EVALUATION OF THUNDERSTORM PREDICTORS IN FINLAND FROM ECMWF REANALYSES AND LIGHTNING LOCATION DATA

Peter Ukkonen

25. toukokuuta 2015

Ohjaajat: Antti Mäkelä, Marja Bister
Tarkastajat: Marja Bister, Heikki Järvinen
While numerical weather forecasts have improved dramatically in recent decades, forecasting severe weather events remains a great challenge due to models being unable to resolve convection explicitly. Forecasters commonly utilize large-scale convective parameters derived from atmospheric soundings to assess whether the atmosphere has the potential to develop convective storms. These parameters are able to describe the environments in which thunderstorms occur but relate to actual thunderstorm events only probabilistically.

Roine (2001) used atmospheric soundings and thunderstorm observations to assess which from a variety of stability indices were most successful in predicting thunderstorms in Finland, and found that Surface Lifted Index, CAPE and the Showalter index were most skillful based on the data set in question. This study aims to extend the assessment of thunderstorm predictors to atmospheric reanalyses, by utilising model pseudo-soundings. Reanalyses such as ERA-Interim use sophisticated data assimilation schemes to reconstruct past atmospheric conditions from historical observational data. In addition to a large sample size, this approach enables examining the use of other large-scale model parameters, which are hypothesized to be associated with convective initiation, as supplemental forecast parameters.

Using lightning location data and ERA-Interim reanalysis fields for Finnish summers between 2002 and 2013, it is found that the Lifted Index (LI) based on the most unstable parcel in the lowest 300 hPa has the highest forecast skill among traditional stability indices. By combining this index with the dew point depression at 700 hPa and low-level vertical shear, its performance can be further slightly increased. Moreover, vertically integrated mass flux convergence between the surface and 500 hPa calculated from the ERA-I convergence seems to have high association with thunderstorm occurrence when used as a supplementary parameter.

Finally, artificial neural networks (ANN) were developed for predicting thunderstorm occurrence, and their forecast skill compared to that of stability indices. The best ANN found, utilizing 11 parameters as input, clearly outperformed the best stability indices in a skill score test; achieving a True Skill Score of 0.69 compared to 0.61 with the most unstable Lifted Index. The results suggest that ANNs, due to their inherent nonlinearity, represent a promising tool for forecasting of deep, moist convection.
1 Introduction

Past studies on severe weather and convective parameters have often focused on severe convective storms, where parameters such as CAPE (convective available potential energy) and vertical wind shear have been shown to be able to characterize severe convective events such as tornadoes and large hail. In Finland, very high-CAPE environments which give birth to severe storms such as found in the United States seldom occur. For forecasters, it is useful to know which indices have the best skill in predicting the occurrence of thunderstorms in the specific region they operate in.

Globally, thunderstorms and lightning clearly favour continental tropical regions (Figure 1.1) and certain extratropical regions such as the Himalayas and Florida. Further away from the tropics, thunderstorms have seasonal occurrence, while in the tropics they occur practically all year around. Maritime areas are not favourable for thunderstorms due to the weaker nature of convection, which is the driving force in the electrification processes in the cloud. In Finland, annual lightning rates are clearly low in global terms. However, this is partly due to the seasonality, and also in Finland, thunderstorms are a major cause of weather-related damage.

Assessments of the forecast skill of convective parameters not only help forecasters, but such studies can also gain insight into the physical aspects of thunderstorm environments. The objective of this work is therefore to assess which thermodynamic parameters are efficient for predicting summer thunderstorms in Finland. This is carried out by a variety of methods: i) calculation of commonly used skill scores to assess the relative forecast skill of stability indices in a dichotomous (yes/no) forecasting scheme, ii) calculation of thundery case probability as a function of one or two parameters, and iii) training of artificial neural networks (ANN) for forecasting thunderstorms using various thunderstorm parameters. Additionally, mean soundings for thundery cases are compared with non-thundery soundings to determine the typical environments which result in the initiation of thunderstorms in Finland.

These methods are applied on lightning location data and ECMWF ERA-Interim reanalyses. Using atmospheric reanalyses to calculate stability indices has two major advantages: it allows for a large dataset of high spatial and temporal resolution compared to radiosonde data, and it enables examining the use of supplemental forecast parameters which cannot be calculated from sounding data. These factors are very important for the neural network experiment in particular, which is designed
to find a good set of ANN inputs from a large pool of thermodynamic, but also non-thermodynamic, parameters.

The structure of this work is as follows. Chapter 2 provides an overview of the theory of thunderstorm development, namely moist thermodynamics and relevant stability concepts, but also including a summary of processes related to convective initiation. Chapter 3 concerns the use of stability indices and other parameters in forecasting thunderstorms. In Chapter 4, the data and methodology is described. Chapter 5 is a summary of three previously made stability index studies in Europe, to provide some context for the results of this study, which are presented in Chapter 6. The results and how they might be affected by the methodology are discussed further in Chapter 7. Finally, conclusions are described in Chapter 8.

Figure 1.1 The global average annual flash rate. The unit is flashes km$^{-2}$ yr$^{-1}$. Courtesy: NASA
2 Theory of thunderstorm development

A thunderstorm is a cloud that produces thunder, i.e. is electrified enough to produce lightning (MacGorman and Rust, 1998). This cloud is known as Cumulonimbus. Thunderstorms are convective clouds which are formed by moist air rising upwards and condensing into liquid water, resulting in a release of latent heat which feeds the updraft further. A defining characteristic of the thunderstorm cloud is its vertical dimension; being of the same order of magnitude as its horizontal dimension, thunderstorm clouds are much taller than other types of clouds. The vertical motions associated with thunderstorms are very strong; often exceeding 10 m s\(^{-1}\) for ordinary thunderstorms, while updraft speeds of up to 50 m s\(^{-1}\) have been observed for severe thunderstorms (Williams, 1995).

Thunderstorms are the manifestation of deep, moist convection in the atmosphere. Emanuel (1994) defines convection in the context of atmospheric sciences as a class of relatively small-scale, thermally direct circulations which result from the action of gravity upon an unstable vertical distribution of mass. Even this restricted definition, which excludes things such as Hadley circulations and sea breezes, encompasses a great variety of atmospheric phenomena of varying scales. The interactions between larger and smaller scales, along with the critical influence of phase changes of water, make convection a very challenging subject in the atmospheric sciences.

In this chapter, a brief review of the theoretical aspects of thunderstorm development is presented, with an emphasis on basics of moist thermodynamics and instability, and omitting complex aspects of cumulus convection such as theory of mixing. The difficult issue of convective initiation is also discussed (Section 2.3).

2.1 Moist thermodynamics

2.1.1 Unsaturated, adiabatic process

For a parcel of dry air undergoing an adiabatic process there will be no heat exchange with the surroundings, and the first law of thermodynamics can be written in the form

\[ c_p D \ln T - R D \ln p = 0 \]  

(2.1)

This may be integrated to obtain the potential temperature
\[ \theta = T \left( \frac{p_0}{p} \right) \frac{R_d}{c_{pd}} \]  

(2.2)

where \( R_d \) is the gas constant for dry air and \( c_{pd} \) the heat capacity of dry air in constant pressure.

Thermodynamics of dry and moist (but unsaturated) air differ in that effective heat capacities are influenced by the presence of water vapour (Emanuel, 1994). Approximate formulae for specific heats of the moist air at constant volume and constant pressure may be deduced from the first law of thermodynamics and are, respectively,

\[ c'_{v} = c_{vd} (1 + 0.94r) \]
\[ c'_{p} = c_{pd} (1 + 0.85r) \]  

(2.3)

where \( c_{vd} \) and \( c_{pd} \) are the corresponding values for dry air, and \( r \) is the mixing ratio of water vapor.

Using the effective gas constant

\[ R' = R_d \frac{1 + r/\epsilon}{1 + r} \]  

(2.4)

and the effective heat capacity, equation 2.2 becomes

\[ \theta = T \left( \frac{p_0}{p} \right) \frac{\kappa'}{c_{pd}} \approx T \left( \frac{p_0}{p} \right) \kappa', \]  

(2.5)

where \( \kappa' = \kappa (1 - 0.24r) \) and \( \kappa = \frac{R_d}{c_{pd}} \). The variation in \( \kappa \) is less than 1% and usually ignored (Smith, 1997).

Replacing temperature by the virtual temperature which takes into account the effect of water vapour on density, we also define a virtual potential temperature

\[ \theta_v = T_v \left( \frac{p_0}{p} \right) \frac{\kappa'}{c_{pd}}. \]  

(2.6)

When a parcel of unsaturated air is lifted adiabatically, it expands and cools, conserving its \( \theta_v \) to a very good approximation. The rate at which its temperature decreases with height is called the dry adiabatic lapse rate, \( \Gamma_d \). Assuming hydrostatic equilibrium (a balance between the gravity force and vertical pressure gradient force) for moist air, it can be shown that (Emanuel, 1994).
\[
\Gamma_d = -\left(\frac{dT}{dz}\right)_{dq=0} = \frac{g}{c_{pd}} \left[ 1 + \frac{r}{1 + r \left( \frac{c_{pd}}{c_{pd}} \right)} \right].
\] (2.7)

Here, "dry" means that there is no condensation occurring. Often, the dry adiabatic lapse rate is presented assuming \( r = 0 \), whereby \( \Gamma_d = \frac{g}{c_{pd}} \approx 9.8 \, ^\circ C/km \). The dry adiabatic lapse rate is therefore approximately constant throughout the lower atmosphere.

### 2.1.2 Saturated, pseudoadiabatic process

As the parcel of air is lifted adiabatically, its mixing ratio \( r \) is conserved during the ascent, but the saturated mixing ratio \( r^* = r^*(p,e^*) \) decreases. Equivalently, the temperature of the adiabatically rising parcel decreases faster than its dew-point temperature. Thus, the parcel will eventually become saturated. The level at which this occurs is called the \textit{lifting condensation level} (LCL). Owing to the release of latent heat from condensation, the rate at which the temperature of the parcel falls beyond the LCL is less than \( \Gamma_d \). Assuming that all condensed liquid water precipitates out of the air immediately (making the process irreversible), it can be shown that the lapse rate accounting for latent heat release is

\[
\Gamma_s = -\left(\frac{dT}{dz}\right) = \Gamma_d \left[ 1 + \frac{\frac{L_v r}{R_d T}}{1 + \frac{L_v^2 (1+r/\epsilon)}{R_e T^2 (c_{pd} + r c_{pd})}} \right].
\] (2.8)

Because the liquid water removes a small amount of heat from the system, this process is not entirely adiabatic and \( \Gamma_s \) is called \textit{pseudoadiabatic lapse rate}. If the treatment is not simplified in this way, a similar formula to 2.8 is obtained but which includes an additional term accounting for the heat capacity of liquid water. The \textit{moist adiabatic lapse rate} (MALR) or \textit{saturated adiabatic lapse rate} (SALR) usually refers to this reversibly defined lapse rate but differs from the pseudoadiabatic lapse rate by less than 1% (Emanuel, 1994). According to Holton (1992), \( \Gamma_s \) ranges from about 4\(^\circ\)C/km in warm humid air masses in the lower troposphere to 6-7\(^\circ\)C/km in the mid-troposphere.
2.2 Static stability of the moist atmosphere

2.2.1 Buoyancy

Archimedes discovered that an isolated body of density $\rho_1$ that is immersed in a fluid of density $\rho_2$ will experience a force equal to difference between the weight of the body and that of the fluid it displaces. Expressing this force per unit mass of the immersed body, we obtain the buoyancy acceleration $B$. We can equivalently consider a hypothetical isolated parcel of air that is displaced vertically, its buoyancy given by

$$B = g \frac{\rho - \rho_p}{\rho_p}$$

where $\rho$ is the density of the ambient environment and $\rho_p$ the density of the air parcel. The buoyancy acceleration represents the action of gravity on density anomalies. Thus, the vertically displaced air parcel is positively buoyant if it has a smaller $\rho$ than the environment.

Considering an ideal gas and neglecting the contribution of pressure perturbations, the buoyancy acceleration may be written in terms of temperature alone

$$B = g \frac{T_p - T_a}{T_a}$$

This is a good approximation in the atmosphere, where maximum velocity variations are generally substantially subsonic (Emanuel, 1994).

For moist air, we can take into account the effect of moisture by utilizing the virtual temperature definition,

$$B = g \frac{T_{vp} - T_{va}}{T_{va}} = g \frac{\theta_{vp} - \theta_{va}}{\theta_{va}}$$

2.2.2 The parcel method

The stability of the atmosphere to convection is often analyzed using parcel theory. This method evaluates the buoyancy of a parcel displaced a finite vertical distance under a reversible or pseudoadiabatic process (Emanuel, 1994). The static stability is assumed to only depend on the buoyancy, so that the vertical momentum equation is written

$$\frac{dw}{dt} = B = g \frac{T_{vp} - T_{va}}{T_{va}}.$$
In this formulation, we have neglected pressure perturbation gradient forces, viscosity, and the Coriolis force.

The parcel method is useful since it can be used to assess conditional instability, whereby a displacement is stable if the parcel remains unsaturated, but ultimately becomes unstable if saturation occurs. A parcel of air that is pseudoadiabatically lifted from the boundary layer will eventually saturate at the LCL. From there on, it will ascend along the moist adiabat on a thermodynamic diagram, i.e. following the pseudoadiabatic lapse rate. The parcel will remain negatively buoyant (requiring forced lift) unless, at some point, it reaches a level where it has a higher virtual temperature than its environment. This is called the level of free convection (LFC). Beyond the LFC, the parcel will accelerate vertically due to positive buoyancy until it reaches its level of neutral buoyancy (LNB), also known as the equilibrium level (EL). The cloud top of the thunderstorm will be at the EL, often close to the tropopause, except where parcels overshoot their EL due to momentum from strong updrafts in what is known as an overshooting top. A prominent and long-lasting overshooting top is a sign that the thunderstorm is severe (NOAA, 2011).

We can vertically integrate the buoyancy (equation 2.11) of the parcel lifted from level $i$ to its equilibrium level to obtain the total amount of energy available for convection, called the convective available potential energy (CAPE)

$$\text{CAPE}_i = g \int_{p_i}^{p_{EL}} \frac{(T_v - T_v')}{T_v'} dp$$

where $p_i$ is the pressure at the initial parcel level and $p_{EL}$ the pressure at its EL. Using the hypsometric equation, this can be written

$$\text{CAPE}_i = R_d \int_{p_i}^{p_{EL}} (T'_v - T_v) d(ln p) \quad (2.13)$$

In a thermodynamic diagram whose coordinates are linear in temperature and log $p$, CAPE is proportional to the area enclosed by the temperature curves of the parcel and of the environment (Figure 2.1). We can divide this area into positive and negative areas on the sounding (PA, NA), where positive refers to the positively buoyant portion between the LFC and the EL, and negative to the negatively buoyant portion between the initial parcel level and the LFC (Emanuel, 1994):
\[ NA_i = -R_d \int_{P_i}^{PLFC} (T'_v - T_v) \, d(\ln p) \]  
\[ PA_i = R_d \int_{PLFC}^{PEL} (T'_v - T_v) \, d(\ln p) \]  

(2.14)

The negative area can be regarded as a potential barrier to convection, preventing it from occurring spontaneously (Emanuel, 1994). For this reason it is commonly referred to as *convective inhibition* (CIN). In the absence of CIN, any positive CAPE would be released the moment it came into being, so that it is the existence of CIN that allows CAPE to accumulate and be released explosively at a later time.

Although Emanuel (1994) defines CAPE as the difference of the negative and positive areas \( (CAPE_i = PA_i - NA_i) \), it’s commonly used to mean the positive area, especially in a forecasting context. Therefore

\[ CIN_i = NA_i = -R_d \int_{PLFC}^{P_i} (T'_v - T_v) \, d(\ln p) \]  
\[ CAPE_i = PA_i = R_d \int_{PEL}^{PLFC} (T'_v - T_v) \, d(\ln p). \]  

(2.15)
Figure 2.1 Pseudoadiabatic parcel ascent illustrated on a skew T-logp diagram. We use a surface-based parcel starting from 1000 hPa. The parcel, when lifted (solid red line), will follow the dry adiabat until it eventually reaches its lifted condensation level (LCL, dashed blue) at approximately 920 hPa. Thereafter, it ascends pseudoadiabatically. If it reaches its level of free convection (LFC, solid blue), it will experience positive buoyancy until it reaches its equilibrium level (where the parcel temperature once again equals the environmental temperature, at approximately 200 hPa). The buoyant energy available is called the CAPE or positive area, since it is proportional to the area between the parcel curve and the sounding from the LFC to the EL. The potential energy needed to lift the parcel to its LFC is called convective inhibition (CIN), or negative area. This is an example sounding from the data set which was associated with thunderstorms.

CAPE depends on the initial parcel and which thermodynamic process is used in displacing the parcel (Emanuel, 1994). Basing calculations on pseudoadiabatic ascent leads to significantly larger buoyancy and CAPE compared to reversible ascent (Smith, 1997).

2.2.2.1 Limitations

Although parcel theory is useful, it is a very simplified treatment of buoyancy. It views ascending parcels as a perturbation in a homogenous base state (the environment) and is therefore not an accurate representation of reality, where rising plumes of air physically interact with their (heterogenous) environment. Markowski and
Richardson (2010, hereafter MR10) list several factors not taken into account in simple parcel theory, which alter the effective buoyancy (p. 44-47):

1. **Perturbation pressure gradient forces.** In derivation of parcel theory, pressure perturbations are neglected twice: in the approximation of the buoyancy force (2.10), and in the vertical velocity momentum equation (2.12). The latter omission of the perturbation pressure gradient force is more problematic (Doswell and Markowski, 2004), and according to MR10, the force is not negligible in general. A bubble of warm (cold) air that is associated with an upward (downward) -directed buoyancy force tends to be associated with a downward (upward) -directed perturbation pressure gradient force. The latter force is only small compared to the buoyancy force as long as the temperature anomaly is relatively narrow, since for a wide warm or cold air bubble, more air needs to be pushed out of the way. Therefore the vertical perturbation pressure gradient force tends to partially offset the positive buoyancy acceleration.

2. **Entrainment.** The inflow of environmental air into the cloud is known as *entrainment* and the outflow cloudy air is known as *detrainment* (de Rooy et al., 2013). These mixing processes typically reduce the buoyancy of rising air parcels. MR10 explain that entrainment depends e.g. on updraft tilt, so that the larger the tilt of an updraft, and the smaller its width, the larger the entrainment rate will be. Such a tilt can be caused by a strong vertical wind shear. Because some degree of mixing always occurs in the real atmosphere, environmental conditions above the initial parcel level do matter, despite not being considered in simple parcel theory and therefore in parameters such as CAPE. Due to mixing, dry air above the boundary layer can inhibit convective initiation, especially if CAPE values are low.

3. **Hydrometeors.** For pseudoadiabatic ascent, it is assumed that hydrometeors instantly fall out of a rising, saturated parcel, so that the buoyancy is not affected of the weight of the condensates. In reversible ascent, condensates remain in the parcel, which reduces buoyancy, however the condensates also carry heat, which increases the buoyancy. MR10 assert that the net effect is generally a net reduction of buoyancy in the lower troposphere and a net increase in the upper troposphere. Pseudoadiabatic and reversible ascent both
represent idealized extremes, so that in general a buoyancy somewhere in be-
tween is realized. CAPE calculations are generally based on pseudoadiabatic
ascent using the integrated temperature or virtual temperature excess.

4. **Freezing.** Freezing of water droplets is an additional source of positive buoy-
ancy not considered in the pseudoadiabatic treatment. However, the latent
heat of freezing is only a small fraction of that of condensation, so that MR10
considers this a relatively minor limitation. However, Williams and Renno
(1993) showed that the inclusion of the ice phase has a large effect on CAPE
calculations in the tropical atmosphere.

5. **Compensating subsidence.** Another example of physical interaction be-
tween the environment and the parcel, compensating subsidence within the
surrounding air can affect the buoyancy and/or the perturbation pressure field.

The limitations of parcel theory have considerable implications. Emanuel (1994)
points out that the parcel method only enables to determine an upper bound on the
potential energy available for convection. The actual amount of potential energy
available when considering the system as a whole can be substantially less than
 parcel CAPE. The dependence of CAPE on the choice of the parcel is obviously a
limitation in itself. For daytime soundings, CAPE generally decreases with height
of the lifted parcel. Many authors and forecasters therefore base calculations on
some average environmental profile near the surface, i.e. in the lowest 500m. This
accounts for entrainment to some extent, but it assumes that the entrainment rate
is constant in the layer (MR10, p193).

### 2.2.3 Different types of instability associated with convection

Having defined pseudoadiabatic parcel ascent and the concept of buoyancy, we can
now discuss instability. Although there exists a large variety of different kinds of
instabilities in the atmosphere, here we consider static instability (also called hy-
drostatic, vertical instability) which gives rise to atmospheric convection. This is a
condition in which air will rise freely on its own due to positive buoyancy. There
are three commonly used instability concepts associated with static stability: conditional instability, latent instability and potential instability. Because their definitions
have often been confused with one another in literature, it is worthwhile to clarify
them in an attempt to instil proper usage of terminology. This was already done
comprehensively by Schultz et al. (2000).
2.2.3.1 Conditional instability

The concept of conditional instability is tied to the parcel method. The commonly used definition of conditional instability considers the environmental lapse rate: if it lies between the dry- and moist-adiabatic lapse rate, columns of air are said to be conditionally unstable, i.e. the instability is conditional to the saturation of the air parcel. The criteria for absolute instability is then that the environmental lapse rate is steeper than the dry-adiabatic lapse rate, and if the environmental lapse rate is less than the moist-adiabatic lapse rate it absolutely stable.

A conditionally unstable atmospheric layer will usually be far from saturated, but the instability can be released by a rising parcel, originating e.g. from near the surface. Upon saturation, the temperature of the parcel will decrease less rapidly with height compared to the (conditionally unstable) environment, so that it will eventually gain positive buoyancy.

