

## Certamen 2 - MAT467

ESTUDIANTE: CLEMENTE FERRER

- 1. a)** En un contexto espacio-temporal, demuestre que

$$\gamma(\mathbf{h}, u) = C_s(\mathbf{0})\gamma_t(u) + C_t(0)\gamma_s(\mathbf{h}) - \gamma_s(\mathbf{h})\gamma_t(u).$$

**Solución:** Asumiendo una función de covarianza de la forma  $C(\mathbf{h}, u) = C_s(\mathbf{h})C_t(u)$ , se tiene que

$$\begin{aligned} \gamma(\mathbf{h}, u) &= C(\mathbf{0}, 0) - C(\mathbf{h}, u) \\ &= C_s(\mathbf{0})C_t(0) - C_s(\mathbf{h})C_t(u) \\ &= C_s(\mathbf{0})C_t(0) - C_s(\mathbf{h})C_t(u) + C_s(\mathbf{0})C_t(u) - C_s(\mathbf{0})C_t(u) \\ &\quad + C_s(\mathbf{h})C_t(0) - C_s(\mathbf{h})C_t(0) + C_s(\mathbf{0})C_t(0) - C_s(\mathbf{0})C_t(0) \\ &= C_s(\mathbf{0})C_t(0) - C_s(\mathbf{0})C_t(u) + C_t(0)C_s(\mathbf{0}) - C_t(0)C_s(\mathbf{h}) \\ &\quad - C_s(\mathbf{0})C_t(0) + C_s(\mathbf{0})C_t(u) + C_s(\mathbf{h})C_t(0) - C_s(\mathbf{h})C_t(u) \\ &= C_s(\mathbf{0})(C_t(0) - C_t(u)) + C_t(0)(C_s(\mathbf{0}) - C_s(\mathbf{h})) \\ &\quad - (C_s(\mathbf{0}) - C_s(\mathbf{h}))(C_t(0) - C_t(u)) \\ &= C_s(\mathbf{0})\gamma_t(u) + C_t(0)\gamma_s(\mathbf{h}) - \gamma_s(\mathbf{h})\gamma_t(u) \end{aligned}$$

Donde en la última igualdad se usaron las propiedades vistas para el variograma temporal y espacial

$$\gamma_t(u) = C_t(0) - C_t(u) \quad \text{y} \quad \gamma_s(\mathbf{h}) = C_s(\mathbf{0}) - C_s(\mathbf{h}).$$

**Observación:** El espíritu de la demostración se basa en sumar ceros convenientes como se puede apreciar en la tercera igualdad.

- b)** Desde a) ¿es posible deducir que si la función de covarianza es separable, entonces el variograma respectivo también es separable?

**Solución:** En a) usamos una función de covarianza separable (en el sentido multiplicativo) y obtuvimos un variograma en función de los variogramas marginales (espaciales y temporales). Por lo anterior, entendiendo *separabilidad* del variograma como la nula interacción simultánea entre el espacio y el tiempo, concluimos que sí es posible deducir la afirmación del enunciado.

- 2.** Considere la siguiente función de covarianza

$$C(\mathbf{h}, u) = k_1C_s(\mathbf{h})C_t(u) + k_2C_s(\mathbf{h}) + k_3C_t(u),$$

donde  $C_s(\cdot)$  es una función de covarianza espacial,  $C_t(\cdot)$  es una función de covarianza temporal, y  $k_1 > 0$ ,  $k_2 > 0$  y  $k_3 > 0$ , son constantes que garantizan la validez de  $C(\mathbf{h}, u)$ . Demuestre que para un proceso espaciotemporal estacionario,

$$\gamma(\mathbf{h}, u) = (k_2 + k_1C_t(0))\gamma_s(\mathbf{h}) + (k_3 + k_1C_s(\mathbf{0}))\gamma_t(u) - k_1\gamma_s(\mathbf{h})\gamma_t(u),$$

donde  $\gamma_s$  y  $\gamma_t$  son los correspondientes semivariogramas,  $C(\mathbf{0}, 0)$  es el sill global de  $\gamma$ ,  $C_s(\mathbf{0})$  es el sill de  $\gamma_s$  y  $C_t(0)$  es el sill de  $\gamma_t$

