Korean Journal of Environmental Biology

Original article

https://doi.org/10.11626/KJEB.2021.39.4.526

Korean J. Environ. Biol.
39(4) : 526-535 (2021)
ISSN 1226-9999 (print)
ISSN 2287-7851 (online)

농지 공간격자 자료의 층화랜덤샘플링: 농업시스템 기후변화 영향 공간모델링을 위한 국내 농지 최적 층화 및 샘플 수 최적화 연구

이민영, 김용은[1], 홍진솔, 조기종*

고려대학교 환경생태공학과, [1]고려대학교 오정리질리언스연구원

A stratified random sampling design for paddy fields: Optimized stratification and sample allocation for effective spatial modeling and mapping of the impact of climate changes on agricultural system in Korea

Minyoung Lee, Yongeun Kim[1], Jinsol Hong and Kijong Cho*

Department of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea
[1]Ojeong Resilience Institute, Korea University, Seoul 02841, Republic of Korea

***Corresponding author**
Kijong Cho
Tel. 02-3290-3064
E-mail. kjcho@korea.ac.kr

Received: 29 November 2021
Revised: 17 December 2021
Revision accepted: 20 December 2021

Abstract: Spatial sampling design plays an important role in GIS-based modeling studies because it increases modeling efficiency while reducing the cost of sampling. In the field of agricultural systems, research demand for high-resolution spatial data-based modeling to predict and evaluate climate change impacts is growing rapidly. Accordingly, the need and importance of spatial sampling design are increasing. The purpose of this study was to design spatial sampling of paddy fields (11,386 grids with 1 km spatial resolution) in Korea for use in agricultural spatial modeling. A stratified random sampling design was developed and applied in 2030s, 2050s, and 2080s under two RCP scenarios of 4.5 and 8.5. Twenty-five weather and four soil characteristics were used as stratification variables. Stratification and sample allocation were optimized to ensure minimum sample size under given precision constraints for 16 target variables such as crop yield, greenhouse gas emission, and pest distribution. Precision and accuracy of the sampling were evaluated through sampling simulations based on coefficient of variation (CV) and relative bias, respectively. As a result, the paddy field could be optimized in the range of 5 to 21 strata and 46 to 69 samples. Evaluation results showed that target variables were within precision constraints ($CV < 0.05$ except for crop yield) with low bias values (below 3%). These results can contribute to reducing sampling cost and computation time while having high predictive power. It is expected to be widely used as a representative sample grid in various agriculture spatial modeling studies.

Keywords: spatial sampling, climate change, agriculture modeling, sampling cost, high-resolution spatial data

서 론

기후변화는 농업시스템 전반에 걸쳐 영향을 미치고 있으며, 국내외에서는 기후변화가 농업에 미치는 영향을 평가하기 위해 모형 개발 및 시뮬레이션 연구들이 활발히 이루어지고 있다(Hatfield *et al.* 2020). GIS 기반 공간모델링은 가장 흔히 활용되는 모델링 기법으로, 작물수량 변동, 해충 분포 변화, 농경지 온실가스배출량 변화 예측 등 농업시스템을 구성하는 구성성분들에 대한 기후변화의 시공간적 영향예측 및 평가에 활용되고 있다(Moore *et al.* 2017; Tonnang *et al.* 2017; Zhang *et al.* 2020). 기후변화 공간모델링은 일반적으로 대량의 자료를 필요로 한다. 높은 해상도의 시공간적 자료의 이용은 공간예측모형의 예측 범위와 수준을 넓히고, 모형의 예측 정확도 및 정밀도를 향상시킬 수 있으며, 고성능 모형 개발을 가능하게 한다(Folberth *et al.* 2012). 하지만, 고해상도 자료 기반 공간모델링은 모형의 시뮬레이션 또는 구축 과정에서 많은 계산 시간과 자원이 소요된다는 문제가 있다(Hijmans *et al.* 2005). 특히, 전지구적 또는 지역적 범위의 기후변화 영향 공간모델링의 경우, 이러한 이유로 인해 고해상도 자료의 활용이 극히 제한적인 실정이다.

기후변화 영향 모델링 시 공간정보의 효과적인 활용을 위해서 공간샘플링(Spatial sampling) 기술이 사용될 수 있다. 공간샘플링이란 공간자료에서 모집단의 특성치를 대변하는 표본을 추출하는 일로, 샘플링 기법의 한 영역이다(Wang *et al.* 2012). 공간샘플링은 공간모델링 연구에 활용되어 샘플링의 비용을 감소시키고 모형의 효율성을 향상시키는 역할을 할 수 있다. 최근 GIS 기술의 발전 및 빅데이터 시대의 도래와 함께 고해상도의 시공간적 자료의 축적이 급격히 증가하고 있으며, 이에 대한 연구적 수요와 활용 또한 증가하고 있는 추세이다(Metzger *et al.* 2013; Goyal *et al.* 2017). 이에 따라 고해상도 자료 이용의 비용 및 시간 문제 극복 측면에서 공간샘플링 기법의 중요성과 필요성이 강조되고 있다(Wang *et al.* 2012).

