Supplementary material for Gourdji et al., "North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion"

Supplement A - Geostatistical inversion system of equations

Minimizing the geostatistical inversion objective function in equation (1) with respect to fluxes, $\mathbf{s}$, and drift coefficients, $\boldsymbol{\beta}$, yields the following system of linear equations:

$$\begin{bmatrix} \mathbf{HQH^T + R} & \mathbf{HX} \\ (\mathbf{HX})^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^T \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{HQ} \\ \mathbf{X}^T \end{bmatrix} \quad \text{(A1)}$$

which is solved for the weights $\boldsymbol{\Lambda}$ $(m \times n)$ and Lagrange multipliers $\mathbf{M}$ $(p \times m)$ that are used to define the estimated fluxes ($\hat{\mathbf{s}}$) and their posterior covariance ($\mathbf{V}_{\hat{\mathbf{s}}}$) as:

$$\hat{\mathbf{s}} = \boldsymbol{\Lambda}\mathbf{z} \quad \text{(A2)}$$

$$\mathbf{V}_{\hat{\mathbf{s}}} = -\mathbf{XM} + \mathbf{Q} - \mathbf{QH}^T\boldsymbol{\Lambda}^T \quad \text{(A3)}$$

Estimates of the drift coefficients, $\widehat{\boldsymbol{\beta}}$, and their uncertainty covariance ($\mathbf{V}_{\widehat{\beta}}$) are obtained as:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{Q^{-1}X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1}\boldsymbol{\Lambda}\mathbf{z} \quad \text{(A4)}$$

$$\mathbf{V}_{\widehat{\beta}} = (\mathbf{X}^T\mathbf{H}^T(\mathbf{HQH}^T + \mathbf{R})^{-1}\mathbf{HX})^{-1} \quad \text{(A5)}$$

where the diagonal elements of $\mathbf{V}_{\widehat{\beta}}$ represent the uncertainties of the drift coefficients, and the off-diagonal terms in $\mathbf{V}_{\widehat{\beta}}$ represent their error covariances.

The estimated fluxes ($\hat{\mathbf{s}}$) can also be expressed in a form more similar to that used in synthesis Bayesian inversions, as the sum of a deterministic component ($\mathbf{X}\widehat{\boldsymbol{\beta}}$), i.e. the estimated model of the trend of the flux distribution, and a stochastic component that is a function of the *a priori* correlation structure in $\mathbf{Q}$:

$$\hat{\mathbf{s}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{Q}\mathbf{H}^T(\mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R})^{-1}\left(\mathbf{z} - \mathbf{H}\mathbf{X}\widehat{\boldsymbol{\beta}}\right) \quad \text{(A6)}$$

Supplement B – Atmospheric CO_2 data and boundary conditions

CO_2 mixing ratio data processing and filtering

After preliminary data filtering to exclude low-quality flags and other obvious errors or anomalous spikes in the data, continuous measurements from all towers were averaged to a three-hourly timescale for use in the inversions. Aircraft data were included for measurements below 4000m in altitude, where the strongest influence from surface CO_2 fluxes on the continent is expected (Gerbig et al., 2003a).

With a perfect transport model, it would be possible to use all available well-calibrated CO_2 measurement data in an inversion to help improve the atmospheric constraint on flux estimates. In fact, inversions using synthetic data with perfect transport yield better flux estimates with more defined spatial features as measurements are added throughout the day and night (Mueller et al., in prep.). However, including measurements during times of the day that are subject to systematic transport model errors can lead to biased flux estimates, and potentially misleading scientific conclusions (Lin and Gerbig, 2005; Prather et al., 2008).

Following previous inversion studies (e.g. Peters et al., 2007; Schuh et al., 2010; Göckede et al., 2010a), we therefore rely primarily on afternoon CO_2 measurements when vertical convective mixing is strongest, and the height of the PBL is generally well-represented by WRF-STILT (Zhao et al., 2009). Specifically, 3-hourly averages centered at 1 and 4pm were included year-round for all towers, as well as some morning and evening data (3-hourly averages centered at 10am and 7pm) during the height of the growing season when the air should also be well-mixed. Also, following data selection criteria for the CarbonTracker system (Peters et al., 2007), some night-time data (3-hourly averages centered on 1 and 4am) were included for the two tall towers (LEF and WKT, Table 1), where towers sample the residual mixed layer representative of CO_2 fluxes from the previous day.

