On climate information for Crop Modeling

Hirofumi Sakuma

APL/CVPARP YES/JAMSTEC

(Potential) Collaborators: National Institute for Agro-Environmental Sciences, Japan International Research Center for Agricultural Sciences, Tottori Univ., Univ. of the Free State and ARC.

Outline of talk

- Background & recent activities in Japan.
- Major issues in coupling climate and crop models.
- Brief introduction of Bayesian statistics.
- On large-scale crop model based on BS.
- Introduction of a recent work done in climate change research, which must be applied to short-term climate variations problem.
- Short remarks on the relation with SATREPS SA.

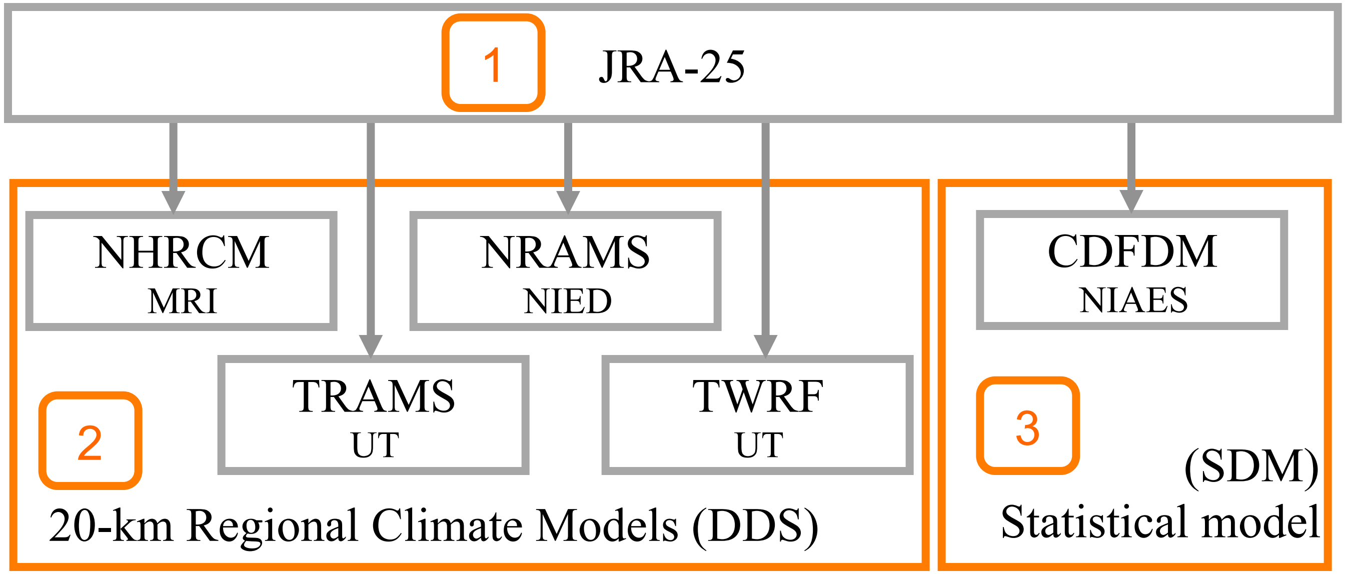
Background

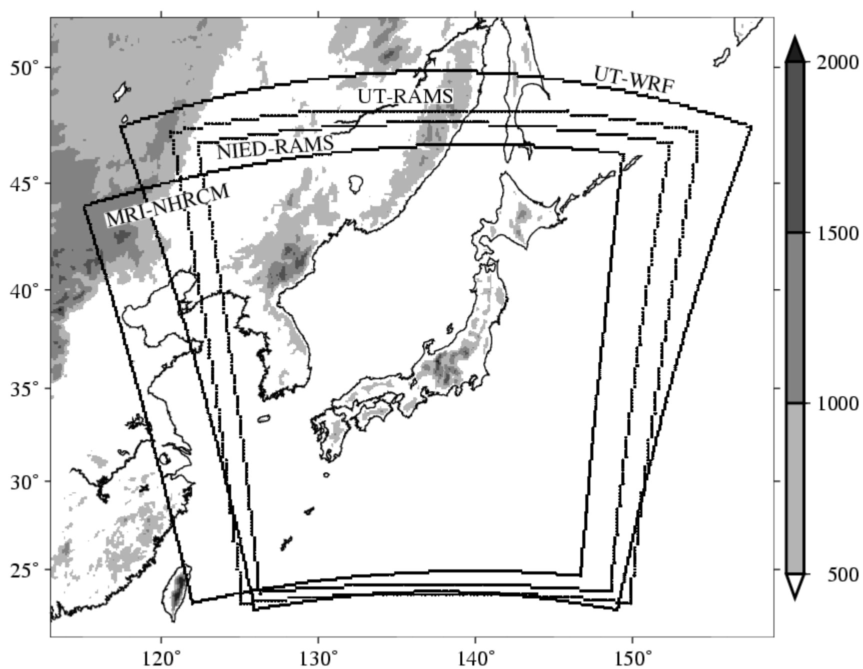
Climate-related risk management in agricultural production is an important issue for the world community

