

A Statistical View of the Transient Signals that Support a Wireless Call

A. Buvanewari, John M. Graybeal, David A. James, Diane Lambert,
Chuanhai Liu and W. Michael MacDonald

May 3, 2007

Abstract

A wireless call in a cellular network requires the cooperation of a wireless enabled device such as a cell phone and a dynamic set of base stations, antennas and other network elements, with control of the call changing in response to changes in signal strength and the location of the mobile. The signaling between the network and mobile that is needed to manage the call generates a huge amount of signal strength data, some of which is seen only by the mobile placing the call and some of which is seen only by the network. This paper describes some of the complexities of the signal strength data and provides a statistical model of the signals from the network to the wireless device that takes the time dependent, spatial, and multivariate nature of the call into account. An approach to estimating the model parameters online as network data are collected, which would be useful for network monitoring, is also described and applied to a set of call data obtained from an active commercial CDMA (Code Division Multiple Access) network.

Keywords: Event histories, multivariate time series, network monitoring, online estimation, signal propagation, spatial models.

A. Buvanewari, John M. Graybeal and W. Michael MacDonald are Members of Technical Staff in wireless engineering research departments at Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ. David James is Associate Director of Biostatistics at Novartis Pharmaceuticals in Hanover, NJ. Diane Lambert is Research Scientist at Google, New York, NY. Chuanhai Liu is Professor of Statistics at Purdue University, West Lafayette, IN. James, Lambert and Liu were previously Members of Technical Staff at Bell Laboratories. We thank Paul Polakos and Larry Drabeck for their many helpful insights into wireless networks and calls.

1 Introduction

A wireless network is never idle. Each *transmit antenna* on a base station in the network constantly emits a pilot signal, and each cell phone or other wireless device that is turned on constantly monitors all the pilot signals that it receives. When a cell phone initiates or receives a call, the base station with the strongest signal at the phone takes control of the call and becomes the *primary base station* for that call. Only the primary base station handles the voice traffic for the call, but the cell phone continues to monitor pilot signals from other base stations as directed by the network. If the mobile moves away from the base station or network conditions change, control of the call passes smoothly to another base station whose pilot signal the mobile is already monitoring. The list of pilot signals that the phone is monitoring plays a key role in how the call is managed, and this list can change rapidly throughout a call.

From the perspective of the network, a wireless call is a set of transient signals between the network and the phone whose sources and strengths change in response to the location of the phone and network conditions. Understanding these signals is important for understanding call quality and network efficiency. Generally, a strong pilot signal is better for call quality, but pilot signals can interfere with each other so two strong pilot signals may not be better than one. Of course, which signals are strong changes over time and with location, so understanding the dynamics of call quality requires understanding the dynamics of the pilot signals that the phone receives during a call. This set of transient signals also determines network efficiency. Each change to the set of pilot signals monitored at the mobile, either dropping or adding a signal to the list of monitored signals, incurs a cost to the network. Each cost is slight, but there are many calls and many changes to the list of pilot signals monitored during each call so the total cost to the network of actively maintaining a list for each call is large. Optimizing the parameters that determine whether a signal is added to or dropped from the call requires understanding the statistical behavior of the signal strengths at the phone.

This paper explores the nature of the data that can be collected on the strengths of all the received pilot signals at the mobile during a wireless call or data session and provides a statistical model that can be updated online as new measurements are received. Section 2 provides more background on wireless network performance and introduces two sources of received pilot signal strength data in wireless networks: the network and drive tests. The network collects the current received signal strength readings from the mobile whenever certain thresholds on signal strength are crossed (see Section 3). In drive testing a van with specialized equipment continually captures all the received signal strengths for a call

made while the van drives through an area (see Section 4). The drive test data are too costly to obtain routinely, but they are useful for model exploration. Understanding the complete time series from drive testing is a critical step to understanding the event-based data that are routinely collected by the network and could be used for call management. A model for the multiple, transient received signal strengths of a call is justified in Section 5. This model takes the network characteristics, antenna properties, and signal propagation into account. A method of online estimation of the model parameters is described in Section 6. The model fit for a set of drive test data taken on a commercial network is presented in Section 7. Section 8 further discusses the value of knowing the statistical behavior of wireless signals and concludes with open problems.

2 Cellular Networks

2.1 Network Elements

Connecting and maintaining a wireless call or data session requires the ongoing cooperation of the user's *mobile station* (e.g., cell phone) and one or more base stations that are at most a few kilometers away. In a CDMA (Code Division Multiple Access) network, each transmit antenna on a base station emits a *pilot signal* or beacon at constant power and fixed frequency, and each mobile that is turned on scans for pilot signals. The mobile maintains a list of pilot signals that it can receive, dropping base stations that it can no longer hear and adding base stations as instructed by the network. The details of how the list of pilots changes depend on the specific wireless network technology in use.

The area around a base station is typically divided into three sectors that are labeled α , β , γ . Each sector has one transmit antenna and one or two receive antennas. The antennas are usually directional and together cover 360 degrees around the base station. The location, height, azimuth (horizontal angle), tilt (vertical angle), beam width, and power of the transmit antennas determine the area or *footprint* in which the base station can provide reliable communication. The footprint can be complex because buildings and land formations affect the propagation of radio signals, sometimes impeding them and sometimes strengthening them by reflection.

The *mobile switching center* (MSC) is the final element in a cellular network. The MSC is connected to the base stations and to wired networks and the internet by cables. It is the MSC that directs the base stations and manages call processing.

2.2 Call Dynamics: Managing Quickly Varying Radio Signals

When a call is in progress (where calls are understood to include data sessions), the information that is needed to maintain and control the call and the encoded content of the call are sent on radio waves that travel between the mobile and a set of antennas within its reach. Each connection between a mobile and a sector is called a *leg*, and each leg is comprised of two unidirectional links: the *forward link*, which carries signals from the transmit antenna for the sector to the mobile, and the *reverse link*, which carries signals sent by the mobile to a receive antenna for the sector. A call can have multiple legs simultaneously if it is in contact with more than one sector, and the set of active legs can change many times during a call as the set of monitored sectors changes. Legs are dropped by the network when the forward link signal at the mobile becomes too weak, and legs can be added when the forward link signal becomes sufficiently strong at the mobile. (Calls can also be dropped because the signal generated by the mobile, which the network does not control and does not know, is too weak.) The strengths of the legs of a call fluctuate throughout the call even if the mobile is stationary because of strong scattering from the surrounding environment.

The quality of a call clearly depends on the quality of both the forward and reverse links. However, the network knows the quality of only the forward link to the mobile. At certain events described in Section 2.4, the mobile reports the strength of all the pilot signals that it is monitoring to the base station, so the base station knows the quality of the forward links at the mobile at those times. Because our focus is on network management and not device performance, we use only the forward link data available to the network, but we recognize that this misses the effect on call quality of the signals sent by the mobile.

Finally, we note that our data were obtained from a commercial CDMA network. In first generation cellular systems, a mobile call is assigned a narrow frequency band for voice or data transfer that is reserved for its exclusive use until the call ends. Second generation and later systems increase network capacity by multiplexing or allowing multiple users to share a broader frequency band simultaneously. In time-domain multiple access (TDMA) networks, users who share a frequency band transmit and receive during pre-arranged, non-overlapping time slots. In CDMA networks, each user is assigned an orthogonal filter or basis function called a Walsh code that is convolved with the traffic signal so multiple users can share the same frequency band simultaneously. Digital decoders (inverse filters) are used to recover the original signals. For more background on wireless networks, see Lee (1995). Although this paper considers only CDMA networks, our statistical analysis does not depend on the details of the encoding or frequency allocation and should be relevant for other wireless technologies like UMTS

(Universal Mobile Telecommunications Service) that also provide multiple links per call.

2.3 The Need for New Metrics for Wireless Network Performance

The performance of commercial cellular networks is routinely summarized by aggregate metrics, such as the hourly fraction of failed call attempts, the hourly fraction of initiated calls that terminate abnormally (drop), and the hourly number of Walsh codes in use. Such metrics aggregate performance over many calls and callers and do not describe the experience a single user has on a per-call basis. Aggregates that average over all calls in a time period also fail to capture the interactions between simultaneous calls that can affect network dynamics. Finally, coarse summaries like fraction of dropped calls are of limited value for evaluating the performance of networks with significant data traffic. Data sessions (e.g., web browsing) have infrequent bursts of activity that put high demands on network resources for short periods of time. These transient demands are important to the customer's network experience, but time averages smooth them out and thus hide them. As service providers attract more data users to their networks, the need for more finely grained metrics increases.

Further, cellular networks may soon be able to self-optimize and adapt to changing conditions on short time scales. Such online tuning will require more detailed knowledge of call behavior than aggregate metrics provide. (See Borst et al (2005) and Buvanewari et al (2005) for more information about future cellular network optimization and management services.) In current networks, the ability to rescue a call that is about to be dropped requires the ability to predict the next state of a call in progress, which requires an understanding of the strengths of its traffic bearing and signalling channels that aggregated performance metrics cannot provide.

