Fit Random Forest Model

Fits a random forest model to data in a table.

Random forest (Breiman, 2001) is machine learning algorithm that fits many classification or regression tree (CART) models to random subsets of the input data and uses the combined result (the forest) for prediction. For a detailed description of random forests and practical advice their application in ecology, see Cutler et al. (2007).

This tool fits a classification or regression forest using either the R randomForest package (Liaw and Wiener, 2002) which implements Breiman's classic algorithm, or the cforest function from the R party package (Hothorn et al, 2006; Strobl et al, 2007; Strobl et al, 2008).

A principal feature of random forests is their ability to estimate the importance of each predictor variable in modeling the response variable. Strobl et al. (2007, 2008) found that the randomForest package produces poor estimates in certain scenarios. The party package provides a solution that uses conditional inference trees and importance estimates, making it an attractive alternative to randomForest. The party package does suffer from two drawbacks, however: it does not produce the same diagnostic plots as randomForest, and it requires more processing time and much more memory than randomForest. If the input table has thousands of records, the party package may simply not have enough memory to run.

References

Breiman, L. (2001). Random forests. Machine Learning, 45: 5-32.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. (2007). Random Forests for Classification in Ecology. Ecology 88: 2783-2792.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. (2006). Survival Ensembles. Biostatistics 7: 355-373.

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. R News 2: 18-22.

Strobl, S., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional Variable Importance for Random Forests. BMC Bioinformatics 9:307.

Strobl, S., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. BMC Bioinformatics 8:25.

R party package documentation

R randomForest package documentation

Command line syntax

RandomForestModelFitToArcGISTable_GeoEco <inputTable> <outputModelFile> <formula> {ntree} {mtry} {randomForest | party} {where} {replace} {cfMaxSurrogate} {seed} {importance} {useScaledImportance} {useConditionalImportance} {xColumnName} {yColumnName} {zColumnName} {mColumnName} {writeSummaryFile} {writeImportancePlot} {writePartialDependencePlots} {png | emf} {res} {width} {height} {pointSize} {bg}

Parameters

Expression	Explanation
<inputTable>	ArcGIS table, table view, feature class, or feature layer containing the data to which the model should be fitted.
<outputModelFile>	Output file to receive the fitted model. The file will not be in a user-readable format. After the model is fitted, you can provide the file to other tools that perform further analysis or visualization of the fitted model. It is suggested, but not required, that you give the file an .Rdata extension.
<formula>	Formula that specifies the table fields that are the response variable and predictor variables of the model. To fit a regression forest, the formula must be of the form: response ~ predictor1 + predictor2 + ... + predictorN where response and predictor1 ... predictorN are fields of the table. To fit a classification forest, the formula must be of the form: factor(response) ~ predictor1 + predictor2 + ... + predictorN The use of the R factor() function on the response variable designates it as a categorical variable and causes a classification forest to be built for it. Above, "response" must be a field name. It may not be an R expression. This prohibits certain shortcuts sometimes available in R, such as fitting a binary

	classification using a response expression such as factor(X > 10). To do that, add a new field, set it to the result of X > 10, and then use the new field as the response variable.
	
	The field names are case sensitive. If any field used in the formula is NULL for a given row, that row will not be used in fitting the model.
	
	For example, if you have a field Presence that indicates the categorical presence or absence of a species (1 or 0) and you want to model it in terms of sampled environmental covariates stored in the SST, ChlDensity, and Depth fields, you would use the formula:
	
	`factor(Presence) ~ SST + ChlDensity + Depth`
	
	By default, all predictors are treated as continuous variables. To indicate that a predictor should be treated as a categorical variable, use the factor function. For example, if SubstrateType is an integer code that should be treated as categorical:
	
	`factor(Presence) ~ SST + ChlDensity + Depth + factor(SubstrateType)`
	
	Additional syntax may be possible depending on which R package is used to fit the model. Please see the documentation for the R packages for details.

