# A Study of the Cross-Scale Causation and Information Flow in a Stormy Model Mid-Latitude Atmosphere

**X. San Liang**

Center for Ocean-Atmosphere Dynamical Studies, Nanjing Institute of Meteorology, 219 Ningliu Blvd, Nanjing 210044, China; x.san.liang@gmail.com

check for updates

**Abstract:** A fundamental problem regarding the storm–jet stream interaction in the extratropical atmosphere is how energy and information are exchanged between scales. While energy transfer has been extensively investigated, the latter has been mostly overlooked, mainly due to a lack of appropriate theory and methodology. Using a recently established rigorous formalism of information flow, this study attempts to examine the problem in the setting of a three-dimensional quasi-geostrophic zonal jet, with storms excited by a set of optimal perturbation modes. We choose for this study a period when the self-sustained oscillation is in quasi-equilibrium, and when the energetics mimic the mid-latitude atmospheric circulation where available potential energy is cascaded downward to smaller scales, and kinetic energy is inversely transferred upward toward larger scales. By inverting a three-dimensional elliptic differential operator, the model is first converted into a low-dimensional dynamical system, where the components correspond to different time scales. The information exchange between the scales is then computed through ensemble prediction. For this particular problem, the resulting cross-scale information flow is mostly from smaller scales to larger scales. That is to say, during this period, this model extratropical atmosphere is dominated by a bottom-up causation, as collective patterns emerge out of independent entities and macroscopic thermodynamic properties evolve from random molecular motions. This study makes a first step toward an important field in understanding the eddy–mean flow interaction in weather and climate phenomena such as atmospheric blocking, storm track, North Atlantic Oscillation, to name a few.

**Keywords:** causality; information flow; multiscale interaction; self-organization; storm; atmospheric jet stream; weather and climate patterns

## 1. Introduction

The atmospheric motion is rich in scale. In many cases, the formation of weather and climate patterns can be attributed to the interaction between a few scale ranges such as that between synoptic eddies and the jet stream, or that among the synoptic eddies, the low-frequency variability, and the mean flow, just as the storm track in the Northern Hemisphere (e.g., [1–3]), the blocking high (e.g., [4,5]), the sudden stratospheric warming, e.g., [6,7], and references therein), the North Atlantic Oscillation (e.g., [8–13]), to name but a few. This makes multiscale interaction a central issue in dynamic meteorology.

An important problem in multiscale interaction is how energy is transferred across scales; this transfer is closely related to the fundamental processes in atmospheric flows, namely, barotropic instability and baroclinic instability. The extratropical atmosphere is special in that, while on the whole the available potential energy is cascaded toward smaller scales, the kinetic energy is inversely transferred upward to larger scales (e.g., [14,15]). More specifically, there exists a symbiotic relationship between the synoptic-scale and planetary-scale disturbances. In a 3-scale setting, Cai and Mak [1] found that the former are produced and maintained through extracting energy from the zonal flow via Reynolds stress; they then supply energy to the latter through upscale energy transfer, while the latter form regions of enhanced baroclinicity where the former preferentially grow. This kind of energetic cycle has been employed to interpret many weather/climate phenomena. In the case of atmospheric blockings, for example, Holopainen and Fortelius [16] identified an enhanced transfer of eddy kinetic energy (KE) to the mean flow over the storm tracks. Hansen and Chen [17] found that the

nonlinear interaction between the cyclone-scale and planetary-scale waves is essential to the Atlantic blocking, while baroclinic amplification plays the most important role in the formation of the Pacific one. After the maintenance stage, the eddy kinetic energy is converted back to the eddy available potential energy (APE), leading to the decay of the event [18].

Another example is the inverse relationship between the boreal wintertime Pacific jet strength and storm-track intensity. It has long been observed that, in winters when the jet stream over the North Pacific is extremely strong, the storm track is however unexpectedly weak. This observation, which is at odds with the prediction of the classical linear baroclinic instability theory, has caught enormous attention from the atmospheric community. For example, based on energy budget diagnostics, Chang [19] and Nakamura and Sampe [20] found that, in that case, the synoptic waves tend to be trapped in the upper troposphere and hence are less efficient to tap APE from the background baroclinicity. In some other studies, it has been argued (e.g., [21,22]) that the inverse KE transfer may suppress the APE release, and hence contribute to the inverse relationship. Considering that atmospheric processes, cyclogenesis in particular, tend to be localized in space and time, Liang and Robinson [23] established, on the basis of the multiscale window transform of Liang and Anderson [24], a methodology for local multiscale atmospheric energetics analysis. The resulting interscale energy transfers turn out to bear a Lie bracket form, reminiscent of the Poisson bracket in Hamiltonian dynamics, and, furthermore, it naturally has the property that the energy redistributed among scales remains conserved (see [25]). To distinguish them, these energy transfers have been called "canonical transfers" ever since. With them, it is shown that, at each spatial location, there exists a local Lorenz cycle which allows one to trace the origin of the observed, if any, energy burst processes. Recently, by analyzing the local Lorenz cycles, it is found that, in boreal winters, the storms over the North Pacific actually are generated at latitudes far northward of the jet core [26]. This greatly lowers the relevance of the jet strength to the storm-track intensity, and hence the inverse relationship is actually not much a surprise. That is to say, the linear baroclinic instability theory may still hold [27], and the inverse relationship should be attributed to internal dynamics.

Energy and entropy are two parallel basic notions in physics. Naturally, the transfer of entropy, i.e., information transfer or information flow, across scales must make another fundamental problem in multiscale interaction. This problem, however, has been mostly overlooked in the past, mainly due to a lack of appropriate theory and research methodology. (So far, only a few studies have touched this issue, e.g., [28,29]). In physics, entropy may make an objective functional to be optimized to determine or regulate the distribution of energy. Without knowledge of the information flow across scales, the understanding of multiscale interaction is then incomplete. For example, in a classical energetics analysis, the emergence of a structure on another scale is driven by Reynolds stress or Reynolds stress-like quantities, which are essentially correlations between perturbation fields (e.g., [30]). While causation implies correlation, correlation does not necessarily imply causation ([31]; see below in Section 2. Thus, it is likely that an interaction which has a two-way energy flow may in fact have a one-way causation (as the example in the following; see below). We hence want to give this an investigation, in the hope of shedding some light on the dark side of the multiscale interaction problem. We emphasize that this is just the first step to exploring a rather profound field; it is by no means our intention to solve all the problems. In the following we first give a brief introduction of a recently developed theory of information flow, then introduce the atmospheric model (Section 3) and its solution. The model is reduced approximately to a low-dimensional dynamical system (Section 4), and with it the information flows across the scales are computed through ensemble prediction (Section 5). As we will see, remarkably, the system nearly has a bottom-up causation, consistent with how a reductionist in evolutionary biology would view the emergence of a higher-level organization out of independent, lower-level entities. This study is summarized in Section 6.

## 2. A Brief Review of the Theory of Causation and Information Flow That Pertains to This Study

Causal inference is a problem lying at the heart of science. In many disciplines, it makes a direct research objective. It is also an important topic in philosophy, as it forms a "guide to higher understanding" [32]. However, it is very challenging; in fact, it has been identified as "one of the biggest challenges" in the science of big data [33]. During the past thirty years, it has come to a consensus that the widely applicable notion in physics, namely, information flow or information transfer as it may appear in the literature, is logically associated with causality: The latter is the key to the former, while the former provides not only the magnitude but also the direction for the latter.

Realizing that information flow is a real physical notion, Liang and Kleeman [34] put the problem on a rigorous footing, and obtained in a closed form the information flow between the components of a 2D dynamical system. This formalism was soon generalized by Majda and Harlim [35] to a setting with two subspaces. Recently, it was successfully extended by Liang [36] to systems of arbitrary dimensionality. The following is a brief review of the work that pertains to this study.

We begin by stating a principle or an observational fact about causality:

> *If the evolution of an event, say, $X_1$, is independent of another one, $X_2$, then the causality from $X_2$ to $X_1$ is zero.*

Since it is the only quantitatively stated fact about causality, all previous empirical/half-empirical causality formalisms have attempted to verify it in applications. Considering its importance, it has been referred to as the **principle of nil causality** [36]. Recently, Smirnov [37] systematically examined the traditional formalisms, i.e., transfer entropy analysis and/or Granger causality testing, and found that they cannot verify the principle in a wide range of situations; similar conclusions have been drawn by Lizier and Prokopenko [38]. We will see soon below that, within our framework, this principle turns out to be a proven theorem.

Now, consider an $n$-dimensional continuous-time stochastic system for state variables $\mathbf{x} = (x_1, \ldots, x_n)$

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}}, \tag{1}$$

where $\mathbf{F} = (F_1, \ldots, F_n)$ may be arbitrary nonlinear functions of $\mathbf{x}$ and $t$, $\dot{\mathbf{w}}$ is a vector of white noise, and $\mathbf{B} = (b_{ij})$ is the matrix of perturbation amplitudes which may also be any functions of $\mathbf{x}$ and $t$. Here, we adopt the convention in physics and do not distinguish deterministic and random variables; in probability theory, they are usually distinguished with capital and lower-case symbols. Assume that $\mathbf{F}$ and $\mathbf{B}$ are both differentiable with respect to $\mathbf{x}$ and $t$. We then have the following theorem [36]:

**Theorem 1.** *For the system (1), the rate of information flowing from $x_j$ to $x_i$ (in nats per unit time) is*

$$
\begin{aligned}
T_{j \to i} &= -E\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{n-2}}\frac{\partial(F_i \rho_{\setminus j})}{\partial x_i}d\mathbf{x}_{\setminus i \setminus j}\right] + \frac{1}{2}E\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{n-2}}\frac{\partial^2(g_{ii}\rho_{\setminus j})}{\partial x_i^2}d\mathbf{x}_{\setminus i \setminus j}\right] \\
&= -\int_{\mathbb{R}^n}\rho_{j|i}(x_j|x_i)\frac{\partial(F_i\rho_{\setminus j})}{\partial x_i}d\mathbf{x} + \frac{1}{2}\int_{\mathbb{R}^n}\rho_{j|i}(x_j|x_i)\frac{\partial^2(g_{ii}\rho_{\setminus j})}{\partial x_i^2}d\mathbf{x},
\end{aligned} \tag{2}
$$

*where $d\mathbf{x}_{\setminus i \setminus j}$ signifies $dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_{j-1}dx_{j+1} \ldots dx_n$, $E$ stands for mathematical expectation, $g_{ii} = \sum_{k=1}^n b_{ik}b_{ik}$, $\rho_i = \rho_i(x_i)$ is the marginal probability density function (pdf) of $x_i$, $\rho_{j|i}$ is the pdf of $x_j$ conditioned on $x_i$, and $\rho_{\setminus j} = \int_{\mathbb{R}}\rho(\mathbf{x})dx_j$.*

If $T_{j \to i} = 0$, then $x_j$ is not causal to $x_i$; otherwise, it is causal, and the absolute value measures the magnitude of the causality from $x_j$ to $x_i$. For discrete-time mappings, the information flow is in a much more complicated form; see [36].

