An Empirical Approach for Estimating Macroturbulent Heat Transport Conditional upon the Mean State

UTE LUKSCH AND HANS VON STORCH

Meteorologisches Institut der Universität Hamburg, Hamburg, Germany

(Manuscript received 14 January 1998, in final form 17 August 1998)

ABSTRACT

A stochastic specification for monthly mean wintertime eddy heat transport conditional upon the monthly mean circulation is proposed. The approach is based on an analog technique. The nearest neighbor for the monthly mean streamfunction (at 850 and 300 hPa) is searched for in a library composed of monthly data of a 1268-yr control simulation with a coupled ocean–atmosphere model. To reduce the degrees of freedom a limited area (the North Atlantic sector) is used for the analog specification. The monthly means of northward transient eddy flux of temperature (at 750 hPa) are simulated as a function of these analogues.

The stochastic model is applied to 300 years of a paleosimulation (last interglacial maximum around 125 kyr BP). The level of variability of the eddy heat flux is reproduced by the analog estimator, as well as the link between monthly mean circulation and synoptic-scale variability. The changed boundary conditions (solar radiation and CO_2 level) cause the Eemian variability to be significantly reduced compared to the control simulation. Although analogues are not a very good predictor of heat fluxes for individual months, they turn out to be excellent predictors of the distribution (or at least the variance) of heat fluxes in an anomalous climate.

1. Introduction

The analog technique has a long history in weather forecasting; the idea is to analyze the present state and to find in a library of past cases that one that resembles most closely the present state. This closest case is the *analogue*, and the properties of this analogue are then assigned to the present state. In the case of day-to-day or seasonal forecasting, this property is the trajectory emanating from the analogue in the next few days or in the upcoming season (Barnett and Preisendorfer 1978; Drosdowsky 1994; just to mention a few). The analog technique for weather forecasting purposes has been widely discredited. For instance, Petterssen (1957) summarized the state of the art: ''While types and analogues . . . have been found useful as background information . . . , forecasts with acceptable tolerance do not appear to be obtainable on the basis of analogues alone.'' Theoretical studies have shown that the library must be enormously large to guarantee a reasonable skill for forecast on the daily timescale (i.e., returning a trajectory reasonably close to the actual trajectory; Lorenz 1969; van den Dool 1994).

In spite of this general failure of the analog method for predictive purposes, the analog technique has been used for a number of useful purposes recently. Fraedrich (1986) and Fraedrich and Leslie (1991) studied the forecast problem in a phase space setup. They examined the divergence of trajectories emanating from a state and its analogue and estimated the dimension of the attractor. In a recent study, Fraedrich and Rückert (1998) determined a metric leading to optimal forecasts. This metric describes a subspace of the full phase space, and Fraedrich argues that this subspace is an approximation of the attractor of the weather system. Another application deals with the specification problem in downscaling (Zorita et al. 1995). In this case, the property of the analogue is a variable insufficiently represented in climate modeling, such as precipitation. The analogue is determined in terms of a variable well simulated by a climate model, for instance, a large-scale pressure map.

In the present study, we deal with the specification problem of determining complex variables with the help of known distributions. Such problems occur when turbulent fluxes are to be parameterized in reduced models or atmospheric feedbacks are to specified as a function of sea surface temperature as in Barnett et al. (1993; Tropics) and Xu et al. (1998; North Pacific). We deal exemplarily with the specification of midlatitude macroturbulent meridional heat flux by means of the time-mean circulation. More specifically, we ask for a random function relating the monthly mean streamfunction $\overline{\Psi}$ and the monthly mean of poleward eddy heat transport

Corresponding author address: Dr. Ute Luksch, Meteorologisches Institut der Universität Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany.
E-mail: Luksch@elkrz.de

in the Northern Hemisphere $\overline{v'T'}$. That is, we adopt the view that the transport is *not determined* by the mean state but only *conditioned* by it. In probabilistic terms, the random vector **X**, a vector of scalar random variables that are the result of the same experiments, (here, monthly mean of poleward eddy heat transport in the North Atlantic sector) is conditioned upon the mean state (cf. von Storch 1997). This may formally be written as

$$\mathbf{X} \sim \mathbf{P}(\overline{\mathbf{\Psi}}), \tag{1}$$

where **P** represents a probability distribution that depends in some way on the variable $\overline{\Psi}$.

Earlier, linear regression analyses have been made linking the mean state and the (conditional) mean transport (Lau 1988). In our terminology, such a regression establishes a linear *deterministic* link,

$$E(\mathbf{X}) = \mathbf{S}\overline{\mathbf{\Psi}}, \tag{2}$$

with the expectation operator E and some suitable matrix **S**. In the present analysis, the probabilistic approach [(1)] is adopted and executed by means of an analog approach. A similar concept is used by Da Costa and Vautard (1997). They used 25 years of observed data to close their low-order model for diagnosing weather regimes with the analog technique.

By using the analog method, several advantages are gained. First, automatically a probabilistic specification is obtained, and second, there is no need for explicitly specifying the conditional probability distribution **P**, which is an unsolved problem, nor the deterministic linkage **S**. The application of the analog method requires a large library of cases, and we are in the lucky situation of being among the first having access to large sets of pairs of fields $\overline{\Psi}$ and $\overline{v'T'}$, which have been simulated in ultralong (thousands of years) simulations with quasi-realistic coupled ocean–atmosphere models (e.g., Manabe and Stouffer 1994; von Storch et al. 1997). Thus, we do not claim to be more insightful than earlier scientists but only that we luckily have access to powerful tools that were unavailable just a few years earlier.

Equation (1) offers an intriguing perspective. It may be used as a parameterization of macroturbulence in an atmospheric model operating with time-mean fields as state variables, without calculating the detailed synoptic activity. Such a model automatically would be a *stochastic climate model* (Hasselmann 1976), being able to generate long-term variations as a response to short-term random fluctuations. Simple closure relations for the horizontal eddy fluxes cannot be derived from eddy flux equations (Opsteegh and van den Dool 1979). An empirical approach by analogues makes sense if the variability of monthly or seasonal mean patterns is not completely caused by day-to-day weather fluctuations. The path in the phase space is realistic but not applicable to the development of the real atmosphere (Opsteegh and van den Dool 1979). Therefore, we cannot expect very high pattern correlation coefficients as a measure for the difference between original and estimated fields. But we can expect that the connection between low-frequency and bandpass-filtered variability measured, for example, by the canonical correlation analysis (CCA) is maintained by the analog system. The approach is tested with an independent dataset where the variability is changed in comparison with the library. This test is stronger than splitting the control run in two halves or taking the target month out.

The paper is organized as follows. The analog model and the data basis (library and an independent test set) are described in sections 2 and 3. Measures of skill are briefly introduced in section 4. Afterward, the stochastic model is tested with data from a GCM simulation of the last interglacial period, using a library of a multicentury simulation of the present state (section 5). At the end the results are summarized and an outlook is offered.

2. The analog specification

The analog specification is done in the phase space spanned by EOFs; that is, the originally gridded data are expanded into a series of EOFs:

$$\overline{\mathbf{\Psi}} = \sum_{i=1}^{N_\psi} \alpha_i^\Psi \mathbf{e}_i^\Psi. \tag{3}$$

In this expansion, the EOFs are normalized, $\|\mathbf{e}_i^\Psi\| = 1$, so that the variance of α_1^ψ is maximum, that of α_2^Ψ is second largest, and so on. To find better analogues the analog specification area is restricted the North Atlantic sector, which is defined as the domain lying between 20° and 70°N latitude and 110°W and 40°E longitude. Here N_Ψ is the number of grid points in this limited area, which may be an overestimation of the theoretical degree of freedom. All data are anomalies; the long-term mean of the control run is subtracted.

