdot y(t) > 0)$.

Como antes, $dx(t)/dt$ y $dy(t)/dt$ denotan el cambio en el tamaño de la población de las especies correspondientes y considerando los puntos anteriores, un modelo que representa el cambio en el tamaño de las poblaciones de estas especies que interaccionan es:

$$\begin{aligned} \frac{d}{dt}x(t) &= r_1x(t) - \alpha x(t)y(t) \\ \frac{d}{dt}y(t) &= -r_2y(t) + \beta x(t)y(t) \end{aligned} \quad (19)$$

donde r_1 es la tasa de crecimiento de la presa, r_2 la tasa de muerte del depredador, α es el impacto en la presa por la interacción de ambas, y β el impacto en el depredador por la misma interacción.

Por ejemplo, consideremos el modelo depredador-presa con $r_1 = 2$, $\alpha = 2$, $r_2 = 1$, $\beta = 1$ y con poblaciones iniciales $x(0) = 1$ y $y(0) = 3$. Usando el método de Euler en el intervalo de tiempo $0 \leq t \leq 20$ con el tamaño de paso de 0.01, la Figura 11 muestra la solución aproximada. Como vemos, este modelo da una representación simplificada, pero efectiva, de las interacciones entre las poblaciones de presas y depredadores, lo que permite entender los patrones de crecimiento/decrecimiento básicos de estas relaciones.

Ahora que ya contamos con los modelos y una primera herramienta para analizarlos (análisis numéricos), veamos con un poco más de detalle usando un ejemplo un poco más elaborado, cómo podemos usar lo que hemos aprendido.

5.10 Sección 6: Modelado de Abundancia y Riqueza (MAR)

Los ejemplos vistos en las secciones anteriores, tratan de describir algunas propiedades básicas que regulan el crecimiento de las poblaciones. Sin embargo, esas poblaciones suelen encontrarse en un ecosistema mucho más complejo, donde coexisten e interactúan múltiples especies. Dichas interacciones suelen ser descritas por medio de redes. Estas redes describen diferentes tipos de interacciones, por ejemplo, entre polinizadores y plantas, entre herbívoros y plantas, interacciones de cooperación o comensalismo, entre otras. Uno de los tipos de redes ecológicas más estudiadas son las redes alimentarias. Las redes alimentarias son fundamentales en ecología debido a que a través de ellas es

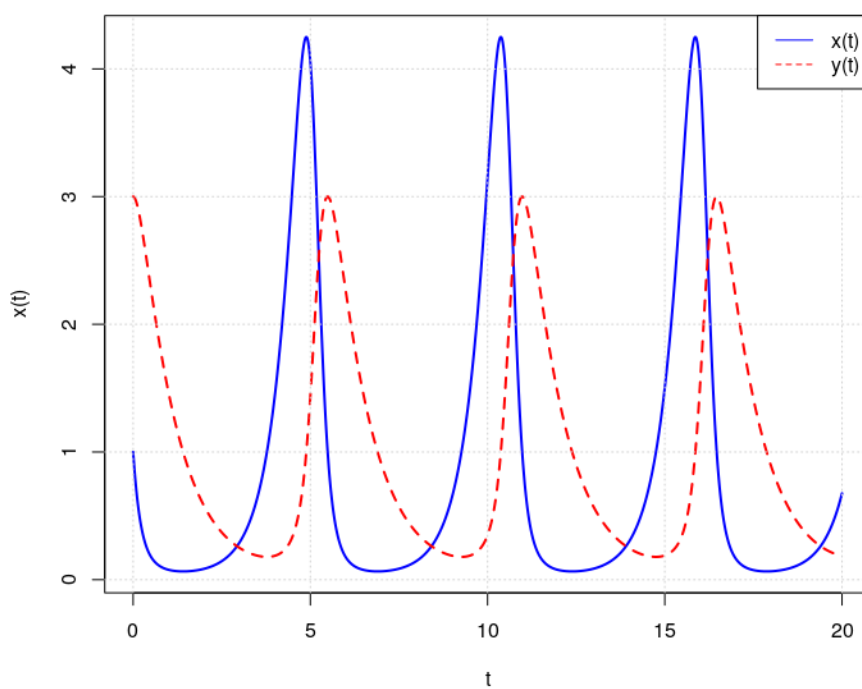


Figure 5.8: Solución del modelo depredador presa para $r_1 = 2$, $\alpha = 2$, $r_2 = 1$, $\beta = 1$ y con poblaciones iniciales $x(0) = 1$ y $y(0) = 3$.

posible describir y analizar la dirección y flujo de cada población en un ecosistema, por ejemplo, estudiar la abundancia y riqueza de las especies en una población.

En esta sección presentamos un modelo que describe, por medio de redes alimentarias, la abundancia de las especies en función de varios parámetros. La solución de este sistema será a través de simulaciones numéricas como las vistas en las secciones anteriores. Iniciamos con algunos conceptos útiles para comprender el Modelo de Abundancia y Riqueza (MAR).

5.10.1 6.1. Conceptos básicos

Las redes alimentarias (RAs) son definidas como aquellas que describen las relaciones de interacción trófica (quién come a quién, depredador-presa) entre diversas poblaciones; son una representación estructural de la comunidad.

A través de las RAs es posible estudiar propiedades importantes de la comunidad como su dinámica poblacional o su biodiversidad. La forma más sencilla de representar una RA es a través un grafo dirigido o digrafo (Bondy y Murty, 1976), donde los nodos representan a las especies y los enlaces la relación trófica. Dado que se está describiendo una interacción de tipo depredación, es posible distinguir a las especies como presas y depredadores, siendo entonces la RA una red que describe la transferencia de nutrientes entre las especies de distintas poblaciones biológicas.

Los detalles sobre cómo construir dichas redes están fuera del alcance de este trabajo. Por este motivo, aquí sólo introduciremos los conceptos necesarios para los fines de este capítulo. Sin embargo, el lector interesado puede consultar otras referencias sobre el tema (e.g., Cohen et al. 1990; Williams y Martínez 2000; Abrica-Jacinto et al. 2023). Aquí utilizaremos el modelo de red alimentaria llamado de nicho, el cual produce redes con propiedades observadas empíricamente (Williams y Matríguez, 2008).

5.10.1.1 Matriz de interacciones (tróficas)

Uno de los conceptos principales que se requiere para construir una red alimentarias es el de matriz de interacciones tróficas. La matriz de interacciones tróficas, A , describe las relaciones tróficas entre las especies y se define como sigue: $A = (a_{i,j})$ es una matriz de tamaño $S \times S$, con S el número de especies, con $a_{i,j} = 1$ si y sólo si la especie j consume a la especie i ; y $a_{i,j} = 0$ si la especie j no consume a la especie i . En el caso de que $a_{i,i} = 1$, describe una relación de canibalismo. A los números $a_{i,j}$ s de la matriz se le conocen como elementos de la matriz.

Por medio de esta matriz de interacciones es posible visualizar una red alimentaria. Por ejemplo la Figura 12 muestra una red con 8 especies que interaccionan

de manera trófica, mostrando cómo se ve en la naturaleza, cómo es la matriz de interacciones y cómo se representa por medio de un digrafo. Notemos que la dirección del enlace, por ejemplo, de i_1 a i_2 , indica el flujo de energía de la especie i_1 a la especie i_2 .



Figure 5.9: Representación de una red alimentaria en la naturaleza, por medio de la matriz de interacciones y cómo un digrafo; respectivamente.

