Submitted to the Annals of Statistics

PETER HALL'S MAIN CONTRIBUTIONS TO DECONVOLUTION

By Aurore Delaigle*

University of Melbourne

Peter Hall died in Melbourne on January 9, 2016. He was an extremely prolific researcher and contributed to many different areas of statistics. In this paper, I talk about my experience with Peter and I summarise his main contributions to deconvolution, which include measurement error problems and problems in image analysis.

1. My experience with Peter. I met Peter for the first time at a workshop in Belgium in 2001 when I was a PhD student. I was a nobody and he was a god in Statistics, but he took the time to discuss my work, which left me with the impression that he was a very nice person. That was the general experience that young researchers had when meeting Peter. Even if you were very insignificant in a department with many senior statisticians, he managed to make you feel included just through the shear warmth of his personality. Indeed, Peter was a wonderful person. He was gentle, generous, passionate, enthusiastic, optimistic and very supportive. He not only inspired me, but he also had a massive impact on hundreds of other young statisticians all over the world.

Two years after I met Peter for the first time, he visited the University of California at Davis where I was a postdoc. We were office neighbours, and again I was amazed at how gentle and accessible he was. He discussed with me as if we were equal, which made me feel very comfortable. I told him about the topic of my PhD thesis (deconvolution), and a few days later, when I got into my office, there was a 20 page document waiting for me on my chair. It was Peter who had essentially written a paper on a new problem in the area, and he was asking if I would be interested in joining him to work on it. This is a story I share with many. He continuously had academic visitors from overseas who talked to him about a problem they had. Often he got very enthusiastic about it, or about a modified version of it, and soon after there was a 20 page document with a solution and long proofs waiting under their door.

*Supported by a grant and a future fellowship from the Australian Research Council.

Keywords and phrases: Berkson errors, errors-in-variables, image analysis, measurement errors, nonparametric smoothing

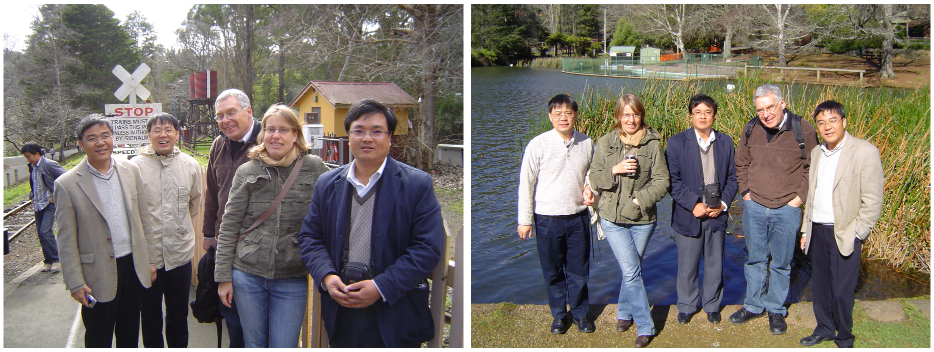

Fig 1: Left: Jianqing Fan, Jiashun Jin, Peter Hall, Aurore Delaigle and Maozai Tian on a trip to Puffing Billy. Right: lunch break on that day.

Peter was extremely prolific, his work was deep and very creative, and the breadth of problems he tackled was very unique. Although I worked with him continuously, he kept on managing to surprise me regularly with some of his unbelievably creative and beautiful ideas. I often told him: "how on earth did you even get to think about this in the first place?" It seemed like this came out of nowhere, I was wondering how such inventive things could come out of someone's brain. Peter also had the reputation of being able to prove almost any theoretical result, which was essentially true. He was a problem solver; he really enjoyed the challenge of solving a new problem and he was absolutely passionate about science and mathematics in general.

Peter won the most prestigious awards a statistician can get, including Fellowships of the Royal Society of London and of the Australian Academies of Science and Social Sciences, the election to Foreign associate of the US National Academy of Sciences and to Officer of the Order of Australia. Yet, he was the most unassuming person I've ever met.

He liked having visitors around the department, which he found very stimulating. He had lunch with them every day, but at lunch he much preferred talking about the news, politics, trains, planes or cats than about statistics. When he had time on the week-end, he enjoyed taking them to the countryside of his beautiful Australia that he loved so much. In Melbourne, his favourite activity with visitors was to take them for a ride on a steam train called "Puffing Billy". We went there many times after he acquired a digital camera, and his biggest pleasure was to photograph the locomotive under every possible angle. He developed his passion for trains and photography at a young age. It is him who introduced photography to his sister Fiona Hall, who later became a distinguished artist in Australia and of whom he

Fig 2: Peter Hall, Aurore Delaigle and Raymond Carroll on a trip to Puffing Billy.

was very proud (Delaigle and Wand, 2016).