Finally, some of the flask and aircraft data collected on the Pacific coast were excluded, given that a substantial misfit was found at these sites between measured CO_2 mixing ratios

and bottom-up fluxes transported forward to the measurement locations. This may be due to difficulties in modeling coastal transport (Riley et al., 2005), or the coarse resolution of WRF-STILT in these locations (i.e. 40km) relative to the scale of meteorological phenomena (Ahmadov et al., 2009).

In addition to systematic transport model errors, using observed continuous CO_2 measurements in a regional inversion is subject other challenges. First, instruments are subject to failure, creating gaps in the concentration record for some towers. For example, at Harvard Forest, 45% of the potential measurements are missing for the year, particularly in the early part of the year, while at Moody, Texas, observations are missing for January, August, and most of September. Other shorter data gaps occur throughout the year for all towers.

Second, some of the variability in the measurements is due to very local influences (Gerbig et al., 2009) that cannot be resolved by the transport model, the driving meteorological data, or the flux estimation grid, leading to representation errors in the inversion. These data should ideally be excluded, although there is no perfect method for identifying purely local influence. Therefore, with the exception of filtering out extreme events that appear to be plumes from nearby point sources (i.e. short-term increases >30 ppm over background air), this study did not attempt to filter the data for local variability before averaging to the 3-hourly resolution. This decision also leverages the fact that the WRF-STILT model, by simulating transport at relatively high spatiotemporal resolution in a Lagrangian framework, should be better able to resolve local influences relative to coarse-grid Eulerian models (Rastigejev et al., 2010).

Third, measurements are influenced by both land and oceanic fluxes, although oceanic flux locations were not explicitly estimated in this study. This is an issue particularly for coastal locations, where mixing with ocean air can dilute the influence of land fluxes, thereby potentially yielding misleading land flux estimates if this is not taken into account in the inversion framework. Therefore, filtering was applied to exclude measurements with primary sensitivity to ocean fluxes (defined as greater than 90% of the total sensitivity coming from ocean regions or an integrated land footprint less than 0.5 ppm/(μmol $m^{-2}s^{-1}$)). After all data

filters were applied, the number of data points included in the inversion per tower, accounting for data gaps, is shown with other tower information in Table B1.

Table B1: Measurement locations, along with other identifying characteristics of the sites and data included in the inversion. The first 9 locations in the table have continuous data, while the last two represent flask and aircraft measurements from multiple locations across the continent.

Measurement site code	Site name	Site latitude/ longitude	Altitude above ground level (m)	Night-time data included	Number of observations for year
LEF	Park Falls, Wisconsin	45.93N, 90.27W	396	1 and 4 am year-round	1500
WKT	Moody, Texas	31.32N, 97.33W	457	1 and 4 am year-round	959
SBL	Sable Island, Nova Scotia	43.93N, 60.02W	25	None	663
BRW	Barrow, Alaska	71.32N, 156.60W	10	None	248
ARM	Norman, Oklahoma	36.62N, 97.50W	60	None	879
HFM	Harvard Forest, Massachusetts	42.54N, 72.17W	30	None	558
AMT	Argyle, Maine	45.03N, 68.68W	107	None	795
FRD	Fraserdale, Ontario	49.84N, 81.52W	40	None	878
CDL	Candle Lake, Saskatchewan	53.99N, 105.12W	30	None	927
FLA	Flask samples from 6 sites	*See Figure 1*	0 to 4	None	153
AIR	Aircraft vertical profiles from 15 sites	*See Figure 1*	139 to 3999	None	943

Boundary conditions

The empirical (EMP) North American boundary condition dataset is similar to the boundary condition developed by Gerbig et al. (2003b), and nominally represents a time-varying curtain (latitude, longitude, altitude) of atmospheric CO_2 mixing ratios through the middle of the Pacific Ocean. This curtain was created by interpolating un-polluted surface and

