Data assimilation in operator algebras

David Freeman[a], Dimitrios Giannakis[a,b,1] iD, Brian Mintz[a], Abbas Ourmazd[c] iD, and Joanna Slawinska[a] iD

We develop an algebraic framework for sequential data assimilation of partially observed dynamical systems. In this framework, Bayesian data assimilation is embedded in a nonabelian operator algebra, which provides a representation of observables by multiplication operators and probability densities by density operators (quantum states). In the algebraic approach, the forecast step of data assimilation is represented by a quantum operation induced by the Koopman operator of the dynamical system. Moreover, the analysis step is described by a quantum effect, which generalizes the Bayesian observational update rule. Projecting this formulation to finite-dimensional matrix algebras leads to computational schemes that are i) automatically positivity-preserving and ii) amenable to consistent data-driven approximation using kernel methods for machine learning. Moreover, these methods are natural candidates for implementation on quantum computers. Applications to the Lorenz 96 multiscale system and the El Niño Southern Oscillation in a climate model show promising results in terms of forecast skill and uncertainty quantification.

data assimilation | operator algebras | quantum information | Koopman operators | kernel methods

Since its inception in weather forecasting (1) and object tracking problems (2), sequential data assimilation, also known as filtering, has evolved into an indispensable tool in forecasting and uncertainty quantification of dynamical systems (3, 4). In its essence, data assimilation is a Bayesian inference framework: Knowledge about the state of the system at time t is described by a probability distribution p_t. The system dynamics acts on probability distributions, carrying along p_t to a time-dependent family $p_{t,\tau}$, which can be used to forecast observables of the system at time $t + \tau$, $\tau \geq 0$. When an observation is made, at time $t + \Delta t$, the forecast distribution $p_{t,\Delta t}$ is updated in an analysis step using Bayes' rule to a posterior distribution $p_{t+\Delta t}$, and the cycle is repeated.

In real-world applications, the Bayesian theoretical "gold standard" is seldom feasible to employ due to a variety of challenges, including high-dimensional nonlinear dynamics, nonlinear observation modalities, and model error. Weather and climate dynamics (5) represent a classical application domain where these challenges are prevalent due to the extremely large number of active degrees of freedom (which necessitates making dynamical approximations such as subgrid-scale parameterization) and nonlinear equations of motion and observation functions (which prevent direct application of Bayes' rule). Addressing these issues has stimulated the creation of a broad range of data assimilation techniques, including variational (6), ensemble (7), and particle (8) methods.

In this paper, we examine Bayesian data assimilation and its representation through finite-dimensional computational methods from an algebraic perspective. Our formulation employs different levels of description, depicted schematically in Fig. 1. We begin by assigning to a measure-preserving dynamical flow $\Phi^t : X \to X$, $t \in \mathbb{R}$, an algebra of observables (complex-valued functions of the state) $\mathfrak{A} = L^\infty(X, \mu)$, where X is the state space and μ the invariant measure. This algebra is a commutative, or abelian, von Neumann algebra (9) under pointwise function multiplication. The state space of $\mathfrak{A}$, denoted as $S(\mathfrak{A})$, is the set of continuous linear functionals $\omega : \mathfrak{A} \to \mathbb{C}$, satisfying the positivity condition $\omega(f^*f) \geq 0$ for all $f \in \mathfrak{A}$ and the normalization condition $\omega\mathbf{1} = 1$. Here, * denotes the complex conjugation of functions, and $\mathbf{1}$ is the unit of $\mathfrak{A}$, $\mathbf{1}(x) = 1$ for all $x \in X$. Every probability density $p \in L^1(X, \mu) \equiv \mathfrak{A}_*$ induces a state $\omega_p \in S(\mathfrak{A})$ that acts on $\mathfrak{A}$ as an expectation functional, $\omega_p f = \int_X fp \, d\mu$. Such states ω_p constitute the set of normal states of $\mathfrak{A}$, denoted as $S_*(\mathfrak{A})$.

Elements of $\mathfrak{A}$ evolve under the Koopman operator $U^t : \mathfrak{A} \to \mathfrak{A}$, which is the composition operator by the dynamical flow, $U^t f = f \circ \Phi^t$. Moreover, algebra states evolve under the transfer operator $P^t : S(\mathfrak{A}) \to S(\mathfrak{A})$, which is the adjoint of the Koopman operator, $P^t \omega = \omega \circ U^t$ (10–12). The space of normal states $S_*(\mathfrak{A})$ is invariant under P^t, and we have $P^t \omega_p = \omega_{U^{-t}p}$ for every probability density $p \in \mathfrak{A}_*$. In this picture, the evolution $p_t \mapsto p_{t,\tau}$ of the forecast density is represented by dynamics on $S_*(\mathfrak{A})$

[1]To whom correspondence may be addressed. Email: dimitrios.giannakis@dartmouth.edu.
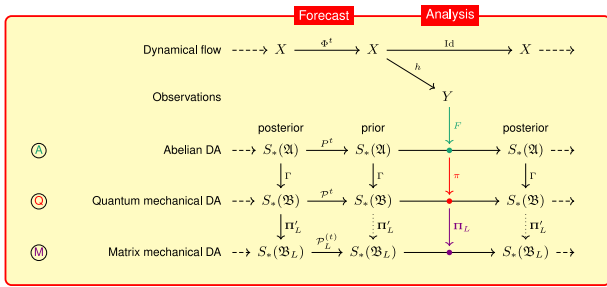
Fig. 1. Schematic representation of the abelian and nonabelian formulations of sequential data assimilation (DA), showing a forecast–analysis cycle. The *Top* row of the diagram shows the dynamical flow $\Phi^t : X \to X$. The second row shows the observation map $h : X \to Y$ used to update the state of the DA system in the analysis step. The rows labeled Ⓐ, Ⓠ, and Ⓜ show the abelian, infinite-dimensional nonabelian (quantum mechanical), and finite-dimensional nonabelian (matrix mechanical) DA systems, respectively. In Ⓐ, the forecast step is carried out by the transfer operator $P^t : S_*(\mathfrak{A}) \to S_*(\mathfrak{A})$ acting on states of the abelian algebra $\mathfrak{A}$. The analysis step (green dot) is represented by an effect-valued map $F : Y \to \mathfrak{A}$ that updates the state given observations in Y. In Ⓠ, the forecast step is carried out by the transfer operator $\mathcal{P}^t : S_*(\mathfrak{B}) \to S_*(\mathfrak{B})$ acting on states of the nonabelian operator algebra $\mathfrak{B}$. The analysis step (red dot) is carried out by an effect $\mathcal{F} : Y \to \mathfrak{B}$ given by the composition of F with the regular representation π of $\mathfrak{A}$ into $\mathfrak{B}$ (red arrow). The state space $S_*(\mathfrak{A})$ is embedded into $S_*(\mathfrak{B})$ by means of a map Γ, which is compatible with both forecast and analysis; Eqs. **4** and **9**. This compatibility is represented by the commutative loops between Ⓐ and Ⓠ having Γ as a vertical arrow. To arrive at the matrix mechanical DA, Ⓜ, we project $\mathfrak{B}$ into an L^2-dimensional operator algebra $\mathfrak{B}_L$ using a positivity-preserving projection Π_L. The composition of this projection with $\mathcal{F}$ leads to an effect $\mathcal{F}_L : Y \to \mathfrak{B}_L$ employed in the analysis step (purple arrow and dot). Moreover, Π_L induces a state space projection $\Pi_L' : S_*(\mathfrak{B}) \to S_*(\mathfrak{B}_L)$ and a projected transfer operator $\mathcal{P}_L^{(t)} : S_*(\mathfrak{B}_L) \to S_*(\mathfrak{B}_L)$ employed in the forecast step. Vertical dotted arrows indicate asymptotically commutative relationships that hold as $L \to \infty$.

under the transfer operator, $\omega_{p_{t,\tau}} = P^\tau \omega_{p_t}$. Moreover, Bayesian analysis, $p_{t,\Delta t} \mapsto p_{t+\Delta t}$, is represented by projective conditioning of the state. Together, these two steps encapsulate classical data assimilation within an abelian algebraic setting, labeled Ⓐ in Fig. 1.

The next level of our framework, labeled Ⓠ in Fig. 1, generalizes data assimilation by embedding it in a nonabelian operator algebra. Operator algebras form the mathematical backbone of quantum mechanics (13)—one of the most successful theories in physics. Quantum information theory and quantum probability provide a unified mathematical framework to characterize the properties of information transfer in both abelian and nonabelian systems through maps (quantum operations) acting on elements of the algebra and the corresponding states (14–16). Our nonabelian formulation Ⓠ is based on the von Neumann algebra $\mathfrak{B} \equiv B(H)$ of bounded linear operators on the Hilbert space $H = L^2(X, \mu)$, equipped with operator composition as the algebraic product. The state space $S(\mathfrak{B})$ is defined as the space of continuous, positive, normalized functionals analogously to $S(\mathfrak{A})$; that is, every state $\omega \in S(\mathfrak{B})$ satisfies $\omega(A^*A) \geq 0$ and $\omega I = 1$, where $*$ denotes the operator adjoint on $B(H)$ and I is the identity operator. Analogously to $\mathfrak{A}$, $S(\mathfrak{B})$ has a subset of normal states, $S_*(\mathfrak{B})$, induced in this case by trace-class operators. Specifically, letting $\mathfrak{B}_* \equiv B_1(H)$ denote the space of trace-class operators in $B(H)$, every positive operator $\rho \in \mathfrak{B}_*$ of unit trace induces a state $\omega_\rho \in S_*(\mathfrak{B})$ such that $\omega_\rho A = \mathrm{tr}(\rho A)$. Such operators ρ are called density operators and can be thought of as nonabelian analogs of probability densities $p \in \mathfrak{A}_*$. As we will see below, analogs of the transfer operator and Bayesian update described above for $S_*(\mathfrak{A})$ are naturally defined for $S_*(\mathfrak{B})$.

To arrive at practical data assimilation algorithms, we project (discretize) the infinite-dimensional system on $\mathfrak{B}$ to a system on a finite-dimensional subalgebra $\mathfrak{B}_L \subset \mathfrak{B}$, which is concretely represented by an $L \times L$ matrix algebra (Ⓜ in Fig. 1). We show that this approach leads to computational techniques which are well suited for assimilation of high-dimensional observables, while enjoying structure-preservation properties that cannot be obtained from orthogonal projections of abelian function spaces. Moreover, by virtue of being rooted in linear operator theory, these methods are amenable to consistent data-driven approximation using kernel methods for machine learning.

Previous Work. Recently, an operator-theoretic framework for data assimilation, called quantum mechanical data assimilation (QMDA) (17), was developed using ideas from Koopman operator theory (10, 12) in conjunction with the Dirac–von Neumann axioms of quantum dynamics and measurement (18). In QMDA, the state of the data assimilation system is a density operator ρ_t acting on $H = L^2(X, \mu)$ (rather than a classical probability density $p_t \in L^1(X, \mu)$), and the assimilated observables are multiplication operators in $B(H)$ (rather than functions in $L^\infty(X, \mu)$). Between observations, ρ_t evolves under the induced action of the transfer operator, and the forecast distribution of observables is obtained as a quantum mechanical expectation with respect to ρ_t. During observations, the density operator ρ_t is updated projectively as a von Neumann measurement, which is the quantum analog of Bayes' rule. QMDA has a data-driven formulation based on kernel methods for Koopman operator approximation (19–21), which was shown to perform well in low-dimensional applications. Meanwhile, the paper (22) has shown that Koopman operators of systems with pure point spectra can be approximated on quantum computers using shallow quantum circuits, offering an exponential computational advantage over classical deterministic algorithms for Koopman operator approximation.

Contributions. We provide a general algebraic framework that encompasses classical data assimilation and QMDA as particular instances (abelian and nonabelian, respectively). The principal distinguishing aspects of this framework are as follows:

1. *Dynamical consistency.* We employ a dynamically consistent embedding of abelian data assimilation Ⓐ into the nonabelian framework Ⓠ. As in ref. 17, observables in $\mathfrak{A}$ are mapped into multiplication operators in $\mathfrak{B}$, but here, we also employ an embedding $\Gamma : S_*(\mathfrak{A}) \to S_*(\mathfrak{B})$ that is compatible with the transfer operator (see the commutative loops between Ⓐ and Ⓠ in Fig. 1). This allows us to study QMDA in relation to the underlying classical theory and establish the consistency between the two approaches.

2. *Effect system.* In both the abelian and nonabelian settings, the analysis step, given observations in a space Y acquired through an observation map $h : X \to Y$, is carried out using quantum effects (loosely speaking, algebra-valued logical predicates) (14). In the abelian case, the effect $F : Y \to \mathfrak{A}$ is induced by a kernel feature map. In the nonabelian setting, F is promoted to an operator-valued feature map $\mathcal{F} : Y \to \mathfrak{B}$; see the column in the schematic of Fig. 1 labeled "Analysis." Our use of feature maps enables assimilation of data of arbitrarily large dimension, overcoming an important limitation of the original QMDA scheme (17) (which becomes prohibitively expensive for high-dimensional observation maps).

