Identification of the honey bee swarming process by analysing the time course of hive vibrations.

Martin Bencsik[1], Joseph Bencsik, Michael Baxter[1], and
Andrei Lucian, Julien Romieu and Mathias Millet.

[1]College of Science,
Nottingham Trent University
Clifton Lane, NG11 8NS Nottingham
United Kingdom

Corresponding author: Martin Bencsik, School of Science and Technology, Nottingham Trent University, Clifton campus, NG11 8NS Nottingham, UK, phone (44)115 848 8057, fax (44) 115 848 66 36, e.mail: martin.bencsik@ntu.ac.uk

ABSTRACT

Honey bees live in groups of approximately 40 000 individuals and go through their reproductive cycle by the swarming process, during which the old queen leaves the nest with numerous workers and drones to form a new colony. In the spring time, many clues can be seen in the hive, which sometimes demonstrate the proximity to swarming, such as the presence of more or less mature queen cells. In spite of this the actual date and time of swarming cannot be predicted accurately, as we still need to better understand this important physiological event. Here we show that, by means of a simple transducer secured to the outside wall of a hive, a set of statistically independent instantaneous vibration signals of honey bees can be identified and monitored in time using a fully automated and non-invasive method. The amplitudes of the independent signals form a multi-dimensional time-varying vector which was logged continuously for eight months. We found that combined with specifically tailored weighting factors, this vector provides a signature highly specific to the swarming process and its build up in time, thereby shedding new light on it and allowing its prediction several days in advance.

The output of our monitoring method could be used to provide other signatures highly specific to other physiological processes in honey bees, and applied to better understand health issues recently encountered by pollinators.

Key words: **honey bee;swarming;vibrations;pca**.

1. INTRODUCTION

Honey bees live in societies of several thousands of individuals and form the object of much scientific work due to the complexity and sophistication of their behaviour. Following the major advances in this field brought by Karl von Frish (Frish, 1967), extensive work has been done focusing on further understanding the way honey bees communicate, in particular with acoustic noise and body vibrations.

In the waggle dance 280 Hz sounds are produced modulated by a 15 Hz body wag (Michelsen A., 1999, pp. 111-131), performed directly on the comb (although not involving abdominal contact). It has been shown (Tautz J., 1996) that the substrate on which the dance is performed affects the quality of the communication process (although plastic comb foundations do not (Seeley T.D. *et al.*, 2005), that bee legs are most sensitive to propagating (in the comb) longitudinal vibrations between 30 and 100 Hz (Sandeman D.C. *et al.*, 1996), that vibrations found in the comb between 200 and 300 Hz are significantly enhanced during the waggle phase (Nieh J. and Tautz J., 2000), and that phase reversal of the propagating wave occurs in close proximity to the dancer, at a distance at which followers are seen to be most frequently attracted to the messenger (Tautz J. *et al.*, 2001).

Studies of sound radiated by bees have focused on the sound field of single bees (Michelsen A. *et al.*, 1987), and it appears that follower bees tend to place their antennae in the zone of maximum acoustical short-circuiting. The properties of 'piping', a specific sound pulse which has been shown to be intimately linked with the preparation for swarming (Seeley T.D. and Tautz J., 2001; Visscher P.K. and Seeley T.D., 2007), has also been studied in terms of acoustic noise, both for 'worker piping'

(Seeley T.D. and Tautz J., 2001; Visscher P.K. and Seeley T.D., 2007) and 'queen piping' (Kirchner W.H., 1997, pp. 273-300). Measurements of the sound radiated by a collection of bees has been reported in the scientific literature (Ferrari, 2008), by placing a microphone on the top of the frames of a hive. This only provides crude and unreliable short term recordings, although the authors have some evidence that these can indicate the preparation for the swarming process. Similar microphone measurements are also found in a few patents (Bromenshenk *et al.*, 2007; Etter R. *et al.*, 2007; Bromenshenk *et al.*, 2009) including exemplification data by which the inventors demonstrate that upon exposure to predatory mites or to sub-lethal concentrations of a few specific airborne toxicants, differing signatures may be extracted from the measured honey bee sound. A separate patent (Woods, 1957) claims that a set of simple band pass filters applied to the acoustic noise of bees may be a good indicator of the vicinity of the swarming process but no exemplification data is made available to the reader to substantiate the claim.

Very recently, a device for the long term logging of important features of honey bee sounds in a hive has also been successfully developed (Atauri, 2009).

Those honey bees that are managed by humans often live in wooden hives, and these emit acoustic noise that can clearly be heard if the ear is located near enough to it. In periods of very high activity, the mechanical vibrations sustained in the hive's outer box can also be sensed simply by placing the hand on the wood.

In our work these vibrations were sensed by two accelerometers secured onto the outer wall of two separate hives, approximately 3 meters away from one other, each comprising of a healthy colony of *Apis mellifera* honey bees. We are focusing on analysing this ‘by-product’ of honey bee activity, to show that it comprises of complex global information, rather than taking the microscopic approach of exploring the insect’s individual communication process.

We present one method for processing this raw data in order to extract a specific time varying signature comprising ten independent components, and demonstrate that this can be applied to identify and predict the swarming process several days in advance.

2. MATERIALS AND METHODS

2.1. Honey bees under investigation.

Two *Langstroth* hives were monitored in Jarnioux, France (Latitude (DMS) 45° 58' 0N Longitude (DMS) 4° 37' 60E Altitude (meters) 271). They were located in close (10 meters) proximity to a house comprising a computer indoors (Pentium III 800 MHz CPU running a Linux operating system). A simple roof was placed above the monitored hives to prevent vibrations caused by rain drops directly falling onto them. The two colonies were looked after with minimal interference, although they were fed during the winter.

Swarming was witnessed by one of the authors on several occasions:

- 3rd of May for hive 1 (11:30 am) and hive 2 (5:30 pm),
- 30th of April, 2nd and 11th of May for hive 2.

We found out that by mistake a queen excluder had been placed at the bottom of the broodbox of the second hive. This caused six unsuccessful swarms (three of which were witnessed) for that hive until the queen was left to go by opening a hole in the front wall, generating a final seventh swarm (also witnessed). The first hive swarmed twice, one of which was witnessed.