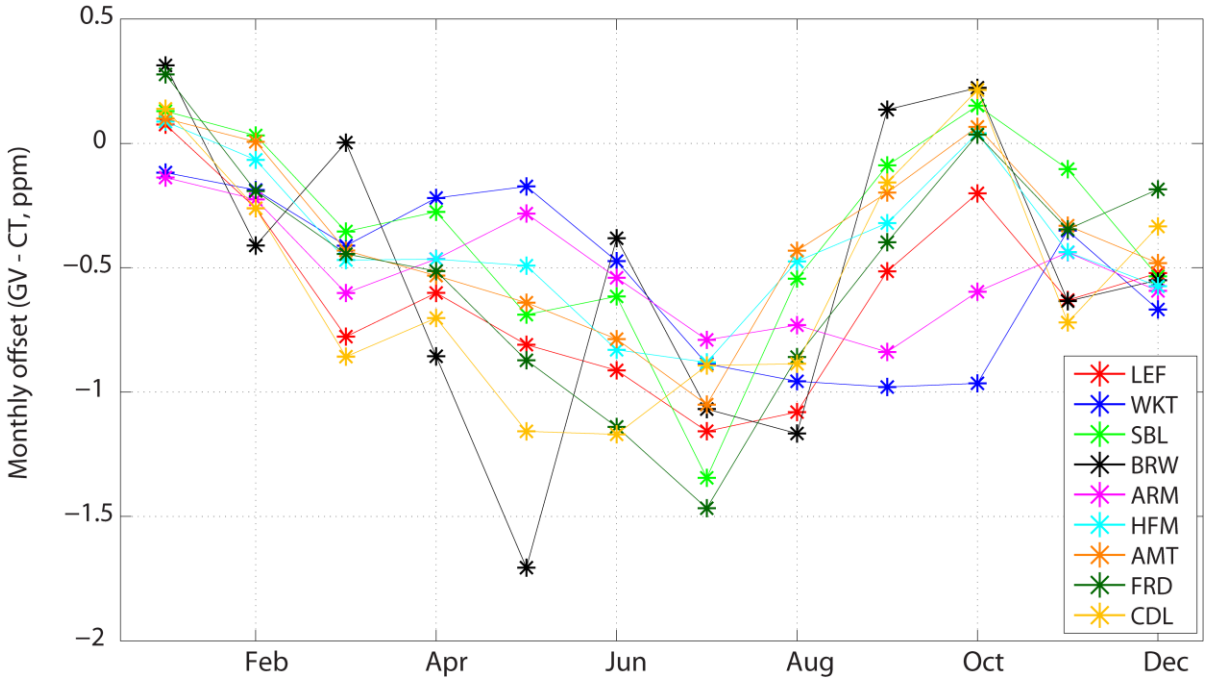
aircraft measurements of CO_2 from the NOAA-ESRL Cooperative Air Sampling Network (Tans & Conway, 2005) in both space and time, in a manner similar to the GLOBALVIEW data products (Masarie & Tans, 1995; GLOBALVIEW-CO2).

The EMP boundary condition is strictly valid only for air masses reaching the continent from the West. Relatively little air enters the study area from the East, and given that the dominant outflow from North America is to the East, un-polluted aircraft measurements of CO_2, which are representative of background air, are lacking over the Atlantic. The summer monsoonal circulation brings substantial inflow from the tropical Atlantic and Gulf of Mexico, and particularly affects observations in the Southeastern U.S. (e.g. at Moody, Texas). However, across the southern boundary, available data suggest that longitudinal gradients are less than 1 ppm. At mid- and high-latitudes, differences between Atlantic and Pacific sites are larger and seasonally varying. Future versions of the EMP boundary condition will attempt to account for Atlantic vs. Pacific differences, but knowledge of the boundary values is limited by the sparseness of available data, particularly aloft. Meridional transport by synoptic eddies in the Arctic should perhaps also be accounted for in future work (Parazoo et al., 2011).

The CarbonTracker (CT) boundary conditions represent gridded 4-dimensional (i.e. latitude, longitude, altitude and time) modeled CO_2 concentrations from the CarbonTracker data assimilation system (Peters et al., 2007, 2010a). Therefore, in contrast to the EMP dataset, the CT boundary conditions contain both synoptic and longitudinal variability. However, these modeled CO_2 concentrations are subject to biases in the inferred global fluxes resulting from the data assimilation system. In fact, the EMP boundary condition dataset was developed in response to known seasonal biases in the CT CO_2 fields, particularly during the Northern hemisphere growing season, for measurement locations upwind of North America (Peters et al., 2010b).