Given any circulation pattern, say $\overline{\Psi}_0$, its analogue is its nearest neighbor in terms of some distance in the phase space. In general, the distance between any two fields $\overline{\Psi}_1$ and $\overline{\Psi}_2$ will be of the form

$$d_\gamma(\overline{\mathbf{\Psi}}_1, \overline{\mathbf{\Psi}}_2) = \sum_{i=1}^{N_\psi} \gamma_i[\alpha_i^{\Psi_1} - \alpha_i^{\Psi_2}]^2 \tag{4}$$

with suitably chosen weights γ_i. In our application, we take into account only the first n_Ψ EOFs, so that $\gamma_i = 0$ for $i > n_\Psi$. All other weights are set to 1. Note that the expected difference $(\alpha_i^{\Psi_1} - \alpha_i^{\Psi_2})^2 = 2\ \mathrm{var}(\alpha_i^\Psi)$ between two randomly selected fields $\overline{\Psi}_1$ and $\overline{\Psi}_2$ is decreasing with increasing index i, so that the contributions of the low-indexed EOFs are more important than those of high-indexed EOFs.

With this definition, the analogue $\overline{\mathbf{\Psi}}_*$ of $\overline{\mathbf{\Psi}}_0$ satisfies

$$d_\gamma(\overline{\mathbf{\Psi}}_0, \overline{\mathbf{\Psi}}_*) = \min_{\overline{\mathbf{\Psi}} \in \mathcal{L}} d_\gamma(\overline{\mathbf{\Psi}}_0, \overline{\mathbf{\Psi}}), \tag{5}$$

where $\mathcal{L}$ is the *library*, that is, the collection of all circulation fields available. After this analogue of $\overline{\Psi}$

(monthly mean streamfunction at 300 and 850 hPa) is determined, a dynamically consistent macroturbulent heat transport $\overline{v'T'}$ (monthly mean values at 750 hPa) may be determined as the heat transport observed (or simulated) together with $\overline{\Psi}_*$, that is,

$$\widehat{\overline{\mathbf{v}'\mathbf{T}'}} = \overline{\mathbf{v}'\mathbf{T}'_*}. \quad (6)$$

The estimated flux is the monthly mean anomaly from the long-term mean of the control run for the neighbor ''$*$''. The full field $\overline{v'T'_*}$ is used without truncation to preserve variance. The estimator [(6)] depends on the choice of the ''key'' variable $\overline{\Psi}$ and the available library of cases, and on the parameters n_Ψ and the weights γ_i.

For the following, we need to define some technical nomenclature. The field, for which a consistent property—in this case the monthly mean streamfunction $\overline{\Psi}$—is sought, is called the *key field*, while the $\overline{v'T'}$ field is called the *property field.* The consistent property field assigned to the key field is the *estimated property field.*

In the basic setup, the complete key fields $\overline{\Psi}$ need not to be stored in the library but only the first n_Ψ EOF coefficients. Since $n_\Psi \ll N_\Psi$ the amount of required storage is significantly reduced, and also the computational efforts for finding the minimum [(5)] are considerably eased. In a similar manner, the required disk space for the property field library may be reduced significantly (for details, see the appendix).

Note that no feature of the property field enters the analog determination. Indeed, when an analogue is found for some key field, then any property attached to the analogue may be seen as an estimate of that property of the original key variable. This means, in practical terms, that when we deal with climate model data, we obtain not only an estimate of the meridional heat transport but also of all other properties simulated by the climate model, be it the angular momentum, regional soil moisture, or intramonthly extreme values. However, when the metric used to determine the analogue is optimized for returning best fits of the property field, as in case of Fraedrich and Rückert (1998), an analogue suitable for $\overline{v'T'}$ is no longer automatically also suitable for, say, soil moisture.

3. Ultralong ocean–atmosphere GCM simulations

For the success of our study the availability of a large library of dynamically consistent cases is mandatory. We use two such ultralong simulations, namely, a 1268-yr simulation of the present climate (''control run''; von Storch et al. 1997) and a 500-yr simulation of the last interglacial about 125 kyr BP (''Eemian run''; Montoya et al. 1998). Boundary conditions of the Eemian experiment differ from those of the control run in two ways: CO_2 is reduced (control run, 330 ppmv; Eemian run, 267 ppmv) and the orbit parameters are changed. During Eem (approximately 125 kyr BP) the obliquity and eccentricity was increased and the perihelion occurred in Northern Hemisphere summer (rather than winter, as today) causing an amplification (reduction) of the seasonal cycle of insolation in the Northern (Southern) Hemisphere. The proposed 6-m sea level rise for the Eemian results in negligible changes in the T21 land–sea grid. From the control run the library is derived, and the data from the Eemian run are used for the determination of the skill of the estimator (see section 4).

In the remainder of this section, we describe the setup of the two simulations, demonstrate the ability of the control run to generate a quasi-realistic link between mean states and the meridional heat transport, and describe the modifications in the meridional heat transport simulated in the Eemian run. In section 5 we will show that our analog estimator is able not only to reproduce the link between mean states and the heat transport but also the modification of the heat transport variability in the Eemian run.

Both long-term simulations are done with the coupled ocean–atmosphere Hamburg climate model (ECHAM1/LSG; Cubasch et al. 1992), which is driven by a prescribed annual cycle of solar radiation and a fixed concentration of atmospheric carbon dioxide. The ocean and the atmosphere interact freely in this model, and dynamically inactive flux correction is adopted for preventing the model from ''running away.'' The resolution of the GCM is around 5.6° in the longitudinal and meridional directions both in the atmosphere and in the ocean. In spite of this rather coarse resolution the North Atlantic storm track is simulated reasonably well in terms of its pattern, but its strength is underestimated compared to observations. Also, the link between monthly mean states and the monthly midtropospheric heat transports are simulated reasonably well (see Fig. 1; cf. Lau 1988).

The 1268-yr control run (von Storch et al. 1997) is integrated with the present-day annual solar radiation cycle and an atmospheric CO_2 concentration of 330 ppm; the Eemian run (Montoya et al. 1998) was initiated with the state of the control run in the year 600, when the control run has long reached a statistical equilibrium state in terms of atmospheric variables, and then integrated with a CO_2 concentration of 267 ppm and radiation reconstructed for the time about 125 000 years before present (the ''Eem'').

In the present study only the winter months December, January, and February are used. During the first centuries, the model simulations drift toward their statistical equilibria. Therefore, only the 2880 cases from the winter months of the years 309–1268 were placed into the library of cases. Similarly, the test with the Eemian simulation was limited to the 900 winter months of the last 300 years of that run.

