Supervised Machine Learning for Analysing Spectra of Exoplanetary Atmospheres

Pablo Márquez-Neila\textsuperscript{1,2}, Chloe Fisher\textsuperscript{2}, Raphael Sznitman\textsuperscript{3}, Kevin Heng\textsuperscript{2,3}

\textsuperscript{1} University of Bern, ARTORG Center for Biomedical Engineering, Murtenstrasse 50, CH-3008, Bern, Switzerland
\textsuperscript{2} University of Bern, Center for Space and Habitability, Gesellschaftstrasse 6, CH-3012, Bern, Switzerland
\textsuperscript{3} Corresponding author: kevin.heng@ph.unibe.ch

The use of machine learning is becoming ubiquitous in astronomy \cite{11,12,3}, but remains rare in the study of the atmospheres of exoplanets. Given the spectrum of an exoplanetary atmosphere, a multi-parameter space is swept through in real time to find the best-fit model \cite{4,5,6}. Known as “atmospheric retrieval”, it is a technique that originates from the Earth and planetary sciences \cite{7}. Such methods are very time-consuming and by necessity there is a compromise between physical and chemical realism versus computational feasibility. Machine learning has previously been used to determine which molecules to include in the model, but the retrieval itself was still performed using standard methods \cite{8}. Here, we report an adaptation of the “random forest” method of supervised machine learning \cite{9,10}, trained on a pre-computed grid of atmospheric models, which retrieves full posterior distributions of the abundances of molecules and the cloud opacity. The use of a pre-computed grid allows a large part of the computational burden to be shifted offline. We demonstrate our technique on a transmission spectrum of the hot gas-giant exoplanet WASP-12b using a five-parameter model (temperature, a constant cloud opacity \(X\)) and the volume mixing ratios \(NH_3, H_2O, CH_4\) and hydrogen cyanide \(HCN\), and a constant cloud opacity \(\eta_0\). Given that these 5 parameters represent continuous data, we make use of “regression trees” rather than decision trees (which are used for discrete data) in our random forest \cite{11}. The splitting is performed so as to maximize the gain in information entropy \cite{14}. Since decision trees are sensitive to slight changes in the training set, they are suitable for use with the bootstrapping method, which constructs the decision tree by randomly drawing from the training set \cite{13}.

The training set consists of 80,000 synthetic WFC3 transmission spectra, each described by 5 parameters: the temperature \(T\), volume mixing ratios (relative abundances by number) of water \((X_{H_2O})\), ammonia \((X_{NH_3})\) and hydrogen cyanide \((X_{HCN})\), and a constant cloud opacity \(\eta_0\). Within the training set, each spectrum was a way of splitting a training set into subsets based on common characteristics of its members \cite{13}. The splitting is performed so as to maximize the gain in information entropy \cite{14}. Since decision trees are sensitive to slight changes in the training set, they are suitable for use with the bootstrapping method, which constructs the decision tree by randomly drawing from the training set \cite{14}.

The training set consists of 80,000 synthetic WFC3 transmission spectra, each described by 5 parameters: the temperature \(T\), volume mixing ratios (relative abundances by number) of water \((X_{H_2O})\), ammonia \((X_{NH_3})\) and hydrogen cyanide \((X_{HCN})\), and a constant cloud opacity \(\eta_0\). Given that these 5 parameters represent continuous data, we make use of “regression trees” rather than decision trees (which are used for discrete data) in our random forest \cite{11}. For each spectrum, the values of the 5 parameters are randomly generated either from a log-uniform (volume mixing ratios and cloud opacities) or uniform (temperature) distribution. In addition to adopting the same wavelength range and 13 bins of the measured WASP-12b WFC spectrum \cite{12}, we assume a noise floor of 50 parts per million (ppm) on the transit depth.

In a general machine-learning situation, each member of a training set is associated with a number of characteristics known as “features” (in the jargon of machine learning), e.g., color, height, type of terrain. For a spectrum, the features are the number of data points it contains. Here, the WFC3 spectrum has 13 features or binned data points. Within the training set, each spectrum is identified by its values of the 5 parameters. The training set of 80,000 synthetic spectra resides in a 13-dimensional space, where each dimension corresponds to a wavelength bin. Along the axis of each dimension is a continuous range of values of the transit radii. The goal is to relate an entry (a synthetic spectrum) in this 13-
dimensional space with the range of values of each of the 5 parameters. We accomplish this by sub-dividing the 13-dimensional space into patches or islands, which is handled using a regression tree. Each patch encompasses some subset of the training set, from which the variance in their transit radii may be computed. The sub-division of the 13-dimensional space is done so as to minimize the sum of the variances of all of these patches. This is conceptually equivalent to maximizing the gain in information entropy for discrete data [14].