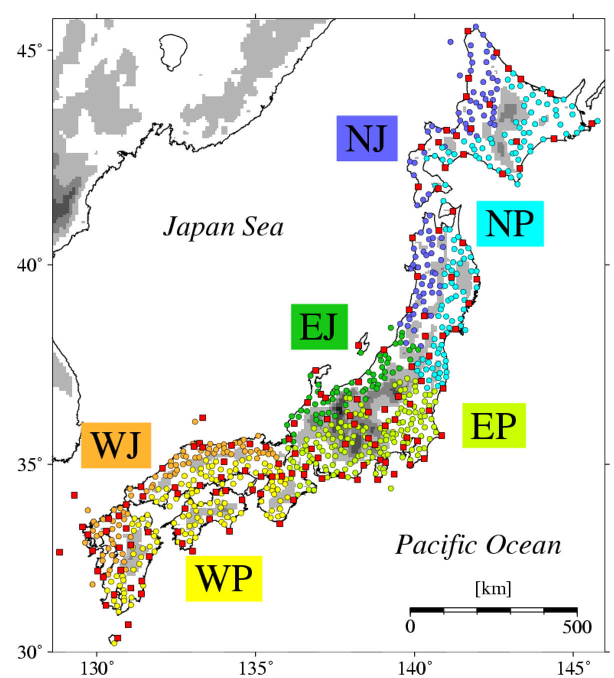
Recent (Japanese) activities in this field cover:

I. Probabilistic impact assessment considering uncertain factors both in climate and agricultural activities.
II. Creation of climate change scenario around Japan and associated impact studies on rice production.
III. *Development of large-scale crop model utilizing Bayesian approach.*

Experimental Design







Major issues in coupling climate and crop models

- For actual risk assessment, climate variations or seasonal data with spatio-temporarily fine resolution is necessary.
- The order of horizontal resolution state-of-the-art high resolution CGCMs employ is 110～300km, which is quite coarse from the viewpoint of application.
- CGCMs output have bias to be removed.
- Suitable downscaling becomes necessary which provides sufficiently high resolution as well as bias-reduced output.
- There are couple of downscaling methods which have both merits and demerits.
- In evaluating downscaling techniques, some preliminary studies had performed as a part of IPCC activities in Japan.

On Bayes' Theorem

A distinct characteristics of Bayesian statistics is that it gives the probability of causes: $P(H_i \mid A)$ for a given event A

Let Ω be a sample space (on which σ-algebra is defined) of the union of causes H_i for certain event A, that is to say, H_i is the partition of Ω,

$$H_i \cap H_j = \phi \quad ; \quad H_1 \cup H_2 \cup \cdots \cup H_k = \Omega$$

On Bayes' Theorem (continued)

then, Bayes' theorem says that the inverse (or posterior) probability is $P(H_i | A)$ given by

$$P(H_i | A) = \frac{P(H_i)P(A|H_i)}{\sum_j P(H_j)P(A|H_j)}$$

In most cases, we know $P(A | H_i)$ rather than $P(H_i | A)$

Correspondence

Probability Theory	Statistics
H_i : Cause	population parameter
A : Event	Sample
$P(A \mid H_i)$: Conditional Probability	Likelihood

Notion of Bayesian Updating:

In Bayesian statistics, utilizing the series of the occurrence of events A_k, posterior probability $P(H_i \mid A_k)$ is "updated" starting from a prior probability $P(H_i)$.

Example : spam mail filter

$$P(spam|word) = \frac{P(spam)P(word|spam)}{P(word)}$$

On large-scale crop model based on BS

(Parameters in large-scale crop must be different from those in field-scale models)

1) Phenological development components (flowering, heading, maturity are influenced by environmental conditions and characteristics of cultivar. Large-scale model attempts to simulate a typical phonological development *averaged over* cultivars, cultivation practices and local climate conditions.

2) Dry matter production components (leaf area index etc.)

