

Authors' Responses to the Comments from Editor and Referees

General reply

Based on the comments from editor and referees, we carefully reconsidered our manuscript. There are too many revisions such that the structure of the new manuscript is messy. Therefore both the tracked and the clean versions were provided to make reading as easy as possible. All the positions we mention in this response letter are according to the pages and lines of the clean version.

The revisions mainly consist of six parts:

1. Structure changes. First, we added a new Sect. 2.3 titled "Traditional formulation of data assimilation in the Bayesian theorem framework" in the revised manuscript and the corresponding text in Sect. 3.2 was removed, because this material is also a basic knowledge of our study. Second, according to editor's suggestion, we added examples to further explain the theoretical results. All the examples were organized in the new Sect. 3.4 titled "Examples: the stochastic radiative transfer equation (SRTE)".
2. Rewriting Sect.1 and Sect. 4. We rewritten the introduction to make the text more concise and comprehensible. The advances and scientific problem in our study were clearly presented. The title of Sect.4 was changed to "Discussion & Conclusions", and accordingly there are only two subsections left. In the revised manuscript, Sect.4.1 titled "Discussion" mainly focused on the necessity of methodology, the advantages and limitations of our study, and Sect.4.2 titled "Conclusions" restated the major results.
3. Revisions according to the comments from editor and referees. Detail information can be found in the point-by-point responses to editor and referees.
4. The formulation of scale transformation between the scales of initial state and forecasting operator was removed (In the end of Sect. 3.3). This formulation was presented in the previous manuscript. After lots of reconsiderations and discussion, we believed that this formulation is not necessary to be introduced in this paper because the main scientific problem is to formulate the error in the

update step, and the error caused by the scale difference between initial state and forecasting operator will not have a significant impact on the formulation of the update step.

5. To make one term in this paper only presents a single object, some similar terms in different fields were again clarified. Please check them according to the following form.

	<i>Terms in this paper</i>	<i>Explanation</i>
1	Measure	Measure is a term for measure theory, and Observation is an estimation of the value for a geophysical variable.
	Observation/Observe	“Measurement” have the similar meaning with “Observation’ but may be confused with “Measure”, so “Measurement” will not be introduced. Also, “measurement error” will be replaced by “instrument error”.
2	Footprint	A footprint is the observation footprint, and space is the measure space or state/observation space that a geophysical variable can evolve. We try to avoid the use of “field” or “region” to indicate the similar meaning.
	Space	
3	Variable	Variable is the geophysical variable or variable in state vector that can be observed by Earth observations. Parameter cannot be observed. In the mathematical formula, parameter also refer to the argument of a function. Random process indicates the stochastic process or Ito process only on the condition that a rigorous mathematical expression is involved.
	Parameter	
	Random process	

6. Other modifications that are based on the updated cognition on our study can also be found in the revised manuscript.

Responses to the Editor's Comments

We thank Dr. Olivier Talagrand once again for taking his valuable time to review our revised manuscript and given recognition of the improvement in our study. Based on all the comments from editor and referees, we have made major revisions of our manuscript. Here are the point-by-point responses to the new comments from editor.

1. *“Thinking of the general significance of your paper, I actually think as Editor that it would be useful to add a simple illustrative example, which would show explicitly how considering a stochastic scale transformation can impact the assimilation and the probability distribution that it produces. You have removed the section dealing with the Radiative Transfer Equation (RTE) which was included in the previous version of your paper, considering it was not closely tied to the other sections of the paper. I do not think it was explicit enough to show the impact on assimilation of a stochastic scale transformation, but it could possibly be used, with appropriate modifications, for that purpose. Without making it a condition for acceptance of the paper, I think an appropriate illustrative example, whether based on the RTE or not, would make the paper more understandable.”*

Response:

We agree with you. An example can make our study more understandable. With some modification, we first added an example to explain how the scales of system state and observation can obey the one-dimensional rule based on the scales presented in Figure 1. Then the formula of likelihood function was deduced to two different cases. The first one is that the observation is the same physical quantity as the state. In the second case, a nonlinear observation operator, i.e., a stochastic radiative transfer equation (SRTE) is used as another example. One thing should be noticed in the second case is that we only offered the Ito process-formed state, observation and observation operator. These functions can be used to further deduce the likelihood function according to Eq. (22).

Changes in the manuscript:

The example with Stochastic Radiative Transfer Equation in the previous manuscript is a little complicated. So we reduced its expression. Besides, using the scales presented in Figure 1, we also added some more simple examples. Both the simple examples and the example based on SRTE are introduced in Sect. 3.4 titled “Examples: the stochastic radiative transfer equation (SRTE)”.

2. P. 5, l. 14, and p. 8, l.5. I suggest to add indices i as follows

Response:

Yes, this makes the formula more clearly. We also add the indices i in the last paragraph of Sect. 3.1.

3. P. 6, l. 17. I presume you mean $\forall t1 > s1 \geq t2 > s2$, the increments $W(t1) - W(s1)$ and $W(t2) - W(s2)$ are independent.

Response:

Your suggestion makes the definition easy to understand, so we accept it and revised it in the new version.

4. P. 8, l. 15, ... a unit interval ... \rightarrow ... the unit square ...

Response:

Thanks for your suggestion. We also revised the other 3 words with the similar problem.

5. Probabilities are denoted $p()$ in some places (eq. 8, p. 10 for instance), and $P()$ in other places (p. 11, ll. 2-4). Please use consistent notations.

Response:

Thanks for your comment. We have made the revision accordingly.

Responses to Referee 1's Comments

We thank the anonymous Referee once again for taking his/her valuable time to review our revised manuscript and given recognition of the improvement in our study. Here are the point-by-point responses to the new suggestions.

1. Lebesgue measurable subset

1. On page 5: *"Generally a Lebesgue measure on R^n assumes that A is any subset of R^n ." This is not true. The σ -algebra \mathcal{F} in the definition of the triplet $(\Omega, \mathcal{F}, \mu)$ (line 6 on p5) that defines the measure does not include all subsets of R^n . There are subsets of R^n to which a Lebesgue measure cannot be consistently assigned. Construction of these so-called "unmeasurable sets" is described in the standard texts, as the authors know.*
2. Same paragraph: *instead of "Thus if A is any subset of R^n , one can collect..." I suggest "for any $A \in \mathcal{F}$, one can collect ..."*
3. p5, toward the bottom: *"The Lebesgue measure of any subset in R^n also coincides with its volume." Again, there are subsets of R^n to which a Lebesgue measure cannot be consistently assigned.*
4. p8: *"...any bounded closed domain A " As before, A must be Lebesgue measurable.*
5. p11 lines 7-8 *instead of " $A \in R^n$ ", you want " $A \in \mathcal{L}^2$ "*

Response:

Comment 1: Yes, you are right. The expression in here is not rigorous. Here A should be a Lebesgue measurable subset of R^n .

Comment 2, 3 and 5: Yes, these sentences will be more rigorous based on your suggestions. We have made revisions accordingly.

Comment 4: Based on my knowledge, a bounded closed domain is Lebesgue measurable. So, I think this sentence might be right. Besides, in this sentence, A indicates the observed space of a footprint-scale observation, which are Lebesgue measurable.

Changes in the manuscript:

Comment 1. This sentence was removed. To make this statement more clear, in the next paragraph, we added some sentences behind the Lebesgue σ -algebra: "The construction of the Lebesgue σ -algebra

is based on the Caratheodory condition (Bartle, 1995, definition 13.3). Fortunately, almost all of the observation footprints and model units are finite and closed, leading them to be Lebesgue measurable. This consequently ensures the Lebesgue measure m^n is a measure and the triple $(R^n, \mathcal{L}^n, m^n)$ is a measure spaces.”

Comment 2. In this paragraph, if the first sentence was removed, then we only talk about the outer measure. I think the definition is valid for the outer measure of an arbitrary subset of R^n , so maybe no change is needed in here.

Comment 3. This sentence was changed to “The Lebesgue measure of a Lebesgue measurable subset in R^n also coincides with its volume”.

Comment 5: “ $A \in R^2$ ” was changed to “ $A \in \mathcal{L}^2$ ”

2. The definition of scale

p8, lines 4-5: “We use Lebesgue measure on R^2 , i. e., $\mu_{iv}(A) = m^2(A) = \inf \left\{ \sum_{i=1}^{+\infty} I^2(A_i) \right\}$ where ...
 From a geometric perspective, the measure function refers to the shape of the subset, and the scale further indicates the size.” OK, I understand that, say, in Figure 1, you mean to say that disks C_1 , C_2 and C_3 have the same shape, but they have different scales because they are different sizes. But the Lebesgue measure is the area, and you have defined it as it is defined in the books. By your definition of m^2 , referring to Figure 1, $m^2(C_2) > m^2(C_1)$, but on line 10 you write “ $m^2_{C_1} = m^2_{C_2} = m^2_{C_3}$ because they are the same function.” I don't understand this at all. You mention a function you call f but it plays no part in the definition of m^2 . The statement “ $m^2_{C_1} = m^2_{C_2} = m^2_{C_3}$ ” is inconsistent with the stated definition of m^2 .

I'm guessing that C_1 , C_2 and C_3 are examples of footprints. If so, would you please say this explicitly?

What, exactly, are the functions associated with C_1 , C_2 and C_3 ?

Response:

I am so sorry that I had not make it much clearer. Your understanding is correct. We defined the scale as the Lebesgue measure with respect to the observation footprint, i. e., $s = m^2(A)$. Therefore, there are two elements in this definition, the Lebesgue measure function $m^2(\cdot)$ and the observation footprint A . The two scales, s_1 and s_2 , are equal to each other happens only when they have the same Lebesgue measure functions and the same observation footprints, say, $m^2_{s_1}(\cdot) = m^2_{s_2}(\cdot)$ and $A_{s_1} = A_{s_2}$. In Figure

1, $m_{C_1}^2$, $m_{C_2}^2$ and $m_{C_3}^2$ have the same Lebesgue measure functions, that is, they all refer to a disc footprint, so $m_{C_1}^2(\cdot) = m_{C_2}^2(\cdot) = m_{C_3}^2(\cdot)$. Your challenge is reasonable, here we made a mistake that we left out the note (\cdot) in this formula, which misled the readers to believe that the Lebesgue measures of disc measurements C_1 , C_2 and C_3 are equal in value. Here the Lebesgue measure functions $m_{C_1}^2(\cdot)$, $m_{C_2}^2(\cdot)$ and $m_{C_3}^2(\cdot)$ are associated with C_1 , C_2 and C_3 .

We introduced the real function f because it is involved in the general definitions of Lebesgue integral and Lebesgue integration by substitution. However, f is equal to 1 in the definitions of scale and scale transformation because the Lebesgue measure in R^2 is area. At p8, lines 9 and line 12, we have stated that $f \equiv 1$.

C_1 , C_2 and C_3 are indeed the examples of footprints, we have clarified this in the revised manuscript.

Changes in the manuscript:

The formula $m_{C_1}^2 = m_{C_2}^2 = m_{C_3}^2$ was changed to $m_{C_1}^2(\cdot) = m_{C_2}^2(\cdot) = m_{C_3}^2(\cdot)$. In addition, we revised Fig.1 and removed the diamond observation footprint $D1$ and $D2$ to make Fig.1 more concise. We also added some necessary words in this paragraph to make the explanation of Fig. 1 clearer. Please find the detailed information in the new manuscript.

Responses to Prof. van Leeuwen's Comments

We thank Prof. van Leeuwen once again for taking his valuable time to review our revised manuscript and provide us some very thoughtful and constructive comments. The point-by-point responses to the new suggestions were classified by what these comments focus on.

1. Scale

1. *Abstract: Define the spatial scale issue, Do you mean that models and observations define scale differently? Or do you mean that scale is not well defined in general, which then hampers model-observation comparisons and data assimilation?*
2. *P8, 11: scale is defined as an area, so has physical dimension m^2 . Typically scale is defined in terms of distance. That might be mentioned.*
3. *P8, 14: The standard scale depends on the units used, it is a different thing using meters of millimetres. Is this a useful definition? Or do you assume that all physical scales have been normalised? Please clarify.*

Response:

The understanding of scale issue and the definitions of scale and scale transformation play an important role in our work. If the observation footprints or model units are changed and associated variables present heterogeneities, the scale issue is inevitable. However, scale is not well understood currently. Defining scale in terms of distance is not adequate because distance is a one-dimensional quantity but scale generally refers to a two- or three-dimensional space. We believe the scale is related to the shape and size of observation footprint or model unit, so the Lebesgue measure on R^2 was used to define it.

Another reason for defining the scale is values for variables may change with scale in most of Earth observations and simulations. Scale may change much complicated (for example, from an irregular observation footprint to a square observation footprint), so how to quantify this change must be based on a rigorous definitions of scale and scale transformation.

Therefore, for *comment 1*, we think that scale is not well defined in general. Meanwhile, in the studies of model-observation comparisons or adding a new observation into data assimilation system, the

transformations between different scales result in remarkable error (also can be seen as a part of representativeness error). However, scale transformation was also not fully addressed.

For *comment 2*, as we have mentioned, distance is a one-dimensional quantity and unable to meet the definition of scale.

For *comment 3*, similarly, to use meters of millimetres is more reasonable for one-dimensional scale. But the scales that related to Earth observations, simulations and data assimilation should be regarded as two-dimensional or three-dimensional. We introduced unit area A_0 . The standard scale, defined as the area of A_0 , is significant because, on the one hand, it is a standard reference, by which one can make a quantitative comparison between different scales; on the other hand, the standard scale can be seen as a origin if we treat scale similarly to other physical quantities, such as time. Consequently we can develop the Brownian motion and stochastic calculus based on scale. Brownian motion and stochastic calculus both begin with the standard scale (for example, Lemma 1 and Eq. 12).

We introduced the standard scale, but that is not necessary to require all physical scales to be normalized by standard scale. However, if it is in need of a rigorous formulation of scale-dependent error, it's better to look to standard scale.

Changes in the manuscript:

We revised the manuscript accordingly as follows,

For comment 2, In paragraph 1, Sect. 1, we added some text after the first sentence: “Scale is traditionally defined in terms of distance, which is not adequate both because distance is a one-dimensional quantity but scale generally refers to a two- or three-dimensional space, and because scale may change much complicated (for example, form an irregular observation footprint to a square observation footprint).”.

For comment 3, detail explanation was appended after the definition of the standard scale: “The standard scale can be regarded as a basic unit of scale in two-dimensions. It presents a standard reference, by which one can make a quantitative comparison between different scales. The standard scale is also the origin of scales that let scales vary similarly to other physical quantities, such as time.” Additionally, after the paragraph on “Remark on Lemma 1”, we added “Therefore, beginning with the standard scale, the Brownian motion and stochastic calculus with respect to scale can be further developed.”

2. Language improvements

1. P1, 25: *'increases with the difference'*
2. P4, 16: *'the definition of scale uses measure theory'*
3. P6, 19: *typo in first equation: 'the first s1' should be 's2'.*
4. P10, 16: *'discovered' should perhaps be 'described the origin of'*
5. P13, 13 *Typo in equation (15): $X(s_X) = X_0 + \dots$*

Response:

Thanks for your comments on language improvements. We have made revisions accordingly except comment 1, which was replaced by the new introduction.

3. Variables and parameters

1. P1, 28: *parameters are chosen constants in time, variables are varying in time. So I guess the authors mean 'geophysical variables', or perhaps both. This runs through the whole manuscript.*
2. P2, 14: *Parameters cannot be collected. Values for variables can be collected. And they are not collected by Earth Observation techniques but by Earth observations.*
3. P11, 7: *the definition of 'variable' is unconventional and perhaps misleading. An element of state vector X is typically called a variable in the data-assimilation literature. Another name would be preferable.*
4. P11, 7: *What is the exact relation between state vector X and variable V ?*

Response:

Thanks for these comments. It is really helpful that variable and parameter are different and we indeed mixed them up. As you stated, the term "geophysical parameter" is usually regarded as a spatial and temporal constant, which cannot be observed. But variable changes with space and time, and its value can be observed.

For *comment 3*, the term "variable" is a common concept both in geophysics and mathematics. In measure theory, it indicates a real-value function on a probability space (Ω, \mathcal{F}, P) , but in data assimilation "variable" means a little differently. Here we in fact defined a general geophysical variable,

therefore, according to your advice, in the new manuscript, we use geophysical variable instead of “variable”.

For *comment 4*, in our study, the V is the stochastic version of an element of state vector X . We try to further introduce the mathematical definition of geophysical variable in the sense of measure theory (see page 10, line 6 and 7), and then study the Ito process-formed geophysical variable (see Eq. (9) and Eq. (10)).

Changes in the manuscript:

For *comment 1 and 2*, the term “*geophysical parameters*” was revised as “*geophysical variables*”, and “parameter” in the new manuscript only refers to “argument of a function”. How to distinguish the other similar terms was also presented in part 5, General reply of this response letter.

For *comment 3 and 4*, V was revised as “*geophysical variable*”. After the definition of geophysical variable, we also appended “In n -dimensional data assimilation, a geophysical variable V is related to an element of state vector X at a specific scale s and time t ”. Other related text will be revised in the new manuscript as well.

4. Stochastic differential equation with respect to scale

1. *P11, 23: The authors introduce a stochastic process in scale space that operates at a time instance, so time is a constant, and the variable changes due to a process in scale space. Is this interpretation correct, and if so, what is this process physically? This is a crucial point for me, and the point where I get lost.*
2. *P12: I understand this as a formal derivation of the stochastic process in scale space up to line 23 (but, as mentioned, I don't understand the physical process behind this).*
3. *P13, 4: What does this equation mean? That $X(s)$ changes due to changes in scale in the analysis step? If so, ϕ should also depend on Y , or at least on the scales in Y . Or does this equation describe a scale relation in X ? So how X depends on scale? If so, what is the stochastic forcing? Also, how to choose or estimate ϕ and σ in an application?*

Response:

Yes, this is the most important problem in our study. In paragraph 2, Sect. 1, we stated that the scale issue is related to spatial heterogeneities and dynamic process variations among different scales. That's to say, if the study region is not homogeneous, the values of a variable that observed

at the same place may present differently between large scale and small scale (for example, between the larger footprint $C2$ and the smaller footprint $C3$ in Figure 1). Some physical processes also vary among different scales. For example, except the ones we mentioned in paragraph 2, Sect. 1, ground water flow process is governed by Darcy's law at the macro-scale and by the Navier–Stokes equations at the pore-scale (Narsilio, et al. 2009). The validity of Planck's law also depends on the scale (Li, et al. 1999).

Therefore, to understand the physical processes behind the scale issue should both consider the heterogeneities and the changes of dynamic processes among different scales. However, based on associated literatures, most of them are not very clear, let alone to model these physical processes in a general theory study. Therefore, a sophisticated formulations is conceptualized in our manuscript but we believe this problem needs further study to make it more concrete.

For *comment 3*, we think that in the analysis step, time is invariant, but the state X in the state space is mapped to the observational space, i.e., the scale of X changes from s_X (scale of state space) to s_Y (scale of observation space). This process can be regarded as an Ito process of state with respect to scale, which can also be formulated by Eq. (15). Based on Eq. (12) (the integral form of Ito process), state X in the state space is $X(s_X) = X_0 + \int_{s_0}^{s_X} \varphi(u)du + \int_{s_0}^{s_X} \sigma(u)dW(u)$ and state in the observational space is $X(s_Y) = X_0 + \int_{s_0}^{s_Y} \varphi(u)du + \int_{s_0}^{s_Y} \sigma(u)dW(u)$. Apparently, X depends on scale.

In Eq. (13), as we stated, $\varphi(s)$ is the scale-dependent drift rate from standard scale to a specific scale, for example, s_X or s_Y . $\varphi(s)$ accords to the physical processes of state with respect to scale, which currently may be hard to be formulated. $\sigma(s)$ can be regarded as the stochastic perturbation with respect to scale, which is needed to be further investigated. However, if assuming the perturbation at the scales is totally random or Gaussian, then $\sigma(u) = 1$. In our study, only the simplest case, i.e., $\varphi = 0$ and $\sigma = 1$, was considered. This means only the Gaussian perturbation presents when the scale is changed. However, the result (Eq. (23)) is still complicated.

Changes in the manuscript:

For *comment 1* and *comment 2*, we added some necessary explanations after Eq. (10): “To formulate $\varphi(s)$ should consider both the spatial heterogeneities and physical process variations among different scales. However, neither of them is well understood in a general theory study.

Therefore $\varphi(s)$ is conceptualized in Eq. (10).” And after Eq. (13) and Eq. (14), we stated “ $\varphi(s)$ also implies the heterogeneities and physical processes from standard scale to a specific scale, which currently maybe hard to be formulate.”

For *comment 3*, we added text after Eq. (14): “Therefore, according to Eq. (12), a state is $X(s_X) = X_0 + \int_{s_0}^{s_X} \varphi(u)du + \int_{s_0}^{s_X} \sigma(u)dW(u)$ in the state space and is $X(s_Y) = X_0 + \int_{s_0}^{s_Y} \varphi(u)du + \int_{s_0}^{s_Y} \sigma(u)dW(u)$ in the observation space. These formulas prove that the value of state varies with the changes of scale.”

5. Data assimilation

1. P12: *I don't understand what assumption 1 means, 'in data assimilation' seems rather vague to me.*

Assumption 2 assumes that the model has a constant grid in space and time. That should perhaps me mentioned explicitly. So the grid is not finite element or finite volume, or adaptive in time.

Assumption 3: What does 'scale dependent' mean here? That the scale changes when applying Bayes Theorem?

2. P13, 9: *The 'authors state that 'Assumption 3 implies that the scales of the state and observation are invariant when observational information is added in the analysis step'. I don't see that, please clarify.*

Response:

For assumption 1, the scale transformation between state space and observation space of data assimilation obeys a one-dimensional rule. The one-dimensional rule is defined in Sect. 3.1 and can make scale change in a sense of geometrical similarity (for example, form a square observation footprint to a smaller square observation footprint). By this assumption, the formulations of scale transformation in data assimilation can be extremely reduced, but turn out the same conclusions with the one without any assumption about scale transformation.

For assumption 2, in the forecasting step, the model unit and state scale are both supposed to be same and invariant. There is no scale transformation in this step.