Peter loved animals, and cats in particular. He often told me that when he got home, he used to stroke his cat Pumpkin 150 times, and that she complained if he stopped before 150. When he was at the Australian National University, he also developed an interest in cockatoos. He liked their company so much that he had bags of bird seeds in his office, from which he would feed the cockatoos through his window. Many visitors still remember the noise of the cockatoos knocking on his window, begging for food.

Peter was someone really special. His sheer presence made the whole atmosphere around him peaceful, joyful and exciting. He was an extraordinary, kind, gentle and generous person, of the type most people do not even have the chance to meet once in their lifetime. I feel blessed for having had him as a friend, collaborator and mentor for many years, but I miss him terribly, and I will miss him for the rest of my life.

2. Peter and errors-in-variables deconvolution problems.

2.1. *Introduction.* Peter made important and groundbreaking contributions to deconvolution problems in statistics, also referred to as nonparametric errors-in-variables or measurement errors problems. For an excellent introduction to measurement errors problems, the reader is referred to Carroll, Ruppert, Stefanski and Crainiceanu (2006). In the measurement errors literature, one can principally distinguish two types of errors called classical errors and Berkson errors. Peter contributed to nonparametric density and regression estimation problems for the two types of errors.

In its most basic form, the classical errors-in-variables problem can be described as follows. Suppose we are interested in estimating the density f_X of a variable X, but we can observe only an i.i.d. sample $W_1, \dots, W_n$ where, for each i,

$$W_i = X_i + U_i \,. \tag{2.1}$$

Here, the X_i's are i.i.d. with unknown density f_X, the U_i's are i.i.d. with symmetric known or estimable density f_U, and the X_i's are independent of the U_i's. In this model, the U_i's typically represent measurement errors made when collecting the data. If we let f_W denote the density of the W_i's, then f_W is the convolution of f_X and f_U, that is $f_W = f_X * f_U$, and we have to de-convolve this equation in order to estimate f_X from an estimator of f_W, whence the name "deconvolution".

Throughout this article, ϕ_T will denote the characteristic function of a random variable T or the Fourier transform of a function T. Assuming that

$$\inf_{t\in\mathbb{R}} |\phi_U(t)| > 0 \,, \tag{2.2}$$

and using the fact that $\phi_W = \phi_X \phi_U$, combined with the Fourier inversion theorem, Carroll and Hall (1988) and Stefanski and Carroll (1990) proposed the deconvolution kernel density estimator of f_X, defined by:

$$\hat{f}_X(x) = \int e^{-itx} \hat{\phi}_W(t) \phi_K(ht)/\phi_U(t)\, dt = (nh)^{-1} \sum_{j=1}^{n} K_U\Big(\frac{x - W_j}{h}\Big) \,, \tag{2.3}$$

$$\text{where} \quad K_U(x) = (2\pi)^{-1} \int e^{itu} \phi_K(t)/\phi_U(t/h)\, dt \,. \tag{2.4}$$

Here $\hat{\phi}_W(t) = n^{-1} \sum_{j=1}^{n} e^{itW_j}$ denotes the empirical characteristic function of the W_i's, $h > 0$ is a smoothing parameter called bandwidth, and ϕ_K denotes the Fourier transform of a function K called kernel, and used to dampen the effect of the unreliability of $\hat{\phi}_W(t)$ for $|t|$ large.

In the regression context, the classical errors-in-variables problem consists in estimating a regression curve m from i.i.d. data $(W_1, Y_1), \dots, (W_n, Y_n)$ generated by the model

$$Y_i = m(X_i) + \epsilon_i \,,\ W_i = X_i + U_i \,, \tag{2.5}$$

where the W_i's are as in (2.1),the X_i's, the U_i's and the ϵ_i's are completely independent, and the ϵ_i's are i.i.d. with mean zero and finite variance. Recall the definition of K_U in (2.4). Fan and Truong (1993) proposed the following kernel estimator of $m(x)$:

$$\hat{m}(x) = \sum_{j=1}^{n} Y_j K_U\Big(\frac{x - W_j}{h}\Big) \Big/ \sum_{j=1}^{n} K_U\Big(\frac{x - W_j}{h}\Big) \,. \tag{2.6}$$

In the Berkson error model, the roles of X_i and W_i are reversed compared to the classical error model. Specifically, in the Berkson errors-in-variables regression model, we wish to estimate a regression curve $m(x) = E(Y|X = x)$, but we observe only i.i.d. data $(W_1, Y_1), \ldots, (W_n, Y_n)$, where, for each i,

$$Y_i = m(X_i) + \epsilon_i \, , \; X_i = W_i + U_i \, , \tag{2.7}$$

with the W_i's, the U_i's and the ϵ_i's completely independent. The unobservable X_i's are i.i.d. with unknown density f_X, the errors U_i are i.i.d. with known symmetric density f_U, and the ϵ_i's are i.i.d. with mean zero and finite variance.

