Analyse P(D) des cartes SPIRE

Sommaire

4.1 Présentation de l'analyse P(D)

4.1.1 Principe

Dans les chapitres précédents, nous avons vu qu'il est difficile d'étudier directement les sources les plus faibles à cause de la confusion. La méthode de l'empilement permet de sonder ces populations à condition d'avoir au préalable des informations venant d'autres longueurs d'onde. La méthode du P(D) est complémentaire à l'empilement. En effet, elle se base uniquement sur l'histogramme d'une (ou plusieurs) cartes infrarouges, dont la forme dépend uniquement de la PSF, du bruit instrumental et des comptages[1]. Dans le cas où la PSF et le bruit instrumental sont bien connus, il est possible de retrouver les comptages qui sont compatibles avec l'histogramme observé. Le principe est donc simple, mais la mise en application l'est beaucoup moins.

4.1.2 Quelques exemples illustratifs

La manière dont les comptages de sources sont liés à l'histogramme des pixels[2] d'une carte est non triviale (voire Sect. 4.2). La Fig. 4.1 illustre ce lien. Nous avons pris un modèle simple de comptages en $S^{-1.5}$ en dessous de 1 mJy et $S^{-3.5}$ au delà (en noir). On suppose également que les cartes sont à moyenne nulle[3]. On voit tout de suite qu'en dehors des hauts flux, les comptages et l'histogramme ont des comportements assez différents. En effet, une source affecte le signal dans plusieurs pixels, avec un effet qui décroît avec la distance entre la source et le pixel. De plus, il peut y avoir plusieurs sources faibles par pixel. C'est pour ces raisons que l'histogramme du flux des sources (les comptages) est assez différent de celui du signal dans les pixels.

La forme des comptages de sources affecte fortement l'histogramme d'une carte. Si on augmente le nombre de sources brillantes, en utilisant un comportement en $S^{-2.5}$ au delà de 10 mJy (courbe en bleu de la Fig. 4.1), la forme de l'histogramme est fortement affectée à haut flux (D>10 mJy). En effet, plus de sources brillantes implique plus de pixels brillants. En revanche, le pic de l'histogramme est très peu affecté, car les ailes des sources brillantes, peu nombreuses, n'affectent que légèrement l'histogramme. Si on augmente le nombre de sources à bas flux, en utilisant un comportement en $S^{-2.5}$ en dessous de 1 mJy (courbe en rouge de la Fig. 4.1), on élargit le pic de la distribibution, alors

---

[1] En fait, les propriétés de regroupement des sources modifient également légèrement cet histogramme. Néanmoins, cet effet est souvent négligeable.

[2] On confondra ici l'histogramme des pixels avec la densité de probabilité du signal dans un pixel. En effet, l'histogramme moyen est exactement la densité de probabilité dans les pixels à une constante de normalisation près.

[3] Cette convention est souvent utilisée quand le niveau zéro du signal est inconnu. C'est le cas pour *Herchel* et la plupart des observations *Spitzer*

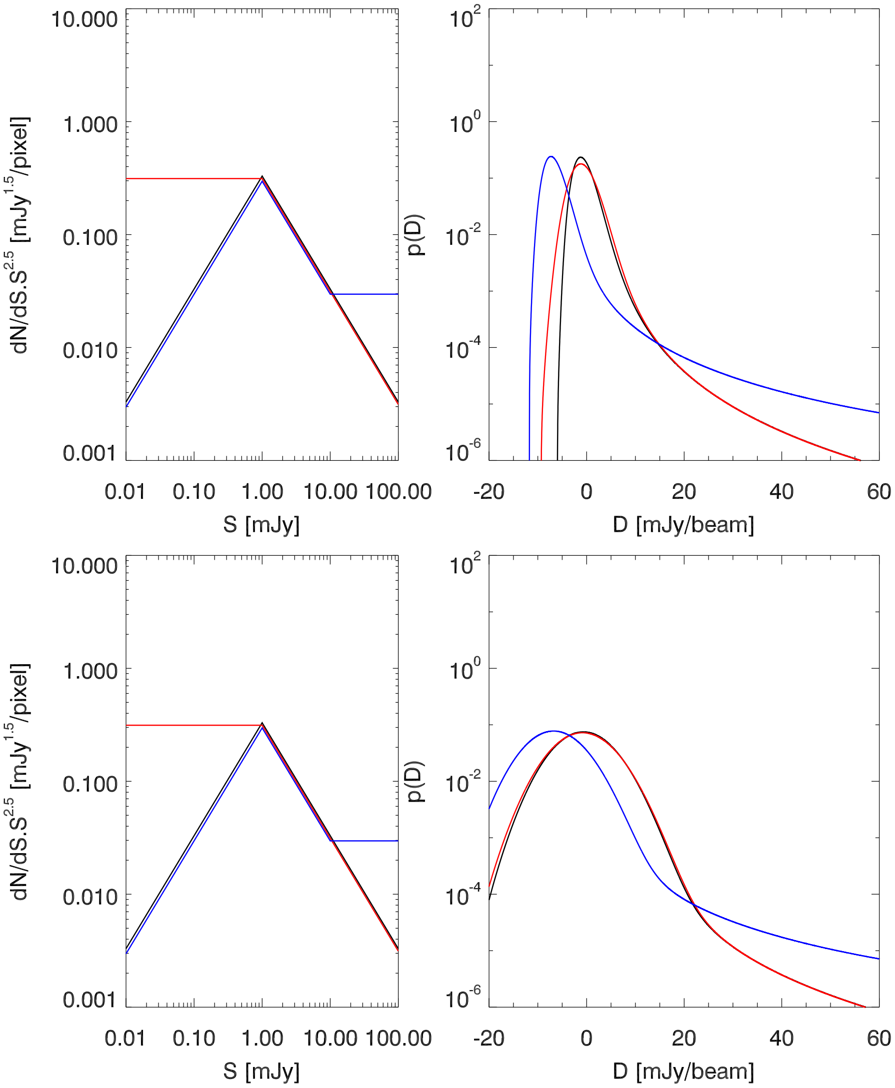

FIGURE 4.1 – Histogramme d'une carte infrarouge en fonction des comptages, avec une PSF gausienne de 3 pixels de largeur à mi-hauteur et un bruit instrumental nul (en haut) et de 5 mJy (en bas). A gauche : Trois modèles de comptages différentiels normalisés en noir, rouge et bleu. A droite : Densité de probabilité du signal dans un pixel de la carte.

que les haut flux sont très peu affectés. Ces sources de flux faibles sont très nombreuses (plusieurs sources par *beam*), ne sont pas détectables individuellement dans les images, mais engendrent des fluctuations du fond. Si on augmente le nombre de sources faibles, ces fluctuations augmentent (comme $\sqrt{N}$ dans le cas simpliste où toute les sources ont le même flux).

Cet exemple simple illustre les principes de base de l'analyse P(D). Les pixels brillants permettent de poser des contraintes sur les sources brillantes, alors que la largeur du pic à bas flux contraint les comptages de sources faibles. L'histogramme est très sensible aux comptages à haut flux, mais moins aux comptages à bas flux, plus difficiles à estimer. Enfin, un bruit instrumental important (graphique du bas) élargit l'histogramme et s'il est trop fort, peut empêcher de discriminer deux pentes de comptages à bas flux (courbes noire et rouge de la Fig. 4.1). Il est donc important pour bien contraindre les bas flux d'utiliser des observations très profondes.

4.1.3 Mise en oeuvre

Contrairement aux méthodes de comptages classiques pour lesquels on mesure simplement un nombre de sources par bin de flux, l'analyse P(D) nécessite de supposer une représentation paramétrique des comptages. On peut par exemple supposer une simple loi de puissance. Toutefois, ce type de fonction simple n'est pas représentatif des comptages réels, et le meilleur ajustement en loi de puissance peut être très différent de la réalité. Patanchon *et al.* (2009) a proposé d'utiliser une description paramétrique des comptages plus élaborée. Des noeuds sont placés à des flux fixés. La valeur des comptages différentiels (dN/dS) à chaque noeud est un paramètre libre, et sont reliés entre eux par des lois de puissances.

Les paramètres du modèle de comptages sont ajustés afin de reproduire les histogrammes observés. On peut par exemple utiliser une méthode MCMC (*Monte Carlo Markov Chain*, Chib et Greenberg (1995)) pour étudier précisément les dégénérescences entre paramètres et produire des intervalles de confiance réalistes. Cette technique consiste à explorer de manière semi-aléatoire l'espace de paramètres. A chaque pas, un déplacement aléatoire se fait dans l'espace des paramètres. La nouvelle position est conservée si la vraisemblance est meilleure qu'au pas précédent. Elle est conservée avec une probabilté $L(n)/L(n-1)$ (rapport entre la nouvelle et l'ancienne vraisemblance) dans le cas contraire. Pour une chaîne suffisamment longue (hypothèse d'ergodicité), les différentes réalisations ont la même distribution que la densité de probabilité a posteriori des paramètres du modèle de comptages. Cette propriété rend extrêmement commode la détermination des intervalles de confiance sur chacun des paramètres.

4.2 Le formalisme du P(D)

La densité de probabilité du signal dans un pixel peut se calculer analytiquement à partie des comptages, de la PSF et du bruit instrumental. Dans cette section, nous allons montrer comment ils sont reliés entre eux.

4.2.1 Cas d'une PSF ponctuelle

Le nombre de source par unité de flux et par unité d'angle solide est notée dN/dS, où S est le flux. On prendra la taille du pixel comme unité d'angle solide. Le nombre moyen $\overline{n_k}$ de sources observées dans un intervalle de flux k compris entre S_k et $S_k + \Delta_S$ est

$$\overline{n_k} = \frac{dN}{dS}\Delta S. \tag{4.1}$$

Nous supposons que les sources sont non corrélées spatialement. Le nombre de sources observées dans une tranche de flux k (n_k) suit donc une loi de Poisson :

$$p_{n_k}(n_k) = \frac{\overline{n_k}^{n_k}}{n_k!}e^{-\overline{n_k}}. \tag{4.2}$$

Le flux total reçu par un pixel, D, vaut

$$D = \sum_k n_k S_k = \sum_k D_k, \tag{4.3}$$

où D_k est la contribution au flux total des sources dont le flux tombe dans le k-ième intervalle de flux. La densité de probabilité de D_k est

$$p_{D_k}(D_k) = \frac{1}{S_k} \times p_{n_k}(D_k/S_k). \tag{4.4}$$

Celle de D est donnée par (on peut démontrer que la densité de probabilité d'une somme de variable aléatoire est la convolution de leurs densité de probabilité)

$$p_D = ... * p_{D_{k-1}} * p_{D_k} * p_{D_{k+1}} * .... \tag{4.5}$$

Dans l'espace de Fourier, le produit de convolution devient une simple multiplication, et on a donc

$$\hat{p}_D = \otimes_{k=1}^{N_{bin}} \hat{p}_{D_k}, \tag{4.6}$$

où $\hat{p}$ est la transformée de Fourier de p (appelé aussi fonction caractéristique). On peut monter que

$$\hat{p}_{n_k}(\omega) = exp(\overline{n_k}(e^{i\omega} - 1)), \tag{4.7}$$

puis,

$$\hat{p}_{D_k}(\omega) = exp(\overline{n_k}(e^{i\omega S_k} - 1)). \tag{4.8}$$

En combinant les équations 4.1, 4.6 et 4.8, on trouve alors

$$\hat{p}_D(\omega) = \otimes_{k=1}^{N_{bin}} exp(\overline{n_k}(e^{i\omega S_k} - 1)) = exp\left(\sum_{k=1}^{N_{bin}} \frac{dN}{dS} \Delta S\,(e^{i\omega S_k} - 1)\right). \tag{4.9}$$

Puis, on passe à la limite continue :

$$\hat{p}_D(\omega) = exp\left(\int_0^\infty \frac{dN}{dS}(e^{i\omega S} - 1)dS\right). \tag{4.10}$$

En applicant le transformée de Fourier inverse, on obtient finalement

$$p_D(D) = \frac{1}{2\pi}\int_0^\infty exp\left(\int_0^\infty \frac{dN}{dS}(e^{i\omega S} - 1)dS\right) e^{-i\omega D} d\omega. \tag{4.11}$$

4.2.2 Cas d'une PSF entendue positive

Lorsque la PSF n'est pas ponctuelle, une source n'a pas besoin d'être centrée sur un pixel pour contribuer au signal reçu par celui-ci. On note F le flux effectif vu par un pixel. Pour une source de flux S à une distance angulaire θ du pixel considéré, il vaut

$$F = S \times PSF(\theta), \tag{4.12}$$

où PSF est la réponse impulsionnelle de l'instrument. Le nombre de sources par unité de flux effectif et par pixel, n, est donné par

$$n(F) = \int\int \frac{dN}{dF}\Big|_F d\Omega. \tag{4.13}$$

Et l'équation 4.12 donne

$$\frac{dN}{dF}\Big|_F = \frac{dN}{dS}\Big|_{S=\frac{F}{PSF(\theta)}} \times \frac{dS}{dF}\Big|_{S=\frac{F}{PSF(\theta)}} = \frac{dN}{dS}\Big|_{S=\frac{F}{PSF(\theta)}} \times \frac{1}{PSF(\theta)}. \tag{4.14}$$

En combinant les équations 4.13 et 4.15, on trouve

$$n(F) = \int\int \frac{dN}{dS}\Big|_{S=\frac{F}{PSF(\theta)}} \times \frac{1}{PSF(\theta)} d\Omega. \tag{4.15}$$

La suite du calcul est similaire à celui de la section 4.2.1 en remplaçant dN/dS par n, et S par F. On obtient finalement

$$p_D(D) = \frac{1}{2\pi}\int_0^\infty exp\left(\int_0^\infty n(F)(e^{i\omega F} - 1)dF\right) e^{-i\omega D} d\omega. \tag{4.16}$$

