

# Chapter 3

## Analysis of Variation of Thermal Performance Curves

The caterpillar data in Figure 1.1 are an example of Thermal Performance curves (TPCs). In turn, (TPCs) are particular examples of a larger class of curves collected in biology called Continuous Reaction Norm Curves (CRNs). Examples of CRNs and some terminology in evolutionary biology is given in Section 3.1 and illustrated with the caterpillar example. Representing CRNs of different genotypes as curves offers the main advantage of visualizing the change of a trait in a population as the environment changes and also as the genotype changes. The genetic variation in CRNs is important as is illustrated in Section 3.2 because it affects the evolutionary response of a population under selection. We illustrate our method for the decomposition of the genetic variation into modes of biological interest using our method in Section 3.5. A paper presenting some of the results presented in this Chapter [Izem, Marron, and Kingsolver (2004)] won the student paper competition in Graphics and Computational Statistics, it will be published in the Proceeding of the American Statistical Association.

## 3.1 Continuous Reaction Norm Curves

Growth rate as a function of temperature, body weight as a function of diet quality, plant height as a function of the amount of soil nutrients, and photosynthetic rate as a function of light intensity are examples of continuous reaction norms (CRNs). Growth rate, body weight, plant height, and photosynthetic rate are examples of phenotypic traits as we see in Subsection 3.1.1. CRNs are important for biologists to understand the effect and interaction of the genetic make-up and the environment on Phenotypic traits of an individual. Thermal performance curves are particular examples of CRNs which are formally defined in Subsection 3.1.4. Most of the definitions in this section could be found in [Lynch and Walsh (1998)].

### 3.1.1 Phenotypic trait

In the caterpillar data, the growth rate  $z$  of the caterpillar is a measure of gained mass in a unit of time. Because the mass of a caterpillar is a physical characteristic, and the growth rate is a function of mass then the mass and the growth rate are called *phenotypic traits* in evolutionary biology. In general, *a phenotypic trait of an organism is any observable physical characteristic of that organism*. Other phenotypic traits of caterpillars include initial body mass, body dimensions, color and pupal mass. Examples of phenotypical traits for a plant are height, color and photosynthetic rate.

### 3.1.2 Phenotype, genotype, and environment

The growth rate of a caterpillar depends on its genetic make up and also on the environment. For the caterpillar data, each individual's growth rate was measured in a controlled lab environment where only the temperature varied. In general, we can consider the value  $z$  that a phenotypic trait takes for an individual or a family as a realization of a random variable

$Z$ . The value of  $Z$  depends on the two random effects  $g$  and  $e$ . Where  $g$  and  $e$  are the genetic and the environmental effects respectively. Evolutionary biologists suppose that the genetic and the environmental effects are independent. We can think of *the genotype as the inheritable information carried in the genes of an organism*, and *the phenotype as the visible or external physical manifestation of an organism*. The phenotype is a compounding of genetic and environmental effects.

### 3.1.3 Continuous Reaction Norms: $z = Z_g(e)$

*A reaction norm describes the phenotype that a genotype produces across a range of environments (Woltereck, 1909; Johannsen, 1909, 1911)*. So, for a fixed genotype  $g$ , the change of a trait as a function of some aspect of the environment  $e$  is called a reaction norm. When the aspect of the environment  $e$  changes continuously, the reaction norm is called continuous reaction norm. In this dissertation, we consider the cases where  $z$  is a continuous function of some aspect of the environment. So, although in most data the trait is measured at only some environments, the underlying relationship between the trait and the environment is considered continuous. This continuity allows us to address the problem in a Functional Data Analysis framework in Section 3.3.

### 3.1.4 Thermal performance curve (TPC), $z = Z_g(T)$

The caterpillar data is an example of thermal performance curves, where the growth rate is a trait which measures a performance. It is also an example of CRN where the phenotypic trait is the growth rate and the aspect of the environment of interest is the temperature. *A thermal performance curve is a continuous reaction norm where the trait measures a performance of the organism, and the aspect of the environment of interest is the temperature*. A main feature

of TPCs for a population, that we see in the caterpillar data, is that they have a common shape. Each individual curve increases slowly, tend to reach a maximum and decreases rapidly.

## 3.2 Genetic Tradeoffs and Response to Selection

Each curve in the caterpillar data represents the mean growth rate across individuals in the same family, where a family is a set of offspring of the same parents. In general, a family could be any set of individuals with similar genotype. Thus, the visible variation in the 32 curves representing 32 different families is the between-family variation which corresponds to the genetic variation in the population. This genetic variation is important for evolutionary biologists because different modes of genetic variation in a population induce different evolutionary responses of the population under selection. In particular, evolutionary biologists are interested in three different modes of variation which correspond to three different genetic tradeoffs: faster-slower, hotter-colder, and generalist-specialist [Kingsolver et al. (2001)]. Each of these genetic tradeoffs would induce a different evolutionary response of a population under selection. Selection in this case refers to natural selection or also, as often used in agriculture, to human selection through breeding.

### 3.2.1 Three genetic tradeoffs: faster-slower, hotter-colder, and generalist-specialist

Figure 3.1 illustrates the three genetic modes of variation of interest shown in the first column of plots. For each genetic mode the corresponding evolutionary response of the population under selection is shown in the second column. For illustration, the selection criterion in Figure 3.1 is to select for families with higher performance (higher growth rate) at temperatures 37° Celsius, and the curves of those families are represented in the *darker shade* in this

figure. In the first row, the mode of variation is the **vertical shift** of curves or the tradeoff between slower growth rate at all temperatures and higher growth rate at all temperatures (**i.e Faster-Slower tradeoff**). In the second row, the mode of variation is the **horizontal shift** of curves or the tradeoff between higher growth rate at hotter temperatures and higher growth rate at colder temperatures (**i.e Hotter-Colder tradeoff**). In the third row, the mode of variation is the change in width or **generalist-specialist tradeoff** where a generalist corresponds to a TPC that has average growth rate in a wide range of temperatures and a specialist corresponds to a TPC that has high growth rate over a narrow range of temperatures. Biologists understand the generalist-specialist as a change in width of a common template shape with a constraint involving a *constant area under the curve*. Since the common shape function can take negative values, we translate this constraint in models 3.1 and 3.2 by keeping the positive area under the curve constant.

### 3.2.2 Evolutionary response to selection

As we see in Figure 3.1, when the mode of variation is vertical shift, and families with highest performance at 37° Celsius are selected then the population after selection will have higher average performance over all the temperature range compared to the population before selection. However, when the mode of variation is the horizontal shift or the generalist-specialist and families with highest performance at 37° Celsius are selected, then the population after selection will have higher average performance only at some temperatures and will have lower average performance at other temperatures compared to the population before selection.

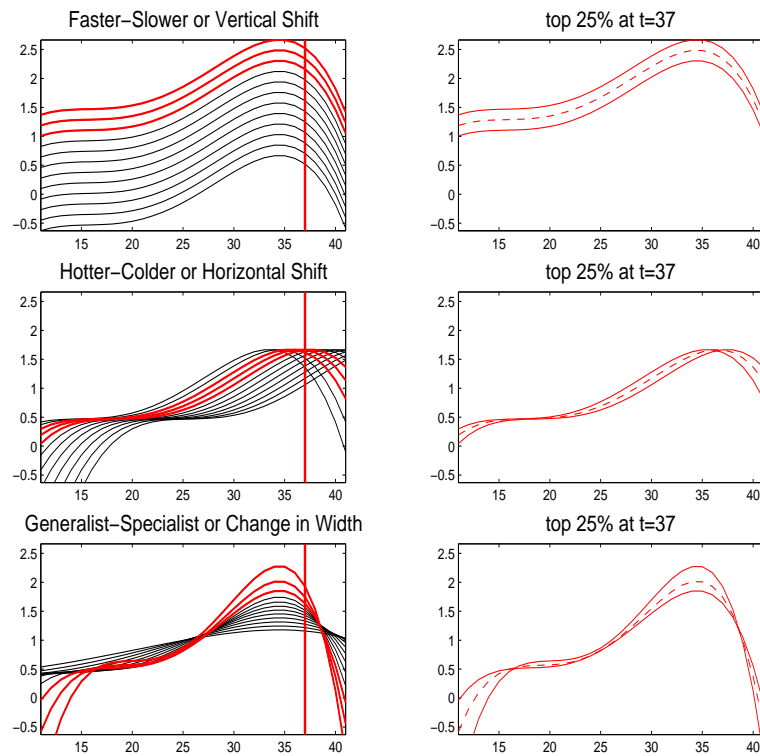


Figure 3.1: Illustration of three modes of genetic variation of interest and its consequence on evolutionary response under selection at 37 degrees Celsius. *Rows:* Illustrate the three genetic modes of variation of interest, from top to bottom: vertical shift, horizontal shift, and generalist-specialist. *Left column:* Genetic variation in the population before selection, the selection criterion is to select at 37° Celsius (the vertical line) the top 25% of families with highest performance (the darker shade curves). *Right column:* Genetic variation in the population after selection.

### 3.2.3 Principal component analysis applied to the data

One usual FDA method, for the analysis of variance, is to apply Principal Component Analysis (PCA) to the data set. A question of interest is, can PCA successfully decompose the variation in the data into the three biological modes of interest? Figure 3.2 shows the result of a PCA on the data, we see in the first row the data, the pointwise average of the data, and the mean residual. The following rows show projections in the eigen-directions ordered in descending percent of the mean residual sum of squares explained. So, in the first column we find the six principal components directions, i.e eigenvectors, and the percent of total variation explained by each direction. In the second column we find the projection of the data onto each direction, and finally in the third column we find the cumulative residuals.

The first principal component direction PC1 explains most of the variation (about 65%) in the data. The shape of PC1 indicates that most of the variation in the data occurs at the two last temperature measurements. This mode of variation tends to obscure interesting biological variation. In particular, does the PC1 direction correspond to a horizontal change in optimal growth rate, or a Hotter-Colder trade off, or does it correspond to a Generalist-Specialist tradeoff? PCA seeks linear directions of maximal variation which turns out to be different from the directions of biological interest.

## 3.3 Shape invariant models

We use a shape invariant model, which uses the common shape feature in the data and parameterizes the three modes of interest, as a first fitting step to the data. When the parameters of variation are specified, we illustrate in Chapter 4 how to decompose the variation.

