Doswell, C. A. III, and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic J. Severe Storms Meteor.*, **1**(3)**,** 1–22.

Electronic Journal of
SEVERE STORMS METEOROLOGY

On the Use of Indices and Parameters in Forecasting Severe Storms

CHARLES A. DOSWELL III
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

DAVID M. SCHULTZ
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/National Severe Storms Laboratory, Norman, Oklahoma

(Submitted 11 May 2006; in final form 05 November 2006)

ABSTRACT

This paper describes our concept of the proper (and improper) use of diagnostic variables in severe-storm forecasting. A framework for classification of diagnostic variables is developed, indicating the limitations of such variables and their suitability for operational diagnosis and forecasting. The utility of diagnostic variables as forecast parameters is discussed, in terms of what we consider to be the relevant issues in designing new diagnostic variables used for making weather forecasts. Finally, criteria required to determine whether a new diagnostic variable represents an effective forecast parameter are proposed. We argue that many diagnostic variables in widespread use in forecasting severe convective storms have not met these criteria for demonstrated utility as forecast parameters.

1. Introduction

Operational and research meteorologists often refer to diagnostic variables, such as convective available potential energy (CAPE) or the supercell composite parameter (SCP; Thompson et al. 2003), as forecast parameters. We contend that they are not necessarily forecast parameters; rather, they constitute a set of diagnostic variables. For most such variables, their forecast value generally has not been established via rigorous verification. That operational forecasters have used many of them for decades is not sufficient, in our opinion, to establish their capability as forecast parameters.

Furthermore, diagnostic variables can lead to faulty perceptions of the state of the atmosphere, owing to various issues tied to their computation and representativeness. Although there is nothing inherently wrong with diagnostic variables, forecasters need to be aware of the limitations on their use. For example, one of the most common and well established among these is CAPE in various forms. Monteverdi et al. (2003) showed that for at least one type of severe weather forecasting (tornadoes in California), CAPE proved to be of little value in discriminating tornado cases from nontornadic cases when tested as a forecast parameter. The forecasts based on CAPE were verified against observed events, whereas another variable (0–1-km shear) appeared to be capable of making such discriminations.

Diagnostic variables have a long history in association with forecasting severe convection (Schaefer 1986; Johns and Doswell 1992). It is evident that the value of such variables is strongly associated with their capacity to summarize in a single number (or field variable)

Corresponding author address: Dr. Charles A. Doswell III, The University of Oklahoma/CIMMS, 120 David L. Boren Blvd, Suite 2100, Norman, OK 73072-7304.
E-mail: cdoswell@hoth.gcn.ou.edu

some characteristic of the severe storm environment. Rather than having to consider the full complexity of four-dimensional atmospheric data, it is often regarded as a benefit for forecasters working with hard deadlines to be able to distill that complexity into a single variable.

Forecasters and researchers generally acknowledge that any single diagnostic variable considered in isolation has little forecast value. Nevertheless, in our experience, we have seen instances where forecasters, often under forecast deadline pressure, will make forecast decisions based heavily, if not primarily, on a single diagnostic variable. It has been our observation that forecasters are most prone to rely heavily on a single diagnostic variable in the context of determining the general likelihood of severe weather. For example, when CAPE or strong vertical wind shear is found to be absent at the diagnosis time, the likelihood of severe convective weather subsequently is sometimes dismissed as unlikely. An important concern is that most of the widely used diagnostic variables have not been validated as proper forecast parameters. (This process for validation is defined in section 2). As diagnostic variables, they can be useful in assessing quantitatively the state of the atmosphere at the time of their calculation, but their capability to inform forecasters about weather in the future can be quite limited, at best.

One purpose of this paper is to propose a classification scheme for diagnostic variables in use for severe weather forecasting, in order to understand their characteristics and limitations. Another goal is to address the issue of what it takes to validate the value of a variable as a proper forecast parameter. We are not seeking to discourage the use of diagnostic variables, per se, in forecasting severe convective storms. Rather, we seek to develop a perspective through which their value to forecasting can be maximized.

2. Diagnostic variables

What specifically do we mean by a *diagnostic variable*? A diagnostic variable is some quantity, valid at a specific instant in time, which either is a basic observed variable (e.g., pressure, temperature, wind, humidity) or can be calculated from those variables. The relationship between diagnosis and prognosis, as developed by Doswell (1986), can be expressed mathematically as follows. Let $\mathbf{\Phi}$ be an n-dimensional vector of atmospheric state variables

$$\mathbf{\Phi} = \left(\phi_1, \phi_2, \ldots, \phi_n\right).$$

In general, this vector is a function of its location in space, denoted by the position vector $\mathbf{X}$, and is known at some particular time, $t = t_0$, such that

$$\mathbf{\Phi} = \mathbf{\Phi}\left(\mathbf{X}, t = t_0\right) \equiv \mathbf{\Phi}_0 .$$

A basic principle of numerical weather prediction (NWP) is that the information about the state of the atmosphere at some initial time can be used to make a forecast of the state of the atmosphere at some future time in the following way:

$$\mathbf{\Phi}\left(\mathbf{X}, t = t_0 + \delta t\right) = \mathbf{\Phi}\left(\mathbf{X}, t = t_0\right) + \delta t \left.\frac{\partial \mathbf{\Phi}}{\partial t}\right|_{t=t_0} .$$

Thus, the state of the atmosphere at some future time is the sum of its state at the initial time (the diagnosis) and the product of the time step, δt, with the local time rate of change of $\mathbf{\Phi}$, where

$$\frac{\partial \mathbf{\Phi}}{\partial t}\left(\mathbf{X}, t\right) = f\left[\mathbf{\Phi}\left(\mathbf{X}, t\right)\right].$$

That is, the time trend of the atmospheric state variables at any point in space at a particular time is a function of the current spatial distribution of those state variables. In NWP, this time trend is expressed in the form of a set of governing equations used for a particular NWP model. Thus, although it is strictly true that a forecast can be calculated from the state variable vector $\mathbf{\Phi}_0$ at the initial time, we exclude from consideration as diagnostic variables the explicit calculation of local time tendencies for atmospheric state variables via this NWP-like process. This does not, as we will show, exclude as diagnostic variables the estimation of time tendencies by other means.

Establishing a quantitative understanding of the state of the atmosphere at a given time using diagnostic variables is possible. Qualitative understanding, in turn, can be substantially enhanced by this quantitative analysis. For example, a forecaster can simply look at the plotted winds on a surface chart and recognize zones of strong rotation (vorticity) and convergence, but without a quantitative evaluation of the vorticity and convergence values, a forecaster's sense of the strength of those variables is obviously subjective. A quantitative assessment is generally preferred over subjective assessment in forecasting; numerical values of variables often matter to the forecast decision..

We define a diagnostic variable being used as a *forecast parameter* as one that allows a forecaster to make an accurate weather forecast based on the *current* values of that variable. For a proper forecast parameter, there should be a time-lagged correlation between the parameter and the weather being forecast. This is to be distinguished from using a diagnostic variable calculated from a forecast of the state variables (say, from NWP model gridded forecast fields) valid at some future time to make a forecast of the weather for that valid time. How accurately an NWP model forecasts that diagnostic variable is a very different issue from how accurate a forecast can be using only the current distribution of that variable. A diagnostic variable might be very good at discriminating between different weather events when known at the time of the event, but still be ineffective as a forecast parameter because its values *before* the events occur make no directly useful statement about the future weather.

Obviously, diagnostic variables change with time, and indeed, a description of the time tendency for some diagnostic variable can be an important diagnostic variable in its own right and may even be valuable as a proper forecast parameter. The accuracy of forecasts made using a diagnostic variable as a forecast parameter will tend to decay with time. Some forecast parameters might permit reasonably accurate forecasts only for a short period very close to the time of their diagnosis, whereas others might show high lagged correlation with the weather for a relatively long time.

Therefore, what we are concerned with herein is a forecast by a human forecaster that makes a statement about the future weather (say, the occurrence of severe storms), rather than a statement about the distribution of atmospheric state variables. For a variable to be useful in making a weather forecast by a human, it must be shown rigorously that there is some measure of forecast skill when the current values of the variable are used to make that weather forecast. This argument will be developed in detail in what follows.

When calculating diagnostic variables, forecasters unaware of the caveats associated with the particular variable being analyzed can be misled. What are the unique sensitivities in those calculations? How much confidence can one put in the numbers and how they might change over time? If diagnostic variables are to be used properly, forecasters need to be aware of the story behind each of them. In order to discuss diagnostic variables and their limitations, we propose a classification scheme as follows.

a. Simple observed variables

Simple observed variables are those measured by meteorological instruments. Strictly speaking, most modern meteorological instruments make electronic measurements (e.g., resistance or capacitance) that are calibrated to provide readouts of the meteorological variables (e.g., pressure, temperature, dewpoint, wind direction and speed).

b. Simple calculated variables

Calculated variables are not observed directly but are computed from the raw measurements using relatively simple conversion formulae. Calculated variables typically involve combinations of two or more observed variables, but they are not combined in an arbitrary fashion. Rather, they are combined in ways having a physical basis. Such calculated variables are usually sought because they have some valuable physical property, such as being conserved under certain reasonable assumptions. For example, at the surface, the temperature and dewpoint temperature are the common observed variables. However, for reasons discussed in Sanders and Doswell (1995), mixing ratio and potential temperature are conserved variables that incorporate the pressure observations, as well as the temperature and dewpoint. The formulae for calculating the mixing ratio and potential temperature are determined by the laws governing the physical properties of air parcels under the assumption of adiabatic flow.

c. Derivatives or integrals (spatial or temporal) of simple observed or calculated variables