2.2. Vibration data collection

A 5 mm deep cavity was drilled in the wood of the back of each hive, in the centre, with a diameter causing a tight fit to the accelerometers. The transducers (Isotron 7259B-10, Endevco, San Juan Capistrano, U.S.A.) were connected to a dual channel conditioner (Nexus, Brüel and Kjær; Nærum, Denmark) residing between the two hives, and encased in a water tight acrylic box to minimise moisture ingress into the electronics. Having checked that the signal above 4000 Hz was negligible, the analogue output channels of the conditioner were fed into the sound card of the indoors computer, by means of individual shielded coaxial cables, for 16 bit, 8000 Hz sampling rate digitisation and storage on a 500 GB external hard disk.

Four very gentle knocks were supplied by hand to each hive, and we could not measure any crosstalk from the two resulting digitised waves, although both hives sat on the same wooden beams, approximately one meter above the ground. The waves recorded from the knocks were additionally used to estimate the resonant frequencies of the hives under investigation. We also checked that large external acoustic noise such as lorries passing nearby (~ 100 m away) did not induce measurable signal in our recordings. The vibrations (i.e. a deformation wave propagating in wood/wax) that are sensed are therefore mainly coming from (i) the acoustic noise (a pressure wave propagating in air) from the honey bees inducing vibrations into the structure of the hive and (ii) the body motion of the individual bees inducing vibrations through their legs residing on the wax/wood.

The computer started logging the vibrations in individual files of time duration of one hour each, from the 1st of November 2008 until the 9th of April 2009, when a power cut interrupted the recording (well before the swarming season). It was launched again on the 11th of April 2009 until a second power cut stopped and damaged the computer on the 17th of June 2009.

2.3. Data processing

The time course of the measured vibrations mostly comes from several thousands of honey bee individuals which may be seen as a collection of transient pulsed oscillators with random individual phases. The raw signal is therefore too noisy for direct analysis. By computing short frequency spectra that are averaged together, a very clear curve with pronounced peaks may be rapidly obtained, which suggests that the vibrations induced by the oscillators have similar spectral features in spite of their lack of phase coherence. The time duration of an individual sample for computing one instantaneous spectrum determines its frequency resolution whilst the time duration over which it is averaged dictates the signal to noise ratio of the signal. A good compromise was found for a frequency resolution of 20 Hz and an averaging time of 510 seconds.

Using a purpose-built computer script (in matlab®, as is for all software presented in this work), these averaged spectra were computed for the entire data set, and stacked in one day long images called spectrograms for visual inspection.

The next processing task consisted in extracting an instantaneous 'feature' that one can use as a 'signature' of a specific state of activity of the honey bee hive. To date we have used (Baxter, 2003) principal component analysis (PCA).

Given n observations on p variables X_i, $i = 1, \ldots, p$, PCA as most commonly used involves a linear transformation to p orthogonal variables, Z_i, the principal components, If $\mathbf{X}$ is the original n by p data matrix then $\mathbf{Z}$, the n by p matrix of principal component scores, contains exactly the same information as $\mathbf{X}$. A major use of PCA is as a dimension reduction technique, whereby the first r columns of $\mathbf{Z}$ can be used to 'approximate' $\mathbf{X}$, where r is much less than p. This, allied to the orthogonality of the Z_i, allows patterns in the data to be explored much more readily in r or fewer dimensions than with the original data matrix. It is sometimes possible to assign a 'meaning' to the components, but this is not essential for the method to be useful.

Specializing to our application, the rows and columns of $\mathbf{X}$ correspond to frequency and time, and can be interchanged. Entries in the data matrix correspond to (averaged) amplitudes. Both frequency and time have a natural ordering that can be exploited in interpreting the principal components, which we call 'eigenspectra' in this study. Often, as is the case for our data, relatively few components are needed to explain most of the variation in the data, and time plots based on the first few components can identify normal hive activity that changes when, for example swarming is about to occur.

The entire time series of n = 37856 spectra for each hive was analysed by PCA on centred but unstandardised data, using a separate script which takes 53 ms to converge (for one hive) on a 2.4 GHz CPU computer. Following this, any averaged spectrum can be expressed as a linear combination of the Z_i eigenspectra, to which the mean spectrum (over the entire data set) must be added.

The weighting factors required for a specific 'reconstruction' are called scores, and at any point in time an instantaneous spectrum can be perfectly described by the linear combination of the corresponding p instantaneous scores and p eigenspectra (which are valid for the entire data set), where p is the number of digital points in one spectrum (p = 200 for the compromise that we chose in this study. Higher p requires longer spectral averaging time, in order to keep the same Signal to Noise). Alternatively, an approximate description of an instantaneous spectrum may be obtained by reducing the number of scores and eigenspectra (e.g. from p = 200 to r = 10) used in the linear combination.

3. RESULTS

3.1. Spectrograms.

The one day-long spectrograms of the two hives are shown in Figure 1 for a specific day. For both hives four to five 'bands' are clearly seen, the strongest being around

2000 Hz. Small frequency differences can be seen between the two hives, resulting from a combination of different hives and, perhaps, from differences coming from the honey bee colonies themselves. Continuous activity in the night can be seen as well as amplitude variations over two orders of magnitudes. The second hive provides approximately twice more signal than the first one.

3.2. PCA filtering of honey bee signal.

We established that at any point in time the linear combination of the first ten scores and eigenspectra gives an excellent reproduction of the original spectrum (Figure 2). The eigenspectra were also inspected visually and those corresponding to orders higher than ten essentially contain noise. In our data the information relevant to honey bee activity can be approximated very well by a PCA analysis using $r = 10$.

3.3. Eigenspectra.

The eigenspectra for both hives exhibit remarkable similarities, and those of hive 1 are shown in Fig. 3. Fortunately, the effect of eigenspectra with order higher than one can often be interpreted in terms of three majors features: shifting specific peaks (corresponding to a change in signal frequency), changing their relative amplitudes (corresponding to the strength of the vibration at a given frequency) or width (corresponding to the time duration of the pulsed oscillators causing the vibrations). Comparison with the spectrum of the hive resonances shows marked discrepancies, but we have not attempted at compensating our data for the natural 'transfer function' that the hive provides between the honey bee vibrations and our measurements.

3.4. Scores.

Visual inspection of the spectrograms mostly reveals very large variations of the signal's overall amplitude. These variations are captured in the time course of the scores on the first order principal component (PC), or eigenspectrum. The PC scores with a higher order (two and above) reveal subtle variations in the signal (Fig. 4) that are not obvious to the naked eye on the spectrograms, but highly relevant to honey bee activity, as will be demonstrated in this work.

To allow suitable visualisation of the results, the first order score in Figure 4 has been scaled down so as to reach the same maximum value as the second one, whilst the others have not been scaled. For the first hive and second hive respectively, on average, the ratio of the first to the second order score is found to be six and twenty. This is in good agreement with the visual inspection of the hives, as by opening them prior and following the study it was quite clear that the second one was more populated and 'stronger' than the first one.

