Apéndice A

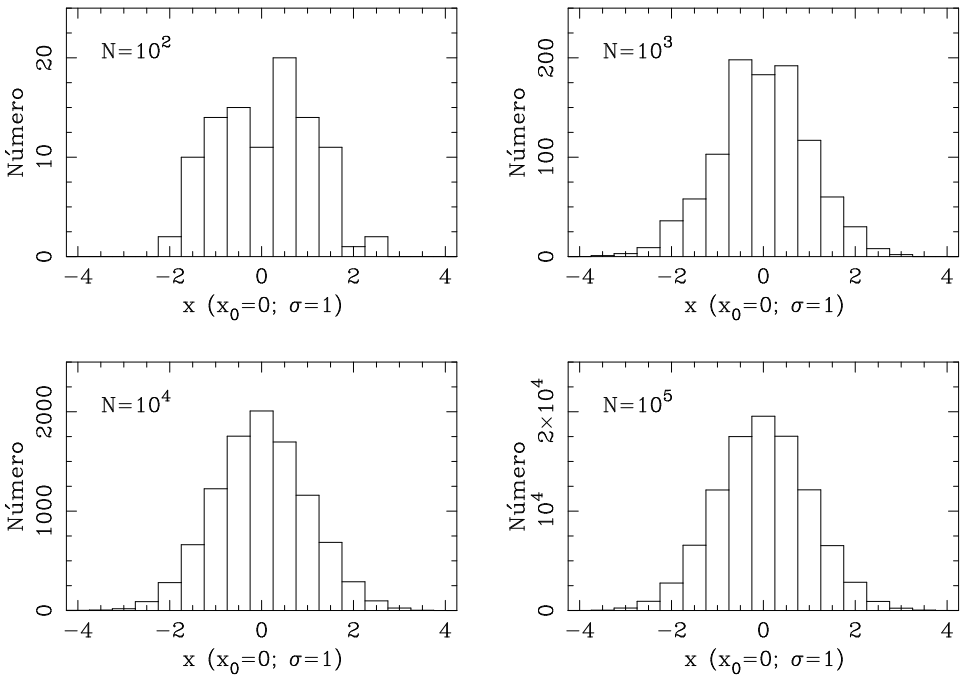
Métodos estadísticos

A.1 Método de Monte Carlo

En esta sección se describen los métodos estadísticos de análisis multivariable utilizados para derivar las propiedades de una población estelar compuesta a partir de la medida de un determinado número de observables y las predicciones de modelos de síntesis evolutiva.

En primer lugar y mediante el uso de un método de Monte Carlo se reconstruyó la distribución de los N observables. Se asumieron como valores centrales de las distribuciones los observables medidos y como σ las correspondientes desviaciones estándar de los mismos. En la Figura A.1 se muestra la reconstrucción de las distribuciones correspondientes utilizando diferente número de puntos, desde 10^2 hasta 10^5. Una vez generadas las distribuciones correspondientes a estos observables en el espacio de N dimensiones, se compararon las mismas con diferentes modelos de síntesis evolutiva. Puesto que los parámetros de estos modelos son la edad (t), la fuerza del brote ($M_{brote}/M_{total}=b$) y la metalicidad (Z), este método permitió obtener las distribuciones (t,b,Z) para cada región de formación estelar o galaxia analizada (véase los Capítulos 4 y 5).

Figura A.1: Ejemplos de reconstrucción de una función gausiana unidimensional a partir del método de Monte Carlo para diferente número de puntos simulados (10^2, 10^3, 10^4 y 10^5).