2.2. *Peter's first influential contribution to classical error problems.* Peter's first work in the area was the influential Carroll and Hall (1988) paper. There, the authors were the first to establish minimax convergence rates for nonparametric estimation of the density f_X in the model at (2.1). Let $C_k(B)$ denote the class of k-times differentiable densities f such that $\|f\|_\infty \leq B$ and $\|f^{(k)}\|_\infty \leq B$, and for each $f_X \in C_k(B)$, let $\hat{f}_X(x_0)$ denote any nonparametric estimator of $f_X(x_0)$, constructed from the W_i's at (2.1), where x_0 is a fixed real number.

The main result of the paper states that if, for a sequence of positive constants $a_n, n \geq 1$ we have

$$\liminf_{n\to\infty} \inf_{f_X \in C_k(B)} P_{f_X}\left\{|\hat{f}_X(x_0) - f_X(x_0)| \leq a_n\right\} = 1 \text{ for each } B > 0,$$

then if f_U is a standard normal density, $\lim_{n\to\infty} (\log n)^{k/2} a_n = \infty$ and if f_U is such that $|\phi_U(t)|$ decreases like $|t|^{-\alpha}$ as $|t| \to \infty$, then $\lim_{n\to\infty} n^{k/(2k+2\alpha+1)} a_n = \infty$. In other words, in the class of densities in $C_k(B)$, no nonparametric estimator can converge at a faster rate than $(\log n)^{-k/2}$ in the normal error case, and than $n^{-k/(2k+2\alpha+1)}$ in the algebraically decaying case. Moreover, the deconvolution kernel estimator reaches those rates.

The distinction between these two rates of decay has now become standard in the deconvolution literature. Error densities whose Fourier transform decays exponentially fast are usually referred to as supersmooth error densities, and error densities whose Fourier transform decays algebraically fast are referred to as ordinary smooth error densities; see Fan (1991), who generalised the results of Carroll and Hall (1988) to those two classes of errors. In the supersmooth error case, unless the density f_X is itself supersmooth, the convergence rates of nonparametric estimators are only logarithmic, whereas in the ordinary smooth error case, these rates are polynomial in n.

2.3. *Classical error problems with unknown error distribution.* Some of Peter's most important contributions to deconvolution focus on relaxing the assumption that the error density f_U in the model at (2.1) is known.

Diggle and Hall (1993) were among the first to relax this assumption. In this paper, the authors assume that, in addition to the sample $W_1, \ldots, W_n$, an i.i.d. sample $U_1, \ldots, U_m$, with $U_i \sim f_U$, is also available. Using this additional sample, they estimate the unknown $\phi_U(t)$ by $\hat{\phi}_U(t) = m^{-1} \sum_{j=1}^m e^{itU_j}$, and then replace ϕ_U in (2.3) by $\hat{\phi}_U$. Moreover, instead of using commonly employed finite order kernels, they use the infinite order sinc kernel K, defined by $\phi_K(t) = 1\{|t| \leq 1\}$. They derive asymptotic properties of their density estimator in this case, from which they conclude that, as long as $m \neq o(n)$, estimating ϕ_U has no first order asymptotic effect on the mean squared error of the estimator of f_X. This problem was taken up later by other authors, including Neumann (1997).

Discouraged by the slow convergence rates in the case of normally distributed errors, Peter essentially stopped working in the area for nearly ten years. However, in 2002, he considered a more optimistic model where the variance of the errors is regarded as tending to zero as sample size increases. That is, $\mathrm{var}(U_i) \to 0$ as $n \to \infty$. To justify this assumption, we can view the asymptotic behaviour of an estimator as a way to reflect the estimator's behaviour when the sample becomes ideal. In the traditional sense, "ideal" means "sample size tending to infinity". In the measurement error setting, it is sometimes reasonable to regard an ideal sample as a sample whose size increases, but also whose error contamination decreases. Asymptotics based on the assumption that $\mathrm{var}(U_i) \to 0$ as $n \to \infty$ can suitably reflect the finite sample scenario where the error variance is relatively small compared to the variance of the X_i's. In other words, methods that have good theoretical properties under this scenario can work reasonably well in practice (indeed, better than standard deconvolution approaches) as long as $\mathrm{var}(U_i)$ is relatively small. In other cases, they can produce seriously biased estimator (Delaigle, 2008).