4.2.3 Effet du filtrage

Lors de la fabrication des cartes, un filtrage passe haut peut être appliqué. Une PSF prenant en compte ce filtrage peut posséder des parties négatives. Or, l'équation 4.15 n'est pas définie pour $PSF(\theta) = 0$. Par conséquent, il est nécessaire de séparer la contribution positive et la contribution négative de la PSF :

$$D = D_+ - D_-. \tag{4.17}$$

La densité de probabilité de D sera donc

$$p_D(D) = P_{D_+}(D) * P_{D_-}(-D). \tag{4.18}$$

On définit alors

$$n_+(F) = \int\int_{PSF>0} \frac{dN}{dS}\Big|_{S=\frac{F}{PSF_+(\theta)}} \times \frac{1}{PSF_+(\theta)} d\Omega, \tag{4.19}$$

et,

$$n_-(F) = \int\int_{PSF<0} \frac{dN}{dS}\Big|_{S=\frac{F}{PSF_-(\theta)}} \times \frac{1}{PSF_-(\theta)} d\Omega, \tag{4.20}$$

où PSF_- est la valeur absolue des parties négatives de la PSF. On a alors la densité de probabilité de la contribution positive et négative (D_+ et D_-) au signal d'un pixel :

$$p_{D_+}(D_+) = \frac{1}{2\pi} \int_0^\infty exp\left(\int_0^\infty n_+(F)(e^{i\omega F} - 1)dF\right) e^{-i\omega D_+} d\omega \tag{4.21}$$

puis,

$$p_{D_-}(D_-) = \frac{1}{2\pi} \int_0^\infty exp\left(\int_0^\infty n_-(F)(e^{i\omega \times (-F)} - 1)dF\right) e^{-i\omega D_-} d\omega. \tag{4.22}$$

En combinant les relations 4.18, 4.21 et 4.22, et en utilisant les propriétés de la transformée de Fourier du produit de convolution, on obtient finalement

$$p_D(D) = \frac{1}{2\pi} \int_0^\infty exp\left(\int_0^\infty n_+(F)(e^{i\omega F} - 1)) \times exp(\int_0^\infty n_-(F)(e^{-i\omega F} - 1)dF\right) e^{-i\omega D} d\omega, \tag{4.23}$$

simplifiable en

$$p_D(D) = \frac{1}{2\pi} \int_0^\infty exp(\int_0^\infty [n_+(F)e^{i\omega F} + n_-(F)e^{-i\omega F} - n_+(F) - n_-(F)]dF) e^{-i\omega D} d\omega. \tag{4.24}$$

4.2.4 Effet du bruit instrumental

Dans le cas d'une carte réelle, le bruit instrumental s'ajoute au signal dû aux galaxies. On supposera ici ce bruit gaussien :

$$p_B(B) = \text{normalisation} \times exp(-\frac{B^2}{2\sigma^2}). \tag{4.25}$$

La transformée de Fourier d'une gaussienne étant une gaussienne, on a

$$\hat{p}_B(\omega) = \text{normalisation} \times exp(-\frac{\omega^2\sigma^2}{2}). \tag{4.26}$$

La densité de probabilité de $D_{gal+bruit}$ ($D_{gal+bruit} = D + B$) sera donc donnée par

$$p_{D_{gal+bruit}} = p_D * p_B, \tag{4.27}$$

soit dans l'espace de Fourier :

$$\hat{p}_{D_{gal+bruit}} = \hat{p}_D \times \hat{p}_B. \tag{4.28}$$

Finalement, en transformant le produit d'exponentielles en exponentielle d'une somme, on a

$$p_{D_{gal+bruit}}(D_{gal+bruit}) = \frac{1}{2\pi}\int_0^\infty exp\left(\int_0^\infty [n_+(F)e^{i\omega F} + n_-(F)e^{-i\omega F} - n_+(F) - n_-(F)]dF - \frac{\omega^2\sigma^2}{2}\right)e^{-i\omega D_{gal+bruit}}d\omega \tag{4.29}$$

4.2.5 Prise en compte d'une calibration relative

Les instruments de *Spitzer*, *Herschel* et BLAST font de la photométrie relative. Leur signal zéro est choisi arbitrairement. En général, on choisit le niveau zéro de manière à avoir une carte à moyenne nulle. On appelle μ la différence entre le zéro absolu (après soustraction des avant-plans zodiacaux et galactiques) et le zéro de la carte. La distribution $p_{D_{mes}}$, observée pour une calibration relative sera liée à p_D par

$$p_{D_{mes}}(D_{mes}) = p_D(D_{mes} + \mu), \tag{4.30}$$

soit dans l'espace de Fourier

$$\hat{p}_{D_{mes}}(\omega) = \hat{p}_D(\omega) \times e^{i\mu\omega}. \tag{4.31}$$

La dentité de probabilité de D_{mes} peut donc se calculer en utilisant

$$p_{D_{mes}}(D_{mes}) = \frac{1}{2\pi}\int_0^\infty exp\left(\int_0^\infty [n_+(F)e^{i\omega F} + n_-(F)e^{-i\omega F} - n_+(F) - n_-(F)]dF - \frac{\omega^2\sigma^2}{2} + i\mu\omega\right)e^{-i\omega D_{mes}}d\omega. \tag{4.32}$$

4.2.6 Développement d'un code d'analyse P(D)

J'ai développé des outils en langage IDL permettant de réaliser une analyse P(D). Ces outils utilisent une paramétrisation des comptages similaire à celle de Patanchon *et al.* (2009). Toutefois, une option permet de relier les noeuds avec une *spline* à la place d'une loi de puissance. Il est également possible de prendre en compte l'incertitude sur le bruit instrumental, en autorisant ce bruit à varier, mais en pénalisant les valeurs éloignées de la valeur mesurée. Un second terme additionnel peut être ajouté pour forcer un accord entre le modèle de comptages et le niveau du fond infrarouge mesuré par FIRAS. L'ajustement des paramètres est réalisé par méthode MCMC en utilisant l'algorithme Metropolis-Hastings. La loi de déplacement dans l'espace des paramètres est dans un premier temps choisie en calculant la matrice de Fisher du problème à une position initiale choisie manuellement (typiquement un modèle de comptages). Un première chaîne courte (∼10 000 pas) est lancée. Sa matrice variance-covariance est calculée. La nouvelle loi de déplacement est la loi gaussienne ayant la même matrice variance-covariance. On lance alors une nouvelle chaîne plus longue. On vérifie la convergence de celle-ci en utilisant les critères de Dunkley *et al.* (2005), qui utilisent les spectres de puissance des réalisations des paramètres.

Le calcul rapide de la vraisemblance du modèle à chaque pas pilote totalement la vitesse du programme. J'ai écrit un programme rapide permettant de réaliser cette opération. La première étape consiste à calculer $n_+(F)$ et $n_-(F)$ (voir Eqs. 4.19 et 4.20). Pour gagner du temps, ces fonctions sont calculées sur une grille logarithmique. Un histogramme de la PSF est utilisée pour calculer l'intégrale, au lieu de sommer sur tous les pixels de la PSF. On interpole ensuite ces fonctions sur une grille régulière afin de calculer sa transformée de Fourier dans Eq. 4.23. On soustrait alors la moyenne du signal à la grille en D, afin d'obtenir une distribution à moyenne nulle. Puis, on intègre par la méthode des trapèzes la fonction p(D) dans les bins de flux utilisés pour construire l'histogramme moyen des cartes analysées, afin d'en déduire la vraisemblance du modèle comptages testé (Patanchon *et al.* (2009))

$$ln(L) = \sum_i (n_i ln(p_i) - ln(n_i!)) + ln(N!), \tag{4.33}$$

où L est la vraisemblance, n_i le nombre de pixel de la carte dans le i-ième intervalle de flux, p_i est la valeur moyenne du nombre de pixels dans le même intervalle de flux, et N le nombre total de pixels.

4.3 Résultats : application aux données *Herschel*/SPIRE

Cette partie reprend les résultats que j'ai obtenus dans le cadre de la collaboration HerMES. Ces résultats ont été publiés dans Glenn *et al.* (2010). Ce résumé est centré sur ma contribution personnelle.

4.3.1 Motivations

L'analyse P(D) des données BLAST ne permettait pas de bien contraindre les comptages à bas flux. *Herschel*, grâce à une résolution angulaire et une sensibilité supérieure, permet d'obtenir des images d'une bien meilleure qualité. Lors de la phase de démonstration scientifique, quatre champs extragalactiques ont été observés par le programme HerMES[4] (programme de temps garanti de l'instrument SPIRE) : GOODS-N (champs profond), Lockman-North (champs intermédaire), FLS (champs large) et Lockman SWIRE (champs très large). La combinaison de champs larges et profonds fournit de fortes contraintes sur les sources faibles et brillantes, permettant de mesurer les comptages avec une précision et une profondeur inégalée, qui vont fortement contraindre les populations infrarouges à haut *redshift*.

4.3.2 Le problème de la paramétrisation des comptages

Le choix de la position des noeuds utilisés lors de l'analyse P(D) est le fruit d'un compromis entre deux effets. Si les noeuds sont trop éloignés les uns des autres, le modèle paramétrique ne peut pas reproduire fidèlement la forme complexe des comptages. En revanche, une trop grande proximité entre deux noeuds induit de fortes dégénérescences, et fait exploser les barres d'erreur.

On peut tester la pertinence d'un choix de noeuds en utilisant les comptages issus de modèles (Lagache *et al.* (2004), Le Borgne *et al.* (2009), Valiante *et al.* (2009)). On calcule les histogrammes attendus pour le modèle initial. Puis, on construit un second modèle de comptages en plaçant les noeuds exactement sur la courbe du modèle initial et en utilisant une interpolation entre les noeuds avec une loi de puissance ou une *spline*. Puis, on calcule la différence entre les histogrammes des deux modèles. Ces considérations nous ont montré qu'il est nécessaire d'utiliser environ 3 noeuds par décades, et qu'une interpolation *spline* est plus adéquate qu'une interpolation en loi de puissance. Néanmoins, aucune paramétrisation n'est parfaite, et la position d'un noeud ne peut en aucun cas être considérée comme une mesure des comptages à un flux précis. Les points P(D) ne doivent donc pas être ajustés tels quels par un modèle d'évolution.

4.3.3 Le problème du filtrage

Le formalisme du P(D) suppose que les sources sont distribuées de manière poissonienne. En réalité, les sources infrarouges ont tendance à se regrouper (Lagache *et al.* (2007),Viero *et al.* (2009)). Ceci élargit le pic de l'histogramme. En effet, la largeur de ce

---

[4] *Herschel Multi-tiered Extragalactic Survey*

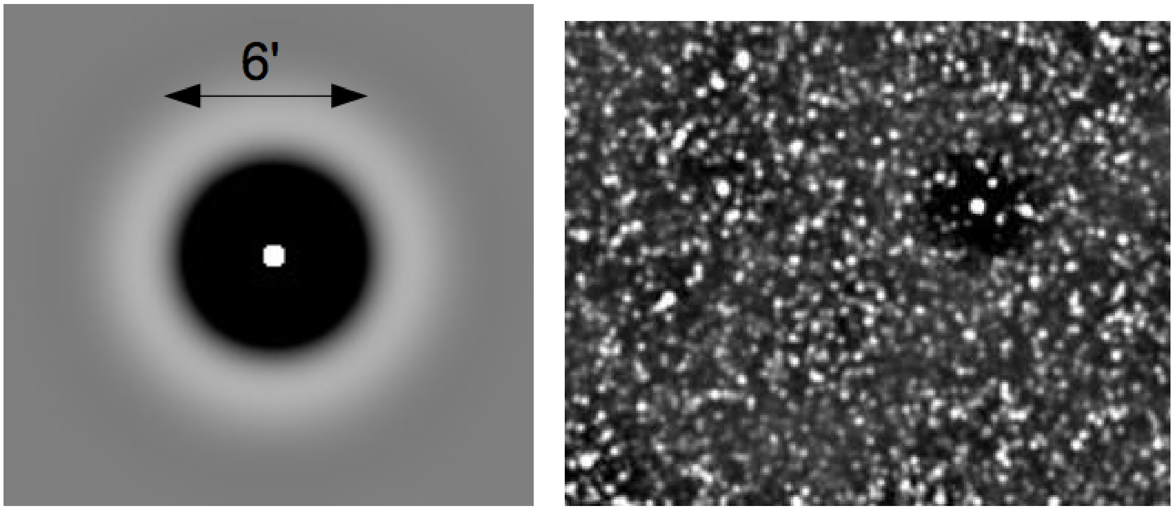

FIGURE 4.2 – (A gauche) : PSF après filtrage. (A droite) : carte simulée après filtrage.

pic est liée, en l'absence de bruit instrumental, au spectre de puissance du fond infrarouge (Takeuchi et Ishii (2004), Patanchon *et al.* (2009)) :

$$\sigma = \int\int P(k)\widetilde{PSF} d^2k, \tag{4.34}$$

où P(k) est le spectre de puissance des cartes infrarouges et $\widetilde{PSF}$ la transformée de Fourier de la PSF. Lagache *et al.* (2007) sépare le spectre de puissance en plusieurs contributions : une contribution poissonienne qui ne dépend pas de l'échelle, et une partie due au regroupement des sources qui domine au delà d'une dizaine d'arcminutes. La partie due au regroupement n'est pas prise en compte par le formalisme du P(D). On peut par exemple comparer la variance entre sous-champs de 500×500 pixels dans Lockman SWIRE. La variance sur le nombre de sources dans un bin de flux est environ deux fois plus grande que pour une loi de Poisson. En revanche, elle est poissonienne si on soustrait la moyenne de chaque sous-carte ou si on filtre les grandes échelles. Ceci confirme que les effets du regroupement sur l'histogramme agissent essentiellement aux grandes échelles.

