Convective self-aggregation as a paradigm to understand tropical cloud organization: representation by models, quantification by metrics

Giovanni Biagioli
LMD/IPSL, CNRS, Paris (France)

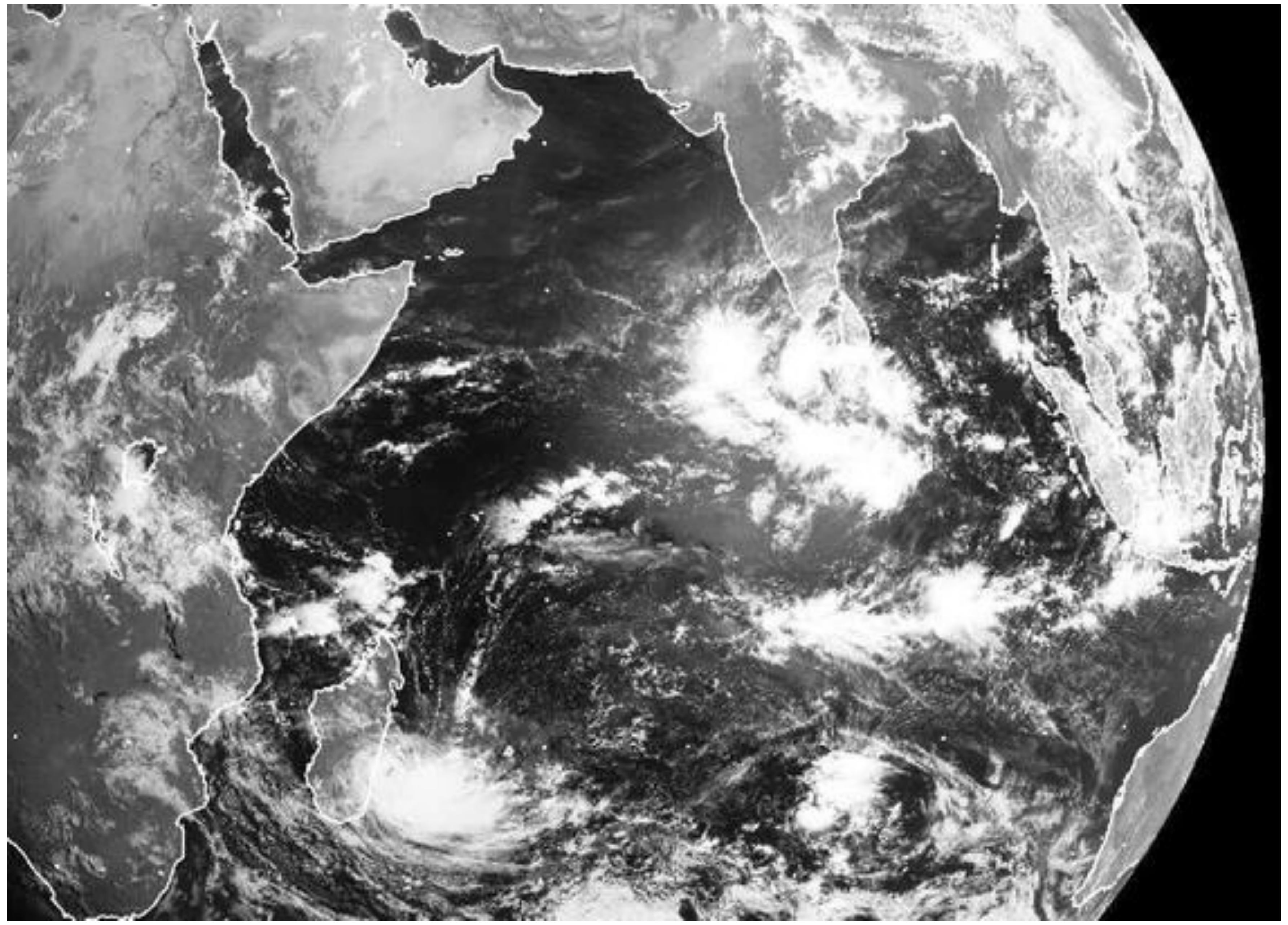
2025 UTCC PROES Meeting
Paris, 21 May 2025

Deep convective organization in the Tropics

Organized deep convection is ubiquitous in the tropical atmosphere.

- Strong impacts on the circulation, radiation budgets, hydrological cycle.

- Extreme events often associated with organized convective systems.
  MCSs contribute ~50% of total tropical rainfall (Nesbitt et al., 2006; Tan et al., 2013).

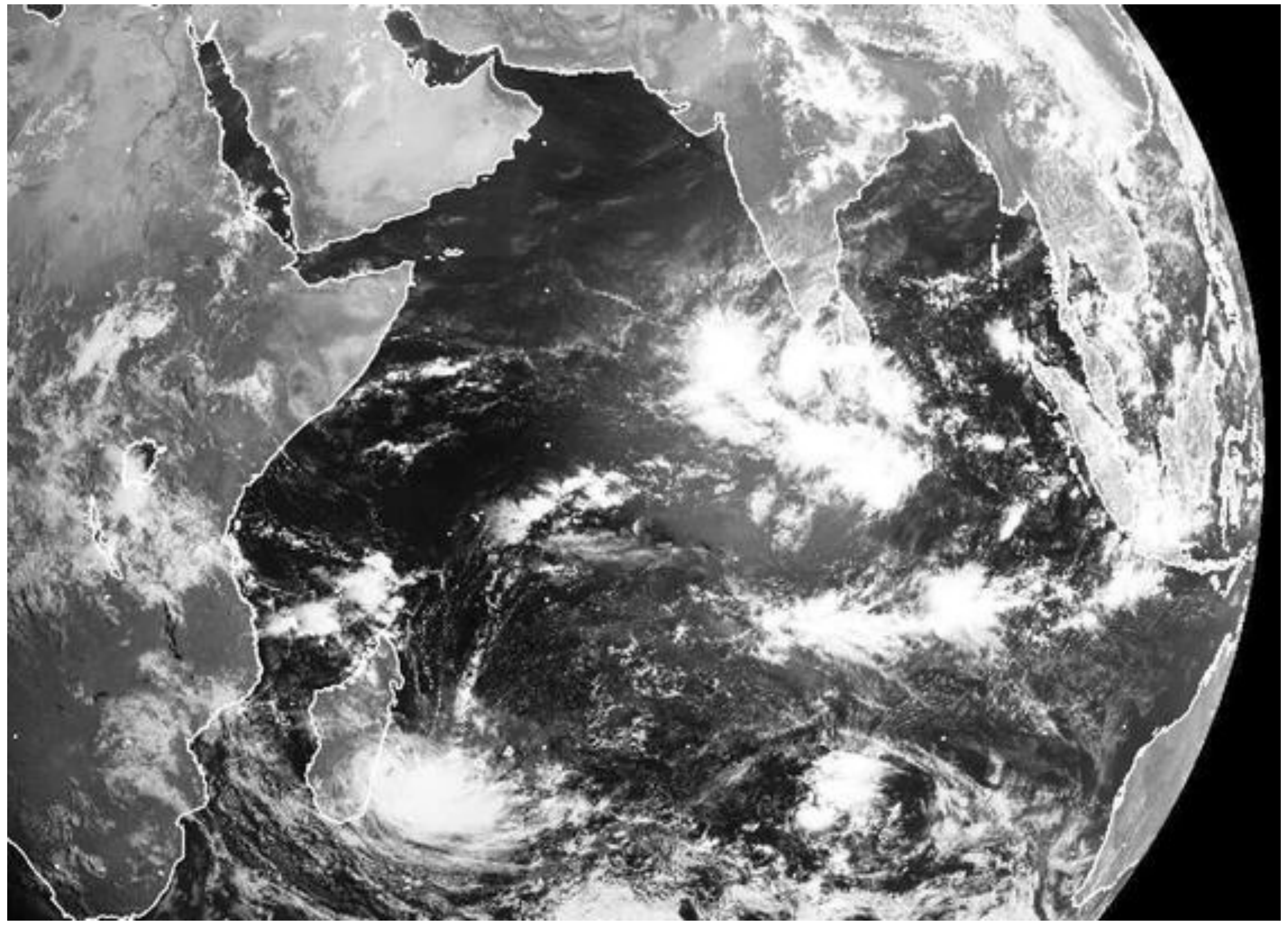


From *Holloway et al. (2017)*

Deep convective organization in the Tropics

Organized deep convection is ubiquitous in the tropical atmosphere.

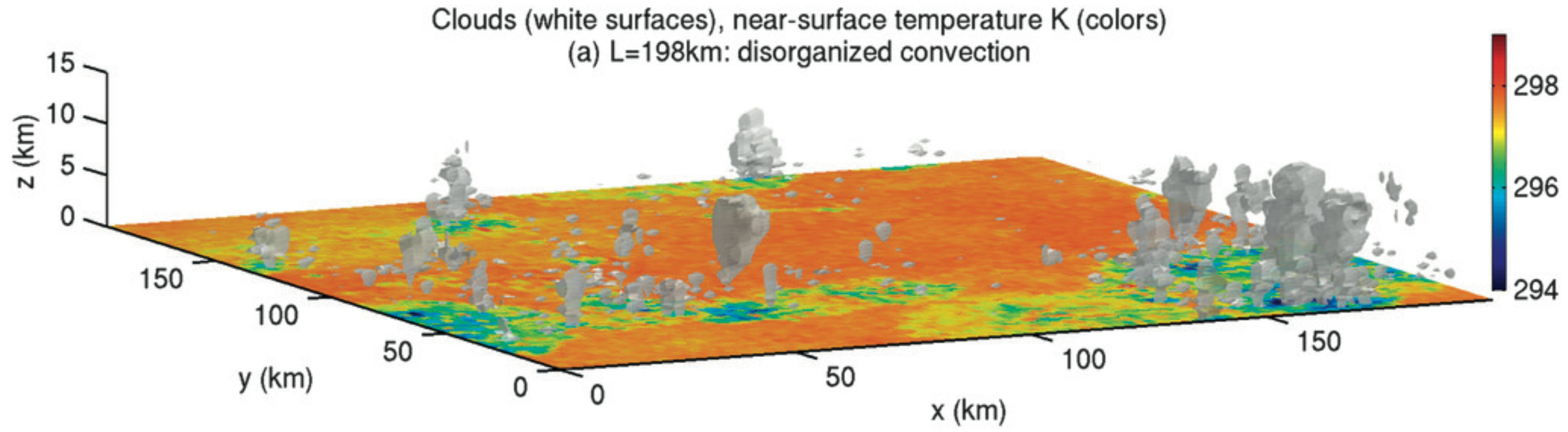
- Strong impacts on the circulation, radiation budgets, hydrological cycle.
- Extreme events often associated with organized convective systems.
  MCSs contribute ~50% of total tropical rainfall (Nesbitt et al., 2006; Tan et al., 2013).
- What is convective organization in practice?
  One possible definition: "**nonrandomness in meteorological fields in convecting regions**" (Mapes and Neale, 2011).
- **No unanimous definition of convective organization ➔** difficult to quantify in models and observations.



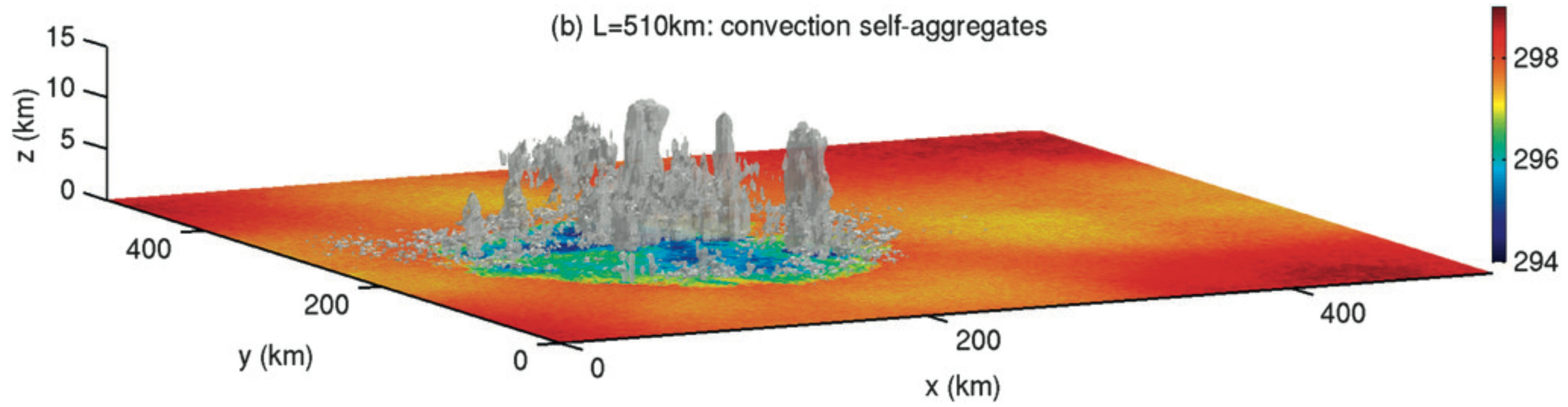
From *Holloway et al. (2017)*

An intriguing behavior: convective self-aggregation

One particular regime of organization found in simulations is the spontaneous aggregation (**self-aggregation**) of convection.



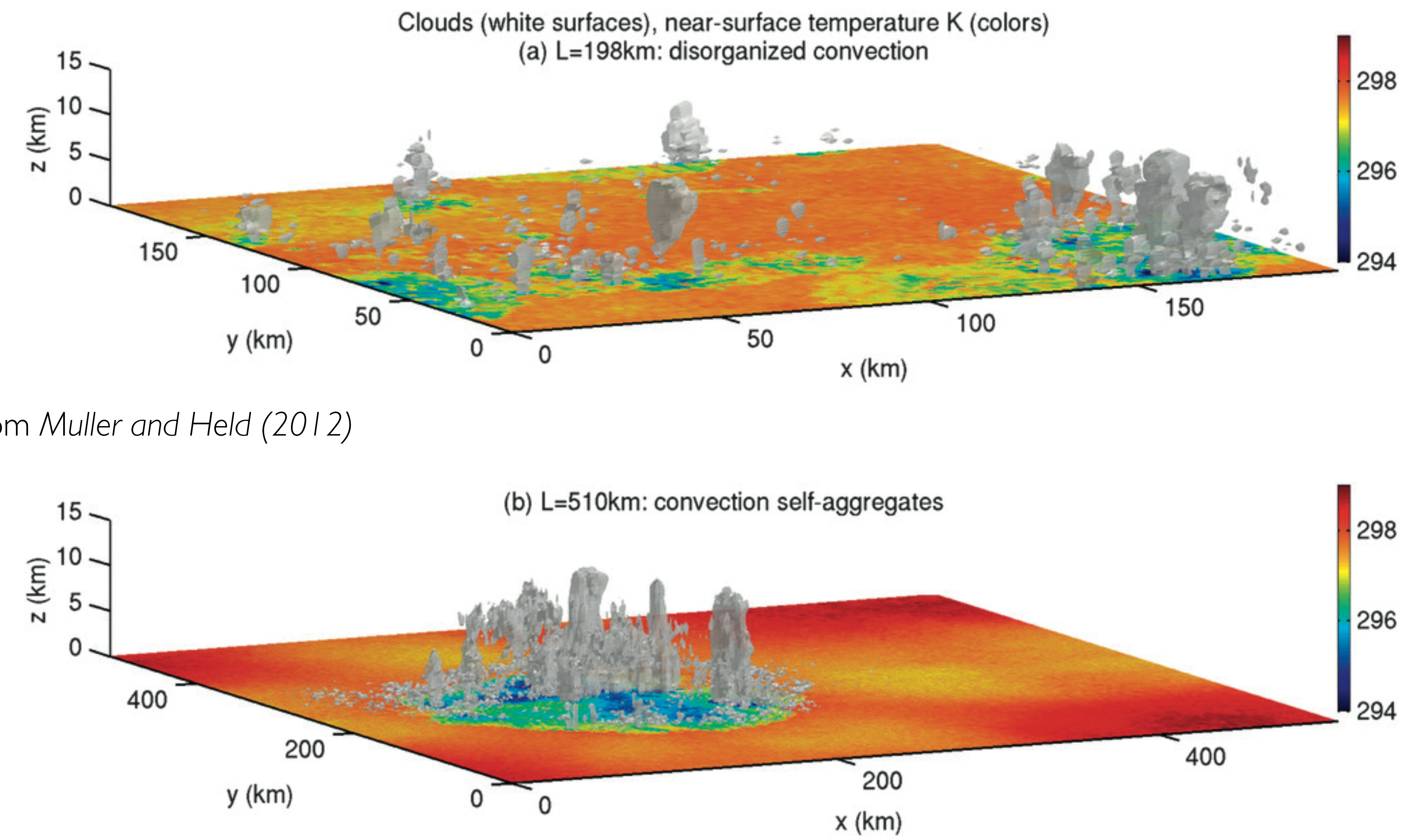
From *Muller and Held (2012)*



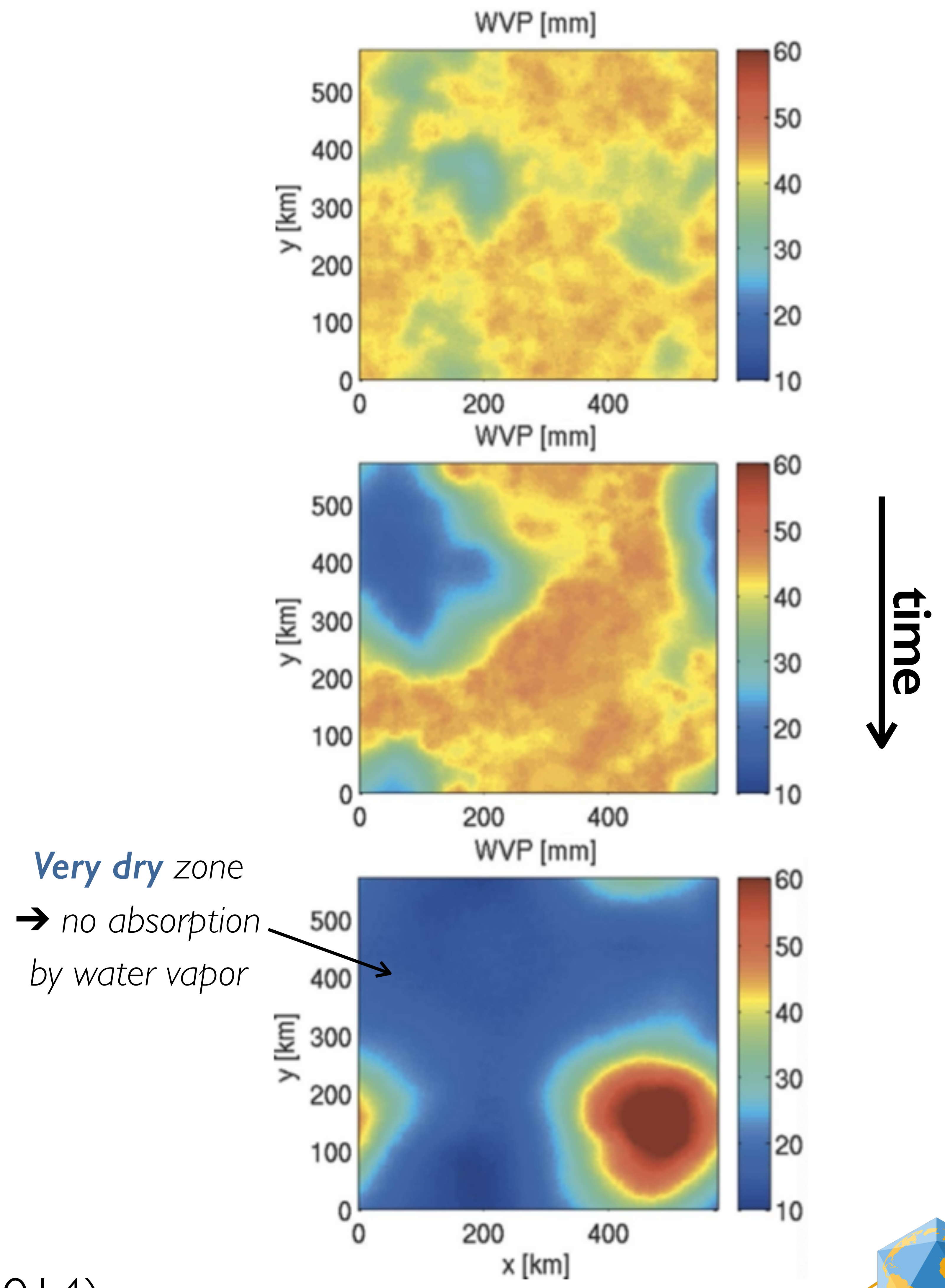
Why is it such a hot topic in current climate research?

An intriguing behavior: convective self-aggregation

One particular regime of organization found in simulations is the **spontaneous** aggregation (**self-aggregation**) of convection.



From *Muller and Held (2012)*



From *Bretherton et al. (2005)*

Why is it such a hot topic in current climate research?

- It can act as a safety valve to regulate tropical climate.
- It has the potential to provide a negative feedback to global warming (Emanuel et al., 2014).

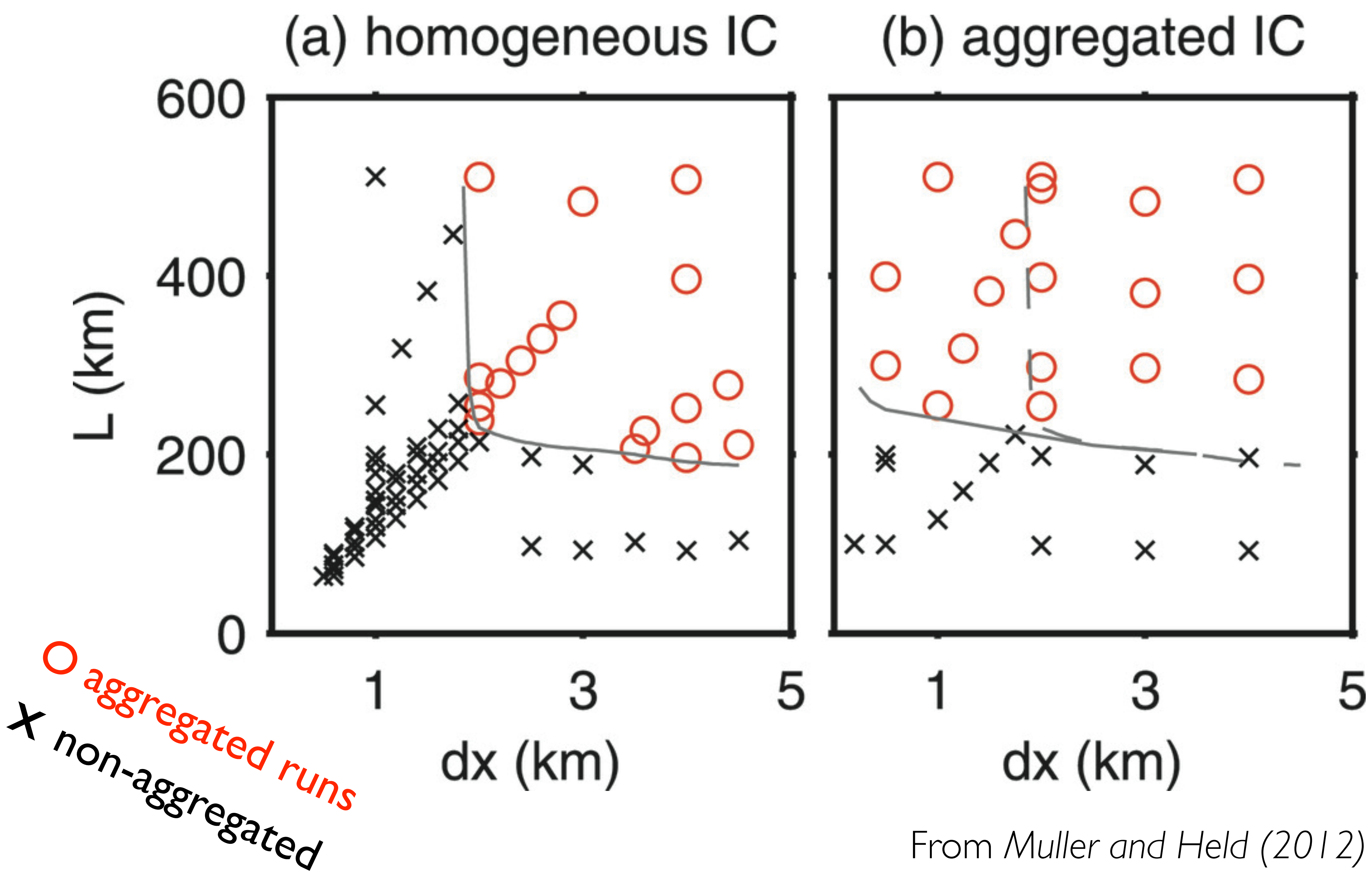
So far so good, but…

- The occurrence of self-aggregation is dependent on the model setup and/or representation of physics.

- Models generally acknowledge the role of **diabatic processes** (radiative/surface flux feedbacks)…

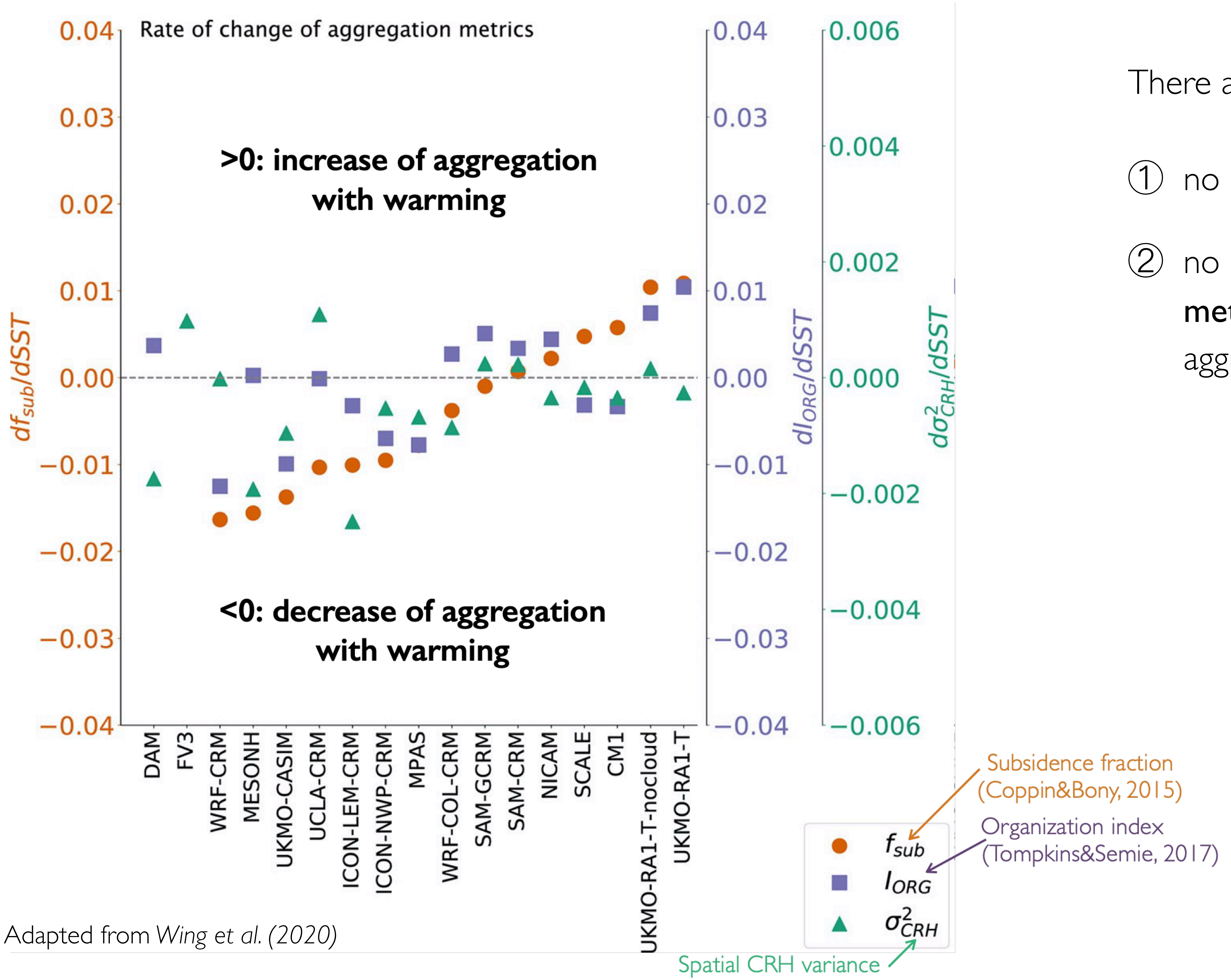
…but show **little consensus** about

- degree and strength of aggregation
- temperature dependence of self-aggregation (found also in snowball-Earth simulations!)
- and sometimes even about whether they undergo aggregation or not at all for a given experimental framework!



