

## **Responses to the editor's comments**

We appreciate the kind considerations from Dr. Olivier Talagrand on our manuscript in NPG Discussion.

We made major revisions to our original manuscript based on all the comments from the editor and referees. The main amendments are as follows (please forgive us for not marking revisions in the revised manuscript; many changes were made, and marking all revisions would make the paper a mess and difficult to read):

### **1. Structural changes**

We thoroughly revised the structure to make our manuscript more understandable. The title of the revised manuscript was rewritten as “Formulation of Scale Transformation in a Stochastic Data Assimilation Framework”. We made this significant modification for the following reasons. First, defining the scale and scale transformation laid a foundation for our study and makes our work distinct from the previous studies. Second, the original title was insufficient because we did not reformulate the framework of a stochastic data assimilation, which was used only to investigate the expression of errors that are determined by scale transformation. Therefore, the new title is more suitable.

First, we rewrote Sect. 1. We reduced the original text and added some necessary reviews on representativeness error. Second, Sect. 2 was also changed to “Basic knowledge”; we stated the basic knowledge of measure theory in Sect. 2.1, and the basic knowledge of stochastic calculus in Sect. 2.2. Our study was mainly presented in Sect. 3. In Sect. 3.1, we defined the scale and scale transformation with Lebesgue measure. In Sect. 3.2, we introduced the definition of stochastic variables in data assimilation. The errors from scale transformation in a data assimilation framework were presented in Sect. 3.3. Then, Sect. 4 was divided into three subsections, namely, “Summary”, “Discussion” and “Next Step”. In the last section, we listed the basic notations in measure theory and stochastic calculus and the new notations in our study.

In addition, we removed Sect. 4.4, “An example: the stochastic radiative transfer equation (SRTE)”, which was not closely tied to the other sections of this study. We shifted Sect. 4.5 to Sect. 3.4, “Extension to n-dimensional data assimilation”.

## 2. Full introduction of Lebesgue measure to define “scale”

In the original version, we developed a vector-valued measure to define the “scale” (see Sect. 2 in the original manuscript). Referee 1 argued that this measure mismatches with his/her basic knowledge of measure theory and leads to a poor understanding of the manuscript. Dr. Talagrand also stressed this problem. We accept this advice. After careful consideration, we decided to abandon the original idea and completely change the definition of the “scale” by further introducing Lebesgue measure.

**Changes in the manuscript:** Correspondingly, Sect. 2 was completely rewritten. In Sect. 2.1, which was retitled “Measure theory”, we introduced some basic concepts such as  $\sigma$ -algebra, measure, measure space, Lebesgue measure, Lebesgue integral, and so on. In Sect. 3.1, “Definition of scale”, we used Lebesgue measures to investigate the structures of scales and scale transformation.

## 3. Previous studies on representativeness error were included and compared to our study

Prof. van Leeuwen noted that no reviews of representativeness error exist and that this topic requires a comparison between our study and previous work on representativeness error. We added the corresponding content in the revised manuscript.

**Changes in the manuscript:**

- (1) The introduction mainly considered previous work on representation error.
- (2) Sect. 4.2 was added to compare our study with the available works on representation error.

Please find this information in the revised manuscript (main contents are in Sect. 1 and Sect. 4.2).

## 4. Equation recompilation

We reedited all the equations with Microsoft equation editor. The typesetting of the mathematical equations in the revised manuscript is better than that in the original manuscript.

## 5. Notation consistency

Both the referees and Dr. Talagrand emphasized that the notations in the original manuscript were inconsistent. We seriously addressed this issue. The corresponding changes in the new version are as follows:

Previous notation	Current notation
Scale change, change of scale	Scale transformation

Observation/measurement region	Observation footprint
Model units	Model unit
System state	State
dynamic model (operator), physical model (operator)	Forecasting model (operator)
System state space	Model space
Instrument error	Measurement error

## 6. New Notations

Dr. Talagrand advised us to clearly explain the new notations that were introduced in our manuscript. We accept this advice. In the revised manuscript, we expanded the “Notation” section and classified all the notations into two types. One was the basic notations in measure theory and stochastic calculus, which were given only a full name, and their definitions can be found in Sect. 2.1 and Sect. 2.2. The other type was the new notations. We offered their full names and their detailed explanations and indices in the manuscript. Please find more information in the revised manuscript (Sect. 5).

## 7. Other changes

Many other small presentation changes were made in the current manuscript. Some of these changes involved our re-examination, and some were provided by a professional native English speaking team. We hope that these changes improved the manuscript substantially to meet the quality standards of NPG. However, the changes were not presented as revisions in our revised manuscript because so many revisions that presenting them all would make the paper a mess.

## Responses to Referee 1

We thank the anonymous Referee 1 for taking his/her valuable time to review our manuscript and provide us some very thoughtful and constructive comments. We apologize for our first response letter, which now seems inadequate. We had to update our reply after thoroughly changing this manuscript, especially the introduction of measure theory and definitions of scale and scale transformation. Here are the new point-by-point responses to the comments from Referee 1 (please forgive us for not marking revisions in the revised manuscript; many changes were made, and marking all revisions would make the paper a mess and difficult to read).

