

Etapa 3 Interpolarea spațio-temporală a temperaturii aerului pe teritoriul României

Alexandru Dumitrescu

Cuprins

1	Introducere.....	2
2	Evaluarea metodelor de interpolare spațială.....	3
2.1	Metoda anomaliilor	4
2.1.1	Hărțile zilnice de temperatura aerului.....	4
2.1.2	Hărțile abaterilor orare de temperatura aerului	5
2.2	Metoda spațio-temporală.....	7
2.3	Metode de validare a rezultatelor interpolării	7
2.4	Rezultate	8
2.4.1	Analiza exploratorie	8
2.4.2	Analiza estimărilor metodelor de interpolare	11
3	Estimarea spațio-temporală la scară fină a datelor orare de temperatura aerului	13
4	Validarea setului de date gridate de temperatura orară a aerului	15
5	Temperaturi extreme caracteristice și indici termici extremi.....	19
5.1	Temperaturi extreme caracteristice.....	20
5.1.1	Mediile lunare ale temperaturilor extreme zilnice	20
5.1.2	Numărul de zile cu diferite temperaturi caracteristice	22
5.2	Valuri de căldură și frig.....	26
6	Publicarea online a seturilor de date gridate și a manualelor de utilizare a produselor 27	
	Bibliografie.....	30
	Anexe	33
	Anexa 1: Descrierea datelor gridate.....	33
	Anexa 2: Exemple de manipulare a datelor gridate utilizând limbajul R.....	34

1 Introducere

În meteorologie și climatologie, informația geo-spațială de bază este constituită din date prevalate punctual, în amplasamentele unde sunt efectuate măsurătorile și observațiile meteorologice. Pentru obținerea informației din arealele unde nu sunt amplasate stațiile meteorologice, se utilizează metode de interpolare spațială (geostatistice), având la bază o serie de funcții matematice, cu ajutorul cărora se poate estima valoarea unui punct dat în funcție de măsurătorile efectuate în imediata apropiere a aceluși punct. Pentru îmbunătățirea acurateții estimărilor metodelor de interpolare spațială, se pot utiliza variabile explicative extrase din modelul numeric altitudinal al terenului (MNA), din prelucrarea imaginilor satelitare și a datelor furnizate de radarele meteorologice (Aalto et al. 2013; Dumitrescu and Birsan 2015; Dumitrescu et al. 2016; Dumitrescu et al. 2017; Dumitrescu et al. 2020b).

În această etapă a proiectului Interpolarea Spațio-Temporală La Scară Fină A Temperaturii Aerului Pentru Teritoriul României (High resolution gridded air temperature dataset over Romania - HiGridTemp) s-au realizat seturi de date gridate orare privind temperatura aerului (Taer) pe teritoriul României, la o rezoluție spațială de 1000 m × 1000 m, pentru perioada 01 ianuarie 2009 - 31 decembrie 2017. Datele obținute în Etapa 1, respectiv seturi de date orare de temperatura aerului omogenizate statistic, obținute din măsurători de suprafață, efectuate între anii 2009-2017 (Dumitrescu et al. 2020c), au fost interpolate spațial utilizându-se o metodă geostatistică multi-variată, care a luat în calcul și variabilele explicative obținute în Etapa 2, respectiv predictorii derivației din MNA și datele de temperatură a suprafeței subiacente (LST), extrase din procesarea imaginilor satelitare SEVIRI (Dumitrescu et al. 2020a; LSAF 2018).

Seturi de date gridate privind temperatura orară a aerului s-au mai realizat și pentru teritoriul Germaniei, pentru intervalul 1995 - 2012, utilizându-se o metodă hibridă de spațializare, care îmbină regresia liniară multiplă cu metoda de interpolare bazată pe inversul distanței (Krähenmann et al. 2018). Meyer et al. (2019) au aplicat metoda de regresie Random Forest și informațiile furnizate de canalele spectrale MSG SEVIRI pentru estimarea spațială a temperaturii orare a aerului în Africa de Sud. Lussana et al. (2019) aplică filtrul Kalman pentru estimarea spațială a temperaturii orare a aerului pentru un an de zile (31 iulie 2017 - 31 iulie 2018), utilizând date măsurate la stațiile meteorologice din Norvegia și ansamblu de prognoze meteo, furnizate de modelele numerice. Date orare gridate de temperatura aerului la o rezoluție spațială grosieră pot fi obținute, la scară globală sau continentală, și din modele de reanaliză climatică ERA5 Land¹, COSMO-REA6² și UERRA³.

¹ <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview>

² <https://reanalysis.meteo.uni-bonn.de/?COSMO-REA6>

³ <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-uerra-europe-single-levels?tab=overview>

2 Evaluarea metodelor de interpolare spațială

În această etapă au fost analizate două categorii de metode de interpolare: (1) metoda anomaliilor, bazată pe calculul diferențelor dintre mediile zilnice și valorile orare de temperatură, și (2) metoda spațio-temporală Regression Kriging (RK), care cuantifică la nivel orar relațiile din temperatura aerului și variabilele auxiliare selectate (SEVIRI LST și predictorii derivați din MNA). Având în vedere volumul mare de date disponibile, evaluarea metodelor de interpolare s-a realizat pe un subset de doi ani de date orare (2016-2017), suficient pentru a surprinde la nivel național particularitățile variabilității spațio-temporale a parametrului analizat (temperatura orară a aerului).

În Figura 1 este descrisă schematic metodologia aplicată în evaluarea metodelor de interpolare selectate analiză în această etapă.

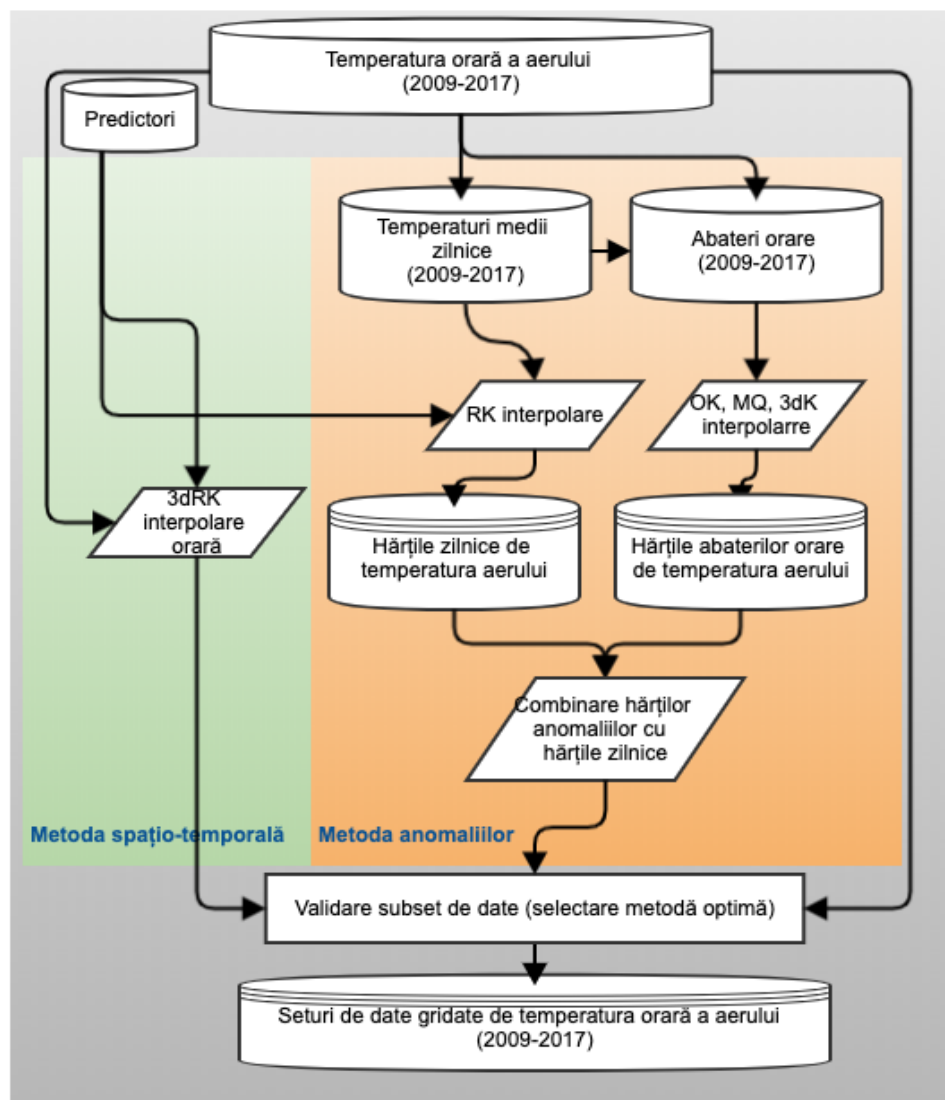


Figura 1: Fluxul de lucru utilizat în evaluarea metodelor de interpolare spațială.

2.1 Metoda anomaliilor

Această procedură presupune estimarea spațio-temporală a unui parametru selectat în funcție de valoarea mediată a acestuia pe subseturi de date (Brinckmann et al. 2016; Dumitrescu et al. 2016; Haylock et al. 2008). Deoarece relațiile statistice dintre variabilele climatice și predictorii luați în calcul sunt mai puțin intense la scară temporală detaliată, într-o primă etapă se obține distribuția spațială parametrului agregat, pe un interval de timp dat (an, lună sau zi), în funcție de predictorii selectați. În a doua etapă, se cuantifică spațial anomaliile față de situația medie, la scara temporală selectată pentru realizarea gridurilor. Prin combinare celor două hărți, se obține harta finală pentru un pas de timp dat.

Astfel, pentru obținerea estimărilor spațiale privind temperatura orară aerului, au fost parcurse următoarele etape:

1. interpolarea spațială la o rezoluție spațială de 1000 m × 1000 m a valorilor medii zilnice corespunzătoare fiecărei luni și an;
2. interpolarea spațială a abaterilor orare față de media zilnică, pentru fiecare oră din intervalul 2016-2017;
3. obținerea seturilor de date spațio-temporale prin combinarea celor două suprafețe obținute în etapele 1 și 2.

Pentru realizarea hărților zilnice, s-a utilizat metoda de interpolare spațială Regression Kriging (RK). Pentru alegerea metodei optime aplicată în spațializarea abaterilor, au fost testate, prin utilizarea unui set de date independent, trei metode de interpolare: multicuadratică (MQ), kriging normal (OK) și 3d Kriging (3DK).

Rezultatele însumării hărților zilnice cu hărțile anomaliilor, obținute prin cele trei metode, vor fi notate în etapa de evaluare a performanțelor metodelor cu următoarele acronime RKa-MQ, RKa-OK și RKa-3DK.

2.1.1 Hărțile zilnice de temperatura aerului

Pentru obținerea temperaturilor medii zilnice interpolate cu metoda RK, s-au utilizat și următoarele date auxiliare: date orare LST, obținute din procesarea imaginilor SEVIRI, și date derivate din MNA (altitudinea, distanța față de Marea Neagră și numărul pixelilor reprezentând suprafața urbană).

RK este o metodă multivariată, care poate lua în calcul una sau mai multe variabile cu o distribuție continuă în spațiu (modelul numeric altimetric, imagini satelitare etc.). Estimările finale rezultă din însumarea suprafeței determinată prin metoda celor mai mici pătrate (aplicată la regresia multiplă liniară - RML) cu suprafața obținută prin interpolarea spațială a reziduurilor regresiei, utilizând metoda Kriging. Primul pas al acestei metode constă în validarea statistică a modelului determinist, în sensul verificării intensității relațiilor dintre predictorii și variabila dependentă. Alegerea celui mai bun model de regresie se realizează prin procedura de regresie automată pas cu pas (stepwise regression). În cazul metodei RK, matricea punctelor de grilă a regresiei multiple reprezintă variabilitatea la scară mare a

parametrului analizat, în funcție de variabilele explicative, reziduurile interpolate constituind particularitățile locale modelate cu ajutorul semivariogramei (Hengl et al. 2007):

$$\hat{Z}(s_0) = \sum_{k=1}^p \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^N \lambda_i \cdot e(s_i)$$

unde $\hat{\beta}_k$ reprezintă coeficienții modelului de regresie, q_k valoarea predictorului în punctul localizat prin coordonatele s_0 pentru care se estimează o nouă valoare, iar λ_i coeficienții de ponderare ai reziduurilor regresiei $e(s_i)$ având coordonatele s_i . Coeficienții regresiei pot fi obținuți fie prin metoda simplă a celor mai mici pătrate, fie prin aplicarea modelului de regresie generalizat.

Metoda RK a fost aplicată cu succes în spațializarea datelor de temperatura aerului în regiunile cu topografie complexă, bazându-se pe variabile explicative derivate din MNA și din procesarea imaginilor satelitare (Brinckmann et al. 2016; Eiselt et al. 2017; Kilibarda et al. 2014; Perčec Tadić 2010; Ruiz-Álvarez et al. 2019).

2.1.2 Hărțile abaterilor orare de temperatura aerului

În această etapă de realizare a proiectului au fost plicate trei metode de interpolare: multicuadratică (MQ), kriging normal (OK) și 3d Kriging (3dK).

Metoda MQ face parte din clasa de metode exacte de interpolare Funcțiile Radiale de Bază (RBF), care este foarte asemănătoare cu cele din familia Kriging, doar că nu beneficiază de aportul analizei structurii spațiale a datelor prin intermediul semi-variogramei. Forma generală a acestei categorii de interpolatori este definită de (Johnston et al. 2001):

$$\hat{Z}(s_0) = \sum_{i=1}^N \omega_i \phi(\|s_i - s_0\|) + \omega_{n+1}$$

unde $\phi(r)$ este funcția de bază radială, $r = \|s_i - s_0\|$ distanța radială dintre punctul pentru care se calculează o nouă valoare s_0 și punctele cu valori măsurate s_i iar ω simbolizând ponderile care vor fi estimate.

Valoarea ponderii fiecărui punct folosit în interpolare rezultă în urma rezolvării unui sistem de ecuații utilizând calculul matricial de tipul:

$$\begin{pmatrix} \phi & 1' \\ 1 & 0 \end{pmatrix} \begin{pmatrix} w \\ \omega_{n+1} \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}$$

cu ϕ = matricea distanțelor dintre punctele cu valori cunoscute căreia i se aplică o funcție de bază radială, z = vectorul cu distanțele dintre locația aleasă pentru estimare și punctele cu măsurători, căruia i se aplică aceeași funcție radială, w = ponderile estimate iar ω_{n+1} = reziduurile.

Funcția radială MQ este dată de relația

$$\phi(r) = (r^2 + \sigma^2)^{-1/2}$$

Parametrul de netezire σ poate fi ales diferit, fie prin calcularea sumei minime a pătratelor erorilor rezultate în urma aplicării procedurii de validare încrucișată, fie prin calcularea unei valori în funcție de numărul de puncte utilizate și extinderea acestora în spațiu sau în mod subiectiv direct de către utilizator.

Metoda OK calculează ponderile pe baza funcțiilor care iau în calcul și configurația spațială a datelor (Isaaks and Srivastava 1989). Primul pas în interpolarea prin metoda OK constă în analiza interdependenței spațiale a setului de date, realizată prin construirea semivariogramei punctelor eșantionate. Prin construirea semivariogramei se poate extrage structura spațială a coordonatelor utilizate în model. Semivariograma este alcătuită din două componente: semivariogramă experimentală (empirică) și modelul semivariogramei. Semivariograma experimentală rezultă în urma calculării varianței fiecărui punct măsurat, față de celelalte puncte utilizate în spațializare. Ea se calculează cu ajutorul formulei (Pebesma 2004):

$$\hat{\lambda}(\bar{h}_j) = \frac{1}{2N_j} \sum_{i=1}^{N_j} (Z(s_i) - Z(s_i + h))^2$$

unde N_j un set de perechi de locații separate de distanța h și $\bar{h} =$ media distanțelor dintre perechile distincte N_j . Modelul teoretic aproximează semivariograma experimentală, acesta alegându-se în funcție de traiectoria liniei care se potrivește cel mai bine punctelor reprezentate pe graficul semi-variogramei.

