

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ЗАДАЧАХ ОПТИМИЗАЦИИ АГРОТЕХНОЛОГИЧЕСКИХ РЕШЕНИЙ И ОЦЕНКИ ПРОСТРАНСТВЕННОЙ НЕОДНОРОДНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ УГОДИЙ ПО УРОЖАЙНОСТИ

Касабуцкий А.Ф., к.ф.м-н., доц., Серебрякова Н.Г., к.п.н., доц. (БГАТУ, Минск)

Введение

В развитии производства продукции растениеводства перспективным направлением является переход к дифференцированным технологиям точного земледелия [1]. В его основе лежит управление продуктивностью посевов, учитывающее пространственно-временную вариабельность среды обитания растений. Центральное место в обосновании применения системы точного земледелия занимают вопросы его информационного обеспечения. Особенно велика роль информационного обеспечения земледелия на уровне конкретного хозяйства для решения плановых технологических и оперативных задач по управлению производственным процессом сельскохозяйственных культур. Объем и качество информационного обеспечения на этом уровне напрямую зависят от наличия и возможностей физико-технических и программных средств.

Концептуальная схема по совершенствованию информационного обеспечения точного земледелия представлена на рисунке 1 [2].

Вероятностно-статистическое моделирование предполагает создание математических моделей и разработку математических методов нахождения оптимальных или близких к оптимальным решений в условиях наличия стохастической неопределенности.

Основная часть

1 Математическая постановка и решение оптимизационной задачи выбора сроков проведения агротехнологических операций

Оптимизация сроков проведения агротехнологических операций является важнейшим этапом в реализации точного земледелия [3]. Уже на стадии планирования агротехнологий возникает задача оценки момента времени проведения той или иной операции, при этом, как правило, можно оценить ожидаемые потери за единицу времени, связанные с завышением или, наоборот, с занижением оценки времени проведения необходимого мероприятия. В некоторых случаях эти потери могут быть выражены непосредственно в денежных единицах (например, простой техники; потери, связанные с необходимостью привлечения дополнительной техники, рабочей силы и т.д.). В других случаях величину потерь можно получить в результате экспертной оценки относительной нежелательности ошибки, связанной с завышением оценки момента времени проведения мероприятия по сравнению с ее занижением.

В задачах такого типа часто заранее можно указать временные границы проведения мероприятий. Проблема сводится к оценке момента времени проведения необходимых работ внутри некоторого временного интервала.

Предположим, что в качестве оценки момента времени проведения агротехнологического мероприятия выбран момент времени $x \in [a, b]$ внутри фиксированного заданного временного промежутка $[a, b]$, причем наилучший момент времени проведения этого мероприятия τ является случайной величиной с известной функцией распределения $F(t)$ и, вообще говоря, может не принадлежать промежутку $[a, b]$. Предположение о случайности момента времени τ моделирует неопределенность, связанную с наличием разнообразных факторов, оказывающих значимое влияние и трудно прогнозируемых на практике.

Предположим, что c – величина потерь за единицу времени, связанных с занижением оценки, а l – величина потерь за единицу времени, связанных с завышением оценки.

Показано, что

Ожидаемые средние потери $Q(x)$ составят следующую величину

$$Q(x) = \tilde{n} \int_x^{\infty} (\min(t, b) - x) dF(t) + l \int_{-\infty}^x (x - \max(t, a)) dF(t),$$

Поставим задачу минимизации ожидаемых потерь:

$$x^* = \arg \min_{x \in [a, b]} Q(x)$$

Пусть x_p – решение уравнения: $F(x_p) = \frac{c}{c+l}$.

Теорема 1. Если строго возрастающая функция распределения $F(t)$ непрерывна, то решение задачи (2.2) определяется выражением

$$x^* = \begin{cases} b, & \text{если } x_p > b \\ x_p, & \text{если } a < x_p \leq b \\ a, & \text{если } x_p \leq a \end{cases}$$