### 1. General reply

In the original version, we only introduced the concept of measure and developed a vector-valued measure to define “scale” (see Sect. 2 in the original manuscript). However, this new measure did not exhibit enough rigorousness. After careful consideration, we decided to abandon the original idea and completely change the definition of “scale” by further introducing Lebesgue measure for the following reasons:

- (1) Lebesgue measure is generally accepted. Its definition is sound and its geometric meaning is similar with scale. Therefore, Lebesgue measure can be potentially applied in mathematical formalism of scale (in the original manuscript, we attempted to develop a measure that belonged to Earth observations and simulations, but the result was not better than that obtained using Lebesgue measure).
- (2) Demonstrating scale transformation is important in addition to the scale. This notion coincides with the concept of change of variable in the Lebesgue integral (in the revised manuscript, this concept was called Lebesgue integration by substitution). In the original manuscript (see Sect. 2.2), we implied the concept of scale transformation with abundant discussion. However, we introduced this concept in the revised manuscript (see Sect. 3.1) following Lebesgue measure for simplicity (see Sect. 3.1 in the revised manuscript).
- (3) Based on the original idea, the explanation and instances of this new measure contained abundant content, which made our presentation redundant and hard to understand. Therefore, we thoroughly modified this content by introducing Lebesgue measure and other associated concepts. Being

similar to our study in terms of scale and scale transformation, this content could provide a more concise and rigorous presentation. Moreover, the length of this article was reduced because no extra explanation was needed.

**Changes in the manuscript:** Correspondingly, the paper was completely rewritten. The title of the revised manuscript was rewritten as “Formulation of Scale Transformation in a Stochastic Data Assimilation Framework”. We made this significant modification for the following reasons. First, defining the scale and scale transformation laid a foundation for our study and makes our work distinct from the previous studies. Second, the original title was insufficient because we did not reformulate the framework of a stochastic data assimilation, which was used only to investigate the expression of errors that were determined by scale transformation. Therefore, the new title is more suitable.

Sect. 1 was reorganized and Sect. 2 was retitled as “Basic knowledge”, which mainly introduced the basic concepts and theorems of measure theory and stochastic calculus. Sect. 3 was retitled as “Reformulation of scale transformation in a data assimilation framework”, where we first defined some essential concepts, such as the scale, scale transformation and variables. Then, we established a Bayesian description of data assimilation with time- and scale-dependent stochastic processes and formulated the effect of scale transformations on the posterior probability of the state.

In Sect. 2.1, which was retitled “Basic knowledge of measure theory”, we introduced some basic concepts such as  $\sigma$ -algebra, measure, measure space, Lebesgue measure, Lebesgue integral, and so on. Two main references were used: “Billingsley, P.: Probability and Measure, 2nd ed., John Wiley & Sons, New York, 1986.”, and “Bartle, R. G.: The Elements of Integration and Lebesgue Measure, Wiley, New York, 1995.” The latter might be the book that Referee 1 recommended in the interactive comment. Indeed, some terminological incongruences exist between these two books, so we tried our best to make the exposition acceptable and explicit.

In Sect. 3.1, which was retitled “Definition of scale”, we mainly developed the structures of “scale” and “scale transformation” by Lebesgue measure. Scale is logically similar to Lebesgue measure and some technicalities were also included in the previous section, so this section is more concise than that in the original manuscript. In addition, the revised definition of scale is also valid for the following sections of our study.

Please find the detailed information in the revised manuscript.

## 2. Language problem

*“Unfortunately, the exposition is extremely hard to follow, and, after reading the article through carefully, I still don’t understand what the authors are doing. Part of the problem is probably the language. The manuscript would profit considerably if the authors could find a sympathetic native English speaker to read it over.”*

**Response:** Because we are non-native English speakers, our manuscript was thoroughly edited by a manuscript service company, American Journal Experts, before submission (Certificate Verification Key: 53B9-681E-0C9B-BC77-FF4C). However, the exposition could be improved further. The revised manuscript was totally rewritten and was re-edited by a professional native English speaking team. We ensure that we will provide a modified version of this manuscript with higher-quality English expression in the future.

## 3. Reference problems

*“More importantly, I find much of the exposition puzzling. I don’t know the book by Billingsley. In my day students in the USA learned measure theory from the texts by Bartle and Royden, and, relative to my background, much of the material is written very unconventionally.”*

*“The Bayesian expression of DA in terms of the stochastic calculus appears in many places. The authors should consult the volume by Jazwinski and the recent work of P. J. van Leeuwen and M. Bocquet.”*

**Response:** The literature by Billingsley, Bartle and Royden are all classic works of measure theory (the texts that you mentioned may be *The Elements of Integration and Lebesgue Measure* by Bartle R. G. and *Real analysis* by Royden H. L.) and were all frequently cited according to the search results of Google Scholar. In addition, we received help from the volume “Stochastic Processes and Filtering Theory” by Jazwinski A. H. (1970) during our study, which will be listed in ‘References’ in the revised manuscript. We also appreciate the recommended literature regarding Bayesian data assimilation in terms of stochastic calculus. This literature introduced the latest advances in this research field, and the related papers will be cited in our present and future works.

#### 4. Why measures were defined as a vector-valued set function in the manuscript

*“The usual intuition for the concept of measure is that measure is a generalization of the concepts of length, area and volume, and is thus a scalar valued set function. The authors’ response at the end of section 2.1 to Prof. Talagrand’s comment is inadequate. There is nothing intrinsically wrong with choosing a vector valued measure, but that choice requires more explanation than simply “the measure correspondingly turns to ...” The authors should explain why they want to define measure as a vector valued set function, rather than simply defining the measure of a rectangle in Euclidean space as its area. Again, maybe Billingsley defines it differently, but the Lebesgue measure of a rectangle is its area, not a vector whose components are the lengths of its sides as the authors assert on line 16 of page 6.”*

**Response:** Dr. Talagrand also stressed this problem. In the revised manuscript, we completely changed the definition of scale by introducing Lebesgue measure. Detailed information can be found in “General reply” in this response and in the revised manuscript (main contents are in Sect. 2.1 and Sect. 3.1).

#### 5. Definition of scale

*“I do not understand the definition of scale. First, measure is a function whose domain is the sigma field  $F$ , as noted at the very beginning of section 2.1. The integrals in the definition of scale, line 16, page 7, don’t make sense to me.  $A_0$  is a specific set. (Pardon my TeX, I don’t know how to make superscripts, subscripts or special characters). In the analysis texts I learned from,  $\mu(A_0)$  is the area of the set  $A_0$ . I don’t understand the expression  $\mu(A_0)dA_0$ . I cannot make sense of the second integral. The domain of the measure  $\mu$  is the  $\sigma$ -field  $F$ , as noted in the beginning of section 2.1. The Lebesgue measure is not a point function”*

**Response:** In the revised manuscript, we completely changed the definition of scale by introducing Lebesgue measure. The scale and scale transformation were formulated by Lebesgue integral and change of variable in the Lebesgue integral. Detailed information can be found in “General reply” in this response letter and in the revised manuscript (main contents are in Sect. 2.1 and Sect. 3.1).

## 6. Regarding Figure 1

*“It would help a great deal if there were more explanation of figure 1. In particular, after reading and rereading the last paragraph on page 8, I can’t understand how  $C_2$  can have the same measure as  $C_1$  and  $C_3$ , and  $D_1$  has the same measure as  $D_2$ , though they have the same “scale.” The problem may be the terminology: As I recall my long-ago analysis classes, the common intuition for measure was that measure corresponds to area, and, in particular, the Lebesgue measure of a geometrical figure in the plane is its area.”*

**Response:** We feel very sorry because the description of this figure was not enough to help you understand. As stated in the above response, we defined scale by Lebesgue measure. Thus, the explanation of Figure 1 was changed accordingly. The conclusion is more reasonable and concise because it was established by the definition of Lebesgue measure.

**Changes in the manuscript:** The paragraph before Figure 1 was updated as follows:

“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 interval  $A_0$ . Let  $m_{C_1}^2$ ,  $m_{C_2}^2$  and  $m_{C_3}^2$  be the Lebesgue measures of disc measurements  $C_1$ ,  $C_2$  and  $C_3$ , respectively, and let  $m_{D_1}^2$  and  $m_{D_2}^2$  be the Lebesgue measures of diamond measurements  $D_1$  and  $D_2$ . Then,  $m_{C_1}^2 = m_{C_2}^2 = m_{C_3}^2$  because they are the same function. 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,  $A_i + 2 = [x: x_k + 2, x_k \in A_i, k = 1, 2]$  and  $A_i \times 3 + 2 = [x: x_k \times 3 + 2, x_k \in A_i, k = 1, 2]$ , which proves that  $m_{C_1}^2$ ,  $m_{C_2}^2$  and  $m_{C_3}^2$  collect the desirable set based on the same scheme, so they are identical. Additionally,  $\sum I^2(A_i \times 3 + 2)$  is much larger than  $\sum I^2(A_i)$ . 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.”

## 7. Inconsistency problems

*“Finally, the manuscript seems inconsistent with itself. As examples, consider the abstract. “...measure theory was used to propose [a definition of] spatial scale ...[and the] Jacobian matrix [was used] to describe the change of scale. The Jacobian matrix is introduced on page 7 in the well known change of variables formula, change of scale by the Jacobian matrix is defined on page 8, and the Jacobian is not mentioned again until the summary. No further discussion of the effects of change of scale appears. Again, in the abstract, “...the variation range of this type of error is proportional to the scale gap, ...” I’m sure I’m not the only reader for whom the phrase “scale gap” conjures up ideas of inertial range from turbulence theory and similar notions. The term “scale gap” is never mentioned in the body of the article.”*

