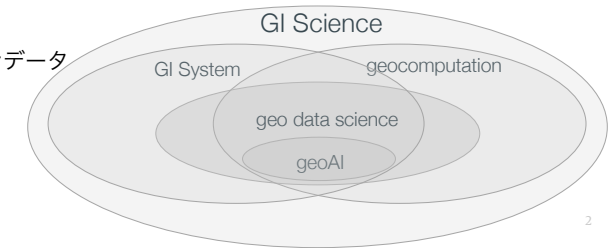


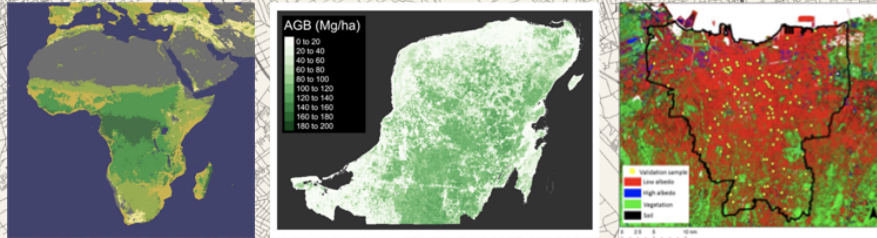
堤田成政（つつみだ なるまさ）

研究興味対象

- 地理情報科学
 - GIS, リモートセンシング, ジオコンピューテーション, 空間統計学
- 人間活動の環境インパクト評価
 - 土地利用・被覆
- オープンサイエンス
 - オープンソース・オープンデータ



土地被覆分類データの空間精度・空間誤差



堤田 成政
埼玉大学 大学院理工学研究科
数理電子情報部門情報領域



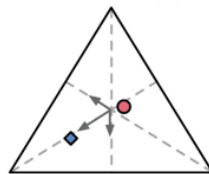
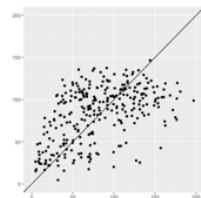
ラスタ型土地被覆データの誤差評価

データ制作とは独立した検証データをもちいる

- カテゴリーデータ
 - 全体精度 (overall accuracy)
データを正しく分類できた割合
 - ユーザー精度 (user's accuracy)
分類結果のうち、あるクラスに正しく分類できた割合
 - プロデューサー精度 (producer's accuracy)
検証サンプルのうちあるクラスに正しく分類できた割合
- 連続型データ
 - 相関係数 (r)
 - 平均誤差 (ME)
 - 平均絶対誤差 (MAE)
 - 平均二乗誤差平方根 (RMSE)
- 比率データ
 - Aitchison距離
 - など

	Validation				Total	User
	Class A	Class B	Class C			
Classified	Class A	B _{AA}	B _{AB}	B _{AC}	B _{A+}	B _{AA} /B _{A+}
	Class B	B _{BA}	B _{BB}	B _{BC}	B _{B+}	B _{BB} /B _{B+}
	Class C	B _{CA}	B _{CB}	B _{CC}	B _{C+}	B _{CC} /B _{C+}
Total	B _{+A}	B _{+B}	B _{+C}		n	
Producer	B _{AA} /B _{+A}	B _{BB} /B _{+B}	B _{CC} /B _{+C}			

Overall accuracy = $(B_{AA} + B_{BB} + B_{CC})/n$; n is the total number of classified or validation data.



<https://elifesciences.org/articles/21887>

精度評価の課題

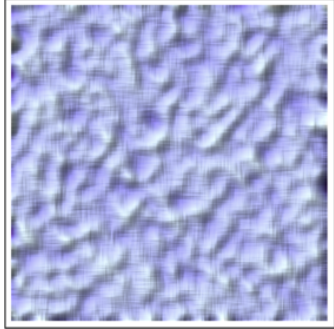
- 分類誤差がランダムに分布しているとは限らない
- 精度指標は正しさ（誤差）の平均値
 - 2次元→1次元
 - 空間的な精度の偏りを表現できない

