

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Produção

Um Sistema utilizando *Data Mining*
para o Zoneamento e Monitoramento Agrícola no
Brasil:
Um caminho para a qualidade de Software

Dissertação de Mestrado

Paulo Roberto Lobão Lima

Florianópolis

2001

PAULO ROBERTO LOBÃO LIMA

**Um Sistema utilizando *Data Mining*
para o Zoneamento e Monitoramento Agrícola no Brasil:
Um caminho para a qualidade de Software**

Esta dissertação foi julgada adequada e aprovada para obtenção
Mestre em Engenharia de Produção no
Programa de Pós-Graduação em Engenharia de Produção
da Universidade Federal de Santa Catarina

Florianópolis, 11 de julho de 2001.

Profº. Ricardo Miranda Barcia, Ph. D.
Coordenador do Curso

BANCA EXAMINADORA

Profª. Christiane Gresse Von Wangenheim, Drª.
Orientadora

Profª. Elizabeth Sueli Specialski, Drª.

Profº Alejandro Martins, Dr.

SUMÁRIO

LISTA DE FIGURAS	v
RESUMO.....	vi
ABSTRACT.....	vi
1 - INTRODUÇÃO.....	1
1.1 - DESENVOLVENDO SISTEMAS VISANDO USABILIDADE.....	1
1.2 - ÁREA DE ATUAÇÃO	3
1.3 - JUSTIFICATIVAS.....	4
1.4 - OBJETIVOS.....	5
1.5 - PREMISSAS	8
1.6 - CENÁRIOS E REQUISITOS DO PROJETO	8
2 - FUNDAMENTAÇÃO TEÓRICA	10
2.1 - USABILIDADE.....	11
2.2 - QUALIDADE.....	11
2.3 - SEMIÓTICA	13
2.3.1 - <i>Os componentes de um sinal.....</i>	<i>13</i>
2.3.2 - <i>A semiótica computacional.....</i>	<i>14</i>
2.4 - ESTRUTURAS COGNITIVAS.....	15
2.4.1 - <i>A disciplina de interação humano-computador.....</i>	<i>15</i>
2.4.2 - <i>As abordagens cognitivas</i>	<i>17</i>
2.4.3 - <i>As etapas do modelo de NORMAN.....</i>	<i>18</i>
2.4.4 - <i>A comunicação designer-usuário</i>	<i>20</i>
2.5 - INTERFACES INTELIGENTES	22
2.5.1 - <i>Tecnologia de sistema especialista.....</i>	<i>26</i>
2.5.2 - <i>Interface com o usuário.....</i>	<i>27</i>
2.5.3 - <i>Alguns problemas fundamentais.....</i>	<i>28</i>
2.5.4 - <i>Frontends inteligentes.....</i>	<i>30</i>
2.5.5 - <i>Arquitetura de um sistema amigável.....</i>	<i>32</i>
2.5.6 - <i>Agentes inteligentes</i>	<i>34</i>
2.6 - DATA WAREHOUSE.....	35
3 - PROPOSTA DE DATA MINING PARA O ZONEAMENTO E MONITORAMENTO AGRÍCOLA	38
3.1 - REDUÇÃO DOS RISCOS CLIMÁTICOS (RESULTADOS OBTIDOS).....	38
3.1.1 - <i>Áreas homogêneas.....</i>	<i>39</i>
3.1.2 - <i>Probabilidades de fatores climáticos.....</i>	<i>43</i>
3.1.3 - <i>Simulações probabilísticas</i>	<i>45</i>
3.2 - PROCESSO DE DATA MINING.....	49
3.2.1 - <i>Algoritmos de mineração.....</i>	<i>50</i>
3.2.2 - <i>Passos do processo de Data Mining proposto.....</i>	<i>56</i>
4 - VALIDAÇÃO	58
4.1 - VISÃO GERAL DO SISTEMA PROPOSTO	78
5 - CONCLUSÃO.....	80
BIBLIOGRAFIA.....	83

LISTA DE FIGURAS

FIGURA 1: Reestruturação do PROAGRO	4
FIGURA 2: Esquema semiótico de PEIRCE (1995)	13
FIGURA 3: Funções de um sinal	14
FIGURA 4: O <i>design</i> de sistemas e a interação usuário-sistema	16
FIGURA 5: O modelo metacomunicativo da engenharia semiótica	21
FIGURA 6: Consultores computadorizado	24/25
FIGURA 7: Arquitetura de um sistema de consultoria conversacional (SCHWARTZ, 1988)	34
FIGURA 8: Ciclo do projeto do <i>Data Warehouse</i> (KIMBALL, 1998).....	36
FIGURA 9: Áreas homogêneas	40
FIGURA 10: Gráfico média de precipitação pluviométrica	41
FIGURA 11: Áreas	42
FIGURA 12: Tabela de distribuições das estações	43
FIGURA 13: Gráficos das zonas A e E	44
FIGURA 14: Gráficos das zonas A e E (ajustado)	45
FIGURA 15: Gráfico do ISNA.....	48
FIGURA 16: Gráfico do ISNA (comportamento)	48
FIGURA 17: Gráfico do ISNA (probabilidade)	49
FIGURA 18: Switch de Software como Processo de <i>Data Mining</i>	56
FIGURA 19: Visão geral do sistema proposto.....	78
ANEXOS:	87

RESUMO

Visando atingir um padrão de qualidade, os profissionais da área de desenvolvimento de software tem a incumbência de construir sistemas computacionais como ferramentas que sejam úteis para os usuários. O objetivo é chamar a atenção para o desenvolvimento centrado no usuário, no qual a interface de usuário do sistema tem uma importância fundamental.

Conhecimentos de outras áreas em especial das ciências humanas são necessárias para o desenvolvimento de sistemas. A qualidade dos produtos e serviços oferecidos depende da capacitação desses profissionais que irão interpretar os verdadeiros anseios dos clientes ou usuários com o máximo de qualidade.

Palavra-chave : Qualidade, Interface, Desenvolvimento de Software, Clientes e Usuários.

ABSTRACT

In order to reach a quality standard, the professionals who work with development of software are supposed to create computer's systems with tools that fully attempt the users necessity. The main goal is to draw attention to the development centralized in the user, whose interface has a great importance.

Knowledge from other areas, specially the human science, is necessary to develop the system. The products and service's quality offered depend on the capacity of these professionals, who will interpret the real urge from clients and users with the highest quality.

Key word: Quality, Interface, Development of Software, Clients and Users.

INTRODUÇÃO

Desde o princípio a industrialização sempre foi um tema inquietante em nossa história. No novo século, somos movidos pela necessidade ainda mais acentuada, de produzir cada vez mais em menos tempo, com maior qualidade e inovação a custos baixos.

Nem mesmo as técnicas básicas de estruturação de projetos de software são seguidas em sua plenitude pelos desenvolvedores de softwares. Menor importância ainda recebe o projeto de interfaces de utilização. As técnicas de projeto centrado no usuário são pouco adotadas pelos desenvolvedores de software, especialmente no Brasil.

Este trabalho procura demonstrar, como o software deve ser amigável utilizando técnicas de interfaces, usabilidade e agentes inteligentes para que o resultado não seja a frustração dos usuários diante das dificuldades de obterem no software o apoio necessário para a execução de suas tarefas, o que, em situações extremas, leva ao abandono.

A inundação da informação tem intensificado o desejo para uma assistência computadorizada de como tomar decisão e solucionar os problemas. Dentro do ambiente de negócio, onde o acesso rápido aos dados críticos é essencial para tomar a decisão correta.

1.1 Desenvolvendo sistemas visando usabilidade

Cientistas têm argumentado que sistemas computacionais devem ser projetados como uma ferramenta intelectual que aumenta a capacidade humana. WINOGRAD (1996) ressalta que novas tecnologias serão mais eficazes quando projetadas para aumentar, ao invés de substituir, as habilidades dos usuários. Ele denomina desafio de usabilidade o projeto de novas tecnologias que buscam explorar ao máximo as habilidades dos usuários na

criação de ambientes de trabalho mais eficazes e produtivos. Este desafio, abandona a tradicional perspectiva que considera o usuário como peça integrante do sistema, no qual os únicos critérios são aqueles que visam minimizar o esforço físico dos usuários. Alguns fatores de usabilidades que devem ser considerados:

- ✓ facilidade de aprendizado;
- ✓ facilidade de uso;
- ✓ satisfação do usuário;
- ✓ flexibilidade; e
- ✓ produtividade.

NORMAN (1986) tem enfatizado que a tecnologia de construção de artefatos cognitivos deve ser projetada com o objetivo de ajudar as pessoas a serem mais espertas, eficientes e inteligentes.

As habilidades dos usuários, a situação de uso e o contexto onde eles estão envolvidos são fundamentais para a usabilidade e devem ser considerados. Sendo assim, é necessário uma clara definição do escopo global dos trabalhos, visando a criação de uma estrutura gerencial capaz de desenvolver, controlar, implantar e manter o projeto dentro das condições estabelecidas nas especificações, tais como:

- ✓ perfil e tipo do cliente ou usuário;
- ✓ clara definição dos conjuntos de objetivos, metas e especificação de requisitos;
- ✓ disponibilidade tecnológica e de pessoal qualificado;
- ✓ capacitação da equipe, e
- ✓ relação com empresas e parceiros.

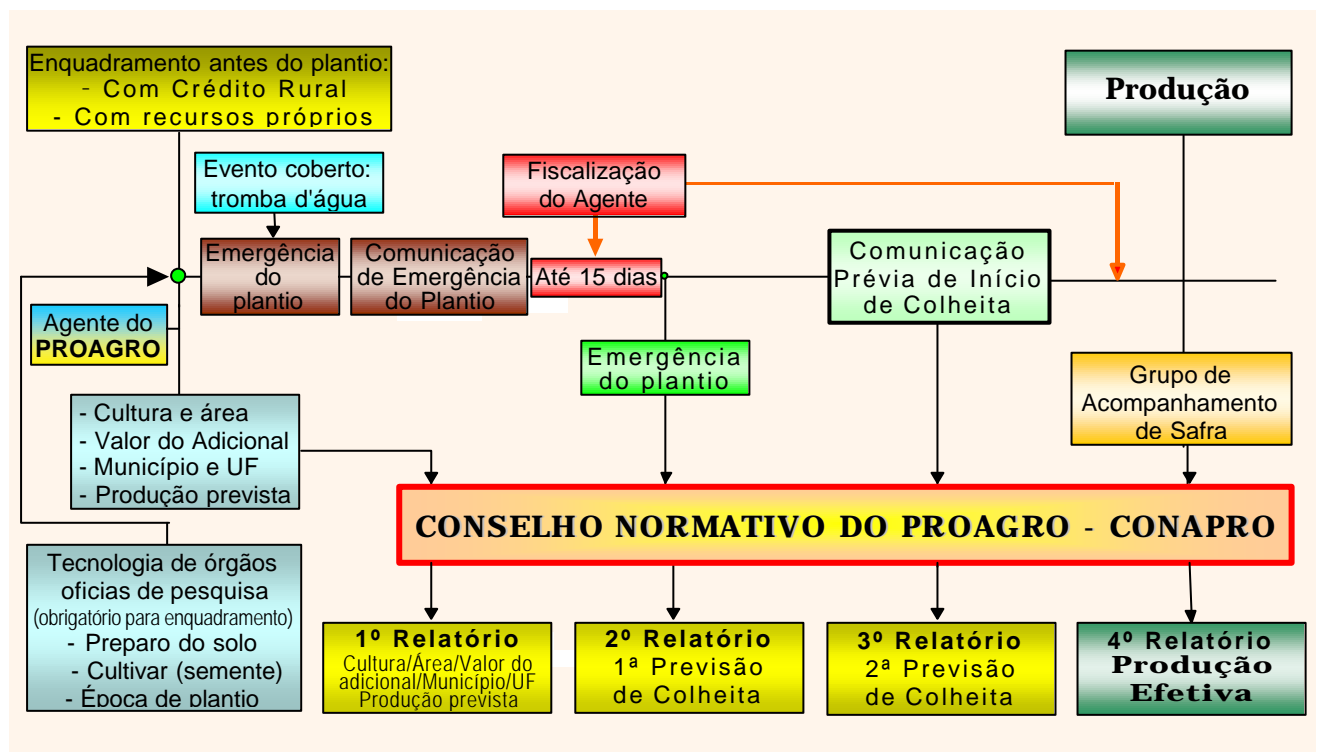
1.2 Área de atuação

O Programa de Zoneamento Agrícola do Ministério da Agricultura e do Abastecimento, coordenado pela Secretaria da Comissão Especial de Recursos - CER/PROAGRO, firma-se como valioso instrumento de apoio à Política Agrícola do Governo Federal, bem como, de difusor de tecnologia e indispensável suporte para a tomada de decisões no âmbito do PROAGRO. Para acompanhar os resultados desse projeto, instituiu-se o serviço de Monitoramento das operações enquadradas no PROAGRO dentro do Zoneamento Agrícola que, conta com a importante parceria da FINATEC, da EMBRAPA, do INMET, da ANEEL, do IAPAR, da EPAGRI-SC e da UNICAMP, que emprestam suas experiências nas áreas de pesquisa agropecuária e climatológica, visando alcançar os resultados esperados.

Os projetos, com base em dados técnico-científicos, oferecem orientações de períodos de plantio por município, para cada cultura/cultivar e tipos de solo, de modo a evitar-se as adversidades climáticas (seca, chuvas excessivas, geada e granizo), responsáveis por significativo percentual de perdas na agricultura e reflexos negativos no abastecimento e nos preços dos produtos.