Estimarea într-un nou amplasament se realizează prin însumarea valorilor ponderii punctelor multiplicat cu valoarea Z datelor folosite în interpolare (Johnston et al. 2001):

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i), \sum_{i=1}^N \lambda_i = 1$$

Funcțiile de ponderare OK iau în calcul atât distanța, cât și aranjarea geografică a datelor. Valoarea ponderii fiecărui punct folosit în interpolare rezultă în urma rezolvării unui sistem de ecuații prin calculul matricial de tipul:

$$\begin{pmatrix} C & 1' \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} c \\ 1 \end{pmatrix}$$

matricea C reprezentând covarianțele dintre punctele cu valori cunoscute, vectorul c fiind constituit din covarianțele dintre punctele cu valori cunoscute și punctul cu valoare necunoscută, $\lambda =$ vectorul coeficienților de ponderare Kriging, iar $m =$ multiplicatorul Lagrange utilizat în minimizarea erorilor prin relația:

$$\sigma^2 = \sum_{i=1}^N \lambda_i c + m$$

Metoda 3DK consideră timpul ca o a treia dimensiune ortogonală, reclassificată în metri pentru a se alinia la unitățile spațiale. O caracteristică importantă a acestei metode este că

predicțiile din cubul spațiu-timp se bazează doar pe un model de semi-variogramă pentru întreaga perioadă sau sub-perioade de analiză, în timp ce modelele clasice de interpolare Kriging necesită câte o semi-variogramă pentru unitatea de timp (Pebesma and Heuvelink 2016). Pentru obținerea hărților abaterilor orare față de media zilnică, s-au construit semi-variograme utilizându-se subseturi de date lunare.

2.2 Metoda spațio-temporală

Relațiile statistice dintre variabila analizată și predictorii selectați au fost cuantificate cu modelul RML pentru fiecare pas de timp (oră), iar reziduurile au fost interpolate printr-o metoda de interpolare care ține cont și de componenta timp (3dRK). Prin aplicarea metodei *stepwise regression* de tip retrograd, s-au selectat, pentru fiecare caz în parte (oră), predictorii relevanți din punct de vedere statistic.

Interpolarea reziduurilor regresiiilor s-a realizat utilizându-se semi-variogramele calibrate automat⁴, conform funcției *fit.variogram()* implementată în biblioteca R *gstat* (Pebesma 2004).

2.3 Metode de validare a rezultatelor interpolării

Pentru alegerea metodei optime de interpolare a datelor de temperatură, s-a aplicat procedura de eșantionare, setul de date inițial fiind împărțit în două subseturi: de calibrare a metodelor (70%) și de testare a acurateții metodelor (30%).

Cuantificarea diferențelor dintre estimări și datele observate s-a realizat cu ajutorul indicatorilor de măsurare a erorilor (Fox 1981):

- eroarea medie (ME) reprezintă mediile diferențelor dintre valorile estimate și cele măsurate

$$ME = \frac{1}{N} \sum_{i=1}^N (Ps_i - Os_i)$$

- eroarea medie absolută (MAE) reprezintă mediile diferențelor absolute dintre valorile estimate și cele măsurate

$$MAE = \frac{1}{N} \sum_{i=1}^N |Ps_i - Os_i|$$

- eroarea pătratică medie (RMSE) este senzitivă la prezența erorilor mari, procesul de ridicare la pătrat atribuind ponderi disproporționate reziduurilor

⁴ <https://www.r-spatial.org/r/2016/02/14/gstat-variogram-fitting.html>

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (P_{S_i} - O_{S_i})^2 \right)^{\frac{1}{2}}$$

Diagrama de tip box-plot furnizează informații privind tendința centrală și forma distribuției șirului de date, rezultate prin rezumarea celor 5 valori ce definesc o distribuție: valoarea minimă, prima quartilă, mediana, a treia quartilă și valoarea maximă (McGill et al. 1978). Mediana este marcat pe diagramă prin creștătura din lateralul dreptunghiurilor, cele două quartile delimitând dimensiunea dreptunghiului. Valorile aberante, cerculețele pline, sunt delimitate de pragul $3xD$ (D reprezentând distanța dintre prima și a treia quartilă), reprezentată pe diagramă prin lungimea liniei verticale.

Diagrama de tip Taylor a fost de asemenea utilizată analiza calitativă a rezultatelor metodelor de interpolare (Taylor 2001). Conform acestei reprezentări grafice, similaritatea dintre 2 seturi de date (observat vs. estimat) este cuantificată prin calcularea coeficientului de corelație Pearson, erorii pătratice medii (RMSE) și a deviației standard.

2.4 Rezultate

2.4.1 Analiza exploratorie

Într-o primă etapă, s-a verificat în ce măsură predictorii selectați explică variabilitatea spațio-temporală a temperaturii aerului. Au fost luate în calcul date geo-spațiale derivate din MNA sau din informații obținute din prelucrarea imaginilor satelitare, utilizate în mod frecvent ca variabile auxiliare în metodele de interpolare spațială. Astfel, au fost testate corelațiile dintre temperatura orară măsurată la stațiile meteorologice (Taer), pe de-o parte, și datele SEVIRI LST, altitudine, distanța față de Marea Neagră și numărul pixelilor încadrați ca suprafață urbană pe o rază de 3 km, pe de altă parte. Asocierea și relațiile statistice dintre variabilele explicative și temperatura aerului au fost cuantificate calitativ prin utilizarea diagramei de împrăștiere a valorilor (scatterplot), dar și cantitativ, prin calcularea unor indicatori statistici (coeficientul de corelație Pearson și panta liniei de regresie).

Analiza distribuției comune de împrăștiere dintre Taer și SEVIRI LST, pe sezoane și trepte de altitudine, relevă o asociere statistică intensă a celor două variabile (Figura 2). Coeficienții de corelație (Corr) obținuți pe subseturi de date sunt mai mari de 0.720, valorile maxime fiind calculate din datele corespunzătoare sezoanelor intermediare (primăvara și toamna), indiferent de treapta de altitudine analizată. Valorile pantelor dreptelor de regresie și concentrarea punctelor în jurul dreptei de pante de regresie confirmă dependența strânsă dintre variabilele analizate, cele mai multe puncte fiind concentrate de-a lungul liniei de regresie.

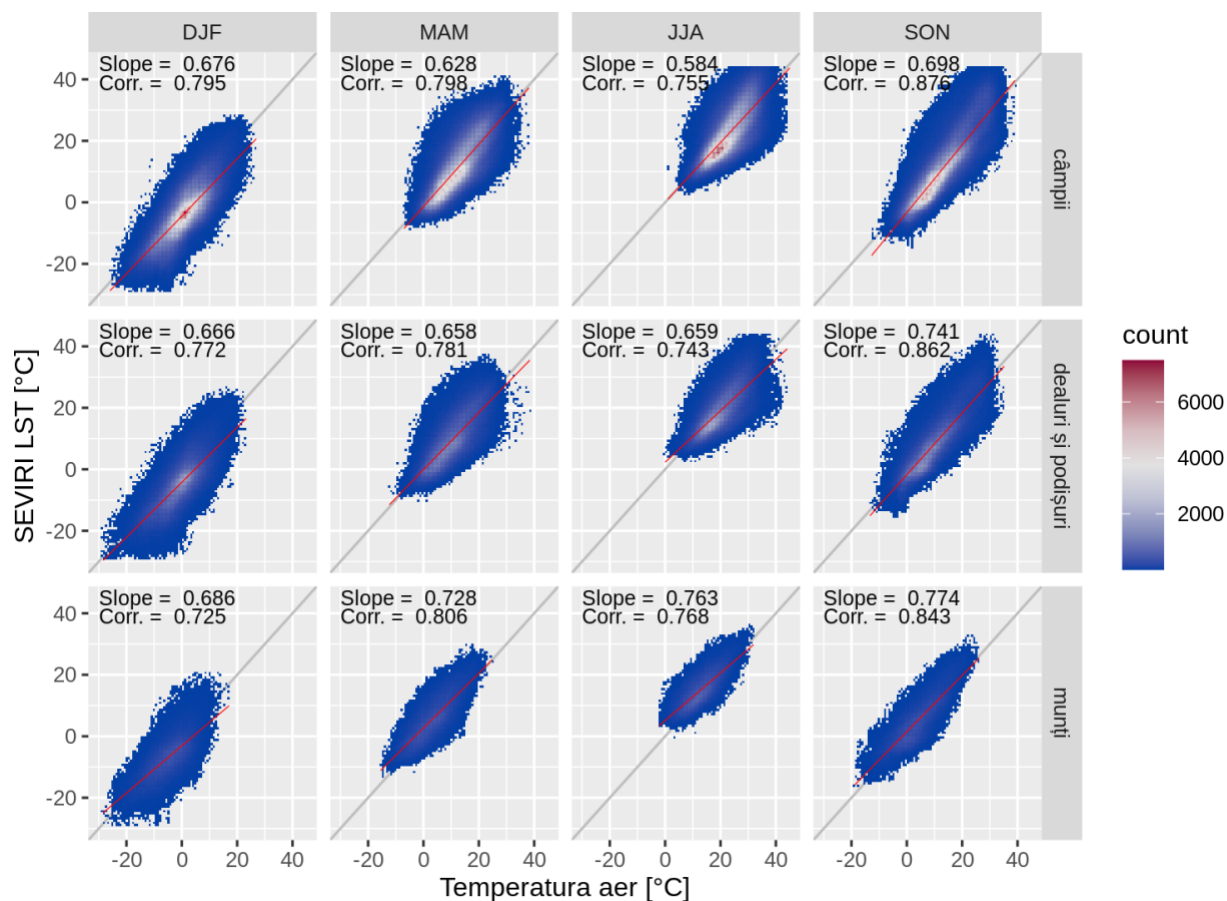


Figura 2: Temperatura aerului vs. temperatura de la nivelul suprafeței terestre active pentru teritoriul României extrasă din datele complete SEVIRI LST.

Calculații pentru fiecare anotimp, coeficienții de corelație Pearson (Corr) dintre Taer și variabilele MNA evidențiază o dependență strânsă între variabila analizată și altitudine, cu o valoare maximă calculată din datele corespunzătoare sezonului cald (Tabelul 1). Distanța față de mare influențează semnificativ distribuția spațială a Taer vara și toamna, iar suprafețele urbane vara.

Tabel 1: Coeficienții de corelație Pearson calculați între datele Taer orare și variabilele explicative derivate din MNA (2016 - 2017).

Sezon	Altitudine	Distanța față de Marea Neagră	Număr pixeli suprafață urbană
DJF	-0.260	-0.086	0.066
MAM	-0.367	-0.064	0.098
JJA	-0.440	-0.152	0.113
SON	-0.300	-0.119	0.069

Deoarece metoda de interpolare spațială bazată pe calculul anomaliilor ține cont de distribuția spațială a temperaturii medii zilnice, au fost testate și corelațiile dintre datele medii zilnice de Taer, SEVIRI LST și predictorii MNA (Tabelul 2). La scară temporală zilnică,

se observă o dependență mai strânsă între Taer și variabilele auxiliare (LST și MNA), mai ales primăvara și vara pentru predictorii derivați din MNA.

Tabel 2: Coeficienți de corelație Pearson calculați între datele Taer zilnice și variabilele explicative derivate din MNA (2016 - 2017).

Sezon	SEVIRI LST	Altitudine	Distanța față de Marea Neagră	Număr pixeli suprafață urbană
DJF	0.880	-0.284	-0.094	0.072
MAM	0.918	-0.451	-0.078	0.120
JJA	0.866	-0.626	-0.216	0.161
SON	0.944	-0.338	-0.134	0.077

Prin utilizarea modelelor de regresie multiplă liniară, s-a analizat pentru fiecare pas de timp (zilnic și orar) în ce măsură toți predictorii explică variabilitatea spațială a temperaturii aerului. Utilizându-se procedura de regresie automată pas cu pas (stepwise regression) (Venables and Ripley 2013), pentru fiecare caz în parte a fost aleasă combinația optimă de predictorii, care poate fi utilizată în estimarea spațială a temperaturii.

Rezultatele modelelor de regresie orare și zilnice sunt sintetizate în diagrama de tip box-plot (Figura 3). Indiferent de luna analizată, se constată că estimările modelelor de regresie bazate pe date medii zilnice sunt superioare celor obținute din date orare, cu valori medii ale coeficientului de determinare ajustat R^2 întotdeauna mai mari. Astfel se justifică analiza unei metode hibride de interpolare spațială (abaterile orare față de media zilnică), bazate pe hărțile cu distribuția spațială a temperaturilor zilnice. Cele mai mici valori ale R^2 au fost obținute în sezonul rece, când gradul de acoperire cu nori este maxim și frecvent se produce fenomenul de inversiune termică, care modifică gradientul termic vertical, dar și determină unele limitări în estimarea temperaturii suprafeței active prin instrumente de teledetecție.

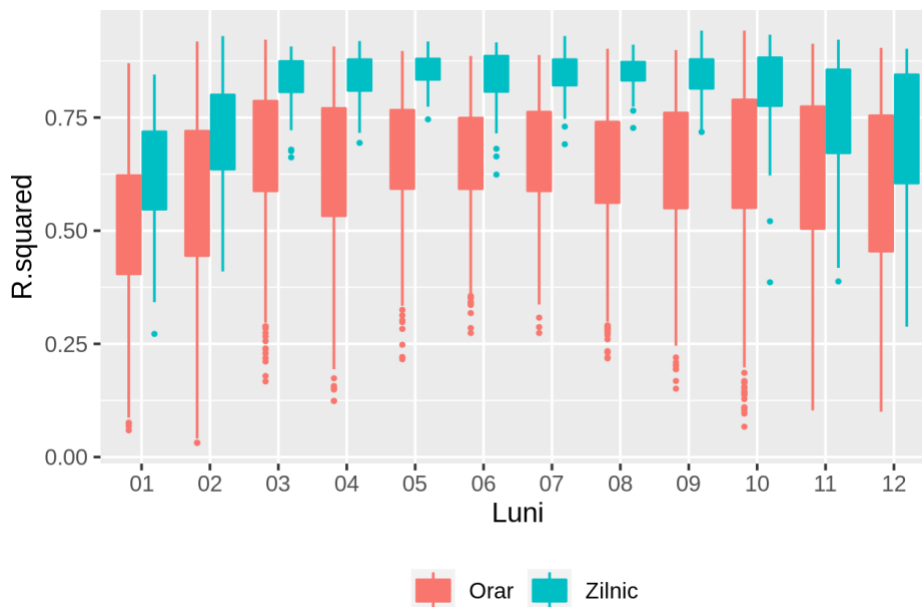


Figura 3: Diagrama de tip box-plot a coeficienților de determinare ajustați R^2 , calculați din subseturi de date orare și zilnice.

2.4.2 Analiza estimărilor metodelor de interpolare

În această etapă, au fost evaluate estimările metodelor de interpolare prezentate în secțiunea anterioară (RKa-MQ, RKa-OK, RKa-3DK și 3dRK). Analiza estimărilor celor patru metode de spațializare s-a realizat prin utilizarea unui set de date care nu a fost luat în calcul în realizarea interpolărilor. Din setul de date inițial, 70% au fost selectate aleator pentru calibrarea metodelor selectate, iar 30% pentru evaluarea acurateții estimărilor (validare).

Prin calculul indicatorilor de măsurare a erorilor de estimare, se remarcă o diferențiere între metodele bazate pe mediile zilnice și metoda 3dRK, aceasta din urmă obținând erorile de estimare cele mai mari, indiferent de sezonul analizat (Tabelul 3). Dintre metodele bazate pe abaterile față de media zilnică, conform indicatorilor calculați pentru fiecare sezon, cele mai bune rezultate sunt obținute de metoda RKa-OK, mai ales toamna și în sezonul rece. Metoda RKa-MQ obține erori de estimare mai mici ca RKa-3DK, cu valori CORR, MAE și RMSE mai bune în toate cazurile analizate.

Tabel 3: Valorile coeficienților de corelație Pearson (CORR), erorilor medii absolute (MAE) și erorilor pătratic medii (RMSE) obținute dintre datele măsurate și datele estimate prin patru metode de interpolare. Validarea metodelor s-a realizat pe un subset de date care nu a fost utilizat ca date de intrare în metodele de interpolare spațială

Metoda	CORR				MAE				RMSE			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
3DRK	0.91	0.936	0.911	0.95	1.984	1.723	1.846	1.734	2.623	2.292	2.465	2.319
RKa-3DK	0.95	0.936	0.899	0.954	1.409	1.658	1.924	1.554	1.959	2.304	2.609	2.218
RKa-OK	0.957	0.953	0.927	0.965	1.307	1.424	1.645	1.367	1.829	1.964	2.226	1.936
RKa-MQ	0.954	0.943	0.91	0.959	1.344	1.519	1.774	1.448	1.893	2.166	2.472	2.118

Analiza cu ajutorul diagramei de tip Taylor a estimărilor celor patru metode relevă unele diferențieri în funcție de luna analizată, diferențele maxime între cele două categorii de metode evidențiindu-se mai ales iarna, când valorile coeficienților *Pearson* 3DRK sunt mai mici de 0.9, în comparație cu celelalte metode când se obțin valori între 0.92 și 0.94 (Figura 4). Metodele bazate pe calculul abaterilor obțin aproximativ aceleași rezultate, indiferent de indicatorul sau luna analizată, remarcându-se totuși aproximări mai bune ale celor trei indicatori obținute de metoda RKa-OK.

