



Mathematisch-Naturwissenschaftliche Fakultät

Walter Acevedo | Sebastian Reich | Ulrich Cubasch

Towards the assimilation of tree-ring-width records using ensemble Kalman filtering techniques

Suggested citation referring to the original publication:

Climate Dynamics 46 (2016) 1909–1920

DOI <https://doi.org/10.1007/s00382-015-2683-1>

ISSN (print) 0930-7575

ISSN (online) 1432-0894

Postprint archived at the Institutional Repository of the Potsdam University in:

Postprints der Universität Potsdam

Mathematisch-Naturwissenschaftliche Reihe ; 892

ISSN 1866-8372

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-436363>

DOI <https://doi.org/10.25932/publishup-43636>

Towards the assimilation of tree-ring-width records using ensemble Kalman filtering techniques

Walter Acevedo¹ · Sebastian Reich² · Ulrich Cubasch¹

Received: 1 July 2014 / Accepted: 22 May 2015 / Published online: 10 June 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract This paper investigates the applicability of the Vaganov–Shashkin–Lite (VSL) forward model for tree-ring-width chronologies as observation operator within a proxy data assimilation (DA) setting. Based on the principle of limiting factors, VSL combines temperature and moisture time series in a nonlinear fashion to obtain simulated TRW chronologies. When used as observation operator, this modelling approach implies three compounding, challenging features: (1) time averaging, (2) “switching recording” of 2 variables and (3) bounded response windows leading to “thresholded response”. We generate pseudo-TRW observations from a chaotic 2-scale dynamical system, used as a cartoon of the atmosphere-land system, and attempt to assimilate them via ensemble Kalman filtering techniques. Results within our simplified setting reveal that VSL’s nonlinearities may lead to considerable loss of assimilation skill, as compared to the utilization of a time-averaged (TA) linear observation operator. In order to understand this undesired effect, we embed VSL’s formulation into the framework of fuzzy logic (FL) theory, which thereby exposes multiple representations of the principle of limiting factors. DA experiments employing three alternative growth rate functions disclose a strong link between the lack of smoothness of the growth rate function and the loss of optimality in the estimate of the TA state.

Accordingly, VSL’s performance as observation operator can be enhanced by resorting to smoother FL representations of the principle of limiting factors. This finding fosters new interpretations of tree-ring-growth limitation processes.

Keywords Proxy forward modeling · Data assimilation · Fuzzy logic · Ensemble Kalman filter · Paleoclimate reconstruction

1 Introduction

In recent years the assimilation of proxy data into climate models has emerged as a very promising way to bring physical consistency to paleoclimate reconstructions, as well as to reduce their uncertainty levels. “High resolution” proxy records (Hughes and Ammann 2009) offer information about inter-annual and slower internal climate variabilities which in principle could be constrained by a suitable DA scheme. As a result, several research groups have been actively working with the purpose of paving the way towards an eventual paleo-reanalysis (Hughes et al. 2010; Guiot et al. 2009; Brönnimann 2011).

To date, several very diverse DA schemes have been tested on climate models—such as pattern nudging (von Storch et al. 2000), forcing singular vectors (Barkmeijer et al. 2003), adjoint method (Kurahashi-Nakamura et al. 2014), particle filter (Dubinkina and Goosse 2013; Mathiot et al. 2013) and ensemble Kalman filter (Bhend et al. 2012; Pendergrass et al. 2012)—providing very encouraging results regarding the potential constraint of the inter-annual internal variability. As for the observations employed in these experiments, the dominant approach has been to use either pseudo-proxies or statistically reconstructed climate

This research work was produced within the Helmholtz graduate research school GeoSim.

✉ Walter Acevedo
walter.acevedo@met.fu-berlin.de

¹ Institut für Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany

² Institut für Mathematik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

time series, both considering a univariate linear relationship between climate state and climate-driven proxy signals.

Nowadays, the many developments in the booming field of high-resolution paleoclimatology (Hughes and Ammann 2009) have made evident the complexity of climate proxy systems. The formation of a proxy record might comprise biological, physical, and chemical mechanisms, each of these able to introduce non-linear processes. Accordingly, there is a growing interest in adopting more realistic approaches regarding the relation between proxy data and climate forcing (Evans et al. 2013). One of the most promising approaches in this direction is “Proxy Forward Modeling” (Hughes et al. 2010; Evans et al. 2013), which takes climate forcing as input data and generate artificial proxy records that can be directly compared to actual proxy records. This strategy is diametrically opposed to the traditional inverse approach where climate conditions are directly inverted from proxy data.

Proxy forward models can be used for different purposes such as model-paleodata comparison in the proxy space (Evans et al. 2013) and prediction of future evolution of proxy archives (Vaganov et al. 2006). On the other hand, despite its forth direction, forward models can also be used for reconstruction purposes by resorting to probabilistic inversion strategies, such as Bayesian hierarchical modeling (Tolwinski-Ward et al. 2014), Markov Chain Monte Carlo method (Boucher et al. 2014) and DA (Hughes et al. 2010). Contributing to the development of the latter approach is the objective of the present paper.

Different proxy data kinds react in distinctive ways to specific sets of climate variables. Accordingly, the assimilation of a given proxy type should be pursued on its own merits, using a forward model able to simulate the climate recording processes specific to that particular kind of proxy record. Nowadays, there exist forward models for many different proxy types, most notably TRW (Vaganov et al. 2006; Evans et al. 2006; Tolwinski-Ward et al. 2011), tree-ring isotopes (Roden et al. 2000; Danis et al. 2012; Evans 2007), coral isotopes (Thompson et al. 2011), ocean sediments (Heinze 2001; Schmidt 1999) and stalagmite isotopes (Baker et al. 2012). There are also initiatives to model the transport of isotopes within the atmosphere (Sturm et al. 2010). Depending on the particular application and the availability of data, proxy forward models assume different complexity levels which go from the minimalistic linear pseudo-proxies to comprehensive models that simulate the proxy generation process as realistically as possible.

Following this train of thought, we will investigate the role of the observation operator in the assimilation of the most traditional climate proxy: TRW. This proxy record type is particularly suitable for DA purposes given that it exhibits the largest spatial coverage, albeit strongly biased towards the Northern hemisphere, as well as one of the

highest and most stable temporal resolutions, among the different paleoclimate proxies currently available. Additionally, tree-ring growth forward models have reached a rather mature state of development. In particular within climate applications, there exists a sufficiently realistic yet affordable forward model for TRW chronologies called Vaganov–Shashkin–Lite (VSL) (Tolwinski-Ward et al. 2011, 2013). This model has been shown to skillfully simulate the climate-driven tree-ring growth for very diverse species and climate regimes, both at continental (Tolwinski-Ward et al. 2011) and global scales (Breitenmoser et al. 2014). With that it presents itself as a perfect candidate for a TRW observation operator within a paleo-DA setting. Despite the aforementioned assets, VSL still presents several challenging aspects regarding DA, most notably:

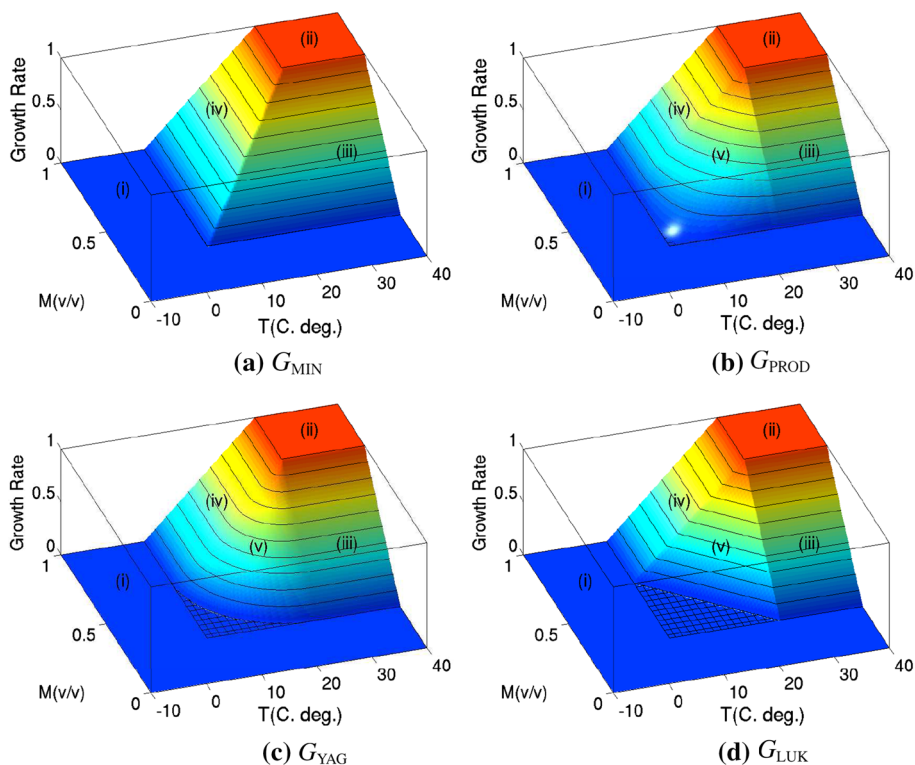
1. Time Averaging: TRW is the accumulated tree ring growth during the entire growth season.
2. “Thresholded Response” (Evans et al. 2013): trees become insensitive to climate variability during dormancy and optimal growth.
3. Abrupt shifting of recorded variable: Growth is limited by either temperature or moisture, hence transitions between limitation regimes happen suddenly.

We claim that the last of these hindrances is not a necessary component of the principle of limiting factors but rather a consequence of the specific formulation of the VSL model. This can be naturally improved for DA purposes within the framework of FL theory.

FL mimics the kind of reasoning achieved by human language, and by doing so it provides a mathematical framework where definite conclusions can be drawn from imprecise data and vague knowledge of the underlying mechanisms. Accordingly, FL constitutes a powerful modeling approach in areas dealing with complex systems. Since its introduction in the 1960s (Zadeh 1975), FL has greatly influenced many applied disciplines, most notably control theory (Nguyen et al. 2002). Within the environmental sciences, FL has also found numerous applications, including ecological and hydrological modeling (Marchini 2011; Salski 2006; Se 2009). Regarding climate proxy forward modeling, however, FL theory has not yet been explicitly employed. As we will show below, several of the tree growth modeling concepts have direct “fuzzy” analogues and in particular, VSL can be completely reinterpreted in terms of FL.

The purpose of this paper is then to assess the impact of VSL’s nonlinear aspects within a simplified, coupled DA setting as a stepping stone towards more realistic experiments involving atmospheric and coupled climate models. We will describe the TRW forward model VSL (Sect. 2.1), introduce the necessary concepts of FL

Fig. 1 Growth rate surfaces for different representations of the fuzzy intersection operation. **a** minimum, **b** product, **c** Yager and **d** Lukasiewicz t-norms. Growth regimes: (i) dormancy, (ii) optimal, (iii) moisture limited, (iv) temperature limited and (v) temperature-moisture limited. Crossed hatched areas correspond to additional dormancy conditions introduced by the t-norm used. (Threshold values: $T_1 = 5$ C.deg, $T_2 = 25$ C.deg, $M_1 = 0.3$ v/v and $M_2 = 0.70$ v/v)



(Sect. 2.2) and examine VSL from the perspective of FL (Sect. 2.3). Afterwards, we present the driving dynamical system (Sect. 2.4), the DA method considered (Sect. 2.5) and define the experimental setting (Sect. 2.6). The results are shown in Sect. 3, followed by a discussion in Sect. 4.

2 Material and methods

2.1 Tree-ring-width (TRW) forward model

The VSL model for TRW chronologies offers an intermediate complexity approach between ecophysiological and completely data-driven models (Tolwinski-Ward et al. 2011; Tolwinski-Ward 2012), where the climate-driven component of tree-ring growth is parametrized by way of a simple representation of the “principle of limiting factors” (Fritts 1976). This biological concept states that the pace at which a plant develops is controlled by the single basic growth resource, typically either energy or water, that is in shortest supply. Within VSL, the limiting factors relevant for tree-ring growth are near-surface air temperature T and soil moisture M . Their impacts on tree development are accounted for via the “growth response” functions g_T and g_M . Making use of the piece-wise linear “standard ramp” function (Tolwinski-Ward et al. 2014)

$$\Psi(u) = \begin{cases} 0 & \text{if } 0 \geq u \\ u & \text{if } 0 < u \leq 1 \\ 1 & \text{if } u > 1, \end{cases}$$

VSL’s growth responses at a particular time¹ can be expressed as:

$$g_T = \Psi\left(\frac{T - T^L}{T^U - T^L}\right) \tag{1}$$

and

$$g_M = \Psi\left(\frac{M - M^L}{M^U - M^L}\right). \tag{2}$$

Here T^L and M^L denote lower thresholds below which there is no growth and trees are said to be dormant, while T^U and M^U designate upper thresholds above which tree growth is optimal. Given a pair of g_T and g_M values, the corresponding growth rate G_{MIN} is determined by the smallest growth response, i.e.,

$$G_{MIN} = \min\{g_T, g_M\}, \tag{3}$$

¹ VSL computes monthly response functions from monthly T and M time series. In our experiments we use a dynamical system whose time variable is non-dimensional and therefore we do not make explicit mention of the time resolution in this paper.

which defines a non-smooth surface in the $T - M$ plane, displaying 4 distinct growth regimes: (1) dormancy, (2) optimal growth, (3) temperature limited and (4) moisture limited (see Fig. 1a). Afterwards, VSL generates the TRW for each year via the yearly time integration of the growth rate function modulated by the relative local insolation I .

$$\text{TRW}_n = \int_{t_n - \tau}^{t_n} G_{\text{MIN}}(t)I(t) dt. \quad (4)$$

Finally, simulated TRW series are standardized so as to allow for comparison against actual TRW chronologies, which are given as non-dimensional index series.

For climate forcing trajectories confined within either temperature- or moisture-limited zones (see Fig. 1a), VSL's formulation reduces to an univariate linear approach. In this case a single input variable is imprinted in the final TRW chronology, as opposed to the general case where tree growth is limited by T and M in a fluctuating fashion resulting in a mixed chronology. This "response-switching" mechanism (Tolwinski-Ward et al. 2014) presents a challenge to inverse problem applications of VSL. As a matter of fact, in the context of simultaneous temperature-moisture reconstructions via Bayesian Hierarchical modeling, it has been linked to non-Gaussian, bimodal posterior probability densities (Tolwinski-Ward et al. 2014).

2.2 Basic fuzzy logic concepts

As opposed to the classical Boolean logic, where a proposition can be either true or false, FL allows for partial truth values between 1 (completely true) and 0 (completely false), and then it provides a rigorous mathematical foundation for the implementation of approximate reasoning systems (Zadeh 1975).

Assuming U as the set of all elements or conditions considered in a particular application, e.g., all the climate conditions a tree can be subjected to, a classical set A in U can be defined by a membership function μ_A that assigns to each element of U a number from the set $\{0, 1\}$. Here 0 (1) represents non-membership (membership) to the set A . This definition can be rephrased in classical logic terms by interpreting the value $\mu_A(x)$ as the truth value of the proposition "x is in A", where x denotes a generic element of U . Fuzzy set theory generalizes the classical one by extending the range of membership functions to the real interval $[0, 1]$. This modification allows the existence of elements with partial membership: $\{x : 0 < \mu_A(x) < 1\}$, for which the proposition "x is in A" is partially true (Zadeh 1965).

Fuzzy membership functions can adopt very different shapes depending on the particular application. Piecewise-linear forms—such as ramps, triangles and trapezoids—are frequently used because of their simplicity in terms of implementation and interpretation. On the other hand,

when smoothness is a concern, membership functions with continuous derivative are preferred, e.g., Gaussian and sigmoidal functions.

Given two fuzzy sets A and B , the intersection set $A \wedge B$ (or equivalently the truth value of the composed statement "x is in A" AND "x is in B") can be found by the following expression:

$$\mu_{A \wedge B} = T(\mu_A, \mu_B) \quad (5)$$

where T is a function, technically referred to as "t-norm" (Nguyen et al. 2002), satisfying the following requirements:

Commutativity: $T(a, b) = T(b, a)$,

Associativity: $T(a, T(b, c)) = T(T(a, b), c)$,

Monotonicity: $T(a, b) \leq T(c, d)$ if $a \leq c$ and $b \leq d$, (6)

Number 1 is an identity element : $T(a, 1) = a$,

which determine the admissible representations of the fuzzy intersection, or equivalently the fuzzy AND operation.

T-norms constitute a rather ample class of functions, from which the first and most popular example is the minimum function, also called Gödel t-norm (Zadeh 1965). Abundant types of t-norms have been studied (Nguyen et al. 2002) and, similarly to membership functions, the selection of the particular t-norm depends heavily on the problem considered.

2.3 VSL from the fuzzy set perspective

Bearing in mind the concepts introduced in Sect. 2.2, VSL's growth response g_T (g_M) can be understood as a membership function defining the fuzzy set S_T (S_M) of climate states whose temperature (moisture) values allow tree growth. Furthermore, the growth rate G_{MIN} can be viewed as a t-norm that determines the fuzzy intersection set $S_{T \wedge M}$ in the $T - M$ plane, or analogously the degree of truth of the statement " T and M conditions allow tree growth" at a particular time.

Therefore, we claim that VSL's formulation of the principle of limiting factors—which is inherited from the full Vaganov–Shashkin (VS) model (Vaganov et al. 2006)—essentially determines the fuzzy set of temperature-moisture conditions that are favorable for tree growth. The final two steps of VSL's TRW calculation, i.e., insolation modulation and time integration, can also be translated to fuzzy terms as examples of aggregation and defuzzification methods (Nguyen et al. 2002). Furthermore, the whole mathematical formulation of the VSL model can be viewed as a zero-order Sugeno-type fuzzy inference system (Sugeno 1985) with $2m$ arguments (monthly temperature and moisture values for m months) and one output (tree-ring width).

Regarding the application of VSL as observation operator within a DA setting, we find the use of a minimum t-norm

problematic due to the consequent exclusive growth limitation. Given that the growth rate at an specific time is only determined by the most limiting factor, the transitions between growth-limited regimes necessarily take place in a sharp manner. This abrupt response-switching mechanism, as we will show below, can significantly deteriorate the performance of an Ensemble Kalman Filter. Consequently, we make use of the freedom we have to select the fuzzy intersection representation and study three alternative t-norms (see Fig. 1):

1. the algebraic product:

$$G_{\text{PROD}} = g_T \cdot g_M, \quad (7)$$

2. Yager t-norm of second order:

$$G_{\text{YAG}} = \max \left\{ 0, ((1 - g_T)^2 + (1 - g_M)^2)^{1/2} \right\}, \quad (8)$$

3. Lukasiewicz t-norm:

$$G_{\text{LUK}} = \max \{ 0, g_T + g_M - 1 \}. \quad (9)$$

All of these new representations of the limiting factor's principle exhibit an additional regime where T and M concurrently limit tree-ring growth (see zones (v) in Fig. 1), which allows a gradual transition between temperature- and moisture-limited modes and thus a progressive alternation of the recorded variable. In the case of G_{YAG} this transition is perfectly smooth while for G_{PROD} there are still two subtle edges in the growth rate surface (see Fig. 1b). Nonetheless, the product t-norm offers high stability to input variable errors and membership function selection, given that it exhibits the smallest average sensitivity among all t-norms (Nguyen et al. 2002). Finally, for the Lukasiewicz t-norm, the temperature-moisture limited growth area presents a perfectly linear dependence, unfortunately at the expense of increasing the sharpness of the above mentioned edges.

One important feature of the growth rate surfaces obtained using Yager and Lukasiewicz t-norms is the appearance of new points in the $T - M$ plane leading to null growth rates, despite none of their corresponding growth responses is null (see Fig. 1c, d). This new dormancy area, which is due to the presence of the maximum function in Eqs. 8 and 9, constitutes a source of additional lower response thresholding, as we will show below.

