Table 2
Detection Features*

Feature	Value
(Intercept)	1
mag	e^i_m
depth	e^i_d
dist	Δ_{ik}
dist0	$\mathcal{N}(\Delta_{ik}, 0, 5)$
dist35	$\mathcal{N}(\Delta_{ik}, 35, 20)$
dist40	$\mathcal{N}(\Delta_{ik}, 40, 20)$
dist12520	$\mathcal{N}(\Delta_{ik}, 125, 20)$
dist12540	$\mathcal{N}(\Delta_{ik}, 125, 40)$
dist 145	$\mathcal{N}(\Delta_{ik}, 145, 10)$
dist170	$\mathcal{N}(\Delta_{ik}, 170, 20)$
dist175	$\mathcal{N}(\Delta_{ik}, 175, 30)$
mag6	$\mathcal{N}(e^i_m, 6, 5.5)$
mag68	$\mathcal{N}(e^i_m, 6, 8)$
md	$(7 - e^i_m) \cdot \Delta_{ik}$

*List of features used for computing the probability of detecting an arrival from an event i with magnitude e^i_m, depth e^i_d, and distance Δ_{ik} from station k. Here $\mathcal{N}(x, \mu, \sigma)$ is the standard Gaussian density with mean μ and standard deviation σ measured at x.

in the event magnitude would result in the odds of detection increasing by a multiplicative constant. The exact constant is the exponentiation of the corresponding weight of the magnitude feature, and this is station dependent. In fact, if the weight was negative then the detections odds would decrease; this is, in fact, the case for the travel-time feature.

Directly estimating the feature weights μ_d^{wjk} is not always possible because many of the station–phase combinations have very little data. To deal with this data sparsity we used a hierarchical Bayesian procedure (Gelman *et al.*, 2004), which posits that the weight for a station–phase is drawn from a global prior for that phase. This global prior is, in turn, drawn from a weakly informative prior, as follows:

$$\mu_d^{wjk} \sim \mathcal{N}(\mu_d^{wj}, \sigma_d^{wj}) \qquad \mu_d^{wj} \sim \mathcal{N}(0, 100)$$
$$(\sigma_d^{wj})^{-2} \sim \Gamma(0.01, 100),$$

where $\mathcal{N}$ and Γ are the Gaussian and the gamma distributions parameterized by their mean and standard deviation, and shape and scale, respectively. Estimation of parameters follows a coordinate ascent procedure. For each phase j, we initialize $\mu_d^{wj} = 0$ and $\sigma_d^{wj} = 1$, and then alternately optimize $\mu_d^{wjk}{:}\forall w, k$, $\mu_d^{wj}{:}\forall w$, and $\sigma_d^{wj}{:}\forall w$ until convergence. In each maximization step, the optimal value of μ_d^{wjk} is computed by second-order optimization, while the remaining values have a closed-form solution.

For each phase, the previously described procedure ensures that if a station has a lot of data (both detections and nondetections), then its feature weights will be determined almost entirely by its own data, whereas feature weights for data-poor stations tend toward a global average obtained from all stations. Thus, if most data-rich stations

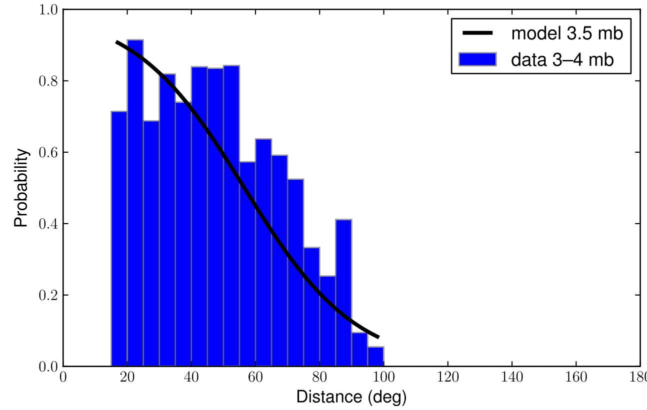

Figure 5. Conditional detection probabilities for the P phase of surface events between 3 and 4 mb at station ASAR. The color version of this figure is available only in the electronic edition.

have a positive weight for event magnitude for the P phase, then a station with very little data will also have a positive weight for this feature.

In Figure 5 we show the model prediction that is, the probability of detection, for one phase at a station.

Arrival Attributes. If event i's jth phase is detected by a station k, we define Λ^{ijk} as the corresponding arrival. For notational convenience, we define $\Lambda^{ijk} = \zeta$ whenever this phase is not detected. Our model specifies a probability distribution for the measured attributes of this arrival: the onset time Λ_t^{ijk}, the azimuth Λ_z^{ijk}, the slowness Λ_s^{ijk}, the amplitude Λ_a^{ijk}, and the assigned phase label Λ_h^{ijk}. The arrival time is

$$\Lambda_t^{ijk} = e_t^i + I_T^j(e_d^i, \Delta_{ik}) + r_t^{ijk},$$

where I_T^j is the prediction from the IASPEI travel-time model (a function of the event depth and distance to the station), and r_t^{ijk} is a random residual. The residual distribution accounts for the inhomogeneities in the Earth's crust, which allow seismic waves to travel faster or slower than the IASPEI prediction. This distribution also accounts for any systematic biases in picking seismic onsets from waveforms. Whereas most authors assume Gaussian residuals, an assumption implicit in the use of quadratic cost functions in the GA system, we find experimentally that travel time and indeed most other residuals are distributed according to a Laplacian method. The parameters of a Laplacian distribution are the mean μ_t^{jk} and scale b_t^{jk}, and the distribution is given by

$$P_{\phi,t}^{jk}(\Lambda_t^{ijk}|e^i) = \frac{1}{2b_t^{jk}}\exp\left(-\frac{|\Lambda_t^{ijk} - e_t^i - I_T(e_d^i, \Delta_{ik}) - \mu_t^{jk}|}{b_t^{jk}}\right).$$

Similarly, the arrival azimuth and slowness follow a Laplacian distribution: