

UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN (FAMAF)

ANÁLISIS Y COMPARACIÓN DE MEDIDAS E  
ÍNDICES DE SIMILARIDAD  
PARA IMÁGENES DIGITALES

TESIS PRESENTADA POR MARÍA LUCÍA PAPPATERRA  
PARA OBTENER EL GRADO DE MAESTRÍA EN ESTADÍSTICA APLICADA

BAJO LA SUPERVISIÓN DE LA DRA. SILVIA MARÍA OJEDA

2019

---

# Agradecimientos

Quisiera agradecer a mi directora Silvia por su paciencia y sus ideas, sin su ayuda y conocimiento esta tesis no hubiera sido posible.

A los maestros que me cruce en este largo camino académico.

A mi papá Fernando, por haberme inculcado el amor por el conocimiento. A mi mamá Mercedes, por sus consejos, su apoyo incondicional y su paciencia. Todo lo que hoy soy es gracias a ellos.

A mis amigas y amigos, por estar siempre a mi lado.

A mis familiares, por su apoyo.

A aquellos que siguen estando cerca de mi y que le regalan a mi vida algo de ellos.

---

# Índice general

<b>Agradecimientos</b>	<b>2</b>
<b>1 Introducción</b>	<b>6</b>
1.1 Motivación de esta tesis	6
1.1.1 Presentación del problema	6
1.1.2 Objetivos	9
1.2 ¿Qué es una imagen?	10
1.2.1 1 bit: blanco o negro	11
1.2.2 Escala de grises con 8 bits	12
1.2.3 Formato RGB de 24 bits	12
1.3 El sistema visual humano (HVS)	13
1.3.1 Características psicofísicas del HVS	14
<b>2 Medidas de Calidad de una Imagen</b>	<b>16</b>
2.1 Medidas de referencia completa (FR-IQA: Full Reference Image Quality Assessment)	16
2.1.1 El error cuadrático medio (MSE) y la relación de cociente señal-ruido	17
2.1.2 Medida de calidad del ruido (NQM)	19
2.1.3 Índice de similitud estructural (SSIM)	19
2.1.4 Multi-Scale SSIM (MS-SSIM)	22
2.1.5 Medidas direccionales: El Índice CQ	22
2.1.6 Medidas direccionales: FR-IQA basadas en la magnitud del gradiente	24
2.1.7 Fidelidad a la información visual (VIF)	28
2.2 Medidas de referencia reducida (RR-IQA: Redude Reference Image Quality Assessment)	28
2.3 Medidas de referencia nula (NR-IQA: Null Reference Image Quality Assessment)	29
2.4 Ejemplo de aplicación	29
<b>3 La Base TID2013</b>	<b>31</b>
3.0.1 El Índice MOS	31
3.1 Los distintos tipos de distorsiones	32
3.1.1 Distorsión #1: Ruido Gaussiano Aditivo	35

3.1.2	Distorsión #2: Ruido aditivo en la componente cromática . . . . .	35
3.1.3	Distorsión #3: Ruido correlacionado espacialmente . . . . .	36
3.1.4	Distorsiones #4 y #5: Ruido enmascarado y Ruido de alta frecuencia . . . . .	37
3.1.5	Distorsión #6: Ruido sal y pimienta o Ruido de impulso . . . . .	38
3.1.6	Distorsión #7: Ruido de cuantificación . . . . .	39
3.1.7	Distorsión #8: Desenfoque Gaussiano . . . . .	40
3.1.8	Distorsión #9: Eliminación de ruido . . . . .	41
3.1.9	Distorsiones #10, #11, #12 y #13: Compresión JPEG, Compresión JPEG2000, Transmisión de errores JPEG y Transmisión de errores JPEG2000 . . . . .	42
3.1.10	Distorsión #14: Patrón de ruido sin excentricidad . . . . .	44
3.1.11	Distorsión #15: Distorsiones locales de bloques de diferentes intensidades . . . . .	45
3.1.12	Distorsiones #16 y #17: Cambio de valor medio y Cambio de contraste . . . . .	48
3.1.13	Distorsión #18: Cambio en la saturación del color . . . . .	49
3.1.14	Distorsión #19: Ruido Gaussiano Multiplicativo . . . . .	49
3.1.15	Distorsión #20: Ruido de confort . . . . .	49
3.1.16	Distorsión #21: Compresión de imágenes ruidosas . . . . .	50
3.1.17	Distorsión #22: Cuantificación del color con tramado . . . . .	51
3.1.18	Distorsión #23: Aberraciones cromáticas . . . . .	52
3.1.19	Distorsión #24: Sparse sampling and reconstruction . . . . .	54
<b>4</b>	<b>Obtención y análisis de los datos</b>	<b>56</b>
4.1	Breve análisis descriptivo . . . . .	57
<b>5</b>	<b>Correlación</b>	<b>64</b>
5.1	Técnicas no paramétricas de correlación . . . . .	71
5.2	Correlación de Spearman . . . . .	71
5.2.1	Prueba de Hipótesis . . . . .	72
5.3	Correlación de Kendall . . . . .	73
5.3.1	Prueba de Hipótesis . . . . .	75
5.4	El coeficiente w de Kendall . . . . .	76
5.4.1	Prueba de concordancia de Kendall . . . . .	77
<b>6</b>	<b>Concordancia</b>	<b>79</b>
6.1	Matriz de confusión . . . . .	79
6.2	Coefficiente kappa de Cohen . . . . .	87
6.3	Coefficiente pi de Scott . . . . .	89
6.4	Coefficiente kappa de Fleiss . . . . .	90
6.4.1	Ecuaciones . . . . .	91

---

<b>7</b>	<b>Análisis por grupos de distorsiones</b>	<b>95</b>
<b>8</b>	<b>Conclusiones y trabajos a futuro</b>	<b>104</b>
<b>A</b>	<b>Diagramas de dispersión de las distorsiones #1 a #9</b>	<b>112</b>
<b>B</b>	<b>Diagramas de dispersión de las distorsiones #10 a #17</b>	<b>139</b>
<b>C</b>	<b>Diagramas de dispersión de las distorsiones #18 a #24</b>	<b>163</b>
<b>D</b>	<b>Códigos en Python para obtener los coeficientes de correlación</b>	<b>184</b>
D.1	Función que le asigna a cada valor su rango ordenándolos de mayor a menor . . . . .	184
D.2	Función que calcula la correlacion de Spearman . . . . .	185
D.3	Función que calcula el total de pares concordantes y el total de pares discordantes	186
D.4	Función que cuenta la cantidad de pares enlazados que tiene cada grupo de pares enlazados . . . . .	186
D.5	Función que calcula el coeficiente de correlación de Kendall . . . . .	187
D.6	Función que calcula el coeficiente de concordancia w de Kendall . . . . .	189
<b>E</b>	<b>Matrices de Confusión para cada grupo de distorsiones</b>	<b>191</b>
<b>F</b>	<b>Códigos en Python para obtener los coeficientes de concordancia</b>	<b>216</b>
F.1	Función que calcula el kappa de Cohen y el pi de Scott entre dos variables . . . . .	216
F.2	Función que calcula índices de concordancia entre varias variables . . . . .	217
F.3	Función que calcula el kappa de Fleiss . . . . .	218
<b>G</b>	<b>Medidas de Resumen para cada grupo de distorsiones</b>	<b>221</b>

---

# Capítulo 1

## Introducción

### 1.1. Motivación de esta tesis

El enorme desarrollo de los recursos tecnológicos de las últimas décadas, ha sido un factor determinante en la construcción e implementación computacional de diferentes índices y medidas de calidad para cuantificar la similaridad entre dos imágenes digitales. Si bien las propuestas son innumerables, resultan escasos e incompletos los estudios de comparación y clasificación de estas herramientas. Esta tarea es importante, porque el concepto *similitud* referido a imágenes digitales es muy amplio y las metodologías sugeridas deben dar respuestas en diferentes escenarios donde se plantean disparidades y semejanzas según múltiples y variados arquetipos de distorsión. Esta tesis plantea un estudio intenso del desempeño de las medidas e índices de similaridad propuestos hasta el momento para relacionar imágenes digitales; propone comparar el desempeño de estos procesos teniendo en cuenta su capacidad para emular el Sistema Visual Humano (HVS) al vincular una imagen con una distorsión de ella, bajo distintos esquemas y niveles de cambio. La iniciativa se manifiesta a favor de señalar a los mejores procedimientos de comparación de imágenes digitales, por sus propiedades matemáticas y estadísticas y por su solidez en cuanto a su capacidad de interpretar al HVS.

#### 1.1.1. Presentación del problema

En la mayoría de los problemas en procesamiento de imágenes digitales, surge la necesidad de comparar imágenes; si diferentes algoritmos se aplican a una imagen con un propósito específico, por ejemplo recuperación de información perdida, será requisito disponer de una herramienta precisa que permita evaluar tales algoritmos; la base de comparación entre estos procesos es precisamente la comparación entre las imágenes de salida de cada algoritmo y una imagen que se toma como referencia.

A pesar de la frecuencia con que se requiere comparar imágenes digitales, esta tarea no resulta en absoluto trivial; en parte porque tal comparación involucra el concepto matemático de *medida*

y además conlleva implícitamente la exigencia de interpretar al HVS [1],[2].

*Similaridad, Asociación y Calidad* constituyen nociones fuertemente relacionadas al comparar imágenes digitales. La evaluación de la calidad de una imagen (IQA, por sus siglas en inglés: Image Quality Assessment), se vincula a la valoración de calidad de la imagen que se deriva del juicio humano; a nivel general, IQA se puede clasificar en **IQA subjetiva** e **IQA objetiva** [3]. La perspectiva subjetiva es ampliamente aceptada como el enfoque más exacto para medir la calidad, ya que el ojo humano es el receptor final de la mayoría, sino de todos, los sistemas de comunicación visual [4].

La contribución más significativa de IQA subjetiva en las últimas décadas es probablemente la construcción de las bases de datos, constituidas por imágenes digitales con diversos tipos de distorsiones y sus valoraciones subjetivas registradas con el índice MOS (mean opinion score). Las bases de datos más utilizadas son Laboratory for Image & Video Engineering (LIVE) [5], Tampere Image Database 2008 (TID2008) [6], Tampere Image Database 2013 (TID2013) [7] [8] y Categorical Subjective Image Quality (CSIQ) [9]. Esas bases de datos facilitan en gran medida la investigación de propuestas de evaluación objetiva de la calidad de imágenes (IQA objetiva) [10].

El procedimiento para la obtención del índice MOS, la organización de las bases de datos TID2008 y TID2013, y los tipos de distorsiones que puede sufrir una imagen son aspectos que se abordan con mayor detalle en el capítulo 3.

La evaluación IQA subjetiva, a pesar de ser el enfoque más fiel al HVS, a menudo resulta lento, caro, y laborioso; por lo tanto no es adecuado para la cuantificación de la calidad de imágenes en general, ni para la evaluación de la similitud entre imágenes dentro de sistemas de comunicación visual. Como alternativa muchos investigadores se han dedicado a la exploración de algoritmos en el marco de IQA objetiva [4].

Para el diseño de una medida objetiva de calidad de una imagen, puede utilizarse información sobre la imagen original (de la cual no siempre se dispone), información sobre el tipo de proceso de distorsión aplicado sobre la misma y/o conocimiento sobre cómo opera el HVS. Además, estos procedimientos se clasifican en tres grupos: Referencia completa (FR-IQA: Full Reference Image Quality Assessment), Referencia reducida (RR-IQA: Redude Reference Image Quality Assessment) y Referencia nula (NR-IQA: Null Reference Image Quality Assessment), en función de la disponibilidad de la imagen de referencia [11], [12].

La lista de índices, métricas <sup>1</sup> y pseudo-métricas desarrolladas hasta el presente para evaluar similaridad y calidad en imágenes es muy extensa y se encuentra en progreso permanente; solo por mencionar algunos, entre los procedimientos desarrollados en el marco FR-IQA se encuentran: el error cuadrático medio (MSE) que es el más ampliamente utilizado a pesar de su comprobado bajo rendimiento [12], la relación de cociente señal-ruido (SNR) definida en base al MSE y medidas

---

<sup>1</sup>Los términos métrica o medida se utilizarán a lo largo de este trabajo para referirse a los índices y procedimientos para determinar la calidad objetiva de una imagen y no deben ser interpretados en el sentido matemático estricto de la teoría de la medida.

relacionadas como PSNR (Peak Signal to Noise Ratio) y WSNR que agrega pesos. La medida de calidad del ruido (NQM) [13] definida según el modelo de contrastes de Peli [14]. El índice de calidad universal (UQI) y el índice de similaridad estructural (SSIM) [15] [16] que tienen en cuenta la información estructural de la escena global, y sus derivados MS-SSIM (Multi-Scale SSIM) [17], WSSIM [18], IW-SSIM [19], SW-SSIM [20], entre otros. Las medidas direccionales como el índice CQ [21] y CQ-max [22] y direccionales basadas en la magnitud del gradiente como las GSM (Gradient Magnitude Similarity Mean) y GMSD (Gradient Magnitude Similarity Deviation) [23] o las FSIM (Feature Similarity Index) [24] y GSIM (Gradient Similarity Index) [25]. Fidelidad a la información visual (VIF) [26] es una medida definida en base a la teoría de la información. Las métricas GES [27] y LGPS [28] fueron propuestas en el dominio de la transformada de Gabor; MAD (Most apparent distortion) [29] es una medida que utiliza estrategias basadas en la apariencia. La medida IPAD (Image quality model based on phase and amplitude differences) [30] utiliza el análisis de amplitud y fase. Teorías basadas en las neurociencias hicieron grandes aportes al diseño IQA, por ejemplo la medida IGM (internal generative mechanism) [31], que utiliza parámetros psicofísicos.

Otras medidas desarrolladas en el marco de RR-IQA son: el índice FEDM (Free energy based distortion metric) [32], RRED (RR entropic-difference indexes) [33], FTQM (Fourier Transform based quality measure) [34]. También existen medidas basadas en el SSIM, entre ellas RR-SSIM [35] y SDM (Structural degradation model) [36].

La mayoría de los sistemas de comunicación visual no cuentan con una imagen de referencia ni dispone de información sobre la misma, por eso, en el marco NR-IQA se han desarrollado algunas medidas teniendo en cuenta ciertos tipos de distorsiones en particular, como los estimadores de ruido blanco propuestos en [37], [38] [39] y el JPEG quality estimator [40]. Otras medidas NR-IQA de propósito general son DIIVINE [41], BLINDS-II [42] y BRISQUE [43] que extraen características de la imagen distorsionada y luego realizan una regresión; QAC (quality-aware clustering) [44] y NIQE (natural image quality evaluator) [45] que no utilizan evaluaciones humanas.

En el capítulo 2 se definen y caracterizan los índices y medidas más importantes.

De la exposición precedente, puede verse que existe una enorme cantidad de índices, medidas y métricas para determinar la calidad objetiva de una imagen, sin embargo, no se ha realizado hasta el presente una comparación exhaustiva de estas alternativas. Este trabajo de comparación es importante, porque diferentes medidas estudian la similaridad entre imágenes enfatizando distintos aspectos y propiedades de las mismas y sus distorsiones. Por ejemplo, las medidas de similaridad espacial direccionales, evalúan y analizan similaridad en una dirección específica, mientras que SSIM y sus posteriores versiones hasta 2012, analizan la similaridad desde un punto de vista global. Surge entonces una pregunta natural: ¿Qué índice de similaridad conviene utilizar? No tenemos una respuesta; sin embargo, un punto de partida inicial es disponer de un estudio comparativo de estas herramientas, evaluando su desempeño frente a distorsiones de interés de las imágenes en las bases de imágenes disponibles, anteriormente mencionadas y usando la interpretación del sistema



visual humano (HVS) dada a partir del MOS.

Restringiéndonos al marco de FR-IQA, que es ampliamente utilizado para evaluar algoritmos de procesamiento de imágenes al medir la calidad de las imágenes de salida, nos proponemos realizar este trabajo de análisis y comparación a través de métodos estadísticos.

### 1.1.2. Objetivos

A continuación se presentan los objetivos propuestos para esta tesis.

#### 1.1.2.1. Objetivos Generales

- 1) Realizar aportes relevantes dentro del área del Análisis y Procesamiento Estadístico de Imágenes que permitan determinar procedimientos óptimos para la comparación de imágenes digitales.
- 2) Elaborar un estudio comparativo intenso de las propuestas desarrolladas hasta el presente para medir similaridad y evaluar calidad en imágenes digitales.

#### 1.1.2.2. Objetivos Específicos

- 1) Realizar una detallada búsqueda bibliográfica sobre el estado del arte en lo referente a la evaluación de calidad y medidas de Similitud de imágenes digitales, partiendo de [1], [12], [15], [11], [46], [47].
- 2) Realizar una selección de los índices y medidas de Similitud y calidad más utilizados y de mayor empleo en situaciones prácticas, distinguiendo entre aquellos que incorporan la información contextual, tal como procede el Sistema Visual Humano, de aquellos que no lo hacen. Como punto de partida, consultar [48].
- 3) Utilizar las implementaciones computacionales disponibles de los procedimientos señalados en 2), y desarrollar las que falten.
- 4) Evaluar los índices y medidas seleccionados en alguna de las bases de datos disponibles y habitualmente usadas, [6], [7] y [5].
- 5) Comparar el desempeño de estos métodos con el MOS (Mean Opinion Score).
- 6) Posicionar estos índices en un ranking, en base al estudio comparativo efectuado.
- 7) Analizar, ventajas, desventajas, fortalezas y debilidades de los índices estudiados.
- 8) Completar el análisis estadístico realizado en [8], agregando los índices direccionales que no fueron incluidos en el mismo.

## 1.2. ¿Qué es una imagen?

Para el desarrollo de los siguientes capítulos es necesario introducir el concepto de imagen y sus representaciones.

Desde una perspectiva formal, una imagen digital constituye una realización  $x$  de un proceso aleatorio  $X : \Omega \rightarrow E^W$ , donde  $E$  es un conjunto acompañado por una estructura de  $\sigma$ -álgebra, siendo  $W$  la “ventana de observación de la imagen  $x$ ” [49], [50]. Desde el punto de vista de un observador común, se trata de una escena provista de ciertos atributos que le son propios y la caracterizan, tales como bordes y patrones de textura.

El término imagen se refiere a una función de intensidad bidimensional  $x(i, j)$ , donde  $i$  y  $j$  son las coordenadas espaciales y el valor de  $x$  en un punto  $(i, j)$  es proporcional a la intensidad o nivel de gris de la imagen en ese punto. Se denotará indistintamente y dependiendo del contexto y de la nomenclatura utilizada por los métodos que vamos a definir, como  $x(i, j)$ ,  $y(i, j)$ ,  $f(i, j)$ ,  $f(x, y)$ , etc.

Podemos representar a una imagen como una matriz numérica donde cada elemento, denominado píxel (picture element) tendrá un valor asignado que se corresponde con el nivel de luminosidad del punto correspondiente en la escena captada. Esta matriz se denomina mapa de bits (bitmap en inglés); un esquema de la misma puede apreciarse en la Figura 1.1.

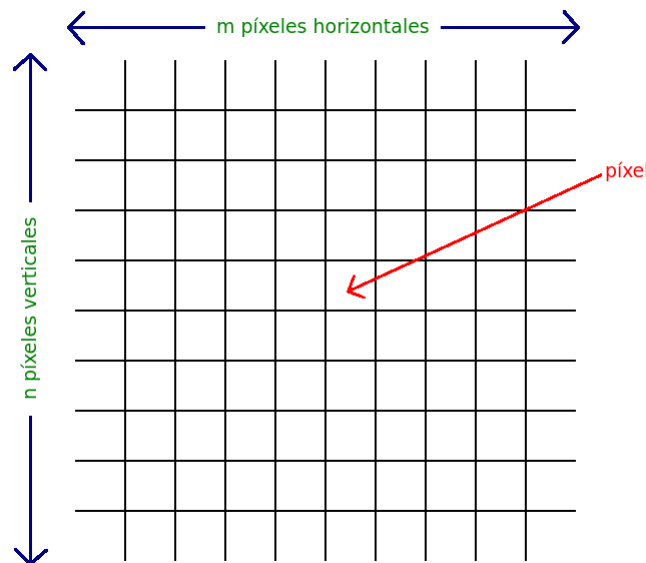


Figura 1.1: Mapa de bits, representación bidimensional de una imagen

En base a esta representación bidimensional, definimos un sistema de coordenadas como se muestra en la Figura 1.2, que necesitaremos para definir algunos conceptos en el próximo capítulo.

A lo largo de este trabajo por lo general representaremos a las imágenes como matrices bidimensionales, aunque también podemos representarlas como un vector unidimensional al acomodar

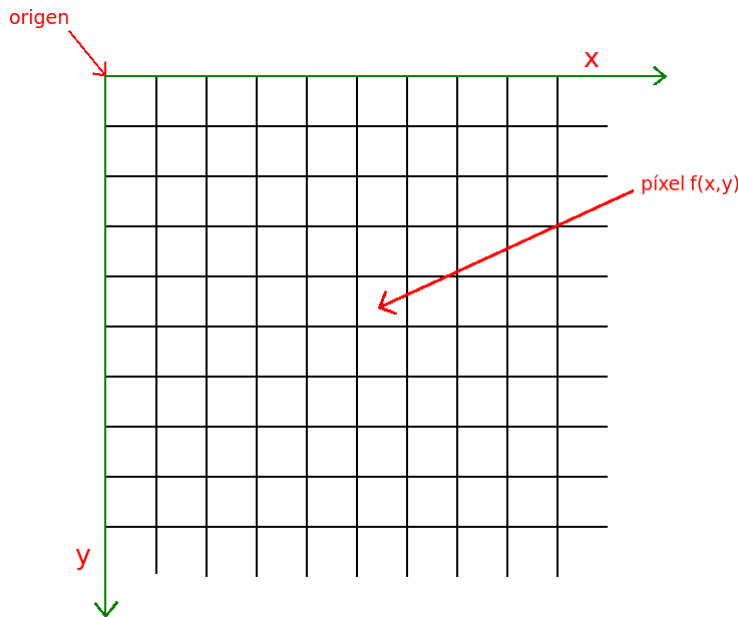


Figura 1.2: Convención de ejes utilizada para la representación de imágenes digitales.

en orden las filas <sup>2</sup> de la representación bidimensional en un vector:

$$x = \{x_i : i = 1, 2, \dots, N\} \quad (1.1)$$

donde  $N$  es el número de muestras de la imagen (píxeles),  $N = nm$ , y  $x_i$  es la intensidad del  $i$ -ésimo elemento de la muestra o píxel <sup>3</sup>.

En algunos casos será más conveniente la representación unidimensional de una imagen, como veremos más adelante.

Cada píxel en un mapa de bits contiene cierta información, generalmente interpretada como información de color. El contenido de la información es siempre del mismo tipo para todos los píxeles en un mapa de bits (o vector) en particular [51].

La cantidad de información sobre el color puede ser la que requiera la aplicación, pero existen algunos estándares; los principales se describen a continuación.

### 1.2.1. 1 bit: blanco o negro

Este es el contenido de información más pequeño posible que se puede guardar para cada píxel. El mapa de bits resultante se denomina monocromo o blanco y negro. Los píxeles con un 0 serán negros, mientras que los píxeles con un 1, blancos, aunque si bien solo son posibles dos estados, éstos podrían interpretarse como dos colores cualesquiera, 0 se asigna a un color y 1 se asigna a otro color.

<sup>2</sup>También podría hacerse por columnas.

<sup>3</sup> $x_1, x_2, \dots, x_m$  corresponde a la primer fila de la representación bidimensional,  $x_{m+1}, x_{m+2}, \dots, x_{2m}$  a la segunda y así sucesivamente.

### 1.2.2. Escala de grises con 8 bits

En este caso, cada píxel toma 1 byte (8 bits) de almacenamiento, lo que da como resultado 256 estados diferentes. Si estos estados se asignan a una escala de grises de negro a blanco, como se muestra en la Figura 1.3, el mapa de bits se denomina imagen en escala de grises. Por convención, 0 es normalmente negro y 255 blanco. Los niveles de gris son los números intermedios, por ejemplo, en una escala lineal, 127 será un nivel de gris del 50 %.



Figura 1.3: Escala de grises

El **rango dinámico** es la capacidad de captar el detalle en las luces y en las sombras dentro una misma imagen. El número de bits utilizado por píxel, determina el máximo rango dinámico  $L$  que se puede representar fielmente; por ejemplo, para imágenes en escala de grises de 8 bits el rango va de 0 a 255, siendo  $L = 255$  el máximo rango dinámico.

### 1.2.3. Formato RGB de 24 bits

Para representar una imagen en color pueden utilizarse tres bitmaps, también conocidos como tres bandas, a las que denominaremos R, G, B. Ahora hay 8 bits asignados a cada componente, rojo (R), verde (G) y azul (B), como se muestra en la Figura 1.4. En cada componente, el valor 0 se refiere a ninguna contribución de ese color, mientras que 255 se refiere a la contribución totalmente saturada de ese color. Como cada componente tiene 256 estados diferentes, hay un total de 16,777,216 colores posibles.

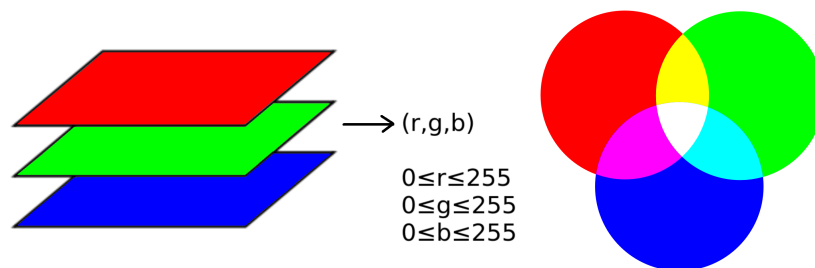


Figura 1.4: El espacio del color RGB

En el espacio RGB, cualquier color se representa como un punto dentro de un cubo de color con los ejes ortogonales R, G, B, como puede apreciarse en la Figura 1.5. Esta idea es un concepto fundamental en gráficos por computadora.

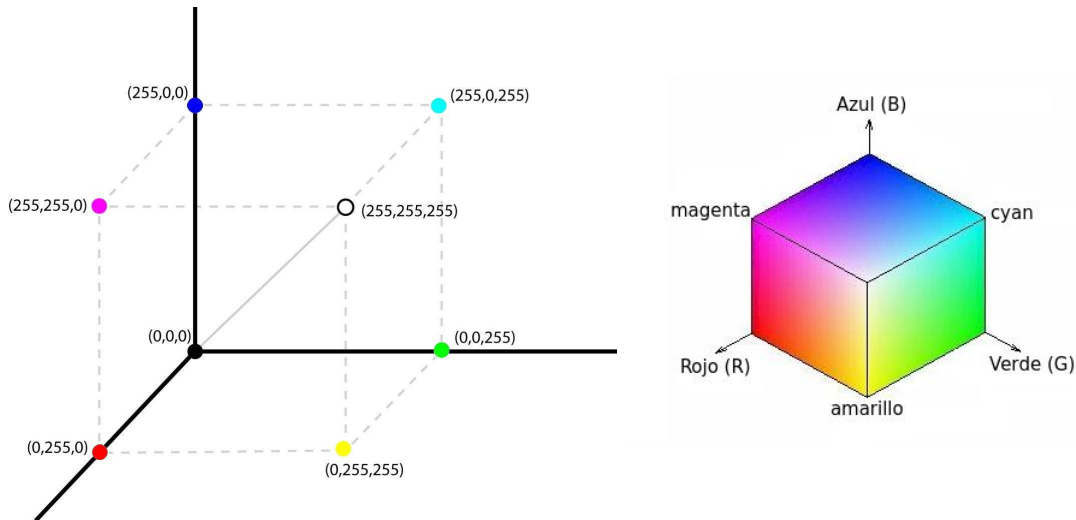


Figura 1.5: Cubo RGB. Los valores grises forman una línea recta de negro a blanco a lo largo de la diagonal del cubo,  $R = G = B$ .

Existen otros métodos para representar imágenes digitales, pero los hasta aquí descritos serán suficiente para abordar el trabajo que nos proponemos.

### 1.3. El sistema visual humano (HVS)

Un concepto fundamental para el trabajo que nos proponemos es el de Sistema Visual Humano (HVS por sus siglas en inglés *Human Visual System*) y su funcionamiento.

Es útil concebir al HVS como un sistema de información. Existen modelos del HVS que describen el funcionamiento de este sistema, generalmente implementados en un proceso secuencial como se muestra en la Figura 1.6 [52].



Figura 1.6: Diagrama de bloques de un modelo HVS típico

Estos modelos deben dar cuenta de ciertos efectos psicofísicos de la visión humana. Pero la mayoría de las propiedades visuales del HVS no son intuitivas. Incluso cuando se han caracterizado por experimentos psicofísicos, la evidencia fisiológica es la única forma de comprender el fenómeno por completo, esta anatomía básica del HVS sería como el “hardware” del sistema, aunque no nos adentraremos en su descripción (para ampliar puede consultar [4]), en cambio nos restringiremos a exponer las características más importantes del HVS.

### 1.3.1. Características psicofísicas del HVS

- ***Sensibilidad al contraste***

La sensibilidad al contraste es la capacidad de percibir ligeros cambios en la luminancia entre regiones que no están separadas por fronteras definidas y es tan importante como la capacidad de percibir contornos nítidos de objetos relativamente pequeños.

Uno de los problemas más importantes en el modelado de HVS se refiere a la disminución de la sensibilidad para frecuencias espaciales más altas. Este fenómeno se parametriza mediante la función de sensibilidad de contraste (CSF). El modelado correcto del CSF es especialmente difícil para imágenes en color. La sensibilidad humana al contraste también depende de la frecuencia temporal de los estímulos. La interacción entre frecuencias espaciales y temporales se puede describir mediante funciones de sensibilidad de contraste espacio-temporal [52].

- ***Adaptación a la luz***

Para el HVS la luminancia no es lineal, este fenómeno es conocido como *adaptación a la luz* o *enmascaramiento de la luminancia*. La percepción humana de la luminancia obedece a la ley de Weber que puede ser descripta como:

$$\frac{\Delta I}{I} = K \quad (1.2)$$

donde  $I$  es la luminancia de fondo,  $\Delta I$  es un incremento notable de la luminancia y  $K$  es una constante denominada fracción de Weber.

Esta ley se cumple para un amplio rango de luminancias y deja de valer cuando la luz es demasiado baja o alta.

- ***Efecto de enmascaramiento***

El enmascaramiento ocurre cuando un estímulo que es visible por sí mismo no puede ser detectado debido a la presencia de otro. A veces ocurre el efecto opuesto, la facilitación: un estímulo que no es visible por sí mismo puede detectarse debido a la presencia de otro. En el marco del procesamiento de imágenes, es útil pensar que la distorsión o el ruido de decodificación es enmascarado (o facilitado) por la imagen original o secuencia de fondo. El enmascaramiento explica por qué las distorsiones similares son perturbadoras en ciertas regiones de una imagen, mientras que apenas notable en otro lugar, un ejemplo de este fenómeno puede verse en la Figura 1.7.

- ***Visión foveal***

Debido a las distribuciones no uniformes de los receptores de cono y las células ganglionares en la retina, cuando se fija en un punto de la imagen o escena, ese punto de fijación y la

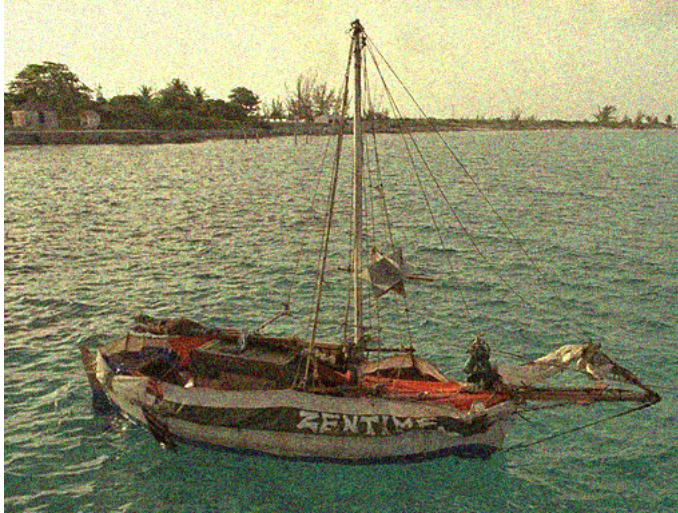


Figura 1.7: Efecto de enmascaramiento. Imagen de la base TID 2013 con ruido añadido. Para el observador el ruido es mucho más notorio en el cielo que en el agua debido al efecto de enmascaramiento.

zona que lo rodea se muestran con resolución espacial más alta, esta resolución disminuye rápidamente en las regiones distantes al punto de fijación. En la Figura 1.8 puede apreciarse un ejemplo de este efecto visual. En el lenguaje de la ciencia de la visión, la visión de alta resolución cerca del centro de fijación de un observador se denomina *visión foveal*, mientras que la visión progresivamente de menor resolución lejos de la fijación se denomina *visión periférica*.



Figura 1.8: Visión Foveal. A la izquierda la imagen original, a la derecha la imagen con el efecto de visión foveal, donde el punto de fijación es el ojo de la mujer.

Teniendo en cuenta estas, entre otras, propiedades y limitaciones del HVS se pueden diseñar aplicaciones de procesamiento de imágenes, por ejemplo medidas IQA-objetivas, más eficientes.

---

## Capítulo 2

# Medidas de Calidad de una Imagen

Las propuestas de índices, métricas y pseudo-métricas para evaluar similaridad y calidad en imágenes son muy amplias y se encuentran en permanente desarrollo. En este capítulo se describirán algunas de las medidas objetivas de evaluación de calidad y similaridad en imágenes más utilizadas.

Tres tipos de información pueden utilizarse para el diseño de una medida objetiva de calidad de una imagen: información sobre la imagen original (no siempre disponible), información sobre el tipo de proceso de distorsión aplicado sobre la imagen y conocimiento sobre cómo opera el sistema visual humano (HVS). Luego, estos procedimientos se clasifican en tres grupos en función de la disponibilidad de la imagen de referencia: a) Referencia completa (FR-IQA: Full Reference Image Quality Assessment), b) Referencia reducida (RR-IQA: Redude Reference Image Quality Assesment) y c) Referencia nula (NR-IQA: Null Reference Image Quality Assesment) [11], [12].

### 2.1. Medidas de referencia completa (FR-IQA: Full Reference Image Quality Assesment)

En términos generales hay dos estrategias para el diseño de FR-IQA. La primera estrategia (bottom-up framework) simula las diversas etapas de procesamiento del sistema visual humano (HVS), incluido el efecto de enmascaramiento visual (visual masking effect), la sensibilidad de contraste, las diferencias apenas perceptibles, etc. Sin embargo el HVS es muy complejo y nuestro conocimiento actual sobre él dista mucho de ser suficiente para construir una medida IQA-objetiva precisa, acorde a esta estrategia. La segunda estrategia (top-down framework) tiene como objetivo modelar la función general de HVS en base a algunas suposiciones globales sobre el mismo. Las métricas que describiremos a continuación se basan en su mayoría en esta segunda estrategia.



### 2.1.1. El error cuadrático medio (MSE) y la relación de cociente señal-ruido

Una de las medidas más sencilla y ampliamente utilizada para calcular la calidad de una imagen, es sin duda el MSE. Esta herramienta mide esencialmente la diferencia de intensidad promedio entre una imagen distorsionada (imagen  $y$ ) y otra de referencia (imagen  $x$ ). Está dada por:

$$\text{MSE} = \frac{1}{nm} \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} (x_{i,j} - y_{i,j})^2 \quad (2.1)$$

donde las imágenes  $x$ ,  $y$  están representadas como arreglos bidimensionales de tamaño  $n \times m$ ,  $x_{i,j}$  e  $y_{i,j}$  indican respectivamente la intensidad de las imágenes en la posición  $(i, j)$ .

Dado que esta medida no utiliza información posicional de la imagen, puede ser definida utilizando una representación vectorial unidimensional de  $x$  e  $y$  como se describió en el capítulo anterior:  $x = \{x_i : i = 1, 2, \dots, N\}$  e  $y = \{y_i : i = 1, 2, \dots, N\}$ , donde  $x_i$  e  $y_i$  indican respectivamente la intensidad de las imágenes  $x$  e  $y$  en la posición  $i$  y  $N$  es el número total de píxeles de ambas imágenes,  $N = nm$ . Por lo tanto el MSE se calcula de la siguiente manera:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (2.2)$$

Algunos investigadores prefieren utilizar la raíz del error cuadrático medio, esta métrica se denota RMSE:

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (2.3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (2.4)$$

Otra medida de similaridad emparentada con el MSE es la relación cociente señal ruido.

Si consideramos que para obtener la imagen distorsionada  $y$  se agregó un ruido (distorsión) a la imagen original o de referencia  $x$ , es decir  $y = x + r$  (visto por píxel  $y_i = x_i + r_i$ ), la relación de cociente señal ruido, SNR (por sus siglas en inglés Signal to Noise Ratio) se define como la relación entre la potencia de la señal y la potencia del ruido:

$$\text{SNR} = 10 \log_{10} \left( \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N r_i^2} \right) \quad (2.5)$$

Esta definición es una adaptación directa de la definición física de potencia eléctrica.

Notar que la potencia es la norma al cuadrado de la señal, la cual muestra la media de la

desviación cuadrada que tiene la señal respecto de cero.

Entonces, si

$$|x|^2 = \sum_{i=1}^N x_i^2 \quad (2.6)$$

luego,

$$\text{SNR} = 10 \log_{10} \left( \frac{|x|^2}{\sum_{i=1}^N (x_i - y_i)^2} \right) \quad (2.7)$$

donde  $|x|^2$  es el cuadrado de la norma de la señal (imagen) de referencia.

Por lo tanto, una formula equivalente para el SNR es:

$$\text{SNR} = 10 \log_{10} \left( \frac{|x|^2}{N \cdot \text{MSE}} \right) \quad (2.8)$$

En general, las métricas basadas en la relación de cociente señal-ruido pueden ser definidas en base al MSE.

El PSNR (Peak Signal to Noise Ratio), se define similarmente al SNR, pero utilizando el cuadrado de la máxima intensidad de la señal en lugar del promedio del cuadrado de cada píxel:

$$\text{PSNR} = 10 \log_{10} \left( \frac{L^2}{\text{MSE}} \right) \quad (2.9)$$

donde  $L$  es el máximo rango dinámico (255 para imágenes en escala de grises de 8 bits).

Este criterio es menos estricto y se verifica que  $\text{PSNR} \geq \text{SNR}$ , solo serán iguales si la señal original es constante y con amplitud máxima. Hay que tener en cuenta que aunque la varianza de una señal constante es nula, su potencia no lo es; el nivel de dicha señal constante hace una diferencia en SNR pero no en PSNR.

En el caso de SNR estamos viendo qué tan fuerte es la señal y qué tan fuerte es el ruido, suponiendo que no hay circunstancias especiales. En el caso de PSNR, interesa el pico de la señal porque podemos estar interesados en el ancho de banda de la señal o el número de bits que necesitamos para representarla. Esto es mucho más específico que la SNR pura y puede encontrar muchas aplicaciones razonables, por ejemplo en la compresión de imágenes. Para el PSNR lo importante es qué tan bien surgen las regiones de alta intensidad de la imagen a través del ruido, y presta mucha menos atención a cómo se desempeña en baja intensidad.

Otra medida es el WSNR (weighted signal to noise ratio) que agrega pesos, dado que del sistema sensorial, sabemos que en el oído o la visión, no todas las frecuencias tienen el mismo peso perceptual.

$$\text{WSNR} = 10 \log_{10} \left( \frac{\sum_{i=1}^N w_i x_i^2}{\sum_{i=1}^N w_i (x_i - y_i)^2} \right) \quad (2.10)$$

donde  $w_i$  le da menor peso a las frecuencias menos importantes.

Para mayores detalles sobre estas medidas se puede consultar [12] y [13].

### 2.1.2. Medida de calidad del ruido (NQM)

Las medidas SNR y PSNR asumen que la distorsión solo es causada por ruido aditivo independiente de la señal, en [13] se plantea que es necesario tener en cuenta también otras fuentes de degradación. Apoyandose en la observación de que los efectos psicovisuales del filtrado y el ruido están separados, se definió la métrica NQM (Noise Quality Measure) utilizando el modelo de contrastes de Peli [14].

En lugar de calcular una imagen residual  $x - y$  como se hizo para las métricas SNR y PSNR, se calcula una imagen restaurada modelo procesando la imagen original a través de un algoritmo de restauración (que se basa en el modelo de contrastes de Peli) utilizando los mismos parámetros que se utilizaron al restaurar la imagen degradada.

Entonces, si  $O_s(x, y)$  es la imagen restaurada modelo obtenida a partir de la imagen original y  $I_s(x, y)$  es la imagen restaurada obtenida a partir de la imagen distorsionada, NQM se define como:

$$\text{NQM} = 10 \log_{10} \left( \frac{\sum_x \sum_y O_s^2(x, y)}{\sum_x \sum_y (O_s(x, y) - I_s(x, y))^2} \right) \quad (2.11)$$

Notar que esta fórmula es equivalente a la presentada en (2.7) utilizando  $O_s$  e  $I_s$  en lugar de  $x$  e  $y$ .

NQM demostró un mejor rendimiento que PSNR para imágenes distorsionadas por ruidos aditivos.

Las medidas expuestas hasta el momento pueden interpretarse fácilmente y son simples de calcular, pero ignoran por completo la influencia sustancial de la posición del píxel en la imagen en la evaluación de la calidad y/o similaridad, lo que determina que su correlación con la evaluación del sistema visual humano (MOS) sea pobre [12], Para superar estas limitaciones Wang et al. propusieron una medida llamada índice de similaridad estructural (SSIM por sus siglas en inglés Structural Similarity Index) [15], [16].

### 2.1.3. Índice de similaridad estructural (SSIM)

Esta medida se basa en la razonable hipótesis de que la percepción visual humana está fuertemente adaptada para extraer información estructural de una escena y por ende, una medida de la similaridad estructural es una buena aproximación de la calidad percibida en una imagen.

El índice SSIM es una función entre dos imágenes  $x$  e  $y$ ; si asumimos que una de estas imágenes tiene calidad perfecta, a la cual llamaremos imagen de referencia, el índice SSIM puede ser

interpretado como una medida de calidad de la otra imagen.

Este índice separa la medición de la similaridad entre dos imágenes en tres componentes: la luminancia, el contraste y la similaridad estructural de la siguiente manera:

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (2.12)$$

donde  $l$ ,  $c$  y  $s$  definen respectivamente la relación de luminancia, contraste y comparación estructural entre  $x$  e  $y$ , mientras que  $\alpha$ ,  $\beta$ ,  $\gamma$  son parámetros que ajustan la relativa importancia de cada componente; en general y por convención se establece que  $\alpha = \beta = \gamma = 1$  como una forma de reducir la cantidad de parámetros en (2.12).

La luminancia local de cada señal (ventanas tomadas en las mismas posiciones en  $x$  e  $y$  respectivamente) se estima mediante la intensidad media:

$$\hat{\mu}_x = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.13)$$

La función de comparación entre luminancias  $l(x, y)$  es entonces, una función de  $\hat{\mu}_x$  y  $\hat{\mu}_y$ :

$$l(x, y) = l(\hat{\mu}_x, \hat{\mu}_y) = \frac{2\hat{\mu}_x\hat{\mu}_y + C_1}{\hat{\mu}_x^2 + \hat{\mu}_y^2 + C_1} \quad (2.14)$$

donde  $C_1$  es una constante que se incluye para evitar inestabilidad si  $\hat{\mu}_x^2 + \hat{\mu}_y^2$  es muy cercano a 0, específicamente:  $C_1 = (K_1L)^2$ , donde  $L$  es el máximo rango dinámico y  $K_1 \ll 1$  es una constante pequeña.

Si extraemos la intensidad media de la señal:  $x - \hat{\mu}_x$ , la desviación estandar  $\hat{\sigma}_x$  es una estimación del contraste de la señal.

$$\hat{\sigma}_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu}_x)^2} \quad (2.15)$$

La comparación de contrastes  $c(x, y)$  es entonces, una función de  $\hat{\sigma}_x$  y  $\hat{\sigma}_y$ :

$$c(x, y) = c(\hat{\sigma}_x, \hat{\sigma}_y) = \frac{2\hat{\sigma}_x\hat{\sigma}_y + C_2}{\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + C_2} \quad (2.16)$$

donde  $C_2 = (K_2L)^2$  con  $K_2 \ll 1$  se introduce para evitar la división por valores muy cercanos a 0.

Finalmente, para realizar la comparación estructural, normalizamos la señal por su media y su desviación estandar, es decir, reemplazamos las imágenes  $x$  e  $y$  por  $\frac{x-\hat{\mu}_x}{\hat{\sigma}_x}$  e  $\frac{y-\hat{\mu}_y}{\hat{\sigma}_y}$  respectivamente.

Como la correlación entre  $\frac{x-\hat{\mu}_x}{\hat{\sigma}_x}$  e  $\frac{y-\hat{\mu}_y}{\hat{\sigma}_y}$  y entre  $x$  e  $y$  es la misma, luego la componente estructural puede ser definida de la siguiente manera:

$$s(x, y) = \frac{\hat{\sigma}_{xy} + C_3}{\hat{\sigma}_x \hat{\sigma}_y + C_3} \quad (2.17)$$

donde  $C_3 = C_2/2$  es una constante para evitar divisiones por valores cercanos a 0 y  $\hat{\sigma}_{xy}$  es la covarianza entre  $x$  e  $y$ :

$$\hat{\sigma}_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) \quad (2.18)$$

$\hat{\sigma}_{xy}$  mide la tendencia de  $x$  e  $y$  de variar juntas y es por lo tanto una indicadora de la similaridad estructural.

Es importante resaltar que estas tres componentes son relativamente independientes, lo cual tiene un sentido físico, pues los cambios en la luminancia y el contraste tienen escaso impacto en la estructura de los objetos en la escena visual.

Luego, de acuerdo a la ecuación (2.12), la fórmula para el SSIM será entonces:

$$\text{SSIM}(x, y) = \frac{(2\hat{\mu}_x \hat{\mu}_y + C_1)(2\hat{\sigma}_{xy} + C_2)}{(\hat{\mu}_x^2 + \hat{\mu}_y^2 + C_1)(\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + C_2)} \quad (2.19)$$

En el caso en que  $C_1 = C_2 = 0$  (es decir  $K_1 = K_2 = 0$ ), esta medida se denomina índice de calidad universal UQI (por sus siglas en inglés Universal Quality Index):

$$\text{UQI} = \frac{4\hat{\sigma}_{xy} \hat{\mu}_x \hat{\mu}_y}{(\hat{\sigma}_x^2 + \hat{\sigma}_y^2)(\hat{\mu}_x^2 + \hat{\mu}_y^2)} \quad (2.20)$$

Este índice se mueve en el rango de  $[-1, 1]$ , toma el valor 1 cuando las dos imágenes son idénticas, toma el valor  $-1$  si  $y_i = 2\bar{x} - x_i, \forall i = 1, 2, \dots, N$

UQI fue la versión preliminar del SSIM propuesta en [15], si bien se desempeña mejor que el MSE, produce resultados inestables cuando  $\hat{\mu}_x^2 + \hat{\mu}_y^2$  o  $\hat{\sigma}_x^2 + \hat{\sigma}_y^2$  son muy cercanas a 0, SSIM supera esta cuestión al introducir las constantes  $C_1$  y  $C_2$ ; en [16] se propone utilizar  $K_1 = 0,01$  y  $K_2 = 0,03$ .

El índice SSIM verifica las siguientes propiedades que son deseables en una medida de similitud:

1. Simetría:  $\text{SSIM}(x, y) = \text{SSIM}(y, x)$ . Cambiar el orden de las señales no afecta la medida.
2. Está acotado superiormente:  $\text{SSIM}(x, y) \leq 1$ . Una cota superior indica cuán cerca están dos señales de ser perfectamente idénticas. Notar que otras medidas de similaridad, como por ejemplo la SNR, no son acotadas.
3. Unicidad del máximo:  $\text{SSIM}(x, y) = 1$  si y solo si  $x = y$ . La medida perfecta solo se obtiene en el caso que las señales sean idénticas, es decir, esta medida de similaridad cuantifica cualquier variación que pueda existir entre las señales [4].

El índice SSIM superó ampliamente a los índices MSE y PSNR en la base de datos LIVE [5]. Pero a pesar del éxito que tuvo en esta base de datos, su desempeño en bases de datos más recientes (como la TID2008 [6], TID2013 [7] [8] y CSIQ [9]) dista de ser ideal. Por este motivo se han desarrollado otros índices FR-IQA, que se presentan como optimizaciones del índice SSIM; por ejemplo Multi-Scale SSIM (MS-SSIM) [17], [53], [54], [55], [56] y el índice WSSIM [18] que agrega pesos a los factores que intervienen en la definición del índice SSIM usando un mapeo de prominencias <sup>1</sup>, en sus tres versiones:  $S_NW$ -SSIM [57], [58], IW-SSIM (statistical information-content weighted SSIM) [19] y la SW-SSIM (structural similarity weighted SSIM) [20].

#### 2.1.4. Multi-Scale SSIM (MS-SSIM)

Esta forma más avanzada de SSIM, es un método conveniente para incorporar detalles de las imágenes en diferentes resoluciones. En [17] se presentó un método de multiescala para el SSIM que iterativamente aplica un filtro de paso bajo y luego realiza un downsamplig (aumento de resolución) de factor 2 a la imagen filtrada. La imagen distorsionada original se considera la escala 1, la escala más alta es la  $M$ , la cual se obtiene luego de  $M - 1$  iteraciones.

Para la  $j$ -ésima escala se calculan la comparación de contraste, formula (2.16), y la comparación estructural, formula (2.17), que denotaremos como:  $c_j(x, y)$  y  $s_j(x, y)$  respectivamente. La comparación luminancia se calcula con la fórmula (2.14) solo para la escala  $M$  y se denota por  $l_M(x, y)$ .

El MSSIM se obtiene entonces combinando estos componentes en diferentes escalas:

$$\text{MSSIM}(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (2.21)$$

Al igual que en (2.12),  $\alpha_M$ ,  $\beta_j$  y  $\gamma_j$  se usan para ajustar la relativa importancia de cada componente; sólo por simplicidad, para todo  $j$ ,  $\alpha_M = \beta_j = \gamma_j$  y los parámetros de distintas escalas se relacionan entre si mediante:  $\sum_{j=1}^M \gamma_j = 1$ ; queda entonces, determinar la importancia relativa entre las escalas. En [17] se propone una calibración. El SSIM es un caso especial del MSSIM utilizando una sola escala.

Esta medida también verifica las tres propiedades presentadas en la sección anterior.

#### 2.1.5. Medidas direccionales: El Índice CQ

En el año 2012, Ojeda et al. [21] propusieron utilizar el coeficiente de codispersión espacial en lugar del coeficiente de correlación en la definición del índice SSIM. Esta iniciativa dio lugar a una

<sup>1</sup>Un mapa de prominencia (saliency map) es un tipo de segmentación de imágenes. El resultado del mapa es un conjunto de contornos extraídos de la imagen donde cada uno de los píxeles en una región es similar con respecto a alguna característica o propiedad calculada, como el color, la intensidad o la textura. El objetivo es simplificar la representación de una imagen por algo que sea más significativo y fácil de analizar.

nueva y prometedora medida IQA objetiva a la que se denominó índice CQ.

Definimos  $\gamma(h)$ , el variograma cruzado de los procesos estacionarios débiles  $X(t)$  y  $Y(t)$ , con  $t \in D \subset \mathbb{Z}^d$ ,  $d \in \mathbb{N}$  como:

$$\gamma(h) = \mathbb{E}[(X(t+h) - X(t))(Y(t+h) - Y(t))] \quad (2.22)$$

donde  $t, t+h \in D$ .

El coeficiente de codispersión es una normalización de  $\gamma(h)$ :

$$\rho(h) = \frac{\gamma(h)}{\sqrt{V_X(h)V_Y(h)}} \quad (2.23)$$

donde  $V_X(h) = \mathbb{E}(X(t+h) - X(t))^2$  y similarmente para  $V_Y(h)$ .

Es importante notar que  $\gamma(h)$ , y por lo tanto  $\rho(h)$ , también depende de  $X(t)$  e  $Y(t)$ , aunque por simplicidad solo aparezca  $h$  en la notación.

Se verifica que:

$$|\rho(h)| \leq 1 \quad (2.24)$$

En condiciones regulares se puede calcular el coeficiente de codispersión muestral:

$$\hat{\rho}(h) = \frac{\sum_{t,t+h \in D'} a_t b_t}{\sqrt{\hat{V}_X(h)\hat{V}_Y(h)}} \quad (2.25)$$

con  $t = (t_1, t_2)$ ,  $h = (h_1, h_2)$ ,  $D' \subseteq D$ ,  $\#D' < \infty$ ,

$a_t = X(t_1 + h_1, t_2 + h_2) - X(t_1, t_2)$ ,  $b_t = Y(t_1 + h_1, t_2 + h_2) - Y(t_1, t_2)$ ,

$\hat{V}_X(h) = \sum_{t,t+h \in D'} a_t^2$  y  $\hat{V}_Y(h) = \sum_{t,t+h \in D'} b_t^2$

Es importante notar que  $\hat{\rho}(h)$ , también depende de  $X(t)$  e  $Y(t)$ , aunque por simplicidad solo se nombre a  $h$  en la notación.

Los autores demuestran que  $\hat{V}_X(h)$  y  $\hat{V}_Y(h)$  son consistentes y asintóticamente normales (propiedades estadísticas que permiten construir intervalos de confianza y test de hipótesis para el coeficiente).

Dadas dos secuencias:  $x = \{X(t) : t = 1, 2, \dots, N\}$  e  $y = \{Y(t) : t = 1, 2, \dots, N\}$ , el índice SSIM definido en (2.19) es capaz de capturar solo la asociación lineal entre  $x$  e  $y$  pero no puede explicar otros tipos de relaciones entre estas secuencias, por ejemplo, la asociación espacial en una dirección específica  $h$ . El coeficiente CQ se define de una forma similar al SSIM, buscando superar esta limitación, de la siguiente forma:

$$\text{CQ}_h(x, y) = l(x, y) \cdot c(x, y) \cdot \hat{\rho}(h) \quad (2.26)$$

donde  $l$  es la relación de luminancia,  $c$  el contraste, al igual que en la fórmula del SSIM, pero

se cambia la componente  $s$  definida en términos de la correlación, por una estimación de  $\rho(h)$ , el coeficiente de codispersión en la dirección  $h$ .

Este índice tiene en cuenta la asociación espacial en una dirección específica  $h$  entre una imagen degradada y la imagen original sin modificar, y permite detectar similaridad oculta en esa dirección. Además la propuesta posee interesantes propiedades matemáticas y se están analizando sus ventajas como estructura estadística [47].

Recientemente, en 2015, se propuso otro índice capaz de cuantificar la similitud local máxima entre dos imágenes, una vez evaluada la similaridad CQ en un cierto conjunto de direcciones de interés. Este índice, denominado CQ-max [22], posee al igual que CQ, propiedades matemáticas deseables y está en estudio su desempeño sobre diferentes bases de datos.

### 2.1.6. Medidas direccionales: FR-IQA basadas en la magnitud del gradiente

En los últimos años, muchos investigadores se han dado cuenta de la importancia de las características de bajo nivel para estudiar la similaridad entre imágenes, lo que ha dado lugar a las *FR-IQA basadas en la magnitud del gradiente* [25], [23].

El gradiente de una imagen  $f(i, j)$  en un punto  $(i, j)$  se define como un vector bidimensional dado por la ecuación (2.27):

$$G(f(i, j)) = (G_h, G_v) = \left( \frac{\partial f(i, j)}{\partial h}, \frac{\partial f(i, j)}{\partial v} \right) \quad (2.27)$$

donde  $G_h$  y  $G_v$  son las derivadas parciales de la función intensidad de la imagen en las direcciones horizontal y vertical respectivamente. Por lo tanto  $G$  apunta en la dirección de variación máxima de  $f$  en el punto  $(i, j)$  por unidad de distancia.

La magnitud del gradiente ( $G$ ) se calcula como:

$$G = \sqrt{G_h^2 + G_v^2} \quad (2.28)$$

Es decir, dada una imagen digital, la magnitud del gradiente se define como la raíz cuadrada de los cuadrados de los gradientes direccionales en dos direcciones ortogonales, estas direcciones pueden ser, por ejemplo, los ejes de la convención de coordenadas definida en el capítulo anterior, ver Figura 1.2.

Muchas medidas FR-IQA siguen un modelo de dos pasos: primero se calcula un mapa de calidad local LQM (por sus siglas en inglés Local Quality Map) al comparar localmente la imagen distorsionada con la de referencia a través de una función de similaridad (LQM es una nueva imagen que refleja la calidad local de cada zona de la imagen distorsionada). Finalmente se calcula el índice de calidad a partir de la imagen LQM, por ejemplo promediando todos los valores de sus píxeles. Este proceso de dos pasos se describe en la Figura 2.1. La operación realizada en el



último paso para obtener el valor final que determina el índice, es denominada *estrategia de pooling* en la literatura (del inglés *pooling strategy*). Muchas de las métricas basadas en la magnitud del gradiente siguen este esquema y utilizan G para definir un LQM.

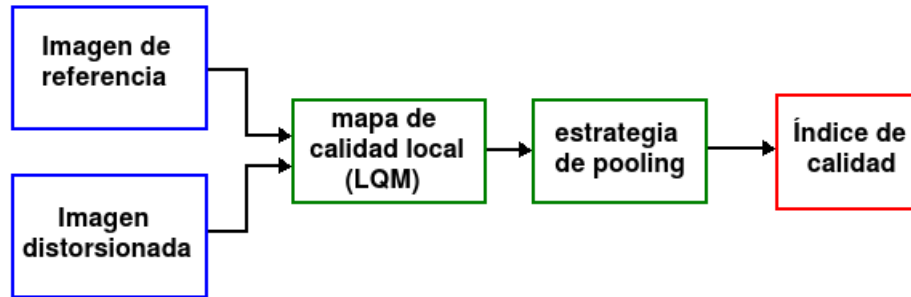


Figura 2.1: Modelo de dos pasos para FR-IQA

En [23] se exponen dos métricas que siguen este modelo, las presentaremos a continuación.

El gradiente de una imagen se calcula usualmente convolucionando la imagen con un filtro lineal <sup>2</sup>, por ejemplo con un filtro de Roberts, Sobel, Scharr [46] o Prewitt [59].

En [23] se propone utilizar un filtro de Prewitt (aunque otros filtros arrojan resultados similares), el mismo se define de la siguiente manera:

$$h = \begin{bmatrix} \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix}$$

en la dirección horizontal, y

$$v = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

en la dirección vertical.

La convolución se realiza píxel a píxel como se ejemplifica en la Figura 2.2.

Como puede verse en el ejemplo que ilustra la Figura 2.2, el filtro lee sucesivamente, de izquierda a derecha y de arriba a abajo, todos los píxeles del área de acción. Multiplica el valor de cada uno de ellos por el valor correspondiente del kernel y suma los resultados. El nuevo píxel obtenido tiene un valor de 7:  $46 \cdot \frac{1}{3} + 42 \cdot 0 + 40 \cdot -\frac{1}{3} + 55 \cdot \frac{1}{3} + 50 \cdot 0 + 46 \cdot -\frac{1}{3} + 58 \cdot \frac{1}{3} + 56 \cdot 0 + 52 \cdot -\frac{1}{3} = 7$ .

Para obtener el siguiente valor, la máscara se mueve a la siguiente posición del nuevo píxel y se repite el proceso, después de haber barrido todas las posibles posiciones, el resultado es una imagen

<sup>2</sup>Una convolución es un operador matemático que transforma dos funciones  $f$  y  $g$  en una tercera función que en cierto sentido representa la magnitud en la que se superponen  $f$  y una versión trasladada e invertida de  $g$ . En procesamiento de imágenes el operador que se utiliza para convolucionar la función de intensidad de una imagen, se denomina filtro, máscara o kernel de convolución y usualmente puede ser representado por una matriz  $3 \times 3$ .

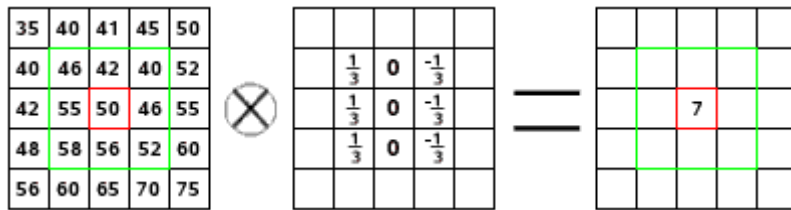


Figura 2.2: Ejemplo de convolución. A la izquierda la matriz que representa parte de una imagen; el píxel al cual se le aplicará la operación de convolución se encuentra indicado con rojo; el área de acción del filtro está delimitada por un borde verde. En el medio, el kernel, y a la derecha, el resultado de convolución.

gradiente. Es preciso tener en cuenta que en los bordes de la imagen los valores del gradiente no se pueden calcular por sobrepasar las máscaras de la propia imagen.

Convolucionando con  $h$  y  $v$  la imagen de referencia  $x$ , obtenemos respectivamente el gradiente horizontal  $G_h$  y el gradiente vertical  $G_v$  de la imagen  $x$ .

$$G_h(x) = x \otimes h \tag{2.29}$$

$$G_v(x) = x \otimes v \tag{2.30}$$

donde el signo  $\otimes$  indica la operación de convolución.

El **mapa de magnitud del gradiente** de la imagen  $x$  se calcula de acuerdo a la fórmula (2.28) de la siguiente manera:

$$Gx(i) = \sqrt{(x \otimes h)^2(i) + (x \otimes v)^2(i)} \tag{2.31}$$

donde  $i$  indica la ubicación o píxel  $i$  en la imagen  $x$ <sup>3</sup>.

Este mapa es básicamente una nueva imagen en escala de grises, obtenida a partir de la imagen  $x$ .

Análogamente para la imagen distorsionada  $y$ :

$$Gy(i) = \sqrt{(y \otimes h)^2(i) + (y \otimes v)^2(i)} \tag{2.32}$$

Con los mapas de la magnitud del gradiente de las imágenes  $x$  e  $y$ , se define el **mapa de similitud de la magnitud del gradiente** (**GMS** por sus siglas en inglés: Gradient Magnitude Similarity Map) de la siguiente manera:

$$GMS(i) = \frac{2Gx(i)Gy(i) + C}{(Gx(i))^2 + (Gy(i))^2 + C} \tag{2.33}$$

donde  $C$  es una constante positiva para obtener estabilidad numérica (evitar divisiones por 0).

<sup>3</sup>Notar que aquí se está utilizando la representación unidimensional de la imagen.

Claramente, en el caso  $x = y$ ,  $\text{GMS}(i) = 1$ , que es el mayor valor posible.

Este GMS es utilizado como LQM. La estrategia de pooling más sencilla es tomar el promedio de los valores de calidad local como el índice de calidad final, la métrica GSM (Gradient Magnitude Similarity Mean) sigue esta estrategia, por lo tanto se define como:

$$\text{GSM} = \frac{1}{N} \sum_{i=1}^N \text{GMS}(i) \quad (2.34)$$

donde  $N$  es el número total de píxeles de la imagen. Un valor de GSM mayor indica mayor calidad en la imagen.

El proceso de promediar asume que todos los píxeles tienen la misma importancia en la imagen, sin embargo, diferentes regiones contribuyen de distinta manera en la calidad global de una imagen. Muchos investigadores han dedicado esfuerzos a asignar un peso a los valores de calidad local antes de promediarlos, esto daría un resultado un poco más preciso, aunque puede ser computacionalmente muy costoso.

Una imagen generalmente tiene una variedad de estructuras locales en su escena, cuando la imagen es distorsionada, las distintas estructuras locales sufrirán diferentes degradaciones en la magnitud del gradiente. La métrica GMSD (Gradient Magnitude Similarity Deviation), propuesta en [23], señala que la distribución espacial de los niveles de distorsión tiene impacto en la calidad de percepción, es decir, la distorsión distribuida desigualmente degrada la calidad visual con mayor severidad. GMSD utiliza la desviación estandar como estrategia de pooling para obtener el índice de calidad final, basándose en la idea de que la variación global de la calidad local de la imagen es un reflejo de la calidad final percibida en la misma. Esta métrica se define como:

$$\text{GMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{GMS}(i) - \text{GSM})^2} \quad (2.35)$$

El valor de GMSD refleja el rango de severidad de distorsión de una imagen, por lo tanto, cuanto mayor sea el GMSD, mayor el rango de distorsión y por lo tanto menor la calidad percibida. Esta métrica es tanto eficiente computacionalmente, como muy efectiva para predecir calidad.

Otras medidas basadas en la magnitud del gradiente son: el FSIM (Feature Similarity Index) [24] que utiliza la fase de congruencia y la magnitud del gradiente para caracterizar la calidad de imagen local. El GSIM (Gradient Similarity Index) [25] que mide los cambios del gradiente de similitud entre imágenes. Un modelo local sintonizado-global (local-tuned-global model) que utiliza información del gradiente fue propuesto en [60].

### 2.1.7. Fidelidad a la información visual (VIF)

Al considerar a IQA como un problema de teoría de la información, la métrica VIF (Visual Information Fidelity) [26] está definida como el cociente entre la información en común de las dos imágenes y la información de la primera imagen:

$$\text{VIF} = \frac{I(x, y + n)}{I(x, x + n)} \quad (2.36)$$

donde  $n$  es un ruido gaussiano blanco estacionario con varianza  $\sigma_n^2$ .

El denominador representa la cantidad de información que el HVS puede extraer de la imagen original, mientras que el numerador representa la cantidad de información que el HVS puede extraer de la imagen distorsionada. El cociente entre esas dos cantidades, es por lo tanto, una medida de la cantidad de información en la imagen distorsionada relativa a la imagen de referencia y demostró estar bien correlacionada con la calidad visual.

Otras medidas desarrolladas en el marco de FR-IQA son: las métricas GES [27] y LGPS [28] que fueron propuestas en el dominio de la transformada de Gabor. MAD (Most apparent distortion) [29] que utiliza estrategias basadas en la apariencia. La medida IPAD (Image quality model based on phase and amplitude differences) [30] que utiliza el análisis de amplitud y fase. Teorías basadas en las neurociencias hicieron grandes aportes al diseño IQA, por ejemplo la medida IGM (internal generative mechanism) [31], que utiliza parámetros psicofísicos [17].

## 2.2. Medidas de referencia reducida (RR-IQA: Reduced Reference Image Quality Assessment)

Para cuando la disponibilidad de información de referencia es parcial (RR-IQA) se ha propuesto el índice FEDM (Free energy based distortion metric) [32] basado en el principio de energía libre que explica y unifica varias teorías físicas y biológicas sobre el comportamiento, percepción y aprendizaje. Este principio asume como hipótesis básica que el proceso cognitivo humano es controlado por un modelo generativo interno del cerebro. Este modelo generativo es básicamente un modelo probabilístico que puede separarse en un término potencial y un término a priori. La percepción visual consistiría en invertir este término potencial para inferir las posibles escenas a posteriori. Dado que el modelo generativo no es universal siempre existirá una brecha entre la escena y la predicción. El índice FEDM simula el modelo generativo interno del cerebro humano.

Otros índices de RR-IQA son RRED (RR entropic-difference indexes) [33], FTQM (Fourier Transform based quality measure) [34]. También existen medidas basadas en el SSIM, RR-SSIM [35] y SDM (Structural degradation model) [36].

## 2.3. Medidas de referencia nula (NR-IQA: Null Reference Image Quality Assessment)

En la mayoría de los sistemas de comunicación visual, no se cuenta con ningún tipo de información sobre la imagen de referencia. Para abordar este problema NR-IQA analiza algunas distorsiones específicas; en [40] se propuso un estimador de calidad para el tipo de distorsión JPEG (JPEG quality estimator) y se han desarrollado numerosos algoritmos que estiman el ruido blanco [37], [38] [39]. También se han propuesto medidas NR-IQA de propósito general. Estas últimas pueden clasificarse en dos tipos. Las primeras extraen características de la imagen distorsionada y luego realizan una regresión, como por ejemplo DIIVINE [41], BLIINDS-II [42], y BRISQUE [43]. Las segundas no utilizan evaluaciones humanas, como por ejemplo QAC (quality-aware clustering) [44] y NIQE (natural image quality evaluator) [45] que fue desarrollada para estimar las desviaciones de regularidades estadísticas observadas en imágenes naturales sin ningún conocimiento previo del contenido de la imagen o tipos de distorsión.

## 2.4. Ejemplo de aplicación

A continuación presentamos, a modo de ejemplo y para ilustrar la importancia de la evaluación de la calidad de una imagen (IQA), una aplicación al problema de Fusión de Imágenes. Un objetivo de gran interés en procesamiento de imágenes es la estimación de una imagen que no se observa, utilizando información parcial disponible acerca de ella. En el marco de RR-IQA objetiva, un problema en estudio es la construcción de índices de similitud capaces de colaborar en la estimación de una imagen no observada, de la que sólo se conocen dos versiones distorsionadas de ella, una de las cuales es direccional. Para fijar ideas hemos tomado la imagen I01 de la base TID2008, y dos distorsiones de ella: Additive Gaussian noise (I01AGN) y JPEG2000 transmission errors (I01JPG2000).

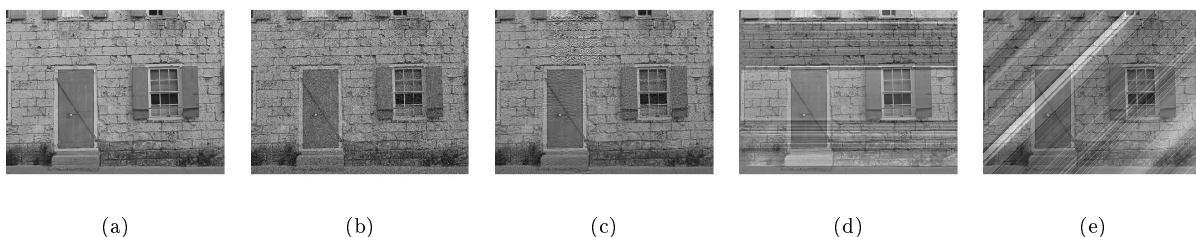


Figura 2.3: (a) Imagen Original I01 (b) I01AGN (c) I01JPG2000 (d) I01h10 (e) I01h11. Las imágenes (b) - (e) son distorsiones de (a) de distinto tipo y nivel.

Luego generamos a partir de I01, dos distorsiones direccionales, una en la dirección horizontal

con  $h_1=(1,0)$  (I01h10) y otra en la dirección diagonal con  $h_2=(1,1)$  (I01h11) mediante el algoritmo 1 presentado en [61] con  $\alpha = 1$ . El objetivo es estimar la imagen I01, a partir de una distorsión de las disponibles en la base TID2008 y alguna de las dos distorsiones direccionales. Asumiremos también que la dirección en que se produce la distorsión direccional es desconocida, por ejemplo, estimar I01 a partir de Additive Gaussian noise y la contaminación direccional diagonal. Esta tarea ha sido recientemente abordada a partir del índice CQ-max [61].

---

# Capítulo 3

## La Base TID2013

La base de datos Tampere Image Database 2013 (TID2013) [7], es una base de datos de imágenes creada para la evaluación de FR-IQA, su antecesora fue la base TID2008 [6]. Se decidió trabajar con la base de imágenes TID2013 ya que la mayoría de los índices y medidas FR-IQA desarrollados en los últimos años han sido probados en la misma. TID2013 contiene 3000 imágenes distorsionadas obtenidas a partir de las 25 imágenes de referencia que se muestran en la Figura 3.1. Contiene las mismas imágenes patrón que se utilizaron para la base TID2008. Lo novedoso de TID2013 respecto a TID2008 es que se agregaron 7 nuevos tipos de distorsiones y un nivel más de distorsión para cada una de ellas. Así en TID2013 se consideraron 24 tipos de distorsiones en total, con 5 niveles de distorsión cada una.

### 3.0.1. El Índice MOS

Cada una de las imágenes distorsionadas cuenta con un valor MOS asociado. El MOS, por sus siglas en inglés *Mean Opinion Score*, es un índice que indica la calidad subjetiva de una imagen; esto es, la calidad de una imagen percibida por los seres humanos. Para obtener estos valores fue necesario realizar 985 experimentos subjetivos, tanto presenciales como online, con voluntarios de cinco países (Finlandia, Francia, Italia, Ucrania y Estados Unidos) [7] [8].

La metodología para obtener el MOS fue la misma que se utilizó para la base TID2008 [6]. A cada voluntario se le asignó solo una de las 25 imágenes de referencia, por lo que cada uno trabajó sólo con 120 imágenes (24 distorsiones por 5 niveles). Las comparaciones fueron realizadas de la siguiente manera: se le mostraron al observador tres imágenes: la original y dos distorsiones; el mismo tenía que decidir cuál de las dos distorsiones era más fiel a la original; la ganadora sumaba 1 punto. Esta competición fue organizada como en el sistema suizo de ajedrez <sup>1</sup>, de manera que cada una de las 120 imágenes participara en 9 comparaciones. Por lo tanto, cada observador, para una sola imagen tuvo que realizar 540 ( $120 \times 9 \div 2$ ) comparaciones de pares, lo cual le requirió un tiempo de aproximadamente 17 minutos. Para obtener el puntaje o score de una imagen se

---

<sup>1</sup>Ver [https://es.wikipedia.org/wiki/Sistema\\_suizo](https://es.wikipedia.org/wiki/Sistema_suizo)



Figura 3.1: Imágenes de referencia de la base TID2013. La última es una imagen creada artificialmente.

sumaron todos los puntos obtenidos durante estas nueve competencias, por lo que este puntaje es un número mayor o igual a 0 y menor o igual a 9. Los resultados anormales (outliers) fueron eliminados. Finalmente se promediaron todos los puntajes asignados a una imagen (más de 30 para cada una) para obtener su MOS, que también será un número entre 0 y 9. Valores más altos se corresponden con una mayor calidad.

Es notable que no se obtuvieron MOS iguales a 0 o a 9 para ninguna imagen; para más detalles ver el histograma, diagrama de caja y diagrama de dispersión de las Figuras 3.2 y 3.3. Más aún, no se obtuvieron valores de MOS superiores a 7.5, esto sugiere que no fue sencillo para los participantes del experimento elegir la imagen de mayor calidad durante cada comparación, sobre todo entre imágenes distorsionadas de alta calidad.

### 3.1. Los distintos tipos de distorsiones

Los creadores de las bases TID2008 y TID2013 consideraron que para cada característica del HVS, la base de datos debía contener, al menos, un tipo de distorsión que permita estimar cómo la misma influye en la calidad visual de la imagen. Además estimaron deseable que la base de datos contenga distorsiones típicas en la práctica, como las que se originan debido a compresión, eliminación de ruido, errores de transmisión de datos, entre otras causas posibles [6].



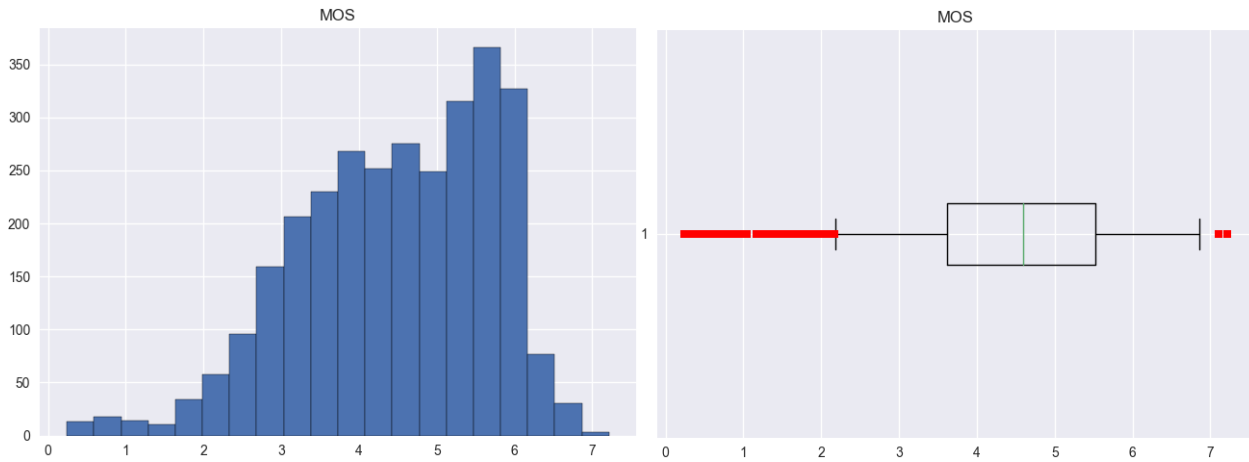


Figura 3.2: Histograma y diagrama de caja del MOS. Se observa que la distribución del MOS es asimétrica

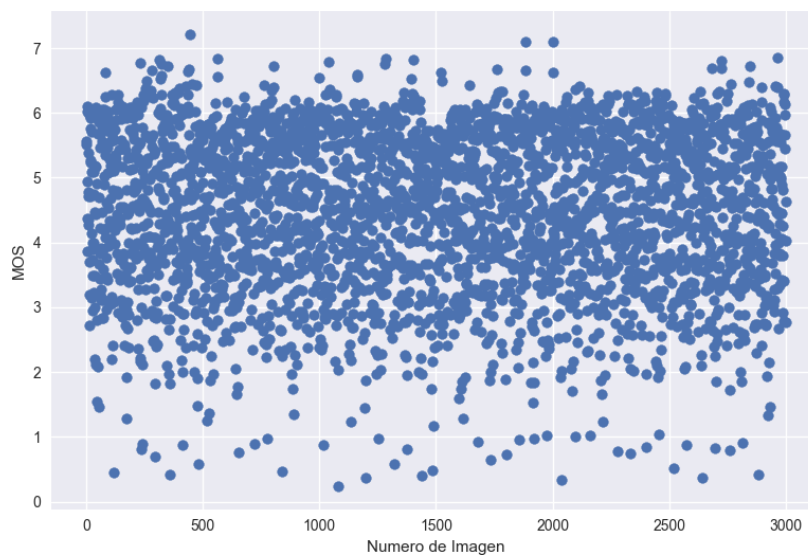


Figura 3.3: Diagrama de dispersión del MOS. Ninguna imagen obtuvo un MOS igual a 0 o mayor a 7.5

A continuación describiremos brevemente en qué consiste cada una de las 24 distorsiones utilizadas en la base TID2013. También se presentan los diagramas de dispersión e histograma del MOS para cada distorsión por separado en las Figuras 3.4 a 3.28.

Para cada distorsión hay en la base de datos 125 imágenes distorsionadas por ella, ya que son 5 niveles de distorsión para cada una de las 25 imágenes de referencia. En el diagrama de dispersión de cada distorsión se incluyen estas 125 imágenes agrupadas en grupos de 5 puntos, los 5 niveles de distorsión correspondiente a la misma imagen. Por ejemplo, en la Figura 3.4, podemos ver, en la figura de la izquierda, el típico diagrama de dispersión con los 125 puntos correspondientes a todas las imágenes afectadas por la distorsión de tipo #1. Abajo de este se observa otra versión de este mismo diagrama agregando segmentos que unen los 5 niveles de distorsión de una misma imagen. Para lograr mayor claridad, agregamos también colores a los niveles de distorsión: el

amarillo corresponde al nivel #1, el verde al nivel #2, el celeste al nivel #3, el azul al nivel #4 y el violeta al nivel #5. Así, por ejemplo, para el cuarto segmento de cinco puntos unidos, el punto amarillo corresponde a la versión distorsionada de la imagen original #4, por el nivel #1 de la distorsión de tipo #1, el punto verde corresponde a la versión distorsionada de la imagen original #4, por el nivel #2 de la distorsión de tipo #1 y así sucesivamente.

Para el resto de las distorsiones, en las Figuras 3.5 a 3.28, se resumió solo incluyendo este último diagrama de dispersión, ya que es el que contiene mayor información. La finalidad de agregar estos segmentos y colores es indicar visualmente si el MOS crece o disminuye a medida que aumenta el nivel de distorsión. Nuestra intuición diría que debe disminuir, pero en muchos casos no ha sido así.

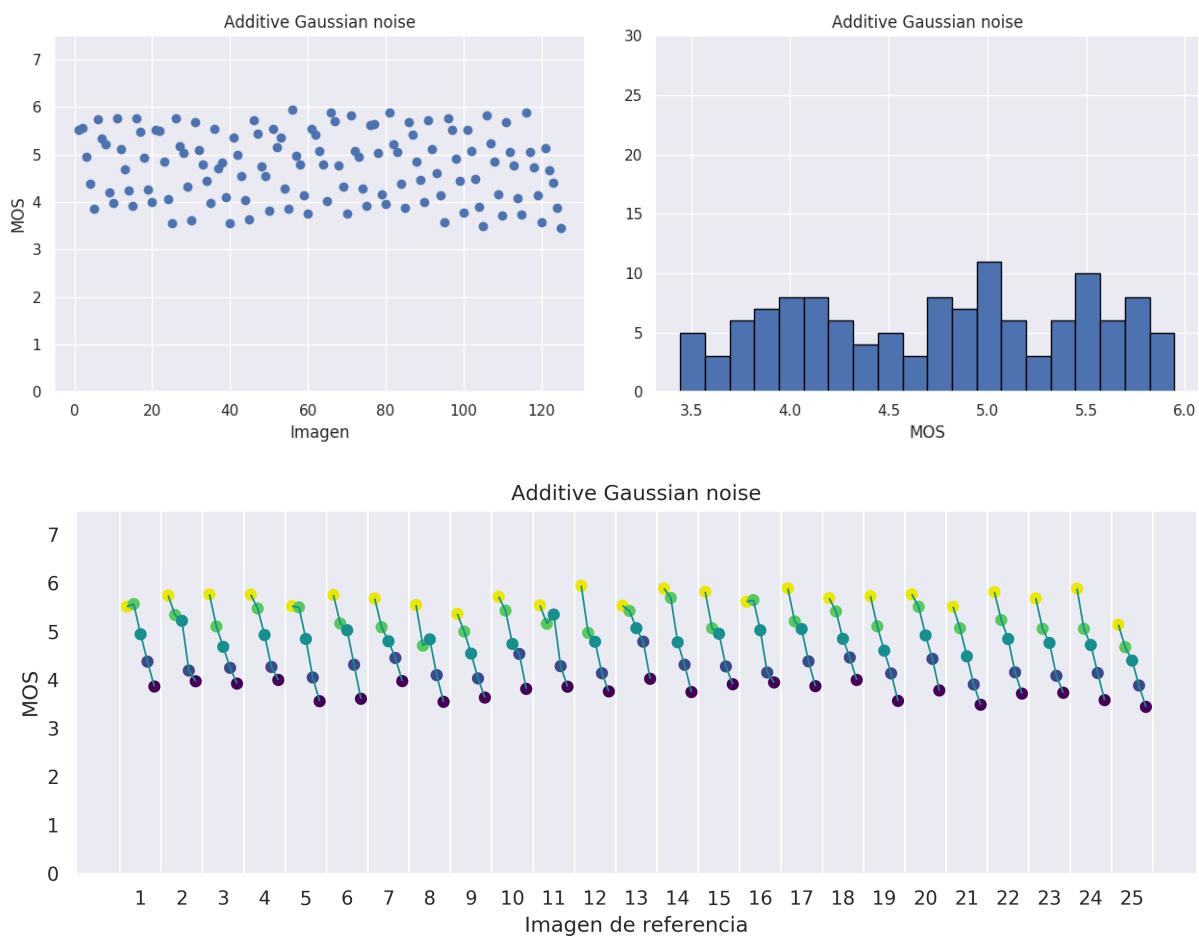


Figura 3.4: Diagramas de dispersión e Histograma del MOS para la distorsión de tipo #1. Se observa que, en general, a medida que aumenta el nivel de distorsión, disminuye el MOS, aunque hay algunas excepciones, por ejemplo, para el onceavo segmento (que une las cinco imágenes distorsionadas de la imagen original #11), el nivel #3 de contaminación presenta un MOS más alto que el nivel #2.

### 3.1.1. Distorsión #1: Ruido Gaussiano Aditivo

En el ruido de tipo gaussiano, todos y cada uno de los píxeles que componen la imagen cambian su valor, de acuerdo con una distribución normal o gaussiana con media cero. Se dice aditivo porque el ruido será sumado a la señal transmitida, por lo tanto la señal recibida será  $y(t) = x(t) + n(t)$ , donde  $x(t)$  es la señal original limpia y  $n(t)$  es el ruido.

El ruido blanco es una señal aleatoria, caracterizada porque sus valores en instantes de tiempo distintos no tienen relación alguna entre sí, es decir, no existe correlación estadística entre sus valores.

El ruido aditivo de media cero se genera a menudo en la práctica cuando se adquieren imágenes y comúnmente se modela como un ruido gaussiano blanco, su percepción visual está vinculada a la capacidad de adaptabilidad y robustez del HVS.

Este tipo de distorsión se incluye en la mayoría de los estudios de efectividad de una métrica de calidad y es, probablemente, uno de los pocos casos en que las métricas MSE y PSNR presentan una buena coincidencia con el HVS [6].

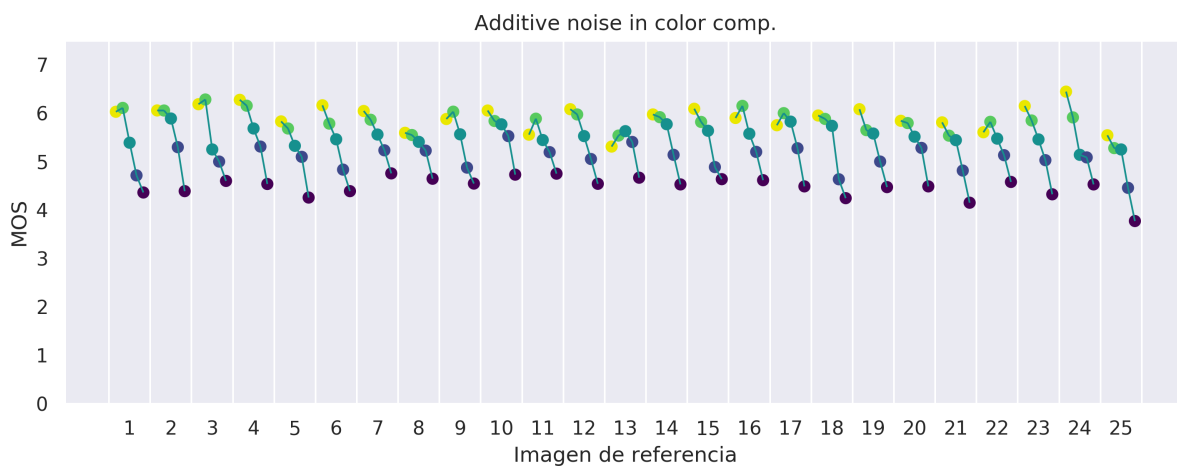


Figura 3.5: Diagrama de dispersión para la distorsión de tipo #2. Se observa que, en general, a medida que aumenta el nivel de distorsión, disminuye el MOS, con algunas excepciones.

### 3.1.2. Distorsión #2: Ruido aditivo en la componente cromática

En las imágenes digitales el mayor nivel de detalle consiste principalmente de las variaciones en el brillo (luminancia) en lugar de las variaciones de tono (crominancia). Dado que cualquier algoritmo de reducción de ruido debe tratar de eliminar el ruido sin sacrificar los detalles de la escena fotografiada, se corre el riesgo de una mayor pérdida de detalle en la reducción de ruido de luminancia que en la reducción de ruido de crominancia. Además, el ruido cromático de las imágenes se encuentra más desagradable que el ruido de luminancia, las manchas de color se consideran como algo más artificial, en comparación con la apariencia granulada de ruido

de luminancia, comparable con el grano de una película. Por estas dos razones, la mayoría de algoritmos de reducción de ruido fotográfico dividen la imagen en las componentes de crominancia y luminancia. La mayoría de programas informáticos dedicados de reducción de ruido permite al usuario controlar la reducción del ruido de crominancia y de luminancia por separado.

Este tipo de distorsión, como su nombre lo indica, consiste en aplicar un ruido aditivo más intenso en la componente de crominancia que en la componente de luminancia, se ha agregado para probar la correspondencia de la métrica de calidad a testear, con la propiedad conocida del HVS para no percibir igualmente las distorsiones en los componentes de brillo (luminancia) y color (crominancia) [6].

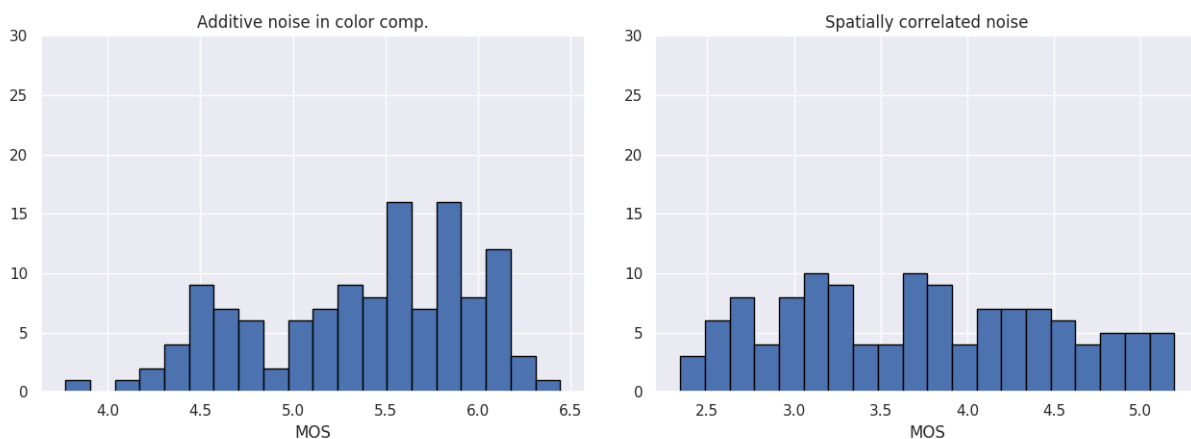


Figura 3.6: Histogramas del MOS para las distorsiones de tipo #2 y #3

### 3.1.3. Distorsión #3: Ruido correlacionado espacialmente

Muy a menudo, el ruido aditivo no se puede considerar como espacialmente no correlacionado (ruido blanco) [6]. Este ruido es una distorsión típica en las fotografías digitales actuales.

Una imagen tomada con una cámara digital es un mosaico, corrompida por un ruido con propiedades estadísticas bastante complicadas, aunque en su versión original este ruido es prácticamente independiente del espacio. Sin embargo, en el proceso de conversión de imágenes desde el mosaico original a datos en forma de raster se utiliza interpolación no lineal y entonces el ruido se correlaciona espacialmente. La eliminación de ruido y la compresión de imágenes en estos casos son tareas bastante complejas.

La consideración del ruido correlacionado espacialmente permite, además, verificar la correspondencia de las métricas con la sensibilidad de frecuencia espacial inherente al HVS [6].

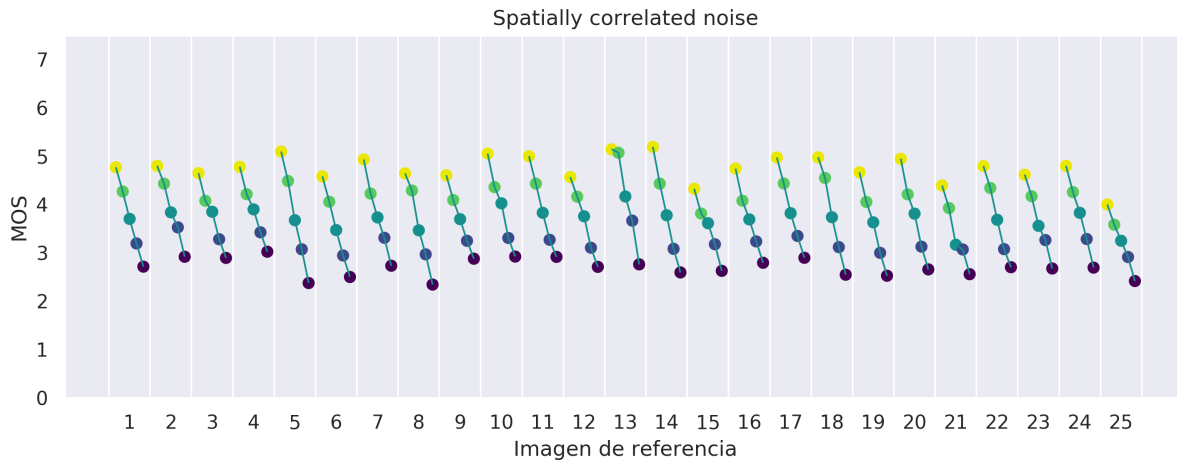


Figura 3.7: Diagrama de dispersión para la distorsión de tipo #3. Se observa que, para las 25 imágenes, a medida que aumenta el nivel de distorsión, disminuye considerablemente el MOS.

### 3.1.4. Distorsiones #4 y #5: Ruido enmascarado y Ruido de alta frecuencia

El ruido de paso bajo correlacionado espacialmente no es el único caso de ruido aditivo que no es blanco. El ruido enmascarado y los ruidos de alta frecuencia son otros tipos de distorsiones que también permiten analizar la adecuación de las métricas con respecto a la sensibilidad de contraste local y la sensibilidad de frecuencia espacial características del HVS. Dichas distorsiones son típicas para una amplia clase de tareas prácticas, como la compresión de imágenes con pérdidas y, especialmente, las marcas de agua digitales. Tal como lo demostraron algunos estudios, muchos índices de calidad conocidos, desafortunadamente, no tienen suficientemente en cuenta estas peculiaridades del HVS [6].



Figura 3.8: Diagrama de dispersión para la distorsión de tipo #4.

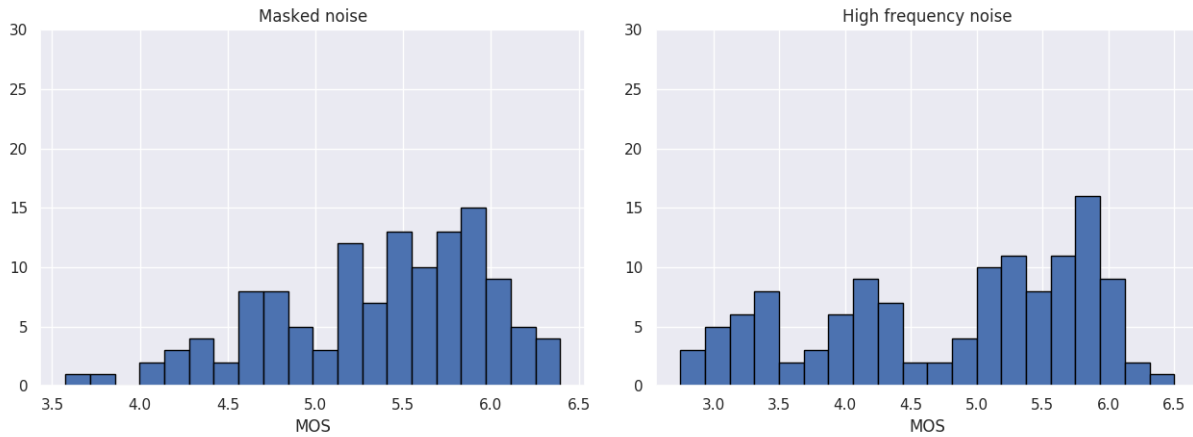


Figura 3.9: Histogramas del MOS para las distorsiones #4 y #5.

