[American Journal of Science, Vol. 316, October, 2016, P. 746–777, DOI 10.2475/08.2016.02]

# A DATA-DRIVEN SPLINE MODEL DESIGNED TO PREDICT PALEOCLIMATE USING PALEOSOL GEOCHEMISTRY

GARY E. STINCHCOMB*,**,†, LEE C. NORDT**, STEVEN G. DRIESE**, WILLIAM E. LUKENS**, FORREST C. WILLIAMSON***, and JACK D. TUBBS***

**ABSTRACT. Paleosols (fossil soils) are abundant in the sedimentary record and reflect, at least in part, regional paleoclimate. Paleopedology thus offers a great potential for elucidating high resolution, deep-time paleoclimate records. However, many fossil soils did not equilibrate with climate prior to burial and instead dominantly express physical and chemical features reflective of other soil forming factors. Current models that use elemental oxides for climate reconstruction bypass the issue of soil-climate equilibration by restricting datasets to narrow ranges of soil properties, soil-forming environments and mean annual precipitation (MAP) and mean annual temperature (MAT). Here we evaluate a data-driven paleosol-paleoclimate model ($PPM_{1.0}$) that uses subsoil geochemistry to test the ability of soils from wide-ranging environments to predict MAP and MAT as a joint response with few initial assumptions.**

**The $PPM_{1.0}$ was developed using a combined partial least squares regression (PLSR) and a nonlinear spline on 685 mineral soil B horizons currently forming under MAP ranging from 130 to 6900 mm and MAT ranging from 0 to 27 °C. The PLSR results on 11 major and minor oxides show that four linear combinations of these oxides (Regressors 1-4), akin to classic oxide ratios, have potential for predicting climate. Regressor 1 correlates with increasing MAP and MAT through Fe oxidation, desilication, base loss and residual enrichment. Regressor 2 correlates with MAT through temperature-dependent dissolution of Na- and K-bearing minerals. Regressor 3 correlates with increasing MAP through decalcification and retention of Si. Regressor 4 correlates with increasing MAP through Mg retention in mafic-rich parent material. The nonlinear spline model fit on Regressors 1 to 4 results in a Root Mean Squared Error ($RMSE_{MAP}$) of 228 mm and $RMSE_{MAT}$ of 2.46 °C. $PPM_{1.0}$ model simulations result in Root Mean Squared Predictive Error ($RMSPE_{MAP}$) of 512 mm and $RMSPE_{MAT}$ of 3.98 °C. The RMSE values are lower than some preexisting MAT models and show that subsoil weathering processes operating under a wide range of soil forming factors possess climate prediction potential, which agrees with the state-factor model of soil formation. The nonlinear, multivariate model space of $PPM_{1.0}$ more accurately reflects the complex and nonlinear nature of many weathering processes as climate varies. This approach is still limited as it was built using data primarily from the conterminous USA and does not account for effects of diagenesis. Yet, because it is calibrated over a broader range of climatic variable space than previous work, it should have the widest array of potential applications. Furthermore, because it is not dependent on properties that may be poorly preserved in buried paleosols, the $PPM_{1.0}$ model is preferable for reconstructing deep time climate transitions. In fact, previous studies may have grossly underestimated paleo-MAP for some paleosols.**

Key words: weathering model, climosequence, soil geochemistry, Partial Least Squares

## INTRODUCTION

Decadal- to centennial-scale climate change is occurring within the context of longer-term millennial-, orbital- and tectonic-scale climate processes. Yet, little is

* Watershed Studies Institute & Department of Geosciences, Murray State University, Murray, Kentucky 42071
** Terrestrial Paleoclimatology Research Group, Department of Geosciences, Baylor University, Waco, Texas 76798
*** Department of Statistical Science, Baylor University, Waco, Texas 76798
† Corresponding author: gstinchcomb@murraystate.edu; phone: 270-809-6761

known about the interaction between short- and long-term climate forcings and associated responses (National Research Council, 2011). The deep-time terrestrial record is a reservoir of historical transitions in short- and long-term climate and paleoclimate research requires new mineral proxies that will improve the precision and accuracy of detecting and measuring such changes (*ibidem*).

Paleosols, lithified or buried soils that document a past landscape, are one means of modeling paleoclimate (Retallack, 2001; Sheldon and Tabor, 2009). Paleosols are important paleoclimate archives because they are (1) abundant in the rock record (Retallack, 1997, 2001), (2) found in a wide variety of environments, reflecting regional climate variations (Yaalon, 1971; Birkeland, 1999; Bradley, 1999; Retallack, 2001; Holliday, 2004; Sheldon and Tabor, 2009; Schaetzl and Thompson, 2015), and (3) records of "Critical Zones" of the Earth's past, preserving the *in situ* substrate of ancient landscapes (Amundson and others, 2007; Nordt and Driese, 2013). The conceptual framework for using paleosol geochemistry as an environmental and climate proxy is based on the fact that (1) the proportion of chemical elements within a soil are determined by the starting parent material composition and subsequent modification due to pedogenic processes, and (2) these processes, in turn, evolve as the soil forms through time within an open system that is largely dictated by climate, organisms, relief, parent material, and the duration of pedogenesis (Jenny, 1941).

Paleopedologists have tried to minimize uncertainty in paleoclimate model development using a variable-selection approach, where specific geochemical oxides were selected based on fundamental weathering principles (for example, base loss) or the soils examined were limited to a particular suite of characteristics (for example, soil order and horizon type) and/or soil-forming factor constraints (Retallack, 1994; Sheldon and others, 2002; Nordt and Driese, 2010a; Oskarsson and others, 2012). Models derived from variable-selection can lead to misapplication, where incorrect assumptions are made regarding the paleosol being measured. This limitation is perhaps the most understated and most difficult problem faced by pedologists and geochemists working with modern and ancient soils. This problem leads us to pose two important research questions relevant to the paleopedological and terrestrial paleoclimate communities: (1) how does pedogenesis vary with respect to climate, given few constraints on soil forming factors, and (2) how well can a comprehensive mineral soil model predict climate that requires few assumptions *a priori*?

This study addresses these questions by using a combined partial least squares (PLS) regression and semiparametric spline modeling approach on 685 mineral soil B horizons from a wide range of soil forming environments, currently weathering under MAP ranging from 130 to 6900 mm and MAT ranging from 0 to 27 °C. The results of the combined PLSR-spline technique are used to model climate for soils that may or may not have achieved steady state with respect to the current climate. We present the first version of a paleosol-paleoclimate model ($PPM_{1.0}$) that accounts for uncertainty in paleosol forming factors, unknown steady-state conditions and other pedogenic processes difficult to account for in the stratigraphic record.

## BACKGROUND

The conceptual framework of most paleosol-based paleoclimate studies relies upon the fundamental state-factor theory for modern soils (Jenny, 1941, 1980):

$$S \text{ or } s = f(cl, o, r, p, t, \ldots), \quad (1)$$

where $S$ is the soil and $s$ is any soil property and $cl$ is climate, $o$ is organisms, $r$ is relief, $p$ is parent material, and $t$ is time. The ellipsis represents other possible unaccounted for factors influencing $S$. Jenny (1980) then showed that $s$ could be studied as it varies in relation to a single factor by holding other factors constant, expressed mathematically as the factor-function approach:

$$s = f(cl, u)_{o, r, p, t}. \quad . \quad . \quad . \quad (2)$$

In this example, $s$ only varies as a function of $cl$, while $o$, $r$, $p$ and $t$ are held constant, termed a *climofunction*. A *climosequence* is when this mathematical expression is applied using a suite of soils that only vary based on climate. The term $u$ is a disturbance or uncertainty term that was added to recognize that in modern-day soil sequences the factors assumed to be held constant, do indeed vary (Jenny, 1980 and references therein). Despite this error, the sequence approach (for example, climo-, bio-, topo-, litho- and chronosequence) has proven to be a valuable soil geomorphic tool and is widely used to assess how soil properties vary with respect to environmental factors (Jenny, 1941, 1980; Birkeland, 1999; Schaetzl and Thompson, 2015).

For paleosols, it is assumed that equation 1 can be applied in a similar manner throughout much of geologic time (principle of uniformitarianism). Thus modern climosequence relationships that relate $s$ to $cl$ (eq. 2) are applied to paleosols using an environmental-factor approach (*sensu* Retallack, 1994), where $s$ is measured in the paleosol using a variety of bulk and isotope geochemical, (magnetic) mineral, and physical proxies (Maynard, 1993; Maher and Thompson, 1995; Maher, 1998; Ludvigson and others, 1998, 2013; Sheldon and others, 2002; Retallack 2005; Dworkin and others, 2005; Nordt and others, 2006; Sheldon, 2006; Sheldon and Tabor, 2009 and references therein; Nordt and Driese, 2010a, 2010b; Passey and others, 2010; Retallack and Huang, 2010; Gulbranson and others, 2011; Gallagher and Sheldon, 2013; Hyland and others, 2015). As with all paleoclimatic signals from proxy data, the paleosol-climate signal is entangled with "noise" from the effects of non-climatic factors (Bradley, 1999). In the case of soils and paleosols, this "noise" is from changing $o$, $r$, $p$, and $t$ as represented by $u$ (eq. 2). Additionally, paleosols may have experienced diagenesis.

One approach is to pre-filter a dataset before building a model and acknowledge model limitations. This approach, either explicitly or implicitly, presents a set of assumptions that must be met prior to using the model. For example, the model cannot be used on laterites (CIA-K, Sheldon and others, 2002); or the model is constrained to forested soils (Gallagher and Sheldon, 2013) or calcareous soils (Retallack, 2005), and so on. To our knowledge no one has tried an approach where all soils, regardless of conditions, are included to develop the model. Thus, this "noise", or uncertainty, is incorporated into the model, which will theoretically provide some first-order assessment of climate when little is known about the target paleosol.

Paleosol-based climate models have limitations. Proxies for mean annual precipitation (MAP) that rely on the bulk geochemistry of B horizons (CIA-K, Sheldon and others, 2002; CALMAG, Nordt and Driese, 2010a) are limited to specific soil orders and MAP ranges between ~200 to 1600 $yr^{-1}$. This MAP interval is bounded asymptotically by the accumulation and loss of base cations relative to alumina at low and high MAP, respectively (Sheldon and others, 2002). Similarly, bulk geochemical proxies for mean annual temperature (MAT) are either limited by soil order (for example, PWI of Gallagher and Sheldon, 2013) or temperature range with high standard error (2–22 °C MAT, ± 4.4, salinization ratio, Sheldon and others, 2002, see Sheldon and Tabor, 2009 for error). By limiting application of a proxy to specific soil orders, one must make *a priori* assumptions that may not be valid, including potential mistakes in paleosol classification or the range of paleosol formation times over which the proxies may be applied may not be valid.

Single-mineral paleoclimate proxies, such as the $\delta^{18}O$ of pedogenic carbonates (Cerling, 1984; Dworkin and others, 2005) or pedogenic siderites (Ludvigson and others 1998; Fernandez and others, 2014), formed in specific geologic and climatic settings, have strong seasonal biases in time of formation (Breecker and others, 2009; Fernandez and others, 2014) and have the potential to recrystallize during diagenesis.

One recent attempt to reconcile this issue in paleosol climate proxies is the goethite-hematite ratio (G/H) of Hyland and others (2015), which relates increasing abundance of goethite to increasing MAP across a wide range of soils and MAP values (100–3300 mm $yr^{-1}$) with relatively low error. However, for paleosols that have experienced more than modest burial, the G/H ratio may be reset by diagenesis, thus limiting the applications of the approach to Cenozoic paleosols, or to cases where the basin had experienced minimal burial. Concern has also been raised regarding the difficulty of measuring the true concentration of pedogenic goethite using the methods of Hyland and others (2015) (see Maxbauer and others, 2016).

Many paleosols occur in depositional environments where soil formation and burial are occurring episodically on $10^1$ to $10^3$ year cycles (Atchley and others, 2013); thus, the absence of sufficient formation time should be a concern for those wishing to predict paleoclimate using paleosols. Of particular note are Cambic horizons (Bw), which are weakly-developed subsoil horizons commonly observed in alluvial basins (Retallack, 2001; Atchley and others, 2013). Many Cambic horizons are not likely to have reached more than a quasi-steady state with a long-term environmental factor like climate. Paleosol profiles that experienced cumulization and have no apparent horizonation are also problematic, as they represent aggradation rates either matching or outpacing top-down weathering. The inverse relationship between sedimentation rate and pedogenic maturity has been used to refine stratigraphic age models by allocating appropriate time intervals to paleosol successions in alluvial basins (Kraus, 1999). Thus, the link between soil maturity and sedimentation suggests that at least some of the paleosol properties are dominated by the duration of pedogenesis at the outset and may not be at steady state with environmental factors such as climate. Even if careful paleopedologists note that a paleosol reflects the average climate during the duration of its development; the duration may be too short and the geochemical proxy used to infer paleoclimate may not be an accurate reflection of that climate (fig. 1).

The preceding overview demonstrates that there is an irreducible uncertainty when using a paleosol to predict paleoclimate, akin to the error, $u$, noted by Jenny in modern soils (1980). To overcome these issues, we begin with a conceptual framework that (1) the knowledge of paleosol steady state, with respect to paleoclimate, is unknown and (2) error ($u$) like that observed in modern soils related to non-constant environmental factors, $o$, $r$, $p$ and $t$, also affects paleosol properties ($s$). This framework suggests that a diverse range of modern soils with no constraints on steady-state development or environmental factors may show no correlation with climate, and that a mineral-soil model built using a diverse range of soils that predicts climate will perform poorly.

## MATERIAL AND METHODS

A data-driven paleosol-paleoclimate model ($PPM_{1.0}$) was developed that uses subsoil geochemistry to predict MAP and MAT as a joint response. The URL for accessing the $PPM_{1.0}$ model code, example data and online tutorial can be found here: http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb/, and in Appendix A.

### *Data Compilation*

Modern surface soils and state-factor data were compiled and adapted from five online databases and previous studies (table 1) to develop a paleoclimate model and to characterize the range of environmental factors for soils in the dataset. Bulk soil-geochemical and soil-characterization data (s) were acquired from the (1) Natural Resources Conservation Service (NRCS) online soil geochemical database (Burt and others, 2003; Wilson and others, 2008) and the (2) NRCS National Soils Information System (NASIS) online database. The NRCS soil geochemical data were originally

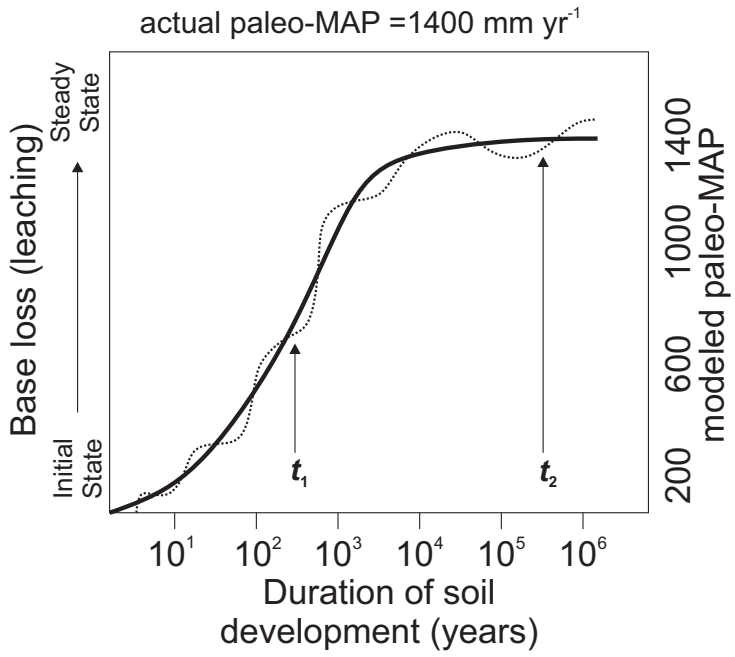


Fig. 1. Conceptual diagram showing one example of the limitations of using soil properties like standard bulk geochemical methods to model paleo-MAP in a paleosol. This figure is a modification of Yaalon's (1971) figure depicting how a soil property (in this case, base loss) approaches steady state over time. This hypothetical example considers a paleosol that weathered under a paleo-MAP of 1400 mm $yr^{-1}$. With the influx of precipitation to the soil system, transformations such as hydrolysis will occur and result in base-cation loss via leaching. As hydrolysis and leaching remove base cations from the soil during the initial 100 to 1000 years of soil development (for example, $t_1$), the base loss and modeled paleo-MAP increase; yet, the actual MAP remains unchanged. It is not until $\sim \geq 10^4$ years (for example, $t_2$) that the modeled paleo-MAP inferred from base loss reflects the actual value, 1400 mm $yr^{-1}$. This simplified scenario does not consider complicating factors such as climate change, post-burial diagenesis, and other environmental changes during pedogenesis. It is probable that even careful paleopedologists work in depositional environments where the paleosol may be buried and shut off from pedogenesis in $10^3$ years (Atchley and others, 2013), recording a base loss that is not steady-state with MAP. With enough passage of time, the soil may reach steady state with respect to MAP.

collected to assess elemental concentrations with respect to anthropogenic additions, evaluate elemental source, and relate bulk geochemistry to other soil properties (Burt and others, 2003). Unlike Burt and others (2003), this study examines the uppermost B horizons from a sample of 685 pedons queried from the NRCS online soil geochemical database (table 1). The uppermost B horizon was chosen because (1) A horizons are preserved less in the rock record (Tabor and Myers, 2015) and (2) the geochemistry reflects pedogenic processes more so than underlying horizons more affected by parent material.