2.4 Dynamical system

Soil processes are known to have a significantly longer memory than atmospheric ones (Wanru and Dickinson 2004) and, as a consequence their corresponding variables present significantly different time scales. In order to mimic this multi-scale environmental driving on tree

growth, we utilize the 2-scale coupled model introduced by Lorenz (1996), given by the following nonlinear equations:

$$\begin{aligned} \dot{T}_i &= T_{i-1}(T_{i+1} - T_{i-2}) - T_i - \frac{hc}{b} \sum_{j=1}^n M_{j,i} + F, \\ \dot{M}_{j,i} &= cbM_{j+1,i}(M_{j-1,i} - M_{j+2,i}) - cM_{j,i} + \frac{hc}{b} T_i, \end{aligned} \quad (10)$$

where $i = 1, \dots, m$, $j = 1, \dots, n$, F is a constant forcing term, c and b are factors controlling the time scale and amplitude of the component M , and h is the coupling constant. Additionally, T and M variables are assumed to be periodic, i.e., $T_1 = T_{n+1}$; $M_1 = M_{n+1}$. Model state trajectories are obtained numerically via the standard Runge-Kutta method of fourth order with time step $\Delta t = 0.01$, where the model variables are represented by 64-bit FORTRAN real variables. The results shown in this paper correspond to the parameter setting $m = 40$, $n = 1$, $F = 8.0$, $c = 0.5$, $b = 1.0$ and $h = 1.0$, which is representative of our whole numerical exploration. For these parameter values, the faster component T oscillates with roughly 2.5 times the amplitude of the slow component M . Moreover, the predictability horizon for instantaneous quantities is roughly four times longer in the slow component than in the fast one (see Fig. 2a). This configuration is meant to resemble the typical ratios of predictability time scale and amplitude between near-surface air temperature and soil moisture variables. Notice that the use of the letters T and M to denote fast and slow variables is made only for the sake of notation simplicity and there is no actual relationship with the details of atmosphere-land interaction.

2.5 Data assimilation method

In order to reconstruct the model state, we make use of the Ensemble Kalman Filter (EnKF) (Burgers et al. 1998). Providing a complete account of the EnKF formulation is beyond the scope of the present paper, thus we only describe the main aspects of the technique and refer the reader to Hamill (2006) for the details. EnKF relies on the assumption that the Probability Density Function (PDF) of the model state is a Gaussian distribution, which can be represented by a collection of model states, known as an ensemble. The best estimate of the state is given by the ensemble mean and the ensemble spread is used to assess the uncertainty of the estimate. EnKF estimation proceeds in cycles comprising two steps:

1. Forecast step: an initial ensemble, which represents the current knowledge of the state, is propagated in time using the model dynamics to generate a forecast ensemble.

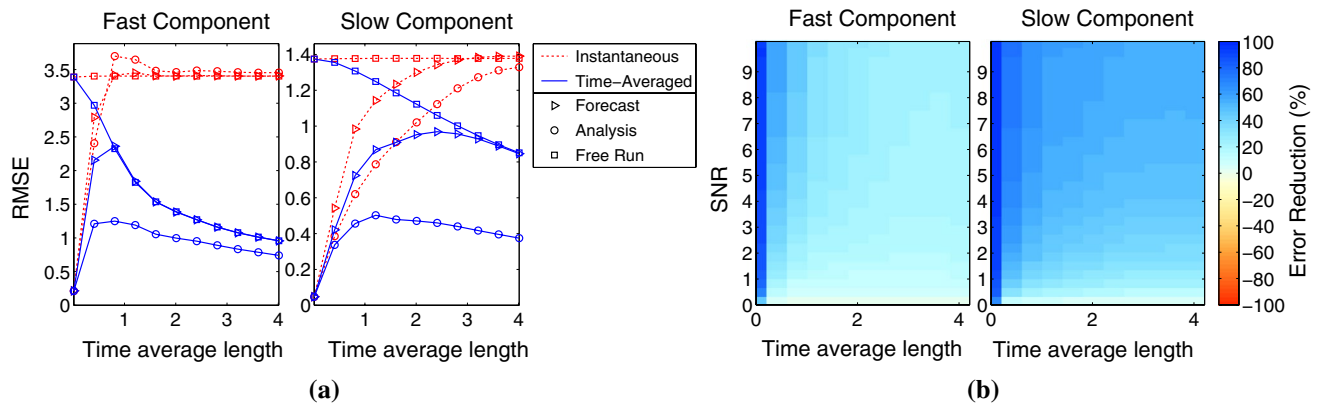


Fig. 2 Error statistics for DA experiments assimilating TA linear observations. **a** RMS error vs averaging period τ for SNR = 10. Free Run curves correspond to an ensemble run without DA. **b** TA Analysis Error Reduction (regarding a free run) versus τ and signal-to-noise ratio

2. Analysis step: available observations and their corresponding uncertainties are combined probabilistically with the forecast ensemble to obtain an updated ensemble, called the analysis. The latter is used subsequently as initial condition for the next DA cycle.

In the limit of infinitely large ensembles, EnKF provides the best linear unbiased estimate of the state under the Gaussian assumption. However, in practice for high-dimensional geophysical applications only small ensembles are affordable, and then EnKF methods suffer from two typical problems: (1) underestimation of the forecast uncertainty and (2) spurious correlations between variables at points very distant from each other. These conditions are normally alleviated by way of two empirical remedies: (1) covariance inflation, to increase the spread of the ensemble, and (2) covariance localization, to discard large-range correlations, respectively.

The EnKF forecast step allows model nonlinearity given that the whole model, potentially nonlinear, is utilized to evolve the ensemble. The analysis step, in turn, is linear due to the Gaussian assumption and then its performance might be deteriorated by nonlinear effects coming from the model or the observation operator. Notice that there exist fully nonlinear DA methods, however for high-dimensional applications they are currently unwieldy and appear to remain that way in the foreseeable future. Consequently, there is a strong interest in adapting Gaussian DA techniques to work into nonlinear settings.

Finally, in order to handle the time integrated nature of TRW chronologies, we employ a variation of the EnKF algorithm, known as time-averaged EnKF, introduced in Dirren and Hakim (2005) and used subsequently in

Huntley and Hakim (2010), Pendergrass et al. (2012) and Steiger et al. (2014). This technique allows to estimate the TA component of the model state trajectory by updating the TA ensemble trajectories (see Pendergrass et al. 2012 for details).

2.6 Experimental setting

2.6.1 Nature and ensemble runs

We perform standard “perfect model” DA experiments comprising (1) a “nature” run, from which pseudo-TRW observations are produced, (2) a DA ensemble run, where the pseudo-TRW observations are assimilated, and (3) a free ensemble run, intended to serve as reference.

The nature run consists in a single model trajectory initialized from a random sample of the model climatology, which is used as the “true” model state trajectory. The DA ensemble run is obtained by performing a sequence of EnKF analysis cycles while the free ensemble run is produced by just freely evolving the ensemble under the action of the model dynamics. Both ensemble runs are initialized identically from a random sampling of the model climatology that does not include the initial condition of the nature run. The model climatology is obtained by starting a model trajectory from a random initial condition and storing the model state every 5000 time steps.

In our experiments we use ensembles with 20 members and a typical EnKF setting consisting of a stochastic analysis step (Burgers et al. 1998), multiplicative inflation between 0 and 5 %, and covariance localization using the Gaspari–Cohn function (1999) with localization radius $r = 2$, assuming a grid spacing $\Delta x = 1$.

