

# Assimilação de dados em modelos agrícolas baseados em processos

---

1100222 - Modelagem de Crescimento de Culturas Agrícolas  
LEB5048 - Modelagem de Culturas Agrícolas I

Izael Martins Fattori Junior  
Prof. Fábio Ricardo Marin

11/05/2023  
Piracicaba

# O que é assimilação de dados?

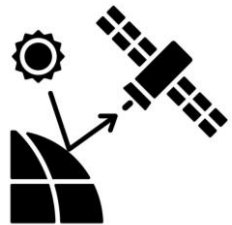
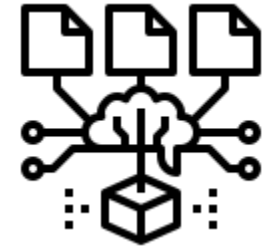
---

É um conjunto de **métodos** empregados para realizar adequadamente a inserção de **dados observados** num **sistema operacional de previsão**, com intuito de melhorar as previsões.

(Huang et al., 2019; Jin et al., 2019)

# O que é assimilação de dados?

Sistema operacional de previsão (modelos agrícolas baseados em processos)



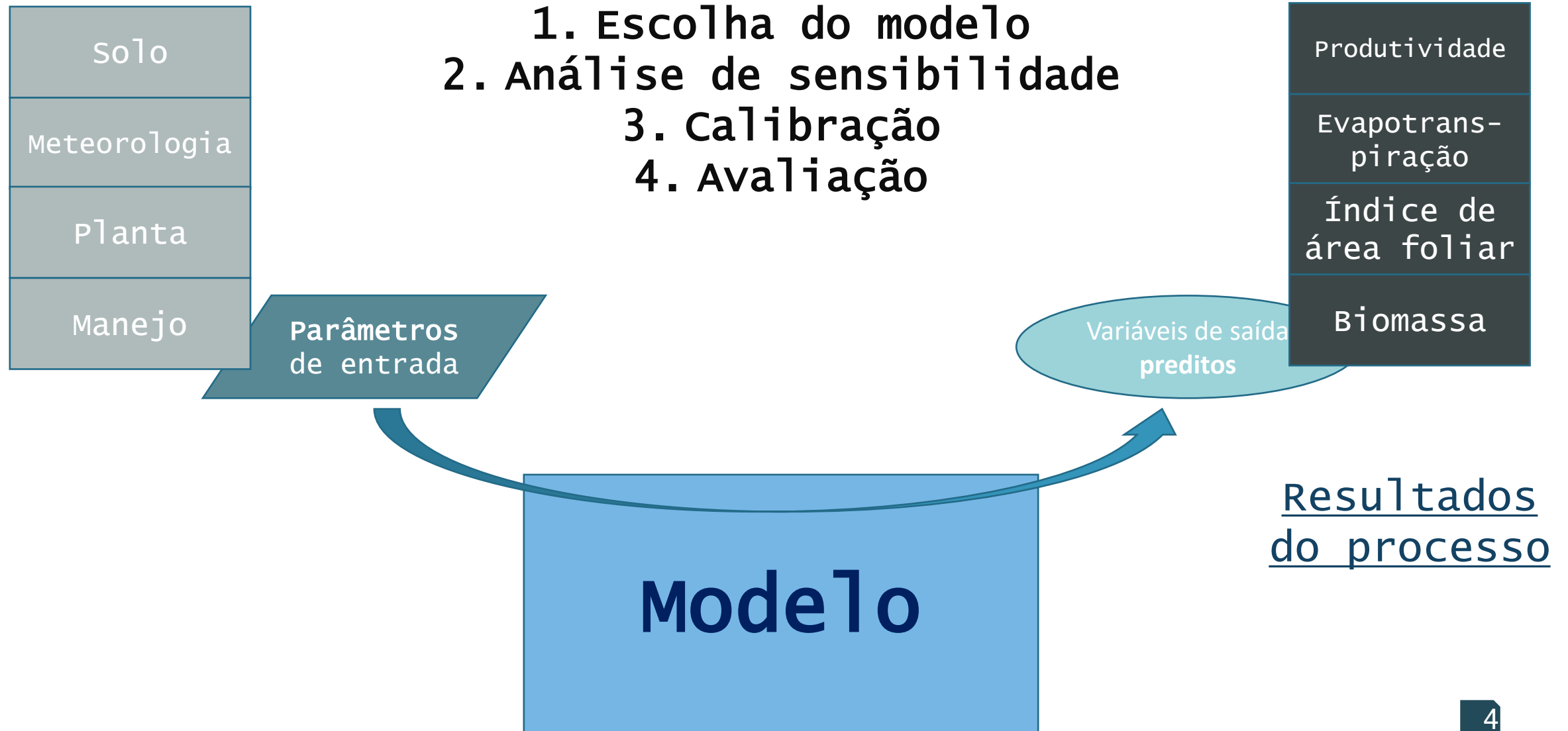
Dado observado (dado assimilado)

Métodos para assimilação de dados



(Huang et al., 2019; Jin et al., 2019)

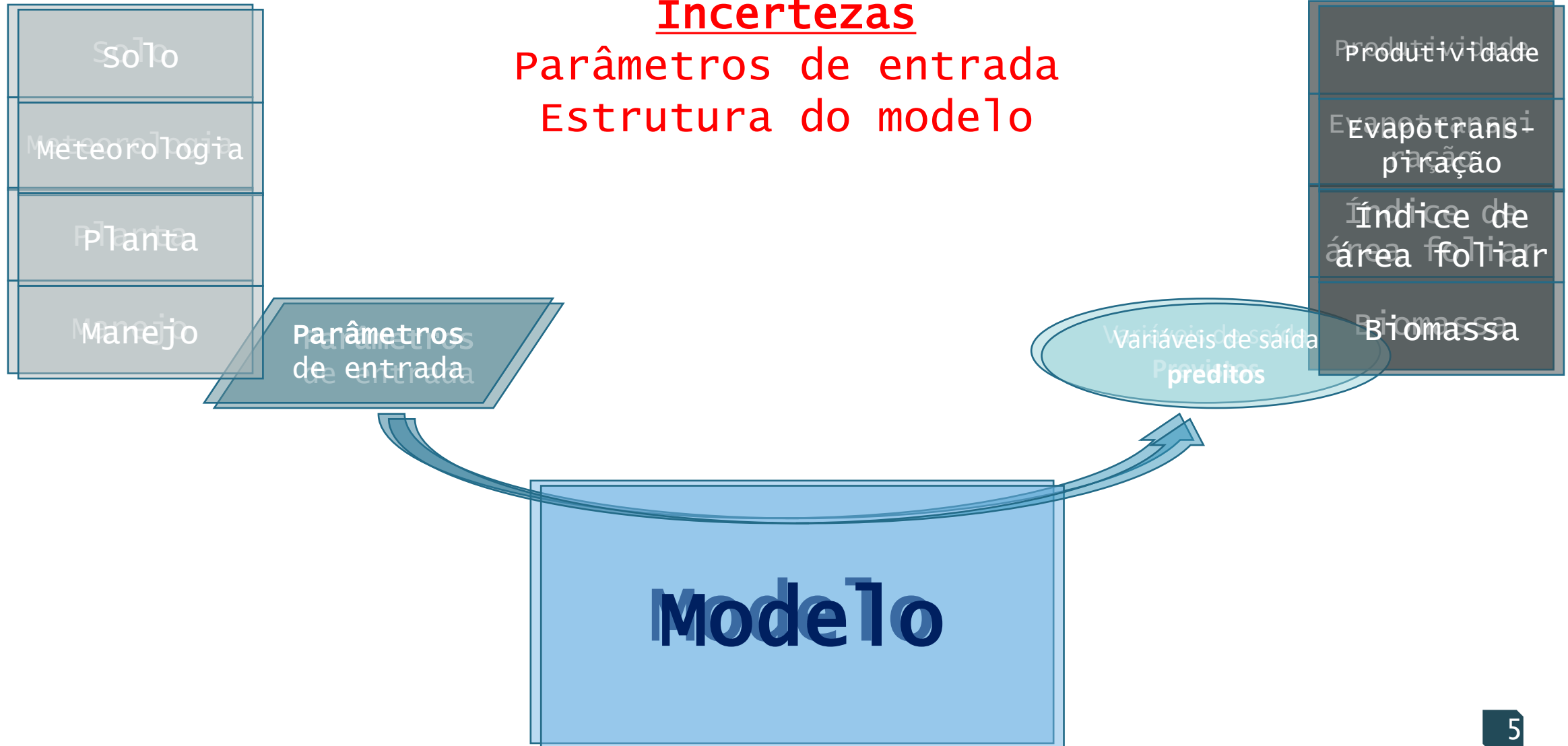
# Modelos entradas e saídas



# Modelos entradas e saídas

## Incertezas

Parâmetros de entrada  
Estrutura do modelo



# Incertezas na estrutura

- DSSAT: plataforma de modelos específicos;
- WOFOST: um modelo para várias culturas.

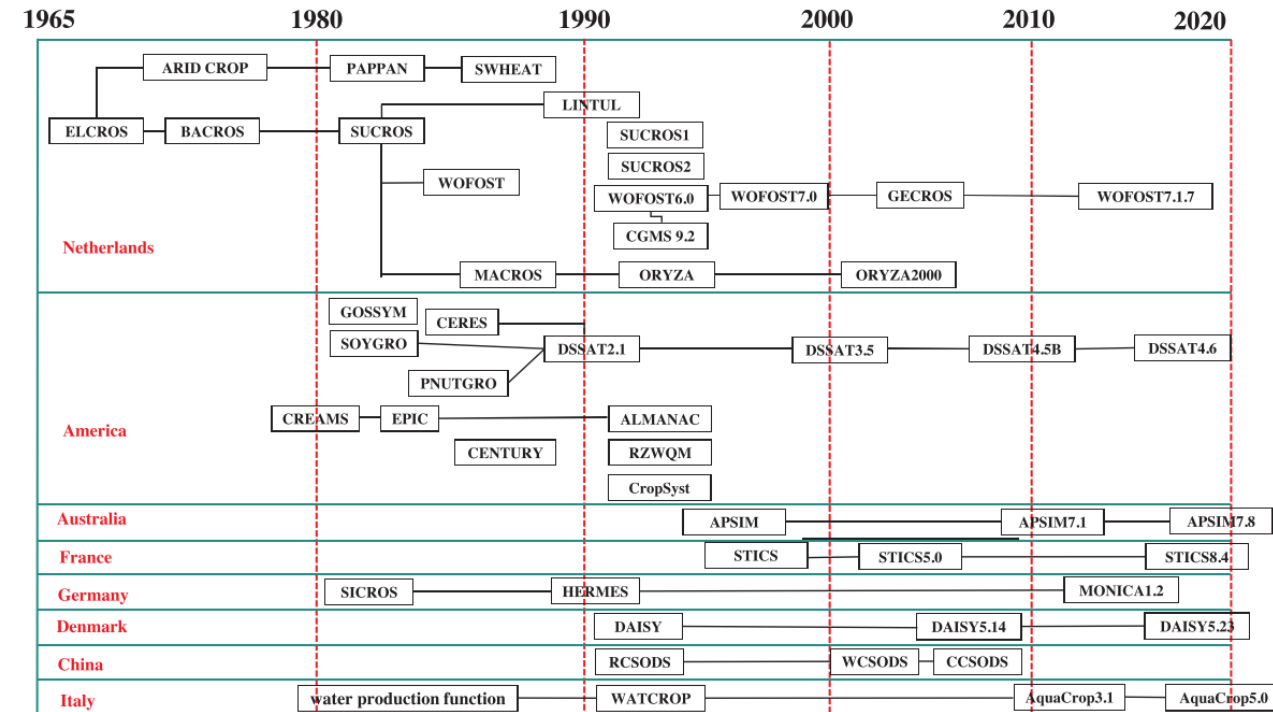


Fig. 1. Development of main crop models over time. Note: horizontal continuous lines indicate the development of new crop models.



# Incertezas nos parâmetros

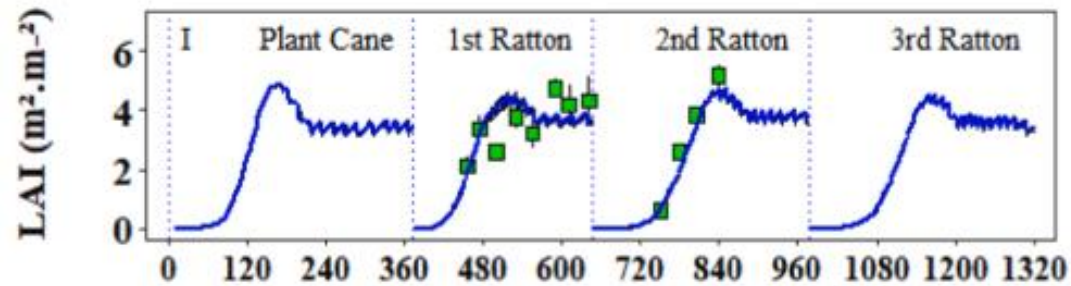
- Modelos determinísticos **determinam** os resultados de forma exata a partir das condições iniciais de um problema;
- A partir de um conjunto de dados de entrada, vamos ter sempre os mesmos resultados;
- **Não consideram qualquer distribuição de probabilidade**, variância ou elemento aleatório, nas simulações.