Climate data (*cl*) associated with each pedon were determined from the Parameter Regression Independent Slope Model (PRISM) (Daly and others, 1994) in ArcMap10 using the 1971 to 2000 observation period (table 1), which has an output spatial resolution of ~4 km. Land cover, which serves as a proxy for vegetation (*o*), was retrieved for each pedon from the United States Geological Survey (USGS) National Land Cover Dataset (NLCD) of 2001 (Homer and others, 2007). Pedon elevation and slope (*r*) were determined from both the National Elevation Dataset (NED) 10-m resolution digital elevation models, available from the U. S. Geological Survey, and NASIS site descriptive data. Parent material (*p*) type for each pedon was obtained using

TABLE 1

*Data used to build the* $PPM_{1.0}$.

| Data type | Variable of interest | Link | Source |
|---|---|---|---|
| Whole-soil (bulk) geochemistry | *s* | http://www.nrcs.usda.gov/wps/portal/nrcs/detail/wv/soils/?cid=nrcs142p2_053632 | Burt and others, 2003; Wilson and others, 2008 |
| Soil characterization data | *s* | http://websoilsurvey.nrcs.usda.gov/ | NRCS website |
| Climate data | *cl* | http://www.prism.oregonstate.edu/normals/ | Daly and others, 1994 |
| Land cover | *o* | http://landcover.usgs.gov/ | Homer and others, 2007 |
| Elevation and slope | *r* | http://ned.usgs.gov/ | USGS NED website |
| Parent material | *p* | http://websoilsurvey.nrcs.usda.gov/ | NRCS and NASIS websites |
| Duration of soil formation | *t* | See publications; http://websoilsurvey.nrcs.usda.gov/ | Pavich and others, 1985; Muhs and others, 2001; Bacon and others, 2012; NRCS and NASIS websites |

the NASIS online database. Variations in the duration of soil formation (*t*) were inferred through soil series descriptions and previous work (Pavich and others, 1985; Muhs and others, 2001; Bacon and others, 2012).

Eleven geochemical oxides ($Al_2O_3$, $ZrO_2$, $TiO_2$, $Fe_2O_3$, $P_2O_5$, MnO, CaO, MgO, $Na_2O$, $K_2O$, and $SiO_2$) were selected from the NRCS Soil Geochemistry Spatial Dataset (Burt and others, 2003; Wilson and others, 2008) as potential explanatory variables for this study. Exchangeable element concentrations were not included because they are no longer extractable after consolidation and rock formation.

The bulk oxides were measured on the fine-earth fraction (<2 mm) using bulk samples by pedogenic horizon. Acid digestion was used to dissolve the samples and the analytes were measured using ICP-AES (Soil Survey Staff, 2014). The eleven oxides, which included all major and some minor oxides with the exception of C and S, were chosen simply on the basis of being the most available data within the NRCS dataset. Removing geochemical data *a priori* would have contradicted the assumption-limited approach of this study. Calcium oxide (CaO) was log-transformed using $\log_e$, ln(CaO), to create a more symmetrical distribution similar to the distributions of the other oxides so that the normality assumptions required for the partial least squares regression approach (below) would be met.

### *Data Analysis*

A partial least squares regression (PLSR) for this paper was generated using SAS software, Version 9.2 of the SAS system for PC (SAS Institute, Inc., 2008), on the geochemical data of the 685 uppermost B horizons, to reduce the dimensions of the eleven explanatory variables, while at the same time seeking to maximize correlations

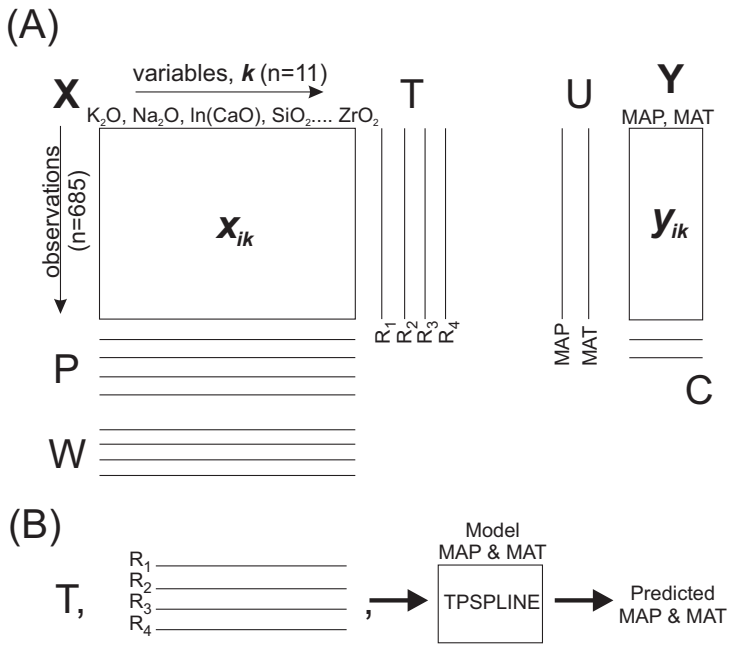


Fig. 2. Illustration of data matrices in the PLSR process (after Wold and others, 2001) and thin-plate spline, TPSPLINE, model development. See text for variable definitions and equations used to relate the descriptor matrix, **X** to the response matrix, **Y**. (A) For this study, **X** consists of 685 modern pedon observations each with 11 geochemical oxide variables. For each pedon observation, there is a corresponding MAP and MAT observation in **Y**. The PLSR was used in this study to reduce the dimensions of the data and remove collinearity between descriptors while at the same time maximizing covariance with the combined response **Y**. This is accomplished using an algorithm that creates transposed loadings on the **X** matrix, **P** and **W**, and the **Y** matrix, **C**. The resulting matrix, **T**, consists of four regressors determined using equation 6: $\mathbf{Y} = \mathbf{TC} + \mathbf{F}$, where **C** are the loadings determined so that the residual error, **F**, between observed and predicted climate, is minimized. (B) The **T** is input into a TPSPLINE model to predict MAP and MAT.

with the joint response variables, mean annual precipitation (MAP) and mean annual temperature (MAT). The PLSR procedure extracts successive linear combinations of the observable variables (that is, eleven oxides), resulting in latent, or "hidden", variables that explain response (MAP, MAT) variation (Wold and others, 2001), which is an important advantage over other dimension-reduction techniques (for example, Principal Components Analysis). These linear combinations are referred to herein as "regressors" and are estimations of the latent variables and/or their rotations in a projected space, reducing the dimensions of the data (fig. 2). The regressors generated by PLSR are analogous to the geochemical weathering ratios that correlate with climate.

In this study, we have a predictor matrix, ***X***, composed of 11 geochemical oxides from 685 B horizons (size = 11 × 685). The regressors are estimated using the following formula (expressed in matrix form):

$$\mathbf{T} = \mathbf{XW}^*, \tag{3}$$

where **T** is the regressors, **X** is the matrix of 11 geochemical oxides (descriptor variables), and **W** is the matrix of transformed PLSR weights (**W***) of the geochemical oxides. The **T** values are then multiplied by loadings, **P**, so that residuals, **E**, are minimized:

$$\mathbf{X} = \mathbf{TP} + \mathbf{E}. \quad (4)$$

The response matrix in this study, **Y**, is composed of two response variables **U** (MAP and MAT), that are multiplied by loadings, **C**, so that residuals, **G**, are minimized:

$$\mathbf{Y} = \mathbf{UC} + \mathbf{G}. \quad (5)$$

The regressors are determined using the following equation:

$$\mathbf{Y} = \mathbf{TC} + \mathbf{F}, \quad (6)$$

where **T** is multiplied by **C**, so that the difference between the observed and modeled climate response (that is, residual error), **F**, is minimized (fig. 2). PLS also weights the response variables to improve correlations between the **X** and **Y** matrices. Cross-validation and the Predicted Residual Sum of Squares (PRESS) statistic were used to choose the number of regressors and to prevent over-fitting (Allen, 1974).

We used the Pearson product-moment correlation coefficient (Pearson's *r*) to describe associations among regressors, oxides, characterization data and climate. Although PLSR and Pearson's *r* only measure linear correlation between variables, the paleoclimate model (described below) was built using semiparametric regression, which has a nonlinear component. Thus, the strength of linear correlation between *r* and regressors, oxides, characterization data and climate does not affect the final semiparametric model. Lastly, it's worth stressing that there is no absolute cutoff that allows us to define Pearson's *r* values as "weakly correlated", "moderately correlated", and so on. Although there is a significance test for Pearson's *r*, we chose not to run this here because it detracts from a primary goal of this study – to introduce a new multivariate, nonlinear spline model for predicting paleoclimate using paleosol geochemistry.

### *Model Development and Validation*

Climate (MAP and MAT) was modeled as a joint response on the PLSR regressors using a semiparametric thin-plate spline (TPSPLINE) model generated using SAS software, Version 9.2 of the SAS system for PC (SAS Institute, Inc., 2008). The TPSPLINE is based on the work of Duchon (1976, 1977) and Meinguet (1979) and described in detail in SAS Institute Inc. (2008) User's Guide. The calculations used to express the relationship between MAP and MAT and the four regressors are described below.

The TPSPLINE approach in $PPM_{1.0}$ uses the PLS regressors ($R_1$-$R_4$) and these are plotted in a *d*-dimensional space, $H_m$ (that is, Hilbert Space), of functions that have partial derivatives of total order *m*, where $m \leq 2^d - 1$. In the case of the $PPM_{1.0}$ model, $d = 4$ ($R_1$, $R_2$, $R_3$ and $R_4$) and $m = 15$. The derivation of the TPSPLINE model that is the $PPM_{1.0}$ is detailed below.

The PLS-regressors, $R_1$, $R_2$, $R_3$ and $R_4$ are fit to MAP and MAT using the following semiparametric model:

$$y_i = f(\mathbf{R}_i) + \varepsilon_i,\ i = 1,\ldots n, \quad (7)$$

where $y_i$ are the combined responses MAP and MAT. Here, $f(\mathbf{R}_i)$ is a nonparametric function in the model and the domain $\boldsymbol{R}_i$ are the smoothing variables, $R_1$, $R_2$, $R_3$ and $R_4$. The function, *f*, is an element within $H_m$; and lastly, $\varepsilon_i$, $i = 1, \ldots, n$, is the independent, zero-mean random errors from the sample, $n = 685$.

The function, *f*, in equation (7) is estimated by setting a fixed smoothing parameter, $\lambda$ (see text on $\lambda$ below), and minimizing a penalized least squares function using the following form:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(R_i))^2 + \lambda J_m(f) \quad (8)$$

Here, $\lambda J_m(f)$ is a penalty term that imposes smoothness on $f$. In the case of $PPM_{1.0}$, modeled MAT, $\lambda$ is $4.07*10^{-6}$ and $\lambda$ for MAP is $3.68*10^{-7}$. Thus, the $PPM_{1.0}$ $\lambda \approx 0$ and essentially yield no penalty term, hence the solution is mainly the penalized least squares model form with the emphasis on the semiparametric function of the $R_1$-$R_4$ smoothing variables. Derivation of $\lambda J_m(f)$ and the basis functions that connect individual thin-plate splines in the overall model are detailed in Appendix B.

We use a Monte Carlo method to assess error. The resulting model was validated using a repeated random sub-sampling validation based on 10,000 simulations. The number of simulations were chosen based on the complexity of the model, number of model parameters and range of possible values for the parameters. Repeated random sub-sampling validation is a simulation-based approach to assess how the model will generalize to an independent data set. The model presented uses the full data set to estimate climate (MAP and MAT), however this gives little insight as to how the model will perform on another data set. Repeated random sub-sampling validation randomly selects a subset of the data and performs the same model fit procedure (PLSR followed by TPSPLINE) on this random subset of the data, which is referred to as the "training data set" because it informs the model. The remaining data that are not part of the training data set are used to test the fit. This random subset, referred to as the "validation data set", is run through the model created from the training data, and climate predictions are obtained. Then, the error between the true climates of the validation data set with the predicted climates can be measured. By repeating this procedure many times, each time randomly partitioning the data into testing and validation data sets, the approximate distribution of prediction error can be assessed. For this simulation study, we used 90 percent of the data to create the training data set and the remaining 10 percent to create the validation data set, repeating the process for 10,000 simulations. The validation error, reported here as Root Mean Squared Prediction Error (RMSPE), is an approximation for the average error we would expect in the climate predictions for external data sets.