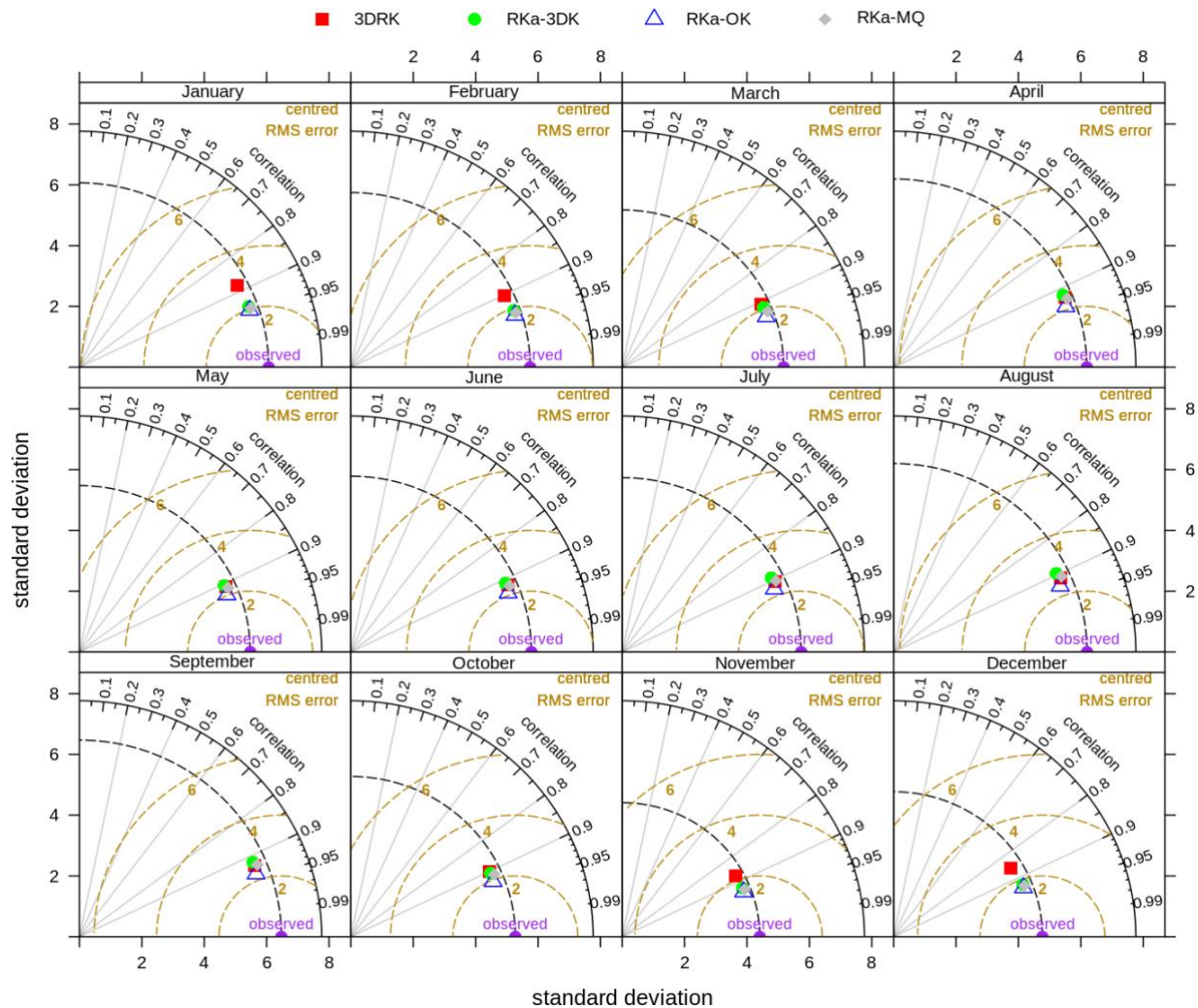


Figura 4: Diagrama de tip Taylor realizată din datele măsurate și datele estimate prin patru metode de interpolare.

Cu ajutorul diagramei de tip box-plot, s-a verificat pentru fiecare sezon distribuția statistică a diferențelor dintre valorile estimate de cele patru metode și valorile măsurate la stațiile meteorologice (Figura 5). Se constată că majoritatea erorilor de estimare se încadrează între prima și a treia cuartilă, cu observația că valorile medianelor sunt mai apropiate de zero în cazul metodelor bazate pe anomalii zilnice. Exceptând iarna și toamna, când metoda 3DRK conduce la valori mai mici ale diferențelor maxime dintre valorile observate și estimate (cerculețe umplute), interpolările bazate pe anomalii se apropie mai mult de măsurători, indiferent de indicatorul statistic reprezentat grafic pe diagramă (mediană, quartile, $3 \times$ distanța intra-cuartile). Diferențierea rezultatelor obținute de RKa-MQ, RKa-OK și RKa-3DK este greu de realizat prin această metodă, remarcându-se valori aproximativ egale atât ale valorilor maxime, cât și ale medianei și cuartilelor.

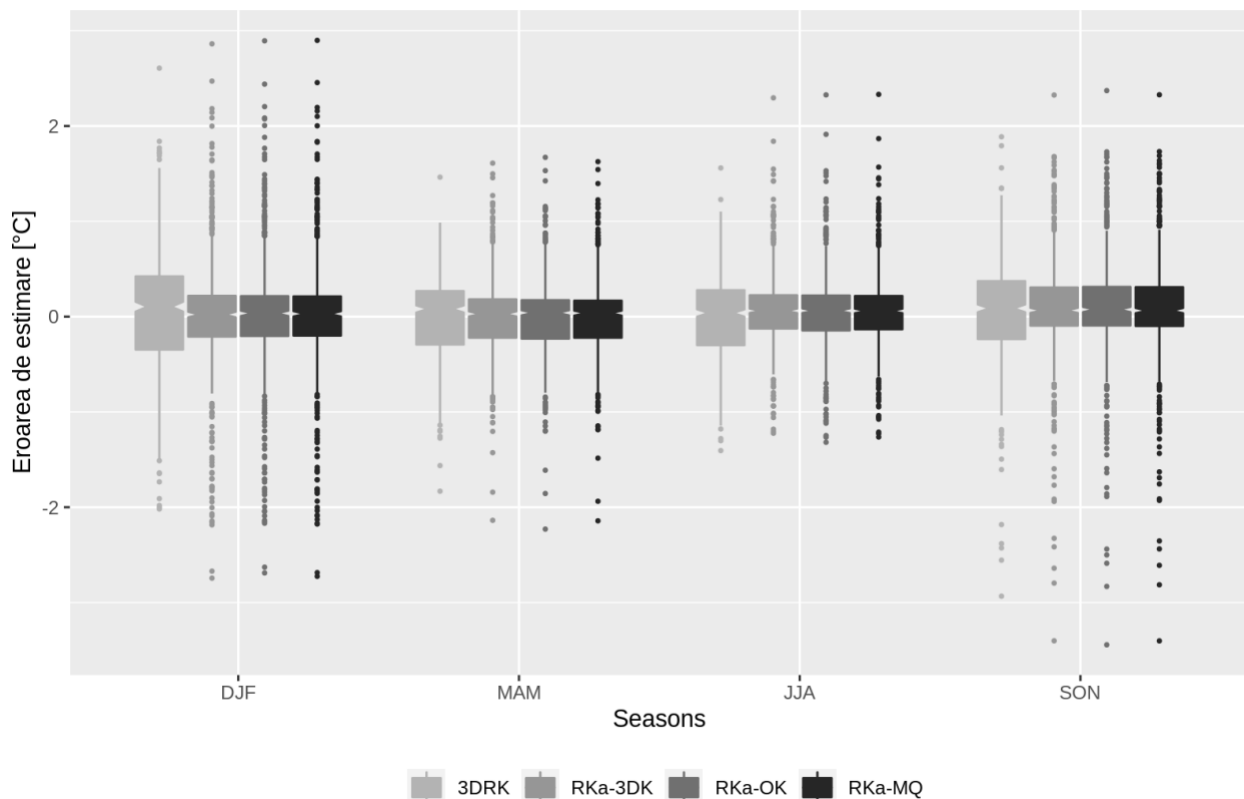


Figura 5: Diagrama de tip boxplot a diferențelor dintre valorile estimate prin patru metode de interpolare și măsurate la stațiile meteorologice.

Deoarece s-a constatat că indiferent de indicatorul de evaluare a estimărilor analizat (CORR, RMSE, MAE, diagrama Taylor), erorile de cele mai mici sunt obținute prin metoda RKa-OK, s-a decis utilizarea acesteia în realizarea seturilor de date gridate de temperatură orară a aerului.

3 Estimarea spațio-temporală la scară fină a datelor orare de temperatura aerului

Seturile de date gridate au fost interpolate la o rezoluție spațială de 1000 m × 1000 m, în proiecție cartografică Stereografică 1970⁵. Pentru facilitarea utilizării datelor gridate alături de alte seturi de date geo-spațiale, care au o extindere mai mare decât teritoriul României, s-a decis arhivarea acestora în format netCDF, în proiecție Geografică WGS84⁶, la o rezoluție spațială de 0.01° latitudine × 0.01° longitudine (aproximativ 1000 m).

⁵ <https://epsg.io/3844>

⁶ <https://epsg.io/4326>

Datele orare de tip raster au fost grupate în fișiere zilnice multidimensionale de tip netCDF⁷, care conțin informația interpolată de temperatura aerului pentru fiecare pas de timp (24 de ore). NetCDF (Network Common Data Form) este un format de fișier utilizat de comunitatea științifică pentru stocarea datelor geo-spațiale multidimensionale, cum ar fi variabilele meteorologice (temperatura, umezeala, presiunea, viteza vântului și direcția etc.) Fiecare dintre aceste variabile poate fi afișată printr-o dimensiune (timpul sau strat vertical). În cadrul fișierelor netCDF dimensiunea spațială este definită doar o singură dată și este valabilă pentru toate datele corespunzătoare fiecărui pas de timp, prin urmare și spațiul ocupat de aceste tipuri de fișiere pe mediile de stocare este mai redus în comparație cu fișierele raster unidimensionale (GeoTIFF, Esri Grid etc.).

Interfețe pentru accesarea fișierelor de tip netCDF sunt implementate atât în pachetele principalelor limbaje de programare (R, Python, Matlab, IDL, Octave), dar și în cele mai cunoscute aplicații SIG (Quantum GIS, ArcGIS, Saga GIS, Panoply etc.). Panoply⁸ este una dintre cele mai ușor de utilizat aplicații pentru vizualizarea fișierelor de tip netCDF, având implementate atât date geo-spațiale de tip vector (granițele țărilor și continentelor, limitele oceanelor etc), cât și palete de culori reprezentative pentru parametrii meteorologici. Un exemplu de vizualizare cu Panoply a datelor obținute în această etapă este prezentat în Figura 6, pentru care s-a utilizat o paletă de culori sugestivă pentru parametrul vizualizat, dar și alte informații geo-spațiale disponibile în aplicație (granițele țărilor).

În *Anexa 2* sunt prezentate câteva posibilități de descărcare și postprocesare a datelor netCDF orare utilizând limbajul R.

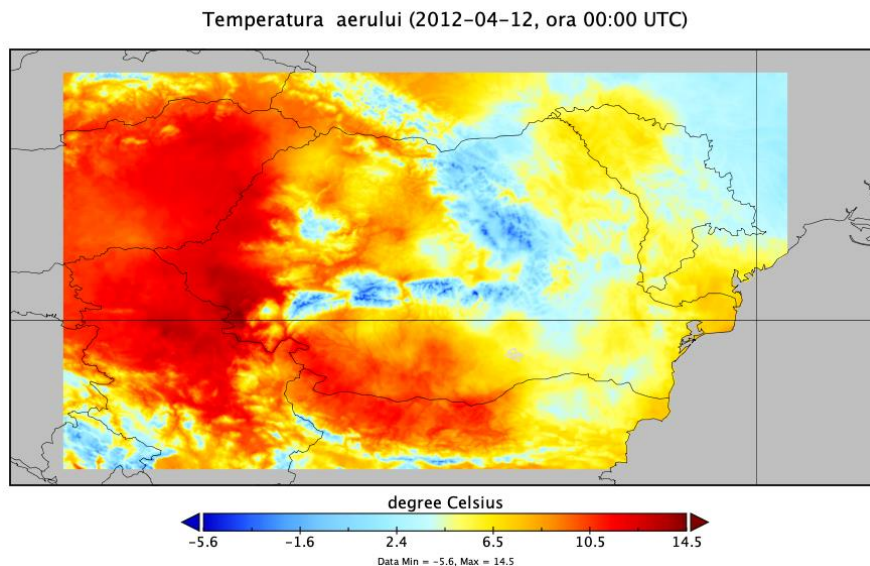


Figura 6: Distribuția spațială a temperaturii orare a aerului (2012-04-12, ora 00:00 UTC).

⁷ https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html

⁸ <https://www.giss.nasa.gov/tools/panoply/>

4 Validarea setului de date gridate de temperatura orară a aerului

În această secțiune, seturile de date gridate orare au fost comparate cu alte seturi de date din aceeași categorie, realizate la nivel național și continental, a căror extindere spațială acoperă în totalitate zona de studiu: ROCADA (Dumitrescu and Birsan 2015), E-OBS (Cornes et al. 2018) și MiKlip (Brinckmann et al. 2016). Caracteristicile celor trei seturi de date sunt prezentate în Tabelul 4. Rezoluția spațială a datelor selectate este de aproximativ 10 km pentru ROCADA și E-OBS, și de aproximativ 5 km pentru MiKlip. E-OBS și MiKlip sunt realizate la scară continentală, iar ROCADA la scară națională. Gridurile realizate la scară continentală au utilizat date de la stațiile din România care fac parte din schimbul internațional de date meteorologice (27 stații), iar setul de date ROCADA s-a bazat pe toate măsurătorile de la stațiile care au cel mult de 30 % date lipsă în intervalul 1961 - 2010 (150 stații).

Tabel 4: Seturi de date gridate selectate pentru analiza comparativă.

	Rezoluția spațială	Rezoluția temporală	Perioada de timp
ROCADA	0.1° × 0.1°	zilnic	1961 - 2013
E-OBS	0.1° × 0.1°	zilnic	1950 - prezent
MiKlip	0.044° × 0.044°	zilnic	2001 -2010

Deoarece nu au fost identificate măsurători de temperatura aerului de la stații care să nu fie luate în calcul în obținerea celor patru seturi de date gridate, s-a realizat doar o evaluare calitativă a rezultatelor. De asemenea, având în vedere rezoluția temporală zilnică a celor trei seturi de date selectate, datele gridate orare (Taer) au fost mediate la scară temporală zilnică pentru analiza comparativă. Astfel, au fost realizate reprezentări cartografice ale temperaturilor medii pentru patru zile: 15 ianuarie 2010, 15 aprilie 2010, 15 iulie 2010 și 15 octombrie 2010.

Din analiza figurilor 7, 8, 9 și 10 se observă că rezoluția spațială reprezintă principalul factor limitativ în estimarea spațială a valorilor de temperatură. Acest aspect este evident mai ales în zonele caracterizate printr-o fragmentare mare a reliefului, unde un pixel cu rezoluția de aproximativ 10 km se extinde atât peste zone culmile munților, cât și peste văile din apropiere. Prin urmare, estimările spațiale la rezoluția spațială de 0.1° din zona montană înaltă, extrase din ROCADA și E-OBS, sunt reprezentative pentru două zone topografice cu caracteristici diferite, ceea ce nu corespunde realității. Această limitare este mai puțin evidentă la datele MiKlip, a căror rezoluție spațială este de aproximativ 5 km. Avantajul rezoluției spațiale detaliate este foarte bine relevat de setul de date Taer, distribuția spațială a temperaturii aerului ținând cont de fragmentarea accentuată a reliefului din zonele montane.

De asemenea, numărul stațiilor meteo de la care s-au utilizat date pentru realizarea estimărilor spațiale este determinant. În figurile 7 și 10 se observă că în hărțile corespunzătoare MiKlip și E-OBS, care se bazează pe date de la un număr mic de stații meteo, nu se regăsesc valorile de temperatură scăzute din zonele montane înalte (ex. Munții Țarcu și Munții Făgăraș), în comparație cu celelalte două seturi de date, în realizarea cărora au fost

integrate date de la toate stațiile disponibile în Munții Carpați. De exemplu, în luna ianuarie s-au înregistrat valori mai mici de -12 °C, care sunt reprezentate pe hărțile Taer și ROCADA, dar care nu apar în celelalte două seturi de date (MiKlip și E-OBS), unde valorile minime de temperatură sunt de până la -10 °C.

