Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Genome-enabled Prediction of Complex Traits with Kernel Methods: What Have We Learned?

D. Gianola[1]**, G. Morota**[1] **and J. Crossa**[2]
[1]University of Wisconsin-Madison, USA, [2]International Maize and Wheat Improvement Center, Mexico

ABSTRACT: Complex traits are presumably affected by several genomic regions acting in some concerted manner (epistasis), by non-linearity between genome and phenotypes stemming from enzyme kinetics, and by interactions with environmental forces. Prompted by these considerations, non-parametric approaches entered into quantitative genetics early in the 21st century, and a decade of experience has been accumulated, mostly in animals and plants. Some developments are reviewed in this paper, and areas for additional investigation are discussed.

Keywords: complex traits; genome-enabled prediction; non-parametric regression

Introduction

The emergence of a massive number of molecular markers has prompted an enormous amount of research aimed at exploring genome-wide associations with continuous and meristic (typically disease) traits, an inferential task, as well as prediction of outcomes. Inference is cursed by dimensionality (the number of markers, *p*, exceeds the number of observations in the sample, *n*). Following Meuwissen et al. (2001), genome-enabled prediction has become important in animal and plant breeding, and it may play a role in personalized medicine, e.g., de los Campos et al. (2010a) and Vazquez et al. (2012). A relevant area of research is that of finding a "prediction machine" that is flexible and robust (stable) with respect to type of input data, gene action and environmental circumstances, and that is amenable to routine computing as required in applied genetics.

Much effort has been done in developing Bayesian linear regression models differing in their prior distribution, with the objective of capturing varied forms of "genetic architecture". Gianola (2013) argued that this is possibly futile for complex traits because at most *n* parameters can be identified in a likelihood, so that *p-n* complementary "parameters" are redundant (although their posterior distributions exist if priors are proper). The process of Bayesian learning is imperfect here because the prior **IS** influential with mathematical certainty. In a view of this, Kempthorne (1972) stated: *"reporting only of Bayesian estimates, each based on the prior of the person who obtained them, will butcher the processes of science"*. Apart from this serious matter, a linear regression model should not be taken more seriously than as a local approximation. As a metaphor, it may be useful to regard the earth as flat within the range of our visual field, and plan accordingly; however, this perception would be mechanistically inaccurate. In that vein, quantitative genetics is mainly based on the assumption of a linear function of allelic effects and, when the number of loci goes to infinity (plus linkage equilibrium, LE, assumptions) the infinitesimal model results. With a finite number of loci (presumably "causal variants"), substitution effects can be viewed as partial derivatives on allelic content. This linear approximation accounts for most genetic variance even in well-characterized epistatic systems (Hill et al., 2008). While this finding reassures us that the additive model offers a good first-order local approximation, it does set an upper limit to the potential of our theory from a discovery perspective: almost everything turns out to be additive when, in fact, it is not. As Rousseau suggested in "*La Nouvelle Heloise"*, obviously in another context, the additive model seems to deny what it is, explaining what it is not.

Based on the above, it is doubtful whether it may ever be possible to understand complex traits using multiple linear regressions. We have argued (Gianola et al., 2006; Gianola and van Kaam, 2008) that machine learning methods have been used successfully in fields where either complex problems (e.g., image reconstruction) or a lack of good theory (e.g., economics) are encountered. Based on our experience with reproducing kernel Hilbert spaces regression or RKHS (Gonzalez-Recio et al., 2008, 2009; de los Campos et al., 2009, 2010; Perez-Rodriguez et al., 2012) and neural networks (Gianola et al., 2011; Okut et al., 2011, 2013; Gonzalez-Camacho et al., 2012; Tusell et al., 2013) we have found that the latter are unstable in performance as their propensity to over-fitting is hard to temper by regularization, unless variable selection is conducted concomitantly. On the other hand, kernel-based methods have been used extensively for regression and classification (Wahba, 1990; Vapnik, 1998; Shawe-Taylor and Cristianini, 2004), because of their capacity for delivering accurate predictions if properly tuned. Based on our empirical experience with animal and plant data sets, it appears that RKHS is one of the best prediction machines for genome-based prediction.

In this paper, we review essentials of the RKHS approach and recent developments, discuss some of the evidence and suggest areas where additional research may be fruitful.

