Heterogeneously hierarchical approximation with compactly supported Radial basis functions

International Conference of Kernel-Based Approximation Methods in Machine Learning

May 19-21, 2017
South China Normal University
Guang Zhou

Shengxin.Zhu(at)xjtlu.edu.cn

Department of Mathematics

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Xi'an Jiaotong-Liverpool University（XJTLU)

XJTLU only
11 years old

- **September 2004: Agreement signed**
- **May 2006: Licenced by MoE，160+ first cohort students**
- **Over 12k registered students, near 9,000 on XJTLU campus (including 570+ international students, 525 master students, and 113 PhD students)**

➢ **Over 70 undergraduate and postgraduate degrees: STEM, Architecture, Business, social & humanity**

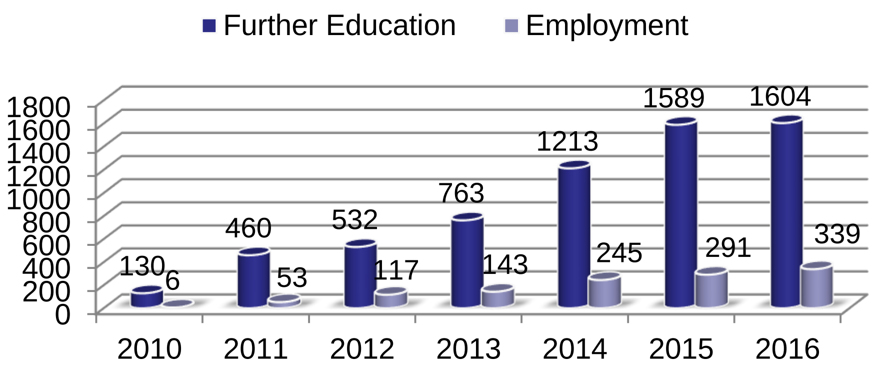
➢ **Dual Degrees from XJTLU & UoL**
   - **7th UG cohort graduate in July 2015/16**
   - **MSc graduate since January 2014**
   - **PhD graduate since June 2014**

- **Plan: 18k registered students, 15k (2-3k overseas, 2-3k PG) students on campus**

Graduates statistics:

from ordinary to extraordinary

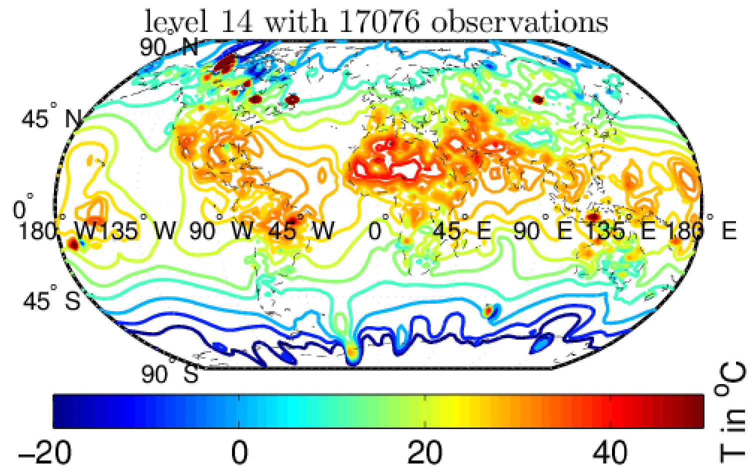
2010-2015 cohorts of graduates



Year	2010	2011	2012	2013	2014	2015	2016
Further Education	130	460	532	763	1213	1589	1604
Employment	6	53	117	143	245	291	339
Undetermined	-	-	-	-	26	28	39
Total	136	513	649	906	1484	1908	1982

- **Around 80% entered world class universities to continue postgraduate studies**
- **10-15%** entered **the world top10** universities.
- More than **50%** entered **the world top 100** universities.
- Above 10% entered market and many of the graduates are **employed by the world 500 companies**
- Imperial College (EEE) announced: **For Chinese applicants studying in China**, only consider applications from students studying: **C9 or 985, or XJTLU.**

Outline: CSRBFs why and how?



- Introduction and background
- Compactly supported RBFs
- Motivation for
  - hierarchical approach
  - heterogeneously hierarchical approach
- How
  - Interpolation with CSRBFs with different shape.
  - Heterogeneously hierarchical approach
- Numerical examples

Heterogeneously hierarchical is referred to as using RBFs with different scales/shapes on the same level in a ML approach

Introduction and background Ⅰ

- RBFs： many applications by our speakers （& Poggo &Girosi 1990, Regularization algorithms for learning that are equivalent to multilayer networks, *Science*)
- Why RBF so popular? Micchelli’s work 1986 for solvability for Multiquadrics (MQs) and a class of RBFs （Mairhuber-Curtis theorem 1956, Haar 1917)
- MQs is still the best globally RBFs for most cases (Used by B. Hon, Ling)
- Why MQ so effective (brief explanation): (Sandwell,1987).
- Biharmonic splines $r^3\ in\ R$, $r^2(\log r\ -1)$ in R^2 （Matlab griddata v4）, r in R^3

$$\lim_{c\to 0}\sqrt{c^2+r^2}=r$$

Xi'an Jiaotong-Liverpool University 西交利物浦大学

Introduction and background II：CSRBFs

1. Sparsity/scalability
   - Askey's power functions $\max(0,(1-||x||)^{\ell})$
   - Truncated Gaussian
   - Easy to construct
2. Positivity/Conditioning Positivity: Wu & Wendland's contribution
   - Wu functions (Wu1995)
   - Wendland functions(Wendland 1995, only 4 pages !)
   - Missing Wendland functions(Shaback 2011)
   - Unified by the associated Legendre functions of the 1st kind and the Gamma functions (S. Hubbert 2012)
   - Relationship between Wendland functions and Gaussian (Chernih, Sloan & Womersley, 2014)
   - Difficult & painful to read; deep & beautiful mathematics

Introduction and background II：CSRBFs

3. Accuracy: Positive definite CSRBFs matters

- Wu, 1997, positive definite CSRBFs violated the Strang-Fix(SF) condition (next slide).
- The SF condition is a necessary and sufficient condition for convergence.
- SF condition: convergence requires the interplant reproduce polynomials.
- There is no positive definite CSRBFs which satisfy the SF in R^d for $d \geq 2$ (Wu 1997)
- Positive definite functions can not reproduce a constant functions (Gaussian can't not recover a constant function Buhmann 1990)

Introduction and background II: CSRBFs

The Strang-Fix Condition (1971):

For any $p \geq q \geq 0$, the following condition are equivalent:

1. φ lies in H_c^q, $\hat{\phi}(0) \neq 0$, but has zeros of order at least $p+1$ at other points of $2\pi Z^n$.
2. φ lies in H_c^q, and for $|\alpha| \leq p$, the function $\sum j^\alpha \varphi(t-j)$ is a polynomial in $t_1, \ldots, t_d$ with leading term Ct^α, $C \neq 0$.
3. φ is a distribution with compact support, and for each u in H^{p+1} there are weights w_j^h such that $h \to 0$

$$||u - \sum w_j^h \varphi_j^h||_{H^s} \leq c_s h^{p+1-s} ||u||_{H^{p+1}} \ , s \leq p$$

