Swarm Data Mining for the Fine Structure of Thermals

Alfred Ultsch
Databionics Research Group
University of Marburg, Germany
ultsch@ulweb.de

Accurate thermal models can help to optimize the design of sailplanes. Theoretical models should be based on empirical evidence. However, very few measurements on the structure of thermals are published. This paper uses data mining techniques on data collected by swarms. In this case the swarm consists of the world’s best pilots in the world’s best gliders competing in a world championship at Uvalde, Texas in 2012. It is pointed out how the data collected by this swarm in the form of ICG files (i.e., GPS recordings) may be processed in order to yield the vertical speed of the air in thermals. This resulted in about 100 hours of data on thermals. From this data a model of the fine structure of thermals could be derived consisting of three components: a Gaussian representing the buoyancy, a vortex modeling entrainment and a border vortex caused by the difference in speed between the air inside the thermal and the surrounding air.

Introduction

The Diana-2 sailplane was the winner’s sailplane in the 15 meter class of the 2012 World Gliding Championship in Uvalde, Texas. The design of this sailplane has been optimized using a model structure of thermals [1]. However, as Kubrynski puts it: “a very limited amount of measured data is available in the literature, making this problem even more difficult” [1]. This means, only a handful of measurements form the empirical basis of present day’s thermal models. On the other hand, the flight data of most cross country flights and all of the competition flights are measured and logged every second using GPS devices. These in-flight measurements are well documented in the form of IGC-files [2]. This paper reports the methods and results of mining this data with the aim of deriving a thermal model with a much broader empirical basis, i.e., about 100 hours of measured flights spent spiraling in thermals.

Flight Data and Soaring Conditions

In order to obtain comparable flight data the flights in Ventus-2, ASW-27 and ASW-29 (15m) sailplanes in the 15 meter class of the Uvalde competition were used. Only flights that finished with a rank of not more than 25 (of 37 participants) on each day were used. To the best of the author’s knowledge, all pilots launched at maximum takeoff weight resulting in the same (i.e., maximal) wing loading of the sailplanes. This assured that all pilots performed on a very high level of competence and that all sailplanes were configured as best as possible. All flights took place during August 5th to 18th, 2012. Start and finish airport was Garner Airfield with an elevation of 287m (942ft). Elevation of the terrain ranged from 200–700m (700–2300ft). Only flight data after the start of the race were used. The observed thermals were all between 5.2 and 6.9 hours after sunrise in Uvalde. The thermals are localized in an area between 98.5 and 100.6 degrees West and 27.5 to 30.5 North. Weather conditions in the semi-arid climate zone of south western Texas during that period were rather constant with temperatures around 40°C, with 0–3 octas of Cu and a base of 2300–3000m. Wind direction was mostly in the range of 250–350 degrees. Wind speed was Gaussian-distributed with a mean of 19 km/h ±9 km/h standard deviation. The average integrated net climb speed of all thermals followed a Gaussian distribution (N(m,s)) with a mean m of 3.0 m/sec and standard deviation s of 0.81 m/sec. The GPS altitudes of the first and last fixes in the helix (centered spiraling and climbing in circles) were N(1800,360) [m] (first) and N(2200,370) [m] (last). Height gain in these helices ranged from 200 to 1600m. The thermal data collected at Uvalde in this way are referred to as U-thermals in this paper.

Data Mining Methods

Climbs in the IGC files of the flights were identified as successive periods of flight where at least 250m of altitude was gained and the engine noise level indicated normal gliding flight. In these climb periods a circling flight was identified by at least three full circles in the same direction (see Fig. 1).

Wind direction and speed during the spiraling was estimated from the average movement of the centers of these full circles. These wind data were subtracted from the x and y Gauss-Krüger coordinates of the GPS fixes. All fixes were normalized to a one second interval using local spline interpolation for the coordinates and GPS altitude h if necessary.