**Solución:** De manera directa, note que

$$\begin{aligned}
\gamma(\mathbf{h}, u) &= C(\mathbf{0}, 0) - C(\mathbf{h}, u) \\
&= k_1 C_s(\mathbf{0}) C_t(0) + k_2 C_s(\mathbf{0}) + k_3 C_t(0) - k_1 C_s(\mathbf{h}) C_t(u) - k_2 C_s(\mathbf{h}) - k_3 C_t(u) \\
&= k_1 C_s(\mathbf{0}) C_t(0) + k_2 C_s(\mathbf{0}) + k_3 C_t(0) - k_1 C_s(\mathbf{h}) C_t(u) - k_2 C_s(\mathbf{h}) - k_3 C_t(u) \\
&\quad + k_1 C_t(0) C_s(\mathbf{h}) - k_1 C_t(0) C_s(\mathbf{h}) + k_1 C_t(u) C_s(\mathbf{0}) - k_1 C_t(u) C_s(\mathbf{0}) \\
&\quad + k_1 C_t(0) C_s(\mathbf{0}) - k_1 C_t(0) C_s(\mathbf{0}) \\
&= k_2 C_s(\mathbf{0}) - k_2 C_s(\mathbf{h}) + k_1 C_t(0) C_s(\mathbf{0}) - k_1 C_t(0) C_s(\mathbf{h}) \\
&\quad + k_3 C_t(0) - k_3 C_t(u) + k_1 C_t(0) C_s(\mathbf{0}) - k_1 C_t(u) C_s(\mathbf{0}) \\
&\quad - k_1 C_t(0) C_s(\mathbf{0}) + k_1 C_t(u) C_s(\mathbf{0}) + k_1 C_t(0) C_s(\mathbf{h}) - k_1 C_t(u) C_s(\mathbf{h}) \\
&= (k_2 + k_1 C_t(0))(C_s(\mathbf{0}) - C_s(\mathbf{h})) + (k_3 + k_1 C_s(\mathbf{0}))(C_t(0) - C_t(u)) \\
&\quad - k_1(C_s(\mathbf{0}) - C_s(\mathbf{h}))(C_t(0) - C_t(u)) \\
&= (k_2 + k_1 C_t(0))\gamma_s(\mathbf{h}) + (k_3 + k_1 C_s(\mathbf{0}))\gamma_t(u) - k_1\gamma_s(\mathbf{h})\gamma_t(u)
\end{aligned}$$

Donde en la última igualdad se usaron las propiedades vistas para el variograma temporal y espacial

$$\gamma_t(u) = C_t(0) - C_t(u) \quad \text{y} \quad \gamma_s(\mathbf{h}) = C_s(\mathbf{0}) - C_s(\mathbf{h}).$$

**Observación:** El espíritu de la demostración, al igual que en el ejercicio anterior, se basa en sumar ceros convenientes como se puede apreciar en la tercera igualdad.

3. Considere el conjunto de imágenes tomadas por una de las cámaras del instituto Harvard Forest, USA, entre el 1 y 5 de octubre cada año entre los años 2008 al 2022. Es decir, esta base de datos contiene 15 imágenes tomadas aproximadamente en la misma fecha del año, durante 15 años consecutivos. Las imágenes originales y también versiones recortadas de las mismas pueden ser descargadas desde el sitio siguiente.

Las versiones recortadas consisten en imágenes más pequeñas de las zonas de interés de las imágenes originales, evitando la parte superior (cielo). El objetivo es estudiar el nivel de verde de estas imágenes.

Una imagen es un arreglo que consiste en tres imágenes del mismo tamaño que representan los colores red (R), green (G) y blue (B). Luego, lo que nosotros observamos cuando se usa algún comando para graficar una imagen en el plano es una combinación lineal de estos tres colores pixel a pixel.