3.5. Identification of the swarming process.

The relevance of the time course of the scores was specifically tested on the process of swarming. The swarming events that have been witnessed were first inspected on the spectrograms and on the time course of the scores. These plots were also inspected after the 11th of May and before the 3rd of May, where no swarming was witnessed or possible (winter time). When described by the time course of the scores, the swarming events were found to exhibit a unique set of combined features (Fig. 5) never seen together in the winter days or those following the 11th of May: a gradual, exponential-like divergence of most of the scores over approximately five to ten hours, followed by a peaked amplitude over a single point in time (i.e. over 8.5 mins), followed by a reversal of polarity of some scores and an extremely low set of amplitudes thereafter (except for scores directly expressing frequency shifts).

A simple quantitative study of the vibrational measurements of the building-up of the swarming event was undertaken for the first hive, which swarmed “naturally”. The second hive’s swarming data is affected by the misplaced queen excluder, and is therefore biased. For hive 1, the witnessed swarming was selected, and the 5 hours of spectrogram immediately preceding the highest vibration at the end of the divergence of the data was selected. The cross-correlation function between this section of data and the entire data set was calculated as a function of time, and this did not yield specific peaks in the vicinity of the swarming. The cross-correlation function was then calculated for each individual score of the same section of data. A clear build-up of the amplitude of the function was seen within several days preceding swarming for the scores of order 5, 9 and 10 (Figure 6), whilst those maxima seen elsewhere for the scores with other orders did not seem to relate to the swarming.

We then investigated the relative importance of specific scores, with regards to detecting the swarming event from our data. A script was written to compute a cross-correlation function combining each of the ten scores with independent weighting factors. The script uses the “fminsearch” Matlab® algorithm and attempts at maximising and minimising the function when using a section respectively during swarming and outside swarming. The section that we selected to maximise the function was limited at the end of the witnessed swarm, and its start can be chosen at any time between one to ten days prior to it, without changing the result. The selected section for minimisation was limited by the end of the data, and can be started at any time between two to ten days before, which correspond to 45 days and 35 days after the first and second peak of the function. The results shown in Figure 7 suggest that swarming can be detected several days in advance and that a secondary swarm (not witnessed) emerged ten days after the first one, as is often seen in hives that are not interfered with (Winston, 1987).

3.6. Sunrise activity.

For both hives on most of our spectrograms acquired after the 10th of March (see Fig. 1. and Fig. 2.) a clear peak of vibrational activity can be seen in the early hours of the morning. This was further investigated by looking at the time course of individual scores (Fig. 8.), which we superimposed with the time course of the sunrise*.

4. Discussion and further work.

We have shown that a non-invasive and automated method can provide to be a useful monitoring device, by analysis of the time course of the normal vibrations recorded on the wall of a honey bee hive.

We chose to use accelerometers in our work, which, to our knowledge, have never been tried as hive monitoring sensors. Although not published in the professional literature, we know from our own experience and that of others (Michelsen A., 2005) that **microphone** measurements are only sensitive to a few bees in the close vicinity of the microphone, that bees passing nearby can cause huge transient signals, that unwanted external noises such as bird singing or airplanes or motorcars can often be picked up, and that long-term monitoring is severely restricted due to the natural propensity bees exhibit to coat with propolis any body that is alien to their natural habitat. The published work that involves microphone measurements is restricted to relatively short sequences (Bromenshenk *et al.*, 2007; Etter R. *et al.*, 2007; Bromenshenk *et al.*, 2009), and/or single bee measurements (Michelsen A. *et al.*, 1987), and/or measurements done in a hive which probably senses a few hundred bees (Ferrari, 2008), which are continuously moving in and out of the volume to which the microphone is sensitive. Vibrational data has also been measured by **video** analysis of transparent hives (Schneider, 2004; Brennan, 2007). Such measurement requires specialised transparent hive hardware and cannot easily be undertaken in the natural darkness to which honey bees are used. Lengthy measurements end up in very large data sets, and analysis is, to our knowledge, only possible by human ‘manual’ analysis. Another possibility is to measure substrate born vibrations with **laser Doppler vibrometry** (Nieh J. and Tautz J., 2000; Tautz J., 2001; Seeley, 2005), which is sensitive to the solid boundary displacement. This technique is exquisitely sensitive to the smallest relative displacement between the sample and the vibrometer, and has not, to our knowledge, been applied to either *in-situ* or long-term measurements. It is however expensive, suffers from slow drifts due to temperature changes, and from any optical interference such as dirt particles gradually building up on the light path. A less expensive approach, for individual insects, is to measure vibrations using the **stylus** of a ceramic cartridge in contact with a substrate (Reader and Duce, 2009). Such an approach relies on the measurement of displacement of the substrate, and is generally thought less sensitive, with a lower signal-to-noise ratio, than similar techniques employing an accelerometer.

Without the need for combining our method with weather or other measurements, we have shown that a quantity highly specific to the swarming process is seen to grow increasingly strong, several days in advance of the witnessed swarm. The identification of swarming in the vibration data requires PCA analysis, with the extraction of PC scores of fairly high orders.

We have discovered a peak in honey bee vibrational activity, early in the morning, which matches the sun rise timings, which we have not seen reported in the literature elsewhere.

Our spectra reveal vibrations coming from honey bees, but the information is severely weighted by the resonances of the hive's solid structure. In spite of this we have found remarkable similarities in some of our spectral patterns with that of other workers. Whilst we see pronounced peaks at 250, 500, 750 and 2000 Hz, others have mostly focused on acoustic noise peaks, and have found similar low frequency peaks. Dietlein (Dietlein, 1985) has found noise to be concentrated around 300, 410 and 510 Hz, Eren's data (Eren, 1997) suggests peaks at 235 ± 35 Hz and 425 ± 25 Hz, and the latest published spectra (Ferrari, 2008) exhibits clear peaks at 150 Hz and 300 Hz. Clearly this suggests that our two first peaks are directly due to honey bee vibrations whilst the higher frequencies are most probably harmonics of the hive structure's resonant modes, indirectly stimulated by honey bee activity.