Time and space derivatives and integrals of the observed or calculated variables form the next class of diagnostic variables. In effect, these diagnostic quantities allow estimates of the terms in formulae that might arise in mathematical descriptions of atmospheric structure. An important caveat in the calculation of such diagnostic variables is the inevitable truncation error that arises from the limited temporal and spatial resolution of our meteorological information, whether it be observed or modeled. If we were to calculate, say, the Eulerian time rate of change

of the 500-hPa height at some location where soundings are taken, the true instantaneous value is not known, simply because soundings generally are taken at 12-h intervals. Of course, estimating that same variable at some point other than where a sounding is launched[1] is problematic because of the small number of locations where soundings are made. The usual approach is to make estimates using available information; this typically involves making assumptions about the behavior of the fields where we have no information. Realizing the limitations of those calculations is critical. Just because the calculations were performed by computer is no guarantee that they are even remotely accurate.

d. Combined variables

Next is the class of *combined* diagnostic variables. Many ways exist to take two or more diagnostic variables and combine them in some way that might be more useful for some specific purpose than the raw observations or simple derivatives and integrals of those variables. Moisture flux convergence (MFC), discussed in detail by Banacos and Schultz (2005), is an example of a combined variable. The formulation of MFC can vary from one application to another, but it is often formulated as the finite difference calculation of the quantity:

$$\mathrm{MFC} = -\nabla_h \bullet \left(r\mathbf{V}_h\right) = \underbrace{r\,\nabla_h \bullet \mathbf{V}_h}_{1} + \underbrace{\mathbf{V}_h \bullet \nabla_h r}_{2}$$

where r is the mixing ratio, ∇_h is the horizontal gradient operator, and $\mathbf{V}_h$ is the horizontal wind vector. This calculation involves a calculated variable (the mixing ratio) and a spatial derivative of an observed variable (the horizontal wind vector). MFC is often calculated using the relatively dense surface observations, and, as used in forecasting severe storms, MFC is purported to show where ascent is occurring in the presence of surface moisture. Banacos and Schultz (2005) emphasize four points about MFC relevant to this discussion.

1. The scientific justification for using MFC as a forecast tool for convective initiation is inadequate.
2. The putative value of MFC as a forecast variable has never been firmly established by a careful statistical verification study. Its popularity is based almost entirely on anecdotal evidence and heuristic arguments.
3. Term 1 on the rhs of the MFC equation associated with horizontal divergence is characteristically *much* larger than term 2, associated with moisture advection. The MFC field tends to look very much like that of the horizontal divergence field alone on mesoscale and smaller scales.
4. MFC is an inadequate tool as a forecast parameter because it *combines* two of the three ingredients required for deep, moist convection (e.g., Johns and Doswell 1992). The two components of the major term in this combination, the mixing ratio and the divergence field, can evolve quasi-independently. What this variable shows is where those two main components overlap at the time of the analysis. Thus, MFC is first and foremost a diagnostic variable.

CAPE is another example of a combined variable; its calculation involves a vertical integral of multiple state variables and incorporates a number of assumptions (e.g., Doswell and Rasmussen 1994; Doswell and Markowski 2004). Nevertheless, that calculation can be represented by its most basic components: large CAPE is generally observed where low-level moisture is found in the presence of conditionally unstable lapse rates in the lower mid-troposphere. As with MFC, these constituent fields can evolve quasi-independently and can be superimposed by differential advection processes. Prior to that superposition, the air streams carrying conditionally unstable lapse rates and low-level moisture (typically at different levels in the atmosphere advecting variables in different directions and at different speeds) have not yet interacted and so little or no CAPE is found.

The presence of CAPE indicates when its constituents overlap, but an absence of CAPE prior to that superpositioning cannot be used to infer large CAPE will not be present in the future. CAPE depicts where moisture and conditional instability are already superimposed, not where they will (or will not) be superimposed in the future. A comparable statement can be made about MFC. This aspect of combined variables is an important element for understanding the difference between a diagnostic variable and a proper forecast parameter.

[1] In certain situations, the displacement of the sounding instrument from its original launch location has to be accounted for in the calculations.

e. Indices

Indices, the final class of diagnostic variables, can be broken down into two distinct subclasses: indices based on physically based formulae and indices representing more or less arbitrary combinations of diagnostic variables. This is a complex topic and is the subject of a wholly separate section (section 4, below). Before we discuss indices, however, we consider several issues associated with diagnostic variables that affect their utility in diagnosing the current state of the atmosphere.

3. Issues affecting the suitability of diagnostic variables

All diagnostic variables are subject to error, but not all are equally error-prone. Errors can be decomposed into *measurement* errors, which have a number of sources that are generally instrument-dependent (Brock and Richardson 2001, their section 1.1.3), and *sampling* errors (associated with the finite number of observations). Measurement and sampling errors are associated with much of the *volatility* of diagnostic variables. Here we use volatility to mean that the variable can vary considerably over both time and space as a result of sensitivity to both measurement and sampling errors. Some diagnostic variables are inherently more volatile than others.

Consider CAPE as an example of a combined variable, and compare it to the Showalter (1947) index. Both of these purport to be variables pertinent to forecasting severe convective weather, and both are based on simple parcel theory. The Showalter index is determined by the temperature difference between a hypothetical parcel lifted from 850 hPa to 500 hPa (a calculated diagnostic variable) and that at 500 hPa (an observed diagnostic variable). The Showalter index was originally proposed because it uses only three observed variables: 500-hPa and 850-hPa temperatures, and 850-hPa dewpoint temperature. At the time, manual processing of soundings meant that the mandatory pressure-level data were available more quickly than the rest of the sounding. Calculating an index requiring only mandatory-level data was advantageous because it gave forecasters a quick look at the convective instability before the whole sounding came in later.

The temperature of a parcel lifted from 850 to 500 hPa depends strongly on the 850-hPa dewpoint depression. When that value is large, unless the 850–500-hPa lapse rate is very nearly dry adiabatic, the parcel is likely to arrive at 500 hPa colder than the observed 500-hPa temperature because it will have followed a dry adiabat for most of its ascent. Conversely, when the 850-hPa temperature–dewpoint spread is small, the parcel will ascend mostly along a moist adiabat and is likely to be warmer than the observed 500-hPa temperature, unless the lapse rate is less than moist adiabatic. So far, so good, but consider the example shown in Fig. 1a. If, as in this case, the low-level moisture decreases rapidly just below 850 hPa, the Showalter index may be calculated correctly, but its implications about the convective

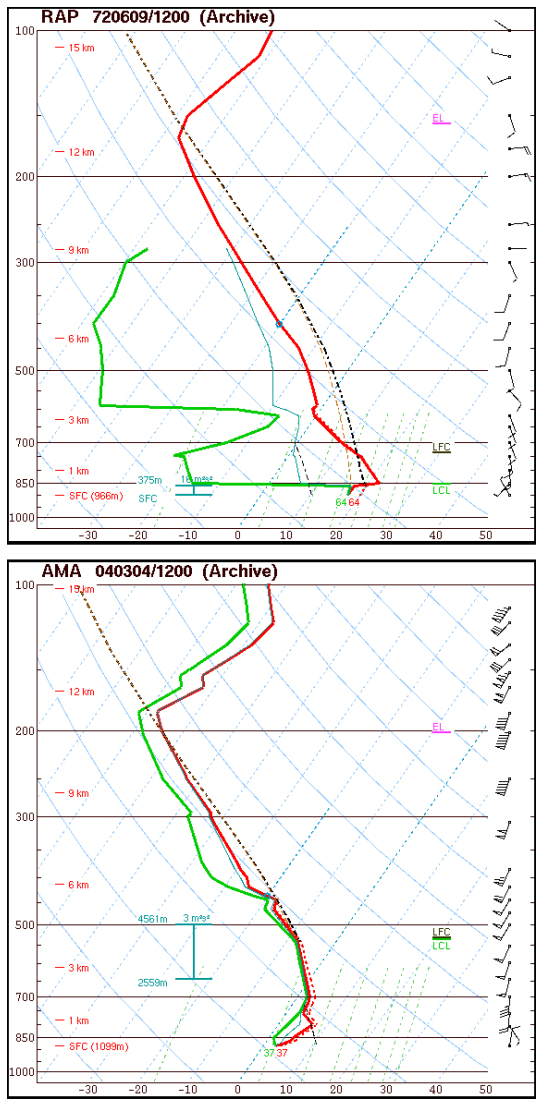

Figure 1. Rawinsonde plots on a Skew-*T*–log *p* diagram for (a) Rapid City, SD on 09 June 1972, and (b) Amarillo, TX on 04 March 2004. See text for discussion.

instability of the atmosphere can be misleading. This example was taken on the morning of a devastating flash flood in the vicinity that evening—the calculated Showalter index nominally would indicate little or no chance for thunderstorms simply because of the low dewpoint temperature at 850 hPa.

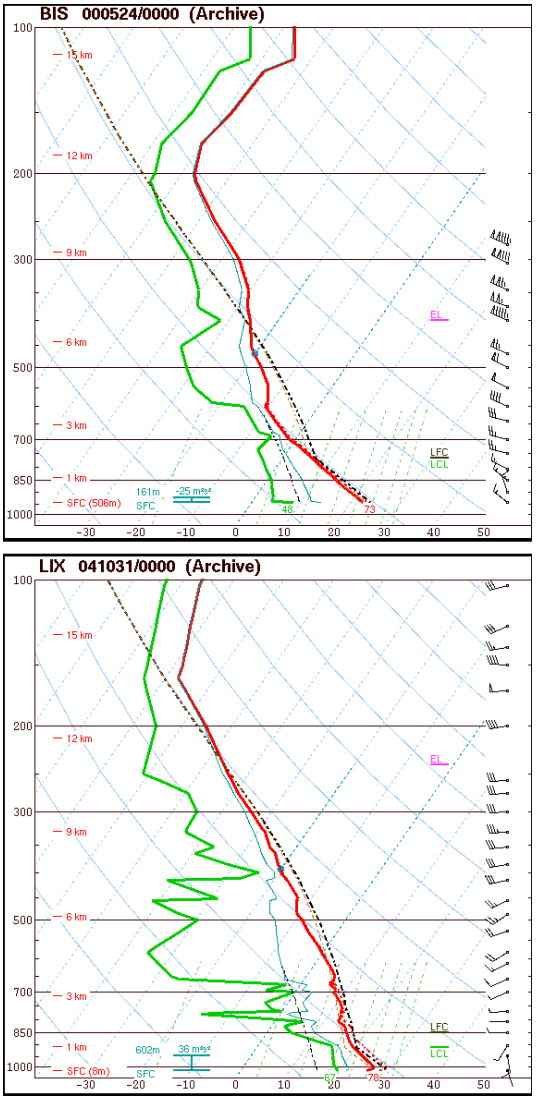

Figure 2. Soundings with very nearly equal CAPE values, but showing distinctly different vertical distributions of CAPE—(a) for Bismarck, ND at 00 UTC on 24 May 2000, and (b) for Slidell, LA at 00 UTC on 31 October 2004.