In Hall and Simar (2002) and Carroll and Hall (2004), the authors consider two estimation problems under this small error variance assumption where $\mathrm{var}(U_i) \to 0$ as $n \to \infty$. Aware of the fact that it is often merely an approximation to the truth, they argue that instead of attempting to consistently estimate f_X directly, using methods which have poor convergence rates, one should instead estimate an approximation to f_X obtained under that assumption, but which can be estimated at standard error-free nonparametric rates. In addition, instead of requiring knowledge of the whole error density, the small error assumption permits to develop approaches that require only a few low order moments of the U_i's. In Hall and Simar (2002), the goal is to estimate changepoints and discontinuities of f_X from data generated by (2.1). Under the small error variance assumption, the authors derive estimators that converge at standard polynomial rates, rather than the slow typical deconvolution rates. In Carroll and Hall (2004), the authors propose two nonparametric estimators of f_X (kernel and orthogonal series) constructed from data generated by (2.1). Under the small error variance assumption, they show that these estimators converge at fast algebraic rates.

These two papers resparkled Peter's interest in deconvolution problems, to which he made contributions until the end of his life. In Delaigle, Hall and Meister (2008), Peter tackled again the case where the error density f_U is unknown in model (2.1), this time assuming that replicated contaminated measurements of the X_i's are available. That is, for each i, we observe $W_{ij} = X_i + U_{ij}$, where $j \geq 2$ and the X_i's and the U_{ij}'s are totally independent, with $U_{ij} \sim f_U$. Noting that, for $j \neq k$, we have $W_{ij} - W_{ik} = U_{ij} - U_{ik}$, and recalling that f_U is symmetric, we can construct a consistent estimator of $|\phi_U|^2$ from the $W_{ij} - W_{ik}$'s. Assuming that $\phi_U(t) \geq 0$ for all t, we deduce an estimator $\hat{\phi}_U$ of ϕ_U, which can replace ϕ_U in the estimator at (2.3). More precisely, to avoid getting too close to zero, the authors replace ϕ_U by $\hat{\phi}_U + \rho$, where $\rho \geq 0$ is a small ridge parameter. Proceeding similarly, they also extend the regression estimator at (2.6) to this context. They show, in both the density and the regression cases, that estimating ϕ_U only has second order impact on the asymptotic properties of the curve estimators, although in the ordinary smooth case, for this to hold they require f_X to be a little smoother than f_U. This problem was also studied in Li and Vuong (1998) in a more complex setting, but under a set of assumptions that are difficult to satisfy.

Peter's last paper in the area (Delaigle and Hall, 2016) was one of his favourite contributions to deconvolution. In that paper, he considers the density deconvolution problem in the difficult case where f_U is unknown and no additional data are available. Several authors (Butucea and Matias, 2005, Meister, 2006 and Butucea, Matias and Pouet, 2008) had considered this problem before, but under the assumption that f_U belonged to a known parametric family. In Delaigle and Hall (2016), the only assumptions about f_U are that it is symmetric and satisfies (2.2). Arguing that the real world is dominated by irregular distributions, the density f_X is assumed to be sufficiently irregular for it to be distinguishable from the nice symmetric error density f_U. Specifically, it is assumed that f_X cannot be expressed as a mixture of two densities, one of which is symmetric. Peter showed that, under this assumption, the density of X can be estimated from its phase function, which itself can be easily estimated from the W_i's. The authors propose a data-driven method that gives surprisingly good results.

2.4. *Other contributions to classical error problems.* In Hall and Meister (2007), the authors relax assumption (2.2) by proposing a ridge-based procedure. As in Stefanski and Carroll (1990), their method is based on the Fourier inversion theorem, but unlike the deconvolution kernel estimator at (2.3), the authors regularise their estimator of f_X through a positive ridge parameter function $\rho(t)$. Let $\hat{\phi}_W$ as defined above, and let $r \geq 0$ be a tuning parameter. Using data generated by the model at (2.1), in order to avoid dividing by a number too close to zero, they propose to estimate $f_X(x)$ by

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\hat{\phi}_W(t)\bar{\phi}_U(t)|\phi_U(t)|^r}{\max\{|\phi_U(t)|, \rho(t)\}^{r+2}}\, dt\,.$$

They also suggest a version of their estimator in the regression case at (2.5), and establish optimality of their estimators in a wide variety of settings.