3) Yield formation components (high and low temperature stresses are included in this component)

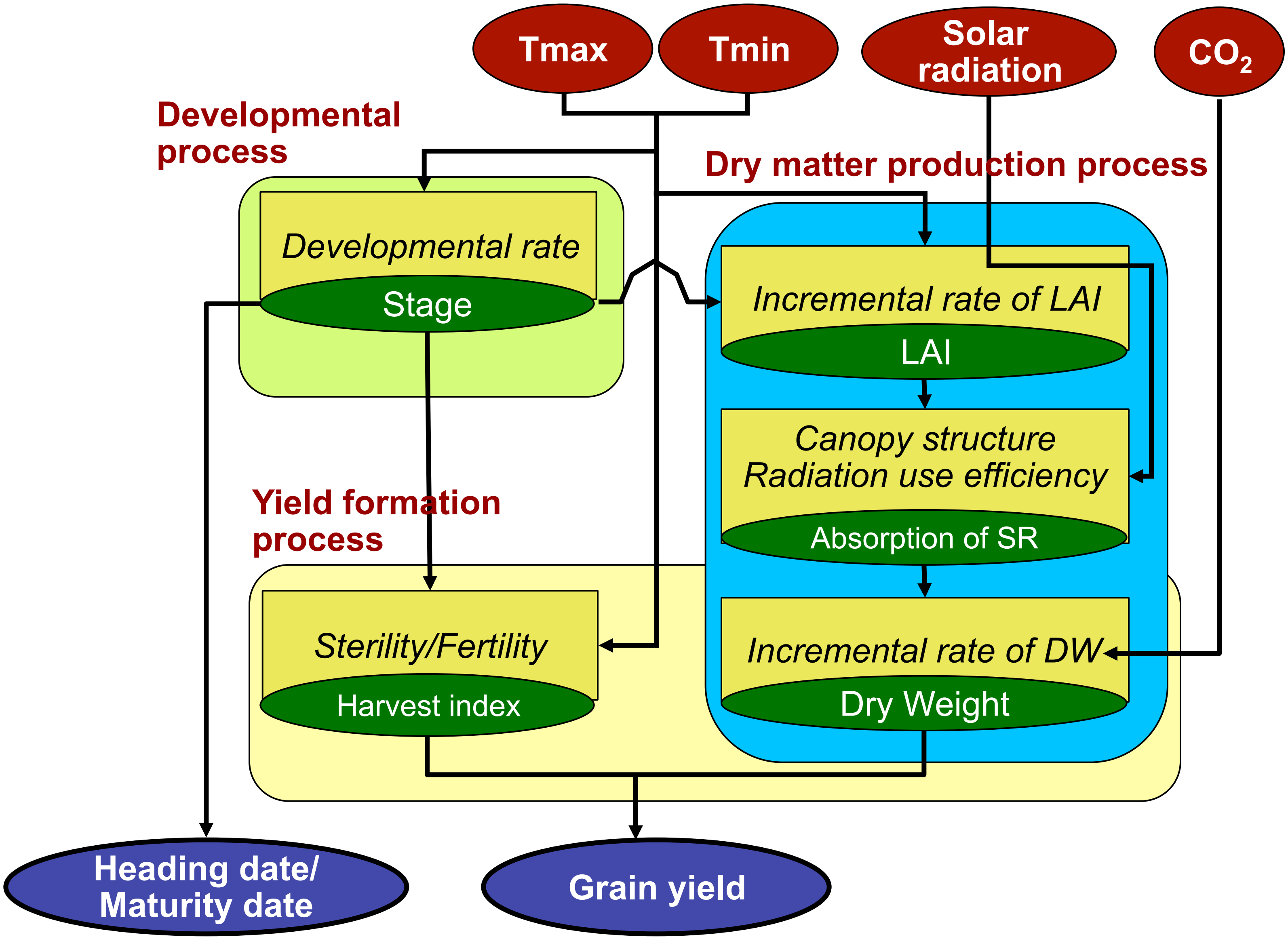
Difficulties in crop yield simulation on a large-scale

- A field-scale → Climate-model grid-scale
  - (≈tens meters) → (≈hundreds kilometers)
- Spatial heterogeneity of cultivation management
  - Cultivar, crop calendar, timings and intensities of irrigation and fertilization etc.
- Data limitation
  - Availability is limited for spatially-detailed data on cultivation management and its time change;
  - Spatially-fine data derived from remote sensing has a potential but its time period is limited
- How to model the spatial- and temporal-variation of crop yield on a large-scale?