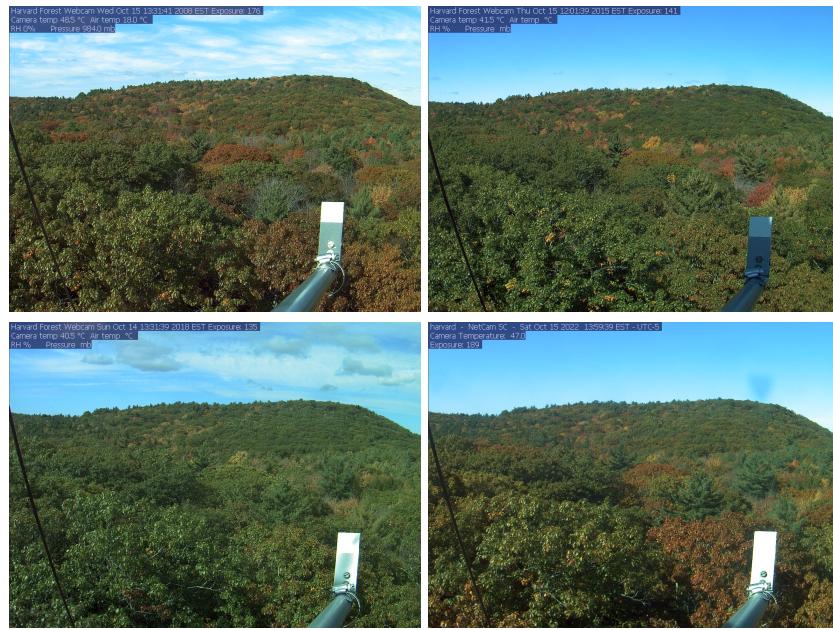
- Como nos interesa sólo el nivel de verde construya para cada imagen recortada, su correspondiente imagen del color verde normalizado ( $G_{cc}$  green chromatic coordinate) usando la fórmula para cada pixel

$$G_{cc}(i, j) = \frac{G(i, j)}{G(i, j) + R(i, j) + B(i, j)}$$

**Solución<sup>1</sup>:** Antes de realizar el análisis, es importante visualizar algunas mediciones y motivar la resolución del problema. La figura 1 muestra la primera y última fotografía de la base de datos respectivamente.

---

<sup>1</sup>Los códigos usados pueden hallarse en el siguiente GitHub: [clementeferrer](#) en la carpeta MAT467 – Estadística Espacial



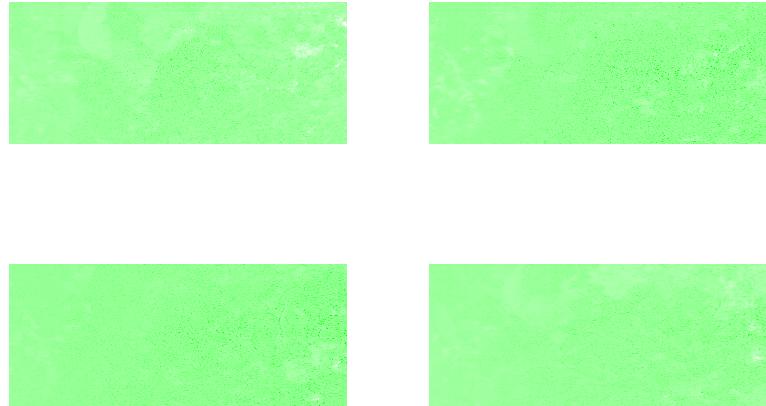
**Figura 1.** Imágenes del instituto Harvard Forest durante los años 2008, 2015, 2018 y 2022.

Ahora bien, el nivel de verde de las especies arbóreas y arbustivas es vital para la vida. Zhu et al. (2016) concluyeron la relación directa entre este factor y la disminución de CO<sub>2</sub> de la atmósfera. Por lo anterior, descartaremos del estudio la zona del cielo en cada imagen y trabajaremos con fotografías recortadas como muestra la figura 2.



**Figura 2.** Imágenes recortadas durante los años 2008, 2015, 2018 y 2022.

Dado que nos interesa sólo el nivel de verde para cada imagen recortada, se obtuvo su correspondiente imagen del color verde normalizado  $G_{cc}$ . Esto fue realizado en **R** y se obtuvieron los siguientes resultados



**Figura 3.** Imagen  $G_{cc}$  de los años 2008, 2015, 2018 y 2022.

**Observación:** Los colores RGB pueden verse como vectores dentro del cubo unitario, es decir,  $v \in [0, 1]^3 \subset \mathbb{R}^3$ , donde  $v = (R, G, B)$ . Al normalizar el color verde se colapsan las coordenadas restantes,  $v_{gcc} = \left(0, \frac{G}{\|G\|}, 0\right)$ . Así, solo estudiamos la *saturación* del color resultante que genera dicho proceso en vez de todos los canales de la imagen. Es por ello que se agregó manualmente una capa de color arbitrario para facilitar la visualización. Esto no entorpece el análisis por la discusión previa.