Periods where the flight path was not turning with at least a
rate of 3°/sec (centering) were excluded. This led to a total helix-shaped flight periods of 56.9 h (Ventus-2), 22.1 h (ASW-27) and 24.8 h (ASW-29). The centers of the helix were estimated using the method of Kasa [3] using successive points indicating a direction change of at least 360°. The distance from this center is used as the momentary radius of the turn. This is denoted as Radius [m] in the following figures. Figure 2 shows the coordinates of the wind corrected helix as part of the spiraling flight shown in Fig. 1. For all helix data of the same sailplane type the turn Radius [m] as well as the successive altitude differences DH [m] were calculated for each successive seconds. The altitude gain was compensated for total-energy. Within a range of 50–290 m, the DH values were binned with respect to the radius in bins of 1m. For Radius > 290m, a bin width of 5m was used. This assured at least 50 data points in each bin. Averaging over these bins resulted in the data of Fig. 3.

From the radius of the turn and the momentary speed of the sailplane the bank angle could be calculated. Using the L/D of the particular aircraft the sink rate in the turn was estimated. Adding this velocity to the sailplane’s vertical speeds led to an estimation of the momentary encountered total vertical speed, see Fig. 4.

From Swarm Data to Vertical Speeds of Thermals

All the data considered above were not collected with the aim to measure the vertical velocity of air in a thermal. This is a typical example of so-called “swarm data.” A swarm of world class pilots in high performance sailplanes were sent out to find the best thermals, center as efficiently as possible and try to find the best possible climbing method. The aim of each member of this swarm is to win the competition. This only can be done by making the best (i.e., most efficient) use of every meteorological situation. Hence our data do not come from carefully designed experiments. However, we can assume that the pilots fly as well as they can. This means the measures of Fig. 4 can be considered to squeeze the maximum out of the meteorology and the maximum out of the sailplane. This gives a hint on how to rescale the data. The vertical speed “far away from the center of the thermal” i.e. for a turn radius > 300m can be associated with a vertical air speed of zero (see the horizontal line in Fig. 4). The rescaled data are presented in Fig. 5.

From 300m inward to 130 m (vertical lines in Fig. 5), the
vertical speed increases from 0 up to a maximum of 4.4 m/sec. A linear model of the data in the range 130–230m results in a gradient of 3.6 m/s per 100m (solid line in Fig. 5). For a turn radius of less than 130m the vertical speed is constant respectively decreasing (dots in Fig. 5). Within these distances from center of the thermal the limiting factor, however, is the performance limit of the aircraft. In order to obtain a smaller turn radius, pilots must turn at a steeper angle. This increases the sink speed, as the well known turn polar shows. This means at these small radii the increased sink due to the bank angle in the turn compensates or exceeds the additional lift of the thermal.

**The Core of the Thermals**

Since all three sailplane types (Ventus-2, ASW-27 and ASW-29) showed the same drop in performance, another type of measurements is necessary to explore the core of thermals. For this the same type of swarm data gathering and analysis was done on paragliders. Paragliders operate in a speed range of 30 to 55 km/h and are able to fly very small turning circles. Martin Serner, a high performance German paraglider pilot, provided us with data from 15 long distance cross country flights. The flights covered distances of more than 200km in the German lowlands in the area between Cottbus and Lüsse. This resulted in a total of 4.5 h of helix flight of paragliders. Fig. 6 shows the results of swarm data analysis in this case. It can be seen that paragliders reach their performance limits at a turn radius of about 30 m. This figure demonstrates that the vertical speed of the air will increase as we get close to the core of the thermal.

In order to extrapolate the data to the core of the thermal we now know, that the vertical speed profile is increasing. The fine structure of Fig. 6 in the range 30-40 m indicates, however, that this extrapolation is not linear.
Figure 7 Linear (dashed) and Gaussian Model (solid line) for the core of thermals at Uvalde.