Part of the goal of this paper is to start the process of developing models for forward link signals that will provide insight into how to develop better metrics for network performance and network adaptation.

2.4 Measurements from Wireless Networks

Routinely obtaining signal strength data for monitoring commercial cellular networks is not easy. The mobile continuously monitors the relative strengths of the pilot signals it receives during a call but it does not record them. Time series of received signal strengths can be obtained only from specialized test mobiles using software and equipment that are impractical for all but small, infrequent "drive test" studies. Even then, drive testing has limited value for understanding ongoing network performance because drive test data are collected on only one mobile taking one path through one area of one network under the set of traffic and network conditions that are in effect at test time. Drive testing fails to capture

the broad range of network conditions and user behaviors, such as in-building vs. outdoor use, that are critical to system performance.

In contrast, the MSC sees the received signal strength at the mobile for each leg of each call that it manages whenever there is a change in signal strength that could lead to the addition or deletion of a leg for the call. At other times, the cost of transmitting the data from the mobile to the base station over the air is too high relative to its value. The MSC data are accessible (although not trivially so) because the MSC, like other switches and routers, can be passively monitored. That is, the mobile sees a complete time series of received signal strength but it is highly impractical to capture these data routinely, while the MSC has an incomplete but accessible view of all calls under all network conditions. Note that although the MSC has a limited view of the mobile, the MSC data are used to manage the call and, to a large degree, determine its fate.

Our ultimate goal is to have a statistical model of a call that can be estimated solely from MSC data and that can be used to identify calls that are about to fail but can be rescued or to adapt network parameters online, for example. This paper provides a first step in that direction by describing the nature of wireless data, introducing a statistical model of the pilot signal strengths received in the multiple legs of a call throughout its duration, and showing how the parameters of the model can be estimated from the time series data collected by a mobile. The next step of estimating the parameters of the model from MSC data will be explored in a later paper. Although our time series model requires data that are not available to the MSC for call management, it is realistic in the sense that it accounts for the geometry of a network, the interaction of network elements, and the dynamics of adding and deleting legs during a call. While there has been much work on models of radiowave signal propagation (see, for example, Jakes (1974) and Patzold (2002)), we believe that this is the first model that takes the energy received in the multiple legs of a call into account and that is based on extensive measurements on a commercial network.

3 The Network View of a Wireless Call: Event Data

When a call initiates, the MSC assigns multiple traffic channels or *legs* to the mobile to reduce the chance of losing the call; the set of participating legs is called the *active set*. One leg is selected as the *primary leg*. Any other active leg, whether for another sector of the same base station or a sector of a different base station, is called *secondary*. A secondary leg may become stronger than the primary leg, but that does not change its status as a secondary leg. A call may have many secondary legs, and

these can start and end throughout the call as the relative strengths of the signals that the mobile receives increase and decrease. When a leg in the active set weakens so much that it is no longer viable, it is dropped from the active set. Similarly, legs can be added to the active set during the call. Any change in the set of active set is called a *handoff* or *handover*. The maximum number of active legs for a call is a configurable parameter that may vary across the network.

The mobile reports its received signal strengths for the active legs of a call to the primary base station, and thus to the MSC that manages the call, whenever one of these received strengths crosses a threshold. When the signal crosses the T_{add} threshold from below, it can be added to the active set. When the signal crosses the T_{drop} threshold from above and stays below T_{drop} for a specified length of time, it is dropped from the active set unless it is the only signal in the active set. Each threshold crossing is called a *handoff trigger* and we call the time of the threshold crossing a *trigger time*. There can be many handoff triggers during a call even if the mobile is stationary. For example, a nine day study of four base stations and one frequency band in a small urban region of a commercial network generated about 2.7 million handoff triggers during 30 million seconds of call time.

The births and deaths of legs between a mobile and a sector are controlled by their relative signal strength E_c/I_0 at the mobile, where E_c/I_0 for a transmit antenna A is defined as the ratio of the energy E_c received in the pilot signal of antenna A to the total energy (I_0) that the mobile receives, summing over all pilots (including A) and the thermal noise in the receiver; E_c/I_0 is expressed in $10\log_{10}$ (decibel) units. Relative and not absolute signal strength is measured because a signal from one antenna interferes with the signals for all other antennas. Note that the ratio is one (0 dB) only when there is just one signal and no thermal noise.

Figure 1 shows the trigger-driven E_c/I_0 data for the end of a call that terminated abnormally. The times of the triggers are indicated by faint vertical lines that extend through all the panels. The call shown has active legs with the α and β sectors of base station 1 and the β sector of base station 7. The top panel shows the relative pilot strength E_c/I_0 at the mobile for each of these links as reported by the mobile at trigger times. The horizontal bar in the middle of the top panel shows the lower threshold T_{add} and upper threshold T_{drop} on E_c/I_0 beyond which the mobile can request that legs be added or dropped. If E_c/I_0 rises above the top of the band (T_{add}), then the leg for that sector can be added to the active set if the active set is not full or to the *candidate set* otherwise. A leg can be dropped if its E_c/I_0 falls below the bottom of the gray band (T_{drop}) and stays there for a specified duration, as long as there is another active leg with a larger E_c/I_0 . Legs are usually added quickly when they rise above T_{add} , but they are not dropped as soon as they fall below T_{drop} to avoid reacting to short, temporary

changes in the environment, such as a passing truck. The time out period for dropping a leg is a network configurable parameter, typically about 4 seconds.

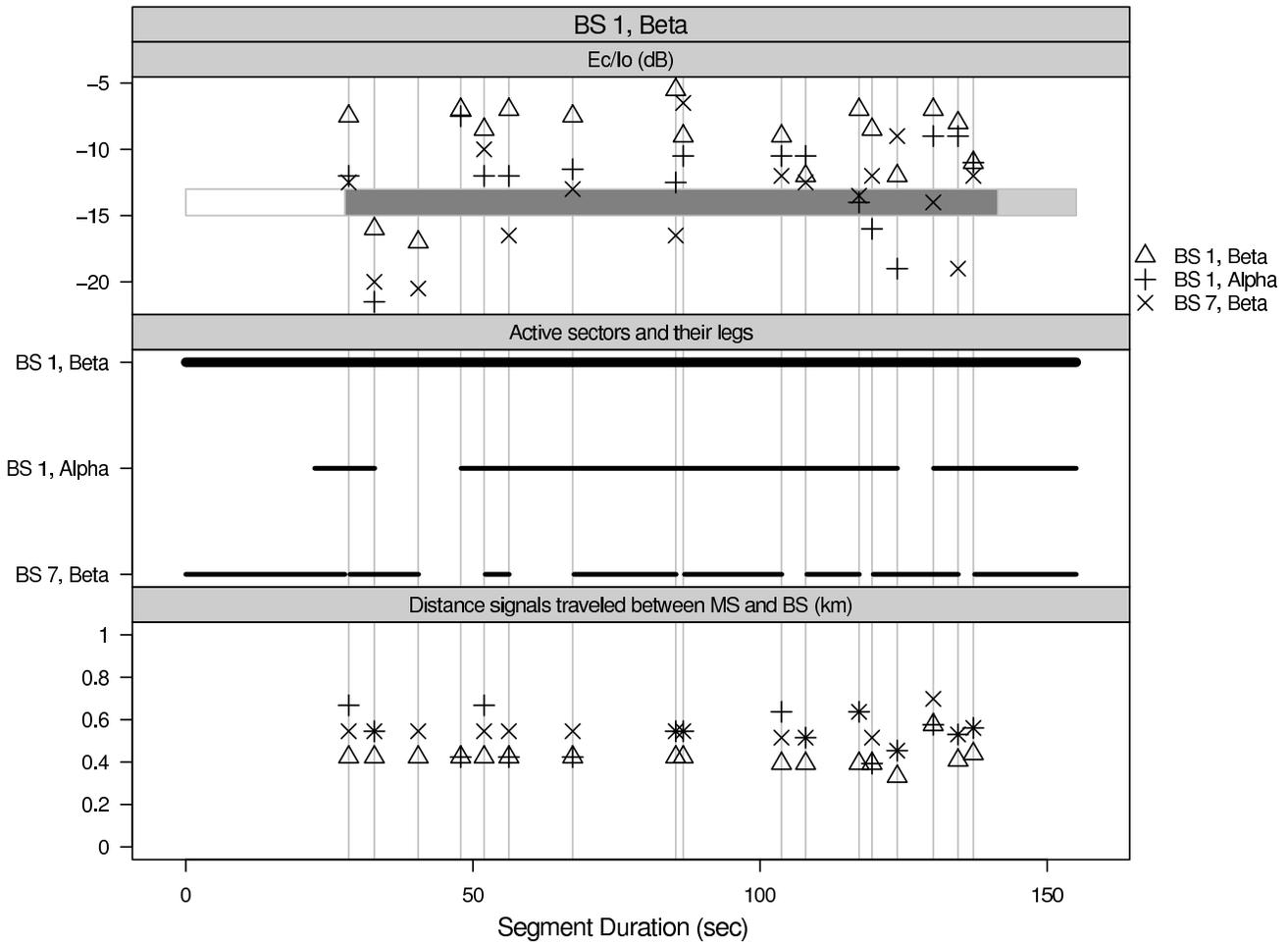


Figure 1: The MSC view of a call during a period when sector β of Base Station 1 was active; the dark horizontal band in the top panel indicates the period during which this sector was primary. The panels from top to bottom show relative signal strengths of the active legs at trigger times (marked by vertical dotted lines), the births and deaths of active legs for each sector, and distance (computed from signal round trip time) for the active legs.