In the future it would be desirable to relate the specific vibrations to physical processes involving honey bee activities, such as body and wing vibrations, and to better understand the means by which they are conveyed to the transducers (wax, wood, resonances of the structures). Vibrations in two other tangential directions and also in separate locations could provide improved or novel information. Other processes than the swarming could be explored and monitored, as well as other social insects which live in large enough groups (wasps, bumble-bees). The process of swarming itself might be better understood by combining this study with a weather monitoring logging device, and with measurements relating the hive's population with the magnitude of the vibrations. Using pattern recognition algorithms that are more sophisticated than the simple cross-correlation function is also a way forward when corroborative measurements are available such as video monitoring. Finally, by using hives in well controlled laboratory environment, very low frequencies (frequencies lower than 120 Hz have been dismissed in this study) could be sensed, such as those used for other important processes clarified by other workers (Schneider, 2007).

Acknowledgements

We are grateful to Dr. Robert Morris, NTU, who wrote the Linux script continuously storing the digitised data in individual short files, * to Dr. Daniel Brown, NTU, who worked out the sun rise times from the location of the hives, to Antoine and Joseph Bencsik for looking after the 8 month-long continuous recording, and to NTU who partly funded this work through a SPUR summer work bursary which Andrei Lucian benefited from.

References

Atauri D. and Llorente J., 2009. Platform for bee-hives monitoring based on sound analysis. A perpetual warehouse for swarm's daily activity. Spanish Journal of Agricultural Research. 7 (4), 824-828.

Baxter, M., 2003. Statistics in Archeology. John Wiley and Sons.

Brennan B., 2007. Abdominal wagging in the social paper wasp Polistes dominulus : Behavior and substrate vibrations, Ethology, 113, 692-702.

Bromenshenk et al., 2007. US 2007/0224914 A1.

Bromenshenk et al., 2009. US 7549907 B2.

Dietlein D.G., 1985. A method for remote monitoring of activity of honeybee coloniess by sound analysis. Journal of Apicultural Research, 24 (3), 176-183.

Etter R. et al., 2007. CA 2 573 049.

Ferrari S. et al., 2008. Monitoring of swarming sounds in bee hives for early detection of the swarming period. Computers and Electronics in Agriculture. 64, 72-77.

Kirchner W.H., 1997. Orientation and Communication in Arthropods. (Lehrer M. Ed.) Birkenhäuser Verlag, Basel.

Michelsen A. et al., 1987. The acoustic near-field of a dancing honeybee. Journal of Comparative Physiology A, 161, 633-643.

Michelsen A., 1999. The Dance Language of Honeybees: Recent Findings and Problems, in: The Design of Animal Communication (MD Hauser and M Konishi, Eds.) MIT Press.

Michelsen A., 2005. Personal communication.

Nieh J. and Tautz J., 2000. Behaviour-locked signal analysis reveals weak 200-300 Hz comb vibrations during the honeybee waggle dance. Journal of Experimental Biology, 203, 1573-1579.

Reader T. and Duce I.R., 2009. Intraguild interactions promote assortative mating and affect sexual attractiveness in a phytophagous fly, Biological Journal of the Linnean Society, 98, 171-180.

Sandeman D.C. et al., 1996. Transmission of vibration across honeycombs and its detection by bee leg receptors. Journal of Experimental Biology, 199, 2585-2594.

Schneider S.S. and Lewis L.L., 2004. The vibration signal, modulatory communication and the organisation of labour in honey bees, Apis mellifera. Apidologie 35, 117-131.

Schneider S., 2007. Vibration Signals and the Organisation of Labour in Honey Bee Colonies. Bee Craft, 20-25.

Seeley T.D. and Tautz J., 2001. Worker piping in honey bee swarms and its role in preparing for liftoff. Journal of Comparative Physiology A, 187, 667-676.

Seeley T.D. et al., 2005. Does plastic comb foundation hinder waggle dance communication?. Apidologie, 36, 513-521.

Tautz J., 1996. Transmission of vibration across honeycombs and its detection by bee leg receptors. Journal of Experimental Biology, 199, 1375-1381.

Tautz J. et al., 2001. Phase reversal of vibratory signals in honeycomb may assist dancing honeybees to attract their audience. Journal of Experimental Biology, 204, 3737-3746.

Visscher P.K. and Seeley T.D., 2007. Coordinating a group departure: who produces the piping signals on honeybee swarms?. Behavioral Ecology and Sociobiology, 61, 1615-1621.

Von Frisch K., 1967. The dance language and orientation of bees. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Winston M.L., 1987. The Biology of the Honey Bee. Harvard University Press, Cambridge.

Woods E.F., 1957. US 2,806,082.

FIGURE CAPTIONS

Figure 1. **Honey bee hive vibrations.** One day long spectrogram of the data collected on hive one (top) and hive two (bottom) on the 15th of April 2009. The colour linearly codes the amplitude of the acceleration (in arbitrary units, but identically scaled for both data sets), for a time and a frequency given by the horizontal and vertical coordinates.

Figure 2. **PCA filtering of hive vibrations.** The data from hive 1 on the 15th of April 2009 is projected onto its set of eigenspectra. On the top, all of the 200 scores are used to reconstruct the data, ending up in the same spectrogram as seen in Fig. 1. On the bottom plot, 10 scores (and associated spectograms) only are used. The PCA filtering process removes unwanted noise, simplifies the analysis of the extracted data, and preserves excellent temporal and frequency resolutions, unlike Fourier or spatial convolution types of filtering.

Figure 3. **Eigenspectra of honey bee hive vibrations.** When comparing the two hives, some strong features are found to be common to the two sets of eigenspectra: (i) they exhibit oscillations around zero except for the first one which corresponds to the overall amplitude of vibrations, (ii) four peaks are clearly identified around 500 Hz, 750 Hz, 2000 Hz and 3200 Hz, (iii) a pair of eigenspectra (here, No 3 and 5) causes a shift of the frequency in two peaks in the same (No 3) or in opposite directions (No 5), whilst (iv) another pair (here, No 2 and 4) allow the respective amplitudes of the peaks to be adjusted independently. Some mild differences are also seen, probably due to a combination of the hive natural resonances and the signature of the vibrations of a colony being specific to a particular colony (those spectra shown here come from hive

1). The entire data set (37856 spectra each averaged over 8.5 minutes) was supplied to our software to extract this data. The dotted line give an idea of the mechanical resonances of the hive ensemble, obtained by Fourier transformation of the pulses recorded with the hand knocks, although frequency-dependent attenuation coefficients also participate in the shape of that curve.