샘플링 기법으로는 일반적으로 랜덤샘플링, 규칙샘플링, 층화샘플링 등이 있으며, 공간샘플링에서는 층화랜덤샘플링(Stratified random sampling)이 가장 흔히 사용되고 있다. 층화랜덤샘플링이란 자료를 계층화하고 각 계층에서 일정 수의 샘플을 랜덤샘플링하는 기법으로 공간자기상관성을 가지는 공간자료의 경우 랜덤샘플링에 비해 샘플링의 정확도와 효율성을 향상시킬 수 있다(Aoyama 1954). 이외에도 다양한 공간샘플링 기법들이 연구되어 왔으며, 예측대상 또는 연구목적에 따른 적절한 공간샘플링의 설계는 공간예측모형의 효율성을 제고하고 성능을 향상시킬 수 있다(Stein and Ettema 2003). 한편, 공간샘플링의 샘플 수는 샘플링의 시간과 비용에 중요한 영향을 미친다. 최적의 샘플 수 결정은 공간샘플링의 비용 효율을 극대화시킬 수 있으며 뿐만 아니라 모형의 성능을 개선시키는 데에도 역할을 한다는 점에서 공간샘플링 영역에서 중요시되고 있다(Gonzalez and Eltinge 2010). 공간샘플링은 수학의 통계적 기법의 영역을 넘어 현재 여러 학문분야에서 다양한 목적으로 활용되고 있다(Wang *et al.* 2012; Metzger *et al.* 2013). 농업분야에서는 기후변화에 따른 작물수량 예측을 위한 농지 계층화, 온실가스 메타모형 구축 등을 목적으로 기후변화 공간모델링 연구 시 공간샘플링 기법이 일부 활용된 바 있다(McCallion 1992; Perlman *et al.* 2014; Bussel *et al.* 2016; Zhao *et al.* 2016). 국내에서 공간샘플링 기법을 공간모델링 연구에 활용한 사례는 현재까지 알려지지 않았다. 공간적 계층화와 관련하여 공간적 기상자료에 대한 클러스터 분석 연구가 일부 수행된 바 있지만, 이는 공간적 기상자료에 대한 일차적인 분석에 그치며 기후변화 공간모델링 연구를 위한 공간샘플링 연구로 확장되지는 않고 있는 실정이다(Joo *et al.* 2009; Yeo 2011).

국내 기상청에서는 다양한 공간해상도(135 km, 12.5 km, 1 km)의 미래 기후 전망자료를 배포하고 있다. 특히, 1 km 해상도 남한상세 기후전망자료는 파편화된 국내 농지의 특성을 고려하였을 때 농업분야 공간모델링 연구에 활용되기에 적절한 자료라고 할 수 있다. 기초자료제공 측면에서 국내 농업시스템 공간모델링 연구분야는 기후변화 연구의 우수한 연구기반을 갖추고 있다고 할 수 있다. 하지만, 실제 고해상도 공간자료 활용 시 발생하는 시뮬레이션 시간 소요 및 비용 문제로 인해 이러한 자료들에 대한 적극적인 활용은 어려운 실정이다. 따라서, 국내의 농업분야 기후변화 공간모델링 연구 활성화 및 고해상도 공간자료의 효과적인 활용을 위해서는 농업시스템 공간샘플링 연구가 선행될 필요가 있다.

본 연구는 국내 농지 모집단의 공간샘플링 연구를 통해 농업분야 기후변화연구의 공간자료 활용의 효율성을 제고하고자 하였다. 이에 따라, 본 연구는 국내 농지를 기상 및 토양 특성에 따라 계층화하였으며, 층화랜덤샘플링을

Table 1. Descriptions of a total of 29 stratification variables used in the construction of stratified random sampling design

Attribute	Variable	Name	Description (unit)
Climate	x1	tmax1	Average max temperature during period 1 (°C)
	x2	tmax2	Average max temperature during period 2 (°C)
	x3	tmin1	Average min temperature during period 1 (°C)
	x4	tmin2	Average min temperature during period 2 (°C)
	x5	tmean1	Average mean temperature during period 1 (°C)
	x6	tmean2	Average mean temperature during period 2 (°C)
	x7	tmean3	Average mean temperature during period 3 (°C)
	x8	diur1	Mean diurnal range during period 1 (°C)
	x9	diur2	Mean diurnal range during period 2 (°C)
	x10	srad1	Accumulated solar radiation during period 1 (MJ m^{-2})
	x11	srad2	Accumulated solar radiation during period 2 (MJ m^{-2})
	x12	srad3	Accumulated solar radiation during period 3 (MJ m^{-2})
	x13	prec1	Accumulated precipitation during period 1 (mm)
	x14	prec2	Accumulated precipitation during period 2 (mm)
	x15	prec3	Accumulated precipitation during period 3 (mm)
	x16	ftemp1	Average mean temperature for 7 days after first fertilizer application (°C)
	x17	ftemp2	Average mean temperature for 7 days after second fertilizer application (°C)
	x18	ftemp3	Average mean temperature for 7 days after third fertilizer application (°C)
	x19	BIO2	Mean diurnal range (0.1°C)
	x20	BIO3	Isothemality (= mean diurnal range/annual temperature range)
	x21	BIO5	Max temperature of warmest month (0.1°C)
	x22	BIO6	Min temperature of coldest month (0.1°C)
	x23	BIO12	Annual precipitation (mm)
	x24	BIO13	Precipitation of wettest month (mm)
	x25	BIO14	Precipitation of driest month (mm)
Soil	x26	SOC	Soil organic carbon (%)
	x27	pH	Potential of hydrogen
	x28	Den	Soil bulk density (Mg m^{-3})
	x29	Clay	Clay content (%)

Notes: period 1, period between transplanting date and anthesis date; period 2, period between anthesis date and maturity date; period 3, period between transplanting date and maturity date

기반으로 공간샘플링의 비용 효율을 극대화하기 위해 최적 층화 및 샘플 배정 및 샘플 수 최적화를 수행하였다.

재료 및 방법

1. 공간샘플링 설계 및 공간자료 수집

기상 및 토양인자들은 농업의 결과물(e.g., 작물수량, 농지 온실가스 배출)에 영향을 미치는 주요 영향인자들이다. 유사한 기상 및 토양 특성을 가지는 농지 공간격자들에서는 유사한 농업의 결과물들이 산출될 것으로 기대할 수 있다. 이에 따라, 국내 농지를 기상 및 토양 특성에 따라 계층화하고 농업의 결과물에 대한 예측이 가능하도록 공간샘플링을 설계하였다.