Let  $z_{i,j}$  be the growth rate of family  $i$  measured at temperature  $j$ . Since the reaction

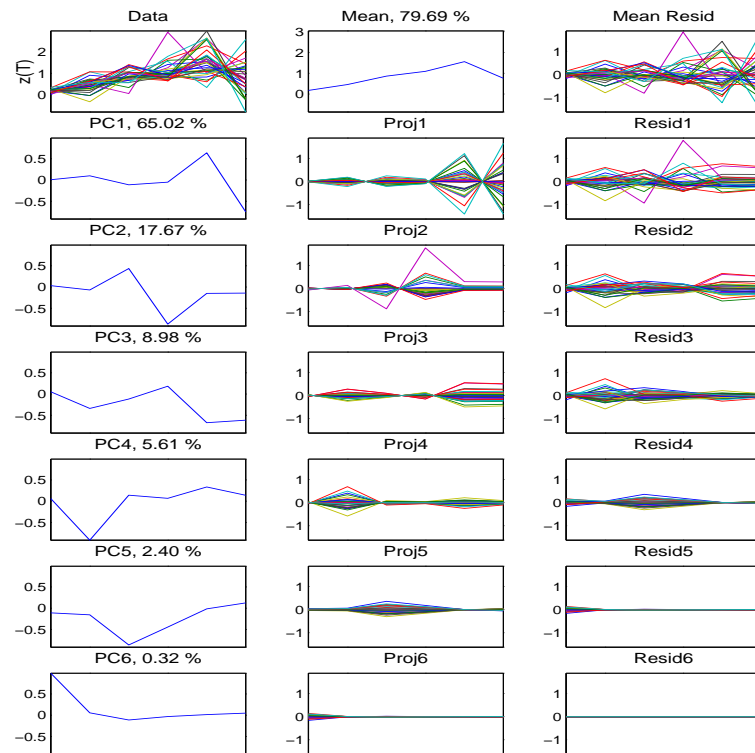


Figure 3.2: Results of the Principal Component Analysis on the caterpillars data. In the first row, from left to right, we see the family curves, the family mean growth rate, and the residuals (family curves from which we subtracted the mean growth rate). In rows 2 to 7, from left to right, we see a principal component direction, projection of the data onto each direction, and the cumulative residuals. The first principal component direction PC1 explains most of the variation (about 65%) in the data. The shape of PC1 indicates that most of the variation in the data occurs at the two last temperature measurements.

norm is continuous (i.e: the relationship between the growth rate and the temperature is continuous), the growth rate of a family  $i$  could be represented by a continuous curve  $z_i$  such as

$$z_{i,j} = z_i(t_j) + \epsilon_{i,j}; \quad j \in \{1, \dots, 6\}, \text{ and } i \in \{1, \dots, 32\}$$

where  $\epsilon_{i,j}$  are either errors from the measurements, or reflects environmental factors other than temperature on the growth rate (these factors were controlled, to the extent possible, by performing all experiments in a lab chamber). In this section, each mode is first modelled separately, and finally all modes are considered simultaneously. In all models,  $z$  represents a nonlinear common shape function.

### 3.3.1 Faster-Slower and Hotter-Colder

The vertical shift or Faster-Slower mode is modelled as

$$z_i(t) = z(t) + h_i.$$

The value of  $h_i$  characterizes the vertical shift of individual  $i$ ,  $1 \leq i \leq n$ . Note that no matter what the function  $z$  is, the vertical shift mode of variation is linear<sup>1</sup>. The horizontal shift or Hotter-Colder mode is modelled as

$$z_i(t) = z(t - m_i), \forall t \in T.$$

If we assume that  $z$  has a peak at 0, then the value of  $m_i$  characterizes the location of the TPC peak of individual  $i$ ,  $1 \leq i \leq n$ . Since  $z$  is nonlinear, this mode of variation is nonlinear.

---

<sup>1</sup>See in Chapter 4 for a visual representation of a linear and nonlinear mode

### 3.3.2 Generalist-Specialist

The generalist-specialist mode is not as easy to characterize as the faster-slower or the hotter-colder mode. We parameterize this variation with a width parameter  $w_i$  characterizing the width of family or individuals  $i$  as

$$z_i(t) = w_i z(w_i(t - m)), \forall t \in T. \quad (3.1)$$

Since  $z$  is nonlinear, then this mode of variation is nonlinear. We assume that the common shape  $z$  has a maximum  $c > 0$  at 0, i.e.  $\max_t z(t) = z(0) = c$ , and at least two zeros of opposite signs. Figure 3.3 illustrates the change of the function  $w.z(wt)$  as  $w$  changes. In this figure, we see that as  $w$  increases the *peak* of the function increases, the *width* decreases, and the positive area under the curve is the same. To define the positive area under the curve mathematically, let  $c_1$  be the largest negative zero of  $z$  and  $c_2$  be the smallest positive zero of  $z$ . The positive area under the  $z(t)$  curve is  $A = \int_{c_1}^{c_2} z(t)dt$ . Similarly, the positive area under the  $wz(wt)$  curve is  $A_w = \int_{\frac{c_1}{w}}^{\frac{c_2}{w}} w * z(wt)dt$ , by a change of variable  $t' = wt$  we have that

$$A_w = \int_{c_1}^{c_2} z(t')dt'$$

So,  $A_w = A$  for all  $w$ .

### 3.3.3 Model with three simultaneous variations

Now that all pieces of the puzzle have been laid in the previous models, we can model the simultaneous three modes of variation with the following 3-parameter shape invariant model

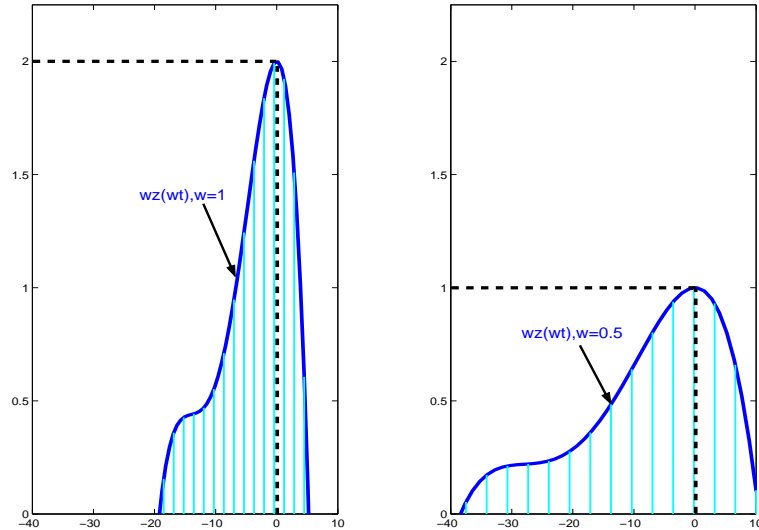


Figure 3.3: An example of a common shape  $z(t)$ , and the effect of  $w$  on the function  $w * z(wt)$ . *Left:*  $w * z(wt)$  for  $w = 1$ . *Right:*  $w * z(wt)$  for  $w = 0.5$ . The common shape function  $z$  has its maximum at 0, in this example the maximum value is  $c$  is 2. The positive area under the curve for  $w = 1$  and  $w = 0.5$  is shaded. As the parameter  $w$  changes, the function  $w * z(wt)$  has the same positive area.

where each mode is parameterized by one parameter

$$z_i(t) = w_i z(w_i(t - m_i)) + h_i; i \in \{1, \dots, 32\} \quad (3.2)$$

In this model,  $z$  represents the common shape of the curves. The parameters  $(h_i, m_i, w_i)$  are respectively the height, the location and the width parameters characterizing the variation of the curve of family  $i$  from the common shape  $z$  along the modes of interest. More precisely, the height  $h_i$  parameterizes the vertical shift mode of variation, the location  $m_i$  parameterizes the horizontal shift mode of variation and the width  $w_i$  parameterizes the generalist-specialist mode of variation. As in the previous models,  $z$  has a maximum  $c > 0$  at 0 and at least two zeros of opposite signs. Equation 3.2 holds true for all temperatures  $t$ , and in particular at the six recorded temperatures  $t_1, \dots, t_6$ .

Since both the horizontal shift and generalist-specialist are nonlinear modes, the space of

variation is not linear but is a curved manifold of dimension three in  $\mathbb{R}^6$ . The new method for decomposition and quantification of the variation in the data is presented in Chapter 4 and Chapter 5. However results of the fit of the model to the data and decomposition of variation are presented in this Chapter in Section 3.5.

### 3.4 Choice of the template shape

In Chapter 4 and Chapter 5, we assume that the template shape is known. In practice the template shape is estimated from the data. The larger the number of temperature measurements, the better our estimate of the common shape can be. Several nonparametric methods were proposed to estimate curves of common shape in [Kneip and Gasser (1988)], [Kneip et al. (2000)], [Wang and Gasser (1997)], and [Wang and Gasser (1999)]. Since we have 6 growth rate measurements for each curve, we fitted the model 3.2 with a template shape  $z$  as a polynomial of degree four. The parameters of this polynomial and the parameters  $(w_i, m_i, h_i)$  for each curve were optimized to minimize the Sum of Squared Errors (SSE) where

$$SSE = \sum_{i=1}^n \sum_{j=1}^d \|z_{i,j} - w_i z(t_j - m_i) - h_i\|^2 \quad (3.3)$$

if the families are not weighted by sample size and

$$SSE = n * \sum_{i=1}^n \sum_{j=1}^d \frac{n_i}{\sum_k^n n_k} \|z_{i,j} - w_i z(t_j - m_i) - h_i\|^2 \quad (3.4)$$

if the families are weighted by sample size. The main advantage of a polynomial template shape for this data is its simplicity, which results in a quickly convergent fitting algorithm, a unique solution, and avoids over-fitting. However, polynomials of lower degree do not fit a rapid change in the curves very well. Details of the fitting algorithm to the data are in

Appendix Section 6.1.

## 3.5 Results

The results of the fit to the data with the chosen template shape are presented in this section. We first describe several visual assessments of the fit and then discuss the table which presents the decomposition.