For assumption 3, the term “scale dependent” means that the state, observation and observation operator are all dependent on scale, and they can vary with scale. But when the Bayesian theorem is applied, the scales of state and observation are actually not changed, and the scale transformation

only involves in the process that mapping the state vector from state space to observation space. Then we get an estimation of observation $H(X(s_x))$ in the observation space which is related to the state $X(s_x)$ defined in the state space.

Changes in the manuscript:

We add some necessary text after all the assumptions, which is supposed to be more explicit to express our intention.

For assumption 1, we added “In assumption 1, the one-dimensional rule ensures that scale changes in a sense of geometrical similarity (for example, from a larger square observation footprint to a smaller square observation footprint, or from C_2 to C_3 as presented in Figure 1). Additionally, the formulations of scale transformation can be extremely reduced, but turn out the same conclusions with the one without any assumption about scale transformation.”

For assumption 2, it stated that “Assumption 2 indicates that the model unit and state scale are both supposed to be the same and invariant in space and time. So, there is no scale transformation in the forecasting step”.

For assumption 3, after Eq. (14), the sentence “Assumption 3 implies that the scales of the state and observation are invariant when observational information is added in the analysis step” missed some necessary information. This sentence was replaced by “The scale transformation only involves in the process that mapping the state vector from state space to observation space”.

3. P13, 22: *What does this equation mean? Is this the prior marginal pdf of X in scale space?*

I am lost as this page. Where is the measurement uncertainty? That should also appear somewhere in $p(y|x)$. I see only the scale part of the error. Is the assumption that the scale part is dominant? And again, what is this process in scale space that happens at observation time prior to calculating $p(y|x)$. Note that assumption 2 mentions explicitly that the scale of state and observation do not change before the observation time.

Response:

The equations mean the prior pdfs of state and observation with respect to scale in state space and observation space, respectively. They are different from the pdfs with respect to time, because their means are equal to the value at the standard scale and variances depends on the differences

between standard scale and the state space or observation space. These two prior pdfs are introduced into the Bayesian theorem that reformulated by scale.

We first clarify that the measurement error has little impact on the error caused by scale transformation (it will be also clarified in the new manuscript, see paragraph 3, Sect. 1). Actually, in some data assimilation literatures, for example, Lorenc (1995) and van Leeuwen (2014), the observational error is composed of two individual errors on the Gaussian assumption: measurement error and representativeness error (the error caused by scale transformation is the major component of representativeness error). Both of them are equally important. Meanwhile measurement error is independent with scale transformation. Therefore, it is not necessary to introduce the measurement error when formulate the scale transformation in data assimilation, but that should be stated in the manuscript.

Assumption 2 and 3 mentioned that both model unit and scale of state do not change before the analysis step, but if mapping the state vector from state space to observation space, the scale transformation occurs, and the state X , observation Y and observation operator $H(\cdot)$ are all dependent on scale. Then the error caused by scale transformation can be formulated with $Y(s_Y) - H(s_X, X(s_X))$.

Changes in the manuscript:

For the equation, explanation was appended: “Eq. (17) and Eq. (18) are the prior PDFs of state and observation with respect to scale in state space and observation space, respectively. Compared with the PDFs with respect to time, their expectations are equal to the value at the standard scale, and the variances depend on the differences between the standard scale and the scale in state or observation space. These two prior PDFs are introduced into the Bayesian theorem that reformulated by scale. ”

For the issue on measurement error, we appended some necessary text in paragraph 3, Sect. 1 to clarify that the measurement error is not necessary to introduce in this paper: “The representativeness error and instrument error make up the observation error of data assimilation. Under the Gaussian assumption, they are independent of each other (Lorenc, 1995; van Leeuwen, 2014). This study will not introduce the instrument error when formulate the scale transformation in data assimilation.” Here “measurement error” is replaced by “instrument error” to avoid being confused with “measure”.

6. Data assimilation and stochastic calculus

1. *My main issue is that I don't understand the stochastic equation in scale space, neither where it comes from nor how it helps solve the representation error problem.*
2. *P13, 28: I'm not sure what happens here. Why is there a stochastic equation for H ? Why not use the pdf of X directly to find the uncertainty in H ?*
3. *I seem to miss something fundamental related to the stochastic equation in scale space and hope the authors can clarify that. What I would understand is a transformation from state to observation space, which might be modelled by a stochastic process. The rationale for that is that the state and/or observation subgrid processes are unknown and treated randomly. For that, only equations (19) and (20) are needed, although I still don't see why a stochastic differential equation is used to model this transformation, why not define it directly as a nonlinear function from state to observation space? Then the likelihood can be formulated.*

Response:

Thanks for your comments. We believe it is worth to use the stochastic approach to solve the representation error problem based on the following reasons:

First, using the Ito process and stochastic calculus is essentially consistent with the definitions of scale, scale transformation and geophysical variable. In Sect. 3.1, the scale was defined as the Lebesgue measure with respect to the observation footprint, scale transformation presents the change between two different scales, and geophysical variable is a real mapping function on R . All of them are associated with corresponding measure spaces $(\Omega, \mathcal{F}, \mu)$. Therefore, it is natural to regard the state space and observation space as two different measure spaces, respectively, and each element of state (or observation) vector can be seen as a geophysical variable that mapping the state (or observation) measure space onto R . Correspondingly, stochastic calculus, which is defined for integrals of random processes with respect to random processes, was adopted.

Second, understanding the scale transformation between different scales can be further improved by stochastic calculus. As we stated, to map the state vector from state space to observation space should consider the transformation of scales, heterogeneities and physical processes. This can be illustrated by Eq. (12).

$$V(s) = V_0 + \int_{s_0}^s \varphi(u)du + \int_{s_0}^s \sigma(u)dW(u)$$

Eq. (12) is integrated from s_0 to s , which presents the scale transformation. The integral term $\varphi(u)$ combines physical processes with heterogeneities.

Third, using stochastic calculus can formulate the scale-dependent errors. The results are presented in Eq. (23) and Eq. (25), which are derivate from Eq. (20). Therefore, we believe that all the equations are needed.

Further, compared with nonlinear function, the stochastic equation can offer a more general framework for scale issue and representativeness error. For example, we used the one-dimensional rule to simplify the scale transformation. However, that is the simplest situation. If the scale changes randomly, say, from an irregular footprint to another irregular footprint, the stochastic equation can offer a double-integral or multiple-integral to further study the scale issue, such as

$$V(x, y) = V_0 + \iint \varphi(x, y)dxdy + \iint \sigma(x, y)dW_1(x)dW_2(y)$$

where $W_1(x)$ and $W_2(y)$ are two independent Brownian Motion.

Changes in the manuscript:

To explicitly explain why using stochastic calculus is benefit to understanding the scale transformation and the representation error, we reorganized the section of “Discussion”. In Sect. 4.1 titled “Discussion”, we restated the advantages of this study:

“The reasons that the methodology focuses on a stochastic framework are: First, the stochastic data assimilation framework is essentially consistent with the conceptions of scale and scale transformation. Both of them are associated with corresponding measure spaces $(\Omega, \mathcal{F}, \mu)$. Therefore, it is natural to regard the state space and observation space as two different measure spaces, respectively, and each element of state (or observation) vector can be seen as a geophysical variable that mapping the state (or observation) measure space onto R . Correspondingly, as the integrals of random processes with respect to random processes, stochastic calculus was adopted ultimately. Second, using stochastic calculus can also formulate the errors caused by scale transformations. The study proceeds with and improves the understanding of representativeness error in terms of scale. Results did not only prove the conventional point that the uncertainties of these errors mainly depend on the differences between scales, but indicated that the first-order

differential of the nonlinear observation operator should also be incorporated in representativeness error. Last, stochastic calculus can be extended to meet a general scale transformation and formulate corresponding representativeness error. This was unattainable in previous work. For example, if the scale changes randomly, say, from an irregular footprint to another irregular footprint, the stochastic equation can offer a multiple-integral to present this kind of a scale transformation, such as $V(x, y) = V_0 + \int_{y_0}^y \int_{x_0}^x \varphi(x, y) dx dy + \int_{y_0}^y \int_{x_0}^x \sigma(x, y) dW_1(x) dW_2(y)$, where $W_1(x)$ and $W_2(y)$ are two independent Brownian Motion.”

7. Other problems

1. P2, 15: *The ‘therefore’ is not a logical consequence. What does ‘them’ refer to?*

Response:

Here the term “them” refer to “forcing data and system states”.

Changes in the manuscript:

We have rewritten the Sect. 1 titled “introduction”. According to your comments, this sentence was changed to “Geophysical data are typically observed by various Earth observations, therefore to update the observation data in a data assimilation system may result in scale transformations between observation space and system state space”.

2. P2, 16: *The fact that observation operators are nonlinear and complex has in principle nothing to do with mismatches between model units and observation footprints. The logical connection is not clear.*

Changes in the manuscript:

We realized this problem and have made some revision according to your comments. Related text will be changed to “If observation operator is strongly nonlinear and complex, errors caused by scale transformation is even more serious”.

3. P2, 16: *Model units should be defined a bit better (I know this is difficult because it is unclear what scales a model with a certain grid box size represents.)*

Response:

We agree with you that to well define the model units is a little difficult, and its definition may vary with branches of geoscience. So in the definition of scale (page 8, line 10, previous manuscript), we have exemplified the model units.

Changes in the manuscript:

According to the previous problem, the term “model units” was removed in this sentence.

We further defined the model unit in paragraph 1, Sect. 2 as “The model unit is a specified subspace where a geophysical variable evolves in the model space. It could be a point, a rectangular grid, or an irregular unit such as a response unit (watershed, landscape patch and so on)”.

4. *P3, 4: Van Leeuwen also discussed the spatial and temporal resolution differences as giving rise to representation error, and Lorenc 1986 was the first to discuss the observation operator as source of an extra error on top of the measurement error in data assimilation.*

Response:

Thanks for the guidance. The related references have been cited in the revised manuscript.

Changes in the manuscript:

Related text will be changed to “An important concept that is related to scale transformation in data assimilation is “representativeness error”, which is associated with the inconsistency in spatial and temporal resolutions between states, observations and operators (Lorenc, 1986; Janjić and Cohn, 2006; van Leeuwen, 2014; Hodyss and Nichols, 2015)”.

New reference:

Lorenc A C. Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 1986, 112(474):1177–1194.

5. *P3, 7: ‘According to the above...’ The land surface dynamical processes have not been discussed, so the logical link is missing.*

Response:

In the new manuscript, the Sect. 1 was rewrote, so the related text about land surface dynamics is removed.

6. P4, 7: *Data assimilation does not necessarily result in first and second moments, the solution is the full pdf. For instance, data assimilation can describe multimodal pdfs.*

Response:

Thanks for the guidance. In the new manuscript, the Sect. 1 was rewrote, and the incorrect sentences was also removed.

7. P7: *I assume $\phi(t)$ is deterministic in eq (1), it is the drift term, so 'transition probability' is perhaps misleading.*

Response:

We agree with you. The term '*transition probability*' will be changed to "drift rate" in the new version. Correspondingly, the term '*volatility*' is also changed to "volatility rate".

8. P11, 15: *'the variable varies with scale because of the scale issue' is unnecessary vague. Perhaps remove this part of the sentence?*

Response:

The unnecessary vague part will be removed in the new version.

9. P11, 31: *What are the exact relations between M and p , and η and q ?*

Response:

As we stated, Eq. (6) (In the revised manuscript, it changed to Eq. (5)) is a discrete-time forecasting system, and Eq. (9) is a continuous-time Ito process that obtaining the prediction of state. So $p(t)$ can be regarded as a continuous-time version of M that obtaining the state on the interval $[0, t]$. η is the model error, and $q(t)$ can be seen as the error caused by evolution of time. So $q(t)$ is one part of η .

Generally $q(t)dW(t)$ is Gaussian (Apte, et al., 2007) and can hardly be used to study the representativeness error. That is also one of the reasons that we define the scale and formulate stochastic processes with respect to scale.

Changes in the manuscript:

We removed this paragraph, and made the new paragraph instead:

"Eq. (9) can be regarded as a continuous-time version of Eq. (5), i.e., to estimate the state is equal to the integral of Eq. (9) over a time interval. Here $p(t)$ indicates the physical process with respect to

time, and $q(t)$ is the error only caused by the evolution of time, so model error η in Eq. (5) contains more parts than $q(t)$. Eq. (10) implies that the value and variance of a geophysical variable may change if the scale changes. ... Particularly, if the study region is homogeneous, then the values of a variable that observed at the same place are identical between large scale and fine scale, and $\varphi(s)$ can be left out. $\sigma(s)$ is the error caused by the scale transformation.”

References

- Li, X. W., Wang, J. D., and Strahler, A. H.: Scale effect of planck’s law over nonisothermal blackbody surface, *Sci. China*, 42, 652-656, 1999
- Narsilio, G. A., Buzzi, O., Fityus, S., et al.: Upscaling of Navier–Stokes equations in porous media: Theoretical, numerical and experimental approach. *Computers & Geotechnics*, 36,1200–1206, 2009.
- Lorenc, A.C.: *Atmospheric Data Assimilation*, Meteorological Office, Bracknell, 1995.
- van Leeuwen, P. J.: Representation errors and retrievals in linear and nonlinear data assimilation, *Q. J. Roy. Meteor. Soc.*, 141, 1612-1623, doi: 10.1002/qj.2464, 2014.
- Apte, A., Hairer, M., Stuart, A. M. and Voss, J.: Sampling the posterior: An approach to non-Gaussian data assimilation, *Physica D*, 230, 50-64, doi: 10.1016/j.physd.2006.06.009, 2007.

Formulation of Scale Transformation in a Stochastic Data Assimilation Framework

Feng Liu^{1,3} and Xin Li^{1,2,3}

5 Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, 730000, China

²Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences

³University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence to: Xin Li (lixin@lzb.ac.cn)

10 **Abstract:** Understanding of the errors caused by spatial scale transformation in Earth observations and simulations requires a rigorous definition of scale. These errors are also an important component of representativeness errors in data assimilation. Several relevant studies have been conducted, but the theorization of the scale associated representativeness errors is still not well developed. We addressed these problems by reformulating the data assimilation framework using measure theory and stochastic calculus. First, the measure theory is used to propose that the spatial scale is the Lebesgue measure with respect to
15 the observation footprint or model unit, and the Lebesgue integration by substitution is used to describe the scale transformation. Second, a scale-dependent geophysical variable is defined to consider the heterogeneities and dynamic processes. Finally, the structures of scale-dependent errors are studied in the Bayesian framework of data assimilation based on stochastic calculus. All the results were presented on the condition that scale is one-dimensional, and the variations in these errors depend on the difference between scales. This new formulation provides a more general framework to understand the representativeness error
20 in a nonlinear and stochastic sense and is a promising way to address the spatial scale issue.

1 Introduction

The spatial scale in Earth observations and simulations refers to the observation footprint or model unit in which a geophysical variable is observed or modelled (scale is used below to abbreviate spatial scale). Scale is traditionally defined in terms of distance, which is not adequate both because distance is a one-dimensional quantity but scale generally refers to a
25 two- or three-dimensional space, and because scale may change much complicated (for example, from an irregular observation footprint to a square observation footprint). Generally, scale is not explicitly expressed in the dynamics of a geophysical variable, partially because a rigorous definition of scale is difficult to find, except for an intuitive conception (Goodchild and

Proctor, 1997) and certain qualitative classifications of scale (Vereecken et al., 2007). This reflects the complexity of scale and requires a more rigorous mathematical conceptualization of scale.

Scale transformation of a geophysical variable may result in significant errors (Famiglietti et al., 2008; Crow et al., 2012; Gruber et al., 2013; Hakuba et al., 2013; Huang et al., 2016; Li and Liu, 2016; Ran et al., 2016). These errors are mainly caused
5 by the strong spatial heterogeneities (Miralles et al., 2010; Li, 2014) and irregularities (Atkinson and Tate, 2000) that are associated with geophysical variables across different scales, and are also closely related to dynamic variations, e. g., hydrological (Giménez et al., 1999; Vereecken et al., 2007; Merz et al., 2009; Narsilio, et al. 2009), soil (Ryu and Famiglietti, 2006; Lin et al., 2010) and ecological (Wiens, 1989) processes. How to develop mathematical tools to elucidate the scale transformation has yet to be fully addressed.

10 Data assimilation could be an ideal tool to explore the scale transformation because it presents a unified and generalized framework in Earth system modelling and observation (Talagrand, 1997). Geophysical data are typically observed by various Earth observations, therefore to update the observation data in a data assimilation system may result in scale transformations between observation space and system state space. If observation operator is strongly nonlinear and complex, errors caused by scale transformation is even more serious (Li, 2014). An important concept that is related to scale transformation in data
15 assimilation is “representativeness error”, which is associated with the inconsistency in spatial and temporal resolutions between states, observations and operators (Lorenç, 1986; Janjić and Cohn, 2006; van Leeuwen, 2014; Hodyss and Nichols, 2015), and the missing physical information that is related to numerical operator compared to the ideal operator (van Leeuwen, 2014), such as the discretization of a continuum model or neglect of necessary physical processes. The representativeness error and instrument error make up the observation error of data assimilation. Under the Gaussian assumption, they are independent
20 of each other (Lorenç, 1995; van Leeuwen, 2014). This study will not introduce the instrument error when formulate the scale transformation in data assimilation.

Recently, approaches have been developed to assess representativeness error. Janjić and Cohn (2006) studied representativeness error by treating system state as the sum of resolved and unresolved portions. Bocquet et al. (2011) used a pair of operators, namely, restriction and prolongation, to connect the relationship between the finest regular scale and a coarse
25 scale, and determined representativeness error using a multi-scale data assimilation framework. van Leeuwen (2014)

considered two complicated cases, i.e., conducting the observation vector in a finer resolution compared with system state vector and assimilating the retrieved variables. Their solutions were formulated using an agent in observation or state space, and a particle filter was proposed to treat the nonlinear relationship between observations, states and retrieved values. Hodyss and Nichols (2015) also estimated the representativeness error by investigating the difference between the truth and the
5 inaccurate value that is generated by forecasting model.

Although these approaches explored the structure of representativeness error and offered various solutions, improvements are still necessary to investigate what is the exact expression of errors caused by scale transformation in data assimilation. The authors believe that these approaches are optimal in linear systems, but may not be suitable when observations are heterogeneous and sparse, or when operators are nonlinear between states and observations, although the general equations in
10 the nonlinear case were given. Without taking heterogeneities and nonlinear operators into account, representativeness error cannot be fully understood. However, heterogeneity varies depending on the situation and is difficult to be formulated in a general theory study.

Data assimilation studies based on stochastic processes (Miller, 2007; Apte et al., 2007) or stochastic dynamic model (Miller et al., 1999; Eyink et al., 2004) have been proposed recently. Compared to deterministic models, stochastic data assimilation
15 is more applicable in an integrated and time-continuous theoretical study (Bocquet et al., 2010), and creates an infinite sampling space of the system state (Apte et al., 2007). Although the theorems of calculus that are based on stochastic processes (or stochastic calculus) are different from those of ordinary calculus, these advantages suggest that stochastic data assimilation offers a more general framework to study scale transformation.

We attempt to explore the mathematic definitions of scale and scale transformation, and then formulate the errors caused by
20 scale transformation in a general theory study on stochastic data assimilation. The next section introduces the basic concepts and theorems of measure theory, stochastic calculus and data assimilation. In Sect. 3, we present the definitions of scale and scale transformation. The posterior probability of system state was also reformulated by scale transformation in a stochastic data assimilation framework. In the final section, the contributions and deficiencies of this study were discussed.

2 Basic knowledge

25 The scale greatly depends on the geometric features of a certain observation footprint or model unit. The model unit is a specified subspace where a geophysical variable evolves in the model space. It could be a point, a rectangular grid, or an irregular unit such as a response unit (watershed, landscape patch and so on). We offer a solution in which the definition of

scale uses measure theory and the expression of geophysical variable as a stochastic process uses stochastic calculus. Therefore, we first introduce several basic concepts of measure theory and stochastic calculus.

2.1 Measure theory

Let Ω be an arbitrary nonempty space. \mathcal{F} is a **σ -algebra** (or **σ -field**) of subsets of Ω that satisfies the following conditions:

- 5 (i) $\Omega \in \mathcal{F}$, and the empty set $\Phi \in \mathcal{F}$;
(ii) $A \in \mathcal{F}$ implies that its complementary set $A^c \in \mathcal{F}$;
(iii) $A_1, A_2, \dots \in \mathcal{F}$ implies their union $A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

A set function μ of \mathcal{F} is called a **measure** if it satisfies the following conditions:

- (1) $\mu(A) \in [0, \infty)$ and $\mu(\Phi) = 0$;
10 (2) If $A_1, A_2, \dots \in \mathcal{F}$ is any disjoint sequence and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, μ is countably additive such that $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$.

If $\mu(\Omega) = 1$, μ can be replaced by the probability measure p , and if μ is finite, p can be calculated as $p(A) = \mu(A)/\mu(\Omega)$.

The triples $(\Omega, \mathcal{F}, \mu)$ and (Ω, \mathcal{F}, p) are the **measure space** and **probability measure space**, respectively.

Let Ω be the set of real numbers R and σ -algebra \mathcal{B} be **Borel algebra**, which is generated by all closed intervals in R . Then
15 $\forall A = [a, b] \in \mathcal{B}$, a **Lebesgue measure** on R is defined as $I(A) = b - a$. Intuitively, the Lebesgue measure on R coincides with length.

An **n-dimensional Lebesgue volume** is defined to measure the standard volumes of subsets in R^n based on $I^n(A) = \prod_{k=1}^n (b_k - a_k)$, where $A = [x: a_k \leq x_k \leq b_k, k = 1, 2, \dots, n]$ is an n-dimensional regular cell in R^n . The n-dimensional Lebesgue volume is an ordinary volume, such as length (n=1), area (n=2) and volume (n=3).

20 Next, the **outer measure** is defined as $m^n(A) = \inf\{\sum_{i=1}^{+\infty} I^n(A_i)\}$, where $\inf\{\cdot\}$ is the infimum, $A_i = [x: a_{i,k} \leq x_k \leq b_{i,k}, k = 1, 2, \dots, n]$ is the n-dimensional regular cell in R^n , and $A \subseteq \bigcup_{i=1}^{+\infty} A_i$. Thus, if A is any subset of R^n , one can collect many sets of n-dimensional regular cells $\{A_i\}$ to cover A . Among them, the outer measure denotes the set whose union has the smallest n-dimensional Lebesgue volume.