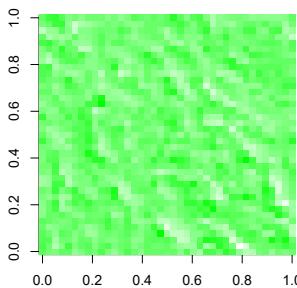
- b) Considere la imagen  $G_{cc}$  correspondiente al año 2018. Elija una subimagen de  $40 \times 40$  y ajuste un modelo autoregresivo espacial a la subimagen. Reporte las estimaciones de los parámetros y obtenga y grafique la imagen residuo.

**Solución:** La subimagen seleccionada correspondió al cuadrado de  $40 \times 40$  en la esquina inferior izquierda de las fotografías recortadas.

Además, se ajustó el modelo autoregresivo espacial estudiado por Reinsel & Basu (1993):

$$X(i, j) = \phi_1 X(i - 1, j) + \phi_2 X(i, j - 1) + \epsilon(i, j),$$

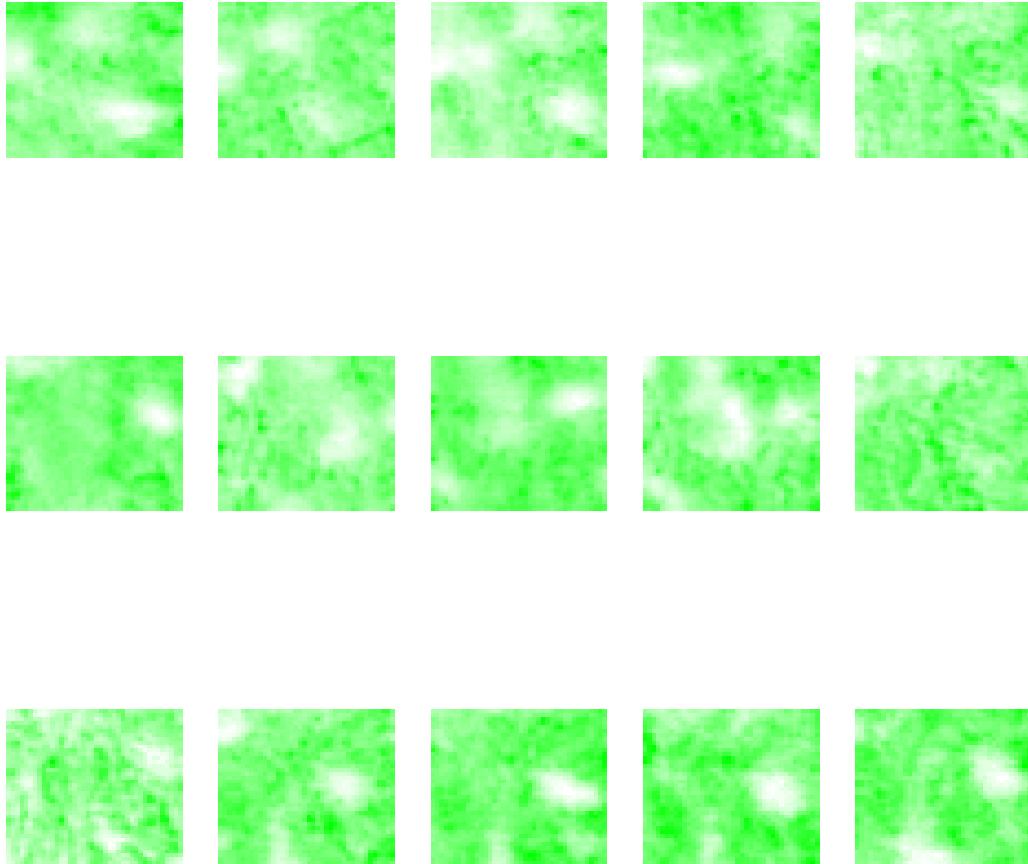
donde  $\{\epsilon(i, j)\}$  es un ruido blanco con varianza  $\sigma^2$ . El código fue realizado en su totalidad y se obtuvo que  $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2) = (0,4056380; 0,5944509)$ , donde se observó un patrón regular en los residuos como se aprecia en la figura 4.



**Figura 4.** Proceso autoregresivo ajustado a la imagen  $G_{cc}$  del 2018.

- c) Fije una grilla de tamaño  $40 \times 40$  dentro de cada imagen recortada que permanezca fija en el tiempo y grafique dicha secuencia de imágenes ( $G_{cc}$ ) en el tiempo como realizaciones de un proceso espacio temporal con  $t = 1, \dots, 15$ .