3. *Positivity-preserving discretization.* The discretization procedure leading to the finite-dimensional scheme Ⓜ has the important property that positive elements of $\mathfrak{B}$ are mapped into positive elements of $\mathfrak{B}_L$. Moreover, the transfer operator on $S_*(\mathfrak{B})$ is mapped into a completely positive, trace-nonincreasing map, so the finite-dimensional data assimilation system is a quantum operation. We call this system "matrix mechanical." By virtue of these properties, the sign of sign-definite observables, i.e., observables which are either positive or negative, is preserved. Relevant examples include positive physical quantities such as mass and temperature but also statistical quantities such as probability density or standard deviation, which are useful for uncertainty quantification. We emphasize that the approach of first embedding classical data assimilation in $\mathfrak{A}$ to the nonabelian operator setting of $\mathfrak{B}$ and then projecting to the finite-dimensional system on $\mathfrak{B}_L$ is important in positivity preservation.

4. *Data-driven formulation.* The matrix mechanical system Ⓜ on $\mathfrak{B}_L$ admits a data-driven approximation in which all operators are represented in a kernel eigenbasis learned from time-ordered training data, without requiring a priori knowledge of the equations of motion. In the limit of large training data, predictions made by the data-driven assimilation system converge to those from Ⓜ, which in turn converge to those from the infinite-dimensional system Ⓠ on $\mathfrak{B}$ as $L \to \infty$.

5. *Route to quantum computing.* QMDA is well suited for implementation on quantum computers as we demonstrate here with simulated quantum circuit experiments. Our approach provides a route to quantum algorithms that sequentially alternate between unitary evolution and projective measurement to perform inference and prediction of classical dynamics with quantum computers.

To place this work in the context of our previous work (17, 22), we note that ref. 17 proposed QMDA as an ad hoc data assimilation scheme, without attempting to connect it to the underlying classical Bayesian framework. Moreover, ref. 17 did not investigate the positivity-preserving aspects of QMDA, nor did it employ an operator-valued feature map in the analysis step. Ref. 22 developed a scheme for quantum computational simulation without addressing data assimilation. There, the focus was on representing classical dynamics via quantum circuits of low complexity (depth); to that end, the scheme was limited to systems with pure point spectra, which do not include systems with mixing (chaotic) dynamics that lie in the scope of QMDA.

Embedding Data Assimilation in Operator Algebras

Consider a dynamical flow $\Phi^t : X \to X$, $t \in \mathbb{R}$, on a completely metrizable, separable space X with an ergodic, invariant, Borel probability measure μ. The flow induces Koopman operators $U^t : f \mapsto f \circ \Phi^t$, which are isomorphisms of the $L^p(X, \mu)$ spaces with $p \in [1, \infty]$. The flow also induces transfer operators $P^t : L^p(X, \mu)^* \to L^p(X, \mu)^*$ on the dual spaces $L^p(X, \mu)^*$, given by the adjoint of the Koopman operator, $P^t\alpha = \alpha \circ U^t$. Under the canonical identification of $L^p(X, \mu)^*$, $p \in [1, \infty)$, with finite, complex Borel measures with densities in $L^q(X, \mu)$, $\frac{1}{p} + \frac{1}{q} = 1$, the transfer operator is identified with the inverse Koopman operator; that is, for $\alpha \in L^p(X, \mu)^*$ with density $\varrho = \frac{d\alpha}{d\mu} \in L^q(X, \mu)$, $\alpha_t := P^t\alpha$ has density $\varrho_t = \frac{d\alpha_t}{d\mu} \in L^q(X, \mu)$ with

$\varrho_t = U^{-t}\varrho$. In what follows, $\|f\|_{L^p(X,\mu)} = (\int_X |f|^p \, d\mu)^{1/p}$ for $p \in [1, \infty)$ and $\|f\|_{L^\infty(X,\mu)} = \lim_{p\to\infty}\|f\|_{L^p(X,\mu)}$ denote the standard $L^p(X, \mu)$ norms.

Among the $L^p(X, \mu)$ spaces, $H := L^2(X, \mu)$ is a Hilbert space, and $\mathfrak{A} := L^\infty(X, \mu)$ is an abelian von Neumann algebra with respect to function multiplication and complex conjugation. In particular, for any two elements $f, g \in \mathfrak{A}$, we have

$$\|fg\|_{\mathfrak{A}} \le \|f\|_{\mathfrak{A}}\|g\|_{\mathfrak{A}}, \quad \|f^*f\|_{\mathfrak{A}} = \|f\|_{\mathfrak{A}}^2, \qquad \textbf{[1]}$$

making $\mathfrak{A}$ a C^*-algebra, and moreover, $\mathfrak{A}$ has a predual $\mathfrak{A}_* := L^1(X, \mu)$ (i.e., a Banach space whose dual is $\mathfrak{A}$), making it a von Neumann algebra. We let $\langle f, g \rangle = \int_X f^*g \, d\mu$ denote the inner product on H. On H, the Koopman operator is unitary, $U^{t*} = U^{-t}$.

Embedding Observables. Let $\mathfrak{B} := B(H)$ be the space of bounded operators on H, equipped with the operator norm, $\|A\|_{\mathfrak{B}} = \sup_{f \in H} \frac{\|Af\|_H}{\|f\|_H}$. This space is a nonabelian von Neumann algebra with respect to operator composition and adjoint. That is, for any $A, B \in \mathfrak{B}$, we have

$$\|AB\|_{\mathfrak{B}} \le \|A\|_{\mathfrak{B}}\|B\|_{\mathfrak{B}}, \quad \|A^*A\|_{\mathfrak{B}} = \|A\|_{\mathfrak{B}}^2,$$

which is the nonabelian analog of Eq. **1** making $\mathfrak{B}$ a C^*-algebra. Moreover, $\mathfrak{B}$ has a predual, $\mathfrak{B}_* := B_1(H)$, making it a von Neumann algebra. Here, the space of trace-class operators $B_1(H) \subseteq B(H)$ is equipped with the norm $\|A\|_1 := \operatorname{tr}\sqrt{A^*A}$, which can be thought of as a nonabelian analog of $L^1(X, \mu)$. The unitary group of Koopman operators U^t on H induces a unitary group $\mathcal{U}^t : \mathfrak{B} \to \mathfrak{B}$ (i.e., a group of linear maps mapping unitary operators to unitary operators), which acts by conjugation, i.e.,

$$\mathcal{U}^t A = U^t A U^{t*}. \qquad \textbf{[2]}$$

The abelian algebra $\mathfrak{A}$ embeds isometrically into $\mathfrak{B}$ through the map $\pi : \mathfrak{A} \to \mathfrak{B}$, such that πf is the multiplication operator by f, $(\pi f)g = fg$. This map is injective, and satisfies $\pi(fg) = (\pi f)(\pi g)$, $\pi(f^*) = (\pi f)^*$ for all $f, g \in \mathfrak{A}$. Thus, π is a *-representation, preserving the von Neumann algebra structure of $\mathfrak{A}$. The representation π is also compatible with Koopman evolution, in the sense that $\mathcal{U}^t \circ \pi = \pi \circ U^t$ holds for all $t \in \mathbb{R}$. Equivalently, we have the commutative diagram

$$
\begin{array}{ccc}
\mathfrak{A} & \xrightarrow{U^t} & \mathfrak{A} \\
{\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\
\mathfrak{B} & \xrightarrow{\mathcal{U}^t} & \mathfrak{B}
\end{array},
$$

which shows that π provides a dynamically consistent representation of observables of the dynamical system in $\mathfrak{A}$ as elements of the nonabelian operator algebra $\mathfrak{B}$. In Fig. 1, we refer to the level of description involving $\mathfrak{B}$ as quantum mechanical due to the central role that operator algebras play in the algebraic formulation of quantum mechanics (13). In particular, Eq. **2** is mathematically equivalent to the Heisenberg picture for the unitary evolution of quantum observables (here, under the Koopman operator).

Embedding States. A dual construction to the representation $\pi : \mathfrak{A} \to \mathfrak{B}$ of observables can be carried out for states. Let $\omega_p \in S_*(\mathfrak{A})$ be a normal state induced by a probability density

$p \in \mathfrak{A}_*$. Since p is a positive function with $\|p\|_{\mathfrak{A}_*} = 1$, we have that $\sqrt{p}$ is a real unit vector in H, and thus $\rho = \langle \sqrt{p}, \cdot \rangle \sqrt{p}$ is a rank-1 orthogonal projection. Every such projection is a density operator inducing a normal state $\omega_\rho \in S_*(\mathfrak{B})$, where

$$\omega_\rho A = \mathrm{tr}(\rho A) = \langle \sqrt{p}, A\sqrt{p} \rangle, \quad \forall A \in \mathfrak{B}. \qquad [3]$$

Such states ω_ρ induced by unit vectors in H are called vector states. In fact, ω_ρ is a pure state, i.e., it is an extremal point of the state space $S(\mathfrak{B})$ (which is a convex set). Defining the map $\Gamma : S_*(\mathfrak{A}) \to S_*(\mathfrak{B})$ such that $\Gamma(\omega_p) = \omega_\rho$, one can readily verify that Γ is compatible with the regular representation π; i.e., for every observable $f \in \mathfrak{A}$ and probability density $p \in \mathfrak{A}_*$,

$$\omega_p f = \Gamma(\omega_p)(\pi f). \qquad [4]$$

Next, analogously to the transfer operator $P^t : S(\mathfrak{A}) \to S(\mathfrak{A})$, we define $\mathcal{P}^t : S(\mathfrak{B}) \to S(\mathfrak{B})$ as the adjoint of $\mathcal{U}^t : \mathfrak{B} \to \mathfrak{B}$ from Eq. 2, $\mathcal{P}^t \omega = \omega \circ \mathcal{U}^t$. Note that $\mathcal{U}^t$ and $\mathcal{P}^t$ form a dual pair, i.e., $(\mathcal{P}^t \omega)A = \omega(\mathcal{U}^t A)$ for every state $\omega \in S(\mathfrak{B})$ and element $A \in \mathfrak{B}$. Moreover, if $\omega_\rho \in S_*(\mathfrak{B})$ is induced by a density operator $\rho \in \mathfrak{B}_*$, then $\mathcal{P}^t \omega_\rho = \omega_{\rho_t}$, where ρ_t is the density operator given by $\rho_t = \mathcal{U}^{-t}\rho = U^{t*}\rho U^t$.

In quantum mechanics, the evolution $\rho \mapsto \rho_t$ is known as the Schrödinger picture, and it is the dual of the Heisenberg picture from Eq. 2. In the particular case that $\rho = \langle \xi, \cdot \rangle \xi$ is a vector state induced by $\xi \in H$ (which would be called a wavefunction in quantum mechanical language), we have $\rho_t = \langle \xi_t, \cdot \rangle \xi_t$, where $\xi_t = U^{t*}\xi$. Using this fact and Eq. 3, it follows that Γ is compatible with the evolution on $S_*(\mathfrak{A})$ and $S_*(\mathfrak{B})$ under the transfer operator; that is, $\mathcal{P}^t \circ \Gamma = \Gamma \circ P^t$. This relation is represented by the commutative diagram

$$
\begin{array}{ccc}
S_*(\mathfrak{A}) & \xrightarrow{\;P^t\;} & S_*(\mathfrak{A}) \\
{\scriptstyle \Gamma}\downarrow & & \downarrow{\scriptstyle \Gamma} \\
S_*(\mathfrak{B}) & \xrightarrow{\;\mathcal{P}^t\;} & S_*(\mathfrak{B})
\end{array}
\qquad [5]
$$

which also captures the correspondence between the abelian and quantum mechanical forecast steps in Fig. 1.

Probabilistic Forecasting. In both abelian and nonabelian data assimilation, we can describe probabilistic forecasting of observables of the dynamical system using the formalism of positive operator-valued measures (POVMs) (23). First, we recall that an element a of a C^*-algebra $\mathfrak{W}$ is i) self-adjoint if $a^* = a$; ii) positive (denoted as $a \geq 0$) if $a = b^*b$ for some $b \in \mathfrak{W}$; and iii) a projection if $a^* = a = a^2$. Supposing that $\mathfrak{W}$ is also a von Neumann algebra, a map $E : \Sigma \to \mathfrak{W}$ on the σ-algebra of a measurable space (Ω, Σ) is said to be a POVM if i) for every set $S \in \Sigma$, $E(S) \geq 0$; ii) $E(\Omega) = I$, where I is the unit of $\mathfrak{W}$; and iii) for every countable collection $S_1, S_2, \ldots$ of disjoint sets in Σ, $E(\bigcup_i S_i) = \sum_i E(S_i)$, where the sum converges in the weak-* topology of $\mathfrak{W}$ (i.e., for every $\gamma \in \mathfrak{W}_*$, $E(\bigcup_i S_i)\gamma = \lim_{n\to\infty} \sum_{i=1}^n E(S_i)\gamma$.) These properties imply that for every $\gamma \in \mathfrak{W}_*$, the map $\mathbb{P}_{E,\gamma} : \Sigma \to \mathbb{C}$ given by

$$\mathbb{P}_{E,\gamma}(S) = E(S)\gamma, \qquad [6]$$

is a complex normalized measure. In particular, if γ induces a normal state $\omega_\gamma \in S_*(\mathfrak{W})$, then $\mathbb{P}_{E,\gamma}$ is a probability measure on Ω. We say that the POVM E is a projection-valued measure (PVM) if $E(S)$ is a projection for every $S \in \Sigma$.