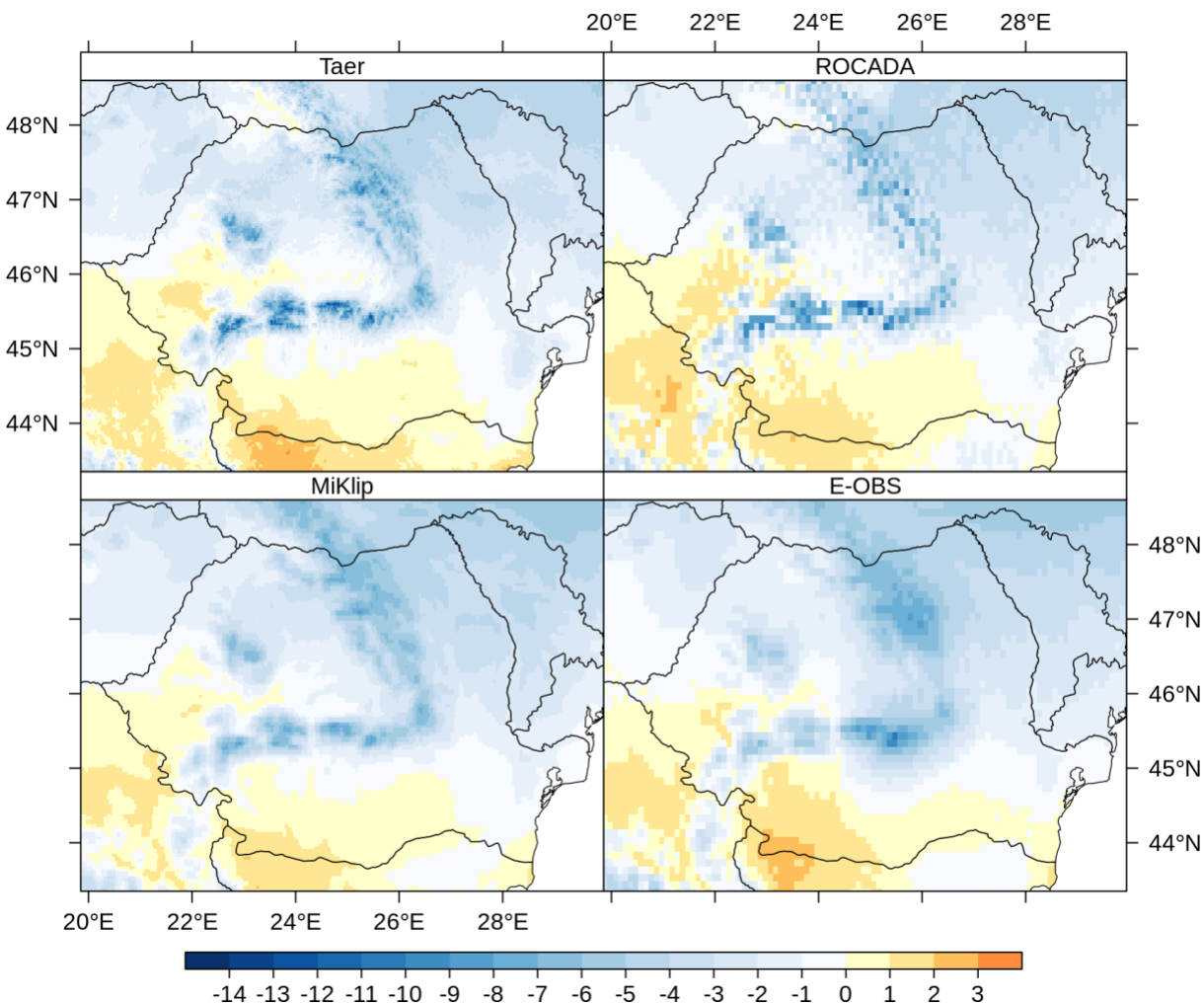


Figura 7: Distribuția spațială a temperaturii aerului extrasă din patru seturi de date gridate (15 ianuarie 2010): Taer (obținut din medierea datele orare), ROCADA, MiKlip și E-OBS.

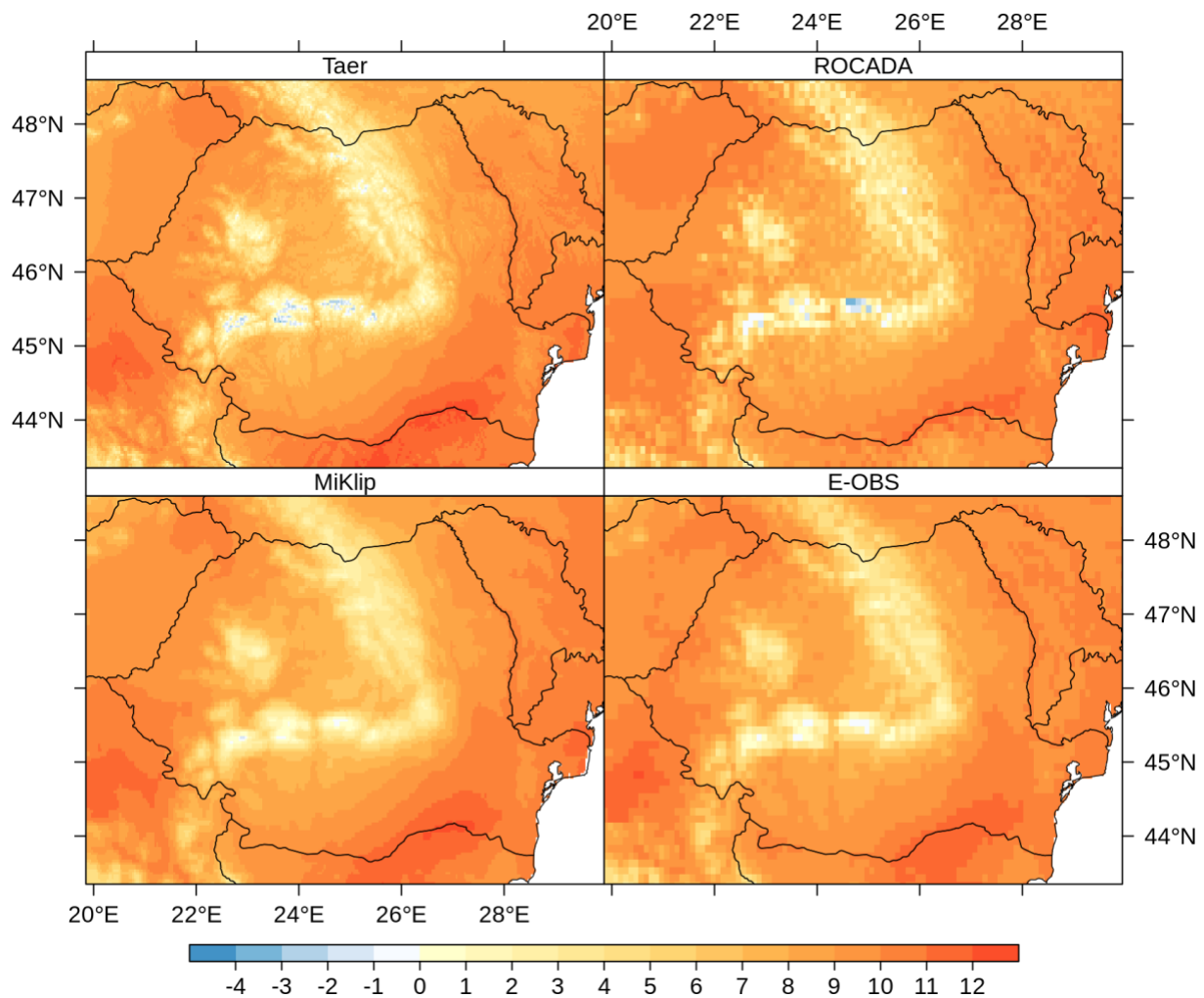


Figura 8: Distribuția spațială a temperaturii aerului extrasă din patru seturi de date gridate (15 aprilie 2010): Taer (obținut din medierea datele orare), ROCADA, MiKlip și E-OBS.

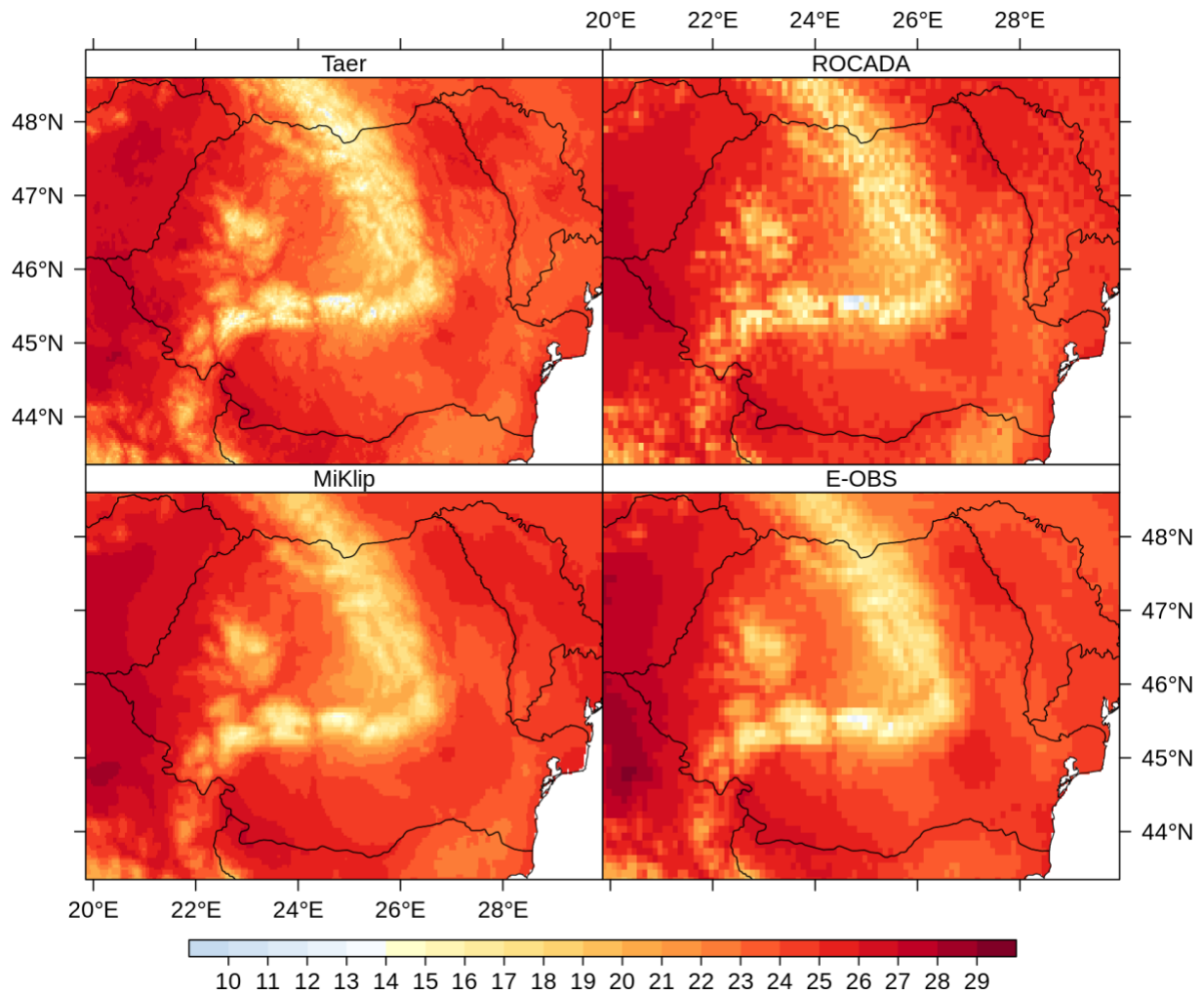


Figura 9: Distribuția spațială a temperaturii aerului extrasă din patru seturi de date gridate: (15 iulie 2010), Taer (obținut din medierea datele orare), ROCADA, MiKlip și E-OBS.

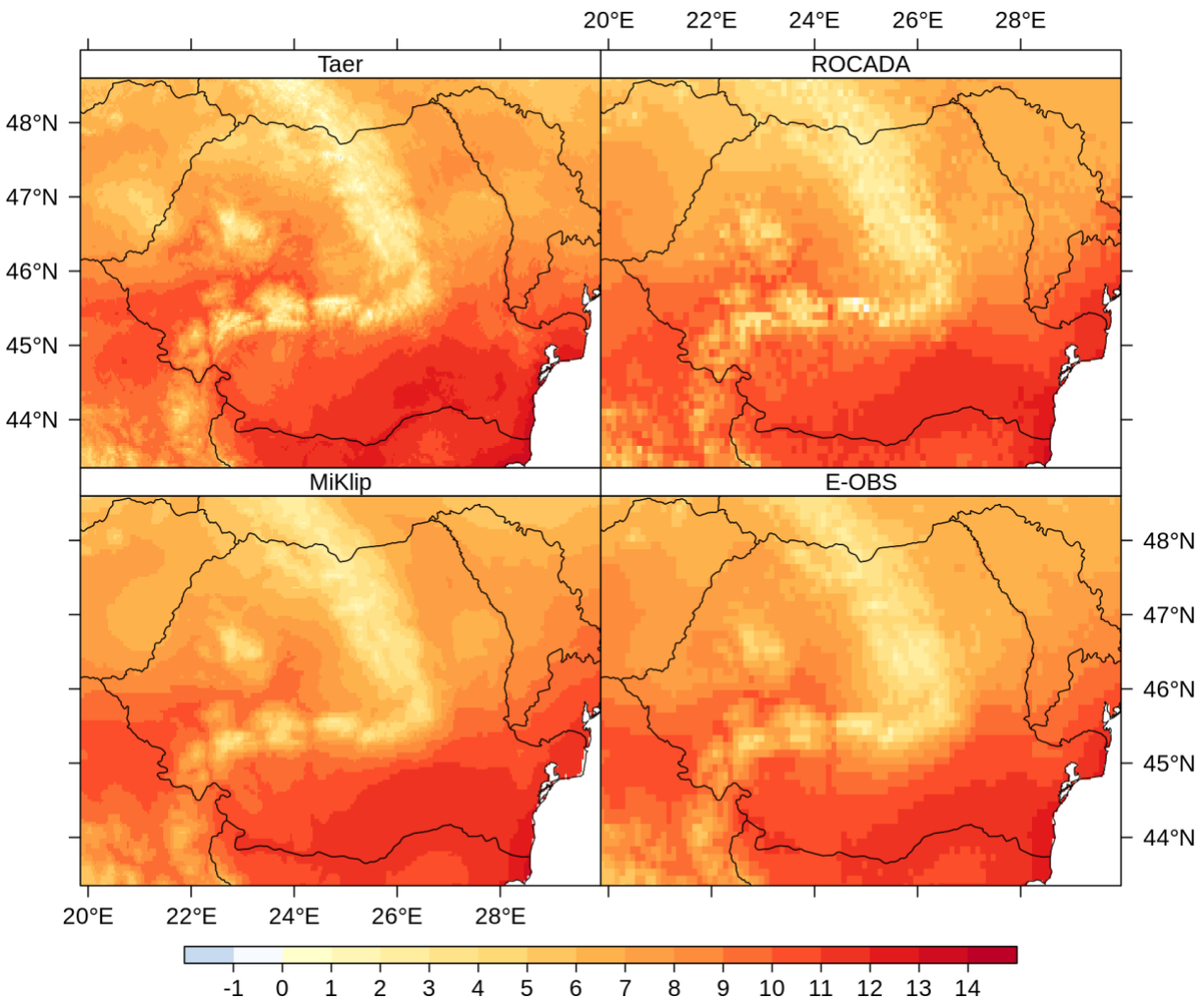


Figura 10: Distribuția spațială a temperaturii aerului extrasă din patru seturi de date gridate (15 octombrie 2010): Taer (obținut din medierea datele orare), ROCADA, MiKlip și E-OBS.

Cele mai sugestive reprezentări cartografice, indiferent de zona geografică analizată, sunt obținute din datele Taer, obținute la o rezoluție spațială suficient de detaliată și care se bazează pe date obținute de la un număr suficient de mare de stații meteorologice, localizate în toate arealele reprezentative ale zonei de studiu. Gradul mare de detaliere a distribuției spațiale a temperaturii aerului se remarcă nu numai în zona arcului carpatic, dar și în zonele joase de câmpie. Calitatea superioară a acestui set de date este determinată și de utilizarea ca informație auxiliare a temperaturii suprafeței active, extrasă din imaginile satelitare SEVIRI.