Schultz et al. (2000) point out that the lapse-rate definition of conditional instability does not fit the definition of an instability and is really a statement about uncertainty about instability. A foremost problem with the lapse-rate definition is that it is entirely possible to have a situation where the parcel stability differs from lapse-rate stability, so that an ascending parcel can be unstable (have positive buoyancy) with respect to a layer that is absolutely stable. For most purposes, the authors therefore recommend the available-energy definition of conditional instability (also known as latent instability). However, the authors also acknowledge that the lapse-rate definition of conditional instability can be useful in some contexts, e.g. when using an ingredients-based methodology for forecasting deep moist convection or DMC (Sect. 2.3.3).

2.2.3.2 Latent instability

According to Schultz et al., the correct use of the term latent instability involves the available-energy definition, so that a conditionally unstable atmosphere with positive CAPE is viewed as having latent instability.

2.2.3.3 Potential instability

Potential instability considers the lapse rate of the equivalent potential temperature $\theta_e$, i.e. the potential temperature an air parcel would have if all its water vapour were
condensed and the resulting latent heat were used to warm the parcel (Emanuel, 1994). Three possible states of potential instability are:

\[ \frac{\partial \theta_e}{\partial dz} < 0 \quad \text{potentially unstable} \]

\[ \frac{\partial \theta_e}{\partial dz} = 0 \quad \text{potentially neutral} \]

\[ \frac{\partial \theta_e}{\partial dz} > 0 \quad \text{potentially stable} \]

Let us consider a potentially unstable atmospheric layer. If such a layer is lifted, the bottom of the layer will eventually become saturated before the top of the layer. Upon further lift, it will then cool at the moist adiabatic lapse rate, while the top of the layer cools at the dry adiabatic lapse rate. This implies destabilization, since the top is cooling at a faster rate. Regardless of initial stratification, sufficient lifting of the layer can therefore cause the environmental lapse rate to become absolutely unstable.

Potential instability has been cited as an important mechanism for DMC and particularly useful in situations when convection can develop over broad regions as a result of large-scale lifting of deep layers (Trier, 2003). However, MR10 (p. 188) seem less optimistic about its actual relevance to DMC. The authors reason that although potential instability is often present in environments where DMC develops, it usually does not play a role in the destabilization that is associated with DMC. If such were the case, the emergence of widespread stratiform clouds should precede the eruption of thunderstorm clouds, but this is not usually observed.

### 2.3 Convective initiation

Forecasting convective weather is one of the most difficult tasks in weather forecasting. This difficulty is largely due to the fact that the mere presence of CAPE is an insufficient condition for the initiation of DMC. Fundamentally, convective initiation (CI) requires that air parcels reach their LFC and remain positively buoyant over a significant upward excursion. The former typically requires some forced ascent owing to the presence of at least some convective inhibition (CIN). In addition to the presence of a mechanism to provide such a forcing, destabilization leading to a reduction of CIN (and increase in CAPE) often also plays a critical part in CI.

Convective initiation is fundamentally a mesoscale process. However, because the mesoscale is by definition "between scales" i.e. smaller than the synoptic scale
but larger than the microscale, it is linked to the large-scale environment. The large-scale setting can often prime the mesoscale for CI e.g. by way of large-scale mean ascent which tends to reduce CIN and deepen the low-level moist layer (MR10, p. 183). Mean subsidence has the opposite effect. Specific large-scale environments can also be associated with various small-scale phenomena linked to CI; for example small-scale horizontal convective rolls are associated with boundary layer wind shear and have been identified as CI triggers (Weckwerth and Parsons, 2006). However, lifting mechanisms are generally distinctly mesoscale processes.

Below, we consider destabilization and lifting mechanisms separately to provide an overview of some of the major processes involved. However, these are clearly linked, since any process that forces the ascent of air is also associated with some degree of destabilization.

### 2.3.1 Thermodynamic destabilization processes

CAPE and CIN are both sensitive to 1) lower-tropospheric moisture and 2) lapse rate of temperature (the temperature of the parcel can be regarded as being included in the latter). Bluestein and Jain (1985) calculated that for a typical squall-line sounding in the US, an increase in the mixing ratio of 1 gkg$^{-1}$ or increase in temperature of 1 °C of the parcel results in an increase in CAPE of 500-600 m$^2$s$^{-2}$. For CIN, the analysis is more complicated, since the sensitivity of CIN to temperature and moisture depends on the sounding. In the typical case of a well-mixed PBL, CIN is generally more sensitive to temperature than moisture (Smith, 1997). Often, however, dramatic changes in CIN and CAPE are a result of sharp increases in low-level moisture, and not lapse rate (MR10, p. 188).

Forecasting the evolution of thermodynamic stability is challenging, since there can be multiple simultaneous processes at play that influence both lapse rates and moisture, sometimes in opposite directions (Trier 2003). Here we omit writing the tendency equations for lapse rate and moisture, but consider the physical processes involved.

**Diabatic heating.** Differential diabatic heating that decreases with height will increase the lapse rate. During the day, the air near the surface will warm, which results in a decrease in low-level stability.

**PBL turbulence.** Associated with the diurnal heating are turbulent eddies, which tend to mix the air within the PBL. In a well-mixed daytime PBL, water vapour is transported upward, and $q$ will typically decrease within the PBL. Such
daytime drying and warming of the PBL typically reduces CIN, but can also reduce CAPE for sharp decreases of $r$ above the PBL (Trier, 2003).

**Horizontal advections.** Cold (warm) air advection which increases (decreases) with height will lead to destabilization. In environments where this occurs, such as within the warm sector of baroclinic waves, regions of low-level warm advection maxima often experience large-scale ascent, which also contributes to destabilization (Trier, 2003).

**Vertical motions.** Mean ascent is always associated with adiabatic cooling. Large-scale lifting of lower tropospheric layers will increase CAPE and reduce CIN of the PBL by cooling the layers above it, and also by reducing the stability of pre-existing stable layers through stretching, since large-scale ascent generally increases with height in the lower troposphere (Trier 2003). Although synoptic-scale vertical motions are small in magnitude, *persistent* large-scale ascent is effective in reducing CIN (MR10, p. 188).

### 2.3.2 Lifting mechanisms

The accumulation of CAPE is associated with the presence of CIN, and rarely does destabilization lead to CIN being entirely absent. Therefore, in order for parcels to reach their LFC so that convection can be initiated, there must exist a process to provide a forcing to overcome the potential energy barrier. As mentioned, the rising motions associated with synoptic-scale processes are usually too slow to lift a parcel to its LFC in the required time. Therefore, lifting mechanisms associated with DMC are generally subsynoptic. Convection is known to often initiate along air-mass *boundaries*, characterized by large gradients of density, i.e. temperature and/or moisture. Boundaries are usually easily identifiable by operational observing systems, but forecasting is complicated by mesoscale processes, which result in convection only developing along small segments of boundaries and not along their entire length (MR10, p. 189).

Some major lifting mechanisms associated with initiation of DMC include:

**Forced mesoscale ascent.** Isentropic mesoscale ascent refers to air traveling along an upward-sloping isentropic surface, which can directly initiate convection by saturating lower-tropospheric air, or reduce CIN. Isentropic lift may be orographically forced or associated with the overrunning of statically stable air masses (Trier 2003).
**Solenoidal circulations.** Solenoidal circulations are thermally direct atmospheric circulations that are forced by horizontal temperature gradients. They often result from differential surface heating. Two such circulations are land/sea breezes and mountain/valley circulations. Both may initiate convection at the ascending branch of the circulation. The outflow boundaries of pre-existing storms, also known as gust fronts, also belong to this class. The leading edge of evaporatively cooled thunderstorm outflow creates a marked temperature contrast and can trigger further convection owing to intense vertical motions. However, this mechanism seems to depend on the environmental wind profile, so that low-level shear is required to balance the shear induced by an outflow boundary; in such a case, long-lived mature convective systems can be maintained (MC10, p. 254-257; Weckwerth and Parsons, 2006). Advancing cold fronts or other regions of pronounced frontogenesis may behave similarly to gust fronts and provide forced ascent and deepening of the moist layer to initiate DMC. It is worth noting that although cold fronts are synoptic-scale disturbances, their cross-front dimensionality is mesoscale.

**Horizontal convective rolls.** Horizontal convective rolls (HCRs) are a convective PBL-based circulation manifested as counterrotating vortices with a horizontally oriented axis (Weckwerth and Parsons, 2006). HCRs, also known as cloud streets, can extend hundreds of kilometers as rows of cumulus clouds aligned parallel to the mean boundary layer wind. HCRs are associated with surface-layer heat flux and low-level vertical wind shear. Regions where HCR updrafts intersect convergence zones of larger mesoscale boundaries are optimal locations for convective development (Weckwerth and Parsons, 2006). The interaction of HCRs with sea breeze fronts especially is well-established to be associated with CI.

**Drylines.** Drylines are air mass boundaries characterized by large horizontal moisture gradients (MR10, p.132-139). They separate air with abundant water vapour, typically originating from large bodies of warm water, and dry, continental air (e.g. from higher elevations). They are common in southern plains of the United States in the spring and summer, when moist air from the gulf of Mexico in the east meets dry desert air from the west. Dryline strength is highly correlated with large-scale confluence associated with synoptic-scale processes (MR10). Drylines are well-known for frequently initiating convective storms, and according to MR10, they are probably associated with most of the major tornado outbreaks in the central US. The mechanisms for this are not yet exactly understood, but both large- and small scale variability along drylines is believed to play a role in CI (Rye and Duda,
Large-scale effects include enhanced convergence from a localized turning of surface wind, associated with a pressure gradient force caused e.g. by differential heating. Small-scale effects include the presence of horizontal convective rolls and misocyclones (see below).

There are also other phenomena associated with convective initiation, such as small-scale vertical vortices known as misocyclones, and mesoscale gravity waves, which may provide a CI trigger but conversely can also be initiated themselves by convective storms (Coleman and Knupp, 2006). In general, areas of kinematic or thermodynamic inhomogeneities along mesoscale boundaries are favourable for CI. Boundaries are associated with low-level convergence, which acts to deepen the moist layer and provide lift. Hence, MR10 state that the most effective strategy for forecasting convective initiation may be looking for regions with positive CAPE, low CIN, and persistent low-level convergence.

2.3.3 Ingredients-based approach

An ingredients-based approach has been used to forecast deep, moist convection (e.g Johns and Doswell, 1992). It identifies three necessary ingredients that must be in place for DMC to be initiated: i) conditional instability, ii) a moist layer of sufficient depth at lower levels and iii) a source of lift. Some authors cite the first condition as conditional, latent and/or potential instability but the authors themselves refer to conditional instability (e.g. Doswell et al., 1996) or steep lapse rates (Doswell et al., 1992) which are synonymous when using the lapse-rate definition of conditional instability. Looking at the ingredients, it’s clear that the existence of an LFC - and therefore CAPE - is associated with the first two, while the third condition refers to a mechanism to force a parcel through a negatively buoyant layer to its LFC.

In ingredients-based forecasting of DMC, forecasters can track the evolution of the ingredients independently, and look for the time and space where all three come together. The benefit of this approach is a focus on the physical processes.

Lock and Houston (2014) suggested a slight modification to this methodology by considering four principal factors, in pairs of two: i) Buoyancy and dilution, ii) lift and inhibition. This approach is intuitive, since the factors are undoubtedly linked: "lift and inhibition are paired since the amount of inhibition is what determines if a given amount of lift is sufficient to initiate a thunderstorm", as explained by the authors. Meanwhile, dilution (caused by entrainment/detrainment) is related to the "moist layer of sufficient depth" in the three ingredients of Johns and Doswell.
Given an environment that is more prone to dilution, stronger positive buoyancy is needed for DMC to initiate.
3 Using indices and parameters to forecast thunderstorms

3.1 Stability indices

*Stability indices* (other terms commonly in use include *instability index, convective index, thunderstorm index*) have been a cornerstone in the forecasting of convection for many decades (Doswell, 1996). They usually quantify two of the three preconditions required for the initiation of deep, moist convection, namely (conditional) instability and low-level moisture. Omitting the third factor (a lifting mechanism), these parameters reflect the potential for thunderstorm development based on the large-scale thermodynamic environment, but cannot be used to pinpoint the time and place where a convective storm might be initiated. The task of a convective index can thus be seen as being able to effectively sample the thermodynamic instability associated with thunderstorms.

3.1.1 Uses and limitations of stability indices

In spite of their prevalence in the forecasting of convection, the simplicity of these parameters is both an appeal and a pitfall, and forecasters and researchers generally acknowledge that any single variable considered in isolation has limited forecast value, even disregarding the fact that stability indices do not account for lift. A comprehensive discussion about the uses and limitations of indices in forecasting severe storms, which can largely be extended to the forecasting of thunderstorms in general, can be found by Doswell and Schultz (2006). The authors emphasize that no single parameter is likely to ever represent a perfect forecasting tool and that inspecting the full atmospheric sounding is always more valuable to forecasters. This is because indices are one-dimensional representations of atmospheric instability, which depends on the vertical structure of temperature and moisture. No single index can fully capture all the important details of the lapse rate structure; indeed, most indices are tied to two specific pressure levels and therefore only measure a very specific part of the atmospheric stability (stability of a given layer, respective to a given parcel).

Nonetheless, indices are generally considered to have some degree of skill in forecasting convection and characterizing convective environments. As such, they can at least be helpful to quickly identify regions of elevated probability for thunderstorm
occurrence. It is once again stressed that indices are most useful when considered in combination and not just relying on a single index. Therefore, the aim of a study such as this is not to find the "ideal" one-size-fits-all thunderstorm index for Finland, because none such exists. For example, an index that was designed to predict frontal thunderstorms shouldn’t be used to predict air-mass thunderstorms, and it may be unlikely that any single index will have the highest skill in predicting both. However, comparing the overall forecast skill of indices is especially useful to learn which of them capture the typical convective environments in a given region, and to determine which of similar indices are the most skillful. For example, there are many indices which measure surface-based instability, and it is important to properly validate which of them has most forecasting power based on a large and reliable data set.

3.1.2 The use of diagnostic variables as forecast parameters

Doswell and Schultz (2006) discussed what it means for a diagnostic variable (such as a stability index) to be used as a forecast parameter. The latter was defined as a parameter that allows a forecaster to make an accurate weather forecast based on the current values of the variable, implying a time-lagged correlation between the parameter and the event. The authors contrasted this to using a model-forecasted diagnostic variable valid at a future time to make a forecast of the weather for that future time. This is a valid distinction to make, and a given stability index might be more suited to one of these uses than the other. For example, an index which is based on conditions at the surface will naturally be sensitive to solar heating and therefore benefit greatly from using the forecasted daytime temperature, since the time-lagged correlation to the event (thunderstorm) will be very low for longer time periods; i.e. calculating such an index from a night-time sounding to forecast afternoon thunderstorms the next day will undoubtedly have bad results.

However, with the proliferation of more sophisticated NWP models, it bears asking if the use of stability indices as forecast parameters as in the example above has much relevance anymore. If a NWP model is able to forecast a convective parameter reasonably well, then it is arguably of little value to ask which index shows the best lagged correlation to a convective event, if another index (that can be accurately forecasted using the NWP model) shows an even higher non-lagged correlation to the event. If this is the case, then it is better to ask which parameters
are best at characterizing the preconvective environment immediately prior to the onset of convection. The forecasting of the variables is then left to models.

### 3.1.3 Commonly used indices

All of the commonly used stability indices are either directly or indirectly based on parcel theory. Most indices simply evaluate the temperature difference between the environment at a given level (usually 500 hPa) and an air parcel that is pseudoadiabatically lifted to this level from the boundary layer.

The formulae and description for a few classic stability indices are presented below to provide some examples. The remaining of the 28 thunderstorm indices featured in this study are described in Appendix A.

**Showalter index (SI)**

Likely the first stability index that ever published, and one that is still used today, is the Showalter index (Showalter, 1953). SI is defined as the difference between the temperature at 500 hPa and the temperature of an air parcel that is lifted pseudoadiabatically to 500 hPa from 850 hPa,

$$SI = T_{500} - T'_{850hPa\rightarrow500hPa}$$

(3.1)

According to Showalter, values of +3 °C or less are "quite likely to produce thunderstorms".

**Lifted index (LI)**

$$LI = T_{500} - T'_{i\rightarrow500hPa}$$

(3.2)

The Lifted Index is very similar to SI. In fact, it was originally developed by Galway (1956) to improve upon SI for the prediction of latent instability during afternoon hours by using the forecast maximum temperature near the surface. Since then, LI has largely been used as an observed parameter instead of using the forecasted maximum temperature, based on various parcels $i$:

- SLI (Surface-based Lifted Index) or $LI_{sfc}$: using a parcel with the temperature, dew-point temperature and pressure at 2 m above the surface.

- $LI_{50}$, using a parcel with the mean temperature, dew-point temperature and pressure of the layer in the lowest 50 hPa.
– LI_{100}, using a parcel with the mean temperature, dew-point temperature and pressure of the layer in the lowest 100 hPa.

– LI_{\text{mu}}, using the parcel with the highest $\theta_e$ in the the lowest 300 hPa.

In this study, we use the above described parcels for both LI and CAPE. Hence, CAPE_{sfc}, also known as SBCAPE (Surface Based CAPE) is CAPE calculated using a surface-based parcel, while CAPE_{\text{mu}} is the same as MUCAPE (Most Unstable CAPE). CAPE_{50} and CAPE_{100} are often referred to as mixed layer or mean layer CAPE (MLCAPE).

Jefferson Index (JI and JI_{mod})

$$JI = 1.6 \cdot \theta_{w900} - T_{500} - 11$$ (3.3)

The Jefferson Index was designed to improve upon another index, the Rackliff index ($RI = \theta_{w900} - T_{500}$), so as to depend less on the temperature of the air mass (Jefferson, 1963a). $\theta_w$, the wet bulb potential temperature, is conserved for reversible adiabatic processes (American Meteorological Society, 2015). Higher values of JI indicate higher possibility for thunderstorms.

$$JI_{mod} = 1.6 \cdot \theta_{w900} - T_{500} - 0.5 (T_{700} - T_d700) - 8$$ (3.4)

Jefferson developed the index further since it was found that the index over-forecast thunderstorms when the 900-500 hPa layer was dry (Peppier, 1988). $JI_{mod}$ takes therefore into account the humidity at the 700 hPa level (dew point depression $T - T_d$ is inversely proportional to relative humidity).

3.2 Parameters related to convective initiation

3.2.1 Moisture and mass flux convergence

A diagnostic measure that has been widely used in short-term forecasting of convective initiation is horizontal moisture flux convergence (MFC). This is a term in the conservation of water vapour equation and has the expression

$$MFC = -\nabla \cdot (r \nabla h) = -\nabla_h \cdot \nabla r - r \nabla \cdot \nabla_h,$$  (3.5)
where \( r \) is the mixing ratio and \( \mathbf{V}_h \) is the horizontal wind vector. The first term is the advection term which represents the advection of humidity, and the second the convergence term

\[
-r \nabla \cdot \mathbf{V}_h = -r \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right),
\]

(3.6)

where \( u \) and \( v \) are the \( x \)- and \( y \)-components of the wind velocity, respectively. On the scale of fronts, the convergence term is an order of magnitude larger than the advection term (Banacos and Schultz, 2005).

MFC can be integrated in the vertical to obtain VIMFC (vertically integrated MFC, where the advection term is dropped),

\[
VIMFC = -\frac{1}{g} \int_{p_1}^{p_0} r \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) dp.
\]

(3.7)

The role of moisture convergence for convective initiation is explained by MR10 (p. 195) as follows. Horizontal MFC is only one term in the expression for local tendency of water vapour, which is also affected by vertical MFC and sources/sinks (evaporation/condensation). Horizontal and vertical MFC always oppose each other, and in such a way that the convergence term in equation 3.5 is exactly cancelled out, so that only moisture advection remains. In the absence of evaporation, water vapour mixing ratio can only increase locally due to advection, but advection cannot generate local extrema. Therefore, moisture convergence cannot produce local moisture maxima, a mistake made by some authors who interpret moisture pooling as an explicit result of horizontal MFC. The fact that locally large mixing ratios at the surface are often found within convergence zones is instead explained by the deepening of boundary layer moisture. Increased moisture depth means that vertical mixing does not decrease surface moisture as effectively, which results in locally elevated surface moisture concentration.

Although moisture convergence is therefore related to CI, the underlying cause is the upward motion associated with horizontal mass convergence, which in turn is well correlated to moisture convergence. Banacos and Schultz (2005) examined the use of MFC for forecasting DMC through case studies and found that horizontal mass convergence is at least as effective as surface MFC in identifying boundaries. In the absence of a robust theory that justifies the use of MFC in forecasting deep, moist convection, the way this parameter combines moisture and convergence infor-
mation can be considered rather arbitrary and inconsistent with ingredients-based methodology.

Although the use of horizontal mass convergence is scientifically better justified as it is an indirect measure of lift, it suffers from many of the same problems as MFC, including scale-dependency (observational networks are generally unable to resolve convergence at scales most relevant for CI) and the lack of a simple one-way relationship between horizontal mass convergence and cumulus convection (Banacos and Schultz, 2005). To elucidate on the latter the authors established some conceptual models of convective initiation as it relates to convergence (Figure 3.1).

![Figure 3.1](image)

**Figure 3.1** Possible relationships between subcloud horizontal mass convergence and cumulus convection. Adopted from Banacos and Schultz (2004). a) Surface convergence maximum is associated with DMC, b) the convergence maximum is associated with shallow convection due to a capping inversion and/or midlevel subsidence, c) the convergence maximum is located near change in boundary layer depth, d) the convergence maximum is rooted above the PBL.