RKHS in a Nutshell

General. Most quantitative genetic methodology assumes (for a covariate-homogeneous population) that the phenotype is the result of at most three "independent" factors: genotype (G), environment (E) and genotype-environment interaction (GE). If G and E are associated, no orthogonal partition of variance is attainable; the same happens with lack of LE: the attribution of variance to a given locus is ambiguous (e.g., Gianola et al., 2013). Ignoring GE, one writes $\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e},$ in an obvious notation. Since $\boldsymbol{g}$ is not observable, proxy variables or instruments are used, such as markers $\boldsymbol{x}$ and we replace $\boldsymbol{g}$ by some function of markers $f(\boldsymbol{x})$. Then, the model residual changes to $\boldsymbol{\varepsilon} = \boldsymbol{g} - f(\boldsymbol{x}) + \boldsymbol{e}$, where $\boldsymbol{g} - f(\boldsymbol{x})$ is a misspecification error, expected to behave non-randomly, as opposed to $\boldsymbol{e}$. Breiman (2001) emphasized that it is crucial to ensure that $\boldsymbol{\varepsilon}$ behaves randomly and has suggested "debiasing" techniques that animal and plant breeders have ignored in their obsession with *"bigger is better"*. The standard representation of $f(\boldsymbol{x})$ is often a linear regression on markers or a linear regression on additive relationships (de los Campos et al., 2009; Gianola et al., 2011).

The basic idea underpinning RKHS (Kimeldorf and Wahba, 1971) is that, given $\mathbf{x}$, the best (in some precisely defined sense) approximation to $\boldsymbol{g}$ can be found by solving the random effects model $\boldsymbol{y} = \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{h})\boldsymbol{\alpha} + \boldsymbol{\epsilon}$, where $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{h})$ is an $n \times n$ positive semi-definite symmetric matrix and $\boldsymbol{h}$ is a vector of one or more "bandwidth" parameters; $\boldsymbol{\alpha} \sim (\boldsymbol{0}, \boldsymbol{K}^{-1}\sigma_{\alpha}^2)$ is a vector of regression coefficients (we will often ignore $\boldsymbol{h}$ in subsequent notation, but it "is there") with σ_{α}^2 being a variance parameter; $\boldsymbol{\epsilon} \sim (\boldsymbol{0}, \boldsymbol{R}\sigma_{\epsilon}^2)$, where σ_{ϵ}^2 is the residual variance and $\boldsymbol{R}$ is some matrix (typically an identity matrix). It is crucial to recognize some points: 1) the problem reduces to finding n regression coefficients instead of p, as in the family of linear regressions known as the "Bayesian alphabet". 2) There is a "primal" and a "dual" representation of the problem. For example (e.g., de los Campos et al., 2009), it can be shown that GBLUP and "ridge regression-BLUP" are primal and dual representations of a problem where the kernel is proportional to $\boldsymbol{XX}'$, where $\boldsymbol{X}$ is an $n \times p$ marker matrix. Likewise, if a numerator relationship matrix $\boldsymbol{A}$ is adopted as kernel, one ends up with representations of BLUP involving either the mixed model equations of Henderson, or the "strong arm" approach involving the inverse of the phenotypic covariance matrix. This is a consequence of the fact that the primal and dual representations induce the same marginal and conditional distributions. That is why one can easily go from GBLUP to BLUP of marker effects, and vice-versa. This result, shown by Henderson (1977) was rediscovered by Goddard (2008) and by Janss et al. (2012); sometimes it is useful to revisit "old". 3) The RKHS paradigm does not guide on how $\boldsymbol{K}$ is to be chosen, a crucial issue in arriving at good predictions.

Kernel forms. The kernel creates a similarity in features among individuals, even if genetically unrelated. For example, if additive relationships (ignoring inbreeding, for simplicity) are inputs, a valid kernel is $\boldsymbol{A} = \{a_{ij}\}$ and another kernel could be $\boldsymbol{K} = \left\{\exp(-h\frac{a_{ij}{}^2}{\max\left(a_{ij}{}^2\right)})\right\}$ where *max* (.) is the maximum $a_{ij}{}^2$ in the data set. The latter kernel produces a non-linear transformation of relationships; this may (may not) deliver a better predictive performance, and whether or not this conveys meaning with respect to some theory is immaterial from a predictive perspective. Valente et al. (2014, this volume) argue that causality (inference) and predictive tasks are different matters. It is well known (Takezawa, 2005) that a "causal model" may provide bad predictions if too richly parameterized, because of over-fitting propensity. The issue here is to exploit complexity, as opposed to understanding it, which might be a formidable task, e.g., gene × gene × gene × gene × gene interactions (the Krebbs cycle includes 12 different enzymes, and recall that "one gene-one enzyme").