Figure 4. **Spectrogram and scores.** On the top, the one day long spectrogram of the data collected on hive 2 on the 6th of May 2009 is shown, together with the time-course of the six most important scores (bottom) corresponding to the same day of activity. The first order score spans much larger values than the others, and has been scaled down (see text). A score taking a value of zero means that the corresponding eigenspectrum is not needed at that time, as its contribution exactly matches that of the mean (over the entire data set) spectrum. A score with, respectively, a positive or negative value indicates that the corresponding eigenspectrum provides a contribution larger or weaker than that required for the mean spectrum. One score does not correspond to a specific honey bee vibrational activity, but any score provides a contribution to the total signal that is statistically independent from any other score, since the eigenspectra have been ensured to be all orthogonal to one another by PCA.

Figure 5. **Time course of scores on six different days.** The four witnessed swarms for hive 2 are shown from **A** to **D**, that witnessed for hive 1 is shown in **E**, whilst a 'normal' day's data well after the swarming season in shown in **F**, for comparison purposes. The set of features (see text) seen for the swarming events are common to all five witnessed swarming events (irrespective of the hive), and are never seen outside the swarming

season. The swarming signatures in **A**, **B** and **E** are remarkably similar (and happen at the same time in the day, between 11 am and 1 pm), whilst those in **C** and **D**, occurring later in the day (between 4 pm and 5 pm), are probably strongly affected by the queen 'trapped' in the hive.

For all plots the first score has been scaled here as in Fig. 4.

Figure 6. **Time course of the swarming cross-correlation function for the 9th order score of the first hive.** The red line indicates the day of the swarming event that has been witnessed, which also comprises the five hours of data used for computing the cross-correlation function. The function is negative at any time in the recording until 11 days before swarming where it becomes positive. The first maximum is extremely close to the swarming event, the second one occurs ten days later where a secondary swarm might have occurred. The curve is remarkably insensitive to the time duration of the selected swarming section, provided that it is between two to ten hours, and demonstrates the emergence of a vibrational 'pattern' highly specific to the swarming process.

Figure 7. **Time course of the weighted cross-correlation function for the first hive.** The weighting factors have been optimised as indicated in the text. The colour codes the cross-correlation function, in arbitrary units. On the left, all values are shown, clearly demonstrating a discontinuity of the data to negative values after the 23rd of May. On the right, all negative values have been forced to take the deep blue color, so that on the basis of that threshold an 'alarm' relevant to the swarming event can be triggered with

increasing confidence from mid-April onwards. A few false alarms appear in winter but they are not followed by an increasing value of the function.

Figure 8. **colour-coded time course of the third order score (hive 1).** A clear peak in amplitude is seen early in the morning, from the 10th of March onwards, which follows very well the timings of the sun rise, which is displayed as a white line. The peak lasts a few minutes and is also followed by a return to normality. The first day on the horizontal axis corresponds to the 3rd of November 2008.