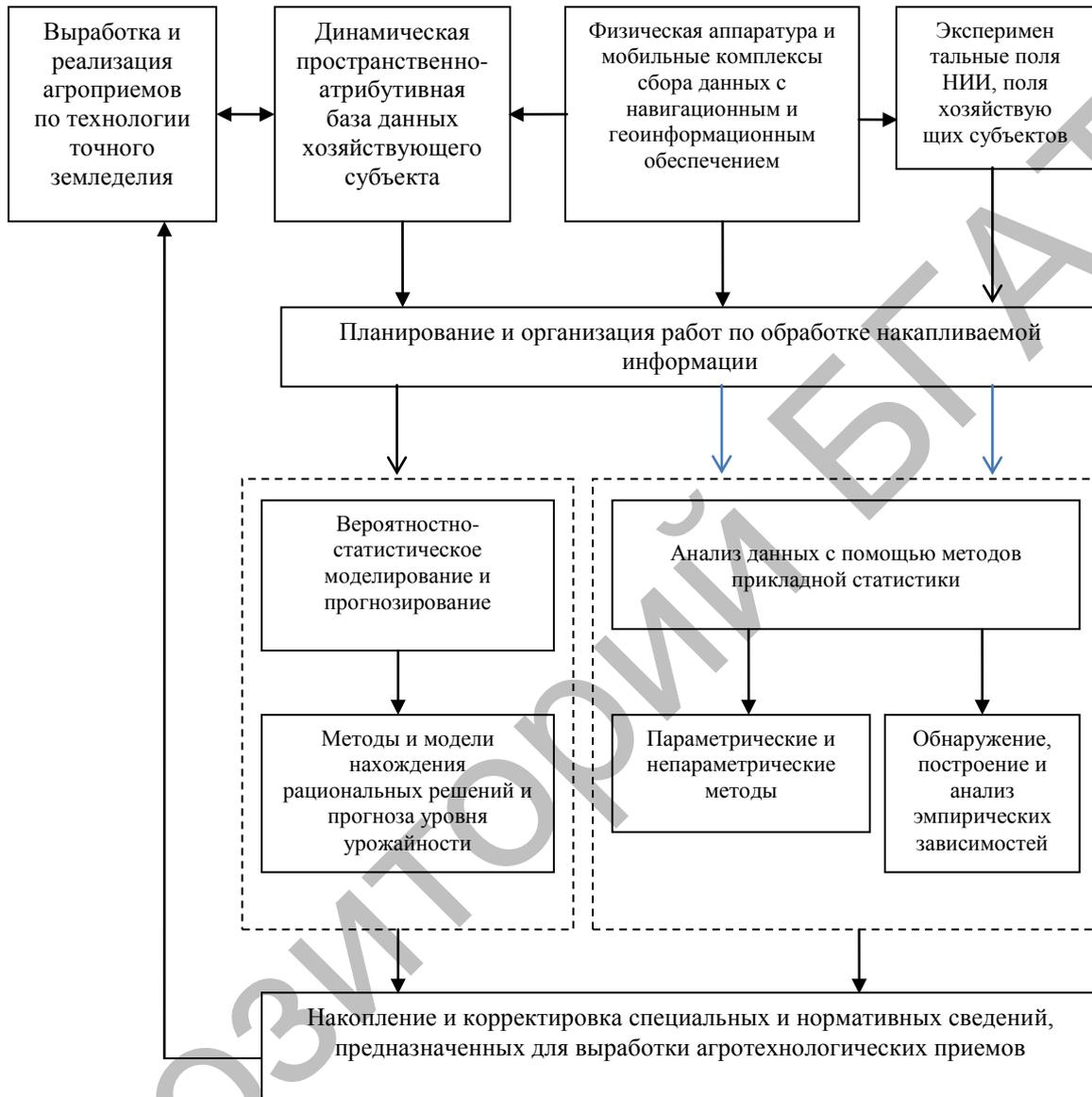


Рисунок 1 - Концептуальная схема совершенствования информационного обеспечения точного земледелия

Теорема 1 дает решение оптимизационной задачи построения оптимальной оценки момента времени по выбранному критерию оптимальности. Если функция распределения $F(t)$ неизвестна, то следует построить статистическую оценку функции распределения $F(t)$ по имеющимся опытным данным, либо использовать минимаксный подход.

Суть минимаксного подхода заключается в решении следующей оптимизационной задачи:

$$x^{**} = \arg \inf_{x \in \{a,b\}} \sup_{F \in f} Q(x, F), \quad (1)$$

где f - множество всех функций распределения, $Q(x, F) \equiv Q(x)$.

Теорема 2. Решение задачи (1) задается выражением

$$x^{**} = \frac{c}{c+l}b + \frac{l}{c+l}a.$$

Рассмотрим случай, когда функция распределения $F(t)$ представима в виде конечной смеси известных

непрерывных строго возрастающих функций распределения $F_1(t), \dots, F_m(t)$:

$$F(t) = \sum_{i=1}^m \delta_i F_i(t),$$

где весовые множители $p_i > 0, i=1, \dots, m$ и $\sum_{i=1}^m \delta_i = 1$, то есть весовые множители образуют вероятностное распределение.