Indices can be affected dramatically when soundings rise through precipitation (as in Fig. 1b), or from errors that just happen to fall at mandatory pressure levels, thereby rendering calculation of parameters such as the Showalter index unrepresentative or even meaningless. This sensitivity to a small number of observations occurs in part because the Showalter index involves a differential quantity, and derivatives generally are much more sensitive to errors than are the basic quantities being differenced.

On the other hand, the CAPE calculation is a vertical integral that uses measurements at more than two levels. By virtue of being an integral quantity, rather than a derivative, calculation of CAPE is inherently less sensitive to small differences that might arise from measurement errors. Unfortunately, that apparently useful property of integration renders its values non-unique; that is, you can get the same CAPE value from distinctly different vertical distributions of a parcel's thermal buoyancy (Fig. 2 - see also Blanchard 1998). It is likely that one would interpret these soundings differently in terms of the weather forecasts, but considering only the CAPE values without seeing the soundings would permit no such discrimination.

Moreover, as discussed in Brooks et al. (1994), finding a *representative* sounding can be something of a challenge, owing to undersampled variability in space and time. This undersampled variability also affects any interpretation of the Showalter index, of course. At least in the short term, little can be done about measurement errors and undersampling. However, when considering how to interpret diagnostic variables, it seems quite unlikely to be able to define a diagnostic variable that distills the relatively rich complexity of a complete sounding into a single number that isn't vulnerable to being unrepresentative under some circumstances. In fact, we assert that *no* single number can replace the value of a forecaster simply looking at the soundings, as well as looking at diverse diagnostic variable computations based on those soundings. Any obvious errors in the sounding, or any characteristics that would render diagnostic variable computations misleading, will be apparent to someone experienced in sounding interpretation.

Further, we suggest that this principle applies to *any* diagnostic variable, not just those based on soundings. The diagnostic variable calculations still add value beyond what our hypothetical forecaster can see simply by viewing the data on a display, but using *only* the diagnostic variables as a substitute for considering all the information in the data is inherently risky in making weather forecasts.

Volatility is associated with the specific way in which the observations are used to construct the

fields of a diagnostic variable. Consider the horizontal divergence, expressed in conventional meteorological notation as:

$$div_h = \nabla_h \bullet \mathbf{V}_h = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right),$$

where (u,v) is the Cartesian coordinate (x,y) form of the horizontal velocity vector. When expressed in terms of natural coordinates—along and normal to the horizontal wind vector—Panofsky (1964, p. 33 ff.) has shown that under "normal circumstances" (i.e., quasi-hydrostatic flow) the horizontal divergence calculation involves the difference between two relatively large numbers, so it can be quite sensitive to small changes in the winds. Note that this sensitivity is not present when calculating the vertical component of the vorticity, ζ, where

$$\varsigma = \mathbf{k} \bullet \nabla \times \mathbf{V} = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right),$$

and where **k** is the unit vector in the vertical—the two terms in this calculation have no characteristic tendency to cancel one another.

Further, a divergence (or vertical vorticity) calculation is dependent on the resolution of the data. To show this, assume a characteristic magnitude for the difference in wind velocity between two points (δV, which is the same order of magnitude as the wind velocity itself, V), separated by a characteristic distance scale L, so that

$$div_h \sim \frac{\delta V}{L},$$

which is a simple scaling law for the divergence (or vorticity) magnitude. The order of δV turns out to be not very scale dependent, typically of order 1–10 m s^{-1}. This characteristic difference rarely reaches 100 m s^{-1} or becomes as small 0.1 m s^{-1}. Thus, over a fairly wide range of scales, the calculated divergence tends to depend most strongly on the distance between sample points.[2]

If divergence is calculated from a sparse network of points, such as rawinsonde sites (L~400 km), a rough order of magnitude for the divergence, according to the above scaling rule, is about (1–10 m s^{-1}) ÷ 400 km = 2.5 x (10^{-6}–10^{-5}) s^{-1}. In contrast, for the network of surface observations (L~100 km), the simple scaling rule gives a value of 1 x (10^{-5}–10^{-4}) s^{-1}, which is four times larger.

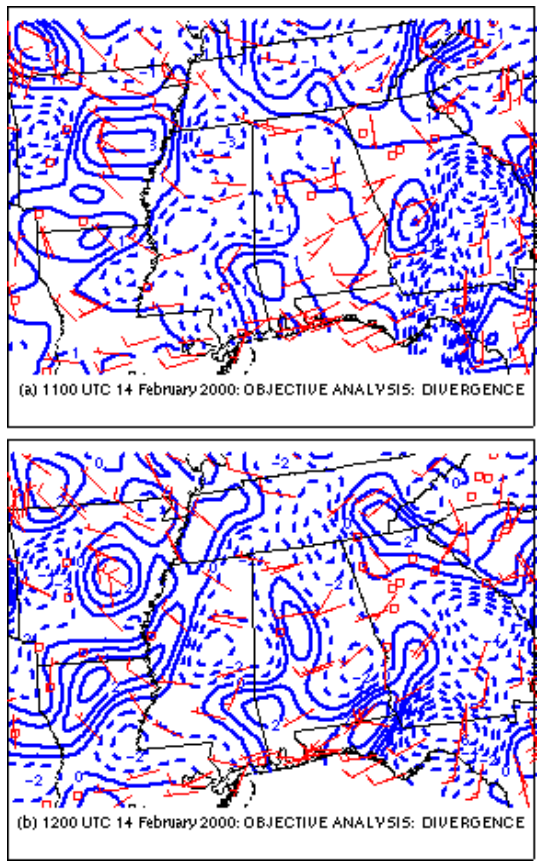

Figure 3. Objective analyses of surface divergence at (a) 1100 UTC and (b) 1200 UTC 14 February 2000, illustrating the volatility of the details. Solid contours are nonnegative divergence, starting at zero, and dashed contours are negative values. Conventional wind barbs show the observations used in the calculations.

According to this simple scaling law, horizontal divergence is inversely proportional to the scale of the data resolution over a wide range of scales.

Like the Showalter index, divergence calculations can be volatile. This volatility also follows from the simple scaling law above: if winds are only known to within 2 m s^{-1}, then a divergence

[2] Note that at *synoptic* scales (where L ~ 1000 km), because the wind is nearly geostrophic, the characteristic divergence magnitude is an order of magnitude less than this simple scaling law predicts because the geostrophic wind calculated on an *f*-plane in many meteorological coordinate systems is nondivergent (see Doswell 1988). This can be accounted for by changing the scaling law to

$$div_h \sim Ro\frac{V}{L}$$

as shown by Haltiner and Williams (1980, p. 57), where Ro is the Rossby number and is of order 10^{-1} on synoptic scales, but dynamics-based scale analysis is beyond the scope of this paper. Again, synoptic-scale vertical vorticity scaling is not subject to this consideration.

estimate based on a station separation of 100 km is known only to within 2 x 10^{-5} s^{-1}. The fields tend to be noisy and behave somewhat erratically over time as a result of undersampled variability in the wind field and measurement errors, although the basic shape of the fields might be fairly consistent from one time to the next (Fig. 3).

To some extent, this volatility can be reduced by heavy smoothing that limits the spatial and/or temporal scales of the features retained to those that can be depicted reliably in the analysis (Doswell 1977). The consistency of the basic shape of the fields, combined with the volatility of the details, is a direct indication of the sensitivity of the details in the field to small changes in the wind.

4. Indices

Having considered several of the issues that confront users of diagnostic variables, we are now prepared to consider the topic of indices. Indices have a long history in severe storms forecasting that perhaps began with the Showalter index (SI). Their use has continued through a growing plethora of constructs, including, among many others, the lifted index (LI), SWEAT index, bulk Richardson number (BRN), energy–helicity index (EHI), Cross Totals (CT), SCP, significant tornado parameter (STP), and enhanced stretching potential (ESP; J. Davies 2005, personal communication). These and selected other indices are listed in Table 1. Indices also have been applied to forecasts other than severe convective weather. One example is the Garcia (1994) method for forecasting snowfall. Wetzel and Martin (2001) and Schultz et al. (2002) discuss the scientific integrity of the Garcia (1994) method and other approaches.

Some indices are associated with a physical argument. The early stability indices (e.g., SI, LI) were based on simple parcel theory, although we have suggested some issues with their use as diagnostic variables previously, to say nothing of their use as forecast variables. The dimensionless BRN is at least related to the true Richardson number, but the actual physical significance of any Richardson number to the physics of deep moist convection is unclear, as the original intent of the Richardson number was to address topics in turbulence theory (Tennekes and Lumley 1972, p. 98 ff.).

Many of these indices, including the SWEAT index, CT, EHI, SCP, and STP, have combined variables in ways that have no physical rationale. In other words, the process of forming sums, products, and ratios has not been done in accordance with a formula originating in the mathematics describing a physical process. Rather, the mathematical expression for the index is more or less arbitrary. Why a sum of two variables divided by a third? Why not one variable raised to a power defined by a second variable multiplied by a third?