The most widely used kernel is the Gaussian. Suppose a DNA sequence is divided into 3 sections, e.g., exonic, intronic and inter-genic, such that an individual has string $(\boldsymbol{x}_E{}', \boldsymbol{x}_I{}', \boldsymbol{x}_{IG}{}')'$. A measure of similarity based on squared Euclidean distance between sequences i and j could take the form $k_{ij} = \left\{exp\left[-\left(\frac{(\boldsymbol{x}_{E,i} - \boldsymbol{x}_{E,j})'(\boldsymbol{x}_{E,i} - \boldsymbol{x}_{E,j})}{h_E}\right)\right] \times exp\left[-\left(\frac{(\boldsymbol{x}_{I,i} - \boldsymbol{x}_{I,j})'(\boldsymbol{x}_{I,i} - \boldsymbol{x}_{I,j})}{h_I}\right)\right] exp\left[-\left(\frac{(\boldsymbol{x}_{IG,i} - \boldsymbol{x}_{I,j})'(\boldsymbol{x}_{IG,i} - \boldsymbol{x}_{I,j})}{h_{IG}}\right)\right]\right\}$. An alternative is the *t*-kernel presented by Tusell et al. (2014). Ignoring the distinction between regions, this kernel for a string with p markers is $k_{ij} = \left[1 + \frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)'\Sigma(\boldsymbol{x}_i - \boldsymbol{x}_j)}{p\vartheta}\right]^{-\frac{(\vartheta+1)}{2}}$ where ϑ (the degrees of freedom) enters as a bandwidth parameter and $\sum$ is a positive-definite matrix of weights. Mathematicians have developed methods for estimating the kernel, e.g., based on the Matérn covariance function. The Gaussian kernel is good enough most often (Ober et al., 2010). When inputs are discrete (e.g., strings of markers), theory rules out continuous kernels. Morota et al. (2013) evaluated diffusion kernels for discrete inputs with animal and plant data, and compared these with the Gaussian. Differences in predictive ability were minimal; this is fortunate because computing the diffusion kernel is time consuming.

Genomic data are heterogeneous and annotation may guide kernel construction. Morota et al. (2014a) used SNP annotation (e.g., SNPs in coding regions or in introns or in inter-genic regions) in a predictive analysis of broiler traits. The different types of genomic information affected predictive performance, but whole-genome prediction (ignoring annotation) was good enough for practical purposes. This may not be a universal finding.

The importance of choosing a good kernel (akin to selecting a good model) is illustrated in a study by Konstantinov and Hayes (2010) with dairy cattle. They compared two RKHS models with GBLUP; the second RKHS implementation was uniformly better than GBLUP, but the first did not deliver a good predictive performance.

Kernels encoding non-additivity. Using dairy cattle data, Morota et al. (2014b) attempted to capture dominance by fitting additive and dominance Gaussian (or parametric) kernels together. For additive kernels (A), coding was 0, 1 and 2 for aa, Aa and AA, respectively. Coding genotypes as -0.5, 0.5 and -0.5, respectively, led to a dominance (D) kernel. The parametric kernels were the standard genomic relationship matrix GA and its dominance counterpart GD derived as in Su et al. (2012). A third kernel (parametric or Gaussian), aimed at capturing additive by dominance epistasis, was constructed by taking Hadamard products of matrices, e.g., GKA#GKD for the Gaussian kernels, following Henderson (1985). In the parametric case this assumes no linkage and linkage LE. The parametric version of the additive by dominance epistasis kernel is GA#GD. Morota et al. (2014b) observed a dominance contribution to variance when estimated breeding value was the target trait, which is counter-intuitive. This perplexing result was investigated by simulation where average adjacent linkage disequilibrium (LD) was 0.18 (r^2 metric). Genotypes under LE were created for n = 4,482, while varying p from 150 to 40,000. The off-diagonals of GD became more strongly correlated with those of GA when a larger number of SNPs was used for constructing the kernels. This highlights that a partition of marked variance into additive and dominance components is difficult to attain under LD, producing misleading results. Correlations between off-diagonals of the additive and dominance relationship matrices were small with LE. Perhaps the variance estimates reported by Su et al. (2012) and Morota et al. (2014b) are affected by lack of orthogonality between kernels. As observed in these studies, a sizable gain cannot be achieved with prediction models aiming at exploiting non-additive genetic variation when naively structured kernels are used. Genomic relationship kernels (standard or Gaussian) that are "orthogonal" to each other may enhance predictive ability but it is unclear how such kernels are to be constructed. Unfortunately, stylized models of quantitative genetics break down under LD (Gallais, 1974; Weir and Cockerham, 1976). For the time being, neither the parametric or non-parametric approaches explored so far can go much beyond exploiting additivity, so Hill et al. (2008) hold again.