Under zero wind conditions we can assume that the distribution of vertical speed within a thermal profile is symmetric. As the central limit theorem suggests, a plausible hypothesis of the shape of the profile is a Gaussian. This hypothesis was also used in the thermal models of Carmichael in 1954 [4]. A Gaussian was adapted using least square optimization to the symmetric values in the range 130–290 m, see Fig. 7. Also a linear model was fitted to these data. The strength of the core at Radius = 0 predicted by the Gaussian is 10.15 [m/sec] (std. deviation = 150) and 8 [m/sec] for the linear model. A comparison to the paraglider data of Fig. 6 suggests that these estimates are unrealistically high. In [5] Gerhard Waibel also insisted that most thermals have a flat top and are rather “hat shaped” than pointed. Therefore a model is sought that delivers a center strength of not significantly more than 6 m/s for U-thermals.

We propose here the GTB model, an additive model of the thermal profile consisting of three components G, T and B:

Component G consists of a profile of vertical velocities that are Gauss distributed. This models the buoyancy caused by the lower density of the air inside the thermal.

Component T consists of a vortex modeling the exchange of air between the thermal and the surrounding air (entrainment). For the radial symmetric case this is a vortex torus centered at the middle of the thermal. At Radius = 0 the vortex has its maximal downspeed. As a formula, this airflow describes a sinusoid which is symmetrical to the y-axis

Component B models the effects encountered at the boundary layer of the thermal. We assume a small band in which a vortex is rolling. In the horizontal case this would be equivalent to a Kelvin-Helmholtz instability. Also in this small band a constant air flow is confined which models the global air mass flow vertically along the thermal. In the model this is a sinusoid of one period confined between Border_min and Border_min plus a constant vertical speed \(w_{\text{border}}\)

Boundary conditions for the model are:

- at Radius = 0 the G and T component sum up to the maximum thermal strength.
- for Radius > \(r_{\text{max}}\) the vertical speed is zero
- the border vortex \(B\) ends at \(r_{\text{max}}\)
- the model should fit the U-thermals

In the case that the border conditions are less important the model reduces to the components \(G\) and \(T\) (GT model).

Using least square techniques the components were fitted to the U-thermals. The results were as follows:

- Maximum thermal strength at Radius = 0: 6.0 m/sec, \(r_{\text{max}} = 310\) m.
- Component G the std. deviation of the Gaussian is 170 m
- Component T A sinusoid with wavelength 140 m. The amplitude of the vertical speed of the entrainment vortex torus is 0.8 m/sec. Entrainment vortex and max buoyancy speed sum up to thermal core velocity of 6 m/sec.
- Component B: this vortex has a phase length of 55m and amplitude of 0.5 m/sec. Contained within the boundaries of vortex is a downstream of air at a speed of 0 m/sec.

This border vortex is responsible for the increased sink at a radius of 300m in Fig. 4 and the increased lift at about 260m.

The GTB model and its components are shown in Fig. 8. Complete formulas and parameters of the models are given in the Appendix. The fine dashed line in Fig. 8 is the buoyancy Gaussian. The entrainment vortex is drawn in the lower part of Fig. 8. In the box at the lower right side, the border vortex can be seen.

**GTB Model Applied to Paragliding Data**

It may be the case that the GTB model is over adapted to U-thermals due to the nature of the swarm data collection in the semi-arid weather condition of south-west Texas. On the other extreme of the speed of uplift strengths and speeds is the data from the paragliders in the German Lowlands. The optimal GTB model for this case gives a maximum thermal strength of 3.2 m/sec, \(r_{\text{max}} = 87\) m, the standard deviation of the buoyancy Gaussian is 40m, the entrainment vortex has an amplitude of 0.4 m/sec, and a period of 31m (Fig. 9). The border vortex has a period of 40m with an amplitude of 1.0 m/sec. The border airflow was upward with 0.1m/sec.
Compared to the U-thermals the maximum strength in the core of these thermals is about half and the diameter is about one third. This may well be explained by the meteorological conditions in northern Germany compared with south-western Texas.