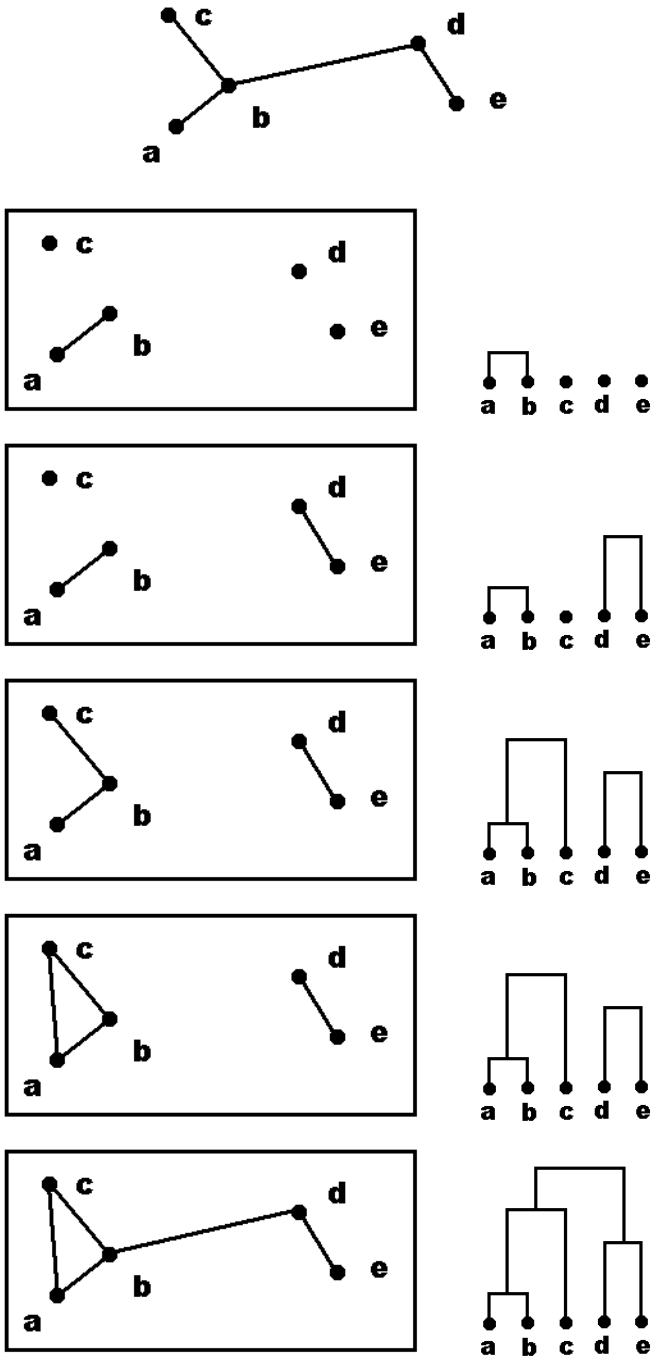


A.2 Análisis de *Clustering*

El valor medio ($< t >$, $< \log b >$,$< \log Z >$) de la distribución de soluciones de cada galaxia puede considerarse como un primer estimador de la mejor solución dentro del espacio (t,$\log b$,$\log Z$), y su desviación estándar (σ_t,$\sigma_{\log b}$,$\sigma_{\log Z}$) de la correspondiente dispersion de los datos. Sin embargo, debido a las degeneraciones existentes entre edad y metalicidad y edad y fuerza de brote, estos valores de la desviación estándar no son siempre representativos de la distribución de soluciones en el espacio (t,$\log b$,$\log Z$). Afortunadamente, estas degeneraciones no cubren el rango completo de edad, fuerza de brote o metalicidad dado por los modelos.

Por lo tanto, se estudió el agrupamiento (o *clustering*) presente en las distribuciones (t,$\log b$,$\log Z$) de cada galaxia o región de formación estelar individual. Para ello se usó un método jerárquico de conexión simple (*single linkage hierarchical*; véase Murtagh y Heck 1987). En primer lugar, (1) se determinaron las distancias entre cada pareja de soluciones, lo que representa un total de $n\times(n-1)/2$ distancias,

Figura A.2: Construcción de un dendograma por el método jerárquico de conexión simple. Adaptada de la Figura 3.3 de Murtagh (1987).



con n=10^3. La distancia entre dos elementos j y k, $d_{j,k}$, se definió como

$$d^2_{j,k} = \sum_{i=1}^{N}(x_{ij} - x_{ik})^2 \tag{A.1}$$

La matriz de distancias obtenida se denominada *dendograma*. A continuación, (2) se encuentra la menor distancia dentro de dicha matriz, $d_{i,k}$. Los puntos i y k son entonces sustituidos por un nuevo punto, $i \cup k$, y el resto de las distancias se actualizan de tal modo que