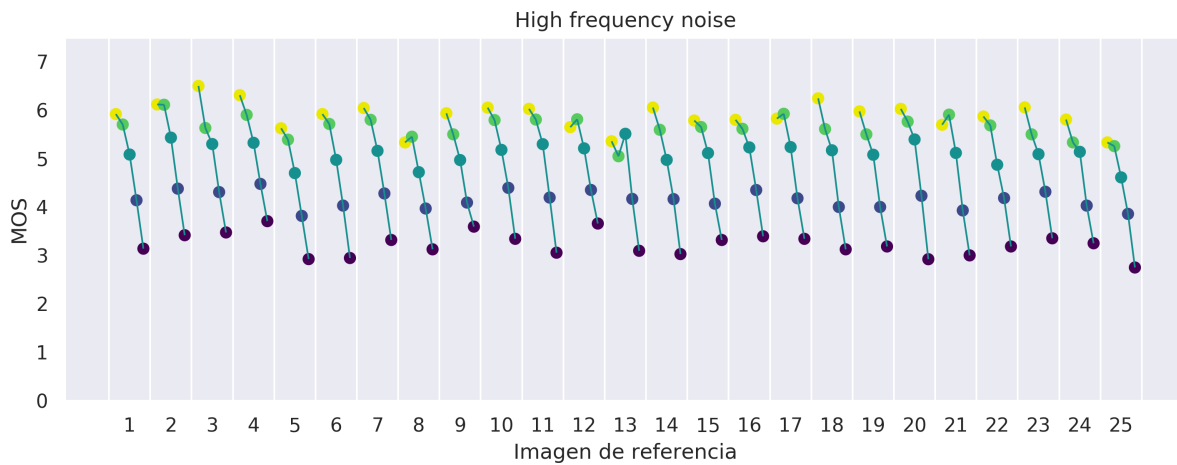


Figura 3.10: Diagramas de dispersión para la distorsión de tipo #5.

### 3.1.5. Distorsión #6: Ruido sal y pimienta o Ruido de impulso

El ruido sal y pimienta se caracteriza principalmente por cubrir de forma dispersa toda la imagen con una serie de píxeles de colores. Este tipo de ruido suele producirse cuando la señal de la imagen es afectada por intensas y repentinas perturbaciones o impulsos, y puede darse particularmente debido a errores de codificación / decodificación en la transmisión de datos.

Esta distorsión fue generada para ser aplicada a las imágenes de las bases TID2008 y TID2013, utilizando un modelo típico de ruido de impulso distribuido uniformemente, como se indica en [6]. En la práctica su estudio es importante para evaluar la efectividad de los métodos para la eliminación del ruido impulsivo y la restauración de imágenes (Inpainting).

Se ha demostrado que, para pequeñas proporciones de ruido impulsivo, los humanos pueden recuperar de forma intuitiva los valores de los píxeles dañados utilizando píxeles vecinos, por lo tanto el estudio de imágenes afectadas por este tipo de distorsión también puede ayudar a evaluar cómo las métricas objetivas modelan esta capacidad del HVS.

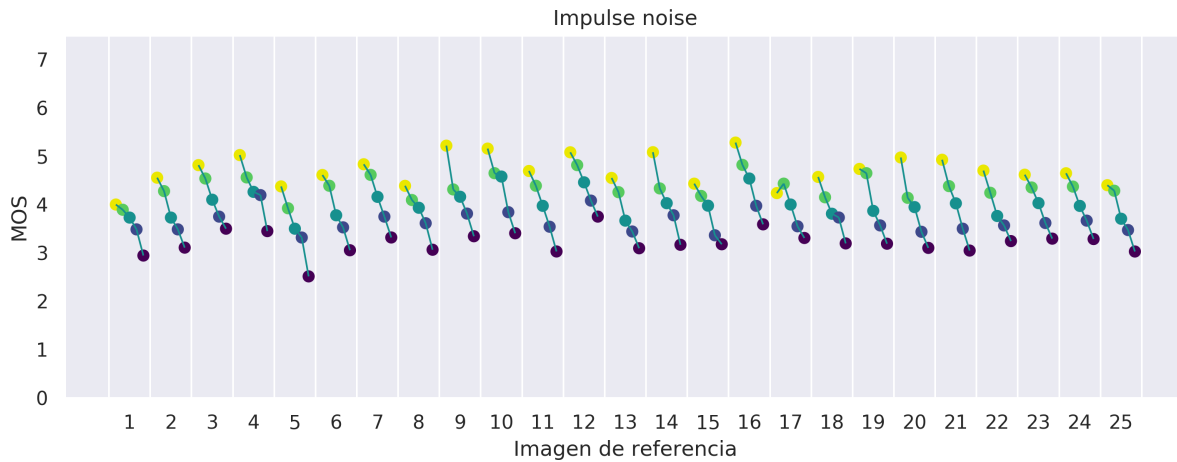


Figura 3.11: Diagrama de dispersión para la distorsión de tipo #6. Es curioso que para el segmento número 17 (imagen original #17) al pasar del nivel #1 al nivel #2 de distorsión, el MOS aumenta, mientras que para el resto de las imágenes al aumentar el nivel de distorsión, el MOS disminuye.

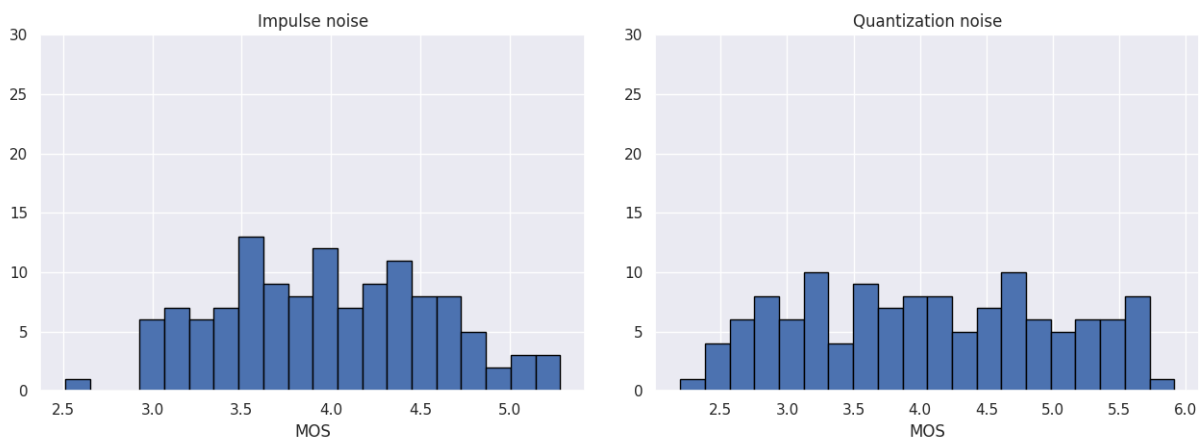


Figura 3.12: Histogramas del MOS para las distorsiones #6 y #7.

### 3.1.6. Distorsión #7: Ruido de cuantificación

Es un error introducido a partir de la conversión analógica a digital. Cuando se convierte de una señal analógica a una señal digital, el error es inevitable. Una señal analógica es continua, con precisión idealmente infinita, mientras que la precisión de la señal digital depende de la resolución de cuantificación o del número de bits del convertidor analógico a digital. La diferencia entre el valor analógico real y el valor digital aproximado debido al redondeo que se produce durante la conversión se denomina error de cuantificación.

En situaciones ideales, el ruido de cuantificación durante la adquisición se distribuye uniformemente.

El ruido de cuantificación no ha recibido demasiada atención en el estudio IQA, aunque esta distorsión se cumple a menudo en la práctica y permite estimar la adecuación de las métricas

objetivas con respecto a varias peculiaridades del HVS como la sensibilidad al color, al contraste local y la frecuencia espacial. La mayoría de las métricas conocidas no evalúan adecuadamente la calidad visual de las imágenes sujetas a este tipo de distorsión [6].

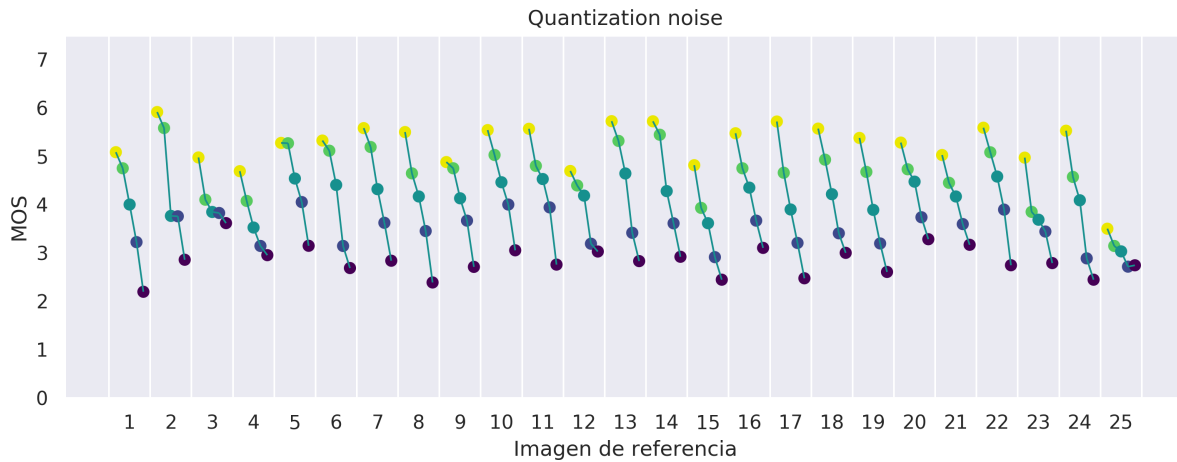


Figura 3.13: Diagrama de dispersión para la distorsión de tipo #7.

### 3.1.7. Distorsión #8: Desenfoque Gaussiano

En el procesamiento de imágenes, un desenfoque gaussiano, también conocido como suavizado gaussiano, es el resultado de difuminar una imagen por una función gaussiana. Es un efecto ampliamente utilizado en software gráfico, típicamente para reducir el ruido de la imagen, así como los detalles.

Matemáticamente, aplicar un desenfoque gaussiano a una imagen es lo mismo que convolucionar la imagen con una función gaussiana. Este proceso también se conoce como transformación Weierstrass bidimensional. Dado que la transformada de Fourier de un gaussiano es otro gaussiano, la aplicación de un desenfoque gaussiano tiene el efecto de reducir los componentes de alta frecuencia de la imagen; un desenfoque gaussiano es, por lo tanto, un filtro de paso bajo.

En la práctica, el efecto mezcla ligeramente los colores de los píxeles vecinos, lo que provoca que la imagen pierda algunos pequeños detalles y, de esta forma, se vea más suave, aunque menos nítida o clara, respecto a los bordes presentes en la imagen, los que se ven afectados.

El efecto visual de esta técnica de borrosidad es un destello suave que se asemeja al de ver la imagen a través de una pantalla translúcida o la sombra de un objeto poco iluminado; se puede comparar al efecto que se produce al tomar una fotografía con una cámara fotográfica desenfocada.

La suavización gaussiana también se utiliza como una etapa de preprocesamiento en algoritmos de visión computarizada (computer vision) con el fin de mejorar las estructuras de imagen a diferentes escalas.



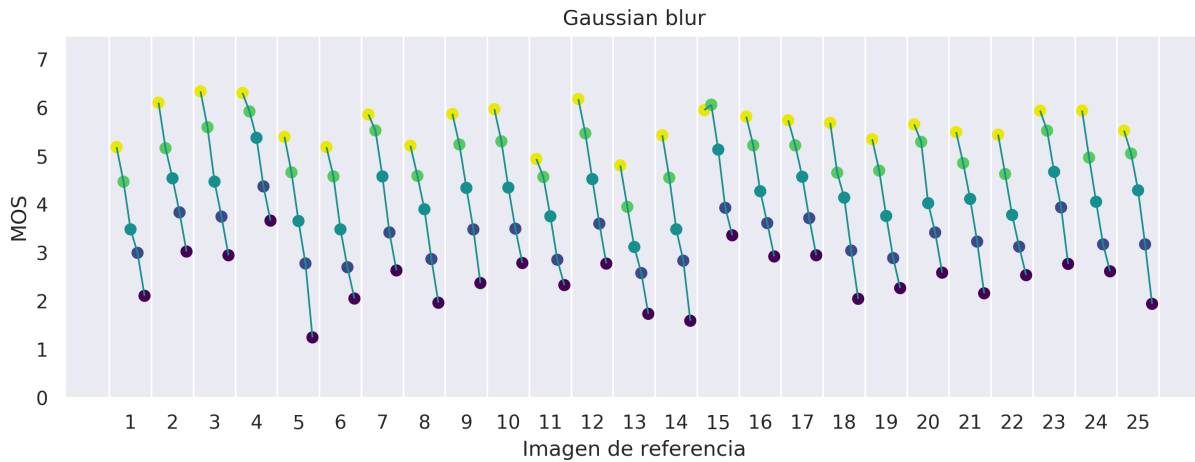


Figura 3.14: Diagrama de dispersión para la distorsión de tipo #8.

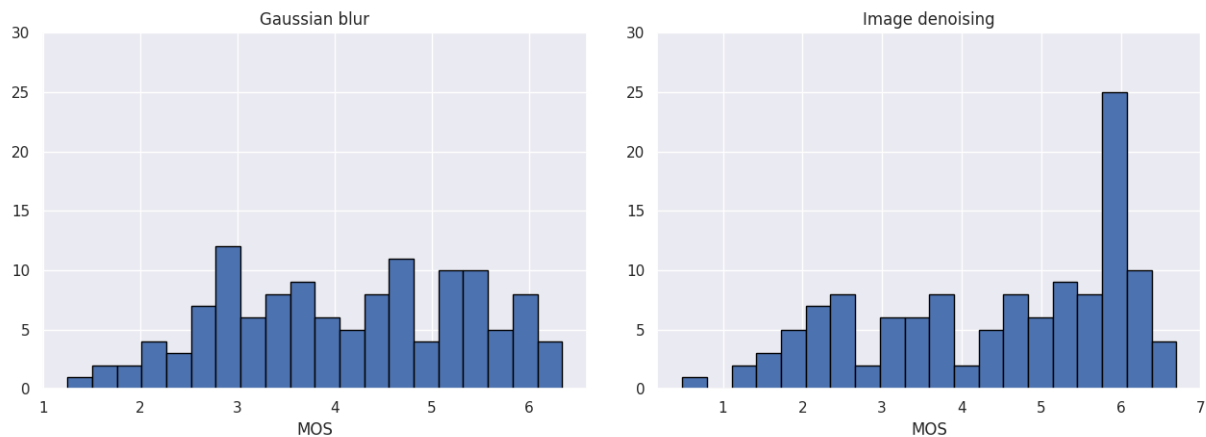


Figura 3.15: Histogramas del MOS para las distorsiones #8 y #9.

### 3.1.8. Distorsión #9: Eliminación de ruido

Otro tipo importante de distorsiones estudiadas recientemente son las distorsiones residuales resultantes después de aplicar diferentes procedimientos de eliminación de ruido.

Como su nombre lo indica, esta distorsión se presenta durante el proceso de remover ruido de una señal, en este caso una imagen, a través de un sistema de filtrado.

El filtrado de imágenes constituye una tarea práctica habitual, para la cual es necesario contar con herramientas adecuadas para evaluar la calidad visual de la imagen filtrada.

A menudo sucede que una imagen filtrada presenta un valor PSNR más alto (2-3 dB más alto) que la original ruidosa correspondiente; al mismo tiempo, esta imagen procesada se ve perceptualmente peor que la ruidosa original. Por lo tanto, en las bases TID2008 y TID2013 se han incluido imágenes generadas al suprimir ruido gaussiano (previamente aplicado) por medio de uno de los filtros de última generación basado en la Transformada de Coseno Discreta 3D (DCT) como se indica en [6], según ciertos estudios este enfoque de filtrado ha producido la mejor calidad visual

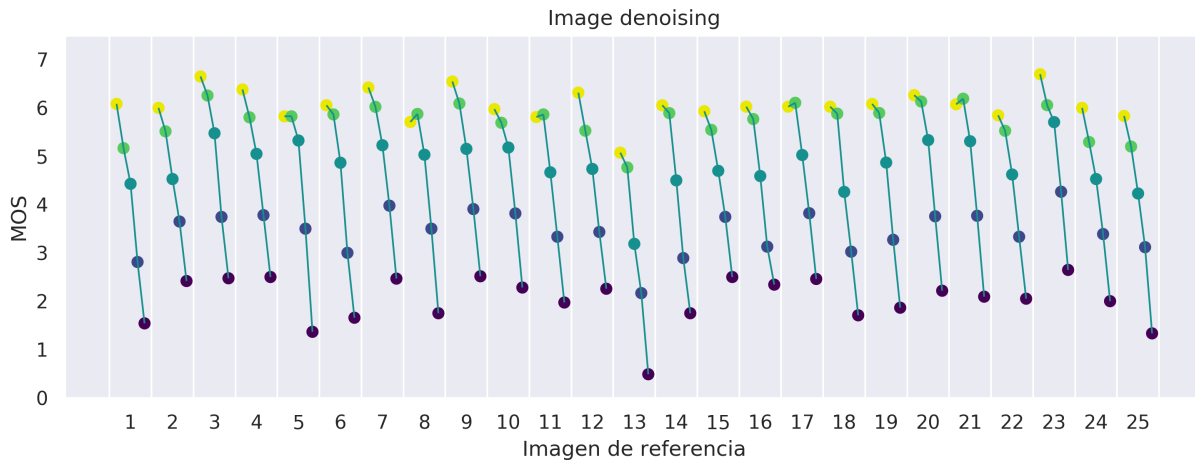


Figura 3.16: Diagrama de dispersión para la distorsión de tipo #9.

posible de las imágenes filtradas entre varios filtros considerados.

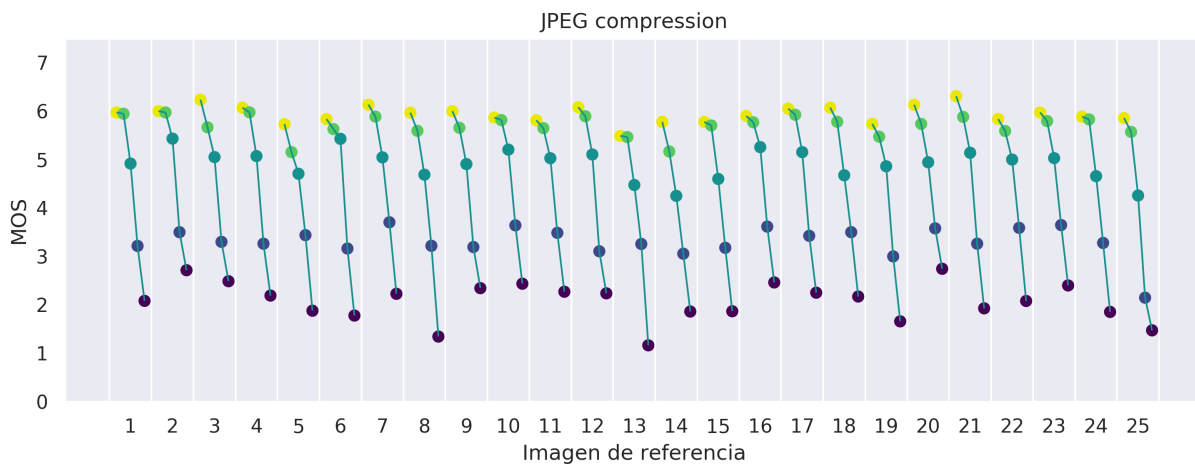


Figura 3.17: Diagrama de dispersión para la distorsión de tipo #10.

### 3.1.9. Distorsiones #10, #11, #12 y #13: Compresión JPEG, Compresión JPEG2000, Transmisión de errores JPEG y Transmisión de errores JPEG2000

JPEG es un estándar de compresión y codificación digital de imágenes creado por el Grupo Conjunto de Expertos en Fotografía o JPEG (Joint Photographic Experts Group) en 1992. JPEG 2000 es otro estándar similar, creado en el año 2000 con la intención de sustituir el formato original de 1992. Ambos son métodos de compresión con pérdidas (lossy compression) muy usados en la práctica, particularmente para aquellas imágenes producidas por la fotografía digital.

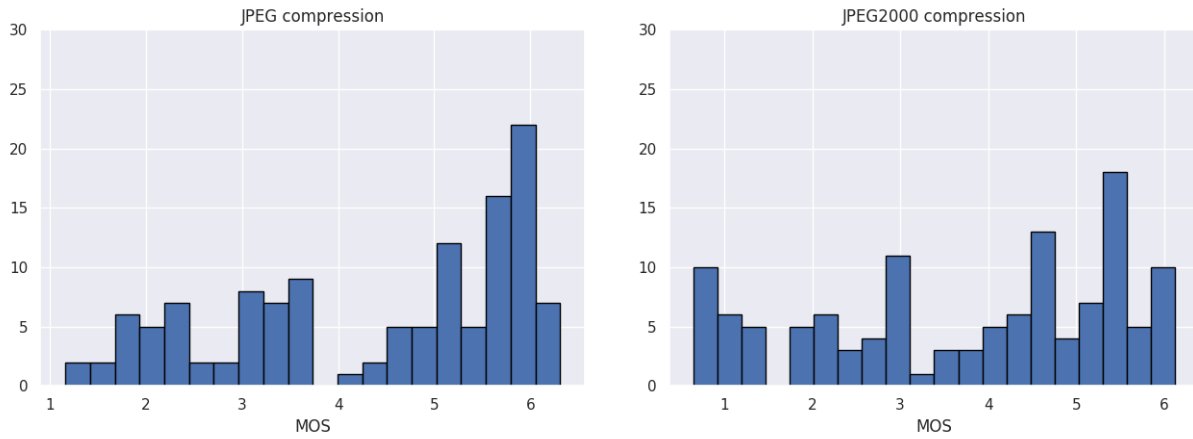


Figura 3.18: Histogramas del MOS para las distorsiones #10 y #11.

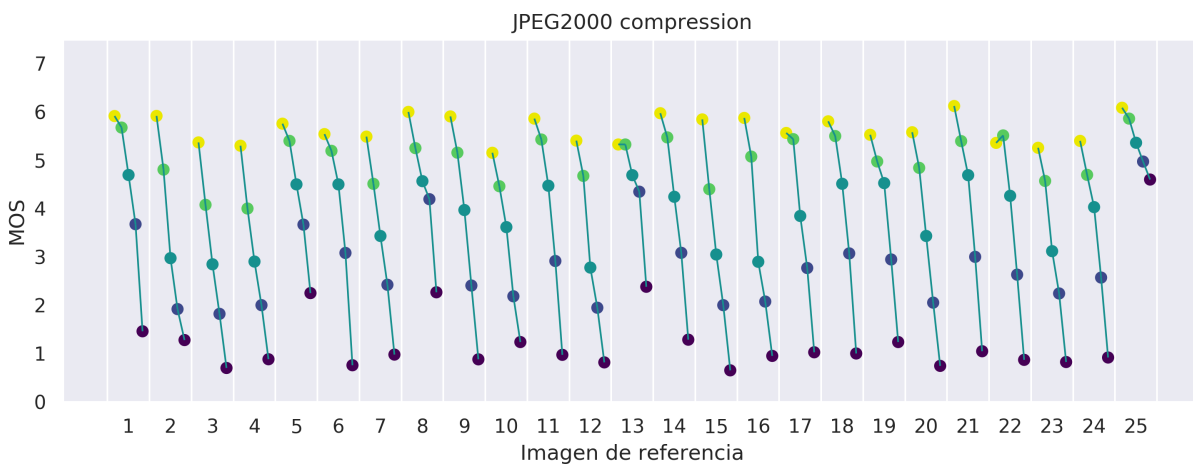


Figura 3.19: Diagrama de dispersión para la distorsión de tipo #11.

Se han incluido en las base de datos las imágenes comprimidas en JPEG y JPEG2000, así como las decodificadas con errores en los canales de transmisión de datos (JPEG transmission errors y JPEG2000 transmission errors respectivamente). Los errores de descodificación se han modelado de tal manera que se ha proporcionado el PSNR requerido para las imágenes decodificadas [6].

Muy a menudo las distorsiones inducidas por tales errores son casi invisibles gracias a su no excentricidad<sup>2</sup>, pues los fragmentos distorsionados pueden ser similares en textura y / o en color al de los fragmentos circundantes y, debido a las peculiaridades del HVS, un ser humano podría no notar tales distorsiones.

<sup>2</sup>Aquí y en lo que sigue el término “excentricidad” se refiere a comportamientos o patrones raros o poco comunes que llaman la atención.

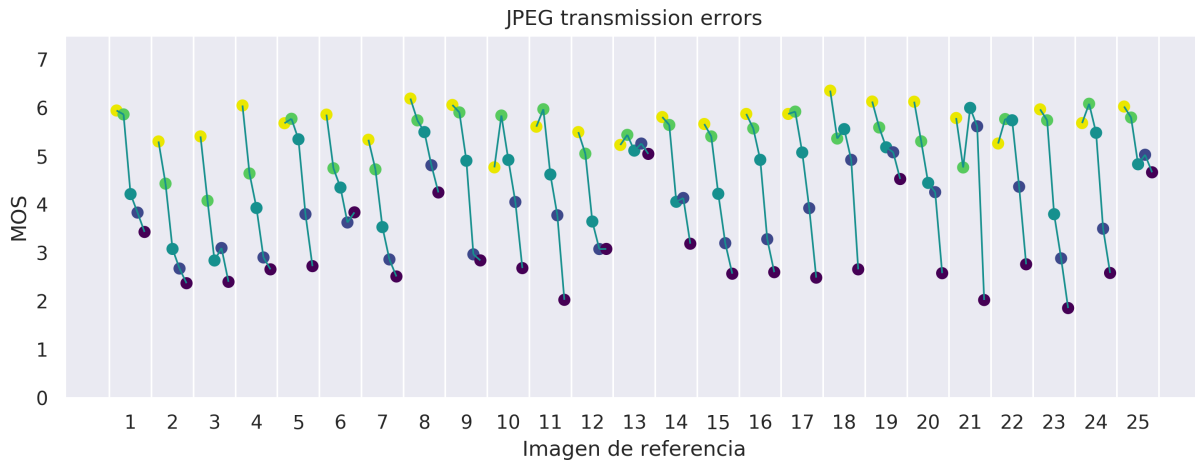


Figura 3.20: Diagrama de dispersión para la distorsión de tipo #12.

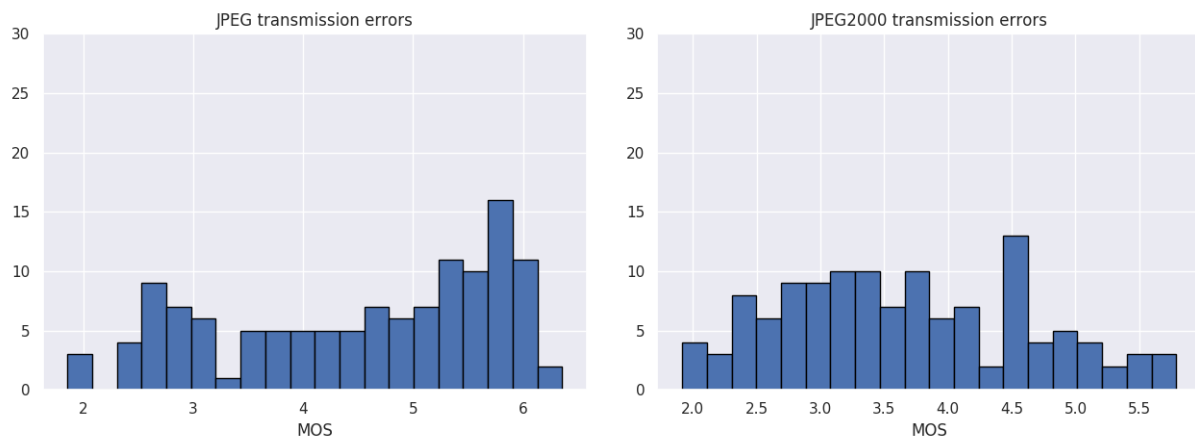


Figura 3.21: Histogramas del MOS para las distorsiones #12 y #13.

### 3.1.10. Distorsión #14: Patrón de ruido sin excentricidad

El HVS por lo general no es sensible a las distorsiones sin excentricidad; considerar este comportamiento en aplicaciones, como la compresión con pérdida de imagen, puede conducir a una mejora de calidad considerable de las imágenes comprimidas con la misma relación de compresión<sup>3</sup>; por esta razón los autores han decidido incluir en las bases de datos imágenes distorsionadas por un tipo específico de artefactos que modelaron y denominaron “patrones de ruido sin excentricidad”[6].

Tales distorsiones se han modelado de la siguiente manera: se ha tomado aleatoriamente un pequeño fragmento de imagen de  $15 \times 15$  píxeles de tamaño en una imagen de referencia y se lo ha copiado en lugar de otro fragmento ubicado cerca (a una distancia de pocos píxeles). Esta operación

<sup>3</sup>Relación que indica, la cantidad de datos finales que quedan después de ser comprimida la información original. Por ejemplo, en una imagen que pesa 100 kB y tras su compresión queda en 50 kB podemos decir que la relación de compresión es del 50%. Esto nos dice que el archivo original ha sido reducido a la mitad.



Figura 3.22: Diagrama de dispersión para la distorsión de tipo #13.

se ha repetido varias veces hasta que se proporciona aproximadamente un PSNR requerido.

Si una imagen está dañada, por ejemplo, por un ruido impulsivo, un ser humano puede detectar fácilmente los píxeles afectados. Pero en el caso de este tipo de distorsión, con frecuencia es difícil identificar distorsiones compactas de un tamaño bastante grande de  $15 \times 15$  píxeles. Esto demuestra la propiedad del HVS de descartar distorsiones sin excentricidad.

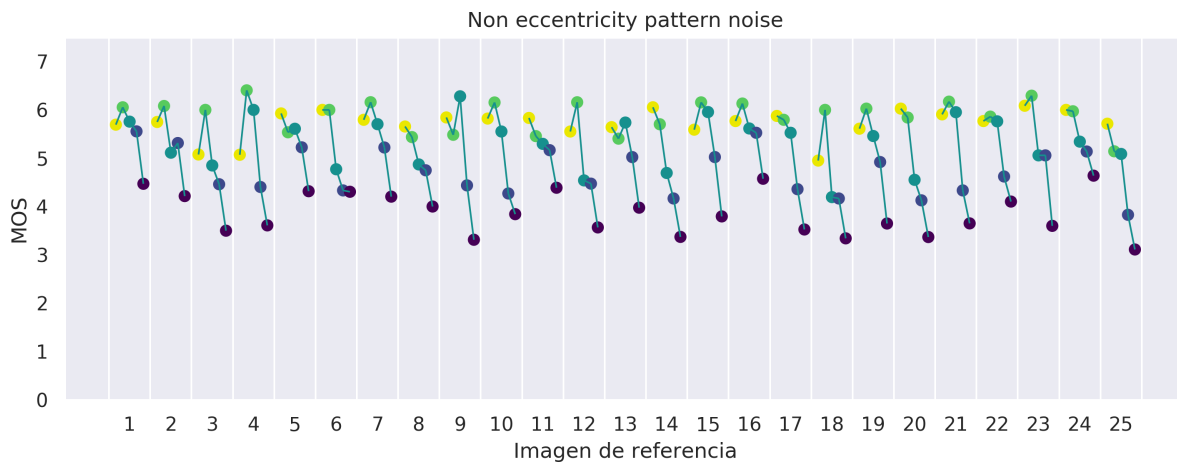


Figura 3.23: Diagrama de dispersión para la distorsión de tipo #14.

### 3.1.11. Distorsión #15: Distorsiones locales de bloques de diferentes intensidades

Otras distorsiones modeladas por los autores y agregadas a TID2008 y TID2013 son las llamadas distorsiones locales de bloques de diferentes intensidades [6].

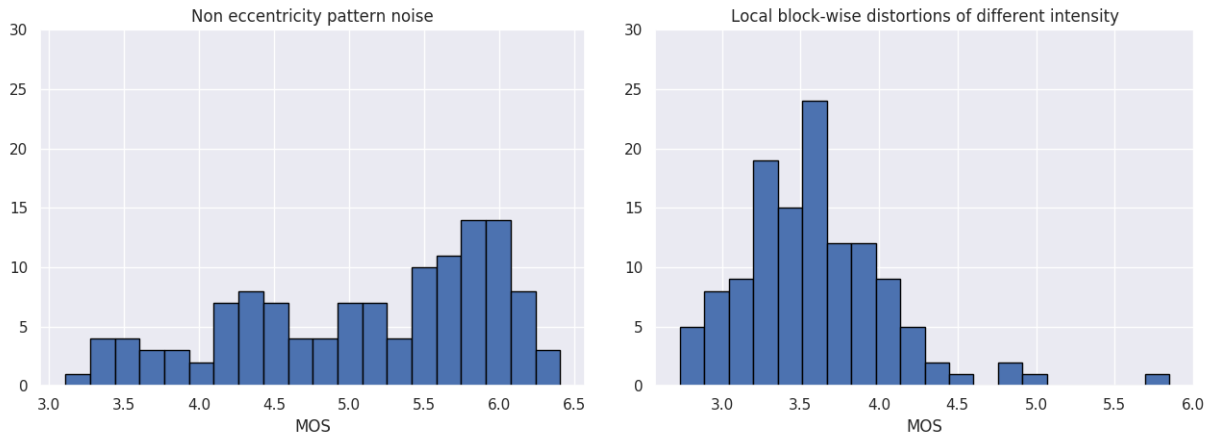


Figura 3.24: Histogramas del MOS para las distorsiones #14 y #15.

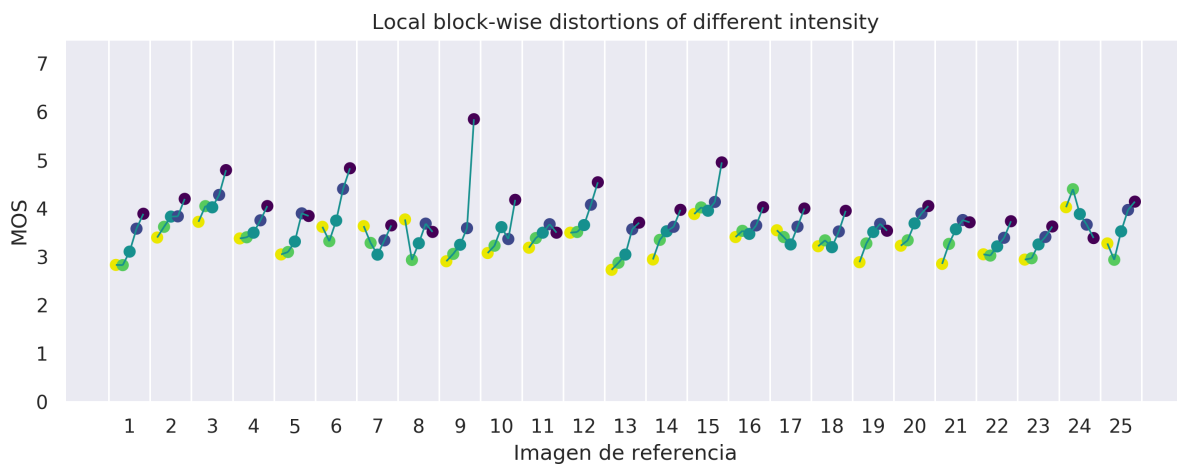


Figura 3.25: Diagrama de dispersión para la distorsión de tipo #15. Al aumentar el nivel de distorsión, la calidad mejora en casi todos los casos.

Se parte del supuesto que en el caso de distorsiones compactas similares a impulsos, el HVS no reacciona a la distorsión en un solo píxel, sino que se enfoca en el área (porcentaje de píxeles) que está afectada por la distorsión.

Estas distorsiones se han modelado de tal manera que bloques de tamaño  $32 \times 32$  píxeles que tienen un color aleatorio arbitrario se colocan aleatoriamente en áreas importantes de una imagen.

Para el primer nivel de distorsión, se han agregado 16 bloques que tienen un color ligeramente diferente del color medio del fragmento reemplazado. Para el segundo nivel de distorsión, la cantidad de dichos bloques fue 8 pero su color difiere más del color medio del fragmento reemplazado. Para los niveles tercero y cuarto, cuatro y dos bloques han sido reemplazados, respectivamente. Sin embargo, para estos bloques, su color difiere incluso más fuertemente de los colores medios de los correspondientes fragmentos reemplazados.

Las diferencias de color e intensidad se han ajustado de tal manera que, independientemente

del número de bloques, se ha proporcionado aproximadamente el mismo PSNR.

Experimentos muestran que una imagen corrompida por dos bloques es percibida como de mejor calidad visual que una imagen distorsionada por 16 bloques, aunque detenta un PSNR más pequeño [6]. Lo más probable es que esto se explique por la incapacidad de HVS para “recuperar” información perdida en lugares distorsionados por bloques independientemente de su color. Entonces parece que, desde el punto de vista de la percepción de la imagen, es mejor si el área total de tales distorsiones por bloques es menor, mientras que el grado de tales distorsiones no tiene tanta importancia.



Figura 3.26: Diagrama de dispersión para la distorsión de tipo #16.

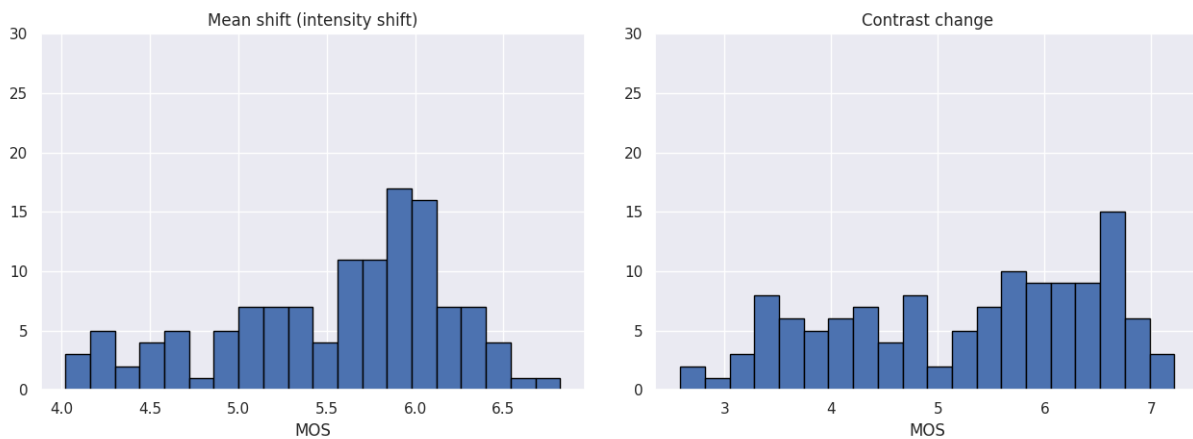


Figura 3.27: Histogramas del MOS para las distorsiones #16 y #17.

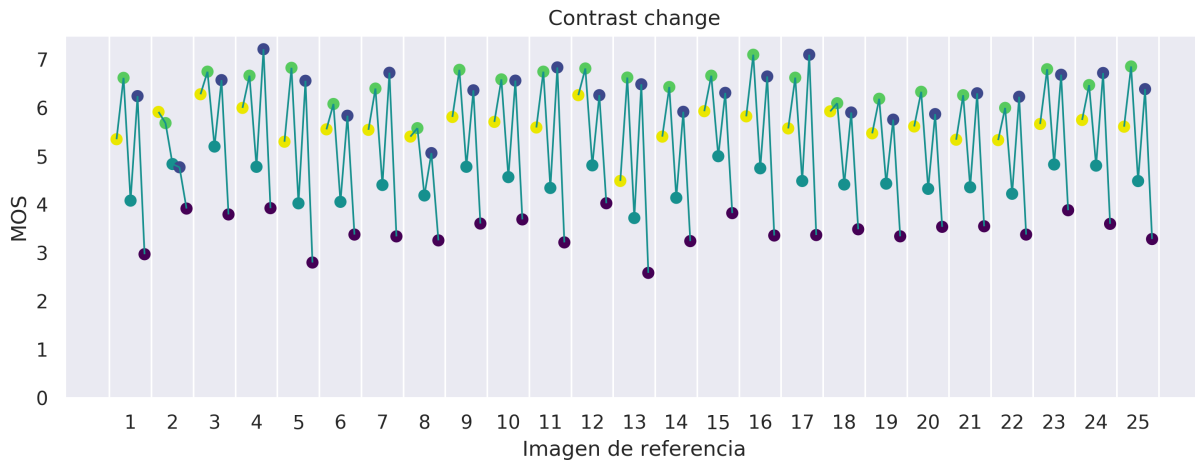


Figura 3.28: Diagrama de dispersión para la distorsión de tipo #17. Una observación interesante es que, en varios casos, al aumentar el nivel de distorsión la calidad de la imagen aumenta y disminuye alternadamente.

### 3.1.12. Distorsiones #16 y #17: Cambio de valor medio y Cambio de contraste

En la base TID2008 para cada distorsión de media y contraste, se han simulado dos cambios a valores más pequeños y dos cambios a valores más grandes. El cambio de valor medio y el cambio de contraste se han realizado con respecto a las imágenes como un todo [6].

Hasta ahora hemos descrito las distorsiones que se presentaron en la base TID2008 [6], que fueron posteriormente también utilizadas para la base TID2013. A continuación se describen las distorsiones nuevas que se agregaron en la base TID2013 [8].

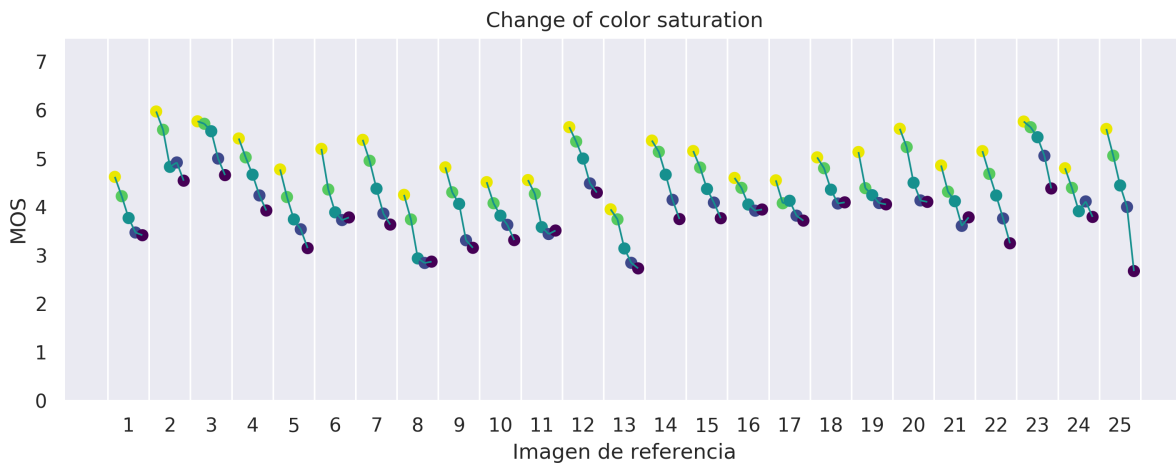


Figura 3.29: Diagrama de dispersión para la distorsión de tipo #18.



### 3.1.13. Distorsión #18: Cambio en la saturación del color

Un cambio en la saturación del color puede venir como resultado de diferentes factores en las etapas de adquisición y procesamiento de imágenes. En particular, puede surgir debido a una gran cuantificación de las componentes de color durante la compresión de imágenes basada en JPEG. Tal distorsión también podría presentarse durante la impresión de imágenes en color. Todas las simulaciones para obtener las imágenes con esta distorsión en la base TID2013 fueron realizadas por los autores de la base TID2013 en Matlab [8].

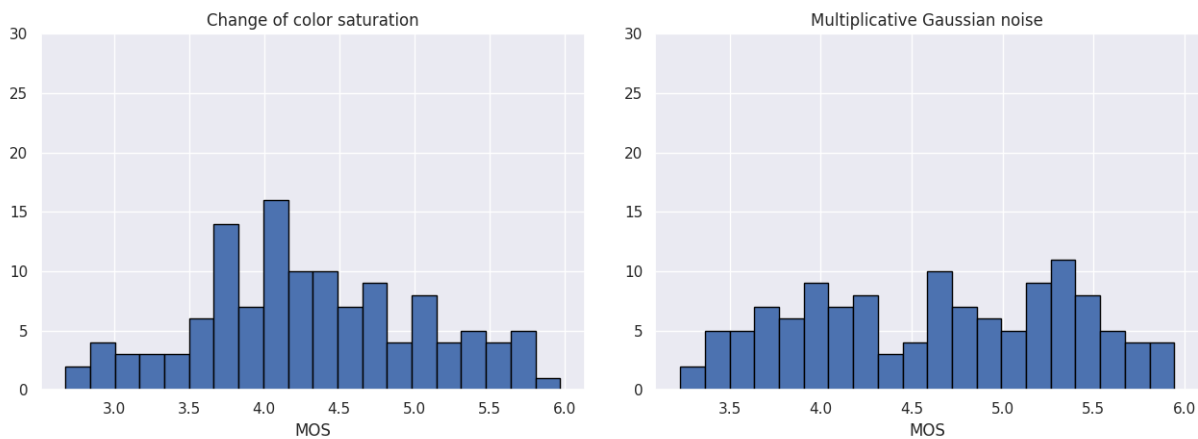


Figura 3.30: Histogramas del MOS para las distorsiones #18 y #19.

### 3.1.14. Distorsión #19: Ruido Gaussiano Multiplicativo

Se trata del ruido de aspecto granulado muchas veces presentes en imágenes de radar o ecografías. Para generarlo se multiplica la señal recibida obteniendo  $y(t) = x(t) * n(t)$ , donde  $x(t)$  es la señal original limpia y  $n(t)$  es el ruido, usualmente modelado con una distribución normal o gaussiana con media cero.

Este tipo de distorsión no fue utilizada en ninguna otra base de imágenes. En la base TID2013 el ruido fue simulado separadamente para cada componente de color RGB, usando la misma varianza  $\sigma^2$  en todas las componentes [8].

### 3.1.15. Distorsión #20: Ruido de confort

Es un tipo muy particular de distorsión. Se sabe que los seres humanos no prestan mucha atención al ruido presente en una imagen; similarmente, los humanos a veces no distinguen realizaciones de textura si estos fragmentos de textura tienen los mismos parámetros. Estas propiedades se aprovechan en la compresión de imágenes para obtener una relación de compresión más alta a la vez que la información descomprimida se ve más natural. Al comparar una imagen distorsionada producida con esta metodología, con su original, se obtiene en general una diferencia notable

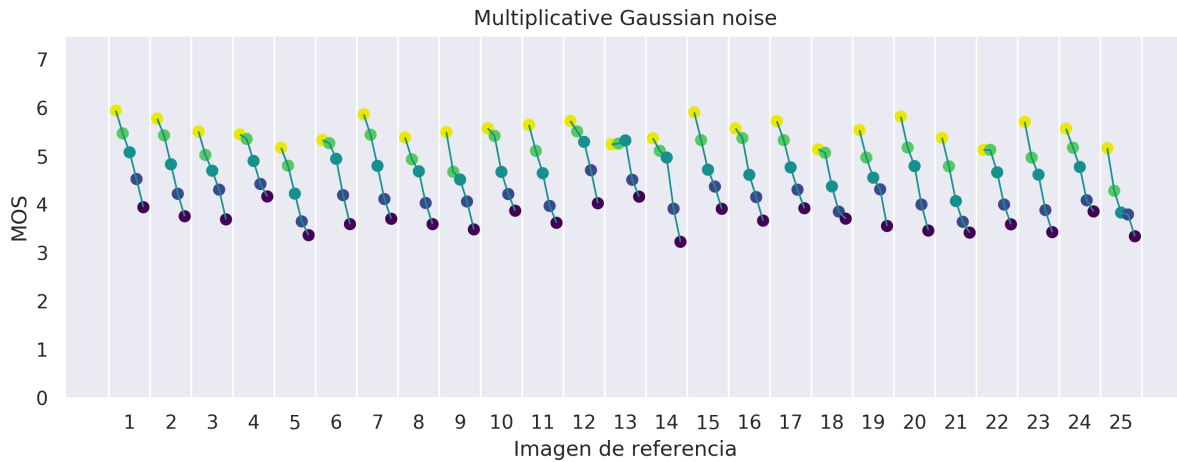


Figura 3.31: Diagrama de dispersión para la distorsión de tipo #19.

en términos de métricas estándares tales como MSE o PSNR, mientras que visualmente estas imágenes suelen verse muy similares [8].

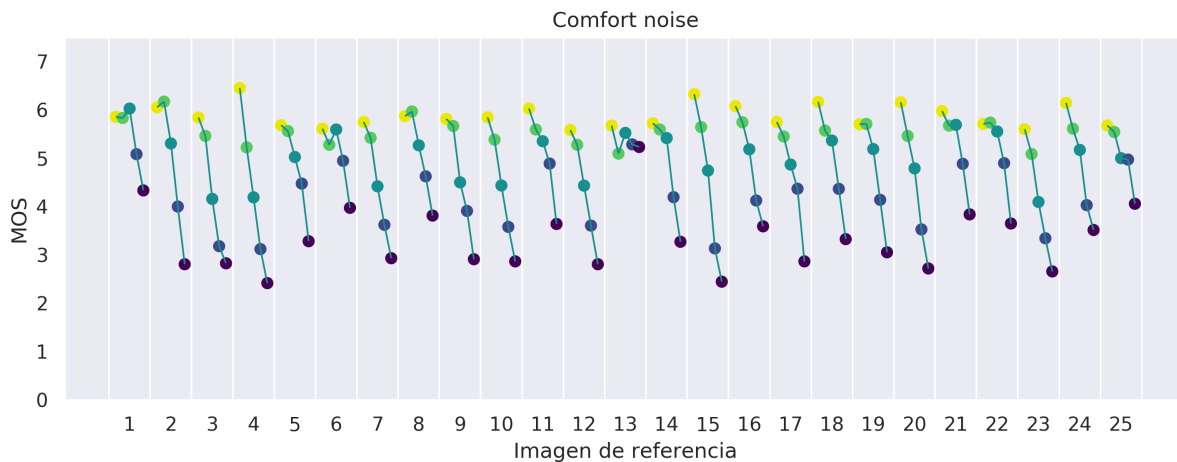


Figura 3.32: Diagrama de dispersión para la distorsión de tipo #20.

### 3.1.16. Distorsión #21: Compresión de imágenes ruidosas

Este tipo de distorsión básicamente combina otros dos tipos de distorsiones: ruido y compresión. Se presenta cuando se comprimen imágenes en condiciones no ideales, por lo que son de mucha importancia en la práctica. Además muchas investigaciones [8] señalan que por lo general hay muchos tipos de distorsiones simultáneamente presentes en imágenes mientras que las bases de datos de imágenes comúnmente contienen imágenes con distorsiones puramente de un solo tipo.

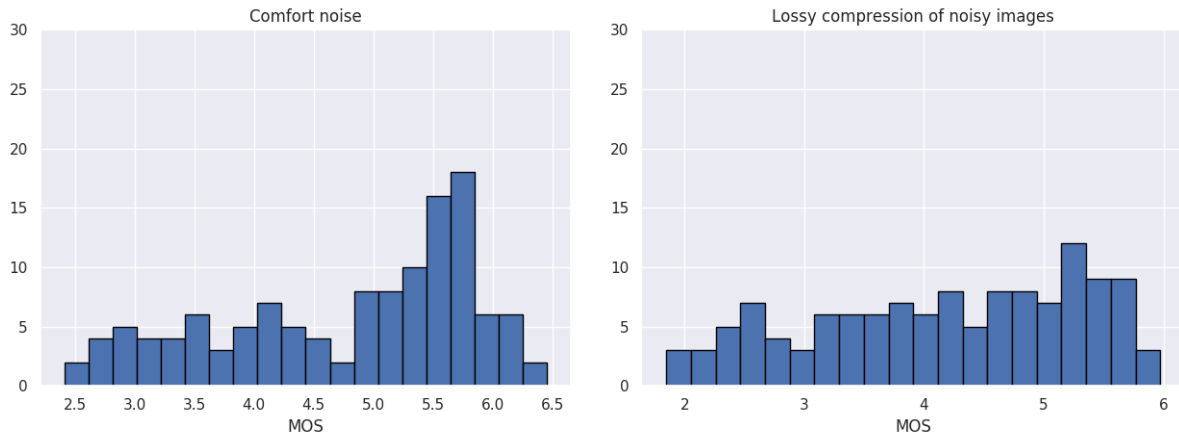


Figura 3.33: Histogramas del MOS para las distorsiones #20 y #21.

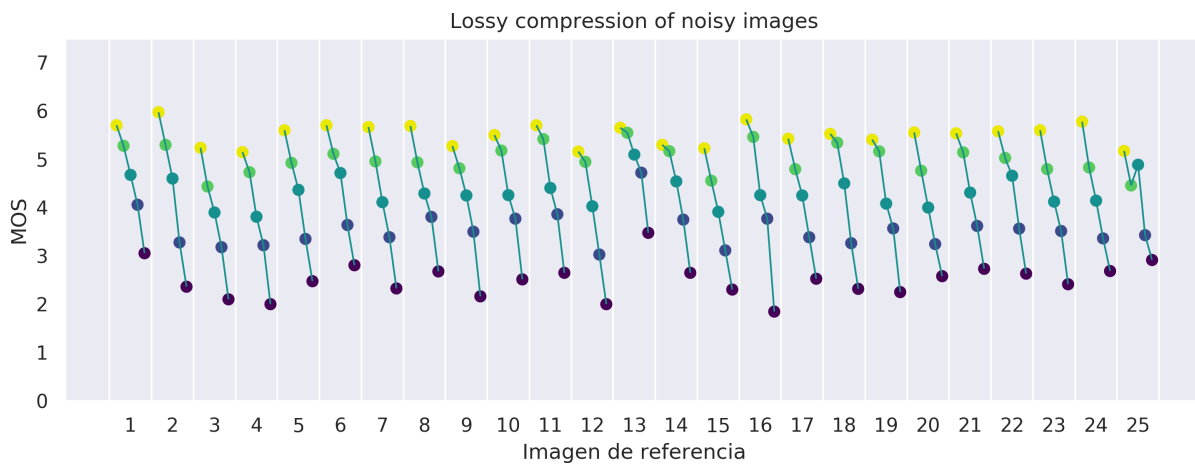


Figura 3.34: Diagrama de dispersión para la distorsión de tipo #21. Al aumentar el nivel de distorsión el MOS disminuye drásticamente en la mayoría de los casos.

### 3.1.17. Distorsión #22: Cuantificación del color con tramado

El tramado (dither en inglés) es una técnica usada para crear la ilusión de profundidad de color en imágenes con una paleta de colores limitada. Muchas veces, esta forma de ruido es aplicada intencionalmente para aleatorizar el error de cuantificación, evitando patrones a gran escala, como bandas de color en las imágenes.

En una imagen tramada, como la que se presenta en la Figura 3.36 a modo de ejemplo, los colores no disponibles en la paleta se aproximan por una difusión de píxeles de color dentro de la gama de colores disponibles. El ojo humano percibe la difusión como una mezcla de los colores dentro de ésta. Las imágenes tramadas, en particular las que tienen relativamente pocos colores, a menudo se distinguen por un grado de granulosidad característico o por un aspecto

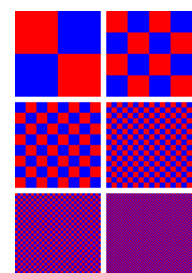


Figura 3.36: Una ilustración de tramado. El rojo y el azul son los únicos colores utilizados, pero a medida que los cuadrados rojos y azules se hacen más pequeños, el cuadrado aparece morado.

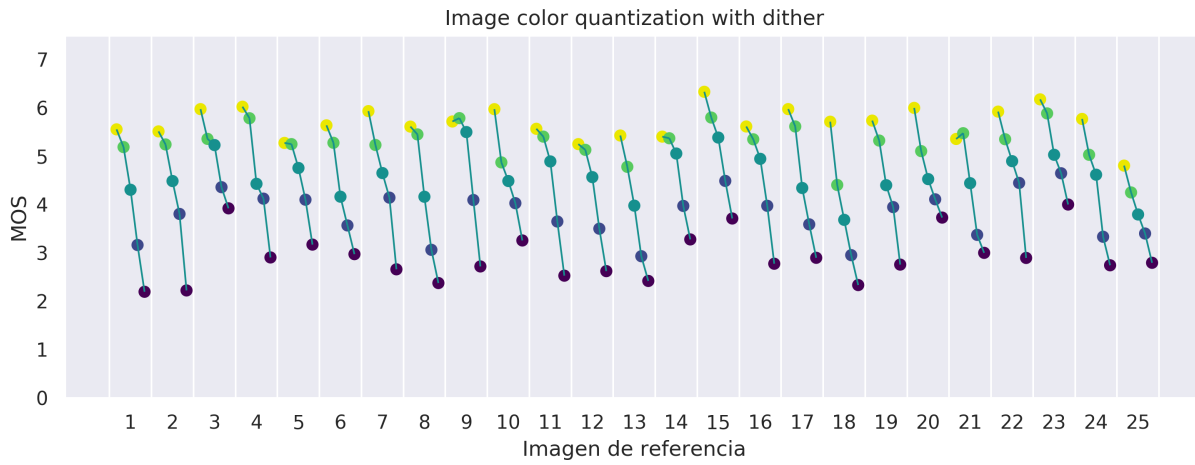


Figura 3.35: Diagrama de dispersión para la distorsión de tipo #22.

moteado. Este tipo de distorsión es típico durante la impresión de imágenes. La misma fue modelada en [8] usando la función *rgb2ind* de Matlab que convierte una imagen RGB a la imagen indexada usando dither.

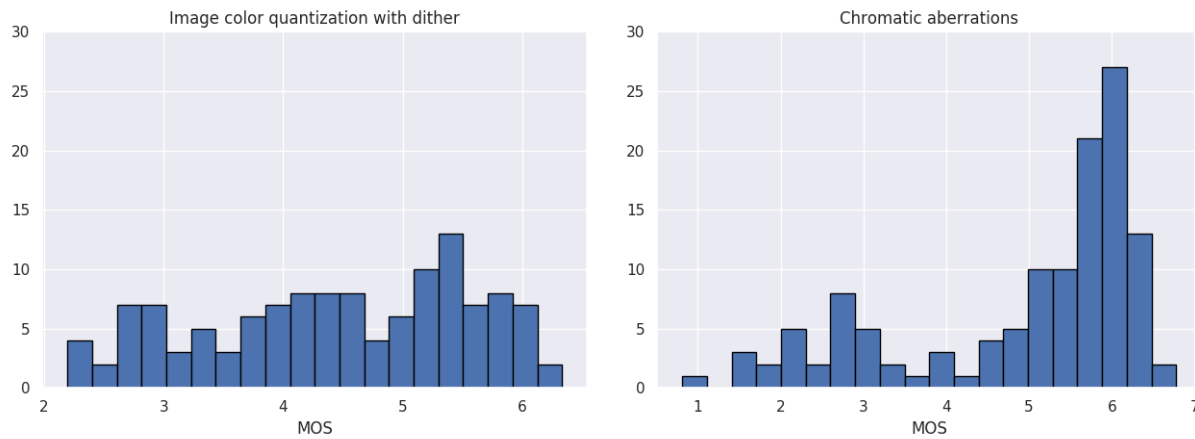


Figura 3.37: Histogramas del MOS para las distorsiones #22 y #23.

### 3.1.18. Distorsión #23: Aberraciones cromáticas

Este tipo de distorsión puede presentarse durante el proceso de obtener o transformar imágenes. Es una distorsión muy notable, especialmente en zonas de alto contraste y si el nivel de distorsión es alto. La distorsión se modeló mediante desplazamientos (*shifting*) mutuos de las componentes RGB respecto de ellas mismas [8]. Además, se ha producido un ligero desenfoque (*blurring*) de los componentes desplazados.

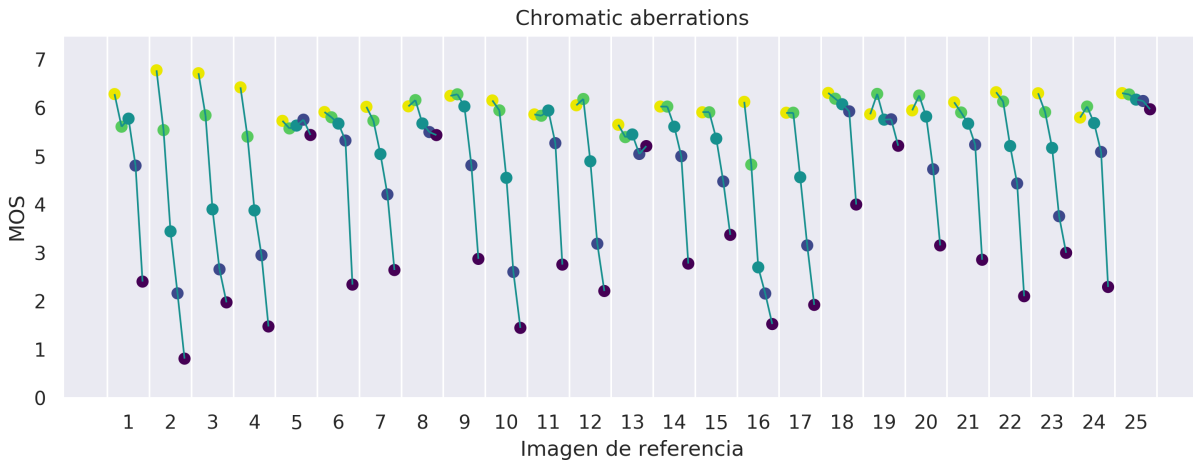


Figura 3.38: Diagrama de dispersión para la distorsión de tipo #23.

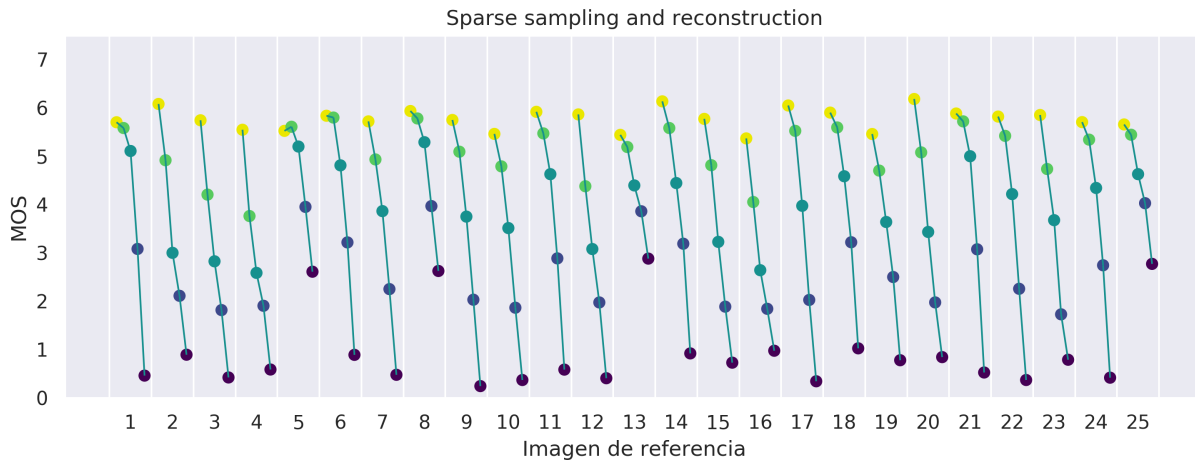


Figura 3.39: Diagrama de dispersión e Histograma del MOS para la distorsión de tipo #24. Se observa que para los niveles de distorsión altos el MOS es considerablemente pequeño.

### 3.1.19. Distorsión #24: Sparse sampling and reconstruction

Este tipo de distorsión está relacionada con el *compressive sensing*, también llamado *compressive sampling* o *sparse sampling*, y depende del método de *compressive sensing* utilizado [8].

Ponomarenko et al. [8] clasifican a estas 24 distorsiones, según sus características, en los siguientes grupos: Ruidosas, Actuales, Simples, Exóticas, Nuevas y Color, como se muestra en el Cuadro 3.1.

Número	Distorsión	Ruidosa	Actual	Simple	Exótica	Nueva	Color
1	Additive Gaussian noise	+	+	+	-	-	-
2	Additive noise in color comp.	+	-	-	-	-	+
3	Spatially correlated noise	+	+	-	-	-	-
4	Masked noise	+	+	-	-	-	-
5	High frequency noise	+	+	-	-	-	-
6	Impulse noise	+	+	-	-	-	-
7	Quantization noise	+	-	-	-	-	+
8	Gaussian blur	+	+	+	-	-	-
9	Image denoising	+	+	-	-	-	-
10	JPEG compression	-	+	+	-	-	+
11	JPEG2000 compression	-	+	-	-	-	-
12	JPEG transm. errors	-	-	-	+	-	-
13	JPEG2000 transm. errors	-	-	-	+	-	-
14	Non ecc. pattern noise	-	-	-	+	-	-
15	Local block-wise dist.	-	-	-	+	-	-
16	Mean shift (intensity shift)	-	-	-	+	-	-
17	Contrast change	-	-	-	+	-	-
18	Change of color saturation	-	-	-	-	+	+
19	Multipl. Gaussian noise	+	+	-	-	+	-
20	Comfort noise	-	-	-	+	+	-
21	Lossy compr. of noisy images	+	+	-	-	+	-
22	Image color quant. w. dither	-	-	-	-	+	+
23	Chromatic aberrations	-	-	-	+	+	+
24	Sparse sampl. and reconstr.	-	-	-	+	+	-

Cuadro 3.1: Grupos de distorsiones

El subconjunto “Ruido” contiene diferentes tipos de ruido y distorsiones en el procesamiento de imágenes convencional; el subconjunto “Actual” se refiere a los tipos de distorsiones más comunes en la práctica del procesamiento de imágenes, incluida la compresión; el subconjunto “Simple” incluye solo tres tipos estándar de distorsiones; el subconjunto “Exótico” corresponde a distorsiones que ocurren con poca frecuencia pero que se encuentran entre las “más difíciles” para las métricas de calidad visual. Además de estos subconjuntos estudiados anteriormente en la base TID2008 [6], se considerará también el subconjunto “Nuevo” que incluye los siete nuevos tipos de distorsiones introducidos en TID2013. El subconjunto “Color” se refiere a los tipos de distorsiones que se encuentran de una u otra manera conectado con cambios de contenido de color.

En la Figura 3.29 se ilustra a través de un diagrama de Venn como se clasifican las 24 distorsiones descritas en estos seis subconjuntos de distorsiones.

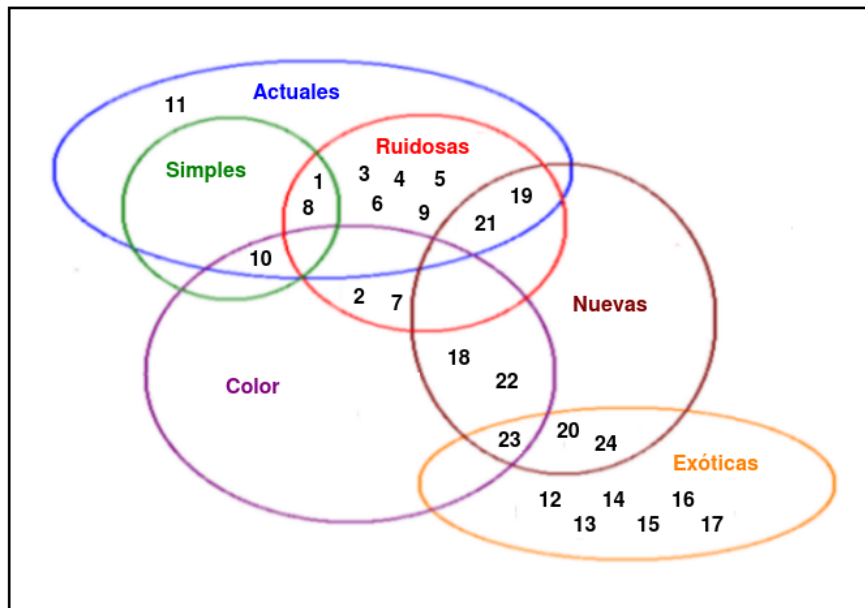


Figura 3.40: Diagrama de Venn con los grupos de distorsiones

---

# Capítulo 4

## Obtención y análisis de los datos

Obtener y analizar los datos requirió realizar un arduo trabajo de programación, el lenguaje elegido para hacerlo fue Python 2.7 [62]. Python es un lenguaje de programación interpretado, de tipado dinámico y una sintaxis que favorece un código legible; posee una licencia de código abierto (software libre) y es multiplataforma. Además se trata de un lenguaje multiparadigma, que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional.

Contamos con la implementación de las métricas MSE, RMSE, PSNR, SNR, WSNR, UQI, VIF, NQM, SSIM, MSSIM, Índice CQ, GSM, GMSD en este lenguaje, algunas fueron obtenidas de [63] mientras que otras las desarrollamos en base a los códigos disponibles en Matlab (software privativo).

Nombre	Imagen	Distorsión	Nivel	MOS	MOS-std	MSE	RMSE	...
i01-01-1.bmp	1	1	1	5,51429	0,13013	32,03753	5,66017	...
i01-01-2.bmp	1	1	2	5,56757	0,16008	63,92944	7,99559	...
i01-01-3.bmp	1	1	3	4,94444	0,15186	129,79863	11,39292	...
i01-01-4.bmp	1	1	4	4,37838	0,11046	257,90276	16,05935	...
i01-01-5.bmp	1	1	5	3,86486	0,10229	514,64925	22,68588	...
i01-02-1.bmp	1	2	1	6,02857	0,16896	30,91275	5,55992	...
i01-02-2.bmp	1	2	2	6,10811	0,17001	61,64238	7,85127	...
i01-02-3.bmp	1	2	3	5,38889	0,16795	123,78750	11,12598	...
.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	...

Cuadro 4.1: Matriz 3000 x 19 con los datos a analizar

Para generar los datos fue necesario desarrollar una rutina que computa cada una de estas medidas entre la imagen original y su distorsión para las 3000 imágenes distorsionadas de la base TID2013. Este proceso requirió varias horas de ejecución en una computadora estandar. Se generó una matriz de datos de 3000 por 19: las 3000 filas se corresponden con cada una de las 3000 imágenes distorsionadas; las columnas se corresponden con las variables: nombre y número de la



imagen, número de distorsión, nivel de distorsión, valor en el MOS, desviación estandar del MOS y en el resto de las columnas el valor obtenido por cada métrica al comparar la imagen con la original, como se muestra en el Cuadro 4.1.

Nombre	Imagen	Distorsión	Nivel	MOS	MOS-std	MSE	RMSE	...
i01-06-1.bmp	1	6	1	4,00000	0,11467	30,14498	5,49044	...
i01-06-2.bmp	1	6	2	3,89189	0,16566	65,25448	8,07802	...
i01-06-3.bmp	1	6	3	3,72973	0,10571	131,88924	11,48430	...
i01-06-4.bmp	1	6	4	3,48571	0,08448	256,30729	16,00960	...
i01-06-5.bmp	1	6	5	2,94444	0,11072	517,54441	22,74960	...
i02-06-1.bmp	2	6	1	4,55556	0,12696	46,53351	6,82155	...
i02-06-2.bmp	2	6	2	4,27778	0,10838	92,89449	9,63818	...
i02-06-3.bmp	2	6	3	3,72973	0,11241	189,48243	13,76526	...
.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	...

Cuadro 4.2: Matriz 125 x 19 con los datos para la distorsión #6

Con el fin de estudiar el desempeño de las métricas para cada tipo particular de distorsión, se generaron también 24 submatrices, una por cada distorsión, de 125 por 19: las 125 filas se corresponden con las imágenes distorsionadas por una misma distorsión, ya que son 25 imágenes originales con 5 niveles de distorsión cada una; las 19 columnas se corresponden con las mismas variables que en la matriz anterior. En el Cuadro 4.2 se muestra el formato que tendrá cada una de estas matrices.

Estas matrices obtenidas son los datos sobre los cuales se sustentará todo el análisis que sigue.

### 4.1. Breve análisis descriptivo

Una vez obtenidos los datos se calcularon algunas medidas de resumen, las mismas pueden apreciarse en los Cuadros 4.3 y 4.4. También se obtuvieron los histogramas y diagramas de caja con los 3000 valores correspondientes a cada métrica, los mismos pueden verse en las Figuras de 4.1 a 4.13.

Al igual que el MOS, los índices considerados no siguen distribuciones normales o conocidas, siendo la mayoría asimétricas, como se observa en los histogramas y diagramas de caja.

Para descartar la normalidad de los datos aplicamos el test de normalidad de Shapiro-Wilks y el test de Kolmogorov-Smirnov de bondad de ajuste utilizando las funciones `stats.shapiro()` [64] y `stats.kstest()` [65] del paquete **stats** de **scipy**.

El test de Shapiro-Wilks tiene como hipótesis nula que los datos fueron extraídos de una distribución Normal. Mientras que la hipótesis nula del test de Kolmogorov-Smirnov es que dos distribuciones son idénticas, en este caso se comparó la distribución Normal y la distribución del

Métrica	Media	Desviación Estandar	Varianza
MOS	4,47466	1,23967	1,53679
MSE	233,78198	283,37769	80302,91403
RMSE	13,37625	7,40661	54,85794
PSNR	26,86361	4,71222	22,20503
SNR	20,20610	4,97254	24,72613
WSNR	31,78474	9,41328	88,60992
UQI	0,64451	0,26199	0,06864
VIF	0,59575	0,20843	0,04344
NQM	24,08157	8,48928	72,06788
SSIM	0,85043	0,14097	0,01987
MSSIM	0,91083	0,09287	0,00863
CQ(1,1)	0,79928	0,23052	0,05314
GMSM	0,95762	0,05434	0,00295
GMSD	0,06911	0,06426	0,00413

Cuadro 4.3: Medidas de Resumen del MOS y las métricas

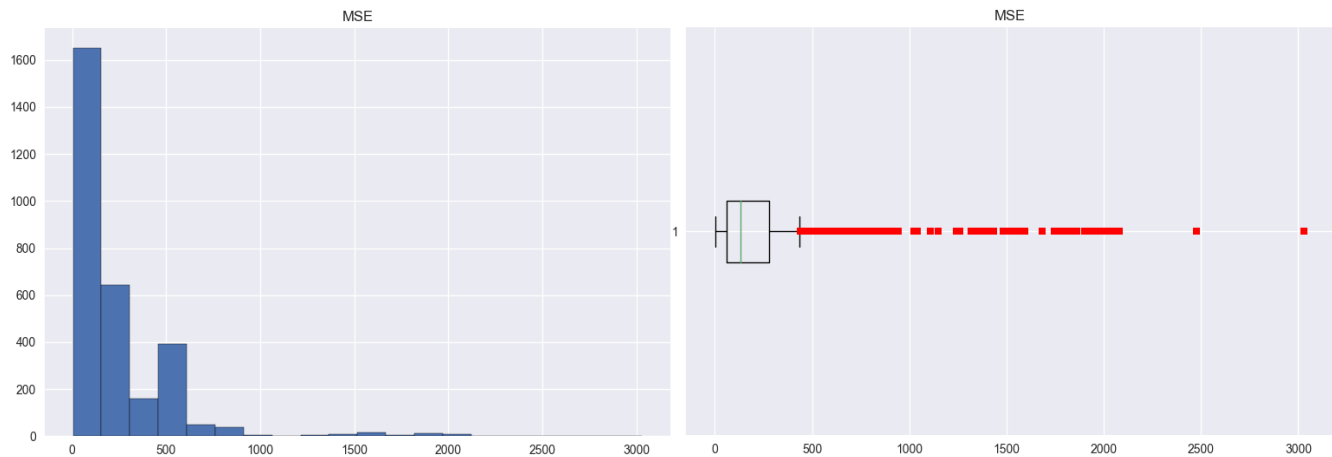


Figura 4.1: MSE

MOS; y la distribución Normal con la distribución de los datos obtenidos al aplicar cada métrica a las 3000 imágenes, respectivamente. Ambos tests dieron p-valores menores a 0,0001 para todas las métricas y para el MOS, indicando que las hipótesis nulas deben ser rechazadas y que por lo tanto los datos no son normales.

Hay dos formas de abordar esta situación, una opción podría ser intentar normalizar los datos con una transformación de Box-Cox, la otra recurrir a técnicas estadísticas no paramétricas.

Se transformaron los datos, 3000 valores de MOS y 3000 por cada métrica, mediante la transformación de Box-Cox utilizando la función `stats.boxcox()` [66], al aplicar ambos tests de normalidad a los datos transformados se concluyó que tampoco eran normales.

Recurriremos entonces al análisis de los datos con técnicas estadísticas no paramétricas como se explica en los siguientes capítulos.

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,24242	3,61111	4,60000	5,52632	7,21429
MSE	4,12544	61,60102	129,82053	277,06905	3029,35779
RMSE	2,03112	7,84863	11,39388	16,64539	55,03960
PSNR	13,3173	23,70492	26,99737	30,23492	41,97610
SNR	5,90372	16,29856	19,97778	24,28589	34,26859
WSNR	5,54221	24,31280	32,13409	39,01546	56,82755
UQI	-0,59986	0,45584	0,68147	0,87359	0,99843
VIF	0,02084	0,43516	0,59239	0,72822	1,10361
NQM	1,65970	17,92705	25,03627	30,66008	46,02755
SSIM	0,09164	0,78036	0,90021	0,95782	0,99869
MSSIM	0,30043	0,87240	0,94940	0,97843	0,99980
CQ(1,1)	-0,75296	0,71836	0,88843	0,96103	1,00008
GSM	0,53680	0,94252	0,98016	0,99375	0,99999
GMSD	0,00014	0,01586	0,04806	0,10982	0,29871

Cuadro 4.4: Medidas de Resumen del MOS y las métricas

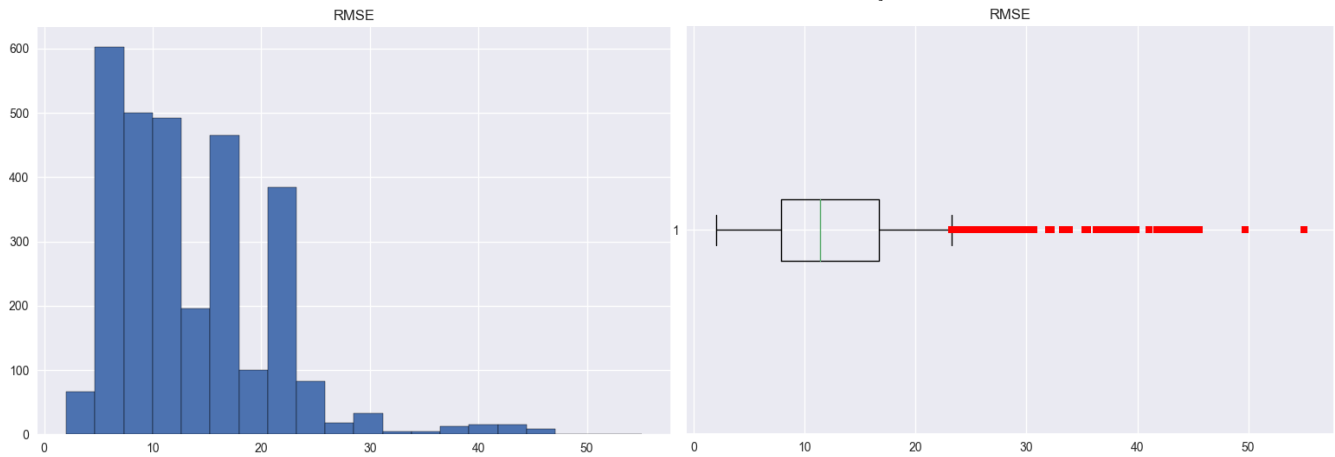


Figura 4.2: RMSE

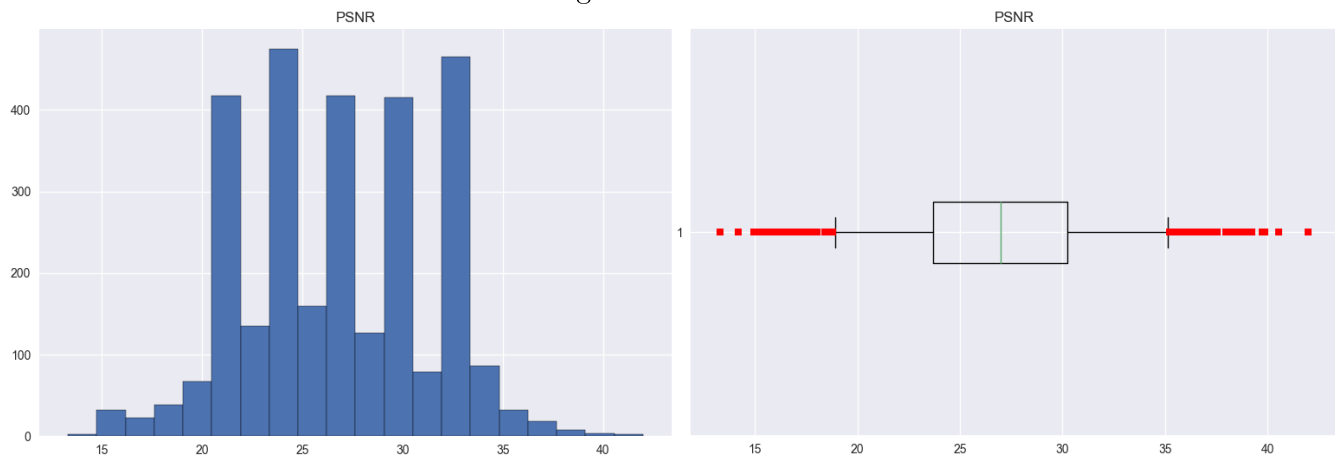


Figura 4.3: PSNR

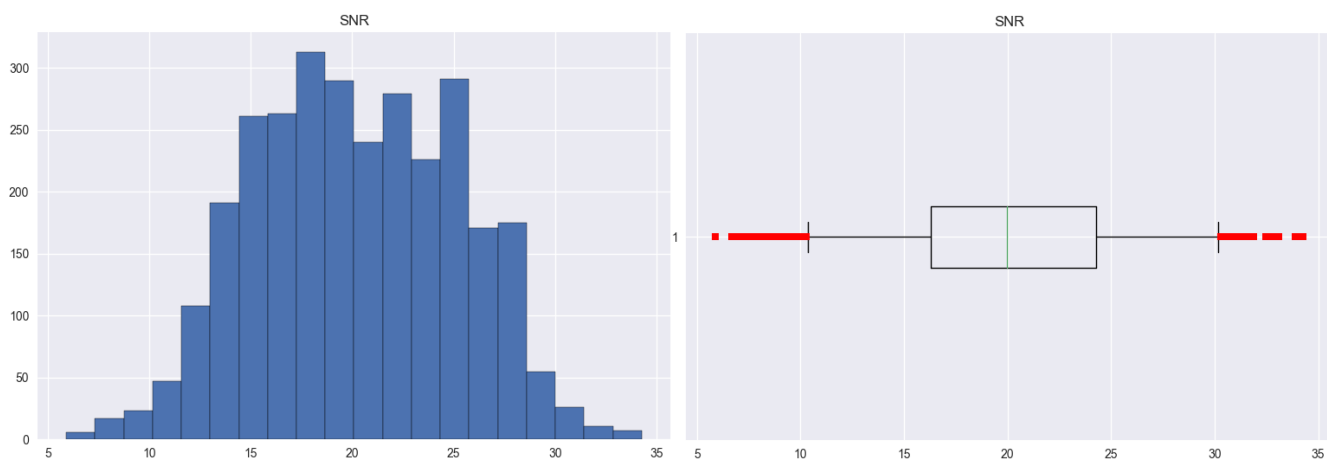


Figura 4.4: SNR

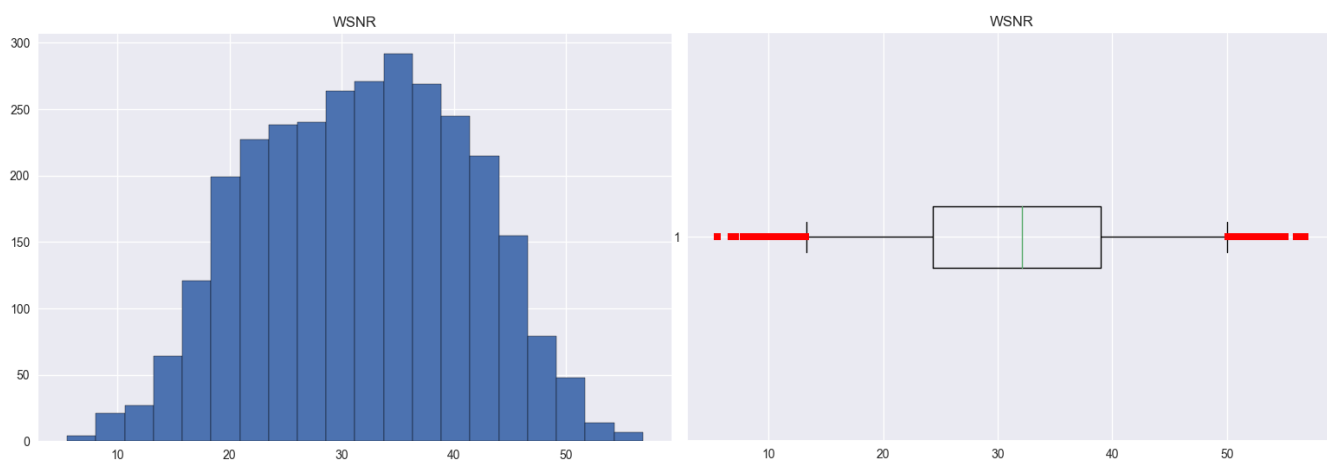


Figura 4.5: WSNR

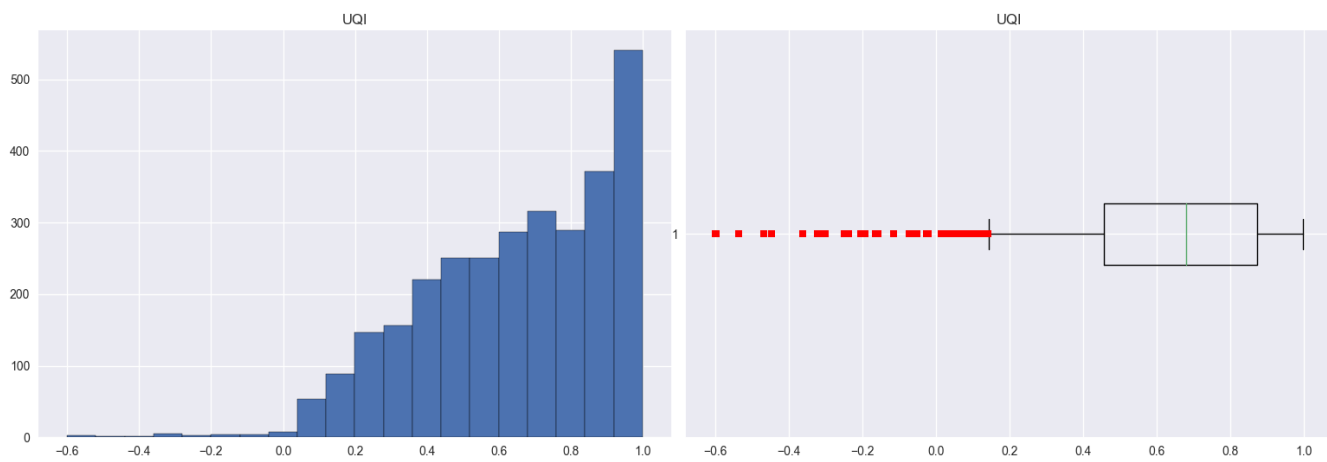


Figura 4.6: UQI

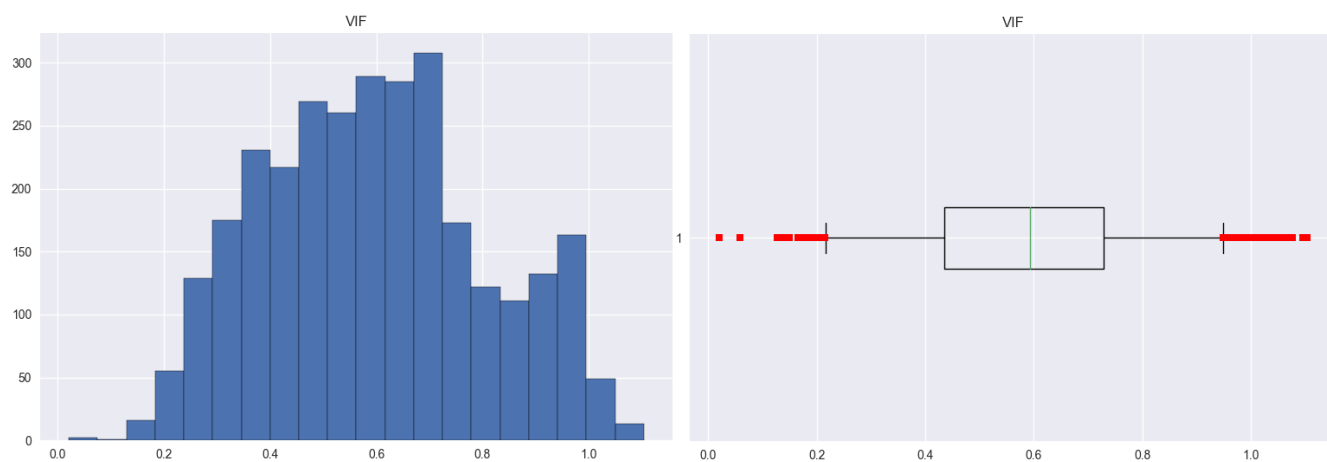


Figura 4.7: VIF

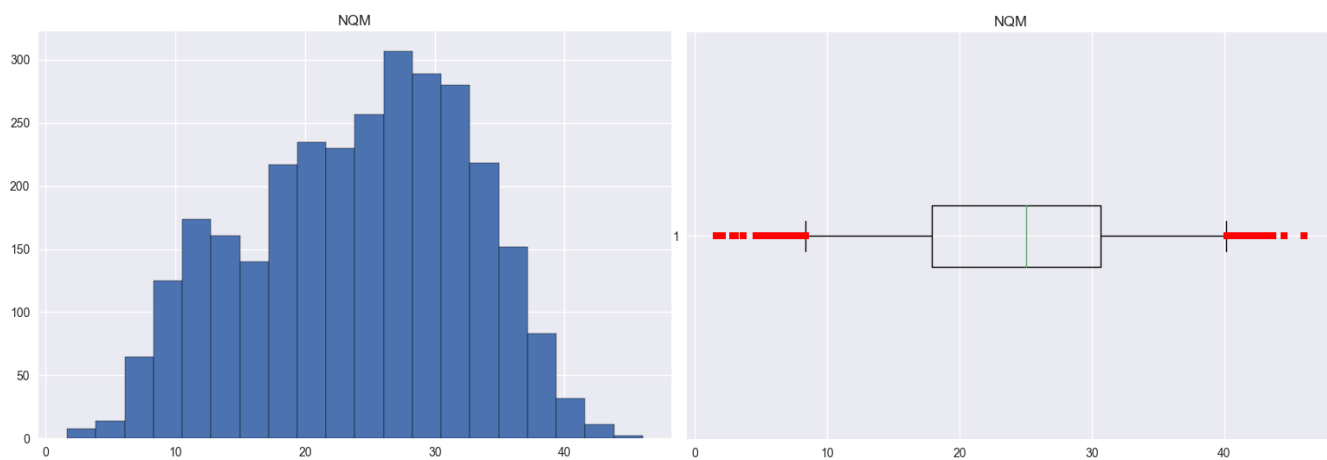


Figura 4.8: NQM

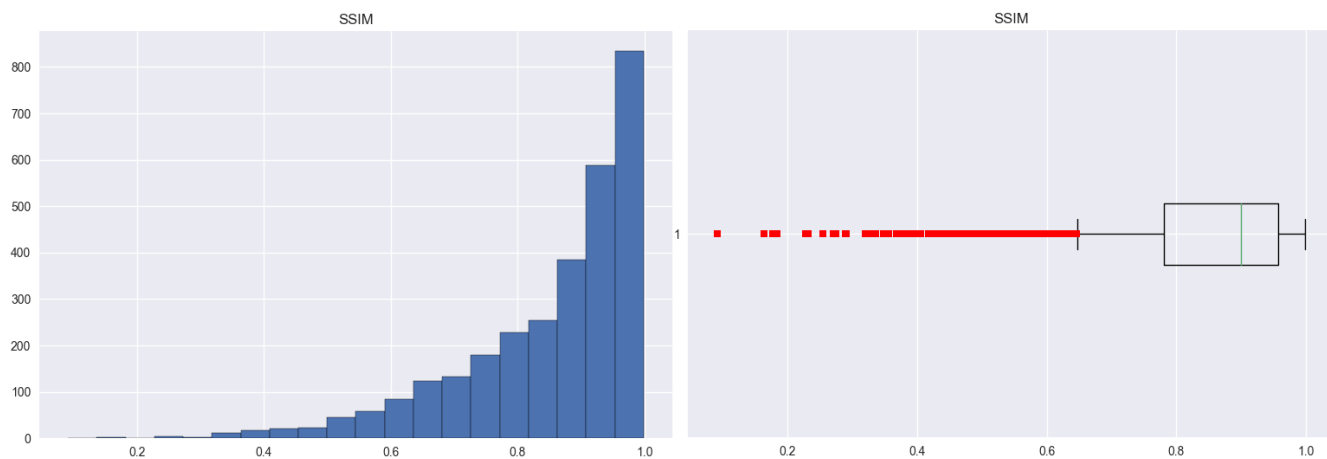


Figura 4.9: SSIM

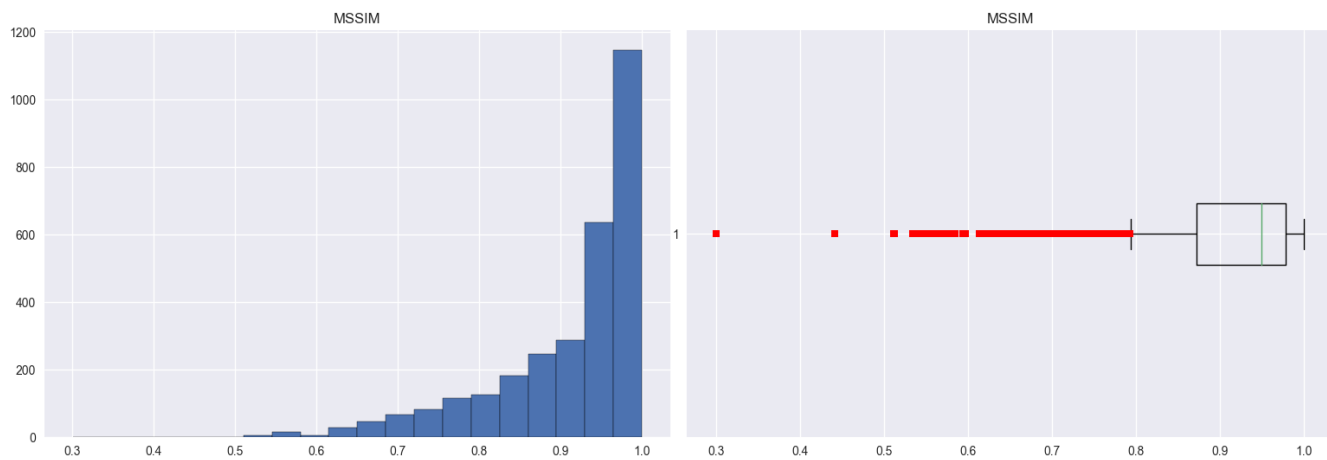


Figura 4.10: MSSIM

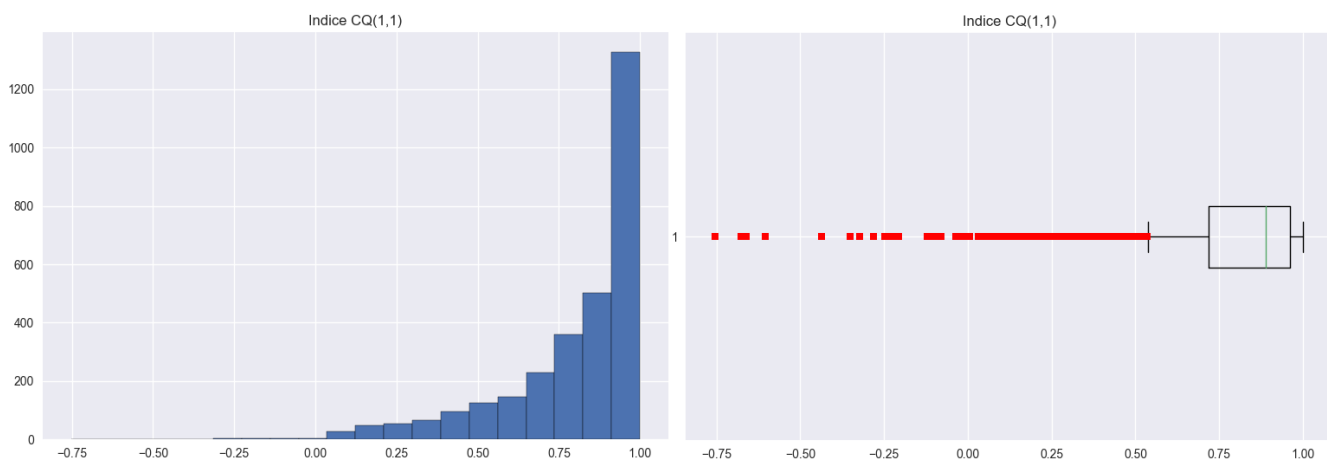


Figura 4.11: Indice CQ(1,1)

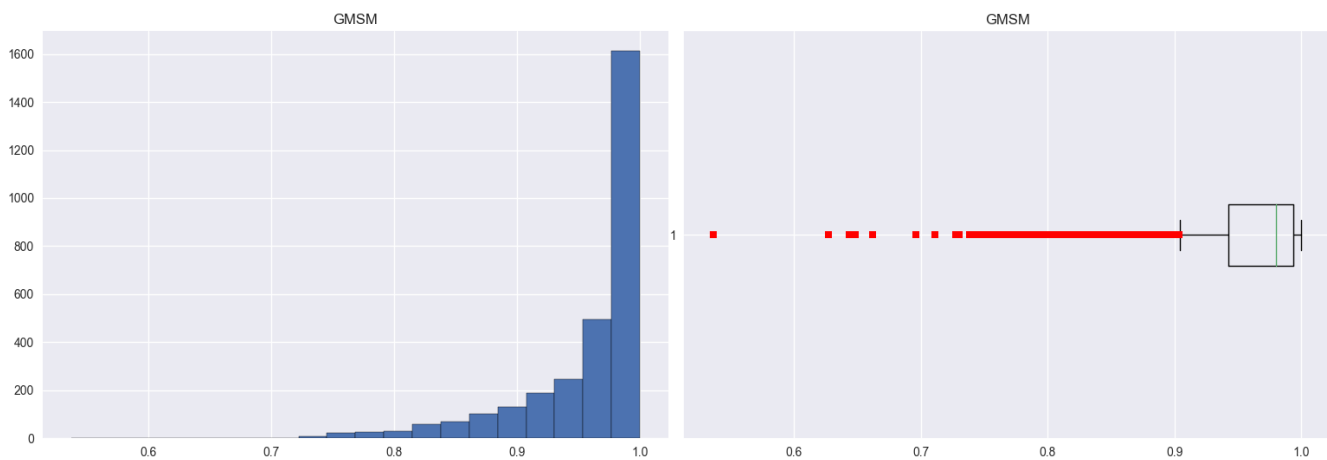


Figura 4.12: GSM

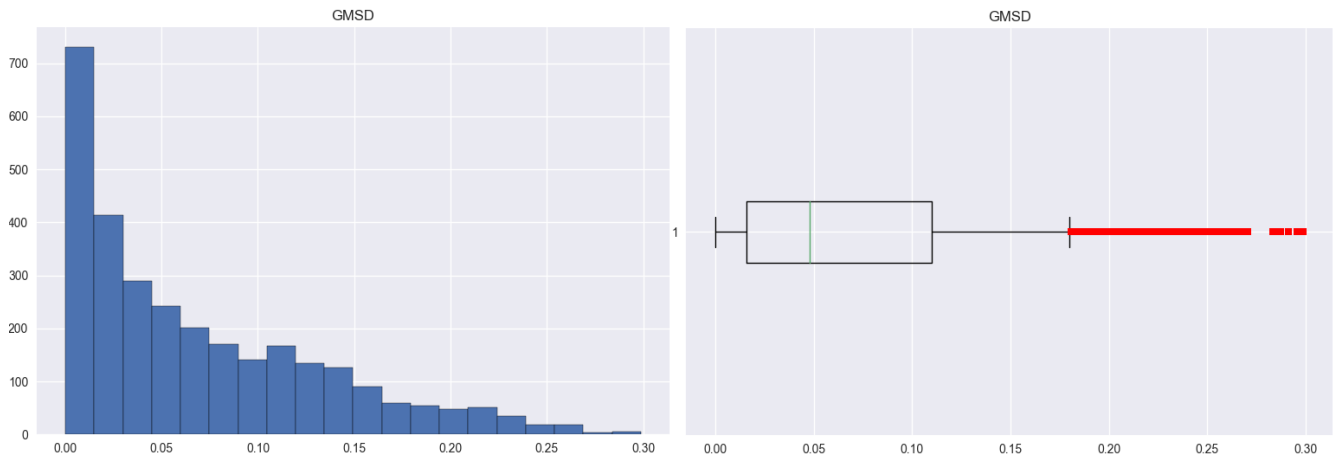


Figura 4.13: GMSD

---

# Capítulo 5

## Correlación

Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables,  $X$  e  $Y$ , existe una correlación positiva entre ellas si al aumentar los valores de  $X$  lo hacen también los de  $Y$ , equivalentemente, si cada vez que los valores de  $X$  aumentan, los de  $Y$  disminuyen, entonces existe una correlación negativa entre las variables. Es decir, la correlación entre dos variables estadísticas indica si existe una proporcionalidad entre ellas, aunque es importante tener en cuenta que la correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad.