1 km 공간해상도의 국내 농지 모집단 공간격자자료에 대해, 층화랜덤샘플링을 기반으로 하는 공간샘플링을 설계하였다. 국내 농지는 1 km 공간해상도 수준에서 11,386개 격자(내륙지역)로 이루어졌다. 단, 제주, 울릉, 독도를 포함한 도서지역들은 1 km 공간해상도 수준에서 농지를 포함하지 않으므로 본 연구에서 제외되었다. 기초 공간자료로서 1 km 해상도의 기상(최고기온, 최저기온, 평균기온, 강수량, 일사량), 토양(토양유기물, 토양산성도, 용적밀도, 점토함량), 및 농업의 결과물 자료(i.e., 작물(작물수량, 필요관개용수량, 증발산량), 농지 온실가스 배출(CO_2 배

출량, CH_4 배출량, N_2O 배출량), 해충(벼멸구(p01), 애멸구(p02), 이화명나방(p03), 배줄기굴파리(p04), 혹명나방(p05)의 분포확률 및 발생 세대 수))를 수집하였다(Table 1). 기상자료는 기상청에서 제공하는 RCP (Representative Concentration Pathways, 대표농도경로) 시나리오 1 km 해상도 남한상세 자료(2026~2035/2046~2055/2076~2085년)를 활용하였으며, 본 연구에서는 온실가스 저감정책이 상당히 실현되는 RCP 4.5 및 현재 추세대로 온실가스가 배출되는 RCP 8.5 자료를 활용하였다(data from Web site of the Korea Meteorological Administration, http://www.climate.go.kr/). 일사량의 경우, 12.5 km 한반도 자료를 이중선형보간법을 통해 1 km 자료로 변환하여 사용하였으며, 전체 기상자료들은 작물생육시기별 자료로 재구성하여 사용하였다(Table 1). 토양자료는 농촌진흥청 "흙토람"에서 제공하는 토양통 자료를 받아 1 km 격자형 자료로 변환하여 사용하였다(data from Web site of the Korean Soil Information System, http://soil.rda.go.kr/) (Table 1). 농업의 결과물 자료들은 수집된 기상 및 토양자료를 기반으로 작물생산성 모형(DSSAT; Decision Support System for Agrotechnology Transfer), 토양 온실가스 모형(DNDC; Denitrification and Decomposition), 해충 모형(MaxEnt; Maximum Entropy model)을 구동하여 얻은 모형 예측 결과 자료를 활용하였다(Table 2).

기후변화는 장기적 현상으로 기후변화 하에서의 기상 및 토양 특성에 따른 농지 특성화 역시 장기적으로 형성되는 특성으로 볼 수 있다. 이에 따라, 농지 특성화는 매년 시시각각 변하지 않고 수십 년 단위(또는 연대 수준)에서 장기적으로 변화하는 것으로 가정하였으며, 2030, 2050, 2080년대에 대한 연대별 공간샘플링을 설계하였다. 기후시나리오별로 각 연대의 기상 및 토양 특성에 따른 농지 계층화 및 샘플 수 최적화를 수행하였으며, 연대 자료는 Bussel *et al.* (2016)의 방식과 유사하게 각 연대의 10년 자료(2026~2035/2046~2055/2076~2085년)를 평균하여 사용하였다.

2. 최적 층화 및 샘플 배정 및 샘플 수 최적화

층화랜덤샘플링의 최적 층화 및 샘플 배정 최적화는 목표 변수들에 대한 주어진 정밀도 제한 내에서 샘플링 비용을 최소화하는 방향으로 진행된다(Ballin and Barcaroli 2013). 층화 변수로는 총 29가지 기상 및 토양인자들이 사용되었으며(Table 1), 목표 변수로는 총 16가지 농업 결과물 인자들이 사용되었다(Table 2).

초기 계층에서 시작하여 최적화의 각 단계별로 층화랜덤샘플링을 수행하여 총 분산(Eq. 1)과 샘플링 비용(Eq.

Table 2. Descriptions of a total of 16 target variables used in the construction of stratified random sampling design

Attribute	Variable	Name	Description (unit)
Crop	y1	yield	Potential yield (kg ha^{-1})
	y2	irrig_amt	Irrigation amount (mm)
	y3	et	Evapotranspiration (mm)
Greenhouse gas emission	y4	CO_2	CO2 emission (kgC ha^{-1})
	y5	CH_4	CH4 emission (kgC ha^{-1})
	y6	N_2O	N2O emission (kgC ha^{-1})
Pest	y7	p01di	Potential distribution of p01
	y8	p02di	Potential distribution of p02
	y9	p03di	Potential distribution of p03
	y10	p04di	Potential distribution of p04
	y11	p05di	Potential distribution of p05
	y12	p01ng	Potential number of generations of p01
	y13	p02ng	Potential number of generations of p02
	y14	p03ng	Potential number of generations of p03
	y15	p04ng	Potential number of generations of p04
	y16	p05ng	Potential number of generations of p05

Notes: p01, *Nilaparvata lugens*; p02, *Laodelphax striatellus*; p03, *Chilo suppressalis*; p04, *Chlorops oryzae*; p05, *Cnaphalocrocis medinalis*

2)을 계산한다. 계층화의 초기값(단위 격자들이 속하는 초기 계층)은 층화 변수들에 대한 K-means 클러스터링을 수행하여 결정하였다.

$$\mathrm{VAR}(\hat{Y}_g) = \sum_{h=1}^{H} N_h^2 \left(1 - \frac{n_h}{N_h}\right) \cdot \frac{S_{h,g}^2}{n_h} \quad \text{(Eq. 1)}$$

h는 계층($h = 1, \ldots, H$), g는 목표 변수, N_h, $S^2_{h,g}$, n_h는 각각 각 계층에서의 모집단, 분산, 샘플링 수를 나타낸다.

$$C(n_1, \ldots, n_H) = C_0 + \sum_{h=1}^{H} C_h n_h \quad \text{(Eq. 2)}$$

h는 계층($h = 1, \ldots, H$), C_0는 고정상수, n_h는 각 계층에서의 샘플링 수, C_h는 샘플당 인터뷰 비용을 나타낸다.

