Supplementary Materials: Characterization of Background Temperature Dynamics of a Multitemporal Satellite Scene through Data Assimilation for Wildfire Detection

Gustave Udahemuka [1,*], Barend J. van Wyk [2], and Yskandar Hamam [1,3]

S1. Materials and Methods

S1.1. Modeling - *Pixel's background description*

The parameters of a Diurnal Temperature Cycle (DTC) model can be considered to have a long-term trend (i.e., an average behavior over time) and a short-term [daily] variation (i.e., a variability about that average). The expected DTC parameters of a pixel location can be taken as the long-term trend and determined by averaging parameters of a number of valid DTCs. In a real-time configuration, the expected parameters of an observed DTC can then be determined once before the start of the observed DTC and kept constant throughout the observed DTC. Therefore, in such case, the time would take a detector of a fire feature in a DTC to adapt to the *non-fire* change in the DTC is more than a day, and after a fire – especially if a fired pixel location has burned areas – or after an abrupt change in land cover (or the atmosphere), the time to adapt would be longer. One appropriate option to adapt to the *non-fire* change much faster would be to estimate the DTC parameters at short intervals. In this study, the DTC parameters are updated each time a valid sample is observed; and the process generating each of the DTC parameters is represented by a first-order Markov process (with noise assumed additive) that is expressed by

$$\boldsymbol{x}_k = f_k(\boldsymbol{x}_{k-1}) + \boldsymbol{v}_{k-1} \tag{S1}$$

where $\boldsymbol{x}_k$ denotes the state variable of the process at time step k (or at discrete time t_k: $\boldsymbol{x}_k \equiv \boldsymbol{x}(t_k)$) and is a vector containing the parameters of the DTC. Function $f_k(\cdot)$ is a forecast model operator (or state transition function) that determines the time evolution of the state. The model is assumed to contain errors $\boldsymbol{v}_k \equiv \boldsymbol{v}(t_k)$ – due, for example, to neglected physics and numerical approximations. The form of the function $f_k(\cdot)$ can be approximated theoretically or empirically. In this study, the DTC parameter vector is modeled as a Gaussian random walk process; therefore, the state equation given by Equation (S1) becomes

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \boldsymbol{v}_{k-1}; \tag{S2}$$

with $\boldsymbol{x}_k$ containing the forecast model prognostic variables (free DTC parameters) given by $\boldsymbol{x}_k \equiv (T_{0,k}, T_{a,k}, t_{m,k}, t_{s,k}, \omega_{1,k}, \omega_{2,k})$ for the BER06 DTC model [1] used in this study. The interpretation of the forecast model variables is given in Table S1. The model diagnostic variable β_k is obtainable from $\boldsymbol{x}_k$ by assuming continuity at the start of night time attenuation t_s and is given by

$$\beta_k = \frac{\omega_{2,k}}{\pi} \tan^{-1}\left(\frac{\pi}{\omega_{2,k}}(t_{s,k} - t_{m,k})\right) \tag{S3}$$

where $\tan^{-1}(\cdot) = 1/\tan(\cdot)$. The process noise vector $\boldsymbol{v}_k$ (i.e., the uncertainties in the DTC parameters) is assumed to be a Gaussian white-noise random variable with density $p_{\boldsymbol{v}_k} \sim \mathcal{N}(\boldsymbol{v}_k; \boldsymbol{0}, \boldsymbol{Q}_k)$ where $\boldsymbol{Q}_k$ is the model error covariance matrix. $\boldsymbol{v}_k$ represents change of parameters due to unmodeled effects such as frequent or occasional changes in, among others, cloud condition, soil moisture, wind condition, sunglint condition, land cover condition (e.g., land surface emissivity) on a pixel location during times of *non-fire* . Theoretically, there are discrepancies between parameters of consecutive days due to, among others, the difference in DTC daylight width (ω_1 and ω_2), and the difference in DTC full width (i.e., the length of a DTC from its start to the start of the next DTC) due to the variation in the sunrise times of consecutive days on a given pixel location [2].

Table S1. Symbols used in the physics-based DTC models.

Parameter	Meaning
T_0 (K)	'residual' temperature (at sunrise approximately)
T_a (K)	temperature amplitude
ω (hh : minutes)	half-period of cosine/sine term
t_m (hh : minutes)	time of maximum temperature
t_s (hh : minutes)	start of decay function (start of night time attenuation)
β (hh : minutes)	attenuation constant
t_{sr} (hh : minutes)	sunrise time
t_{tsr}(hh : minutes)	'thermal' sunrise time

t represents time (and is expressed in (hh : minutes)). A pseudo-physics-based approach [1] uses (ω_1, ω_2) instead of a single ω, one as a quarter-period of cosine/sin term before $t = t_m$ (i.e., ω_1) and another as a quarter-period (ω_2) after $t = t_m$.

The process model generates at time t_k an observation $\boldsymbol{y}(t_k)$, a vector made of the observed brightness temperature $y_1(t_k)$ taken from Spinning Enhanced Visible and Infrared Imager (SEVIRI) IR 3.9 and the observed offset $y_2(t_k)$ of the thermal sunrise (t_{tsr}) from the sunrise (t_{sr}) of the day. The transformation from model space to observation space is achieved through the measurement equation given by

$$\boldsymbol{y}_k = h_k(\boldsymbol{x}_k) + \boldsymbol{n}_k \tag{S4}$$

where $\boldsymbol{y}_k \equiv \boldsymbol{y}(t_k)$ denotes the observation at time step k (i.e., at time t_k), and $\boldsymbol{n}_k \equiv \boldsymbol{n}(t_k)$ represents the measurement noise. Function $h_k(\cdot)$ is the observation (or forward) operator that transforms vectors in model space, into their corresponding vector in observation space. The state-space model is then represented by Equation (S1) and Equation (S4). Transformation from the model space to the observed temperature is realized through BER06 model [1]. Hence, the measurement equation (Equation (S4)) is developed into

$$\begin{aligned} y_1(t_k) &= n_1(t_k) + h_{1,k}(\boldsymbol{x}_k) \\ &= n_1(t_k) + \begin{cases} T_{0,k} + T_{a,k}\cos\left(\frac{\pi}{\omega_{1,k}}(t_k - t_{m,k})\right) & t_{tsr,v} \le t_k < t_{m,k} \\ T_{0,k} + T_{a,k}\cos\left(\frac{\pi}{\omega_{2,k}}(t_k - t_{m,k})\right) & t_{m,k} \le t_k < t_{s,k} \\ T_{0,k} + T_{a,k}\cos\left(\frac{\pi}{\omega_{2,k}}(t_{s,k} - t_{m,k})\right) e^{-\frac{t_k - t_{s,k}}{\beta_k}} & t_{s,k} \le t_k < t_{tsr,v+1} \end{cases} \end{aligned} \tag{S5}$$

and

$$\begin{aligned} y_2(t_k) &= h_{2,k}(\boldsymbol{x}_k) + n_2(t_k) \\ &= c(t_k) = t_{m,k} - \frac{\omega_{1,k}}{2} - t_{sr,k} + n_2(t_k) \end{aligned} \tag{S6}$$

where t_k is a time between the start of the current DTC, $t_{tsr,v}$, and the start of the next DTC, $t_{tsr,v+1}$, while k is a time step that starts from the launch of the first data assimilation cycle. The offset of the thermal sunrise from the sunrise $c(t_k)$ is equal to $t_{tsr,k} - t_{sr,k}$ (the observed $t_{tsr,k}$ and $t_{sr,k}$ are the current values, i.e., $t_{tsr,v}$ and $t_{sr,v}$, that change to $t_{tsr,v+1}$ and $t_{sr,v+1}$, respectively, at the start of the next cycle). The observed offset c_k is the mean of the recent observed 7 offsets c_k's, including the offset of the current DTC. The measurement noise (i.e., the uncertainties in observations) $\boldsymbol{n}_k$ is assumed to be Gaussian white-noise random variable and statistically independent of $\boldsymbol{v}_k$, with density $p_{n_k} \sim \mathcal{N}(\boldsymbol{n}_k; \boldsymbol{0}, \boldsymbol{R}_k)$ where $\boldsymbol{R}_k$ is the observational error covariance matrix. The measurement noise $\boldsymbol{n}_k$ represents satellite sensor errors (navigation errors), errors in observation operator $h_k(\cdot)$ and the presence of subpixel-scale variability not represented in the pixel-average values of brightness temperatures (commonly known in data assimilation literature as representativeness error), and in general, the short term random errors that do not contribute to change of the forecast model (e.g., the DTC model fit on different observed DTCs can yield same DTC parameters). The forecast state $\boldsymbol{x}_k^f$ is transformed to the forecast observation by $h_k(\cdot)$. The true state before the start of the initial data assimilation cycle (i.e., the start of the simulation) $\boldsymbol{x}_0^t$ is assumed Gaussian distributed and $\{\boldsymbol{v}_k, \boldsymbol{n}_k, \boldsymbol{x}_0^t\}$ are assumed mutually, statistically

independent. At a given time while in the current DTC, the predicted offset $\hat{c}(t_k)$ is the expected offset of the next cycle and is used to derive the expected thermal sunrise of the next cycle $t_{tsr,v+1}$. DTC parameters are estimated at each time step and having forecast DTC parameters, the value of the forecast temperatures can be calculated.