### 3.5.1 Fitted curves

In Figure 3.4 we see the results of the optimization, i.e the 32 polynomials of degree 4 fit for the curves using model 3.2. Since both plots in this Figure are on the same scale, we can see that the range of variation in the polynomial fits closely matches the range of variation in the data. Figure 3.5 gives a closer look at how much of the variation is explained by the model. In this figure, the data are displayed in the *lighter shade* and the discretized projection in the *darker shade*, we see that the fits are close to the data and that our model explains most of the variation in the data.

Figure 3.6 and Figure 3.7 check the one-to-one fit in which each data curve in lighter shade is compared to its smooth polynomial fit in darker shade. We see that the model fits very closely for most families. Note that in Figure 3.6 some families (for example, Families 5, 9 and 14) seem to have one main peak and one lower peak. We see in the first line of Figure 3.7 that the polynomial doesn't fit well for families 22, 23 and 24. These families have a TPC that increases and/or decreases rapidly and the polynomial fit doesn't model this rapid change very well.

Figure 3.8 gives an important visual assessment of the residuals in the model and how well the template shape fits the data. In this figure, each data curve  $i$  was *rescaled* with respect to the

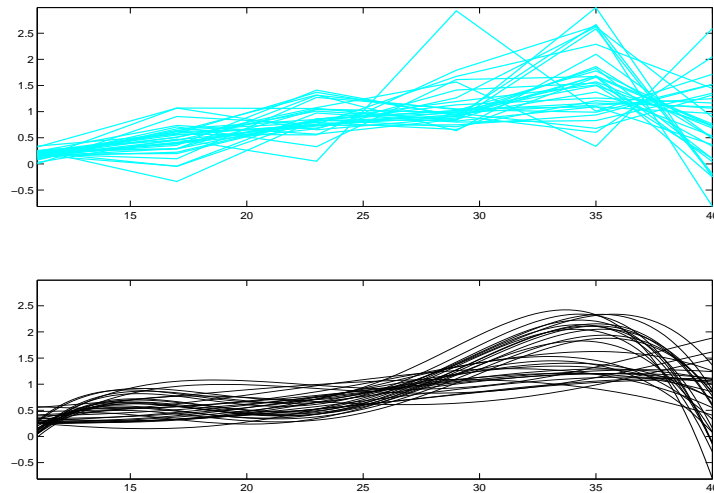


Figure 3.4: *Above:* The caterpillar’s TPCs. *Below:* Fitted polynomials of degree 4 to the data using Model 3.2. Note that both plots have the same axis scale, and that the variation in the fit seems to closely match the variation in the data.

three estimates of the parameters  $(w_i, m_i, h_i)$  and all the rescaled discrete data is compared to the continuous common shape  $z(t)$ . More precisely, each rescaled growth rate  $z_{i,j}/w_i$  was plotted at the rescaled temperature  $T_{i,j}$  such as  $T_i = w_i(t_j - m_i) + h_i$ . We see in this figure that the fitted common shape is a good approximation of the common shape of the curves, and that not very much of the variation remains after fitting the model.

### 3.5.2 Decomposition of the variation

The decomposition of the variation into the modes of interest by our Template Mode of Variation (TMV) method, displayed in Table 3.1, confirms our visual assessment of the fit. As we see in this table, the model explains more than two thirds (68.28%) of the total variation in the data. This surprisingly high percentage shows the importance of these biologically interpretable modes of variation. A surprising result for evolutionary biologists was that most of the variation was explained by the two nonlinear modes of variation: the generalist-specialist (39.19%) and the horizontal shift (17.12%) and the vertical shift explained only

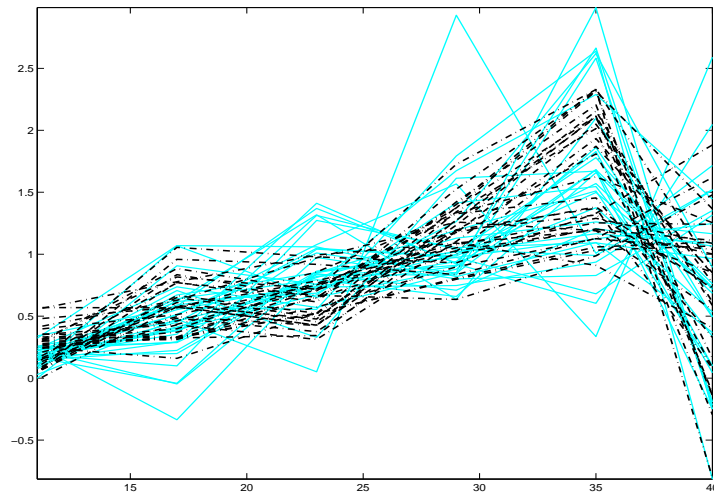


Figure 3.5: Caterpillar TPCs and fitted curves on the same scale. *Lighter shade curves*: The caterpillar's TPCs. *Darker shade curves(-.-)*: Polynomials fits discretized at six temperatures. The fit by Model 3.2 explains most of the variation in the data.

Table 3.1: Decomposition of the total variation in caterpillars data using Template Mode of Variation (without sample size correction)

Mode of variation	$\widehat{RSS}(\%)$
Generalist Specialist	39.19
Horizontal Shift	17.12
Vertical Shift	11.28
Model total	68.28
Error	31.72

about 11.26% of the total variation in the data. So, according to our initial motivation for this analysis described in Section 3.1, we conclude from this result that **selecting families with highest performance in a certain range of temperatures will not result after selection in a population with high performance over all ranges of temperatures.**

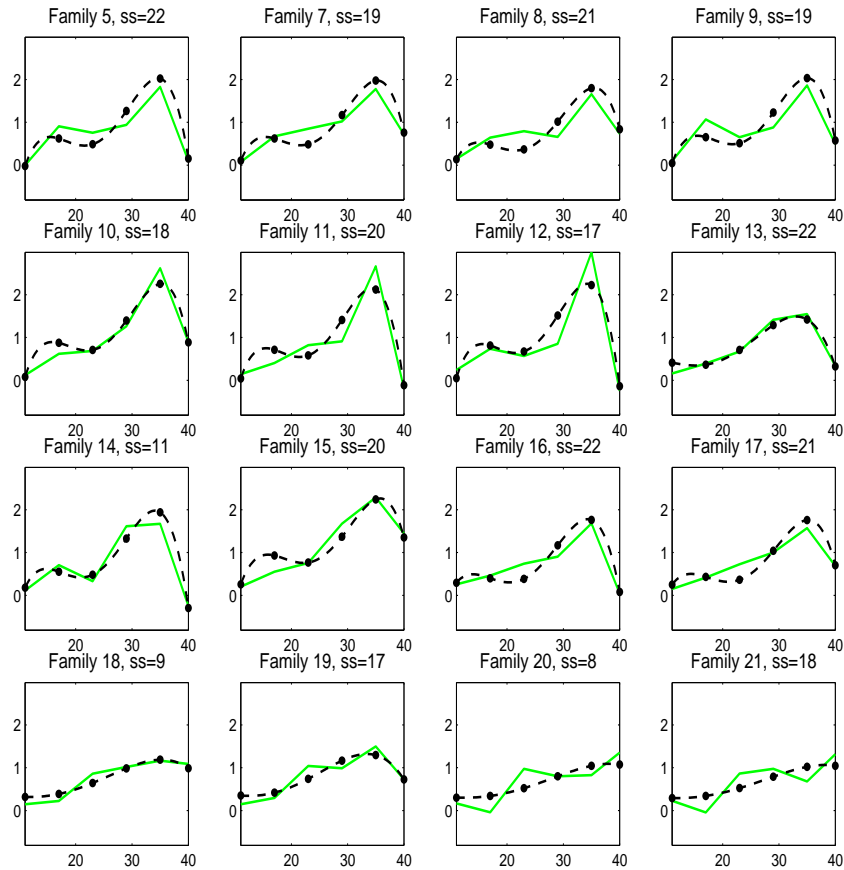


Figure 3.6: The 16 subplots represent families 5-21 (the title line for each subplot has the family label given by the experimenter and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that some families have only one peak (families 13, 18, and 19) and most families have a high peak and a low peak. The polynomial fits the data very closely.

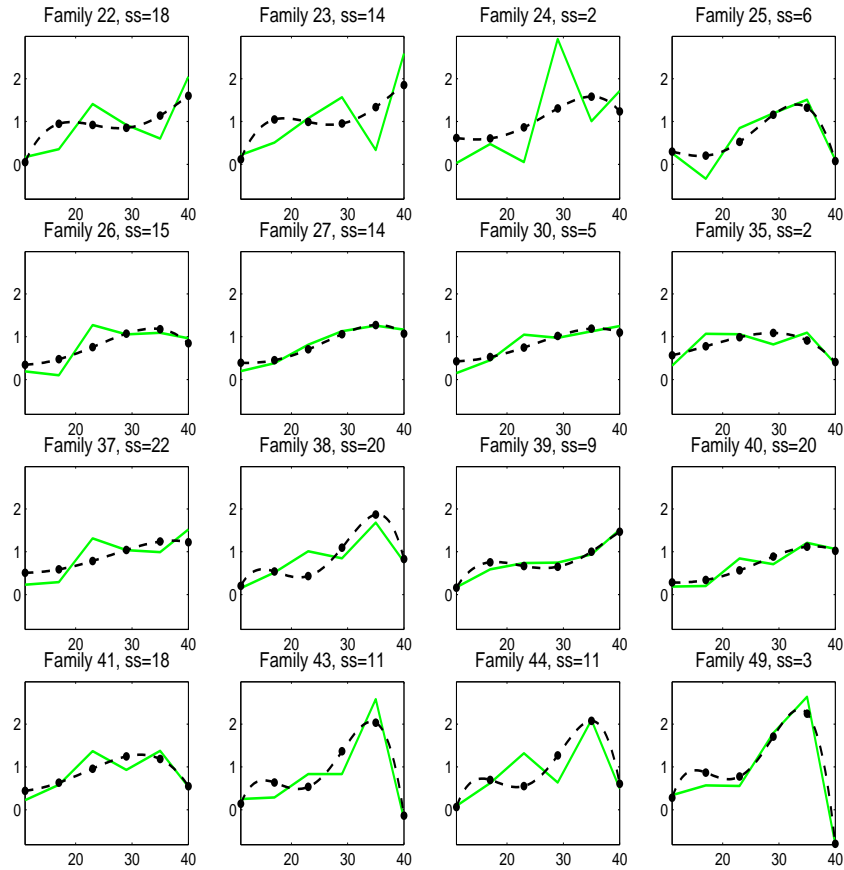


Figure 3.7: Individual fits of caterpillar TPCs. The 16 subplots represent families 22-49 (the title line for each subplot has the family label given by experimenter and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial doesn't fit well TPC's of families with rapid increase and/or decrease (i.e families 22, 23 and 24). For other families, the polynomial fits the data very closely.