{ntree}	Number of trees to grow. In the random forests literature, this is referred to as the ntree parameter. Larger number of trees produce more stable models and covariate importance estimates, but require more memory and a longer run time. For small datasets, 50 trees may be sufficient. For larger datasets, 500 or more may be required. Please consult the random forests literature for extensive discussion of this parameter (e.g. Cutler et al., 2007; Strobl et al., 2007; Strobl et al., 2008).
{mtry}	Number of variables available for splitting at each tree node. In the random forests literature, this is referred to as the mtry parameter. The default value of this parameter depends on which R package is used to fit the model: • randomForest - For classification models, the default is the square root of the number of predictor variables (rounded down). For regression models, it is the number of predictor variables divided by 3 (rounded down). • party - The default is always 5. There is extensive discussion in the literature about the influence of mtry. Cutler et al. (2007) reported that different values of mtry did not affect the correct classification rates of their model and that other performance metrics (sensitivity, specificity, kappa, and ROC AUC) were stable under different values of mtry. On the other hand, Strobl et al. (2008) reported that mtry had a strong influence on predictor variable importance estimates. Due to the conflicting evidence reported in the literature, we suggest you start with the default value (i.e. leave this parameter blank) but review the literature carefully and form your own opinion about what value might be suitable for your specific model.
{randomForest | party}	R package to use when fitting the model. • randomForest - The randomForest package (Liaw and Wiener, 2002) which implements Breiman's classic algorithm. This is the default. • party - The party package, from which the cforest function (Hothorn et al, 2006; Strobl et al, 2007; Strobl et al, 2008) will be used to fit the model. This package provides better results in certain situations, particularly in estimating predictor variable importance, at the cost of requiring more processing time and much more memory than randomForest. If the input table has thousands of records or more, the cforest function may simply not have enough memory to run. Also, this package does not produce the diagnostic plots that randomForest produces. For more information on the two packages, please see the randomForest package documentation and the party package documentation.
{where}	SQL WHERE clause expression that specifies the subset of rows to process. If this parameter is not provided, all of the rows will be processed. If this parameter is provided but the underlying database does not support WHERE clauses, an error will be raised.

	The exact syntax of this expression depends on the underlying database. ESRI recommends you reference fields using the following syntax: • If you're querying ArcInfo coverages, shapefiles, INFO tables or dBASE tables (.dbf files), enclose field names in double quotes in the SQL expression: "MY_FIELD". • If you're querying Microsoft Access tables or personal geodatabase tables, enclose field names in square brackets: [MY_FIELD]. • If you're querying ArcSDE geodatabase tables, an ArcIMS feature class, or an ArcIMS image service sublayer, don't enclose field names: MY_FIELD.
{replace}	If True, individual trees of the forest are built using sampling with replacement. If False, the default, the trees are built using sampling without replacement. In the random forests literature, this is referred to as the replace parameter. Please see Strobl et al. (2007) for a discussion of its effects. Although the classic randomForest package used a default value of True for this parameter, we opted to use a default value of False after reviewing the findings of Strobl et al.
{cfMaxSurrogate}	Number of surrogate splits to evaluate. This parameter is only used when the R party package is used to fit the model. The default value is determined by the party package itself. At the time of this writing it was 0, which meant that surrogate splits were not evaluated by default. If surrogate splits are evaluated, the model may use them to estimate values for predictor variables that are missing data. Please see the party package documentation for more information; also, Hapfelmeier (2012) may be useful. References Hapfelmeier, A. (2012). Random Forest variable importance with missing data. Technical Report Number 121, 2012. Department of Statistics, University of Munich, Germany.
{seed}	Random number seed to use. If a value is provided, it will be used to initialize R's random number generator before the model is fitted. If a value is not provided (the default), the random number generator will be initialized from the current time. This parameter is provided so you can precisely control the model fitting process, if desired. Because random forests rely on random selections of data, the default behavior is for a different forest to be built every time you run the tool. To override this, you can specify the seed for the random number generator. This will cause the same exact sequence of random selections to be performed every time you run the tool (assuming do not chang the input data or any other parameters).
{importance}	If True, the default, the importance of each predictor variable will be estimated and reported. If False, variable importance will not be estimated. If the R randomForest package is used to fit the model, two estimates of importance will be reported, permutation accuracy and node impurity, as described below. If the party package is used, only permutation accuracy will be reported. *Permutation Accuracy* Permutation accuracy is the method that is most often recommended for estimating variable importance in random forests. The basic idea is to see how much worse the model performs when each predictor variable is assigned random but realistic values and the rest of the variables are left unchanged. The worse the model performs when a given predictor variable is randomized, the more important that variable is in predicting the response variable. The estimate is computed as follows. First, for each tree in the forest, the prediction error is calculated on the out-of-bag portion of the data. Next, for each variable, the same calcuation is performed using a random permutation of the values of that variable. Finally, for each variable, the differences in prediction errors are averaged over all trees. For classification trees, the result is the mean decrease in prediction accuracy (i.e. the mean descrease in the percentage of observations classified correctly), reported on a 0 to 1 scale (with 1 representing 100%). In the Variable Importance table output by the tool, this is the MeanDecreaseAccuracy column.