Una forma de observar gráficamente la correlación entre dos variables es a través de un diagrama de dispersión, vinculando los valores de las dos variables  $X$  e  $Y$  en pares homónimos  $(x, y)$ , los mismos pueden ser graficados formando un conjunto de puntos en el plano cartesiano; de esta forma la relación entre las dos variables cuantitativas queda representada mediante la curva que mejor ajuste esa nube de puntos.

En caso de que las métricas PSNR, SNR, WSNR, UQI, PBVIF, NQM, SSIM, MSSIM, Índice CQ y GSMG estuvieran correlacionadas con el MOS, debería identificarse una correlación positiva, mientras que las métricas MSE, RMSE y GMSD deberían presentar una correlación negativa, esto se deduce de las definiciones de cada métrica, presentadas en el capítulo 2, y de los resultados obtenidos hasta el momento.

Se realizaron los diagramas de dispersión entre el MOS y cada métrica para analizar si existe correlación. Se asignó un color a cada tipo de distorsión para hacer más sencilla la interpretación visual por distorsión. Los mismos pueden apreciarse en las Figuras 4.1 - 4.13. También realizamos y analizamos los diagramas de dispersión de las 13 métricas y el MOS para cada una de las 24 distorsiones individualmente; estos 312 gráficos se encuentran en los Apéndices A, B y C.



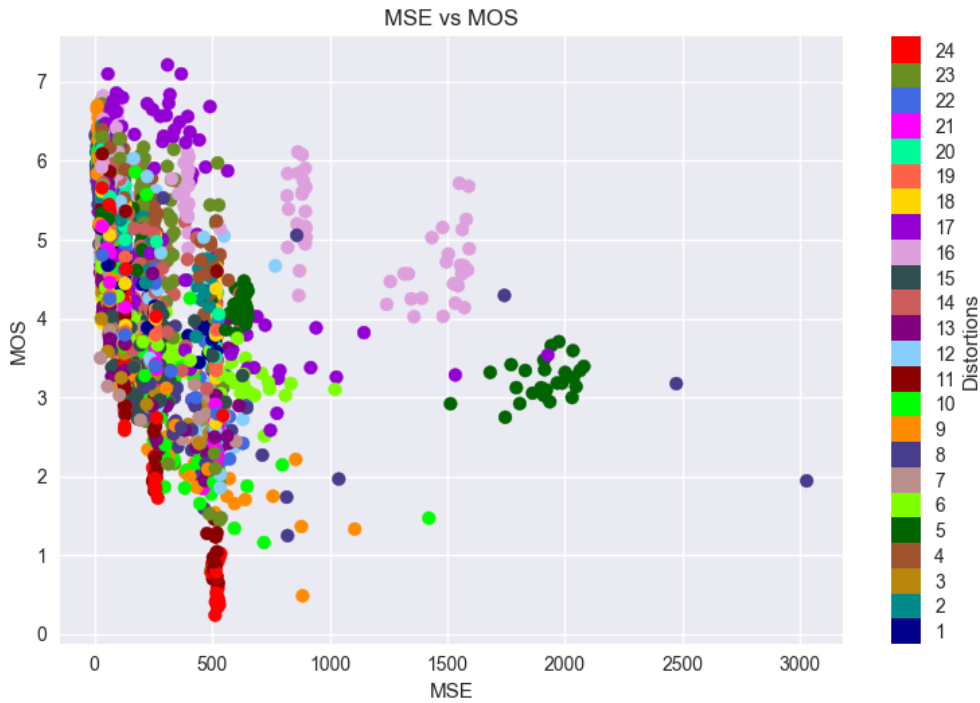


Figura 5.1: Diagrama de dispersión MSE vs MOS. En general no parece haber una correlación significativa con el MOS, aunque habría que analizar para cada tipo de distorsión por separado.



Figura 5.2: Diagrama de dispersión RMSE vs MOS. No parece haber una correlación significativa con el MOS, pero podría haberla para alguna distorsión en particular, por ejemplo la #6. También puede observarse que los puntos correspondientes a las distorsiones #11, #16 y #24, están alineados horizontalmente por grupos.

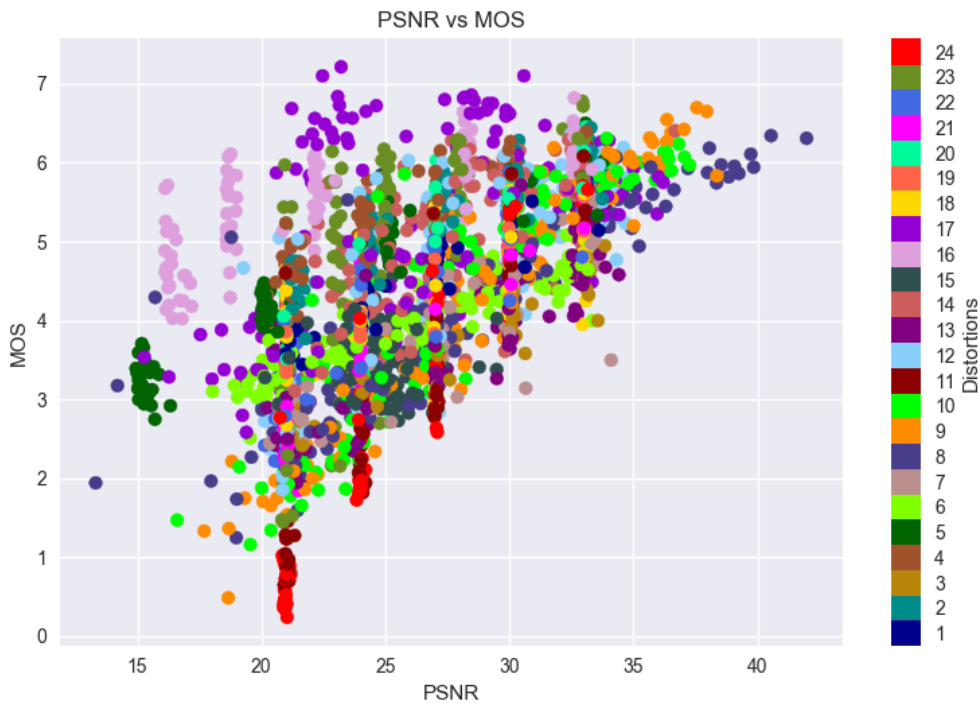


Figura 5.3: Diagrama de dispersión PSNR vs MOS. Se observa que con las distorsiones #6, #9 y #10 la métrica presenta una mayor correlación con el MOS. Al igual que en el diagrama anterior, los puntos correspondientes a las distorsión #11, #16 y #24, están alineados horizontalmente por grupos. Los puntos de la distorsión #5 también están agrupados.

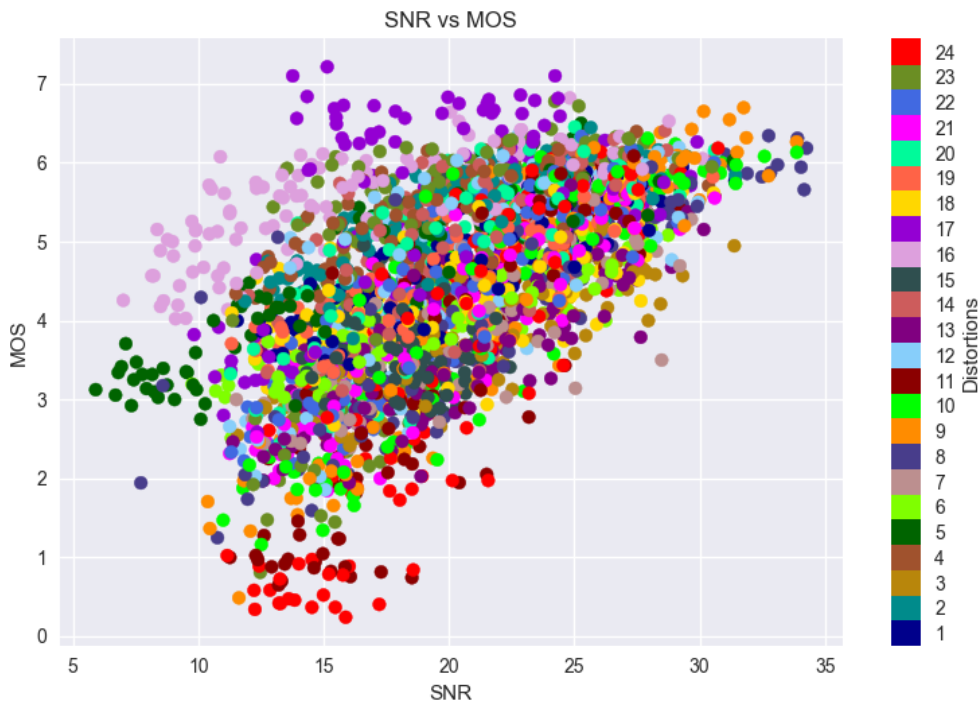


Figura 5.4: Diagrama de dispersión SNR vs MOS. No se percibe una tendencia clara, hay que analizar para cada distorsión en particular.

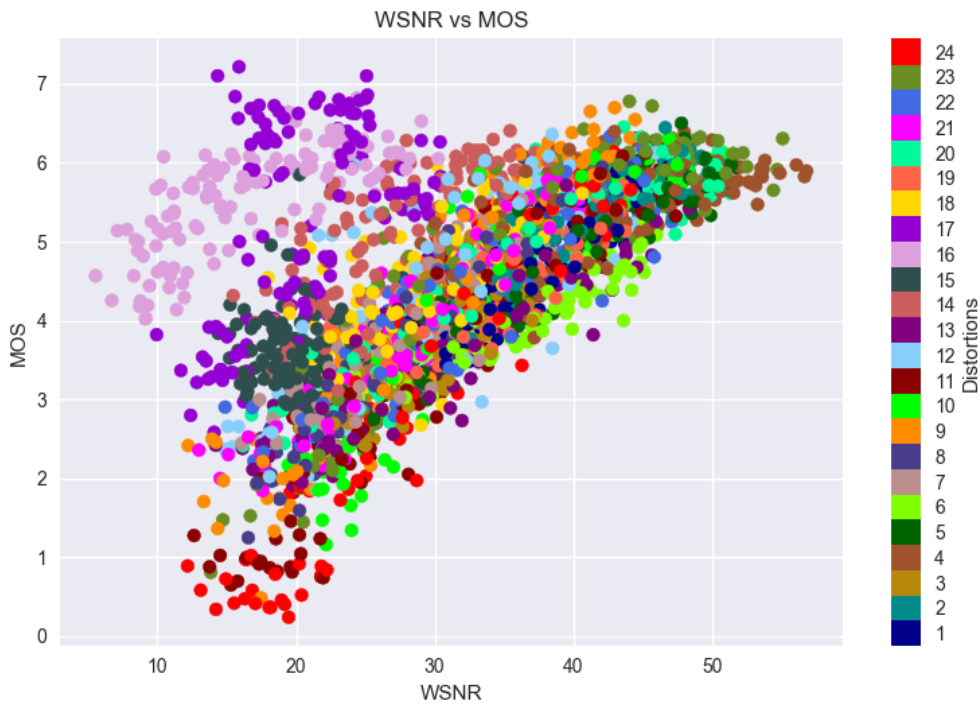


Figura 5.5: Diagrama de dispersión WSNR vs MOS. Puede apreciarse que para las distorsiones #15, #16 y #17, la métrica presenta un comportamiento particular. Sacando estos puntos, el resto parece seguir una correlación lineal positiva.

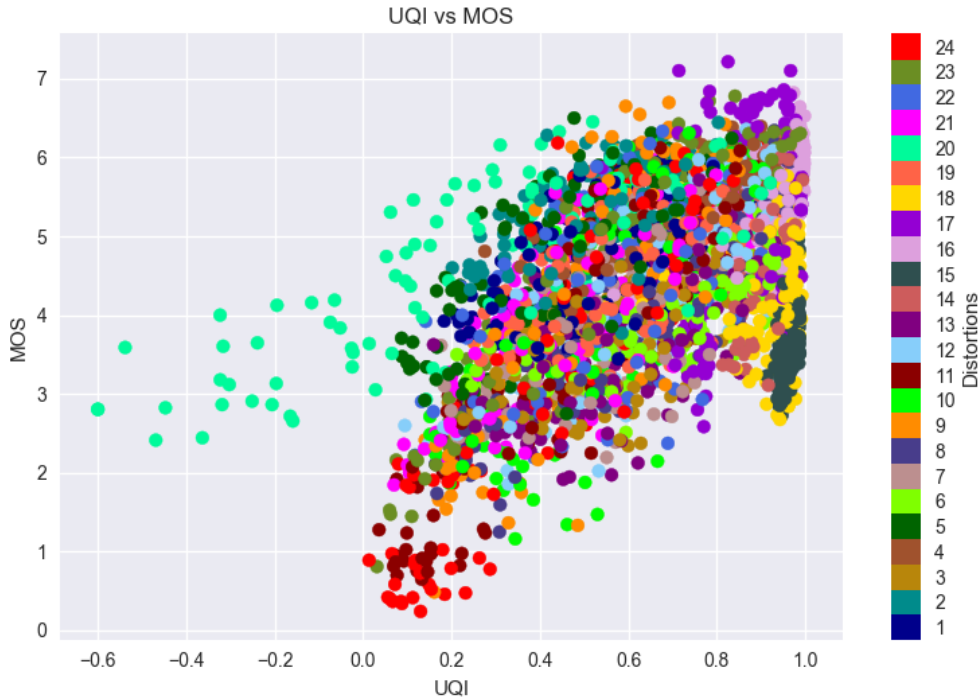


Figura 5.6: Diagrama de dispersión UQI vs MOS. Se observa que los puntos correspondientes a la distorsión #20 presentan otro tipo de dispersión y son los únicos que toman valores negativos. Por otra parte, los puntos correspondientes a las distorsiones #15 y #18 están todos concentrados en una zona. El resto de los puntos podrían presentar una correlación positiva.

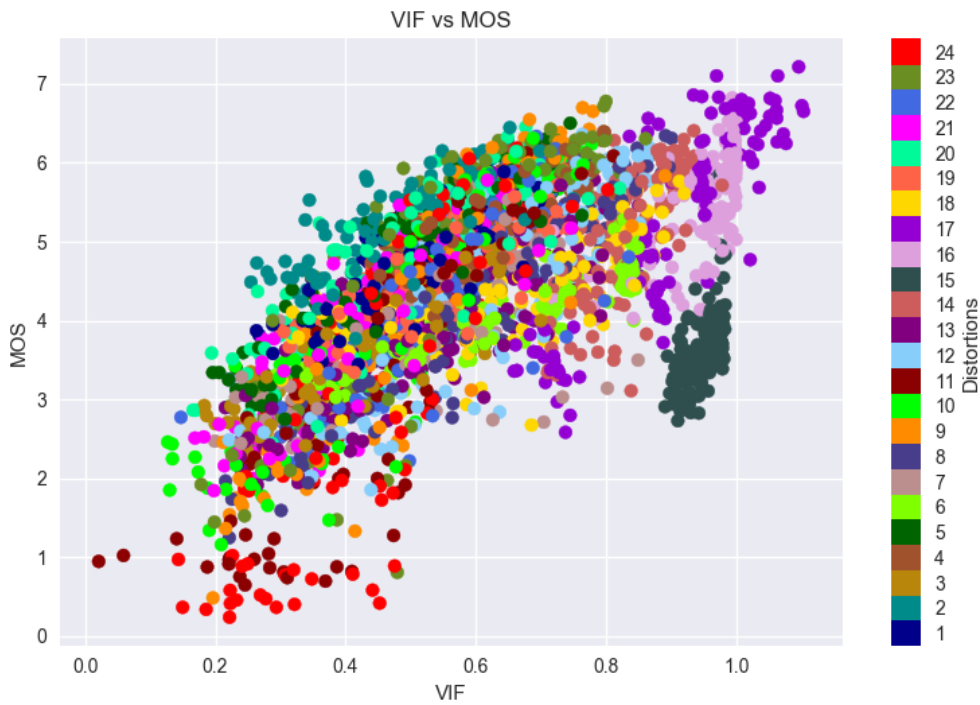


Figura 5.7: Diagrama de dispersión VIF vs MOS. Puede apreciarse que los puntos de las distorsiones #15 y #16 presentan un comportamiento atípico en relación a los otros, al concentrarse todos en una zona.

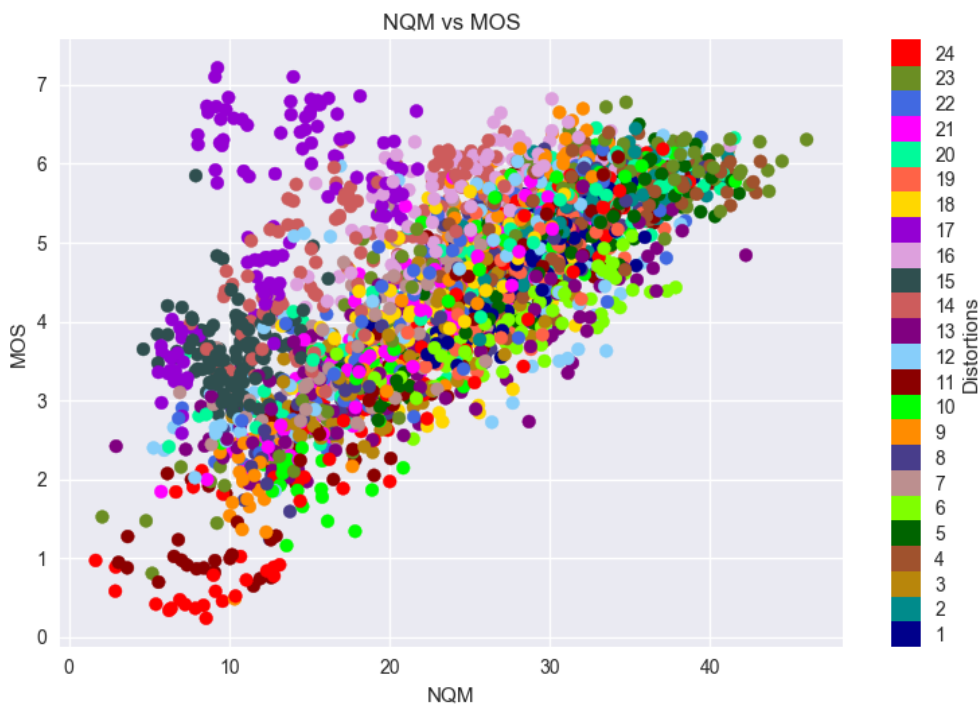


Figura 5.8: Diagrama de dispersión NQM vs MOS. Gráficamente puede verse que sacando los puntos de las distorsiones #14, #15, #16 y #17, el resto parece estar correlacionado positivamente con el MOS. Los puntos de la distorsión #17 presentan una mayor dispersión que el resto, mientras que los de la #15 están concentrados en una zona.



Figura 5.9: Diagrama de dispersión SSIM vs MOS. En general no se percibe una tendencia clara, hay que analizar para cada distorsión en particular. Los puntos de la distorsión #15 están concentrados en una zona.

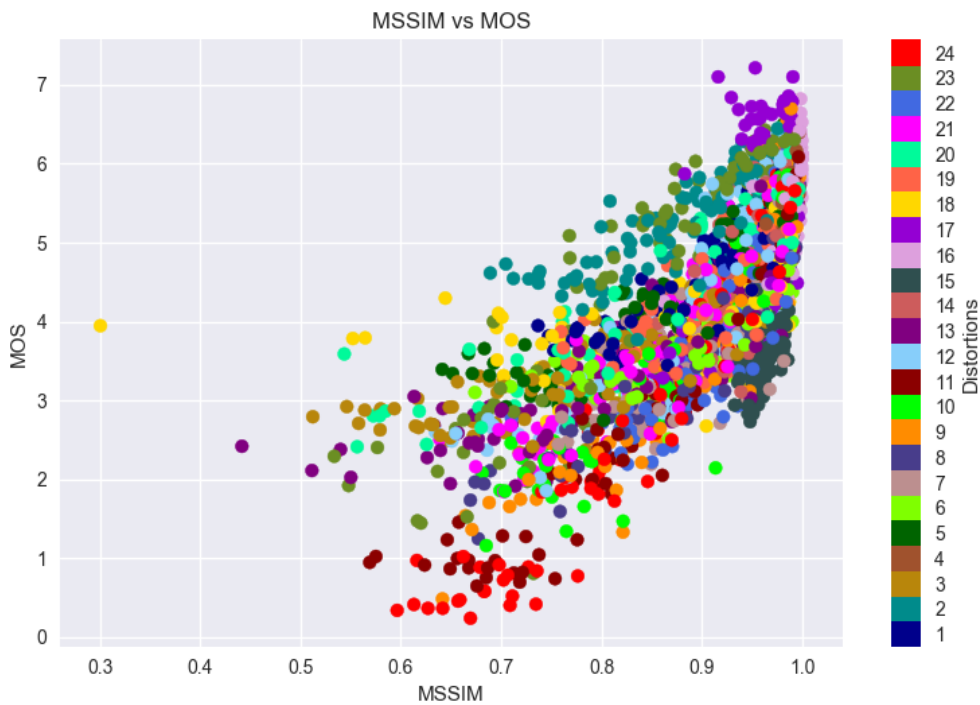


Figura 5.10: Diagrama de dispersión MSSIM vs MOS. La tendencia general parece ser una correlación lineal positiva. Los puntos de la distorsión #18 presentan una tendencia particular con valores atípicos. Los puntos de la distorsión #15 parecen estar todos concentrados en una zona.

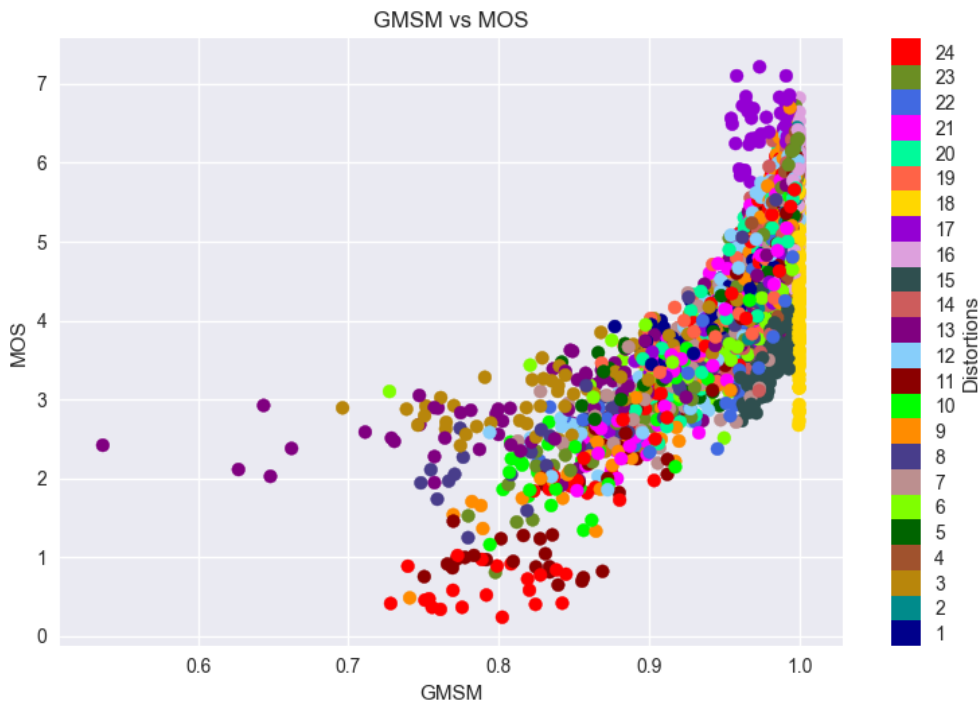


Figura 5.11: Diagrama de dispersión GMSM vs MOS. Se observa que la métrica toma el mismo valor en todas las imágenes distorsionadas con la distorsión #18, por lo que claramente esta métrica no es buena para este tipo de distorsión. Los puntos correspondientes a la distorsión #15 parecen estar todos concentrados en una misma zona. Los puntos correspondientes a la distorsión #13 presentan una tendencia distinta del resto. Para las otras distorsiones, en general parece darse una correlación lineal positiva.

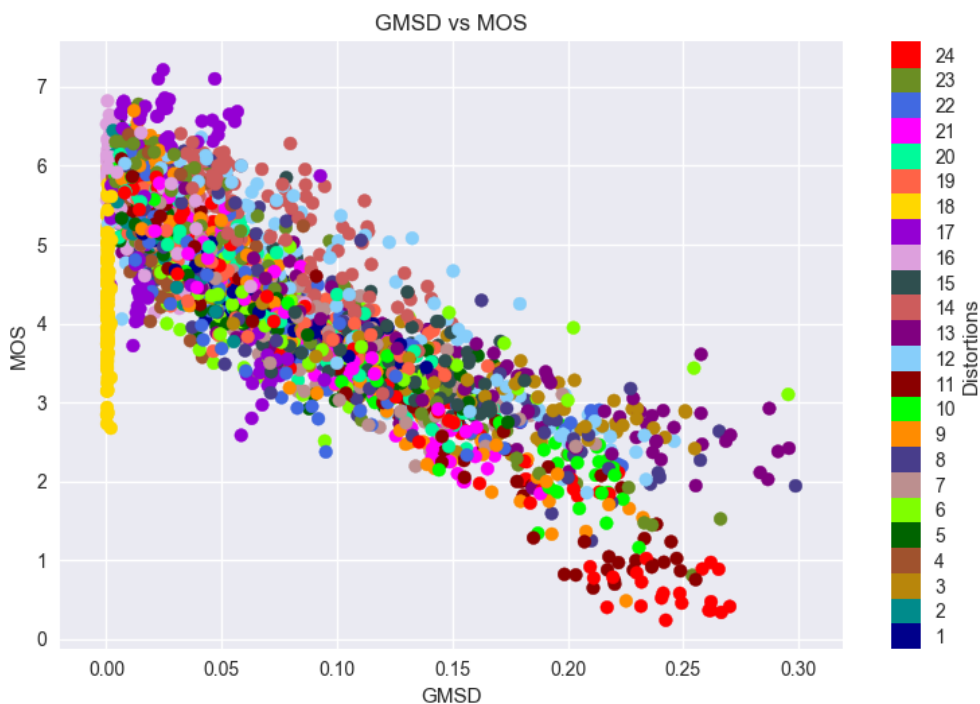


Figura 5.12: Diagrama de dispersión GMSD vs MOS. Ocurre lo mismo que con la métrica GMSM para la distorsión #18. Para el resto de los puntos se observa una correlación lineal negativa.

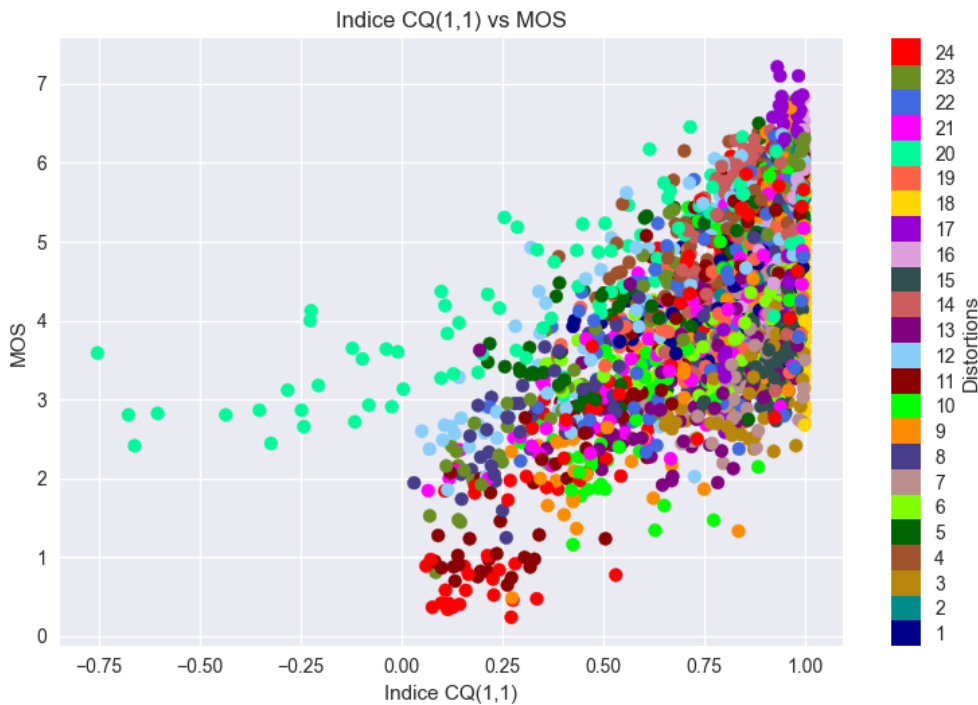


Figura 5.13: Diagrama de dispersión Índice CQ(1,1) vs MOS. Puede verse que la distorsión #20 presenta una dispersión diferente del resto y es la única que toma valores negativos. Para el resto de los puntos la tendencia no es clara, deben observarse cada caso por separado.

## 5.1. Técnicas no paramétricas de correlación

El coeficiente de correlación más ampliamente utilizado es el coeficiente de correlación de Pearson, el mismo exige como supuesto que la distribución conjunta de las dos variables involucradas sea una normal bivariada. No siendo este el caso de las variables que nos interesa analizar, el MOS con cada una de las métricas implementadas, hemos recurrido a otras técnicas no paramétricas de correlación, donde la distribución de las variables no es importante, entre las que se destacan los coeficientes de correlación de Spearman y Kendall. Estos coeficientes se basan en una ordenación de los datos en rangos, lo cual evita también el problema de comparar variables con escalas muy diferentes.

## 5.2. Correlación de Spearman

La correlación  $\hat{\rho}$  de Spearman es una medida no paramétrica de la monotonicidad de la relación entre dos conjuntos de datos [67]. A diferencia de la correlación de Pearson, no supone que ambos conjuntos de datos se distribuyen normalmente.

Sean  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  observaciones del vector aleatorio  $(X, Y)$ . Para poder calcular  $\hat{\rho}$ , primero los datos son reemplazados por su respectivo orden, al cual denominaremos rango

<sup>1</sup>, al ser ordenados de menor a mayor. Se tiene que considerar la existencia de datos idénticos (denominados lazos) al asignar los rangos, en este caso se ajusta asignando a todos los valores el promedio de los rangos correspondientes, por ejemplo si hay cuatro valores repetidos a los cuales se les debía asignar los rangos 6, 7, 8, 9, se le asigna el promedio de estos rangos a todos, es decir a esos cuatro valores se les asigna un rango de 7,5.

La ecuación para calcular el coeficiente  $\rho$  es entonces:

$$\hat{\rho} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \tag{5.1}$$

donde  $n$  es la cantidad de pares ordenados en la muestra, en nuestro caso  $n = 3000$  y  $d_i$  es la diferencia entre rangos de cada par  $(x_i, y_i)$ ,  $d_i = R(x_i) - R(y_i)$ .

Al igual que otros coeficientes de correlación, éste varía entre  $-1$  y  $+1$  y un coeficiente igual a  $0$  implica que no hay correlación. Los valores positivos cercanos a  $1$  indican que cuando una variable aumenta, también lo hace la otra, mientras que los negativos cercanos a  $-1$  establece que cuando una variable aumenta, la otra disminuye. Las correlaciones de  $-1$  o  $+1$  implican una relación monotónica exacta. Todas las interpretaciones posibles se encuentran en el Cuadro 5.1 <sup>2</sup>.

Coeficiente de correlación estimado en el intervalo:	Correlación
$(0; 0,2] \cup [-0,2; 0)$	Muy pobre
$(0,2; 0,4] \cup [-0,4; 0 - ,2)$	Pobre
$(0,4; 0,6] \cup [-0,6; -0,4)$	Moderada
$(0,6; 0,8] \cup [-0,8; -0,6)$	Fuerte
$(0,8; 1] \cup [-1; -0,8)$	Muy fuerte

Cuadro 5.1: Interpretación del coeficiente de correlación estimado

Los resultados obtenidos para la correlación de Spearman entre el MOS y cada métrica pueden apreciarse en Cuadro 5.2. El MSE y el RMSE obtuvieron los mismos coeficientes, ya que el RMSE es la raíz cuadrada del MSE, mientras que los coeficientes del PSNR tienen el mismo módulo que los coeficientes del MSE, pero distinto signo. Los módulos de todos los coeficientes se mueven en un rango de  $0,5$  a  $0,8$ . El coeficiente de módulo más cercano a  $1$  lo obtuvo el GMSD, mientras que el más cercano a  $0$  fue el del UQI.

### 5.2.1. Prueba de Hipótesis

El coeficiente de correlación de Spearman tiene asociada una prueba de hipótesis bilateral cuya hipótesis nula es que los dos conjuntos de datos no están correlacionados:

- Hipótesis nula ( $H_0$ ): No existe correlación entre las variables  $X$  e  $Y$ ,  $\rho = 0$ .

<sup>1</sup>Rank o ranking en inglés.

<sup>2</sup>No hay acuerdo en la comunidad científica sobre cuál es la forma precisa de interpretar los valores intermedios de los coeficientes de correlación, esta interpretación fue tomada de [68] y la usaremos para todos los que presentaremos, pero la misma es arbitraria, en [69] se propone otra.



- Hipótesis alternativa ( $H_a$ ): Existe correlación entre las variables  $X$  e  $Y$ , es decir, el coeficiente  $\rho$  es significativamente distinto de cero,  $\rho \neq 0$ .

El estadístico de prueba:

$$t = \frac{\hat{\rho}}{\sqrt{(1 - \hat{\rho}^2)/(n - 2)}} \tag{5.2}$$

tiene una distribución aproximada  $t$  de Student con  $n - 2$  grados de libertad bajo la hipótesis nula de correlación 0.

El  $p$ -valor obtenido indica aproximadamente la probabilidad de que un sistema no correlacionado produzca conjuntos de datos que tengan una correlación de Spearman al menos tan extrema como la calculada a partir de estos conjuntos de datos. Todos los  $p$ -valores dieron menor que 0,0001 indicando que los coeficientes de correlación son significativamente no nulos.

### 5.3. Correlación de Kendall

La  $\hat{\tau}$  de Kendall es una medida de la correspondencia entre dos rankings, es decir evalúa la similitud de los ordenamientos de los datos cuando se clasifican por cada tipo de ranking, en este caso el MOS y la métrica considerada.

Este coeficiente se basa más en la jerarquización de las observaciones que en los propios datos, esto hace que la distribución de  $\hat{\tau}$  sea independiente de la que presentan las variables  $X$  e  $Y$ , siempre y cuando los datos representados por estas dos variables sean independientes y continuos.

Este coeficiente se define como:

$$\hat{\tau} = \frac{S_c - S_d}{\frac{n(n-1)}{2}} \tag{5.3}$$

donde  $n$  es el número de datos  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  tenidos en cuenta y  $S_c$  y  $S_d$  son el total de pares concordantes y el total de pares discordantes respectivamente, los cuales se describen a continuación.

Notar que  $\frac{n(n-1)}{2}$  es la cantidad de conjuntos de pares que se pueden formar con  $n$  datos.

Un par de observaciones  $\{(x_i, y_i), (x_j, y_j)\}$  con  $i \neq j$ , se dice concordante si los ordenamientos según los rankings por  $X$  y por  $Y$  coinciden, es decir si se cumplen  $x_i > x_j$  y  $y_i > y_j$  ó  $x_i < x_j$  y  $y_i < y_j$ . El par será discordante si se verifica que  $x_i > x_j$  y  $y_i < y_j$  o que  $x_i < x_j$  y  $y_i > y_j$ . Si  $x_i = x_j$  o  $y_i = y_j$ , el par no es ni concordante, ni discordante, en este caso se dice que el par de datos  $\{(x_i, y_i), (x_j, y_j)\}$  está enlazado. Notar que el par  $\{(x_i, y_i), (x_j, y_j)\}$  será concordante cuando  $x_j - x_i$  y  $y_j - y_i$  tengan el mismo signo y discordante en caso de tener el signo contrario.

$S = S_c - S_d$  mide la fuerza de la relación entre las variables  $X$  e  $Y$ , una preponderancia de pares concordantes resulta en un gran valor positivo de  $S$  que indica una fuerte relación positiva entre  $X$  e  $Y$ ; una preponderancia de pares discordantes resulta en un gran valor negativo de  $S$

que indica una fuerte relación negativa entre  $X$  e  $Y$ . Pero como medida de la fuerza de la relación entre dos variables  $S$  tiene una desventaja, su rango depende de el tamaño de la muestra  $n$ , siendo el máximo valor posible  $\frac{n(n-1)}{2}$  y el mínimo  $-\frac{n(n-1)}{2}$ . Si estandarizamos  $S$  dividiendo por  $\frac{n(n-1)}{2}$ , obtenemos la fórmula para  $\hat{\tau}$  dada en (5.3) que verifica  $-1 \leq \hat{\tau} \leq 1$ . El valor máximo 1 se obtiene si todos los pares son concordantes y el mínimo  $-1$  si todos los pares son discordantes [70].

El coeficiente  $\hat{\tau}$  se interpreta igual que  $\hat{\rho}$ , valores cercanos a 1 indican una fuerte asociación directa, valores cercanos a  $-1$  una fuerte asociación inversa, valores cercanos a 0 indican ausencia de correspondencia. Para más detalles referirse nuevamente al Cuadro 5.1.

Dado que nuestras variables presentan pares de datos enlazados (por ejemplo, hay imágenes que tienen el mismo MOS, otras imágenes el mismo valor para cierta métrica), utilizaremos la versión  $\tau_B$  del coeficiente que realiza los ajustes necesarios para tratar estos casos:

$$\hat{\tau}_B = \frac{S_c - S_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \tag{5.4}$$

donde,  $n_0 = \frac{n(n-1)}{2}$ ,  $n_1 = \sum_i \frac{t_i(t_i-1)}{2}$ ,  $n_2 = \sum_j \frac{u_j(u_j-1)}{2}$  y

$t_i$  = número de datos enlazados en el  $i$ -ésimo grupo de lazos del conjunto  $X$ ,

$u_j$  = número de datos enlazados en el  $j$ -ésimo grupo de lazos del conjunto  $Y$ .

Para profundizar el análisis se calcularon las correlaciones de Kendall de cada métrica con el MOS, las mismas pueden apreciarse en el Cuadro 5.2.

Métrica	Corr. Kendall	Correlación	Corr. Spearman	Correlación
MSE & RMSE	-0,4958	Moderada	-0,6869	Fuerte
PSNR	0,4958	Moderada	0,6869	Fuerte
SNR	0,4607	Moderada	0,6491	Fuerte
WSNR	0,4938	Moderada	0,6382	Fuerte
UQI	0,3695	Pobre	0,5239	Moderada
VIF	0,4669	Moderada	0,6338	Fuerte
NQM	0,5348	Moderada	0,7126	Fuerte
SSIM	0,4457	Moderada	0,6273	Fuerte
MSSIM	0,5921	Moderada	0,7909	Fuerte
CQ(1,1)	0,4291	Moderada	0,6008	Fuerte
GMSM	0,6132	Fuerte	0,7884	Fuerte
GMSD	-0,6339	Fuerte	-0,8044	Muy Fuerte

Cuadro 5.2: Coeficientes de Kendall y Spearman entre el MOS y cada métrica.

En general los coeficientes de Kendall dieron valores mucho más bajos que los de Spearman. El MSE y el RMSE obtuvieron los mismos coeficientes, mientras que PSNR coficientes con el mismo módulo que los de estas dos. Las métricas GMSM y GMSD son las que parecen estar más

correlacionadas con el MOS según este coeficiente, mientras que, al igual que lo que sucedió con el coeficiente de Spearman, UQI es la que presenta menor correlación con el MOS.

Según ambos coeficientes, las métricas más correlacionadas con el MOS serían GSM, GMSD, MSSIM y NQM, mientras que la menos correlacionada sería la UQI. Pero antes de asumir estas conclusiones, es necesario probar que los coeficientes obtenidos son estadísticamente no nulos con una prueba de hipótesis.

### 5.3.1. Prueba de Hipótesis

Al igual que el coeficiente de correlación de Spearman, el coeficiente de correlación de Kendall tiene asociada una prueba de hipótesis bilateral cuya hipótesis nula es la ausencia de asociación entre las variables.

- Hipótesis nula ( $H_0$ ): No existe correlación entre las variables  $X$  e  $Y$ ,  $\tau = 0$ .
- Hipótesis alternativa ( $H_a$ ): Existe correlación entre las variables  $X$  e  $Y$ , es decir, el coeficiente  $\tau$  es significativamente distinto de cero,  $\tau \neq 0$ .

El test correspondiente a la versión  $\tau_B$  de la  $\tau$  de Kendall que explica los lazos, tiene el estadístico de prueba  $z_B$ .

$$z_B = \frac{S_c - S_d}{\sqrt{v}} \tag{5.5}$$

donde  $v = \frac{v_0 - v_t - v_u}{18 + v_1 + v_2}$ ,

$$v_0 = n(n - 1)(2n + 5),$$

$$v_t = \sum_i t_i(t_i - 1)(2t_i + 5),$$

$$v_u = \sum_j u_j(u_j - 1)(2u_j + 5),$$

$$v_1 = \frac{\sum_i t_i(t_i - 1) \sum_j u_j(u_j - 1)}{2n(n - 1)},$$

$$v_2 = \frac{\sum_i t_i(t_i - 1)(t_i - 2) \sum_j u_j(u_j - 1)(u_j - 2)}{9n(n - 1)(n - 2)}$$

El estadístico  $z_B$  tiene la misma distribución que  $\tau_B$  y es aproximadamente igual a una distribución normal estandar cuando los dos conjuntos considerados son estadísticamente independientes.

Al realizar esta prueba todos los p-valores dieron menos que 0,0001 indicando que los coeficientes obtenidos y presentados en el Cuadro 5.2 son estadísticamente no nulos.

Tanto el  $\tau$  de Kendall como el  $\rho$  de Spearman pueden ser formulados como casos especiales de un coeficiente de correlación entre dos variables más general denominado Rank Correlation.

Los resultados obtenidos hasta el momento son para pares de variables, el MOS y una métrica. Un segundo objetivo es hacer un análisis multivariado de la correlación entre todas las variables; es decir, el MOS y todas las métricas implementadas. Para eso necesitamos generalizar el concepto de correlación y asimilarlo al de concordancia, que será explicado en mayor detalle en el próximo capítulo. Por ahora presentamos la siguiente prueba de correlación entre múltiples variables.

## 5.4. El coeficiente $w$ de Kendall

Este coeficiente presentado por Kendall et al. [71] es una medida del grado de acuerdo o concordancia entre  $m$  conjuntos de  $n$  rangos. Por ejemplo, para un grupo de  $n$  objetos evaluados por  $m$  jueces, el coeficiente provee información sobre el grado de acuerdo entre los  $n$  rangos otorgados por los jueces. Además puede probarse que este coeficiente está relacionado linealmente con el valor medio de los coeficientes de correlación de Spearman entre todos los pares de las clasificaciones sobre las que se calcula.

La fórmula para calcularlo es:

$$w = \frac{12S}{m^2(n^3 - n) - m \sum_{j=1}^m L_j} \quad (5.6)$$

donde,  $w$  es el coeficiente de concordancia de Kendall,  $n$  es el tamaño de la muestra en función del número de tripletes, tetrapletes, quintupletes, etc.,  $m$  es el número de variables incluidas. En nuestro caso  $n = 3000$  tuplas de  $m = 14$  elementos, el valor del MOS y de cada métrica para una de las observaciones.

Supongamos que el evaluador  $j$  otorga el rango  $r_{ij}$  al objeto  $i$ , entonces el rango total otorgado al objeto  $i$  es:

$$R_i = \sum_{j=1}^m r_{ij} \quad (5.7)$$

El valor medio de estos rangos totales es:

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i \quad (5.8)$$

$S$ , la suma de los cuadrados de las diferencias observadas con respecto a un promedio, es definido como:

$$S = \sum_{i=1}^n (R_i - \bar{R})^2 \quad (5.9)$$

$L_j$  es la sumatoria de los lazos o empates entre los rangos y está definida como:

$$L_j = \sum_{k=1}^{g_j} (t_k^3 - t_k) \quad (5.10)$$

Donde  $g_j$  es el número de grupos de rangos repetidos según el ranking realizado por el evaluador  $j$  y  $t_k$  es el número de rangos repetidos en el  $k$ -ésimo grupo de lazos.

Entonces  $L_j$  es el factor de corrección para el conjunto de rangos del juez  $j$ . Notar que si para el evaluador  $j$  no hay rangos empatados, entonces  $L_j = 0$

$w$  está linealmente relacionado con el promedio de los  $\binom{m}{2}$  coeficientes de Spearman que se pueden calcular tomando de a dos variables, pero una diferencia que presenta con este coeficiente es que este último varía de  $-1$  a  $1$ , mientras que el  $w$  de Kendall siempre da un valor positivo entre  $0$  y  $1$ ; es decir, si la evaluación de cada juez a los  $n$  objetos es similar, entonces la  $w$  es igual a  $1$ , en cambio si hay un total desacuerdo, entonces el  $w = 0$ . Sin embargo, hay que tener en cuenta que un  $w = 0$  puede indicar que los atributos a evaluar son ambiguos o están pobremente definidos, consecuentemente, no se puede discriminar y por tanto hay discordancia.

Para calcular el coeficiente  $w$  entre el MOS y todas las métricas, fue necesario programarlo en Python, ya que ninguno de los paquetes disponibles de este lenguaje cuenta con una función que lo calcule, para más detalles referirse al Apéndice D donde se encuentra detallado y explicado el código generado.

Obtuvimos como resultado:

$$w = 0,6725 \quad (5.11)$$

Este es el  $w$  de Kendall entre las 14 variables analizadas.

Para interpretar este resultado nos remitimos nuevamente al Cuadro 5.1, considerando solo los intervalos con valores positivos, concluimos entonces que la correlación entre las 14 variables es fuerte.

### 5.4.1. Prueba de concordancia de Kendall

Se trata de una prueba multivariada [71], donde las hipótesis a considerar son:

- Hipótesis nula ( $H_0$ ): Entre las variables citadas no existe correlación o asociación.
- Hipótesis alternativa ( $H_a$ ): Existe una asociación entre las variables.

Esta prueba es una generalización del test de Friedman (una prueba estadística no paramétrica similar a la ANOVA paramétrica para medidas repetidas, que se usa para detectar diferencias entre tratamientos) y se basa en el coeficiente  $w$  previamente definido, que es una normalización del estadístico de Friedman entre 0 y 1.

El estadístico de prueba es:

$$c^2 = m(n - 1)w \quad (5.12)$$

donde  $n$  es el número de imágenes,  $m$  el número de métricas (incluyendo al MOS) y  $w$  el coeficiente de concordancia de Kendall.

$c^2$  tiene una distribución aproximada Chi-cuadrada con  $n - 1$  grados de libertad.

Con la prueba de concordancia de Kendall buscamos descubrir si existe asociación entre todas las variables: el MOS y las métricas, es decir, si el coeficiente  $w$  obtenido en (5.11) es significativamente distinto de 0.

Realizamos esta prueba y obtuvimos un p-valor menor que 0,0001, por lo que la hipótesis nula se rechaza y concluimos que existe asociación entre las 14 variables analizadas.

También es posible calcular el coeficiente  $w$  para subgrupos de variables, por ejemplo para el MOS y NQM, MSSIM, GSM, GMSD, que son las métricas que obtuvieron los coeficientes de Spearman y de Kendall más altos. El  $w$  entre el MOS y estas métricas es:

$$w = 0,8125 \quad (5.13)$$

y el p-valor de la prueba correspondiente es menor que 0,0001. Concluimos entonces, que el acuerdo mejora si consideramos este grupo reducido de variables.

---

# Capítulo 6

## Concordancia

En estadística, la confiabilidad entre evaluadores o la concordancia es el grado de acuerdo entre los evaluadores. A continuación se presentan una serie de conceptos y técnicas que servirán para evaluar el grado de concordancia entre las métricas estudiadas y el MOS. Para poder aplicar estos conceptos a los datos, fue necesario transformar las variables continuas en variables categóricas, como se explicará más adelante.

### 6.1. Matriz de confusión

Una matriz de confusión (Kohavi y Provost, 1998) contiene información sobre las clasificaciones reales y predichas hechas por un sistema de clasificación. El rendimiento de estos sistemas se evalúa comúnmente utilizando los datos de la matriz.

A modo de ejemplo, la Figura 6.1 muestra la matriz de confusión para un clasificador de dos clases para un test que tiene dos resultados posibles: positivo (P) y negativo (N) [72]. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real.

		<i>P</i>	<i>N</i>
Clase Real	<i>P</i>	Verdadero Positivo	Falso Negativo
	<i>N</i>	Falso Positivo	Verdadero Negativo
		Clase Predicha	

Figura 6.1: Ejemplo matriz de confusión para dos clases

La matriz de confusión normalizada es la matriz que se obtiene al dividir cada elemento por

el total que suma la fila en que se encuentra, de esta forma se obtiene sobre el total real en cada categoría, la proporción que fue clasificada en cada categoría por el predictor.

Para poder aplicar este concepto a los datos de interés fue necesario clasificar los valores del MOS de cada imagen de la base TID2013 en tres categorías: “calidad baja”, “calidad media” y “calidad alta”. El criterio utilizado fue el que se presenta en [8] : valores en el intervalo  $[0; 3,94)$  se consideraron de “calidad baja”, valores en el intervalo  $[3,94118; 5,25000)$  de “calidad media” y en el intervalo  $[5,25000; 9,00000)$  de “calidad alta”, como se ilustra en a Figura 6.2. De esta forma el MOS clasifica 999 imágenes como de “calidad baja”, 1000 imágenes como de “calidad media” y 1001 imágenes como de “calidad alta”. Esta clasificación basada en el MOS es la clasificación real, mientras que cada métrica será considerada un predictor.

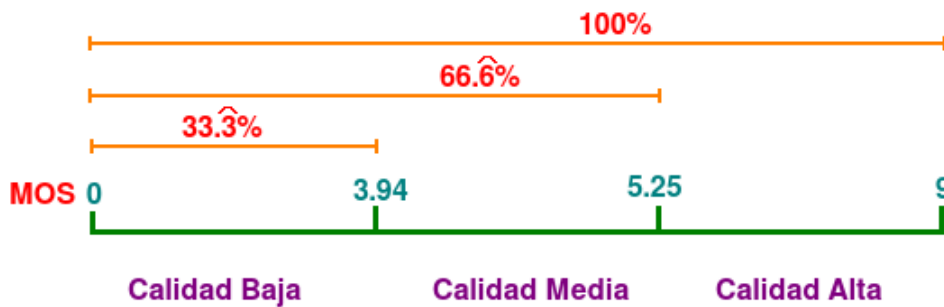


Figura 6.2: Categorización del MOS según terciles

El criterio para establecer en cuál de estas tres categorías clasifica cada métrica a una imagen según el valor obtenido, se definió siguiendo la misma lógica, se eligieron como extremos del intervalo que define la categoría “calidad media”, los percentiles  $\frac{1}{3}$  y  $\frac{2}{3}$ , de esta manera, aproximadamente 1000 imágenes fueron clasificadas como “calidad baja”, 1000 como “calidad media” y otras 1000 como “calidad alta” por cada métrica.

Una vez definida esta clasificación, se obtuvieron las matrices de confusión, y su versión normalizada, para cada métrica vs el MOS, las mismas pueden apreciarse en las Figuras 6.3 - 6.13. En el Apéndice E se encuentran las matrices de confusión para cada grupo de distorsiones presentado en el Capítulo 3.



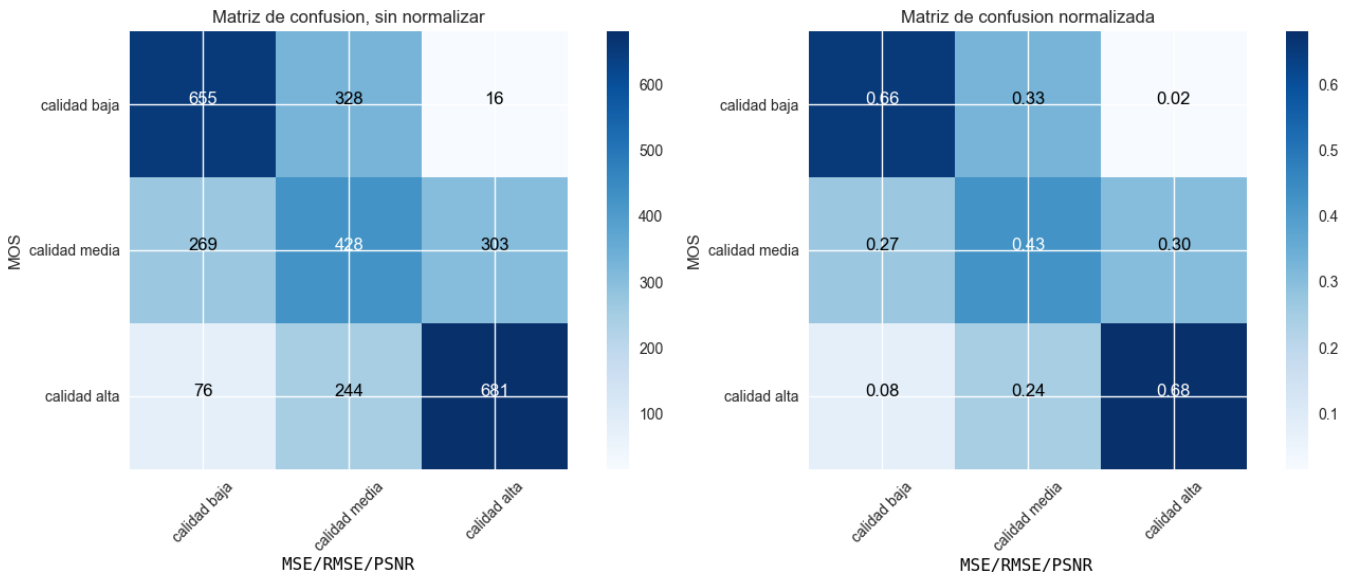


Figura 6.3: A la izquierda, matriz de confusión para las métricas MSE, RMSE y PSNR (que es la misma). A la derecha, la matriz de confusión normalizada. La escala de azules ilustra la densidad de cada casillero; cuanto más oscura sea la diagonal y más claro el resto de los casilleros, mejor será el desempeño de la métrica. Observando la matriz normalizada se puede concluir que la métrica MSE predice correctamente “calidad alta” el 68% de las veces, “calidad media” el 43% de las veces y “calidad baja” el 66% de las veces. El 33% de las imágenes de “calidad baja” son clasificadas como “calidad media”, el 27% de las imágenes de “calidad media” son clasificadas como “calidad baja”, mientras que el 30% de las imágenes de “calidad media” son clasificadas como “calidad alta” y el 24% de las imágenes de “calidad alta” fueron clasificadas como de “calidad media”.

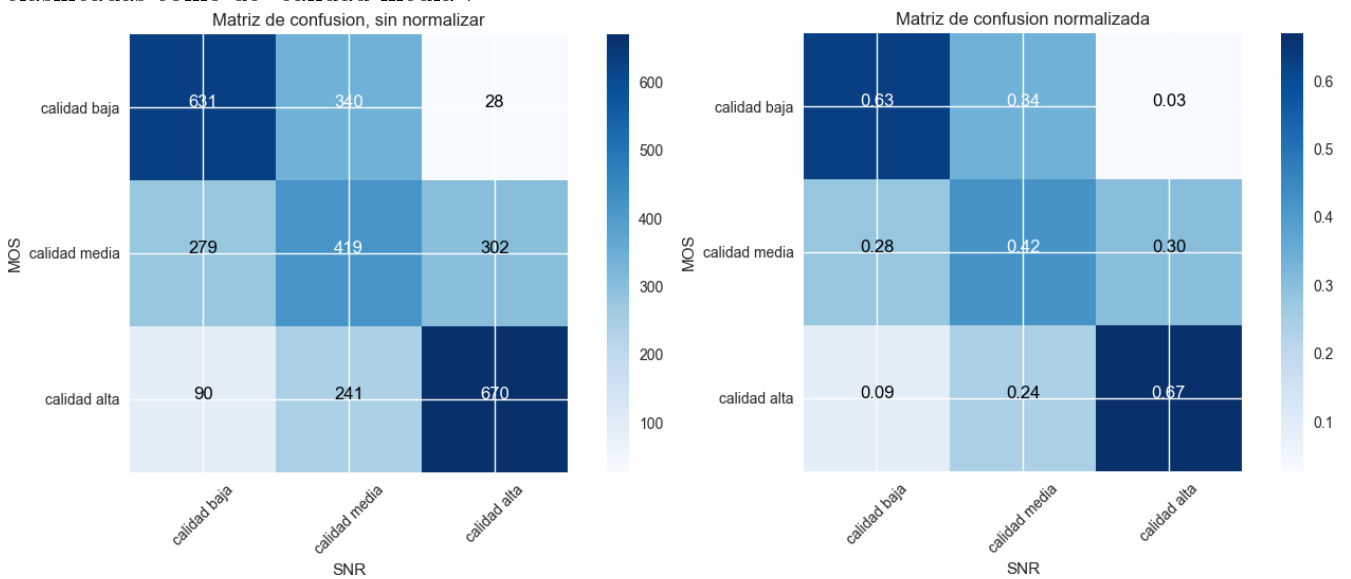


Figura 6.4: A la izquierda, matriz de confusión para la métrica SNR. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 67% de las imágenes de “calidad alta”, el 42% de las imágenes de “calidad media” y el 63% de las imágenes de “calidad baja”, son clasificadas correctamente. Los resultados son muy similares a las métricas MSE, RMSE y PSNR.

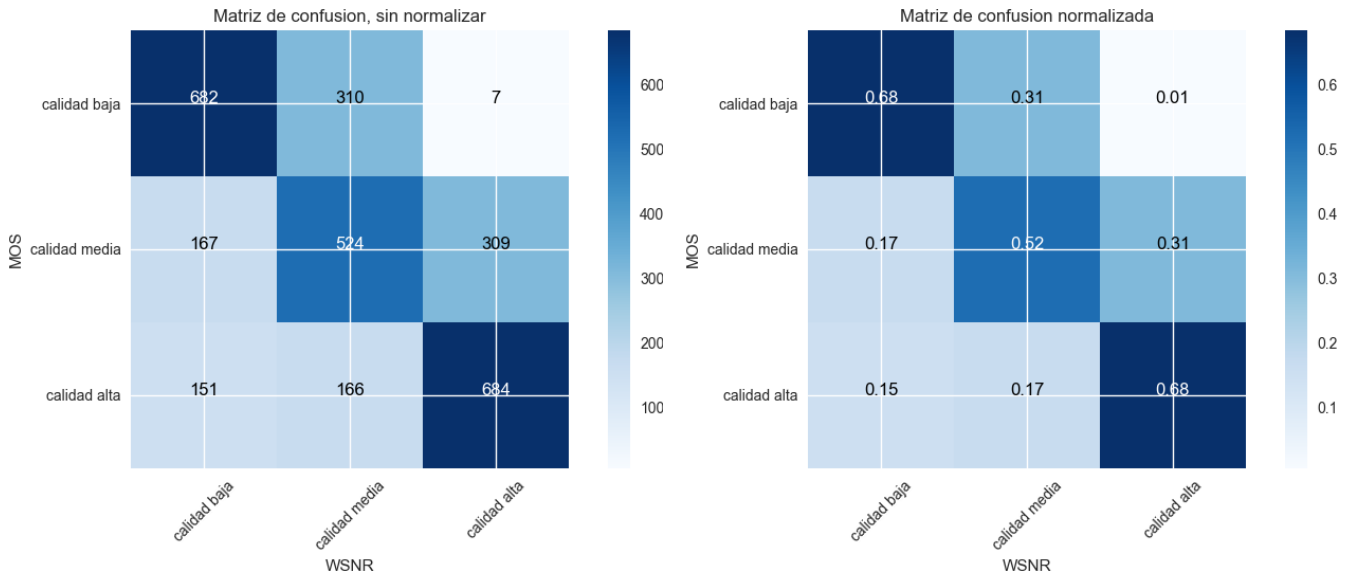


Figura 6.5: A la izquierda, matriz de confusión para la métrica WSNR. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 68 % de las imágenes de “calidad alta”, el 52 % de las imágenes de “calidad media” y el 68 % de las imágenes de “calidad baja”, son clasificadas correctamente. El 31 % de las imágenes de “calidad baja” son clasificadas como de “calidad media”, mientras que el 31 % de las imágenes de “calidad media” son clasificadas como “calidad alta”.

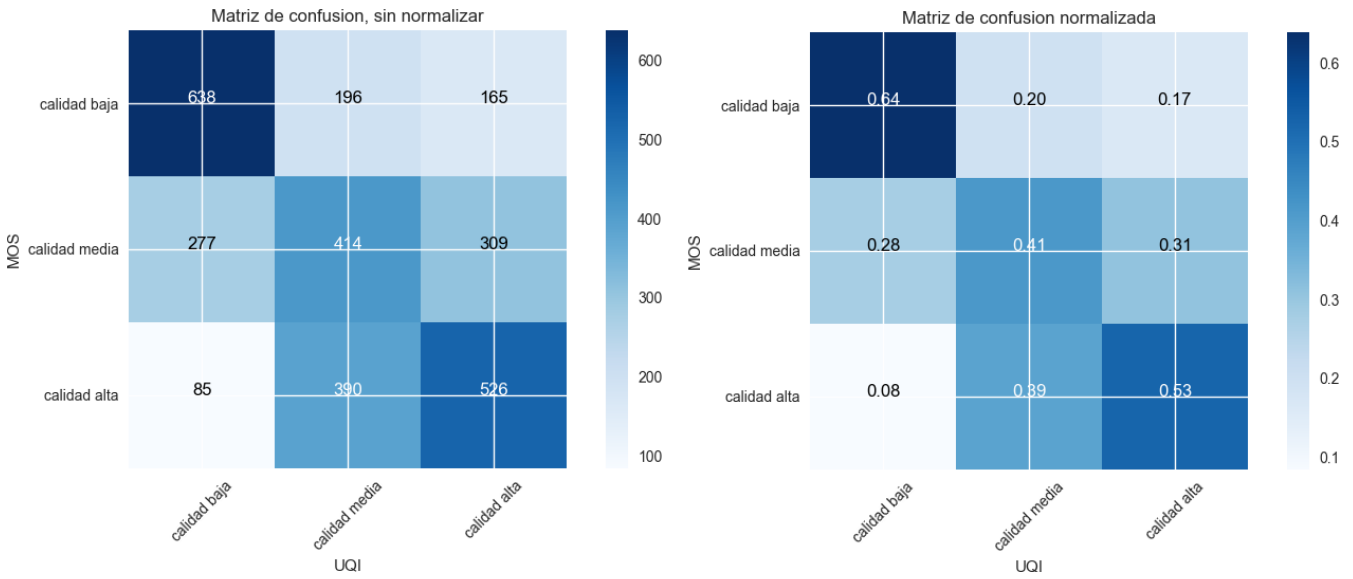


Figura 6.6: A la izquierda, matriz de confusión para la métrica UQI. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 53 % de las imágenes de “calidad alta”, el 41 % de las imágenes de “calidad media” y el 64 % de las imágenes de “calidad baja”, son clasificadas correctamente. El 39 % de las imágenes de “calidad alta” son clasificadas como de “calidad media”.

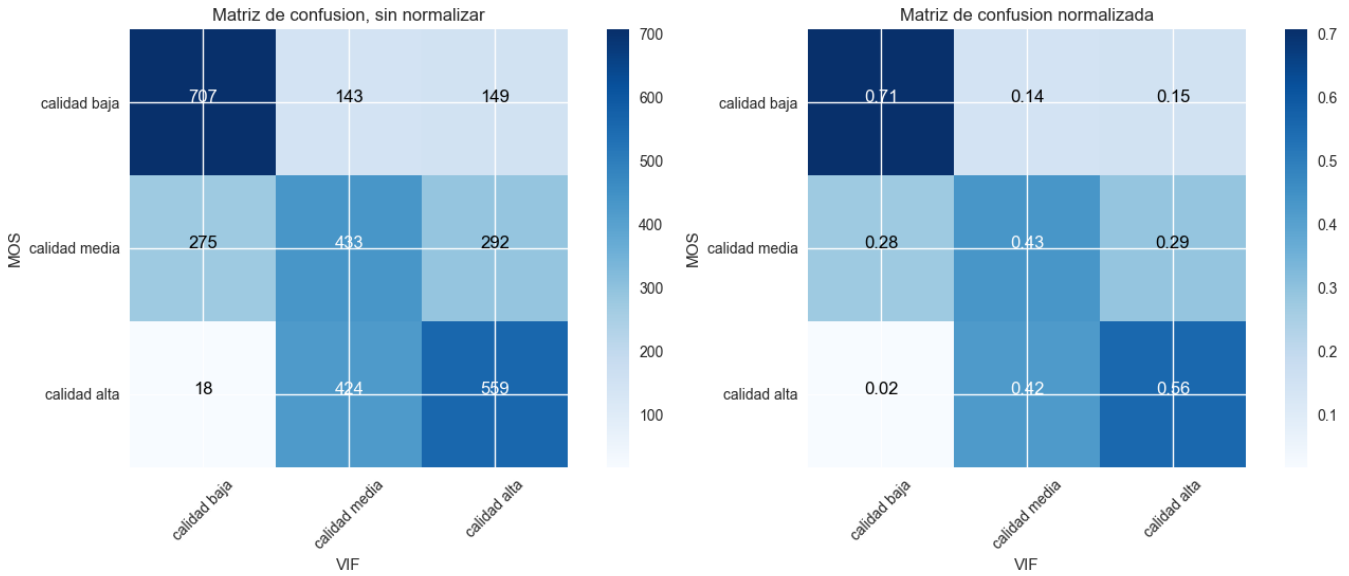


Figura 6.7: A la izquierda, matriz de confusión para la métrica VIF. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 56 % de las imágenes de “calidad alta”, el 43 % de las imágenes de “calidad media” y el 71 % de las imágenes de “calidad baja”, son clasificadas correctamente.

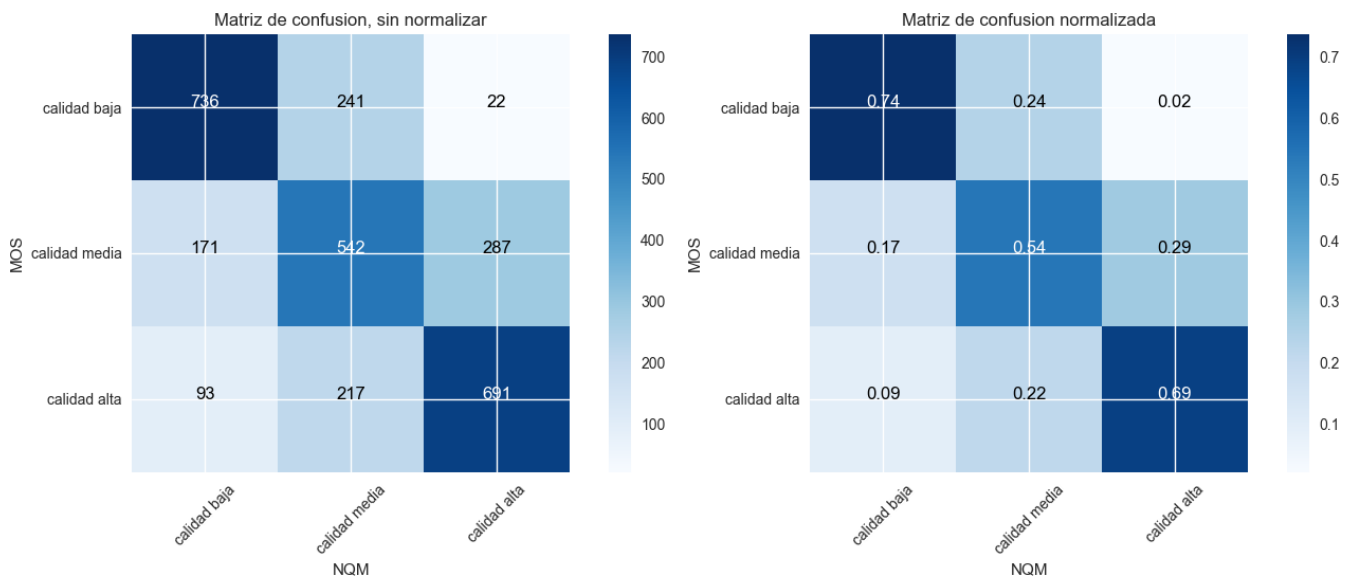


Figura 6.8: A la izquierda, matriz de confusión para la métrica NQM. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 69 % de las imágenes de “calidad alta”, el 54 % de las imágenes de “calidad media” y el 74 % de las imágenes de “calidad baja”, son clasificadas correctamente.

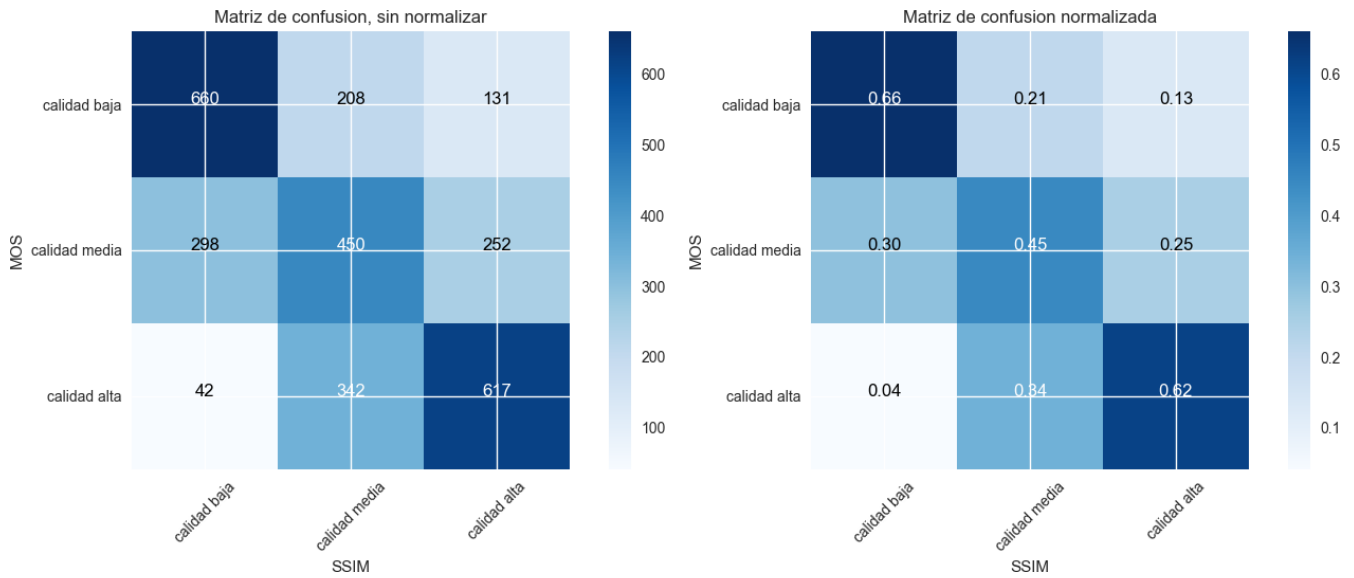


Figura 6.9: A la izquierda, matriz de confusión para la métrica SSIM. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 62 % de las imágenes de “calidad alta”, el 45 % de las imágenes de “calidad media” y el 66 % de las imágenes de “calidad baja”, son clasificadas correctamente.

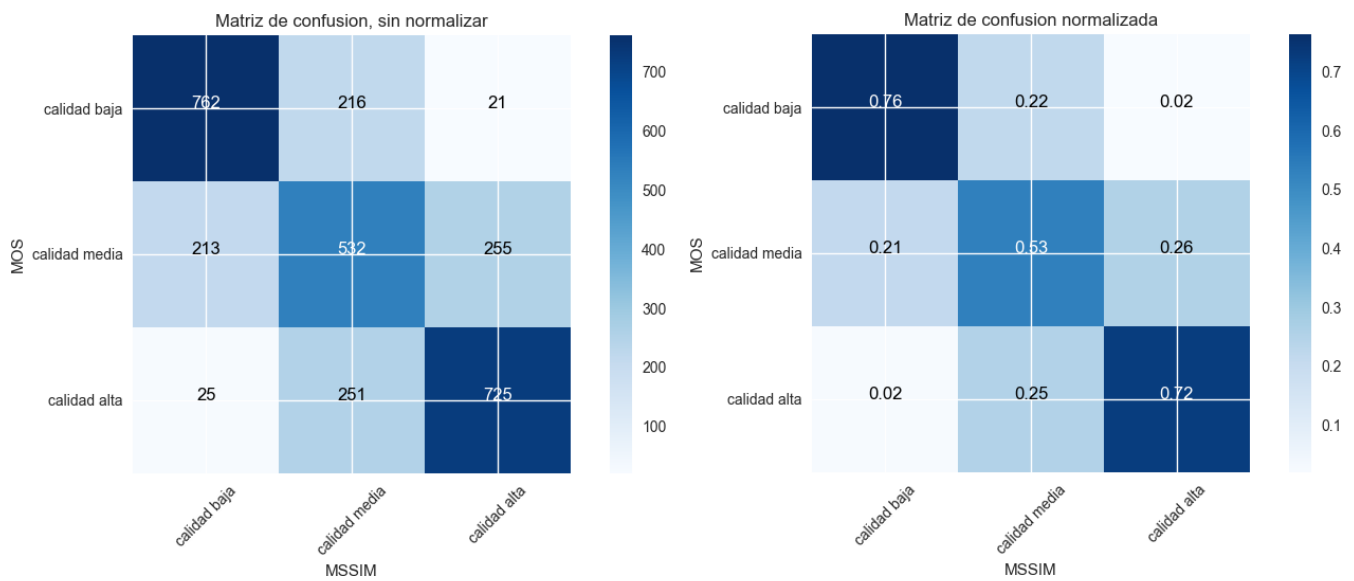


Figura 6.10: A la izquierda, matriz de confusión para la métrica MSSIM. A la derecha, la matriz de confusión normalizada. De la matriz normalizada se observa que el 72 % de las imágenes de “calidad alta”, el 53 % de las imágenes de “calidad media” y el 76 % de las imágenes de “calidad baja”, son clasificadas correctamente.

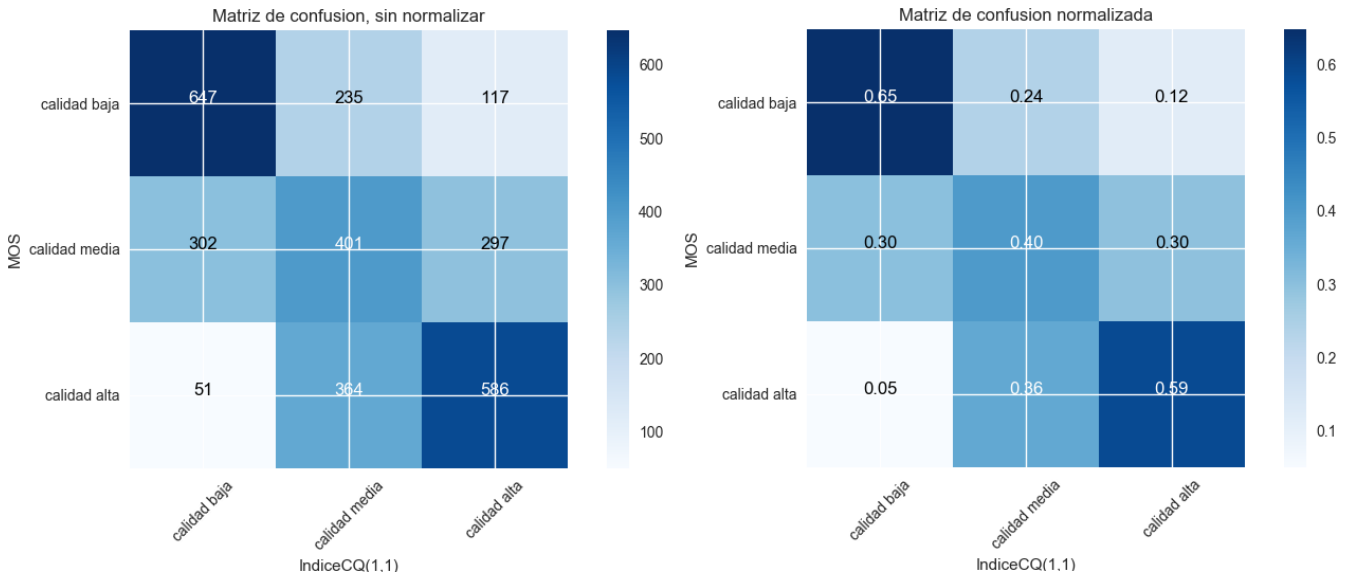


Figura 6.11: Izquierda: matriz de confusión para el Índice CQ(1,1). Derecha: matriz de confusión normalizada. Puede observarse que el 59 % de las imágenes de “calidad alta”, el 40 % de las imágenes de “calidad media” y el 65 % de las imágenes de “calidad baja”, son clasificadas correctamente.

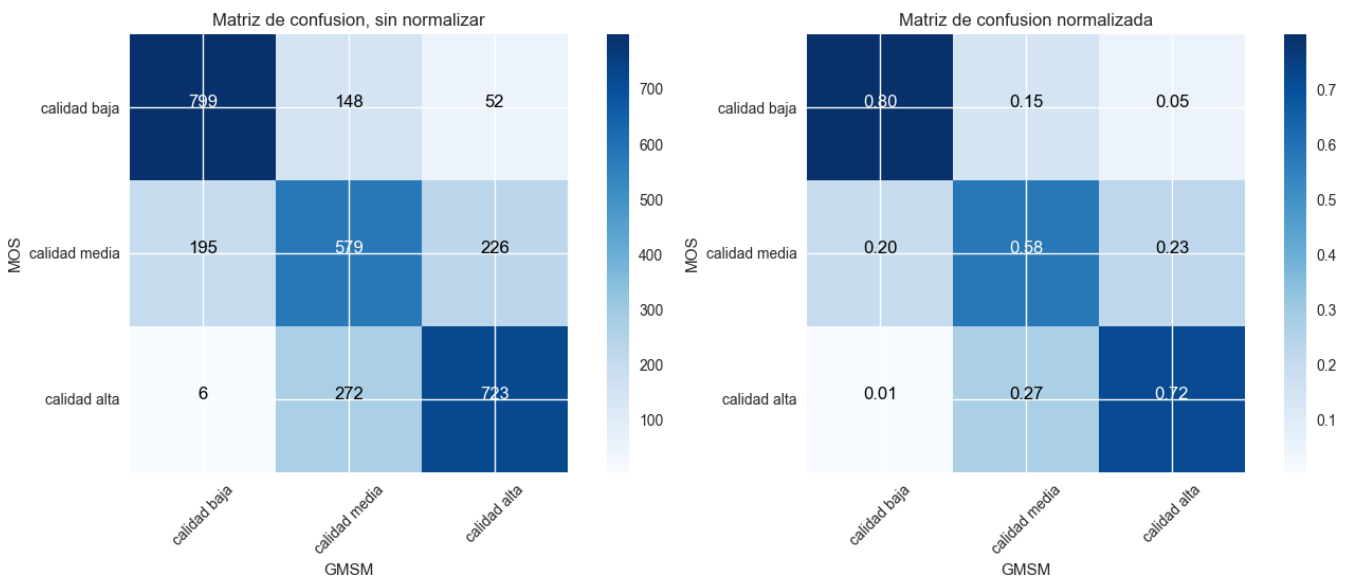


Figura 6.12: Izquierda: matriz de confusión para la métrica GSM. Derecha: matriz de confusión normalizada. Puede observarse que el 72 % de las imágenes de “calidad alta”, el 58 % de las imágenes de “calidad media” y el 80 % de las imágenes de “calidad baja”, son clasificadas correctamente.

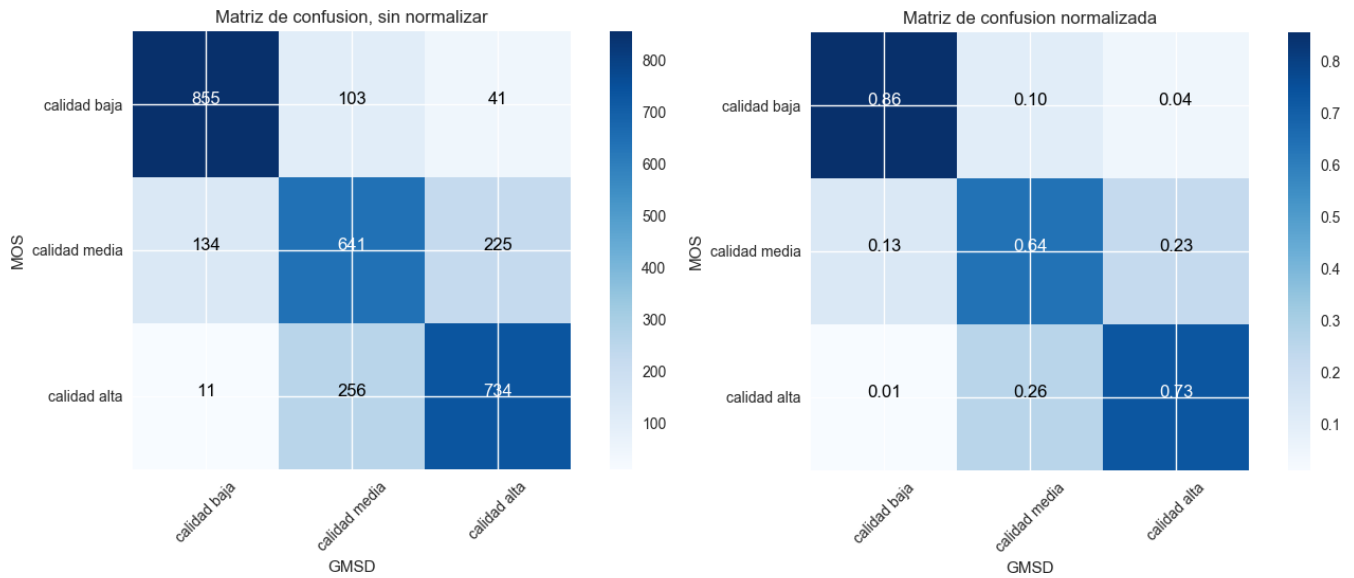


Figura 6.13: Izquierda: matriz de confusión para la métrica GMSD. Derecha: matriz de confusión normalizada. Puede observarse que el 73 % de las imágenes de “calidad alta”, el 64 % de las imágenes de “calidad media” y el 86 % de las imágenes de “calidad baja”, son clasificadas correctamente.

En el Cuadro 6.1 se presentan los porcentajes de imágenes de calidad baja, media y alta que fueron clasificadas correctamente según cada métrica, estos valores se obtuvieron de las diagonales de las matrices de confusión normalizadas.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	66 %	43 %	68 %
SNR	63 %	42 %	67 %
WSNR	68 %	52 %	68 %
UQI	64 %	41 %	53 %
VIF	71 %	43 %	56 %
NQM	74 %	54 %	69 %
SSIM	66 %	45 %	62 %
MSSIM	76 %	53 %	72 %
CQ(1,1)	65 %	40 %	59 %
GMSM	80 %	58 %	72 %
GMSD	86 %	64 %	73 %

Cuadro 6.1: Porcentaje de imágenes bien clasificadas por métrica, según calidad.

Puede observarse que las métricas GMSD y GMSM clasifican correctamente una proporción alta de imágenes de calidad alta y baja, y esta proporción baja considerablemente cuando se trata de imágenes de calidad media, aunque siguen siendo los porcentajes más altos de imágenes de calidad media bien clasificadas entre todas las métricas. En general, la columna de calidad media es la que presenta porcentajes más bajos de clasificaciones correctas. Las métricas MSE y SNR funcionan mejor cuando se aplican a imágenes de calidad alta, la WSNR clasifica correctamente la

misma proporción de imágenes de calidad baja y alta, mientras que el resto tiene mejor rendimiento en imágenes de calidad baja.

Ordenando estos porcentajes de mayor a menor, podemos establecer cuáles métricas clasifican mejor según la calidad de la imagen, los resultados se muestran en el Cuadro 6.2.

calidad baja	calidad media	calidad alta
GMSD	GMSD	GMSD
GMSM	GMSM	GMSM & MSSIM
MSSIM	NQM	NQM
NQM	MSSIM	MSE & WSNR
VIF	WSNR	SNR

Cuadro 6.2: Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

Las métricas que clasificaron mejor fueron, primera la GMSD y segunda la GMSM. Siguen luego la MSSIM y la NQM. Esto coincide con lo obtenido en el capítulo anterior. Sin embargo en este análisis no estamos teniendo en cuenta qué ocurre con las clasificaciones incorrectas, por ejemplo, no sería deseable que una métrica clasifique una proporción grande de imágenes de alta calidad como de baja calidad o viceversa.

## 6.2. Coeficiente kappa de Cohen

El coeficiente kappa de Cohen [73] mide el acuerdo entre las correspondientes clasificaciones de dos evaluadores que clasifican  $n$  elementos en  $c$  categorías mutuamente excluyentes. Esta medida estadística ajusta el efecto del azar en la proporción de la concordancia observada para elementos cualitativos (variables categóricas). En general se cree que es una medida más robusta que el simple cálculo del porcentaje de concordancia, ya que tiene en cuenta el acuerdo que ocurre por casualidad [67].

La ecuación que lo define es la siguiente:

$$\kappa = \frac{P_a - P_e}{1 - P_e} \tag{6.1}$$

O la expresión equivalente:

$$\kappa = 1 - \frac{1 - P_a}{1 - P_e} \tag{6.2}$$

donde  $P_a$  es el acuerdo observado relativo entre los evaluadores, y  $P_e$  es la probabilidad hipotética de acuerdo por azar, utilizando los datos observados para calcular las probabilidades de que cada observador clasifique aleatoriamente cada categoría.

$1 - P_e$  se interpreta entonces como el grado de acuerdo que es alcanzable por encima del azar, mientras que  $P_a - P_e$  es el grado de acuerdo realmente alcanzado por encima del azar.

En general, para  $c$  categorías y  $n$  elementos,

$$P_e = \frac{1}{n^2} \sum_c n_{c1}n_{c2} \quad (6.3)$$

donde  $n_{cj}$  es la cantidad de veces que el evaluador  $j$  clasifica en la categoría  $c$ . Este índice solo mide el acuerdo entre dos clasificadores. A modo de ejemplo, y para que se entienda mejor el concepto, se presenta a continuación el cálculo paso a paso del índice kappa de Cohen entre el MOS y la métrica GMSD.

