On Estimating the CO_2 Fluxes for Whole Agricultural and Forest Area in South Korea using Flux Measurement-based Data-driven Technology

Sungsik Cho[1, 2], Minseok Kang[1*], Kazuhito Ichii[3], Joon Kim[1, 2, 4], Jung-Hwa Chun[5], Hyun Seok Kim[1, 2, 6], Sung-Won Choi[1], Yohana M. Indrawati[2], Seung-Hoon Lee[2], and Jongho Kim[1]

[1]National Center for AgroMeteorology, South Korea, [2]Interdisciplinary Program in Agricultural & Forest Meteorology, Seoul National University, South Korea, [3]Center for Environmental Remote Sensing, Chiba University, Japan, [4]Department of Landscape Architecture and Rural Systems Engineering/ Institutes of Green Bio Science and Technology, Seoul National University, South Korea, [5]National Institute of Forest Science, South Korea, [6]Department of Forest Science, Seoul National University, South Korea

(Corresponding email : ms-kang@ncam.kr)

Problem statement

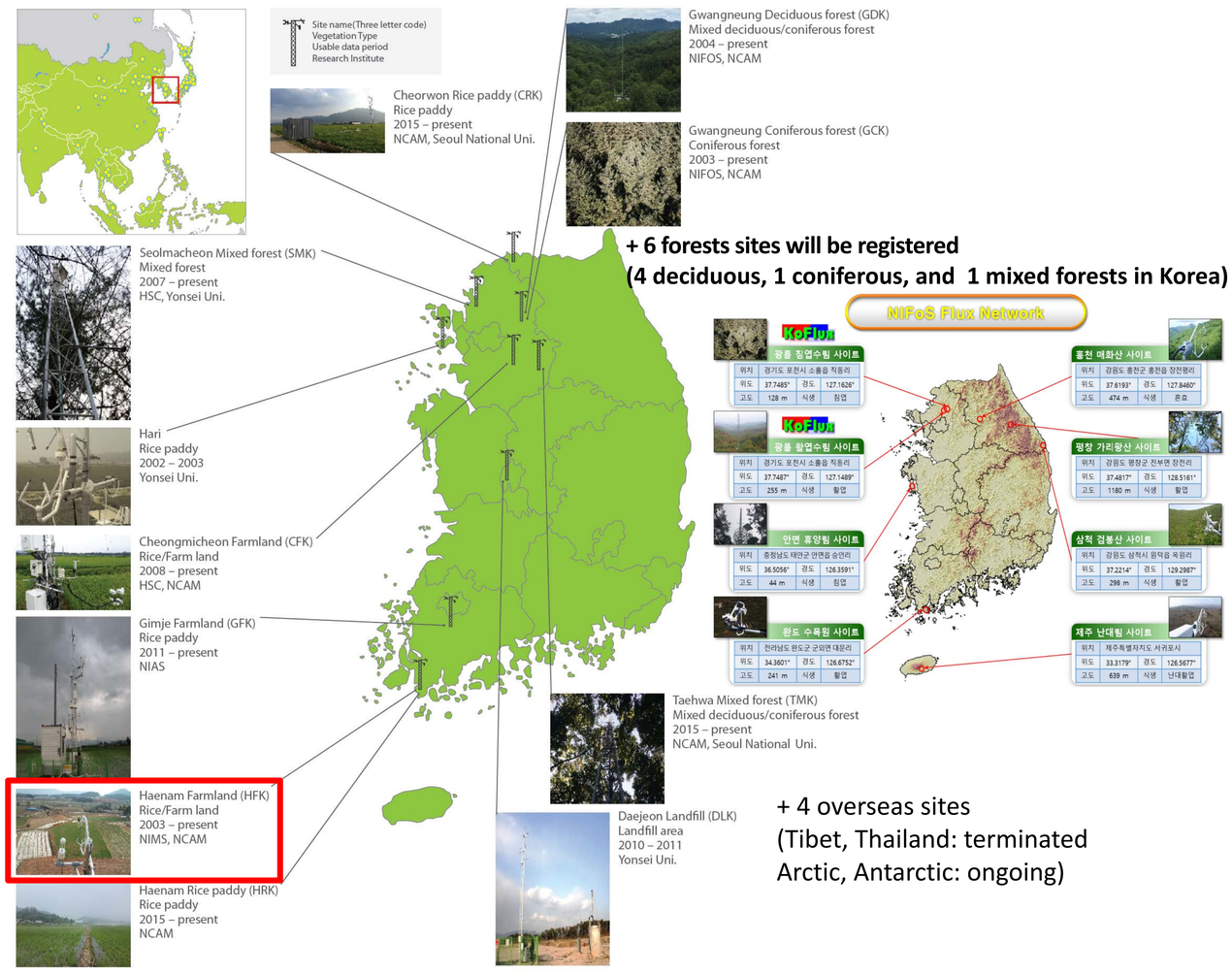

Fig. 1. KoFlux Network.

- KoFlux is a regional Korean network of long-term micrometeorological flux measurement sites, launched in 2002, to measure the CO_2/H_2O/energy flux of Korean agricultural lands and forests.
- The eddy covariance method that directly observes the net ecosystem exchange of mass and energy between the surface and the atmosphere can be alternative.
- However, the eddy covariance-based monitoring systems are expensive to establish and maintain, so the number of observation systems have to be limited. Therefore, the methods of upscaling via linkage between observation data and satellite data (or modeling) are widely used.