The horizontal bar in the top panel, including the white portion of the bar on the left and the light portion of the bar on the right, shows the start and end of the leg for the β sector of base station 1, which is henceforth denoted by BS-1 β . The darker gray of the bar indicates the period during which the BS-1 β leg was primary. Because the BS-1 β leg was not primary at time zero in this panel, we know that it started as a secondary leg and became primary through a handoff 25 seconds after it began, at the time

of the first trigger shown. The E_c/I_0 for BS-1 β drops below T_{drop} at the next trigger at 30 seconds, but there is no stronger signal then and so the BS-1 β leg was not dropped. At about 40 seconds, all the legs strengthen, with BS-1 β continuing to be the strongest signal through the end of the period shown. The light gray period on the right end of the bar indicates a timer that is started by the MSC when contact with the mobile is lost. The call was considered lost at the end of the timer, although the signal on the forward channel to the mobile appears to be strong at the end of the call. The reverse link from the mobile to the base station may have been weak, though, causing the call to end. Both the reverse and forward links are needed to maintain a call.

The middle panel of Figure 1 shows the lifetimes of the legs for each sector during the period that the BS-1 β leg was primary. In all, 12 active legs were created, eight of which were for BS-7 β . One of the gaps between dropping and re-adding BS-7 β to the active set (and the gap between the corresponding trigger times) is imperceptible in Figure 1 and another gap is nearly imperceptible. Transient weak legs such as these can be kept in the active set to protect against a rapid drop in E_c/I_0 for the strongest signal. The question of whether cycling BS-7 β in and out of the active set is worth the trade-off in reliability versus cost is one that has not been quantified.

Finally, the bottom panel of Figure 1 shows the distance of the mobile to the primary antenna as computed from the time it took a signal to travel from the transmit antenna to the mobile and then back to the receive antenna. This round trip time or distance can change even if the mobile is stationary due to random fluctuations in the environment and to temporary obstructions between the mobile and antenna that affect the signal path. In this example, it appears that the mobile is either not moving or moving slowly, although it is perhaps possible that it is fortuitously moving on a path of constant mean round trip time. If that is not the case, then the addition and deletion of legs is probably in response to changing network and traffic conditions, not in response to the location of the mobile.

Clearly, the anatomy of a call can be complex, especially when the mobile is moving. There can be many active legs and many “candidate” sectors that could support active legs should an active leg be dropped. (To simplify the plot, measurements from the candidate sectors are not shown in Figure 1.) The mobile also receives signals from neighboring sectors that could become candidate legs. These too are not shown in Figure 1. No leg may last throughout an entire call. But no matter how complex the call or network dynamics, it is still the measurement of E_c/I_0 that determines which legs are active, when a leg should be dropped or a new leg added, and the quality that the user experiences throughout the call. Thus, the remainder of this paper focuses on E_c/I_0 .

4 The Mobile View of a Wireless Call: Time Series Data

Relative signal strength E_c/I_0 at the mobile and the location of the mobile as determined by GPS (Global Positioning System) were obtained every second of a 2.5 hour call made mid-day from a mobile in a car driving through an urban area with low rise buildings. The path of the car and its speed throughout the period are shown in Figure 2, where speed is computed by taking a moving average of the distance traveled between one second measurements with a window of 30 seconds. The car moved about 5 mph for almost half the period. It rarely exceeded 15 mph.

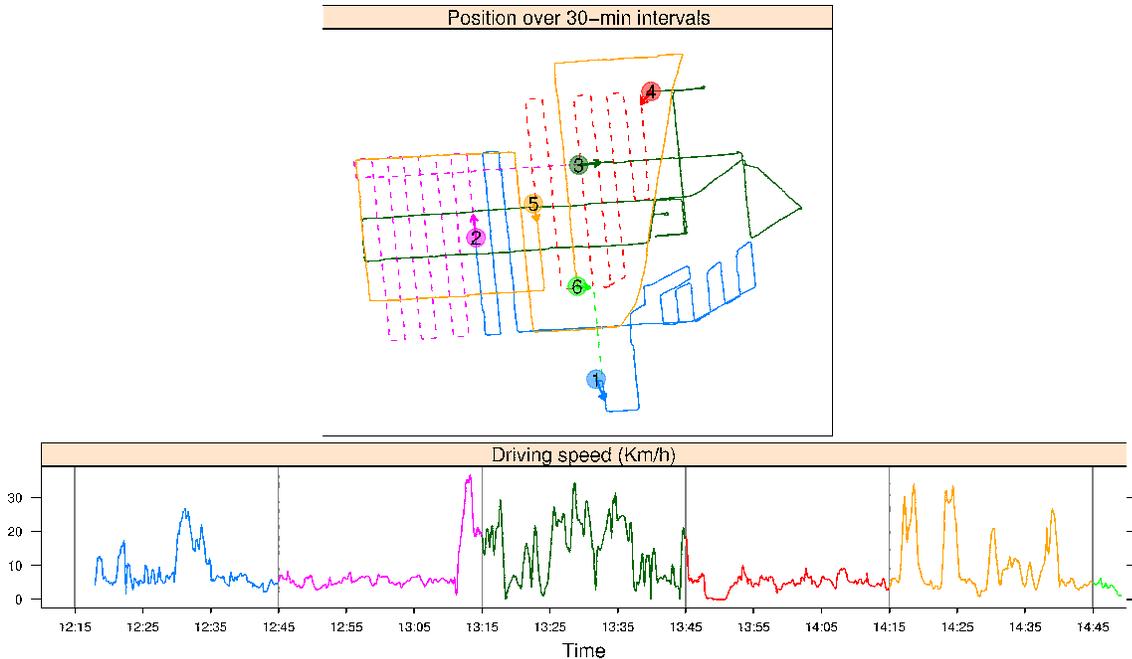


Figure 2: The mobile's path (top) and speed (bottom) during the drive test. Small dots in the top panel show the location of the mobile every minute; large dots show its location at the start of each 30 minute interval. The numbers in the large dots indicate the sequence number for the interval.

Figure 3 shows the footprints of four of the base stations that supported the drive test call. Columns represent base stations; rows represent their sectors. The location of a base station relative to the drive test region is shown as a large dot, and the azimuth (center) of a transmit antenna beam for a sector is shown as a long thick arrow. The shorter arrows around the beam depict the beam width, which is the angular range for which E_c/I_0 drops off by at most a factor of 2 from its peak at the azimuth.

The dark points on each map in Figure 3 show the footprint of a sector in the sense that the sector had a primary leg at least once during the drive test at each location marked by a dark point. As would