We use three variables in an area covering the North Atlantic in this study: the monthly mean meridional heat transport $\overline{v'T'}$ at 750 hPa due to synoptic disturbances, and the monthly mean streamfunction at 300 and 850 hPa. The meridional heat transport is calculated from

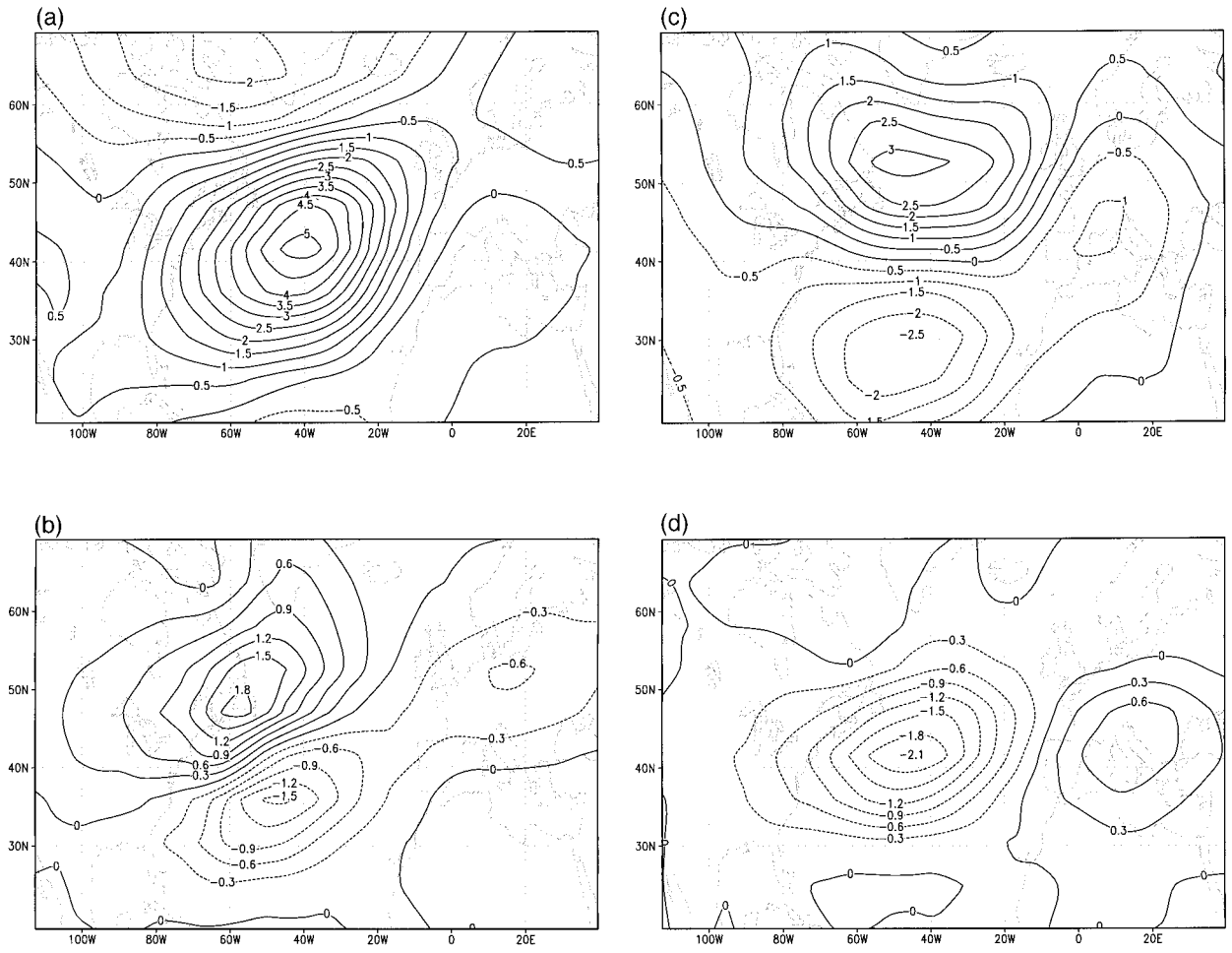

FIG. 1. First two pairs of CCA patterns of anomalous monthly $\overline{\Psi}_{850}$ (units: 10^6 m^2 s^{-1}) (upper) and anomalous monthly $\overline{v'T'}$ at 750 hPa (units: K m s^{-1}) (lower) in the North Atlantic sector derived from 2880 winter months simulated in the control run. (left) The first pair with a correlation of 0.85; they represent 29% of the $\overline{\Psi}_{850}$ variance and 17% of the $\overline{v'T'}$ variance. (right) The second pair with a correlation of 0.74 and 17% and 15% described variances.

time-filtered daily time series of anomalous (deviation from the monthly mean) northward wind and of anomalous temperature. The filter is the one proposed by Blackmon (1976) for retaining synoptic variability on timescales between 2.5 and 6 days. The long-term mean of the control run is subtracted from all data, from the control run, as well as from the Eemian run.

The state of the atmosphere in the North Atlantic region in the Eemian simulation deviates moderately but statistically significantly from that in the control run. In particular, the mean strength of the Atlantic storm track is reduced (not shown). The interannual variability in terms of monthly mean streamfunction and $\overline{v'T'}$ is significantly reduced over the Atlantic (Fig. 2).

For the successful application of the estimator [(6)] to the Eemian run, it is required that the EOFs derived from the control run are capable of representing sufficiently large amounts of variance of the key field in the Eemian run. For the success of the analog estimator it would also be desirable that the $\overline{v'T'}$ EOFs of the control run are efficiently representing variance in the Eemian run. That this is so is demonstrated by Table 1.

The projection of the Eemian data on the control run EOFs explains nearly the same amount of variance as the projection on the Eemian run EOFs. The pattern correlation coefficients control run EOFs versus Eemian run EOFs are very high (about 0.8 to 0.97 for the first five patterns). If the EOF space is totally changed then it is impossible to find neighbors by minimizing the distance in the EOF space.

4. Measures of skill

We quantify the skill of the estimator [(6)] with a number of standard measures (cf. von Storch and Zwiers 1998). One measure is the *pattern* or *anomaly correlation coefficient A* defining the spatial variability at a special month. If $\mathbf{X}$ and $\hat{\mathbf{X}}$ are fields with spatial means $\overline{X}$ and $\overline{\hat{X}}$, then the anomaly correlation coefficient of $\mathbf{X}$ and $\hat{\mathbf{X}}$ is

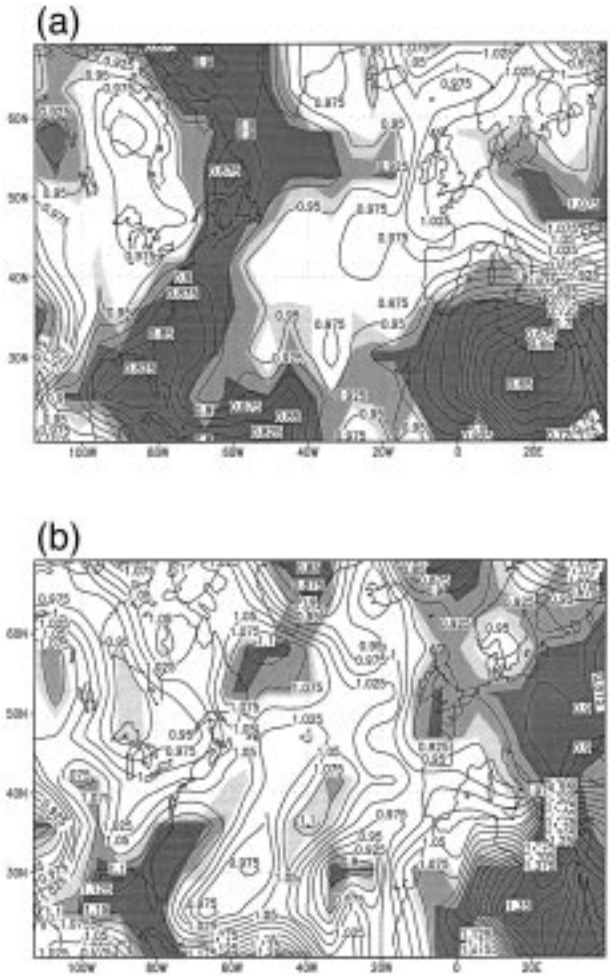

FIG. 2. Ratios of standard deviations of meridional eddy heat transport $\overline{v'T'}$. Areas where the ratio is significantly larger (or smaller) than 1 according to a conventional F test operating at a significance level of 90%, 95%, and 99% are lightly, medium, and heavily shaded, respectively. All data are from winter. (upper) ECHAM/LSG simulation Eemian vs control (σ_E/σ_C); (lower) ANALOG simulation vs ECHAM/LSG for the Eemian simulation (σ_A/σ_E).