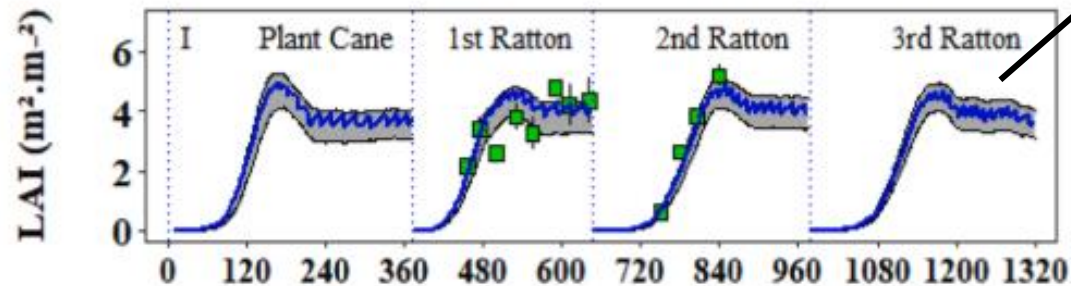
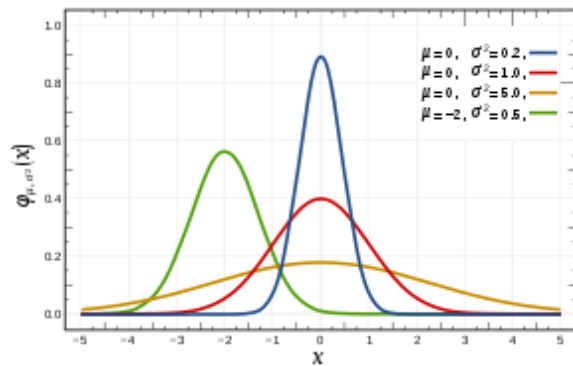
# Modelos determinísticos

Simulação determinística

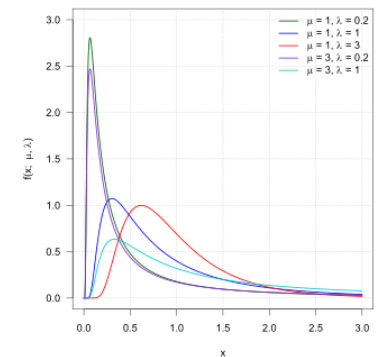
Dados de entrada pontuais



Simulação estocástica



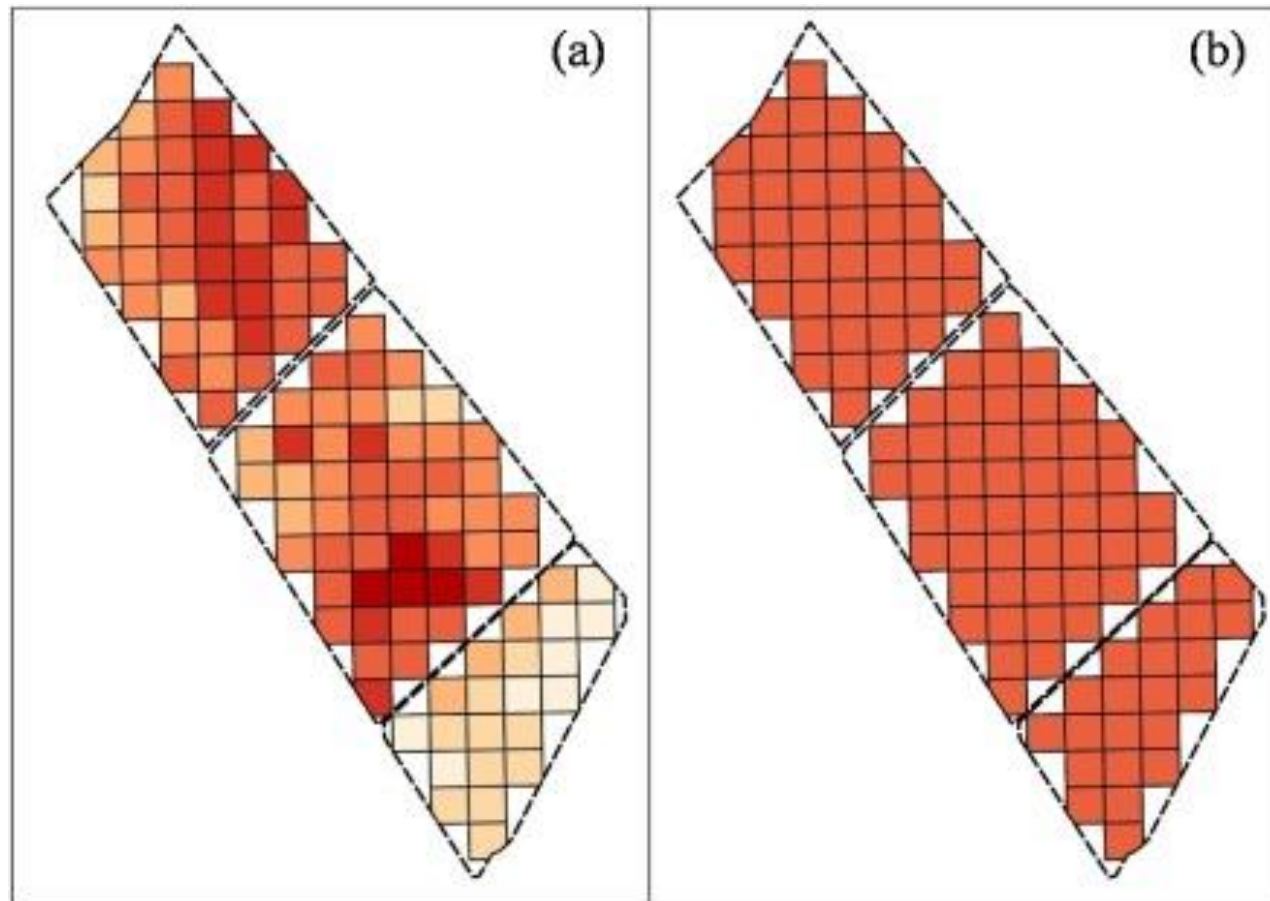
Conjunto de possíveis resultados



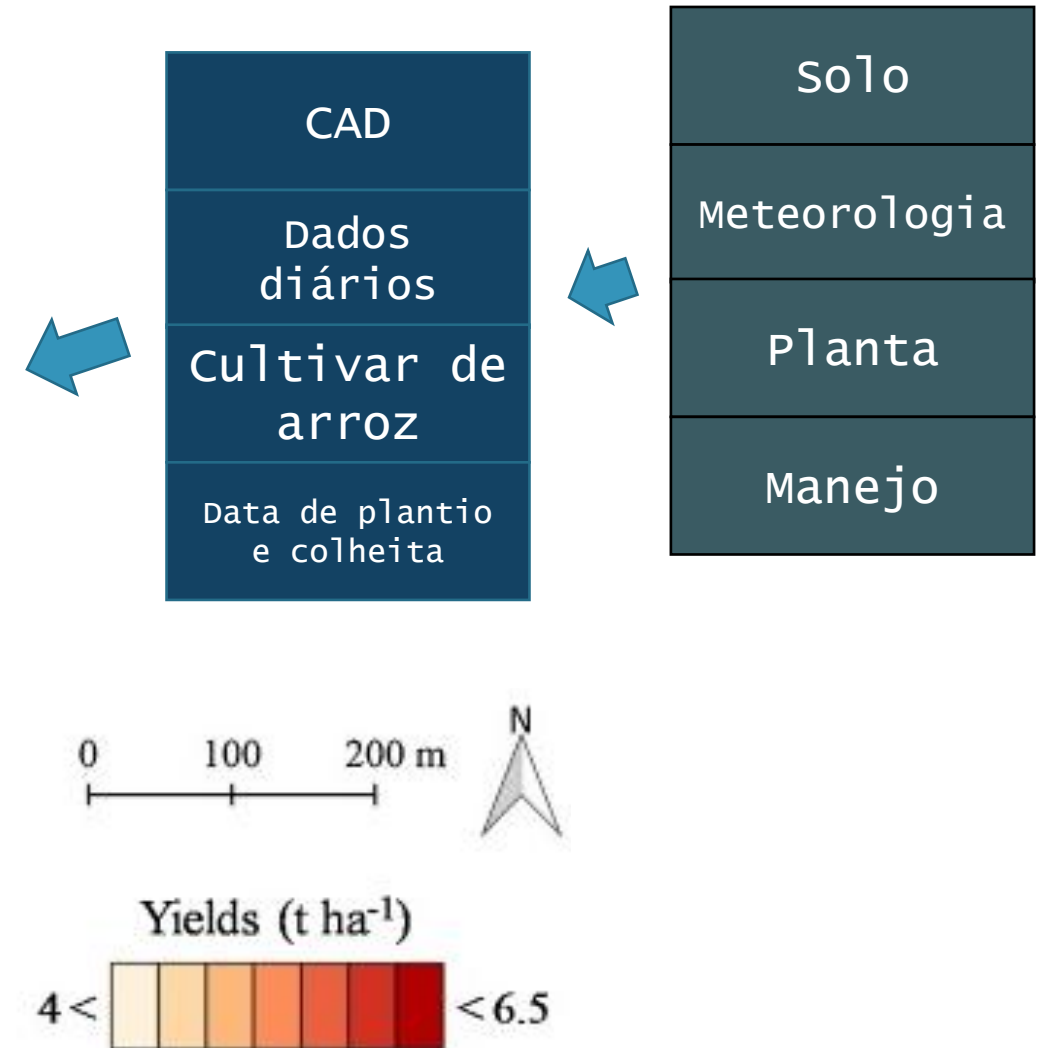
(Marin et al., 2017; Pereira et al., 2021)



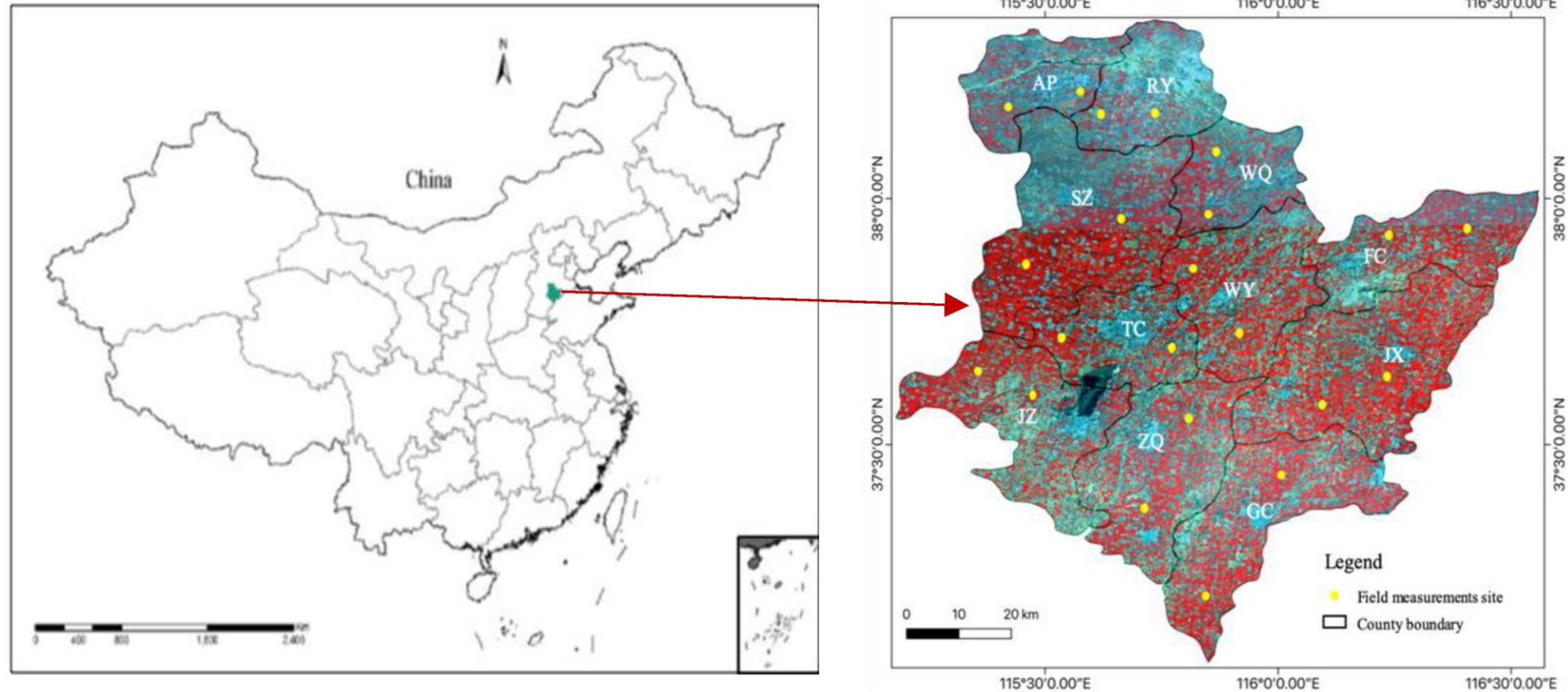
# Simulações determinísticas



(Gilardelli et al., 2019)

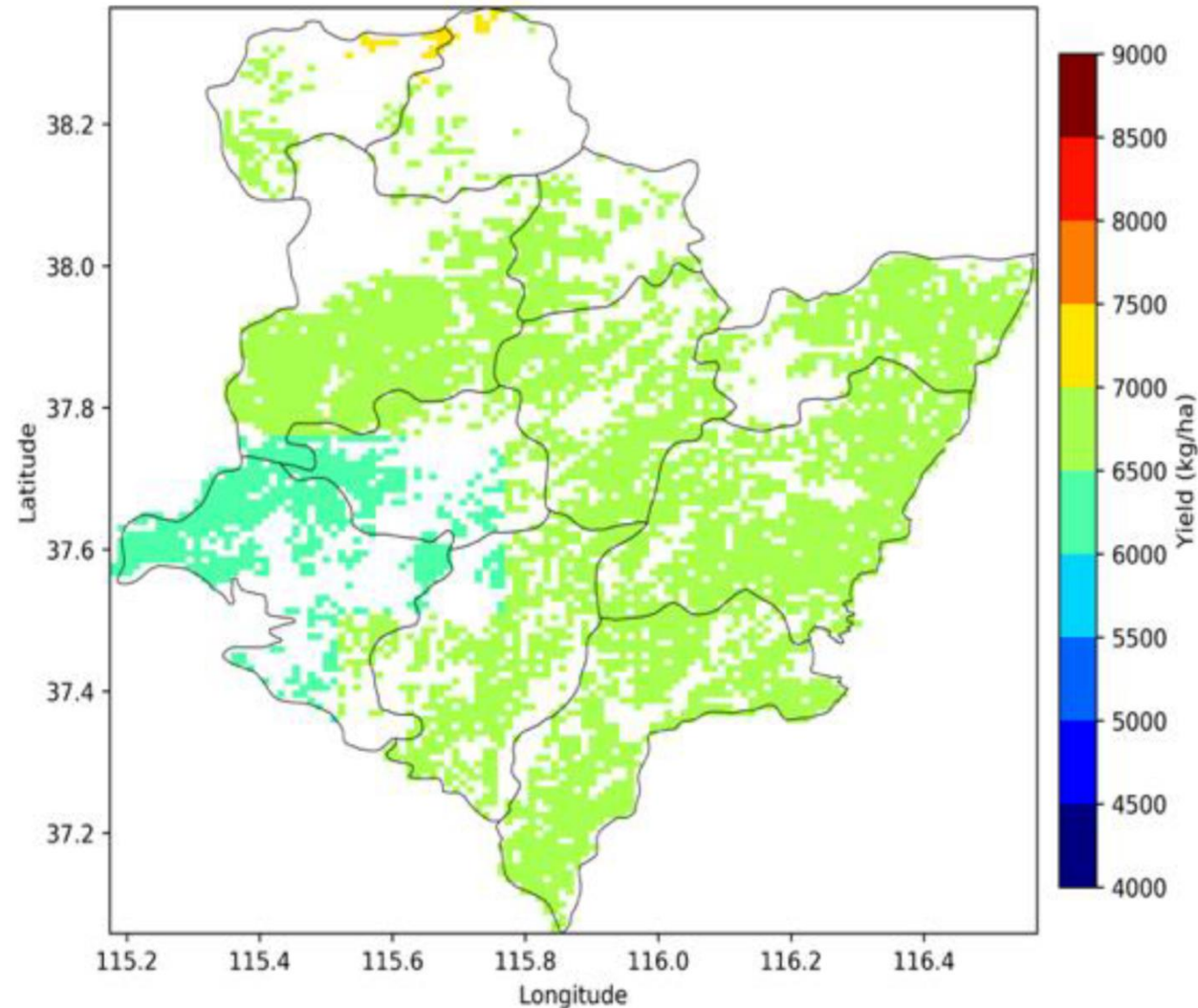


# Simulações determinísticas



Mapa da área de estudo e campos com dados observados de trigo em 2017 (Pan et al., 2019)