Past studies investigating the use of moisture or mass flux convergence for predicting CI have almost exclusively looked at convergence at or near the surface; probably largely explained by availability of data. Recently, the use of vertically integrated moisture flux convergence (VIMFC) as a thunderstorm predictor was examined by van Zomeren and van Delden (2007) using six-hourly ECMWF weather analysis data. The authors found that VIMFC (integrated between 1000 hPa and
700 hPa) alone did not perform well as a dichotomous thunderstorm predictor. However, when combined with a stability index to calculate thunderstorm probability as a function of these two parameters, very-high thundery case probabilities (around 90%) were reached for environments with a high positive VIMFC and large instability.

3.2.2 Other parameters

Lock and Houston (2014) investigated the ability of various thermodynamic parameters to distinguish between initiation and non-initiation of convective storms. The parameters were calculated from 20-km grid data and intended to represent buoyancy, dilution, lift and inhibition. Parameters featured in Lock and Houston (2014) that are also calculated in this work are:

**ACBL lapse rate.** Lapse rate of the active cloud-bearing layer (ACBL, the atmospheric layer above the LFC, defined here as a 100 hPa -deep layer); a measure for buoyancy and dilution. Numerical experiments using an idealized cloud-resolving model have suggested this parameter may be important for CI (Houston and Niyogi, 2007). Dilution can be considered to increase the height of the LFC of the parcel by reducing parcel moisture and thereby promoting evaporation and cooling the parcel. How much the LFC is increased depends on the environmental lapse rate of the ACBL, since for a smaller ACBL lapse rate, the environmental temperature above the LFC will be warmer and thus the diluted parcel must be lifted higher before it can become positively buoyant (Houston and Niyogi, 2007).

**VWS<sub>ACBL</sub>.** Vertical wind shear (VWS) in the active cloud-bearing layer. This can have an inhibiting effect on convection by promoting entrainment (Ch. 2.2).

**VWS<sub>subLFC</sub>.** Vertical wind shear in the subcloud layer (below LFC). Low-level shear is hypothesized to be associated with CI by promoting updrafts along gust fronts and other boundaries due to its association with horizontal convective rolls (Ch. 2.2).

**Δz<sub>LFC</sub>.** LFC height minus initial parcel height. Related to the depth of lift needed to initiate convection. This was among the parameters with highest discriminatory power in Lock and Houston (2014).

**CAPE<sub>LCL</sub>.** LCL to LCL +2km CAPE. Sum of the CAPE and CIN in a 2 km -deep layer based at the LCL. Buoyancy metric for the layer associated with the critical early stages of convection.
3.3 Using Artificial Neural Networks to forecast thunderstorms

Having presented a large number of parameters which should at least in theory have some value for forecasting thunderstorms, another interesting problem presents itself: can we use multiple parameters to construct a forecasting tool that with superior skill compared to any single stability index? This is a *multivariate analysis* problem, since we are interested in forecasting thunderstorm occurrence from more than one predictor variable. A multivariate analysis tool which is particularly powerful for modelling complex, nonlinear relationships is the artificial neural network (ANN). ANN’s are a type of machine learning algorithm and thus, a branch of artificial intelligence. ANN’s owe their name to being (very crudely) modelled after biological neural networks, such as the brain. Neural networks can be used for both regression and classification problems, of which the prediction of thunderstorm occurrence falls into the latter category.

Although ANN’s have seen some utilization in the atmospheric sciences for a few decades now (Manzato, 2005; Gardner and Dorling, 1998), overall they have gathered quite little attention (Maqsood et al., 2004), considering that their intrinsic nonlinearity suggests feasibility for forecasting small-scale weather events. Previously, Manzato (2005, 2007) developed ANNs for short-term forecasting of thunderstorms with good results, with successful operational implementation by a regional meteorological office in Italy. Given the opportunity arising in this work to develop a neural network for forecasting thunderstorms from ERA-reanalysis parameters as ANN inputs, by utilizing the Matlab Neural Network Toolbox, we thus decide to carry out a small ANN experiment similarly to Manzato (2005). A brief description of the ANN is provided below, while the experiment setup is described in the next chapter.

An introduction to the multilayer perceptron (MLP), the type of ANN used here, can be found in Gardner and Dorling (1998), who reviewed its use in the atmospheric sciences. The MLP is essentially a system comprising of multiple levels of logistic regression models (Bishop, 2006). It connects or maps inputs to outputs by one or more layer of interconnected computation nodes (the neurons). This is illustrated in Figure 3.2. In the MLP, also known as a feed-forward neural network, information moves forward from one layer to the next. The nodes are connected to each other by weights, and the sum of the inputs to each node is modified by a
non-linear function known as an activation or transfer function, to form the output signal (Gardner and Dorling, 1998). This function is usually the logistic sigmoid or the hyperbolic tangent. Although the input signal only moves one way - forward - in the feed-forward network, the actual training (in the supervised learning scheme, i.e. training by labelled examples) is performed by means of a back-propagation algorithm. Here, the final output, after traversing through all the hidden layers (i.e. the one or more layers between the input and output layers, see Fig. 3.2), is compared with the correct answer to compute an error function. The error is then fed back through the network and the algorithm adjusts the weights to reduce the value of the error function by a small amount. This is done by using the procedure known as gradient descent, in which the local gradient of the error surface is calculated and weights adjusted in the direction of the steepest local gradient. If there are \( k \) number of inputs, the gradient descent takes place in \( k \)-dimensional space. To try to avoid ending up in a local minimum of the error function, various tricks can be used.

During the computationally intensive training process, the error should eventually converge to a minimum number. However, it is very important to ensure the ANN does not suffer from overfitting, whereupon the model could be made complex enough to fit the training data extremely well, but at the loss of generalization, i.e. the ability to make accurate predictions for new unseen data. Although overfitting becomes less of a problem the more data is available for training, it is very hard to know just how much data is needed to make the risk of overfitting insignificant. Therefore, despite the large data set (1.6 million pseudo-soundings) we utilize a standard technique for ensuring good generalization, called early-stopping (Section 4.4.4).
Figure 3.2  A multilayer perceptron with two hidden layers (from Gardner and Dorling, 1998).
4 Data and methods

4.1 Datasets

4.1.1 FMI’s lightning location network

Observations of thunderstorms in Finland have been collected since 1887 by the Finnish Meteorological Institute and its predecessors. FMI first obtained an automatic ground lightning location system in 1984. Since then, the observational network has undergone a number of improvements. Major changes have included upgrading the ground lightning location system in 1997 to a new system (IMPACT) and the installation of a cloud lightning system in 2001 called SAFIR, which however seized operations in 2011. Importantly, lightning location at FMI has been done in co-operation with Nordic countries for many years. Co-operation began with Norway in 2001, then included Sweden (since 2002), Estonia (2005) and Lithuania (2014). This comprises the NORDLIS (Nordic Lightning Information system) observational network, in which the sensors in every member country are utilized for dramatically improved performance and coverage (Figure 4.1). The total number of NORDLIS sensors is now more than 30 (Mäkelä 2013).

The system locates individual lightning strokes. A flash can comprise of several strokes, the number of which is expressed by multiplicity. Since the lightning flash is a more widely used and arguably more suitable meteorological and climatic quantity than a stroke, multiplicity is ignored in this work.

Relevant aspects of the lightning location system in this work are the detection efficiency (not all flashes are located with the system) and location error. Because the sensors are not uniformly distributed, these depend on the region. For the purposes of this study, the quality of the lightning location data can be considered homogenous over Finland in the period 2002-2013. Therefore, the variations of the detection efficiency and location error are neglected.
4.1.2 The ECMWF ERA-Interim reanalysis

ERA-Interim is the most recent global atmospheric reanalysis produced by European Centre for Medium-Range Weather Forecasts (ECMWF), replacing ERA-40 (Dee et al., 2011). It represents a major upgrade over ERA-40, using 4-dimensional variational analysis (4D-Var) for data assimilation. ERA-Interim covers the period from 1979 onwards, as it is continuously updated in near real-time. The spatial resolution is 0.75 degrees at 37 atmospheric levels.

All indices and parameters are calculated from ERA-Interim fields of geopotential height, temperature, specific humidity, two-dimensional wind components and divergence, given at a level spacing of 25 hPa across the lower and 50 hPa across most of the upper troposphere. Additionally, surface parameters of temperature, dew point temperature, mean sea level pressure and 10 m wind components are utilized. The reanalysis provides a best guess of the state of the atmosphere at 6-h intervals, which can be considered as pseudo-soundings (referred to in this work often imply as soundings) for the calculation of all indices and parameters used in this study.
The data used in this work covers the main thunderstorm season months (May - August) between 2002 and 2013. Although lightning location data spans further back than 2002, this year was chosen because the lightning location network went through major changes in 2001 and the quality of data is much improved from 2002 onwards. The data set can be considered a large one at any rate, with the total number of flashes nearing 3 million. Using a time span of 12 years is also sufficient to capture a variety thunderstorm environments in Finland and thus be considered climatologically reasonably representative. The reason that this is important is that year-to-year variation for example in the number of flashes observed in a summer is typically very large, and particular summers can be characterized by flow regimes and thunderstorms of a particular type.

4.2 Definition of a thunderstorm event

The reanalysis provides atmospheric model soundings, and the objective is to compare soundings associated with thunderstorms to those that are not. How a thunderstorm (or "thundery", from hereafter) event and null (non-thundery) event is defined based on the lightning location data is somewhat ambiguous. The second issue is associating a sounding with a convective event. Here, the spatial and temporal resolution of the sounding data is important. Soundings associated with convective events that are used in convective research are often referred to as proximity soundings (Brooks et al., 2003). Proximity soundings are often used to study the large-scale environments associated with severe weather. Fundamentally, the sounding needs to be representative of the atmospheric conditions that the thunderstorm (or particular convective event of interest) developed in. Previous authors have used different criteria for a proximity sounding, being largely dependent on the data. In this study, both the horizontal and temporal resolution of the reanalysis which provides model soundings are deemed small enough to utilize all lightning observations. Associating a thunderstorm event with the nearest model sounding in space-time is an easy starting point and used by e.g. Brooks (2009), but raises the issue of whether this is appropriate when assessing the skill of parameters as thunderstorm predictors, since convection inevitably leads to stabilization. Associating thundery events (based on lightning data) to nearest model soundings in space-time can at worst mean that the sounding sampled conditions 3 hours after the event (lightning flash), which would not be representative of preconvective conditions. However, the
The alternative way of relating soundings to events while utilizing all observations is associate the events with the last sounding available before the event. This way, events fall 0-6 hours after soundings. In 6 hours, conditions can change significantly by diurnal heating, frontal passages, etc. Because of this, it is difficult to say which approach is more appropriate \textit{a priori}. Thus, both criteria are used in this study.

This leaves the question of how a thunderstorm event is defined based on the lightning flash data. The amount of false detections is deemed low enough to use a lightning threshold of 1 or more flashes to designate a thunderstorm event. Remaining cases comprise the very large null data set.

The criteria used to define thundery and non-thundery soundings thus become

1. Thundery if at least 1 lightning flash was located in the grid area within 3 hours of the model sounding, otherwise non-thundery (THUN1).

2. Thundery if at least 1 lightning flash was located in the grid area within the 6 hours \textit{following} the model sounding, otherwise non-thundery (ThUN2).

The data consists mainly of cloud-to-ground (CG) flashes. Approximately 23% of signals in the data set were interpreted as intracloud (IC) flashes. Because of the physical similarities of thunderclouds leading to IC and CG flashes, these signals were chosen not to be filtered out in spite of cloud-to-ground lightning be the phenomena of interest for society at large. Additionally, the ability of the lightning location system to distinguish between C-G and IC is limited, so discarding IC flashes would erroneously lead to some CG events being ignored.

Roine (2001) argued that thunderstorms should be separated into frontal and air-mass thunderstorms to increase the representativeness of various indices. Since it is impossible to assess thunderstorm type on the basis of lightning location data and these kind of classifications are not readily available, no such classification is done here. However, since pseudo-soundings are available at 00, 12, 18 and 00 UTC, we can easily analyze the results separately for daytime and nocturnal cases.
The total number of pseudo-soundings in the dataset is just over 1.6 million, of which 6.1% were classified as thundery. Of all thundery cases, 72% occurred between 06 and 18 UTC, or between 9 a.m. and 9 p.m. local time. Therefore, just over a quarter of thunderstorm events were nocturnal.

4.3 Calculation of parameters

This work was carried out using MATLAB. All parameters which are based on a pseudo-adiabatically lifted parcel were calculated using equations 2.2 and 2.8 where ascent was simulated on a 1 hPa vertical resolution. The environmental temperatures and humidities were interpolated on the 1 hPa resolution from the pressure level and surface variables using a cubic spline interpolation.

For the calculation of CAPE parameters, the LFC was required to be above LCL, i.e. it was defined as the first level above the LCL where the parcel virtual temperature exceeded the environmental virtual temperature. LCL height and $\theta_e$ were calculated using the empirical formulas of Bolton (1980).

4.4 Data analysis

4.4.1 Forecast verification in a dichotomous forecasting scheme

The process of determining the quality of a forecast is known as forecast verification (Wilks, 2006). The traditional use of stability indices for forecasting convective weather falls into the domain of a dichotomous (two-class) forecasting scheme, where the predictand is the occurrence vs. non-occurrence of the convective weather event. The predictor is in our case a stability index, which can take a range of values. One way of assessing the forecast skill of a stability index is to make it dichotomous as well by determining a threshold value that divides its range into two parts, "thundery" and "non-thundery". This dichotomous forecasting setting is illustrated in a contingency table (4.3.1.1) while the goodness of the forecast can be estimated using skill scores (4.3.2.1).

4.4.1.1 Contingency table

In the simple forecasting setting where both the forecast (I) and predictand i.e. observation (J) can only take on two possible values (yes/no), there are $I \times J = 2 \times$
2 possible combinations of forecast and event pairs that can be displayed in a 2 x 2 contingency table (Figure 4.2).

We follow the description of the contingency table given by Wilks (2006). Of the four possible combinations of forecast and event outcome ($n = a + b + c + d$), $a$ gives the amount of times of the $n$ total cases that the event was successfully forecast, usually called hits. If the event was forecasted to occur but did not, it was a false alarm (category $b$). An event that occurred despite not having been forecast to occur is a miss (category $c$). Finally, there are $d$ instances where the event was correctly forecast not to occur (correct rejection or correct negative).

![Figure 4.2 The 2 x 2 contingency table. Adapted from Wilks (2006).](image)

A number of scalar measures exist for different attributes of the contingency table (such attributes are e.g. accuracy and bias; for further information see Wilks, 2006). These measures describe the forecast quality in some way. The dimensionality of the 2 x 2 contingency table is $(I \times J) - 1 = 3$, meaning that complete specification of forecast performance requires a minimum of 3 verification measures. Thus, any single measure is an incomplete measure of forecast "goodness", and as remarked by Wilks, there can be differing views of what constitutes a good forecast. Here we first describe some commonly used scalar verification measures of forecast performance which are not skill scores, described in the next section, to emphasize that there is a distinction (some authors have erroneously used the term skill score to describe all verification measures).
The most direct measure of forecast accuracy is the proportion correct, i.e. fraction of the \( n \) forecasts that were correctly forecast

\[
PC = \frac{a + d}{n}
\]  

(4.1)

Another measure of forecast accuracy is the threat score, also known as critical success index (CSI)

\[
CSI = \frac{a}{a + b + c}
\]

CSI has occasionally been used as a skill score but is not suited to this task as it does not take into account correct rejections.

The most widely used measure of reliability is the false alarm ratio (FAR), also known as probability of false alarm (POFA)

\[
FAR = POFA = \frac{b}{a + b}
\]

FAR is the proportion of "yes" forecasts that turned out to be wrong, i.e. the number of false alarms divided by all "yes" forecasts. A small FAR is preferred.

FAR (false alarm ratio) is sometimes confused with the false alarm rate, also known as probability of false detection (POFD), the recommended term to avoid confusion.

\[
POFD = \frac{b}{b + d}
\]

POFD is the ratio of false alarms to the total number of non-occurrences.

Finally, we have the hit rate,

\[
H = HIT = POD = \frac{a}{a + c}
\]

The hit rate is the ratio of correct "yes" forecasts to the number of times the event occurred. In other words, given that an event occurs, it gives the likelihood that it is also forecast, and therefore it is also called probability of detection (POD).

### 4.4.1.2 Skill scores

Scalar measures of "overall" forecast skill have been developed for convenience, although such measures are inherently incomplete representations of forecast performance in a higher-dimensional setting (Wilks 2006). Skill scores measure the
forecast accuracy relative to some reference model, e.g. climatology. The two perhaps most widely used skill scores for verifying categorical forecasts are the Heidke Skill Score (HSS) and True Skill Statistic (TSS).

The Heidke Skill Score,

\[
HSS = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)},
\]

is a skill measure based on the proportion correct (equation 4.1) measure of accuracy. The reference measure in HSS is the proportion correct that would be achieved by random forecasts that are statistically independent of the observations (Wilks, 2006). Perfect forecasts receive a HSS value of 1, forecasts equivalent to the reference forecast receive a zero, and forecasts worse than the reference forecasts have negative HSS values.

True Skill Statistic (TSS), also known as the Peirce skill score (after its original author) or Kuiper's performance index, is formulated as

\[
TSS = \frac{ad - bc}{(a + c)(b + d)} = POD - POFD.
\]

TSS is similar to HSS but uses the sample climatology as the reference forecast, equivalently, the hit rate in the denominator is that for random forecasts which are constrained to be unbiased (Wilks, 2006). Again, perfect forecasts receive TSS = 1, reference forecasts TSS = 0, and forecasts inferior to reference forecasts receive negative scores.

To verify the forecast skill of stability indices using skill scores, we need to choose a single measure of forecast skill that can be evaluated for the range of values of the predictor. The "optimal" threshold value of the stability index is then the value at which the skill score attains its highest value, and the relative forecast skill of indices can be assessed by the maximum skill score value attained. Which measure do we choose? Doswell et al. (1990) favoured HSS for rare-event forecasting arguing that TSS is prone to hedging, defined by the authors as "deviating from the forecaster’s true beliefs in order to increase the verification score". However, TSS has the significant benefit of not depending on event frequency or sample size (Huntrieser et al., 1997) which is important so that results can be compared to other studies. For these reasons, both TSS and HSS are used for assessment of indices. In this work, it was found that HSS approaches TSS when the event sample size approaches the null event sample size, becoming equal for \( \frac{N_{\text{event}}}{N_{\text{null}}} = 1 \).
Calculated thresholds are "optimal" for "yes/no" forecasting based on one forecast skill measure. Results are likely to differ using another skill score which emphasize a different aspect of forecast performance. However, the relative ranking of indices based on maximisation of a skill score should ideally be at least somewhat consistent across skill scores, to the degree to which they are useful measures of forecast skill.

4.4.2 Thundery Case Probability

After calculating the skill statistics of the convective parameters – testing the parameters in a dichotomous forecasting scheme – we might also be interested in how the parameters fare as non-dichotomous predictors. Although dividing index ranges into thundery and non-thundery values in a dichotomous scheme is consistent with real-world use where forecasters often have to make a yes/no prediction (issue a warning or not), such forecasts are in reality done using more than just one parameter. Moreover, the division can be considered rather arbitrary in the absence of a physical basis for dividing index values into thundery and non-thundery categories.

Thus, the “thundery case probability” (TCP) as a function of the thunderstorm parameter is calculated, following the work of Haklander and Van Delden (2003). TCP indicates how the probability for a thunderstorm to occur changes as the index increases or decreases. Here, TCP is calculated by taking the moving average of event occurrence after sorting indices in descending order. To obtain a somewhat smooth distribution, a large window size is used for the moving average (N = 3000 or larger), consistent with the sizeable data set. Features of the TCP function associated with a good predictive skill in this probabilistic scheme are a high maximum TCP, monotonicity, and a steep gradient.

It is also possible to calculate TCP as a function of more than one variable, depicting TCP in multi-dimensional space. This can be useful for examining how thunderstorm probability behaves empirically as a function of different factors associated with convective initiation. In this work, we wish to examine TCP in two-dimensional space, using a stability index and convergence to quantify stability and lift, similarly to van Zomeren and van Delden (2007). This is done by dividing the whole data set into appropriate bins in 2D space and calculating TCP for each bin where the number of cases exceeds some minimum number in order to limit uncertainty.
4.4.3 Mean temperature soundings

Following Huntrieser et al. (1996) and Roine (2001), the mean temperature and dew-point depression soundings for thundery (at least 1 flash observed) and non-thundery (no flashes observed) cases is calculated. By calculating the difference in the mean soundings, the climatological characteristics of the vertical distribution of temperature and moisture in environments associated with thunderstorms can be revealed.

4.4.4 Neural network development

Artificial Neural Networks are developed for the forecasting of thunderstorm occurrence (1/0, based on the THUN2 classification) in the following way. First, the database is randomly divided into training (3/5), validation (1/5) and testing (1/5) subsets. The training subset is used for training the network i.e. computing the gradient and updating the network weights and biases, to minimize an error function. In this work, cross-entropy error (CEE) is used, since this is considered a more appropriate measure of error than mean squared error for classification problems (Bishop, 2006). The validation subset is used for the early-stopping method of improving network generalization (the default in the Matlab Neural Network Toolbox). The early-stopping method entails monitoring the error for the validation subset during training and stopping training when the validation error begins to rise (a sign of overfitting). Additionally, the validation error is used for model selection: specifically, we utilize a modified version of the input-selection algorithm of Manzato (2005) to find a good set of input parameters for our neural network (see below). Finally, the testing subset is an independent dataset used neither for training nor model selection, and can therefore be used to evaluate the final performance in an objective manner. In our case, this is done by comparing the forecast skill of the ANN to the stability indices by means of a skill score test.