**Corollary 1.** [39] *When $n = 2$,*

$$T_{2 \to 1} = -E\left[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right] + \frac{1}{2}E\left[\frac{1}{\rho_1}\frac{\partial^2(g_{11}\rho_1)}{\partial x_1^2})\right]. \tag{3}$$

In the absence of noise, this is precisely the result of [34] based on a heuristic argument.
There is a nice property for the above information flow:

**Theorem 2.** (Principle of nil causality) *If in Equation (1) neither $F_1$ nor $g_{11}$ depends on $X_2$, then $T_{2 \to 1} = 0$.*

Note that this is precisely the principle of nil causality. Remarkably, here it appears as a proven theorem, while the classical ansatz-like formalisms fail to verify it in many problems (e.g., [37]).

In the case with only two time series (no dynamical system is given), we have the following result [31]:

**Theorem 3.** *Given two time series $X_1$ and $X_2$, under the assumption of a linear model with additive noise, the maximum likelihood estimator (mle) of the rate of information flowing from $X_2$ to $X_1$ is*

$$\hat{T}_{2\to1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2},$$

(4)

*where $C_{ij}$ is the sample covariance between $X_i$ and $X_j$, and $C_{i,dj}$ the sample covariance between $X_i$ and a series derived from $X_j$ using the Euler forward differencing scheme: $\dot{X}_{j,n} = (X_{j,n+k} - X_{j,n})/(k\Delta t)$, with $k \geq 1$ some integer.*

Equation (4) is rather concise in form; it only involves the common statistics, i.e., sample covariances. In other words, a combination of some sample covariances will give a quantitative measure of the causality between the time series. This makes causality analysis, which otherwise would be complicated with the classical empirical/half-empirical methods, very easy. Nonetheless, note that Equation (4) cannot replace (3); it is just the mle of the latter. A statistical significance test must be performed before a causal inference is made based on the computed $T_{2\to1}$. For details, refer to [31].

Considering the long-standing debate ever since George Berkeley in 1710 over correlation versus causation, we may rewrite (4) in terms of linear correlation coefficients, which immediately implies [31]:

*Causation implies correlation, but correlation does not imply causation.*

In fact, suppose there is no correlation between $X_1$ and $X_2$, $C_{12} = 0$; then, by Equation (4), $\hat{T}_{2\to1} = 0$. However, from $\hat{T}_{2\to1} = 0$, one cannot come up with $C_{12} = 0$.

Causality can be normalized in order to reveal the relative importance of a causal relation. However, the normalization is by no means as trivial as that for covariance, considering that information flow is asymmetric in direction ($T_{2\to1} \neq T_{1\to2}$ in general), and, in addition, there is no such property like a Cauchy–Schartz inequality that makes it possible for covariance to be normalized. In [40], a way of normalization is given, but a complete solution is yet to be sought.

The above formalism has been validated with many benchmark systems (e.g., [36]) such as Baker transformation, Hénon map, Kaplan–Yorke map, and Rössler system, to name a few. Particularly, Equation (4) has been validated with touchstone problems where the traditional Granger causality test and transfer entropy analysis fail. An example is the highly chaotic anticipatory system problem described in [41], which with Equation (4) turns out not to be a problem at all.

The formalism has been successfully applied to the studies of many real world problems, among them are the causal relation between El Niño-Indian Ocean Dipole [31], tropical cyclone genesis prediction [42], near-wall turbulence [28], global climate change ([43,44]), and financial time series analysis [40], to name but a few. Here, we particularly want to mention the study by Stips et al. [43] who, through examining with Equation (4) the causality between the $CO_2$ index and the surface air temperature, identified a reversing causal relation with time scale. They found, during the past century, that $CO_2$ emission indeed drives the recent global warming; the causal relation is one-way, i.e., from $CO_2$ to global mean atmosphere temperature. Moreover, they were able to find how the causality is distributed over the globe, thanks to the quantitative nature of our formalism. However, on a time scale of 1000 years or over, the causality is completely reversed; that is to say, on a paleoclimate scale, it is global warming that causes $CO_2$ concentration to rise!

## 3. A Quasi-Geostrophic Atmospheric Model

### 3.1. The Governing Equation

Consider a three-dimensional (3D) quasi-geostrophic (QG) model on a $\beta$-plane within a channel between latitudes $y = \pm1$ (cf. [45,46]):

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = -r\zeta + A_H \nabla_H{}^2 \zeta,$$

(5)

where $\psi$ is streamfunction,

$$\zeta = \nabla_H{}^2\psi + \frac{\partial}{\partial z}\left(\frac{F_r^2}{N^2}\frac{\partial\psi}{\partial z}\right) \tag{6}$$

potential vorticity, $J$ the Jacobian operator such that $J(\psi,\zeta) = \frac{\partial\psi}{\partial x}\frac{\partial\zeta}{\partial y} - \frac{\partial\zeta}{\partial x}\frac{\partial\psi}{\partial y}$, $N$ the buoyancy frequency, and $F_r$ the rotational internal Froude number. On the right-hand side, the first term stands for the Rayleigh-type friction and the second is the horizontal dissipation of the potential vorticity. In this study, these two terms are set to be zero ($r = A_H = 0$). This equation together with the following boundary conditions:

$$\psi = \text{const} \qquad y = \pm 1, \tag{7}$$

$$\psi \text{ is periodic in } x, \tag{8}$$

$$\frac{\partial}{\partial t}\left(\frac{F_r^2}{N^2}\frac{\partial\psi}{\partial z}\right) + J\left(\psi, \left(\frac{F_r^2}{N^2}\frac{\partial\psi}{\partial z}\right)\right) = \begin{cases} 0, & z = 1, \\ -J(\psi, b), & z = 0, \end{cases} \tag{9}$$

where $b$ is the bottom relief, forms the problem that we are about to solve. In Equation (9), at $z = 1$, the atmosphere is taken as a rigid lid (vertical velocity $w(z = 1) = 0$); at $z = 0$, a flat bottom is considered and hence $b = 0$.

We choose to solve the equation for the perturbation around the mean state $\bar{\psi}$. The mean state is a zonally homogeneous jet

$$\bar{U} = \bar{U}(y, z). \tag{10}$$

It is easy to verify that it satisfies the QG equation for an ideal fluid. Here, $\bar{U}$ horizontally is assumed to be a cosine jet within $[-L, L]$, $L \leq 1$, as that in [47],

$$\bar{U} = \mathscr{L}(z) \cdot \cos^2\frac{\pi y}{2L} = \frac{1 + \cos\frac{\pi}{L}y}{2} \cdot \mathscr{L}(z), \qquad y \in [-L, L]; \tag{11}$$

outside $[-L, L]$, the fluid is motionless. In this study, $L$ is chosen to be 1, $\mathscr{L}(z)$ is prescribed such that $\frac{\partial\bar{U}}{\partial z} = 0$ near $z = 0$ and $z = 1$, and is maximized in the upper troposphere. From $\bar{U}$, the mean states

$$\bar{\psi}(y, z) = -\int_{-1}^{y} \bar{U}(y, z)dy, \tag{12}$$

$$\bar{\zeta}(y, z) = \bar{\psi}_{yy} + \frac{\partial}{\partial z}\left(\frac{F_r^2}{N^2}\frac{\partial\bar{\psi}}{\partial z}\right) \tag{13}$$

can be easily obtained, and

$$\frac{\partial\bar{\zeta}}{\partial y} = -\frac{\partial S}{\partial z}\frac{\partial\bar{U}}{\partial z} - S\frac{\partial^2\bar{U}}{\partial z^2} - \frac{\partial^2\bar{U}}{\partial y^2} = -S_z\bar{U}_z - S\bar{U}_{zz} - \bar{U}_{yy}, \tag{14}$$

where we have shortened $F_r^2/N^2$ as $S(z)$.

Let $\psi = \bar{\psi} + \psi'$, $u = \bar{U} + u'$, $\zeta = \bar{\zeta} + \zeta'$. The perturbation equation is

$$\frac{\partial\zeta'}{\partial t} + (\bar{U} + u')\frac{\partial\zeta'}{\partial x} + v'\frac{\partial\zeta'}{\partial y} + [\beta - (S_z\bar{U}_z + S\bar{U}_{zz} - \bar{U}_{yy})]\frac{\partial\psi'}{\partial x} = 0. \tag{15}$$

Correspondingly, the boundary conditions are changed to: $\psi'$ vanishes at $y = \pm 1$, and

$$\frac{\partial S\psi'_z}{\partial t} + (\bar{U} + u')\frac{\partial(}{\partial z}S\psi'_z) + v'S\bar{U}_z + v'\frac{\partial S\psi'_z}{\partial y} = 0 \tag{16}$$

at $z = 0, 1$. Note in this study $\bar{U}_z = 0$ at $z = 0, 1$, hence the third term vanishes. Thus, this is simply the horizontal advection of $S\psi'_z$. If initially $\psi'_z = 0$, and there is no energy flux toward the upper and lower

boundaries, it will remain unperturbed. To simplify the problem, we then set the vertical boundary condition as

$$\frac{\partial \psi'}{\partial z} = 0, \qquad \text{at } z = 0, 1. \tag{17}$$

As we will see later, though with such a rather strong condition, the result does reproduce the expected energetics typical of the large-scale mid-latitude atmospheric motion.

### 3.2. Model Setup

In this study, we choose a mesh with spacings of $\Delta x = 0.2$, $\Delta y = 0.04$, and $\Delta z = 0.2$, which results in a grid with $50 \times 51 \times 5$ points. Choose a vertical profile for the basic flow $\mathscr{U}(z_k) = (0.2, 0.2, 0.6, 1, 1)$ such that $\partial \bar{U} / \partial z = 0$ at $z = 0$ ($k = 1$) and $z = 1$ ($k = 5$). To determine $S(z) = F_r^2 / N^2$, first notice that $F_r = f_0 L_0 / N_0 H_0$ is the rotational Froude number; usually, it is taken as 1. We hence only need to pay attention to the buoyancy frequency $N$. Scale it by $N_0$. In dimensional form, it is defined as

$$N_0 N = \left( -\frac{g}{\rho} \frac{\partial \bar{\rho}}{\partial z} \right)^{1/2}, \tag{18}$$

where $\rho$ is the density of the fluid. While for oceans this can be directly computed, for the atmosphere, it is usually converted into another form

$$N_0 N = \left( \frac{g}{\bar{\theta}} \frac{\partial \bar{\theta}}{\partial z} \right)^{1/2} = \left( g \frac{\partial \log \bar{\theta}}{\partial z} \right)^{1/2}. \tag{19}$$

Here, $\bar{\theta}$ is potential temperature; it is the temperature of an air parcel moving adiabatically to some reference pressure (usually 1000 hPa), i.e., a temperature with the effect of pressure change excluded:

$$\bar{\theta} = \bar{T} \left( \frac{P_0}{\bar{P}} \right)^{R/c_p} \equiv \bar{T} \left( \frac{1000}{\bar{P}} \right)^{\kappa_d}, \tag{20}$$

where $\kappa_d \approx 0.286$, $R = 8.314 \, \text{J/mol} \cdot K = 287 \, \text{J/kg} \cdot K$ is the ideal gas constant, and $c_p = 1004 \, \text{J/kg} \cdot K$ the specific heat capacity at constant pressure. This yields

$$N_0^2 N^2 = \frac{g}{\bar{T}} \left( \frac{\partial \bar{T}}{\partial z} + \frac{g}{c_p} \right), \tag{21}$$

where $-\frac{\partial \bar{T}}{\partial z}$ is the lapse rate.