Multi-kernel models. Perhaps the more robust approach to building a RKHS machine is a multi-kernel specification. There is little theory on this approach, seemingly recent in machine learning (Bach et al., 2004; Gönen and Alpaydin, 2011). Gianola and van Kaam (2008) described an implicit multi-kernel using both pedigree and markers, and de los Campos et al. (2010b) formalized the approach by fitting several Gaussian kernels differing in bandwidth parameter. Here, a "global" kernel creates strong commonality among all individuals, whereas a "local" kernel allow information borrowing only from individuals that are very similar, e.g., in molecular profiles. The idea is that some kernel captures a part of a pattern that is not detected by other kernels. An example of a multi-kernel model is given by $\boldsymbol{y} = \boldsymbol{Ag} + \boldsymbol{K}(\boldsymbol{x},\boldsymbol{h})\boldsymbol{\alpha} + \boldsymbol{\epsilon}$, where $\boldsymbol{g} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{A}^{-1}\sigma_a^2)$, $\boldsymbol{\alpha} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{K}^{-1}\sigma_\alpha^2)$ and $\boldsymbol{\epsilon} \sim (\boldsymbol{0}, \boldsymbol{R}\sigma_\epsilon^2)$ are mutually independent. Note that $\boldsymbol{Ag} = \boldsymbol{u}_p$ has the same distribution as the infinitesimal breeding value, as captured by a pedigree. If $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{h}) = \boldsymbol{G}$ is a genomic relationship matrix constructed using additive codes for markers $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{h})\boldsymbol{\alpha} = \boldsymbol{u_m}$ is interpretable as a molecularly marked breeding value. Sometimes, concern is expressed about redundancy between $\boldsymbol{A}$ and $\boldsymbol{G}$**.** Apart from the fact that such view is debatable, why would one refuse using the two kernels together if this leads to a better predictive performance? If the objective is to obtain more accurate predictions, stricture from theory may not be useful. Examples of how combined use of $\boldsymbol{A}$ and $\boldsymbol{G}$ enhances predictive ability are in Erbe et al. (2010) and Rodriguez-Ramilo et al. (2014). The latter authors employ an approach similar to the one in de los Campos et al. (2010), but differs in that, instead of fitting 2 variance components (one per kernel), a single variance plus a parameter λ that "averages" $\boldsymbol{A}$ and $\boldsymbol{G}$ are used. The weight placed on marker versus pedigree-based information was inferred from a Bayesian MCMC model, and their method was assessed with a many SNPs, a large sample and 5 dairy traits. Results indicated that when a larger weight was given to $\boldsymbol{A}$ the predictive correlation was lower than when more weight was placed on $\boldsymbol{G}$. Importantly, the posterior mean of λ was always near the maximum of 1 (all weight on $\boldsymbol{G}$).

Consider the following multi-kernel representation with *c+2* kernels: $\boldsymbol{y} = \boldsymbol{Ag} + \boldsymbol{G\alpha}_g + \sum_{\boldsymbol{i=1}}^{\boldsymbol{c}} \boldsymbol{K_i}(\boldsymbol{x},\boldsymbol{h_i})\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}$**.**

Kernel $\boldsymbol{A}$ captures infinitesimal effects; $\boldsymbol{G}$ accounts for additive effects of markers, and the c kernels $\boldsymbol{K_i}$ could be Gaussian kernels with varying bandwidth parameters. Even if inputs are just additive marker codes, the Gaussian kernels make non-linear transformations of such inputs, hopefully capturing epistasis that might be relevant to the prediction problem. It may seem mystifying that a kernel on additive inputs encodes epistasis. An explanation can be motivated by considering a two-locus linear model with interaction, as follows: $y = x_1\beta_1 + x_2\beta_2 + x_1x_2\beta_{12} + e$, where the *x's* denote the number of copies of a certain allele at the appropriate locus. There is epistasis here because the effect of an allelic substitution at the first locus depends on the number of copies of alleles at the second locus (and vice-versa), that is, $\frac{\delta y}{\delta x_1} = \beta_1 + x_2\beta_{12}$. If a Gaussian kernel is employed, the model for an individual possessing marker string $\boldsymbol{x}$ becomes $y(\boldsymbol{x}) = \sum_{i=1}^{n} exp\left(\frac{(\boldsymbol{x}-\boldsymbol{x_i})'(\boldsymbol{x}-\boldsymbol{x_i})}{h}\right) + \epsilon$, and an allelic substitution at marker j in the string has impact $\frac{\delta y}{\delta x_j} = 2\sum_{i=1}^{n} exp\left(\frac{(\boldsymbol{x}-\boldsymbol{x_i})'(\boldsymbol{x}-\boldsymbol{x_i})}{h}\right)\left(x_j - x_{ij}\right)$. Hence, the non-linear transformation encodes a type of epistasis that is not represented by the linear modes of, say, Cockerham (1954).