目的

土地被覆データにおける精度（誤差）評価指標の空間拡張

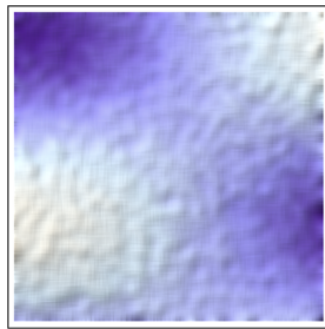
空間モデルが考慮する空間的特性

空間的従属性
Spatial dependence



<https://stats.stackexchange.com/questions/18406/what-is-the-difference-between-spatial-dependence-and-spatial-heterogeneity>

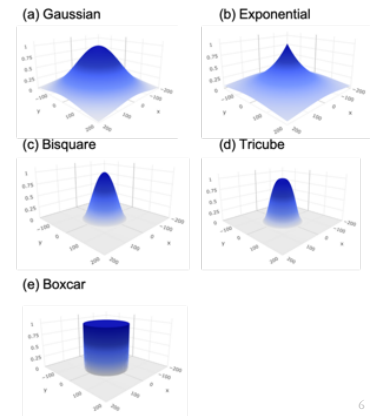
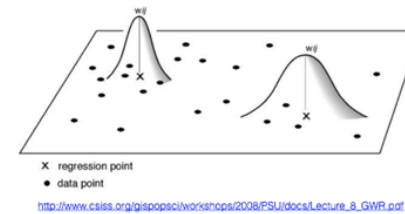
空間的従属性 + 空間的異質性
Spatial dependence + heterogeneity



5

地理的加重モデルとは？

- Geographically weighted (GW) model
 - 空間的異質性の把握
 - さまざまな統計モデル、基本統計量に適用可能
 - 距離減衰移動カーネル

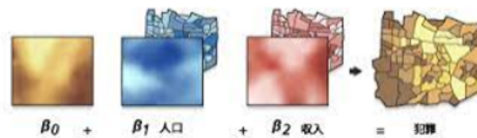


6

例：地理的加重回帰

地理的加重回帰

- 回帰係数が場所ごとに変わる（ローカル）



<https://desktop.arcgis.com/ja/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm>

7

線形回帰

- 回帰係数が一定（グローバル）

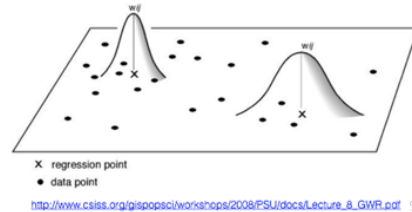
地理的加重モデルとは？

- 地理的加重要約統計量
 - GW mean, variance, correlation, ...
- 地理的加重回帰モデル
 - regression, semiparametric regression, sparse regression, ...
- 地理的加重一般化線形モデル
 - Poisson regression, logistic regression, ...
- 地理的加重主成分分析
 - PCA, non-negative PCA, ...

8

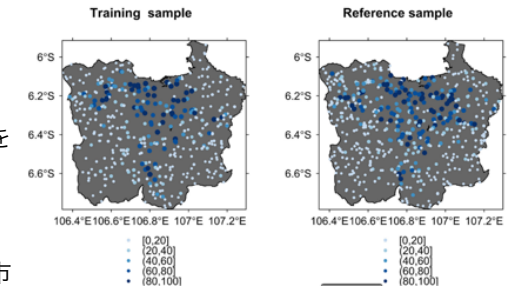
空間精度・誤差分析手法

- 連続型データ
 - GW error diagnostics
 - Pearson's correlation coefficient
 - Mean Signed Deviation (MSD)
 - Mean Absolute Error (MAE)
 - Root Mean Squared Error (RMSE)
- カテゴリーデータ
 - GW correspondence matrix
 - Overall accuracy
 - User's accuracy
 - Producer's accuracy
- 比率データ
 - GW total composite error analysis
 - Aitchison distance