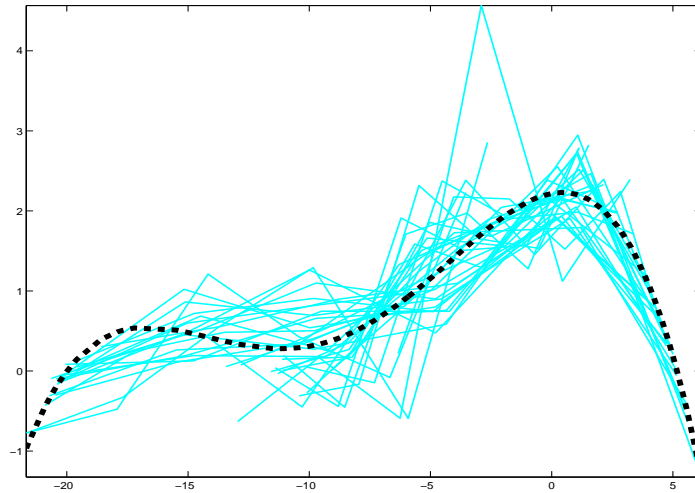


Figure 3.8: Rescaled caterpillar TPCs compared to the fitted common template shape  $z(t)$ . *Lighter shade curves*: Rescaled TPCs. Each TPC was standardized with respect to the estimates of height, location and width parameters from the fit to Model 3.2. *Darker shade curve (-)*: The fitted common shape  $z(t) = 10^{-2}(-2 * 10^{-2}t^4 - 0.61t^3 - 5.77t^2 + 5.45t + 256)$ . We see that the fitted common shape fits well the common shape of the TPCs and the residual variation around the common shape after fitting the model is small.

### 3.6 Discussion

We presented in this Chapter a decomposition and quantification of biological modes of variation for a set of TPCs, which was not previously possible, using the new method we developed: TMV. As we see in Subsection 3.6.1, we applied this method to a different set of TPCs, collected on wasps, and we obtained a similar decomposition with similar conclusions for evolutionary biologists. TMV can be readily applied to other continuous reaction norms as well as to other functional data. Future work is to introduce some improvements of the method and test sensitivity to violation of some assumptions. In particular, since the wasp data that we see in Figure 3.16 have more temperature measurements than the caterpillar data, we fitted the data with two polynomials of different degrees and compared the decomposition in Subsection 3.6.1. Such comparison gives an insight on the sensitivity of the decomposition to

the template shape. In Subsection 3.6.2, we explore the fitted parameters in the caterpillar data. Finally, in Subsection 3.6.3 we compare the within to the between family variation. Further goals are to generalize TMV to decompose the variation in the data into any modes of variation of interest that could be represented in a shape invariant model and to apply TMV to decompose variation in any predetermined modes in functional data.

### 3.6.1 Sensitivity of decomposition to template shape

As we see in Figure 3.6, 3.7, and 3.8, the fitted polynomial is a good fit to the data, and further work is to check sensitivity of this decomposition to the template shape. We see in Section 3.6.4 results of this method on a different set of TPCs collected on wasp. Because the trait was measured at seven temperatures in the wasp data, we fit the model with a polynomial of degree 4 and also with a polynomial of degree 6. Although, as we can predict, the polynomial of degree 6 explains more of the variability in the data, the decomposition of the variation of the two fits is comparable. As a polynomial of degree higher fits always the data better (i.e have a lower  $SSE$ ), we can base our template shape selection and decomposition on minimizing  $SSE$  and penalizing for the total number of parameters in the fit. It would also be of interest to check the decomposition for more temperatures and for a varying window of temperature measurements.

### 3.6.2 Parameters of Variation, Family Clusters

The relationship between the parameter estimates are shown in Figure 3.9. Since each parameter represents a mode of variation, the distribution of the parameters and the relationship between the parameters of variation are an indicator of how much variation there is along each mode and how they vary together. We see in Figure 3.9 two different groups that are

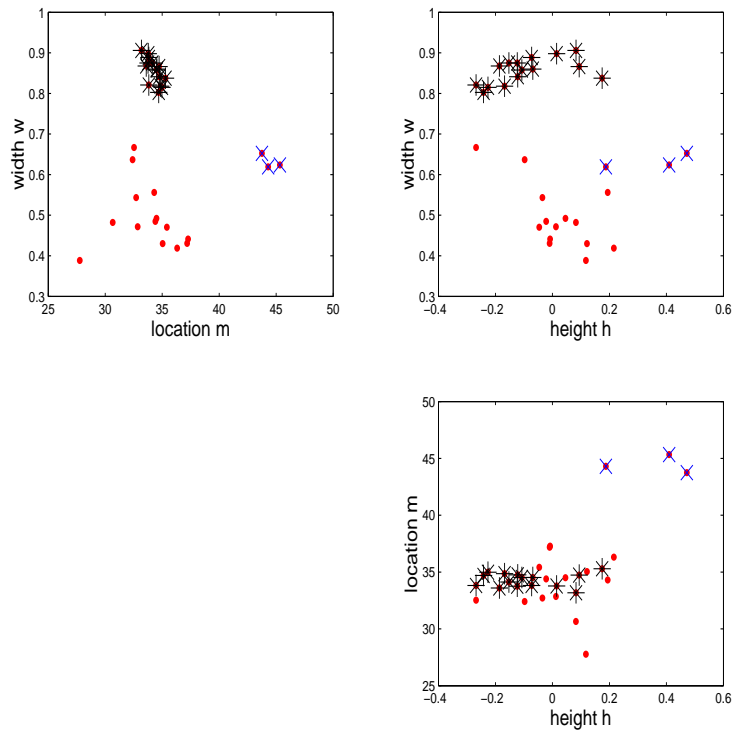


Figure 3.9: Values and Relationship between the fitted parameters ( $w, m, h$ ) for the caterpillar data. We identify two groups in particular: the first (star) is the group with high width parameter. The second (cross) is the group with high location parameter.

separate from most families, one with a high width parameter and the other with a high location parameter. Figures 3.10 and 3.11 are subsets of Figures 3.6 and 3.7 which highlight the curve fits for these families.

### 3.6.3 Within Family Variation

Because families in the caterpillar data do not all have the same sample size, the fit and decomposition was redone by incorporating a sample size correction to the different  $RSS$  and, as we see in Table 3.2, very similar results to the ones without the sample size correction were obtained. Since each curve in this data is the arithmetic *average* growth rate among offspring in a family, further work would be to use the estimate of the Fréchet mean among

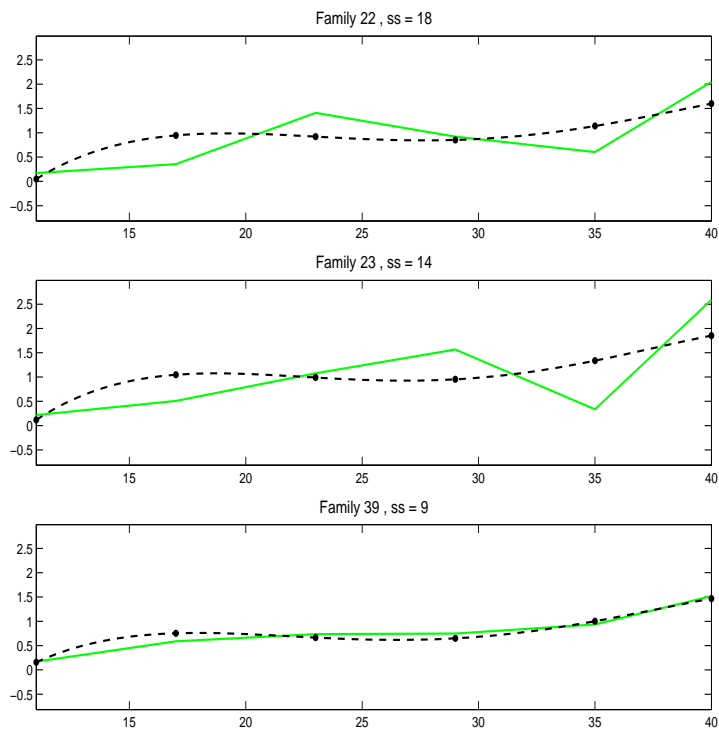


Figure 3.10: Data and fitted polynomial for a subset of caterpillar families with higher location parameter  $m$ . These are a subset of Figure 3.7 which highlight the three families (22, 23, and 39) with higher fitted location parameter  $m$ . The fitted location parameters for these families were the group (cross) in Figure 3.9. We see that the polynomial doesn't fit well the dip in the data (between the fourth and sixth measurement).