S1.2. Data assimilation

In the context of numerical prediction, data assimilation is used to determine the initial state of a forecast model given observations; and is implemented to assist a numerical prediction system to produce accurate estimates of the condition (or the evolution) of a flow of an atmospheric or oceanic system or other systems [3]. At the initial time (also known as *analysis time*) t_0, the data assimilation is used to combine a first guess estimate of the model variable at time t_0 ($\boldsymbol{x}^b$, the best estimate without consideration of observations at the initial time t_0 and is also known as *background state* or prior) and observed data at t_0 (i.e., $\boldsymbol{y}_o$), so to generate the best initial estimate of the state ($\boldsymbol{x}^a$, also known as *analysis state*) at t_0 and its quality if possible, for a better initialization of the forecasts at t_0. During assimilation, the quality of the background state and observations must be considered to weigh between the background and the observations, which one contributes more to the analysis of a field of model variables. Data assimilation considers the fact that, globally, observations are limited and available at irregularly spaced points. Currently, short-range forecasts as a first guess estimate of the state at the initial time are universally adopted in analysis (or data assimilation) cycles, especially in operational numerical predictions [4,5]. At the completion of the data assimilation cycle, the analysis estimate is used to derive the short-range forecast for the next data assimilation cycle and the long-range forecast (an operational forecast). The three commonly used data assimilation methods that consider the temporal evolution of observations and produce a smooth and continuous evolution in the temporal dimension are Four-Dimensional Variational Assimilation (4D-Var), Ensemble Kalman Filter (EnKF), and Sampling Importance Resampling (SIR) particle filter. Hereafter in this section, the *background state* refers only to the first guess at an analysis time, and the background temperature of a given pixel (a term commonly used in fire detection products) is only referred to as the *non-fire* temperature.

Variational method in four dimension (i.e., 4D-Var) [e.g., 6–9] assimilates global observations of the 3-D meteorological field and adds the deviations of observations distributed over a *time* interval (assimilation window t_0 to t_{N_a}) from the model integrated in the interval from the initial time of the window t_0 to each observation time in the assimilation window. Observations are therefore assimilated into the forecast model at their observation times. The approach minimizes a cost function that penalizes both misfit to the valid observations (e.g., *non-fire* observations) in 4D and deviation from the background. The inclusion of the estimate of model error ($\boldsymbol{v}_k$ in Equation (S2)) into a 4D-Var data assimilation problem results in a variational assimilation with the forecast model used as a weak dynamical constraint and referred to as weak-constraint 4D-Var [10]. For serial uncorrelated model errors, serial uncorrelated observation errors, and uncorrelated model and observation errors (and the true state at t_0 uncorrelated to both model errors and observation errors), the state estimation of the forecast model with model error for the weak-constraint 4D-Var problem is given by the cost function:

$$\begin{aligned} J\left(\boldsymbol{x}_0, \boldsymbol{v}_0, \cdots, \boldsymbol{v}_{N_a-1}\right) &= \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^{\mathsf{T}} \boldsymbol{B}_0^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) \\ &\quad + \frac{1}{2}\sum_{n=0}^{N_a}\left[h_n\left(\boldsymbol{x}_n\right) - y_n\right]^{\mathsf{T}} \boldsymbol{R}_n^{-1}\left[h_n\left(\boldsymbol{x}_n\right) - \boldsymbol{y}_n\right] \\ &\quad + \frac{1}{2}\sum_{n=1}^{N_a} \boldsymbol{v}_{n-1}^{\mathsf{T}} \boldsymbol{Q}_{n-1}^{-1} \boldsymbol{v}_{n-1} \\ &= J_b + J_o + J_M \end{aligned} \tag{S7}$$

where n is the time step number in the assimilation window, and the control variable is made of $\boldsymbol{x}_0$ (the model state vector at t_0) and model errors $\left\{\boldsymbol{v}_n^i\right\}_{i=0}^{N_a-1}$ ($\boldsymbol{v}_{n-1} = \boldsymbol{x}_n - f_n\left(\boldsymbol{x}_{n-1}\right)$). $\boldsymbol{x}_0^b$ is the background state at the start of the window t_0, $\boldsymbol{B}_0$ is the background error covariance matrix, $\boldsymbol{y}_n$ is the observation, $h_n(\cdot)$ is the observation operator, $\boldsymbol{R}_n$ is the observation error covariance and $\boldsymbol{Q}_n$ is the model error covariance matrix. J_b is the background component of the cost function, J_o is the observation cost function (for observations distributed within the data assimilation window) and J_M is the contribution to the cost function of model error (random or systematic error). For data assimilation using the weak-constraint 4D-Var window of length N_a, $N_a + 1$

best-fit variables must be determined (i.e., the analysis at t_0 [$\boldsymbol{x}_0^a$] and N_a model errors; equivalently ($\boldsymbol{x}_0^a, \cdots, \boldsymbol{x}_{N_a}^a$)). Apart from the formulation of the commonly used weak-constraint 4D-Var presented by Equation (S7), known as the weak-constraint 4D-Var with model error forcing, other practical formulations of the weak constraint 4D-Var exist [11]. For a tractable 4D-Var analysis (i.e., quadratic cost function that gives obviously a single minima), errors are modeled as independent Gaussian random noise, and model errors and observation errors are uncorrelated (and the true state at t_0 uncorrelated to both model errors and observation errors). Under these circumstances, the posterior distribution (or analysis distribution) can be parametrized by only the mean and the covariance. In this study, a one-dimensional variational data assimilation method compatible with 4D-Var is used where only the time dimension is considered – usually in data assimilation, a name has been established for a variational data assimilation method in one vertical dimension (usually pressure levels) as 1D-Var data assimilation, commonly used for 1D-Var retrievals. The horizontal and vertical dimensions are not considered because an image pixel is processed independently of its neighbor pixels and only one vertical level is considered. A sliding assimilation time-window that moves by one sample from one analysis cycle to the next analysis cycle (i.e., a sequential "analysis cycles") is implemented, and the assimilation window is in actual fact extending from the launch of simulation (or indefinitely in the past); and such implementation is equivalent to a fixed-lag Kalman smoother. However, analysis is implemented when a *non-fire* sample is observed and is implemented with only *non-fire* observations (i.e., *fire* observations assumed to have an infinite variance). In this study, the weak-constraint 4D-Var is solved using Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System - Accelerated Representer (NAVDAS-AR) [12,13], a representer-based algorithm that implements an observation space search; coupled Euler-Lagrange equations are derived and solved using representers. A gradient-based search method is used to derive the minimum point of the cost function given that the state-space model introduced in Section S1.1 presents a forecast model function that is smooth and a non-linear observation function that is at least of class C^2. The minimum point is obtained through a gradient-based method that has been used with the NAVDAS-AR, namely, the flexible (preconditioned) conjugate gradient (untruncated version) [13,14]. The Jacobian of the nonlinear forward observation operator $h_k(\cdot)$ (an observation model that varies with time) is needed in the computation of the gradient of the cost function. It has elements $\widehat{\boldsymbol{H}}_{k,i,j} = \frac{\partial h_{k,i}}{\partial \boldsymbol{x}_{k,j}}\big|_{\boldsymbol{x_k}=\boldsymbol{x}_k^f}$ that yield an approximate linear observation operator at a time step k given by

$$\widehat{\boldsymbol{H}}_k = \begin{cases} \begin{bmatrix} 1 & 0 \\ \cos\left(\frac{\pi}{\omega_{1,k}}(t_k - t_{m,k})\right) & 0 \\ T_{a,k}\frac{\pi}{\omega_{1,k}}\sin\left(\frac{\pi}{\omega_{1,k}}(t_k - t_{m,k})\right) & 1 \\ 0 & 0 \\ T_{a,k}\frac{\pi}{\omega_{1,k}^2}t_k\sin\left(\frac{\pi}{\omega_{1,k}}(t_k - t_{m,k})\right) & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}^{\mathsf{T}} & t_{tsr,v} \le t_k < t_{m,k} \\ \begin{bmatrix} 1 & 0 \\ \cos\left(\frac{\pi}{\omega_{2,k}}(t_k - t_{m,k})\right) & 0 \\ T_{a,k}\frac{\pi}{\omega_{2,k}}\sin\left(\frac{\pi}{\omega_{2,k}}(t_k - t_{m,k})\right) & 1 \\ 0 & 0 \\ 0 & -\frac{1}{2} \\ T_{a,k}\frac{\pi}{\omega_{2,k}^2}t_k\sin\left(\frac{\pi}{\omega_{2,k}}(t_k - t_{m,k})\right) & 0 \end{bmatrix}^{\mathsf{T}} & t_{m,k} \le t_k < t_{s,k} \\ \begin{bmatrix} 1 & 0 \\ \cos\left(\frac{\pi}{\omega_{2,k}}(t_{s,k} - t_{m,k})\right)\gamma_k & 0 \\ T_{a,k}\frac{\pi}{\omega_{2,k}}\gamma_k\sin\left(\frac{\pi}{\omega_{2,k}}(t_{s,k} - t_{m,k})\right) & 1 \\ 0 & 0 \\ 0 & -\frac{1}{2} \\ T_{a,k}\frac{\pi}{\omega_{2,k}^2}t_{s,k}\gamma_k\sin\left(\frac{\pi}{\omega_{2,k}}(t_{s,k} - t_{m,k})\right) & 0 \end{bmatrix}^{\mathsf{T}} & t_{s,k} \le t_k < t_{tsr,k} \end{cases} \tag{S8}$$