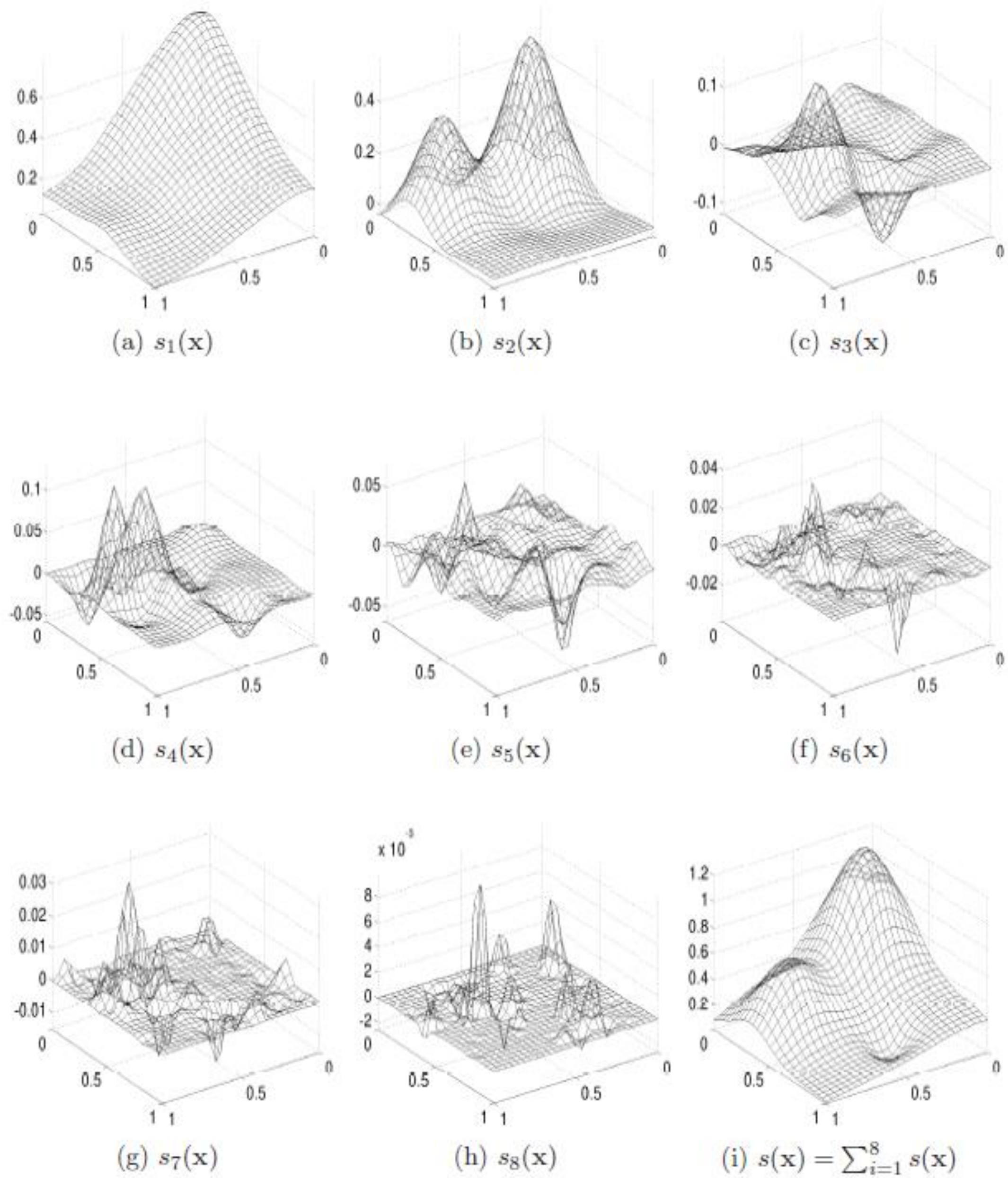
and $\sum \left|w_j^h\right|^2 \leq K ||u||_{H^0}^2$

G. Strang and G.Fix. A Fourier analysis of the finite element variational method, 1971
Also: I.J. Schoenberg 1946 seminal paper on cardinal splines

Motivation for multi-step and multi-scale

- Approximation quality for CSRBFs is bad. (any ?).
  1. The native space for pdf is small
  2. How to interpret Wendland's convergence results for positive definite CSRBFs?
     - Based on Wu's IMA paper, use positive definite CSRBFs as conditionally positive definite ones
     - Wendland states in his book *Scattered Data Approximation: keeping the support fixed and refine the interpolation points*
     - *Converge in which norm? is the convergence we want?*

Multistep or multiscale approach is one of several ways to improve approximation quality for positive definite CSRBFs.



(a) $s_1(\mathbf{x})$ (b) $s_2(\mathbf{x})$ (c) $s_3(\mathbf{x})$

(d) $s_4(\mathbf{x})$ (e) $s_5(\mathbf{x})$ (f) $s_6(\mathbf{x})$

(g) $s_7(\mathbf{x})$ (h) $s_8(\mathbf{x})$ (i) $s(\mathbf{x}) = \sum_{i=1}^{8} s(\mathbf{x})$

Overview of multi-step or multiscale method

$X_1 \subset \cdots \subset X_\ell = X, r_0 = f$

For i=1 to ℓ do

$$s_i|_{X_i} = r_{i-1}|_{X_i}$$

$$r_i = r_{i-1} - s_i|_X$$

endFor

$$S(x) = \sum_{1}^{\ell} s_\ell(x)$$

$$s(x) = \sum_{i=1}^{\ell} \sup|s_i(x)| \frac{s_i(x)}{\sup|s_i(x)|}$$

Frequency → Frequencies

Motivation for multi-step and multi-scale II

- Floater and Iske 1996 introduced multistep method
- Multi-level *Stationary* does not converge (Fasshauer's book(2007), p.277)
  Notes: stationary, keeps the support proportional to the fill distance h
- Fasshauer & Jerome 1999, Multistep approximation algorithms: improved convergence rates through postconditioning with smooth kernels
- Le Gia & Sloan & Wendland 2010 set up multiscale analysis in Sobolev spaces on sphere
- Alex Townsend 2012 ; Patricio Fawell 2014
- After this we can say CSRBF useful.

So what's next problem?

Motivation for heterogeneously hierarchal

- Multi-level stationary approach
  - **have a good scability, storage $\boldsymbol{O(N)}$**
  - **Does not converge**
- Multi-level non-stationary approach
  - **Converge**
  - **Scarify scability, the bandwith of the interpolation matrix increases as the data set are refined.**
  - **Storage $\boldsymbol{O(N^{2-\delta})}$, $\boldsymbol{0 \leq \delta < 1}$.**

Can we balance these two approaches?

Heterogeneously hierarchal RBFs

HHRBF 3 components

- Using different shapes on each level
- Fast Hierarchical data sets construction
- Fast sparse kernel summations

HHRBF: Convexity and solvability of CSRBFS with different shapes (Zhu & Wathen 2014)

Theorem on convexity (Z&W 2014) For any Wendland function $\phi_{d,k}$, there exists a positive real number $q \in [0,1)$, such that $\phi_{d,k}$ is convex on $[q, 1]$.

Proof requires another two lemma

THM (Z&W 2014 For any Wendland function $\phi_{d,k}(r)$, there exists a positive number m and c , such that for $\phi_{d,k}(r) \leq c(1-r)^m$

Proof trial

Estimation of the convex interval for Wendland functions

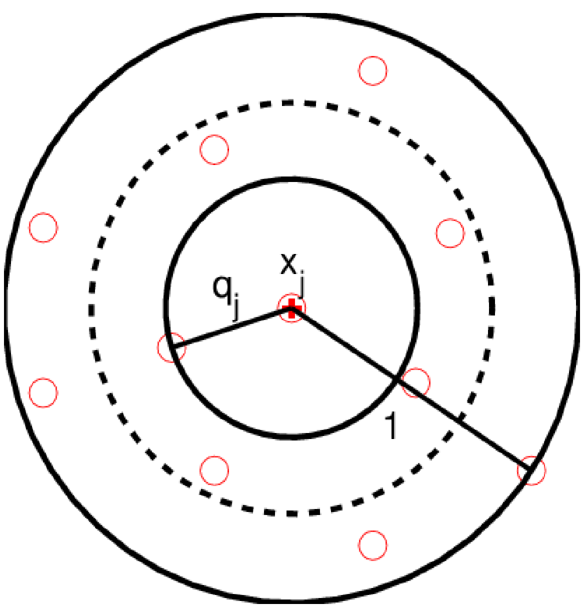
Table 1. Estimations of the convex interval for Wendland functions

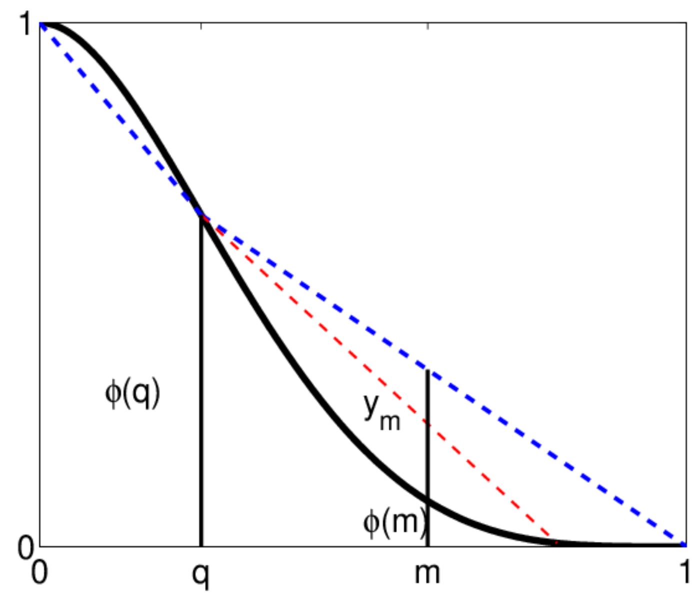
Smoothness	Function	convex interval
C^0	$\phi_{1,0} = (1-r)_+$	$[0,1]$
C^2	$\phi_{1,1} = (1-r)_+^3(3r+1)$	$[\frac{1}{3},1]$
C^4	$\phi_{1,2} = (1-r)_+^5(8r^2+5r+1)$	$[0.2760,1]$
C^0	$\phi_{3,0} = (1-r)_+^2$	$[0,1]$
C^2	$\phi_{3,1} = (1-r)_+^4(4r+1)$	$[\frac{1}{4},1]$
C^4	$\phi_{3,2} = (1-r)_+^6(35r^2+18r+3)$	$[0.2356,1]$
C^0	$\phi_{5,0} = (1-r)_+^3$	$[0,1]$
C^2	$\phi_{5,1} = (1-r)_+^5(5r+1)$	$[\frac{1}{5},1]$
C^4	$\phi_{5,2} = (1-r)_+^7(16r^2+7r+1)$	$[0.2056,1]$

HHRBF: Convexity and solvability of CSRBFS with different shapes (Zhu & Wathen 2014)