Methodologies to model crop yield on a large-scale

- Omit the spatially-fine processes
  - Regression models (Lobell and Asner 2003)
  - Empirical approach e.g., AEZ (Fischer et al. 2002)
- Develop a new large-scale model
  - Process-based but simple enough to avoid the need for many location-specific inputs
    - GLAM (Challinor et al., 2004)
    - Hasegawa et al. (2008)
    - MCWLA (Tao et al., 2009)
- Upscale a existing smaller scale model
  - Re-calibrate a field-scale model with inputs on a large-scale and inclusion of spatial heterogeneity
    - PRYSBI (Iizumi et al., 2009a)
    - GSWAT (the U.S. crop model)

Modeling prefectural paddy rice yield in Japan



(Horie et al., 1995)

Parameters analyzed

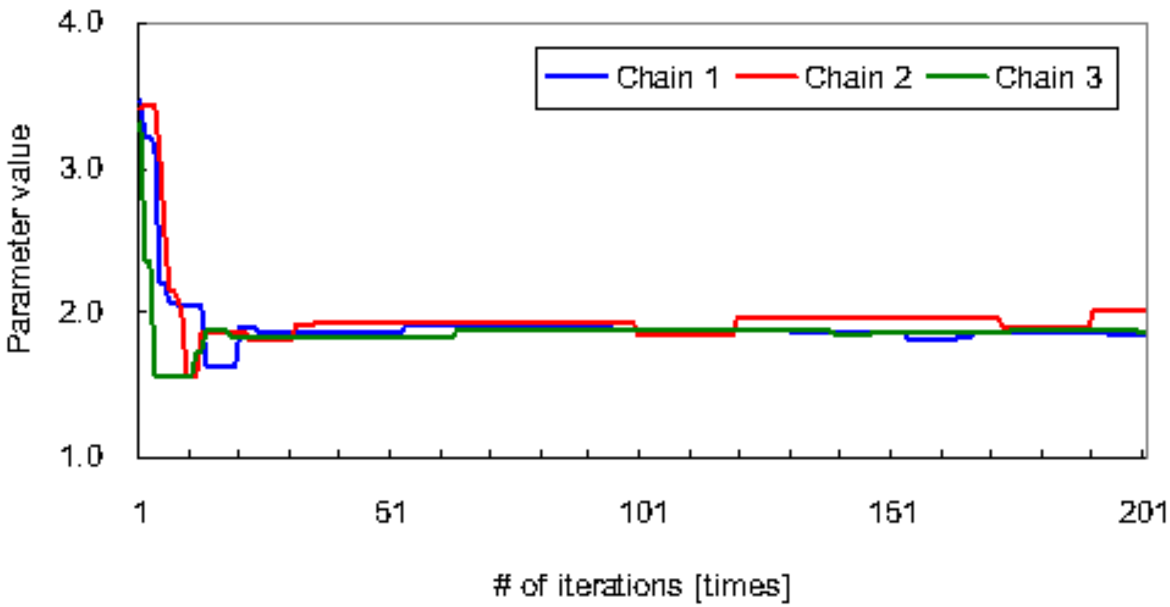
Abbr.	Definition
DVI_0 (day^{-1})	Initial developmental index (DVI)
G (day)	Minimum number of days required for heading under 350 ppm of atmospheric CO_2 concentration
A_T (-)	Sensitivity of developmental rate (DVR) to air temperature
T_h (°C)	Air temperature at which DVR is half of the maximum rate at the optimum temperature
B_L (-)	Sensitivity of DVR to day length
L_c (hr)	Critical day length
DVI^* (day^{-1})	Value of DVI at which point the crop becomes sensitive to the photoperiod
LAI_0 (-)	Initial leaf area index
DW_0 (g m^{-2})	Initial dry weight
T^* (°C)	Base air temperature for calculating cooling degree days
C_{cool} (-)	Curvature factor of spikelet sterility caused by low temperature
C_{hot} (-)	Curvature factor of spikelet sterility caused by high temperature
τ (-)	Technical coefficient