where $^\intercal$ is the transpose operator, $\gamma_k = e^{-\frac{t_k - t_{s,k}}{\beta_k}}$, $\boldsymbol{x}_k \equiv (T_{0,k}\ , T_{a,k}\ , t_{m,k}, t_{s,k}, \omega_{1,k}, \omega_{2,k},)^\intercal$. The covariance $\boldsymbol{B}_0$ is kept constant in time and is calculated once for the whole region of interest. $\boldsymbol{B}_0$ is given by Equation (S26). The estimation of the background error covariance $\boldsymbol{B}_0$ and the model error covariance $\boldsymbol{Q}_n$ that are kept constant in time, and the estimation of the observation covariance $\boldsymbol{R}_n$ that varies with observed DTC are described in Section S1.3. From one DTC to the next, the estimated $\boldsymbol{R}_k$ of the outgoing DTC is considered in each assimilation cycle until the start of the sliding assimilation window is completely out of the outgoing DTC, and $\boldsymbol{R}_k$ is always constant in the assimilation window. For the first analysis (at launch of the simulation), $\boldsymbol{x}_0^b = \boldsymbol{x}_1^f$ is assumed Gaussian distributed with mean and covariance specified in Section S1.3. The choice of the length of an assimilation window is crucial as a short assimilation window is required for a fast adaptation to the change on land and in the atmosphere, and a long assimilation window is needed for a robust initialization of a forecast but risks introducing the problem of multiple minima due to nonlinearity in state space model or time varying error statistics [e.g., 15]. In this study, a data assimilation window of length $N_w = 6$ (i.e., equal to the number of degrees of freedom of the forecast model) is considered; and given that, the 6 samples must be under *non-fire* condition, the window can extend further in the past to accommodate 6 *non-fire* samples. It is assumed that for the first 5 samples after the start of the simulation (i.e., the first analysis cycle), no pixel location is burning, and any subsequent sixth sample to include in the assimilation window for a given pixel location must first be evaluated to determine whether is a *fire* pixel given a *non-fire* temperature fit and observation at the time of the sixth sample. In this study, The NAVDAS-AR's stopping criteria (for notations, refer to [12,16]) are for the outer loop: maximum number of outer-iterations = $N_a + 1$ or $\frac{\|(\delta \boldsymbol{x}^p)^j + \boldsymbol{g}\|}{\|\boldsymbol{x}^j\|} < 0.001$. The stopping criteria for the inner loop are: maximum number of inner-iterations equals to the double of the dimension of the observation space or $\frac{\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{\beta}^j\|}{\|\boldsymbol{b}\|} < 0.001$ where $\boldsymbol{b} = \boldsymbol{y} - \mathcal{H}(\boldsymbol{x}^{j-1}) - \boldsymbol{H}^{j-1}\delta(\boldsymbol{x}^p)^j$ and $\boldsymbol{A} = \boldsymbol{H}^{j-1}(\boldsymbol{P}^b)^{j-1}[\boldsymbol{H}^{j-1}]^\intercal + \boldsymbol{R}$.

The EnKF method, originally developed by [17], approximates via Monte Carlo (MC) methods, both the forecast probability density function (pdf) that represents the background state and the analysis pdf (the posterior) with the densities represented by ensemble of model states. An ensemble of N_e data assimilation cycles, carried out simultaneously, assimilates the same real observation. After completing the ensemble of analyses at time t_{k-1}, i.e., $\{\boldsymbol{x}_{k,i}^a\}_{i=1}^{N_e}$ (i for ensemble member i), the "forecast" step of EnKF advances the forecast and the forecast error covariance matrix at the next analysis time t_k (i.e., EnKF is made of sequential "analysis cycles"). The forecast estimate of the state at current initial time t_k given by $\boldsymbol{x}_{k,i}^f = f_{k-1}(\boldsymbol{x}_{k-1,i}^a)$ is a short-range model forecast same as the background state $\boldsymbol{x}_0^b$ mentioned for 4D-Var in Equation (S7) when the background state is taken as the short-range forecast. At the assimilation of observations, various implementations of EnKF algorithm update ensembles (i.e., generate analysis ensemble) differently. One group implements a stochastic update of the ensembles (stochastic approach) while another implements a deterministic update of the ensembles. This study implements the EnKF algorithm with perturbed observations (perturbed observation filters) [e.g., 18,19]) – the main method of the group whose ensembles are updated stochastically. Ensemble members are updated during analysis using a random perturbed observation $\{\boldsymbol{y}_{k,i}\}_{i=1}^{Ne}$, which is created by adding random noise to the real observation $\boldsymbol{y}_k = \boldsymbol{y}_k^0$ (i.e., $\boldsymbol{y}_{k,i} = \boldsymbol{y}_k + \boldsymbol{\varepsilon}_i$ with $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k)$ and $\boldsymbol{R}_k$ is the observational error covariance matrix) [18,19]. For a state space model given by Equation (S1) and Equation (S4), given the ensemble forecast (i.e., prior ensemble) at t_k, $\{\boldsymbol{x}_{k,i}^f\}_{i=1}^{Ne}$, and the current random perturbed observation $\{y_{k,i}\}_{i=1}^{Ne}$ as well as the second-order statistics of the forecast estimate and observation errors, the data assimilation by EnKF with perturbed observations combines the short-range forecast (i.e., the background state) and the observations to produce the analysis estimate of the state at time t_k ($\boldsymbol{x}_k^a$) given by

$$\boldsymbol{x}_{k,i}^a = \boldsymbol{x}_{k,i}^f + \boldsymbol{G}_{f,k}[\boldsymbol{y}_{k,i} - h_k(\boldsymbol{x}_{k,i}^f)] \quad \forall\ i = 1 : N_e \tag{S9}$$

where $\boldsymbol{G}_{f,k}$ is a weight matrix (also referred to as Kalman gain) at time k. The analysis point estimate is computed as $\overline{\boldsymbol{x}_k^a} = \frac{1}{N_e}\sum_{i=1}^{N_e}\boldsymbol{x}_{k,i}^a$. Equation (S9) expresses the statistical interpolation mechanism in data assimilation [20]. In this study, the weight matrix $\boldsymbol{G}_{fk}$ is derived according to [21], given by

$$\boldsymbol{G}_{fk} = \boldsymbol{P}_k^f \boldsymbol{H}^\intercal (\boldsymbol{R}_k + \boldsymbol{H}_k \boldsymbol{P}_k^f \boldsymbol{H}_k^\intercal)^{-1} \tag{S10}$$

where

$$\overline{\boldsymbol{x}_k^f} = \frac{1}{N_e}\sum_{i=1}^{N_e}\boldsymbol{x}_{k,i}^f \tag{S11}$$

$$\overline{h_k(\boldsymbol{x}_k^f)} = \frac{1}{N_e}\sum_{i=1}^{N_e}h_k(\boldsymbol{x}_{k,i}^f) \tag{S12}$$

$$\boldsymbol{X}_{k,i}^f = \boldsymbol{x}_{k,i}^f - \overline{\boldsymbol{x}_k^f} \quad \forall\, i = 1:N_e \tag{S13}$$

$$\boldsymbol{HX}_{k,i}^f = h_k(\boldsymbol{x}_{k,i}^f) - \overline{h_k(\boldsymbol{x}_k^f)} \quad \forall\, i = 1:N_e \tag{S14}$$

$$\boldsymbol{P}_k^f\boldsymbol{H}_k^\intercal = \frac{1}{N_e-1}\sum_{i=1}^{N_e}\boldsymbol{X}_{k,i}^f[\boldsymbol{HX}^f]_{k,i}^\intercal \tag{S15}$$

$$\boldsymbol{H}_k\boldsymbol{P}_k^f\boldsymbol{H}_k^\intercal = \frac{1}{N_e-1}\sum_{i=1}^{N_e}\boldsymbol{HX}_{k,i}^f[\boldsymbol{HX}^f]_{k,i}^\intercal \tag{S16}$$