Ressalte-se que as informações oriundas desse trabalho serão divulgadas e disponibilizadas às associações de produtores, entidades de assistência técnica e extensão rural, agentes financeiros, cooperativas, secretarias de agricultura e entidades públicas e privadas ligadas ao setor agrícola, de modo a atualizar os agricultores que já aderiram ao zoneamento agrícola e, também, para aqueles que ainda não o fizeram, disponibilizando este importante instrumento de avanço tecnológico.

Ressalte-se, ainda, que coube ao Ministério da Agricultura e do Abastecimento, através da Secretaria da Comissão Especial CER/PROAGRO, sanar uma ação administrativa que já deveria estar em execução a bastante tempo, e cujo descaso ocasionou prejuízos incomensuráveis aos cofres público, conforme diagrama abaixo:



1.3.1.1 FIGURA 1 – Reestruturação do PROAGRO

- ✓ precipitação pluviométrica;
 - ✓ evapotranspiração;
 - ✓ ventos fortes e vendavais;
 - ✓ geadas;
 - ✓ chuvas fortes na colheita;
 - ✓ granizo;
 - ✓ tromba d'água; e
 - ✓ incidência de seca.
- 6) Coletar informações pluviométricas no DNAEE e no INMET;
 - 7) Analisar e criticar os dados pluviométricos;
 - 8) Organização de um banco de dados referente a pedidos de cobertura securitária de riscos agrícolas no Brasil, com abrangência particular no PROAGRO;
 - 9) Implantação de um sistema computacional (interfaces) para cálculo das probabilidades de sinistros agrícolas e das probabilidades de perdas na safra;
 - 10) Concepção e implantação de um sistema computacional que permita identificar os municípios cujas perdas de safra sejam significativamente superiores às previstas estatisticamente;
 - 11) Estabelecer análise freqüencial e a espacialização dos dados;
 - 12) Imprimir e disponibilizar mapas de ocorrência de veranicos nas regiões brasileiras;
 - 13) Tabular análise freqüencial da precipitação pluviométrica para períodos de 10, 15 e 30 dias, com freqüência de 20, 50 e 80%, caracterizando precisamente os anos e utilizando-se das mesmas informações do item anterior, com impressão de mapas;

- 14) Avaliar e regionalizar as chuvas de maior intensidade no território nacional, utilizando-se das mesmas informações dos itens precedentes, com impressão de mapas;
- 15) Regionalizar e otimizar as datas de plantio para as culturas de trigo, para a região centro – sul brasileira, e de arroz, feijão, milho, soja e algodão para todo o território nacional;
- 16) Estabelecer cálculos de evapotranspiração potencial para as estações agrícolas;
- 17) Coleta de informações agronômicas das culturas de feijão, milho, arroz, soja, trigo;
- 18) Cálculo e estabelecimento da capacidade de retenção dos principais solos cultivados;
- 19) Simulação do modelo do balanço hídrico para culturas escolhidas para 9 datas de plantio diferentes, 3 ou 4 tipos de solo e culturas precoces, normais e tardias;
- 20) Espacialização dos resultados dos índices de penalização obtidos do balanço hídrico; e
- 21) Impressão e disponibilização de mapas com datas de plantio otimizadas para as culturas de arroz, feijão, milho e soja e algodão em todo território nacional, e para a cultura do trigo na região centro - sul brasileira.

Portanto, trata-se, de um trabalho importantíssimo para o país, onde o seu desenvolvimento e execução possibilitarão a definitiva implantação de uma metodologia atuarial adequada e informatizada, sobre a qual já nos referimos e cuja ausência limita o desenvolvimento da Seguridade Agrícola no Brasil.

1.5 Premissas

Não faz parte do escopo deste trabalho provar as premissas abaixo. Futuros estudos, ao analisarem os resultados aqui obtidos, devem levar em conta que se assume como verdadeiro que:

- a) o método está sendo implementado através da elaboração e do uso de modelos matemáticos-estatísticos não triviais, devendo se transformar no primeiro passo para uma revisão profunda dos preceitos atuariais vigentes, afim de torná-los adequados à realidade do setor agrícola;
- b) os princípios ergonômicos propostos pela norma ISO 9241, podem ser tão gerais quanto qualquer conjunto de critérios proposto. O grau de detalhamento de outros critérios existentes ou que venham a ser propostos será determinante para obtenção de maior ou menor precisão em uma avaliação de usabilidade;
- c) restrição das causas de cobertura do PROAGRO; e
- d) redução das alíquotas de adicional do PROAGRO;

1.6 Cenários e requisitos do projeto

No futuro próximo, com o termino deste trabalho, teremos um sistema gerenciador computadorizado que irá não só dar ao executivo informação básica, mas também irá monitorar planos de lucros, análise de resultado com divergências sem qualquer demora quando elas ocorrerem, determinar causas possíveis destas divergências, projetar seus efeitos e recomendar opções para resolver estes problemas. É claro, que a decisão final deverá ser sempre feita pelo diretor ou pelo executivo.

Nós iremos explorar no Capítulo 2 – Fundamentação Teórica – alguns dos conceitos de interfaces, usabilidade, qualidade e questões técnicas circundantes ao desenvolvimento de um sistema com habilidade automática de planejar estratégias de alto nível ou fazer análise.

Como resultados esperados, podemos apontar:

- ✓ redução das solicitações de cobertura por eventos climáticos sinistrantes. Na safra de inverno e safra verão, em operações monitoradas e nos pleitos de
- ✓ inibição e a conseqüente diminuição das solicitações fraudulentas, uma vez que o acompanhamento sistemático das ocorrências climáticas e o monitoramento das operações securitárias, em todas as suas fases, coíbem a realização de pagamentos indevidos;
- ✓ disponibilização de informação gerências necessárias a uma melhor gestão do PROAGRO; e
- ✓ disponibilização de informações sistematizadas que possibilitam avaliação permanente dos resultados do Zoneamento Agrícola e seu constante

Principais requisitos à serem utilizados pelo projeto:

- ✓ equipamentos de monitoração para coleta de dados;
- ✓ sistemas de automação agrícola (Robótica);
- ✓ agricultura de precisão (nas áreas de engenharia agrícola, engenharia florestal, ciência do solo, ciência dos alimentos e fitotecnia);
- ✓ sistemas de controle de ambientes;
- ✓ GIS - Sistemas de informações geográficas;
- ✓ GPS – Sistema de posicionamento global; e
- ✓ sistemas de apoio à decisão aplicados às diversas áreas da agricultura.

FUNDAMENTAÇÃO TEÓRICA

Normalmente, se faz muita confusão com a terminologia usada. Palavras como interação, interface e aplicação precisam ser compreendidas. Algumas pessoas confundem interface com interação, ou interface com aplicação, ou até mesmo não sabem diferenciar se a interface é apenas o software ou se envolve outros componentes adicionais de hardware e

Um sistema interativo permite que o seu funcionamento possa ser controlado por seus usuários. Num sistema interativo é importante distinguir a interface de usuário, que determina um modelo de interação, do seu núcleo funcional, que determina a sua funcionalidade.

A *aplicação de software* pode ser vista como uma *virtualidade*, ou seja, uma máquina virtual destinada a resolver problemas específicos de seus usuários. São máquinas que operam sobre o domínio conceitual de informação e conhecimento. Uma aplicação de software, como entidade virtual, existe apenas em um sistema computacional. O hardware funciona como um meio que possibilita a existência desta máquina simbólica.

O termo *funcionalidade* se refere àquilo **que** um programa faz e, no caso de software interativo, o que ele deve oferecer para seus usuários. Funcionalidade é diferente de *funcionamento* que se refere a **como** um software realiza as suas funções. A funcionalidade é determinada pelo *modelo funcional da aplicação*, também chamado de modelo de funcionalidade. Na engenharia de software, o modelo de funcionalidade é elaborado na fase de *especificação funcional de software* a partir das informações obtidas durante a *análise de requisitos*.

1.7 Usabilidade

WINOGRAD (1996) apresenta uma proposta mais radical e inovadora para usabilidade. Do ponto de vista dele, usabilidade deve ser vista como a qualidade que ao mesmo tempo satisfaz as necessidades do usuário, se acopla às suas capacidades e conhecimentos, considera o impacto da tecnologia no contexto de trabalho e integra o usuário em tal contexto. Neste sentido, sistemas computacionais devem ser vistos como artefatos cognitivos, ferramentas intelectuais ou mídia.

Segundo MORAN (1988) e NORMAN (1986), o modelo conceitual do usuário, é o conhecimento necessário para o usuário adquirir competência para utilizar o sistema, onde envolve tudo aquilo o que se pode fazer (*a funcionalidade*) e como se pode interagir (o *modelo de interação*). Para o design da aplicação de software é fundamentalmente o design do modelo de interação e do modelo funcional da aplicação.

Durante o processo de interação com o sistema o usuário precisa recorrer ao seu modelo conceitual de usabilidade, para poder desempenhar as ações que acionam as funções do sistema necessárias para atingir suas metas.

O *design de interface de usuário* é, portanto, o processo de concepção dos objetos de software e hardware que determinam os modos e as estruturas de interação (o modelo de

1.8 Qualidade

Qualidade é um outro conceito que está relacionado com as características específicas e peculiares de um determinado bem ou serviço, e que de acordo com um padrão de referência, pode ou não satisfazer as necessidades do usuário (LAWRENCE, 1994).

Sendo definida como a propriedade peculiar de um determinado bem ou serviço, segundo sua necessidade e utilidade para o usuário, avaliada pela relação benefício/custo (CAMPOS, 1994), conforme os conceitos a seguir:

- ✓ adequabilidade;
- ✓ confiabilidade;
- ✓ cumprimento da missão;
- ✓ eficiência;
- ✓ estética;
- ✓ funcionalidade;
- ✓ garantia;
- ✓ integração contextual;
- ✓ manutenibilidade;
- ✓ resistência;
- ✓ velocidade de resposta; e
- ✓ satisfação.

Tão importante quanto o próprio conceito de qualidade é o estabelecimento de um Índice de Suficiência da Qualidade - ISQ. Este índice determina até que ponto deve-se investir na qualidade, sem que haja comprometimento das necessidades básicas do usuário. As necessidades podem ser explícitas ou implícitas (ATAIR, 1993). As explícitas referem-se àquelas manifestadas e de maior percepção por parte do usuário, geralmente são de natureza estrutural, exigidas por normas ou procedimentos específicos, e nem sempre são notadas ou percebidas pelo usuário, visto que não fazem parte do conhecimento deste. O desenvolvedor do sistema tem uma maior ingerência na tomada de decisões relacionadas com as necessidades implícitas, devido sua responsabilidade no desempenho global do sistema. O ISQ favorece o desenvolvimento de sistemas pela

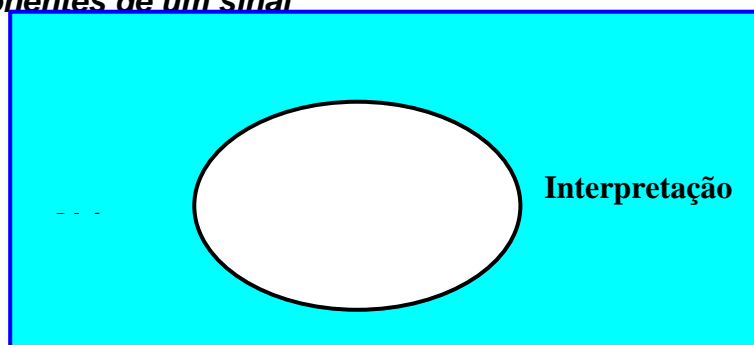
metodologia da prototipação, pois pode ser escolhido para cada um dos diferentes módulos que compõem o sistema completo, e pode variar entre cada um deles (STRAFACCI, 1996).

A definição do Índice de Suficiência da Qualidade depende fundamentalmente da condição social, cultural e funcional do usuário ou do solicitante de um bem ou serviço (STRAFACCI, 1996).

1.9 Semiótica

Semiótica: ciência geral das linguagens, foi lançada por PEIRCE (1995) no início do século XX, com enfoque no estudo da lógica dos sistemas de sinais: linguagens, códigos, sinalização, etc., tratando a interação humano-computador como um processo de comunicação entre dois sistemas cognitivos que fazem uso de informação simbólica.

1.9.1 Os componentes de um sinal

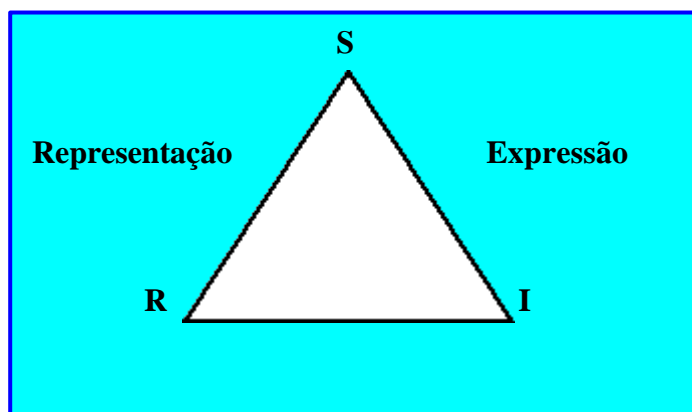


1.9.1.1 FIGURA 2 – Esquema semiótico de Peirce (1995)

A tríade de PEIRCE mostrada na Figura acima, é uma representação dos componentes dos sinais e de suas inter-relações. Ele envolve um sinal (ou expressão) **S**, um

objeto de referência (ou conteúdo) **R**, e uma interpretação **I**. Segundo este esquema, um sinal ocorre somente quando ele for interpretado na mente de uma pessoa.

Ele é visto como conhecimento quando integrar as funções de expressão conforme HJELMSLEV (1969) e de representação de modo a passar algum conhecimento geral ou específico sobre o objeto representado.