Pour cette raison, nous avons filtré les échelles supérieures à 6 arcminutes dans les cartes et les PSFs. Ce filtrage permet également de réduire l'effet des cirrus qui dominent la puissance aux très grandes échelles (>30 arcmin). En pratique, ce filtrage est réalisé de la manière suivante. On calcule la transformée de Fourier de la carte $\tilde{m}$, où on a mis à zéro les pixels non observés. On construit un masque où les échelles supérieures à 6 arcminutes sont à 0 et les autres à 1, qu'on lisse avec un noyau gaussien de 1.8 arcmin. Ce lissage limite les oscillations autour des sources brillantes. On multiplie ensuite $\tilde{m}$ par ce masque lissé, et on réalise la transformée inverse. La Fig. 4.2 montre l'effet du filtrage sur la PSF et sur la carte. La PSF est peu modifiée aux petites échelles, mais présentent de larges

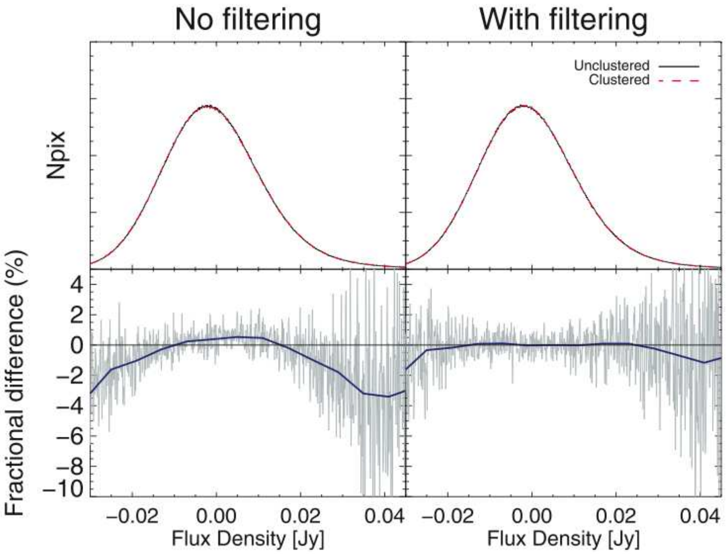

FIGURE 4.3 – (Panneaux du haut) Histogramme des pixels de cartes simulées du champ Lockman SWIRE avec (tirets rouges) et sans regroupement (ligne noire). (Panneaux du bas) : Différence entre les histogrammes avec et sans regroupement. (Panneaux de gauche) : Cartes non filtrées. (Panneaux de droite) : Cartes filtrées.

ailes négatives qui sont également visibles dans la carte. Notre filtrage étant parfaitement linéaire, il est pris en compte naturellement par le formalisme du P(D), si on utilise une carte et une PSF filtrées de la même manière.

Nous avons validé cette méthode en comparant deux simulations avec et sans regroupement, basée sur les même comptages (voire Fig. 4.3). L'écart entre les deux histogrammes est bien plus petit dans le cas où les grandes échelles ont été filtrées. J'ai par la suite réalisé des simulations avec regroupement des champs observés à partir du travail de Fernandez-Conde *et al.* (2008). J'ai alors appliqué mes outils d'analyse P(D) et retrouvé des comptages en accord avec les comptages initiaux de la simulation. Ceci confirme l'efficacité de la méthodologie choisie.

Le filtrage des grandes échelles n'est pas le seul filtrage à tester. En effet, Patanchon *et al.* (2009) ont également lissé leur carte par la PSF, afin de réduire le bruit instrumental et améliorer la précision des comptages obtenus. Toutefois, les champs profonds SPIRE (GOODS-N et Lockman-N) sont peu affectés par le bruit instrumental. Dans ce cas, j'ai montré qu'il n'est pas nécessaire de lisser la carte par la PSF en utilisant une analyse de Fisher.

4.3.4 Résultats

Nous avons appliqué ces méthodes aux données de la phase de démonstration scientifique de l'instrument SPIRE. Nous avons fait le choix de n'utiliser que 3 champs extragalactiques sur 4. En effet, le champs FLS est contaminé par un cirrus. Son utilisation présentait donc des risques d'effets systématiques. De plus, ce champ peu profond et moins large que Lockman SWIRE n'améliore que très faiblement les contraintes sur les comptages. Nous avons donc ajusté simultanément les trois champs suivant : GOODS-N, Lockman-N et Lockman SWIRE.

Nous avons développé deux codes d'analyse distincts afin de s'assurer de la qualité de nos résultats : mon code IDL utilisant une interpolation spline entre les noeuds, et le code d'Alex Conley en C++ optimisé pour une interpolation en loi de puissance. Les deux codes ont fourni des résultats compatibles (voir Fig. 4.4). A l'exception du premier et du dernier noeud, tous les autres sont incompatibles avec zéro. Nous sommes donc en mesure de présenter une estimation robuste des comptages entre 2 mJy et 1 Jy[5], sondant un domaine de flux jamais exploré à ces longueurs d'ondes. En effet, à cause de la confusion, les comptages résolus sont très difficiles à estimer en dessous de 20 mJy (Oliver *et al.* (2010)).

Ces résultats peuvent être comparés avec les mesures réalisées par BLAST (Fig. 4.4, panneau du haut). Les comptages que j'ai mesurés dans les données BLAST sont en bon accord avec l'analyse P(D) des données *Herschel*. En revanche, le pic des comptages différentiels normalisés de Patanchon *et al.* (2009) est trop haut. Ceci est probablement dû à un effet de paramétrisation. En effet, ils n'utilisent qu'un seul noeud entre 0.1 et 100 mJy, alors que les comptages évoluent rapidement dans cette région. On peut également comparer nos résultats aux comptages de sources résolues par *Herschel*. L'accord avec l'analyse de Oliver *et al.* (2010) des mêmes champs est très bon. Il y a également un bon accord avec les comptages H-ATLAS (16 deg^2) de Clements *et al.* (2010). Enfin, nous avons comparé nos mesures avec une large collection de modèles. Aucun d'eux n'est en mesure de reproduire les comptages dans les trois bandes. Ceci montre à quel point l'analyse P(D) fournit des contraintes fortes sur l'évolution des galaxies infrarouges.

4.4 Conclusion

Dans ce chapitre, nous avons présenté le formalisme sur lequel se base la méthode du P(D). Cette méthode se révèle très puissante pour mesurer les comptages dans des régimes de confusion extrême. Contrairement à l'analyse par empilement, cette méthode ne

---

[5] La position des noeud étant arbitraire, la postion du dernier noeuds bien contraint l'est tout autant. En revanche, différents tests nous ont montré qu'il est impossible de contraindre plus de deux noeuds en dessous de 5 mJy qui est le bruit de confusion à 1-σ de SPIRE.

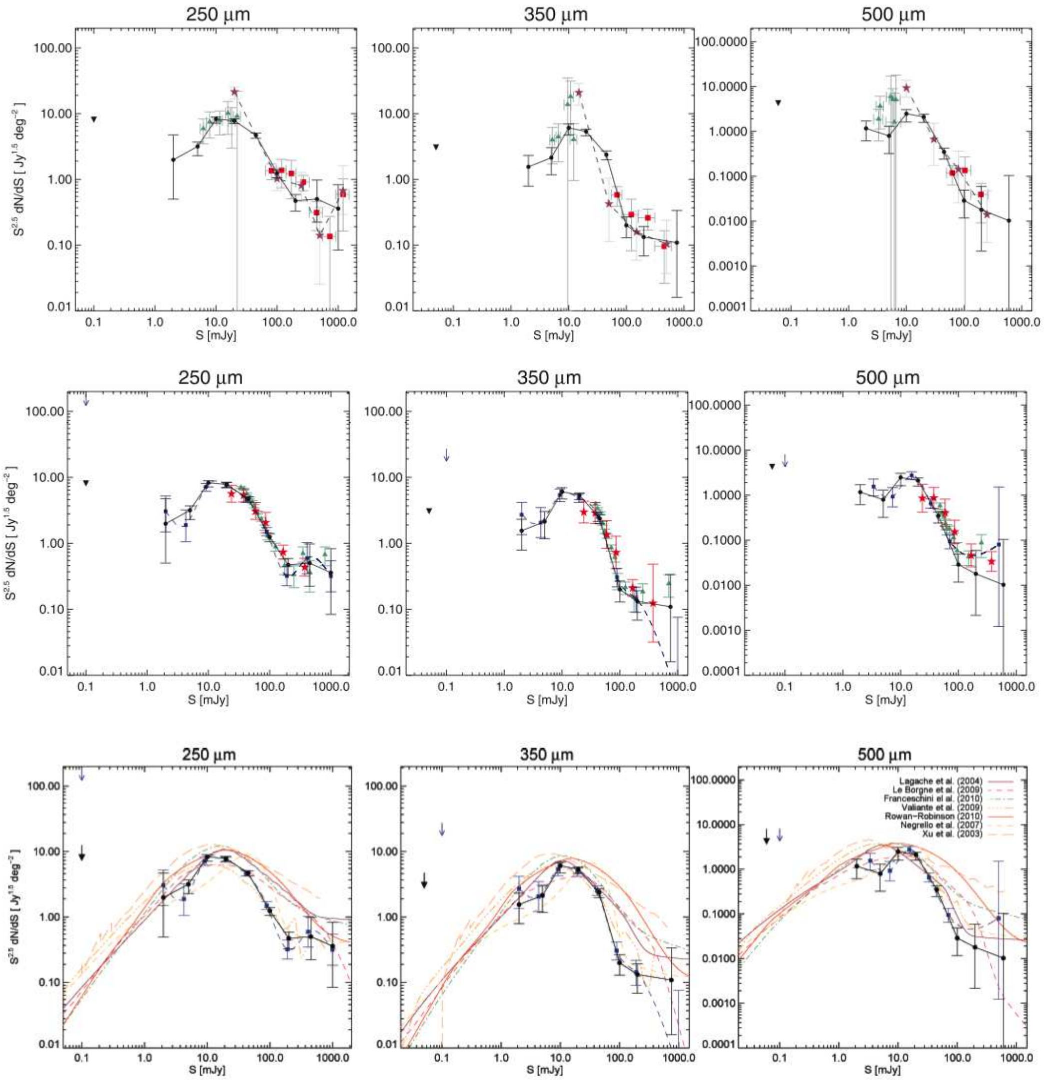

FIGURE 4.4 – Comptages différentiels normalisés à 250, 350 et 500 μm. L'analyse P(D) des données *Herschel* est en bleu (mon code) et en noir (code de Alex Conley). (Panneau du haut) : Comparaison avec les mesures BLAST : comptages résolus en rouge, comptages par empilement en vert (Béthermin *et al.* (2010b)), et analyse P(D) en violet (Patanchon *et al.* (2009)). (Panneau du mileu) : Comparaison avec les comptages résolus *Herschel* dans HerMES (Oliver *et al.* (2010) en rouge et dans H-ATLAS en vert (Clements *et al.* (2010))). (Panneau du bas) : Comparaison avec une collection de modèles.

suppose aucun lien entre longueurs d'onde. Ceci permet de confirmer avec une méthode et des données différentes ce qui a été obtenu par empilement dans les données BLAST (voir Chap. 3). De plus, les points obtenus par analyse P(D) des données SPIRE sont beaucoup plus contraignants pour les modèles que ce qui a été obtenu avec la méthode de l'empilement dans les données BLAST.

Toutefois, la mise en oeuvre d'une analyse P(D) n'est pas triviale, et l'interprétation des résultats doit être faite avec précaution. Par exemple, le choix de la position des noeuds utilisés affecte fortement le résultat. Le pic des comptages différentiels normalisés trouvé par Patanchon *et al.* (2009) est ainsi trop haut d'un facteur 2. Le choix des noeuds utilisés pour réaliser l'analyse P(D) des données HerMES cherche à limiter ce type d'effets, bien qu'un biais de paramétrisation soit inévitable. En fait, l'idéal serait d'ajuster directement les paramètres d'un modèle d'évolution sur l'histogramme des cartes.

4.5 Article : *HerMES : deep galaxy number counts from a P(D) fluctuation analysis of SPIRE Science Demonstration Phase observations*

Un article synthétisant les travaux réalisés sur l'analyse P(D) dans le cadre de la collaboration HerMES a été publié dans MNRAS.

Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 30 September 2010 (MN LaTeX style file v2.2)

HerMES: Deep Galaxy Number Counts from a $P(D)$ Fluctuation Analysis of SPIRE Science Demonstration Phase Observations

J. Glenn,[1]⋆ A. Conley,[1] M. Béthermin,[2] B. Altieri,[3] A. Amblard,[4] V. Arumugam,[5] H. Aussel,[6] T. Babbedge,[7] A. Blain,[8] J. Bock,[8,9] A. Boselli,[10] V. Buat,[10] N. Castro-Rodríguez,[11,12] A. Cava,[11,12] P. Chanial,[7] D.L. Clements,[7] L. Conversi,[3] A. Cooray,[4,8] C.D. Dowell,[8,9] E. Dwek,[13] S. Eales,[14] D. Elbaz,[6] T.P. Ellsworth-Bowers,[1] M. Fox,[7] A. Franceschini,[15] W. Gear,[14] M. Griffin,[14] M. Halpern,[16] E. Hatziminaoglou,[17] E. Ibar,[18] K. Isaak,[14] R.J. Ivison,[18,5] G. Lagache,[2] G. Laurent,[19] L. Levenson,[8,9] N. Lu,[8,20] S. Madden,[6] B. Maffei,[21] G. Mainetti,[15] L. Marchetti,[15] G. Marsden,[16] H.T. Nguyen,[9,8] B. O'Halloran,[7] S.J. Oliver,[22] A. Omont,[23] M.J. Page,[24] P. Panuzzo,[6] A. Papageorgiou,[14] C.P. Pearson,[25,26] I. Pérez-Fournon,[11,12] M. Pohlen,[14] D. Rigopoulou,[25,27] D. Rizzo,[7] I.G. Roseboom,[22] M. Rowan-Robinson,[7] M. Sánchez Portal,[3] B. Schulz,[8,20] Douglas Scott,[16] N. Seymour,[24] D.L. Shupe,[8,20] A.J. Smith,[22] J.A. Stevens,[28] M. Symeonidis,[24] M. Trichas,[7] K.E. Tugwell,[24] M. Vaccari,[15] I. Valtchanov,[3] J.D. Vieira,[8] L. Vigroux,[23] L. Wang,[22] R. Ward,[22] G. Wright,[18] C.K. Xu[8,20] and M. Zemcov[8,9]