In quantum mechanics, a triple (Ω, Σ, E) where E is a POVM is referred to as an observable. We alert the reader to the fact that in dynamical systems theory, an observable is generally understood as a function $f : X \to V$ on state space X taking values in a vector space, V. Thus, in situations where the space of dynamical observables forms an algebra (e.g., the $\mathfrak{A} = L^\infty(X, \mu)$ algebra corresponding to $V = \mathbb{C}$), the term "observable" is overloaded, and its meaning must be understood from the context.

Given a POVM (Ω, Σ, E) as above, and a bounded, measurable function $u : \Omega \to \mathbb{C}$, we define the integral $\int_\Omega u(\omega) \, dE(\omega)$ as the unique element a of $\mathfrak{W}$ such that for every $\gamma \in \mathfrak{W}_*$, $a\gamma = \int_\Omega u(\omega) \, d\mathbb{P}_{E,\gamma}(\omega)$. If a is a self-adjoint element of $\mathfrak{W}$, i.e., $a^* = a$, the spectral theorem states that there exists a unique PVM $E : \mathcal{B}(\mathbb{R}) \to \mathfrak{W}$ on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ of the real line such that $a = \int_\mathbb{R} \omega \, dE(\omega)$.

In abelian data assimilation, the self-adjoint elements are the real-valued functions f in the von Neumann algebra $\mathfrak{A}$, and every such f has an associated PVM $E_f : \mathcal{B}(\mathbb{R}) \to \mathfrak{A}$. Explicitly, we have $E_f(S) = \chi_{f^{-1}(S)}$, where $\chi_{f^{-1}(S)} : X \to \mathbb{R}$ is the characteristic function of the set $f^{-1}(S) \subseteq X$. If, at time t, the data assimilation system is in a state $\omega_{p_t} \in S_*(\mathfrak{A})$ induced by a probability density $p_t \in \mathfrak{A}_*$, then the forecast distribution for f at lead time $\tau \geq 0$ is $\mathbb{P}_{f,t,\tau} \equiv \mathbb{P}_{E_f,p_\tau}$, where $p_{t,\tau}$ is the probability density associated with the state $\omega_{p_{t,\tau}} = P^\tau \omega_{p_t}$.

The forecast distribution $\mathbb{P}_{f,t,\tau}$ is equivalent to the distribution obtained via classical probability theory. That is, given an observable $f \in L^\infty(X, \mu)$, the density $p_{t,\tau} \in L^1(X, \mu)$ induces a probability measure on $\mathbb{R}$ such that $\mathrm{Prob}(S) = \int_{f^{-1}(S)} p_{t,\tau} \, d\mu$ is the probability that f takes values in a set $S \in \mathcal{B}(\mathbb{R})$. It follows by definition of $\mathbb{P}_{f,t,\tau}$ that $\mathrm{Prob}(S) = \mathbb{P}_{f,t,\tau}(S)$.

In the non-abelian setting of $\mathfrak{B}$, the spectral theorem states that for every self-adjoint operator $A \in \mathfrak{B}$, there exists a unique PVM $E_A : \mathcal{B}(\mathbb{R}) \to \mathfrak{B}$, such that $A = \int_\mathbb{R} y \, dE_A(y)$. If, at time t, the nonabelian data assimilation system is in a normal state ω_{ρ_t} induced by a density operator $\rho_t \in \mathfrak{B}_*$, then the forecast distribution for A at lead time $\tau \geq 0$ is given by $\mathbb{P}_{A,t,\tau} \equiv \mathbb{P}_{E_A,\rho_{t,\tau}}$, where $\rho_{t,\tau}$ is the density operator associated with $\omega_{\rho_{t,\tau}} = \mathcal{P}^\tau \omega_{\rho_t}$. This distribution is compatible with the embeddings of states $\Gamma : S_*(\mathfrak{A}) \to S_*(\mathfrak{B})$ and observables $\pi : \mathfrak{A} \to \mathfrak{B}$ introduced above. That is, for every observable $f \in \mathfrak{A}$, probability density $p_{t,\tau} \in S_*(\mathfrak{A})$, and Borel set $S \in \mathcal{B}(\mathbb{R})$, we have $\mathbb{P}_{f,p_{t,\tau}}(S) = \mathbb{P}_{\pi f, \rho_{t,\tau}}(S)$, where $\Gamma(\omega_{p_{t,\tau}}) = \omega_{\rho_{t,\tau}}$.

Representing Observations by Effects. For a unital C^*-algebra $\mathfrak{W}$, an effect is an element $e \in \mathfrak{W}$ satisfying $0 \leq e \leq I$. Intuitively, one can think of effects as generalizations of logical truth values, used to model outcomes of measurements or observations (24, 25). In Boolean logic, truth values lie in the set $\{0, 1\}$. In fuzzy logic, truth values are real numbers in the interval $[0, 1]$. In unital C^*-algebras, the analogs of truth values are elements e satisfying $0 \leq e \leq I$ (26). We denote the set of effects in a C^*-algebra $\mathfrak{W}$ as $\mathcal{E}(\mathfrak{W})$. It can be shown that $\mathcal{E}(\mathfrak{W})$ is a convex space, whose extremal points are projections. Given a state $\omega \in S(\mathfrak{W})$ and an effect $e \in \mathcal{E}(\mathfrak{W})$, the number $\omega e \in [0, 1]$ is called the validity of e. Note that every effect $e \in E(\mathfrak{W})$ induces a binary POVM $E : \{0, 1\} \to \mathfrak{W}$ such that $E(\{1\}) = e$ and $E(\{0\}) = I - e$.

Suppose now that $\mathfrak{W}$ is a von Neumann algebra, let $\omega_\rho \in S_*(\mathfrak{W})$ be a normal state induced by an element $\rho \in \mathfrak{W}_*$, and let $e \in \mathcal{E}(\mathfrak{W})$ be an effect. If the validity ωe is nonzero, we can

define the conditional state $\omega_\rho|_e \in S_*(\mathfrak{W})$ as the normal state induced by $\rho|_e \in \mathfrak{W}_*$, where

$$\rho|_e = \frac{\sqrt{e}\rho\sqrt{e}}{\omega_\rho e}. \qquad [7]$$

The map $\omega_\rho \mapsto \omega_\rho|_e$ generalizes the Bayesian conditioning rule employed in the analysis step of classical data assimilation.

As an example, let $\mathfrak{W} = \mathfrak{A}$ and $\chi_S : X \to \{0, 1\}$ be the characteristic function of measurable set (event) $S \in \mathcal{B}(X)$. According to Bayes' theorem, if $p \in \mathfrak{A}_*$ is a probability density and $\int_S p \, d\mu > 0$, the conditional density of p given S is

$$q = \frac{p\chi_S}{\int_X p\chi_S \, d\mu} = \frac{\sqrt{\chi_S}p\sqrt{\chi_S}}{\int_X p\chi_S \, d\mu}. \qquad [8]$$

Since $\chi_S(x) \in \{0, 1\}$ for every $x \in X$, it follows that χ_S is an effect in $\mathfrak{A}$, and since $\int_X p\chi_S \, d\mu = \omega_\rho \chi_S$, the Bayesian formula above is a special case of Eq. 7 with $p|_{\chi_S} = q$. Note that to obtain the second equality in Eq. 8, we made use of the commutativity of function multiplication, which does not hold in a nonabelian algebra.

An important compatibility result between effects in the abelian algebra $\mathfrak{A}$ and effects in the nonabelian algebra $\mathfrak{B}$ is as follows: The regular representation $\pi : \mathfrak{A} \to \mathfrak{B}$ maps the effect space $\mathcal{E}(\mathfrak{A})$ into the effect space $\mathcal{E}(\mathfrak{B})$. As a result, and by virtue of Eq. 4, for every normal state $\omega_\rho \in S_*(\mathfrak{A})$ and effect $e \in \mathcal{E}(\mathfrak{A})$, the conditioned state $\omega_\rho|_e$ satisfies

$$\Gamma(\omega_\rho|_e) = (\Gamma\omega_\rho)|_{\pi e}. \qquad [9]$$

This means that conditioning by effects in $\mathcal{E}(\mathfrak{A})$ consistently embeds to conditioning by effects in $\mathcal{E}(\mathfrak{B})$.

Next, let Y be a set. In first-order logic, a predicate is a map $F : Y \to \{0, 1\}$ such that $F(y) = 1$ means that the proposition $F(y)$ is true, and $F(y) = 0$ means that it is false. In fuzzy logic, predicates are generalized to maps $F : Y \to [0, 1]$. In quantum logic, predicates are represented by effect-valued maps $F : Y \to \mathcal{E}(\mathfrak{W})$. Applying Eq. 7 for $e = F(y)$ leads to the update rule $p \mapsto p|_{F(y)}$, which represents the conditioning of the normal state associated with p by the truth value of the proposition $F(y)$ associated with $y \in Y$.

In our algebraic data assimilation framework, we use an effect-valued map to carry out the analysis step given observations of the system in a space Y (Fig. 1). Specifically, let $h : X \to Y$ be a measurable observation map, such that $y = h(x)$ corresponds to the assimilated data given that the system is in state $x \in X$. Let $\psi : Y \times Y \to [0, 1]$ be a measurable kernel function on Y, taking values in the unit interval. Every such kernel induces an effect-valued map $F : Y \to \mathcal{E}(\mathfrak{A})$ given by $F(y) = \psi(y, h(\cdot))$. Possible choices for ψ include bump kernels—in such cases, $F(y)$ can be viewed as a relaxation of a characteristic function χ_S of a set S containing $h^{-1}(\{y\})$ (*SI Appendix*, Eq. S25).

If, immediately prior to an observation at time $t + \Delta t$, the abelian data assimilation system has state $\omega_{p_{t,\Delta t}} \in S_*(\mathfrak{A})$ (recall that $p_{t,\Delta t}$ is the forecast density for lead time Δt initialized at time t), and $F(y)$ has nonzero validity with respect to $\omega_{p_{t,\Delta t}}$, our analysis step updates $\omega_{p_{t,\Delta t}}$ to the conditional state $\omega_{p_{t,\Delta t}}|_{F(y)} \equiv \omega_{p_{t+\Delta t}}$ using Eq. 7. In the nonabelian setting, we promote F to the operator-valued function $\mathcal{F} : Y \to \mathcal{E}(\mathfrak{B})$ with $\mathcal{F} = \pi \circ F$ and use again Eq. 7 to update the prior state $\omega_{\rho_{t,\Delta t}} \in S_*(\mathfrak{B})$ to $\omega_{\rho_{t,\Delta t}}|_{\mathcal{F}(y)} \equiv \omega_{\rho_{t+\Delta t}}$; see the Analysis column of the schematic in Fig. 1. By Eq. 9, the abelian and nonabelian analysis steps are

mutually consistent, in the sense that if $\omega_{\rho_{t,\Delta t}} = \Gamma(\omega_{p_{t,\Delta t}})$, then for every observable $f \in \mathfrak{A}$, we have $\omega_{\rho_{t+\Delta t}}(\pi f) = \omega_{p_{t+\Delta t}} f$.

We should note that the effect-based analysis step introduced above can naturally handle data spaces Y of arbitrarily large dimension, overcoming an important limitation of the QMDA framework proposed in ref. 17. It is also worthwhile pointing out connections between effect-valued maps and feature maps from RKHS theory (27): If ψ is positive-definite, there is an associated RKHS $\mathcal{H}$ of complex-valued functions on X with $w(x, x') := \psi(h(x), h(x'))$ as its reproducing kernel. The map F then takes values in the space $\mathcal{E}(\mathfrak{A}) \cap \mathcal{H}$ and is thus an instance of a feature map. In the nonabelian case, one can think of $\mathcal{F}$ as an operator-valued feature map.

Positivity-Preserving Discretization. The abelian and nonabelian formulations of data assimilation described thus far employ the infinite-dimensional algebras $\mathfrak{A}$ and $\mathfrak{B}$, respectively. To arrive at practical computational algorithms, these algebras must be projected to finite dimensions, carrying along the associated dynamical and observation operators to finite-rank operators. We refer to this process as discretization.