Рассмотренная вероятностная модель часто встречается на практике и соответствует случаю, когда генеральная совокупность представляет собой смесь нескольких относительно однородных совокупностей. В рассматриваемой задаче можно, например, считать, что $m=3$. При этом первая совокупность состоит из «хороших» лет, то есть лет с высокой урожайностью данной культуры, вторая совокупность состоит из «средних» лет, то есть лет, когда урожайность соответствует среднему уровню, и третья совокупность состоит из «плохих» лет, когда урожайность оказывается низкой.

Введем дискретную случайную величину с распределением

$$p\{v=i\}=p_i, i=1, 2, \dots, m.$$

Введем следующие определения:

$$x^* = \arg \min_{x \in [a,b]} Q(x, F),$$

$$x_i^* = \arg \min_{x \in [a,b]} Q(x, F_i),$$

$$\tilde{x}^* = x_1^* \cdot I\{v=1\} + x_2^* \cdot I\{v=2\} + \dots + x_m^* \cdot I\{v=m\},$$

$$\text{где } I\{v=i\} = \begin{cases} 1, & \text{если } v=i \\ 0, & \text{если } v \neq i \end{cases}$$

Здесь x^* - оптимальное решение для всей смеси F , x_i^* - оптимальное решение, когда точно известен номер совокупности i , \tilde{x}^* - оптимальное решение, когда точно идентифицируются события $\{v=i\}, i=1, \dots, m$.

Теорема 3. Если строго возрастающие функции распределения $F_1(t), F_2(t), \dots, F_m(t)$ непрерывны, то $M\tilde{Q}(\tau, \tilde{x}^*) \leq Q(x^*, F)$.

Возникает задача статистического оценивания теперь уже всех функций распределения $F_i(t), i=1, \dots, m$ или статистического оценивания соответствующих квантилей. Предпочтительнее построить оценки распределений, так как это позволит варьировать величины возможных потерь и оценить степень изменчивости оптимальной оценки. Оценка параметров смеси и построены оптимальные решения по найденным оценкам.

2 Оценка пространственной изменчивости и выделение однородных зон на сельскохозяйственном поле по урожайности

В основе точного земледелия лежит представление о возможности значительного повышения урожаев, существенной экономии ресурсов и снижения антропогенной нагрузки на окружающую среду путем дифференциации агротехнологий в соответствии с пространственной изменчивостью почвенных и иных факторов продуктивности в пределах отдельного сельскохозяйственного поля. Ключевым понятием этой концепции является адаптация элементов агротехнологии к внутривидовому варьированию условий, т.е. приспособление системы хозяйствования к пространственной неоднородности конкретного поля. Очевидно, что с точки зрения такого подхода первостепенное значение приобретает более глубокое изучение самой пространственной неоднородности сельскохозяйственных угодий, разработка методов ее количественного описания и выделения границ изменчивости на заданной территории.

2.1 Оценка биоэквивалентности двух участков на сельскохозяйственном поле

Предположим, что на некотором поле можно выделить два участка A и B . Требуется принять решение о степени однородности или неоднородности этих участков между собой по уровню средней урожайности некоторой культуры за несколько лет на этих участках. Предполагается, что каждый из участков состоит из большого числа небольших по площади делянок, средняя урожайность представляет собой суммарный урожай всех делянок, поделенный на их количество. Биоэквивалентность этих участков означает совпадение распределений этих случайных величин (средних урожайностей) или достаточную близость этих распределений. В дальнейшем будем считать эти случайные величины взаимно независимыми.

