Detection Theory: the Short Form

Note: the following statements are generic—they may or may not apply to the specific source detection algorithm of your choice. They are simply meant to help build your intuition.

The Ingredient(s)

- A N-dimensional event list or binned "image." (And an exposure map, knowledge of the PSF, *etc.*, if appropriate.)

The Detection "Tool"

- Some function that is localized (*i.e.*, non-zero only over some characteristic scale) within at least some subset of the dimensions.

The Hypotheses

- M_0: the data in a given pixel are (Poisson-)sampled from the background.
- M_1: the data are a sum of (Poisson) samples from the background and an astronomical source.

The Five-Fold Path

For a given source detection algorithm, an analyst might follow this five-fold path to source detection Nirvana.[1]

- Select an appropriate function scale, σ.
  (If one is attempting to detect a point source, this would be some encircled-energy radius of the PSF.)
- Estimate the background amplitude, $\hat{B}$.
- Determine the value of a selected model comparison test statistic, T_o.
- Determine the significance, α:

$$\alpha \;=\; \int_{T_o}^{\infty} dT\, p(T|f[\hat{B}])$$

- Compare α to a pre-determined threshold significance value α_o.

If $\alpha < \alpha_o$, the pixel is associated with a source!

---

[1] In the following, we assume that while this analyst will treat an entire multi-pixel image, she only really cares about one pixel, (i, j); this assumption allows us to ignore messy subscripts that would otherwise appear everywhere!

Potholes on the Five-Fold Path

However, the analyst may need an SUV to drive along the five-fold path, because, for instance:

- the determination of $\hat{B}$ from raw data that consists of both source and background counts is a *nasty* statistical problem that has no known best solution;
- there is no model comparison test statistic T that has been proven to be "most powerful";
- the PSDs from which observed values T_o are sampled generally cannot be represented with analytic functions, except perhaps in the high-counts limit–hence simulations are necessary to determine α; and
- because most astronomical sources will be associated with multiple pixels, a distinction must be made between, *e.g.*, the number of false *pixels* associated with significance α_o, and the number of false *sources*;

However, the deepest potholes are associated with instrumental issues such as hot pixels, vignetting, varying non-Gaussian PSFs, dither, *etc.*—items that will be dealt with later in this class.

Classic Detection: CELLDETECT

The Function(s)

- Two box functions with unit amplitude, co-aligned and centroided at pixel (i, j). These boxes are of size $d \times d$ (smaller) and $b \times b$ (larger), and the number of counts within each box are D_d and D_b, respectively.

The Determination of $\hat{B}$

- This is done by assuming (a) the truth of the alternative model, and (b) that the source is point-like:

$$\begin{aligned} D_d &= \alpha \hat{S} + \hat{B} \\ D_b &= \beta \hat{S} + \left(\frac{b}{d}\right)^2 \hat{B} \end{aligned}$$

  where α and β are the integrals of the PSF within each box, respectively.

The Model Comparison Test Statistic

- $T_o = \hat{S}/\sigma_{\hat{S}}$, *i.e.*, the "SNR."

Associating a Pixel with a Source

- If SNR $\geq$ SNR$_{\rm thr}$, accept. For *Chandra*, one determines SNR$_{\rm thr}$ from a table, as it is a function of exposure time (*i.e.*, expected background amplitude) and off-axis angle (*i.e.*, most appropriate scales d and b); it corresponds to a number of false *sources*.