- We intend to estimated the carbon uptake of the whole agricultural and forest area in South Korea using the flux data from the observation network and data-driven approach.

Objective

- To reliably estimate the CO_2 fluxes for whole agricultural and forest area in South Korea using the flux data from the observation networks and the data-driven approach (i.e., the scaling technology via the linkage between satellite images and in-situ measurements using machine learning technique).

Experimental design

- Machine learning algorithm: Support vector regression (Ichii et al., 2017)
- Target variables: (Daily) ET (mm), GPP (g C m^{-2} s^{-1}), RE (g C m^{-2} s^{-1}), NEE (g C m^{-2} s^{-1})
- Machine learning algorithm: Support vector regression (Ichii et al., 2017)
- Input variables: Downward shortwave radiation, Air temperature (Tair, daytime), Vapor pressure deficit (VPD, daytime), Precipitation (PPT), Soil water content (SWC or LSWI), Leaf area index (LAI), Fuzzy transformation of the cultivation.

Results (case study 1) – Site level estimation in HFK (Temporal extrapolation)

- In KoFlux, the Haenam Farmland (HFK) site has the longest record (from July 2002 to present) of carbon/water/energy flux measurement using eddy covariance technique in agricultural lands including rice paddy, beans and sesame fields.
- The long-term database at the HFK is vital to better understand how the farmlands have adapted and been managed with natural and/or human disturbances at various time and spatial scales. Unfortunately, there are long gaps particularly in 2007 and 2014, which hinder the researchers from analyzing the decade-longtime series data.

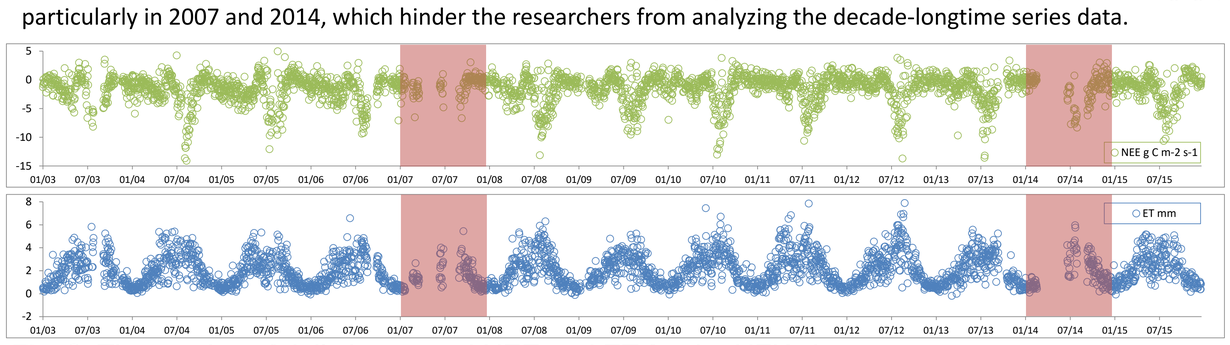

Fig. 2. Time series of daily integrated NEE and ET for the HFK site.