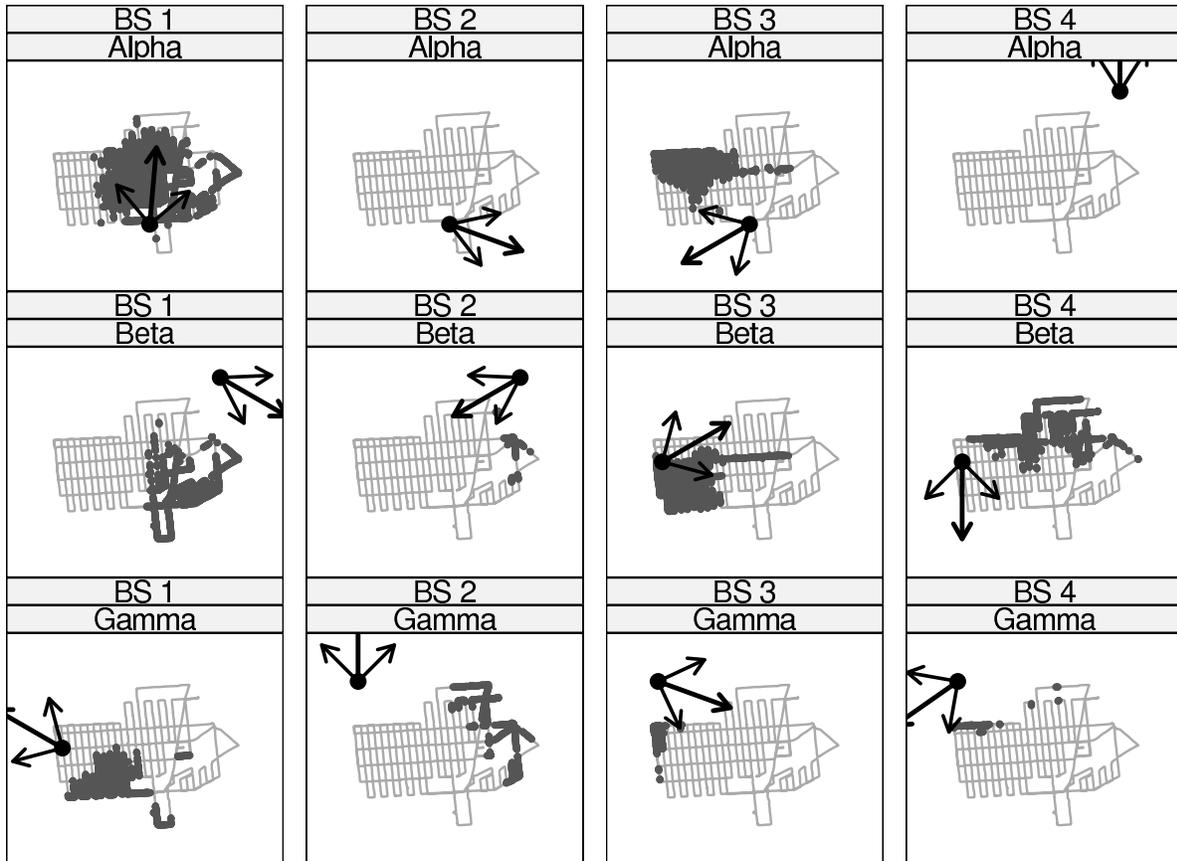


Figure 3: The footprints of several sectors during the drive test. The large dot at the end of an arrow indicates the antenna location. The dark arrow indicates the azimuth of the beam, and the lighter arrows indicate the beam width. A smaller black dot on the drive path indicates that the sector supported a primary leg for the mobile at least once at that location during the drive test.

be expected from their locations and directions, BS-2 α and BS-4 α never supported primary legs for the drive test call, while BS-4 β supported primary legs when the mobile was in the north of the region and BS-1 β supported primary legs when the mobile was in the southeast. Surprisingly, BS-1 γ was primary for some locations directly behind it and thus opposite to the azimuth of its transmit antenna beam. Some weak signal “leakage” behind the antenna would be expected, but not at the large distances seen in Figure 3. Moreover, several of the sectors shown, such as BS-1 γ , BS-2 β , and BS-4 γ , were dominant in non-contiguous regions. These patterns may be caused by reflections from buildings that are directly in front of the antenna beam, for example. Finally, note that BS-1 β has a long reach along a road that is perpendicular to the azimuth of its beam. The complexity of the patterns shown in Figure 3 suggest that modeling signal propagation is not easy.

Finally, Figure 4 shows five minutes of the mobile’s data. The call had 28 active legs, denoted by colored curves, in this five minute period, although most were only slightly above T_{add} and lasted just a few seconds, so are barely discernible. One leg was dominant throughout much of the period, but around 12:22 it became noticeably weaker than other legs. The candidate signals (black circles) are not yet actively supporting this call but might be strong enough to support a call. (The maximum numbers of active and candidate legs are network parameters chosen by the service provider.) Signals from neighboring sectors (gray circles) could be added to the candidate set. The neighbor list is typically the set of sectors adjacent to the primary sector, and so is important for maintaining call quality when the mobile is moving quickly. However, the neighbor and candidate pilot signals contribute to interference for the active legs.

Active legs tend to be stronger than candidate legs, but there are exceptions in Figure 4. Also, E_c/I_0 sometimes changes rapidly, perhaps in response to passing interference (such as a truck) or changes in the landscape (such as occur when turning the corner of a building). Moreover, there are many triggers, or threshold crossings, even though Figure 2 shows that the mobile traveled only a short distance in this five minute period.

The lower panel of Figure 4 shows the number of neighboring signals received by the mobile throughout the call. Whenever there is a trigger, the mobile re-sets all stored values of E_c/I_0 for the neighbors to a minimum (-32.5 dB) and the base station transmits a new neighbor list to the mobile. The rising slope after the triggers indicates how quickly the mobile finds signals with E_c/I_0 above -32.5 dB after the trigger.

The remainder of this paper builds a model of E_c/I_0 for a set of active, candidate and neighbor list signals and validates it on the data collected during the 2.5 hour drive test. A model of raw signal

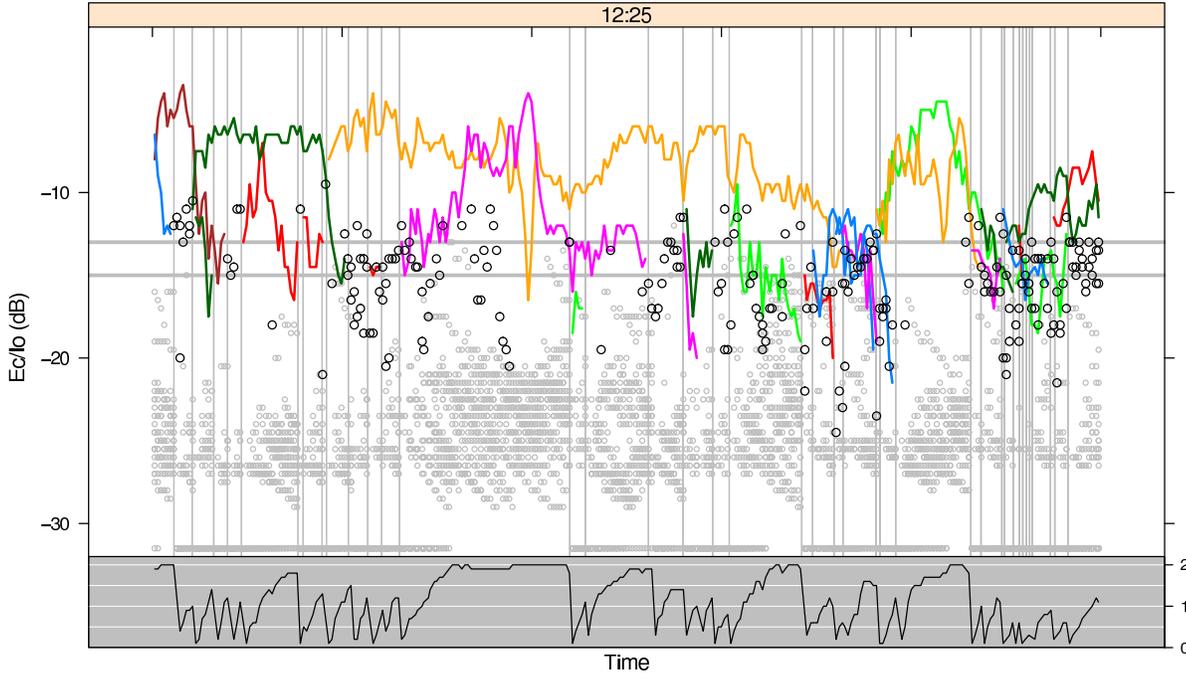


Figure 4: Time series of E_c/I_0 over a five minute period for active legs (colored) and for candidate (black) and neighboring (gray) sectors. Vertical lines denote trigger times at which E_c/I_0 values are reported to the MSC. Horizontal lines show the lower threshold T_{drop} and upper threshold T_{add} for dropping and adding active legs. Values of -32dB, if recorded at all, are censored from below so might be even smaller.

strength for one leg of a call would be simpler and more standard (see Jakes (1974) and Patzold, 2000), but unrealistic because only relative signal strength E_c/I_0 is known to the MSC that manages the call and the multiple legs of a call interact and together determine the quality of the call.

5 A Model of Relative Forward Link Signal Strength at the Mobile

In this section we build a model for relative signal strength that incorporates network topology and signal propagation. Specifically, we allow relative signal strength at a location to depend on its distance from the antenna and its angular deviation from the azimuth of the beam. More precisely, the intercept of our model depends on the deviation from the azimuth of the beam and the rate of degradation of relative signal strength with distance also depends on the deviation from the azimuth. Both the intercept and slope (degradation) parameters are taken to be nonparametric functions of deviation from the azimuth of the beam. We also incorporate the fact that signal strength does not appear to continue to increase as the antenna is approached but seems to flatten out at a maximum value at a small distance from the antenna. This is not surprising because the antenna is unlikely to be at the same height as the mobile and our model (and data) considers only planar distance from the antenna.

The relative strength E_c/I_0 of a signal from transmit antenna i , $1 \leq i \leq I$ in decibels is defined as

$$10 \log_{10} \left(\frac{E_{it}}{E_{0t} + \sum_{j=1}^I E_{jt}} \right),$$

where E_{jt} , $j = 1, \dots, I$ is the energy in the signal from antenna j at time t and E_{0t} represents thermal noise plus all sources of signal energy not accounted for by the I known pilot signals, including the term E_{it} in the numerator.