EnKF produces state-dependent uncertainty (i.e., provides analysis perturbations that represent analysis error covariance, and ensemble forecasts that represent the forecast error covariance matrix). While the forecast error covariance matrix of the standard 4D-Var at the start of the assimilation window t_0 (i.e., the background error covariance $\boldsymbol{B}_0$) is assumed constant and the forecast error covariance evolves only and implicitly in the assimilation window [20,22], the analysis by EnKF replaces the constant background error covariance matrix by a forecast error covariance matrix that evolves in time from one analysis to another. The ensemble-evolved forecast error covariance matrix $\boldsymbol{P}_k^f$ at time step k can be estimated from the N_e forecasts $\{\boldsymbol{x}_{k,i}^f\}_{i=1}^{N_e}$. From one data assimilation cycle to the next, EnKF updates the forecast error; and the forecast error at a given analysis time depends on the initial (analysis) error at the last analysis, on the errors introduced by the forecast model from one data assimilation cycle to the next, and the model errors. The EnKF was designed with the aim of reducing the computational complexity of the standard Kalman filter (or the extended Kalman filter for nonlinear state-space model) in case of a model with a very large number of degrees of freedom. The standard Kalman filter requires a cost increase of the order of the number of degrees of freedom of the model. This cost emanates from the forward integration in time of the full analysis error covariance matrix using a linear forecast model operator, which is of the size of the model space, to update the forecast error covariance. The EnKF reduces the cost of the standard Kalman filter by deriving a low-rank error covariance approximated using ensemble with very few members compared to the size of the model space. The performance of EnKF depends on the number of ensemble members. However, the EnKF has been proven to perform relatively well with 10 to 100 ensemble members for a model with a number of degrees of freedom greater than 10^5 (e.g., [18,23,24]). The EnKF accommodates nonlinearity in a state-space model by statistical linearization of nonlinear functions through ensemble (an implicit linearization). In all different implementations of EnKF, for a tractable EnKF analysis, errors are modeled as independent Gaussian random noise, and model error and observation error are uncorrelated at each time step (the true state at initial analysis cycle is uncorrelated to both model errors and observation errors). The posterior distribution (or analysis distribution) is then parametrized by only the mean and the covariance. The estimate of model error and observation error covariances $\boldsymbol{Q}$ and $\boldsymbol{R}_k$, respectively, are explained in Section S1.3. At the launch time of the EnKF, $\boldsymbol{x}_1^f$ is assumed Gaussian distributed with mean and covariance specified in Section S1.3, and realizations of $\boldsymbol{x}_1^f$ form the ensemble of background field valid at the launch time. The forecast model (see Equation (S2)) has 6 degrees of freedom per pixel (i.e., number of state variables). In this study, the number of ensemble members is considered to be $N_e = 51$ (few ensemble members), with the expectation of a full-rank forecast error covariance. Noting that, the full benefit of the EnKF can be exploited if the number of ensemble members would be better less than 6. Otherwise, the number can be kept to 51 and, at the same time, take into consideration the spatial correlation between neighbor pixels to have a state variable whose size equals the product of 6 and the number of pixels.

The SIR particle filter uses MC method to estimate, at time step k (i.e., observation time t_k), the true analysis pdf (i.e., the posterior pdf of a state of a dynamical system $p(\boldsymbol{x}_k|\mathbb{Y}_k)$, where $\boldsymbol{x}_k$ is the state at time step k and $\mathbb{Y}_k = \{y_1, y_2, \cdots, y_k\}$ is the set of observed data from the launch time of the filter until the present time step k) through a recursive Bayesian estimation [25]. The SIR approximates the true analysis pdf by a discrete weighted

distribution represented by a set of N_p random samples of the state and their associated weights. The true analysis pdf is estimated using a sequential version of the Importance Sampling (IS) approach (i.e., Sequential Importance Sampling (SIS)) that represents the density at time step k by a set of N_p MC samples (i.e., particles) of the state $\{x_{k,i}\}_{i=1}^{N_p}$ and their associated importance weights $\{w_{k,i}\}_{i=1}^{N_p}$, and propagates particle points and their associated weights from one time step to the next (e.g., from $k-1$ to k for the case of the propagation in time) continuously as observations are received sequentially. The approximation to the analysis density at time step k is given by

$$p\left(\boldsymbol{x}_k|\mathbb{Y}_k\right) \approx \sum_{i=1}^{N_p} \frac{w_{k,i}\delta\left(\boldsymbol{x}_k - \boldsymbol{x}_{k,i}\right)}{\sum_{j=1}^{N_p} w_{k,j}} \tag{S17}$$

where $\delta(\cdot)$ is the Dirac delta function and the normalized importance weight $w_{k,i}/\sum_{j=1}^{N_p} w_{k,j}$ is an approximation to the analysis probability of the particle $\boldsymbol{x}_{k,i}$. Different importance densities have been used in approximation of the true analysis density in case of SIR particle filter, and in this study the (state) transition prior probability distribution $p\left(\boldsymbol{x}_k|\boldsymbol{x}_{k-1,i}\right)$ as a choice of the importance (proposal) density to draw sample particles from is considered (same as in the standard SIR particle filter presented by [25]). The particle $\boldsymbol{x}_{k,i}$ and the unnormalized importance weights $w_{k,i}$ (associated with a particle $\boldsymbol{x}_{k,i}$) at the observation of $\boldsymbol{y}_k$ are then given by

$$\begin{cases} \boldsymbol{x}_{k,i} \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1,i}) : \boldsymbol{x}_{k,i}^f = f_k\left(\boldsymbol{x}_{k-1,i}^a\right) \\ w_{k,i} \propto w_{k-1,i} p(\boldsymbol{y}_k|\boldsymbol{x}_{k,i}) \end{cases} \tag{S18}$$

Given samples $\{\boldsymbol{x}_{k-1,i}\}_{i=1}^{N_p}$ that represent the analysis estimate at $k-1$ (i.e., $p\left(\boldsymbol{x}_{k-1}|\mathbb{Y}_{k-1}\right)$), the transition prior density will propose new samples $\{\boldsymbol{x}_{k,i}\}_{i=1}^{N_p}$ (i.e., $\boldsymbol{x}_{k,i}^f = f_{k-1}\left(\boldsymbol{x}_{k-1,i}^a\right)$). Then, to complete one iteration of the recursion, the importance weights are computed where the likelihood function or measurement density $p\left(\boldsymbol{y}_k|\boldsymbol{x}_k\right)$ corresponds to the measurement equation given by Equation (S4) with known statistics of $\boldsymbol{n}_k$, i.e., known $p(\boldsymbol{n}_k)$. The recursion forms the SIS part of the SIR particle filter and it assumes that the sequence $\{\boldsymbol{v}_{k-1}, k \in \mathbb{Z}^+\}$ and the sequence $\{\boldsymbol{n}_k, k \in \mathbb{Z}^+\}$ are mutually uncorrelated and the initial value of the state $\boldsymbol{x}_0$ (with known distribution) is uncorrelated to both $\boldsymbol{v}_k$ and $\boldsymbol{n}_k$. The standalone SIS algorithm suffers mainly from an undesirable effect named weight degeneracy (a phenomenon that is confirmed when only one or few particles are significant and others have negligible weights) [26], and the standard SIR [25] mitigates the degeneracy problem by including a resampling step after each SIS implementation. In the resampling step at time step k, a new set of N_p particles are reproduced from the SIS set at time step k by replacing the SIS particles with small weights with replicates of the particles with large weights [27], and results in importance weights $w_{k-1,i} = 1/N_p \ \forall i$. In this study, N_p particles and their associated weights are computed in one data assimilation cycle of the SIR particle filter, and are given by

$$\begin{cases} \boldsymbol{x}_{k,i} \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1,i}) : \boldsymbol{x}_{k,i}^f = f_k\left(\boldsymbol{x}_{k-1,i}^a\right) + g(\boldsymbol{v}_{k-1}) \\ w_{k,i} \propto p(\boldsymbol{y}_k|\boldsymbol{x}_{k,i}) = e^{-\frac{1}{2}\left(\boldsymbol{y}_k - h_k\left(\boldsymbol{x}_{k,i}^f\right)\right)^{\mathsf{T}} \boldsymbol{R}_k^{-1}\left(\boldsymbol{y}_k - h_k\left(\boldsymbol{x}_{k,i}^f\right)\right)} \end{cases} \tag{S19}$$

that results in a set $\{\boldsymbol{x}_{k,i}, w_{k,i}\}_{i=1}^{N_p}$; and this step is followed by a normalization of importance weights:

$$w_{k,i} = w_{k,i}/\textstyle\sum_{j=1}^{N_p} w_{k,j}, \tag{S20}$$

and then resampling stage to derive particles that form the analysis ensemble:

$$\left[\{\boldsymbol{x}_{k,i}^a, 1/N_p\}_{i=1}^{N_p}\right] = \text{resample}\left[\{\boldsymbol{x}_{k,i}^f, w_{k,i}\}_{i=1}^{N_p}\right]. \tag{S21}$$