The resulting methodology for developing a good-as-possible neural network for predicting thunderstorm occurrence (given computational and other restraints) in this work is based on the following steps:

1) Gather a set of "promising" inputs. These constitute various stability indices and measures of moisture, convective inhibition and also lift, represented by vertically integrated mass and moisture flux convergences. Parameters were picked on the basis of the results from TCP and skill score analyses, physical reasoning, the work of Manzato (2005), and/or the linear correlation coefficient between the
parameter and thunderstorm occurrence. The complete list of \( N = 37 \) tested inputs is given in Appendix B.

2) Looping over all of the \( N \) parameters in the pool, train \( M = 20 \) neural networks using the \textit{scaled conjugate gradient} method (function \textit{trainscg} in Matlab) with the given parameter as single input. Each time a neural network is retrained, weights are initialized differently. Training multiple neural networks is a simple method of improving generalization and increasing the likelihood of finding the global minimum.

The networks are trained using \( H \) number of hidden neurons and one hidden layer. Because of computational restraints, \( H \) is fixed at this stage to

\[
H = H_0 = \text{Integer}\left[\left(\frac{I - 1}{2}\right) + 1\right],
\]

where \( I \) is the number of neural network inputs (\( I = 1 \) initially). Manzato found this \( H \) value to lead to the best results. No manual pre-processing of inputs was performed before network training; however, the Matlab feed-forward function automatically applies some pre-processing, such as \([-1,1]\) min-max normalization.

3) Out of the \( N \) parameters, choose the parameter that led to the smallest average validation CEE (for the \( M \) networks) to be the neural network input.

4) Find the second-best input from the remaining \( N - I \) parameters based on the smallest resulting validation CEE, and add it to the set of \( I \) network inputs.

5) Keep increasing progressively the number of inputs until the validation CEE does not drop any more by at least 0.5\% upon adding the next-best input.

Although this method should lead to a good set of ANN inputs, it does not guarantee the best combination of parameters, as it does not try all the possible combinations of inputs \( I \), yet alone all possible combinations of \( I \) and \( H \).

Having used the above algorithm of finding a suitable combination of inputs, another algorithm is then used to find the optimal number of hidden neurons. This algorithm trains networks with varying number of hidden neurons: \( H = H_0, H_0 + 1, \ldots, I+2 \), with \( H_0 \) unchanged from before (given the much larger data set compared to Manzato, it is expected that the optimal number of hidden neurons will be higher). For each \( H \), the \( M = 30 \) neural networks are trained, again using scaled gradient descent, and using the same index division every time. Among the networks with the \( H \) which led to the smallest average validation CEE across the \( M \) differently initialized networks, the network with the lowest CEE is chosen. This represents
the final neural network for predicting thunderstorm occurrence, whose performance is then assessed using the skill score test.

The aim of this neural network experiment is twofold: 1) compare the dichotomous forecast skill of the best network found to that of the best-performing stability indices alone, and 2) use the input selection algorithm to gain insight into convective initiation in Finland. It is hoped that the selection of parameters by the algorithm, from a large pool of variables, can give useful information about the most effective ways of quantifying principal factors associated with convective initiation: buoyancy, dilution, inhibition, and lift. This secondary objective, however, is extremely limited by the methodology used: the spatial and temporal resolution of the reanalysis in particular, constrains the possibilities of examining convective initiation processes in this work. It is assumed that only buoyancy and dilution are factors which can be robustly assessed in this framework, as these are associated with the larger-scale thermodynamic environment, which the reanalysis is assumed to capture well. Still, based on the strong dependence of TCP on the ERA-I convergence parameters, one or more of them could well be picked by the algorithm.
5 Previous comparative studies of stability indices in Europe

The physical laws of the atmosphere are independent of space and time. However, because stability indices are imperfect and do not capture all of the physical factors involved in thunderstorm development, their efficiency in predicting thunderstorms varies with region. An example of this is the discovery by Brooks (2009) that for a given combination of high CAPE and deep tropospheric wind shear, the probability for a significant severe storm is considerably higher for Europe than for the US (note that these conditions are met much more seldom in Europe, and thus severe convective storms are more frequent in the US). Many stability indices have even been designed originally to predict thunderstorms in a specific region or of a specific type, such as the Boyden index, that was devised to predict frontal thunderstorms in Britain. Stability indices have often been used without proper validation about their suitability for the region in question. Several authors have stressed that to be effective, a stability index must be adapted to the local climatology of convection (Dalla Fontana, 2008).

Below are summarized the results of three studies where the relative skill of various stability indices to predict thunderstorms was assessed. These studies were chosen on the basis of region and having focused on the occurrence of thunderstorms in general, and not severe convective storms or particular convective weather. In addition, the methodologies are reasonably similar to that of this work and should thus enable a meaningful comparison of results.

5.1 Huntrieser et al (1996)

Huntrieser et al. (1996) carried out an extensive study of thunderstorm indices for Switzerland. The preconvective environment on thunderstorm days was investigated by utilizing a variety of methods on thermodynamic and kinematic parameters calculated from atmospheric soundings, observations from a mesoscale network (ANETZ), radar images and insurance company reports of hail damage. Radiosoundings in Payerne at 0000 and 1200 UTC were assumed to represent preconvective conditions for thunderstorms occurring between 1240 and 2340 UTC over the whole observation area of around 250 km × 100 km in northern Switzerland. The study period encompassed the months of May-August between 1985 and 1989. Huntrieser et al.
calculated skill scores and probability distributions to compare the skill of 16 differ-
ent indices to predict isolated and widespread thunderstorms. The indices featured
in the study were K-, TT-, DCI-, KO-, BI-, HI-, SLI- and SWEAT -indices and
four variations of CAPE and the Showalter Index. In addition, mean temperature
soundings and hodographs for days with and without thunderstorms were compared
to gain insight into the characteristic thermodynamic and kinematic environments
associated with thunderstorms.

Skill scores indicated that the original Showalter Index (SI850) was the best index
for prediction of a non-thundery or thundery day based on 0000 UTC soundings.
At 1200 UTC the best index was the SWEAT index followed closely by SI850.
To distinguish between isolated and widespread thunderstorms, CAPE and DCI
calculated from afternoon soundings proved to be most effective.

Mean temperature soundings and hodographs showed thundery days to have
markedly different thermodynamic and kinematic profiles. The most significant
temperature differences were found for the layer between 1.5 and 2 km (850 hPa).
For humidity, the layer between 3 and 4 km (600-700 hPa) proved to be decisive, the
dew point depression being approximately 7 C lower at 1200 UTC on thundery days.
Hodographs also revealed notably different wind conditions, in particular thundery
days were characterized by a distinct low-level jet at 3 km.

Based on these findings, Huntrieser et al. developed a new thunderstorm index
especially designed for northern Switzerland, the SWISS index. This index combined
a traditional stability index with a humidity and wind shear term. Verification using
an independent testing dataset confirmed that the SWISS index had significantly
higher skill in predicting thunderstorms than other indices.

5.2 Roine (2001)

Roine (2001) compared the skill of stability indices calculated from both day- and
night-time soundings to predict afternoon thunderstorms in central and southern
Finland. 25 different indices were compared to SYNOP observations of thunder-
storms between 9 and 18 UTC (12 and 21 local time) and within a radius of 170
and 120 km, respectively, of the two sounding stations in central and southern Fin-
land. Additionally, lightning location data was used, and a thunderstorm event
was defined as at least one flash being observed in the study area by a SYNOP
station or located by the lightning location system. The study period comprised
the main thunderstorm season months (May, June, July, August) of 1998 and 2000. Roine separated thunderstorms into frontal and air-mass thunderstorms based on frontal analyses done by on-duty meteorologists at FMI. A thunderstorm day was designated as frontal if any front was analyzed within the study area between 9 and 18 UTC and remaining thunderstorm days were associated with air-mass thunderstorms. Frontal thunderstorms comprised 31% of thundery cases and air-mass thunderstorms the remaining 69%.

Similarly to Huntrieser et al. (1996), Roine used skill scores to compare the relative skill of the stability indices as dichotomous (i.e. yes/no) thunderstorm predictors, ranking the indices based on their True Skill Score (TSS). His main results were:

- For daytime soundings for all thundery days (both frontal and air-mass), SLI, CAPE$_{sfc}$ and Showalter Index with the parcel based at the LCL had the highest predictive skill. Linear correlation coefficients between observations and index-based forecasts were deemed surprisingly high at 0.65 - 0.69. Using 00 UTC soundings to predict thunderstorms for the following day, forecasts still had relatively good skill, the best index PI (potential wet bulb index) reaching a correlation coefficient of 0.55. This index is based on conditions at the 850 hPa (and 500 hPa) level, so that it is not very sensitive to diurnal heating.

- For predicting air-mass thunderstorms, the results were particularly good. The best indices calculated from daytime soundings were the original Showalter Index and SLI. Showalter Index (LCL) and LI using the forecasted maximum daytime temperature had the highest skill using night-time soundings. Correlation coefficients between forecasts and observations were 0.71-0.73 for daytime and 0.60-0.61 for night-time soundings.

- As expected, results were markedly different for frontal thunderstorms. Skill scores were lower, indicating that stability indices fare worse for predicting frontal thunderstorms. The best indices were PI, Jefferson and Boyden index, and SLI for 1200 UTC soundings.

- Roine (2001) also investigated the use of relative vorticity at 500 hPa as a supplemental forecast parameter to be used with stability indices, by requiring the relative vorticity to be neutral or positive in the study area for a thunderstorm to be predicted. Overall, this deteriorated the forecasts. Although the false
alarm rate decreased, the probability of detection decreased even more. This indicated that thunderstorms can develop even if the 500 hPa flow is neutral or even weakly anticyclonic.

5.3 Haklander and Van Delden (2003)

Haklander and Van Delden (2003) compared the forecast skill of 32 different thunderstorm predictors for the Netherlands. Indices were derived from six-hourly rawinsonde observations at De Bilt, Netherlands. The data set used was rather large compared to previously mentioned studies, utilizing 11 495 soundings for all months of the year between January 1993 and December 2000. Thunderstorm activity was determined using data from the UK Met Office Arrival Time Difference (ATD) lightning detection and location system. Any detected lightning activity within 100 km from De Bilt during the 6 h following a sounding served as the dichotomous predictand.

To assess the forecast skill of the 32 predictors, Haklander and Van Delden (2003) introduced the Normalized Skill Score (NSS), which combined the commonly used True Skill Score (TSS) and Heidke Skill Score (HSS). Indices were ranked according to their maximum NSS score, attained at the value near which both the TSS and HSS reached their respective maximum values. No separation of thundery cases into further categories such as winter or summer was done. The index which showed the highest skill in a thundery/non-thundery forecast scheme based on the NSS was the Lifted Index: \( \text{LI}_{100}, \text{LI}_{50} \) and \( \text{LI}_{\text{mu}} \) were the best three indices, followed by the SWISS\(_{12} \) index of Huntrieser et al. (1996). SWISS\(_{12} \), however is very similar to SLI, which itself ranked 5th. The probability of detection (POD) and false alarm rate (FAR) for \( \text{LI}_{100} \) at its maximum NSS was 0.66 and 0.59, respectively. CAPE\(_{\text{MU}} \) was ranked 10th and CAPE\(_{\text{sfc}} \) was among the poorer indices in the study at 20th place.

In addition to examining the use of indices in a dichotomous forecasting scheme, Haklander and Van Delden (2003) estimated a "thundery case probability" (TCP) as a function of each thunderstorm index. This was done by compiling an ordered list of each index, marking cases as either "thundery" or "non-thundery", and taking the moving average of thundery event occurrence using 200 cases to obtain a somewhat smooth distribution. In this forecasting scheme, SLI was among the indices with the highest maximum TCP values, reaching a TCP of 65.5% at mean SLI of -4.1 °C (standard deviation of 0.9 °C). Other parameters which reached a maximum TCP
of above 60% were CAPE$_{50}$, CAPE$_{sfc}$, SWISS$_{12}$ and the Adedokun2 index, which measures the buoyancy of an air parcel that is lifted pseudoadiabatically from the surface to 2m similarly to SLI, but is calculated using wet-bulb potential temperatures. All of these indices represent latent instability close to the surface. The authors pointed out that although CAPE did not excel as a dichotomous predictor in the study, it did reach high thundery case probabilities.
6 Results

6.1 Skill scores

6.1.1 Verification measures as a function of threshold value

Figure 6.1 illustrates how the verification measures and skill scores behave as a function of the threshold value of the predictor, which is SLI in this example. If a high (stable) SLI value is chosen for the threshold value so that a thunderstorm is forecasted whenever SLI falls below this value, the probability of detection (POD) will be high, but so will the false alarm ratio (FAR). By maximising a skill score (TSS or HSS) we try to find a suitable balance between detection and overforecasting. However, it is clear from the graph that the optimal threshold values given by their maxima differ quite significantly, so that HSS emphasizes a low FAR, while TSS emphasizes a high POD. Using the threshold value of SLI < 1.5 °C for forecasting thunderstorms as indicated by TSS, the POD and FAR are both a high 80%. Using the HSS-based threshold of SLI < -1 °C, the FAR is lower at 66%, but the probability that a thunderstorm event that occurs is also forecast (POD) is less than 50%.

![Image of Figure 6.1](image-url)

**Figure 6.1** POD, FAR, TSS, HSS and CSI as a function of the threshold value of the SLI index in a dichotomous forecasting scheme. The optimal forecasting threshold is given by the maxima of the skill score measure (TSS or HSS).
6.1.2 Verification using True Skill Statistic

The optimal thresholds for dichotomous forecasts and rankings of indices according to TSS value are presented in Table 6.1. For these results, the THUN1 thundestorm event criteria was used (at least 1 flash observed within 3 hours of the sounding). The index which has the best forecast skill according to the maximum TSS is \( L_{\text{mu}} \) with a TSS of 0.646 at the threshold \( L_{\text{mu}} < 0.9 \, ^{\circ}C \). It leads the next best index significantly, which is "lightning-CAPE" based on the most unstable parcel. This is simply the slice of total MUCAPE between the -10 \( ^{\circ}C \) and -40 \( ^{\circ}C \) levels. The fact that the maximum TSS for this parameter is quite a bit lower at 0.611 is almost entirely due to \( L_{\text{mu}} \) having a higher POD (0.88 compared to 0.84), illustrating the importance of POD on TSS. The threshold value for lightning-CAPE is very low at > 5.4 J/kg; the interpretation is that the existence of essentially any positive buoyant energy in this layer is a good predictor for lightning activity. In addition to the -10...-40 \( ^{\circ}C \) formulation for lightning-CAPE, the parameter was recalculated with using the following layers: -10...-30 \( ^{\circ}C \), and 0...-30 \( ^{\circ}C \). These formulations led to marginally lower maximum TSS compared to using the -10...-40 \( ^{\circ}C \) layer.

The third best index is SWISS\(_{12}\) with a maximum TSS of 0.609. This index uses a surface-based parcel (SLI), however, it also depends on the dew-point depression at 650 hPa and is one of the only indices which incorporates a wind shear term \( \text{VWS}_{0.3} \) (vertical wind shear between the surface and 3 km above ground).

Also performing well are the indices JI, THOM, \( L_{50} \) and MUCAPE. The Jefferson Index (JI) depends on buoyancy of a parcel at the 900 hPa level respective to 500 hPa, and additionally incorporates the dew-point depression at 700 hPa. THOM and \( L_{50} \) share the 5th place with TSS = 0.594. Since THOM is simply formulated as \( THOM = KI - L_{50} \), the indication is that incorporating the KI index has no benefit. However, when using the THUN2 thunderstorm event criteria, where the model soundings used for calculating the indices always fall (0-6 hours) prior to events, the results are generally similar but do deviate slightly when it comes to THOM and \( L_{50} \) (Table 2). Here, THOM comes in at fourth place, while \( L_{50} \) is ranked a distant 11th, with maximum TSS values of 0.569 and 0.552, respectively.

The other index whose forecast skill relative to other indices is different with the THUN2 criteria is SLI. This index is ranked 5th using THUN2 proximity criteria but 10th with THUN1. However, even for this index, the maximum TSS value is lower using the THUN2 criteria. This result holds for \textit{all} indices and is fairly significant, with the mean maximum TSS being 0.545 using the THUN1 criteria and
0.520 using the THUN2 criteria. The optimal threshold value according to TSS is identical for almost all indices irrespective of which proximity criteria is used.

<table>
<thead>
<tr>
<th>Index</th>
<th>threshold</th>
<th>TSS</th>
<th>HSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI&lt;sub&gt;inu&lt;/sub&gt;</td>
<td>&lt; 0.9</td>
<td>0.646</td>
<td>0.248</td>
<td>0.876</td>
<td>0.803</td>
<td>0.192</td>
<td>0.348</td>
<td>0.127</td>
</tr>
<tr>
<td>MUCAPE&lt;sub&gt;10...40&lt;/sub&gt;</td>
<td>&gt; 5.4</td>
<td>0.611</td>
<td>0.237</td>
<td>0.838</td>
<td>0.808</td>
<td>0.185</td>
<td>0.33</td>
<td>0.211</td>
</tr>
<tr>
<td>SWISS&lt;sub&gt;12&lt;/sub&gt;</td>
<td>&lt; 1.8</td>
<td>0.610</td>
<td>0.240</td>
<td>0.839</td>
<td>0.809</td>
<td>0.184</td>
<td>0.329</td>
<td>0.110</td>
</tr>
<tr>
<td>JI</td>
<td>&gt; 27.5</td>
<td>0.598</td>
<td>0.212</td>
<td>0.857</td>
<td>0.824</td>
<td>0.171</td>
<td>0.312</td>
<td>0.098</td>
</tr>
<tr>
<td>THOM</td>
<td>&gt; 19.9</td>
<td>0.594</td>
<td>0.204</td>
<td>0.866</td>
<td>0.83</td>
<td>0.166</td>
<td>0.307</td>
<td>0.1</td>
</tr>
<tr>
<td>LI&lt;sub&gt;50&lt;/sub&gt;</td>
<td>&lt; 2.2</td>
<td>0.594</td>
<td>0.234</td>
<td>0.816</td>
<td>0.809</td>
<td>0.183</td>
<td>0.323</td>
<td>0.105</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>&gt; 113</td>
<td>0.592</td>
<td>0.247</td>
<td>0.798</td>
<td>0.8</td>
<td>0.19</td>
<td>0.33</td>
<td>0.207</td>
</tr>
<tr>
<td>SI&lt;sub&gt;LCLmod&lt;/sub&gt;</td>
<td>&lt; 3.2</td>
<td>0.59</td>
<td>0.206</td>
<td>0.855</td>
<td>0.828</td>
<td>0.167</td>
<td>0.307</td>
<td>0.103</td>
</tr>
<tr>
<td>MUCAPE&lt;sub&gt;500&lt;/sub&gt;</td>
<td>&gt; 100</td>
<td>0.588</td>
<td>0.234</td>
<td>0.808</td>
<td>0.809</td>
<td>0.183</td>
<td>0.323</td>
<td>0.18</td>
</tr>
<tr>
<td>SLI</td>
<td>&lt; 1.5</td>
<td>0.586</td>
<td>0.238</td>
<td>0.8</td>
<td>0.806</td>
<td>0.185</td>
<td>0.323</td>
<td>0.104</td>
</tr>
<tr>
<td>SI&lt;sub&gt;LCL&lt;/sub&gt;</td>
<td>&lt; 3.0</td>
<td>0.584</td>
<td>0.211</td>
<td>0.836</td>
<td>0.824</td>
<td>0.17</td>
<td>0.307</td>
<td>0.103</td>
</tr>
<tr>
<td>SI&lt;sub&gt;550&lt;/sub&gt;</td>
<td>&lt; 3.7</td>
<td>0.58</td>
<td>0.198</td>
<td>0.855</td>
<td>0.833</td>
<td>0.162</td>
<td>0.299</td>
<td>0.102</td>
</tr>
<tr>
<td>JI&lt;sub&gt;mod&lt;/sub&gt;</td>
<td>&gt; 27.6</td>
<td>0.571</td>
<td>0.19</td>
<td>0.855</td>
<td>0.838</td>
<td>0.158</td>
<td>0.292</td>
<td>0.089</td>
</tr>
<tr>
<td>KO</td>
<td>&lt; 0.2</td>
<td>0.552</td>
<td>0.185</td>
<td>0.836</td>
<td>0.841</td>
<td>0.155</td>
<td>0.283</td>
<td>0.112</td>
</tr>
<tr>
<td>PI</td>
<td>&lt; 1.3</td>
<td>0.546</td>
<td>0.172</td>
<td>0.855</td>
<td>0.849</td>
<td>0.147</td>
<td>0.274</td>
<td>0.1</td>
</tr>
<tr>
<td>EI</td>
<td>&lt; 3.6</td>
<td>0.544</td>
<td>0.17</td>
<td>0.857</td>
<td>0.85</td>
<td>0.146</td>
<td>0.273</td>
<td>0.111</td>
</tr>
<tr>
<td>SWISS&lt;sub&gt;50&lt;/sub&gt;</td>
<td>&lt; 7.1</td>
<td>0.544</td>
<td>0.18</td>
<td>0.832</td>
<td>0.843</td>
<td>0.152</td>
<td>0.278</td>
<td>0.1</td>
</tr>
<tr>
<td>NCAPE&lt;sub&gt;50&lt;/sub&gt;</td>
<td>&gt; 0.4</td>
<td>0.542</td>
<td>0.212</td>
<td>0.771</td>
<td>0.822</td>
<td>0.169</td>
<td>0.294</td>
<td>0.197</td>
</tr>
<tr>
<td>DCI</td>
<td>&gt; 7.3</td>
<td>0.522</td>
<td>0.17</td>
<td>0.819</td>
<td>0.849</td>
<td>0.146</td>
<td>0.265</td>
<td>0.113</td>
</tr>
<tr>
<td>KO</td>
<td>&gt; 21.7</td>
<td>0.513</td>
<td>0.151</td>
<td>0.851</td>
<td>0.861</td>
<td>0.136</td>
<td>0.253</td>
<td>0.089</td>
</tr>
<tr>
<td>CAPE&lt;sub&gt;500&lt;/sub&gt;</td>
<td>&gt; 141</td>
<td>0.502</td>
<td>0.247</td>
<td>0.665</td>
<td>0.792</td>
<td>0.188</td>
<td>0.302</td>
<td>0.194</td>
</tr>
<tr>
<td>BI</td>
<td>&gt; 95.1</td>
<td>0.492</td>
<td>0.142</td>
<td>0.839</td>
<td>0.866</td>
<td>0.131</td>
<td>0.242</td>
<td>0.091</td>
</tr>
<tr>
<td>CAPE&lt;sub&gt;50&lt;/sub&gt;</td>
<td>&gt; 43.0</td>
<td>0.486</td>
<td>0.225</td>
<td>0.666</td>
<td>0.808</td>
<td>0.175</td>
<td>0.284</td>
<td>0.207</td>
</tr>
<tr>
<td>TT</td>
<td>&gt; 46.1</td>
<td>0.465</td>
<td>0.124</td>
<td>0.854</td>
<td>0.876</td>
<td>0.121</td>
<td>0.225</td>
<td>0.061</td>
</tr>
<tr>
<td>YI</td>
<td>&gt; 0.7</td>
<td>0.464</td>
<td>0.138</td>
<td>0.802</td>
<td>0.868</td>
<td>0.128</td>
<td>0.23</td>
<td>0.067</td>
</tr>
<tr>
<td>YI&lt;sub&gt;mod&lt;/sub&gt;</td>
<td>&gt; 2.8</td>
<td>0.463</td>
<td>0.137</td>
<td>0.802</td>
<td>0.868</td>
<td>0.128</td>
<td>0.229</td>
<td>0.067</td>
</tr>
<tr>
<td>CAPE&lt;sub&gt;100&lt;/sub&gt;</td>
<td>&gt; 22.1</td>
<td>0.462</td>
<td>0.212</td>
<td>0.647</td>
<td>0.816</td>
<td>0.167</td>
<td>0.269</td>
<td>0.208</td>
</tr>
<tr>
<td>RI</td>
<td>&gt; 34.3</td>
<td>0.421</td>
<td>0.111</td>
<td>0.818</td>
<td>0.883</td>
<td>0.114</td>
<td>0.203</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Table 6.1 Ranking of stability indices based on maximum TSS value using the THUN1 thunderstorm event criteria (at least 1 flash observed within 3 hours of the sounding). Given are the optimal thresholds and verification measures at the maximum TSS. The second last column gives the linear correlation coefficient between forecasts and events when both are either "yes" or "no" (1/0). The last column gives the linear correlation coefficient between the index and number of flashes observed (within 3 hours of the sounding).