To match the two conditions of a measure, one can define the outer measure m^n as a **Lebesgue measure** on measure spaces $(R^n, \mathcal{L}^n, m^n)$, where \mathcal{L}^n is the **Lebesgue σ -algebra** of R^n . The construction of the Lebesgue σ -algebra is based on the Caratheodory condition (Bartle, 1995, definition 13.3). Fortunately, almost all of the observation footprints and model units are finite and closed, leading them to be Lebesgue measurable. This consequently ensures the Lebesgue measure m^n is a measure and the triple $(R^n, \mathcal{L}^n, m^n)$ is a measure spaces. The Lebesgue measure of a Lebesgue measurable subset in R^n also coincides with its volume.

The n-dimensional **Lebesgue integral** in $(R^n, \mathcal{L}^n, m^n)$ is $\int f dm^n$, where f is a real function on R^n . The Lebesgue integral can be further denoted by $\int f dm^n = \int f(x) dx$, where $x \in R^n$ and $x = (x_1, \dots, x_n)$.

In the two-dimensional case ($n = 2$), the Lebesgue integral is

$$\iint_A f(x_1, x_2) dx_1 dx_2,$$

where $A \in \mathcal{L}^2$. Next, we consider the **Lebesgue integration by substitution** on R^2 . Let $T(x_1, x_2) = [t_1(x_1, x_2), t_2(x_1, x_2)] = [y_1, y_2]$ be a one-to-one mapping of a subset X onto another subset Y on R^2 . Assuming that T is continuous and has a continuous partial derivative matrix $T_x = \begin{pmatrix} \partial t_1 / \partial x_1 & \partial t_1 / \partial x_2 \\ \partial t_2 / \partial x_1 & \partial t_2 / \partial x_2 \end{pmatrix}$, then

$$\iint_Y f(y_1, y_2) dy_1 dy_2 = \iint_X f(T(x_1, x_2)) |J(x_1, x_2)| dx_1 dx_2,$$

where the Jacobian determinant $|J(x_1, x_2)| = |\det T_x| = \begin{vmatrix} \partial t_1 / \partial x_1 & \partial t_1 / \partial x_2 \\ \partial t_2 / \partial x_1 & \partial t_2 / \partial x_2 \end{vmatrix}$. If T is linear, the integral reduces to

$$\iint_Y f(y_1, y_2) dy_1 dy_2 = |J(x_1, x_2)| \iint_X f(T(x_1, x_2)) dx_1 dx_2.$$

By so, any observation footprint or model unit can be regarded as a Lebesgue measurable subset in a two-dimensional space R^2 .

Additional details regarding measure theory can be found in the literature (for example, Billingsley, 1986; Bartle, 1995).

2.2 Stochastic calculus

We then introduce some necessary concepts and theorems of stochastic calculus. All the classic theorems are introduced without proofs; their detailed derivations can be found in the literature (Itô, 1944; Karatzas et al., 1991; Shreve, 2005).

Stochastic calculus is defined for ordinary integrals with respect to stochastic processes. One of the simplest stochastic processes is **Brownian motion**. The Brownian motion W that is defined on a probability measure space (Ω, F, p) is characterized as follows:

- 1) $W(0) = 0$.
- 2) $\forall t_1 > s_1 \geq t_2 > s_2 > 0$, the increments $W(t_1) - W(s_1)$ and $W(t_2) - W(s_2)$ are independent.
- 3) $\forall t > s \geq 0$, $W(t) - W(s) \sim N(0, t - s)$.

The last two conditions represent that $\forall t_2 > s_2 > t_1 > s_1 \geq 0$, $W(t_2) - W(s_2)$ and $W(t_1) - W(s_1)$ are independent Gaussian random variables. Additionally, W is related to the probability measure p .

Stochastic calculus based on Brownian motion produces **Ito process**. The differential form of the time-dependent **Ito process** is

$$dI = \varphi(t)dt + \sigma(t)dW(t), \quad (1)$$

where $\varphi(t)$, $\sigma(t)$ and $W(t)$ are the drift rate, volatility rate and Brownian motion, respectively. The integral form of Eq. (1) is

$$I(t) = I(0) + \int_0^t \varphi(u)du + \int_0^t \sigma(u)dW(u). \quad (2)$$

Theorem 1: For any Ito process defined as in Eq. (1), the **quadratic variation** that is accumulated on the interval $[0, t]$ is

$$[I, I](t) = \int_0^t \sigma^2(u)du, \quad (3)$$

and the **drift** of Eq. (1) is $I(0) + \int_0^t \varphi(u)du$.

As distinguishing features of stochastic calculus, quadratic variation and drift can be regarded as stochastic versions of the variance and expectation, respectively. That is, the variance and expectation are instances of their random variable stochastic counterparts within a certain integral path. Therefore, rather than being constants, quadratic variation and drift are given in terms of probability.

Theorem 2 (Ito's Lemma): If the partial derivatives of function $f(u, x)$, viz. $f_u(u, x)$, $f_x(u, x)$ and $f_{xx}(u, x)$ are defined and continuous, and if $t \geq 0$, we have

$$f(t, x(t)) = f(0, x(0)) + \int_0^t f_u(u, x(u))du + \int_0^t f_x(u, x(u))\sigma(u)dW(u) + \int_0^t f_x(u, x(u))\varphi(u)du + \frac{1}{2} \int_0^t f_{xx}(u, x(u))\sigma^2(u)du. \quad (4)$$

Ito's Lemma is typically used to build the differential of a stochastic model with Ito processes. In this study, Ito's Lemma is applied to study the scale-dependent relationship between the observation and state, and the errors caused by scale transformation.

2.3 Traditional formulation of data assimilation in the Bayesian theorem framework

We use the well accepted Bayesian theorem of data assimilation (Lorenz, 1995; van Leeuwen, 2015) to investigate its time- and scale-dependent errors. State and observation are first assumed to be one-dimensional. In Sect. 3.5, the results are extended to n -dimensional state vectors and observation vectors.

5 A nonlinear forecasting system can be described by

$$X(t_k) = M_{k-1:k}(X(t_{k-1})) + \eta(t_k), \quad (5)$$

where $M_{k-1:k}(\cdot)$, $X(t_k)$ and $\eta(t_k)$ represent a nonlinear forecasting operator that transmits the state from the discrete time $k - 1$ to k , the state with prior probability distribution function (PDF) $p(X)$, and the model error at time k , respectively.

If a new observation is available at time k , the observation system is given by

10
$$Y^o(t_k) = H_k(X(t_k)) + \varepsilon(t_k), \quad (6)$$

where $H_k(\cdot)$, $Y^o(t_k)$ and $\varepsilon(t_k)$ represent the nonlinear observation operator, true observation with prior PDF $p(Y)$, and observation error at time k , respectively.

Previous studies (e.g., Janjić and Cohn, 2006; Bocquet et al. 2011) described the origin of the components of $\varepsilon(t_k)$ and $\eta(t_k)$, such as white noise, the discretization error of a continuum model, the errors that are caused by missing physical processes, and scale-dependent bias. In this study, we assume that both forecasting and observation operators are perfect models, so errors that are caused by missing physical processes are discarded.

According to Bayesian theory, the posterior PDF of the state based on the addition of a new observation into the system is

$$p(X|Y) = p(Y|X)p(X)/p(Y), \quad (7)$$

where $p(X|Y)$ is the posterior PDF that presents the PDF value of state X given an available observation Y . $p(Y|X)$ is a likelihood function, which is the probability that an observation is Y given a state X . $p(X)$ and $p(Y)$ are the prior PDF values of the state and observation, respectively. Here, $p(X)$ is supposed to be known and $p(Y)$ is a normalisation constant (van Leeuwen, 2014). The aim of data assimilation is equivalent to finding the posterior PDF $p(X|Y)$.

3 Reformulation of scale transformation in data assimilation framework

3.1 Definition of scale

25 We define the scale based on the measure theory that was introduced in Sect. 2. The relationship between Lebesgue measure in $(R^2, \mathcal{L}^2, m^2)$ and scale is firstly introduced by the following measures of Earth observations.

- (i) Measure of a single point observation: When the observation footprint is very small and homogeneous, we assume that its footprint approaches zero and its measure is accordingly zero under the condition of the Lebesgue measure.

(ii) Measure in a line: The measure is a one-dimensional Lebesgue measure.

(iii) Measure of a rectangular pixel (for example, remote sensing observation): $\forall A = [x: a_k \leq x_k \leq b_k, k = 1, 2]$, it is a two-dimensional Lebesgue volume, i.e., $\mu_{iii}(A) = I^2(A) = \prod_{k=1}^2 (b_k - a_k)$.

(iv) Measure of a footprint-scale observation: The footprint is any bounded closed domain A , which is not necessary to be regular rectangles, but as circles or ellipses. We use Lebesgue measure on R^2 , i.e., $\mu_{iv}(A) = m^2(A) = \inf \left\{ \sum_{i=1}^{+\infty} I^2(A_i) \right\}$, where $A_i = [x: a_{i,k} \leq x_k \leq b_{i,k}, k = 1, 2]$ and $A \subseteq \bigcup_{i=1}^{+\infty} A_i$. Obviously, measure (i)~(iii) are the special cases of the measure of a footprint-scale observation.

Actually, all the above measures mainly depend on the shape and size of A . The Lebesgue measure on R^2 coincides with the area, so the Lebesgue integral of $\mu_{iv}(A)$ is $\iint_A dx_1 dx_2$, where the real function $f \equiv 1$.

Now, we can generalize the above examples by defining the **scale** as the Lebesgue measure with respect to the observation footprint. This definition can also be extended to a certain model unit. Thus, for any subset $A \in \mathcal{L}^2$, the scale is $s = m^2(A) = \iint_A dx_1 dx_2$, where the real function $f \equiv 1$. From a geometric perspective, the measure function $m^2(\cdot)$ refers to the shape of the subset, and the scale further indicates its size.

We represent the scale as s , and let $s_0 = m_0^2(A_0) = \iint_{A_0} dx_1 dx_2 = 1$ be the **standard scale**, where $A_0 = [x: 0 \leq x_k \leq 1, k = 1, 2]$ is the unit square in R^2 . The standard scale can be regarded as a basic unit of scale. It presents a standard reference, by which one can make a quantitative comparison between different scales. The standard scale is also the origin of scales that let scales vary similarly to other physical quantities, such as time.

We can further define **scale transformation**. For $\forall A_1, A_2 \in \mathcal{L}^2$, if there are two different scales, $s_1 = m^2(A_1) = \iint_{A_1} dx_1 dx_2$ and $s_2 = m^2(A_2) = \iint_{A_2} dy_1 dy_2$, then we can obtain $s_2 = \iint_{A_2} dy_1 dy_2 = \iint_{A_1} |J(x_1, x_2)| dx_1 dx_2$ based on Lebesgue integration by substitution, where the Jacobian matrix $J(x_1, x_2)$ represents the geometric transformation from A_1 to A_2 . In particular, if $J(x_1, x_2) = \text{diag}(\xi, \xi), \xi \in R$, which also indicates that the geometric transformation is linear, then the following expression is valid based on Lebesgue integration by substitution:

$$s_2 = |J(x_1, x_2)| \iint_{A_1} dx_1 dx_2 = \xi^2 s_1, \quad (8)$$

where s_1 and s_2 represent the change of **one-dimensional rule**.

If two scales follow the one-dimensional rule, they are geometrically similar. This rule simplifies scale as a one-dimensional variable that corresponds to the scale transformations between most remote sensing images with various spatial resolutions. For example, $\forall A = [x: a \leq x_k \leq b, k = 1,2]$, where A and the unit square A_0 are geometrically similar, and the scale $s = \mu_{iii}(A)$ can be expressed by the one-dimensional rule of scale transformation: $s = \mu_{iii}(A) = |J(x_1, x_2)| \iint_{A_0} dx_1 dx_2 =$

5 $(b - a)^2 s_0$. For another example, let $s = \iint_A dy_1 dy_2$ be the scale of a disc footprint A with radius r . The mapping function between A and A_0 is $T(x_1, x_2) = [rx_1 \cos(2\pi x_2), rx_1 \sin(2\pi x_2); 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1] = [y_1, y_2]$, and the Jacobian

determinant $|J(x_1, x_2)| = \begin{vmatrix} r \cos(2\pi x_2) & -2\pi r x_1 \sin(2\pi x_2) \\ r \sin(2\pi x_2) & 2\pi r x_1 \cos(2\pi x_2) \end{vmatrix} = 2\pi r^2 x_1$. Therefore, $s = \iint_A dy_1 dy_2 =$

$\iint_{A_0} |J(x_1, x_2)| dx_1 dx_2 = \pi r^2 s_0$, which is equal to its area. However, s_0 and s do not obey one-dimensional rule because the Jacobian matrix is not diagonal.

10 The Layer 1 in Figure 1 shows the relationship between the Lebesgue measure and scale. The measure space $\Omega = [x: 0 \leq x_k \leq 4, k = 1,2]$ is regularly divided by the unit square A_0 . Let scales $s_{C1} = m_{C1}^2(C1)$, $s_{C2} = m_{C2}^2(C2)$ and $s_{C3} = m_{C3}^2(C3)$ be the Lebesgue measures of disc observation footprints $C1, C2$ and $C3$, respectively. Then, $m_{C1}^2(\cdot) = m_{C2}^2(\cdot) = m_{C3}^2(\cdot)$ because they are the same Lebesgue measure functions. That is, if $\{A_i\}$ is the set with the smallest volume that covers $C1$, then similar sets $\{A_i + 2\}$ and $\{A_i \times 3 + 2\}$ can be used (with the origin located in the upper-left corner) to cover C_3

15 and C_2 with the smallest volumes, respectively. Here, $A_i + 2 = [x_i: x_{i,k} + 2, x_{i,k} \in A_i, k = 1,2]$ and $A_i \times 3 + 2 = [x_i: x_{i,k} \times 3 + 2, x_{i,k} \in A_i, k = 1,2]$, which proves that functions $m_{C1}^2(\cdot)$, $m_{C2}^2(\cdot)$ and $m_{C3}^2(\cdot)$ collect the desirable set based on the same scheme, so they are identical. Additionally, $s_{C2} = m_{C2}^2(C2) = \sum I^2(A_i \times 3 + 2)$ is much larger than $s_{C1} = m_{C1}^2(C1) = \sum I^2(A_i)$ and $s_{C3} = m_{C3}^2(C3) = \sum I^2(A_i + 2)$. Therefore, the scale of $C2$ is not equal to the two other scales because the volumes of their subsets are different. However, their scales are governed by one-dimensional rules because their measures

20 are identical and the Jacobian matrices between them are diagonal.

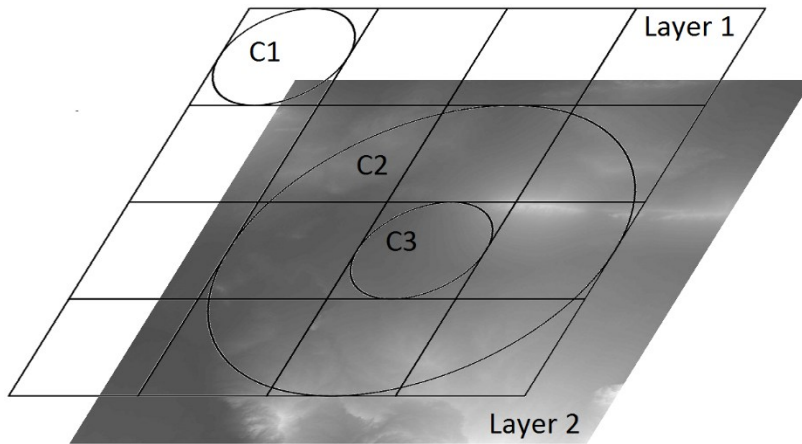


Figure 1. Diagram of the relationships among a Lebesgue measure, scale and geophysical variable

3.2 Stochastic variables in data assimilation

Instead of using Eq. (5) and (6), which are discrete in time, we use Ito process-formed expressions with the one-dimensional
 5 infinitesimals ds and dt to formulate a continuous-time (or continuous-scale) state and observation.

Geophysical variable can be regarded as a real function $V(s, t)$, and it maps the space $(R^2, \mathcal{L}^2, m^2)$ onto R , where s is the
 scale, $s = m^2(A)$, $A \in \mathcal{L}^2$, and t is the time. In n-dimensional data assimilation, a geophysical variable V is related to an
 element of state vector X at a specific scale s and time t .

In Figure 1, Layer 2 presents a heterogeneous geophysical variable in the entire region. If aggregating Layer 2 into Layer 1
 10 and let each pixel intensity is the value for a geophysical variable in each pixel, then the measure space Ω is heterogeneous. A
 geophysical variable represents a spatial average in a specific observation footprint with a specific scale. Therefore, the
 geophysical variables in $C1$ and $C3$ are not equal because their observation footprints are different, and the geophysical
 variables in $C2$ and $C3$ are also different because the scale changes. The former introduces that the geophysical variables vary
 with location, and the latter states that the geophysical variables are scale-dependent.

15 If the statistical properties of the geophysical variable are available, we can construct an explicit stochastic equation for
 the geophysical variable. We introduce the time-dependent Ito process Eq. (1) to define the geophysical variable process:

$$dV = p(t)dt + q(t)dW(t). \quad (9)$$

Similarly, the geophysical variable is supposed to evolve via a stochastic process, for which the dynamic process and
 uncertainty are allowed to vary with scale:

$$20 \quad dV = \varphi(s)ds + \sigma(s)dW(s), \quad (10)$$

where $\varphi(s)$ and $\sigma(s)$ are the scale-based drift rate and volatility rate, respectively. The geophysical variable is a probabilistic
 process with respect to scale and thus has scale-dependent errors, where the scale should shift forward or backward based on
 the condition that the scale follows the one-dimensional rule.

Eq. (9) can be regarded as a continuous-time version of Eq. (5), i.e., to estimate the state is equal to the integral of Eq. (9) over a time interval. Here $p(t)$ indicates the physical process with respect to time, and $q(t)$ is the error only caused by the evolution of time, so model error η in Eq. (5) contains more parts than $q(t)$. Eq. (10) implies that the value and variance of a geophysical variable may change if the scale changes. To formulate $\varphi(s)$ should consider both the spatial heterogeneities and physical process variations among different scales. However, neither of them is well understood in a general theory study. Therefore $\varphi(s)$ is conceptualized in Eq. (10). Particularly, if the study region is homogeneous, then the values of a variable that observed at the same place are identical between large scale and fine scale, and $\varphi(s)$ can be left out. $\sigma(s)$ is the error caused by the scale transformation.

The state in the forecasting step can be expressed by Eq. (9) because only time is involved. In the analysis step of data assimilation, the state does not pertain to time, and we assume that the scale has a quantifiable effect on the errors in this step; thus, both the states and observations can be defined by Eq. (10).

3.3 Expression of scale transformation in a stochastic data assimilation framework

First, we provide the following lemma.

Lemma 1: For $\forall s_0 > 0$, let $W^*(0) = W(s_0) - W(s_0), \dots, W^*(s) = W(s_0 + s) - W(s_0)$; then, $W^*(s), s \geq 0$ is a Brownian motion.

Proof. First, $W^*(0) = W^*(s_0) - W^*(s_0) = 0$. $\forall s_{i+1} > s_i \geq 0, i = 1, 2, 3, \dots$, $W^*(s_{i+1}) - W^*(s_i) = [W(s_0 + s_{i+1}) - W(s_0)] - [W(s_0 + s_i) - W(s_0)] = W(s_0 + s_{i+1}) - W(s_0 + s_i)$, which suggests that the increments $[W^*(s_{i+1}) - W^*(s_i)]$ are equal to $[W(s_0 + s_{i+1}) - W(s_0 + s_i)]$ and are independent Gaussian distributed. Therefore, $W^*(s), s \geq 0$ is a Brownian motion, with $E[W^*(s_{i+1}) - W^*(s_i)] = 0$ and $Var[W^*(s_{i+1}) - W^*(s_i)] = s_{i+1} - s_i$. Q. E. D.

Remark on Lemma 1: Note that in the definition of Brownian motion, the parameter starts at zero. However, the scale is realistically greater than zero, which results that it cannot be directly applied in Brownian motion. So, Lemma 1 is logical because it implies that $W(s), s \geq s_0$ is an equivalent expression of $W^*(s), s \geq 0$. Therefore, beginning with the standard scale, the Brownian motion and stochastic calculus with respect to scale can be further developed.

In the following content, we use Brownian motion with a parameter that starts at s_0 to define the scale-dependent geophysical variables; therefore, the classic expressions above are changed. According to Lemma 1, Eq. (3) is given by

$$[I, I](s) = \int_{s_0}^s \sigma^2(u) du, \quad (11)$$

Additionally, the integral form of the Eq. (10) is as follows:

$$V(s) = V_0 + \int_{s_0}^s \varphi(u) du + \int_{s_0}^s \sigma(u) dW(u), \quad (12)$$

where $V_0 = V(s_0)$ and the drift of Eq. (12) is

$$V_0 + \int_{s_0}^s \varphi(u) du.$$

Similarly, Eq. (4) becomes

$$f(s, V(s)) = f(s_0, V(s_0)) + \int_{s_0}^s f_u(u, V(u))du + \int_{s_0}^s f_x(u, V(u))\sigma(u)dW(u) + \int_{s_0}^s f_x(u, V(u))\varphi(u)du + \frac{1}{2} \int_{s_0}^s f_{xx}(u, V(u))\sigma^2(u)du.$$

Now, we make the following assumptions.

5 **Assumption 1:** The scale transformations between the state and observation spaces of data assimilation obey the one-dimensional rule as defined in Sect. 3.1.

Assumption 2: In the forecasting step, the model unit equals the scale of the state space, and both of them are constant.

Assumption 3: In the analysis step, the state, observation and observation operator are scale dependent. Only one observation is added into the data assimilation system at a time.

10 In assumption 1, the one-dimensional rule ensures that scale changes in a sense of geometrical similarity (for example, from a larger square observation footprint to a smaller square observation footprint, or from C_2 to C_3 as presented in Figure 1). Additionally, the formulations of scale transformation can be extremely reduced.

Assumption 2 indicates that the model unit and state scale are both supposed to be the same and invariant in space and time. So, there is no scale transformation in the forecasting step. Thus, Eq. (9) can adequately describe this step.