In Delaigle, Hall and Müller (2007), the authors consider a subtle variant of the Berkson model at (2.7). As in the Berkson model, they observe data $(W_1, Y_1), \ldots, (W_n, Y_n)$ and are interested in estimating the curve $m(x) = E(Y_i|X_i = x)$, where W_i, X_i and U_i are as in (2.7). However, a crucial difference with (2.7) is that the Y_i's satisfy $Y_i = g(W_i) + \eta_i$, where the η_i's are i.i.d. with zero mean. Thus instead of being generated by the X_i's, the Y_i's are generated by the W_i's. This difference makes the problem in Delaigle, Hall and Müller (2007) much simpler than the Berkson one, and the authors propose a nonparametric estimator of m that converges at the parametric rate. That work was taken further in Carroll, Delaigle and Hall (2009), where the authors consider more general similar phenomena in a prediction setting.

In Hall and Ma (2007), the authors propose a bootstrap procedure for testing whether a regression curve is polynomial, using data generated by the model at (2.5). As a side result, they also suggest a nonparametric estimator of the cumulative distribution function (cdf) of the X_i's, a problem which was later studied in depth by Hall and Lahiri (2008), where the authors also propose moment and quantile estimators. In Hall and Lahiri (2008), the authors showed that, in the ordinary smooth error case, as long as the distribution of the X_i's is sufficiently smooth, the cdf can be estimated at the parametric convergence rate. They also make the striking discovery that the convergence rate is not always the same at all points.

In Delaigle and Hall (2008), the authors propose an approximation method for selecting smoothing parameters in general deconvolution problems. In standard error-free nonparametric curve estimation problems, a standard and popular approach to choosing smoothing parameters is the so called plug-in method. It consists in constructing an estimator of the smoothing parameter that minimises the asymptotic mean squared error of the nonparametic curve estimator. A difficulty in the errors-in-variables context is that, in some problems (e.g. for the regression estimator at (2.6)), this asymptotic mean squared error is so complex that it is not possible to construct a reasonable estimator of it. To overcome these difficulties, Delaigle and Hall (2008) suggest applying simulation extrapolation (SIMEX) methods to smoothing parameter choice. SIMEX methods were originally introduced by Cook and Stefanski (1994) in the parametric context. Delaigle and Hall (2008) proved that, although SIMEX methods generally provide non-consistent nonparametric curve estimators (Staudenmayer and Ruppert, 2004), when used appropriately they yield bandwidths of the right order, which, in turn, result in consistent nonparametric curve estimators.

Their method consists of two steps (simulation and extrapolation), which we explain in the density estimation case, for simplicity: (i) generate data which contain more noise than the W_i's: for $i = 1, \dots, n$, let $W_i^* = W_i + U_i^*$ and $W_i^{**} = W_i^* + U_i^{**}$, with $U_i^* \sim f_U$ and $U_i^{**} \sim f_U$. Note that W_i^* and W_i^{**} denote contaminated versions of, respectively, W_i and W_i^*, which are all available. (ii) Consider temporarily that, instead of f_X, the densities of interest are f_W and f_{W^*}, and construct their deconvolution kernel estimators $\hat{f}_W$ and $\hat{f}_{W^*}$ using the contaminated data W_i^* and W_i^{**}, respectively. Construct also standard kernel estimators $\tilde{f}_W$ and $\tilde{f}_{W^*}$ using the error-free versions W_i and W_i^*. Since $\tilde{f}_W$ and $\tilde{f}_{W^*}$ converge faster to f_W and f_{W^*} than $\hat{f}_W$ and $\hat{f}_{W^*}$ do, bandwidths h^* and h^{**} that are appropriate for $\hat{f}_W$ and $\hat{f}_{W^*}$ can be defined by $h^* = \text{argmin} \int (\hat{f}_W - \tilde{f}_W)^2$ and $h^{**} = \text{argmin} \int (\hat{f}_{W^*} - \tilde{f}_{W^*})^2$. Since W^{**} and W^* measure W^* and W in the same way as W measures X, then it is reasonable to expect that h^{**} measures h^* in the same way as h^* measures h, where h is a bandwidth appropriate for computing $\hat{f}_X$ at (2.3). This motivates taking $h = h^{**}/(h^*)^2$. The same ideas can be used to select the smoothing parameters of other errors-in-variables problems. See for example Delaigle and Hall (2011) and Delaigle, Hall and Jamshidi (2015)