In theory, epistasis is important for phenotypic prediction, but less so for selective breeding. This is because epistasis involving additive effects is transmitted from parents to offspring according to the rule $\left(\frac{1}{2}\right)^m$. For instance, for additive × additive × additive epistasis, *m=3*. This rule holds in some ideal world with absence of selection, mutation, assortative mating and linkage (Kempthorne, 1954), so it is not known with certainty how epistasis is transmitted in finite populations undergoing artificial or natural evolutionary processes. The *"additivity rules the waves"* from Hill et al. (2008) should be taken with caution, as it may be more reflective of limitations of variance components models than of the biology underpinning a trait. Prudent construction of predictive machines should not rule out epistasis *a priori.*

Akdemir (2014) compared single with multi-kernel approaches in five data sets representing wheat, mice, barley, rice and maize, and concluded that "similar or better accuracies were observed for a number of populations compared to single kernel models".

Parameter estimation and predictive ability. Bayesian implementations of RKHS are in Gianola and van Kaam (2008) and can be run in the BGLR package of de los Campos and Perez-Rodriguez (http://bglr.r-forge.r-project.org). The Bayesian approach requires creative assignment of priors to all parameters. An alternative is a likelihood-based machinery. If regression coefficients are taken as jointly normal, the likelihood function stems from the marginal distribution $\boldsymbol{y}|\sigma_\alpha^2, \sigma_\epsilon^2, \boldsymbol{h} \sim (\boldsymbol{0}, \boldsymbol{K}\sigma_\alpha^2 + \boldsymbol{R}\sigma_\epsilon^2)$, and the variance and bandwidth parameters can be estimated by maximum likelihood. Liu et al. (2007) give a fairly standard implementation of restricted maximum-likelihood with an application to prostate cancer. They had data on age of patient and tumor differentiation plus information on some pathway involving 5 genes. The pathway information was embedded into a Gaussian kernel with a single bandwidth parameter. In our experience, the asymptotic correlation between estimates of h and of the variance σ_α^2 is strong, which is also revealed when MCMC scans for these two parameters are compared. Also, the bandwidth parameter may be estimated with great uncertainty, so it seems dangerous to take a point estimate as if it were "true". A practical course of action is use of a grid of h values and, for each h, estimate σ_α^2 and carry out a cross-validation. After all, our concern is predictive ability as opposed to inference since, as argued before, it is doubtful that much inferential meaning can be extracted from complex systems via blatantly over-parameterized models.

Assuming values of the dispersion and bandwidth parameters have been arrived at, the BLUP of $\boldsymbol{\alpha}$ and of the marked signal $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{h})\boldsymbol{\alpha}$ can be obtained by the standard "strong-arm" method:

$$BLUP(\boldsymbol{\alpha}|\sigma_\alpha^2,\sigma_\epsilon^2,\boldsymbol{h})=\left(\boldsymbol{K}+\boldsymbol{R}\frac{\sigma_\epsilon^2}{\sigma_\alpha^2}\right)^{\boldsymbol{-1}}\boldsymbol{y}\,,$$

This shows that the inverse of $\boldsymbol{K}$ is not needed. If one uses Henderson's equations (assume fixed effects have been accounted for in some pre-processing, but of course can be included in the model), the RKHS solution can also be calculated as

$$\hat{\alpha}=\left(\boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{K}+\boldsymbol{K}\frac{\sigma_\epsilon^2}{\sigma_\alpha^2}\right)^{\boldsymbol{-1}}\boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{y}$$

$$=\left(\boldsymbol{R}^{-1}\boldsymbol{K}+\boldsymbol{I}\frac{\sigma_\epsilon^2}{\sigma_\alpha^2}\right)^{\boldsymbol{-1}}\boldsymbol{R}^{-1}\boldsymbol{y}.$$

Since BLUP is linearly invariant, the genetic signal is fitted as $\boldsymbol{K}\hat{\alpha}$. Extension of likelihood-BLUP machinery to multiple kernels is not innovative as the principles of mixed model methodology have been well established by decades.