**Response:** You mentioned two inconsistency problems. The first is why the Jacobian matrix and the change in scale (the former was included in the introduction of the change of variable in the Lebesgue integral, and the latter was renamed as “scale transformation”) disappeared after Sect. 2. Actually, these concepts were not omitted but simplified by the one-dimensional rule (defined in Sect. 3.1 and Eq. (5)) to suit stochastic calculus. Then, we used this simplified version of scale changes to investigate the uncertainties in data assimilation (we also mentioned this problem in the second paragraph of Sect. 5 in the original manuscript). Although the one-dimensional transformation maybe the simplest case, the results were still complicated and some new components of uncertainty were discovered (Eq. 21~27). However, fully studying the scale transformation using Jacobian matrices is comprehensive and universal, and it will be launched in our following work rather than in this study. The second inconsistency is the phrase “scale gap” in the abstract. We apologize for the trouble with this word. Here, the term “scale gap” refers to the quadratic variation between  $s_X$  and  $s_Y$  (see Eq. 25~27). However, fully explaining this term in the abstract was difficult, so we had to use the term “scale gap”. Replacing this term with “the difference between scales” may be more practical.

## Responses to Referee 2

We thank Prof. van Leeuwen for his valuable comments and suggestions. Most importantly, he provided some valuable literature on representation error, which convinced us to reconsider our study in light of these previous works and helped us improve our study in line with the latest advances in studying representation error. Our point-by-point responses to the comments from Prof. van Leeuwen are as follows (please forgive us for not marking revisions in the revised manuscript; many changes were made, and a revised version would make the paper a mess and difficult to read):

### 1. General reply

In the original version, we considered some results from other researchers, including basic knowledge of scale, scale problem in geophysical processes, and data assimilation with respect to stochastic processes. However, we did not consider the representation error, which was also noted by Prof. van Leeuwen in his interactive comment.

In fact, we also devoted ourselves to studying the representation error. Some results of how to understand and quantify representation error with real world experiments have been published or are being considered for publication:

- Li, X.: Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems, *Sci. China Ser. D*, 2014.
- Huang, G., Li, X., et al.: Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products, *Remote Sens. Environ.*, 2016.
- 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)
- Ran, Y. H., Li, X., et al.: Spatial representativeness and uncertainty of eddy covariance carbon flux measurement for upscaling net ecosystem productivity to field scale, *Agric. Forest Meteorol.*, 2016.

We used “representativeness errors” in our previous studies, and we believe that the representation error is not limited to data assimilation. This factor may remarkably affect many fields that are associated with Earth observations and simulations.

We did not explicitly introduce representation error in the original manuscript because this study mainly focuses on the errors from scale transformation, which is only one component of representation error. However, we accept the advice from Prof. van Leeuwen to consider representation error in our study. According to this comment, some valuable works examined representation error in data assimilation, and considering these works should substantially improve our study.

**Changes in the manuscript:** In the revised manuscript, the corresponding text was added and the structure of the manuscript was reworked. The main changes are as follows:

- (1) The title of the revised manuscript was rewritten as “Formulation of Scale Transformation in a Stochastic Data Assimilation Framework”. We made this significant modification for the following reasons. First, defining the scale and scale transformation laid a foundation for our study and makes our work distinct from previous studies. Second, the original title was insufficient because we did not reformulate the framework of a stochastic data assimilation, which was used only to investigate the expression of errors that were determined by scale transformation. Therefore, the new title is more suitable.
- (2) We rewrote the introduction. We reduced the original text and added some necessary reviews on representation error. The fundamental causes of representation error, the difference between representation error and the scale-dependent error, the latest advances in studying the representation error in data assimilation, and the room to improve these studies were summarized.
- (3) Sect. 2 was also changed. This section’s name was changed to “Basic knowledge”. We stated the basic knowledge of measure theory in Sect. 2.1, and we introduced the basic knowledge of stochastic calculus in Sect. 2.2. Our study was mainly presented in Sect. 3. In Sect. 3.1, we defined the scale and scale transformation with Lebesgue measure. In Sect. 3.2, we introduced the definition of stochastic variables in data assimilation. The errors from scale transformation in a data assimilation framework were presented in Sect. 3.3. Then, Sect. 4 was divided into three subsections, namely, “Summary”, “Discussion” and “Next step”. In the last section, we listed the basic notations in measure theory and stochastic calculus, and the new notations in our study. In addition, we removed Sect. 4.4, “An example: the stochastic radiative transfer equation (SRTE)”, which was not closely tied to the other sections of this study. We shifted Sect. 4.5 to Sect. 3.4, “Extension to n-dimensional data assimilation”.