Para calcular el porcentaje de acuerdo observado ( $P_a$ ) debemos sumar los valores en la diagonal de la matriz de confusión (Figura 6.13) y dividir por el total:

$$P_a = \frac{855 + 641 + 734}{3000} = \frac{2230}{3000} = 0,7433 \quad (6.4)$$

Para poder calcular la probabilidad de acuerdo por azar ( $P_e$ ) es preciso observar que:

1. el MOS clasifica como “calidad baja” 999 elementos, mientras que el GMSD 1000, por lo tanto la probabilidad de que ambos clasifiquen como “calidad baja” por azar es:

$$\frac{999}{3000} \times \frac{1000}{3000} = \frac{999 \times 1000}{3000^2} = 0,111 \quad (6.5)$$

2. equivalentemente, la probabilidad de que el GMSD y el MOS, clasifiquen como “calidad media” por azar es:

$$\frac{1000 \times 1000}{3000^2} = 0,1111 \quad (6.6)$$

3. la probabilidad de que ambos clasifiquen como “calidad alta” por azar es:

$$\frac{1001 \times 1000}{3000^2} = 0,1112 \quad (6.7)$$

Por lo tanto la probabilidad de acuerdo por azar es:

$$P_e = 0,111 + 0,1111 + 0,1112 = 0,3333 \quad (6.8)$$

Aplicando la fórmula (6.1), el índice kappa entre el MOS y GMSD es:

$$\kappa = \frac{P_a - P_e}{1 - P_e} = \frac{0,7433 - 0,3333}{1 - 0,3333} = 0,615 \quad (6.9)$$

La forma de interpretar el índice es la siguiente, si los evaluadores están completamente de acuerdo, entonces  $\kappa = 1$ , si no hay acuerdo entre los calificadores distinto al que cabría esperar



por azar, entonces  $\kappa = 0$ . Las interpretaciones intermedias se encuentran detalladas en el Cuadro 6.3 [68] <sup>1</sup>.

$\kappa$	Interpretación
$\leq 0$	Poco acuerdo
$(0; 0,2]$	Leve acuerdo
$(0,2; 0,4]$	Acuerdo justo
$(0,4; 0,6]$	Acuerdo moderado
$(0,6; 0,8]$	Acuerdo substancial
$(0,8; 1]$	Acuerdo casi perfecto

Cuadro 6.3: Interpretación del coeficiente

Podemos concluir entonces que entre la métrica GMSD y el MOS existe un acuerdo substancial.

Los índices kappa de Cohen entre el MOS y el resto de las métricas fueron calculados utilizando el paquete *nltk* de Python <sup>2</sup>, los mismos se muestran en el Cuadro 6.4 completando el Cuadro 5.2 del capítulo anterior. Los códigos desarrollados para calcular estos índices se encuentran en el Apéndice F.

Métrica	Corr. Kendall	Corr. Spearman	Kappa Cohen / Pi Scott	Acuerdo con el MOS
MSE & RMSE	-0,4958	-0,6869	0,3820	Justo
PSNR	0,4958	0,6869	0,3820	Justo
SNR	0,4607	0,6491	0,3600	Justo
WSNR	0,4938	0,6382	0,4450	Moderado
UQI	0,3695	0,5239	0,2890	Justo
VIF	0,4669	0,6338	0,3495	Justo
NQM	0,5348	0,7126	0,4845	Moderado
SSIM	0,4457	0,6273	0,3635	Justo
MSSIM	0,5921	0,7909	0,5095	Moderado
CQ(1,1)	0,4291	0,6008	0,3170	Justo
GMSM	0,6132	0,7884	0,5505	Moderado
GMSD	-0,6339	-0,8044	0,6150	Substancial

Cuadro 6.4: Todas las medidas son con respecto al MOS.

### 6.3. Coeficiente pi de Scott

Presentado por William Scott [74], es un estadístico similar al kappa de Cohen, solo difiere de este en el cálculo de la probabilidad de acuerdo por azar  $P_e$ .

El coeficiente pi de Scott hace la suposición de que los evaluadores tienen la misma distribución de respuestas, lo que hace que el kappa de Cohen sea un poco más informativo.

<sup>1</sup>Ver nota al pie n° 2 del capítulo anterior.

<sup>2</sup>O equivalentemente *sklearn*, se obtienen los mismos resultados.

$$\pi = \frac{P_a - P_e}{1 - P_e} \tag{6.10}$$

$P_e$  es calculado utilizando proporciones conjuntas. La fórmula general es:

$$P_e = \sum_c \left( \frac{n_{c1} + n_{c2}}{2n} \right)^2 \tag{6.11}$$

Se entenderá mejor con un ejemplo. Calculemos el pi de Scott entre el MOS y la métrica GMSD.

Para calcular las proporciones conjuntas se suma la cantidad que clasificó en cada categoría cada evaluador y se divide por el total de clasificaciones hechas por ambos (en este caso 6000, 3000 cada uno). Luego se toma el cuadrado de cada proporción conjunta, la suma de estos es la probabilidad de acuerdo por azar  $P_e$ . Todos estos cálculos están detallados en el Cuadro 6.5.

Clasificación	MOS	GMSD	Proporción Conjunta	Proporción Conjunta al cuadrado
Calidad baja	999	1000	$\frac{1999}{6000} = 0,3332$	0,111
Calidad media	1000	1000	$\frac{2000}{6000} = 0,3333$	0,1111
Calidad alta	1001	1000	$\frac{2001}{6000} = 0,3335$	0,1112
Total	3000	3000		0,3333

Cuadro 6.5: Tabla para calcular las proporciones conjuntas y la probabilidad de acuerdo por azar

En este caso, si bien el cálculo de  $P_e$  se hace de otra forma, obtenemos el mismo resultado que antes, por lo tanto los coeficientes kappa de Cohen y pi de Scott entre el MOS y GMSD nos dan el mismo valor. Lo mismo ocurrió al calcular el coeficiente pi entre el MOS y las otras métricas si redondeamos a cuatro dígitos decimales. Esto se debe a que tanto  $n_{c1}$  como  $n_{c2}$  son iguales, o casi iguales, a 1000 en todos los casos, por lo tanto

$$\left( \frac{n_{c1} + n_{c2}}{2} \right)^2 \approx \left( \frac{2n_{c1}}{2} \right)^2 = (n_{c1})^2 \approx n_{c1}n_{c2}$$

y entonces las fórmulas (6.3) y (6.11) son equivalentes.

A pesar de que, en este caso, a los fines prácticos no hay diferencias entre un coeficiente y el otro, fue necesario introducir este nuevo concepto porque el pi de Scott se extiende a más de dos evaluadores en forma de kappa de Fleiss, y estamos interesadas en obtener un índice de concordancia multivariado entre todas las métricas o un subconjunto de ellas.

## 6.4. Coeficiente kappa de Fleiss

Esta generalización del coeficiente pi de Scott [75], es una medida estadística para evaluar la confiabilidad del acuerdo entre un número fijo de evaluadores al asignar calificaciones categóricas a un número de ítems a clasificar. La medida calcula el grado de acuerdo en la clasificación sobre lo

que se esperaría por casualidad o azar, dando una puntuación de cuánta homogeneidad o consenso hay en las calificaciones otorgadas por los evaluadores.

Se puede interpretar que expresa la medida en que la cantidad observada de acuerdo entre los evaluadores excede lo que se esperaría si todos los evaluadores hicieran sus calificaciones completamente al azar.

Es útil para refinar las herramientas de evaluación, por ejemplo, determinando si una escala particular es apropiada para medir una variable particular, si varios evaluadores no están de acuerdo, o bien la escala es defectuosa o los evaluadores necesitan ser nuevamente entrenados.

Este coeficiente por lo general es utilizado cuando las categorías de clasificación son nominales, aunque claramente puede aplicarse a categorías ordinales también.

En nuestro caso, el acuerdo se puede pensar de la siguiente manera: si un número fijo de métricas (las 13 métricas más el MOS) asignan calificaciones numéricas (1: calidad baja, 2: calidad media, 3: calidad alta) a un número fijo de imágenes (las 3000 imágenes de la base TID2013), entonces el kappa dará una medida de la consistencia de estas calificaciones.

### 6.4.1. Ecuaciones

El coeficiente es definido [67], al igual que los anteriores, por la ecuación:

$$\kappa = \frac{P_a - P_e}{1 - P_e} \quad (6.12)$$

sólo que esta vez  $P_e$  y  $P_a$  serán calculados teniendo en cuenta la relación entre todos los pares de evaluadores posibles.

La forma de interpretar el índice es la misma que antes; si los evaluadores están en completo acuerdo, entonces  $\kappa = 1$ ; si no hay acuerdo más que el que podría esperarse por azar,  $\kappa \leq 0$ . Para más detalles de cómo interpretar este coeficiente referirse nuevamente a el Cuadro 6.3.

Veremos a continuación cómo definir  $P_e$  y  $P_a$ .

Sea  $n$  el número total de ítems a clasificar, sea  $m$  el número de valoraciones por ítem, y sea  $c$  el número de categorías que pueden asignarse. Los ítems son indexados por  $i = 1, \dots, n$  y las categorías son indexadas por  $k = 1, \dots, c$ . Sea  $n_{ik}$  el número de evaluadores que al  $i$ -ésimo ítem le asignaron la  $k$ -ésima categoría, claramente  $0 \leq n_{ik} \leq 14$ .

En nuestro caso los “ítems” son las imágenes, por lo tanto  $n = 3000$ , cada imagen tiene  $m = 14$  valoraciones, la clasificación hecha por el MOS y la clasificación hecha por el resto de las 13 métricas.

Las categorías son: “calidad baja (1)”, “calidad media (2)” y “calidad alta (3)”; por lo tanto  $c = 3$ .

Primero se calcula  $p_k$ , la proporción de todas las asignaciones que fueron a la  $k$ -ésima categoría:

$$p_k = \frac{1}{nm} \sum_{i=1}^n n_{ik} \quad (6.13)$$

Por definición:

$$1 = \sum_{k=1}^c p_k \quad (6.14)$$

Esto se desprende del hecho que para todo  $i$ :

$$\sum_{k=1}^c n_{ik} = m \quad (6.15)$$

luego

$$\sum_{k=1}^c p_k = \sum_{k=1}^c \frac{1}{nm} \sum_{i=1}^n n_{ik} \quad (6.16)$$

$$= \frac{1}{nm} \sum_{i=1}^n \sum_{k=1}^c n_{ik} \quad (6.17)$$

$$= \frac{1}{nm} \sum_{i=1}^n m \quad (6.18)$$

$$= \frac{1}{nm} nm = 1 \quad (6.19)$$

$P_e$  queda definido de la siguiente manera:

$$P_e = \sum_{k=1}^c p_k^2 \quad (6.20)$$

Definimos  $P_i$  como la medida en que los evaluadores están de acuerdo para el  $i$ -ésimo sujeto, esto vendría a ser cuántos pares de evaluador-evaluador están de acuerdo, en relación con el número de todos los posibles pares de evaluador-evaluador.

$$P_i = \frac{1}{m(m-1)} \sum_{k=1}^c n_{ik} (n_{ik} - 1) \quad (6.21)$$

$$= \frac{1}{m(m-1)} \sum_{k=1}^c n_{ik}^2 - n_{ik} \quad (6.22)$$

$$= \frac{1}{m(m-1)} \left( \left( \sum_{k=1}^c n_{ik}^2 \right) - m \right) \quad (6.23)$$

$P_a$  queda definido entonces como la media de los  $P_i$

$$P_a = \frac{1}{n} \sum_{i=1}^n P_i \tag{6.24}$$

$$= \frac{1}{nm(m-1)} \left( \left( \sum_{i=1}^n \sum_{k=1}^c n_{ik}^2 \right) - nm \right) \tag{6.25}$$

En nuestro caso de estudio,  $n = 3000$ ,  $m = 14$  y  $c = 3$ , por lo tanto,

$$p_k = \frac{1}{42000} \sum_{i=1}^{3000} n_{ik} \tag{6.26}$$

más específicamente,

$$p_1 = \frac{1}{42000} \sum_{i=1}^{3000} n_{i1} \tag{6.27}$$

$$p_2 = \frac{1}{42000} \sum_{i=1}^{3000} n_{i2} \tag{6.28}$$

$$p_3 = \frac{1}{42000} \sum_{i=1}^{3000} n_{i3} \tag{6.29}$$

y

$$P_e = \sum_{k=1}^3 p_k^2 = p_1^2 + p_2^2 + p_3^2 \tag{6.30}$$

por otro lado, según la ecuación (6.25),

$$P_a = \frac{1}{546000} \left( \left( \sum_{i=1}^{3000} \sum_{k=1}^3 n_{ik}^2 \right) - 42000 \right) \tag{6.31}$$

El coeficiente kappa de Fleiss se calculó utilizando el paquete *nltk* de Python, el código correspondiente se presenta en el Apéndice F, el resultado obtenido fue:

$$\kappa = 0,4137 \tag{6.32}$$

Por lo que concluimos que el MOS junto con todas las métricas seleccionadas tienen un acuerdo moderado.

Si ahora calculamos el coeficiente kappa de Fleiss para el MOS y el subconjunto de las métricas que, en base a todos los análisis hechos hasta el momento, consideramos han tenido el mejor

desempeño, estas serían: NQM, MSSIM, GSM y GMSD, obtenemos:

$$\kappa = 0,5326 \tag{6.33}$$

El acuerdo mejora, pero sigue siendo moderado.

---

# Capítulo 7

## Análisis por grupos de distorsiones

Se presenta a continuación un análisis por grupos de distorsiones. Como se planteó en el Capítulo 3, las distorsiones se pueden dividir en grupos según sus características [8], como se indica en el Cuadro 7.1 y la Figura 7.1.

Número	Distorsión	Ruidosa	Actual	Simple	Exótica	Nueva	Color
1	Additive Gaussian noise	+	+	+	-	-	-
2	Additive noise in color comp.	+	-	-	-	-	+
3	Spatially correlated noise	+	+	-	-	-	-
4	Masked noise	+	+	-	-	-	-
5	High frequency noise	+	+	-	-	-	-
6	Impulse noise	+	+	-	-	-	-
7	Quantization noise	+	-	-	-	-	+
8	Gaussian blur	+	+	+	-	-	-
9	Image denoising	+	+	-	-	-	-
10	JPEG compression	-	+	+	-	-	+
11	JPEG2000 compression	-	+	-	-	-	-
12	JPEG transm. errors	-	-	-	+	-	-
13	JPEG2000 transm. errors	-	-	-	+	-	-
14	Non ecc. pattern noise	-	-	-	+	-	-
15	Local block-wise dist.	-	-	-	+	-	-
16	Mean shift (intensity shift)	-	-	-	+	-	-
17	Contrast change	-	-	-	+	-	-
18	Change of color saturation	-	-	-	-	+	+
19	Multipl. Gaussian noise	+	+	-	-	+	-
20	Comfort noise	-	-	-	+	+	-
21	Lossy compr. of noisy images	+	+	-	-	+	-
22	Image color quant. w. dither	-	-	-	-	+	+
23	Chromatic aberrations	-	-	-	+	+	+
24	Sparse sampl. and reconstr.	-	-	-	+	+	-

Cuadro 7.1: Grupos de distorsiones

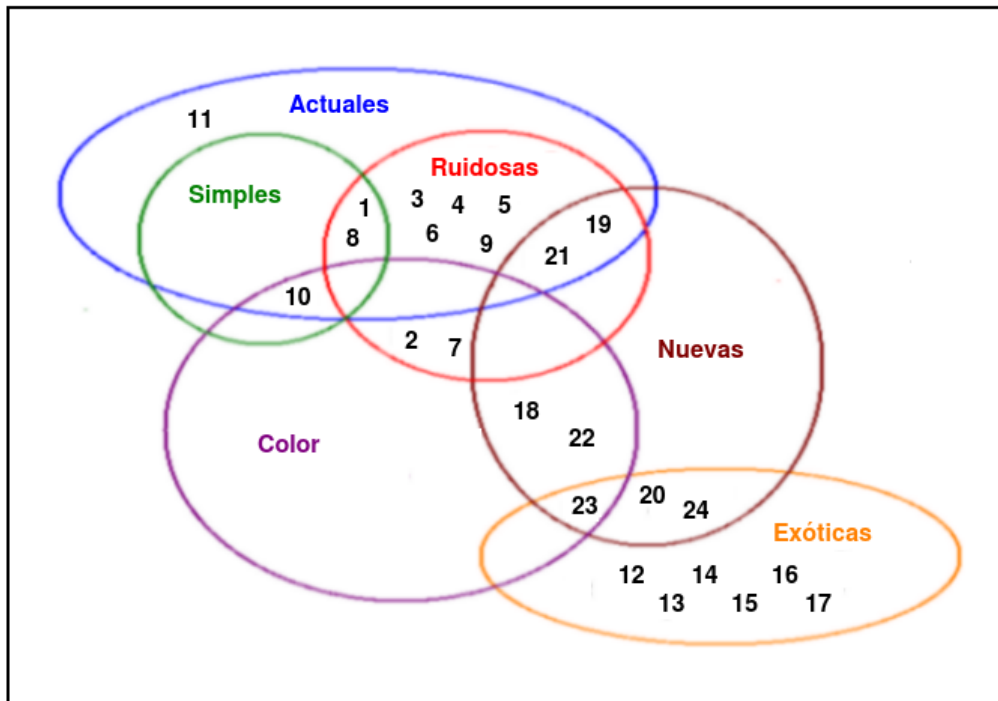


Figura 7.1: Diagrama de Venn con los grupos de distorsiones

1. **Distorsiones Ruidosas.** En este grupo se encuentran todas las distorsiones que implican agregar o quitar ruido a la imagen afectada, estas son las distorsiones #1, #2, #3, #4, #5, #6, #7, #8, #9, #19 y #21.
2. **Distorsiones Actuales.** Se refiere a los tipos de distorsiones más comunes en la práctica del procesamiento de imágenes, por ejemplo la compresión (distorsiones #10 y #11), también se encuentran en este grupo las distorsiones #1, #3, #4, #5, #6, #8, #9, #19 y #21.
3. **Distorsiones Simples.** Este grupo incluye solo las tres distorsiones básicas: Ruido Gaussiano Aditivo (distorsión #1), Desenfoque Gaussiano (distorsión #8) y Compresión JPG (distorsión #10).
4. **Distorsiones Exóticas.** Este subconjunto corresponde a distorsiones que ocurren con poca frecuencia pero que se encuentran entre las “más difíciles” para las métricas de calidad visual [8]. Esta compuesto por las distorsiones #12, #13, #14, #15, #16, #17, #20, #23 y #24.
5. **Distorsiones Nuevas.** Es el grupo formado por las distorsiones #18, #19, #20, #21, #22, #23, #24, que son las distorsiones nuevas que se han agregado a la base TID2013 con respecto a la base TID2008.
6. **Distorsiones del Color.** Se trata de las distorsiones que están relacionadas con los cambios de contenido de color de la imagen, estas son las distorsiones #2, #7, #10, #18, #23 y #24.



En el Apéndice G se presentan las medidas de resumen para cada grupo de distorsiones.

En los Cuadros 7.2 y 7.3 se presentan respectivamente, los coeficientes de correlación de Spearman y de Kendall entre cada métrica y el MOS separando los datos por grupo de distorsiones.

Métrica	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
MSE & RMSE	-0,7691	-0,7839	-0,8759	-0,5621	-0,7772	-0,7360	-0,6869
PSNR	0,7691	0,7839	0,8759	0,5621	0,7772	0,7360	0,6869
SNR	0,7207	0,7446	0,8352	0,5384	0,7323	0,6764	0,6491
WSNR	0,8711	0,8691	0,9227	0,4298	0,8872	0,9014	0,6382
UQI	0,6030	0,6403	0,7348	0,4718	0,5191	0,5092	0,5239
VIF	0,7525	0,8002	0,8456	0,5150	0,7659	0,7056	0,6338
NQM	0,8482	0,8507	0,8882	0,6116	0,8762	0,8937	0,7126
SSIM	0,6753	0,7204	0,7669	0,5372	0,7446	0,6766	0,6273
MSSIM	0,8096	0,8727	0,8861	0,7391	0,7996	0,7481	0,7909
CQ(1,1)	0,6510	0,6532	0,8359	0,6457	0,5285	0,4822	0,6008
GMSM	0,8928	0,8863	0,9474	0,7974	0,6473	0,6006	0,7884
GMSD	-0,9187	-0,9150	-0,9415	-0,8452	-0,6512	-0,5922	-0,8044

Cuadro 7.2: Correlación de Spearman entre las métricas y el MOS por grupos de distorsiones

Métrica	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
MSE & RMSE	-0,5619	-0,5762	-0,6892	-0,3923	-0,5760	-0,5373	-0,4958
PSNR	0,5619	0,5762	0,6892	0,3923	0,5763	0,5373	0,4958
SNR	0,5160	0,5383	0,6305	0,3725	0,5307	0,4806	0,4607
WSNR	0,6827	0,6813	0,7551	0,3046	0,7054	0,7210	0,4938
UQI	0,4194	0,4494	0,5230	0,3311	0,3650	0,3613	0,3695
VIF	0,5575	0,6010	0,6452	0,3663	0,5644	0,5157	0,4669
NQM	0,6557	0,6598	0,6997	0,4336	0,6924	0,7080	0,5348
SSIM	0,4777	0,5151	0,5610	0,3784	0,5347	0,4794	0,4457
MSSIM	0,6092	0,6757	0,6971	0,5444	0,5978	0,5481	0,5921
CQ(1,1)	0,4659	0,4692	0,6361	0,4589	0,3791	0,3446	0,4291
GMSM	0,7093	0,7004	0,7966	0,6071	0,5202	0,4848	0,6132
GMSD	-0,7461	-0,7408	-0,7948	-0,6528	-0,5248	-0,4723	-0,6339

Cuadro 7.3: Correlación de Kendall entre las métricas y el MOS por grupos de distorsiones

De los resultados expuestos en los Cuadros 7.2 y 7.3 se deduce que las métricas correlacionan mejor con el MOS cuando las imágenes a considerar han sido afectadas por distorsiones Simples, ya que las métricas calculadas sobre este conjunto de imágenes presentan, en casi todos los casos, el coeficiente de Spearman o de Kendall con el MOS más cercano a 1 o a -1 en comparación a los valores obtenidos para los otros grupos de distorsiones. Por otro lado las distorsiones Exóticas muestran los coeficientes más alejados de 1 y -1 para casi todas las métricas excepto para las direccionales: Índice CQ, GMSM y GMSD. A su vez, estas tres métricas tienen su peor desempeño

con las distorsiones del Color seguidas por las Nuevas.

En el Apéndice E se encuentran las matrices de confusión de cada métrica por cada uno de los seis grupos de distorsiones. En los Cuadros 7.4 y 7.5 se presentan respectivamente, los índices de concordancia kappa de Cohen y pi de Scott para cada grupo de distorsiones.

Métrica	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
MSE & RMSE & PSNR	0,4334	0,4502	0,5223	0,2826	0,4702	0,3944	0,3820
SNR	0,3755	0,3935	0,4542	0,2985	0,4445	0,3552	0,3600
WSNR	0,4747	0,4377	0,5423	0,3335	0,5977	0,6389	0,4450
UQI	0,2820	0,3151	0,4042	0,2842	0,2273	0,2386	0,2890
VIF	0,3517	0,4082	0,4844	0,2669	0,3197	0,2735	0,3495
NQM	0,5120	0,4907	0,5454	0,3919	0,6043	0,6429	0,4845
SSIM	0,3316	0,3805	0,4523	0,3205	0,3994	0,3631	0,3635
MSSIM	0,5102	0,5826	0,6446	0,4365	0,5050	0,4270	0,5095
CQ(1,1)	0,3539	0,3405	0,5373	0,2916	0,2982	0,3528	0,3170
GMSM	0,6083	0,5801	0,7517	0,4953	0,4693	0,5676	0,5505
GMSD	0,7048	0,6958	0,7518	0,5327	0,5750	0,5908	0,6150

Cuadro 7.4: Índice kappa de Cohen de cada métrica con el MOS en cada grupo de distorsiones

Métrica	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
MSE & RMSE & PSNR	0,4309	0,4463	0,5188	0,2728	0,4681	0,3943	0,3820
SNR	0,3732	0,3900	0,4512	0,2921	0,4444	0,3550	0,3600
WSNR	0,4673	0,4272	0,5365	0,3157	0,5970	0,6383	0,4450
UQI	0,2761	0,3126	0,4017	0,2785	0,2239	0,2360	0,2890
VIF	0,3461	0,4054	0,4822	0,2551	0,3117	0,2628	0,3495
NQM	0,5084	0,4849	0,5425	0,3833	0,6036	0,6427	0,4845
SSIM	0,3268	0,3781	0,4513	0,3137	0,3970	0,3596	0,3635
MSSIM	0,5093	0,5813	0,6437	0,4340	0,5039	0,4223	0,5095
CQ(1,1)	0,3538	0,3398	0,5342	0,2896	0,2942	0,3493	0,3170
GMSM	0,6074	0,5788	0,7512	0,4918	0,4685	0,5659	0,5505
GMSD	0,7047	0,6957	0,7517	0,5290	0,5736	0,5878	0,6150

Cuadro 7.5: Índice pi de Scott de cada métrica con el MOS en cada grupo de distorsiones

Los índices kappa de Cohen y pi de Scott son prácticamente iguales. Si nos restringimos a las distorsiones Simples, las métricas presentan generalmente valores de estos índices más altos, excepto la WSNR y la NQM. Si nos restringimos a las distorsiones Exóticas los valores son casi siempre los más bajos, excepto para UQI, MSSIM y GMSM.

En el Cuadro 7.6 se presenta un ranking de las mejores métricas para cada grupo de distorsiones. Este ranking se estableció redondeando los coeficientes de Spearman a un dígito decimal, las métricas que obtuvieron un coeficiente con el MOS de valor absoluto 0,9 para un cierto grupo de

distorsiones se consideran como la primera opción para ese grupo de distorsiones; las que obtuvieron un coeficiente de módulo 0,8, la segunda, y así sucesivamente hasta la quinta opción para módulos de 0,5 que fue el más bajo obtenido. Cómo el MSE, RMSE y PSNR tienen los mismos módulos de  $\rho$  en todos los casos, solo se incluyó al MSE por simplicidad.

	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
Primeras $ \rho  = 0,9$	WSNR, GMSM, GMSD	WSNR, NQM, MSSIM, GMSM, GMSD	MSE		WSNR, NQM	WSNR, NQM	
Segundas $ \rho  = 0,8$	MSE, VIF, NQM, MSSIM	MSE, VIF	WSNR, NQM, MSSIM, GMSM, GMSD	MSSIM, GMSM, GMSD	MSE, VIF, MSSIM		MSSIM, GMSM, GMSD
Terceras $ \rho  = 0,7$	SNR, SSIM, CQ(1,1)	SNR, SSIM, CQ(1,1)	SNR, VIF, SSIM, CQ(1,1)	NQM, CQ(1,1)	SNR, SSIM, GMSD	MSE, SNR, VIF, SSIM, MSSIM	MSE, NQM
Cuartas $ \rho  = 0,6$	UQI	UQI	UQI	MSE, SNR, UQI, VIF, SSIM	GMSM	GMSM, GMSD	SNR, WSNR, VIF, SSIM, CQ(1,1)
Quintas $ \rho  = 0,5$				WSNR	UQI, CQ(1,1)	UQI, CQ(1,1)	UQI

Cuadro 7.6: Ranking de métricas para cada grupo de distorsiones según el coeficiente de Spearman

En el Cuadro 7.7 se establece un ranking similar, esta vez utilizando el coeficiente de Kendall. Los resultados son muy cercanos a los obtenidos en el ranking anterior.

Se realizó el mismo proceso con el índice kappa de Cohen y el índice pi de Scott, como los resultados fueron muy similares, se resume en una sola clasificación que puede apreciarse en el Cuadro 7.8.

En base a los rankings establecidos, definimos uno nuevo que resume a los tres en el Cuadro 7.9.

En el Cuadro 7.10 se presenta el coeficiente de concordancia multivariado  $w$  de Kendall por grupo de distorsiones. Todos los p-valores asociados dieron 0, indicando que los coeficientes obtenidos son significativos.

	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
Primeras $ \tau  = 0,8$			WSNR, GMSM, GMSD				
Segundas $ \tau  = 0,7$	WSNR, NQM, GMSM, GMSD	WSNR, NQM, MSSIM, GMSM, GMSD	MSE NQM, MSSIM	GMSD	WSNR, NQM	WSNR, NQM	
Terceras $ \tau  = 0,6$	MSE, VIF, MSSIM	MSE, VIF	SNR, VIF, SSIM, CQ(1,1)	MSSIM, GMSM	MSE, VIF, MSSIM		MSSIM, GMSM, GMSD
Cuartas $ \tau  = 0,5$	SNR, SSIM, CQ(1,1)	SNR, SSIM, CQ(1,1)	UQI	CQ(1,1)	SNR, SSIM, GMSM, GMSD	MSE, SNR, VIF, SSIM, MSSIM, GMSM, GMSD	MSE, SNR, WSNR, VIF, NQM
Quintas $ \tau  = 0,4$	UQI	UQI		MSE, SNR, VIF, NQM, SSIM	UQI, CQ(1,19)	UQI	UQI, SSIM, CQ(1,1)
Sextas $ \tau  = 0,3$				WSNR, UQI		CQ(1,1)	

Cuadro 7.7: Ranking de métricas para cada grupo de distorsiones según el coeficiente de Kendall

Por grupos de distorsiones se obtuvieron valores de  $w$  más altos que considerando todos los tipos de distorsiones, excepto para las distorsiones Exóticas que obtuvo un valor de  $w$  menor al  $w$  global.

En el Cuadro 7.11 se presenta el índice de concordancia multivariado kappa de Fleiss por grupo de distorsiones.

Al igual que lo que ocurrió con el  $w$  de Kendall, todos los grupos tienen un kappa de Fleiss mayor que el global, excepto el grupo de distorsiones Exóticas que presenta un valor menor. Según el Cuadro 6.1, el acuerdo entre las métricas es moderado en estos grupos, y justo en el grupo de las distorsiones Exóticas.

Al realizar el análisis por grupos de distorsiones es pertinente considerar sólo las métricas que funcionan mejor para ese grupo de distorsiones. Basandonos en el Cuadro 7.9, calcularemos el  $w$  de Kendall y el kappa de Fleiss por grupo de distorsiones solamente para las métricas que quedaron primeras y segundas en el ranking 7.9. Los resultados se muestran en los Cuadros 7.12 y 7.13.

	Ruidosas	Actuales	Simples	Exóticas	Nuevas	Color	Todas
Primeras $\kappa = 0,8$ o $0,7$	GMSD	GMSD	GMSM, GMSD				
Segundas $\kappa = 0,6$	GMSM	MSSIM, GMSM	MSSIM		WSNR, NQM, GMSD	WSNR, NQM, GMSM, GMSD	GMSM, GMSD
Terceras $\kappa = 0,5$	WSNR, NQM, MSSIM	MSE, NQM	MSE, SNR, WSNR, VIF, NQM, SSIM, CQ(1,1)	GMSM, GMSD	MSE, MSSIM, GMSM		NQM, MSSIM
Cuartas $\kappa = 0,4$	MSE, SNR, VIF, CQ(1,1)	SNR, WSNR, VIF, SSIM	UQI	NQM, MSSIM	SNR, SSIM	MSE, SNR, SSIM, MSSIM, CQ(1,1)	MSE, SNR, WSNR, SSIM
Quintas $\kappa = 0,3$	UQI, SSIM	UQI, CQ(1,1)		MSE, SNR, WSNR, UQI, VIF, SSIM, CQ(1,1)	VIF, CQ(1,1)	VIF	UQI, VIF, CQ(1,1)
Sextas $\kappa = 0,2$					UQI	UQI	

Cuadro 7.8: Ranking de métricas para cada grupo de distorsiones según el índice kappa de Cohen y el el índice pi de Scott

En todos los casos se obtuvieron coeficientes de concordancia  $w$  de Kendall más cercanos a 1 cuando nos restringimos a las métricas que vienen demostrando mejor desempeño para un determinado grupo de distorsiones.

En el caso del kappa de Fleiss, al restringirnos a las métricas más significativas para cada grupo, obtuvimos índices un poco más altos, aunque en casi todos los casos el acuerdo sigue siendo moderado.

	Ruidosas	Actuales	Simplees	Exóticas	Nuevas	Color	Todas
Primeras	GMSD, GMSM, WSNR	GMSD, GMSM, MSSIM, NQM	GMSD, GMSM	GMSD, GMSM	WSNR, NQM	WSNR, NQM	GMSD, GMSM, MSSIM
Segundas	NQM, MSSIM	WSNR, MSE	MSSIM, WSNR, MSE, NQM	MSSIM	MSE, MSSIM, VIF, GMSD, GMSM	MSE, SNR, SSIM, MSSIM	NQM
Terceras	MSE, VIF	VIF, SNR, SSIM, CQ(1,1)	SNR, VIF, SSIM, CQ(1,1)	NQM, CQ(1,1)	SNR, SSIM	GMSM, GMSD, VIF	SSIM, WSNR, MSE, SNR
Cuartas	SNR, SSIM, CQ(1,1), UQI	UQI	UQI	MSE, SNR, WSNR, UQI, VIF, SSIM	CQ(1,1), UQI	CQ(1,1), UQI	VIF, CQ(1,1), UQI

Cuadro 7.9: Ranking resumido para cada grupo de distorsiones.

Grupo de distorsiones	w de Kendall
Ruidosas	0,7837
Actuales	0,8005
Simplees	0,8575
Exóticas	0,5952
Nuevas	0,7519
Color	0,7094
Todas	0,6725

Cuadro 7.10: Índice de concordancia w de Kendall de las métricas con el MOS por grupo de distorsiones

Grupo de distorsiones	Kappa de Fleiss
Ruidosas	0,4450
Actuales	0,4637
Simplees	0,5238
Exóticas	0,3377
Nuevas	0,4813
Color	0,4379
Todas	0,4137

Cuadro 7.11: Índice kappa de Fleiss de las métricas con el MOS por grupo de distorsiones

Grupo de distorsiones	w de Kendall	Métricas consideradas
Ruidosas	0,9024	GMSD, GSM, WSNR, NQM, MSSIM
Actuales	0,8746	GMSD, GSM, MSSIM, NQM, WSNR, MSE
Simples	0,9110	GMSD, GSM, MSSIM, WSNR, MSE, NQM
Exóticas	0,8726	GMSD, GSM, MSSIM
Nuevas	0,7909	WSNR, NQM, MSE, MSSIM, VIF, GMSD, GSM
Color	0,8080	WSNR, NQM, MSE, SNR, SSIM, MSSIM
Todas	0,8125	GMSD, GSM, MSSIM, NQM

Cuadro 7.12: Coeficiente de concordancia w de Kendall de las métricas con el MOS por grupo de distorsiones. Se consideraron sólo las métricas más apropiadas para cada grupo.

Grupo de distorsiones	Kappa de Fleiss	Métricas consideradas
Ruidosas	0,5666	GMSD, GSM, WSNR, NQM, MSSIM
Actuales	0,5274	GMSD, GSM, MSSIM, NQM, WSNR, MSE
Simples	0,6003	GMSD, GSM, MSSIM, WSNR, MSE, NQM
Exóticas	0,5084	GMSD, GSM, MSSIM
Nuevas	0,4989	WSNR, NQM, MSE, MSSIM, VIF, GMSD, GSM
Color	0,4748	WSNR, NQM, MSE, SNR, SSIM, MSSIM
Todas	0,5326	GMSD, GSM, MSSIM, NQM

Cuadro 7.13: Índice kappa de Fleiss de las métricas con el MOS por grupo de distorsiones. Se consideraron sólo las métricas más apropiadas para cada grupo.

---

## Capítulo 8

### Conclusiones y trabajos a futuro

Si se conoce el tipo de distorsión que ha sufrido la imagen a evaluar, este trabajo brinda un criterio para considerar las métricas que presentan un mejor rendimiento, pues del análisis por grupos de distorsiones realizado, podemos concluir cuál es la métrica que mejor funciona para cada grupo de distorsiones, y cuáles le siguen en segundo y tercer lugar. En el Cuadro 8.1 se exhiben las conclusiones a las que hemos llegado.

Grupo	Primera	Segunda	Tercera
Ruidosas	GMSD	GMSM	WSNR
Actuales	GMSD	GMSM	MSSIM
Simples	GMSM	GMSD	WSNR
Exóticas	GMSD	GMSM	MSSIM
Nuevas	WSNR	NQM	MSSIM
Color	WSNR	NQM	MSSIM
Todas	GMSD	GMSM	MSSIM, NQM

Cuadro 8.1: Métrica más apropiada por grupo de distorsión.

El grupo de distorsiones Exóticas fue el más problemático, ya que es en el cual la mayoría de las métricas tienen el peor desempeño. En el caso de la compresión de imágenes (distorsiones #10 y #11), muy habitual en la práctica, la métrica GMSD resultó ser la más adecuada. Es notable que, para los grupos de distorsiones Nuevas y del Color, la métrica más apropiada es la WSNR, seguida de la NQM, a diferencia de los otros grupos donde siempre es preferible utilizar una direccional.

No obstante en la práctica es muy difícil conocer el tipo de distorsión por la cual ha sido contaminada una imagen, se necesita para ello información adicional, por caso sobre el método de captación y/o transmisión, entre otros. Por ejemplo, si se trata de una imagen enviada por mail la misma estará afectada por la compresión, además de posiblemente otras distorsiones.



Si bien frecuentemente no vamos a conocer el tipo de distorsión que afectó una imagen, este trabajo revela que, desconociendo este factor, las herramientas que mostraron mejor rendimiento en la comparación son las direccionales. Pues en base a todos los análisis realizados a lo largo de este trabajo, podemos concluir que de todas las métricas consideradas, la que tiene un mejor desempeño en general es la GMSD y en segundo lugar la GSM; en tercer lugar tenemos la métrica MSSIM y luego la NQM.

Aunque estas son las medidas que presentaron mayor correlación con el MOS en el grupo de Todas las imágenes, es decir independientemente del tipo de distorsión aplicada, esto no implica que las distorsiones que afectaron a estas imágenes tengan algún factor direccional, sino que las métricas basadas en la magnitud del gradiente lograron captar la similitud con la imagen original más frecuentemente que otras métricas, independientemente del tipo de distorsión que las haya afectado. Nos queda entonces, como trabajo a futuro, dilucidar qué es lo que hace a este tipo de medidas tan efectivas para captar similitud.

Un aporte interesante de esta tesis es haber completado el estudio de Ponomarenko et al. incorporando las métricas direccionales GMSD, GSM y CQ(1,1) al cuadro propuesto en [8]. Un hallazgo importante es que precisamente dos de estas nuevas métricas son las que mejor correlacionan con el MOS tanto en general como por grupo de distorsión excepto en los grupos Nuevas y de Color.

En el caso de la métrica direccional CQ, solo hemos considerado la dirección (1, 1), nos proponemos como trabajo a futuro indagar sobre un conjunto de direcciones de interés y obtener una función resumen (promedio, mediana, máximo, mínimo) de estos valores de CQ en varias direcciones.

A lo largo de este trabajo se utilizaron una gran variedad de índices de correlación y concordancia, queda también, como trabajo a futuro, compararlos y analizar la correspondencia entre ellos.

Por último, otro trabajo a futuro es proponer una nueva métrica para evaluar similaridad y calidad en imágenes. Como hemos visto, muchas de las métricas analizadas presentan deficiencias, como es el caso del MSE, el SNR, entre otros. También hemos hallado métricas con mejor desempeño, sin embargo el desafío de encontrar una herramienta que a pesar de la distorsión que sufra la imagen, sea capaz de asociarla a su imagen original, no es un tema cerrado. Por otra parte, hasta el momento son pocas las iniciativas que se han concentrado en las propiedades estadísticas, de manera que proponer una medida que cumpla con los objetivos anteriores sigue siendo un tema atractivo y de interés para la matemática y la estadística, como así también del análisis y procesamiento de imágenes.

---

# Bibliografía

- [1] Lin, W., and Kuo, J., 2011. Perceptual visual quality metrics: A survey. *J. Vis. Commun. Image R.* 22(2011) 297-312.
- [2] Pappas, T. and Safranek, R., 2000. Percentual criteria for image quality evaluation, *Handbook of Image Video Processing*, A. C. Bovik, Ed. New-York: Academic.
- [3] Wang Z., Bovik A., 2009. Mean squared error: love it or leave it? - a new look at fidelity measures, *IEEE Signal Process. Magazine*.
- [4] Zhou Wang and Alan C. Bovik. *Modern Image Quality Assessment*.
- [5] Sheikh H. R., Wang Z., Cormack L., and Bovik A. C. LIVE image quality assessment. Database Release 2, [Online]. Available: <http://live.ece.utexas.edu/research/quality>.
- [6] Ponomarenko N., Lukin V., Zelensky A., Egiazarian A., Carli M., and Battisti F., 2009. TID2008- A database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, vol. 10, pp. 30 & 45.
- [7] Ponomarenko N., Ieremeiev O., Lukin V., Egiazarian K., Jin L., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., and Jay Kuo C., 2013. Color image database TID2013: Peculiarities and preliminary results, 4th European Workshop on Visual Information Processing EUVIP2013, pp.106-111.
- [8] Nikolay Ponomarenko and Lina Jin and Oleg Ieremeiev and Vladimir Lukin and Karen Egiazarian and Jaakko Astola and Benoit Vozel and Kacem Chehdi and Marco Carli and Federica Battisti and C.-C. Jay Kuo, 2014. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, pp 57-77, vol. 30.
- [9] Larson, E. C., and Chandler, D. M. Categorical image quality (CSIQ) database, [Online]. Available: <http://vision.okstate.edu/csi>.
- [10] Deng, Ch., Ma L., Lin, W., and Ngan K., 2015. *Visual Signal Quality Assessment. Quality of Experience (QoE)*. Springer International Publishing.

- 
- [11] Deng, Ch., Ma L., Lin, W., and Ngan K., 2015. Visual Signal Quality Assessment. Quality of Experience (QoE). Springer International Publishing.
- [12] Wang Z., Bovik A., 2009. Mean squared error: love it or leave it? - a new look at fidelity measures, *IEEE Signal Process. Magazine*.
- [13] Niranjana Damera-Venkata, Thomas D. Kite, Wilson S. Geisler, Brian L. Evans, and Alan C. Bovik, Apr. 2000. Image Quality Assessment Based on a Degradation Model. *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636-650.
- [14] E. Peli, Oct. 1990. Contrast in complex images, *J. Opt. Soc. Amer. A*, vol. 7, pp. 2032-2039.
- [15] Wang Z. and Bovik A., 2002. A universal image quality index. *IEEE Signal Processing Letters*, vol. 9, pp. 81-84.
- [16] Wang Z. Bovik, A., Sheikh, H. and Simoncelli, E. 2004, Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process*, vol. 13, pp. 600-612.
- [17] Wang Z., Simoncelli E. P., and Bovik A. C., 2003. Multi-scale structural similarity for image quality assessment. *IEEE Asilomar Conference Signals, Systems and Computers*, pp. 1398-1402.
- [18] Liu H. and Heynderickx I., 2011. Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971-982.
- [19] Wang Z. and Li Q., 2011. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185-1198.
- [20] Gu K., Zhai G., Yang X., Zhang W., and Liu M., 2013. Structural similarity weighting for image quality assessment, *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, pp. 1-6.
- [21] Ojeda Silvia M., Vallejos Ronny O., Lamberti Pedro W., 2012. Measure of similarity between images based on the codispersion coefficient. *Journal of Electronic Imaging*.
- [22] Pistonessi S., Martinez J., Ojeda Silvia M., Vallejos Ronny O., 2015. A Novel Quality Image Fusion Assessment Based on Maximum Codispersion. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 20th Iberoamerican Congress, CIARP 2015 Montevideo, Uruguay*.
- [23] Xue W., Zhang L., Mou X., and Bovik A. C., 2014. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684-695.

- 
- [24] Zhang L., Zhang L., Mou X., and Zhang D., 2011. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378-2386.
- [25] Liu A., Lin W., and Narwaria M., 2012. Image quality assessment based on gradient similarity. *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500-1512.
- [26] Sheikh H. R., and Bovik A. C., 2006. Image information and visual quality, *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444.
- [27] Zhai G., Zhang W., Yang X., Yao S., and Xu Y., 2006. GES: a new image quality assessment metric based on energy features in Gabor transform domain, *Proc. IEEE Int. Symposium on Circuits and Systems*, pp. 1715-1718.
- [28] Zhai G., Zhang W., Xu Y., and Lin W., 2007. LGPS: Phase based image quality assessment metric. *IEEE Workshop on Signal Processing Systems*, pp. 605-609.
- [29] Larson E. C. and Chandler D. M., 2010. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, vol. 19, no. 1.
- [30] Zhang F., Jiang W., Autrusseau F., and Lin W., 2014. Exploring V1 by modeling the perceptual quality of images, *Journal of Vision*, vol. 14, no. 1, pp. 1-14.
- [31] Wu J., Lin W., Shi G., and Liu A., 2013. Perceptual quality metric with internal generative mechanism. *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43-54
- [32] Zhai G., Wu X., Yang X., Lin W., and Zhang W., 2012. A psychovisual quality metric in free-energy principle. *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41-52.
- [33] Soundararajan R. and Bovik A. C., 2012. RRED indices: Reduced-reference entropic differencing for image quality assessment. *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517-526.
- [34] Narwaria M., Lin W., McLoughlin I. V., Emmanuel S., and Chia L. T., 2012. Fourier transform-based scalable image quality measure. *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3364-3377.
- [35] Rehman A. and Wang Z., 2012. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378-3389.
- [36] Gu K., Zhai G., Yang X., and Zhang W., 2013. A new reduced-reference image quality assessment using structural degradation model. *Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 1095-1098.
- [37] Zoran D. and Weiss Y., 2009. Scale invariance and noise in natural images. *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2209-2216.

- 
- [38] Zhai G. and Wu X., 2011. Noise estimation using statistics of natural images. Proc. IEEE Int. Conf. Image Process., pp. 1857-1860.
- [39] Liu X., Tanaka M., and Okutomi M., 2013. Single-image noise level estimation for blind denoising. IEEE Trans. Image Process., vol. 22, no. 12, pp. 5226-5237.
- [40] Wang Z., Sheikh H. R., and Bovik A. C., 2002. No-reference perceptual quality assessment of JPEG compressed images. Proc. IEEE Int. Conf. Image Process., pp. 477-480.
- [41] Moorthy A. K. and Bovik A. C., 2011. Blind image quality assessment: From scene statistics to perceptual quality. IEEE Trans. Image Process., pp. 3350-3364, vol. 20, no. 12.
- [42] Saad M. A., Bovik A. C., and Charrier C., 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Trans. Image Process., pp. 3339-3352, vol. 21, no. 8.
- [43] Mittal A., Moorthy A. K., and Bovik A. C., 2012. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process., pp. 4695-4708, vol. 21, no. 12.
- [44] Xue W., Zhang L., and Mou X., 2013. Learning without human scores for blind image quality assessment. Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recognition, pp. 995-1002.
- [45] Mittal A., Soundararajan R., and Bovik A. C., 2013. Making a completely blind image quality analyzer. IEEE Signal Process. Letters, pp. 209-212, vol. 22, no. 3.
- [46] Jähne B., Haubecker H., and Geibler P., 1999. Handbook of Computer Vision and Applications. New York: Academic.
- [47] Vallejos, R., Mancilla, D., Acosta, J., 2016. Image Similarity Assessment Based on Coefficients of Spatial Association. J Math Imaging Vis DOI 10.1007/s10851-016-0635-y
- [48] Goshtasby, A. (2012) Image Registration. Principles, Tools and Methods. Springer-Verlag London.
- [49] Allende, H., Galbiati, J., Vallejos, R., 2001. Robust Image Modelling on Image Processing. Pattern Recognition Letters, 22, 1219-1230.
- [50] Cliff, A., Ord, J., 1981. Spatial processes: Models and applications. Pion Ltd., London.
- [51] Paul Bourke, Peter Diprose and Bill Rattenbury, November 1993. A Beginners Guide to Bitmaps.
- [52] Nadenau Marcus J., Winkler Stefan, Alleysson David, Kunt Murat. Human Vision Models for Perceptually Optimized Image Processing - A Review

- [53] Liu M., Zhai G., Gu K., Xu Q., Yang X., Sun X., Chen W., and Zuo Y., 2013. A new image quality metric based on MIx-Scale transform, Proc. IEEE Workshop on Signal Processing Systems, pp. 266-271.
- [54] Peli E., 1990. Contrast in complex images. Journal of Optical Society of America, vol. 7, pp. 2032-2040.
- [55] Gu K., Zhai G., Yang X., and Zhang W., 2013. Self-adaptive scale transform for IQA metric, Proc. IEEE Int. Symp. Circuits and Syst., pp. 2365-2368
- [56] Gu K., Zhai G., Liu M., Xu Q., Yang X., and Zhang W., 2013. Adaptive high-frequency clipping for improved image quality assessment, Proc. IEEE Visual Communications and Image Processing, pp. 1-5.
- [57] Gu K., Zhai G., Yang X., Chen L., and Zhang W., 2012. Nonlinear additive model based saliency map weighting strategy for image quality assessment. IEEE International Workshop on Multimedia Signal Processing, pp. 313-318.
- [58] Itti L., Koch C., and Niebur E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, pp. 1254-1259.
- [59] J. M. S. Prewitt, "Object enhancement and extraction," in Picture Processing and Psychopictorics, Academic Press, 1970.
- [60] Gu K., Zhai G., Yang X., Zhou J., Gu X., and Zhang W., 2014. An efficient color image quality metric with local-tuned-global model, Proc. IEEE Int. Conf. Image Process.
- [61] Ojeda Silvia M., Britos Grisel M., Vallejos Ronny. An image quality index based on coefficients of spatial association with an application to image fusion. Spatial Statistics 23 (2018) 1-16.
- [62] <https://www.python.org/>
- [63] <https://bitbucket.org/kuraiev/pymetrikz>
- [64] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html>
- [65] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html>
- [66] <https://docs.scipy.org/doc/scipy-0.19.1/reference/generated/scipy.stats.boxcox.html>

- 
- [67] Badii, M.H., A. Guillen, O.P. Lugo Serrato and J.J. Aguilar Garnica UANL, San Nicolás, N.L., México, mhbadiaz@gmail.com. *Correlación No-Paramétrica y su Aplicación en la Investigaciones Científicas (Non-Parametric Correlation and Its Application in Scientific Research)*. Daena: International Journal of Good Conscience. 9(2)31-40. Agosto 2014. ISSN 1870-557X.
- [68] Landis J. R, Koch G. G. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.
- [69] Mukaka M. M. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*; 24(3):69-71 September 2012.
- [70] Noether G. E. Why Kendall Tau? *Teaching Statistics*; 3(2):41-3 May 1981.
- [71] Kendall, M. G.; Babington Smith, B. (Sep 1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*. 10(3):275-287.
- [72] Fawcett, Tom (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*. 27 (8): 861-874.
- [73] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- [74] Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19(3), 321-325.
- [75] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5 pp. 378-382.