$$d_{i\cup k,j} = min\ \{d_{i,j}, d_{k,j}\} \tag{A.2}$$

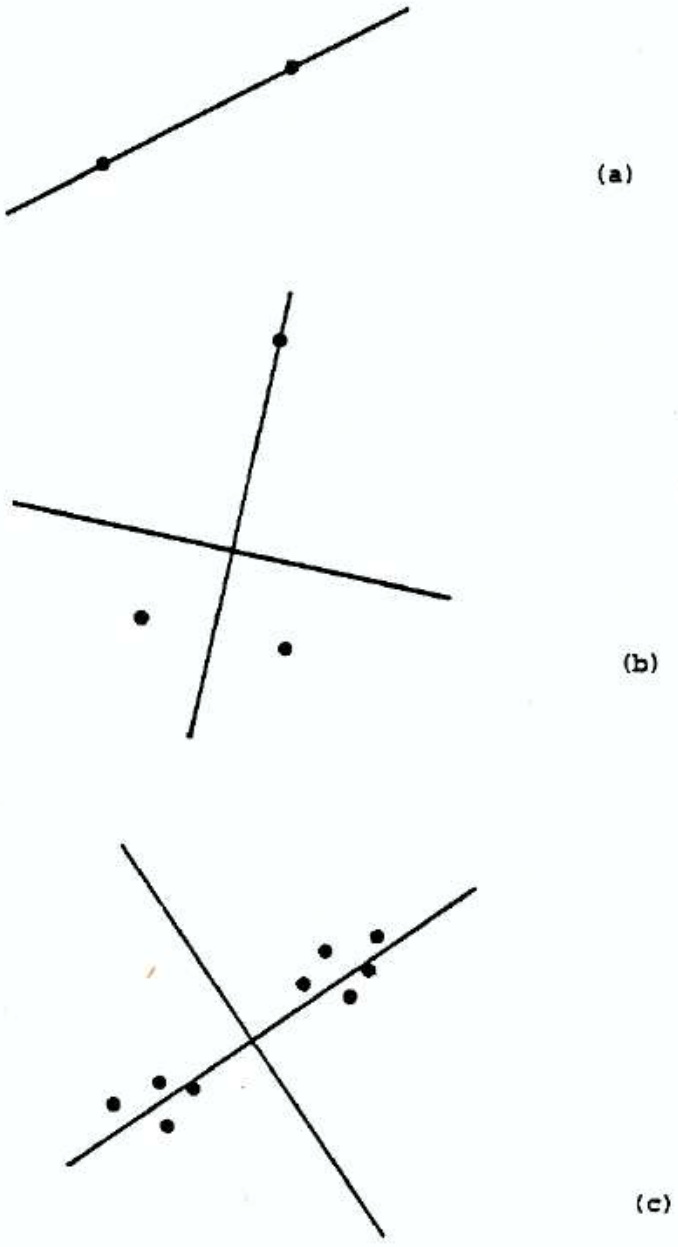
para todo los puntos $j \neq i, k$.

Entonces, (4) las distancias $d_{i,j}$ y $\mathrm{d}_{k,j}$, para todo j, son eliminadas de la matriz, puesto que ya no son necesarias. Se repite a continuación el proceso volviendo al punto (1), así hasta que se recuperan el número de cúmulos de soluciones deseado. En la Figura A.2 se muestra un ejemplo gráfico y simplificado del análisis de *clustering* llevado a cabo. Al final de este apéndice se da el listado del programa en lenguaje `fortran77` desarrollado para llevar a cabo este análisis.

A.3 Análisis de la Componente Principal (PCA)

El resultado de este análisis de *clustering* (véase las Secc. 4.8 y 5.2.4) sugiere que la formación de estos cúmulos es debida a la discretización en metalicidad de los modelos de síntesis evolutiva de Bruzual y Charlot (1996). Se analizó entonces la distribución de las soluciones dentro de cada uno de los cúmulos de soluciones individuales. La discretización en fuerza de brote y edad son comparables, siendo aproximadamente 0.04 dex en fuerza de brote y 0.05 dex en edad. Por tanto, en este caso lo más apropiado es utilizar el método de Análisis de la Componente Principal (PCA). El PCA básicamente determina, dentro de un espacio R^N, el conjunto de N ejes ortogonales que mejor reproducen la distribución de nuestros datos (véase la Figura A.3 para tres ejemplos sencillos). El primero de los nuevos ejes (PCA1), es decir, la componente principal, deberá atravesar la nube formada por el total de los puntos lo más cerca posible a todos ellos, describiendo así la mayor fracción de la varianza de los datos.

Figura A.3: Algunos ejemplos simples del resultado del Análisis de la Componente Principal. **a)** La componente PCA1 para dos puntos es la recta que une esos puntos. **b)** La componente PCA1 une el punto medio de los dos puntos más cercanos con el tercer punto, esto permite reducir la suma cuadrática de las desviaciones de los tres puntos respecto a dicho eje. **c)** En el caso de dos conjuntos de soluciones muy separadas la componente PCA1 se situará a lo largo del eje que une dichos conjuntos de soluciones. Todos estos ejemplos representan caso extremos dentro de un espacio de soluciones genérico. Adaptada de la Figura 2.2 de Murtagh (1987).



