

Tehnici de portare a datelor - Raport tehnic -

Elaborat în cadrul contractului **34PCCDI/2018**

Instrumente de modelare a proceselor de interfață Apă-Sol-Plante pentru administrarea inteligentă și durabilă a bazinelor hidrografice și a ecosistemelor dependente de apa subterană

Proiect P1 SITRA-CLIM

Autori: Stelian Ion
Dorin Marinescu
Ștefan Gicu-Cruceanu

Instituția: Institutul de Statistică Matematică și Matematică Aplicată
“Gheorghe Mihoc - Caius Iacob” al Academiei Române

1 Introducere

Procesele bio-fizico-chimice, observațiile directe și modelele teoretice aferente au un spectru larg de scări spațio-temporale de reprezentare. Metodologiile de colectare a datelor diferă de la un domeniu la altul, fiecărui domeniu fiindu-i specifică o anumite rețetă de distribuție spațială a punctelor de observație sau o anumite frecvență temporală a acestora. În plus, anumite modele matematice necesită distribuții spațiale continue ale unor variabile care nu pot fi măsurate decât într-un număr limitat de puncte. Rezultă astfel că este utilă elaborarea unor instrumente matematice de prelucrare și analiză a datelor care să ofere posibilitatea prelucrării lor pentru cerințe multiple și grade diferite de complexitate și acuratețe.

În acest raport analizăm două probleme legate de transferul de informații, și anume:

1. *Cunoscând valorile unei variabile spațiale la nivelul unității de discretizare, să se determine variația ei în interiorul unității.*
2. *Cunoscând valorile unei variabile spațiale într-un număr de puncte dintr-un domeniu dat, să se determine valorile ei în oricare alt punct din domeniu.*

Prima problemă este tipică cazului datelor obținute prin observații satelitare. A doua problemă este specifică monitoringului integrat, iar în acest caz, o mare parte din date sunt obținute prin măsurători punctuale și este de dorit ca ele să fie utilizate de cât mai mulți utilizatori.

Analiza și prelucrarea datelor spațiale este un subiect vast în științele mediului. Au fost elaborate modele de analiză dintre cele mai diverse așa încât o inventariere a lor depășește cadrul acestui raport, pentru aceasta recomandăm lucrarea [7].

Soluția primei probleme se bazează pe folosirea unor variabile auxiliare măsurate la un nivel de rezoluție mai fin. Algoritmii de reconstrucție presupune existența unei relații de dependență funcțională între cele două variabile și utilizând această relație, variabila primară se coboară la scara mai fină.

Metodele de a rezolva a două problemă pot fi clasificate în trei categorii (a) geostatistice, (b) deterministe și (c) combinate, [7]. Pentru metodele geostatistice recomandăm softul cu surse libere *R* [1] care conține mai multe module de prelucrare a datelor spațiale, **gstat**, **sp**, **rgeos**. În cadrul contractului am dezvoltat o metodă nouă de interpolare a datelor arbitrar distribuite. Metoda noastră permite rezolvarea a două probleme legate de interpolarea datelor: distribuția neregulată a punctelor de observație și problema interpolării unui număr foarte mare de date. În acest sens amintim două metode utilizate: metoda rețelelor neuronale și metoda dezvoltărilor ierarhice.

Ambele metode utilizează dezvoltarea funcției interpolatoare în baza generată de o funcție radială

$$f(\mathbf{x}) = \sum_{i \in I} \alpha_i \Phi(\|\mathbf{x} - \mathbf{c}_i\|)$$

În relația precedentă, $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ este funcția generatoare, $\{\mathbf{c}_i\}_{i \in I}$ sunt punctele de colocație, iar α_i sunt coeficienții dezvoltării ce urmează a fi determinați pe baza datelor.

În metoda *rețelelor neuronale*, [2], punctele de colocație diferă de punctele de observație atât ca număr cât și ca poziție, determinarea lor făcându-se în procesul de învățare.

Metoda *dezvoltărilor ierarhice*, [3], punctele de colocație sunt submulțimi ale punctelor de observație organizate ierarhic $I_1 \subset I_2 \subset \dots$.

Ambele metode au câte un punct critic, numărul de punctelor de colocație, în prima metodă, respectiv alegerea șirului ierarhic în a doua metodă.

Metoda dezvoltată de noi rezolvă ambele probleme prin introducerea unei scheme de partiționare a domeniului spațial de analiză și utilizarea spațiului spline wavelets ca spațiu de aproximare.

Raportul include:

1. suportul matematic al schemelor de portare;
2. aplicații numerice.

Toate schemele de calcul au fost incluse în softul ASTERIX-IADS dedicat analizei și prelucrării datelor spațiale. Acest soft este în dezvoltare și testare.

2 Aplicații ale softului ASTERIX-IADS

Exemplele considerate și rulate în această secțiune au drept scop validarea numerică a algoritmilor de calcul și testarea softului. Softul a fost proiectat astfel încât să necesite un număr cât mai redus de parametri. Vom prezenta câteva aplicații însoțite de indicații referitoare la datele de intrare și posibilele domenii de aplicare.

2.1 Probleme de “down-scaling”

Detalii teoretice legate de schema numerică pot fi găsite în secțiunea *Downscaling Schemes* a capitolului *Data porting technique across scales. Applications to Soil-Plant-Atmosphere System* din acest raport.

Aplicație. *Determinarea valorilor unei variabile spațiale f pe o grilă de înaltă rezoluție cunoscând distribuția ei pe o grilă de joasă rezoluție.*

Soluție. Trecerea de la o rezoluție joasă la o rezoluție înaltă se face cu ajutorul unei variabile auxiliare g distribuită pe grila fină și aflată într-o relație funcțională cu f .

Date de intrare. Distribuția celor două variabile și relația funcțională.

Domeniu de aplicabilitate. Această problemă este tipică pentru datele obținute prin măsurători satelitare. De exemplu, pentru umiditatea solului există determinări în banda de radio-frecvență la o rezoluție de $40km \times 40km$. Temperatura la suprafața solului și indicele de vegetație normalizat pot fi măsurate la o rezoluție de $1km \times 1km$. Utilizând aceste mărimi, se determină umiditatea solului la scara de $1km \times 1km$, [4, 5, 6].

Formula de trecere de la o rezoluție la alta este dată de

$$f_{HR} = f_{LR} + \partial_g f(g_{HR} - g_{LR}),$$

unde indicii HR și LR desemnează scara de înaltă și respectiv joasă rezoluție pentru valorile variabilelor f și g , iar $\partial_g f$ este derivata lui f în raport cu g . Evaluarea acestei ultime mărimi este punctul critic al metodei.

Pentru validare am folosit următoarele date

$$\Omega = [-1, 1] \times [-1, 1] \in \mathbb{R}^2, \quad g(x, y) = a_1 + a_2 \cos \alpha(x + y), \quad f(x, y) = g^2(x, y), \quad (1)$$

unde Ω reprezintă domeniul de definiție al variabilelor f și g .

Pentru a determina eroarea de reconstrucție a variabilei f am folosit date exacte ale ei pe cele două grile precum și valori reconstruite ale ei cu ajutorul formulei (1), evaluarea derivatei fiind considerată exactă. Rezultatele obținute sunt trecute în tabelul 2.1 și figura 2.1.

Grilă "grosieră"	Grilă rafinată		
	8×8	16×16	32×32
4×4	0.036	0.042	0.044
	0.10	0.12	0.12
16×16	0.06	0.06	0.06
	0.16	0.17	0.17
32×32	0.077	0.089	0.093
	0.07	0.09	0.09
64×64	0.021	0.026	0.028
	0.02	0.02	-

Table 1: Analiza erorii de reconstrucție pe baza relației funcționale. Frecvența $\alpha = 4\pi$ și coeficientul $a_1 = 0.5$ ai funcției de modelare g sunt fixați, iar amplitudinea a_2 este variabilă. În acest tabel sunt trecute erorile de reconstrucție pentru două valori ale amplitudinii a_2 : $a_2 = 0.3$ în prima linie și $a_2 = 0.5$ în a doua linie a casetelor.