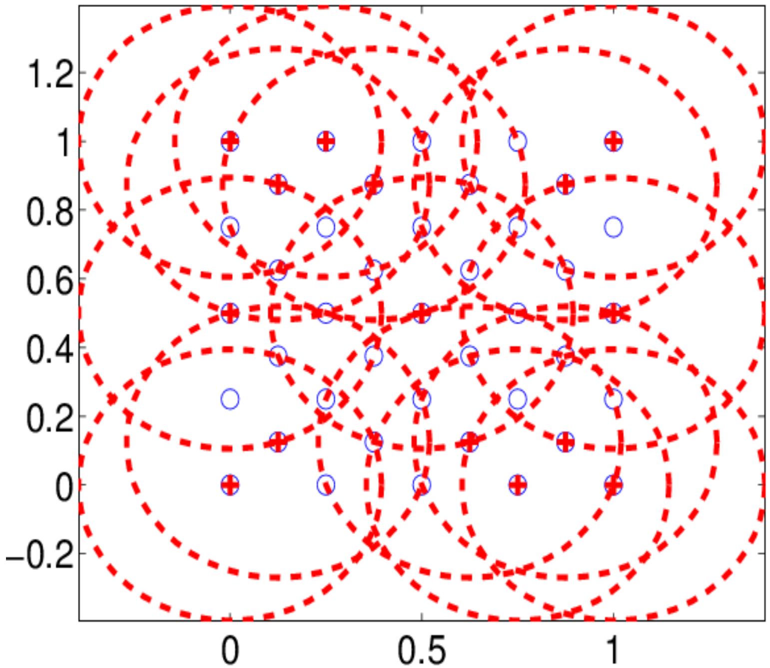
by the convexity and lemma in last slide， give a condition to guarantee the interpolation matrix is column diagonally dominant

Theorem on solvability (Z&W 2014)， Basic idea illustration, column diagonally dominant