15 Based on assumption 3, the analysis step is related to the scale. According to Eq. (10), the state and observation in the analysis step are as follows:

$$dX = \varphi_X(s)ds + \sigma_X(s)dW(s) \quad (13)$$

and

$$dY = \varphi_Y(s)ds + \sigma_Y(s)dW(s), \quad (14)$$

20 where $\varphi_X(s)$, $\sigma_X(s)$, $\varphi_Y(s)$ and $\sigma_Y(s)$ represent the scale-dependent drift rates and volatility rates of state X and observation Y , respectively. $\varphi(s)$ also implies the heterogeneities and physical processes from standard scale to a specific scale, which currently maybe hard to be formulate. $\sigma(u)$ can be regarded as the stochastic perturbation with respect to scale. Therefore, according to Eq. (12), a state is $X(s_X) = X_0 + \int_{s_0}^{s_X} \varphi(u)du + \int_{s_0}^{s_X} \sigma(u)dW(u)$ in the state space and is $X(s_Y) = X_0 + \int_{s_0}^{s_Y} \varphi(u)du + \int_{s_0}^{s_Y} \sigma(u)dW(u)$ in the observation space. These formulas prove that the value of state varies with the changes of scale.

25 The scale transformation only involves in the process that mapping the state vector from state space to observation space. For simplicity, assume the scale-based drift rates of the state and observation do not exist, which leads to $\varphi_X(s) = 0$ and $\varphi_Y(s) = 0$. If the noises are Gaussian, we have $\sigma_X(s) = \sigma_Y(s) = 1$.

Based on the above discussion, the differential and integral forms of the state are

$$30 \quad dX = dW(s) \text{ and } X(s_X) = X_0 + \int_{s_0}^{s_X} dW(s). \quad (15)$$

For the observation, we have

$$dY = dW(s) \text{ and } Y(s_Y) = Y_0 + \int_{s_0}^{s_Y} dW(s) \quad (16)$$

In Eq. (15) and Eq. (16), the time t is omitted, and s_X, s_Y, X_0 and Y_0 represent the scale of the state space, scale of the observation space, state in s_0 and observation in s_0 , respectively.

The Bayesian equation of data assimilation (Eq. (7)) produces the posterior PDF $p(X|Y)$ that is associated with the likelihood function $p(Y|X)$ and the distributions of the state and observation. Theorem 1 and Eqs. (15)~(16) yield $X \sim N(X_0, \int_{s_0}^{s_X} ds)$ and $Y \sim N(Y_0, \int_{s_0}^{s_Y} ds)$ under the condition that the variances exist. In addition, assumption 1 states that the scales vary in one-dimensional space, which results in

$$X \sim N(X_0, s_X - s_0) \quad (17)$$

$$\text{and } Y \sim N(Y_0, s_Y - s_0). \quad (18)$$

Eq. (17) and Eq. (18) are the prior PDFs of state and observation with respect to scale in state space and observation space, respectively. Compared with the PDFs with respect to time, their expectations are equal to the value at the standard scale, and the variances depend on the differences between the standard scale and the scale in state or observation space. These two prior PDFs are introduced into the Bayesian theorem that reformulated by scale.

Then, we calculate the posterior PDF.

The scale-dependent observation operator is $H(s, x)$, which suggests that the observation operator and its parameters are both susceptible to the scale. If $H(s, x)$ is defined, its continuous partial derivatives are $H_s(s, x)$, $H_x(s, x)$ and $H_{xx}(s, x)$. In line with Ito's Lemma, we get an estimation of observation in the observation space, which is related to the state $X(s_X)$ defined in the state space

$$\begin{aligned} H(s_X, X(s_X)) &= H(s_0, X_0) + \int_{s_0}^{s_X} H_s(u, X(u)) du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \\ &= H(s_0, X_0) + \int_{s_0}^{s_X} \left[H_s(u, X(u)) + \frac{1}{2} H_{xx}(u, X(u)) \right] du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u). \end{aligned} \quad (19)$$

Assumption 1 suggests that the observation and model spaces have the same probability measure; thus, the Brownian motions in these two spaces are equivalent. Let Eq. (16) – Eq. (19), and we obtain

$$\begin{aligned} Y(s_Y) - H(s_X, X(s_X)) &= Y_0 + \int_{s_0}^{s_Y} dW(u) - \left[H(s_0, X_0) + \int_{s_0}^{s_X} H_s(u, X(u)) du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right] \\ &= Y_0 - H(s_0, X_0) + \int_{s_0}^{s_Y} dW(u) - \left[H(s_X, X(s_X)) - H(s_0, X(s_0)) \right] - \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du - \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) \\ &= Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right] + \left\{ \int_{s_0}^{s_Y} dW(u) - \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) \right\}. \end{aligned} \quad (20)$$

Equation (20) can be regarded as an Ito process, and its drift is

$$Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right]. \quad (21)$$

The integral term in Eq. (21) is the difference in the first-order differential observation operator between the state scale s_X and the standard scale s_0 . This term illustrates that the mapping process should consider not only the observation operator

but also the first-order differential term when state is mapped to the observation space. The former is typically determined from the literature, whereas the latter was derived in this study for the first time. This result prompted us to further consider the first-order differential of the observation operator when calculating the representativeness error.

The quadratic variation of Eq. (20) is

$$5 \quad (s_Y - s_0) + \int_{s_0}^{s_X} H_x^2(u, X(u)) du. \quad (22)$$

This equation suggests that the uncertainty in the observation error includes both the difference between scales s_Y and s_0 and the change in the observation operator from scale s_X to s_0 . Therefore, Eq. (21) and Eq. (22) can be combined to produce

$$p(Y|X) = N\left(Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du\right], (s_Y - s_0) + \int_{s_0}^{s_X} H_x^2(u, X(u)) du\right). \quad (23)$$

Based on Eqs. (17), (18) and (23), $p(Y|X)$, $p(X)$ and $p(Y)$ are stochastic functions that depend on the scale; thus, the
10 posterior PDF of the state is scale-dependent as well.

In particular, if Y is a direct observation, which means the observation is of the same physical quantity and scale as the state, viz. $H(s, X(s)) = X(s)$. The result becomes

$$Y(s_Y) - X(s_X) = \begin{cases} Y_0 - X(s_X) + W(s_Y) - W(s_X), & s_Y > s_X \\ Y_0 - X(s_X) + W(s_X) - W(s_Y), & s_X > s_Y \end{cases} \quad (24)$$

$$\text{and } p(Y|X) = N\{Y_0 - X(s_X), |s_Y - s_X|\}. \quad (25)$$

15 The quadratic variation in Eq. (22) can be further described by the scale from s_X to s_Y . Under the condition $s_Y > s_X$ and because $W(s_Y) - W(s_X)$ and $W(s_X) - W(s_0)$ are independent, the quadratic variation of Eq. (20) is

$$s_Y - s_X + \int_{s_0}^{s_X} [1 - H_x(u, X(u))]^2 du. \quad (26)$$

Similarly, if $s_X > s_Y$, the quadratic variation of Eq. (20) is

$$\int_{s_0}^{s_Y} (1 - H_x(u, X(u)))^2 du + \int_{s_Y}^{s_X} H_x^2(u, X(u)) du. \quad (27)$$

20 The significance of Eqs. (20)~(27) is that the effect of scale on the posterior PDF can be determined quantitatively. In addition to the model error and instrument error (both of them were not introduced explicitly in this study because they have little influence on the error caused by scale transformation), a new type of error in data assimilation was discovered in the analysis step. The expectation of the posterior PDF may vary with the scale of the state space if Y is an indirect observation, and the variance of the drift depends on the difference between s_Y and s_X (based on Eq. (26) and Eq. (27)) or among s_0 , s_Y
25 and s_X (based on Eq. (22)). In addition, if Y is a direct observation (Eq. (24) and Eq. (25)), the expectation of the posterior PDF is the difference between Y and X , and the variance is equal to the increment of Brownian motion with respect to the

scale. Additionally, if the results are not derived from assumption 1, i.e., the scale varies randomly, the posterior PDF is more complex because the Jacobian matrix in Lebesgue integration of scale transformation is arbitrary.

3.4 Examples: the stochastic radiative transfer equation (SRTE)

To explicitly show how the stochastic scale transformations impact on assimilation, we introduce an illustrative example based on the scales presented in Figure 1. Assuming that in the analysis step, the state is with the standard scale s_0 , whose observation footprint is the unit square A_0 . If the scale of observation space is s_{C1} and its observation footprint is the disc C_1 , then the Jacobian matrix of the transformation between the scales of state space and observation space is not diagonal according to the statements in Sect. 3.1, leading the two scales do not obey the one-dimensional rule and against assumption 1. However, if let the scales of of state space and observation space are s_{C1} and s_{C2} , respectively, the assumption 1 is met and it can be counted that $s_X = s_{C1} = \frac{\pi}{4} s_0$ and $s_Y = s_{C2} = \frac{9\pi}{4} s_0$.

Now the scales of state space and observation space obey the one-dimensional rule, and then we further presume that the measure space Ω in Figure 1 is free of the spatial heterogeneities and dynamic process variations depending on scale. Consequently, the drift rate $\varphi(s) = 0$. If denoting the value of state in the standard scale is X_0 , then the prior PDF of state is $X \sim N\left(X_0, \frac{\pi}{4} s_0 - s_0\right)$ according to Eq. (17)., Noting that $\frac{\pi}{4} s_0 - s_0$ is not a real number and only indicates the variation when the scale changes.

If $H(s, X(s)) = X(s)$, the observation is the same physical quantity as the state, and according to Eq. (25), the likelihood function is $p(Y|X) = N\{Y_0 - X(s_X), |s_Y - s_X|\} = N\{Y_0 - X(s_X), |s_{C2} - s_{C1}|\} = N\left\{Y_0 - X(s_X), \left|\frac{9\pi}{4} s_0 - \frac{\pi}{4} s_0\right|\right\}$.

To formulate the likelihood function in the case that the observation is different from the state, the SRTE will be employed in the following text. The SRTE is a stochastic integral-differential equation that describes the radiative transfer phenomena through a stochastically mixed immiscible media. Scientists have developed analytical or numerical methods for finding the stochastic moments of the solution, such as the ensemble-averaged or variance of the radiation intensity (Pomraning, 1998; Shabanov et al., 2000; Kassianov et al., 2011).

Consider the general expression of the SRTE (leave out the scattering and emission),

$$-\mu \frac{dI(\tau)}{d\tau} = -I(\tau), \quad (28)$$

where $I(\tau)$, μ and τ are the radiation intensity, coefficient of radiation direction and optical depth, respectively. The analytical solution of Eq. (28) is $I(\tau) = -I(0)e^{\tau/\mu}$.

To tie into more substantial random optical properties of transfer media, such as absorption and scattering, the optical depth τ is assumed to be stochastic. So it suggests that optical depth is a scale-dependent Ito process and can be expressed as

$$d\tau(s) = \varphi_\tau(s)ds + \sigma_\tau(s)dW(s), \quad (29).$$

This causes the radiation intensity depend on scale.

SRTE can be considered as a concrete instance of stochastic observation operator by defining $H(s, x(s)) = I(x) = I(0)e^{x/\mu}$. Therefore, $H_s(s, x(s)) = 0$, $H_x(s, x(s)) = \frac{1}{\mu}I(0)e^{x/\mu}$ and $H_{xx}(s, x(s)) = \frac{1}{\mu^2}I(0)e^{x/\mu}$. Based on Ito's Lemma,

$$\begin{aligned} dI(\tau(s)) &= dH(s, \tau(s)) = H_s(s, \tau(s))ds + H_x(s, \tau(s))d\tau(s) + \frac{1}{2}H_{xx}(s, \tau(s))d\tau(s)d\tau(s) \\ &= \frac{1}{\mu}I(0)e^{\tau(s)/\mu}d\tau(s) + \frac{1}{2\mu^2}I(0)e^{\tau(s)/\mu}d\tau(s)d\tau(s) \\ &= \frac{1}{\mu}I(\tau(s))d\tau(s) + \frac{1}{2\mu^2}I(\tau(s))d\tau(s)d\tau(s) \quad (30) \\ &= \left(\frac{1}{\mu}I(\tau(s)) \right) \sigma_\tau(s)dW(s) + \left(\frac{1}{\mu}I(\tau(s)) \right) \varphi_\tau(s)ds + \left(\frac{1}{2\mu^2}I(\tau(s)) \right) \sigma_\tau^2(s)ds \\ &= \left(\frac{\sigma_\tau(s)}{\mu} \right) I(\tau(s))dW(s) + \left(\frac{\sigma_\tau^2(s)}{2\mu^2} + \frac{\varphi_\tau(s)}{\mu} \right) I(\tau(s))ds \end{aligned}$$

Radiation intensity is a scale-dependent Ito process. The difference between Eq. (30) and the general Ito process is that there is a primitive function $I(\tau(s))$ in the integral term. Therefore, the uncertainty of the radiation intensity is more complex because it is related to both the change of scale and the primitive function.

Integrating both sides of Eq. (30) yields the general solution of the radiation intensity,

$$I(\tau(s)) = C \cdot \exp \left[\int \left(\frac{\sigma_\tau(s)}{\mu} \right) dW(s) + \int \left(\frac{\sigma_\tau^2(s)}{2\mu^2} + \frac{\varphi_\tau(s)}{\mu} \right) ds \right], \quad (31)$$

where the constant $C \in R$. Eq. (31) further indicates that $I(\tau(s))$ is a scale-dependent Ito process. Considering that the optical depth τ is the state, the radiation intensity I is the observation and $I(\tau(s))$ is the observation operator, then the above results in Sect. 3.3 (For example, Eq. (20)) could be easily applied here to study the posterior PDF of data assimilation.

3.5 Extension to n-dimensional data assimilation

In the above discussion, we assumed that only one state existed in data assimilation. However, numerous states typically exist. This section further introduces the **product spaces** to extend the one-dimensional stochastic data assimilation to n-dimensions.

5 Assume that the independent states X_k are the variables of the measure spaces $(\Omega_k, \mathcal{F}_k, \mu_k)$, $k = 1, 2, \dots, n$, and $(\Omega^n, \mathcal{F}^n)$ is the product space, where $\Omega^n = \prod_{k=1}^n \Omega_k$ and $\mathcal{F}^n = \prod_{k=1}^n \mathcal{F}_k$. According to Fubini's theorem (Billingsley, 1986), only one product measure μ^n in $(\Omega^n, \mathcal{F}^n)$ exists, such that $\mu^n(\prod_{k=1}^n A_k) = \prod_{k=1}^n \mu_k(A_k)$, where $A_k \in \mathcal{F}_k$.

We define the state vector $X^n = (X_1, X_2, \dots, X_n)^T$ as a variable vector of the product measure space $(\Omega^n, \mathcal{F}^n, \mu^n)$. In particular, if all the scales obey the one-dimensional rule, we have

$$10 \quad \mu^n \left(\prod_{k=1}^n A_k \right) = \prod_{k=1}^n \xi_k^2 \mu_0(A_k) = \left(\prod_{k=1}^n \xi_k \right)^2 \mu_0^n \left(\prod_{k=1}^n A_k \right).$$

This expression proves that the product measure also obeys a one-dimensional rule. However, the above results may not hold without the assumption that the states X_k are independent.

As discussed in Sect. 2.1, the Lebesgue measure m^2 is a measure and the triple $(R^2, \mathcal{L}^2, m^2)$ is a measure space. Therefore, the above extension is reasonable.

15 The analysis of a single state can also be applied to finite multiple states in the product measure space.

4 Discussion & Conclusions

4.1 Discussion

Our study offered a stochastic data assimilation framework to formulate the errors that are caused by scale transformation. The necessity of the methodology, the difference to previous works by other investigators, and the advantages and limitations of this study are discussed as follows.

The reasons that the methodology focuses on a stochastic framework are: First, the stochastic data assimilation framework is essentially consistent with the conceptions of scale and scale transformation. Both of them are associated with corresponding measure spaces $(\Omega, \mathcal{F}, \mu)$. Therefore, it is natural to regard the state space and observation space as two different measure spaces, respectively, and each element of state (or observation) vector can be seen as a geophysical variable that mapping the state (or observation) measure space onto R . Correspondingly, as the integrals of random processes with respect to random processes, stochastic calculus was adopted ultimately. Second, using stochastic calculus can also formulate the errors caused by scale transformations. The study proceeds with and improves the understanding of representativeness error in terms of scale.

Results did not only prove the conventional point that the uncertainties of these errors mainly depend on the differences between scales, but indicated that the first-order differential of the nonlinear observation operator should also be incorporated in representativeness error. Last, stochastic calculus can be extended to meet a general scale transformation and formulate corresponding representativeness error. This was unattainable in previous work. For example, if the scale changes randomly, say, from an irregular footprint to another irregular footprint, the stochastic equation can offer a multiple-integral to present this kind of a scale transformation, such as $V(x, y) = V_0 + \int_{Y_0}^Y \int_{X_0}^X \varphi(x, y) dx dy + \int_{Y_0}^Y \int_{X_0}^X \sigma(x, y) dW_1(x) dW_2(y)$, where $W_1(x)$ and $W_2(y)$ are two independent Brownian Motion.

The significant of this work is: We developed a more rigorous formulation of scale and the scale transformation based on Lebesgue measure, which places the related conceptions in a rigorous mathematical framework and then conduces new understanding of the errors caused by scale transformation. In addition, due to the Ito process-formed state and observation, a stochastic data assimilation framework was proposed by considering the nonlinear operators, heterogeneity of a geophysical variable and a general Gaussian representativeness error. Scale transformation is also nonlinear if the one-dimensional rule is not involved. Additionally, Ito processes-formed state and observation offer the drift rate (i.e., $\varphi(s)$ in Eq. (10)) to formulate the heterogeneity associated with scale transformation. It also permits the representativeness error to be general Gaussian in this framework. If all the integrands in Eq. (13) and Eq. (14) are nonlinear functions instead of constants (in this study we let $\varphi_x(s) = 0$, $\varphi_y(s) = 0$ and $\sigma_x(s) = \sigma_y(s) = 1$ for simplicity), then these two equations are integrated over the field of Brownian motion, and state and observation are the general Gaussian processes of scale. Based on these functions, representativeness error is a general Gaussian process.

As a theoretical exploration towards scale transformation and stochastic data assimilation, there is still big room for improvement. First, we reduced the scale transformation by one-dimensional rule, and let the variables in data assimilation evolve regularly according to assumptions 1~3. So, only the ideal result was investigated. Therefore, an in-depth and comprehensive exploration should be conducted in future to describe other situations in the real world. However, either an arbitrary scale transformation or the geophysical variable without ignoring the drift rates will deduce lengthy results. Therefore, the second improvement focuses on how to make the formulation more concise. Last, noting that all the results in our framework were given in terms of probability, it is necessary to implement the real-world applications of these theoretical results, such as introducing some concrete dynamic models to formulate the Ito process-formed geophysical variable of scale.

4.2 Conclusions

In this study, we mainly addressed two basic problems associated with scale transformation in earth observation and simulation. First, we produced a mathematical formalism of scale and scale transformation by employing measure theory. Second, we demonstrated how scale transformation and associated errors could be presented in a stochastic data assimilation framework.

We revealed that the scale is the Lebesgue measure with respect to the observation footprint or model unit. Scale is related to the shape and size of a footprint, and scale transformation depends on the spatial change between different footprints. We then defined the geophysical variable, which further considers the heterogeneities and physical processes. A geophysical variable consequently expresses the spatial average at a specific scale.

5 We formulated the expression of scale transformation and investigated the error structure that is caused by scale transformation in data assimilation using basic theorems of stochastic calculus. Formulations explicate that the first-order differential of the nonlinear observation operator should be considered in representativeness error, and the uncertainty of representativeness error is directly associated with the difference between scales. A concrete physical models (SRTE) was introduced to demonstrate the results when observation operator is nonlinear. Extension the results to n-dimensional stochastic data assimilation was also presented.

This work conducted a theoretical exploration of formulating the errors caused by scale transformation in stochastic data assimilation framework. We hope that the stochastic methodology can essentially benefit the study on these errors.

5 Notation

5.1 Basic notations

15	Ω	Non empty space
	\mathcal{F}	σ -algebra
	μ	Measure
	dV	Variable process
	$W(s)$	Brownian motion
20	$(\Omega, \mathcal{F}, \mu)$	Measure space
	I^n	N-dimensional Lebesgue volume
	m^n	Lebesgue measure or an outer measure on R^n
	\mathcal{L}^n	Lebesgue σ -algebra of R^n
	$\int f dm^n$	Lebesgue integral
25	$ J(\cdot) $	Jacobian determinant
	$(\Omega^n, \mathcal{F}^n)$	Product space

5.2 New notations

Notation	Name	Explanation	Index
s	Scale	The observation footprint or model unit to observe or model a geophysical variable	Sect. 1 & Sect. 3.1

A_0	Unit square in R^2		Sect. 3.1
s_0	Standard scale	A Lebesgue integral of A_0 is the unit area	Sect. 3.1
	One-dimensional rule	Two scales are geometrically similar	Eq. (8)
V	Geophysical variable	Estimation of a variable at a specific scale	Sect. 3.2
dX	State process	Ito process-formed state	Eq. (13)
dY	Observation process	Ito process-formed observation	Eq. (14)
X_0	State in s_0		Eq. (15)
Y_0	Observation in s_0		Eq. (16)
s_X	Scale of state space		Eq. (15)
s_Y	Scale of observation space		Eq. (16)

Acknowledgements

We thank the editor-in-chief of NGP, Prof. Talagrand, and his kind help and valuable comments on our manuscript. We also thank Dr. van Leeuwen and another anonymous reviewer for their valuable comments and suggestions. This work was supported by the NSFC projects (grant numbers 91425303 & 91625103) and the CAS Interdisciplinary Innovation Team of the Chinese Academy of Sciences.

References

- Apte, A., Hairer, M., Stuart, A. M. and Voss, J.: Sampling the posterior: An approach to non-Gaussian data assimilation, *Physica D*, 230, 50-64, doi: 10.1016/j.physd.2006.06.009, 2007.
- Atkinson, P. M. and Tate, N. J.: Spatial scale problems and geostatistical solutions: a review, *Prof. Geogr.*, 52, 607-623, doi: 10.1111/0033-0124.00250, 2004.
- Bartle, R. G.: *The Elements of Integration and Lebesgue Measure*, Wiley, New York, 1995.
- Billingsley, P.: *Probability and Measure*, 2nd ed., John Wiley & Sons, New York, 1986.
- Bocquet, M., Pires, C. A., Wu, L.: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation, *Mon. Weather Rev.*, 138, 2997-3023, doi: 10.1175/2010MWR3164.1, 2010.
- Bocquet, M., Wu, L., Chevallier, F.: Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism, *Q. J. Roy. Meteor. Soc.*, 137, 1340-1356, doi: 10.1002/qj.837, 2011.
- Crow, W. T., Berg, A. A., and Cosh, M. H., et al.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, *Rev. Geophys.*, 50, 3881-3888, doi: 10.1029/2011RG000372, 2012.

