

Influencia de la altitud en la composición química de los aceites de oliva vírgenes de Andalucía. Ecuaciones matemáticas de clasificación

Por L. Ferreiro y R. Aparicio*

Instituto de la Grasa y sus Derivados (CSIC).

Avda. Padre García Tejero, 4. 41012 SEVILLA (España)

RESUMEN

Influencia de la altitud en la composición química de los aceites de oliva vírgenes de Andalucía. Ecuaciones matemáticas de clasificación.

La influencia del clima en la composición química del aceite de oliva virgen es un campo de investigación que ha sido escasamente estudiado. En el presente trabajo se expone el estudio de esta relación, realizado para Andalucía, a través de un parámetro indirecto como es la altitud de las zonas de cultivo. Se ha encontrado que parámetros como los esteroides, y algunos alcoholes e hidrocarburos triterpénicos presentan variaciones sensibles en función de la altitud. Estas variaciones han servido para diferenciar los aceites procedentes de olivares de valle de los procedentes de montaña, utilizando técnicas de análisis discriminante y de regresión. Las ecuaciones obtenidas se han verificado con aceites de las cosechas de 1988 y 1989, encontrándose una significativa capacidad predictiva (más del 90%) cuando la variedad del olivar es una constante.

PALABRAS-CLAVE: Aceite de oliva virgen - Altitud (influencia de) - Análisis multivariante - Andalucía - Composición química.

SUMMARY

Influence of altitude on chemical composition of Andalusian virgin olive oil. Mathematical classification equations.

The changes of chemical composition due to climatic influence, in virgin olive oil, is a research subject which has been scarcely developed. This paper studies this relation with Andalusian oils by altitude as an indirect parameter related to climate of their olive grow zones. It was found that sterols and some triterpenic alcohols and hydrocarbons show changes in relation with altitude. The oils were classified into two groups of altitude, valley and mountains, by mean of discriminant and regression analysis procedures. The rules were verified with samples harvested in 1988 and 1989. It was found a significative predictive power (more than 90%) when the variety is Picual.

KEY-WORDS: Altitude (influence of) - Andalusia - Chemical composition - Multivariate analysis - Virgin olive oil.

1. INTRODUCCION

En numerosas ocasiones los investigadores han planteado la posible correlación entre la composición química del aceite de oliva y el lugar de procedencia del aceite, especialmente en lo referente a la altitud o al clima. Es sabido que el clima, sobre todo por su importancia en la maduración, tiene una influencia apreciable en la compo-

sición química de los aceites vegetales. En diferentes cultivos se ha encontrado una relación clara entre variables climáticas y composición química de la planta o productos derivados de ella.

En este sentido, Lotti et al. (1973) han demostrado la influencia de la latitud sobre la composición del aceite de cacahuete, encontrando que el porcentaje de ácidos grasos sigue la ley de Ivanov: mayor cantidad de ácido linoleico y menor cantidad de ácido oleico en climas más fríos. Asimismo, la composición esteróica también sufre variaciones considerables: aumenta la cantidad de beta-sitosterol y disminuyen las de campesterol y estigmasterol a medida que el clima es más frío. Este mismo autor (Lotti et al., 1973) ha puesto de relieve que, en el aceite de semilla de *Momordica balsamina*, existe un pequeño aumento del ácido punícico a medida que los climas son más fríos, mientras el ácido linoleico presenta oscilaciones irregulares.

Por otra parte, en los últimos años se ha incrementado el número de trabajos dirigidos a interrelacionar la composición de los alimentos con el lugar de procedencia. En general, pretenden encontrar funciones de clasificación geográficas, ya sea con métodos estadísticos (Aparicio et al., 1987; Eddib y Nickless, 1987; Leardi y Paganuzzi, 1987; Eteviand et al., 1989; Aparicio et al., 1990; Aparicio et al., 1991a) o, más allá de éstos métodos, desde el campo de la inteligencia artificial (Aparicio, 1988).

Más recientemente (Armanino et al., 1989), se ha señalado la incidencia de la altitud en la composición química de los aceites de oliva vírgenes y su importancia. Mediante la representación gráfica de resultados estadísticos, los autores analizaron la incidencia de la altitud del cultivo en la composición química de los aceites y su posible capacidad discriminadora en métodos de clasificación geográfica, especialmente entre los aceites próximos al mar y los cultivados en las estribaciones de los Alpes.

Este trabajo estudia, siguiendo ese camino, los parámetros químicos con mayor incidencia en la discriminación entre aceites procedentes de diferentes altitudes, que básicamente coinciden con los indicados por Armanino et al. (1989) y Aparicio et al. (1991a; 1991b).