For the troposphere, usually the lapse rate can be roughly taken as a constant $\frac{\partial \bar{T}}{\partial z} \approx -0.65 \,°\text{C}/100 \, \text{m} = 6.5 \times 10^{-3} \,°\text{C/m}$. That is to say,
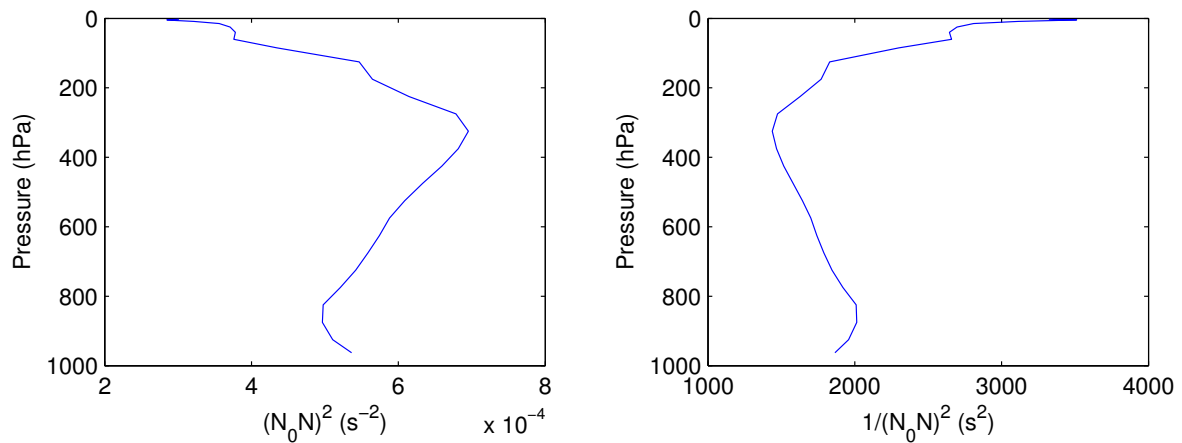
$$\bar{T} = \bar{T}_0 - 6.5 \times 10^{-3} z. \tag{22}$$

Thus,

$$\begin{aligned} N^2(z) &= \frac{g}{\bar{T}} \left( \frac{\partial \bar{T}}{\partial z} + \frac{g}{c_p} \right) \\ &\approx \frac{0.032}{\bar{T}_0 - 6.5 \times 10^{-3} z}. \end{aligned} \tag{23}$$

In Figure 1, the vertical profile of $N_0^2 N^2$ for the atmosphere between latitudes $20°$ N–$60°$ N is computed. Furthermore, its reciprocal $N_0^{-2} N^{-2}$ is shown.

**Figure 1.** Vertical profile of $N_0^2 N^2$ for the atmosphere between $20°$ N–$60°$ N.

By Figure 1, $N_0^2 N^2$ varies from $5 \times 10^{-4}$ s$^{-2}$ near the bottom to $7 \times 10^{-4}$ s$^{-2}$ at tropopause. If normalized by $N_0^2$, this means it varies from 1 to 1.4. Considering that the lapse rate $-\frac{\partial \bar{T}}{\partial z}$ is nearly a constant $0.65 °C/100$ m, $N_0^2/N^2 = S(z)$ $(F_r^2 = 1)$ and hence almost decreases linearly with $z$, with a rate of $(1 - \frac{1}{1.4})/5 = 0.057$ per level. In a brief summary, Table 1 gives a list of the parameters for the model.

**Table 1.** Model parameters.

| | |
|---|---|
| Mesh | $50 \times 51 \times 5$ |
| $\Delta x$ | 0.2 |
| $\Delta y$ | 0.04 |
| $\Delta z$ | 0.2 |
| $\Delta t$ | 0.02 |
| $\beta$ | 1 |
| $\mathscr{X}(z_k)$ | 0.2, 0.2, 0.6, 1, 1 |
| $S(z_k)$ | 0.943, 0.886, 0.828, 0.771, 0.714 |

*3.3. Solution Strategy*

The problems (15)–(17) is solved with a leap-frog scheme. To suppress the computational mode that may arise due to the time splitting, the result at each integration step is filtered in time as follows:

$$\psi^n = (1 - \alpha)\psi^n + \frac{\alpha}{2}(\psi^{n+1} + \psi^{n-1}) \tag{24}$$

with a weak filter coefficient $\alpha = 0.01$. This is equivalent to a weak dissipation of the flow. At each step, it is required to solve the 3D elliptic equation

$$\nabla_H{}^2 \psi + \frac{\partial}{\partial z}\left(S(z)\frac{\partial \psi}{\partial z}\right) = \zeta \tag{25}$$

for $\psi$ subject to

$$\begin{cases} \psi_z(z = 1) = 0, \\ \psi_z(z = 0) = 0, \end{cases} \tag{26}$$

in the vertical, a no-flux condition in $y$, and a periodic condition in $x$. We may separate $z$ from $(x, y)$ to convert the 3D equation to a set of 2D equations. Separation of variables results in an eigenvalue problem

$$\frac{d}{dz}\left(S(z)\frac{d\theta_k(z)}{dz}\right) = \lambda_k \theta_k(z) \tag{27}$$

together with a boundary condition

$$\frac{d\theta_k}{dz} = 0, \qquad \text{at } z = 0, 1. \tag{28}$$

This is a Sturm–Liouville problem (cf. [48]), and it can be proven that the resulting eigenvectors $\{\theta_k\}$ form an orthogonal set. The set is complete and can be normalized. Thus, it can be made as an orthonormal basis. Expand $\psi'$ and $\zeta'$ (time-dependence suppressed for notational simplicity) with the basis:

$$\psi'(x,y,z) = \sum_{k=1}^{n} \tilde{\Phi}_k(x,y) \cdot \theta_k(z), \tag{29}$$

$$\zeta'(x,y,z) = \sum_{k=1}^{n} \tilde{Z}_k(x,y) \cdot \theta_k(z), \tag{30}$$

where $n$ is the number of levels (of the discretized model) in the vertical direction, and substitute them into the original Equation (25) to get

$$\sum_{k=1}^{n} \nabla_H{}^2 \tilde{\Phi}_k \cdot \theta_k(z) + \sum_{k=1}^{n} \tilde{\Phi}_k \frac{\partial}{\partial z}\left[S(z)\frac{\partial \theta_k(z)}{\partial z}\right] = \sum_{k=1}^{n} \tilde{Z}_k(x,y)\theta_k(z). \tag{31}$$

By the orthonormality of $\{\theta_k\}$, the original 3D equation is transformed into $n$ decoupled 2D equations:

$$\nabla_H{}^2 \tilde{\Phi}_k + \lambda_k \tilde{\Phi}_k = \tilde{Z}_k, \tag{32}$$

which can be solved individually.

The eigenvalue problems (27)–(28) are solved with the parameters as listed in Table 1. The resulting eigenvalues $\lambda_k$, $k = 1, \ldots, 5$, are all negative; see Figure 2. The corresponding eigenvectors $\theta_k$ are displayed in Figure 3; it is easy to verify that they are orthonormal (cf. [48]).

When $N^2$ is constant, the problem can be solved analytically (e.g., [49]). In that case, the most rudimentary mode is barotropic and the remaining ones baroclinic. In this case, $\theta_1$ is the very barotropic mode, and $\theta_3$, $\theta_4$, $\theta_5$ are approximately sinusoidal and hence are just like the baroclinic modes. Here, as $N^2$ varies with height, mode 2 somehow has a different form. Note that $c_k = 1/\sqrt{-\lambda_k}$ corresponds to the phase speed of mode k.
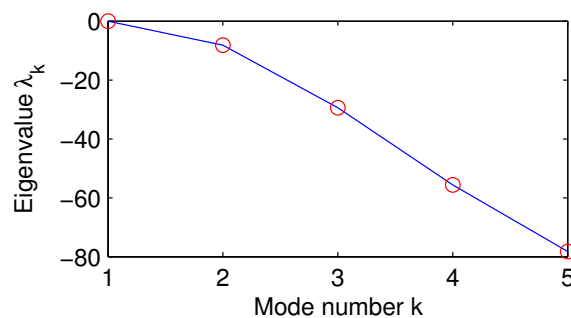


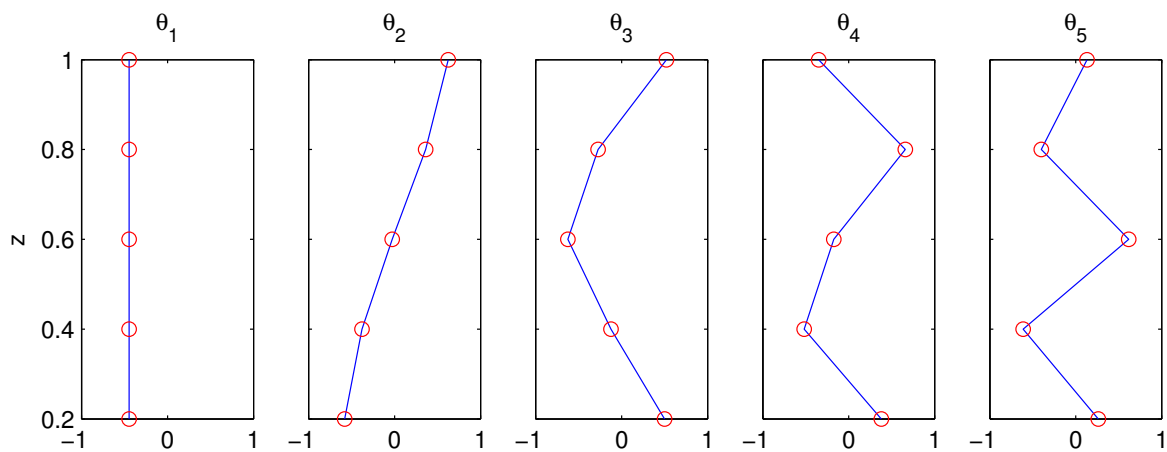**Figure 2.** The computed eigenvalues $\lambda_k$.