Typically, predictive ability is assessed using some cross-validation scheme (its design is more an art than a science) where one divides data into training and testing sets, with random repeats or bootstrapping, to measure uncertainty. The model is fitted with training data and phenotypes are "masked" in the testing set. Phenotypes $\boldsymbol{y}_{Test}$ in the testing set are predicted as $\hat{\boldsymbol{y}}_{Test}=\boldsymbol{K}_{Test,Train}$ $\hat{\alpha}$. Subsequently, $\boldsymbol{y}_{Test}$ and $\hat{\boldsymbol{y}}_{Test}$ are used to measure "accuracy" of prediction, often via some correlation metric. A large predictive correlation does not necessarily reflect accuracy of prediction. For instance, methods A and B for predicting time to death after a cancer is metastasized may produce a correlation of 0.8, but say that A over-predicts by an average of 2 years. Obviously, B is accurate whereas A is not. Also, animal and plant breeders are often interested in tail or "center" behavior. Hence, examination of confusion matrices (Jimenez-Montero et al., 2013) or use of classification algorithms (Ornella et al., 2013) may be more informative than across the board mean squared error or correlation.

Model averaging. In Bayesian and classical theory (e.g., Hoeting et al., 1999; Sorensen and Gianola, 2002) it is well established that averaging over models can produce more accurate predictions (in the sense of closeness, not correlation) than use of a single model. Claeskens and Hjort (2008) wrote: *"Most [model] selection strategies work by assigning a certain score to each candidate model. In some cases there might be a clear winner, but sometimes these scores might reveal that there are several candidates that do almost as well as the winner. In such cases there may be considerable advantage in combining inference output across these best models".* Tusell et al. (2014) examined the ability of predicting yet-to-be observed litter size (pig) and grain yield (wheat) records using several RKHS regression models with different numbers of Gaussian or *t* kernels. Predictions were combined using three different types of model averaging: (i) mean of predicted phenotypes obtained in each model, (ii) weighted average using the reciprocal of mean squared error in a tuning set (training-tuning-testing design) as weight, or (iii) using the marginal likelihood as weight, estimated via MCMC. Phenotypes were 2598, 1604 and 1879 average litter size records from three commercial pig lines and wheat grain yield of 599 lines evaluated in four macro-environments. SNPs from the PorcineSNP60 BeadChip and 1447 DArT markers were the inputs for the pig and wheat data analyses, respectively. Gaussian and *t* kernels had similar predictive performance. Multi-kernel RKHS regression models increased the predictive correlation of RKHS by 0.05 when 3 Gaussian or *t* kernels were fitted simultaneously. None of the averaging strategies improved the predictive correlations attained with the multi-kernel kernel fitting. Carre et al. (2014, this volume) combined predictions from this study into a predictive meta-algorithm, finding a marginal improvement in predictive ability, in terms of correlation, mean-squared error or area under the curve.

Tentative conclusion: multi-kernel RKHS fitting constitutes a practical and robust predictive strategy.

Evidence

What follows is not the result of a comprehensive review of literature, but is fairly prototypical of findings obtained so far. While there have been fairly extensive comparisons among members of the Bayesian alphabet, e.g., Lehermeier et al. (2013), which also includes an assessment of sensitivity with respect to priors, similar studies involving RKHS are lacking, especially with animals. Gonzalez-Recio et al. (2008, 2009) found a slightly better predictive ability of RKHS over parametric methods for early mortality and feed efficiency in broilers. However, differences were within the range of the noise stemming from the cross-validation distribution. Heslot et al. (2012) compared many prediction methods, including ridge-regression BLUP, Bayes C-pi and RKHS (neural networks and support vector machines were included as well) using 18 plant breeding data sets. On average, most methods produced the same predictive correlations (see Table 2 of their paper), corroborating the view in Gianola (2013) that well-constructed predictive machines differ by little. However, an "across the board" evaluation is not the best way of comparing methods. For example, using figures from Heslot et al. (2012), if a scatter plot is made for the 18 pairs of predictive correlations, RKHS was better than either ridge regression BLUP (mild differential shrinkage of markers) or Bayes C-pi (a "variable selection" procedure, although not strictly so because every marker receives a non-zero posterior probability of inclusion in the model) in 16 of such comparisons. This suggests that some methods are consistently better, given a specific prediction problem.