Based on exploratory data analysis (not shown here), we find that

$$x_{it} = \ln(E_{it})$$

is reasonably well-modeled by an autoregressive (AR) model of order one with mean μ_{it} , variance σ_e^2 , and autocorrelation coefficient ϕ ,

$$x_{it} = \mu_{it} + \phi(x_{i,t-1} - \mu_{i,t-1}) + e_{it},$$

where $-1 < \phi < 1$ and the e_{it} , $1 \leq i \leq I$, are independent normal($0, \sigma_e^2$) random variables. (The description of the model fit in Section 7 also suggests that the simple AR(1) model is adequate for our data.)

A bit of algebra shows that with

$$u_{it} = \ln \left(\frac{E_{it}}{\sum_{j=0}^I E_{jt}} \right)$$

and

$$y_{it} = \ln \left(\frac{e^{u_{it}}}{1 - \sum_{j=1}^I e^{u_{jt}}} \right),$$

we have

$$y_{it} = x_{it} - x_{0t} = \mu_{it} + \phi(y_{i,t-1} - \mu_{i,t-1}) + \epsilon_{i,t} - \phi z_{t-1} + z_t, \quad (1)$$

where $z_t = -x_{0t}$ is thermal noise plus the signal strengths of unmeasured antennas. We assume that z_t is AR(1) with mean zero, variance σ_z^2 and autoregression coefficient ρ the same for all antennas. The same ρ and σ_z^2 are used for all measured signals because all are affected by the same set of measured and unmeasured signals. Note that the y_{it} 's, $i = 1, \dots, I$, which are based on relative signal strength, are correlated over antennas (space) and time, but the raw signal strengths x_{it} are correlated only over time. Finally, y_{it} , which is an observable function of E_c/I_0 , is AR(1) and has mean μ_{it} .

To be realistic, the μ_{it} in equation (1) must depend on distance from the antenna and angular deviation from the direction of the antenna beam. To explore the nature of these effects, compute a standardized location for each single observed y_{it} taken during the drive test call discussed in Section 4, whether for an active or candidate leg, by placing the transmit antenna for y_{it} at the origin and pointing it due east along the positive horizontal axis. Having placed all the y_{it} on a standardized map, we partition the map into small regions and compute an empirical cumulative distribution function (ecdf) of y_{it} in each. Figure 5 shows the right tail of the ecdf. For example, at least 85% of the measurements for a 90 degree antenna beam are below -20 and at most 9% are above -10. (Beam widths were taken from network configuration files.)

As would be expected, Figure 5 suggests that relative signal strength is largest at the azimuth of the beam and falls off with both distance from the antenna and deviation from the azimuth and that relative signal strength is roughly symmetric around the azimuth. There are discrepancies from this simple pattern that are likely caused by the physical environment in this portion of the network, though. For example, as in Figure 3, there are strong signals behind the antenna – far outside the nominal beam width – perhaps because there are buildings in front of the antennas that reflect or refract the beam. Simple signal propagation models would not predict strong signals as far behind the antenna as seen in Figure 5. Also note that Figure 5 suggests that relative signal strength does not continue to increase as the mobile approaches the antenna along a ray within the antenna beam. This is perhaps because signal

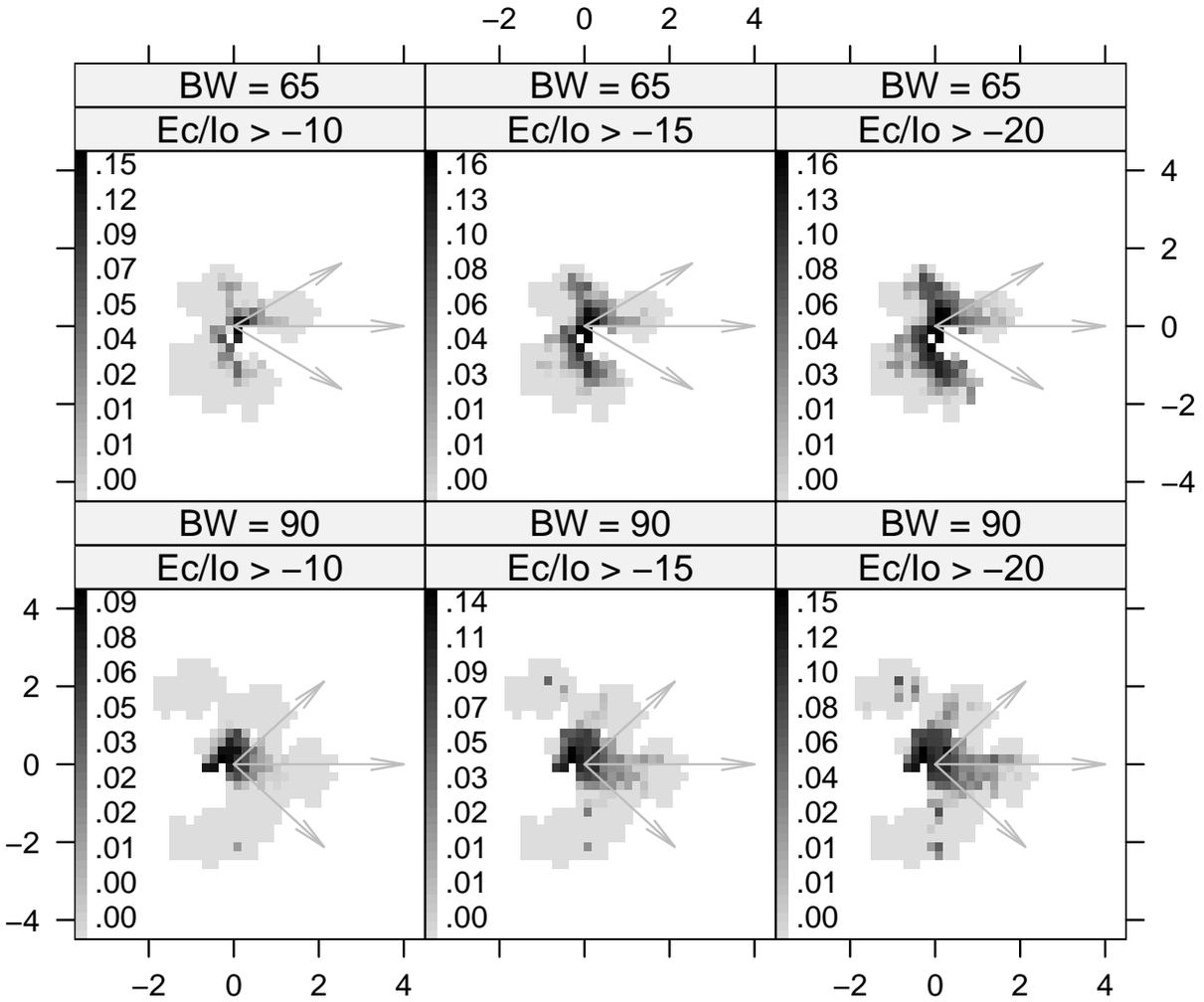


Figure 5: The distribution of y_{it} over 2.5 hours of a drive test call as a function of distance (in kilometers) from the base station and deviation from the azimuth of the beam for 13 antennas with 90 degree beams (bottom row) and 11 antennas with a 65 degree beams (top row). The panels show the fraction of E_c/I_0 measurements above a threshold in small regions.

strength also depends on the angle in the vertical plane and Figure 5 ignores vertical distance from the antenna. Signal strength also appears to be roughly symmetric around the azimuth of the beam.

Standard signal propagation models (e.g., Jakes (1974) and Patzold (2002)) assume that average signal strength in free space measured in decibels degrades with the log of distance from the antenna. The intercepts and slopes of these models depend on the environment; e.g., they are different for low density rural and for high-density urban environments. For directional antennas, signal strength also degrades as the mobile moves away from the azimuth of the antenna beam. A simple model of the mean that is roughly consistent with these predictions and with Figure 5 is

$$\mu_{it} = \alpha_i(\theta_{it}) + \beta(\theta_{it}) \ln(\sqrt{d_{max}^2 + (d_{it} - d_{max})^2}), \quad (2)$$

where d_{it} is the distance of the mobile from antenna i at time t , d_{max} is the distance below which signal strength no longer increases (here assumed to be 0.15 km), and θ_{it} is the absolute angular deviation of the mobile from the azimuth of the beam in the horizontal plane (ignoring any effects of height). Angular deviation is normalized to lie between 0 and 1, controlling for beam width. More precisely, we define the angular deviation θ of an angle a from the azimuth of a beam of width $2b$ by

$$\theta(a) = \begin{cases} \frac{|a|}{2b} & \text{if } |a| \leq b \\ 1/2 \left(1 + \frac{|a| - b}{\pi - b}\right) & \text{if } b < |a| \leq \pi. \end{cases}$$

The only antenna-specific effect in equation (2) is the intercept $\alpha_i(\theta)$, which represents the transmit power of antenna i . In other words, antenna power varies across antennas and depends on deviation from the azimuth in the horizontal plane. The rate $\beta(\theta)$ at which signal strength falls off with log distance at any given angular deviation θ from the azimuth is the same for all antennas, however, regardless of their power. Note that the rate at which signal strength degrades with distance from the antenna depends on the deviation from the azimuth of the beam.