**Figure 3.** The eigenvectors $\theta_k(z)$ corresponding to the eigenvalues $\lambda_k$ in Figure 2.

*3.4. Initialization*

Different initial disturbances will in general yield different solutions, many of which may be eventually damped out. In order to obtain a quasi-equilibrium oscillatory state, there exists some "optimal perturbation." To find this, linearize Equation (15) to get

$$\frac{\partial \zeta'}{\partial t} = -\bar{U}\frac{\partial}{\partial x}\left[\nabla_H^2 \psi' + \frac{\partial}{\partial z}\left(S\frac{\partial \psi'}{\partial z}\right)\right] - \left(\frac{\partial \bar{\zeta}}{\partial y} + \beta\right)\frac{\partial \psi'}{\partial x}. \tag{33}$$

This together with the boundary condition

$$\begin{cases} \psi' = 0 & \text{at } y = \pm 1, \\ \psi' \text{ is periodic in } x, \\ \frac{\partial \psi'}{\partial z} = 0 & \text{at } z = 0, 1, \end{cases} \tag{34}$$

forms a linear system. Write (33) and (34) as

$$\frac{\partial \zeta'}{\partial t} \equiv \mathscr{L}_\zeta \psi', \tag{35}$$

and write $\zeta'$ together with (34) as

$$\zeta' = \nabla_H^2 \psi' + \frac{\partial}{\partial z}\left(S\frac{\partial \psi'}{\partial z}\right) \equiv \mathscr{L}\psi'. \tag{36}$$

Then, the linearized perturbation equation becomes

$$\frac{\partial}{\partial t}\mathscr{L}\psi' = \mathscr{L}_\zeta \psi'. \tag{37}$$

As $\mathscr{L}$ is linear and independent from $t$, it commutes with $\partial/\partial t$. We hence obtain the following linear dynamical system for the perturbation field $\psi'$:

$$\frac{\partial \psi'}{\partial t} = \left(\mathscr{L}^{-1} \circ \mathscr{L}_\zeta\right)\psi'. \tag{38}$$

The discretized version of (38) is denoted as:

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \tag{39}$$

where **u** is a vector of the values of $\psi'$. Initialized with a vector $\mathbf{u}_0$, its solution is

$$\mathbf{u} = e^{\mathbf{A}t}\mathbf{u}_0. \tag{40}$$

The optimal perturbation corresponds to the largest singular value of the matrix $e^{\mathbf{A}t}$. To see this, it suffices to consider one particular time, say, $t = 1$. Perform a singular value decomposition of $e^{\mathbf{A}\cdot 1}$. Perturbations of the modal forms corresponding to singular values greater than 1 will grow. Here, the singular values become smaller than 1 after the modal number $m > 4333$, as shown in Figure 4. In order to have the disturbance grow, we need to choose the modes with numbers lower than 4333.
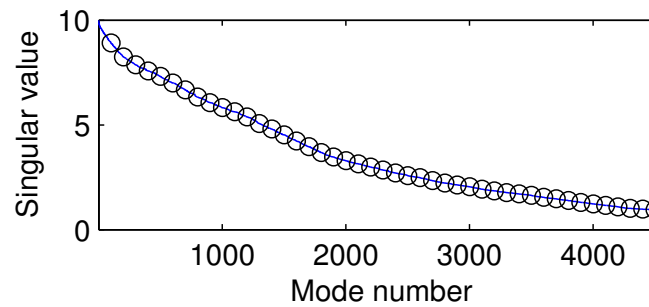


**Figure 4.** Singular value versus modal number.

Displayed in Figures 5 and 6 are modes 2, 100, 500, and 1000 (mode 1 is trivial). Clearly, they have different structures in both horizontal and vertical directions. Theoretically, the lower the modes, the more efficient the perturbation. However, since this is just a linear solution, the evolution after the initial perturbation may not grow as expected after nonlinearity takes effect. Here, we find that modes 100, 500, and 1000, among others, are satisfactory. In the following, we choose mode 1000 as the perturbation to initialize the system.
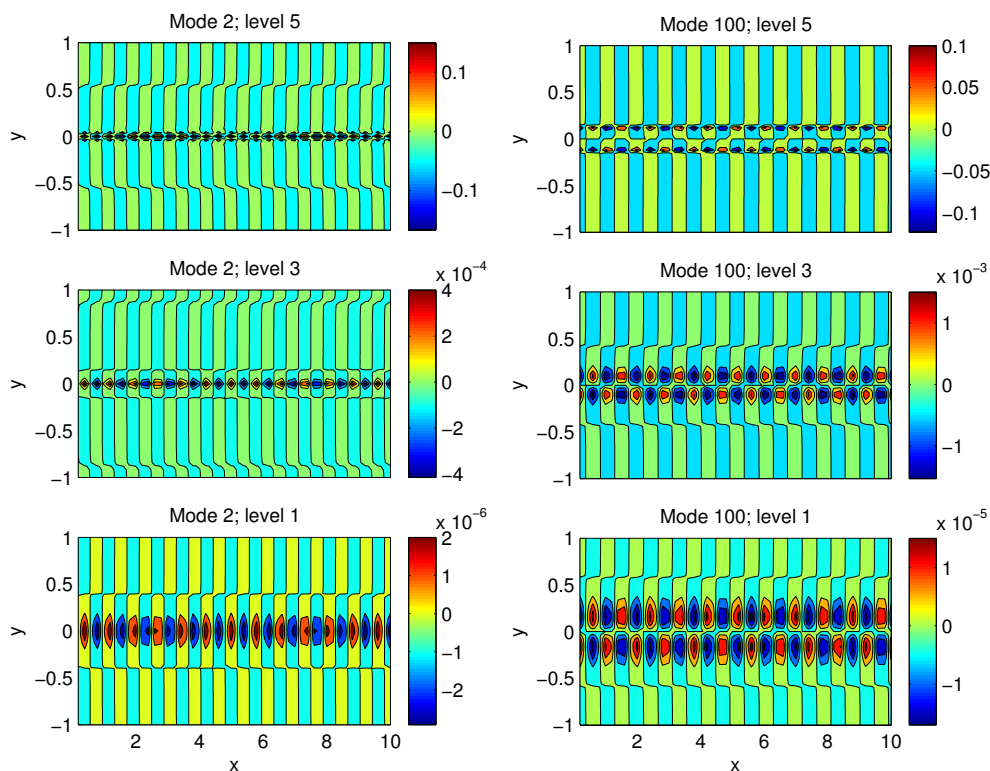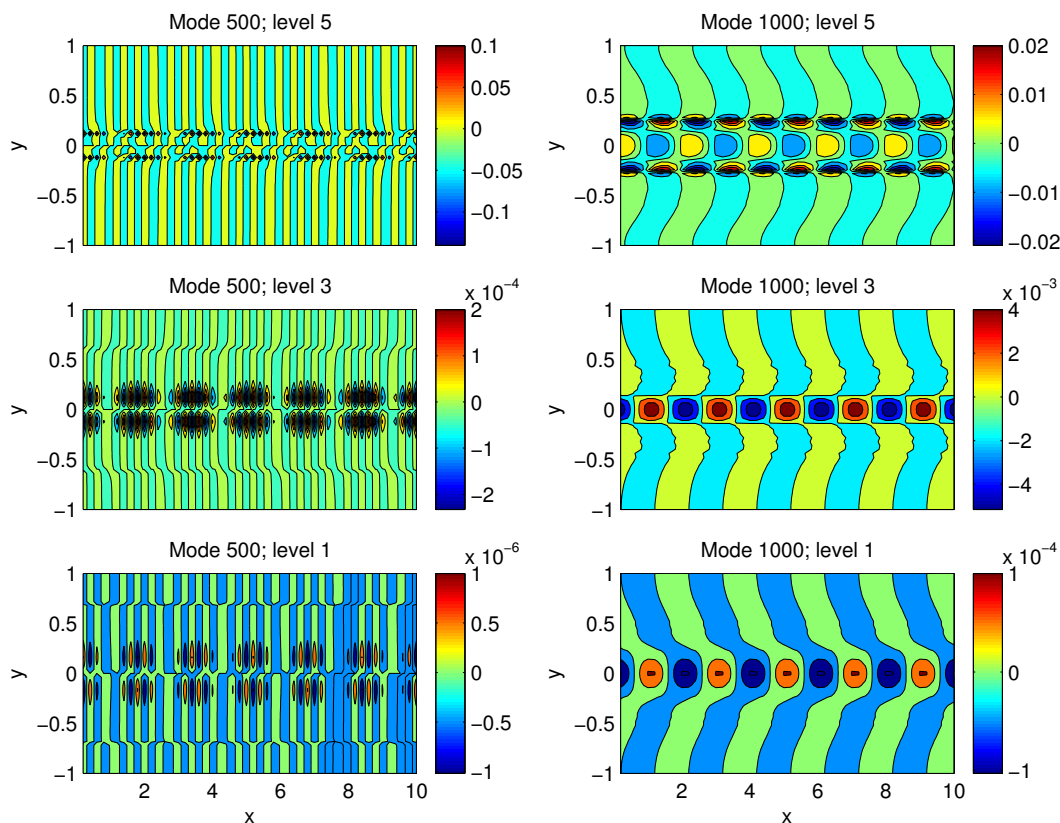


**Figure 5.** Perturbation modes.

**Figure 6.** Perturbation modes (cont'd).

## 4. Reduced Model

### 4.1. Results of the Quasi-Geostrophic Model

After initialization, the model reaches a quasi-equilibrium after some 400,000 steps. Figure 7 shows the evolution of total perturbation kinetic energy. Consider for our purpose the time interval between steps 500,000 and 650,000. We choose such an interval because (1) it is not too large; otherwise, too many processes may be involved and hence the model cannot be reduced much, (2) the processes during the interval appear to be stationary. Another reason is that the energetic cycle mimicks well that in the mid-latitude atmosphere. To see this, apply a multiscale window transform (MWT) to separate the process. MWT is a functional analysis tool developed by Liang and Anderson [24] which can decompose a function space into a direct sum of orthogonal subspaces, each with an exclusive range of scales, while preserving the local properties of the functions. Such a subspace is called a "scale window". Originally, MWT is developed for a physically consistent expression of multiscale quadratic quantities, and hence to make multiscale energetics analysis possible. Traditionally, filters have been widely used for multiscale studies in atmospheric research, but, in a rigorous sense, the traditional filters are generally incapable of representing multiscale energy, which is a concept in phase space (it is connected to energy in the physical sense thanks to the Parseval identity), while the outputs of traditional filters are fields in physical space. Liang and Anderson [24] found that, for a class of specially designed orthogonal filters, there exists a transform-reconstruction pair, i.e., a pair of MWT and multiscale window reconstruction (MWR), just as Fourier transform and inverse Fourer transform. An MWR is just like a filtered quantity, while the corresponding MWT coefficient squared gives the energy on the scale window of concern. With MWT and MWR, it has been established that, for an atmospheric/oceanic flow, at each location, there exists a local Lorenz cycle consisting of three conservative processes. The resulting energy transfers have been referred to as canonical transfers; they all bear a Lie bracket form, in contrast to the classical emprically-obatained energy transfers. A comprehensive introduction of the theory is beyond the scope of this study; for details, see [25].