El análisis de estos parámetros químicos, teniendo en cuenta las comarcas olivareras de donde procedían, indujo a pensar que podría existir una relación verificable entre la altura sobre el nivel del mar y algunos parámetros químicos. Se escogió la altura como parámetro externo para estudiar la influencia del ambiente en la composición química del aceite de oliva, por el hecho de ser el más fácilmente disponible y, además, por estar relacionado con variables climáticas más complejas como, por ejemplo, las temperaturas medias, máximas y mínimas.

Este trabajo determina los parámetros químicos capaces de discriminar entre aceites procedentes de altitudes inferiores a 400 m, por un lado y superiores a 700 m, por otro, y formula ecuaciones lineales capaces de discriminar ambos grupos.

Posteriormente, se estima la bondad de cada una de esas ecuaciones con otras muestras de la misma procedencia, pero correspondientes a otras cosechas. También se analizan muestras de otras zonas, de la misma o diferente cosecha, de las que se conoce su altitud mediante la información cartográfica contenida en el Mapa Topográfico de Andalucía, editado por la Consejería de Política Territorial de la Junta de Andalucía.

2. PARTE EXPERIMENTAL

2.1. Metodología química

Se han analizado 53 parámetros químicos pertenecientes a las series de ácidos grasos, esteroides, metil-esteroides, alcoholes triterpénicos, n-alcoholes e hidrocarburos, más el fitol, eritrodilol y aldehído oleanólico (Aparicio et al., 1991a). Los métodos de cuantificación se describen someramente en Aparicio et al. (1991b-c).

2.2. Métodos estadísticos

Con la finalidad de obtener los parámetros químicos que varían con la altitud, se han estudiado sus distribuciones mediante el test de la *t* de Student para la diferencia de las medias de cada parámetro en cada grupo de altitud.

Para el estudio de los grupos de altitud se realiza un contraste T^{**2} de Hotelling, (BMDP, 1981), que establece si se puede rechazar o no la hipótesis nula de la igualdad de las medias de los grupos que se comparan (Bizquera, 1989).

Mientras que, con carácter predictivo, se han utilizado métodos multivariantes (regresión y discriminante) por pasos, que construyen las ecuaciones óptimas para la

clasificación de las muestras en los grupos que se comparan, seleccionando los parámetros más convenientes.

Estos procedimientos, al operar por pasos, permiten obtener los mejores parámetros discriminantes (Tabachnick y Fidell, 1983). Los cuales, a su vez, deben estar comprendidos, en su mayor parte, en el conjunto de los previamente seleccionados, confirmando por tanto los resultados del análisis previo de los parámetros químicos. Un estudio detallado de los métodos aplicados en este trabajo puede verse en Tabachnick y Fidell (1983), Forina et al., (1987) y Bizquera (1989).

3. RESULTADOS Y CONCLUSIONES

3.1. Parámetros químicos relacionados con la altitud

Los métodos citados en primer lugar en el apartado anterior, pueden emplearse para obtener información sobre la variabilidad de los parámetros químicos según la altitud, información que no ha sido descrita en la bibliografía consultada.

Por ello se ha comenzado por análisis estadísticos elementales de los datos disponibles, tal como la construcción de histogramas y la realización de contrastes de igualdad de las medias de los grupos para los distintos parámetros químicos, cuyos resultados permiten una descripción suficientemente detallada de la relación entre altitud y composición de los aceites.

Con este análisis se pretendió descubrir la posible existencia de parámetros químicos, cuya variación pudiera ser indicio de la variación de factores climáticos relacionados con la altitud de la zona de procedencia del aceite.

En una primera etapa, se agruparon las muestras por altitud, distinguiendo tres grandes grupos cuyos límites se establecieron en alturas de 400 y 700 metros sobre el nivel del mar. El primer grupo estaba formado por muestras recogidas hasta 400 metros, el segundo entre 400 y 700, y el tercero por alturas superiores a 700. Para cada una de ellas se construyeron los histogramas que informan de la distribución de los valores de cada parámetro.

Los histogramas, correspondientes a cada grupo, se compararon entre sí observándose que, para algunos parámetros, sus distribuciones presentan diferencias respecto a sus medias y sus rangos de variación. En la Tabla I se observa la tendencia a disminuir a medida que aumenta la altitud del beta-sitosterol, estigmasterol, campesterol, esteroides totales, palmitoleico, y de los hidrocarburos heptadeceno, heneicosano y trideceno.