CAPE calculated using surface and mixed level parcels are among poorer overall thunderstorm predictors in Finland according to TSS. CAPE<sub>100</sub> received a low TSS
of 0.462, only one index (RI) performing worse. However, CAPE parameters are better predictors for thunderstorm intensity than other indices based on the linear correlation coefficient between the index and number of flashes observed (best is lightning-CAPE with \( r = 0.211 \), best non-CAPE index is LI\(_{\text{mu}}\) with \( r = 0.127 \)).

<table>
<thead>
<tr>
<th>Index</th>
<th>threshold</th>
<th>TSS</th>
<th>HSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI(_{\text{mu}})</td>
<td>&lt; 0.9</td>
<td>0.616</td>
<td>0.238</td>
<td>0.847</td>
<td>0.808</td>
<td>0.186</td>
<td>0.333</td>
</tr>
<tr>
<td>SWISS(_{12})</td>
<td>&lt; 1.6</td>
<td>0.591</td>
<td>0.224</td>
<td>0.829</td>
<td>0.8157</td>
<td>0.178</td>
<td>0.317</td>
</tr>
<tr>
<td>MUCAPE(_{-10...40^\circ\text{C}})</td>
<td>&gt; 5.4</td>
<td>0.582</td>
<td>0.227</td>
<td>0.81</td>
<td>0.813</td>
<td>0.179</td>
<td>0.316</td>
</tr>
<tr>
<td>THOM</td>
<td>&gt; 19.8</td>
<td>0.569</td>
<td>0.195</td>
<td>0.845</td>
<td>0.835</td>
<td>0.161</td>
<td>0.294</td>
</tr>
<tr>
<td>LI</td>
<td>&lt; 1.5</td>
<td>0.568</td>
<td>0.233</td>
<td>0.783</td>
<td>0.808</td>
<td>0.182</td>
<td>0.314</td>
</tr>
<tr>
<td>JI</td>
<td>&gt; 27.5</td>
<td>0.563</td>
<td>0.199</td>
<td>0.829</td>
<td>0.832</td>
<td>0.163</td>
<td>0.294</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>&gt; 113</td>
<td>0.563</td>
<td>0.236</td>
<td>0.77</td>
<td>0.806</td>
<td>0.184</td>
<td>0.315</td>
</tr>
<tr>
<td>LI(_{\text{CL} \text{mod}})</td>
<td>&lt; 3.4</td>
<td>0.56</td>
<td>0.188</td>
<td>0.844</td>
<td>0.838</td>
<td>0.157</td>
<td>0.288</td>
</tr>
<tr>
<td>MUCAPE(_{500})</td>
<td>&gt; 107</td>
<td>0.559</td>
<td>0.232</td>
<td>0.77</td>
<td>0.808</td>
<td>0.181</td>
<td>0.311</td>
</tr>
<tr>
<td>LI(_{50})</td>
<td>&lt; 3.9</td>
<td>0.555</td>
<td>0.18</td>
<td>0.855</td>
<td>0.844</td>
<td>0.152</td>
<td>0.282</td>
</tr>
<tr>
<td>LI(_{CL})</td>
<td>&lt; 3.3</td>
<td>0.551</td>
<td>0.186</td>
<td>0.833</td>
<td>0.839</td>
<td>0.156</td>
<td>0.283</td>
</tr>
<tr>
<td>JI(_{\text{mod}})</td>
<td>&gt; 27.4</td>
<td>0.541</td>
<td>0.174</td>
<td>0.842</td>
<td>0.847</td>
<td>0.149</td>
<td>0.274</td>
</tr>
<tr>
<td>SWISS(_{500})</td>
<td>&lt; 7.1</td>
<td>0.525</td>
<td>0.175</td>
<td>0.814</td>
<td>0.845</td>
<td>0.149</td>
<td>0.269</td>
</tr>
<tr>
<td>KO</td>
<td>&lt; 0.2</td>
<td>0.525</td>
<td>0.177</td>
<td>0.81</td>
<td>0.844</td>
<td>0.15</td>
<td>0.27</td>
</tr>
<tr>
<td>PI</td>
<td>&lt; 1.3</td>
<td>0.524</td>
<td>0.164</td>
<td>0.839</td>
<td>0.853</td>
<td>0.143</td>
<td>0.263</td>
</tr>
<tr>
<td>EI</td>
<td>&lt; 3.6</td>
<td>0.521</td>
<td>0.162</td>
<td>0.839</td>
<td>0.854</td>
<td>0.142</td>
<td>0.262</td>
</tr>
<tr>
<td>NCAPE(_{\text{mu}})</td>
<td>&gt; 0.4</td>
<td>0.52</td>
<td>0.205</td>
<td>0.75</td>
<td>0.825</td>
<td>0.165</td>
<td>0.283</td>
</tr>
<tr>
<td>DCI</td>
<td>&gt; 6.4</td>
<td>0.511</td>
<td>0.157</td>
<td>0.834</td>
<td>0.856</td>
<td>0.14</td>
<td>0.256</td>
</tr>
<tr>
<td>KI</td>
<td>&gt; 21.5</td>
<td>0.498</td>
<td>0.145</td>
<td>0.846</td>
<td>0.864</td>
<td>0.133</td>
<td>0.245</td>
</tr>
<tr>
<td>CAPE(_{500})</td>
<td>&gt; 140</td>
<td>0.494</td>
<td>0.245</td>
<td>0.658</td>
<td>0.793</td>
<td>0.187</td>
<td>0.299</td>
</tr>
<tr>
<td>BI</td>
<td>&gt; 95.1</td>
<td>0.481</td>
<td>0.141</td>
<td>0.829</td>
<td>0.866</td>
<td>0.13</td>
<td>0.238</td>
</tr>
<tr>
<td>TT</td>
<td>&gt; 46.1</td>
<td>0.446</td>
<td>0.119</td>
<td>0.835</td>
<td>0.878</td>
<td>0.119</td>
<td>0.216</td>
</tr>
<tr>
<td>YI(_{\text{mod}})</td>
<td>&gt; 2.6</td>
<td>0.443</td>
<td>0.127</td>
<td>0.802</td>
<td>0.873</td>
<td>0.123</td>
<td>0.218</td>
</tr>
<tr>
<td>YI</td>
<td>&gt; 0.6</td>
<td>0.442</td>
<td>0.128</td>
<td>0.797</td>
<td>0.873</td>
<td>0.123</td>
<td>0.218</td>
</tr>
<tr>
<td>CAPE(_{50})</td>
<td>&gt; 43</td>
<td>0.432</td>
<td>0.202</td>
<td>0.616</td>
<td>0.821</td>
<td>0.161</td>
<td>0.254</td>
</tr>
<tr>
<td>CAPE(_{100})</td>
<td>&gt; 20</td>
<td>0.408</td>
<td>0.183</td>
<td>0.604</td>
<td>0.834</td>
<td>0.15</td>
<td>0.235</td>
</tr>
<tr>
<td>RI</td>
<td>&gt; 34.1</td>
<td>0.39</td>
<td>0.099</td>
<td>0.812</td>
<td>0.889</td>
<td>0.108</td>
<td>0.188</td>
</tr>
<tr>
<td>Persistence</td>
<td></td>
<td>0.313</td>
<td>0.313</td>
<td>0.355</td>
<td>0.645</td>
<td>0.216</td>
<td>0.313</td>
</tr>
</tbody>
</table>

Table 6.2 Same as Table 6.1, but using the THUN2 thunderstorm event criteria (at least 1 flash observed within 6 hours after the sounding).

When using the THUN2 criteria (Table 6.2), the forecast skill of indices can also be compared to a persistence forecast, where the thunderstorm activity in the
previous 6 hour period is used for forecasting subsequent activity. According to TSS, a persistence forecast is worse than all index-based forecasts (TSS = 0.313, while the maximum TSS for the worst-performing index, RI, is 0.390). However, the linear correlation coefficient between the yes/no forecast and event occurrence is higher for a persistence forecast than for almost all stability indices.

6.1.3 Verification using Heidke Skill Score

We repeat the analysis using maximum HSS to assess the forecast skill of indices (Table 6.3). It can be seen that changing the skill score to HSS does not drastically change the overall ranking, with LI\textsubscript{mu}, SWISS\textsubscript{12} and variants of MUCAPE again ranking high. LI\textsubscript{mu} once again comes at first place with a HSS = 0.348 for the optimal threshold of $<-1.2$ °C for forecasting thunderstorm occurrence for the 6 hours following the analysis. Although the rankings are similar, the optimal thresholds do change significantly towards more thundery index values, reflecting a decrease in FAR at the expense of POD. For all of the indices, the POD at the maximum HSS value is less than 0.5, while the FAR values range from approximately 0.67 for the best-performing indices to 0.873 for RI, again ranking last. Hence, FAR is now greater than POD for all of the indices at the optimal threshold according to HSS. When TSS was used, the POD values were just about higher than FAR for the higher ranking indices. What is not shown in any of the tables is POFD (probability of false detection), but this can be calculated by subtracting POD with TSS. At the maximum HSS, the POFD is close to zero for better-performing indices, e.g. 0.065 for LI\textsubscript{mu}, while at the threshold values given by TSS, POFD generally falls between 0.2 and 0.3.

Comparing tables 6.3 and 6.2, we find large differences in index ranking e.g. for NCAPE\textsubscript{mu} (ranked 6th and 18th with HSS and TSS, respectively), CAPE\textsubscript{sfc} (7th and 21st), CAPE\textsubscript{50} and CAPE\textsubscript{100} which are much better according to HSS, and JI (14th and 6th). Since a persistence forecast has a relatively low FAR (0.645), it is ranked surprisingly high based on HSS, at 9th place! According to the Heidke Skill Score, a persistence forecast is therefore superior to forecasts made using most stability indices, e.g. JI, which ranked 6th using TSS compared to the persistence forecast at last place. However, the persistence forecast has a low POD (0.355). Indices which are relatively poor predictors for thunderstorm occurrence in Finland according to both skill scores are RI, YI and YI\textsubscript{mod}, TT, BI and KI. This is not surprising, given that these indices were developed with entirely different regions.
(e.g. Japan with YI) or specific thunderstorm types (RI and JI were designed to predict air-mass thunderstorms) in mind.

<table>
<thead>
<tr>
<th>Index</th>
<th>threshold</th>
<th>HSS</th>
<th>TSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LI_{00}</td>
<td>&lt; −1.2</td>
<td>0.348</td>
<td>0.428</td>
<td>0.493</td>
<td>0.67</td>
<td>0.246</td>
<td>0.356</td>
</tr>
<tr>
<td>2</td>
<td>SLI</td>
<td>&lt; −1.1</td>
<td>0.342</td>
<td>0.398</td>
<td>0.456</td>
<td>0.663</td>
<td>0.24</td>
<td>0.346</td>
</tr>
<tr>
<td>3</td>
<td>MCAPE_{-10...-40°C}</td>
<td>&gt; 155</td>
<td>0.337</td>
<td>0.41</td>
<td>0.474</td>
<td>0.675</td>
<td>0.239</td>
<td>0.345</td>
</tr>
<tr>
<td>4</td>
<td>MCAPE</td>
<td>&gt; 408</td>
<td>0.337</td>
<td>0.40</td>
<td>0.461</td>
<td>0.671</td>
<td>0.238</td>
<td>0.342</td>
</tr>
<tr>
<td>5</td>
<td>SWISS_{12}</td>
<td>&lt; −1.3</td>
<td>0.334</td>
<td>0.41</td>
<td>0.470</td>
<td>0.678</td>
<td>0.237</td>
<td>0.341</td>
</tr>
<tr>
<td>6</td>
<td>NCAPE_{nu}</td>
<td>&gt; 0.8</td>
<td>0.333</td>
<td>0.376</td>
<td>0.43</td>
<td>0.663</td>
<td>0.233</td>
<td>0.335</td>
</tr>
<tr>
<td>7</td>
<td>CAPE_{500}</td>
<td>&gt; 401</td>
<td>0.331</td>
<td>0.369</td>
<td>0.423</td>
<td>0.662</td>
<td>0.231</td>
<td>0.333</td>
</tr>
<tr>
<td>8</td>
<td>MCAPE_{500}</td>
<td>&gt; 293</td>
<td>0.325</td>
<td>0.399</td>
<td>0.466</td>
<td>0.687</td>
<td>0.23</td>
<td>0.332</td>
</tr>
<tr>
<td>9</td>
<td>Persistence</td>
<td></td>
<td>0.313</td>
<td>0.313</td>
<td>0.355</td>
<td>0.645</td>
<td>0.216</td>
<td>0.313</td>
</tr>
<tr>
<td>10</td>
<td>LI_{50}</td>
<td>&lt; 0.0</td>
<td>0.295</td>
<td>0.351</td>
<td>0.415</td>
<td>0.705</td>
<td>0.208</td>
<td>0.3</td>
</tr>
<tr>
<td>11</td>
<td>CAPE_{50}</td>
<td>&gt; 165</td>
<td>0.286</td>
<td>0.32</td>
<td>0.377</td>
<td>0.7</td>
<td>0.201</td>
<td>0.288</td>
</tr>
<tr>
<td>12</td>
<td>THOM</td>
<td>&gt; 27.3</td>
<td>0.281</td>
<td>0.36</td>
<td>0.436</td>
<td>0.728</td>
<td>0.201</td>
<td>0.291</td>
</tr>
<tr>
<td>13</td>
<td>CAPE_{100}</td>
<td>&gt; 101</td>
<td>0.274</td>
<td>0.301</td>
<td>0.356</td>
<td>0.706</td>
<td>0.192</td>
<td>0.275</td>
</tr>
<tr>
<td>14</td>
<td>JI</td>
<td>&gt; 29.5</td>
<td>0.273</td>
<td>0.361</td>
<td>0.442</td>
<td>0.738</td>
<td>0.197</td>
<td>0.284</td>
</tr>
<tr>
<td>15</td>
<td>SI_{50}</td>
<td>&lt; 1.6</td>
<td>0.273</td>
<td>0.348</td>
<td>0.423</td>
<td>0.733</td>
<td>0.196</td>
<td>0.282</td>
</tr>
<tr>
<td>16</td>
<td>DCI</td>
<td>&gt; 16.4</td>
<td>0.272</td>
<td>0.331</td>
<td>0.4</td>
<td>0.727</td>
<td>0.194</td>
<td>0.278</td>
</tr>
<tr>
<td>17</td>
<td>SI_{LCL}</td>
<td>&lt; 0.9</td>
<td>0.262</td>
<td>0.355</td>
<td>0.44</td>
<td>0.75</td>
<td>0.19</td>
<td>0.274</td>
</tr>
<tr>
<td>18</td>
<td>SI_{LCLmod}</td>
<td>&lt; 1.2</td>
<td>0.261</td>
<td>0.375</td>
<td>0.47</td>
<td>0.757</td>
<td>0.191</td>
<td>0.278</td>
</tr>
<tr>
<td>19</td>
<td>SWISS_{00}</td>
<td>&lt; 4.1</td>
<td>0.255</td>
<td>0.345</td>
<td>0.431</td>
<td>0.754</td>
<td>0.186</td>
<td>0.267</td>
</tr>
<tr>
<td>20</td>
<td>EI</td>
<td>&lt; −0.3</td>
<td>0.25</td>
<td>0.335</td>
<td>0.42</td>
<td>0.757</td>
<td>0.182</td>
<td>0.261</td>
</tr>
<tr>
<td>21</td>
<td>KO</td>
<td>&lt; −2.8</td>
<td>0.241</td>
<td>0.36</td>
<td>0.463</td>
<td>0.775</td>
<td>0.179</td>
<td>0.26</td>
</tr>
<tr>
<td>22</td>
<td>BI</td>
<td>&gt; 28.7</td>
<td>0.228</td>
<td>0.292</td>
<td>0.372</td>
<td>0.768</td>
<td>0.167</td>
<td>0.235</td>
</tr>
<tr>
<td>23</td>
<td>YI</td>
<td>&gt; 2.4</td>
<td>0.181</td>
<td>0.286</td>
<td>0.404</td>
<td>0.802</td>
<td>0.153</td>
<td>0.215</td>
</tr>
<tr>
<td>24</td>
<td>YI_{mod}</td>
<td>&gt; 4.4</td>
<td>0.18</td>
<td>0.289</td>
<td>0.411</td>
<td>0.821</td>
<td>0.143</td>
<td>0.2</td>
</tr>
<tr>
<td>25</td>
<td>TT</td>
<td>&gt; 48.7</td>
<td>0.148</td>
<td>0.35</td>
<td>0.567</td>
<td>0.855</td>
<td>0.131</td>
<td>0.197</td>
</tr>
<tr>
<td>26</td>
<td>RI</td>
<td>&gt; 35.7</td>
<td>0.119</td>
<td>0.303</td>
<td>0.547</td>
<td>0.873</td>
<td>0.115</td>
<td>0.165</td>
</tr>
</tbody>
</table>

Table 6.3 Same as Table 6.2 (using THUN2 criteria where soundings fall 0-6 hours before events), but the ranking, optimal thresholds, and verification measure values are based on maximum HSS.

6.1.4 Diurnal variation

The results presented thus far describe the overall skill of stability indices to predict the occurrence of May-August thunderstorms in Finland without any specification of thunderstorm type or time of day, given that all lightning observations and model
soundings were included. However, distinguishing between 00, 06, 12 and 18 UTC soundings can be important due to the diurnal cycle of solar heating, which manifests itself as an increase in latent instability. Largest instabilities are therefore often found for surface-based parcels in the afternoon, whereas nocturnal thunderstorms are often associated with elevated convection, whereupon the most unstable level is found at higher levels. Both the optimal thresholds and relative forecast skill of indices are expected to have a diurnal dependency, given that indices are tied to specific parcels. Therefore, we repeat the analysis for 12 UTC and 00 UTC model soundings separately. Table 6.4 shows the ranking of indices when only 00 UTC (3 AM local time) soundings are included. Since the THUN1 criteria is used, this corresponds to thunderstorms occurring between 00 and 06 local time.

As expected, the best indices for predicting night-time thunderstorms are based on a most unstable parcel. LI_{mu} once again clearly leads other indices, with the maximum TSS unchanged from before (0.646). The THOM index (THOM = KI - LI_{so}) is ranked third at TSS = 0.613, KI itself at 8th place, and SI_{850} at 6th place with TSS = 0.587. These indices all depend on the static stability between 850 and 500 hPa (and THOM also depends on conditions near the surface). The indication is that considering parcels up to 850 hPa is in most cases adequate to assess the potential for convection initiating from above the surface, although LI_{mu} which uses the most unstable parcel in the lowest 300 hPa does have a significantly higher POD than THOM (0.857 and 0.822, respectively) at a similar FAR. Normalized MUCAPE, defined as the total MUCAPE divided by the depth of the free convective layer (the layer between the LFC and EL), is also an effective dichotomous predictor of night-time thunderstorms, coming in at 7th place with TSS = 0.581, considerably higher than when all soundings were included (18th with TSS = 0.542 as seen in Table 6.1).