To motivate our approach, we recall the definitions of quantum operations and channels (15): A linear map $\mathcal{T} : \mathfrak{W}_2 \to \mathfrak{W}_1$ between two von Neumann algebras $\mathfrak{W}_1$ and $\mathfrak{W}_2$ is said to be a quantum operation if i) $\mathcal{T}$ is completely positive, i.e., for every $n \in \mathbb{N}$, the tensor product map $\mathcal{T} \otimes \mathrm{Id}_n : \mathbb{M}_n(\mathfrak{W}_2) \to \mathbb{M}_n(\mathfrak{W}_1)$ is positive, where $\mathbb{M}_n(\mathfrak{W}_1)$ and $\mathbb{M}_n(\mathfrak{W}_2)$ are the von Neumann algebras of $n \times n$ matrices over $\mathfrak{W}_1$ and $\mathfrak{W}_2$, respectively; ii) $\mathcal{T}$ is the adjoint of a map $\mathcal{T}_* : \mathfrak{W}_{1*} \to \mathfrak{W}_{2*}$ such that $\omega_{\mathcal{T}_*\rho} \mathbf{1} \leq 1$ for every normal state $\omega_\rho \in S_*(\mathfrak{W}_1)$. If, in addition, $\omega_{\mathcal{T}_*\rho} \mathbf{1} = 1$, $\mathcal{T}$ is said to be a quantum channel.

In quantum theory, operations and channels characterize the transfer of information in open and closed systems, respectively. Here, the requirement of complete positivity of $\mathcal{T} : \mathfrak{W}_1 \to \mathfrak{W}_2$ (as opposed to mere positivity) ensures that $\mathcal{T}$ is extensible to a state-preserving map between any two systems that include $\mathfrak{W}_1$ and $\mathfrak{W}_2$ as subsystems. If $\mathfrak{W}_1$ is abelian, then positivity and complete positivity of $\mathcal{T}$ are equivalent notions. If $\mathfrak{W}_2 = B(H_2)$ for a Hilbert space H_2, Stinespring's theorem (28) states that $\mathcal{T}$ is completely positive if and only if there is a Hilbert space H_1, a representation $\varpi : \mathfrak{W}_1 \to B(H_1)$, and a bounded linear map $V : H_2 \to H_1$ such that $\mathcal{T}a = V^*\varpi(a)V$.

It follows from these considerations that the Koopman operator $U^t : \mathfrak{A} \to \mathfrak{A}$ is a quantum operation (since U^t is positive, the transfer operator preserves normal states, and $\mathfrak{A}$ is abelian), and so is $\mathcal{U}^t : \mathfrak{B} \to \mathfrak{B}$ (by Stinespring's theorem). In fact, U^t and $\mathcal{U}^t$ are both quantum channels. It is therefore natural to require that the discretization procedure leads to a quantum operation in both of the abelian and nonabelian cases. A second key requirement is that the discretization procedure is positivity preserving; that is, positive elements of the infinite-dimensional algebra are mapped into positive elements of the finite-dimensional algebra associated with the projected system. This requirement is particularly important when modeling physical systems, where failure to preserve signs of sign-definite quantities may result in loss of physical interpretability and lead to numerical instabilities (29). Our third requirement is that the finite-dimensional approximations converge in an appropriate sense to the original system as the dimension increases. One of the main perspectives put forward in this paper is that the construction of discretization schemes meeting these requirements is considerably facilitated by working in the nonabelian setting of $\mathfrak{B}$ rather than the abelian setting of $\mathfrak{A}$.

First, as an illustration of the fact that a "naive" projection will fail to meet our requirements, consider the Koopman operator $U^t : H \to H$. Fix an orthonormal basis $\{\phi_0, \phi_1, \ldots\}$ of H with $\phi_l \in \mathfrak{A}$, and let $\Pi_L : H \to H$ be the orthogonal projection that maps into the L-dimensional subspace $H_L := \mathrm{span}\{\phi_0, \ldots, \phi_{L-1}\}$. A common approach to Koopman and transfer operator approximation (30, 31) is to orthogonally project elements of H to elements of H_L, $f \mapsto f_L := \Pi_L f$ and similarly approximate U^t by the finite-rank operator $U_L^{(t)} := \Pi_L U^t \Pi_L$. The rank of $U_L^{(t)}$ is at most L, and it is represented in the $\{\phi_l\}$ basis by an $L \times L$ matrix $\boldsymbol{U}$ with elements $U_{ij} = \langle \phi_i, U^t \phi_j \rangle$. Note the inclusions $H_L \subset \mathfrak{A} \subset H$ and that H_L and $\mathfrak{A}$ are invariant subspaces of H under $U_L^{(t)}$. Moreover, $U_L^{(t)}$ maps $f \in H$ to $g = U_L^{(t)} f \in H_L$ such that $g = \sum_{i,j=0}^{L-1} \phi_i U_{ij} \hat{f}_j$, where $\hat{f}_j = \langle \phi_j, f \rangle$. Letting $\boldsymbol{f} = (\hat{f}_0, \ldots, \hat{f}_{L-1})^\top$ and $\boldsymbol{g} = (\hat{g}_0, \ldots, \hat{g}_{L-1})^\top$ with $\hat{g}_l = \langle \phi_l, g \rangle$ be the L-dimensional column vectors giving the representation of $\Pi_L f$ and g in the $\{\phi_l\}$ basis of H, respectively, we can express the action of $U_L^{(t)}$ on f as the matrix–vector product $\boldsymbol{g} = \boldsymbol{U f}$.

Unfortunately, such methods are not positivity preserving; that is, if f is a positive function in $\mathfrak{A}$, $\Pi_L f$ need not be positive. A classical example is a tophat function on the real line, which develops oscillations to negative values upon Fourier filtering (the Gibbs phenomenon). Even if f is a positive function in the finite-dimensional subspace H_L (so that $\Pi_L f = f$), the function $g = U_L^{(t)} f$ need not be positive. Thus, standard discretization approaches based on orthogonal projections fail to meet the requirements laid out above.

Next, we turn to positivity-preserving discretizations utilizing the abelian algebra $\mathfrak{A}$, as opposed to the Hilbert space H. Recalling that the projections in $\mathfrak{A}$ are the characteristic functions of measurable sets, let S be a measurable subset of X, and consider the multiplication operator $M_S : \mathfrak{A} \to \mathfrak{A}$ such that $M_S f = \chi_S f$. The map M_S is positive, and the projected Koopman operator, $M_S U^t M_S$ is a quantum operation. However, in order for M_S to be a discretization map, we must have that its range is a finite-dimensional algebra. This is equivalent to asking that the restriction of μ to S is supported on a finite number of atoms, i.e., measurable sets that have no measurable subsets of positive measure. This is a highly restrictive condition that fails to hold for broad classes of dynamical systems (e.g., volume-preserving flows on manifolds), so the abelian algebra $\mathfrak{A}$ does not provide an appropriate environment to perform discretizations meeting our requirements.

We now come to discretizations based on the operator algebra $\mathfrak{B}$. Working with $\mathfrak{B}$ allows us to use both Hilbert space techniques to construct finite-rank operators by orthogonal projection and algebraic techniques to ensure that these projections are positivity preserving. With H_L as above, consider the von Neumann algebra $\mathfrak{B}_L := B(H_L)$. This algebra has dimension L^2 and is isomorphic to the algebra $\mathbb{M}_L \equiv \mathbb{M}_L(\mathbb{C})$ of $L \times L$ complex matrices. In particular, each element $A \in \mathfrak{B}_L$ is represented by a matrix $\boldsymbol{A} \in \mathbb{M}_L$ with elements $A_{ij} = \langle \phi_i, A \phi_j \rangle$. Correspondingly, we refer to data assimilation based on $\mathfrak{B}_L$ as matrix mechanical; see Ⓜ in Fig. 1.

Next, note that $\mathfrak{B}_L$ can be canonically identified with the subalgebra of $\mathfrak{B}$ consisting of all operators A satisfying $\ker A \supseteq H_L^\perp$ and $\mathrm{ran} A \subseteq H_L$. Thus, we can view the projection $\boldsymbol{\Pi}_L : \mathfrak{B} \to \mathfrak{B}$ with $\boldsymbol{\Pi}_L A = \Pi_L A \Pi_L$ as an operator from $\mathfrak{B}$ to $\mathfrak{B}_L$.

By Stinespring's theorem, $\boldsymbol{\Pi}_L$ is completely positive. As a result, i) the projection $A \in \mathfrak{B} \mapsto \boldsymbol{\Pi}_L A \in \mathfrak{B}_L$ is positivity preserving, and thus, so is the projected representation $\pi_L : \mathfrak{A} \to \mathfrak{B}_L$ with $\pi_L = \boldsymbol{\Pi}_L \circ \pi$; and ii) the projected Koopman operator $\mathcal{U}_L^{(t)} : \mathfrak{B}_L \to \mathfrak{B}_L$ with $\mathcal{U}_L^{(t)} A = U_L^{(t)} A U_L^{(t)*}$ is a quantum operation. Moreover, since $\{\phi_l\}$ is an orthonormal basis of H, for any $f \in H$, we have $\lim_{L \to \infty} \Pi_L f = f$. This implies that for every $A \in \mathfrak{B}$, the operators $A_L = \boldsymbol{\Pi}_L A \in \mathfrak{B}_L$ converge strongly to A, i.e., $\lim_{L \to \infty} A_L g = Ag$, for all $g \in H$. In particular, $\pi_L f$ with $f \in \mathfrak{A}$ converges strongly to πf. Further details on these approximations can be found in *SI Appendix*, sections 2.D–2.H. Note that, in general, $\pi_L f$ is not a multiplication operator. That is, the act of embedding $\mathfrak{A}$ in the nonabelian algebra $\mathfrak{B}$ using $\pi : \mathfrak{A} \to \mathfrak{B}$ and then projecting to the finite-dimensional subalgebra $\mathfrak{B}_L$ using $\boldsymbol{\Pi}_L : \mathfrak{B} \to \mathfrak{B}_L$ is not equivalent to projecting $\mathfrak{A}$ into H_L using Π_L and then embedding H_L into $\mathfrak{B}$ using π.

Consider now a normal state $\omega_p \in S_*(\mathfrak{A})$ induced by a probability density $p \in \mathfrak{A}_*$, and let $\omega_\rho = \Gamma(\omega_p)$ be the associated normal state on $\mathfrak{B}$ obtained via Eq. 3. For L sufficiently large, $C_L(\rho) := \boldsymbol{\Pi}_L \rho$ is nonzero, and thus, $\rho_L = \boldsymbol{\Pi}_L \rho / C_L(\rho)$ is a density operator in $\mathfrak{B}_L$ inducing a state $\omega_{\rho_L} \in S_*(\mathfrak{B}_L)$, which extends to $S_*(\mathfrak{B})$. In Fig. 1, we denote the map $\omega_\rho \mapsto \omega_{\rho_L}$ as $\boldsymbol{\Pi}_L'$. By construction, the state ω_{ρ_L} satisfies $\omega_{\rho_L} A = \omega_\rho(\boldsymbol{\Pi}_L A)/C_L(\rho)$ for all $A \in \mathfrak{B}$. Setting, in particular, $A = \pi f$ with $f \in \mathfrak{A}$, it follows from Eq. 4 and the strong convergence of $\pi_L f$ to πf that

$$\lim_{L \to \infty} \omega_{\rho_L}(\pi_L f) = \omega_\rho(\pi f) = \omega_p f; \qquad [10]$$

SI Appendix, section 2.E. It should be kept in mind that, aside from special cases, ω_{ρ_L} is not the image of a state $\omega_{p_L} \in S_*(\mathfrak{A})$ under Γ for a probability density $p_L \in \mathfrak{A}_*$; that is, in general, ω_{ρ_L} is a "nonclassical" state. Note also that ω_{ρ_L} is a vector state (Eq. 3) induced by the unit vector $\xi_L = \Pi_L \sqrt{p} / \|\Pi_L \sqrt{p}\|_H$, which, as just mentioned, is generally not the square root of a probability density.