**Solución:** La fijación de la grilla para la secuencia de imágenes se muestra en la figura 5.



**Figura 5.** Grilla  $40 \times 40$  para cada imagen  $G_{cc}$  desde 2008 hasta 2022.

- d) Usando la secuencia de 15 imágenes construida en c) ajuste un modelo espaciotemporal de la forma

$$Z(\mathbf{s}, t) = \mu(t) + \epsilon(\mathbf{s}, t),$$

donde  $\mu(t) = \beta_0 + \beta_1 t$ , y el proceso  $\epsilon(\mathbf{s}, t)$  es un ruido Gausiano con media nula y matriz de covarianza  $\Sigma$ . (En su implementación use el modelo de covarianza espacio-temporal de Gneiting). Reporte las estimaciones de los parámetros y un breve análisis de residuos. Comente sobre la calidad del ajuste.

**Solución:** Se utilizó la librería GeoModels creada por Bevilacqua et al. (2018). Además, el modelo implementado posee la estructura de covarianza de Gneiting (2002) en el ruido Gaussiano, i.e.

$$C(\mathbf{s}, t, \boldsymbol{\theta}) = \frac{\sigma^2}{(|t|^{2\gamma} + 1)^\tau} \exp\left(\frac{-c\|\mathbf{s}\|^{2\gamma}}{(|t|^{2\gamma} + 1)^{\beta\gamma}}\right),$$

donde  $\boldsymbol{\theta} = (\tau, \gamma, c, \beta, \sigma^2)$ . Sin embargo, dada la tendencia lineal de la media, en verdad buscamos estimar  $\boldsymbol{\vartheta} = (\tau, \gamma, c, \beta, \sigma^2, \beta_0, \beta_1)$ . Para ello, dada la realización

espacio-temporal  $\{Z(s_i, t_l)\}$ , sea  $f_Z(z_{il}, z_{jk})$  la densidad Gaussiana de un par de observaciones  $Z(s_i, t_l)$  y  $Z(s_j, t_k)$ . Usaremos, la *función de verosimilitud de a pares* definida por Bevilacqua & Gaetan (2015) para realizar las estimaciones pertinentes:

$$pl(\vartheta) = \sum_{i,j,l,k \in D} \log(f_Y(y_{il}, y_{jk})) w_{ijkl},$$

donde  $w_{ijkl}$  son pesos no negativos que no dependen de  $\vartheta$  tales que

$$w_{ijkl} = \begin{cases} 1 & \|s_i - s_j\| < d_s, |t_l - t_k| < d_t \\ 0 & \text{en otro caso.} \end{cases}$$

La estimación mediante  $pl(\vartheta)$  fue ejecutada usando la función GeoFit de la librería mencionada anteriormente y los resultados fueron:

```
#####
Maximum Composite-Likelihood Fitting of Gaussian Random Fields

Setting: Marginal Composite-Likelihood
Model: Gaussian
Type of the likelihood objects: Pairwise
Covariance model: Gneiting
Optimizer: BFGS

Number of spatial coordinates: 24000
Number of dependent temporal realisations: 15
Type of the random field: univariate
Number of estimated parameters: 7

Type of convergence: Successful
Maximum log-Composite-Likelihood value: 3525641.22

Estimated parameters:
  mean      mean1    power_s    power_t    scale_s    scale_t     sill
0.3881062  0.0002751  0.6571025  0.2547918  0.4502391  2.0078524  0.0010223
#####
```

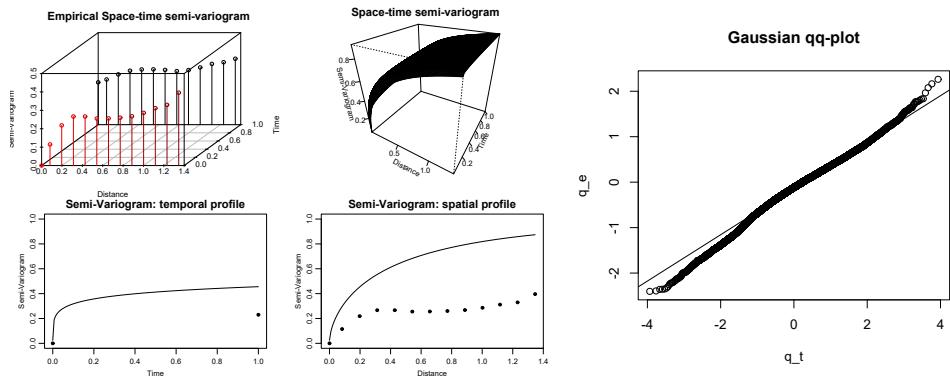
**Figura 6.** Output GeoFit

Ahora bien, dadas las estimaciones de los parámetros, los residuos estimados satisfacen