---

# Apéndice A

## Diagramas de dispersión de las distorsiones #1 a #9

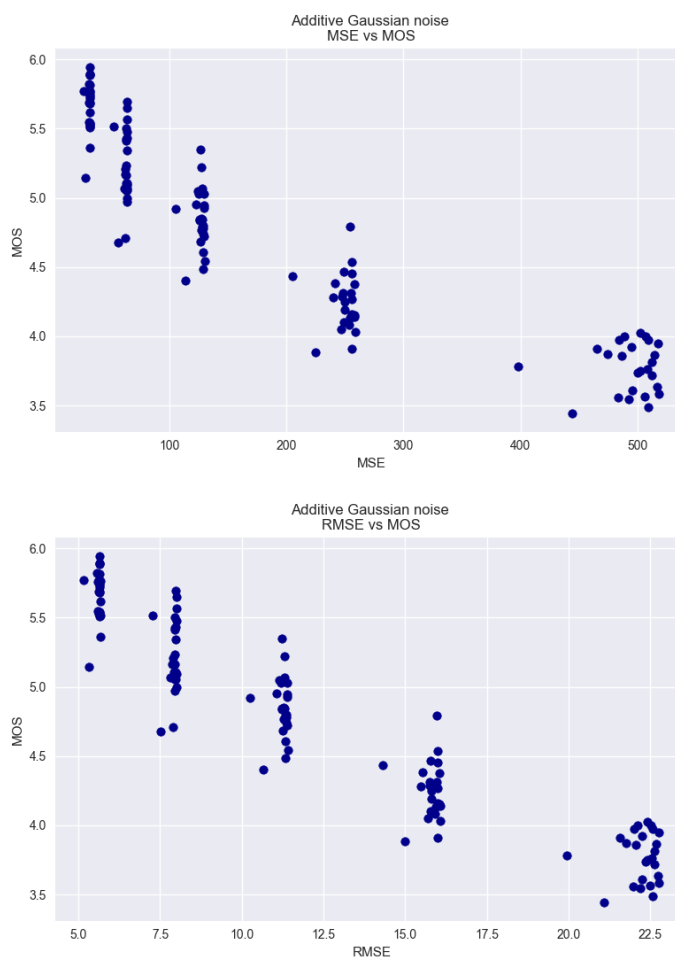


Figura A.1: Distorsión #1: Additive Gaussian noise



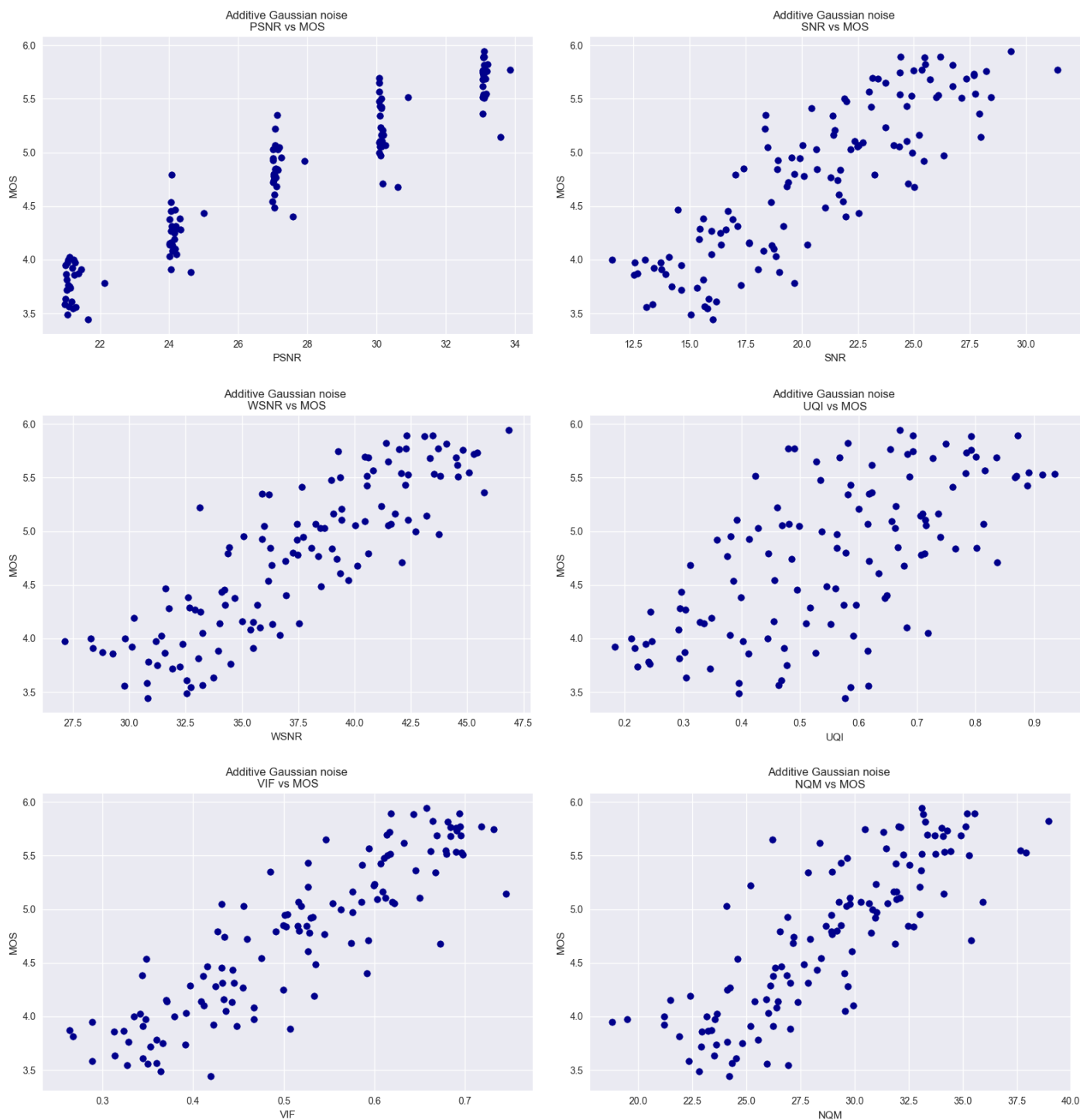


Figura A.2: Distorsión #1: Additive Gaussian noise

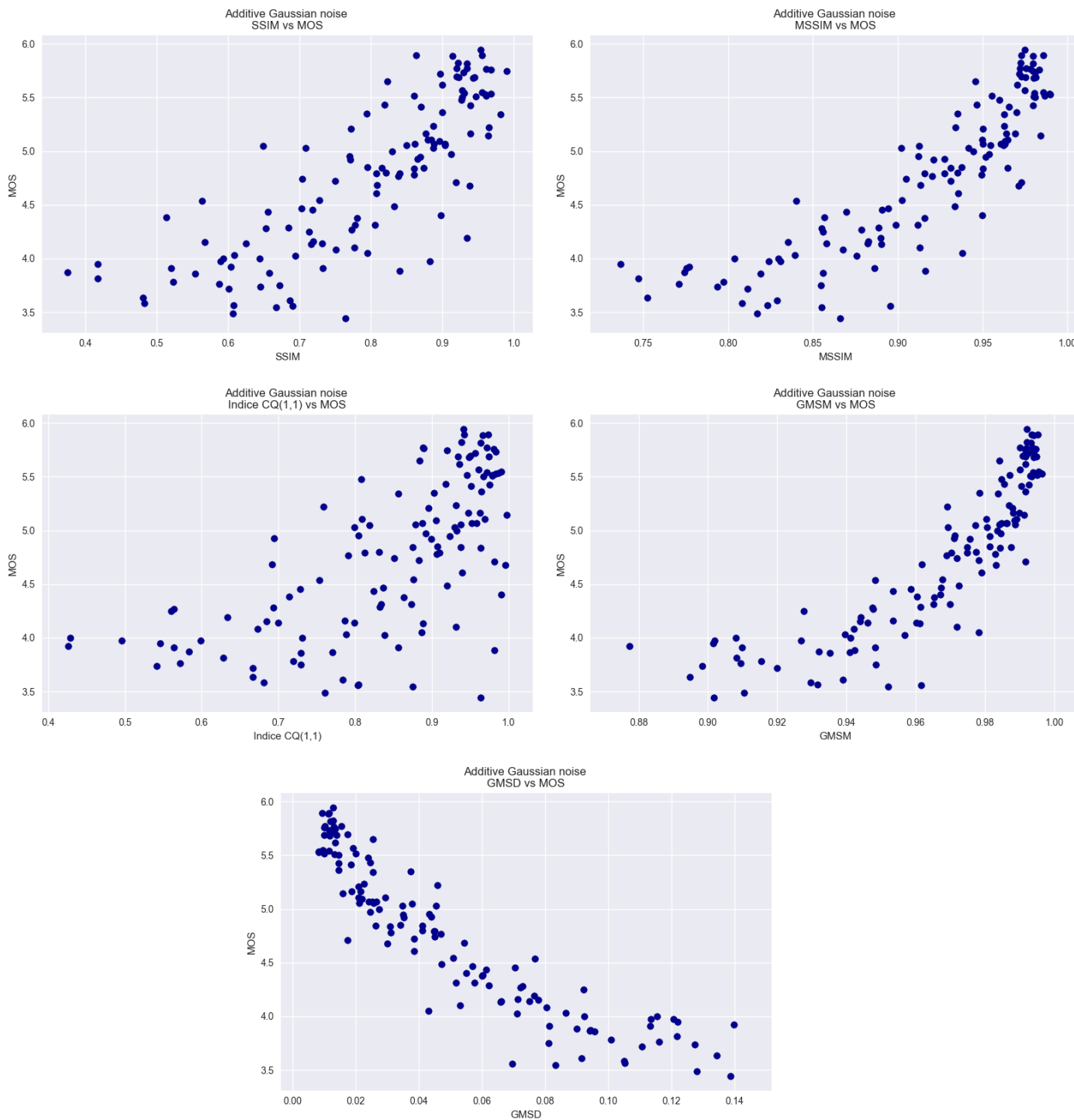


Figura A.3: Distorsión #1: Additive Gaussian noise

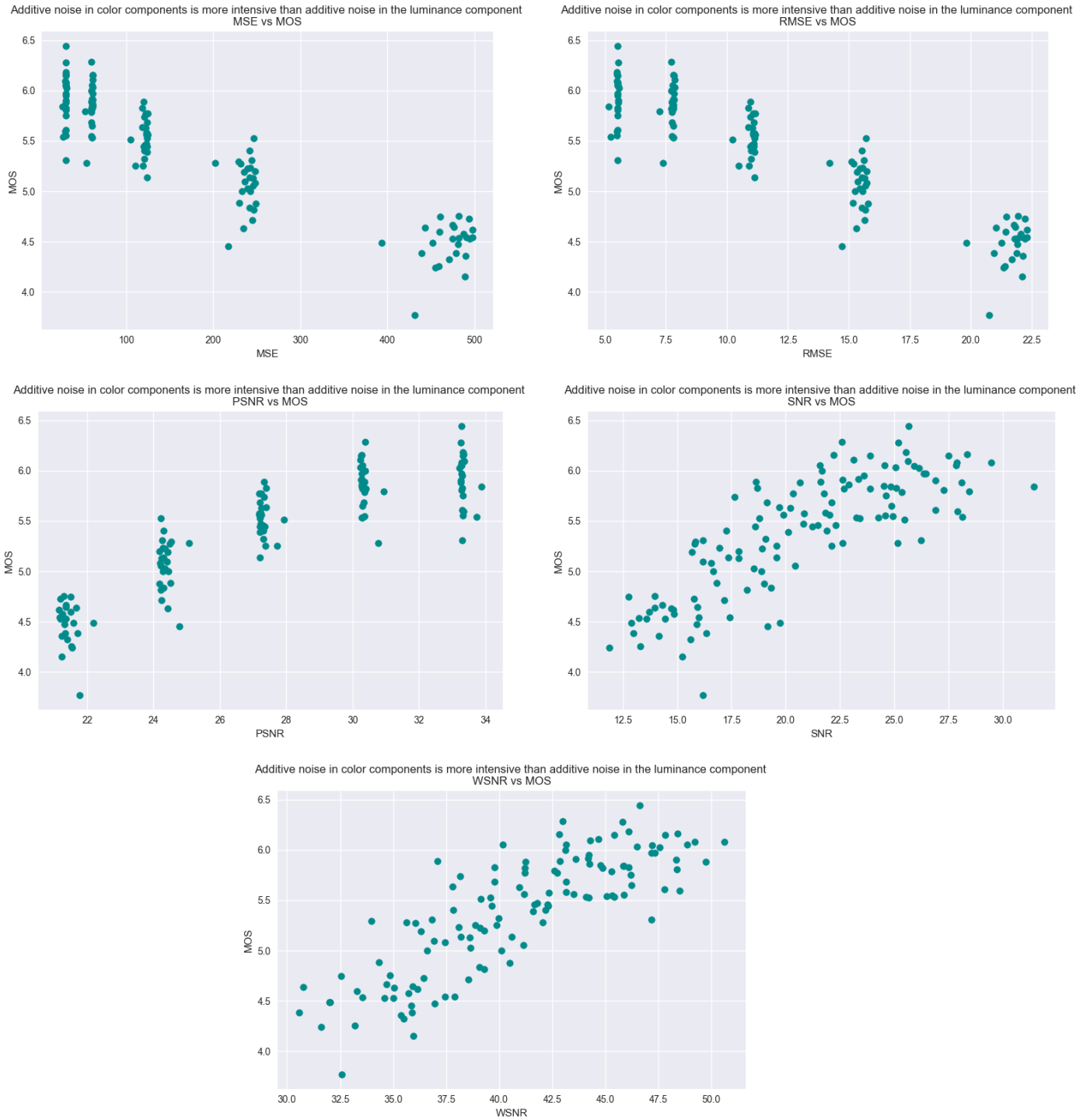


Figura A.4: Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component

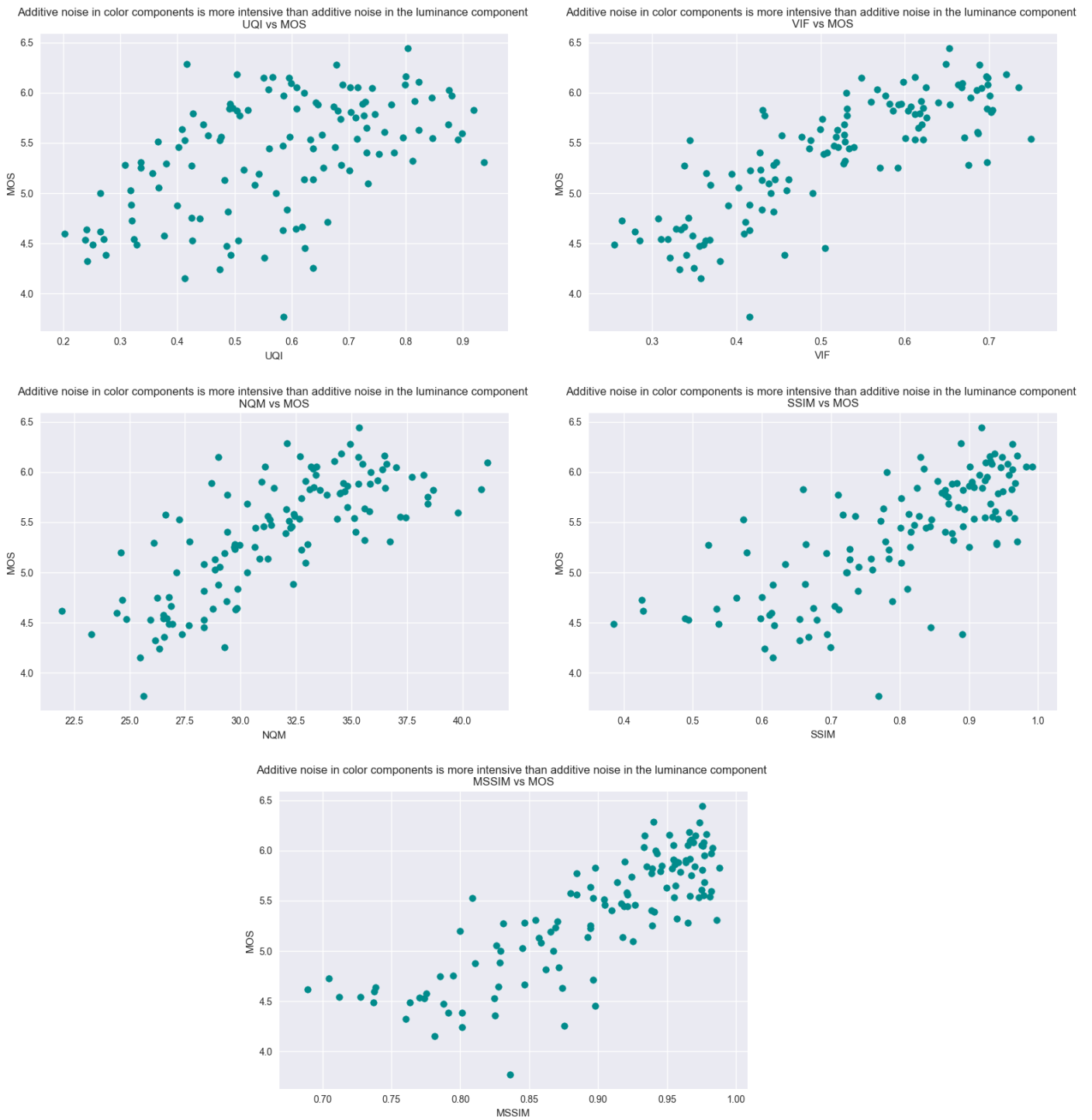


Figura A.5: Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component

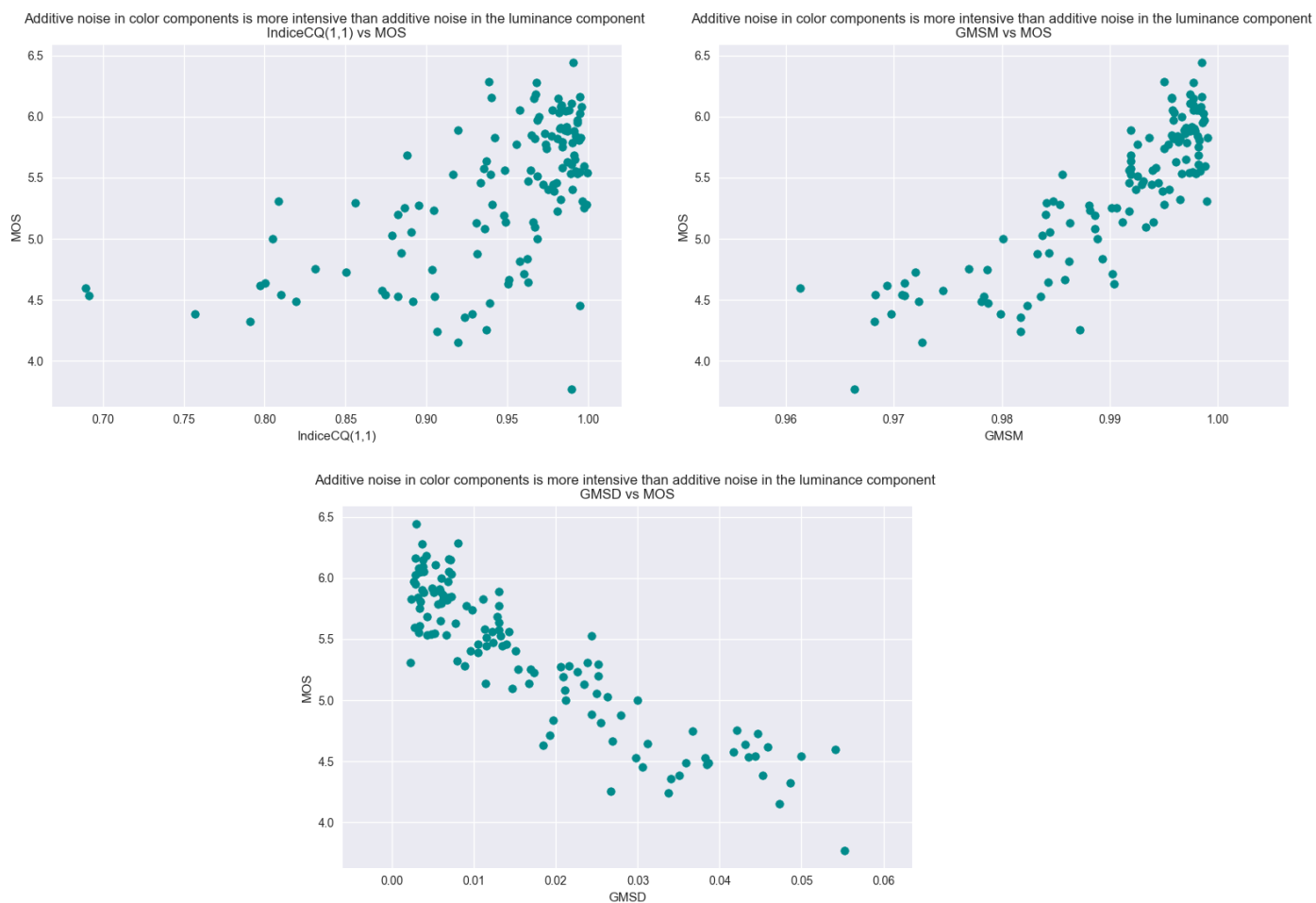


Figura A.6: Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component

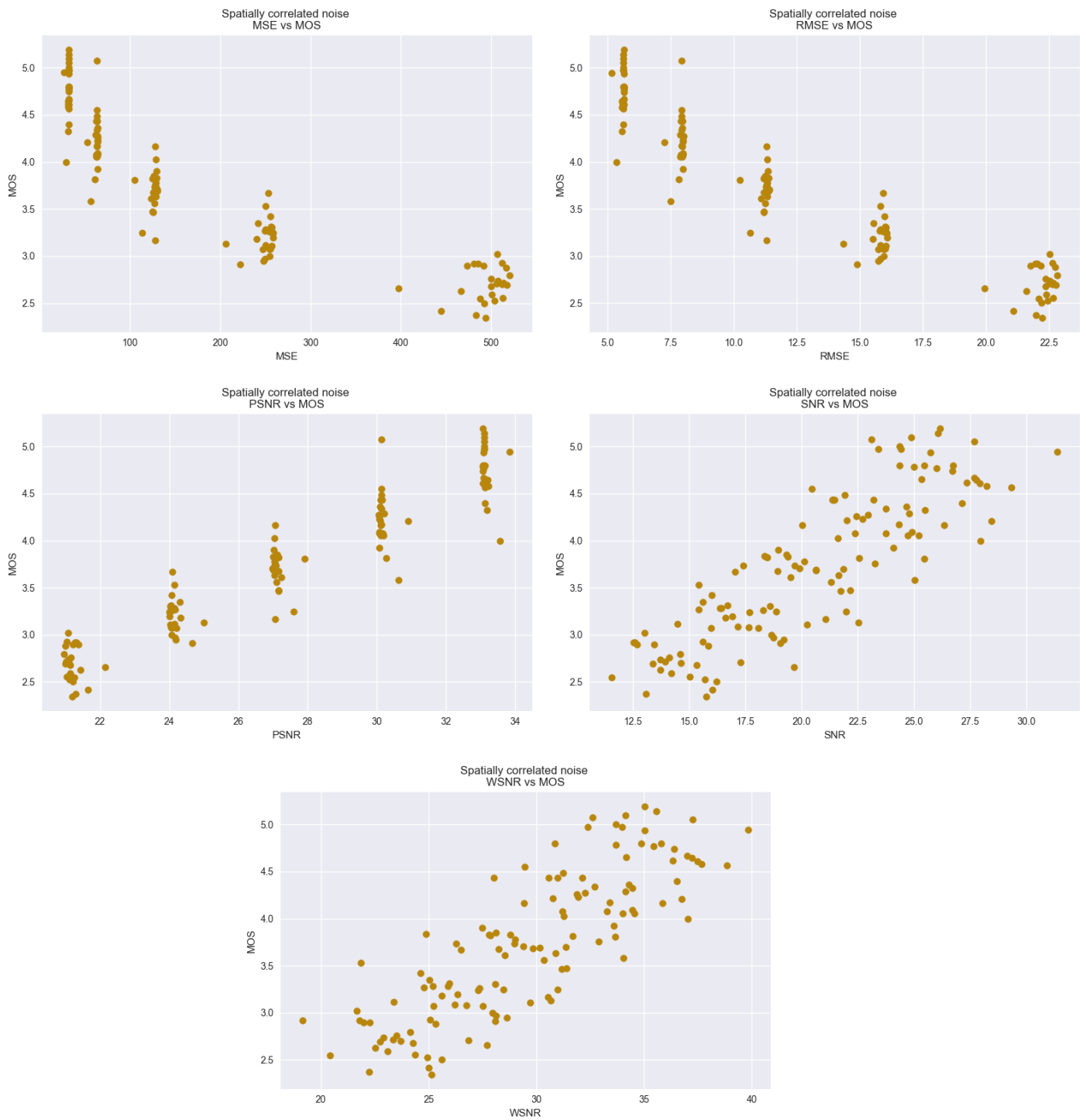


Figura A.7: Distorsión #3: Spatially correlated noise

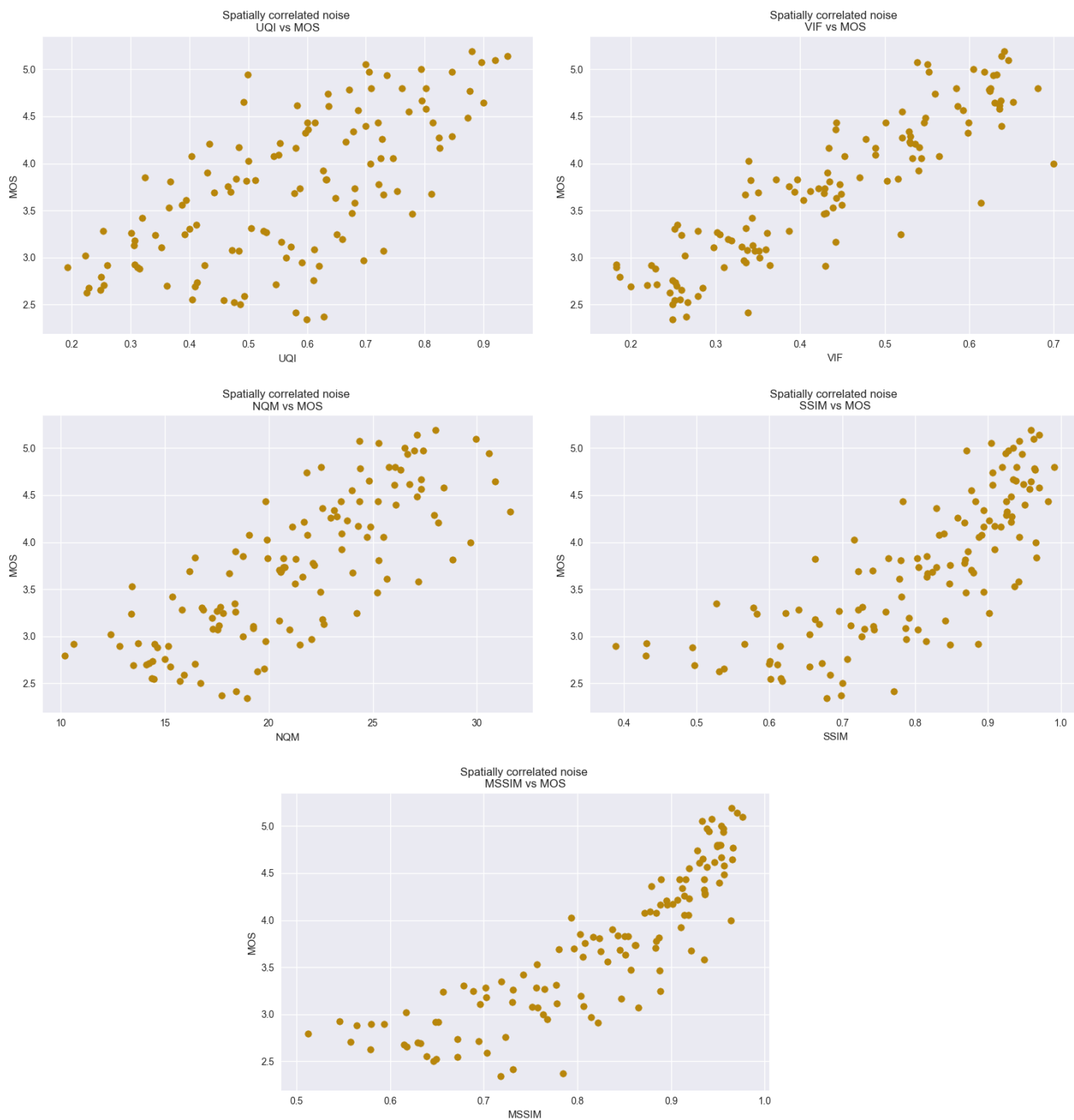


Figura A.8: Distorsión #3: Spatially correlated noise

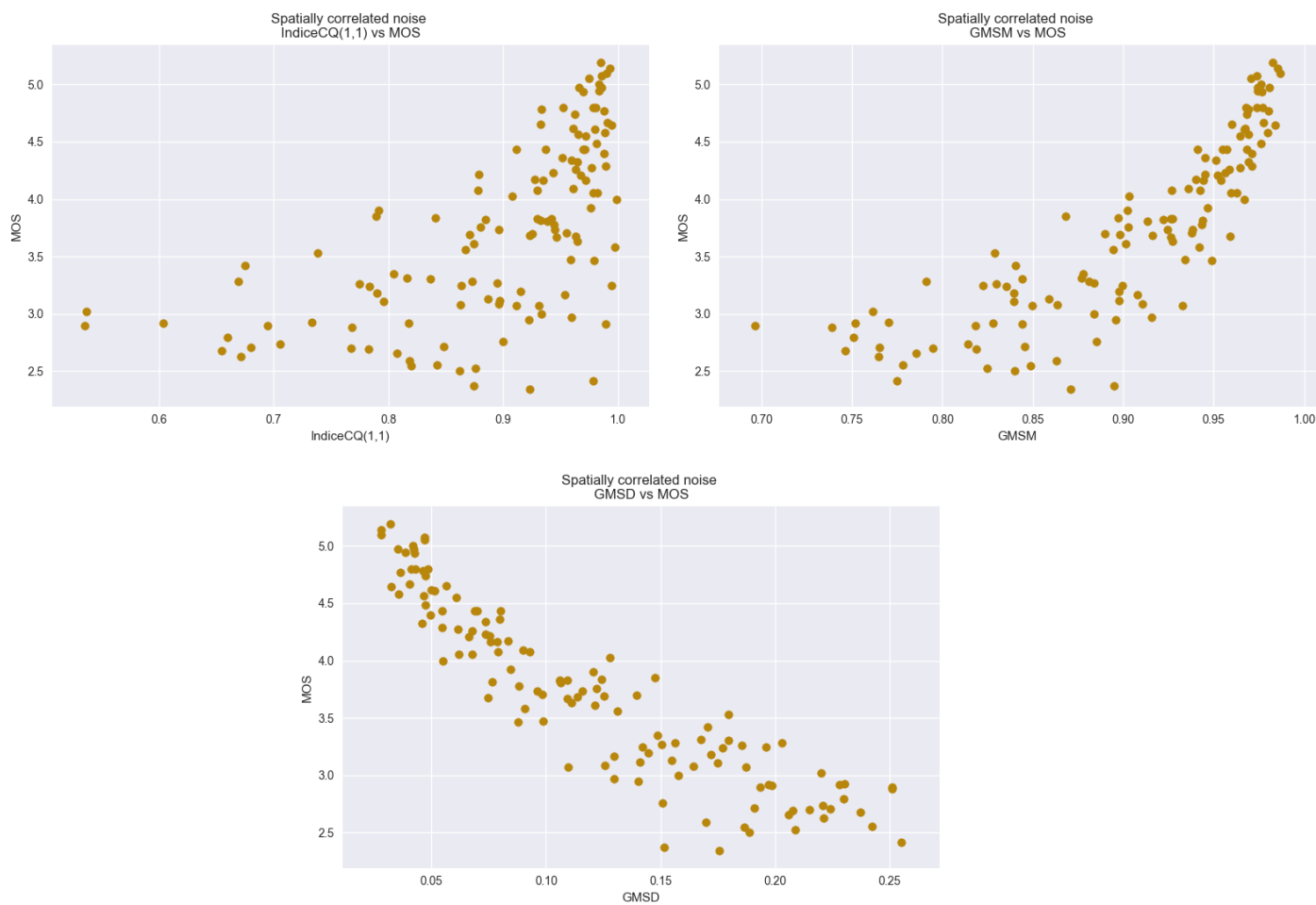


Figura A.9: Distorsión #3: Spatially correlated noise



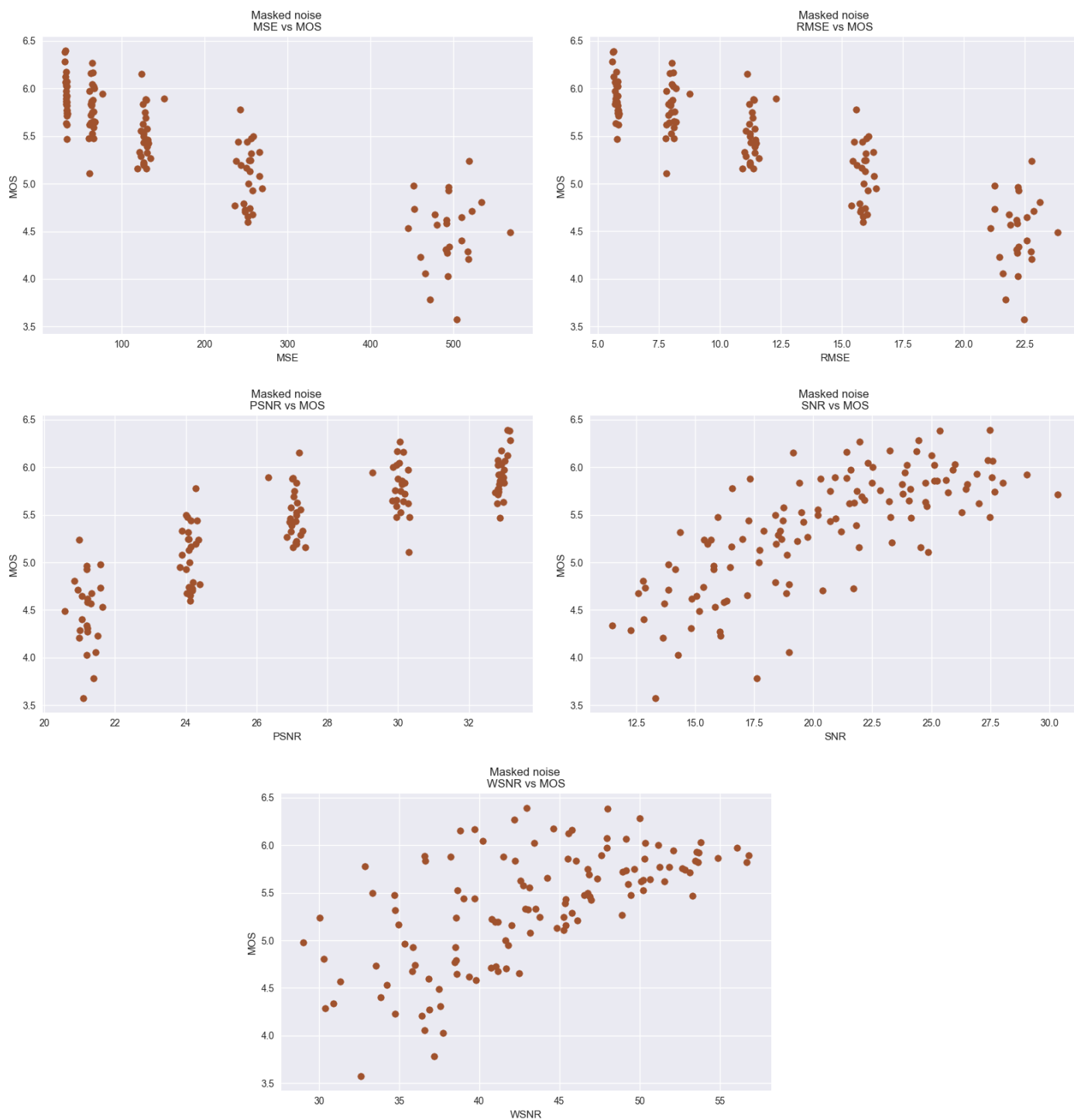


Figura A.10: Distorsión #4: Masked noise

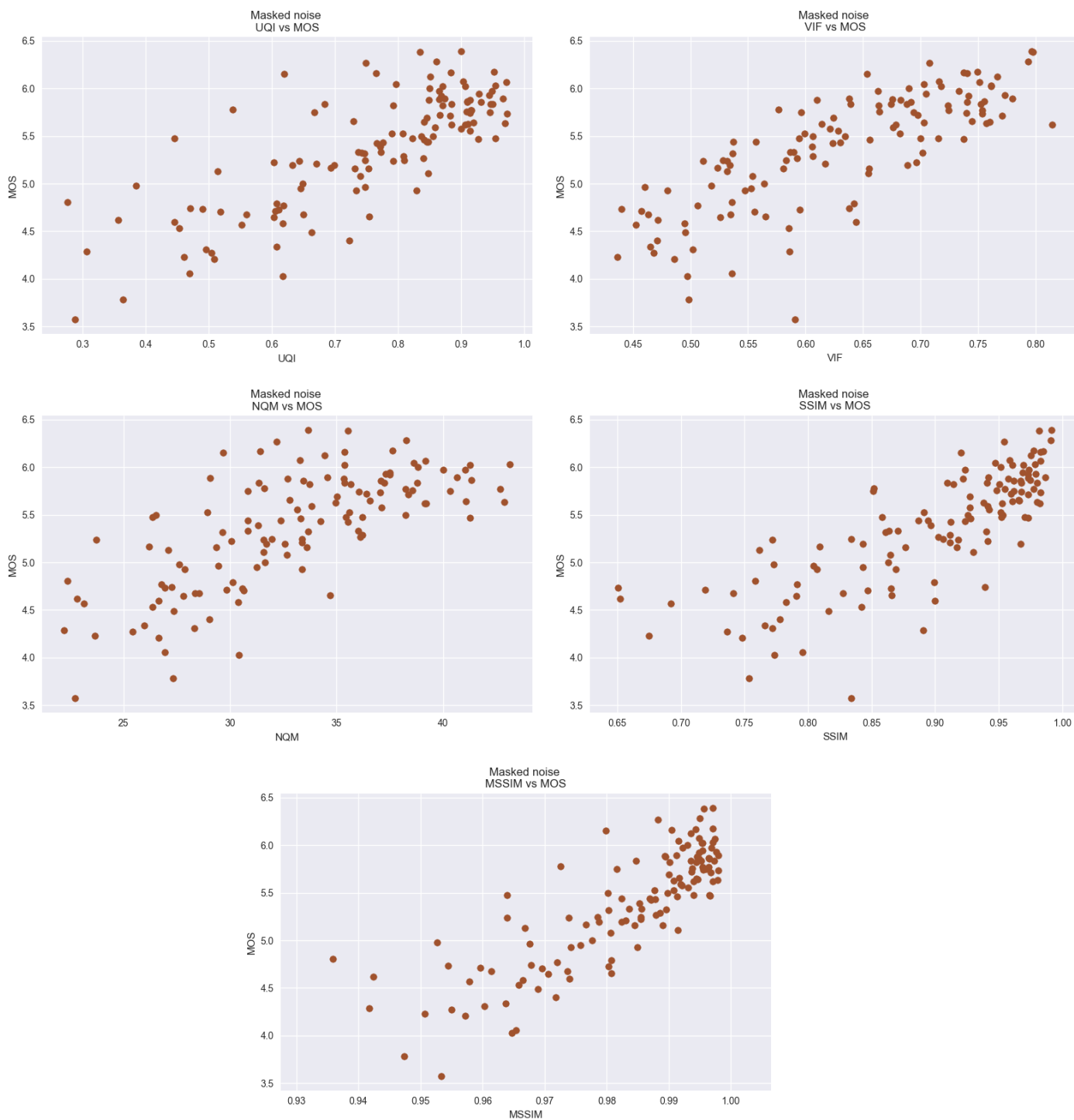


Figura A.11: Distorsión #4: Masked noise

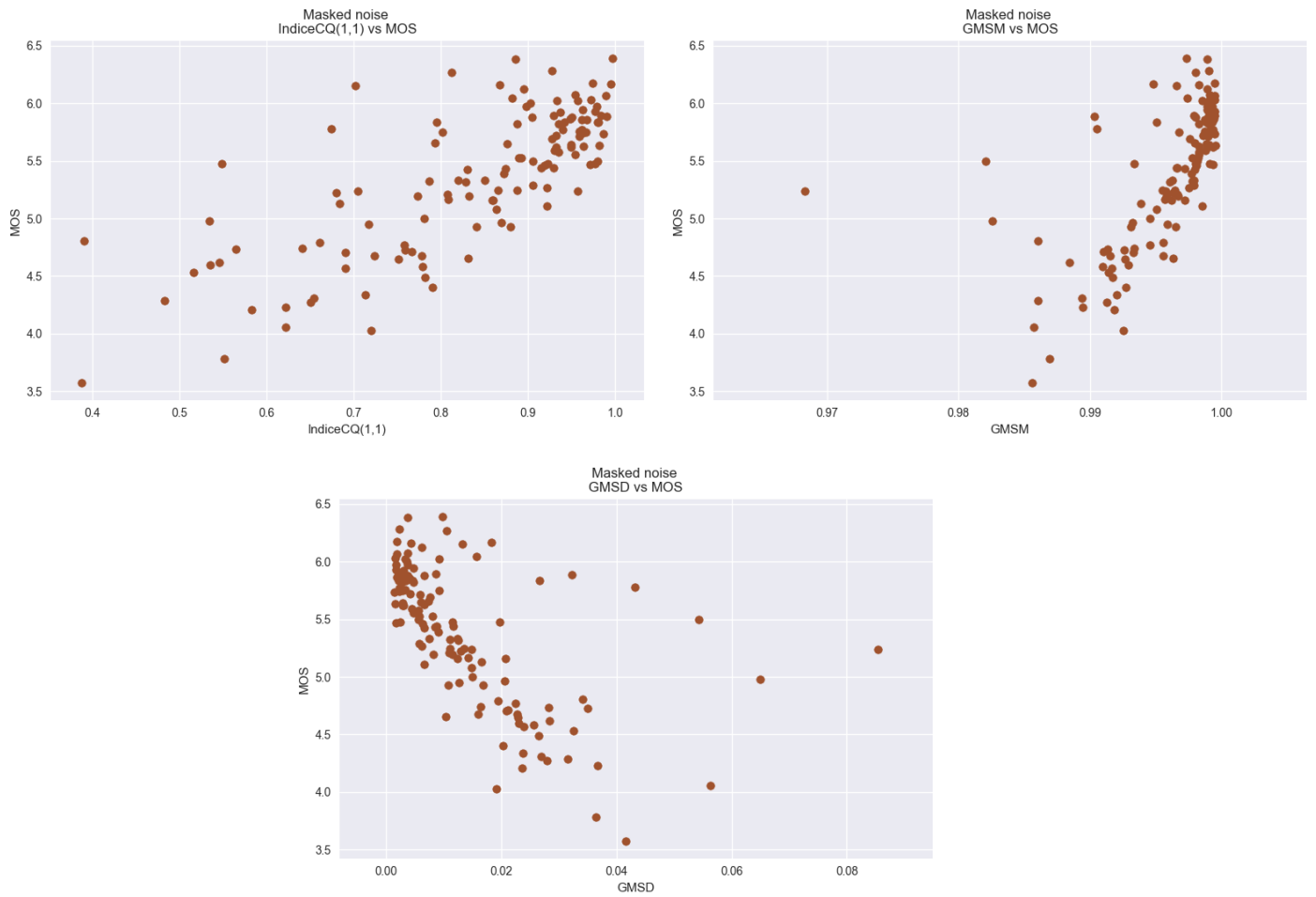


Figura A.12: Distorsión #4: Masked noise

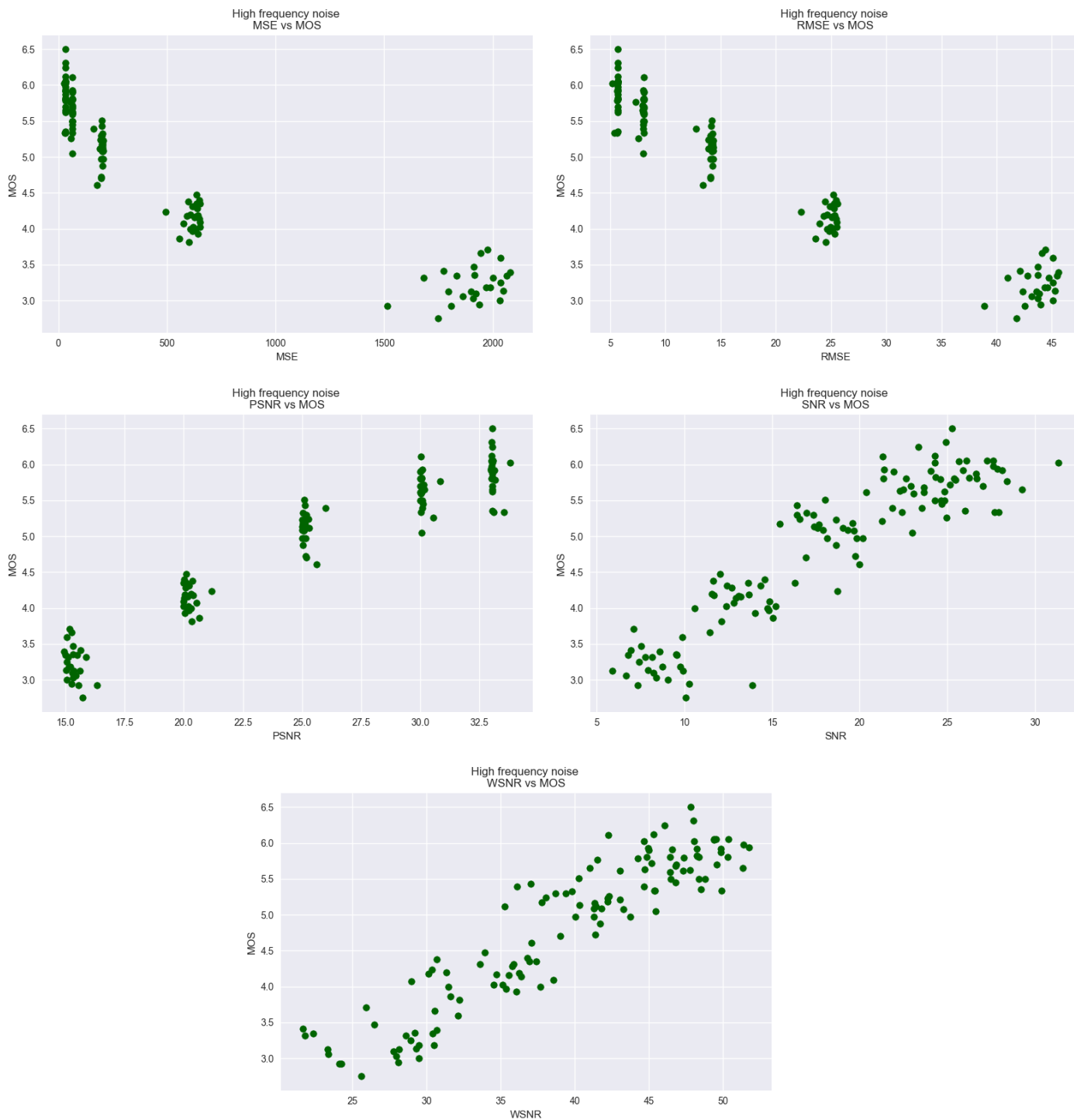


Figura A.13: Distorsión #5: High frequency noise

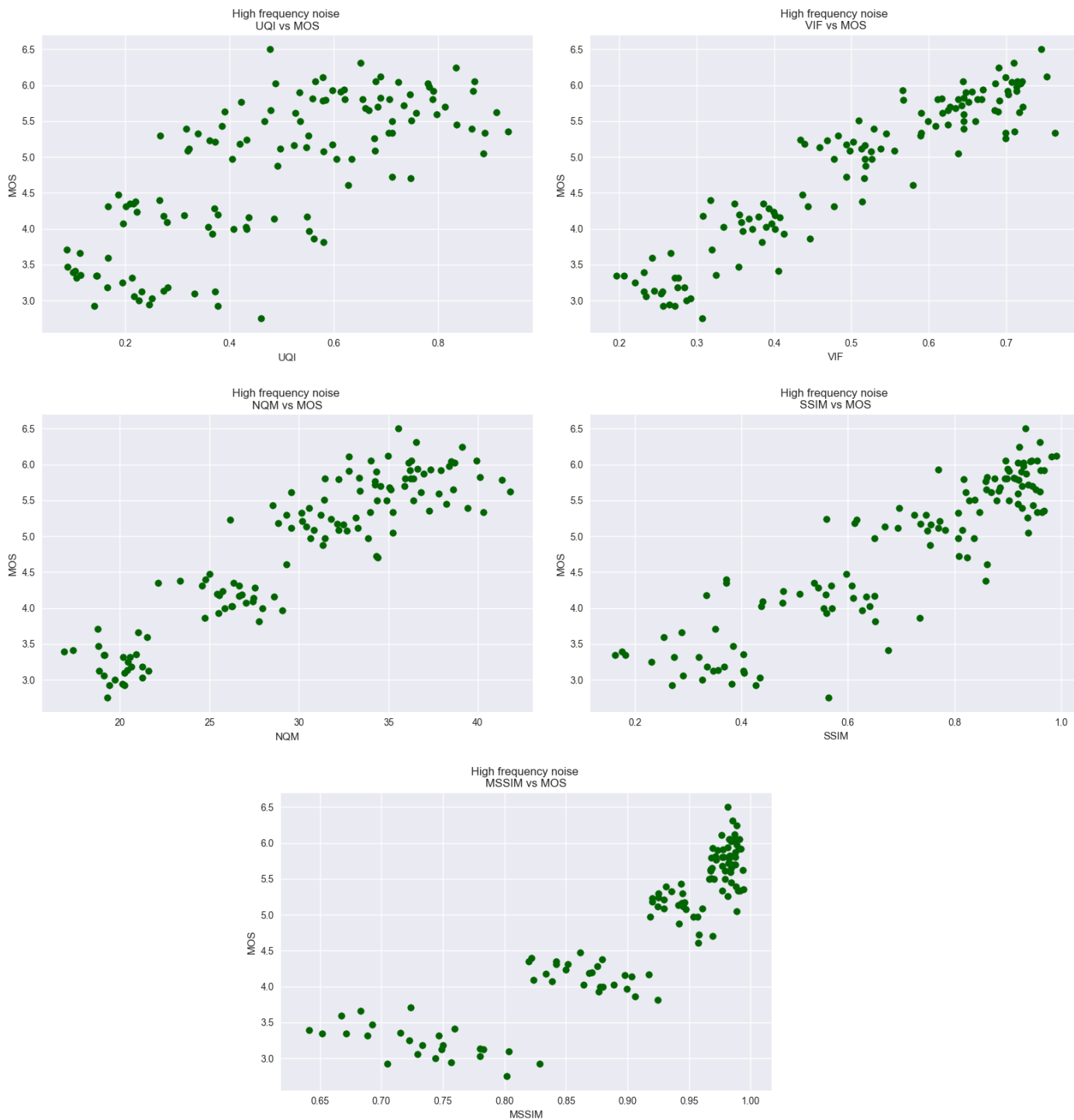


Figura A.14: Distorsión #5: High frequency noise

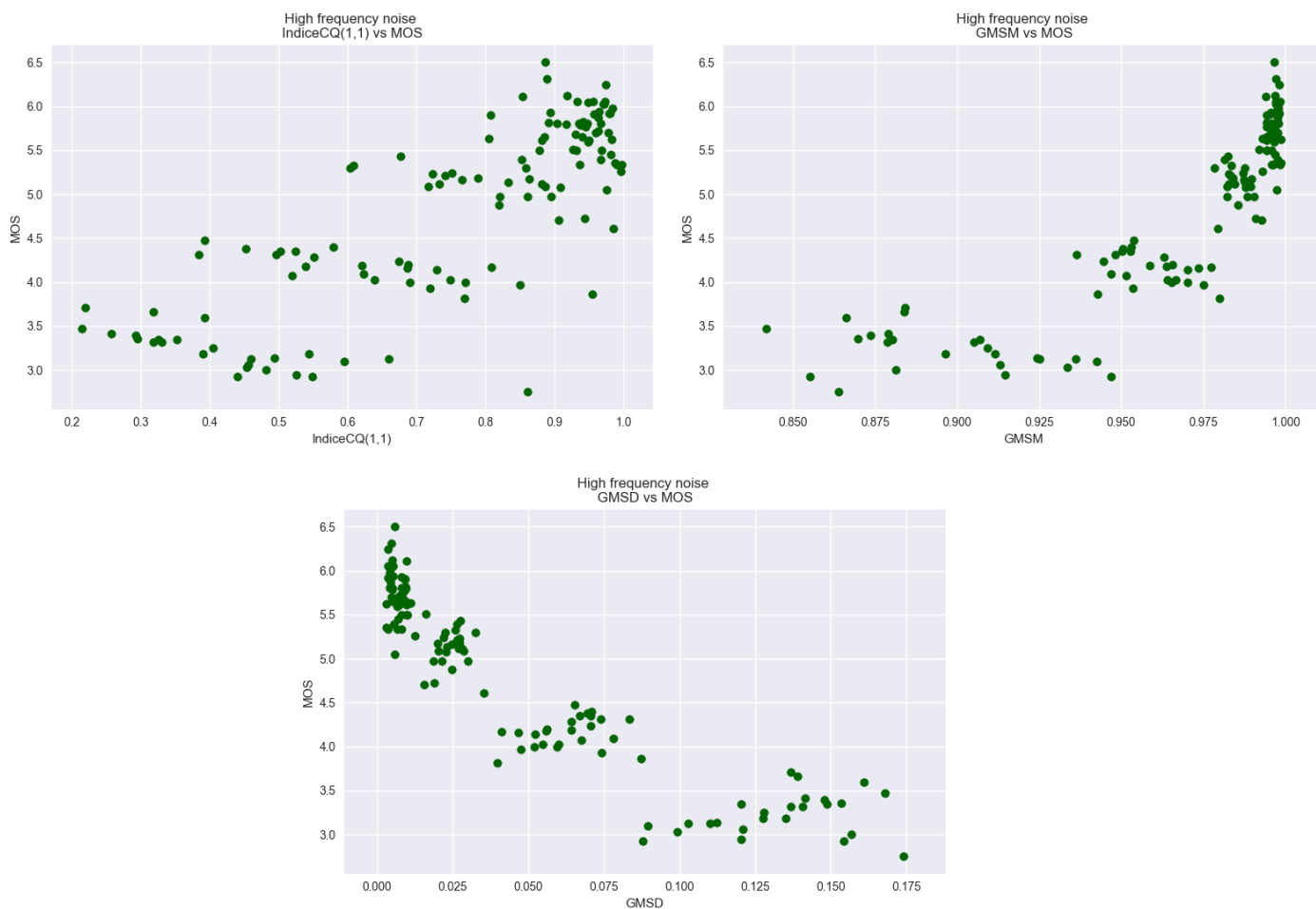


Figura A.15: Distorsión #5: High frequency noise

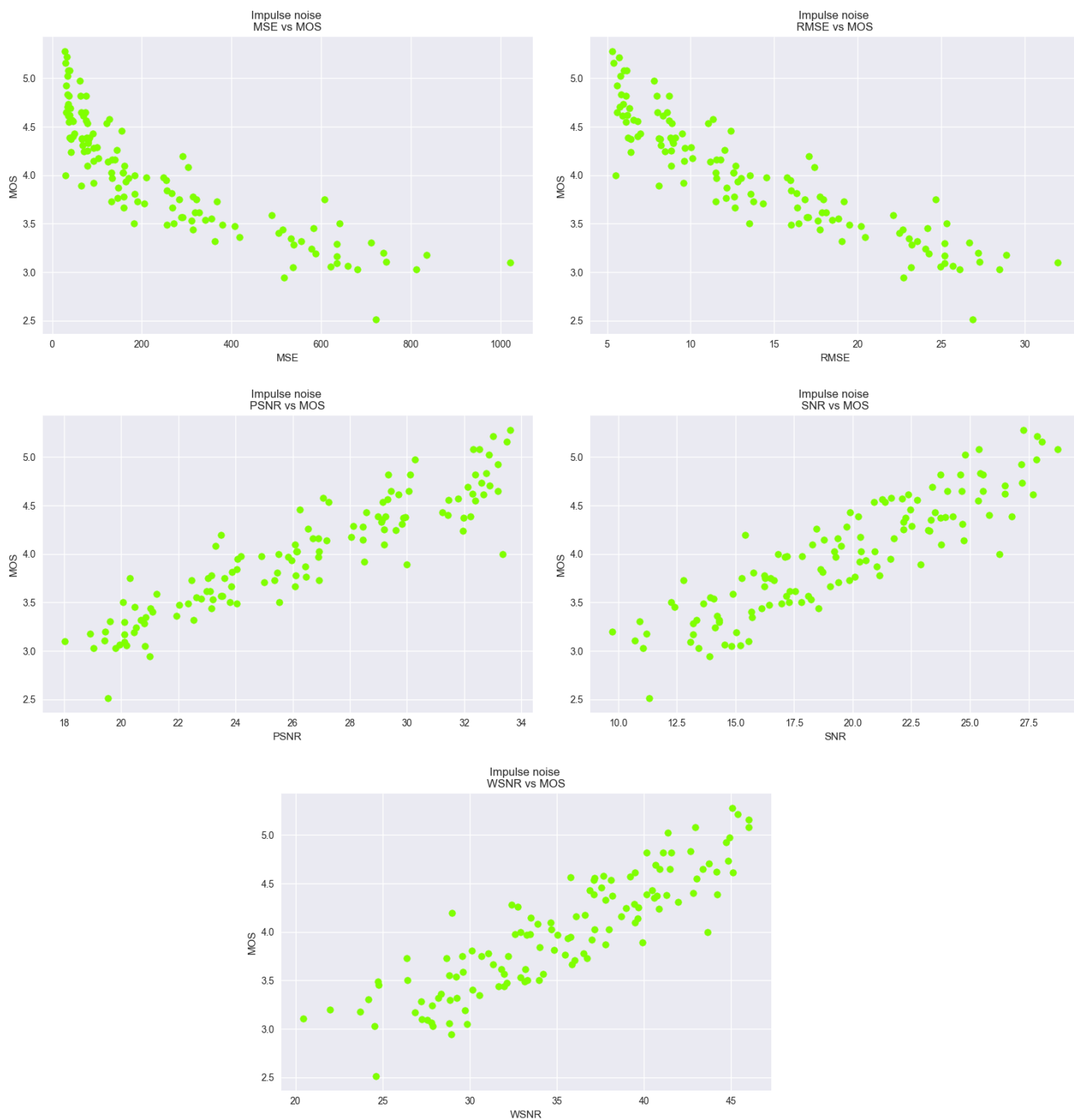


Figura A.16: Distorsión #6: Impulse noise

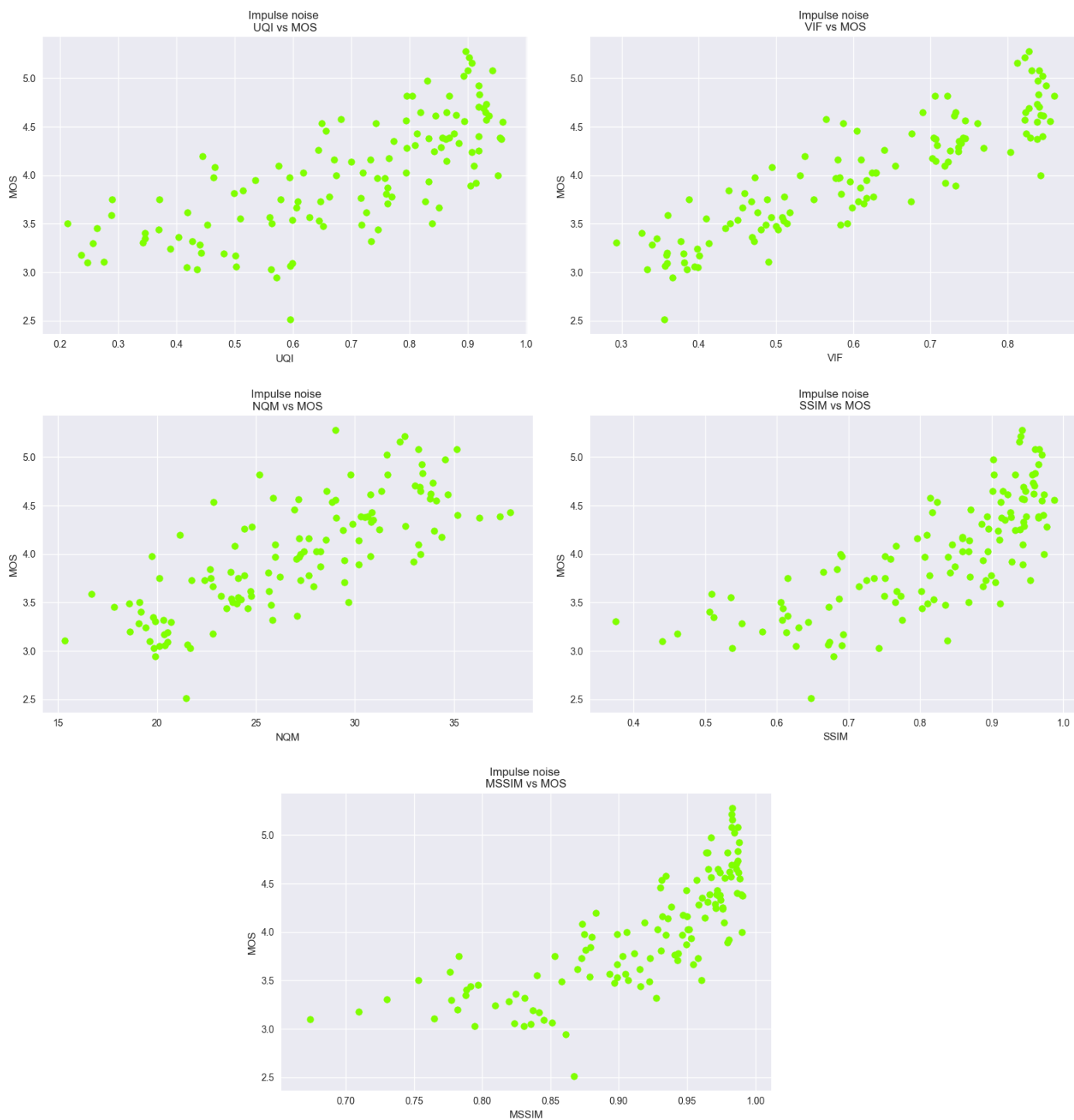


Figura A.17: Distorsión #6: Impulse noise



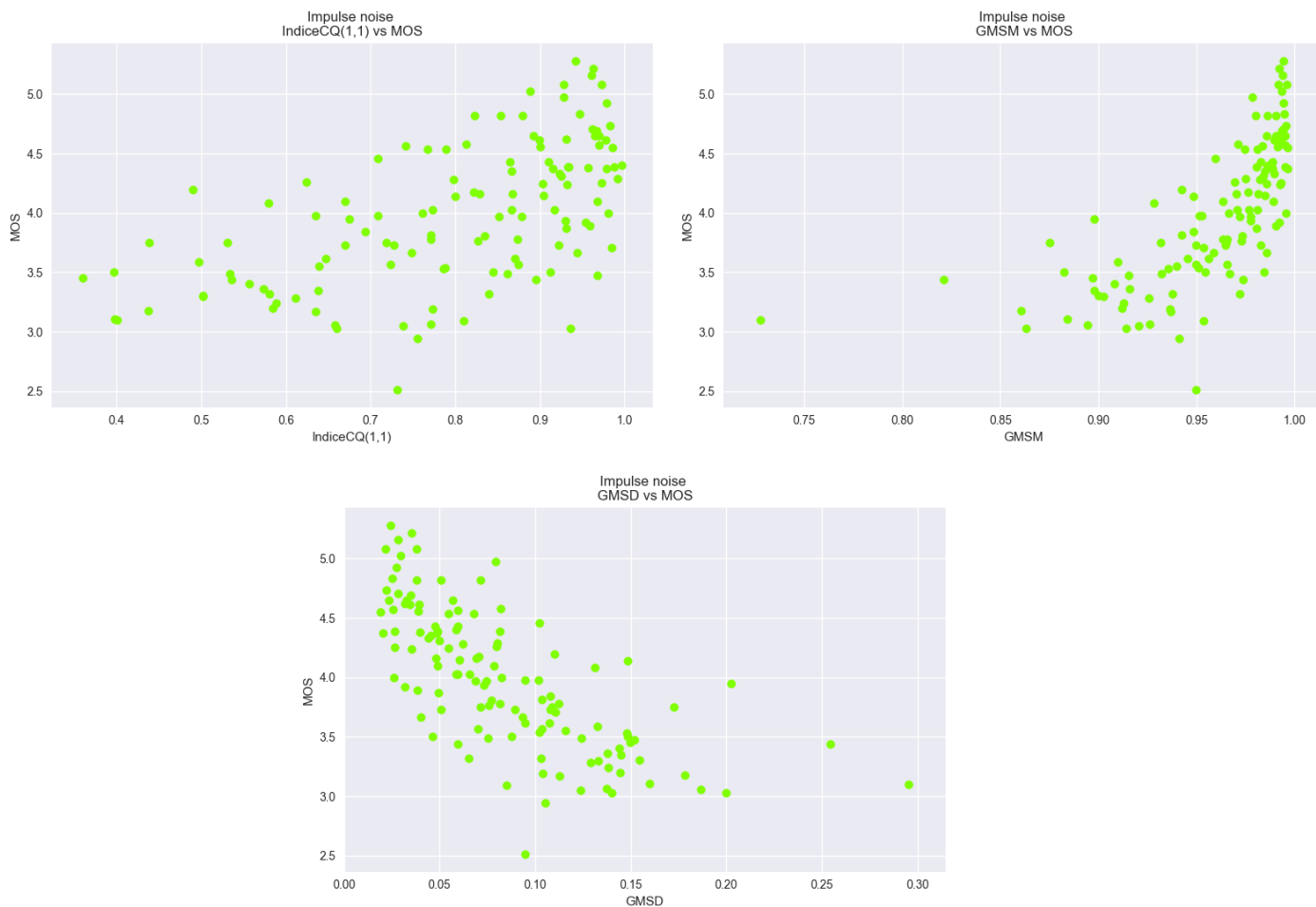


Figura A.18: Distorsión #6: Impulse noise

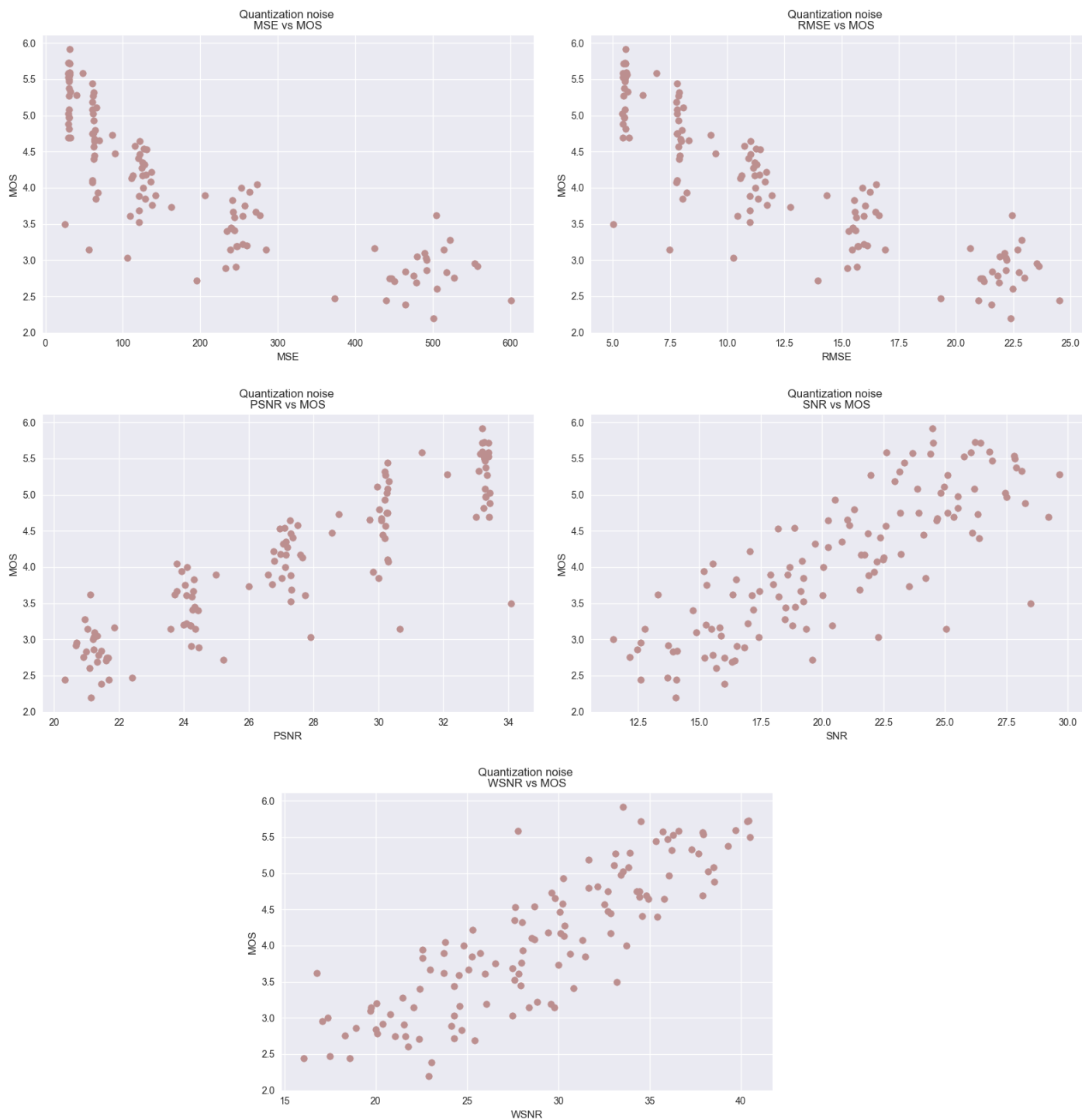


Figura A.19: Distorsión #7: Quantization noise

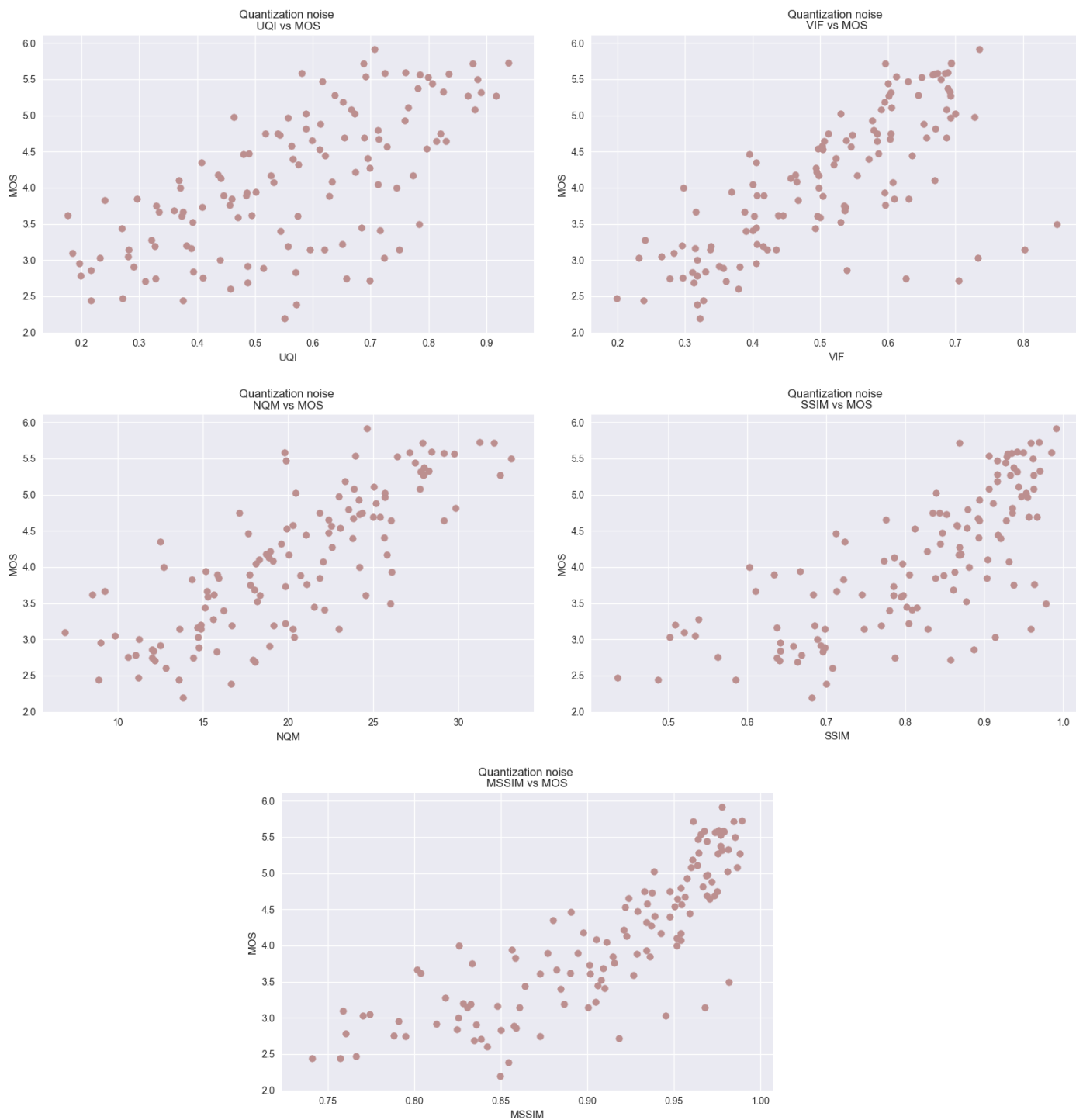


Figura A.20: Distorsión #7: Quantization noise

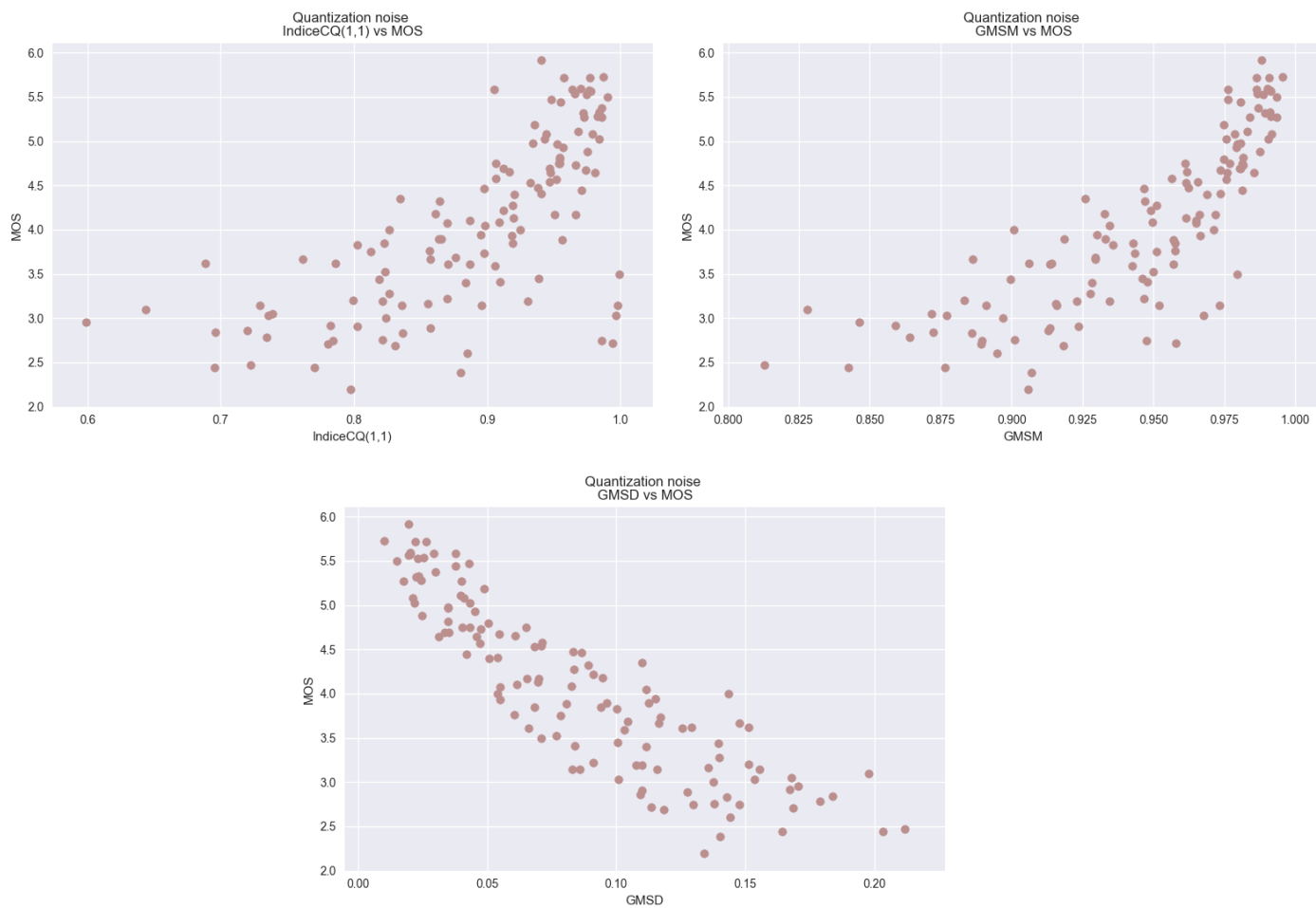


Figura A.21: Distorsión #7: Quantization noise

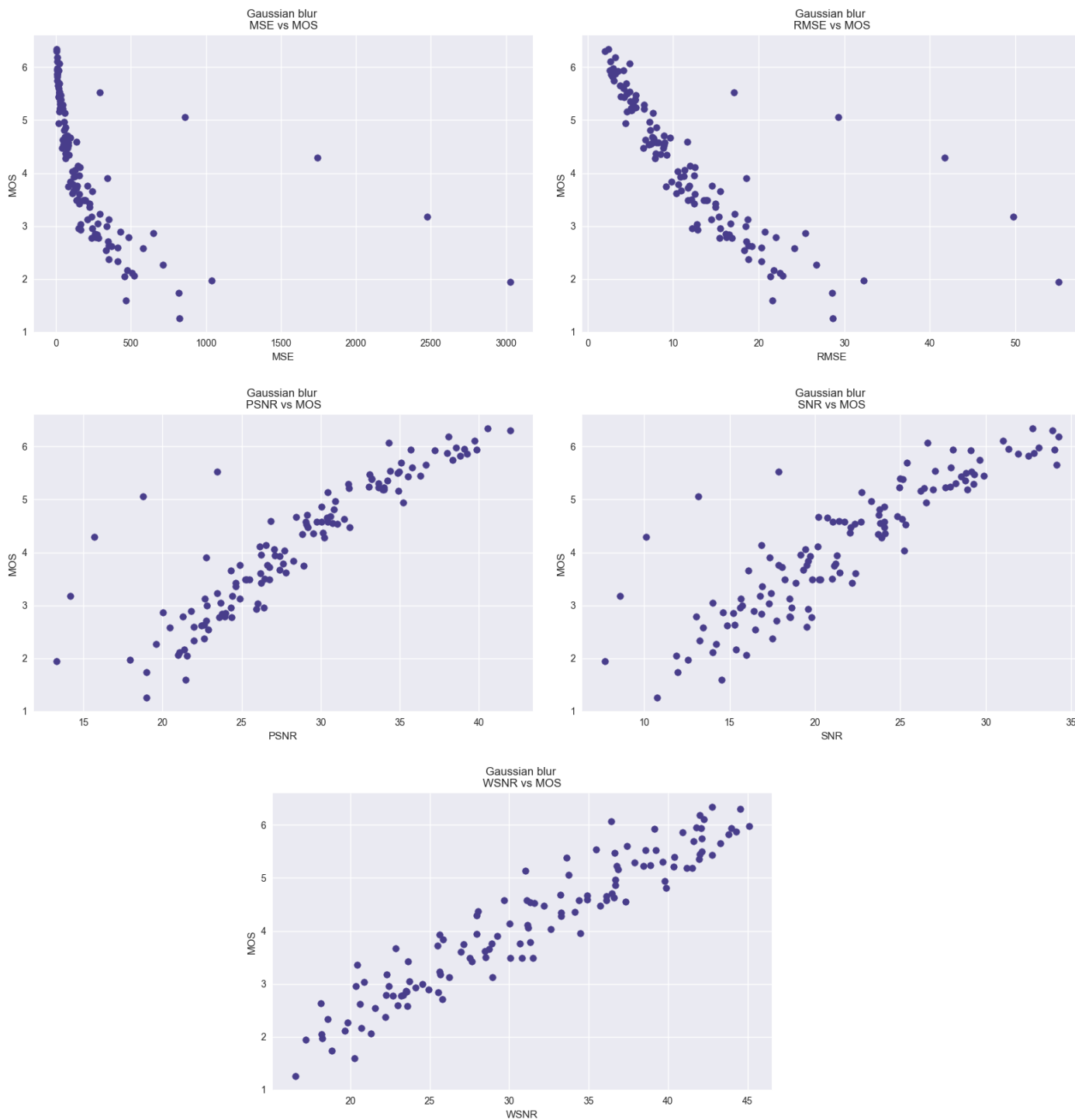


Figura A.22: Distorsión #8: Gaussian blur

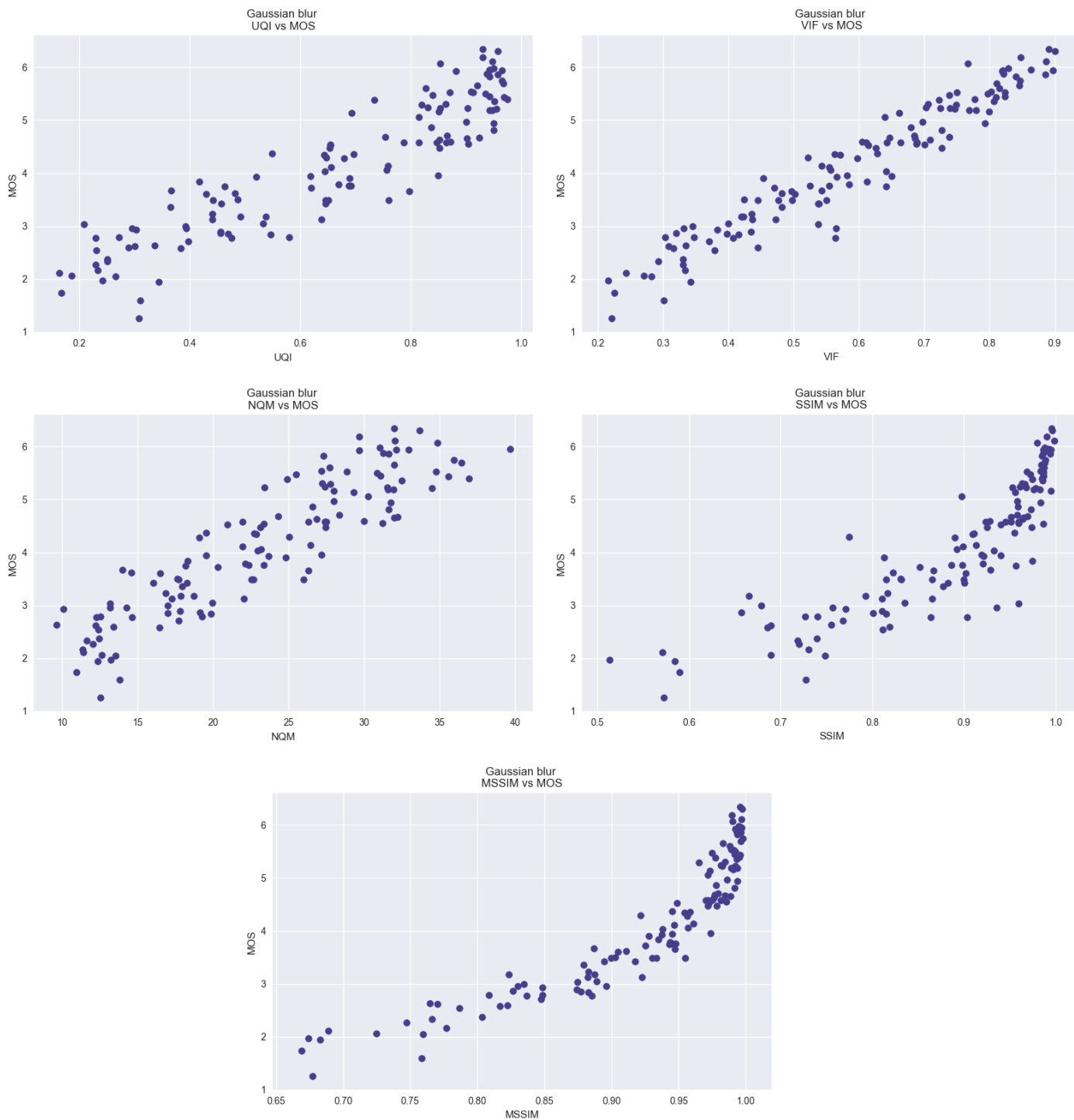


Figura A.23: Distorsión #8: Gaussian blur

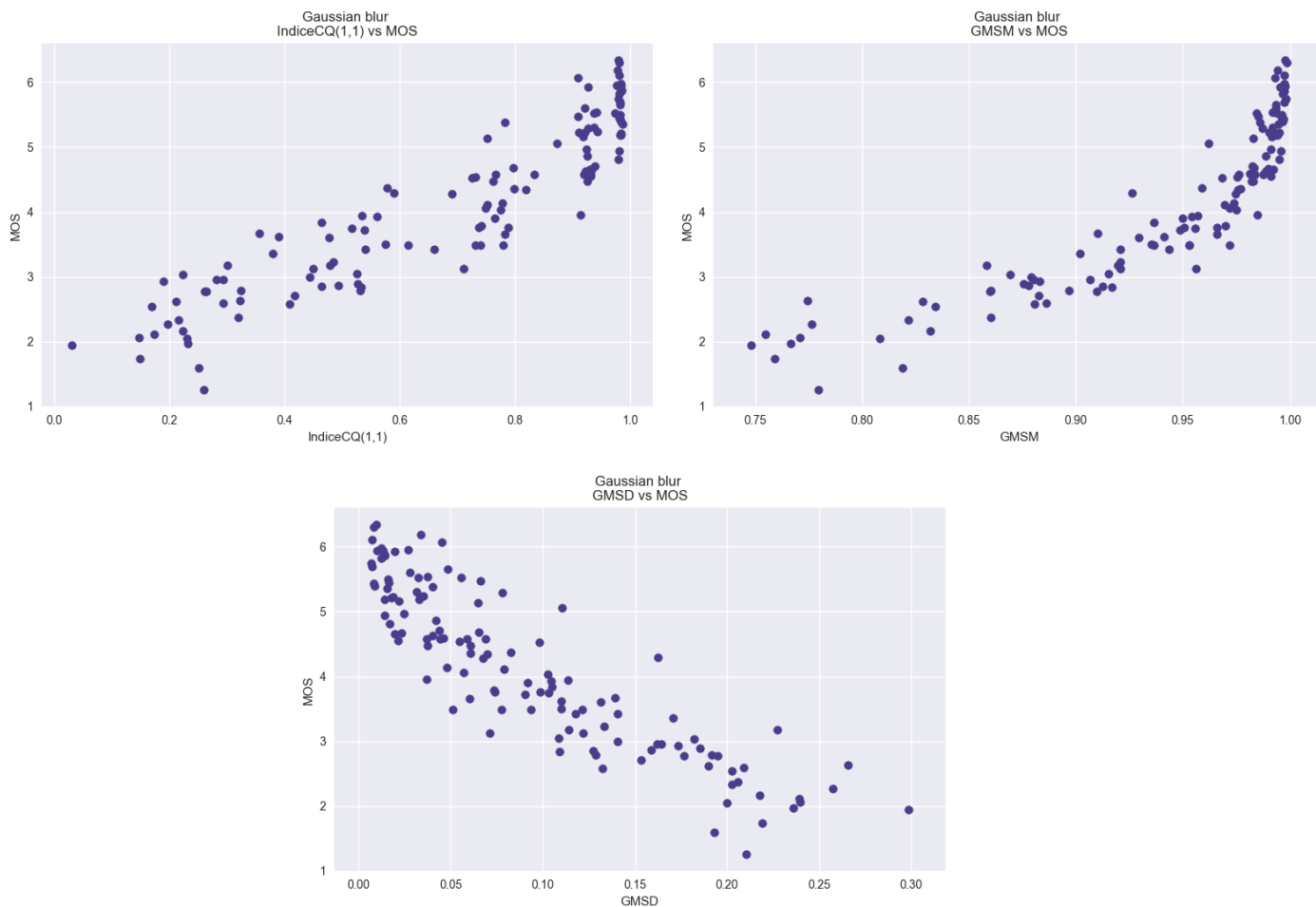


Figura A.24: Distorsión #8: Gaussian blur

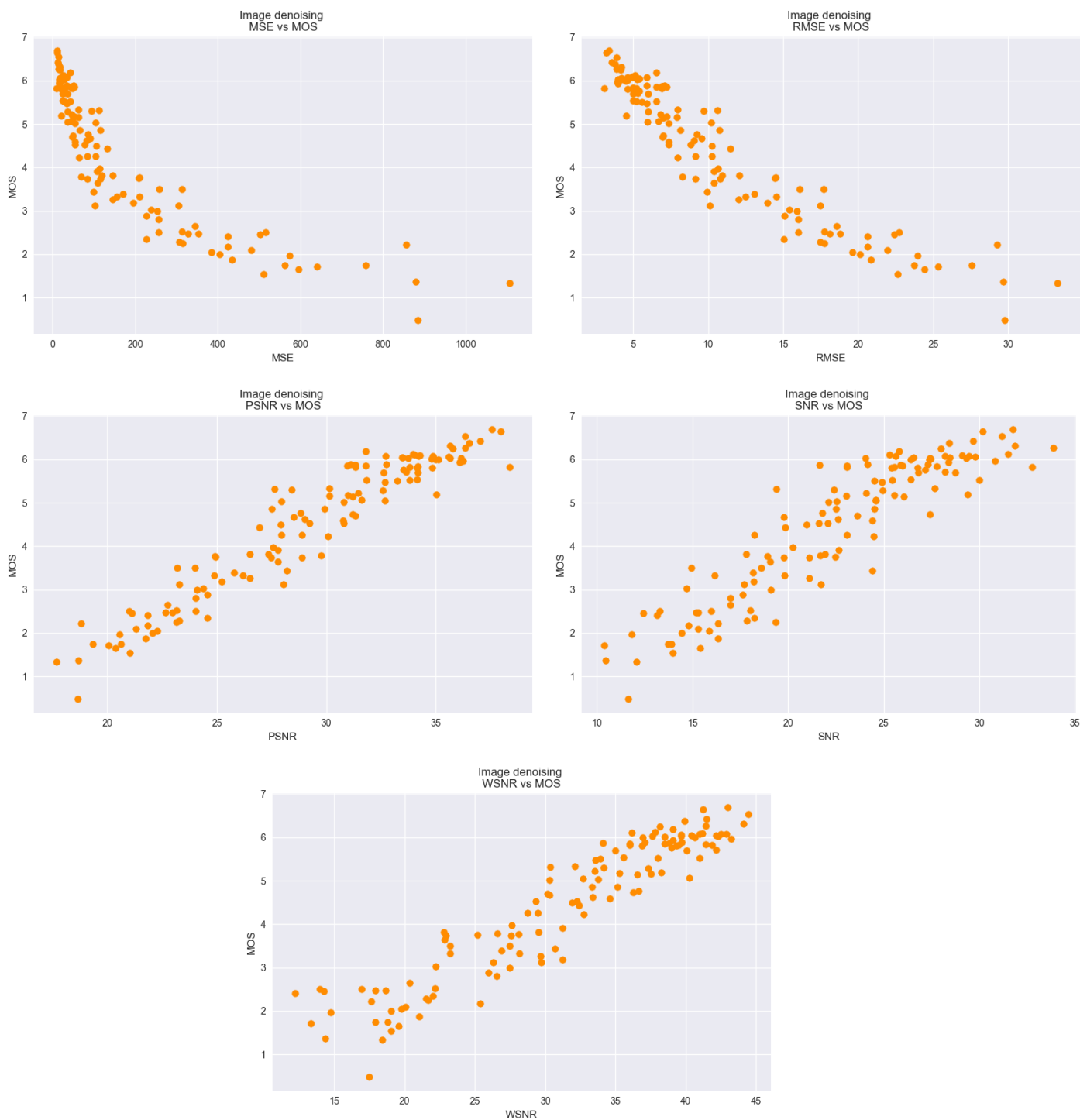


Figura A.25: Distorsión #9: Image denoising



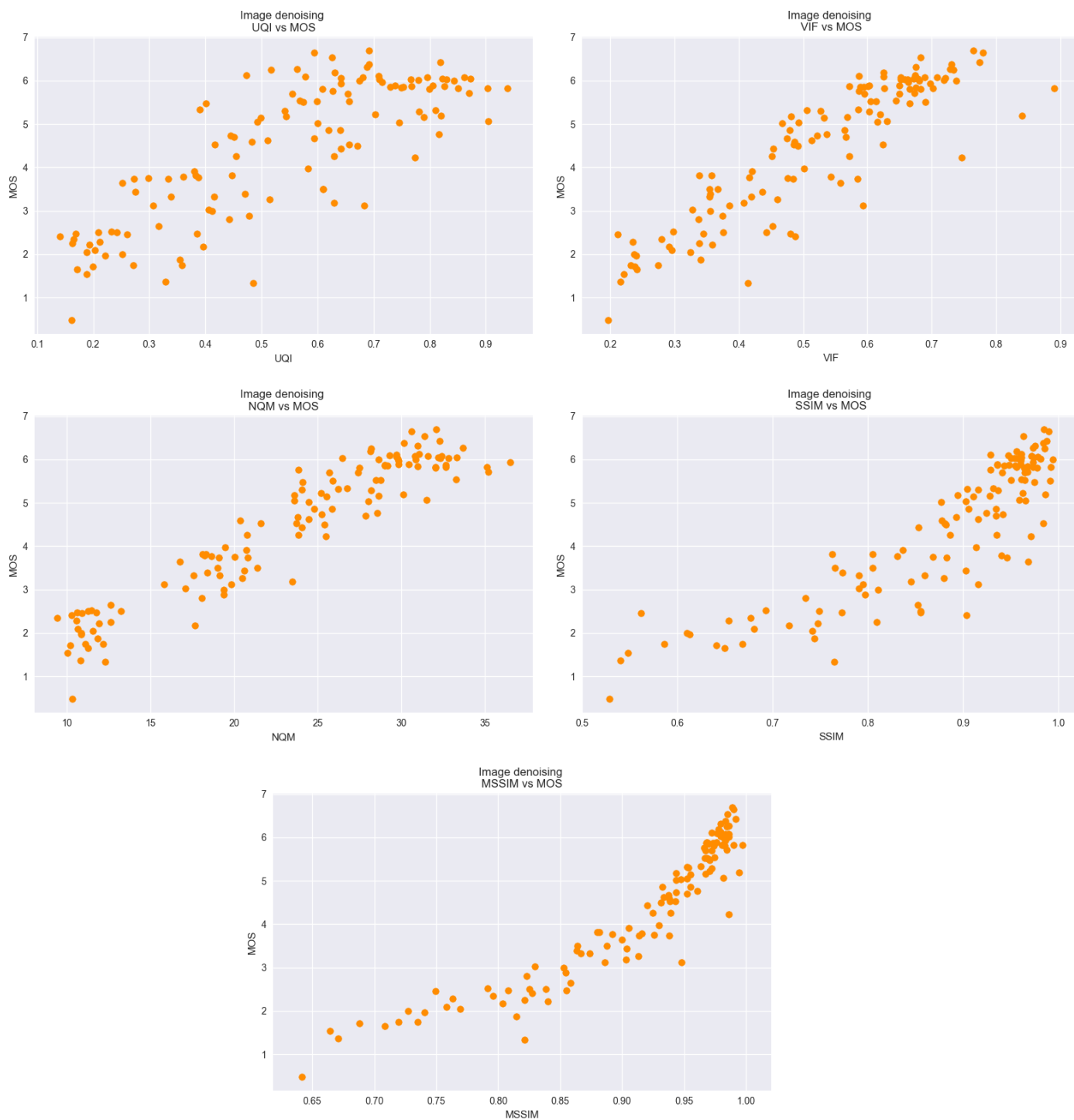


Figura A.26: Distorsión #9: Image denoising

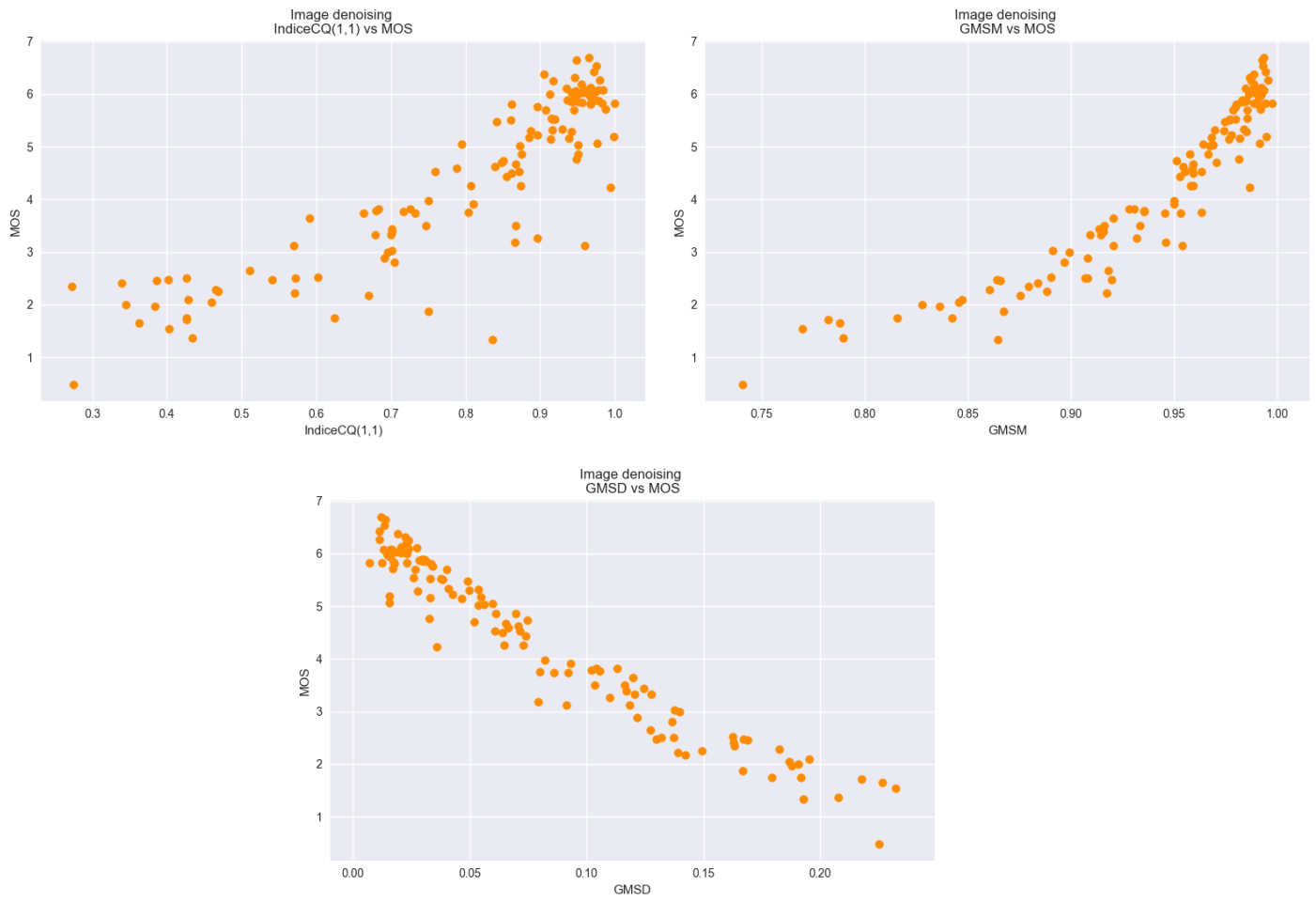


Figura A.27: Distorsión #9: Image denoising

---

## Apéndice B

### Diagramas de dispersión de las distorsiones #10 a #17

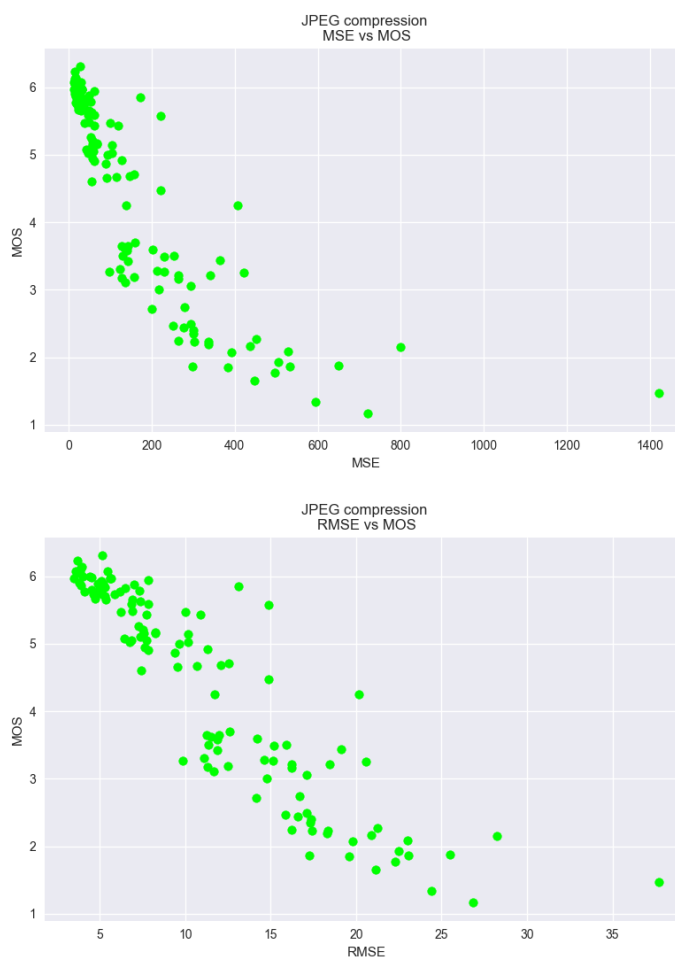


Figura B.1: Distorsión #10: JPEG compression

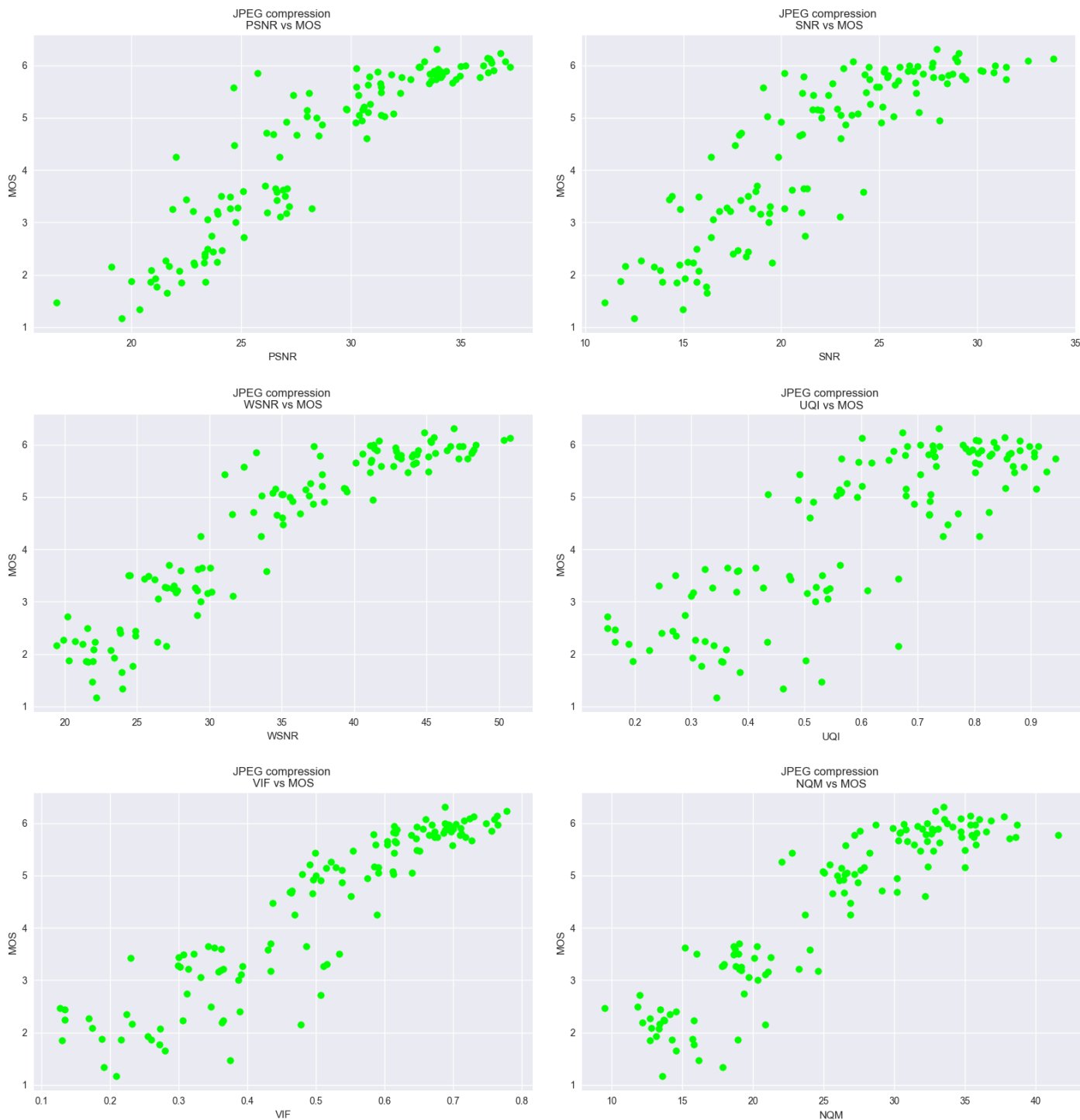


Figura B.2: Distorsión #10: JPEG compression

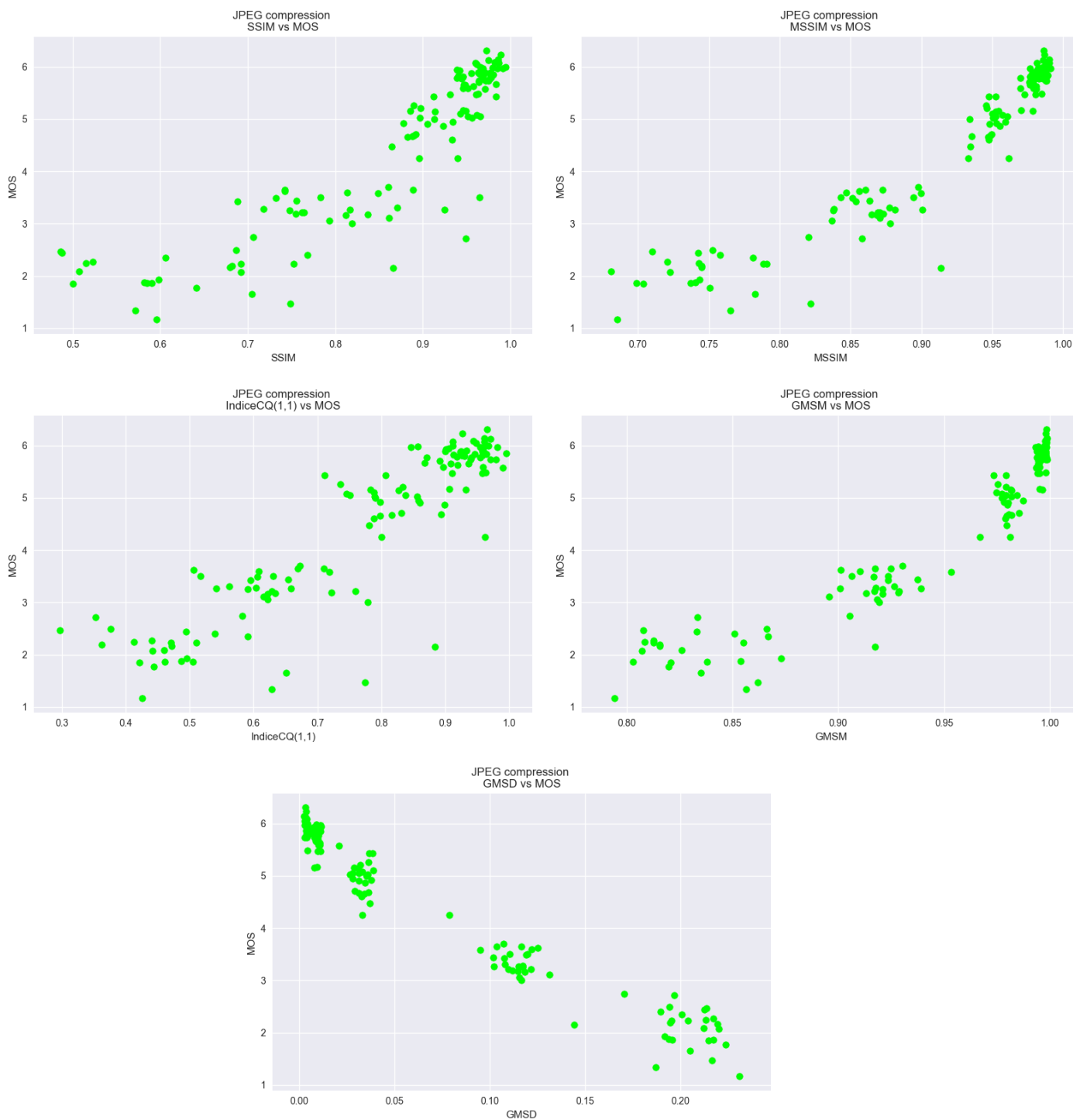