[1] *Dept. of Astrophysical and Planetary Sciences, CASA 389-UCB, University of Colorado, Boulder, CO 80309, USA*
[2] *Institut d'Astrophysique Spatiale (IAS), bâtiment 121, Université Paris-Sud 11 and CNRS (UMR 8617), 91405 Orsay, France*
[3] *Herschel Science Centre, European Space Astronomy Centre, Villanueva de la Cañada, 28691 Madrid, Spain*
[4] *Dept. of Physics & Astronomy, University of California, Irvine, CA 92697, USA*
[5] *Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[6] *Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu - CNRS - Université Paris Diderot, CE-Saclay, pt courrier 131, F-91191 Gif-sur-Yvette, France*
[7] *Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[8] *California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA*
[9] *Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[10] *Laboratoire d'Astrophysique de Marseille, OAMP, Université Aix-marseille, CNRS, 38 rue Frédéric Joliot-Curie, 13388 Marseille cedex 13, France*
[11] *Instituto de Astrofísica de Canarias (IAC), E-38200 La Laguna, Tenerife, Spain*
[12] *Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38205 La Laguna, Tenerife, Spain*
[13] *Observational Cosmology Lab, Code 665, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*
[14] *Cardiff School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA, UK*
[15] *Dipartimento di Astronomia, Università di Padova, vicolo Osservatorio, 3, 35122 Padova, Italy*
[16] *Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada*
[17] *ESO, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany*
[18] *UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[19] *Southwest Research Institute, Boulder, Colorado 80302, USA*
[20] *Infrared Processing and Analysis Center, MS 100-22, California Institute of Technology, JPL, Pasadena, CA 91125, USA*
[21] *School of Physics and Astronomy, The University of Manchester, Alan Turing Building, Oxford Road, Manchester M13 9PL, UK*
[22] *Astronomy Centre, Dept. of Physics & Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[23] *Institut d'Astrophysique de Paris, UMR 7095, CNRS, UPMC Univ. Paris 06, 98bis boulevard Arago, F-75014 Paris, France*
[24] *Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK*
[25] *Space Science & Technology Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK*
[26] *Institute for Space Imaging Science, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada*
[27] *Astrophysics, Oxford University, Keble Road, Oxford OX1 3RH, UK*
[28] *Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK*

30 September 2010

ABSTRACT

Dusty, star forming galaxies contribute to a bright, currently unresolved cosmic far-infrared background. Deep *Herschel*-SPIRE images designed to detect and characterize the galaxies that comprise this background are highly confused, such that the bulk lies below the classical confusion limit. We analyze three fields from the HerMES programme in all three SPIRE bands (250, 350, and 500 μm); parameterized galaxy number count models are derived to a depth of $\sim$ 2 mJy/beam, approximately 4 times the depth of previous analyses at these wavelengths, using a $P(D)$ (probability of deflection) approach for comparison to theoretical number count models. Our fits account for 64, 60, and 43 per cent of the far-infrared background in the three bands. The number counts are consistent with those based on individually detected SPIRE sources, but generally inconsistent with most galaxy number counts models, which generically overpredict the number of bright galaxies and are not as steep as the $P(D)$-derived number counts. Clear evidence is found for a break in the slope of the differential number counts at low flux densities. Systematic effects in the $P(D)$ analysis are explored. We find that the effects of clustering have a small impact on the data, and the largest identified systematic error arises from uncertainties in the SPIRE beam.

Key words: Submillimeter Galaxies – Cosmology: Observations

1 INTRODUCTION

The cosmic far-infrared background (hereafter CFIRB) provides unique information on the history of energy injection in the Universe by both star formation and active galactic nuclei. First detected by the *COBE* satellite (Puget et al. 1996; Fixsen et al. 1998), the CFIRB contains a large amount of energy, indicating that the total luminosity from thermal dust emission is comparable to the integrated UV/optical energy output of galaxies (Guiderdoni et al. 1997).

Galaxy surveys, both from the ground (with SCUBA, LABOCA, Bolocam, AzTEC, and MAMBO at 850 μm, 870 μm, 1.1 mm, 1.1 mm, and 1.3 mm, respectively) and from space using *IRAS* (at 12, 25, 60 and 100 μm), *ISO* (at 15, 90, and 170 μm), and *Spitzer* (at 3.6 to 160 μm), found high number counts compared to non-evolving galaxy number counts models. This implied that strong evolution of the source populations must have occurred, challenging contemporary galaxy evolution models (Saunders 1990; Scott et al. 2002; Lagache et al. 2003). Deeper number counts test galaxy formation models more severley. By stacking *Spitzer* MIPS 24 μm sources, at least 80% of the CFIRB was resolved at 70 μm and 65% at 160 μm (Dole et al. 2006; Béthermin et al. 2010a). A small fraction (10-20%) has been resolved in the submillimeter in blind sky surveys from ground-based observatories, but it is possible to go deeper by taking advantage of gravitational lensing. At 850 μm this approach has resolved 60% or more of the background in small fields (Smail et al. 2002; Zemcov et al. 2010).

A $P(D)$ – probability of deflection – analysis of Bolocam observations of the Lockman Hole (Maloney et al. 2005) demonstrated that a fluctuation analysis can provide more stringent constraints on source number counts than those derived by extracting individual sources, for which the threshold must be set high enough to ensure a minority of false detections. $P(D)$ techniques were first developed for application to radio observations (Scheuer 1957), but have since been widely applied to other regimes. $P(D)$ was used to account for the majority of the X-ray background long prior to the availability of sufficiently deep imaging to resolve individual sources (Barcons 1994), to extend deep infrared counts (Oliver et al. 1997), and in the sub-mm to SCUBA (Hughes et al. 1998), LABOCA (Weiß et al. 2009), and AzTEC (Scott et al. 2010) data. The depth of a $P(D)$ analysis is set by the flux density at which the number of sources per beam becomes large. The resulting contribution to the $P(D)$ becomes that of a Poissonian distribution with a large mean, which becomes difficult to distinguish from the nearly-Gaussian instrumental noise. An often-used rule of thumb for the maximum depth is one source per beam, but the precise limit depends on the survey area, the shape of the underlying counts, and how precisely the instrumental noise is known. In practice, for rapidly rising source counts at faint fluxes, this is considerably deeper than the limits for a source-extraction approach. Fluctuation analyses are well-suited to determination of source number counts in the case where the dynamic range of detected sources is not large because of confusion. Deep number counts are interesting because they allow us to measure the sources responsible for the bulk of the CFIRB, and because they probe intrinsically fainter galaxies which may have better matching counterparts in the local Universe.

Recently, a $P(D)$ analysis was performed on 250, 350, and 500 μm observations of a 10 deg^2 field (GOODS-S) with a 0.8 deg^2 deep inner region from the balloon-borne BLAST telescope, using duplicate SPIRE detector technology (Patanchon et al. 2009, hereafter P09). Differential number counts were estimated down to 20, 15, and 10 mJy in the three bands, respectively. Below these thresholds, upper limits were provided. Combined with 24 μm observations, Devlin et al. (2009) concluded that a large fraction ($> 1/2$) of the CFIRB comes from galaxies with $z > 1.2$. Also from BLAST observations, Marsden et al. (2009) concluded that 24 μm-selected galaxies can account for the entire CFIRB based on a stacking analysis. These results confirm that fluctuation and stacking analyses have substantial power in elucidating the sources of the CFIRB. Such techniques will also be necessary for SPIRE observations because galaxy models predict that at the confusion limit SPIRE is expected to resolve only a small fraction of the CFIRB (Lagache et al. 2003; Fernandez-Conde et al. 2008). A recent source extraction-based analysis of the SPIRE Science Demonstration Phase (SDP) data – the same data used in this paper – directly resolved 15, 10, and 6 per cent of the CFIRB at 250, 350, and 500 μm, respectively (Oliver et al. 2010). At shorter wavelengths, Berta (2010) directly

---

⋆ E-mail: `jason.glenn@.colorado.edu`

resolved 52 and 45 per cent of the CFIRB at 100 and 160 μm using *Herschel*-PACS SDP data.

2 DATA

The observations used in this analysis were obtained with the SPIRE instrument (Griffin et al. 2010) on the *Herschel* Space Observatory (Pilbratt et al. 2010) as part of the HerMES programme[1] (Oliver et al. 2010, in prep) during the SDP. SPIRE observes simultaneously in three passbands: 250, 350 and 500 μm. The on-orbit beam sizes, including the effects of the scanning strategy, are 18.1, 25.2, and 36.6 arcseconds, respectively, with mean ellipticities of 7 - 12%. The calibration is based on observations of Neptune, and is described in Swinyard et al. (2010). Observations of five fields were obtained during SDP, but only three are used in this analysis: GOODS-N, Lockman-North, and Lockman-SWIRE. Their properties are summarized in table 1. The Lockman-North region are contained within the shallower Lockman-SWIRE field. The HerMES SDP fields omitted from this analysis are: FLS, which was left out because it is the same depth as the much-larger Lockman-SWIRE field and is significantly contaminated by infrared cirrus, and Abell 2218, because the strong lensing in this field complicates the interpretation of the background number counts.

The detector (bolometer) timelines were processed using the standard SPIRE pipeline, which detects cosmic rays and removes instrumental signatures and temperature drifts (Dowell et al. 2010). The maps were produced using the SMAP package (Levenson et al. 2010, in prep) using 1/3 beam-FWHM pixels (6, 8 1/3, and 12 arcsec); this is a compromise between adequately sampling the beam and maintaining even coverage over the map. Samples flagged as contaminated by cosmic-rays were excluded. Each map was masked to form an even-coverage region and was mean subtracted. In addition to the even-coverage mask, a small amount of additional masking was required, as there are five resolved sources in the Lockman-SWIRE field. These sources are relatively bright, but are not the brightest in the field. Since the $P(D)$ formalism is based on unresolved point sources, we mask these objects with a 2 arcmin circular mask, and then correct our final number counts using the measured flux of each excluded source. Our instrumental noise estimates are based on the technique of Nguyen et al. (2010), and are assumed to have 5% uncertainty, which represents only the uncertainty for a fixed calibration. The overall SPIRE calibration error is discussed in § 5.1. The resulting pixel flux density histograms are shown in figure 1.

Smoothing the maps by the beam (via cross-correlation) is beneficial for finding isolated point sources. For BLAST observations, Chapin et al. (2010) show that the standard point-source-optimized filter should be modified in the presence of confusion noise. However, there is no guarantee this will benefit a $P(D)$ analysis. Smoothing the map has the effect of broadening the effective beam, which decreases the depth that the $P(D)$ can probe, while also reducing (but correlating) the instrumental noise. We empirically determined if smoothing is beneficial for our analysis by fitting simulated data with and without smoothing and comparing the scatter in the recovered model parameters to the error estimates, and found that it helps for all but our deepest map (GOODS-N); note that our GOODS-N data are several times deeper (relative to the confusion noise) than the deepest BLAST observations. It is likely that some amount of spatial de-convolution would be beneficial for the GOODS-N field, but since this would significantly complicate the instrumental noise properties of our maps, and hence require extensive testing, we do not pursue de-convolution here, but do beam-smooth the two shallower fields. In addition, we have applied a 6 arcmin high-pass spatial filter to our maps to reduce the effects of clustering on our results. The motivation for this is discussed in § 3.3.

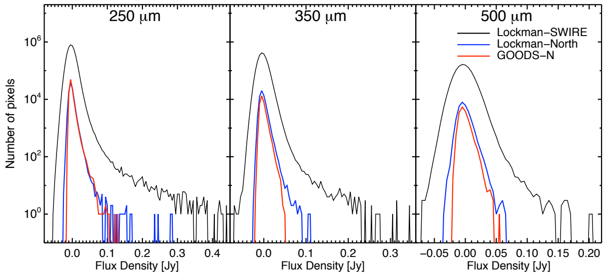

Figure 1. Histogram of pixel flux densities for the three fields considered in this paper in 5 mJy bins. The Lockman-SWIRE field is considerably shallower than the others, but it is the only field large enough to probe the bright end of the source distribution. The maps have been mean subtracted. The binning here is for display purposes and does not correspond to the binning used in the actual analysis, which is much finer. In all cases the pixel histograms show clear non-Gaussianity despite the Gaussian nature of the noise, indicating that a significant point-source contribution is present.

A $P(D)$ analysis is critically dependent on measurements of the effective beam (which includes the effects of the map making and reduction, as well as any smoothing applied). Our beam map is based on in-flight fine-scan (highly oversampled) observations of Neptune with a large number of repeats and small offsets between each scan. These observations allowed us to measure the beam with finer resolution than our data maps to properly match the SPIRE calibration (which is timestream rather than map based), and to build beam maps for the individual bolometers. We corrected these observations for the relative motion of Neptune during the scans using the HORIZONS[2] ephemeris computation service at the orbit of *Herschel*. The Neptune observations are deep enough that the third Airy ring is clearly detected for the array-averaged beams. As discussed later (§ 5.1), beam uncertainties are our largest identified systematic.

3 METHOD

We first describe the basic $P(D)$ framework, then discuss our particular implementations and the filtering we have applied to limit the effects of clustering.