The period of the border vortex is also as big as in the U-thermals. However, the amplitude of the border vortex is twice as much as in the U-thermals. This may point to a bias in measuring this vortex using gliders or other airplanes.

Due to their mass and speeds (500+ kg, 100+ km/h) it could be that this vortex is traversed too quickly by aircraft and is not captured to its full extent.

**Comparison with Published Measurements of Thermals**

As pointed out above, very few actual measurements of thermals are published. One of the first is taken from the works of Konovalov [6] cited after Waibel [5]. In Fig. 10 the single core type (Type a) is compared with the GTB model of the U-thermals (gray line).

Gerhard Waibel also pointed out an Idaflieg publication [7] with the measurement of a thermal. The model used there was a Fourier analysis (sum of sinusoids).

In [8] flight data were collected by a specially instrumented Blanik glider flying over Rogers Dry Lake to the north of the Edwards Air Force Base in the Mojave desert in California in September 2006. The GTB model is compared with these data in Fig. 12.
In the measuring flight of C. Lindeman [9] two thermals could be identified, see Fig. 13. The GTB models of the U-thermals were scaled such that the x-axis, i.e. the width of the thermals, are the same.

Figs. 10 to 13 show a fairly good coincidence with the presented model. This is quite surprising since the origin of the measurements is Russia, Germany and California. This may point that the average Uvalde-thermal as presented here (U-thermal) may represent some universal characteristics.

Comparison to Published Thermal Models

One of the first models of the thermal comes from the Carmichael’s publication in 1954 [4]. Compared to our model, these model distributions are too narrow. However, Carmichael derives the central Gaussian from the idea of a thermal jet stream.

According to Kubrynski [1], Horstmann’s [10] models are probably the most realistic approach. They include four standard thermal profiles: combination of strong (2) and weak (1) and wide (B) and narrow (A) thermals, see Fig. 14.

Kubrynski [1] specified three thermal families: A (narrow), B (wide) and C (middle thermal).

Compared to both models the GTB model is much wider. In the GTB model the zone of rising air has a total diameter of about 500m.

It is remarkable that GTM-model of the U-thermals seems to have the same gradient as the A-types of the Horstmann profiles. This may have contributed to the successful optimization of the Diana-2 sailplane described in [1].

“Hat” Type Thermals

Gerhard Waibel triggered the research presented in this paper. In his talk at the OSTIV Congress in Uvalde 2012 he pointed out that our thermal models do not fit to the experience of pilots. He claimed, that the thermals are more of a “hat” type i.e. do have a flat or even impressed core. Gerhard also cited H. W. Grosse and G. Stich, who favor these hat-types of thermals. In the data of Childress [8] two measured thermals appear which fit this hat-type. These can be modeled with GTB using a strong entrainment component $T$. Figures 15 and 16 show a possible fitting to such hat-thermal data from the flight measurements of [8] Figures 3 (page 12) and 7 (page 14).

The GTB model explains the flat top respectively the dip in the core of a thermal as the interference of the entrainment vortex with the buoyancy Gaussian.

Gerhard Waibel also insisted that at the border of the thermal there seems to be sink. His hand drawings of a thermal profile resemble the typical “Mexican hat” function. The sink at the border is explained in GTB by a border vortex caused by friction. This vortex can be observed in the U-thermals as well as in the paraglider data in Germany and also in the actual flight measurement data of Frey (Fig. 11), Lindemann (Fig. 13) and Childress (Figs. 12, 16, and 17),
Discussion

There are extremely few measurement data published on thermals. So the data base for models of thermals is rather poor. However, the design of better gliders in the future calls for models that fit the nature and not what theoreticians think. The amount of data that in principle can be obtained by the swarm data mining method on logged IGC files of flights is enormous. While the published flight data sum up to only some minutes in thermals, the GTB model is derived using about 100 hours of flights flown spiraling in thermals.