FIGURES

Figure 1

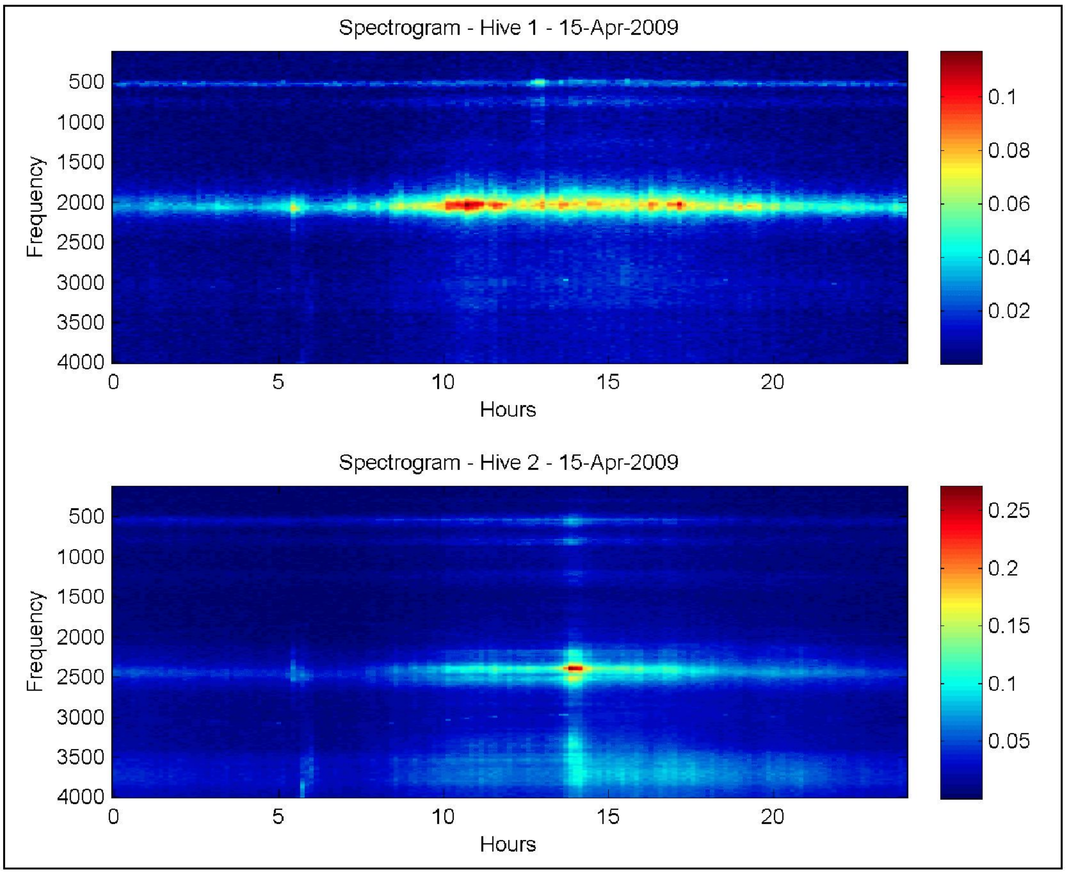


Figure 2

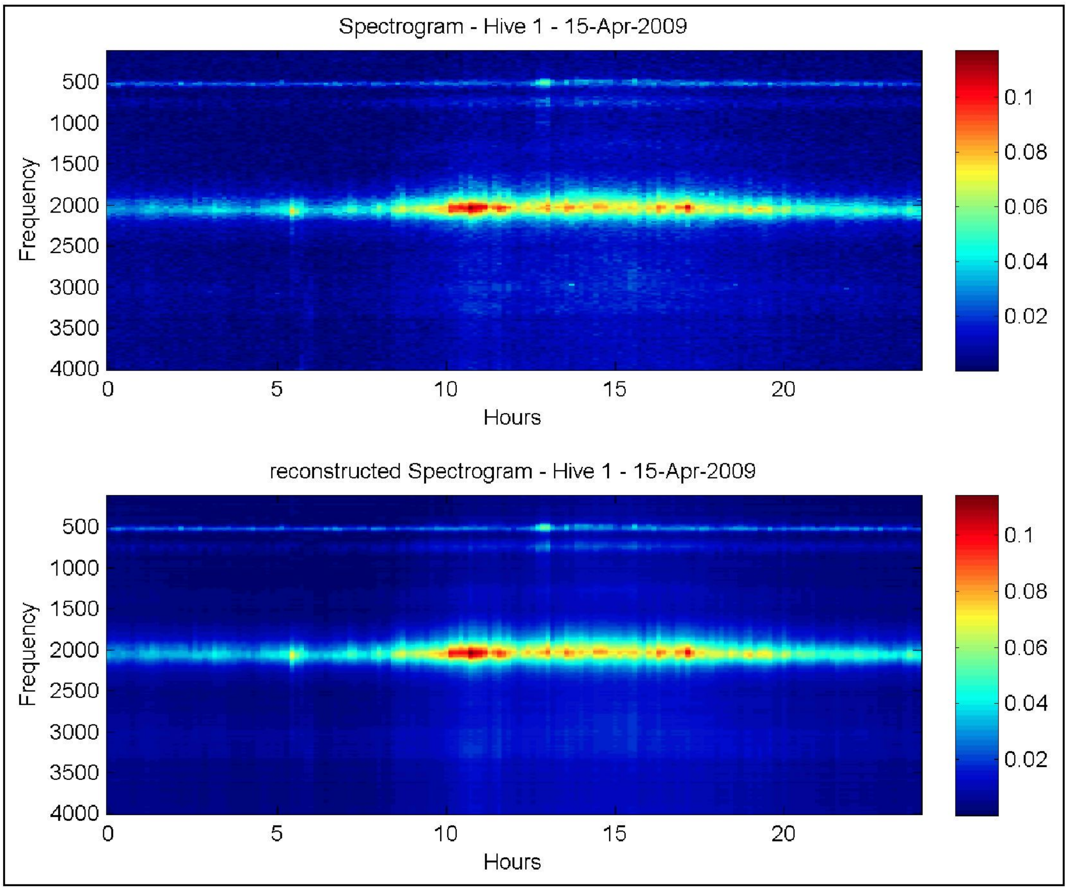
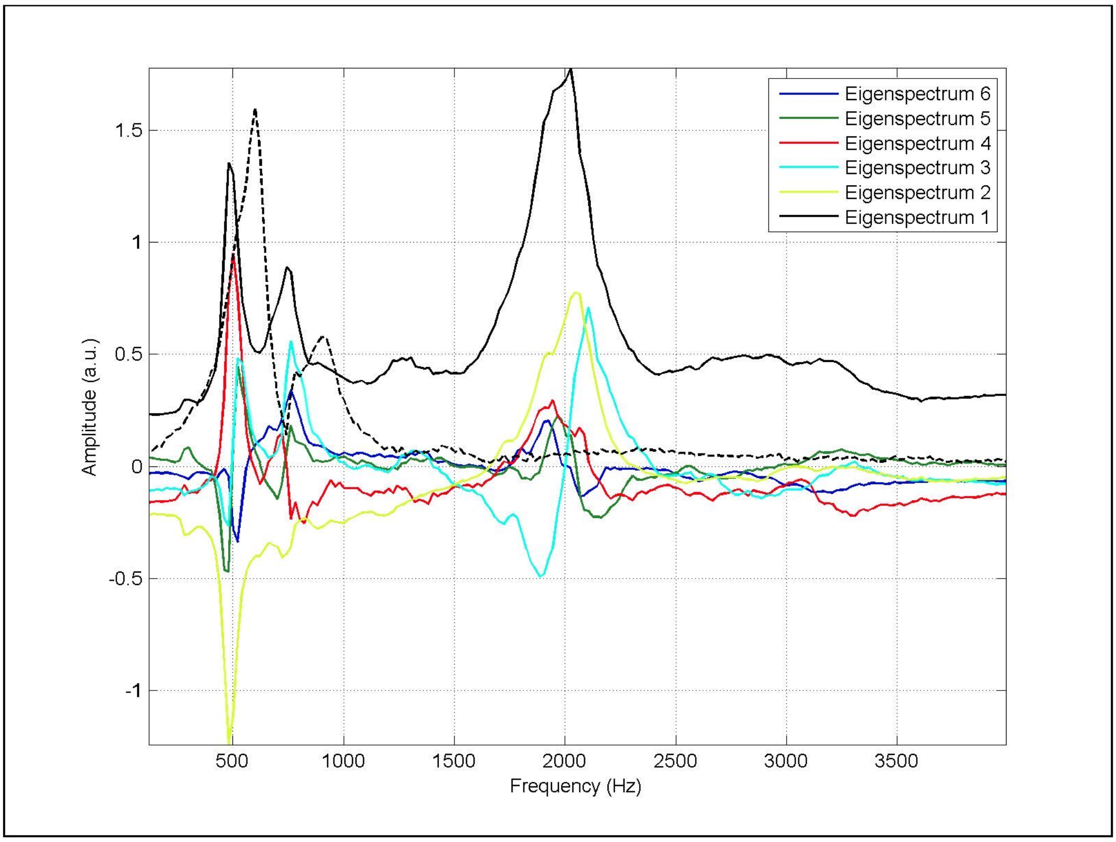

Spectrogram - Hive 1 - 15-Apr-2009
Frequency
500
1000
1500
2000
2500
3000
3500
4000
0
5
10
15
20
Hours
0.1
0.08
0.06
0.04
0.02
reconstructed Spectrogram - Hive 1 - 15-Apr-2009
Frequency
500
1000
1500
2000
2500
3000
3500
4000
0
5
10
15
20
Hours
0.1
0.08
0.06
0.04
0.02

Figure 3



Eigenspectrum 6
Eigenspectrum 5
Eigenspectrum 4
Eigenspectrum 3
Eigenspectrum 2
Eigenspectrum 1
1.5
1
0.5
0
-0.5
-1
Amplitude (a.u.)
500
1000
1500
2000
2500
3000
3500
Frequency (Hz)