Tabla I
Distribuciones de algunos parámetros químicos para los tres grupos de altitud. F y P calculados para probabilidad de la cola mayor de 99% (análisis de la varianza)

Tamaño muestra	Nivel de altitud			F	P
	Nivel 1 <400 m.	Nivel 2 400-700	Nivel 3 >700 m		
TRIDECENO				24,34	>0,0001
Media	0,568	0,505	0,365		
Dev. Est.	0,094	0,077	0,089		
COPAENO				22,05	>0,0001
Media	0,026	0,034	0,059		
Dev. Est.	0,005	0,009	0,022		
BETA-SITOSTEROL				18,55	>0,0001
Media	132,922	125,068	110,338		
Dev. Est.	14,708	9,419	8,997		
SUMA DE ESTEROLES				17,17	>0,0001
Media	145,977	138,026	121,796		
Dev. Est.	15,988	11,030	9,421		
CICLO-ARTANOL				14,64	>0,0001
Media	10,481	16,155	19,439		
Dev. Est.	3,108	5,073	5,709		
24-METILEN-CICLO-ARTENOL				13,83	>0,0001
Media	61,043	97,914	120,355		
Dev. Est.	14,127	24,991	44,994		
CAMPESTEROL				11,96	0,0001
Media	4,882	4,642	4,123		
Dev. Est.	0,576	0,419	0,405		
MUUROLENO				11,42	0,0001
Media	0,003	0,005	0,015		
Dev. Est.	0,004	0,004	0,011		
PALMITOLEICO				9,92	0,0003
Media	0,819	0,676	0,621		
Dev. Est.	0,132	0,110	0,142		
ESTIGMASTEROL				6,60	0,0031
Media	1,687	1,406	1,226		
Dev. Est.	0,432	0,403	0,293		
FITOL				5,90	0,0051
Media	3,268	2,797	3,943		
Dev. Est.	0,573	0,450	1,330		
HEPTADECENO				5,85	0,0056
Media	0,029	0,027	0,023		
Dev. Est.	0,006	0,002	0,006		
HENEICOSANO				4,73	0,0138
Media	0,043	0,041	0,036		
Dev. Est.	0,007	0,009	0,005		

Por el contrario, el ciclo-artenol, 24-metilen-ciclo-artenol, muuroleño y copaeno tienden a aumentar con la altitud.

Sin embargo, además de observar las diferencias existentes es preciso determinar si son suficientes como para decidir si corresponden a grupos distintos. El test t-Student se ha utilizado para analizar su capacidad discriminadora. En la Tabla II se dan los valores de la t para el contraste entre el grupo de altitud inferior a 400 metros y el de más de 700 metros, junto con el nivel de significación correspondiente.

Tabla II
T de Student y grado de significación de las variables más importantes al comparar los niveles inferior y superior de altitud

Variable	t	P	Variable	t	P
palmitoleico	4,65	0,0001	beta-sitosterol	4,78	0,0001
margárico	-0,42	0,6792	delta-5-avenast	1,28	0,2158
margaroleico	1,10	0,2806	taraxerol	-2,45	0,0209
linoléico	-1,82	0,0777	danmaradienol	-1,78	0,0840
aráquico	1,56	0,1276	butirospermol	-2,44	0,0205
gadoleico	1,45	0,1560	metilenlanoste	-2,17	0,0373
behénico	0,88	0,3830	ciclo-artenol	-4,04	0,0003
escualeno	-0,05	0,9610	24-metilenciclo	-4,24	0,0002
cicloeucaleol	-2,21	0,0345	trideceno	5,43	>0,0001
citrostadienol	-1,76	0,0878	copaeno	-5,17	>0,0001
fitol	-1,90	0,0692	muuroleño	-4,54	0,0001
eritrodol	0,25	0,8050	heptaceno	2,56	0,0155
campesterol	3,91	0,0006	heneicosano	3,02	0,0057
estigmasterol	3,65	0,0011	tricosano	1,78	0,0851

En esta tabla se puede observar que los parámetros más significativos son, en principio: trideceno, copaeno, heneicosano y muuroleño (hidrocarburos), beta-sitosterol, campesterol y estigmasterol (esteroles), ciclo-artenol, 24-metilen-ciclo-artenol (alcoholes triterpénicos) y ácido palmitoleico.

En el mismo sentido, puede emplearse la distribución F, como el valor de la F_{to_entr} (BMDP, 1981), para detectar los parámetros más significativos. Atendiendo a su grado de significación y de acuerdo con la escala clasificatoria descrita en (Aparicio et al., 1991a), los parámetros se agrupan en:

- Completamente significativos: 24-metilen-ciclo-artenol, copaeno, beta-sitosterol.
- Muy significativos: Linoleico.
- Bastante significativos: Cicloartenol.
- Algo significativos: Fitol, palmitoleico.

3.2. Clasificación de aceites por la altitud de la zona de procedencia.

La otra cuestión que analiza este trabajo es la posibilidad de clasificar las muestras de aceite de oliva según la altitud de los cultivos de olivar de donde proceden. Si esto fuera posible se obtendrían criterios para predecir el nivel de altitud de una muestra dada, conociendo los valores de los parámetros químicos antes citados.

Para resolver este problema se ha recurrido al análisis lineal discriminante y la regresión lineal usando el nivel de altitud como variable discreta, que equivale a un análisis discriminante por otro procedimiento que, sin embargo, se justifica por el interés que tiene obtener otras ecuaciones que permitan la clasificación de las muestras, y por la posibilidad de confirmar los resultados con independencia de los métodos.