최적화 과정은 유전적 알고리즘(genetic algorithm)에 따라 다음 세대의 적합도(fitness)를 높여가는 방향, 즉, 목적 함수(objective function)인 샘플링 비용(sampling cost; Eq. 2)을 최소화하는 방향으로 진행되었다(Schmitt 2001; Ballin and Barcaroli 2013) (Eq. 3). 최적화가 진행되는 동안 각 반복(iteration)단계에서 각 개체(individual) (i.e., stratification)의 적합도를 계산하며, 적합도가 높은(샘플링 비용이 낮은) 유전체(genome) (i.e., atomic strata)는 다음 세대에 전달, 적합도가 낮은 유전체는 교배(Crossover), 돌연변이(Mutation) 과정을 거쳐 다음 세대에 전달되었다. 목표 변수별로 기대되는 샘플링 분산의 상한선(샘플링 정밀도 제한)을 설정한 후, 베델 알고리즘(Bethel 1989)에 따라 층화된 농지에 대해 샘플링의 비용을 최소로 하는 최적 샘플 배정을 구해 나간다(Eq. 3). 이때, 기대되는 샘플링 분산의 상한선 설정은 목표 변수들의 척도에 의존하지 않도록 변동계수를 기반으로 한다.

$$\begin{cases} C_0 + \sum_{h=1}^{H} C_h n_h \rightarrow min \\ CV(\hat{Y}_1) \leq U_1 \\ CV(\hat{Y}_2) \leq U_2 \\ \cdots \\ CV(\hat{Y}_G) \leq U_G \end{cases} \quad \text{(Eq. 3)}$$

h는 계층($h = 1, \ldots, H$), C_0는 고정상수, n_h는 각 계층에서의 샘플링 수, C_h는 샘플당 인터뷰 비용, G는 목표 변수, U_G는 샘플링의 정밀도 제한, CV(Coefficient of variation)는 변동계수를 나타낸다.

본 연구에서는 모든 목표 변수들에 대해서 정밀도 제한은 변동계수 0.05 수준으로 설정하였다.

3. 공간샘플링 평가

공간샘플링의 정밀도와 정확도는 각각 변동계수와 상대적 편향(Relative bias = (distribution mean − true value)/true value)을 기반으로 평가되었다(Ballin and Barcaroli 2013). 평가를 위해 기후시나리오별, 연대별 최적 층화 및 샘플 배정 최적화 결과에 대해 100번, 10번의 두 가지 층화랜덤샘플링 시뮬레이션을 수행하였으며, 각각에 대해 평가를 수행하였다. 전체 목표변수들에 대해 기후시나리오별, 연대별로 변동계수와 상대적 편향을 계산하였으며(Tables S1~4), 기후시나리오별로 각 연대의 평가값을 평균내어 정밀도(상대적 편향의 절대값의 평균값을 사용)와 정확도를 평가하였다. 본 연구의 전체 과정은 R version 3.5.3 (R Core Team 2018) 및 RStudio version 1.4.1717 (RStudio Team 2021)를 이용하여 이루어졌다.

결과 및 고찰

국내 농지 공간격자 모집단(11,386개 격자)에 대한 기후시나리오 및 연대별 계층화 및 샘플 배정 및 샘플 수 최적화 결과, 전체 농지는 평균적으로 약 10개 계층(범위: 5~21), 59개 샘플(범위: 46~69) 수준에서 최적화되었다(Table 3, Figs. 1, 2).

계층 수는 RCP 8.5 시나리오(연대 평균 약 13계층)에서가 RCP 4.5 시나리오(연대 평균 약 6계층)에서보다 약 2배 정도 더 많았다(Table 3, Fig. 1a, d, and g, and Fig. 2a, d,

Table 3. Optimized stratification and sample size for spatial sampling of domestic paddy fields (11,386 grids with 1 km spatial resolution). Optimization was conducted by RCP scenario (RCP 4.5/RCP 8.5) by year (2030s/2050s/2080s)

RCP scenario	Years	Strata	Samples
RCP 4.5	2030s	8	63
	2050s	6	55
	2080s	5	46
	Mean	6	55
RCP 8.5	2030s	7	69
	2050s	12	56
	2080s	21	63
	Mean	13	63
Total mean		10	59