Upon setting up the regression tree, we use it in tandem with a bootstrapping method. To train each regression tree, we randomly draw from the 80,000 synthetic spectra in the training set. Upon each draw, the drawn synthetic spectrum is placed back into the training set, allowing for it to be drawn more than once. Each regression tree may be visualized as being a predictive “voter”, who returns the ranges of parameter values given the 13 data points of the measured WFC3 transmission spectrum. While a single regression tree produces predictions with large uncertainties, random forests mitigate this pitfall by combining the responses of multiple trees. We performed tests that indicate a convergence of these predictions using 1000 regression trees (see Methods). Using 1000 regression trees to form a random forest, we are able to compute the posterior distributions of the parameters [15].

Figure 1 shows the posterior distributions of the temperature, cloud opacity and volume mixing ratios of water, ammonia and hydrogen cyanide. The retrieved water volume mixing ratio (log $X_O^2$) and temperature ($T = 952^{+151}_{-110} K$) values are broadly consistent with the previous analysis [12]. A non-zero cloud opacity (log $k_0 = -2.3^{+1.3}_{-1} \pm 0.9$) is necessary to flatten the spectral continuum blueward of the 1.4 μm water feature. The degeneracies between the temperature, molecular abundances and cloud opacity are consistent with physical intuition. As the temperature increases linearly, the molecular opacities increase exponentially, a property that may be compensated by an order-of-magnitude decrease in the volume mixing ratio. An increasing temperature also reduces the differences in opacity between the temperature, molecular abundances and cloud opacity are consistent with physical intuition.
the flatness of the spectral continuum. The two reddest data points are most constraining for hydrogen cyanide.

There are straightforward extensions of random-forest retrieval for which no conceptual obstacles exist. We have demonstrated the method on a spectrum with 13 data points, but the random forest method implies that random-forest retrieval is applicable to future James Webb Space Telescope (JWST) spectra spanning a broader range of wavelengths with ~100–1000 data points [29]. The information content analysis may be used to influence observational campaigns and the design of spectrographs, depending on the intended scientific goal.

Another straightforward extension is to train a random forest once and apply it to an ensemble of spectra. In the current study, we picked a specific object (WASP-12b) to demonstrate our method. There is no conceptual obstacle to making model grids where the surface gravity is allowed to vary. The random forest is trained on this larger grid, but the value of the surface gravity may be fixed to the measured value of a specific object during analysis with no need for retraining. In the study of stars and brown dwarfs, model grids spanning different ages, luminosities, radii, gravities and cloud configurations have traditionally been used to analyse ensembles of objects [27–29]. It is conceivable that one may use model grids produced by different research groups to perform retrievals, even if the computer codes used to generate these grids are proprietary.

For the current study, we have showed that more sophisticated models are not necessary to analyse the WFC3 spectrum of WASP-12b. However, there is nothing that prevents one from considering more sophisticated models. For example, using the non-isothermal model of [11] in tandem with the non-grey cloud model of [30] would add 4 more parameters to the retrieval. A longstanding shortcoming of atmospheric retrieval is its isothermal, isobaric and hosts a grey cloud. Using the nested sampling method [17–18] [33, 34, 32], we have performed regular retrievals, which indicate that non-isothermal behavior and non-grey clouds are not necessary to explain the data given its current level of quality and sophistication. We include the opacities of water (H2O), hydrogen cyanide (HCN) and ammonia (NH3), computed using the ExoMol spectroscopic line lists [35–37] as input and in the standard way, meaning that the opacities are products of the integrated line strength and line shape, and the line shapes are assumed to be truncated Voigt profiles [35–37]. For each model, we randomly pick values of the parameters over the following ranges: \( T = 500–2900 \text{ K}, X_{\text{H2O}} = 10^{-13}–1, X_{\text{HCN}} = 10^{-13}–1, X_{\text{NH3}} = 10^{-13}–1, \kappa_0 = 10^{-10}–10^2 \text{ cm}^2 \text{ g}^{-1}. \) The surface gravity of WASP-12b is taken to be 977 cm s\(^{-2}\) [10]. The spectroscopic database used to construct the NH3 opacities does not exist for temperatures above 1600 K [35]. For computational reasons, we set the NH3 opacity to be zero and the volume mixing ratio to be small (\(10^{-16}\)) if the temperature exceeds this threshold. Fortunately, ammonia is expected to be a minor species at high temperatures, where the dominant nitrogen carrier is instead expected to be molecular nitrogen [31–32]. The “features” are the 13 values of the transit radius, across wavelength, associated with each transmission spectrum. One may visualize 13 columns, each with 80,000 values of the transit radius. One then visualizes a 13-dimensional space, where each dimension is marked by a set of numerical thresholds. Boundaries in this 13-dimensional space are drawn based on splitting the training set in order to minimize the total variance. Each time a boundary is drawn, one is splitting the regression tree. Once the reduction in the variance of the tree node is negligibly small (0.01 in our case), we stop splitting the training set. Tree pruning methods are not used. Each time the tree is split, only a random subset (4, which is about \(\sqrt{13}\)) of the 13 spectral bins is used. In other words, both the members of the 80,000 training set, as well as the subset of spectral bins associated with each member, are randomly drawn in order to decrease the correlations between predictions from dif-