Por otra parte, las ecuaciones obtenidas —o reglas de decisión—, se han almacenado en la base de conocimientos del sistema experto que permite clasificar aceites por su procedencia, variedad mayoritaria y sistema de extracción (Aparicio et al., 1991a, 1991b).

Los resultados obtenidos se detallan a continuación, con especial referencia a los coeficientes de correlación parcial de cada parámetro, que indican su importancia predictiva, de correlación lineal múltiple, que expresa en que medida se ajusta la función teórica obtenida a los hechos reales, y el coeficiente de determinación, que mide el porcentaje de la varianza explicada por la función.

3.2.1. Análisis de regresión

En primer lugar, se ha comprobado que las tendencias observadas en el estudio de los parámetros químicos se confirman en el análisis de regresión lineal por pasos, ya que son seleccionados los mismos parámetros químicos cuya relación con la altitud era más significativa.

En segundo lugar, se han podido determinar las zonas entre las que se dan las diferencias más apreciables y las ecuaciones que permiten diferenciarlas. Para ello se consideró: a) tres niveles de altitud, y b) dos niveles, con los resultados que se explican a continuación.

Inicialmente se partió de un conjunto de parámetros que, a raíz de la observación de los histogramas, parecían ser los más útiles, y de tres grupos de muestras correspondientes a tres niveles de altitud (de menos de 400 m, entre 400 y 700 m y de más de 700 m). Esta división del conjunto de muestras en tres grupos demuestra que los parámetros químicos, copaeno y 24-metilen-ciclo-artanol, son sensibles a la variación de la altura. Sin embargo, dada la gradualidad de sus variaciones, no es posible establecer límites claros entre grupos de altitud muy próximos, es decir, entre el primer y el segundo grupo, y entre éste y el tercero.

Consiguiendo, para lograr una observación más clara se prescindió del nivel intermedio, de esta manera se pudieron comprobar claras diferencias entre los aceites procedentes de zonas bajas (<400 m), y los de zonas donde el olivar crece a alturas superiores a los 700 metros sobre el nivel del mar. Prueba de ello es el resultado de aplicar el test T^{**2} de Hotelling a éstos dos grupos, que indica que la hipótesis nula de la igualdad de las medias de los dos grupos puede ser rechazada con 99,6% de probabilidad y, por tanto, que la diferencia entre los grupos es suficientemente significativa (Tabla III). En esta misma tabla se dan los resultados del test de Hotelling para las comparaciones con el nivel intermedio de altitud, la F de Fisher y el grado de significación.

Tabla III

Resultados del test de Hotelling usando las variables discriminantes determinadas mediante la F_{to_enter}

Grupos contrastados	D**2 de Mahalanobis	T**2 de Hotelling	F de Fisher	Grado de significación	Grados de libertad
<400 y 400-700	7,05	46,98	5,10	0,0022	7,24
<400 y >700	13,13	112,59	13,16	>0,0001	7,27
400-700 y >700	5,65	42,37	4,84	0,0016	7,19

Hipótesis nula: ambos grupos tienen medias iguales para todas las variables

Con el fin de buscar informaciones complementarias, se realizaron dos análisis de regresión lineal, uno considerando la serie de los hidrocarburos y otro prescindiendo de ellos. De esta forma se tienen dos ecuaciones complementarias para determinar las altitudes de las zonas de procedencia de las muestras. Si las dos ecuaciones dan como resultado la clasificación de una muestra en una zona determinada, será más probable que la predicción sea acertada que si no coinciden.

Teniendo en cuenta los hidrocarburos se ha obtenido la ecuación de regresión ec.1 (Tabla V), cuyo coeficiente de correlación múltiple es $R=0,85$. En este caso, Tabla IV, se determinaron unos coeficientes parciales de correlación de 0,70 para el copaeno, y 0,65 para el 24-metilen-ciclo-artanol, algo más altos que los obtenidos considerando tres grupos.

Tabla IV

Resumen de las ecuaciones de regresión, coeficientes de correlación parcial de los parámetros y de la ecuación, y Error Típico de la Estimación (E.T.E.)