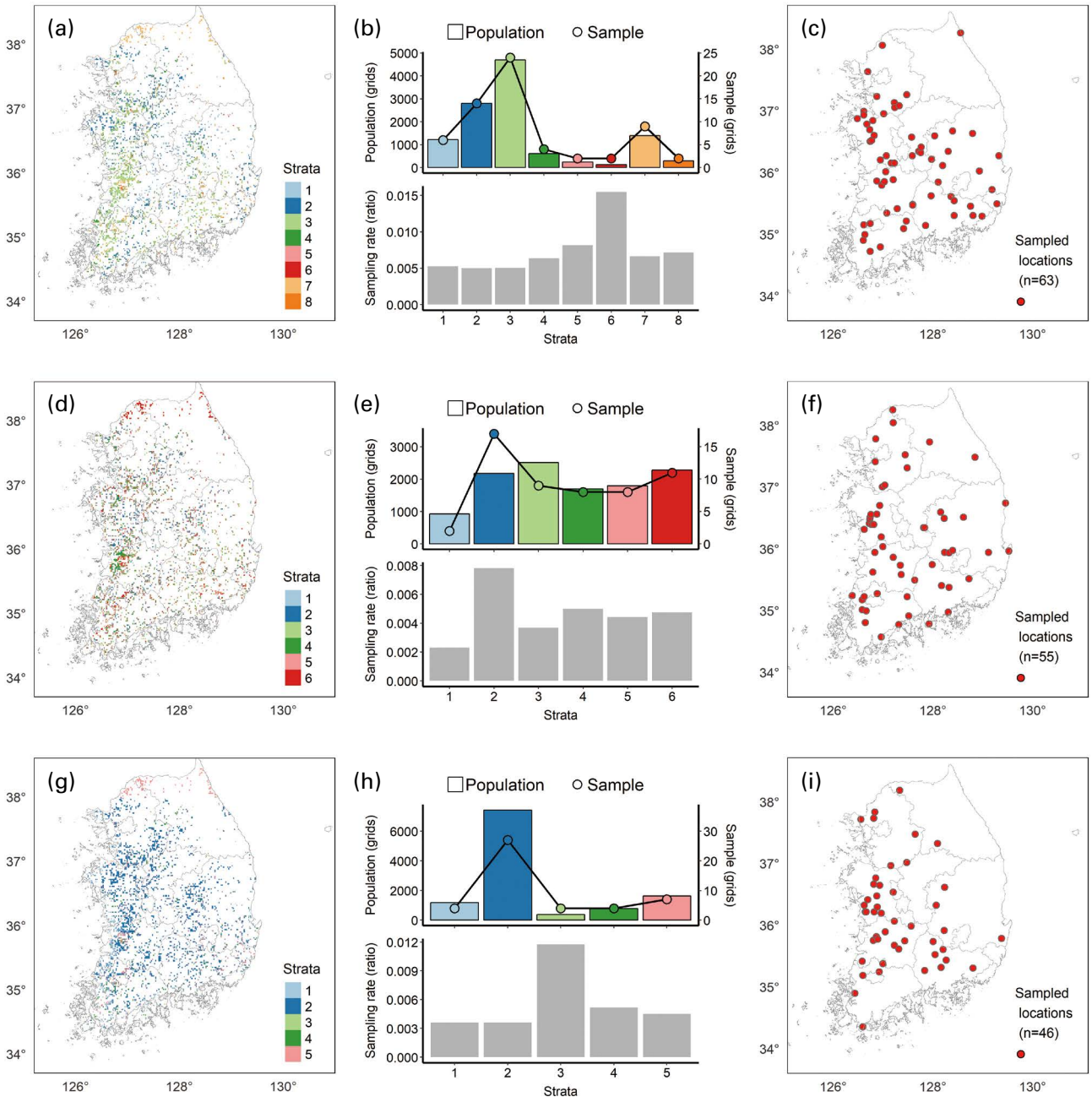

Fig. 1. Results of stratified random sampling design for domestic paddy fields (11,386 grids with 1 km spatial resolution) by years (2030s/2050s/2080s; each presented by first/second/third rows) under RCP 4.5 scenario. Panels in the first, second, and third columns present results of optimized stratification ((a), (d), and (g)), sample allocation, sampling rate (top and bottom, respectively) ((b), (e), and (h)), and examples of sampled locations ((c), (f), and (i)), respectively.

and g). 많은 수의 계층은 농지 특성 구분의 세분화를 의미하며, 계층화 결과들은 RCP 8.5 시나리오에서 국내 농지가 RCP 4.5 시나리오에서보다 기후 및 토양 특성에 따른 구분이 더 세분화되는 경향이 있음을 보여주었다. 시간에 따라서는 2080년대로 갈수록 RCP 8.5 시나리오에서는 계층 수가 증가, RCP 4.5 시나리오에서는 계층 수가 감소하는 경향이 있었다(Table 3, Fig. 1a, d, and g, and Fig. 2a, d, and g).

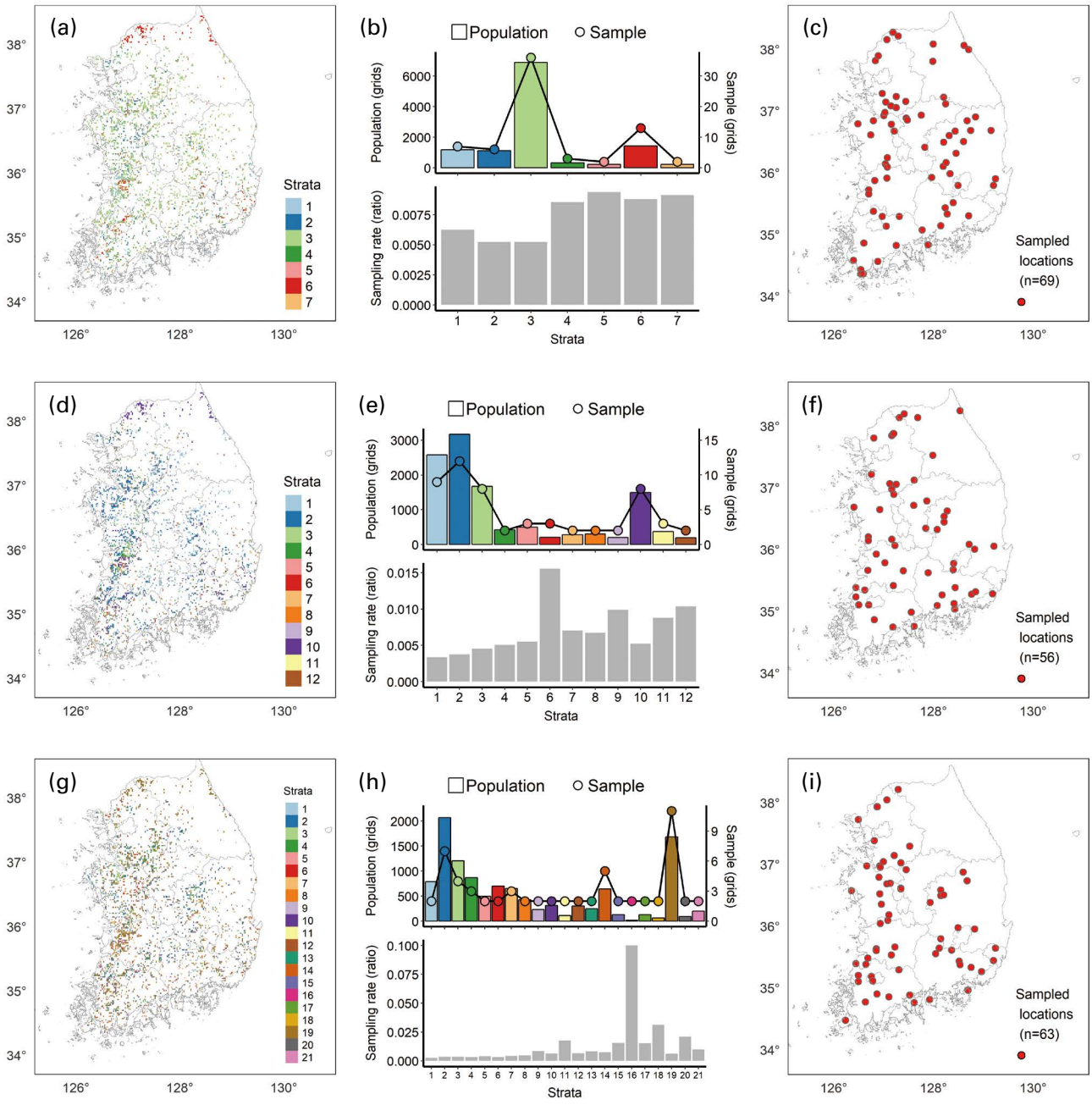

Fig. 2. Results of stratified random sampling design for domestic paddy fields (11,386 grids with 1 km spatial resolution) by years (2030s/2050s/2080s; each presented by first/second/third rows) under RCP 8.5 scenario. Panels in the first, second, and third columns present results of optimized stratification ((a), (d), and (g)), sample allocation, sampling rate (top and bottom, respectively) ((b), (e), and (h)), and examples of sampled locations ((c), (f), and (i)), respectively.