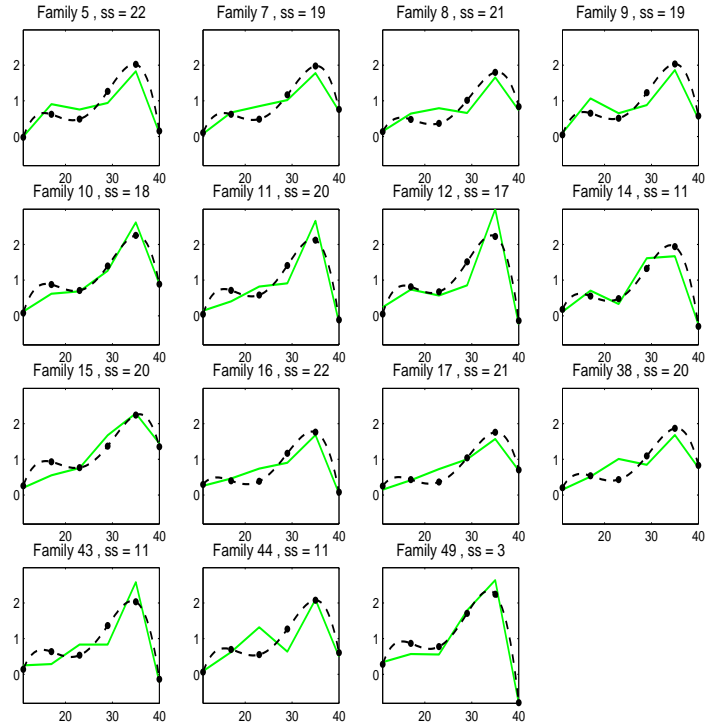


Figure 3.11: Data and fitted polynomial for a subset of families with higher width parameter. These are a subset of Figure 3.6 and Figure 3.7 which highlight the group of families with higher fitted width parameter  $w$ , these families are the group (star) in Figure 3.9. We see that the polynomial fit well this group of data and that all curves show two maxima, and high value at the second maximum.

Table 3.2: Decomposition of the total variation for the caterpillars data using Template Mode of Variation (with sample size correction, see equation 3.4)

Mode of variation	$\widetilde{RSS}(\%)$
Generalist Specialist	38.04
Horizontal Shift	16.23
Vertical Shift	12.63
Model total	66.90
Error	33.10

each family rather than the arithmetic average and to compare the between family variation to the within family variation along the modes of interest. Figures 3.12 and 3.13, show the variation within each family and the sample size within each family. Note that the axes are the same to make the plots comparable.

For each family, the within standard deviation at each temperature is compared to the between family standard deviation at each temperature in Figures 3.14 and 3.15. We see that the between standard deviation is high for all families regardless of the sample size. The between family variation increases with the temperature and so does the within family variation. However, the between family variation increases steadily whereas, in most families, the variation at the fifth temperature measurement is the highest.

#### 3.6.4 Application to other reaction norm curves

Since the results of the decomposition in this data were surprising to evolutionary biologists, analysis of similar TPCs using TMV was performed to check if the low contribution of the vertical shift mode of variation is a common trend. The data in Figure 3.16 represents a set of TPCs collected on female wasps. The response of interest in this data is the wasp velocity (how fast the wasp walks a fixed distance), see [Gilchrist (1996)] for a detailed description of the design and the data. The wasp velocity was measured at 7 temperatures ranging from

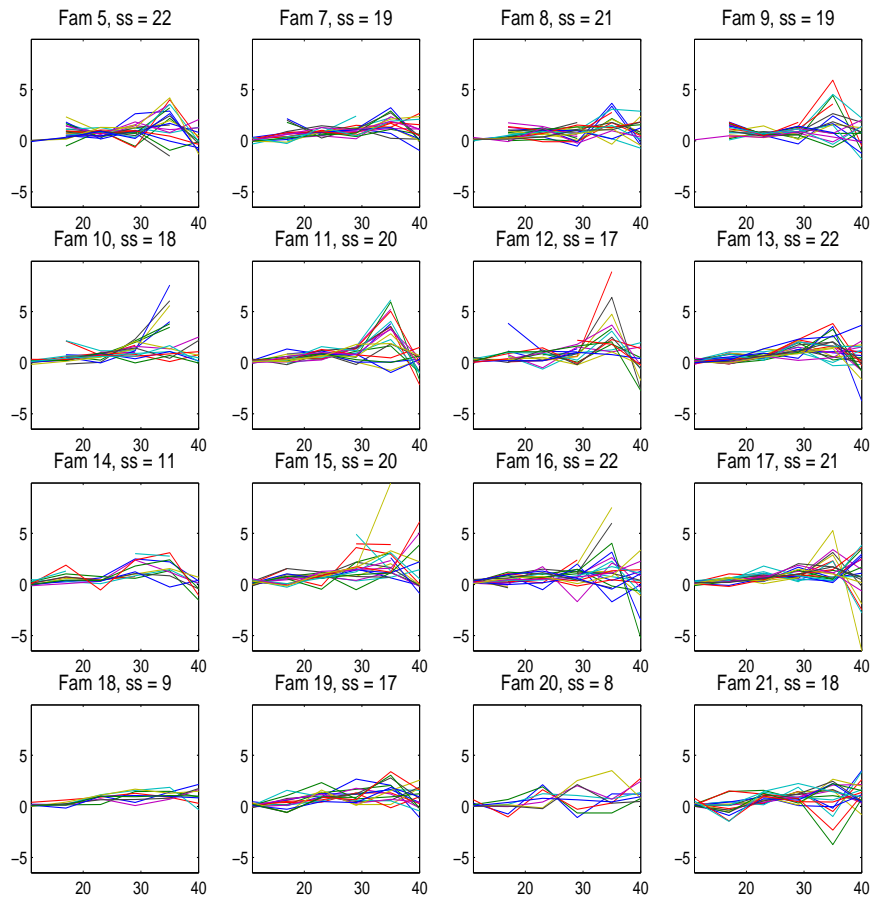


Figure 3.12: Individual TPCs in caterpillar families 5-21 (the title line for each subplot has the family label given by experimenter and the sample size denoted by  $ss$  for each family). Each subplot shows the within family variation, note that the axis scale are the same to make the families comparable. Sample sizes vary from 8 to 22, and most families have more than 15 individuals. We see in families (5-10) missing values at temperature  $11^{\circ}$  C. As in the between family variation, we see that most of the within family variation is on the warmer temperatures. Within family variation is non-negligible and it varies widely from one family to the other

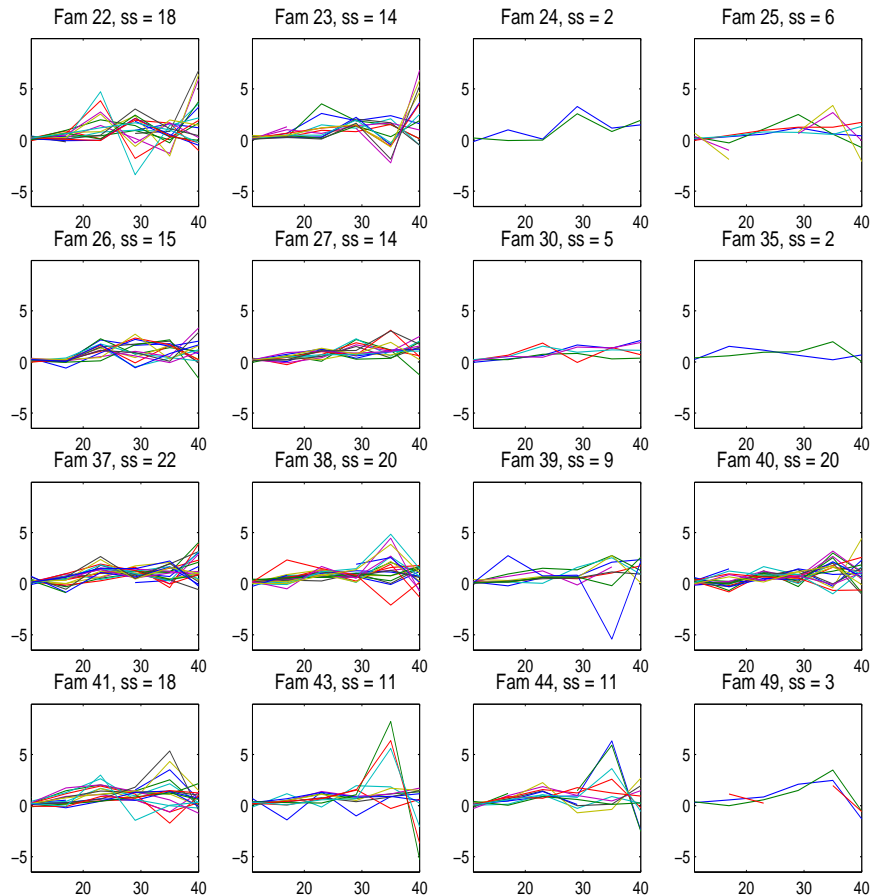


Figure 3.13: Individual TPCs in caterpillar families 22-49 (the title line for each subplot has the family label given by experimenter and the sample size denoted by  $ss$  for each family). Each subplot shows the within family variation, note that the axis scale are the same to make the families comparable. Sample sizes vary from 2 to 22, and some families (24, 25, 30, 35, 49) have fewer than 7 individuals. As in the between family variation, we see that most of the within family variation is on the warmer temperatures. Within family variation is non-negligible and it varies widely from one family to the other.

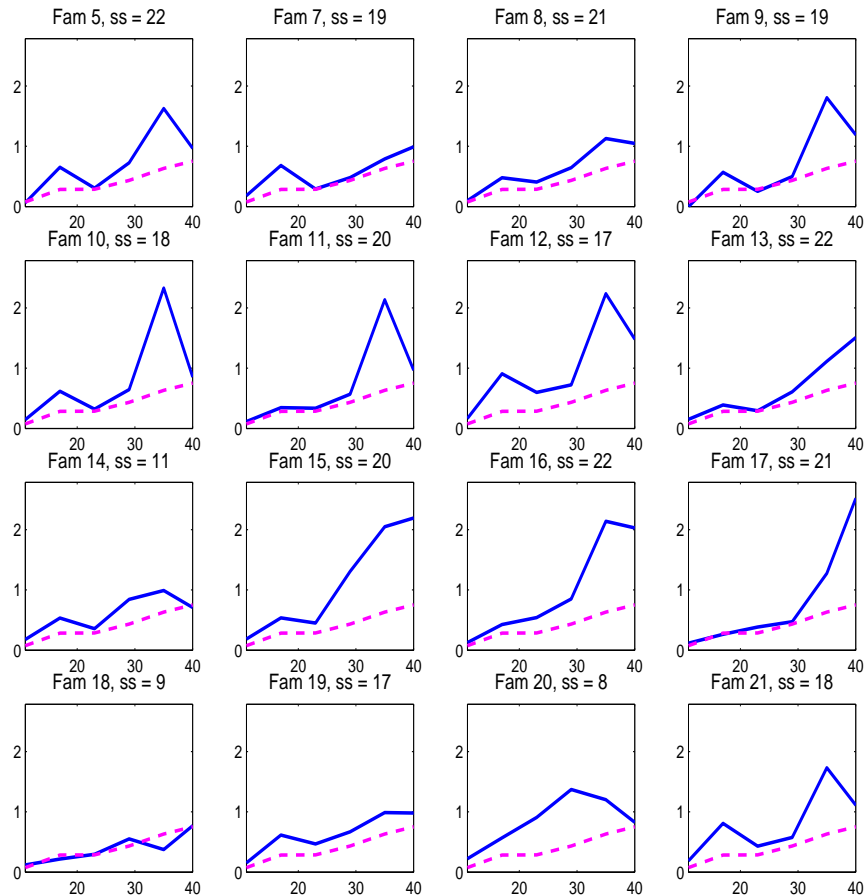


Figure 3.14: Standard deviation at each temperature within families 5-21 (*solid line*) and standard deviation at each temperature between families (*dashed line*). The title line gives the family tag given by the experimenter as well as the sample size (ss) in each family. The within family variation is non-negligible, it is much higher than the between family variation at higher temperatures for most families, regardless of the sample size, except for families 7, 8, 14, 18, and 19.