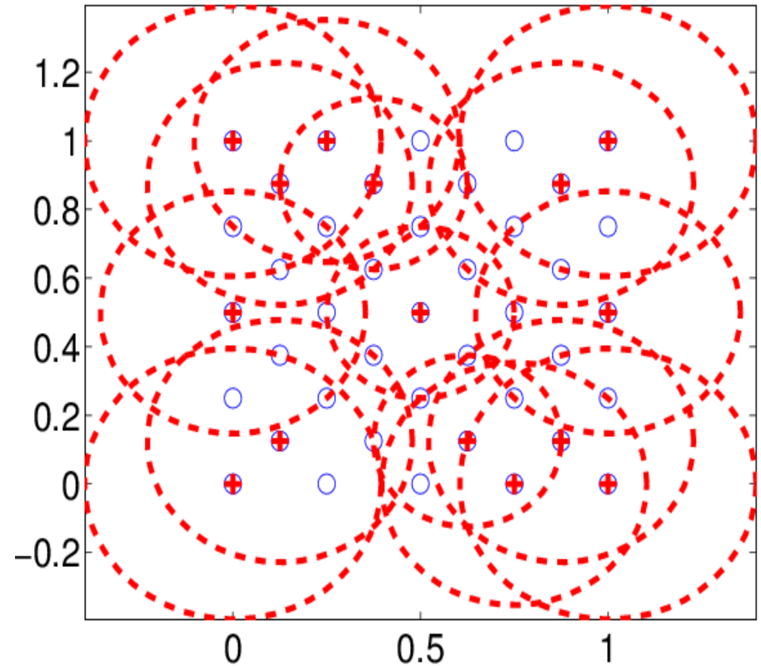




With the same shapes v.s. different shapes on tensor grids

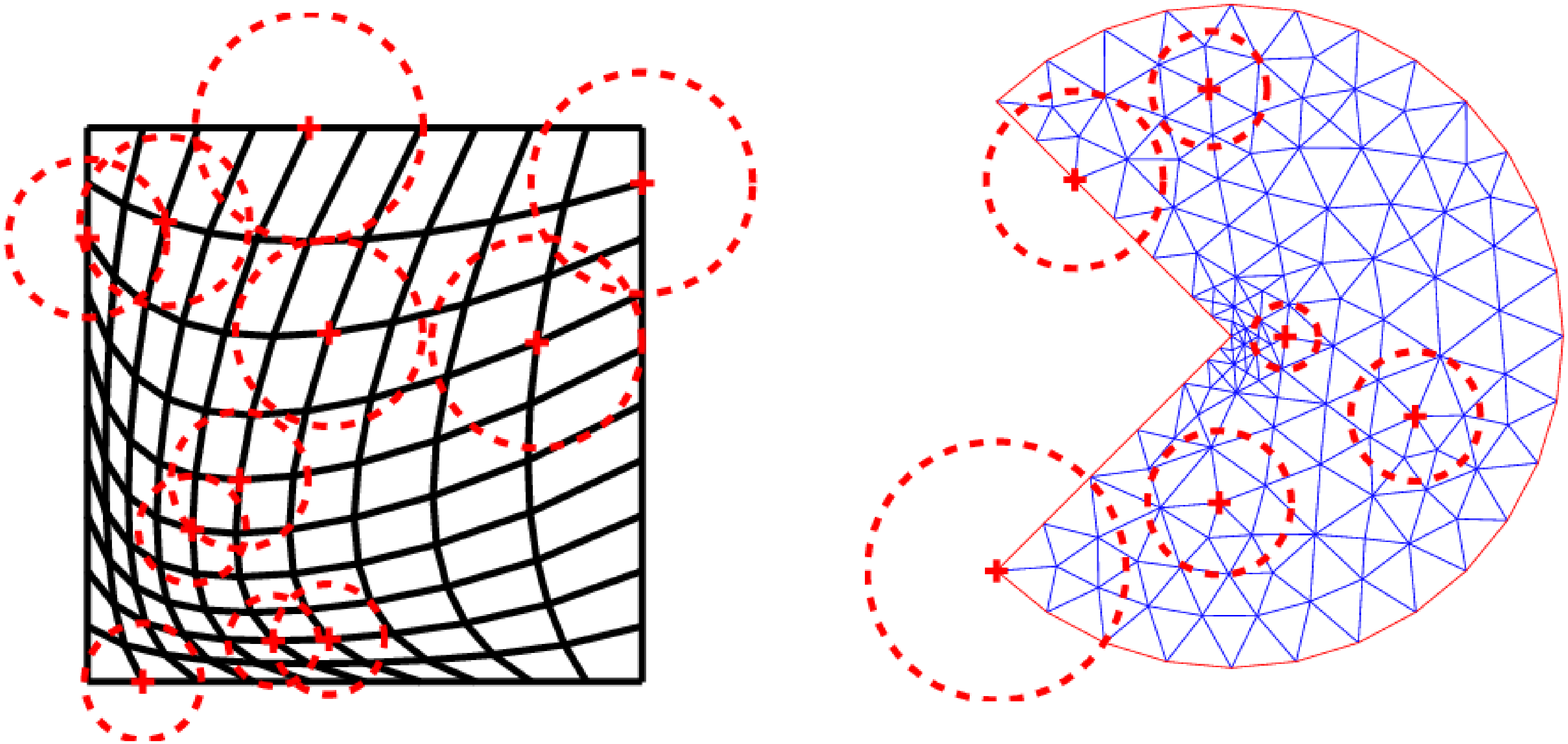


With the same shapes



With different shapes

Algorithms to determine the radii of CSRBFs
Two examples



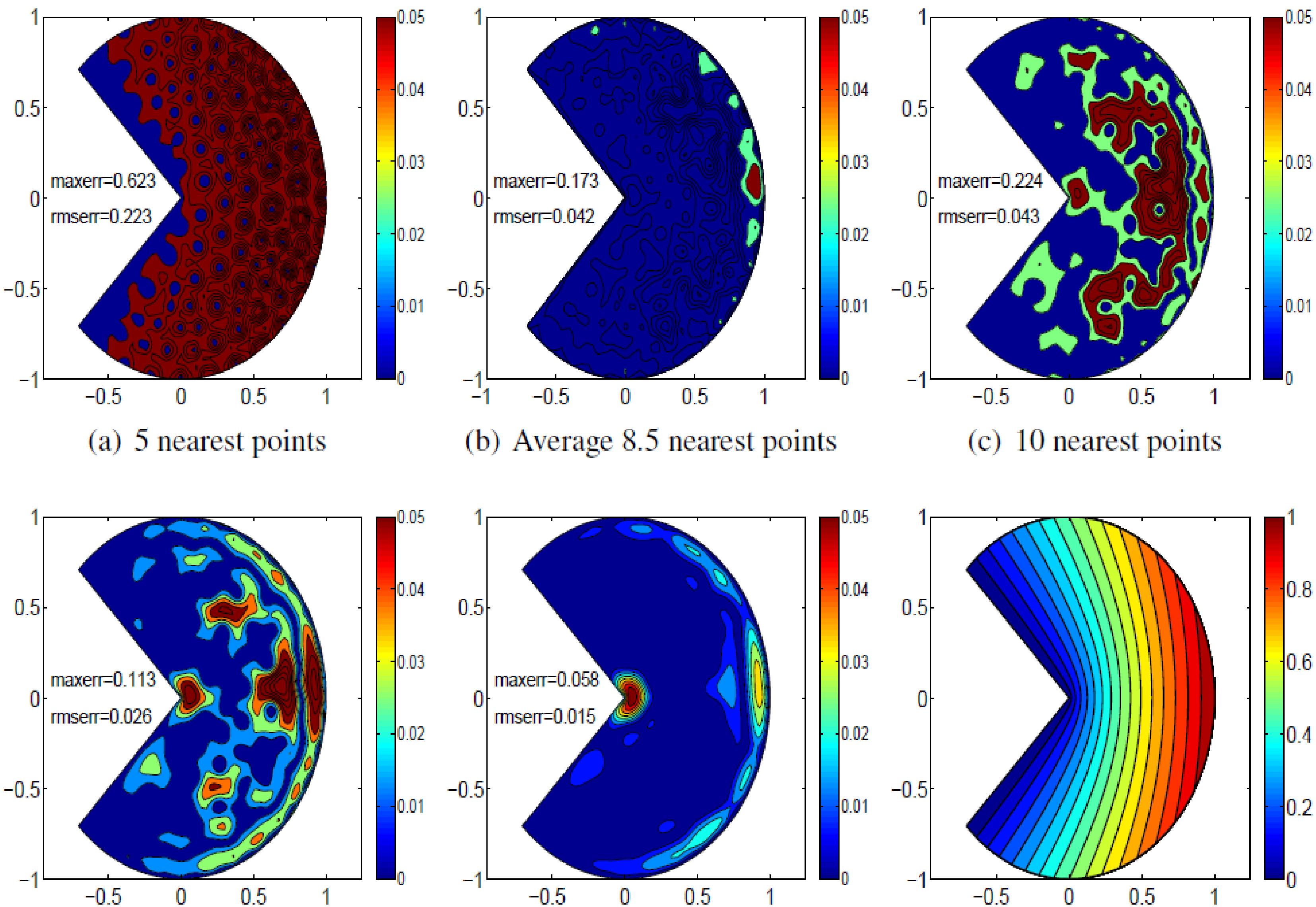
Well condition interpolation matrix
The shape/support of CSRBF varies a lot.

Table 2. Information of several test problems by Algorithm 2

Prob	A		Eig Bounds		n_j			Scale Range (R_j)			
	size	nnz	$\frac{\lambda_{\max}}{\lvert\lambda\rvert_{\min}} \leq$	$\mathtt{Re}(\lambda) \geq$	min	max	mean	min	max	mean	max/min
HC	252	1704	4.0	0.45	6	8	6.7	0.1095	0.3568	0.2275	3.3
T1	117	989	7.6	0.23	6	14	8.5	0.2163	0.4	0.2824	1.8
T2	2050	18350	10.8	0.17	4	14	9.0	3.8e-4	0.1515	5.0e-2	391.9
B4	1351	9911	10.7	0.17	3	16	7.3	3.5e-3	1.6e-2	1.0e-2	4.6
B3	5643	42825	8.8	0.20	3	17	7.6	1.6e-3	8.7e-3	5.3e-3	5.6
B2	24425	183988	12.0	0.15	3	21	7.5	1.3e-4	4.2e-3	2.6e-3	32.2
B1	105615	740493	18.9	0.10	3	21	7.0	5.8e-6	2.9e-3	1.2e-3	510.3
D3	15563	117362	10.1	0.18	3	24	7.5	2.6e-4	5.8e-3	3.51-3	22.0
D2	68830	512376	14.1	0.13	3	19	7.4	1.5e-4	3.4e-3	1.7e-3	22.2
D1	300298	2136353	13.3	0.14	2	24	7.1	2.1e-5	2.3e-3	7.9e-4	109.8

B: Stanford Bunny point cloud
D: Stanford Dragon point cloud

Comparison with approximation quality: better



(a) 5 nearest points (b) Average 8.5 nearest points (c) 10 nearest points

(d) 16 nearest points (e) 40 nearest points (f) Test function

Increasing any computational workload ? Almost no!

Table 4. Timing results (in seconds) for C/C++ implementation

Prob	N	knn	check	sparse	ilu	gmres	R1 (‰)	R2 (‰)
B4	1351	0.0130	0.0002	0.0015	0.0007	0.0054	3.28	1.05
B3	5643	0.0543	0.0009	0.0068	0.0031	0.0169	4.50	1.21
B2	24425	0.2576	0.0040	0.0342	0.0162	0.0726	4.50	1.15
B1	105615	1.2357	0.0180	0.1546	0.0801	0.4029	3.73	1.05
D3	15563	0.1641	0.0027	0.0213	0.0087	0.0543	4.29	1.19
D2	68830	0.7776	0.0119	0.1023	0.0495	0.2118	4.55	1.15
D1	300298	3.5881	0.0521	0.5041	0.2335	0.9574	4.37	1.09

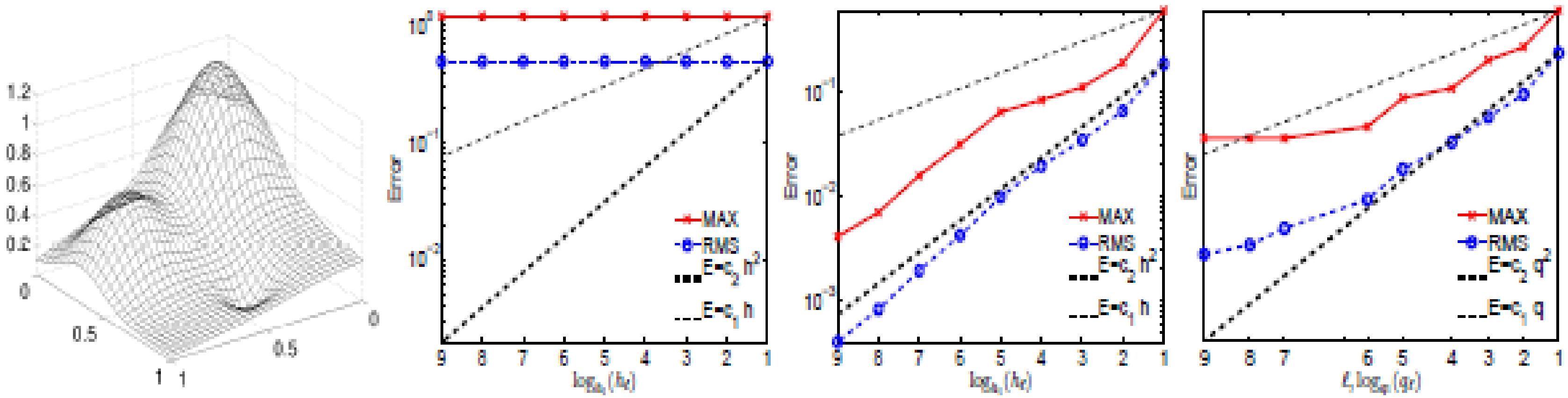
Knn: data querying
Check: determine the radii of all the basis functions
Spare: formation the sparse matrix
R1: check time / linear solver time
R2: check time/ overall time

Comparison on convergence: set up

level	N	Using CSRBFs with different shapes						M1
		n_j			R_j			R_ℓ
		max	min	mean	max	min	mean	
1	9	8	6	6.7	1.1180	0.7071	1.0199	1.1180
2	13	8	6	6.8	0.7906	0.5000	0.7169	0.7906
3	25	8	7	7.5	0.5590	0.3536	0.4567	0.5590
4	41	8	6	7.2	0.3953	0.2500	0.3046	0.3953
5	81	8	7	7.9	0.2795	0.1768	0.2072	0.2795
6	145	8	6	7.6	0.1976	0.1250	0.1384	0.1976
7	289	8	7	8.0	0.1398	0.0884	0.0967	0.1398
8	545	8	6	7.8	0.0988	0.0625	0.0658	0.0988
9	1089	8	7	8.0	0.0699	0.0442	0.0464	0.0699

Two approach share almost the same storage and scalability
M1: multilevel stationary approach
HHRBF.