온실가스 배출이 현재 추세대로 진행될 경우 국내 농지 특성화는 더 다양해질 것으로 예상되며, 온실가스 저감정책이 상당히 실현될 경우 반대로 농지 특성화가 더 단순화될 것으로 예상된다. 본 연구에서는 시간에 따라 토양 조건은 일정하게 유지되는 것으로 가정하였다. 따라서, 시간에 따른 계층 수 변화는 기후인자들의 영향으로 해석될 수 있다. 계층 수 증가는 기상인자들이 더 넓은 범위를 갖게 되거나 패턴이 다양해지면서 나타날 수 있다 (Ballin and

Barcaroli 2013). 본 연구에서 계층 수 증가는 내륙 지방의 계층 세분화에 의해 일어났으며, 경기 및 강원 북부 지역의 계층화는 비교적 일정하게 유지되는 경향이 있었다(Fig. 2a, d, and g). 이에 따라, RCP 8.5 시나리오에서 시간에 따라 농지 계층화가 세분화되는 현상은 내륙 지방의 기후인자들의 범위 확대 및 패턴 다양화에 따른 영향으로 해석될 수 있다. 추후 연구들에서는 각 계층의 기상 및 토양 특성에 대한 추가 분석을 수행함으로써 계층 세분화에 영향을 준 주요 특성 인자들이 무엇인지를 파악할 필요가 있다.

샘플 배정 및 샘플 수 최적화 결과, 국내 농지에 대한 공간샘플링은 모집단(11,386개 격자)의 0.6% 이하 샘플 수(46~69개) 수준에서 최적화되었다(Table 1, Fig. 1c, f, and i, and Fig. 2c, f, and i). 공간샘플링의 정밀도 및 정확도는 각각 100번, 10번의 층화랜덤샘플링을 수행하여 평가하였으며(Fig. 3, Tables S1~4), 정밀도 평가 결과 작물수량을 제외한 전체 목표변수들은 정밀도 제한 범위(변동계수 0.05) 내의 값을 가졌다(Fig. 3a, c, Tables S1, 3). 또한, 100번 및 10번의 샘플링을 기반으로 한 각각의 평가결과는 서로 유사한 경향을 보여주었다(Fig. 3a, c). 이는 10번 정도의 낮은 횟수 샘플링에 대해서도 100번 샘플링한 것과 유사하게 충분히 높은 정밀도를 가질 수 있음을 나타낸다. 전체 목표변수들에 대한 샘플링의 정확도 평가 결과, 상대적 편향은 100번 및 10번의 샘플링에 대해 각각 약 0.004 (0.4%), 0.03 (3%) 이하의 값들을 가졌다(Fig. 3b, d, Tables S2, 4). 정밀도와 달리 정확도는 100번 정도의 높은 횟수의 샘플링 시 낮은 횟수 샘플링에 비해 약 10배 정도 더 높은 정확도를 가질 수 있는 것을 확인할 수 있었다. 한편, 10번의 샘플링에서 대부분의 목표 변수들은 기후시나리오에 관계없이 상대적 편향 0.01 (1%) 이하 수준의 높은 정확도를 나타냈으나, 일부 목표 변수들(i.e., y1 (작물수량), y6 (N_2O 배출량), y7 (벼멸구 분포확률))은 RCP 8.5 시나리오에서 샘플추정량의 모집단 평균에 대한 예측력이 상대적으로 낮은 것으로 나타났다(Fig. 3d). 하지만, 이들 목표 변수들 역시 100번의 샘플링에서는 기후시나리오에 관계없이 모두 0.003 (0.03%) 이하로 낮은 수준의 상대적 편향을 갖는 것으로 확인되었다(Fig. 3b). 이에 따라, y1, y6, y7에