Variables	Coeficientes de correlación parcial		
	Ec.1	Ec.2	Ec.3
Copaeno	0,70	—	—
24-metilen-ciclo	0,65	0,65	—
Beta-sitosterol	—	-0,70	-0,70
Fitol	—	0,30	—
Ciclo-artanol	—	—	0,69
Palmitoleico	—	—	-0,57
Linoleico	—	—	0,33
R	0,85	0,85	0,90
R**2	0,72	0,72	0,81
E.T.E.	0,27	0,28	0,23

Tabla V
Ecuaciones de regresión y discriminantes

$Y = 0,51 + 0,0055 * (24\text{-metilen-ciclo-artenol}) + 12,24 * \text{copaeno}$	(ec.1)
$Y = 2,16 + 0,0057 * (24\text{-metilen-ciclo-artenol}) - 0,013 * \text{beta-sitosterol} + 0,119 * \text{fitol}$	(ec.2)
$Y = 2,49 - 1,21 * \text{palmitoleico} + 0,23 * \text{linoleico} - 0,01 * \text{beta-sitosterol} + 0,02 * \text{cicloartenol}$	(ec.3)
$Y = 4,72 + 0,024 * (24\text{-metilen-ciclo-artenol}) - 54,16 * \text{copaeno}$	(ec.4)
$Y = -2,55 + 0,52 * (24\text{-metilen-ciclo-artenol}) + 0,06 * \text{beta-sitosterol} + 6,73 * \text{fitol}$	(ec.5)

El intervalo de confianza es, en este caso, de 0,55, y los límites de confianza se sitúan entre 0,45 y 1,55 para el primer grupo, y 1,45-2,55 para el segundo grupo, para un nivel de confianza del 95,0%.

Respecto al análisis de regresión prescindiendo de los hidrocarburos, se obtuvo la ecuación ec.2 (Tabla V), con un coeficiente de correlación lineal múltiple $R=0,85$, igual al de la ec. 1, y los de correlación lineal simple, 0,30 para el fitol -0,70 para el beta-sitosterol y 0,65 para el 24-metilen-ciclo-artenol.

El intervalo de confianza para las predicciones hechas con esta función de regresión se sitúa entre 0,43 y 1,57 para el primer grupo, y entre 1,43 y 2,57 para el segundo grupo, con un nivel de confianza del 95%.

Sin embargo, los resultados de este análisis, a pesar de ser aceptables, manifiestan una tendencia, en el grupo de altitudes superiores, a dar resultados más bajos de los que corresponden a su verdadero nivel de altitud.

Por último, se realizó un análisis de regresión partiendo del conjunto de los 37 parámetros disponibles (sin utilizar los hidrocarburos), que permite utilizar una información matemática como es la covarianza entre todos estos parámetros, que no se ha tenido en cuenta en las anteriores ecuaciones. La ecuación resultante, ec.3 (Tabla V), tiene un coeficiente de correlación lineal múltiple $R=0,90$, al que corresponde un 81% de varianza explicada. En consecuencia, los resultados mejoran respecto a las ecuaciones anteriores.

El intervalo de confianza es de 0,471, siendo los límites de confianza: 0,529 y 1.471 para el nivel 1, y 1.529 y 2.471 para el nivel 2.

3.2.2. Análisis discriminante

También se ha estudiado la posibilidad de distinguir, aplicando el análisis discriminante a los parámetros más correlacionados con la altitud, los aceites que provienen de una zona de altitud baja o alta.

Siguiendo el mismo proceso que en la regresión lineal, se trató de encontrar funciones matemáticas que permitieran diferenciar los aceites por los tres niveles distintos de altitud. La ecuación canónica resultante permite clasificar con una probabilidad predictiva del 80,9%. Siendo los aceites del grupo más alto y más bajo los que mejor se clasifican, y los del grupo intermedio los que más difícilmente se consiguen clasificar, lo cual es lógico, pues, como se ha indicado, la influencia de la altura en la composición de los aceites se supone que es gradual y no

discreta. Por consiguiente, un análisis discriminante de dos grupos permitiría, como en el caso de la regresión, obtener separaciones más nítidas entre los aceites de los dos grupos menos próximos de altitudes.

La ecuación canónica que discrimina entre los dos grupos de altitudes, ec.4 (Tabla V), usa los parámetros químicos copaeno y 24-metilen-ciclo-artenol, como términos de la ecuación, del mismo modo que en su homónima de regresión, y la probabilidad predictiva asociada a la ecuación es del 94,3%, lo que constituye un buen resultado.

Si este mismo análisis discriminante se hace sin contar con los datos de hidrocarburos, se obtiene una ecuación canónica, ec.5 (Tabla 5), en la que fitol y beta-sitosterol reemplazan al copaeno, alcanzando la misma probabilidad predictiva (94,3%).

3.3. Verificación de las ecuaciones

Las ecuaciones fueron obtenidas de un universo muestral que comprende la producción del año 1988 de la provincia de Jaén, cuya variedad mayoritaria es picual.

Estas ecuaciones se han aplicado a otro conjunto de muestras procedentes de cooperativas, de ésta y otras provincias, situadas en zonas de altitud conocida; de manera que se pueda contrastar el nivel de altitud asignado por estos criterios, con el que realmente se sabe que les corresponde. Los resultados permiten evaluar la validez de las ecuaciones a través del porcentaje de predicción correcta, que debe ser del orden de la capacidad predictiva teórica estimada por los métodos estadísticos.