However, the data analysis presented here relies heavily on the extraction and preprocessing of the data. As we all know from spectacular events, such as the explosion of the Ariane 5 at first launch, software implementation may be erroneous. So, the best practice would be, that an independent research group repeats the presented data analysis.

The comparison of GTB with the reported data is astonishingly consistent (see Figs. 10 to 13). In comparison to the model profiles (Figs. 17 and 14) the U-thermal data suggest a diameter of the thermal of approximately 600 m including the border vortex. The paraglider data, however, fit to a thermal diameter of approximately 200m. This may be explained by difference in meteorological conditions.

The weather conditions at Uvalde may not represent a typical flying day. In [11] and [12] the author has investigated the strengths of more than 10,000 thermals in August in Bavaria, Germany. The distribution of the thermal strengths follows extremely precisely a squared Gaussian (see Fig. 15).

From this analysis it can be conjectured that the meteorological conditions that produced U-thermals where the pilots found an average integrated lift of 3m/sec occur in less than 10% of all cases (dashed line in Fig. 18). However, most national and international competitions are timed and placed such that these very good conditions are likely to occur.

To explain the general vertical speed within the thermal using a Gaussian stems from the theory of jets, see [13] for a review. Carmichael writes “The shape is pure conjecture but at least the qualitative experience of pilots do not refute the theoretical guidance given by the turbulent free jet.” [4, page 9].

Figure 19 shows the distribution profiles of speeds in a vertical jet with a slight divergence [14]. In [11] it is observed
that the square root of thermal strength is very precisely Gauss-
distributed. As a deeper reason behind this fact, it can be as-
sumed that the radius of the superheated, respectively super-
humid, thermal air bubble in the ground layer is Gauss dis-
tributed.

The G-Component of the GTB model can be thought of as the
“average” of the speed profiles shown in Fig. 19. However, U-
thermal data can hardly be explained by one Gaussian alone. A
single Gaussian would have a too pointed top in the middle and
too heavy tails. Refer to the dashed line in Fig. 7. The addition
of an entrainment vortex compensated for this.

Many pilots report an increased sink rate just “before the ther-
mal begins” The paraglider pilot Martin Serner described even a
feeling as “being drawn into the thermal.” A border vortex could
explain these effects.

One can imagine a series of vertical vortices, similar to the
horizontal Helmholtz-Kelvin types, running on the outer border
of the thermal (see Fig. 20).

The usage of vortices for a model of thermals is not new.
Müller and Kottmeier, for example, compare in [9] thermals
with pipes (constant uplift) and bubbles, i.e. toroid vortices. The
GTB approach may allow quantifying the mixture ratio of both
phenomena.

**Types of Thermals?**

data collected around 1960, to Horstmann 1976 [10], several re-
searchers have suggested that there are different types of ther-
mals: wide vs. narrow, weak vs. strong, pointed top vs. flat top
and specific mixtures of this type. See, for example, Fig. 14 for
such types. For the U-thermals, the average integrated net lift,
which the pilots found, follows a Gaussian N(3.0,0.81) m/sec
data not shown). Fig. 21 shows the distribution of the average
helix radius on all 1211 measured Verrutz-2 helices. These fit
well to a Gaussian distribution with a slight overrepresentation
at about 146 m. At that radius the pilots experienced the best
lift. Most spiraling was done in a range of 130 to 200 m.