Let now $\mathcal{P}_L^{(t)} : S(\mathfrak{B}_L) \to S(\mathfrak{B}_L)$ with $\mathcal{P}_L^{(t)} \omega = \omega \circ \mathcal{U}_L^{(t)}$ be the projected transfer operator on $S(\mathfrak{B}_L)$. Unless H_L is a U^t-invariant subspace, $\mathcal{P}_L^{(t)} \circ \boldsymbol{\Pi}_L'$ is not equal to $\boldsymbol{\Pi}_L' \circ \mathcal{P}^t$; see the dashed arrow in the third column of the schematic in Fig. 1. Nevertheless, we have the asymptotic consistency $\lim_{L \to \infty} \big((\mathcal{P}_L^t \circ \boldsymbol{\Pi}_L') \omega_\rho\big) A_L = (\mathcal{P}^t \omega_\rho) A$, which holds for all $\omega_\rho \in S_*(\mathfrak{B})$ and $A \in \mathfrak{B}$; *SI Appendix*, section 2.I. Applying this result for $A = \pi f$ and $\omega_\rho = \Gamma(\omega_p)$, with $f \in \mathfrak{A}$ and $\omega_p \in S_*(\mathfrak{A})$, it follows that

$$\lim_{L \to \infty} \big((\mathcal{P}_L^{(t)} \circ \boldsymbol{\Pi}_L') \omega_\rho\big)(\pi_L f) = (\mathcal{P}^t \omega_\rho)(\pi f) = (P^t \omega_p) f. \qquad [11]$$

Eq. 11 implies that the matrix mechanical data assimilation scheme consistently recovers forecasts from data assimilation in the abelian algebra $\mathfrak{A}$ in the limit of infinite dimension L.

In *SI Appendix, section 2.F.2*, we describe how for any self-adjoint element $A \in \mathfrak{B}$, the spectral measures of $\boldsymbol{\Pi}_L A$ converge to the spectral measure of A. Since $\pi f \in \mathfrak{B}$ is self-adjoint whenever $f \in \mathfrak{A}$ is real, the spectral convergence of $\pi_L f$ to πf implies that the forecast distributions $\mathbb{P}_{\pi_L f, t, \tau}$ induced by $\omega_{\rho_{t, \tau, L}} = (\mathcal{P}_L^{(\tau)} \circ \boldsymbol{\Pi}_L') \omega_{\rho_{t, L}}$ consistently recover the forecast distributions $\mathbb{P}_{\pi f, t, \tau}$ and $\mathbb{P}_{f, t, \tau}$ from the infinite-dimensional quantum mechanical and abelian systems, respectively.

With a similar approach (*SI Appendix*, section 2.J), one can deduce that the analysis step is also consistently recovered: Defining the effect-valued map $\mathcal{F}_L : Y \to \mathcal{E}(\mathfrak{B}_L)$ with $\mathcal{F}_L = \mathbf{\Pi}_L \circ \mathcal{F}$, it follows from Eq. **9** and Eq. **10** that for every $f \in \mathfrak{A}$ and $\omega_p \in S_*(\mathfrak{A})$,

$$\lim_{L \to \infty} \omega_{\rho_L}|_{\mathcal{F}_L(y)}(\pi f) = \omega_\rho|_{\mathcal{F}(y)}(\pi f) = \omega_p|_{F(y)}f, \qquad [12]$$

where $\omega_\rho = \Gamma(\omega_p)$ and $\omega_{\rho_L} = \mathbf{\Pi}'_L \omega_{\rho_L}$, so the matrix mechanical analysis step is asymptotically consistent with the infinite-dimensional quantum mechanical and abelian analyses.

On the basis of Eqs. **11** and **12**, we conclude that as the dimension L increases, the matrix mechanical data assimilation scheme is consistent with the abelian formulation of sequential data assimilation. Moreover, the discretization leading to this scheme is positivity preserving, and the projected Koopman operator $\mathcal{U}_L^{(t)}$ is a quantum operation. Thus, matrix mechanical data assimilation provides a nonabelian, finite-dimensional framework that simultaneously meets all of the requirements listed in the beginning of this subsection.

Data-Driven Approximation. The matrix mechanical data assimilation scheme described above admits a consistent data-driven approximation using kernel methods for machine learning (17, 20, 22). The data-driven scheme employs three, possibly related, types of training data, all acquired along a dynamical trajectory $X_N = \{x_0, x_1, \ldots, x_{N-1}\} \subset X$ with $x_n = \Phi^{n \Delta t}(x_0)$, where $\Delta t > 0$ is a sampling interval: i) samples $y_n = h(x_n)$ from the observation map $h : X \to Y$; ii) samples $f_n = f(x_n)$ from the forecast observable $f \in \mathfrak{A}$; iii) samples $z_n = z(x_n)$ from a map $z : X \to Z$, used as proxies of the dynamical states x_n. If the x_n are known, we set $Z = X$ and $z = \mathrm{Id}$. Otherwise, we set $Z = Y^{2Q+1}$ for a parameter $Q \in \mathbb{N}$ and define z as the delay-coordinate map $z(x) = (h(\Phi^{-Q \Delta t}(x)), h(\Phi^{(-Q+1) \Delta t}(x)), \ldots, \Phi^{Q \Delta t}(x))$, giving $z_n = (y_{n-Q}, y_{n-Q+1}, \ldots, y_Q)$. By delay-embedding theory (32), for sufficiently large Q and typical observation maps h and sampling intervals Δt, z is an injective map.

The dynamical trajectory x_n has an associated sampling measure $\mu_N := \sum_{n=0}^{N-1} \delta_{x_n}/N$ and a finite-dimensional Hilbert space $\hat{H}_N := L^2(X, \mu_N)$. By ergodicity, as N increases, the measures μ_N converge to the invariant measure μ in weak-* sense, so we can interpret $\hat{H}_N$ as a data-driven analog of the infinite-dimensional Hilbert space H (*SI Appendix*, section 1). Given the training data $z_0, z_1, \ldots, z_{N-1}$, and without requiring explicit knowledge of the underlying states x_n, we use kernel integral operators to build an orthonormal basis $\{\phi_{0,N}, \ldots, \phi_{L-1,N}\}$ of an L-dimensional subspace $H_{L,N} \subseteq \hat{H}_N$ that plays the role of a data-driven counterpart of H_L. More specifically, the basis elements $\phi_{l,N}$ are eigenvectors of a kernel integral operator $K_N : \hat{H}_N \to \hat{H}_N$ induced by a kernel function $\kappa : Z \times Z \to \mathbb{R}$. The operator K_N is represented by an $N \times N$ kernel matrix $\boldsymbol{K}_N$ constructed from the training data z_n; *SI Appendix*, section 2.A. We let $\mathfrak{B}_{L,N} = B(H_{L,N})$ be the L^2-dimensional algebra of linear maps on $H_{L,N}$, which, as in the case of $\mathfrak{B}_L$, is isomorphic to the matrix algebra $\mathbb{M}_L$.

Every operator employed in the matrix-mechanical scheme described in the previous section has a data-driven counterpart, represented as an $L \times L$ matrix with respect to the $\{\phi_{l,N}\}$ basis. Specifically, the projected Koopman operator $U_L^{(t)}$ at time $t = q \Delta t$, $q \in \mathbb{Z}$, is replaced by an operator $U_{L,N}^{(q)} \in \mathfrak{B}_{L,N}$

induced by the shift map on the trajectory x_n (30), with a corresponding quantum operation $\mathcal{U}_{L,N}^{(q)} : \mathfrak{B}_{L,N} \to \mathfrak{B}_{L,N}$. Moreover, the projected multiplication operator $\pi_L f$ is replaced by $\pi_{L,N} \hat{f}_N \in \mathfrak{B}_{L,N}$, and the effect-valued map $\mathcal{F}_L$ by a map $\mathcal{F}_{L,N} : Y \to \mathcal{E}(\mathfrak{B}_{L,N})$. Here, $\hat{f}_N$ is the restriction of f on $\hat{X}_N$. Further details are provided in *SI Appendix*, sections 2.D–2.J.

The data-driven scheme is positivity preserving and constitutes a quantum operation analogously to the matrix mechanical scheme. Moreover, by results on spectral approximation of kernel integral operators (33) and ergodicity of the dynamics, the kernel matrices $\boldsymbol{K}_N$ exhibit spectral convergence in the large-data limit, $N \to \infty$, to a kernel integral operator $K : H \to H$ in a suitable sense (*SI Appendix*, Theorem 1). Correspondingly, all matrix representations of operators, and thus all predictions made by the data-driven scheme, converge to the predictions of the matrix mechanical scheme Ⓜ in Fig. 1. Overall, we obtain a data-driven, positivity-preserving, and asymptotically consistent data assimilation scheme. The data requirements and computational complexity of this scheme are comparable to standard kernel methods for supervised machine learning (*SI Appendix*, section 2.K).

Lorenz 96 Multiscale System

As our first numerical example, we apply QMDA to assimilate and predict the slow variables of the Lorenz 96 (L96) multiscale system (34). This system was introduced by Lorenz in 1996 as a low-order model of atmospheric circulation at a constant-latitude circle. The dynamical degrees of freedom include K slow variables $\mathsf{x}_1, \ldots, \mathsf{x}_K$, representing the zonal (west to east) component of the large-scale atmospheric velocity field at K zonally equispaced locations. Each slow variable x_k is coupled to J fast variables $\mathsf{y}_{1,k}, \ldots, \mathsf{y}_{J,k}$, representing small-scale processes such as atmospheric convection. The dynamical state space is thus $X = \mathbb{R}^{J(K+1)}$ with $x = (\mathsf{x}_k, \mathsf{y}_{j,k})_{j,k=1,1}^{J,K} \in X$.

The governing equations are

$$\dot{\mathsf{x}}_k = -\mathsf{x}_{k-1}(\mathsf{x}_{k-2} - \mathsf{x}_{k+1}) - \mathsf{x}_k + \mathsf{F} + \frac{\mathsf{h}_\mathsf{x}}{J} \sum_{j=1}^{J} \mathsf{y}_{j,k},$$

$$\dot{\mathsf{y}}_{j,k} = \frac{1}{\varepsilon}\left(-\mathsf{y}_{j+1,k}(\mathsf{y}_{j+2,k} - \mathsf{y}_{j-1,k}) - \mathsf{y}_{j,k} + \mathsf{h}_\mathsf{y}\mathsf{x}_k\right), \qquad [13]$$

$$\mathsf{x}_{k+K} = \mathsf{x}_k, \quad \mathsf{y}_{j,k+K} = \mathsf{y}_{j,k}, \quad \mathsf{y}_{j+J,k} = \mathsf{y}_{j,k+1},$$

where the parameter F represents large-scale forcing (e.g., solar heating), h_x and h_y control the coupling between the slow and fast variables, and ε is a parameter that controls the timescale separation between the fast and slow variables. The governing equations for x_k feature large-scale forcing, F, a quadratic non-linearity, $-\mathsf{x}_{k-1}(\mathsf{x}_{k-2} - \mathsf{x}_{k+1})$, representing advection, a linear damping term, $-\mathsf{x}_k$, representing surface drag, and a flux term, $\mathsf{h}_\mathsf{x}\sum_{j=1}^{J}\mathsf{y}_{j,k}/J$, representing forcing from the fast variables. The terms in the $\mathsf{y}_{j,k}$ equations have similar physical interpretations. In general, the dynamics becomes more turbulent/chaotic as F increases.

Here, we focus on the chaotic dynamical regime studied in refs. 35 and 36 with $K = 9$, $J = 8$, $\varepsilon = 1/128$, $\mathsf{F} = 10$, $\mathsf{h}_\mathsf{x} = -0.8$, and $\mathsf{h}_\mathsf{y} = 1$. In this regime, ε is sufficiently small so that the dynamics of the $(\mathsf{x}_1, \ldots, \mathsf{x}_K)$ variables is approximately Markovian. We consider that the observation map $h : X \to Y$ projects the state vector $x \in X$ to the slow variables, i.e., $Y = \mathbb{R}^K$

and $h(x) = y := (x_1, \ldots, x_K)$. Our forecast observable $f \in \mathfrak{A}$ is the first slow variable, $f(x) = x_1$.

Training. We employ a training dataset consisting of $N = 40,000$ samples $y_0, \ldots, y_{N-1} \in Y$ and $f_0, \ldots, f_{N-1} \in \mathbb{R}$ with $y_n = h(x_n)$, $f_n = f(x_n)$, and $x_n = \Phi^{n\Delta t}(x_0)$, taken at a sampling interval $\Delta t = 0.05$. To assess forecast skill, we use $\hat{N} = 7,000$ samples $\hat{y}_0, \ldots, \hat{y}_{\hat{N}-1}$ with $\hat{y}_n = h(\hat{x}_n)$ and $\hat{x}_n = \Phi^{n\Delta t}(\hat{x}_0)$, taken on an independent dynamical trajectory from the training data. The data $z_0, z_1, \ldots, z_{N-1} \in Z$ for computation of the data-driven basis $\{\phi_{l,N}\}$ of $H_{L,N}$ consist of snapshots of the slow variables, $z_n = y_n$. That is, we have $Z = \mathbb{R}^{(2Q+1)K} = Y$ with $Q = 0$ delays and $z = \text{Id}$. This choice is motivated by the fact that the evolution of y_n is approximately Markovian for $\varepsilon \ll 1$, and the forecast observable $f(x) = x_1$ depends on $x \in X$ only through $y = h(x)$. Following ref. 21, we compute the $\phi_{l,N}$ using a variable-bandwidth Gaussian kernel (37) with a bistochastic normalization (38). Further details on this kernel and the L96 data are provided in *SI Appendix, sections 2.B and 5.A*, respectively.