**Methods**

For the physics input, we choose to use a previously validated analytical formula to convert the temperatures, molecular opacities and relative abundances of molecules into transit radii [11]. The simplicity of this forward model allows us to straightforwardly diagnose problems and understand trends in the posterior distributions. We use the simplest incarnation of this formula, which assumes that the atmosphere is isothermal, isobaric and hosts a grey cloud. Using the flatness of this formula, which assumes that the atmosphere is isothermal, isobaric and hosts a grey cloud. Using the

Data Availability: The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Code Availability: The code used to generate all random-forest retrievals can be accessed at [https://github.com/exoclime](https://github.com/exoclime).

Correspondence and request for materials should be made to K.H. We acknowledge partial financial support from the Center for Space and Habitability (P.M.N. and K.H.), the University of Bern International 2021 Ph.D Fellowship (C.F.), the PlanetS National Center of Competence in Research (K.H.), the Swiss National Science Foundation (R.S., C.F. and K.H.), the European Research Council via a Consolidator Grant (K.H.), the Swiss-based MERAC Foundation (K.H.), the PlanetS National Science Foundation (R.S., C.F. and K.H.), the European Research Council via a Consolidator Grant (K.H.), the Swiss-based MERAC Foundation (K.H.), and the Swiss National Science Foundation (K.H.).

P.M.N. led the development of computer codes used for this study, performed the machine-learning related calculations, participated in the experimental design and made the majority of the figures. C.F. computed the grid of atmospheric models used as the training set, participated in the experimental design and performed the nested-sampling retrievals. R.S. co-led the scientific vision and experimental design, and co-wrote the manuscript. K.H. co-led the scientific vision and experimental design and led the writing and typesetting of the manuscript.
ferent trees. For a pedagogical summary of the ran-

dom forest method, please see [14]. The implemen-
tations of the random forest method and $R^2$ metric
are from the open-source scikit.learn library in the
Python programming language.

It has been previously shown that the random for-
est method is capable of handling systems with 1000–
10,000 features and tree depths of several tens to hun-
dreds [21, 22, 23, 24, 25]. Our current problem has
13 features and the regression trees have, on average,
about 19,000 nodes and depths of 14.

To check the robustness of our results with respect to
our implementation of the random forest method, we
examine retrieval outcomes with different num-
bers of regression trees. Like before, we train on
80,000 synthetic spectra and then use it to analyse
20,000 more synthetic spectra. Figure 5 shows that
the outcomes of these mock retrievals converge when
the number of trees used exceeds about 100. In
the same figure, we also checked the retrieval outcomes
with different levels of assumed noise floors. For each
of the 13 data points in the synthetic WFC3 spec-
tra, we assume a Gaussian uncertainty on the transit
depths with full widths at half-maximum of 10, 50
and 100 ppm, which represent ideal, typical and eas-
dy attainable conditions. As expected, the variance
associated with the true versus predicted values of the
five parameters decreases (i.e., the coefficient of de-
termination increases) when the assumed noise floor
is lower. As a further check, we first train a random
forest on a model grid with an assumed noise floor of
50 ppm and use it to analyse mock data with assumed
noise floors of 10, 50 and 100 ppm. The resulting $R^2$
values are 0.676, 0.651 and 0.586, respectively, for the
joint predictions.

We also ran the same mock retrievals for model
grids where the atmosphere contains water only ver-
sus one that contains hydrogen cyanide and ammo-
nia (without water), as shown in Figure 6. In
the former case, the retrievals return $X_{\text{HCN}} \sim 10^{-8}$ and $X_{\text{NH}_2} \sim 10^{-10}$ even when neither molecule is present
in the mock spectra, which is consistent with our
finding in Figure 4 that volume mixing ratios below
$\sim 10^{-9}$ indicate non-detections of these molecules.
In the latter case, we obtain $X_{\text{H}_2}\text{O} \sim 10^{-8}$, which is consistent with the non-detection of water.

As a final test and precursor for future studies,
we generated mock JWST-like data in the NIRSpec
range of wavelengths (0.8 to 5.0 $\mu$m) at a resolution
of 100 (not shown). Despite the increase in the num-
ber of features (data points) from 13 to 181, the time
needed to train the random forest on 80,000 mock
spectra only increased by a factor of 4 (without any
try to parallelise the computation). The time
needed for interpreting the additional 20,000 mock
spectra (termed “testing”) is virtually the same in
both cases. Furthermore, we note that both the train-
ing and testing steps are highly parallelisable.

To determine the spectral resolution used for our
opacities, we ran retrievals with resolutions of 1, 2,
5 and 10 cm$^{-1}$ assuming an isothermal atmosphere
containing grey clouds and all three molecules. Re-

trieval practitioners typically use a spectral resolu-
tion of 1 cm$^{-1}$ for their opacities [43, 14, 15],
although it is not uncommon for workers to not state
the spectral resolution used. For these 4 resolutions,
the retrievals are shown in Figure 7. The correspond-
ing retrieved parameter values are tabulated. Based
on this resolution test, we adopt 5 cm$^{-1}$ as our spec-
tral resolution for the opacities.

We assume pressure broadening to be negligible.
Since the inferred atmospheric temperature does not
fall well below 1000 K and the volume mixing ratios
are typically much smaller than unity, this is not an
unreasonable assumption [33]. Operationally, to im-
plement this assumption we assume a pressure of 1
mbar when computing the opacities. As is accepted
practice [40], our ignorance of the physics of pressure
broadening forces us to truncate the Voigt profile at
some distance from line center. We have made an ad
hoc choice of 100 cm$^{-1}$, but since pressure broaden-
ing is assumed to be negligible this has little to no
effect on the outcome.

To check our assumption of a constant/grey cloud
opacity, we ran another retrieval calculation with the
non-grey cloud model of [32]. The Bayes factor for
the pair of models with grey versus non-grey clouds
is 0.6 (with the former having a higher Bayesian evi-
dence), which implies there is no evidence for the data
favouring the non-grey over the grey cloud model [33].
In fact, we note that the model with non-grey clouds
and water only has the same Bayesian evidence as one
with grey clouds and all three molecules. Similarly,
the Bayes factor for a pair of models with isothermal
versus non-isothermal atmospheres (both with wa-
ter only) is 0.7, implying a lack of evidence for non-
isothermal behavior. The latter has lower Bayesian
evidence and was computed using the non-isothermal
analytical formula derived by [11].

References
[1] Banerji, M. et al. Galaxy Zoo: reproducing
galaxy morphologies via machine learning. Mon.
Not. R. Astron. Soc., 406, 342–353 (2010).
[2] Graff, P., Feroz, F., Hobson, M.P., & Lasenby,
A. SKYNET: an efficient and robust neural net-
work training tool for machine learning in astronomy. *Mon. Not. R. Astron. Soc.*, 441, 1741–1759 (2014).

[3] Pearson, K.A., Palafox, L., Griffith, C.A. Searching for exoplanets using artificial intelligence. *Mon. Not. R. Astron. Soc.*, 474, 478–491.

[4] Madhusudhan, N., & Seager, S. A Temperature and Abundance Retrieval Method for Exoplanet Atmospheres. *Astrophys. J.*, 707, 24–39 (2009).

[5] Benneke, B., & Seager, S. Atmospheric Retrieval for Super-Earths: Uniquely Constraining the Atmospheric Composition with Transmission Spectroscopy. *Astrophys. J.*, 753, 100 (2012).

[6] Line, M.R. et al. Information Content of Exoplanetary Transit Spectra: An Initial Look. *Astrophys. J.*, 749, 93 (2012).

[7] Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice* (World Scientific, 2000).

[8] Waldmann, I.P. *Dreaming of Atmospheres*. *Astrophys. J.*, 820, 107 (2016).

[9] Ho, T.K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 832–844 (1998).

[10] Breiman, L. Random Forests. *Machine Learning*, 45, 5–32 (2001).