Although it is easy to argue that the transmit power $\alpha_i(\theta)$ should be decreasing in θ and the degradation rate $\beta(\theta)$ should be increasing in θ (because signals degrade faster away from the azimuth), the exact dependence of α_i and β on θ is unknown. In what follows we assume that α_i and β are smooth, symmetric functions of θ that can be represented by a Fourier series on $(0, \pi)$ with only a few cosine terms. To allow strong signals directly behind the antenna, we model the intercept and slope coefficients as Fourier series with terms $\cos(k\pi\theta^2)$ for $k = 0, \dots, K$. Here we take $K = 4$. To summarize, the transmit power and signal degradation parameters are smooth functions of absolute deviation from the azimuth of the beam.

6 Online Estimation

6.1 Observed and Imputed Data

The location of a mobile and its signal strength readings were recorded for each second of the 2.5 hour drive test call discussed in Section 4. The distance to each antenna and the angular deviation from the azimuth of the antenna in the horizontal plane were computed from the GPS measured location of the mobile and the location and direction of the antenna, which were taken from network configuration files. Here we use model (1) with mean (2) to describe the measurements for the 25 sectors that contributed the most data. This set includes active, candidate and neighboring links.

There are two sources of missing signals y_{it} . First, E_c/I_0 is missing for all neighboring sectors from the time a trigger occurs until the mobile receives a new neighbor list and re-scans the sector. Here we assume that neighboring sectors that drop at trigger times are missing at random. This assumption is roughly right because signals in neighboring sectors are highly unlikely to cause triggers, although it is not exactly right because the neighboring sectors and active sectors may both be affected by whatever caused the trigger. Second, previously measured signals become missing when they fall below -32 dB since this is a limit of detection for these data.

The complete data would be $\{(Y_t, z_t) : Y_t = (y_{1t}, \dots, y_{It})'; t = 1, \dots, T\}$, where I is the total number of sectors the mobile was in contact with during the call. The missing data consist of the missing values y_{it} and every z_t because z_t is an unmeasured source of variability. Starting from $z_1 = 0$, we sequentially impute z_t and y_{it} when it is also missing using their predictive distributions. (See Little and Rubin (1987) for a full discussion of imputation.) More precisely, at time t , the predictive distribution of z_t conditional on z_{t-1} is normal($\rho z_{t-1}, \sigma_z^2$), so we impute z_t with a random draw from this normal distribution using mean and variance estimates based on the sufficient statistics, which are given in Section 6.2 below. We then impute the missing value $y_{mis,t}$ using the imputed values of z_{t-1} and z_t in the AR(1) model given by equation (1), replacing ρ and σ_z^2 with estimates based on their sufficient statistics, as described next in Section 6.2. Note that the mean of y_{it} is not constant and several consecutive y_{it} can be missing.

6.2 Updating the Estimated Mean

The power level $\alpha_i(\theta)$ and degradation rate $\beta(\theta)$ in the mean model (2) are assumed to change smoothly with absolute deviation from the azimuth in the horizontal plane by assuming that both are linear functions of K terms of $\cos(k\pi\theta^2)$, $k = 0, \dots, K$ for a small K . The power levels $\alpha_i(\theta)$ are allowed differ-

ent intercepts but otherwise they have the same dependence on θ . Thus, the mean μ_{it} has $2(K+1)+I-1$ parameters: $(\alpha_{01}, \dots, \alpha_{0I}, \alpha_1, \dots, \alpha_K, \beta_0, \dots, \beta_K)$. These can be estimated online (sequentially) by the method of least squares as follows.

First, define $\tilde{X}_k = \cos(k\pi\theta^2)$ and write the vector of covariates for one relative signal strength measurement as $X = (\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_K)'$. Then

$$\mu_{it} = X_{it}' (\alpha_{0i}, \alpha_1, \dots, \alpha_K)' + g_{it} X_{it}' (\beta_0, \dots, \beta_K)'$$

where $g_{it} = \ln \sqrt{d_{max}^2 + (d_{it} - d_{max})^2}$, $i = 1, \dots, I$ and d_{it} is the distance of the mobile from antenna i at time t . We then work with the ‘‘complete’’ data (observed and imputed data), with missing values imputed as described in Section 6.1.

To update the estimates of $(\alpha_1, \dots, \alpha_I)$ and β , first define the $K \times K$ matrices

$$\begin{aligned} S_{XX'}^{(it)} &= S_{XX'}^{(i,t-1)} + X_{it} X_{it}' \\ S_{gXX'}^{(it)} &= S_{gXX'}^{(i,t-1)} + g_{it} X_{it} X_{it}' \\ S_{g^2XX'}^{(t)} &= S_{g^2XX'}^{(t-1)} + \sum_{i=1}^I g_{it}^2 X_{it} X_{it}' \end{aligned}$$

and the column vectors of length K

$$\begin{aligned} S_{XY}^{(it)} &= S_{XY}^{(i,t-1)} + X_{it} y_{it} \\ S_{gXY}^{(t)} &= S_{gXY}^{(t-1)} + \sum_{i=1}^I g_{it} X_{it} y_{it}, \end{aligned}$$

where all elements of each matrix and vector are initialized to zero at $t = 0$. Given a matrix A , denote its ij^{th} element $A[i, j] = A_{ij}$, its i^{th} row $A[i,] = (A_{i1}, \dots, A_{iK})$, its i^{th} row without its first element as $A[i, -1] = (A_{i2}, \dots, A_{iK})$, its i^{th} column without its first element as $A[-1, i]$ and the matrix without its first row and first column as $A[-1, -1]$. Next define the $(2K + I - 1) \times (2K + I - 1)$ matrix

$$S_{XX'} = \begin{bmatrix} S_{XX'}^{(1t)}[1, 1] & \dots & 0 & S_{XX'}^{(1t)}[1, -1] & S_{gXX'}^{(1t)}[1,] \\ & \ddots & & \vdots & \vdots \\ 0 & \dots & S_{XX'}^{(It)}[1, 1] & S_{XX'}^{(It)}[1, -1] & S_{gXX'}^{(It)}[1,] \\ S_{XX'}^{(1t)}[-1, 1] & \dots & S_{XX'}^{(It)}[-1, 1] & \sum_{i=1}^I S_{XX'}^{(it)}[-1, -1] & \sum_{i=1}^I S_{gXX'}^{(it)}[-1,] \\ S_{gXX'}^{(1t)}[1,] & \dots & S_{gXX'}^{(It)}[1,] & \sum_{i=1}^I S_{gXX'}^{(it)}[1, -1] & S_{g^2XX'}^{(t)} \end{bmatrix}$$

and the column vector of length $2K + I - 1$

$$S_{XY} = \left(S_{XY}^{(1t)}[1], \dots, S_{XY}^{(It)}[1], \left[\sum_{i=1}^I S_{XY}^{(it)}[-1] \right]', [S_{gXY}^{(t)}]' \right)'.$$

S_{XX} and S_{XY} are the sufficient statistics for $(\alpha_1(\theta), \dots, \alpha_I(\theta), \beta(\theta))$. Finally, the estimated regression coefficients corresponding to the Fourier series model can be updated at time t using the method of least squares equation

$$(\hat{\alpha}_{1,1}, \dots, \hat{\alpha}_{1,I}, \hat{\alpha}_2, \dots, \hat{\alpha}_I, \hat{\beta}_1, \dots, \hat{\beta}_K)' = S_{XX}^{-1} S_{XY}.$$

6.3 Updating the Variance and Autoregression Coefficient

Define D to be the difference matrix with $D_{i,i} = 1$, $D_{i+1,i} = -1$, and $D_{ij} = 0$ if $j > i + 1$ or $j < i$. Then integrating out the noise terms e_{it} and z_t gives

$$D(y_t - \mu_t) = \phi D(y_{t-1} - \mu_{t-1}) + De_t. \quad (3)$$

Because $De_t \sim N(0, \sigma_e^2 DD')$, it follows that

$$\phi \sum_t (y_{t-1} - \mu_{t-1})' D(DD')^{-1} D(y_{t-1} - \mu_{t-1}) = \sum_t (y_{t-1} - \mu_{t-1})' D(DD')^{-1} D(y_t - \mu_t),$$

which gives a set of sufficient statistics for ϕ given μ and an obvious (and simple) way to update ϕ . The estimate of σ_e^2 is updated using equation (3). To reduce the effect of imputed values, we apply equation (3) and update the sufficient statistics and ϕ only when both y_t and y_{t-1} are observed; otherwise set $\hat{\sigma}_e^2$ and $\hat{\phi}$ equal to their most recent values.