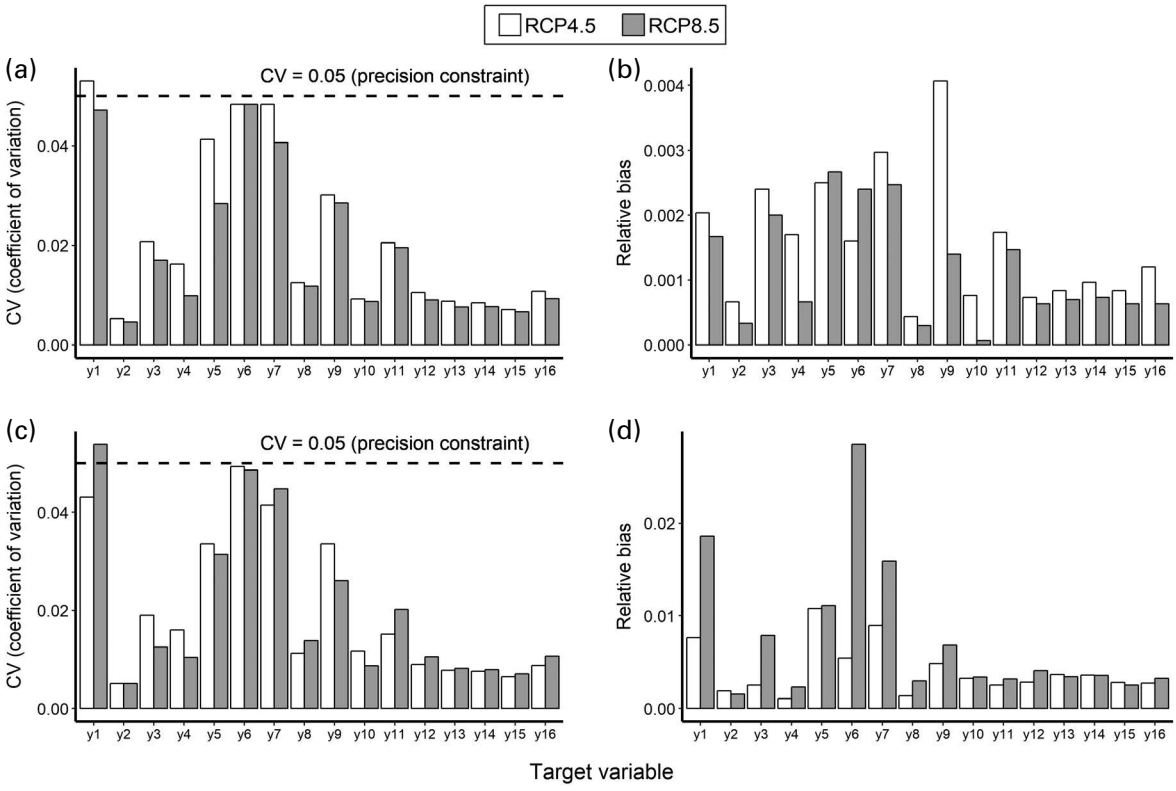

Fig. 3. Evaluation of precision ((a) and (c)) and accuracy ((b) and (d)) of spatial sampling based on 100 ((a) and (b)) and 10 ((c) and (d)) stratified random sampling simulations. Evaluations were conducted for all 16 target variables. Descriptions of target variables are provided in Table 2. The coefficient of variation (CV) and relative bias are used as indicators of precision and accuracy evaluation, respectively.

대한 특히 높은 정확도를 요구하는 샘플링이 필요한 경우에는, 샘플링 횟수를 증가시킴으로써 원하는 수준의 정확도를 얻을 수 있을 것으로 사료된다.

농지 공간샘플링에 관한 기존 연구들에서는 적절한 샘플 수를 결정하기 위해 샘플링 수를 변경해가며 샘플링의 정확도를 평가하고, 정확도 평가 결과에 따라 최적의 샘플 수를 결정하는 방식을 사용해왔다(van Bussel *et al.* 2016; Zhao *et al.* 2016). 본 연구에서는 최적화 알고리즘을 통해 최적의 샘플 수를 결정하는 방식을 사용함으로써 수동으로 샘플 수를 결정해야 하는 번거로움을 덜 수 있었다. 한편, 공간샘플링 결과들은 추후 활용 목적에 따라 더 많은 비용 절감이 필요할 경우 최적화 과정의 반복(Iteration) 수 증가를 통해 샘플 수를 감소시켜 활용할 수 있으며 또는 더 높은 정확도가 필요할 경우에는 샘플 수 조정(Adjustment)을 통해 샘플 수를 증가시켜 활용하는 것이 가능하다(Ballin and Barcaroli 2013). 최적화 기반의 공간샘플링은 유동성 및 효율성이 높으며, 따라서 많은 연구들에서 공간샘플링 설계 시 효과적으로 활용될 수 있을 것으로 생각된다.

많은 통계조사연구들에서 샘플 수의 결정은 샘플링 비용 절감 측면에서 최대의 관심사였다(Cochran 1977). 최근에는 GIS 기술의 발전에 따라 공간자료를 활용한 연구들이 증가하고 있으며, 이에 따라 공간모델링 분야에서 역시 비용 절감을 위한 공간샘플링의 설계 및 샘플 수 최적화 문제가 주요 관심사로 떠오르고 있다(Wang *et al.* 2010). 더욱이, 4차 산업혁명 및 빅데이터 시대가 도래하면서, 축적된 공간자료의 효과적·효율적 활용을 위한 적절한 공간샘플링의 설계는 필수적으로 요구되고 있다(Goyal *et al.* 2017). 본 연구의 국내 고해상도 농지 공간격자자료의 공간적 계층화 및 샘플 수 최적화 결과는 농업분야 내 기후변화 공간예측모형 연구들에 활용되어 시뮬레이션 비용 절감 및 계산 시간 단축에 기여할 수 있을 것으로 기대된다. 뿐만 아니라 모형 개발에 활용될 경우, 개발모형의 구축 효율 및 성능 향상에도 기여할 수 있을 것으로 기대된다.

본 연구의 공간샘플링 설계 시 농업시스템 내 작물, 해충, 온실가스 분야 대표 모형들의 입출력변수들을 활용하여 층화 변수와 목표 변수를 구성하였으며, 이로써 샘플링 격자를 활용한 농업시스템의 결과물들에 대한 전반적인 예측이 가능하도록 하였다. 기존 농업연구들에서 공간샘플링은 주로 단일 목표 변수에 대한 예측을 목적으로 하는 공간모델링 연구들에서 활용되어 왔으며, 특히 작물수량 예측을 위한 공간모델링이 주를 이루어 왔다(Bussel *et al.* 2016; Zhao *et al.* 2016). 기존 연구들에서의 샘플링 격자들은 해당 변수에 대한 예측 능력만을 가졌기 때문에 여러 분야에서 활용되기에는 어려움이 있었다. 하지만, 실제 농업시스템은 작물생산 이외에도 다양한 결과물과 부산물을 동반하며 농업 분야에서의 공간샘플링 결과의 폭넓은 연구적 활용을 위해서는 농업 내 다양한 부문에서 보편적 활용이 가능한 공간샘플링을 설계할 필요가 있다. 이러한 관점에서 본 연구의 결과물들은 국내 농지를 두루 대변하는 공간샘플링으로써 농업 내 다양한 분야의 공간모델링 연구들에서 대표 샘플 격자로서의 폭넓은 활용이 가능할 것으로 기대된다.