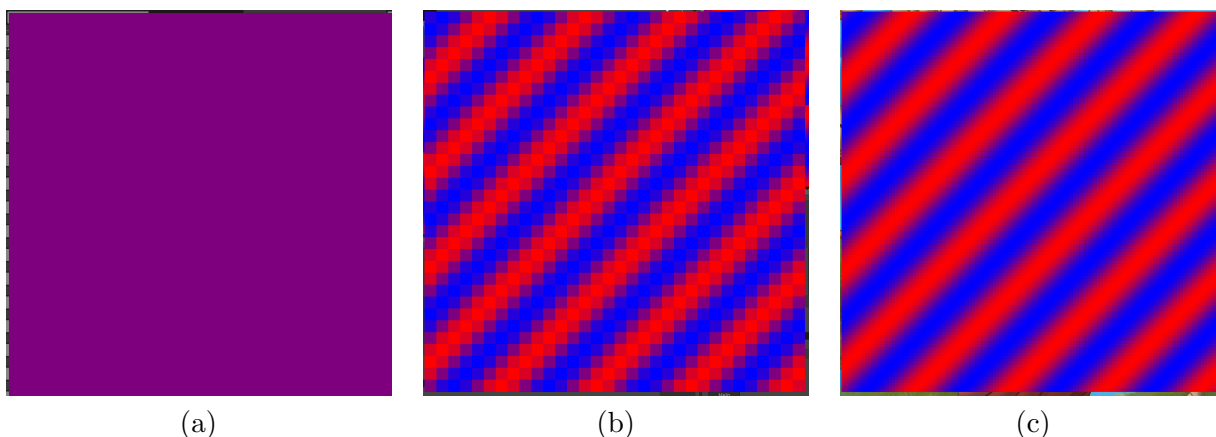


Figure 1: Trei distribuții ale valorilor funcției f : cea originală, pe nivelul de joasă rezoluție (a) și cele reconstruite pe grile 8×8 (b) și 32×32 (c).

Comentarii:

1. În cazul în care cele două variabile sunt liniar dependente, atunci reconstrucția este exactă.
2. Acuratețea reconstrucției depinde de nivelul de reconstrucție și de variația lui g .
3. În cazul în care nu se cunoaște relația dintre cele două variabile, coeficientul $\partial_g f$ este necunoscut și deci trebuie definit un procedeu de estimare a derivatei funcționale folosind informații suplimentare. Aceasta este o problemă deschisă.

2.2 Probleme de interpolare a datelor arbitrar distribuite

Metoda dezvoltată în cadrul proiectului este o combinație a metodei de interpolare *thin plate spline* cu o metodă de tip *spline wavelets*. Utilizarea bazelor wavelets în științele mediului este nouă, de văzut lucrările [8, 9].

Principalele avantaje ale metodei sunt:

1. nu sunt impuse restricții asupra regularității repartiției punctelor spațiale;
2. comportament local;
3. posibilitatea eliminării unor perturbații aleatoare ale datelor;
4. modelarea la scări multiple.

Mai multe detalii matematice legate de schema numerică se găsesc în secțiunea *Upscaling Schemes* a capitolului *Data porting technique across scales. Applications to Soil-Plant-Atmosphere System* din acest raport.

Aplicație. *Cunoscând valorile unei variabile spațiale într-un număr de puncte dintr-un domeniu dat, să se determine valorile ei în oricare alt punct din domeniu.*

Soluție. Metoda de interpolare spațială, TPS-MRA, formulele (21-25) din secțiunea menționată mai sus.

În cele ce urmează ilustrăm performanțele algoritmului LRS și MRA-SD folosind un caz teoretic. Datele sunt generate de funcția lui Franke, [10]

$$\begin{aligned}
 f(x, y) = & 0.75 * \exp \left(- \left(\frac{9x-2}{2} \right)^2 - \left(\frac{9y-2}{2} \right)^2 \right) + \\
 & + 0.75 * \exp \left(- \left(\frac{9x+1}{7} \right)^2 - \frac{9y+1}{10} \right) + \\
 & + 0.5 * \exp \left(- \left(\frac{9x-7}{2} \right)^2 - \left(\frac{9y-2}{2} \right)^2 \right) - \\
 & - 0.2 * \exp \left(- (9x - 4)^2 - (9y - 7)^2 \right)
 \end{aligned}$$

definită pe domeniul $D = [-1, 1] \times [-1, 1]$. Punctele de observație \mathbf{x}_i sunt selectate aleatoriu în domeniul D , iar valorile observate sunt $z_i = f(\mathbf{x}_i)$. Ordinul funcției thin plate spline este $q = 3$. Performanțele schemei sunt date în tabelul 2, unde

$$e_2^{TPS} = \frac{\sqrt{\sum_i (\mathcal{Q}^I f(\mathbf{x}_i) - f(\tilde{\mathbf{x}}_i))^2}}{2^J}, \quad e_2^{MRA} = \frac{\sqrt{\sum_i (\mathcal{Q}f^j(\mathbf{w}_i) - f(\mathbf{w}_i))^2}}{M},$$

cu \mathbf{w}_i formând o rețea regulată cu $M = 61 \times 61$ noduri pe D . J este nivelul maxim de rezoluție determinat automat în funcție de distribuția punctelor de observație. Nodurile $\tilde{\mathbf{x}}_i$ sunt punctele de colocație pentru interpolarea spline, sv reprezintă valoarea de trunchiere în șirul valorilor singulare din algoritmul LRS, iar j reprezintă nivelul funcțiilor de scalare în reprezentarea wavelet.

Reamintim că pentru nivelul maxim J calculat, aproximarea spline wavelet este dată de

$$Qf^J(x, y) = \sum_{k,l} a_{k,l}^J \varphi_{k,l}^J(x, y), \quad (2)$$

iar pentru nivele inferioare $4 \leq j \leq J$ avem o aproximare similară

$$Qf^{J,j}(x, y) = \sum_{k,l} a_{k,l}^{J,j} \varphi_{k,l}^j(x, y), \quad (3)$$

în care coeficienții $\mathbf{a}^{J,j}$ se obțin prin algoritmi de deconstrucție și reconstrucție ai lui \mathbf{a}^J .

N	sv, j							
	$\mu = 0, j = J$		$\mu = 1e - 4, j = J$		$\mu = 0, j = 4$		$\mu = 1e - 4, j = 4$	
	e_2^{TPS}	e_2^{MRA}	e_2^{TPS}	e_2^{MRA}	e_2^{TPS}	e_2^{MRA}	e_2^{TPS}	e_2^{MRA}
25	1.0e - 1,	9.8e - 1	1.0e - 1,	1.0e - 1	1.0e - 1,	9.8e - 1	1.0e - 1,	1.0e - 1
100	4.0e - 2,	4.1e - 2	5.0e - 2,	4.7e - 2	4.0e - 2,	4.1e - 2	5.0e - 2,	4.6e - 2
2500	9.1e - 4,	1.6e - 4	1.6e - 3,	1.1e - 3	9.1e - 4,	2.0e - 3	1.6e - 3,	2.1e - 3
100000	1.5e - 4,	5.0e - 4	2.1e - 4,	5.4e - 4	1.5e - 4,	1.9e - 3	2.1e - 4,	1.9e - 3
1000000	1.8e - 7,	2.1e - 7	4.1e - 7,	3.5e - 7	1.8e - 7,	1.9e - 3	4.1e - 7,	1.9e - 3

Table 2: Metoda de interpolare spațială TPS-MRA. e_2^{TPS} măsoară acuratețea interpolării thin plate spline, iar e_2^{MRA} măsoară acuratețea metodei combinate thin plate spline – spline wavelet.

Analiza multirezoluție a datelor discrete permite rezolvarea a trei probleme principale:

1. *Problema redundanței.* Cele un milion de puncte conțin aproape tot atâta informație cât pot stoca cei 17×17 coeficienți $\mathbf{a}^{J,4}$.
2. *Problema selecției scării specifice a variabilei observate.* Prin analiza mărimii coeficienților wavelets se poate detecta nivelul de la care aceștia devin neglijabili, acest nivel reprezentând scara spațială specifică variabilei analizate.
3. *Problema vitezei de procesare a datelor.* Odată rezolvată problema interpolării, procesarea funcției calculate devine simplă: funcțiile bazei au suport compact, iar pentru un punct \mathbf{x} sunt doar 16 astfel de funcții ale căror suporturi îl conțin pe \mathbf{x} .

Aplicație. Setul de date $\{P_i(x_i, y_i), \tilde{z}_i\}_{i=1,N}$ conține o serie de erori în valorile z_i măsurate. Se cere determinarea valorilor funcției “adevărate”.