1.9.1.2

FIGURA 3 – Funções de um sinal

1.9.2 A semiótica computacional

A semiótica computacional proposta por ANDERSEN (1993) é a utilização de sinais computacionais na sociedade atual. Suas propostas são baseadas em duas das interpretações do esquema estruturalista elaborado por HJELMSLEV (1969) e a tríade de PEIRCE. No centro de sua perspectiva está o indivíduo, considerado como o criador, o intérprete e a referência dos sinais. Assim, não se pode dizer que um projetista conceba sinais. Ele propõe sinais, que em algumas circunstâncias se realizam, mas que em muitas outras nunca atingem a realização. Programar, no sentido semiótico do termo é, segundo ANDERSEN (1993) usar o computador para tentar dizer algo às pessoas. Deste modo, os sinais computacionais são definidos como sinais candidatos. Eles dependem do usuário para se realizarem como sinais.

Desta forma, o computador é visto essencialmente, como um meio para a comunicação. Em um sistema informatizado é o projetista que atua como emissor ou receptor.

Nesta visão o computador não possui as faculdades de um emissor ou de um receptor, ao contrário de pessoas, que articulam uma linguagem mesmo sem conhecer seu **programa** ou gramática.

A interface humano-computador é vista como componente e passa a ser entendida como processo de um sistema. Segundo esta definição pode-se afirmar que um sistema informatizado possui inúmeras interfaces, uma vez que cada usuário entra em contato com uma coleção diferente de sinais os quais ele interpreta de uma maneira particular. A relação que se estabelece entre o usuário e as partes perceptíveis do sistema faz com que uma nova interface nasce do sistema informatizado cada vez que ele é utilizado.

ANDERSEN (1993) sugere que a descrição de um sistema interativo baseada em cenas deve ser feita em dois níveis. O primeiro descreve cenas genéricas, como por exemplo aquelas ligadas ao gerenciamento do sistema de janelas, da manipulação de arquivos, dos dispositivos de entrada e saída, etc. Num segundo nível ocorre a descrição das cenas associadas às tarefas específicas de um aplicativo.

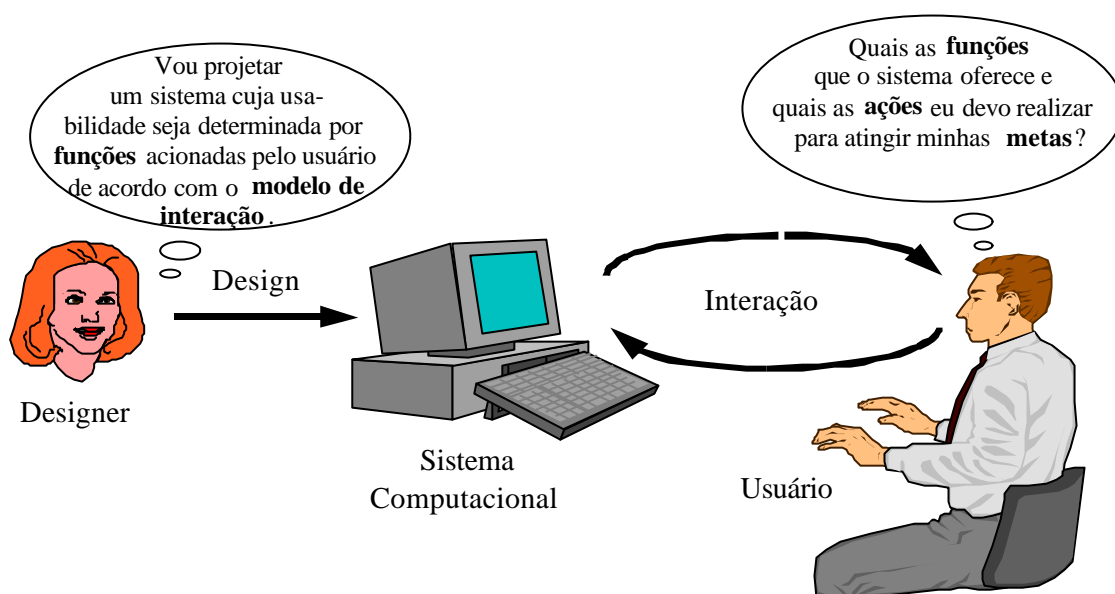
Assim a concepção de um sistema interativo pode se realizar como um processo de inserção de novas cenas em um livro já escrito e comercializado, como aquelas definidas nos sistemas MS-Windows, X-Windows, MacApp, etc.

1.10 Estruturas cognitivas

1.10.1 A disciplina de interação humano-computador

A capacidade e limitação física do ser humano requerem o desenvolvimento de um hardware de interface específico. Estudos de como o ser humano enxerga ou escuta e suas habilidades manuais são fontes de conhecimento para a tecnologia de entrada e saída.

Da mesma forma, o design do modelo de interação requer conhecimento proveniente de estudos de psicologia que indiquem capacidades de memorização, raciocínio e aprendizado, dentre outras faculdades mentais.



1.10.1.1 FIGURA 4 – O design de sistemas e a interação usuário-sistema

O estudo de como os sentidos e as capacidades motoras, permitem às pessoas utilizarem máquinas e ferramentas complexas já era objeto de estudo da disciplina *Fatores humanos* (E.U.A), ou *Ergonomia* (Europa), que surgiu nos anos 80, com rótulo de Interação entre Seres Humanos e Sistemas Computacionais (IHC).

O IHC é a disciplina que visa fornecer aos pesquisadores e desenvolvedores de sistemas explicações e previsões para fenômenos de interação usuário-sistema e resultados práticos para o design da interface de usuário, buscando desenvolver modelos teóricos da

performance e cognição humana, bem como técnicas efetivas para avaliar a usabilidade (LINDGARD, 1994).

Um dos objetivos dos pesquisadores envolvidos com IHC, foi chamar a atenção dos desenvolvedores, para considerar no desenvolvimento de sistemas, não apenas o sistema e sua interface, mas também, usuários envolvidos e o processo de interação entre eles. A esta proposta de mudança do foco e perspectiva deu-se o nome de Design de Sistemas Centrado no

Portanto, o IHC é multidisciplinar e está relacionada com áreas tais como: ciência da computação, que sustenta o conhecimento tecnológico para a engenharia de software e hardware; ergonomia, para o estudo do desempenho físico do usuário; psicologia e ciência cognitiva, que oferecem conhecimento sobre comportamento e habilidades perspectivas e cognitivas, bem como técnicas de análise e avaliação empírica; sociologia, para ajudar o designer a entender o contexto da interação; design gráfico, para a produção dos elementos gráficos apresentados; marketing, para a sua comercialização; antropologia, para o estudo de características humanas; e várias outras. PREECE (1994) acrescenta ainda engenharia,

O IHC deve contribuir com a reunião de modelos teóricos e métodos práticos para design que possam ser aplicados sem o conhecimento específico em cada área. A característica multidisciplinar abre espaço para se envolver novas disciplinas.

1.10.2 As abordagens cognitivas

As abordagens dominantes que têm caracterizado IHC são as cognitivas (PREECE, 1994). Os resultados delas são de longe mais numerosos do que os de qualquer outra. As abordagens cognitivas têm por objetivo a aplicação de estudos de psicologia cognitiva, ciência

cognitiva e inteligência artificial para a compreensão das capacidades e limitações da mente

O objetivo é oferecer uma abordagem mais subjetiva (interpretativa) dos processos mentais, examinando o papel do conhecimento, compreensão, aprendizado e significado na

A idéia básica é que modelos cognitivos que descrevam tais processos e estruturas, como os de recordação, interpretação, planejamento e aprendizado, podem indicar para pesquisadores e projetistas quais as propriedades que os modelos de interação devem ter de maneira que ela possa ser desempenhada mais facilmente.

Um dos trabalhos pioneiros na aplicação de psicologia cognitiva a IHC é o de MORAN (1994), cujo objetivo é a utilização de técnicas cognitivas para a construção de modelos que possibilitem aos designers fazerem previsões de desempenho dos usuários e, conseqüentemente, oferecer condições para a tomada de decisões.

NORMAN (1986) é um pesquisador da área que vem influenciando o trabalho de vários outros pesquisadores. NORMAN (1986) apresentou um modelo das atividades do usuário na interação com o sistema que funciona como um quadro teórico de referência, uma vez que vários outros modelos podem ser alocados em sua moldura.

1.10.3 As etapas do modelo de NORMAN

O modelo de NORMAN (1986) permite ao usuário mapear suas metas em comandos e funções do sistema. Segundo esta proposta, o papel do designer está em desenvolver uma interface que permita ao usuário, durante o processo de interação, adquirir um modelo mental correspondente ao modelo do sistema. Para isto são necessários modelos cognitivos das atividades mentais. Ele considera que este modelo mental do usuário é adquirido durante a

interação com a *imagem do sistema* - a interface, o sistema de ajuda e toda a documentação do sistema.

Como fundamentação teórica, o modelo de NORMAN (1986) nos revela as atividades desempenhadas pelo usuário, permitindo identificar quais elementos devem ser projetados e qual o impacto que eles terão no desempenho do usuário:

- 1) ***Estabelecimento da meta*** → O usuário utiliza um sistema como ferramenta para realizar uma tarefa. Para isto ele deve estabelecer uma meta. Então, o usuário deve ter conhecimento a respeito dos objetos do sistema e como eles podem ser modificados. O designer deve preocupar-se em deixar claro para o usuário quais são os objetos do sistema e que tarefas podem ser realizadas;
- 2) ***Planejamento ou formulação da intenção*** → A meta que o usuário deseja apenas pode ser atingido por alguma das funções do sistema. Caso exista esta função a meta será atingida diretamente por ela. Senão, o usuário deve realizar um planejamento mental no qual a meta inicial deve ser dividida em submetas, sucessivamente, até que para cada uma delas exista uma função do sistema que permita atingi-la. Este planejamento só é possível se o usuário conhecer as funções que o sistema oferece (denominado de funcionalidade);
- 3) ***Especificação da estrutura de ações*** → uma vez realizado o planejamento, o usuário precisa especificar as ações que controlam a função planejada. Estas ações podem estar estruturadas em seqüência, combinação simultânea, repetição, etc.;
- 4) ***Execução*** → As atividades anteriores são a preparação mental para a execução física propriamente. Nesta etapa existe o esforço físico de acionar o hardware e os *widgets* da interface. Cabe ao designer minimizá-lo, bem como aumentar a

produtividade do usuário através de um projeto de comandos e de hardware de interfaces adequados;

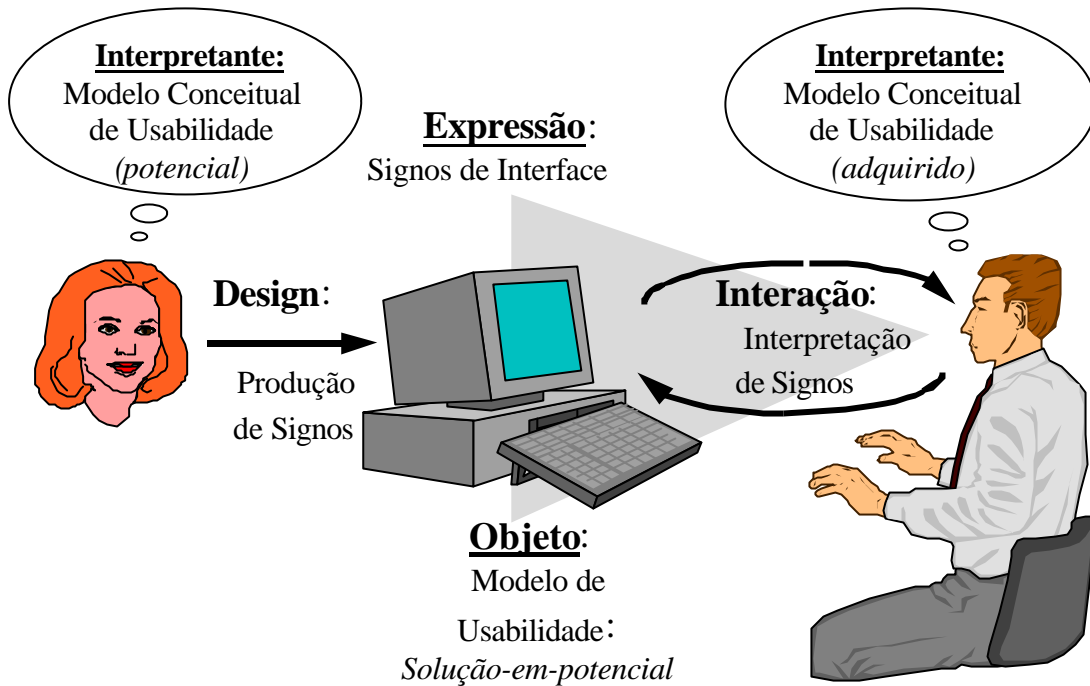
- 5) **Percepção** → Escolher o canal sensorial adequado (normalmente o visual e o auditivo) e qual a distinção naquele canal chamará a atenção do usuário;
- 6) **Interpretação** → Aquilo que é percebido pelo usuário precisa ser interpretado para que possa ser avaliado. Cada distinção num canal sensorial - símbolo, palavra escrita, ruídos, palavra sonora, ícone - quando interpretado pelo usuário é um signo que representa algo sobre o sistema e deve permitir ao usuário avaliar se a meta foi atingida;
- 7) **Avaliação** → A avaliação é realizada através da comparação daquilo que foi interpretado com a meta que havia sido estabelecida inicialmente.

A perspectiva cognitiva mostra que a facilidade de aprendizado deve ser solucionada com o design de interfaces que proporcionem o menor esforço cognitivo para as atividades e na aquisição do modelo conceitual. O objetivo destas abordagens é descobrir o conhecimento teórico a respeito das atividades mentais que orientam o design da interface para usabilidade.

1.10.4 A comunicação designer-usuário

Como na perspectiva central da engenharia semiótica o designer comunica-se com o usuário através da interface, portanto, ele deve projetar a interface de maneira a melhor transmitir o potencial de aplicação do sistema.