Markov Chain Monte Carlo (MCMC) technique

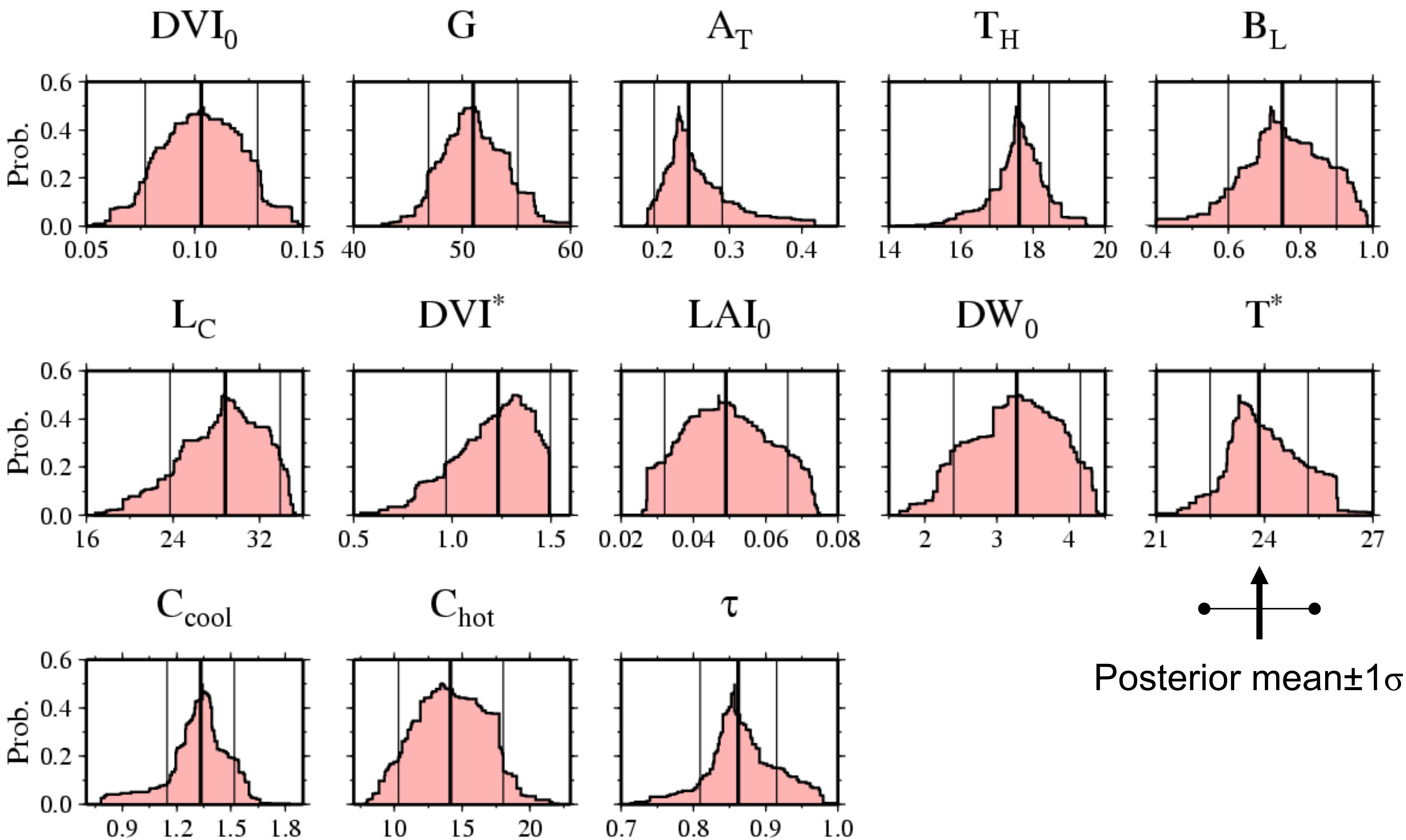
- Bayes' theorem

$$\underbrace{p(param \mid Data)}_{\text{Posterior distribution}} \propto \underbrace{\pi(Data \mid param)}_{\text{Likelihood function}} \underbrace{q(param)}_{\text{Prior distribution}}$$

- Application of Metropolis-Hastings (M-H) algorithm
  - Calibration: 13 odd-years (1979, 1981, …, 2001, 2003);
  - Verification: 25 years (1979-2003)