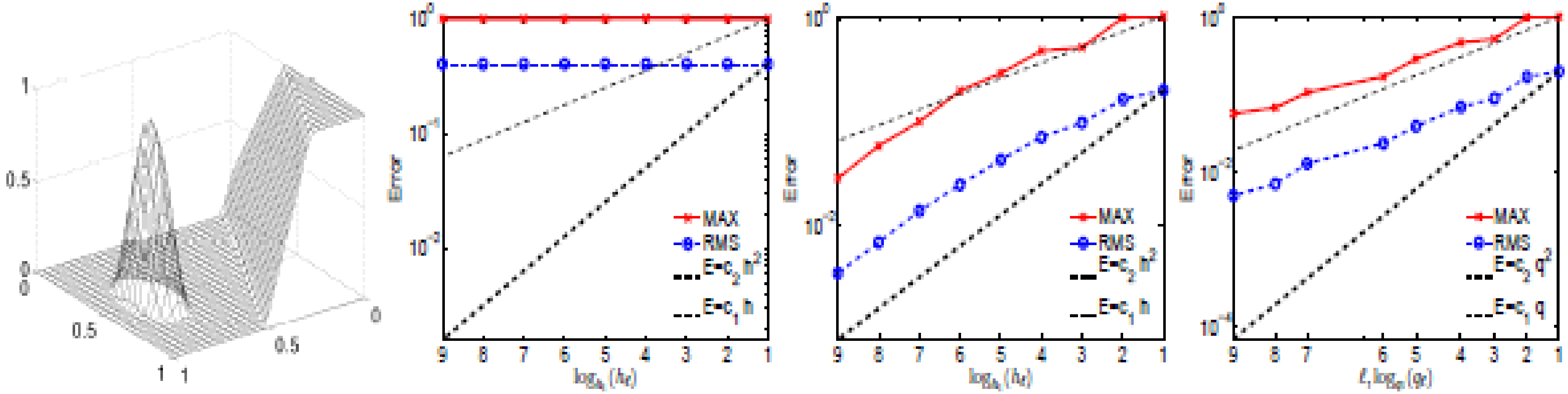
Comparison on convergence



(a) F1 (b) F1 (c) F1 (d) F1

(a) Test function： Frank function
(b) Multilevel stationary interpolation does not converge
(c) HHRBF on regular nested data sets (Red ℓ_∞) blue (RMS)
(d) HHRBF on scattered nested data sets
(e) All other 12 test functions share similar result.

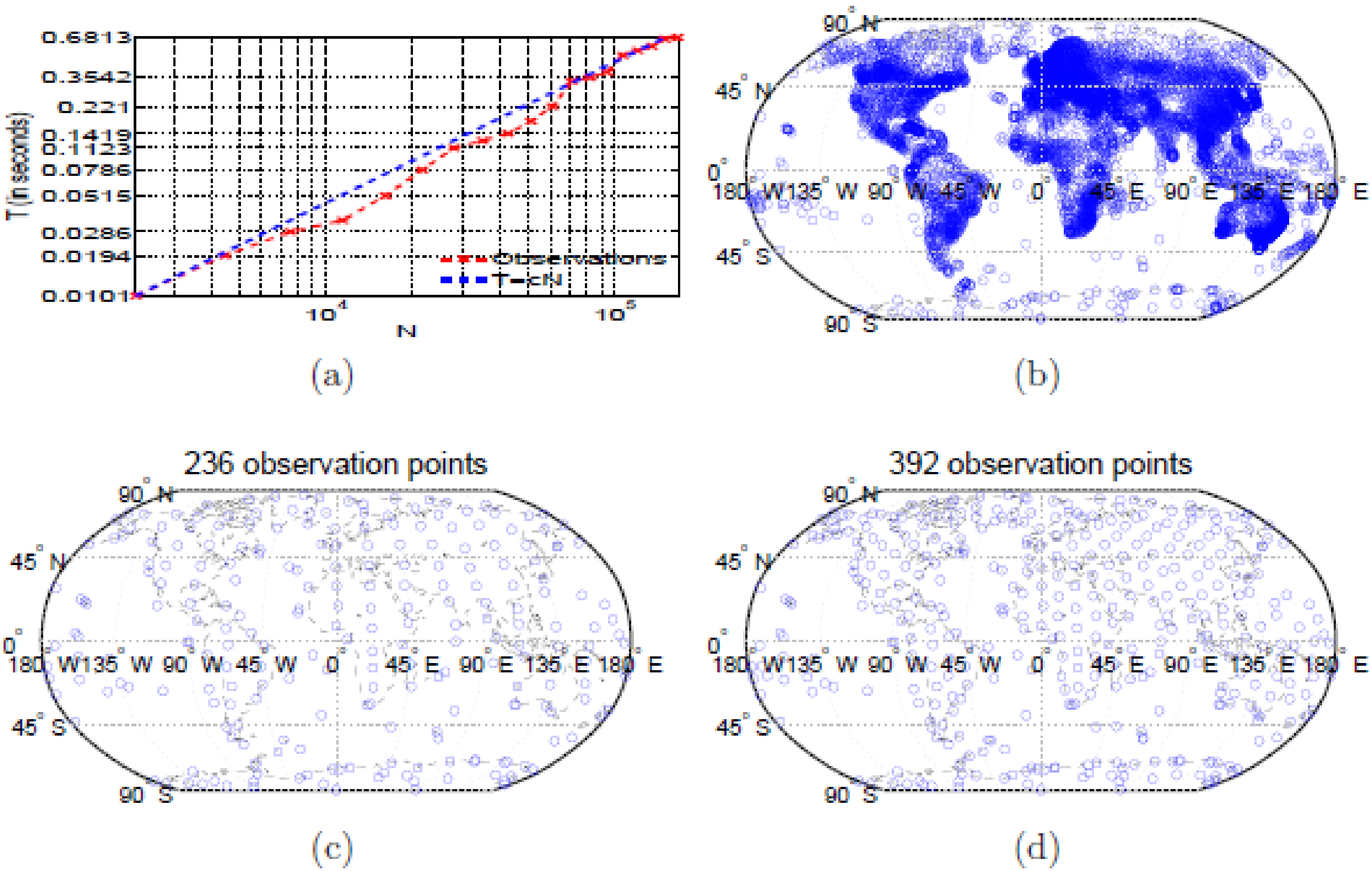
Comparison on convergence



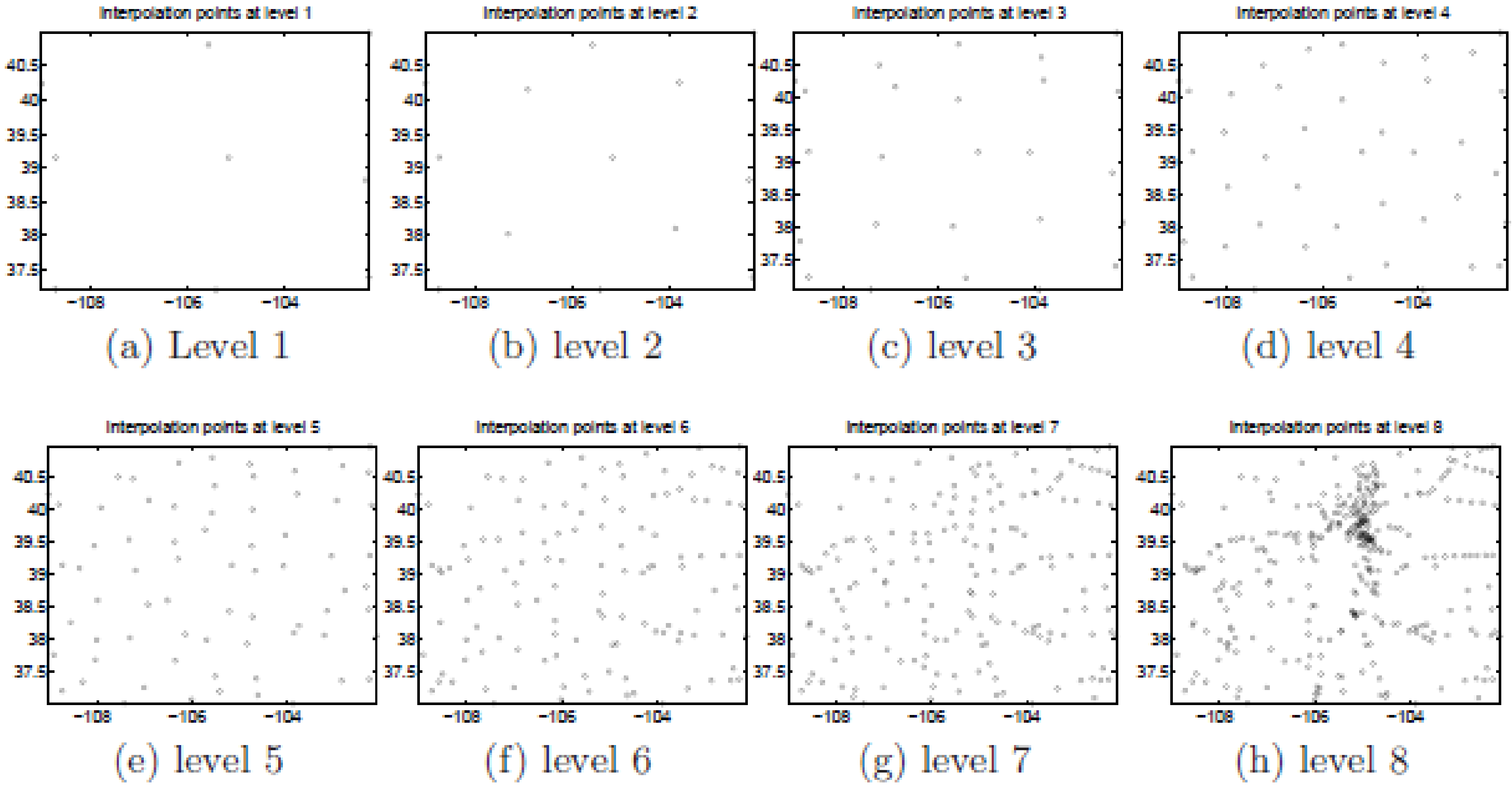
(i) F11 (j) F11 (k) F11 (l) F11

(a) Test function： Frank function
(b) Multilevel stationary interpolation does not converge
(c) HHRBF on regular nested data sets (Red ℓ_∞) blue (RMS)
(d) HHRBF on scattered nested data sets
(e) All other 12 test functions have similar result.