As we have mentioned for the case of simple diagnostic variables calculated from observations (cf. section 2b), combining two or more variables in a way that has a physical basis affects the interpretation and use of the resulting variable. If a variable is conserved during certain physical processes, for example, that is quite relevant to its application in diagnosis or forecasting.

At issue is whether or not the variable can be related to physical principles. Examples of diagnostic variables based on physical principles might be something like the static stability time tendency, potential vorticity, or energy dissipation rate. Combining two or more variables in an arbitrary way leaves open many questions and makes it difficult to relate the variable to any physical understanding of the process. The individual diagnostic variables used to form an index may have physical relevance to the problem at hand, but when a specific formula combining them is unphysical, this can be problematic.

Furthermore, the physical dimensions of these indices may or may not make any physical sense. For example, the ratio of CAPE to shear (a ratio used in one form or another for several indices within Table 1) has dimensions of J kg^{-1} s, the product of energy per unit mass and time, whereas the *product* of CAPE and shear has dimensions of J kg^{-1} s^{-1}. The former has no obvious physical interpretation, whereas the latter has dimensions of energy per unit mass per unit time, which at least can be related to terms in an energy budget. Thus, the product of CAPE and shear might yield a more physically meaningful parameter than the quotient of CAPE and shear.

Table 1. A selection of indices commonly used in the United States for severe storm forecasting. In the formulae, T denotes a temperature and D denotes a dewpoint temperature in ºC, with a subscript indicating at what mandatory pressure level (in hPa) this value is to be taken from; α denotes the specific volume and a subscript *lp* denotes a value associated with a lifted parcel; LFC stands for a lifted parcel's level of free convection and EL stands for its equilibrium level. For the Lifted index, the lifted parcel is for a surface parcel with *forecast* properties at a representative time of day. For the SWEAT index, V denotes a wind speed (in knots), and ΔV denotes a wind direction difference (in degrees). For the Bulk Richardson number, $\overline{U}$ denotes the density-weighted speed of the mean vector wind in the layer 0–6 km, and U_0 denotes the speed of the mean vector wind in the layer from the surface to 500 m—the quantity $\left(\overline{U}-U_0\right)$ is sometimes referred to as the "BRN shear". For the storm-relative helicity, **C** denotes the storm motion vector. See Thompson et al. (2003) for an explanation of symbols used for the SCP and STP calculations.

Index Name	**Formula**	**Reference**
Showalter index	SI = $T_{500} - T_{lp(850\ hPa)}$	Showalter (1947)
Lifted index	LI = $T_{lp(fcst\ surface)} - T_{500}$	Galway (1956)
K-index	$\mathrm{K} = \left(T_{850} - T_{500}\right) + D_{850} - \left(T_{700} - D_{700}\right)$	George (1960, pp. 407–415)
Convective available potential energy	$\mathrm{CAPE} = \int_{LFC}^{EL} \left(\alpha_{lp} - \alpha\right) dp$	Glickman (2000, p. 176)
Vertical totals	VT = $T_{850} - T_{500}$	Miller (1972)
Cross totals	CT = $D_{850} - T_{500}$	Miller (1972)
Total totals	TT = VT +CT	Miller (1972)
SWEAT index	SWEAT = 20(TT–49ºC) + 12D_{850} + 2V_{850} + V_{500} + 125[sin($\Delta V_{500\text{-}850}$) + 0.2]	Miller (1972)
Bulk Richardson number	$\mathrm{BRN} = \dfrac{\mathrm{CAPE}}{\frac{1}{2}\left(\overline{U} - U_0\right)^2}$	Weisman and Klemp (1982)
Storm-relative helicity	$\mathrm{SRH} = -\int_{z_o}^{z} \mathbf{k} \bullet \left[\left(\mathbf{V}_h - \mathbf{C}\right) \times \dfrac{\partial \mathbf{V}_h}{\partial z} \right] dz$	Davies-Jones et al. (1990)
Energy-helicity index	$\mathrm{EHI} = \dfrac{(\mathrm{CAPE})(\mathrm{SRH})}{160{,}000}$	Hart and Korotky (1991)
Supercell composite parameter	$\mathrm{SCP} = \left(\dfrac{\mathrm{MUCAPE}}{1000\ \mathrm{J\,kg^{-1}}}\right)\left(\dfrac{\mathrm{SRH}_{0-3\,\mathrm{km}}}{100\ \mathrm{m^2\,s^{-2}}}\right)\left(\dfrac{\overline{U} - U_0}{40\ \mathrm{m\,s^{-1}}}\right)$	Thompson et al. (2003)
Significant tornado parameter	$\mathrm{STP} = \left(\dfrac{\mathrm{MLCAPE}}{1000\ \mathrm{J\,kg^{-1}}}\right)\left(\dfrac{\mathrm{SHR}_{0-6\,\mathrm{km}}}{20\ \mathrm{m\,s^{-1}}}\right)\left(\dfrac{\mathrm{SRH}_{0-1\,\mathrm{km}}}{100\ \mathrm{m^2\,s^{-2}}}\right)\left(\dfrac{(2000\,\mathrm{m} - \mathrm{MLLCL})}{1500\ \mathrm{m}}\right)$	Thompson et al. (2003)
Enhanced stretching potential	$\mathrm{ESP} = \left(\left.\dfrac{\partial T}{\partial z}\right\|_{0-2\,\mathrm{km}} - 7^{\circ}\mathrm{C\,km^{-1}}\right)\left(\dfrac{\mathrm{MLCAPE}_{3\,\mathrm{km}}}{50\ \mathrm{J\,kg^{-1}}}\right)$	Davies (2005 – personal communication)

a. Example of representative problems with indices: EHI

In order to demonstrate these issues using one of these indices, we use the EHI as a representative example, although any of the above-listed indices would reveal similar problems. As described in Rasmussen and Blanchard (1998, p. 1154), "This index [using the 0–1-km layer] is used operationally for supercell and tornado forecasting, with values larger than 1.0 indicating a potential for supercells, and EHI > 2.0 indicating a large probability of supercells." Our problems with EHI center around five issues: combination of ingredients, arbitrary construction, choice of scaling constant, ambiguous physical meaning, and lack of proper validation.

First, EHI is a combination of two separate variables that may not even be collocated during the event and can evolve separately, as discussed in section 3. For instance, although large CAPE and large SRH are both pertinent to supercell forecasting, they need not be precisely collocated in space—many severe storm forecasters believe CAPE and SRH need simply be in proximity to each other, perhaps with some overlap. Therefore, a variable based on the combination of these two variables may not adequately reflect the true potential for storms.

Further, EHI is the combination of two ingredients in an unphysical, arbitrary fashion. Can it be shown that the formation of a supercell depends in some well-defined physical way on the product of CAPE and SRH? Although we cannot exclude such a possibility in the future, as of this writing, this product has not been shown to be physically pertinent, in the sense of appearing in some physically-relevant mathematical formula. Schultz et al. (2002) made a similar argument about the lack of scientific justification in reference to the PVQ parameter (the product of PV_{es} and the divergence of Q when both are negative) defined and proposed by Wetzel and Martin (2001).

Consider the following argument. Imagine a hypothetical world where the EHI's two constituent parameters were all that is needed to forecast supercell tornadoes perfectly. Then picture a two-dimensional (2D) phase space of SRH and CAPE in which some irregular region within this phase space was associated with the occurrence of supercell tornadoes. By assumption, such a scenario would represent a perfect forecasting tool: if the values of SRH and CAPE, fall within this presumably complex region of 2D phase space, supercell tornadoes *always* occur. Anywhere outside of this region, supercell tornadoes *never* occur. Use of the EHI compresses the information within this hypothetical 2D space into a single number, eliminating our ability to apply the information in this phase space.

Furthermore, EHI is scaled by 160,000 somewhat arbitrarily, so that whether the value of the EHI is 1 or 100 is similarly arbitrary. EHI's scaling constants are associated with a "standard" value for CAPE of 1000 J kg^{-1} and for SRH of 160 m^2 s^{-2}—note that these units are actually equivalent and so EHI has dimensions of $(J\ kg^{-1} = m^2\ s^{-2})^2$, which has little obvious physical interpretation. For indices of this sort, the scaling constants are determined by what the developer felt to be typical values for the input variable. What is typical for some variable associated with severe weather in one part of the world may not be typical elsewhere. Hence, this can lead to incorrect interpretations of the index when used outside of the region in which it was developed (Tudurí and Ramis 1997). It is reasonable, we believe, to ask that a forecast parameter's utility and interpretation should not vary from one location to another.

Moreover, consider our hypothetical world again, in which the constituent parameters for EHI can be used to construct a 2D phase space that includes some complex region wherein supercell tornadoes always occur. It is conceivable that *some* transformation of the EHI's constituent variables (CAPE and SRH) would convert the complex region into a simple one—say, a circle. However, it seems highly unlikely that the existing scaling and unphysical construction of EHI could correspond to such a transformation.

This leads to the next point: how does a forecaster interpret EHI values? Put another way, does a situation where EHI=2 mean that supercells are twice as likely as a situation where EHI=1? If the CAPE doubles, doubling the EHI (assume SRH remains unchanged), does this imply twice the chance of supercells? There might be some empirical way to determine the significance of these values, but there is no physical rationale for interpreting them. Is the relationship between EHI and severe weather linear or nonlinear? How does the lagged correlation between EHI and the observed weather

vary as a function of the lag time? Such questions ought to be of concern to any severe weather forecaster, but are certainly not readily answerable by the method EHI was constructed.