Now, perform a two-scale window decomposition, and choose the longest scale to be the whole interval. (This can be done by setting the lowest scale window index to be zero; see [24]). With this

setting, compute the canonical transfers using the localized multiscale energetics of Liang (2016a) [25]. The horizontally integrated canonical transfers are shown in Figure 8, where positive values indicate a transfer of energy from the mean to the perturbation. Clearly, here the transfer of available potential energy overwhelmingly dominates that of kinetic energy (two orders larger), and, on the whole, the former is cascaded downward (left panel), while the latter is transferred inversely upward (right panel). This seems to agree with what has been observed in the mid-latitude atmosphere (e.g., [14,15]).
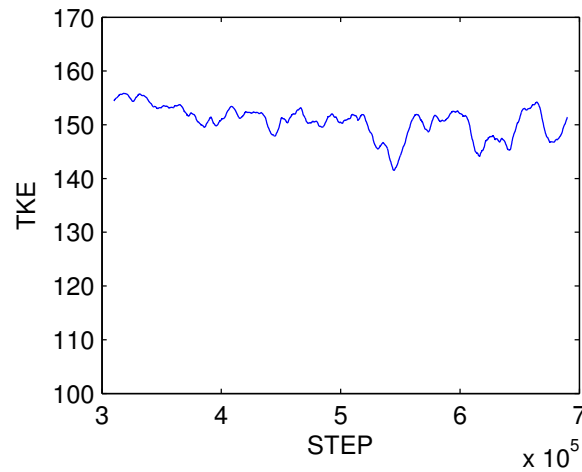


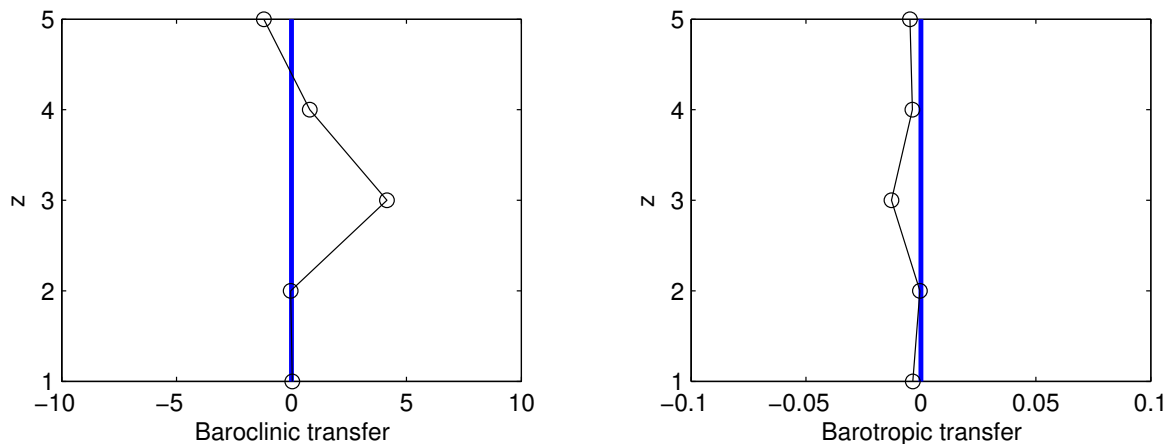**Figure 7.** Evolution of total perturbation kinetic energy.



**Figure 8.** Barotropic and baroclinic canonical transfers from the mean state to the perturbation field.

*4.2. Principal Component Analysis*

With the modeled 4D field $\psi(x, y, z; t)$ on the chosen time interval, we perform a principal component (PC) analysis, or empirical orthogonal function (EOF) analysis as it is called. The eigenvalues $\lambda$ are shown in Figure 9a. Obviously, the first three modes possess most of the variance. Displayed in Figure 9b are some of the corresponding PCs. It appears that the first and second PCs approximately are in quadrature phase; they should form a harmonic subsystem. The third and the fourth have similar frequencies, though that of the latter is a little higher.
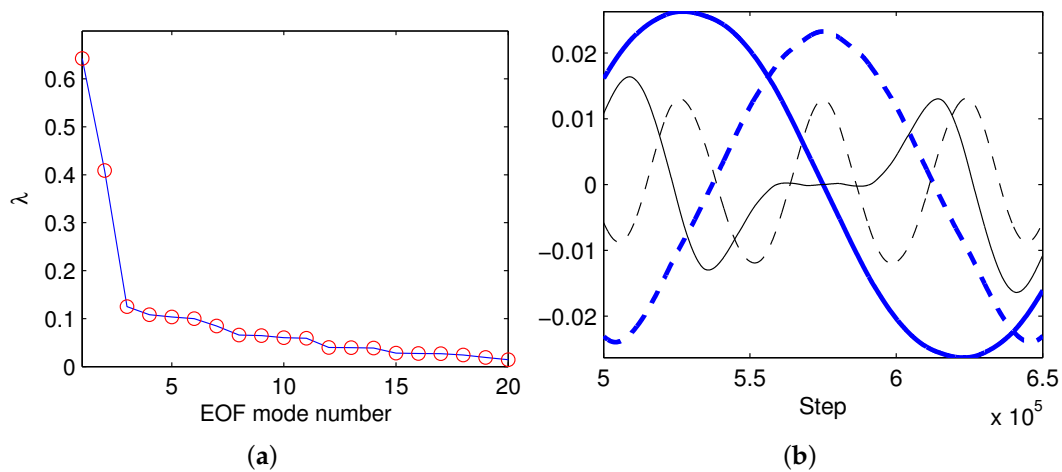
**Figure 9.** (**a**) eigenvalue $\lambda$ vs. EOF mode number; (**b**) first four principal components: mode 1 (thick solid), mode 2 (thick dashed), mode 3 (thin solid), and mode 4 (thin dashed).

### 4.3. Model Reduction

The EOF modes form an orthonormal basis for $\psi$. In this subsection, we use the basis to reduce the original governing Equation (15) into a low-dimensional dynamical system.

With the operator $\mathscr{L}$ as defined in (36), Equation (15) becomes

$$\frac{\partial \psi'}{\partial t} = -\mathscr{L}^{-1}[J(\psi', \mathscr{L}\psi')] - \mathscr{L}^{-1}\left[U\frac{\partial \mathscr{L}\psi'}{\partial x}\right] - \mathscr{L}^{-1}\left[(\beta + \bar{\zeta}_y)\frac{\partial \psi'}{\partial x}\right]. \tag{41}$$

In the equation, $\mathscr{L}^{-1}$ is the inverse of $\mathscr{L}$. Expand $\psi$ with $\{e_m\}_{m=1,2,\ldots}$ and truncate at $m = 4$ to get:

$$\psi = \sum_{m=1}^{4} p_m(t)e_m(x, y, z). \tag{42}$$

Substitution of this expansion into (41) yields

$$\sum_m \frac{dp_m}{dt}e_m =$$
$$-\mathscr{L}^{-1}[J(\sum_i p_i e_i, \mathscr{L}\sum_j p_j e_j)] - \mathscr{L}^{-1}U\frac{\partial \mathscr{L}\sum_i p_i e_i}{\partial x} - \mathscr{L}^{-1}[\beta + \bar{\zeta}_y]\frac{\partial \sum_i p_i e_i}{\partial x}$$
$$= -\sum_i\sum_j p_i p_j \mathscr{L}^{-1}[J(e_i, \mathscr{L}e_j)] - \sum_i p_i \mathscr{L}^{-1}\left(U\frac{\partial \mathscr{L}e_i}{\partial x}\right) - \sum_i p_i \mathscr{L}^{-1}[(\beta + \bar{\zeta}_y)\frac{\partial e_i}{\partial x}]. \tag{43}$$

Since $\{e_m\}$ is orthonormal, taking an inner product on both sides with $e_m$ results in

$$\begin{aligned}
\frac{dp_m}{dt} &= -\sum_i\sum_j p_i p_j \left\langle e_m, \mathscr{L}^{-1}[J(e_i, \mathscr{L}e_j)]\right\rangle \\
&\quad -\sum_i p_i \left\langle e_m, \mathscr{L}^{-1}\left(U\frac{\partial \mathscr{L}e_j}{\partial x}\right)\right\rangle \\
&\quad -\sum_i p_i \left\langle e_m, \mathscr{L}^{-1}\left[(\beta + \bar{\zeta}_y)\frac{\partial e_i}{\partial x}\right]\right\rangle \\
&= \sum_i\sum_j \alpha^N_{m,i,j} p_i p_j + \sum_i \alpha^L_{m,i} p_i, \qquad m, i, j = 1, \ldots, 4,
\end{aligned} \tag{44}$$

where

$$\alpha^N_{m,i,j} = -\left\langle e_m, \mathscr{L}^{-1}[J(e_i, \mathscr{L}e_j)]\right\rangle \tag{45}$$

are the quadratic term (nonlinear) coefficients, and

$$\alpha^L_{m,i} = -\left\langle e_m, \, \mathscr{L}^{-1}\left[U\frac{\partial \mathscr{L}e_i}{\partial x} + (\beta + \bar{\zeta}_y)\frac{\partial e_i}{\partial x}\right]\right\rangle \tag{46}$$

are the linear term coefficients.

By computation, the linear coefficients $\alpha^L_{m,i}$ are

| | | | | |
|---|---|---|---|---|
| $\alpha^L_{1,i}$ | $-3.4755100 \times 10^{-2}$ | $1.423829$ | $-1.3951972 \times 10^{-2}$ | $0.1284307$ |
| $\alpha^L_{2,i}$ | $-1.417942$ | $-1.3652847 \times 10^{-2}$ | $-0.1443377$ | $-3.5396349 \times 10^{-3}$ |
| $\alpha^L_{3,i}$ | $-1.3292501 \times 10^{-2}$ | $0.1588890$ | $2.1233991 \times 10^{-2}$ | $0.4809141$ |
| $\alpha^L_{4,i}$ | $-0.1304244$ | $-6.8064928 \times 10^{-3}$ | $-0.4592572$ | $7.3776743 \times 10^{-3}$ |