Using the basis vectors, we compute $L \times L$ matrix representations of the projected Koopman operators $U_{L,N}^{(\tau_j)}$ for lead times $\tau_j = j\Delta t$ with $j \in \{0, 1, \ldots, J_f\}$, $J_f = 150$ (*SI Appendix, Algorithm S9*). Moreover, using the $\phi_{l,N}$ and the training samples f_n, we compute the $L \times L$ matrix representation $\boldsymbol{A}_{L,N}$ of the operator $A_{L,N} := \pi_{L,N} f$ associated with the forecast observable. To evaluate forecast distributions for f, we compute the PVM $E_{A_{L,N}}$ of $A_{L,N}$, which amounts to computing an eigendecomposition of $\boldsymbol{A}_{L,N}$ (*SI Appendix, Algorithm S7*). To report forecast probabilities, we evaluate $E_{A_{L,N}}$ on a collection of bins $S_1, \ldots, S_M \subset \mathbb{R}$ of equal probability mass in the equilibrium distribution of f. As our observation kernel $\psi : Y \times Y \to [0, 1]$, we use a variable-bandwidth bump function. The corresponding effect-valued map $\mathcal{F}_{L,N} : Y \to \mathcal{E}(\mathfrak{B}_{L,N})$ is represented by a matrix-valued function; further details are provided in *SI Appendix, section 2.J*. In Figs. 2 and 3, we show results for Hilbert space dimension $L = 2,000$, though forecast skill does not change appreciably for values of L in the range 500 to 2,000 (*SI Appendix, Fig. S1*).

Data Assimilation. We perform data assimilation experiments initialized with the pure state $\omega_0 \equiv \omega_{\rho_0} \in S(\mathfrak{B}_{L,N})$ induced by the density operator $\rho_0 = \langle \mathbf{1}_X, \cdot \rangle_{\hat{H}_N} \mathbf{1}_X \in \mathfrak{B}_{L,N}$. We interpret this state as an uninformative equilibrium state, in the sense that i) $\omega_0 A_{L,N} = \text{tr}(\rho_0 A_{L,N}) = \bar{f}_N$, where $\bar{f}_N = \sum_{n=0}^{N-1} f_n / N$ is the empirical mean of f, and ii) ω_{ρ_0} is invariant under the action of the transfer operator, i.e., $\mathcal{P}_{L,N}^{(t)} \omega_0 := \omega_0 \circ \mathcal{U}_{L,N}^{(t)} = \omega_0$.

Starting from ω_0, QMDA produces a sequence of states $\omega_0, \omega_1, \ldots, \omega_{\hat{N}-1}$ by repeated application of the forecast–analysis steps, as depicted schematically in Fig. 1 and in pseudocode form in *SI Appendix, Algorithm S1*. Specifically, for $n \in \{1, \ldots, \hat{N}-1\}$, we compute ω_n by first using the transfer operator to compute the state $\omega_{n-1,1} := \mathcal{P}_{L,N}^{(\Delta t)} \omega_{n-1}$ (which is analogous to the prior in classical data assimilation) and then applying the effect map to observation $\hat{y}_n$ to yield $\omega_n = \omega_{n-1,1}|_{\mathcal{F}_{L,N}(\hat{y}_n)}$ (which is analogous to the classical posterior). For each $n \in \{0, \ldots, \hat{N}-1\}$, we also compute forecast states $\omega_{n,j} = \mathcal{P}_{L,N}^{(\tau_j)} \omega_n$ and associated forecast distributions $\mathbb{P}_{n,j}$ for the observable f. We evaluate $\mathbb{P}_{n,j}$ on the bins S_m and normalize the result by the corresponding bin size, $s_m := \text{length}(S_m)$ to produce discrete
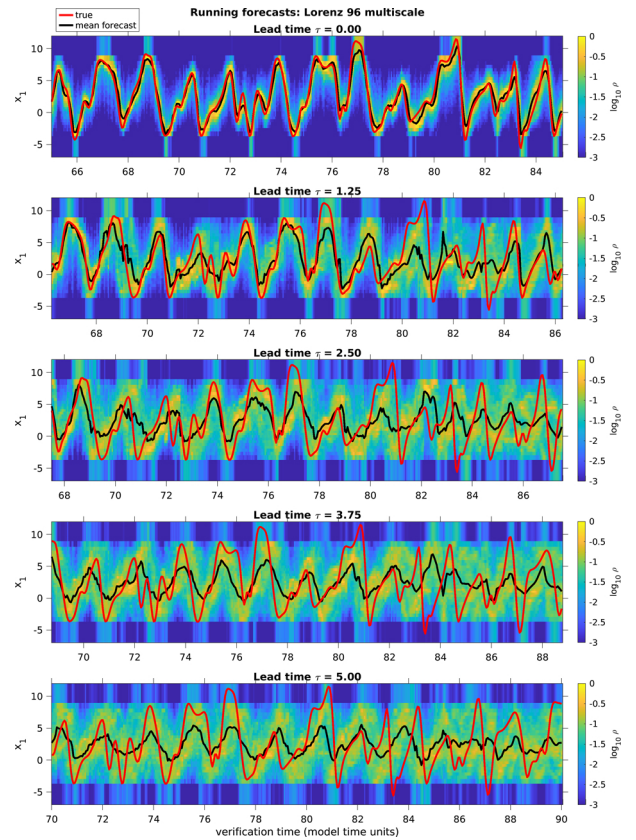

Fig. 2. Running QMDA forecasts of the x_1 variable of the L96 multiscale system in a chaotic regime. The panels show the true x_1 evolution (black lines), the logarithm of the discrete forecast probability density $\varrho_{n,j}$ (colors), and the corresponding forecast mean (red lines) as a function of verification time for lead times in the range 0 to 5 model time units (*Top* to *Bottom*). The assimilated observable is the K-dimensional vector $(x_1, \ldots, x_K)$ of the L96 slow variables.

probability densities $\varrho_{n,j} = (\varrho_{n,j,0}, \ldots, \varrho_{n,j,M-1})$ with $\varrho_{n,j,m} := \mathbb{P}_{n,n}(S_m)/s_m$. We also compute the forecast mean and standard deviation, $\bar{f}_{n,j} = \omega_{\rho_{n,j}} A_{L,N}$ and $\sigma_{n,j} = (\omega_{\rho_{n,j}}(A_{L,N})^2 - \bar{f}_{n,j}^2)^{1/2}$, respectively. We assess forecast skill through the normalized mean square error (NRMSE) and anomaly correlation (AC) scores, computed for each lead time τ_j by averaging over the $\hat{N}$ samples in the verification dataset (*SI Appendix, section S4*).

Fig. 2 shows the forecast probability densities $\varrho_{n,j}$ (colors), forecast means $\bar{f}_{n,j}$ (black lines), and true signal $\hat{f}_{n+j}$ (red lines), plotted as a function of verification time t_{n+j} over intervals spanning 20 time units for representative lead times τ_j in the range 0 to 5 time units. The corresponding NRMSE and AC scores are displayed in Fig. 3. Given the turbulent nature of the
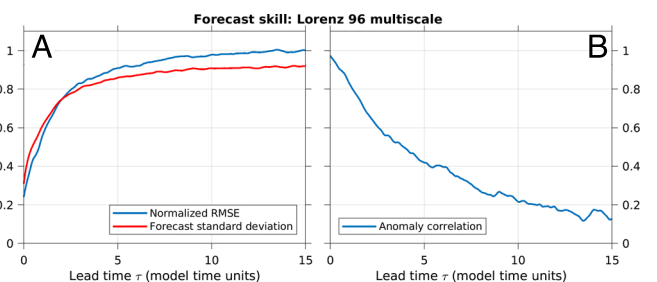

Fig. 3. NRMSE (*A*) and AC score (*B*) of the L96 forecasts from Fig. 2.

dynamics, we intuitively expect the forecast densities $\varrho_{n,j}$ to start from being highly concentrated around the true signal for small τ_j and progressively broaden as τ_j increases (i.e., going down the panels of Fig. 2), indicating that the forecast uncertainty increases. Correspondingly, we expect $\bar{f}_{n,j}$ to accurately track the true signal for small τ_j and progressively relax toward the equilibrium mean $\int_X f \, d\mu$.

The results in Figs. 2 and 3 are broadly consistent with this behavior: The forecast starts at $\tau_j = 0$ from a highly concentrated density around the true signal (note that Fig. 2 shows logarithms of $\varrho_{n,j}$), which is manifested by low NRSME and large AC values in Fig. 3 of approximately 0.24 and 0.98, respectively. As τ_j increases, the forecast distribution broadens, and the NRMSE (AC) scores exhibit a near-monotonic increase (decrease). In Fig. 3A, the estimated error based on the forecast variance $\sigma_{n,j}$ is seen to track well the NRMSE score, which indicates that the forecast distribution $\varrho_{n,j}$ well represents the true forecast uncertainty. It should be noted that errors are present even at time $\tau_j = 0$, particularly for periods of time where the true signal takes extreme positive or negative values. Such reconstruction errors are expected for a fully data-driven driven method applied to a system with a high-dimensional attractor. Overall, the skill scores in Fig. 3 are comparable with the results obtained in ref. 36 using the kernel analog forecasting (KAF) technique (39).

El Niño Southern Oscillation

The El Niño Southern Oscillation (ENSO) (40) is the dominant mode of interannual (3- to 5-y) variability of the Earth's climate system. Its primary manifestation is an oscillation between positive sea surface temperature (SST) anomalies over the eastern tropical Pacific Ocean, known as El Niño events, and episodes of negative anomalies known as La Niñas (41). Through atmospheric teleconnections, ENSO drives seasonal weather patterns throughout the globe, affecting the occurrence of extremes such as floods and droughts, among other natural and societal impacts (42). Here, we demonstrate that QMDA successfully predicts ENSO within a comprehensive climate model by assimilating high-dimensional SST data.

Our experimental setup follows closely ref. 43, who performed data-driven ENSO forecasts using KAF. As training and test data, we use a control integration of the Community Climate System Model Version 4 (CCSM4) (44), conducted with fixed preindustrial greenhouse gas forcings. The simulation spans 1,300 y, sampled at an interval $\Delta t = 1$ month. Abstractly, the dynamical state space X consists of all degrees of freedom of CCSM4, which is of order 10^7 and includes variables such as density, velocity, and temperature for the atmosphere, ocean, and sea ice, sampled on discretization meshes over the globe. Since this simulation has no climate change, there is an implicit invariant measure μ sampled by the data, and we can formally define the algebras $\mathfrak{A}$ and $\mathfrak{B}$ associated with the invariant measure as described above.

In our experiments, the observation map $h : X \to Y$ returns monthly averaged SST fields on an Indo-Pacific domain; that is, we have $Y = \mathbb{R}^d$, where d is the number of surface ocean gridpoints within the domain. We have $d = 44{,}414$, so these experiments test the ability of QMDA to assimilate high-dimensional data. However, note that h is a highly noninvertible map since Indo-Pacific SST comprises only a small subset of CCSM4's dynamical degrees of freedom. As our forecast observable $f \in \mathfrak{A}$ we choose the Niño 3.4 index—a commonly

used index for ENSO monitoring defined as the average SST anomaly over a domain in the tropical Pacific Ocean. Large positive (negative) values of Niño 3.4 represent El Niño (La Niña) conditions, whereas values near zero represent neutral conditions. Additional information on the CCSM4 data is included in *SI Appendix*, section 5.B.

Following ref. 43, we use the SST and Niño 3.4 samples from the first 1,100 y of the simulation as training data and the corresponding samples for the last 200 y as test data. Thus, with the notation of the previously described L96 experiments, our training data are $y_n = h(x_n)$ (Indo-Pacific SST) and $f_n = f(x_n)$ (Niño 3.4) for $n \in \{0, \dots, N - 1\}$ and $N = 1{,}100 \times 12 = 13{,}200$, and our test data are $\hat{y}_n = h(x_{n+N})$ and $\hat{f}_n = f(x_{n+N})$ for $n \in \{0, \dots, \hat{N} - 1\}$ and $\hat{N} = 200 \times 12 = 2{,}400$. Here, $x_n = \Phi^{n\Delta t}(x_0) \in X$ is the (unknown) dynamical trajectory of the CCSM4 model underlying our training and test data. Using the SST samples y_n, we build the training data z_n using delay-coordinate maps with parameter $Q = 5$; i.e., the data z_n used for building the basis of $H_{L,N}$ are SST "videos" that span a total of $2Q + 1 = 11$ months and have dimension $11d \simeq 4.9 \times 10^5$. We compute the basis $\{\phi_{l,N}\}$ using a kernel κ that depends on the pairwise Euclidean distances between points in Z as well as Niño 3.4 trajectories evaluated on these points (*SI Appendix* Eq. S5). This approach improves the ability of the basis vectors to capture covariability between Indo-Pacific SST fields and the Niño 3.4 index, leading to a modest improvement of short-term forecast skill and a more significant improvement of uncertainty quantification over kernels that depend only on SST. Aside from the different kernel κ, the procedure for initializing and running QMDA is identical to the L96 experiments.