Perez-Rodriguez et al. (2012) compared the Bayesian LASSO, Bayesian ridge regression, Bayes A and Bayes B with RKHS, Bayesian regularized neural networks (BRNN), and radial basis function neural networks (RBFNN). Models were compared using 306 elite wheat lines genotyped with 1717 markers and days to heading (DTH) and grain yield (GY) as traits, measured in each of 12 environments. The non-linear models had better overall predictive ability than the linear regressions. Results in Table 2 of their paper speak by themselves.

Genotyping-by-sequencing (GBS) can deliver marker genotypes with less ascertainment bias than SNP arrays. Crossa et al. (2013) evaluated methods for incorporating GBS information, and compared these with pedigree models for predicting genetic values of lines from two maize populations using different traits measured in different environments. Methods were compared using non-imputed, imputed, and GBS-inferred haplotypes of different lengths. GBS and pedigree data were incorporated into statistical models using either GBLUP or RKHS and prediction accuracy was assessed via cross-validation. GBLUP and RKHS models with pedigree with non-imputed and imputed GBS data provided the best predictive correlations for the three traits in their experiment 1, whereas for experiment 2 RKHS provided slightly better predictions than GBLUP for drought-stressed environments, and both models provided similar predictions in well-watered environments. This illustrates again that RKHS can be at least as good as GBLUP.

Gonzalez-Camacho et al. (2012) using high density markers evaluated RBFNN, RKHS, and the additive Bayesian LASSO on 21 maize data sets. Results indicated a slightly but consistently higher predictive correlation of RKHS (0.552) over RBFNN (0.542) and the Bayesian LASSO model (0.542).

In plant breeding there has been much interest in incorporating genotype × environment interaction in marker-assisted prediction models and in structuring environmental information. Jarquin et al. (2013) used a "reaction norm" model in which a matrix of similarities among environments Ω was introduced. This method was applied to 139 wheat lines genotyped with 2,395 markers with 68 environmental conditions modeled. Genotype × environment interaction was fitted by constructing a Hadamard product matrix, essentially a RKHS representation. Predictive ability was much increased by accommodating the environmental and interaction inputs

Conclusion

There does not seem to a uniformly best prediction machine: predictive ability varies with species, environment and trait. Often, variation is due to huge cross-validation noise. As illustrated in the study of human stature by Makowsky et al. (2011), a model with 400,000 markers captured about 80% of the variability in training data, but not more than 20% of that present in testing sets. With such an enormous noise, it is dangerous to ascribe meaning in terms of "genetic architecture" to what might be a transient predictive superiority of a method based on some arbitrary prior, constructed to reflect some idealized vision on gene action. Rather, as noted by Breiman (1996), it is crucial to be cognizant of the sensitivity of predictions to data idiosyncrasy. The latter can be tempered by combining bootstrapping with some robust method, and experience suggests that the latter is multi-kernel RKHS regression. Gianola et al. (2014) found that "bagging" (bootstrap aggregated sampling) may temper over-fitting without damaging predictions, as well as produce candidate specific measures of cross-validation reliability.

Quantitative genetics is primarily a descriptive and predictive science but, arguably, has not been too effective for discovery of genes, especially when compared with the astonishing record of molecular genetics. On the other hand, the explosive availability of genomic and post-genomic data provides means for refining and enhancing prediction of complex traits, an exciting area *per se*, but not too amenable to reductionist reasoning or experimentation. The greatest opportunities for this predictive approach seem to reside in veterinary, human and plant protection applications, as society is increasingly concerned with the potential adverse effects from carbon, methane and water footprints of production agriculture.

Literature Cited

Akdemir, D. (2014). arXiv: 1402.2026v1 [stat.AP]

Bach, F. R., Lanckriet, G. R. G., and Jordan, M. I. (2004). Proc. 21st Int. Conf. Mach. Learn., p. 6

Breiman, L. (1996). Mach. Learn, 24: 123-140.

Breiman, L. (2001). Mach. Learn, 45: 261-277.

Carre, C. et al. (2014). Proc. WCGALP2014.

Claeskens, G., and Hjort N. L. (2008). Cambridge, 320 pp.

Cockerham, C. C. (1954). Genetics, 39:859–882.