After each SIR analysis, particles are propagated forward for the next assimilation cycle (sequential "analysis cycles"). The sample impoverishment is a loss of diversity among the particles explained by the fact that particles with large weights are statistically selected many times during resampling, and the resultant sample set contains then many repeated points and few distinct points. As the number of distinct points decrease, the particles can

eventually coalesce into a single particle. To counter the sample impoverishment, a small perturbation $g(\boldsymbol{v}_{k-1})$ (see Equation (S19)) is added to the resampled particles and is given by

$$g(\boldsymbol{v}_k) = \frac{\boldsymbol{v}_k}{\text{erf}^{-1}(1-\frac{1}{N})\sqrt{2\rho(\boldsymbol{Q}_k)}} \tag{S22}$$

where N is the total number of generated random values at time step k (the product of the number of pixels, the number of particles and the state dimension) and $\rho(\boldsymbol{Q})$ is the spectral radius of the model error covariance matrix $\boldsymbol{Q}$. In this study, the resampling step is implemented using systematic resampling scheme [28]. The standard SIR particle filter [25] with the systematic resampling that corresponds to the algorithm presented by [29] is therefore the algorithm used in this study. The multinomial resampling is used in the standard SIR particle filter but is more expensive per operation compared to the systematic resampling [e.g, 30,31] – systematic resampling is also less expensive per operation compared to other commonly used resampling algorithms for particle filters – and the quality of samples generated by the systematic resampling is as good as the quality of samples generated by the multinomial resampling [e.g, 30–32]. The assumptions needed in order to use the standard SIR are weak compared to the assumptions required by other particle filter algorithms [33,34]. However, as the resampling is applied at every iteration, it is more prone to problems introduced by the resampling step such as the sample impoverishment. Same as the EnKF, the background (or forecast) error covariance evolves with time in the SIR particle filter through the forecast distribution (i.e., the prior distribution) $p(\boldsymbol{x}_k|\mathbb{Y}_{k-1})$. The EnKF assumes the analysis distribution to be always Gaussian. However, the assumption is highly violated in case of strong nonlinearity in a state-space model or long analysis steps (distance between two analysis cycles)even under Gaussian environment. The SIR particle filter, on the other hand, represents any analysis density, which is generally of an arbitrary function form (i.e., not necessary Gaussian) under Gaussian environment for nonlinear state-space models or non-Gaussian environment. The particle filters have a limit on parallelization due to the resampling step. The model error is accounted for in this study, and the estimate of model error and observation error covariances $\boldsymbol{Q}$ and $\boldsymbol{R}_k$, respectively, are explained in Section S1.3 and the background (forecast) state at the first analysis cycle, $\boldsymbol{x}_1^f$, is assumed Gaussian distributed with mean and covariance specified in Section S1.3. With the number of degrees of freedom being 6 (number of state variables) per pixel for the dynamical model in Equation (S2), the same number of ensemble members considered for EnKF, are used for SIR particle filter in this study, i.e., the number of particle $N_p = 51$.

A new analysis is implemented at each time an observation is flagged as a *non-fire* observation (i.e., *fire*-contaminated samples are not assimilated). The analysis procedure during a possible fire condition is detailed in Algorithm S1.

Algorithm S1. Data assimilation under possible *fire* conditions.

1: Initialize the time step k, and also derive $\boldsymbol{Q}$, $\boldsymbol{B}_0$ and initial state at $k = 1$ as explained in Section S1.3.
2: **for** each time step k **do**
3: Retrieve background (or predicted) estimates – i.e., $\boldsymbol{x}_0^b = f_k(\boldsymbol{x}_{0,k-1}^a)$ and $\boldsymbol{B}_0$ for 4D-Var, or $\{\boldsymbol{x}_{k,i}^f\}_{i=1}^{51}$ drawn from state transition prior distribution $p(x_k|x_{k-1,i})$ for EnKF and SIR.
4: Collect $\boldsymbol{y}_k$ for EnKF and SIR while for 4D-Var, 5 recent *non-fire* observations $\{\boldsymbol{y}_{k-5}, \cdots, \boldsymbol{y}_{k-1}\}$ must also be collected.
5: **if** $\boldsymbol{y}_k$ is a *fire*-contaminated observation (see Algorithm 1 in the main article) **then**
6: No data assimilation
7: **else**
8: $\boldsymbol{R}_k \Leftarrow \boldsymbol{R}^*$ ($\boldsymbol{R}^*$ is constant throughout the current observed DTC. Its estimation is explained in Section S1.3.)
9: Compute the analysis $\{\boldsymbol{x}_{0,k}^a, \cdots, \boldsymbol{x}_{5,k}^a\}$ for 4D-Var or $\{\boldsymbol{x}_{k,i}^a\}_{i=1}^{51}$ for EnKF and SIR – i.e., evaluate Equation (S7) for 4D-Var using NAVDAS-AR algorithm [12], Equation (S9) for EnKF; and Equations (S19), (S20), (S21) for SIR.
10: **end if**
11: **end for**

S1.3. Derivation of pixel non-fire -state parameters

The measurement error $[n_1(t_k), n_2(t_k)]^\intercal$ (see Equations (S5) and (S6)) provides the error in the observation vector $\boldsymbol{y}_k$ formed by the observed brightness temperature and the observed shift of the thermal sunrise from the sunrise (the offset of t_{tsr} from t_{sr}). In this study, the two elements of the measurement error are assumed uncorrelated and the measurement error variance-covariance $\boldsymbol{R}^*$ during data assimilation cycles in the current DTC of a given pixel location, is then a 2×2 diagonal matrix with element $\boldsymbol{R}^*_{1,1}$ being the variance of $n_1(t_k)$, considered to be the sample variance of the misfits of observations on a DTC model (in this case, the BER06 model) from previous valid 7 DTCs (valid according to the conditions set on DTC parameters, presented in Table S2). Element $\boldsymbol{R}^*_{2,2}$ is the sample variance of the shifts $c(t_k)$'s in previous valid 7 DTCs. $\boldsymbol{R}^*$ is constant during all data assimilation cycles implemented in a single DTC of a pixel location. It is estimated in 15 minutes before the start of the current DTC, and is the only parameter that is learned online.

Table S2. Conditions that prompts the removal of a given DTC from the training set.

'Rejection conditions'
$t_m <$ solar noon
$t_m >$ solar noon $+ 6$
$t_m > t_s - 1$
$T_a \leq 0$
$t_s <$ sunset $- 3$
$t_s >$ sunset $+ 3$
$\beta > 100$
$\omega_1 < 0$

The constraints are linked by an *OR* operation.

The forecast model has 6 degrees of freedom (i.e., is made of 6 free DTC parameters), and the statistical uncertainty in the model, $\boldsymbol{v}_k$ (see Equation (S2)), is represented by a 6×6 model error variance-covariance matrix $\boldsymbol{Q}_k$ defined on a pixel location and kept constant on a pixel location throughout the simulation. Different methods can be used to estimate the covariance matrix, and the natural estimate is the sample covariance as an unbiased estimate of the actual covariance matrix. The size of the sample needed to compute the sample covariance matrix of an n-dimensional random variable can be set to a value greater than the minimum sample size that guarantees a certain quality or level of accuracy (with a high probability) of the sample covariance. This minimum value depends on the probabilistic distribution and independence of the samples, and the size of the population from which the samples are being drawn [35]. A pixel in a land surface image acquired by a sensor at the Top Of the Atmosphere (TOA) can be described using sample data provided by the sensor on the pixel location over a certain period. To ensure that enough information on the pixel location is extracted, the period must be a compromise between a too-short period that increases the probability of extracting only cloudy measurements and a too-long period that ignores changes in the land cover on the pixel location. The problem is then one of defining the population. In this study, the covariance matrix $\boldsymbol{Q}_k$ is estimated using DTCs extracted over 15 days (i.e., assumption of no abrupt change in land cover/atmosphere over the period) ending on 2007/August/05 (excluding this date) in the region of interest, and only valid DTCs in a single-pixel location according to conditions set in Table S2 are chosen to estimate the covariance matrix in the pixel location. A sample size of 15 DTCs (i.e., if all DTCs in 15 days are valid) complies with the minimum sample size required to capture the covariance matrix in ensemble-based filters, EnKF-based and sigma-point Kalman filter-based methods [18,23,24,36,37]. Sampling error in the estimate of $\boldsymbol{Q}_k$, due to the small sample size, is reduced, and the estimate improved by estimating $\boldsymbol{Q}_k$ using a pooling mechanism [38]. This mechanism is one of the methods that are commonly used to estimate the covariance matrix of an n-dimensional random variable in case the sample size is less than n. The reason for its use in this study is that it is simple and has a theoretical justification when the dimensionality of the random variable is large. The method also gives potentially a full rank covariance matrix estimate and does not exhibit a full or partial loss of information on the covariance matrix of a pixel location. The pooled sample (variance-) covariance matrix of a pixel at location (i, j) is determined by taking into consideration its 3×3 neighborhood and is defined as

$$\boldsymbol{Q}_{(i,j),k} \equiv \boldsymbol{Q}_{(i,j)} = \frac{\sum_{p=-1}^{1}\sum_{q=-1}^{1} N_{(i+p,j+q)}\boldsymbol{S}_{(i+p,j+q)}}{\sum_{p=-1}^{1}\sum_{q=-1}^{1} N_{(i+p,j+q)} - 9} \tag{S23}$$

where $N_{(i,j)}$ and $\boldsymbol{S}_{(i,j)}$, are the number of valid DTCs and the sample covariance matrix (unbiased version), respectively, at location (i,j). The sample covariance matrix $\boldsymbol{S}_{(i,j)}$ of an individual pixel location (i,j) (considering the information received only from (i,j)) is given by

$$\boldsymbol{S}_{(i,j)} = \frac{1}{N_{(i,j)}-1}\sum_{d=1}^{N_{(i,j)}} (\boldsymbol{p}_{(i,j),d} - \bar{\boldsymbol{p}}_{(i,j)})(\boldsymbol{p}_{(i,j),d} - \bar{\boldsymbol{p}}_{(i,j)})^{\mathsf{T}} \tag{S24}$$

where $\boldsymbol{p}_{(i,j),d}$ is a 6-dimensional vector of the free DTC parameters, of a single DTC on day d at the pixel location (i,j). It is derived from the robust fitting of the observed DTC on the DTC model BER06 . In this study, the robust fitting is achieved using the Geman-McClure error function and the Nelder-Mead simplex optimization method. $\bar{\boldsymbol{p}}_{(i,j)}$ is the sample mean vector from $N_{(i,j)}$ DTCs. The pooled estimate at (i,j) is a consistent estimator of the actual covariance matrix at (i,j), only when the actual covariance matrices in all pixel locations in a 3×3 neighborhood of location (i,j) are equal. Though pixels in a 3×3 METEOSAT Second Generation (MSG) Level 1.5 data used in this study overlap (due to an oversample factor of 1.6), statistical independence between observations of neighbor pixels in a 3×3 can hold under *non-fire* and no-cloud conditions due to the coarse spatial resolution of the MSG pixels and due to the atmosphere. In case the DTC parameters from neighbor pixels are uncorrelated, the sampling error in the pooled estimate reduces.

