Formulation of Scale Transformation in a Stochastic Data Assimilation Framework

Feng Liu[1,3] and Xin Li[1,2,3]

[1]Key Laboratory of Remote Sensing of Gansu Province, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, 730000, China
[2]Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, P. R. China
[3]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), 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 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 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 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 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 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) $A_1, A_2, \cdots \in \mathcal{F}$ implies their union $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.

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

(1) $\mu(A) \in [0, \infty)$ and $\mu(\Phi) = 0$;

(2) If $A_1, A_2, \cdots \in \mathcal{F}$ 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 $(\Omega, \mathcal{F}, P)$ 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 $\forall A = [a, b] \in B$, 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 $A = [x: a_k \leq x_k \leq b_k, k = 1,2, \cdots, n]$ 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, $A_i = [x: a_k \leq x_k \leq b_k, \mathrm{k} = 1,2, \cdots, \mathrm{n}]$ 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, 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 Lebesgue measure of 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 $x = (x_1, \cdots, x_n)$.

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

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

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

2.2 Stochastic calculus

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. The Brownian motion W that is defined on a probability measure space (Ω, F, P) is characterized as follows:

1) $W(0) = 0$.
2) $\forall t > s \geq 0$, 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

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

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

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

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

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

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 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

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

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

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. 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): $\forall A = [x: a_k \leq x_k \leq b_k, \mathrm{k} = 1,2]$, 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 $J(x_1, x_2) = diag(\xi, \xi), \xi \in R$, 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 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_{C1}^2, m_{C2}^2 and m_{C3}^2 be the Lebesgue measures of disc measurements 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, $m_{C1}^2 = m_{C2}^2 = m_{C3}^2$ 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_{C1}^2, m_{C2}^2 and m_{C3}^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.

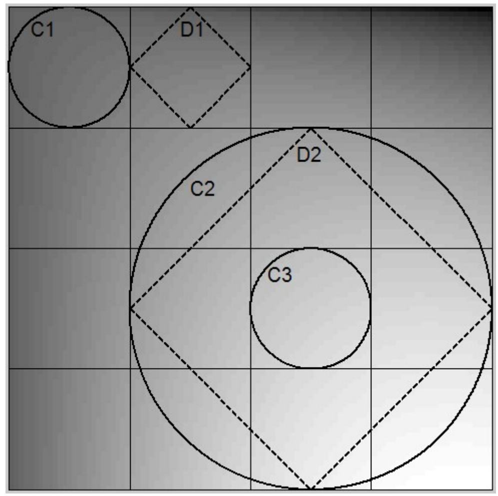

Figure 1. Diagram of the Relationships among a Lebesgue Measure, Scale and Variable

3.2 Stochastic variables in data assimilation

We introduce the widely accepted Bayesian theorem of data assimilation (Lorenc, 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

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

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

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

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; 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.

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

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

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 $p(X|Y)$.

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.

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.

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:

$$dV = p(t)dt + q(t)dW(t). \tag{9}$$

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

$$dV = \varphi(s)ds + \sigma(s)dW(s), \tag{10}$$

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.

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.

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

Proof. First, $W^*(0) = W^*(s_0) - W^*(s_0) = 0$. $\forall s_{i+1} > s_i \geq 0, i = 1,2,3, \ldots$, $W^*(s_{i+1}) - W^*(s_i) = [W(s_0 + s_{i+1}) - W(s_0)] - [W(s_0 + s_i) - W(s_0)] = W(s_0 + s_{i+1}) - W(s_0 + s_i)$, 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 $Var[W^*(s_{i+1}) - W^*(s_i)] = s_{i+1} - s_i$.□

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; 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

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

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.

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 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:

$$dX = \varphi_X(s)ds + \sigma_X(s)dW(s) \tag{13}$$

and

$$dY = \varphi_Y(s)ds + \sigma_Y(s)dW(s), \tag{14}$$

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

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

For the observation, we have

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

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 $X \sim N\left(X_0, \int_{s_0}^{s_X} ds\right)$ 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

$$X \sim N(X_0, s_X - s_0) \tag{17}$$

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

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

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

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

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

$Y(s_Y) - H\big(s_X, X(s_X)\big)$

$= Y_0 + \int_{s_0}^{s_Y} dW(u) - \Big[H(s_0, X_0) + \int_{s_0}^{s_X} H_s\big(u, X(u)\big)\,du + \int_{s_0}^{s_X} H_x\big(u, X(u)\big)\,dW(u) + \frac{1}{2}\int_{s_0}^{s_X} H_{xx}\big(u, X(u)\big)\,du\Big]$

$= Y_0 - H(s_0, X_0) + \int_{s_0}^{s_Y} dW(u) - \big[H\big(s_X, X(s_X)\big) - H\big(s_0, X(s_0)\big)\big] - \frac{1}{2}\int_{s_0}^{s_X} H_{xx}\big(u, X(u)\big)\,du - \int_{s_0}^{s_X} H_x\big(u, X(u)\big)\,dW(u)$

$= Y_0 - \Big[H\big(s_X, X(s_X)\big) + \frac{1}{2}\int_{s_0}^{s_X} H_{xx}\big(u, X(u)\big)\,du\Big] + \Big\{\int_{s_0}^{s_Y} dW(u) - \int_{s_0}^{s_X} H_x\big(u, X(u)\big)\,dW(u)\Big\}.$ (20)

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

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

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

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

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

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

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. $H\big(s, X(s)\big) = X(s)$. The result becomes

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

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

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

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

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

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

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

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

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:

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

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 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 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.

5 Notation

5.1 Basic notations

Ω	Observational region
$\mathcal{F}$	σ-algebra
μ	Measure
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
$\mathcal{L}^n$	Lebesgue σ-algebra of R^n
$\int f dm^n$	Lebesgue integral
$\lvert J(\cdot) \rvert$	Jacobian determinant
$(\Omega^n, \mathcal{F}^n)$	Product space

5.2 New notations

Notation	Name	Explanation	Index
$\boldsymbol{s}$	Scale	The observation footprint or model unit to measure or evaluate a geophysical parameter	Sect. 1 & Sect. 3.1
$\boldsymbol{A_0}$	Unit interval in R^2		Sect. 3.1
$\boldsymbol{s_0}$	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)
$\boldsymbol{V}$	Variable	Estimation of a geophysical parameter at a specific scale	Sect. 3.2
$\boldsymbol{dX}$	State process	State in the sense of the Ito process	Eq. (13)
$\boldsymbol{dY}$	Observation process	Observation in the sense of Ito process	Eq. (14)
$\boldsymbol{X_0}$	State in s_0		Eq. (15)
$\boldsymbol{Y_0}$	Observation in s_0		Eq. (16)
$\boldsymbol{s_X}$	Scale of state		Eq. (15)
$\boldsymbol{s_Y}$	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.

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.

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.

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 Ustinh, 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.

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.

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.

Lorenc, A.C.: Atmospheric Data Assimilation, Meteorological Office, Bracknell, 1995.

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.

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.

Wiens, J. A.: Spatial scaling in ecology, Funct. Ecol., 3, 385-397. doi: 10.2307/2389612, 1989.