## RESULTS

### *Description of the PPM v. 1.0 Soil Dataset*

The 685 uppermost B horizon samples are from 10 different soil orders at depths ranging from ~3 to 130 cm with > 50 percent within 40 cm of the ground surface (figs. 3 and 4; supplementary table 1, http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb). Over half of the dataset is comprised of Alfisols, Mollisols, and Inceptisols, while the remaining portion is from Spodosols, Andisols, Oxisols, Vertisols, Aridisols, Ultisols and Entisols. The latter soil order only includes coarse-grained B horizons, where the soil may still classify as an Entisol (Soil Survey Staff, 1999 – see chapter on Entisols). The only USDA soil orders not represented here are Histosols because they contain no B horizons, and Gelisols because no climate data were available for the one uppermost B horizon sample. When only the first horizon suffix is considered, the uppermost B horizons are dominated by Bt and Bw descriptors (> 70% of the samples). This group of Bt and Bw horizons includes: Btg, Btk, Btn, Bts, Btv, Btx, Bwg, Bwk, and Bwu horizons (table 2). The sample of Bg horizons, common in poorly-drained soils, comprises <5 percent of the total sample population. Other horizons in the data set include: B, Bgy, Bh, Bhs, Bk, Bkm, Bkn, Bkq, Bks, Bky, Bo, Bs, Bss, By, Byz and Bz. The complete nomenclature for the uppermost individual B horizons is too complicated to group and describe statistically. However, the subordi-

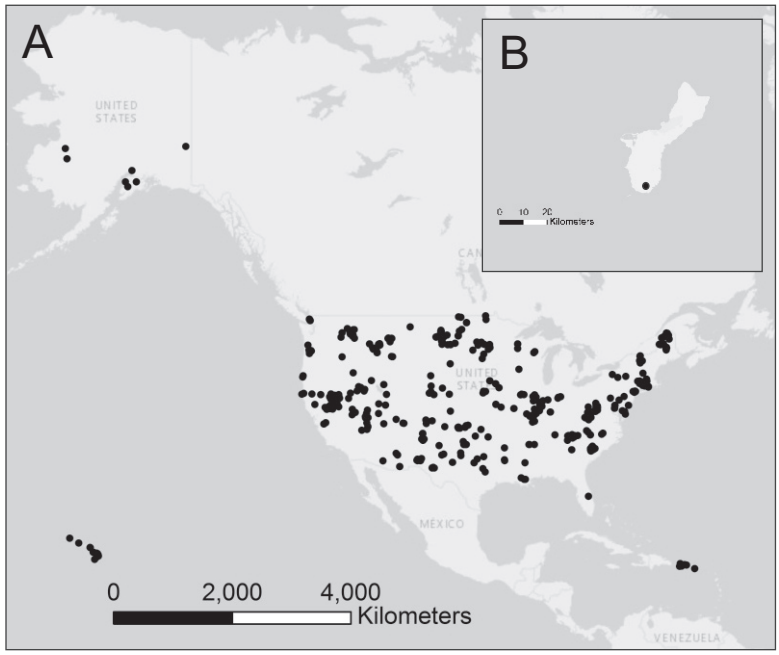


Fig. 3. Map showing location of pedons (black dots) used in this study. (A) The distribution of pedons is geographically biased towards the conterminous USA and samples from Alaska, Hawaii, Puerto Rico and (B) Guam. The pedon locations were obtained from the NRCS geochemical database (Burt and others, 2003; Wilson and others, 2008).

nate B horizons listed here suggest a diverse range of subsoil properties characterize the data set. The uppermost B horizons have a wide variety of NRCS soil textures ranging from sand to clay.

The B horizons sampled here formed under a variety of soil-forming factors. Among the pedons sampled, MAP (*cl*) ranges from ~130 to 6866 mm $yr^{-1}$ and MAT ranges from ~0 to 27 ° C (fig. 4). Although we did not model effective precipitation, the pedons fall within regions ranging from arid (low effective precipitation) to tropical (high effective precipitation). Land cover (*o*) varies widely, with many of the soils forming under forests, grassland, pasture hay, cultivated crops, or scrub. Roughly half of the pedons have a slope <5 percent at elevations ranging from near sea level to greater than 2000 m (*r*). Parent material (*p*) types also vary widely with > 50 percent forming on alluvium or residuum. Based on soil series descriptions and mineralogical family taxonomic associations from B horizons, the pedons have assumed parent material compositions that include, but not limited to: mixed alluvium, quartz-rich material, carbonate, serpentinite, and felsic to mafic volcanic ash, which come from a wide array of sedimentary, metamorphic and igneous rocks.

The duration of soil formation (*t*) for the 685 pedons sampled is mostly unknown; however, some general observations can be made (supplementary table 1, http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb). Official series descriptions for many of the pedons suggest that soil formation durations range from decades to millions of years. Entisols and Inceptisols forming in recent alluvium (for example, Limerick and Chewacla Series) suggest a duration age of decades to centuries. At least 8 to 10 percent of the database contains soils formed on late Pleistocene loess mantling high terraces and uplands throughout the Midwestern

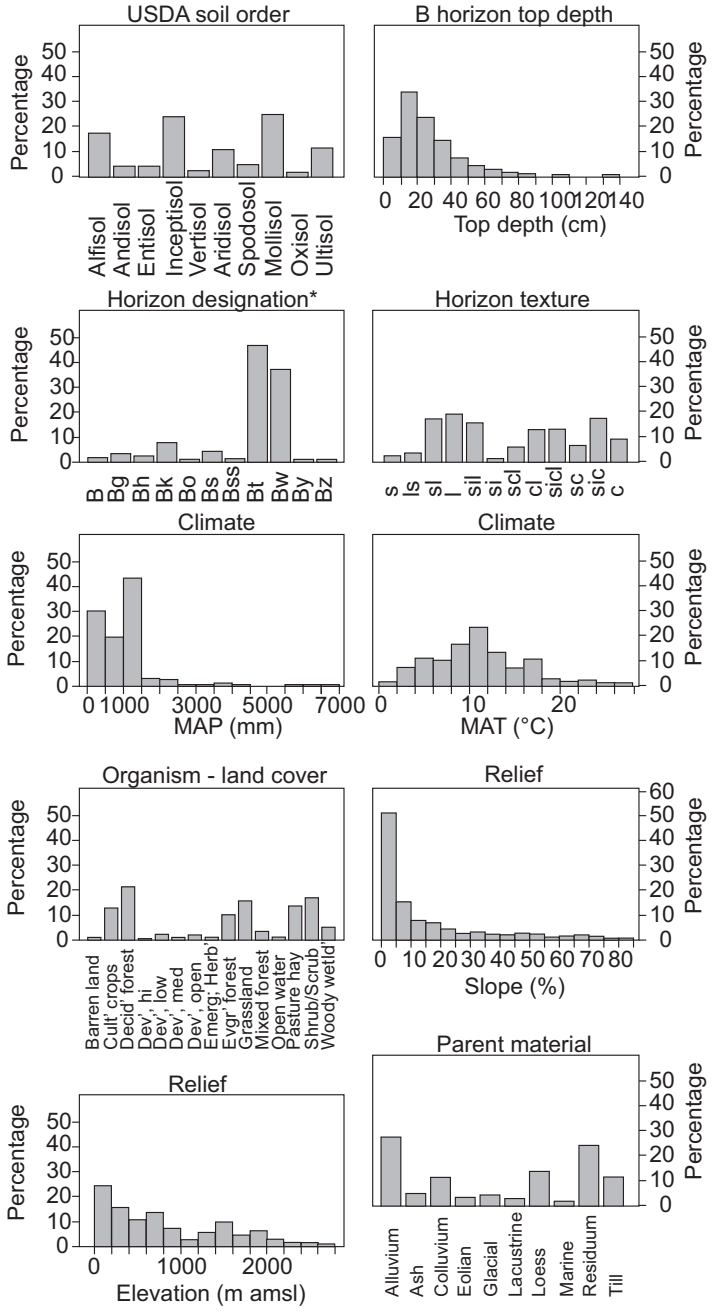


Fig. 4. Summary of 685 uppermost B horizons sampled and associated environmental factor used in this study. * Horizon designations shown are simplified by including only the first horizon modifier, if present. See Soil Survey Division Staff (1993) for horizon descriptions. For horizon textures: l = loam, sic = silty clay, sl = sandy loam, sil = silt loam, sicl = silty clay loam, cl = clay loam, c = clay, sc = sandy clay, scl = sandy clay loam, ls = loamy sand, s = sand, si = silt.

TABLE 2

*Soil B horizon suffix symbols found in the $PPM_{1.0}$ dataset and their associated characteristics (table adapted from Schaeztl and Thompson, 2015).*

| Suffix | Description |
|---|---|
| g | Presence of gleying – where Fe has been reduced or removed usually due to stagnant water; Munsell colors commonly have a low chroma (≤2) |
| h | Accumulations of amorphous and dispersible organic material |
| k | Accumulation of secondary pedogenic $CO_3$, commonly as $CaCO_3$ |
| m | Cemented or indurated horizon that implies root restrictions |
| n | Accumulation of pedogenic exchangeable Na, commonly as an Na-salt |
| o | Residual accumulation of sesquioxides |
| s | Accumulation of Fe and Al sesquioxides that are often amorphous |
| ss | Presence of pedogenic slickensides |
| t | Accumulation of silicate clay as clay coatings or as lamellae |
| u | Presence of human remains, i.e., artifacts |
| v | Presence of plinthite – a firm, Fe-rich, humus-poor material that is often red and irreversibly hardened when exposed and allowed to dry |
| w | Reddish color and some soil structure but with little evidence of illuviated material |
| x | Presence of a fragipan or fragic properties – pedogenically induced brittleness, firmness and can have high bulk density |
| y | Accumulation of pedogenic gypsum |
| z | Accumulation of salt(s) more soluble than gypsum |

United States (for example, Menfro series). Many of these soils have estimated soil formation durations of 12,000 to 14,000 years, and could also have received dust additions throughout the Holocene (Muhs and others, 2001). Cosmogenic dating of Alfisols and Ultisols forming on stable Piedmont uplands has estimated residence times on the order of $10^5$ to $10^6$ years (Pavich and others, 1985; Bacon and others, 2012). This variation in soil ages will undoubtedly influence the results of the model. Yet, by incorporating soils of varying ages, fewer assumptions must be met for the model to perform well in an application using paleosols (similar to Sheldon and others, 2002).

### *Partial Least Squares Regression*

Partial Least Squares Regression (PLSR) on the eleven geochemical oxides from the soil B horizons identified four regressors ($R_1$, $R_2$, $R_3$, $R_4$) to be the best linear predictors of the joint response (MAP and MAT) (figs. 5 and 6). Although the PLSR program generated more regressors than $R_1$-$R_4$, adding an additional regressor, for example, $R_5$, only slightly improved the TPSPLINE model performance. All four regressors correlate significantly with either MAP or MAT (Pearson's $r$, $p<0.05$); and thus, all eleven oxides possess value in linking geochemistry and climate (fig. 5, table 3). Therefore, no oxide was omitted from the $PPM_{1.0}$ model.

The correlation loadings show how much variation in each variable is accounted for by each regressor. The loadings are plotted here as correlation coefficient, $r$, and are examined to clarify how Regressors 1 to 4 load the oxides to maximize covariance with MAP and MAT (figs. 5 and 6). Note that the correlation loadings are similar to **P** and **C**, and some oxide loadings for each of the four regressors cannot be explained

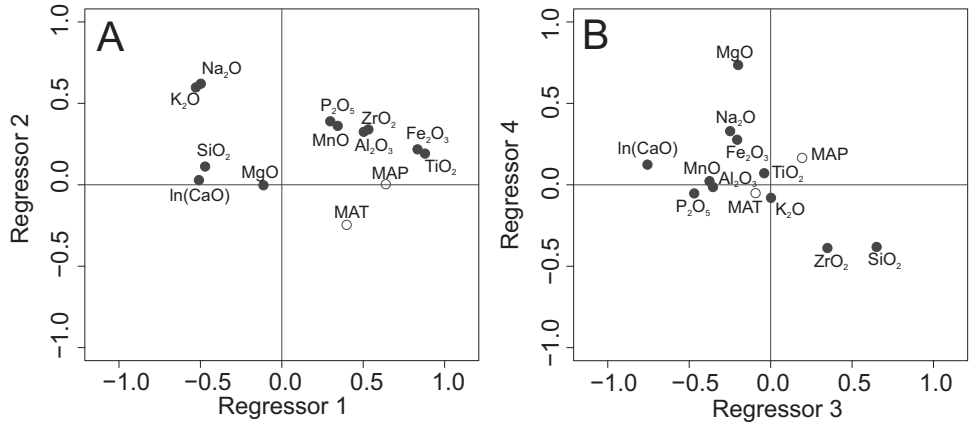


Fig. 5. Correlation loading plots for independent (geochemical oxides) and dependent (MAP and MAT) variables by regressor. Regressors are plotted against one another to emphasize differences in element loadings and correlation with MAP and MAT. The loadings show the amount of variability in each geochemical oxide that is accounted for by the Regressors, 1–4, and the degree to which the Regressors explain the response variables, MAP and MAT. See table 2 for corresponding $r$ values. Note that $r$ does not directly reflect oxide or climate loadings, **P'** or **C'**, that were determined in an 11-dimensional space using Partial Least Squares Regression (PLSR).

through pedogenesis rather, this is likely the result of the mathematics of the loading process that maximizes covariance between **X** and **Y** scores. These four regressors are orthogonal with respect to each other (that is, they are not correlated to each other), and so links between weathering processes and MAP and MAT for Regressors 1 to 4 are, *a priori*, statistically independent. This differs from all previous approaches using bulk-soil geochemical data to devise paleoclimate proxies, which all possess some degree of multicollinearity among independent variables related to the fact that geochemical oxide constituents are mass fractions that sum to one. Each regressor has less predictive power than the regressor before it.

Correlation loadings for Regressor 1 show that it is directly correlated with MAP and MAT, having Pearson's $r$ values of 0.65 and 0.41, respectively (figs. 5 and 6, table 3). Thus, Regressor 1 from the PLSR results contains the most predictive potential with respect to MAP and MAT (fig. 6). Correlation coefficients for Regressor 1 show $Fe_2O_3$, $TiO_2$, $ZrO_2$, $Al_2O_3$, $Na_2O$, $K_2O$, ln(CaO) and $SiO_2$ contribute the most to its prediction strength, where Pearson's $r > 0.4$. Here, $Fe_2O_3$, $Al_2O_3$, $TiO_2$, and $ZrO_2$ loadings are positively correlated with Regressor 1 (and MAP and MAT), whereas $Na_2O$, $K_2O$, ln(CaO) and $SiO_2$ are negatively correlated. Notably, the base oxides [$Na_2O$, $K_2O$, ln(CaO)] inversely correlate with MAT and MAP; whereas, the refractory oxides positively correlate with MAP and MAT.

Regressors 2 to 4 possess some potential for predicting MAP and/or MAT, albeit notably less than Regressor 1 (fig. 6). The results show that Regressor 2 is inversely correlated with MAT, Pearson's $r = -0.25$, and shows no correlation with MAP, Pearson's $r < 0.01$. Correlation coefficients for Regressor 2 show that $Na_2O$, $K_2O$, $P_2O_5$ and MnO contribute the most to its prediction strength, where Pearson's $r > 0.36$. Although Regressor 2 is positively correlated with ten of the eleven oxides, it is most strongly correlated with $Na_2O$ and $K_2O$. $P_2O_5$ and MnO concentrations also provide some weight to Regressor 2, where Pearson's $r \geq 0.36$.

Regressor 3 is positively correlated with MAP (Pearson's $r = 0.21$), but shows a weak correlation with MAT, where Pearson's $r = 0.08$. Correlation coefficients for Regressor 3 show that ln(CaO) and $SiO_2$ contribute the most to its composition, where

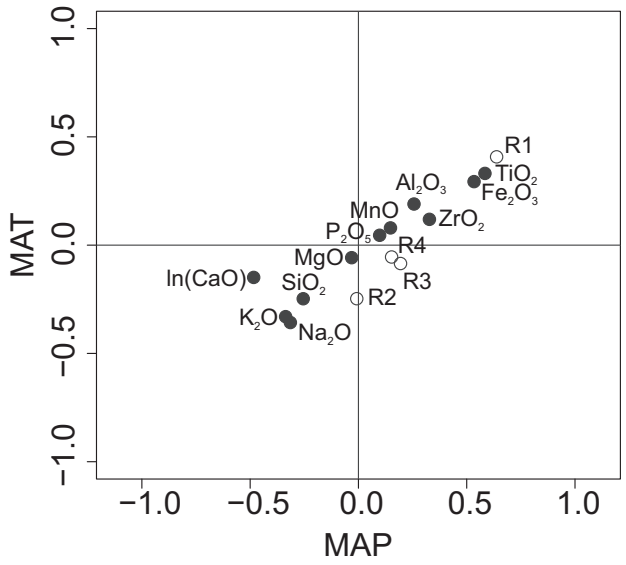


Fig. 6. Correlation loading plots for oxides and Regressors 1–4 by MAP and MAT. The loadings show the amount of variability in each geochemical oxide that is accounted for by the Regressors, 1–4, and the degree to which the Regressors explain the response variables, MAP and MAT. See table 2 for corresponding *r* values. The base oxides inversely correlate with MAT and MAP; whereas, the refractory oxides directly correlate with MAP and MAT, suggesting a fundamental temperature and precipitation weathering control on base loss and retention of refractory oxides. Note that *r* do not directly reflect oxide or climate loadings, **P'** or **C'**, that were determined in an 11-dimensional space using Partial Least Squares Regression (PLSR).