Background error variance-covariance $\boldsymbol{B}_0$ used in 4D-Var (see Equation (S7)) can be computed using different methods [39]. In this study, it is estimated based on one of the operational approaches used to derive the Three-Dimensional Variational Assimilation (3D-Var) background error variance-covariance presented by [40], namely, the National Meteorological Center (NMC) (the current National Centers for Environmental Prediction (NCEP)) method, which is widely adopted and referenced by different centers that conduct researches on numerical weather predictions or that operate numerical weather predictions, and it is given by

$$\boldsymbol{B}_0 \approx \lambda\mathbb{E}\left[\left[\boldsymbol{x}^f(48\text{ h}) - \boldsymbol{x}^f(24\text{ h})\right]\left[\boldsymbol{x}^f(48\text{ h}) - \boldsymbol{x}^f(24\text{ h})\right]^{\mathsf{T}}\right], \tag{S25}$$

where λ is a scale parameter and $\boldsymbol{x}^f(t_k)$ is t_k-forecast. $\boldsymbol{B}_0$ estimate is a statistic of the difference between 24-hour and 48-hour forecasts (two short-range forecasts, from 24-hour-apart initialization times, verifying at the same time). Based on Equation (S25) and the forecast model given by Equation (S2), $\boldsymbol{B}_0$ is found by the covariance of the difference between parameters (here, referring specifically to the free DTC parameters) of two consecutive DTCs of a single-pixel location. $\boldsymbol{B}_0$ (a 6×6 matrix) for an individual pixel in the study area is evaluated by randomly drawing 2,500 land pixel locations among land pixel locations not affected by fire throughout 15 days ending on 2007/August/05 (excluding this date), according to MODIS MOD14/MYD14 and EUMETSAT FIR products. It is computed as the sample covariance matrix (unbiased version) of the vector difference between the fits of two, randomly chosen, consecutive valid DTCs per single-pixel location in the period considered, and yields

$$\boldsymbol{B}_0 = \begin{bmatrix} 9.8178 & -5.5937 & -0.0562 & 0.2861 & -2.4879 & 0.2050 \\ -5.5937 & 9.6852 & 0.1932 & -0.7335 & 1.3157 & 0.7264 \\ -0.0562 & 0.1932 & 0.0938 & -0.0529 & 0.2396 & -0.2255 \\ 0.2861 & -0.7335 & -0.0529 & 0.4740 & -0.2062 & 0.3343 \\ -2.4879 & 1.3157 & 0.2396 & -0.2062 & 1.8044 & -0.8073 \\ 0.2050 & 0.7264 & -0.2255 & 0.3343 & -0.8073 & 1.9948 \end{bmatrix} \tag{S26}$$

$\boldsymbol{B}_0$ is constant in time and is the same for all pixel locations on the whole study region. The scale parameter λ in Equation (S25) can also be considered, but instead, this study looked on case correlation between DTC's parameters is considered or ignored, and better results which are reported in this study were found when the correlation is ignored (off-diagonal elements of the covariance matrix are set to zero).

At the launch time of the simulation (i.e., the first data assimilation cycle), the background field (in case of 4D-Var) or model forecast (in case of ensemble-based filters) is assumed Gaussian distributed. The background estimate $\boldsymbol{x}_0^b$ or the expected model forecast estimate $\mathbb{E}\left[\boldsymbol{x}_1^f\right]$ of a pixel location is set to the mean vector of the free parameters of the valid DTCs in 15 days of the pixel location ending on 2007/August/05 (excluding this date). The covariance matrix of the error in the predicted estimate at launch, or a priori estimate covariance $\boldsymbol{P}_1^f$ is equal to the estimated process noise covariance matrix $\boldsymbol{Q}$, and serves in the generation of the initial ensemble (particles) for the ensemble-based methods at the start of the simulation.

S1.4. Ensemble forecasting-based change detection

After each data assimilation cycle (i.e., having established the initial condition of a forecast), a numerical prediction system aims to predict or forecast far-off (i.e., over a long model integration time) with a small error. Albeit errors in initial conditions and deficiency of the model, that cannot totally be removed and restrict the limit of predictability (of atmospheric or oceanic system or other systems), some forecast techniques that account for error in initial conditions and the model errors have been designed to improve predictability (i.e., achieve long-range predictability and together, gain a skillful forecast). The more preferable technique is the ensemble forecasting [41,42]. The technique propagates the uncertainty, and its forecasts are derived from various initial conditions and/or models. In this study, the ensemble forecasting method produces an ensemble that, at initial/analysis time, reflects the uncertainty in the analysis. The EnKF and the SIR particle filter (i.e., methods that provide MC approximations to the analysis) produce an ensemble of analyses that serve as a basis of ensemble forecasting – the ensemble analysis is also propagated forward for the next assimilation cycle. The initialization of ensemble forecasts from the weak-constraint 4D-Var analysis is achieved, in this study, through perturbations of the initial condition. These perturbations are generated using one of the operational ensemble forecasting methods, the breeding approach [43,44]. An N_e-member ensemble consists of one control forecast (control run) and $(N_e - 1)$ perturbed members (perturbed forecasts). The breeding method seeks to create perturbations for the initial condition that exhibit fast growth during the forecast. The faster the initial error/perturbation growth, the better the perturbations for the ensemble prediction – the quality of ensemble is measured by the skill of the mean of ensemble forecasts to approximate more closely the true evolution [41]. In this study, the ensemble initial conditions are created by generating perturbations in the background state at launch time (i.e., around $\boldsymbol{x}_b$ in the first analysis cycle). $\frac{1}{2}(N_e - 1)$ MC samples (random perturbations) are drawn from a Gaussian distribution with zero mean and covariance matrix $\boldsymbol{P}_1^f = \boldsymbol{B}_0$, and are added and subtracted to $\boldsymbol{x}_0^b$. The forecast model is then integrated from the control and from the perturbed initial conditions; and perturbations around the control run scaled at the end of assimilation window of each 4D-Var analysis (i.e., at observation time of the latest assimilated data) to construct $(N_e - 1)$ perturbed members at analysis time (i.e., a set $\{\boldsymbol{x}_{k,i}^a\}_{i=2}^{N_e}$). The perturbations are vectors of equal size measured by a norm $\|\cdot\|_{\boldsymbol{W}}$ defined by a weight matrix $\boldsymbol{W}$ given by $\boldsymbol{W}^\intercal\boldsymbol{W} = \boldsymbol{A}_k^{-1}$, where $\boldsymbol{A}_k$ is an estimate of the analysis error covariance matrix computed at the end of an assimilation window of an analysis cycle. $\boldsymbol{A}_k$ can be estimated using different methods, each with its specific assumptions (see [45] and references therein). In this study, the analysis error covariance of the extended Kalman filter with a constant forecast error covariance $\boldsymbol{B}_0$ is adopted. It is a simple approximation given by

$$\boldsymbol{A}_k = \left(\boldsymbol{I} - \boldsymbol{B}_0\widehat{\boldsymbol{H}}_k^\intercal(\widehat{\boldsymbol{H}}_k\boldsymbol{B}_0\widehat{\boldsymbol{H}}_k^\intercal + \boldsymbol{R}_k)^{-1}\widehat{\boldsymbol{H}}_k\right)\boldsymbol{B}_0 \tag{S27}$$

where $\widehat{\boldsymbol{H}}_k$ is the Jacobian matrix of the forward observation operator $h_k(\cdot)$ (given by Equation (S8)) evaluated at $\boldsymbol{x}_k^f$. The norm of the i^{th} perturbation vector $\boldsymbol{p}_i$ is defined by $\|\boldsymbol{p}_i\|_{\boldsymbol{W}}^2 = \boldsymbol{p}_i^\intercal\boldsymbol{W}^\intercal\boldsymbol{W}\boldsymbol{p}_i$. The i^{th} perturbed member determined by adding the scaled perturbation to the (unperturbed) analysis $\boldsymbol{x}_k^a$ – i.e., adding to the initial of the control forecast – is given by