From *Muller and Held (2012)*

Models and metrics do not agree



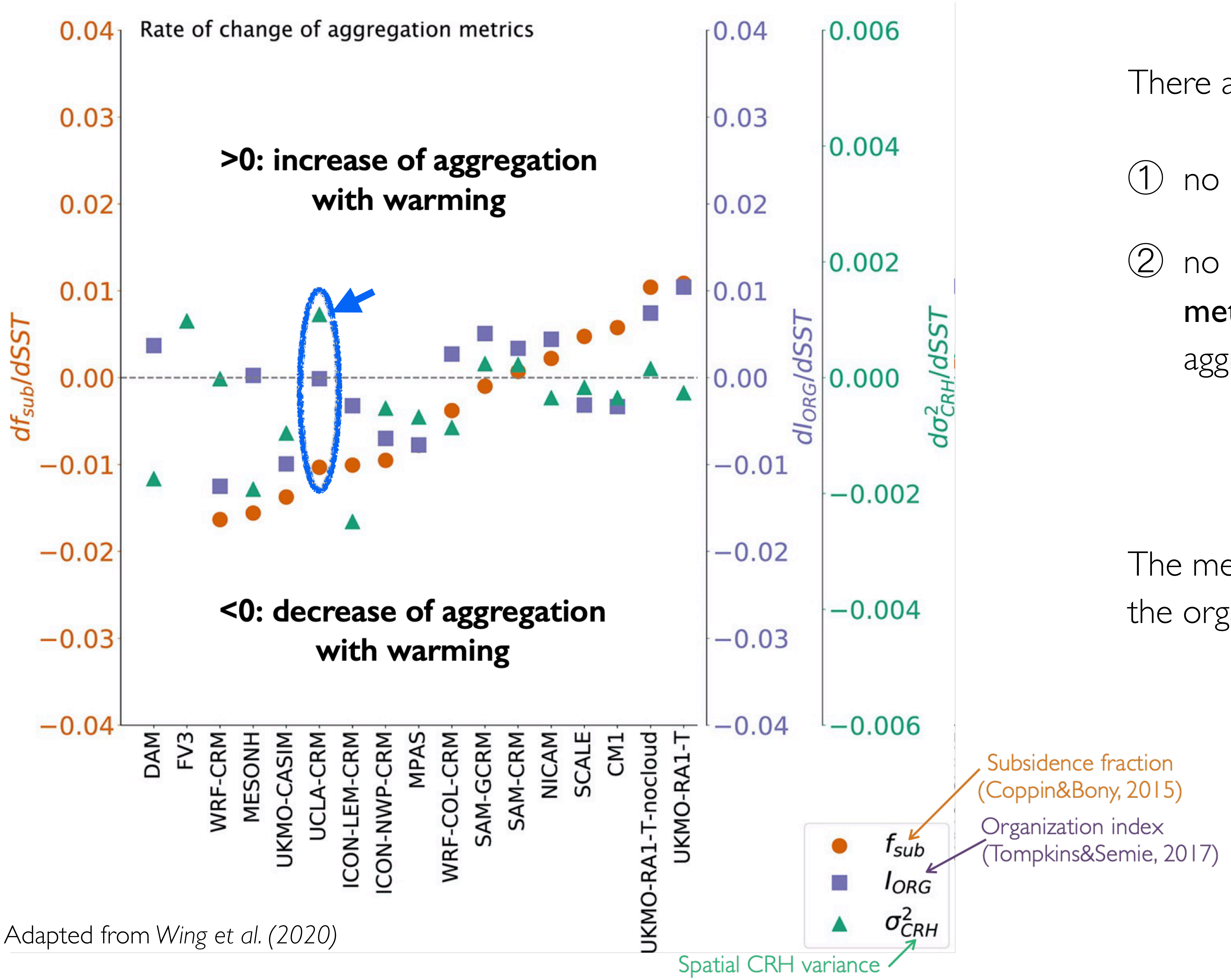
Adapted from *Wing et al. (2020)*

There are **two** sources of uncertainty:

① no consensus among the **models**

② no consensus among the **organization metrics** about the degree/strength of self-aggregation.

Models and metrics do not agree



Adapted from *Wing et al. (2020)*

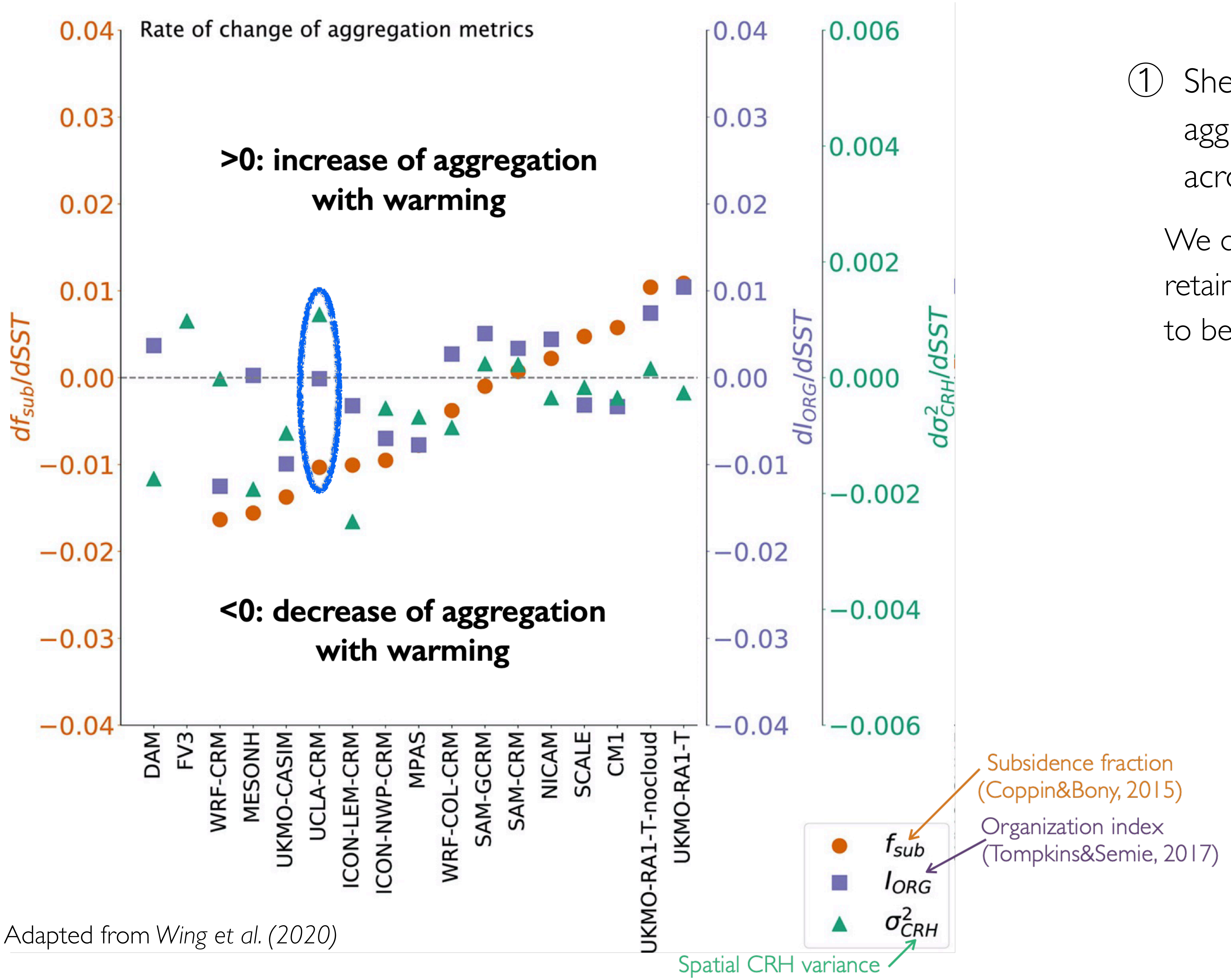
There are **two** sources of uncertainty:

① no consensus among the **models**

② no consensus among the **organization metrics** about the degree/strength of self-aggregation.

The metrics reflect different conceptual views of the organization process.

Our research objectives



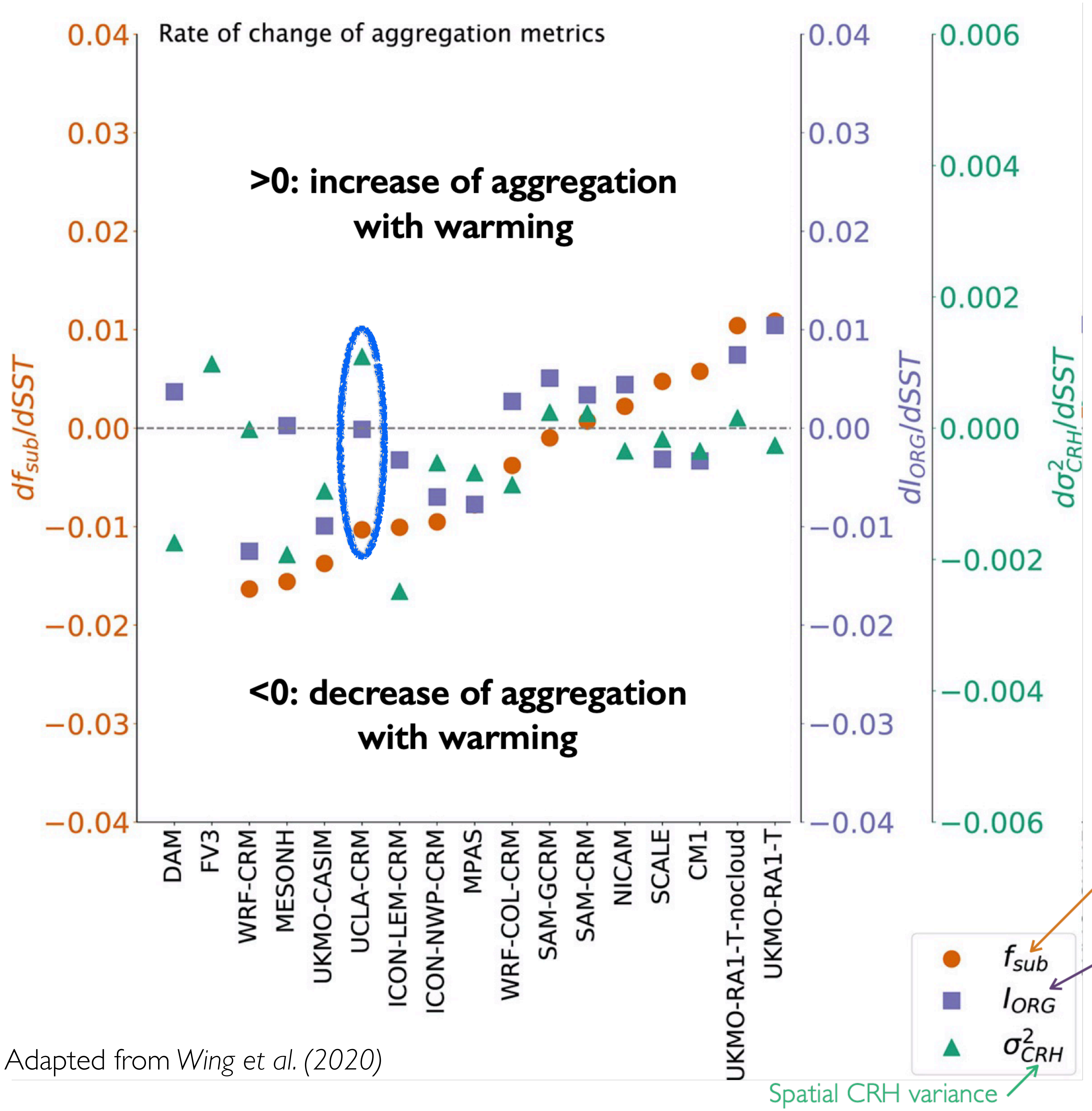
Adapted from *Wing et al. (2020)*

① Shed light on the sensitivities of self-aggregation in CRMs and the differences across the models

We developed a **minimal-physics model** that retains the key aspects of the more complex ones, to be used as a diagnostic tool.

Biagioli and Tompkins, JAMES, 2023

Our research objectives



Adapted from *Wing et al. (2020)*

① Shed light on the sensitivities of self-aggregation in CRMs and the differences across the models

We developed a **minimal-physics model** that retains the key aspects of the more complex ones, to be used as a diagnostic tool.

Biagioli and Tompkins, JAMES, 2023

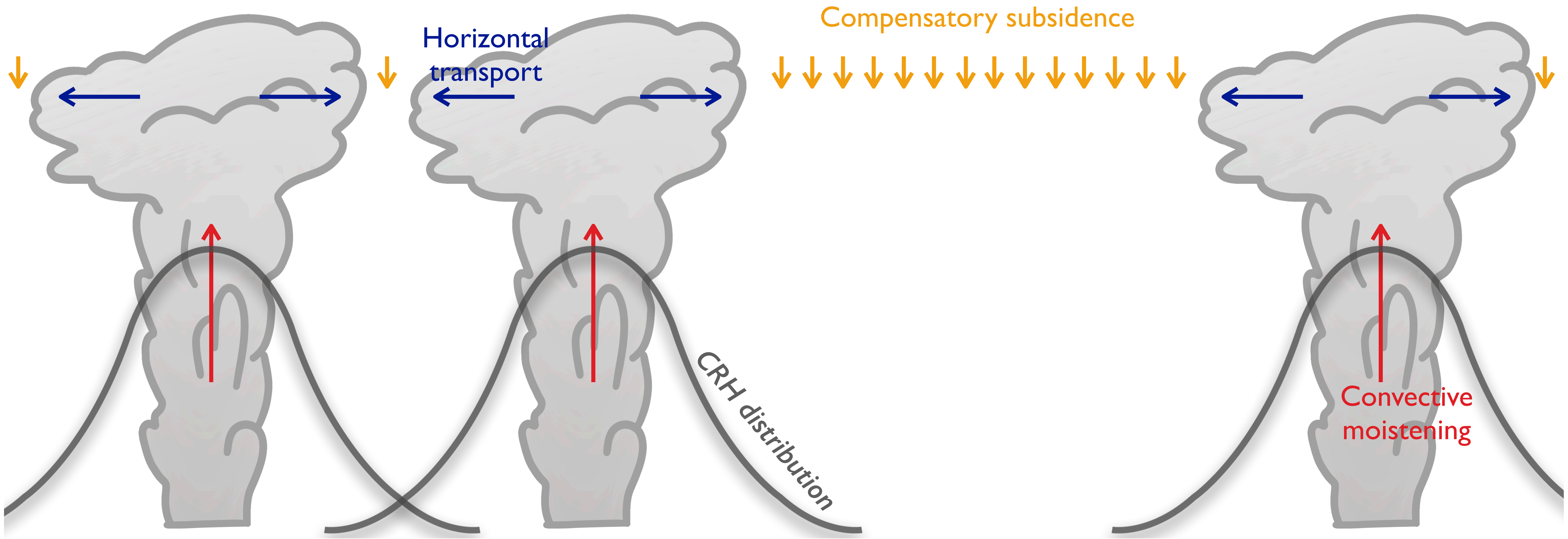
② Better characterize the level of organization in models and observations

We defined a **new organization metric** that amends many drawbacks of a widely used one and captures organization over a range of spatial scales.

Biagioli and Tompkins, JAS, 2023

A 2D stochastic reaction-diffusion model of the tropical atmosphere

- Model closely related to that of Craig and Mack (2013).
- It represents the effects of convective moistening, lateral transport and subsidence drying on the tropical column relative humidity (CRH), $R = R(\boldsymbol{x},t)$, budget.
- Governing equation integrated on a 2D mesh of grid points, using **CRM-like domain sizes** and **resolutions**. Convective activity treated **stochastically**.



LMD

Adding stochastic effects: the selection of convective of cells

① **How many** cells do we select to develop convection?

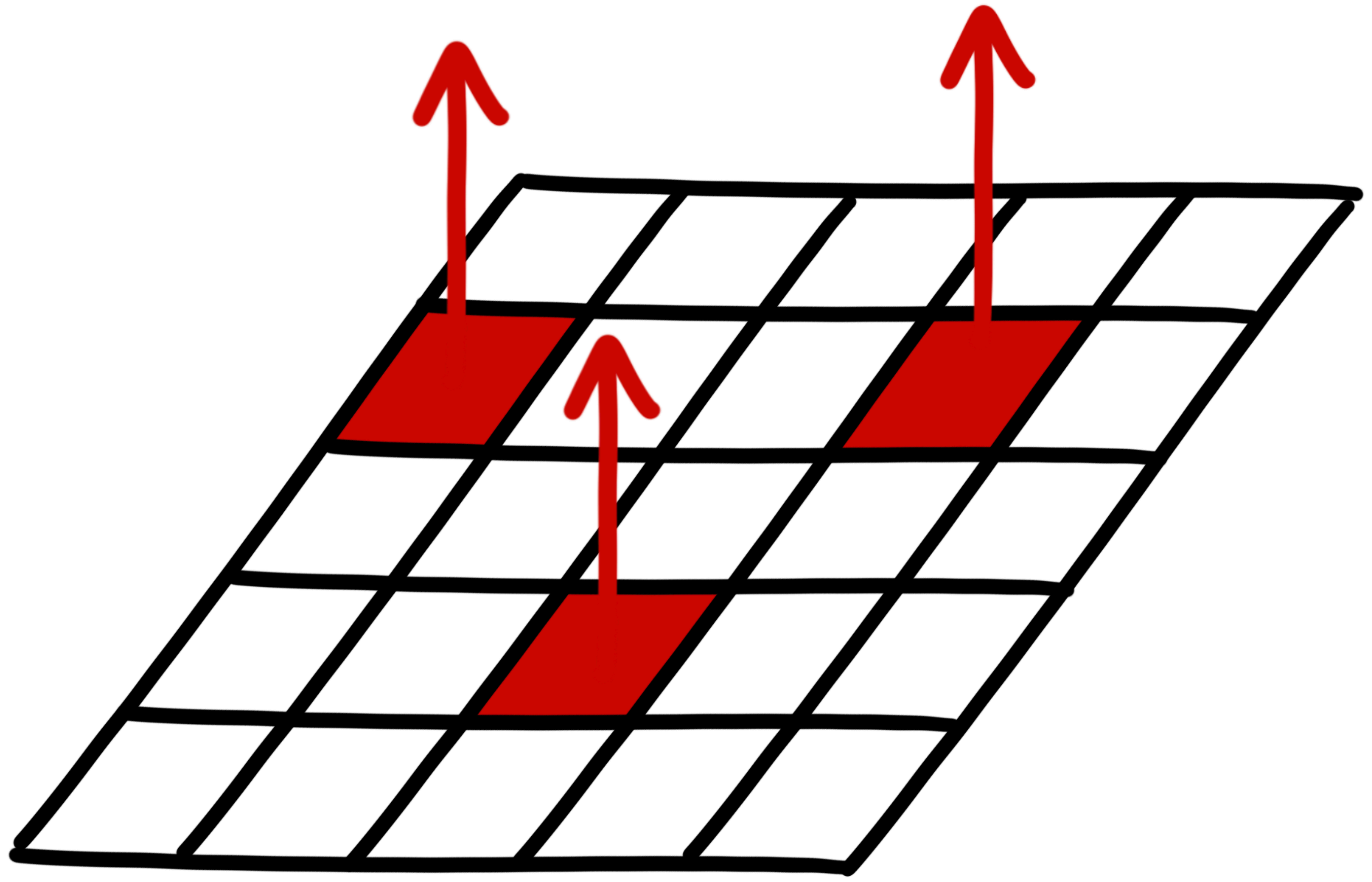
The convective fraction σ*, hence the number* N_c *of convective cells, is such as to obey continuity (externally imposed constraint):*

$$0 = \overline{w} = \sigma w_c + (1 - \sigma) w_{sub} \Rightarrow \sigma = \frac{|w_{sub}|}{w_c}.$$

② **How** do we select the cells that develop convection?

According to weighted random sampling, with humidity-dependent probabilities (based on P-R relationship by Bretherton et al., 2004)

$$p_C(R) = C \exp\left(a_d R\right).$$



The ***parameter*** a_d measures the strength of the convection-water vapor feedback

A 2D stochastic reaction-diffusion model of the tropical atmosphere

- Model closely related to that of Craig and Mack (2013).
- It represents the effects of convective moistening, lateral transport and subsidence drying on the tropical column relative humidity (CRH), $R = R(\boldsymbol{x},t)$, budget.
- Governing equation integrated on a 2D mesh of grid points, using **CRM-like domain sizes** and **resolutions**. Convective activity treated **stochastically**.
- In formulae (continuous form),

CONVECTIVE MOISTENING

$$\frac{\partial R}{\partial t} = \boxed{\frac{1}{\tau_{\mathbf{c}}}(R_{\mathbf{c}} - R)\mathscr{I}_{\mathbf{c}}} + K\,\nabla^2 R - \frac{R}{\tau_{\mathrm{sub}}}$$

new

(fast) convective moistening
characteristic timescale
(constant)

Relaxation target
(constant)

Indicator function
=1 in convective cells
=0 in non-convective cells

LMD

A 2D stochastic reaction-diffusion model of the tropical atmosphere

- Model closely related to that of Craig and Mack (2013).
- It represents the effects of convective moistening, lateral transport and subsidence drying on the tropical column relative humidity (CRH), $R = R(\mathbf{x},t)$, budget.
- Governing equation integrated on a 2D mesh of grid points, using **CRM-like domain sizes** and **resolutions**. Convective activity treated **stochastically**.
- In formulae (continuous form),

$$\frac{\partial R}{\partial t} = \underbrace{\boxed{\frac{1}{\tau_c}(R_c - R)\mathcal{I}_c}}_{\text{CONVECTIVE MOISTENING}} + \underbrace{\boxed{K\,\nabla^2 R}}_{\text{LATERAL TRANSPORT}} - \frac{R}{\tau_{sub}}$$

Horizontal diffusivity
(parameter)

A 2D stochastic reaction-diffusion model of the tropical atmosphere

- Model closely related to that of Craig and Mack (2013).
- It represents the effects of convective moistening, lateral transport and subsidence drying on the tropical column relative humidity (CRH), $R = R(\mathbf{x},t)$, budget.
- Governing equation integrated on a 2D mesh of grid points, using **CRM-like domain sizes** and **resolutions**. Convective activity treated **stochastically**.
- In formulae (continuous form),

$$\frac{\partial R}{\partial t} = \underbrace{\boxed{\frac{1}{\tau_c}(R_c - R)\mathscr{I}_c}}_{\text{CONVECTIVE MOISTENING}} + \underbrace{\boxed{K\nabla^2 R}}_{\text{LATERAL TRANSPORT}} \underbrace{\boxed{-\frac{R}{\tau_{sub}}}}_{\text{SUBSIDENCE}}$$

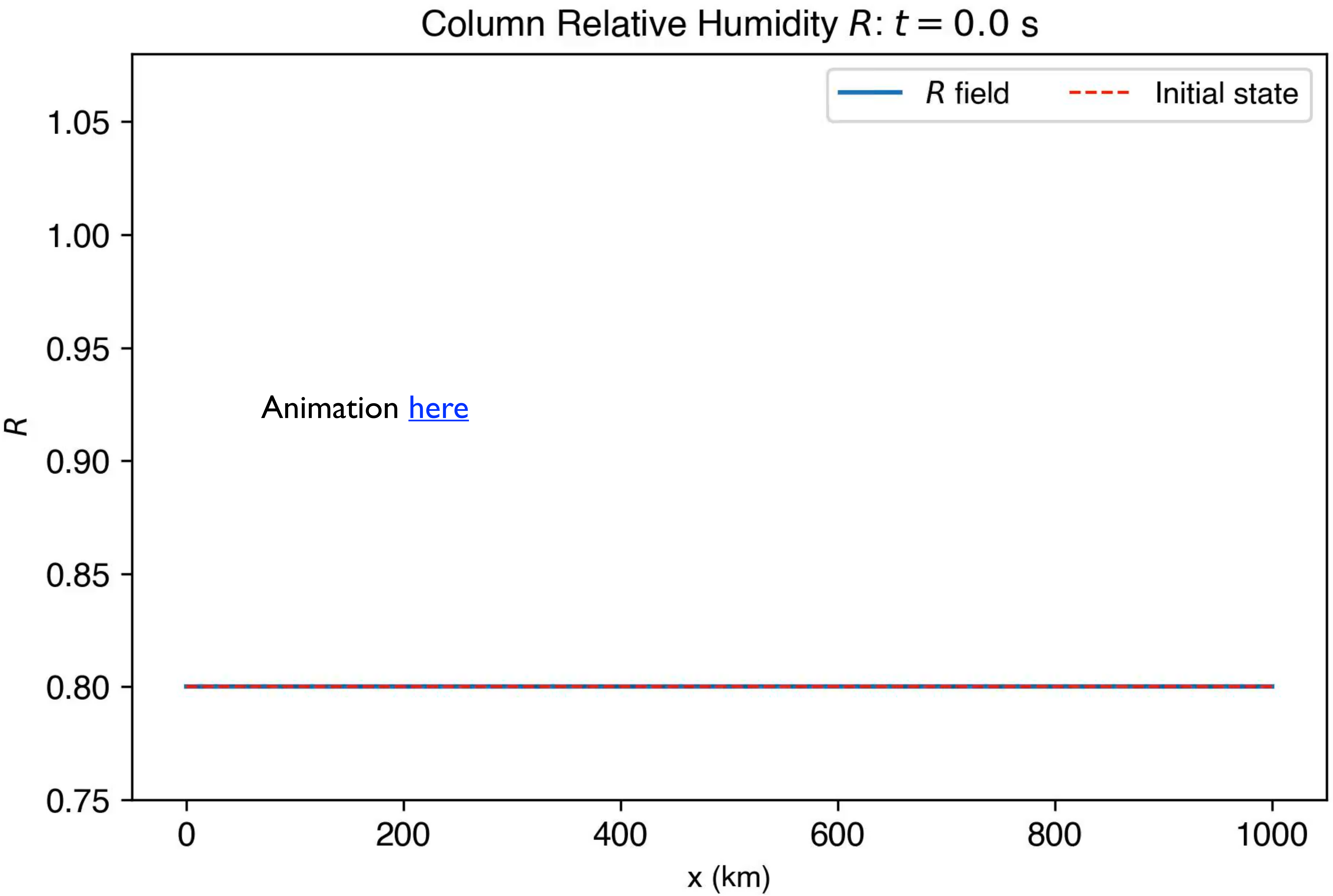
(slow) drying timescale

(parameter)

$\tau_{sub} \gg \tau_c$

A closer look at the model's physics

$K = 10000\ \mathrm{m^2 s^{-1}}$, $\tau_{sub} = 15$ days, $a_d = 14.72$, $\Delta x = 1$ km, $\Delta t = 12.5$ s



- Convective spikes
- Subsidence in the far-field
- Convection more easily develops in moister-than-average areas, as per the functional form of $p_C(R)$
- Subsidence and diffusion create moist halos around convection: **area of influence** $K\tau_{sub}$ (units m^2)

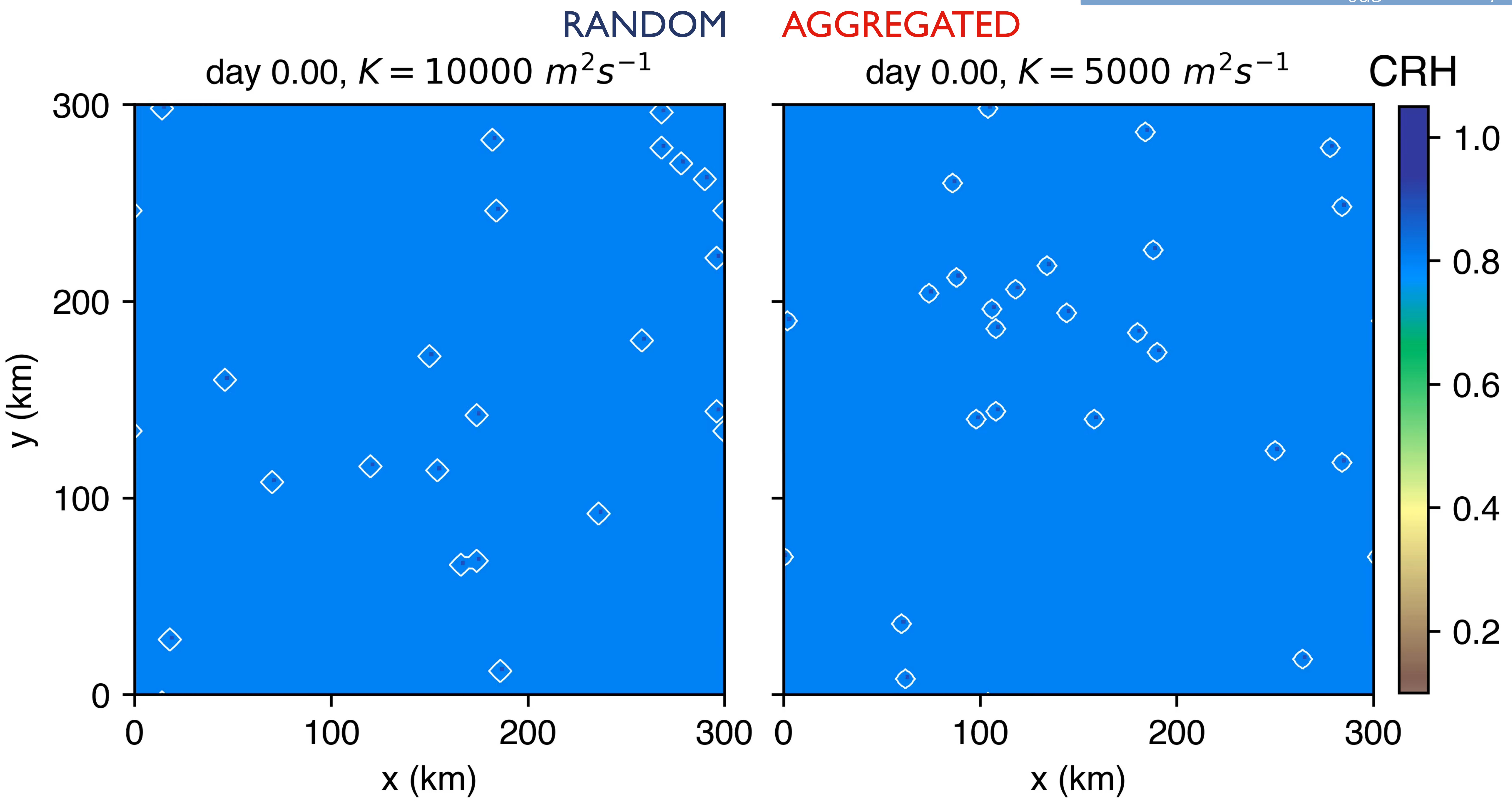
The model mimics convective clustering

Depending on the parameter settings, the model produces random/aggregated states, **similar to those seen in CRMs**.