Finally, the imputed noise sequence z_t is used to update ρ and σ_z^2 . First, $\sum z_{t-1}^2$ and $\sum z_{t-2}z_{t-1}$ are updated, and then the estimates of ρ and σ_z^2 are computed as

$$\hat{\sigma}_t^2 = t^{-1} \sum_{j=1}^t z_j^2 \quad \text{and} \quad \hat{\rho} = \frac{\sum_{j=2}^t z_{j-1}z_j}{\sum_{j=1}^t z_j^2}.$$

The last two estimates can be updated sequentially in the obvious way.

7 Model Fit

The AR(1) assumption was studied by analyzing pieces of call segments that appear to be stationary (plots not shown here) and was found to be appropriate for these data. Higher order AR models did not fit better than the simpler AR(1). The autocorrelation and standard deviation estimates for model (1) at

the end of the drive test call ($t = 9167$ seconds) are $\hat{\rho} = 0.86$, $\hat{\phi} = 0.95$, $\hat{\sigma}_e = 0.29$, and $\hat{\sigma}_z = 0.30$. The correlation across time and the correlation across antennas are both large, suggesting that modeling individual call segments independently would be inadequate. Contours for the estimated means are shown in Figure 6, with locations and orientations translated as in Figure 5. Note that the model has identified the strong signal strength behind the antenna, although this effect was not explicitly included in the model.

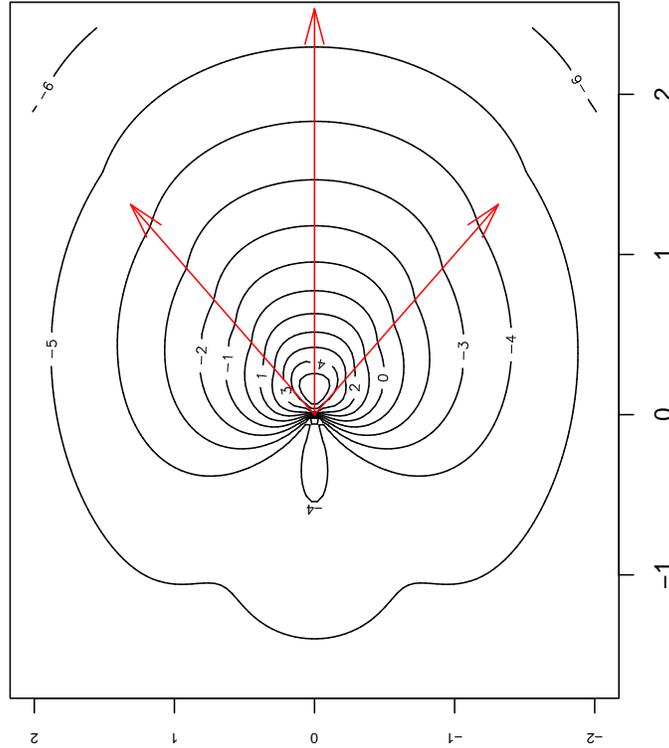


Figure 6: Estimated mean around an antenna located at the origin with a beam pointing down the positive horizontal axis. Distances along the axes are in km.

The estimated antenna-specific intercepts $a_i(\theta)$ and common slope $b(\theta)$ in the mean (2), which are assumed to be symmetric functions of θ , are shown in Figure 7. The intercept curves are parallel functions of θ , so the transmit power away from the azimuth follows the same pattern for all antennas. The intercept curves also suggest that power close to the antenna stays nearly constant for $|\theta|$ roughly between 0 and 0.3 and then degrades nearly linearly at larger deviations from the azimuth. The slope curve $\beta(\theta)$ shows that the relative signal strength decreases with log distance at increasing rates as the azimuth is approached for absolute angular deviations below 0.2, but beyond that the effect of

log distance weakens (that is, the slope moves towards zero.) These are plausible effects, and it is encouraging that the spline model has discovered them.

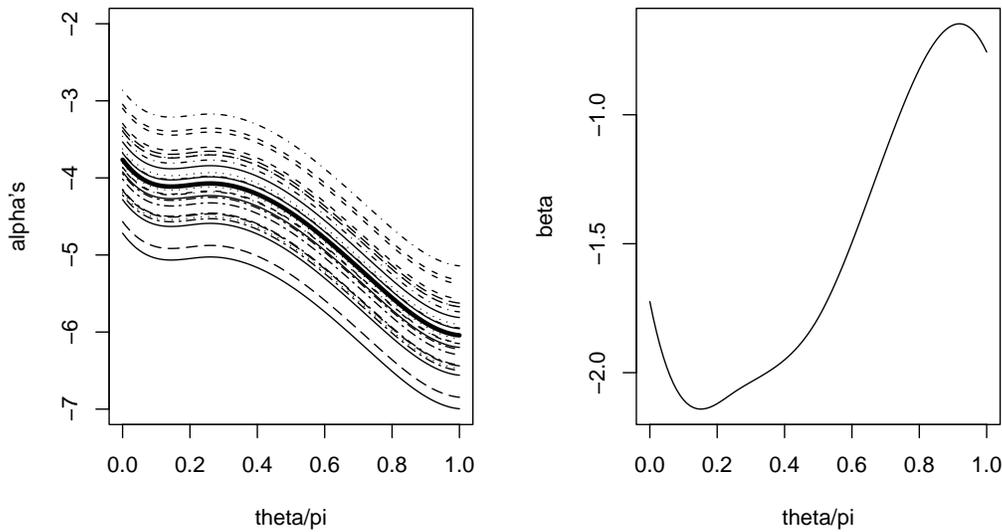


Figure 7: Estimated intercepts $\alpha_i(\theta)$ and slope $\beta(\theta)$ as a function of angular deviation θ from the azimuth in the horizontal plane, as estimated from data collected at the mobile during a 2.5 hour long drive test call. The dark middle line in the left plot shows the mean over the 25 antennas.

Figure 8 shows that the residuals from the fitted model are generally well-behaved. They are centered at zero and approximately symmetric as a function of either distance from the antenna or deviation from the azimuth of the beam. There is some evidence of a scale change with distance that is not captured by the model, though, since the quartiles generally shrink towards zero for larger distances for all angular deviations from the azimuth, but there is also less data at the larger distances so these differences may not be meaningful. Similarly, the large number of outliers in the residual boxplots is not a concern. There are about 642,000 residuals spread over 44 boxplots, or about 15,000 residuals per boxplot on average. Using the usual boxplot parameter settings, we'd expect about 100 outliers per boxplot even if the residuals are approximately normally distributed.

The residuals against the estimated mean (Figure 9) are generally well-behaved but show some evidence of bias. Overall, the model slightly underestimates weak signals and overestimates strong signals, suggesting that the mean should be a weaker function of distance than the logarithmic, perhaps changing the specification near the antenna and at the extreme distances only. Nevertheless, although the model can be improved, it appears to be a reasonable first approximation to the multiple signal strengths seen at the mobile during a call.

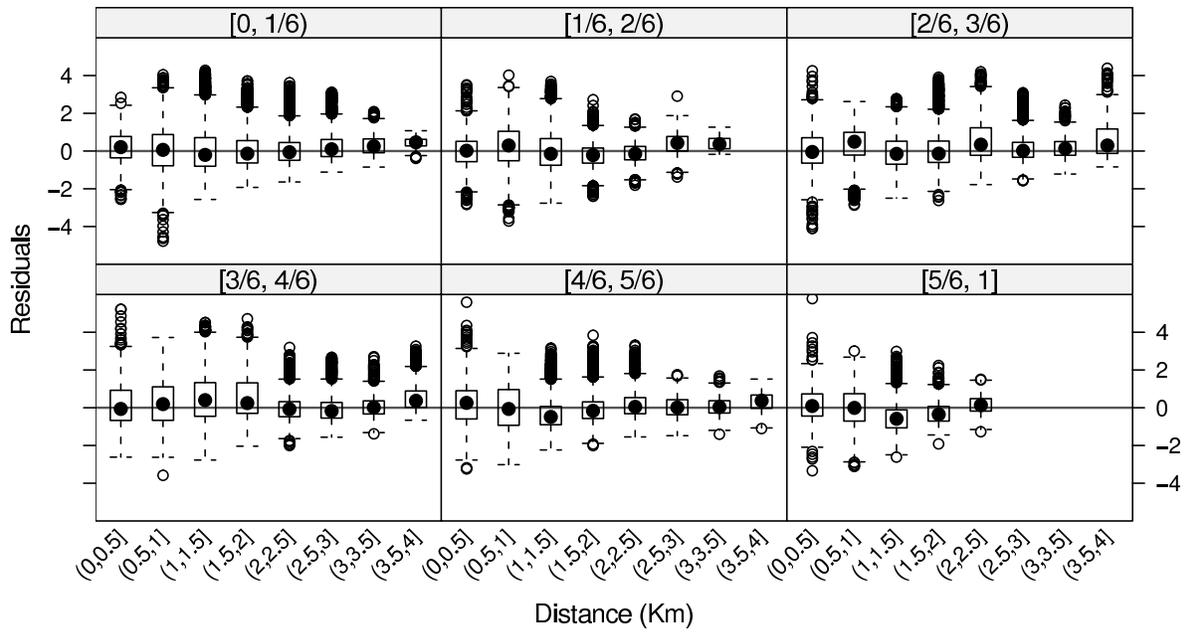


Figure 8: Residuals as a function of angular deviation $|\theta|$ from the azimuth of the antenna beam.