Figura B.3: Distorsión #10: JPEG compression

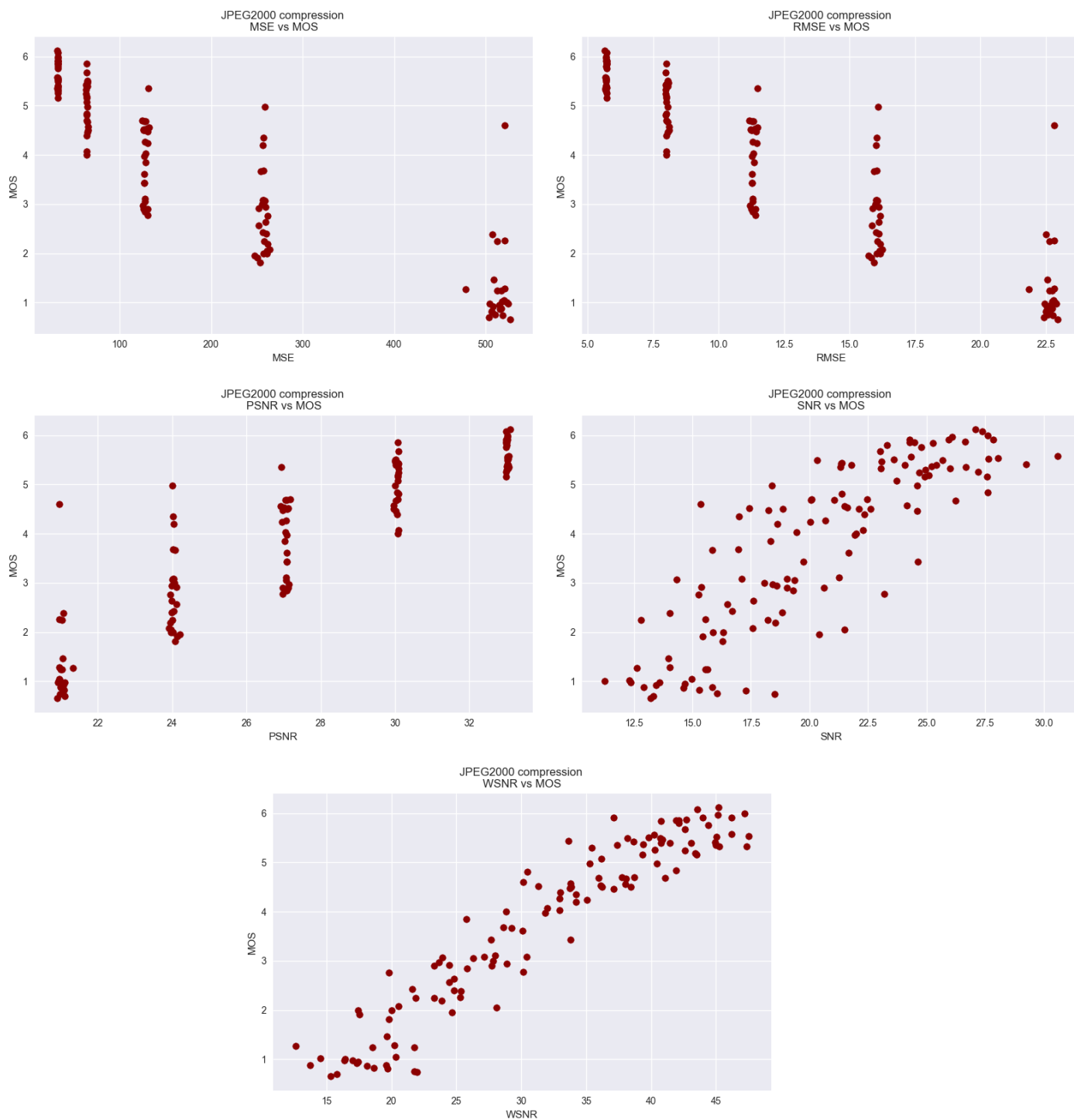


Figura B.4: Distorsión #11: JPEG2000 compression

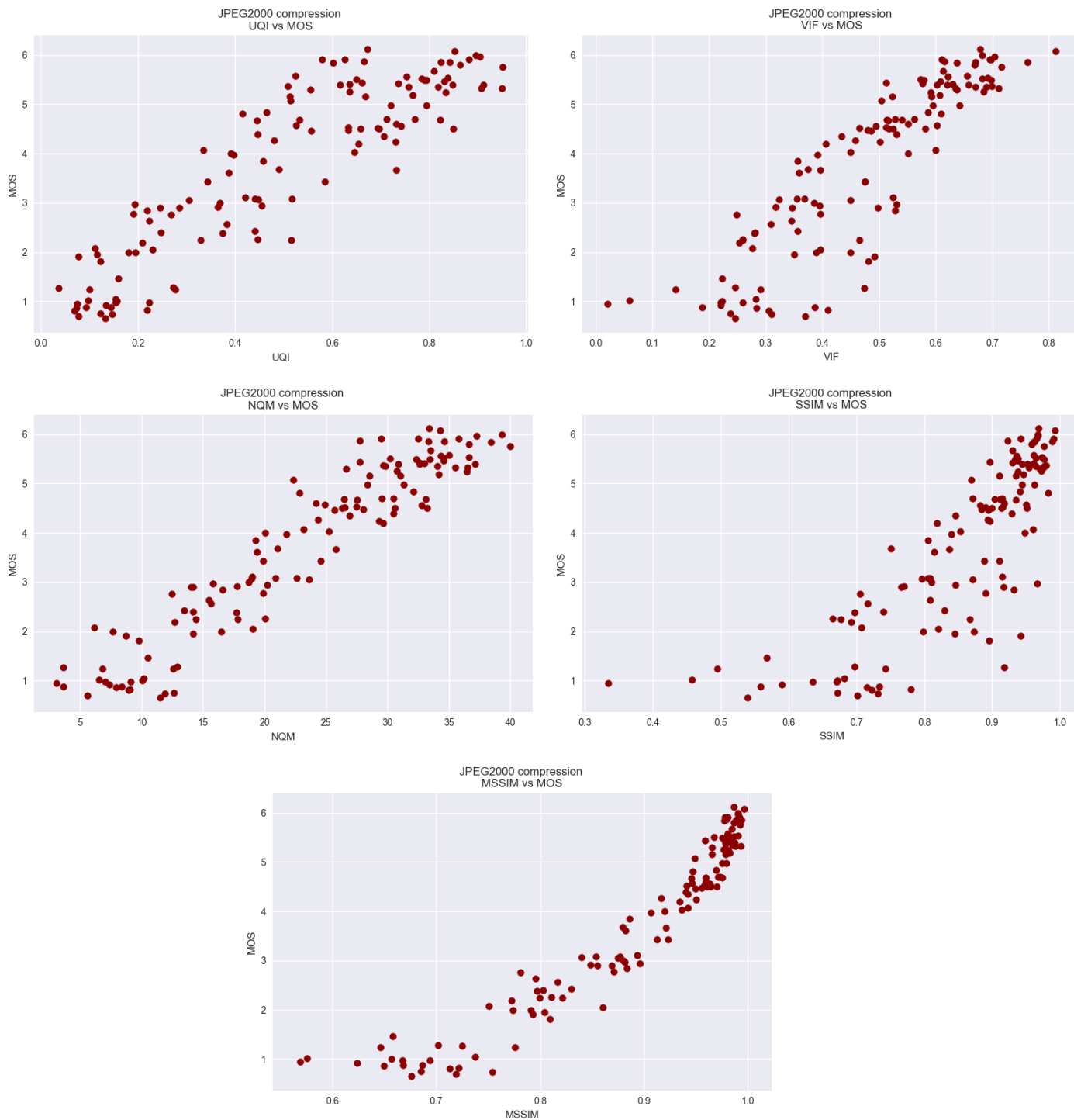


Figura B.5: Distorsión #11: JPEG2000 compression

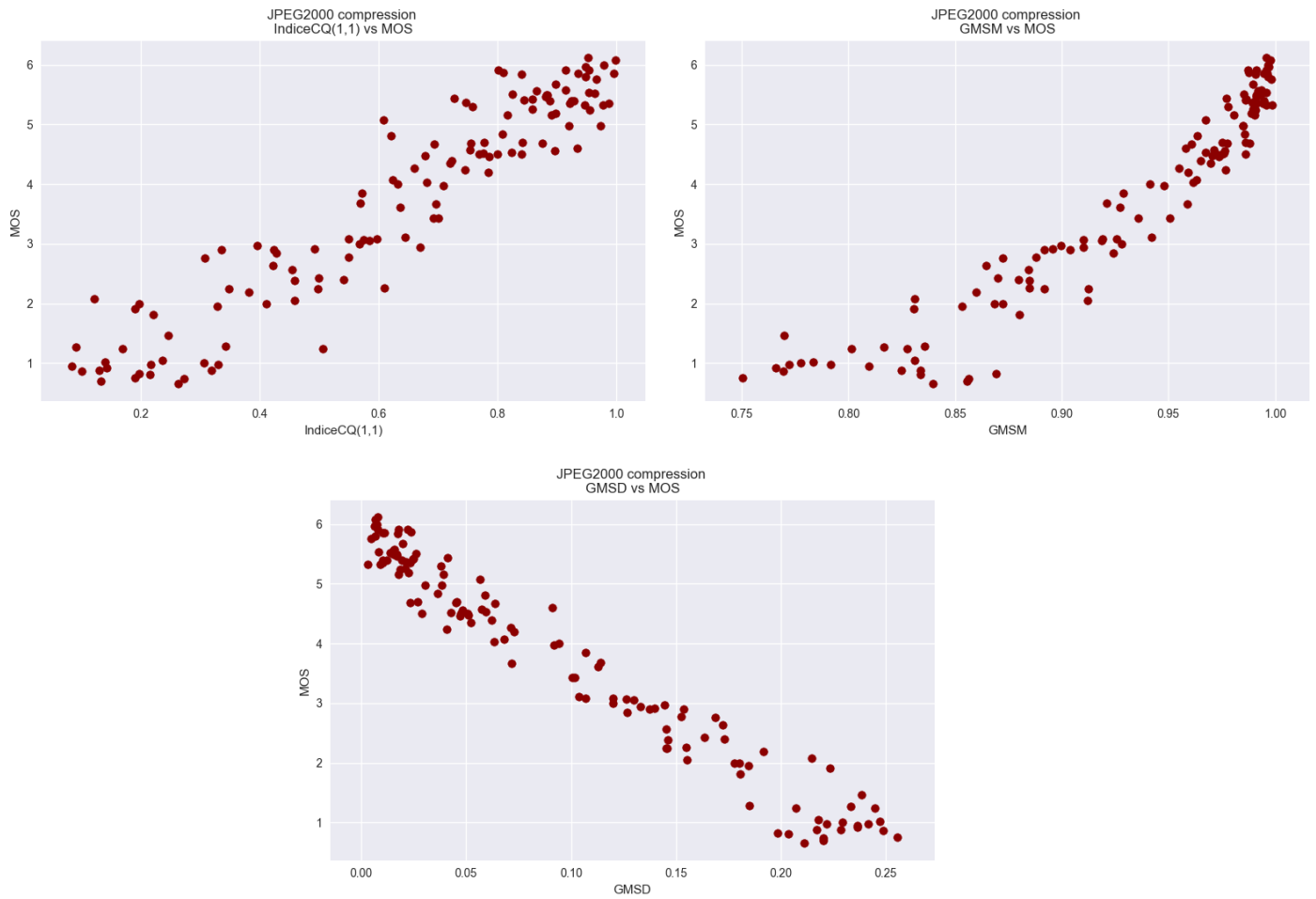


Figura B.6: Distorsión #11: JPEG2000 compression



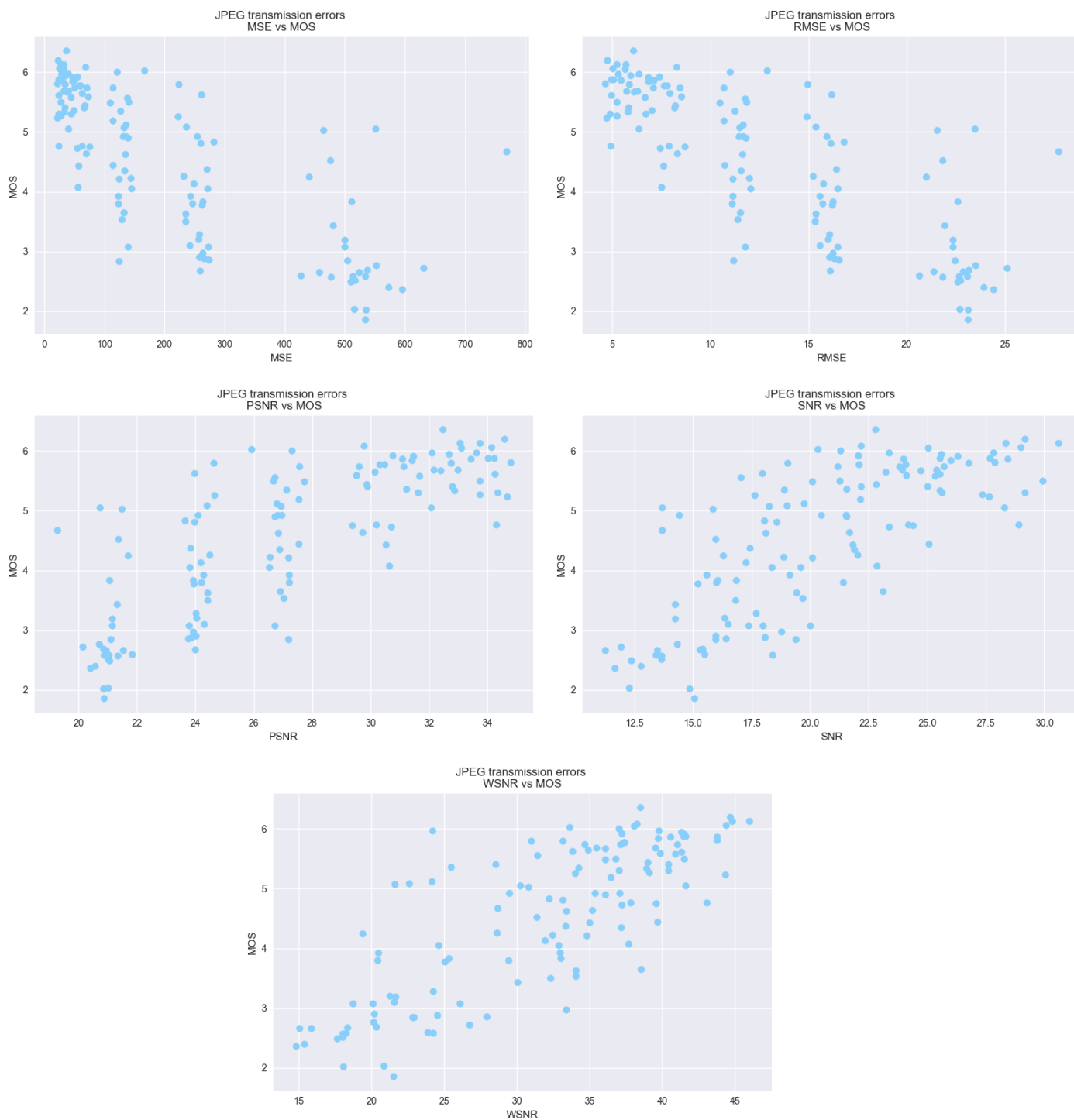


Figura B.7: Distorsión #12: JPEG transmission errors

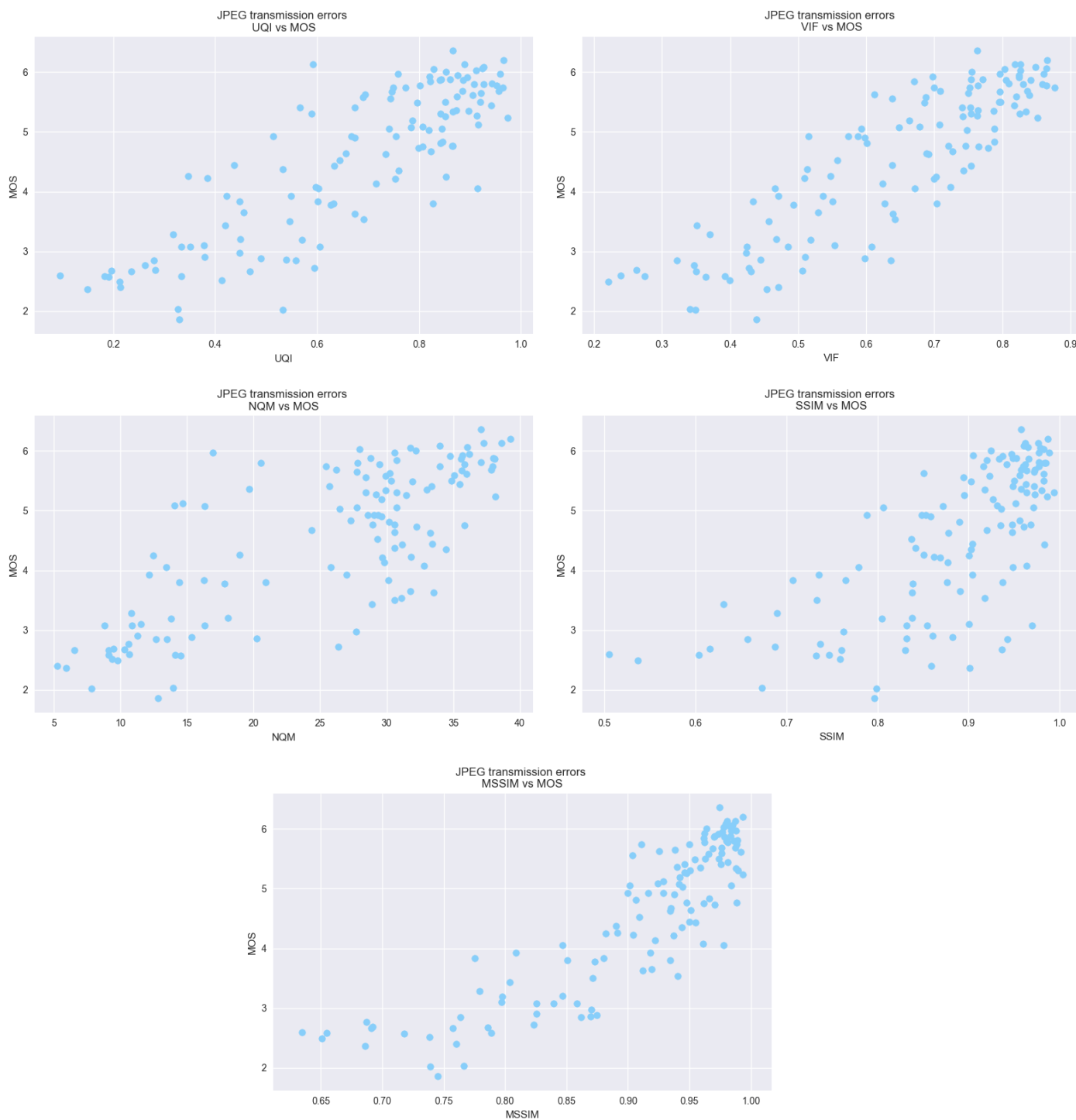


Figura B.8: Distorsión #12: JPEG transmission errors

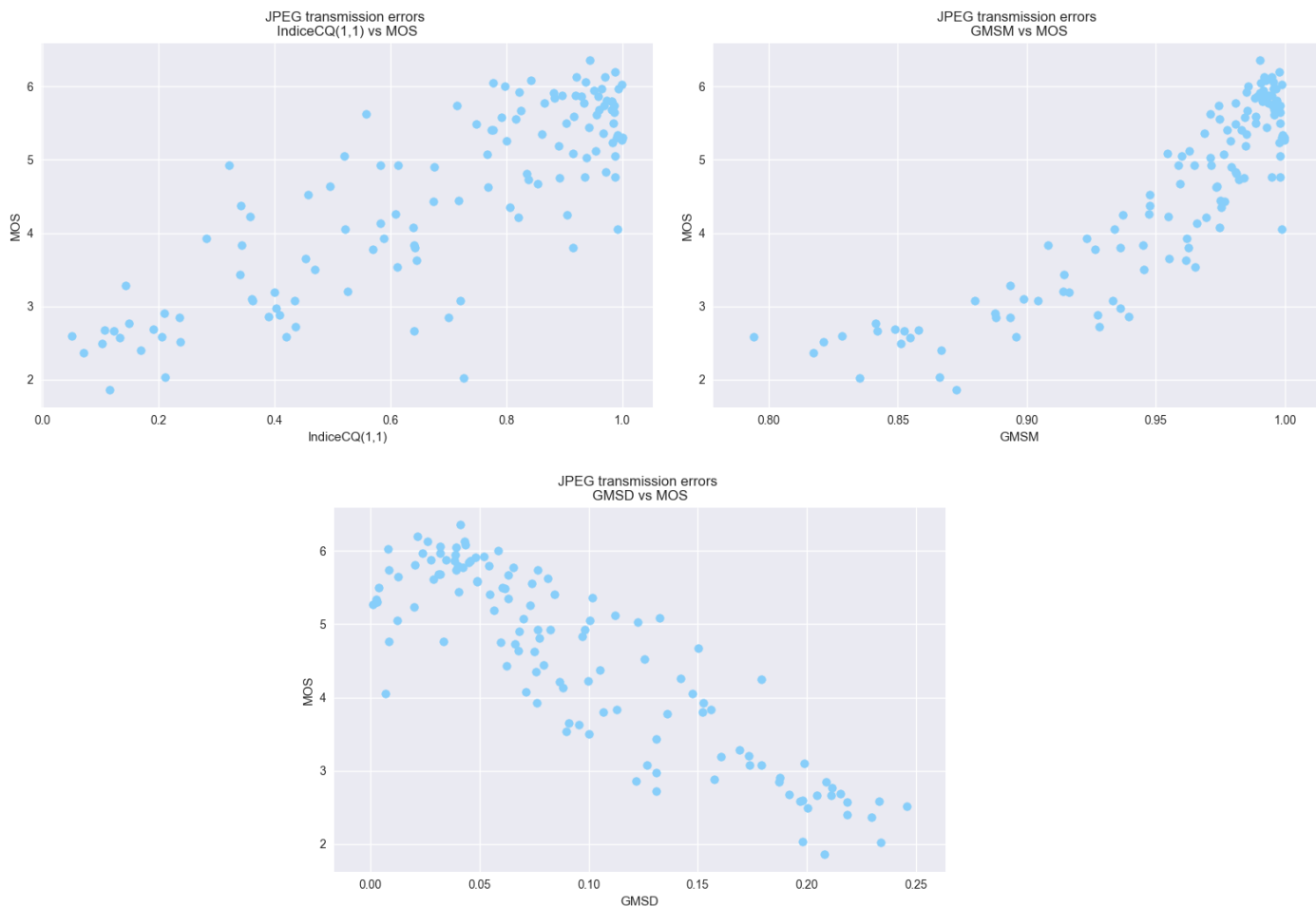


Figura B.9: Distorsión #12: JPEG transmission errors

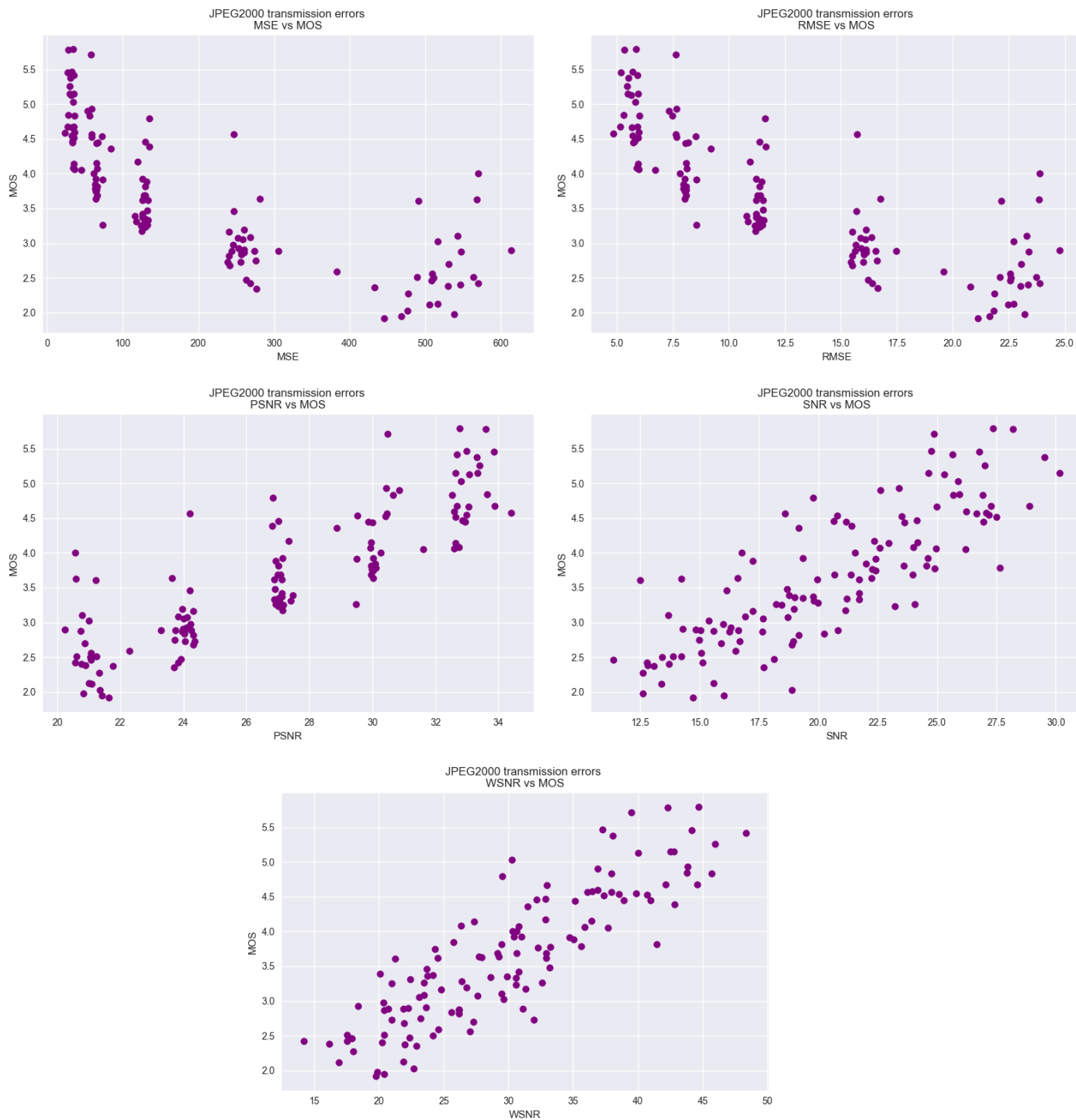


Figura B.10: Distorsión #13: JPEG2000 transmission errors

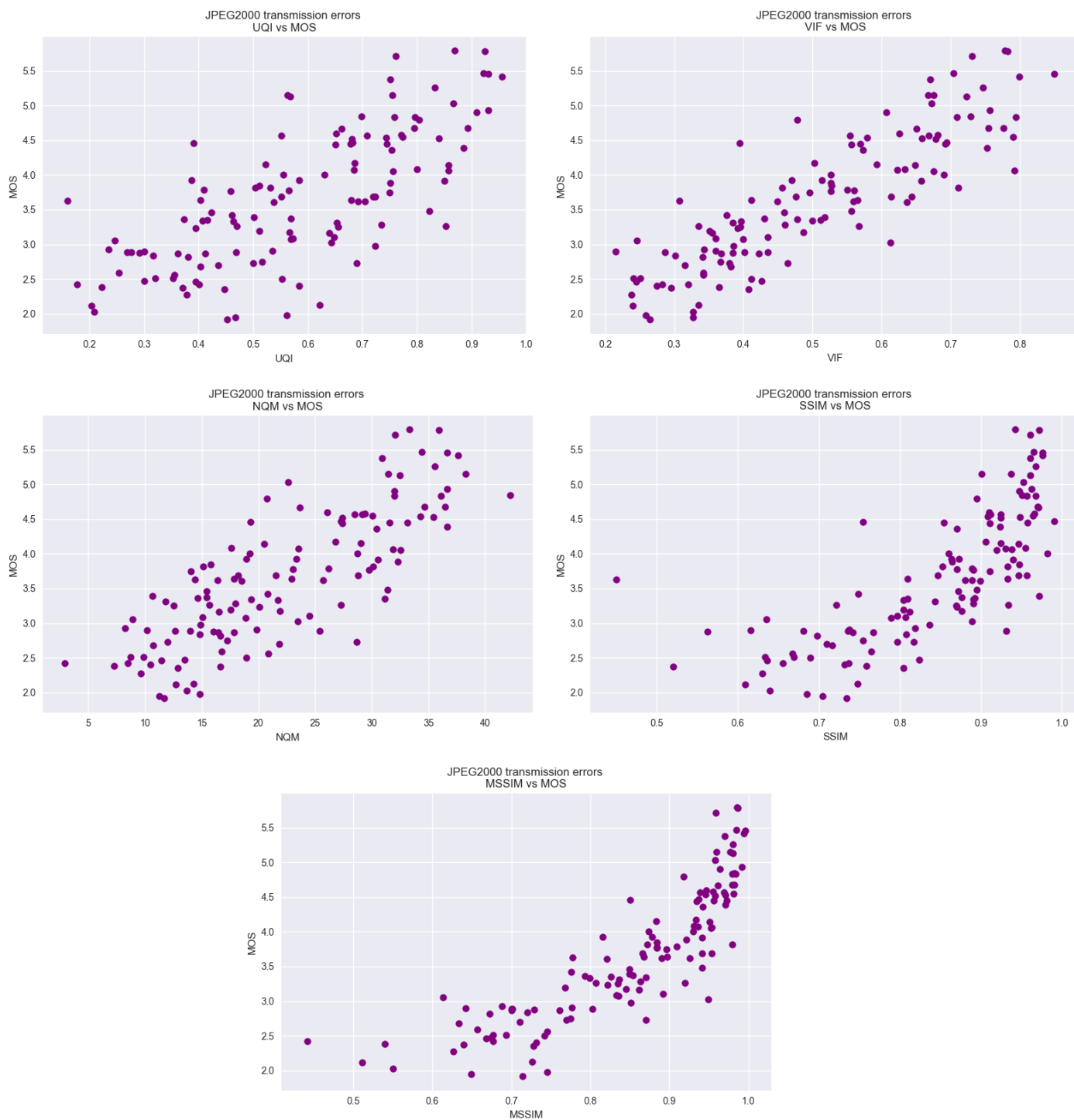


Figura B.11: Distorsión #13: JPEG2000 transmission errors

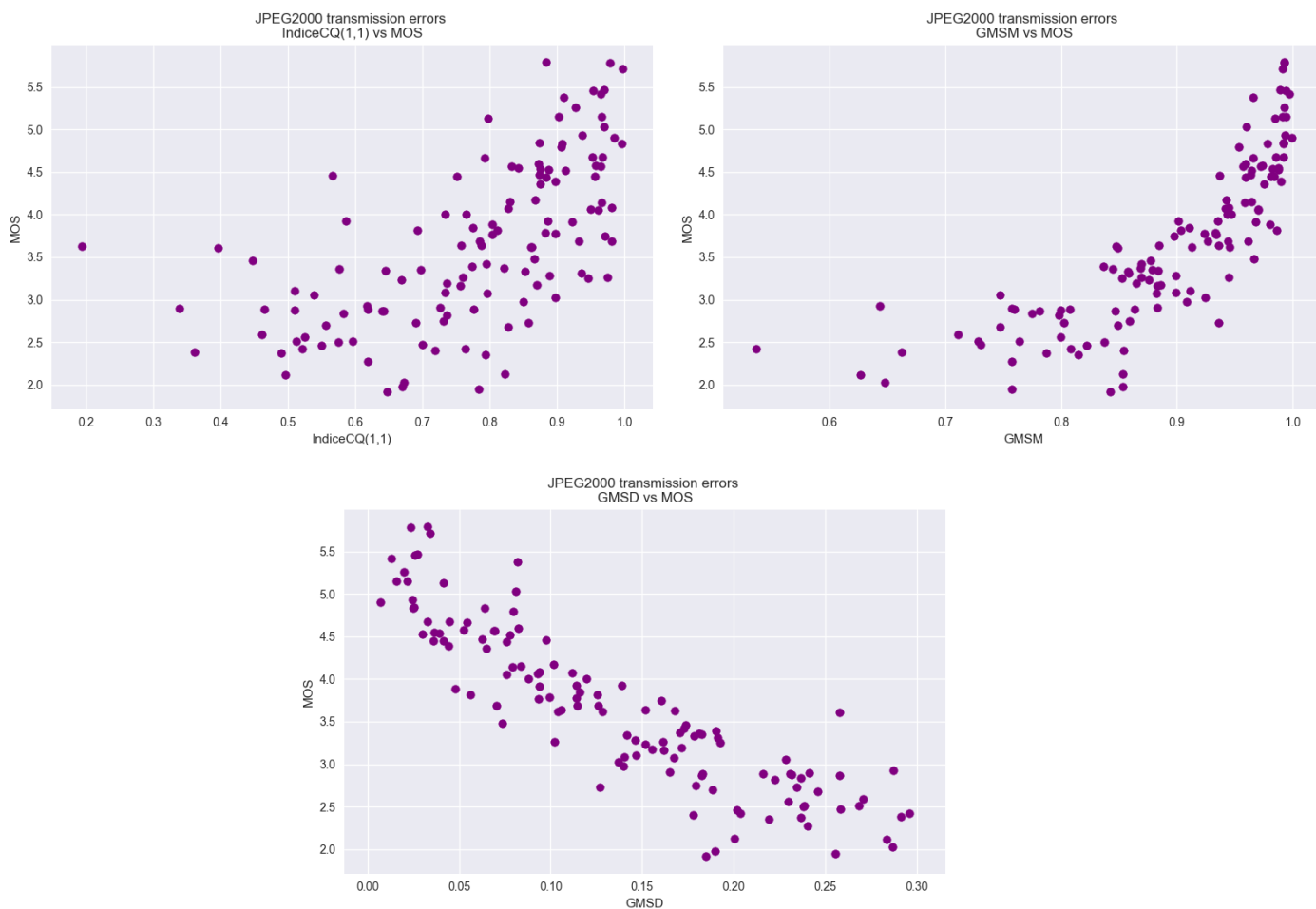


Figura B.12: Distorsión #13: JPEG2000 transmission errors

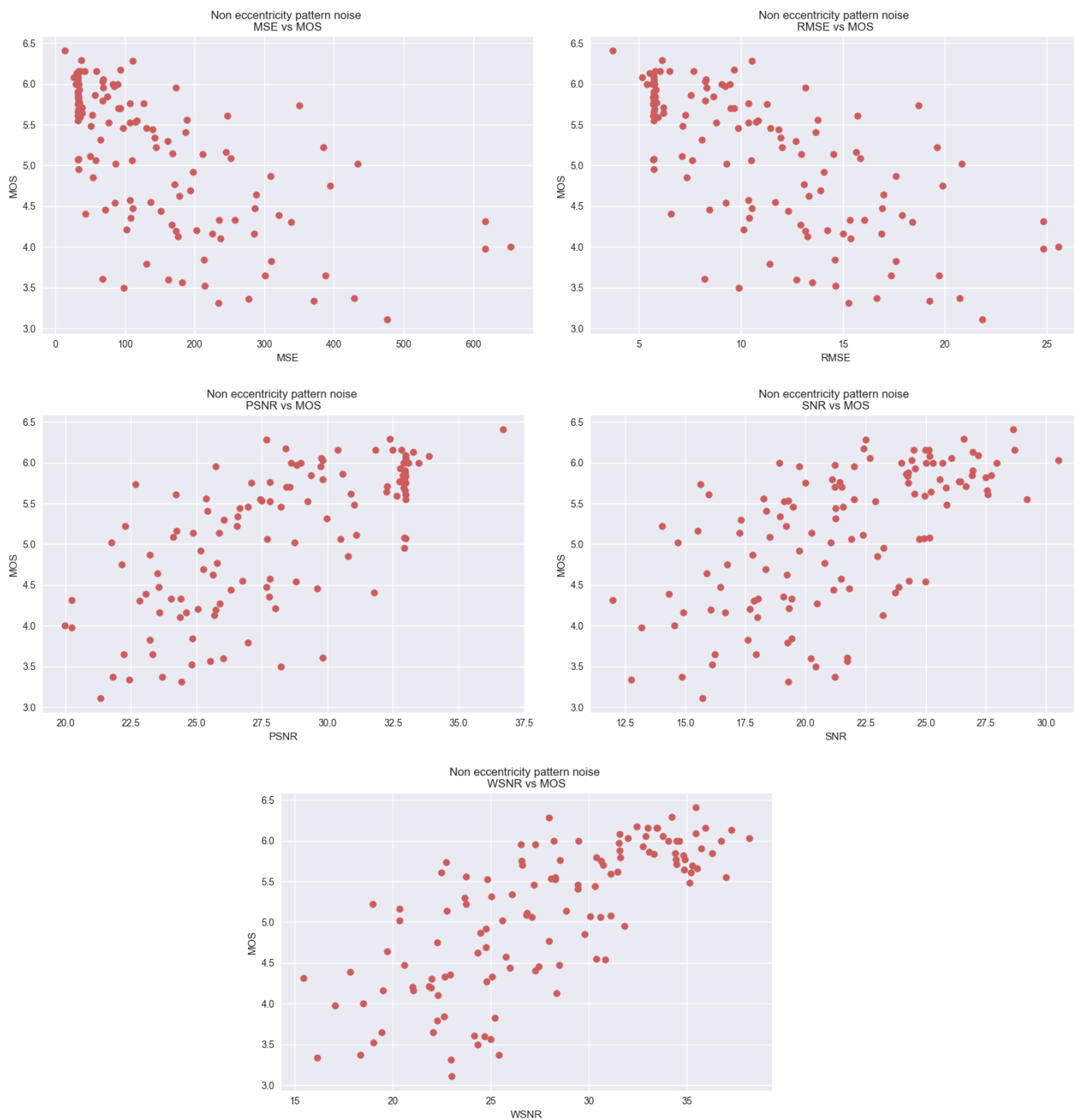


Figura B.13: Distorsión #14: Non eccentricity pattern noise

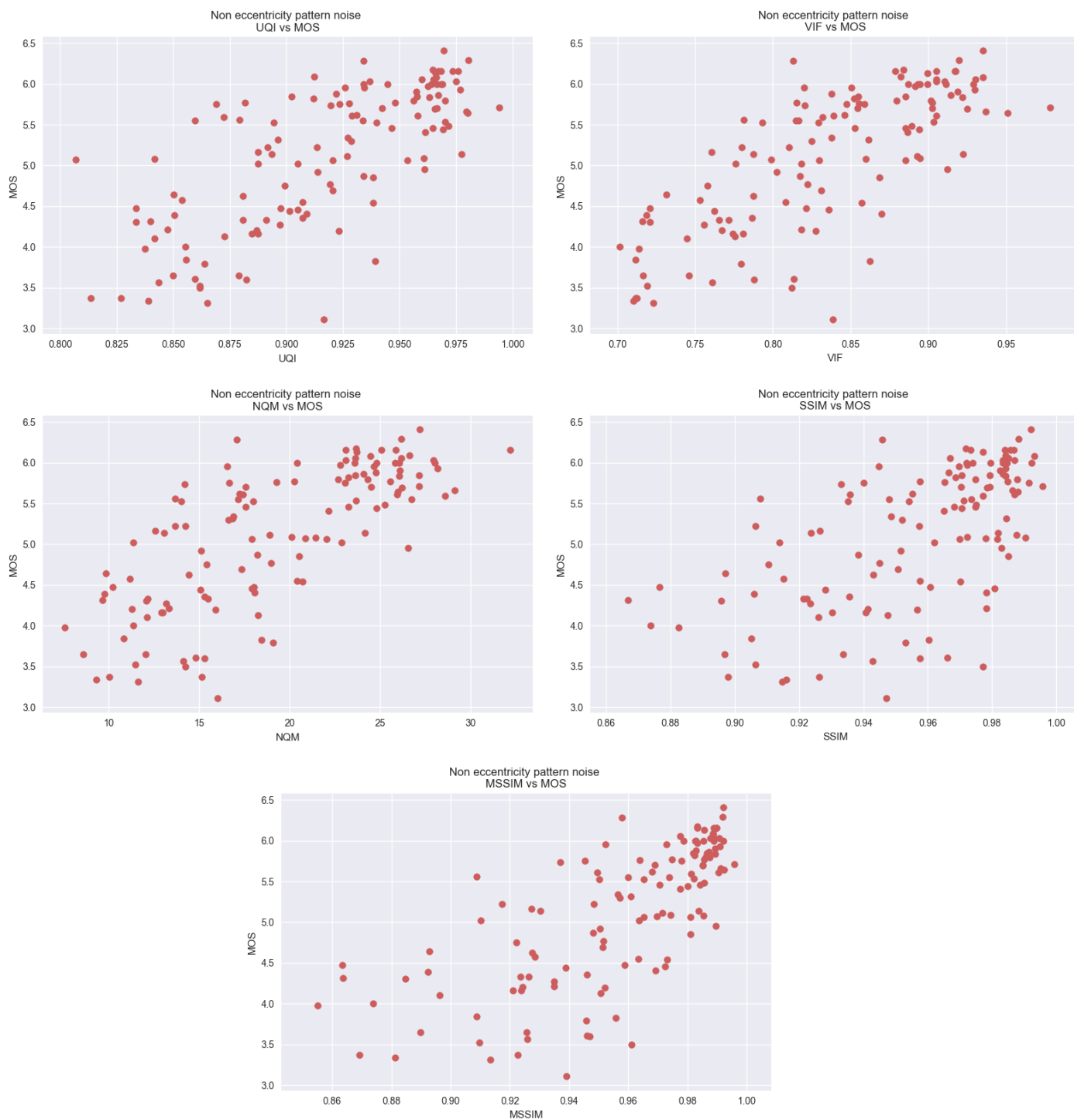


Figura B.14: Distorsión #14: Non eccentricity pattern noise



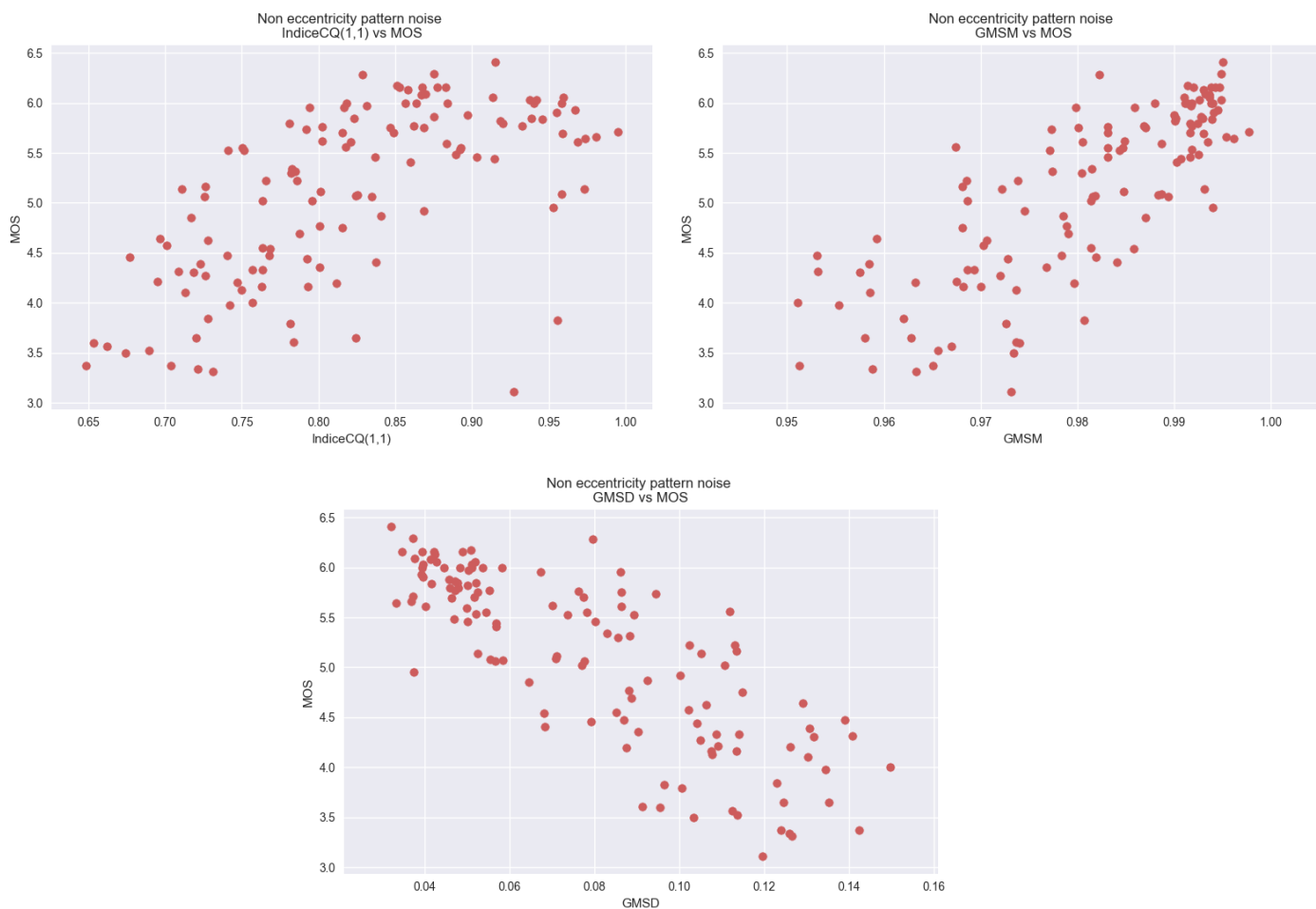


Figura B.15: Distorsión #14: Non eccentricity pattern noise

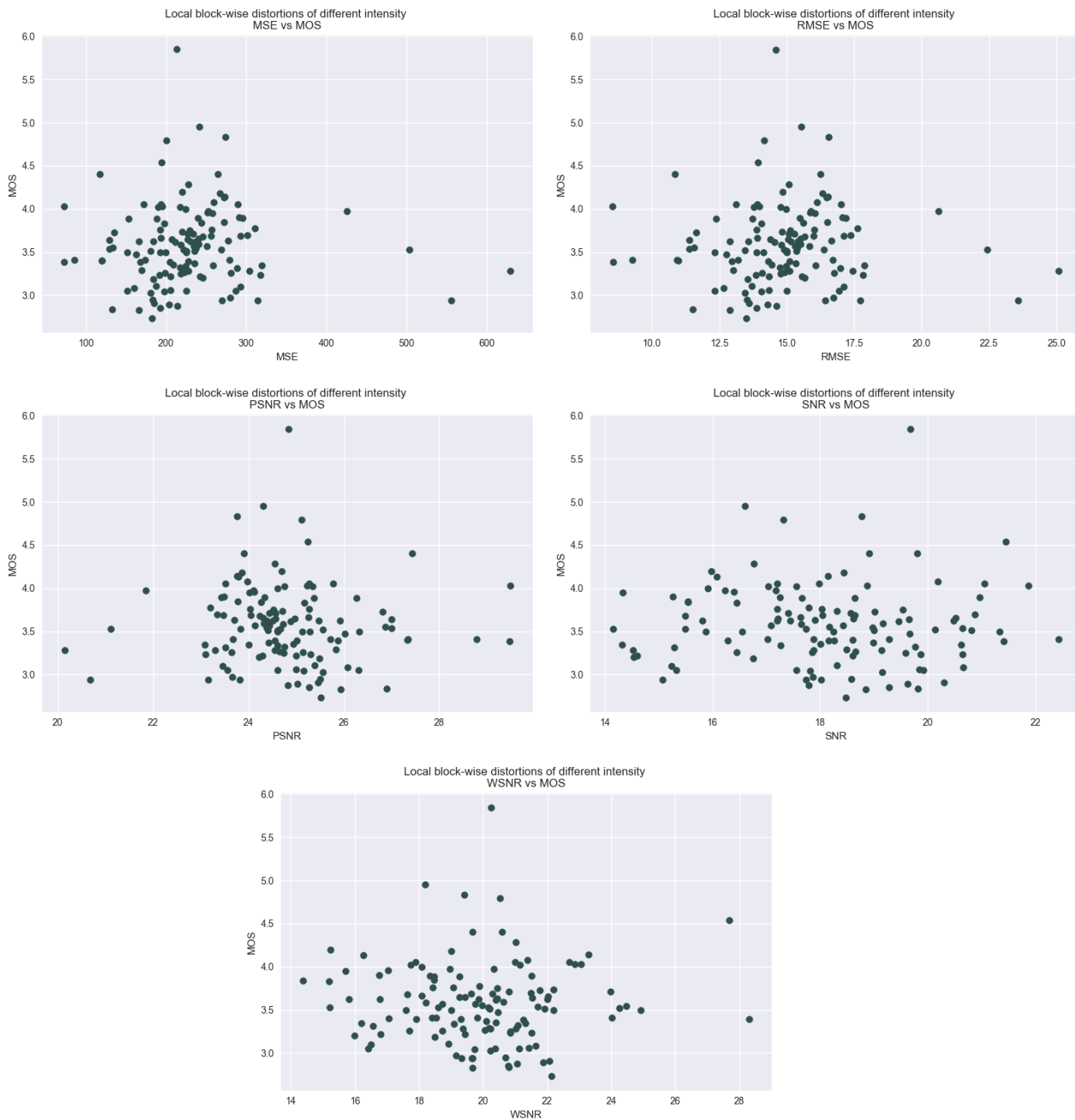


Figura B.16: Distorsión #15: Local block-wise distortions of different intensity

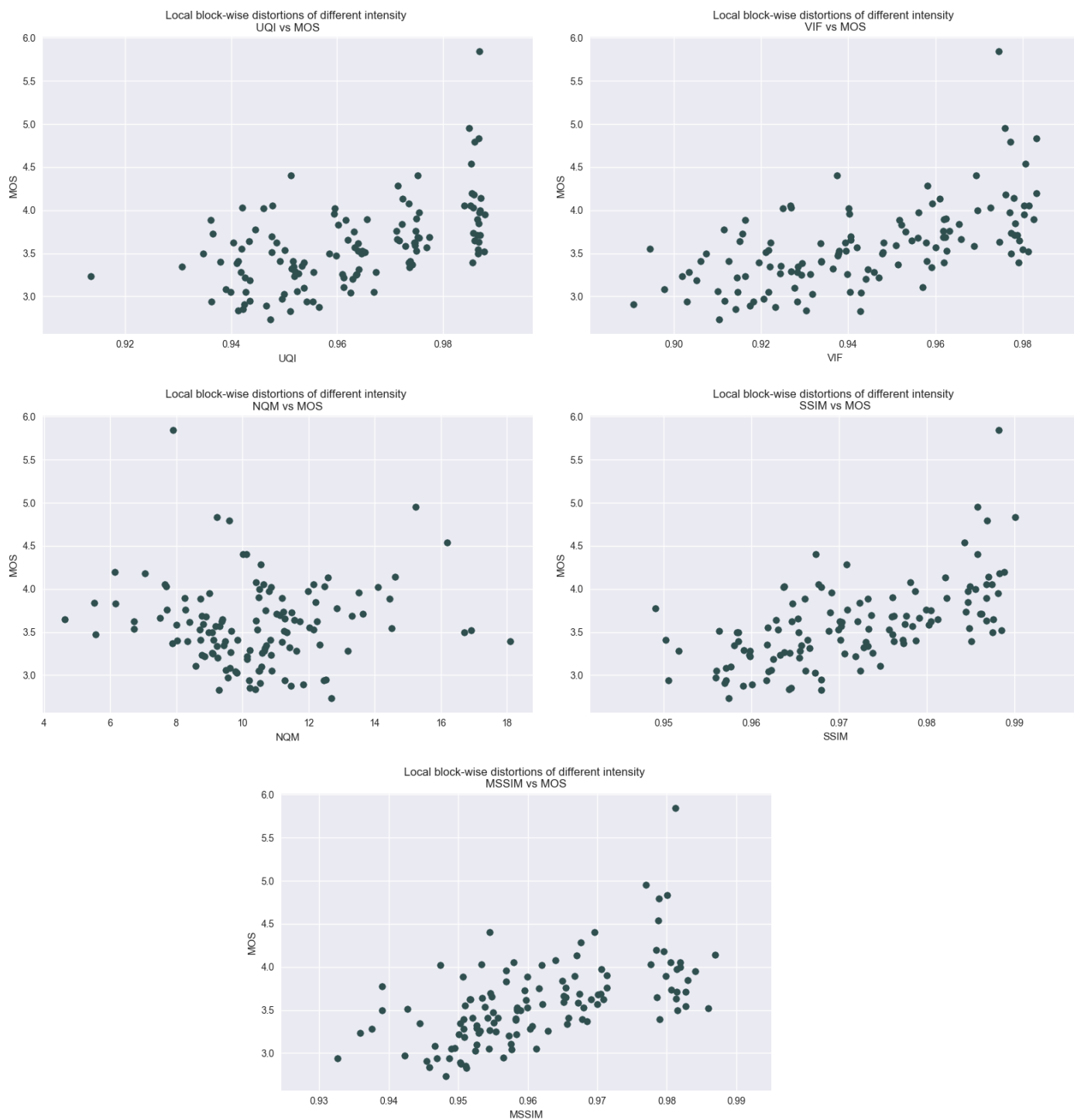


Figura B.17: Distorsión #15: Local block-wise distortions of different intensity

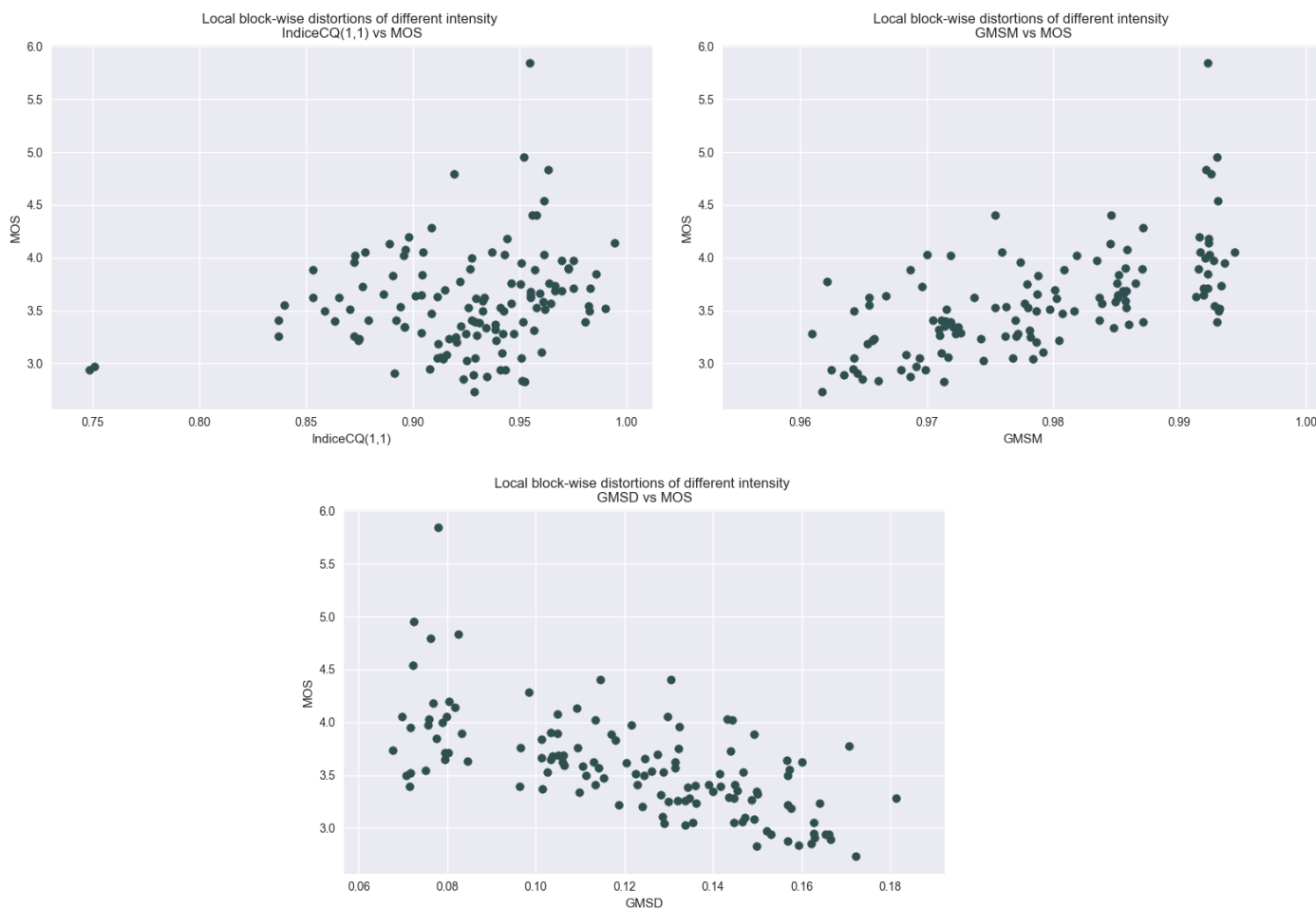


Figura B.18: Distorsión #15: Local block-wise distortions of different intensity

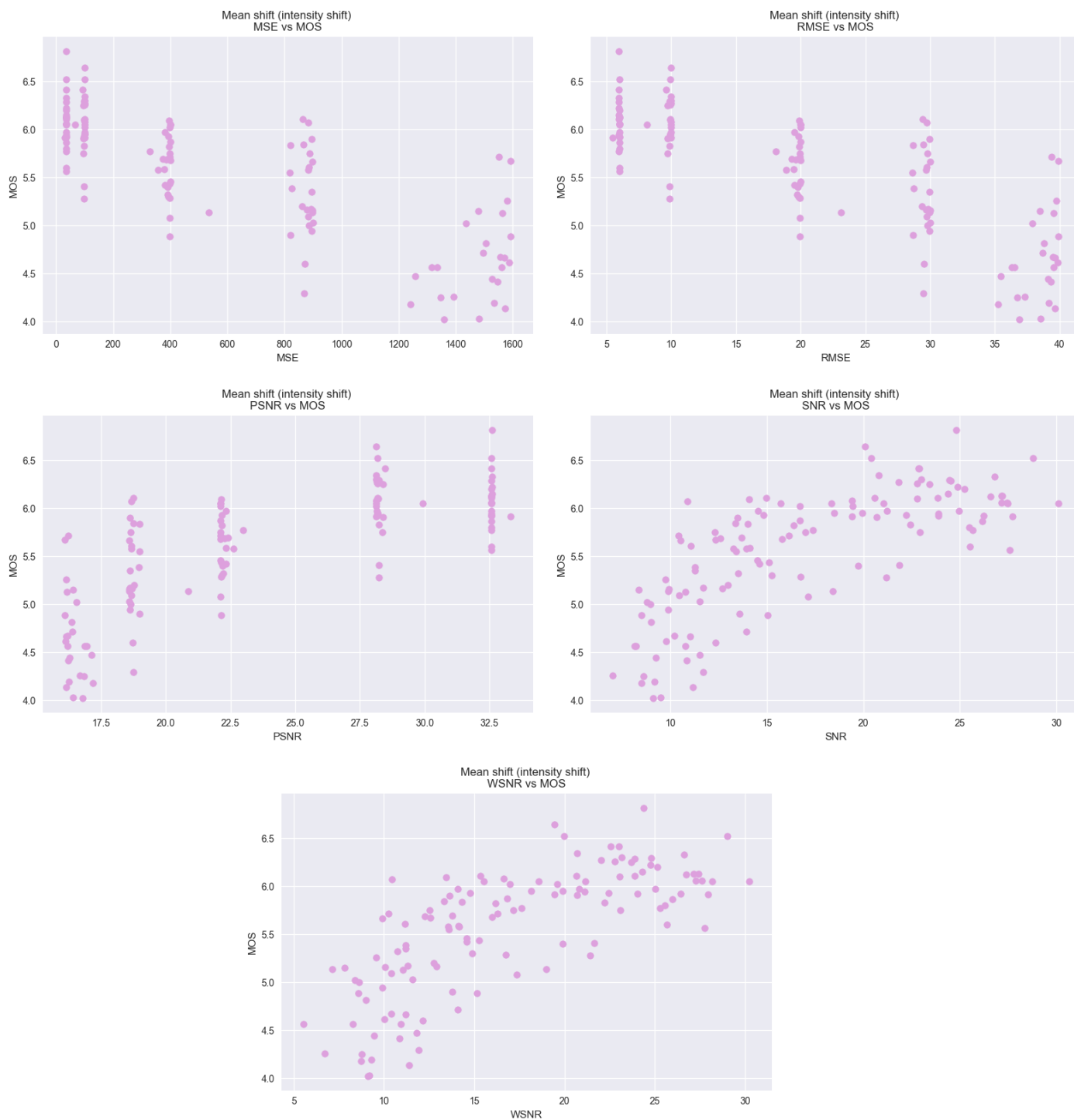


Figura B.19: Distorsión #16: Mean shift

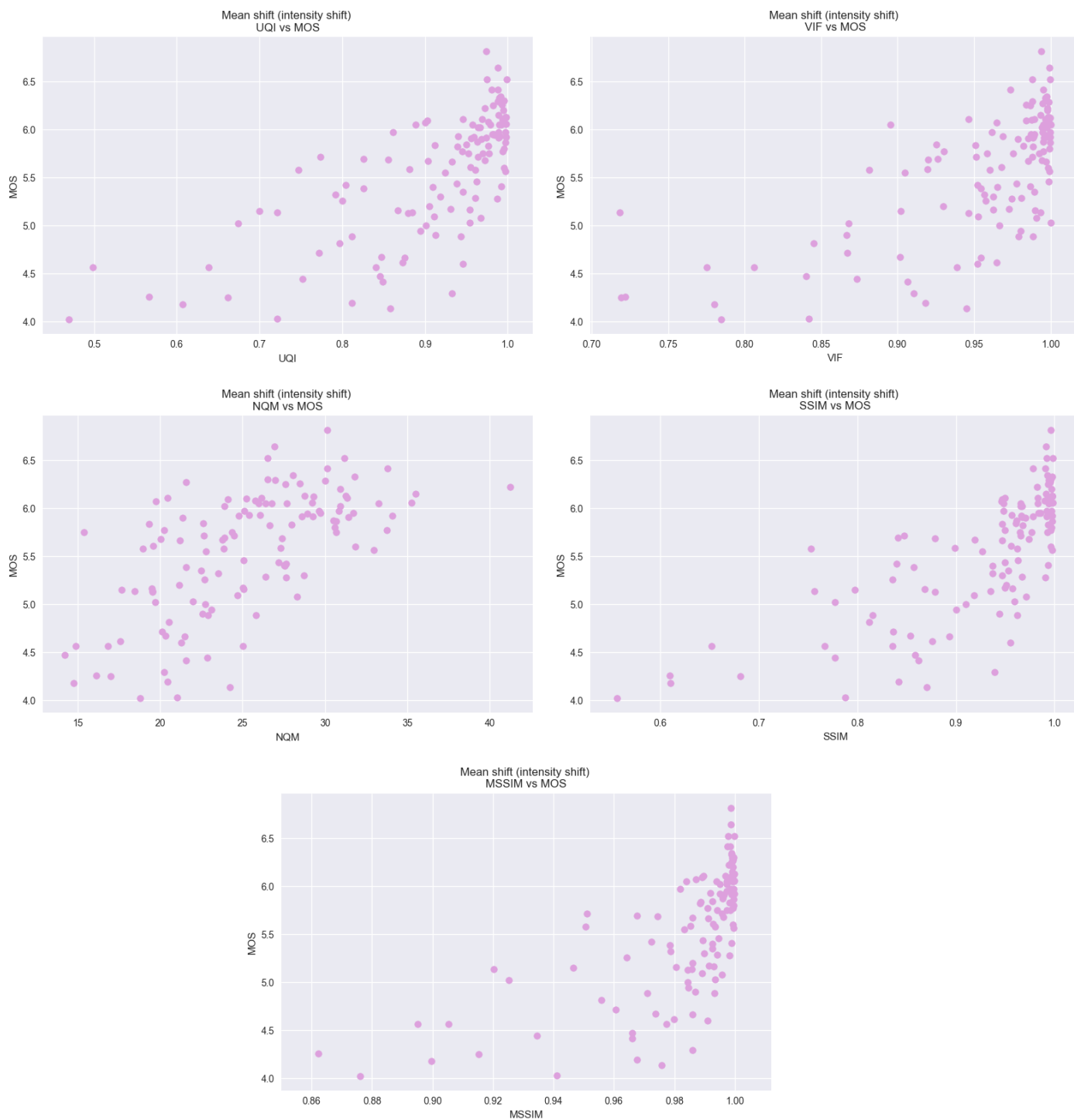


Figura B.20: Distorsión #16: Mean shift

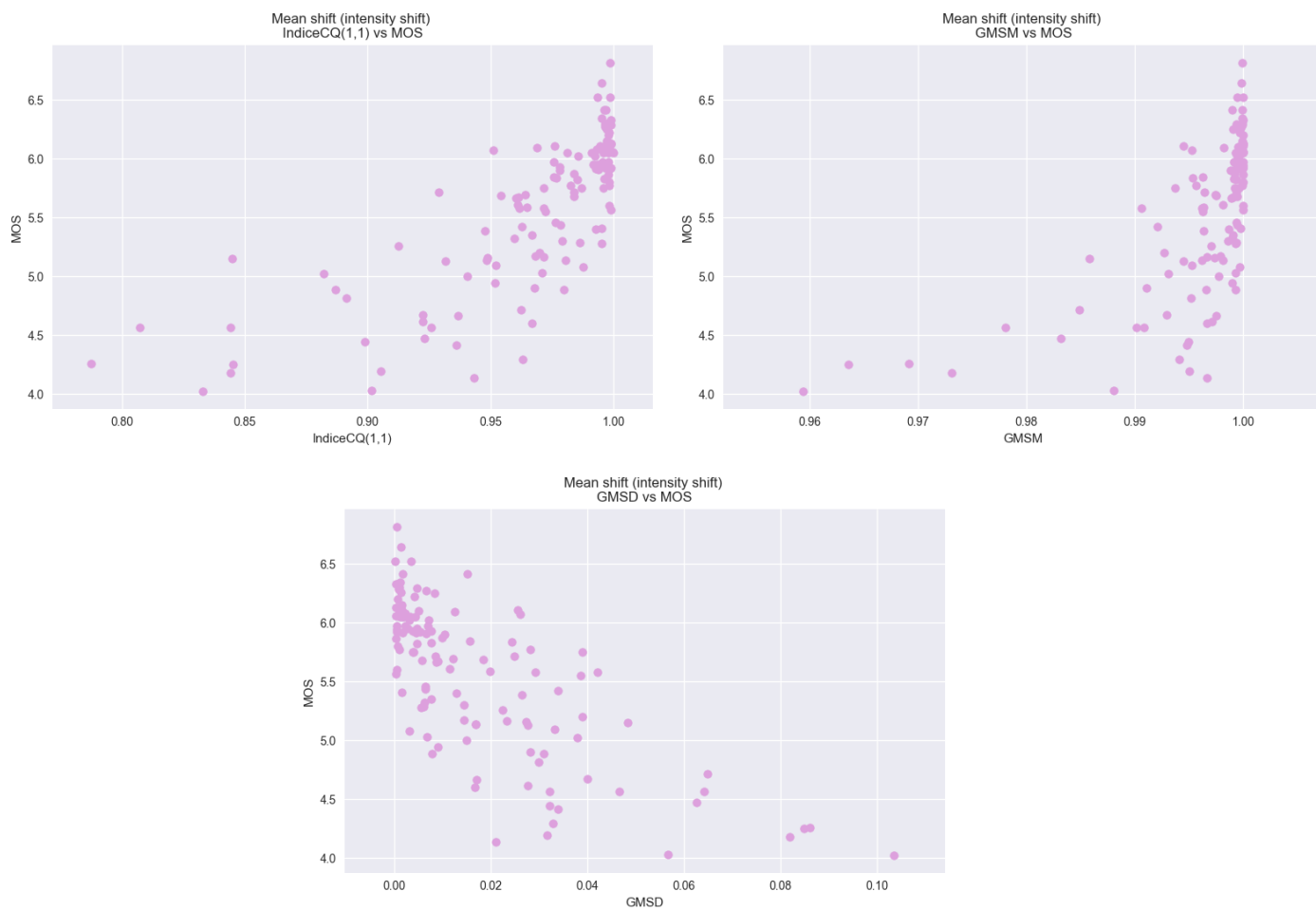


Figura B.21: Distorsión #16: Mean shift

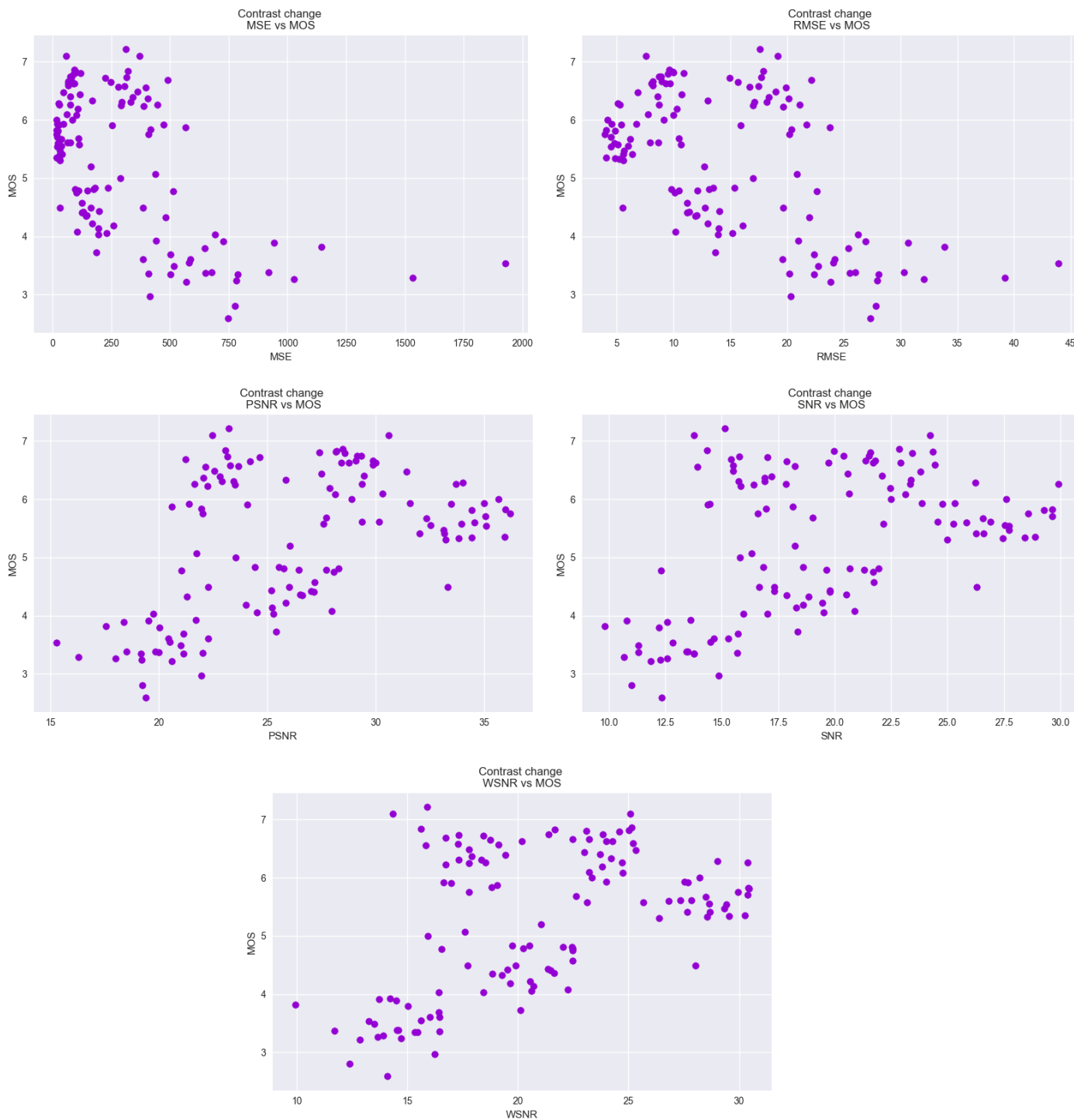


Figura B.22: Distorsión #17: Contrast change



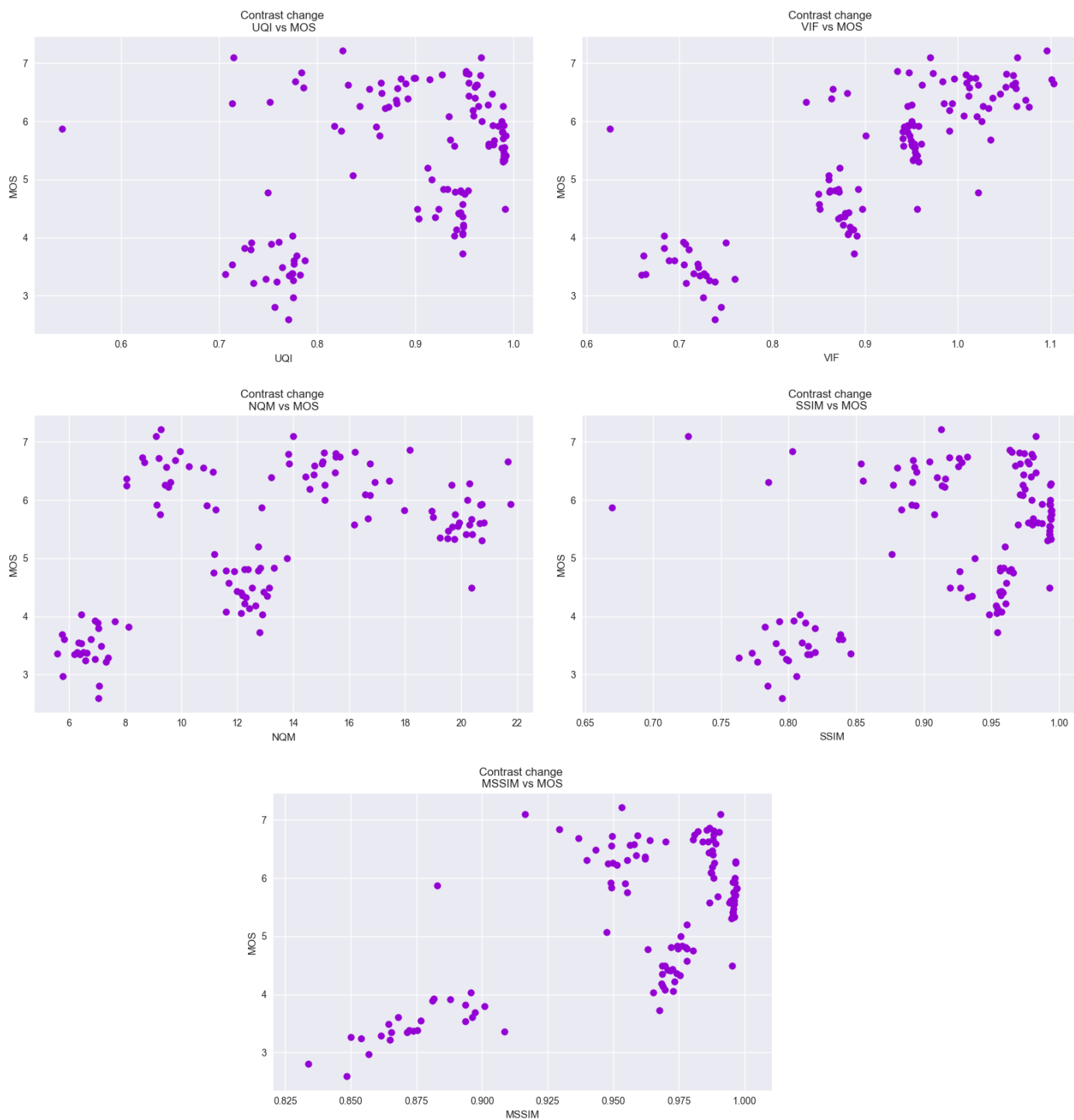


Figura B.23: Distorsión #17: Contrast change

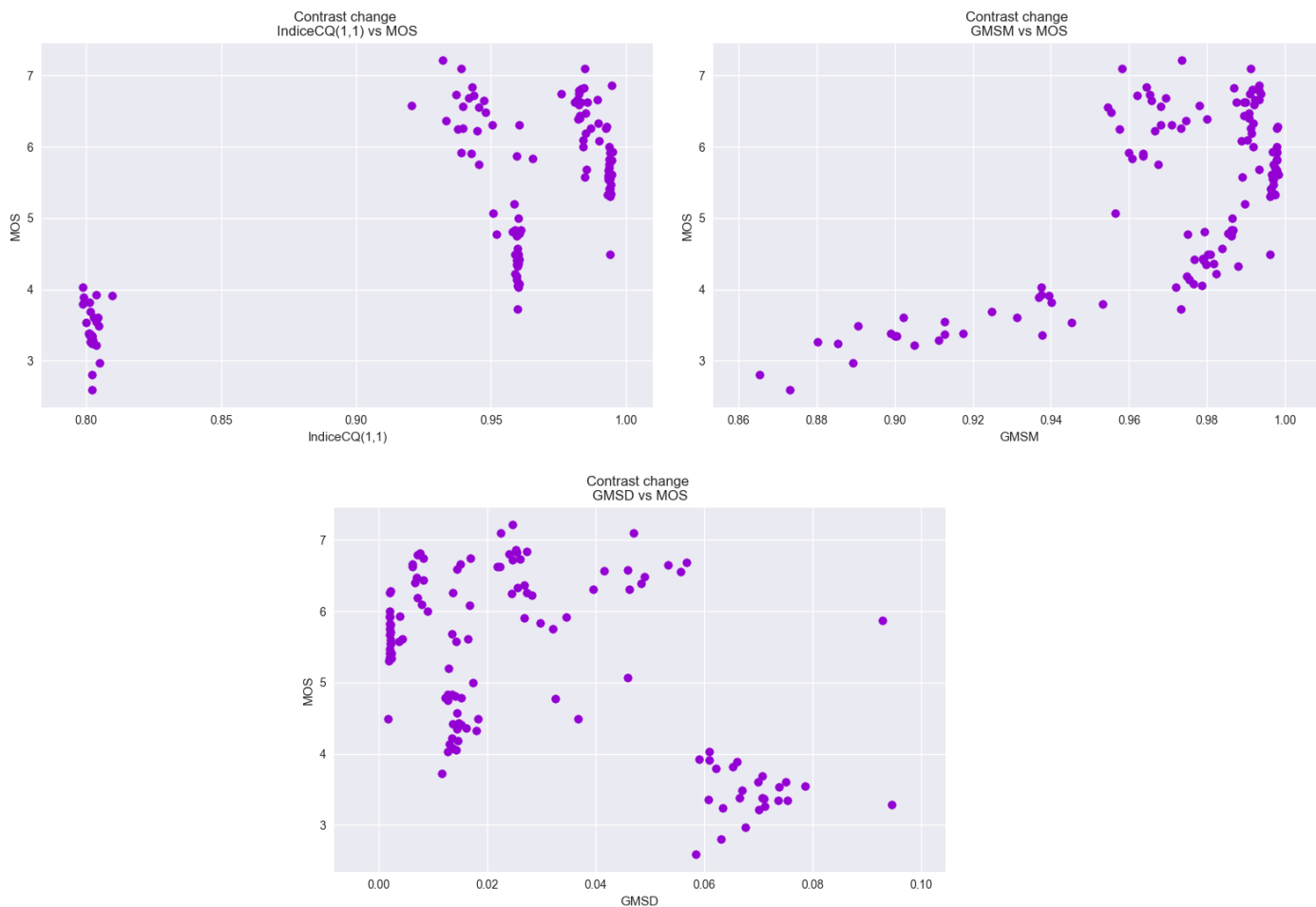


Figura B.24: Distorsión #17: Contrast change

---

## Apéndice C

# Diagramas de dispersión de las distorsiones #18 a #24

En este apéndice se presentan los diagramas de dispersión de las distorsiones nuevas para cada métrica y el MOS.

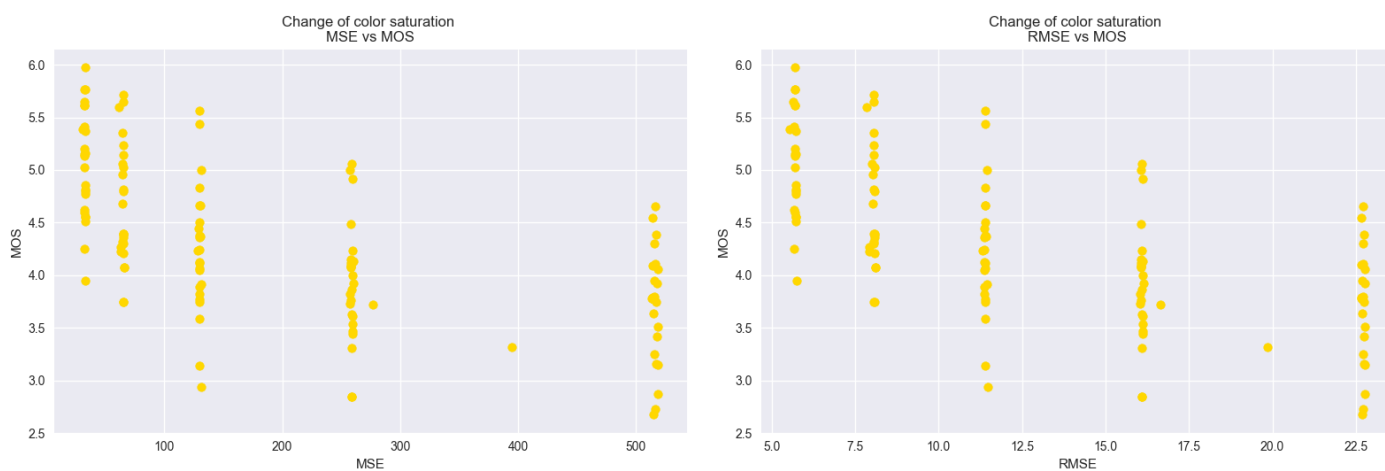


Figura C.1: Distorsión #18: Change of color saturation. Ninguna métrica parece ser aceptable para este tipo de distorsión.

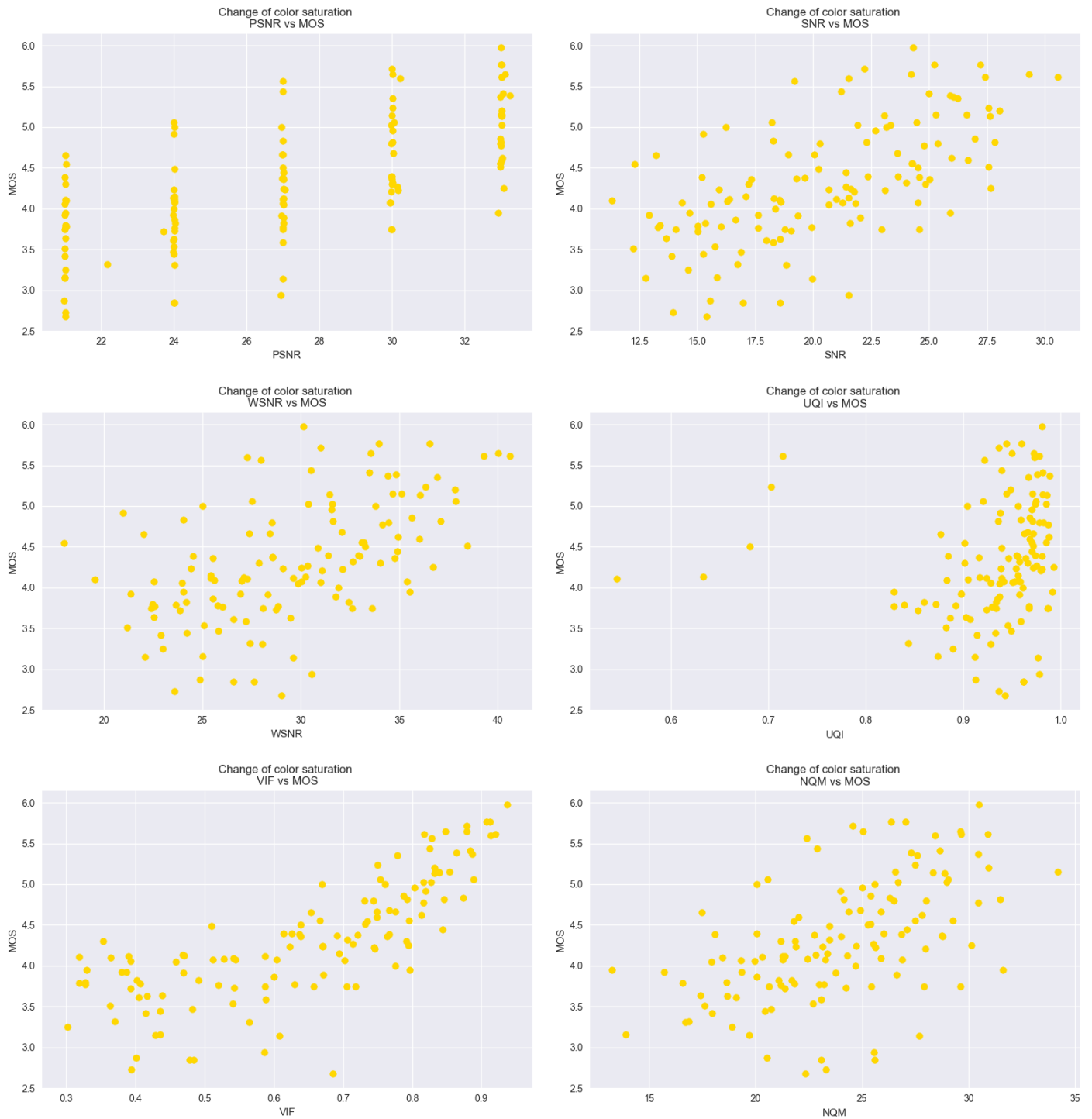


Figura C.2: Distorsión #18: Change of color saturation. Ninguna métrica parece ser aceptable para este tipo de distorsión. Aunque si tuvieramos que elegir una la más aceptable sería la VIF.

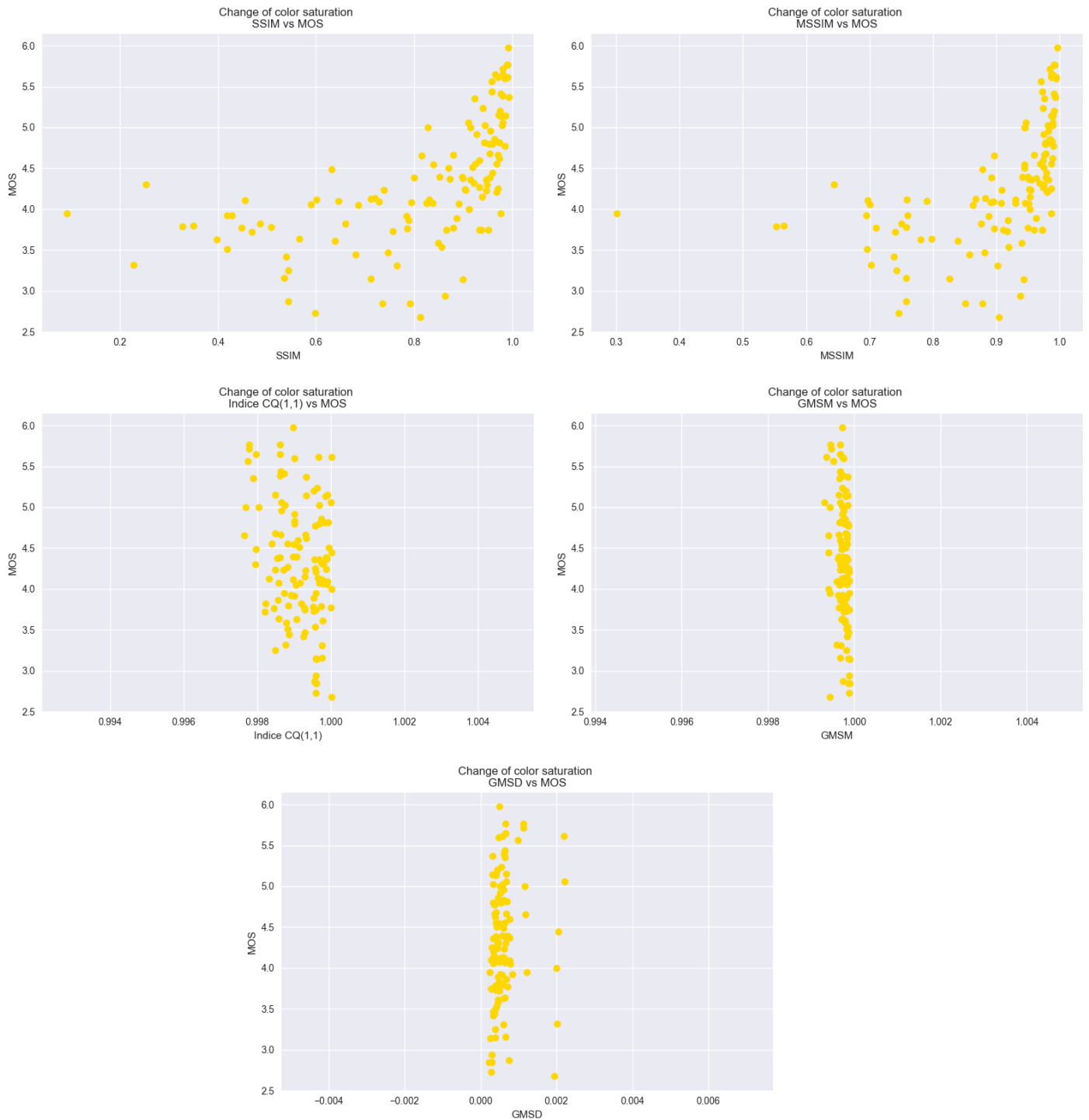


Figura C.3: Distorsión #18: Change of color saturation. Ninguna métrica parece ser aceptable para este tipo de distorsión.

