New gap filling strategies for long-period flux data gaps using data-driven approach

Minseok Kang[1], Kazuhito Ichii[2], Joon Kim[1,3,4], and Yohana M. Indrawati[4]

1National Center for AgroMeteorology, Seoul, South Korea

2Center for Environmental Remote Sensing, Chiba University, Chiba, Japan

3Department of Landscape Architecture and Rural Systems Engineering/ Institutes of Green Bio Science and Technology, Seoul National University, Seoul, South Korea

4Interdisciplinary Program in Agricultural & Forest Meteorology, Seoul National University, Seoul, South Korea

(Correspondence: ms-kang@ncam.kr)

KoFlux - This year marks the 15th year anniversary.

- **KoFlux** is a Korean network of micrometeorological tower sites that use eddy covariance methods to monitor the cycles of carbon, water and energy between the atmosphere and the key terrestrial ecosystems in Korea

Photo of “KoFlux Kick-Off” during the 2nd AsiaFlux workshop held in Jeju Island, Korea in 9-11 January 2002.

Mission and Purposes

KoFlux embraces the mission of AsiaFlux

"thinking community, learning frontiers"

to bring Asia's key ecosystems under observation to ensure quality and sustainability of life on earth

The main purposes of KoFlux are to provide

(1) an infrastructure to monitor, compile, archive and distribute data for the science community and

(2) a forum and short courses for the application and distribution of knowledge and data between scientists including practitioners

Objectives

Specific objectives are

(1) to quantify carbon/water/energy cycles

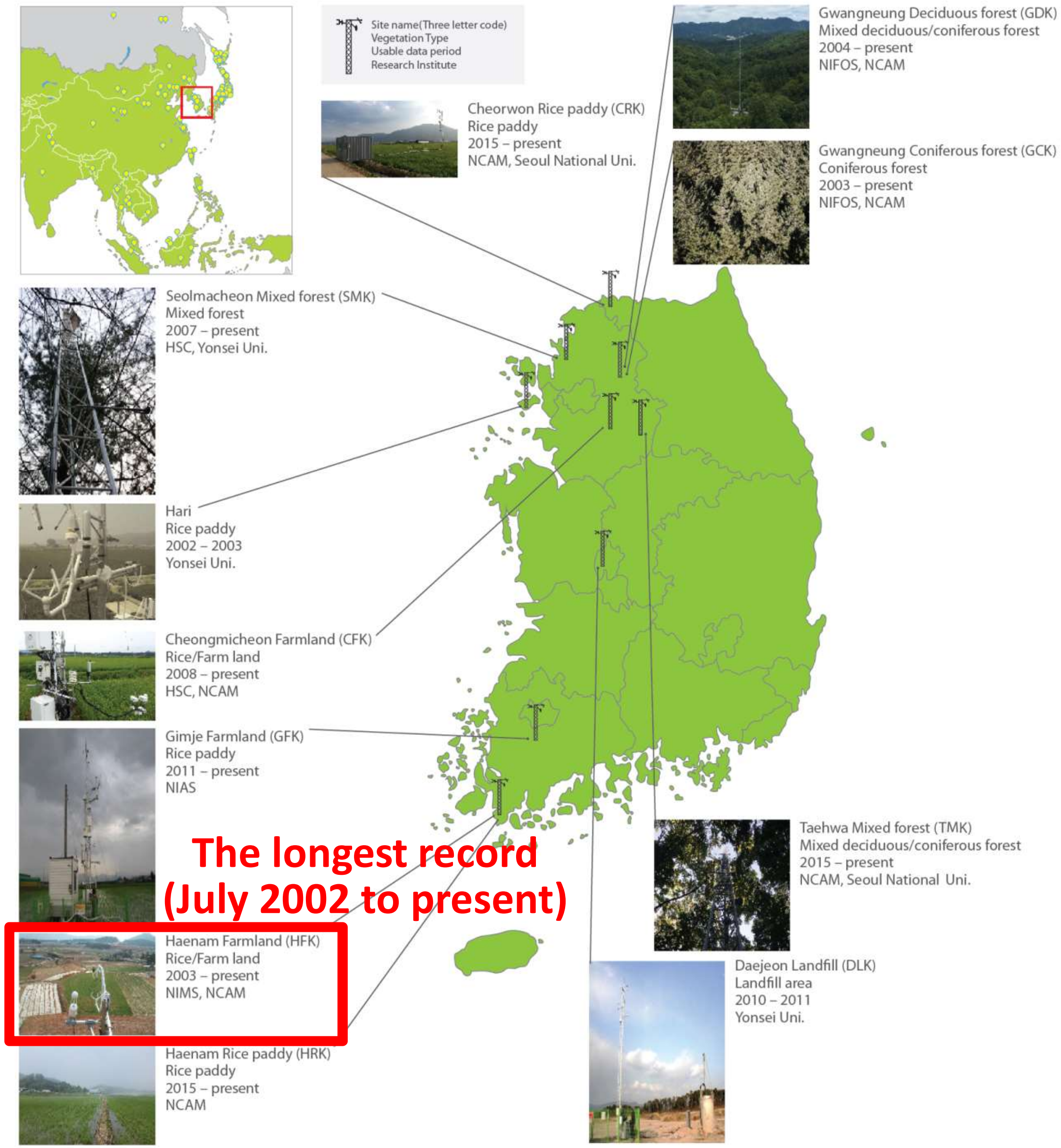
(2) to understand mechanisms and processes that drive carbon/water/energy cycles