Pearson's $r > |0.66|$. Regressor 3 is positively correlated with $SiO_2$ and negatively correlated with ln(CaO). $P_2O_5$, MnO and $ZrO_2$ concentrations also provide some weight to Regressor 3, where Pearson's $r \geq 0.36$. Like Regressor 3, Regressor 4 is positively correlated with MAP, where Pearson's $r = 0.16$, but shows little to no correlation with MAT, where Pearson's $r = -0.05$. Correlation coefficients for Regressor 4 show that MgO contributes the most to its prediction strength, where Pearson's $r = 0.74$. $ZrO_2$ and $SiO_2$ contribute some weight, where Pearson's $r > |0.38|$. MgO is positively correlated with Regressor 4 and MAP, whereas $ZrO_2$ and $SiO_2$ are negatively correlated.

## DISCUSSION

### *Regressors 1 to 4, Molecular Weathering Ratios, MAP and MAT*

Given the population of soils used in this study and their diverse range of environmental state factors, one would assume no correlation with *cl* because of random variation or "noise" from soil-forming factors other than *cl*, notably: *o*, *r*, *p*, *t*, . . . . The PLSR results show correlations between soil geochemistry and climate through 6 dimensions (4 Regressors, 2 responses), as opposed to 13 dimensions (11 oxides, 2 responses). The relationship between the PLS regressors, pedogenesis, and climate were examined using an independently measured suite of physical and chemical soil characterization data and ionic potential values for individual metals using regression analysis.

A correlation of MAP, MAT, Regressors 1 to 4 and published molecular weathering ratios was performed on the 685 B horizons to test the suitability of using the Regressors 1 to 4 to model climate compared to previously established paleoclimate

Table 3

*Correlation coefficients, Pearson's r, for Regressors1-4 ($R_1$-$R_4$), geochemical oxide, and MAP and MAT loadings.*

| Variable | $R_1$ | $R_2$ | $R_3$ | $R_4$ | MAP | MAT |
|---|---|---|---|---|---|---|
| $Fe_2O_3$ | 0.85 | 0.22 | -0.19 | 0.28 | 0.55 | 0.33 |
| MnO | 0.36 | 0.36 | -0.36 | 0.02 | 0.16 | 0.08 |
| $P_2O_5$ | 0.31 | 0.39 | -0.46 | -0.05 | 0.11 | 0.05 |
| $SiO_2$ | -0.47 | 0.11 | 0.66 | -0.38 | -0.25 | -0.25 |
| $TiO_2$ | 0.89 | 0.19 | -0.03 | 0.07 | 0.60 | 0.33 |
| $ZrO_2$ | 0.54 | 0.34 | 0.36 | -0.39 | 0.34 | 0.12 |
| $Al_2O_3$ | 0.52 | 0.32 | -0.34 | -0.01 | 0.27 | 0.19 |
| $Na_2O$ | -0.48 | 0.62 | -0.24 | 0.33 | -0.30 | -0.36 |
| MgO | -0.10 | 0.00 | -0.19 | 0.74 | -0.02 | -0.06 |
| $K_2O$ | -0.51 | 0.60 | 0.01 | -0.08 | -0.33 | -0.33 |
| ln(CaO) | -0.49 | 0.03 | -0.74 | 0.12 | -0.47 | -0.15 |
| $R_1$ | 1.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.41 |
| $R_2$ | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | -0.25 |
| $R_3$ | 0.00 | 0.00 | 1.00 | 0.00 | 0.21 | -0.08 |
| $R_4$ | 0.00 | 0.00 | 0.00 | 1.00 | 0.16 | -0.05 |
| MAP | 0.65 | 0.00 | 0.21 | 0.16 | 1.00 | 0.33 |
| MAT | 0.41 | -0.25 | -0.08 | -0.05 | 0.33 | 1.00 |

proxies. The results show that Regressor 1 has a higher correlation coefficient, *r*, with respect to MAP and MAT than any other published molecular weathering ratio utilized to assess weathering or to predict climate (table 4). Regressor 1 is similar with respect to the product index, developed by Reiche (1950), which tracks the slow dissolution of silica and enrichment of immobile metals and iron oxides.

Regressor 2 is correlated with the Paleosol Weathering Index (PWI), a paleothermometer developed by Gallagher and Sheldon (2013), which tracks the concentration of K, Na, Mg and Ca as a function of temperature with the reasoning that in cooler climates there is less weathering of these mobile elements and that they are more likely to accumulate in salts. Regressor 3 more closely tracks the CALMAG proxy than any other regressor (Nordt and Driese, 2010a). This relationship suggests Regressor 3, much like CALMAG, is tracking the loss of $Ca^{2+}$ through leaching of calcium carbonate and exchangeable $Ca^{2+}$ with increasing MAP (*ibidem*). Unlike CALMAG, however, Regressor 3 does not significantly weight Mg. The varying concentrations of Mg play a greater role in Regressor 4, which closely tracks $Fe_2O_3/Al_2O_3$ ratio of Retallack (2001).

The advantage of the PLSR analysis is that the Regressors 1-4 track changes in MAP and MAT as with other published molecular weathering ratios, but has higher correlation coefficients with MAP and MAT. It is possible the correlations may improve for Regressors 1-4 or other ratios if nonlinear models are applied (see next section).

*Regressors 1 to 4 Covariance with Climate: Underlying Mechanisms*

In order to appreciate how the Regressors maximize covariance with climate their composition is examined here using both correlation plots of regressor loadings (**P**) and regressor scores (**T**). Regressor oxide loadings with Pearson's $r > 0.4$ can be explained in terms of first-order weathering principles. For all four regressors, climate-

TABLE 4

*Correlation coefficients, Pearson's r, for MAP, MAT, Regressor 1-4 scores ($R_1$-$R_4$) and published molecular weathering ratios.*

| Parameter or ratio | Common name | MAP | MAT | $R_1$ | $R_2$ | $R_3$ | $R_4$ | Source |
|---|---|---|---|---|---|---|---|---|
| MAP (mm $y^{-1}$) | Mean annual precipitation | - | 0.33 | 0.61 | 0.12 | 0.19 | 0.11 | Daly and others, 1994 |
| MAT (ºC) | Mean annual temperature | 0.33 | - | 0.42 | 0.21 | -0.12 | -0.06 | Daly and others, 1994 |
| $R_1$ | Regressor 1 | **0.61** | **0.42** | - | 0.00 | 0.00 | 0.00 | This study |
| $R_2$ | Regressor 2 | 0.12 | 0.21 | 0.00 | - | 0.00 | 0.00 | This study |
| $R_3$ | Regressor 3 | 0.19 | -0.12 | 0.00 | 0.00 | - | 0.00 | This study |
| $R_4$ | Regressor 4 | 0.11 | -0.06 | 0.00 | 0.00 | 0.00 | - | This study |
| $(Al_2O_3+K_2O)/(MgO+CaO+Na_2O)$ | Residual Index (RI) | 0.33 | 0.35 | 0.48 | 0.35 | 0.12 | -0.28 | Vogt, 1927 |
| $[SiO_2/(TiO_2+Fe_2O_3+SiO_2+Al_2O_3)]*100$ | Product Index (PI) | -0.49 | -0.34 | **-0.88** | 0.27 | 0.10 | -0.08 | Reiche, 1950 |
| $SiO_2/Al2O_3$ | Ruxton Ratio (R) | -0.12 | -0.01 | -0.31 | 0.43 | 0.13 | 0.16 | Ruxton, 1968 |
| $(CaO+MgO+Na_2O+K_2O)/Al_2O_3$ | Bases:alumina | -0.02 | 0.02 | -0.05 | 0.07 | -0.39 | 0.08 | Birkeland, 1999 |
| $(CaO+MgO+Na_2O+K_2O)/(Al_2O_3+TiO_2+Fe_2O_3)$ | Bases:$R_2O_3$ | -0.02 | 0.02 | -0.05 | 0.07 | -0.39 | 0.07 | Birkeland, 1999 |
| $SiO_2/Fe_2O_3$ | Silica:iron | -0.15 | 0.01 | -0.31 | 0.39 | 0.06 | 0.03 | Birkeland, 1999 |
| $SiO_2/(Al_2O_3+Fe_2O_3)$ | Silica:sesquioxides | -0.14 | 0.00 | -0.33 | 0.44 | 0.12 | 0.08 | Birkeland, 1999 |
| $SiO_2/(Al_2O_3+TiO_2+Fe_2O_3)$ | Silica:$R_2O_3$ | -0.17 | -0.01 | -0.37 | 0.45 | 0.11 | 0.09 | Birkeland, 1999 |
| $(Na_2O+K_2O)/Al_2O_3$ | Alkali:alumina | -0.11 | -0.09 | -0.22 | -0.06 | -0.27 | 0.13 | Retallack, 2001 |
| $Na_2O/K_2O$ | Soda:potash | -0.01 | -0.10 | -0.08 | -0.26 | -0.27 | 0.27 | Retallack, 2001 |
| $Na_2O/Al_2O_3$ | Soda:alumina | -0.06 | -0.06 | -0.14 | -0.05 | -0.29 | 0.11 | Retallack, 2001 |
| $(Fe_2O_3+MnO)/Al_2O_3$ | Iron Manganese:alumina | 0.39 | 0.13 | 0.55 | -0.11 | 0.08 | 0.56 | Retallack, 2001 |
| $Fe_2O_3/Al_2O_3$ | Iron:alumina | 0.40 | 0.14 | 0.55 | -0.09 | 0.10 | **0.58** | Retallack, 2001 |
| $Al_2O_3/(Al_2O_3+CaO+MgO+Na_2O+K_2O)*100$ | Chemical Index of Alteration (CIA) | 0.46 | 0.27 | 0.55 | 0.29 | 0.51 | -0.46 | Nesbitt and Young, 1982 |
| $Al_2O_3/(Al_2O_3+CaO+MgO)*100$ | CALMAG | 0.41 | 0.14 | 0.37 | 0.22 | **0.66** | -0.44 | Nordt & Driese, 2010a |
| [(4.20*Na)+(1.66*Mg)+(5.54*K)+(2.05*Ca)]*100 | Paleosol Weathering Index (PWI) | -0.43 | -0.25 | -0.49 | **-0.50** | -0.55 | 0.32 | Gallagher and Sheldon, 2013 |
| $Al_2O_3/(Al_2O_3+CaO+MgO+Na_2O)*100$ | CIA minus potash (CIA-K) | 0.43 | 0.22 | 0.46 | 0.30 | 0.57 | -0.47 | Harnois, 1988; Sheldon and others, 2002 |

The strongest correlation (maximum $|r|$) per column is in bold. Note that $R_1$ has the strongest correlation for both MAP and MAT, compared to all other Regressors and published geochemical oxide ratios.

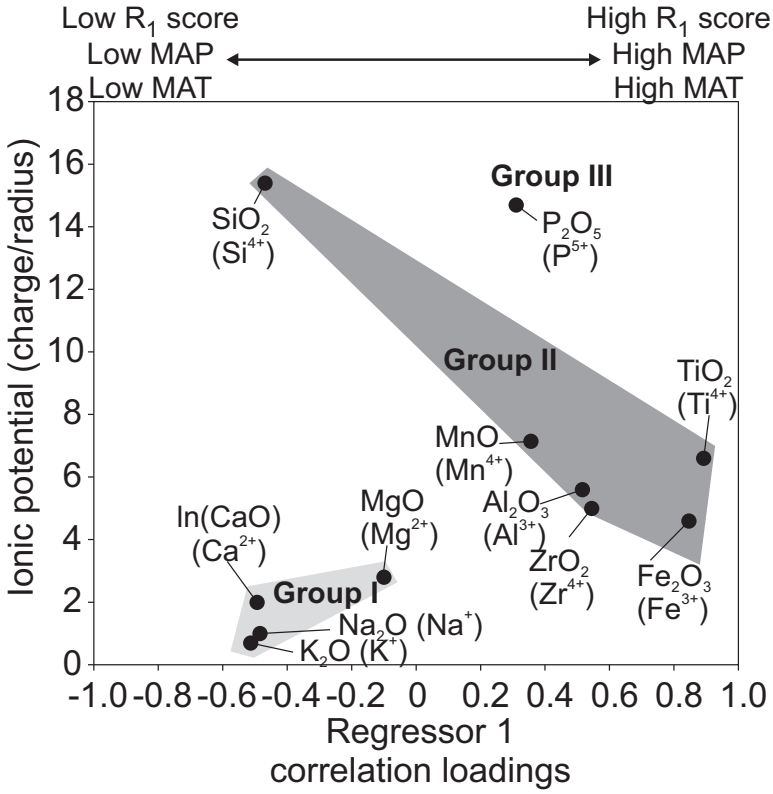


Fig. 7. Ionic potential with respect to PLS Regressor 1 oxide correlation loadings. Regressor 1 ($R_1$) directly correlates with MAP and MAT. Lower $R_1$ scores (low MAP, low MAT) correspond with more negative $R_1$ loadings. Immobile metals (Al, Ti, Zr, Fe), which cluster in the insoluble hydroxides field on the periodic table (Group II), directly correlate with Regressor 1 values and thus have higher concentrations at higher MAP and MAT. Mobile elements (K, Na, Ca, Mg), which cluster in the soluble cation field of the periodic table (Group I), inversely correlate with Regressor 1 and have higher concentrations in lower MAP and MAT settings. The Si behaves like the bases because of desilication in high MAP and MAT environments. The clustering of metals by ionic potential as it relates to MAP and MAT suggests that Regressor 1 utilizes first-order thermodynamics-based controls on weathering as they relate to climate.

driven pedogenic explanations for oxides with correlation coefficients $<0.4$ remain uncertain. The significance of Pearson's $r = 0.4$ remains unclear, yet it appears to be a threshold value and the strength and direction of those $<0.4$ $r$ values are difficult to explain pedogenically. Thus, the approach taken was to focus on the oxides more highly correlated with the regressors while acknowledging that oxides with low correlation were likely weighted to achieve orthogonality between regressors. In some cases, oxides with slightly lower $r$ values (such as Pearson's $r = |0.35|$ for ln(CaO) in Regressor 1) can be explained by pedogenic processes; in other cases, similar $r$ values seem to show arbitrary correlations.

*Regressor 1.*—Regressor 1 has a strong direct relationship with MAT and MAP, whereby the subsoil experiences iron oxidation, base loss, acidification and retention of refractory elements with increasing temperature and precipitation. Regressor 1 correlation coefficients for individual oxides show a systematic relationship when plotted versus the ionic potential of the metal component within the oxide (fig. 7). Defined as the ionic charge divided by the ionic radius, ionic potential provides a sense of how easily an ion can be weathered from a mineral in solution (Goldschmidt, 1937).

Negative Regressor 1 correlation loadings are associated with the low ionic potentials observed for alkali and alkaline earth metals. The low ionic potential translates to weaker bond strength and increased mobility under lower pH conditions. MgO has ~zero loading in Regressor 1 and we infer that it is due to its significant loading in Regressor 4 and the role in weathering of basic volcanic ash (see *Regressor 4*) and the fact that primary sources of Mg are mafic minerals, smectite, some salts and carbonate. Low pH systems like those emphasized in Regressor 1 have minimal carbonate, salt and primary minerals from mafic systems.