5.10.1.2 Categorización de especies en redes tróficas

La representación matricial de una RA contribuye a entender la estructura y composición de la red alimentaria. Por ejemplo, es posible distinguir a través de ciertos indicadores (Abrica-Jacinto et al., en revisión) diversos tipos de especies. Una categorización sencilla para distinguir especies en relaciones tróficas se basa en distinguir a las especies como basales (B), intermedias (I) y superiores o tops (T), definidas de la siguiente manera:

- a) Especie Basal (B): son especies que sólo son consumidas por otras especies, y por lo tanto, no consumen a otras. En el digrafo se ven como los nodos que sólo tienen enlaces de salida. Los elementos en la matriz para una especie basal i son: $a_{ji} = 0$ para toda especie j y $a_{ij} = 1$ para al menos una especie j .
- b) Especie Intermedia (I): son especies que son consumidas por otras especies, y además consumen a otras especies, es decir, “las que comen y son comidas”. En el digrafo son nodos que tienen enlaces de entrada y salida; por lo tanto $a_{ji} = 1$ y $a_{ik} = 1$ para al menos una especie j y k .
- c) Especie Tops o superiores (T): son especies que sólo consumen a otras especies, es decir, “sólo comen”. En el digrafo son los nodos que sólo tienen enlaces de entrada y ninguno de salida; por lo tanto $a_{ji} = 1$ para al menos una especie j y $a_{ik} = 0$ para toda especie k .

En el ejemplo de la Figura 12, el conjunto de especies basales es $[i_1, i_2]$ (triángulos verdes), el conjunto de especies intermedias es $[i_3, i_4, i_5, i_6, i_7]$ (rectángulos azules), y el conjunto de especies tops es i_8 (círculo rojo).

Una de las propiedades de interés en este tipo de redes es la conectividad, mencionada en Martínez (1991, 1993) al decir que la conectividad (dirigida) es un patrón robusto a la agregación trófica y generalmente comparable entre redes alimentarias. La conectividad nos dice que tan densa es la red, es decir, qué tan conectados están los nodos (especies) en la red. Una conectividad alta, nos hablaría de una red altamente conectada. La conectividad C se calcula como todas las interacciones reales (i.e, el número de links L del digrafo) dividido por todas las interacciones posibles (S^2 si se considera canibalismo, $S(S-1)$ en caso contrario; con S el número de especies). Por ejemplo, en la Figura 12, la conectividad equivale a $L/S(S-1) = 9/8(7) \approx 0.16$. Por lo tanto, la conectividad es calculada como:

$$C = L/S^2, \text{ con canibalismo;} \\ C = L/S(S-1), \text{ sin canibalismo.}$$

5.11 Modelo de nicho

Existen diversos algoritmos (modelos) que generan redes alimentarias a través de matrices de interacciones tróficas. El propósito de éstos modelos es reproducir las interacciones tróficas de forma simplificada. En la literatura se han utilizado modelos como: el aleatorio, el de cascada y el de nicho (Cohen et al., 1990; Williams y Martínez, 2000; 2008). Cada uno de ellos se definen utilizando el número de especies (S) y la conectividad de la red (C), siendo diferenciados por cómo definen la probabilidad de interacción (p) entre las especies. Por ejemplo, el modelo aleatorio define la probabilidad de interacción igual a la conectividad de la red ($p = C$), sin considerar otras propiedades o características biológicas de las especies o de la interacción. El modelo de cascada (Cohen et al., 1990) sugiere un “orden jerárquico” en la alimentación (interacción) de las especies: la especie j en esta jerarquía o cascada puede alimentarse de cualquier otra especie i con $i < j$, es decir, no puede alimentarse de especies “superiores” a ella. Este orden no significa que j se alimente de i , solo que j puede alimentarse de i , y para ello se define la probabilidad de interacción como $p = 2CS/((S-1))$. Como vemos, este modelo caracteriza la interacción entre las especies por medio de un orden (aleatorio) sin otras propiedades biológicas, sin embargo, en este sentido es mejor que el aleatorio. El modelo de nicho, por su lado, considera propiedades de las especies de mayor interés. Aquí, cada especie i se caracteriza por un valor de nicho (n_i : número aleatorio entre (0,1)), luego, las especies consumen a aquellas especies que están dentro de su intervalo de comida, es decir, dentro de su dieta. Para definir el valor del intervalo se hace uso de la función de probabilidad beta $p(x) = \beta(1-x)^{\beta-1}$ con $\alpha = 1$ y $\beta = \alpha/2C - \alpha$, de modo que $r_i = x \cdot n_i$ con $0 \leq x \leq 1$ y x en $p(x)$, para cada especie i . Por

ejemplo, la Figura 13 muestra dos redes alimentarias con los mismos valores de nicho, $S = 10$, $C = 0.2$ y para diferentes valores de $x = 0.2513$ (izquierda) y 0.8233 (derecha). Como vemos, si x es grande la conectividad de la red es mayor, presentándose casi en todas las especies el canibalismo.

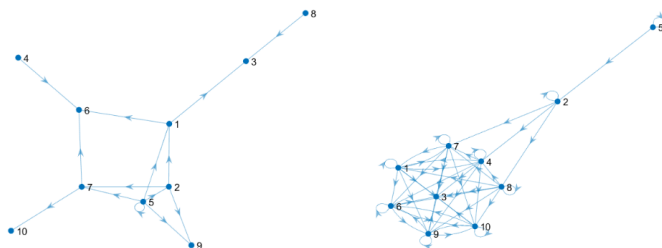


Figure 5.10: Redes construidas con el modelo de nicho con $S = 10$, $C = 0.2$ y $x = 0.2513$ (izquierda) y $x = 0.8233$ (derecha).

5.11.1 Abundancia y Riqueza

La riqueza se define como el número de especies diferentes y no extintas en la red alimentaria. La abundancia cuantifica la cantidad de elementos de cada especie. En ese sentido, la riqueza es una propiedad general en la red, y la abundancia una propiedad local, ya que corresponden a cada especie. Por ejemplo, consideremos la red de la Figura 12. Aquí se tiene una red con $S = 8$ especies diferentes y supongamos que la variable $x_i \in [0, 1]$ representa la abundancia de la especie i . Podemos expresar la abundancia de la RA como un vector de la forma:

$$X = (x_1, x_2, \dots, x_S)$$

Supongamos que $x_1 = 0.5, x_2 = 0.4, x_{(3-7)} = 0.3$ y $x_8 = 0.6$. En este caso la riqueza $Riq = 8$, en general, $0 \leq Riq \leq S$. Supongamos además que a través de ciertas interacciones durante un periodo de tiempo, se llega a que $x_1(t) = 0.25$, $x_2(t) = 0.5$, $x_{(3-8)} = 0$, entonces $Riq(t) = 2$.

5.11.2 Tasa de crecimiento intrínseco de las especies

En cualquier especie, el número de individuos cambia con el tiempo como respuesta a diversos factores (natalidad, mortalidad, competencia, depredación, clima, temperatura, etc.). La tasa de crecimiento intrínseca para una especie i se define como la natalidad menos la mortalidad, y la denotaremos como r_i .

5.11.3 Capacidad de carga

La capacidad de carga (K) es la cantidad máxima de individuos en una población que el hábitat puede soportar, es decir, el tamaño de la población que puede existir o que puede soportar el ambiente.