# Simulações determinísticas



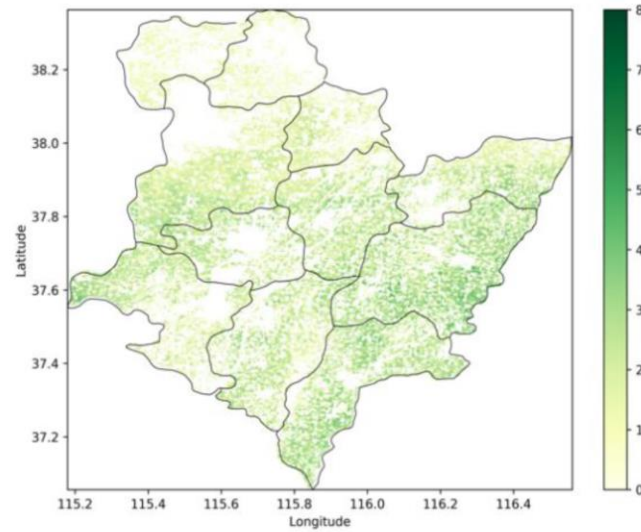
Produtividade  
de trigo  
simulado pelo  
modelo WOFOST

(Pan et al., 2019)

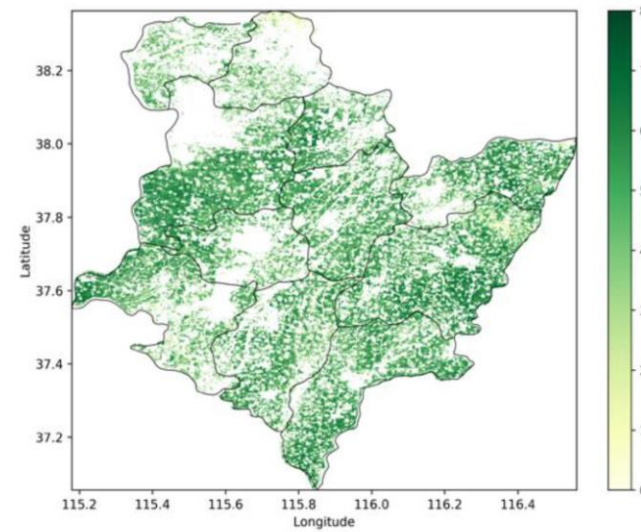


# Simulações determinísticas

Índice de área foliar



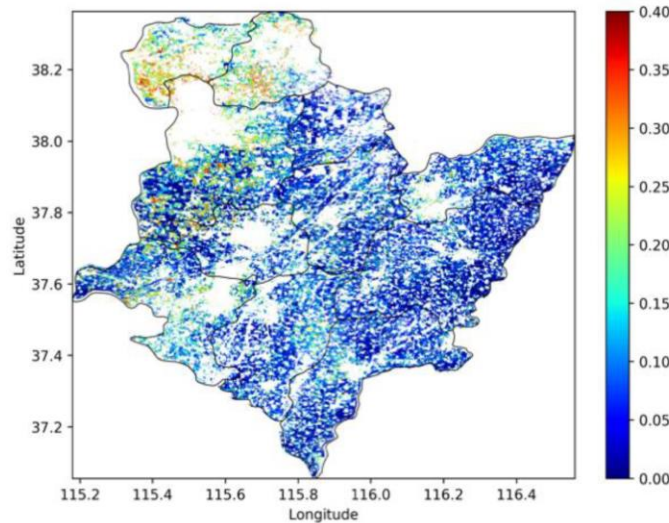
(a)



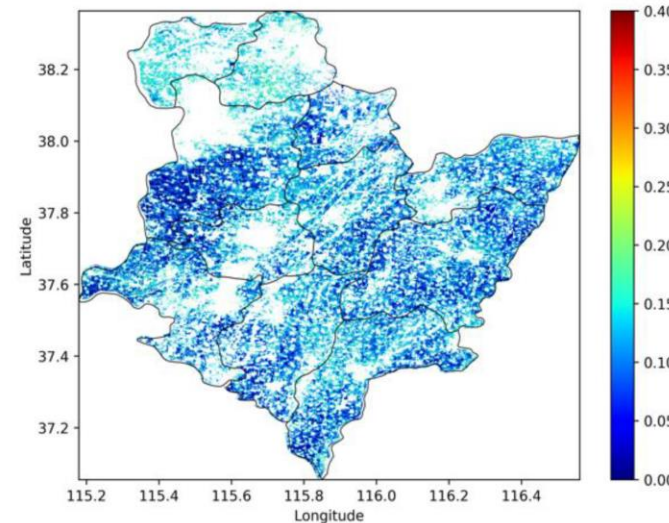
(b)

Valores obtidos do Sentinel-2

Umidade do solo



(c)



(d)

Valores obtidos do Sentinel-1A e 2

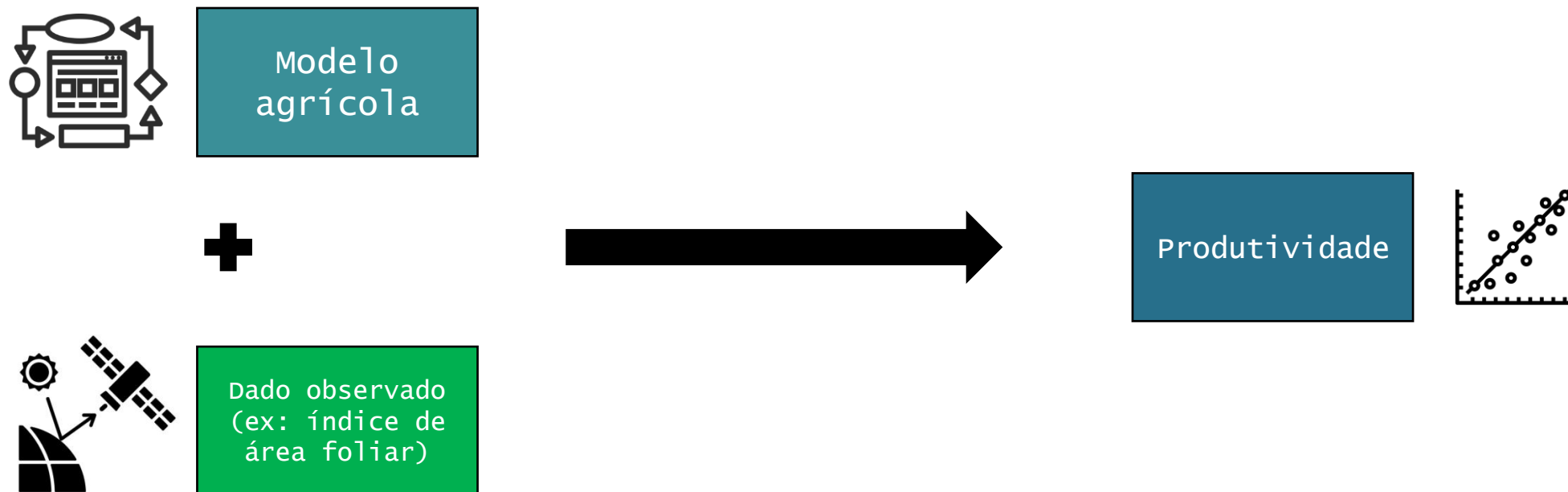
(Pan et al., 2019)

# Simulações determinísticas

- Modelos são **determinísticos**!
- A **escala das simulações** está relacionada com a **escala dos dados de entrada**;
- Modelos não conseguem simular fatores **redutores** (pragas, doenças, diferentes manejos).

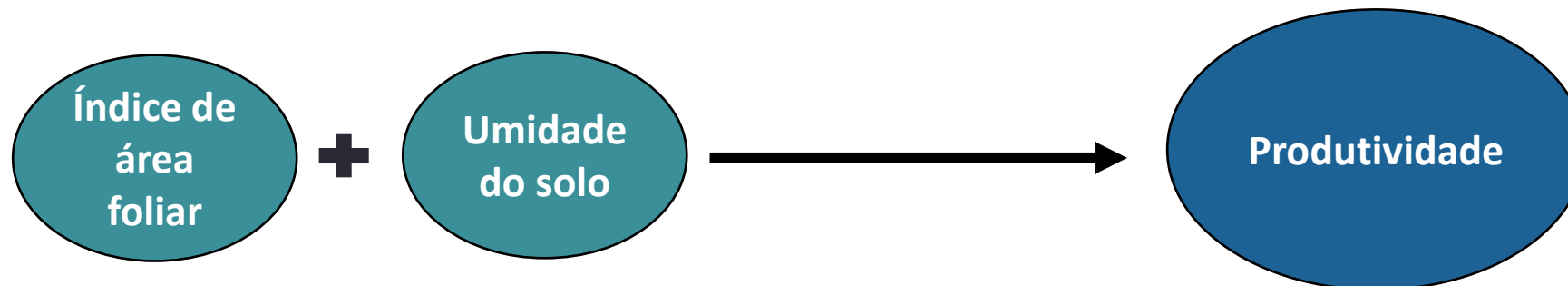
# Assimilação de dados

- Metodologia para **melhorar** as **previsões** de um modelo agrícola para uma **variável específica**.



# Variáveis a serem assimiladas

- Qualquer variável de estado **simulada** pelo modelo pode ser assimilada;
- Importante! A escolha da variável a ser assimilada depende da **correlação** com a variável de interesse a ser simulada.



Fenologia

Evapotranspi-  
ração

fPAR

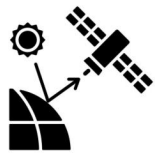
Altura de  
planta

Umidade do  
solo

# Estimando variáveis observadas



Nível de campo;



Uso de sensoriamento remoto;  
Orbitais ou sub-orbitais (VANT);  
Modelos espectrais (índice de vegetação).



# Nível de campo

---

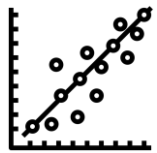
- Uso de sensores de campo (ex: LAI-2200, soda TRD);
- Dados biométricos (IAF, Biomassa, altura de planta);
- Baixo erro, importante para estudos que querem controlar o erro do dado observado;
- Importante para treinar modelos.



# Sensoriamento remoto



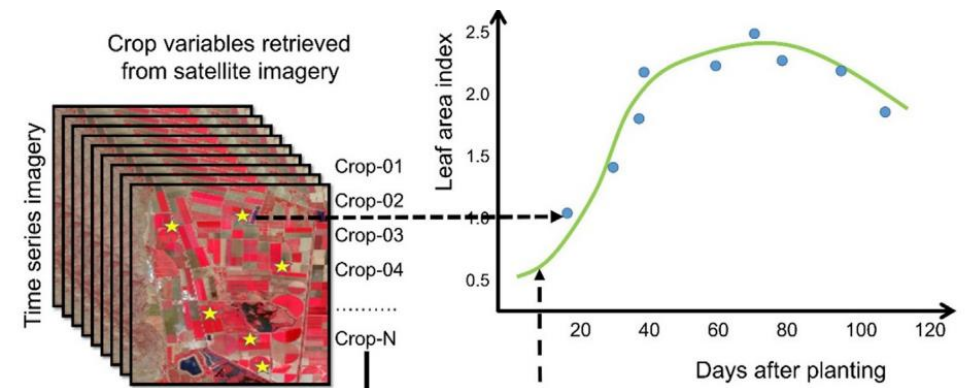
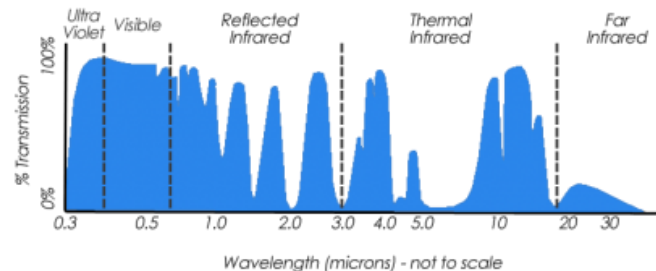
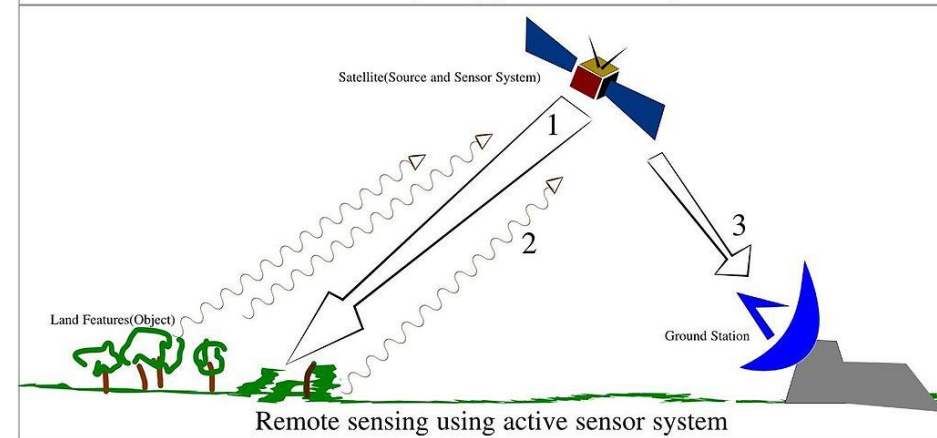
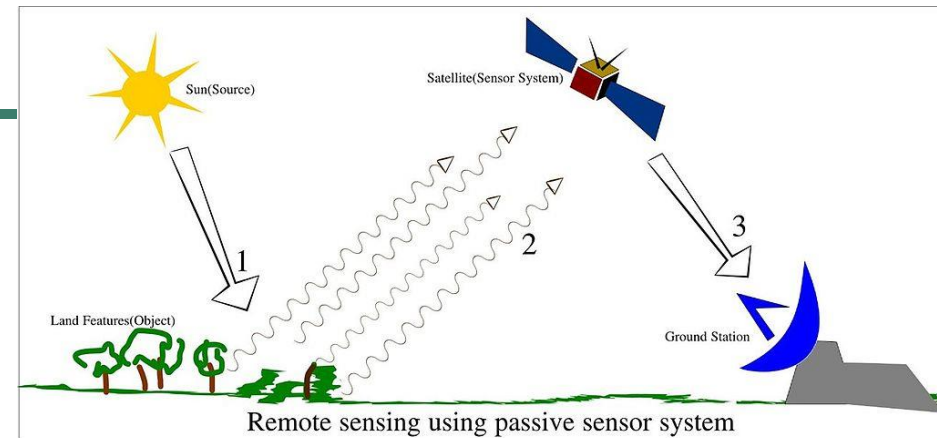
Correlação com dados de sensoriamento remoto;



Modelos lineares, logaritmos, exponenciais...;