	When the R randomForest package is used, this estimate is also reported for each class; these estimates precede the MeanDecreaseAccuracy column. For regression trees, the result is the percentage increase in mean squared errors, reported on a 0 to 100 scale (with 100 representing 100%). In the Variable Importance table output by the tool, this is the %IncMSE column. In both cases, higher values indicate more important variables. *Node Impurity* This is an alternative method for estimating variables in random forests. It is only provided by the randomForest package. Strobl et al. (2007) report that this method is biased in certain ways and recommend against it. Please see that paper for a detailed analysis and description. From the randomForest package documentation: this method reports the total decrease in node impurities from splitting on each variable, averaged over all trees. For classification trees, the result is the mean decrease in the Gini index, and reported in the MeanDecreaseGini column. For regression trees, the result is measured by the residual sum of squares, and reported in the IncNodePurity column.
{useScaledImportance}	This parameter is used only when if the model is fitted with the R randomForest package. If True, the Permutation Accuracy estimates of predictor variable importance are scaled by (divided by) the standard deviation of the differences in prediction errors. If False, the default, the Permutation Accuracy estimates are not scaled. Strobl and Zeileis (2008) reported that a scaling approach can lead to undesirable results. Following their advice, we recommend against scaling. References Strobl, C. and Zeileis, A. (2008). Danger: High Power! - Exploring the Statistical Properties of a Test for Random Forest Variable Importance. Proceedings of the 18th International Conference on Computational Statistics, Porto, Portugal.
{useConditionalImportance}	This parameter is used only when if the model is fitted with the R party package. If True, conditional variable importance is computed. If False, the default, the traditional permutation accuracy is computed, as is done by the randomForest package. A problem with the traditional permutation accuracy method is that it does not account for autocorrelations between predictor variables. When predictor variables are autocorrelated, the traditional method exhibits a bias toward these variables and inflates their estimated importances. A principal feature of the R party package is the ability to estimate conditional variable importance and thereby reduce or eliminate this bias. This method is described in detail by Strobl et al. (2008). We chose not to enable the conditional method by default because we found it requires quite a lot of processing time and memory for large datasets. We suggest you try first without using the conditional method, then again with it enabled.
{xColumnName}	Name to use in the formula for the X coordinates of point features. If the input table is a point feature class or layer, the X coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
{yColumnName}	Name to use in the formula for the Y coordinates of point features. If the input table is a point feature class or layer, the Y coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
{zColumnName}	Name to use in the formula for the Z coordinates of point features. If the input table is a point feature class or layer that has Z coordinates, the Z coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
{mColumnName}	Name to use in the formula for the measure values of point features. If the input table is a point feature class or layer that has measure values, the measure values will be extracted from the points and be accessible in the