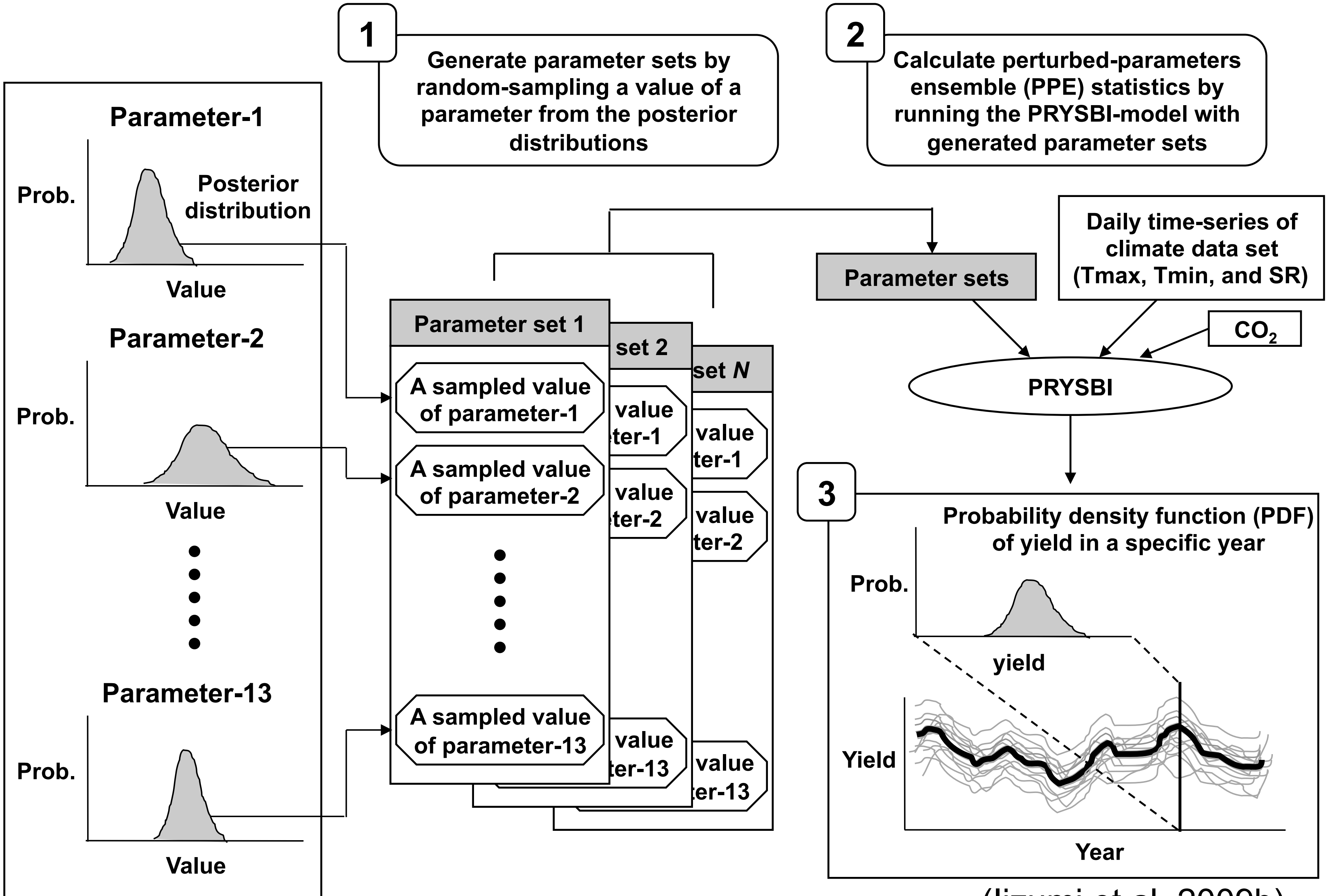


Obtained posterior distributions of parameter values

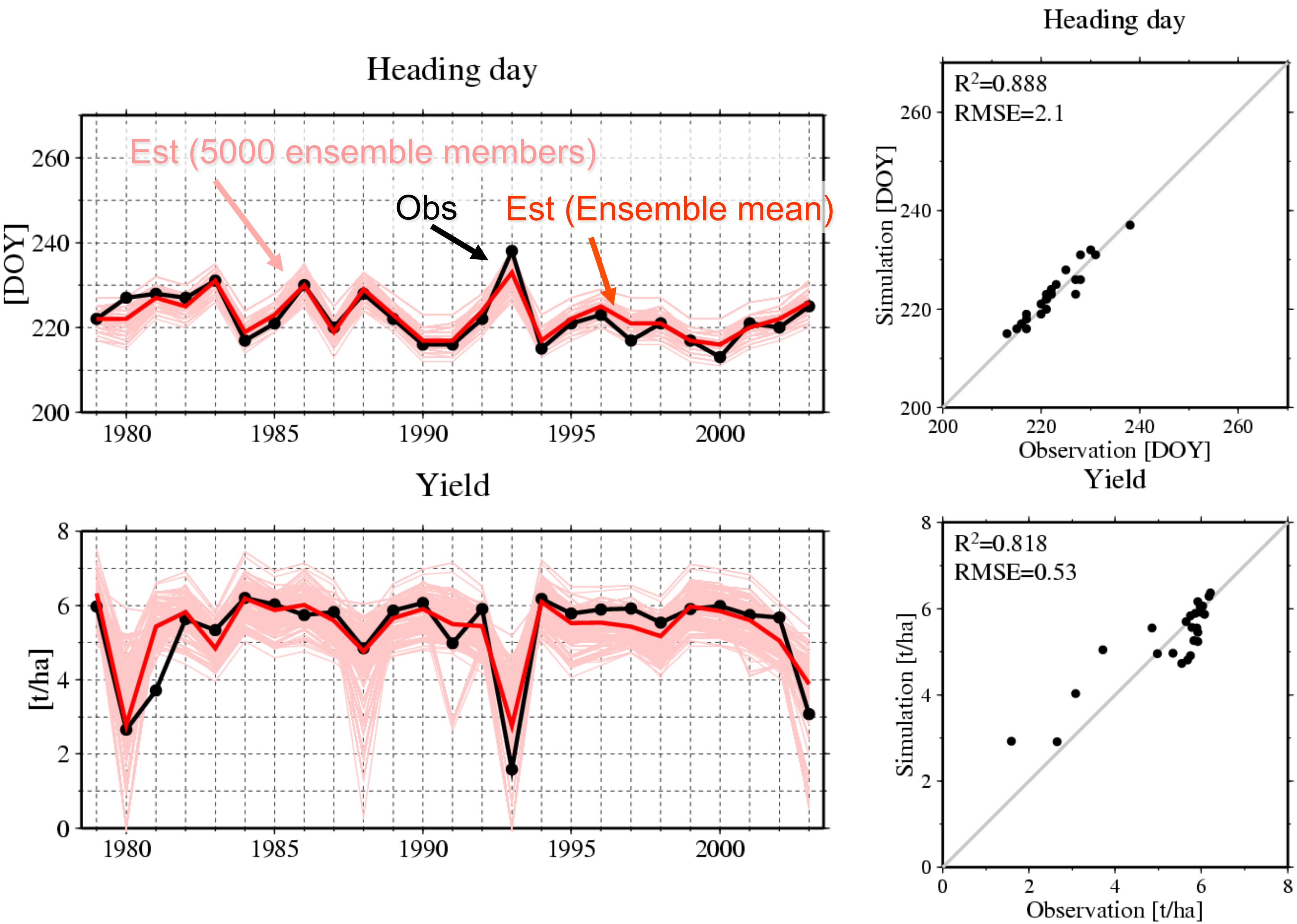