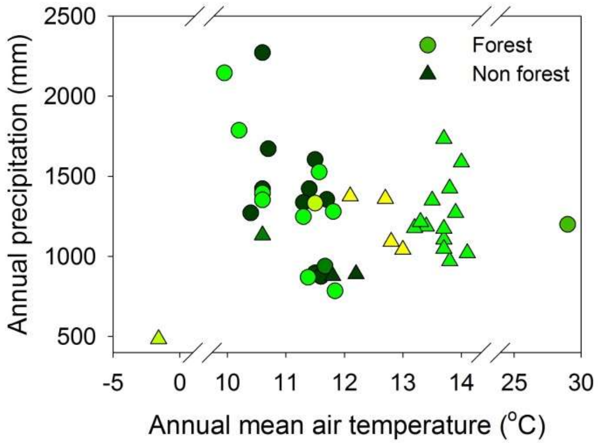
(3) to identify impacts of Asian monsoon on carbon/water/energy cycles

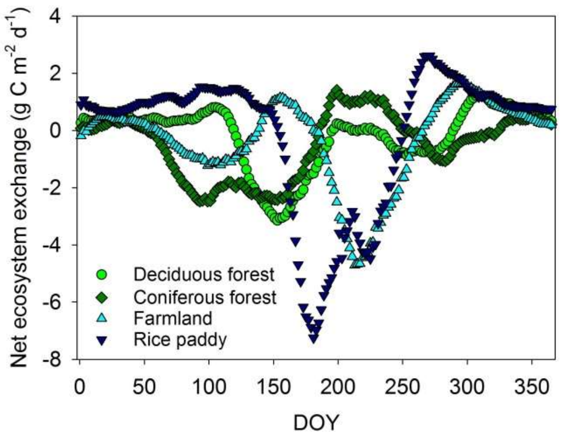
(4) to conduct research and analysis for sustainability of ecological-societal system based on resilience thinking

(5) to train professional manpower

KoFlux Network







+ 5 overseas sites
(Tibet, Thailand, Ieodo: terminated
Arctic, Antarctic: ongoing)

Next 15 years – Rural Systems Visioneering

We try to help rural off-grid villages to overcome their fundamental concerns: (1) lack of resources, (2) lack of infrastructure, (3) lack of quality education and training, and (4) lack of motivation, vision and its engineering (i.e., visioneering).

Adapted from Joon Kim's presentation, May 18 2017

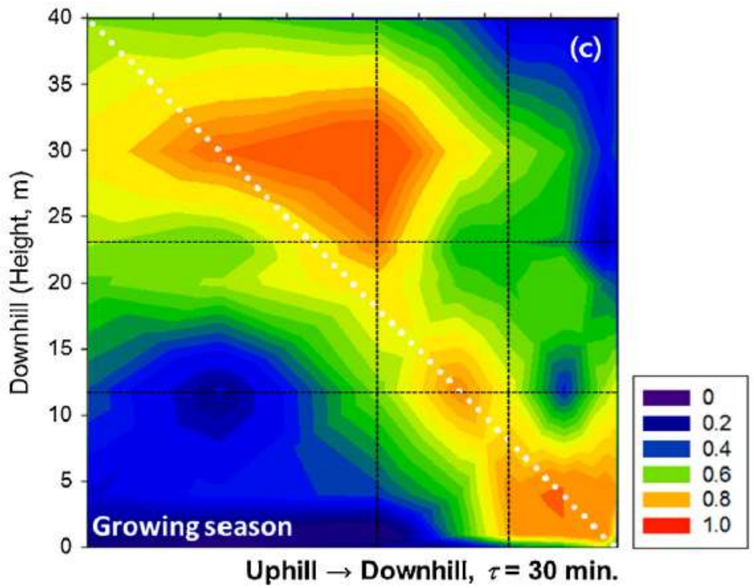
1st stage of the next 15 years – Capacity building

(1) Flux measurement '*everywhere, all of the time*'

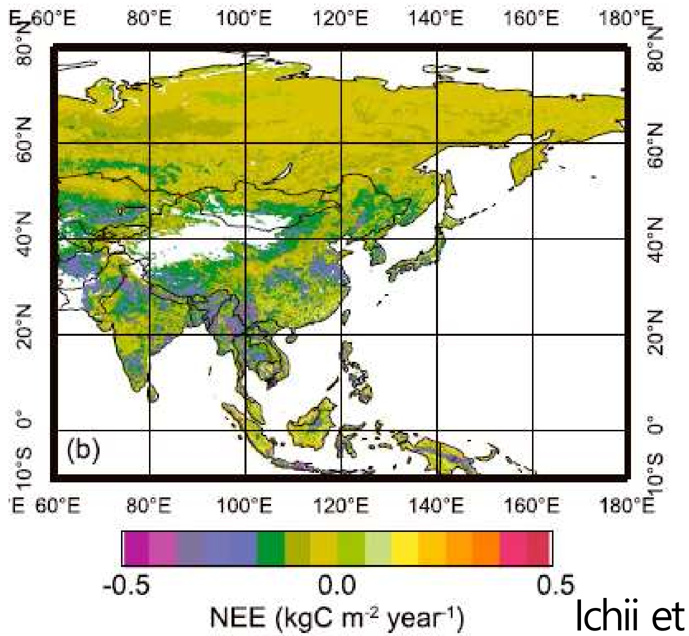
Secure measurement techniques for more various environment with higher accuracy