Предположим, что имеются две выборки, представляющие собой значения средних урожайностей участков A и B за несколько предыдущих лет: $Y_{A1}, \dots, Y_{An}, Y_{B1}, \dots, Y_{Bn}$. Представляется также допустимым предположить, что распределения введенных случайных величин Y_A и Y_B являются нормальными с параметрами μ_A, μ_B (математические ожидания) и σ_A^2, σ_B^2 (дисперсии). В качестве меры сходства случайных величин Y_A, Y_B , естественно выбрать вероятность

$$F_{A,B}(t) = P\{|Y_A - Y_B| \leq r\} \quad (2)$$

Если участки A и B биологически эквивалентны, то, как уже отмечалось выше, распределения средних урожайностей должны быть одинаковыми или достаточно близкими, поэтому вероятность (2.4) следует сравнивать с вероятностью

$$F_A(t) = P\{|Y_A - Y_{\bar{A}}| \leq r\}, \quad (3)$$

где случайная величина $Y_{\bar{A}}$ независима от Y_A и имеет такое же распределение. В качестве величины r в формулах (2) и (3) [3], выберем $t\sigma_A\sqrt{2}$. Тогда вероятности (2) и (3) можно переписать, используя функции стандартного нормального распределения, следующим образом

$$F_{A,B}(t) = P\{|Y_A - Y_B| \leq t\sigma_A\sqrt{2}\} = \Phi\left(\frac{t\sigma_A\sqrt{2} + \mu_B - \mu_A}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right) - \Phi\left(\frac{-t\sigma_A\sqrt{2} + \mu_B - \mu_A}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)$$

$$F_A(t) = P\{|Y_A - Y_{\bar{A}}| \leq t\sigma_A\sqrt{2}\} = \Phi(t) - \Phi(-t),$$

где $\Phi(t)$ – функция стандартного нормального распределения.

Близость функций $F_{A,B}(t)$ и $F_A(t)$ свидетельствует о том, что участки A и B можно считать биоэквивалентными. Более строго биоэквивалентность можно определить в терминах разности или отношения введенных функций, для чего необходимо ввести пороговые значения f_d для разности и f_r для отношения.

Будем говорить, что участки A и B биоэквивалентны, если

$$F_{A,B}(t) - F_A(t) \geq f_d \quad \text{или} \quad F_{A,B}(t) / F_A(t) \geq f_r,$$

где $f_d < 0$, $f_r < 1$.

Очевидно, что большее значение вероятности означает более высокую степень «близости» между средними урожайностями, что говорит о «более высокой биологической эквивалентности участков».

Используя имеющиеся выборки, построим оценки неизвестных параметров $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$, после чего, подставив их в функцию $F_{A,B}(t)$, получим статистическую оценку искомой вероятности: $F_{A,B}^*(t)$. А далее можно производить сравнение найденной оценки с $F_A(t)$. Заметим, что функция $F_A(t)$ полностью известна. Проблема заключается в том, что вместо истинной вероятности $F_{A,B}(t)$ в сравнении используется статистическая оценка $F_{A,B}^*(t)$. В связи с этим можно предложить следующий алгоритм, использующий идею статистического моделирования выборок из нормальных распределений с параметрами $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$ (отдельно для участка A и участка B).

Алгоритм:

- 1) По полученным в результате многолетних наблюдений опытными данным строятся оценки $\mu_A^*, \mu_B^*, \sigma_A^{2*}, \sigma_B^{2*}$.
- 2) На компьютере моделируются две выборки: выборка Y_{A1}, \dots, Y_{An} из нормального распределения с параметрами μ_A^*, σ_A^{2*} и выборка Y_{B1}, \dots, Y_{Bn} из нормального распределения с параметрами μ_B^*, σ_B^{2*} .
- 3) По полученным выборкам строится новая оценка $F_{A,B}^{**}(t)$, после чего для ранее выбранного порогового значения f_d (или f_r) производится проверка условия биоэквивалентности (2.6) (в форме разности или отношения).
- 4) Пункты 2 и 3 многократно повторяются и подсчитывается относительная частота выполнения условия биоэквивалентности.
- 5) Принимается решение о принятии гипотезы биоэквивалентности участков или об отклонении этой гипотезы.

Все вычисления производятся специальной программой, разработанной в среде *Maple*. Программная реализация позволяет при задаваемых различных уровнях урожайности проводить компьютерный анализ и оценку степени однородности тех или иных попарно сравниваемых участков на заданном сельскохозяйственном поле.

2.2 Выделение однородных зон на сельскохозяйственном поле по урожайности отдельных участков

Современные технологии точного земледелия позволяют получать данные по урожайности с точной привязкой к координатам каждого отдельного участка на поле. Рассмотрим поле, состоящее из большого числа элементарных участков (размер участка - несколько квадратных метров), по каждому из которых известна урожайность. На основе урожайностей по отдельным участкам конкретного года требуется произвести

Секция 4: Информационные технологии в АПК

разбиение поля на относительно однородные кластеры (зоны). Перенумеровав участки и зафиксировав урожайность на каждом из них, получаем массив данных, в котором содержится потенциальная информация об однородных зонах на поле. Предполагается, что урожайность внутри каждой из зон примерно одинакова, причем полное совпадение урожайностей невозможно.