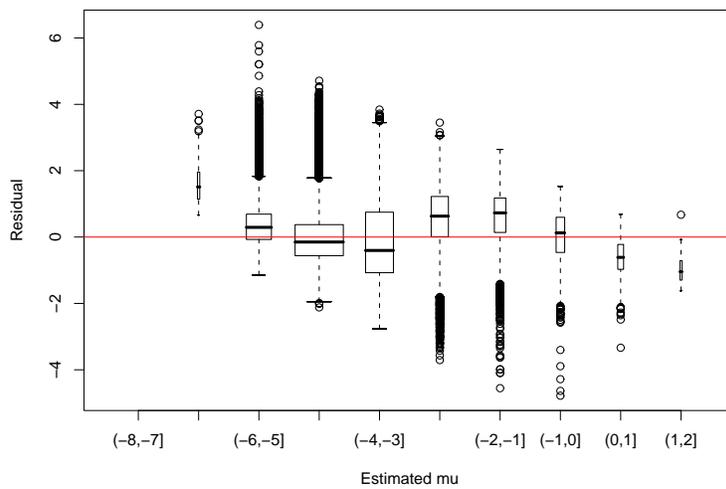


Figure 9: Residuals as a function of distance from the antenna.

Finally, Figure 10 shows the results of simulating 2.5 hours of 25 time series of relative signal strength measurements from the estimated model. The simulated E_c/I_0 's have roughly the same pattern as the curves from which the model was estimated, with sharp drops and quick rises in signal strength, but there are some differences from the raw data. In particular, the simulated signal strength curves for different antennas appear to be more highly correlated than the observed signal strengths were.

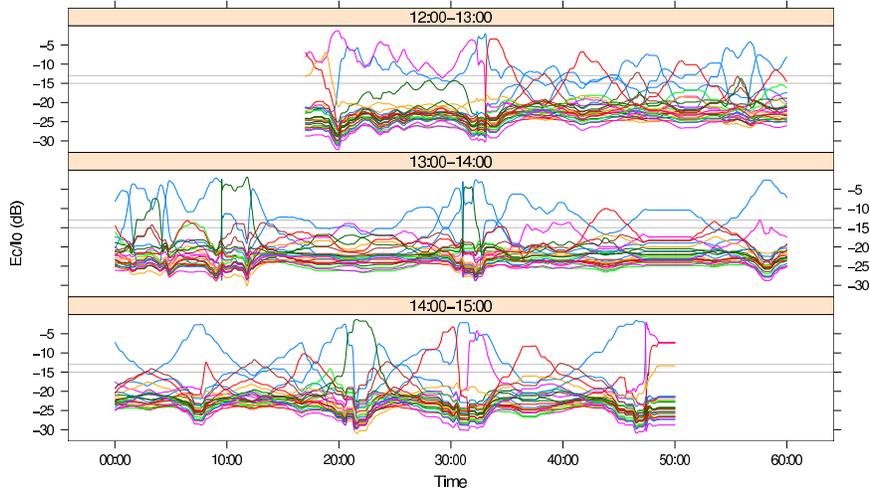


Figure 10: A set of 25 simulated signal strengths using the model estimated from the 2.5 hours of drive test data.

Overall, the residual plots and simulated data suggest that although model (1) with mean (2) is simple and tractable, it captures the main features in the data reasonably well.

8 Discussion and Open Problems

8.1 Possible Applications

First, we believe that the ability to visualize the birth and death of legs of a call gives new insights into wireless networks. Although the engineering behind the creation and destruction of legs is well-understood, graphics like our Figure 1 show that the process can be very nervous. Most graphics for wireless network consider data at a much higher level than that of the mobile; e.g., maps of signal strength. While these are valuable, looking at call details can reveal how much work a network is putting into managing calls.

Models of independent radio signals that take physical topography into account are used extensively in network design and engineering to identify coverage holes, which are areas in which all signals are

weak. Our model has a different goal. Rather than describe the strength of a signal throughout a fixed environment, we describe the relative strengths of the signals at one mobile that is not in a constant environment. In other words, getting the signal strength right over the area is not our goal; getting the time behavior at a moving location right is. A description of signals at a phone like our model provides is not needed for coarse measures like coverage holes, but it is critical for understanding how networks manage parameters like the thresholds T_{add} and T_{drop} , the size of the active and candidate sets, and the length of the neighbor list that affect call quality and call survival. A model of the ongoing behavior of signals at the phones in a network also enables adapting network parameters online; for example, providing more dynamic lists of neighboring base stations or fine-tuning transmit power as traffic densities change.

Secondly, models of the transient signals of a wireless call may make it possible to predict the state of a call over a short time frame and identify if it is about to fail because of problems on the forward link. There will always be calls that the network cannot save; for example, it may be impossible to save a call when a mobile phone enters an elevator. But there may be other calls that can be rescued by instructing the mobile to add sectors to its neighbor list, for example. A new list might contain stronger signals that could help a call in distress.

Finally, as cellular networks improve, dropped calls become rare, making binary summaries like call failure less relevant. More nuanced metrics of call quality like the fraction of time that at least one of the relative signal strengths is strong may take their place. Understanding the nature of the call signals, as well as being able to visualize them at trigger times, will then become more important.

8.2 Improvements to the Model

Signal propagation models commonly assume that signals degrade logarithmically with distance from the source. We offer two generalizations. First, the intercept and slope of our signal propagation model depend smoothly on deviation from the center of the beam, with the exact nature of this dependence left unspecified. Moreover, our model explicitly includes the many signals that the mobile receives. The fact that the model is multivariate is important for two reasons. First, the set of signals, not just one primary signal, determines the quality of the call as the mobile moves or network conditions change. Second, the signals interfere with each other and thus any one signal contributes to the noise for all the other signals received by the mobile. The model fit for the 2.5 hour drive test call discussed in this paper shows that the model is promising and warrants testing on other commercial wireless data.

Of course, any model can be improved. A better fit might be obtained by using a different model of

the mean, perhaps one that includes deviation from the azimuth in the vertical plane or one that assumes that degradation with distance is faster than logarithmic in the right tail. A Bayesian version of the model would better accommodate uncertainty about the model parameters over time, space and calls, at the cost of perhaps greatly increasing computing time and memory. Exponentially weighted moving averaging could be used in the least squares equations to accommodate changes in the Fourier models for the intercepts $\alpha_i(\theta)$ and slope $\beta(\theta)$ in the signal propagation model over time. A higher dimensional model that considered multiple calls simultaneously might be more appropriate but also substantially more time-consuming to fit. A richer imputation scheme could be used to model the signals that drop at triggers and those that are no longer detectable. A more realistic correlation structure could be defined. Before exploring these extensions, it would be important to test the current model against data for other networks and time periods. This would give some insight into which extensions are most needed. In any case, the advantage of the current model is that it is simple to explain and to update.

A much more ambitious goal would be to estimate the parameters of the model from the data that are available to the MSC that manages the call. These event-based (trigger driven) measurements miss most of the time series data that are collected in drive testing. Developing estimates from threshold crossings is generally difficult.

Additional modeling will also be needed as cellular networks become able to collect more data. Current CDMA networks require a pilot signal only at the base station, and its signal strength can be measured at the mobile to determine the quality of the forward link. We model only the forward link because only the forward signal strength is reported to the MSC. Modeling the signal energy on the forward and reverse links simultaneously would give a more complete view of wireless calls, especially in light of the fact that some call failures are due to problems on the reverse link rather than the forward link. But modeling the reverse link will require, at the least, a bivariate time series model for each leg of the call, and then a multivariate model across legs. Moreover, the marginal model for a reverse link may not resemble the model we have proposed for the forward link. Both current and future wireless networking technologies present many fascinating, challenging statistical problems. We hope that statisticians will take them on.

References

Buvaneswari, A., Ravishankar, B., Graybeal, J. M., Haner, M. and Rittenhouse, G. (2005). New optimization and management services for 3G wireless networks using CELNET Explorer, *Bell Labs Tech-*

nical Journal, 9, 101-115.

Borst, Simon C., Buvanewari, A., Drabeck, L. M., Flanagan, M. J., Hampel, G. K., Haner, M., MacDonald, W. M., Polakos, P. A., Rittenhouse, G., Saniee, I., Weiss, A., and Whiting, P. A. (2005). Dynamic optimization in future cellular networks, *Bell Labs Technical Journal*, 10, 99-119.

Jakes, William C. (1974) *Microwave Mobile Communications*, 2nd ed.. IEEE Press.

Lee, William C. Y. (1995) *Mobile Cellular Telecommunications*, 2nd ed.. McGraw-Hill.

Little, R. J. A and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. Wiley& Sons.

Patzold, Matthias (2002). *Mobile Fading Channels*. New York: Wiley&Sons.