Likewise, from Equation (45), the coefficients for the quadratic terms are computed as follows:

| $\alpha^N_{1,i,j}$ $(m=1)$ | | | |
|---|---|---|---|
| $-7.5820560 \times 10^{-4}$ | $-3.2869954 \times 10^{-2}$ | $3.5282248 \times 10^{-3}$ | $4.9599465 \times 10^{-3}$ |
| $4.1258920 \times 10^{-2}$ | $1.6017430 \times 10^{-3}$ | $4.2935433 \times 10^{-3}$ | $4.6059955 \times 10^{-5}$ |
| $-2.4720350 \times 10^{-2}$ | $1.6584761 \times 10^{-2}$ | $-9.0469969 \times 10^{-3}$ | $7.8747235 \times 10^{-3}$ |
| $-2.3384064 \times 10^{-2}$ | $-9.9051173 \times 10^{-4}$ | $-1.1057421 \times 10^{-2}$ | $-5.0229015 \times 10^{-4}$ |

| $\alpha^N_{2,i,j}$ $(m=2)$ | | | |
|---|---|---|---|
| $-2.2212272 \times 10^{-3}$ | $-1.6904000 \times 10^{-2}$ | $-2.1221174 \times 10^{-3}$ | $-1.2588169 \times 10^{-4}$ |
| $2.0407232 \times 10^{-2}$ | $-1.0814944 \times 10^{-3}$ | $3.0389655 \times 10^{-3}$ | $1.2576354 \times 10^{-3}$ |
| $-3.3598884 \times 10^{-3}$ | $-2.4367530 \times 10^{-2}$ | $2.1593377 \times 10^{-3}$ | $4.7868756 \times 10^{-3}$ |
| $-1.0774944 \times 10^{-2}$ | $-4.2805336 \times 10^{-3}$ | $-4.0617036 \times 10^{-3}$ | $-2.7537413 \times 10^{-4}$ |

| $\alpha^N_{3,i,j}$ $(m=3)$ | | | |
|---|---|---|---|
| $-9.9601485 \times 10^{-3}$ | $-7.9435982 \times 10^{-2}$ | $-1.4256314 \times 10^{-2}$ | $-6.8288655 \times 10^{-3}$ |
| $8.4469259 \times 10^{-2}$ | $-1.6651559 \times 10^{-3}$ | $6.7729242 \times 10^{-2}$ | $-2.0513053 \times 10^{-3}$ |
| $1.2046070 \times 10^{-2}$ | $-6.1348621 \times 10^{-2}$ | $-7.0710764 \times 10^{-3}$ | $9.7337542 \times 10^{-3}$ |
| $-1.4477096 \times 10^{-3}$ | $1.6220149 \times 10^{-2}$ | $-4.4799279 \times 10^{-3}$ | $3.3035765 \times 10^{-3}$ |

| $\alpha^N_{4,i,j}$ $(m=4)$ | | | |
|---|---|---|---|
| $-3.0739678 \times 10^{-3}$ | $-0.1182990$ | $1.2369854 \times 10^{-2}$ | $1.4575672 \times 10^{-2}$ |
| $0.1257915$ | $-2.2905001 \times 10^{-3}$ | $-2.0050334 \times 10^{-2}$ | $9.9308938 \times 10^{-3}$ |
| $-4.3914612 \times 10^{-2}$ | $5.8852371 \times 10^{-2}$ | $9.3064429 \times 10^{-3}$ | $-3.0944983 \times 10^{-2}$ |
| $-4.8473705 \times 10^{-2}$ | $-4.0869847 \times 10^{-2}$ | $7.3413953 \times 10^{-2}$ | $-2.7747510 \times 10^{-3}$ |

Note that in reconstructing $\psi$ there is actually a mean part $\bar{\psi} \equiv p_0$ to be added. That is to say,

$$\psi = p_0 + \sum_i p_i e_i.$$

Theoretically, this part should vanish in the system, but, in reality, it may not. If it is added to Equation (43), then

$$\sum_m \frac{dp_m}{dt}e_m =$$
$$-\mathscr{L}^{-1}[J(\sum_i p_i e_i, \mathscr{L}\sum_j p_j e_j)] - \mathscr{L}^{-1}U\frac{\partial \mathscr{L}\sum_i p_i e_i}{\partial x} - \mathscr{L}^{-1}[\beta + \bar{\zeta}_y]\frac{\partial \sum_i p_i e_i}{\partial x}$$
$$-\mathscr{L}^{-1}J(p_0, \mathscr{L}\sum_j p_j e_j) - \mathscr{L}^{-1}J(\sum_i p_i e_i, \mathscr{L}p_0) - \mathscr{L}^{-1}U\frac{\partial \mathscr{L}p_0}{\partial x} - \mathscr{L}^{-1}\left[(\beta + \bar{\zeta}_y)\frac{\partial p_0}{\partial x}\right]. \tag{47}$$

The second line is the new term in comparison to the original one. Thus, the following

$$-\left\langle e_m, \, \mathscr{L}^{-1}[J(p_0, \mathscr{L}e_i) + J(e_i, \mathscr{L}p_0)]\right\rangle$$

should be added to the above coefficients $\alpha^L$. In addition, there exists an nonautonomous term

$$b_m = - \left\langle e_m, \ \mathscr{L}^{-1} \left[ U \frac{\partial \mathscr{L} p_0}{\partial x} + (\beta + \bar{\zeta}_y) \frac{\partial p_0}{\partial x} \right] \right\rangle.$$

However, here it is shown that all these are negligible. Thus, it is adequate to use the above autonomous system in the following studies.

## 5. Information Flow between the Scales of the Model Atmosphere

As we showed above, the interactions among the first four EOF modes can be utilized to study the multiscale interactions typical in the problem of concern, as the modes occur on different time scales. In order to examine the information flow between the modes, we make random draws for $(p_1, p_2, p_3, p_4)$ from a pool of values, and then, starting from these initial conditions, run forward the system to generate an ensemble of solutions. Assume that the initial values obey a normal distribution with a mean vector $(0.1, 0.1, 0.1, 0.1)$ and a $4 \times 4$ identity covariance matrix. Here, the variance is set rather small in order for the trajectories to stay under effective control. The sample space is assumed to be $[-6, 6] \times [-6, 6] \times [-6, 6] \times [-6, 6]$, which makes sense if we do not make too long an integration, as made evident in Figure 10, where the trajectory of a sample path is plotted. The space is discretized using a spacing $\Delta = 0.2$ (the same for the four dimensions), and the probability density functions are then estimated at each time step by counting the bins in the coarse-grained space.
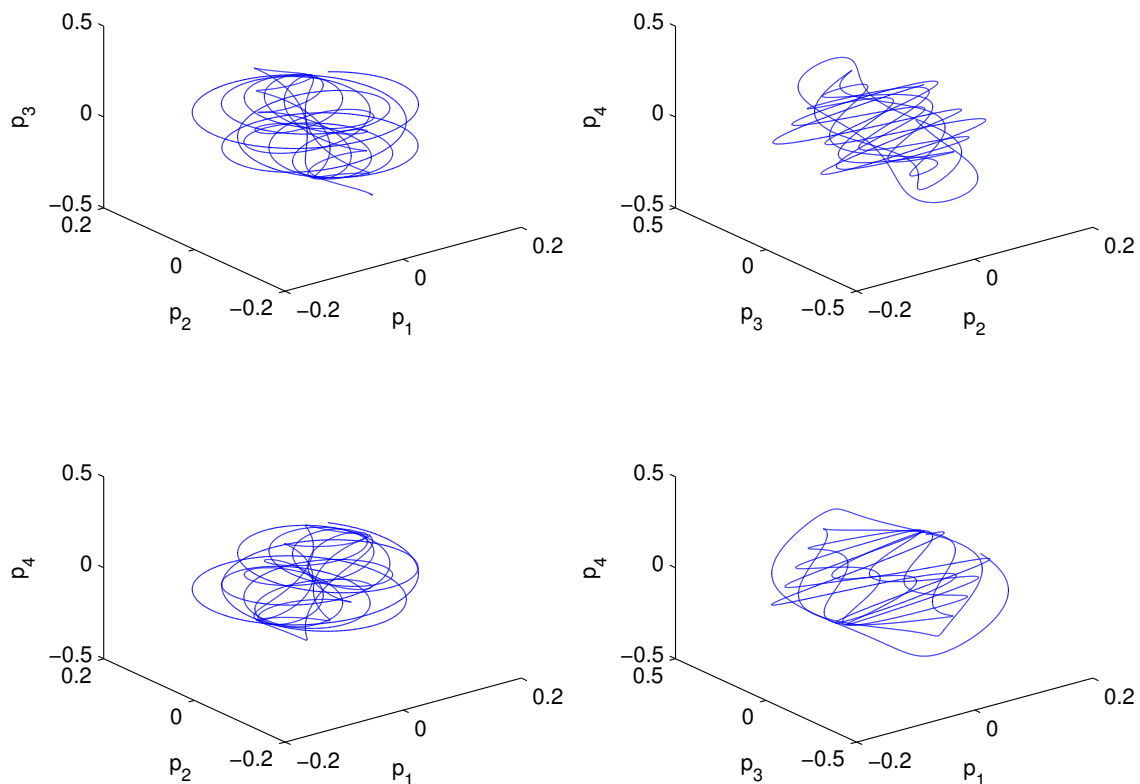


**Figure 10.** The trajectory of a sample path.

To compute the information flows among the four components, recall the deterministic version of Equation (2)

$$T_{j \to i} = - \int_{\mathbb{R}^n} \rho_{j|i}(p_j | p_i) \frac{\partial F_i \rho_i}{\partial p_i} d\mathbf{p}. \tag{48}$$

When the system is initialized with values in a rather limited domain, the integration can be easily evaluated. In this study, $\mathbb{R}^4$ is replaced by $[-6,6] \times [-6,6] \times [-6,6] \times [-6,6]$. The computed information flow evolutions vs. time are plotted in Figure 11.
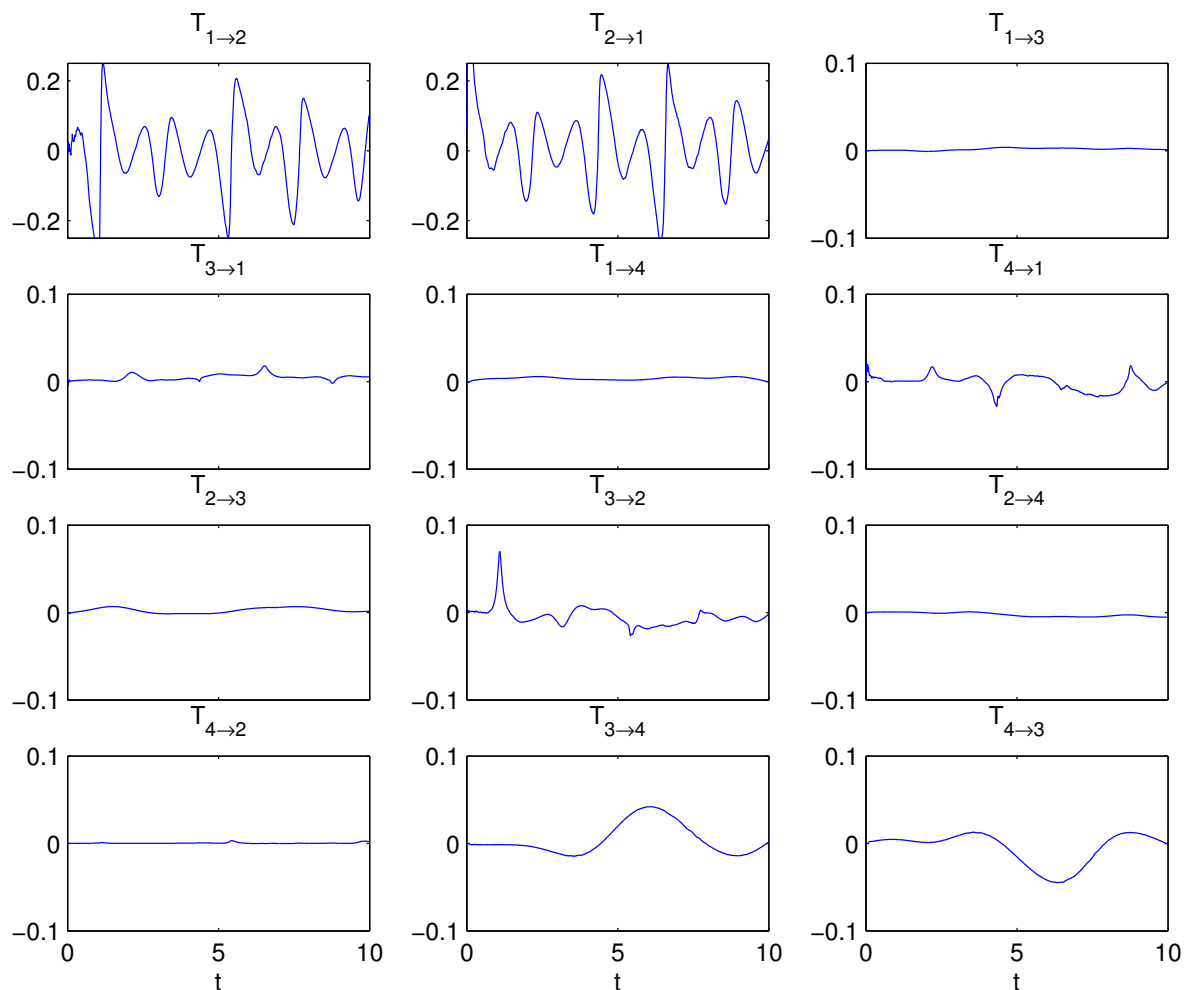