A interface de usuário veicula uma Mensagem do Designer, realizando um ato comunicativo unidirecional, do designer para o usuário, e é portanto um Signo.



1.10.4.1 FIGURA 5 - O modelo metacomunicativo da engenharia semiótica.

- ✓ significado da Mensagem do Designer (seu *objeto*) é o Modelo de Usabilidade;
- ✓ significado atribuído pelo usuário (seu *interpretante*) é uma abstração do Modelo de Usabilidade, e será referido como o Modelo conceitual de usabilidade do
- ✓ significado pretendido pelo designer (seu *interpretante intencional*) é uma abstração do Modelo de Usabilidade, e será referido como o Modelo conceitual de usabilidade do designer (potencial);
- ✓ significado pretendido pelo designer e o significado atribuído pelo usuário à mensagem da interface não são necessariamente idênticos.

O designer pode comunicar com modelo potencial de usabilidade usando diferentes *mídia* tais como o manual do usuário, sistemas de ajuda e outros.

1.11 Interfaces inteligentes

A interface é tanto um *meio* para a interação usuário-sistema, quanto uma *ferramenta* que oferece os instrumentos para este processo comunicativo. Quando se considera a aplicação como máquina(s) virtual(is), a interface pode ser considerada ainda como um *ambiente virtual para ações*.

O *modelo de interação* é o conjunto de protocolos que permite ao usuário interagir com a aplicação. Estes protocolos muitas vezes são determinados por uma linguagem, e por

O usuário precisa conhecer o modelo de interação para poder interagir como o sistema. MORAN (1994) denomina-o de modelo conceitual do usuário e afirma que o design da interface é o design deste modelo. Entretanto é importante diferenciar o modelo conceitual do usuário do modelo que é formalmente definido através da linguagem de

Um dos maiores problemas que as empresas enfrentam hoje, é a administração das informações, que as vezes tem uma maior concentração na gerência, causando deslumbramento e deixando-os sem saber como lidar com tais informações. A questão é, que tipo de sistema pode ou irá fornecer o suporte gerencial requerido para apoiar as decisões imaginadas?

Historicamente, as empresas tem utilizado sistema de gerenciador de banco de dados (DBMS) e sistema de suporte a decisão (DSS) para gerenciar, monitorar, entender e interpretar o desempenho associado. Isto tem sido o papel do DBMS para servir como uma coleção de pontos para o desempenho dos dados e para prover instrumentos que descrevam situações do passado e do presente. O DSS, tem sido usado para examinar o futuro, isto é, para avaliações de impactos de decisões em potencial, para avaliar caminhos alternativos de

ação e para mensurar negócios com algumas situações diferentes. Alguns pesquisadores trabalham na área de suporte executivo, isto é, conforme as atividades, requisitos, habilidades e estilo do trabalho pessoal do executivo que são, quase sempre, insuficientes para requerer uma nova forma de tecnologia da informação para lidar com estas tarefas (ROCKHART, 1988). Esta tecnologia dar-se o nome de sistema de informação executiva (EIS).

A presente geração de EIS são basicamente janelas de frontend's para DBMS e DSS de computadores de grande porte. O ponto principal desses sistemas são customizados, relatórios com situação resumida e gráficos que são usados para monitorar variações de ações planejadas. A maior forma de fornecer análise por estas ferramentas é o *drill-down*. Não há uma exibição da informação resumida, mas ele fornece um explanação generalizada e computadorizada dos resultados. Alguns EIS, portanto, não fornecem virtualmente suporte para monitorar automaticamente as variações complexas ou rumos para determinar as implicações dos resultados, ou ainda, recomendar sugestões dos caminhos das ações. Estas tarefas são de domínio de quem planeja ou analisa.

A fim de fornecer suporte automatizado para estas tarefas, precisamos, antes de mais nada, automatizar o conhecimento de quem planeja ou de quem analisa; ambos dominam o conhecimento e o conhecimento sobre como utilizam as ferramentas analíticas. Estes conhecimentos disponíveis pelo planejador e o pelo analisador são:

- ✓ determinar qual o dado para ser investigado ou monitorado;
- ✓ determinar qual ferramenta para ser usada na narração e análise dos dados;
- ✓ acompanhar os requisitos das análises;
- ✓ interpretar e entender os resultados analíticos; e
- ✓ sugerir recomendações no curso da ação;

Em outras palavras, o tipo de sistema que nós necessitamos criar é um assistente amigável ou consultor (KEEN, 1983).

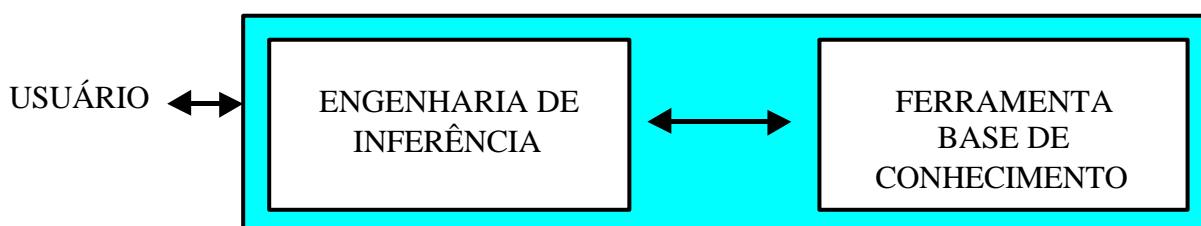
O desenvolvimento de um consultor amigável é um avanço em vários caminhos.

Algumas das alternativas abordadas são:

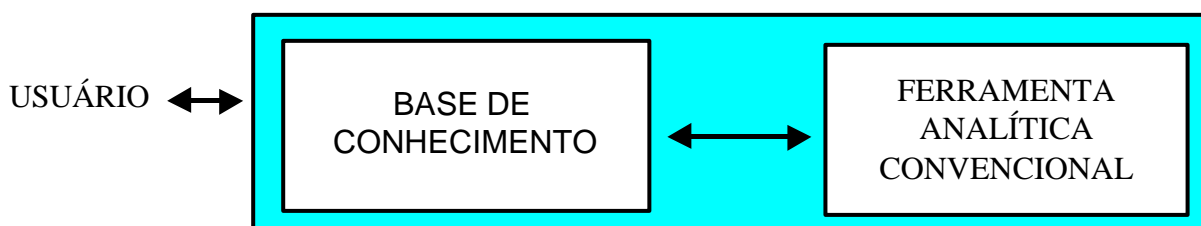
- ✓ sistemas especialistas (ES) num domínio específico;
- ✓ frontend's inteligentes (IFE) servindo de interface para ferramentas analíticas mais convencionais;
- ✓ sistema de Consultoria Conversação ou amigável (CAS) baseado na interface da linguagem natural construída ao redor do sistema de base de conhecimento; e
- ✓ agentes Inteligentes (IA) usado para monitorar eventos e ações nas tarefas automatizadas baseadas nesses eventos.

Os diagramas da Figura 6 resume algumas das diferentes arquiteturas de muitas abordadas. Desta Figura, nós podemos ver que todas as abordagens representam um sistema de base de conhecimento para o papel de um consultor. O que os distingue são os tipos de conhecimentos embutido no sistema.

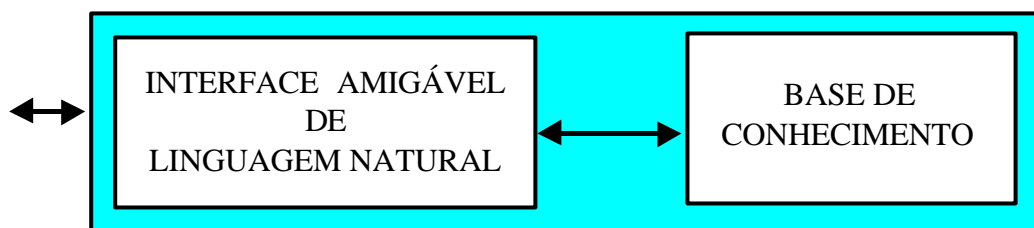
Sistemas especialistas



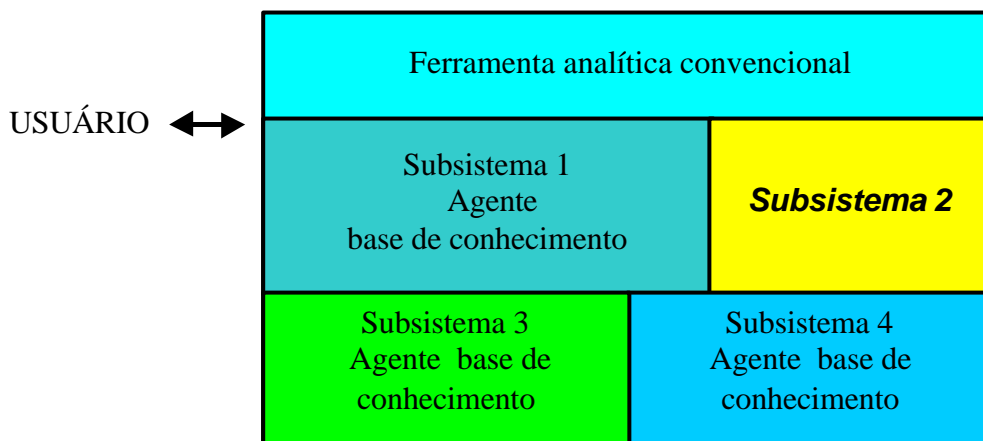
Frontends inteligentes



Sistema de consultoria amigável (Conversacional)



Agentes inteligentes



1.11.1.1 FIGURA 6 - Consultores computadorizados

Segundo ROCKHART e DELONG (1988) estudos primários de EIS, num rápido exame da existência de sistemas mostram que quase todos eles são usados para monitorar e determinar as causas do comportamento da empresa. O início deste processo é um relatório consolidado, examinado pela direção diariamente, semanalmente ou mensalmente. Resumidamente, assim se apresenta:

- 1) Descreve e monitora as mudanças nas variáveis importantes;
- 2) Interpretam e explicam as causas destas mudanças;
- 3) Projetam o impacto da continuação destas mudanças;
- 4) Determinam se algo deverá ser feito para alterar as mudanças;

Em algum ponto do projeto uma importante questão será respondida, qual o tipo de conhecimento é necessário para acompanhar estas tarefas?

1.11.2 Tecnologia de sistema especialista

A teoria diz que um consultor amigável computadorizado pode ser criado pelo um número de produção de ES dentro de domínio especializado, combinando-os dentro de um simples sistema.

Tipicamente, aplicações ES não só dentro mas também fora da área afins estão sendo *shell*. Indiferente das complexidades, quase todos os *shells* contêm ferramentas para codificar o conhecimento, engenharia de inferência para ponderar estes conhecimentos, explicação facilitada para traçar a linha do motivo do sistema e interface para manipular o diálogo entre o construtor e o sistema e entre o usuário e o sistema.

A maioria dos *shells* conhecidos são codificados em uma linguagem que é muito similar para uma programação de linguagem básica. Assim a interface entre o construtor e o sistema é muito desejável em ambiente de programação com editores especializados, corretivos, etc. O processo de transferência e transformação da solução do problema especializado tem sido visto como um gargalo crítico no desenvolvimento do ES. Este gargalo, tem sido discutido em detalhes por outros autores BUCHANAN (1983), JUNGDUCK e COURTNEY (1987) e KING (1987).

As interfaces estão concentradas entre o usuário e a aplicação criada pelo o especialista e o engenheiro do conhecimento. O que é crucial para o ES na abordagem é a tomada explícita que a maioria dos elementos importantes do sistema tem na grande armazenagem do domínio do conhecimento específico, tal conhecimento, é a chave para locação de variáveis importantes e interpretação de resultados.

1.11.3 Interface com o usuário

A maioria das aplicações ES na área aparecem para os usuários em uma das duas formas. A primeira, são aqueles sistemas que apresentam ao usuário uma forma de questionário ou negócio, onde um grande número de itens de dados são requeridos do usuário no momento.

Mais freqüente, são os ES da área onde o usuário é confrontado por uma máquina que o entrevista ou tem uma sessão de perguntas e respostas. Baseado na resposta do usuário, a máquina decide qual conclusão deve tirar e o qual será a próxima questão. No final, o usuário é apresentado com um relatório escrito de questões e conclusões, também com um log de arquivo das perguntas e respostas e todas anotações feitas.

A segunda, em vez de tipo de resposta, o usuário simplesmente aplica a importância do controle do cursor (isto é, acima, baixo, esquerda, direita), para fazer uma seleção entre as escolhas fixadas. Da mesma forma, no final da sessão, o usuário pode questionar como uma conclusão em particular foi inferida e o que acontecerá com as conclusões se um dos valores de dados forem modificados (*What- If*).

Independente se a interação captura lugares por intermédio de uma forma de negócio ou sessão de perguntas e respostas ou se, ainda o sistema tem estabelecidas características presentes ou ausentes, de uma visão estereótipo, o modelo máquina homem numa ES, são baseadas essencialmente na mesma visão. Em quase todas elas tem um Shell que é usado para criar um padrão de máquina especialista que freqüentemente diagnosticam, soluções de problemas e/ou recomendações dentro de um domínio positivamente restrito. Enquanto o usuário inicia a sessão, o diálogo é quase totalmente com a máquina, ficando o usuário com o

papel reduzido para responder questões quando, e somente quando, o sistema requerer dados de entrada para tirar conclusões.