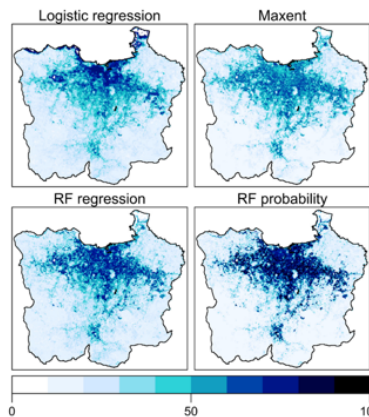


連続型データ

- 2012年のインドネシア・ジャカルタの都市（非透水被覆：Impervious surface）被覆率推定
- MODIS MOD13Q1 EVI (231 m、16 days interval)の年最小、最大、平均、標準偏差と教師データ434点をもちいて、Logistic regression、Random Forest regression、RF probability、Maxentにより推定
- それぞれのモデルから推定された都市被覆率図を独立検証データ（550点）で誤差評価



従来の誤差評価



	msd	mae	rmse	r	Moran's I [*]
Logistic regression	-3.63	15.77	21.87	0.72	0.11
MaxEnt	-7.57	15.83	22.74	0.72	0.11
RF regression	-2.95	15.51	21.34	0.73	0.06
RF probability	-5.12	15.85	24.22	0.71	0.05

* All p-values for estimates of Moran's I are less than 0.05.

Tsutsumida et al., (2019) International Journal of Applied Earth Observation and Geoinformation, 74, 259-268.

GW msd, mae, rmse, r

$$gw. msd(x_i, y_i) = \frac{\sum_{j=1}^n \omega_{ij} (y_j - x_j)}{\sum_{j=1}^n \omega_{ij}}$$

$$gw. mae(x_i, y_i) = \frac{\sum_{j=1}^n \omega_{ij} |y_j - x_j|}{\sum_{j=1}^n \omega_{ij}}$$

$$gw. rmse(x_i, y_i) = \sqrt{\frac{\sum_{j=1}^n \omega_{ij} (y_j - x_j)^2}{\sum_{j=1}^n \omega_{ij}}}$$

$$gw. cor(x_i, y_i) = \frac{c(x_i, y_i)}{s(x_i)s(y_i)}$$

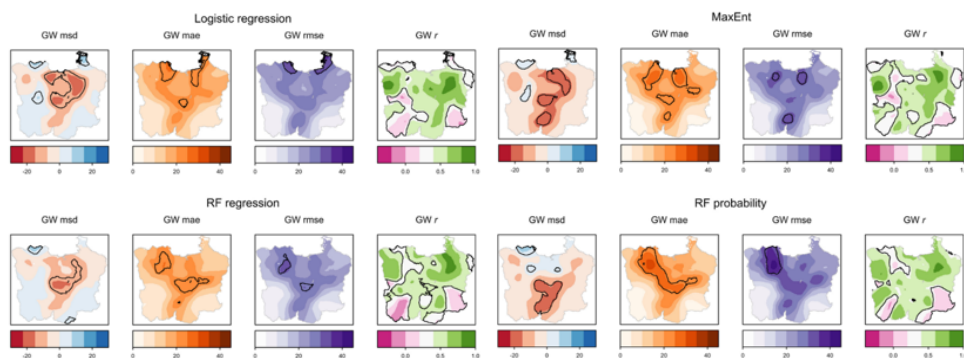
$s(x_i)$: GW standard deviation

$c(x_i, y_i)$: GW covariance

$$\omega_k = \begin{cases} \left(1 - \left(\frac{d_k}{b}\right)^2\right)^2 & \text{if } |d_k| < b, \\ 0 & \text{otherwise} \end{cases}$$

10% adaptive kernel

結果：空間精度評価（連続型データ）



Tsutsumida et al., (2019) International Journal of Applied Earth Observation and Geoinformation, 74, 259-268.

13

結果：空間精度評価（連続型データ）

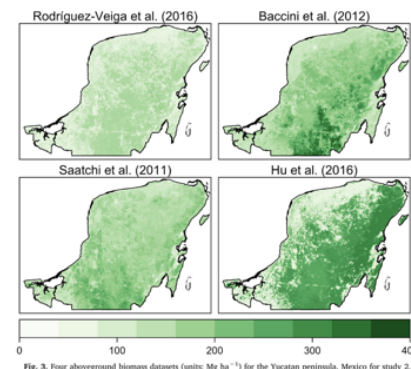


Fig. 3. Four aboveground biomass datasets (units: Mg ha^{-1}) for the Yucatan peninsula, Mexico for study 2.