(Ex. Aomori)

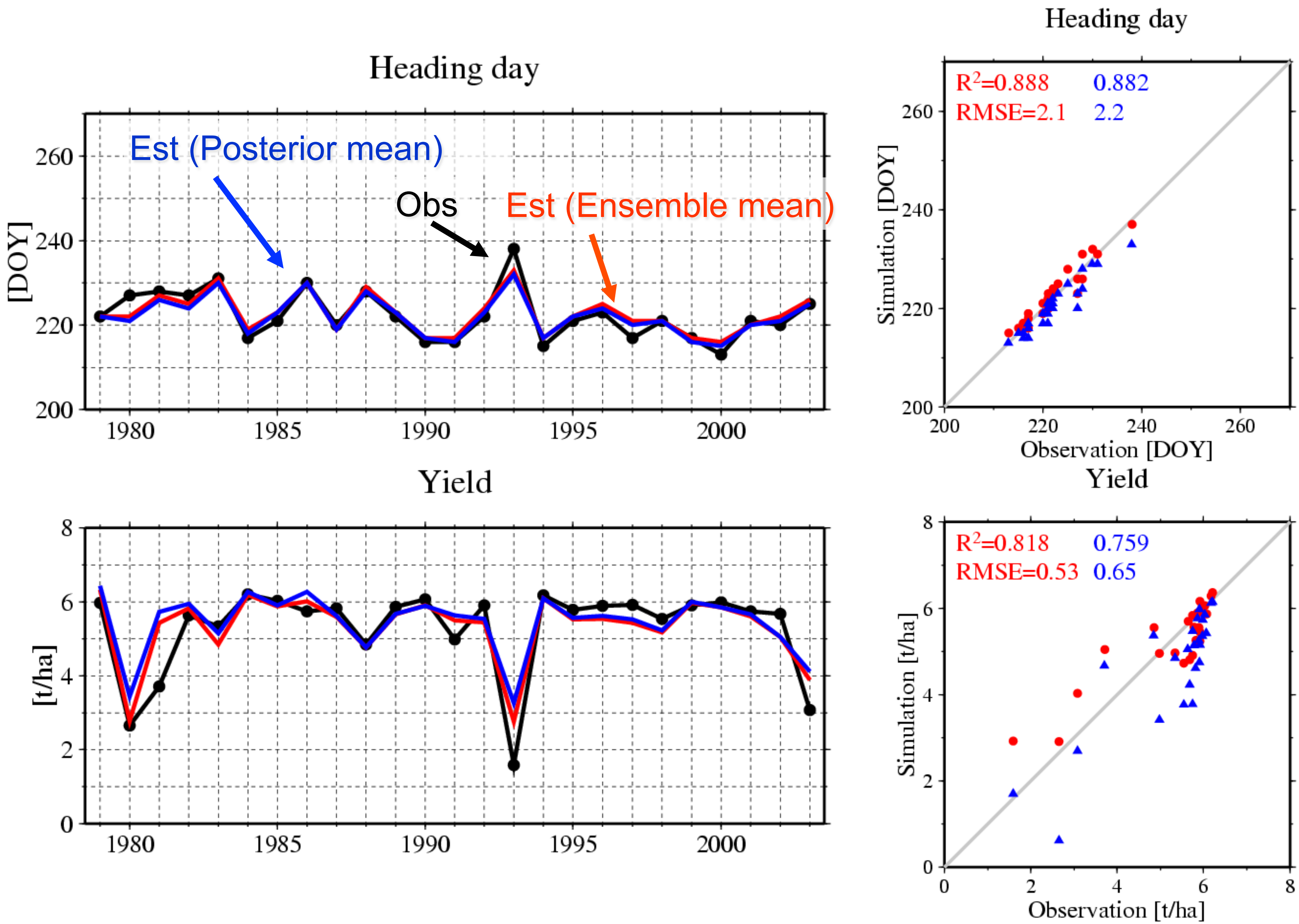
Perturbed-parameter ensemble approach



(Iizumi et al. 2009b)