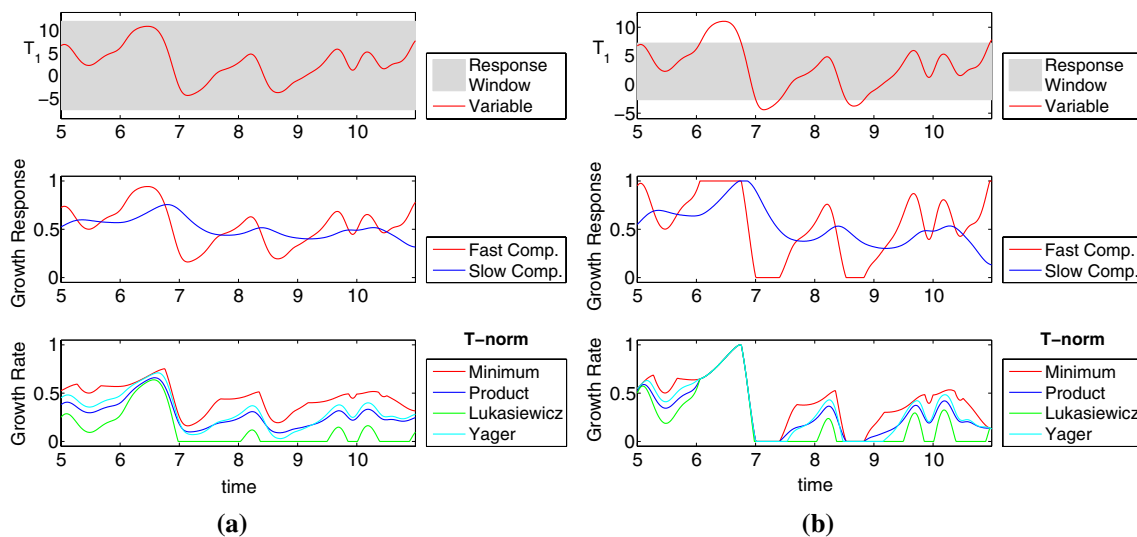


Fig. 3 Example time series of the quantities involved in the generation of pseudo-TRW observations. (For simplicity the quantities corresponding to the slow component are not shown in the upper panels). **a** Wide response windows, **b** narrow response windows

2.6.2 Observation generation

We utilize the VSL model to generate pseudo-TRW observations using the T and M time series of our nature run. To this end, we follow the steps outlined in Sect. 2.1² and add a final contamination step with unbiased Gaussian noise. We set the period between observations equal to the length of the integration period τ . Since the variance of the clean pseudo-TRW observations changes with τ , we set the variance of the noise so as to keep a constant Signal-To-Noise ratio (SNR). Our focus in this paper are the additional challenging features of VSL as an observation operator, therefore for our experiments we use optimistic observability conditions, i.e., observations at every grid point and $SNR = 10$, unless otherwise stated.

Given that our interest lies in the non-linear combination of 2 variables performed by VSL, we present results corresponding to mixed pseudo-TRW observations, obtained by setting the growth response windows to the values:

$$\begin{aligned}
 [T^L, T^U] &= [\bar{T} - R^A \sigma(T), \bar{T} + R^B \sigma(T)], \\
 [M^L, M^U] &= [\bar{M} - R^A \sigma(M), \bar{M} + R^B \sigma(M)].
 \end{aligned}
 \tag{11}$$

Here $\bar{*}$ and $\sigma(*)$ denote the mean and standard deviation values of the quantity $*$, estimated during the nature run; while R^A and R^B are constants that control the position of upper and lower response thresholds, respectively. We will focus on two particular configurations for the response

functions: (1) wide windows ($R^A = R^B = 3.0$) and (2) narrow windows ($R^A = R^B = 1.5$). In the first case, most of the climatological values of T and M are contained within the response windows and hence the response functions basically act as linear rescaling operators (see Fig. 3a). Contrarily, in the second case, T and M time series are constantly cropped by the response functions and thus the recording thresholding becomes significant (see Fig. 3b).

Additionally, with the purpose of studying the influence of the representation of the principle of limiting factors on the performance of the DA scheme, we change the original growth rate function G_{MIN} by each of the alternative functions introduced in Sect. 2.3. Furthermore, in order to have a reference that allows us to assess the loss of optimality due to the nonlinearity of the pseudo-TRW observation operators, we perform an additional ensemble run which assimilates TA linear observations, obtained with the following instantaneous observation operator:

$$H_L = g_T + g_M, \quad R^A = R^B = 3.0.
 \tag{12}$$

This setting corresponds to the wide response window configuration where the response thresholding is negligible. In these conditions, as mentioned in the previous paragraph, the response functions become linear.

2.6.3 Diagnostic statistics

Given that the impact of time averaging on the error levels depends strongly on the time scale of the model internal variability, components with different time scales are expected to react differently to the same time averaging

² Given that the forcing term of the dynamical system used is constant, we replace the insolation term I by the unity.

length. Accordingly, we analyze separately the error levels of slow and fast components by means of their overall Root-Mean-Square (RMS) errors. These diagnostic quantities are obtained by calculating the RMS error of each model variable during 5×10^4 analysis cycles, and afterwards averaging within each component. The first 5×10^3 analysis cycles of the runs are regarded as spin-up period, and therefore they are not considered in the statistics.

3 Results

3.1 Assimilation of time-averaged linear observations

3.1.1 Instantaneous quantities

Due to the low-pass filtering action of the time-averaging operator, TA observations get gradually decorrelated from the instantaneous variables as τ increases. As a consequence, errors in the estimate of the instantaneous state tend to grow by increasing τ . This process can be readily seen in the red lines of Fig. 2a, where the instantaneous forecast RMS error for the fast (slow) component converge to the free run values at around $\tau = 0.8$ ($\tau = 2.8$). Analysis errors undergo an equivalent process but at a slightly slower pace. Notice that for the fast component the analysis error curve presents a clear overshoot around $\tau = 1.0$, where the analysis skill is negative and then the DA run performs worse than the free ensemble run.

3.1.2 Time-averaged quantities

As opposed to the instantaneous state, the TA state remains correlated to the TA observations regardless of the τ value. Consequently, TA analysis error curves remain below their corresponding free run curves for much longer averaging periods than the instantaneous error curves (see Fig. 2a). Notice that the TA analysis skill is moderately sensitive to the noise level and can be considerably diminished for low SNR values (see Fig. 2b).

Regarding TA forecasting skill, despite the considerable TA analysis skill for all τ values considered, the TA forecast errors saturate and reach the free run values at around $\tau = 1.0$ ($\tau = 4.0$) for the fast (slow) component (see Fig. 2a). This situation, where a DA method presents TA analysis skill for averaging periods where the TA forecasting skill is completely lost, has been previously observed in studies applying EnKF techniques on TA quantities (Huntley and Hakim 2010; Bhend et al. 2012; Pendergrass et al. 2012; Steiger et al. 2014). DA performed under these circumstances is currently labeled as “offline”. This term is used to indicate that, under the randomizing action of

the chaotic model dynamics, at assimilation time the prior ensemble is completely decorrelated from the previous analysis ensemble. As a consequence, the observational information cannot accumulate over time, as opposed to the typical application of DA for short-range prediction.