$$A_{X\hat{X}} = \frac{E[(\mathbf{X} - \overline{X})(\hat{\mathbf{X}} - \overline{\hat{X}})]}{\{E[(\mathbf{X} - \overline{X})^2] \cdot E[(\hat{\mathbf{X}} - \overline{\hat{X}})^2]\}^{1/2}}. \quad (7)$$

For each field we get one of these numbers, so that we end up with a distribution of anomaly correlation coefficients when we apply our estimator to many cases.

Another measure is the *skill score*, which compares an estimator with the performance of some simple reference estimator. In the present problem, we are in need of an estimator that has the same local variance as the original property field, that is, $\mathrm{var}(\overline{v'T'}) = \mathrm{var}(\widehat{\overline{v'T'}})$. The only simple reference estimator fulfilling this requirement is random choice, so that the skill scores for each random variable of the random vector $\mathbf{X}$ read

TABLE 1. Amounts of explained variance of $\overline{\Psi}$ (monthly mean streamfunction at 850 and 300 hPa) and $\overline{v'T'}$ (monthly means at 750 hPa) in the control run (cont) and Eemian run (Eem) represented by the first $i = 1, \ldots, 10$ control run EOFs.

EOF	$\overline{v'T'}$		$\overline{\Psi}$		EOF	$\overline{v'T'}$		$\overline{\Psi}$	
i	cont	Eem	cont	Eem	i	cont	Eem	cont	Eem
1	20	18	21	17	6	4	4	4	4
2	17	16	19	20	7	4	4	4	3
3	8	8	12	12	8	3	3	3	3
4	7	8	9	9	9	3	3	3	3
5	5	5	5	4	10	2	2	2	3
1–5	57	55	66	62	1–10	73	71	82	78

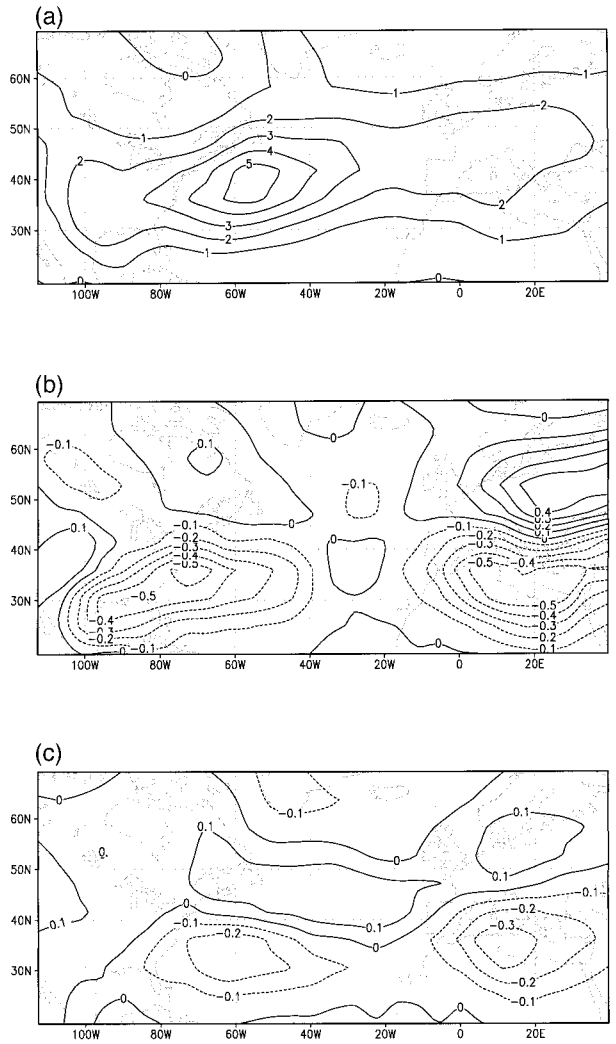

FIG. 3. The $\overline{v'T'}$ fields. (top) Long-term mean of the control run; (middle) average over all $\overline{v'T'}$ anomalies of the Eemian run; (bottom) average all analogues.