5 Temperaturi extreme caracteristice și indici termici extremi

În această secțiune, au fost calculate din datele gridate orare (Taer) temperaturi extreme caracteristice (maxime, minime, amplitudini, zile cu îngheț, nopți geroase, zile de iarnă, zile de vară, zile tropicale, nopți tropicale) și indicatori termici (valuri de căldură, valuri de frig) din seturile de date orare gridate. În mod curent, parametrii menționați sunt calculați din

PN-III-P1-1.1-PD-2016-1579 Interpolarea spațio-temporală la scară fină a temperaturii aerului 19 pentru teritoriul României

valori instantanee, măsurate cu termometrele de maximă și de minimă, care se produc la un moment dat și între orele de observații (Sandu et al. 2008). În cazul de față, valorile maxime și minime au fost obținute pentru fiecare zi din 24 de fișiere de tip raster, corespunzătoare măsurătorilor orare, efectuate cu termometrul ordinar la stațiile meteorologice.

5.1 Temperaturi extreme caracteristice

5.1.1 Mediile lunare ale temperaturilor extreme zilnice

Pe baza valorilor zilnice extreme s-au obținut temperaturile medii maxime și medii lunare. În această secțiune au fost analizate mediile multianuale ale temperaturilor minime corespunzătoare lunii ianuarie și mediile multianuale temperaturilor maxime din luna iunie. Ecartul (amplitudinea) a fost calculat ca diferența dintre cele două medii multianuale (maxima lunii iulie și minima lunii ianuarie).

Repartiția teritorială a temperaturilor minime zilnice în luna ianuarie este influențată în principal de stratul de zăpadă care persistă în sezonul rece pe culmile Munților Carpați, unde în medie s-au produs valori mai mici de $-10.0\text{ }^{\circ}\text{C}$ (Figura 11). Valori mici ale temperaturilor minime (sub $-8\text{ }^{\circ}\text{C}$) s-au produs și în depresiunile intracarpatiche, ca o consecință a frecvențelor inversiuni termice generate de răcirile advecive și radiative. Cele mai ridicate medii ale temperaturilor minime, cuprinse între $-1.0\text{ }^{\circ}\text{C}$ și $0.0\text{ }^{\circ}\text{C}$, se produc în zona litoralului, ca efect al căldurii cedate de apa din bazinul Mării Negre.

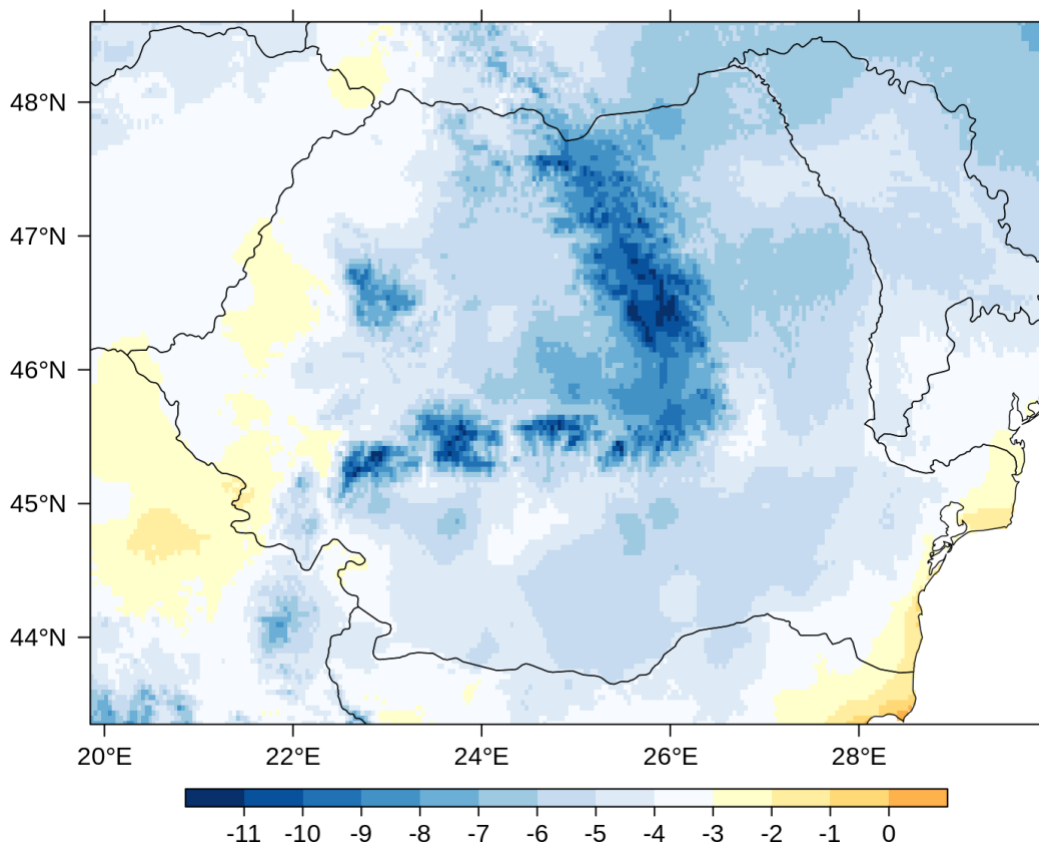


Figura 11: Mediile lunare ale temperaturii minime zilnice a aerului în ianuarie (2009 - 2017).

În iulie (Figura 12), cele mai mari valori ale temperaturilor medii maxime, peste 31 °C, se observă în extremitatea sudică a Câmpiei Române, în câteva areale de-a lungul Dunării. Pe cea mai mare parte a Câmpiei Române valorile sunt cuprinse între 30.0 °C și 31.0 °C. Valorile cele mai mici sunt înregistrate în zona montană, unde vârfurile mai înalte din Munții Carpați sunt înconjurare de izoterma de 12.0 °C.

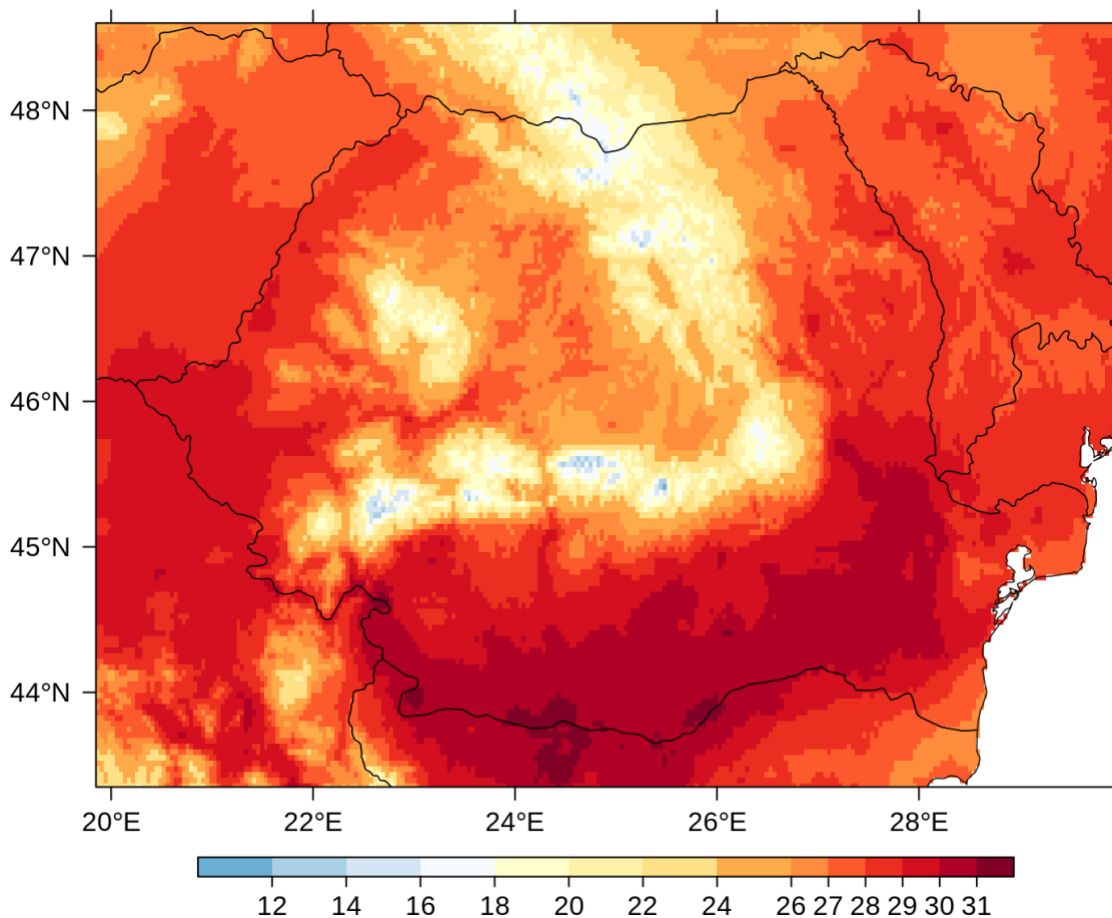


Figura 12: Mediile lunare ale temperaturii maxime zilnice a aerului în iulie (2009 - 2017).

Ecartul (amplitudinea) mediei temperaturilor extreme zilnice exprimă contrastul de temperatură dintre iarnă și vară și indică gradul de continentalism al climei. Conform Figurii 13, cele mai mari diferențe ale mediei temperaturilor extreme zilnice s-au semnalat în partea central-sudică a Câmpiei Române și în sudul Podișului Moldovei, unde depășesc 34.0 °C. Pe cea mai mare parte a teritoriului țării amplitudinea înregistrează valori mari, cuprinse între 30.0 °C și 36.0 °C. Rolul de moderator climatic al Mării Negre determină valori reduse ale ecartului (mai mici de 30.0 °C), dar la numai 60 km distanță de Marea Neagră amplitudinea crește până la 35.0 °C.

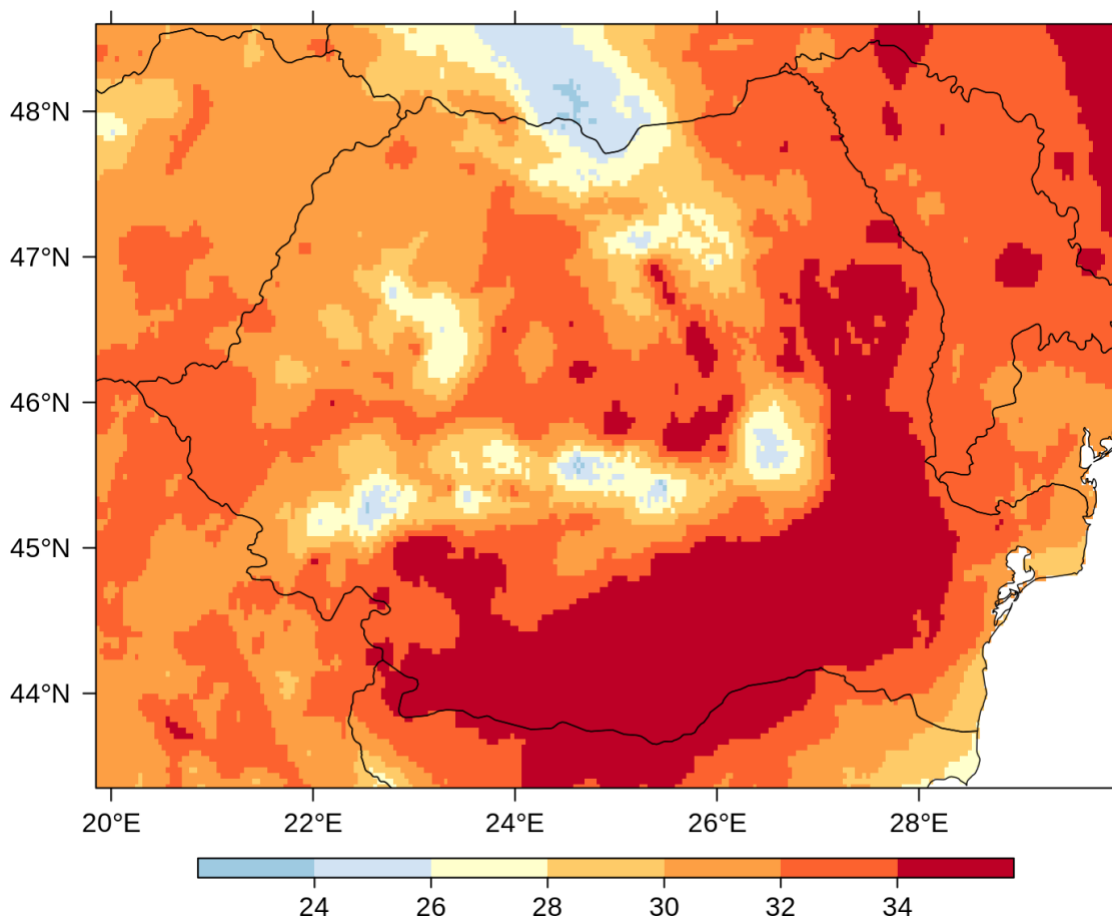


Figura 13: Ecartul (amplitudinea) temperaturilor extreme zilnice (2009 - 2017).

5.1.2 Numărul de zile cu diferite temperaturi caracteristice

Pe baza datelor gridate de temperatură minimă și maximă zilnică a aerului, a fost calculat numărul mediu anual de zile cu diferite temperaturi caracteristice. În funcție de anumite praguri termice, peste care temperatura aerului a urcat sau a coborât, s-au stabilit o serie de categorii convenționale de zile sau nopți (Sandu et al. 2008): zile cu îngheț (temperatura minimă a aerului ≤ 0 °C), nopți geroase (temperatura minimă a aerului ≤ -10 °C), zile de iarnă (temperatura maximă a aerului ≤ 0 °C), zile de vară (temperatura maximă a aerului ≥ 25 °C), zile tropicale (temperatura maximă a aerului ≥ 30 °C), nopți tropicale (temperatura minimă a aerului ≥ 20 °C). În figurile 14, 15, 16, 13, 17, 18 și 19 sunt reprezentate distribuțiile teritoriale a mediilor multianuale ale parametrilor menționați mai sus. Se remarcă rolul decisiv al caracteristicilor reliefului în distribuția teritorială a numărului mediu multianual de zile cu temperaturi caracteristice. Vizibile sunt și influențele suprafețelor întinse de apă, cum sunt marile râuri și Marea Neagră. De asemenea, se constată și unele diferențieri între zonele situate în estul și vestul țării, aflate sub influențe climatice diferite: continentalism în est, oceanic și mediteranean în vest.

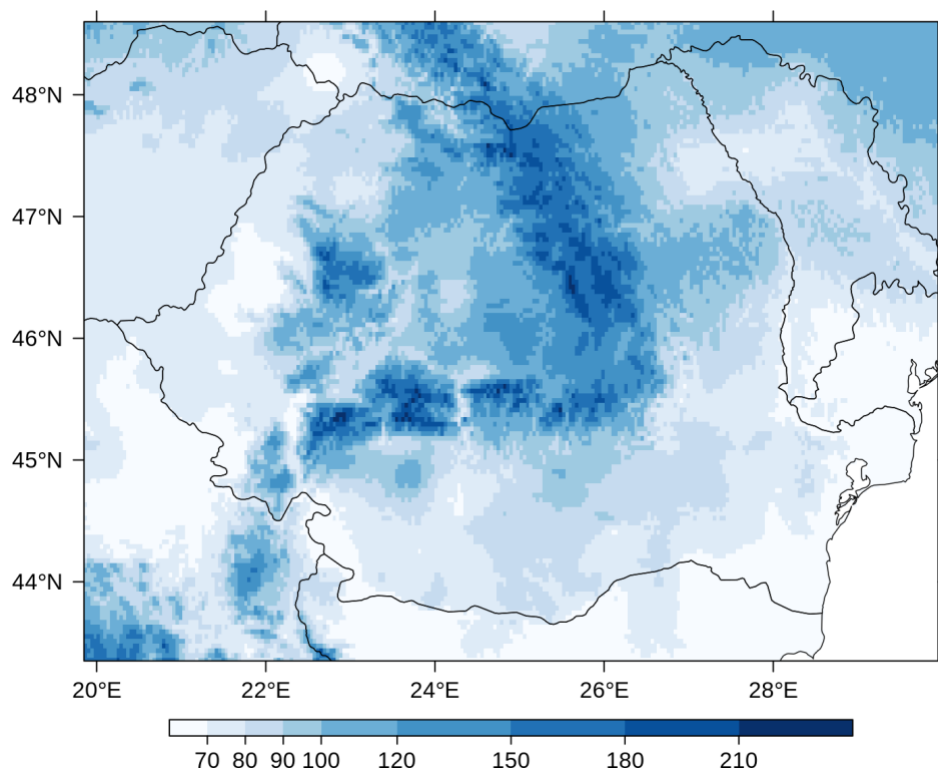


Figura 14: Numărul mediu anual de zile cu îngheț (temperatura minimă a aerului $\leq 0\text{ }^{\circ}\text{C}$) (2009 - 2017).