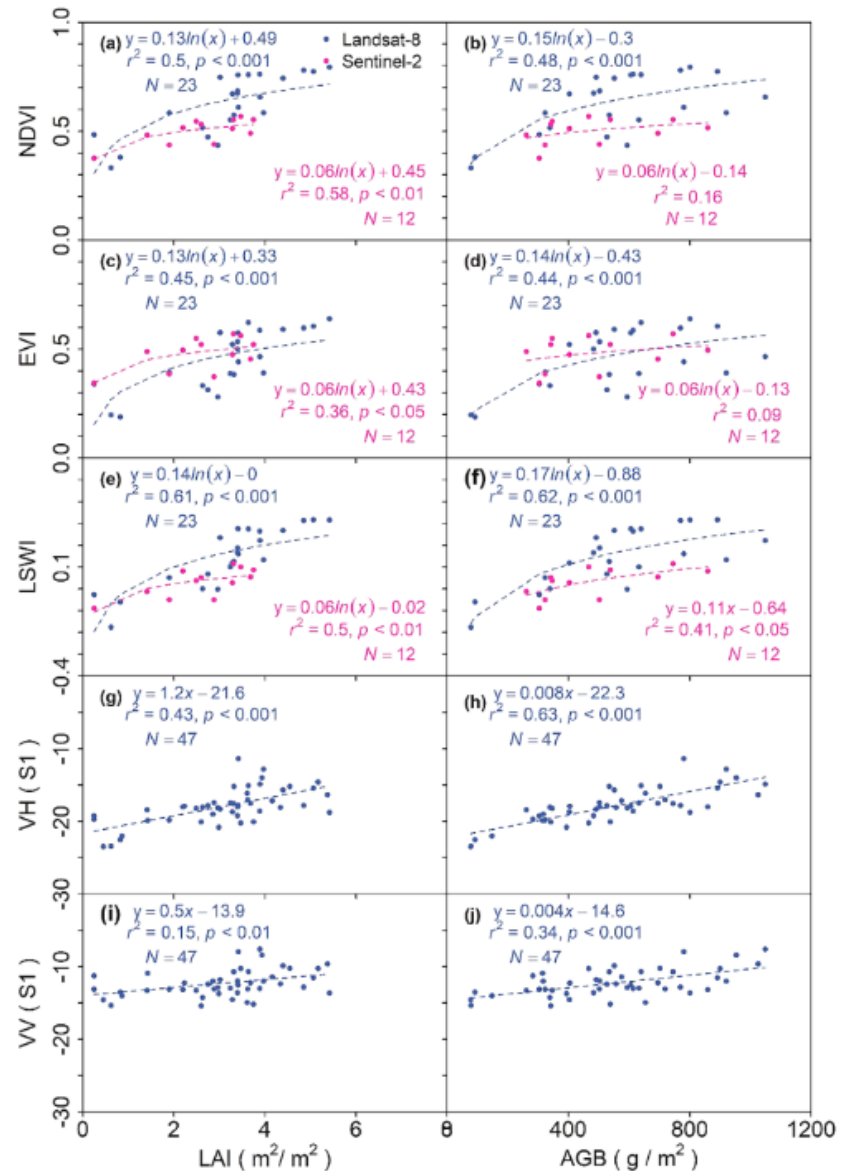
Depende de dados de qualidade.



# Modelos estatísticos

Relação entre dados obtidos por sensoriamento remoto e parâmetros biofísicos do pasto

(Wang et al., 2019)



# Fontes de dados

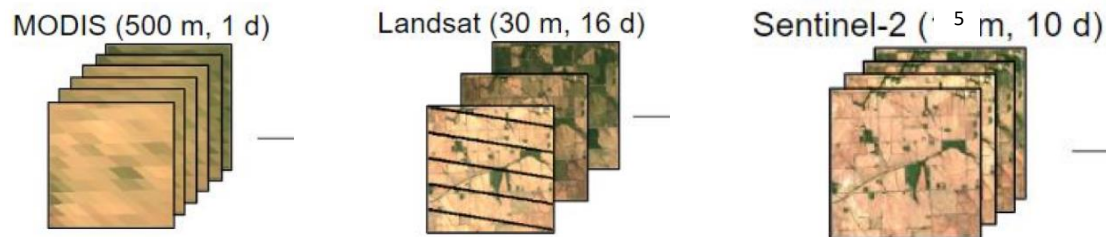
Primeiro autor	Ano	Modelo agrícola	Cultura	Metodo de assimilação	Variavel assimilada	Fonte dos dados			Escala	País
						IAF	US/ET	AP		
de Wit	2007	WOFOST	Milho/Trigo	EnKF	US	-	ERS1/2		País	Europa
Wu	2011	WOFOST	Trigo	EnKF	IAF	MODIS	-		Regional	China
Ines	2013	DSSAT-CSM-Maize	Milho	EnKF	IAF, US	MODIS	AMSR-E		Regional	EUA
Zhao	2013	pyWOFOST	Milho	EnKF	IAF	MODIS	-		Regional	China
Li	2014	WOFOST	Milho	EnKF	IAF	Landsat ETM+	-		Regional	China
Li	2015	DSSAT-CERES	Milho	Recalibração	IAF, BPA	Hyperspectral sensor	-		Talhão	China
Huang	2015	WOFOST	Trigo	4DVar	IAF	MODIS/Landsat	-		Regional	China
Huang	2015	WOFOST	Trigo	SCE	IAF, ET	MODIS	MOD			
Huang	2016	WOFOST	Trigo	EnKF	IAF	Landsat ETM+	-		Regional	China
Xie	2017	CERES-Wheat	Trigo	EnKF	IAF, US	Landsat	Landsat ETM+, OLI and TIRS (or TM)		Regional	China
Novelli	2019	EPIC	Trigo	Recalibração	IAF	Sentinel-2	-		Talhão	Austria
Nearing	2012	DSSAT	Maize	EnKF	IAF, US	Sitético	Sitético		Experimental	-
Kang	2019	SAFY	Maize	EnKF	IAF	Landsat 5, 7 e 8			Regional	EUA
Zhuo	2019	WOFOST	Trigo	EnKF	US	-	Sentinel - 1 e 2		Regional	China
Hu	2019	WOFOST	Cana-de-açúcar	EnKF	IAF, US	LAI 2200 Li-cor	TRIME-PICO-IPH		Experimental	China
Pan	2019	WOFOST	Trigo	EnKF	IAF, US	Sentinel-2	Sentinel-1		Regional	China
Yu	2020	WOFOST	Cana-de-açúcar	ES	AP			Experimental e UAV	Experimental	China
Ma	2021	WOFOST	Trigo	EnKF	NDVI	HJ-1A e B			Regional	China
Giardelli	2019	WARM	Arroz	Calibration	LAI	Landsat-7, 8 e Sentinel 2			Talhão	Italia
Pan	2019	WOFOST	Trigo	EnKF	IAF, US	Sentinel 1 e 2	-		Regional	China

IAF - Índice de área foliar; US - Umidade do solo;

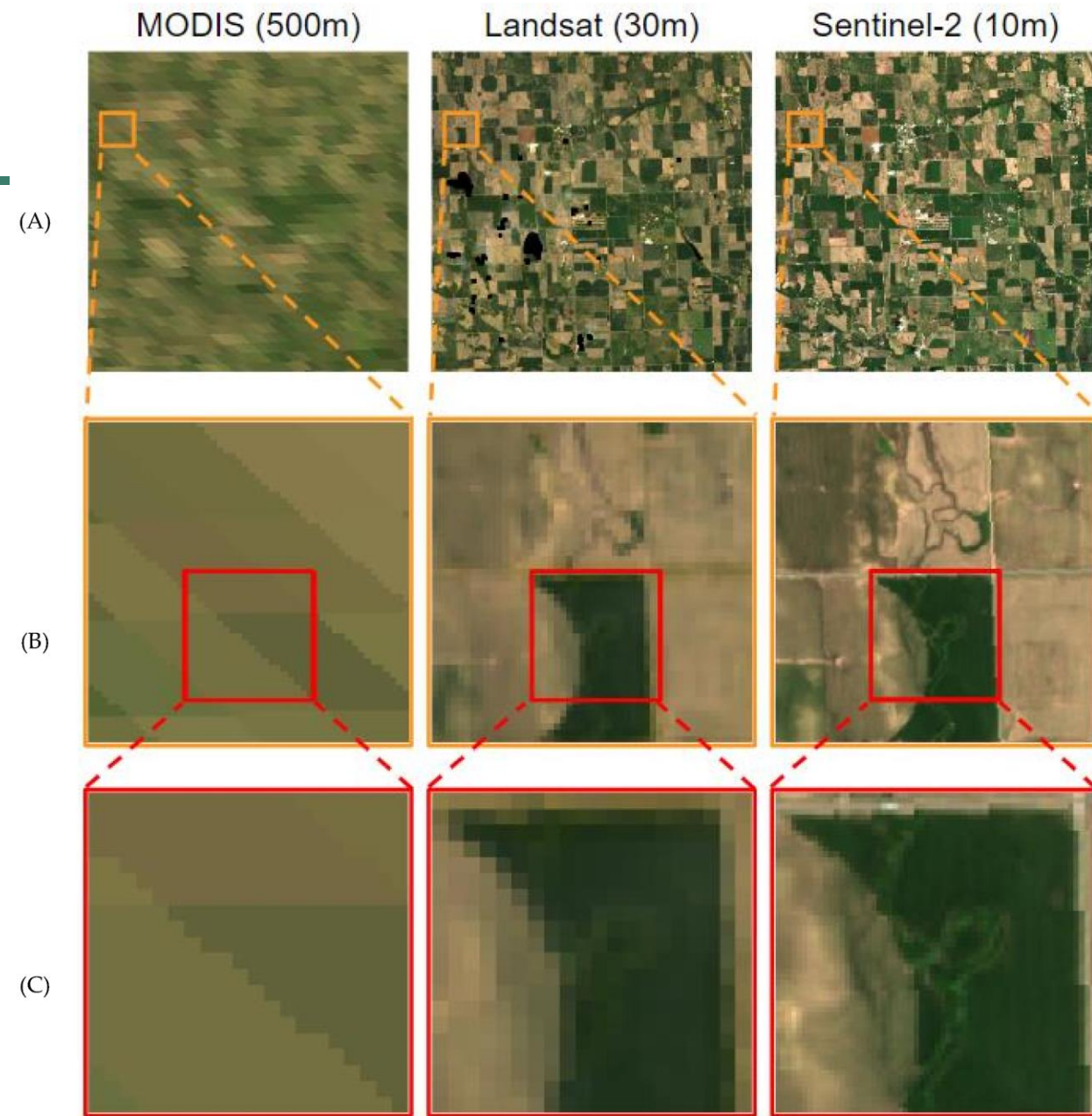
SCE - shuffled complex evolution method—University of Arizona

# Fontes de dados

- Resolução espacial;
- Resolução temporal;
- Resolução espectral;
- Ruídos nos dados (nuvens).



(Lou et al., 2020)



# Métodos

---

- Método forçado ou direto (**forcing methods**);
- Método da calibração (**calibration method**);
- Método da atualização (**updating methods**).



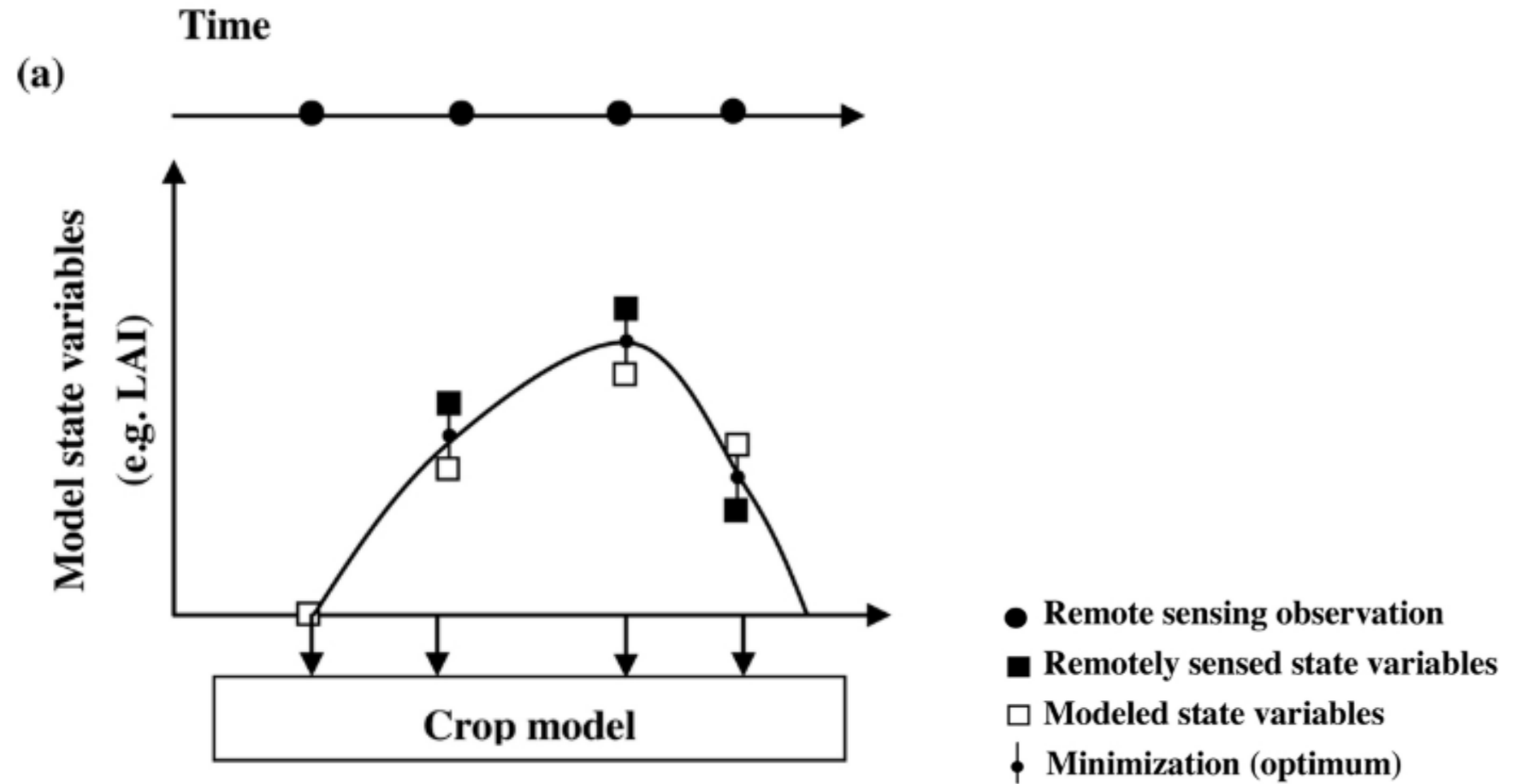
# Calibration methods

---

- Consiste em **ajustar** os **parâmetros iniciais** do modelo para se obter o melhor ajuste entre o **dado observado** e a **variável observada**;
- Escolha dos parâmetros mais **significativos** relacionados a variável assimilada;
- Assume que o modelo consegue representar todas as incertezas do sistema.