Conversely, high positive Regressor 1 oxide correlation coefficients (high MAP; high MAT) are associated with high ionic potentials observed among the chemically inert refractory metals, Ti and Zr, and the more immobile Al and $Fe^{3+}$. These relationships suggest that as Regressor 1 (and therefore as MAP and MAT) increases, the higher precipitation and temperatures combined lead to increased water flow and dissolution of mobile metals from primary aluminosilicate and clay minerals, and retention of immobile metals and Fe- and Al-sesquioxides in well-drained soil-forming environments (White and Blum, 1995; Chadwick and Chorover, 2001; Egli and others, 2004; Rasmussen and others, 2011). Although Bg horizons are present in the dataset, they likely reflect seasonal drainage and thus have little effect on this aspect of regressor loading.

The negative loading of $SiO_2$, where the $Si^{4+}$ has a high ionic potential, suggests that Regressor 1 maximizes prediction of MAP and MAT especially for extensively weathered environments that have experienced desilication, where the leaching of $Si^{4+}$ from the soil combined with ferrallitization accumulates sesquioxides (fig. 7) (Esrawan and Bin, 1978; Buol and Esrawan, 2000).

These observations strongly resemble the fundamental Jackson-Sherman weathering stages as they relate to climate (Jackson and Sherman, 1953). Early stage weathering includes minimally weathered soils with high bases that may have experienced dry, limited leaching, and/or cold environments. Advanced-stage weathering typically has a climax mineralogy largely composed of kaolinite and the metal oxides hematite, goethite, gibbsite, and rutile, and few to no bases due to the extensive leaching and oxidation typical in warm-wet climates. These data lead one to surmise that Regressor 1 maximizes prediction in soils forming under higher than average MAP and MAT. With high rainfall and extensive leaching, silicic acid concentrations will decrease due to less soluble phases (quartz). As a result, the 1:1 kaolinite is the stable clay mineral phase at $\log H_4SiO_4 < -4$ (Sparks, 2002). At lower silicic acid concentrations in the soil solution, $\log H_4SiO_4 < -5.3$, gibbsite becomes the more dominant stable phase over kaolinite.

Fundamental relationships between Regressor 1, geochemical weathering principles, and climate are also apparent in correlations with soil characterization data commonly used to assess pedogenic processes (fig. 8). Properties that correlate positively with Regressor 1 are citrate-dithionite extractable Fe ($Fe_d$) and Al ($Al_d$), and water retention (WRT). These data show that with increasing MAP and MAT, repeated wet-dry cycles alter Fe(II) minerals and release Fe, forming pedogenic iron segregations and coatings ($Fe_d$) (Nordt and Driese, 2009). The frequent wet-dry cycles and exposure to oxic conditions immobilizes Fe(III). The Fe content of the soil is enriched as the soil matures and base loss increases. The mobilization and removal of Fe in poorly drained soils is not a defining process in this database because the number of Bw and Bt horizons of well-drained soils far outweighs the number of Bg horizons common in seasonally wet soils where Fe leaching is unlikely. Furthermore, Regressor 1 utilizes iron concentration and not variations in iron mineralogy. Thus, the higher concentrations of iron will only indicate warm and seasonally wet climates; whereas variations in the ratio of hematite to goethite, regardless of $Fe_2O_3$ concentration, can resolve dry climates (high hematite to goethite) from wet climates (low hematite to

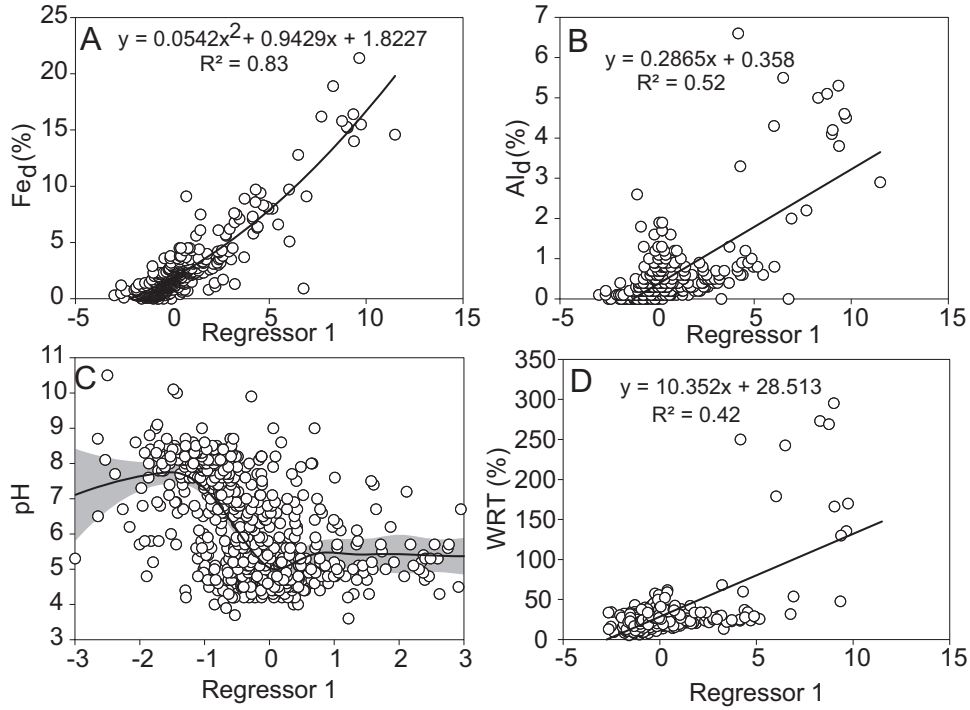


Fig. 8. Correlations PLS Regressor 1 scores. (A) Pedogenic Fe ($Fe_d$) is positively correlated with Regressor 1 and suggests that pedogenic Fe segregations (for example, magnetite, goethite and hematite) increase in concentration with increasing MAP and MAT. (B) Pedogenic Al ($Al_d$) is similar to $Fe_d$ and shows a similar relationship with respect to Regressor 1. High $Al_d$ and $Fe_d$ are common in subsoils that accumulate pedogenic Fe and Al through podsolization (McKeague and Day, 1966). (C) Soil pH (1:1 $H_2O$) shows a sigmoidal correlation with Regressor 1 suggesting that the correlation with MAP is partly driven by alkalinization or acidification. (D) Water retention (WRT) is positively correlated with Regressor 1. Higher WRT values are commonly associated with higher clay and organic matter concentrations. The higher Regressor 1 scores, and therefore higher MAP and MAT, coincides with more clay-rich soils with higher subsoil organic content either due to texture, podsolization or both.

goethite) (Kampf and Schwertmann, 1983; Ji and others, 2004; Hyland and others, 2015).

The relationship between WRT and Regressor 1 correlation coefficients is positive because higher WRT values are commonly associated with higher clay and organic matter concentrations, although the relationship between these three variables is not straightforward (Rawls and others, 2003). The higher Regressor 1 scores, and therefore higher MAP and MAT, coincides with more clay-rich soils with higher subsoil organic content either due to texture, podsolization or both. Conversely, pH correlates negatively with Regressor 1. The decrease in pH with increasing MAP and MAT provides further support that the previously discussed relationships were due to hydrolysis and loss of bases, secondary clay mineral formation (increasing WRT), Fe oxidation and Fe and Al accumulation in the subsoil due to podsolization (Jenny, 1941; van Breeman and others, 1983).

The resemblance between Regressor 1 and pH shows a relationship similar to that discussed by Chadwick and Chorover (2001), where buffering occurs along the high and low pH end due to $CaCO_3$ and aluminosilicate clays, respectively (fig. 5C). The 685 B horizons used in this study show a sigmoidal relationship with respect to pH. The spline curve resembles the Acid Neutralizing Capacity curve and the pedogenic

threshold concept at alkaline and moderately to strongly acidic pH ranges (van Breeman and others, 1983; Chadwick and Chorover, 2001).

*Regressor 2.*—The two main components in Regressor 2, $Na_2O$ and $K_2O$, have an inverse relationship with MAT that is related to the temperature-dependent dissolution of salt and feldspar minerals within the subsoil (White and Blum, 1995; Egli and others, 2004; Williams and others, 2010). At cooler temperatures, $Na^+$ and $K^+$ are less weathered and retained in salts. The inverse relationship between temperature and Na was also observed in a data compilation of upland regolith profiles (Rasmussen and others, 2011) and in a latitudinal transect of weathered shale of similar composition and age (Dere and others, 2013). The observation of temperature-dependent leaching of Na and K shown in Regressor 2 is also an independent confirmation of the "salinization" climofunction found by Sheldon and others (2002), which relates the ratio $[(Na_2O+K_2O)/Al_2O_3]$ with temperature and has been used as a paleothermometer in paleosol studies (Sheldon and Tabor, 2009; Nordt and others, 2015).

Phosphorus, $P_2O_5$, is notable in Regressor 2 and increases with decreasing MAT. On average about 50 percent of all P in soils is stored in OM. The remainder is primarily in mineral form inherited from parent material; whereas some P is complexed on clay colloids (Brady and Weil, 2007). Thus, increasing concentrations of P with decreasing MAT is likely related to higher organic matter due to lower aboveground net primary productivity, and also to biogeochemical processes being suppressed due to lower annual temperatures (Reich and Oleksyn, 2004).

*Regressor 3.*—Regressor 3 is negatively correlated with $CaCO_3$ concentration, meaning that as MAP increases the dissolution of $CaCO_3$ increases within the uppermost B horizon (fig. 9). These results suggest that Regressor 3 is tracking decalcification. Climate has a significant impact on the distribution of calcium carbonate in desert soils (Gile and others, 1966; Mayer and others, 1988) and this appears to be the case for a portion of the soils examined here. Regressor 3 is positively correlated with $SiO_2$ (fig. 5), and it appears that the correlation between Regressor 3 and MAP is largely based on weighting soil B horizons on extreme ends of the weathering spectrum, with calcareous soils forming in low-MAP environments and $SiO_2$-rich soils (likely quartz-rich) forming in higher-MAP environments that are not too tropical, where Si is being leached extensively from the profile (for example, Oxisols).

Samples with the largest negative Regressor 3 score are those with high ln(CaO) values and low MAP (fig. 9). Soil horizons with the highest negative factor weights in Regressor 3 are a mixture of various types of Bk and By horizons, or other B horizons formed on calcic parent material under low MAP. This suggests that both the presence of pedogenic carbonate and calcic parent materials not being leached is an important component of the third latent variable. Examining horizon trends in Regressor 1, it is noted that the partitioning of calcic horizons is less evident (fig. 9D).

*Regressor 4.*—Regressor 4 shows the importance of the increase in MgO and decrease in $SiO_2$ with increasing precipitation by means of differential weathering in mafic parent materials and basic volcanic ash. A regression analysis of MAP on Regressor 4, when grouped by soils forming in ash versus soils forming in other parent materials, shows that the soils forming in ash have a positive correlation (Pearson's $r = 0.39$; $p < 0.001$) (fig. 10), whereas soils forming in parent material types other than ash show no correlation (Pearson's $r = 0.01$; $p = 0.84$). Furthermore, when the soils are grouped by soil order, Andisols have a positive correlation with MAP ($r = 0.38$) (fig. 10).

The relative stability of minerals weathering in volcanic ash, from low to high are: volcanic glass < feldspar (andesine-labradorite) ≤ hypersthene-augite < magnetite (Aomine and Wada, 1962). In this weathering scheme, increasing precipitation alters and eventually removes the volcanic glass relative to Mg-rich minerals typically found

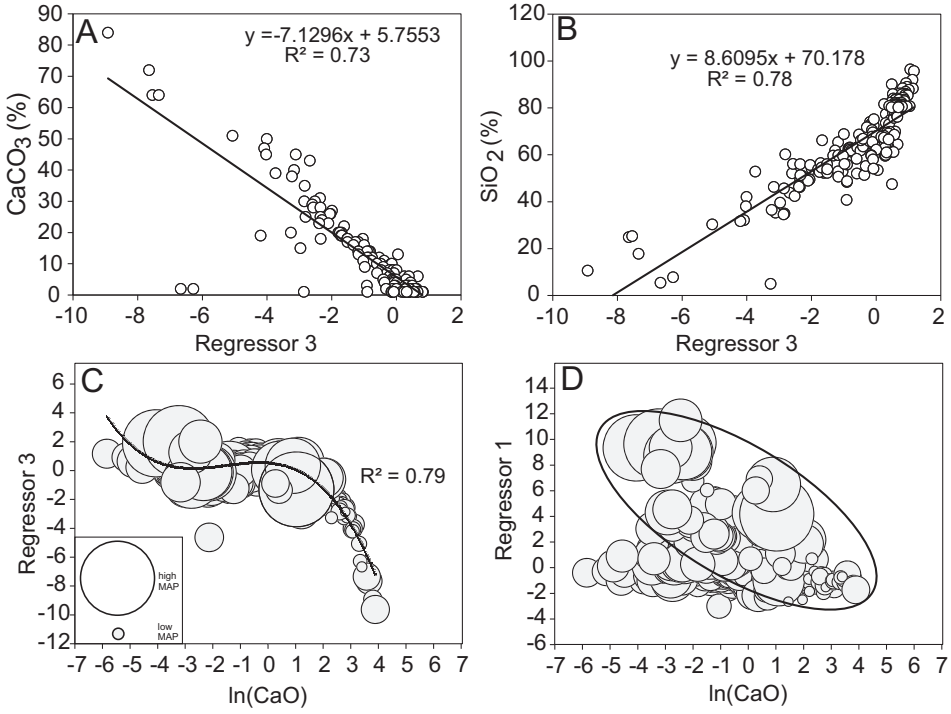


Fig. 9. PLS Regressor 3 correlations. (A) $CaCO_3$ is negatively correlated with Regressor 3 and the non-calcic soils have a low absolute Regressor 3 score. (B) $SiO_2$ directly correlates with Regressor 3. (C) Regressor 3 has a negative cubic correlation with ln(CaO). Large absolute ln(CaO) values coincide with calcic soils weathering under low MAP; whereas many of the small absolute ln(CaO) values weathering under high MAP have very low Regressor 3 scores and show no correlation with ln(CaO). (D) Unlike Regressor 3, Regressor 1 has a weak negative linear correlation with ln(CaO) ($R^2 = 0.12$). The region enclosed in the circle shows that non-calcic soils weathering in high MAP environments have large Regressor 1 scores.

in more mafic-rich ash parent materials. In more humid areas, higher organic matter (OM) contents complex $Al^{3+}$ and $Fe^{3+}$, thereby inhibiting the formation of gibbsite and aluminosilicate clay minerals. The Mg content is higher than Si because the adsorption of $Al^{3+}$ impedes aluminosilicate clay formation and the $H_4SiO_4^{2-}$ is leached from the profile. Preferential volumetric loss of $SiO_2$ over MgO can be observed in a climosequence study of soils forming on basaltic lava flows and tephra in Hawaii (Chadwick and others, 2003–see table 4). The weathering of basaltic glass in Iceland showed that soluble Si content was not related to weathering intensity or pH, whereas Mg was retained in Mg-rich smectite as weathering progressed (Crovosier and others, 1992; Stefánnson and Gíslason, 2001). Stefánsson and Gíslason (2001) suggested that imogolite/allophane and short-range order Fe oxides were the initial products of weathering, with Ca-Fe-Mg smectites becoming more common with increasing soil age.