(2) Flux estimation '*everywhere, all of the time*'

Secure connection techniques between in-situ observation and remote sensing for extrapolating flux measurement



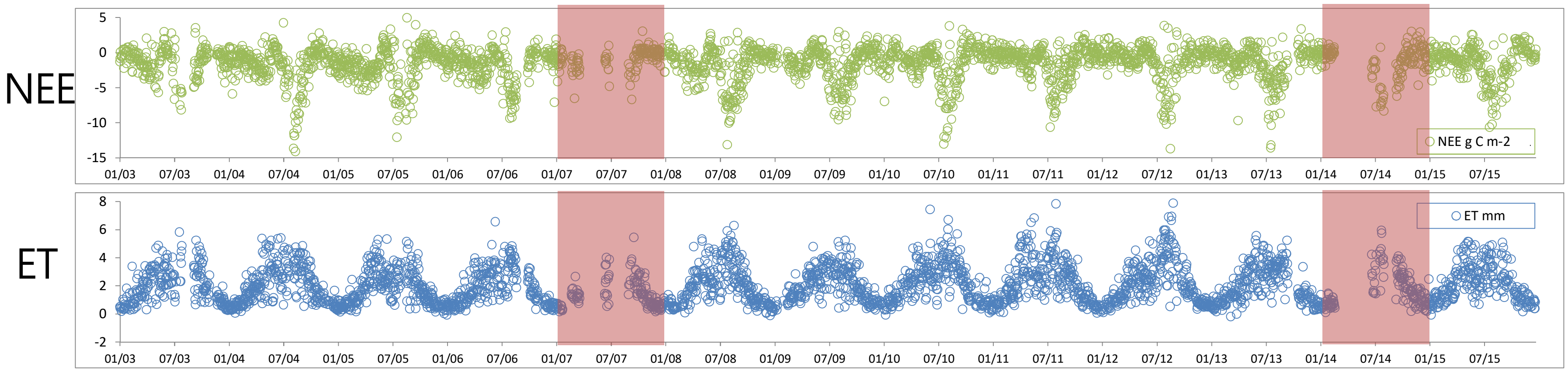
Kang et al. (2017)



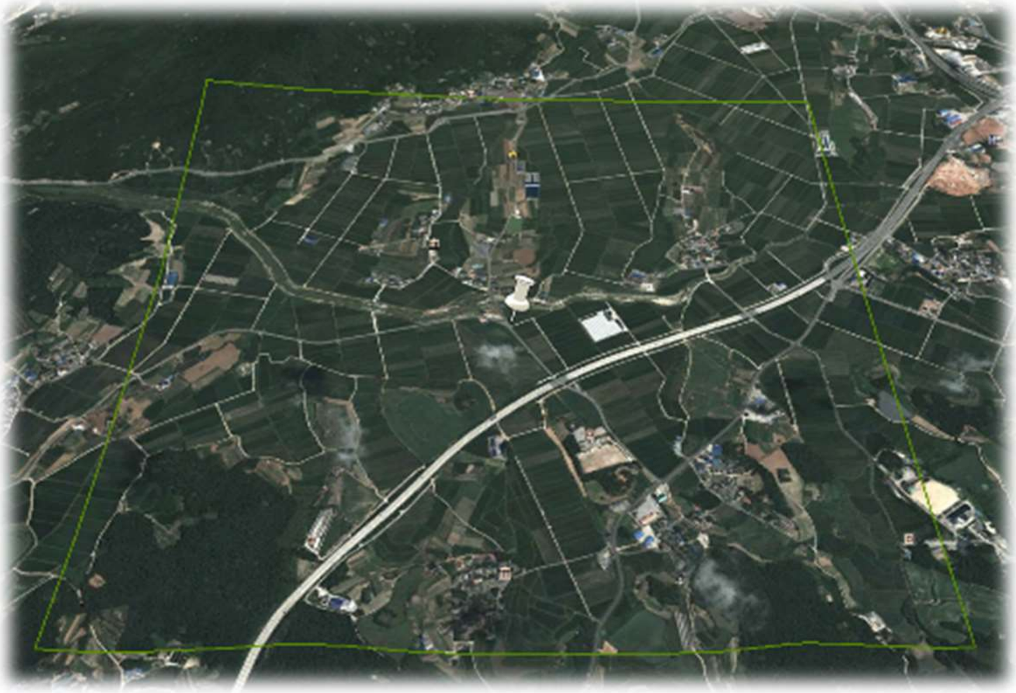
Ichii et al. (2017)

Problem statement

- The HFK site is located in a typical Korean farmland.
- 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.
- The long gaps mainly due to power break in 2007 and 2014 hinder the researchers from analyzing the decade-longtime series data.
- The general gap-filling method is impractical to apply for such long gaps.
- Data-driven approach is used to estimate terrestrial CO_2/H_2O fluxes. Such an approach can be applied to our case after appropriate modifications.
- In this presentation, we evaluate an applicability of data-driven approach to the filling of long gaps in flux data (i.e., GPP, RE, NEE, and ET).