(Jin et al., 2018)

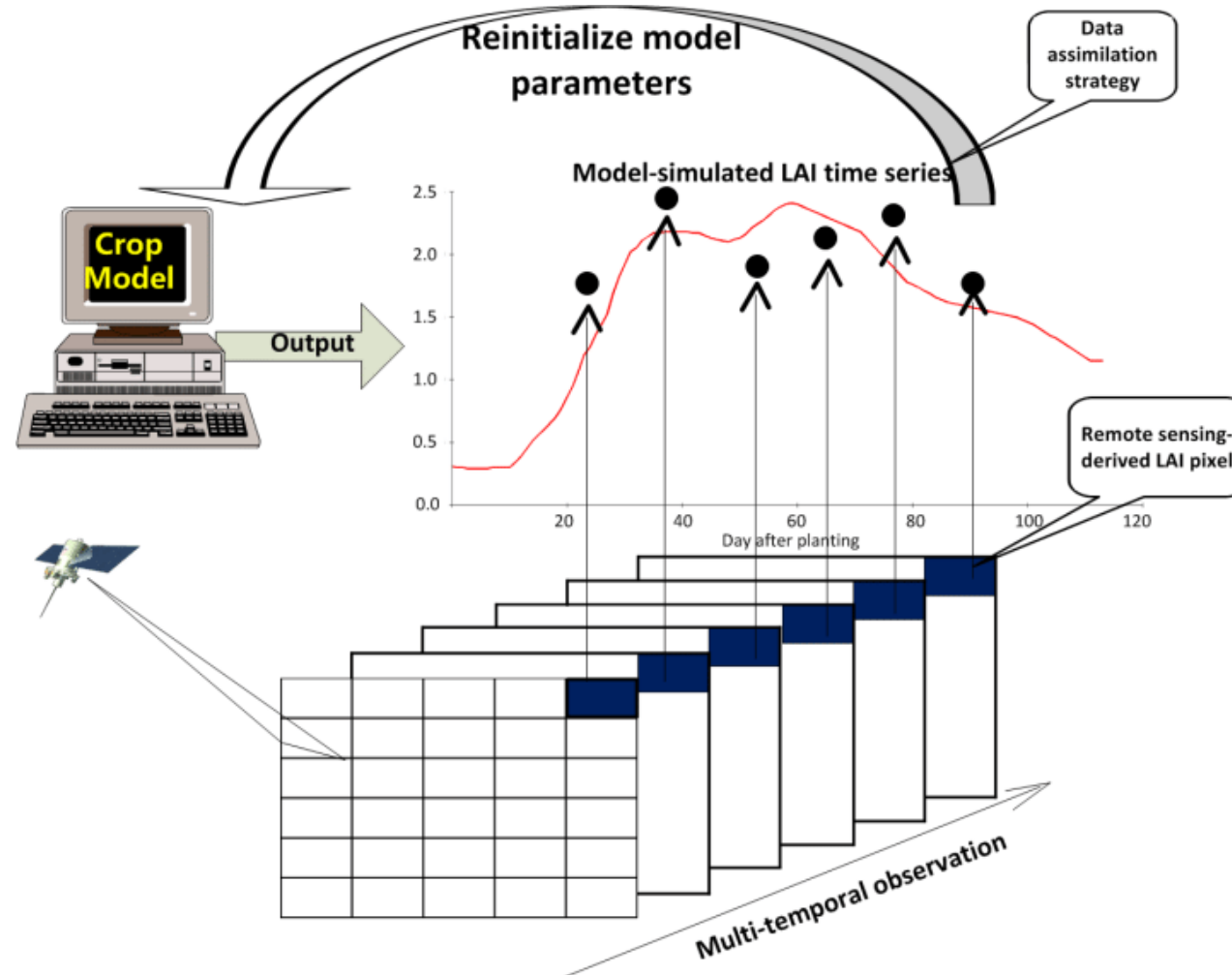
# Calibration methods



(Jin et al., 2018)



# Calibration methods

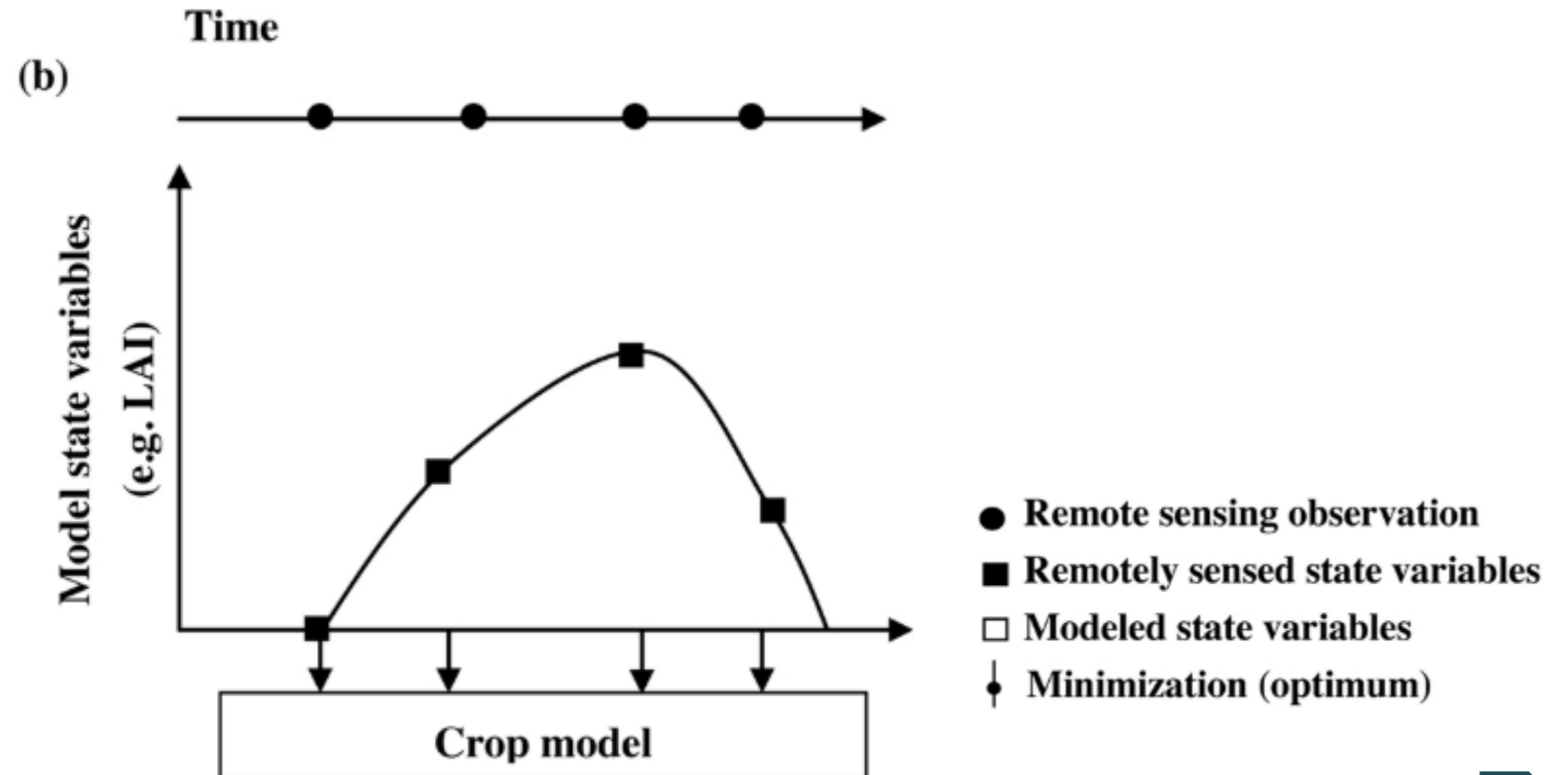


Exemplo de um esquema de assimilação de dados pelo método da calibração

(Pan et al., 2017)

# Forcing methods

- **Substitui** o dado observado **diretamente** nas variáveis de estado do modelo, durante a simulação;



(Jin et al., 2018)

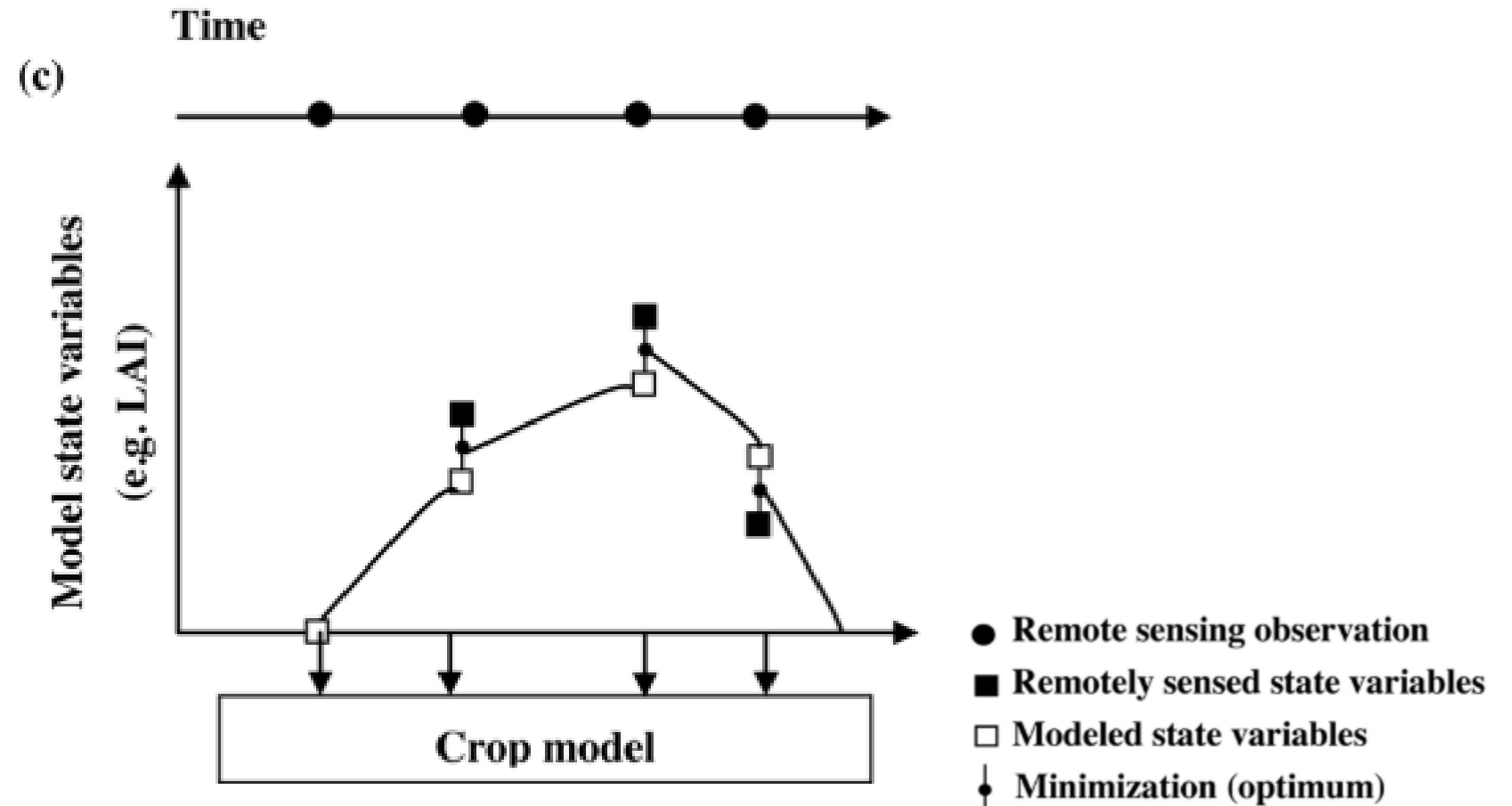
# Updating methods

- Esse método supõe que **diminuir as incertezas** da simulação de uma variável no tempo  $t$  vai **melhorar a acurácia** das **simulações dos dias seguintes**;
- Para isso o método de atualização **considera** tanto as **incertezas** dos **dados observados**, **parametrização e estrutura do modelo**, para **atualizar** as variáveis de estado do modelo **sequencialmente**, quando observações estão disponíveis.



(Jin et al., 2018; Hu et al., 2019)

# Updating methods



(Jin et al., 2018; Hu et al., 2019)

# Updating methods

Algorithm	Basic formula	Algorithm mechanism and characteristics	References
KF	$X_k^f = M_{k-1,k} X_{k-1}^e$ $X_k^e = X_k^f + K_k (Y_k^o - H_k X_k^f)$ $K_k = (H_k P_k)^T [H_k (H_k P_k)^T + R_k]^{-1}$	The observation value is used to adjust model state value of the trajectory when the observed value exists, the current time optimal state values is obtained, and then the current state estimation value of the model is reinitialized; continuing forward integration, until there are the next observational data. It can predict real-time and update the state, but it isn't suitable when the $H_k$ and $M_k$ is nonlinear.	Aubert et al. (2003) and Pellenq and Boulet (2004)
EnKF	$X_{i,k+1}^f = M_{k,k+1} (X_{i,k}^e + w_{i,k}), w_{i,k} \sim N(0, Q_k)$ $X_{i,k+1}^e = X_{i,k+1}^f + K_{k+1} [Y_{k+1}^o - H(X_{i,k+1}^f) + v_{i,k}]$ $K_{k+1} = P_{k+1}^f H^T (H P_{k+1}^f H^T + R_k)^{-1}$ $X_{i,k+1}^e = \frac{1}{N} \sum_{i=1}^N X_{i,k+1}^e$	The ensemble forecast and Kalman filter are combined to calculate the prediction error covariance by Monte-Carlo method. It can be used for nonlinear systems of data assimilation. It is computationally less intensive and is easy for parallel computation.	Evensen (1994), Crow and Wood (2003) and Schläpfer and Richter (2002)
3DVAR	$J(X) = (X - X^b)^T B^{-1} (X - X^b) + (Y - HX)^T R^{-1} (Y - HX)$	All observation of data assimilation window T is used to adjust the predictive values of trajectory in model, constructing the cost function $J(X)$ shows the error between the analysis field and true value, solving the minimization optimal solution of cost function. Because the cost function is very difficult to calculate directly, it needs the help of the gradient function $\Delta J(X)$ and the adjoint model.	Lorenc et al. (2000)
4DVAR	$\Delta J(X) = 2B^{-1}(X - X^b) - 2H^T R^{-1}(Y - HX)$ $J(X) = (X - X^b)^T B^{-1} (X - X^b) + \sum_{k=0}^T (Y_k - H_k(M_k(M_{k-1}(\dots(M_1(X))))))^T R^{-1} (Y_k - H_k(M_k(M_{k-1}(\dots(M_1(X))))))$ $\Delta J(X) = 2B^{-1}(X - X^b) - 2 \sum_{k=0}^T \left\{ M_k^T \dots M_{k-1}^T M_k^T H_k^T R_k^{-1} [Y_k - H_k(M_k(M_{k-1}(\dots(M_1(X)))))] \right\}$	The change of state $M_t$ with time is considered based on the 3DVAR, The value of optimal estimation of state at the time of T is comprehensive consideration $M_t$ results. It also needs the help of the gradient function and adjoint model. It is more computationally intensive because of the consideration of $M_t$ .	Trémolet (2007)
PF	$p(X_k^e   X_{1:k}^e) \approx \sum_{i=1}^N w_{i,k} \delta(X_k^e - X_{i,k}^e)$ $w_{i,k} \propto \frac{p(X_{i,k}^e   X_{1:k}^e)}{q(X_{i,k}^e   X_{1:k}^e)}, \sum_{i=1}^N w_{i,k} = 1$	The posterior probability distribution is sampled with a group of particles, and then the posterior probability distribution is analyzed according to sampling point. It isn't subject to the limitation of nonlinear system and the assumptions of Gauss distribution. It is suitable for parallel computation.	Moradkhani et al. (2005)
HBM	$p(\theta_D, \theta_P, Y X) \approx p(X_1 Y_1, \theta_D) \dots p(X_n Y_n, \theta_D) p(Y_1 \theta_P, Y_2, \dots, Y_n) \dots p(Y_n \theta_P) p(\theta_D) p(\theta_P)$	The data assimilation is split into data, process and parameters of the three layers. The conditional probability model is defined for each layer. The data assimilation is converted to the problem of reasoning process and the posterior probability distribution of parameters in the given data conditions.	Sahu et al. (2009) and Plant and Holland (2011)