- (4) At the end of Sect. 3.3, we considered a related problem regarding what the error would be if the initial state is not at the scale of the forecasting operator. The results are also based on our framework with respect to scale.
- (5) An extra subsection was added to Sect. 4, which was titled “Discussion”. In this section, we made further comparisons between our framework and previous treatments of representation error. After a careful literature survey, we believe that the improvement from our study is a general framework of stochastic data assimilation, which was presented to investigate scale-dependent errors. First, our framework is based on nonlinear forecasting operators and observation operators. The scale transformation, which is similar to the relationship between the model space and observation space, is also nonlinear. For simplicity, we let the scale transformation obey the one-dimensional rule (See Sect. 3.1), which is equal to a linear change in scale. However, our framework does not exclude nonlinear scale transformation, and one can track a more complicated integral path of the scale for this purpose. Second, our framework permits general Gaussian representation error by introducing a non-constant volatility function in Ito-formed states and observations. Finally, we further considered the heterogeneity of geophysical parameters. This improvement highly depended on the first two advantages, i.e., either the nonlinear scale transformation or general Gaussian representation error.

Please find the detailed information in the revised manuscript.

## 2. Language problem

*“The language is also not standard.”*

**Response:** Indeed, we are non-native English speakers. Therefore, our manuscript was completely re-edited by a professional native English speaking team. We tried our best to provide a high-quality English expression manuscript that could be understood by average readers in data assimilation.

We updated the data assimilation terminologies by referring to some classic literature, such as

- Talagrand, O.: Assimilation of observations, an introduction, *J. Meteorol. Soc. JPN*, 75, 191-209, 1997.
- Ide, K., Ghil, M., Lorenc, A. C.: Unified Notation for Data Assimilation: Operational, Sequential and Variational, *J. Meteorol. Soc. JPN*, 75, 181-189, 1997.

The references that were provided by Prof. van Leeuwen were used to check the language. Correspondingly, some notations were changed as follows:

Previous notation	Current notation
Observation/Measurement region	Observation footprint
Model units	Model unit
System state	State
dynamic model (operator), physical model (operator)	Forecasting model (operator)
System state space	Model space
Instrument error	Measurement error

Additionally, some notations from measure theory and stochastic calculus, as well as the new notations that were defined in our study may seem strange to ordinary readers in data assimilation. These notations were all listed and provided corresponding explanations in Sect. 5, “Notation”, in the revised manuscript.

### 3. Innovation of this study

*“There are several recent works that describe the filtering problem for stochastic processes, and I’m not sure the results of the authors are new in this respect.”*

**Response:** Indeed, abundant literature exists on data assimilation based on stochastic processes. However, our study is very different from the available works and deduced some remarkable results. The previous works were mostly based on stochastic processes with respect to time. Our study, which introduced measure theory to investigate the natural structure of scale and treated scale variations similarly to time variations, constructed a framework that was based on stochastic calculus with respect to scale. Some components of representation error are associated with scale, so our study provided a more reasonable framework that can further improve our understanding of the structure of representation error in data assimilation.

#### 4. Necessity of introducing measure theory

*“I don’t see the need for the measure theory developed here.”*

**Response:** The scale in representation error is important, but a rigorous mathematical definition of scale is lacking. Simply regarding the resolution as the scale is not reasonable. For example, if two different observation footprints exist, one is a disc field with a diameter of 1 km and the other is a square field with side length of 1 km. Obviously, they have the same resolutions, but their scale is different. We introduced measure theory to further mathematically describe the geometric characteristics of an observation footprint or model unit, as well as the scale transformation. Measure theory laid the foundation of this study in terms of scale-dependent error in data assimilation.

#### 5. Advantages of our study

*“My suggestion is that the authors consider these references and reconsider their findings in light of those works. What has been missed by the examples of papers referred above? Why is the new framework needed, using language that the average NPG reader interested in data assimilation can understand?”*

**Response:** The advantages of our study include the general framework of stochastic data assimilation to investigate the scale-dependent errors, i.e., a framework that includes the nonlinear operators and nonlinear scale transformation (such as the relationship from model space to observation space), the general Gaussian representation error, and considering the heterogeneity of geophysical parameters. Although some of these factors were simplified to linear situations for more clear results, nonlinear results could also be deduced from this framework.

Please find the detailed information in the general reply of this response letter and the revised manuscript (main contents in Sect. 1 and Sect. 4.2).

# Formulation of Scale Transformation in a Stochastic Data Assimilation Framework

Feng Liu<sup>1,3</sup> and Xin Li<sup>1,2,3</sup>

<sup>1</sup>Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, 730000, China

<sup>2</sup>Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, P. R. China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, P. R. China

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

**Abstract:** The understanding of uncertainties in Earth observations and simulations has been hindered by the spatial scale issue. In addition, errors that are caused by spatial scale transformation are an important component of representativeness errors in data assimilation. Several relevant studies have been conducted, but these errors still exceed the abilities of current theory because of the associated nonlinearity. Thus, we attempt to address these problems. 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, Lebesgue integration by substitution is used to describe the scale transformation. Second, a scale-dependent variable is defined to further consider the heterogeneities. 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 one-dimensional, the variation in this type of error is proportional to the difference between scales. This new methodology can expand the understanding and treatment of the representativeness error in data assimilation and may be able 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 transformations 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. Empirical studies have been conducted only recently. 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 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 variations in land surface systems, which include hydrology (Giménez et al., 1999; Vereecken et al., 2007; Merz et al., 2009),  
5 soil science (Ryu and Famiglietti, 2006; Lin et al., 2010), radiative transfer (Jacquemoud 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 difficult to find, except for an intuitive conception (Goodchild and Proctor, 1997) and certain qualitative classifications of  
10 scale (Vereecken et al., 2007). This gap partially reflects the complexity of this problem and requires corresponding mathematical tools to elucidate the “scale”.