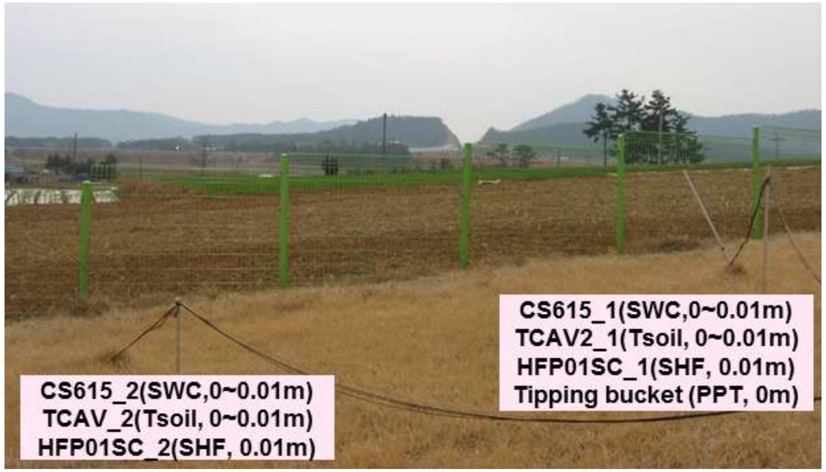
Site description



Location	Southwestern end of Korean Peninsula (34.55˚N, 126.57˚E, 13.74 m a. s. l.)
Topography	flat terrain except the southeast (slope ±4˚)
Land cover	the mixture of rice paddies and various seasonally cultivated crops within 300 meters, >300 meters rice paddies prevailed in the south and the west, scattered residential areas, roads and isolated forests.
Dominant species	Rice and seasonally cultivated crops (e.g., corn, sesame, chili)
Canopy height	1 m
Soil types	silt loam to loam (sand 38.5%, clay 30.0%)
Climate	hot-humid summer and cool-dry winter
Annual mean Temperature*	13.3 ℃
Annual Precipitation*	1306 mm

- **The HFK site is located in a typical Korean farmland which is characterized by mosaic patches of various agricultural lands.**

Measurement



Is flux observation in long-term networks actually science?

(Schmid, 2015)

- **Difference in scientific strategy**
- General observations: ‘controlled experiment' to examine questions, hypotheses and predictions. "what if...?"
- Observations in long-term networks: minimize the influence on the measurement to maximize their external validity. “what happens next?” or “how did this happen?"
- **External validity**
- Ecosystem-Atmosphere fluxes observations in long-term networks identify trends, temporal scales of variability, spatial patterns and variations.
- **Required condition**
- The data are comparable (comparability, compare "apples to apples")
- A necessary condition is the compatibility of the sensors and procedures.
- To achieve comparability and compatibility, a useful tool is standardization.

Objective / Materials and Methods

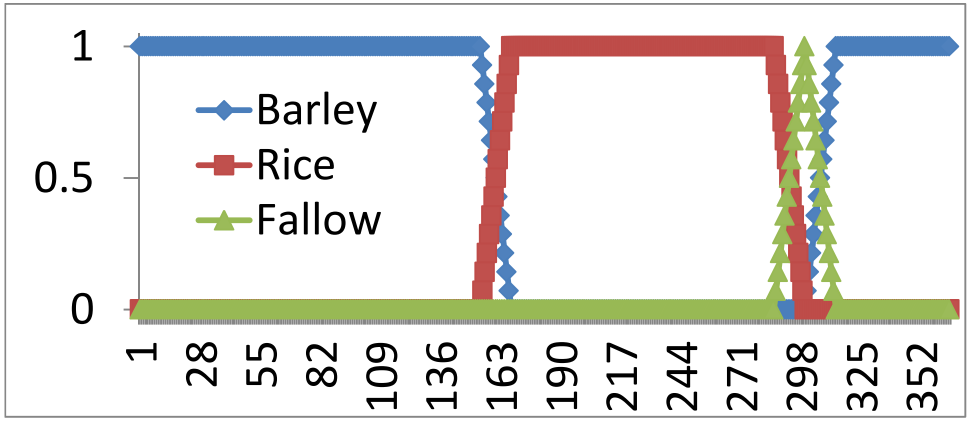
- **Objective**

  To propose a new gap-filling strategy for long-period flux data gaps
- **Materials and Methods**

  Machine learning algorithm: Support vector regression (Ichii et al., 2017)