Due to these biases in CT, there is a systematic offset between the two boundary condition datasets, with the influence at the observational sites from the EMP dataset on average about 0.5 ppm lower than the values from CT, with this offset somewhat higher during the growing season and slightly lower for measurement locations on the East coast. (Please see Figure B1 for the average monthly offset by measurement tower.)

Figure B1: Comparison of the average monthly offsets between the EMP and CT boundary condition CO_2 time series at the nine continuous measurement towers. (Positive values imply that EMP is higher.)



Supplement C – Atmospheric transport model

STILT

The Lagrangian particle dispersion model (LPDM) employed in this work is the Stochastic Time Inverted Lagrangian Transport (STILT) model, run in the time-reversed (receptor-oriented) mode. The STILT model is based on the HYSPLIT model (Draxler and Hess, 1998) and has been developed specifically for inverse greenhouse gas flux estimates (Lin et al., 2003; Nehrkorn et al., 2010). Recent examples of its application at continental and regional scales are given by Kort et al. (2008, 2010), Zhao et al. (2009), Gourdji et al. (2010), and Göckede et al. (2010a). As in all LPDMs, atmospheric dispersion in STILT is simulated by tracking a large set of tracer particles, with each particle transported by mean winds obtained from a meteorological model plus an unresolved, turbulent (subgrid) velocity component computed using the

parameterizations of Hanna (1982). The inclusion of both the mean and stochastic wind components (whose interactions are the basic cause of dispersion in the atmosphere) sets the LPDM approach apart from conventional trajectory models that employ mean winds only, and thus cannot properly simulate dispersion or surface interactions (Stohl, 1998; Stohl et al., 2003; Uliasz and Pielke, 1990; Uliasz, 1994). The LPDM approach prevents particle tracks from intersecting the surface for numerical reasons (e.g., a strong descent forcing a particle to a level below the surface) as is common in trajectory models. Instead, the turbulent winds invoked by the LPDM ensure that particles travel through the planetary boundary layer (PBL) in a physically reasonable way.

In this application, the STILT model transports ensembles of 500 particles backwards in time 10 days from each of a set of receptor points (in this case, from a set of towers at 3-hourly resolution) using the input meteorological data. For each receptor, we calculate the response of the target gas concentration at the receptor point to surface sources ("footprint"), in units of ppmv/(μmol m^{-2} s^{-1}). The footprint, which represents the adjoint of the transport field, is calculated by counting the number of particles in a surface-influenced region (defined as a fraction of the estimated PBL height) and the time spent in the region (for details, see Lin et al., (2003)). When multiplied by an *a priori* field of surface flux, the footprint gives the associated contribution to the mixing ratio measured at the receptor.

The STILT model has undergone a number of upgrades, most recently a merger with parts of the latest HYSPLIT code, resulting in improved handling of nested fields. The model is being developed by a worldwide consortium, with the most up-to-date software and information available at www.stilt-model.org.

Meteorological Input

For the inversions described in this paper, the STILT model has been driven by meteorological fields from a version of the WRF model (Skamarock and Klemp, 2008) customized for STILT and other transport models (Nehrkorn et al., 2010). In particular, these WRF outputs include convective mass fluxes that are used directly in the STILT dispersion calculations, and time-averaged mass fluxes (rather than instantaneous advective velocities) are

used to drive STILT, which results in good mass conservation (a critical consideration for inverse flux estimates). We have employed version 2.2 of WRF (www.wrf-model.org) over a domain covering North America, with two nested 10- and 40-km grids, as shown in **Error! Reference source not found.**. NARR fields have been used for initial and lateral boundary conditions and for analysis nudging of WRF. To prevent drift of the WRF forecasts from the analyses, forecasts were reinitialized every 24 hours (at 00 UTC). Forecasts were run out to 30 hours, but only hours 7-30 from each forecast were used to avoid spin-up effects during the first 6 hours from each forecast. Model fields were output hourly and archived in the native WRF netcdf format. For use in the STILT model, a subset of model fields was archived in the (compressed) ARL format, at a substantial savings in storage.

A summary of numeric and physics options used in these runs is given in Table C1.

Table C1: Numeric and physics options for WRF runs.