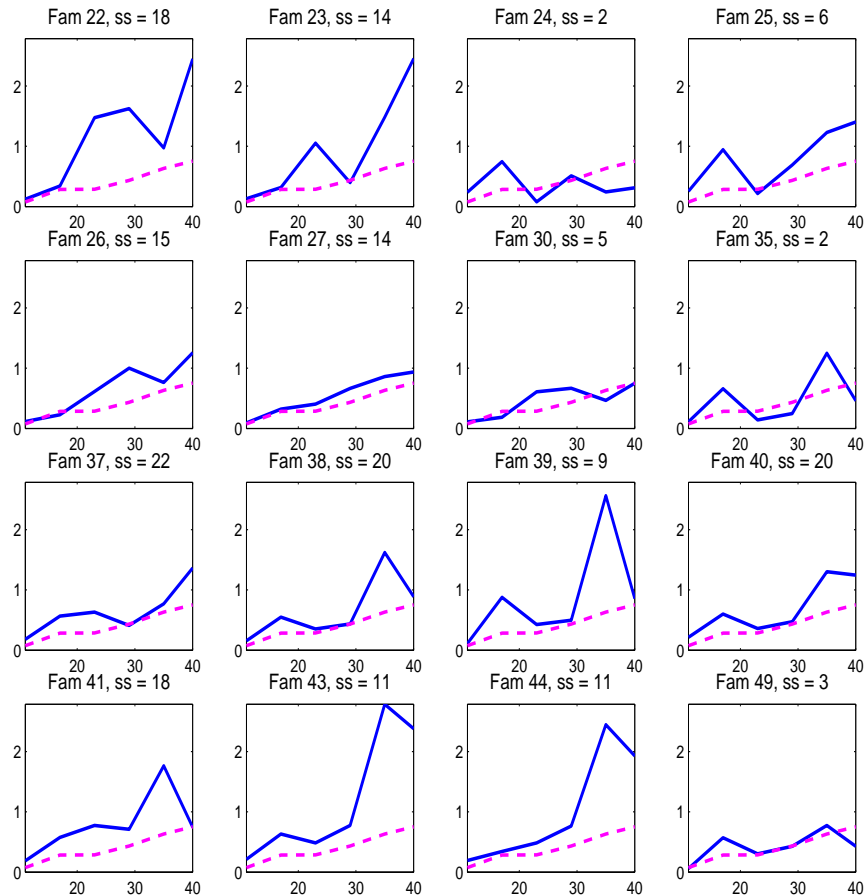


Figure 3.15: Standard deviation at each temperature within families 22-49 (*solid line*) and standard deviation at each temperature between families (*dashed line*). The title line gives the family tag given by the experimenter as well as the sample size (ss) in each family. The within family variation is comparable to the between family variation for families 24, 26, 27, 30, 37, 49. It is considerably higher than the between family variation at higher temperatures, regardless of the sample size, especially for families 22, 23, 38, 39, 43, and 44.

Table 3.3: Decomposition of the total variation in wasp data using Template Mode of Variation (with sample size correction)

<b>Mode of variation</b>	$\widetilde{RSS}(\%)$
Generalist Specialist	20.12
Horizontal Shift	16.7
Vertical Shift	36
Model total	72.82
Error	27.18

12°C to 36°C. This data has 87 families, which is higher than in the caterpillar data. Note however that the number of individuals in each family, shown in the title line of subplots in Figures 3.18, 3.19, 3.20, 3.21, 3.22, and 3.23, is smaller than in the caterpillar data. The numerical results of the decomposition are in Table 3.3. We see in Figure 3.17 that the model explains most of the variability in the data. A closer look to the one to one fit in Figures 3.18, 3.19, 3.20, 3.21, 3.22, and 3.23 shows that the polynomial fit the data very closely. As we see in Figure 3.24 and Figure 3.25, the fits of a polynomial of degree 4 and of a polynomial of degree 6 are almost equivalent. Table 3.3, shows the decomposition of variation. Note that, for this data set also, the model explains a lot of the variability in the data (73%), and the two non-linear modes account for a lot of the variability (37%). However, a noticeable difference in this case is that the vertical shift accounts for almost half of the variability explained by the model.

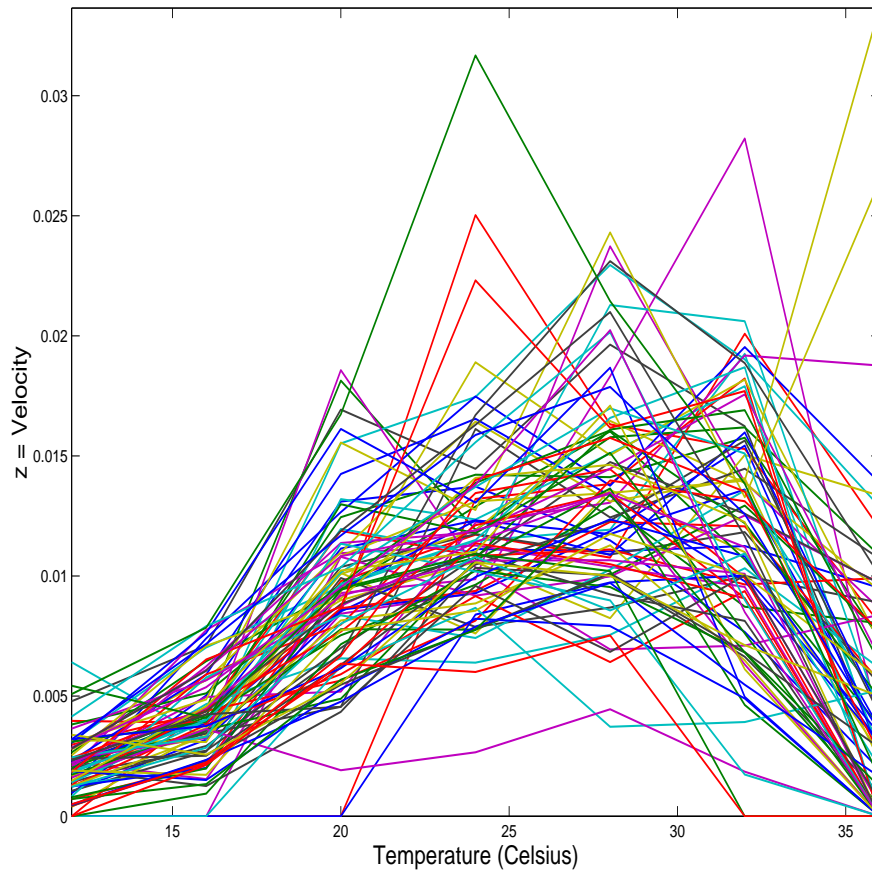


Figure 3.16: TPCs of female wasps. The velocity of each wasp was measured at 7 different temperatures ranging from  $12^{\circ}\text{C}$  to  $36^{\circ}\text{C}$ . There are 87 curves corresponding to 87 different families.

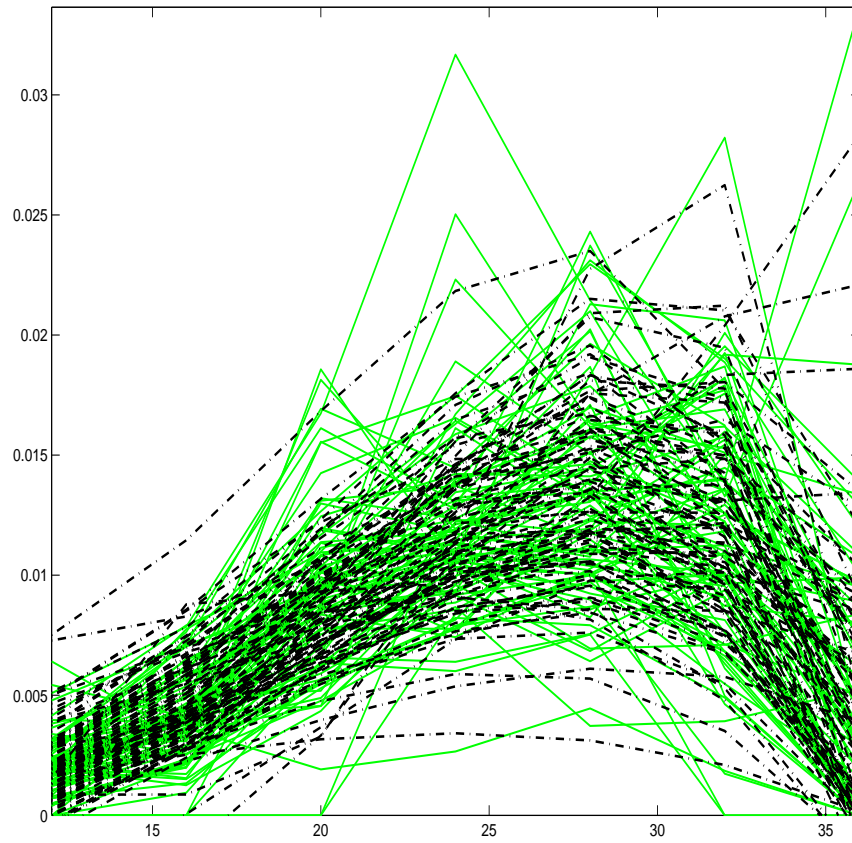


Figure 3.17: TPCs of wasps and fitted curves on the same scale. *Lighter shade curves*: The wasp's TPCs. *Darker shade curves(-.-)*: Polynomials fits discretized at seven temperatures. The fit by Model 3.2 explains most of the variation in the data.