Por su parte, en la Figura A.4 se muestran las componentes principales para los tres cúmulos de soluciones de una distribución (t,$\log b$,$\log Z$) hipotética. En la práctica, el análisis de la componente principal (Morrison 1976) se puede realizar a partir de la matriz de varianza-covarianza de la muestra o de la correspondiente matriz de correlación, donde los términos $(j,k)^{th}$ de las mismas, se definen, respectivamente, como

$$c_{jk} = \frac{1}{N}\sum_{i=1}^{N}(r_{ij} - \overline{r}_j) \times (r_{ik} - \overline{r}_k) \tag{A.3}$$

$$\rho_{jk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(r_{ij} - \overline{r}_j)(r_{ik} - \overline{r}_k)}{s_j s_k} \tag{A.4}$$

siendo

$$s_j^2 = \frac{1}{N}\sum_{i=1}^{N}(r_{ij} - \overline{r}_j)^2 \tag{A.5}$$

En nuestro caso hemos empleado la matriz de coeficientes de correlación. Una vez obtenida dicha matriz, se resuelve la ecuación de autovalores $\rho u = \lambda u$ y se obtienen los autovectores y autovalores. El cociente entre un autovalor y la suma de todos los autovalores, $\lambda_i / \sum_{i=1}^{N} \lambda_i$, da la contribución del nuevo eje, definido por el correspondiente autovector, a la varianza total de los datos. Por lo tanto, el autovector con el mayor autovalor representa la componente principal e indica cual es la degeneración dominante dentro de cada cúmulo de soluciones.

Figura A.4: Distribución hipotética de soluciones (t,log b,log Z). PCA$_1$, PCA$_2$ y PCA$_3$ representan en este caso las componentes principales (PCA1) de cada uno de los cúmulos de soluciones.

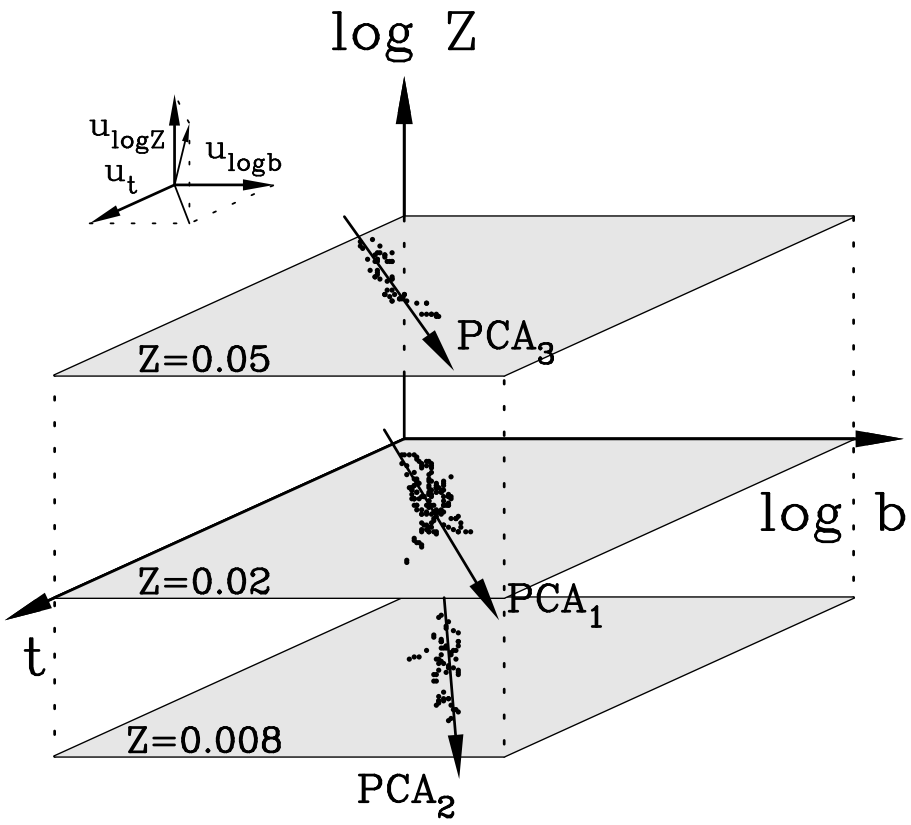


Figura A.5: Subrutina para el análisis de agrupamiento del espacio de soluciones.