3.1 Description of $P(D)$

If $dN/dS\,(S)$ is the differential number counts per solid angle and S the flux density, then the mean number density of sources per unit solid angle with observed flux densities between x and $x + dx$ is

$$R(x)\;dx = \int_{\Omega} \frac{dN}{dS}\left(\frac{x}{b}\right) b^{-1}\, d\Omega\, dx \tag{1}$$

---

[1] http://www.hermes.sussex.ac.uk/

[2] http://ssd.jpl.nasa.gov/?horizons

Table 1. HerMES SPIRE Science Demonstration Phase observations used in this paper

Field	Size deg^2	RA deg^2	Dec deg^2	Scan Rate arcsec/s	Repeats	σ_{250} mJy/beam	σ_{350} mJy/beam	σ_{500} mJy/beam
GOODS-N	0.29	189.23	62.24	30	30	1.77	1.59	1.89
Lockman-North	0.41	161.50	59.02	30	7	3.58	3.16	4.41
Lockman-SWIRE	13.6	162.0	58.11	60	2	9.47	8.47	11.99

The σ values for each band and field are the instrumental noise per pixel before any filtering or smoothing is applied. The confusion noise (the signal in this analysis) is ~ 6 mJy/beam in all bands (Nguyen et al. 2010).

where b is the beam function (not necessarily peak normalized). Ignoring clustering, the probability distribution of sources is Poissonian. The probability distribution function (pdf) for the observed flux in each sky area unit (usually a map pixel) is the convolution of the pdfs for each flux interval over all fluxes; this quantity is called the $P(D)$. Rewriting the above in terms of characteristic functions and denoting the inverse Fourier transform by F_{ω}^{-1},

$$P(D) = F_{\omega}^{-1}\left[\exp\left(\int_0^{\infty} R(x)\exp(i\omega x)\,dx - \int_0^{\infty} R(x)\,dx\right)\right]. \qquad (2)$$

The mean of the $P(D)$ is

$$\mu = \int xR\,dx = \int b\,d\Omega \int S\;dN/dS(S)\;dS,$$

and the variance is

$$\sigma_P^2 = \int b^2 d\Omega \int S^2 dN/dS(S)\;dS.$$

For real observations, the instrumental noise contribution must also be included. Our observations are not sensitive to the mean flux in the maps. Therefore, it is useful to subtract off the mean of the $P(D)$ during construction. Only in the case of very simple models for dN/dS combined with trivial beams is it possible to compute $P(D)$ analytically – an example is given in Scheuer (1957), but even this is only valid for a restricted range of parameters. For an effective beam that is not strictly positive (due to filtering, for example), the $P(D)$ is the convolution of the individual $P(D)$s for the positive beam and for the negative beam (P09). Azimuthally averaging the beam does not preserve the $P(D)$, so it is necessary to use the full 2D beam map.

The log likelihood ($\log\mathcal{L}$) of a dataset relative to a particular model is given (to within an irrelevant normalizing constant) by

$$\log\mathcal{L} = \sum_i \log P(D_i),$$

where D_i is the value of the i^{th} pixel. Usually it is more convenient to bin the data. As long as the individual bins are small compared to the width of the $P(D)$, the two formulations are practically equivalent. Then

$$\log\mathcal{L} = \sum_k h_k \log P(x_k), \qquad (3)$$

where h_k is the number of pixels in the histogram bin centred at flux density x_k. The alternative of using the χ^2 as the fit statistic under-weights bins with a small number of pixels in them because the uncertainty in such a bin is not well modeled by $\sqrt{h_k}$, and is not recommended.

This treatment assumes that different pixels are uncorrelated, which is not true unless the beam is much smaller than a pixel. A source at one location will affect neighbouring values over an area about equal to the area of the beam. The result is that, if applied naively, fits based on the above likelihood will underestimate the model errors. Properly treating this effect requires developing the $P(D)$ formalism in terms of multi-variate Poisson distributions, which is computationally infeasible. P09 recommend dividing the likelihood by the beam area in pixels ($\mathcal{L} \mapsto \mathcal{L}/A_b$) in order to correct for this effect, which amounts to approximately correcting the likelihood for the number of independent samples in the map. This is an ad hoc approach, but in the absence of a better alternative, we have adopted a similar method. However, based on Monte-Carlo simulations of synthetic data sets, we find that this correction factor is overly conservative, as discussed below.

Another approximation in the above treatment which is not valid for real data is that the sources are not Poisson distributed due to clustering. Our approach to this effect is described in § 3.3.

3.2 Implementation

We have developed two independent $P(D)$ analysis packages and checked them against each other. Given the large number of parameters and the non-linear nature of the problem, both make use of Markov Chain Monte Carlo (MCMC) methods. Overviews of MCMC methods can be found in MacKay (2002); Lewis & Bridle (2002); Dunkley et al. (2005). The important aspects of an MCMC implementation are the burn-in criterion and the proposal density. The burn-in criterion is the rule used to determine whether the fit has converged on the region of maximum likelihood. Once the fit has converged, subsequent steps are drawn from the posterior probability of the model given the data, and only these steps are used to measure the errors and values of the parameters. The proposal density is used to propose the next step in the Markov Chain from the current step. Any proposal density that can visit all valid parameters is correct, but a well chosen density can dramatically improve the efficiency of the fitting procedure.

The first code is written in IDL[3] and the burn-in criterion is based on the power spectrum of the single chain (Dunkley et al. 2005). The first chain step within $\Delta\log\mathcal{L} = 2$ of the best-fit parameters is taken as the start of the converged sampling. A Fisher-matrix approximation to the $P(D)$ fit is used for the proposal density. This code interpolates in log space from a moderate number (~ 80) of flux densities to calcluate the $P(D)$. The other code is written in C++, and is explicitly parallel. It uses the Gelman-Rubin criterion for burn-in (Gelman & Rubin 1992), which is based on computing the variance between chains and directly provides the point of convergence. This code does not use interpolation when computing the $P(D)$, but supports a more limited range of models. The proposal density is a multi-variate Gaussian estimated from the previous fit steps, and is frozen in at burn-in. We have checked these codes against each other on simulated data, and find good agreement.

---

[3] Interactive Data Language: http://www.ittvis.com/ProductServices/IDL.aspx

Our $P(D)$ methodology is almost identical to that described in P09 except as follows. First, P09 explicitly fit for the mean of pixel values in the map (μ). Since we can analytically predict the mean of the $P(D)$ for a given set of model parameters, we simply shift the mean to zero explicitly during construction. The input map is also mean-subtracted, and the uncertainty in this subtraction contributes negligibly to our error budget. Second, P09 fit to the instrument noise explicity for each field except for the deepest section of their map. Instead, we marginalize over the noise for all fields in our full fits, but use the measurements of Nguyen et al. (2010) as a prior, assuming an Gaussian uncertainty of 5%. At low flux densities, the number of sources per beam is large, and hence the contribution to the $P(D)$ is almost Gaussian. Therefore, the values of dN/dS for the faintest flux densities probed and the noise level are nearly degenerate, and hence fixing the noise will tend to under-estimate the uncertainties in the model parameters at the faint end.

We have developed a simple simulation framework to test these codes and their sensitivity to various effects such as $1/f$ noise. As inputs we consider two types of catalogs that should be representative of the sub-mm sky: the P09 models, and the simulations of Fernandez-Conde et al. (2008). The fits to the P09 models are easier to compare with the inputs, but the Fernandez-Conde et al. (2008) models include clustering effects.

A fake sky is generated from the input catalogue, and scanned using the pointing information from the actual SPIRE observations. Different noise levels (white and $1/f$) can be specified. These data are then run through the same map-making pipeline as the real data. In addition to the simulated science data, we also simulate observations of Neptune using the same framework to determine the beams we use when fitting the simulated data. These simulations use simple Gaussian beams with FWHMs similar to those measured on-orbit, and account for characteristics of the data introduced by the mapping pipeline, but do not simulate errors in the lower-level SPIRE pipeline (pointing errors, crosstalk-corrections, etc.). We use them to quantify the effects of $1/f$ noise, uneven coverage, clustering, and smoothing by the beam on our maps. The SPIRE $1/f$ knee frequency is a few mHz, corresponding to a spatial scale of approximately 3 degrees for a scan speed of 30 arcseconds per second, and our map-making algorithm reduces this already small amount as discussed in Levenson et al. (2010, in prep). We find that the remaining amount, as well as the uneven coverage, introduces negligible bias in our fits, but that clustering can have measureable effects on our largest maps, as discussed in § 3.3.

In addition, we have determined the appropriate correction for pixel-pixel correlations using the same framework and a large number of simulated HerMES datasets. We find that the correct normalization factor varies with signal-to-noise ratio of the map, and whether it has been additionally smoothed with the beam. If the map is beam-smoothed, then the beam area factor is approximately correct, if slightly conservative for deeper fields; note that all of the maps in P09 were beam-smoothed. However, for deep, unsmoothed maps, this procedure clearly overestimates the uncertainties (by about a factor of 2 for GOODS-N). Rather than derive individual correction factors for each field, we have taken the more conservative approach of finding the largest correction factor (which therefore increases the uncertainties the most) for all of our fields, and applying it to all un-smoothed data. For the GOODS-N and Lockman-North observations, the correct normalization factor (without smoothing) is less than $A_b/3$. Because we do not have an exact formulation for this correction, we conservatively adopt $2A_b/5$. For the smoothed observations, we adopt the A_b normalization, also conservatively; this means that the two Lockman fields have the same correction factors, but the (un-smoothed) GOODS-N field has a different one.

3.3 Filtering

Clustering will affect the $P(D)$ distribution in two ways. First, the presence of clustering implies sample variance effects, so that the SDP fields may not be representative of the all-sky number counts. Second, the fact that the underlying counts are not Poisson distributed would change the shape of the $P(D)$ distribution even if we were somehow lucky enough to select a precisely average region of sky. This effect can be modeled if all of the n-point statistics of the source distribution are known (Barcons 1992). The effect on the width of the $P(D)$ is discussed in Appendix A of P09, although clustering is not purely limited to changing the width of the distribution. Only the 2-point function has been measured for the population sampled by SPIRE, and even this is not known at the flux densities important for our results. The first issue is discussed in § 5.1, and the second here. There are two effects: clustering on small scales between individual SMGs, and clustering on larger scales between groups of SMGs.

The framework for the clustering contribution to the $P(D)$ is given in Barcons (1992); Takeuchi & Ishii (2004). The contribution to the n^{th} moment is proportional to $\int P_n(\mathbf{k})\,\tilde{b}(\mathbf{k})^n\, d^2\mathbf{k}$, where $P_n(\mathbf{k})$ is the power spectrum of the n-point correlation function, and $\tilde{b}(\mathbf{k})$ is the Fourier transform of the beam. $\tilde{b}$ falls rapidly with $|\mathbf{k}|$ (e.g., an 18 arcsecond FWHM Gaussian beam has a 1/e value at k = 1.2 arcmin^{-1}, and the higher powers fall off even more rapidly). Thus, small scale clustering, which is implied by the measurements of, e.g., Blain et al. (2004), is filtered out by the beam on scales of less than about one to two arcminutes in our data.

Generically, P_n falls rapidly with $|\mathbf{k}|$, suggesting that high-pass filtering the maps may mitigate large-scale clustering effects. In particular, in the far-IR the power spectrum of the two point correlation function (P_2) shows excess power above Poissonian noise at scales larger than $10'$ (Lagache et al. 2007; Viero et al. 2009; Cooray et al. 2010). In order to reduce ringing, our filter consists of a high-pass filter with a turn on at 6 arcmin convolved with a σ = 1.8 arcmin Gaussian. Only the Lockman-SWIRE field is large enough to be significantly affected because the other fields are not much larger than this scale. Since the benefit of $P(D)$ analyses is at faint flux densities where most of the CFIRB arises, and the shallow Lockman-SWIRE field has little constraining power here, our main scientific results are minimally affected by non-Poissonian clustering effects even if we ignore them. In fact, we find that the differences between fits to simulated data with and without clustering are well within the statistical errors even without filtering.

Analysis of simulated data from Fernandez-Conde et al. (2008), which has linear clustering based on the assumption that infrared galaxies are tracers of dark-matter fluctuations, shows that a high-pass filter is quite effective at removing clustering signal for this data set. We construct two sets of simulated maps: one with clustering, and another using the same catalogue but with clustering removed by randomizing the source positions. We then compare fits and pixel histograms for both maps. Because filtering will affect the $P(D)$ even in the absence of clustering, comparing these to unclustered, unfiltered maps is not useful. The fits recover the input model accurately in both cases, whereas if we do not filter dN/dS is slightly underestimated at low flux densities for the largest maps. Smaller maps show no evidence for bias. A pixel histogram from such a simulation is shown in figure 2. Such filtering is also effective at removing infrared cirrus, although we have not tested this

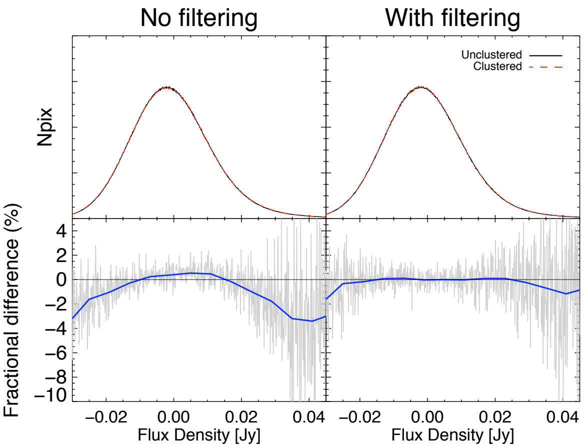

Figure 2. Comparison of the pixel histogram for simulated Lockman-SWIRE data with and without clustering, in the absence of high-pass filtering (left) and with a high-pass filter applied as described in the text (right). The top panel shows the resulting pixel histograms, and the bottom shows the fractional difference between the clustered and unclustered pixel histograms. The thick line in the bottom panel shows the average difference smoothed over a 3 mJy scale. Without filtering, there is a clear trend in the fractional difference, but with filtering the difference is consistent with zero for the bulk of the pixel fluxes. Even without filtering, the effect on the measured counts is smaller than the statistical errors.

explicitly in terms of the recovered fit parameters. However, it is possible that clustering signal on scales between one and six arcminutes could affect our results. This regime is currently not well characterized and thus we could not model it quantitatively in our analysis.

4 MODEL

The best approach for comparing a particular model to SPIRE data using $P(D)$ is to generate pixel histograms as a function of the model parameters and compare directly with our data. However, not all models have smoothly adjustable parameters, and furthermore, if the model is a poor fit to the data this may provide little insight as to at which flux densities the model disagrees with observations. Hence, we have followed P09 and fit simple, non-physical parametric models to our data. These models are defined by the values of the differential number counts dN/dS at a set of fixed flux densities (knots). Observationally we can never do more than place a lower limit on the total number of sources fainter than S, $N(<S)$ because we can never measure all the way down to zero flux density, but dN/dS is better behaved because it only depends on the number of sources in some small range. The actual fit parameters are $\log_{10} dN/dS$ at the knot positions. The differential number counts must become shallower than S^{-2} at low flux densities or else the contribution to the CFIRB diverges. However, this turn-over may lie below the flux densities probed by our data. Therefore, in order to avoid biasing our fits by excluding models which are too steep within the range of our measurements (i.e., would overpredict the CFIRB if integrated down to zero flux density), we assume that the number counts outside the largest and smallest knot are zero; the problem of choosing the limits is discussed below.