[11] Heng, K., & Kitzmann, D. The theory of transmission spectra revisited: a semi-analytical method for interpreting WFC3 data and an unresolved challenge. *Mon. Not. R. Astron. Soc.*, 470, 2972–2981.

[12] Kreidberg, L. et al. A Detection of Water in the Transmission Spectrum of the Hot Jupiter WASP-12b and Implications for Its Atmospheric Composition. *Astrophys. J.*, 814, 66 (2015).

[13] Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. *Classification and Regression Trees* (Wadsworth International Group, 1984).

[14] Kelleher, J., Mac Namee, B., D’Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics* (MIT Press, 2015).

[15] Criminisi, A., Shotton, J., & Konukoglu, E. Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Microsoft Research technical report*, TR-2011-114 (2011).

[16] Trotta, R. Bays in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49, 71–104.

[17] Skilling, J. Nested Sampling for General Bayesian Computation. *Bayesian Analysis*, 1, 833–860.

[18] Feroz, F., & Hobson, M.P. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Mon. Not. R. Astron. Soc.*, 384, 449–463.

[19] Batalha, N.E., & Line, M.R. Information Content Analysis for Selection of Optimal JWST Observing Modes for Transiting Exoplanet Atmospheres. *Astron. J.*, 153, 151 (2017).

[20] Howe, A.R., Burrows, A., & Deming, D. An Information-theoretic Approach to Optimize JWST Observations and Retrievals of Transiting Exoplanet Atmospheres. *Astrophys. J.*, 835, 96 (2017).

[21] Hastie, T., Tibshirani, R., & Friedman, J. *The Elements of Statistical Learning* (Springer, 2001).

[22] Smutzman, R., Becker, C., Fleuret, F., & Fua, P. Fast Object Detection with Entropy-Driven Evaluation. *The IEEE Conference on Computer Vision and Pattern Recognition*, 3270–3277 (2013).

[23] Zikic, D., Glocker, B., & Criminisi, A. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Medical Image Analysis*, 18, 1262–1273 (2014).

[24] Ricke, N. et al. Surgical Tool Tracking and Pose Estimation in Retinal Microsurgery. *Medical Image Computing and Computer Assisted Intervention*, 266–273 (2015).

[25] Zhang, L., Varadarajan, J., Suganthan, P.N., Ahuja, N., & Moulin, P. Robust Visual Tracking Using Oblique Random Forests. *The IEEE Conference on Computer Vision and Pattern Recognition*, 5825–5834 (2017).

[26] Greene, T.P. et al. Characterizing Transiting Exoplanet Atmospheres with JWST. *Astrophys. J.*, 817, 17 (2016).

[27] Marley, M.S. et al. Atmospheric, Evolutionary, and Spectral Models of the Brown Dwarf Gliese 229 B. *Science*, 272, 1919–1921 (1996).

[28] Burrows, A. et al. A Non-gray Theory of Extrasolar Giant Planets and Brown Dwarfs. *Astrophys. J.*, 491, 856–875 (1997).

[29] Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P.H. Evolutionary models for low-mass stars and brown dwarfs: Uncertainties and limits at very young ages. *Astron. Astrophys.*, 382, 563–572 (2002).

[30] Kitzmann, D., & Heng, K. Optical properties of potential condensates in exoplanetary atmospheres. *Mon. Not. R. Astron. Soc.*, 475, 94–107 (2018).
[31] Feroz, F., Hobson, M.P., & Bridges, M. Multi-Nest: an efficient and robust Bayesian inference tool for cosmology and particle physics. Mon. Not. R. Astron. Soc., 398, 1601–1614 (2009).

[32] Buchner, J. et al. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. Astron. Astrophys., 564, A125 (2014).

[33] Barber, R.J., Tennyson, J., Harris, G.J., & Tolchenov, R.N. A high-accuracy computed water line list. Mon. Not. R. Astron. Soc., 368, 1087–1094 (2006).

[34] Barber, R.J. et al. ExoMol line lists - III. An improved hot rotation-vibration line list for HCN and HNC. Mon. Not. R. Astron. Soc., 437, 1828–1835 (2014).

[35] Yurchenko, S.N., Barber, R.J., & Tennyson, J. A variationally computed line list for hot NH3. Mon. Not. R. Astron. Soc., 413, 1828–1834 (2011).

[36] Yurchenko, S.N., Tennyson, J., Barber, R.J., & Thiel, W. Vibrational transition moments of CH4 from first principles. Journal of Molecular Spectroscopy, 291, 69–76 (2013).