Finally, how has the forecast value of EHI been validated? Rasmussen (2003) described his evaluation of EHI for three classes of observed proximity soundings: supercells with significant tornadoes (those rated F2 and greater), supercells without significant tornadoes (no tornadoes reported, or only F0 or F1 tornadoes), and nonsupercell convection, defined in Rasmussen and Blanchard (1998). The weather events occurred anywhere from three hours before to six hours after the nominal 00 UTC sounding time. Rasmussen (2003) found, "only 25% of [proximity soundings from supercells without significant tornadoes] had EHI > 0.5, whereas nearly 2/3 of the [proximity soundings from supercells with significant tornadoes] had values this large." Given the number of soundings in each dataset from Rasmussen and Blanchard (1998), this means about 30 events occurred in each category—a relatively small sample size. Taken together, these results indicate that when a supercell occurs with an EHI > 0.5 there is about a 50% chance of it producing a significant tornado. This figure amounts to a conditional probability, where the condition is the presence of a supercell.

Rasmussen and Blanchard (1998) used a contingency table and scatter diagrams (see section 5) to evaluate the diagnostic potential for EHI and several other candidate variables. But their study evaluated indices from proximity soundings on the basis of observed events occurring in a time period around the sounding's nominal time. Therefore, it is not truly an analysis of the forecast potential for the variables considered—it is instead directed at a related problem: How well do *diagnosed* values of the indices discriminate among the observed events? This sort of analysis does not consider the topic of the lagged correlation between the proposed forecast parameter and the forecast severe weather events.

b. Pros and cons regarding the use of indices

We have shown, using the example of EHI, what types of problems can arise with the use of indices constructed in an arbitrary, unphysical way. We recognize that there are advantages as well as drawbacks to their use in severe weather forecasting.

As already noted, many severe weather forecasters already recognize the risks in relying on a single forecast parameter. Certainly most would never consider using a single variable to determine their forecast, but our experience suggests that *some* forecasters might be tempted to do so, perhaps because the use of some diagnostic variable (such as CAPE) is so widespread (e.g., the "disengaged" forecasters studied by Pliske et al. 2004). The pervasive use and ready availability of diagnostic variables is a trap for the unwary. In situations where the time pressure becomes intense, some might be inclined to do so in the interest of making a quick decision. Such a practice is antithetical to good forecasting, in general.

In a related line of reasoning, the press of time can encourage severe weather forecasters' use of indices and other diagnostic variables to obtain a quick look at the data for the purpose of identifying the hot spots upon to focus more attention on in a diagnosis. By itself, this is a reasonable strategy. However, a serious forecaster is not likely to get much forecasting help from such a cursory consideration of atmospheric structure. There is nothing inherently wrong with a quick look, unless the forecaster limits the diagnosis to that. For all the reasons we have described, a conscientious weather forecaster always should try to find the time to do a comprehensive analysis of the data. We believe that using ingredients-based forecasting methods (e.g., Doswell et al. 1996) is a scientifically sound way to keep the diagnosis within practical time limitations in an operational forecasting environment.

It also can be argued from a purely utilitarian perspective that if a forecast parameter works successfully in forecasting, no matter *how* it was derived, it seems unreasonable to ask forecasters to cease using it. We don't disagree with this at all, especially when proven forecast parameters based on physical arguments are either unavailable or demonstrably inferior to a nonphysically constructed variable. If it can be shown rigorously that an arbitrarily constructed index does indeed have forecast utility (see the following section), we do not advocate ignoring its proven value to the challenge of weather forecasting—unless a physically-based forecast parameter is known to be superior for forecasting. Generally, physical reasoning is always preferable for the construction of diagnostic variables and indices, owing to the relative ease with which such forecast parameters can be interpreted and applied globally.

5. Evaluation of forecast utility for a candidate prognostic variable

Because many diagnostic variables with potential forecast utility never have been tested rigorously as forecast parameters in their own right, we develop herein a general description of what we believe are the requirements that a proper forecast parameter would have to meet. As already discussed, diagnostic variables have their own specific purposes, but a diagnostic variable might also have the capability to make a reasonably accurate and perhaps even skillful prediction of the weather at some future time. See Murphy (1993) for a discussion of the difference between accuracy and skill.

One way to conduct a rigorous assessment of a variable as a forecast parameter is to use a developmental data set to form a classic 2 x 2 contingency table, the standard verification table when considering a dichotomous (yes/no) forecast for some dichotomous event (Wilks 2006, p. 260 ff). One example is given by Monteverdi et al. (2003) for tornadoes in California (see their Table 3 and Fig. 8). To create such a table for a potential forecast parameter, begin with choosing a threshold value for the candidate variable—forecast "yes" if the variable is at or above the threshold, and "no" if the variable is below the threshold. An assessment of the accuracy of the forecasts using the developmental dataset would be done by filling in the contingency table using that threshold. Optimizing the choice for the threshold value of the variable using the so-called Relative (or Receiver) Operating Characteristic curves associated with signal detection theory is possible; see Wilks (2006, p. 294 ff.) for more information.

The accuracy of the forecasts using that threshold can be assessed using standard methods based on the contingency table. The *skill* of the forecasts is determined by comparing the accuracy of the proposed forecasts based on use of that variable against the accuracy of some standard forecasting method (e.g., climatology or persistence, or some other forecast scheme, such as Model Output Statistics). If the forecast scheme using the proposed diagnostic variable shows statistically significant skill in comparison with some standard method, then it can be considered a useful forecast parameter.

To do a thorough assessment, however, another dataset is needed that is completely independent of the developmental dataset—in other words, a wholly different set of cases than those used for development and testing of the threshold values for the variable. If the results using the verification dataset are comparable to those found from the developmental data, confidence in the use of the variable as a forecast variable is correspondingly high. If there is a statistically significant difference between the results from the two datasets, then perhaps a larger sample is needed, but in any case, confidence in the forecast value of the variable (and its associated threshold value) is correspondingly low.

Two concerns often arise when a forecast parameter is proposed. First, many assessments of potential forecast parameter are done with a small number of cases, perhaps as few as one. Knowledge of how to determine an appropriate sample size is outside the scope of this paper; see the discussion of hypothesis testing by Wilks (2006, chapter 5). It is incumbent on the developer of a forecast parameter to provide a reasonably thorough test using a robust sample of enough cases. Second, many attempts to validate the utility of some variable as a forecast parameter make the logical mistake of considering only values of the parameter when forecast events are known to have occurred. Diagnostic variables are of little use in forecasting until they can be shown to discriminate successfully between events and nonevents. Correct predictions of nonevents are not inevitably easy (e.g., Doswell et al. 2002) although in forecasting severe storms (relatively rare events), many nonevents are obvious.

Further, some diagnostic variables may have value as forecast parameters, but it needs to be shown just how far in advance of the event they exhibit forecast accuracy and/or skill. That is, the accuracy of any diagnostic variable used as a forecast parameter is likely to increase as the time lag between the diagnosis and the event decreases, but this may not necessarily be a simple relationship. Contingency tables and full assessments as described above would have to be developed for a variety of diagnosis times relative to the beginning of the forecast events—say, 12, 6, 3, and 1 h before the actual events begin. Alternatively, an analysis of the time-lagged correlation between the forecast parameter and the observed weather could be carried out at a variety of lag times. However it is done, the accuracy of a candidate variable as a forecast parameter should be known as a function of time

before the event. Quantitative knowledge of forecast accuracy as a function of lead time is obviously important when using a diagnostic variable as a forecast parameter.

Another way to verify the potential of a forecast variable would be to construct a multidimensional scatter diagram (say, for the case of two dimensions, CAPE and shear) in which both events and nonevents are plotted with respect to observed values for the diagnostic variables. Using this plot, a probability of occurrence of the weather event as a function of its location in the scatter diagram could developed, perhaps facilitated by the use of kernel density estimation methods (e.g., Ramsay and Doswell 2005). Examples of such verification are found in Rasmussen (2003) and Brooks et al. (2003), although this method would require diagnostic variable values prior to the event, rather than proximity data.

Of course, the preceding does not present the *only* ways to assess the effectiveness of a proposed forecast variable. Many other methods could be used, but it does represent the level of rigor we believe is necessary before asserting that a diagnostic variable has real value as a true forecast parameter.

6. Conclusions

In our experience, many severe weather forecasters and researchers are seeking a "magic bullet" when they offer yet another combined variable or index for consideration, whether or not they realize it. If some single variable or combination of variables made forecasting so simple, then the need for human forecasters effectively vanishes. There may be other reasons for the demise of human forecasters, but distilling the complex atmosphere with its nonlinear, possibly chaotic, interactions into an all-encompassing variable seems improbable. Any forecaster seeking to find such a variable is not only unlikely to be successful, but, if success were achieved, the need for a human forecasting that event vanishes! Should advances in the science of severe convective storms ever produce such a forecast parameter, or should NWP models become near-perfect in terms of forecasting severe convection, then the need for human forecasters will indeed disappear (e.g., Doswell 2004), whatever our wishes might be. But it is our belief that this is not very likely to happen soon. Even if such an unlikely development ever occurs, in the interim, it remains incumbent on forecasters to use the information at their disposal as effectively as possible.

Acknowledgments. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA17RJ1227, U.S. Department of Commerce. We would like to thank our reviewers for their careful and thoughtful reviews and helpful suggestions, which have improved our presentation.

REFERENCES

Banacos, P. C., and D. M. Schultz, 2005: The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives. *Wea. Forecasting,* **20,** 351–366.

Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting*, **13**, 870–877.

Brock, F. V., and S. J. Richardson, 2001: *Meteorological Measurement Systems*. Oxford University, 290 pp.

Brooks, H. E, C. A. Doswell III, and J. Cooper, 1994: On the environments of tornadic and non-tornadic mesocyclones. *Wea. Forecasting,* **10,** 606–618.

————, J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94.

Davies-Jones, R., D. Burgess, and M. Foster, 1990: Test of helicity as a tornado forecast parameter. Preprints, *16th Conf. on Severe Local Storms*, Kananaskis Park, AB, Canada, Amer. Meteor. Soc., 588–592.

Doswell, C. A. III, 1977: Obtaining meteorologically significant surface divergence fields through the filtering property of objective analysis. *Mon. Wea. Rev*., **105**, 885–892.