Предположим, например, что все поле можно условно разделить на пять кластеров (пять зон относительно однородности). Первый кластер соответствует очень благоприятным условиям для произрастания данной культуры; второй кластер включает участки с хорошими условиями произрастания данной культуры; третий кластер соответствует в целом удовлетворительным условиям; четвертый кластер включает неблагоприятные зоны произрастания культуры и, наконец, пятый кластер включает участки с очень плохими условиями. Конечно, в действительности для конкретного поля количество однородных зон может быть меньше пяти, в идеальном случае все поле целиком может представлять собой одну однородную зону.

Общая математическая модель задачи кластеризации урожайности на поле, имеет вид:

$$f(x) = \sum_{j=1}^k p_j \cdot f(x, \theta_j),$$

где: k - число кластеров (зон однородности); $f(x, \theta_j)$ - плотность нормального распределения (описывает закон распределения урожайности x внутри зоны однородности с номером j); $\theta_j = (a_j, \sigma_j^2)$ - неизвестные параметры (математическое ожидание и дисперсия) распределения компоненты с номером j ; p_j - удельный вес зоны однородности (относительная доля наблюдений из этой зоны по отношению к общему числу наблюдений).

Для решения задачи кластеризации разработан алгоритм адаптивного вероятностного обучения, использующий разделение смеси вероятностных распределений, в котором отдельные компоненты моделируются нормальными распределениями с различными параметрами. При этом математическое ожидание соответствует средней урожайности с участка внутри одной зоны, а дисперсия характеризует разброс внутри зоны однородности. Веса компонент указывают относительный размер соответствующего кластера. Урожайности на первоначально выделенных участках следует использовать для оценки параметров изучаемой смеси распределений. Может оказаться, что в процессе разделения смеси распределений (оценки параметров смеси) количество компонент уменьшится, что повлечет за собой уменьшение количества кластеров на поле. В идеальной ситуации, для однородного поля останется только лишь одна компонента. Алгоритм был реализован в виде программы, которая используется в информационном обеспечении прецизионного внесения удобрений.

Заключение

Таким образом, в исследовании предложена математическая постановка оптимизационной задачи выбора оптимального момента времени для проведения агротехнологической операции в условиях стохастической неопределенности, объективно присутствующей в практике применения агротехнологий и обусловленной действием большого числа разнообразных причин. Рассмотрены следующие возможные случаи:

1.1. Для известной функции распределения оптимального момента времени проведения агротехнологической операции аналитически найдено решение сформулированной оптимизационной математической задачи. Полученный результат довольно просто реализуется на практике для наиболее часто применяемых вероятностных распределений.

1.2. Для неизвестной функции распределения оптимального момента времени проведения агротехнологической операции предложен минимаксный подход, в рамках которого аналитически найдено оптимальное решение, легко реализуемое на практике.

1.3. С целью уменьшения статистической неопределенности, связанной с выбором момента времени проведения агротехнологической операции, рассмотрена ситуация, когда функция распределения оптимального момента времени проведения агротехнологической операции представима в виде конечной смеси непрерывных строго возрастающих функций распределения, соответствующих различным условиям произрастания сельскохозяйственной культуры. Для этого случая также найдено оптимальное решение.

1.4. В рамках рассматриваемой математической постановки решена задача нахождения оценки оптимального момента времени проведения агротехнологической операции по накопленной многолетней статистической информации. Рассмотрен конкретный пример.

2. На основе биометрического подхода к определению биоэквивалентности двух относительно больших участков на сельскохозяйственном поле по урожайности некоторой сельскохозяйственной культуры за несколько лет с использованием статистического моделирования выборок из нормальных распределений разработан алгоритм оценки, выполнена его программная реализация.

3. Предложен математико-статистический метод выделения относительно однородных зон на сельскохозяйственном поле по урожайности отдельных небольших участков на поле за один год и разработан адаптивный вероятностный алгоритм решения этой задачи на основе разделения конечной смеси распределений. Создана программа выделения однородных технологических зон по продуктивности, получаемой автоматически с помощью зернового комбайна, оснащенного датчиками урожайности, бортовым компьютером и системой GPS.