Tsutsumida et al., (2019) International Journal of Applied Earth Observation and Geoinformation, 74, 259-268.

14

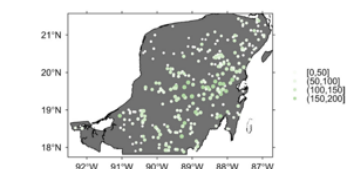


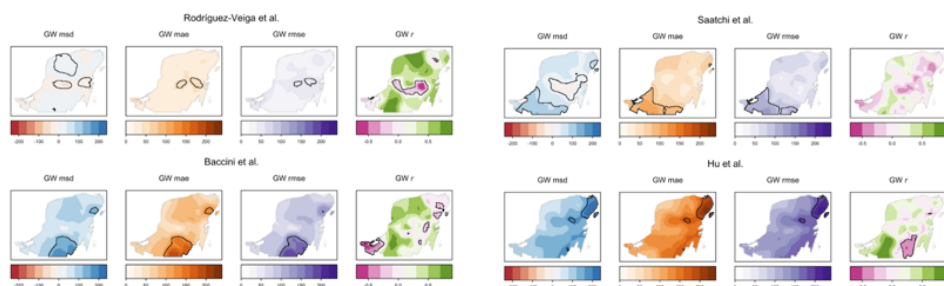
Fig. 4. The spatial distribution of in-situ reference sample points for forest aboveground biomass data (units: Mg ha^{-1}) in the Yucatan peninsula, Mexico.

Table 3
Global diagnostics and Moran's I in forest aboveground biomass datasets for study 2.

	msd	mae	rmse	r	Moran's I
Rodriguez-Veiga et al.	-2.05	31.52	39.42	0.50	0.08
Baccini et al.	-86.36	89.10	105.58	0.36	0.35
Saatchi et al.	-45.26	58.78	70.69	0.08	0.29
Hu et al.	-136.35	141.76	153.75	0.36	0.24

* All p -values for estimates of Moran's I are less than 0.05.

結果：空間精度評価（連続型データ）



Tsutsumida et al., (2019) International Journal of Applied Earth Observation and Geoinformation, 74, 259-268.

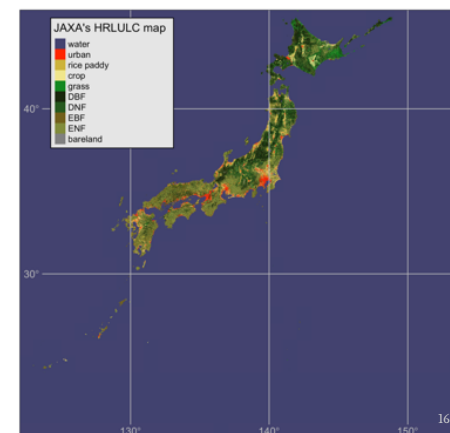
15

カテゴリーデータ

JAXA高解像度土地利用土地被覆図 ver. 18.03 (2014-2016)

- 30m 空間解像度
- 10クラス
 - water; urban; rice paddy; crop; grass; DBF; DNF; EBF; ENF; and bareland
- 使用データ
 - ALOSPALSAR2, Suomi NPP, OSM data, etc.
- モデル
 - カーネル法にもとづく確率密度推定 (橋本ら, 2014)