3.2 Assimilation of pseudo-TRW observations

As expected, assimilation experiments involving pseudo-TRW observations presented higher error levels than the ones already analyzed for linear observations. The recording efficiency of the different growth rate functions considered strongly depends on the characteristics of the response windows.

3.2.1 Wide response windows

For this configuration of the response thresholds, as discussed in Sect. 2.6, g_T and g_M reduce to linear rescaling operators and therefore the only nonlinearities present in the pseudo-TRW observations are those introduced by the particular t-norm used to represent the fuzzy intersection.

Figure 3a shows how, in absence of recording thresholding coming from the response functions, G_{PROD} generates a completely smooth growth rate curve, as opposed to G_{MIN} whose corresponding line exhibits a cusp at every intersection of the growth response functions. G_{YAG} also leads to very smooth g curves, however, in this case there is already a subtle lower thresholding arising from the additional dormancy area generated by this t-norm, as discussed in Sect. 2.3. Finally, for G_{LUK} this lower response thresholding effect is remarkably stronger.

Concerning the filter performance detriment caused by the nonlinearities of pseudo-TRW observations, Fig. 4a reveals that the t-norms yielding the smoothest growth rate curves (G_{YAG} and G_{PROD}) are the ones displaying the smallest optimality losses, as measured by the error increases with regard to the ensemble run assimilating linear observations. G_{MIN} showed considerably larger error increase values, particularly for short time averaging periods, while G_{LUK} clearly presented the poorest performance, with an error increase value for the slow component of around 40 % in the offline zone.

3.2.2 Narrow response windows

For this configuration of the response thresholds, the recording thresholding becomes evident in all growth rate curves (see Fig. 3b) and accordingly the optimality losses become more pronounced (see Fig. 4b). Interestingly, G_{MIN} and G_{LUK} curves are barely affected by the addition of recording thresholding. Accordingly, the performance level

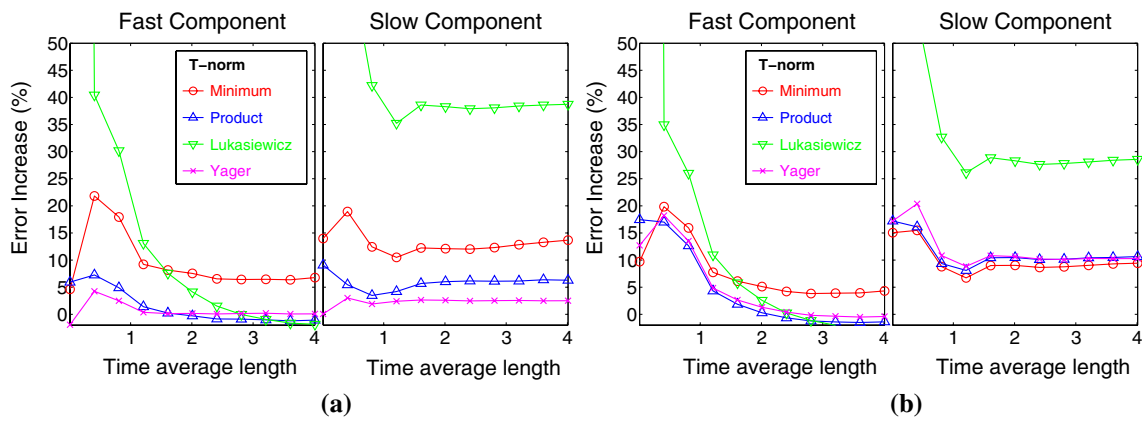


Fig. 4 TA analysis error increase (with reference to an ensemble run assimilating linear observations) versus averaging period τ . **a** Wide response windows, **b** narrow response windows

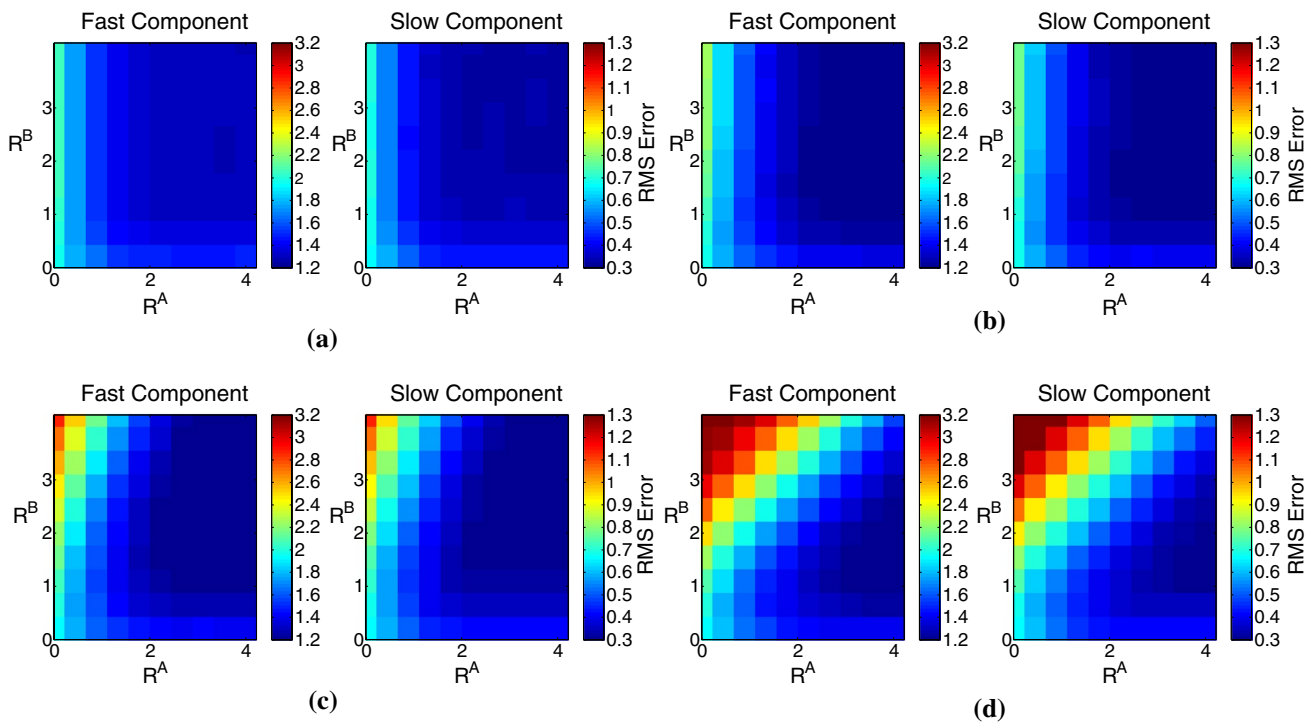


Fig. 5 Dependency of the TA analysis RMS error on the growth response thresholds (see Eq. 11) for the different growth rate functions considered ($\tau = 2.0$). **a** Minimum t-norm, **b** product t-norm, **c** Yager t-norm, **d** Lukasiewicz t-norm

of all growth rate functions considered becomes comparable, with the only exception being G_{LUK} whose operation for the slow component remains especially poor.