É claro, que no final da análise o usuário tem as opções de ignorar e modificar as conclusões produzidas pela máquina. ROTH (1987) tem chamado este modelo de interação o **modelo prótese**. Aqui, a máquina é vista como um substituto ou remédio para uma

1.11.4 Alguns problemas fundamentais

Aplicações de ES são núcleos conhecidos. No desenvolvimento de ES, o enfoque está na conquista da razão do motivo do especialista, enquanto o usuário não especialista, está submetido implicitamente ou explicitamente a um papel de importância secundária. No ES encontramos uma variedade de problemas:

- ✓ **Linguagem:** as linguagens de entrada e saída das aplicações estão às vezes repousada em termos e conceitos onde o entendimento é somente de um profissional ou especialista de aplicação.
- ✓ **Controle:** mesmo se as instruções sejam muito elaboradas ou computadas em uma trajetória, elas são limites para quebra se seguido dentro de uma forma de repetição. Em qualquer sistema temos dificuldades de construir um mecanismo que enfrente as novas situações e que adaptem as condições especiais ou contextos, ou que reconheçam erros humanos ou, ainda, as correções (*bug*) de instruções.
- ✓ **Condição do Sistema:** a fim de utilizar uma ferramenta, o usuário deve conhecer fronteiras e limites, capacidades, os lados das ações, pré condições, pós

condições, etc. Diálogo entre perguntas e respostas fornecem ao usuário a

- ✓ ***Superfície de comunicação:*** quando as perguntas e respostas formam a base da comunicação, o sistema de conhecimento não está detalhado. Esta superfície, a máquina tem muito pouco conhecimento ou nenhum caminho. Qualquer determinação na intenção do usuário, metas ou planos podem ser o núcleo da tarefa ou problema para serem resolvidos.
- ✓ ***Capacidade do Usuário:*** desenvolvedores de sistema de softwares (processadores de textos, banco de dados, etc.), tem aceitado que os limites dos usuários são amplos e em muitos casos os usuários são capazes de apreenderem e adaptarem-se ao sistema. Em ES, se o aprender é considerado tudo, o enfoque está na habilidade do sistema de expandir e modificar este conhecimento do domínio e não na capacidade do usuário de melhorar à sua performance.
- ✓ ***Metas do Sistema:*** no ES, a ênfase está no arquivamento do conhecimento e no esclarecimento da definição da solução. Os usuários são questionados por liderança de conceitos (para fornecer informação contextual, direcionado para a importância do tópico, para auxiliar no prognóstico dos resultados e criticar as soluções oferecidas pelos outros). Dentro de um resumo, eles são requeridos para agir como consultores.
- ✓ ***Especialista:*** quando os usuários de ES consideram a si mesmo.

Baseados nestes argumentos, o enfoque de ES está também na narração, a linguagem que eles falam não é a linguagem do executivo e eles fornecem aos usuários poucos caminhos para explorarem alternativas de saídas.

1.11.5 Frontends *inteligentes*

Muitos dos problemas e tarefas do lado dos administradores tem sido rotineiramente endereçados pelos planejadores e analistas como os convencionais DBMS, DSS, pacotes de estatística analítica e camadas de extensão. Diferentemente dos ES *shells*, estes sistemas analíticos possuem um conjunto de procedimentos computacionais, facilidades de relatórios e rotinas para investigar cenários alternativos. Em comparação do esquema de representação do ES, as várias capacidades destes sistemas são melhores entendidas por um grande número de usuários. O que estes sistemas não tem, portanto, é a habilidade de capturar o conhecimento simbólico e a razão deste conhecimento.

Muitas experiências foram feitas para adicionar ou pelo menos interagir capacidades com estas ferramentas convencionais. Vários esforços tem sido feito para combinar a razão do conhecimento básico com DBMS (KERSHBERG, 1985). Dentro da mesma visão, alguns produtos já estão combinando as características de um ES shell entre um banco de dados ou sistema de arquivos.

O protótipo deverá ser modificado dentro da linguagem de modelagem do IFPS e ser compilado, assim que os usuários criarem bases de conhecimento dentro de uma forma de equação (como as regras IF/THEN). Levando-se em consideração que os dados corporativos residem em banco de dados, inclusive os modelos de planejamento e estatísticas. Devendo-se determinar automaticamente os procedimentos que deverão ser usados, como também, as análises de performance, a interpretação dos resultados e como fazer recomendações.

O uso do conhecimento é um pré-requisito para construir um assistente (ou consultor) de sistemas. Os Frontends Inteligentes (IFE), também conhecido como **pequenos**

sistemas, possuem uma construção representativa de suas próprias funcionalidades, isto é, os sistemas detêm o conhecimento de como resolver determinados problemas e como interpretá-los. A intenção de um IFE é habilitar o usuário a fazer uma utilização mais efetiva dos pacotes de software convencionais através dos frontends.

Segundo CLEAL e HEATON (1988), o perfil da capacidade básica necessária de um IFE para realizar estas tarefas são:

- ✓ cuidado com o diálogo com o usuário;
- ✓ produzir uma especificação dos problemas do usuário;
- ✓ usar a especificação como instruções gerais, para executar o realce do software; e
- ✓ interpretar os resultados produzidos pela análise.

No mundo das decisões, a maioria dos trabalhos de interface de inteligência estão centrados em área de análise estatística. Assim, como, outras questões de base do conhecimento várias técnicas de representações podem ser usadas para criar interfaces inteligentes entre usuários e os sistemas convencionais, tais como:

- ✓ **Regras:** o analista ou planejador de conhecimento é representado por um grande número de leis específicas da situação. A consequência das regras representa os comandos do sistema e procedimentos para realizar análises específicas, enquanto antecedem a representação dos pré-requisitos para cuidar das saídas
- ✓ **Estruturas:** consiste em aparentes campos no banco de dados, salvo aqueles que são muito ricos na estrutura de dados; aspecto de várias facetas. Alguns desses aspectos guardam valores dos dados; outros referenciam procedimentos que são ativados quando determinadas ações são trazidas para os dados; e
- ✓ **Rede de procedimento:** esta é uma coleção de esquemas de recursos usados para representarem planos para cuidar das saídas das tarefas. Muito desses esquemas

têm a sua origem em baseada na linguagem natural. As questões e esquemas de representações são usadas como partes do planejamento de tarefas, que são às vezes ilustradas como o mundo da robótica.

1.11.6 Arquitetura de um sistema amigável

A população de usuários permanecerão limitados, até que os computadores possam processarem linguagem natural. O maior obstáculo no ensino do uso dos computadores, é que os usuários tem que aprender a falar a linguagem do computador e não o inverso. Além disso, o desenvolvimento de sistemas avançados de base de conhecimento, tal como consultor amigável, depende criticamente da compreensão da linguagem natural.

Segundo SOWA (1984):

Para muitas pessoas, linguagem natural é o significativo primário de pensar, aprender e comunicar. Não significa ser geral ou flexível. Cardápios (menus), são bons para selecionar opções, mas eles são desconfortáveis para expressar relação. Equações matemáticas, são bons para relações, mas elas não podem expressar comandos. Linguagem de programação comanda questões, mas elas não podem interroga-las. Linguagem de consulta interrogam questões mas elas não podem dar explicações. Somente linguagem natural, pode oferecer todas estas funções de comunicação humana dentro de uma estrutura de trabalho flexível e comum.

Os conceitos são expressos dentro de vários caminhos, assim como:

- ✓ mundo do conhecimento é a chave do entendimento (SCHWARTZ, 1988);
- ✓ robusto entendimento da linguagem natural não pode ocorrer dentro da ausência de uma base de conhecimento concreta de uma área relevante (SCHAWARTZ, 1988);
- ✓ A comunicação da linguagem natural está na atividade total do intelecto (HENDRIX, 1985); e

- ✓ (Mesmo uma simples base de dados), o conhecimento em vez dos dados, é necessário interpretar e responder diversas consultas na linguagem natural (KAY, 1984).

Geralmente, um usuário entra num diálogo com um computador com propostas específicas na mente, tais como:

- 1) receber um assistente na solução de um problema;
- 2) aprender se a máquina tem acesso a um determinado ponto em particular da informação, tendo a máquina que executa-lo;
- 3) a máquina deverá executar uma ação específica; e
- 4) a máquina tem que explicar estas ações ou partes dessas informações.

Se o argumento é uma falsa conclusão, então o computador deverá ser capaz de admitir. Se um requisito ou argumento demonstrar que o usuário é confuso ou trabalha com preposições incorretas, então o sistema provavelmente necessitará de retornar e reexplicar a si mesmo ou no mínimo, corrigir o usuário. Mesmo quando o argumento é um simples requisito para informação o computador ainda necessitará de atender todas estas nuances.

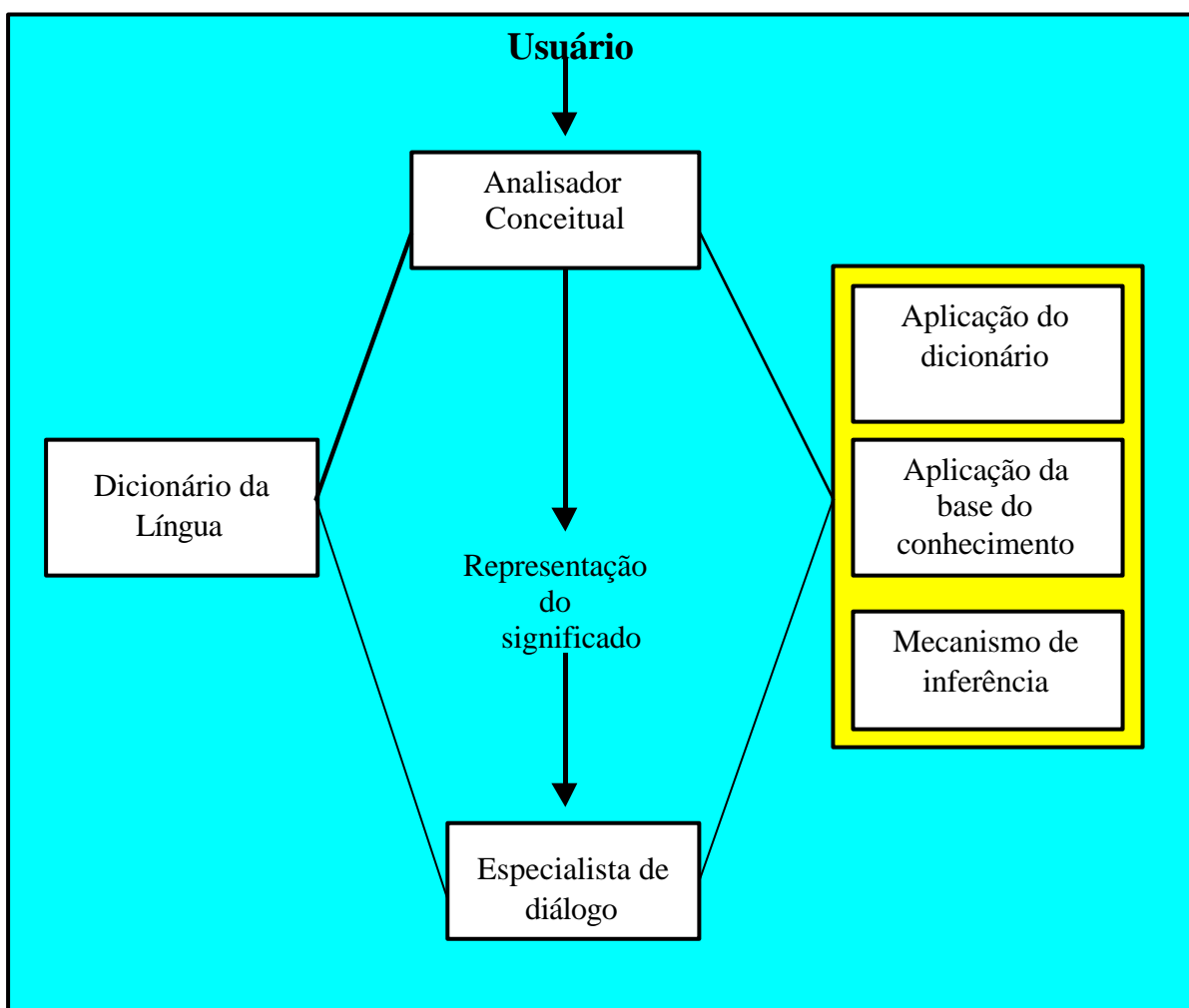
Recentemente sistemas cognitivos tem se dedicado na criação de um sistema de consultoria amigável (CAS), distante deles vem os desenvolvedores de ES.

O CAS é capaz de manipular, não somente perguntas de banco de dados padrão, mas também, questões sobre recomendações, explicações e mudanças na percepção do usuário.

A Figura 7, indica o trabalho do analisador de conceitos, produzindo um conhecimento interno, representando o argumento do usuário. Tal representação é chamada de

(SCHANK, 1981). Em uma representação é produzido a tarefa de manipular o diálogo (também conhecido como uma meta de monitoração), para traçar a conversação e produzir respostas. A arquitetura confirma a suposição de que um âmbito

extenso do conhecimento, incluindo o domínio, são requisitos para orientar a forma da



1.11.6.1

1.11.6.2 FIGURA 7 – Arquitetura de um sistema de consultoria amigável

1.11.6.3 (SCHWARTZ, 1988)

2.5.6 – Agentes Inteligentes

O agente age como o representante do usuário, gerando os eventos importantes para a atenção e performance de ações no interesse do mesmo. Nesses termos de funcionalidade, agentes partem de uma importante característica como os *frontends* inteligentes.

Eles fornecem um significado de captura do conhecimento do uso para um assistente ou para um consultor. A diferença é, em vez de comandos, é dada uma linguagem para o planejador ou analista (assistente) embutir o seu próprio conhecimento.

Teoricamente, agentes inteligentes podem servir uma multidão de objetos dentro de um EIS. O papel de monitorar, portanto, aparece mais promissor, independente se um executivo é visto como um trabalhador de problema ou um fazedor de decisão.