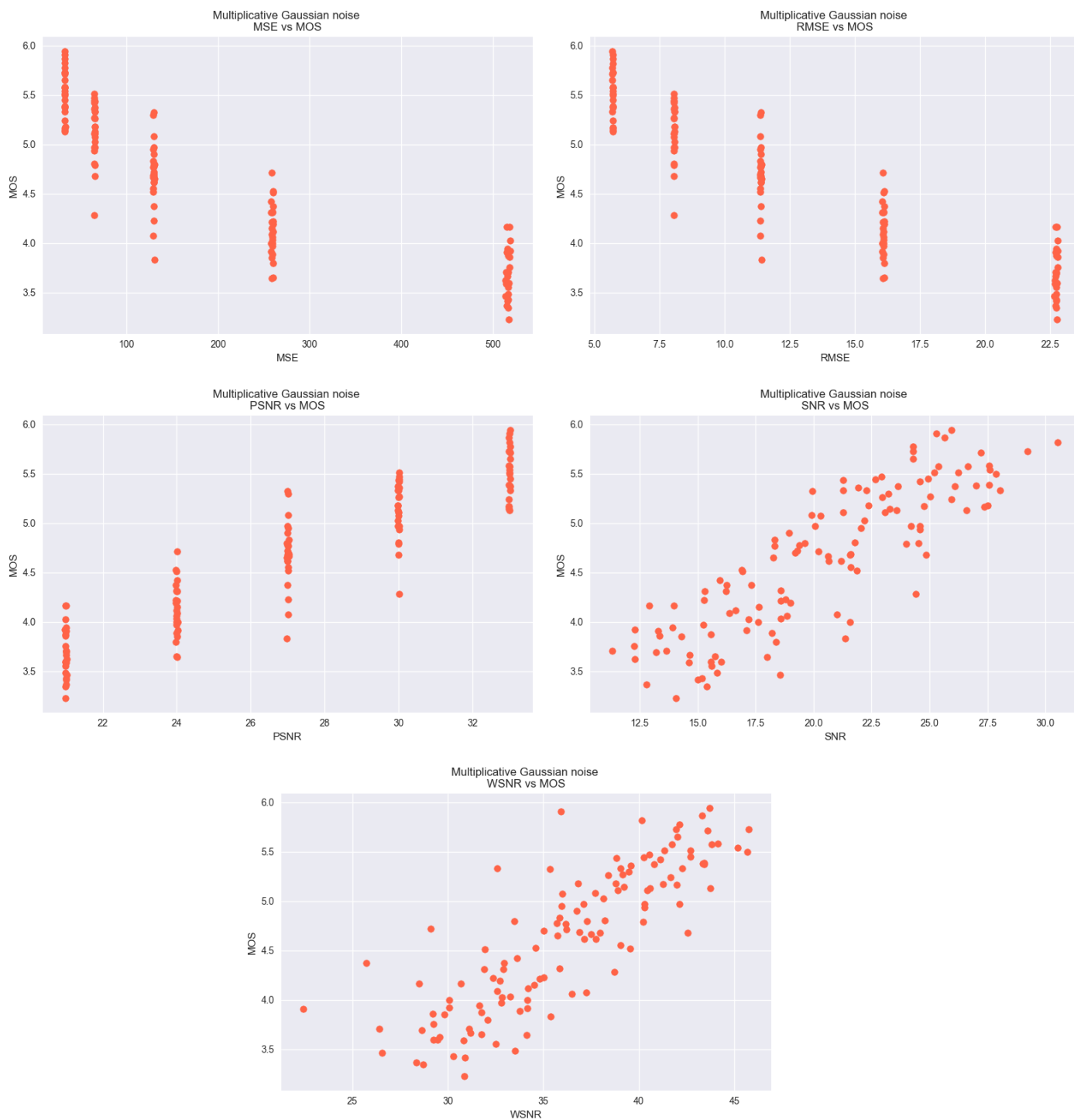


Figura C.4: Distorsión #19: Multiplicative Gaussian noise

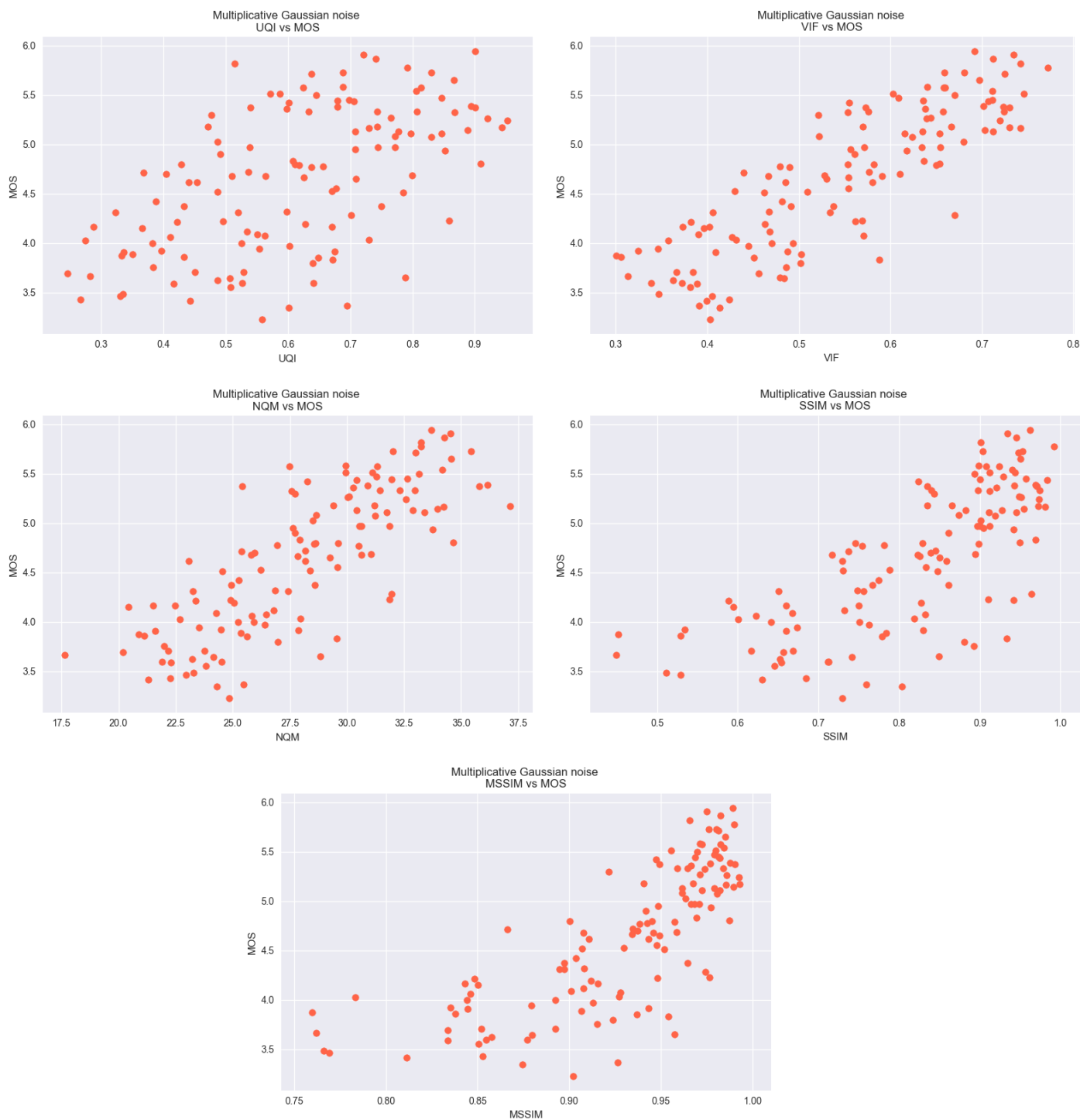


Figura C.5: Distorsión #19: Multiplicative Gaussian noise

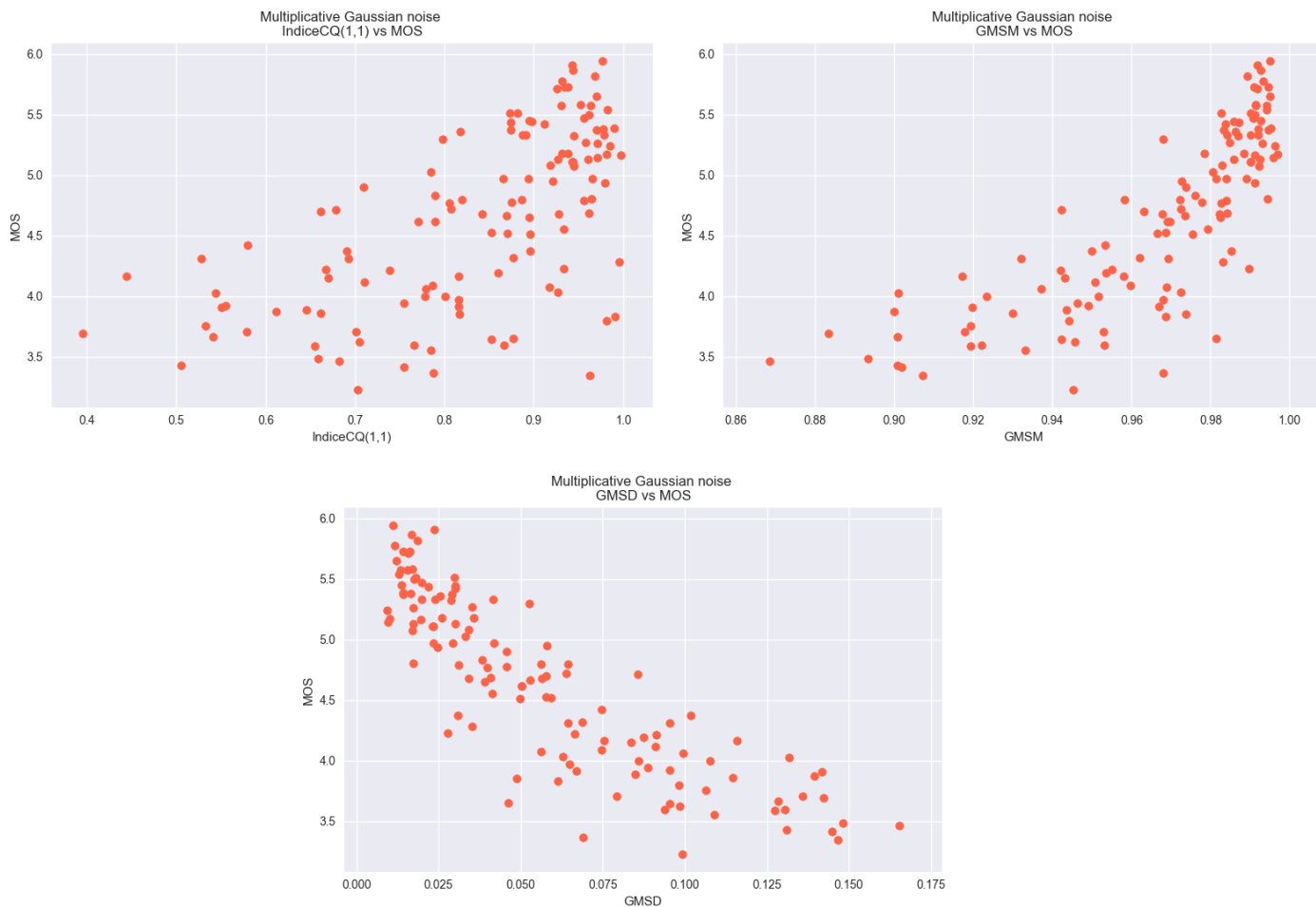


Figura C.6: Distorsión #19: Multiplicative Gaussian noise



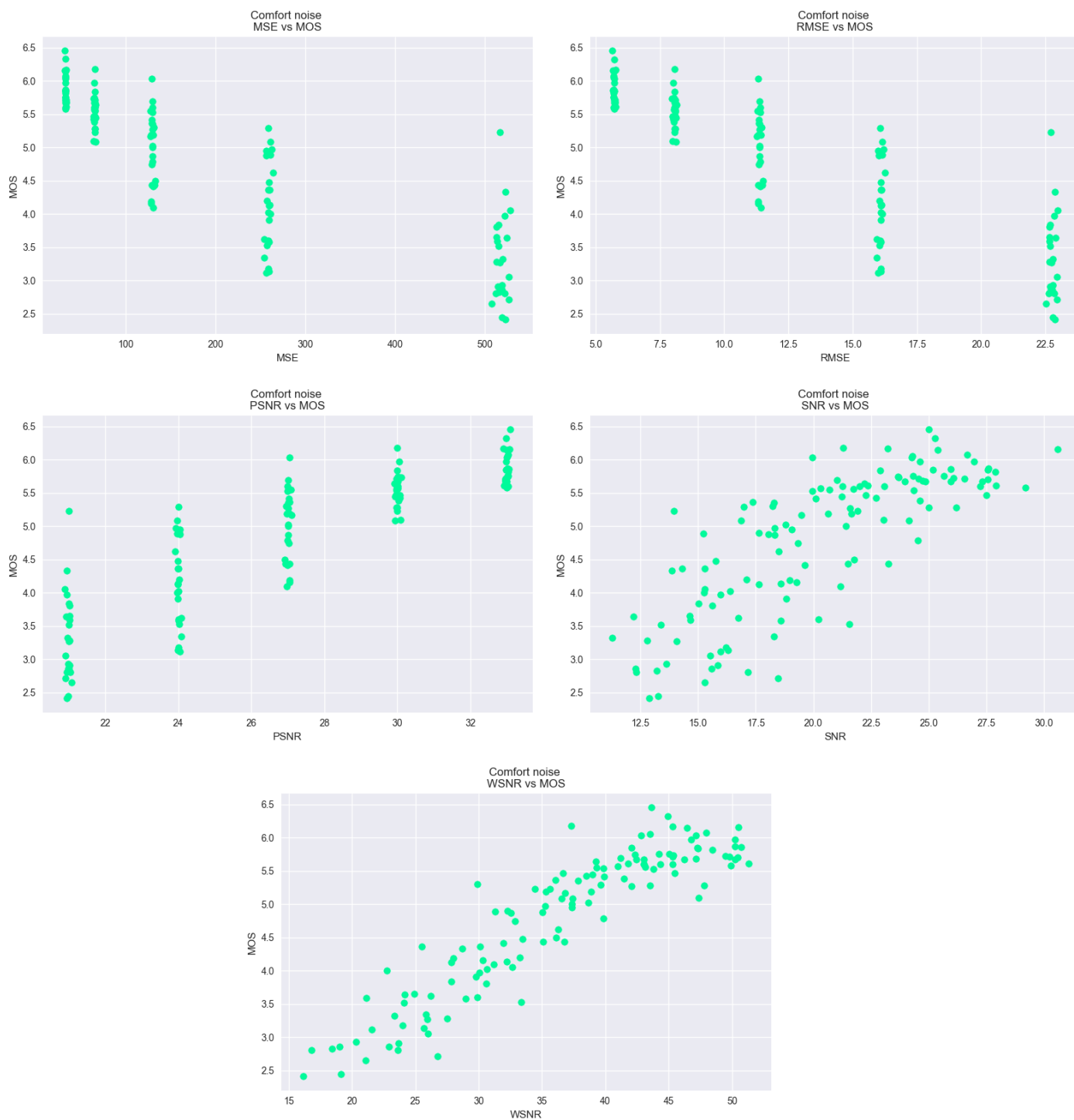


Figura C.7: Distorsión #20: Comfort noise

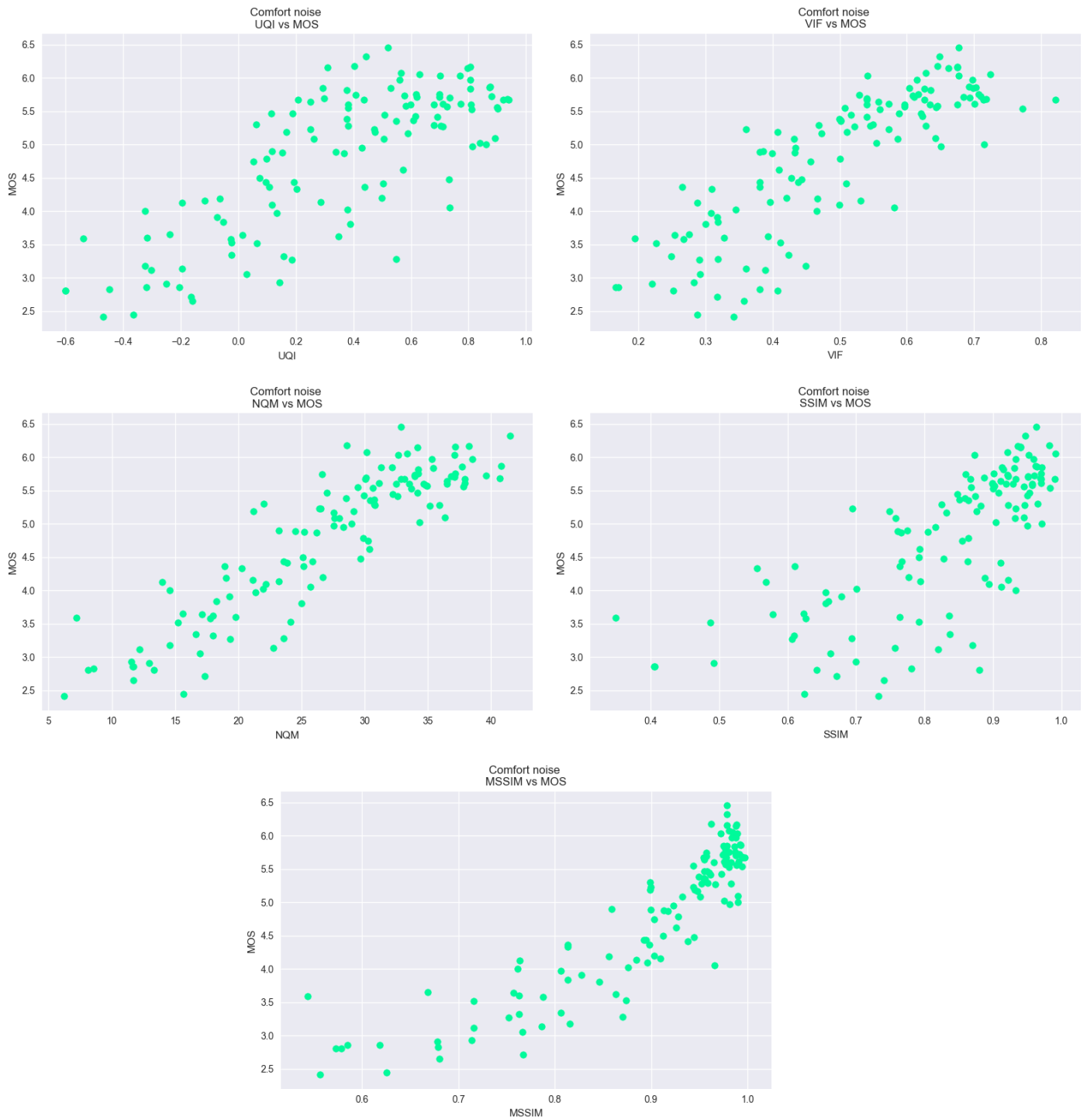


Figura C.8: Distorsión #20: Comfort noise

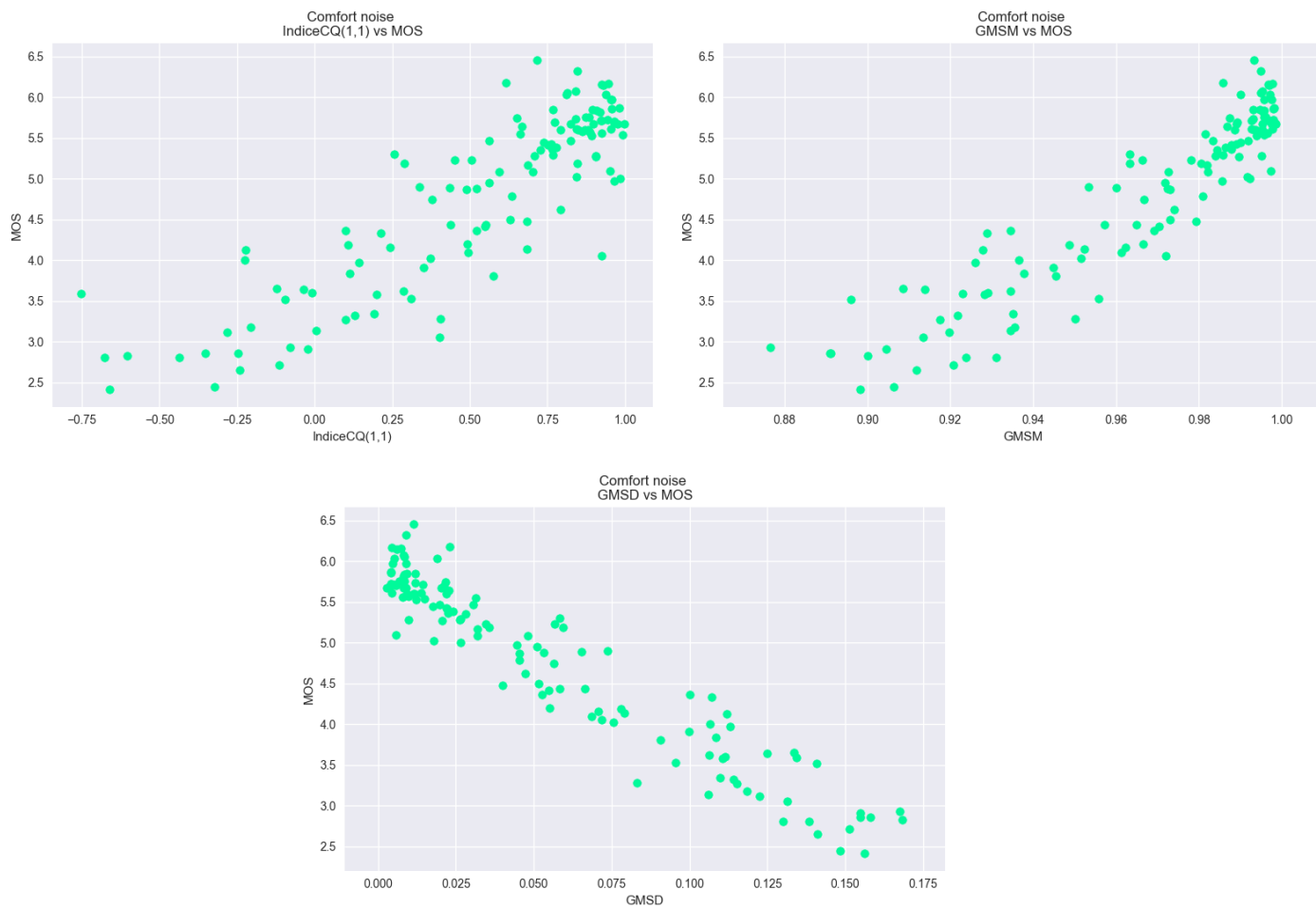


Figura C.9: Distorsión #20: Comfort noise

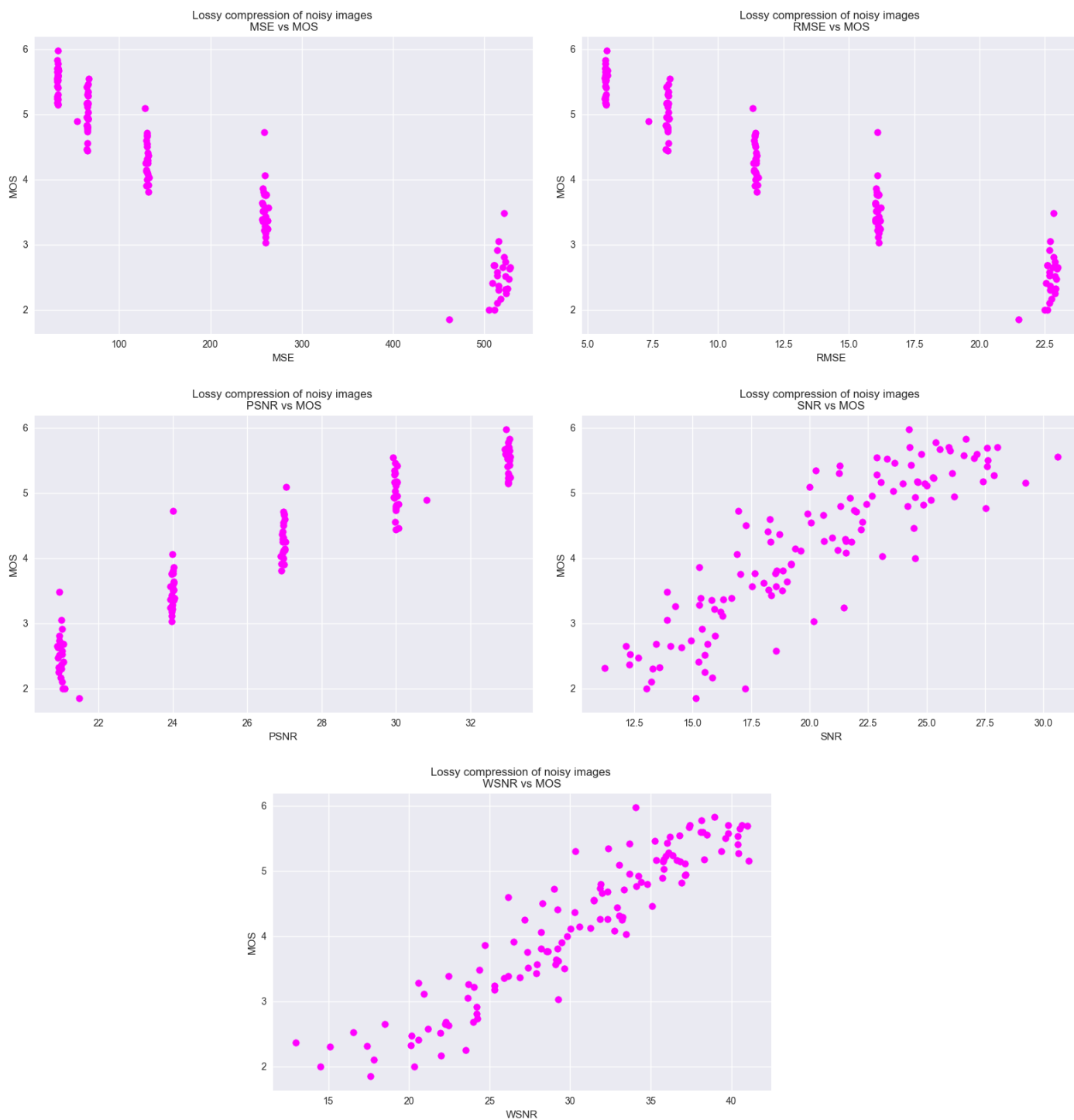


Figura C.10: Distorsión #21: Lossy compression of noisy images

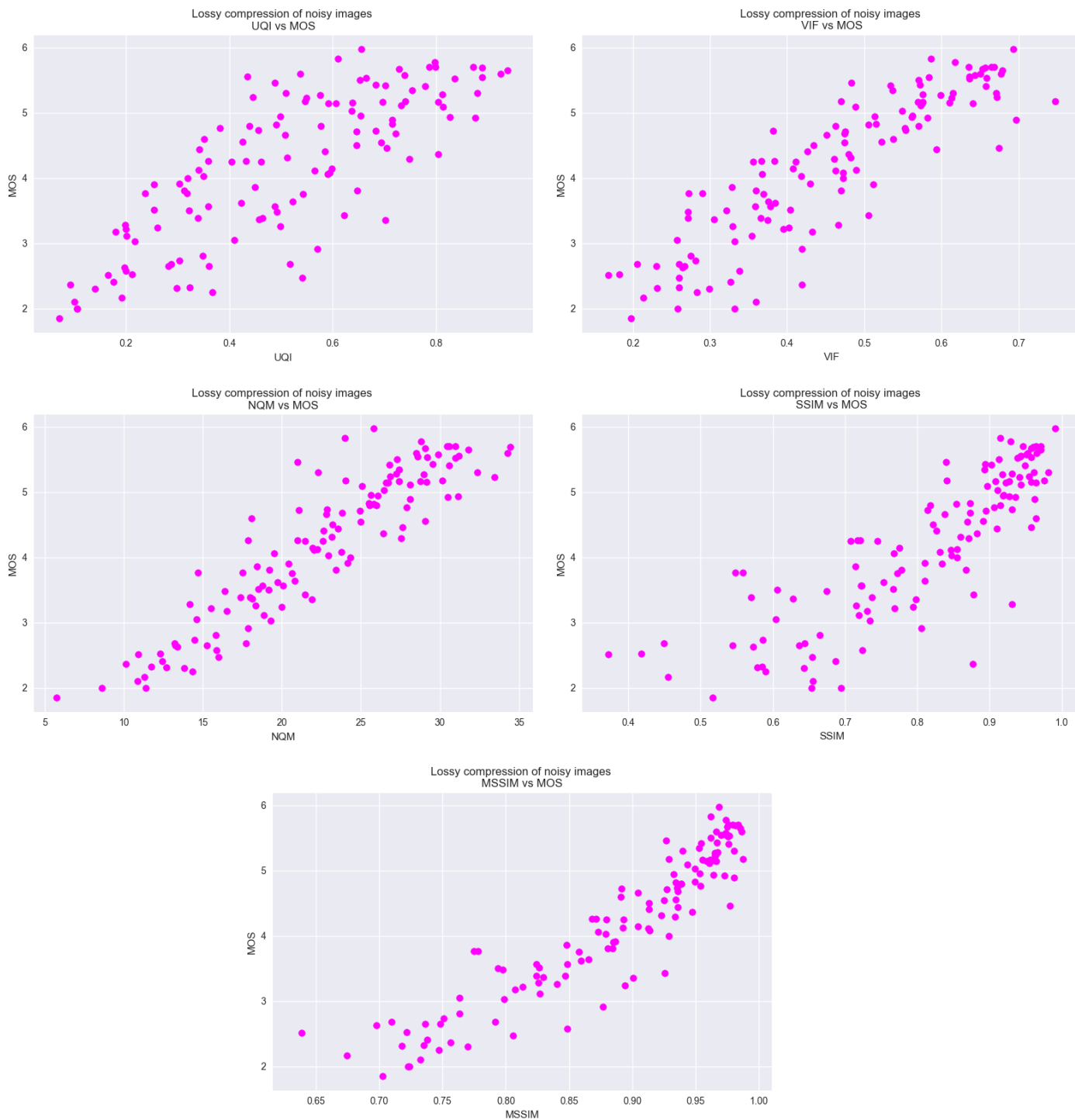


Figura C.11: Distorsión #21: Lossy compression of noisy images

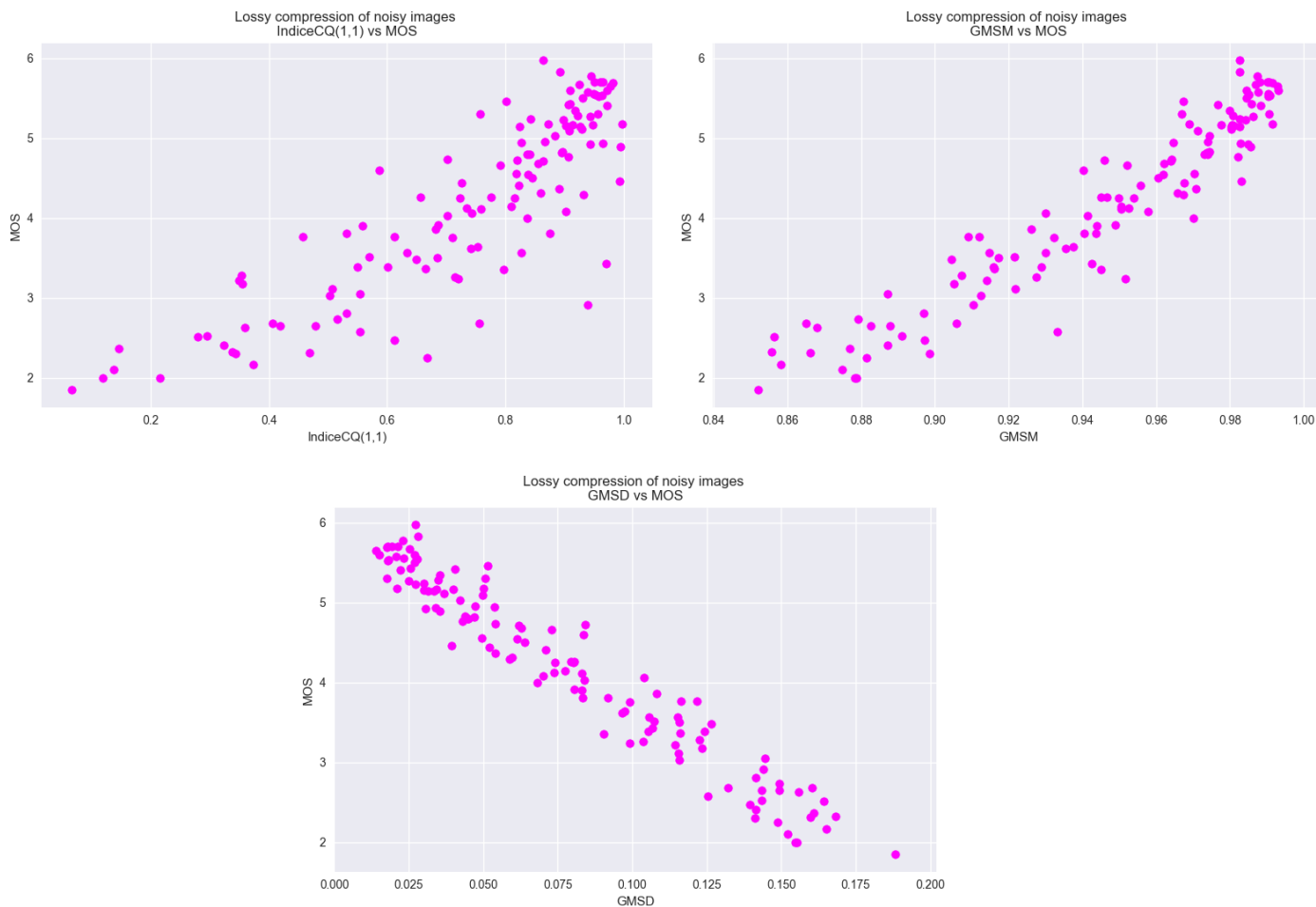


Figura C.12: Distorsión #21: Lossy compression of noisy images

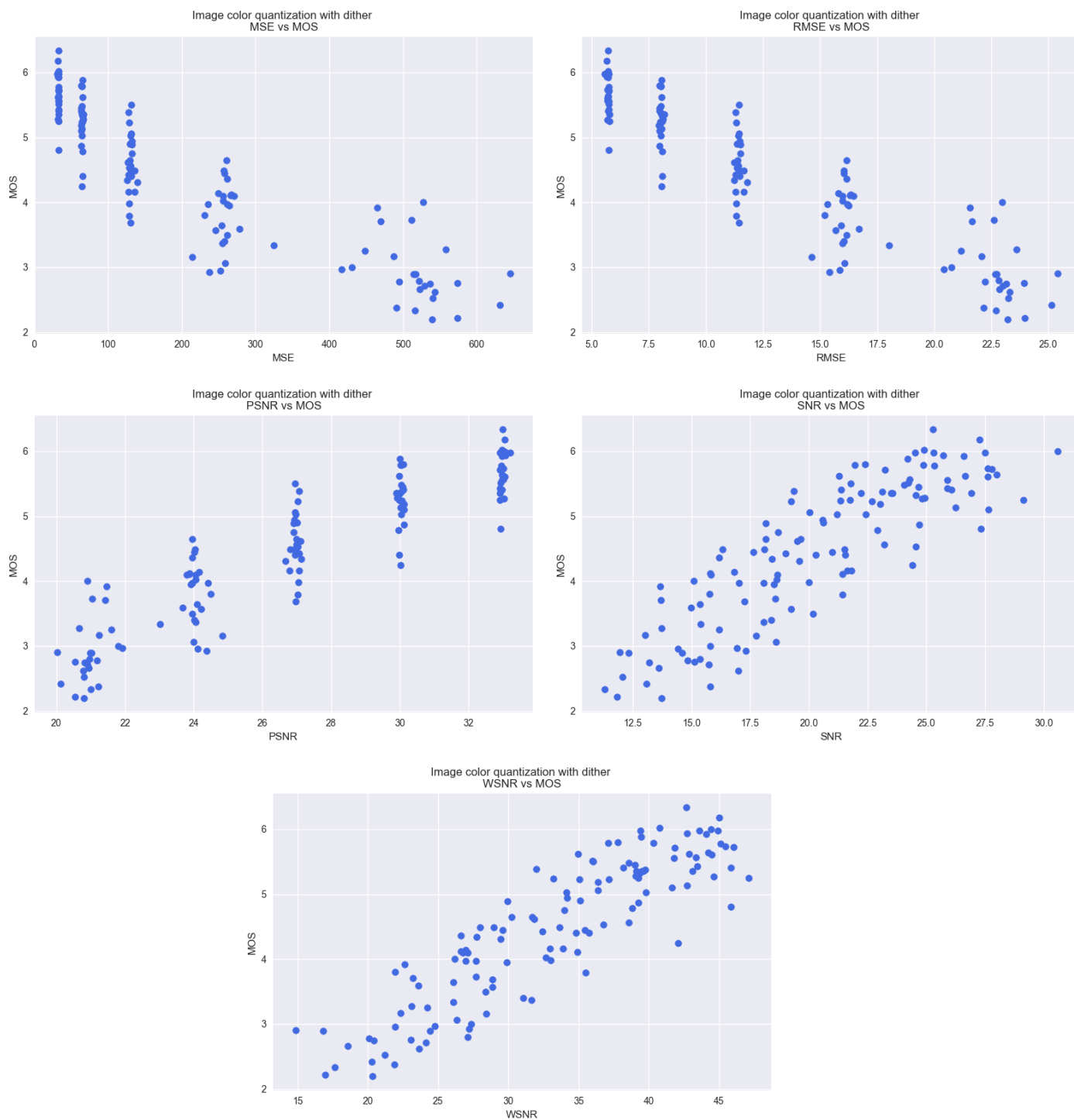


Figura C.13: Distorsión #22: Image color quantization with dither

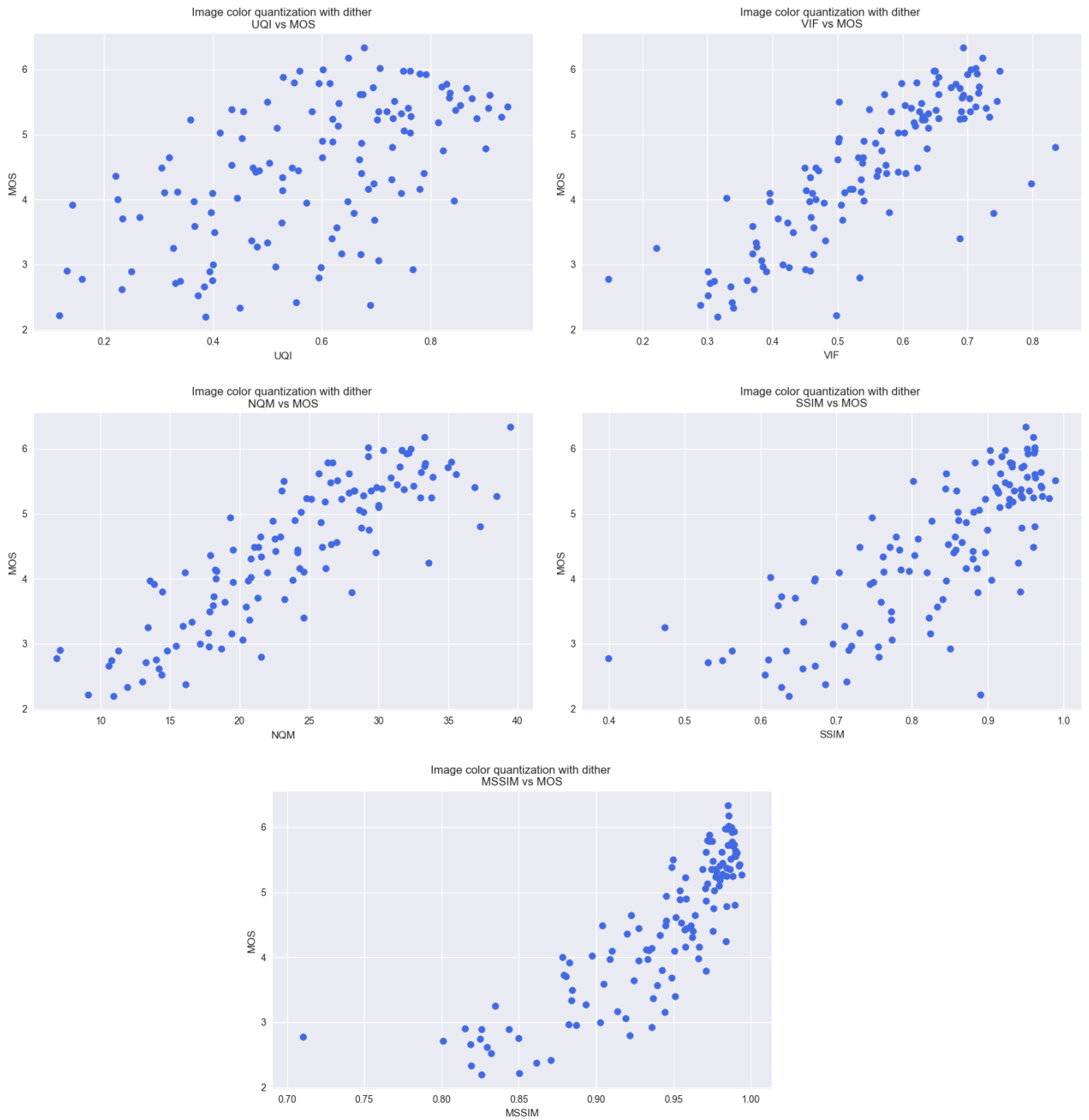


Figura C.14: Distorsión #22: Image color quantization with dither



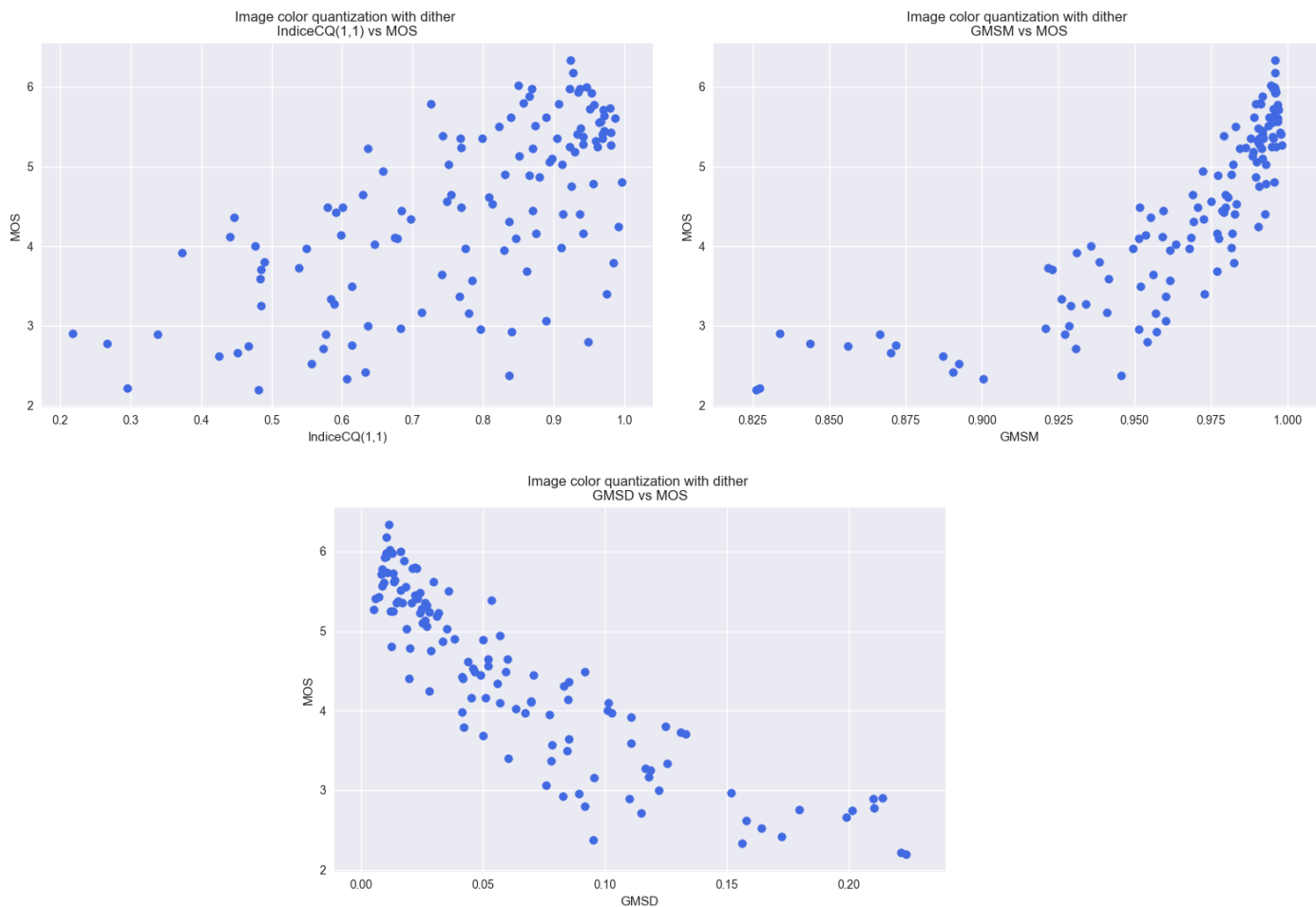


Figura C.15: Distorsión #22: Image color quantization with dither

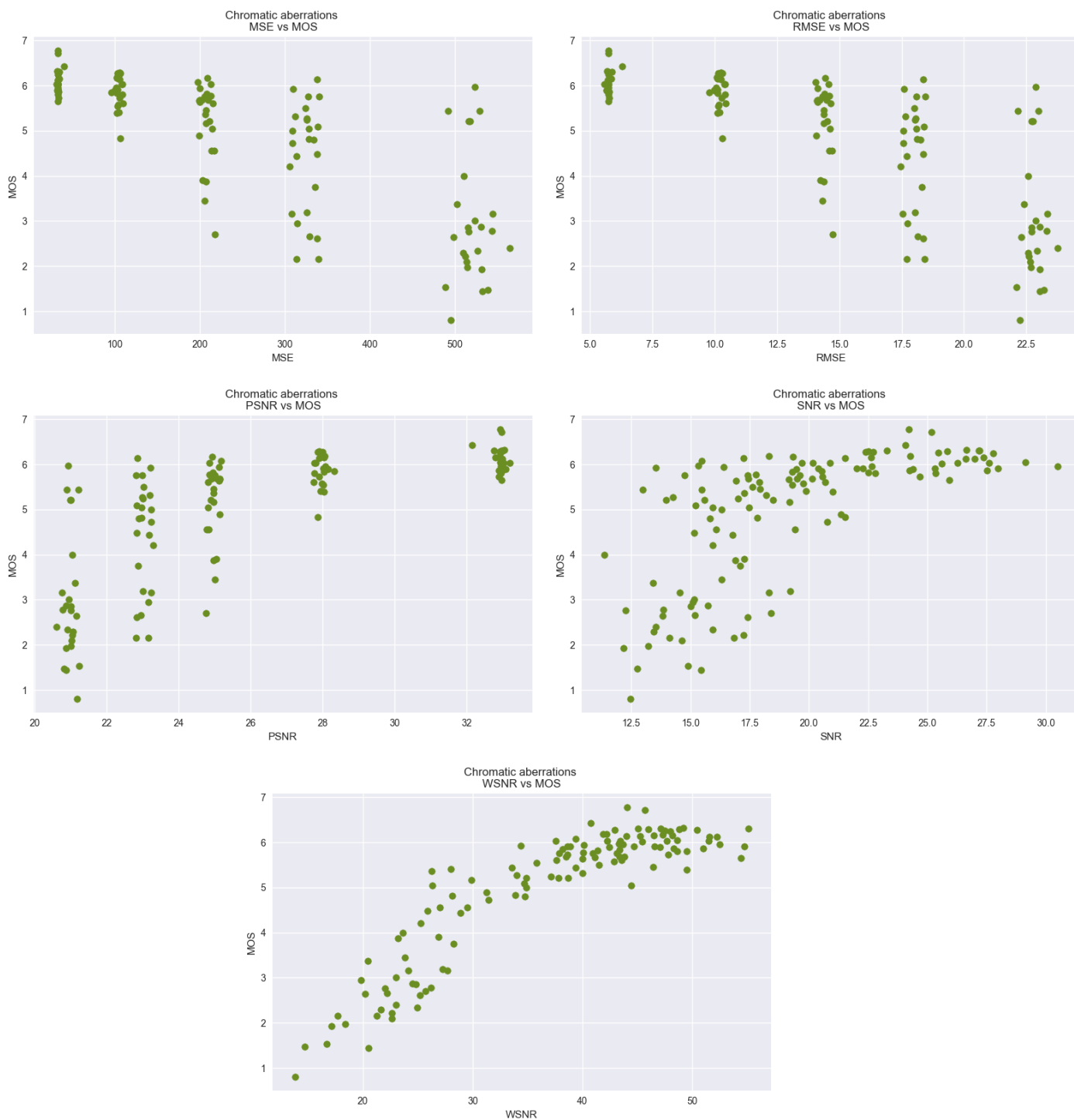


Figura C.16: Distorsión #23: Chromatic aberrations

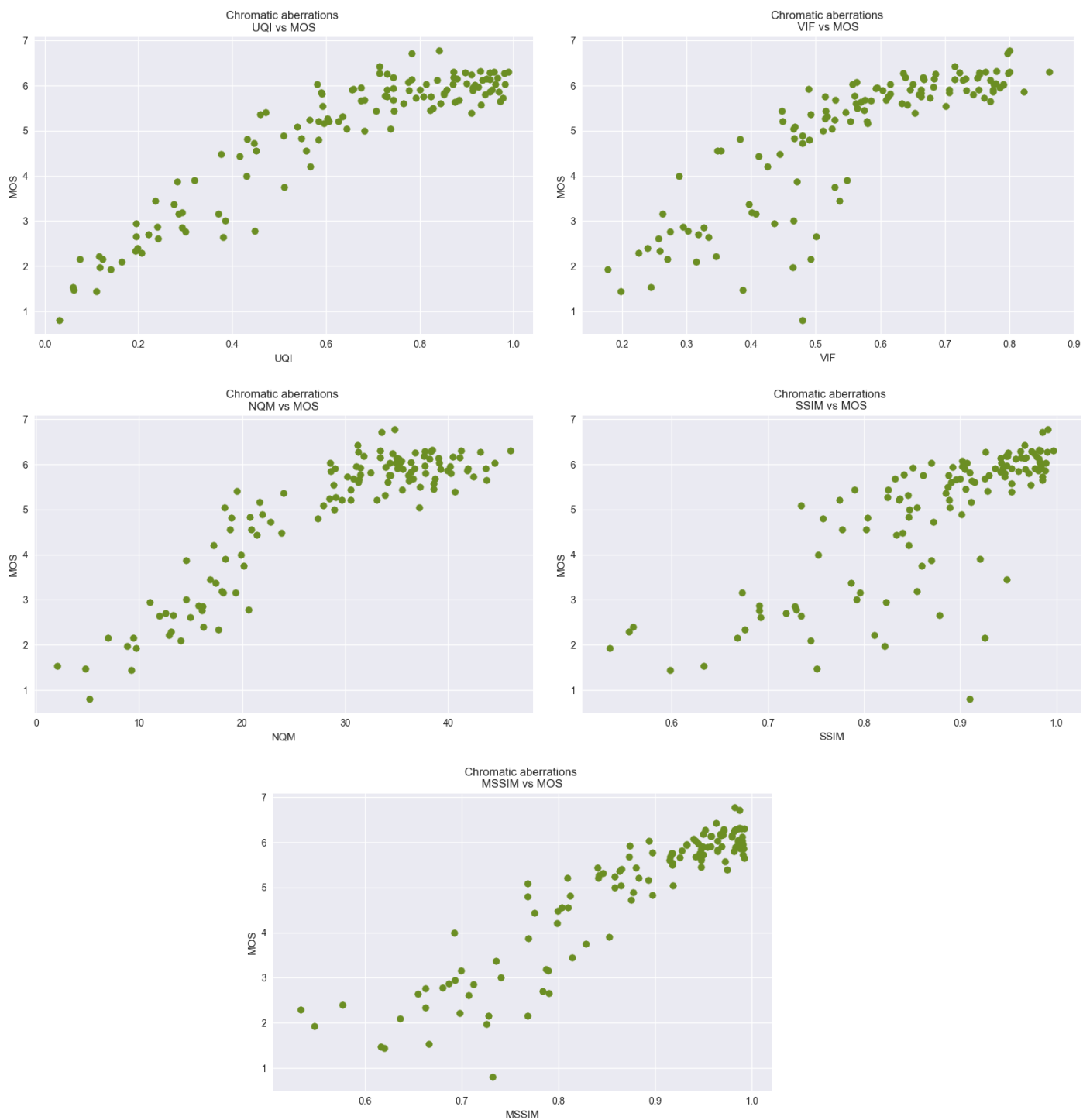


Figura C.17: Distorsión #23: Chromatic aberrations

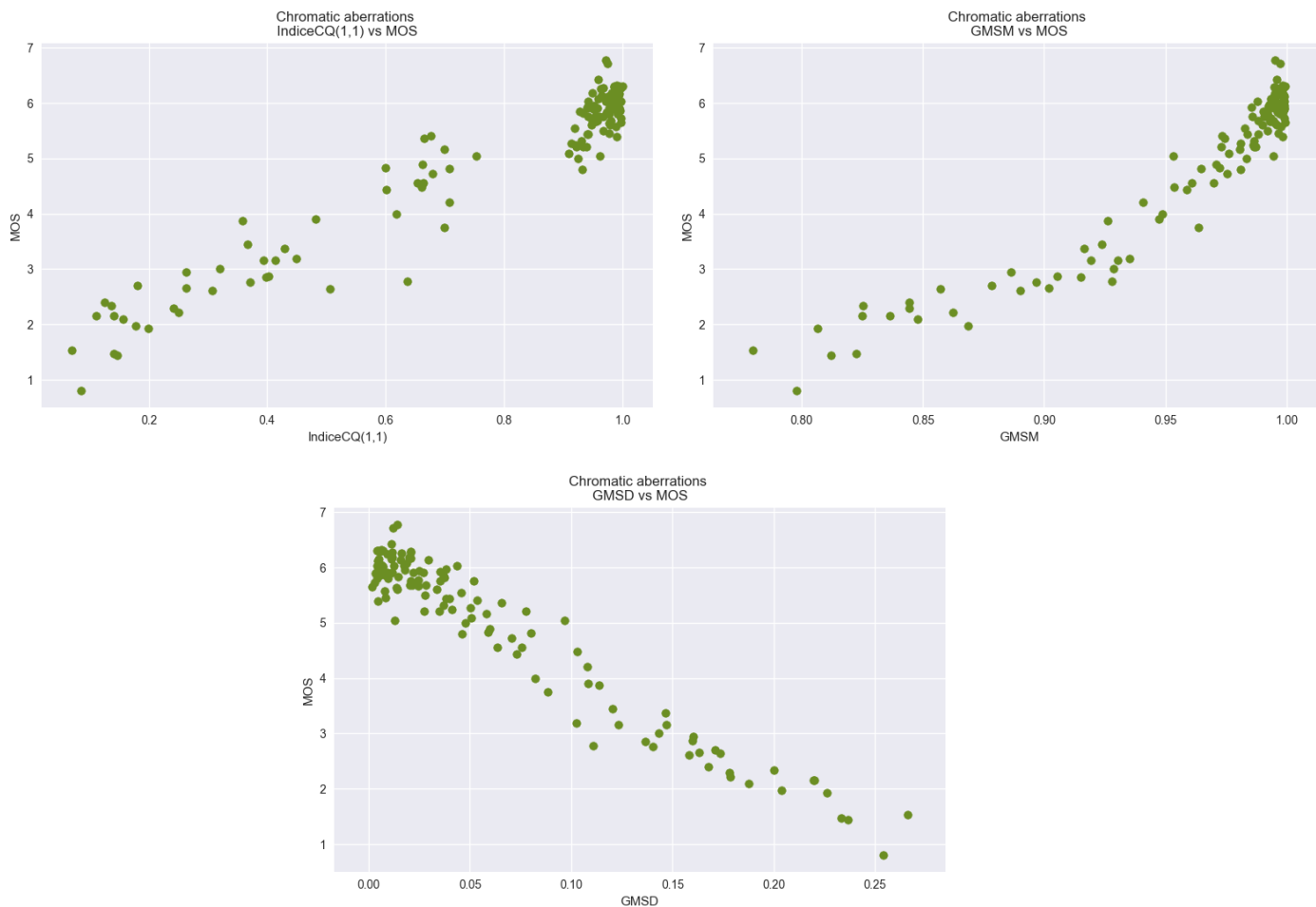


Figura C.18: Distorsión #23: Chromatic aberrations

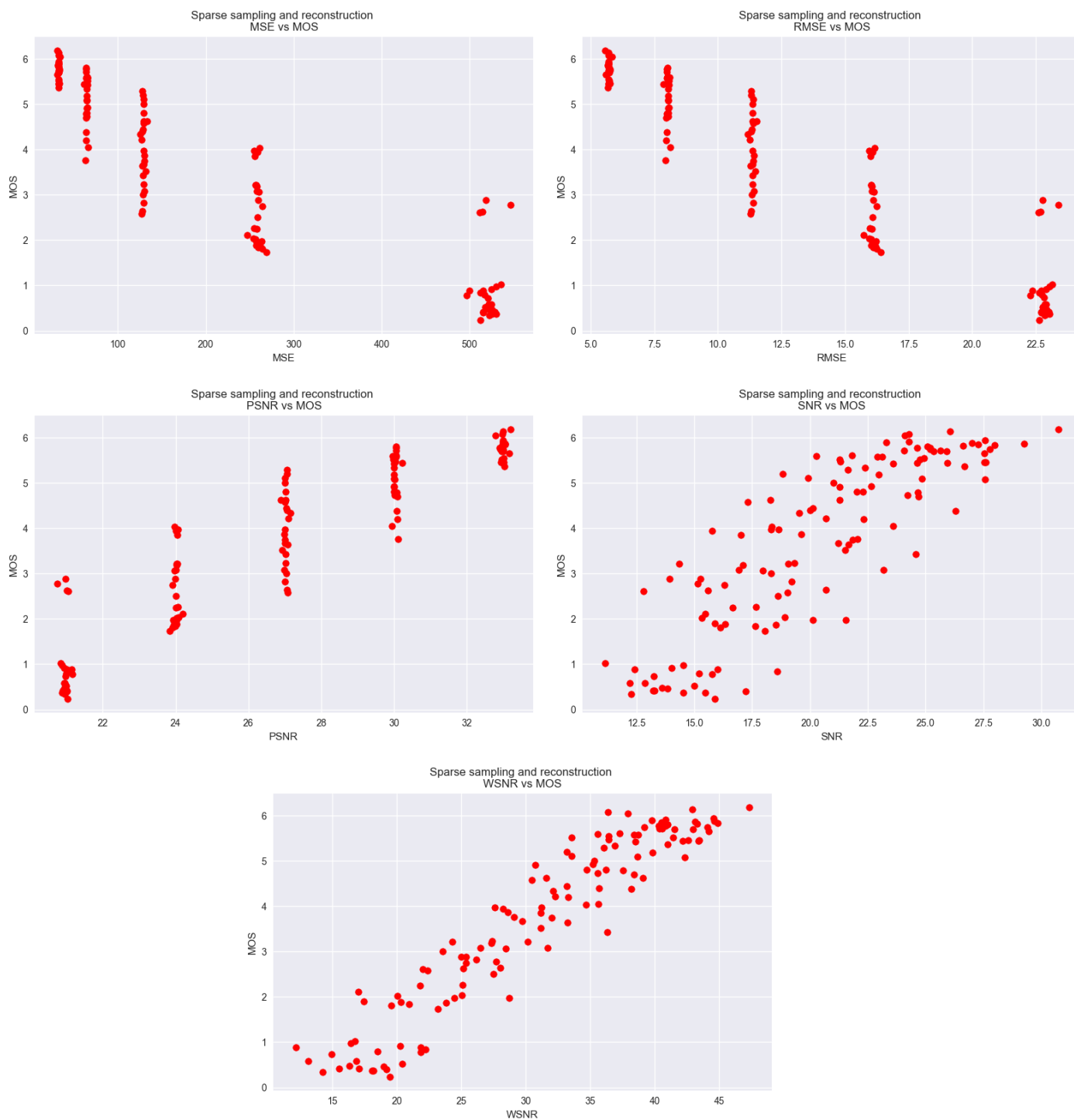


Figura C.19: Distorsión #24: Sparse sampling and reconstruction

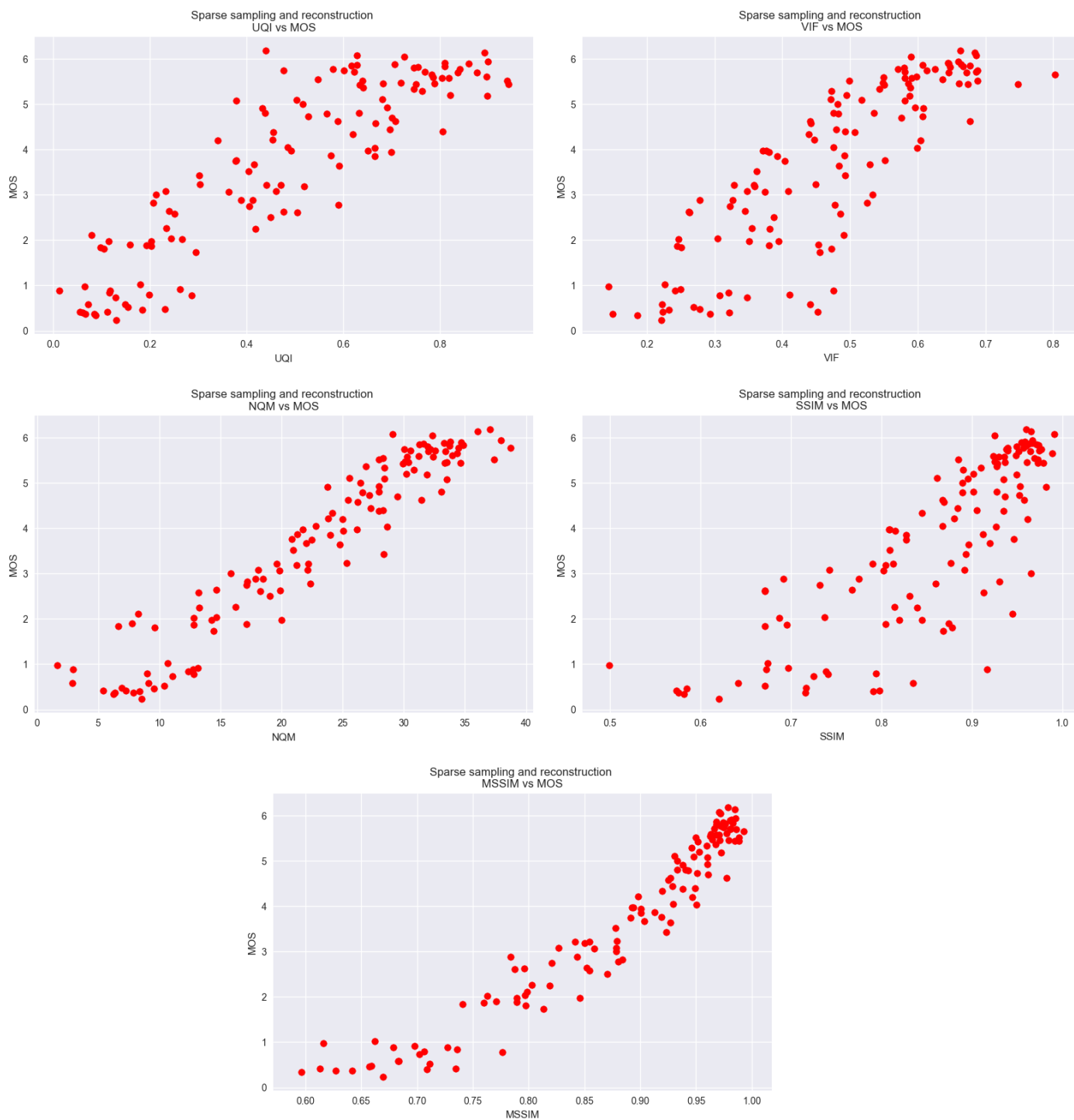


Figura C.20: Distorsión #24: Sparse sampling and reconstruction

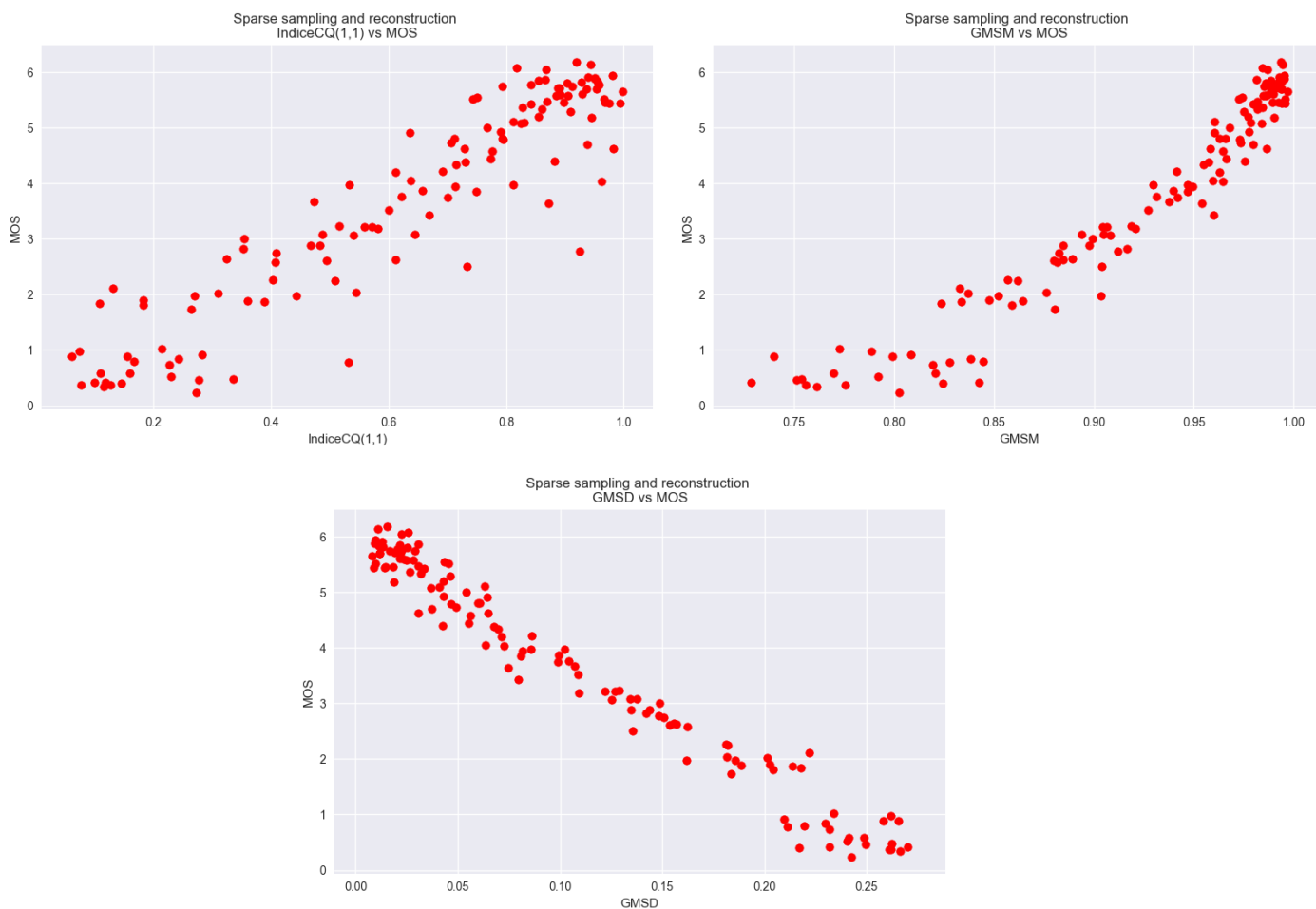


Figura C.21: Distorsión #24: Sparse sampling and reconstruction

---

## Apéndice D

# Códigos en Python para obtener los coeficientes de correlación

En este Apéndice desarrollo los códigos que fueron utilizados para obtener los coeficientes de correlación presentados en el Capítulo 5.

Fueron necesarios los siguientes paquetes:

```
import numpy as np
from scipy import stats
```

### D.1. Función que le asigna a cada valor su rango ordenándolos de mayor a menor

```
def rank(array):

    lis = array.tolist()

    order = array.argsort()

    rank = np.empty((array.shape[0],1))

    r = 1
    for elem in order:
        rank[elem,] = r
        r = r + 1
```



```
for elem in lis:
    c = lis.count(elem)
    if c > 1:
        v = 0
        for x in xrange(c):
            p = lis.index(elem)
            lis[p] = -10001
            v = v + rank[p,]
        v = v/c
    for elem in lis:
        if elem == -10001:
            p = lis.index(elem)
            lis[p] = -(p+1000)*v
            rank[p,] = v

return rank
```

## D.2. Función que calcula la correlacion de Spearman

```
def rho(mos, metric):

    rank_mos = rank(mos)
    rank_metric = rank(metric)

    dc = (rank_mos - rank_metric)**2

    n = mos.shape[0]

    rho = 1 - 6*sum(dc)/(n*(n**2 - 1))

    t = abs(rho[0])/(np.sqrt((1-rho[0]**2)/(n-2)))

    df = n-2

    T = stats.t(df)
```

```
p = 2*(1 - T.cdf(t))  
  
return (rho[0], p)
```

### D.3. Función que calcula el total de pares concordantes y el total de pares discordantes

```
def con_discon(rank1, rank2):  
    n = rank1.shape[0]  
    SC = 0  
    SD = 0  
    for i in xrange(n):  
        C = 0  
        D = 0  
        for j in xrange(i+1, n):  
            if (rank1[i] < rank1[j] and rank2[i] < rank2[j]) or  
                (rank1[i] > rank1[j] and rank2[i] > rank2[j]):  
                C = C + 1  
            elif (rank1[i] < rank1[j] and rank2[i] > rank2[j]) or  
                (rank1[i] > rank1[j] and rank2[i] < rank2[j]):  
                D = D + 1  
        SC = SC + C  
        SD = SD + D  
    return (SC, SD)
```

### D.4. Función que cuenta la cantidad de pares enlazados que tiene cada grupo de pares enlazados

```
def count(rank):  
    g = []  
    lisrank = rank.tolist()  
    for elem in lisrank:  
        i = lisrank.count(elem)  
        if i>1 :
```

```
        g.append(i)
        for x in xrange(i):
            lisrank.remove(elem)

    return g
```

Esta función devuelve una lista cuyos elementos son el total de pares enlazados de cada grupo de pares enlazados. El tamaño de la lista es la cantidad de grupos de pares enlazados.

## D.5. Función que calcula el coeficiente de correlación de Kendall

```
def tauB(mos, metric):

    rank_mos = rank(mos)
    rank_metric = rank(metric)

    C = con_discon(rank_mos, rank_metric)[0]
    D = con_discon(rank_mos, rank_metric)[1]

    n = mos.shape[0]
    n0 = n*(n-1)/2

    t = count(rank_mos)
    u = count(rank_metric)

    n1 = 0
    for elem in t:
        n1 = n1 + elem*(elem - 1)/2

    n2 = 0
    for elem in u:
        n2 = n2 + elem*(elem - 1)/2

    tau = (C - D)/(np.sqrt((n0-n1)*(n0-n2)))

    v0 = n*(n-1)*(2*n+5)
```

```
vt = 0
for elem in t:
    vt = vt + elem*(elem - 1)*(2*elem + 5)

vu = 0
for elem in u:
    vu = vu + elem*(elem - 1)*(2*elem + 5)

v1 = (2*n1*2*n2)/(2*n*(n-1))

n11 = 0
for elem in t:
    n11 = n11 + elem*(elem - 1)*(elem - 2)

n22 = 0
for elem in u:
    n22 = n22 + elem*(elem - 1)*(elem - 2)

v2 = (n11*n22)/(9*n*(n-1)*(n-2))

v = (v0 - vt - vu)/(18 + v1 + v2)

zB = abs(C - D)/np.sqrt(v)

mu = 0
sigma = 1
normal = stats.norm(mu, sigma)

p = 2*(1 - normal.cdf(zB))

return (tau, p)
```

El paquete **stats** de **scipy** trae ya implementadas las funciones para calcular los coeficientes de Spearman y de Kendall con sus respectivas pruebas: *stats.spearmanr()* y *stats.kendalltau()*, sin embargo no hay ningún paquete que traiga una función que calcule el coeficiente w de concordancia de Kendall, es por eso que implementé nuevamente las funciones de Spearman y Kendall, que si bien fueron menos eficientes que las de **scipy.stats** obtuve los mismos resultados, para poder en base a ellas desarrollar la siguiente función.

## D.6. Función que calcula el coeficiente de concordancia w de Kendall

```
def wk(*items):

    ranks = []
    m = 0
    for item in items:
        m = m + 1
        ranks.append(rank(item))

    n = items[0].shape[0]

    Ris = []
    for i in xrange(n):
        Ri = 0
        for elem in ranks:
            Ri = Ri + elem[i]
        Ris.append(Ri)

    Rbar = sum(Ris)/n

    S = 0
    for Ri in Ris:
        S = S + (Ri - Rbar)**2

    L = 0
    for elem in ranks:
        g = count(elem)
        Lj = 0
        for t in g:
            Lj = Lj + (t**3 - t)
        L = L + Lj

    w = 12*S/((m**2)*(n**3 - n) - m*L)

    c2 = m*(n-1)*w
```

```
df = n-1

chi2 = stats.chi2(df)

p = 1 - chi2.cdf(c2)

return (w, p)
```

## Apéndice E

# Matrices de Confusión para cada grupo de distorsiones

En este apéndice se presentan las matrices de confusión de cada métrica por cada uno de los seis grupos de distorsiones. Para cada grupo de distorsiones se estableció un ranking de las métricas que clasificaron correctamente el mayor número de imágenes según la calidad de las mismas.

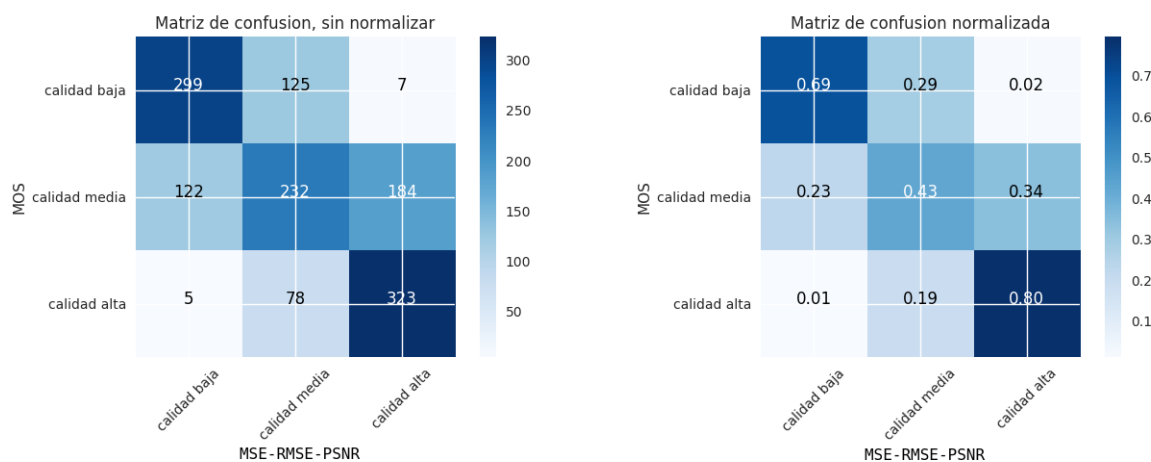


Figura E.1: Grupo 1: Distorsiones Ruidosas. A la izquierda se encuentran la matriz de confusión para las métricas MSE, RMSE y PSNR, a la derecha, la matriz de confusión normalizada.

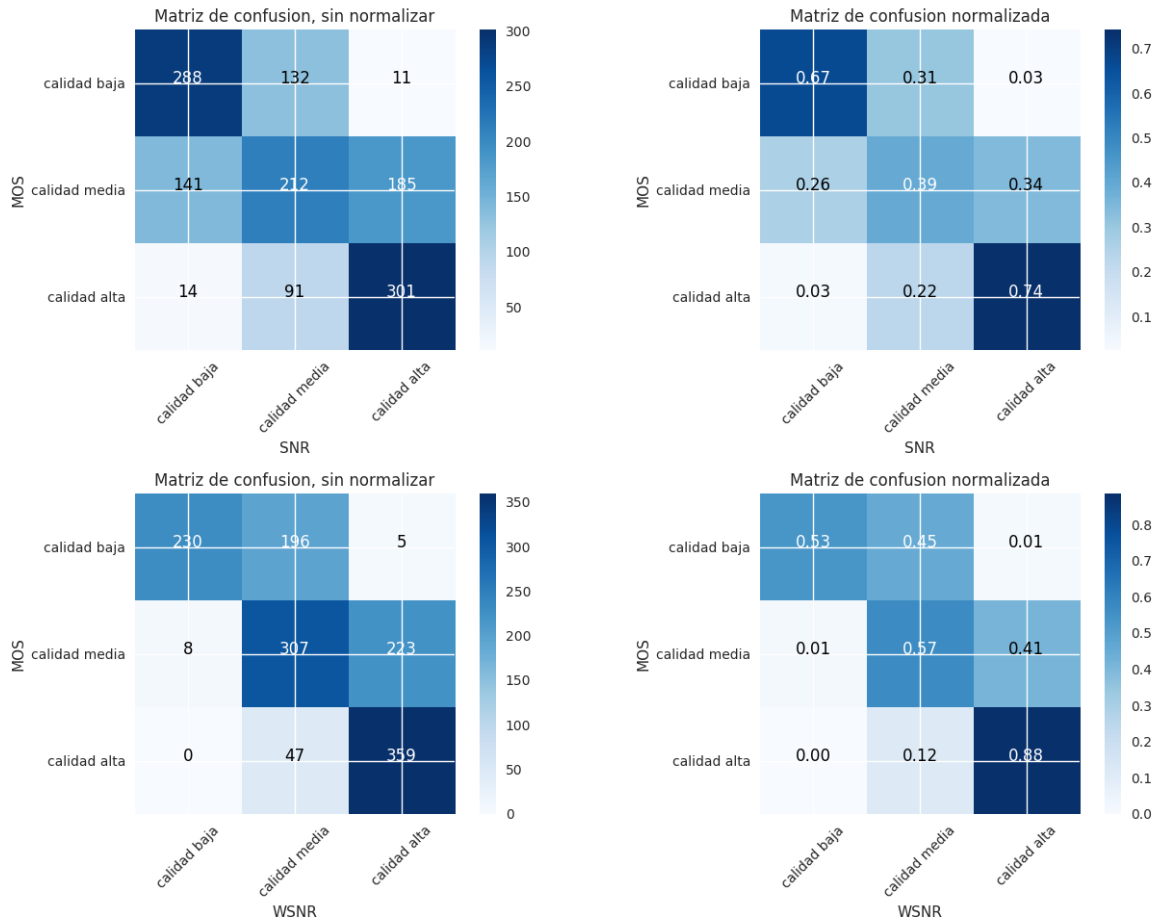


Figura E.2: Grupo 1: Distorsiones Ruidosas. Métricas SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	69 %	43 %	80 %
SNR	67 %	39 %	74 %
WSNR	53 %	57 %	88 %
UQI	73 %	49 %	37 %
VIF	83 %	51 %	39 %
NQM	61 %	60 %	84 %
SSIM	78 %	43 %	48 %
MSSIM	83 %	54 %	69 %
CQ(1,1)	66 %	47 %	61 %
GMSM	92 %	61 %	72 %
GMSD	88 %	74 %	82 %

Cuadro E.1: Grupo 1: Distorsiones Ruidosas. Porcentaje de imágenes bien clasificadas por métrica, según calidad.



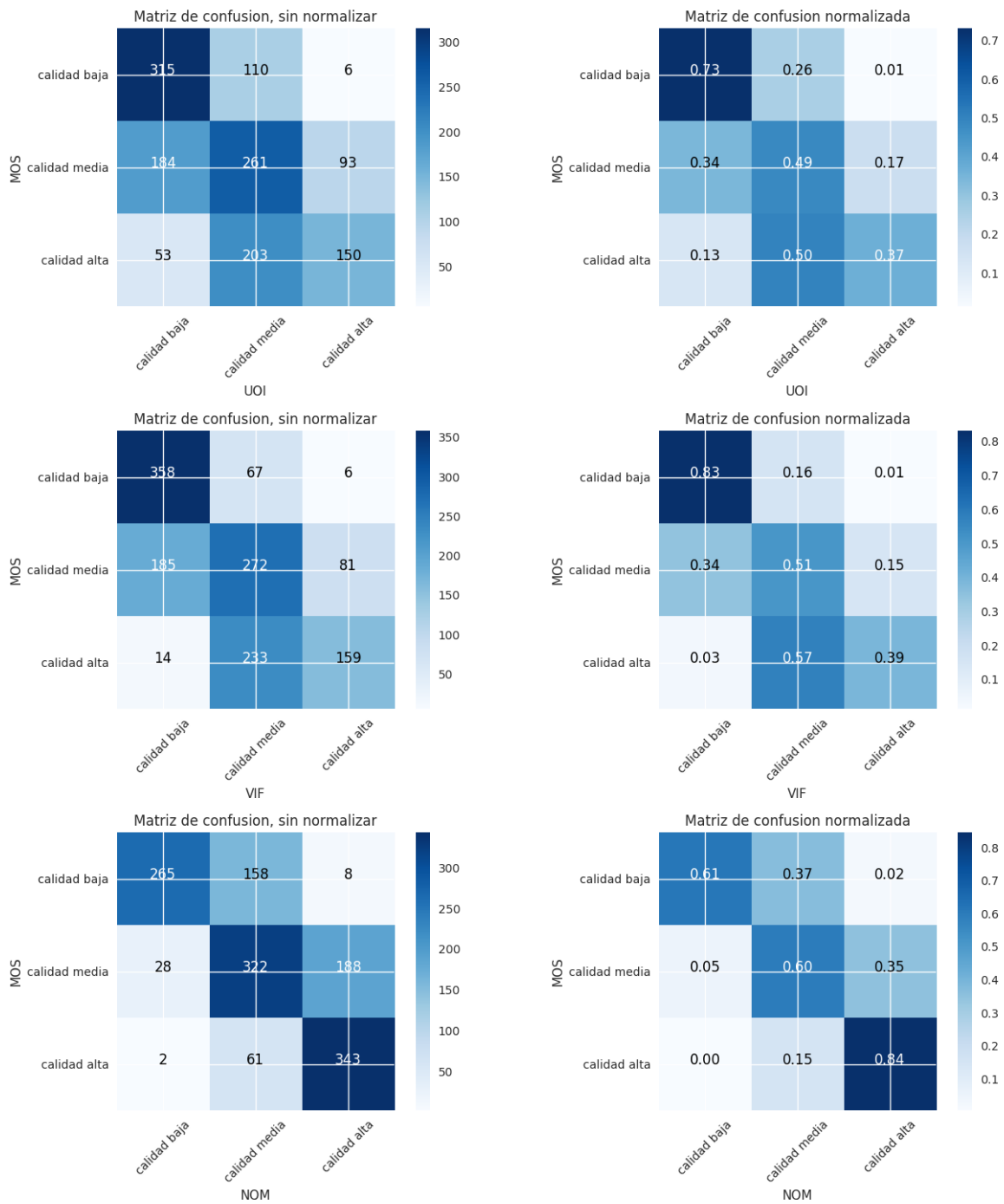


Figura E.3: Grupo 1: Distorsiones Ruidosas. Métricas UOI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	GMSM	GMSD	WSNR
Segundas	GMSD	GMSM	NQM
Terceras	MSSIM & VIF	NQM	GMSD
Cuartas	SSIM	WSNR	MSE & RMSE & PSNR
Quintas	UOI	MSSIM	SNR

Cuadro E.2: Grupo 1: Distorsiones Ruidosas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

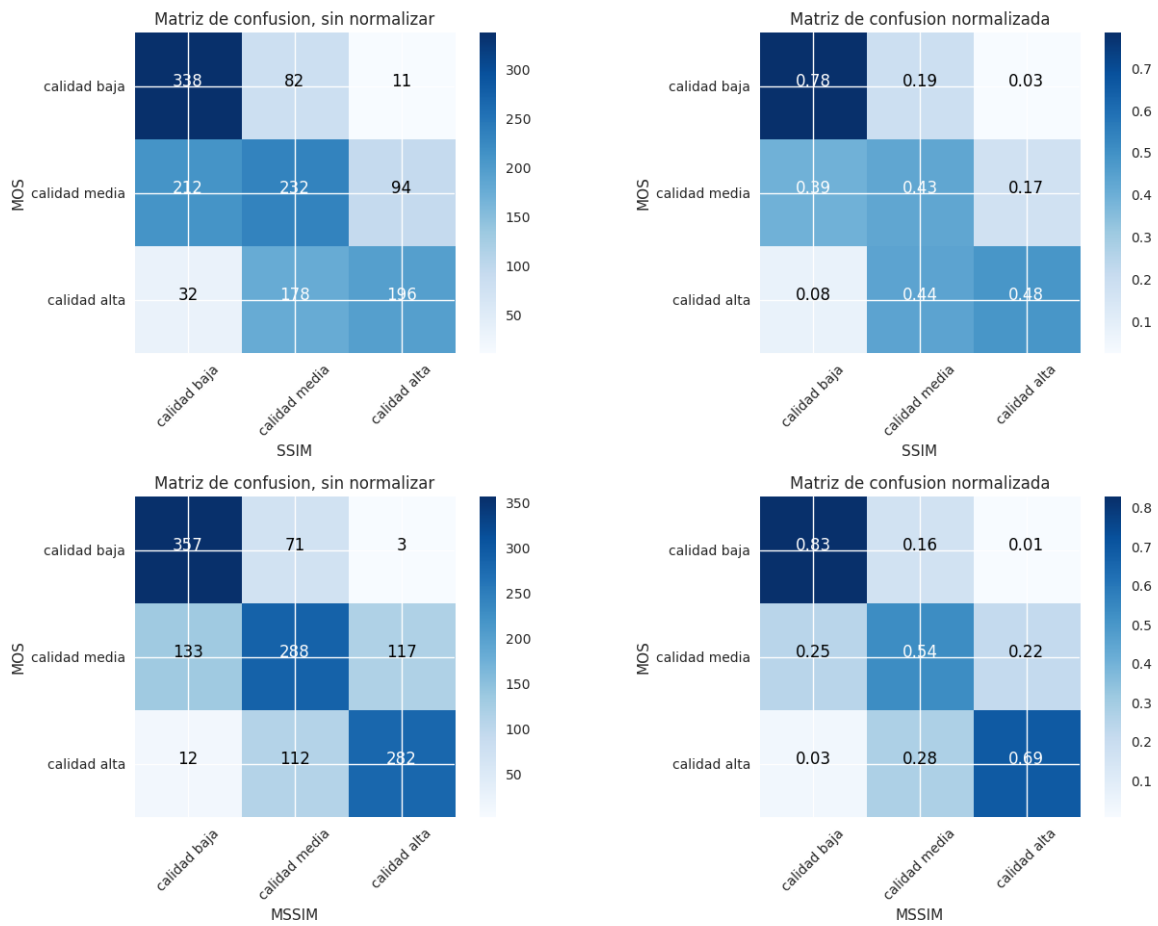


Figura E.4: Grupo 1: Distorsiones Ruidosas. Métricas SSIM y MSSIM.

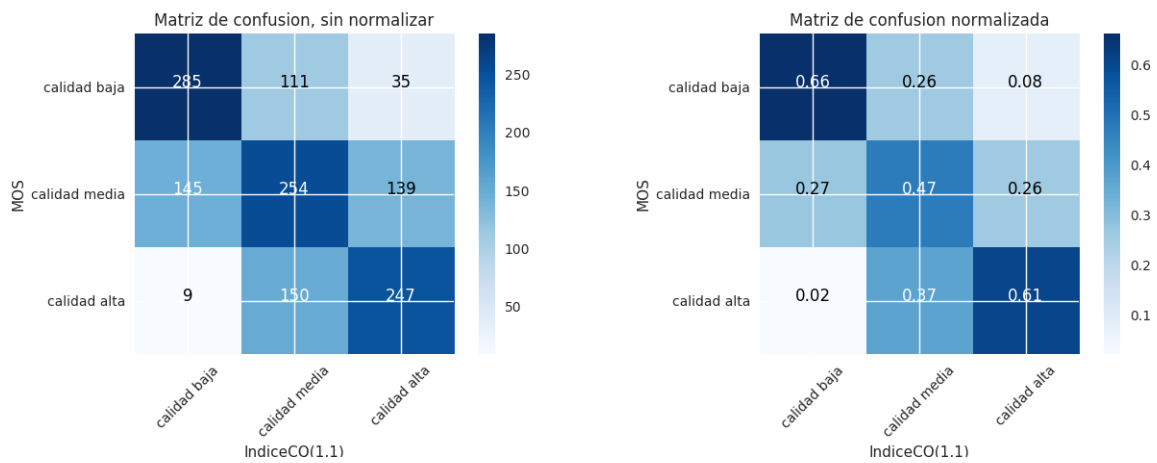


Figura E.5: Grupo 1: Distorsiones Ruidosas. Índice CQ(1,1).

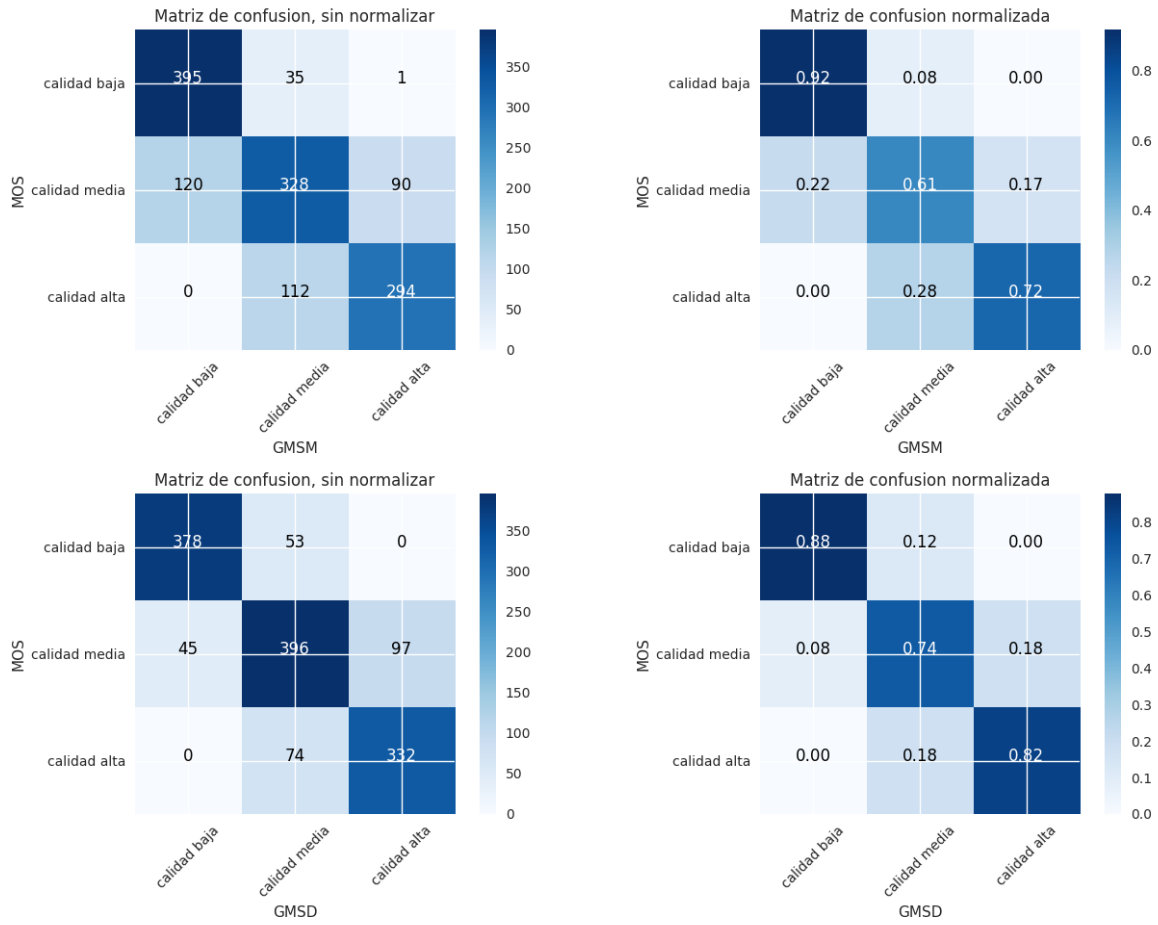


Figura E.6: Grupo 1: Distorsiones Ruidosas. Métricas GSM y GMSD.

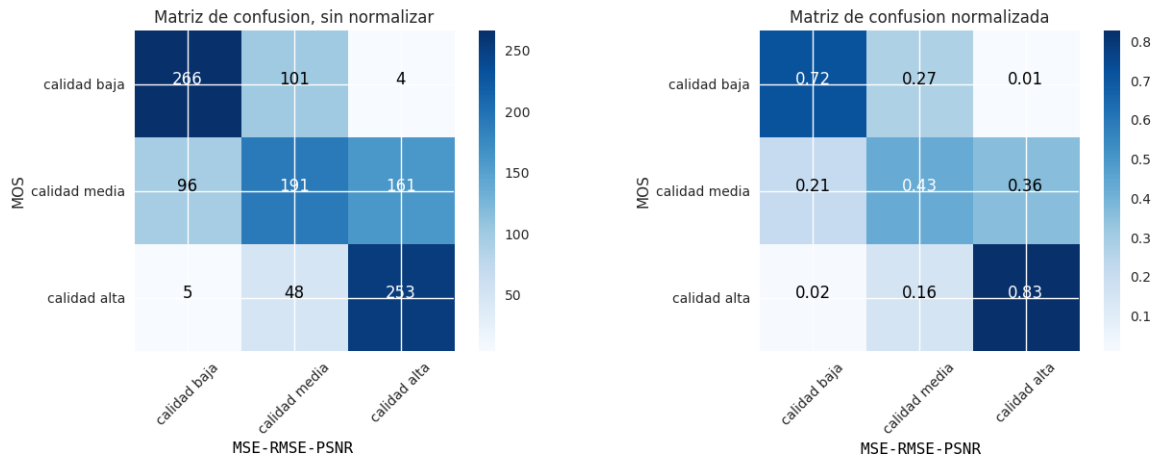


Figura E.7: Grupo 2: Distorsiones Actuales. A la izquierda se encuentran la matriz de confusión para las métricas MSE, RMSE y PSNR, a la derecha, la matriz de confusión normalizada.

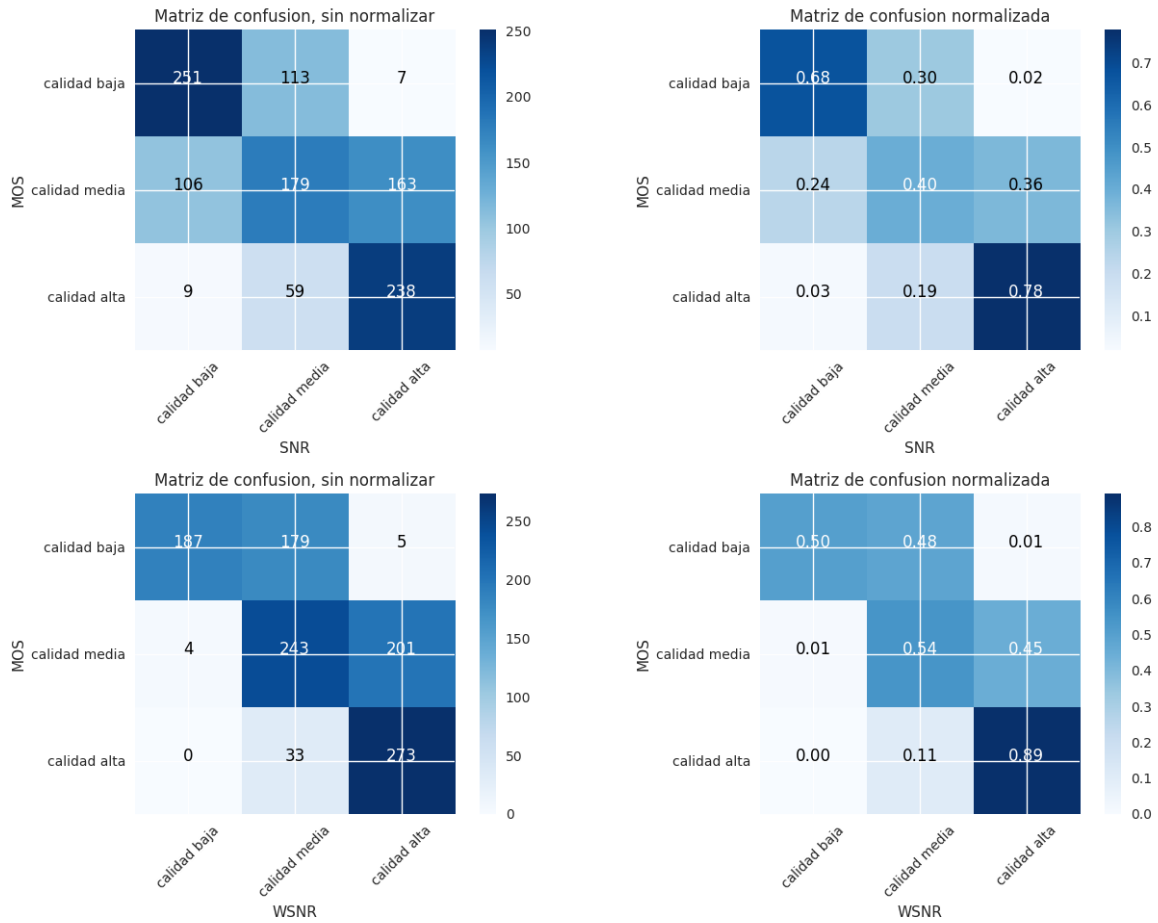


Figura E.8: Grupo 2: Distorsiones Actuales. Métrica SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	72 %	43 %	83 %
SNR	68 %	40 %	78 %
WSNR	50 %	54 %	89 %
UQI	74 %	49 %	42 %
VIF	85 %	53 %	44 %
NQM	58 %	59 %	86 %
SSIM	78 %	45 %	57 %
MSSIM	84 %	56 %	81 %
CQ(1,1)	72 %	45 %	54 %
GMSM	92 %	58 %	70 %
GMSD	89 %	73 %	79 %

Cuadro E.3: Grupo 2: Distorsiones Actuales. Porcentaje de imágenes bien clasificadas por métrica, según calidad.

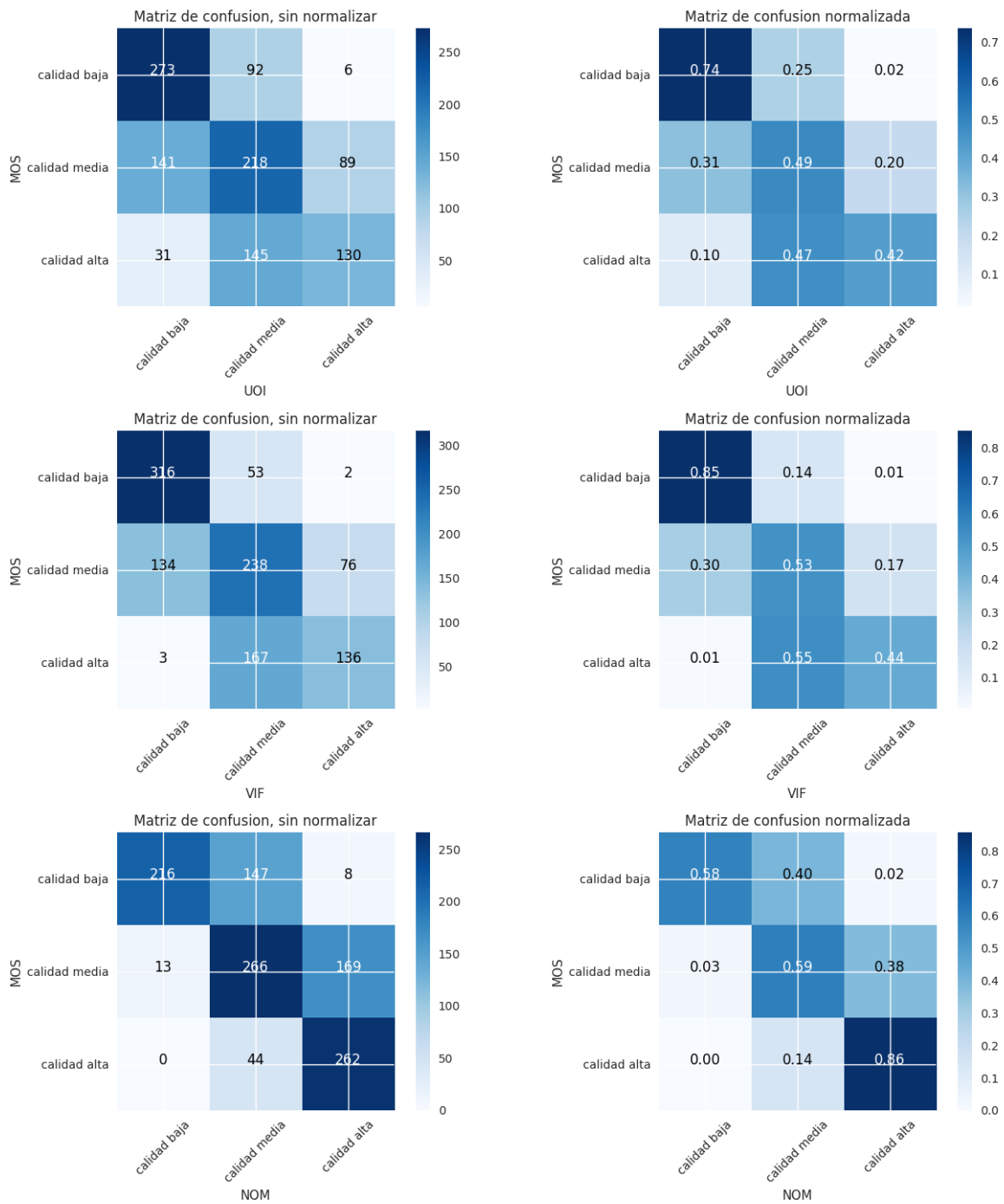


Figura E.9: Grupo 2: Distorsiones Actuales. Métricas UOI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	GMSM	GMSD	WSNR
Segundas	GMSD	NQM	NQM
Terceras	VIF	GMSM	MSE & RMSE & PSNR
Cuartas	MSSIM	MSSIM	GMSD
Quintas	SSIM	WSNR	SNR

Cuadro E.4: Grupo 2: Distorsiones Actuales. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

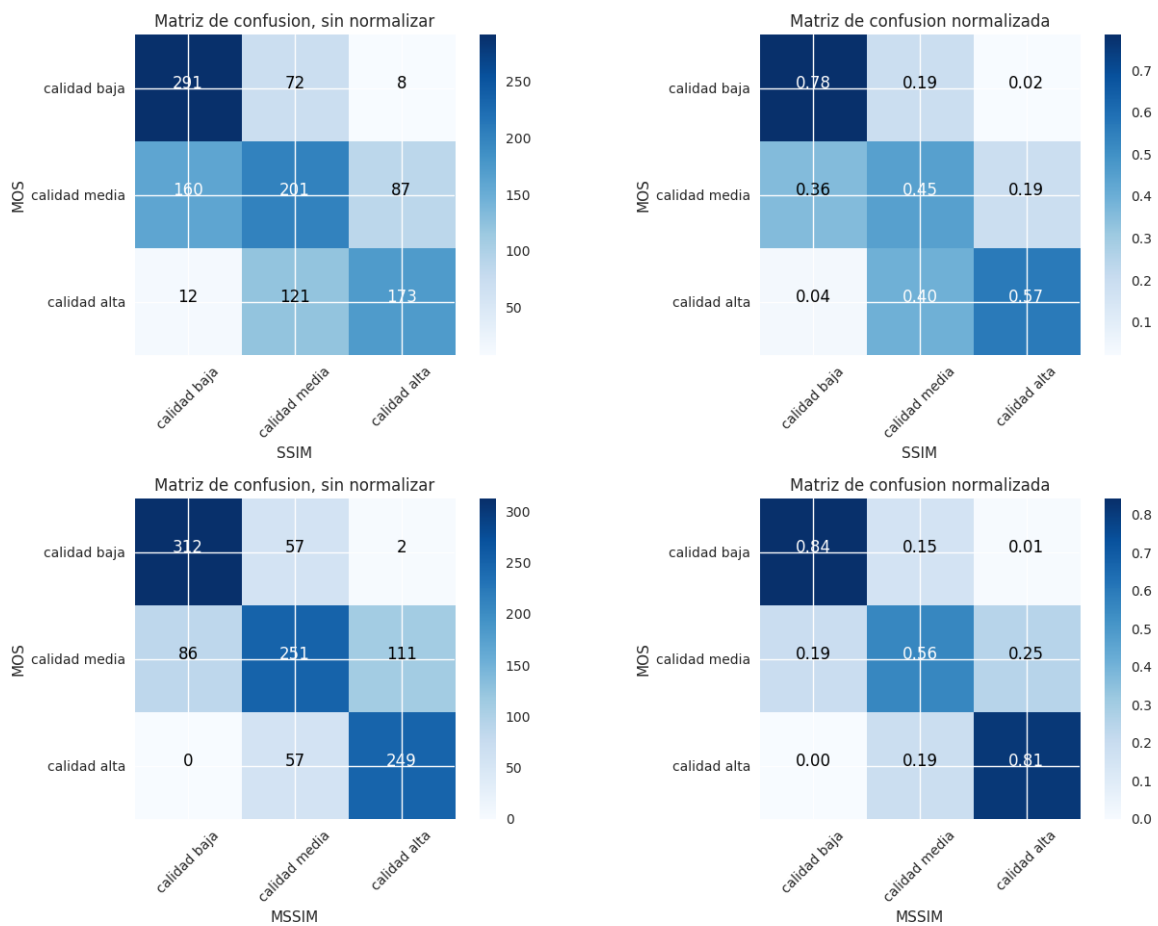


Figura E.10: Grupo 2: Distorsiones Actuales. Métricas SSIM y MSSIM.

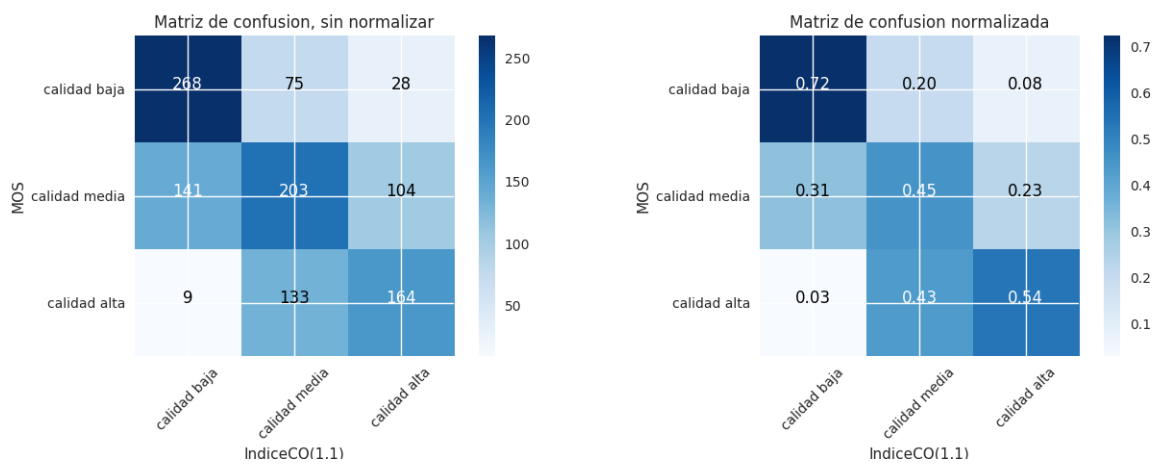


Figura E.11: Grupo 2: Distorsiones Actuales. Índice CQ(1,1).

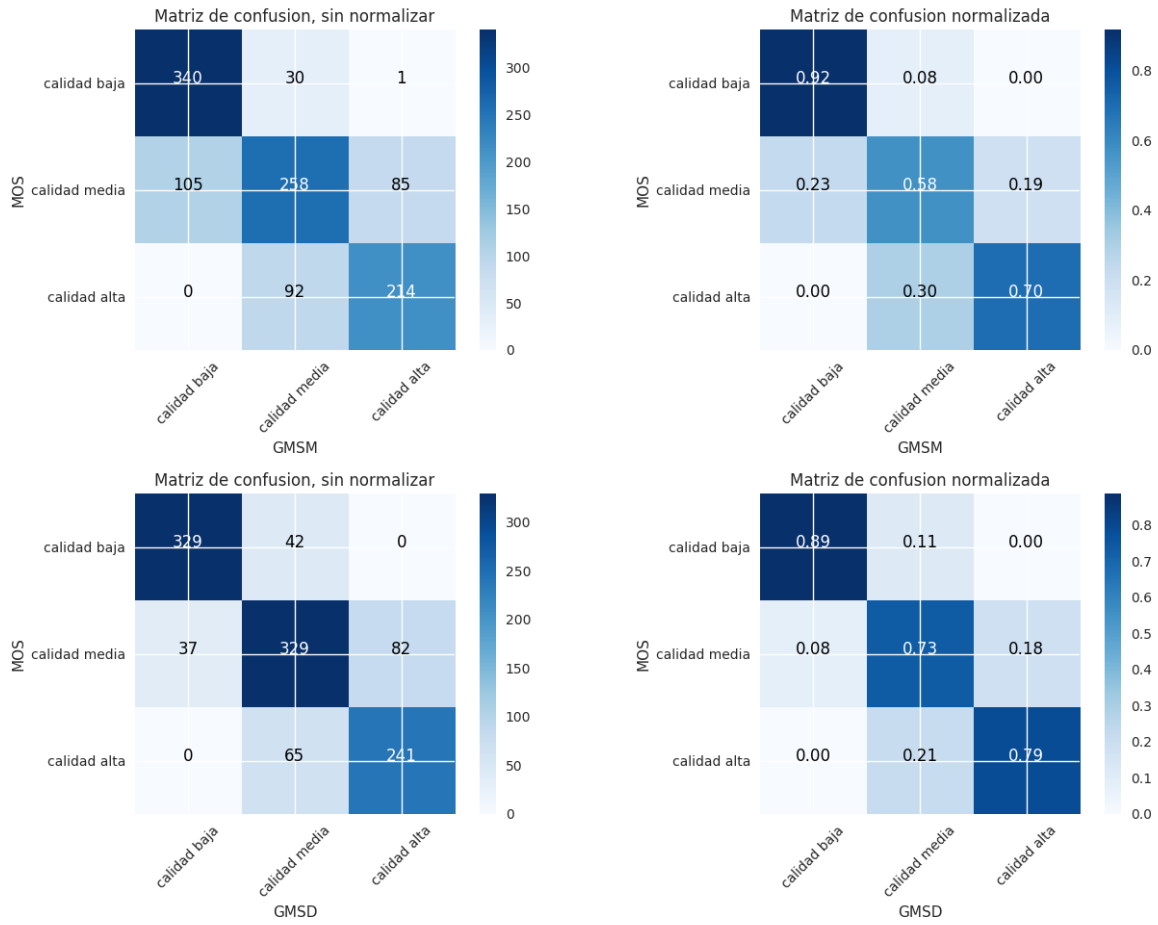


Figura E.12: Grupo 2: Distorsiones Actuales. Métricas GSM y GMSD.

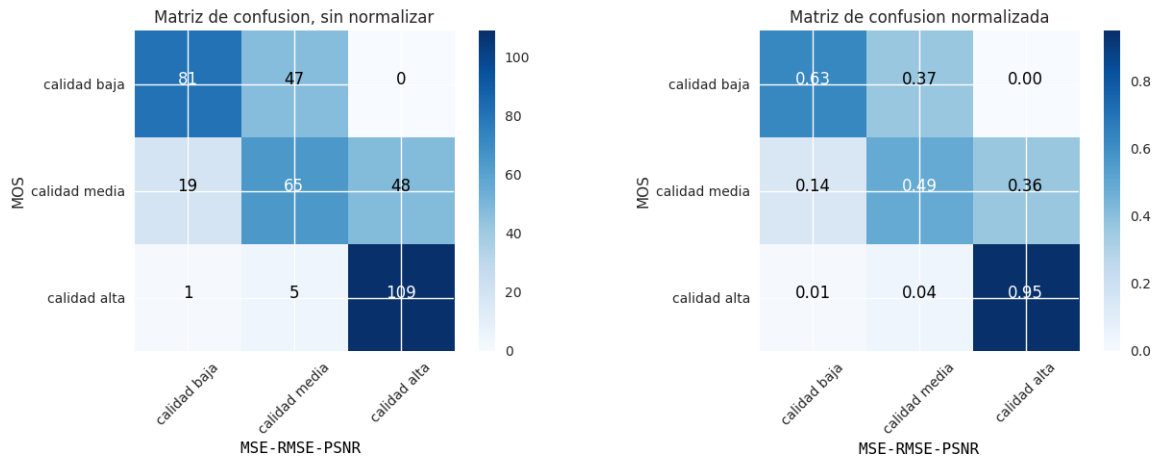


Figura E.13: Grupo 3: Distorsiones Simples. A la izquierda se encuentran la matriz de confusión para las métricas MSE, RMSE y PSNR, a la derecha, la matriz de confusión normalizada.

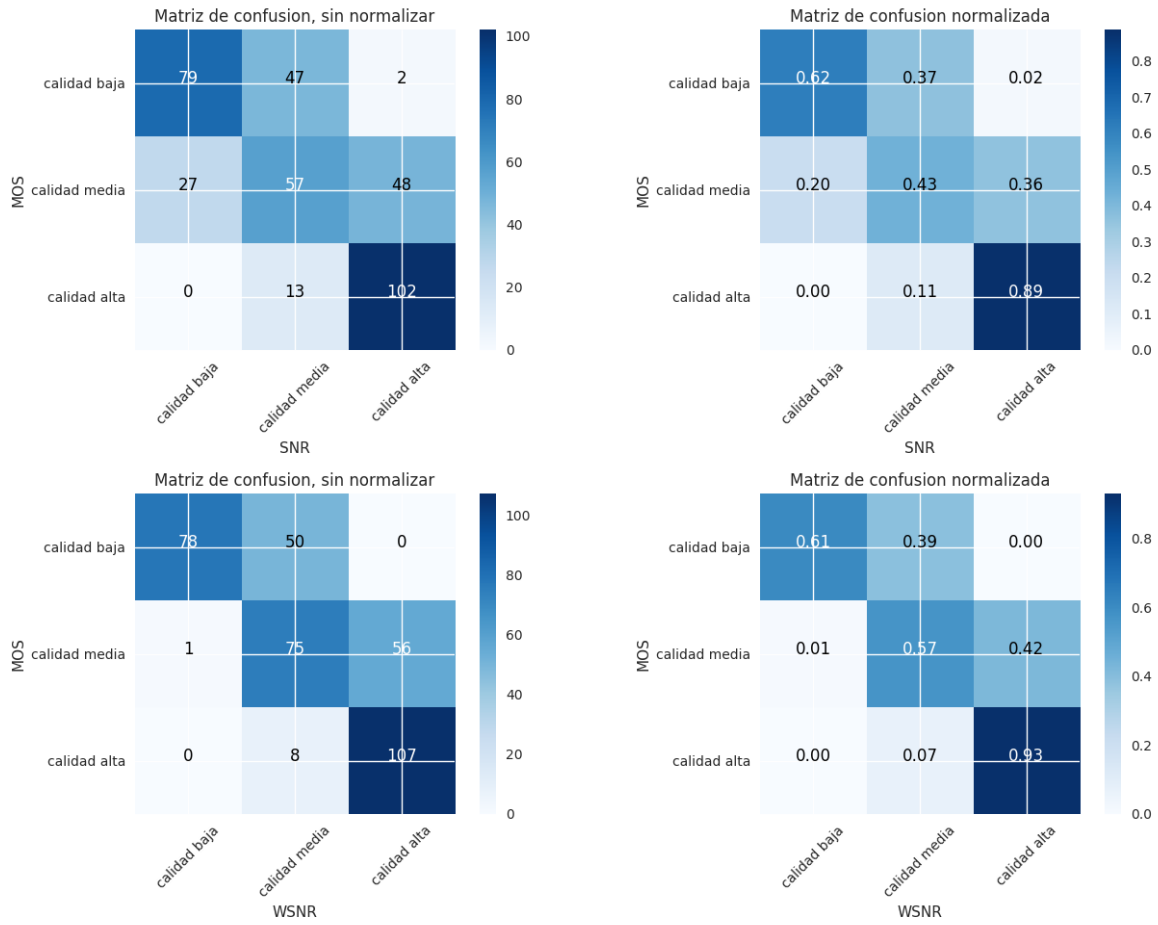


Figura E.14: Grupo 3: Distorsiones Simples. Métricas SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	63 %	49 %	95 %
SNR	62 %	43 %	89 %
WSNR	61 %	57 %	93 %
UQI	83 %	49 %	49 %
VIF	84 %	58 %	56 %
NQM	69 %	57 %	85 %
SSIM	74 %	43 %	75 %
MSSIM	86 %	57 %	88 %
CQ(1,1)	95 %	56 %	57 %
GMSM	96 %	70 %	85 %
GMSD	92 %	74 %	84 %

Cuadro E.5: Grupo 3: Distorsiones Simples. Porcentaje de imágenes bien clasificadas por métrica, según calidad.



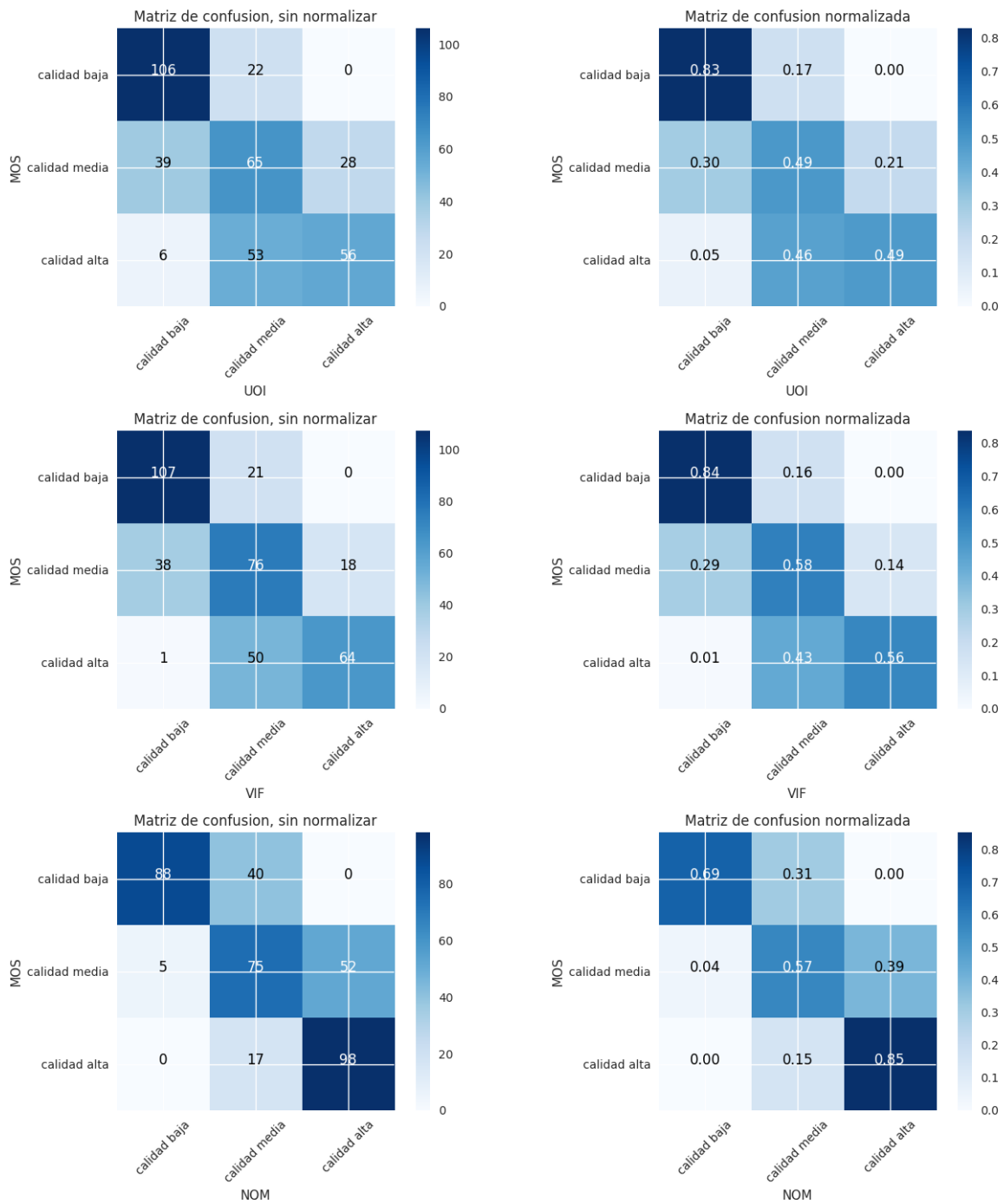


Figura E.15: Grupo 3: Distorsiones Simples. Métricas UOI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	GSM	GMSD	MSE & RMSE & PSNR
Segundas	CQ(1,1)	GSM	WSNR
Terceras	GMSD	VIF	SNR
Cuartas	MSSIM	WSNR & NQM & MSSIM	MSSIM
Quintas	VIF	CQ(1,1)	NQM & GSM

Cuadro E.6: Grupo 3: Distorsiones Simples. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

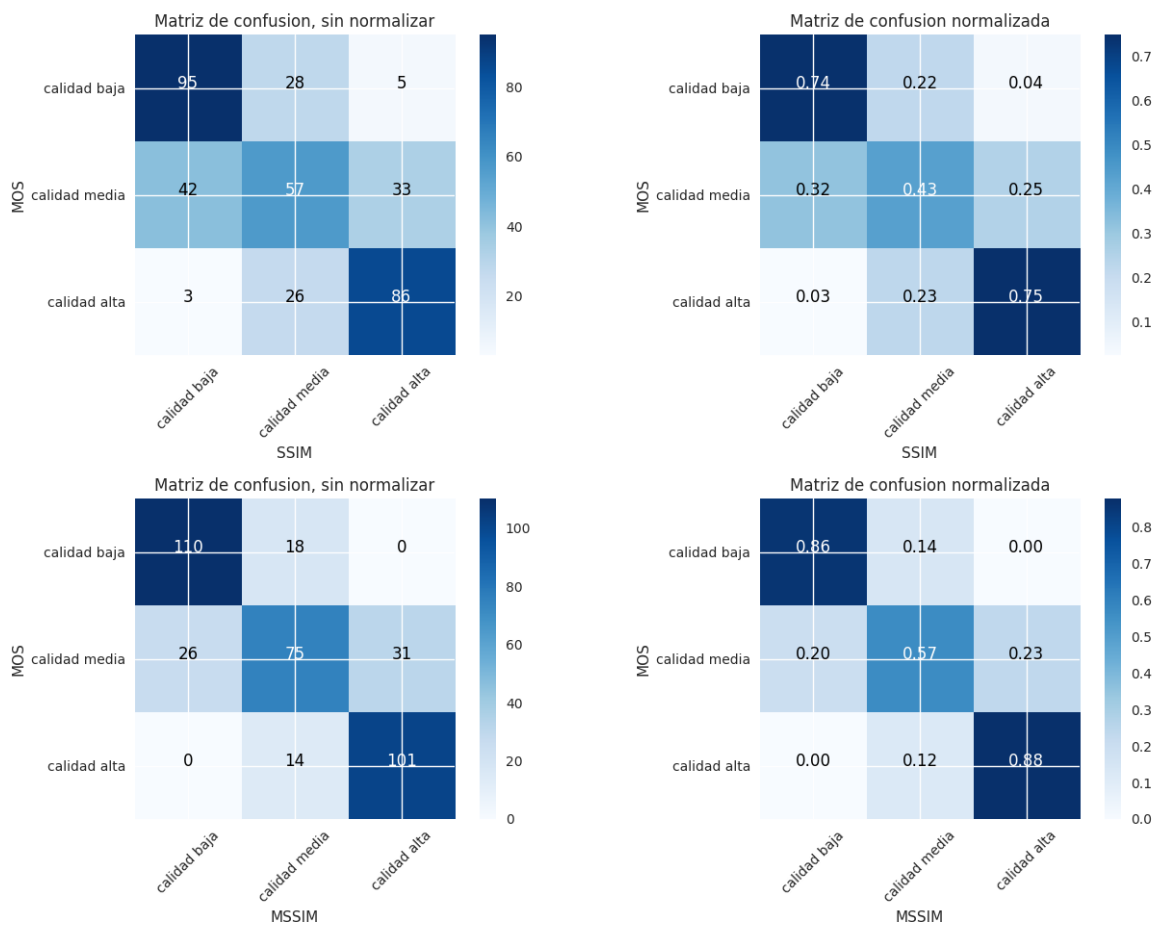


Figura E.16: Grupo 3: Distorsiones Simples. Métricas SSIM y MSSIM.

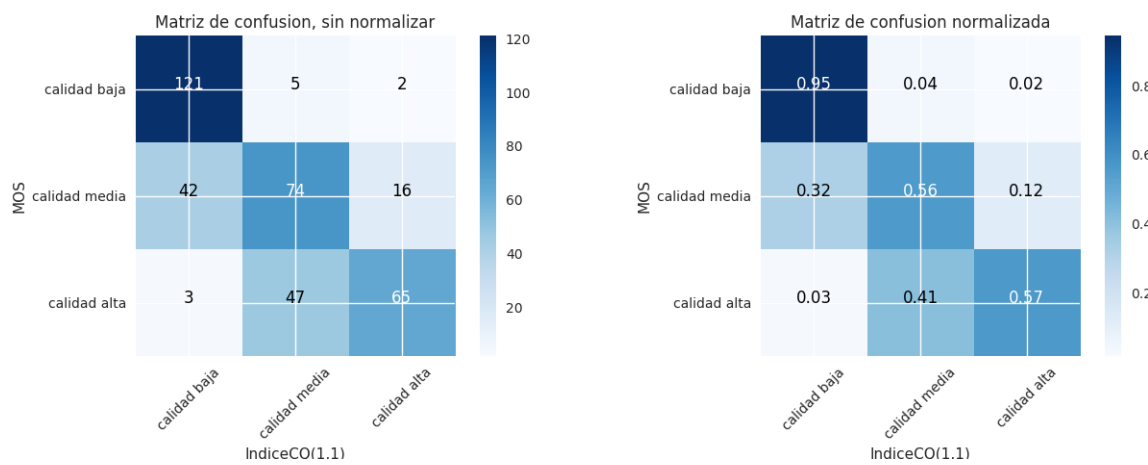


Figura E.17: Grupo 3: Distorsiones Simples. Índice CQ(1,1).