DEFAULT EXPERIMENTAL SETUP

- Doubly periodic domain with size $L = 300$ km, spacing $\Delta x = 2$ km. Horizontally homogeneous initialization, $R_0 = 0.8$.
- $K = \mathcal{O}(10^4)$ m²s⁻¹ $\tau_{\mathrm{sub}} \sim 16$ days $a_d = 14.72$ (TRMM v7)

Animation [here]



The model mimics convective clustering and reproduces many aspects of CRMs

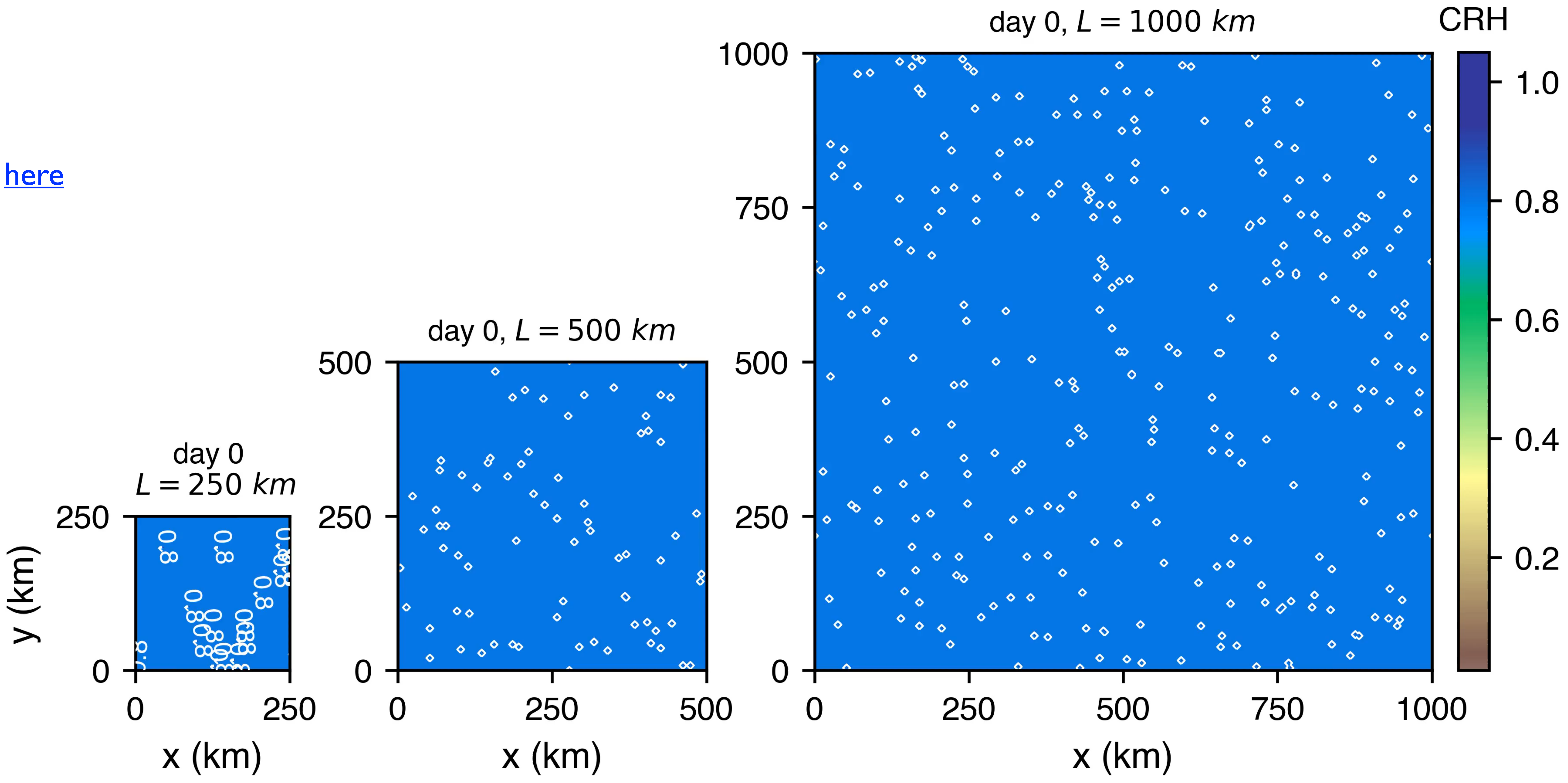
- Perturbing the model key parameters can trigger aggregation.

  Aggregation favored by - weaker diffusion (lower K)

  - stronger subsidence (shorter τ_{sub})
  - stronger convection-vapor feedback (larger a_d)
  - larger domains (larger L), **as in CRMs**

The model mimics convective clustering and reproduces many aspects of CRMs

Animation [here]



INCREASING DOMAIN SIZE

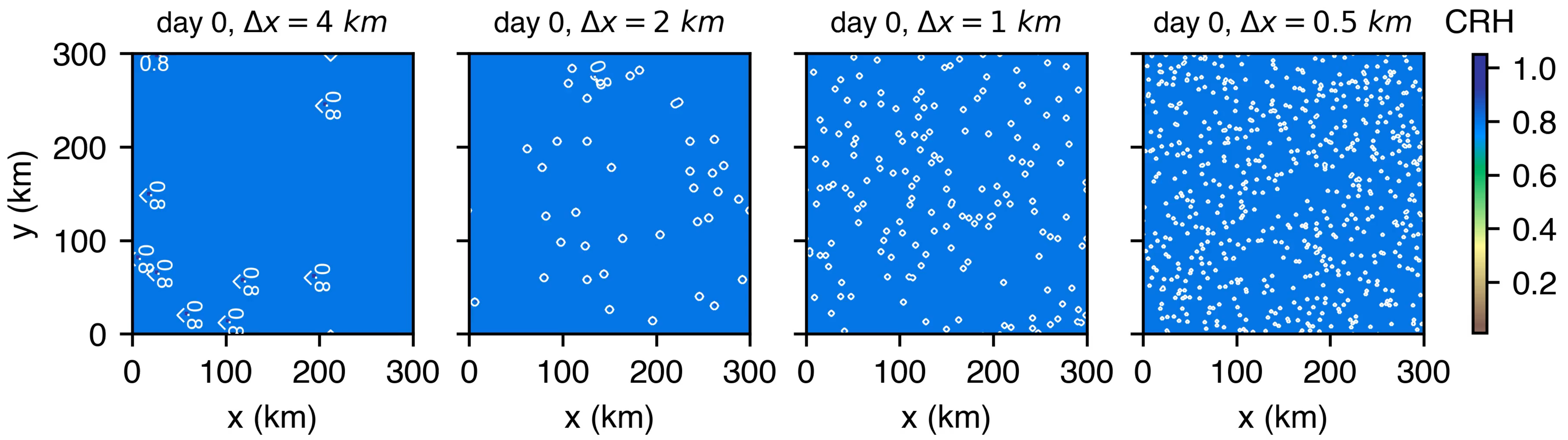
AGGREGATION MORE LIKELY

The model mimics convective clustering and reproduces many aspects of CRMs

- Perturbing the model key parameters can trigger aggregation.

  Aggregation favored by - weaker diffusion (lower K)

  - stronger subsidence (shorter τ_{sub})
  - stronger convection-vapor feedback (larger a_d)
  - larger domains (larger L), **as in CRMs**
  - coarser resolutions (larger Δx), **as in CRMs**



Animation [here]

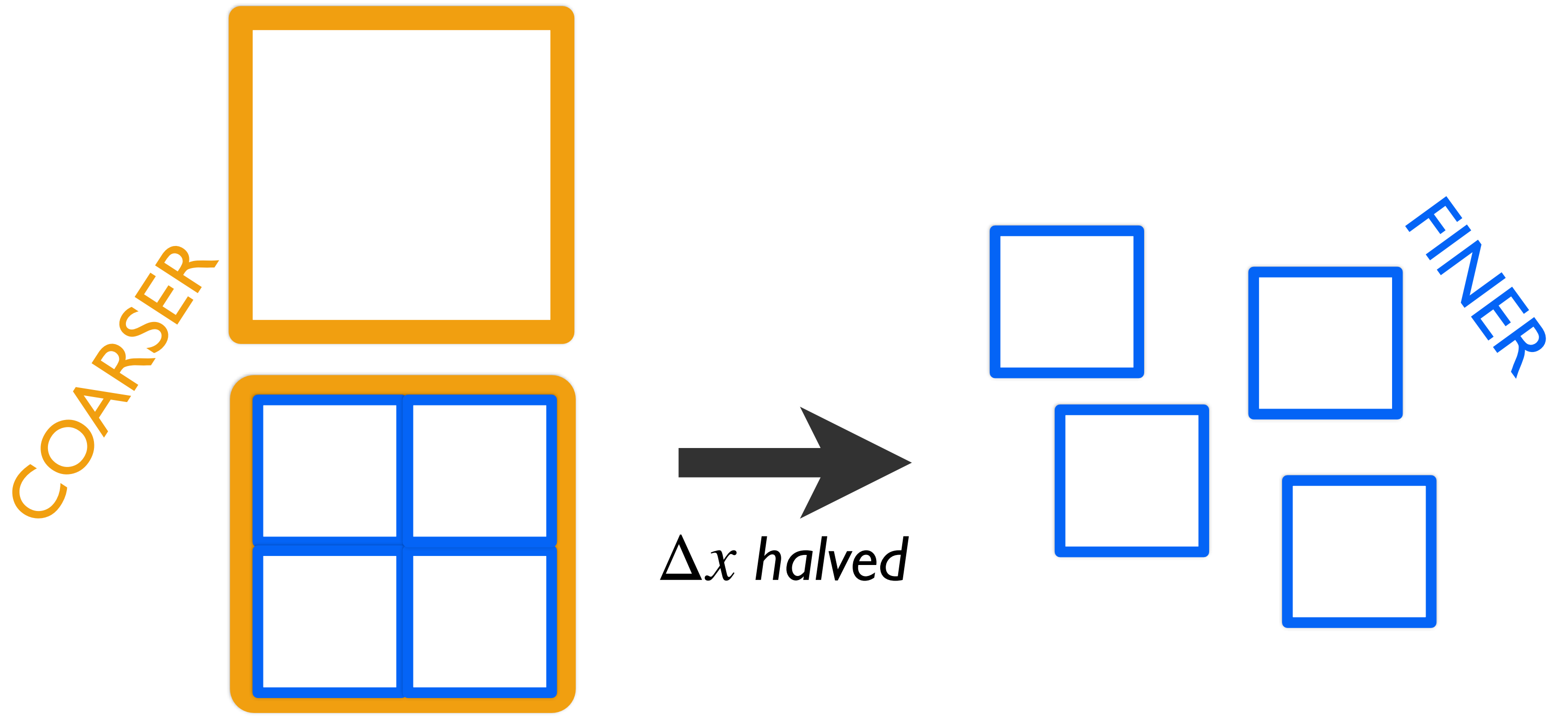
INCREASING RESOLUTION

AGGREGATION LESS LIKELY

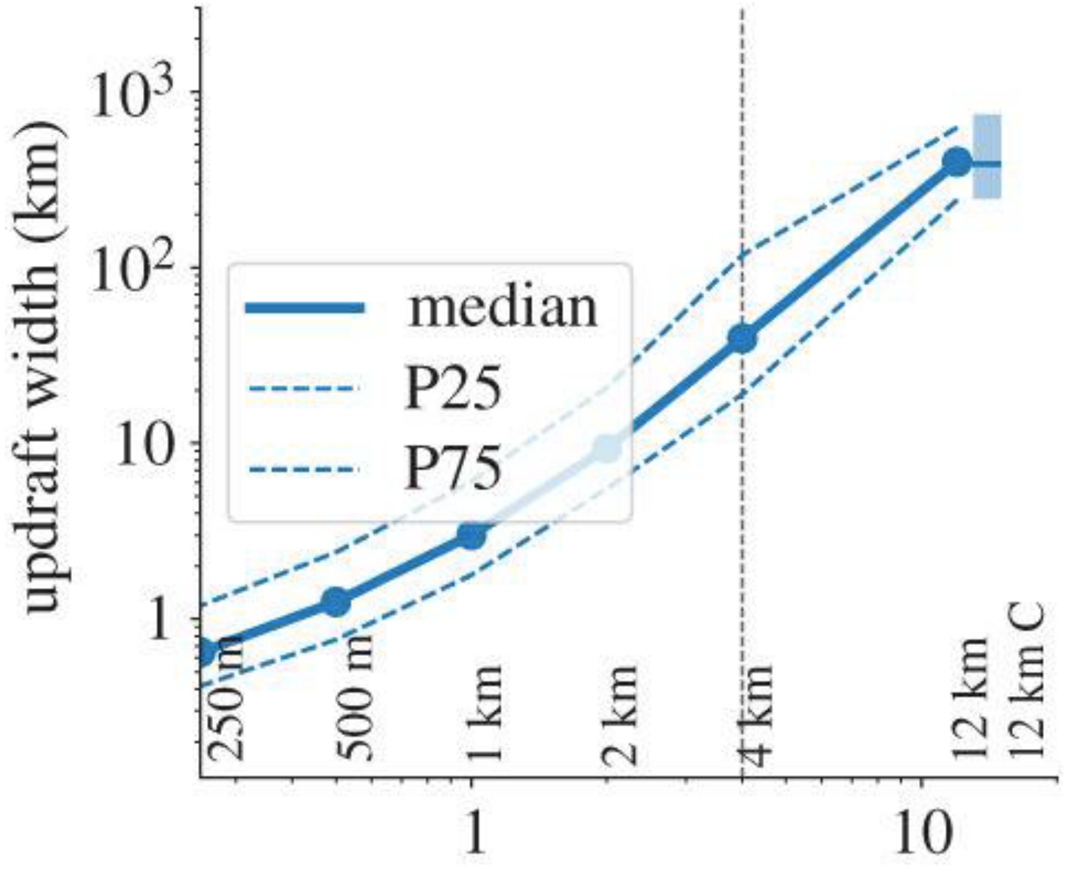
A new explanation for resolution sensitivity of self-aggregation in CRMs

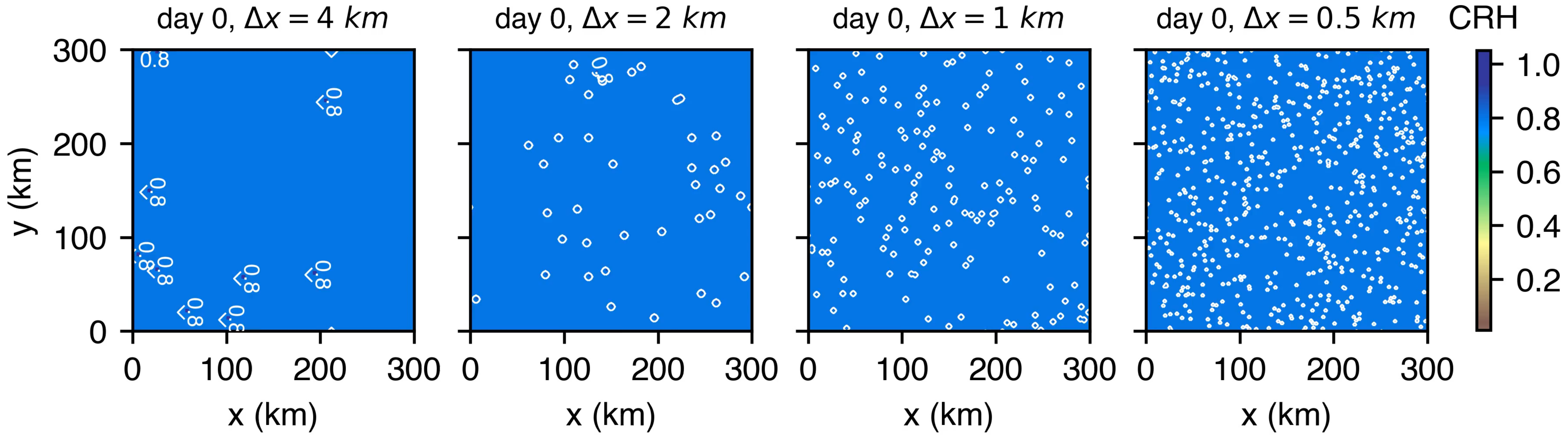
The mass conservation argument constrains the convective fraction σ, not number/size of convective cells.

Higher resolutions ➔ more (smaller) updrafts ➔ smaller inter-convective spacings ➔ aggregation inhibited.



Our hypothesized mechanism is consistent with Sueki et al. (2019), *Prein et al. (2021)*: the updraft width depends on the resolution.





Animation [here]

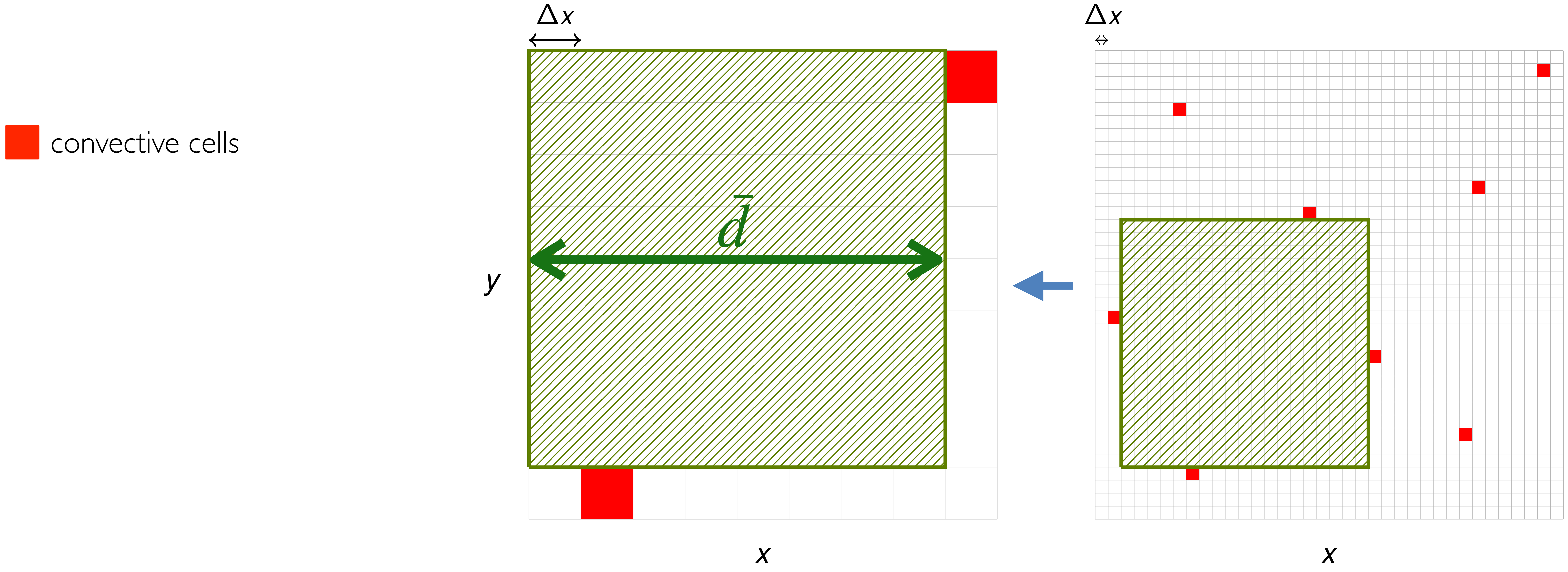
INCREASING RESOLUTION

AGGREGATION LESS LIKELY

The largest clear-sky patch in the pre-organization phase

A key quantity is the *expected* size $\bar{d}$ of the maximum convection-free area in the pre-onset (random) phase.

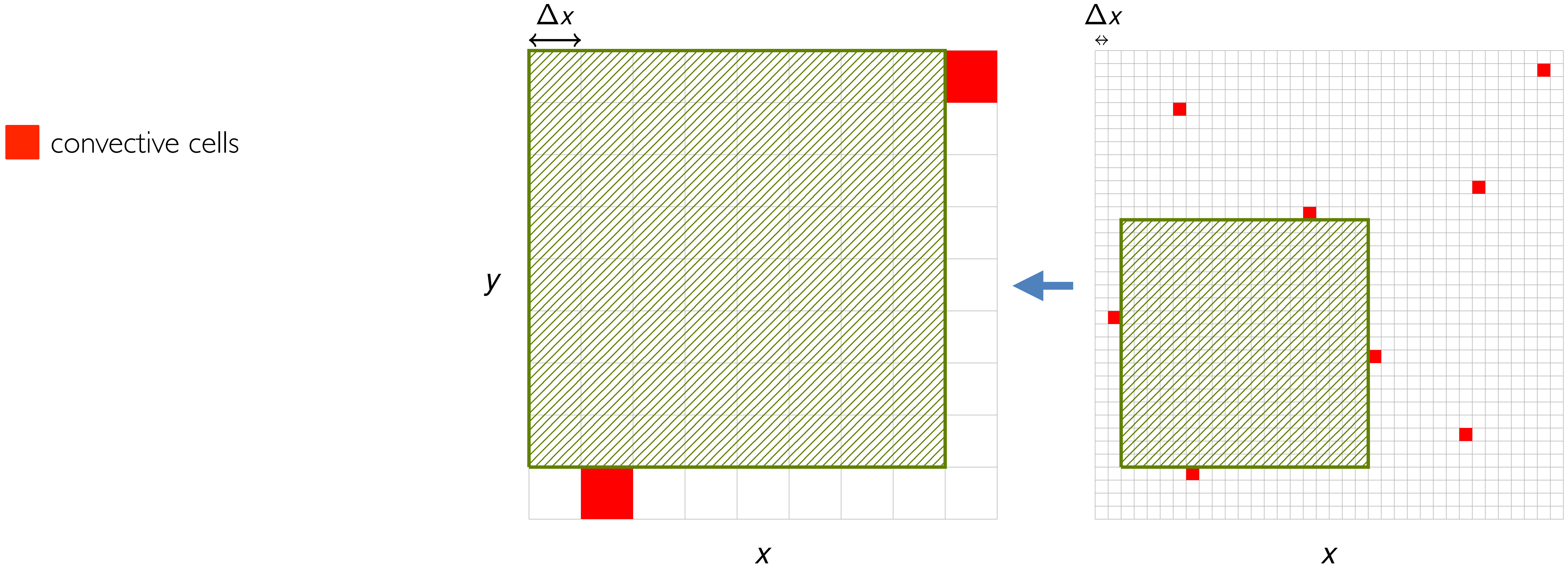
Few updrafts ➔ large convection-free areas ➔ subsidence dominates in the far-field ➔ convection-free areas further dry ➔ aggregation **more** likely.



The largest clear-sky patch in the pre-organization phase

A key quantity is the *expected* size $\bar{d}$ of the maximum convection-free area in the pre-onset (random) phase.

Few updrafts ➔ large convection-free areas ➔ subsidence dominates in the far-field ➔ convection-free areas further dry ➔ aggregation **more** likely.



Given N_c, L, Δx, the quantity $\bar{d}$ can be analytically calculated.

Its relevance to self-aggregation onset in CRMs has been confirmed by Casallas et al. (2025).

Key factors driving the transition to aggregated convection

A key quantity is the *expected* size $\bar{d}$ of the maximum convection-free area in the pre-onset (random) phase.

Few updrafts ➔ large convection-free areas ➔ subsidence dominates in the far-field ➔ convection-free areas further dry ➔ aggregation **more** likely.

The other key ingredient is the area of influence $K\tau_{\text{sub}}$ on the moisture field of a single deep convective event.

Large $K\tau_{\text{sub}}$ ➔ efficient environmental moistening/weak drying ➔ humidity halos enlarged ➔ aggregation **less** likely.

A dimensionless parameter to predict the development of self-aggregation

A key quantity is the *expected* size $\bar{d}$ of the maximum convection-free area in the pre-onset (random) phase.

Few updrafts ➔ large convection-free areas ➔ subsidence dominates in the far-field ➔ convection-free areas further dry ➔ aggregation **more** likely.

The other key ingredient is the area of influence $K\tau_{\text{sub}}$ on the moisture field of a single deep convective event.

Large $K\tau_{\text{sub}}$ ➔ efficient environmental moistening/weak drying ➔ humidity halos enlarged ➔ aggregation **less** likely.

Putting pieces together, we introduce a dimensionless parameter, the **aggregation number** N_{ag}

$$N_{\text{ag}} = \frac{K\tau_{\text{sub}}}{L^2}\frac{1}{\frac{\bar{d}}{L}}\frac{1}{a_d^2} \Rightarrow$$

Rescaled area of influence

Rescaled size of max clear-sky patch

Assessed through fits from model data

$$\boxed{N_{\text{ag}} = \frac{K\tau_{\text{sub}}}{a_d^2 L\bar{d}}}$$

LMD

A dimensionless parameter to predict the development of self-aggregation

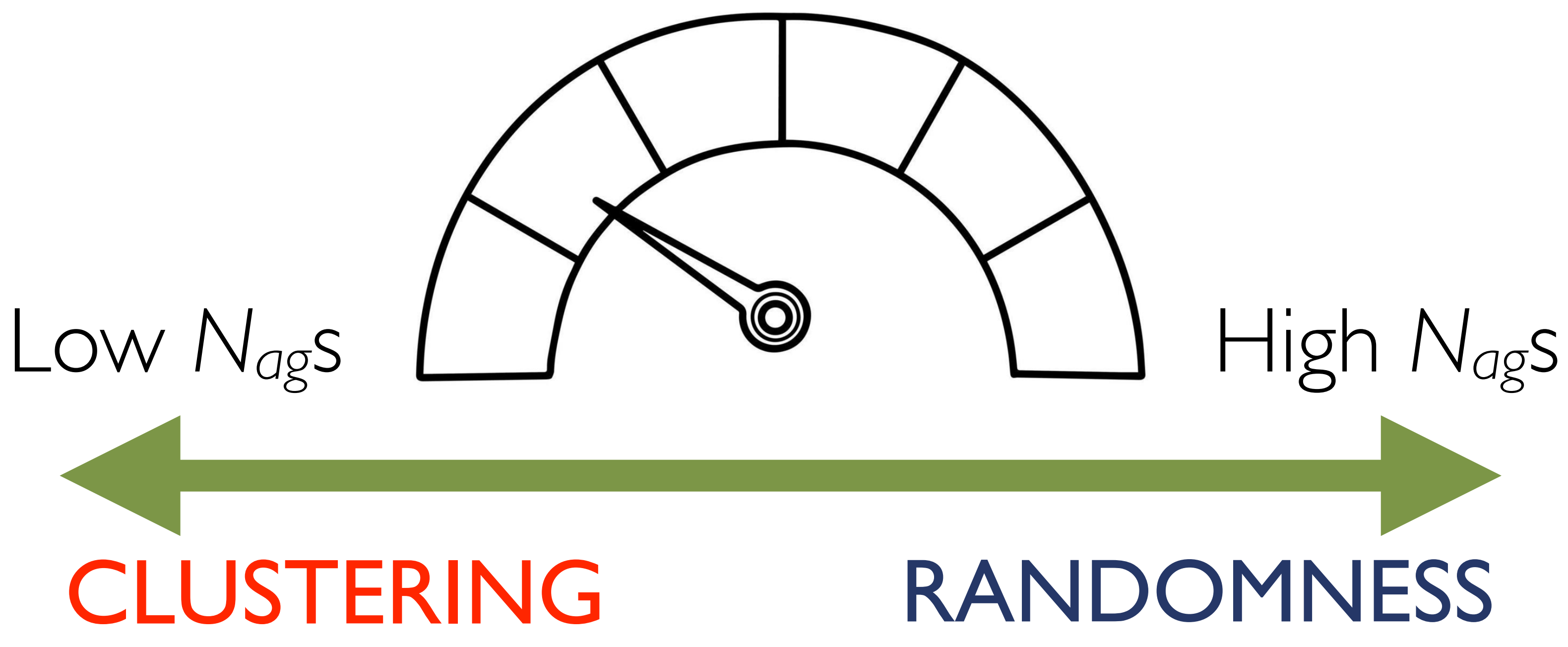
A key quantity is the *expected* size $\bar{d}$ of the maximum convection-free area in the pre-onset (random) phase.