Figure 4

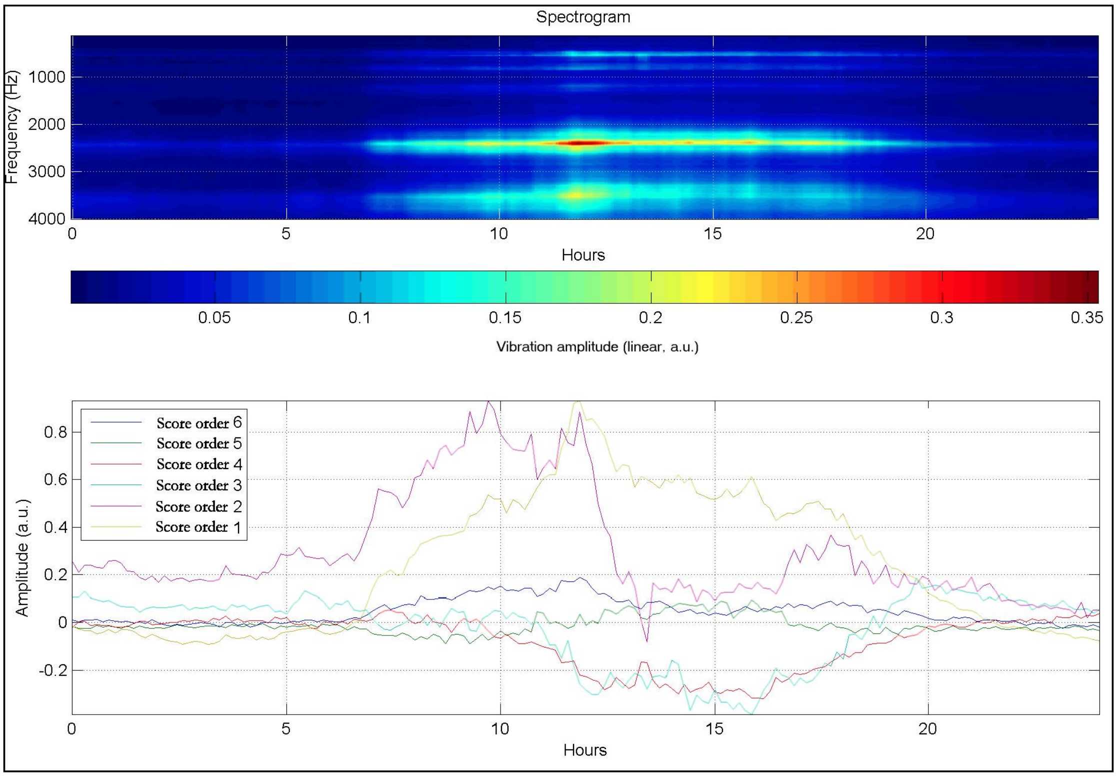

Spectrogram
Frequency (Hz)
1000
2000
3000
4000
0
5
10
15
20
Hours
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Vibration amplitude (linear, a.u.)
Score order 6
Score order 5
Score order 4
Score order 3
Score order 2
Score order 1
Amplitude (a.u.)
0.8
0.6
0.4
0.2
0
-0.2
0
5
10
15
20
Hours

Figure 5

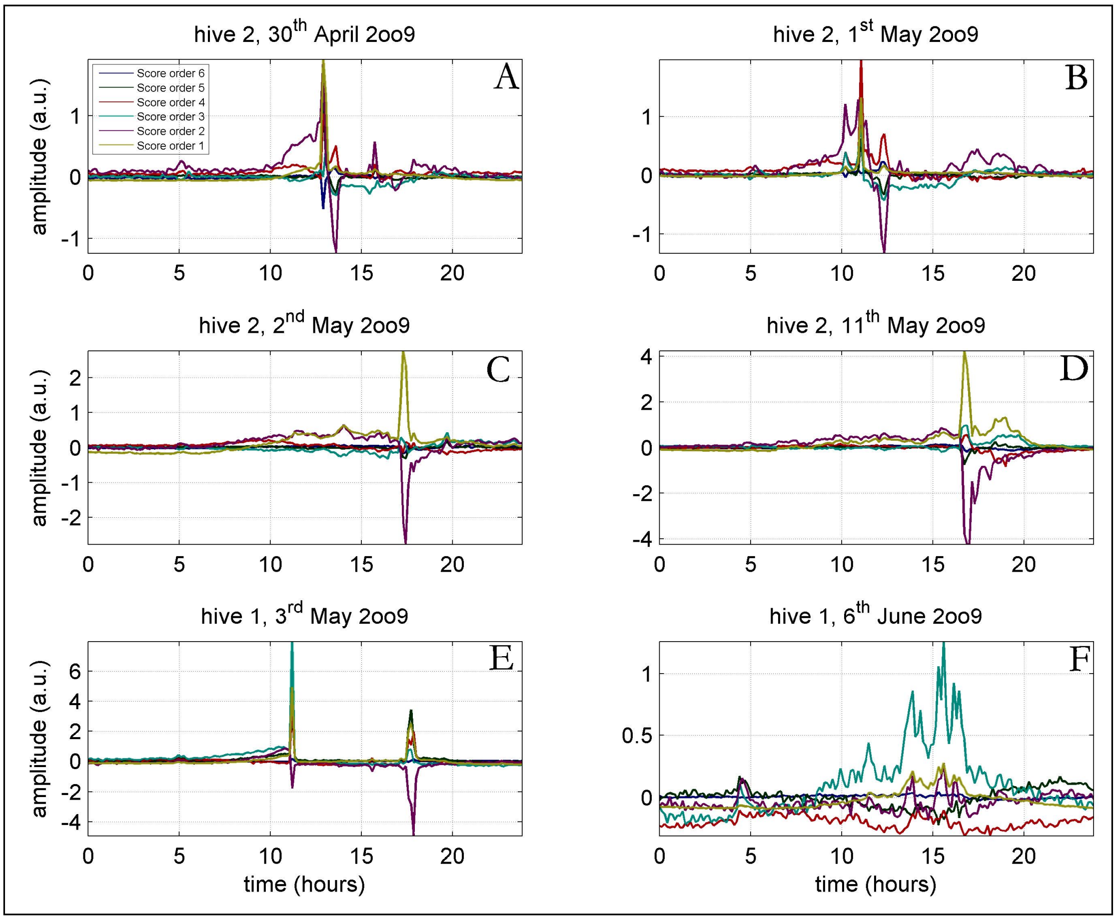

hive 2, 30th April 2oo9
A
Score order 6
Score order 5
Score order 4
Score order 3
Score order 2
Score order 1
amplitude (a.u.)
1
0
-1
0
5
10
15
20
hive 2, 1st May 2oo9
B
1
0
-1
0
5
10
15
20
hive 2, 2nd May 2oo9
C
amplitude (a.u.)
2
1
0
-1
-2
0
5
10
15
20
hive 2, 11th May 2oo9
D
4
2
0
-2
-4
0
5
10
15
20
hive 1, 3rd May 2oo9
E
amplitude (a.u.)
6
4
2
0
-2
-4
0
5
10
15
20
time (hours)
hive 1, 6th June 2oo9
F
1
0.5
0
0
5
10
15
20
time (hours)