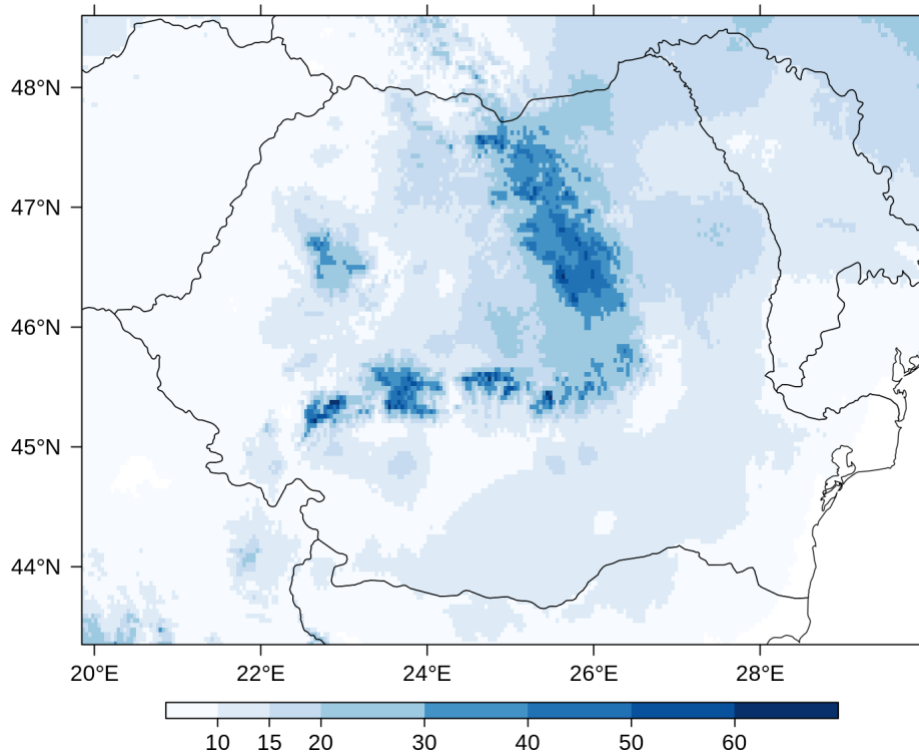


Figura 15: Numărul mediu anual de nopți geroase (temperatura minimă a aerului $\leq -10\text{ }^{\circ}\text{C}$) (2009 - 2017).

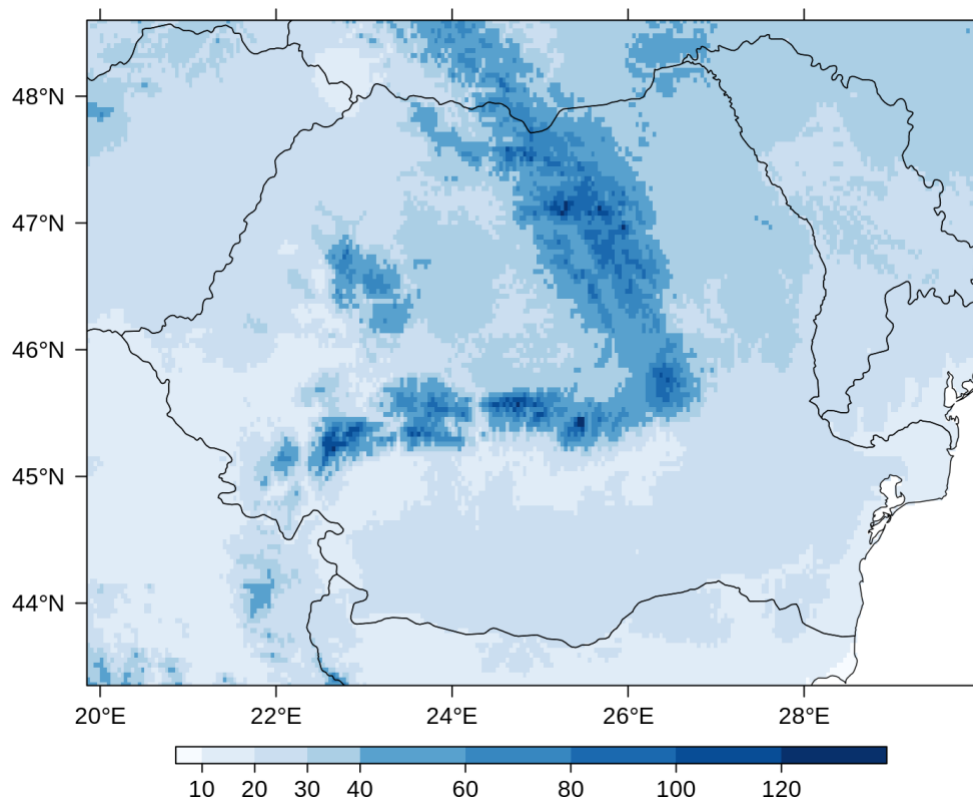


Figura 16: Numărul mediu anual de zile de iarnă (temperatura maximă a aerului ≤ 0 °C) (2009 - 2017).

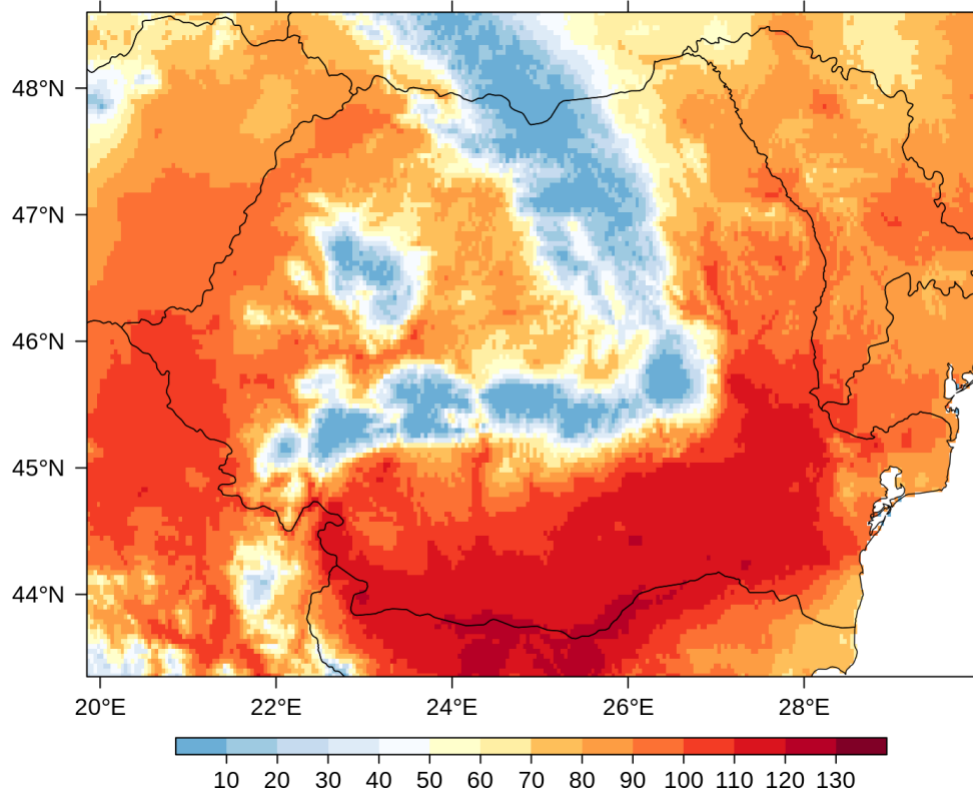


Figura 17: Numărul mediu anual de zile de vară (temperatura maximă a aerului ≥ 25 °C) (2009 - 2017).

PN-III-P1-1.1-PD-2016-1579 Interpolarea spațio-temporală la scară fină a temperaturii aerului 24 pentru teritoriul României

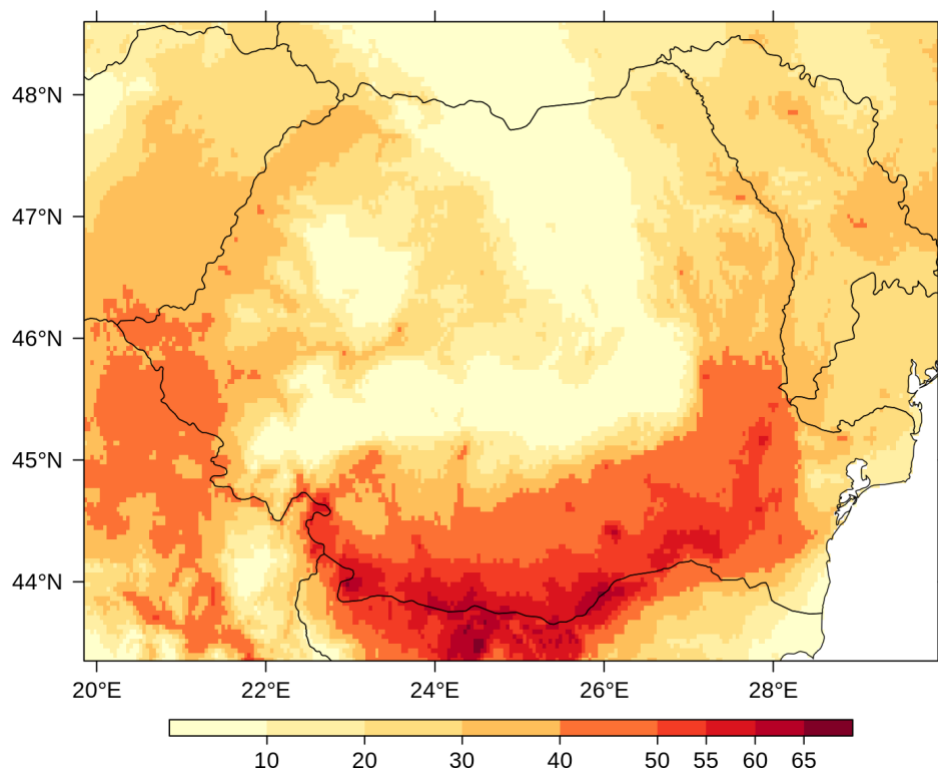


Figura 18: Numărul mediu anual de zile tropicale (temperatura maximă a aerului ≥ 30 °C) (2009 - 2017).

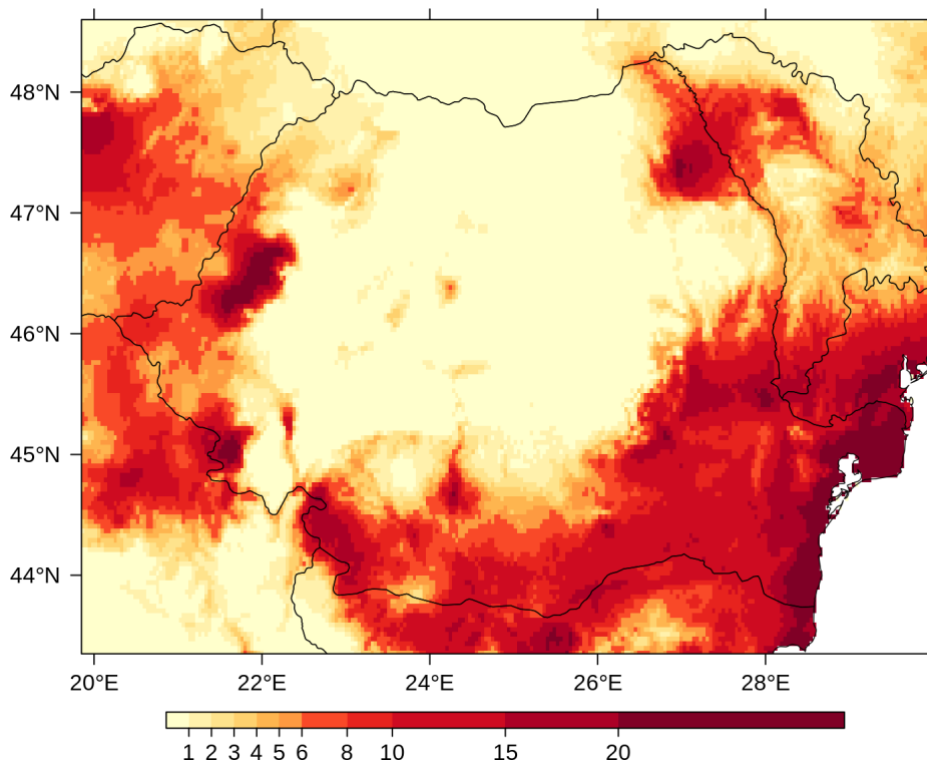


Figura 19: Numărul mediu anual de nopți tropicale (temperatura minimă a aerului ≥ 20 °C) (2009 - 2017).

5.2 Valuri de căldură și frig

Valurile de căldură și de frig sunt generate de o succesiune de mai multe zile în care temperaturile se mențin foarte ridicate sau foarte scăzute. Din cauza seriei temporale reduse a datelor gridate, nu s-a putut utiliza metodologia bazată pe calculul percentilelor (Pacific Climate Impacts Consortium 2020), prin urmare valul de căldura a fost considerat ca intervalul de minim 2 zile cu temperaturi maxime cel puțin egale sau mai mari decât 37°C (Bojariu et al. 2015), iar valul de frig a fost calculat ca intervalul cu minim 2 zile când temperatura minimă este mai mică sau egală decât -15°C.

Valorile de căldură se produc pe teritoriul României frecvent în zonele de câmpie din sudul, estul și vestul țării (Figura 20). Astfel, în anul 2012, în sudul Câmpiei Române s-au înregistrat între 12 și 16 cazuri, iar în anul 2017 s-au semnalat între 4 - 6 cazuri pe regiuni întinse din Câmpia Olteniei și Câmpia Banatului. Valorile de frig se produc cel mai frecvent în zona Arcului Carpatic, dar s-a constatat că în anul 2012 pe regiuni întinse din Câmpia Română și Podișul Moldovei s-au consemnat nopți consecutive cu temperaturi foarte scăzute (Figura 21).

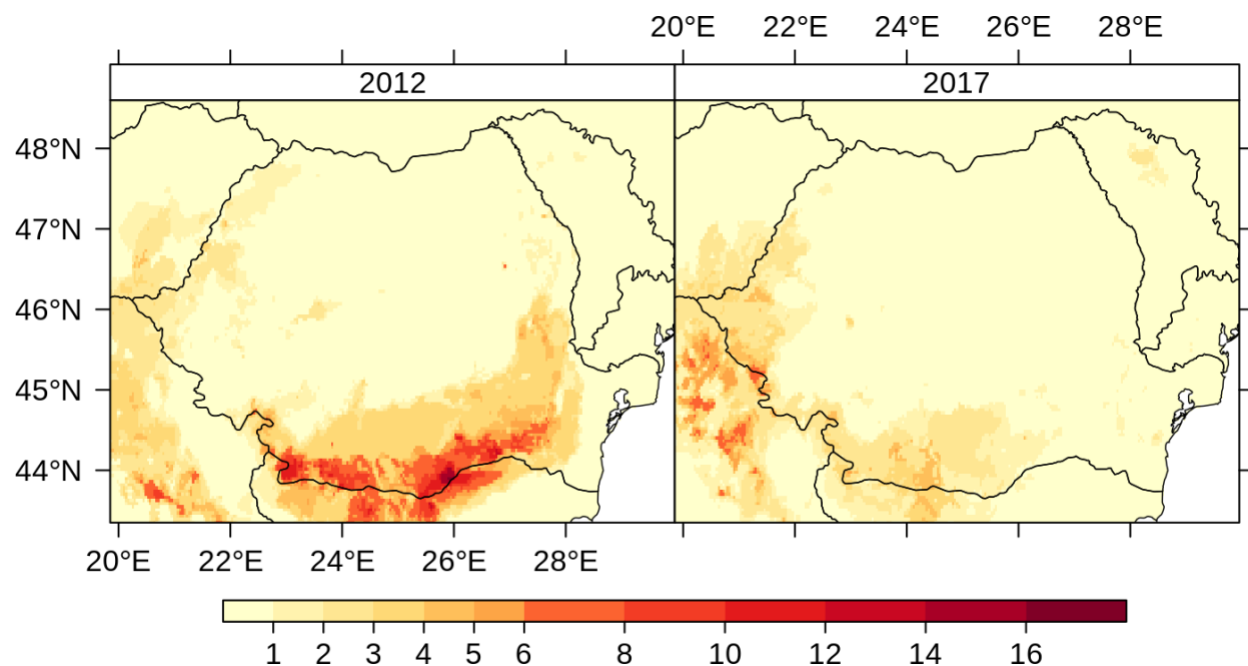


Figura 20: Numărul anual de cazuri cu valuri de căldură (2012 și 2017).