As the volcanic ash chemical stability scheme above (Aomine and Wada, 1962) suggests, CaO and $Na_2O$ from plagioclase feldspar weathering should have been weighted heavily in Regressor 4. However, because the PLSR process avoids multicollinearity among regressors during dimension reduction, CaO and $Na_2O$ already contribute to the Regressor 1, 2 and 3 loading schemes. Mg-silicates (pyroxenes) relative to volcanic glass would be the next logical choice when trying to model volcanic weathering relative to MAP. Also, there is less chemically-resistant quartz in

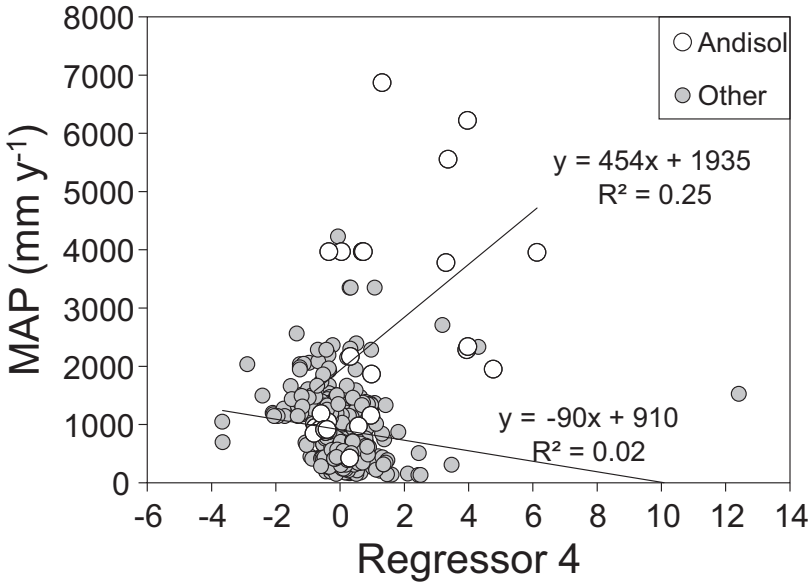


Fig. 10. Correlation of MAP with respect to Regressor 4 scores, by Andisol (white circles) and all other soil orders (gray circles).

mafic-rich parent materials and the $SiO_2$ is more in Fe- and Mg-rich silicate minerals (for example, pyroxenes).

*Paleoclimate Model Development using Thin-Plate Spline*

Climate (MAP and MAT) was modeled as a joint response on Regressors 1 to 4 (scores from 685 B horizons) using a multivariate thin-plate spline (TPSPLINE) model using SAS software (see Appendices B through E for $PPM_{1.0}$ user instructions, code, example data and links to online supplementary material, http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb, and a tutorial video) (SAS Institute, Inc., v. 9.2). The resulting Paleosol-Paleoclimate Model developed at Baylor University ($PPM_{1.0}$) lies in 6-dimensional space and therefore cannot be seen in its entirety.

Figure 11 shows three of the six dimensions of the full TPSPLINE model. The six dimensions of the full model can be placed in one of two categories: the 4-dimensional predictor space (Regressors 1–4) and the 2-dimensional outcome space (MAP and MAT). For all samples represented in our data set, we have values for each of these six dimensions (the four predictor variables come from the results of the PLSR analysis). We can plot up to three dimensions at a time in a 3D plot. Figure 11 shows MAP by Regressors 1 and 3. The points represent the observed MAP response, whereas the surface represents the model fit (using TPSPLINE). The vertical lines are the MAP residuals – the distance between the recorded MAP values and the model-fitted MAP values.

The $PPM_{1.0}$ estimates MAP with a root mean square error (RMSE) = ±228 mm and MAT with a RMSE = ±2.46° C (fig. 12, table 5). For temperature fit, modeling the bivariate, or joint, response (MAP and MAT), is superior to modeling the univariate response (MAT). This improvement in modeling MAT is rooted in the positive correlation (Pearson's $r = 0.33$) between the two climate parameters (table 3 and 4), and suggests that precipitation and temperature combines both chemical reaction rate

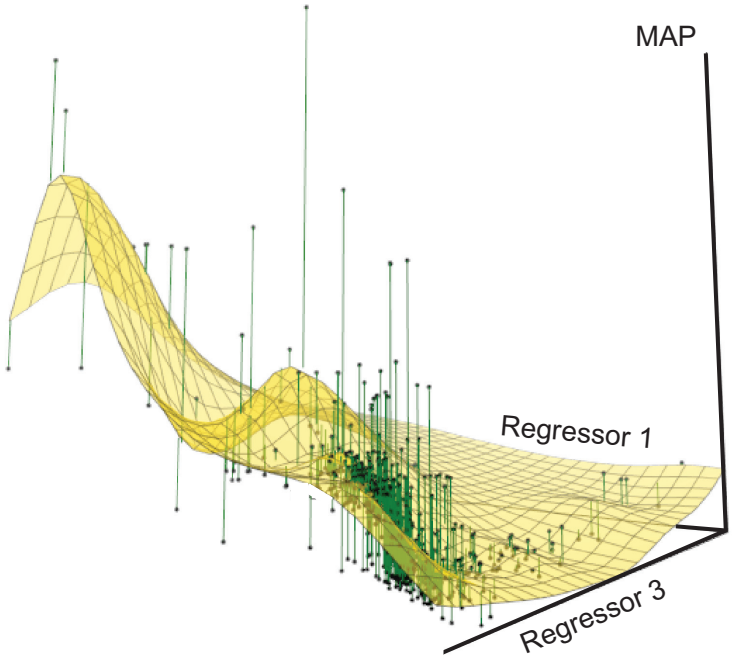


Fig. 11. A three-dimensional interpolated surface plot of MAP on Regressors 1 and 3. This illustration provides one possible three-dimensional perspective of the $PPM_{1.0}$, TPSPLINE model behavior. The full TPSPLINE model reported in this study consists of six dimensions. The predictor space is 4-dimensional, consisting of Regressors 1, 2, 3 and 4; and the response space is 2-dimensional, MAP and MAT. This figure shows only MAP based on Regressors 1 and 3.

and weathering in the profile. Furthermore, when residuals of predicted MAP and MAT are plotted with respect to one another they are uniformly scattered about zero, suggesting that $PPM_{1.0}$ is not biased towards predicting high-MAP, high-MAT or low-MAP, low-MAT climates.

### *Validations and Limitations of the $PPM_{1.0}$*

*Simulations and independent validations.*—The $PPM_{1.0}$ model was validated using a repeated random subsampling on 10,000 simulations. The results show prediction error estimates for MAP and MAT, separately, that are notably higher than the RMSE for both responses combined (table 4). Although the RMSE values are lower for CIA-K and the salinization ratio (Sheldon and others, 2002), one major disadvantage of these models is the limited range of response: $MAP_{range}$ = 179–1564 mm $y^{-1}$, $MAT_{range}$ = 1.6–23.4 °C. These ranges offer a much narrower climate prediction window than the $PPM_{1.0}$, which has a $MAP_{range}$ = 130–6866 mm $y^{-1}$ and $MAT_{range}$ = 0–26.8 °C and encompasses more soil orders and soil-forming factors, and a greater geochemical oxide range than Marbut (1935) data. The wide range of data used to construct the model likely contributes to the high predictive error in the model. But the wide range also improves overall model robustness by preventing gross underestimation of MAP and MAT.

To determine the utility of the combined PLSR and TPSPLINE approach in modeling climate using soil geochemical data, we applied the $PPM_{1.0}$ on the previously published subset of Marbut (1935) data (Sheldon and others, 2002) with one caveat. The element, $ZrO_2$, was removed from the $PPM_{1.0}$ because it was not measured in the

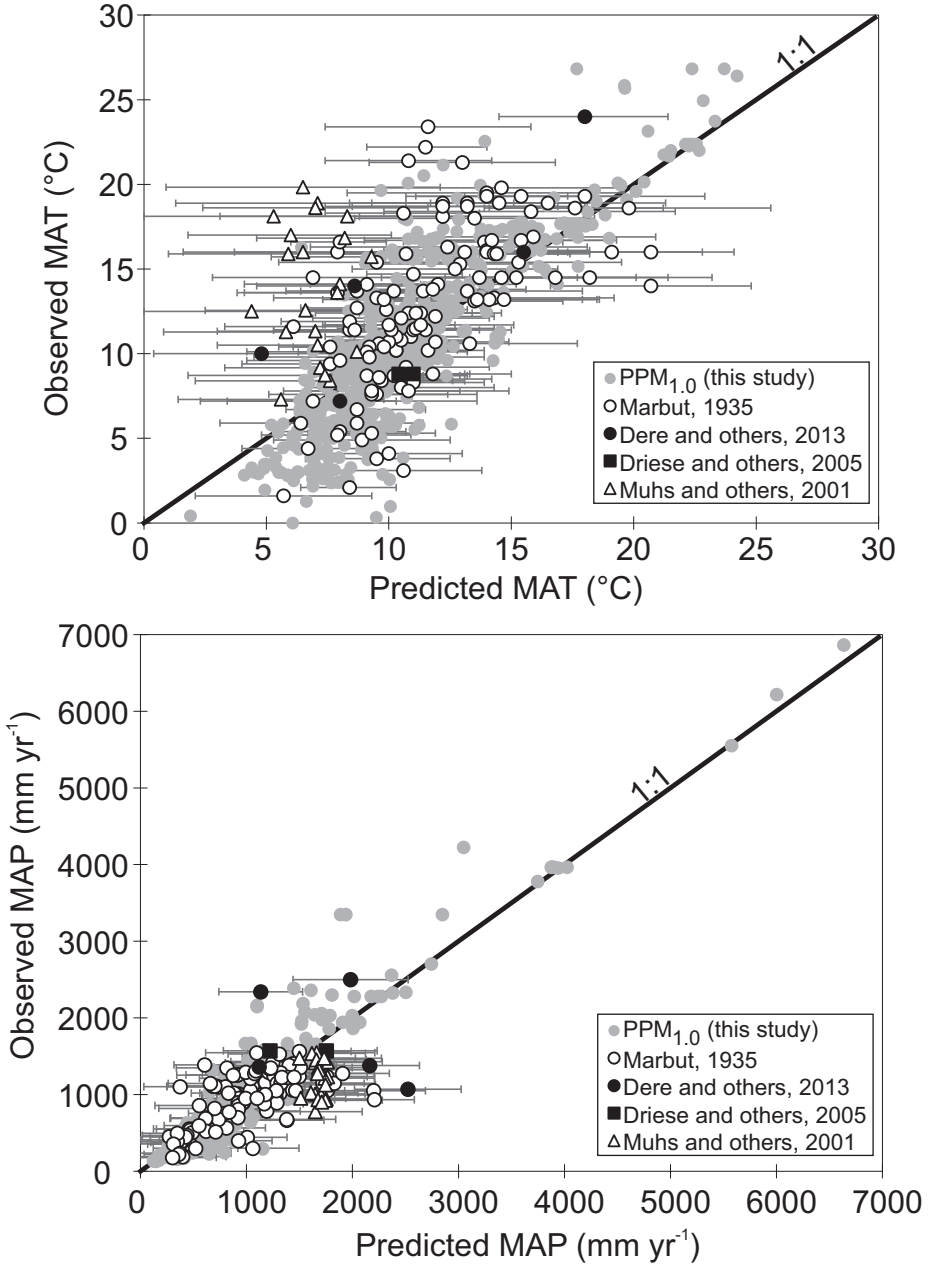


Fig. 12. Independent validation of MAP and MAT using the $PPM_{1.0}$. Observed MAT and MAP versus $PPM_{1.0}$ predicted MAP and MAT by individual study. The data are from a soil province study in the USA (Marbut, 1935), shale climosequence study (Dere and others, 2013), alluvial terrace soil study in the eastern USA (Driese and others, 2005), and a Midwestern USA soil-loess study (Muhs and others, 2001). The thick black line is 1:1 correlation between observed and predicted MAP and MAT.

Table 5

*Results from the independent validation of $PPM_{1.0}$ and errors for previously published models that use paleosol geochemistry to predict paleoclimate.*

| Model | $RMSE_{MAP}$ | $RMSE_{MAT}$ | $RMSPE_{MAP}$ | $RMSPE_{MAT}$ |
|---|---|---|---|---|
| $PPM_{1.0}$ | 228 | 2.46 | 512 | 3.98 |
| CIA-K | 181 | - | - | - |
| Alkali:alumina | - | 4.4 | - | - |
| CALMAG | 108 | - | - | - |
| PWI | - | 2.1 | - | - |

RMSE = Root Mean Squared Error, RMSPE = Root Mean Squared Prediction Error, Model source: $PPM_{1.0}$ = this study; CIA-K = Sheldon and others, 2002; Alkali:alumina = Sheldon and others, 2002; CALMAG = Nordt and Driese, 2010; PWI = Gallagher and Sheldon, 2013.

Marbut (1935) study. The $PPM_{1.0}$ results for Marbut data (from Sheldon and others, 2002) are RMSE MAP = 122.65 mm $yr^{-1}$ and RMSE MAT = 2.76 °C, which improves upon previously reported errors using the Marbut data set (see Sheldon and others, 2002 – MAP RMSE = 173 mm $yr^{-1}$; and Sheldon and Tabor, 2009, MAT RMSE = 4.4 °C). These results increase our confidence that the combined PLSR and TPSPLINE approach is an appropriate choice for modeling the diverse range of soils in this study. Although PLSR and TPSPLINE results used here yield higher MAP error than Sheldon and others (2002) reported MAP, the MAT error is lower than previously reported (Sheldon and Tabor, 2009).

The performance of $PPM_{1.0}$ was also assessed using shale climosequence soils (Dere and others, 2013), eastern USA alluvial terrace soils (Driese and others, 2005) and a latitudinal loess transect along the Mississippi valley (Muhs and others, 2001). These surface soils were actively weathering during sample collection and have all of the necessary geochemical oxides and the MAP and MAT are known.

The results of the independent validation show that larger predicted MAP and MAT values trend further away from the observed MAP and MAT values (that is, further away from 1:1 line) (fig. 12). The $PPM_{1.0}$ underestimates MAT for the Mississippi loess. This is likely related to the surprisingly high $Fe_2O_3/TiO_2$ values observed in the northern (colder) portion of the transect, which may be related to Holocene dust reactivation and deposition (Muhs and others, 2001).

*Model limitations.*—Despite the widely ranging environmental factors of the 685 B horizons used to build the $PPM_{1.0}$ model (fig. 4), modeling MAT for soils with more recent parent material additions remains a challenge. Alternatively, the poor prediction of MAT from the Mississippi loess may be due to analytical method used to generate the whole-soil geochemistry (X-ray fluorescence) versus the method used by the NRCS (ICP-AES). The NRCS training data used here are not anhydrous, that is, not corrected for loss-on-ignition (LOI), and thus users should not use the model if their data were LOI corrected.

There is a persistent overestimation of MAT for soils forming below ~6 °C, which suggests the soils have chemical weathering features more consistent with higher MAT values. A residual error analysis showed that the model consistently over-predicted MAT for soils weathering in glacial and subglacial till, with overestimated MATs of 1.79 and 2.98 °C, respectively. A correction factor was built into the PPM1.0 where the user can note glacial/subglacial till parent material, invoking the correction (see Appendix B and C).

This overestimation of MAT is at odds with a recent paleothermometer developed for forested paleosols that excluded soils below 6 °C because of slow weathering rates

(Gallagher and Sheldon, 2013). Chemical weathering in cold environments has been documented in the hyporheic exchange zone (Gooseff and others, 2002), and so chemical weathering may be concentrated in this zone adjacent to the stream (floodplain) and might explain the overestimation of MAT due to base loss and release of silica. With regards to the $PPM_{1.0}$ soils, it is possible that organic-rich podsols with a surplus of water can translocate abundant $Fe_2O_3$ to the subsoil, despite the low temperatures. This pedogenic pathway would result in an overestimation of MAT, given its direct correlation with MAT in Regressor 1. We hypothesize that more training set soils weathering in cold climates will likely yield new regressors that will improve overall $PPM_{1.0}$ performance.