Data assimilation presents 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 are typically collected by various Earth observation  
15 techniques or from data products; therefore, scale mismatch may arise between them. Furthermore, this problem is even more common between the model units and observation footprints of measurements 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  
20 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  
25 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 in data assimilation is “representativeness error”, which is associated with the inconsistency in spatial and temporal resolutions between states, observations and operators (Janjić and Cohn, 2006; Hodyss and Nichols, 2015), and 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 observations and simulations and inevitably results in representativeness error.

Recently, approaches have been developed to assess representativeness error. Janjić and Cohn (2006) 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, the observations had a finer resolution than the model. In the other, the retrieved variables, which represented different dynamic processes, were assimilated. Their solutions were formulated using an agent variable in observation or model 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 the difference between the truth and the inaccurate value that is forecasted by the model.

Overall, these approaches explored the structure of representativeness error and offered various solutions. However, improvements can still be made. 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 scale, or when operators are nonlinear between states and observations. In previous studies, 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, heterogeneity varies situationally and is difficult to formulate in 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. 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 errors in data assimilation. The next section introduces the basic concepts and theorems of measure theory and stochastic calculus. In Sect. 3, we present some essential concepts, such as scale, scale transformation and variable, which form the basis for the subsequent study. In Sect. 4, 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 contribution of this study is presented in light of previous work. 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. We offer a solution in which the definition of scale must use the measure theory and the expression of geophysical parameter as a stochastic process must use stochastic calculus. Therefore, we first introduce several basic concepts of measure theory and stochastic calculus.

### 2.1 Measure theory

Let  $\Omega$  be an arbitrary nonempty space.  $\mathcal{F}$  is a  **$\sigma$ -algebra** (or  **$\sigma$ -field**) of subsets of  $\Omega$  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  $A_1 \cup A_2 \cup \dots \in \mathcal{F}$ .

A set function  $\mu$  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  $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ ,  $\mu$  is countably additive such that  $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$ .

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

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

Let  $\Omega$  be the set of real numbers  $R$  and  $\sigma$ -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.

10 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, we define the **outer measure** as  $m^n(A) = \inf\{\sum_{i=1}^{+\infty} I^n(A_i)\}$ , where  $\inf\{\cdot\}$  is the infimum,  $A_i = [x: a_k \leq x_k \leq b_k, k = 1, 2, \dots, n]$  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  $A$ .  
15 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, 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  $\sigma$ -algebra** of  $R^n$ . The Lebesgue measure of any subset in  $R^n$  also coincides with its volume.  
20

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

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

5 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

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

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

## 2.2 Stochastic calculus

10 We have incorporated some necessary concepts and theorems of stochastic calculus. All the classic theorems have been 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 with respect to stochastic processes, such as Brownian motion. **Brownian motion** is one of the simplest stochastic processes.

15 The Brownian motion  $W$  that is defined on a probability measure space  $(\Omega, F, P)$  is characterized as follows:

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

The last two conditions represent that  $\forall t_2 > s_1 > t_1 > s_1 \geq 0$ ,  $W(t_2) - W(s_2)$  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  $P$ .

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 transition probability, volatility and Brownian motion, respectively. The integral form of

25 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  $I(0) + \int_0^t \varphi(u) du$ .

5 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 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  
 10 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

$$f(t, x(t)) = f(0, x(0)) + \int_0^t f_u(u, x(u)) du + \int_0^t f_x(u, x(u)) \sigma(u) dW(u) + \int_0^t f_x(u, x(u)) \varphi(u) du +$$

$$15 \frac{1}{2} \int_0^t f_{xx}(u, x(u)) \sigma^2(u) du . \quad (4)$$

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

### 3 Reformulation of scale transformation in data assimilation framework

#### 20 3.1 Definition of scale

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

(i) Measure of a single point measurement: 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.

25 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.

(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 measurement: The observed space of a footprint measurement is any bounded closed domain  $A$ , most of which are not regular rectangles, such 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  $A_i = [x: a_k \leq x_k \leq b_k, k = 1, 2]$  and  $A \subseteq \bigcup_{i=1}^{+\infty} A_i$ . Obviously, measure (i)~(iii) are the special cases of the measure of a footprint measurement.

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  $\mu_v(A)$  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 cover 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 in  $R^2$ .