Subrutina clan.f Página 1/2

```
      SUBROUTINE CLAN(M,N,A,LEV,NF,F)

C Programa para realizar CLUSTER ANALYSIS
C M=Dimension de la matriz (VARIABLES)
C N=Dimension de la matriz (PUNTOS)
C A=Matriz

C DEFINICION DE VARIABLES

      INTEGER N,I,J,K,L,M,LEV
      REAL A(M,N),D(N,N),DMIN
      INTEGER X(N-1),Y(N-1),F(N,N),NF(N)

C CREACION DE LA MATRIZ DE DISTANCIAS

      DMIN=1.E20
      N1=0
      DO I=1,N
         DO J=1,I
            D(I,J)=0.0
            IF (J.NE.I) THEN
               DO K=1,M
                  D(I,J)=D(I,J)+(A(M,I)-A(M,J))**2.
               ENDDO
               D(I,J)=SQRT(D(I,J))
               IF (D(I,J).LT.DMIN) THEN
                  DMIN=D(I,J)
                  X(1)=MIN(I,J)
                  Y(1)=MAX(I,J)
               ENDIF
            ELSE
               D(J,J)=0.0
            ENDIF
            D(J,I)=D(I,J)
         ENDDO
      ENDDO

      DO I=1,N
      NF(I)=1
      F(I,NF(I))=I
      ENDDO

C ESCRITA MATRIZ DE DISTANCIAS

C COMIENZO DE LA REDUCCION DEL DENDOGRAMA

      DO L=1,N-LEV

         DO I=1,N+1-L
            IF (I.LT.X(L)) THEN
            NF(I)=NF(I)
               DO J=1,I
                  D(I,J)=D(I,J)
                  D(J,I)=D(J,I)
               ENDDO
            ENDIF

            IF (I.EQ.X(L)) THEN
            DO K=1,NF(Y(L))
            F(I,NF(I)+K)=F(Y(L),K)
            ENDDO
            NF(I)=NF(I)+NF(Y(L))
```

Figura A.5: (cont.) Subrutina para el análisis de agrupamiento del espacio de soluciones.

Subrutina clan.f Página 2/2

```
               DO J=1,I
                  D(I,J)=MIN(D(X(L),J),D(Y(L),J))
                  D(J,I)=D(I,J)
               ENDDO
            ENDIF

            IF (I.LT.Y(L).AND.I.GT.X(L)) THEN
               NF(I)=NF(I)
               DO J=1,I
                  IF (J.EQ.X(L)) THEN
                     D(I,J)=MIN(D(I,X(L)),D(I,Y(L)))
                  ELSE
                     D(I,J)=D(I,J)
                  ENDIF
                  D(J,I)=D(I,J)
               ENDDO
            ENDIF

            IF (I.GT.Y(L)) THEN
            DO K=1,NF(I)
            F(I-1,K)=F(I,K)
            ENDDO
            NF(I-1)=NF(I)
               DO J=1,I-1
                  IF (J.LT.X(L)) THEN
                     D(I-1,J)=D(I,J)
                     D(J,I-1)=D(I-1,J)
                  ENDIF
                  IF (J.EQ.X(L)) THEN
                     D(I-1,J)=MIN(D(I,X(L)),D(I,Y(L)))
                     D(J,I-1)=D(I-1,J)
                  ENDIF
                  IF (J.LT.Y(L).AND.J.GT.X(L)) THEN
                     D(I-1,J)=D(I,J)
                     D(J,I-1)=D(I-1,J)
                  ENDIF
                  IF (J.GT.Y(L)) THEN
                     D(I-1,J-1)=D(I,J)
                     D(J-1,I-1)=D(I-1,J-1)
                  ENDIF
               ENDDO
            ENDIF
         ENDDO

C FINAL DE LA REDUCCION DEL DENDOGRAMA

         DMIN=1.E20
         DO I=1,N-L
            DO J=1,I
               IF (I.NE.J) THEN
                  IF (D(I,J).LT.DMIN) THEN
                     DMIN=D(I,J)
                     X(L+1)=MIN(I,J)
                     Y(L+1)=MAX(I,J)
                  ENDIF
               ENDIF
            ENDDO
         ENDDO

      ENDDO

      RETURN
      END
```