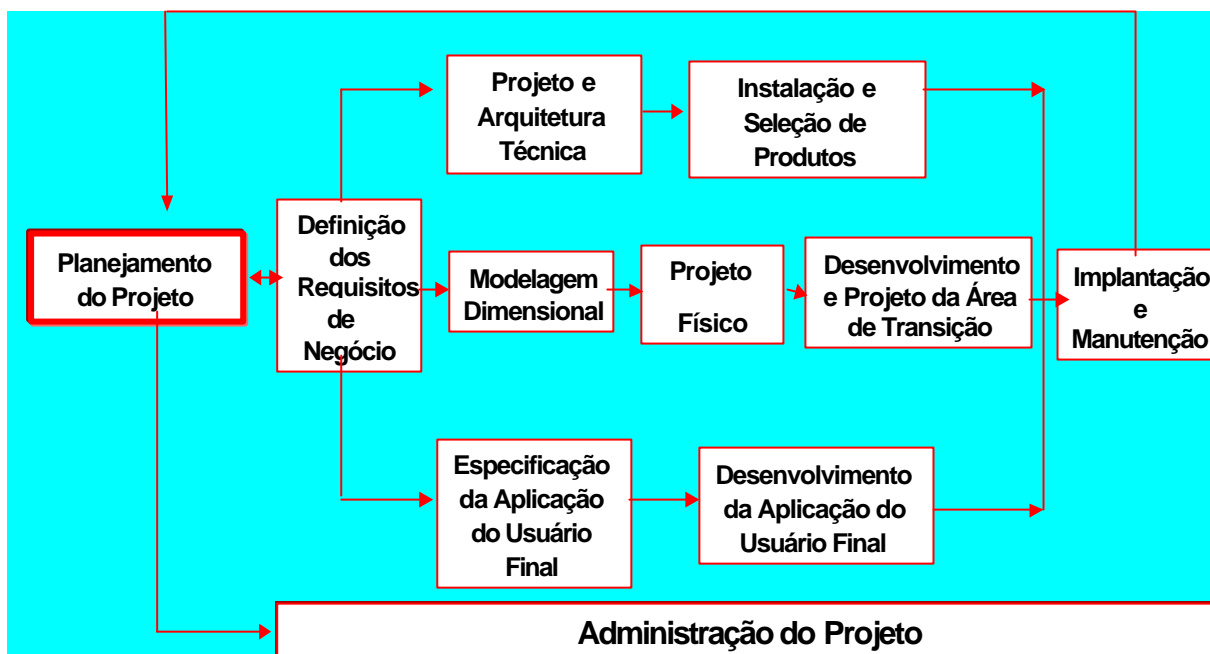
A interface WIMP (*windows, icons, mouse e pointer's*), ao redor do protótipo, está construindo um fácil aprendizado e encorajando o desenvolvimento de aplicações que requerem uma ação mínima por parte do usuário.

Atualmente, são desenvolvidos modelos ou imitações de como navegar conhecimentos computadorizados, em alguns dos protótipos e produtos que trabalham com domínios especializados (BRAND, 1987).

1.12 Data Warehouse

O potencial de mineração de dados pode ser melhorado se os dados apropriados tiverem sido coletados e armazenados em um *Data Warehouse* (*RDMS – Relational Database Management System*). Um *Data Warehouse* é uma nova e poderosa tecnologia, tornando possível a extração de dados operacionais e superação de inconsistências entre formatos de dados legados (INMON, 1997).

Assim como é possível a integração de dados através da empresa independentemente da localização, formato ou requerimentos de comunicação, também é possível a incorporação de informações adicionais.



1.12.1.1 FIGURA 8 - Ciclo de Projeto do Data Warehouse (KIMBALL, 1998)

Em outras palavras um *Data Warehouse* fornece dados que já estão transformados e resumidos, tornando apropriado um ambiente para as aplicações DSS (sistemas de suporte à decisão) e EIS (sistemas de informações executivas) mais eficientes (KIMBALL, 1998a). A primeira etapa, importante no processo de mineração de dados é organizar grandes volumes de dados em alguma forma de categoria que facilite a busca, interpretação e

O elemento importante de um *Data Warehouse* é que os dados são armazenados em diferentes níveis de detalhamento, permitindo o acesso rápido aos mesmos. O *Data Warehouse* vai nos permitir que consultas de análise sejam respondidas, ao invés de **simples agregações**, tais como; “*qual é o período de plantio para uma determinada região?*”,

“qual a cultivar é mais plantada na América do Sul e como isto, está relacionado com as importações e exportações brasileiras?”.

Os objetivos principais da mineração de dados são as previsões e as descrições. As previsões fazem uso de variáveis existentes no banco de dados para prever valores desconhecidos ou futuros. As descrições são voltadas para as buscas de padrões descrevendo os dados e a subsequente apresentação para a interpretação do usuário. A relativa ênfase entre previsão e descrição varia de acordo com o sistema de mineração de dados utilizado (KIMBALL, 1998b).

Alguns algoritmos de mineração de dados abaixo podem ser utilizados para resolver problemas específicos (exemplos na área agrícola estão citados no capítulo 3):

- Associações;
- Classificação ou Geração de perfis;
- Padrões seqüenciais;
- Agrupamento;
- Redes neurais;
- Indução;
- Árvore de decisão;
- Visualização de dados;
- Aprendizagem induzida;
- Estatística.

PROPOSTA DE *DATA MINING* PARA O ZONEAMENTO E MONITORAMENTO AGRÍCOLA

1.13 Redução dos riscos climáticos (resultados obtidos)

O projeto de zoneamento agrícola no Brasil objetiva o *“desenvolvimento de estudos de regionalização dos sinistros climáticos no Brasil, visando minimizar as perdas na produção agrícola, disponibilizando ao produtor rural técnicas que permitiriam fugir de riscos climáticos oriundos do regime de chuva”* (ROSSETTI, 1999).

Esta meta estava de acordo com a proposta desenvolvimentista do Governo, pois visa a tecnificação mínima da agricultura e a adoção de qualidade na atividade produtiva, garantindo sua sustentabilidade.

Os órgãos inicialmente afinados ao projeto foram: a FINATEC/UnB, a EMBRAPA, através do CPAC e posteriormente de todos os outros Centros de Pesquisas espalhados pelo território nacional, o INMET e o DNAEE (atual ANEEL). Logo a seguir, juntaram-se aos primeiros o IAPAR/PR, a EPAGRI/SC, a UNICAMP/SP e o IAC/SP.

Para sua operacionalização, registre-se que o DNAEE (Departamento Nacional de Água e Energia Elétrica do Ministério do Meio Ambiente e Recursos Hídricos) é o gestor de um excelente banco de dados pluviométricos, que foi, como está sendo, utilizado para determinar as análises climáticas em base estatística. O INMET (Instituto Nacional de

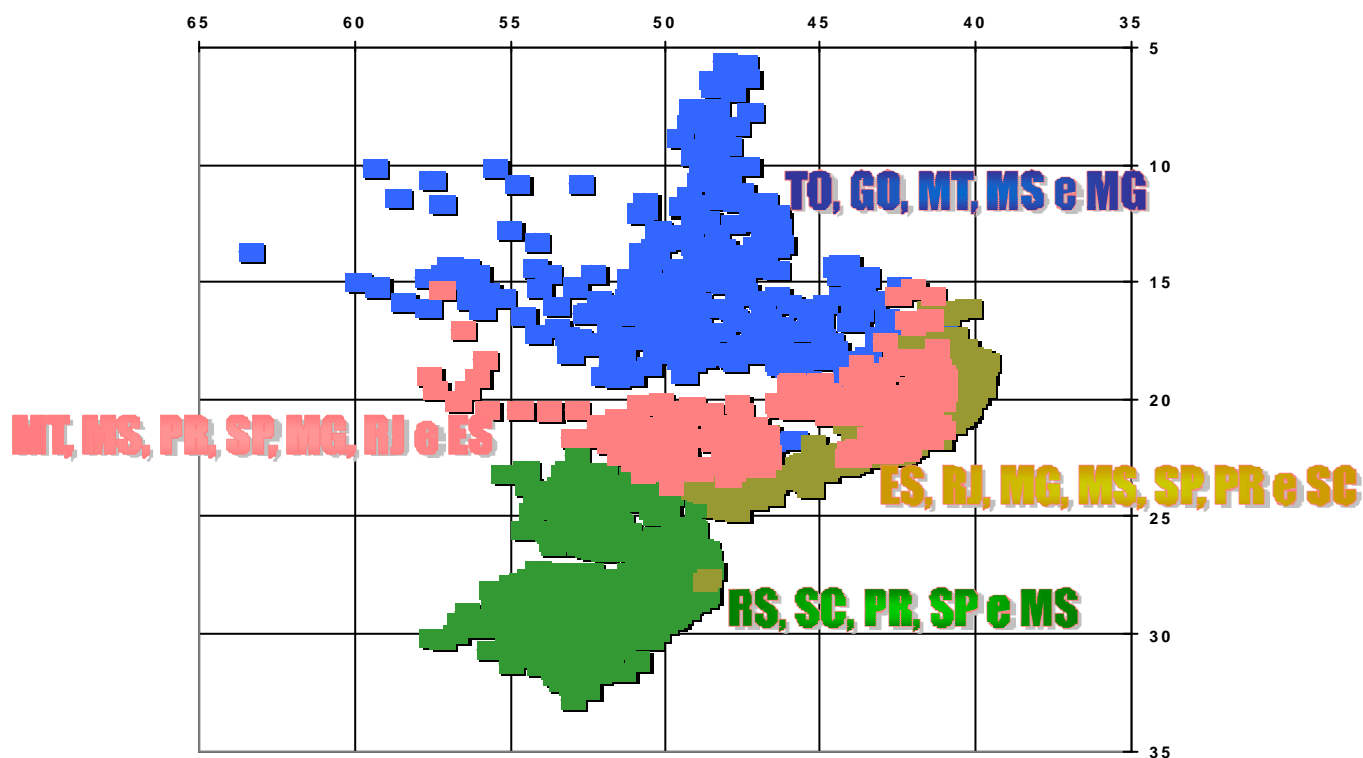
odos de cinco dias (pêntadas). Esse fato não produz resultados distintos dos que seriam obtidos com dados diários, conforme a metodologia apresentada por KELLER (1999). Já os resultados referentes a relações Probabilísticas baseiam-se em modelos obtidos a partir da análise de dados considerados dia a dia.

1.13.1 Áreas homogêneas

Os resultados da aplicação da Análise de Agrupamento (*Cluster Analysis*) permitiram identificar a existência de quatro grandes regiões climatológicas, nas quais a precipitação pluviométrica possui regime probabilístico assemelhado. Dessas regiões, indicadas por **Área 1**, **Área 2**, **Área 3** e **Área 4**, as duas primeiras possuem regimes pluviométricos extremos, enquanto que as demais são regiões de regimes intermediários.

A Figura 9 apresenta a distribuição espacial dessas regiões, a partir da localização geográfica das estações meteorológicas que as compõem.

Figura 1 : Distribuição Espacial das Áreas Homogêneas

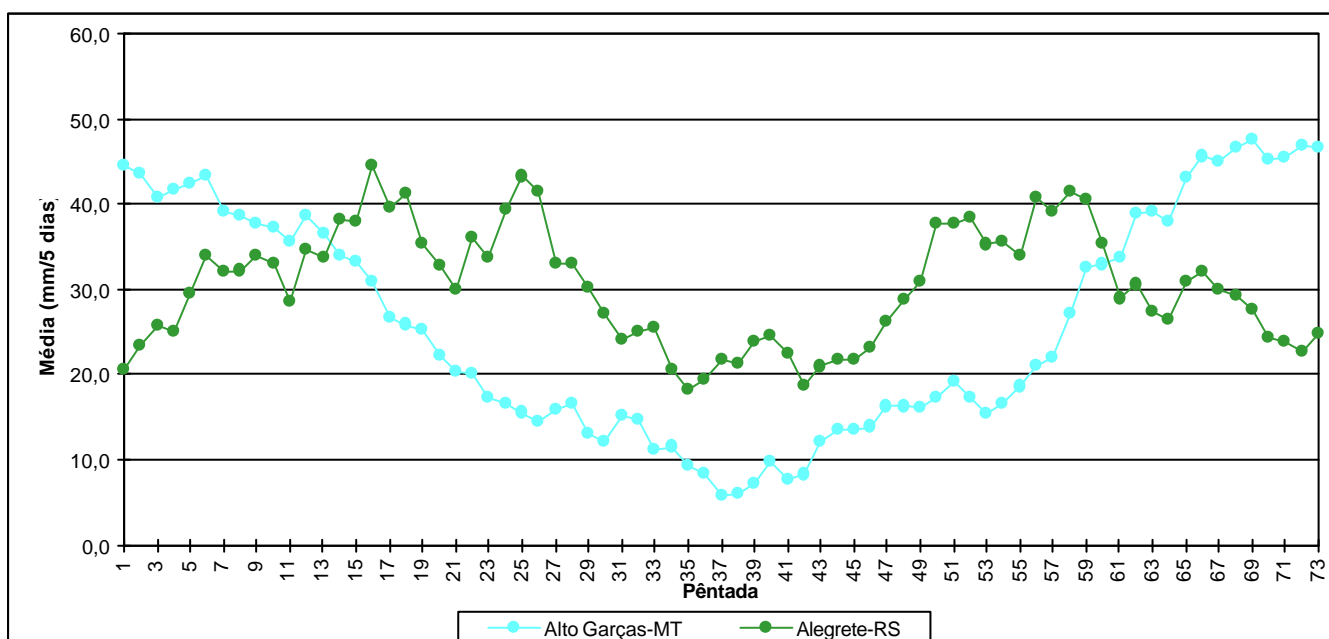


1.13.1.1 FIGURA 9 – Áreas homogêneas

Uma análise detalhada dos valores assumidos pelas variáveis classificatórias nas consideradas, permite constatar que o comportamento conjunto das mesmas

possui um elevado poder de discriminação, capaz de distinguir claramente os diferentes regimes pluviométricos a que estão submetidas essas regiões.