Soluție Metoda de interpolare spațială, TPS-MRA, cu filtrare.

Datele au fost generate utilizând funcția lui Nielsen

$$f(x, y) = 0.5y \cos^4(a * (x^2 + y - 1)).$$

Cele $N = 250000$ puncte de observație sunt distribuite aleatoriu (uniform) în domeniul $D = [0, 1000] \times [0, 1000]$. Valorile măsurate sunt date de

$$z_i = f(x_i, y_i) * (1 + 0.1(U() - 0.5)),$$

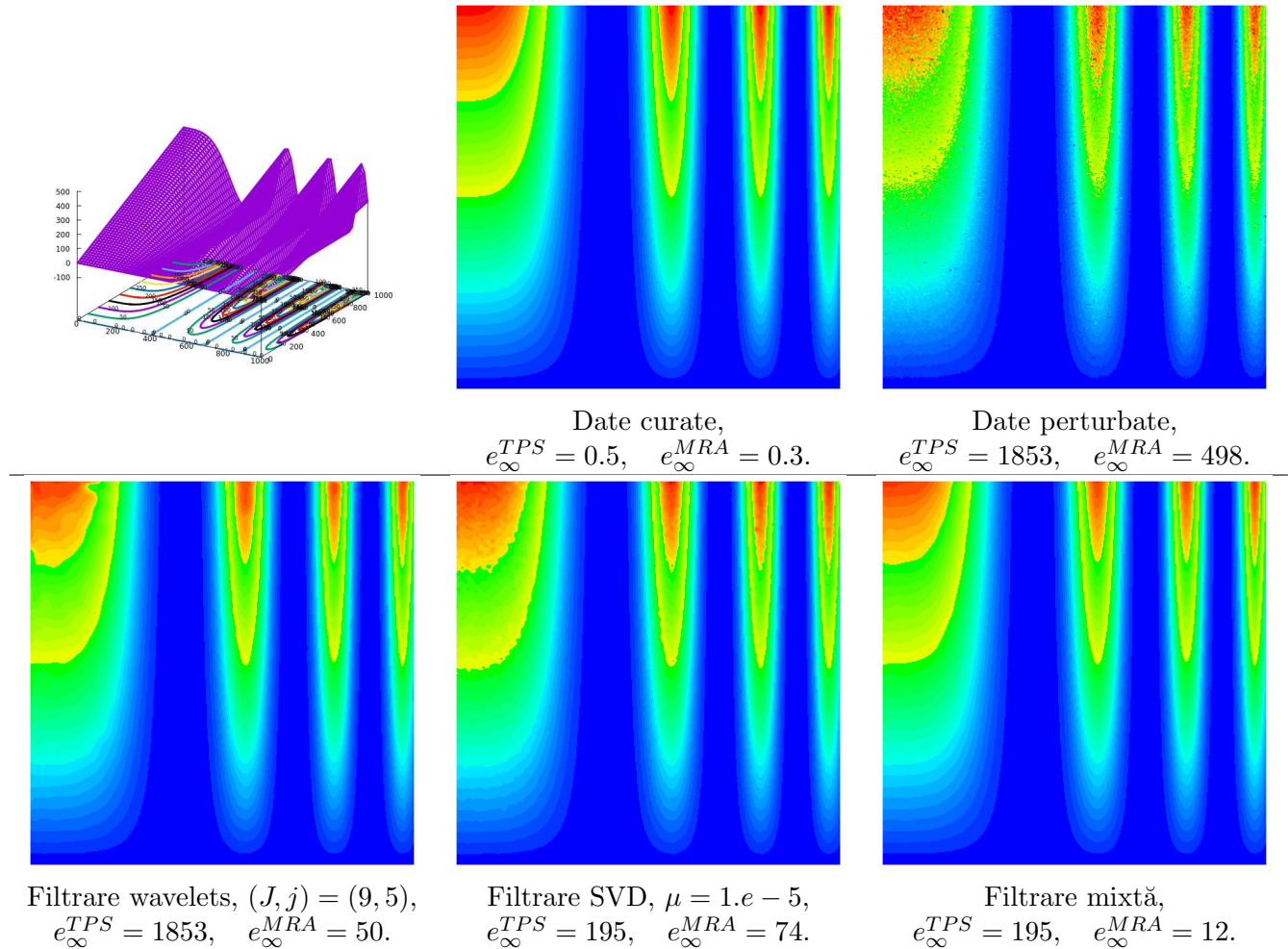


Figure 2: Aplicație la procedura de filtrare a datelor cu perturbații. e_{∞} reprezintă eroarea în norma $\|\cdot\|_{\infty}$ a interpolării față de funcția neperturbată folosind thin plate spline (e_{∞}^{TPS}) și spline wavelets (e_{∞}^{MRA}). Valoarea maximă observată a fost 523, iar cea minimă 0.

unde $U()$ este distribuția uniformă pe intervalul $(0,1)$ și $a = 1.0e - 5$. Toate datele sunt mărimi adimensionale.

Comentarii:

1. Interpolarea cu funcții thin plate spline a datelor cu erori induce erori foarte mari.
2. Metoda combinată thin plate spline – spline wavelet oferă un instrument performant de interpolare a datelor spațiale cu erori în date.

3 Concluzii

Reconstruirea unui câmp discret sau continuu pentru o observabilă, plecând de la valorile ei măsurate pe o anumită rețea este în general o problemă delicată. Acest câmp pe care îl construim este afectat de distribuția datelor măsurate, numărul acestora, eroarea de măsurare etc. Nu se poate da o rețetă universală de reconstrucție a observabilei care să fie optimă pentru orice tip de date măsurate. În funcție de tipul de date, problema trebuie rezolvată prin metode specifice. Instrumentele matematice expuse aici tratează probleme de prelucrare și analiză a datelor cu o structură des întâlnită în științele mediului.

În secțiunea 2.1 am răspuns la următoarea problemă: date fiind două rețele regulate de date, una grosieră unde cunoaștem valorile unui câmp f , alta mai fină unde sunt date valorile altui câmp g și cunoscând o relație funcțională între aceste două câmpuri, să se reconstruiască valorile lui f în rețeaua mai fină. Această metodă este utilă, de exemplu, pentru prelucrarea datelor colectate prin intermediul sateliților.

În secțiunea 2.2 am prezentat o metodă de interpolare și analiză a datelor pe rețele neregulate cu o tehnică bazată pe o combinație între interpolarea thin plate spline și analiza multirezoluție cu spline wavelets, care în plus permite filtrarea erorilor de un anumit tip pentru datele măsurate. Metoda aceasta se pretează pentru date provenite din măsurători efectuate în teren.

References

- [1] R Core Team (2019), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>.
- [2] T. Poggio, F.. Girosi, *Networks for Approximation and Learning*, Proceedings of the IEEE, vol. 78, No. 9, September 1990.
- [3] Armin Iske, *Hierarchical Scattered Data Filtering for Multilevel Interpolation Schemes*, in Mathematical Methods for Curves and Surfaces, eds. Tom Lyche and L.L. Schumaker, Vanderbilt University Press, Nashville, TN., 2000
- [4] Andreas Colliander *et al.*, *Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI during SMAPVEX15*, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 14, NO. 11, NOVEMBER 2017
- [5] Merlin, O., J. P. Walker, A. Chehbouni, and Y. Kerr (2008a), *Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency*, Remote Sens. Environ., 112(10), 3935–3946.
- [6] Fang, B., V. Lakshmi, R. Bindlish, and T.J. Jackson, *Downscaling of SMAP soil moisture using land surface temperature and vegetation data*, Vadose Zone J. 17:170198, 2018, doi:10.2136/vzj2017.11.0198
- [7] Jin Li, Andrew D. Heap, *Spatial interpolation methods applied in the environmental sciences: A review*, Environmental Modelling&Software 53, (2014) 173–189

-
- [8] Mitasova, H., Mitas, L., Brown, W.M., Gerdes, D.P., Kosinovsky, I., Baker, T., 1995. *Modelling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS*, Int. J. Geogr. Inf. Syst. **9** (4),(1995), 433–446.
- [9] Junjie Wang, Tiejun Wang, Tiezhu Shu, Guofeng Wu, and Andrew K. Skidmore, *A Wavelet-Based Area Parameter for Indirectly Estimating Copper Concentration in Carex Leaves from Canopy Reflectance*, Remote Sens., **7**, 15340-15360; doi:10.3390/rs71115340, 2015.
- [10] Franke, R., and Neilson, G., *Smooth interpolation of large sets of scattered data*, International Journal of Numerical Methods in Engineering 15, 1691–1704, 1980.