$$\hat{\epsilon}(s, t) = Z(s, t) - \hat{\mu}(t).$$

Luego, se debe cumplir que la media de estos debe ser cercana a cero para un correcto supuesto de normalidad. En particular, se obtuvo que  $\mathbb{E}[\hat{\epsilon}(s, t)] = -0.1565987$ . Más aún, la figura 7 presenta los semivariogramas para los residuos y el respectivo QQPlot.

A partir de lo anterior, se concluyó que la normalidad de los residuos no es clara, puesto que la media no es lo suficientemente cercana a cero y el ajuste del semivariograma no es preciso. Se propone aplicar una transformación a las imágenes recortadas en grillas para obtener normalidad o considerar otro modelo de covarianza como el de Wendland.



**Figura 7.** Semivariogramas y QQPlot para los residuos

- e) Usando el modelo ajustado en la parte d) prediga el valor medio de la imagen en el instante  $t = 16$ .

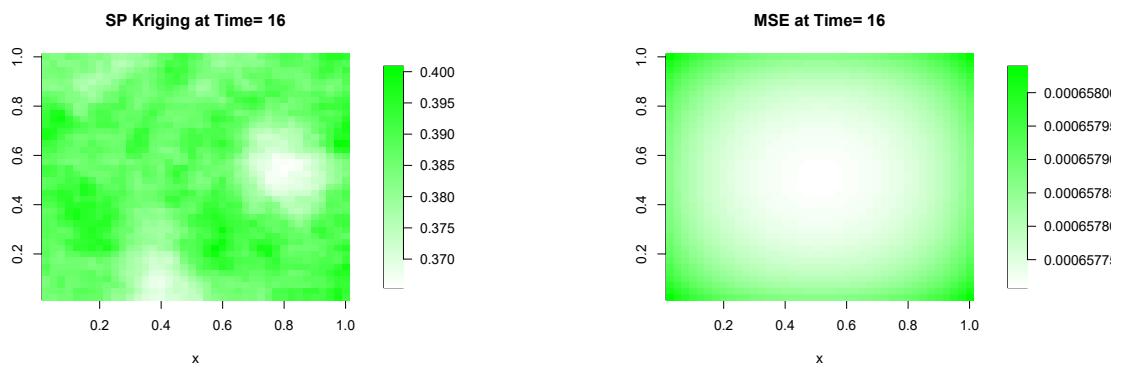
**Solución:** De acuerdo a Bevilacqua et al. (2018), la predicción óptima para un proceso espacio-temporal está dada por

$$\widehat{Z}(\mathbf{s}_0, t_0) = \hat{\mu}(t) + \sum_{l=1}^T \sum_{i=1}^N \lambda_{l,i} [\hat{\epsilon}(\mathbf{s}_i, t_l)],$$

donde el vector de pesos  $\boldsymbol{\lambda} = (\lambda_{1,1}, \dots, \lambda_{T,N})$  es dado por  $\boldsymbol{\lambda} = R^{-1}\mathbf{c}$  y

- $\mathbf{c} = (\text{cor}(Z(s_0, t_0), Z(s_1, t_1)), \dots, \text{cor}(Z(s_0, t_0), Z(s_N, t_T)))^T$ .
- $R = \left[ [\text{cor}(Z(s_i, t_l), Z(s_j, t_k))]_{l,k=1}^T \right]_{i,j=1}^N$  es la matriz de correlación.

En particular, la predicción fue realizada a partir del estimador  $\hat{\vartheta}$  y la función GeoKrig. Los resultados, junto con el MSE se presentan en la figura 8.



**Figura 8.** Predicción para 2023 y MSE asociado.

A modo de conclusión, puede notarse que la predicción para el año 2023 presenta un patrón similar a los últimos tres años, destacando un leve aumento en la saturación del color. Además, el error cuadrático medio posee una tendencia radial, lo que motiva a extender el análisis realizado usando coordenadas curvilíneas en los modelos a desarrollar.