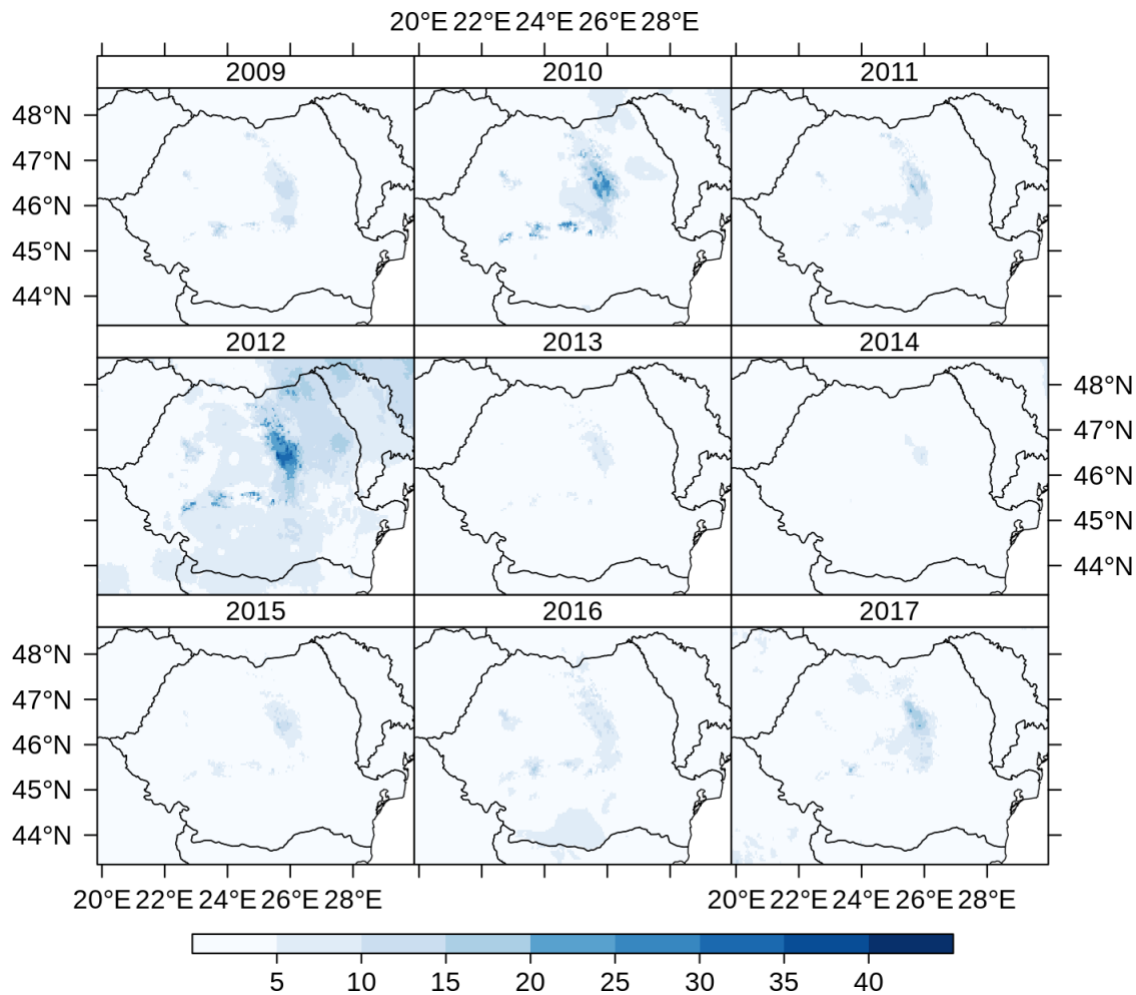


Figura 21: Numărul anual de cazuri cu valuri de frig (2012 - 2017).

6 Publicarea online a seturilor de date gridate și a manualelor de utilizare a produselor

Seturile de date gridate cu distribuția spațială a temperaturii orare sunt disponibile prin serviciul de stocare online a fișierelor Google Drive. Prin accesarea linkului https://drive.google.com/drive/u/1/folders/1qUDQgwgYHNklAyGopIxB4_IDhK8T7RR0 utilizând un browser web (Chrome, Safari etc.), datele pot fi obținute atât prin descărcarea directorului, care conține toate datele, cât și ca fișiere separate, disponibile în subdirectoare. Datele sunt structurate în două subdirectoare principale: *hourly* (<https://drive.google.com/drive/u/1/folders/1pLAuVTBKcukPQ2mw84JfxcvHk0UKUIV2>) și *daily* (<https://drive.google.com/drive/u/1/folders/10gQWWwMZUNwffqojw868pRD6eNGAnsJV>).

În subdirectorul *hourly* se găsesc fișierele netCDF care conțin gridurile orare pentru perioada 2009 - 2017, iar în subdirectorul *daily* sunt arhivate fișierele netCDF care conțin sintezele obținute prin procesarea datelor orare.

Datele netCDF orare sunt arhivate în sudirectoare corespunzătoare fiecărui an, prin urmare folderul *hourly* conține nouă subdirectoare 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016 și 2017 (Figura 22).

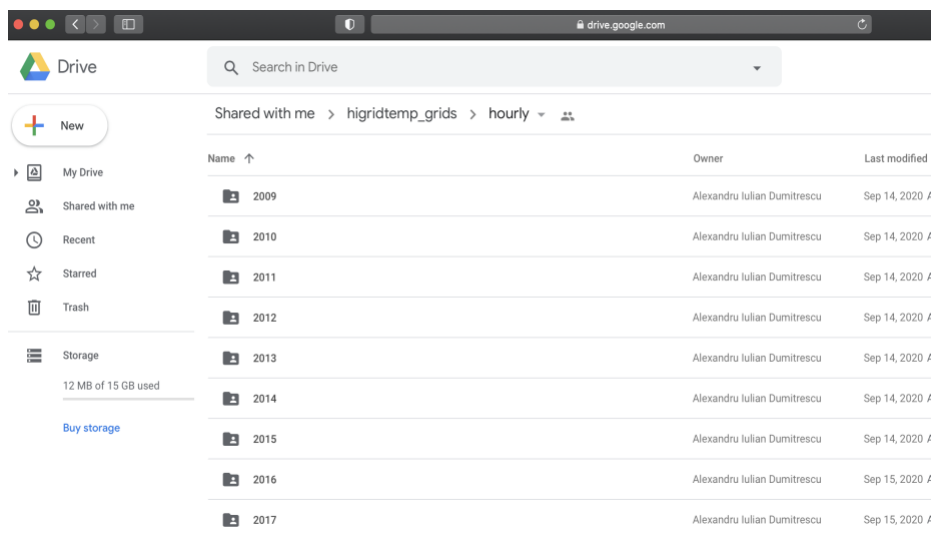


Figura 22: Structura subdirectoarelor în care sunt arhivate fișierele netCDF cu date orare de temperatura aerului.

Datele netCDF obținute din prelucrarea datelor orare sunt arhivate în următoarele subdirectoare: *indicatori* (valuri de căldură și valuri de frig), *nopti_geroase*, *nopti_tropicale*, *range* (amplitudinea), *tmax*, *tmin*, *zile_iarna*, *zile_inghet*, *zile_tropicale*, *zile_vara* (Figura 23).

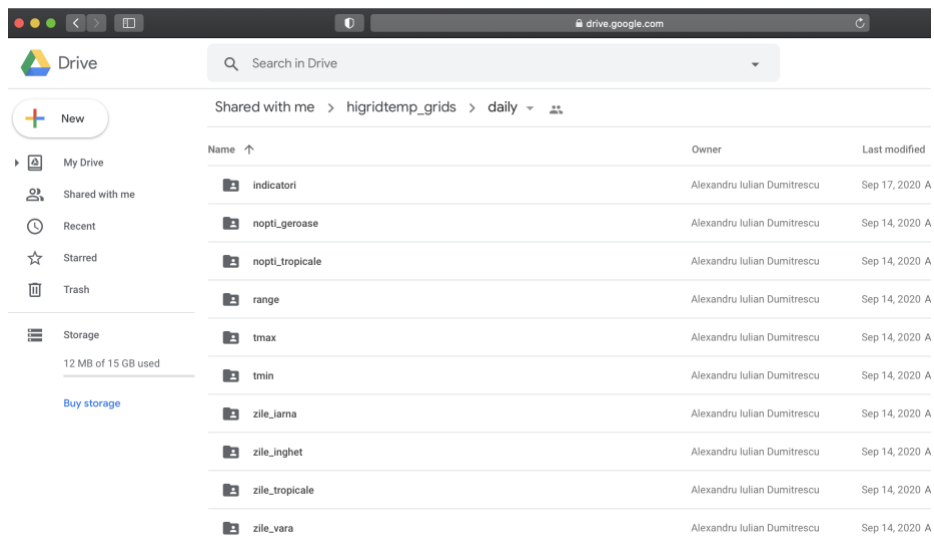


Figura 23: Structura subdirectoarelor în care sunt arhivate fișierele netCDF obținute din prelucrarea datelor orare.

În *Anexa 1* sunt descrise caracteristicile seturile de date gridate, care pot fi descărcate de pe Google Drive, iar în *Anexa 2* sunt prezentate câteva posibilități de obținere și manipulare programatică a datelor gridate, utilizând funcții disponibile în limbajul de programare R.

Bibliografie

Aalto J, Pirinen P, Heikkinen J, Venäläinen A. Spatial interpolation of monthly climate data for finland: Comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*. Springer; 2013;112(1-2):99–111.

Becker RA, Chambers JM. Design of the s system for data analysis. *Communications of the ACM*. ACM New York, NY, USA; 1984;27(5):486–95.

Bojariu R, Bîrsan M-V, Cică R, Velea L, Burcea S, Dumitrescu A, et al. Schimbările climatice–de la bazele fizice la riscuri și adaptare. București, Editura Printech. 2015;

Brinckmann S, Krähenmann S, Bissolli P. High-resolution daily gridded data sets of air temperature and wind speed for europe. *Earth System Science Data*. Copernicus GmbH; 2016;8(2):491.

Cornes RC, Schrier G van der, Besselaar EJM van den, Jones PD. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres* [Internet]. 2018;123(17):9391–409. Available from:

<https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD028200>

D’Agostino McGowan L, Bryan J. googledrive: An Interface to Google Drive [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=googledrive>

Dumitrescu A, Birsan M-V. ROCADA: A gridded daily climatic dataset over romania (1961–2013) for nine meteorological variables. *Natural Hazards*. Springer; 2015;78(2):1045–63.

Dumitrescu A, Birsan M-V, Manea A. Spatio-temporal interpolation of sub-daily (6 h) precipitation over Romania for the period 1975–2010. *International Journal of Climatology*. Wiley Online Library; 2016;36(3):1331–43.

Dumitrescu A, Birsan M-V, Nita I-A. A Romanian daily high-resolution gridded dataset of snow depth (2005–2015). *Geofizika*. Andrija Mohorovicic Geophysical Institute; 2017;34(2):275–95.

Dumitrescu A, Brabec M, Cheval S. Statistical gap-filling of sevir land surface temperature. *Remote Sensing* [Internet]. MDPI AG; 2020a Apr;12(9):1423. Available from: <http://dx.doi.org/10.3390/rs12091423>

Dumitrescu A, Brabec M, Matreata M. Integrating ground-based observations and radar data into gridding sub-daily precipitation. *Water Resources Management* [Internet]. 2020b; Available from: <https://doi.org/10.1007/s11269-020-02622-4>

Dumitrescu A, Cheval S, Guijarro JA. Homogenization of a combined hourly air temperature dataset over romania. *International Journal of Climatology*. Wiley Online Library; 2020c;40(5):2599–608.

Eiselt K-U, Kaspar F, Mölg T, Krähenmann S, Posada R, Riede JO. Evaluation of gridding procedures for air temperature over southern africa. *Advances in Science and Research* [Internet]. 2017;14:163–73. Available from: <https://asr.copernicus.org/articles/14/163/2017/>

Fox DG. Judging Air Quality Model Performance. *Bulletin of the American Meteorological Society*. 1981;62:599–609.

- Haylock M, Hofstra N, Klein Tank A, Klok E, Jones P, New M. A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*. Wiley Online Library; 2008;113(D20).
- Hengl T, Heuvelink GB, Rossiter DG. About regression-kriging: From equations to case studies. *Computers & Geosciences*. Elsevier; 2007;33(10):1301–15.
- Henry L, Wickham H. purrr: Functional Programming Tools [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=purrr>
- Hijmans RJ. raster: Geographic Data Analysis and Modeling. 2020.
- Isaaks EH, Srivastava RM. An introduction to applied geostatistics. Oxford University Press; 1989;
- Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N. Using ArcGIS Geostatistical Analyst. Esri Redlands; 2001.
- Kilibarda M, Hengl T, Heuvelink GB, Gräler B, Pebesma E, Perčec Tadić M, et al. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*. Wiley Online Library; 2014;119(5):2294–313.
- Krähenmann S, Walter A, Brienen S, Imbery F, Matzarakis A. High-resolution grids of hourly meteorological variables for Germany. *Theoretical and Applied Climatology* [Internet]. 2018 Feb;131(3):899–926. Available from: <https://doi.org/10.1007/s00704-016-2003-7>
- LSAF T. Product User Manual Land Surface Temperature (LST), PRODUCTS: LSA-001 (MLST), LSA-050 (MLST-R), LSA-003 (DLST), LSA-002 (ELST). version 3.0. EUMETSAT LSA SAF (Satellite Application Facility on Land Surface Analysis); 2018.
- Lussana C, Seierstad IA, Nipen TN, Cantarello L. Spatial interpolation of two-metre temperature over norway based on the combination of numerical weather prediction ensembles and in situ observations. *Quarterly Journal of the Royal Meteorological Society* [Internet]. 2019;145(725):3626–43. Available from: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3646>
- Mcgill R, Tukey JW, Larsen WA. Variations of box plots. *The American Statistician* [Internet]. Taylor & Francis; 1978;32(1):12–6. Available from: <https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1978.10479236>
- Meyer H, Schmidt J, Detsch F, Nauss T. Hourly gridded air temperatures of south africa derived from msg seviri. *International Journal of Applied Earth Observation and Geoinformation* [Internet]. 2019;78:261–7. Available from: <http://www.sciencedirect.com/science/article/pii/S0303243419300315>
- Pacific Climate Impacts Consortium DB for the. Climdex.pcic: PCIC implementation of climdex routines [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=climdex.pcic>
- Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* [Internet]. 2018;10(1):439–46. Available from: <https://doi.org/10.32614/RJ-2018-009>
- Pebesma E, Bivand RS. Classes and methods for spatial data in R. *R news*. 2005;5(2):9–13.

- Pebesma E, Heuvelink G. Spatio-temporal interpolation using gstat. *RFID Journal*. 2016;8(1):204–18.
- Pebesma EJ. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*. Elsevier; 2004;30(7):683–91.
- Perčec Tadić M. Gridded Croatian climatology for 1961–1990. *Theoretical and Applied Climatology* [Internet]. 2010;102(1):87–103. Available from: <https://doi.org/10.1007/s00704-009-0237-3>
- Pierce D. Ncdf4: Interface to unidata netCDF (version 4 or earlier) format data files [Internet]. 2019. Available from: <https://CRAN.R-project.org/package=ncdf4>
- R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: <https://www.R-project.org/>
- Ruiz-Álvarez M, Alonso-Sarria F, Gomariz-Castillo F. Interpolation of instantaneous air temperature using geographical and modis derived variables with machine learning techniques. *ISPRS International Journal of Geo-Information* [Internet]. MDPI AG; 2019 Aug;8(9):382. Available from: <http://dx.doi.org/10.3390/ijgi8090382>
- Sandu I, Pescaru VI, Poiană I, Geicu A, Cîndea I, Țâștea D. *Clima României*. Editura Academiei Romane, București. 2008;
- South A. rnatuarearth: World Map Data from Natural Earth. 2017.
- Taylor KE. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* [Internet]. 2001;106(D7):7183–92. Available from: <http://dx.doi.org/10.1029/2000JD900719>
- Venables WN, Ripley BD. *Modern applied statistics with S-PLUS*. Springer Science & Business Media; 2013.
- Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=dplyr>

Anexe

Anexa 1: Descrierea datelor gridate

Datele realizate în proiectul HiGridTemp sunt disponibile în mod gratuit. Mai multe informații pot fi obținute de la investigatorul principal: - Alexandru Dumitrescu; - email: alexandru.dumitrescu@gmail.com

Acoperirea spațială

- 18.8° - 30.5° longitudine estică;
- 42.6° - 49.0° latitudine nordică.

Acoperirea temporală

- 2009-01-01 00:00 - 2017-12-31 23:00

Rezoluție spațială

- 0.01° longitudine și 0.01° latitudine, aproximativ 1 km x 1 km.

Rezoluție temporală

- orară.