Data porting technique across scales. Applications to Soil-Plant-Atmosphere System*

Stelian Ion, Dorin Marinescu, Ștefan Gicu-Cruceanu

Technical report[†]

Abstract

The transfer information across scales is of great importance in environmental sciences. The data collection from direct observations is the only fundamental information that one can have about a natural phenomenon. Based on this direct information, the environmental scientists test hypotheses or make scenarios about how the natural systems work. The components of a natural system interact at different levels and in order to explain this, one needs models at different scales. In this report we address the problem of building up an interface that allows different scaled models to access a given set of observation data.

Mathematical Subject Classification(MSC2010): 65D05, 15A24, 90C05, 90C31

Key words: downscaling, upscaling, remote sensing, scatter data, thin plate spline

Contents

1	Introduction	2
2	Downscaling Schemes	2
2.1	The Functional Relationship Method	3
3	Upscaling Schemes	5
3.1	Thin Plate Spline Interpolation Method	6
3.2	Cubic Spline Approximation of Scatter Data	8
3.2.1	Multiresolution Analysis (MRA) of Scatter Data	9
4	Conclusions	10
5	Appendix	11

*This work was supported by a grant of Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0721/34PCCDI/2018 within PNCDI III.

[†]This report was elaborated on the base of the activity within the Project 1 SITRA-CLIM.

1 Introduction

Scatter data interpolation is a problem of representation: given a set of discrete data which are assumed to be the values of a certain quantity measured on a set of locations, one wishes to know the values of that quantity at a new location. The first step in solving this problem is to choose a class of functions inside which the representation is sought. This is a matter of modeling, not a mathematical problem. The second step is to effectively determine the element of this functional space that approximates the scatter data. This is a problem of mathematics and there is a large bibliography concernig the solution of the approximation problem. In Section 3.1 we present an approximation method with radial basis functions.

In this report, we propose a method of representing the scatter data in the cubic spline wavelet space. Throughout this paper, by scatter data we mean a set of data

$$\mathcal{S} := \{(x_i, y_i), z_i\}_{i=\overline{1, N}}, \quad \text{with } (x_i, y_i) \in D \subset \mathbb{R}^2, \quad z_i \in \mathbb{R}. \quad (1)$$

We assume there is a function $f(x, y)$ defined everywhere on D , called *support function*, such that

$$f : D \rightarrow \mathbb{R}, \quad f(x_i, y_i) = z_i. \quad (2)$$

This function f is generally unknown. To exactly recover f from the scatter data \mathcal{S} is an ill-posed problem: there is an infinity of functions that coincide on \mathcal{S} .

To deal with a treatable problem, one searches for a model function $\mathcal{Q}f$ in a certain functional space that preserves certain properties of f and approximates the scatter data \mathcal{S} .

2 Downscaling Schemes

Downscaling means to disaggregate the scale information to smaller scales. In the case of data stored on a net of cells, this requires to guess the variation of the field inside a cell on which one knows some averaged value of that field. This kind of problem is encountered, for example, in the remote sensing applications, [3, 4, 5].

Mathematically, one faces the following problem: if a partition $\Pi = \{\omega_i\}_{i=\overline{1, N}}$ of a domain Ω and the values $\{f_i\}_{i=\overline{1, N}}$ of a field \mathcal{F} over this partition Π are given, then find the values of the field \mathcal{F} on a new given partition $\mathcal{T} = \{\Delta_a\}_{a=\overline{1, M}}$ than is finer than Π .

To be mathematically treatable, one needs to make certain hypotheses concerning the field \mathcal{F} , the domain Ω and the partitions Π and \mathcal{T} .

Physical continuum model

- 1) Ω is a domain of \mathbb{R}^n .
- 2) $\mathcal{F} : \Omega \rightarrow \mathbb{R}$ is a continuous function.

Rough and refined partitions

$$3) \quad \Omega = \bigcup_{i=1}^N \omega_i = \bigcup_{a=1}^M \Delta_a$$

4	2	7	2	
1	6			
4	3	4	8	
6	5	3		
4			6	

Figure 1: Downscaling the gridded data. On each cell from the coarser grid, one knows an averaged value. Inside the gross cell, the field can variate and the problem is to find this variation.

- 4) For any $a = \overline{1, M}$ there is a unique element $\omega_{i(a)}$ of Π such that $\Delta_a \subset \omega_{i(a)}$ and for any $i = \overline{1, N}$ there is a subset $A(i) \subset \{1, \dots, M\}$ such that $\omega_i = \bigcup_{a \in A(i)} \Delta_a$.

The first two assumptions affirm that we work in the frame of continuum physics. The third assumption ensures us that there is no gap in the information concerning the field \mathcal{F} and the two partitions cover the same domain. The fourth assumption sets the frame for downscaling problems.

2.1 The Functional Relationship Method

The method assumes the existence of two fields related by a functional relation. One of the fields is known on the finer partition and its values are used as pattern for the variation of the other field, [3, 4].

Let $\{\Pi, f\}$ and $\{\mathcal{T}, g\}$ be given. This means we know the discrete values $\{f_i\}_{i=\overline{1, N}}$ and $\{g_a\}_{a=\overline{1, M}}$. One defines the downscaled values of the field \mathcal{F} by

$$f_a^* = f_{i(a)} + \partial(f, g)|_{i(a)} (g_a - \tilde{g}_{i(a)}), \quad a = \overline{1, M}, \quad (3)$$

where

- $i(a)$ is the index corresponding to the domain ω_i that includes the subdomain Δ_a ,
- $\tilde{g}_{i(a)}$ represents an average of the all values g_b with $b \in A(i(a))$:

$$\tilde{g}_{i(a)} = \frac{1}{m(\omega_{i(a)})} \sum_{b \in A(i(a))} m(\Delta_b) g_b, \quad (4)$$

where m is a measure on Ω used to calculate the averaged values;

- $\partial(f, g)|_{i(a)}$ is an estimate of the derivative of the field \mathcal{F} with respect to \mathcal{G} evaluated at $\tilde{g}_{i(a)}$. It is obvious that $i(b) = i(a)$, $\forall b \in A(i(a))$.

Relatively to the downscaling scheme (3), a fundamental question can now be raised: how far the guessed values are from the true values of the field \mathcal{F} ? We answer to the question in a specific case when the discrete values of the two fields are given by a space averaged method.

Let $f(x)$ and $g(x)$ be the values of the two fields at a point $x \in \Omega$. We introduce the space averaged quantities of functions f and g as

$$f_i = \frac{1}{m(\omega_i)} \int_{\omega_i} f(x) dx, \quad f_a = \frac{1}{m(\Delta_a)} \int_{\Delta_a} f(x) dx, \quad g_a = \frac{1}{m(\Delta_a)} \int_{\Delta_a} g(x) dx.$$

Assume that the two fields are functional dependent and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be the expression of this dependence

$$f(x) = h(g(x)), \quad \forall x \in \Omega. \quad (5)$$

Using a classical Taylor formula, one can write

$$f(x) = h(g(x)) = h(g_0) + h'(g_0)(g(x) - g_0) + \frac{1}{2}h''(\xi_x)(g(x) - g_0)^2, \quad (6)$$

where g_0 is an arbitrary number.