  Target variables: (Daily) ET (mm), GPP (g C m-2), RE (g C m-2), NEE (g C m-2)

  Input drivers: Downward shortwave radiation (Rsdn), 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



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**

In-situ

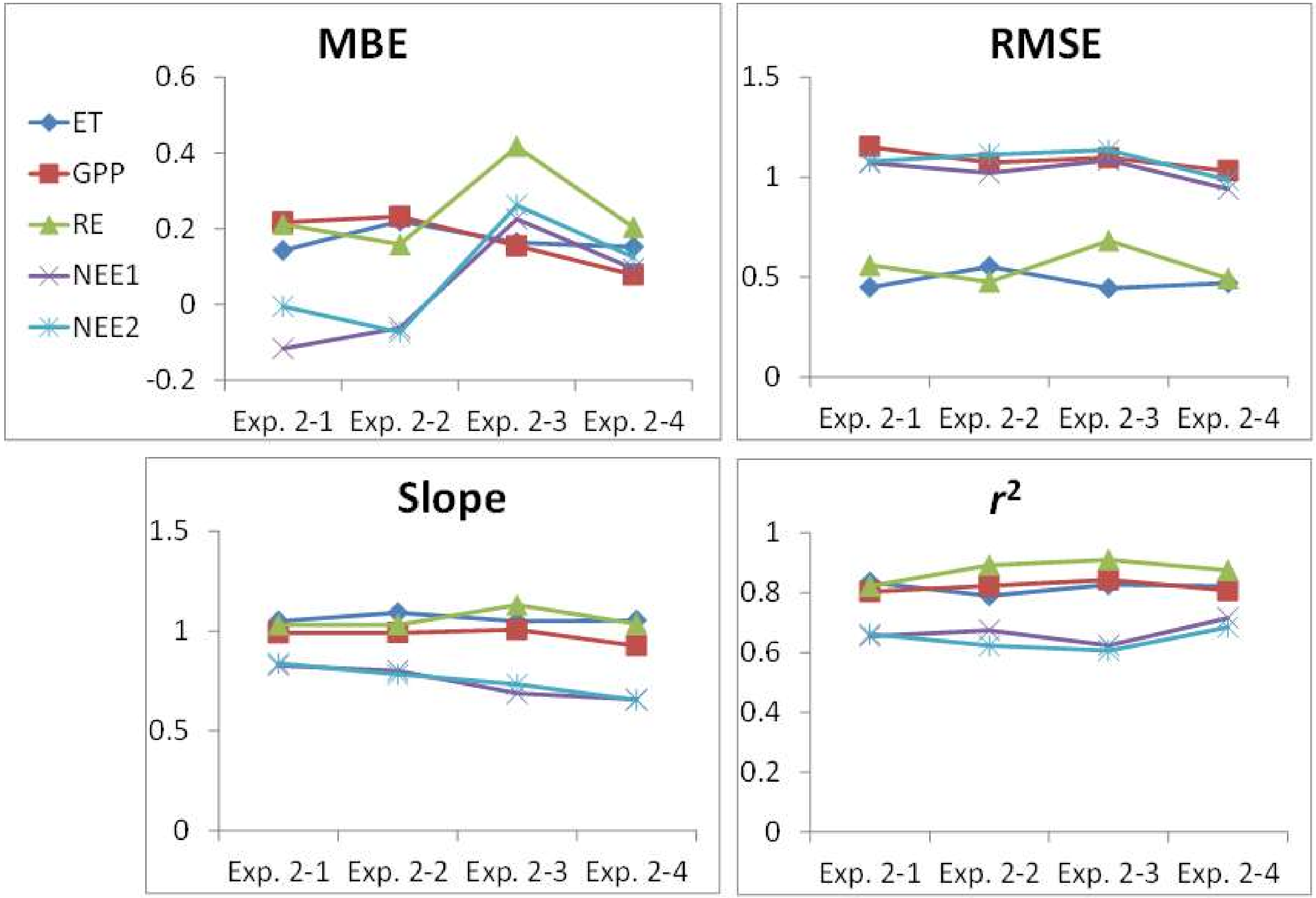
Exp. 1-1	ET	GPP	RE	NEE1*	NEE2*
MBE	-0.13	-0.11	-0.14	-0.01	-0.03
RMSE	0.47	1.17	0.45	1.32	1.22
Slope	0.91	0.91	0.98	0.52	0.58
r^2	0.86	0.83	0.91	0.63	0.68

RS

Exp. 1-2	ET	GPP	RE	NEE1	NEE2
MBE	-0.19	-0.09	-0.14	0.07	-0.05
RMSE	0.92	1.57	0.57	1.75	1.57
Slope	0.79	0.88	0.97	0.28	0.41
r^2	0.30	0.67	0.85	0.31	0.44