En este sentido, interesa comprobar, en primer lugar, si las ecuaciones son independientes del año de producción; en segundo lugar, si se puede extender su campo de aplicación a todas las provincias estudiadas o sí, por el contrario, sólo son válidas para la provincia de Jaén; y por último, si la aplicación a variedades distintas de la estudiada, picual, afecta a su validez.

Para ello se han utilizado los conjuntos de comprobación que se detallan a continuación.

3.3.1. Muestras de una cosecha posterior y la misma procedencia geográfica

Para comprobar la independencia de los resultados respecto al año de producción, se utilizó un conjunto de prueba de 33 muestras de la provincia de Jaén, correspondientes a la producción de 1989. Las muestras proce-

dían de las mismas cooperativas que se utilizaron para construir las ecuaciones, aunque con muestras de la cosecha de 1988.

Previamente, se comprobó que el análisis discriminante y la regresión por pasos seleccionaban los mismos parámetros químicos en la producción de 1989 que en la del año anterior, cuando se toman estas muestras como conjunto de partida, lo cual es indicativo de que la capacidad discriminante de los parámetros empleados es invariante con el tiempo.

Inicialmente, el conjunto de prueba estaba constituido por 15 muestras del grupo 1 (altitud: 200-400 m.) y 18 del grupo 2 (altitud: más de 700 m). Sin embargo, después de un análisis de los casos se detectaron 4 "outliers" o valores atípicos que fueron eliminados del estudio.

3.3.1.1. Verificación de las ecuaciones discriminantes

La ecuación canónica 5 (ec.5) permitió clasificar correctamente todas las muestras del primer grupo. En cambio, las muestras del segundo grupo originaron más errores en su clasificación. De las 14 muestras 11 han sido correctamente clasificadas, mientras que las otras 3 lo han sido erróneamente en el grupo 1. Analizando las clasificaciones erróneas, se observa que algunas de ellas pueden explicarse por la proximidad de su altitud a los límites de su grupo. Este es el caso en las cooperativas de la comarca de La Loma, correspondientes a las localidades de Sabiote e Iznatoraf, cuyas zonas de recolección tiene una altitud promedio de 750 m. Teniendo en cuenta estos errores, la ecuación clasifica correctamente el 89,6% de las muestras.

3.3.1.2. Verificación de las ecuaciones de regresión

Las ecuaciones de regresión también se han utilizado para clasificar las muestras de la producción de 1989. La ecuación 2 clasifica adecuadamente el 93,1% de las muestras, frente al 96,5% que se alcanza aplicando la ecuación 3.

Como primera conclusión de la aplicación de los procedimientos discriminante y de regresión, se puede considerar que sus ecuaciones son válidas, con independencia del año de producción.

3.3.2. Muestras de distinta procedencia geográfica y de la misma cosecha

Otro grupo de muestras de prueba utilizado para verificar las ecuaciones fue el constituido por muestras de provincias distintas a la de Jaén, correspondientes al año 1988. En este caso se afronta una dificultad nueva respecto al caso anterior, ya que las ecuaciones se han establecido a partir de la provincia de Jaén, que presentaba la ventaja de cultivar como variedad única la aceituna picual. La razón de la elección de este punto de partida era evidente, se trataba de hacer mínimo el número factores que influyen en la variabilidad de los datos químicos.

Los resultados para el conjunto de prueba se detallan a continuación, según la clase de ecuaciones aplicadas.

3.3.2.1. Verificación de las ecuaciones discriminantes

Como conjunto de prueba del análisis discriminante,

se han utilizado 31 muestras de cooperativas que no corresponden a la provincia de Jaén, de las que se posee información sobre la altitud media de la zona, siendo 16 de ellas de variedad picual.

La ecuación 4 clasificó correctamente 12 muestras de la variedad picual, las cuatro muestras clasificadas incorrectamente corresponden a dos cooperativas de Baena, a Aguilar de la Frontera y a Montilla. Las tres primeras se explican por la proximidad de la altitud a la que se ha considerado límite del grupo.

La ecuación canónica 5, en cambio, logró clasificar correctamente todas las muestras de variedad picual, excepto las correspondientes a una de las cooperativas de Baena y la de la Rambla.

Además de la cercanía a los límites, cabe una explicación más, relacionada con la composición varietal de los aceites, que se expone en el siguiente epígrafe. En este sentido, si bien la variedad mayoritaria de las aceitunas utilizadas en la extracción del aceite ha sido la picual, existe una proporción de aceitunas de otras variedades que, a veces, es considerable como ocurre en el caso de las muestras de Aguilar de la Frontera, Montilla y La Rambla.

En conclusión, la ecuación 4 clasifica correctamente el 75% de las muestras, inferior al 87,5% que se obtiene aplicando la ecuación 5.