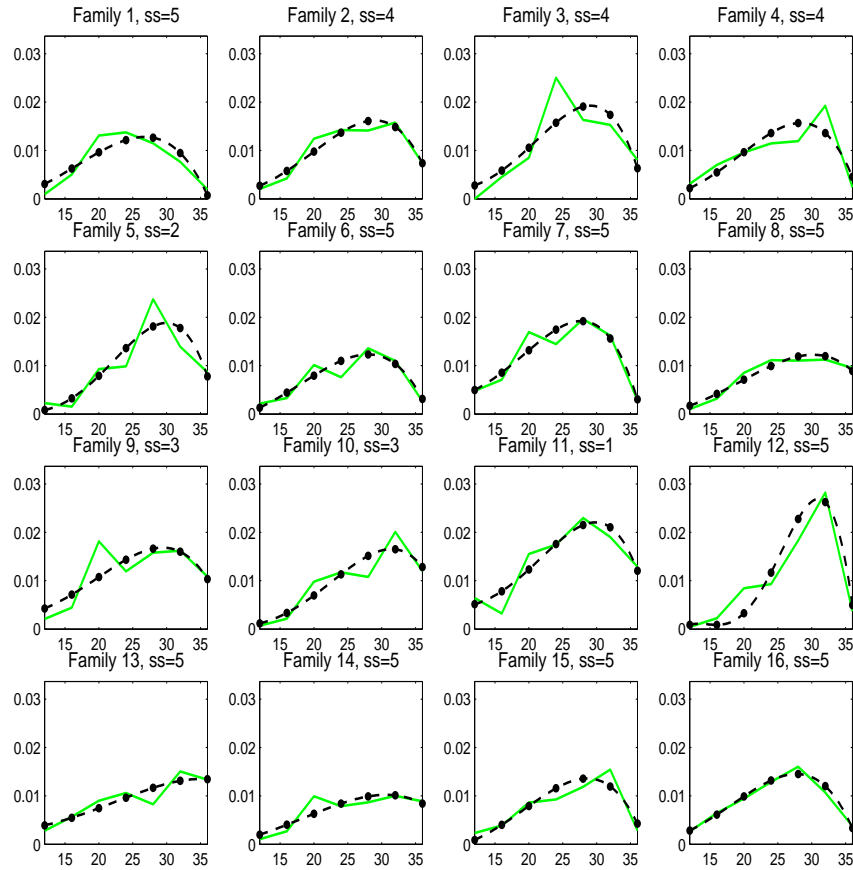


Figure 3.18: Individual fits for wasp data. The 16 subplots represent families 1-16 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

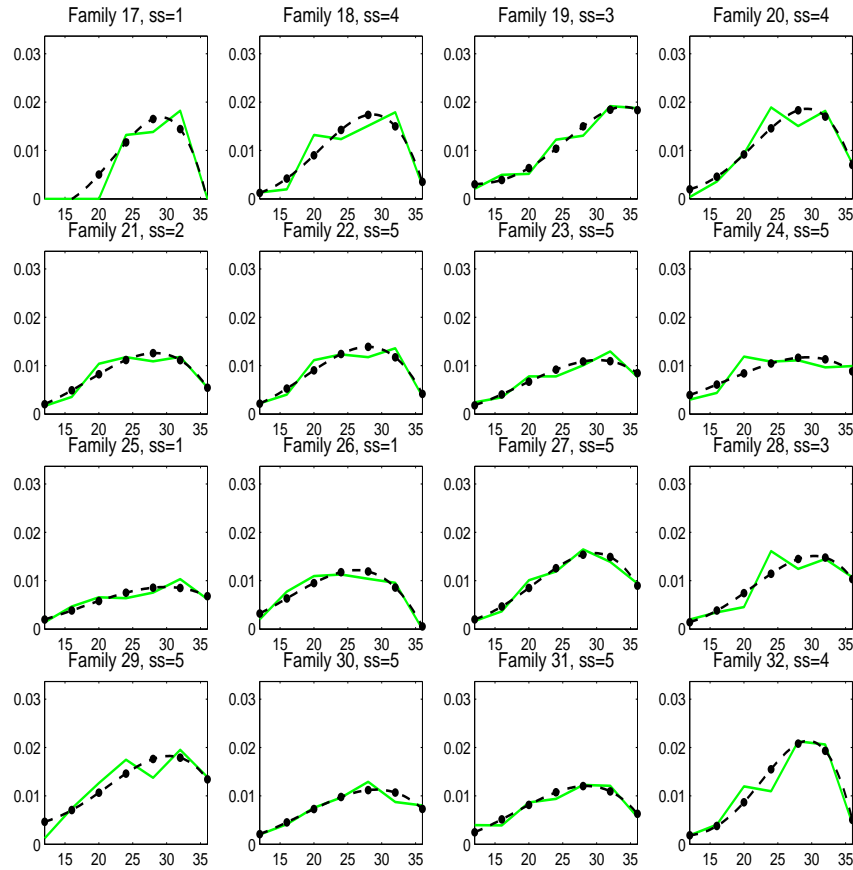


Figure 3.19: Individual fits for wasp data. The 16 subplots represent families 17-32 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

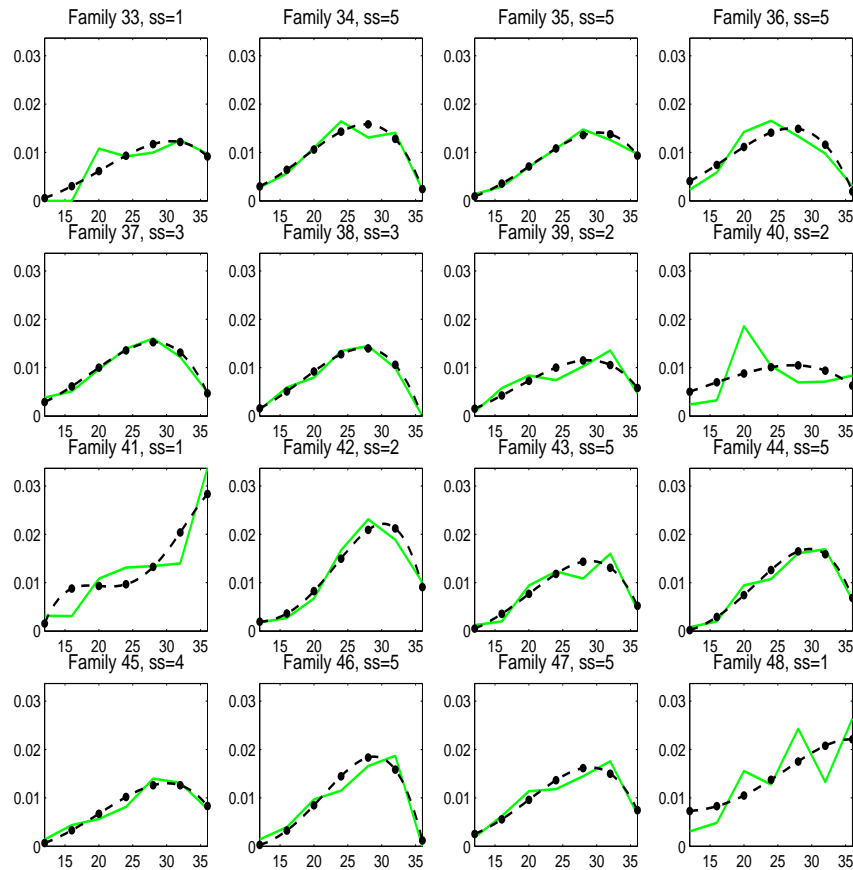


Figure 3.20: Individual fits for wasp data. The 16 subplots represent families 33-48 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

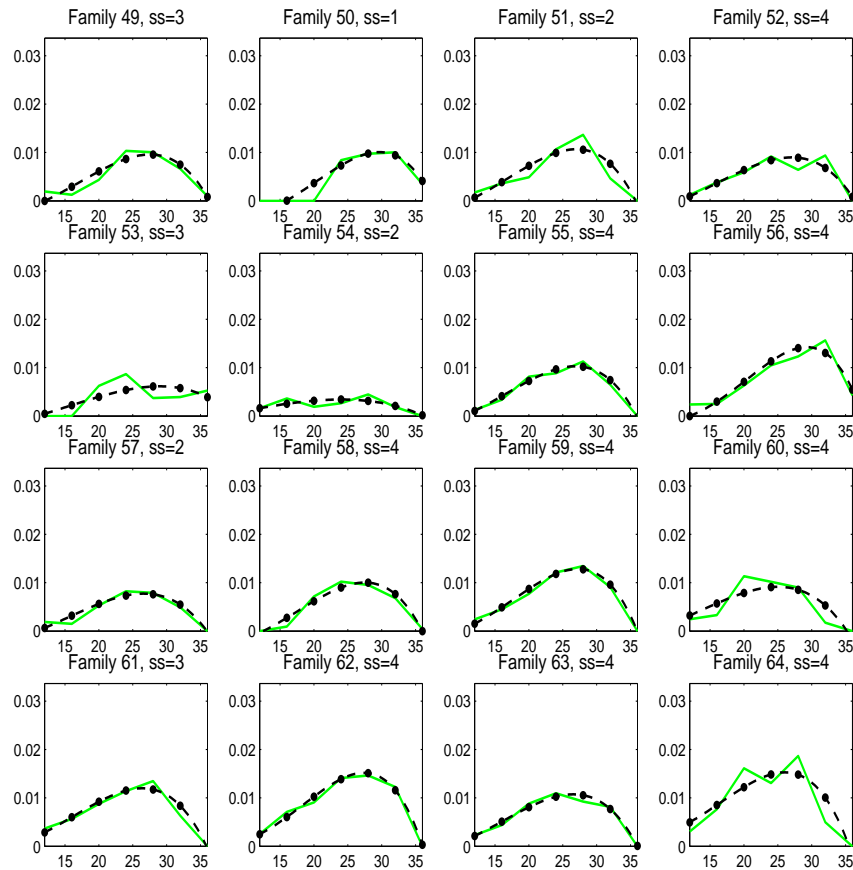


Figure 3.21: Individual fits for wasp data. The 16 subplots represent families 49-64 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