————, 1986: Short range forecasting. *Mesoscale Meteorology and Forecasting*. P. Ray, Ed., Amer. Meteor. Soc., 689–719.

————, 1988: Comments on "An improved technique for computing the horizontal pressure-gradient force at the earth's surface." *Mon. Wea. Rev*., **116**, 1251–1254.

————, 2004: Weather forecasting by humans—Heuristics and decision making. *Wea. Forecasting*, **19**, 1115–1126.

————, and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting,* **9,** 625–629.

————, and P. M. Markowski, 2004: Is buoyancy a relative quantity? *Mon. Wea. Rev.*, **132,** 853–863.

————, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11,** 560–581.

————, D. V. Baker and C. A. Liles 2002: Recognition of negative mesoscale factors for severe weather potential: A case study. *Wea. Forecasting*, **17**, 937–954.

Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.*, **43**, 528–529.

Garcia, C., Jr., 1994: Forecasting snowfall using mixing ratios on an isentropic surface—An empirical study. NOAA Tech. Memo. NWS CR-105, PB 94-188760 NOAA/NWS, 31 pp. [Available from NOAA/National Weather Service Central Region Headquarters, Kansas City, MO 64106-2897.]

George, J. J., 1960: *Weather Forecasting for Aeronautics*. Academic Press, 673 pp.

Glickman, T. S., Ed., 2000: *Glossary of Meteorology*. 2nd ed., Amer. Meteor. Soc., 855 pp.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*. 2d ed., John Wiley & Sons, 477 pp.

Hart, J. A., and W. Korotky, 1991: The SHARP workstation v1.50 users guide. National Weather Service, NOAA, U.S. Department of Commerce, 30 pp. [Available from NWS Eastern Region Headquarters, 630 Johnson Ave., Bohemia, NY 11716.]

Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7,** 588–612.

Miller, R. C., 1972: Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. Tech. Report 200(R), Headquarters, Air Weather Service, Scott Air Force Base, IL 62225, 190 pp.

Monteverdi, J. P., C. A. Doswell III, and G. S. Lipari, 2003: Shear parameter thresholds for forecasting tornadic thunderstorms in northern and central California. *Wea. Forecasting*, **18,** 357–370.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.

Panofsky, H., 1964: *Introduction to Dynamic Meteorology*. Pennsylvania State University, 243 pp.

Pliske, R. M., B. Crandall, and G. Klein, 2004: Competence in weather forecasting. *Psychological Investigations of Competence in Decision Making*. K. Smith, J. Shanteau, and P. Johnson, Eds., Cambridge University Press, 40–68.

Ramsay, H., and C. A. Doswell III, 2005: A sensitivity study of hodograph-based methods for estimating supercell motion. *Wea. Forecasting*, **20,** 954–970.

Rasmussen, E. N., 2003: Refined supercell and tornado forecast parameters. *Wea. Forecasting*, **18,** 530–535.

————, and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13,** 1148–1164.

Sanders, F., and C. A. Doswell III, 1995: A case for detailed surface analysis. *Bull. Amer. Meteor. Soc.*, **76,** 505–521.

Schaefer, J. T., 1986: Severe thunderstorm forecasting: A historical perspective. *Wea. Forecasting*, **1,** 164–189.

Schultz, D. M., J. V. Cortinas Jr., and C. A. Doswell III, 2002: Comments on "An operational ingredients-based methodology for forecasting midlatitude winter season precipitation." *Wea. Forecasting*, **17,** 160–167.

Showalter, A. K., 1947: A stability index for forecasting thunderstorms. *Bull. Amer. Meteor. Soc.*, **34**, 250–252.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18,** 1243–1261.

Tennekes, H., and J. L. Lumley, 1972: *A First Course in Turbulence*. MIT Press, 300 pp.

Tudurí, E., and C. Ramis, 1997: The environments of significant convective events in the western Mediterranean. *Wea. Forecasting*, **12**, 294–306.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520.

Wetzel, S. W., and J. E. Martin, 2001: An operational ingredients-based methodology for forecasting midlatitude winter season precipitation. *Wea. Forecasting*, **16,** 156–167.

————, and ————, 2002: Reply. *Wea. Forecasting*, **17,** 168–171.

Wilks, 2006: *Statistical Methods in the Atmospheric Sciences*, 2d ed. Academic Press, 627 pp.

REVIEWER COMMENTS

[Authors' responses in *blue italics*.]

REVIEWER A (Richard L. Thompson):

Initial Review:

Recommendation: Accept with major revision

The paper provides an overview of several common parameters/indices used in severe storm forecasting, and outlines a process for developing proper forecast parameters. On the surface, this topic appears worthy of consideration, but I have several concerns regarding the tone of the paper and the apparent motivation of the authors. Without providing supporting evidence, the authors infer that a substantial number of forecasters have no clue how to interpret the convective parameters/indices they discuss. Can the authors site anything more specific than a few vague references? It is important to establish reasonable motivation for this work because that motivation identifies the target audience. At this point, it could be anyone in severe storm meteorology, but this work seems most appropriate for undergraduate meteorology students. I can only speak for myself, but I found this paper to be mildly degrading to professional forecasters.

We regret that you had this reaction to the manuscript, as that surely was not our intention. Our intention was to educate forecasters, not degrade them. Our intended audience was anyone in severe storms meteorology, as we believe that even experts would benefit from revisiting the ideas in this manuscript. In the revised manuscript, we have softened the tone and rewritten the introduction so that our intentions and intended audience are more clear.

As far as whether we have any documentation on this, we have no formal references. We have clarified our perceptions as our own, rather than from some formal refereed literature, which almost certainly does not exist. We believe it is reasonable to allow us the privilege of making a nonspecific observation of things we've seen personally. We would prefer not to point a finger at specific individuals, even if we could. Nevertheless, even if this interpretation of our perception is incorrect across the board for all forecasters, as it surely is, we still feel our message is relevant, as noted by one of the other reviewers of this manuscript.

The authors must provide a more balanced account of parameter and index use by forecasters.

We have put more emphasis on the value of parameters up front in the paper, hoping to provide a more balanced account. However, this article is intended as a position piece, so we should be allowed some license to present a different side of the argument than is currently seen in most papers. A balanced perspective can be derived by consulting our references, many of which have managed to present a very rosy picture of their proposed variables, which is also far from a balanced perspective.

They paint a very limited and unrepresentative picture of the problem, and almost completely overlook the spatial composite aspects of many parameters.

We disagree with this statement. For example, we have discussed the spatial patterns associated with divergence and the issue with overlap of parameters in the construction of composite parameters. In fact, many diagnostic parameters are presented and evaluated in the literature in the context of proximity soundings to storms. Therefore, although we have some discussion of the spatial aspects, we are nevertheless discussing these parameters the way many people use them and the way the discussion is framed in these papers.

They focus on EHI as a specific example, stating that similar problems plague a majority of the parameters. I am quite familiar with the significant tornado parameter (STP, Thompson et al. 2003), so I will use that as a counter example to their EHI discussion. The STP is a product of four normalized ingredients that have shown some ability to discriminate between tornadic and nontornadic supercells: 0-1 km AGL storm-

relative helicity (SRH), 0-6 km bulk wind difference, 100 mb mean parcel CAPE, and 100 mb mean parcel LCL height. Independent studies, utilizing independent data sets, each confirmed the ability of these ingredients to discriminate between supercells and nonsupercells, as well as tornadic and nontornadic supercells, in a diagnostic sense.

Each one of these variables alone cannot adequately discriminate among tornadic and nontornadic supercells (e.g., Rasmussen and Blanchard 1998). If a single variable were an adequate discriminator, then presumably the STP wouldn't be needed. CAPE, on its own, has not shown much ability to discriminate between tornadic and nontornadic supercells (e.g., Rasmussen and Blanchard 1998; Monteverdi et al. 2003). There are questions about the pertinence of LCL height on its own – it appears to be more useful in the USA than in other countries. Hence, we're not so easily convinced about the individual discriminatory power of these constituent variables as the reviewer seems to be. We cited a specific study (Monteverdi et al. 2003) to support our contention. Have similar studies of these terms individually shown discriminatory power capable of being operationally useful? If so, we're unaware of them.

In any case, the reviewer is missing our point about why the construction of composite parameters is inappropriate. We will take that to be our fault for not being more clear. Consider the following. Imagine a world where the STP's four constituent parameters were all that were needed to forecast supercell tornadoes perfectly. Imagine a 4D phase space of SRH, shear, CAPE, and LCL in which some convoluted 4D volume within this phase space was associated with the occurrence of supercell tornadoes. Presumably, such a scenario would represent a perfect forecasting tool: find the values of SRH, shear, CAPE, and LCL, and, if they fit inside this convoluted volume of phase space, supercell tornadoes always occur. Anywhere outside of this volume, supercell tornadoes never occur. The construction of the STP takes all the rich information contained within the 4D space and compresses it into a single number, eliminating our ability to use the complex 4D volume occupied by tornadic supercells in this phase space. This is analogous to allowing a single value of CAPE to represent the detailed information in a vertical sounding. Thus, we feel composite parameters are necessarily more inferior to a better understanding of the 4D phase space.

The STP provides a simple means of compositing the ingredients, without all of the mess of a "spaghetti" chart.

We disagree. See above. Also, we have addressed this issue in the rewritten manuscript.

Also, the STP purports to diagnose the supercell tornado threat based on current conditions, or *expected* conditions at some point in the future.

There's a big difference between those two. If STP has some value as a diagnosis of current conditions, it might also have some value as a forecast variable. Which use are you describing here?

I am aware of no attempts to forecast tornadoes in the afternoon based on the morning values of STP.

That doesn't mean that such efforts don't exist. We stand by our statement.

A more common (and reasonable) approach is to examine the ingredients independently (and in combination) early in the day, and then account for any observed or expected changes in the ingredients as the day progresses. Composite indices can be very useful in identifying the degree of "overlap" in the ingredients, and how the spatial distribution of ingredients is changing with time.