- The general gap-filling method is impractical to apply for such long gaps. Recently, data-driven approach (i.e., interpolating/extrapolating EC measurements via available networks temporally/ spatially using machine learning technique) is used to estimate terrestrial CO_2/H_2O fluxes. Such an approach can be applied to our case after appropriate modifications.
- In this case study, we evaluate an applicability of data-driven approach to the filling of long gaps in flux data (i.e., gross primary production (GPP), ecosystem respiration (RE), net ecosystem exchange (NEE), and evapotranspiration (ET)).

Table 1. Hypotheses and experimental design.

Hypothesis	Exp. no.	Target year	Training year	Input source
Hyp. 1: Estimation using in-situ measurement data for input to machine learning is more reasonable than that using remote sensing (or modeling) data	1-1	2010	2009, 2011	In-situ
	1-2			RS (and model)
Hyp. 2: Closer (to gaps) training dataset for machine learning results in better estimation	2-1	2009	2008, 2010	In-situ
	2-2		2006, 2011	
	2-3		2005, 2012	
	2-4		2004, 2013	
Hyp. 3: Longer training dataset for machine learning results in better estimation	3-1	2009	2008 (or 2010)	In-situ
	3-2		2008, 2010	
	3-3		2006, 2008, 2010, 2011	
	3-4		2005, 2006, 2008, 2010, 2011, 2012	
	3-5		2004, 2005, 2006, 2008, 2010, 2011, 2012, 2013	

Hyp. 1: Using In-situ input data performs better than using RS input data — Accepted

Table 2. Statistical parameters for error assessment of the experiments 1-1 and 1-2. MBE and RMSE indicate mean bias error and root mean square error, respectively. Slope and r^2 are from linear regression analysis.

Exp. 1-1	ET	GPP	RE	NEE1*	NEE2*	Exp. 1-2	ET	GPP	RE	NEE1	NEE2
MBE	-0.13	-0.11	-0.14	-0.01	-0.03	MBE	-0.19	-0.09	-0.14	0.07	-0.05
RMSE	0.47	1.17	0.45	1.32	1.22	RMSE	0.92	1.57	0.57	1.75	1.57
Slope	0.91	0.91	0.98	0.52	0.58	Slope	0.79	0.88	0.97	0.28	0.41
r^2	0.86	0.83	0.91	0.63	0.68	r^2	0.30	0.67	0.85	0.31	0.44

*NEE1: Directly estimated, NEE2: Estimated RE – Estimated GPP

Hyp. 2: Closer (to gap) training dataset results in better estimation — Rejected

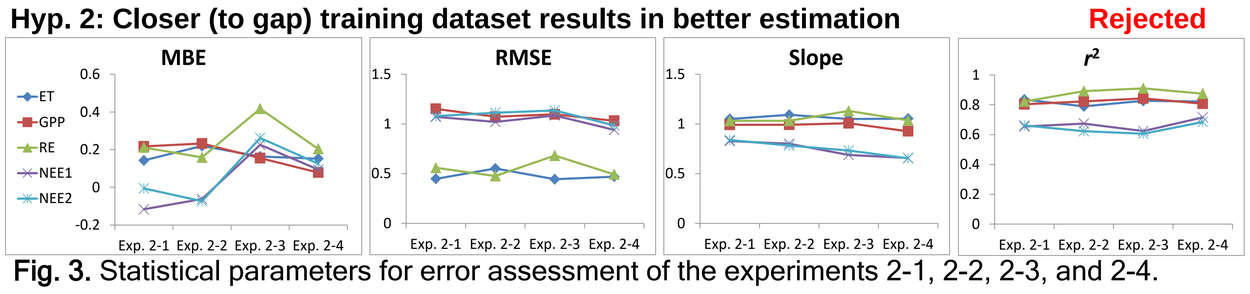

Fig. 3. Statistical parameters for error assessment of the experiments 2-1, 2-2, 2-3, and 2-4.