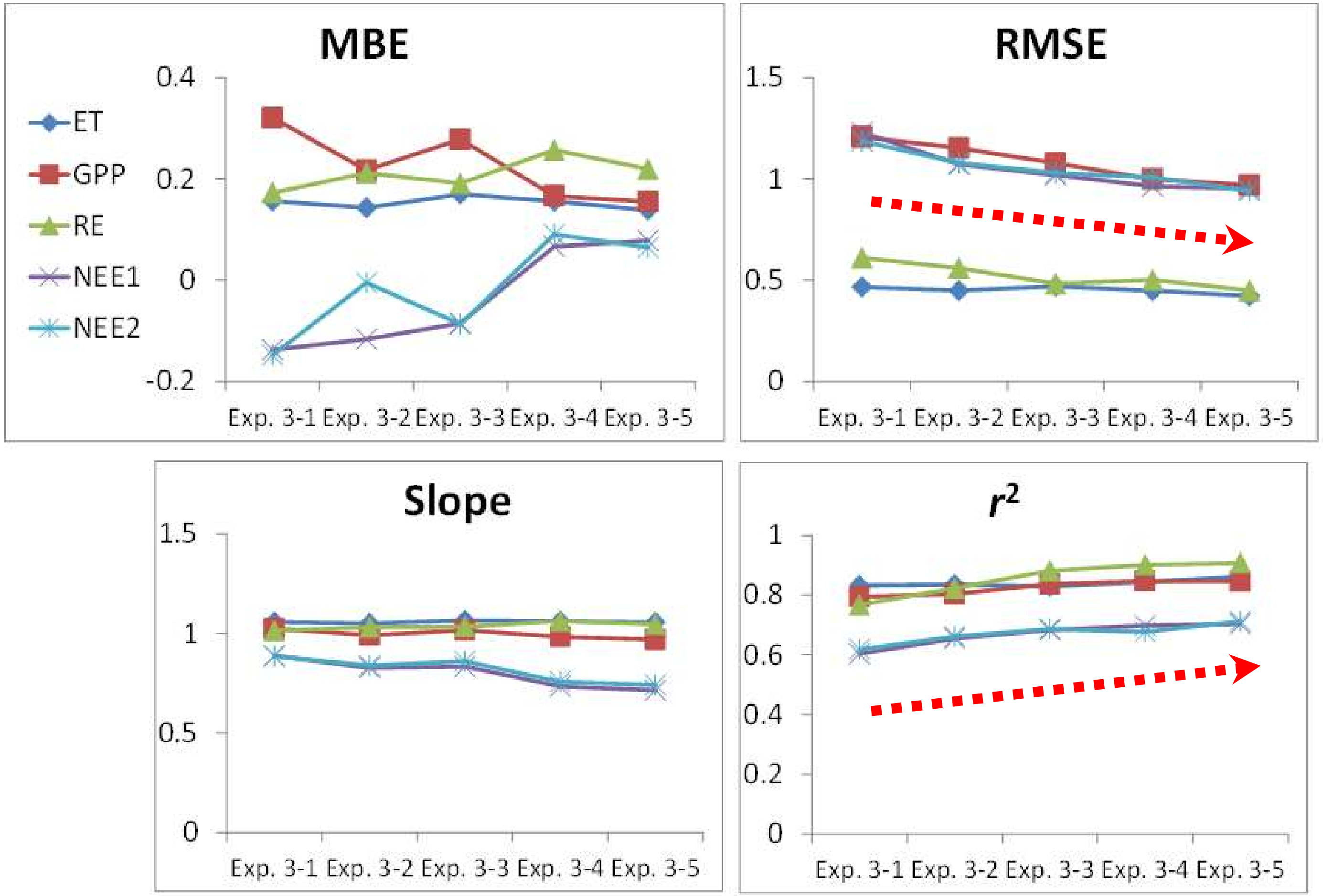
*NEE1: Directly estimated, NEE2: Estimated RE – Estimated GPP

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



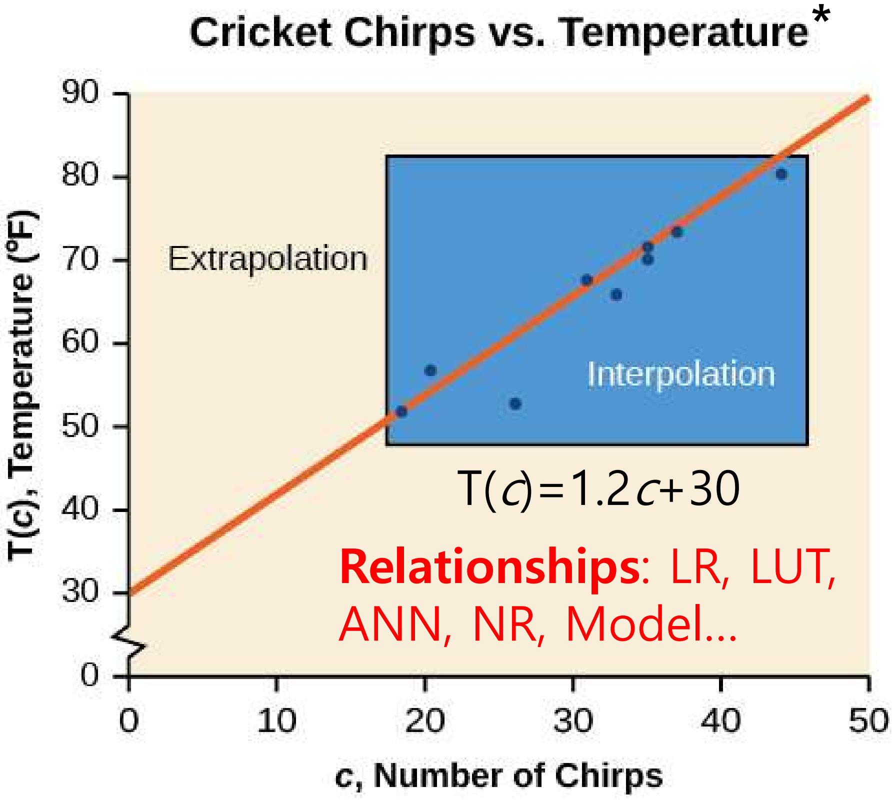
Hyp. 3: Longer training dataset results in better estimation