Few updrafts ➔ large convection-free areas ➔ subsidence dominates in the far-field ➔ convection-free areas further dry ➔ aggregation **more** likely.

The other key ingredient is the area of influence $K\tau_{\text{sub}}$ on the moisture field of a single deep convective event.

Large $K\tau_{\text{sub}}$ ➔ efficient environmental moistening/weak drying ➔ humidity halos enlarged ➔ aggregation **less** likely.

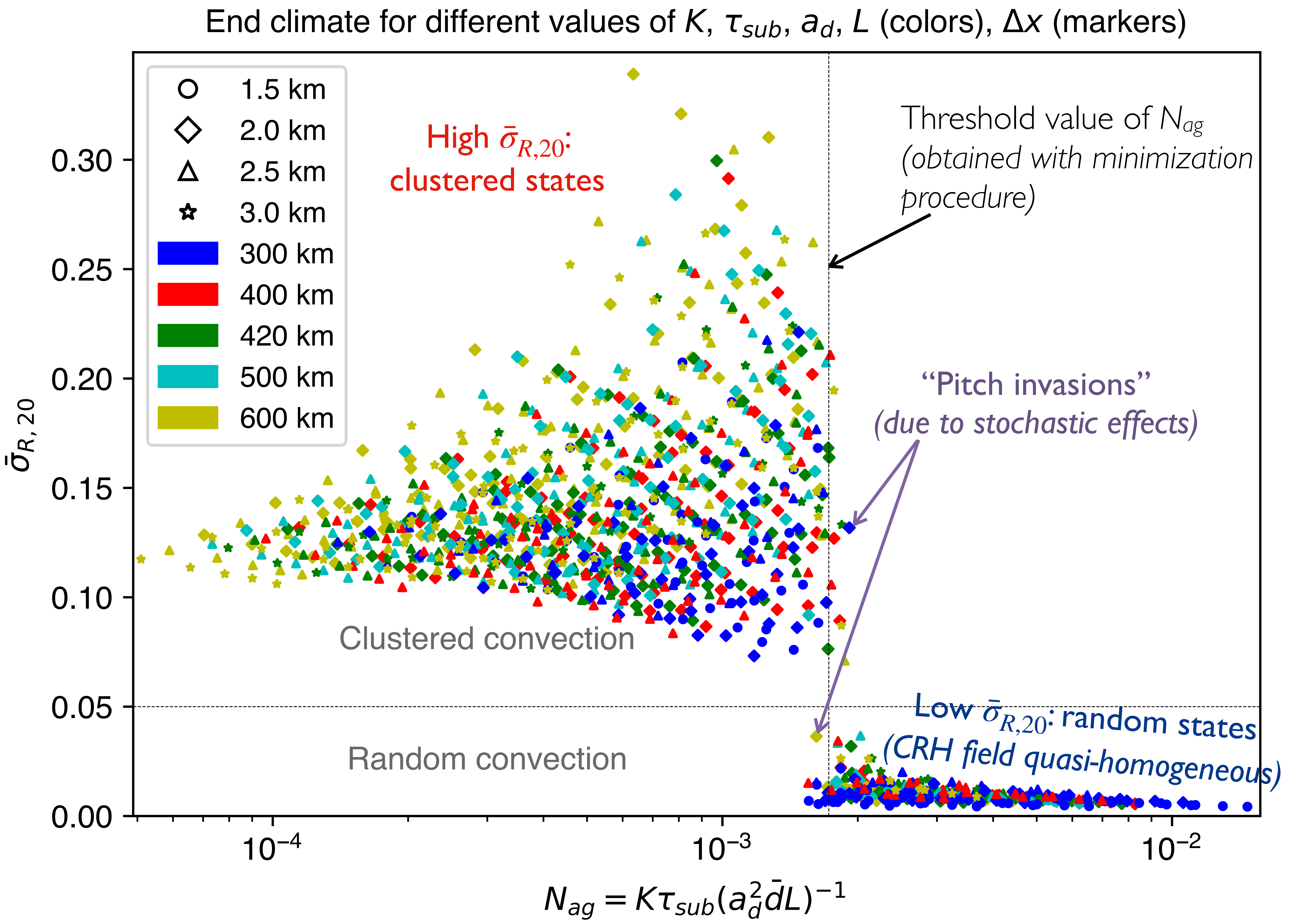
Putting pieces together, we introduce a dimensionless parameter, the **aggregation number** N_{ag}



$$\boxed{N_{\text{ag}} = \frac{K\tau_{\text{sub}}}{a_d^2 L \bar{d}}}$$

LMD

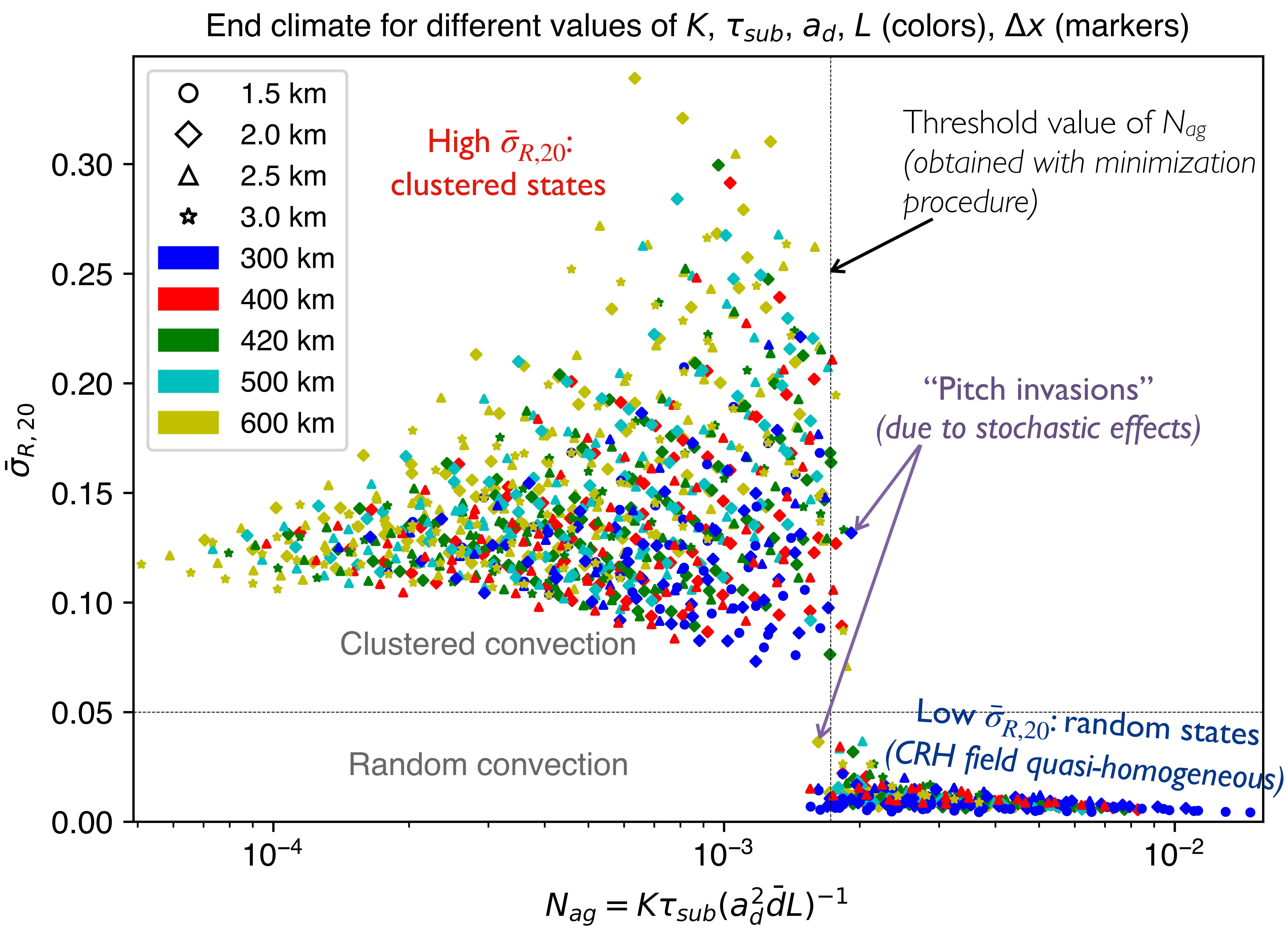
N_{ag} has a predictive power



Ensemble of $\mathcal{O}(1000)$ simulations.

Each symbol represents an experiment with its own N_{ag} vs $\bar{\sigma}_{R,20}$, spatial CRH standard deviation averaged over last 20 (out of 180) days of simulation.

N_{ag} has a predictive power



Ensemble of $\mathcal{O}(1000)$ simulations.

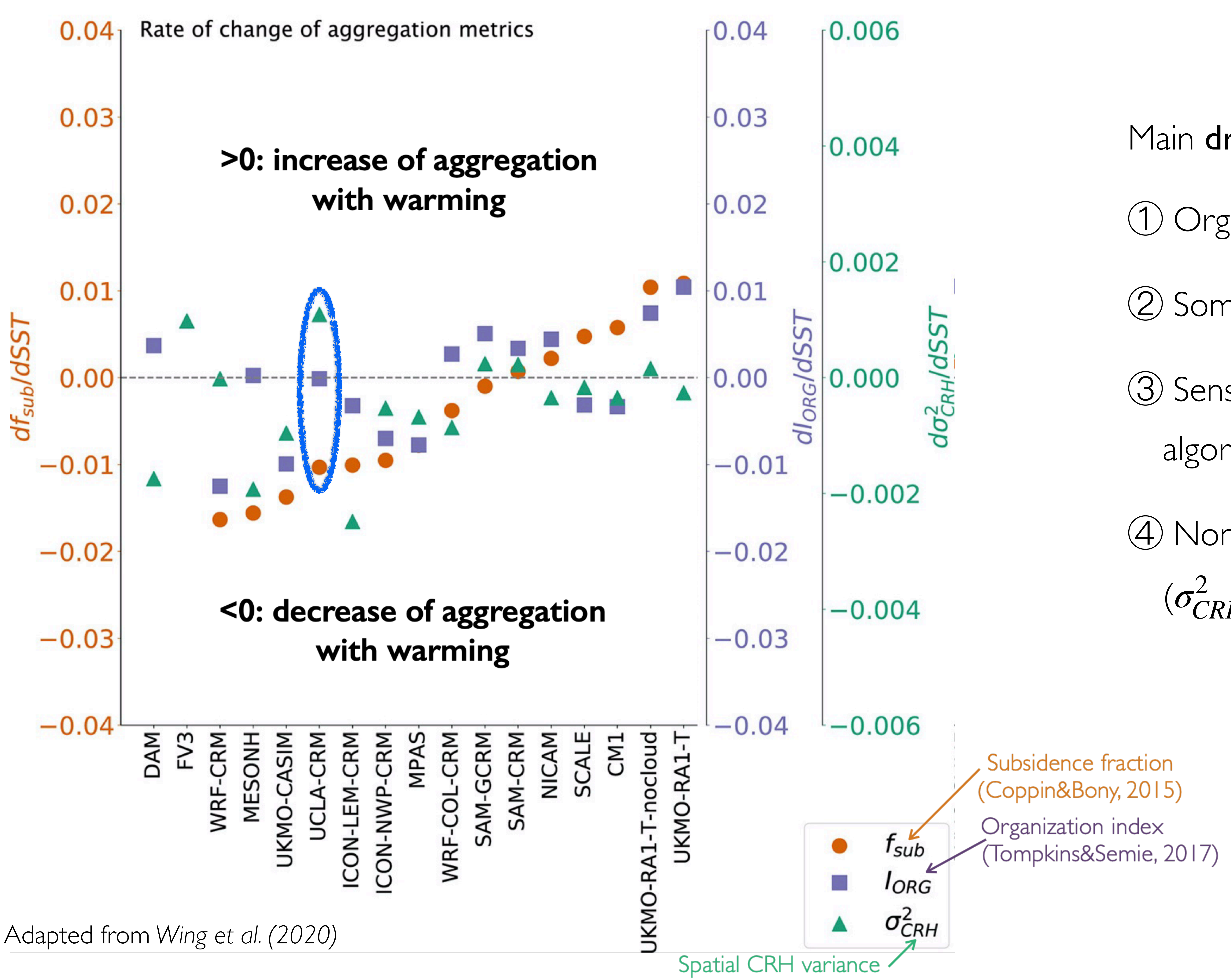
Each symbol represents an experiment with its own N_{ag} vs $\bar{\sigma}_{R,20}$, spatial CRH standard deviation averaged over last 20 (out of 180) days of simulation.

***N_{ag} robustly indicates which model and experiment setups result in aggregation** (i.e., those for which $N_{ag} < N_{ag,crit} \sim 1.72 \times 10^{-3}$).*

If we can project the initial random state fields and diagnose K, τ_{sub} and a_d from CRMs, N_{ag} should predict if a specific run is expected to cluster or not.

LMD

Back to organization indices: some room for improvement?



Adapted from *Wing et al. (2020)*

Main **drawbacks** of the existing metrics:

① Organization measured in a <u>relative</u> sense.

② Some spatial scales are "favored" (e.g., I_{org}).

③ Sensitivity to details of calculation algorithm.

④ Non-linear assessment of organization (σ^2_{CRH}).

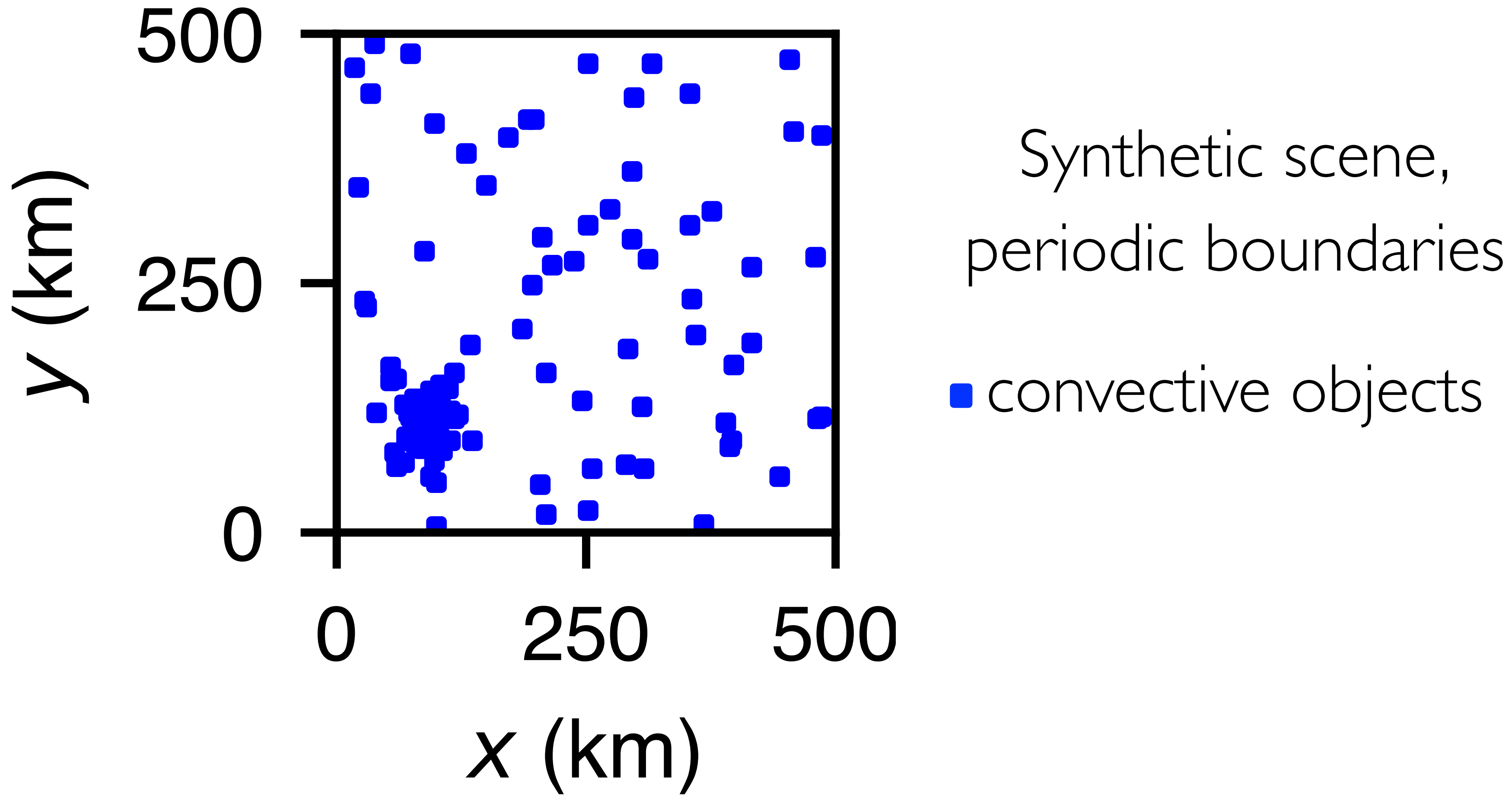
We conducted a systematic review of existing indices to date, see *Biagioli and Tompkins, JAS, 2023*.

See also *Mandorli and Stubenrauch, GMD, 2024*, for an assessment of metrics

LMD

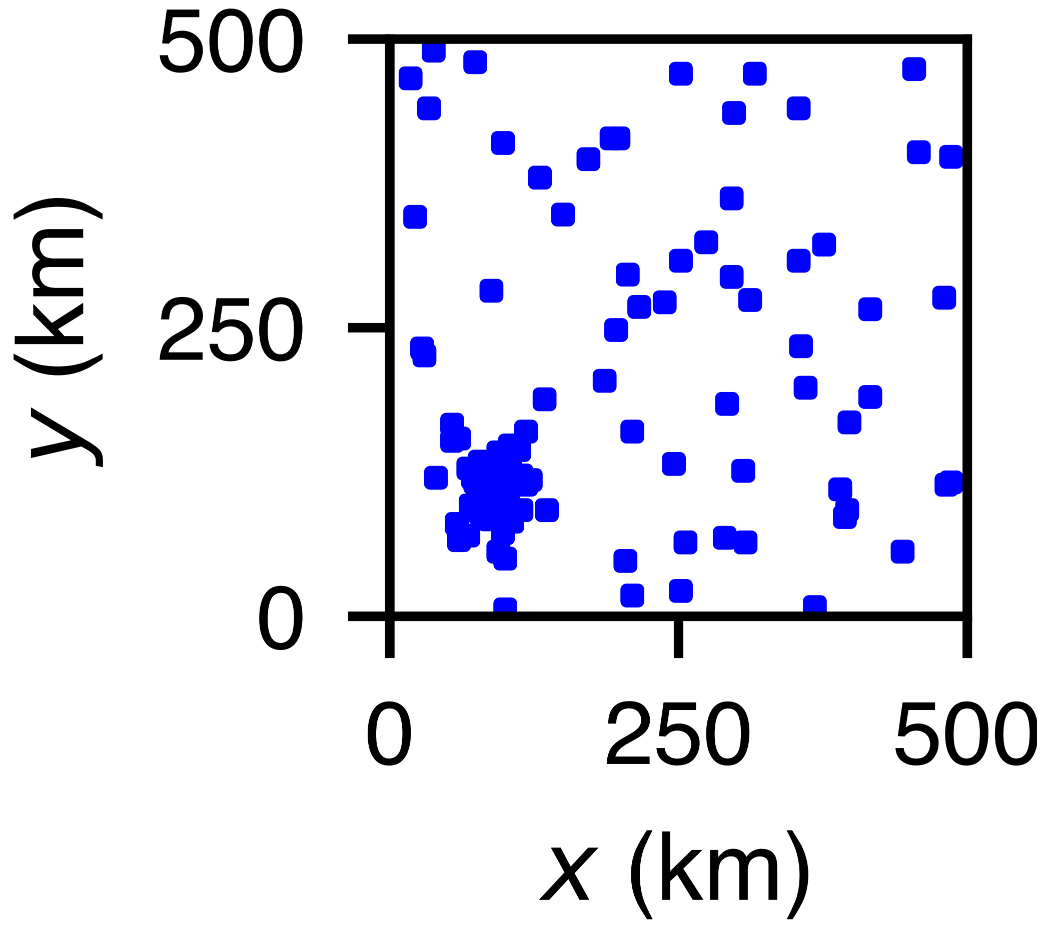
A focus on I_{org}

Given a cloud field scene,



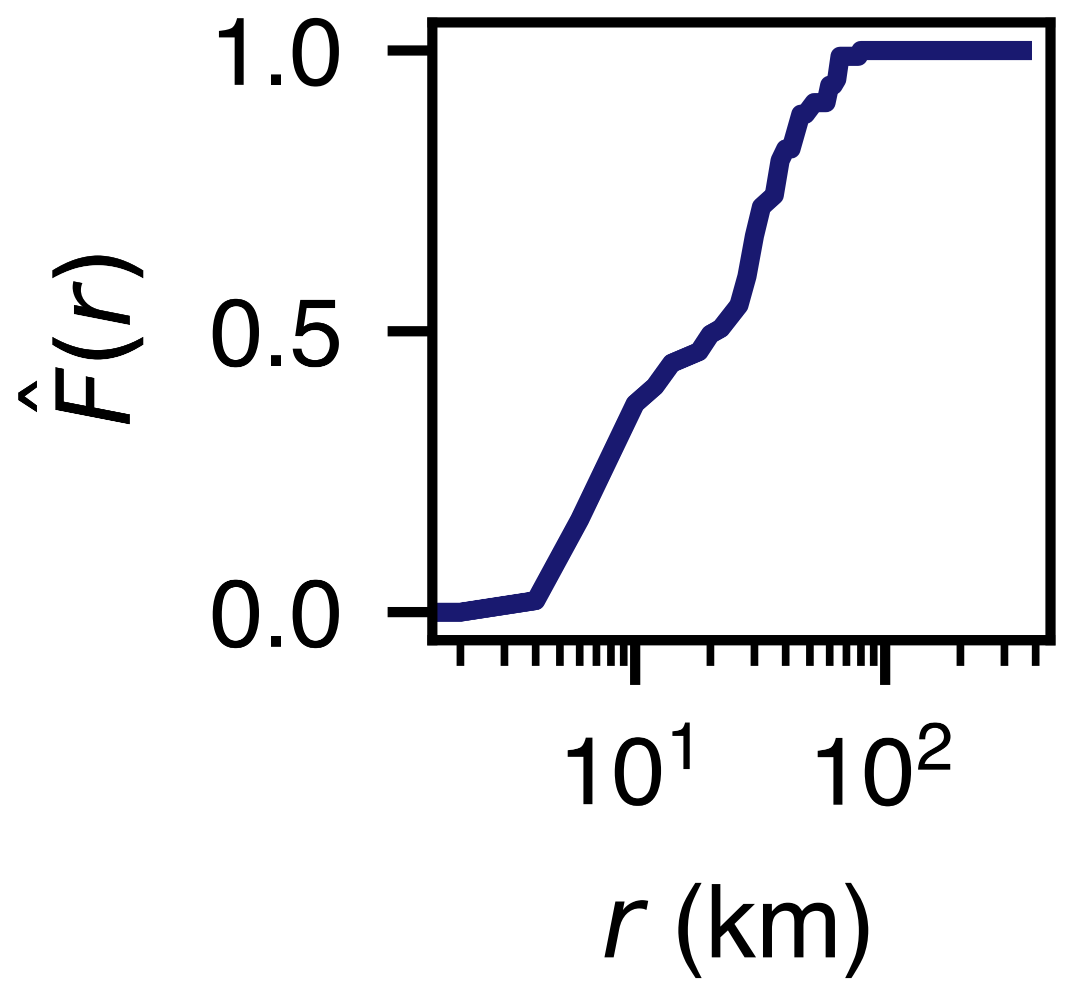
A focus on I_{org}

Given a cloud field scene,



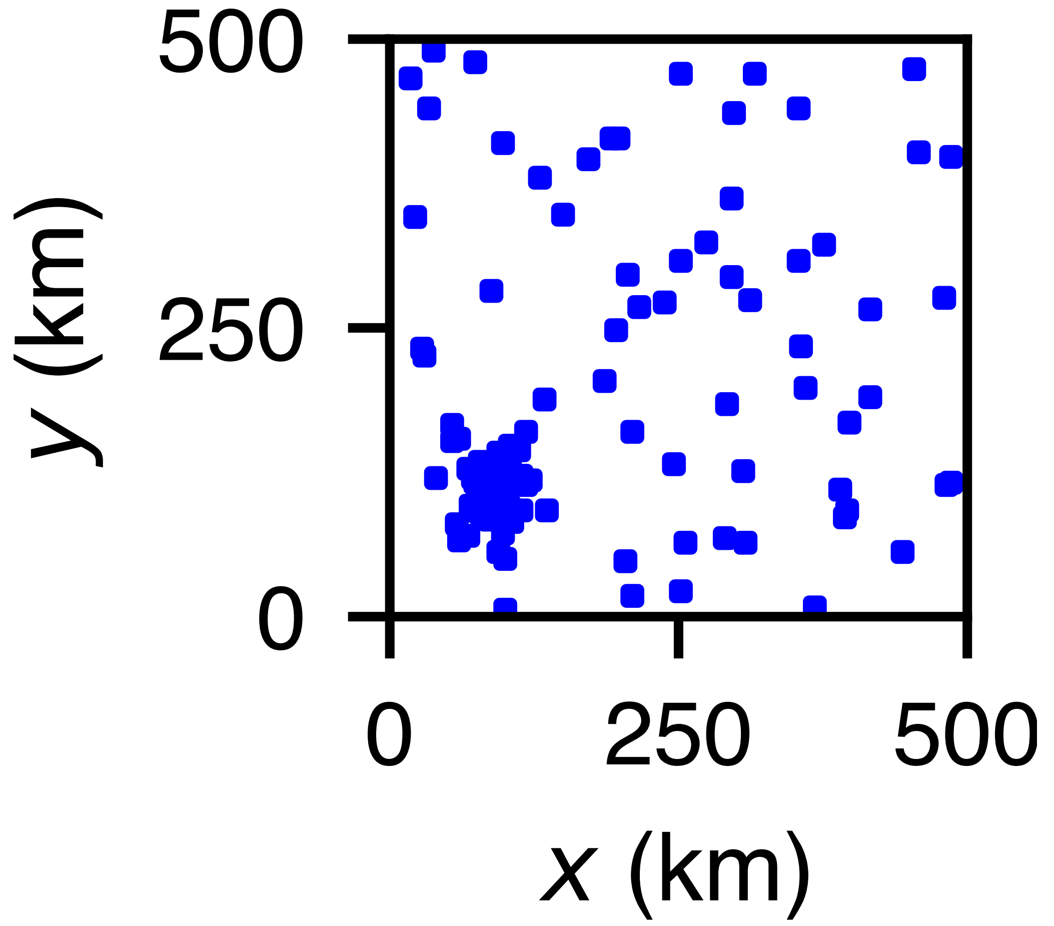
we compute two cumulative distribution functions (CDFs)…