Como se pode observar no Gráfico da Figura 10, a precipitação média do Mato Grosso-MT diminui a partir do início do ano até atingir um valor mínimo na metade do ano, voltando então a crescer até o final do ano. Já a precipitação pluviométrica na estação do Rio Grande do Sul-RS possui comportamento inteiramente diferente, pois apresenta, alternadamente, dois períodos de máximo e um de mínimo durante o ano. Essas características são típicas de todas as regiões climatológicas nas quais estão situadas as estações meteorológicas.



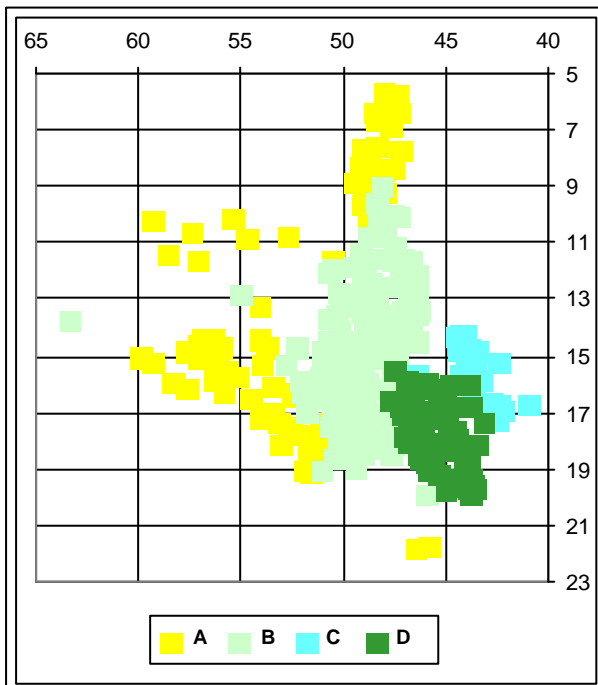
1.13.1.2 FIGURA 10 – Gráfico média da precipitação pluviométrica

Posteriormente, dentro de cada área, as estações foram novamente submetidas à upamento, resultando nas sub-famílias, designadas por:

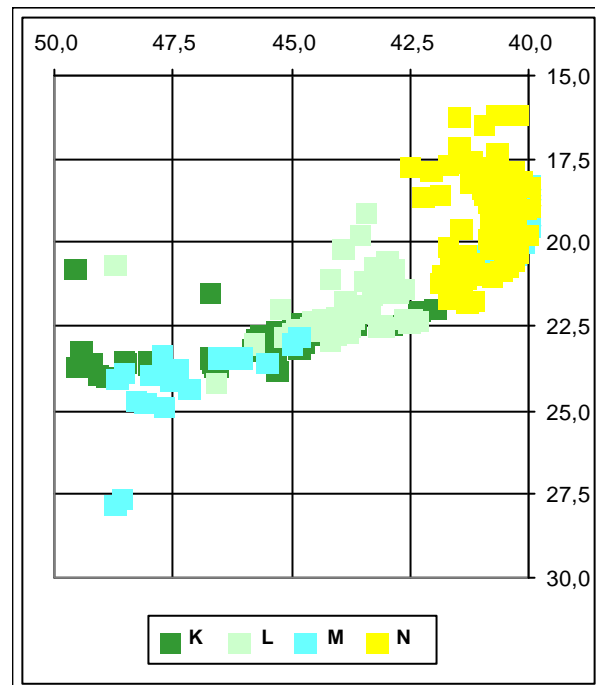
- ✓ **Zona A, Zona B, Zona C e Zona D** na Área 1;
- ✓ **Zona E, Zona F, Zona G, Zona H, Zona I e Zona J**, na Área 2;

- ✓ Zona K, Zona L, Zona M e Zona N, na Área 3; e
- ✓ Zona O, Zona P e Zona Q, na Área 4.

Área 1

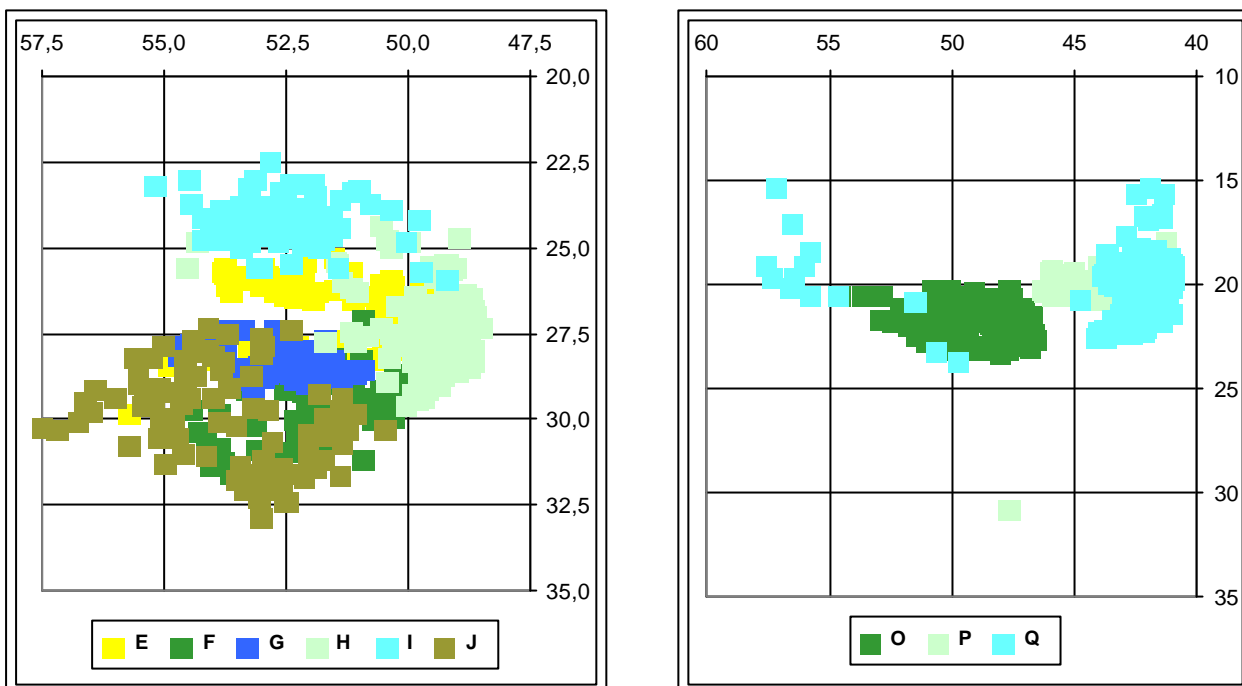


Área 3



1.13.1.3 Área 2

1.13.1.4 Área 4



1.13.1.5 FIGURA 11– Áreas

A Figura 11 mostra a distribuição espacial dessas zonas dentro de suas respectivas
 ização geográfica das estações que as compõem.

UF	Área 1				Área 2						Área 3				Área 4			Total
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
TO	24	31																55
GO	15	102		4														121
MT	39	5															1	45
MS	2								3						8		10	23
MG	2	2	21	63							1	21		20		59	62	251
ES													29	60				98
RJ											9	34		6			40	89
SP								1	1		58	9	17		85	1		172
PR					36			20	79						2		2	139
SC					20			106					2					128
RS					23	74	51	5		84								237
	82	140	21	67	79	74	51	132	83	84	68	64	48	86	95	60	124	1.358

1.13.1.6 FIGURA 12 – Tabela de distribuição das Estações

A Figura acima apresenta a distribuição do número de estações, por área/zonas da Federação.

1.13.2 Probabilidades de fatores climáticos

Adotou-se a família de distribuições Johnson S_B para a determinação das densidades de probabilidade da precipitação pluviométrica positiva. A Figura 13 mostra um gráfico com quatro das 1241 distribuições identificadas nesse estudo (73 pântadas vezes 17 zonas correspondentes às zonas homogêneas A e E, e às pântadas n.º 01 (de 01/Jan a 05/Jan) e n.º 37 (de 30/Jun a 04/Jul).

Nesse gráfico, pode-se observar o perfil característico das distribuições de probabilidades da precipitação pluviométrica, no qual se destacam a probabilidade de dias secos, a densidade de probabilidade da precipitação no intervalo $(0, \lambda)$ onde λ é um parâmetro fixado empiricamente, e a probabilidade de precipitações pluviométricas superiores a λ .

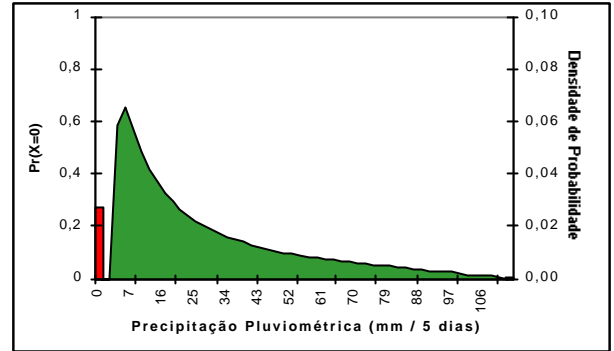
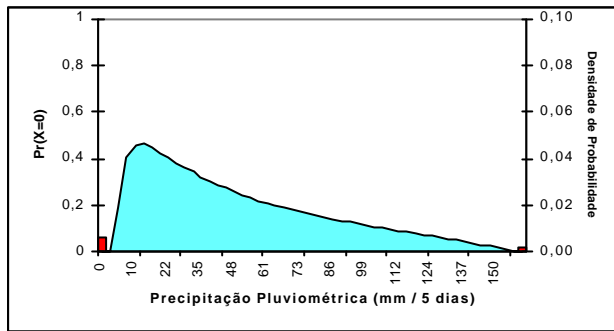
É de se destacar, também, que em todas as distribuições nas quais o limite λ foi fixado, a probabilidade de precipitação superior a λ não excedeu a 2, ficando a grande maioria em torno de 0,5.

Quanto ao grau de ajustamento dessas distribuições teóricas aos dados pluviométricos observados, na Figura 14 o gráfico exemplifica, com muita clareza, a grande aderência da família de Johnson S_B a situações e formas distintas.

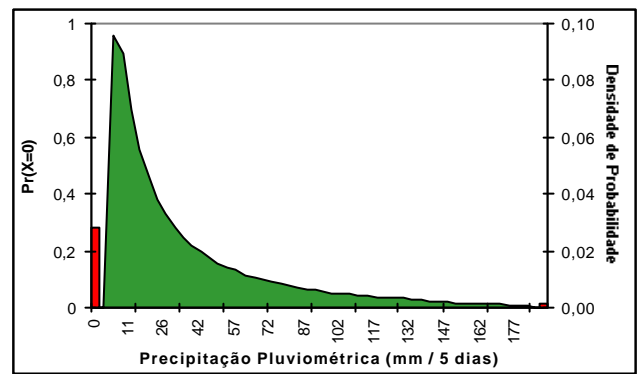
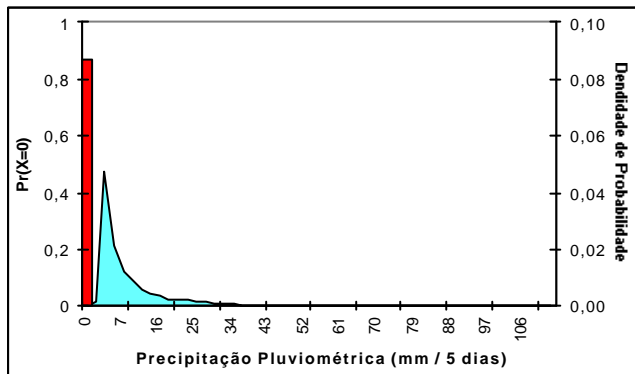
Área 1 Zona A

Área 2 Zona E

Pântada 01 – 1º a 5 de janeiro

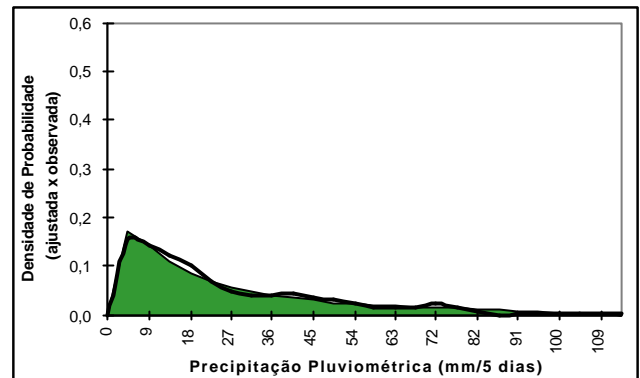
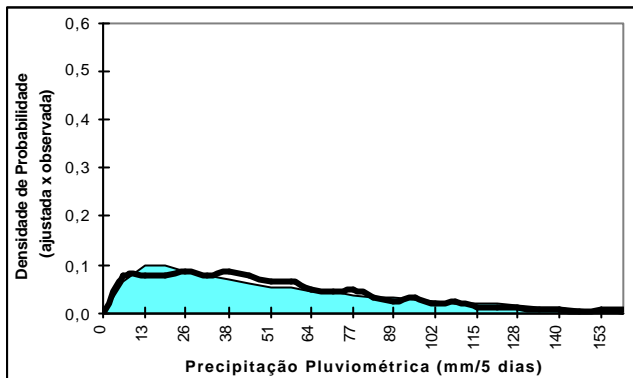