5.12 5.2. El modelo MAR

En ecología, indicadores como la abundancia y riqueza de especies nos dan información sobre la diversidad de especies en un determinado sistema ecológico. La abundancia, recordemos, es definida como el número de individuos de una especie y la riqueza el como el número de especies. El siguiente ejemplo busca explicar, por medio de una ecuación diferencial, cómo la abundancia de una especie cambia en función del tiempo y de las interacciones entre los diversos individuos de diferentes especies. Tales interacciones las definimos utilizando una red alimentaria, la cual es posible describir por medio de una matriz de interacciones $A = (a_{ij})$. Para describir la abundancia de las especies a lo largo del tiempo, se define el siguiente sistema de EDO, no lineal y de 2do orden (Abrica-Jacinto et al, 2023b; en revisión) con $i \in 1, 2, \dots, S$ y S el número total de especies:

$$\frac{d}{dt}x_i(t) = x_i(t)\left[\left(1 - \frac{x_i(t)}{K_i}\right) * (r_i + e_i \sum_{j=1}^S a_{ji}x_j(t)) - \sum_{j=1}^S a_{ij}X_j(t)\right] \quad (20)$$

Para cada especie i se tiene: $x_i \in [0, 1]$ es el valor de la abundancia, $K_i \in [0, 1]$ es la capacidad de carga, r_i denota la tasa de crecimiento y e_i la eficiencia en el consumo. La eficiencia cuantifica el beneficio obtenido de la especie al consumir todas sus presas. Siguiendo el trabajo de Abrica-Jacinto et al (2023b; en revisión) y haciendo una simplificación en los valores de los parámetros, consideremos las siguientes condiciones para cada tipo de especies:

- Especies basales: si i es una especie basal, entonces $r_i = 0.8$, y claramente, $e_i = 0$.

- Especies intermedias: si i es una especie intermedia, entonces $r_i = -0.2$, y $e_i = 0.66$.
- Especies superiores: si i es una especie superior, entonces $r_i = -0.2$, y $e_i = 0.85$.

Supongamos que tenemos una RA descrita por la Figura 14, con $S = 10$, $C = 0.2$, generada con el modelo de nicho. Las especies basales son $\{3,8,10\}$, las especies intermedias son $\{1,2,4,5,6,7\}$ y la especie top es $\{9\}$. Para este experimento supondremos que las tasas de crecimiento intrínseco para las especies basales es del mismo valor; así como para las especies intermedias y tops.

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}_{S \times S}$$

Figure 5.11: Red alimentaria (modelo de nicho) con $S = 10$ especies, conectividad de $C = 0.2$; especies basales (3, 8, 10), especies intermedias (1, 2, 4, 5, 6, 7) y especie top (9).

Con las condiciones mencionadas, consideremos además que la capacidad de carga $K = 1$ es igual para todas las especies y como condición inicial de la abundancia el vector:

$$X(0) = (0.534, 0.808, 0.995, 0.876, 0.035, 0.501, 0.876, 0.995, 0.4321, 0.875)$$

La Figura 15 muestra la evolución en la abundancia para cada una de las especies. El código en R es el siguiente:

```
# arreglo para la matriz de interacciones: red alimentaria
aa = c(0,0,0,0,0,0,1,0,1,0,
```

```

0,0,0,1,1,1,1,0,0,0,
0,1,0,0,1,1,0,0,0,0,
1,0,0,0,1,0,1,0,0,0,
0,0,0,0,0,0,0,0,1,0,
0,0,0,0,0,0,0,0,1,0,
0,0,0,0,0,0,0,0,1,0,
0,1,0,0,1,1,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,
0,0,0,0,1,1,0,0,0,0)
#
# la matriz de interacciones
A <- matrix(aa, nrow=10, ncol=10,byrow=FALSE)
#
# abundancia inicial de la poblacion
X0 <- c(0.534,0.808,0.995,0.876,0.035,0.501,0.876,0.995,0.4321,0.875)
#
# tasa de crecimiento de las especies
r <- c(-0.02,-0.02,0.8,-0.02,-0.02,-0.02,-0.02,0.8,-0.02,0.8)
# numero de especies
S <- nrow(A)
# eficiencia de consumo
e <- c(0.66, 0.66, 0.00, 0.66, 0.66, 0.66, 0.66, 0.00, 0.85, 0.00)
# capacidad de carga
K <- rep(1, S)
#
#
#####
# Funcion del sistema
iter_Euler <- function(X0,K,R,A,Ti,Tf,num){

  # numero de especies
  S <- dim(A)[1]

  ## tamaño de paso
  h <- (Tf-Ti)/(num)

  # Matriz que contiene los cambios en las abundancias
  # para todas las especies a lo largo del tiempo
  evol_cont <- matrix(data=0, nrow=num+1,S)

  # corresponde al valor inicial de las especies
  evol_cont[1,] = X0

  ### Iteraciones
  for (t in 2:(num+1)){

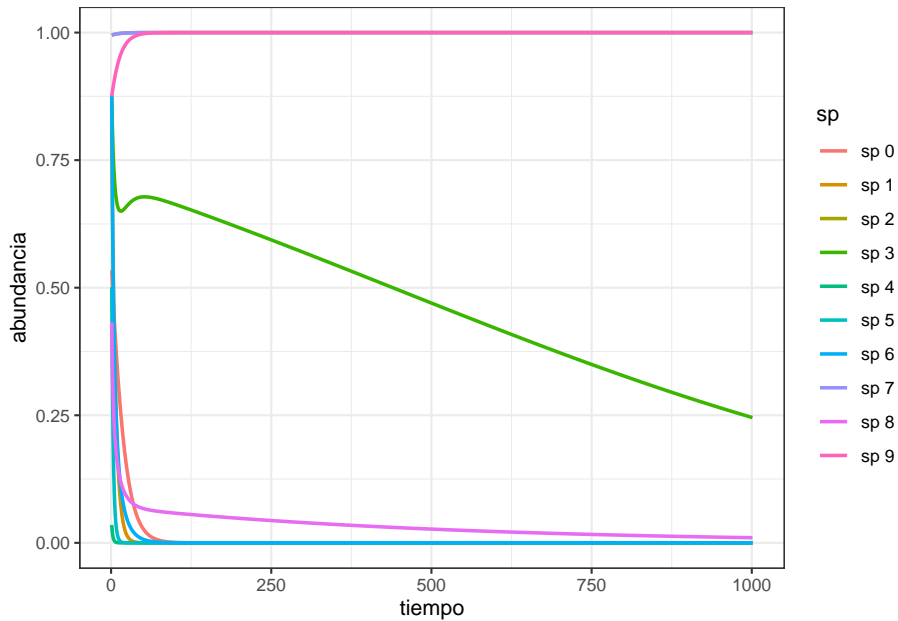
```

```

for (i in 1:S){
  cap <- 1-(evol_cont[t-1,i]/K[i])
  pos <- sum(A[,i]*evol_cont[t-1,])
  wa <- 1 * A[i,]
  neg <- sum(wa * evol_cont[t-1,])
  f1 <- evol_cont[t-1,i] * (cap * (R[i] + e[i]*pos) - neg)
  evol_cont[t,i] <- evol_cont[t-1,i] + h*f1
}

}
# Devuelve una matriz con las especies en las columnas
# y el número de iteraciones en las filas
evol_cont
}
#
# tiempo para la dinamica y numero de pasos
Ti <- 0
Tf <- 100
num <- 1000
# la solucion del sistema de ecuaciones
sol = iter_Euler(X0,K,r,A,Ti,Tf,num)
#
# # paquetes:
# install.packages("tidyr")
# install.packages("ggplot2")
# # librerias:
library(tidyr)
library(ggplot2)
#
# Grafica de evolucion temporal del sistema
data <- data.frame(c(1:nrow(sol)),sol)
colnames(data) <- c("tiempo", paste(rep("sp",10),c(0:9)))
data.graficar <- pivot_longer(data,cols=2:11,names_to="sp",
                             values_to="abundancia")
ggplot(data.graficar, aes(x = tiempo, y = abundancia))+
  geom_line(aes(color = sp),linewidth=0.85) +
  theme_bw()