- Eyink, G. L., Restrepo, J. M., and Alexander, F. J.: A mean field approximation in data assimilation for nonlinear dynamics, *Physica D*, 195, 347-368, doi: 10.1016/j.physd.2004.04.003, 2004.
- Famiglietti, J.S., Ryu, D., Berg, A.A., Rodell, M., and Jackson, T.J.: Field observations of soil moisture variability across scales, *Water Resour. Res.* 44, doi: 10.1029/2006WR005804, 2008.
- 5 Giménez, D., Rawls, W. J., and Lauren, J. G.: Scaling properties of saturated hydraulic conductivity in soil, *Geoderma*, 27, 115-130, 1999.
- Goodchild, M. F. and Proctor, J.: Scale in a digital geographic world, *Geographical & Environmental Modelling*, 1, 5-23, 1997.
- Gruber, A., Dorigo, W. A., Zwieback, S., Xaver, A., and Wagner, W.: Characterizing coarse-scale representativeness of in situ soil moisture measurements from the international soil moisture network, *Vadose Zone J.*, 12, 522-525, doi: 10.1002/jgrd.50673, 2013.
- 10 Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements, *J. Geophys. Res-Atmos.*, 118, 8585–8597, 2013.
- Hodyss, D. and Nichols, N. K.: The error of representation: Basic understanding, *Tellus A*, 66, 1-17, doi: 10.3402/tellusa.v67.24822, 2015.
- 15 Huang, G., Li, X., Huang, C., Liu, S., Ma, Y., and Chen, H.: Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products, *Remote Sens. Environ.*, 181, 198-206, doi: 10.1016/j.rse.2016.04.001, 2016.
- Itô, K.: Stochastic integral, *P. JPN. Acad.*, 22, 519-524, 1944.
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., and Ustin, S. L.: Prospect and sail models: a review of use for vegetation characterization, *Remote Sens. Environ.*, 113, S56-S66, 2009.
- 20 Janjić, T. and Cohn, S. E.: Treatment of Observation Error due to Unresolved Scales in Atmospheric Data Assimilation, *Mon. Weather Rev.*, 134, 2900-2915, doi: 10.1175/MWR3229.1, 2006.
- Jazwinski, A. H.: Stochastic processes and filtering theory, Academic Press, New York, 1970.
- Karatzas, I. and Shreve, S.E.: Brownian Motion and Stochastic Calculus, 2nd ed., Springer-Verlag, New York, 1991.
- 25 Kassianov, E. and Veron, D.: Stochastic radiative transfer in Markovian mixtures: past, present, and future, *J. Quant. Spectrosc. Radiat. Transf.* 112, 566-576, doi: 10.1016/j.jqsrt.2010.06.011, 2011.
- Li, X.: Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems, *Sci. China Ser. D*, 57, 80-87, doi: 10.1007/s11430-013-4728-9, 2014.
- Li, X. and Liu, F.: Can Point Measurements of Soil Moisture Be Used to Validate a Footprint-Scale Soil Moisture Product?, *IEEE Geosci. Remote S.*, 2016. (Submitted)
- 30 Lin, H., Flühler, H., Otten, W., and Vogel, H. J.: Soil architecture and preferential flow across scales, *J. HYDROL.*, 393, 1-2, doi: 10.1016/j.jhydrol.2010.07.026, 2010.
- Lorenc, A.C.: Atmospheric Data Assimilation, Meteorological Office, Bracknell, 1995.
- Lorenc, A. C.: Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc.*, 112, 1177–1194, 1986.

- Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modelling, *Water Resour. Res.*, 45, 627-643, doi: 10.1029/2009WR007872, 2009.
- Miller, R. N., Carter, E. F., and Blue, S. T.: Data assimilation into nonlinear stochastic models, *Tellus A*, 51, 167-194, doi: 10.1034/j.1600-0870.1999.t01-2-00002.x, 1999.
- 5 Miller, R. N.: Topics in data assimilation: stochastic processes, *Physica D*, 230, 17-26, doi: 10.1016/j.physd.2006.07.015, 2007.
- Miralles, D. G., Crow, W. T., and Cosh, M. H.: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, *J. Hydrometeorol.*, 11, 1423-1429, 2010.
- Narsilio, G. A., Buzzi, O., Fityus, S., et al.: Upscaling of Navier–Stokes equations in porous media: Theoretical, numerical and experimental approach. *Computers & Geotechnics*, 36, 1200–1206, 2009.
- 10 Pomraning, G.C.: Radiative transfer and transport phenomena in stochastic media, *Int. J. Eng. Sci.*, 36, 1595–1621, doi: 10.1016/S0020-7225(98)00050-0, 1998.
- Ran, Y. H., Li, X., Sun, R., Kljun, N., Zhang, L., Wang, X. F., and Zhu, G. F.: Spatial representativeness and uncertainty of eddy covariance carbon flux measurement for upscaling net ecosystem productivity to field scale, *Agric. Forest Meteorol.*, doi: 10.1016/j.agrformet.2016.05.008, 2016.
- 15 Ryu, D. and Famiglietti, J. S.: Multi-scale spatial correlation and scaling behavior of surface soil moisture, *Geophys. Res. Lett.*, 33, 153-172, 2006.
- Shabanov, N. V., Knyazikhin, Y., Baret, F., and Myneni, R. B.: Stochastic modeling of radiation regime in discontinuous vegetation canopies, *Remote Sens. Environ.*, 74, 125-144, doi: 10.1016/S0034-4257(00)00128-0, 2000.
- Shreve, S.E.: *Stochastic Calculus for Finance II*, Springer-Verlag, New York, 2005.
- 20 Talagrand, O.: Assimilation of observations, an introduction, *J. Meteorol. Soc. JPN*, 75, 191-209, 1997.
- van Leeuwen, P. J.: Representation errors and retrievals in linear and nonlinear data assimilation, *Q. J. Roy. Meteor. Soc.*, 141, 1612-1623, doi: 10.1002/qj.2464, 2014.
- van Leeuwen, P. J.: Nonlinear data assimilation for high-dimensional systems, in *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, vol. 2, Springer-Verlag, New York, 2015.
- 25 Vereecken, H., Kasteel, R., Vanderborght, J., and Harter, T.: Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: a review, *Vadose Zone J.*, 6, 1-28, doi: 10.2136/vzj2006.0055, 2007.
- Wiens, J. A.: Spatial scaling in ecology, *Funct. Ecol.*, 3, 385-397. doi: 10.2307/2389612, 1989.

Formulation of Scale Transformation in a Stochastic Data Assimilation Framework

Feng Liu^{1,3} and Xin Li^{1,2,3}

Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, 730000, China

²Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, ~~Beijing 100101, P. R. China~~

³University of Chinese Academy of Sciences, Beijing 100049, ~~P. R. China~~

Correspondence to: Xin Li (lixin@lzb.ac.cn)

Abstract: ~~The understanding of uncertainties—the errors caused by spatial scale transformation~~ in Earth observations and simulations ~~requires a rigorous definition of scale, has been hindered by the spatial scale issue. In addition, errors that are caused by spatial scale transformation~~ These errors are also an important component of representativeness errors in data assimilation. Several relevant studies have been conducted, but ~~the theorization of these—the scale associated representativeness errors is still exceed the abilities of current theory not well developed because of the associated nonlinearity. Thus, we attempt~~ to address these problems ~~by reformulating the data assimilation framework using measure theory and stochastic calculus.~~ First, the measure theory is used to propose ~~a mathematical definition such~~ that the spatial scale is the Lebesgue measure with respect to the observation footprint or model unit. ~~Then, and the~~ Lebesgue integration by substitution is used to describe the scale transformation. Second, a scale-dependent ~~variable~~ geophysical variable is defined to ~~further~~ consider the heterogeneities ~~and dynamic processes.~~ Finally, the structures of scale-dependent errors ~~in nonlinear and general Gaussian senses~~ are studied in the Bayesian framework of data assimilation based on stochastic calculus. ~~If we restrict the scale to be~~ All the results were presented on the condition that scale is one-dimensional, ~~and~~ the variations in ~~this type of these errors is proportional to~~ depend on the difference between scales. This new ~~methodology formulation can expand the understanding and treatment of~~ provides a more general framework to understand the representativeness error in ~~a nonlinear and stochastic sense~~ data assimilation ~~and may be able to~~ and is a promising way to address the spatial scale issue.

1 Introduction

~~Scientists have devoted considerable attention to understanding uncertainties in Earth observations and simulations. However, uncertainties that are caused by spatial scale—(scale is used below to abbreviate spatial scale)—transformations result in significant errors in understanding geophysical variables and have yet to be fully addressed. Here, the spatial scale refers to the observation footprint or model unit in which a geophysical parameter can be measured or evaluated modelled. Empirical~~

studies have been conducted only recently. Scientists have devoted considerable attentions to understanding uncertainties associated to scale in Earth observations and simulations. Investigations Studies have found that the uncertainty increases with increases in the difference between spatial scales (Famiglietti et al., 2008; Crow et al., 2012; Gruber et al., 2013; Hakuba et al., 2013; Huang et al., 2016; Li and Liu, 2016; Ran et al., 2016). This uncertainty that is associated with the spatial scale (for
5 brevity, the term “scale” is used to refer to the spatial scale below) results in significant errors in understanding geophysical parameters.

—The scale issue is mainly derived from the strong spatial heterogeneities (Miralles et al., 2010; Li, 2014) and irregularities (Atkinson and Tate, 2000) that are associated with geophysical parameters across different scales, and both the spatial heterogeneities and irregularities vary nonlinearly with scale. In addition, the scale issue is closely related to dynamic process
10 variations in land surface systems, which include hydrology (Giménez et al., 1999; Vereecken et al., 2007; Merz et al., 2009), soil science (Ryu and Famiglietti, 2006; Lin et al., 2010), radiative transfer (Jaquemoud et al., 2009) and ecology (Wiens, 1989).

—A mathematic conceptualization of scale is extremely important to study Earth observations and simulations. However, traditionally, scale is not explicitly expressed in geosystem dynamics and its measurement. A rigorous definition of scale is
15 difficult to find, except for an intuitive conception (Goodchild and Proctor, 1997) and certain qualitative classifications of scale (Vereecken et al., 2007). This gap partially reflects the complexity of this problem and requires corresponding mathematical tools to elucidate the “scale”. The spatial scale in Earth observations and simulations refers to the observation footprint or model unit in which a geophysical variable is observed or modelled (scale is used below to abbreviate spatial scale). Scale is traditionally defined in terms of distance, which is not adequate both because distance is a one-dimensional
20 quantity but scale generally refers to a two- or three-dimensional space, and because scale may change much complicated (for example, form an irregular observation footprint to a square observation footprint). Generally, scale is not explicitly expressed in the dynamics of a geophysical variable, partially because a rigorous definition of scale is difficult to find, except for an intuitive conception (Goodchild and Proctor, 1997) and certain qualitative classifications of scale (Vereecken et al., 2007). This reflects the complexity of scale and requires a more rigorous mathematical conceptualization of scale.

Scale transformation of a geophysical variable may result in significant errors (Famiglietti et al., 2008; Crow et al., 2012; Gruber et al., 2013; Hakuba et al., 2013; Huang et al., 2016; Li and Liu, 2016; Ran et al., 2016). These errors are mainly caused by the strong spatial heterogeneities (Miralles et al., 2010; Li, 2014) and irregularities (Atkinson and Tate, 2000) that are associated with geophysical variables across different scales, and are also closely related to dynamic variations, e. g., hydrological (Giménez et al., 1999; Vereecken et al., 2007; Merz et al., 2009; Narsilio, et al. 2009), soil (Ryu and Famiglietti, 2006; Lin et al., 2010) and ecological (Wiens, 1989) processes. How to develop mathematical tools to elucidate the scale transformation has yet to be fully addressed.

Data assimilation ~~could be an ideal tool to explore the scale transformation presents~~ because it presents a unified and generalized framework in Earth system modelling and observation ~~in a unified and generalized framework~~ (Talagrand, 1997) ~~and therefore is an ideal tool to explore scale transformation.~~ In the forecasting operators of data assimilation, scale and associated uncertainties exist in forcing data and parameters, which Geophysical data are typically collected ~~observed~~ by various Earth observations ~~techniques or from data products~~, therefore to update the observation data in a data assimilation system may result in scale transformations between observation space and system state space ~~therefore, scale mismatch may arise between them.~~ Furthermore, this problem is even more common between the model units and observation footprints of measurements. If observation operator is strongly nonlinear and complex, errors caused by scale transformation is even more serious (Li, 2014), because both the forecasting and observation operators in data assimilation are likely strongly nonlinear and complex (Li, 2014). The scale issue cannot be properly treated using traditional linear rules in Earth observations and simulations. The forecasting and observation operators of a data assimilation system are typically deterministic models. Recently, nonlinear dynamic models that were based on stochastic differential equations (SDEs), such as the stochastic Lorenz model (Miller et al., 1999; Eyink et al., 2004), have been studied in assimilation. A data assimilation study that was based on stochastic processes (Miller, 2007; Apte et al., 2007) has also been proposed. Compared to deterministic models, data assimilation that is based on stochastic models is more applicable in an integrated and time-continuous theoretical study (Bocquet et al., 2010), and creates an infinite sampling space of the system state (Apte et al., 2007). However, the theorems of calculus that are based on stochastic processes (or stochastic calculus) are different from those of ordinary calculus. Scale

~~transformations between different components of data assimilation must be reformulated in a stochastic manner to fully present the random and nonlinear geosystem dynamics and observations in a multi-scale data assimilation framework.~~

—An important concept that is related to scale transformation in data assimilation is “representativeness error”, which is associated with the inconsistency in spatial and temporal resolutions between states, observations and operators (Lorenc, 1986;

5 Janjić and Cohn, 2006; van Leeuwen, 2014; Hodyss and Nichols, 2015), and the missing physical information that is related to numerical operator compared to the ideal operator (van Leeuwen, 2014), such as the discretization of a continuum model or neglect of necessary physical processes. ~~The first source of representativeness error is related to scale. According to the above~~

~~discussion, scale issue produces effects on the land surface dynamic process, so we argue that the second is also partly associated with the scale variations in physical processes. Thus, the scale issue is a universal phenomenon in the study of Earth~~

10 ~~observations and simulations and inevitably results in representativeness error.~~The representativeness error and measurement instrument error make up the observation error of data assimilation. Under the Gaussian assumption, they are independent of each other (Lorenc, 1995; van Leeuwen, 2014). This study will not introduce the measurement instrument error when formulate the scale transformation in data assimilation.

Recently, approaches have been developed to assess representativeness error. Janjić and Cohn (2006) studied

15 ~~representativeness error by treating system treated states as the sum of resolved and unresolved portions. This resulted in observation error was the sum of the measurement error and representativeness error.~~ Bocquet et al. (2011) used a pair of operators, namely, restriction and prolongation, to connect the relationship between the finest regular scale and a coarse scale, and determined ~~the scale dependent~~ representativeness error using a multi-scale data assimilation framework. van Leeuwen

(2014) considered two complicated cases: ~~In one, i.e., conducting the observation vector s had in a finer resolution than the model. In the other, compared with system state vector (we below use “observation” and “state” for brevity) and assimilating~~

20 ~~the retrieved variables, which represented different dynamic processes, were assimilated.~~ Their solutions were formulated using an agent ~~variable~~ in observation or model state space, and a particle filter was proposed to treat the nonlinear relationship between observations, states and retrieved values. Hodyss and Nichols (2015) also estimated the representativeness error based on the concept that the main cause of this error is by investigating the difference between the truth and the inaccurate value that

25 ~~is forecasted-generated by the forecasting model.~~

~~Overall, Although~~ these approaches explored the structure of representativeness error and offered various solutions. ~~However,~~ improvements ~~can still be made~~ are still necessary to investigate what is the exact expression of errors caused by scale transformation in data assimilation. The authors believe that these approaches are optimal in linear systems, but may not be suitable when observations are heterogeneous and sparse ~~and thus cannot be averaged to fit model units at a relatively coarse~~ 5 ~~scale,~~ or when operators are nonlinear between states and observations. ~~In previous studies,~~ ~~although the general equations in the nonlinear case were given,~~ the forecasting and observation operators, maps of the resolutions of different variables and models were assumed to be linear. Representativeness error is unavoidable, even if micro-scale observations are averaged over a larger area (van Leeuwen, 2014; Li and Liu, 2016), partly because of the heterogeneity of geophysical parameter. ~~However,~~ Without taking heterogeneities and nonlinear operators into account, representativeness error cannot be fully understood. 10 ~~However,~~ heterogeneity varies ~~depending on the situation~~ situationally and is difficult to be formulated in a general theory ~~an~~ ~~integrated~~ study. ~~We can use semivariogram to quantify the heterogeneity of a geophysical parameter in a special region at a special time, but have no idea how to generalize this result to the entire region and time.~~

~~Data assimilation studies based on stochastic processes (Miller, 2007; Apte et al., 2007) or stochastic dynamic model (Miller et al., 1999; Eyink et al., 2004) have been proposed recently. Compared to deterministic models, stochastic data assimilation~~ 15 ~~is more applicable in an integrated and time-continuous theoretical study (Bocquet et al., 2010), and creates an infinite sampling space of the system state (Apte et al., 2007). Although the theorems of calculus that are based on stochastic processes (or stochastic calculus) are different from those of ordinary calculus, these advantages suggest that stochastic data assimilation offers a more general framework to study scale transformation.~~

~~We attempt to explore the mathematic definitions of scale and scale transformation, and then formulate the errors caused by~~ 20 ~~scale transformation in a general theory study on stochastic data assimilation. We believe that the solution to this problem should begin with an integrated study of all the random evolutions of a parameter in its probability distribution space. Meanwhile, data assimilation also stresses an integrated understanding of the probability distribution function (PDF) of the model space, which results in an estimation of the first and second moments (data value and error information).~~

~~—— In this study, we attempt to explore the mathematic definition of scale and how scale transformation influences the~~ 25 ~~errors in data assimilation.~~ The next section introduces the basic concepts and theorems of measure theory, ~~and~~ stochastic calculus ~~and data assimilation.~~ In Sect. 3, we present ~~some essential concepts, such as~~ the definitions of scale, ~~and~~ scale transformation ~~and variable,~~ which form the basis for the subsequent study. In Sect. 4, ~~The posterior probability of system state was also reformulated by scale transformation in a stochastic data assimilation framework. we establish a Bayesian description of data assimilation with time and scale dependent stochastic processes and investigate the effect of scale transformations on~~

~~the posterior probability of the state. In the final section, the contributions and deficiencies of this study is/were presented in light of previous work discussed. Comments and future work are also summarized.~~

2 Basic knowledge

~~As mentioned above,~~ The scale greatly depends on the geometric features of a certain observation footprint or model unit.

- 5 The model unit is a specified subspace where a geophysical variable evolves in the model space. It could be a point, a rectangular grid, or an irregular unit such as a response unit (watershed, landscape patch and so on). We offer a solution in which the definition of scale ~~must use the~~ uses measure theory and the expression of geophysical ~~parameter~~ variable as a stochastic process ~~must use~~ stochastic calculus. Therefore, we first introduce several basic concepts of measure theory and stochastic calculus.

10 2.1 Measure theory

Let Ω be an arbitrary nonempty space. \mathcal{F} is a **σ -algebra** (or **σ -field**) of subsets of Ω that satisfies the following conditions:

- (i) $\Omega \in \mathcal{F}$, and the empty set $\Phi \in \mathcal{F}$;
- (ii) $A \in \mathcal{F}$ implies that its complementary set $A^c \in \mathcal{F}$;
- (iii) $A_1, A_2, \dots \in \mathcal{F}$ implies their union $A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

15 A set function μ of \mathcal{F} is called a **measure** if it satisfies the following conditions:

- (1) $\mu(A) \in [0, \infty)$ and $\mu(\Phi) = 0$;
- (2) If $A_1, A_2, \dots \in \mathcal{F}$ is any disjoint sequence and $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, μ is countably additive such that $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$.

If $\mu(\Omega) = 1$, μ can be replaced by the probability measure $p\mathcal{P}$, and if μ is finite, $p\mathcal{P}$ can be calculated as $p\mathcal{P}(A) = \mu(A)/\mu(\Omega)$. The triples $(\Omega, \mathcal{F}, \mu)$ and $(\Omega, \mathcal{F}, p\mathcal{P})$ are the **measure space** and **probability measure space**, respectively.

20

Let Ω be the set of real numbers R and σ -algebra \mathcal{B} be **Borel algebra**, which is generated by all closed intervals in R . Then $\forall A = [a, b] \in \mathcal{B}$, a **Lebesgue measure** on R is defined as $I(A) = b - a$. Intuitively, the Lebesgue measure on R **actually** coincides with length.

An **n-dimensional Lebesgue volume** is defined to measure the standard volumes of subsets in R^n based on $I^n(A) = \prod_{k=1}^n (b_k - a_k)$, where $A = [x: a_k \leq x_k \leq b_k, k = 1, 2, \dots, n]$ is an n-dimensional regular cell in R^n . The n-dimensional Lebesgue volume is an ordinary volume, such as length (n=1), area (n=2) and volume (n=3).

~~Generally, a Lebesgue measure on R^n assumes that A is any subset of R^n . First~~Next, we define the **outer measure** is defined as $m^n(A) = \inf\{\sum_{i=1}^{+\infty} I^n(A_i)\}$, where $\inf\{\cdot\}$ is the infimum, $A_i = [x: a_{i,k} \leq x_k \leq b_{i,k}, k = 1, 2, \dots, n]$ is the n-dimensional regular cell in R^n , and $A \subseteq \bigcup_{i=1}^{+\infty} A_i$. Thus, if A is any subset of R^n , one can collect many sets of n-dimensional regular cells $\{A_i\}$ to cover A . Among them, the outer measure denotes the set whose union has the smallest n-dimensional Lebesgue volume.