In summary, the U-thermal data do not suggest any clusters
of thermals. This can be attributed to the very homogeneous
climatic and orographic conditions of the flights. Furthermore
all measurements are all taken in a very short period of only
2 hours of the best flying time during each day. Furthermore
the swarm pilots are conditioned to spiral in only the strongest
thermals that can be found each day in order to win the world
championship. So the U-thermal measurements are highly se-
lective for only the class of strongest thermals. Other thermals
types, in particular in other meteorological conditions, different
orographic situations (mountains vs. flat country) and differ-
cent time of the day (onset of thermals in the morning; evening
thermals/“Umkehrthermik”) may well exist. In order to cap-
ture these thermal types, data from many cross country flights
in different weather conditions and landscapes would be neces-
sary. However, nowadays the policy to obtain our own flight data back from the OLC database (www.onlinecontest.org) for scientific purposes is unfortunately too restrictive to allow such research. A submission of our flight data to an open source database such as, for example, www.skylines-project.org will help to overcome this problem.

Conclusion
This paper sheds some light on the fine structure of thermals. Up to the present only a handful of inflight measurements of thermal data has been published. This analysis is based on averaging over ca. 100 hours of such flying time where the pilots have already centered the thermal and are climbing in circles (Helix). An enormous amount of flight data are there and — except from the OLC database — freely available. An initial approach was presented here on how to use these data for thermal models that describe what is going on in nature.

The result is a mixture model of buoyancy, entrainment and a vortex on the border. The same model could be applied to strong thermals with 6 m/sec core lift and 600m width in south west Texas down to paraglider flights in Germany’s flatlands with 3 m/sec core lift and 200m width.

As soon as more data is available and the data processing is standardized and error proofed, we can expect that theoretical thermal models can be verified using a solid empirical basis stemming from our “swarm data.”

The vertical fine structure of (U-)thermals remains for further research

Acknowledgement
The author wishes to thank Martin Serner, Krzysztof Kubrynski, Christof Maul and Gerhard Waibel for their helpful discussions and comments on draft versions of this paper and Martin Serner in particular for his flight data and sharing his experiences in thermals.

References

GTB Model
The GTB model gives the vertical speed of air as a function of the distance $r$ from the center of a thermal. It is an additive model consisting of a Gaussian $G$ (buoyancy), a large vortex $T$ (mixture/friction) and a border vortex $B$ (Helmholz/Kelvin).

$$w = M_G e^{-r^2/2S_G^2} + w_T + w_B$$

$w$ [m/sec] vertical speed of air
$r$ [m] distance from center of thermal
$M_G$ [m/sec] mean strength of central thermal Gaussian
$S_G$ [m/sec] standard deviation of central thermal Gaussian
$w_T$ [m/sec] amplitude of (entrainment) vortex
$p_T$ [m/sec] width of (entrainment) vortex (period)
$w_B$ [m/sec] amplitude of border vortex
$p_B$ [m/sec] width of border vortex (period)
$w_B$ [m/sec] constant vertical speed of border vortex
$r_{max}$ [m] limit radius of thermal: if $r > r_{max}$, it follows that $w = 0$

Figure 21 Distribution of the Median of all measured Vantus-2 Helix data, dotted: Gaussian N(160, 24)
The model gives the vertical speed of the air as
\[
  w = G(r, M_G, S_G, w_T, p_T, w_B, p_B, w_0B, r_{\text{max}})
= G(r, M_G, S_G) \\
  + T(r, w_T, d_T) \\
  + B(r, w_B, p_B, w_0B, r_{\text{max}}) \ast 1(r \leq r_{\text{max}})
\]
where component G (central Gaussian) is given by
\[
  G(r, M_G, S_G) = N(r, M_G, S_G)
\]
\(N(r, m, s)\) denotes a Gaussian with mean \(m\) and standard deviation \(s\);
Component T (torus) is given by
\[
  T(r, A, p) = -A \ast \cos(\pi/p \ast r)
\]
and Component B (border vortex) is given by
\[
  B(r, A, p, w_0B, r_{\text{max}}) = -A \ast \sin((2\pi)/(r - r_b) + w_0)
\]
with \(r_b = r_{\text{max}} - p/2\) and \(1(E) \equiv 1\) if expression \(E\) is true, 0 otherwise.