```



5.13 Conclusiones/Discusión, posibles cursos lecturas posteriores

En este capítulo hemos cubierto algunas bases para construir, simular y analizar modelos de ED de sistemas o procesos biológicos. En particular vimos que es un sistema dinámico y como se define. Mostramos diferentes variantes que existen para representarlo como un modelo matemático. Debido a que cualquier proceso biológico es en cierto sentido un sistema dinámico, este es susceptible de ser modelado matemáticamente. Posteriormente nos enfocamos en los modelos continuos deterministas, los cuales usan ED. Para esto, explicamos qué son las ED enfocándonos en su relación con las derivadas, su interpretación biológica y describiendo brevemente cómo podemos usar y resolver las ED. Dentro de los posibles métodos de resolución de ED, aquí exploramos los métodos numéricos empleando el algoritmo de Euler, el cual nos permite estudiar el comportamiento del sistema a partir de una condición inicial. A partir de allí nos centramos en la construcción de modelos de dinámica poblacional usando ED y resolviéndolos con el algoritmo de Euler. Finalmente mostramos cómo podemos usar estas herramientas para estudiar un modelo más complejo. Si bien todos los temas de este capítulo se cubren con mayor amplitud en muchos otros libros, este capítulo tiene la particularidad de estar pensado para un público con una formación profesional en un área distinta a las matemáticas, pero que requiere comenzar a emplear modelos matemáticos para su investigación. En ese sentido es laxo y se queda en un nivel superficial en varios de los temas cubiertos, tratando de

llevar al lector de forma eficiente a tener una comprensión de los modelos que le permita profundizar posteriormente si fuese de su interés.

5.13.1 Agradecimientos

Agradecemos a la Dra. Veronica Zepeda por su apoyo en revisión de los códigos escritos en esta sección.

5.14 Bibliografía

Abrica-Jacinto NL, Benítez M, Azpeitia E*. (2023). Matrices de interacciones: conceptos y modelos de redes alimentarias. En revisión.

Abrica-Jacinto, L., Veronica Zepeda, Mariana Benítez y Eugenio Azpeitia. (2023b). Population dynamics in food webs. En revisión.

Allman, E. S., y Rhodes, J. A. (2004). Mathematical models in biology: an introduction. Cambridge University Press.

Alon, U. (2020). An Introduction to Systems Biology. Chapman & Hall/CRC Computational Biology Series. CRC Press.

Bondy, J. A., y Murty, U. S. R. (1976). Graph theory with applications (Vol. 290). London: Macmillan.

Boyce, W. E. (1990). ED y problemas con valores en la frontera. España: Limusa.

Boyce, W. E., DiPrima, R. C., y Meade, D. B. (2021). Elementary differential equations and boundary value problems. John Wiley & Sons.

Brannan, J. R., y Boyce, W. E. (2015). Differential equations: An introduction to modern methods and applications. John Wiley & Sons.

Brodland, G. W. (2015). How computational models can help unlock biological systems. In Seminars in cell & developmental biology (Vol. 47, pp. 62-73). Academic Press.

Camazine, S. Deneubourg, J.-L. Franks, N. Theraulaz, G. Bonabeau, E. (2001) Self-Organization in Biological Systems; Princeton University Press: New Jersey.

Cohen, J. E., Briand, F., y Newman, C. M. (1990). Community food webs: data and theory. Biomathematics, Vol 20. Springer, Berlín.

Diekmann, O., y Heesterbeek, J. A. P. (2000). Mathematical epidemiology of infectious diseases: model building, analysis and interpretation (Vol. 5). John Wiley & Sons.

Fowler, A. C. (1997). *Mathematical models in the applied sciences* (Vol. 17). Cambridge University Press.

Godin, C., Golé, C., & Douady, S. (2020). Phyllotaxis as geometric canalization during plant development. *Development*, 147(19), dev165878.

Goel, N.S., Richter-Dyn N. (2003). *Stochastic Models in Biology*. Blackburn Press

Haefner, J. W. (2005). *Modeling Biological Systems: Principles and Applications*. Alemania: Springer.

Hernández, P. A. Q., Serrano, M. D. C. C., y Oliver, E. B. V. (2008). *Métodos de solución de ecuaciones diferenciales y aplicaciones*. Reverté.

Hiebeler, David E. (2018). *R and Matlab*. CRC Press.

Iglesias, P. A. (2010). Spatial modeling . In P. A. Iglesias & B. P. Ingalls (Eds.), *Control Theory and Systems Biology* (pp. 45 – 68). Cambridge, MA : MIT Press .

Ingalls, B. P. (2013). *Mathematical modeling in systems biology: an introduction*. MIT press.

Jaeger, J., & Reinitz, J. (2006). On the dynamic nature of positional information. *BioEssays*, 28(11), 1102-1111.

Kermack, W. O., y McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772), 700-721.

Kondo, S., & Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *science*, 329(5999), 1616-1620.

Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins.

Martinez, N. D. (1991). Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological monographs*, 61(4), 367-392.

Martinez, N. D. (1993). Effects of resolution on food web structure. *Oikos*, 403-412.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(5560), 459-467.

Miramontes, P., & de Biomatemática, G. (2004). *La biología matemática*. Bautista R. *Las matemáticas y su entorno*. México: Siglo XXI, 47-65.

Mitchell M. (2011). *Complexity: A Guided Tour*. Oxford University Press

Mora, Flores Walter. (2015). *Introducción a los métodos numéricos. Implementaciones en R*, 1era ed. Escuela de Matemáticas, Instituto Tecnológico de Costa Rica. *Revista digital Matemática*: <https://tecdigital.tec.ac.cr/servicios/revistamatematica/>

- Murray J. D. (2002). *Mathematical Biology. Interdisciplinary Applied Mathematics*.
- Nagle, R. K., Saff, E. B., Snider, A. D. (2005). *Ecuaciones diferenciales y problemas con valores en la frontera*. España: Pearson Educación.
- Otto, S. P., y Day, T. (2007). *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press.
- Purcell, E. J., Varberg, D. E., Rigdon, S. E. (2007). *Cálculo*. México: Pearson Educación.
- Stewart, J. (2012). *Cálculo de una variable: Trascendentes Tempranas. Séptima Edición*. México: CENGAGE Learning. <https://www.fbioyf.unr.edu.ar/evirtual/pluginfile.php/107533/course/section/2765/calculo-james-stewart-7ed.pdf>
- Turing, A.M. (1952). The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. B*, 237, 37-72.
- Vandermeer, J., & Perfecto, I. (2017). *Ecological complexity and agroecology*. Routledge.
- Volterra, V. (1927). *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi (Vol. 2)*. Società anonima tipografica "Leonardo da Vinci".
- Yonekura, T., Iwamoto, A., Fujita, H., & Sugiyama, M. (2019). Mathematical model studies of the comprehensive generation of major and minor phyllotactic patterns in plants with a predominant focus on orixate phyllotaxis. *PLoS computational biology*, 15(6), e1007044.
- Williams, R. J., y Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, 404(6774), 180-183.
- Williams, R. J., y Martinez, N. D. (2008). Success and its limits among structural models of complex food webs. *Journal of Animal Ecology*, 512-519.
- Zill, D. G., González Pozo, V. (2002). *Ecuaciones diferenciales con aplicaciones de modelado*. Argentina: Thomson Learning.