~~Both $I(A)$ and $I^n(A)$ are measures because they satisfy the two conditions of a measure. However, the outer measure $m^n(A)$ is not a measure because it is not countably additive. Fortunately, almost all the observed footprints and model units are finite and closed; therefore, as an alternative~~To match the two conditions of a measure, one can define the outer measure m^n as a **Lebesgue measure** on measure spaces $(R^n, \mathcal{L}^n, m^n)$, where \mathcal{L}^n is the **Lebesgue σ -algebra** of R^n . The construction of the Lebesgue σ -algebra is based on the Caratheodory condition (Bartle, 1995, definition 13.3). Fortunately, almost all of the observation footprints and model units are finite and closed, **leading** them to be Lebesgue measurable. This consequently ensures the Lebesgue measure m^n is a measure and the triple $(R^n, \mathcal{L}^n, m^n)$ is a measure spaces. The Lebesgue measure of a Lebesgue measurable ~~any~~ subset in R^n also coincides with its volume.

The n-dimensional **Lebesgue integral** in $(R^n, \mathcal{L}^n, m^n)$ is $\int f dm^n$, where f is a real function on R^n . The Lebesgue integral can be further denoted by $\int f dm^n = \int f(x) dx$, where $x \in R^n$ and $x = (x_1, \dots, x_n)$.

In the two-dimensional case ($n = 2$), the Lebesgue integral is

$$\iint_A f(x_1, x_2) dx_1 dx_2,$$

where $A \in \mathcal{L}^2$. Next, we consider the **Lebesgue integration by substitution** on R^2 . Let $T(x_1, x_2) = [t_1(x_1, x_2), t_2(x_1, x_2)] = [y_1, y_2]$ be a one-to-one mapping of a subset X onto another subset Y on R^2 . Assuming that T is continuous and has a continuous partial derivative matrix $T_x = \begin{pmatrix} \partial t_1 / \partial x_1 & \partial t_1 / \partial x_2 \\ \partial t_2 / \partial x_1 & \partial t_2 / \partial x_2 \end{pmatrix}$, then

$$5 \quad \iint_Y f(y_1, y_2) dy_1 dy_2 = \iint_X f(T(x_1, x_2)) |J(x_1, x_2)| dx_1 dx_2,$$

where the Jacobian determinant $|J(x_1, x_2)| = |\det T_x| = \begin{vmatrix} \partial t_1 / \partial x_1 & \partial t_1 / \partial x_2 \\ \partial t_2 / \partial x_1 & \partial t_2 / \partial x_2 \end{vmatrix}$. If T is linear, the integral reduces to

$$\iint_Y f(y_1, y_2) dy_1 dy_2 = |J(x_1, x_2)| \iint_X f(T(x_1, x_2)) dx_1 dx_2.$$

By so, any observation footprint or model unit can be regarded as a Lebesgue measurable subset in a two-dimensional space R^2 .

Additional details regarding measure theory can be found in the literature (for example, Billingsley, 1986; Bartle, 1995).

10 2.2 Stochastic calculus

We ~~then introduce~~ ~~have incorporated~~ some necessary concepts and theorems of stochastic calculus. All the classic theorems ~~have been~~ ~~are~~ introduced without proofs; their detailed derivations can be found in the literature (Itô, 1944; Karatzas et al., 1991; Shreve, 2005).

~~Compared to ordinary differential and integral calculus,~~ **stochastic calculus** is defined for ~~integrals of stochastic processes~~ ordinary integrals with respect to stochastic processes, ~~such as Brownian motion.~~ One of the simplest stochastic processes is Brownian motion ~~is one of the simplest stochastic processes.~~ The Brownian motion W that is defined on a probability measure space (Ω, F, \mathbb{P}) is characterized as follows:

- 1) $W(0) = 0$.
- 2) $\forall t_1 \neq t_2 > s_1 \neq s_2 \geq t_2 > s_2 > 0$, the increments $W(t_1) - W(s_1)$ and $W(t_2) - W(s_2)$ are independent.
- 20 3) $\forall t > s \geq 0$, $W(t) - W(s) \sim N(0, t - s)$.

The last two conditions represent that $\forall t_2 > s_2 \geq t_1 > s_1 \geq 0$, $W(t_2) - W(s_2)$ and $W(t_1) - W(s_1)$ are independent Gaussian random variables. Additionally, ~~Brownian motion is based on a probability measure space,~~ ~~so~~ W is related to the probability measure \mathbb{P} .

Stochastic calculus based on Brownian motion produces Ito process. The differential form of the time-dependent **Ito process** is

$$25 \quad dl = \varphi(t)dt + \sigma(t)dW(t), \quad (1)$$

where $\varphi(t)$, $\sigma(t)$ and $W(t)$ are the ~~transition probability~~drift rate, volatility rate and Brownian motion, respectively. The integral form of Eq. (1) is

$$I(t) = I(0) + \int_0^t \varphi(u) du + \int_0^t \sigma(u) dW(u). \quad (2)$$

Theorem 1: For any Ito process defined as in Eq. (1), the **quadratic variation** that is accumulated on the ~~scale~~ interval $[0, t]$ is—

$$[I, I](t) = \int_0^t \sigma^2(u) du, \quad (3)$$

and the **drift** of Eq. (1) is $I(0) + \int_0^t \varphi(u) du$.

As distinguishing features of stochastic calculus, quadratic variation and drift can be regarded as stochastic versions of the variance and expectation, respectively. That is, the variance and expectation are instances of their random variable stochastic counterparts within a certain integral path. Therefore, rather than being constants, quadratic variation and drift are given in terms of probability. ~~The quadratic variation is expressed by the second-order variation of a stochastic process, which consequently is 0 in a continuous differentiable random variable. Equation (3) relies on the volatility $\sigma^2(u)$; thus, the quadratic variation varies with the integration path. In addition, a general expression occurs when the integral path is random; that is, Eq. (2) is the curvilinear integral $I(t) = I(0) + \int_L \varphi(u) du + \int_L \sigma(u) dW(u)$, where L is an arbitrary path from 0 to t .~~

Theorem 2 (Ito's Lemma): If the partial derivatives of function $f(u, x)$, viz. $f_u(u, x)$, $f_x(u, x)$ and $f_{xx}(u, x)$ are defined and continuous, and if $t \geq 0$, we have

$$f(t, x(t)) = f(0, x(0)) + \int_0^t f_u(u, x(u)) du + \int_0^t f_x(u, x(u)) \sigma(u) dW(u) + \int_0^t f_x(u, x(u)) \varphi(u) du + \frac{1}{2} \int_0^t f_{xx}(u, x(u)) \sigma^2(u) du. \quad (4)$$

Ito's Lemma is typically used to build the differential of a stochastic model with Ito processes. In this section study, Ito's Lemma is applied to study the scale-dependent relationship between the observations operator and state, as well as and the uncertainties errors that are caused by scale transformation in the analysis step update of the state by introducing new observations.

2.3 Traditional formulation of data assimilation in the Bayesian theorem framework

We use the widely well accepted Bayesian theorem of data assimilation (Lorenc, 1995; van Leeuwen, 2015) to investigate its time- and scale-dependent errors. State and observation are first assumed to be one-dimensional. In Sect. 3.5, the results are extended to n-dimensional state vectors and observation vectors.

Consider a nonlinear forecasting system that is can be described by

$$X(t_k) = M_{k-1:k}(\cdot) X(t_{k-1}) + \eta(t_k), \quad (5)$$

where $M_{k-1:k}(\cdot)$ $X(t_k)$ and $\eta(t_k)$ represent a nonlinear forecasting operator that transits the state from the discrete time $k - 1$ to k , the state with prior probability distribution function (PDF) $p(X)$, and the model error at time k , respectively.

If a new observation is available at time k , the observation system is given by

$$Y^o(t_k) = H_k(X(t_k)) + \varepsilon(t_k), \quad (6)$$

where $H_k(\cdot)$, $Y^o(t_k)$ and $\varepsilon(t_k)$ represent the nonlinear observation operator, true observation with prior PDF $p(Y)$, and observation error at time k , respectively.

5 Previous studies (e.g., Janjić and Cohn, 2006; Bocquet et al. 2011) described the origin of the components of $\varepsilon(t_k)$ and $\eta(t_k)$, such as white noise, the discretization error of a continuum model, the errors that are caused by missing physical processes, and scale-dependent bias. In this study, we assume that both forecasting and observation operators are ~~derived from a perfect models~~, so errors that are caused by missing physical processes are discarded.

According to Bayesian theory, the posterior PDF of the state based on the addition of a new observation into the system is

$$p(X|Y) = p(Y|X)p(X)/p(Y), \quad (7)$$

10 where $p(X|Y)$ is the posterior PDF that presents the PDF value of state X given an available observation Y . $p(Y|X)$ is a likelihood function, which is the probability that an observation is Y given a state X . $p(X)$ and $p(Y)$ are the prior PDF values of the state and observation, respectively. Here, $p(X)$ is supposed to be known and $p(Y)$ is a normalisation constant (van Leeuwen, 2014). ~~So Therefore~~. The aim of data assimilation is equivalent to finding the posterior PDF $p(X|Y)$.

15 3 Reformulation of scale transformation in data assimilation framework

3.1 Definition of scale

We define the scale based on the measure theory that was introduced in ~~section Sect. 2. The relationship between Lebesgue measure in $(R^2, \mathcal{L}^2, m^2)$ and scale is firstly introduced by the following measures of Earth observations. The following measures of Earth observations are considered to connect the Lebesgue measure in $(R^2, \mathcal{L}^2, m^2)$ and scale.~~

- 20 (i) Measure of a single point ~~measurement observation~~: When the observation footprint is very small and homogeneous, we assume that its footprint approaches zero and its measure is accordingly zero under the condition of the Lebesgue measure. ~~However, in the real world, the volume of the observation footprint is not zero; thus, any single point measurement with an absolute zero measure is just an ideal assumption.~~
- (ii) Measure in a line: The measure is a one-dimensional Lebesgue measure.
- 25 (iii) Measure of a rectangular pixel (for example, remote sensing observation): $\forall A = [x: a_k \leq x_k \leq b_k, k = 1, 2]$, it is a two-dimensional Lebesgue volume, i.e., $\mu_{iii}(A) = I^2(A) = \prod_{k=1}^2 (b_k - a_k)$.

(iv) Measure of a footprint-~~scale~~ ~~measurement~~ observation: The ~~observed space of a footprint-scale measurement observation~~ is any bounded closed domain A , ~~most of which are is not necessary to be~~ regular rectangles, ~~such as but~~ as circles or ellipses. We use Lebesgue measure on R^2 , i.e., $\mu_{iv}(A) = m^2(A) = \inf\left\{\sum_{i=1}^{+\infty} I^2(A_i)\right\}$, where $A_i = [x: a_{i,k} \leq x_k \leq b_{i,k}, k = 1,2]$ and $A \subseteq \bigcup_{i=1}^{+\infty} A_i$. Obviously, measure (i)~(iii) are the special cases of the measure of a footprint-~~scale~~ ~~measurement~~ observation.

Actually, all the above measures mainly depend on the shape and size of A . The Lebesgue measure on R^2 coincides with the area, so the Lebesgue integral of $\mu_{iv}(A)$ is $\iint_A dx_1 dx_2$, where the real function $f \equiv 1$.

Now, we can generalize the above examples by defining the **scale** as the Lebesgue measure with respect to the observation footprint. This definition can also be extended to a certain model unit, ~~which could be a point, a rectangular grid, or an irregular unit such as a response unit (watershed, land coverscape patch and so on)~~. Thus, for any subset $A \in \mathcal{L}^2$, the scale is $s = m^2(A) = \iint_A dx_1 dx_2$, where the real function $f \equiv 1$. From a geometric perspective, the measure function $m^2(\cdot)$ refers to the shape of the subset, and the scale further indicates its size.

We represent the scale as s , and let $s_0 = m_0^2(A_0) = \iint_{A_0} dx_1 dx_2 = 1$ be the **standard scale**, where $A_0 = [x: 0 \leq x_k \leq 1, k = 1,2]$ is ~~a unit interval~~ the unit square in R^2 . ~~The standard scale can be regarded as a basic unit of scale-~~ It presents a standard reference, by which one can make a quantitative comparison between different scales. The standard scale is also the origin of scales that let scales vary similarly to other physical quantities, such as time.

We can further define **scale transformation**. For $\forall A_1, A_2 \in \mathcal{L}^2$, if there are two different scales, $s_1 = m^2(A_1) = \iint_{A_1} dx_1 dx_2$ and $s_2 = m^2(A_2) = \iint_{A_2} dy_1 dy_2$, then we can obtain $s_2 = \iint_{A_2} dy_1 dy_2 = \iint_{A_1} |J(x_1, x_2)| dx_1 dx_2$ based on Lebesgue integration by substitution, where the Jacobian matrix $J(x_1, x_2)$ represents the geometric transformation from A_1 to A_2 . In particular, if $J(x_1, x_2) = \text{diag}(\xi, \xi), \xi \in R$, which also indicates that the geometric transformation is linear, then the following expression is valid based on Lebesgue integration by substitution:

$$s_2 = |J(x_1, x_2)| \iint_{A_1} dx_1 dx_2 = \xi^2 s_1, \quad (\S 8)$$

where s_1 and s_2 represent a-the change of one-dimensional rule-change.

If two scales follow the one-dimensional rule, they are geometrically similar. This rule simplifies scale as a one-dimensional variable that corresponds to the scale differences-transformations between most remote sensing images with various spatial resolutions. For example, $\forall A = [x: a \leq x_k \leq b, k = 1,2]$, where A and the unit square interval A_0 are geometrically similar, and the scale $s = \mu_{iii}(A)$ can be expressed by the one-dimensional rule of scale transformation: $s = \mu_{iii}(A) = |J(x_1, x_2)| \iint_{A_0} dx_1 dx_2 = (b - a)^2 s_0$. For another example, let $s = \iint_A dy_1 dy_2$ be a disc measure the scale, where A is of a disc observation—footprint A —with radius r . The mapping function between A and A_0 is $T(x_1, x_2) = [rx_1 \cos(2\pi x_2), rx_1 \sin(2\pi x_2); 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1] = [y_1, y_2]$, and the Jacobian determinant $|J(x_1, x_2)| = \begin{vmatrix} r \cos(2\pi x_2) & -2\pi r x_1 \sin(2\pi x_2) \\ r \sin(2\pi x_2) & 2\pi r x_1 \cos(2\pi x_2) \end{vmatrix} = 2\pi r^2 x_1$. Therefore, $s = \iint_A dy_1 dy_2 = \iint_{A_0} |J(x_1, x_2)| dx_1 dx_2 = \pi r^2 s_0$, which is equal to its area. However, s_0 and s do not obey one-dimensional rule because the Jacobian matrix is not diagonal.

The Layer 1 in Figure 1 shows the relationship between the Lebesgue measure and scale. The measure space $\Omega = [x: 0 \leq x_k \leq 4, k = 1,2]$ is regularly divided by the unit square interval A_0 . Let scales $s_{C1} = m_{C1}^2(C1)$, $s_{C2} = m_{C2}^2(C2)$ and $s_{C3} = m_{C3}^2(C3)$ be the Lebesgue measures of disc measurements observation footprints C_1, C_2 and C_3 , respectively, and let m_{D1}^2 and m_{D2}^2 be the Lebesgue measures of diamond measurements D_1 and D_2 . Then, $m_{C1}^2(\cdot) = m_{C2}^2(\cdot) = m_{C3}^2(\cdot)$ because they are the same Lebesgue measure functions. That is, if $\{A_i\}$ is the set with the smallest volume that covers C_1 , then similar sets $\{A_i + 2\}$ and $\{A_i \times 3 + 2\}$ can be used (with the origin located in the upper-left corner) to cover C_3 and C_2 with the smallest volumes, respectively. Here, $A_i + 2 = [x_i \neq: x_{i,k} + 2, x_{i,k} \in A_i, k = 1,2]$ and $A_i \times 3 + 2 = [x_i \neq: x_{i,k} \times 3 + 2, x_{i,k} \in A_i, k = 1,2]$, which proves that functions $m_{C1}^2(\cdot)$, $m_{C2}^2(\cdot)$ and $m_{C3}^2(\cdot)$ collect the desirable set based on the same scheme, so they are identical. Additionally, $s_{C2} = m_{C2}^2(C2) = \sum I^2(A_i \times 3 + 2)$ is much larger than $s_{C1} = m_{C1}^2(C1) = \sum I^2(A_i)$ and $s_{C3} = m_{C3}^2(C3) = \sum I^2(A_i + 2)$. Therefore, the scale of C_2 is not equal to the two other scales because the volumes of their subsets are different. However, their scales are governed by one-dimensional rules because their measures are identical and the Jacobian matrices between them are diagonal. Similarly, $m_{D1}^2 = m_{D2}^2$; although their scales are different, they obey a one-dimensional rule.

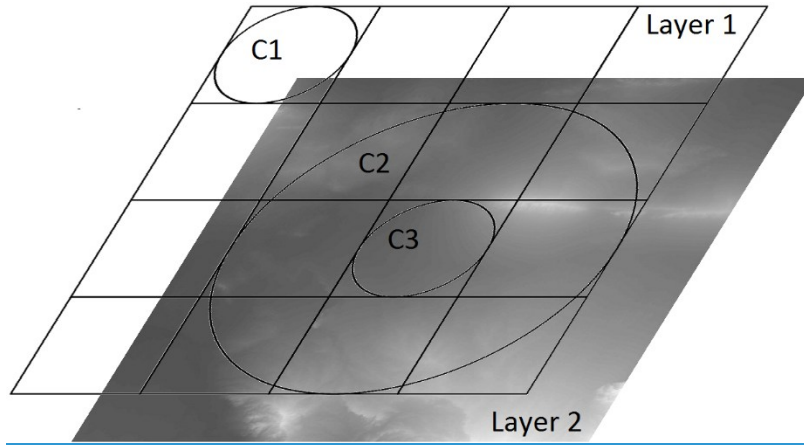


Figure 1. Diagram of the Relationships-relationships among a Lebesgue Measuremeasure, Scale-scale and Variablegeophysical variable

3.2 Stochastic variables in data assimilation

5 We introduce use the widely accepted Bayesian theorem of data assimilation (Lorenz, 1995; van Leeuwen, 2015) to investigate its time- and scale-dependent errors. We assume that both the state vector and observation vector are one-dimensional (in the following text, we use “state” and “observation” for brevity). In Sect. 3.4, the results are extended to n-dimensional state vectors and observation vectors.

— Consider a nonlinear forecasting system that is described by

$$10 \quad X(t_k) = M_{k-1:k}(X(t_{k-1})) + \eta(t_k), \quad (6)$$

where $M_{k-1:k}(\cdot)$, $X(t_k)$ and $\eta(t_k)$ represent a nonlinear forecasting operator that transits the state from the discrete time $k-1$ to k , the state with prior PDF $p(X)$, and the model error at time k , respectively. In addition, if a new observation is available at time k , the observation system is given by

$$15 \quad Y^o(t_k) = H_k(X(t_k)) + \varepsilon(t_k), \quad (7)$$

where $H_k(\cdot)$, $Y^o(t_k)$ and $\varepsilon(t_k)$ represent the nonlinear observation operator, true observation with prior PDF $p(Y)$, and observation error at time k , respectively.

— Previous studies (e.g., Janjić and Cohn, 2006; Boequet et al. 2011) discovered the components of $\varepsilon(t_k)$ and $\eta(t_k)$, such as white noise, the discretization error of a continuum model, the errors that are caused by missing physical processes, and scale-dependent bias. In this study, we assume that both forecasting and observation operators are derived from a perfect model, so the discretization errors and errors that are caused by missing physical processes are discarded.

— According to Bayesian theory, the posterior PDF of the state based on the addition of a new observation into the system is

$$20 \quad p(X|Y) = p(Y|X)p(X)/p(Y), \quad (8)$$

where $p(X|Y)$ is the posterior PDF that presents the PDF value of state X given an available observation Y . $p(Y|X)$ is a likelihood function, which is the probability that an observation is Y given a state X . $P(X)$ and $P(Y)$ are the prior PDF values of the state and observation, respectively. Here, $P(X)$ is supposed to be known and $P(Y)$ is a normalisation constant (van Leeuwen, 2014). The aim of data assimilation is equivalent to finding the posterior PDF $p(X|Y)$.

5 Instead of using Eq. (65) and (76), which are discrete in time, we use Ito process-formed expressions with the one-dimensional infinitesimals ds and dt to formulate a continuous-time (or continuous-scale) state and observation.

Geophysical variable can be regarded as Let a real function $V(s, t)$ be the variable if, and it maps the space $(R^2, \mathcal{L}^2, m^2)$ onto R , where s is the scale, $s = m^2(A)$, $A \in \mathcal{L}^2 R^2$, and t is the time. In n-dimensional data assimilation, a geophysical variable V is related to an element of state vector X at a specific scale s and time t . A variable is an estimation of a geophysical parameter at a specific scale s and time t .

10 In Figure 1, Layer 2 presents a heterogeneous geophysical variable in the entire region. If aggregating in Layer 2 into Layer 1 and let ~~the~~ each pixel intensity is the estimator of value for a geophysical variable parameter in each pixel, then this parameter the measure space Ω is heterogeneous in the entire region. A variable geophysical variable represents an ensemble spatial average in a specific observation footprint with a specific scale. Therefore, the variable geophysical variables in C1 and C3 are
 15 not equal because their observation footprints are different, and the variable geophysical variables in C2 and C3 are also different because the scale changes. The former introduces that the variable geophysical variables that vary with location, and the latter states that the variable geophysical variables are scale-dependent. Therefore, from an Earth observation perspective, a variable is a nonlinear and heterogeneous mapping function of observation footprints onto R at a given scale.

20 The dynamic process of the variable clearly depends on time, and we further assume that the variable varies with scale in view of the scale issue. Furthermore, assuming It is reasonable to a that the variable is random both in time and scale is reasonable because of the uncertainties in Earth observations and simulations. Therefore, if the statistical properties of the variable geophysical variable are available, we can construct an explicit stochastic equation for the variable geophysical variable.

—We introduce the time-dependent Ito process Eq. (1) to define the variable geophysical variable process:

$$dV = p(t)dt + q(t)dW(t). \quad (9)$$

25 Similarly, the variable geophysical variable is supposed to evolve via a stochastic process, for which the dynamic process and uncertainty are allowed to vary with scale:

$$dV = \varphi(s)ds + \sigma(s)dW(s), \quad (10)$$

where $\varphi(s)$ and $\sigma(s)$ are the scale-based transition probability drift rate and volatility rate, respectively. The variable geophysical variable is a probabilistic process with respect to scale and thus has scale-dependent errors, where the scale should shift forward or backward based on the condition that the scale follows the one-dimensional rule.