In Carroll, Delaigle and Hall (2011), the authors considered modified tilted deconvolution estimators, where, instead of giving equal weight n^{-1} to each observation in the estimators at (2.3) and (2.6), the ith observation receives a nonnegative weight p_i. The weights satisfy $\sum_i p_i = 1$, and the p_i's are chosen so that the estimators at (2.3) and (2.6) satisfy a shape constraint. In Delaigle and Hall (2011), the authors consider a heteroscedastic version of the model at (2.5), where the ϵ_i's are replaced by $\sigma(X_i)\eta_i$, with the η_i's i.i.d. with mean zero and variance one, and independent of the X_i's and the U_i's, and with σ a nonnegative function. They propose parametric and nonparametric estimators of σ. In Delaigle, Hall and Jamshidi (2015), the authors construct pointwise confidence bands for the estimator $\hat{m}$ at (2.6). In the error-free case, constructing such bands is complex because of the difficulty of choosing the smoothing parameters in practice, a problem which remains largely unsolved. An interesting aspect of Delaigle et al.'s (2015) contribution is that, exploiting SIMEX ideas from Delaigle and Hall (2008), they manage to derive data-driven smoothing parameters relatively easily.

Other noticeable contributions of Peter to the deconvolution problem include Hall and Qiu (2005), where the authors propose to estimate f_X from data generated by the model at (2.1) using a cosine-series estimator, in the case where f_X is supported on a known compact interval; Delaigle and Hall (2006), where the authors discuss the choice of an optimal kernel for deconvolution; Hall and Maiti (2009) where the authors analyse clustered data using deconvolution techniques; Chen, Delaigle and Hall (2010) where the authors exploit deconvolution techniques for inference in a class of Lévy processes; and Lee, Hall, Shen, Marron, Tolle and Burch (2013), where the authors consider the case where the distribution of X is a mixture of a finite number of discrete atoms and a continuous distribution. They use a sieve estimator, which they compute using penalised likelihood.

2.5. *Peter's main contributions to Berkson error problems.* Peter made several contributions to the nonparametric Berkson errors-in-variables model at (2.7). Estimating f_X from data $W_1, \ldots, W_n$ generated as in (2.7) is trivial because $f_X = f_W * f_U$, where f_U is known and f_W can be estimated directly from the W_i's. By contrast, estimating the regression curve m is complex. To understand this, let $g(w) = E(Y|W = w) = E\{m(X)|W = w\} = m*f_U(w)$. We can estimate g from the (W_i, Y_i)'s, and since we know f_U, in principle we can obtain an estimator of m by deconvolving this equation. However, deconvolving Berkson errors causes a number of difficulties which do not arise when deconvolving classical errors.

Delaigle, Hall and Qiu (2006) highlighted a problem which arises when f_W and f_U are compactly supported. Let $[a_W, b_W]$ and $[-\delta, \delta]$ denote the support of f_W and f_U, respectively, with $\delta > 0$. Since $W_i \in [a_W, b_W]$ for all i, we can estimate $g(\cdot) = E(Y|W = \cdot)$ nonparametrically only on $[a_W, b_W]$. Now $g = m * f_U$, which means that we can estimate m only on $[a_W + \delta, b_W - \delta]$. However often $m(\cdot) = E(Y|X = \cdot)$ has the same support as f_X. Since $X = W + U$ with W and U independent, this support is $[a_W - \delta, b_W + \delta]$. Thus often we cannot estimate m nonparametrically on its entire support.

This example illustrates that, in the Berkson errors-in-variables problem, the regression curve m is not always identifiable, especially when the curves are compactly supported. Under identifiability conditions, Delaigle, Hall and Qiu (2006) propose and study properties of a sine-cosine series estimator of m. They also suggest a kernel estimator, but Fourier transform-based kernel approaches are really the topic of Carroll, Delaigle and Hall (2007). There, the authors also consider a more general case with a mixture of Berkson and classical errors, this time without compact support assumptions. However, the non compactly supported case causes problems too. In order to deconvolve $g = m * f_U$ using Fourier transforms, that is, in order to use an estimator of the type $m(x) = (2\pi)^{-1} \int e^{-itx} \hat{\phi}_g(t)\phi_K(ht)/\phi_U(t)\, dt$, where $\hat{\phi}_g(t) = \int e^{itx} \hat{g}(x)\, dx$ denote the Fourier transform of a nonparametric estimator $\hat{g}$ of g, the estimator $\hat{g}$ needs to be sufficiently good to be integrated over the whole real line. This makes the problem particularly complex.