	formula using the name provided for this parameter.
{writeSummaryFile}	If True, this tool will write summary information about the fitted model to a text file. (This is the same information that the tool outputs as log messages.) The file will have the name X_summary.txt, where X is the name of the output model file, minus any extension.
{writeImportancePlot}	This parameter is used only when if the model is fitted with the R randomForest package. If True, this tool will write a plot of the importance of predictor variables estimated by permutation accuracy to a file having the name X_importance.Y, where X is the name of the output model file minus the extension and Y is the extension of the selected output plot format. If the model is a classification model, the tool will also write a plot of predictor variable importance for each class to a file having the name X_importance_class_C.Y, where C is the name of the class. These plots are just graphical depictions of the predictor importance values reported in the tool's output. They do not contain any information not already present in the tool's output. At the time this tool was implemented, the R party package did not provide a function for producing these plots. In the future, we may adapt the code from the randomForest package to run on models fitted with the party package.
{writePartialDependencePlots}	This parameter is used only when if the model is fitted with the R randomForest package. This option is disabled by default because creating these plots can take a long time. If True, this tool will write a partial dependence plot for each term in the fitted model's formula. Each plot gives a graphical depiction of the marginal effect of a predictor variable on the class probability (for classification models) or response (for regression models). For classification models, a plot is generated for each class and term combination. Please see Cutler et al. (2007) for a detailed discussion of partial dependence plots. At the time this tool was implemented, the R party package did not provide a function for producing these plots. In the future, we may adapt the code from the randomForest package to run on models fitted with the party package.
{png | emf}	Plot file format, one of: • emf - Windows enhanced metafile (EMF) format. This is a vector format that may be printed and resized without any pixelation and is therefore suitable for use in printable documents that recognize this format (e.g. Microsoft Word or Microsoft Visio). • png - Portable network graphics (PNG) format. This is a compressed, lossless, highly portable raster format suitable for use in web pages or other locations where a raster format is desired. Most scientific journals accept PNG; they typically request that files have a resolution of at least 1000 DPI.
{res}	PNG plot file resolution, in dots per inch (DPI). The default is set to a high value (1000) because this is the minimum resolution typically required by scientific journals that accept figures in PNG format. This parameter is ignored for EMF format because it is a vector format.
{width}	Plot file width in thousandths of inches (for EMF format; e.g. the value 3000 is 3 inches) or pixels (for PNG format).
{height}	Plot file height in thousandths of inches (for EMF format; e.g. the value 3000 is 3 inches) or pixels (for PNG format).
{pointSize}	The default pointsize of plotted text.
{bg}	PNG plot file background color. The color must be a valid name in R's color palette, or "transparent" if there is no background color. This parameter is ignored if the plot format file is EMF.

Scripting syntax

RandomForestModelFitToArcGISTable_GeoEco (inputTable, outputModelFile, formula, ntree, mtry, rPackage, where, replace, cfMaxSurrogate, seed, importance, useScaledImportance, useConditionalImportance, xColumnName, yColumnName, zColumnName, mColumnName, writeSummaryFile, writeImportancePlot, writePartialDependencePlots, plotFileFormat, res, width, height, pointSize, bg)

Parameters

Expression	Explanation
Input table (Required)	ArcGIS table, table view, feature class, or feature layer containing the data to which the model should be fitted.
Output model file (Required)	Output file to receive the fitted model. The file will not be in a user-readable format. After the model is fitted, you can provide the file to other tools that perform further analysis or visualization of the fitted model. It is suggested, but not required, that you give the file an .Rdata extension.