We don't have any problem with this, subject to some caveats, which we've provided in the revised manuscript.

To make a weather forecast, one must diagnose the current state of the atmosphere, and then anticipate future changes to the current state. Numerical model guidance can provide an expected state of the atmosphere when focusing on longer time ranges, while extrapolation of observed trends may be preferred for short-term forecasts. Parameters and indices are useful in the initial diagnoses, later prognoses, and in comparing one event to another. Until the science of meteorology arrives at a complete mathematical ex-

pression describing tornadogenesis (and many other processes), severe storm forecasters will necessarily rely on incomplete approximations, including various indices. I applaud the efforts of the authors to encourage forecasters to understand the strengths and weaknesses of indices and parameters used in the forecast process. However, they run the risk of alienating much of their intended audience by placing too much emphasis on improper use of indices, while largely ignoring many positive benefits.

Hopefully, our revised manuscript provides a more balanced argument. Nevertheless, we reserve the right in this manuscript to favor one side more than the other.

Specific Substantive Comments:

1. Introduction, end of 1st paragraph: The authors are quick to discount 30 years of forecaster experience, all because "rigorous verification" has not established these variables as *prognostic*? Why is actual forecast experience worth so little in evaluating the forecast utility of a variable? In the absence of scientific evidence supporting either the forecasters or your claim, this amounts to little more than a difference of opinion!

Forecaster experience has been shown to be an unreliable substitute for actual verification (see Doswell's paper on heuristics). Forecaster confidence grows with experience, but forecast accuracy typically grows fast early in a forecaster's career and then levels off.

2. End of P. 1: Words like "at times, "some", and "occasionally" do not make a convincing argument. Can you cite more specific examples and/or circumstances?

See above argument.

3. Top of P. 2, last sentence: I see no evidence to support your assertion that these diagnostic variables' "capability to inform forecasters about the weather in the future can be quite limited, at best." Why such a narrow view of forecasting? The authors are taking an unrealistic stance by isolating the indices/parameters, and assuming that forecasters simply look at the current value and make a forecast? I have not actually seen any professional meteorologist do this, and I have worked in the NWS for 13 years.

We think this statement is almost certainly untrue, but, of course, we have no way to prove that. Neither does the reviewer have a way to prove that our perspective is incorrect.

4. 2nd to last full paragraph, P. 2: Who actually makes a severe storm forecast based solely on current index values, without considering potential changes in the input variables? The authors seem to misunderstand the practical use of the indices in question, because even the most simplistic forecasters usually consider *trends* in the index, as well. Severe storm forecasting is quite complex, as the authors are well aware, yet they rely on a ridiculously simple (and flawed) forecast process as the motivation for this paper? Instead of degrading the forecasters, the authors should outline a process for developing both proper diagnostic and prognostic variables, and leave it at that.

Clearly, we have failed to convince the reviewer that the intention of our paper is not to degrade the forecaster, but to educate the forecaster to do a better job. From our collective experience educating forecasters, rules of thumb and incorrect use of diagnostic parameters are being employed by some forecasters. Schultz has seen it in the context of winter storm forecasting, and Doswell has seen it in the context of severe-storms forecasting. Unfortunately, some faculty still teach their students these parameters and rules of thumb. Hopefully, the revised manuscript helps clarify our intentions.

5. End of P. 7: I dispute the authors' claim that "these indices…have combined variables in ways that have no physical rationale." In the absence of a "supercell tornado equation", we must deal with incomplete information and understanding. Since you mentioned the STP, I would like to elaborate. The STP is a product of four variables normalized to "typical" values. The primary

function of the STP is to highlight areas where "favorable" ingredients co-exist. Does an STP of 4 mean that associated tornadoes will be twice as "strong" as those associated with an STP of 2? No! The authors seem to be taking the stance that the only value in the index is a strict mathematical interpretation of the numerical value. Larger values of STP do imply a greater probability of significant tornadoes, yet the more important function of the parameter is to highlight areas where multiple, independent ingredients "overlap".

Obviously, we have failed to convince the reviewer about the problems associated with this manner of constructing composite parameters. We agree that STP does indeed show where the constituent variables overlap, and it might indeed be possible to show that larger values of STP are more favorable for supercell tornadoes. If the probability of some event can indeed be shown to increase as some diagnostic variable changes, it seems logical to ask about the nature of that relationship. For example, is it linear or nonlinear? Is it a conserved variable? And so on.

Going back to that 4D volume in phase space for a world in which the STP constituent parameters would be capable of producing perfect forecasts of supercell tornadoes, why should the normalized product of the four variables provide the ideal forecast parameter? This would be true only if the 4D volume had a very simple shape. Surely, the 4D volume in our hypothetical world would be rather complex – it might be possible to transform the constituent variables in some way that would transform the convoluted volume into a simple shape but it seems unlikely that the existing transformation of the constituent variables embodied in the STP can be shown to accomplish such a feat. Therefore, the STP formulation as a product of the axes in the phase space, normalized by arbitrary constants, seems to be, at best, only a marginal attempt to understand the real complexity associated with the shape of that 4D volume. And in reality, of course, it is quite unlikely that such a perfect forecast scheme actually exists.

We believe better approaches, hopefully based on physical arguments, are required. At the very least, STP interpretation and use of it in forecasting hangs on such issues, which is why we're concerned about combinations of variables in which the manner in which they are constructed (i.e., arbitrarily multiplying four variables together) has no physical rationale.

6. Beginning of P. 8: Again, the majority of the indices are not designed to make precise estimates of tornado intensity, longevity, etc. Their primary function is to highlight areas with a favorable combination of known ingredients at the time of the analysis. You could just draw independent contour analyses of SRH, CAPE, LCL, bulk "shear", and then simply overlay all of the individual analyses to form a composite. Essentially, that is what the SCP and STP do for the forecaster. The forecaster must then put forth some effort to explain the combination of parameters, or run the risk of forecast failure when other factors are of greater importance.

See previous responses to similar comments. This is logically equivalent to reducing a sounding to a single variable, say the CAPE. I think a key element of our concern for the use of diagnostic variables in forecasting is precisely what the reviewer is describing. We simply can't endorse this simplification, which is why this paper was written in the first place.

7. Table 1: SCP – the third term should be the complete BRN denominator, normalized to 40 $m^2 s^{-2}$. STP – the $SHR_{0\text{-}6\ km}$ term should denominator is 20 m s^{-1}.

The one in the table is identical to the Eqn (4) in the original reference (Thompson et al. 2003), not as subsequently modified in the current operational incarnation.

8. 2nd paragraph of subsection a, P. 9: In consecutive sentences the authors write that CAPE and SRH "**need not be collocated** in space to a significant extent", and that "many forecasters believe CAPE and SRH need simply to be in proximity to each other, with **some overlap**." These two sentences contradict one another. In the case of EHI, the index will highlight the areas where the "overlap" in CAPE and SRH occurs. If there is *no* overlap, then the index value is zero. I have witnessed severe storm events with relatively small overlap in these parameters, but that is hardly

representative of many important severe weather episodes. I suggest the authors either delete this discussion, or substantially rewrite the entire paragraph.

There's no contradiction. Overlap does not mean collocation. To be collocated, maxima/minima would lie on top of one another, as well as gradients in the field variables.

9. 2nd full paragraph, 2nd column of P. 9: Has anyone ever claimed that tornadoes are twice as likely when EHI=2 versus when EHI=1? It is worth noting that the numerical values of these "arbitrary" indices are not to be taken literally, but that does not mean the numbers are meaningless. An EHI=2 suggests greater overlap in CAPE/SRH than an EHI=1.

See previous comments re this topic.

10. P. 10, first rebuttal, 2nd paragraph: My problem goes back to the authors' motivation for this work – that "some forecasters might indeed be so inclined, perhaps because the use of some diagnostic variable (such as CAPE) is so widespread." This statement is hand waving at best. The authors, again, provide no documentation of an actual problem, they simply speculate that an over-reliance on indices is possible with an (apparently) limited number of forecasters. Even if this "problem" proved to be true, will this paper solve anything? The folks who place excess emphasis on indices are the same ones with little physical understanding of the atmosphere. If you take away their indices, what will they do next? I seriously doubt that most forecasters will experience an epiphany and seek physical truth – they will likely turn to another crutch! Perhaps operational meteorology would have a stronger basis in science if more forecasters were held accountable as *meteorologists*, but that concern extends well beyond this work.

We don't think we have ever advocated taking away anything. It is incumbent on the developers of such indices to produce a proper verification. On that point, we should all agree. Rather, we are arguing that practitioners are vulnerable to misuse of the vast array of indices and parameters. If we concede that forecasters misusing diagnostic variables is acceptable, we are simply conceding that forecasting might best be done totally objectively and eliminate humans altogether. We hope our revised manuscript clarifies our intent.

11. P. 10, second rebuttal: Just how does the authors' response vary from the claims of the "critic"? The authors state "there is nothing wrong with a quick look, unless the forecaster limits his or her diagnosis to that." If the authors believe that a comprehensive analysis of the data is *always* possible in an operational environment, then it is clear that the authors have not spent much time working under standard operational time constraints. Various indices and parameters can help augment analyses of the raw observations, and they do often focus attention on areas where greater focus is necessary. For some reason, the authors seem to believe that too many forecasters look at parameter "bullseyes" and little else.

Both authors have spent time in their careers training forecasters from throughout the NWS. As the reviewer knows very well, the first author has indeed worked for the NWS under standard operational time constraints, and it was those very experiences that have led him to the conclusions we are advocating in our paper. Indeed, we think we have some basis for believing precisely that "too many forecasters look at parameter 'bullseyes' and little else." If the reviewer disagrees, that's a personal choice but does not preclude that we have experiences that lead us to our position.