Chapter 6

Modelos de ecuaciones estructurales (SEM)

Dr. Francisco Mora (IIES UNAM Morelia)

6.1 Presentación

Los modelos de ecuaciones estructurales (SEM por su siglas en inglés) permiten establecer *vínculos de causa y efecto entre múltiples variables*. Una de las “virtudes” del SEM es que se puede evaluar, **como un “todo”**, al conjunto de vínculos entre todas las variables del sistema. Es decir, *permite explorar hipótesis “completas”*, considerando todos los vínculos y el tipo de vínculos especificados.

6.1.1 Partes del modelo SEM.

- 1) *“las variables observadas”*: conjunto de descriptores que se se pueden medir, ya sea como parte de un experimento o como “observación”.
- 2) *“vínculo entre variables”*: conexiones, relaciones entre aquellos pares de variables que me permitan representar o modelar el sistema; estas relaciones son de causa-efecto, i.e. una variable influye o afecta a otra.
- 3) *“variables latentes”*: conjunto de variables compuestas, i.e. conformadas por las variables observadas y que se asume que están influyendo en el sistema, aunque no se midan directamente. Estas variables son uno de los aspectos más relevantes del SEM.

La representación gráfica de los SEM es conceptualmente muy sencilla. Cada variable se representa como una caja (nodo, vértice) y el vínculo entre variables como una flecha. El sentido de la flecha muestra la relación de causa-efecto entre ese par de variables. Por ello es que se pueden representar modelos conceptuales muy complicados. Las variables pueden ser “observadas” o “latentes”, o una combinación de ambas.

Los SEM pueden ser ajustados siguiendo dos estrategias diferentes (Grace et al. 2015): 1) el ajuste de todas las ecuaciones estructurales “al mismo tiempo”, es decir del modelo completo, empleando métodos de varianza-covarianza, y 2) el ajuste por partes, en el que el modelo general se divide en secciones y cada sección se ajusta de manera independiente, para posteriormente “re-agrupar las piezas” y generar el modelo global. A continuación abordaremos esta segunda estrategia de ajuste, empleando para ello un ejemplo sobre las causas de la degradación ambiental.

6.1.2 Caso de estudio

El ejemplo que utilizaremos hace referencia a las causas de la degradación ambiental en nuestro planeta. Para ello emplearemos información de cuatro variables medidas al nivel de país: la población, el producto interno bruto total (PIB), las emisiones de gases de efecto invernadero totales (GEI) y la proporción de energía producida que proviene de fuentes renovables. En particular pondremos a prueba tres diagramas causales diferentes, como se ilustra en la **Figura 1**.

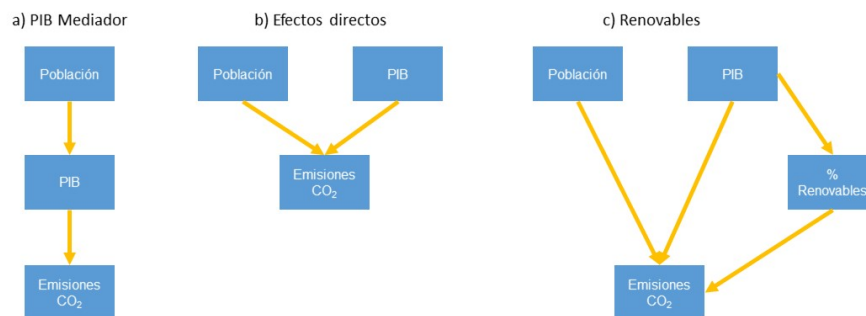


Figure 6.1: Hipótesis de causalidad para explicar las emisiones de gases de efecto invernadero en función del tamaño de la población, el volumen de actividades económicas y la producción de energía limpia a escala de país. a) Hipótesis del PIB como mediador. b) Hipótesis de los efectos directos. c) Hipótesis de las energías renovables.

En general se espera observar los siguientes tipos de asociaciones:

- Las emisiones de gases de efecto invernadero (GEI) se incrementan con el tamaño de la población.
- Las emisiones de GEI se incrementan con el producto interno bruto.
- Las emisiones de GEI se reducen al incrementarse la proporción de energía producida que proviene de fuentes renovables.

6.2 Preliminares

6.2.1 Cargar librerías

Para poder proceder primero cargamos las librerías que vamos a emplear y a leer la tabla de datos

6.2.2 Inspeccionar los datos

Una buena práctica antes de empezar cualquier análisis consiste en revisar los datos para evaluar su calidad en términos del estado de la tabla: si las variables están bien especificadas, qué tanta información faltante hay y cómo está codificada, entre otros. Dos formas de hacerlo es obteniendo resúmenes de la tabla, o generando una inspección gráfica de la misma. Usemos la primera opción empleando la función `summary`, la cual permite generar un resumen de los datos por variable:

```
##      Country                               Income_Region Total_Population
## Length:217                High_income      :80      Min.   :1.028e+04
## Class :character          Low_income       :34      1st Qu.:7.457e+05
## Mode  :character          Lower_middle_income:47      Median :6.038e+06
##                                     Upper_middle_income:56      Mean   :3.274e+07
##                                     3rd Qu.:2.117e+07
##                                     Max.   :1.351e+09
##                                     NA's   :1
## Log10_Total_Population  Total_GDP          Log10_Total_GDP
## Min.   :4.012           Min.   :3.767e+07      Min.   : 7.576
## 1st Qu.:5.872           1st Qu.:6.433e+09      1st Qu.: 9.808
## Median :6.781           Median :2.562e+10      Median :10.409
## Mean   :6.574           Mean   :3.700e+11      Mean   :10.462
## 3rd Qu.:7.326           3rd Qu.:1.930e+11      3rd Qu.:11.286
## Max.   :9.131           Max.   :1.620e+13      Max.   :13.210
## NA's   :1              NA's   :16             NA's   :16
## Renewable_electricity_output Total_GHG_emissions Log10_Total_GHG_emissions
## Min.   : 0.00037           Min.   : 5           Min.   :0.7193
## 1st Qu.: 9.24690           1st Qu.: 11811       1st Qu.:4.0723
## Median : 27.53666           Median : 52747       Median :4.7222
```

## Mean	: 37.23679	Mean	: 286601	Mean	:4.5382
## 3rd Qu.	: 63.23411	3rd Qu.	: 161719	3rd Qu.	:5.2088
## Max.	:100.00000	Max.	:12454711	Max.	:7.0953
## NA's	:47	NA's	:40	NA's	:40

Como podemos observar, esta tabla contiene información de nueve variables. Para cada variable aparece un resumen de los datos, cuyo despliegue depende del tipo de variable. Para la variable *Country* se menciona que su longitud es 217, es decir que hay 217 registros. Las variables son:

- *Country*: nombre del país.
- *Income_region*: Categoría de ingreso según el Banco Mundial para el año 2012. Son cuatro categorías.
- *Total_Population*: población total para el año 2012.
- *Log10_Total_Population*: logaritmo en base 10 de la población total para el año 2012.
- *Total_GDP*: producto interno bruto total para el año 2012.
- *Log10_Total_GDP*: logaritmo en base 10 del producto interno bruto total para el año 2012.
- *Renewable_electricity_output*: porcentaje de la energía total producida por el país que es generada en fuentes renovables.
- *Total_GHG_emissions*: Emisiones totales de CO₂ equivalentes para el año 2012.
- *Total_GHG_emissions*: logaritmo en base 10 de las emisiones totales de CO₂ equivalentes para el año 2012.