A $P(D)$ fit requires that the number counts model is continuous. Therefore, we must choose a method of interpolating between the knots, and for a finite number of knots, the interpretation of our results depends on the interpolation method. We consider two methods of interpolation in this paper: first, as in P09, using power law extrapolation between each knot (these are multiply-broken power-law fits), and second, using a cubic spline in log-log space. The first code supports both methods, and the second only the former. We do not expect the fit parameters (i.e., dN/dS at the knot positions) of these models to be identical, since they have different meaning.

It is important to understand that the results of this paper are model fits. The fit results are not simply dN/dS at the flux densities of the knots, but instead are effectively integral constraints over some region surrounding each knot. Any excursion in the number counts that lies entirely between two knots will affect at least both neighboring knots, and likely others as well. The flux density range that each fit parameter is sensitive to depends on the interpolation scheme, with the spline response more local to the knot. Therefore, simply reading off the values predicted by a theoretical or empirical number counts model at the knot positions and comparing that with our fit parameters is wrong since they are *integral constraints* over a region surrounding the knot. This is also true for more traditional methods (i.e., simple number counts derived from individual galaxy detections) because of the importance of the de-boosting corrections for low signal-to-noise ratio detections. A preferable approach is to first find the best approximation to the differential counts of the theoretical (or empirical) number counts model chosen for comparison using either of our parametric models (for example, by fitting a multiply-broken power law to the dN/dS of the theoretical model giving equal weighting to all fluxes, not just the values at the knots) and comparing the parameters of that approximation to our results.

The highest and lowest knot positions must be chosen with some care because the differential number counts are assumed to be zero outside this range. Our criterion is based on examining the effects of cutting the number counts at a given level on a selection of galaxy evolution models from the literature. We compare the predicted $P(D)$ for each literature model truncated below a specified flux level with the $P(D)$ without truncation, and find that our data are not sensitive to a cut-off of less than 0.1 mJy at 250 μm. A similar analysis shows that truncating the fit above 1 Jy is also undetectable, with similar values for the other passbands. From simulations, we find that we can obtain good constraints if the second knot lies approximately at the 1σ instrumental noise. Because the number counts below our flux limit are unlikely to be well described by a single knot all the way down to 0.1 mJy, the fit value for this point should be treated with care; simulations indicate that this is not a problem for the 2 mJy knot. In order to avoid over-tuning our fits to represent literature models, we adopt approximately logarithmically spaced knots between these extremes. The choice of the number of knots is somewhat arbitrary. Neighboring knots are very strongly correlated, and as the number increases the correlations increase. We have tried to chose the number of knots to be as large as possible while keeping the correlations reasonably small.

5 RESULTS

We fit all three fields simultanously, but each band independently. The uncertainty in the instrumental noise is modeled as a single multiplicative factor having a Gaussian prior with $\sigma = 5\%$. Note that we are making the assumption that the timestream instrumental noise is the same for all three fields as found in Nguyen et al.

Table 3. Differential Number Counts Constraints For a Multiply-Broken Power-Law Model with the FIRAS prior

250 μm		350 μm		500 μm	
Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]
0.1	$< 9.38\,(1\sigma)$	0.05	$< 5.42\,(1\sigma)$	0.05	$< 4.37\,(1\sigma)$
2	$7.39^{+0.06}_{-0.17}$	2	$7.87^{+0.08}_{-0.19}$	2	$7.70^{+0.09}_{-0.21}$
5	$5.83^{+0.15}_{-0.25}$	5	$5.24^{+0.32}_{-0.56}$	5	$5.50^{+0.29}_{-0.52}$
10	$5.978^{+0.027}_{-0.071}$	10	$5.87^{+0.05}_{-0.10}$	10	$5.13^{+0.10}_{-0.21}$
20	$5.13^{+0.02}_{-0.04}$	20	$4.960^{+0.028}_{-0.067}$	20	$4.686^{+0.039}_{-0.094}$
45	$4.041^{+0.014}_{-0.034}$	45	$3.744^{+0.026}_{-0.062}$	45	$2.82^{+0.07}_{-0.15}$
100	$2.591^{+0.027}_{-0.062}$	100	$1.82^{+0.07}_{-0.16}$	100	$1.10^{+0.19}_{-0.35}$
200	$1.42^{+0.06}_{-0.14}$	200	$0.81^{+0.16}_{-0.29}$	200	$-0.08^{+0.54}_{-0.97}$
450	$0.58^{+0.12}_{-0.22}$	750	$-0.69^{+0.16}_{-0.29}$	600	$-1.45^{+0.99}_{-2.01}$
1000	$-0.44^{+0.32}_{-0.62}$				

Marginalized fit parameters for a multiply-broken power-law model from a joint analysis of all three fields including the FIRAS CFIRB prior of Fixsen et al. (1998). Quoted uncertainties are 68.3% confidence intervals, except for the first knot where 1σ upper limits are given. The systematic uncertainties are the same as in table 2.

(2010). In addition to the SPIRE data, we also explore the effects of including the FIRAS CFIRB prior (Fixsen et al. 1998) by integrating $S\,dN/dS$ for our model down to the lowest knot and adding a term to the likelihood that compares that value with the FIRAS measurement and its error. This assumes that the CFIRB is entirely due to discrete sources, and that flux densities outside the range of our model contribute only negligibly. We integrate the Fixsen et al. (1998) spectrum through the SPIRE passbands and adopt the relative errors given in Marsden et al. (2009). The uncertainty in the relative calibration between FIRAS and SPIRE significantly affects the utility of this prior.

The best fit multiply broken power-law fit is compared with the GOODS-N data in figure 3, and the parameters are given in tables 2 and 3, and for the spline interpolation fits in tables 4 and 5. The correlations between adjacent knots are large and negative[4], with typical correlation coefficients of −0.5 to −0.8. The two models are compared with each other in figure 4. The two interpolating models (spline and multiply-broken power-law) produce very similar results. As discussed previously, these are model fits, not independent number counts, and since the parameterizations differ, directly comparing the values at the knot positions is not entirely correct. Nonetheless, the agreement is clear. Also, because the models were fit to the same data their results should not be coadded: they are both presented to demonstrate that similar results are obtained with using independent codes.

Since the agreement is so good, we express no preference of one model versus the other (multiply-broken power-law versus spline). However, we note that the spline model has a narrower flux density window function about each knot, and thus represents the differential number counts of the knot position slightly more accurately locally than the power-law model. For comparison to other number counts models, one can either (i) select the fits with the FIRAS prior, which assumes that the remaining portion of the CFIRB unaccounted for by our priorless fits is encompassed in the range between the upper limit on dN/dS at 0.1 mJy and the 2 mJy knot (at 250 μm) – this method is simpler; or (ii) select the fits without the CFIRB prior, in which case the prior should be applied independently. The latter does not require that the model share the same assumptions about the number counts at low flux densities as our fits.

Table 5. Differential Number Counts Constraints For a Spline Model with the FIRAS prior

250 μm		350 μm		500 μm	
Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]
0.1	$< 8.743\,(1\sigma)$	0.05	$< 7.63\,(1\sigma)$	0.05	$< 6.81\,(1\sigma)$
2	$7.281^{+0.067}_{-0.081}$	2	$7.335^{+0.066}_{-0.077}$	3.4	$6.578^{+0.064}_{-0.083}$
4.3	$6.565^{+0.082}_{-0.094}$	4.3	$6.26^{+0.23}_{-0.26}$	7.3	$5.37^{+0.14}_{-0.22}$
9.1	$5.817^{+0.049}_{-0.052}$	9.1	$5.78^{+0.10}_{-0.11}$	15.5	$4.91^{+0.090}_{-0.093}$
19.5	$5.241^{+0.027}_{-0.028}$	19.5	$4.983^{+0.058}_{-0.063}$	33.2	$3.55^{+0.09}_{-0.10}$
41.8	$4.023^{+0.033}_{-0.031}$	41.8	$3.831^{+0.054}_{-0.055}$	71	$1.82^{+0.15}_{-0.16}$
89.3	$2.786^{+0.045}_{-0.049}$	89.3	$2.13^{+0.10}_{-0.11}$	500	$-0.63^{+1.11}_{-1.80}$
191	$-0.08^{+0.26}_{-0.32}$	191	$0.95^{+0.16}_{-0.18}$		
408	$0.957^{+0.065}_{-0.075}$	1000	$-2.12^{+1.06}_{-1.83}$		
1000	$0.186^{+0.084}_{-0.088}$				

Marginalized fit parameters for a spline interpolation model from a joint analysis of all three fields including the FIRAS CFIRB prior of Fixsen et al. (1998). The systematic uncertainties are the same as in table 4.

Our fits are compared with other measurements in figures 5 and 6. Ignoring the lowest knot (where only an upper limit is available), our fits predict a CFIRB flux density of 0.54 ± 0.08, 0.39 ± 0.06, and 0.16 ± 0.03 MJy sr^{-1} from all sources down to 2 mJy in the three bands; the dominant error in all cases is due to the 15 per cent calibration uncertainty of SPIRE. The contribution from each flux range is shown in figure 7. The CFIRB from Fixsen et al. (1998) integrated over the SPIRE bands is 0.85 ± 0.19, 0.65 ± 0.19, and 0.39 ± 0.10 MJy sr^{-1}, respectively, so our fits therefore account for 64 ± 16, 60 ± 20 and 43 ± 12 percent in the SPIRE 250, 350, and 500 μm bands, respectively. We expect to resolve a smaller fraction of the CFIRB at longer wavelengths because the size of the SPIRE beam is proportional to wavelength, and hence the 500 μm band is more confused. Here the errors are dominated by the uncertainty in the FIRAS measurement. We find marginalized values for the instrumental noise that are 1.02, 1.1, and 1.01 times the values given in table 1 at 250, 350, and 500 μm, respectively, giving a χ^2 of 4.2 for 3 degrees of freedom. Hence, our instrumental noise values are consistent with the Nguyen et al. (2010) prior.

5.1 Systematic Effects

Our basic tool for estimating the importance of a particular systematic is to compute the $\Delta \log \mathcal{L}$ between the $P(D)$ with and without the effect for maps the same size and depth as our data. We use the P09 best fit model as a basis for this computation. Recall that a $\Delta \log \mathcal{L}$ of 0.5 corresponds roughly to a 1σ statistical error.

Because different parts of the map are sampled by different bolometers, and the beam shape varies across the bolometer array, the effective beam will vary over the map. We evaluated this effect by choosing 200 random pixels in our maps and computing the fractional contribution of each bolometer to each pixel. We

---

[4] Covariance matrices are provided at http://www.hermes.sussex.ac.uk/

Table 2. Differential Number Count Constraints For a Multiply-Broken Power-Law Model

250 μm		350 μm		500 μm	
Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]
0.1	$< 11.08\,(1\sigma)$	0.05	$< 11.20\,(1\sigma)$	0.05	$< 11.28\,(1\sigma)$
2	$7.05^{+0.33}_{-0.57} \pm 0.19$	2	$6.94^{+0.13}_{-0.27} \pm 0.11$	2	$6.82^{+0.11}_{-0.25} \pm 0.12$
5	$6.25^{+0.04}_{-0.13} \pm 0.05$	5	$6.08^{+0.13}_{-0.25} \pm 0.08$	5	$5.65^{+0.19}_{-0.38} \pm 0.09$
10	$5.919^{+0.028}_{-0.063} \pm 0.011$	10	$5.78^{+0.05}_{-0.11} \pm 0.04$	10	$5.39^{+0.09}_{-0.18} \pm 0.03$
20	$5.139^{+0.013}_{-0.035} \pm 0.025$	20	$4.976^{+0.026}_{-0.061} \pm 0.024$	20	$4.57^{+0.05}_{-0.12} \pm 0.03$
45	$4.038^{+0.015}_{-0.033} \pm 0.031$	45	$3.742^{+0.026}_{-0.061} \pm 0.051$	45	$2.91^{+0.07}_{-0.16} \pm 0.04$
100	$2.596^{+0.025}_{-0.058} \pm 0.044$	100	$1.80^{+0.07}_{-0.16} \pm 0.10$	100	$0.96^{+0.22}_{-0.38} \pm 0.06$
200	$1.42^{+0.05}_{-0.14} \pm 0.08$	200	$0.87^{+0.14}_{-0.28} \pm 0.08$	200	$0.00^{+0.51}_{-0.92} \pm 0.07$
450	$0.57^{+0.13}_{-0.24} \pm 0.26$	750	$-0.65^{+0.39}_{-0.78} \pm 0.30$	600	$-1.43^{+0.96}_{-2.09} \pm 0.29$
1000	$-0.45^{+0.31}_{-0.60} \pm 0.20$				

Marginalized fit parameters for a multiply-broken power-law model from a joint analysis of all three fields without using the FIRAS CFIRB prior. The quoted uncertainties are the 68.3% confidence intervals for the statistical error followed by the estimated systematic uncertainty, except for the first knot where the 1σ upper limit is given.

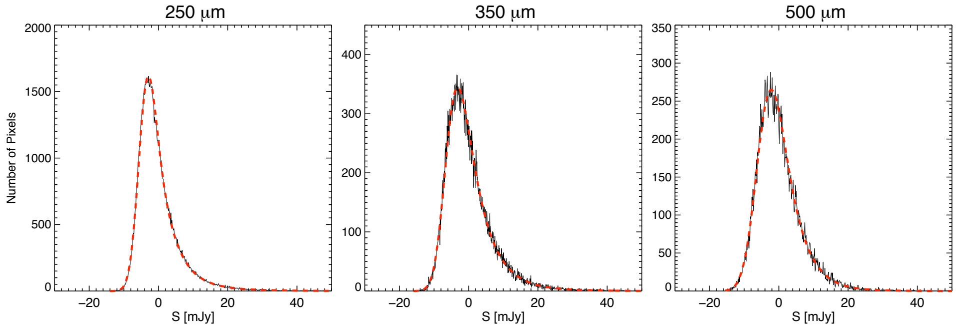

Figure 3. Comparison of the GOODS-N pixel histograms (solid lines) to the best fit model to all three fields (dashed lines) using the multiply-broken power law fit and not including the FIRAS prior.