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:

$$s_2 = |J(x_1, x_2)| \iint_{A_1} dx_1 dx_2 = \xi^2 s_1, \quad (5)$$

where  $s_1$  and  $s_2$  represent a **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 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 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, where  $A$  is a disc observation footprint with radius  $r$ . The mapping function between  $A$  and  $A_0$  is  $T(x_1, x_2) = [x_1 \cos x_2, x_1 \sin x_2; 0 \leq x_1 \leq r, 0 \leq x_2 \leq 2\pi] = [y_1, y_2]$ , and the Jacobian

5 determinant  $|J(x_1, x_2)| = \begin{vmatrix} \cos x_2 & -x_1 \sin x_2 \\ \sin x_2 & x_1 \cos x_2 \end{vmatrix} = x_1$ . Therefore,  $s = \iint_A dy_1 dy_2 = \iint_{A_0} |J(x_1, x_2)| dx_1 dx_2 = \pi r^2$ , which is equal to its area. However,  $s_0$  and  $s$  do not obey one-dimensional rule because the Jacobian matrix is not diagonal.

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 interval  $A_0$ .

Let  $m_{C_1}^2, m_{C_2}^2$  and  $m_{C_3}^2$  be the Lebesgue measures of disc measurements  $C_1, C_2$  and  $C_3$ , respectively, and let  $m_{D_1}^2$  and  $m_{D_2}^2$  be the Lebesgue measures of diamond

10 measurements  $D_1$  and  $D_2$ . Then,  $m_{C_1}^2 = m_{C_2}^2 = m_{C_3}^2$  because they are the same function. 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,  $A_i + 2 = [x: x_k + 2, x_k \in A_i, k = 1, 2]$  and  $A_i \times$

$3 + 2 = [x: x_k \times 3 + 2, x_k \in A_i, k = 1, 2]$ , which proves that  $m_{C_1}^2, m_{C_2}^2$  and  $m_{C_3}^2$  collect the desirable set based on the same scheme, so they are identical. Additionally,  $\sum I^2(A_i \times 3 + 2)$  is much larger than  $\sum I^2(A_i)$ . Therefore, the scale of  $C_2$  is not

15 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,  $m_{D_1}^2 = m_{D_2}^2$ ; although their scales are different, they obey a one-dimensional rule.

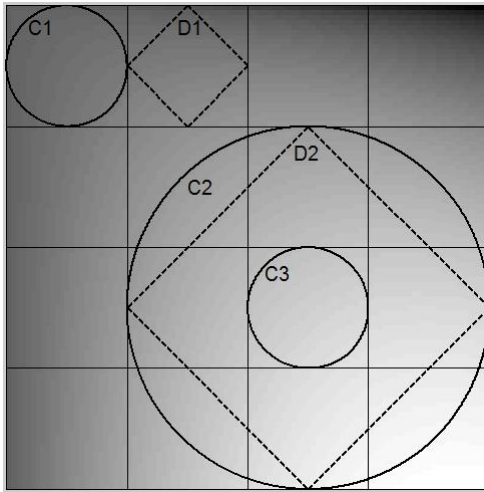


Figure 1. Diagram of the Relationships among a Lebesgue Measure, Scale and Variable

### 3.2 Stochastic variables in data assimilation

We introduce the widely accepted Bayesian theorem of data assimilation (Lorenz, 1995; van Leeuwen, 2015) to investigate its  
 5 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

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

10 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

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

15 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) 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.

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

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



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. (6) and (7), 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 scale) state and observation.

Let a real function  $V(s, t)$  be the **variable** if it maps the space  $(R^2, \mathcal{L}^2, m^2)$  onto  $R$ , where  $s$  is the scale,  $s = m^2(A)$ ,  $A \subset R^2$ , and  $t$  is the time. A variable is an estimation of a geophysical parameter at a specific scale  $s$  and time  $t$ .

10 In Figure 1, let the each pixel intensity is the estimator of a geophysical parameter in each pixel, then this parameter is heterogeneous in the entire region. A variable represents an ensemble average in a specific observation footprint with a specific scale. Therefore, the variables  $C1$  and  $C3$  are not equal because their observation footprints are different, and the variables  $C2$  and  $C3$  are also different because the scale changes. The former introduces the variables that vary with location, and the latter states that the 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.

15 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 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 are available, we can construct an explicit stochastic equation for the variable.

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

$$20 \quad dV = p(t)dt + q(t)dW(t). \quad (9)$$

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

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

25 where  $\varphi(s)$  and  $\sigma(s)$  are the scale-based transition probability and volatility, respectively. The variable is a probabilistic process with respect to scale and thus has scale-dependent errors.

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 case 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.

30 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 can be expressed by Eq. (9). 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 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 data assimilation framework

First, we provide the following lemma.

5 **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  
10 motion, with  $E[W^*(s_{i+1}) - W^*(s_i)] = 0$  and  $Var[W^*(s_{i+1}) - W^*(s_i)] = s_{i+1} - s_i$ .  $\square$

**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  $W^*(s), s \geq 0$ .

In the following content, we use Brownian motion with a parameter that starts at  $s_0$  to define the scale-dependent variables;  
15 therefore some classic expressions above should be 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

$$20 \quad V_0 + \int_{s_0}^s \varphi(u) du.$$

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.

25 **Assumption 1:** The measures of the state and observation in 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, and both 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  
30 independent.