**Figure 11.** The computed information flows among the components of the reduced system. Note the range scale for the first two subplots ($T_{1\to2}$ and $T_{2\to1}$) is twice that for the others.

For a system with four components, by expectation, there are in general $4 \times 3 = 12$ information flows. As we have shown before, the four components make two pairs, i.e., $(p_1, p_2)$ and $(p_3, p_4)$, which essentially represent two scales. Thus, the cross-scale information flows are those between modes (1,2) and modes (3,4). In Figure 11, $T_{1\to2}$ and $T_{2\to1}$ are overwhelmingly large (note the different scale range in the first two subplots); second to them are $T_{3\to4}$, $T_{4\to3}$. By the property of causality (ideally nonzero information flow implies causality), that is to say, $p_1$ and $p_2$ are mutually causal, and so are $p_3$ and $p_4$. These are the information flows within their respective scales. These causal patterns are similar to that between the displacement and linear momentum of a harmonic oscillator, as is shown in Liang [36]. From the table of $\alpha_{m,i}^L$ indeed to the first order, the system is like

$$\frac{d}{dt}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0 & 1.4 \\ -1.4 & 0 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \tag{49}$$

just as the computed $T_{1\to2}$ and $T_{2\to1}$ would imply.

The other information flows are interscale. Strictly speaking, there exist flows in both directions (small-scale⟶large-scale and large-scale⟶small-scale). However, by observation, $|T_{4\to1}|$ and $|T_{3\to2}|$ are much larger than others. This asymmetric flow structure indicates that the causation between the

scales are dominantly one-way, i.e., from higher frequency modes (modes 3 and 4) to lower frequency modes, modes 1 and 2.

It should be mentioned that what has been solved is actually the QG equation for the perturbation field; the mean flow is not included in the four components $(p_1, p_2, p_3, p_4)$ of the reduced system. However, the influence has been embedded in the system. Here, we give it an evaluation.

For notational convenience, let $p_0$ denote the "mean component." Since here the mean flow is prescribed, it does not vary in time, so there is no way to examine the influence of other components on it. That is to say, there is no base to study $T_{i \to 0}$, but nonetheless we can evaluate $T_{0 \to i}$, $i = 1, 2, 3, 4$. We know, from the Bayes' rule, that

$$\rho_{0|i}(p_0|p_i) = \frac{\rho(p_i|p_0)\rho_0(p_0)}{\rho_i(p_i)}. \tag{50}$$

Since the mean flow is prescribed, it is certain; $\rho(p_i|p_0)$ is hence in fact $\rho_i(p_i)$. Thus, the whole term is equal to 1. This substituted into (48) yields

$$\begin{aligned} T_{0 \to i} &= -\int_{\mathbb{R}^n} \rho_{0|i} \frac{\partial F_i \rho_0}{\partial p_i} d\mathbf{p} \\ &= -\int_{\mathbb{R}^n} \frac{\partial F_i \rho}{\partial p_i} d\mathbf{p} \\ &= 0 \end{aligned} \tag{51}$$

by the compactness of $\rho$. That is to say, the information flow between the mean flow and the higher frequency components, if it exists, cannot be toward higher modes. In other words, if existing, it must be one way, i.e., in the direction upward toward the mean.

It should be emphasized that, generally, the mean flow should also have a distribution, and hence the information flow may not be this easy to evaluate. However, in this case, as we have shown in the preceding section, the variation around the mean is so small that it can be neglected in forming the low-dimensional system. Anyhow, for this particular case, by computation the information flow, and hence causation, is essentially one-way, i.e., from high frequency modes to low frequency modes.

We want to mention that here EOF analysis has been used to reduce the model order. The advantages of using it include its orthonormality, the maximization of variance toward lowest modes, etc. The limitations of this approach are also well known. The most serious one is that the EOF modes may not be real modes in the physical sense. Here, what we are investigating is the information exchange between processes on different temporal scales, and, fortunately, the principal components of the lowest modes do reflect such temporal variabilities (Figure 9). However, in a more general situation, this may not be true. We hope some advanced methods, such as the recently developed method by Majda and Qi [50] to efficiently reduce models, can help here.

An alternative approach is to use Equation (4) to estimate from data, rather than directly compute, the information flow, and hence avoid solving a large-dimensional Liouville equation (the curse of dimensionality). However, here comes another issue: Theorem 3 relies on the assumption of Gaussianity. Though (4) has also been successfully applied to some highly nonlinear systems, e.g., the chaotic anticipatory system in [41] (see [31]), caution should be used, as non-Gaussianity may appear significant in realistic atmospheres. However, anyhow, these are topics for future studies; here as the first step, we only consider what we have generated with the QG model.

## 6. Discussion and Conclusions

How processes on different scales interact to form weather and climate patterns is one of the central issues in dynamic meteorology. Traditionally, it is studied by diagnosing the exchange of energy (such as the Lorenz cycle), or, equivalently, momentum/angular momentum, between the scales. However, it has long been realized that just multiscale energetics based on the governing equation may not be enough. In a nonlinear dynamical system, as time moves on, two highly correlated events may soon lose correlation, while, on the other hand, two completely irrelevant events could turn out to be correlated in the end. As remarked by Corning [51], the underlying causal efficacy may actually be missing in the equations or "rules". In addition, in the classical multiscale formalism, cyclogenesis is driven by Reynolds stress, which is essentially the linear correlation between the perturbation fields.

As we proved earlier on, while causation implies correlation, correlation does not necessarily imply causation. That said, the traditional perspective on the problem may be limited.

In physics, entropy is another concept as important as energy. The transference of entropy results in a flow of information, but how information flows or transfers across scales has been overlooked in dynamic meteorology, in contrast to the extensively studied energy transfer. Recently, information flow has been rigorously formulated in the framework of dynamical systems; it proves to satisfy the "principle of nil causality" (see [36]), an observational fact which people endeavor to verify in real applications. In this study, this formalism is applied to study the information flow among the scales within a three-dimensional quasigeostrophic (QG) circulation. The basic flow is a zonal jet mimicking the atmospheric jet stream. We chose a period when the system is in equilibrium with an energetic scenario typical of a mid-latitude atmosphere: the mean state is releasing available potential energy to eddies, while the latter feeds kinetic energy back to the mean state. We first solved the 3D QG equation; then, for the period of concern, performed a principal component analysis and obtained the EOF modes to construct a basis. It has been shown that these modes characterize the desired temporal scales. The state variable, i.e., streamfunction, is then expanded with the aid of the basis, and the expansion is truncated at the fourth term. By inverting a 3D elliptic differential operator, the QG equation is converted into a four-dimensional dynamical system. The study of the information flows among the scales is then converted into the investigation of the information flows among the components of the low-dimensional system.

Initialized with an ensemble of streamfunctions drawn randomly according to a normal distribution, the system is integrated forward and, at each step, a probability density function is estimated, which, by Formula (2), allows us to obtain the desired information flow pairs. By computation mode 1 and mode 2, which represent the long temporal scale, are mutually causal, functioning like the components of a 2D harmonic oscillator; this is also the case for mode 3 and mode 4 that represent the motion on a short scale. These are the information flows within their respective scales. The interscale flows are significant only for that from mode 4 to mode 1 and that from mode 3 to mode 2, i.e., from modal pair (3,4) to modal pair (1,2). In addition, the possibility that the mean state has information flow to these four modes are excluded. That is to say, for this particular problem, the information flow is mostly one-way—from higher frequency modes to lower frequency modes. Hence, for this particular problem, underlying the multiscale interaction is mostly a bottom-up causation.

The bottom-up causation, or the information flow from the low levels to higher levels, is actually seen in many natural and social phenomena. In investigating the transition in biological complexity, for example, a reductionist will view the emergence of new, higher level, aggregate entities as a result of lower level entities (e.g., [52–54]). Similarly, it is found that some simple computer networks may transit from a low traffic state to a high congestion state, entailing a flow of information from a combination of independent objects to a collective pattern representing a higher level of organization. Most of all, in statistical physics [55,56], bottom-up causation lays for it the theoretical foundation, based on which the macroscopic thermodynamic properties can be tracked back to random molecular motions.

However, we did not exclude the existence of information flow the other way around; it is just weak by comparison in this example. Top-down causation has been found in many fields. For example, in community ecology, it has been argued that host community-level structures may determine the disease dynamics and hence control the constituent populations (e.g., [57]). Nonetheless, here we showed that a prescribed mean flow seems to be unlikely to have information flow to the anomalies.