3.3.2.2. Verificación de las ecuaciones de regresión

Por otra parte, se han utilizado 31 muestras de Córdoba, Málaga y Sevilla para verificar las funciones de regresión, con el resultado siguiente: mediante la ecuación de regresión (ec.3) se ha conseguido clasificar correctamente el 81,2% de las muestras de variedad picual, mientras que la predicción ha sido errónea en tres de ellas, que corresponden a las localidades de: Montilla, que se encuentra en el límite, Espejo y Santaella. Además estas tres muestras, según la información almacenada en la base de datos corresponden a mezclas en las que está presente la variedad hojiblanca.

Por otra parte, la ecuación 2 mejora los resultados de la ecuación 3, sólo 6,3% de errores que corresponden a una muestra de La Rambla y otra de Baena, mientras que la ecuación 1 es inaplicable al generar más del 43% de errores.

Atendiendo a otras variedades, las muestras de la variedad hojiblanca son difícilmente clasificables mediante las ecuaciones descritas hasta el momento, puesto que tienden a ser asignadas, casi siempre, al grupo de mayor altitud. Las causas de que éstas ecuaciones no sean efectivas para clasificar las muestras que no pertenecen a la variedad picual, hay que buscarlas en la distribución de los parámetros químicos escogidos para la clasificación. Estos parámetros presentan rangos de variación muy diferentes, en algunos casos, para los mismos niveles de altitud, como puede verse en la Tabla VI, en la cual se comparan las distribuciones de la comarca de Estepa (SEST), Sevilla, y Archidona (MARC), Málaga, ambas de variedad hojiblanca, casi exclusivamente, pero de menor nivel de altitud la primera que la segunda.

Tabla VI

Medias y rangos de variación de los parámetros químicos más usados por las ecuaciones de clasificación

Variedad Provincia nivel de altitud	Hojiblanca		Picual	
	Sevilla bajo	Málaga alto	Jaén bajo alto	
<i>BETA-SITOSTEROL</i>				
media	169,0	138,7	132,9	110,3
máximo	205,3	149,5	163,7	122,6
mínimo	139,6	132,7	114,9	87,2
<i>24-METILEN-CICLO-ARTENOL</i>				
media	74,4	105,9	61,0	120,4
máximo	117,7	131,5	85,8	219,0
mínimo	43,2	79,7	36,6	59,9
<i>COPAENO</i>				
media	0,17	0,32	0,03	0,06
máximo	0,25	0,41	0,04	0,10
mínimo	0,10	0,25	0,01	0,02
<i>FITOL</i>				
media	5,9	5,6	3,3	3,9
máximo	7,0	6,5	4,5	8,3
mínimo	4,9	5,0	2,4	2,6
<i>CICLO-ARTANOL</i>				
media	13,9	17,5	10,5	19,4
máximo	18,5	22,9	15,7	30,4
mínimo	9,1	12,8	5,6	10,0

La variedad hojiblanca tiene un rango de variación de los parámetros discriminantes al que no se puede aplicar las ecuaciones obtenidas. Por consiguiente, la clasificación por altitud de las muestras de esta variedad requiere otras ecuaciones más adecuadas, que están dirigidas a construir relaciones entre parámetros químicos más que ecuaciones matemáticas, y que serán objeto de nuevos trabajos en un futuro próximo.

3.3.3. Muestras de distinta cosecha y de distinta procedencia geográfica

Un paso más en la verificación de las ecuaciones es la aplicación de éstas a muestras de zonas olivareras

que no son de la provincia de Jaén, ni corresponden a la cosecha utilizada para la obtención de las ecuaciones. Los resultados obtenidos mediante las ecuaciones de regresión permiten deducir que, para una misma variedad –picual–, la clasificación correcta es factible en el 90% de los casos con la ec.2 y el 70% con la ec.3, mientras que si se trata de variedades diferentes es muy difícil la predicción de la altitud, por las razones explicadas en el apartado anterior. Asimismo, el análisis discriminante, mediante la ecuación 5 –similar a la ecuación 2–, consigue clasificar el 90% de las muestras.

3.4. Conclusiones

Se ha encontrado que el 24-metilen-ciclo-artanol y el copaeno son los parámetros más estrechamente relacionados con la altitud, seguidos de los esteroides –tanto en su suma total, como considerando campesterol, estigmasterol y beta-sitosterol por separado–, del ciclo-artanol, de los hidrocarburos trideceno y muuroleno, y del ácido palmítico. El fitol, en mucha menor proporción según indican los resultados de la regresión y los histogramas, también presenta alguna relación, cuando se comparan niveles de altitud más dispares, aunque su menor influencia puede deberse a la saponificación del proceso de cuantificación.