Accepted



Nature of gap filling and partitioning

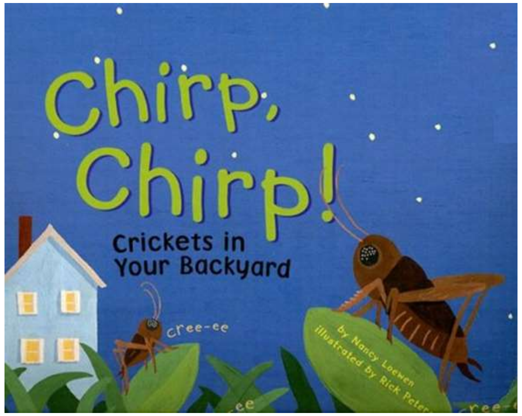
- **Interpolation and Extrapolation**



Target variables: $CO_2/H_2O/CH_4$ fluxes, RE...

Relationships: LR, LUT, ANN, NR, Model...

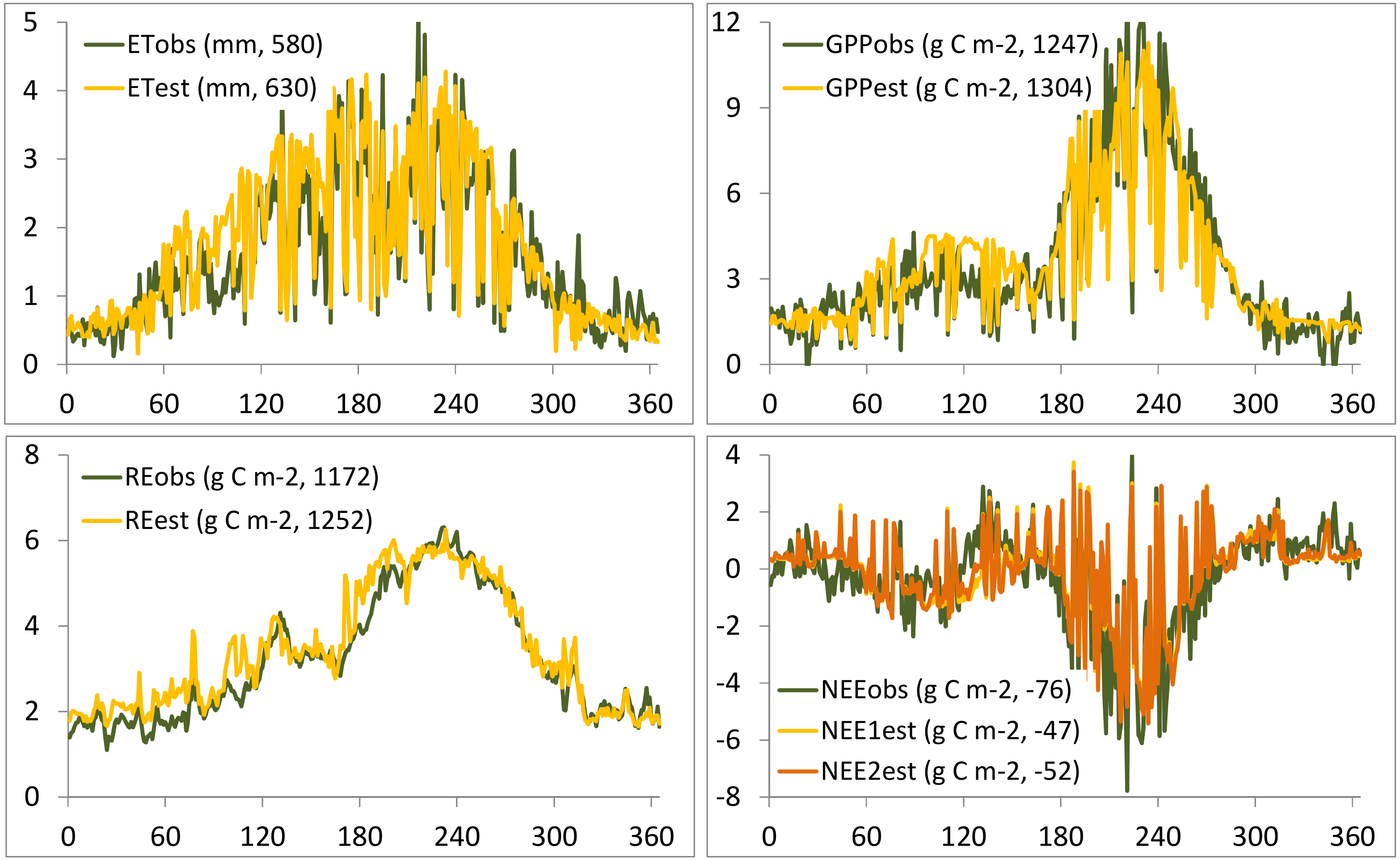
Drivers: R_{sdn}, T_{air}, VPD, SWC...