Formula (Required)	Formula that specifies the table fields that are the response variable and predictor variables of the model.
	To fit a regression forest, the formula must be of the form:
	response ~ predictor1 + predictor2 + ... + predictorN
	where response and predictor1 ... predictorN are fields of the table.
	To fit a classification forest, the formula must be of the form:
	factor(response) ~ predictor1 + predictor2 + ... + predictorN
	The use of the R factor() function on the response variable designates it as a categorical variable and causes a classification forest to be built for it.
	Above, "response" must be a field name. It may not be an R expression. This prohibits certain shortcuts sometimes available in R, such as fitting a binary classification using a response expression such as factor(X > 10). To do that, add a new field, set it to the result of X > 10, and then use the new field as the response variable.
	The field names are case sensitive. If any field used in the formula is NULL for a given row, that row will not be used in fitting the model.
	For example, if you have a field Presence that indicates the categorical presence or absence of a species (1 or 0) and you want to model it in terms of sampled environmental covariates stored in the SST, ChlDensity, and Depth fields, you would use the formula:
	`factor(Presence) ~ SST + ChlDensity + Depth`
	By default, all predictors are treated as continuous variables. To indicate that a predictor should be treated as a categorical variable, use the factor function. For example, if SubstrateType is an integer code that should be treated as categorical:
	`factor(Presence) ~ SST + ChlDensity + Depth + factor(SubstrateType)`
	Additional syntax may be possible depending on which R package is used to fit the model. Please see the documentation for the R packages for details.

Number of trees to grow (Optional)	Number of trees to grow. In the random forests literature, this is referred to as the ntree parameter. Larger number of trees produce more stable models and covariate importance estimates, but require more memory and a longer run time. For small datasets, 50 trees may be sufficient. For larger datasets, 500 or more may be required. Please consult the random forests literature for extensive discussion of this parameter (e.g. Cutler et al., 2007; Strobl et al., 2007; Strobl et al., 2008).
Number of variables available for splitting (Optional)	Number of variables available for splitting at each tree node. In the random forests literature, this is referred to as the mtry parameter. The default value of this parameter depends on which R package is used to fit the model: • randomForest - For classification models, the default is the square

	root of the number of predictor variables (rounded down). For regression models, it is the number of predictor variables divided by 3 (rounded down). • party - The default is always 5. There is extensive discussion in the literature about the influence of mtry. Cutler et al. (2007) reported that different values of mtry did not affect the correct classification rates of their model and that other performance metrics (sensitivity, specificity, kappa, and ROC AUC) were stable under different values of mtry. On the other hand, Strobl et al. (2008) reported that mtry had a strong influence on predictor variable importance estimates. Due to the conflicting evidence reported in the literature, we suggest you start with the default value (i.e. leave this parameter blank) but review the literature carefully and form your own opinion about what value might be suitable for your specific model.
R package to use (Optional)	R package to use when fitting the model. • randomForest - The randomForest package (Liaw and Wiener, 2002) which implements Breiman's classic algorithm. This is the default. • party - The party package, from which the cforest function (Hothorn et al, 2006; Strobl et al, 2007; Strobl et al, 2008) will be used to fit the model. This package provides better results in certain situations, particularly in estimating predictor variable importance, at the cost of requiring more processing time and much more memory than randomForest. If the input table has thousands of records or more, the cforest function may simply not have enough memory to run. Also, this package does not produce the diagnostic plots that randomForest produces. For more information on the two packages, please see the randomForest package documentation and the party package documentation.
Where clause (Optional)	SQL WHERE clause expression that specifies the subset of rows to process. If this parameter is not provided, all of the rows will be processed. If this parameter is provided but the underlying database does not support WHERE clauses, an error will be raised. The exact syntax of this expression depends on the underlying database. ESRI recommends you reference fields using the following syntax: • If you're querying ArcInfo coverages, shapefiles, INFO tables or dBASE tables (.dbf files), enclose field names in double quotes in the SQL expression: "MY_FIELD". • If you're querying Microsoft Access tables or personal geodatabase tables, enclose field names in square brackets: [MY_FIELD]. • If you're querying ArcSDE geodatabase tables, an ArcIMS feature class, or an ArcIMS image service sublayer, don't enclose field names: MY_FIELD.
Sample with replacement (Optional)	If True, individual trees of the forest are built using sampling with replacement. If False, the default, the trees are built using sampling without replacement. In the random forests literature, this is referred to as the replace parameter. Please see Strobl et al. (2007) for a discussion of its effects. Although the classic randomForest package used a default value of True for this parameter, we opted to use a default value of False after reviewing the findings of Strobl et al.
Number of surrogate splits to evaluate (party package only) (Optional)	Number of surrogate splits to evaluate. This parameter is only used when the R party package is used to fit the model. The default value is determined by the party package itself. At the time of this writing it was 0, which meant that surrogate splits were not evaluated by default. If surrogate splits are evaluated, the model may use them to estimate values for predictor variables that are missing data. Please see the party package documentation for more information; also, Hapfelmeier (2012) may be useful. References Hapfelmeier, A. (2012). Random Forest variable importance with missing data. Technical Report Number 121, 2012. Department of Statistics, University of Munich, Germany.
Random number seed (Optional)	