Respecto a la *calidad* de los datos, se observa que casi todas las variables tienen información faltante, codificada como NA. Sin embargo, el número de NA's es mayor en las variables de emisiones. También se observa que excepto la variable *Country* que está almacenada como vector de tipo *character*, y la de categoría de ingreso que está almacenada como *factor*, todas las demás variables son reconocidas como variables numéricas.

Ahora podemos realizar una inspección gráfica de los datos que nos permita obtener una idea de la distribución de frecuencias de las variables y del grado de asociación entre ellas. Pero antes vamos a hacer una pequeña modificación a la tabla para poder proceder con los análisis. Puesto que las relaciones “más evidentes” se observan con la versión logarítmica de las variables continuas, retiraremos de la tabla las variables originales. A la nueva tabla la guardaremos en otro objeto llamado *dat_SEM_log*. Posteriormente correremos de nuevo los gráficos por pares.

```
dat_SEM_log <- dat_SEM %>%
  select(-Country, -Total_Population, -Total_GDP,
         -Total_GHG_emissions) %>%
  mutate(Income_Region = factor(Income_Region, levels=c("Low_income", "Lower_middle_in
```

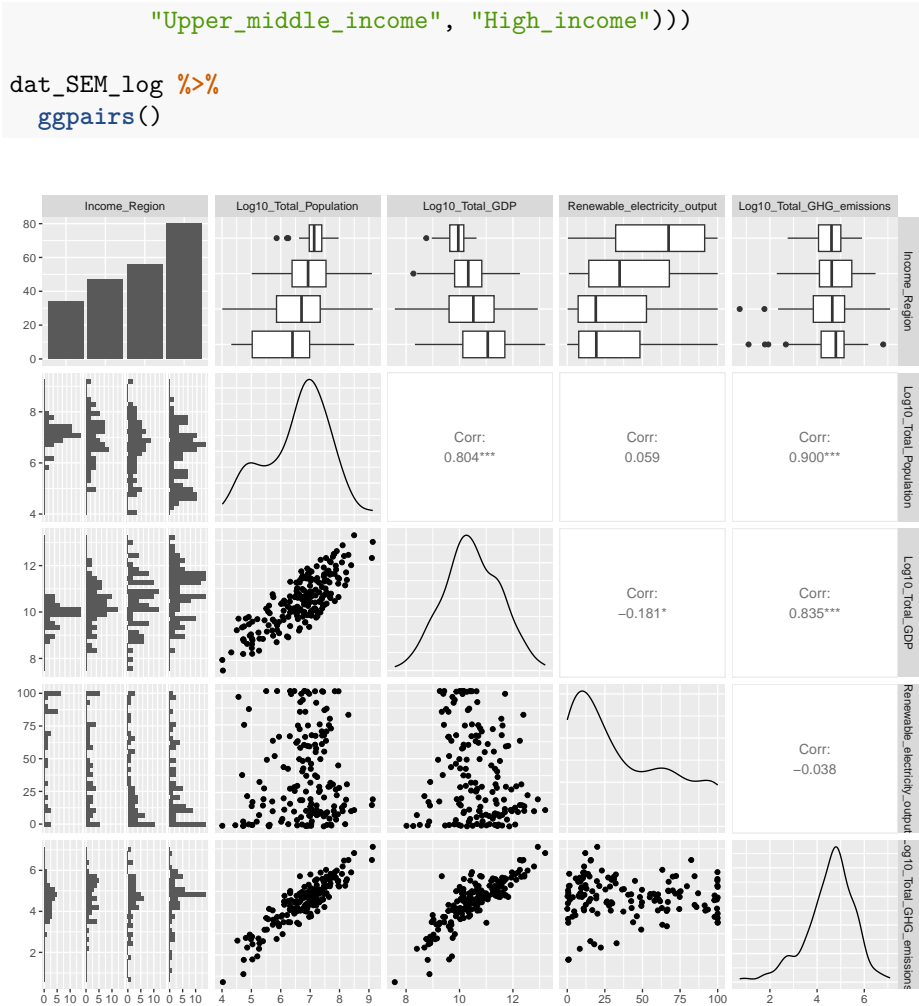



Figure 6.2: Relaciones bivariadas entre todas las variables incluidas en el análisis, una vez las variables cuantitativas fueron transformadas con logaritmo en base 10.

```

ggsave("Figure_2.jpg", ggpairs(dat_SEM_log))

```

Actividad 1 (15 min) : Interpretando gráficas de parejas. Revisa con calma la **Figura 2** y trata de interpretar la información que está mostrada en cada panel. Las siguientes son algunas preguntas que pueden ayudarte a entenderla:

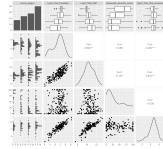


Figure 6.3: Relaciones bivariadas entre todas las variables incluidas en el análisis, una vez las variables cuantitativas fueron transformadas con logaritmo en base 10.

- ¿Qué variable está en el eje x de cada panel? ¿Qué variable está en el eje Y?
- ¿Qué tienen en común todos los paneles de una misma fila? ¿Y los de una misma columna?
- ¿Qué información muestran las gráficas que se encuentran en la diagonal de la matriz de figuras?
- ¿Qué significan los paneles que no tienen gráfica sino un texto?

Reflexiona sobre si las figuras presentan hasta el momento evidencia respecto a los tipos de asociaciones esperadas entre las variables. Discute tus apreciaciones con el grupo.

6.3 Modelo del “PIB Mediador”

Empezaremos nuestro ejercicio de ajuste del SEM con el modelo representado en la **Figura 1a**. En dicho modelo, el efecto del tamaño de la población está medido por el producto interno bruto, de tal forma que al incrementarse el tamaño de la población se incrementa el PIB, lo cual causa a su vez un incremento en las emisiones de CO₂. El PIB actúa entonces como **mediador** del efecto de la población.

6.3.1 Evaluación visual de la forma de las relaciones funcionales

Veamos con más detalle cómo se ve la relación entre estos dos pares de variables. Para ello nos ayudaremos incluyendo en las figuras líneas de tendencia no paramétricas suavizadas.

```
ggsave("Figure_3.jpg",
ggarrange(
```

```

dat_SEM_log %>%
  ggplot(aes(x=Log10_Total_Population, y=Log10_Total_GDP)) +
  geom_point() +
  geom_smooth() +
  labs(title="a) PIB vs. Población"),

dat_SEM_log %>%
  ggplot(aes(x=Log10_Total_GDP, y=Log10_Total_GHG_emissions)) +
  geom_point() +
  geom_smooth() +
  labs(title="b) Emisiones vs. PIB")
), height = 5, width = 10
)

```

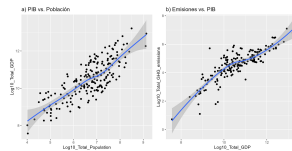


Figure 6.4: **Figura 3.** Relaciones bivariadas entre la población, y el PIB (a) y entre el PIB y las emisiones de GEI (b). Las variables fueron previamente transformadas con logaritmo en base 10.