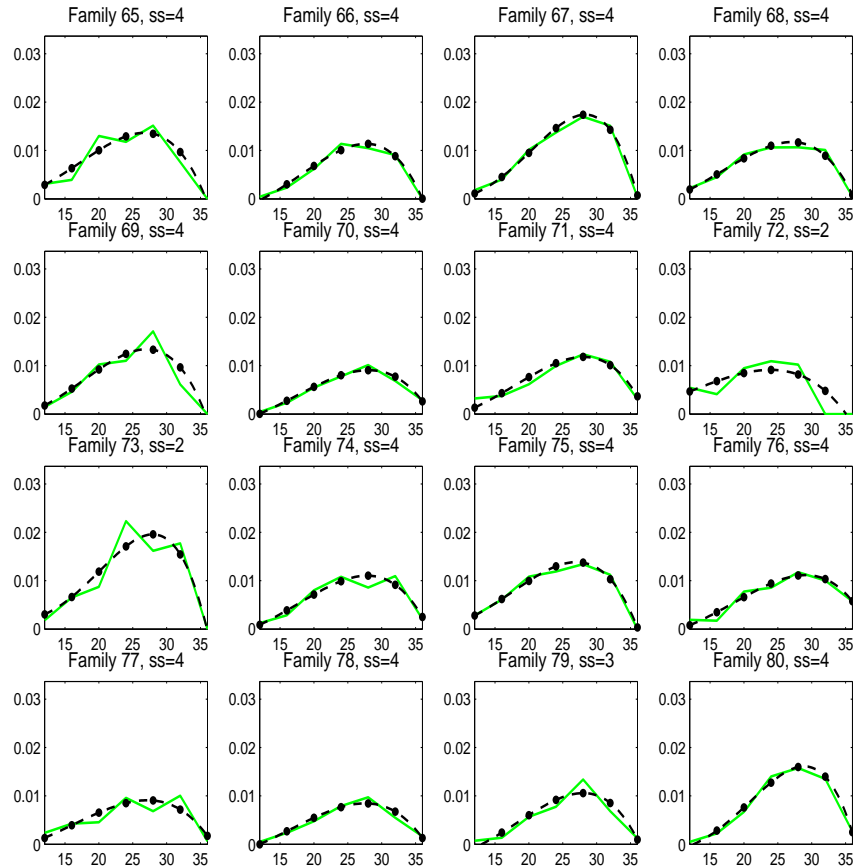


Figure 3.22: Individual fits for wasp data. The 16 subplots represent families 65-80 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

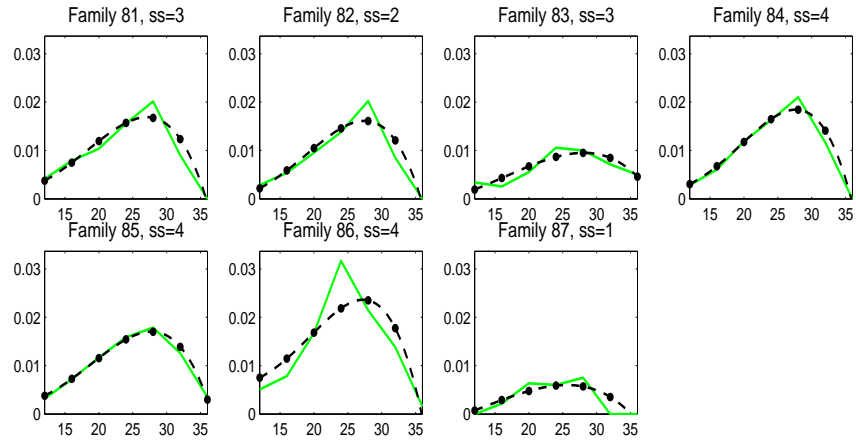


Figure 3.23: Individual fits for wasp data. The 7 subplots represent families 81-87 (the title line for each subplot has the family label and the sample size denoted by  $ss$  for each family). In each subplot, the *lighter shade curve* is the mean curve for one family, and the *darker shade curve* is the polynomial fit for that family. Note that the polynomial fits very closely all TPCs.

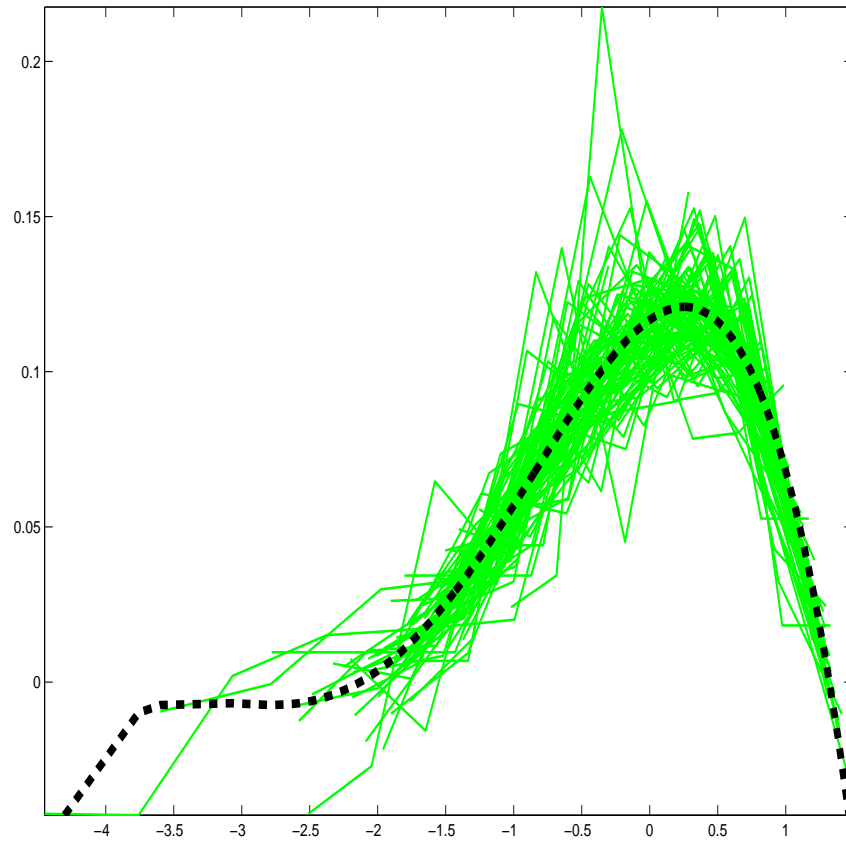


Figure 3.24: Rescaled TPCs of wasp data compared to the fitted common template shape, a polynomial of degree four. *Lighter shade curves*: Rescaled TPCs. Each TPC was standardized with respect to the estimates of height, location and width parameters from the fit to Model 3.2. *Darker shade curve (-)*: The fitted common shape is a polynomial of degree four. We see that the fitted common shape fits well the common shape of the TPCs and the residual variation around the common shape after fitting the model is small.

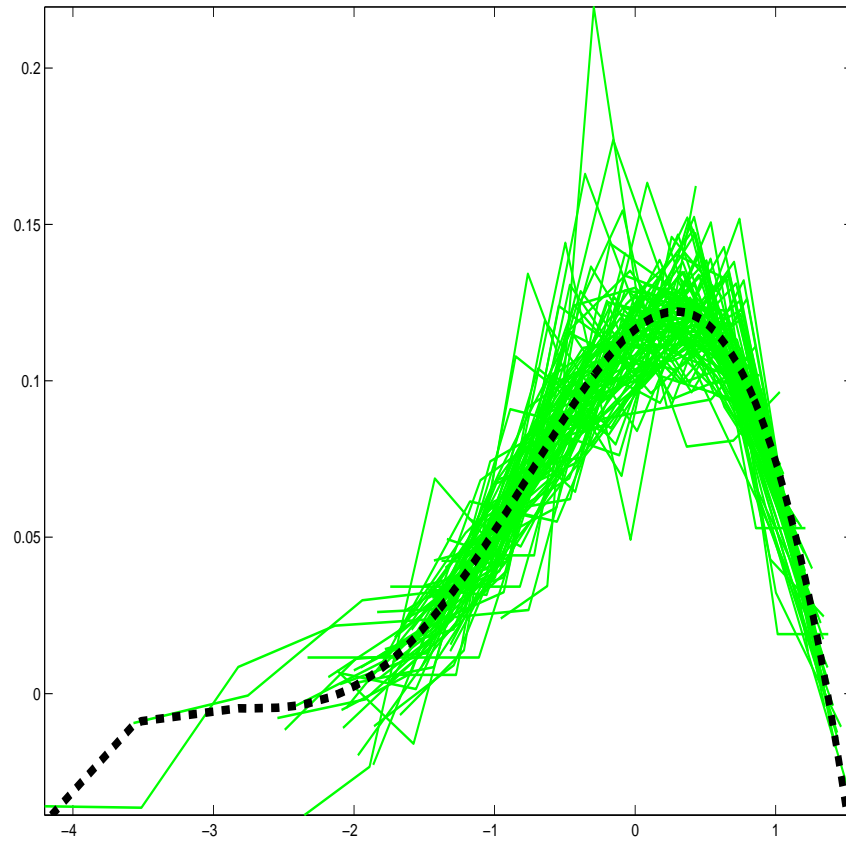


Figure 3.25: Rescaled TPCs of wasp data compared to the fitted common template shape, a polynomial of degree six. *Lighter shade curves*: Rescaled TPCs. Each TPC was standardized with respect to the estimates of height, location and width parameters from the fit to Model 3.2. *Darker shade curve (-)*: The fitted common shape  $z(t)$  is a polynomial of degree six. We see that the fitted common shape fits well the common shape of the TPCs and the residual variation around the common shape after fitting the model is small. Note also that this fit is identical to the fit by polynomial of degree 4 shown in Figure 3.24