	Random number seed to use. If a value is provided, it will be used to initialize R's random number generator before the model is fitted. If a value is not provided (the default), the random number generator will be initialized from the current time. This parameter is provided so you can precisely control the model fitting process, if desired. Because random forests rely on random selections of data, the default behavior is for a different forest to be built every time you run the tool. To override this, you can specify the seed for the random number generator. This will cause the same exact sequence of random selections to be performed every time you run the tool (assuming do not chang the input data or any other parameters).
Estimate predictor variable importance (Optional)	If True, the default, the importance of each predictor variable will be estimated and reported. If False, variable importance will not be estimated. If the R randomForest package is used to fit the model, two estimates of importance will be reported, permutation accuracy and node impurity, as described below. If the party package is used, only permutation accuracy will be reported. *Permutation Accuracy* Permutation accuracy is the method that is most often recommended for estimating variable importance in random forests. The basic idea is to see how much worse the model performs when each predictor variable is assigned random but realistic values and the rest of the variables are left unchanged. The worse the model performs when a given predictor variable is randomized, the more important that variable is in predicting the response variable. The estimate is computed as follows. First, for each tree in the forest, the prediction error is calculated on the out-of-bag portion of the data. Next, for each variable, the same calcuation is performed using a random permutation of the values of that variable. Finally, for each variable, the differences in prediction errors are averaged over all trees. For classification trees, the result is the mean decrease in prediction accuracy (i.e. the mean descrease in the percentage of observations classified correctly), reported on a 0 to 1 scale (with 1 representing 100%). In the Variable Importance table output by the tool, this is the MeanDecreaseAccuracy column. When the R randomForest package is used, this estimate is also reported for each class; these estimates precede the MeanDecreaseAccuracy column. For regression trees, the result is the percentage increase in mean squared errors, reported on a 0 to 100 scale (with 100 representing 100%). In the Variable Importance table output by the tool, this is the %IncMSE column. In both cases, higher values indicate more important variables. *Node Impurity* This is an alternative method for estimating variables in random forests. It is only provided by the randomForest package. Strobl et al. (2007) report that this method is biased in certain ways and recommend against it. Please see that paper for a detailed analysis and description. From the randomForest package documentation: this method reports the total decrease in node impurities from splitting on each variable, averaged over all trees. For classification trees, the result is the mean decrease in the Gini index, and reported in the MeanDecreaseGini column. For regression trees, the result is measured by the residual sum of squares, and reported in the IncNodePurity column.
Use scaled variable importance (randomForest package only) (Optional)	This parameter is used only when if the model is fitted with the R randomForest package. If True, the Permutation Accuracy estimates of predictor variable importance are scaled by (divided by) the standard deviation of the differences in prediction errors. If False, the default, the Permutation Accuracy estimates are not scaled. Strobl and Zeileis (2008) reported that a scaling approach can lead to undesirable results. Following their advice, we recommend against scaling. References Strobl, C. and Zeileis, A. (2008). Danger: High Power! - Exploring the Statistical Properties of a Test for Random Forest Variable Importance. Proceedings of the 18th International Conference on Computational Statistics, Porto, Portugal.
Use conditional variable importance (party package only) (Optional)	This parameter is used only when if the model is fitted with the R party package.