Литература

1. План мероприятий по внедрению систем точного земледелия в Республике Беларусь. Утв. 2 февраля 2012 г. зам. Премьер-министра Республики Беларусь В.Н. Ивановым.
2. Якушев В.П. Статистический анализ опытных данных. Непараметрические критерии. – СПб.: АФИ, 2001. – 61с.
3. Плахотник С.В. Адаптивные методы прогнозирования временных рядов в среде Matlab- Труды III Всероссийской научной конференции «Проектирование научных и инженерных приложений в среде Matlab», СПб, изд. С-Петербург. ун-та, 2007. с.1363-1370.
4. Shall, R. Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. Biometrics 51, 615-626, 2005
5. Буре В.М. Методология статистического анализа опытных данных.-СПб.: Изд. С.-Петербург. ун-та, 2007. 141 с.
6. Буре В.М. Комплекс программ по непараметрической статистике в среде Matlab. С.-Петербург: Изд. С.-Петербург. ун-та, 2007. 84 с.

УДК 004:378.01

ОСОБЕННОСТИ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В АПК» В АГРАРНЫХ УНИВЕРСИТЕТАХ СНГ

Воюш В.И., к.э.н., доц., Шакирин А.И., к.т.н., доц. (БГАТУ, Минск)

Введение

В настоящее время, когда персональный компьютер стал одним из рабочих инструментов, владение технологиями обработки информации является очень важным в трудовой деятельности любого человека. Это предъявляет высокие требования к тем образовательным дисциплинам, которые связаны с информационными технологиями. Это же требует и их постоянного совершенствования, для чего следует учитывать лучший опыт их преподавания. В настоящей статье систематизируются те материалы по данному вопросу, которые были получены из Интернета. Конечно, данный обзор не претендует на то, что анализ всех курсов, читаемых в университетах, является полным, потому что проблематично просмотреть все сайты учебных заведений. К тому же, как отмечено в [1], очень большая часть организаций аграрного профиля стран СНГ не имеет возможности быть представленной в Интернете (правда, это не относится к университетам). А имеющиеся сайты обычно содержат информацию о самой организации (руководство, контакты, структура, далеко не всегда публикации сотрудников), а вот информация о разработках (а для университетов таковыми и являются изучаемые дисциплины) представлена крайне слабо.

Тем не менее, авторы убеждены, что представляемый обзор отображает реальную ситуацию, а «скрытые» (не найденные в Интернете) материалы не искажают общую картину в достаточно большой степени. Были рассмотрены ведущие университеты - РГАУ - МСХА имени К.А. Тимирязева и Санкт-Петербургский государственный аграрный университет.

Основная часть

Как и можно было ожидать, курс «Информационные технологии» читается практически во всех университетах. И если рассматривать чисто информационные технологии, то содержание этого курса примерно одинаковое во всех университетах. Этот курс, в основном, базируется на Windows-приложениях Excel и Access. Эта ситуация отражается и в используемых в настоящее время учебниках по информационным технологиям. Дополнительные дисциплины, которые изучаются в этих курсах, обычно представлены системными или прикладными пакетами.

Ситуация по использованию информационных технологий в АПК несколько иная. Обычно такие дисциплины так и называются: «Информационные технологии в АПК». Эта дисциплина также обычно базируется на Windows-приложениях Excel и Access, но они дополняются изучением и других систем. Часто это просто использование «обычных» технологий применительно к условиям АПК или работа с системными и прикладными программами. Такие курсы изучаются, в частности, в Российском университете дружбы народов, Ростовском институте переподготовки кадров, Ижевской государственной сельскохозяйственной академии, Орловском государственном университете, Петрозаводском государственном университете (там эта дисциплина называется «Информационные технологии в сельском хозяйстве»).

В некоторых университетах изучаются и более специализированные дисциплины. В ведущих университетах: в РГАУ-МСХА имени К.А. Тимирязева читаются курсы «Инновационные технологии в агрономии», «Экономика и управление на предприятии АПК», «Основы информационного обеспечения управления АПК», «Информационные технологии в экономике», в Санкт-Петербургском государственном аграрном университете - курс «Основы точного земледелия».

Наиболее фундаментальной из этих дисциплин выглядит представленная в РГАУ-МСХА имени К.А. Тимирязева «Инновационные технологии в агрономии» (объем 144 часа: лекции – 12, практические занятия - 18, семинары – 6, самостоятельная работа - 108). Целью этого курса является стремление научить самостоятельно обобщать информацию об инновационных технологиях в агрономии, анализировать