Fig. 4 shows the forecast probability density ($\varrho_{n,j}$; colors), forecast mean ($\bar{f}_{n,j}$; black lines), and true signal ($\hat{f}_{n+j}$; red lines) for the Niño 3.4 index as a function of verification time t_{n+j} over 20-y portions of the test dataset for lead times τ_j in the range 0 to 12 mo, obtained for Hilbert space dimension $L = 1{,}000$. The corresponding NRMSE and AC scores are displayed in Fig. 5. The skill scores do not vary significantly for values of L in the interval 500 to 2,000 (*SI Appendix*, Fig. S2).

Qualitatively, the forecast density $\varrho_{n,j}$ displays a similar behavior as in the L96 experiments; that is, it is concentrated around the true signal on short lead times ($\tau_j \lesssim 3$ months) and gradually broadens as forecast uncertainty grows with increasing lead time τ_j due to chaotic climate dynamics. In Fig. 5A, the estimated forecast error based on the forecast variance $\sigma_{n,j}$ agrees reasonably well with the actual NRMSE evolution. Adopting AC = 0.6 as a commonly used threshold for ENSO predictability, we see from the AC results in Fig. 5B that QMDA produces useful forecasts out to $\tau_j \simeq 12$ months. The performance of QMDA in terms of the NRMSE and AC metrics is comparable to that found for KAF in ref. 43, but QMDA has the advantage of producing full forecast probability distributions instead of point estimates. Compared to KAF, QMDA also has the advantage of being positivity preserving. While this property may not be critical for sign-indefinite ENSO indices, there are many climatic variables where sign preservation is particularly important.

Quantum Circuit Implementation

As a demonstration of the potential of QMDA for implementation on quantum computing platforms, we present forecasting results for the L96 multiscale system obtained from quantum circuit simulations performed using the Qiskit Aer Python library
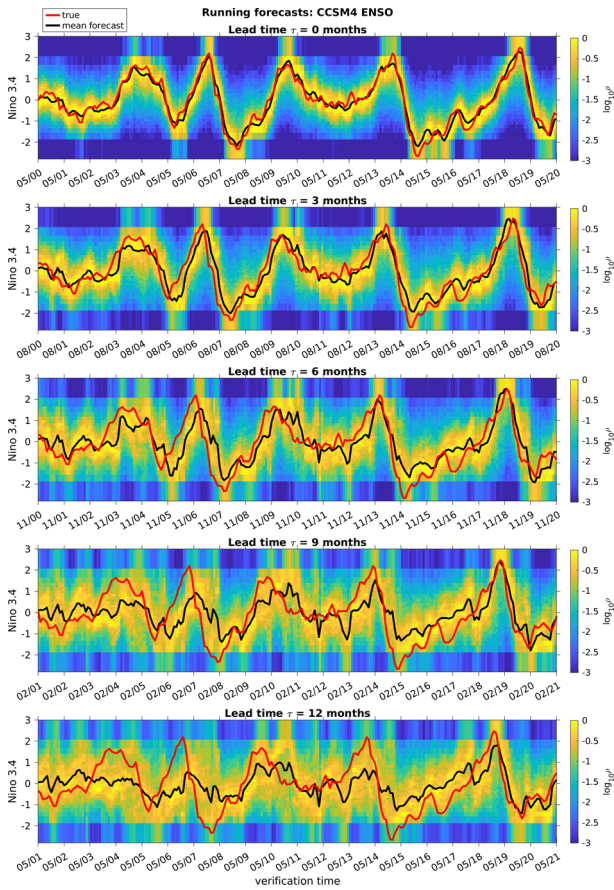
Fig. 4. As in Fig. 2, but for forecasts of the Niño 3.4 index in CCSM4. The assimilated observable is the vector of Indo-Pacific SST gridpoint values. The forecast lead times are in 3-mo increments in the range 0 to 12 mo. The verification times are shown in mm/yy date format relative to an arbitrary year in the verification interval.

(45). Our quantum circuit architecture consists of initialization, Koopman evolution, eigenbasis rotation, and measurement stages, depicted in Fig. 6 for a 4-qubit setup. The initialization and Koopman stages implement the QMDA analysis and forecast steps, respectively, via unitary operations acting on the qubits. The eigenbasis rotation implements a unitary induced by the eigenvectors of the quantum mechanical forecast observable $A_{L,N}$ so that measurement at the output of the circuit samples the desired probability distribution $\mathbb{P}_{n,j}$ at the given initialization time t_n and forecast lead time τ_j.

In more detail, associated with a quantum computational system of $\mathfrak{n}$ qubits is a tensor product Hilbert space $\mathbb{B}_{\mathfrak{n}} = \mathbb{B}^{\otimes \mathfrak{n}}$ of dimension $2^{\mathfrak{n}}$, where $\mathbb{B} = \mathrm{span}\{|0\rangle, |1\rangle\}$ is the two-dimensional Hilbert space generated by $|0\rangle$ ("up") and $|1\rangle$
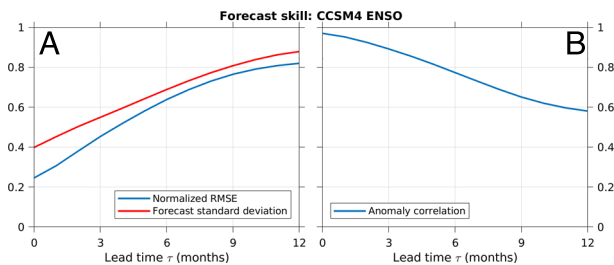

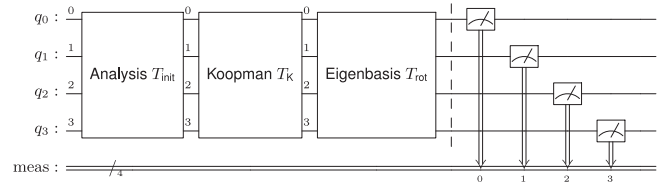**Fig. 5.** NRMSE (A) and AC score (B) of the Niño 3.4 forecasts from Fig. 4.

Fig. 6. Four-qubit circuit implementation of an analysis–forecast QMDA cycle.

("down") vectors (46). The standard basis of $\mathbb{B}_{\mathfrak{n}}$, known as quantum computational basis, has the tensor product form $\{|\boldsymbol{b}\rangle = |b_1\rangle \otimes \cdots \otimes |b_{\mathfrak{n}}\rangle\}_{\boldsymbol{b} \in \{0,1\}^{\mathfrak{n}}}$, indexed by binary strings $\boldsymbol{b} = (b_1, \ldots, b_{\mathfrak{n}})$ of length $\mathfrak{n}$. A (noise-free) quantum computer is represented as a quantum channel $\mathcal{T} : \mathfrak{M}_{\mathfrak{n}} \to \mathfrak{M}_{\mathfrak{n}}$ on the $2^{2\mathfrak{n}}$-dimensional von Neumann algebra $\mathfrak{M}_{\mathfrak{n}} := B(\mathbb{B}_{\mathfrak{n}})$, which is isomorphic to the algebra $\mathbb{M}_{2^{\mathfrak{n}}}$ of $2^{\mathfrak{n}} \times 2^{\mathfrak{n}}$ matrices. Typically, the channel $\mathcal{T}$ has the form $\mathcal{T}A = T^*AT$, where $T : \mathbb{B}_{\mathfrak{n}} \to \mathbb{B}_{\mathfrak{n}}$ is a unitary map. In the circuit shown in Fig. 6, we express T as the composition $T = T_{\mathrm{rot}} \circ T_{\mathrm{K}}(j) \circ T_{\mathrm{init}}(\xi, y)$, where $T_{\mathrm{init}}(\xi, y)$ represents the initialization (analysis) step given a prior state vector $\xi \in H_{L,N}$, and an observation $y \in Y$, $T_{\mathrm{K}}(j)$ represents Koopman evolution over $j \in \mathbb{N}$ forecast timesteps, and T_{rot} is the eigenbasis rotation stage of the circuit.

In an actual quantum computational environment, T_{init}, T_{K}, and T_{rot} would be implemented by combining quantum logic gates through operations such as compositions and tensor products, the goal being to implement T by a circuit of low depth (longest path from input to output). For instance, circuits whose depth scales as a polynomial in $\mathfrak{n}$ allow simulation of quantum states and observables on the exponentially large-dimensional Hilbert space $\mathbb{B}_{\mathfrak{n}}$ in a polynomial running time. Here, we do not address the important questions of how to implement T_{init}, T_K, and T_{rot} efficiently and robustly on an actual quantum computer, so the results presented in this section should be viewed as a proof of concept.

Suppose that the Hilbert space dimension of the matrix mechanical data assimilation system Ⓜ is $L = 2^{\mathfrak{n}}$ for some $\mathfrak{n} \in \mathbb{N}$. Given the data-driven basis $\{\phi_{l,N}\}$ of $H_{L,N}$ indexed by integers $l \in \{0, \ldots, L-1\}$, we define a unitary $W_L : H_L \to \mathbb{B}_{\mathfrak{n}}$ such that $W_L\phi_{l,N} = |\boldsymbol{b}\rangle$, where $\boldsymbol{b} = (b_1, \ldots, b_{\mathfrak{n}})$ is the binary representation of l, i.e., $l = \sum_{i=1}^{\mathfrak{n}} b_i 2^{\mathfrak{n}-i}$. Using W_L, we can encode any $A \in \mathfrak{B}_{L,N}$ into an operator $B = \mathcal{W}_L A := W_L A W_L^* \in \mathfrak{M}_{\mathfrak{n}}$. In particular, if $\omega_\rho \in S(\mathfrak{B}_{L,N})$ is a state induced by a density operator $\rho \in \mathfrak{B}_{L,N}$, then the transformed state $\omega_\sigma \in S(\mathfrak{M}_{\mathfrak{n}})$ with $\sigma = \mathcal{W}_L\rho$ satisfies $\omega_\rho(A) = \omega_\sigma(B)$, so predictions from the matrix mechanical and quantum computational systems are equivalent. If ω_ρ is a vector state induced by unit vector $\xi \in H_{L,N}$, then ω_σ is a vector state induced by $\zeta = W_L\xi \in \mathbb{B}_{\mathfrak{n}}$.

Given a prior state $\omega_{n-1,1} \in S(\mathfrak{B}_{L,N})$ at time t_n with corresponding state vector $\xi_{n-1,1} \in H_{L,N}$, an observation $\hat{y}_n \in Y$, and a forecast lead time τ_j, the circuit in Fig. 6 operates by acting on the computational basis vector $|\boldsymbol{0}\rangle$ (which is the "default" state vector at the start of a quantum computation) by the transformations $T_{\mathrm{init}}(\xi_{n-1,1}, \hat{y}_n)$, $T_{\mathrm{K}}(j)$, and T_{rot}, leading to the state vector $\zeta_{n,j} = T_{\mathrm{rot}} \circ T_{\mathrm{K}}(j) \circ T_{\mathrm{init}}(\xi_{n-1,1}, \hat{y}_n)|\boldsymbol{0}\rangle$. At the end of the computation, a measurement in the quantum computational basis yields a random binary string $\boldsymbol{b}$ with probability $\mathbb{P}_{n,j}(\{\boldsymbol{b}\}) = |\langle \zeta_{n,j}, \boldsymbol{b}\rangle_{\mathfrak{n}}^2|$. It can be shown that $a_l \in \mathbb{R}$, where $l \in \{0, \ldots, L-1\}$ is the integer with binary representation $\boldsymbol{b}$ and a_l is the l-th eigenvalue in the spectrum of $A_{L,N}$ (ordered in increasing order), is a sample from the distribution $\mathbb{P}_{n,j}(\{a_l\})$
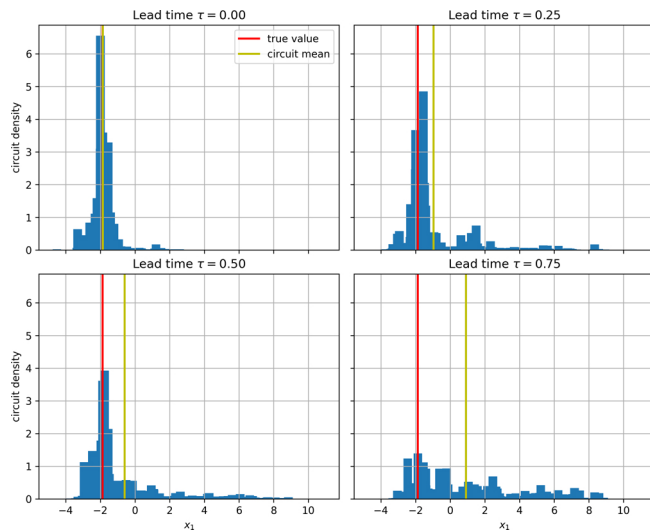
Fig. 7. Simulated quantum circuit prediction of the x_1 variable of the L96 multiscale system using $\mathfrak{n} = 10$ qubits. The histograms show the empirical probability density $\tilde{\varrho}_{n,j}$ computed from $M = 10^6$ shots for lead times in the range 0 to 3.75 time units. The vertical red lines show the true value of x_1 at the verification time, which is held fixed in all experiments. The vertical yellow lines show the empirical mean forecasts.

induced from the spectral measure $E_{A_{L,N}}$ and the quantum state $\omega_{n,j}$. Repeating this procedure over M identically prepared circuits leads to an ensemble of measurements (or "shots") $\{a_{l_1}, \ldots, a_{l_M}\}$, which provides a Monte Carlo approximation of the theoretical forecast distribution $\mathbb{P}_{n,j}$. Further details are provided in *SI Appendix*, section 3.