적 요

공간 샘플링은 공간모델링 연구에 활용되어 샘플링 비용을 줄이면서 모델링의 효율성을 높이는 역할을 한다. 농업분야에서는 기후변화 영향을 예측하고 평가하기 위한 고해상도 공간자료 기반 모델링에 대한 연구 수요가 빠르게 증가하고 있으며, 이에 따라 공간 샘플링의 필요성과 중요성이 증가하고 있다. 본 연구는 국내 농지 공간샘플링 연구를 통해 농업분야 기후변화연구의 공간자료 활용의 효율성을 제고하고자 하였다. 본 연구는 층화랜덤샘플링을 기반으로 하였으며, 1 km 해상도의 농지 공간격자자료 모집단(11,386개 격자)에 대해서 RCP 시나리오별(RCP 4.5/8.5) 연대별(2030/2050/2080년대) 공간샘플링을 설계하였다. 국내 농지는 기상 및 토양 특성에 따라 계층화되었으며, 샘플링 효율 극대화를 위해 최적 층화 및 샘플 배정 최적화를 수행하였다. 최적화는 작물수량, 온실가스 배출량, 해충 분포 확률을 포함하는 16개 목표 변수에 대해 주어진 정밀도 제한 내에서 샘플 수를 최소화하는 방향으로 진행되었다. 샘플링의 정밀도와 정확도 평가는 각각 변동계수(CV)와 상대적 편향을 기반으로 하였다. 국내 농지 공간격자 모집단 계층화 및 샘플 배정 및 샘플 수 최적화 결과, 전체 농지는 5~21개 계층, 46~69개 샘플 수 수준에서 최적화되었다. 본 연구결과물들은 국내 농업시스템 대표 공간격자로써 널리 활용될 수 있을 것으로 기대된다. 또한, 기후변화 영향예측 공간모델링 연구들에 활용되어 샘플링 비용 및 계산 시간을 줄이면서도 모델의 효율성을 높이는 데에 기여할 수 있다.

사 사

본 연구는 과학기술정보통신부의 재원으로 한국연구재단의 지원(NRF-2019R1A2C1009812)을 받아 수행된 연구입니다. 자료를 제공해주신 서울대학교 작물생태정보연구실 김광수 교수님, 현신우 연구원님, 유병현 연구원님, 고려대학교 식물환경학 실험실 김정규 교수님, 민현기 연구원님, 토양환경 및 오염물질 제어 실험실 현승훈 교수님, 황원재 연구원님께 깊은 감사를 드립니다.

REFERENCES

Aoyama H. 1954. A study of stratified random sampling. Ann. Inst. Stat. Math. 6:1–36.

Ballin M and G Barcaroli. 2013. Joint determination of optimal stratification and sample allocation using genetic algorithm. Surv. Methodol. 39:369–393.

Bethel J. 1989. Sample allocation in multivariate surveys. Surv. Methodol. 15:47–57.

Cochran WG. 1977. Sampling Techniques, 3rd ed. Wiley. New York.

Folberth C, H Yang, X Wang and KC Abbaspour. 2012. Impact of input data resolution and extent of harvested areas on crop yield estimates in large-scale agricultural modeling for maize in the USA. Ecol. Model. 235:8–18.

Gonzalez JM and JL Eltinge. 2010. Optimal survey design: A review. pp. 4970–4983. In: Section on Survey Research Methods - JSM. American Statistical Association. Alexandria, VA.

Goyal H, C Sharma and N Joshi. 2017. An integrated approach of GIS and spatial data mining in big data. Int. J. Comput. Appl. 169:1–6.

Hatfield JL, J Antle, KA Garrett, RC Izaurralde, T Mader, E Marshall, ... and L Ziska. 2020. Indicators of climate change in agricultural systems. Clim. Change 163:1719–1732.

Hijmans RJ, SE Cameron, JL Parra, PG Jones and A Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965–1978.

Joo YS, HJ Jung and BJ Kim. 2009. Cluster analysis with Korean weather data: Application of model-based Bayesian clustering method. J. Korean Data Inf. Sci. Soc. 20:57–64.

McCallion T. 1992. Optimum allocation in stratified random sampling with ratio estimation as applied to the northern Ireland December agricultural sample. J. R. Stat. Soc. Ser. C-Appl. Stat. 41:39–45.

Metzger MJ, RG Bunce, RH Jongman, R Sayre, A Trabucco and R Zomer. 2013. A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. Glob. Ecol. Biogeogr. 22:630–638.

Moore FC, ULC Baldos and T Hertel. 2017. Economic impacts of climate change on agriculture: a comparison of process-based and statistical yield models. Environ. Res. Lett. 12:065008.

Perlman J, RJ Hijmans and WR Horwath. 2014. A metamodelling approach to estimate global N_2O emissions from agricultural soils. Glob. Ecol. Biogeogr. 23:912–924.

Schmitt LM. 2001. Theory of genetic algorithms. Theor. Comput. Sci. 259:1–61.

Stein A and C Ettema. 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. Agric. Ecosyst. Environ. 94:31–47.

Tonnang HE, BD Hervé, L Biber-Freudenberger, D Salifu, S Subramanian, VB Ngowi, ... and C Borgemeister. 2017. Advances in crop insect modelling methods - Towards a whole system approach. Ecol. Model. 354:88–103.

Van Bussel LG, F Ewert, G Zhao, H Hoffmann, A Enders, D Wallach, ... and F Tao. 2016. Spatial sampling of weather data for regional crop yield simulations. Agric. For. Meteorol. 220:101–115.

Wang JF, RP Haining and ZD Cao. 2010. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. Int. J. Geogr. Inf. Sci. 24:523–543.

Wang JF, A Stein, BB Gao and Y Ge. 2012. A review of spatial sampling. Spat. Stat. 2:1–14.

Yeo IK. 2011. Clustering analysis of Korea's meteorological data. J. Korean Data Inf. Sci. Soc. 22:941–949.

Zhang J, H Tian, H Shi, J Zhang, X Wang, S Pan and J Yang. 2020. Increased greenhouse gas emissions intensity of major croplands in China: Implications for food security and climate change mitigation. Glob. Change Biol. 26:6116–6133.

Zhao G, H Hoffmann, J Yeluripati, S Xenia, C Nendel, E Coucheney, ... and F Ewert. 2016. Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. Environ. Model. Softw. 80:100–112.