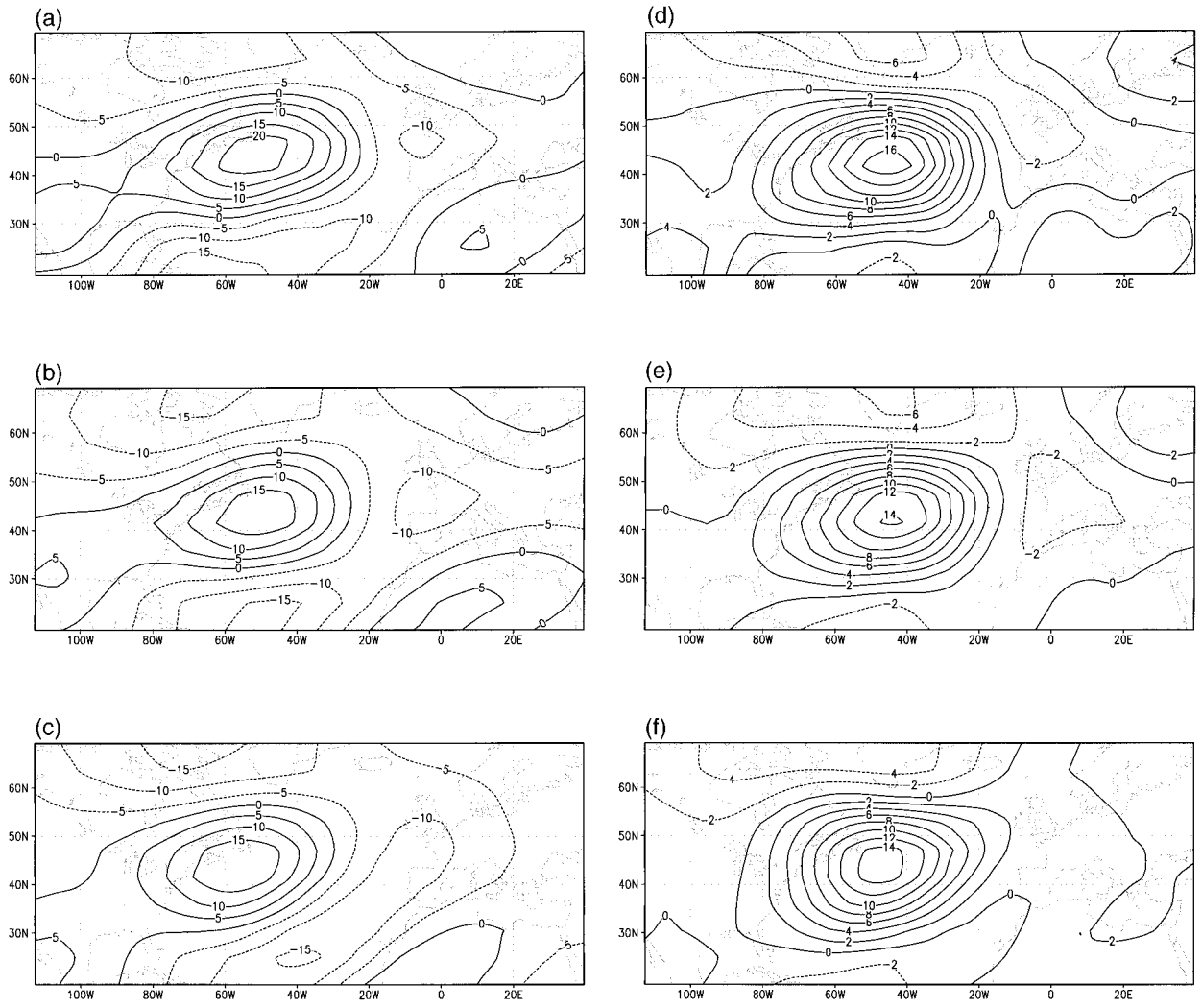

FIG. 4. An example of the anomalous monthly mean streamfunction (the long-term mean of the control run is subtracted; units, 10^6 m^2 s^{-1} simulated in the Eemian run and its analogue derived from the control run library; (left) 300-hPa and (right) 850-hPa streamfunction anomaly; (top) Eemian state, untruncated; (middle) Eemian state, truncated by the first 25 EOFs (search for a minimum in the truncated space); (bottom) control state, truncated by the first 25 EOFs. The $\overline{v'T'}$ fields and $\overline{\Psi}$ fields belonging to the Eemian mean state and to the control run analogue are shown in Fig. 5.

$$B = 1 - \frac{1}{2}\frac{\mathrm{var}(\widehat{\overline{v'T'}} - \overline{v'T'})}{\mathrm{var}(\overline{v'T'})}. \qquad (8)$$

A skill score of 1 indicates that the estimator is perfect, a positive score that the estimator is better than the reference estimator, a skill score of zero that it is as good as the reference, and a negative score that it is worse. Here the temporal variability at a special location is considered.

5. Results: Application to a paleosimulation

After some experimentation, we have selected the streamfunction at 300 and 850 hPa as a key variable for the determination of analogues. The analogues are used to estimate the meridional heat transport $\overline{v'T'}$. Our standard setup was $n_\Psi = 25$ and $\gamma_i = 1$ for $i \le n_\Psi$ and $\gamma_i = 0$ for $i > n_\Psi$ [see (4)].

The analog system is an ''anomaly model,'' and all data are monthly mean anomalies from the long-term climate of the control run. The northward transport of heat in the Eemian run is reduced south of 40°N and increased over Europe. Therefore the mean over all Eemian anomalies are not equal to zero (Fig. 3). This is reproduced by the analog system.

To demonstrate the procedure we present an example of analogues and the analog estimator [(6)] in Figs. 4 and 5. First, an anomalous circulation $\overline{\Psi}$, in terms of 850- and 300-hPa streamfunction, is selected from the Eemian run (top row of Fig. 4). The truncated fields, obtained through projections of the full fields on the first $n_\Psi = 25$ EOFs, are displayed in the middle row.

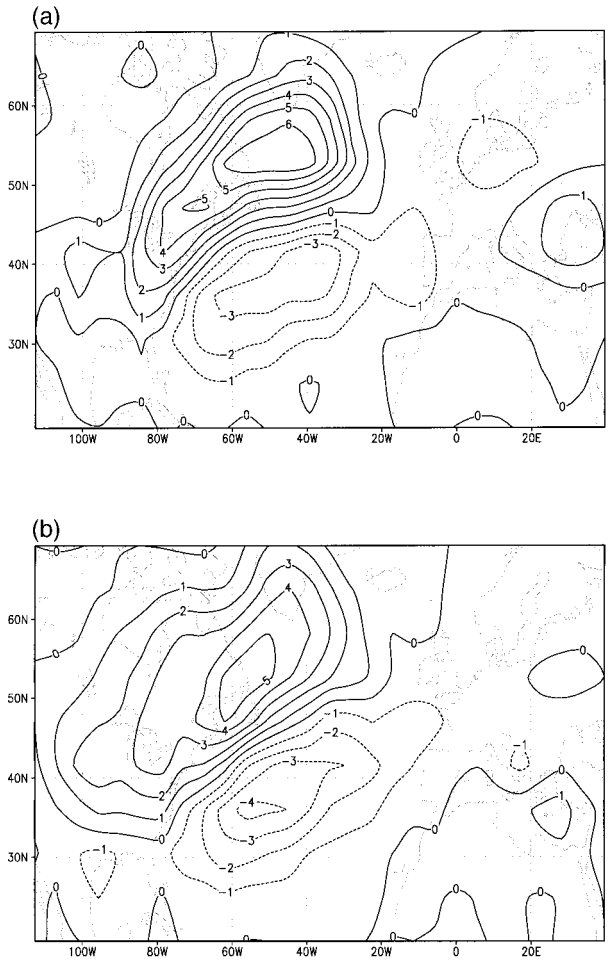

FIG. 5. Anomalous Eemian $\overline{v'T'}$ at 750-hPa distributions belonging to the case shown in Fig. 4 (top) and its estimate derived from the analogue shown in Fig. 4 (bottom); the long-term mean of the control run is subtracted). (top) Eemian state; (bottom) analog estimate.

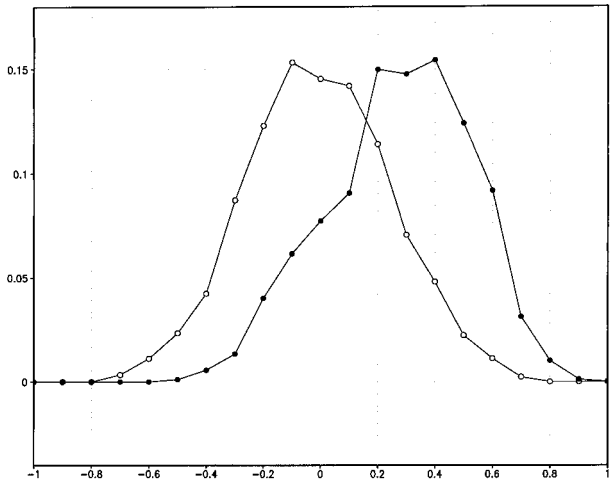

FIG. 6. Distribution of anomaly pattern correlation coefficients between analog estimator and Eemian data from ECHAM/LSG (frequency divided by the total number of experiments). Full points, $\overline{v'T'}$ field; boxes, $\overline{v'T'}$ field, choice of the analogue by chance; open points, $\overline{\Psi}$ field.