The $PPM_{1.0}$ data set consists mostly of MAP values in the range of 400 to 1500 mm and MAT values in the range of 6 to 18 degrees Celsius, not unlike the Marbut (1935) dataset. MAP and MAT values greater than 1500 mm and 18 °C each make up ~10 percent of the $PPM_{1.0}$ dataset. Additionally, the PRISM spatial resolution is ~4 km and this could be a source of error in the $PPM_{1.0}$. Despite these limitations, $PPM_{1.0}$ RMSE values are superior to previous paleosol-MAT models built using a diverse range of soils. The $PPM_{1.0}$ MAP range extends far beyond current limits; and it is possible that previous studies may have underestimated paleo-MAP for some paleosols. Given these points, the $PPM_{1.0}$ should always be the first step when reconstructing climate transitions in geologic time, especially where paleosol properties are uncertain.

### *Application of $PPM_{1.0}$ to a Miocene Paleosol from Equatorial East Africa*

The $PPM_{1.0}$ was applied to a previously studied Miocene paleosol located in equatorial east Africa (Driese and others, 2016). This paleosol is the basal unit at the Ngira locality and consists of a 7.6 m thick succession of polygenetic, kaolinitic Bss horizons weathered on a mixture of basaltic and alluvial parent material. A lithologic discontinuity was documented in the Ngira paleosol and likely marks a burial event where the overlying paleosol is finer-grained than the underlying siltier paleosol. The Ngira paleo-Vertisol has very high $Fe_d$, indicative of Oxisols forming in an udic moisture regime (Udox). Thus, the Ngira was taxonomically classified as an Oxisolic Hapludert, where the Oxisolic modifier of Vertisols is not currently recognized in the USDA soil taxonomy (Soil Survey Staff, 2014).

For the Ngira paleosol, all paleo-MAP and paleo-MAT estimates using CIA-K, CALMAG, PWI and salinization are likely lower-range estimates due to the abundance of kaolinite clay and extreme depletion of base cations (Driese and others, 2016). Paleobotanical-based MAT estimates from a nearby fossil forest are much hotter, ranging from 22.6 °C to 34.5 °C (Michel and others, 2014), compared to the salinization and PWI-based MAT estimates, 16.4 °C and 15.7 °C, respectively (Driese and others, 2016). This discrepancy led Driese and others (2016) to suggest that the geochemical proxies are underestimating MAT.

We used the geochemistry from the Ngira Bss1 - 20 (Ngira-01) as data input into the $PPM_{1.0}$ (http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb/). The output is shown in table 6 and compared to previously published estimates (Driese and others, 2016). The best $PPM_{1.0}$ MAP estimates are higher than MAP and MAT determined from previously published proxies, and the highest $PPM_{1.0}$ MAT estimates are comparable with the paleobotanical-based MAT estimates for this region in Kenya (Michel and others, 2014).

## CONCLUSIONS

The data-driven PLSR and TPSPLINE modeling approach presented here predicts MAT and MAP for paleosols where uncertainty exists in the fossil soil-forming

TABLE 6

*Application of the $PPM_{1.0}$ and comparison of proxy estimates for a Miocene paleosol from equatorial East Africa (Driese and others, 2016). Paleobotanical-based MAT estimates from nearby site range from 22.6 °C to 34.5 °C (Michel and others, 2014).*

| Ngira paleosol horizon | Depth (cm) | MAP (mm $y^{-1}$) | | | | | MAT (°C) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | CIA-K | CALMAG | $PPM_{1.0}$ Low | $PPM_{1.0}$ Best | $PPM_{1.0}$ High | Salinization | PWI | $PPM_{1.0}$ Low | $PPM_{1.0}$ Best | $PPM_{1.0}$ High |
| Bss1 | 20 | 1309±172 | 1538±108 | 1299 | 1769 | 2238 | 16.4±4.4 | 15.7±2.1 | 16.8 | 20.9 | 25.0 |

environment. Despite the large range of variability in climate, biota, topography, parent material, and duration of pedogenesis, the paleosol-paleoclimate model ($PPM_{1.0}$) predicts MAT and MAP, with RMSEs of 228 mm and 2.46 °C, respectively. The model takes advantage of the strong correlation between $Fe_2O_3$, MAT and MAP. This is unique compared to more recent geochemical indices used to predict climate from paleosol geochemistry and resembles more traditional podsolization as well as newer lateritization indices (for example, Babechuck and others, 2014). Additionally, Al was typically used as a refractory element. Here, it is also utilized in podsolized soils. Like previous paleosol-climate proxies, base loss improves MAT and MAP estimates, but the $PPM_{1.0}$ also takes advantage of calcification/decalcification to improve MAP estimates. The $PPM_{1.0}$ also incorporates information on soils formed from mafic materials.

It is likely that if additional geochemical data were added to the $PPM_{1.0}$ training data new regressors may form to improve MAT and MAP prediction for soils in unique environments not well represented in the model space. In particular, more B horizon soils formed at higher MAP and MAT in tropical environments may improve the $PPM_{1.0}$ performance. Other needs include addition of Gelisols and related soils formed in cold-climate (MAT $<$ 0 °C), low-precipitation regions such as those which dominate at high latitudes today. The $PPM_{1.0}$ MAP range extends far beyond current limits; and it is possible that previous studies may have underestimated paleo-MAP for some paleosols. Given these points, the $PPM_{1.0}$ should always be the first step when reconstructing climate transitions in deep time where paleosol properties are uncertain, especially during intervals where ancient climates and atmospheres yield insight into future conditions.

ACKNOWLEDGMENTS

The authors thank Mike Wilson, Julia Kahmann-Robinson, Kimberley Kuijper, Ashlee Dere, Dan Peppe and the Baylor University (Department of Geology and College of Arts and Sciences) for their input, assistance and data during the development of this model. We thank Isabel Montanez, Gregory Retallack and Nathan Sheldon for providing helpful reviews and constructive input on an earlier version of this paper.

## APPENDICES

### APPENDIX A

Link to online supplementary data: http://earth.geology.yale.edu/%7eajs/SupplementaryData/2016/Stinchcomb/.

### APPENDIX B

The smoothing penalty, $\lambda J_m(f)$, and the basis functions used in constructing the $PPM_{1.0}$ are detailed herein. Although $\lambda \approx 0$ in the $PPM_{1.0}$, and hence the product of $\lambda J_m(f) \approx 0$, we show that for the $PPM_{1.0}$, where $d = 4$ and $m = 15$, $J_m(f)$ is defined as:

$$J_{15}(f) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum \frac{15!}{\alpha_1! \ldots \alpha_4!} \left( \frac{\partial^{15} f}{\partial R_1^{\alpha_1} \ldots \partial R_4^{\alpha_4}} \right)^2 dR_1 \ldots dR_4 \tag{9}$$

where $\alpha$ is a coefficient. Because $J_{15}(f) \approx 0$, the $f(\boldsymbol{R}_i)$ in equation (8) can be selected from a reproducing kernel Hilbert space and written as:

$$\hat{f_\lambda}(\boldsymbol{R}_i) = \theta_0 + \sum_{j=1}^{d} \theta_j R_{ij} + \sum_{j=1}^{p} \delta_j B_j(R_j) \tag{10}$$

The $\hat{f_\lambda}(\boldsymbol{R}_i)$ is represented as a combination of a sequence of basis functions, $B_j$, dependent on the location of the smoothing variables $R_1$-$R_4$, in ($\boldsymbol{R}_i$) and the coefficients $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ which are estimated by solving an $n \times n$ system (SAS Institute, Inc., 2008).

The $B_j$ is a logical function that orders the thin-plates in the spline surface, telling the model which plates touch in the overall model. Although not entirely accurate, this basis function can be visualized using figure 11, where each grid plate in the interpolated surface approximates a portion of $f(\boldsymbol{R}_i)$. In the case of figure 11, $f_\lambda(\boldsymbol{R}_i)$ is $f(R_1, R_2)$ that predicts a $y_i$ = MAP. The basis functions tell the model how each grid plate is positioned in this three-dimensional space.

## REFERENCES


Allen, D. M., 1974, The Relationship between Variable Selection and Data Augmentation and a Method for Prediction: Technometrics, v. 16, n. 1, p. 125–127, http://dx.doi.org/10.1080/00401706.1974.10489157

Amundson, R., Richter, D. D., Humphreys, G. S., Jobbágy, E. G., and Gaillardet, J., 2007, Coupling between biota and Earth materials in the Critical Zone: Elements, v. 3, n. 5, p. 327–332, http://dx.doi.org/10.2113/gselements.3.5.327

Aomine, S., and Wada, K., 1962, Differential weathering of volcanic ash and pumice, resulting in formation of hydrated halloysite: American Mineralogist, v. 47, n. 9–10, p. 1024–1048.

Atchley, S. C., Nordt, L. C., Dworkin, S. I., Ramezani, J., Parker, W. G., Ash, S. R., and Bowring, S. A., 2013, A linkage among Pangean tectonism, cyclic alluviation, climate change, and biologic turnover in the Late Triassic: The record from the Chinle Formation, southwestern United States: Journal of Sedimentary Research, v. 83, n. 12, p. 1147–1161, http://dx.doi.org/10.2110/jsr.2013.89

Babechuk, M. G., Widdowson, M., and Kamber, B. S., 2014, Quantifying chemical weathering intensity and trace element release from two contrasting basalt profiles, Deccan Traps, India: Chemical Geology, v. 363, p. 56–75, http://dx.doi.org/10.1016/j.chemgeo.2013.10.027

Bacon, A. R., Richter, D. D., Bierman, P. R., and Rood, D. H., 2012, Coupling meteoric $^{10}$Be with pedogenic losses of $^{9}$Be to improve soil residence time estimates on an ancient North American interfluve: Geology, v. 40, n. 9, p. 847–850, http://dx.doi.org/10.1130/G33449.1

Birkeland, P. W., 1999, Soils and geomorphology: Oxford, Oxford University Press, 430 p.

Bradley, R. S., 1999, Paleoclimatology: Reconstructing Climates of the Quaternary, *in* Dmowska, R., and Holton, J. R., editors, International Geophysical Series, Volume 68: San Diego, Academic Press, 2nd edition, 613 p.

Brady, N. C., and Weil, R. R., 2007, The Nature and Properties of Soils, 14th edition: Upper Saddle River, New Jersey, Prentice Hall, 980 p.

Breecker, D. O., Sharp, Z. D., and McFadden, L. D., 2009, Seasonal bias in the formation and stable isotopic composition of pedogenic carbonate in modern soils from central New Mexico, USA: Geological Society of America Bulletin, v. 121, n. 3–4, p. 630–640, http://dx.doi.org/10.1130/B26413.1

Buol, S. W., and Eswaran, H., 2000, Oxisols: Advances in Agronomy, v. 68, p. 151–195, http://dx.doi.org/10.1016/S0065-2113(08)60845-7

Burt, R., Wilson, M. A., Mays, M. D., and Lee, C. W., 2003, Major and trace elements of selected pedons in the USA: Journal of Environmental Quality, v. 32, n. 6, p. 2109–2121, http://dx.doi.org/10.2134/jeq2003.2109

Cerling, T. E., 1984, The stable isotopic composition of modern soil carbonate and its relationship to climate: Earth and Planetary Science Letters, v. 71, n. 2, p. 229–240, http://dx.doi.org/10.1016/0012-821X(84)90089-X

Chadwick, O. A., and Chorover, J., 2001, The chemistry of pedogenic thresholds: Geoderma, v. 100, n. 3–4, p. 321–353, http://dx.doi.org/10.1016/S0016-7061(01)00027-1

Chadwick, O. A., Gavenda, R. T., Kelly, E. F., Ziegler, K., Olson, C. G., Elliott, W. C., and Hendricks, D. M., 2003, The impact of climate on the biogeochemical functioning of volcanic soils: Chemical Geology, v. 202, n. 3–4, p. 195–223, http://dx.doi.org/10.1016/j.chemgeo.2002.09.001

Crovisier, J. L., Honnorez, J., Fritz, B., and Petit, J. C., 1992, Dissolution of sub glacial volcanic glasses from Iceland: Laboratory study and modeling: Applied Geochemistry, Supplementary Issue, n. 1, 55–81.

Daly, C., Neilson, R. P., and Phillips, D. L., 1994, A statistical-topographic model for mapping climatological precipitation over mountainous terrain: Journal of Applied Meteorology, v. 33, p. 140–158, http://dx.doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

Dere, A. L., White, T. S., April, R. H., Reynolds, B., Miller, T. E., Knapp, E. P., McKay, L. D., and Brantley, S. L., 2013, Climate dependence of feldspar weathering in shale soils along a latitudinal gradient: Geochimica et Cosmochimica Acta, v. 122, p. 101–126, http://dx.doi.org/10.1016/j.gca.2013.08.001

Driese, S. G., Li, Z. -H., and Horn, S. P., 2005, Late Pleistocene to Holocene climate and geomorphic histories as interpreted from 23,000 $^{14}$C yr B.P. paleosol and floodplain soils, southeastern West Virginia, USA: Quaternary Research, v. 63, n. 2, p. 136–149, http://dx.doi.org/10.1016/j.yqres.2004.10.005

Driese, S. G., Peppe, D. J., Beverly, E. J., DiPietro, L. M., Arellano, L. N., and Lehmann, T., 2016, Paleosols and paleoenvironments of the early Miocene deposits near Karungu, Lake Victoria, Kenya: Palaeogeography, Palaeoclimatology, Palaeoecology, v. 443, p. 167–182, http://dx.doi.org/10.1016/j.palaeo.2015.11.030

Duchon, J., 1976, Fonctions-Spline et Esperances Conditionnelles de Champs Gaussiens: Annales Scientifiques de l'Université de Clermont-Ferrand 2 Série Mathématiques, v. 14, p. 19–27.