① The CDF $\hat{F}$ of the nearest-neighbor distances (NNDs) in *that* scene,



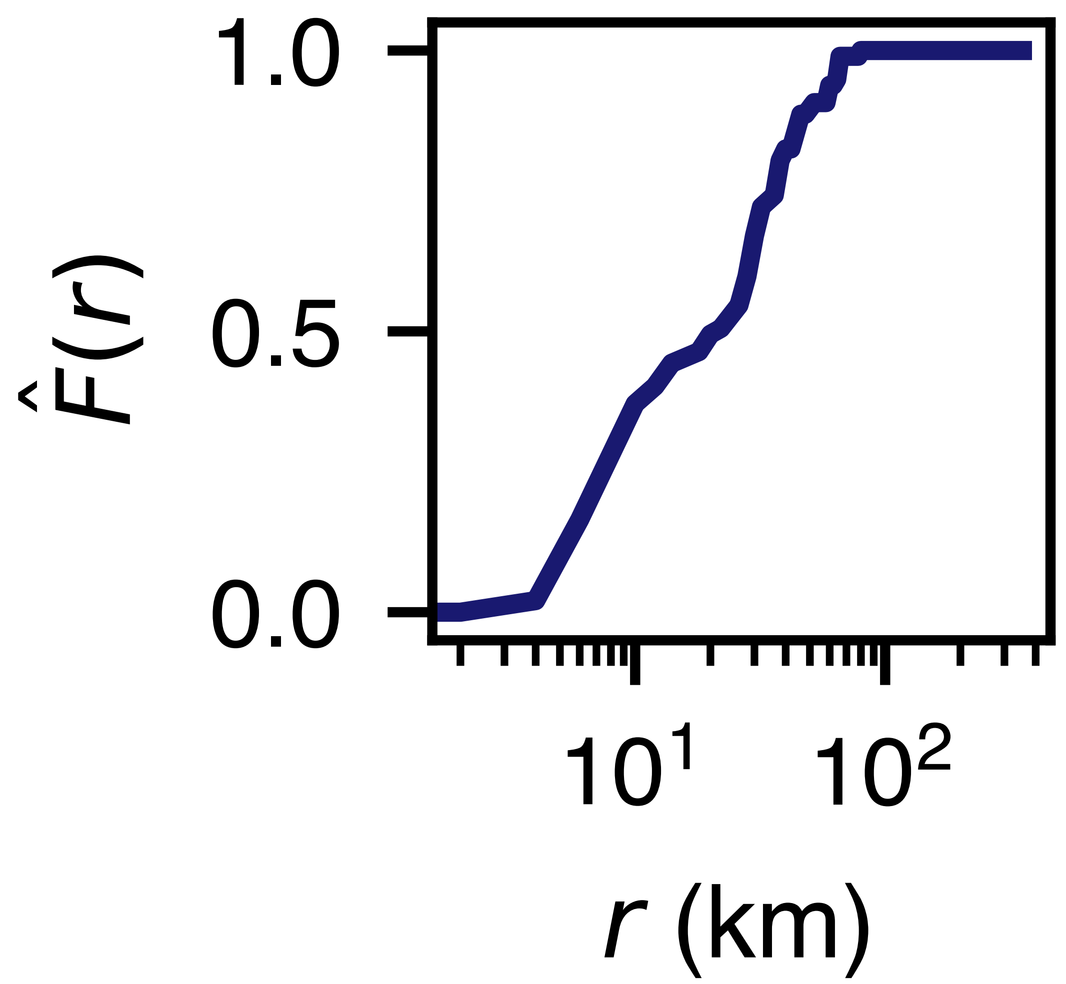
A focus on I_{org}

Given a cloud field scene,

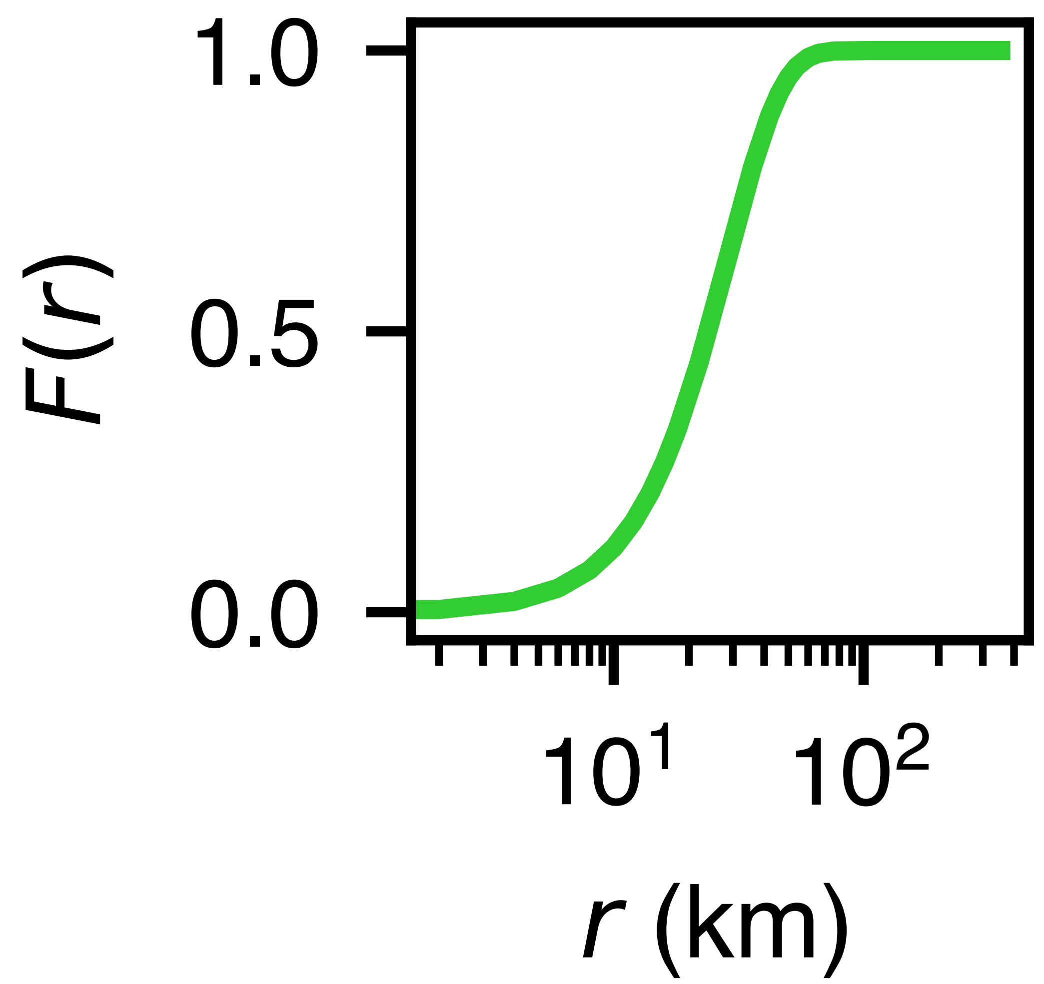


we compute two cumulative distribution functions (CDFs)…

① The CDF $\hat{F}$ of the nearest-neighbor distances (NNDs) in *that* scene,

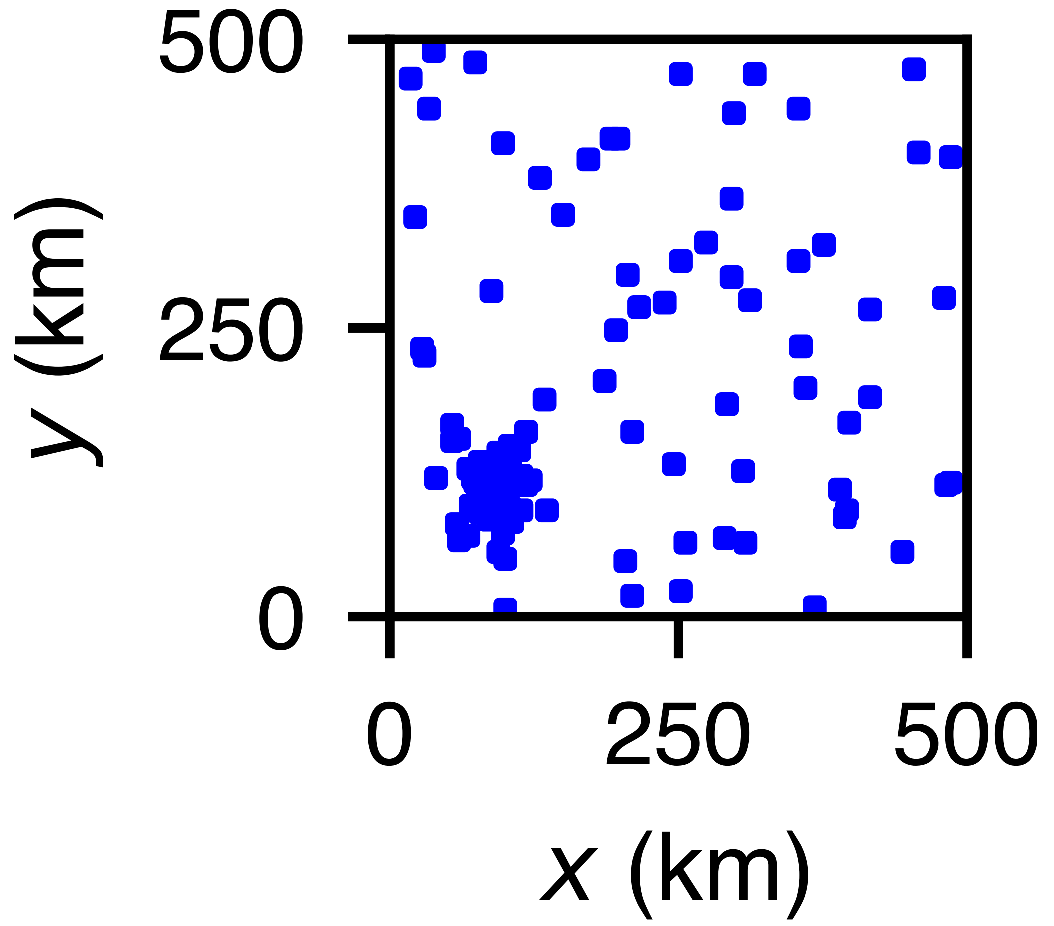


② The CDF F of NNDs if the **same** number of cloud objects were **randomly** distributed,



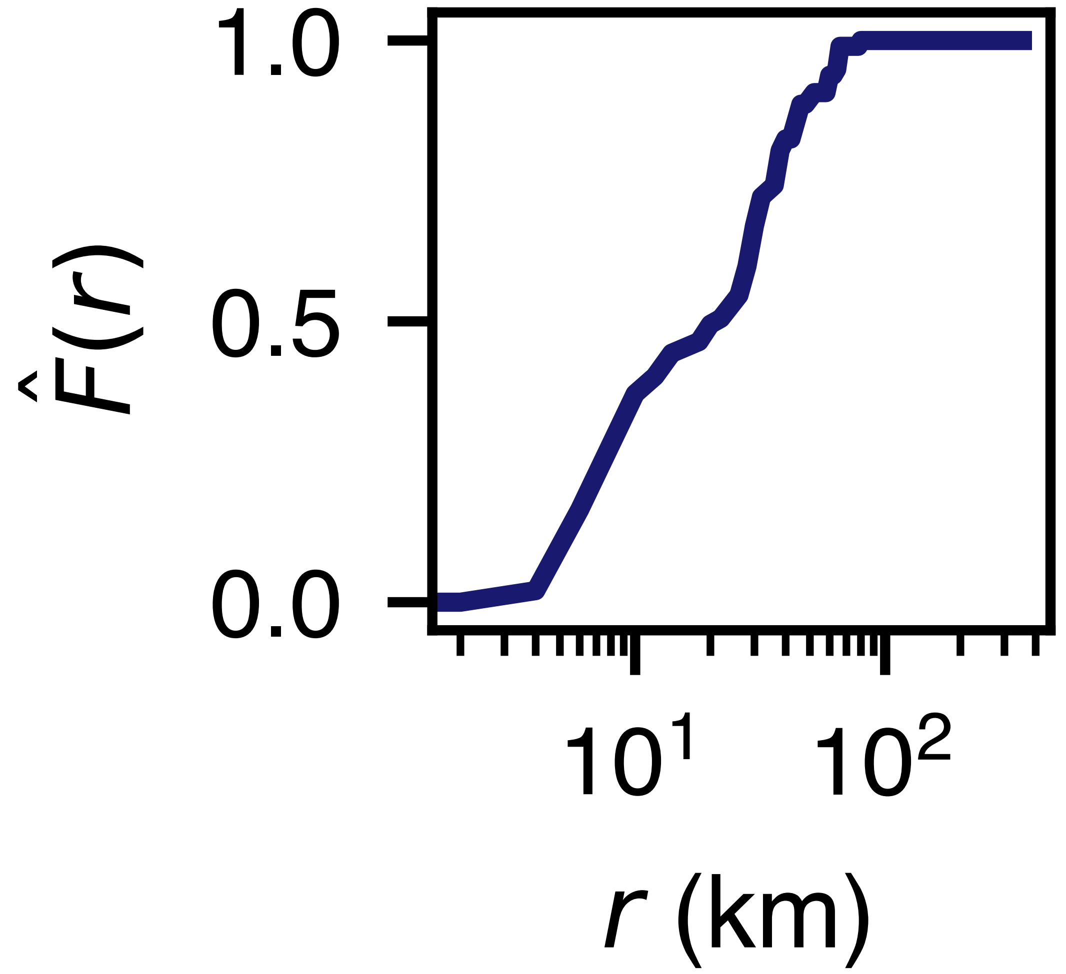
A focus on I_{org}

Given a cloud field scene,

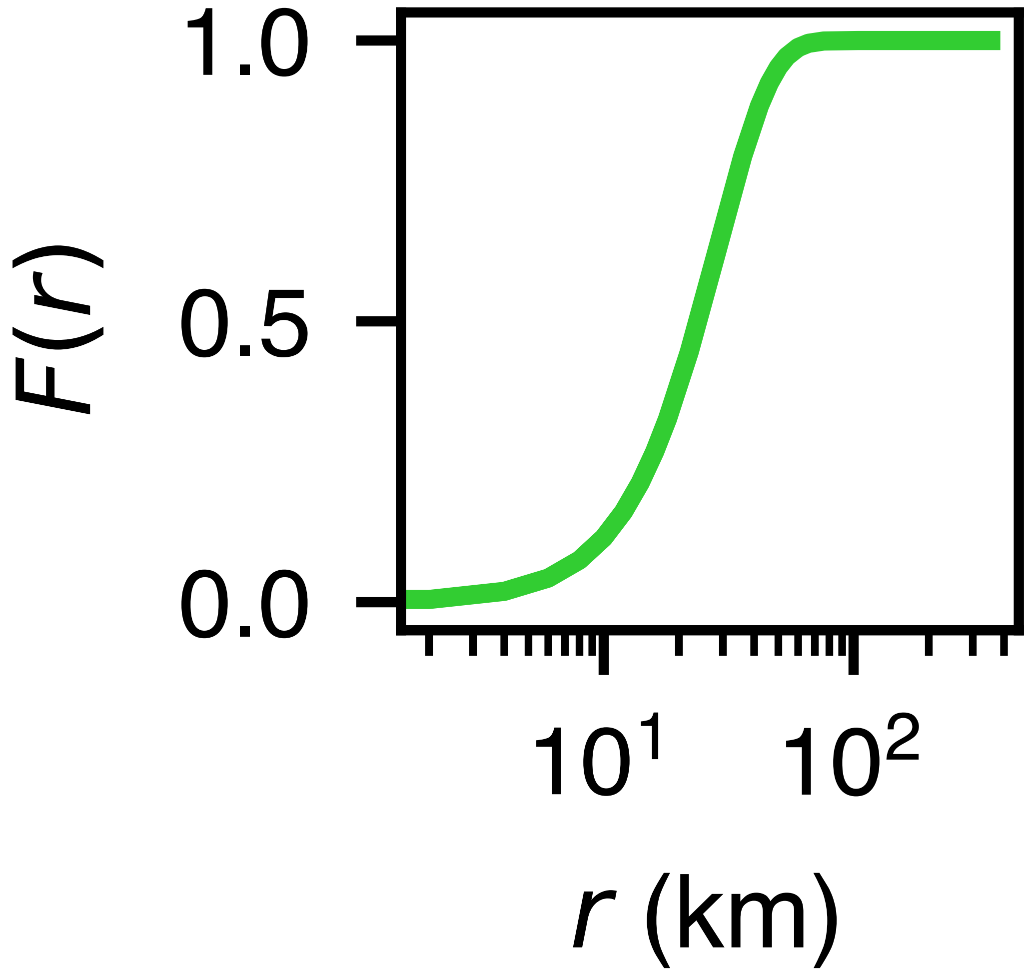


we compute two cumulative distribution functions (CDFs)…

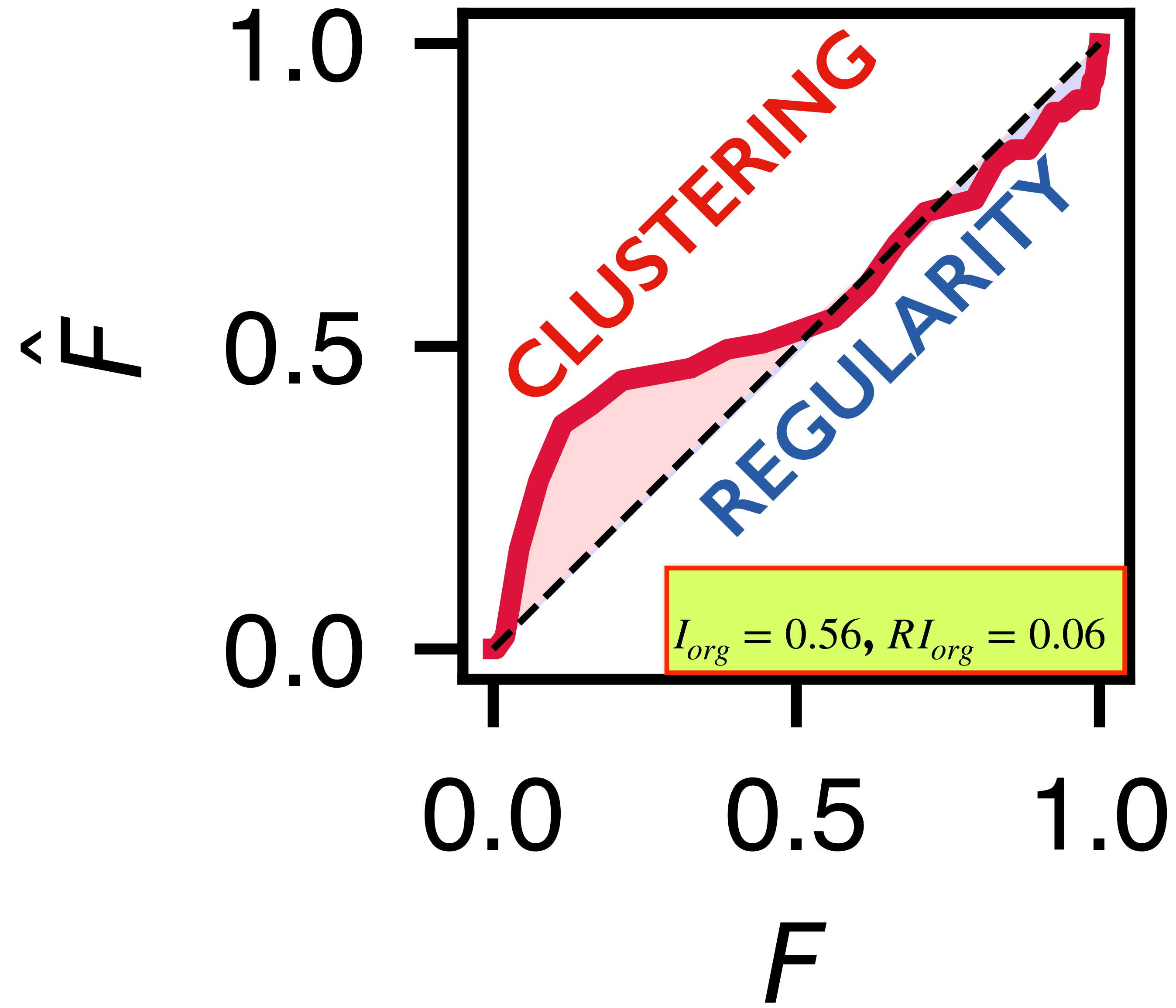
① The CDF $\hat{F}$ of the nearest-neighbor distances (NNDs) in *that* scene,



② The CDF F of NNDs if the **same** number of cloud objects were **randomly** distributed,

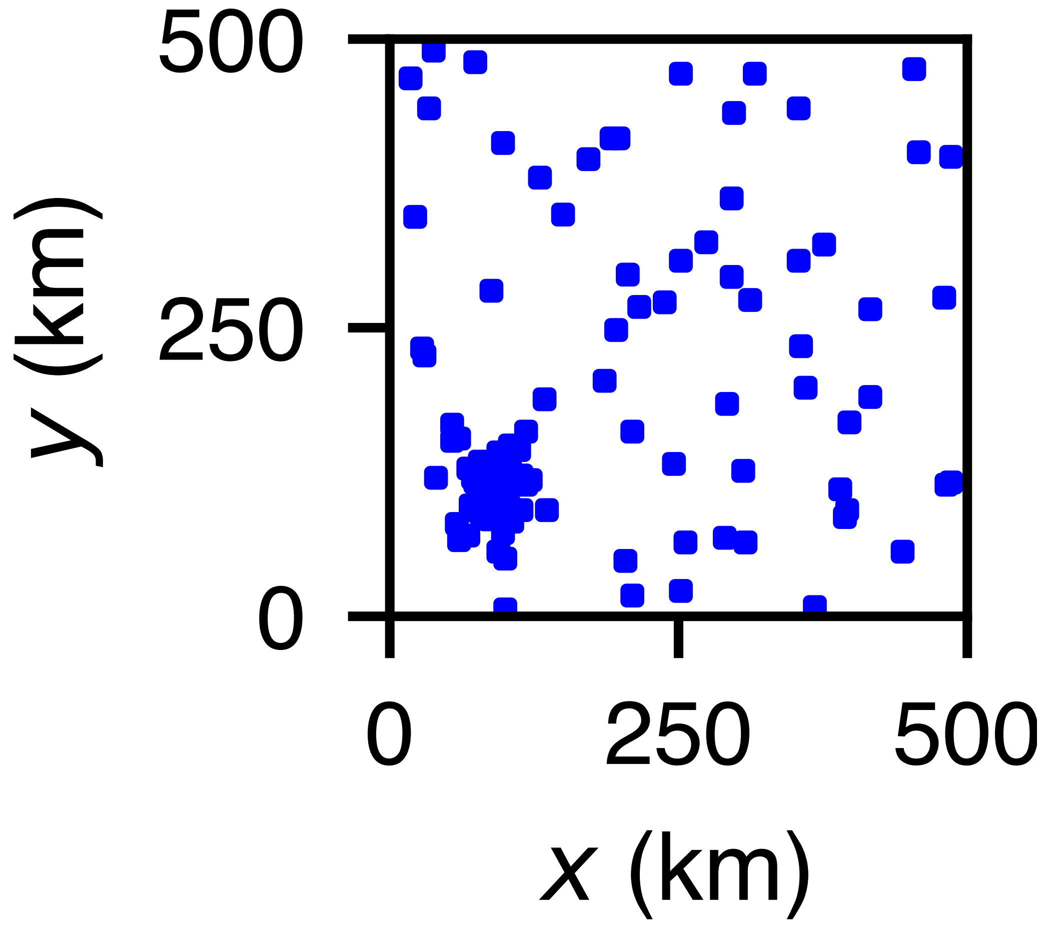


…and we compare them to give I_{org}.



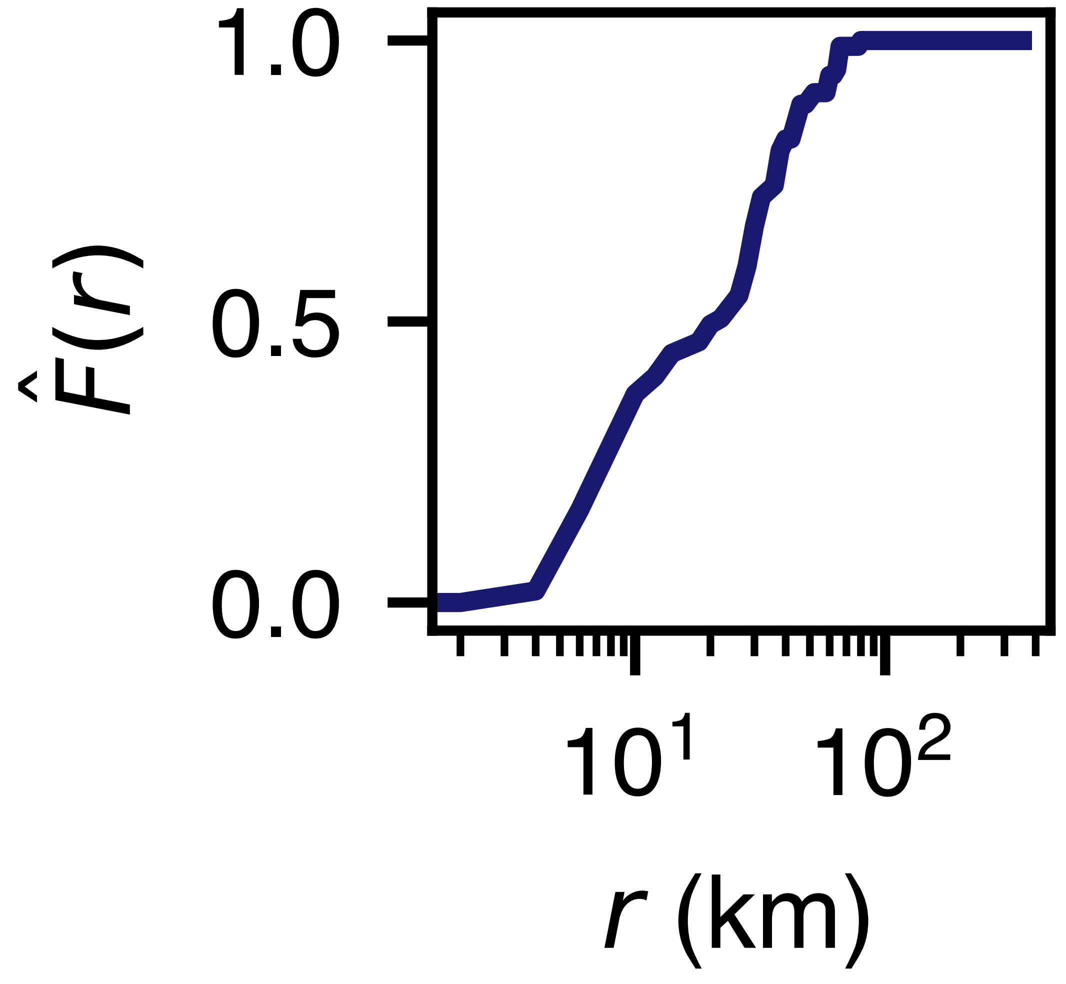
A focus on I_{org}

Given a cloud field scene,

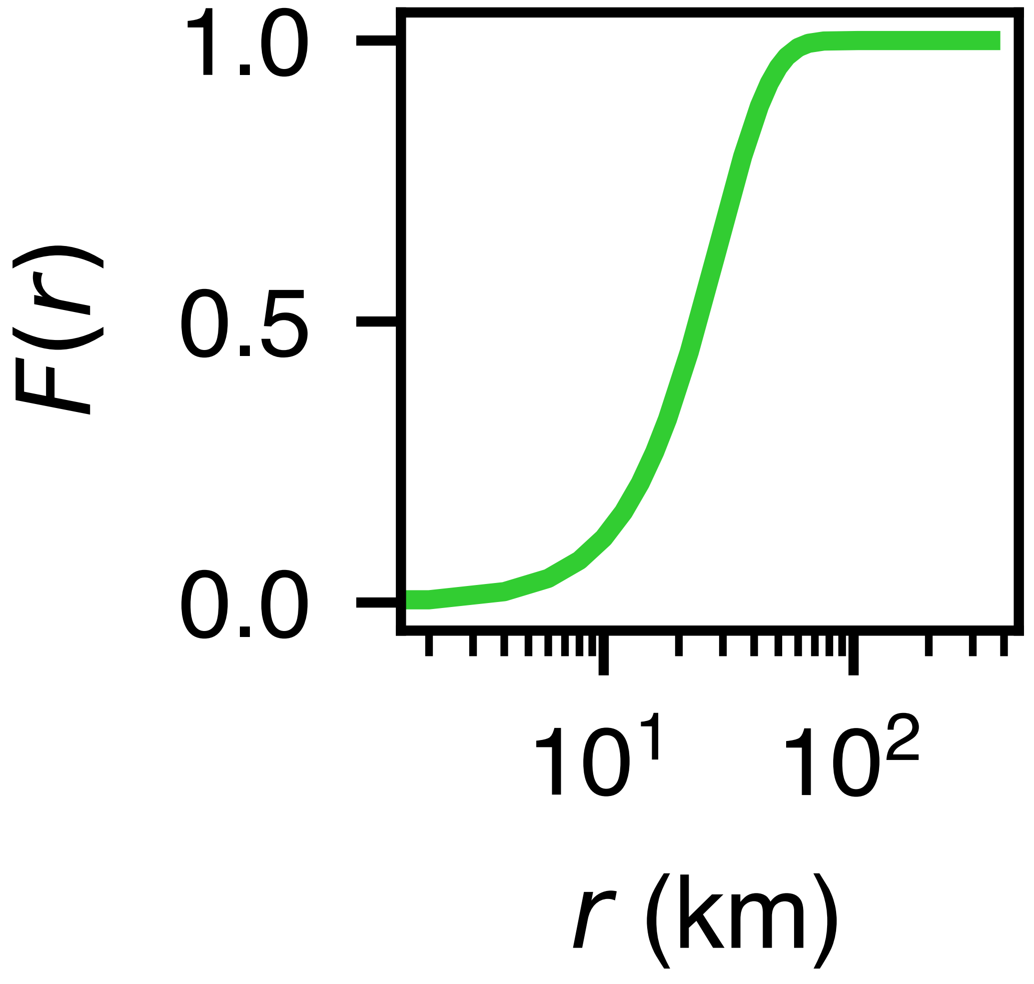


we compute two cumulative distribution functions (CDFs)…

① The CDF $\hat{F}$ of the nearest-neighbor distances (NNDs) in *that* scene,

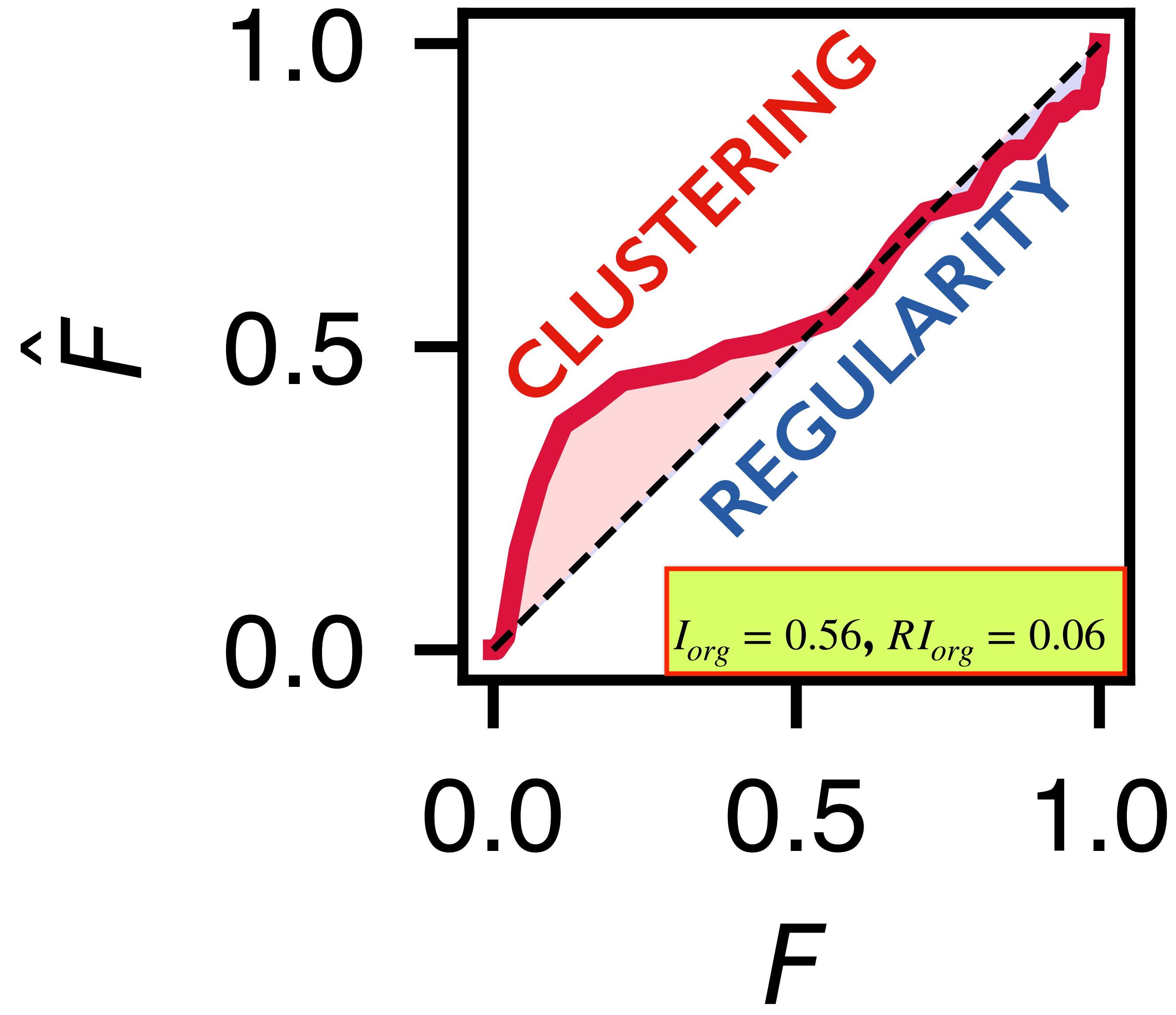


② The CDF F of NNDs if the **same** number of cloud objects were **randomly** distributed,