The poleward heat flux $\overline{v'T'}$ belonging to this circulation is in the top panel of Fig. 5. Then, in the library of cases from the control run the nearest neighbor is found (bottom row of Fig. 4). Together with this anomalous circulation, the poleward heat transport displayed in the bottom panel of Fig. 5 was simulated in the control run. This heat transport is the estimate $\widehat{\overline{v'T'}}$ of the ''true'' heat flux $\overline{v'T'}$. The presented case is a good case, with an anomaly correlation of 0.89 between the true and estimated eddy heat transport field.

The case represents a situation in which the eddy heat flux in the northeast Atlantic area is shifted northward, while over Europe the circulation seems to be ''kind of blocked.'' These features are well reproduced by the analog mean circulation and the estimated $\overline{v'T'}$ field.

The estimator $\widehat{\overline{v'T'}}$ has been determined for all 900 winter months considered in the Eemian simulation, both for $\overline{\Psi}$ and the poleward eddy heat transport. The skill of the estimator is quantified by calculating anomaly correlation coefficients for all 900 cases. Furthermore, the skill score is calculated for all grid points, and the ratios of the variances in the Eemian data and in the estimated fields are calculated.

The distribution of pattern correlation coefficients between true and analog fields are shown in Fig. 6. Maxima are at 0.4. For comparison, the performance of an unconditionally random choice from the library of cases are given by symbols. In that case, the anomaly correlation coefficients are distributed centered around a mildly negative value because these random choices represent the control climate, which is different from the Eemian climate (Figs. 3b,c). Clearly, the analog estimator is superior to the random choice estimator.

The ratio of variance calculated directly from the Eemian run and of the variance of the estimated $\overline{v'T'}$ fields is displayed in Fig. 2 (lower panel). This ratio varies around the ideal value of 1, except for a maximum ratio in the transport variance over North Africa, where the variations of the transport are small in any case so that small changes result in large ratios. Indeed, the above-mentioned reduction of variability of the monthly mean eddy heat transport in the Eemian simulation compared to the control run is reasonably well captured by the analog estimator (Fig. 2, upper panel).

The skill score (Fig. 7) is almost everywhere positive with maximum values around 0.4 for $\overline{v'T'}$.

Finally, the utility of the proposed estimator is tested by checking that the above-documented CCA link between $\overline{\Psi}_{850}$ and poleward heat transport is still valid when both $\overline{\Psi}_{850}$ and $\overline{v'T'}$ are estimated through the analog procedure. The patterns are very similar to the patterns shown in Fig. 1 and are therefore not shown. The correlations derived from the analog estimates amount to 0.86 for the first pair and 0.75 for the second pair, compared to 0.84 and 0.74 in the Eemian data.

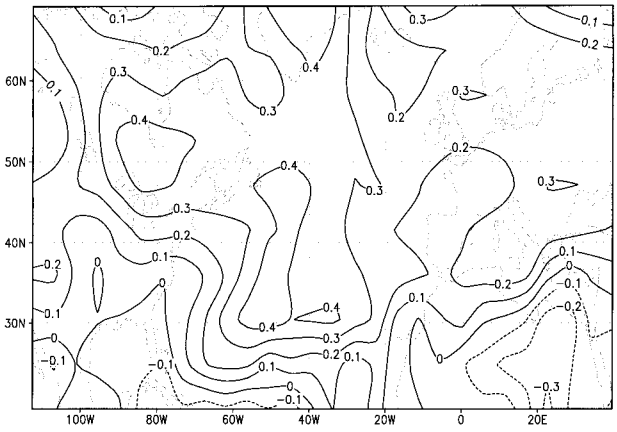

FIG. 7. Skill score for $\overline{v'T'}$ at 750 hPa as simulated in the Eemian run and specified through analogues from the control run.

Both patterns together explain 42%–43% for $\overline{\Psi}_{850}$ variance and 32%–33% for $\overline{v'T'}$ variance.

The time coefficients of the $\overline{v'T'}$ EOFs, derived from the simulation and from the analog estimates, have been compared. The variances of both time series for EOFs $i = 1, \ldots, 30$ are found to be virtually identical (not shown). The correlations between the time series is generally low with maximum values of about 0.5 in the case of the first $\overline{v'T'}$ EOFs. It should be remembered, though, that the estimator [(6)] is not designed for returning maximum correlation but the right variance while maintaining significant correlations. Nevertheless, the distributions of the EOF time coefficients are quite similar (Fig. 8).

Composites of the $\overline{v'T'}$ fields were constructed using an index derived from the monthly mean streamfunction $\overline{\Psi}_{850}$. Here, $\overline{\Psi}_{850}$ data are projected on the first EOF of the control run. The $\overline{v'T'}$ fields averaged for index values greater/smaller than two standard deviations show a nonsymmetric response (Fig. 9). The northward shift