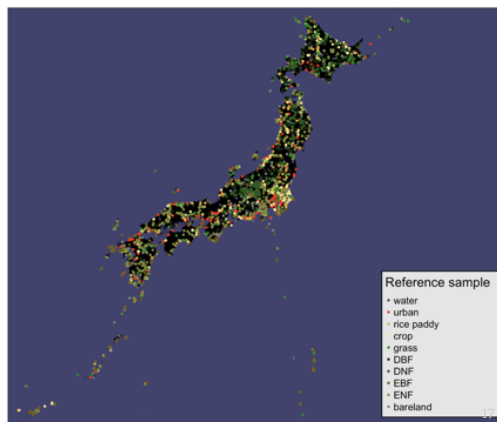
→ 4つの森林クラスの分類誤差評価



検証データ

SACLAJ: Site-based dataset for Assessment of Changing LAnd cover by JAXA

- 現地調査写真・記録データ
- 層別サンプリング
 - 各クラス約300点をランダム抽出



GW correspondence matrix

Conventional correspondence matrix (M)

m_{ij} : the element of a matrix at i th row and j th column of M

$$m_{ij} = \sum_{l=1}^L a_{ij(l)},$$

$$a_{ij(l)} = \begin{cases} 1 & \text{if } c_l = i \text{ and } r_l = j, \\ 0 & \text{otherwise} \end{cases}$$

Forest overall accuracy: $FO = \frac{\sum_{i=f_1}^{f_n} m_{ii}}{\sum_{i=f_1}^{f_n} \sum_{j=f_1}^{f_n} m_{ij}}$

Forest User's accuracy: $FU_i = \frac{m_{ii}}{\sum_{j=f_1}^{f_n} m_{ij}}$

Forest Producer's accuracy: $FP_i = \frac{m_{ii}}{\sum_{j=f_1}^{f_n} m_{ji}}$

Aspatial accuracy assessment approach

GW correspondence matrix

GW correspondence matrix (Comber et al, 2017)

$\rho_{ij(k)}$: the element of a matrix at i th row and j th column at k -th grid

$$m_{ij} = \sum_{l=1}^L a_{ij(l)},$$

$$a_{ij(l)} = \begin{cases} 1 & \text{if } c_l = i \text{ and } r_l = j, \\ 0 & \text{otherwise} \end{cases}$$

$$\rho_{ij(k)} = \sum_{k=1}^K \omega_k a_{ij(k)} \quad \omega_k = \begin{cases} \left(1 - \left(\frac{d_k}{b}\right)^2\right)^2 & \text{if } |d_k| < b, \\ 0 & \text{otherwise} \end{cases}$$

Forest overall accuracy: $GWFO_{(k)} = \frac{\sum_{i=f_1}^{f_n} \rho_{ii}}{\sum_{i=f_1}^{f_n} \sum_{j=f_1}^{f_n} \rho_{ij}}$ 10% adaptive kernel

Forest User's accuracy: $GWFU_{i(k)} = \frac{\rho_{ii}}{\sum_{j=f_1}^{f_n} \rho_{ij}}$

Forest Producer's accuracy: $GWFP_{i(k)} = \frac{\rho_{ii}}{\sum_{j=f_1}^{f_n} \rho_{ji}}$

結果

混合行列

Classified	Reference									user's accuracy (%)
	water	urban	rice paddy	crop	grass	DBF	DNF	EBF	ENF	
water	304	1	0	1	1	0	0	0	0	95.3
urban	1	254	5	3	2	0	0	1	0	85.8
rice paddy	1	3	287	34	6	0	0	0	0	85.4
crop	0	7	5	220	14	0	0	1	0	84.0
grass	0	1	1	28	248	9	3	2	1	78.0
DBF	0	0	1	3	14	238	26	16	6	76.0
DNF	0	0	1	2	0	29	252	7	4	84.6
EBF	1	0	0	2	8	13	4	218	24	79.6
ENF	4	0	0	1	1	11	15	56	266	73.5
bareland	0	34	0	6	6	0	0	0	0	190
producer's accuracy (%)	97.7	84.7	95.7	73.3	82.7	79.3	84.0	72.4	88.4	63.1

- 全体精度: 82.2%
- 森林／非森林の分類精度は高い (97.4%)が森林内クラス分類精度はそこそこ
 - 森林全体精度 (82.2%)

森林ユーザー・プロデューサー精度

	DBF	DNF	EBF	ENF
forest user's accuracy (%)	83.2	86.3	84.2	76.4
forest producer's accuracy (%)	81.8	84.8	73.4	88.7