——— 1977, Splines Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces, *in* Schempp, W., and

Zeller, K., editors, Constructive Theory of Functions of Several Variables, v. 571 of the series Lecture Notes in Mathematics: Berlin, Springer, p. 85–100.
Dworkin, S. I., Nordt, L., and Atchley, S., 2005, Determining terrestrial paleotemperatures using the oxygen isotopic composition of pedogenic carbonate: Earth and Planetary Science Letters, v. 237, n. 1–2, p. 56–68, http://dx.doi.org/10.1016/j.epsl.2005.06.054
Egli, M., Mirabella, A., Mancabelli, A., and Sartori, G., 2004, Weathering of soils in alpine areas as influenced by climate and parent material: Clays and Clay Minerals, v. 52, n. 3, p. 287–303, http://dx.doi.org/10.1346/CCMN.2004.0520304
Eswaran, H., and Bin, W. C., 1978, A study of a deep weathering profile on granite in Peninsular Malaysia: III. Alteration of feldspars: Soil Science Society of America Journal, v. 42, n. 2, p. 154–158, http://dx.doi.org/10.2136/sssaj1978.03615995004200010034x
Fernandez, A., Tang, J., and Rosenheim, B. E., 2014, Siderite 'clumped' isotope thermometry: A new paleoclimate proxy for humid continental environments: Geochimica et Cosmochimica Acta, v. 126, p. 411–421, http://dx.doi.org/10.1016/j.gca.2013.11.006
Gallagher, T. M., and Sheldon, N. D., 2013, A new paleothermometer for forest paleosols and its implications for Cenozoic climate: Geology, v. 41, n. 6, p. 647–650, http://dx.doi.org/10.1130/G34074.1
Gile, L. H., Peterson, F. F., and Grossman, R. B., 1966, Morphological and genetic sequences of carbonate accumulation in desert soils: Soil Science, v. 101, n. 5, p. 347–360, http://dx.doi.org/10.1097/00010694-196605000-00001
Goldschmidt, V. M., 1937, The principles of distribution of chemical elements in minerals and rocks. The seventh Hugo Müller Lecture, delivered before the Chemical Society on March 17th, 1937: Journal of the Chemical Society, p. 655–673, http://dx.doi.org/10.1039/JR9370000655
Gooseff, M. N., McKnight, D. M., Lyons, W. B., and Blum, A. E., 2002, Weathering reactions and hyporheic exchange controls on stream water chemistry in a glacial meltwater stream in the McMurdo Dry Valleys: Water Resources Research, v. 38, n. 12, p. 15-1 – 15-17, http://dx.doi.org/10.1029/2001WR000834
Gulbranson, E. L., Montanez, I. P., and Tabor, N. J., 2011, A proxy for humidity and floral province from paleosols: Journal of Geology, v. 119, n. 6, p. 559–573, http://dx.doi.org/10.1086/661975
Harnois, L., 1988, The CIW Index: A new chemical index of weathering: Sedimentary Geology, v. 55, n. 3–4, p. 319–322, http://dx.doi.org/10.1016/0037-0738(88)90137-6
Holliday, V. T., 2004, Soils in archaeological research: Oxford, Oxford University Press, 464 p.
Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., and Wickham, J., 2007, Completion of the 2001 National Land Cover Database for the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, n. 4, p. 337–341.
Hyland, E. G., Sheldon, N. D., Van der Voo, R., Badgley, C., and Abrajevitch, A., 2015, A new paleoprecipitation proxy based on soil magnetic properties: Implications for expanding paleoclimate reconstructions: Geological Society of America Bulletin, v. 127, n. 7–8, p. 975–981, http://dx.doi.org/10.1130/b31207.1
Jackson, M., and Sherman, G. D., 1953, Chemical weathering of minerals in soils: Advances in Agronomy, v. 5, p. 219–318, http://dx.doi.org/10.1016/S0065-2113(08)60231-X
Jenny, H., 1941, Factors of soil formation: A system of quantitative pedology: New York, Dover Publications, Inc., 281 p.
——— 1980, The soil resource: Origin and Behavior: Berlin, Springer Verlag, 377 p.
Ji, J. F., Chen, J., Balsam, W., Lu, H. Y., Sun, Y. B., and Xu, H. F., 2004, High resolution hematite/goethite records from Chinese loess sequences for the last glacial-interglacial cycle: Rapid climatic response of the East Asian Monsoon to the tropical Pacific: Geophysical Research Letters, v. 31, n. 4, p. L03207, http://dx.doi.org/10.1029/2003GL018975
Kampf, N., and Schwertmann, U., 1983, Goethite and hematite in a climosequence in southern Brazil and their application in classification of kaolinitic soils: Geoderma, v. 29, n. 1, p. 27–39, http://dx.doi.org/10.1016/0016-7061(83)90028-9
Kraus, M. J., 1999, Paleosols in clastic sedimentary rocks: Their geologic applications: Earth-Science Reviews, v. 47, n. 1–2, p. 41–70, http://dx.doi.org/10.1016/S0012-8252(99)00026-4
Ludvigson, G. A., Gonzalez, L. A., Metzger, R. A., Witzke, B. J., Brenner, R. L., Murillo, A. P., and White, T. S., 1998, Meteoric sphaerosiderite lines and their use for paleohydrology and paleoclimatology: Geology, v. 26, n. 11, p. 1039–1042, http://dx.doi.org/10.1130/0091-7613(1998)026<1039:MSLATU>2.3.CO;2
Ludvigson, G. A., Gonzalez, L. A., Fowle, D. A., Roberts, J. A., Driese, S. G., Villarreal, M. A., Smith, J. J., and Suarez, M. B., 2013, Paleoclimatic applications and modern process studies of pedogenic siderite, *in*, Driese, S. G., and Nordt, L. C., editors, New Frontiers in Paleopedology and Terrestrial Paleoclimatology: Paleosols and Soil Surface Analog Systems: SEPM Special Publications, v. 104, p. 79–87, http://dx.doi.org/10.2110/sepmsp.104.01
Maher, B. A., 1998, Magnetic properties of modern soils and Quaternary loessic paleosols: Paleoclimatic implications: Palaeogeography, Palaeoclimatology, Palaeoecology, v. 137, n. 1–2, p. 25–54, http://dx.doi.org/10.1016/S0031-0182(97)00103-X
Maher, B. A., and Thompson, R., 1995, Paleorainfall reconstructions from pedogenic magnetic susceptibility variations in the Chinese loess and paleosols: Quaternary Research, v. 44, n. 3, p. 383–391, http://dx.doi.org/10.1006/qres.1995.1083
Marbut, C. F., 1935, Atlas of American agriculture. III. Soils of the United States: Washington, D. C., Government Printing Office, Series n. 59.
Maxbauer, D. P., Feinberg, J. M., and Fox, D. L., 2016, Magnetic mineral assemblages in soils and paleosols as the basis for paleoprecipitation proxies: A review of magnetic methods and challenges: Earth-Science Reviews, v. 155, p. 28–48, http://dx.doi.org/10.1016/j.earscirev.2016.01.014
Mayer, L., McFadden, L. D., and Harden, J. W., 1988, Distribution of calcium carbonate in desert soils: A model: Geology, v. 16, n. 4, p. 303–306, http://dx.doi.org/10.1130/0091-7613(1988)016<0303:DOCCID>2.3.CO;2

Maynard, J. B., 1993, Chemistry of modern soils as a guide to interpreting Precambrian paleosols: The Journal of Geology, v. 100, n. 3, p. 279–289, http://dx.doi.org/10.1086/629632
McKeague, J. A., and Day, J. H., 1966, Dithionite- and oxalate-extractable Fe and Al as aids in differentiating various classes of soils: Canadian Journal of Soil Science, v. 46, n. 1, p. 13–22, http://dx.doi.org/10.4141/cjss66-003
Meinguet, J., 1979, Multivariate Interpolation at Arbitrary Points Made Simple: Zeitschrift für Angewandte Mathematik und Physik (ZAMP), v. 30, n. 2, p. 292–304, http://dx.doi.org/10.1007/BF01601941
Michel, L. A., Peppe, D. J., Lutz, J. A., Driese, S. G., Dunsworth, H. M., Harcourt-Smith, W. E. H., Horner, W. H., Lehmann, T., Nightingale, S., and McNulty, K. P., 2014, Remnants of an ancient forest provide ecological context for Early Miocene fossil apes: Nature Communications, v. 5, n. 3236, http://dx.doi.org/10.1038/ncomms4236
Muhs, D. R., Bettis, E. A., III, Been, J., and McGeehin, J. P., 2001, Impact of climate and parent material on chemical weathering in loess-derived soils of the Mississippi River Valley: Soil Science Society of America Journal, v. 65, n. 6, p. 1761–1777, http://dx.doi.org/10.2136/sssaj2001.1761
National Research Council, 2011, Understanding Earth's Deep Past: Lessons for Our Climate Future: Washington, D. C., The National Academies Press, 212 p.
Nesbitt, H. W., and Young, G. M., 1982, Early Proterozoic climates and plate motions inferred from major element chemistry of lutites: Nature, v. 299, p. 715–717, http://dx.doi.org/10.1038/299715a0
Nordt, L. C., and Driese, S. G., 2009, Hydropedological model of Vertisol formation along the Gulf Coast Prairie land resource area of Texas: Hydrology and Earth System Sciences, v. 13, n. 11, p. 2039–2053, http://dx.doi.org/10.5194/hess-13-2039-2009
——— 2010a, New weathering index improves paleorainfall estimates from Vertisols: Geology, v. 38, n. 5, p. 407–410, http://dx.doi.org/10.1130/G30689.1
——— 2010b, A modern soil characterization approach to reconstructing physical and chemical properties of paleo-Vertisols: American Journal of Science, v. 310, n. 1, p. 37–64, http://dx.doi.org/10.2475/01.2010.02
——— 2013, Application of the Critical Zone concept to the deep-time sedimentary record: Sedimentary Record, v. 11, n. 3, p. 4–9, http://dx.doi.org/10.2110/sedred.2013.3.4
Nordt, L., Orosz, M., Driese, S., and Tubbs, J., 2006, Vertisol carbonate properties in relation to mean annual precipitation: Implications for paleoprecipitation estimates: The Journal of Geology, v. 114, n. 4, p. 501–510, http://dx.doi.org/10.1086/504182
Nordt, L., Atchley, S., and Dworkin, S., 2015, Collapse of Late Triassic megamonsoon in western equatorial Pangea, present-day American Southwest: Geological Society of America Bulletin, v. 127, n. 11–12, p. 1798–1815, http://dx.doi.org/10.1130/B31186.1
Oskarsson, B. V., Riishuus, M. S., and Arnalds, O., 2012, Climate-dependent chemical weathering of volcanic soils in Iceland: Geoderma, v. 189–190, p. 635–651, http://dx.doi.org/10.1016/j.geoderma.2012.05.030
Passey, B. H., Levin, N. E., Cerling, T. E., Brown, F. H., and Eiler, J. M., 2010, High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates: Proceedings of the National Academy of Sciences of the United States of America, v. 107, n. 25, p. 11245–11249, http://dx.doi.org/10.1073/pnas.1001824107
Pavich, M. J., Brown, L., Valette-Silver, J. N., Klein, J., and Middleton, R., 1985, $^{10}$Be analysis of a Quaternary weathering profile in the Virginia Piedmont: Geology, v. 13, n. 1, p. 39–41, http://dx.doi.org/10.1130/0091-7613(1985)13<39:BAOAQW>2.0.CO;2
Rasmussen, C., Brantley, S., Richter, D., Blum, A., Dixon, J., and White, A. F., 2011, Strong climate and tectonic control on plagioclase weathering in granitic terrain: Earth and Planetary Science Letters, v. 301, n. 3–4, p. 521–530, http://dx.doi.org/10.1016/j.epsl.2010.11.037
Rawls, W. J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M., and Bloodworth, H., 2003, Effect of soil organic carbon on soil water retention: Geoderma, v. 116, n. 1–2, p. 61–76, http://dx.doi.org/10.1016/S0016-7061(03)00094-6
Reich, P. B., and Oleksyn, J., 2004, Global patterns of plant leaf N and P in relation to temperature and latitude: Proceedings of the National Academy of Sciences of the United States of America, v. 101, n. 30, p. 11001–11006, http://dx.doi.org/10.1073/pnas.0403588101
Reiche, P., 1950, A survey of weathering processes and products: University of New Mexico Publications in Geology, n. 3, 95 p.
Retallack, G. J., 1994, The environmental factor approach to the interpretation of paleosols, *in* Amundson, R., Harden, J., and Singer, M., editors, Factors of soil formation – A Fiftieth Anniversary Retrospective: Soil Science Society of America Special Publication, n. 33, p. 31–64, http://dx.doi.org/10.2136/sssaspecpub33.c3
——— 1997, A Colour Guide to Paleosols: Chichester, Wiley, 175 p.
——— 2001, Soils of the past: An Introduction to Paleopedology, Second Edition: Oxford, Blackwell Science, 404 p., http://dx.doi.org/10.1002/9780470698716
——— 2005, Pedogenic carbonate proxies for amount and seasonality of precipitation in paleosols: Geology, v. 33, n. 4, p. 333–336, http://dx.doi.org/10.1130/G21263.1
Retallack, G. J., and Huang, C., 2010, Depth to gypsic horizon as a proxy for paleoprecipitation in paleosols of sedimentary environments: Geology, v. 38, n. 5, p. 403–406, http://dx.doi.org/10.1130/G30514.1
Ruxton, B. P., 1968, Measures of the degree of chemical weathering of rocks: The Journal of Geology, v. 76, n. 5, p. 518–527, http://dx.doi.org/10.1086/627357
SAS Institute Inc., SAS 9.1.3 Help and Documentation, Cary, NC: SAS Institute Inc., 2002–2004.
SAS Institute, Inc., 2008, SAS User's Guide: Statistics, Version 9.3 Edition: Cary, NC: SAS Institute, Inc.
Schaetzl, R. J., and Thompson, M. L., 2015, Soils Genesis and Geomorphology, Second Edition: New York, Cambridge University Press, 778 p.

Sheldon, N. D., 2006, Precambrian paleosols and atmospheric $CO_2$ levels: Precambrian Research, v. 147, n. 1–2, p. 148–155, http://dx.doi.org/10.1016/j.precamres.2006.02.004

Sheldon, N. D., and Tabor, N. J., 2009, Quantitative paleoenvironmental and paleoclimatic reconstruction using paleosols: Earth-Science Reviews, v. 95, n. 1–2, p. 1–52, http://dx.doi.org/10.1016/j.earscirev.2009.03.004

Sheldon, N. D., Retallack, G. J., and Tanaka, S., 2002, Geochemical climofunctions from North American soils and application to paleosols across the Eocene-Oligocene boundary in Oregon: The Journal of Geology, v. 110, n. 6, p. 687–696, http://dx.doi.org/10.1086/342865

Soil Survey Division Staff, 1993, Soil Survey Manual: Washington, D. C., US Government Printing Office, USDA-NRCS Agriculture Handbook 18, 437 p.

Soil Survey Staff, 1999, Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys: Lincoln, Nebraska, Natural Resources Conservation Service, U. S. Department of Agriculture Handbook, n. 436, Second Edition, 886 p.

——— 2014, Kellogg Soil Survey Laboratory Methods Manual, Soil Survey Investigations Report No. 42, Version 5.0, Burt, R., and Soil Survey Staff, editors: Lincoln, Nebraska, U.S. Department of Agriculture, Natural Resources Conservation Service, 1003 p.

Sparks, D. L., 2002, Environmental Soil Chemistry, Second Edition: Amsterdam, Academic Press, 352 p.

Stefánsson, A., and Gislason, S. R., 2001, Chemical weathering of basalts, Southwst Iceland: Effect of rock crystallinity and secondary minerals on chemical fluxes to the ocean: American Journal of Science, v. 301, n. 6, p. 513–556, http://dx.doi.org/10.2475/ajs.301.6.513

Tabor, N., J., and Myers, T. S., 2015, Paleosols as indicators of paleoenvironment and paleoclimate: Annual Review of Earth and Planetary Sciences, v. 43, p. 333–361, http://dx.doi.org/10.1146/annurev-earth-060614-105355

van Breemen, N., Mulder, J., and Driscoll, C. T., 1983, Acidification and alkalization of soils: Plant and Soil, v. 75, n. 3, p. 283–308, http://dx.doi.org/10.1007/BF02369968

Vogt, T., 1927, Sulitjelmafeltets geologi og petrografi: Norges Geologiske Undersokelse, v. 121, p. 1–560.

White, A. F., and Blum, A. E., 1995, Effects of climate on chemical-weathering in watersheds: Geochimica et Cosmochimica Acta, v. 59, n. 9, p. 1729–1747, http://dx.doi.org/10.1016/0016-7037(95)00078-E

Williams, J. Z., Bandstra, J. Z., Pollard, D., and Brantley, S. L., 2010, The temperature dependence of feldspar dissolution determined using a coupled weathering-climate model for Holocene-aged loess soils: Geoderma, v. 156, n. 1–2, p. 11–19, http://dx.doi.org/10.1016/j.geoderma.2009.12.029

Wilson, M. A., Burt, R., Indorante, S. J., Jenkins, A. B., Chiaretti, J. V., Ulmer, M. G., and Scheyer, J. M., 2008, Geochemistry in the modern soil survey program: Environmental Monitoring and Assessment, v. 139, n. 1, p. 151–171, http://dx.doi.org/10.1007/s10661-007-9822-z

Wold, S., Sjostrom, M., and Eriksson, L., 2001, PLS-regression: A basic tool of chemometrics: Chemometrics and Intelligent Laboratory Systems, v. 58, n. 2, p. 109–130, http://dx.doi.org/10.1016/S0169-7439(01)00155-1

Yaalon, D. H., 1971, Soil-forming processes in time and space, *in* Yaalon, D. H., editor, Paleopedology: Origin, nature and dating of paleosols, Papers of the Symposium on the age of parent materials and soils, Amsterdam, Netherlands, August 10-15, 1970: Jerusalem, International Society of Soil Science and Israel Universities Press, p. 29–40.