12. P. 10, third rebuttal: While I agree that physical reasoning is a preferable basis for any parameter, the authors continue to focus too narrowly on their perception of index/parameter use in forecasting, and on the use of diagnostic parameters in a forecast mode. Many of the parameters such as EHI, STP, etc., serve to identify areas where important ingredients are co-located. It should be relatively obvious that a threat does not exist until the ingredients co-exist, thus many parameters do not highlight threat areas prior to a short-term threat. These are diagnostic variables, and as such, I do not expect a "signal" 12 h prior to an event! As mentioned several times in this review, the authors fail to document any cases where forecasters utilize *current* diagnostic parameter val-

ues to erroneously forecast an event in the future. The composite parameters serve as short-term aids in identifying threat areas, while actual forecasts rely more on the evolution of important ingredients. Like it or not, the operational forecast environment is littered with distractions. These distractions, combined with the enormous complexity of severe storm environments, leads to a "marriage of convenience" with various indices. It is not a perfect situation, but the authors have a relatively unrealistic view of forecaster routines/capabilities, as well as forecaster time constraints.

And, as mentioned several times in this response, the reviewer seems determined to force us to provide documentation for personal observations and to play "Monday Morning Quarterback". We feel there is no way to validate these statements of personal observations to the satisfaction of the reviewer. In contrast, the reviewer has no evidence that such techniques are not being employed, other than his personal experience, yet he similarly cannot provide documentation of this. This is not a valid basis for disallowing us to express our interpretation of our personal observations.

13. End of conclusions, P. 12: Have the developers of these parameters/indices ever explicitly claimed to be seeking a "magic bullet"? It should be obvious to any meteorologist that the atmosphere cannot be described by a single value of anything, but that does not mean that a parameter values cannot add to a meteorologist's interpretation of the atmosphere! Many of the problems facing indices are the result of a limited understanding of physical processes. An ingredients based approach is the best we can do at the current time, but our list of ingredients is necessarily incomplete. In other words, there is no unique and fool-proof way to forecast the weather, regardless of personal preferences.

Is it necessary for us to get parameter developers to admit openly to such a goal? The evidence is there for the reader of such articles, whether or not the developers even realize what they are doing. In the same way that developers of MOS are ultimately seeking an objective system capable of replacing human forecasters, whether they admit to, or even realize, that is their goal – many of the developers of unphysical variables and parameters are seeking a magic bullet. Just ask yourself: what would they consider the best they could do? A variable that made perfect forecasts of the phenomenon of interest! The demise of human forecasters focused on that phenomenon would inevitably follow as a result. Why would anyone have to do more than calculate the "magic bullet" parameter?

REVIEWER B (Erik N. Rasmussen):

Initial Review:

Recommendation: Accept with minor revision

I thank the authors for the time and effort they have put into this paper, and hope my comments are useful to some small degree.

In the broadest sense, the authors have written a much-needed criticism of the "magic numbers" approach to forecasting. I think this article will be an appropriate contribution to EJSSM after a little more effort to clarify the arguments and discussion. My comments below will be short on specifics and fairly broad ranging. My hope is that the authors will see something in my comments that triggers some insight into appropriate revisions.

First, except for minor comments later in this review, I am generally pleased with section 3 (Issues affecting the suitability…) and section 4 (Indices). These tend to be a bit pedantic, but I view EJSSM as a journal for the broadest possible audience of severe storms forecasters and researchers. So I feel it is seldom inappropriate to explain basic concepts. Even aging researchers like myself need a refresher now and then, and the brilliant among us can skip the more pedantic discussions.

The fundamental issue seems to be this: what gives a variable value in the severe storms forecast process?

There is no doubt in my mind that some variables have operational value because they distill the complexities of the atmosphere and allow forecasters to focus on the important problems. That must be the reason that many were invented. The authors have provided important advice throughout the paper regarding the potential pitfalls of the distillation process. But even a variable like EHI has some distillation value: when large, it tells us that either/both CAPE and SRH are large. And there are physical reasons why both CAPE and SRH are pertinent in severe storms forecasting.

With time, the complexities of the atmosphere have become more apparent, to the point that one can legitimately ask whether the problem is too complex for human comprehension, and therefore best left to numerical observation and integration leading to explicit forecasts of severe weather. To the extent that we want humans to remain responsible for forecasts, we must continue to employ and improve methods to objectively cull the trivial information, distill the state, and bring focus to our mental processing of information.

The authors have posited that a forecast parameter should have a lagged correlation with the occurrence of severe weather phenomena, and utility at forecasting non-occurrence as well. In one sense, this must be correct: if the correlation were zero, one might as well diagnose a variable that is completely unrelated to severe weather (perhaps the density of meadowlarks in song). And indeed the "lagged" part is important, for there are other tools more suitable for the actual detection of severe weather. And I must agree with the authors that there are tools available for determining the forecast utility with a sufficient amount of rigor, and these should be implemented before the variable is proposed for operational use.

We have incorporated the notion of lagged correlation between forecast parameters and the observed weather in the revised manuscript.

My main discomfort with this paper has to do with the discussion of prognostic variables, integration, etc. Formally in meteorology, a prognostic variable is a variable that can be expressed in a form suitable for numerical integration. I.e., a local tendency can be expressed as contributions from advection and from forcing. The latter is often thought of as sources, sinks, etc. Aside from the fact that I do not think we should be redefining "prognostic variable" as "one that has forecast value", the foregoing possibly has further relevance to this whole discussion. A much improved forecast parameter would be one that was conserved following the motion, except for sources and sinks for which there would be legitimate expressions available. This, at least, would give us a local tendency upon which we could base extrapolation (i.e., is the field increasing or decreasing?). Of course, we can always do temporal differences of adequately observed fields to accomplish this goal.

So the new definition of "prognostic variable" supplied by the authors is bound to be confusing. In fact, the confusion began for me in the first paragraph when it was stated that forecast parameters were not necessarily prognostic variables. This seemed very obvious, because I was not aware of predictive equations for any of the forecast parameters. And I was further under the impression that forecast parameters were exactly that: parameters that are used in the process of making a forecast.

The word choice was something of a struggle, as sensed by the reviewer. We have changed "prognostic variable" to "forecast parameter" – although it's not obvious that this is more than using a thesaurus to avoid a particularly bothersome term. We hope this clears up the concern of the reviewer.

My confusion deepened in Section 2 where it appeared that the authors were trying to establish that there ought to be a basis for numerical integration and prediction of forecast parameters. And I must admit that after reading the summary sentences at the top of the second column of page 2 that I cannot understand what the foregoing material was meant to convey. In the middle of the same column, the authors state that a prognostic variable is one that allows a forecaster to make an accurate weather forecast based on the current value of that variable. That definition is fraught with problems, and my opinion is that it is inadequate for this journal. First, there are forecasters who cannot make accurate forecasts no matter the variable being used. I.e., the definition is dependent on forecaster skill. Second, "accurate" begs definition.

But now that I give this some more thought (and erase the diatribe I just wrote), I think the authors are on to a fairly novel and useful approach. It seems that a forecast parameter could be defined as one that has demonstrable statistical/climatological correlation with the weather that occurs during the relevant time window being forecast. I suppose that is rather obvious, but it is helpful that the authors said it and to proposed

the tests that they did. So my suggestion here is to perhaps find a different term than "prognostic variable" and to skip the discussion at the beginning of Section 2 about NWP. It is not especially relevant, and has some unnecessary confusion potential. In fact, the authors are not talking about integration, but about correlation. These two things put my mind into a completely wrong frame for interpreting the paper. (The authors might want to emphasize the fact that the correlation could possibly be very mysterious, and that given a choice between equally useful variables that have a basis in physical understanding and those with a basis in magic, we might prefer those based in physics.)

In summary, my recommendation would be as follows. All the variables the authors discuss are diagnostic variables… they are quantities that can be diagnosed, via observation or calculation, for current data. These variables become "forecast parameters" if they have demonstrable utility in predictions of future events. Otherwise, they are only "diagnostic parameters".

See the preceding response.

Reviewer C (James H. Henderson):

Initial Review:

Recommendation: Accept with minor revision

After reviewing the paper, I recommend that the paper is acceptable with minor revisions and no further review is requested unless major changes are made in accordance with other reviews (at the discretion of the editor).

Reasons for recommendation:

1. I found the scientific content to be acceptable and had only a couple of comments relating to that content.

a. I am not sure paragraph 2 on page 1 makes the point for the authors. In a well written paper by one of the authors of this reviewed paper entitled On Convective Indices and Sounding Classifications, the refers to a paper written by Tudurí and Ramis (1999) where they discuss the use of indices.....in geographical locations outside their original development. Since CAPE was developed as central U.S. convective tool, it's relevance for use as a forecasting tool in California can be called into question. I would like to have seen some brief discussion in either this paragraph or elsewhere of the geographical misuse of indices. I realize that this was not the main theme of the paper, but I believe it has relevance.

This is a good suggestion and has been incorporated in the revised manuscript.

b. I would like to have seen the section on issue affecting the suitability of diagnostic variable to be broken down into separate parts relating to soundings and the volatility argument. Soundings as have been pointed out by one of the authors should be the backbone of severe weather forecasting and I felt that discussion should stand alone with the section on volatility enhancing the argument. However, having stated the above, the paper does make an excellent case for sounding diagnosis as it stands.

We're reluctant to go off on sounding-based parameters too specifically. We don't disagree with the sentiment that soundings are an important part of severe weather forecasting, but not convinced we want to have a separate section on sounding volatility issues.

[Minor comments omitted…]

Finally, I think this is a very relevant paper, to bring some scientific underpinning to the forecast of deep moist convection and associated severe weather. Having said that, I believe that the history of the development of diagnostic indices has followed the usual rules of the nature of scientific inquiry. The developers collected observations (severe weather event) and developed a hypothesis that stated when this index reached a certain value, some type of severe weather event could occur. I think the authors of this paper have very succinctly shown that the hypothesis lacked a rigorous testing and therefore should be discarded in favor of a more scientific approach to the problem.