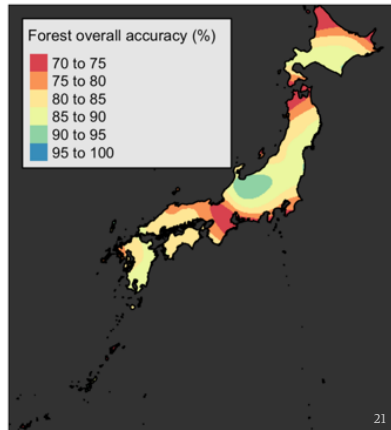
結果

Geographically Weighted approach

GW forest overall accuracy
(Global measure: 82.2%)

- More accurate in central Japan (Nagano/ Toyama)
- Less accurate in Mie, Shiga, Northern edge of Hokkaido, and Northern Aomori

Tsutsumida N., et al (2019) ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences

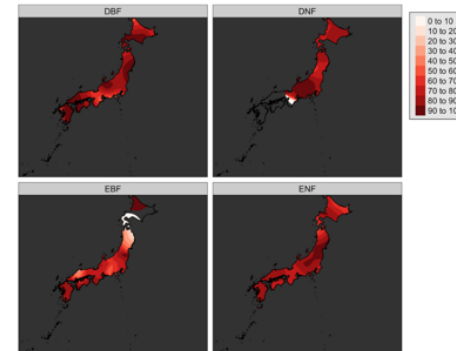


21

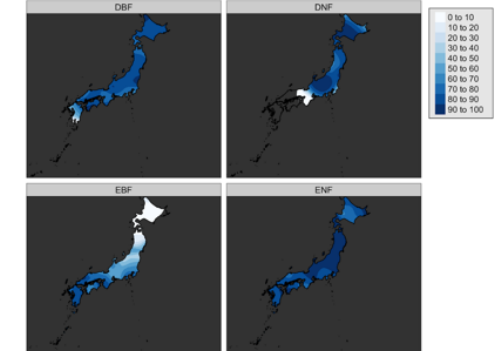
結果

Geographically Weighted approach

GW forest User's accuracy



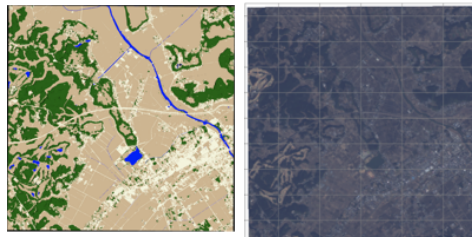
GW forest Producer's accuracy



22

比率データ

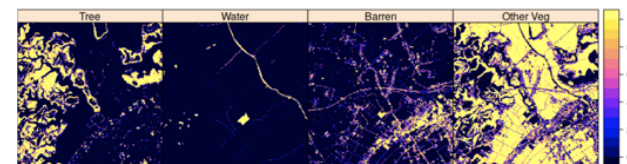
- 千葉県某所
- 使用データ
 - 参照データ: USGS global reference data (2 m)
 - 分類データ: ALOS2 AVNIR-2 (10 m)
- モデル
 - Spectral Unmixing Analysis



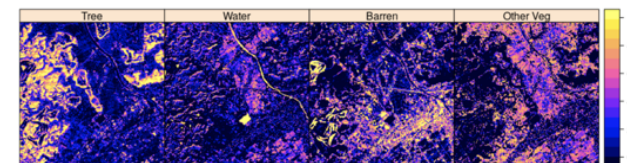
23

比率データ

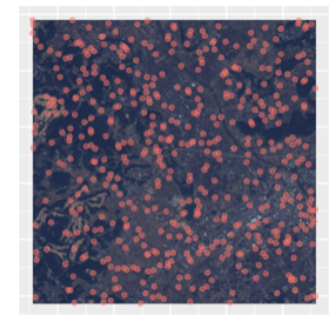
参照比率データ



推定データ



1万点サンプリング



24

GW total composite error

Aitchison distance

$$d_{\text{Ait}}(\mathbf{x}, \mathbf{y}) = d_{\text{Euc}}(\text{clr}(\mathbf{x}), \text{clr}(\mathbf{y}))$$