Table 4. Differential Number Counts Constraints For a Spline Model

250 μm		350 μm		500 μm	
Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]	Knot [mJy]	$\log_{10} dN/dS$ [deg^{-2}Jy^{-1}]
0.1	$< 10.29\,(1\sigma)$	0.05	$< 11.43\,(1\sigma)$	0.05	$< 10.91\,(1\sigma)$
2	$7.26^{+0.10}_{-0.17} \pm 0.19$	2	$7.18^{+0.15}_{-0.28} \pm 0.11$	3.4	$6.36^{+0.13}_{-0.18} \pm 0.10$
4.3	$6.54^{+0.10}_{-0.12} \pm 0.06$	4.3	$6.24^{+0.21}_{-0.21} \pm 0.09$	7.3	$5.31^{+0.19}_{-0.21} \pm 0.05$
9.1	$5.837^{+0.059}_{-0.056} \pm 0.013$	9.1	$5.831^{+0.081}_{-0.090} \pm 0.042$	15.5	$4.961^{+0.074}_{-0.085} \pm 0.029$
19.5	$5.230^{+0.029}_{-0.032} \pm 0.024$	19.5	$4.959^{+0.053}_{-0.061} \pm 0.025$	33.2	$3.511^{+0.094}_{-0.095} \pm 0.034$
41.8	$4.036^{+0.032}_{-0.036} \pm 0.030$	41.8	$3.849^{+0.051}_{-0.050} \pm 0.048$	71	$1.85^{+0.14}_{-0.16} \pm 0.063$
89.3	$2.802^{+0.045}_{-0.050} \pm 0.040$	89.3	$2.11^{+0.10}_{-0.11} \pm 0.076$	500	$-0.64^{+1.25}_{-1.80} \pm 0.028$
191	$0.20^{+0.24}_{-0.34} \pm 0.080$	191	$0.96^{+0.16}_{-0.19} \pm 0.075$		
408	$1.002^{+0.064}_{-0.068} \pm 0.26$	1000	$-2.34^{+1.82}_{-1.92} \pm 0.31$		
1000	$0.18^{+0.09}_{-0.10} \pm 0.20$				

Marginalized fit parameters for a spline model from a joint analysis of all three fields. Quoted uncertainties are as in table 2.

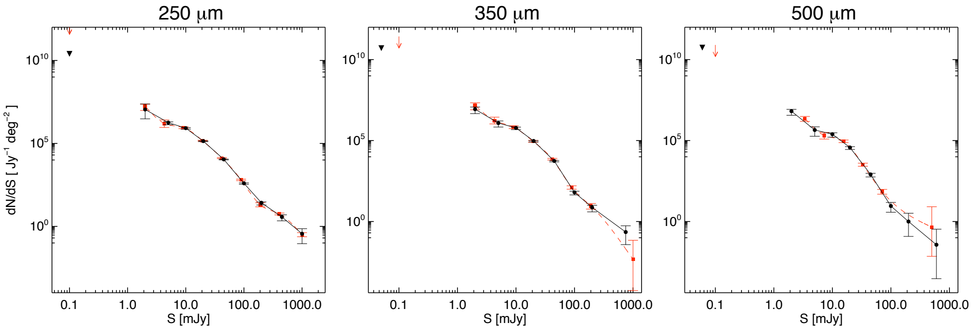

Figure 4. Comparison of the multiply-broken power-law (solid lines) and spline (dashed lines) $P(D)$ fits for the differential number counts to the three SDP fields simultaneously, without the FIRAS prior; 1σ upper limits are shown as arrows. For this comparison, only statistical errors are shown.

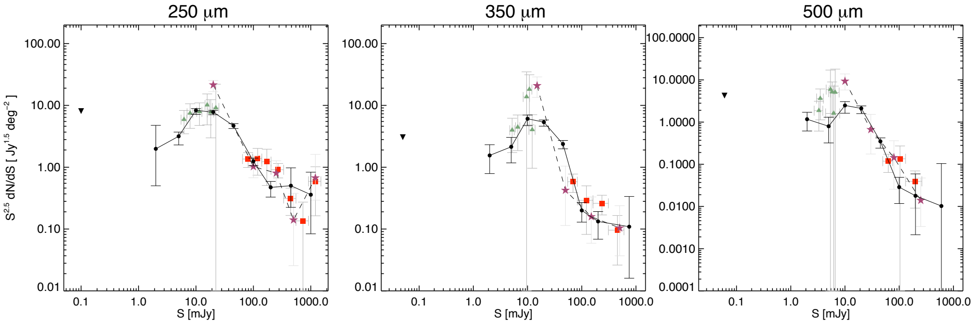

Figure 5. Comparison of Euclidean-normalized multiply-broken power law differential number counts fits (solid lines and circles) with previous balloon-based measurements from BLAST, not using the FIRAS prior. The BLAST $P(D)$ analysis (P09) is shown as dashed lines and stars, the stacking analysis of Béthermin et al. (2010b) as triangles and the source extraction analysis from the same reference as squares. Here the combined statistical and systematic errors are shown.

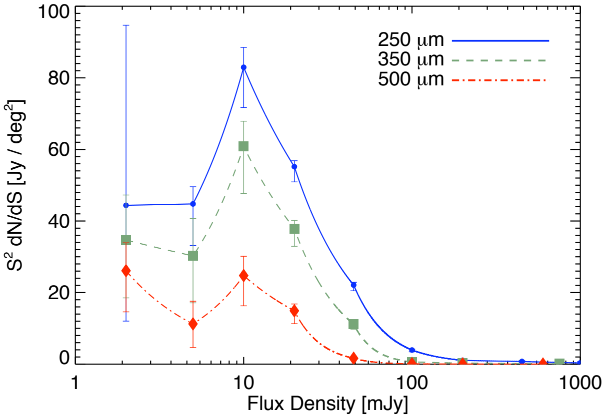

Figure 7. The contribution to the cosmic far-infrared background from each flux range for the multiply-broken power-law model versus the log of the flux density. The results for the spline interpolation model are almost identical. The integral of the curve over the log flux density is proportional to the total flux contribution in each band.

then built per-bolometer maps from our Neptune observations, and combined these to find the effective beam at each of these locations. The beam varies across the map in a complicated fashion because even in our deepest map each pixel only samples a limited subset of bolometers. This produces significant variation in the $P(D)$ with position. In general, the $P(D)$ computed for the bolometer-averaged beam does not have to be the same as the $P(D)$ computed for each bolometer and then averaged across the array (which is the $P(D)$ of the entire map). To evaluate the importance of this variation, we compare the $P(D)$ for the average beam to the $P(D)$ computed for all 200 pixels and then averaged. The $\Delta \log \mathcal{L}$ of this comparison is < 0.01, so for our analysis this is negligible.

Although we masked each map to exclude the low-coverage edges, the HerMES SDP observations were not dithered between repeats so there are significant variations in the number of measurements per pixel even within the high-coverage regions ($\sim$ 20%). This will introduce slight non-Gaussian tails to the instrument noise distribution. We simulated this effect in two ways. First, we generated random realizations of the instrument noise, including the uneven coverage, and compared the $P(D)$ using the resulting noise to the $P(D)$ assuming the noise is purely described by the average σ, and found negligible $\Delta \log \mathcal{L}$. Second, our end-to-end simulations (also including $1/f$ noise) implicitly include uneven coverage effects, and we found no bias in the recovered parameters. Future HerMES observations will include dithering, which will also have the benefit of improving the homogeneity of the beam.

Nguyen et al. (2010) explore the noise characteristics of the SDP maps by carrying out ‘Jack-Knife’ tests on the data. Their findings are generally consistent with the expected noise properties, but it is difficult to rule out some additional level of non-Gaussian

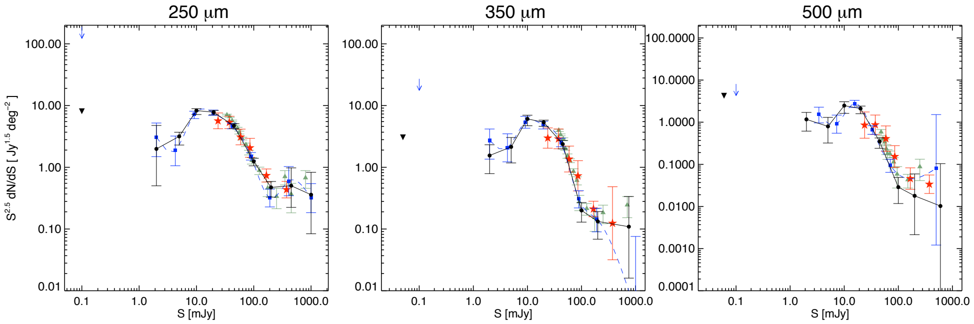

Figure 6. Comparison of Euclidean normalized SPIRE $P(D)$ differential number counts (solid lines/circles and dashed lines/squares for the multiply-broken power law and spline models, respectively, not using the FIRAS prior in both cases) with other SPIRE number counts: first, an analysis of the same data using source extraction techniques (Oliver et al. 2010, red stars), and second the H-ATLAS source extraction on an independent field (Clements et al. 2010, triangles). The errors are the combined statistical and systematic errors.

noise beyond the $1/f$ behavior we have simulated. Directly computing the effects of the Jack-Knife noise histograms on our model shows that any additional non-Gaussianity has negligible effects for the SDP data, down to our lowest constrained knot (2 mJy). This may not be true for future observations where the larger field sizes will reduce the statistical error considerably.

To test the sensitivity to the beam model, we use an alternative set of Neptune observations with a much smaller number of repeats and coarser sampling. Furthermore, the pointing of these observations is not corrected for the small offset between the *Herschel* and SPIRE clocks, and hence they suffer from pointing drift relative to the maps of the science fields[5]. The pipeline nominally corrects for this offset. However, to be conservative, we allow for the possiblity that the pointing drift might not affect the beam maps in the same way as the science maps: we repeat the fits using the alternative beam, and use the difference in the results as an estimate of the beam systematic. This is the dominant identified systematic effect, with $\Delta\mathcal{L} \sim 0.3$. We take the differences between the recovered parameters between the two beams as the systematic error on each knot as given in tables 2 and 4. As our understanding of the SPIRE beams improves, it should be possible to decrease this error.

To further explore issues of pointing drift, we have constructed a simple drift model for the GOODS-N field using Jack-Knife comparisons. We then generate simulated maps with and without applying this model. Because the effect of the model is largely to twist successive observations relative to each other, this has very little effect on the $P(D)$, amounting to $\sim 0.1\sigma$ relative to the statistical errors.

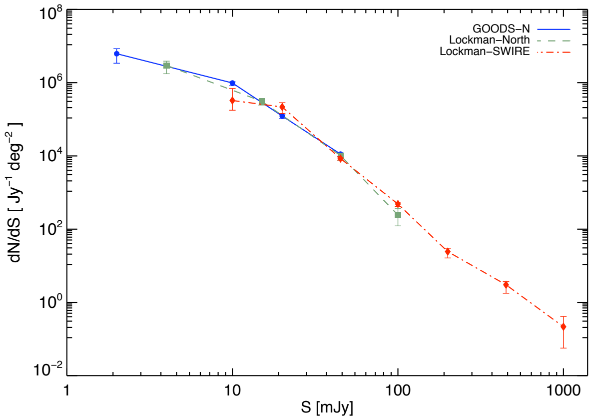

Figure 8. Comparison of 250 μm differential number counts derived from the three fields for the multiply-broken power-law model. Only statistical errors are shown and the FIRAS prior is not included.

While our results should represent the number counts in our fields quite well, sample variance means that they may not perfectly represent the number counts we would obtain with an infinitely large field. If we make the strong assumptions that the SPIRE clustering properties measured in Cooray et al. (2010) apply equally at all flux densities (and, in particular, to depths 10 times greater than they were measured), that the redshift distribution of our sources is independent of measured brightness, and that the source population peaks at $z = 1.5$, then a simple analytic computation suggests that sample variance could be a 20 per cent effect on the total number counts in GOODS-N. More empirically, if we split the Lockman-SWIRE field into sub-regions there is evidence for variation in the pixel histogram from sub-field to sub-field. However, after applying our high-pass filter the variations are no longer statistically significant; that is, the difference between sub-fields lies in the mean rather than the shape of the histograms. Alternatively, mean subtracting each sub-field produces the same effect. This suggests that the fact that our measurements are not sensitive to the mean map values (and therefore are mean-subtracted) may provide some protection against sample variance effects. Since estimating the size of this effect is highly model dependent, we have not attempted to include this in our error budget. We do, however, fit the three fields independently at 250 μm and compare the results, although the different depths complicate this somewhat. For simplicity, we do not marginalize over the instrument noise in this fit, since the only purpose is to compare the three fields. This is shown in figure 8. Within the uncertainties, the fits are consistent. The post-SDP HerMES observations will allow sample variance effects to be better quantified.

We do not include the effects of the SPIRE calibration uncertainty ($\sim$ 15% across all bands) in our error budget, except when using the CFIRB prior or computing the fraction of the FIRAS mea-

---

[5] The SPIRE clock speed differs by a very small amount from the Herschel clock, resulting in a cumulative pointing drift with time in SPIRE maps. The magnitude of the effect is 0.7 arcseconds per hour, with rephasing occuring when "PCAL" internal calibrations are made.

surement accounted for by our model. However, any updates to the SPIRE calibration are easily incorporated into our results without refitting: if the flux scale is multiplied by a factor α the knot positions $K_i \mapsto \alpha K_i$ and the knot values decrease by $-\log_{10}\alpha$.

6 DISCUSSION

In general, our results agree well with those of P09, except for the faintest fluxes fully constrained by their analysis. For example, at 250 μm they find $\log_{10} dN/dS = 5.58^{+0.07}_{-0.11}$ at 20 mJy, while our result is $5.139^{+0.012}_{-0.033} \pm 0.025$. However, as discussed earlier, they did not marginalize over the instrumental noise for their deepest field, so their errors may be somewhat underestimated here. There is also some evidence from simulated data that the small number of knots and knot placement right at the break in the number counts may have biased this knot in the P09 analysis. We find good agreement with the stacking analysis of the BLAST data, but see some mild disagreement at higher fluxes for direct counts of the same data (Béthermin et al. 2010b) as shown in figure 5.