*https://math.libretexts.org/TextMaps/Algebra_Textmaps/Map%3A_Elementary_Algebra_(OpenStax)/04._Linear_Functions/4.3%3A_Fitting_Linear_Models_to_Data

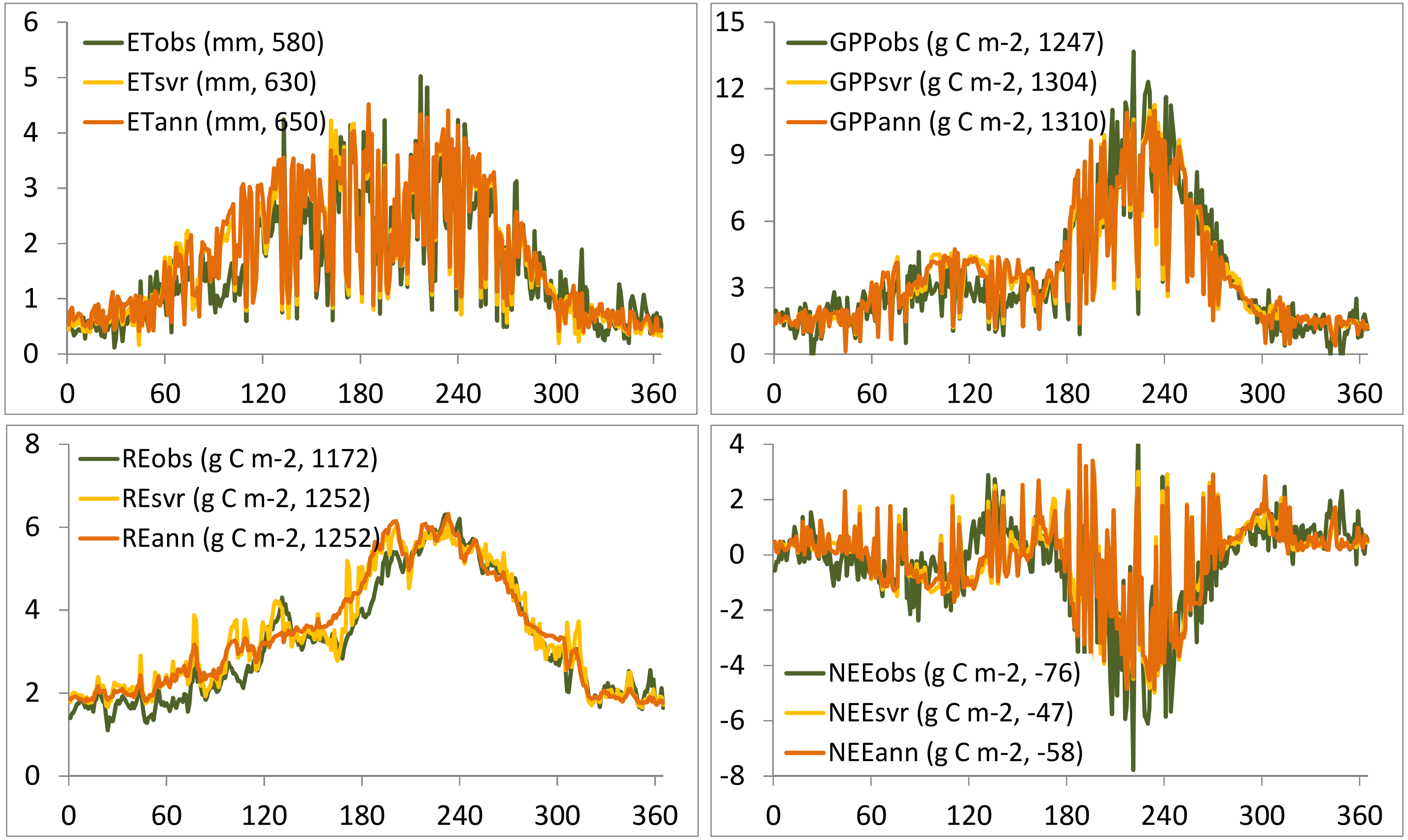
Time series analysis

Acceptable… But, there is room for improvement.



Future plans

- Testing another flux database (a natural deciduous forest, GDK)
- Testing another machine learning technique (artificial neural network)
- Testing that data-driven approach can simulate the interannual variabilities of the fluxes
- Conducting residual analysis




KoFlux
DASAN