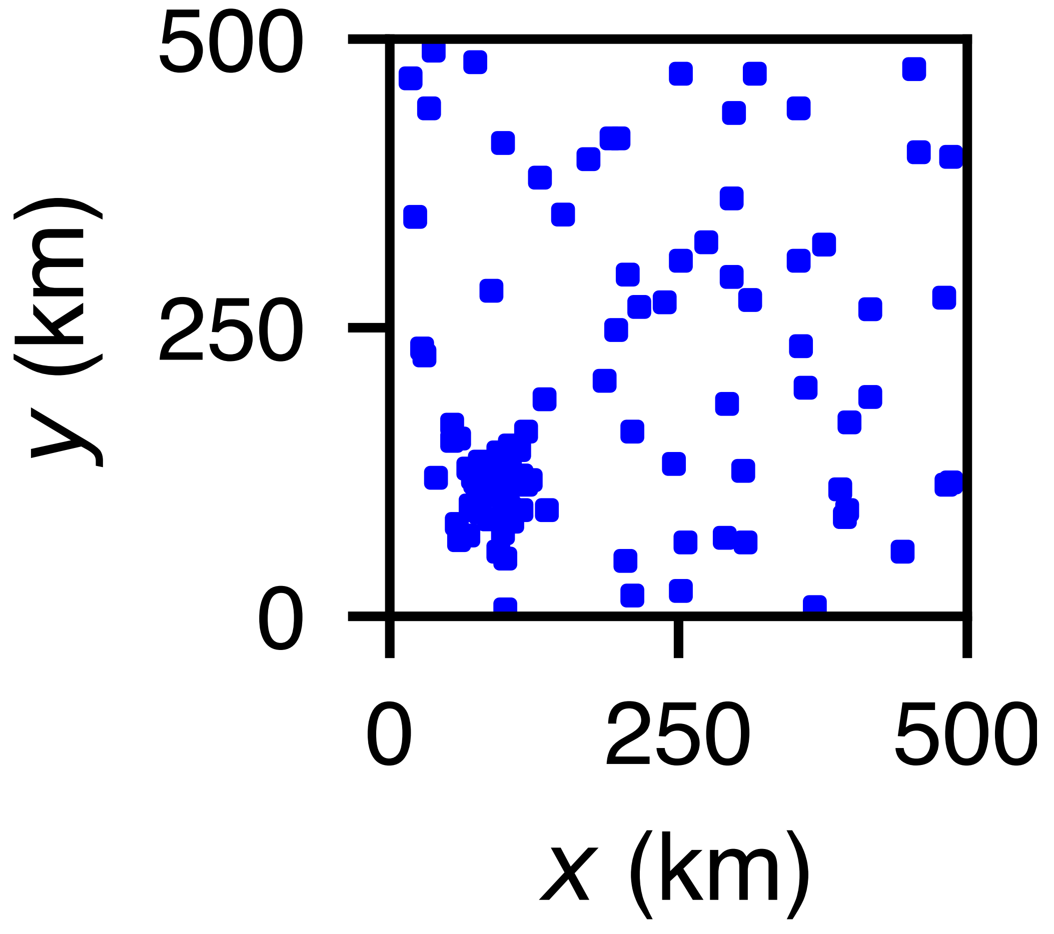
…and we compare them to give I_{org}.

I_{org} is the area under the joint CDF.



A focus on I_{org}

Given a cloud field scene,



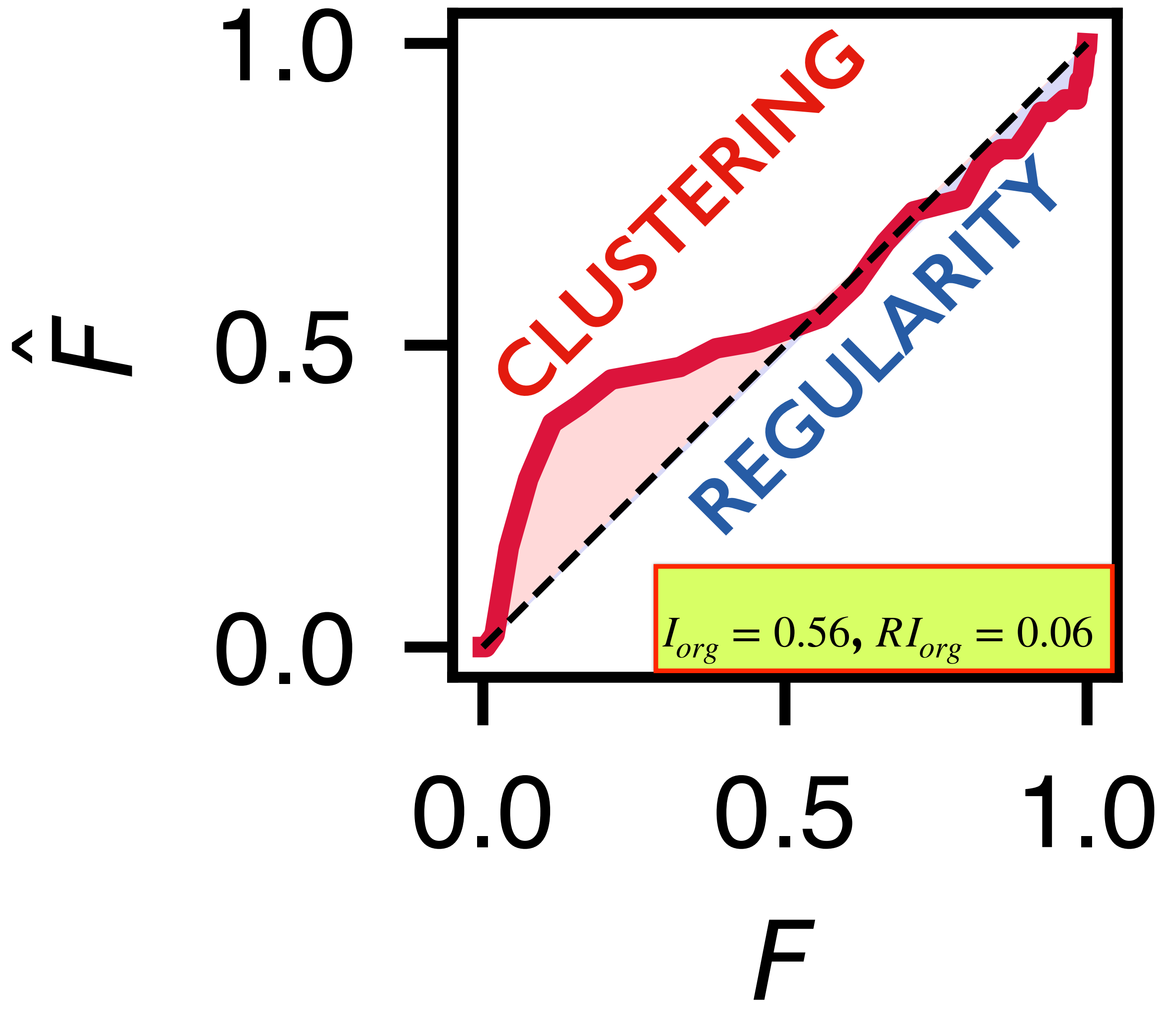
we compute two cumulative distribution functions (CDFs)…

① The CDF $\hat{F}$ of the nearest-neighbor distances (NNDs) in *that* scene,

② The CDF F of NNDs if the **same** number of cloud objects were **randomly** distributed,

…and we compare them to give I_{org}.

I_{org} is the area under the joint CDF.



We also introduce the **relative I_{org}**, RI_{org}, as the integral of $\hat{F} - F$, i.e., $RI_{org} = I_{org} - 0.5$.

I_{org}/RI_{org} categorize scenes as

- **random** ($I_{org} = 0.5$, $RI_{org} = 0$),
- **clustered** ($I_{org} > 0.5$, $RI_{org} > 0$),
- **regular** ($I_{org} < 0.5$, $RI_{org} < 0$).

LMD

Pros and cons of I_{org}

✓ Theoretical null to compare against (Poisson point process)

✓ Measures organization in an <u>absolute</u> sense (cf. point ① listed before)

✗ Sensitive to event number and positions

✗ **Blind to organization beyond the β-mesoscale** $\mathcal{O}$(20-200 km), cf. Orlanski (1975).

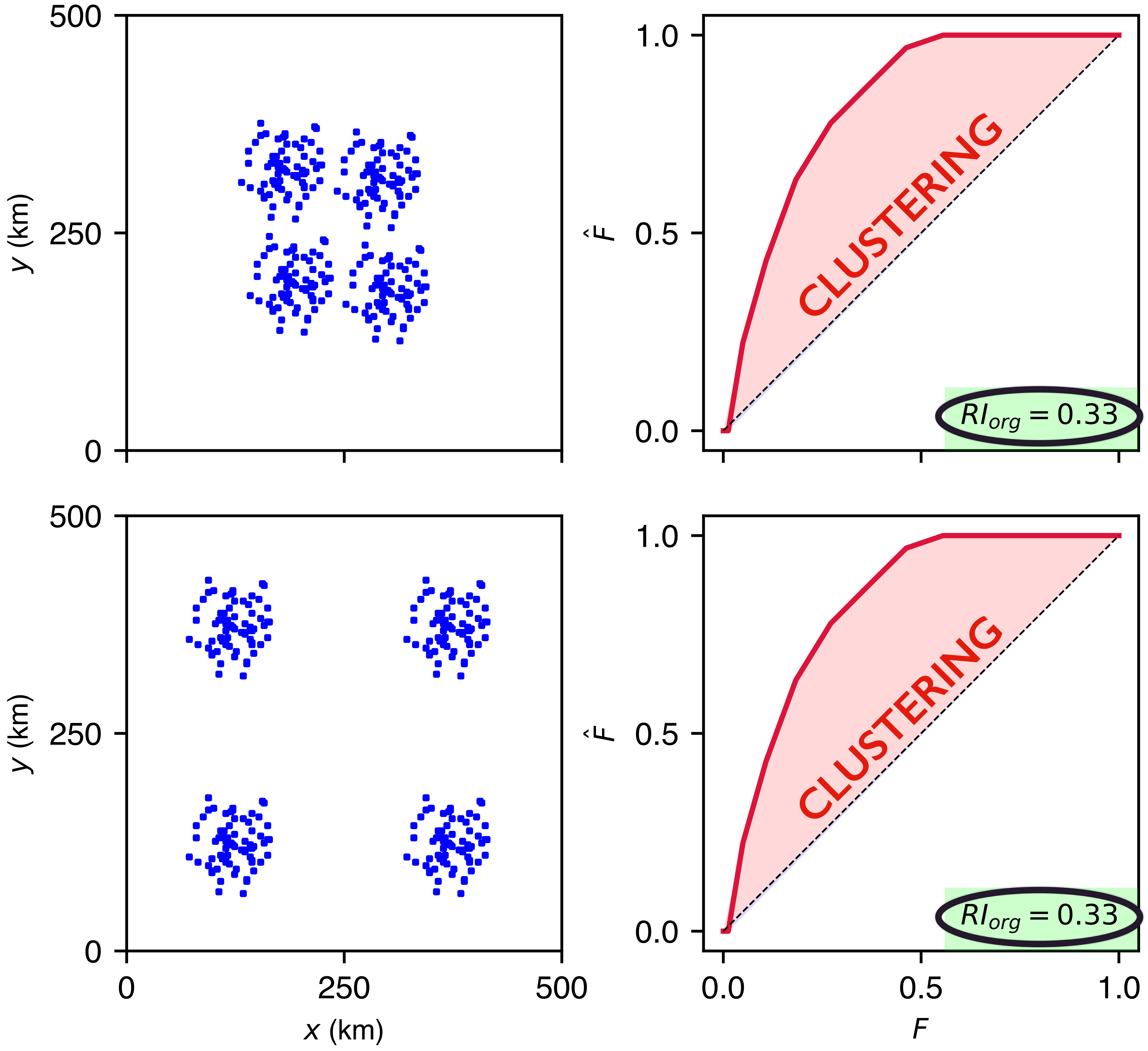
Pros and cons of I_{org}

✓ Theoretical null to compare against (Poisson point process)

✓ Measures organization in an <u>absolute</u> sense (cf. point ① listed before)

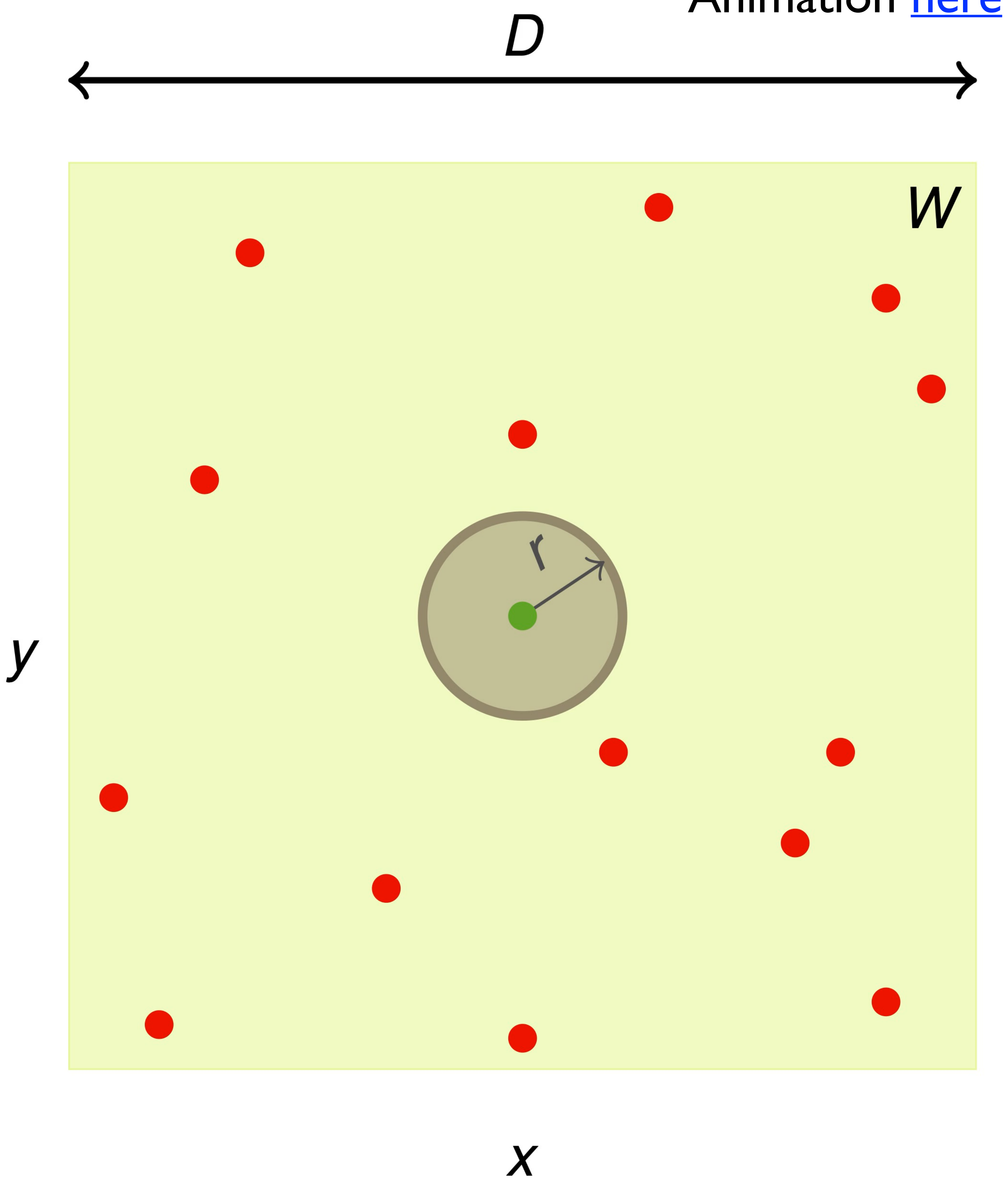
✗ Sensitive to event number and positions

✗ **Blind to organization beyond the β-mesoscale** $\mathcal{O}$(20-200 km), cf. Orlanski (1975).



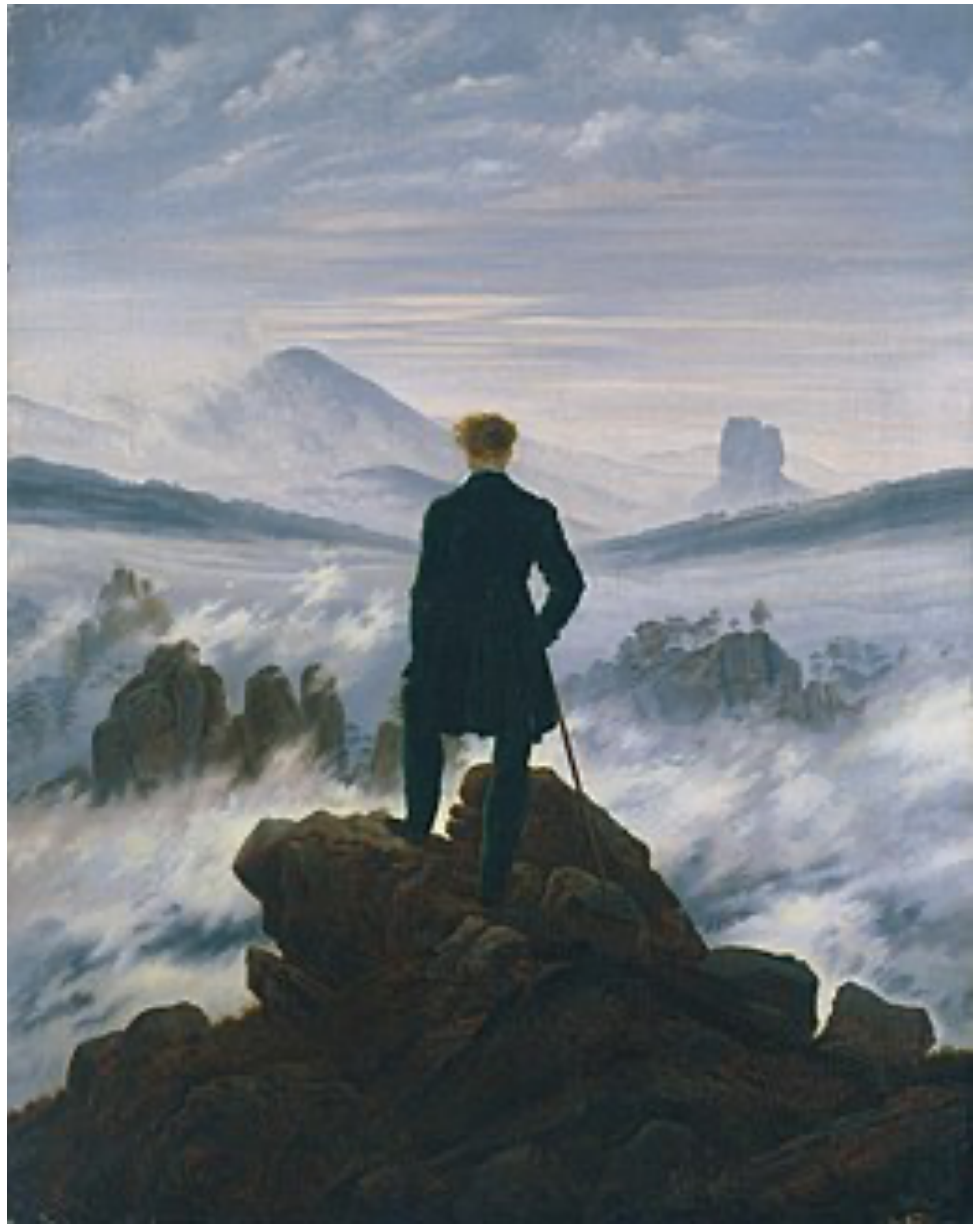
Why don't we broaden our horizons?

Animation here



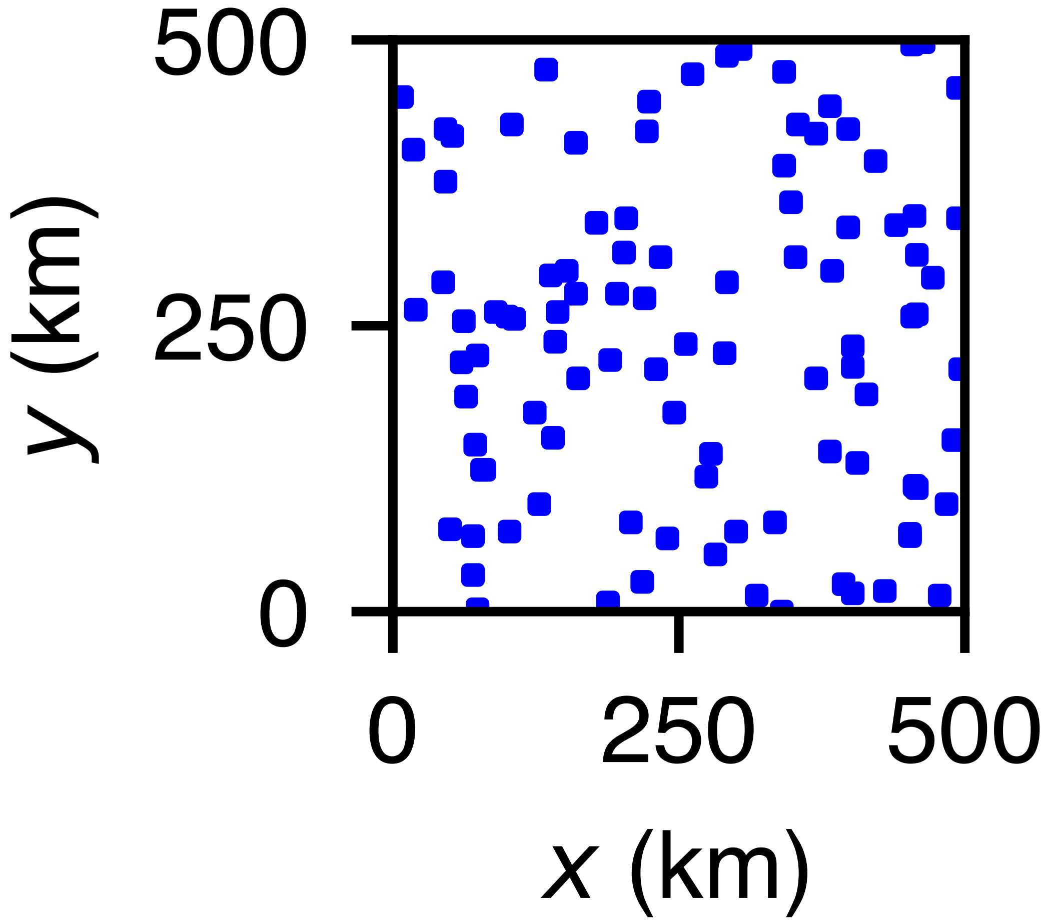
- base point (for neighbor count)
- points of the spatial pattern (i.e., clouds)
- neighbors of ● within distance *r*

We will use a commonly used tool in the statistics analysis of spatial point patterns: the ***L*-function** ~ mean number of neighbors of a cloud object as a function of spatial scale *r*.