3. Peter and deconvolution problems in image analysis. Through his interest in photography, Peter made a number of contributions to image anaylsis, which also extensively use the Fourier inversion techniques employed in the deconvolution problems discussed above. Images are often obtained in a blurred and noisy way. Specifically, letting X denote the ideal image, the observed image Z is often modelled by

$$Z(r) = \int T(u)X(r+u)\, du + \delta(r)\,, \tag{3.1}$$

where T is a point spread function blurring the signal, δ represents additive noise and $r \in \mathbb{R}^d$ (or a bounded subset of $\mathbb{R}^d$). Often, for images, $d = 2$ or 3, but the techniques can be employed for more general d-dimensional signals. Since images are only observed discretely, often the model at (3.1) is replaced by a discrete version of it. Peter worked under the two models (continuous and discrete), and for simplicity, in our discussion below we shall not make the distinction between the two. Often but not always, Peter took T to be a multivariate double exponential point-spread function (or a discrete version of it when working with discrete models), but we shall not specify the form of T in our discussion below.

As in the deconvolution problems discussed above, because of the convolution structure in (3.1), techniques for reconstructing the image X from Z are often based on inverse Fourier transforms. Assuming that the Fourier transform of Z never vanishes, without the error δ and with T known, X could be recovered by direct Fourier inversion of the equation $\phi_X = \phi_Z/\phi_T$. However, the presence of the noise δ makes that inversion unstable, especially at points where ϕ_T is close to zero. One way to address this difficulty is to use some form of regularisation of the Fourier inversion; for example, ϕ_Z/ϕ_T can be replaced by zero when ϕ_T becomes too small.

A first body of work by Peter in the area was dedicated to establishing theoretical properties of such image reconstruction methods. Hall and Titterington (1986) studied theoretical properties of several commonly employed regularised Fourier-based techniques. Other properties, including lower bounds, were established in Hall (1987a). Further properties were also derived in Hall (1987b), where Peter also made the interesting discovery that, in some cases, blurring a blurred image can produce an image of better quality than the originally blurred image. Optimal convergence rates for image recovery were established in Hall (1990), where it was also shown that Fourier-based techniques reach those rates. Further theoretical properties were developed in Hall and Koch (1990), and practical choices for the level of regularisation were suggested in Hall and Koch (1992).

Peter also tackled other related problems. In Hall and Qiu (2007a), the point-spread function T is known up to the value of one or several unknown parameters θ. Motivated by the fact that images often contain sharp edges, they propose to estimate θ by the value $\hat{\theta}$ for which the reconstructed image gives the most plausible edges. In Hall and Qiu (2007b), T is also unknown, but no parametric model is available for it. The authors propose to estimate T in such as way that a test signal X^{test} is best recovered. Motivated by the fact that, in image restoration, it is often desirable for rectangular shapes to be well reconstructed, they suggest taking X^{test} to be a d-dimensional version of a rectangle. Then, to estimate T, they use inverse Fourier inversion techniques, where on this occasion it is X that is known (and equal to X^{test}) and T that is unknown. A problem with this approach is that the Fourier transform of their test signal vanishes periodically, and to overcome this difficulty they use a ridge-based technique similar to the one used in Hall and Meister (2007).

Peter's last contribution to the area was his work in Carroll, Delaigle and Hall (2012), where rather than recovering the image X, the goal was to classify noisy data of the same type as Z into two groups. The authors proposed to use a parametric model for T, where the parameters are chosen so as to minimise a cross-validation estimate of classification error. They showed that, in general, the optimal inversion is not necessarily the one that gives the best image reconstruction.

Acknowledgment. I thank Alicia Nieto-Reyes for helping me prepare my speech at Peter's funeral, which I used as a basis for section 1. I also thank Jiashun Jin for allowing me to use the photos in Fig. 1, Raymond Carroll for the photo in Fig. 2, and Hans Müller for his helpful comments.

References.

[1] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11**, 309–340.

[2] Butucea, C., Matias, C. and Pouet, P. (2008). Adaptivity in convolution models with partially known noise distribution. *Electron. J. Statist.* **2** 897–915.

[3] Carrroll, R.J., Delaigle, A. and Hall, P. (2007). Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *J. Roy. Statist. Soc.* Ser. B **69**, 859–878.

[4] Carrroll, R.J., Delaigle, A. and Hall, P. (2009). Nonparametric prediction in measurement error models. (With discussion.) *J. Amer. Statist. Assoc.* **104**, 993–1003.

[5] Carrroll, R.J., Delaigle, A. and Hall, P. (2011). Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *J. Amer. Statist. Assoc.* **106**, 191–202.

[6] Carrroll, R.J., Delaigle, A. and Hall, P. (2012). Deconvolution when classifying noisy data involving transformations. *J. Amer. Statist. Assoc.* **107**, 1166–1177.

[7] Carrroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.

[8] Carrroll, R.J. and Hall, P. (2004). Low order approximations in deconvolution and regression with errors in variables. *J. Roy. Statist. Soc.* Ser. B **66**, 31–46.