Option	**Description**
Land-surface	Noah land-surface model with Monin-Obukhov surface layer (Ek et al., 2003)
PBL package	Yonsei University (YSU) scheme (Hong et al., 2006)
LW radiation	RRTM (Mlawer et al., 1997)
SW radiation	Goddard (Chou and Suarez, 1994)
Microphysics	Lin et al. (1983); Chen and Sun (2002)
Convection	Grell & Devenyi (2002)
Nesting	One-way
Nudging	u,v,T,q at all levels above PBL, every 3 hours, 1 hour relaxation time
Time stepping	3rd order Runge-Kutta; 4 short time steps per long time step
Advection	5th order horizontal, 3rd order vertical positive definite advection for moisture and scalars
Diffusion	2nd order horizontal diffusion using Smagorinsky first-order closure
Damping	No upper level or vertical velocity damping; default values for divergence and external model damping

Supplement D – Spatial covariance parameters

The monthly flux covariance parameters in $\mathbf{Q}$ optimized using the RML method with the atmospheric data provide insights into the underlying variability of the flux field, and how this variability changes throughout the year. The inferred monthly model-data mismatch parameters in $\mathbf{R}$ are representative of the ability of the inversion to take advantage of surface flux information contained in the measured CO_2 mixing ratios.

Figure D1a shows inferred monthly spatial covariance parameters in $\mathbf{Q}$ from the Simple inversion throughout the year. For this inversion, these inferred parameters represent the

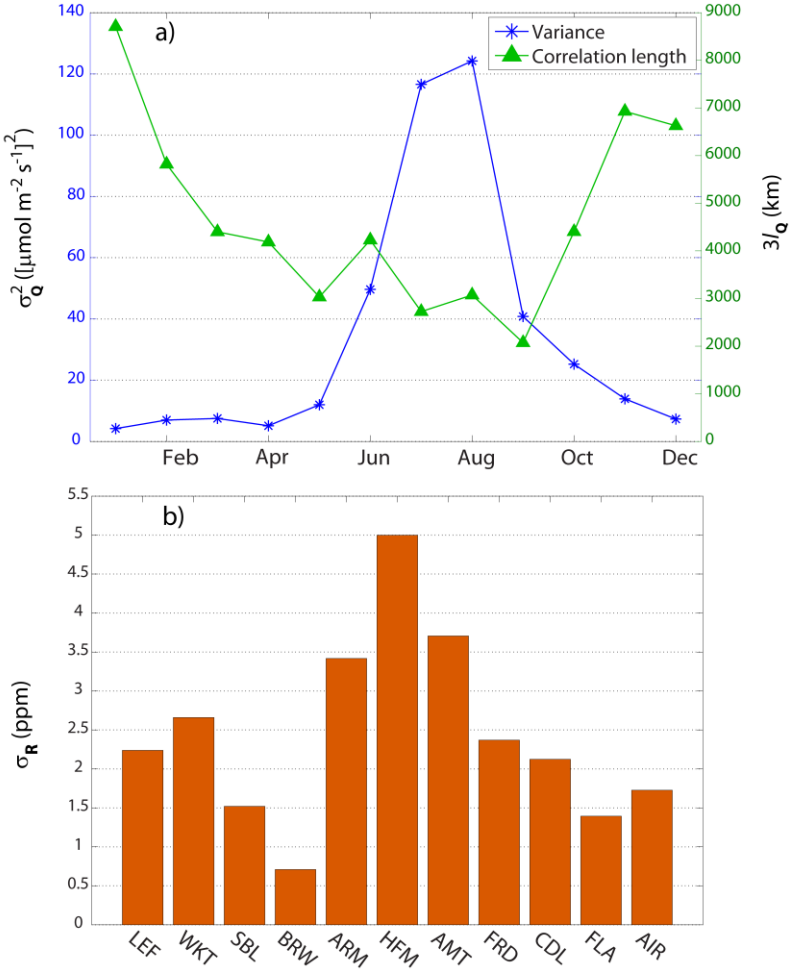
spatial variability of the total biospheric flux, and their seasonality is consistent with that observed in biospheric model output (Huntzinger et al., 2010), with the most variable fluxes in July and August (highest variance and shortest correlation length), and the least variable in the dormant season from November through April (lowest variance and longest correlation length). The inferred correlation lengths range from approximately 2000 km in September to approximately 8500 km in January, with the length scales during the growing season relatively consistent with those from biospheric models (Huntzinger et al., 2010). In the dormant season, the longer correlation lengths from the inversions may potentially be due to higher wind speeds which smear the flux signal from the atmospheric measurements across the continent (Elliott et al., 1986; http://www.windatlas.ca/en/index.php), making the statistical characteristics of the underlying fluxes more difficult to estimate.