Actividad 2 (5 min) : ¿Es la línea recta una buena aproximación?. Revisa con calma las **Figura 3** y discute si la línea recta sería una buena representación o idealización de las asociaciones observadas. Discute tus apreciaciones con el grupo.

6.3.2 Definición de las ecuaciones estructurales

Puesto que tenemos dos **variables endógenas** en nuestros modelos de causalidad, éste tendrá que ser traducido en un modelo de ecuaciones estructurales (SEM) que **incluira dos ecuaciones estructurales, una para cada variable endógena**. La ecuación asociada al PIB incluirá un único predictor, la población, puesto que en el modelo de causalidades (**Figura 1a**) solo hay un **padre** para PIB. De igual forma, la ecuación estructural asociada a las emisiones incluirá un único predictor, el PIB, puesto que ésta solo presenta un **padre** en nuestra hipótesis.

Debemos establecer además cuál es la forma de dichas relaciones. Para efectos del ejercicio, supondremos que ambas relaciones pueden ser satisfactoriamente descritas empleando líneas rectas (Ver **Figura 3**). Las expresiones matemáticas de dichas relaciones seguirían entonces la siguiente fórmula general:

$$y = f(x) = \beta_0 + \beta_1 x + \epsilon$$

Dicha ecuación general aplicada a la construcción de las dos ecuaciones estructurales generaría las siguientes expresiones:

$$\text{Log10}(PIB) = \beta_0 + \beta_1 \text{Log10}(Poblacion) + \epsilon$$

$$\text{Log10}(Emisiones) = \beta_0 + \beta_1 \text{Log10}(PIB) + \epsilon$$

Actividad 3 (10 min) : ¿Cómo se interpretan los parámetros β_0 y β_1 de estas expresiones? ¿Cuál de ellos expresa si hay relación o no entre las variables? ¿Qué significan gráficamente? Discute con tus compañeros.

Para escribir dichas ecuaciones estructurales en R y poder calcular sus parámetros emplearemos en este ejemplo modelos lineales, particularmente la función `lm`. El uso de esta función permitirá ajustar cada uno de esos modelos, es decir, obtener una estimación de sus parámetros (β s). **Es acá donde empezaremos a contrastar nuestras ideas (modelo de causalidad) con los datos, pues el “ajustar” los modelos estadísticos asociados a las ecuaciones estructurales implica estimar los parámetros de las ecuaciones con base en los datos, así como generar pruebas de hipótesis al respecto**

Antes de ajustar los modelos lineales, debemos asegurarnos que los datos que se usan en su ajuste son los mismos. Como recordarás, la tabla de datos tiene datos faltantes en varias columnas, lo que haría que modelos diferentes se ajustaran con conjuntos de datos (países) diferentes. Esto generaría problemas de interpretación en el SEM. **Un procedimiento estándar es entonces asegurarse que no hay datos faltantes en la tabla a usar. Para ello usaremos la función `drop_na` y crearemos una nueva tabla de datos.**

Ahora sí podemos proceder a ajustar las dos ecuaciones estructurales:

Podemos ver el resumen del ajuste de cada ecuación estructural. Veamos el de la primera ecuación:

```
##
## Call:
## lm(formula = Log10_Total_GDP ~ Log10_Total_Population, data = dat_SEM_log_sinNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33707 -0.49682 -0.04983  0.46097  1.21268
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.62999    0.47890   9.668  <2e-16 ***
## Log10_Total_Population 0.87528    0.06785  12.901  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6378 on 139 degrees of freedom
## Multiple R-squared:  0.5449, Adjusted R-squared: 0.5416
## F-statistic: 166.4 on 1 and 139 DF,  p-value: < 2.2e-16
```

Hay varios datos interesantes allí. El primero son las estimaciones de los parámetros del modelo (β_0 y β_1), los cuales aparecen en la columna *Estimate* de la tabla *Coefficients*. Allí podemos ver que $\beta_0 = 4.6299921$, y que $\beta_1 = 0.8752825$.

Actividad 4 (5 min) : ¿Cómo se interpretan los valores de las estimaciones de los parámetros? ¿La relación entre las variables es positiva o negativa?

Otro dato interesante es que la variable Población explica algo más del 50% de la variación observada en la variable PIB. Finalmente, parece que la relación entre las dos variables es estadísticamente significativa, como lo sugiere el valor de p asociado a la pendiente del modelo (p <2e-16).

Actividad 5 (5 min) : Evalúa el resultado de la segunda ecuación estructural empleando la función `summary`. ¿Qué observas? ¿A qué conclusión llegas? ¿Qué tanta variación es explicada?

```
##
## Call:
```

```
## lm(formula = Log10_Total_GHG_emissions ~ Log10_Total_GDP, data = dat_SEM_log_sinNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.03450 -0.31177 -0.03454  0.25992  2.10784
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.81639    0.52822  -7.225 3.02e-11 ***
## Log10_Total_GDP  0.79479    0.04886  16.266 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5446 on 139 degrees of freedom
## Multiple R-squared:  0.6556, Adjusted R-squared:  0.6531
## F-statistic: 264.6 on 1 and 139 DF,  p-value: < 2.2e-16
```

6.3.3 El modelo completo empleando `piecewiseSEM`

Ajustar el modelo de ecuaciones estructurales completo implica no solo ajustar cada una de las ecuaciones estructurales, sino poner a prueba si las flechas que no pinté en mi modelo en realidad no existen, así como generar estadísticos globales para todo el modelo. Si bien cada una de estas cosas puede hacerse “a pie”, existen ya librerías y funciones en R que facilitan todo este proceso y lo hacen expedito. Una de esas librerías se llama `piecewiseSEM` (Lefcheck 2016) y su función principal es `psem`, la cual permite ajustar todas las ecuaciones estructurales con un solo comando. Veamos cómo funciona:

Nótese como es muy sencillo el uso de la función `psem`: lo único que hay que hacer es “darle de comer” los modelos estadísticos asociados a cada una de las ecuaciones estructurales asociadas a mi modelo de causalidad. De hecho, las ecuaciones se podrían describir directamente dentro de la función:

Posteriormente se pueden usar una serie de funciones para evaluar el ajuste del SEM. La más básica de ellas es `summary`. Sin embargo (*por una razón que aún no entiendo*) al correr el resumen del SEM desde un *chunk* **sale un error en el resultado: dice que no hay expresiones de independencia.** :

```
## |
##
## Structural Equation Model of modelo_a
##
```

```

## Call:
##   Log10_Total_GDP ~ Log10_Total_Population
##   Log10_Total_GHG_emissions ~ Log10_Total_GDP
##
##       AIC
## 510.057
##
## ---
## Tests of directed separation:
##
##                                     Independ.Claim Test.Type  DF
##   Log10_Total_GHG_emissions ~ Log10_Total_Population + ...   coef 138
##   Crit.Value P.Value
##     13.0449      0 ***
##
## --
## Global goodness-of-fit:
##
## Chi-Squared = 113.279 with P-value = 0 and on 1 degrees of freedom
## Fisher's C = 115.668 with P-value = 0 and on 2 degrees of freedom
##
## ---
## Coefficients:
##
##           Response                Predictor Estimate Std.Error  DF
##   Log10_Total_GDP Log10_Total_Population  0.8753   0.0678 139
##   Log10_Total_GHG_emissions      Log10_Total_GDP  0.7948   0.0489 139
##   Crit.Value P.Value Std.Estimate
##     12.9012      0      0.7382 ***
##     16.2656      0      0.8097 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##           Response method R.squared
##   Log10_Total_GDP      none      0.54
##   Log10_Total_GHG_emissions  none      0.66