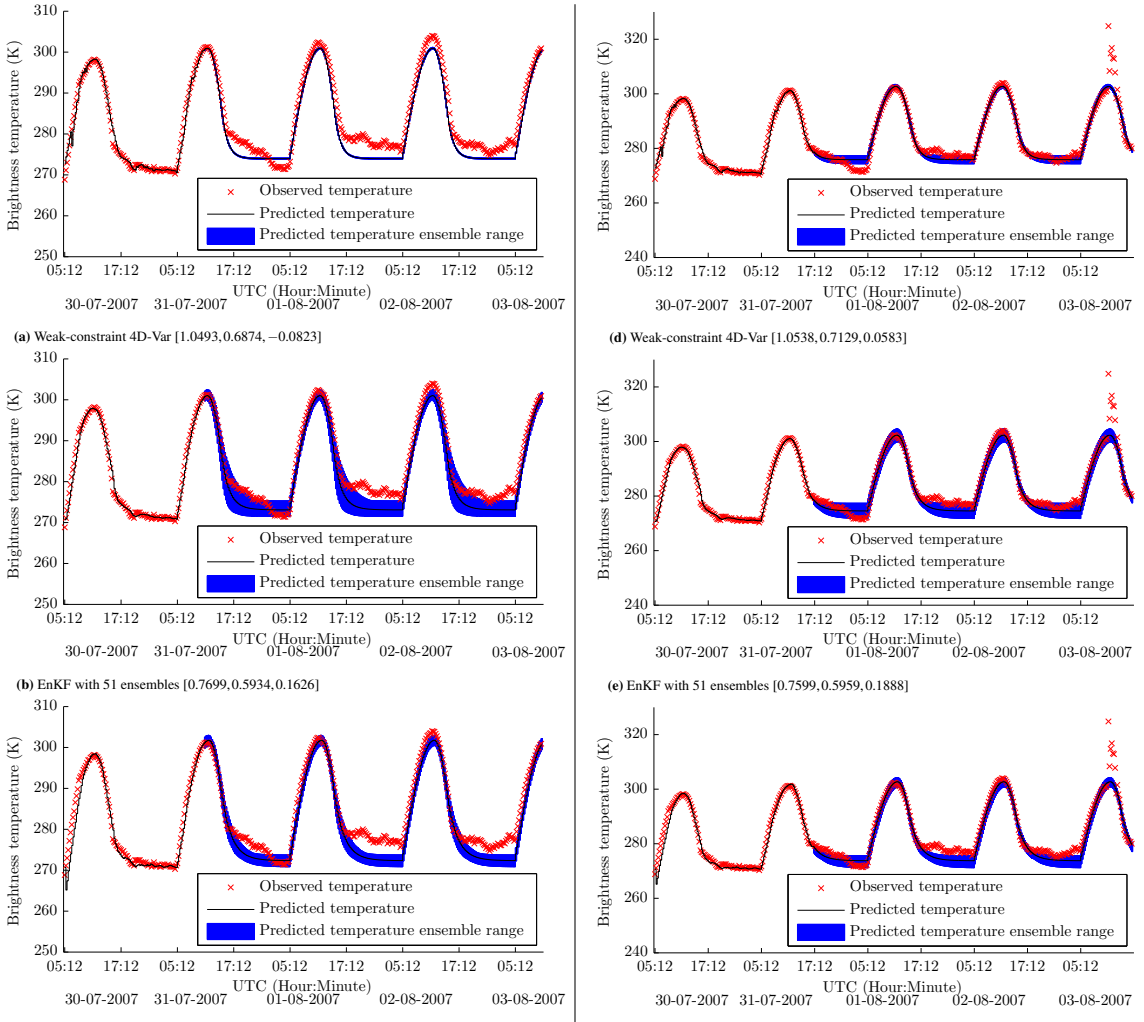
$$\boldsymbol{x}_{k,i}^a = \boldsymbol{x}_k^a + \frac{\boldsymbol{p_i}}{\|\boldsymbol{p_i}\|_{\boldsymbol{W}}} \tag{S28}$$

The number of ensemble members used in the forecast is $N_e = 51$ when the initial condition of the forecast is determined using either 4D-Var, EnKF or SIR particle filter. Though the accuracy or skill of the forecasts also

depends on the number of ensemble members, 51 members are not optimal but sufficient for fulfilling the goal of this initial study while preserving at a feasible level the computational complexity of the algorithms. The ensemble forecasting is implemented pixel-wise, where for each pixel location, 51-member ensemble forecasts are generated independently of the state and observations of its neighbor pixels. Each of the 51 ensemble members is integrated with no data assimilation using the same model configuration (i.e., $\boldsymbol{x}^f_{k+1,i} = f_k(\boldsymbol{x}^a_{k,i})$), starting from the analysis time up until the next *non-fire* brightness temperature is observed, when the forecast is then updated. Predicted values of model variables (free DTC parameters) are produced from the analysis time onwards at each MSG observation times (i.e., 15-minute forecasts for a variable length of time). The shortest-range ensemble forecasts of brightness temperature verifying at time step k in a pixel location (i.e., $y^f_{1,k}$, one of the two elements of the observational forecast computed from the integrated model) represents the distribution of the background (*non-fire*) temperature of the pixel location at time step k, and the mean of this ensemble is the expected value of the background (*non-fire*) temperature of the pixel location at time step k. The expected forecast thermal sunrise $t^f_{tsr,k}$, given the sunrise time of a day $t_{sr,k}$ is given by

$$t^f_{tsr,k} = t_{sr,k} + \frac{1}{N_e}\frac{1}{N_l}\sum_{i=1}^{N_e}\sum_{k=n[t^a_k]+1}^{n[t^e_{tsr,k}]-1} h_{2,k}(\boldsymbol{x}^f_{k,i}) \tag{S29}$$

where $n[\cdot]$ rounds the time to the closest discrete time step from the launch of the simulation, $t^e_{tsr,k} = t_{sr,k} + c_k$ where $t_{sr,k}$ is the incoming sunrise, and c_k is the mean of the recent observed 7 offsets c_k's of the thermal sunrise from the sunrise (including the offset of the current DTC). N_l is the interval length from the analysis time t^a_k to $t^e_{tsr,k}$ (i.e., $N_l = n[t^e_{tsr,k}] - n[t^a_k] - 1$), and N_e is the number of ensemble members. $n[t^f_{tsr,k}]$ is set to $n[t^a_k] + 1$ when $n[t^f_{tsr,k}] < n[t^a_k] + 1$, and set to infinity ($\infty$) when $t^f_{tsr,k} > t^f_k$ (where t^f_k is the forecast verifying time). Figure S1 shows an example of forecasts, when the initial condition of the forecast is derived using the weak constraint 4D-Var with a data assimilation window of length $N_w = 6$ (in the main article, referred to as the 4D-Var-CD$_{(N_w=6)}$ method), the EnKF with number of ensemble $N_e = 51$ (in the main article, referred to as the EnKF-CD$_{(N_e=51)}$ method), and the SIR particle filter with number of particles $N_p = 51$ (in the main article, referred to as the SIR-CD$_{(N_p=51)}$ method). The example displays the ensemble range (lower and upper bounds of the ensemble members) of a quarter-hourly ensemble forecast of brightness temperatures for a forecast window of 3 days. Fire is reported by EUMETSAT FIR product in this example on 3 August 2007 in an MSG pixel location whose center is at $(28.3699^\circ\,\mathrm{S}, 30.3394^\circ\,\mathrm{E})$. This fire that can be observed by visual inspection is reported by EUMETSAT FIR as five fire events, all with high probability (i.e., *probable* fire events), at 11:15 a.m., 11:45 a.m., 12:00 p.m., 12:15 p.m. and 12:30 p.m.



(a) Weak-constraint 4D-Var [1.0493, 0.6874, −0.0823]

(b) EnKF with 51 ensembles [0.7699, 0.5934, 0.1626]

(c) SIR with 51 particles [1.7858, 1.0873, 0.8797]

Example of ensemble generation/forecast for a last analysis at 10:42 a.m. on 31 July 2007 for a Level 1.5 image pixel at $(898, 951)$ whose center is at $(28.3699^\circ\,\mathrm{S}, 30.3394^\circ\,\mathrm{E})$

(d) Weak-constraint 4D-Var [1.0538, 0.7129, 0.0583]

(e) EnKF with 51 ensembles [0.7599, 0.5959, 0.1888]

(f) SIR with 51 particles [1.6535, 1.0044, 0.7246]

Example of an ensemble generation/forecast for a last analysis at 04:42 p.m. on 31 July 2007 for a Level 1.5 image pixel at $(898, 951)$ whose center is at $(28.3699^\circ\,\mathrm{S}, 30.3394^\circ\,\mathrm{E})$.

Figure S1. Example of ensemble forecast from 4D-Var-CD$_{(N_w=6)}$, EnKF-CD$_{(N_e=51)}$ and SIR-CD$_{(N_p=51)}$ analyses. Quantities in square brackets are, respectively, the RMSE, the MAE and the bias between the mean of one-step ahead forecasts and the observations from the launch time of the simulation (i.e., 5:12 a.m. on 30 July 2007) to the time of the last analysis cycle when all observations in this interval have been sequentially assimilated.