Crossa, J. et al. (2013). G3 doi: 10.1534/g3.113.008227.

de los Campos, G., Gianola, D., and Allison, D. B. (2010a). Nat. Rev. Genet., 11: 880–886.

de los Campos, G., Gianola, D., and Rosa G. J. M. (2009) J. Anim. Sci., 87:1883-1887.

de los Campos, G., et al. (2010b). Genet. Res., 92:295–308

Erbe, M. et al. (2010). Proc. WCGALP2010.

Gallais, A. (1974). Biometrics, 30: 429-446

Gianola, D., Fernando, R. L., and Stella, A. (2006). Genetics, 173:1761-1776

Gianola, D., and van Kaam J . T. (2008). Genetics, 178:2289-2303.

Gianola, D. et al. (2011). BMC Genetics 2011, 12:87

Gianola, D. (2013). Genetics, 194:573-596.

Gianola, D., Hospital, F., and Verrier, E. (2013). Theor. Appl. Genet., 126:1457–1472.

Gianola, D. et al. (2014). Plos One (in press).

Goddard, M. E. (2008). Genetica, 136:245-257.

Gönen, M., and Alpaydin, E. (2011). J. Mach. Learn. Res., 12:2211-2268.

Gonzalez-Camacho, J. M. et al. (2012). Theor. Appl. Genet., 125:759-771.

Gonzalez-Recio, O. et al. (2008). Genetics, 178:2305-2313.

Gonzalez-Recio, O. (2009). Gen. Sel. Evol., 41:3.

Henderson, C. R. (1977). J. Dairy Sci., 60:783-787.

Henderson, C. R. (1985). J. Anim. Sci., 60:111-117.

Heslot, N. et al. (2012). Crop Sci., 52:146–160.

Hill, W. G., Goddard, M. E. and Visscher, P. M. (2008). PLoS Genetics, DOI: 10.1371/journal.pgen.1000008

Hoeting, J. A. (1999). Stat Sci., 14:382-417.

Janss, L. et al. (2012). Genetics, 19: 693–704.

Jarquin, D. et al. (2013). Theor. Appl. Genet., DOI 10.1007/s00122-013-2243-1.

Jimenez-Montero, J. A. (2013). J. Dairy Sci., 96:6047-58.

Kempthorne, O. (1954). Proc. Roy. Soc., DOI: 10.1098/rspb.1954.0056.

Kempthorne, O. (1972). J. Royal Stat. Soc. B, 34: 33-37.

Kimeldorf, G., and Wahba, G. (1971). J. Math. Anal. Appl., 33:82-95.

Konstantinov, K. V., and Hayes, B. J. (2010). Proc. WCGALP2010

Lehermeier, C., et al. (2013). SAGMB, 12:375–391.

Liu, D., Lin., X. and Ghosh, D. (2007). Biometrics 63:1079-1088.

Makowsky, R. et al. (2011). PLoS Genet., doi:10.1371/ journal.pgen.1002051.

Meuwissen, T. Hayes, B. J., and Goddard, M. E. (2001). Genetics, 157: 1819–1829

Morota, G. et al. (2013). Gen. Sel. Evol., 45:17

Morota, G. et al. (2014a). BMC Genomics doi:10.1186/1471-2164-15-109.

Morota, G. et al. (2014b). Front. Genet., DOI: 10.3389/fgene.2014.0005

Ober, U. et al. (2010). Genetics, 18:3695-3708

Okut, H. et al. (2011). Genet. Res., 93**:**189-201

Okut , H. et al. (2013). Genet. Sel. Evol. 45: 34.

Ornella, L., et a. (2013). Heredity, doi:10-.1038.

Perez-Rodriguez, P. (2012). G3 DOI: 10.1534.

Rodriguez-Ramilo, S. et al. (2014). PLoS One (in press).

Shawe-Taylor , J., and Cristianini, N. (2004). Cambridge. 474 pp.

Sorensen, D. and Gianola, D. (2002). Springer, 740 pp.

Su, G., et al. (2012). PLoS One. DOI:10.1371

Takezawa, K. (2005). Wiley. 568 pp.

Tusell, L. et al. (2013). Animal 7:1739-1749.

Tusell, L. et al. (2014). J. Anim Breed. Genet., DOI: 10.1111/jbg.12070.

Valente , B. et al. (2014). Proc. WCGALP2014.

Vapnik, V. (1998). Wiley. 768 pp.

Vazquez , A. I. et al. (2012). Genetics, 192: 1493-1502.

Wahba, G. (1990). SIAM, 169 pp.

Weir, B., and Cockerham, C. C. (1976). Proc. Int. Conf. Quant. Genet. Iowa State Press.