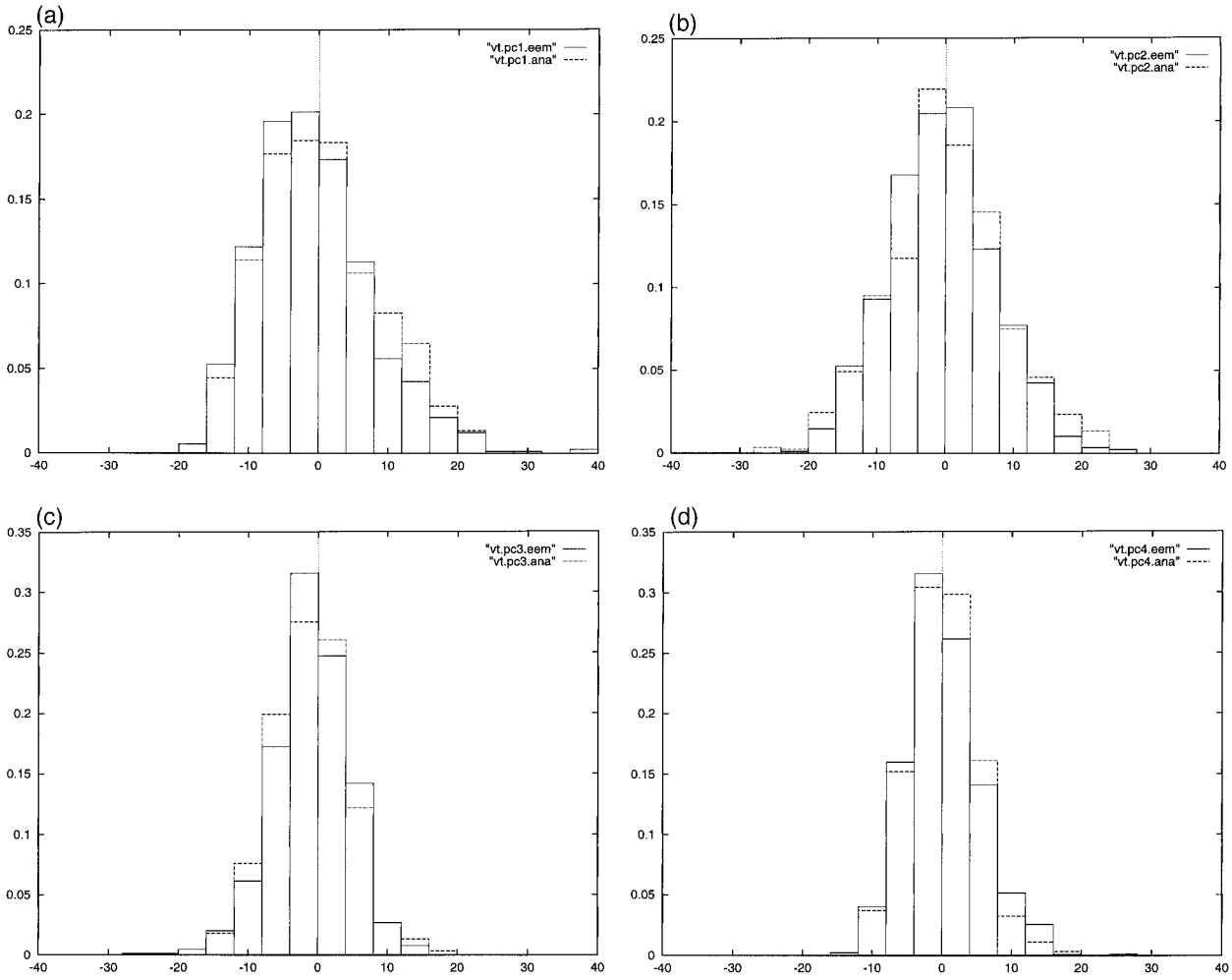

FIG. 8. Distribution of time coefficients. The $\overline{v'T'}$ data are projected on the control run EOFs; Eemian run (dashed line) and the analogues (full line). (top left) EOF 1; (top right) EOF 2; (bottom left) EOF 3; (bottom right) EOF 4.

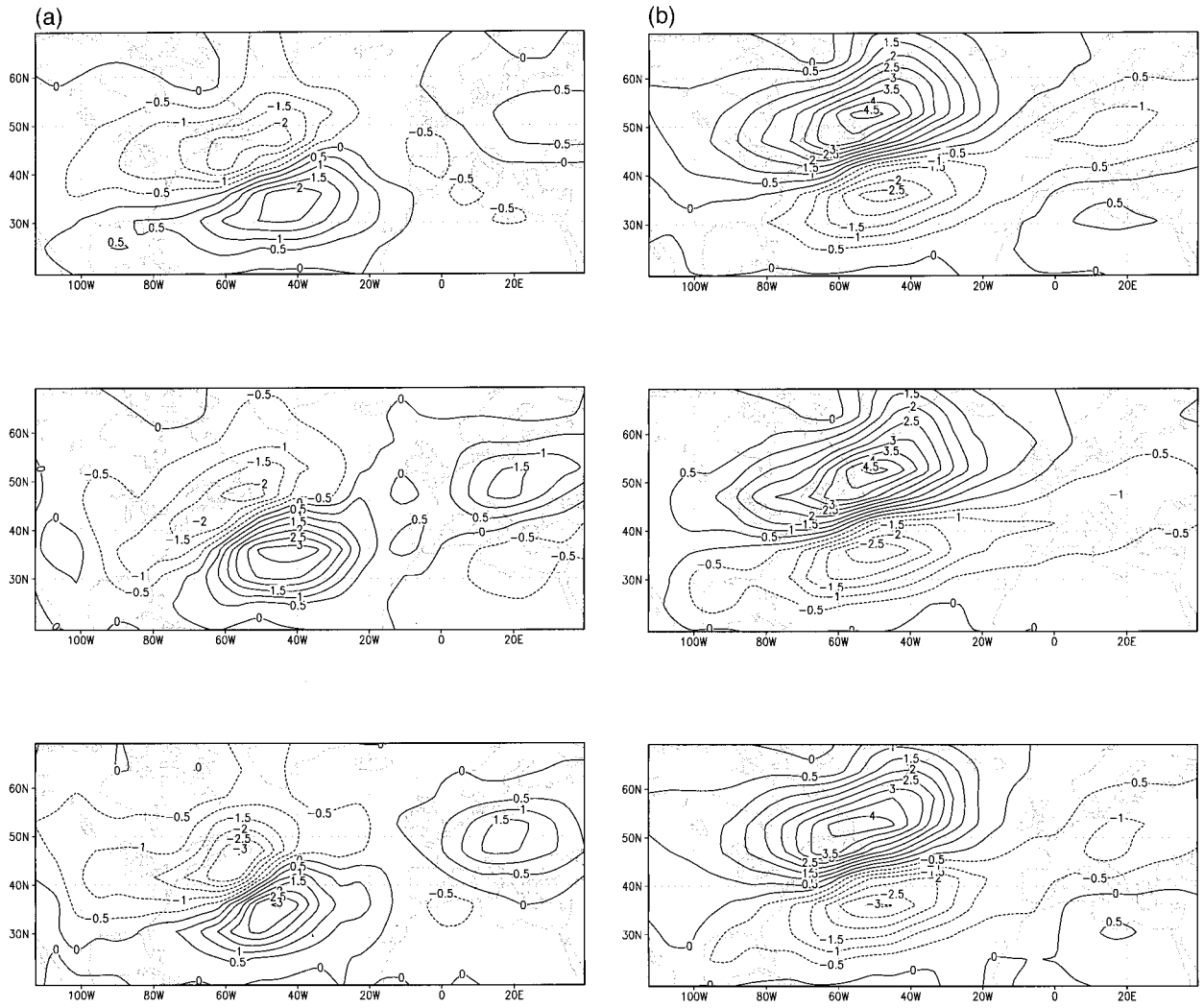

FIG. 9. Composites of the $\overline{v'T'}$ fields. The index is the time coefficient of the streamfunction at 850 hPa projected on the first EOF of the control run. (left) index values greater than +2 standard deviation (average over 92 control/22 Eemian fields). (right) Index values smaller than −2 standard deviation (average over 22 control/8 Eemian fields). (top) Control run; (middle) Eemian run; (bottom) analogues.

of the eddy heat flux is connected with an increase in intensity (Fig. 9, left side: strong positive index) in difference to the southward shift (right side: strong negative index). In the European area there is a stronger intensification of the eddy heat flux for negative index values in the Eemian run (middle panel, right side) but not in the control run (upper panel, right side). This change is reproduced by the analogues (lower panel).

Higher correlation could be obtained if the analog specification would be designed differently. Examples are ''constructed analogues'' (i.e., a linear combination of the library data; van den Dool 1994) and ''analogue/anti-analogues'' (i.e., a combination of neighbors, with weights given by the distance; Livezey and Barnston 1988), or smoothing operations in the analog space, such as kriging (Wackernagel 1994; Biau et al. 1999). However, this gain in correlation is accompanied by a loss of nonlinearity and variance, which would make the scheme unfit to be used as a closure scheme in a simulation model.

The analog technique has been performed with different values of n_Ψ, and the results have been mainly insensitive to this parameter for $n_\Psi \geq 10$.

6. Summary and outlook

In the present analyses, we have tested an analog estimator as a means to determine eddy heat transport fields and other fields (''property fields'') that are consistent with a given monthly mean tropospheric circulation (key field). The estimator is based on an EOF truncation of the key field and a voluminous library of cases of simultaneously simulated key and property fields. Thus, the estimator is not motivated by dynamical insight but entirely based on empirical evidence generated by quasi-realistic climate models.

The proposed scheme has been tested with independent data from a climate model simulation of the last interglacial (Eem, approximately 125 kyr BP). The simulated decrease of eddy activity as well as the link between the mean circulation and the eddy activity simulated in this climate model integration are reproduced by the estimator. In particular it demonstrates that even when analogues are not particularly adept at predicting eddy heat fluxes associated with individual monthly mean states, they can be remarkably accurate predictors of the distribution of monthly mean eddy flux fields that occur in anomalous climates.