Towards the definition of a new index, L_{org}

Given a cloud field scene,

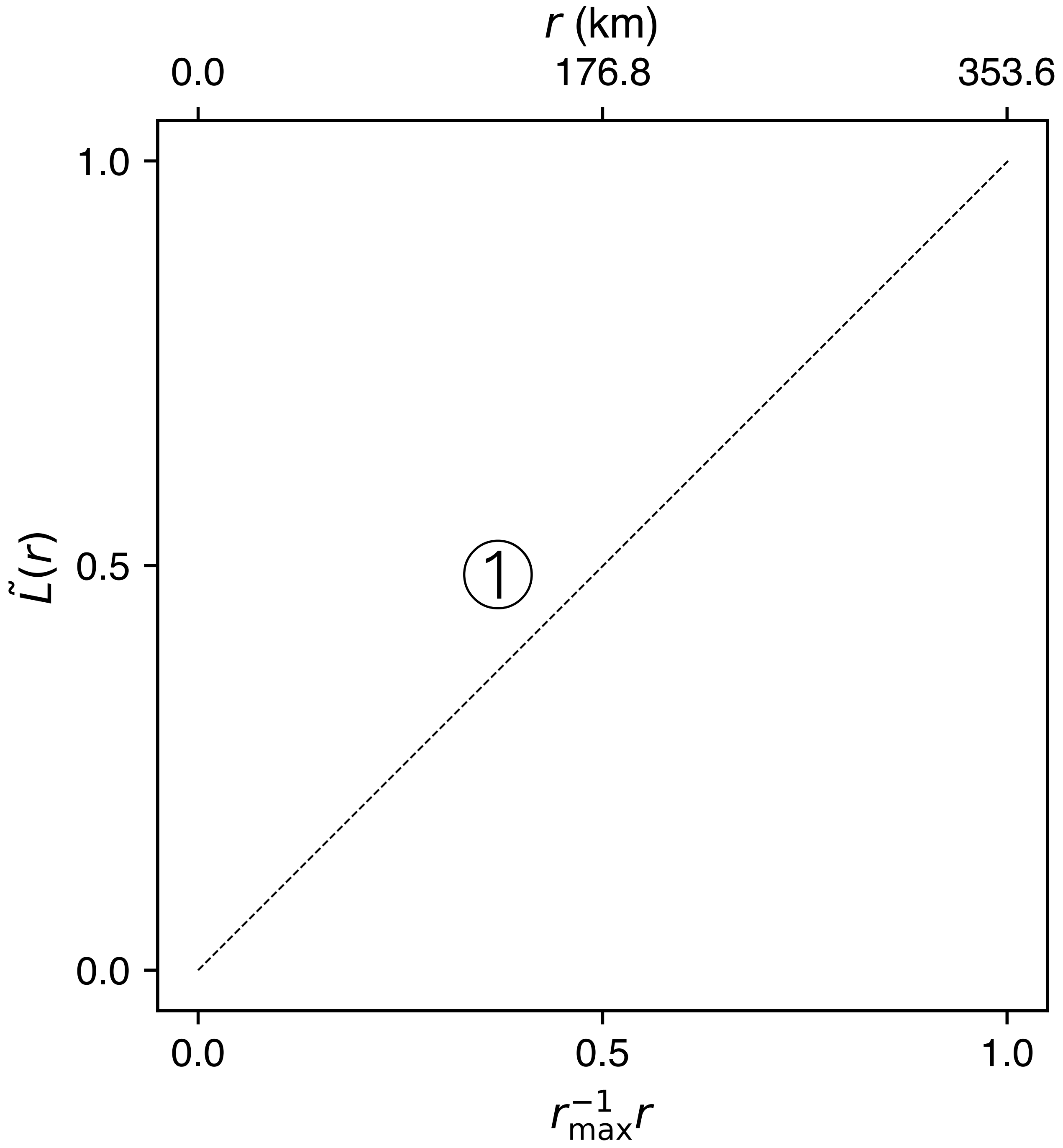


Synthetic scene,
randomly distributed convection,
periodic boundaries

convective objects

we compute

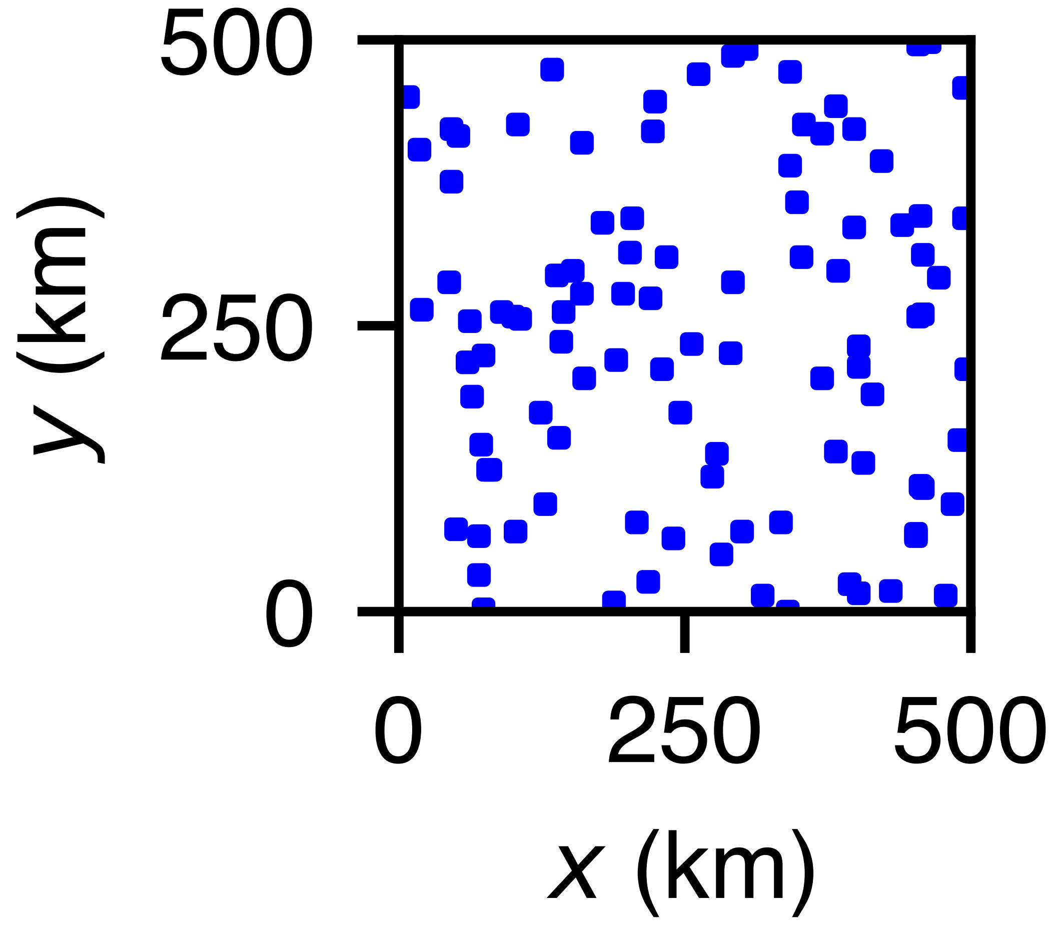
① The theoretical reference L-function $\tilde{L}$,



LMD

Towards the definition of a new index, L_{org}

Given a cloud field scene,



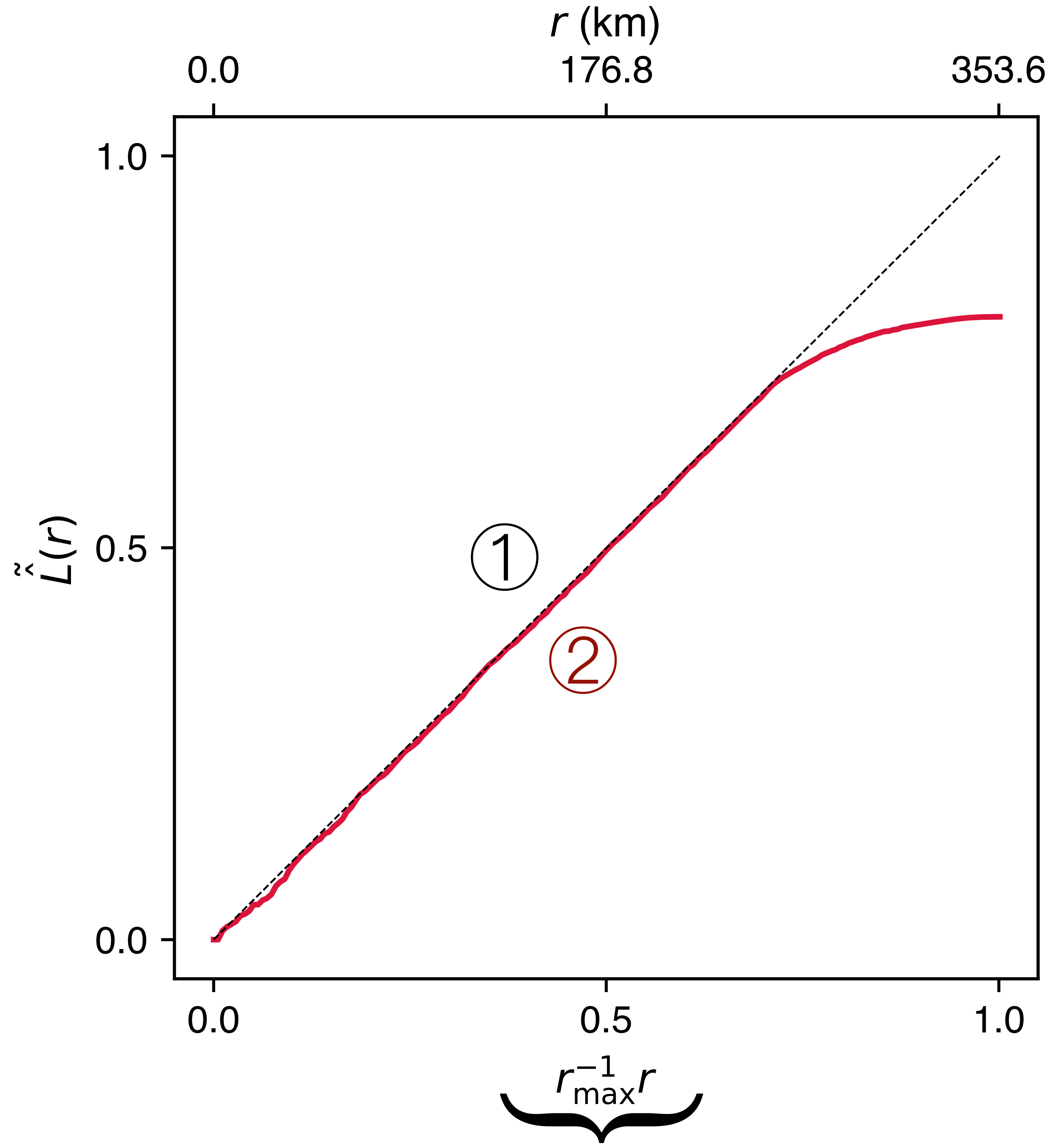
Synthetic scene,
randomly distributed convection,
periodic boundaries

convective objects

we compute

① The theoretical reference L-function $\tilde{L}$,

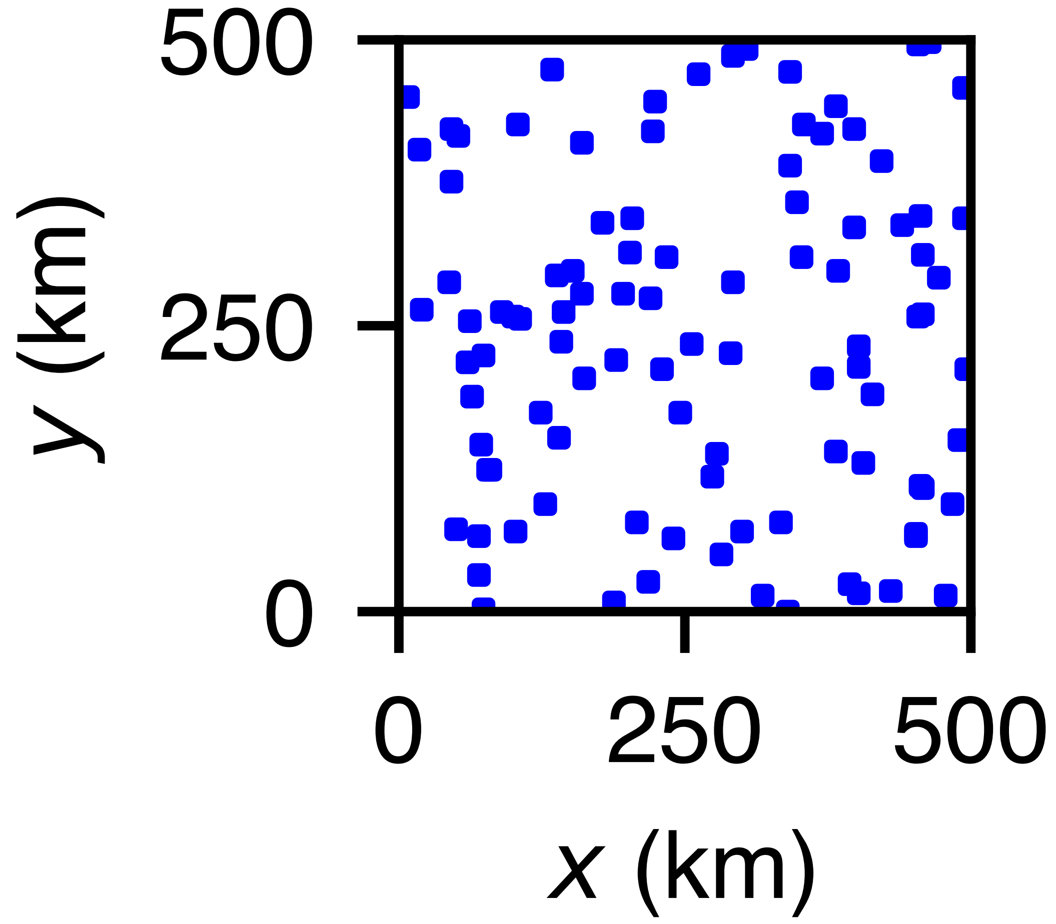
② the L-function $\hat{\tilde{L}}$ derived from the distribution of objects in the scene,



Radii of search discs / max search disc radius in the given domain

Towards the definition of a new index, L_{org}

Given a cloud field scene,



Synthetic scene,
randomly distributed convection,
periodic boundaries

■ convective objects

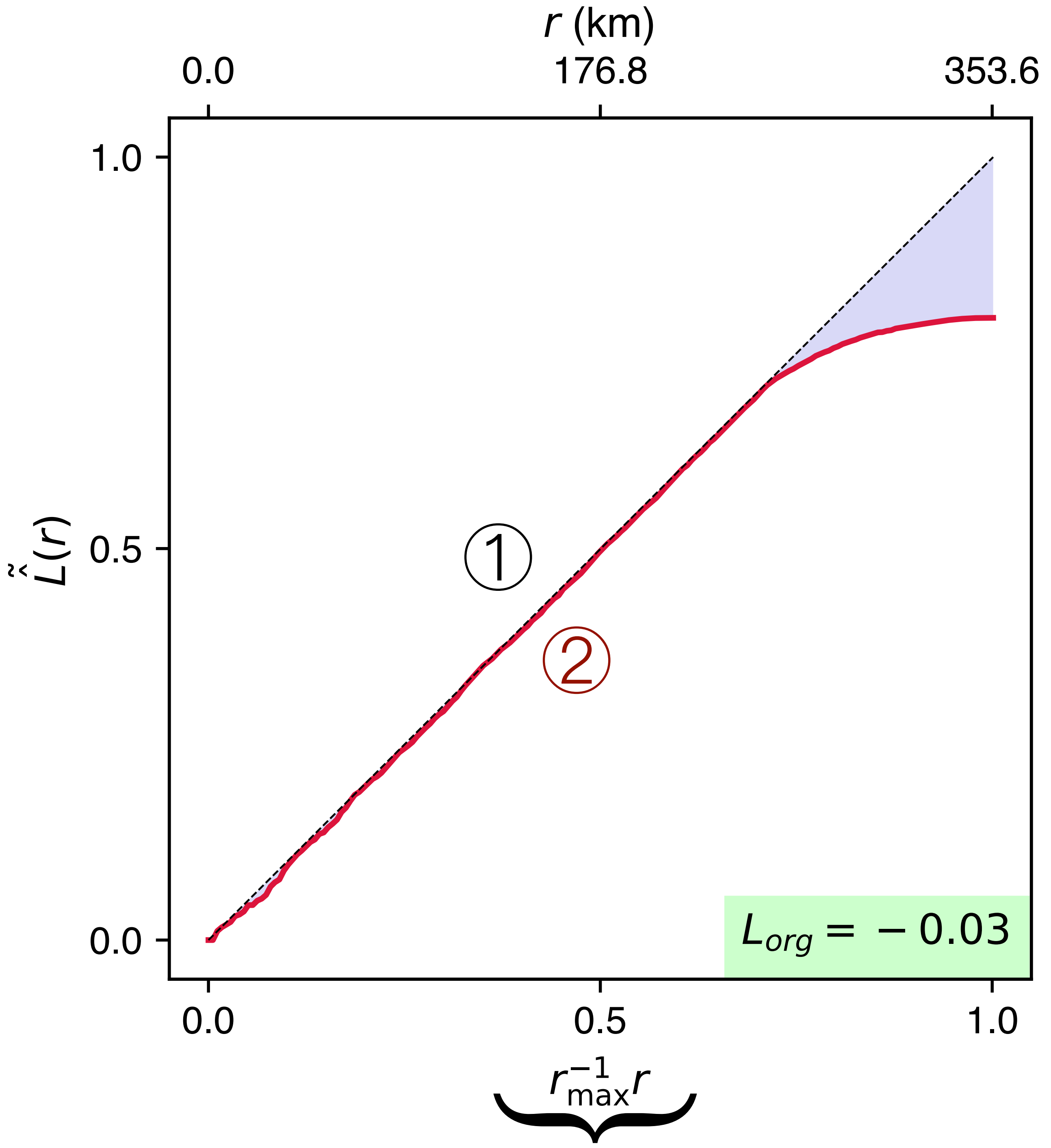
we compute

① The theoretical reference L-function $\tilde{L}$,

② the L-function $\hat{\tilde{L}}$ derived from the distribution of objects in the scene,

③ the integral departure $\hat{\tilde{L}} - \tilde{L}$ to give L_{org}.

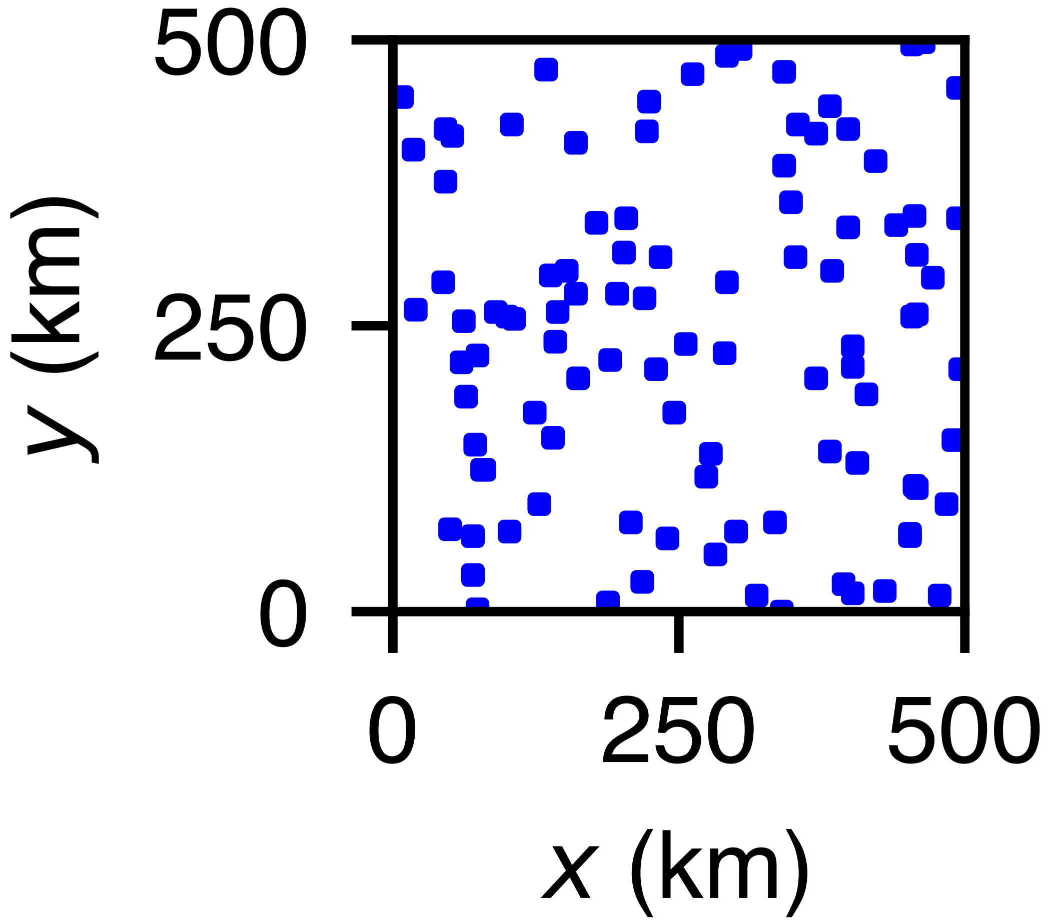
The scenes are classified as - **random** ($L_{org} = 0$),
- **clustered** ($L_{org} > 0$),
- **regular** ($L_{org} < 0$).



Radii of search discs / max search disc radius in the given domain

Towards the definition of a new index, L_{org}

Given a cloud field scene,



Synthetic scene,
randomly distributed convection,
periodic boundaries

convective objects

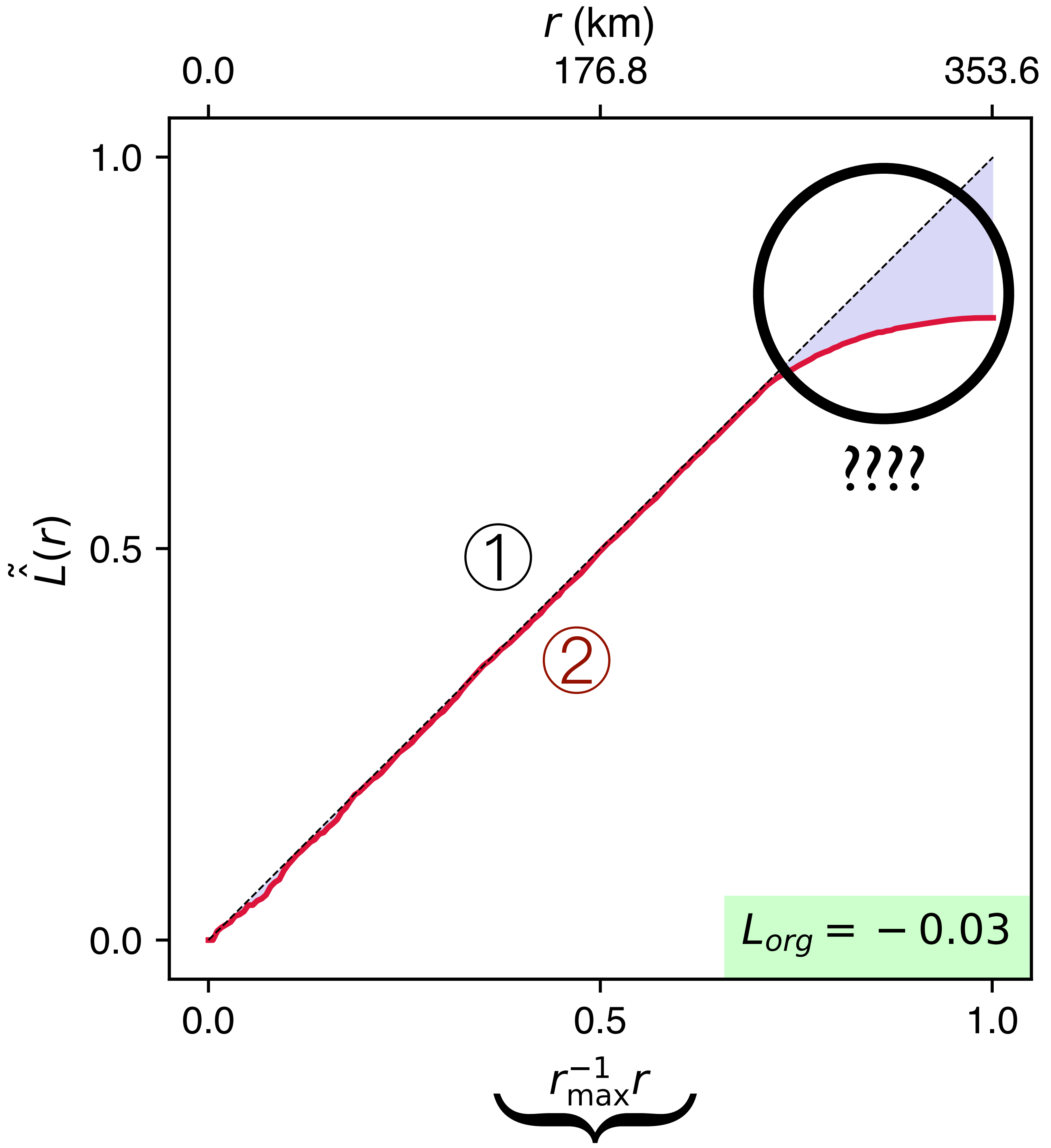
we compute

① The theoretical reference L-function $\tilde{L}$,

② the L-function $\hat{\tilde{L}}$ derived from the distribution of objects in the scene,

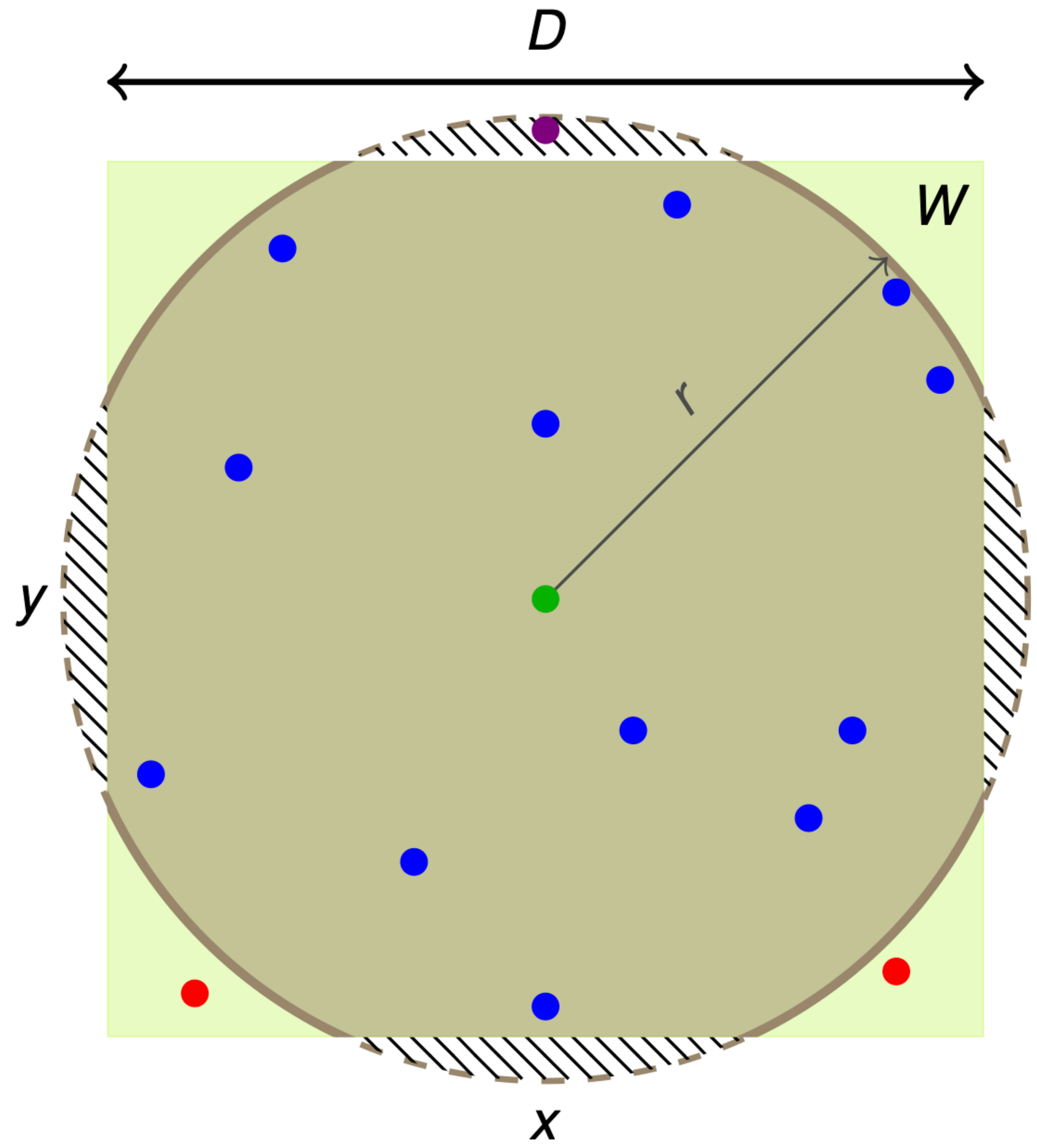
③ the integral departure $\hat{\tilde{L}} - \tilde{L}$ to give L_{org}.

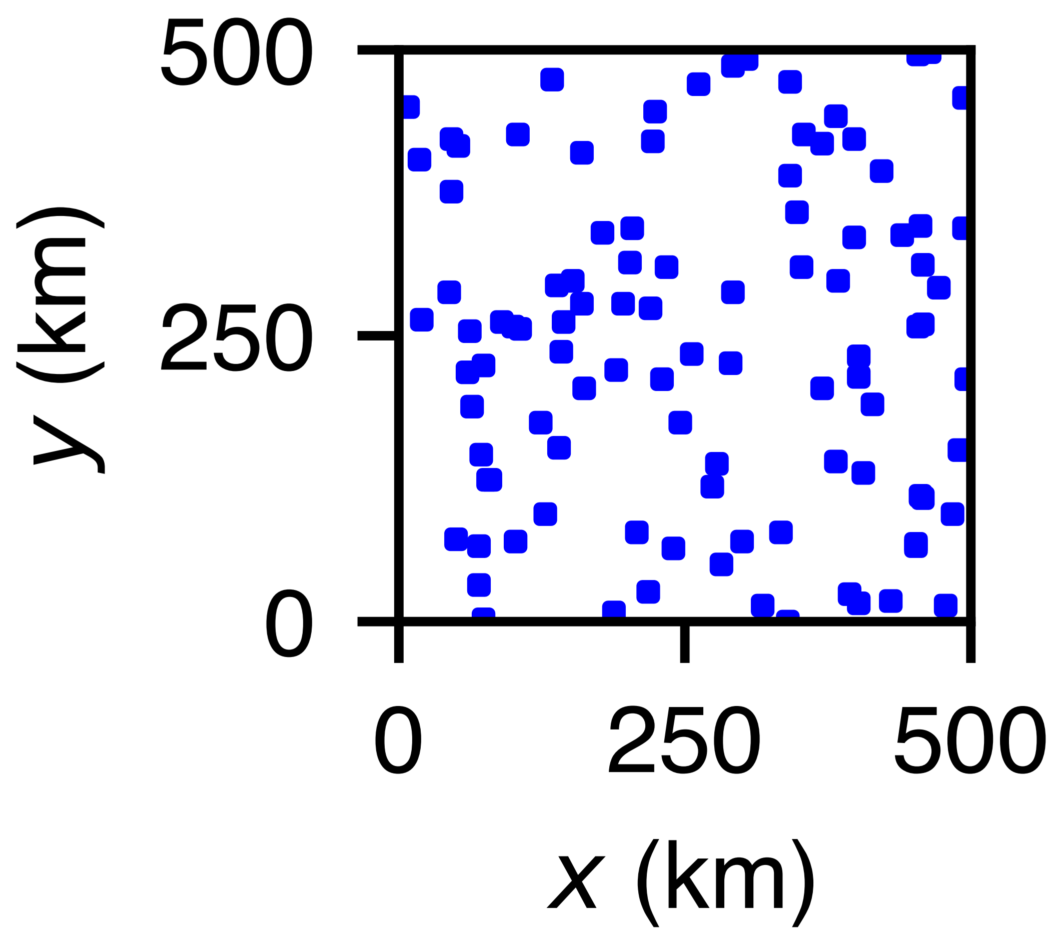
The scenes are classified as - **random** ($L_{org} = 0$),
- **clustered** ($L_{org} > 0$),
- **regular** ($L_{org} < 0$).