```

Dado el problema anterior, debemos de hacer el llamado a la función `summary()` directamente en la consola . El resultado correcto debe de verse en la consola como esto:

152 CHAPTER 6. MODELOS DE ECUACIONES ESTRUCTURALES (SEM)

```

Structural Equation Model of modelo_a

Call:
lmLog10_Total_GDP ~ Log10_Total_Population
lmLog10_Total_GHG_emissions ~ Log10_Total_GDP

AIC
510.057

---
Tests of directed separation:

                                Independ.Claim Test.Type DF Crit.Value P.Value
Log10_Total_GHG_emissions ~ Log10_Total_Population + ...   coef 138   13.0449   0 ***

--
Global goodness-of-fit:

Chi-Squared = 113.279 with P-value = 0 and on 1 degrees of freedom
Fisher's c = 115.668 with P-value = 0 and on 2 degrees of freedom

---
Coefficients:

      Response              Predictor Estimate Std.Error DF Crit.Value P.Value Std.Estimate
Log10_Total_GHG_emissions  Log10_Total_Population  0.8753   0.0678  139   12.9012   0   0.7382 ***
                          Log10_Total_GDP      0.7948   0.0489  139   16.2656   0   0.8097 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

---
Individual R-squared:

      Response method R.squared
Log10_Total_GDP none      0.54
Log10_Total_GHG_emissions none 0.66

```

El resumen del SEM tiene varias partes, a saber:

- *Call*: muestra las ecuaciones estructurales de manera simplificada.
- *Test of directed separation*: La forma de probar si nuestra hipótesis de relaciones de causalidad (modelo de causalidad) es correcta es evaluando si las relaciones que no consideramos en el modelo en realidad no existen. Para el caso de la hipótesis del “PIB Mediador” que estamos evaluando, supusimos que no era necesario incluir la relación directa entre población y emisiones, puesto que el efecto del tamaño de la población sobre las emisiones sería indirecto y estaría completamente mediado por el PIB (**Figura 1**). Entonces lo que se está mostrando acá es una prueba de hipótesis de ello, es decir, de si **además** del efecto a través del PIB, es necesario considerar un efecto directo del PIB sobre las emisiones. Quisiéramos entonces que esta prueba de hipótesis no fuera significativa, con el fin de tener evidencia de que dicha relación no es necesaria.
- *Global goodness-of-fit*: prueba de hipótesis en la que se indica si el modelo de causalidad es coherente con los datos. Específicamente, este estadístico indica si la estructura de relaciones definida (ecuaciones estructurales) es suficiente para explicar las asociaciones observadas entre todos los pares de variables o si, por el contrario, hay evidencia de que hizo falta incluir alguna(s) relación(es) adicional(es) en el modelo. **El modelo de causalidad originalmente propuesto corresponde a la hipótesis nula, por ende, la intención es no rechazar la hipótesis nula (a diferencia de la lógica de la mayoría de las pruebas de hipótesis).**

- *Coefficients*: Es una tabla que muestra los valores de los β_1 's (columna Estimate), es decir, de los parámetros de las ecuaciones estructurales que expresan cuantitativamente *la causalidad* entre las variables. Estas estimaciones de los parámetros están **estandarizadas**: se multiplicaron por la desviación estándar del predictor y se dividieron por la desviación estándar de la variable respuesta. Este re-escalamiento de las estimaciones permite comparar directamente las estimaciones para diferentes relaciones de causalidad.

Adicionalmente, cada parámetro estimado trae asociado una prueba de hipótesis, en la que la hipótesis nula es si el parámetro tiene un valor de cero ($\beta_1 = 0$). Lo que uno esperaría en este caso es rechazar dichas hipótesis nulas, pues ello implicaría que sí hay evidencia en favor de las relaciones causales que se especificaron en las ecuaciones estructurales.

- *Individual R-squared*: Es una tabla que provee los valores de R^2 para cada una de las variables endógenas, es decir, una idea de qué tan bueno es el modelo para explicar cada variable.

Actividad 6 (10 min) : En resumen. A partir de los resultados del ajuste del SEM, regresa al modelo de causalidad original (**Figura 1a**) y establece tus conclusiones respecto a dicho modelo. Luego discútelas con tus compañeros.

6.3.4 Graficando el resultado del SEM

Una forma común de reportar el ajuste de un SEM es presentando el diagrama causal pero “enriquecido” de tal forma que se incluye:

- Información sobre las relaciones de causalidad directa originalmente propuestas, es decir, sobre las flechas originales. Dos piezas de información específica se presentan: la magnitud del efecto de causalidad y la significancia. La magnitud se puede representar tanto presentando el valor del efecto (estimación del parámetro asociado a cada flecha), como modificando el grosor de las flechas para que sea proporcional a dicha magnitud. Para que este último recurso tenga sentido, los parámetros deben de haber sido estandarizados (la opción por default con `piecewiseSEM`), o de lo contrario sus magnitudes no son directamente comparables. También es muy común identificar si la relación es estadísticamente significativa, para lo cual se pueden agregar asteriscos asociados a la estimación del parámetro, y/o colocar las flechas con líneas punteadas cuando la relación resultó no significativa.

- Información sobre las relaciones de causalidad que no fueron propuestas originalmente. En particular, si el ajuste del SEM mostró que alguna relación de causalidad fue originalmente omitida, se debe incluir en la representación esta nueva flecha llamando la atención de que originalmente no hacía parte del modelo, por ejemplo, empleando un color de flecha diferente. A estas flechas también se les adiciona la información de la magnitud y la significancia.
- Información sobre la variación explicada para cada una de las variables endógenas. Dicha información comúnmente se coloca como porcentajes de variación explicada al interior de cada caja.
- Información sobre la prueba de hipótesis del modelo general o el ajuste medido empleando criterios de información tipo AIC. En particular se deben mostrar el valor del estadístico de prueba del *Global goodness-of-fit* y su valor de probabilidad asociado. Se puede incluir también el valor de AIC global del modelo.

Actividad 7 (15 min) : ¡A dibujar!. Emplea una hoja de procesador de texto (ej. Microsoft Word) o de generador de diapositivas (ej. PowerPoint) para generar la representación final del SEM ajustado para el modelo de causalidad de la **Figura 1a**. Para ello trata de tener en cuenta todos los criterios descritos anteriormente.

Actividad 8 (15 min) : Ahora repite todos los procedimientos descritos pero para ajustar el modelo de causalidad hipotetizado en la **Figura 1c**. Sugerencia: no modifiques los chunks de arriba. Más bien, cópialos y pégalos debajo de este texto y modifícalos para desarrollar esta actividad. Recuerda cambiar los nombres a los objetos...

6.4 Bibliografía

Grace, J.B., Scheiner, S.M., Schoolmaster Jr., D.R., 2015. Structural equation modeling: building and evaluating causal models. In: Fox, G.A., Negrete-Yankelevich, S., Sosa, V.J. (Eds.), *Ecological Statistics: From Principles to Applications*. Oxford University Press, Oxford, pp. 169–200.

Jonathan S. Lefcheck, J.S. 2016. PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics *Methods in Ecology and Evolution*, 7, 573–579 doi: 10.1111/2041-210X.12512