Hyp. 3: Longer training dataset results in better estimation — Accepted

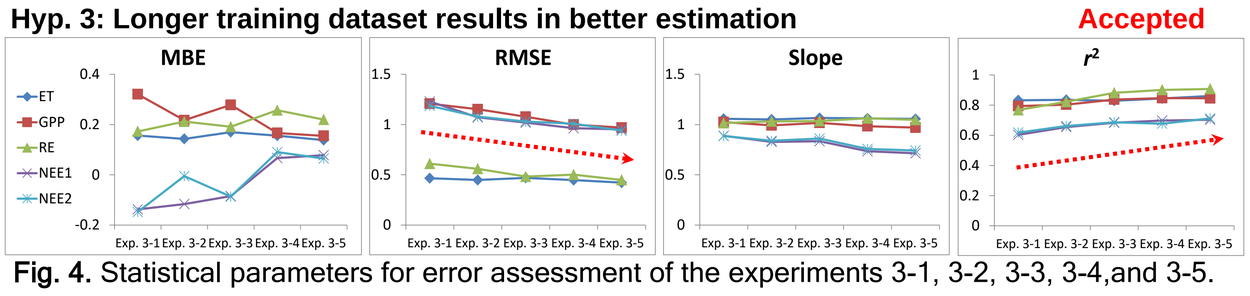

Fig. 4. Statistical parameters for error assessment of the experiments 3-1, 3-2, 3-3, 3-4,and 3-5.

Results (case study 2) – Estimate CO_2 fluxes for whole agricultural and forest area in South Korea (Spatial extrapolation)

- The eddy covariance via available networks can be used to estimate terrestrial CO_2 fluxes by data-driven approach, and it is one of the most effective ways to estimate continental-scale fluxes.
- In this case study, we estimated the CO_2 fluxes for whole agricultural and forest area in South Korea using the flux data from the observation networks and data-driven approach.