3.2.3 Asymmetric impact of lower and upper thresholding

The impact of the response threshold positions on the analysis skill varies strongly between different growth rate function, as shown in Fig. 5. A common feature is the

existence of high error areas for strong lower thresholding conditions, corresponding to small values of R^A . On the other hand, strong upper thresholding, corresponding to small values of R^B , has a much milder impact on the filter performance with the only exception of G_{MIN} which behaves in a very special way for the fast component, presenting its minimum error levels for T^L values around 1.4.

This markedly asymmetric effect of lower and upper response thresholds on the assimilation skill, can be

explained by the fact that the upper parts of g_T and g_M time series are considerably suppressed by the action of the fuzzy intersection operators and become scarcely imprinted in the growth rate time series.

An interesting aspect of the RMS error surfaces for G_{YAG} and G_{LUK} is the presence of a zone of large errors in the left upper corners of the plots. This behavior can be attributed to the additional dormancy area characteristic of these two t-norms, which becomes particularly sizable for large values of R^B (see Sect. 2.3).

In summary, among the four growth rate functions studied, the TA ensemble Kalman filter presented the most consistent performance for G_{PROD} . Representing the fuzzy intersection by means of the product t-norm increases the smoothness of the growth rate time series with regard to the minimum t-norm, without exhibiting the added lower response thresholding present for Yager and Lukasiewicz t-norms. Additionally, its low sensitivity to input data uncertainty (see Sect. 2.3) appears to make the product t-norm particularly robust to the response thresholding phenomenon.

4 Discussion

Our numerical experiments suggest that VSL's nonlinear, time integral formulation of TRW chronologies is compatible in general terms with the TA EnKF technique.

Concerning the switching recording of variables, implied by the principle of limiting factors, we found that it does not necessarily deteriorate the filter performance in a dramatic way. However, when the transition between growth-limited modes occurs abruptly, as in the case of the original growth rate function, the lack of smoothness of the growth rate time series may significantly increase the optimality loss in the estimation of TA model state. Adopting the interpretation of growth responses as membership functions and the minimum function as a t-norm representing a fuzzy intersection operation, it is pertinent to consider other examples of t-norms leading to alternative growth rate functions. Following this train of thought, we found that the smoothness of the particular t-norm employed plays an important role for the operation of the TA EnKF approach. This outcome is consistent with the smooth character of the trajectories of the deterministic model considered. Smooth t-norms allow a progressive switching of the recorded variables that decreases the roughness of the growth rate time series and, ultimately, benefits the filter operation. Among the alternative growth rate functions studied, the one based on the product t-norm provided the best overall performance. We attribute this primacy of the product t-norm to its smoothness and its improved stability to input variable errors and membership function selection, which makes

this t-norm particularly appealing for the observation operator application investigated here.

An important conceptual implication of this viewpoint shift is the relaxation of the growth limitation policy. Smoother t-norms allow tree growth to be jointly limited by two factors at the same time and therefore the principle of limiting factors becomes inclusive, in contrast to the exclusive nature displayed by the original minimum formulation. A natural concern at this point is whether or not this inclusive character can jeopardize VSL's performance as a TRW forward model. This is a question of future research.

With regard to the response thresholding phenomenon, arising from the bounded growth response windows, we found that it influences the assimilation skill in a strongly asymmetric fashion, being the lower thresholding significantly more detrimental than the upper one. This finding is rather fortunate in the view of more realistic DA experiments involving actual TRW chronologies, given the strong insensitivity of VSL to the lower moisture response threshold M^L , observed in Tolwinski-Ward et al. (2013). With that, there exist a relative freedom to set the value of M^L , which could in principle be used to reduce the adverse effects of the lower moisture response thresholding, without compromising the TRW simulation competence of VSL.

Concerning the initial motivation of assimilating actual TRW chronologies into climate models, we deem our results promising in the sense that EnKF techniques appear robust in the face of two strong nonlinearities typically found in process-based tree-ring growth forward models, i.e., switching recording and thresholded response. Nonetheless, real-world applications need to address many additional practical complications, such as (1) the sparse and highly irregular spatial distribution of TRW chronologies, (2) remarkably low Signal-To-Noise ratios, (3) the non-Gaussian nature of the soil moisture variable and (4) the scarcity of significant long-range predictability skill for most of the current climate models. Related to the latter, assimilation of TRW chronologies seems predestined to be performed in offline conditions at first sight, given that the 1-year time-integration period is considerably longer than the useful lead times of existent climate prediction systems. However, the very strong activity in the fields of seasonal and decadal forecasting has made evident the existence of abundant potential sources of climate internal variability with time scales longer than 1 year (Smith et al. 2012). In any case, despite its lack of accumulation of observational information over time, offline DA has already been shown to be more robust than traditional climate field reconstruction techniques based on orthogonal empirical functions (Steiger et al. 2014). Moreover, the implementation and running of offline DA schemes is remarkably cheaper than online approaches.