It is suggested to use this estimator as an empirical means for closing the macroturbulent problem in simplified atmospheric and ocean large-scale modeling. Simple closure relations for the horizontal eddy fluxes cannot be derived from eddy flux equations (Opsteegh and van den Dool 1979). An empirical approach by analogues makes sense if the variability of monthly or seasonal mean patterns is not completely caused by day-to-day weather fluctuations. The path in the phase space is realistic but not applicable to the development of the real atmosphere (Opsteegh and van den Dool 1979). Therefore, we cannot expect very high pattern correlation coefficients as a measure for the difference between original and estimated fields. But we can expect that the connection between low- and high-frequency variability measured, for example, by the CCA is maintained by the analog system. For these first tests of the method we used an independent dataset where the variability is changed in comparison with the library. This is a stronger test than cutting the control run in two halves or taking the target month out. A test with a local regression model (not shown here) gives nearly the same distribution of pattern correlation coefficients as in Fig. 6 but the the CCA patterns are totally changed and the correlation coefficients are much too high. The analog approach will fail if the EOF space is totally changed in the library and the test set. In this case it is impossible to find neighbors by minimizing the distance in the EOF space.

Acknowledgments. The present analysis was done in the framework of the project MILLENNIA funded by the Climate and Environment Programme of the European Commission. We thank Jin-Song von Storch, Marisa Montoya, Eduardo Zorita, and Viatcheslav Kharin for their support. Computational assistance was offered by Matthias Dorn. Finally, we thank the reviewers for their valuable comments.

APPENDIX

A Computationally Efficient Way of Generating Consistent Property Fields

If the number of grid points N_{vT} of the property field $\overline{v'T'}$ is large, it may be computationally inefficient to store all complete fields in the library. In that case, it may be advantageous to expand the property field into an EOF series,

$$\overline{\mathbf{v}'\mathbf{T}'} = \sum_{i=1}^{N_{vT}} \alpha_i^{vT} \mathbf{e}_i^{vT}, \tag{A1}$$

and to keep only the first n_{vT} EOF coefficients, with $n_{vT} \ll N_{vT}$. Then the estimated property field is composed of n_{vT} EOF ''analog'' coefficients and $N_{vT} - n_{vT} - 1$ randomly drawn coefficients:

$$\widehat{\overline{\mathbf{v}'\mathbf{T}'}}_0 = \sum_{i=1}^{m} \alpha_i^{vT_*} \mathbf{e}_i^{vT} + \sum_{i=m+1}^{N_{vT}} \eta_i \mathbf{e}_i^{vT}, \tag{A2}$$

where $\alpha_i^{vT_*}$ is the ith EOF coefficient of the property field $\overline{v'T'_*}$ associated with $\overline{\Psi}_*$, and η_i is randomly drawn from a normal distribution with mean zero and a variance equal to that of the ith EOF coefficient.

The first sum on the right side of (A2) is the analog part, and the second sum is the noise part. This splitting is motivated by the observation that only the most important EOFs contribute skillfully to the estimate. However, a truncation of the property field to the first n_{vT} EOFs would return an estimated property field with an underestimated spatial variance. Therefore, the coefficients of the remaining EOFs are drawn at random.

For this procedure, it is needed that the first n_{vT} EOF coefficients of the property field $\overline{v'T'}$ are kept in the library, and the variances of the remaining $N_{vT} - n_{vT} - 1$ EOF coefficients.

REFERENCES

Barnett, T. P., and R. W. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci*, **35,** 1771–1787.

——, M. Latif, N. Graham, M. Flügel, S. Pazan, and W. White, 1993: ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface temperature with hybrid coupled ocean–atmosphere model. *J. Climate*, **6,** 1545–1566.

Biau, G., E. Zorita, H. von Storch, and H. Wackernagel, 1999: Estimation of precipitation by kriging in the EOF space of the sea level pressure field. *J. Climate*, **12,** 1070–1085.

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.,* **33,** 1607–1623.

Cubasch, U., K. Hasselmann, H. Höck, E. Maier-Reimer, U. Mikolajewicz, B. D. Santer, and R. Sausen, 1992: Time-dependent greenhouse warming computations with a coupled ocean–atmosphere model. *Climate Dyn.,* **8,** 55–69.

da Costa, E., and R. Vautard, 1997: A qualitatively realistic low-order model of the extratropical low-frequency variability built from long records of potential vorticity. *J. Atmos. Sci.,* **54,** 1064–1984.

Drosdowksy, W., 1994: Analog (nonlinear) forecasts of the Southern Oscillation index time series. *Wea. Forecasting*, **9,** 78–84.

Fraedrich, K., 1986: Estimating the dimension of weather and climate attractors. *J. Atmos. Sci.*, **43,** 331–344.

——, and L. Leslie, 1991: Predictability studies of the Antarctic atmosphere. *Aust. Meteor. Mag.,* **39,** 1–9.

——, and B. Rückert, 1998: Metric adaption for analog forecasting. *Physica A,* **254,** 379–393.

Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus,* **28,** 473–485.

Lau, N.-C., 1988: Variability of the observed midlatitude storm tracks in relation to low-frequency changes in the circulation patterns. *J. Atmos. Sci.,* **45,** 2718–2743.

Livezey, R. E., and A. G. Barnston, 1988: An operational multifield analog/antianalog prediction system for United States seasonal temperatures. Part I: System design and winter experiments. *J. Geophys. Res.,* **93** (D9), 10 953–10 974.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26,** 636–646.

Manabe, S., and R. J. Stouffer, 1994: Multiple-century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Climate,* **7,** 5–23.

Montoya, M., T. J. Crowley, and H. von Storch, 1998: Temperatures at the last interglacial simulated by a coupled ocean–atmosphere climate model. *Paleooceanogr.,* **13,** 170–177.

Opsteegh, J. D., and H. M. van den Dool, 1979: A diagnostic study of the time mean atmosphere over nortwestern Europe during winter. *J. Atmos. Sci.,* **36,** 1862–1879.

Petterssen, S., 1957: Weather observations, analysis and forecasting. *Meteor. Monogr.,* No. 3, Amer. Meteor. Soc., 114–151.

van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus,* **46A,** 314–324.

von Storch, H., 1997: Conditional statistical models: A discourse about the local scale in climate modelling. *Monte Carlo Simulations in Oceanography, Proc. 'Aha Huliko'a Hawaiian Winter Workshop,* Manoa, HI, University of Hawaii, 58–59.

——, and F. W. Zwiers, 1998: *Statistical Analysis in Climate Research.* Cambridge University Press, 528 pp.

von Storch, J.-S., V. Kharin, U. Cubasch, G. C. Hegerl, D. Schriever, H. von Storch, and E. Zorita, 1997: A description of a 1260-year control integration with the coupled ECHAM1/LSG general circulation model. *J. Climate,* **10,** 1526–1544.

Wackernagel, H., 1994: Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma,* **62,** 83–92.

Xu, W., T. P. Barnett, and M. Latif, 1998: Decadal variability in the North Pacific as simulated by a hybrid coupled model. *J. Climate,* **11,** 297–312.

Zorita, E., J. P. Hughes, D. P. Lettenmaier, and H. von Storch, 1995: Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J. Climate,* **8,** 1023–1042.