Radii of search discs / max search disc radius in the given domain

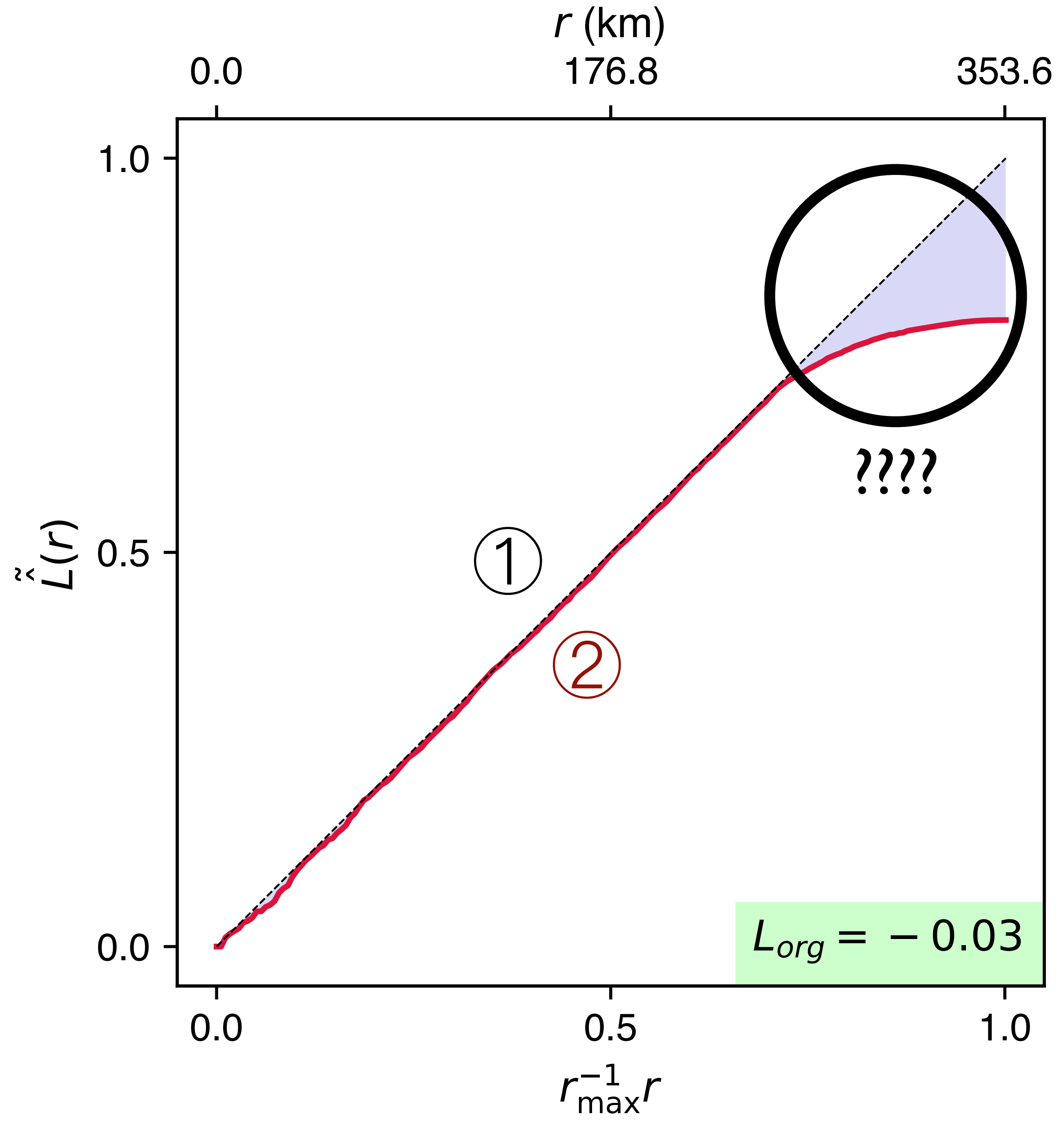
A correction is needed for periodic boundary conditions





Synthetic scene,
randomly distributed convection,
periodic boundaries

■ convective objects

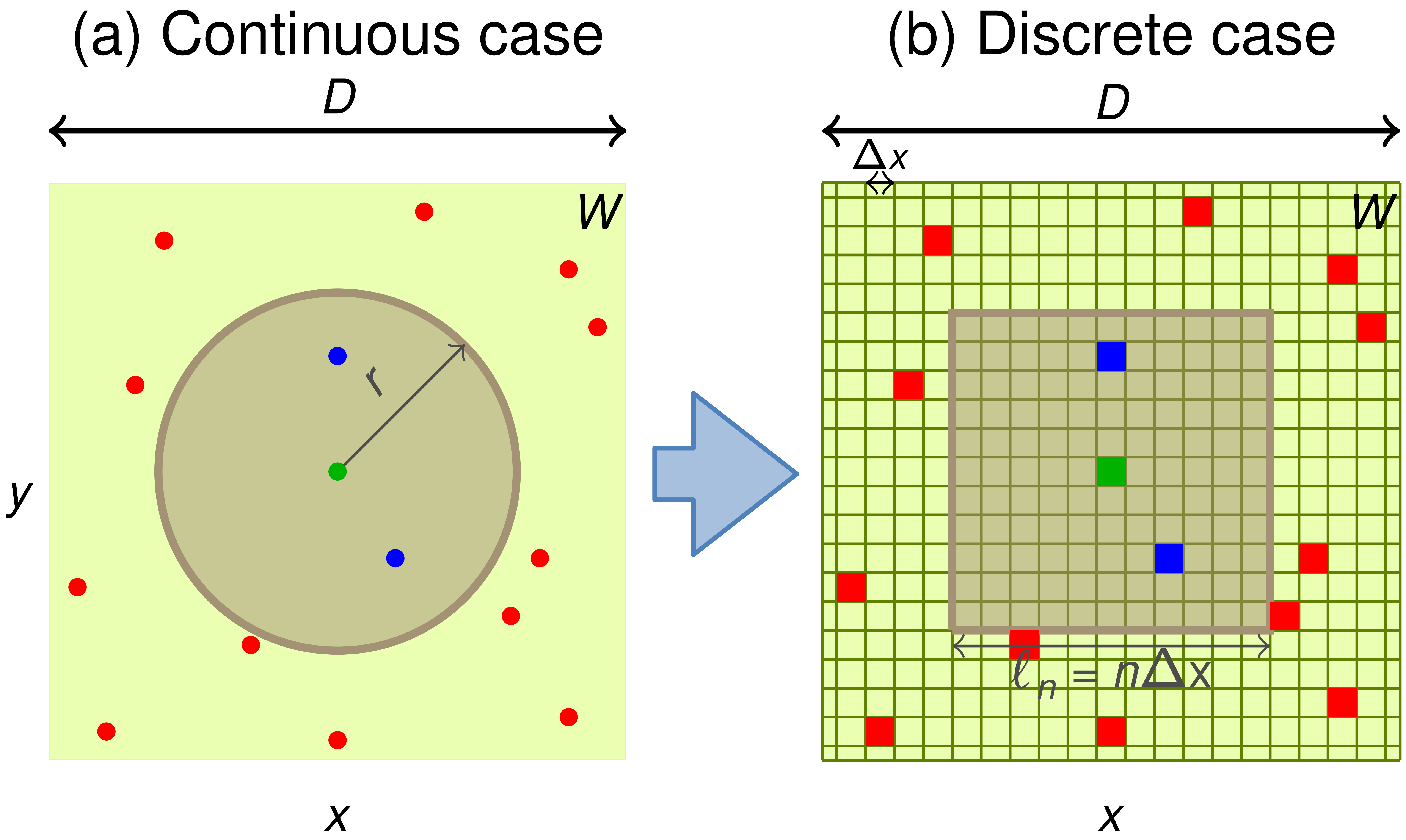


LMD

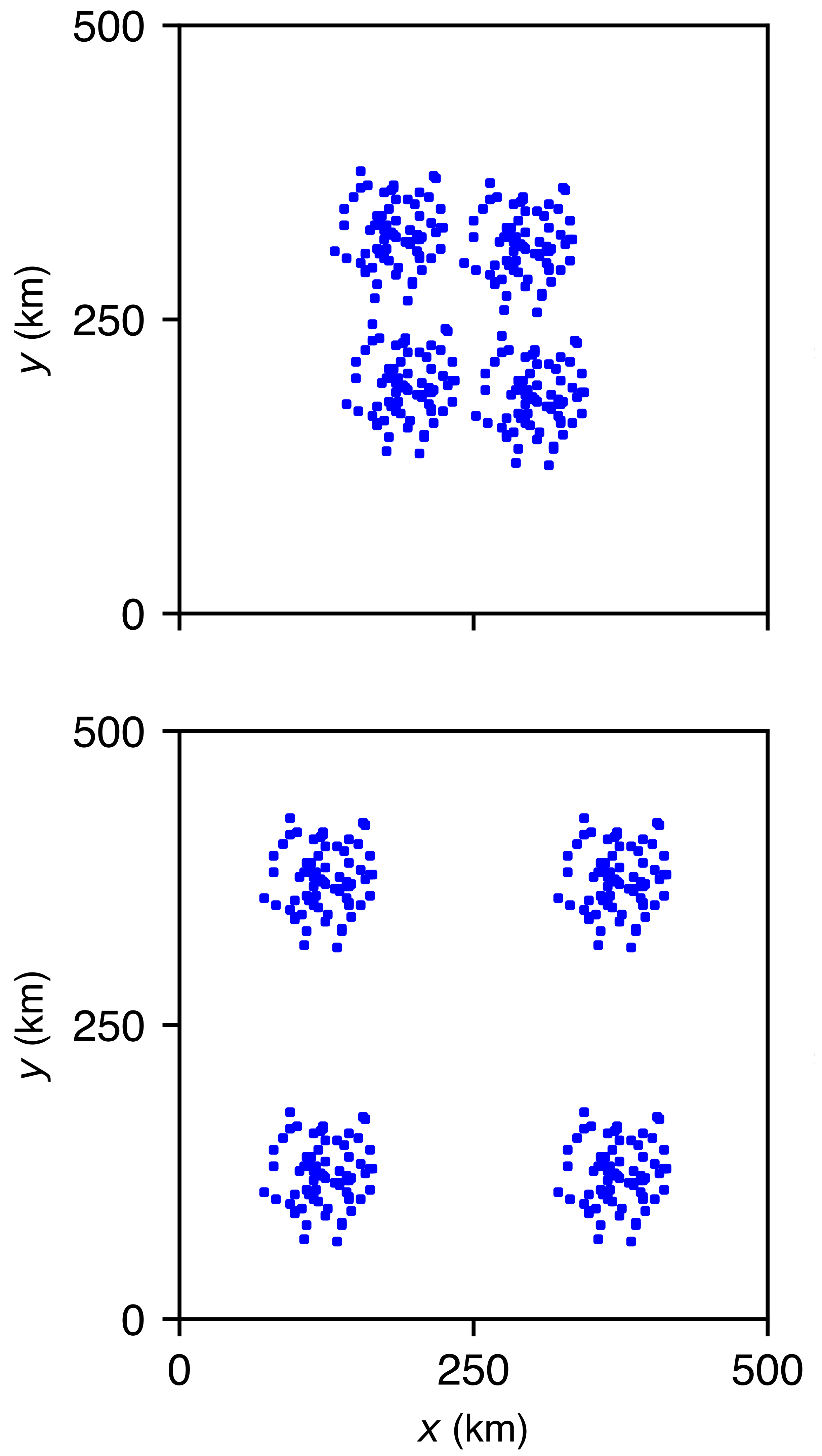
A discrete counterpart for the analysis of gridded data

In practical applications (analysis of model output data/observational datasets) we consider finite, discrete grids.

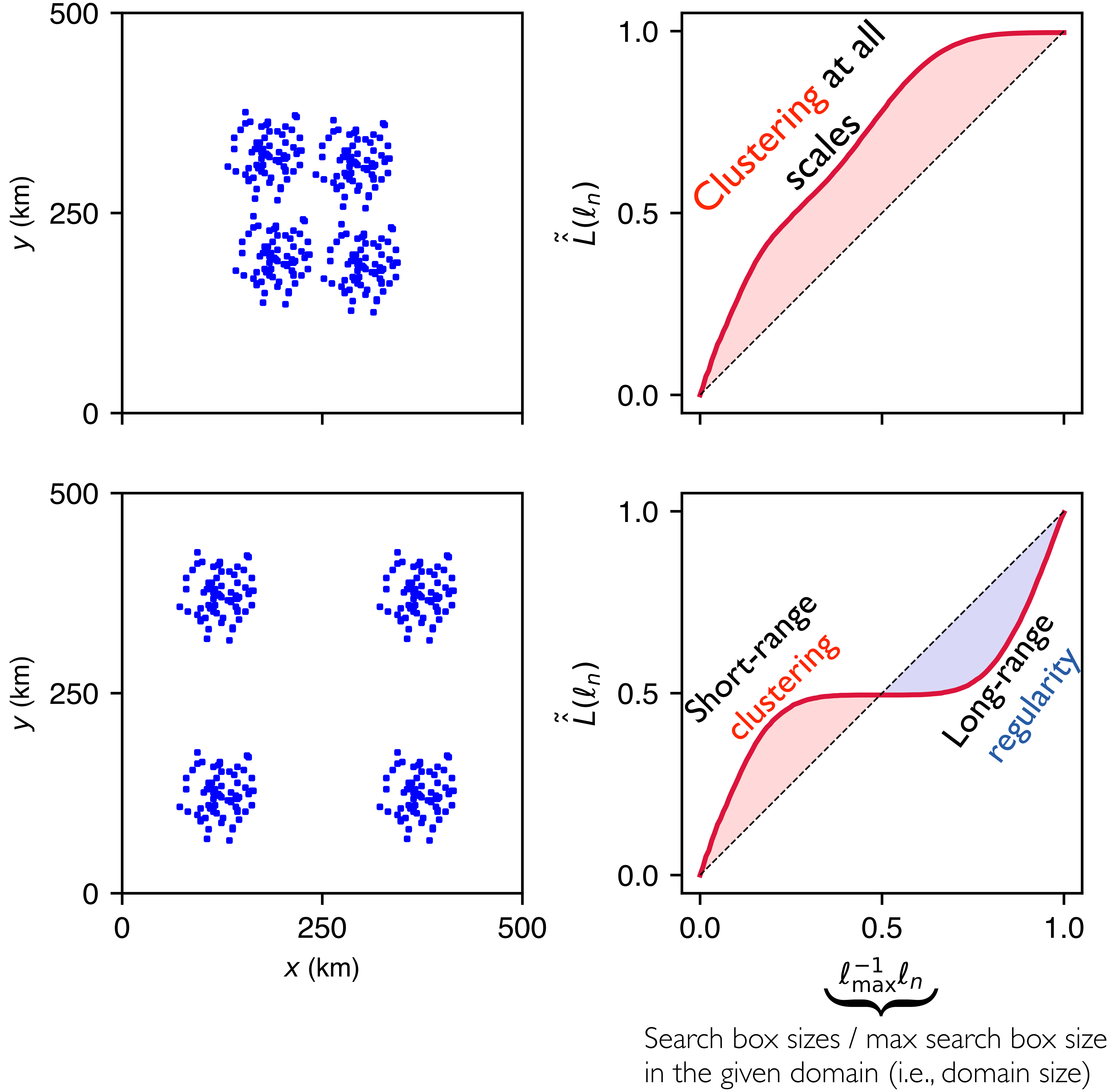
The evaluation of neighbor count is now performed over square observation boxes of size $\ell_n = n\Delta x$.



Capturing organization beyond the β-mesoscale



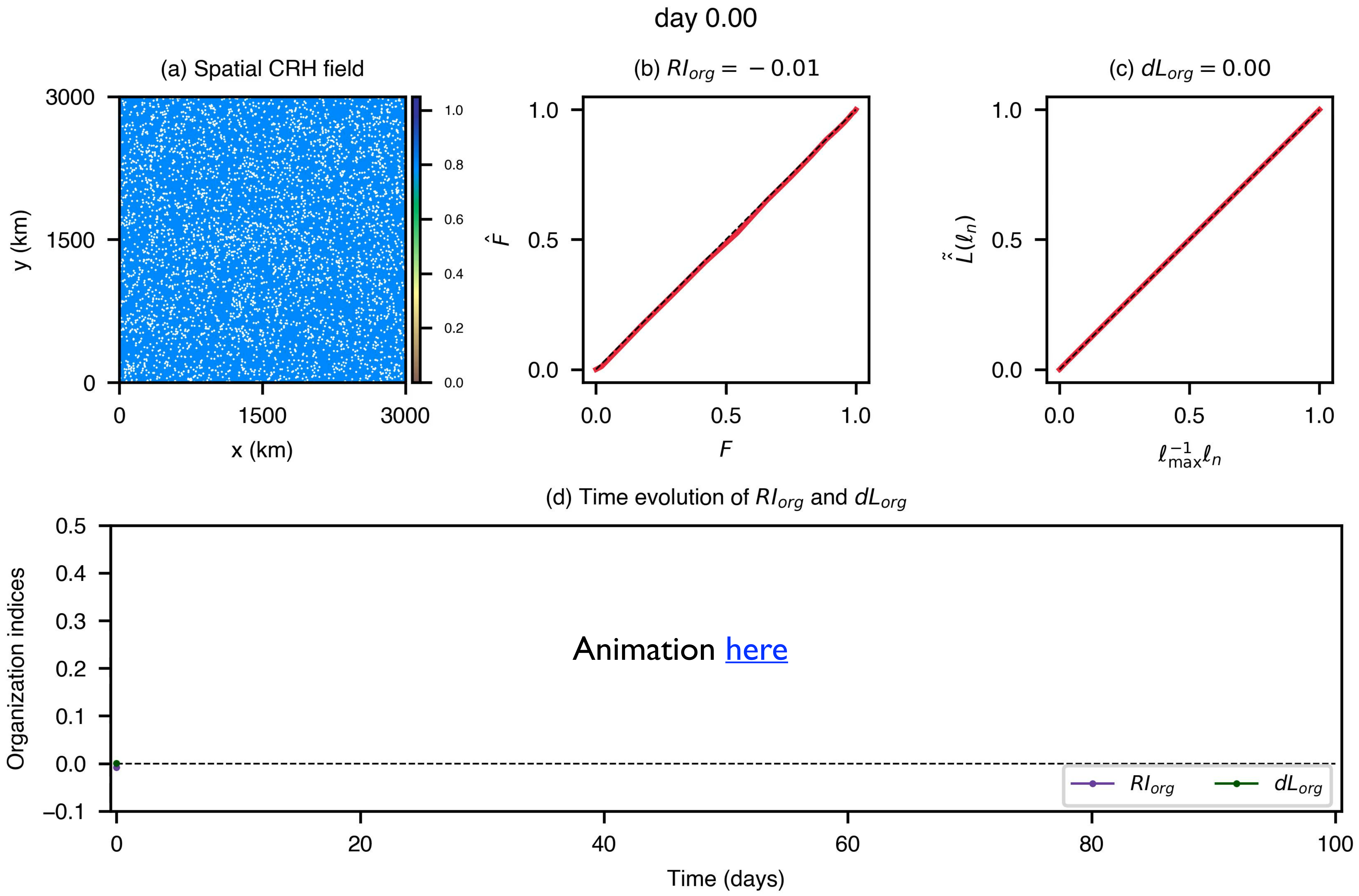
Capturing organization beyond the β-mesoscale



The *L*-functions can capture the different regimes of organization in the short- and long-range.

Application to model output data

- Idealized stochastic model for tropical convection introduced before.
- I_{org} "saturates" almost immediately.
- New index evolves smoothly across various degrees of self-aggregation.

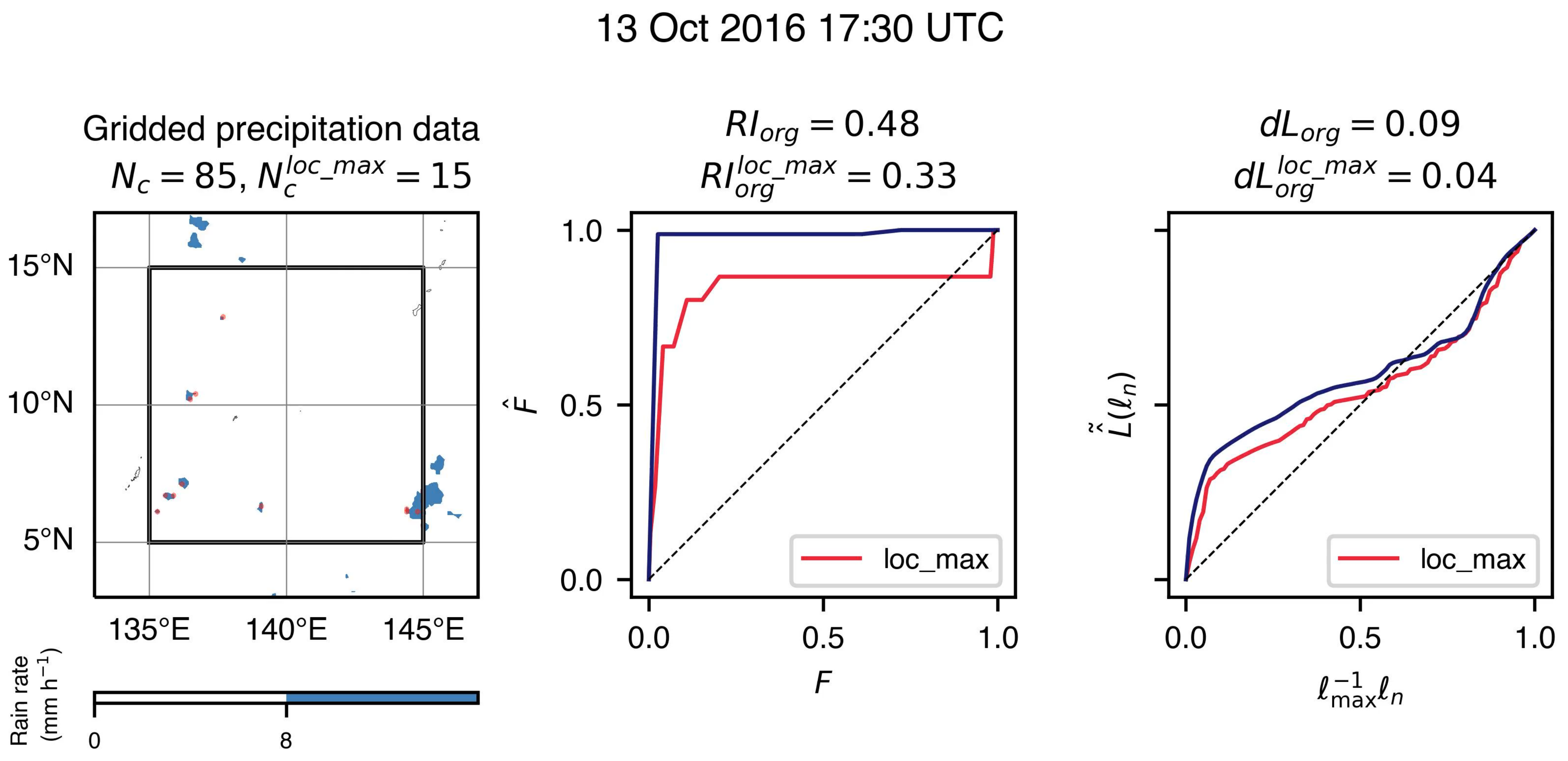


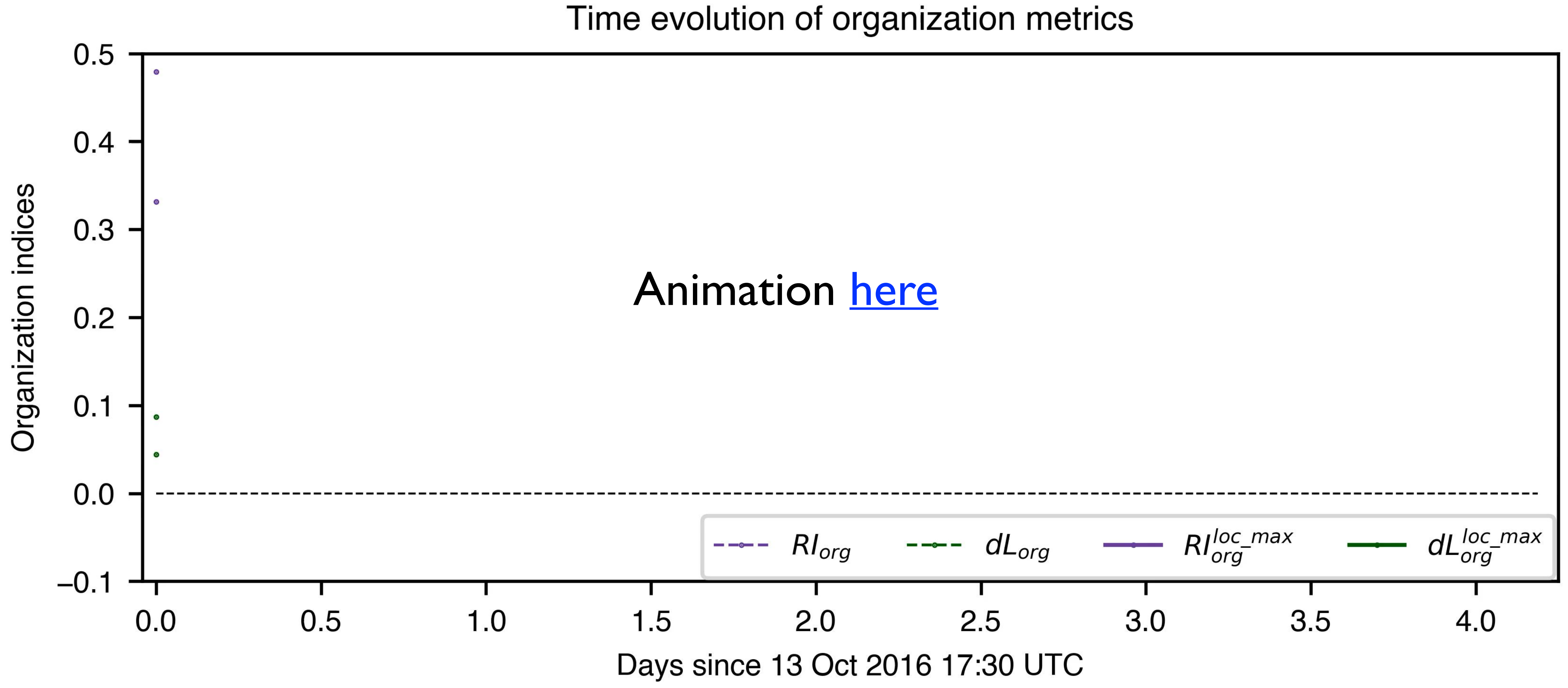
Animation here

Application to observations

- IMERG precipitation data.
- 8 mm/h rain rate threshold to identify convective cells.
- Local maximum method also applied to isolate most vigorous updraft cores (cf. Bony et al., 2020).
- Scenes with <15 convective objects excluded (I_{org} can be very noisy, cf. Semie and Bony, 2020).

New index much less temporally noisy and far more robust to calculation details than I_{org}.





A summary

Can we understand the sensitivities of self-aggregation as found in CRMs with a toy model?

We introduced a model that reduces CRM complexity as much as possible, retaining only the essential physics. The effects of each process can be easily disentangled from the others.

A new explanation is offered for domain size and horizontal resolution sensitivities of self-aggregation found in CRMs.

A new dimensionless parameter predicts self-aggregation onset in the simple model. Is it applicable as a diagnostic to CRMs?

Can we better (or complementarily) measure convective organization in model output and observations?

A new metric, L_{org}, is introduced, which is similar to a popular one, I_{org}, in its theoretical foundations (comparison of two distance distribution functions), with NNCDFs replaced by L-functions.

New metric captures organization over a range of scales and also far more robust to calculation details than I_{org}.

L_{org} suitable for measuring organization strength in model inter-comparison studies and in a wide variety of observations.
Main con: it is more computationally burdensome than I_{org}!