[9] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models*, 2nd Edn. Chapman and Hall CRC Press, Boca Raton.

[10] Chen, S-X., Delaigle, A. and Hall, P. (2010). Nonparametric estimation for a class of Lévy process. *J. Econometrics* **157**, 257–271.

[11] Cook, J. R., and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.* **89**, 1314–1328.

[12] Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica* **18**, 1025–1045.

[13] Delaigle, A. and Hall, P. (2006). On optimal kernel choice for deconvolution. *Statist. Probab. Lett.* **76**, 1594–1602.

[14] Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Amer. Statist. Assoc.* **103**, 280–287.

[15] Delaigle, A. and Hall, P. (2011). Estimation of observation-error variance in errors-in-variables regression. *Statistica Sinica* **21**, 1023–1063.

[16] Delaigle, A. and Hall, P. (2016). Methodology for nonparametric deconvolution when the error distribution is unknown. *J. Roy. Statist. Soc.* Ser. B, **78**, 231–252.

[17] Delaigle, A., Hall, P. and Jamshidi, F. (2015). Confidence bands in nonparametric errors-in-variables regression. *J. Roy. Statist. Soc.* Ser. B **77**, 149–169.

[18] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements.*Ann. Statist.* **36**, 665–685.

[19] Delaigle, A., Hall, P. and Müller, H.-G. (2007). Accelerated convergence for nonparametric regression with coarsened predictors. *Ann. Statist.* **35**, 2639–2653.

[20] Delaigle, A., Hall, P. and Qiu, P. (2006). Nonparametric methods for solving the Berkson errors-in-variables problem. *J. Roy. Statist. Soc.* Ser. B **68** 201–220.

[21] Delaigle, A. and Wand, M.P. (2016). A conversation with Peter Hall. *Statistical Science* **31**, 275–304.

[22] Diggle, P.J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc.* Ser. B, **55**, 523–532.

[23] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
[24] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900–1925.
[25] Hall, P. (1987a). On the amount of detail that can be recovered from a degraded signal. *Adv. Appl. Probab.* **19**, 371–395.
[26] Hall, P. (1987b). On the processing of a motion-blurred image. *SIAM J. Appl. Math.* **47**, 441–453.
[27] Hall, P. (1990). Optimal convergence rates in signal recovery. *Ann. Probab.* **18**, 887–900.
[28] Hall, P. and Koch, I. (1990). On continuous image models and image analysis in the presence of correlated noise. *Adv. Appl. Probab.* **22**, 332–349.
[29] Hall, P. and Koch, I. (1992). On the feasibility of cross-validation in image analysis. *SIAM J. Appl. Math.* **52**, 292–313.
[30] Hall, P. and Lahiri, S. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.* **36**, 2110–2134.
[31] Hall, P. and Ma, Y. (2007). Testing the suitability of polynomial models in errors-in-variables problems. *Ann. Statist.* **35**, 2620–2638.
[32] Hall, P. and Maiti, T. (2009). Deconvolution methods for nonparametric inference in two-level mixed models. *J. Roy. Statist. Soc.* Ser. B **71**, 703–718.
[33] Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution.*Ann. Statist.* **35**, 1535–1558.
[34] Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92**, 135–148.
[35] Hall, P. and Qiu, P. (2007a). Blind deconvolution and deblurring in image analysis. *Statist. Sinica* **17**, 1483–1509.
[36] Hall, P., and Qiu, P. (2007b). Nonparametric estimation of a point spread function in multivariate problems. *Ann. Statist.* **35**, 1512–1534.
[37] Hall, P. and Simar, L. (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *J. Amer. Statist. Assoc.* **97**, 523—534.
[38] Hall, P. and Titterington, D. M. (1986). On some smoothing techniques used in image restoration. *J. Roy. Statist. Soc.* Ser. B **48**, 330–343.
[39] Lee, M., Hall, P., Shen, H., Marron, J. S., Tolle, J. and Burch, C. (2013). Deconvolution estimation of mixture distributions with boundaries. *Electronic J. Statist* **7**, 323–341.
[40] Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators.*J. Multivariate Anal.* **65**, 139–165.
[41] Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica* **16**, 195–211.
[42] Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Statist.* **7** 307–330.
[43] Staudenmayer, J. and Ruppert, D. (2004). Local polynomial regression and simulation-extrapolation. *J. Roy. Statist. Soc.* Ser. B **66**, 17–30.
[44] Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **2**, 169–184.

Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS) and School of Mathematics and Statistics
University of Melbourne, Parkville, Victoria 3010, Australia
E-mail: A.Delaigle@ms.unimelb.edu.au