Acknowledgments We thank Suz Tolwinski-Ward, Michael Evans, Petra Breitenmoser, Jörg Franke and Christian Reick for the very insightful discussions on tree ring forward modeling. We are also grateful to Jeffrey Anderson, Kayo Ide, Georg Gottwald and Katja Matthes for the truly helpful advise concerning paleo-DA. We show our appreciation to Isabel Dorado for her invaluable aid regarding dendroclimatology. We sincerely thank David Volpe for his thorough style correction. WA expresses his deep gratitude to Camila Rodriguez for her unconditional support during the production of this research work.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baker A, Bradley C, Phipps SJ, Fischer M, Fairchild IJ, Fuller L, Spötl C, Azcurra C (2012) Millennial-length forward models and pseudoproxies of stalagmite $\delta^{18}O$: an example from nw scotland. *Clim Past* 8(4):1153–1167
- Barkmeijer J, Iversen T, Palmer TN (2003) Forcing singular vectors and other sensitive model structures. *Q J R Meteorol Soc* 129(592):2401–2423. doi:10.1256/qj.02.126
- Bhend J, Franke J, Folini D, Wild M, Brönnimann S (2012) An ensemble-based approach to climate reconstructions. *Clim Past* 8(3):963–976. doi:10.5194/cp-8-963-2012
- Boucher E, Guiot J, Hatté C, Daux V, Danis PA, Dussouillez P (2014) An inverse modeling approach for tree-ring-based climate reconstructions under changing atmospheric CO_2 concentrations. *Bio-geosciences* 11(12):3245–3258
- Breitenmoser P, Brönnimann S, Frank D (2014) Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies. *Clim Past* 10(2):437–449. doi:10.5194/cp-10-437-2014
- Brönnimann S (2011) Towards a paleoreanalysis? *ProClim-Flash* 1(51):16
- Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. *Mon Weather Rev* 126:1719–1724
- Danis PA, Hatté C, Misson L, Guiot J (2012) Maideniso: a multiproxy biophysical model of tree-ring width and oxygen and carbon isotopes. *Can J Res* 42(9):1697–1713
- Dirren S, Hakim GJ (2005) Toward the assimilation of time-averaged observations. *Geophys Res Lett* 32(4):L04,804. doi:10.1029/2004GL021444
- Dubinkina S, Goosse H (2013) An assessment of particle filtering methods and nudging for climate state reconstructions. *Clim Past* 9(3):1141–1152. doi:10.5194/cp-9-1141-2013
- Evans MN (2007) Toward forward modeling for paleoclimatic proxy signal calibration: a case study with oxygen isotopic composition of tropical woods. *Geochem Geophys Geosyst* 8(7):Q07008. doi:10.1029/2006GC001406
- Evans MN, Reichert BK, Kaplan A, Anchukaitis KJ, Vaganov EA, Hughes MK, Cane MA (2006) A forward modeling approach to paleoclimatic interpretation of tree-ring data. *J Geophys Res* 111:G03008. doi:10.1029/2006JG000166
- Evans M, Tolwinski-Ward S, Thompson D, Anchukaitis K (2013) Applications of proxy system modeling in high resolution paleoclimatology. *Quat Sci Rev* 76:16–28. doi:10.1016/j.quascirev.2013.05.024
- Fritts HC (1976) *Tree rings and climate*. Academic Press, New York
- Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. *Q J R Meteorol Soc* 125(554):723–757
- Guiot J, Wu HB, Garreta V, Hatté C, Magny M (2009) A few prospective ideas on climate reconstruction: from a statistical single proxy approach towards a multi-proxy and dynamical approach. *Clim Past* 5(4):571–583. doi:10.5194/cp-5-571-2009
- Hamill TM (2006) Ensemble-based atmospheric data assimilation. In: Palmer T, Hagedorn R (eds) *Predictability of weather and climate*. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511617652.007 Web
- Heinze C (2001) Towards the time dependent modeling of sediment core data on a global basis. *Geophys Res Lett* 28(22):4211–4214
- Hughes M, Guiot J, Ammann C (2010) An emerging paradigm: process-based climate reconstructions. *PAGES News* 18(2):87–89
- Hughes M, Ammann C (2009) The future of the past—an earth system framework for high resolution paleoclimatology: editorial essay. *Clim Change* 94(3–4):247–259. doi:10.1007/s10584-009-9588-0
- Huntley H, Hakim G (2010) Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Clim Dyn* 35(6):995–1009. doi:10.1007/s00382-009-0714-5
- Kurahashi-Nakamura T, Losch M, Paul A (2014) Can sparse proxy data constrain the strength of the atlantic meridional overturning circulation? *Geosci Model Dev* 7(1):419–432. doi:10.5194/gmd-7-419-2014
- Lorenz EN (1996) Predictability, a problem partly solved. In: *Proceedings of ECMWF seminar on predictability*. ECMWF, Reading, pp 1–19
- Marchini A (2011) Modelling ecological processes with fuzzy logic approaches. In: Jopp F, Reuter H, Breckling B (eds) *Modelling complex ecological dynamics*. Springer, Berlin, pp 133–145
- Mathiot P, Goosse H, Crosta X, Stenni B, Braida M, Renssen H, Van Meerbeeck CJ, Masson-Delmotte V, Mairesse A, Dubinkina S (2013) Using data assimilation to investigate the causes of southern hemisphere high latitude cooling from 10 to 8 ka bp. *Clim Past* 9(2):887–901. doi:10.5194/cp-9-887-2013
- Nguyen HT, Prasad NR, Walker CL, Walker EA (2002) *A first course in fuzzy and neural control*, 1st edn. Chapman and Hall/CRC, London
- Pendergrass A, Hakim G, Battisti D, Roe G (2012) Coupled air-mixed layer temperature predictability for climate reconstruction. *J Clim* 25(2):459–472. doi:10.1175/2011JCLI4094.1
- Roden JS, Lin G, Ehleringer JR (2000) A mechanistic model for interpretation of hydrogen and oxygen isotope ratios in tree-ring cellulose. *Geochimica et Cosmochimica Acta* 64(1):21–35
- Salski A (2006) Ecological applications of fuzzy logic. In: Recknagel F (ed) *Ecological informatics*. Springer, Berlin, pp 3–14
- Schmidt GA (1999) Forward modeling of carbonate proxy data from planktonic foraminifera using oxygen isotope tracers in a global ocean model. *Paleoceanography* 14(4):482–497
- Se Z (2009) *Fuzzy logic and hydrological modeling*. CRC Press, Boca Raton
- Smith DM, Scaife AA, Kirtman BP (2012) What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environ Res Lett* 7(1):015,602
- Steiger N, Hakim G, Steig E, Battisti D, Roe G (2014) Assimilation of time-averaged pseudoproxies for climate reconstruction. *J Clim* 27(1):426–441. doi:10.1175/JCLI-D-12-00693.1
- Sturm C, Zhang Q, Noone D (2010) An introduction to stable water isotopes in climate models: benefits of forward proxy modelling for paleoclimatology. *Clim Past* 6(1):115–129

- Sugeno M (1985) Industrial applications of fuzzy control. Elsevier Science Inc., New York
- Thompson DM, Ault TR, Evans MN, Cole JE, Emile-Geay J (2011) Comparison of observed and simulated tropical climate trends using a forward model of coral $d^{18}O$. *Geophys Res Lett* 38(14):L14,706
- Tolwinski-Ward SE (2012) Inference on tree-ring width and paleoclimate using a proxy model of intermediate complexity. Ph.D. thesis, The University of Arizona
- Tolwinski-Ward SE, Evans MN, Hughes M, Anchukaitis KJ (2011) An efficient forward model of the climate controls on interannual variation in tree-ring width. *Clim Dyn* 36(11–12):2419–2439. doi:[10.1007/s00382-010-0945-5](https://doi.org/10.1007/s00382-010-0945-5)
- Tolwinski-Ward SE, Anchukaitis KJ, Evans MN (2013) Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width. *Clim Past* 9(4):1481–1493. doi:[10.5194/cp-9-1481-2013](https://doi.org/10.5194/cp-9-1481-2013)
- Tolwinski-Ward S, Tingley M, Evans M, Hughes M, Nychka D (2014) Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response. *Clim Dyn*. pp 1–16. doi:[10.1007/s00382-014-2139-z](https://doi.org/10.1007/s00382-014-2139-z)
- Vaganov E, Hughes M, Shashkin A (2006) Growth dynamics of conifer tree rings: images of past and future environments, ecological studies, vol 183. Springer, New York
- von Storch H, Cubasch U, González-Ruoco J, Jones J, Widmann M, Zorita E (2000) Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (datun). In: Proceedings 28–31. 11th symposium on global change studies, AMS, Long Beach, CA, USA.9-14/1 pp 2119–2128
- Wanru W, Dickinson R (2004) Time scales of layered soil moisture memory in the context of land-atmosphere interaction. *J Clim* 17:2752–2764
- Zadeh L (1965) Fuzzy sets. *Inf Control* 8(3):338–353. doi:[10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zadeh LA (1975) Fuzzy logic and approximate reasoning. *Synthese* 30(3–4):407–428