Of course, the result here is just for a particular case with a reduced-order model; in reality, the problem could be very complicated, depending on the stage where the evolving state is. In addition, for simplicity, we have adopted a rigid-lid assumption on the top, and an idealized boundary condition ($\frac{\partial \psi}{\partial z} = 0$, i.e., no density perturbation) at the bottom, although the simplified model does reproduce the desired downward transfer of available potential energy and upward kinetic energy. Nonetheless, the resulting interaction scenario is encouraging, in agreement with those in complex systems, although it is quite different from the corresponding energetic cycle. This result, though preliminary at this stage, may help better understand the mean flow–eddy interaction, gain deeper insight into the phenomena such as cyclogenesis, atmospheric blocking, sudden stratospheric warming, to name a few. On the other hand, the asymmetric causation (mostly bottom-up) provides an observational basis for the parameterization of the subgrid processes in numerical models, such as the stochastic closure scheme of Majda et al. [58]. All of these are interesting and deserve further investigation. We want to emphasize

that information flow is a large field in atmospheric research, and this present study makes only a first attempt; much is yet to be explored in the future.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Cai, M.; Mak, M. Symbiotic relation between planetary and synoptic scale waves. *J. Atmos. Sci.* **1990**, *47*, 2953–2968. [CrossRef]
2. Chang, E.K.M.; Orlanski, I. On the dynamics of a storm track. *J. Atmos. Sci.* **1993**, *50*, 999–1015. [CrossRef]
3. Hoskins, B.J.; Hodges, K.I. New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.* **2002**, *59*, 1041–1061. [CrossRef]
4. McWilliams, J.C. An application of equivalent modons to atmospheric blocking. *Dyn. Atmos. Oceans* **1980**, *5*, 43–66. [CrossRef]
5. Nakamura, H.; Wallace, J.M. Synoptic behavior of baroclinic eddies during the blocking onset. *Mon. Weather Rev.* **1993**, *121*, 1892–1903. [CrossRef]
6. Butler, A.H.; Seidel, D.J.; Hrdiman, S.C.; Butchart, N.; Birner, T.; Match, A. Defining sudden stratospheric warmings. *Bull. Am. Meteor. Soc.* **2015**, *96*, 1913–1928. [CrossRef]
7. Holton, J.R. The dynamics of sudden stratospheric warmings. *Annu. Rev. Earth Planet. Sci.* **1980**, *8*, 169–190. [CrossRef]
8. Barnston, A.G.; Livezey, R.E. Classification, seasonality and persistence of low frequency atmospheric circulation patterns. *Mon. Weather Rev.* **1987**, *115*, 1083–1126. [CrossRef]
9. Franzke, C.; Lee, S.; Feldstein, S.B. Is the North Atlantic Oscillation a breaking wave? *J. Atmos. Sci.* **2004**, *61*, 145–160. [CrossRef]
10. Hurrell, J.W.; Kushnir, Y.; Ottersen, G.; Visbeck, M. An overview of the North Atlantic Oscillation. In *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*; Hurrell, J.W., Kushinir, Y., Ottersen, G., Visbeck, M., Eds.; Geophysical Monograph: Washington, DC, USA, 2003; Voulme 134, pp. 1–35.
11. Jin, F.-F.; Pan, L.L.; Watanabe, M. Dynamics of Synoptic Eddy and Low-Frequency Flow Interaction. Part II: A Theory for Low-Frequency Modes. *J. Atmos. Sci.* **2006**, *63*, 1695–1708. [CrossRef]
12. Robinson, W.A. Does eddy feedback sustain variability in the zonal index? *J. Atmos. Sci.* **1996**, 53, 3556–3569. [CrossRef]
13. Limpasuvan, V.; Hartmann, D.L. Eddies and the annular modes of climate variability. *Geophys. Res. Lett.* **1999**, *26*, 3133–3136. [CrossRef]
14. Green, J.S.A. Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Q. J. R. Meteorol. Soc.* **1970**, *96*, 157–185. [CrossRef]
15. Simmons, A.J.; Hoskins, B.J. The life cycles of some nonlinear baroclinic waves. *J. Atmos. Sci.* **1978**, *35*, 414–432. [CrossRef]
16. Holopainen, E.; Fortelits, C. High-frequency transient eddies and blocking. *J. Atmos. Sci.* **1987**, *44*, 1632–1645. [CrossRef]
17. Hansen, A.R.; Chen, T.-C. A spectral energetics analysis of atmospheric blocking. *Mon. Weather Rev.* **1982**, *100*, 1146–1165. [CrossRef]
18. Ma, J.; Liang, X.S. Multiscale dynamical processes underlying the wintertime Atlantic blockings. *J. Atmos. Sci.* **2017**, *74*, 3815–3831. [CrossRef]
19. Chang, E.K.M.; Lee, S.; Swanson, K.L. Storm track dynamics. *J. Clim.* **2002**, *15*, 2163–2183. [CrossRef]
20. Nakamura, H.; Sampe, T. Trapping of synoptic-scale disturbances into the North-Pacific subtropical jet core in midwinter. *Geophys. Res. Lett.* **2002**, *29*, 1761. [CrossRef]
21. Ioannou, P.; Lindzen, R.S. Baroclinic instability in the presence of barotropic jets. *J. Atmos. Sci.* **1986**, *43*, 2999–3014. [CrossRef]
22. James, I.N. Suppression of baroclinic instability in horizontally sheared flows. *J. Atmos. Sci.* **1987**, *44*, 3710–3720. [CrossRef]

23. Liang, X.S.; Robinson, A.R. Localized multiscale energy and vorticity analysis. I. Fundamentals. *Dyn. Atmos. Oceans* **2005**, *38*, 195–230. [CrossRef]

24. Liang, X.S.; Anderson, D.G.M. Multiscale window transform. *SIAM J. Multiscale Model. Simul.* **2007**, *6*, 437–467. [CrossRef]

25. Liang, X.S. Canonical transfer and multiscale energetics for primitive and quasigeostrophic atmospheres. *J. Atmos. Sci.* **2016**, 73, 4439-4468. [CrossRef]

26. Zhao, Y.-B.; Liang, X.S. On the inverse relationship between the boreal wintertime Pacific jet strength and storm-track intensity. *J. Clim.* **2018**, *31*, 9545–9564. [CrossRef]

27. Penny, S.M.; Battisti, D.S.; Roe, G.H. Examining mechanisms of variability within the Pacific storm track: Upstream seeding and jet-core strength. *J. Clim.* **2013**, *26*, 5242–5259. [CrossRef]

28. Liang, X.S.; Lozano-Durán, A. A preliminary study of the causal structure in fully developed near-wall turbulence. In *Proceedings of the Summer Program*; Center for Turbulence Research: Stanford, CA, USA, 2016; pp. 233–242.

29. Materassi, M.; Consolini, G.; Smith, N.; De Marco, R. Information theory analysis of cascading process in a synthetic model of fluid turbulence. *Entropy* **2014**, *16*, 1272–1286. [CrossRef]

30. Pope, S.B. *Turbulent Flows*; Cambridge University Press: Cambridge, UK, 2013; p. 771.

31. Liang, X.S. Unraveling the cause-effect relation between time series. *Phys. Rev. E* **2014**, *90*, 052150. [CrossRef]

32. Dempster, A.P. Causality and statistics. *J. Stat. Plan. Inference* **1990**, *25*, 261–278. [CrossRef]

33. O'Neil, C.; Schutt, R. *Doing Data Science: Straight Talk from the Frontline*; O'Reilly: Cambridge, MA, USA, 2013.

34. Liang, X.S.; Kleeman, R. Information transfer between dynamical system components. *Phys. Rev. Lett.* **2005**, *95*, 244101. [CrossRef]

35. Majda, A.J.; Harlim, J. Information flow between subspaces of complex dynamical systems. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9558–9563. [CrossRef]

36. Liang, X.S. Information flow and causality as rigorous notions ab initio. *Phys. Rev. E* **2016**, *94*, 052201:1–052201:28. [CrossRef] [PubMed]

37. Smirnov, D.A. Spurious causalities with transfer entropy. *Phys. Rev. E* **2013**, *87*, 042917. [CrossRef] [PubMed]

38. Lizier, J.T.; Prokopenko, M. Differentiating information transfer and causal effect. *Eur. Phys. J. B* **2010**, *73*, 605–615. [CrossRef]

39. Liang, X.S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **2008**, *78*, 031113. [CrossRef]

40. Liang, X.S. Normalizing the causality between time series. *Phys. Rev. E* **2015**, *92*, 022126. [CrossRef] [PubMed]

41. Hahs, D.W.; Pethel, S.D. Distinguishing anticipation from causality: Anticipatory bias in the estimation of information flow. *Phys. Rev. Lett.* **2011**, *107*, 128701. [CrossRef] [PubMed]

42. Bai, C.; Zhang, R.; Bao, S.; Liang, X.S.; Guo, W. Forecasting the tropical cyclone genesis over the Northwest Pacific through identifying the causal factors in the cyclone-climate interactions. *J. Atmos. Ocean Technol.* **2018**, *35*, 247–259. [CrossRef]

43. Stips, A.; Macias, D.; Coughlan, C.; Garcia-Gorriz, E.; Liang, X.S. On the causal structure between $CO_2$ and global temperature. *Nat. Sci. Rep.* **2016**, *6*, 21691. [CrossRef]

44. Vaid, B.H.; Liang, X.S. The changing relationship between the convection over the western Tibetan Plateau and the sea surface temperature in the northern Bay of Bengal. *Tellus A* **2018**, *70*, 1440869. [CrossRef]

45. Majda, A.J.; Wang, X. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*; Cambridge University Press: Cambridge, UK, 2006.

46. Pedlosky, J. *Geophysical Fluid Dynamics*, 2nd ed.; Springer: New York, NY, USA, 1987.

47. Kuo, H.L. Dynamics of quasigeostrophic flows and instability theory. In *Advances in Applied Mechanics*; Yih, C.-S., Ed.; Academic Press: New York, NY, USA, 1973; Volume 13, pp. 247–330.

48. Carrier, G.F.; Pearson, C.E. *Partial Differential Equations: Theory and Technique*; Academic Press: New York, NY, USA, 1976; 320p.

49. Gill, A.E. *Atmosphere-Ocean Dynamics*; Academic Press: New York, NY, USA, 1980; p. 662.

50. Majda, A.J.; Qi, D. Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Rev.* **2018**, *60*, 491–549. [CrossRef]

51. Corning, P.A. The re-emergence of emergence: A venerable concept in search of a theory. *Complexity* **2002**, *7*, 18–30. [CrossRef]

52. Küppers, B. *Information and the Origin of Life*; MIT Press: Cambridge, MA, USA, 1990.

53. Goldenfield, N.; Woese, C. Life is physics: Evolution as a collective phenomenon far from equilibrium. *Ann. Rev. Condens. Matt. Phys.* **2011**, *2*, 375–399. [CrossRef]

54. Murray, J.D. *Mathematical Biology*; Springer: Berlin, Germany, 2000.

55. Allahverdyan, A.E.; Janzing, D.; Mahler, G. Thermodynamic efficiency of information and heat flow. *J. Stat. Mech.* **2009**, *2009*, PO9011. [CrossRef]

56. Crutchfield, J.P.; Shalizi, C.R. Thermodynamic depth of causal states: Objective complexity via minimal representation. *Phys. Rev. E* **1999**, *59*, 275–283. [CrossRef]

57. Johnson, P.T.; Preston, D.L.; Hoverman, J.T.; Richgels, K.L. Biodiversity decreases disease through predictable changes in host community competence. *Nature* **2013**, *494*, 230. [CrossRef]

58. Majda, A.J.; Timofeyev, I.; Vanden-Eijnden, E. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.* **2003**, *60*, 1705–1722. [CrossRef]