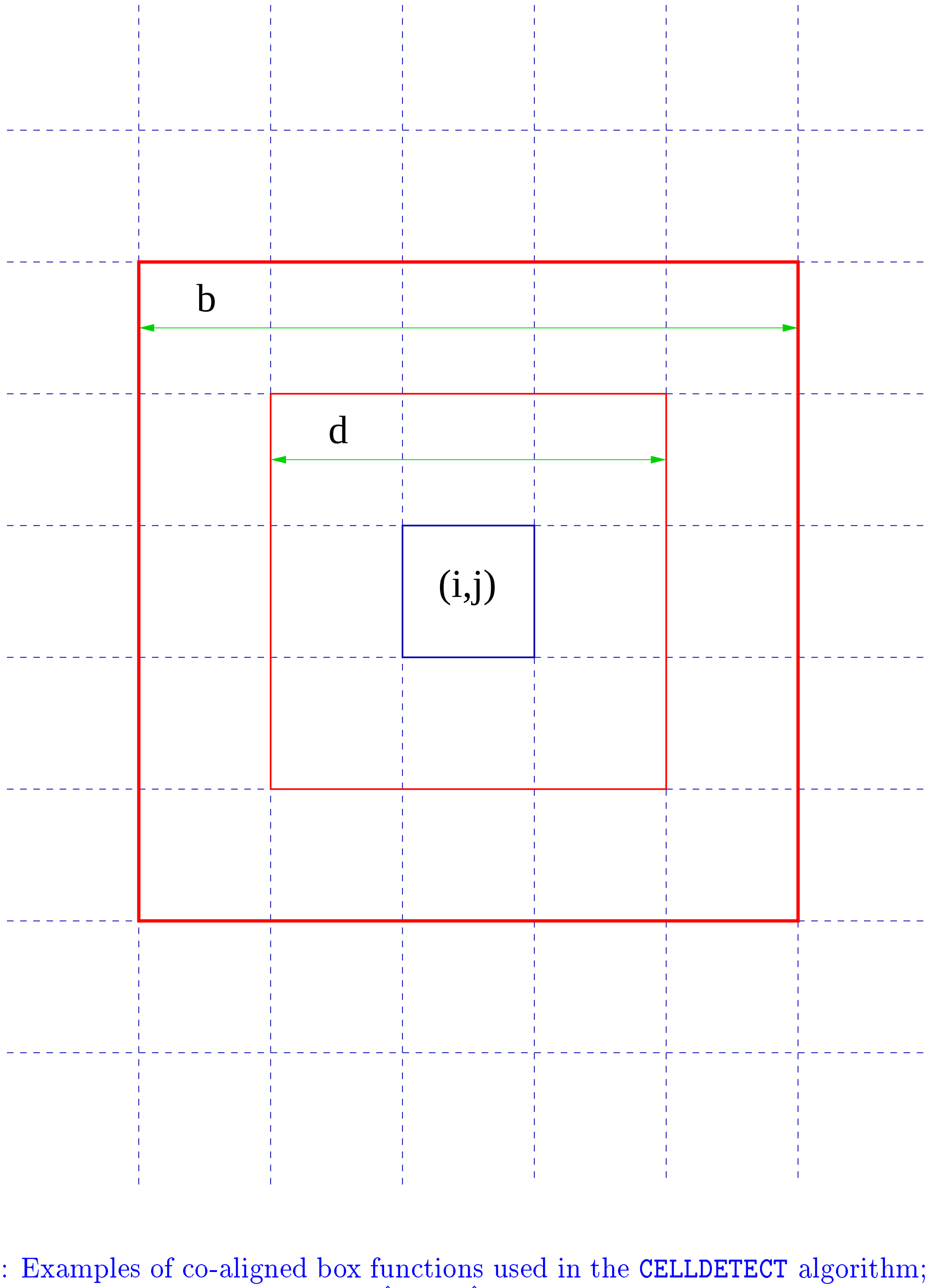

Figure 1: Examples of co-aligned box functions used in the `CELLDETECT` algorithm; the data within the boxes are used to estimate $\hat{S}$ and $\hat{B}$ as described on the previous transparency.

New Detection: WAVDETECT

The Function

- The Marr, or "Mexican Hat" wavelet, $W(\sigma)$, centroided at pixel (i, j). This function is non-zero within a circle of radius $\approx 5\sigma$.

The Determination of $\hat{B}$

- Done by determining the average number of counts per pixel in the wavelet negative annulus, while using it as a weighting function.
- Done iteratively, with source counts removed from the field until the background estimate stabilizes.

The Model Comparison Test Statistic

- $T_o = C_o = \Sigma_{i'}\Sigma_{j'}W_{i-i',j-j'}D_{i',j'}$.

Associating a Pixel with a Source

- If

$$\alpha \;=\; \int_{C_o}^{\infty} dC p(C|2\pi\sigma^2 B) \;<\; \alpha_o\,,$$

accept. A usual choice for α_o is P^{-1}, where P is the number of pixels examined in the image; it thus corresponds to a number of false *pixels*.

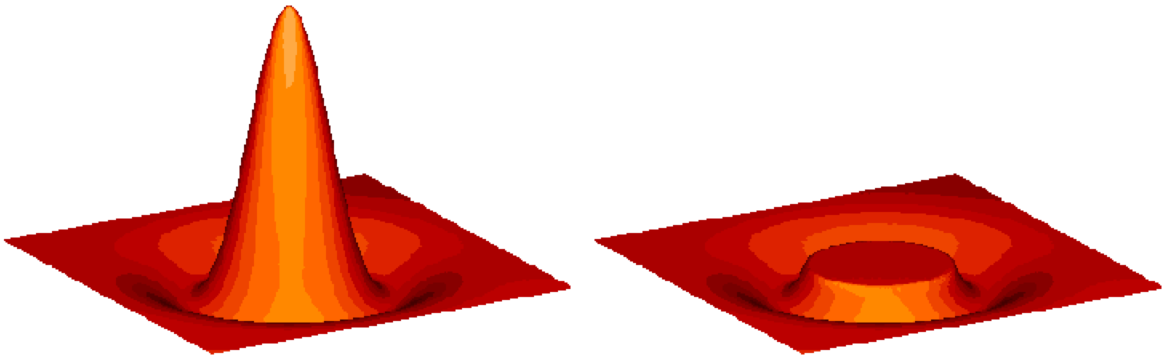

Figure 2: *Left:* The Marr, or "Mexican Hat" wavelet function, used in the **WAVDETECT** algorithm. *Right:* The negative annulus of the wavelet function, used by **WAVDETECT** in background estimation. See Freeman *et al.* (2002, ApJS 138, 185) for more details.