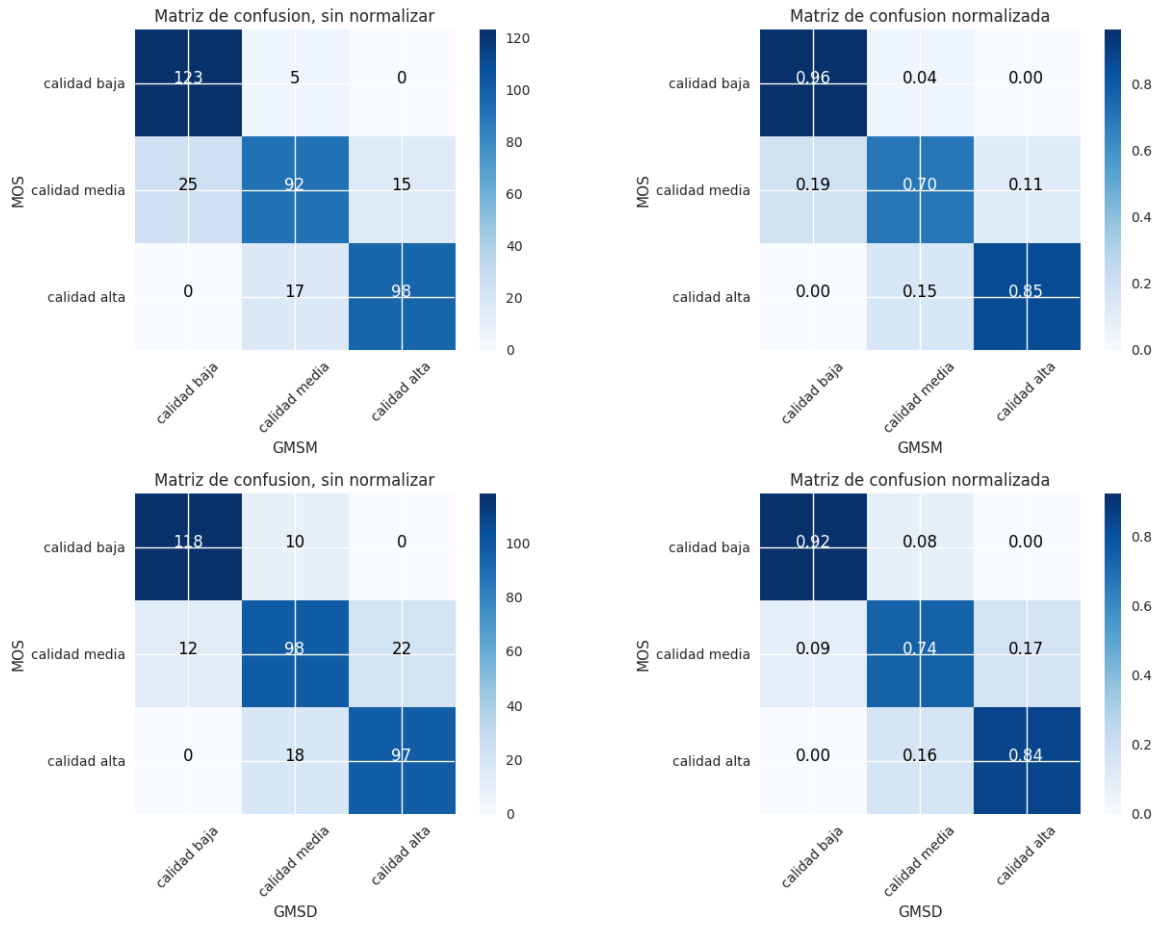


Figura E.18: Grupo 3: Distorsiones Simples. Métricas GSM y GMSD.

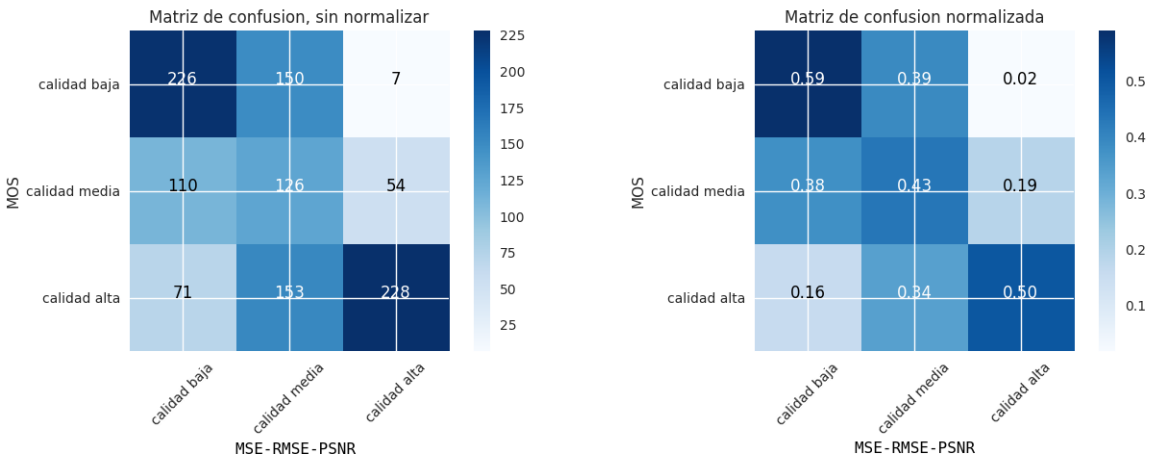


Figura E.19: Grupo 4: Distorsiones Exóticas. Métricas MSE, RMSE y PSNR.

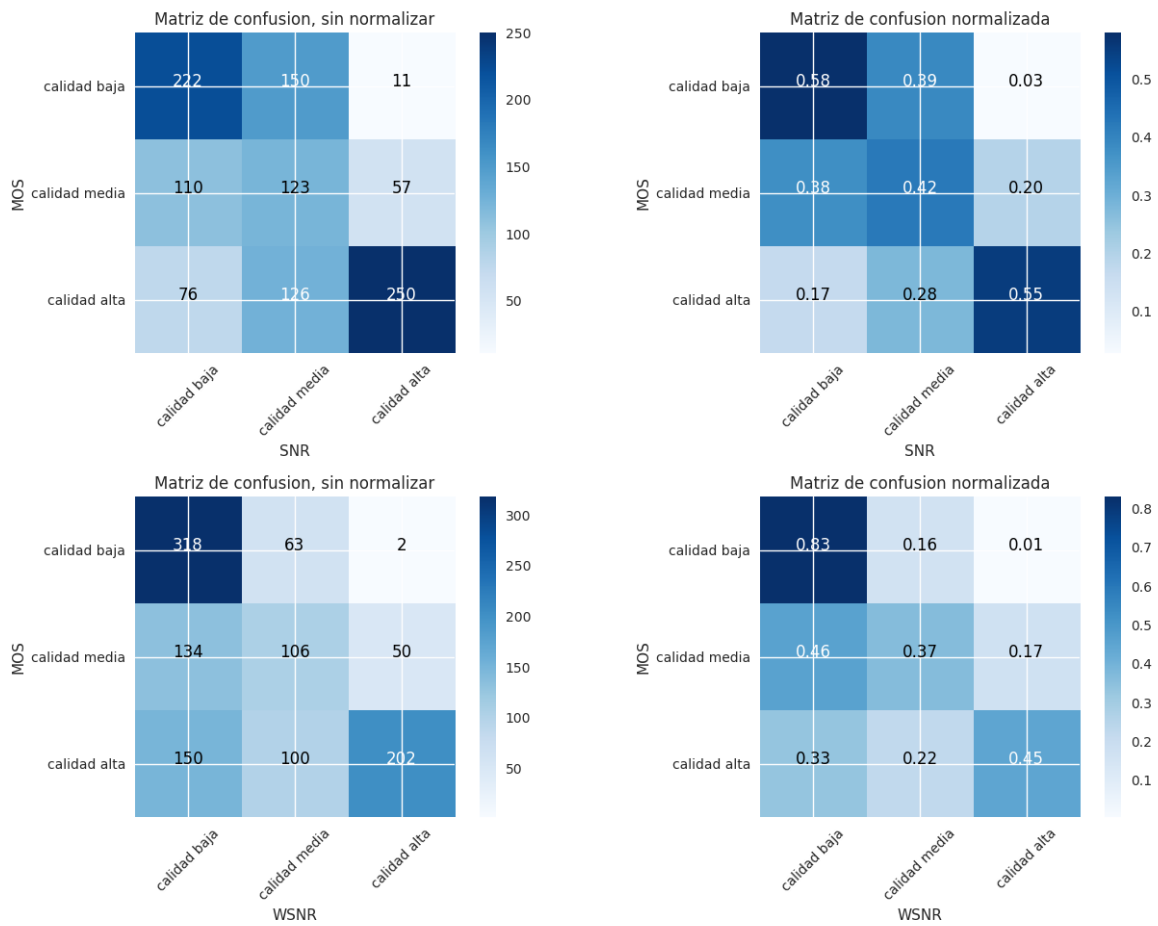


Figura E.20: Grupo 4: Distorsiones Exóticas. Métricas SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	59 %	43 %	50 %
SNR	58 %	42 %	55 %
WSNR	83 %	37 %	45 %
UQI	51 %	32 %	69 %
VIF	51 %	24 %	73 %
NQM	87 %	36 %	52 %
SSIM	47 %	45 %	70 %
MSSIM	65 %	49 %	69 %
CQ(1,1)	59 %	34 %	60 %
GMSM	68 %	59 %	69 %
GMSD	89 %	56 %	60 %

Cuadro E.7: Grupo 4: Distorsiones Exóticas. Porcentaje de imágenes bien clasificadas por métrica, según calidad. Es curioso que VIF clasificó correctamente el mayor número de imágenes de calidad alta y el menor número de imágenes de calidad media.

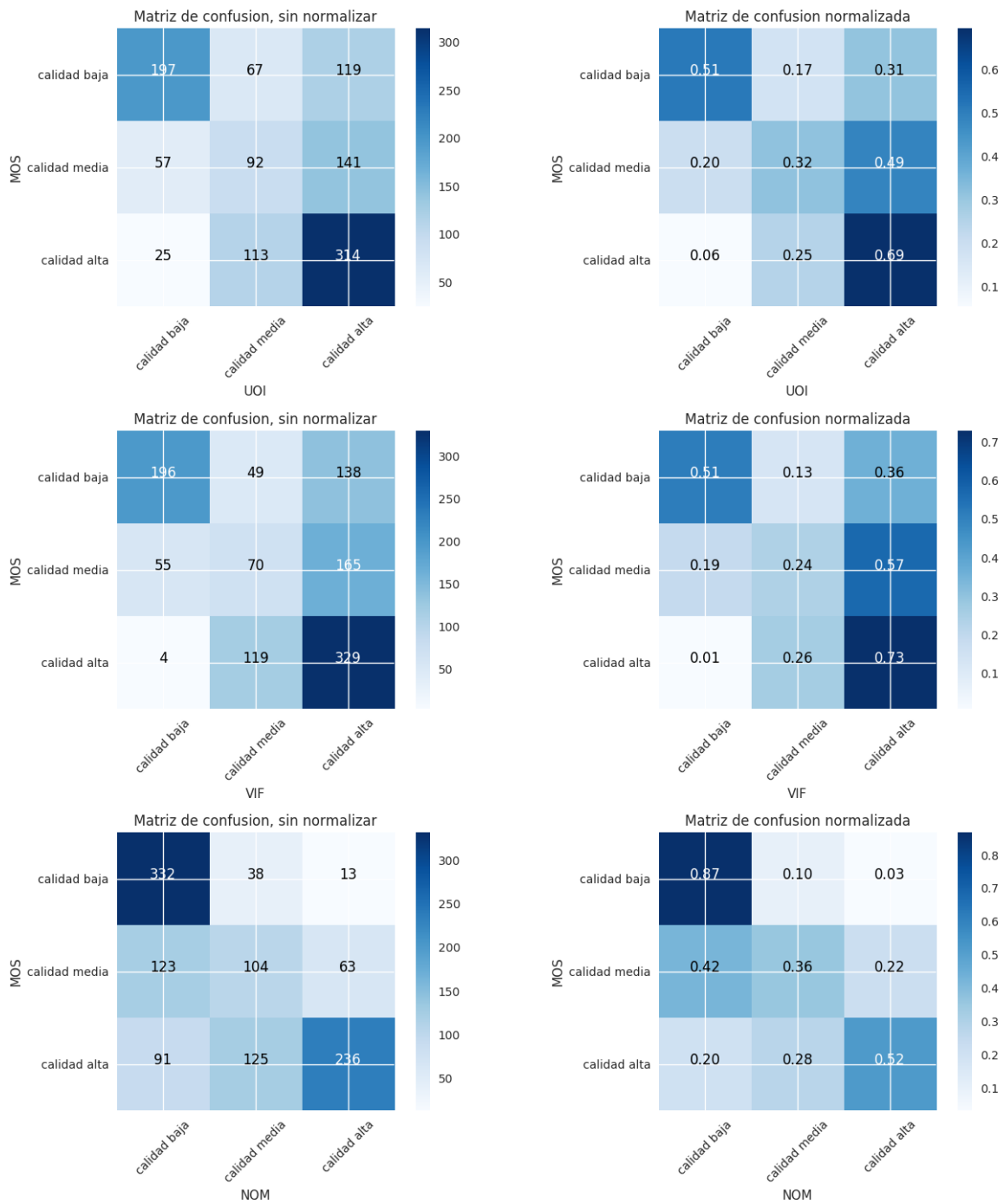


Figura E.21: Grupo 4: Distorsiones Exóticas. Métricas UOI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	GMSD	GMSM	VIF
Segundas	NQM	GMSD	SSIM
Terceras	WSNR	MSSIM	UOI & MSSIM & GMSM
Cuartas	MSSIM	SSIM	CQ(1,1) & GMSD
Quintas	MSE & RMSE & PSNR & CQ(1,1)	MSE & RMSE & PSNR	SNR

Cuadro E.8: Grupo 4: Distorsiones Exóticas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

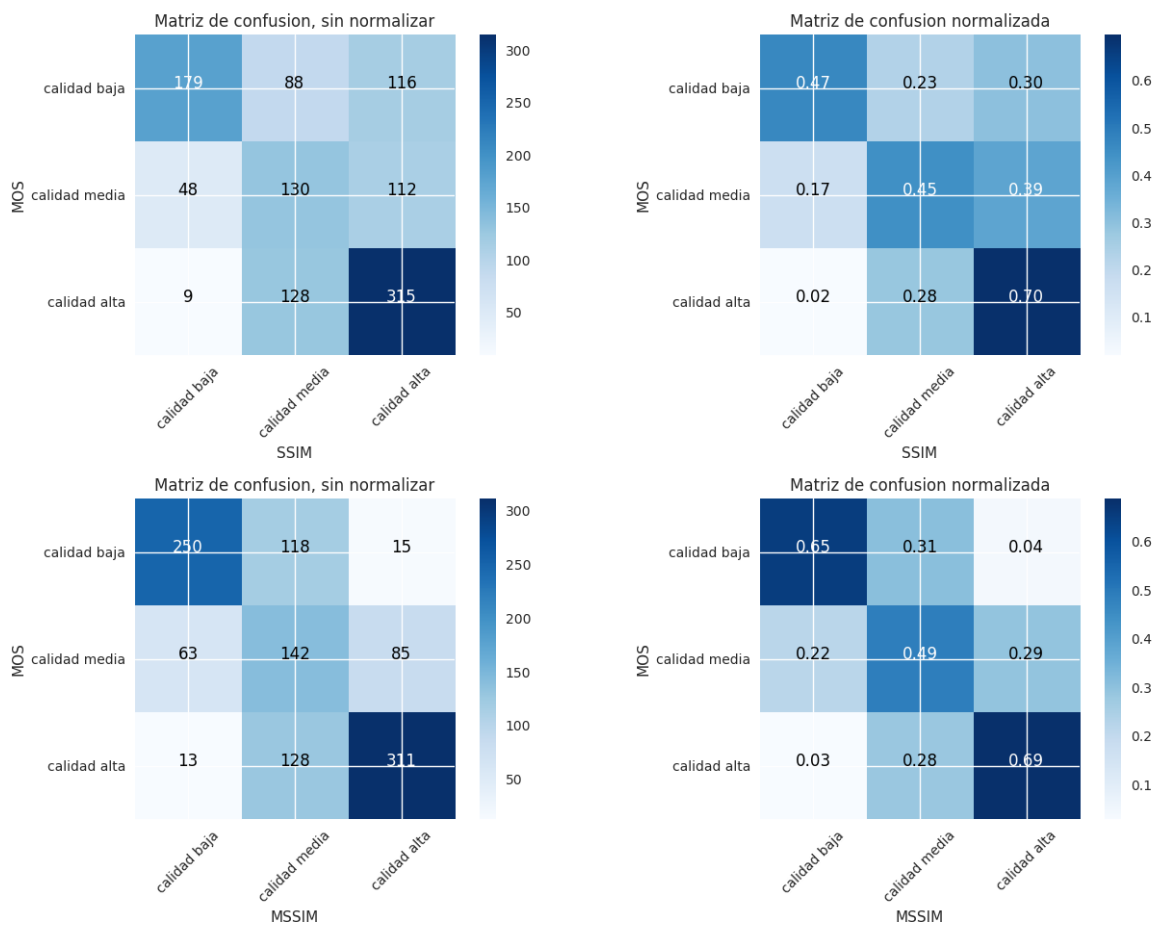


Figura E.22: Grupo 4: Distorsiones Exóticas. Métricas SSIM y MSSIM.

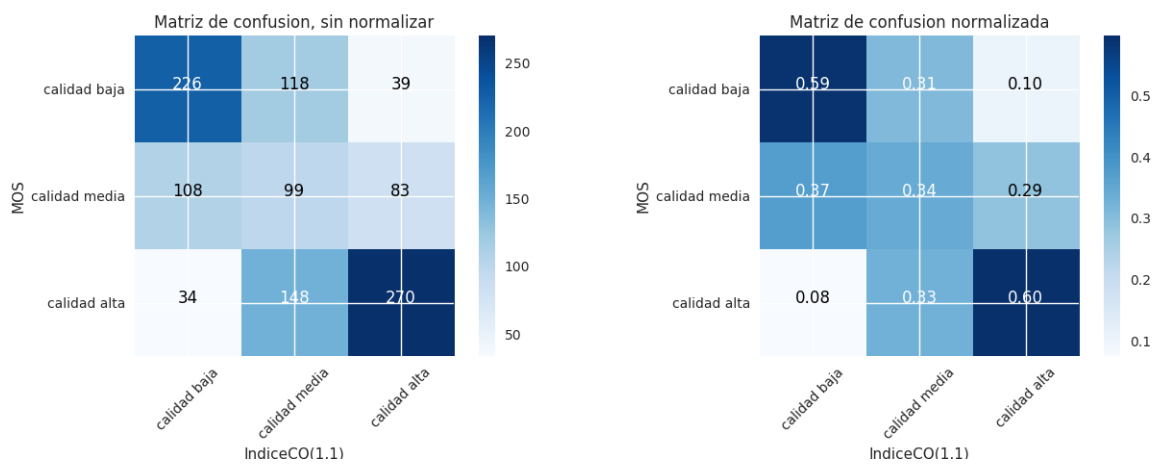


Figura E.23: Grupo 4: Distorsiones Exóticas. Índice CQ(1,1).

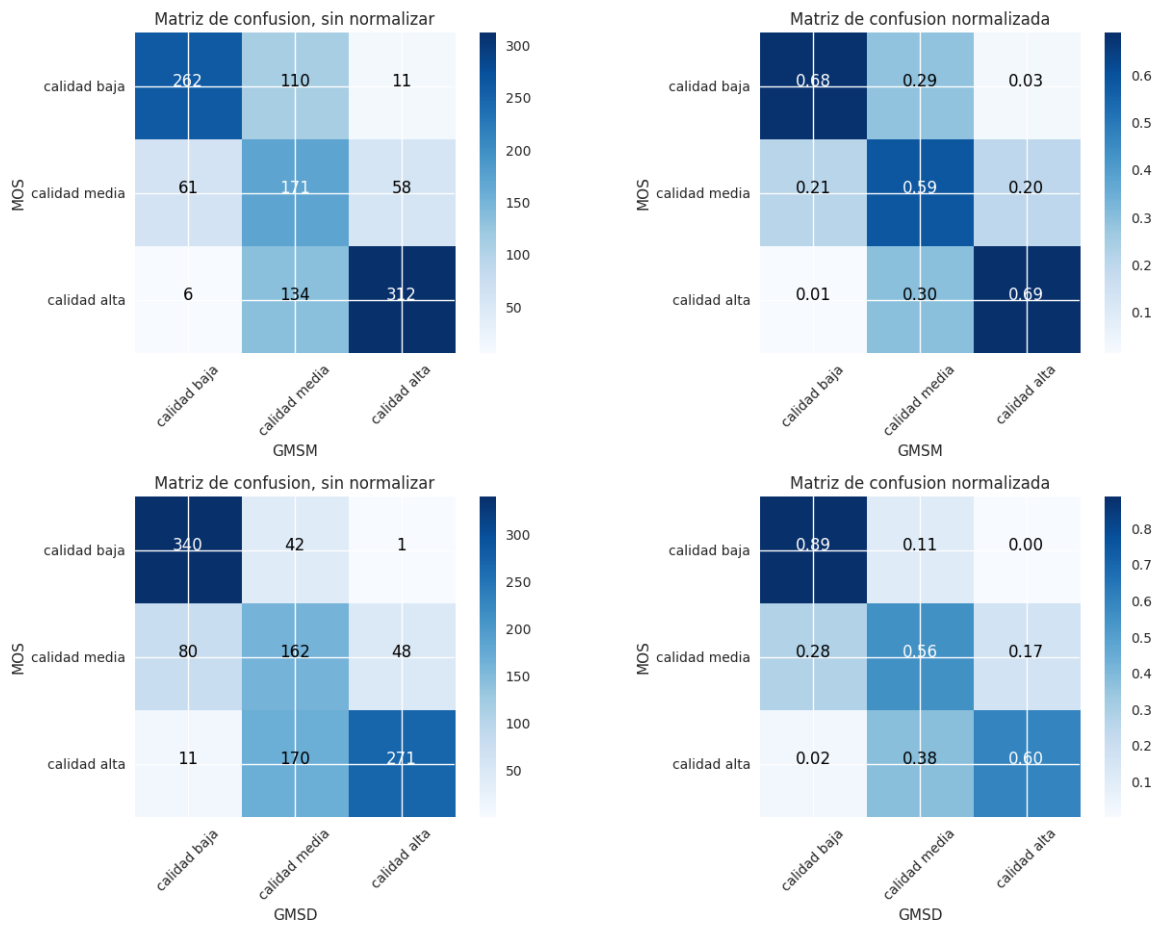


Figura E.24: Grupo 4: Distorsiones Exóticas. Métricas GSM y GMSD.

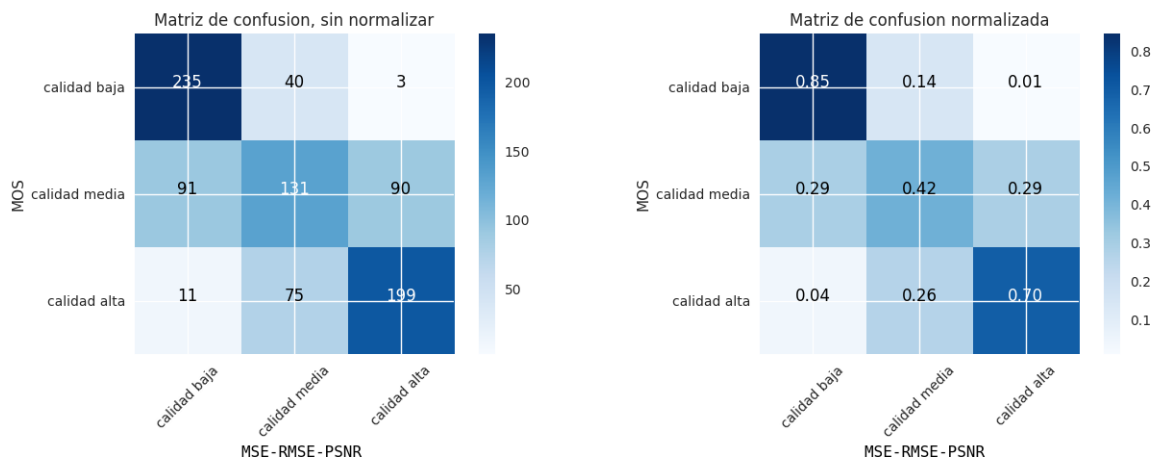


Figura E.25: Grupo 5: Distorsiones Nuevas. Métricas MSE, RMSE y PSNR.

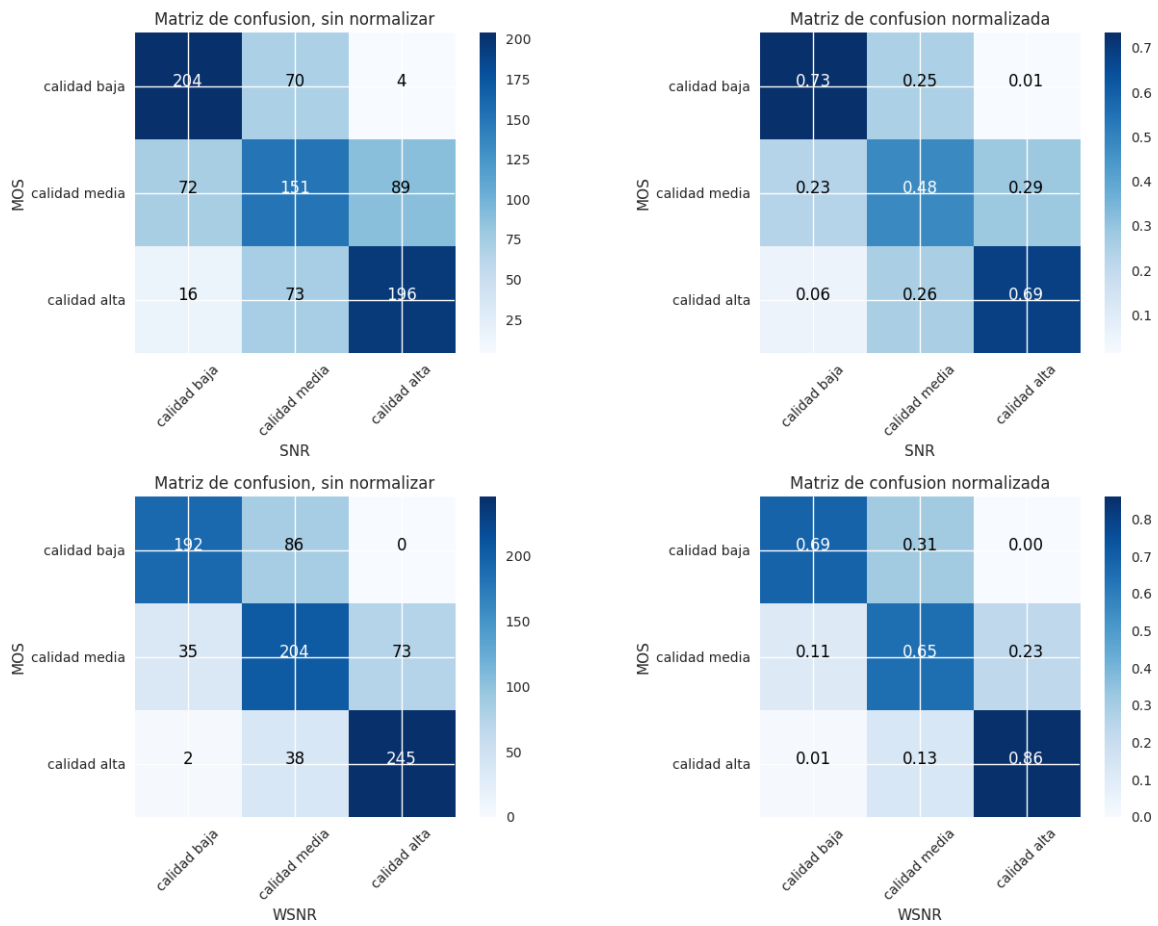


Figura E.26: Grupo 5: Distorsiones Nuevas. Métricas SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	85 %	42 %	70 %
SNR	73 %	48 %	69 %
WSNR	69 %	65 %	86 %
UQI	74 %	37 %	36 %
VIF	86 %	44 %	36 %
NQM	70 %	69 %	83 %
SSIM	79 %	47 %	55 %
MSSIM	86 %	52 %	65 %
CQ(1,1)	77 %	33 %	51 %
GMSM	82 %	49 %	64 %
GMSD	78 %	61 %	77 %

Cuadro E.9: Grupo 5: Distorsiones Nuevas. Porcentaje de imágenes bien clasificadas por métrica, según calidad.



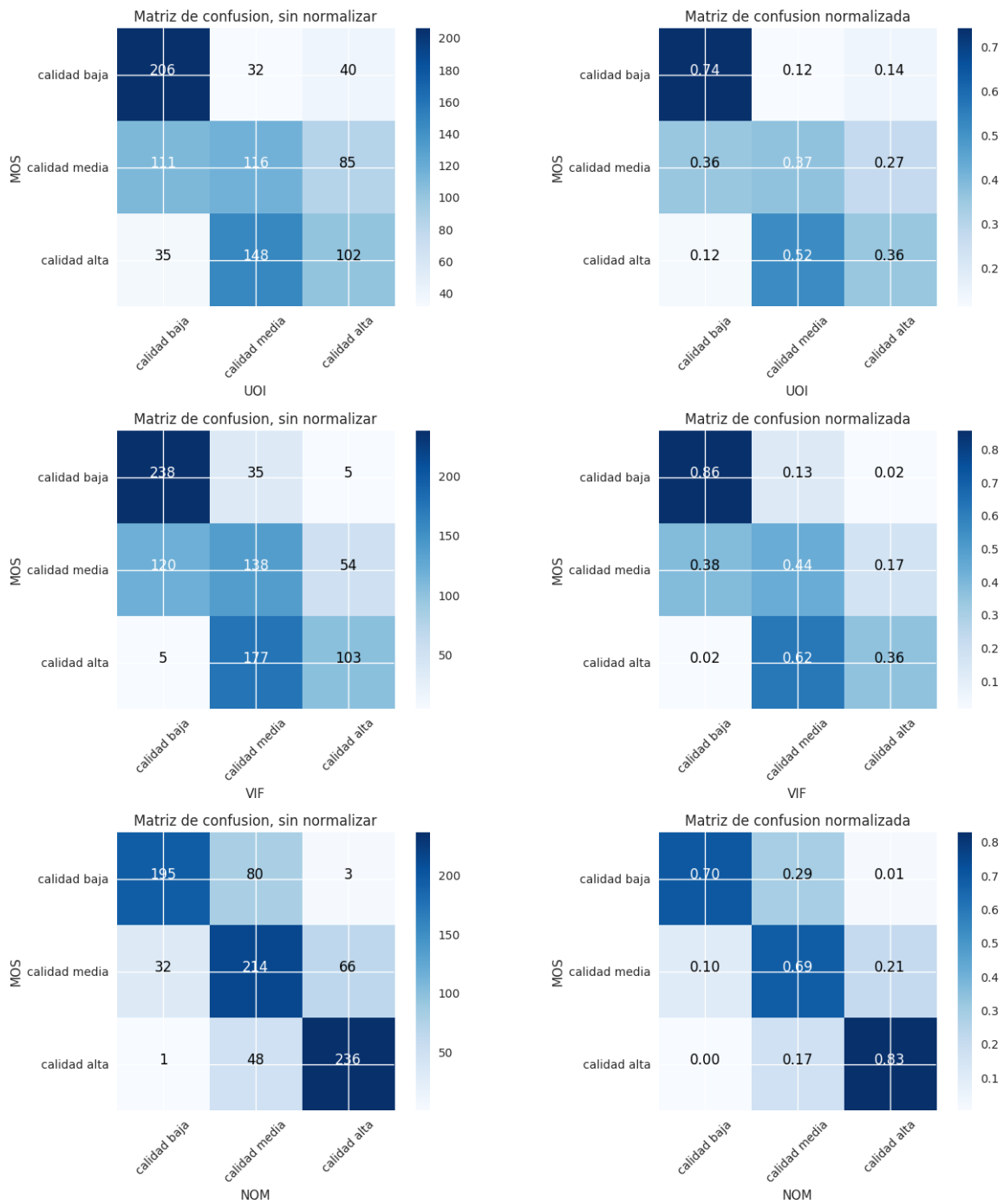


Figura E.27: Grupo 5: Distorsiones Nuevas. Métricas UQI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	VIF & MSSIM	NQM	WSNR
Segundas	MSE & RMSE & PSNR	WSNR	NQM
Terceras	GMSM	MSSIM	GMSD
Cuartas	SSIM	GMSM	MSE & RMSE & PSNR
Quintas	GMSD	SNR	SNR

Cuadro E.10: Grupo 5: Distorsiones Nuevas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

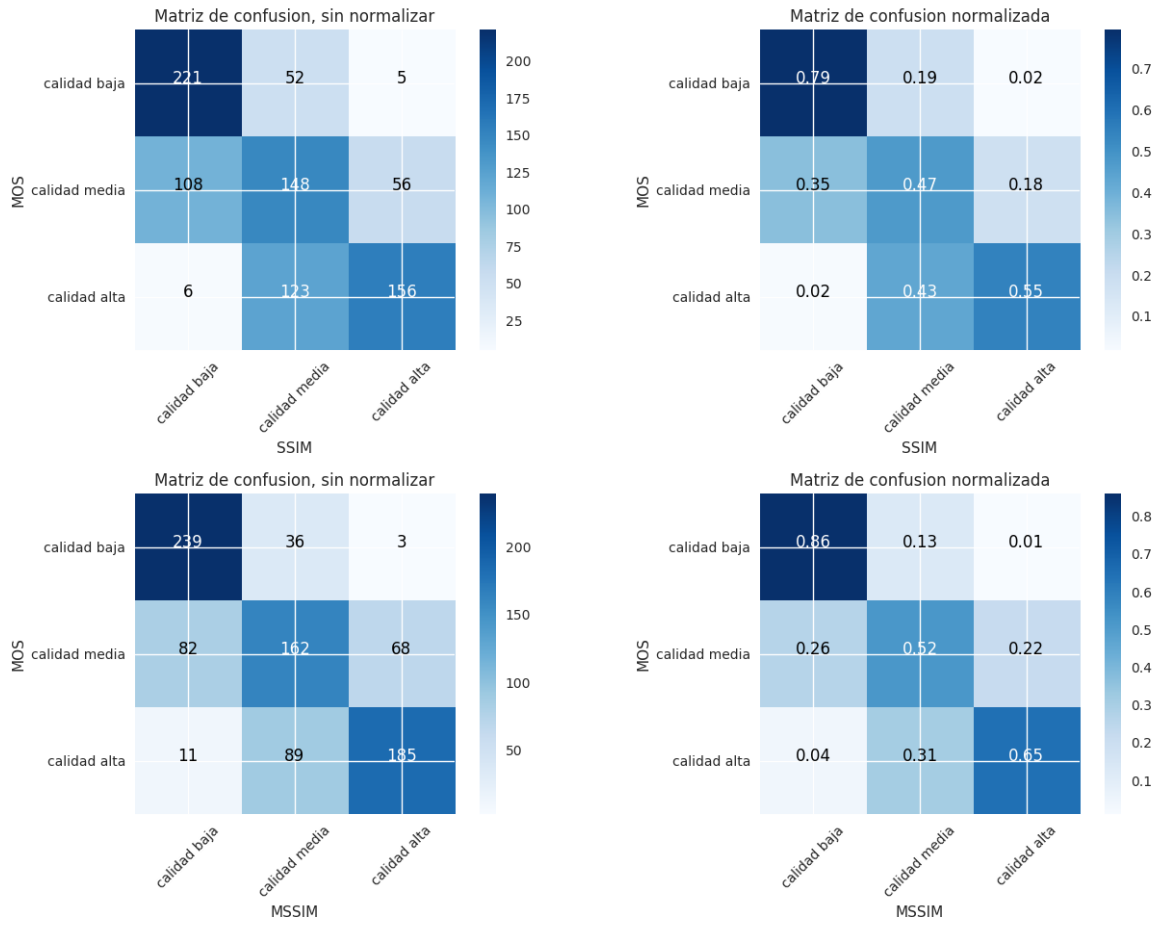


Figura E.28: Grupo 5: Distorsiones Nuevas. Métricas SSIM y MSSIM.

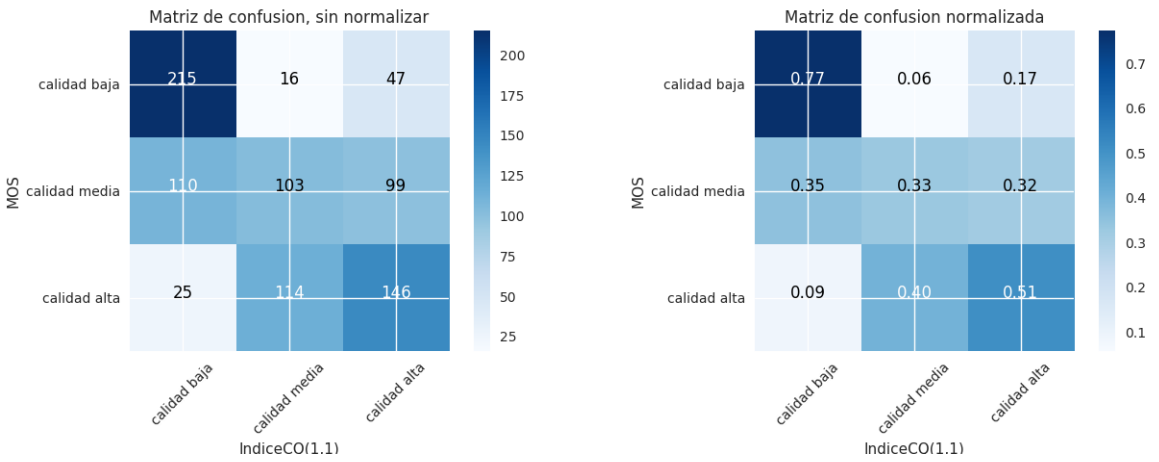


Figura E.29: Grupo 5: Distorsiones Nuevas. Índice CQ(1,1).

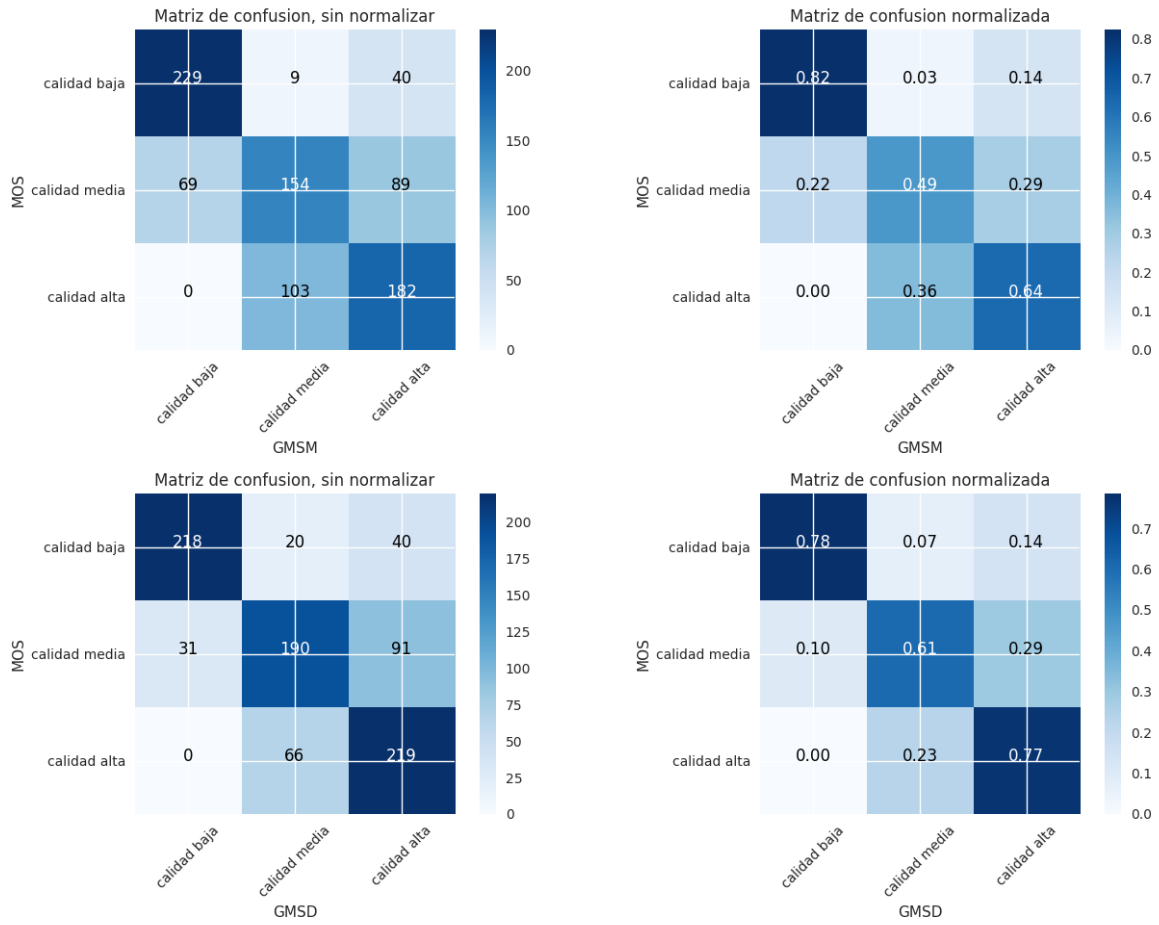


Figura E.30: Grupo 5: Distorsiones Nuevas. Métricas GSM y GMSD.

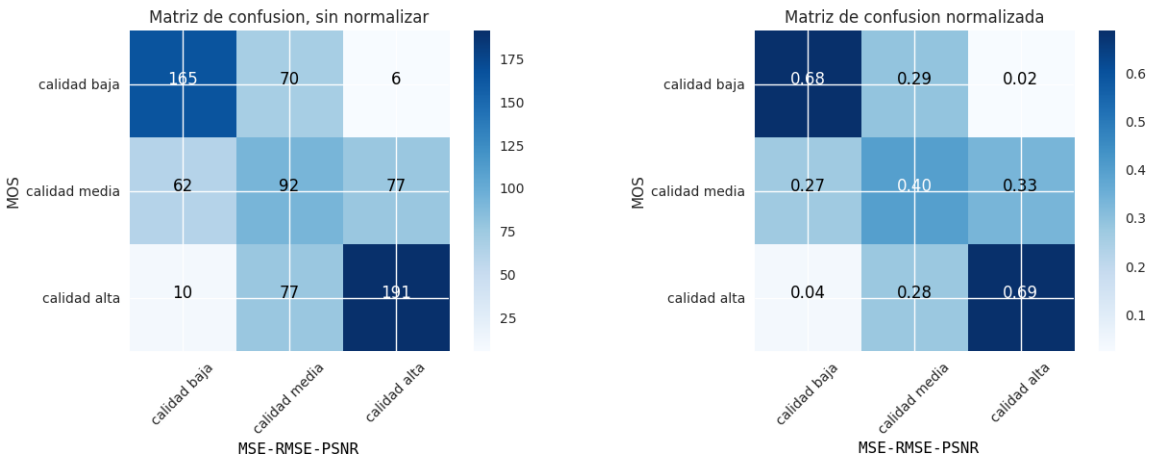


Figura E.31: Grupo 6: Distorsiones del Color. Métricas MSE, RMSE y PSNR.

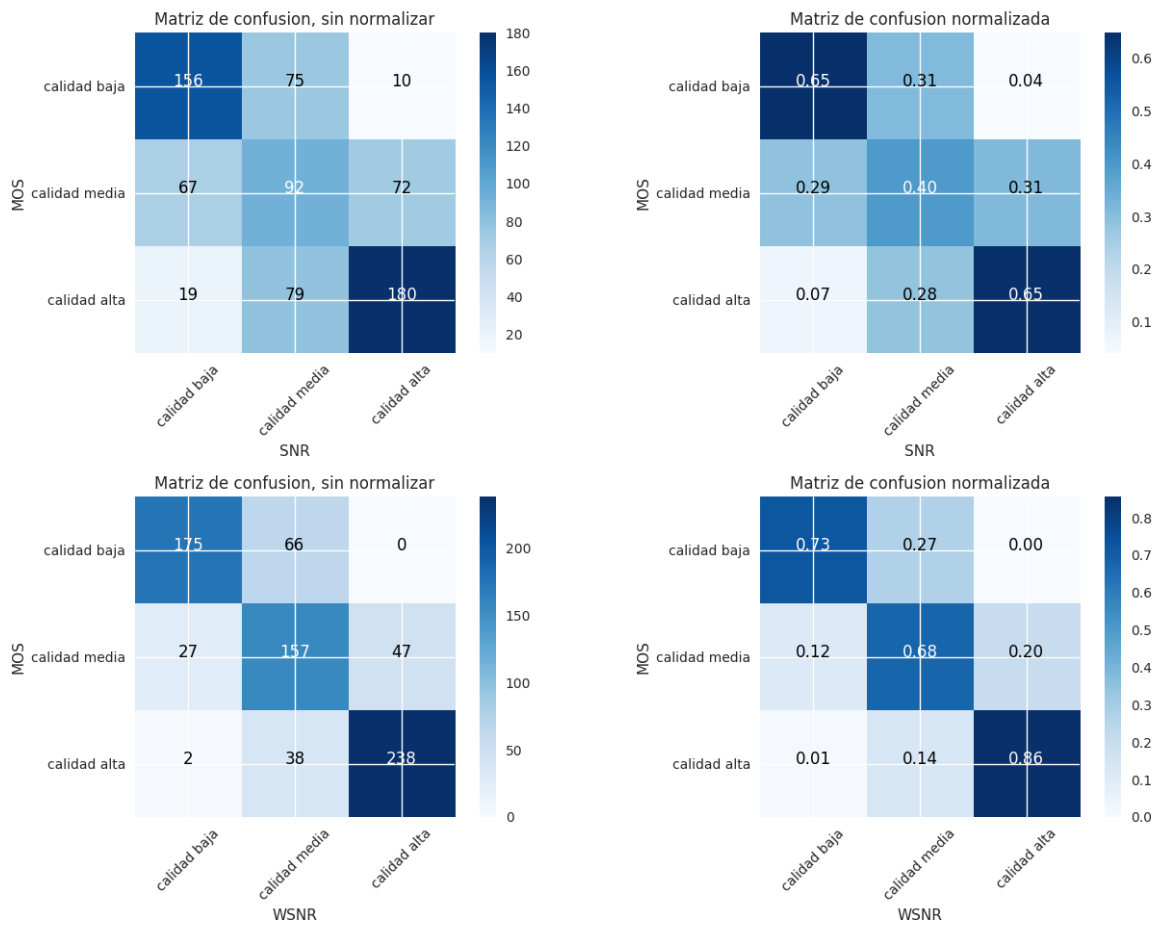


Figura E.32: Grupo 6: Distorsiones del Color. Métricas SNR y WSNR.

Métrica	calidad baja	calidad media	calidad alta
MSE & RMSE & PSNR	68 %	40 %	69 %
SNR	65 %	40 %	65 %
WSNR	73 %	68 %	86 %
UQI	72 %	39 %	37 %
VIF	79 %	39 %	37 %
NQM	76 %	67 %	84 %
SSIM	75 %	45 %	53 %
MSSIM	86 %	45 %	54 %
CQ(1,1)	64 %	35 %	69 %
GMSM	81 %	53 %	79 %
GMSD	78 %	55 %	84 %

Cuadro E.11: Grupo 6: Distorsiones del Color. Porcentaje de imágenes bien clasificadas por métrica, según calidad.

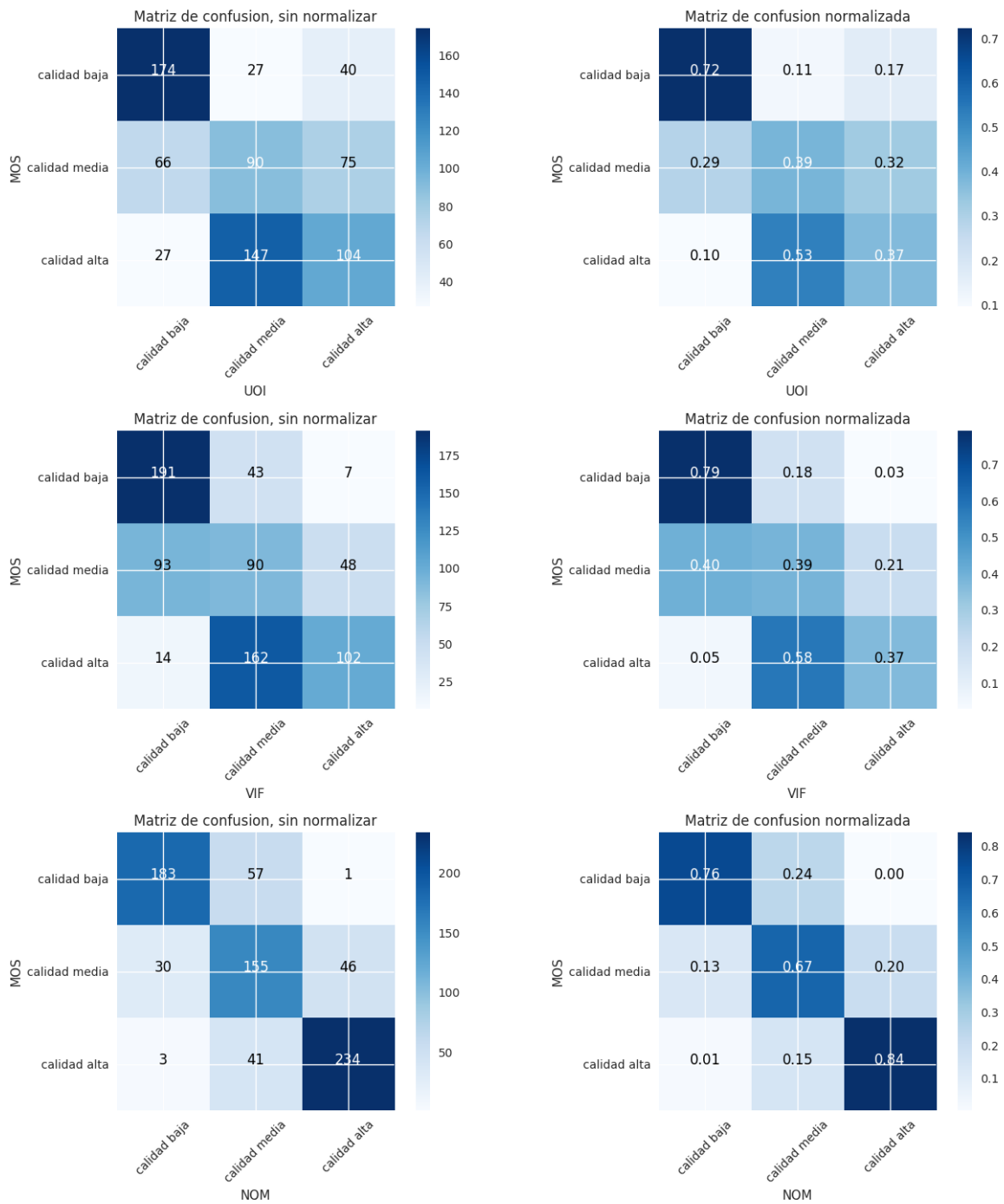


Figura E.33: Grupo 6: Distorsiones del Color. Métricas UOI, VIF y NQM.

Métrica	calidad baja	calidad media	calidad alta
Primeras	MSSIM	WSNR	WSNR
Segundas	GSM	NQM	NQM & GMSD
Terceras	VIF	GMSD	GMSM
Cuartas	GMSD	GMSM	MSE & RMSE & PSNR & CQ(1,1)
Quintas	NQM	SSIM & MSSIM	SNR

Cuadro E.12: Grupo 6: Distorsiones del Color. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad.

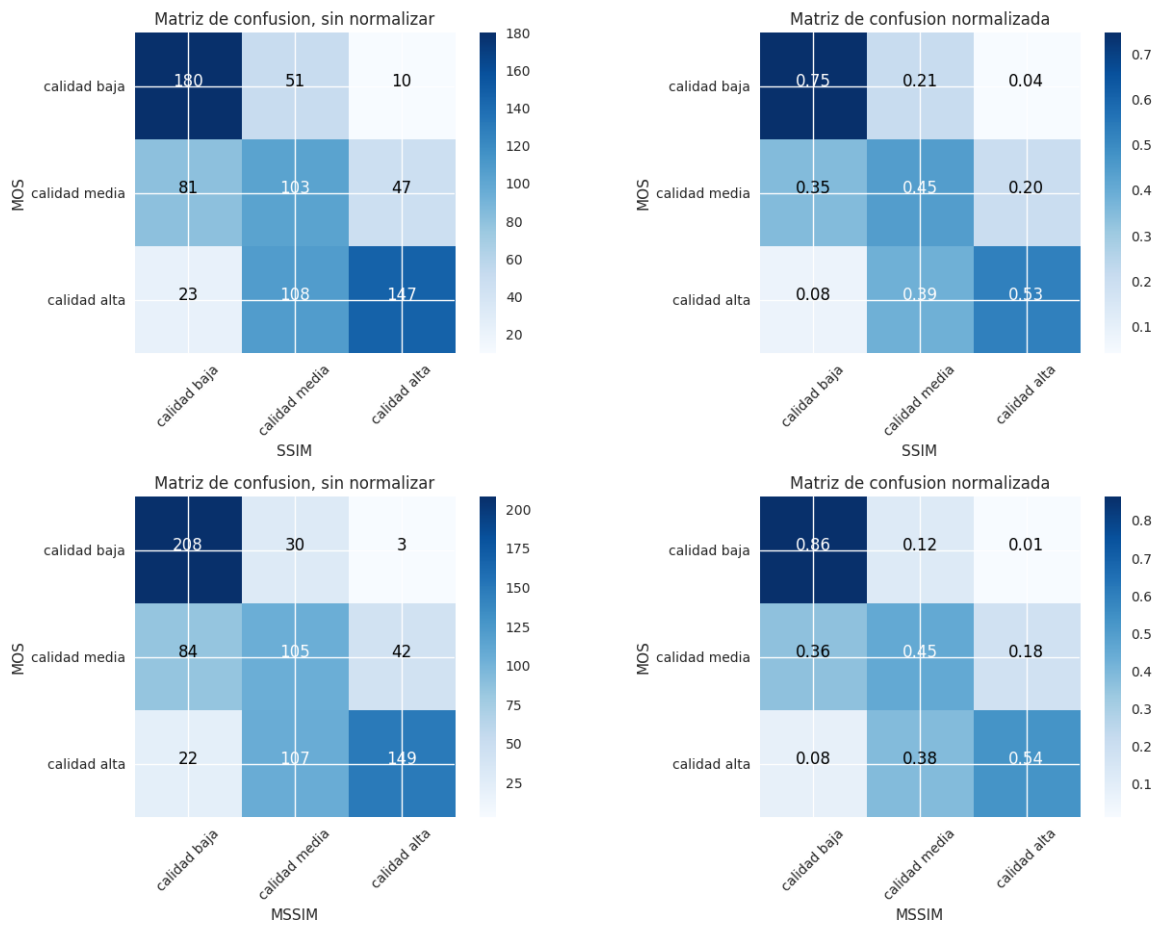


Figura E.34: Grupo 6: Distorsiones del Color. Métricas SSIM y MSSIM.

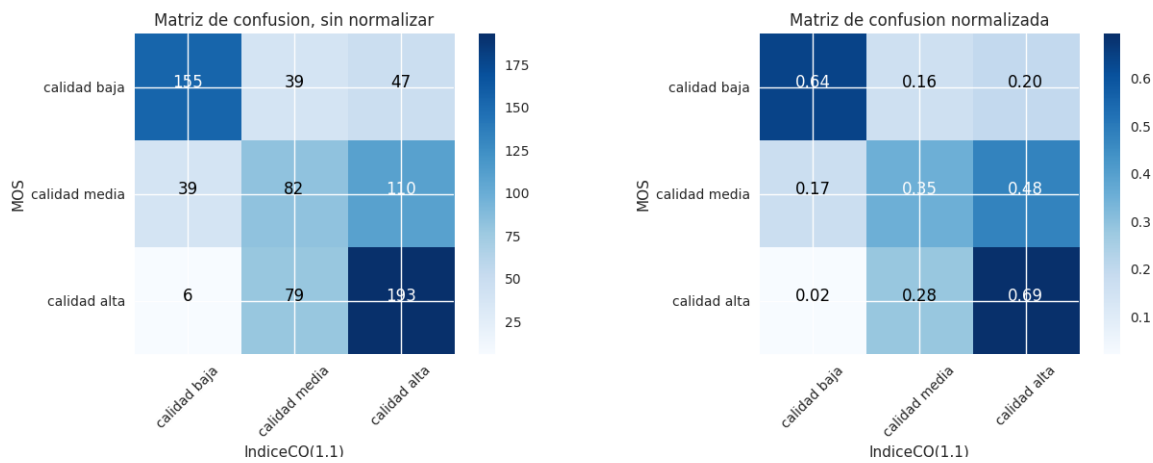


Figura E.35: Grupo 6: Distorsiones del Color. Índice CQ(1,1).

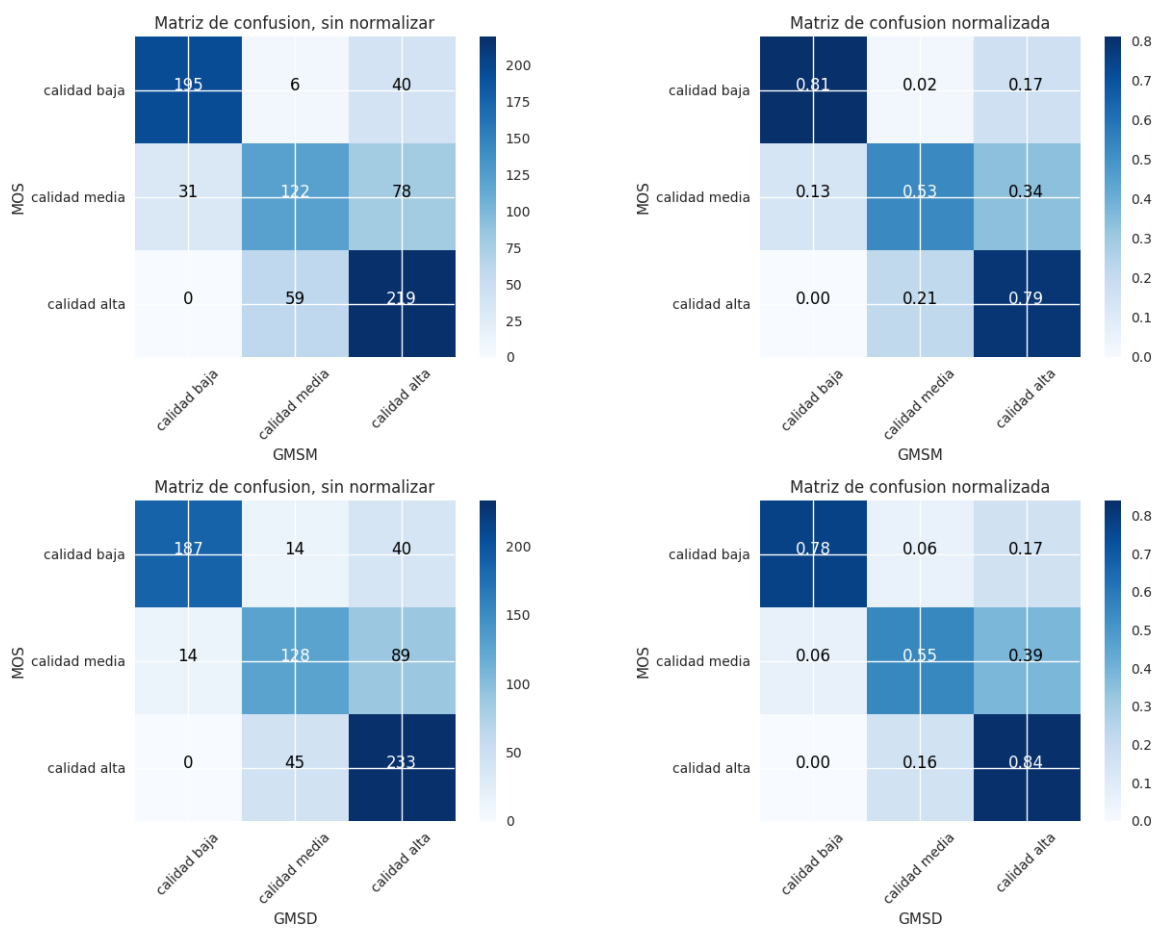


Figura E.36: Grupo 6: Distorsiones del Color. Métricas GSM y GMSD.

---

## Apéndice F

# Códigos en Python para obtener los coeficientes de concordancia

En este Apéndice desarrollo los códigos que fueron utilizados para obtener los coeficientes de concordancia presentados en el Capítulo 6.

Fueron necesarios los siguientes paquetes:

```
import numpy as np
import collections
from nltk import agreement
from nltk.metrics.agreement import AnnotationTask
```

### F.1. Función que calcula el kappa de Cohen y el pi de Scott entre dos variables

```
def kappa_pi(mos, metric):

    rater0 = mos
    rater1 = metric

    n = mos.shape[0]

    taskdata_m = [ ]

    for i in xrange(n):
```



```
        taskdata_m.append([0, str(i) , str(rater0[i])])

    for i in xrange(n):
        taskdata_m.append([1, str(i) , str(rater1[i])])

    ratingtask_m = agreement.AnnotationTask(data=taskdata_m)

    k = ratingtask_m.kappa()

    p = ratingtask_m.pi()

    return (k, p)
```

El kappa de Cohen también puede calcularse usando la función *cohen\_kappa\_score* del paquete `sklearn.metrics` [1], se obtuvieron los mismos resultados.

## F.2. Función que calcula índices de concordancia entre varias variables

```
def multi(*items):

    n = items[0].shape[0]

    taskdata = [ ]

    j = 0
    for item in items:
        rater = item.tolist()
        for i in xrange(n):
            taskdata.append([j, str(i) , str(rater[i])])
        j = j + 1

    ratingtask = agreement.AnnotationTask(data=taskdata)

    kappa = ratingtask.kappa()
    fleiss = ratingtask.multi_kappa()
    alpha = ratingtask.alpha()
```

```
scotts = ratingtask.pi()

return (kappa, fleiss, alpha, scotts)
```

La función *kappa()* devuelve el promedio de todos los kappas entre todos los pares posibles de variables, *multi\_kappa()* realiza el promedio sobre los acuerdos observados y esperados para cada par de variables y luego calcula el índice, *alpha()* calcula el índice alpha de Krippendorff, *pi()* en este caso calcula un multi-pi, que sería precisamente el kappa de Fleiss. Los códigos de estas funciones se encuentran en [2]. Los cuatro índices dieron el mismo resultado redondeando al cuarto dígito decimal cuando se aplicó esta función a todos los datos. Cuando se separaron los datos por grupos de distorsiones, los cuatro índices dieron también valores muy cercanos en todos los casos.

### F.3. Función que calcula el kappa de Fleiss

```
def kappa_fleiss(*items):

    n = items[0].shape[0]

    m = 0
    for item in items:
        m = m + 1

    counter = collections.Counter(items[0])
    c = len(counter)

    N = np.empty((n, c))

    data = np.empty((n, m))

    j = 0
    for item in items:
        data[:,j] = item
        j = j+1

    for i in xrange(n):
        contador = collections.Counter(data[i,:])
```

```
        for j in xrange(c):
            N[i,j] = contador[j+1]

p = []
for x in xrange(c):
    pi = sum(N[:,x])/(n*m)
    p.append(pi)

Pe = 0
for elem in p:
    Pe = Pe + elem**2

S = 0
for i in xrange(n):
    for j in xrange(c):
        S = S + N[i,j]**2

Pbar = (S - n*m)/(n*m*(m-1))

KF = (Pbar - Pe)/(1 - Pe)

return KF
```

Se obtienen exactamente los mismos resultados que con la función multi-pi  $pi()$  del apartado anterior.

---

# Bibliografía

[1] [http://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen\\_kappa\\_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html)

[2] <http://www.nltk.org/api/nltk.metrics.html>

---

## Apéndice G

# Medidas de Resumen para cada grupo de distorsiones

En este apéndice se presentan las medidas de resumen para cada grupo de distorsiones.

Métrica	Media	Desviación Estandar	Varianza
MOS	4,4929	1,0803	1,1671
MSE	233,3646	314,3566	98820,0924
RMSE	13,1521	7,7709	60,3867
PSNR	27,1219	4,8803	23,8174
SNR	20,4643	5,1467	26,4880
WSNR	34,8221	7,9713	63,5421
UQI	0,5897	0,2149	0,0462
VIF	0,5288	0,1555	0,0242
NQM	26,1795	7,0800	50,1270
SSIM	0,8180	0,1518	0,0230
MSSIM	0,9073	0,0871	0,0076
CQ(1,1)	0,8179	0,1869	0,0349
GMSM	0,9566	0,0497	0,0025
GMSD	0,0657	0,0575	0,0033

Cuadro G.1: Medidas de Resumen para el grupo de distorsiones Ruidosas

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,4878	3,7307	4,5897	5,3974	6,6970
MSE	4,1254	59,9195	128,1126	260,8673	3029,3578
RMSE	2,0311	7,7408	11,3187	16,1514	55,0396
PSNR	13,317	23,9666	27,0549	30,3551	41,9761
SNR	5,9037	16,3557	20,4384	24,5835	34,2686
WSNR	12,2130	29,4119	35,3301	40,5288	56,8276
UQI	0,0702	0,4274	0,6058	0,7602	0,9757
VIF	0,1688	0,40631	0,5339	0,6501	0,9009
NQM	5,7457	21,2862	26,8113	31,4439	43,1455
SSIM	0,1630	0,7306	0,8656	0,9380	0,9982
MSSIM	0,5122	0,8592	0,9386	0,9743	0,9980
CQ(1,1)	0,0317	0,7383	0,8874	0,9563	0,9994
GMSM	0,6962	0,9394	0,9762	0,9918	0,9995
GMSD	0,0014	0,0197	0,0471	0,1030	0,2987

Cuadro G.2: Medidas de Resumen para el grupo de distorsiones Ruidosas

Métrica	Media	Desviación Estandar	Varianza
MOS	4,4422	1,0854	1,1781
MSE	243,6986	337,9310	114197,3387
RMSE	13,3294	8,1257	66,0263
PSNR	27,0819	5,0112	25,1118
SNR	20,4243	5,2659	27,7297
WSNR	34,8217	7,9280	62,8533
UQI	0,5948	0,2208	0,0488
VIF	0,5328	0,1593	0,0254
NQM	26,2584	6,9461	48,2484
SSIM	0,8207	0,1549	0,0240
MSSIM	0,9085	0,0904	0,0082
CQ(1,1)	0,7959	0,1965	0,0386
GMSM	0,9541	0,0518	0,0027
GMSD	0,0690	0,0586	0,0034

Cuadro G.3: Medidas de Resumen para el grupo de distorsiones Actuales

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,4878	3,6750	4,5294	5,3333	6,6970
MSE	4,1254	56,6011	129,0402	265,9215	3029,3578
RMSE	2,0311	7,5234	11,3596	16,3071	55,0396
PSNR	13,317	23,8833	27,0236	30,6026	41,9761
SNR	5,9037	16,2691	20,3714	24,5877	34,2686
WSNR	12,2130	29,4412	35,1170	40,4621	56,8276
UQI	0,0702	0,4290	0,6124	0,7772	0,9757
VIF	0,1688	0,4090	0,5376	0,6548	0,9009
NQM	5,7457	21,5000	26,8443	31,4010	43,1455
SSIM	0,1630	0,7373	0,8693	0,9416	0,9982
MSSIM	0,5122	0,8642	0,9433	0,9768	0,9980
CQ(1,1)	0,0317	0,7011	0,8706	0,9455	0,9993
GMSM	0,6962	0,9352	0,9740	0,9915	0,9995
GMSD	0,0014	0,0214	0,0515	0,1056	0,2987

Cuadro G.4: Medidas de Resumen para el grupo de distorsiones Actuales

Métrica	Media	Desviación Estandar	Varianza
MOS	4,4224	1,2285	1,5093
MSE	198,3290	279,7285	78248,0235
RMSE	12,0609	7,2707	52,8633
PSNR	27,9992	5,1885	26,9209
SNR	21,3417	5,3884	29,0351
WSNR	34,4137	7,7525	60,1018
UQI	0,6020	0,2227	0,0500
VIF	0,5343	0,1691	0,0286
NQM	25,8376	7,1738	51,4638
SSIM	0,8450	0,1382	0,0191
MSSIM	0,9137	0,0811	0,0066
CQ(1,1)	0,7615	0,2201	0,0485
GMSM	0,9510	0,0567	0,0032
GMSD	0,0722	0,0669	0,0045

Cuadro G.5: Medidas de Resumen para el grupo de distorsiones Simples

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	1,1628	3,5733	4,6585	5,4942	6,3421
MSE	4,1254	38,3606	124,3457	257,0348	3029,3578
RMSE	2,0311	6,1936	11,1510	16,0323	55,0396
PSNR	13,317	24,0309	27,1845	32,2921	41,9761
SNR	7,7064	17,1792	21,1520	25,2371	34,2686
WSNR	16,5534	28,7781	35,0899	41,0945	50,7694
UQI	0,1516	0,4248	0,6179	0,7996	0,9757
VIF	0,1270	0,4042	0,5381	0,6690	0,9009
NQM	9,5187	20,3190	26,8792	31,7887	41,5838
SSIM	0,3755	0,7555	0,8923	0,9591	0,9982
MSSIM	0,6693	0,8689	0,9470	0,9790	0,9976
CQ(1,1)	0,0317	0,6222	0,8323	0,9375	0,9978
GMSM	0,7483	0,9243	0,9772	0,9931	0,9987
GMSD	0,0027	0,0168	0,0450	0,1126	0,2987

Cuadro G.6: Medidas de Resumen para el grupo de distorsiones Simples

Métrica	Media	Desviación Estandar	Varianza
MOS	4,5644	1,3533	1,8315
MSE	252,9526	278,2028	77396,7834
RMSE	14,0928	7,3720	54,3456
PSNR	26,3094	4,5419	20,6289
SNR	19,6519	4,7816	22,8633
WSNR	27,8786	10,0806	101,6191
UQI	0,7089	0,2978	0,0887
VIF	0,7015	0,2336	0,0546
NQM	21,5085	9,6311	92,7589
SSIM	0,8958	0,1060	0,0112
MSSIM	0,9161	0,0972	0,0094
CQ(1,1)	0,7801	0,2729	0,0745
GMSM	0,9579	0,0583	0,0034
GMSD	0,0779	0,0675	0,0046

Cuadro G.7: Medidas de Resumen para el grupo de distorsiones Exóticas



Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,2424	3,5405	4,7813	5,7222	7,2143
MSE	13,9392	65,0345	169,5715	313,8521	1928,1151
RMSE	3,7335	8,0644	13,0220	17,7159	43,9103
PSNR	15,2795	23,1636	25,8373	29,9994	36,6884
SNR	7,0060	15,9853	19,2842	23,3153	30,7369
WSNR	5,5422	20,2265	25,9979	35,5403	55,1037
UQI	-0,5999	0,5473	0,8253	0,9460	0,9984
VIF	0,1433	0,5116	0,7245	0,9202	1,1036
NQM	1,6597	12,8340	20,7271	29,6503	46,0276
SSIM	0,3486	0,8528	0,9380	0,9704	0,9987
MSSIM	0,4415	0,8848	0,9550	0,9812	0,9998
CQ(1,1)	-0,7530	0,7163	0,8912	0,9604	1,0001
GMSM	0,5368	0,9532	0,9807	0,9939	1,0000
GMSD	0,0001	0,0207	0,0609	0,1225	0,2959

Cuadro G.8: Medidas de Resumen para el grupo de distorsiones Exóticas

Métrica	Media	Desviación Estandar	Varianza
MOS	4,4243	1,2675	1,6066
MSE	205,4650	176,1380	31024,5872
RMSE	12,9846	6,0716	36,8644
PSNR	26,8655	4,2503	18,0654
SNR	20,2080	4,5615	20,8074
WSNR	33,2596	8,3895	70,3840
UQI	0,5833	0,2956	0,0874
VIF	0,5313	0,1612	0,0260
NQM	25,1755	7,9287	62,8635
SSIM	0,8325	0,1403	0,0197
MSSIM	0,8994	0,0994	0,0099
CQ(1,1)	0,7502	0,2942	0,0866
GMSM	0,9608	0,0492	0,0024
GMSD	0,0606	0,0623	0,0039

Cuadro G.9: Medidas de Resumen para el grupo de distorsiones Nuevas

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,2424	3,6493	4,6216	5,4500	6,7778
MSE	30,7233	64,6456	129,8218	261,2981	645,2529
RMSE	5,5429	8,0403	11,3939	16,1647	25,4018
PSNR	20,0335	23,9594	26,9973	30,0254	33,2561
SNR	11,1515	16,2901	20,0270	24,2576	30,7369
WSNR	12,1748	27,1097	33,4398	39,4265	55,1037
UQI	-0,5999	0,3881	0,6229	0,8261	0,9925
VIF	0,1433	0,4055	0,5374	0,6544	0,9368
NQM	1,6597	20,0330	25,7866	30,8050	46,0276
SSIM	0,0916	0,7562	0,8757	0,9427	0,9969
MSSIM	0,3004	0,8536	0,9395	0,9743	0,9969
CQ(1,1)	-0,7530	0,6254	0,8693	0,9668	1,0000
GMSM	0,7282	0,9433	0,9822	0,9953	0,9999
GMSD	0,0002	0,0118	0,0376	0,0972	0,2702

Cuadro G.10: Medidas de Resumen para el grupo de distorsiones Nuevas

Métrica	Media	Desviación Estandar	Varianza
MOS	4,4588	1,3832	1,9131
MSE	196,5269	176,8631	31280,5515
RMSE	12,6174	6,1098	37,3292
PSNR	27,1920	4,4287	19,6126
SNR	20,5345	4,7341	22,4119
WSNR	33,5491	8,7985	77,4127
UQI	0,6278	0,2528	0,0639
VIF	0,5331	0,1674	0,0639
NQM	25,3745	8,3467	69,6677
SSIM	0,8351	0,1436	0,0206
MSSIM	0,8936	0,0980	0,0096
CQ(1,1)	0,8312	0,2282	0,0521
GMSM	0,9606	0,0558	0,0031
GMSD	0,0576	0,0691	0,0048

Cuadro G.11: Medidas de Resumen para el grupo de distorsiones del Color

Métrica	Mínimo	Primer cuartil	Mediana (Segundo Cuartil)	Tercer Cuartil	Máximo
MOS	0,2424	3,5989	4,7297	5,6108	6,7778
MSE	12,2252	60,0018	128,5365	263,2963	1422,0560
RMSE	3,4965	7,7461	11,3374	16,2264	37,7102
PSNR	16,6016	23,9264	27,0405	30,3492	37,2582
SNR	10,9908	16,5153	20,2500	24,5131	33,8781
WSNR	12,1748	26,5613	33,7466	40,5053	55,1037
UQI	0,0138	0,4391	0,6557	0,8530	0,9925
VIF	0,1270	0,4022	0,5352	0,6661	0,9368
NQM	1,6597	19,3288	26,2037	31,8196	46,0276
SSIM	0,0916	0,7551	0,8825	0,9473	0,9969
MSSIM	0,3004	0,8420	0,9326	0,9709	0,9967
CQ(1,1)	0,0614	0,7789	0,9365	0,9864	1,0000
GSM	0,7282	0,9466	0,9866	0,9978	0,9999
GMSD	0,0002	0,0049	0,0266	0,0984	0,2702

Cuadro G.12: Medidas de Resumen para el grupo de distorsiones del Color

---

# Índice de figuras

1.1	Mapa de bits . . . . .	10
1.2	Convención de ejes . . . . .	11
1.3	Escala de grises . . . . .	12
1.4	El espacio del color RGB . . . . .	12
1.5	Cubo RGB . . . . .	13
1.6	Diagrama de bloques de un modelo HVS típico . . . . .	13
1.7	Efecto de enmascaramiento . . . . .	15
1.8	Visión Foveal . . . . .	15
2.1	Modelo de dos pasos para FR-IQA . . . . .	25
2.2	Ejemplo de convolución . . . . .	26
2.3	Fusión de Imágenes . . . . .	29
3.1	Imágenes de referencia de la base TID2013 . . . . .	32
3.2	Histograma y diagrama de caja del MOS . . . . .	33
3.3	Diagrama de dispersión del MOS . . . . .	33
3.4	Diagramas de dispersión e Histograma del MOS para la distorsión de tipo #1 . . . . .	34
3.5	Diagrama de dispersión para la distorsión de tipo #2 . . . . .	35
3.6	Histogramas del MOS para las distorsiones de tipo #2 y #3 . . . . .	36
3.7	Diagrama de dispersión para la distorsión de tipo #3. . . . .	37
3.8	Diagrama de dispersión para la distorsión de tipo #4. . . . .	37
3.9	Histogramas del MOS para las distorsiones #4 y #5. . . . .	38
3.10	Diagramas de dispersión para la distorsión de tipo #5. . . . .	38
3.11	Diagrama de dispersión para la distorsión de tipo #6 . . . . .	39
3.12	Histogramas del MOS para las distorsiones #6 y #7. . . . .	39
3.13	Diagrama de dispersión para la distorsión de tipo #7. . . . .	40
3.14	Diagrama de dispersión para la distorsión de tipo #8. . . . .	41
3.15	Histogramas del MOS para las distorsiones #8 y #9. . . . .	41
3.16	Diagrama de dispersión para la distorsión de tipo #9. . . . .	42
3.17	Diagrama de dispersión para la distorsión de tipo #10. . . . .	42
3.18	Histogramas del MOS para las distorsiones #10 y #11. . . . .	43

3.19	Diagrama de dispersión para la distorsión de tipo #11. . . . .	43
3.20	Diagrama de dispersión para la distorsión de tipo #12. . . . .	44
3.21	Histogramas del MOS para las distorsiones #12 y #13. . . . .	44
3.22	Diagrama de dispersión para la distorsión de tipo #13. . . . .	45
3.23	Diagrama de dispersión para la distorsión de tipo #14. . . . .	45
3.24	Histogramas del MOS para las distorsiones #14 y #15. . . . .	46
3.25	Diagrama de dispersión para la distorsión de tipo #15. . . . .	46
3.26	Diagrama de dispersión para la distorsión de tipo #16. . . . .	47
3.27	Histogramas del MOS para las distorsiones #16 y #17. . . . .	47
3.28	Diagrama de dispersión para la distorsión de tipo #17 . . . . .	48
3.29	Diagrama de dispersión para la distorsión de tipo #18. . . . .	48
3.30	Histogramas del MOS para las distorsiones #18 y #19. . . . .	49
3.31	Diagrama de dispersión para la distorsión de tipo #19. . . . .	50
3.32	Diagrama de dispersión para la distorsión de tipo #20. . . . .	50
3.33	Histogramas del MOS para las distorsiones #20 y #21. . . . .	51
3.34	Diagrama de dispersión para la distorsión de tipo #21. . . . .	51
3.36	Una ilustración de tramado. . . . .	51
3.35	Diagrama de dispersión para la distorsión de tipo #22. . . . .	52
3.37	Histogramas del MOS para las distorsiones #22 y #23. . . . .	52
3.38	Diagrama de dispersión para la distorsión de tipo #23. . . . .	53
3.39	Diagrama de dispersión e Histograma del MOS para la distorsión de tipo #24. . .	53
3.40	Diagrama de Venn distorsiones . . . . .	55
4.1	Histograma y diagrama de caja MSE . . . . .	58
4.2	Histograma y diagrama de caja RMSE . . . . .	59
4.3	Histograma y diagrama de caja PSNR . . . . .	59
4.4	Histograma y diagrama de caja SNR . . . . .	60
4.5	Histograma y diagrama de caja WSNR . . . . .	60
4.6	Histograma y diagrama de caja UQI . . . . .	60
4.7	Histograma y diagrama de caja VIF . . . . .	61
4.8	Histograma y diagrama de caja NQM . . . . .	61
4.9	Histograma y diagrama de caja SSIM . . . . .	61
4.10	Histograma y diagrama de caja MSSIM . . . . .	62
4.11	Histograma y diagrama de caja Índice CQ(1,1) . . . . .	62
4.12	Histograma y diagrama de caja GSM . . . . .	62
4.13	Histograma y diagrama de caja GMSD . . . . .	63
5.1	Diagrama de dispersión MSE vs MOS . . . . .	65
5.2	Diagrama de dispersión RMSE vs MOS . . . . .	65

5.3	Diagrama de dispersión PSNR vs MOS . . . . .	66
5.4	Diagrama de dispersión SNR vs MOS . . . . .	66
5.5	Diagrama de dispersión WSNR vs MOS . . . . .	67
5.6	Diagrama de dispersión UQI vs MOS . . . . .	67
5.7	Diagrama de dispersión VIF vs MOS . . . . .	68
5.8	Diagrama de dispersión NQM vs MOS . . . . .	68
5.9	Diagrama de dispersión SSIM vs MOS . . . . .	69
5.10	Diagrama de dispersión MSSIM vs MOS . . . . .	69
5.11	Diagrama de dispersión GSM SM vs MOS . . . . .	70
5.12	Diagrama de dispersión GSM SD vs MOS . . . . .	70
5.13	Diagrama de dispersión Índice CQ(1,1) vs MOS . . . . .	71
6.1	Ejemplo matriz de confusión para dos clases . . . . .	79
6.2	Categorización del MOS según terciles . . . . .	80
6.3	Matriz de confusión para las métricas MSE, RMSE y PSNR . . . . .	81
6.4	Matriz de confusión para la métrica SNR . . . . .	81
6.5	Matriz de confusión para la métrica WSNR . . . . .	82
6.6	Matriz de confusión para la métrica UQI . . . . .	82
6.7	Matriz de confusión para la métrica VIF . . . . .	83
6.8	Matriz de confusión para la métrica NQM . . . . .	83
6.9	Matriz de confusión para la métrica SSIM . . . . .	84
6.10	Matriz de confusión para la métrica MSSIM . . . . .	84
6.11	Matriz de confusión para el Índice CQ(1,1) . . . . .	85
6.12	Matriz de confusión para la métrica GSM SM . . . . .	85
6.13	Matriz de confusión para la métrica GSM SD . . . . .	86
7.1	Diagrama de Venn distorsiones . . . . .	96
A.1	Distorsión #1: Additive Gaussian noise . . . . .	112
A.2	Distorsión #1: Additive Gaussian noise . . . . .	113
A.3	Distorsión #1: Additive Gaussian noise . . . . .	114
A.4	Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component . . . . .	115
A.5	Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component . . . . .	116
A.6	Distorsión #2: Additive noise in color components is more intensive than additive noise in the luminance component . . . . .	117
A.7	Distorsión #3: Spatially correlated noise . . . . .	118
A.8	Distorsión #3: Spatially correlated noise . . . . .	119

---

A.9	Distorsión #3: Spatially correlated noise . . . . .	120
A.10	Distorsión #4: Masked noise . . . . .	121
A.11	Distorsión #4: Masked noise . . . . .	122
A.12	Distorsión #4: Masked noise . . . . .	123
A.13	Distorsión #5: High frequency noise . . . . .	124
A.14	Distorsión #5: High frequency noise . . . . .	125
A.15	Distorsión #5: High frequency noise . . . . .	126
A.16	Distorsión #6: Impulse noise . . . . .	127
A.17	Distorsión #6: Impulse noise . . . . .	128
A.18	Distorsión #6: Impulse noise . . . . .	129
A.19	Distorsión #7: Quantization noise . . . . .	130
A.20	Distorsión #7: Quantization noise . . . . .	131
A.21	Distorsión #7: Quantization noise . . . . .	132
A.22	Distorsión #8: Gaussian blur . . . . .	133
A.23	Distorsión #8: Gaussian blur . . . . .	134
A.24	Distorsión #8: Gaussian blur . . . . .	135
A.25	Distorsión #9: Image denoising . . . . .	136
A.26	Distorsión #9: Image denoising . . . . .	137
A.27	Distorsión #9: Image denoising . . . . .	138
B.1	Distorsión #10: JPEG compression . . . . .	139
B.2	Distorsión #10: JPEG compression . . . . .	140
B.3	Distorsión #10: JPEG compression . . . . .	141
B.4	Distorsión #11: JPEG2000 compression . . . . .	142
B.5	Distorsión #11: JPEG2000 compression . . . . .	143
B.6	Distorsión #11: JPEG2000 compression . . . . .	144
B.7	Distorsión #12: JPEG transmission errors . . . . .	145
B.8	Distorsión #12: JPEG transmission errors . . . . .	146
B.9	Distorsión #12: JPEG transmission errors . . . . .	147
B.10	Distorsión #13: JPEG2000 transmission errors . . . . .	148
B.11	Distorsión #13: JPEG2000 transmission errors . . . . .	149
B.12	Distorsión #13: JPEG2000 transmission errors . . . . .	150
B.13	Distorsión #14: Non eccentricity pattern noise . . . . .	151
B.14	Distorsión #14: Non eccentricity pattern noise . . . . .	152
B.15	Distorsión #14: Non eccentricity pattern noise . . . . .	153
B.16	Distorsión #15: Local block-wise distortions of different intensity . . . . .	154
B.17	Distorsión #15: Local block-wise distortions of different intensity . . . . .	155
B.18	Distorsión #15: Local block-wise distortions of different intensity . . . . .	156

B.19	Distorsión #16: Mean shift . . . . .	157
B.20	Distorsión #16: Mean shift . . . . .	158
B.21	Distorsión #16: Mean shift . . . . .	159
B.22	Distorsión #17: Contrast change . . . . .	160
B.23	Distorsión #17: Contrast change . . . . .	161
B.24	Distorsión #17: Contrast change . . . . .	162
C.1	Distorsión #18: Change of color saturation . . . . .	163
C.2	Distorsión #18: Change of color saturation . . . . .	164
C.3	Distorsión #18: Change of color saturation . . . . .	165
C.4	Distorsión #19: Multiplicative Gaussian noise . . . . .	166
C.5	Distorsión #19: Multiplicative Gaussian noise . . . . .	167
C.6	Distorsión #19: Multiplicative Gaussian noise . . . . .	168
C.7	Distorsión #20: Comfort noise . . . . .	169
C.8	Distorsión #20: Comfort noise . . . . .	170
C.9	Distorsión #20: Comfort noise . . . . .	171
C.10	Distorsión #21: Lossy compression of noisy images . . . . .	172
C.11	Distorsión #21: Lossy compression of noisy images . . . . .	173
C.12	Distorsión #21: Lossy compression of noisy images . . . . .	174
C.13	Distorsión #22: Image color quantization with dither . . . . .	175
C.14	Distorsión #22: Image color quantization with dither . . . . .	176
C.15	Distorsión #22: Image color quantization with dither . . . . .	177
C.16	Distorsión #23: Chromatic aberrations . . . . .	178
C.17	Distorsión #23: Chromatic aberrations . . . . .	179
C.18	Distorsión #23: Chromatic aberrations . . . . .	180
C.19	Distorsión #24: Sparse sampling and reconstruction . . . . .	181
C.20	Distorsión #24: Sparse sampling and reconstruction . . . . .	182
C.21	Distorsión #24: Sparse sampling and reconstruction . . . . .	183
E.1	Grupo 1: Distorsiones Ruidosas. Matrices de confusión para las métricas MSE, RMSE y PSNR . . . . .	191
E.2	Grupo 1: Distorsiones Ruidosas. Métricas SNR y WSNR. . . . .	192
E.3	Grupo 1: Distorsiones Ruidosas. Métricas UQI, VIF y NQM. . . . .	193
E.4	Grupo 1: Distorsiones Ruidosas. Métricas SSIM y MSSIM. . . . .	194
E.5	Grupo 1: Distorsiones Ruidosas. Índice CQ(1,1). . . . .	194
E.6	Grupo 1: Distorsiones Ruidosas. Métricas GSM y GSM. . . . .	195
E.7	Grupo 2: Distorsiones Actuales. Matrices de confusión para las métricas MSE, RMSE y PSNR . . . . .	195
E.8	Grupo 2: Distorsiones Actuales. Métrica SNR y WSNR. . . . .	196



---

E.9	Grupo 2: Distorsiones Actuales. Métricas UQI, VIF y NQM. . . . .	197
E.10	Grupo 2: Distorsiones Actuales. Métricas SSIM y MSSIM. . . . .	198
E.11	Grupo 2: Distorsiones Actuales. Índice CQ(1,1). . . . .	198
E.12	Grupo 2: Distorsiones Actuales. Métricas GSM y GMSD. . . . .	199
E.13	Grupo 3: Distorsiones Simples. Matrices de confusión para las métricas MSE, RMSE y PSNR . . . . .	199
E.14	Grupo 3: Distorsiones Simples. Métricas SNR y WSNR. . . . .	200
E.15	Grupo 3: Distorsiones Simples. Métricas UQI, VIF y NQM. . . . .	201
E.16	Grupo 3: Distorsiones Simples. Métricas SSIM y MSSIM. . . . .	202
E.17	Grupo 3: Distorsiones Simples. Índice CQ(1,1). . . . .	202
E.18	Grupo 3: Distorsiones Simples. Métricas GSM y GMSD. . . . .	203
E.19	Grupo 4: Distorsiones Exóticas. Métricas MSE, RMSE y PSNR. . . . .	203
E.20	Grupo 4: Distorsiones Exóticas. Métricas SNR y WSNR. . . . .	204
E.21	Grupo 4: Distorsiones Exóticas. Métricas UQI, VIF y NQM. . . . .	205
E.22	Grupo 4: Distorsiones Exóticas. Métricas SSIM y MSSIM. . . . .	206
E.23	Grupo 4: Distorsiones Exóticas. Índice CQ(1,1). . . . .	206
E.24	Grupo 4: Distorsiones Exóticas. Métricas GSM y GMSD. . . . .	207
E.25	Grupo 5: Distorsiones Nuevas. Métricas MSE, RMSE y PSNR. . . . .	207
E.26	Grupo 5: Distorsiones Nuevas. Métricas SNR y WSNR. . . . .	208
E.27	Grupo 5: Distorsiones Nuevas. Métricas UQI, VIF y NQM. . . . .	209
E.28	Grupo 5: Distorsiones Nuevas. Métricas SSIM y MSSIM. . . . .	210
E.29	Grupo 5: Distorsiones Nuevas. Índice CQ(1,1). . . . .	210
E.30	Grupo 5: Distorsiones Nuevas. Métricas GSM y GMSD. . . . .	211
E.31	Grupo 6: Distorsiones del Color. Métricas MSE, RMSE y PSNR. . . . .	211
E.32	Grupo 6: Distorsiones del Color. Métricas SNR y WSNR. . . . .	212
E.33	Grupo 6: Distorsiones del Color. Métricas UQI, VIF y NQM. . . . .	213
E.34	Grupo 6: Distorsiones del Color. Métricas SSIM y MSSIM. . . . .	214
E.35	Grupo 6: Distorsiones del Color. Índice CQ(1,1). . . . .	214
E.36	Grupo 6: Distorsiones del Color. Métricas GSM y GMSD. . . . .	215

## Índice de cuadros

3.1	Grupos de distorsiones . . . . .	54
-----	----------------------------------	----

4.1	Matriz 3000 x 19 con los datos a analizar . . . . .	56
4.2	Matriz 125 x 19 con los datos para la distorsión #6 . . . . .	57
4.3	Medidas de Resumen del MOS y las métricas . . . . .	58
4.4	Medidas de Resumen del MOS y las métricas . . . . .	59
5.1	Interpretación del coeficiente de correlación estimado . . . . .	72
5.2	Coefficientes de Kendall y Spearman entre el MOS y cada métrica . . . . .	74
6.1	Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	86
6.2	Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	87
6.3	Interpretación del coeficiente kappa . . . . .	89
6.4	Coefficientes de Kendall, coeficiente de Spearman, Kappa de Cohen y Pi de Scott entre el MOS y cada métrica . . . . .	89
6.5	Tabla para calcular las proporciones conjuntas y la probabilidad de acuerdo por azar	90
7.1	Grupos de distorsiones . . . . .	95
7.2	Correlación de Spearman entre las métricas y el MOS por grupos de distorsiones .	97
7.3	Correlación de Kendall entre las métricas y el MOS por grupos de distorsiones . .	97
7.4	Índice kappa de Cohen de cada métrica con el MOS en cada grupo de distorsiones	98
7.5	Índice pi de Scott de cada métrica con el MOS en cada grupo de distorsiones . . .	98
7.6	Ranking1 . . . . .	99
7.7	Ranking2 . . . . .	100
7.8	Ranking3 . . . . .	101
7.9	Ranking final . . . . .	102
7.10	Índice de concordancia w de Kendall de las métricas con el MOS por grupo de distorsiones . . . . .	102
7.11	Índice kappa de Fleiss de las métricas con el MOS por grupo de distorsiones . . .	102
7.12	Coefficiente de concordancia w de Kendall de las métricas con el MOS por grupo de distorsiones II . . . . .	103
7.13	Índice kappa de Fleiss de las métricas con el MOS por grupo de distorsiones II . .	103
8.1	Métrica más apropiada por grupo de distorsión. . . . .	104
E.1	Grupo 1: Distorsiones Ruidosas. Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	192
E.2	Grupo 1: Distorsiones Ruidosas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	193
E.3	Grupo 2: Distorsiones Actuales. Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	196

E.4	Grupo 2: Distorsiones Actuales. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	197
E.5	Grupo 3: Distorsiones Simples. Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	200
E.6	Grupo 3: Distorsiones Simples. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	201
E.7	Grupo 4: Distorsiones Exóticas. Porcentaje de imágenes bien clasificadas por métrica, según calidad. Es curioso que VIF clasificó correctamente el mayor número de imágenes de calidad alta y el menor número de imágenes de calidad media. . .	204
E.8	Grupo 4: Distorsiones Exóticas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	205
E.9	Grupo 5: Distorsiones Nuevas. Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	208
E.10	Grupo 5: Distorsiones Nuevas. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	209
E.11	Grupo 6: Distorsiones del Color. Porcentaje de imágenes bien clasificadas por métrica, según calidad. . . . .	212
E.12	Grupo 6: Distorsiones del Color. Ranking métricas que hicieron el mayor número de clasificaciones correctas según calidad. . . . .	213
G.1	Medidas de Resumen para el grupo de distorsiones Ruidosas . . . . .	221
G.2	Medidas de Resumen para el grupo de distorsiones Ruidosas . . . . .	222
G.3	Medidas de Resumen para el grupo de distorsiones Actuales . . . . .	222
G.4	Medidas de Resumen para el grupo de distorsiones Actuales . . . . .	223
G.5	Medidas de Resumen para el grupo de distorsiones Simples . . . . .	223
G.6	Medidas de Resumen para el grupo de distorsiones Simples . . . . .	224
G.7	Medidas de Resumen para el grupo de distorsiones Exóticas . . . . .	224
G.8	Medidas de Resumen para el grupo de distorsiones Exóticas . . . . .	225
G.9	Medidas de Resumen para el grupo de distorsiones Nuevas . . . . .	225
G.10	Medidas de Resumen para el grupo de distorsiones Nuevas . . . . .	226
G.11	Medidas de Resumen para el grupo de distorsiones del Color . . . . .	226
G.12	Medidas de Resumen para el grupo de distorsiones del Color . . . . .	227