Hierarchical data sets construction



Numerical examples

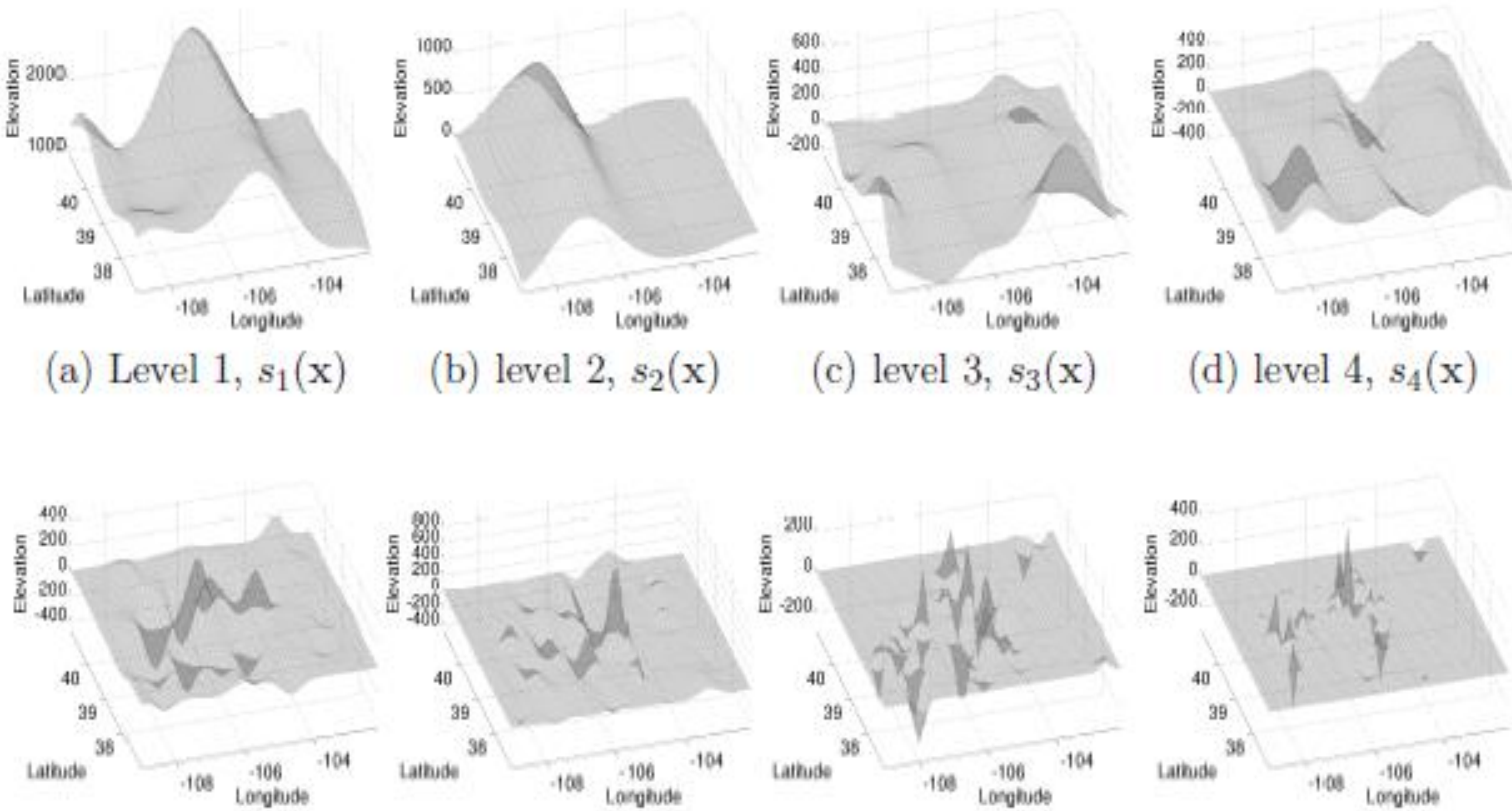


(a) Level 1 (b) level 2 (c) level 3 (d) level 4

(e) level 5 (f) level 6 (g) level 7 (h) level 8

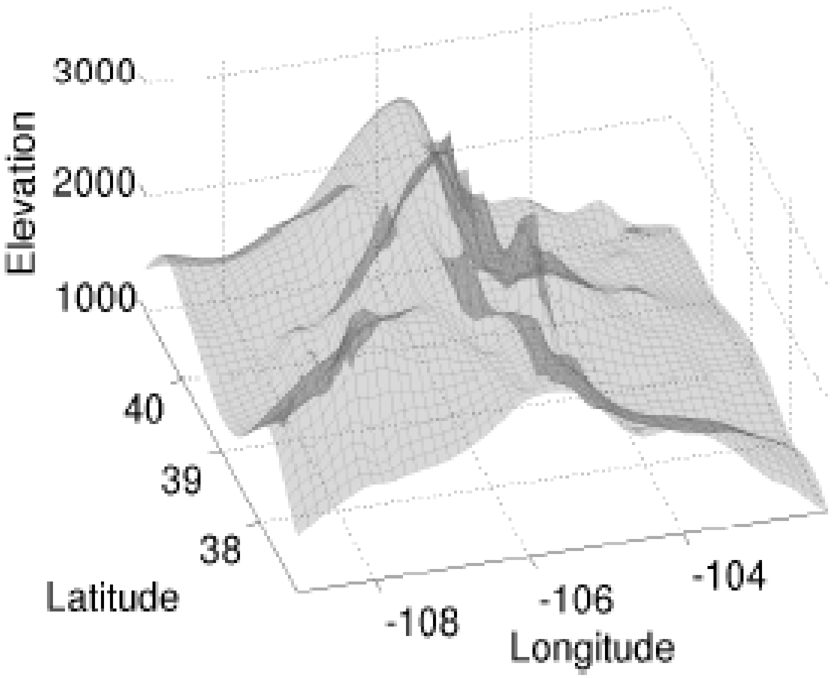
Figure 7.6: Hierarchical data sets of cities in Colorado, USA.

Numerical examples

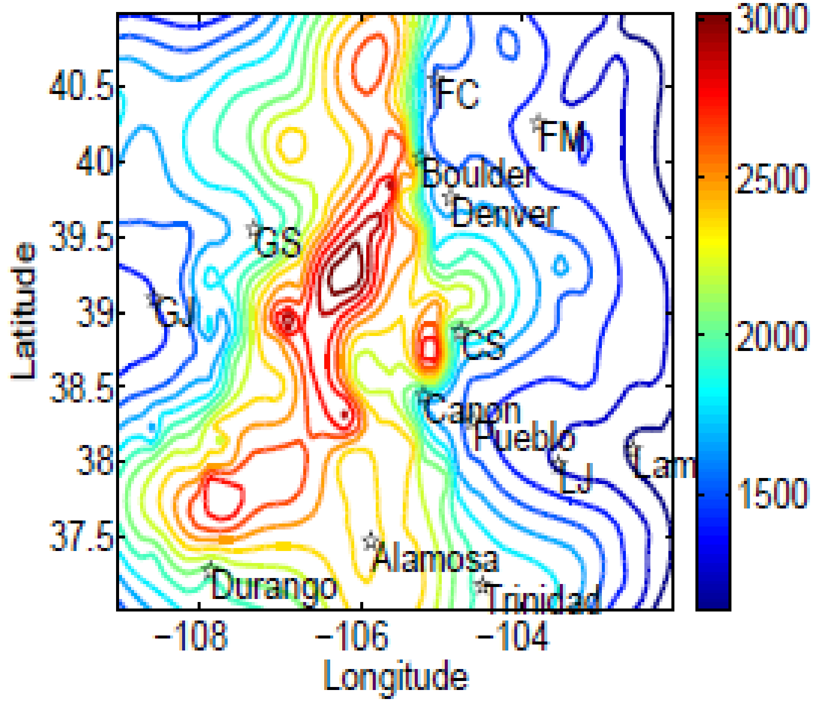


(a) Level 1, $s_1(\mathrm{x})$ (b) level 2, $s_2(\mathrm{x})$ (c) level 3, $s_3(\mathrm{x})$ (d) level 4, $s_4(\mathrm{x})$

(e) level 5, $s_5(\mathrm{x})$ (f) level 6, $s_6(\mathrm{x})$ (g) level 7, $s_7(\mathrm{x})$ (h) level 8, $s_8(\mathrm{x})$