Proiecție

- Geografică WGS84;
- EPSG cod: 4326;
- definiție proj4: *+proj=longlat +datum=WGS84 +no_defs.*

Formatul fișierelor

- netCDF⁹.
- fișierele sunt denumite conform următoarei structuri: *TT_hourly_aaaallzz_v01.nc*, unde *aaaa* reprezintă anul, *ll* luna și *zz* ziua.

Parametri

- temperatura medie orară a aerului (°C);
- numele variabilei în fișierul netCDF: *tair*.

⁹ <https://www.unidata.ucar.edu/software/netcdf/>

Anexa 2: Exemple de manipulare a datelor gridate utilizând limbajul R

Limbajul R (R Core Team 2020) este un mediu relativ nou de analiză statistică și vizualizare a seturilor de date, care a fost dezvoltat în anii '90 în Auckland, Noua Zeelandă, de către Ross Ihaka și Robert Gentleman. Fiind o implementare a limbajului S (Becker and Chambers 1984) în mediul open-source, inițial, limbajul R dispunea numai de tehnici de analiză statistică clasic. Spre sfârșitul anilor '90 apar primele funcții de analiză statistică a seturilor de date geo-spațiale (Venables and Ripley 2013), acestea constituind primul pas în implementarea claselor de date spațiale în R.

Popularitatea limbajului R a fost determinată atât de capacitatea de a rula pe orice tip de platformă (Linux, MacOS X, Windows), cât și de existența a unui număr impresionant de programe suplimentare (biblioteci), care completează sistemul de bază (în momentul de față sunt peste 180 de biblioteci disponibile în categoria *Spatial*¹⁰).

În secțiunile următoare sunt prezentate câteva posibilități de procesare a datelor gridate, utilizând atât funcțiile de bază implementate în limbajul R, dar și în bibliotecile suplimentare *dplyr* (Wickham et al. 2020), *googledrive* (D'Agostino McGowan and Bryan 2020), *purrr* (Henry and Wickham 2020), *rnaturalearth* (South 2017), *sf* (Pebesma 2018), *ncdf4* (Pierce 2019), *sp* (Pebesma and Bivand 2005) și *raster* (Hijmans 2020).

Limbajul R și mediul integrat de dezvoltare (IDE) RStudio se pot descărca de la următoarele adrese web:

- R language: <https://cran.r-project.org/>;
- RStudio: <https://www.rstudio.com/products/rstudio/#Desktop>.

Bibliotecile suplimentare necesare obținerii și procesării datelor gridate se instalează rulând următoarele comenzi în consola IDE:

```
install.packages("dplyr")
install.packages("purrr")
install.packages("googledrive")
install.packages("rnaturalearth")
install.packages("sf")
install.packages("ncdf4")
install.packages("sp")
install.packages("raster")
```

Descărcarea automată a datelor gridate

Pentru descărcarea automată a datelor din mediul de stocare online GoogleDrive, se pot utiliza funcțiile disponibile în biblioteca R *googledrive*¹¹. Utilizatorul care dorește să obțină

¹⁰ <https://cran.r-project.org/web/views/Spatial.html>

¹¹ <https://googledrive.tidyverse.org/>

datele trebuie să configureze biblioteca *googledrive* astfel încât să poată accesa din mediul R fișierele Google Drive printr-un cont activ Google¹².

Utilizând liniile de script de mai jos se pot descărca datele orare din 1 iunie 2017 în directorul de lucru al mediului R.

```
library(dplyr, quietly = T)
library(purrr, quietly = T)
library(googledrive, quietly = T)

# adresa web unde sunt disponibile datele orare
url.hourly <- "https://drive.google.com/drive/u/1/folders/1pLAuVTBKcukPQ2mw84JfxcvHk0
UKU1V2"

# preluați metadatele pentru fișiere prin adresa web
folder.hourly <- drive_get(as_id(url.hourly), verbose = F)

# extrage nume și id url fișiere din Google Drive
nc.files <- drive_ls(folder.hourly, pattern = "nc", recursive = T)

## Items so far:

## 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900
2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 3100 3200 3296 (yearly folders
plus files)

nc.files <- nc.files %>% mutate( date = substr(name, 11, 18)) %>% mutate(date = as.Date
te(date, "%Y%m%d"))
head(nc.files)

## # A tibble: 6 x 4
##   name                id                drive_resource  date
## * <chr>              <chr>              <list>          <date>
## 1 TT_hourly_20171231_v0... 15cwoYCr0uYPoJ4CEtW0VhdUhi... <named list [38... 2017-12-31
## 2 TT_hourly_20171230_v0... 1_-G_N7iwhFvSF6eWtFUcQtKgm... <named list [38... 2017-12-30
## 3 TT_hourly_20171229_v0... 1gYj7GXF8e-RxR7mM6ewq_Nru... <named list [38... 2017-12-29
## 4 TT_hourly_20171228_v0... 1jtNf650w3JAYzZaLKqczvTdpq... <named list [38... 2017-12-28
## 5 TT_hourly_20171227_v0... 1r1n3TqE0XqGwCeXu2Qe_Yo1Xf... <named list [38... 2017-12-27
## 6 TT_hourly_20171226_v0... 1KImDq33vJ43P9sNoF8WEVZFNb... <named list [38... 2017-12-26

# descarcă fișier cu datele orare din 1 iunie 2017
nc_files.subs <- nc.files %>% filter(format(date, "%Y%m%d") %in% "20170601")
walk(nc_files.subs$id, ~drive_download(as_id(.x), overwrite = T))

## File downloaded:
## * TT_hourly_20170601_v01.nc
## Saved locally as:
## * TT_hourly_20170601_v01.nc
```

¹² https://rdr.io/github/tidyverse/googledrive/man/drive_auth.html

Dacă se dorește descărcare datelor orare pentru un an de zile, se vor modifica ultimele două linii, după cum urmează:

```
# descarcă fișier cu datele orare din anul 2017
nc_files.subs <- nc.files %>% filter(format(date, "%Y") %in% "2017")
walk(nc_files.subs$id, ~drive_download(as_id(.x), overwrite = T))
```

De asemenea, modificând aceleași linii de script, se pot obține fișierele din toți anii, luna ianuarie:

```
# descarcă fișier cu datele orare din Luna ianuarie, perioada 2009-2017
nc_files.subs <- nc.files %>% filter(format(date, "%m") %in% "01")
walk(nc_files.subs$id, ~drive_download(as_id(.x), overwrite = T))
```

Extragere date temperatură orară multipuncte

Datele de temperatură orară pot fi extrase în coordonatele unor puncte, care sunt reprezentate grafic și salvate sub formă tabelară utilizând următoarele comenzi R:

```
library(raster)
library(rnaturalearth)
taer <- brick("TT_hourly_20170601_v01.nc")

# extragere timp straturi
time <- names(taer) %>% gsub("X", "", .) %>% as.numeric() %>%
  as.POSIXct(origin = "1970-01-01 00:00.00 UTC", tz = "UTC")
summary(time)

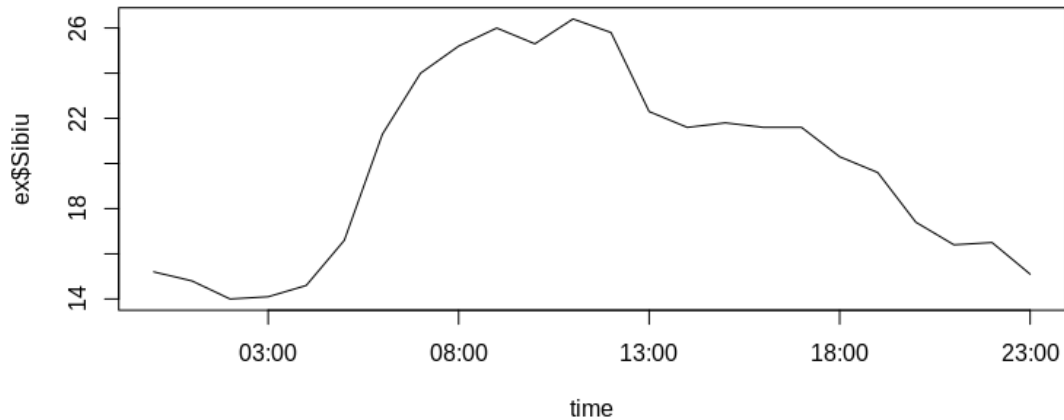
##                Min.                1st Qu.                Median
## "2017-06-01 00:00:00" "2017-06-01 05:45:00" "2017-06-01 11:30:00"
##                Mean                3rd Qu.                Max.
## "2017-06-01 11:30:00" "2017-06-01 17:15:00" "2017-06-01 23:00:00"

# descarcă reședințe de județ
orase <- ne_download(scale = 10, category = 'cultural', type = 'populated_places',
returnclass = "sf") %>% filter(SOV_A3 %in% "ROU")

## OGR data source with driver: ESRI Shapefile
## Source: "/tmp/RtmpSwrpmG", layer: "ne_10m_populated_places"
## with 7343 features
## It has 119 fields
## Integer64 fields read as strings:  wof_id ne_id

# plot(orase["NAME"])

ex <- extract(taer, orase) %>% t() %>% as.data.frame()
names(ex) <- orase$NAME
plot(ex$Sibiu~time, type = "l")
```



```
# scrie datele extrase pentru toate reședințele de județ
write.csv(ex, "temp_orase_20170601.csv")
```

Extragerea datelor de temperatură se poate realiza utilizând fișiere de tip punct SIG (ex. ESRI Shapefile), cu aceeași proiecție cartografică ca a datelor gridate (exemplul de mai sus), sau pot fi extrase prin utilizarea directă a coordonatelor în format numeric (longitudine și latitudine) în funcția *extract*:

```
# extrage datele de temperatură corespunzătoare pixelului cu coordonate 25 Lon și 46 Lat
ex <- extract(taer, cbind(lon = 25, lat = 46)) %>% t() %>% as.data.frame()
names(ex) <- orase$NAME

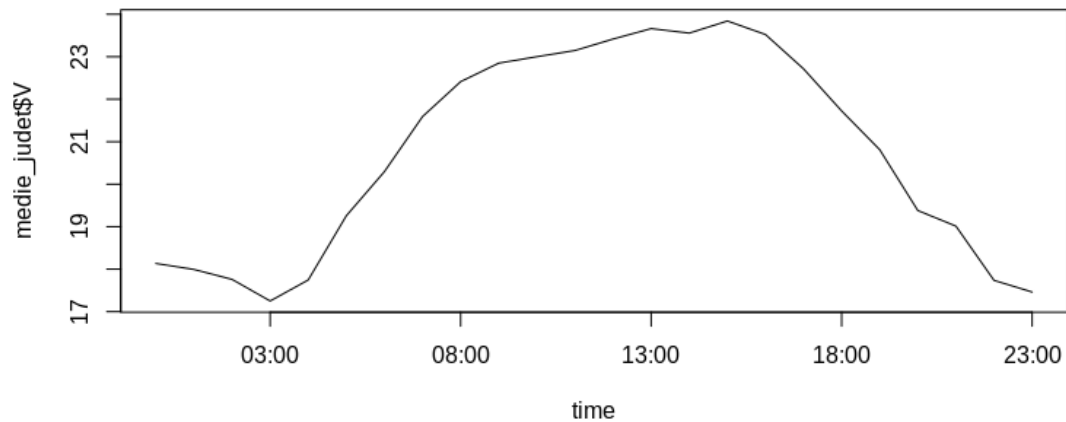
# scrie datele extrase pentru coordonatele cunoscute
write.csv(ex, "temp_25_46_20170601.csv")
```

Extragere date temperatură medie poligon

Datele pot fi extrase și din pixelii care sunt încadrați de un fișier SIG de tip poligon. Valorile pot fi mediate areal, direct în etapa de extragere cu funcția *extract*. În exemplul de mai jos, vor fi extrase și reprezentate grafic datele de temperatură orară din 01 iunie 2017, corespunzătoare județului Buzău.

```
# descarcă limita județelor din România
judet <- ne_states(country = 'romania', returnclass='sf') %>% filter(name %in% "Buzau")

# extrage medie județ Buzău pentru fiecare ora
medie_judet <- extract(taer, judet, fun = mean) %>% t() %>% as.data.frame()
plot(medie_judet$V~time, type = "l")
```

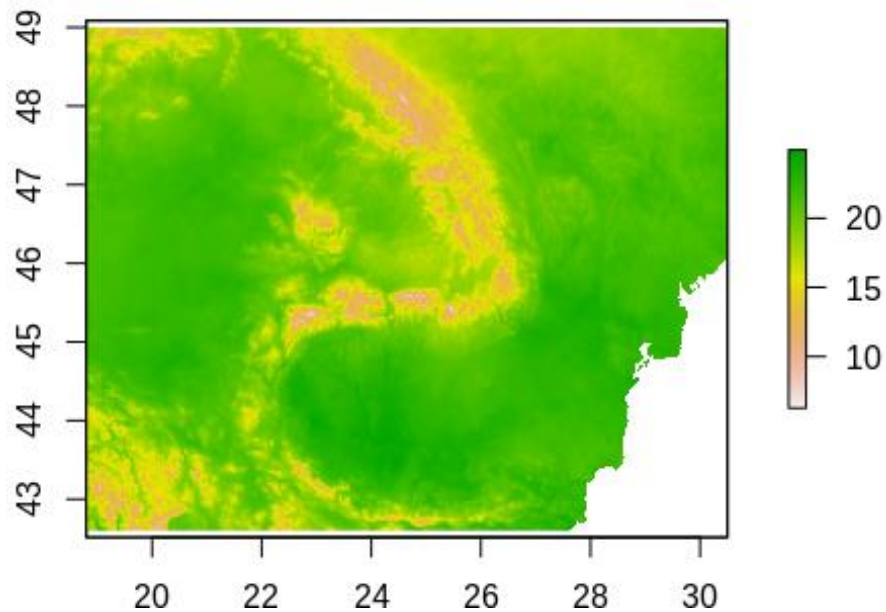


```
# scrie datele extrase ca medie areala pentru județul Buzău
write.csv(ex, "temp_buzau_20170601.csv")
```

Calcul valori zilnice

Pentru calculul mediei zilnice, pentru fiecare punct de grilă, se poate utiliza funcția `mean()` din biblioteca R raster.

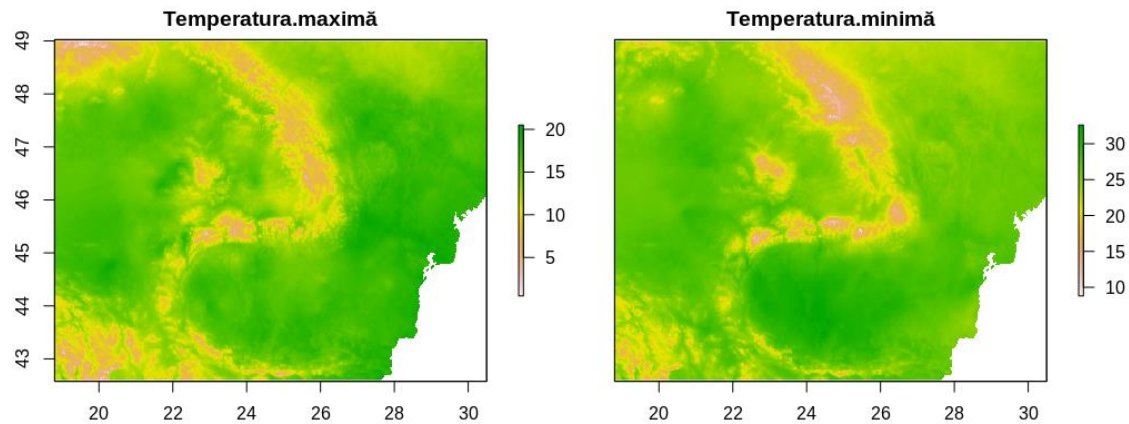
```
# calculează media zilnică 2016-06-01 din fișierul care conține datele orare Taer
taer.mean <- taer %>% mean()
plot(taer.mean)
```



```
# salvează fișierul ca GeoTIFF
writeRaster(taer.mean, "taer_zilnic_20170601.tif", overwrite = T)
```

Pentru calculul valorile maxime se poate utiliza funcția `max()`, iar pentru obținerea minimelor funcția `min()`:

```
# calculează media din fișierul care conține datele orare Taer
taer.min <- taer %>% min()
taer.max <- taer %>% max()
tmin.tmax <- brick(taer.min, taer.max)
names(tmin.tmax) <- c("Temperatura maximă", "Temperatura minimă")
plot(tmin.tmax)
```



În același mod, pot fi aplicate alte funcții pentru prelucrarea datelor gridate orare. Exemple de funcții care sunt implementate în R și pot fi utilizate în procesarea datelor de tip raster sunt disponibile și aici <https://rspatial.org/raster/pkg/4-algebra.html>.