(Jin et al., 2018)

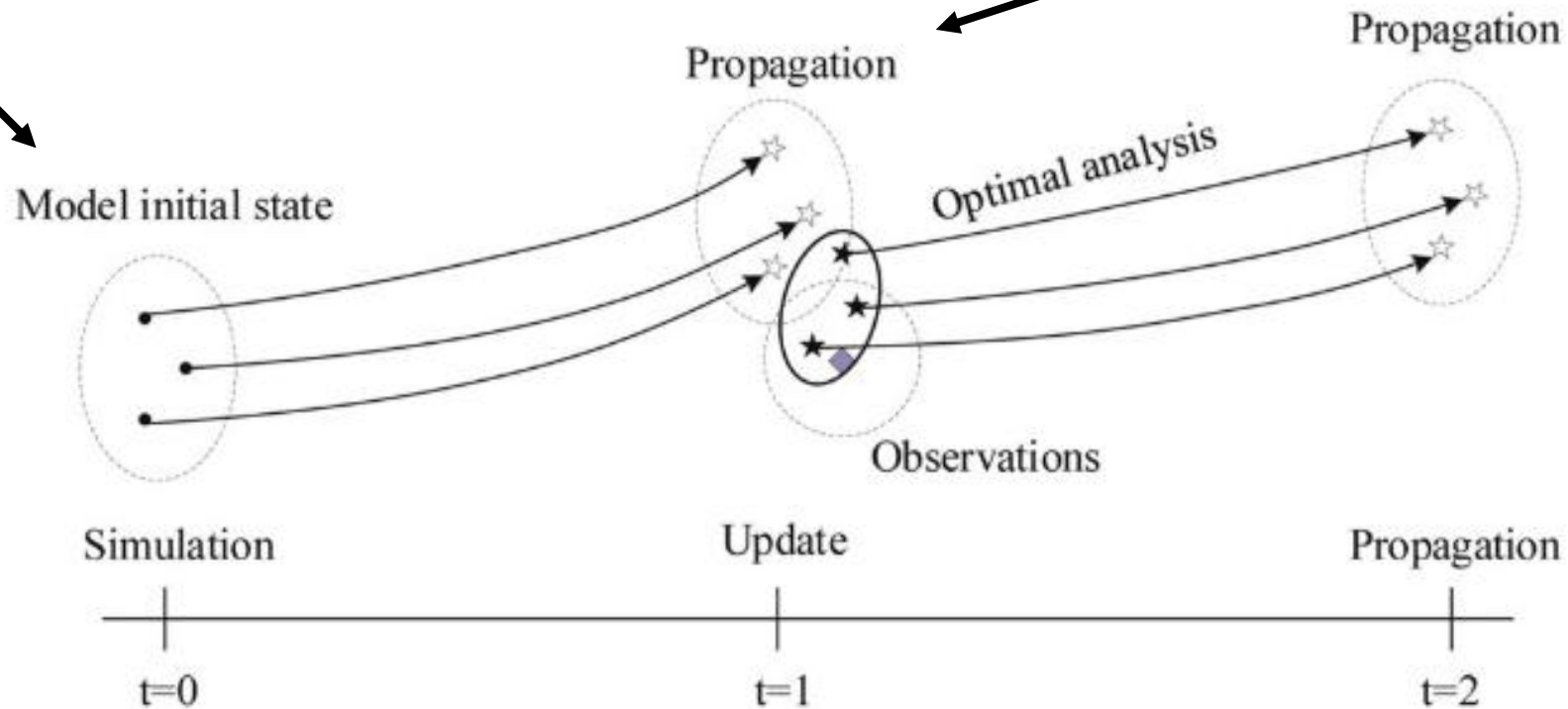
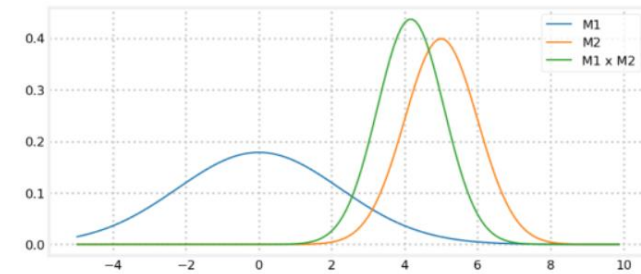
# Updating methods

---

- Ensemble Kalman Filter (EnKF);
- Utiliza o método de **Monte Carlo** para prever o erro de covariância do modelo;
- Podem ser utilizados com **sistemas não lineares**;
- Computacionalmente **eficiente**.

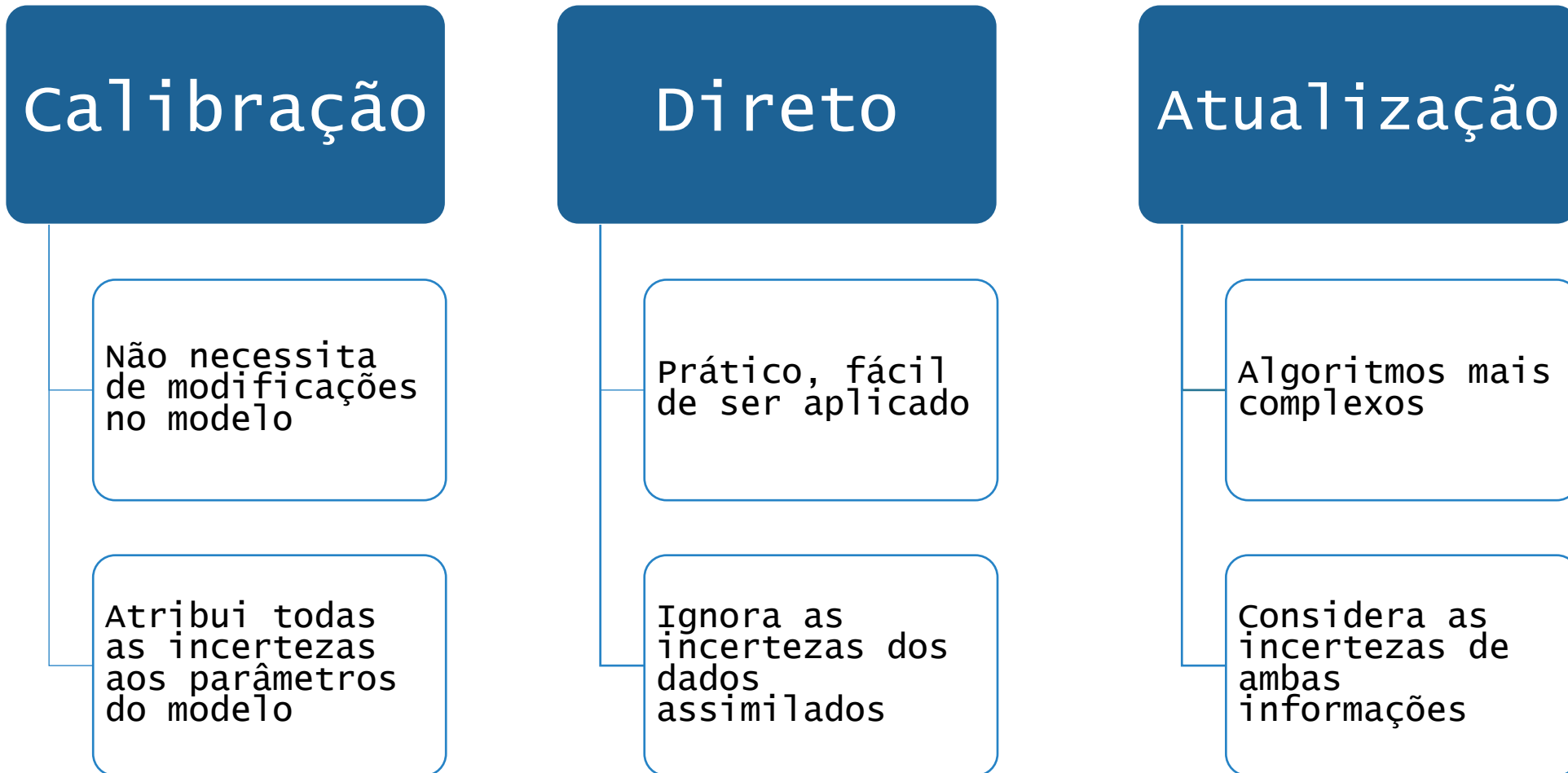
# Updating methods

Método de Monte Carlos



Representação do método de assimilação de dados pelo filtro de Kalman *ensemble*. Adaptado de Huang et al., (2019).

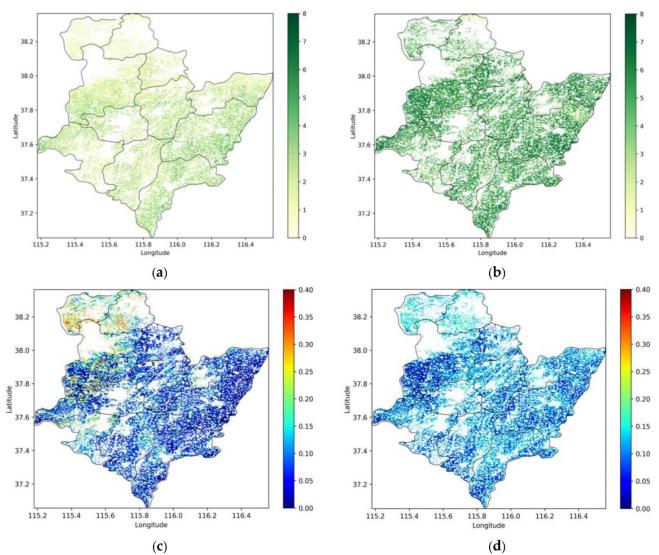
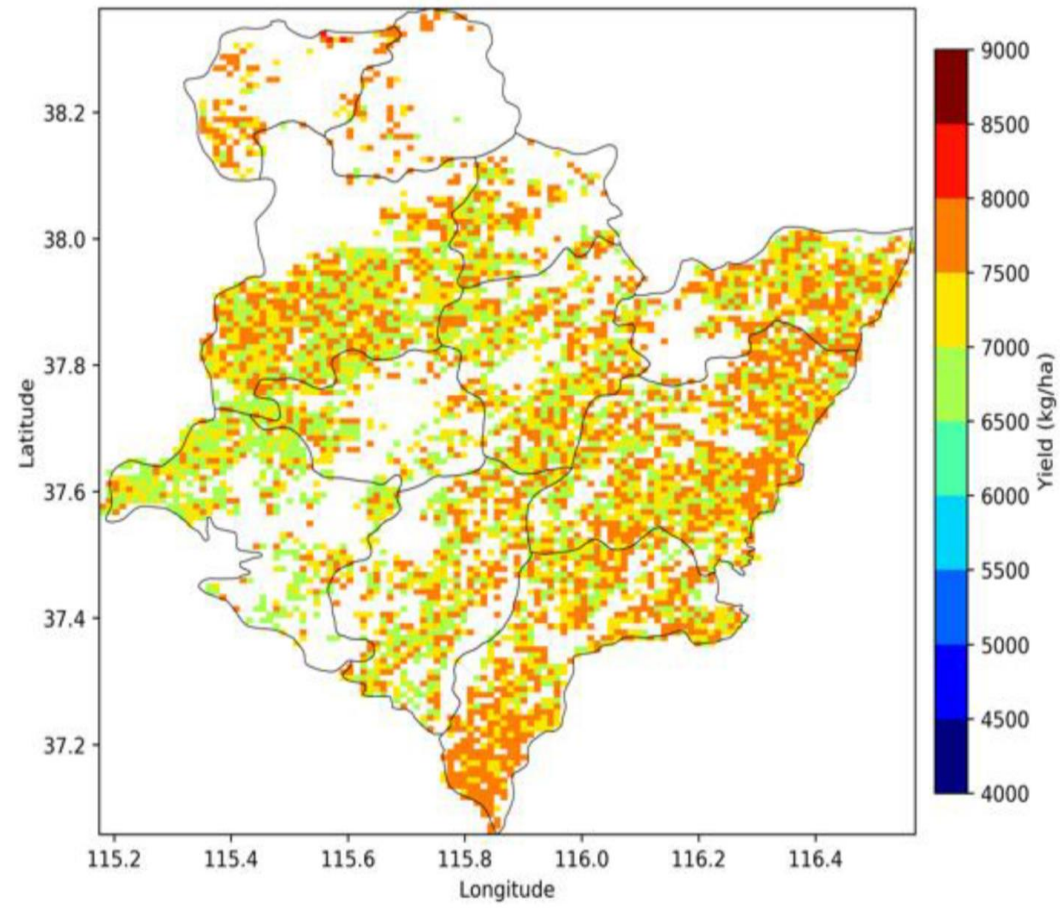
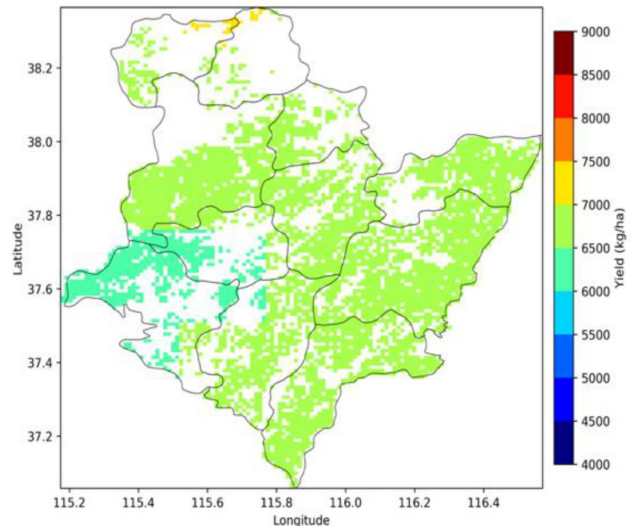
# Método de assimilação



(Hu et al., 2019)



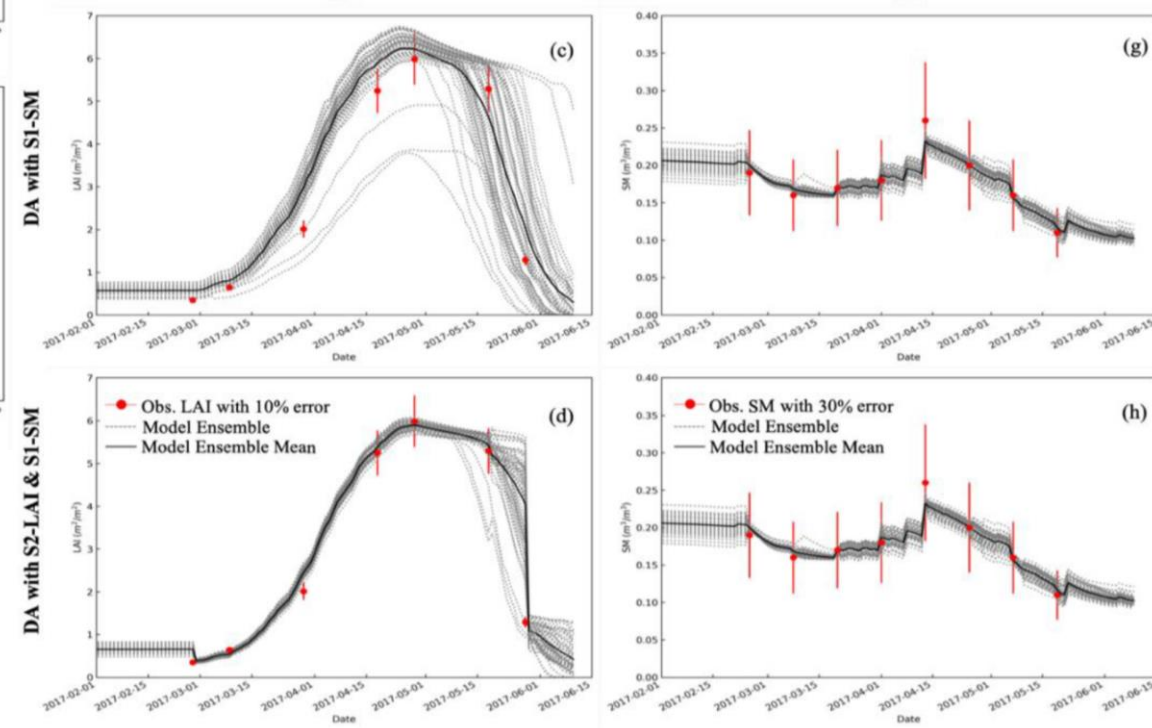
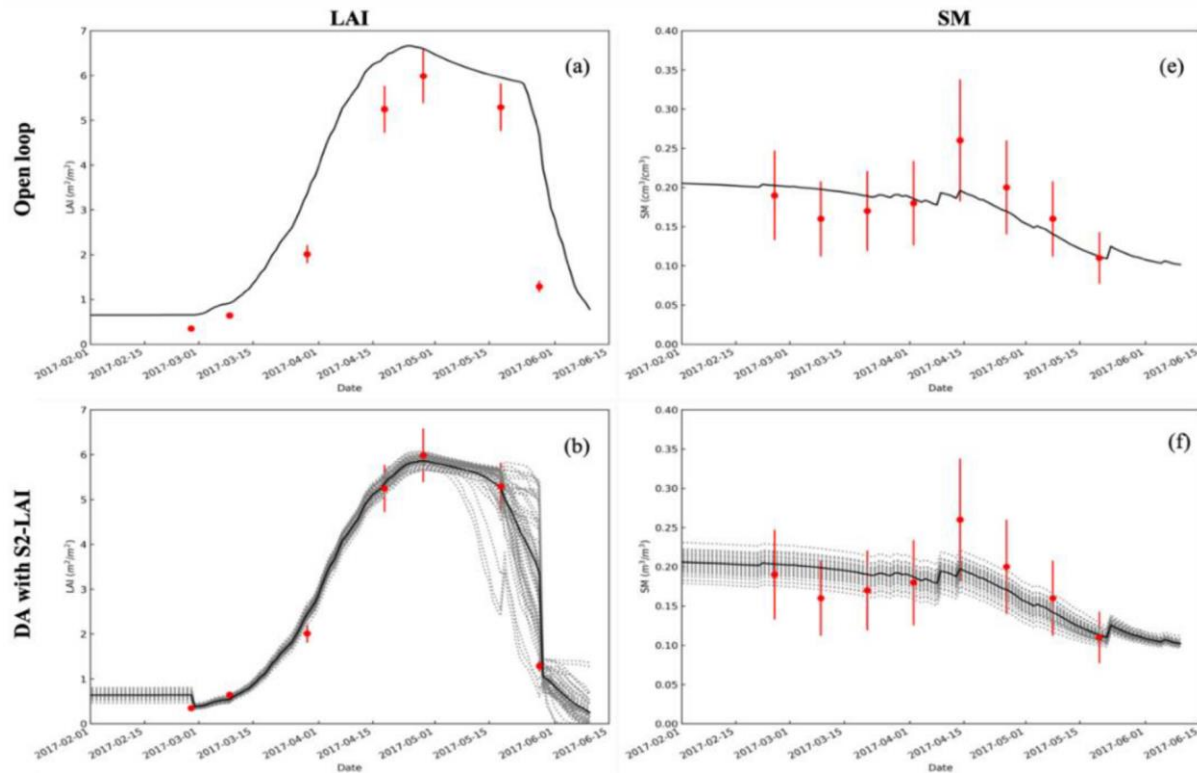
# Aplicação em modelos



(Pan et al., 2019)

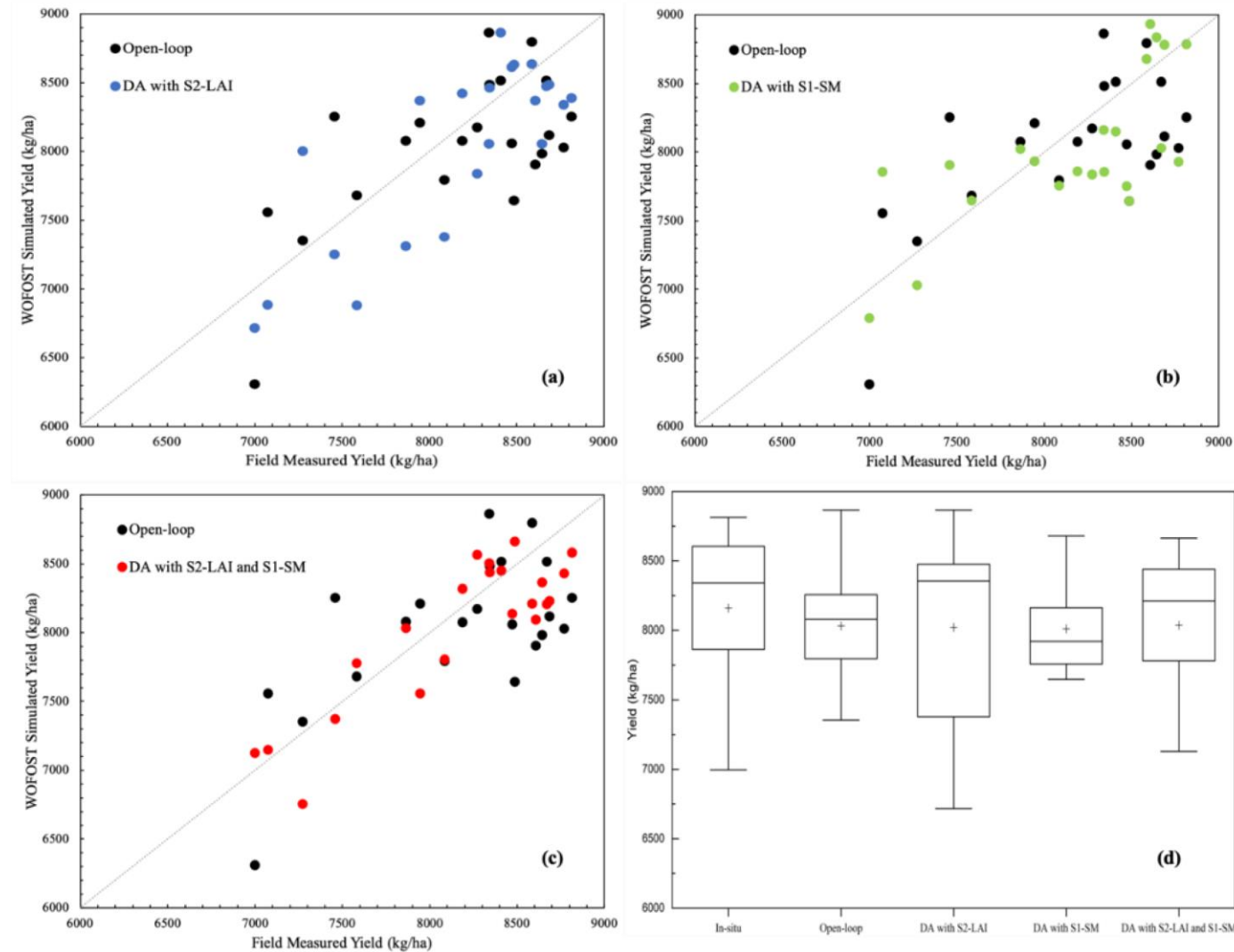
Produtividade simulada pelo modelo WOFOST junto a assimilação de IAF e US.

# Aplicação em modelos



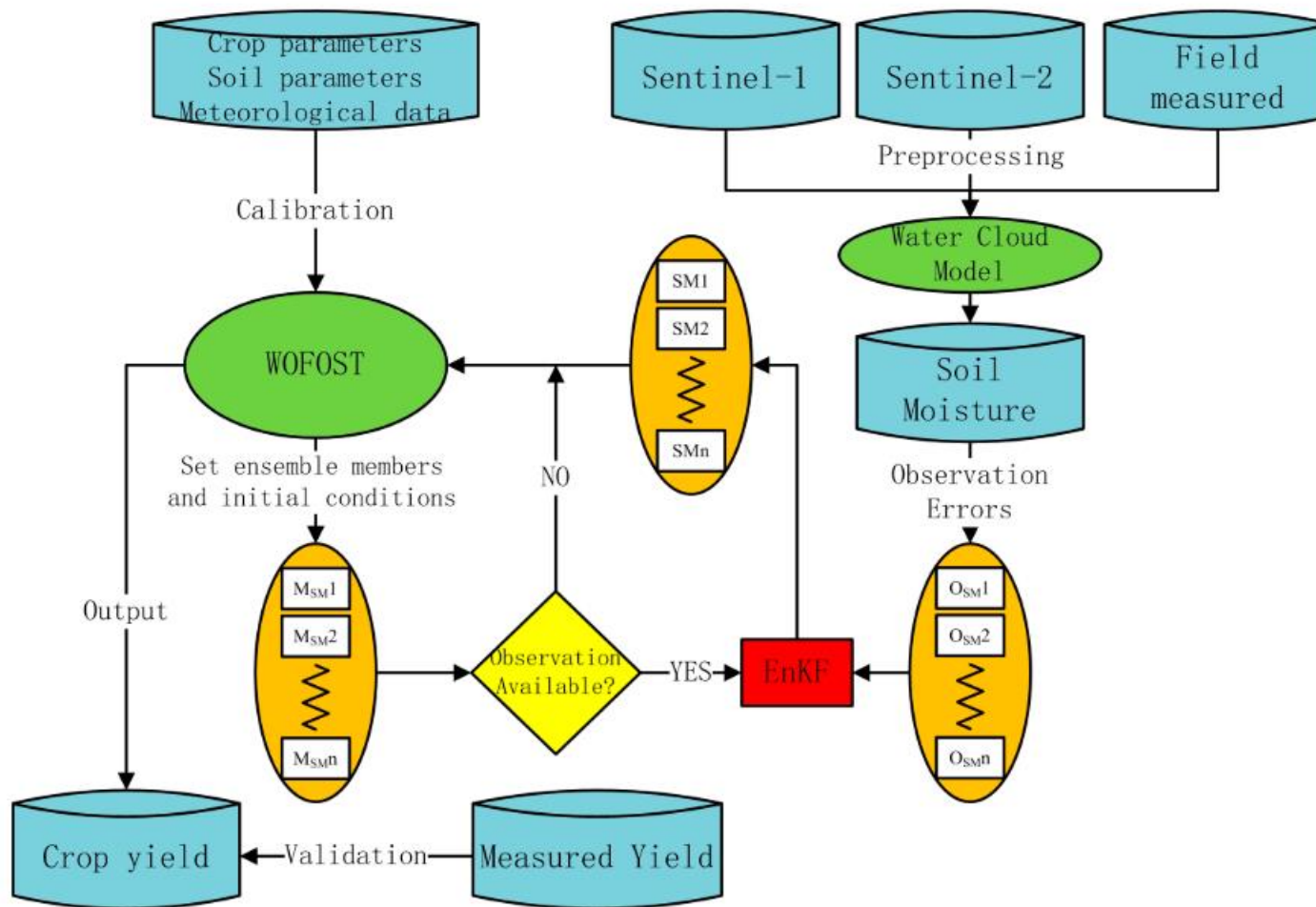
(Pan et al., 2019)

# Aplicação em modelos



(Pan et al., 2019)