The worst index for predicting thunderstorms using 00 UTC soundings is CAPE_{sfc}, with a very low TSS of 0.259, significantly lower than the index with the second poorest performance (RI, TSS = 0.417). Clearly, using surface-based parcels to forecast nocturnal thunderstorms will generally yield very poor results.
For predicting afternoon thunderstorms, the best indices are $\text{LI}_{\text{mu}}$, $\text{SI}_{\text{LCLmod}}$ and $\text{JI}$ according to TSS (Table 6.5). $\text{SI}_{\text{LCLmod}}$ was defined in this work as the temperature difference between the environment and 500 hPa and a parcel that is pseudoadiabatically lifted from the LCL to 500 hPa, when the mean temperature, humidity and pressure of the lowest 50 hPa are used to calculate the LCL.

$\text{LI}_{\text{mu}}$ received a maximum TSS of 0.642 at the optimal threshold of $\text{LI}_{\text{mu}} < 0.5$. $\text{LI}_{\text{mu}}$ (1st) and MUCAPE (9th) both performed better in the TSS ranking than their surface-based equivalents ($\text{SLI}$ at 7th and CAPE$_{\text{sfc}}$ at 15th place). Based on
these results, there is no mean benefit of using a surface-based parcel over the most unstable one; not even in daytime cases, where they are often synonymous with each other anyway.

<table>
<thead>
<tr>
<th>Index</th>
<th>threshold</th>
<th>TSS</th>
<th>HSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LI_mu</td>
<td>&lt; 0.5</td>
<td>0.642</td>
<td>0.350</td>
<td>0.874</td>
<td>0.697</td>
<td>0.290</td>
<td>0.427</td>
<td>0.171</td>
</tr>
<tr>
<td>SI_CL_mod</td>
<td>&lt; 3.3</td>
<td>0.627</td>
<td>0.340</td>
<td>0.861</td>
<td>0.703</td>
<td>0.283</td>
<td>0.416</td>
<td>0.151</td>
</tr>
<tr>
<td>JI</td>
<td>&gt; 27.6</td>
<td>0.624</td>
<td>0.332</td>
<td>0.867</td>
<td>0.709</td>
<td>0.278</td>
<td>0.411</td>
<td>0.140</td>
</tr>
<tr>
<td>SI_LCL</td>
<td>&lt; 3.3</td>
<td>0.621</td>
<td>0.332</td>
<td>0.862</td>
<td>0.709</td>
<td>0.278</td>
<td>0.410</td>
<td>0.151</td>
</tr>
<tr>
<td>LI50</td>
<td>&lt; 2.1</td>
<td>0.621</td>
<td>0.349</td>
<td>0.841</td>
<td>0.695</td>
<td>0.288</td>
<td>0.418</td>
<td>0.150</td>
</tr>
<tr>
<td>SLI</td>
<td>&lt; 0.6</td>
<td>0.617</td>
<td>0.341</td>
<td>0.845</td>
<td>0.701</td>
<td>0.283</td>
<td>0.413</td>
<td>0.149</td>
</tr>
<tr>
<td>MUCAPE_10..-40°C</td>
<td>&gt; 28</td>
<td>0.613</td>
<td>0.348</td>
<td>0.830</td>
<td>0.695</td>
<td>0.287</td>
<td>0.415</td>
<td>0.272</td>
</tr>
<tr>
<td>SWISS12</td>
<td>&lt; 1.1</td>
<td>0.609</td>
<td>0.332</td>
<td>0.843</td>
<td>0.707</td>
<td>0.277</td>
<td>0.405</td>
<td>0.151</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>&gt; 163</td>
<td>0.599</td>
<td>0.343</td>
<td>0.814</td>
<td>0.697</td>
<td>0.283</td>
<td>0.407</td>
<td>0.272</td>
</tr>
<tr>
<td>SI450</td>
<td>&lt; 4.0</td>
<td>0.595</td>
<td>0.291</td>
<td>0.880</td>
<td>0.738</td>
<td>0.253</td>
<td>0.380</td>
<td>0.143</td>
</tr>
<tr>
<td>MUCAPE590</td>
<td>&gt; 146</td>
<td>0.589</td>
<td>0.331</td>
<td>0.813</td>
<td>0.706</td>
<td>0.276</td>
<td>0.397</td>
<td>0.222</td>
</tr>
<tr>
<td>THOM</td>
<td>&gt; 19.3</td>
<td>0.588</td>
<td>0.295</td>
<td>0.863</td>
<td>0.733</td>
<td>0.255</td>
<td>0.379</td>
<td>0.136</td>
</tr>
<tr>
<td>JI_mod</td>
<td>&gt; 27.3</td>
<td>0.578</td>
<td>0.285</td>
<td>0.861</td>
<td>0.741</td>
<td>0.248</td>
<td>0.370</td>
<td>0.122</td>
</tr>
<tr>
<td>KO</td>
<td>&lt; -0.1</td>
<td>0.571</td>
<td>0.281</td>
<td>0.855</td>
<td>0.743</td>
<td>0.246</td>
<td>0.366</td>
<td>0.162</td>
</tr>
<tr>
<td>CAPE_sfc</td>
<td>&gt; 211</td>
<td>0.567</td>
<td>0.341</td>
<td>0.766</td>
<td>0.693</td>
<td>0.280</td>
<td>0.394</td>
<td>0.258</td>
</tr>
<tr>
<td>EI</td>
<td>&lt; 3.6</td>
<td>0.560</td>
<td>0.266</td>
<td>0.862</td>
<td>0.753</td>
<td>0.237</td>
<td>0.355</td>
<td>0.156</td>
</tr>
<tr>
<td>PI</td>
<td>&lt; 1.3</td>
<td>0.556</td>
<td>0.267</td>
<td>0.854</td>
<td>0.752</td>
<td>0.238</td>
<td>0.353</td>
<td>0.140</td>
</tr>
<tr>
<td>SWISS00</td>
<td>&lt; 7.1</td>
<td>0.554</td>
<td>0.282</td>
<td>0.825</td>
<td>0.741</td>
<td>0.246</td>
<td>0.359</td>
<td>0.139</td>
</tr>
<tr>
<td>NCAPE_mw</td>
<td>&gt; 0.5</td>
<td>0.553</td>
<td>0.333</td>
<td>0.753</td>
<td>0.698</td>
<td>0.275</td>
<td>0.385</td>
<td>0.255</td>
</tr>
<tr>
<td>CAPE50</td>
<td>&gt; 32</td>
<td>0.518</td>
<td>0.300</td>
<td>0.737</td>
<td>0.721</td>
<td>0.253</td>
<td>0.354</td>
<td>0.273</td>
</tr>
<tr>
<td>IK</td>
<td>&gt; 20.9</td>
<td>0.502</td>
<td>0.222</td>
<td>0.847</td>
<td>0.780</td>
<td>0.212</td>
<td>0.312</td>
<td>0.118</td>
</tr>
<tr>
<td>TT</td>
<td>&gt; 46.2</td>
<td>0.498</td>
<td>0.212</td>
<td>0.865</td>
<td>0.787</td>
<td>0.206</td>
<td>0.307</td>
<td>0.091</td>
</tr>
<tr>
<td>DCI</td>
<td>&gt; 7.3</td>
<td>0.483</td>
<td>0.219</td>
<td>0.817</td>
<td>0.781</td>
<td>0.209</td>
<td>0.302</td>
<td>0.149</td>
</tr>
<tr>
<td>CAPE100</td>
<td>&gt; 20</td>
<td>0.482</td>
<td>0.306</td>
<td>0.669</td>
<td>0.709</td>
<td>0.255</td>
<td>0.345</td>
<td>0.275</td>
</tr>
<tr>
<td>BI</td>
<td>&gt; 95</td>
<td>0.479</td>
<td>0.201</td>
<td>0.854</td>
<td>0.793</td>
<td>0.200</td>
<td>0.295</td>
<td>0.122</td>
</tr>
<tr>
<td>YI</td>
<td>&gt; 0.7</td>
<td>0.458</td>
<td>0.196</td>
<td>0.825</td>
<td>0.795</td>
<td>0.197</td>
<td>0.283</td>
<td>0.090</td>
</tr>
<tr>
<td>YI_mod</td>
<td>&gt; 2.8</td>
<td>0.457</td>
<td>0.195</td>
<td>0.827</td>
<td>0.795</td>
<td>0.196</td>
<td>0.282</td>
<td>0.090</td>
</tr>
<tr>
<td>RI</td>
<td>&gt; 34.3</td>
<td>0.429</td>
<td>0.173</td>
<td>0.832</td>
<td>0.808</td>
<td>0.185</td>
<td>0.263</td>
<td>0.076e</td>
</tr>
</tbody>
</table>

Table 6.5 Same as Table 6.1, using THUN1 thundery criteria and maximum TSS, but including only 12 UTC soundings.

The mean maximum TSS of indices was 0.527 using 00 UTC soundings and 0.558 using 12 UTC soundings. However, this is a reflection of most indices having been designed to predict PBL-based convection, rather than the predictability of elevated convection (more often associated with 00 UTC soundings) being lower. The
maximum TSS values for LI_{mu} and the MUCAPE parameters are almost identical for daytime and night-time soundings, with the exception of normalized MUCAPE (NCAPE_{mu}), which has a higher TSS using night-time soundings (0.581 compared to 0.553 with 12 UTC soundings).

6.2 Thundery Case Probability

6.2.1 One-dimensional TCP plots

Thunderstorm probabilities (thundery case probability, TCP) were derived as a function of the index values, as described in Sect. 4.4.2. Although LI_{mu} was the most successful dichotomous thunderstorm predictor overall based on skill scores, higher maximum TCP values (above 60%) were obtained with SLI (Figure 6.2), while LI_{mu} reached a maximum TCP of just over 55% (Figure 6.2). Other stability indices which reached maximum TCP values of over 50% were SWISS_{12}, THOM and KI.

As expected, TCP tends to increase monotonically with more thundery index values. However, most indices show a curious drop in TCP for extremely thundery index values. This is can be seen very clearly e.g. LI_{mu} as the TCP, after reaching a maximum at approximately -6 °C, decreases almost constantly for decreasing index

Figure 6.2 TCP (Thundery Case Probability) as a function of Lifted Index using surface and most unstable parcels. The THUN2 criteria was used to distinguish between thunderstorms and null cases. Probabilities were calculated using the moving average of thunderstorm occurrence with a window size of N = 3000.
values. However, it is worth noting that the number of cases having index values this thundery is very small relative to the total number of cases, e.g. there are 3300 cases where LI$_{mu} < -6$ °C, but this corresponds to only 4.4% of all thundery cases, or 0.2% of the total sample size.

CAPE parameters do not show this drop in TCP, except for CAPE$_{50}$ which exhibits a weak one (Figure 6.3). Highest maximum TCP values (around 60%) are found for surface-based CAPE and MUCAPE$_{500}$, which differs from normal MUCAPE in that the integration is stopped at 500 hPa. Other indices which do not clearly show a drop in TCP for extremely thundery index values are KI, YI and YI$_{mod}$.

Slightly higher maximum TCP values were generally reached using the THUN2 criteria over THUN1, especially in the case of surface-based indices.

Finally, Figure 6.4 shows TCP as a function of some vertical wind shear parameters. It can be seen that increasing VWS between 10 m and 3 km and VWS between 3 km and 6 km has the opposite effect on thunderstorm probability, while highest TCP values are reached with the VWS$_{subLFC}$ parameter, defined as the shear between 10 m and the LFC of the most unstable parcel; or 875 hPa, if no LFC exists.
6.3 Spatial distribution

It is of interest to ask to what extent the results presented thus far are spatially homogenous for Finland. Owing to the huge data set, the skill score test can be repeated for individual grid boxes in our data set with decently non-random results. Figure 6.5 shows the spatial distribution of maximum TSS for LI\textsubscript{mu}. Surprisingly, a large degree of regional variance is found. There is a distinct north-south gradient, where significantly higher maximum TSS values of around 0.75 are found for Lapland, whereas the southern coast of Finland shows TSS values as low as 0.5 or lower. Since TSS is a measure of forecast skill, this would perhaps indicate that the predictability of thunderstorm occurrence using stability indices is higher for northern Finland than southern Finland. This result was not confirmed by maximum HSS, which was higher and fairly homogenous over land areas, and lower for maritime grid boxes. However, using HSS is problematic since this skill score is susceptible to the event frequency, shown in Figure 6.6. The larger thunderstorm frequency in the south tends to increase maximum HSS, this is likely the main reason why there is no apparent north-south divide in maximum HSS such that there is for TSS. Secondly, the emphasis of TSS and HSS on POD and FAR, respectively, may explain some of the discrepancy as well.
The right hand side of Figure 6.5 shows the spatial distribution of the optimal thresholds for LI_{\text{mu}} according to TSS. These values are more random than the maximum TSS values, but there is still a fairly clear pattern of significantly lower optimal thresholds for central Finland compared to some other regions, specifically north-eastern Lapland and the Bothnian Bay. However, the large differences between some adjacent grid boxes suggests that this result is associated with considerably more uncertainty than the distribution of forecast skill given by maximum TSS.

Repeating the analysis for a few other indices (the SWISS\textsubscript{12} and SI\textsubscript{850}), it is found that the maximum TSS pattern is highly similar to that obtained with LI_{\text{mu}}. However, the pattern for the optimal threshold values were not. The optimal forecasting threshold for the SWISS\textsubscript{12} according to TSS index was fairly homogenous over Finland, while for SI\textsubscript{850}, a north-south gradient was seen, so that the threshold values for the southern coast were slightly higher overall (around 1 °C) compared to central and northern parts of Finland.
To gain some insight into this issue, TCP was calculated separately for northern and the southern half of Finland, dividing the data into two by latitude (≈ 65° N). The TCP of $LI_{\text{mu}}$ for the southern part of Finland reached a maximum of only 0.5 around $LI_{\text{mu}} = -5$ °C, while the maximum TCP for the northern part of Finland reached almost 0.7 around -6 °C. Similar results were obtained for SWISS$_{12}$.

### 6.3.1 Two-dimensional TCP plots

We wish to repeat the analysis of van Zomeren and van Delden (2007), who found that vertically integrated moisture flux convergence (VIMFC) is a very useful parameter to combine with a stability index, to investigate if their results hold for Finland. Van Zomeren and van Delden (2007) also utilized ECMWF analyses, but with slightly coarser spatial resolution, and had a data set two order of magnitudes smaller. In addition, following the work of Banacos and Schultz (2005), we wish to investigate if horizontal mass convergence is superior or inferior to MFC based on its forecasting ability inferred from thundery case probability (TCP) when combined with a stability index.
Van Zomeren and Van Delden (2007) used VIMFC integrated between 1000 and 700 hPa and "the stability of the most unstable level" or SMUL, which should be nearly identical to LI\textsubscript{mu}, since SMUL was defined at the minimum value of the four Lifted Indices calculated using a 1000, 925, 850 and 700 hPa parcel. Figure 6.7 depicts TCP as a function of LI\textsubscript{mu} and VIMFC integrated between the surface and 300 hPa above the surface. This is preferable to simply using 1000 hPa and 700 hPa, since it was found that the reanalysis provided convergence and mixing ratio values above zero for pressure levels below the surface. Given the large data set, cases were divided into bins using smaller bin widths compared to Van Zomeren and Van Delden (2007), and using a larger minimum bin count (N ≥ 80), i.e. TCP values are only shown for bins where the number of cases is at least 80. We assume that this threshold is ample to ensure decent statistics.

![Figure 6.7](image)

**Figure 6.7** TCP as a function of the most unstable Lifted Index and VIMFC (300 hPa integration depth) using the THUN2 criteria. TCP is shown only for bins which have at least 80 cases.

The results are fairly similar to Van Zomeren and Van Delden (2007). For a given value of LI\textsubscript{mu}, TCP increases with increasing VIMFC, so that very-high TCP values are found for combinations of small stability values and large VIMFC values (TCP > 80% when LI\textsubscript{mu} is between -4 and -5 °C and VIMFC is between 40 and 50 10^{-5} \text{kg m}^{-2} \text{s}^{-1}).
Next, we calculate the same table using vertically integrated horizontal mass convergence (referred to as VIMC here) instead of moisture convergence (Figure 6.8). VIMC was calculated simply by dropping the mixing ratio $r$ from equation 3.7, and similarly to Figure 6.7, we use the THUN2 criteria where the analysis precedes events by 0-6 hours.

Figure 6.8 Same as Figure 6.7, but TCP as a function of the most unstable Lifted Index and VIMC (300 hPa integration depth).

The results indicate that VIMC is equally as useful as VIFMC, if not more, as a supplementary parameter to use with a stability index to forecast thunderstorm occurrence. Although the highest TCP value is slightly higher with VIMFC, the pattern of increasing TCP with increasing VIMC is more consistent than with VIMFC. It is also worth noting that the results are sensitive to choice of bins. Using an even smaller bin width of 5 units on the x-axis led to graphs where highest TCP value was found for VIMC instead of VIMFC (not shown), and extreme TCP values were clearly higher for VIMC (6 bins where TCP exceeded 70%, while the VIMFC-LI mu TCP table only had one such bin). However, given the subjective nature of the interpretation and sensitivity to bin allocation, the results can be considered highly similar overall for VIMC and VIFMC, and it cannot be inferred from these tables.
that either parameter is clearly superior to the other as a non-dichotomous predictor of thunderstorm occurrence.

Figure 6.9 shows TCP as a function of SLI and VIMC calculated with a 300 hPa integration depth. It can be clearly seen that the maximum TCP values are considerably higher when SLI is used instead of $\text{LI}_{\text{mu}}$, as there are several bins with TCP values near 80% or higher. Thus, the combination of strong positive low-level convergence and strong surface-based instability is more likely to lead to thunderstorm activity than the presence of similar convergence and equally strong instability but with respect to a most unstable parcel (which can be the surface parcel, or not).

Finally, we once again use $\text{LI}_{\text{mu}}$ and horizontal mass convergence to calculate TCP but increase the integration depth used to calculate VIMC to 500 hPa. The resulting table (Figure 6.10) suggests that considering a deeper layer near the surface when calculating convergence (at least on the scale of the reanalysis) could be advantageous for forecasting convective initiation, since highest TCP values increase considerably. In addition, the pattern of increasing TCP with increasing VIMC seems to clearer and more consistent for VIMC_{500} than VIMC_{300} (Figure 6.8).
Analysis was also repeated with an integration depth of 700 hPa, with highly similar results to using 500 hPa (not shown).

Figure 6.10 Same as Figure 6.7, but TCP as a function of the most unstable Lifted Index and VIMC (500 hPa integration depth).

Although TCP values nearing or exceeding 80% give the impression of strong forecast value, this might not be the case if the environmental conditions associated with such TCP values are almost never met. Indeed, the distribution of thundery cases corresponding to Figure 6.10, shown in Figure 6.11, clearly illustrates that the combination of such strong positive low-level convergence and a very unstable troposphere is exceedingly rare in Finland. For example, if we consider the four bins with TCP values near or above 80%, the total number of model soundings in the data set corresponding to these bins comprise only 0.48% of all thundery cases. Given that the mean event frequency (or TCP) for the entire data set of 1.6 million cases is 6.1%, the overall frequency of such extreme thundery conditions is thus less than 0.03%.

63
6.4 Mean temperature soundings

The mean temperature and dew-point temperature soundings are presented in Figure 6.12 as the difference between thundery and non-thundery cases, utilizing all soundings in the data set. The temperature curve shows a markedly warmer lower troposphere in thundery compared to non-thundery cases, as the layer below 800 hPa is on average 4-5 degrees warmer. Above this level, the temperature difference drops until 600-700 hPa, where it remains near a rather constant 2 degrees until 300-400 hPa. After this, it drops quickly to near zero.

The humidity profile as indicated by the difference in dew-point temperature (inversely proportional to relative humidity) shows a clear S-shaped curve. A local maxima is found near the surface, where the dew-point temperature is almost 1 degree lower in thundery cases, corresponding to a (slightly) higher relative humidity in thundery cases. Given that the temperature profile is almost constant in the lowest 100 hPa, being much warmer in thundery cases, this must be explained by a large difference in absolute humidity near the surface. A minima in the dew point-temperature difference is found at 900 hPa, where there is hardly any difference between thundery and non-thundery cases. However, since the boundary layer temperature is markedly higher in thundery cases, the absolute humidity content must still be significantly higher.
The largest difference in dew-point temperature is found at 700 hPa, where it is more than 3.5 degrees lower in thundery cases. This reflects a maxima in relative humidity (RH). The maxima may be partially explained by the temperature profile, given that the temperature difference drops quite significantly near the 700 hPa level and that RH increases with decreasing temperature. The existence of a humid layer above the condensation level is possibly of great importance for DMC if mixing would otherwise stop the convection in its early stages.

Comparing our results to those of Roine (2001), the shape of the profiles are very similar, but the values of the differences are not. The contrast is striking for the mean temperatures in the middle and upper atmosphere: in Roine (2001), the temperature difference dropped to below 2 degrees above 800 hPa and even became negative at 500-700 hPa, whereas here, the temperature difference remains strongly positive (more than 2 degrees) up to 300-400 hPa. For humidity however, the differences were larger in Roine (2001), as the dew point depression difference approached a striking 6 degrees at 700 hPa for 12 UTC soundings, and was near 4 degrees for 00 UTC soundings.

Figure 6.12  Difference between the mean temperature soundings (temperature and dew point depression) of thundery and null cases.
6.5 Developing a new index for Finland

Roine (2001) noted the large difference in the dew point depression at 700 hPa between thundery and non-thundery cases, reproduced here, but was rather pessimistic about the utilization of it in stability indices, given that existing indices incorporating it (e.g. KI and JI\textsubscript{mod}) performed relatively poorly in skill score comparisons. However, given that 1) the SWISS\textsubscript{12} index was one of the best-performing indices in our study and this index uses the 700 hPa dew-point depression, and 2) LI\textsubscript{mu} was the best index overall, there may be potential for developing a new index for Finland based on LI\textsubscript{mu} and the 700 hPa dew-point depression.