Let ω_i and $\Delta_a \subset \omega_i$ be fixed. One has

$$f_a = h(g_0) + h'(g_0)(g_a - g_0) + C(g, h'', \Delta_a). \quad (7)$$

The error between the true averaged values f_a and downscaled values f_a^* is given by

$$f_a^* - f_a = f_{i(a)} - h(g_0) + \left(\partial(f, g)|_{i(a)} (g_a - \tilde{g}_{i(a)}) - h'(g_0)(g_a - g_0) \right) - C(g, h'', \Delta_a).$$

If we choose $g_0 = \tilde{g}_{i(a)}$, we obtain

$$f_a^* - f_a = f_{i(a)} - h(\tilde{g}_{i(a)}) + \left((\partial(f, g)|_{i(a)} - h'(\tilde{g}_{i(a)}))(g_a - \tilde{g}_{i(a)}) \right) - C(g, h'', \Delta_a).$$

Taking into account that

$$f_{i(a)} = h(\tilde{g}_{i(a)}) + h'(\tilde{g}_{i(a)})(g_{i(a)} - \tilde{g}_{i(a)}) + C(g, h'', \omega_{i(a)})$$

and defining

$$g_{i(a)} := \tilde{g}_{i(a)},$$

one can write

$$\begin{aligned} f_{i(a)} &= h(\tilde{g}_{i(a)}) + C(g, h'', \omega_{i(a)}), \\ f_a^* - f_a &= \left(\partial(f, g)|_{i(a)} - h'(\tilde{g}_{i(a)}) \right) (g_a - \tilde{g}_{i(a)}) + C(g, h'', \omega_{i(a)}) - C(g, h'', \Delta_a), \end{aligned} \quad (8)$$

where

$$C(g, h'', \omega_{i(a)}) = \frac{1}{2} \int_{\omega_{i(a)}} h''(\xi_x)(g(x) - g_{i(a)})^2 dx.$$

The formula shows that there are three main sources of errors, namely:

- a) the nonlinearity of the functional relation,
- b) variation of the “pattern” function g ,
- c) the error of the approximation of derivative of the functional relation.

We summarize these facts in the following proposition.

Proposition 1 (Accuracy of FDS). *a) If the functional relation is a linear function and if the estimation of derivative is exact, then the downscaling scheme generates the exact space averaging value of the downscaled field*

$$f_a^* = f_a. \quad (9)$$

b) If the estimation of derivative is exact, then the error induced by the downscaling scheme is controlled by the second derivative of h and the variation of the “pattern” function

$$|f_a^* - f_a| \leq \xi(h, \omega_{i(a)}) \left(\frac{1}{m(\Delta_a)} + \frac{1}{m(\omega_{i(a)})} \right) \text{Var}(g, \omega_{i(a)}), \quad (10)$$

where

$$\xi(h, \omega) = \frac{1}{2} \sup_{x \in \omega} |h''(x)|, \quad \text{Var}(g, \omega_i) = \int_{\omega_i} (g(x) - g_i)^2 dx.$$

From the point of view of real world applications, the most important step is to estimate the derivative $|\partial(f, g)|$.

3 Upscaling Schemes

An upscaling scheme means to aggregate the small scale values of a field in order to define the behavior of the field at a larger scale. A common situation (known as the gridding problem) is when one knows the arbitrarily located observations and is interested to define the field on a regular grid. A possible way to solve the gridding problem is to firstly find an everywhere defined function that interpolates the scattered data and then use it to define the regular net values. The critical point of this scheme is the method of interpolation. There are no general rules to choose the best method to interpolate the scattered data. Besides the specificity of the data to be interpolated, a factor that affects the accuracy of the interpolation is the way the points are distributed in the domain. In this sense, there are certain theoretical results that show that an “equidistribution” of the points increases the accuracy of the interpolation, [7].

The problem we deal in this report can be formulated as follows.

Definition 1 (Scattered Data Problem). *Given the scatter data \mathcal{S} find a continuous function $\mathcal{Q}f : D \rightarrow \mathbb{R}$ such that:*

Exact interpolation problem(EI-SDP):

$$\mathcal{Q}f(\mathbf{x}_i) = z_i, \quad i = \overline{1, N}. \quad (11)$$

Quasi interpolation problem(QI-SDP):

$$\mathcal{Q}f(\mathbf{x}_i) \approx z_i, \quad i = \overline{1, N}. \quad (12)$$

In the case of noisy data, there are no reasons to perform an exact interpolation because all data are corrupted by noises. In such case, the quasi interpolation is a better solution for the scattered data problem. The quasi interpolation solution requires to clarify the approximation relation in the definition. For the mathematical aspects of the scattered data problem, the reader is referred to [8, 9], while for applications in environmental sciences we underline [10, 11, 12].

We introduce a method to solve scattered data problem suited to noisy data and (or) to high number of points in data. The method involves two functional spaces of approximation, the space generated by the radial basis functions and the space of spline wavelets. Formally, the approximation scheme read as

$$\mathcal{S} \xrightarrow{P} \mathcal{B} \xrightarrow{Q} \mathcal{W} \quad (13)$$

where \mathcal{B} is the space generated by the radial basis functions and \mathcal{W} is the space of the spline wavelet basis functions. The approximation of the support function f results as the composition of the operators P and Q :

$$\mathcal{Q}f = Q(P(\mathcal{S})). \quad (14)$$

In what follows, we will use the symbol $\stackrel{sdp}{=}$ to denote one of the two relations appearing in definition of the scattered data problem.

3.1 Thin Plate Spline Interpolation Method

In this section we present the construction of the operator $P : \mathcal{S} \rightarrow \mathcal{B}$. Since it can be thought as a solution of the scattered data problem, we will use the notation

$$\mathcal{Q}f = P(\mathcal{S}).$$

The thin plate spline (TPS) belongs to the family of radial basis functions (RBF). Let d be the dimension of the space \mathbb{R}^d and q be an integer number. Let $\mathbb{P}_{d;q} := \{p_a\}_{a=\overline{1, M}}$ be a basis set of polynomials $p_a : \mathbb{R}^d \rightarrow \mathbb{R}$ of degree less than q , where $M = C_{q+d-1}^d$. The thin plate spline generating function of order q is defined by

$$\phi(r) = \begin{cases} r^{2q-d}, & d \text{ odd,} \\ r^{2q-d} \log(r), & d \text{ even.} \end{cases} \quad (15)$$

Consider the family $\{\phi_k\}_{k=\overline{1, N}}$ of the translations of $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ defined by $\phi_k(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{x}_k\|)$. The interpolation of the scatter data \mathcal{S} is sought as a linear combination of the functions $\{\phi_k\}_{k=\overline{1, N}}$ and $\{p_a\}_{a=\overline{1, M}}$

$$\mathcal{Q}f(\mathbf{x}) = \sum_{k=1}^N \alpha_k \phi_k(\mathbf{x}) + \sum_{a=1}^M \beta_a p_a(\mathbf{x}). \quad (16)$$

The coefficients $\{\alpha_k\}_{k=\overline{1,N}}$ and $\{\beta_a\}_{a=\overline{1,M}}$ are determined by solving the linear equations

$$\begin{aligned} \sum_{k=1}^N \alpha_k \phi_k(\mathbf{x}_i) + \sum_{a=1}^M \beta_a p_a(\mathbf{x}_i) &= z_i, \quad i = \overline{1,N}, \\ \sum_{k=1}^N \alpha_k p_a(\mathbf{x}_k) &= 0, \quad a = \overline{1,M}, \quad M = C_{q+d-1}^d. \end{aligned} \quad (17)$$

We resume here two results concerning the resolvability of the system (17) and the accuracy of (16). For this, we first need to introduce two concepts concerning the set of points $\{\mathbf{x}_i\}_{i=\overline{1,N}}$.

Definition 2. We say that the set of points $\{\mathbf{x}_i\}_{i=\overline{1,N}}$ is $\mathbb{P}_{d,q}$ -nondegenerated if and only if 0 is the only polynomial from $\mathbb{P}_{d,q}$ that vanishes on $\{\mathbf{x}_i\}_{i=\overline{1,N}}$.

Definition 3. The density of the set $\{\mathbf{x}_j\}_{j=\overline{1,N}}$ in the domain D is defined by

$$h := \sup_{\mathbf{x} \in D} \min_{j=\overline{1,N}} \|\mathbf{x} - \mathbf{x}_j\|$$

Theorem 1. [7, 8, 9, 14]. Let D be a bounded domain from \mathbb{R}^d , $\{\mathbf{x}_i\}_{i=\overline{1,N}}$ a set of $\mathbb{P}_{d,q}$ -nondegenerated points from D , $\mathcal{Q}f(\mathbf{x})$ the interpolating function defined by (16) and (15) that solves the scattered data problem (11), and h the density of the point locations in D . Then:

(a) the system of equations (17) is uniquely solvable.

(b)

$$\|\mathcal{Q}f - f\|_{\infty, D} \leq C \begin{cases} h\sqrt{\log(h^{-1})}, & 2q - d = 2, \\ \sqrt{h}, & 2q - d = 1, \\ h, & \text{otherwise.} \end{cases}$$

From a practical point of view, the problem of the scattered data points resumes to solve the linear system of equations (17). We accomplish this task using an algorithm adapted from [6]. The algorithm involves a threshold value μ . The solution solves EI-SDP if $\mu = 0$ and QI-SDP if $\mu > 0$.