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

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:

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

5 and

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

where  $\varphi_X(s)$ ,  $\sigma_X(s)$ ,  $\varphi_Y(s)$  and  $\sigma_Y(s)$  represent the scale-dependent transition probabilities and volatilities of state  $X$  and observation  $Y$ , respectively.

Assumption 3 implies that the scales of the state and observation are invariant when observational information is added in the analysis step. For simplicity, assume the scale-based 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 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

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

For the observation, we have

$$15 \quad dY = dW(s) \text{ and } Y(s_Y) = Y_0 + \int_{s_0}^{s_Y} dW(s) \quad (16)$$

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, scale of the observation, state in  $s_0$  and observation in  $s_0$ , respectively.

The Bayesian equation of data assimilation (Eq. (8)) 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

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

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

Thus, the last point is to calculate  $p(X|Y)$ .

25 The scale-dependent observation operator is  $H(s, X(s))$ , which suggests that the observation operator and its parameters are both susceptible to the scale. If  $H(s, X(s))$  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 have

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

30 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

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

5 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 observational 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 observation error.

The quadratic variation of Eq. (20) is

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

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 dependent as well.

In particular, if  $Y$  is a direct 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

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

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

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 error, 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 if  $Y$  is an indirect measurement of  $X$ , 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  $X$  (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 varies randomly, the posterior PDF is more complex because 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  $X(s_i)$ , where  $s_i$  is its scale, and  $X(s_X)$  is the ideal initial state in the model space that is related to  $X(s_i)$ . Then,  $X(s_X)$  has the same scale  $s_X$  as the forecasting operator. From Eq. (28) we obtain the error:

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

where  $M(s_X, X(s_i))$  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  $s_i$ .

### 3.4 Extension to n-dimensional data assimilation

In the above discussion, we assumed that only one variable existed in data assimilation. However, numerous states typically exist. This section further introduces the **product spaces** to extend the one-dimensional 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  $\mu^n$  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 analysis of a single state can also be applied to finite multiple states in the product measure space.

## 4 Summary & Discussion

### 4.1 Summary

In this study, we mainly addressed two basic problems. First, we produced a mathematical formalism of scale. Second, we demonstrated how scale transformation could be evaluated 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, we 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 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. Bocquet 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.

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 path to investigate how heterogeneous observations vary with scale. Our study also permits the representativeness error to be 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  $s_0 = s_Y$ , then Eq. (23) becomes  $P(Y|X) = N\{Y(s_Y) - H(s_Y, X(s_Y)), |s_Y - s_X|\}$ . 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  $M(s_X, X(0, s_i)) - M(s_X, X(0, s_X)) = |W(s_i) - W(s_X)|$  under the linearity assumption, which proves that the error that is associated with the initial state greatly depends on the Brownian motion increment from scale mismatching.

### 4.3 Next step

We conducted an integrated study that considered both the geometric transformation of an observation footprint and the variation in geophysical parameters. This integrated study included all possible situations and predictably conformed to each scale-related case study. That is, a case study could be considered a particular solution to a stochastic calculus equation, for which the scales and scale-dependent Brownian motions evolved in distinct 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.

## 10 5 Notation

### 5.1 Basic notations

	$\Omega$	Observational region
	$\mathcal{F}$	$\sigma$ -algebra
	$\mu$	Measure
15	$dV$	Variable process
	$W(s)$	Brownian motion
	$(\Omega, \mathcal{F}, \mu)$	Measure space
	$I^n$	N-dimensional Lebesgue volume
	$m^n$	Lebesgue measure or an outer measure on $R^n$
20	$\mathcal{L}^n$	Lebesgue $\sigma$ -algebra of $R^n$
	$\int f dm^n$	Lebesgue integral
	$ 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 or evaluate a geophysical parameter	Sect. 1 & Sect. 3.1
$A_0$	Unit 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. (5)
$V$	Variable	Estimation of a geophysical parameter at a specific scale	Sect. 3.2
$dX$	State process	State in the sense of the Ito process	Eq. (13)
$dY$	Observation process	Observation in the sense of Ito process	Eq. (14)
$X_0$	State in $s_0$		Eq. (15)
$Y_0$	Observation in $s_0$		Eq. (16)
$s_X$	Scale of state		Eq. (15)
$s_Y$	Scale of observation		Eq. (16)

---

## Acknowledgements

We thank the editor-in-chief of NGP, Prof. Talagrand, and his kind considerations 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.
- Giménez, D., Rawls, W. J., and Lauren, J. G.: Scaling properties of saturated hydraulic conductivity in soil, *Geoderma*, 27, 115-130, 1999.
- 5 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.
- 10 [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: 15 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.
- [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.](#)
- 20 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.
- 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.
- 25 [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., Flüher, 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.
- 30 [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.
- 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.
- 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.
- Ryu, D. and Famiglietti, J. S.: Multi-scale spatial correlation and scaling behavior of surface soil moisture, *Geophys. Res. Lett.*, 33, 153-172, 2006.
- 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.
- 10 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.
- 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.
- 15 Wiens, J. A.: Spatial scaling in ecology, *Funct. Ecol.*, 3, 385-397. doi: 10.2307/2389612, 1989.