In the NARR inversion, inferred flux variances and correlation lengths (results not shown) are reduced relative to the Simple inversion, by about 13 and 33% respectively. The reduction in the variance parameter shows that the NARR variables are able to explain some, but not all, of the variability in the flux distribution. The reduction in the inferred correlation lengths of the flux residuals implies that the NARR variables are explaining large-scale spatial patterns in the inferred fluxes.

The square root of the average monthly model-data mismatch variances (i.e. $\sigma_{\boldsymbol{R}}$) is shown in Figure D1b for each of the 9 measurement locations and flask and aircraft data, where the yearly average for a given measurement site is weighted by the number of data points included per month. The tower with the highest average model-data mismatch is Harvard Forest. This tower is sited in a highly-productive forested area (Urbanski et al., 2007) about 100 km west of Boston, and even closer to Worcester and Springfield, Massachusetts, while the 1°x1° gridcell containing this site includes several other small towns and developed areas. The difficulty in matching the data at this tower is likely due to spatial aggregation and representation errors associated with nearby fossil fuel emissions and the heterogeneous land cover surrounding the site. The two towers with the lowest model-data mismatch are SBL and BRW, the two coastal sites sampling relatively well-mixed air and/ or low flux variability areas. Similarly, the flask measurements, collected in either MBL or high-altitude locations, have low

model-data mismatch. Interestingly, FRD and CDL, which are short towers in boreal forest regions, have lower model-data mismatch than the three other short towers. This could be due to more homogeneous land cover and low fossil fuel influence in these areas. Also, given their location at higher latitudes where the 1° longitudinal grid-cell size is smaller (see Figure 1), this could point to potential improvements in inversion performance that could be obtained by resolving fluxes at finer spatial scales, perhaps the resolution of the driving winds in the transport model (in this case 40-km or better for all North America).

Figure D1, a) & b): Optimized covariance parameters using the RML algorithm with the atmospheric measurements, for the Simple inversion with GlobalView boundary conditions. a) Monthly flux covariance parameters, i.e. the sill variance (σ^2_Q) and correlation length parameter ($3l$). b) Square root of a weighted average of the monthly model-data mismatch variances (σ_R), (with the weights defined by the number of data-points in each month), for 9 towers, flask and aircraft data. Location information for each measurement code is included in Table B1.



Model-data mismatch can also have significant seasonal variations (results not shown). For example, the optimized model-data mismatch standard deviation at ARM is 1.1 ppm in February, 6.1 ppm in April and 2.4 ppm in July. The high model-data mismatch in April for ARM implies that the strength of local uptake visible in the measurement data in this month, perhaps due to the spring wheat crop planted in the near-vicinity of the tower, is being partially ignored by the inversion. Again, a finer spatial flux resolution could potentially help to more appropriately use the information contained in this data in future inversions.

Supplement E – Branch and Bound variable selection method

The BIC is a criterion-based variable selection method, allowing for the comparison of all possible combinations of a superset of auxiliary variables (Ward, 2008). The original BIC equations from Schwarz (1978) were modified for a geostatistical setup with correlated residuals by Mueller et al., (2010), and were further updated here to be compatible with an inverse modeling formulation using the atmospheric measurements. After these modifications, the criterion that must be minimized for the BIC approach can be expressed as:

$$BIC = \ln|\mathbf{\Psi}| + [\mathbf{z}^{\mathrm{T}}(\mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{HX}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{HX})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathbf{T}}\mathbf{\Psi}^{-1})\mathbf{z}] + p\ln(n) \quad \text{(E1)}$$

where $\mathbf{\Psi} = \mathbf{HQH^T} + \mathbf{R}$, *n* is the number of observations, *p* represents the number of covariates included within a given model ($\boldsymbol{X}$), and all other variables are as described in Section 2.2. Given that comparing all possible models (2^p) quickly becomes computationally expensive and perhaps infeasible for large supersets of variables, the BIC was implemented with a Branch-and-Bound algorithm (Land and Doig, 1960) to help make the problem computationally tractable. This algorithm avoids unnecessary matrix multiplications by eliminating model "branches" as it runs that cannot possibly contain the "best" model (Yadav et al., in prep).