Algorithm [Low Rank Smoother (LRS)]

Step 1: Input the order q of the thin plate spline and set a value $\mu \geq 0$.

Step 2: Read the data $\{\mathbf{x}_j, z_j\}_{j=\overline{1,N}}$.

Step 3: Build the distance matrix $E_{ij} := \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $i, j = \overline{1,N}$.

Step 4: Build the matrix $T_{ia} := p_a(\mathbf{x}_i)$, $i = \overline{1,N}$, $a = \overline{1,M}$.

Step 5: Solve the minimization problem

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \min_{(\boldsymbol{\delta}, \boldsymbol{\eta})} \|\mathbf{y} - \mathbf{E}\boldsymbol{\delta} - \mathbf{T}\boldsymbol{\eta}\|^2 \quad (18)$$

subject to the constraints

$$\mathbf{T}^T \boldsymbol{\delta} = 0. \quad (19)$$

Solving the minimization problem (18, 19)

Step 5.1: Find the singular value decomposition $\mathbf{E} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

Step 5.2: Find the level k such that $\sigma_k > \mu$ and build up the matrices $\mathbf{\Sigma}_k = \text{diag}(\sigma_1, \dots, \sigma_k)$, $\mathbf{V}_k = \text{col}(\mathbf{v}_1, \dots, \mathbf{v}_k)$.

Step 5.3: Find the QR decomposition $\mathbf{V}_k^T \mathbf{T} = \mathbf{Q}\mathbf{R}$.

Step 5.4: Define $\mathbf{Z}_k := \text{col}(\mathbf{q}_{M+1}, \dots, \mathbf{q}_k)$; observe that

$$\mathbf{T}^T \mathbf{Z}_k \boldsymbol{\delta}_k = 0, \quad \forall \boldsymbol{\delta}_k \in \mathbb{R}^{M-k}.$$

Step 5.5: Find the minimum of the unconstrained problem

$$\min_{(\boldsymbol{\delta}_k, \boldsymbol{\eta})} \|\mathbf{y} - \mathbf{E}_k \boldsymbol{\delta}_k - \mathbf{T}\boldsymbol{\eta}\|^2. \quad (20)$$

where $\mathbf{E}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{Z}_k$.

Solving the minimization problem (20)

Step 5.5.1: Build up the matrix $\mathbf{A} = (\mathbf{E}_k | \mathbf{T})$.

Step 5.5.2: The solution of (20) is

$$\begin{pmatrix} \boldsymbol{\delta}_k \\ \boldsymbol{\eta}_k \end{pmatrix} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{z}.$$

Step 6: The solution to the interpolation problem is

$$\boldsymbol{\alpha} = \mathbf{U}_k \mathbf{Z}_k \boldsymbol{\delta}_k, \quad \boldsymbol{\beta} = \boldsymbol{\eta}_k.$$

Comments

1. If $\mu = 0$, then one deals with standard thin plate spline interpolation problem.
2. The method is suitable for “small” number of data, our experience indicates $N \leq 1000$.
3. For large number of data observations, the algorithm becomes time consuming or breaks down.
4. The selection of the threshold value is a crucial step in the defining a “good” interpolating. There is no general method to provide the best values for μ .

3.2 Cubic Spline Approximation of Scatter Data

We remember that our method to solve the scatter data problem involves the operators $P : \mathcal{S} \rightarrow \mathcal{B}$ and $Q : \mathcal{B} \rightarrow \mathcal{W}$ and the solution of the problem is given by the composition $Q \cdot P$. In the section devoted to thin plate interpolation we set up a method to define the operator P . The method works very well when there is a small number of the points in data. If the number of points is very high then the method has no practical use. To surmount such difficulty, our suggestion is to partition the scatter data and then apply the (LRS) algorithm on each element of this partition. We define the $P : \mathcal{S} \rightarrow \mathcal{B}$ in two steps:

Step 1 Define a partition $\{\mathcal{S}_\alpha\}_{\alpha=1}^K$ of the scatter data \mathcal{S}

$$\begin{aligned}
D &= \bigcup_{\alpha=1}^K D_\alpha, \\
\mathcal{P}_\alpha &= \{P_i | P_i \in D_\alpha\}, \quad \mathcal{P} = \bigcup_{\alpha=1}^K \mathcal{P}_\alpha \\
\mathcal{I}_\alpha &= \{i | P_i \in D_\alpha\} \\
\mathcal{S}_\alpha &= \{(P_i, z_i) | i \in \mathcal{I}_\alpha\} \\
M_\alpha &= \#\mathcal{P}_\alpha
\end{aligned} \tag{21}$$

Step 2 For each element of the partition, one constructs an interpolating function $\mathcal{Q}^I f_\alpha$ of the scatter data \mathcal{S}_α and define the interpolating $\mathcal{Q}^I f(\mathbf{x})$ by

$$\begin{aligned}
\mathcal{Q}^I f_\alpha(\mathbf{x}_i) &\stackrel{sdp}{=} z_i, \quad \forall i \in \mathcal{I}_\alpha \\
\mathcal{Q}^I f(\mathbf{x}) &= \sum_{\alpha=1}^K \mathcal{Q}^I f_\alpha(\mathbf{x}) \mathbf{1}_\alpha(\mathbf{x}).
\end{aligned} \tag{22}$$

The function $\mathcal{Q}^I f$ is a continuous function inside each D_α , but not on the entire D : it may have jumps on the boundary of D_α .

3.2.1 Multiresolution Analysis (MRA) of Scatter Data

In what follows, we consider that $D = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. The coordinates of a point \mathbf{x}_i will be denoted by (x_i, y_i) . The cubic spline multiresolution analysis of the space $\mathbb{L}^2([0, 1])$ introduced by Chui and Quack [16] consists in a set of closed finite dimensional subspaces $\mathbf{V}^j([0, 1])$ and $\mathbf{W}^j([0, 1])$, with $j \in \{j_0, j_0 + 1, \dots\}$ ($j_0 \geq 3$) that exhibit the following properties:

1. $\mathbf{V}^j([0, 1]) \subset \mathbf{V}^{j+1}([0, 1])$
2. $\overline{\bigcup_{j=j_0}^{\infty} \mathbf{V}^j([0, 1])} = \mathbb{L}^2([0, 1])$
3. $\mathbf{V}^{j+1}([0, 1]) = \mathbf{V}^j([0, 1]) \oplus \mathbf{W}^j([0, 1])$
4. $\mathbf{V}^{j_0}([0, 1]) \oplus_{j=j_0}^{\infty} \mathbf{W}^j([0, 1]) = \mathbb{L}^2([0, 1])$

The scaling spaces $\mathbf{V}^j([0, 1])$ and the wavelet spaces $\mathbf{W}^j([0, 1])$ are generated by the spline functions $\{\varphi_k^j(x)\}_k$ and $\{\psi_k^j(x)\}_k$, respectively. One constructs a two dimensional cubic spline space defined on $[0, 1] \times [0, 1]$ by taking the tensor product of one dimensional space of cubic spline defined on $[0, 1]$

$$\mathbf{V}\mathbf{V}^j([0, 1]^2) = \mathbf{V}^j([0, 1]) \otimes \mathbf{V}^j([0, 1]).$$

A basis in this space is given by the functions $\varphi_{k,l}^j(x, y)$ defined by the tensor product

$$\varphi_{k,l}^j(x, y) = \varphi_k^j(x) \varphi_l^j(y).$$

By multiresolution analysis of a function $g \in \mathbb{L}^2([0, 1])$, one means to find the decomposition

$$g(x, y) = \sum_{k,l} a_{k,l}^{j_0} \varphi_{k,l}^{j_0}(x, y) + \sum_{j=j_0}^{\infty} \left(\sum_{k,l} b_{k,l}^j \varphi_k^j(x) \psi_l^j(y) + \sum_{k,l} c_{k,l}^j \psi_k^j(x) \varphi_l^j(y) + \sum_{k,l} e_{k,l}^j \psi_k^j(x) \psi_l^j(y) \right). \quad (23)$$

For practical purposes, the infinite sum is truncated at a high level. In the Appendix, we provide more details about the way one can perform a multiresolution analysis.