Fig. 7 shows representative forecast distributions of the x_1 variable of the L96 multiscale system obtained via this approach using $\mathfrak{n} = 10$ qubits (i.e., $L = 2^{10} = 1024$). In these experiments, the verification time is held fixed, so there is a fixed true value $x_1 \approx -1.86$ and the lead time $\tau_j = j \Delta t$ varies from 0 to 0.75 in increments of 0.25. The panels show histograms of the normalized counts $\tilde{\varrho}_{n,j}(l) = M_{njl}/(s_l M)$ centered on $x_1 = a_l$, where M_{njl} is the number of occurrences of eigenvalue a_l of the multiplication operator $A_{L,N}$ in the experiment with lead time τ_j, $M = 10^6$ is the number of shots, and $s_l = (a_{l+1} - a_{l-1})/2$ is an effective bin size. Also shown are the empirical mean $\tilde{f}_{n,j} = \sum_{l=0}^{L-1} a_l M_{njl}/M$ that approximates the forecast expectation $\bar{f}_{n,j}$ and the true value of x_1 (yellow and red lines, respectively).

The time evolution of the histograms illustrates the increase of forecast uncertainty due to chaotic dynamics in conjunction with finite-rank operator approximation. At $\tau_j = 0$, the empirical density is strongly concentrated around the truth, and the empirical mean $\tilde{f}_{n,j} \approx -1.99$ has an approximately 5.9% error. As τ_j increases, the probability density spreads predominantly to larger values of x_1, causing the mean forecast to increasingly deviate from the truth. Intriguingly, the peak of the histograms remains collocated with the truth, which suggests that there may be opportunities to increase skill using an estimator based on the mode of the distribution rather than the mean. Overall, the quantum circuit simulation results are qualitatively consistent with the L96 results from deterministic computation presented earlier, which demonstrates the suitability of QMDA for implementation on quantum computers.

Concluding Remarks

We have developed theory and methods for sequential data assimilation of partially observed dynamical systems using techniques from operator algebra, quantum information, and ergodic theory. At the core of this framework, called quantum mechanical data assimilation (QMDA), is the nonabelian algebraic structure of spaces of operators. One of the main advantages that this structure provides is that it naturally enables finite-dimensional discretization schemes that preserve the sign of sign-definite observables in ways that are not possible with classical projection-based approaches.

We build these schemes starting from a generalization of Bayesian data assimilation based on a dynamically consistent embedding into an infinite-dimensional operator algebra acting on the L^2 space associated with an invariant measure of the system. Under this embedding, forecasting is represented by a quantum operation induced by the Koopman operator of the dynamical system, and Bayesian analysis is represented by quantum effects. In addition to providing a useful starting point for discretizing data assimilation, this construction draws connections between statistical inference methods for classical dynamical systems with quantum information and quantum probability, which should be of independent interest.

QMDA leverages properties of operator algebras to project the infinite-dimensional framework into the level of a matrix algebra in a manner that positive operators are represented by positive matrices, and the finite-dimensional system is a quantum operation. QMDA also has a data-driven formulation based on kernel methods for machine learning with consistent asymptotic behavior as the amount of training data increases. We have demonstrated the efficacy of QMDA with forecasting experiments of the slow variables of the Lorenz 96 multiscale system in a chaotic regime and the El Niño Southern Oscillation in a climate model. QMDA was shown to perform well in terms of point forecasts from quantum mechanical expectations, while also providing uncertainty quantification by representing entire forecast distributions via quantum states.

This work motivates further application and development of algebraic approaches and quantum information to building models and performing inference of complex dynamical systems. In particular, as we enter the quantum computing era, there is a clear need to lay out the methodological and algorithmic foundations for quantum simulation of complex classical systems. Being firmly rooted in quantum information and operator theory, the QMDA framework presented in this paper is a natural candidate for implementation in quantum computers, which we have demonstrated here by means of simulated quantum circuit experiments. As noted in the opening section of the paper, efforts to simulate classical dynamical systems on quantum computers are being actively pursued (22, 47, 48). Porting data assimilation algorithms such as QMDA to a physical quantum computational environment presents new challenges as the iterative nature of the forecast–analysis cycle will require repeated interaction between the quantum computer and the assimilated classical system, possibly using quantum sensors (49). We believe that addressing these challenges is a fruitful area for future research with both theoretical and applied dimensions.

Author affiliations: [a]Department of Mathematics, Dartmouth College, Hanover, NH 03755; [b]Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755; and [c]Department of Physics, University of Wisconsin-Milwaukee, Milwaukee 53211

1. G. P. Cressman, An operational objective analysis system. *Mon. Wea. Rev.* **87**, 367–374 (1959).
2. R. E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
3. A. J. Majda, J. Harlim, *Filtering Complex Turbulent Systems* (Cambridge University Press, Cambridge, 2012).
4. K. Law, A. Stuart, K. Zygalakis, *Data Assimilation: A Mathematical Introduction, Texts in Applied Mathematics* (Springer, New York, 2015), vol. 62.
5. E. Kalnay, *Atmospheric Modeling, Data Assimilation, and Predictability* (Cambridge University Press, Cambridge, 2003).
6. R. N. Bannister, A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteorol. Soc.* **143**, 607–633 (2016).
7. A. R. Karspeck *et al.*, A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. *Quart. J. Roy. Meteor. Soc.* **144**, 2404–2430 (2018).
8. P. J. van Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, S. Reich, Particle filters for high-dimensional geoscience applications: A review. *Quart. J. Roy. Meteorol. Soc.* **145**, 2335–2365 (2019).
9. M. Takesaki, *Theory of Operator Algebras I, Encyclopaedia of Mathematical Sciences* (Springer, Berlin, 2001), vol. 124.
10. B. O. Koopman, Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci. U.S.A.* **17**, 315–318 (1931).
11. V. Baladi, *Positive Transfer Operators and Decay of Correlations, Advanced Series in Nonlinear Dynamics* (World Scientific, Singapore, 2000), vol. 16.
12. T. Eisner, B. Farkas, M. Haase, R. Nagel, *Operator Theoretic Aspects of Ergodic Theory, Graduate Texts in Mathematics* (Springer, Cham, 2015), vol. 272.
13. G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Dover Publications, Mineola, 2009).
14. G. Alber *et al.*, *Quantum Information: An Introduction to Basic Theoretical Concepts and Experiments, Springer Tracts in Modern Physics* (Springer-Verlag, Berlin, 2001), vol. 173.
15. A. S. Holevo, *Statistical Structure of Quantum Theory, Lecture Notes in Physics Monographs* (Springer, Berlin, 2001), vol. 67.
16. M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).
17. D. Giannakis, Quantum mechanics and data assimilation. *Phys. Rev. E* **100**, 032207 (2019).
18. L. A. Takhtajan, *Quantum Mechanics for Mathematicians, Graduate Series in Mathematics* (American Mathematical Society, Providence, 2008), vol. 95.
19. D. Giannakis, J. Slawinska, Z. Zhao, "Spatiotemporal feature extraction with data-driven Koopman operators" in *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges at NIPS 2015, Proceedings of Machine Learning Research*, D. Storcheus, A. Rostamizadeh, S. Kumar, Eds. (PMLR, Montreal, Canada, 2015), vol. 44, pp. 103–115.
20. D. Giannakis, Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.* **47**, 338–396 (2019).
21. S. Das, D. Giannakis, J. Slawinska, Reproducing kernel Hilbert space compactification of unitary evolution groups. *Appl. Comput. Harmon. Anal.* **54**, 75–136 (2021).
22. D. Giannakis, A. Ourmazd, J. Schumacher, J. Slawinska, Embedding classical dynamics in a quantum computer. *Phys. Rev. A* **105**, 052404 (2022).
23. E. B. Davies, J. T. Lewis, An operational approach to quantum probability. *Commun. Math. Phys.* **17**, 239–260 (1970).
24. M. S. Leifer, R. W. Spekkens, Towards a formulation of quantum theory as a causeally neutral theory of Bayesian inference. *Phys. Rev. A* **88**, 052130 (2013).
25. B. Jacobs, F. Zanasi, A predicate/state transformer semantics for Bayesian learning. *Electron. Notes Theor. Comput. Sci.* **325**, 185–200 (2016).
26. S. Gudder, "Quantum probability" in *Handbook of Quantum Logic and Quantum Structures*, K. Engesser, D. M. Gabbary, D. Lehmann, Eds. (Elsevier, Amsterdam, 2007), pp. 121–146.
27. V. I. Paulsen, M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces, Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge, 2016), vol. 152.
28. W. F. Stinespring, Positive functions on C^*-algebras. *Proc. Amer. Math. Soc.* **6**, 211–216 (1955).
29. J. Yuval, P. A. O'Gormann, Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nat. Commun.* **11**, 3295 (2020).
30. T. Berry, D. Giannakis, J. Harlim, Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E.* **91**, 032915 (2015).
31. S. Klus, P. Koltai, C. Schütte, On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comput. Dyn.* **3**, 51–79 (2016).
32. T. Sauer, J. A. Yorke, M. Casdagli, Embedology. *J. Stat. Phys.* **65**, 579–616 (1991).
33. U. von Luxburg, M. Belkin, O. Bousquet, Consistency of spectral clustering. *Ann. Stat.* **26**, 555–586 (2008).
34. E. N. Lorenz, "Predictability of weather and climate" in *Predictability of Weather and Climate*, T. Palmer, R. Hagedorn, Eds. (Cambridge University Press, Cambridge, 1996), pp. 40–58.
35. I. Fatkullin, E. Vanden-Eijnden, A computational strategy for multiscale systems with applications to Lorenz 96 model. *J. Comput. Phys.* **200**, 605–638 (2004).
36. D. Burov, D. Giannakis, K. Manohar, A. Stuart, Kernel analog forecasting: Multiscale test problems. *Multiscale Model. Simul.* **19**, 1011–1040 (2021).
37. T. Berry, J. Harlim, Variable bandwidth diffusion kernels. *Appl. Comput. Harmon. Anal.* **40**, 68–96 (2016).
38. R. Coifman, M. Hirn, Bi-stochastic kernels via asymmetric affinity functions. *Appl. Comput. Harmon. Anal.* **35**, 177–180 (2013).
39. R. Alexander, D. Giannakis, Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *Phys. D* **409**, 132520 (2020).
40. J. Bjerknes, Atmospheric teleconnections from the equatorial pacific. *Mon. Wea. Rev.* **97**, 163–172 (1969).
41. C. Wang *et al.*, *Southern Oscillation (ENSO): A review in Coral Reefs of the Eastern Tropical Pacific: Persistence and Loss in a Dynamic Environment, Coral Reefs of the World*, P. W. Glynn, D. P. Manzello, I. C. Enoch, Eds. (Springer Netherlands, Dordrecht, 2017), vol. 8, pp. 85–106.
42. M. J. McPhaden, S. E. Zebiak, M. H. Glantz, ENSO as an integrating concept in earth system science. *Science* **314**, 1740–1745 (2006).
43. X. Wang, J. Slawinska, D. Giannakis, Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Sci. Rep.* **10**, 2636 (2020).
44. P. R. Gent *et al.*, The community climate system model version 4. *J. Climate* **24**, 4973–4991 (2011).
45. M. S. Anis *et al.*, Qiskit: An open-source framework for quantum computing (2021).
46. M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
47. I. Joseph, Koopman-von Neumann approach to quantum simulation of nonlinear classical dynamics. *Phys. Rev. Res.* **2**, 043102 (2020).
48. J. P. Liu, H. Ø. Kolden, H. K. Krovi, A. M. Childs, Efficient quantum algorithm for dissipative nonlinear differential equations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026805118 (2021).
49. H. Y. Huang *et al.*, Quantum advantage in learning from experiments. *Science* **376**, 1182–1186 (2022).