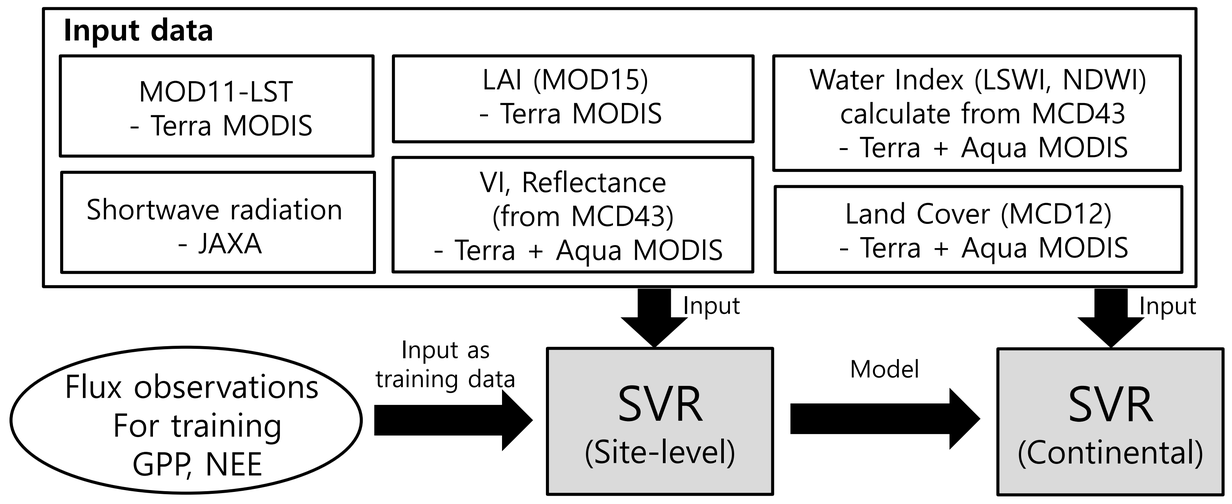

Table 3. List of input parameters of the GPP and NEE models

Target variable	Experiment case	Shortwave radiation	LST	LAI	Vegetation index	Water index	Reflectance & other
GPP	1	○	LSTday	○	NDVI	LSWI	
	2	○	LSTday	○	EVI	LSWI	
	3	○	LSTday	○	NDVI	NDWI	
	4	○	LSTday	○	EVI	NDWI	
	5	○	LSTday				B1-B7
NEE	1	○	LSTave	○	NDVI	LSWI	GPPest
	2	○	LSTave	○	EVI	LSWI	GPPest
	3	○	LSTave	○	NDVI	NDWI	GPPest
	4	○	LSTave	○	EVI	NDWI	GPPest
	5	○	LSTave				GPPest, B1-B7

Table 4. Performance of each SVR

GPP			NEE		
Case	r^2	RMSE	Case	r^2	RMSE
1	0.79	1.20	1	0.81	0.80
2	**0.81**	**1.22**	**2**	**0.83**	**0.75**
3	0.80	1.31	3	0.81	0.79
4	0.81	1.27	4	0.80	0.82
5	0.78	1.30	5	0.80	0.80

- The SVR model were established for estimating GPP and NEE by using input parameters with constraints by GPP and NEE from site observations (table 3).
- For both GPP and NEE predictions at the 8 day temporal scale (table 4), all model setups showed similar performances (r^2 = 0.78 – 0.81 and RMSE = 1.20 - 1.31 g C m^{-2} d^{-1} for GPP and r^2 = 0.80 - 0.83 and RMSE = 0.75 - 0.82 g C m^{-2} d^{-1} for NEE).
- The mean estimated annual GPP and NEE of the South Korea were about 1190 g C m^{-2} $year^{-1}$ and -350 g C m^{-2} $year^{-1}$, respectively.

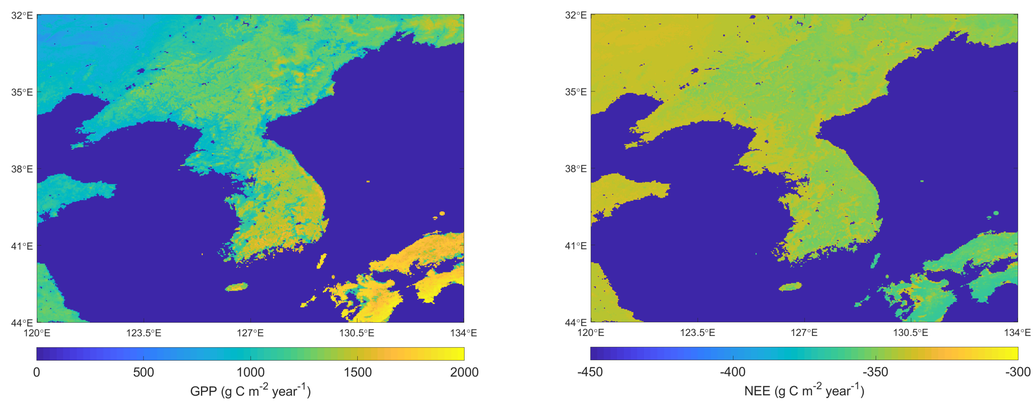

Fig. 5. Spatial distribution of GPP and NEE for whole agricultural and forest area around South Korea

Acknowledgments This work is supported by the Korea Meteorological Administration Research and Development Program under Grant KMIPA 2015-2023, by the Weather Information Service Engine (WISE) project, KMIPA-2012-0001-2, and by the R&D Program for Forest Science Technology (Project No. 2017099A00-1719-BB01) of the Korea Forest Service (Korea Forestry Promotion Institute).