30 First, time is one dimensional and unidirectional, but the scale can shift forward or backward based on the condition that the scale follows the one dimensional rule. Second, Eq. (10) implies that the value and variance of a variable may change if the scale changes. As discussed in Sect. 1, evaluating the heterogeneity in an integrated study is more difficult than in a special

ease study. However, in Eq. (10), one can track a special scale path to obtain the quadratic variation and drift, which indicate the heterogeneity of the variable.

Eq. (9) can be regarded as a continuous-time version of Eq. (5), i.e., to estimate the state is equal to the integral of Eq. (9) over a time interval. Here $p(t)$ indicates the physical process with respect to time, and $q(t)$ is the error only caused by the evolution of time, so model error η in Eq. (5) contains more parts than $q(t)$. Eq. (10) implies that the value and variance of a geophysical variable may change if the scale changes. To formulate $\varphi(s)$ should consider both the spatial heterogeneities and physical process variations among different scales. However, neither of them is well understood in a general theory study. Therefore $\varphi(s)$ is conceptualized in Eq. (10). EspeciallyParticularly, if the study region is ~~mean~~-homogeneous, then the values of a variable that observed at the same place are ~~much~~ identical between large scale and fine scale, and $\varphi(s)$ can be left out. $\sigma(s)$ is the error caused by the scale transformation.

Comparing Eq. (6) and Eq. (9) shows that $M_{k-1:k}(\cdot)$ and $\eta(t_k)$ are associated with $p(t)$ and $q(t)$. The variables in a data assimilation forecasting model The state in the forecasting step can be expressed by Eq. (9) because only time is involved. In the analysis step of data assimilation, the state does not pertain to time, and we assume that the scale has a quantifiable effect on the ~~uncertainties~~-errors in this step; thus, both the states and observations can be defined by Eq. (10). We will try to use this assumption in the following sections.

3.3 Expression of scale transformation in a stochastic data assimilation framework

First, we provide the following lemma.

Lemma 1: For $\forall s_0 > 0$, let $W^*(0) = W(s_0) - W(s_0), \dots, W^*(s) = W(s_0 + s) - W(s_0)$; then, $W^*(s), s \geq 0$ is a Brownian motion.

Proof. First, $W^*(0) = W^*(s_0) - W^*(s_0) = 0$. $\forall s_{i+1} > s_i \geq 0, i = 1, 2, 3, \dots$, $W^*(s_{i+1}) - W^*(s_i) = [W(s_0 + s_{i+1}) - W(s_0)] - [W(s_0 + s_i) - W(s_0)] = W(s_0 + s_{i+1}) - W(s_0 + s_i)$, which suggests that the increments $[W^*(s_{i+1}) - W^*(s_i)]$ are equal to $[W(s_0 + s_{i+1}) - W(s_0 + s_i)]$ and are independent Gaussian distributed. Therefore, $W^*(s), s \geq 0$ is a Brownian motion, with $E[W^*(s_{i+1}) - W^*(s_i)] = 0$ and $Var[W^*(s_{i+1}) - W^*(s_i)] = s_{i+1} - s_i$. Q. E. D.

Remark on Lemma 1: This Lemma is practical because the scale is greater than zero, which does not fit the definition of Brownian motion, whereby the parameter should start at zero. The standard scale s_0 is associated with zero in Lemma 1; thus, it is logical to let $s = 0$ in $W^*(s)$. Lemma 1 further implies that $W(s), s \geq s_0$ is an equivalent expression of $W^*(s), s \geq 0$. Note that in the definition of Brownian motion, the parameter ~~should~~ starts at zero. However, the scale is realistically greater than zero, which results that it cannot be directly applied in Brownian motion. So, Lemma 1 is logical because it implies that $W(s), s \geq s_0$ is an equivalent expression of $W^*(s), s \geq 0$. Therefore, beginning with the standard scale, the Brownian motion and stochastic calculus with respect to scale can be further developed.

In the following content, we use Brownian motion with a parameter that starts at s_0 to define the scale-dependent variablegeophysical variables; therefore, ~~some-the~~ classic expressions above ~~should beare~~ changed. According to Lemma 1, Eq. (3) is given by

$$[I, I](s) = \int_{s_0}^s \sigma^2(u) du. \quad (11)$$

5 Additionally, the integral form of the Eq. (10) is as follows:

$$V(s) = V_0 + \int_{s_0}^s \varphi(u) du + \int_{s_0}^s \sigma(u) dW(u), \quad (12)$$

where $V_0 = V(s_0)$ and the drift of Eq. (12) is

$$V_0 + \int_{s_0}^s \varphi(u) du.$$

Similarly, Eq. (4) becomes

$$10 \quad f(s, V(s)) = f(s_0, V(s_0)) + \int_{s_0}^s f_u(u, V(u)) du + \int_{s_0}^s f_x(u, V(u)) \sigma(u) dW(u) + \int_{s_0}^s f_x(u, V(u)) \varphi(u) du + \frac{1}{2} \int_{s_0}^s f_{xx}(u, V(u)) \sigma^2(u) du.$$

Now, we make the following assumptions.

Assumption 1: The measures of scale transformations between the state and observation ~~in spaces of~~ data assimilation obey the one-dimensional rule as defined in Sect. 3.1.

15 **Assumption 2:** In the forecasting step, the model unit equals the scale of the state space, and both of them are constant.

Assumption 3: In the analysis step, the state, observation and observation operator are scale dependent. Only one observation is added into the data assimilation system at a time, ~~and the states and observations at different times are scale independent.~~

20 In assumption 1, the one-dimensional rule ensures that scale changes in a sense of geometrical similarity (for example, form a larger square observation footprint to a smaller square observation footprint, or from C_2 to C_3 as presented in Figure 1). AlsoAdditionally, the formulations of scale transformation can be extremely reduced.

Assumption 2 indicates that the model unit and state scale are both supposed to be the same-with each other and invariant in space and time. So, there is no scale transformation in the forecasting step. Thus, Eq. (9) can adequately describe this step.

Considering assumption 2, the forecasting step is explicitly free of scale; thus, Eq. (9) can adequately describe this step.

25 Based on assumption 3, the analysis step is related to the scale; ~~thus, some basic definitions should be presented in advance.~~

According to Eq. (10), the state and observation in the analysis step are as follows:

$$dX = \varphi_X(s) ds + \sigma_X(s) dW(s) \quad (13)$$

and

$$dY = \varphi_Y(s) ds + \sigma_Y(s) dW(s), \quad (14)$$

30 where $\varphi_X(s)$, $\sigma_X(s)$, $\varphi_Y(s)$ and $\sigma_Y(s)$ represent the scale-dependent ~~transition probabilities~~ drift rates and volatilities volatility rates of state X and observation Y , respectively. $\varphi(s)$ also implies the heterogeneities and physical processes from standard scale to a specific scale, which currently maybe hard to be formulate. $\sigma(u)$ can be regarded as the stochastic

perturbation with respect to scale. ~~In terms of scale, the state and observation in the analysis step are formulated. Therefore,~~
~~According to Eq. (12), a state is $X(s_X) = X_0 + \int_{s_0}^{s_X} \varphi(u)du + \int_{s_0}^{s_X} \sigma(u)dW(u)$ in the state space and is $X(s_Y) = X_0 + \int_{s_0}^{s_Y} \varphi(u)du + \int_{s_0}^{s_Y} \sigma(u)dW(u)$ in the observation space. These formulas prove that the value of state varies with the changes of scale.~~

5 ~~Assumption 3 implies that the scales of the state and observation are invariant when observational information is added in the analysis step. The scale transformation only involves in the process that mapping the state vector from state space to observation space.~~ For simplicity, assume the scale-based ~~drift rates transition probabilities~~ of the state and observation do not exist, which leads to $\varphi_X(s) = 0$ and $\varphi_Y(s) = 0$. ~~And assuming that~~ If the noises are Gaussian, we have $\sigma_X(s) = \sigma_Y(s) = 1$.

Based on the above discussion, the differential and integral forms of the state are

$$10 \quad dX = dW(s) \text{ and } X(s_X) = X_0 + \int_{s_0}^{s_X} dW(s). \quad (15)$$

For the observation, we have

$$dY = dW(s) \text{ and } Y(s_Y) = Y_0 + \int_{s_0}^{s_Y} dW(s) \quad (16)$$

In Eq. (15) and Eq. (16), the time t is omitted, and s_X, s_Y, X_0 and Y_0 represent the scale of the state space, scale of the observation space, state in s_0 and observation in s_0 , respectively.

15 The Bayesian equation of data assimilation (Eq. (87)) produces the posterior PDF $p(X|Y)$ that is associated with the likelihood function $p(Y|X)$ and the distributions of the state and observation. Theorem 1 and Eqs. (15)~(16) yield $X \sim N\left(X_0, \int_{s_0}^{s_X} ds\right)$ and $Y \sim N\left(Y_0, \int_{s_0}^{s_Y} ds\right)$ under the condition that the variances exist. In addition, assumption 1 states that the scales vary in one-dimensional space, which results in

$$X \sim N(X_0, s_X - s_0) \quad (17)$$

$$20 \quad \text{and } Y \sim N(Y_0, s_Y - s_0). \quad (18)$$

Eq. (17) and Eq. (18) are the prior PDFs of state and observation with respect to scale in state space and observation space, respectively. Compared with the PDFs with respect to time, their expectations are equal to the value at the standard scale, and the variances depend on the differences between the standard scale and the scale in state or observation space. These two prior PDFs are introduced into the Bayesian theorem that reformulated by scale.

25 Then, we calculate the posterior PDF. Thus, the last point is to calculate $p(X|Y)$.

The scale-dependent observation operator is $H(s, X(s)x)$, which suggests that the observation operator and its parameters are both susceptible to the scale. If $H(s, X(s)x)$ is defined, its continuous partial derivatives are $H_s(s, x)$, $H_x(s, x)$ and $H_{xx}(s, x)$. In line with Ito's Lemma, we get an estimation of observation in the observation space, which is related to the state $X(s_X)$ defined in the state space have

$$30 \quad H(s_X, X(s_X)) = H(s_0, X_0) + \int_{s_0}^{s_X} H_s(u, X(u)) du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \\ = H(s_0, X_0) + \int_{s_0}^{s_X} \left[H_s(u, X(u)) + \frac{1}{2} H_{xx}(u, X(u)) \right] du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u). \quad (19)$$

Assumption 1 suggests that the observation and model spaces have the same probability measure; thus, the Brownian motions in these two spaces are equivalent. Let Eq. (16) – Eq. (19), and we obtain

$$\begin{aligned}
& Y(s_Y) - H(s_X, X(s_X)) \\
&= Y_0 + \int_{s_0}^{s_Y} dW(u) - \left[H(s_0, X_0) + \int_{s_0}^{s_X} H_s(u, X(u)) du + \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right] \\
5 \quad &= Y_0 - H(s_0, X_0) + \int_{s_0}^{s_Y} dW(u) - \left[H(s_X, X(s_X)) - H(s_0, X(s_0)) \right] - \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du - \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) \\
&= Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right] + \left\{ \int_{s_0}^{s_Y} dW(u) - \int_{s_0}^{s_X} H_x(u, X(u)) dW(u) \right\}. \quad (20)
\end{aligned}$$

Equation (20) can be regarded as an Ito process, and its drift is

$$Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right]. \quad (21)$$

The integral term in Eq. (21) is the difference in the first-order differential observation operator between the state scale s_X and the standard scale s_0 . This term illustrates that the mapping process should consider not only the observation operator but also the first-order differential term when state is mapped to the observation space. The former is typically determined from the literature, whereas the latter was derived in this study for the first time. This result prompted us to further consider the first-order differential of the observation operator when calculating the [representativeness observation](#)-error.

The quadratic variation of Eq. (20) is

$$15 \quad (s_Y - s_0) + \int_{s_0}^{s_X} H_x^2(u, X(u)) du. \quad (22)$$

This equation suggests that the uncertainty in the observation error includes both the difference between scales s_Y and s_0 and the change in the observation operator from scale s_X to s_0 . Therefore, Eq. (21) and Eq. (22) can be combined to produce

$$p(Y|X) = N \left(Y_0 - \left[H(s_X, X(s_X)) + \frac{1}{2} \int_{s_0}^{s_X} H_{xx}(u, X(u)) du \right], (s_Y - s_0) + \int_{s_0}^{s_X} H_x^2(u, X(u)) du \right). \quad (23)$$

Based on Eqs. (17), (18) and (23), $p(Y|X)$, $p(X)$ and $p(Y)$ are stochastic functions that depend on the scale; thus, the posterior PDF of the state is [scale-scale](#)-dependent as well.

In particular, if Y is a direct [observation measurement](#), which means the observation is of the same physical quantity and scale as the state, viz. $H(s, X(s)) = X(s)$. The result becomes

$$Y(s_Y) - X(s_X) = \begin{cases} Y_0 - X(s_X) + W(s_Y) - W(s_X), & s_Y > s_X \\ Y_0 - X(s_X) + W(s_X) - W(s_Y), & s_X > s_Y \end{cases} \quad (24)$$

$$\text{and } p^{\mathcal{P}}(Y|X) = N\{Y_0 - X(s_X), |s_Y - s_X|\}. \quad (25)$$

25 The quadratic variation in Eq. (22) can be further described by the scale from s_X to s_Y . Under the condition ~~that~~ $s_Y > s_X$ and because $W(s_Y) - W(s_X)$ and $W(s_X) - W(s_0)$ are independent, the quadratic variation of Eq. (20) is

$$s_Y - s_X + \int_{s_0}^{s_X} [1 - H_x(u, X(u))]^2 du. \quad (26)$$

Similarly, if $s_X > s_Y$, the quadratic variation of Eq. (20) is

$$\int_{s_0}^{s_Y} (1 - H_x(u, X(u)))^2 du + \int_{s_Y}^{s_X} H_x^2(u, X(u)) du. \quad (27)$$

The significance of Eqs. (20)~(27) is that the effect of scale on the posterior PDF can be determined quantitatively. In addition to the model error and measurement instrument error (both of them were not introduced explicitly in this study because they have little influence on the error caused by scale transformation), a new type of error in data assimilation was discovered in the analysis step. The expectation of the posterior PDF may vary with the scale of the state space if Y is an indirect measurement of X observation, and the variance of the drift depends on the difference between s_Y and s_X (based on Eq. (26) and Eq. (27)) or among s_0 , s_Y and s_X (based on Eq. (22)). In addition, if Y is a direct measurement of X observation (Eq. (24) and Eq. (25)), the expectation of the posterior PDF is the difference between Y and X , and the variance is equal to the increment of Brownian motion with respect to the scale. Additionally, if the results are not derived from assumption 1, i.e., the measure scale varies randomly, the posterior PDF is more complex because the Jacobian matrix in Lebesgue integration of scale transformation is arbitrary-its integral path is an arbitrary curve.

— However, a problem still exists. If the initial state is not at the scale of the forecasting operator, the corresponding error should also be considered. Similarly, if the forecasting operator $M(s, X(t, s))$ has continuous partial derivatives $M_s(s, x)$, $M_x(s, x)$ and $M_{xx}(s, x)$, then according to Ito's Lemma, we have

$$\begin{aligned} M(s, X(s)) &= M(s_0, X_0) + \int_{s_0}^s M_s(u, X(u)) du + \int_{s_0}^s M_x(u, X(u)) dW(u) + \frac{1}{2} \int_{s_0}^s M_{xx}(u, X(u)) du \\ &= M_0 + \int_{s_0}^s \left[M_s(u, X(u)) + \frac{1}{2} M_{xx}(u, X(u)) \right] du + \int_{s_0}^s M_x(u, X(u)) dW(u) \end{aligned} \quad (28)$$

Assume that the initial state is $X(s_t)$, where s_t is its scale, and $X(s_x)$ is the ideal initial state in the model space that is related to $X(s_t)$. Then, $X(s_x)$ has the same scale s_x as the forecasting operator. From Eq. (28) we obtain the error:

$$\begin{aligned} &M(s_x, X(s_t)) - M(s_x, X(s_x)) \\ &= M_0 + \int_{s_0}^{s_x} M_s(u, X(u)) du + \int_{s_0}^{s_t} \frac{1}{2} M_{xx}(u, X(u)) du + \int_{s_0}^{s_t} M_x(u, X(u)) dW(u) \\ &= \left[M_0 + \int_{s_0}^{s_x} M_s(u, X(u)) du + \int_{s_0}^{s_x} \frac{1}{2} M_{xx}(u, X(u)) du + \int_{s_0}^{s_x} M_x(u, X(u)) dW(u) \right] \\ &= \int_{s_x}^{s_t} \frac{1}{2} M_{xx}(u, X(u)) du + \int_{s_x}^{s_t} M_x(u, X(u)) dW(u) \end{aligned} \quad (29)$$

where $M(s_x, X(s_t))$ and $M(s_x, X(s_x))$ denote the next states that are associated with the true initial state and the ideal initial state, respectively. Based on Eq. (29), the error is an Ito process with a transition probability as the second order differential forecasting operator, and a volatility as the first order differential forecasting operator. Both of these operators are integrated from s_x to s_t .

3.4 Examples: the stochastic radiative transfer equation (SRTE)

To explicitly show how the stochastic scale transformations impact on assimilation, we introduce an illustrative example based on the scales presented in Figure 1 and the above results. Assuming that in the analysis step, the state is with the standard

scale s_0 , whose observation footprint is the unit square A_0 . If the scale of observation space is s_{C1} and its observation footprint is the disc C_1 , then ~~according to the statements in Sect. 3.1~~, the Jacobian matrix of the transformation between the scales of state space and observation space is not diagonal ~~according to the statements in Sect. 3.1~~, ~~which~~ leading the two scales do not obey the one-dimensional rule and ~~go~~ against assumption 1. However, if let the scales of of state space and observation space

5 are s_{C1} and s_{C2} , respectively, the assumption 1 is met and it can be counted that $s_X = s_{C1} = \frac{\pi}{4}s_0$ and $s_Y = s_{C2} = \frac{9\pi}{4}s_0$.

Now the scales of state space and observation space obey the one-dimensional rule, and then we further presume that the measure space Ω in Figure 1 is free of the spatial heterogeneities and dynamic process variations depending on scale. This is

Consequently, the drift rate $\varphi(s) = 0$. If denoting the value of state in the standard scale is X_0 , then ~~according to Eq. (19)~~, the

prior PDF of state is $X \sim N\left(X_0, \frac{\pi}{4}s_0 - s_0\right)$ according to Eq. (1917). , ~~Noting that $\frac{\pi}{4}s_0 - s_0$ in here~~ is not a real number and

10 only indicates the variation when the scale change ds.

If $H(s, X(s)) = X(s)$, the observation is the same physical quantity as the state, and according to Eq. (25), the likelihood

function is $p(Y|X) = N\{Y_0 - X(s_X), |s_Y - s_X|\} = N\{Y_0 - X(s_X), |s_{C2} - s_{C1}|\} = N\left\{Y_0 - X(s_X), \left|\frac{9\pi}{4}s_0 - \frac{\pi}{4}s_0\right|\right\}$.

To formulate the likelihood function in the case that the observation is different from the state, the SRTE will be employed

in the following text. The SRTE is a stochastic integral-differential equation that describes the radiative transfer phenomena

15 through a stochastically mixed immiscible media. Scientists have developed analytical or numerical methods for finding the

stochastic moments of the solution, such as the ensemble-averaged or variance of the radiation intensity (Pomraning, 1998;

Shabanov et al., 2000; Kassianov et al., 2011).

Consider the general expression of the SRTE (leave out the scattering and emission ~~source~~),

$$-\mu \frac{dI(\tau)}{d\tau} = -I(\tau), \quad (28)$$

20 where $I(\tau)$, μ and τ are the radiation intensity, coefficient of radiation direction and optical depth, respectively. The analytical

solution of Eq. (28) is $I(\tau) = -I(0)e^{\tau/\mu}$.

To tie into more substantial random optical properties of transfer media, such as absorption and scattering, the optical depth

τ is assumed to be stochastic. So it suggests that optical depth is a scale-dependent Ito process ~~and can be expressed as~~

$$d\tau(s) = \varphi_\tau(s)ds + \sigma_\tau(s)dW(s), \quad (29).$$

25 ~~which~~ This causes the radiation intensity depend on scale as well (Nevertheless, we assume that the relation between radiation intensity and optical depth is scale invariant).

SRTE can be considered as a concrete instance of stochastic observation operator by defining $H(s, x(s)) = I(x) = I(0)e^{x/\mu}$. Therefore, $H_s(s, x(s)) = 0$, $H_x(s, x(s)) = \frac{1}{\mu}I(0)e^{x/\mu}$ and $H_{xx}(s, x(s)) = \frac{1}{\mu^2}I(0)e^{x/\mu}$. Based on Ito's Lemma,

$$\begin{aligned}
 dI(\tau(s)) &= dH(s, \tau(s)) = H_s(s, \tau(s))ds + H_x(s, \tau(s))d\tau(s) + \frac{1}{2}H_{xx}(s, \tau(s))d\tau(s)d\tau(s) \\
 &= \frac{1}{\mu}I(0)e^{\tau(s)/\mu}d\tau(s) + \frac{1}{2\mu^2}I(0)e^{\tau(s)/\mu}d\tau(s)d\tau(s) \\
 &= \frac{1}{\mu}I(\tau(s))d\tau(s) + \frac{1}{2\mu^2}I(\tau(s))d\tau(s)d\tau(s) \quad . \quad (3230) \\
 &= \left(\frac{1}{\mu}I(\tau(s)) \right) \sigma_\tau(s)dW(s) + \left(\frac{1}{\mu}I(\tau(s)) \right) \varphi_\tau(s)ds + \left(\frac{1}{2\mu^2}I(\tau(s)) \right) \sigma_\tau^2(s)ds \\
 &= \left(\frac{\sigma_\tau(s)}{\mu} \right) I(\tau(s))dW(s) + \left(\frac{\sigma_\tau^2(s)}{2\mu^2} + \frac{\varphi_\tau(s)}{\mu} \right) I(\tau(s))ds
 \end{aligned}$$

5 Radiation intensity is a scale-dependent Ito process. The difference between Eq. (30) and the general Ito process is that there is a primitive function $I(\tau(s))$ in the integral term. Therefore, the uncertainty of the radiation intensity is more complex because it is related to both the change of scale and the primitive function.