Por otra parte, se puede afirmar que es posible predecir si una muestra proviene de una zona de altura inferior a los 400 metros, o de una de más de 700 metros sobre el nivel del mar, con cierta garantía. Mientras que no es probable detectar si un aceite proviene de una zona de altitud entre 400 y 700 metros, en el estado actual de la investigación. Sin embargo, esta posibilidad queda condicionada por la variedad de la aceituna empleada en la obtención del aceite de oliva virgen. En este caso, se ha demostrado la validez de las ecuaciones obtenidas para el aceite de la variedad picual, mientras que las muestras de variedad hojiblanca tienen poca probabilidad de ser clasificadas con acierto.

AGRADECIMIENTO

Este trabajo ha podido realizarse gracias a la colaboración desinteresada de las Cooperativas Olivareras, Delegaciones de Agricultura de Córdoba y Jaén, Agrupaciones de Oleostepa y de Cooperativas de Málaga, junto con la subvención de la Dirección General de Investigación Agraria de la Junta de Andalucía.

Los autores desean expresar su agradecimiento a los Drs. Albi y Cert. Al Dr. Lanzón, que suministró los datos de hidrocarburos del primer año. A D.^a M.^a Soledad Ferreiro, D. Manuel Martín, D. Manuel León, D.^a Rosario González Cordones, D.^a Teresa Rubio Oliver y D. Manuel Rodríguez Aguilar por su inestimable colaboración en la realización de los análisis de las muestras.

Este proyecto ha sido subvencionado por la CICYT, proyecto ALI-0187-CO2-02 y la Dirección General de Investigación y Extensión Agraria de la Junta de Andalucía.

REFERENCIAS

- Aparicio, R.; Albi, T.; Lanzón, A.; Navas, M.A. (1987).- "SEXIA, un sistema experto para la identificación de aceites: Base de datos de zonas olivares".- *Grasas y Aceites* **38**, 9-14.
- Aparicio, R. (1988).- "Characterization of foods by inexact rules: The SEXIA expert system".- *J. Chemometrics* **3**, 175-192.
- Aparicio, R.; Ferreiro, L.; Cert, A.; Lanzón, A. (1990).- "Caracterización de Aceites de Oliva Virgenes Andaluces".- *Grasas y Aceites* **41**, 23-39.
- Aparicio R.; Ferreiro, L.; Rodríguez, J.L. (1991a).- "Caracterización del aceite de oliva de Andalucía".- Junta de Andalucía. Dirección General de Investigación y Extensión Agraria.
- Aparicio, R.; Ferreiro, L.; Leardi, R.; Forina, M. (1991b).- "An approach for building decision rules by chemometric analysis".- *Chemomet. Intellig. Lab. Syst.* **10**, 349-346.
- Aparicio, R.; Sánchez Navarro, M.; Ferreiro, M.S. (1991c).- "Definite influence of the extraction methods on the chemical composition of virgin olive oil".- *Grasas y Aceites* **42**, 356-362.
- Armanino, C.; Leardi, R.; Lanteri, S.; Modi, G. (1989).- "Chemometric Analysis of Tuscan Olive Oils".- *Chemometric and Intelligent Laboratory Systems* **5**, 343-354.
- Bisquerra, R. (1989).- "Introducción conceptual al análisis multivariable".- Vol. 1, Barcelona.
- BMDP Statistical Software.- 1981 edition.- Dept. of Biomathematics, University of California, Los Angeles, CA, USA.
- Eddib, O.; Nickless, G. (1987).- "Elucidation of olive Oil Classification by Chemometrics".- *Analyst* **112**, 391-395.
- Etiévant, P.; Schlich, P.; Cantagrel, R.; Bertrand, A.; Bouvier, J.C. (1989).- "Varietal and Geographic Classification of Red Wines in Terms of Major Acids".- *J. Sci. Food Agric.*, 421-438.
- Forina, M.; Armanino, C.; Lanteri, S.; Calcagno, C.; Tiscornia, E. (1983).- "Valutazione delle caratteristiche chimiche dell'olio di oliva in funzione dell'annata di produzione mediante metodi di classificazione multivariati".- *Riv. Ital. Sostanze Grasse* **60**, 607-613.
- Forina, M.; Lanteri, S.; Armanino, C. (1987).- "Chemometrics in Food Chemistry", en "Topics in current Chemistry".- **141**, 92-143.
- Lotti, G.; Izzo, R.; Nicolini, F. (1973).- "Influenza del clima e dell'andamento della maturazione sull'olio dei semi di *Momordica balsamina* L.". - *Riv. Ital. Sostanze Grasse* **50**, 425.
- Lotti, G.; Baragli, S.; Gentili, M. (1977).- "Composizione acidica e sterolica dell'olio di arachide in funzione del clima".- *Riv. Ital. Sostanze Grasse* **54**, 506.
- Tabachnick, B.G. and Fidell, L.S. (1983).- "Using Multivariate Statistics".- Harper and Row, Publisher, New York.

(Recibido: Noviembre 1991)