[37] Yurchenko, S.N., & Tennyson, J. ExoMol line lists - IV. The rotation-vibration spectrum of methane up to 1500 K. Mon. Not. R. Astron. Soc., 440, 1649–1661 (2014).

[38] Rothman, L.S. et al. The HITRAN Molecular Spectroscopic Database and HAWKS (HITRAN Atmospheric Workstation): 1996 Edition. Journal of Quantitative Spectroscopy & Radiative Transfer, 60, 665–710 (1998).

[39] Grimm, S.L., & Heng, K. HELIOS-K: An Ultra-fast, Open-source Opacity Calculator for Radiative Transfer. Astrophys. J., 808, 182 (2015).

[40] Hebb, L. et al. WASP-12b: The Hottest Transiting Extrasolar Planet Yet Discovered. Astrophys. J., 693, 1920–1928 (2009).

[41] Burrows, A., & Sharp, C.M. Chemical Equilibrium Abundances in Brown Dwarf and Extrasolar Giant Planet Atmospheres. Astrophys. J., 512, 843–863 (1999).

[42] Heng, K., & Tsai, S.-M. Analytical Models of Exoplanetary Atmospheres. III. Gaseous C-H-O-N Chemistry with Nine Molecules. Astrophys. J., 829, 104 (2016).

[43] Line, M.R. et al. A Systematic Retrieval Analysis of Secondary Eclipse Spectra. I. A Comparison of Atmospheric Retrieval Techniques. Astrophys. J., 775, 137 (2013).

[44] Waldmann, L.P. et al. Tau-REx I: A Next Generation Retrieval Code for Exoplanetary Atmospheres. Astrophys. J., 802, 107 (2015).

[45] Lavie, B. et al. HELIOS-RETRIEVAL: An Open-source, Nested Sampling Atmospheric Retrieval Code; Application to the HR 8799 Exoplanets and Inferred Constraints for Planet Formation. Astron. J., 154, 91 (2017).

[46] Sharp, C.M., & Burrows, A. Atomic and Molecular Opacities for Brown Dwarf and Giant Planet Atmospheres. Astrophys. J., 168, 140–166 (2007).
Figure 1: Posterior distributions of the relative molecular abundances (volume mixing ratios), temperature and cloud opacity obtained from the machine-learning retrieval analysis of the WFC3 transmission spectrum of WASP-12b. Shown are the logarithm (base 10) of the volume mixing ratios and cloud opacity. Within each scatter plot, each dot is an individual prediction of a single regression tree in the random forest. The straight lines indicate the median values of the parameters. Note that the volume mixing ratios and cloud opacity are associated with a factor ($P_b/10$ bar) due to the normalization degeneracy.
Figure 2: Same as Figure 1 but for nested-sampling retrieval. Additionally, the insert shows the measured versus best-fit model transmission spectra.
Figure 3: True versus random-forest predicted values of the five parameters in our transmission spectrum model. The coefficient of determination ($R^2$) varies from 0 to 1, where values near unity indicate strong correlations between the predicted and real values of a given parameter, based on the variance of outcomes.

Figure 4: Feature importance plots associated with the machine-learning retrieval analysis of the WFC3 transmission spectrum of WASP-12b. Values along the vertical axis indicate the relative importance of a data point for retrieving the value of a given parameter. Within each panel, the vertical axis values sum up to unity.
Figure 5: Left panel: Coefficient of determination for each of the 5 parameters, as well as for the joint prediction, versus the number of regression trees used in the random forest with an assumed noise floor of 50 ppm. Also included are the joint predictions for assumed noise floors of 10 and 100 ppm. Right panel: Same as Figure 3, but comparing mock retrievals with assumed noise floors of 10 versus 100 ppm. The coefficient of determination ($R^2$) varies from 0 to 1, where values near unity indicate strong correlations between the predicted and real values of a given parameter.

Figure 6: True versus random-forest predicted values of the five parameters in our transmission spectrum model. Left montage: H$_2$O only, where the posterior distributions of HCN and NH$_3$ are shown. Right montage: HCN and NH$_3$ only, where the posterior distribution of H$_2$O is shown.
Figure 7: Nested-sampling retrievals using opacities with spectral resolutions of $1 \text{ cm}^{-1}$ (dotted curves), $2 \text{ cm}^{-1}$ (dot-dashed curves), $5 \text{ cm}^{-1}$ (filled posteriors in colour and solid lines for the median values of parameters) and $10 \text{ cm}^{-1}$ (dashed curves), respectively.