The deepest number counts available at these wavelengths are the result of a semi-traditional source extraction method on the same HerMES data set (Oliver et al. 2010). These are compared in figure 6. Where there is overlap, the agreement is good. A similar analysis was carried out using H-ATLAS SDP data by Clements et al. (2010); this is also shown. Unlike the HerMES source extraction and $P(D)$ analysis, the H-ATLAS counts are 250 μm-selected at all wavelengths, and hence may not entirely probe the same point source population. Nonetheless, again the agreement with the HerMES results is good.

A few features are worth noting. First, we clearly detect a break in the number counts around at 20 mJy in all bands at high significance. However, the SPIRE data alone do not detect the change in slope in dN/dS necessary to keep the CFIRB finite, as the differential counts continue to rise to the lowest limit of our analysis more steeply than S^{-2}. When the FIRAS prior is added, a break is present, but this mostly affects the lowest flux knot, for which we can only provide an upper limit (the effects on the other knots are mostly due to the strong correlations between knots; the FIRAS prior changes the structure of the correlations significantly at low flux densities). Second, there is possible weak ($\sim 1\sigma$) evidence for a 'bump' in the differential counts around 400 mJy at 250 μm. There is no evidence at 350 and 500 μm around this flux density, but the error bars are large. However, this bump is present in the independent H-ATLAS field (Clements et al. 2010) at 250 μm. The cause is unclear; lensing is an intriguing possibility, but we would expect the signature of lensing to be larger at 500 μm due to stronger negative K-correction effects (e.g. Negrello et al. 2007).

We can compare our model fits to the confusion noise estimates of Nguyen et al. (2010) measured using a different technique. The often-used criterion of one source per every n beams is difficult to use, since its translation into the effects on observations depends strongly on the underlying model (Takeuchi & Ishii 2004). Hence, we adopt the square root of the variance of the source contribution to the pixel distribution ($\sigma_{\rm conf}$) as our measure. We find values of $6.5/6.4/6.1 \pm 0.2$ mJy in the three bands, slightly higher than the Nguyen et al. (2010) values, but by less than 2σ.

Our fits are compared with a selection of literature models in figure 9. No currently available model entirely fits our counts, especially when all three bands are considered. There is a general discrepancy in the galaxy number count models in that the theoretical models generically overpredict the number of bright galaxies (in the several × 10 to several × 100 mJy range, limited at the upper end by uncertainties in the $P(D)$ number counts) compared to the number counts from the $P(D)$ analysis. The best match overall across all three SPIRE bands is given by the model of Valiante et al. (2009).

We interpret the discrepancy in the context of the theoretical models of Lagache et al. (2003) and Fernandez-Conde et al. (2008) (figure 9). In Figure 10 the redshifts and FIR luminosities of galaxies are plotted versus their observed flux densities for the Fernandez-Conde et al. (2008) simulations. The transitions from luminous to ultra-luminous infra-red galaxies (LIRGs to ULIRGs, at $10^{12} L_\odot$) with increasing observed flux densities occurs at approximately 12, 6, and 3 mJy, in the 250, 350, and 500 μm SPIRE bands, respectively. Thus, the discrepancy at the bright end likely results from the presence of too many ULIRGs in the theoretical models. It should be noted that the intrinsic luminosities of the underlying galaxy population that contribute to any given bin in observed flux density depend on the redshift distributions and spectral energy distributions; however, the very brightest galaxies are likely either intrisically extremely luminous (ULIRGs or brighter), low redshift, or strongly lensed. There is a large dispersion in redshifts represented by galaxies in each observed flux density bin, with means in the range $z = 1 - 2$, with the average flux densities only mildly inversely correlated with redshifts (due to the negative K-correction).

In all three bands at and below 2 mJy, the $P(D)$-derived number counts are consistent with the theoretical galaxy number count models. This is not surprising because: (i) the theoretical galaxy number count models are constrained not to overpredict the CFIRB, which arises, in large part, from numerous faint galaxies; and (ii) the upper limits of the lowest flux density knot in each band lies well above the theoretical number counts models.

Another more subtle feature also seems apparent in the measured counts. The results from both fitting methods – lending some confidence to their credence – have depressions at the third-lowest flux density knots with respect to the theoretical number count models at low-to-moderate significance (depending on the theoretical model), which are exclusively concave down in this range (a few mJy to a few times 10 mJy). The stacking analysis of Béthermin et al. (2010a) is not deep enough at 250 or 350 μm to verify this feature, although the turnover in $S^{2.5} dN/dS$ from the peak at approximately 10 mJy downward is clear at 250 μm. At 500 μm, the stacking analysis does not display the depression. This flux density range is approximately at the confusion limit (where the flux density is equal to the confusion noise, $\sigma_{\rm conf} = 6$ mJy) and multiple galaxies contribute to the flux density in each beam. Thus, referring to the Fernandez-Conde et al. (2008) models (figure 10), this flux density range corresponds to the transition from ULIRGS to LIRGs, suggesting that LIRG number counts may also be overrepresented by the theoretical galaxy number counts models.

7 CONCLUSIONS

We have measured the differential galaxy number counts from *Herschel*-SPIRE Science Demonstration Phase HerMES observations at 250, 350, and 500 μm using $P(D)$ techniques and two simple parametric models. The number counts were measured down to 2 mJy, approximately a factor of 3 below the 1 σ confusion noise. We find that 64 ± 14 per cent of the measured CFIRB is accounted for by point sources at 250 μm , falling to 43 ± 12 per cent at 500 μm. The errors on the fraction of the CFIRB accounted for by these

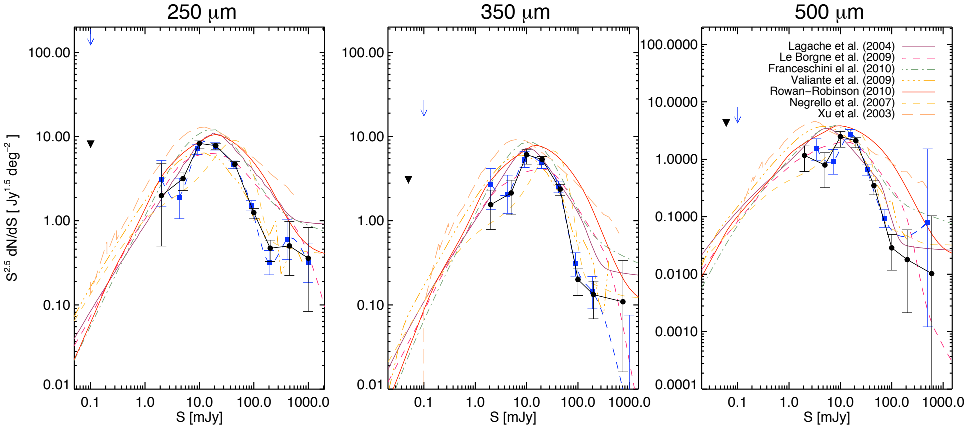

Figure 9. Comparison of our Euclidean normalized differential number counts fits (as in figure 6, and not using the FIRAS prior) to a selection of models from the literature. The error bars are the combined statistical and systematic errors.

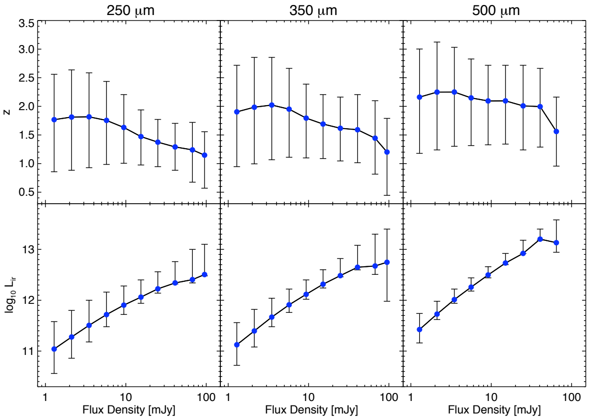

Figure 10. The intrinsic redshifts (upper panels) and infrared luminosities (lower panels) of galaxies as a function of the observed flux densities in the SPIRE 250, 350, and 500 μm bands (left, middle, and right panels, respectively) for the Lagache et al. models (Lagache et al. (2003); Fernandez-Conde et al. (2008)). The error bars give the interquartile range for each bin.

sources are now dominated by those of the FIRAS measurement. However, because of the remaining fraction not accounted for by our fits, this is still not a competitive method for measuring the total CFIRB. We find clear evidence of breaks in the slope of the differential number counts at approximately 10–20 mJy in all bands, which have been hinted at by previous analyses.

Where they overlap, our fits agree well with other *Herschel* results. Comparing with a selection of literature models, however, we find that no model entirely reproduces our observed number counts. As found by Oliver et al. (2010) and Clements et al. (2010), most published models significantly over-predict the number of bright sources at these wavelengths and have shallower slopes. We find somewhat better agreement at fainter fluxes, at or below the break, but the agreement is still not perfect.

Our main systematic uncertainties arise from our understanding of the SPIRE beams. We find that a high-pass filter is effective in removing the signature of clustering from our counts, but in the future it may be preferable to attempt to directly marginalize over clustering using simple models.

These observations represent only ~ 60 hours of the 900 hours of observations that HerMES will ultimately obtain (although not all of these are with SPIRE). The final dataset will cover a wide range of depths and areas. This will significantly increase our ability to constrain dN/dS. Having a number of well-seperated deep fields will also allow a direct measurement of sample variance.

Acknowledgments. The authors would like to thank Guillaume Patanchon and Phil Maloney for many useful discussions. J. Glenn and A. Conley acknowledge support from NASA Herschel GTO grant 1394366, sponsored by the Jet Propulsion Laboratory. SPIRE has been developed by a consortium of institutes led by Cardiff Univ. (UK) and including Univ. Lethbridge (Canada); NAOC (China); CEA, LAM (France); IFSI, Univ. Padua (italy); IAC (Spain); Stockholm Observatory (Sweden); Imperial College London, RAL, UCL-MSSL, UKATC, Univ. Sussex (UK); Caltech, JPL, NHSC, Univ. Colorado (USA). This developement has been supported by national funding agencies: CSA (Canada); NAOC (China); CEA, CNES, CNRS (France); ASI (Italy); MCINN (Spain); SNSB (Sweden); STFC (UK); and NASA (USA). The data presented in this paper will be released through the *Herschel* Database in Marseille, HeDaM (http://hedam.oamp.fr/HerMES).

REFERENCES

Barcons, X. 1992, ApJ, 396, 460
Barcons, X. et al. 1994, MNRAS, 268, 833
Berta, S. et al. 2010, A&A, 518, 30
Béthermin, M. et al. 2010a, A&A, 512, 78
Béthermin, M. et al. 2010b, A&A, 516, 43
Blain, A. et al. 2004, ApJ 611, 725
Chapin, E. L. et al. 2010, ArXiv e-print 1003.2647
Clements, D. L. et al. 2010, A&A, 518, 8
Cooray, A. et al. 2010, A&A, 518, 22
Condon, J. J. 1974, ApJ, 188, 279
Devlin, M. J. et al. 2009, Nature, 458, 737
Dole, H., et al. 2006, A&A, 451, 417
Dowell, C. D. et al. 2010, Space Telescopes and Instrumentation

2010: Optical, Infrared, and Millimeter Wave, Proc. SPIE 7731, in press
Dunkley, J. et al. 2005, MNRAS, 356, 925
Fernandez-Conde, N. et al. 2008, A&A, 481, 885
Fixsen, D.J., Dwek, E., Mather, J.C., Bennett, C.L., & Shafer, R.A. 1998, ApJ, 508, 123
Franceschini, A. et al. 2010, A&A in press
Gelman, A. and Rubin, D. B. 1992, Statistical Science, 7, 457
Griffin, M. et al. 2010, A&A, 518, 3
Guiderdoni, B., Bouchet, F.R., Puget, J.-L., Lagache, G., & Hivon, E 1997, Nature, 390, 257
Hughes, D. H. et al. 1998, Nature, 394, 241
Lagache, G., Dole, H, & Puget, J.-L. 2003, MNRAS, 338, 555
Lagache, G. et al. 2007, ApJ, 665, 89
Laurent, G.T. et al. 2005, ApJ, 623, 742
Le Borgne, D., et al. 2009, A&A, 504, 727
Lewis, A. and Bridle, S. 2002, PhRVD, 66, 103511
MacKay, D. J. C., 2002, Information Theory, Inference, and Learning Algorithms, Cambridge University Press
Maloney, P.R., et al. 2005, ApJ, 635, 1044
Marsden, G., et al. 2009, ApJ, 707, 1729
Negrello, M. et al. 2007, MNRAS, 377, 1557
Nguyen, H. T. et al. 2010, A&A, 518, 5
Oliver, S. J. et al. 1997, MNRAS, 289, 471
Oliver, S. J. et al. 2010, A&A, 518, 21
Patanchon, G. et al. 2009, ApJ, 707, 1750
Pilbratt, G. et al. 2010, A&A, 518, 1
Puget, J.-L. et al. 1996, A&A, 308, L5
Rowan-Robinson, M. 2001, ApJ, 549, 745
Rowan-Robinson, M. 2009, MNRAS, 394, 117
Saunders, W. 1990, MNRAS, 242, 318
Scheuer, P. A. G. 1957, Proc. Camb. Philos. Soc., 53, 764
Scheuer, P. A. G. 1974, MNRAS, 166, 329
Scott, S. E., et al. 2002, MNRAS, 331, 817
Scott, K. S., et al. 2010, MNRAS, 405, 2260
Smail, I., Ivison, R.J., Blain, A.W., & Kneib, J.-P. 2002, MNRAS, 331, 495
Somerville, R. S. et al. 2004, ApJ, 600, 171
Swinyard, B. et al. 2010, A&A, 518, 4
Takeuchi, T. T., & Ishii, T.T. 2004, ApJ, 604, 40
Valiante, E. et al. 2010, ApJ, 701, 1814
Viero, M. P. et al. 2009, ApJ, 707, 1766
Weiß, A. et al. 2009, ApJ, 707, 1201
Xu, C. K. et al. 2003, ApJ, 587, 90
Zemcov, M. et al. 2010, submitted to ApJ