Abbreviations

The following abbreviations are used in the Supplementary Materials:

3D-Var	Three-Dimensional Variational Assimilation
4D-Var	Four-Dimensional Variational Assimilation
DTC	Diurnal Temperature Cycle
EnKF	Ensemble Kalman Filter
IS	Importance Sampling
MAE	Mean Absolute Error
MC	Monte Carlo
MSG	METEOSAT Second Generation
NAVDAS-AR	Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System - Accelerated Representer
NCEP	National Centers for Environmental Prediction
NMC	National Meteorological Center
pdf	probability density function
RMSE	Root Mean Squared Error
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SIR	Sampling Importance Resampling
SIS	Sequential Importance Sampling
TOA	Top Of the Atmosphere

References

1. van den Bergh, F.; van Wyk, M.A.; van Wyk, B.J. A comparison of data-driven and model-driven approaches to brightness temperature diurnal cycle interpolation. In Proceedings of the 17th Symposium of the Pattern Recognition Association of South Africa, Parys, South Africa, 29 November – 1 December 2006; pp. 252–256.
2. Meeus, J. *Astronomical Algorithms*; Willman-Bell: Richmond, VA, USA, 1999.
3. Daley, R. *Atmospheric Data Analysis*; Cambridge Atmospheric and Space Science Series; Cambridge University Press: Cambridge, UK, 1991.
4. Harms, D.E.; Raman, S.; Madala, R.V. An examination of four-dimensional data-assimilation techniques for numerical weather prediction. *Bull. Am. Meteorol. Soc.* **1992**, *73*, 425–440. doi:10.1175/1520-0477(1992)073<0425:AEOFDD>2.0.CO;2.
5. Grell, G.A.; Dudhia, J.; Stauffer, D.R. *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*; Technical Report NCAR/TN-398+STR; National Center for Atmospheric Research: Boulder, CO, USA, 1994.
6. Derber, J.C. A variational continuous assimilation technique. *Mon. Weather Rev.* **1989**, *117*, 2437–2446. doi:10.1175/1520-0493(1989)117<2437:AVCAT>2.0.CO;2.
7. Courtier, P.; Thépaut, J.N.; Hollingsworth, A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **1994**, *120*, 1367–1387. doi:10.1002/qj.49712051912.
8. Li, Z.; Navon, I.M. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.* **2001**, *127*, 661–683. doi:10.1002/qj.49712757220.
9. Cao, Y.; Zhu, J.; Navon, I.M.; Luo, Z. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Methods Fluids* **2007**, *53*, 1571–1583. doi:10.1002/fld.1365.
10. Sasaki, Y. Some basic formalisms in numerical variational analysis. *Mon. Weather Rev.* **1970**, *98*, 875–883. doi:10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2.
11. Trémolet, Y. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 2483–2504. doi:10.1256/qj.05.224.
12. Rosmond, T.; Xu, L. Development of NAVDAS-AR: non-linear formulation and outer loop tests. *Tellus Dyn. Meteorol. Oceanogr.* **2006**, *58*, 45–58. doi:10.1111/j.1600-0870.2006.00148.x.
13. Chua, B.S.; Xu, L.; Rosmond, T.; Zaron, E.D. Preconditioning representer-based variational data assimilation systems: Application to NAVDAS-AR. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*; Park, S.K., Xu, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 307–319. doi:10.1007/978-3-540-71056-1_16.
14. Notay, Y. Flexible conjugate gradients. *SIAM J. Sci. Comput.* **2000**, *22*, 1444–1460. doi:10.1137/S1064827599362314.
15. Pires, C.; Vautard, R.; Talagrand, O. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus* **1996**, *48*, 96–121. doi:10.1034/j.1600-0870.1996.00006.x.
16. Xu, L.; Rosmond, T.; Daley, R. Development of NAVDAS-AR: formulation and initial tests of the linear problem. *Tellus Dyn. Meteorol. Oceanogr.* **2005**, *57*, 546–559. doi:10.3402/tellusa.v57i4.14710.

17. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **1994**, *99*, 10143–10162. doi:10.1029/94JC00572.
18. Houtekamer, P.L.; Mitchell, H.L. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **1998**, *126*, 796–811. doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
19. Burgers, G.; van Leeuwen, P.J.; Evensen, G. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **1998**, *126*, 1719–1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.
20. Kalnay, E.; Li, H.; Miyoshi, T.; Yang, S.C.; Ballabrera-Poy, J. 4-D-Var or ensemble Kalman filter? *Tellus* **2007**, *59A*, 758–773. doi:10.1111/j.1600-0870.2007.00261.x.
21. Houtekamer, P.L.; Mitchell, H.L. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **2001**, *129*, 123–137. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
22. Lorenc, A.C. The potential of the ensemble Kalman filter for NWP–a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3183–3203. doi:10.1256/qj.02.132.
23. Hamill, T.M.; Snyder, C. A hybrid ensemble Kalman filter—3D variational analysis scheme. *Mon. Weather Rev.* **2000**, *128*, 2905–2919. doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.
24. Houtekamer, P.L.; Mitchell, H.L.; Deng, X. Model error representation in an operational ensemble Kalman filter. *Mon. Weather Rev.* **2009**, *137*, 2126–2143. doi:10.1175/2008MWR2737.1.
25. Gordon, N.J.; Salmond, D.J.; Smith, A.F.M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE-Proc.-F (Radar Signal Process.)* **1993**, *140*, 107–113. doi:10.1049/ip-f-2.1993.0015.
26. Doucet, A.; Godsill, S.J.; Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **2000**, *10*, 197–208. doi:10.1023/A:1008935410038.
27. Doucet, A.; Johansen, A.M. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. In *The Oxford Handbook of Nonlinear Filtering*; Crisan, D., Rozovskii, B., Eds.; Oxford University Press: Oxford, UK, 2011; pp. 656–704.
28. Carpenter, J.; Clifford, P.; Fearnhead, P. Improved particle filter for nonlinear problems. *IEEE Proc. Radar Sonar Navig.* **1999**, *146*, 2–7. doi:10.1049/ip-rsn:19990255.
29. Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **2002**, *50*, 174–188. doi:10.1109/78.978374.
30. Hol, J.D.; Schön, T.B.; Gustafsson, F. On resampling algorithms for particle filters. In Proceedings of the IEEE Nonlinear Statistical Signal Processing Workshop (NSSPW), Cambridge, UK, 13–15 September 2006. doi:10.1109/NSSPW.2006.4378824.
31. Sileshi, B.G.; Ferrer, C.; Oliver, J. Particle filters and resampling techniques: Importance in computational complexity analysis. In Proceedings of the Conference on Design and Architectures for Signal and Image Processing (DASIP), Cagliari, Italy, 8 – 10 October 2013.
32. Douc, R.; Cappé, O.; Moulines, E. Comparison of resampling schemes for particle filtering. In Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005), Zagreb, Croatia, 15–17 September 2005; pp. 64–69. doi:10.1109/ISPA.2005.195385.
33. Chopin, N. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Stat.* **2004**, *32*, 2385–2411. doi:10.1214/009053604000000698.
34. Del Moral, P.; Jacod, J. Interacting particle filtering with discrete observations. In *Sequential Monte Carlo methods in practice*; Doucet, A., de Freitas, N., Gordon, N., Eds.; Statistics for Engineering and Information Science; Springer: New York, NY, USA, 2001; pp. 43–75. doi:10.1007/978-1-4757-3437-9_3.
35. Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*; Eldar, Y.C., Kutyniok, G., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 210–268. doi:10.1017/cbo9780511794308.006.
36. Julier, S.J.; Uhlmann, J.K. New extension of the Kalman filter to nonlinear systems. In Proceedings of the Signal processing, sensor fusion, and target recognition VI. International Society for Optics and Photonics, Orlando, FL, USA, 21 – 25 April 1997, Volume SPIE 3068; pp. 182–193. doi:10.1117/12.280797.
37. Julier, S.J.; Uhlmann, J.K. Unscented filtering and nonlinear estimation. *Proc. IEEE* **2004**, *92*, 401–422. doi:10.1109/JPROC.2003.823141.
38. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x.
39. Bannister, R.N. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **2008**, *134*, 1971–1996. doi:10.1002/qj.340.

40. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.
41. Kalnay, E. *Atmospheric Modeling, Data Assimilation, and Predictability*; Cambridge University Press: Cambridge, UK, 2003.
42. Leutbecher, M.; Palmer, T.N. Ensemble forecasting. *J. Comput. Phys.* **2008**, *227*, 3515–3539. doi:10.1016/j.jcp.2007.02.014.
43. Toth, Z.; Kalnay, E. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Meteorol. Soc.* **1993**, *74*, 2317–2330. doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.
44. Toth, Z.; Kalnay, E. Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **1997**, *125*, 3297–3319. doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.
45. Moore, A.M.; Arango, H.G.; Broquet, G. Estimates of analysis and forecast error variances derived from the adjoint of 4D-Var. *Mon. Weather Rev.* **2012**, *140*, 3183–3203. doi:10.1175/MWR-D-11-00141.1.