To do this, we simply take the existing SWISS\textsubscript{12} index of Huntrieser et al. (1997) but use LI\textsubscript{mu} instead of SLI, and the 700-hPa dew-point depression instead of the 650 hPa one. Huntrieser et al. incorporated the vertical wind shear between the surface and 3 km (VWS\textsubscript{0-3}) based on a distinct low-level jet associated with thundery cases in Switzerland revealed by a hodograph analysis. In this work, we have not calculated mean hodographs for thundery and non-thundery cases, but TCP as a function of VWS\textsubscript{0-3} did reveal a monotonically increasing TCP with increasing VWS\textsubscript{0-3} (Figure 6.4). This supports the hypothesis that the vertical wind shear is correlated with thunderstorm development also in Finland.

Thus, a new index is developed similar to the SWISS\textsubscript{12} index but adapted to the local climatology of convection in Finland. To find the optimal coefficients for this index, a logistic regression was performed with thunderstorm occurrence based on the THUN2 classification as the predictor. This provided a good first guess; however, slightly modifying the coefficients revealed that it did not result in the highest maximum TSS value. A combination of trial and error and a test based on the linear correlation coefficient between the index and predictor, resulted in the following linear combination for the seemingly highest TSS value (utilizing all soundings):

\[
FIN = LI\textsubscript{mu} - 0.1 \cdot VWS\textsubscript{0-3} + 0.1 \cdot (T\textsubscript{700} - T_d\textsubscript{700}).
\] (6.1)
The newly developed "FIN" index receives a maximum TSS of 0.663 with the THUN1 criteria, which represents a somewhat meaningful improvement over LI\textsubscript{mu} whose TSS = 0.646 (Table 6.6). At the optimal threshold, POD was unchanged but the false alarm ratio dropped from 0.803 to 0.789 with the incorporation of the wind shear and dew point depression. In a maximum HSS test, FIN was also superior to LI\textsubscript{mu}, obtaining a HSS of 0.366 at threshold < -1.3, while LI\textsubscript{mu} received a HSS of 0.348 (THUN2), again mostly due to lower FAR.

In the development of the index, alternative wind shear parameters were also tested. Specifically, VWS in the subcloud layer (Section 3.2.2) showed promise based on TCP, since this parameter reached a much higher maximum TCP than VWS\textsubscript{0-3} (Figure 6.4). Another parameter, the VWS in the active cloud-bearing layer (3.2.2), exhibited decreasing TCP values for increasing shear (not shown), as hypothesized. However, neither parameter seemed to increase the forecast skill of LI\textsubscript{mu} based on maximum TCP for the combination of terms tested, except VWS\textsubscript{ACBL} which only very marginally increased the maximum TSS of FIN when added as an additional positive term to equation 6.1.

### Table 6.6 Best performing indices according to TSS and the THUN1 criteria with the newly developed FIN index included.

<table>
<thead>
<tr>
<th>Index</th>
<th>threshold</th>
<th>TSS</th>
<th>HSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIN</td>
<td>&lt; 0.6</td>
<td>0.663</td>
<td>0.268</td>
<td>0.873</td>
<td>0.789</td>
<td>0.205</td>
<td>0.332</td>
<td>0.127</td>
</tr>
<tr>
<td>LI\textsubscript{mu}</td>
<td>&lt; 0.9</td>
<td>0.646</td>
<td>0.248</td>
<td>0.876</td>
<td>0.803</td>
<td>0.192</td>
<td>0.348</td>
<td>0.127</td>
</tr>
<tr>
<td>MUCAPE\textsubscript{10$^\circ$C}</td>
<td>&gt; 5.4</td>
<td>0.611</td>
<td>0.237</td>
<td>0.838</td>
<td>0.808</td>
<td>0.185</td>
<td>0.33</td>
<td>0.211</td>
</tr>
<tr>
<td>SWISS\textsubscript{12}</td>
<td>&lt; 1.6</td>
<td>0.609</td>
<td>0.25</td>
<td>0.819</td>
<td>0.799</td>
<td>0.192</td>
<td>0.337</td>
<td>0.109</td>
</tr>
</tbody>
</table>

6.6 Artificial Neural Network

The input-selection algorithm initially chose the following parameters as the first inputs: LI\textsubscript{mu}, ACTP, and VIMFC\textsubscript{700}. What was realized from VIMFC\textsubscript{700} being selected so early on, is that the parameter includes not only lift but moisture information - arguably combining them rather arbitrarily - which explains its success as an ANN input but might not lead to the optimal set of ANN inputs. This is because it can generally be regarded as desirable to keep quantities representing different aspects of CI, such as instability and lift - separate, in the spirit of ingredients-based forecasting (for this reason, the SWISS and FIN index were excluded from the parameter pool). Furthermore, because vertically integrated mass and moisture flux
convergences were strongly linked to TCP previously, we might be interested in the forecast skill of the best neural network found which does not utilize convergence, compared to one that does.

Thus, the script was stopped and re-run with the convergence parameters (VIMC and VIMFC) excluded from the parameter pool. This led to the following combination of inputs (in order of selection): LI$_{mu}$, ACTP, SI$_{850}$, HH, TCW, MRH$_{800-600}$, MRH$_{acbl}$, SLI, VWS$_{sublfc}$ and CIN$_{875}$; a total of 10 parameters. Further parameters decreased the CEE of the network only negligibly, and so the very time-consuming script was stopped. At this point, the convergence parameters were put back to the input pool, and the script modified to continue from the existing input set. This time, VIMC$_{500}$ was chosen as an additional and final input, leading to a more noticeable decrease in validation CEE, for a final CEE of 0.139 (further inputs failed to decrease the CEE by more than 0.5%).

The decrease in cross entropy error as more inputs were added to the network is depicted in Fig. 6.6. In this experiment, ACTP (previous activity, i.e. whether lightning occurred in the previous 6-hour time period) was clearly the parameter which improved performance the most after the principal stability index (LI$_{mu}$). This result was also obtained in Manzato (2005), and is not surprising, given that how well the simple persistence forecast fared in the HSS ranking. After ACTP, the variable slightly edging out the others in terms of performance increase was SI$_{850}$. After this, HH (synoptic sounding hour), the other non-physical parameter in the pool besides ACTP, managed to decrease the CEE noticeably along with TCW. TCW is an ERA-I parameter which estimates the total amount of water (liquid, ice and vapour) in the atmospheric column.

As can be seen from Fig. 6.6, after the first 5 inputs, performance improvements from further input variables starts to level off. It is nonetheless interesting that not only one but two parameters measuring humidity above the boundary layer or the level of the initial parcel were selected next: mean relative humidity (MRH) between 800 and 600 hPa, and MRH in the active cloud-bearing layer (100 hPa -thick layer above the LFC).

SLI and VWS$_{subLFC}$ as the 8th and 9th input variables decreased validation CEE more incrementally; indeed, using only the 7 first inputs would have been a suitable choice for the final network if simplicity was desired, as the CEE at this point was only 4% higher than the final CEE with 11 inputs.

Finally, VWS$_{subLFC}$ and CIN$_{875}$ represent the only variables in the network mea-
suring wind shear and inhibition, respectively. However, CIN$_{875}$ is not a pure measure of inhibition as it defined as CIN-CAPE between the surface and 875 hPa, and its effect on network performance was small. Therefore, convective inhibition measures in this reanalysis-based experiment turned out to be ineffective forecast parameters.

The neural network to be developed with and without VIMC$_{500}$ as the final input were named NN1 and NN2, respectively. When the number of hidden neurons was varied, it was found that the best results were generally obtained with H at or near the number of inputs; significantly higher than in Manzato (2005). For the final network with 11 inputs, H = 12 led to the lowest CEE. A quick test where the network was retrained with a second hidden layer with varying number of hidden neurons revealed that increasing the model complexity further with a second hidden layer generally decreased performance significantly. However, using a second hidden layer with 3 hidden neurons did lead to an incremental decrease in CEE (CEE = 0.138), so this more complex network represents NN1 while NN2 was kept with one hidden layer. Thus, the final good-as-possible neural networks whose dichotomous forecast skill is compared to stability indices are:

- NN1 (with VIMC$_{500}$): 11 inputs, two hidden layers with 12 and 3 hidden neurons, respectively. CEE = 0.138.
NN2 (without VIMC): 10 inputs, one hidden layer with 10 hidden neurons. CEE = 0.141.

The results from the skill score test are presented in Table 6.7. According to maximum TSS, the ANN’s have significantly higher forecasting skill compared to the best stability indices in this study. The two neural networks reached a maximum TSS of 0.68-0.69, compared to 0.626 for the FIN index and 0.613 for LI$_{mu}$. Thus, even without utilizing convergence, the artificial neural network was successfully trained to be able to make meaningfully better yes/no predictions for thunderstorm occurrence, from a range of parameters, compared to any single index. In a maximum HSS test, the ANN’s lead the stability indices more noticeably, reflecting the emphasis of HSS on FAR: HSS = 0.461 and 0.453 for the two networks, compared to 0.346 with LI$_{mu}$. A very significant benefit of using the neural network is that it’s conveniently a nearly perfect tool for assessing the thunderstorm probability; its output gives the TCP directly and reaches a maximum of over almost 0.9 with a large window size for the moving average (N = 3000), and nearly 0.95 with N = 500 (Figure 6.14).

<table>
<thead>
<tr>
<th>Index</th>
<th>Threshold</th>
<th>TSS</th>
<th>HSS</th>
<th>POD</th>
<th>FAR</th>
<th>CSI</th>
<th>r</th>
<th>r flashes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NN2</td>
<td>&gt; 0.05</td>
<td>0.693</td>
<td>0.300</td>
<td>0.886</td>
<td>0.768</td>
<td>0.223</td>
<td>0.543</td>
</tr>
<tr>
<td>2</td>
<td>NN1</td>
<td>&gt; 0.06</td>
<td>0.687</td>
<td>0.306</td>
<td>0.867</td>
<td>0.764</td>
<td>0.227</td>
<td>0.533</td>
</tr>
<tr>
<td>3</td>
<td>FIN</td>
<td>&lt; 0.9</td>
<td>0.626</td>
<td>0.250</td>
<td>0.845</td>
<td>0.799</td>
<td>0.193</td>
<td>0.324</td>
</tr>
<tr>
<td>4</td>
<td>LI$_{mu}$</td>
<td>&lt; 0.9</td>
<td>0.613</td>
<td>0.237</td>
<td>0.845</td>
<td>0.807</td>
<td>0.186</td>
<td>0.333</td>
</tr>
<tr>
<td>5</td>
<td>SWISS12</td>
<td>&lt; 2.1</td>
<td>0.589</td>
<td>0.218</td>
<td>0.838</td>
<td>0.820</td>
<td>0.175</td>
<td>0.314</td>
</tr>
</tbody>
</table>

Table 6.7 Dichotomous forecast skill of the neural network compared to the best-performing indices according to TSS and the THUN2 criteria, utilizing only the independent "testing"subset (1/5 of data set).

TCP as a function of the neural network output for NN2 and VIMC$_{500}$ (Fig. 6.15) similarly reveals extremely high maximum thunderstorm probabilities. In cases where the NN1 output suggests extremely favourable conditions for thunderstorms (NN1 > 0.9), observed thunderstorm probabilities can exceed 90% and for a particular bin where this was combined with strong low-to-mid level convergence, TCP exceeded 95%. Compared to LI$_{mu}$, the bin counts for the extremely thundery bins are also higher, with the combined number of cases for bins where TCP $\geq$ 90% (shown with lighter text color) at 1800, representing 1.84% of all thundery cases. Although these are encouraging results, it is important to remember that the TCP calculation includes all the model soundings, including those that were used to train the network, and so the results might be slightly less impressive for new data.
Finally, it is interesting to note that although the input for NN2 did not include convergence parameters, $\text{VIMC}_{500}$ seems to have only a weak effect on TCP for cases of very thundery ANN output, whereas for $\text{LI}_{\text{nu}}$ (Figure 6.10), a strong dependence of TCP on convergence was found also for the very unstable soundings. When a similar 2D TCP-plot was made with NN1, almost no gradient could be seen along the x-axis, reflecting the fact that the network had learned to fully utilize the convergence parameter.
Figure 6.15  TCP as a function of the output from NN2 (without convergence) and VIMC (500 hPa integration depth) using the THUN2 criteria. TCP is shown only for bins which have at least 80 cases.
7 Discussion

7.1 Stability indices

7.1.1 Which temporal criteria for proximity soundings?

It was found that skill scores were higher across the board when using the THUN1 criteria, where events take place within 3 hours of soundings, compared to THUN2, where soundings precede events by 0 - 6 hours. The increased forecast skill was associated mostly with a higher POD. THUN1 leading to higher skill scores is not entirely surprising, since it is intuitively associated with better representativeness of soundings. The drawback of using THUN2 is that the atmosphere can change quite significantly in 6 hours; caused e.g. by diurnal heating and frontal passages, whereas the importance of model-calculated indices having to strictly sample preconvective conditions is up for debate. When using radiosonde data, it is admittedly important that the soundings are representative of the environment in which the storm formed, instead of a convective cloud. While it is hard to assess to which extent "contamination" of soundings due to destabilization is relevant for reanalysis-based pseudo-soundings; at any rate, it should be less of an issue than when physical soundings are used, given that reanalyses represent area estimates.

7.1.2 Which detection radius for thundery classification?

Both the THUN1 and THUN2 criteria imply that only a pseudo-sounding that is spatially closest to a lightning observation is associated with a thunderstorm. Since the grid area is relatively small (2100 - 3100 km$^2$), this begs the question, is it appropriate to label soundings in grid boxes adjacent to a grid where a flash was located as "non-thundery"? For example, in Haklander (2003), the thundery classification was based on at least one flash being located within a 100-km distance from the sounding station, which is much longer distance than the one we have used (the radius corresponding to the mean grid area is roughly 30 km).

To gain some insight into the effect of this distance on skill score values, a further thundery classification was done, whereupon a sounding in a given grid point is labelled as thundery if at least 1 lightning flash located within an area which includes the adjacent grid boxes. This corresponds to a tripling of the detection radius.

With the THUN1 variant of this classification (lightning observed within 3 hours of the pseudo-sounding), the maximum TSS for LI$_{mu}$ now becomes 0.60 at the op-
timal threshold of $LI_{\mu} < 1.3 \, ^\circ C$, compared to $TSS = 0.646$ from before. The maximum $HSS$ is 0.46 at the optimal threshold of $LI_{\mu} < 0 \, ^\circ C$, compared to $HSS = 0.37$ at the threshold of $< 1.1 \, ^\circ C$ using the normal detection radius. Therefore, increasing the lightning detection radius decreased $TSS$ while increasing $HSS$. This was caused by a significant decrease in FAR for a given threshold value; however, at the optimal threshold according to $TSS$ the $POD$ also decreased from 0.876 to 0.819. As expected, the optimal thresholds changed towards less thundery index values.

The original criteria can generally be considered superior due to a more strict proximity criteria being filled; soundings should be more representative for the environment in which the storm formed. It should also lead to more accurate optimal threshold values for forecasters to use. The decrease in false alarm ratio when the detection radius is increased is expected, since using the original classification will lead to a large number of cases where a thunderstorm is forecast due to a thundery index value but then occurs outside the small radius corresponding to the reanalysis grid size, and is therefore counted as a false alarm. It is important to consider that skill score measures are sensitive to methodologies, including event classification, in these kind of studies.

7.1.3 TSS or HSS?

The emphasis of $TSS$ on POD and of $HSS$ on FAR is seen as a clear separation in the index values where $TSS$ and $HSS$ reach their respective maxima. This is especially striking e.g. for "lightning-CAPE" based on the most unstable parcel, which has an optimal threshold of $> 155 \, J/kg$ based on $HSS$, but just $5.4 \, J/kg$ based on $TSS$. The question of an appropriate balance between detection and over-forecasting of events is difficult since it depends on the valuation of the cost of a missed event, and is therefore not strictly a meteorological issue. In any case, the choice of $TSS$ and $HSS$ did not affect the relative ranking of indices drastically, which indicates that skill scores are useful to assess the forecast skill of thunderstorm predictors.

7.1.4 Spatial variability

It was found that TCP for a given level of instability is higher for Northern Finland. This suggests that convective initiation is more likely in the north when the atmosphere is unstable, which would also explain the higher predictability of thunderstorms in the north based on higher maximum $TSS$. To reaffirm that this is
the case and to explain the reason, further investigation would be required. However, one possible explanation is that thunderstorms in northern latitudes are more likely to be associated with baroclinity and other large scale disturbances whereas in southern latitudes thunderstorms occur more often without a large-scale forcing and is therefore a more stochastic phenomena in these regions (M. Bister, personal communication, September 2014). Finnish Lapland is also possibly more prone to orographic lift due to it being more mountainous.

7.1.5 CAPE

The forecast skill of CAPE has been shown here and in earlier studies to be no better or worse than some "simple"indices such as Lifted Index for forecasting thunderstorm occurrence. This is interesting, because CAPE, as an integrated quantity, is not sensitive to conditions at a specific pressure level (other than that of the parcel), and could therefore be expected to capture atmospheric instability in a more robust manner than non-integrated indices. Specifically, CAPE is a physical requirement for DMC whereas an index such as the Lifted Index can be positive (e.g. the environment is warmer than the parcel at 500 hPa) even if buoyancy is present with respect to other levels than 500 hPa.

Blanchard (1998) discussed the vertical aspects of CAPE and noted that soundings with similar CAPE but different aspect ratios (i.e. whether the profile is "tall and thin" or "short and wide") can exhibit a large range of instability. To address this, Blanchard introduced the normalized CAPE parameter, by dividing CAPE with the depth of the free convective layer (FCL = EL - LFC). While NCAPE (Appendix A) has the benefit of representing the average buoyancy (and vertical velocities) of the FCL, is has largely been shown to perform no better or worse than CAPE as a "yes/no" thunderstorm predictor. In this study, lightning-CAPE, which is the slice of total CAPE above the freezing layer, reached slightly higher maximum skill score values than normal CAPE. Curiously, CAPE$_{500}$ which integrates the buoyancy only up to 500 hPa and thus should differ distinctly from lightning-CAPE, performed just slightly worse than normal CAPE.

Given that the contribution of buoyancy in different levels to the total calculated CAPE in its integration might not reflect a differing physical importance of buoyancy in different layers for DMC, there might be potential for developing the CAPE parameter further, e.g. into variants for predicting specific convective phenomena such as tornadoes, which are known to be associated with the presence of low-level
CAPE especially (Davies, 2004). However, rather than isolating CAPE in a given layer, one idea would be to add a weight to the CAPE formula which depends on height. Obviously, how to choose the weight is problematic. In this work, a quick experiment was done where an arbitrary height-dependent weight which emphasized middle levels (400-500 hPa) was applied to the calculation of CAPE. This led to a very marginally higher TSS score compared to normal CAPE.

7.2 Convective initiation

Although the primary objective of this work was the statistical evaluation of various stability indices, the methodology and particularly the wide range of parameters readily available in the ERA-I reanalysis allowed for investigating the use of other model-provided parameters as thunderstorm predictors. This can be considered a rudimentary study of convective initiation (CI), for which particularly convergence is well-established to be extremely relevant, and thus we calculated thundery case probability (TCP) as a function of vertically integrated horizontal mass/moisture flux convergence and a stability index. However, with these results and those from the ANN experiment, it is crucial to consider that CI encompasses the complex interplay between processes on a wide range of scales. What we have done is linked output from an atmospheric reanalysis to lightning flash observations. What becomes very important is the details of that reanalysis, particularly the horizontal and temporal resolution - since CI and convergence are very much dependent on scale. The time and location where convection is initiated is largely determined by smaller-scale variability in thermodynamic and kinematic environments not expected to be captured by the reanalysis and is therefore not a question looked into here. However, in model simulations by Birch et al. (2014), CI almost always occurred within areas of local-scale convergence (defined as 60 by 60 km). Therefore, although convergence maximas on smaller scales often define the exact location of convective initiation, convergence calculated on larger scales can also have substantial forecasting value.

7.2.1 Moisture and mass flux convergence

Vertically integrated horizontal moisture flux convergence (VIMFC) was shown earlier by Zomeren and van Delden (2007) to be highly useful as a thunderstorm predictor when combined with a stability index. Here we have reproduced the findings of
Zomeren and van Delden (2007), who used ECMWF analyses of similar horizontal resolution. The authors reached higher maximum TCP values but this should be largely attributable to the much-lower minimum bin count requirement. Using mass instead of moisture flux convergence led to fairly similar results, which is expected since MFC is strongly correlated with mass convergence. Since VIMFC is a poor principal thunderstorm predictor (Zomeren and van Delden, 2007), and VIMC performs at least as well as VIMFC as a supplementary thunderstorm predictor, it is easy to see why Banacos and Schultz (2005) spoke in favour of using mass convergence, since this is more consistent with ingredients-based forecasting. Furthermore, in the ANN experiment, VIMC was found to be a more effective input than VIMFC after humidity parameters had already been included.

Another aspect of Banacos and Schultz’ paper which the results presented here are congruous with are the conceptual models of subcloud horizontal convergence as it relates to convection (Fig. 3.2.1). Both cases where surface convergence maximum is associated with shallow convection, and where DMC is associated with convergence maximum rooted above the PBL, may explain why best results with TCP were obtained when convergence was calculated from a deeper layer.

7.2.2 Vertical wind shear

Many authors have suggested that vertical wind shear (VWS) can have the opposite effect on convective initiation depending on the layer of the shear. Specifically, shear in the boundary-layer should be conducive to initiation of convection, while strong shear above the cloud layer is associated with increased entrainment, which is unfavourable for convective development. TCP calculated as a function of different shear parameters seems to support these ideas. The $VWS_{subLFC}$ parameter (shear between 10m AGL and the LFC) specifically showed clearly increasing TCP for large shear values, while TCP decreased almost linearly with increasing $VWS_{3-6}$. Although sub-LFC wind shear reached higher maximum TCP values than $VWS_{0-3}$, only the latter parameter was successfully applied in the new FIN index. At the very least, $VWS_{0-3}$ should be more easily and robustly calculable, as LFC depends on the choice of the parcel, and there is often no LFC to be found. Here, we used LFC of the most unstable parcel, and 875 hPa instead of LFC if no LFC existed (corresponding to the mean $LFC_{\mu}$).