To find the solution of the scatter data problem in a space $\mathbf{V}\mathbf{V}^J([0, 1]^2)$, one must find the coefficients $\{a_{k,l}^J\}_{k,l}$ such that

$$\sum_{k,l} a_{k,l}^J \varphi_{k,l}^J(x_i, y_i) \stackrel{sdp}{=} z_i, \forall i = 1, N.$$

The method we purpose to solve QI-SDP in a space $\mathbf{V}\mathbf{V}^J([0, 1]^2)$ consist of:

Multiresolution Analysis of Scatter Data (MRA-SD)

- a. Choose a resolution level J and define the knots:

$$\tilde{x}_i^J = \frac{i}{2^J}, \quad \tilde{y}_l^J = \frac{l}{2^J}, \quad i, l = \overline{0, 2^J}.$$

- b. Use the function $\mathcal{Q}^I f$ given by (22) and the operator P_S^J given by (31) to evaluate

$$a_{k,l}^J = \sum_{k,l} P_{S_{k,m}}^J P_{S_{l,n}}^J \mathcal{Q}^I f(\tilde{x}_m^J, \tilde{y}_n^J). \quad (24)$$

- c. Define the solution of the quasi interpolation problem by

$$\mathcal{Q}f^J(x, y) = \sum_{k,l} a_{k,l}^J \varphi_{k,l}^J(x, y). \quad (25)$$

Theorem 2 (Quasinterpolant property). *If the interpolating $\mathcal{Q}^I f$ is defined by the thin-plate spline RBF functions, then $\mathcal{Q}f^J$ reproduces the polynomial of the total degree three for any partition $\{\mathcal{S}_\alpha\}_{\alpha=1}^K$ and for any level of resolution J .*

4 Conclusions

We have used two different methods to analyze the environmental data, remote sensing and scatter data interpolation. We focused our attention on some mathematical properties of the methods that are important from a practical application point of view. We find out that the two methods are sufficiently flexible to cope with the environmental data applications.

As further research, we would like to couple scatter data interpolations with remote sensing data analysis in order to increase the accuracy and to validate each of these two methods.

5 Appendix

The basis for the spaces $\mathbf{V}^j([0, 1])$ can be introduced by means of the spline functions [15, 16, 17]:

$$\begin{aligned} N(x) &:= \frac{1}{6} \sum_{l=0}^4 (-1)^l C_4^l (x-l)_+^3, \\ B_k(x) &:= \frac{(-1)^{4+k}}{(3+k)!} \sum_{l=1}^{4+k} (-1)^l l^k C_{4+k}^l (x-l)_+^3, \quad k \in \{-3, -2, -1\}, \end{aligned} \quad (26)$$

where $x_+ := \max\{x, 0\}$.

The wavelet generating function is a spline function of order four

$$\psi(x) := \sum_{l=0}^{10} q_l N(2x-l), \quad (27)$$

where

$$q_l = \frac{(-1)^l}{2^{4-l}} \sum_{n=0}^4 C_4^n N(l-n+1).$$

The scaling base functions of $\mathbf{V}^j([0, 1])$ are given by

$$\begin{aligned} \varphi_k^j(x) &:= N(2^j x - k), \quad k \in \{0, 1, \dots, 2^j - 4\}, \\ \varphi_k^j(x) &:= B_k(2^j x), \quad k \in \{-3, -2, -1\}, \\ \varphi_k^j(x) &:= \varphi_{k-2^j}^j(1-x), \quad k \in \{2^j - 3, 2^j - 2, 2^j - 1\}. \end{aligned} \quad (28)$$

The wavelet space $\mathbf{W}^j([0, 1])$ is the orthogonal space of $\mathbf{V}^j([0, 1])$ in the space $\mathbf{V}^{j+1}([0, 1])$

$$\mathbf{V}^{j+1}([0, 1]) = \mathbf{V}^j([0, 1]) \oplus \mathbf{W}^j([0, 1]). \quad (29)$$

One has the following scaling relations:

$$\begin{aligned} \varphi_k^j(x) &= \sum_{l=-3}^{2^{j+1}-1} p_{k,l} \varphi_l^{j+1}(x), \quad k = \overline{-3, \dots, 2^j - 1}, \\ \psi_k^j(x) &= \sum_{l=-3}^{2^{j+1}-1} q_{k,l} \varphi_l^{j+1}(x), \quad k = \overline{-3, \dots, 2^j - 4}, \end{aligned} \quad (30)$$

where $p_{\cdot, \cdot}$ and $q_{\cdot, \cdot}$ are band matrices, see [17].

The cubic spline space of the functions defined on $[0, 1] \times [0, 1]$ at level j is defined by the tensor product

$$\mathbf{V}\mathbf{V}^j([0, 1]^2) = \mathbf{V}^j([0, 1]) \otimes \mathbf{V}^j([0, 1]).$$

The basis of product scaling functions in $\mathbf{V}\mathbf{V}^j([0, 1]^2)$ is defined by

$$\varphi_{k,l}^j(x, y) = \varphi_k^j(x) \varphi_l^j(y), \quad k = \overline{-3, 2^j - 1}, \quad l = \overline{-3, 2^j - 1}$$

and thus, an element $s \in \mathbf{V}\mathbf{V}^j([0, 1]^2)$ can be written as

$$s = \sum_{k,l} c_{k,l} \varphi_{k,l}^j(x, y).$$

A space $\mathbf{V}\mathbf{V}^{j+1}([0, 1]^2)$ admits an orthogonal decomposition

$$\mathbf{V}\mathbf{V}^{j+1}([0, 1]^2) = \mathbf{V}\mathbf{V}^j([0, 1]^2) \oplus \mathbf{V}\mathbf{W}^j([0, 1]^2) \oplus \mathbf{W}\mathbf{V}^j([0, 1]^2) \oplus \mathbf{W}\mathbf{W}^j([0, 1]^2),$$

where the wavelet subspace are generated by

$$\begin{aligned} \mathbf{V}\mathbf{W}^j([0, 1]^2) &:= \text{span}\{\varphi_k^j(x)\psi_l^j(y)\}_{-3, -3}^{2^j-1, 2^j-4} \\ \mathbf{W}\mathbf{V}^j([0, 1]^2) &:= \text{span}\{\psi_l^j(x)\varphi_k^j(y)\}_{-3, -3}^{2^j-4, 2^j-1} \\ \mathbf{W}\mathbf{W}^j([0, 1]^2) &:= \text{span}\{\psi_l^j(x)\psi_k^j(y)\}_{-3, -3}^{2^j-4, 2^j-4}. \end{aligned}$$

Given a space $\mathbf{V}\mathbf{V}^N([0, 1]^2)$, the multiresolution analysis of it means to decompose it as

$$\mathbf{V}\mathbf{V}^N([0, 1]^2) = \mathbf{V}\mathbf{V}^{j_0}([0, 1]^2) \bigoplus_{j=j_0}^{N-1} (\mathbf{V}\mathbf{W}^j([0, 1]^2) \oplus \mathbf{W}\mathbf{V}^j([0, 1]^2) \oplus \mathbf{W}\mathbf{W}^j([0, 1]^2))$$

The deconstruction algorithm. It is used to find the wavelet components in the multiresolution analysis of a function $f \in \mathbf{V}\mathbf{V}^N([0, 1]^2)$. The algorithm uses the dual components of the function, [17]

$$\begin{aligned} \sum_{k,l} \tilde{a}_{k,l}^{j+1} \tilde{\varphi}_k^{j+1}(x) \tilde{\varphi}_l^{j+1}(y) &= \sum_{k,l} \tilde{a}_{k,l}^j \tilde{\varphi}_k^j(x) \tilde{\varphi}_l^j(y) + \sum_{k,l} \tilde{b}_{k,l}^j \tilde{\varphi}_k^j(x) \tilde{\psi}_l^j(y) + \\ &+ \sum_{k,l} \tilde{c}_{k,l}^j \tilde{\psi}_k^j(x) \tilde{\varphi}_l^j(y) + \sum_{k,l} \tilde{e}_{k,l}^j \tilde{\psi}_k^j(x) \tilde{\psi}_l^j(y) \\ \tilde{a}_{m,n}^j &= \sum_{k,l} \tilde{a}_{k,l}^{j+1} p_{k,m} p_{l,n} \\ \tilde{b}_{m,n}^j &= \sum_{k,l} \tilde{a}_{k,l}^{j+1} p_{k,m} q_{l,n} \\ \tilde{c}_{m,n}^j &= \sum_{k,l} \tilde{a}_{k,l}^{j+1} q_{k,m} p_{l,n} \\ \tilde{e}_{m,n}^j &= \sum_{k,l} \tilde{a}_{k,l}^{j+1} q_{k,m} q_{l,n}, \end{aligned}$$

where $\tilde{\varphi}_k^j$ and $\tilde{\psi}_k^j$ are the dual scaling and wavelet functions to φ_k^j and ψ_k^j in $\mathbf{V}^j([0, 1])$ and $\mathbf{W}^j([0, 1])$, respectively. The coefficients with $\tilde{\cdot}$ are the coefficients from the dual representation.