(i) Terrain of Colorado

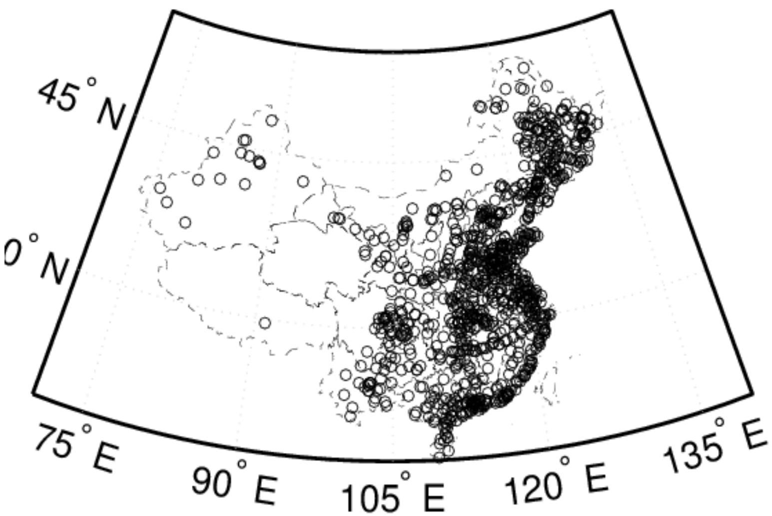


(j) Contour map of Colorado

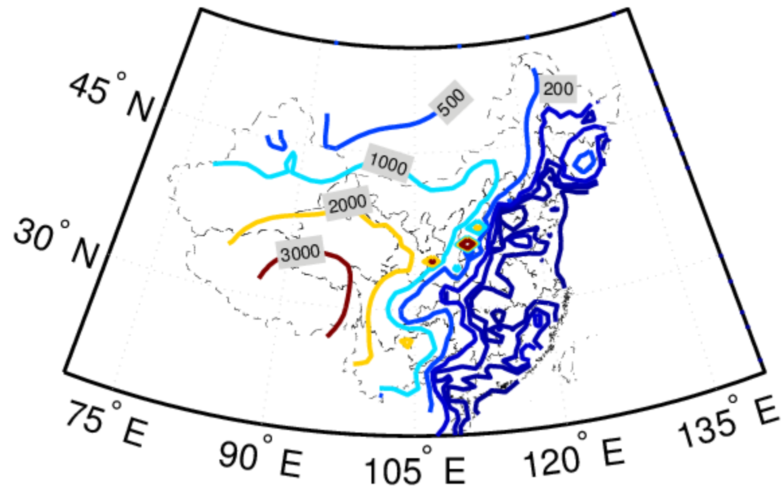
Xi'an Jiaotong-Liverpool University 西交利物浦大学

Example 2:

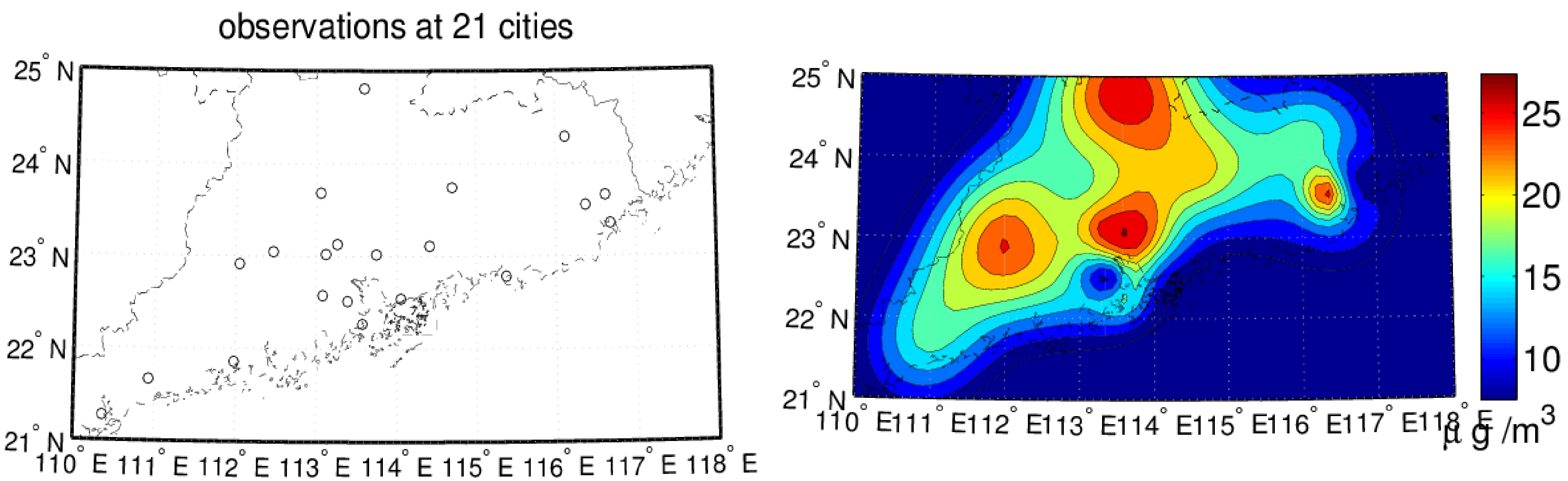
Observations at 914 cities



Interpolation with 9 levels at 914 cities

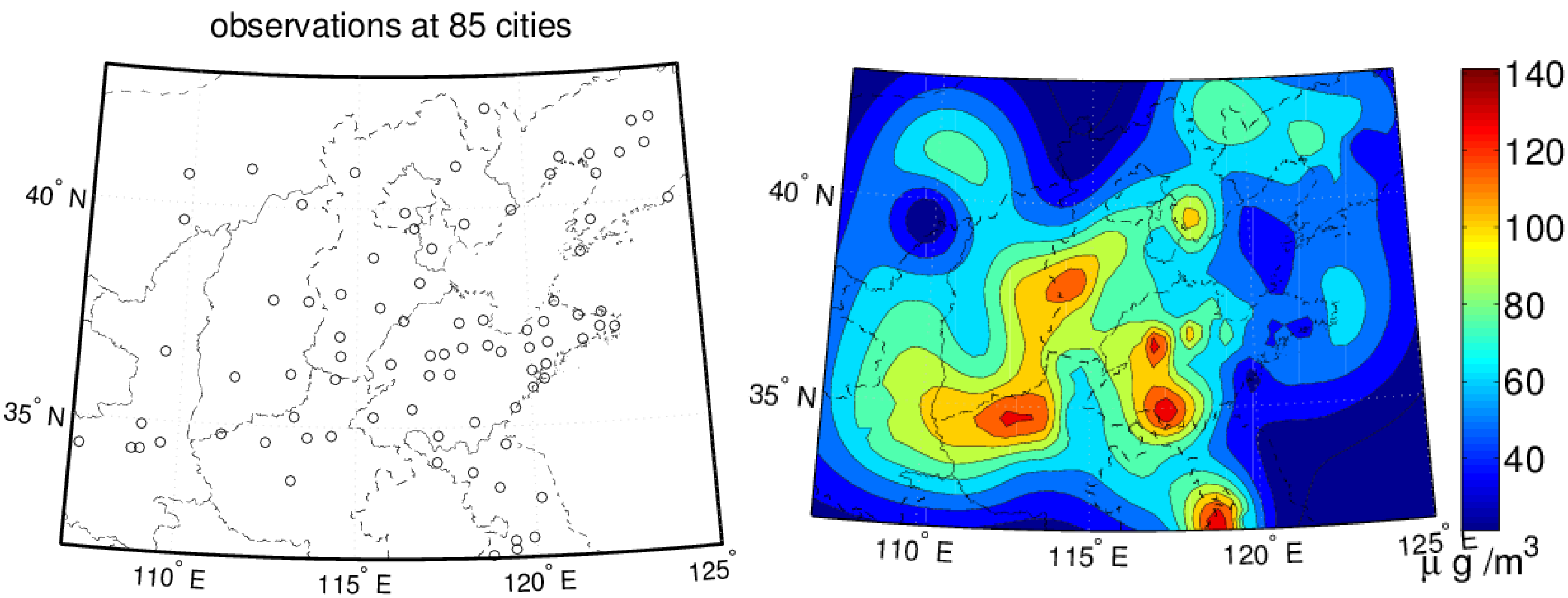


Example 3

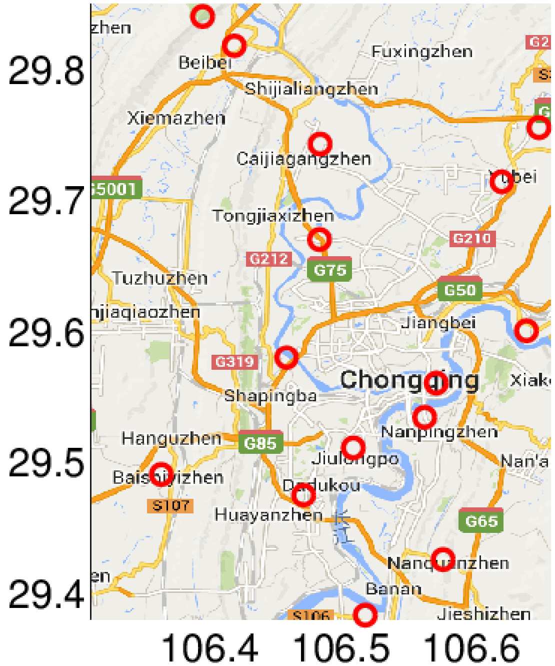


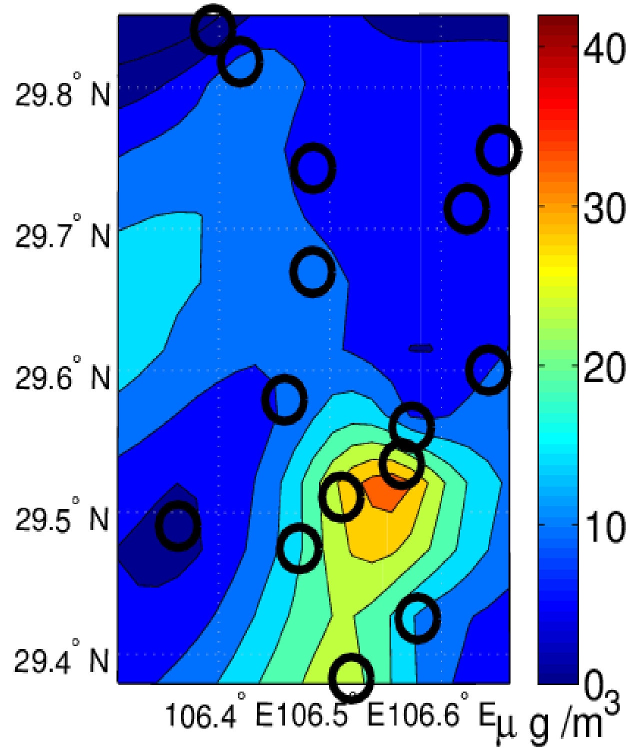
Xi'an Jiaotong-Liverpool University
西交利物浦大学

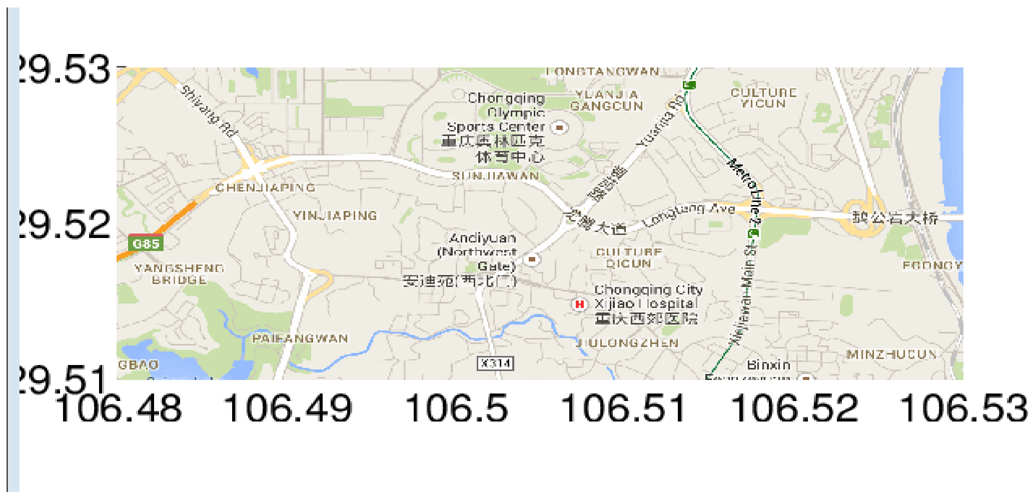
Example 3


observations at 85 cities
40° N
35° N
110° E
115° E
120° E
125°
140
120
100
80
60
40
µ g /m3

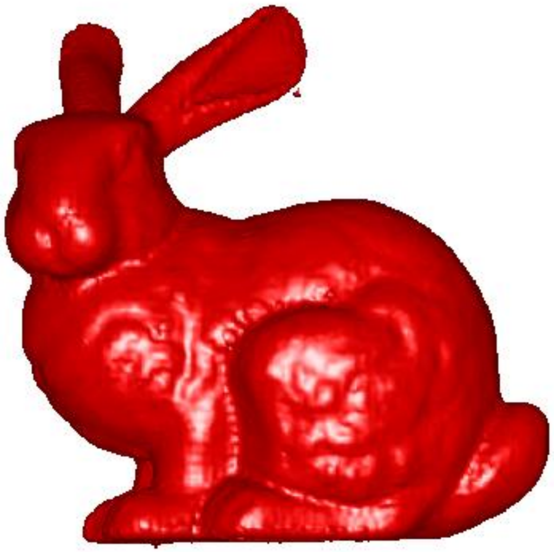
Example 4

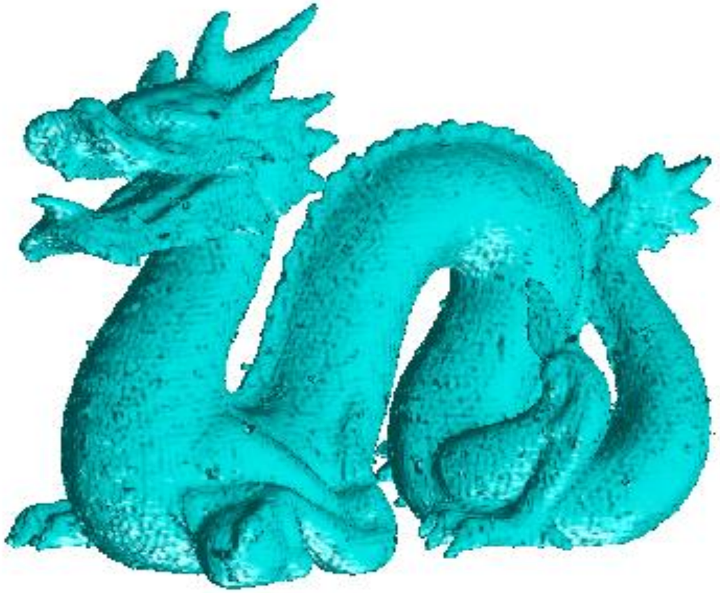

29.8
29.7
29.6
29.5
29.4
106.4
106.5
106.6
Beibei
Fuxingzhen
Shijialiangzhen
Xiemazhen
Caijiagangzhen
Tongjiaxizhen
Tuzhuzhen
Jiangbei
Chongqing
Shapingba
Hanguzhen
Nanpingzhen
Jiulongpo
Dadukou
Huayanzhen
Banan
Jieshizhen
G75
G50
G85
G65
G210
G212
G319
S107
S106


29.8° N
29.7° N
29.6° N
29.5° N
29.4° N
106.4° E
106.5° E
106.6° E
40
30
20
10
0
μ g /m3

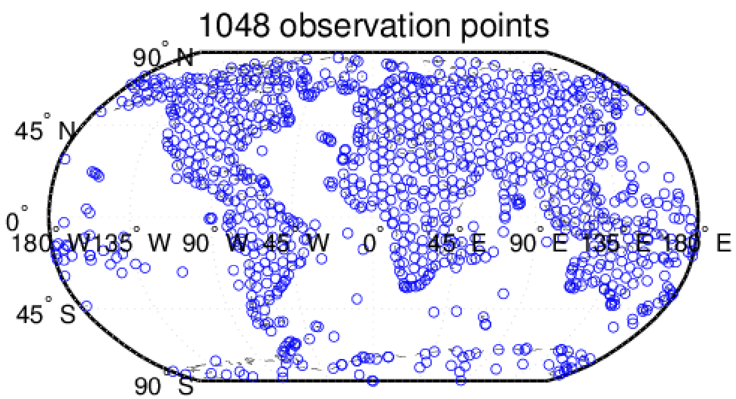

29.53
29.52
29.51
106.48
106.49
106.5
106.51
106.52
106.53

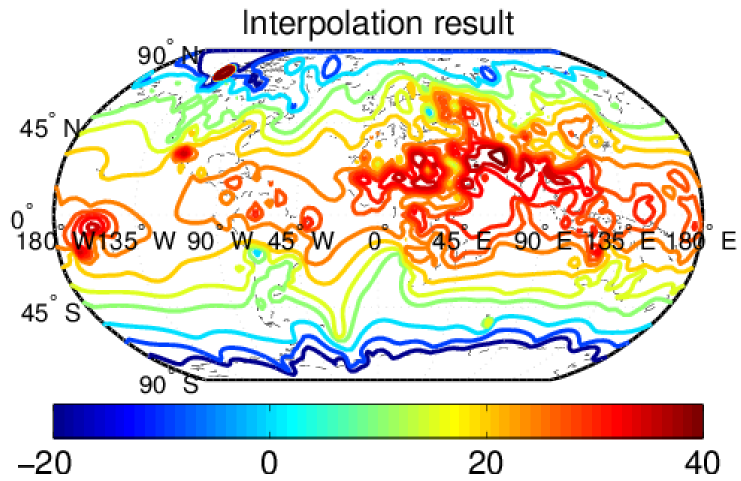
Example 5





Example 6





Temperature June 22, 2014

Conclusions

1. The motivation of HHRBF might be right , we might ask a correct question.
2. Multilevel stationary interpolation does not converge, while HHRBF is observed convengent $O(h^2)$ convergence in ℓ_2 and $O(h)$ in ℓ_∞ , but theoretical results is unknown. We need a proof.
3. HHCSBF, not necessarily radial.

Thank You !

Q & A!

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Xi'an Jiaotong-Liverpool University
西交利物浦大学