Heading day
Est (5000 ensemble members)
Obs
Est (Ensemble mean)
[DOY]
260
240
220
200
1980
1985
1990
1995
2000
Yield
[t/ha]
8
6
4
2
0
Heading day
R²=0.888
RMSE=2.1
Simulation [DOY]
Observation [DOY]
Yield
R²=0.818
RMSE=0.53
Simulation [t/ha]
Observation [t/ha]

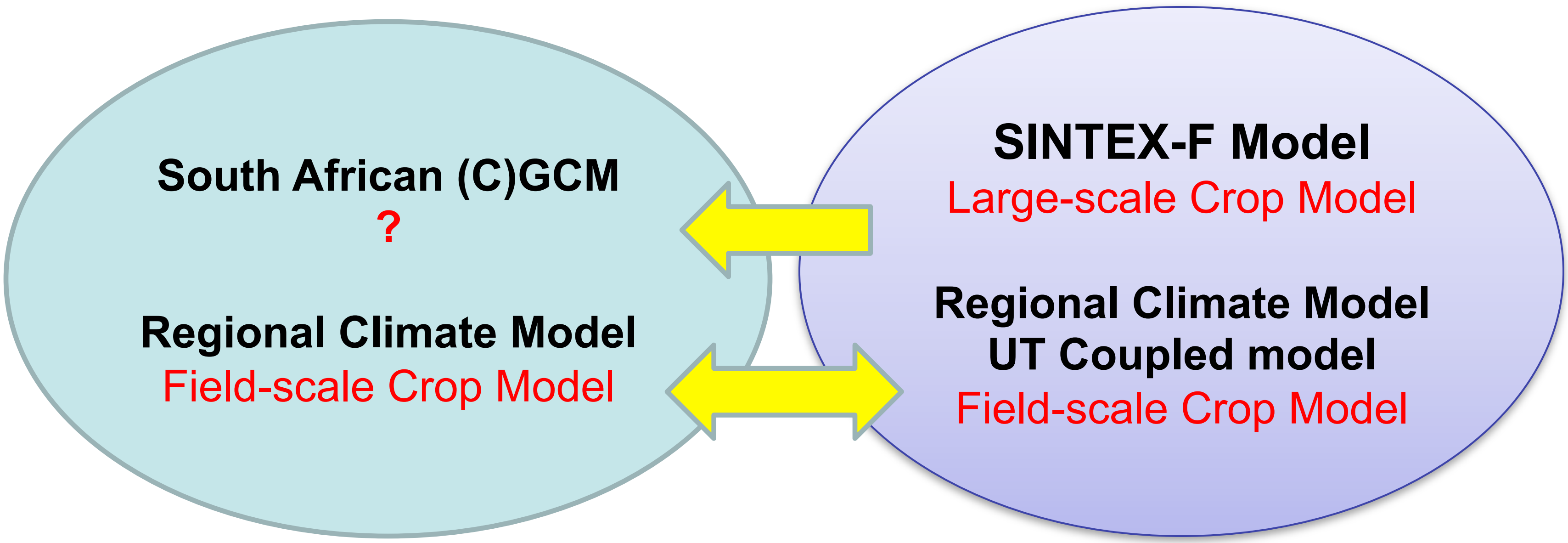
(Ex. Aomori)



Heading day
Est (Posterior mean)
Obs
Est (Ensemble mean)
[DOY]
Yield
[t/ha]
Heading day
R²=0.888 0.882
RMSE=2.1 2.2
Simulation [DOY]
Observation [DOY]
Yield
R²=0.818 0.759
RMSE=0.53 0.65
Simulation [t/ha]
Observation [t/ha]

(Ex. Aomori)

Short remarks on the relation with SATREPS SA



Summary

- The posterior distributions of parameter values were estimated by using the MCMC technique;
- The obtained posteriors includes the information on spatial heterogeneity on a large-scale under given data;
- The perturbed-parameter ensemble approach gives better simulation result than the use of posterior means; this suggests the adequacy of ensemble approach to express the spatial heterogeneity.