As is often the case, inferring causality is problematic, since there are other processes at play as well. For example, low-level shear is weakly positively correlated
with low-level convergence ($r = 0.123$ between VWS$_{0.3}$ and VIMC$_{300}$), which could explain some of the association with TCP.

### 7.3 Neural Networks

The neural network experiment led to quite complex final networks: networks with 10-11 inputs and 10-12 hidden neurons. The fact that the networks were able to add more complexity while still being able to generalize well, compared to Manzato (2005) who ended up with 7 variables and 4 hidden neurons, is likely explained by the much larger data set. The drawback was that the input selection script was very taxing computationally. The `trainseg` network training method in Matlab was used due to it being able to utilize GPU (graphics processing unit) acceleration - something which has facilitated the success of deep neural networks in recent years.

Regarding the final set of inputs, it was surprising to see that variables representing convective inhibition were revealed to be ineffective inputs by the algorithm: the only inhibition measure was CIN$_{875}$ (Appendix B), which includes CAPE as well as CIN and was the last input to be added. Thus, the network learned to forecast thunderstorms using measures of *buoyancy* (LI$_{mir}$, SI$_{850}$ and SLI), non-physical parameters which presumably help classification of storm type or represent the persistence effect (synoptic hour HH and previous activity, respectively), *three* measures of moisture beyond the initial parcel level (TCW, MRH$_{800-600}$ and MRH$_{acbl}$) and thus quantities related to how prone the parcel is for *dilution*; and finally, three parameters related to convective initiation or inhibition (VWS$_{subLFC}$, CIN$_{875}$ and VIMC$_{500}$; of which the latter was the most important). This contrasted with Manzato, who otherwise had similar results but ended up with a convective inhibition measure (CAP) as the third input variable.

Many of these parameters can be considered unorthodox thunderstorm predictors. The fact that three measures of moisture beyond the initial parcel level were selected, suggests that dilution can often be important to consider, although it is disregarded in simple parcel theory. It is hypothesized that in Finland, where CAPE values are often low, dilution by mixing of dry environmental air can often be the cause of DMC ultimately failing to develop.

In essence, the results demonstrate how latent instability and moisture are both crucial for convective processes, which take place in a four-dimensional domain, and for which the thermodynamic favourability cannot possibly be condensed into any
single scalar number. This is of course represented in actual forecasting work, where forecasters rarely use just a single parameter to assess the likelihood for DMC. Furthermore, the following results suggest that artificial neural networks may be useful in an operational context to forecast thunderstorms: i) higher skill score values were attained with the ANN’s compared to any single stability index, ii) the ANN’s were clearly superior to indices for making a probabilistic forecast. We believe that the latter finding is more noteworthy. Although the maximum TSS of the ANN were meaningfully higher than the best index LI\textsubscript{nu}, overall the difference in performance was neither huge or unexpected given that the network utilizes additional information (including moisture and even lift) that, if considered by a human forecaster, would probably result in an even better forecast. In other words, the jump from TSS = 0.61 with LI\textsubscript{nu} to TSS = 0.69 with the ANN is not that large considering the amount of information added with the other 9-10 inputs (albeit, LI\textsubscript{nu} did correlate strongly with all but a handful of the other inputs). However, the finding that the ANN represented a near perfect tool for making a probabilistic forecast (in terms of reaching maximum TCP of close to unity) can be deemed significant. Not only did the ANN outperform indices in this regard by a large margin, a point can also be made about the usefulness of parameters for making probability forecasts compared to yes/no forecasts in general. The fact that severe weather probabilities increase in a gradual manner as a function of indices suggests that using "threshold" values in forecasting is not ideal (Púčík et al., 2013).

Finally, the reanalysis aspect is once again interesting: it is entirely possible that an ANN approach would be even more fruitful given a higher-resolution model, which would capture complex convective initiation processes better.

7.4 Generalizations and uncertainty

To gain a little insight into the reliability of our data and viability of our assumptions, we may calculate how many thundery cases were associated with suspicious i.e. non-thundery index values. Approximately 3.8% of thundery cases (using the THUN1 criteria) in the data set had zero or near-zero MUCAPE (< 5 J/kg), with the majority of said cases having zero MUCAPE. This corresponds to the absence of an LFC with respect to a most unstable parcel (parcel with the highest $\theta_e$ in the lowest 300 hPa). Luckily, this number is fairly low; however, so is the threshold of < 5 J/kg. The proportion of thundery cases with less than 50 J/kg MUCAPE is 10.5%.
There are many reasons why thundery cases could have been associated with zero or near-zero MUCAPE. These are: 1) the model sounding was unrepresentative or inaccurate; 2) the thundery classification was erroneous; or 4) positive-CAPE parcels originated at a really high level, above the lowest 300 hPa. Modifying the thundery criteria to at least 2 flashes located in the grid area, the proportion of cases with < 5 J/kg CAPE drops from 3.8% to 2.5%.

As described in Chapter 2, the lightning location system should not falsely detect virtually any flashes, however the location error can in some cases be significant, in which case lightning could have in reality have occurred in an adjacent grid box. Because the proportion of thundery cases with zero CAPE still stayed relatively high after modifying the flash criteria, the main issue is likely to be representativeness or inadequacies of the pseudo-sounding.

7.4.1 ERA-Interim

Using a reanalysis to construct pseudo-proximity soundings in this kind of work has the major advantage of making it possible to use a much greater number of soundings than would be available if one used traditional rawinsonde soundings. Moreover, it allows for studying the use of parameters which cannot be calculated from rawinsonde soundings. However, the approach does not come without its drawbacks. Basing the analysis on a reanalysis is likely to inject some unique sources of uncertainty into the results. For example, reanalyses have been known to have problems with fields associated with strong vertical gradients (Allen and Karoly, 2014), which the authors identify as an issue specifically for thermal or capping inversions, and is thus relevant for convective initiation. If ERA-I also suffers from poor representation of these kind of environments, it could actually be one of the reasons why CIN parameters were unsuccessful as thunderstorm predictors in this work, based on the ANN experiment.

Verification against rawinsonde profiles was not done here, but has luckily been performed in the case of ERA-I by a few authors. ERA-I represents a state-of-the-science reanalysis, and in general, is able to produce pseudo-soundings consistent with verifying soundings (Bao and Zhang, 2013), with significantly smaller differences between the analyses and soundings compared to earlier reanalyses. In Allen and Karoly (2014), ERA-I produced MLCAPE profiles with only a small positive bias in Australia compared to rawinsondes.
Finally, ECMWF models have been known to have problems with the diurnal cycle of convection not being accurately captured, i.e. occurring several hours too early. In this work, it could present a problem for the interpretation of convergence as a thunderstorm predictor. To ensure that the results were not contaminated by a too-early model convection (which in turn causes an erroneous convergence), TCP in two-dimensional space was recalculated with filtering out thundery events in the first 3 hours following the pseudo-sounding. Luckily, the effect this had on the TCP values was not very high: maximum TCP values were decreased by about 5%, which may already be explained by the decreased representativeness from the time lag. According to P. Bechtold (personal communication to M. Bister, August 9, 2014), the model convection does occur too early in ERA-I, at around local noon, but model convergence caused by convection actually lags the true convergence.
8 Conclusions

An extensive stability index comparison based on ERA-Interim pseudo-soundings and lightning location data from Finland between 2002 and 2013 has been carried out. The best index overall in a yes/no forecasting scheme was the Lifted Index using a most unstable parcel. Given that this index out-performed indices using other parcels even for 12 UTC soundings, the use of the most unstable parcel is generally recommended in Finland. The mixed-layer parcel, which takes entrainment indirectly into account, did not perform very well here. This suggests that quantifying how prone the updraft is to mixing in some other way is probably better. For example, the mean relative humidity in the active cloud-bearing layer was an effective input in the neural network experiment. Regarding these results and others, we have in general a high confidence in the ERA-Interim pseudo-soundings (based on a few earlier studies), although they were not verified against rawinsonde soundings in this work and thus it cannot be ruled out that the reanalysis may involve some biases which make e.g. comparison of parcels difficult.

From the ERA-I convergence fields, we calculated vertically integrated mass and moisture flux convergences and examined their use as a supplemental forecast parameter. It was found that the probability for convective initiation increases both with increasing mass (and moisture) convergence in the low-to-mid troposphere and increasing surface-based instability so that the more unstable the environment, the larger the effect of increasing convergence on thundery case probability. This is an intuitive result, since very unstable environments are often associated with a strong cap, which increases the importance of a lifting mechanism to overcome CIN. Since lift and inhibition are paired concepts, and it is advantageous to treat them as such when forecasting DMC. Regarding the question of whether to use moisture or simple mass convergence, the results presented here speak in favour of using mass convergence - consistent with the ingredients-based forecasting scheme. Finally, considering a deeper layer instead of just the surface convergence may be advantageous.

Vertical wind shear is meaningful not only for the mode of convection, as is well established, but also for its initiation. In this study, parameters related to near-surface wind shear and shear in the cloud-bearing layers were shown to have opposite effect on thundery case probability, and seemed to slightly increase the forecast skill of thunderstorms when added to LI$_{\text{mu}}$ in the formulation of a newly
developed index for predicting thunderstorms in Finland (the FIN index). FIN, which additionally incorporates the dew-point depression at 700 hPa, was shown to perform slightly better as a thunderstorm predictor than the most unstable Lifted Index according to skill score tests.

Finally, a neural network experiment was carried out to investigate if an ANN can be trained to forecast thunderstorm occurrence more skillfully than any single stability index, and to also gain insight into which parameters are relevant for DMC in Finland. The input-selection algorithm picked multiple stability indices as effective ANN inputs, which underscores that it is always important to consider more than just one in convective forecasting, since indices are simple and imperfect representations of instability. Based on the skill score test, the ANNs were superior to any single index for making dichotomous thunderstorm forecasts. Moreover, the ANN is excellent for making a probability forecast, given the the ANN output gives the thunderstorm probability directly and reaches a maximum of over 0.9. The neural networks clearly outperformed indices in this regard, and thus the operational use of model-tuned ANN’s as probabilistic forecasting tools warrants further study.
References


Appendix A

Definition of thunderstorm indices

Boyden Index,

\[ BI = 0.1(Z_{700} - Z_{1000}) - T_{700} - 200 \]

where \( Z \) is the geopotential height. The Boyden Index (Boyden, 1963) does not take moisture into account at all.

Convective Available Potential Energy (CAPE),

\[ CAPE_i = R_d \int_{LFC}^{EL} (T'v - Tv) \, d(ln \, p), \]

where only positive contributions are included (parcel virtual temperature must exceed the environmental virtual temperature, \( T'_v > Tv \)). CAPE (Section 2.2.2.) is a measure of the total potential energy available for convection. Different definitions of CAPE exist in literature but here we have used the above notation and the following parcels: the surface parcel (CAPE\_sfc), most-unstable (MUCAPE, using the parcel with the highest \( \theta_e \) in the lowest 300 hPa from the surface), and a mixed-layer parcel (CAPE\_50, using a parcel with the average temperature, humidity and pressure in the layer consisting the lowest 50 hPa). For MUCAPE, two further variations were also calculated: MUCAPE\_10...-40 \( ^\circ C \) (integrating from -10 \( ^\circ C \) to -40 \( ^\circ C \)), and MUCAPE\_500 (integrating from the LFC to 500 hPa).

Deep Convective Index,

\[ DCI = T_{850} + T_{d850} - SLI. \]

Developed by Barlow (1993) to assess the potential for deep convection.

Energy Index,

\[ EI = \theta_{e500} - \theta_{e850} \]

Jefferson Index (JI and JI\_mod): see Eq. 3.3 and 3.4.
**K Index,**

\[
KI = (T_{850} - T_{500}) + T_d850 - (T_{700} - T_d700)
\]

KI was developed by George (1960) for forecasting air mass thunderstorms. KI increases with increasing static stability between 500 and 850 hPa and increasing moisture at 700 and 850 hPa.

**KO Index,**

\[
KO = 0.5(\theta_{e500} + \theta_{e700}) - 0.5(\theta_{e850} + \theta_{e1000})
\]

KO (Andresson et al., 1989) measures low and mid-level potential instability.

**Lifted Index:** see Eq. 3.2.

**Normalized CAPE,**

\[
NCAPE = CAPE/FCL,
\]

where FCL is the depth of the free convective layer; \( FCL = z_{EL} - z_{LFC} \). NCAPE was developed by Blanchard (1998), who studied the vertical distribution of CAPE. Because CAPE is sensitive not only to buoyancy but to the depth of the integration, it is not a measure of buoyancy as such. A short and wide CAPE profile can lead to much stronger convection than an environment characterized by a tall and thin CAPE, even if they have equal CAPE. To account for impact of the depth of the free convective layer, Blanchard (1998) normalized CAPE by the integration depth, which should result in a better measure of instability. The advantage of NCAPE being more strictly a measure of updraft intensity is reflected in that its unit is meters per second squared, an acceleration (Blanchard, 1998). In this work, the most unstable parcel was used to calculate NCAPE (\( = NCAPE_{mu} \)).

**Potential wet bulb Index or Potential stability Index,**

\[
PI = \theta_{w500} - \theta_{w850}
\]

**Rackliff Index,**

\[
RI = \theta_{w900} - T_{500}
\]

RI was developed by Rackliff (1962) to help forecast air-mass thunderstorms in Great Britain. This index was later developed further into the Jefferson index by

Showalter Index: see Eq. 3.1. Note that LI and SI are almost synonymous: they only differ in which parcel is lifted to 500 hPa. Usually SI refers to specifically to the index calculated using a 850 hPa parcel, so that $SI = SI_{850}$. In this work, the following parcels were used to calculate Showalter Index: $SI_{850}$ (850 hPa parcel), $SI_{LCL}$ (LCL-based parcel, where the LCL is calculated from surface conditions) and $SI_{LCL_{mod}}$, where the LCL is calculated using the mean conditions in the lowest 50 hPa.

SWISS indices,

$$SWISS_{00} = SI_{850} + 0.4 \cdot VW_{S_{3-6}} + 0.1(T_{600} - T_d_{600})$$
$$SWISS_{12} = SLI - 0.3 \cdot VW_{S_{0-3}} + 0.3(T_{650} - T_d_{650})$$

where $VW_{S_{3-6}}$ and $VW_{S_{0-3}}$ is the vertical wind shear between 3 and 6 km above ground level (AGL), and the wind shear between 10 m and 3 km AGL, respectively. In this case, the vertical wind shear is simply defined as the difference in the magnitude of the two wind vectors, so that its unit is [m s$^{-1}$] (or m s$^{-1}$(3 km)$^{-1}$ as per the original reference). The SWISS$_{00}$ and SWISS$_{12}$ indices were developed by Huntrieser et al. (1997) for forecasting subsequent thunderstorms from 00 UTC and 12 UTC soundings, respectively. The SWISS indices decrease with increasing thunderstorm probability. Weak shear between 3 and 6 km AGL, strong shear between 10 m and 3 km AGL, and small dew-point depressions at 600 and 650 hPa should be favourable for thunderstorm initiation. Although Huntrieser used 3 and 6 km AGL for the wind terms, for convenience we used the 700 hPa and 450 hPa levels instead, having calculated these pressure levels to be on average close to 3 and 6 km AGL.

Thompson Index,

$$THOM = KI - LI_{50}.$$ 

A combination of two indices, of which KI neglects latent instability below 850 hPa while $LI_{50}$ measures instability with respect to a near-surface parcel, THOM should
ideally be superior to both indices.

*Total Totals index,*

\[
TT = VT + CT = (T_{850} - T_{500}) + (T_{850} - T_{500}).
\]

A widely used index, TT (Miller, 1967) was originally developed for forecasting thunderstorms in the US (Roine, 2001).

*Yonetani Indices,*

\[
YI = \begin{cases} 
0.966 \cdot \Gamma_L + 2.41(\Gamma_U - \Gamma_W) + 9.66 \cdot \gamma - 15,0 & \text{if } RH > 57 \\
0.966 \cdot \Gamma_L + 2.41(\Gamma_U - \Gamma_W) + 9.66 \cdot \gamma - 16,5 & \text{if } RH \leq 57
\end{cases}
\]

\[
YI_{mod} = \begin{cases} 
0.964 \cdot \Gamma_L + 2.46(\Gamma_U - \Gamma_W) + 9.64 \cdot \gamma - 15,0 & \text{if } RH > 50 \\
0.964 \cdot \Gamma_L + 2.46(\Gamma_U - \Gamma_W) + 9.64 \cdot \gamma - 16,5 & \text{if } RH \leq 50
\end{cases}
\]

where \(\Gamma_L\) and \(\Gamma_U\) are the lapse rates in the 900-850 and 850-500 hPa layers, \(\Gamma_W\) is the moist-adiabatic lapse rate at 850 hPa, RH is the relative humidity of the 900-850 hPa layer and \(\gamma = RH/100\). YI and \(YI_{mod}\) (Jacovides and Yonetani, 1990) assess the potential for thunderstorms on the basis of low-level humidity and latent instability in the cloud-bearing layers being present.
# Appendix B

## Neural network input pool

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTP</td>
<td>Activity in the previous case: 1, if lightning was detected in the previous 6-hour period, 0 otherwise.</td>
</tr>
<tr>
<td>CAPE[LCL]</td>
<td>CAPE-CIN between LCL (mixed-layer parcel [lowest 50 hPa]) and 200 hPa above this.</td>
</tr>
<tr>
<td>CAPE[sfc]</td>
<td>CAPE using a surface-based parcel.</td>
</tr>
<tr>
<td>CAPE[50]</td>
<td>CAPE using a mixed-layer parcel (mean conditions in the lowest 50 hPa)</td>
</tr>
<tr>
<td>CAP</td>
<td>Strength of &quot;cap&quot; (convective inhibition measure). ( \text{CAP} = \theta_{es,100} - \theta_{es,max} ), where ( \theta_{es,100} ) and ( \theta_{es,max} ) are the maximum saturated equivalent potential temperature in the lowest 100 hPa and 500 hPa, respectively.</td>
</tr>
<tr>
<td>CIN[sfc]</td>
<td>CIN using a surface-based parcel.</td>
</tr>
<tr>
<td>CIN[875]</td>
<td>CIN-CAPE between surface and 875 hPa.</td>
</tr>
<tr>
<td>ELR[acbl]</td>
<td>Environmental Lapse Rate in the active cloud-bearing layer (100-hPa thick layer above LFC_{mu}, or 875 hPa, if no LFC exists)</td>
</tr>
<tr>
<td>HH</td>
<td>Hours UTC of sounding (00/06/12/18)</td>
</tr>
<tr>
<td>JI</td>
<td></td>
</tr>
<tr>
<td>KI</td>
<td></td>
</tr>
<tr>
<td>L[limu]</td>
<td></td>
</tr>
<tr>
<td>MRD</td>
<td>Integrated mixing ratio difference between the parcel level (MU) and 100 hPa above this.</td>
</tr>
<tr>
<td>MRH[acbl]</td>
<td>Mean Relative Humidity (%) in the ACBL.</td>
</tr>
<tr>
<td>MRH[1000-500]</td>
<td>Mean Relative Humidity (%) between 1000 and 500 hPa. Arithmetic mean of the column, but with emphasis on low-level humidity, as the vertical resolution is double (25 hPa) below 750 hPa.</td>
</tr>
<tr>
<td>MRH[800-600]</td>
<td>Mean Relative Humidity (%) between 800 and 600 hPa.</td>
</tr>
<tr>
<td>MRH[800-250]</td>
<td>Mean Relative Humidity (%) between 800 and 250 hPa.</td>
</tr>
<tr>
<td>MUCAPE</td>
<td>CAPE using the most unstable parcel in the lowest 300 hPa.</td>
</tr>
<tr>
<td>MUCAPE[-10...40 °C]</td>
<td>CAPE in the layer from -10 to -40 °C.</td>
</tr>
<tr>
<td>NCAPE[Mu]</td>
<td>Normalized MUCAPE.</td>
</tr>
<tr>
<td>PI</td>
<td></td>
</tr>
<tr>
<td>SI[850]</td>
<td></td>
</tr>
<tr>
<td>SIL[CLmod]</td>
<td></td>
</tr>
<tr>
<td>SLI</td>
<td></td>
</tr>
<tr>
<td>TCW</td>
<td>Total Column Water (ERA-I parameter).</td>
</tr>
<tr>
<td>THOM</td>
<td></td>
</tr>
<tr>
<td>VIMFCT</td>
<td>Vertically Integrated Moisture Flux Convergence, using integration depths of 300, 500 and 700 hPa.</td>
</tr>
</tbody>
</table>

B-1
<table>
<thead>
<tr>
<th>VIMC</th>
<th>Vertically Integrated Mass Flux Convergence, using integration depths of 300, 500 and 700 hPa.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VWS$_{acbl}$</td>
<td>Vertical Wind Shear [m s$^{-1}$] in the ACBL.</td>
</tr>
<tr>
<td>VWS$_{subLFC}$</td>
<td>VWS below the LFC (MU).</td>
</tr>
<tr>
<td>VWS$_{0-3}$</td>
<td>VWS between 10 m and 3 km (here 700 hPa) above ground.</td>
</tr>
<tr>
<td>VWS$_{3-6}$</td>
<td>VWS between 3 km and 6 km (here 450 hPa) above ground.</td>
</tr>
<tr>
<td>WVFnorth</td>
<td>Vertical integral of northward water vapour flux (ERA-I parameter).</td>
</tr>
</tbody>
</table>