Integrating both sides of Eq. (30) yields the general solution of the radiation intensity.

$$I(\tau(s)) = C \cdot \exp \left[\int \left(\frac{\sigma_\tau(s)}{\mu} \right) dW(s) + \int \left(\frac{\sigma_\tau^2(s)}{2\mu^2} + \frac{\varphi_\tau(s)}{\mu} \right) ds \right] \quad (31)$$

10 where the constant $C \in R$. Eq. (31) further indicates that $I(\tau(s))$ is a scale-dependent Ito process. Considering that the optical depth τ is the state, the radiation intensity I is the observation and $I(\tau(s))$ is the observation operator, then the above results in Sect. 3.3 (For example, Eq. (20)) could be easily applied here to study the posterior PDF of data assimilation.

3.4.5 Extension to n-dimensional data assimilation

15 In the above discussion, we assumed that only one variable-state existed in data assimilation. However, numerous states typically exist. This section further introduces the product spaces to extend the one-dimensional stochastic data assimilation to n-dimensions.

Assume that the independent states X_k are the variables of the measure spaces $(\Omega_k, \mathcal{F}_k, \mu_k)$, $k = 1, 2, \dots, n$, and $(\Omega^n, \mathcal{F}^n)$ is the product space, where $\Omega^n = \prod_{k=1}^n \Omega_k$ and $\mathcal{F}^n = \prod_{k=1}^n \mathcal{F}_k$. According to Fubini's theorem (Billingsley, 1986), only one product measure μ^n in $(\Omega^n, \mathcal{F}^n)$ exists, such that $\mu^n(\prod_{k=1}^n A_k) = \prod_{k=1}^n \mu_k(A_k)$, where $A_k \in \mathcal{F}_k$.

We define the state vector $X^n = (X_1, X_2, \dots, X_n)^T$ as a variable vector of the product measure space $(\Omega^n, \mathcal{F}^n, \mu^n)$. In particular, if all the scales obey the one-dimensional rule, we have

$$\mu^n \left(\prod_{k=1}^n A_k \right) = \prod_{k=1}^n \xi_k^2 \mu_0(A_k) = \left(\prod_{k=1}^n \xi_k \right)^2 \mu_0^n \left(\prod_{k=1}^n A_k \right).$$

This expression proves that the product measure also obeys a one-dimensional rule. However, the above results may not hold without the assumption that the states X_k are independent.

As discussed in Sect. 2.1, the Lebesgue measure m^2 is a measure and the triple $(R^2, \mathcal{L}^2, m^2)$ is a measure space. Therefore, the above extension is reasonable ~~in our study~~.

~~This~~ The analysis of a single state can also be applied to finite multiple states in the product measure space.

~~4~~ Summary & Discussion

10 ~~4.1~~ Summary

~~In this study, we mainly addressed two basic problems associated with scale transformation in earth observation and simulation. First, we produced a mathematical formalism of scale by employing measure theory. Second, we demonstrated how scale transformation and associated uncertainties could be evaluated presented in a data assimilation framework. Instead of using empirical and qualitative expressions, we employed measure theory and stochastic calculus to define the scale and the evolutions of errors with respect to scale in data assimilation.~~

~~The first problem began with an introduction to measure theory. We revealed that the scale is the Lebesgue measure with respect to the observation footprint or model unit. Scale is related to the shape and size of a space, and scale transformation depends on the spatial change between different scales. The definition of scale transformation is as important as that of scale. This definition was described using a Jacobian matrix and could be further simplified using the one-dimensional rule to suit stochastic calculus. This simplification is reasonable for a large portion of Earth observation data, including remote sensing data, because the scale transformations of those data are geometrically similar. However, an in-depth and comprehensive exploration should be conducted in the future to describe other~~

situations in the real world. We then defined the variable, which further considers the heterogeneities of geophysical parameters. A variable consequently expresses the ensemble average of a geophysical parameter at a specific scale.

—For the second problem, wWe reformulated the expression of scale transformation and investigated the error structure that is caused by scale transformation in data assimilation using basic theorems of stochastic calculus. The new error further Our findings supported previous qualitative knowledge that the observation error is highly related to changes in scale. Understanding the uncertainty of data assimilation based on separating the scale-dependent error from other errors is beneficial. The results can be derived from the one-dimensional simplification of scale transformation, and the variables in data assimilation evolve regularly based on assumptions 1-3. However, these situations may be more complex in the real world.

4.2 Discussion

—Our approach is different from previous work in the literature that studied representativeness error (e.g. Boequet et al., 2011; van Leeuwen, 2014; Hodyss and Nichols, 2015). The basic concept of these studies was to assume that a relationship exists between different variables or operators, and then the relationship was introduced in the Bayesian expression of data assimilation to find the corresponding representativeness error.

—Compared to previous work, our study is significant both in employing rigorous mathematical knowledge and in a more general framework. We contributed the scale transformation to the relationship between model and observation spaces, so we developed the mathematical formalisms of scale and the scale transformation. The definition of scale is

central to this framework. We treated scale variations similarly to time variations, and stochastic calculus based data assimilation was conducted with respect to scale. Fingin a stochastic framework because:

—Our work presents a general framework that benefits the study of data assimilation in a nonlinear and general Gaussian sense. Both the forecasting and observation operators of data assimilation are strongly nonlinear, and the state and observation are generally associated with different geophysical parameters. Therefore, the relationship between them is not linear. We used the nonlinear transformation of scale and stochastic calculus to illustrate this relationship.

—Another advantage is that we considered the heterogeneity of geophysical parameter and a general Gaussian representativeness error, which were included in the reformulation of state and observation. In Sect. 3, both the state and observation with respect to the scale were understood in the Ito sense. Thus, stochastic process offers an infinite probability space of continuous scale paths, and indicates a promising approach to track a specific general Gaussian. In Eq. (13) and Eq. (14), we let $\varphi_x(s) = 0$, $\varphi_y(s) = 0$ and $\sigma_x(s) = \sigma_y(s) = 1$ for simplicity, which caused the state and observation to be Gaussian. However, if all the integrands in Eq. (13) and Eq. (14) are nonlinear functions instead of constants, which makes these two equations integral over the field of Brownian motion, then the state and observation are the general Gaussian processes with respect to scale. These terms finally results in a general Gaussian representativeness error. Note that all the results in our framework were given in terms of probability, not specific values.

—We further continued and improved the representativeness error expression in data assimilation. The nonlinear error that was caused by scale transformation was given in Eq. (23). If we assume that the observation operator and the relationship between the state and observation are linear and expand $H(s_x, X(s_x))$ in Eq. (20) in observation space, i.e., let $s_y = s_x$, then Eq. (23) becomes $P(Y|X) = N\{Y(s_x) - H(s_x, X(s_x)), |s_y - s_x|\}$. Here, we further denote the covariance of representativeness error as the scale difference between the observation and model space $|s_y - s_x|$.

Similarly, Eq. (29) can also be reduced to $M(s_x, X(0, s_x)) - M(s_x, X(0, s_x)) = \text{integral paths}$. Therefore, the stochastic calculus equation provided an infinite space with respect to the variable process $V(t)$, and a case study represented a sampling in this space, whose performance depended on its integral path.

— This study conducted a theoretical exploration. However, applying the above theoretic work to real world data assimilation is challenging. Studies on scale related errors still require further improvements.

4 Discussion & Conclusions

4.1 Discussion

Compared to previous work, our study offered a stochastic data assimilation framework to formulate the errors that are caused by scale transformation. The necessity of the methodology, the difference to previous works by other investigators, and the advantages and limitations of this study are summarized-discussed as follows.

The reasons why that the methodology focuses on a stochastic framework are because: First, the stochastic data assimilation framework is essentially consistent with the conceptions of scale and scale transformation. Both of them are associated with corresponding measure spaces $(\Omega, \mathcal{F}, \mu)$. Therefore, it is natural to regard the state space and observation space as two different measure spaces, respectively, and each element of state (or observation) vector can be seen as a geophysical variable that mapping the state (or observation) measure space onto R . Correspondingly, as the integrals of random processes with respect to random processes, stochastic calculus was adopted ultimately. Second, using stochastic calculus with respect to scale can also formulate the errors caused by scale transformations. The study proceeds with and improves the understanding of representativeness error in terms of scale. Results did not only prove the conventional point that the uncertainties of these errors mainly depend on the differences between scales, but indicated that the first-order differential of the nonlinear observation operator should also be incorporated in representativeness error, which is rarely found in the data assimilation literature. Last, stochastic calculus can be extended to meet a general scale transformation and formulate corresponding representativeness error, which is This was unattainable for in previous work. For example, if the scale changes randomly, say, from an irregular footprint to another irregular footprint, the stochastic equation can offer a multiple-integral to further formulate present this kind of thea scale transformation, such as $V(x, y) = V_0 + \int_{Y_0}^Y \int_{X_0}^X \varphi(x, y) dx dy + \int_{Y_0}^Y \int_{X_0}^X \sigma(x, y) dW_1(x) dW_2(y)$, where $W_1(x)$ and $W_2(y)$ are two independent Brownian Motion.

Our study is: The significant both in forms and functions of this work is: We developed a more rigorous formulation of scale and the scale transformation based on Lebesgue measure, which places the related conceptions in a rigorous mathematical framework and then conduces to new understanding of the errors caused by scale transformation with stochastic calculus. In

5 addition, due to the Ito process-formed state and observation, a stochastic data assimilation framework was proposed by considering the ~~nonlinear operators, heterogeneity of a geophysical variable and a general Gaussian representativeness error.~~ ~~There is no linear assumption of forecasting and observation operators.~~ Scale transformation is also nonlinear as well if the one-dimensional rule is not involved. Additionally, Ito processes-formed state and observation offer the drift rate (i.e., $\varphi(s)$) in Eq. (10)) to formulate the heterogeneity associated with scale transformation. It also permits the representativeness error to be general Gaussian in this framework. If all the integrands in Eq. (13) and Eq. (14) are nonlinear functions instead of constants (in this study we let $\varphi_X(s) = 0$, $\varphi_Y(s) = 0$ and $\sigma_X(s) = \sigma_Y(s) = 1$ for simplicity), then these two equations are integrated over the field of Brownian motion, and state and observation are the general Gaussian processes of scale. Based on these functions, representativeness error is a general Gaussian process.

10 As a theoretical exploration towards scale transformation and stochastic data assimilation, there is still big room for improvement. First, we reduced the scale transformation by one-dimensional rule, and let the variables in data assimilation evolve regularly according to assumptions 1~3. So, ~~only the ideal result was developed~~ ~~investigated~~ investigated. Therefore, an in-depth and comprehensive exploration should be conducted in ~~the~~ future to describe other situations in the real world. However, ~~other situations may be more complex.~~ Either an arbitrary scale transformation or the geophysical variable without ignoring the drift rates will deduce lengthy results. ~~So~~ ~~Therefore~~, the second improvement focuses on how to make the formulation more concise. Last, noting that all the results in our framework were given in terms of probability, it is necessary to implement the real-world applications of these theoretical results, such as introducing some concrete dynamic models to formulate the Ito process-formed geophysical variable of scale.

4.2 Conclusions

20 In this study, we mainly addressed two basic problems associated with scale transformation in earth observation and simulation. First, we produced a mathematical formalism of scale and scale transformation by employing measure theory. Second, we demonstrated how scale transformation and associated errors could be presented in a stochastic data assimilation framework.

25 We revealed that the scale is the Lebesgue measure with respect to the observation footprint or model unit. Scale is related to the shape and size of a footprint, and scale transformation depends on the spatial change between different footprints. We then defined the geophysical variable, which further considers the heterogeneities and physical processes. A geophysical variable consequently expresses the spatial average at a specific scale.

30 We formulated the expression of scale transformation and investigated the error structure that is caused by scale transformation in data assimilation using basic theorems of stochastic calculus. Formulations explicate that the first-order differential of the nonlinear observation operator should be considered in representativeness error, and the uncertainty of representativeness error is directly associated with the difference between scales. A concrete physical models (SRTE) was introduced to ~~explain~~ ~~demonstrate~~ the results when observation operator is nonlinear. ~~How to extend~~ ~~Extension~~ the results to n-dimensional stochastic data assimilation was also presented.

This work conducted a theoretical exploration of formulating the errors caused by scale transformation in stochastic data assimilation framework. We hope that the stochastic methodology can ~~in-essenetialey-~~ benefit the study on these errors.

5 Notation

5.1 Basic notations

5	Ω	<u>Non empty space</u> Observational region
	\mathcal{F}	σ -algebra
	μ	Measure
	dV	Variable process
	$W(s)$	Brownian motion
10	$(\Omega, \mathcal{F}, \mu)$	Measure space
	I^n	N-dimensional Lebesgue volume
	m^n	Lebesgue measure or an outer measure on R^n
	\mathcal{L}^n	Lebesgue σ -algebra of R^n
	$\int f dm^n$	Lebesgue integral
15	$ J(\cdot) $	Jacobian determinant
	$(\Omega^n, \mathcal{F}^n)$	Product space

5.2 New notations

Notation	Name	Explanation	Index
s	Scale	The observation footprint or model unit to measure <u>observe</u> or <u>evaluate-model</u> a geophysical parameter <u>variable</u>	Sect. 1 & Sect. 3.1
A_0	Unit square interval in R^2		Sect. 3.1
s_0	Standard scale	A Lebesgue integral of A_0 is the unit area	Sect. 3.1
	One-dimensional rule	Two scales are geometrically similar	Eq. (58)
V	<u>Geophysical variable</u>	Estimation of a geophysical-variable parameter at a specific scale	Sect. 3.2
dX	State process	State in the sense of the Ito process- <u>formed state</u>	Eq. (13)
dY	Observation process	Observation in the sense of Ito process- <u>formed observation</u>	Eq. (14)

X_0	State in s_0	Eq. (15)
Y_0	Observation in s_0	Eq. (16)
s_x	Scale of state space	Eq. (15)
s_y	Scale of observation space	Eq. (16)

Acknowledgements

We thank the editor-in-chief of NGP, Prof. Talagrand, and his kind ~~considerations~~ [help and valuable comments](#) on our manuscript. We also thank Dr. van Leeuwen and another anonymous reviewer for their valuable comments and suggestions. This work was supported by the NSFC projects (grant numbers 91425303 & 91625103) and the CAS Interdisciplinary
5 Innovation Team of the Chinese Academy of Sciences.

References

- Apte, A., Hairer, M., Stuart, A. M. and Voss, J.: Sampling the posterior: An approach to non-Gaussian data assimilation, *Physica D*, 230, 50-64, doi: 10.1016/j.physd.2006.06.009, 2007.
- Atkinson, P. M. and Tate, N. J.: Spatial scale problems and geostatistical solutions: a review, *Prof. Geogr.*, 52, 607-623, doi:
10 10.1111/0033-0124.00250, 2004.
- Bartle, R. G.: *The Elements of Integration and Lebesgue Measure*, Wiley, New York, 1995.
- Billingsley, P.: *Probability and Measure*, 2nd ed., John Wiley & Sons, New York, 1986.
- Bocquet, M., Pires, C. A., Wu, L.: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation, *Mon. Weather
Rev.*, 138, 2997-3023, doi: 10.1175/2010MWR3164.1, 2010.
- 15 Bocquet, M., Wu, L., Chevallier, F.: Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism, *Q. J. Roy. Meteor. Soc.*, 137, 1340-1356, doi: 10.1002/qj.837, 2011.
- Crow, W. T., Berg, A. A., and Cosh, M. H., et al.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, *Rev. Geophys.*, 50, 3881-3888, doi: 10.1029/2011RG000372, 2012.
- Eyink, G. L., Restrepo, J. M., and Alexander, F. J.: A mean field approximation in data assimilation for nonlinear
20 dynamics, *Physica D*, 195, 347-368, doi: 10.1016/j.physd.2004.04.003, 2004.
- Famiglietti, J.S., Ryu, D., Berg, A.A., Rodell, M., and Jackson, T.J.: Field observations of soil moisture variability across scales, *Water Resour. Res.* 44, doi: 10.1029/2006WR005804, 2008.
- Giménez, D., Rawls, W. J., and Lauren, J. G.: Scaling properties of saturated hydraulic conductivity in soil, *Geoderma*, 27, 115-130, 1999.
- 25 Goodchild, M. F. and Proctor, J.: Scale in a digital geographic world, *Geographical & Environmental Modelling*, 1, 5-23, 1997.

- Gruber, A., Dorigo, W. A., Zwieback, S., Xaver, A., and Wagner, W.: Characterizing coarse-scale representativeness of in situ soil moisture measurements from the international soil moisture network, *Vadose Zone J.*, 12, 522-525, doi: 10.1002/jgrd.50673, 2013.
- Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements, *J. Geophys. Res-Atmos.*, 118, 8585–8597, 2013.
- 5 Hodyss, D. and Nichols, N. K.: The error of representation: Basic understanding, *Tellus A*, 66, 1-17, doi: 10.3402/tellusa.v67.24822, 2015.
- Huang, G., Li, X., Huang, C., Liu, S., Ma, Y., and Chen, H.: Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products, *Remote Sens. Environ.*, 181, 198-206, doi: 10.1016/j.rse.2016.04.001, 2016.
- 10 Itô, K.: Stochastic integral, *P. JPN. Acad.*, 22, 519-524, 1944.
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., and Ustin, S. L.: Prospect and sail models: a review of use for vegetation characterization, *Remote Sens. Environ.*, 113, S56-S66, 2009.
- Janjić, T. and Cohn, S. E.: Treatment of Observation Error due to Unresolved Scales in Atmospheric Data Assimilation, *Mon. Weather Rev.*, 134, 2900-2915, doi: 10.1175/MWR3229.1, 2006.
- 15 Jazwinski, A. H.: Stochastic processes and filtering theory, Academic Press, New York, 1970.
- Karatzas, I. and Shreve, S.E.: Brownian Motion and Stochastic Calculus, 2nd ed., Springer-Verlag, New York, 1991.
- [Kassianov, E. and Veron, D.: Stochastic radiative transfer in Markovian mixtures: past, present, and future. J. Quant. Spectrosc. Radiat. Transf. 112, 566-576, doi: 10.1016/j.jqsrt.2010.06.011, 2011.](#)
- 20 Li, X.: Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems, *Sci. China Ser. D*, 57, 80-87, doi: 10.1007/s11430-013-4728-9, 2014.
- Li, X. and Liu, F.: Can Point Measurements of Soil Moisture Be Used to Validate a Footprint-Scale Soil Moisture Product?, *IEEE Geosci. Remote S.*, 2016. (Submitted)
- Lin, H., Flühler, H., Otten, W., and Vogel, H. J.: Soil architecture and preferential flow across scales, *J. HYDROL.*, 393, 1-2, doi: 10.1016/j.jhydrol.2010.07.026, 2010.
- 25 Lorenc, A.C.: Atmospheric Data Assimilation, Meteorological Office, Bracknell, 1995.
- [Lorenc, A. C.: Analysis methods for numerical weather prediction. Q. J. Roy. Meteor. Soc., 112, 1177–1194, 1986.](#)
- Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modelling, *Water Resour. Res.*, 45, 627-643, doi: 10.1029/2009WR007872, 2009.
- 30 Miller, R. N., Carter, E. F., and Blue, S. T.: Data assimilation into nonlinear stochastic models, *Tellus A*, 51, 167-194, doi: 10.1034/j.1600-0870.1999.t01-2-00002.x, 1999.
- Miller, R. N.: Topics in data assimilation: stochastic processes, *Physica D*, 230, 17-26, doi: 10.1016/j.physd.2006.07.015, 2007.
- Miralles, D. G., Crow, W. T., and Cosh, M. H.: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, *J. Hydrometeorol.*, 11, 1423-1429, 2010.

- Narsilio, G. A., Buzzi, O., Fityus, S., et al.: Upscaling of Navier–Stokes equations in porous media: Theoretical, numerical and experimental approach. *Computers & Geotechnics*, 36, 1200–1206, 2009.
- Pomraning, G.C.: Radiative transfer and transport phenomena in stochastic media, *Int. J. Eng. Sci.*, 36, 1595–1621, doi: [10.1016/S0020-7225\(98\)00050-0](https://doi.org/10.1016/S0020-7225(98)00050-0), 1998.
- 5 Ran, Y. H., Li, X., Sun, R., Kljun, N., Zhang, L., Wang, X. F., and Zhu, G. F.: Spatial representativeness and uncertainty of eddy covariance carbon flux measurement for upscaling net ecosystem productivity to field scale, *Agric. Forest Meteorol.*, doi: [10.1016/j.agrformet.2016.05.008](https://doi.org/10.1016/j.agrformet.2016.05.008), 2016.
- Ryu, D. and Famiglietti, J. S.: Multi-scale spatial correlation and scaling behavior of surface soil moisture, *Geophys. Res. Lett.*, 33, 153-172, 2006.
- 10 [Shabanov, N. V., Knyazikhin, Y., Baret, F., and Myneni, R. B.: Stochastic modeling of radiation regime in discontinuous vegetation canopies, *Remote Sens. Environ.*, 74, 125-144, doi: 10.1016/S0034-4257\(00\)00128-0, 2000.](https://doi.org/10.1016/S0034-4257(00)00128-0)
- Shreve, S.E.: *Stochastic Calculus for Finance II*, Springer-Verlag, New York, 2005.
- Talagrand, O.: Assimilation of observations, an introduction, *J. Meteorol. Soc. JPN*, 75, 191-209, 1997.
- van Leeuwen, P. J.: Representation errors and retrievals in linear and nonlinear data assimilation, *Q. J. Roy. Meteor. Soc.*, 141, 1612-1623, doi: [10.1002/qj.2464](https://doi.org/10.1002/qj.2464), 2014.
- 15 van Leeuwen, P. J.: Nonlinear data assimilation for high-dimensional systems, in *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, vol. 2, Springer-Verlag, New York, 2015.
- Vereecken, H., Kasteel, R., Vanderborght, J., and Harter, T.: Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: a review, *Vadose Zone J.*, 6, 1-28, doi: [10.2136/vzj2006.0055](https://doi.org/10.2136/vzj2006.0055), 2007.
- 20 Wiens, J. A.: Spatial scaling in ecology, *Funct. Ecol.*, 3, 385-397. doi: [10.2307/2389612](https://doi.org/10.2307/2389612), 1989.