The reconstruction algorithm. It allows one to find the scaling component at a higher level N by using the wavelet components and the scaling coefficients at the lower level.

$$\begin{aligned} \sum_{k,l} a_{k,l}^{j+1} \varphi_k^{j+1}(x) \varphi_l^{j+1}(y) &= \sum_{k,l} a_{k,l}^j \varphi_k^j(x) \varphi_l^j(y) + \sum_{k,l} b_{k,l}^j \varphi_k^j(x) \psi_l^j(y) + \\ &+ \sum_{k,l} c_{k,l}^j \psi_k^j(x) \varphi_l^j(y) + \sum_{k,l} e_{k,l}^j \psi_k^j(x) \psi_l^j(y) \end{aligned}$$

$$a_{m,n}^{j+1} = \sum_{k,l} a_{k,l}^j p_{k,m} p_{l,n} + \sum_{k,l} b_{k,l}^j p_{k,m} q_{l,n} + \sum_{k,l} c_{k,l}^j q_{k,m} p_{l,n} + \sum_{k,l} e_{k,l}^j q_{k,m} q_{l,n}.$$

The duality relation is given by

$$\begin{aligned}\tilde{a}_{m,n}^j &= a_{k,l}^j C_{k,m}^j C_{l,n}^j, \\ \tilde{b}_{m,n}^j &= b_{k,l}^j C_{k,m}^j D_{l,n}^j, \\ \tilde{c}_{m,n}^j &= c_{k,l}^j D_{k,m}^j C_{l,n}^j, \\ \tilde{e}_{m,n}^j &= a_{k,l}^j D_{k,m}^j D_{l,n}^j,\end{aligned}$$

where

$$\begin{aligned}C_{k,m}^j &= \int_0^1 \varphi_k^j(x) \varphi_m^j(x) dx, \\ D_{k,m}^j &= \int_0^1 \psi_k^j(x) \psi_m^j(x) dx.\end{aligned}$$

Given a function in $\mathbf{VV}^N([0,1]^2)$, there is a decomposition of it as the orthogonal sum of the scaling components at a low level j_0 and the wavelet components,

$$\begin{aligned}f(x, y) &= \sum_{k,l} a_{k,l}^{j_0} \varphi_{k,l}^{j_0}(x, y) + \\ &+ \sum_{j=j_0}^{N-1} \left(\sum_{k,l} b_{k,l}^j \varphi_k^j(x) \psi_l^j(y) + \sum_{k,l} c_{k,l}^j \psi_k^j(x) \varphi_l^j(y) + \sum_{k,l} e_{k,l}^j \psi_k^j(x) \psi_l^j(y) \right).\end{aligned}$$

A function

$$f(x, y) : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$$

can be approximated by a function of the space $\mathbf{VV}^N([0,1]^2)$ using spline interpolation with knots (x_k, y_l) , $x_k = k/2^N$, $y_l = l/2^N$, $k, l = 0, 2^N$:

$$f_W^N(x, y) = \sum_{k,l} a_{k,l}^N(\mathbf{f}) \varphi_{k,l}^N(x, y).$$

The coefficients $a_{k,l}^N(\cdot)$ can be expressed as

$$a_{k,l}^N = \sum_{m,n} P_{k,m}^N P_{l,n}^N f(x_m, y_n).$$

The ‘‘proctor’’ P^N can be determined such that $f_W^N(x, y)$ interpolates the function f and reproduces the cubic polynomial

$$\begin{aligned}f_W^N(x_k, y_l) &= f(x_k, y_l) \\ f_W^N(x, y) &\equiv f(x, y), \quad \text{if } f \in \pi_3.\end{aligned}$$

A simplified version of P is given in

$$P_S^N = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 & \dots & 0 \\ 7/18 & 18/18 & -9/18 & 2/18 & 0 & \dots & 0 \\ -1/6 & 8/6 & -1/6 & 0 & 0 & \dots & 0 \\ 0 & -1/6 & 8/6 & -1/6 & 0 & \dots & 0 \\ & & & \ddots & & & \\ 0 & \dots & 0 & -1/6 & 8/6 & -1/6 & 0 \\ 0 & \dots & 0 & 0 & -1/6 & 8/6 & -1/6 \\ 0 & \dots & 0 & 2/18 & -9/18 & 18/18 & 7/18 \\ 0 & \dots & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (31)$$

For P_S^N given by (31), the approximation function $f_W^N(x, y)$ loses the interpolating property but still reproduces the cubic polynomial.

References

- [1] Watson D.J., *Comparative physiological studies in the growth of field crops.I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years*, Annals of Botany, **11**, 1947, pp. 41–76.
- [2] Li Shen, Zhaoqin Li, Xulin Guo, *Remote Sensing of Leaf Area Index (LAI) and a Spatiotemporally Parameterized*, International Journal of Applied Science and Technology, **4**(1), 2014,46–61.
- [3] Andreas Colliander *et al.*, *Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI during SMAPVEX15*, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, VOL. 14, NO. 11, NOVEMBER 2017
- [4] Merlin, O., J. P. Walker, A. Chehbouni, and Y. Kerr (2008a), *Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil :q evaporative efficiency*, Remote Sens. Environ., **112**(10), 3935–3946.
- [5] Peng,J., A. Loew, O. Merlin, and N. E. C. Verhoest, *A review of spatial downscaling of satellite remotely sensed soil moisture*, Rev. Geophys., **55**, doi:10.1002/2016RG000543.
- [6] S. N. Wood, *Thin plate regression spline*, J.R. Statist. Soc. B(2003), **65**, Part.1 , pp. 95-114.
- [7] M. J. D. Powell, *The uniform convergence of thin-plate spline interpolation in two dimensions*, Numer. Math. **67**(1994), 107–128.
- [8] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx. **1**(1986), 11–22.
- [9] Z. Wu and R. Schaback (1993), *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal. **13**(1993), 13–27.

-
- [10] M. F. Hutchinson (1995), *Interpolating mean rainfall using thin plate smoothing splines*, International Journal of Geographical Information Systems, 9:4, 385-403, DOI: 10.1080/02693799508902045
- [11] Nynke Hofstra, Malcolm Haylock, Mark New, Phil Jones, and Christoph Frei, *Comparison of six methods for the interpolation of daily, European climate data*, JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D21110, doi:10.1029/2008JD010100, 2008
- [12] D. S. Trossman, L. Thompson, and S. L. Hautala (2011), *Application of Thin-Plate Splines in Two Dimensions to Oceanographic Tracer Data*, J. Atmos. Oceanic Technol., 1522–1538, 28.
- [13] A. Bejancu (1997), *The uniform convergence of multivariate natural spline*, DAMTP, Technical Report, University of Cambridge.
- [14] M.D. Buhmann (2000), *Radial basis functions*, Acta numerica, 1-38.
- [15] Charles K. Chui and Jian-Zhong Wang, *A Cardinal Spline Approach to Wavelets*, Proceedings of the American Mathematical Society, **113**(3), 1994, 785–793.
- [16] C. K. Chui and E. Quak, *Wavelets on a Bounded Interval*, in Numerical Methods of Approximation Theory, vol. 9, eds. D. Braess and L.L. Schumaker, pp. 53-75, Birkhauser, Basel, 1992.
- [17] Ewald Quak and Norman Weyrich, *Decomposition and Reconstruction Algorithms for Spline Wavelets on a Bounded Interval*, Applied and Computational Harmonic Analysis, **1**, 1994, 217–231.
- [18] S. Ion and D. Marinescu, *Spline Wavelets Analysis of Reticulated Function on Bounded Interval*, Math. Rep., Tom 54, vol. 4, No.2, 2002.