有心対数比

(centered log-ratio)変換

$$\text{clr}(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right]^T$$

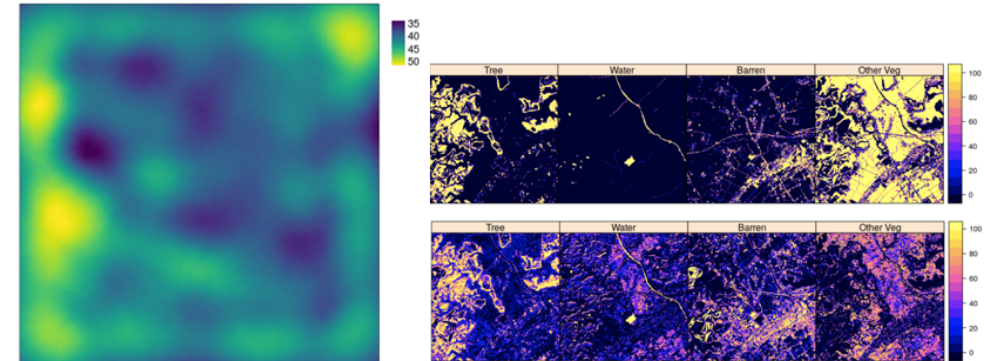
$$g(\mathbf{x}) = \left(\prod_{d=1}^D x_d \right)^{1/D}$$

これに重み (bi-squared, 10% adaptive kernel)をつけ、
各ピクセルiに適用

$$d_{\text{Ait},i}(\mathbf{x}'_i, \mathbf{y}'_i) \quad \begin{aligned} \mathbf{x}'_i &= \omega_k \text{clr}(\mathbf{x}) \\ \mathbf{y}'_i &= \omega_k \text{clr}(\mathbf{y}) \end{aligned} \quad \omega_k = \begin{cases} \left(1 - \left(\frac{d_k}{b}\right)^2\right)^2 & \text{if } |d_k| < b, \\ 0 & \text{otherwise} \end{cases}$$

25

GW total composite error

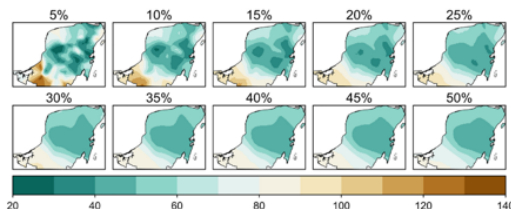


Tsutsunida et al., (2020) IGARSS 2020

26

考察

- GW法によるアプローチの課題
 - カーネルサイズ：空間局所性をどのように決定すべきか
 - 検証データの分布が結果に与える影響
 - i. 階層サンプリングには不向き
 - ii. ランダムサンプリングが望ましい



27

考察

- 空間的従属性の検討
 - クリギング など
- 空間精度（空間誤差）の誤差は？
- 参照データの信頼性

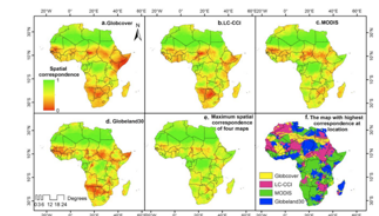


Figure 7. Spatial correspondence of the GLC maps (a-d), maximum correspondence (e) and the map with highest correspondence (f).

Tsendbazar et al. (2015) Spatial Accuracy Assessment and Integration of Global Land Cover Datasets, Remote Sensing

28

まとめ

- 土地被覆データは主題図作成（環境研究など）に不可欠
 - 誤差はその後の分析に大きな影響を及ぼす可能性がある
- 空間データの次元を下げず誤差評価する手法を提案した
 - 二次元→二次元で誤差の視覚化
 - どの地域で誤差が大きい・小さいか
 - ROIにおいて既存プロダクトはどの程度有用か？
 - どの地域の分類結果をアップデートしていくか？

29