Figure 6

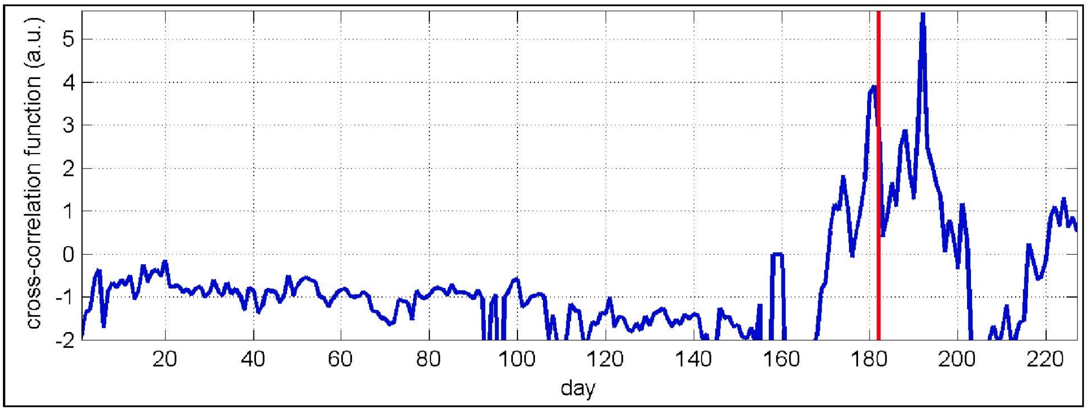


Figure 7

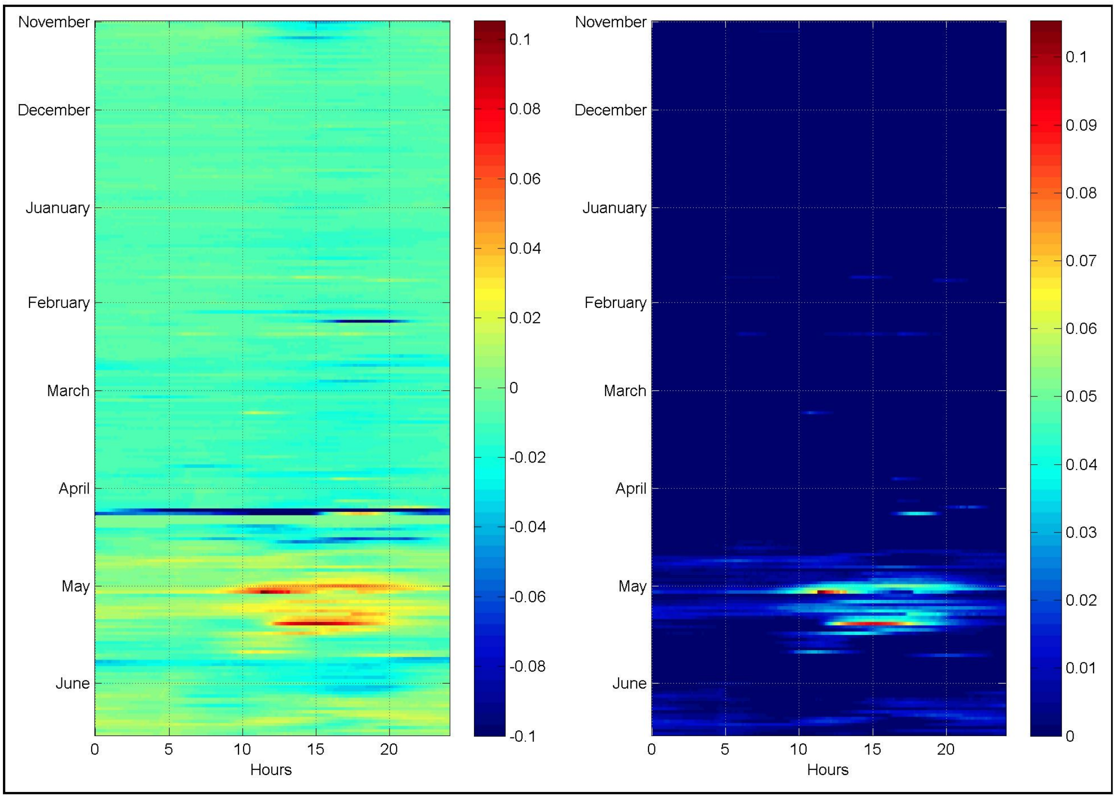
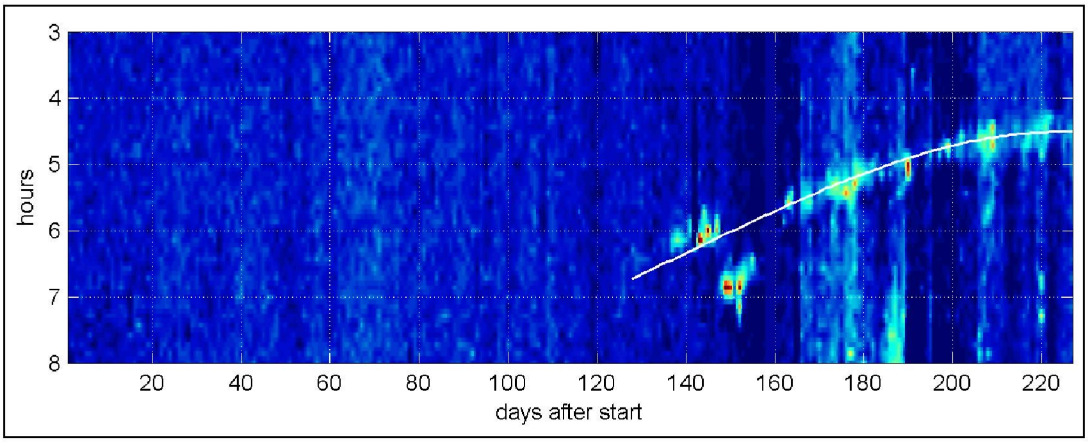


Figure 8


3
4
5
6
7
8
hours
20
40
60
80
100
120
140
160
180
200
220
days after start