	If True, conditional variable importance is computed. If False, the default, the traditional permutation accuracy is computed, as is done by the randomForest package. A problem with the traditional permutation accuracy method is that it does not account for autocorrelations between predictor variables. When predictor variables are autocorrelated, the traditional method exhibits a bias toward these variables and inflates their estimated importances. A principal feature of the R party package is the ability to estimate conditional variable importance and thereby reduce or eliminate this bias. This method is described in detail by Strobl et al. (2008). We chose not to enable the conditional method by default because we found it requires quite a lot of processing time and memory for large datasets. We suggest you try first without using the conditional method, then again with it enabled.
Name to use for X coordinates of points (Optional)	Name to use in the formula for the X coordinates of point features. If the input table is a point feature class or layer, the X coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
Name to use for Y coordinates of points (Optional)	Name to use in the formula for the Y coordinates of point features. If the input table is a point feature class or layer, the Y coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
Name to use for Z coordinates of points (Optional)	Name to use in the formula for the Z coordinates of point features. If the input table is a point feature class or layer that has Z coordinates, the Z coordinates will be extracted from the points and be accessible in the formula using the name provided for this parameter.
Name to use for M values of points (Optional)	Name to use in the formula for the measure values of point features. If the input table is a point feature class or layer that has measure values, the measure values will be extracted from the points and be accessible in the formula using the name provided for this parameter.
Write model summary file (Optional)	If True, this tool will write summary information about the fitted model to a text file. (This is the same information that the tool outputs as log messages.) The file will have the name X_summary.txt, where X is the name of the output model file, minus any extension.
Write predictor variable importance plot (Optional)	This parameter is used only when if the model is fitted with the R randomForest package. If True, this tool will write a plot of the importance of predictor variables estimated by permutation accuracy to a file having the name X_importance.Y, where X is the name of the output model file minus the extension and Y is the extension of the selected output plot format. If the model is a classification model, the tool will also write a plot of predictor variable importance for each class to a file having the name X_importance_class_C.Y, where C is the name of the class. These plots are just graphical depictions of the predictor importance values reported in the tool's output. They do not contain any information not already present in the tool's output. At the time this tool was implemented, the R party package did not provide a function for producing these plots. In the future, we may adapt the code from the randomForest package to run on models fitted with the party package.
Write partial dependence plots (Optional)	This parameter is used only when if the model is fitted with the R randomForest package. This option is disabled by default because creating these plots can take a long time. If True, this tool will write a partial dependence plot for each term in the fitted model's formula. Each plot gives a graphical depiction of the marginal effect of a predictor variable on the class probability (for classification models) or response (for regression models). For classification models, a plot is generated for each class and term combination. Please see Cutler et al. (2007) for a detailed discussion of partial dependence plots. At the time this tool was implemented, the R party package did not provide a function for producing these plots. In the future, we may adapt the code from

	the randomForest package to run on models fitted with the party package.
Plot file format (Optional)	Plot file format, one of: • emf - Windows enhanced metafile (EMF) format. This is a vector format that may be printed and resized without any pixelation and is therefore suitable for use in printable documents that recognize this format (e.g. Microsoft Word or Microsoft Visio). • png - Portable network graphics (PNG) format. This is a compressed, lossless, highly portable raster format suitable for use in web pages or other locations where a raster format is desired. Most scientific journals accept PNG; they typically request that files have a resolution of at least 1000 DPI.
Plot resolution, in DPI (Optional)	PNG plot file resolution, in dots per inch (DPI). The default is set to a high value (1000) because this is the minimum resolution typically required by scientific journals that accept figures in PNG format. This parameter is ignored for EMF format because it is a vector format.
Plot width (Optional)	Plot file width in thousandths of inches (for EMF format; e.g. the value 3000 is 3 inches) or pixels (for PNG format).
Plot height (Optional)	Plot file height in thousandths of inches (for EMF format; e.g. the value 3000 is 3 inches) or pixels (for PNG format).
Default pointsize of plotted text (Optional)	The default pointsize of plotted text.
Plot background color (Optional)	PNG plot file background color. The color must be a valid name in R's color palette, or "transparent" if there is no background color. This parameter is ignored if the plot format file is EMF.