# Aplicação em modelos



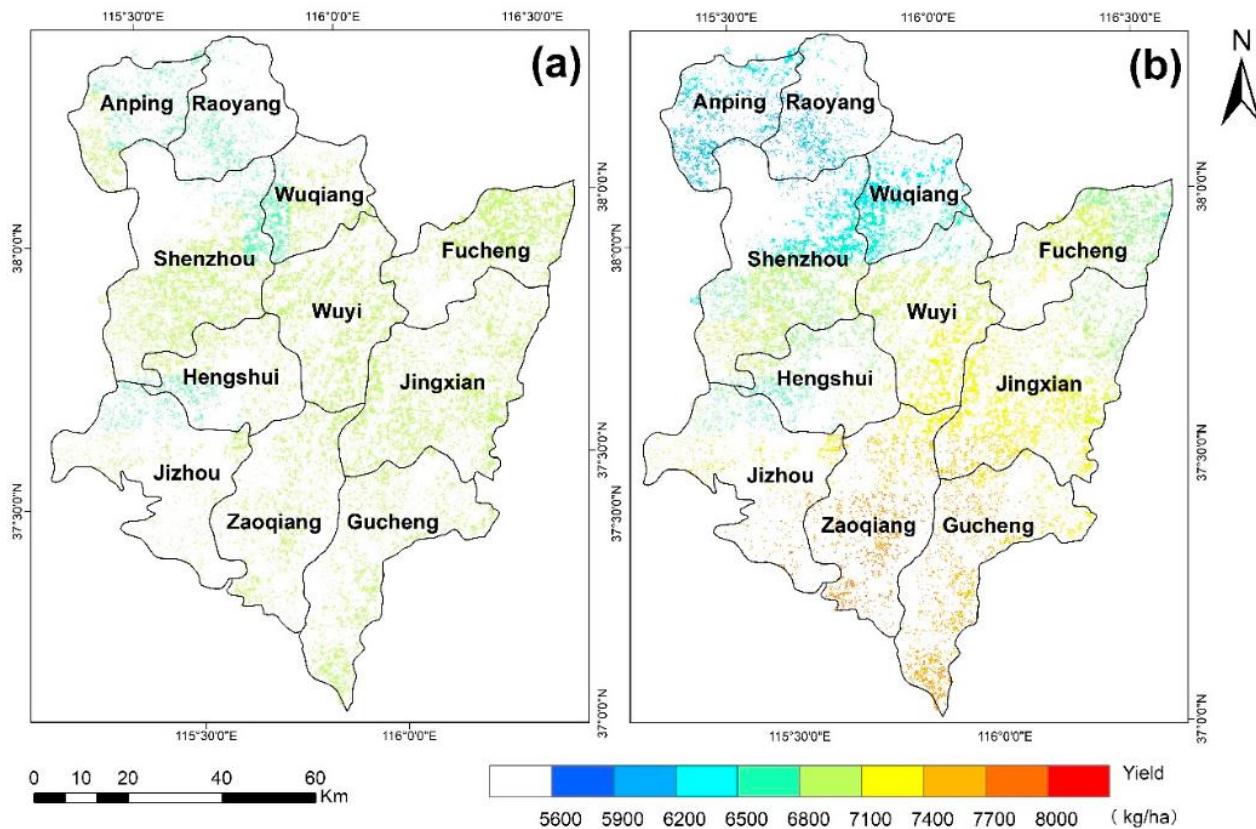
Fluxograma para estimativa de produtividade do trigo usando o algoritmo EnKF.

Figure 3. Flowchart for the winter wheat yield estimation using the EnKF-based assimilation algorithm.

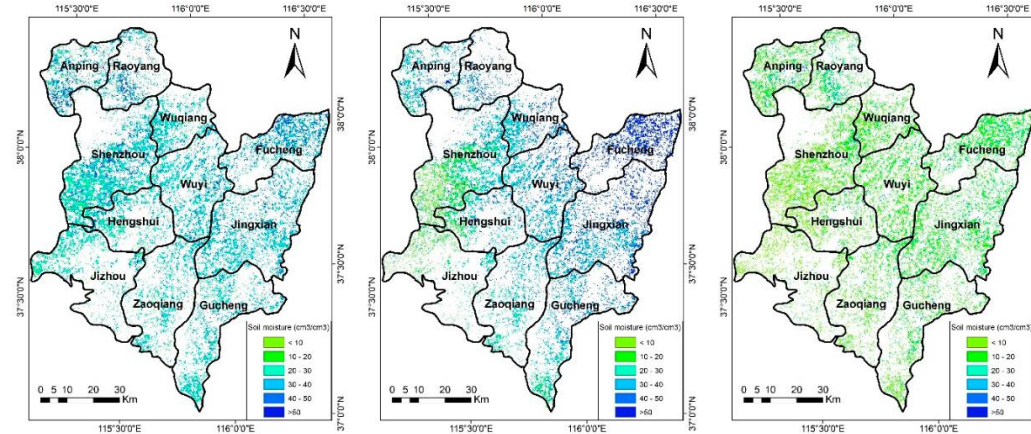
(Zhuo et al., 2019)



# Aplicação em modelos

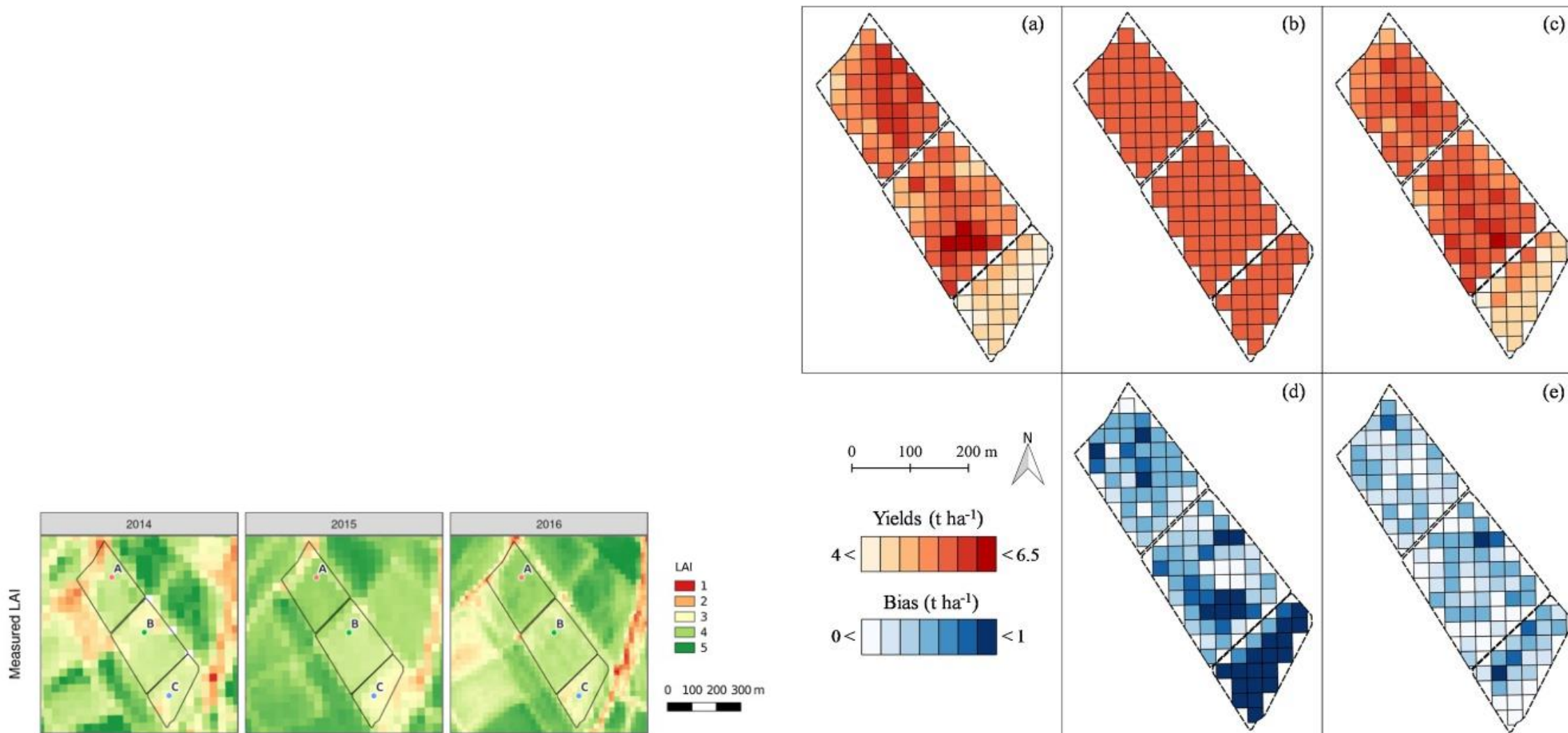


(Zhuo et al., 2019)



Simulação da produtividade do trigo com o modelo WOFOST (a) sem assimilação de dados e (b) com assimilação de umidade do solo.

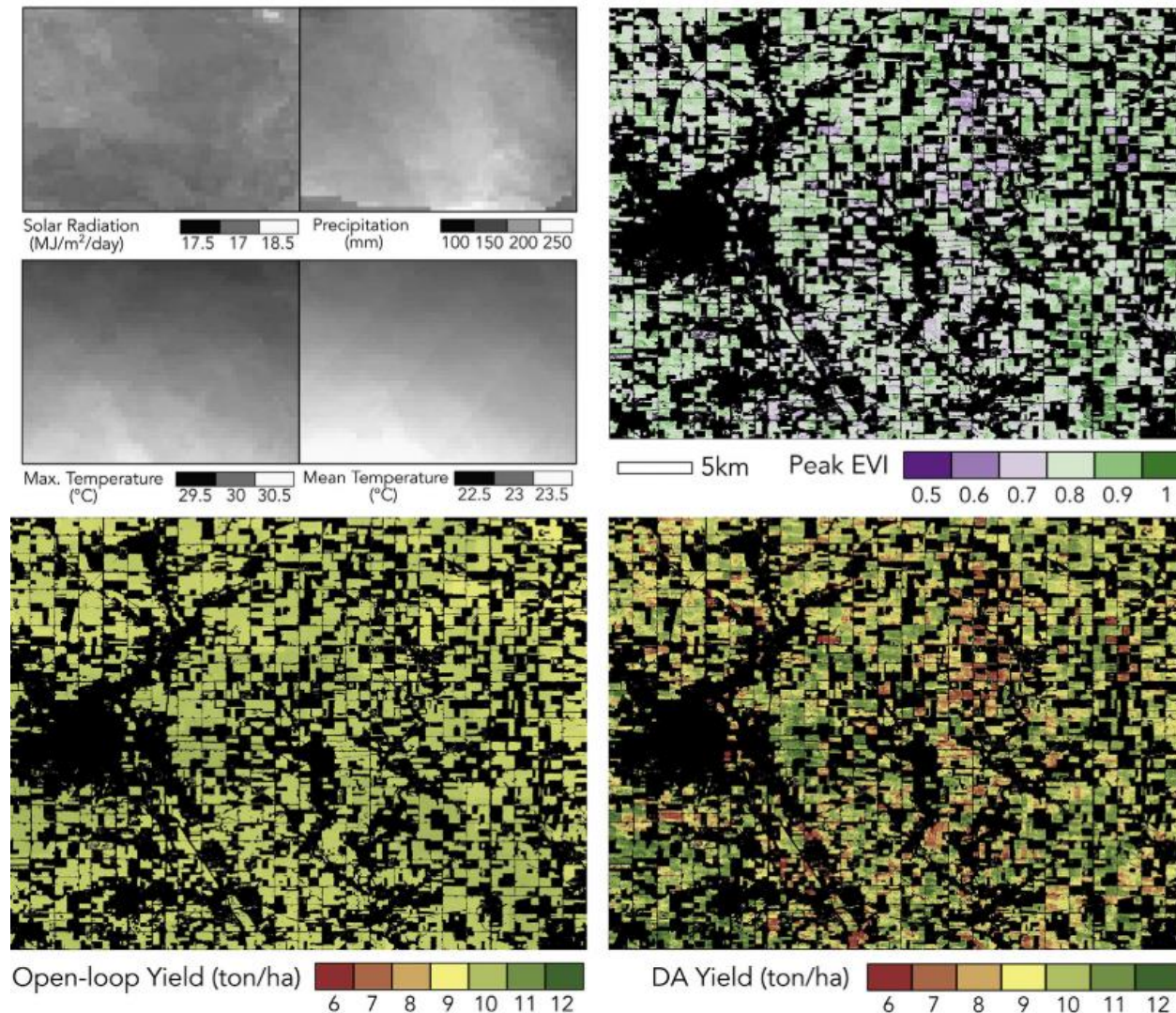
# Aplicação em modelos



(Gilardeelli et al., 2019)



# Aplicação em modelos



(Kang and Özdoğan, 2019)

Fig. 14. Spatial variations of weather features, peak EVI, open-loop yield, and data assimilation yield for Story county, IA in 2012. The weather features (upper left panel) include average daily solar radiation from April to October, cumulative precipitation from June to August, average daily maximum temperature from June to August, and average daily temperature from June to August. Note that the EVI map shows signs of the SLC-off gaps, since only Landsat ETM+ data was available in 2012.

# Aplicação em modelos

---

- Quantidade de dados observados assimilados;
- Estágio fenológico em que os dados são assimilados;
- Calibração prévia do modelo, para representar a variação genotípica.

(Fattori Junior et al., 2022)



# Referências

FATTORI JUNIOR, Izael Martins Fattori; VIANNA, Murilo; MARIN, Fábio Ricardo. Assimilating leaf area index data into a sugarcane process-based crop model for improving yield estimation. *European Journal of Agronomy*, v. 136, p. 126501, 2022. <<https://doi.org/10.1016/j.eja.2022.126501>>

GILARDELLI, Carlo et al. Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *European Journal of Agronomy*, v. 103, p. 108-116, 2019. <<https://doi.org/10.1016/j.eja.2018.12.003>>

HU, Shun et al. Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. *Field Crops Research*, v. 232, p. 49-61, 2019. <<https://doi.org/10.1016/j.fcr.2018.12.009>>

HUANG, Jianxi et al. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, v. 276, p. 107609, 2019. <<https://doi.org/10.1016/j.agrformet.2019.06.008>>

JIN, Xiuliang et al. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, v. 92, p. 141-152, 2018. <<https://doi.org/10.1016/j.eja.2017.11.002>>

LABBE, Roger. Kalman and bayesian filters in python. *Chap*, v. 7, n. 246, p. 4, 2014.

LUO, Yunan et al. STAIR 2.0: A generic and automatic algorithm to fuse modis, landsat, and Sentinel-2 to generate 10 m, daily, and cloud-/gap-free surface reflectance product. *Remote Sensing*, v. 12, n. 19, p. 3209, 2020. <<https://doi.org/10.3390/rs12193209>>

KANG, Yanghui; ÖZDOĞAN, Mutlu. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment*, v. 228, p. 144-163, 2019. <<https://doi.org/10.1016/j.rse.2019.04.005>>

PAN, Zhuokun et al. Remote sensing of agricultural disasters monitoring: recent advances. *economy and society*, v. 10, n. 13, p. 16, 2017. <[https://www.researchgate.net/publication/320548641\\_Remote\\_Sensing\\_of\\_Agricultural\\_Disasters\\_Monitoring\\_Recent\\_Advances](https://www.researchgate.net/publication/320548641_Remote_Sensing_of_Agricultural_Disasters_Monitoring_Recent_Advances)>

WANG, Jie et al. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 154, p. 189-201, 2019. <<https://doi.org/10.1016/j.isprsjprs.2019.06.007>>

ZHUO, Wen et al. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. *Remote Sensing*, v. 11, n. 13, p. 1618, 2019. <<https://doi.org/10.3390/rs11131618>>

# Agradeço a atenção!

---

1100222 - Modelagem de Crescimento de Culturas Agrícolas  
LEB5048 - Modelagem de Culturas Agrícolas I

Izael Martins Fattori Junior  
E-mail: [izael.fattori@usp.br](mailto:izael.fattori@usp.br)

11/03/2022  
Piracicaba