Pêntada 37 - 30 de junho a 4 de julho

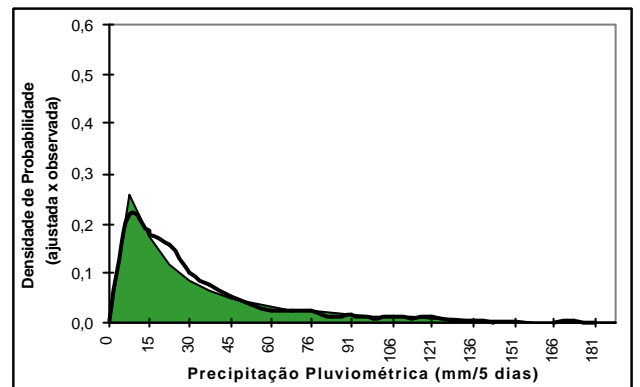
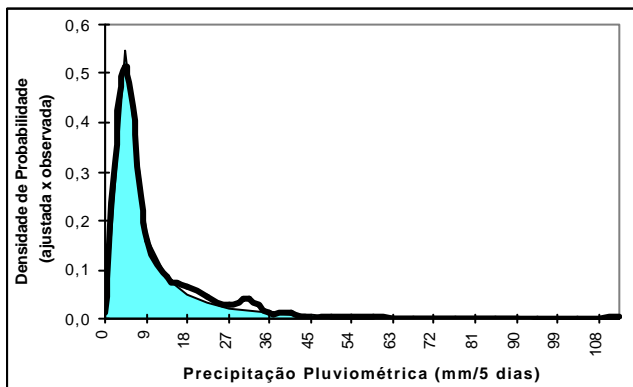


1.13.2.1 FIGURA 13 – Gráfico da Zona A e E

Pêntada 01 - 1º a 5 de janeiro



Pêntada 37 - 30 de junho a 4 de julho



1.13.2.2 FIGURA 14 – Gráfico da Zona A e E (ajustado)

1.13.3 Simulações probabilísticas

Efetua-se a seguir uma simulação da precipitação pluviométrica diária aplicada a um modelo de avaliação do balanço hídrico que utiliza a função de EAGLEMANN (1971), com o objetivo de determinar, mediante a observação do comportamento aleatório do Índice de Satisfação de Necessidade de Água (ISNA), épocas adequadas para o plantio de determinadas culturas, de forma a minimizar os riscos de escassez de chuva em fases críticas de seu desenvolvimento.

rio é baseado no segundo decil dos valores simulados, ou seja, um valor a partir do qual 80% dos demais valores se encontrem acima dele.

A diferença na aplicação das diversas alternativas possíveis está no maior ou menor rigor que se deseja impor ao critério, visando garantir a satisfação das necessidades de água íca. No caso estudado, a *variável critério* definida pelo segundo decil é mais rigorosa do que a definida pela média.

Em ambas as alternativas, considerou-se como comportamento favorável dos valores do índice ISNA ao desenvolvimento da cultura em estudo, a ocorrência do evento:

ser superior a um valor pré-fixado. Esse valor, que aqui denominou-se *ponto de*

decisão, deve ser estabelecido a partir de considerações técnico-científicas associadas à cultura.

O processo de simulação inicia-se pela geração aleatória da precipitação pluviométrica para cada dia do período de interesse, obtendo-se, a partir do modelo de balanço hídrico adotado, os correspondentes valores diários do ISNA.

A seguir, compute-se o valor da *variável critério* e registra-se a ocorrência ou a não-ocorrência do evento favorável ao desenvolvimento da cultura.

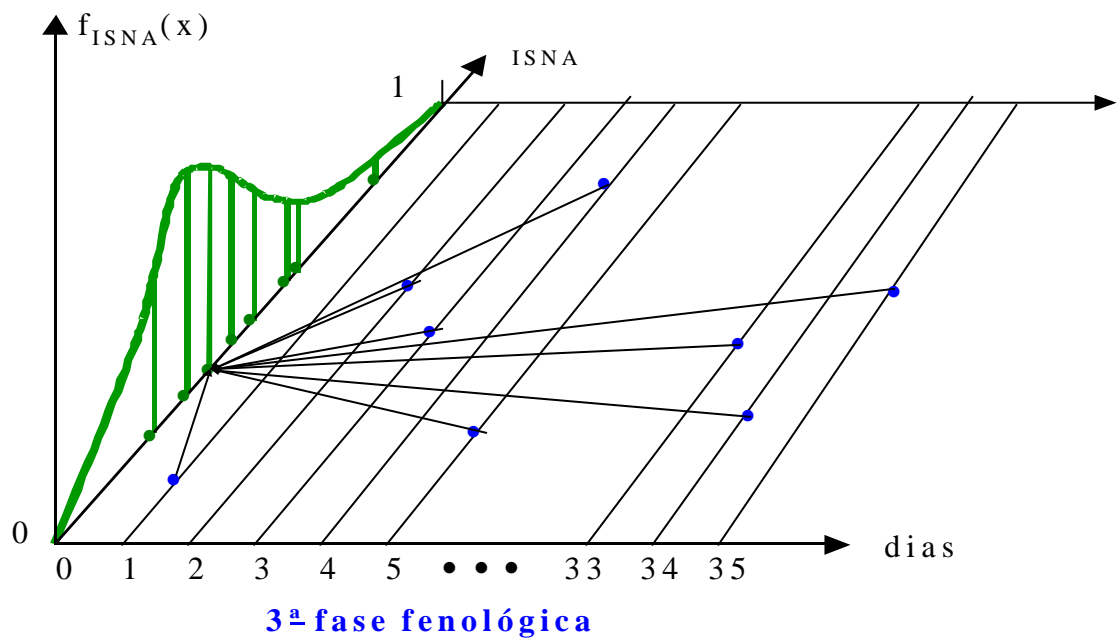
Esse procedimento é repetido tantas vezes quanto se queira, obtendo-se, por fim, dados suficientes para a identificação da distribuição de probabilidade da bem como da probabilidade de ocorrência do evento favorável ao desenvolvimento da cultura, na região determinada e a partir da data de plantio fixada.

Com base nesses resultados, podemos identificar se a lavoura considerada possui boas chances de sucesso, verificando se a probabilidade de ocorrência de condições favoráveis quanto a oferta hídrica no período em estudo é suficientemente grande, como por exemplo, superior a 90%.

Finalmente, variando-se a data de plantio ao longo do ano, pode-se identificar as datas que são favoráveis ao sucesso da lavoura pretendida.

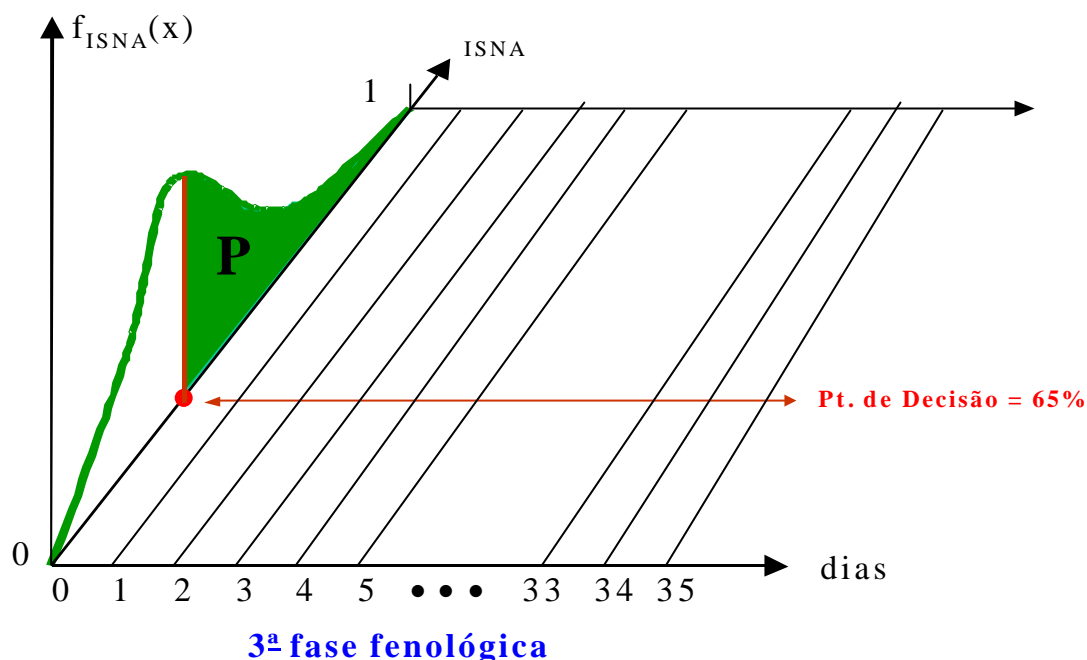
A Figura 15 mostra os valores do ISNA em uma seqüência simulada de 35 dias, - fase fenológica, bem como o valor obtido, nessa seqüência, para a variável critério.

- fase fenológica do desenvolvimento da cultura considerada.



1.13.3.2 FIGURA 16 – Gráfico do ISNA (comportamento)

Na Figura 17 o gráfico, a área sombreada, **P**, representa a probabilidade da *variável critério* assumir um valor superior a 65%, ou seja, representa a probabilidade de ocorrência de envolvimento da cultura na data de plantio fixada.



1.13.3.3 FIGURA 17 – Gráfico do ISNA (probabilidade)

Outras aplicações da simulação poderão ser efetuadas, com o objetivo de estabelecer, para uma cultura em determinada região, as datas de plantio que minimizem os efeitos negativos de outros eventos climáticos dependentes do regime pluviométrico.

O Grupo de Estatística, composto atualmente pelos Professores Thadeu Keller, Paulo Schubnell e Juarez Figueiredo estão propondo à reformulação e o aperfeiçoamento do Sistema de Informações Gerenciais do PROAGRO (isto é, informatização dentro desta proposta de *Data Mining* que está sendo sugerida no capítulo 3), tendo em vista torná-lo apto à coleta, crítica sistemática e processamento de dados ao longo das safras, nos moldes exigidos pela metodologia a ser desenvolvida para análise estatística.

1.14 Processo de *Data Mining*

1.14.1 Algoritmos de mineração

Existem vários algoritmos de mineração de dados utilizados para resolver problemas específicos são categorizados em:

a) Associações

A premissa básica de algoritmos de associação é achar todas as associações em que a presença de um conjunto de itens em uma transação implica em outros itens (KIMBALL, 1988a). Algoritmos de associação têm numerosas aplicações, incluindo planejamento de armazenagem, mala direta para marketing direcionado e planejamento de política agrícola. Por exemplo, a regra de associação deriva a partir da mineração de dados de um banco de dados de transações. A regra de associação poderia ser:

✓ *63% dos produtores que plantam Milho também plantam Feijão.*

O algoritmo produz uma grande quantidade destas regras e cabe ao usuário selecionar o subconjunto de regras que têm graus de confiança maiores e também porcentagem de listas que seguem a esta regra. Podem existir também múltiplas associações tais como:

✓ *52% dos produtores que plantam Milho e Feijão também plantam Melancia.*

É importante para o usuário determinar quando existe algum elemento com chance de correlação (Milho e Feijão estão sendo plantadas) ou quando existe alguma correlação desconhecida mas importante (Melancia também está sendo plantada).

b) Classificação ou Geração de perfis

Algoritmos de classificação ou geração de perfis desenvolvem perfis de diferentes grupos. Para um dado conjunto de registros com seus atributos correspondentes, um conjunto de marcadores (representando registros), e um marcador para cada registro, uma

função de classificação examina o conjunto de registros marcados e produz descrições das características dos registros para cada classe (KIMBALL, 1998a). A regra de classificação poderia ser (esta regra poderia então ser utilizada para a classificação de novos conjuntos de dados):

- ✓ *Agricultores com histórico com uso de crédito agrícola utilizando proagro (programa de seguro agrícola) tem uma taxa de juros de menos de 3%.*

c) Padrões Seqüenciais

Algoritmos de padrões seqüenciais identificam tipos de padrões seqüenciais em restrições mínimas especificadas pelo usuário (KIMBALL, 1988a). Esta técnica procura por eventos que ocorrem em uma seqüência através do tempo.

- ✓ *40% dos importadores de Trigo na safra 2000/2000, que compraram Farinha de Trigo seguidos de Fermento.*

d) Agrupamento

A técnica de agrupamento irá segmentar um banco de dados em subconjuntos ou grupos. Isto pode ser criado estatisticamente ou através da utilização de métodos de indução não supervisionados neurais ou simbólicos (KIMBALL, 1988a). Os vários métodos neurais e simbólicos são distinguidos pelo tipo de valores de atributos que podem ser aceitos (numérico, nominal e objetos estruturados); representação de grupos, e organização de grupos (em hierarquia ou em um nível plano). Esta técnica irá dividir produtores de acordo com os seus padrões, através da criação de conjuntos de grupos que contêm a máxima similaridade com estes e a máxima diferença entre eles. Alguns dos usos pertinentes têm sido na análise de bancos de dados de patentes, pois é de comum acordo que existe um consenso na utilização de palavras-chave; análise de texto para conceitos; compreensão de tipos de produtores em pesquisas de consumidor; e procura por artigos de pesquisa relevantes.

Qualquer algoritmo de mineração de dados é composto de três componentes: representação através de modelo, avaliação do modelo e método de busca. Resumidamente, o modelo deve representar limites flexíveis e suposições adequadas, de uma maneira que os padrões possam ser descobertos. As ferramentas de mineração ou mecanismos de busca são usualmente programas ou agentes automatizados inteligentes, incorporando alguma forma de inteligência artificial em bancos de dados relacionais. Os agentes detectam padrões predefinidos e alertam o usuário sobre variações. Vários tipos de ferramentas são utilizadas na mineração de dados, tais como:

d.1) Redes Neurais

Redes neurais são uma solução computacional que envolve o desenvolvimento de estruturas matemáticas com a habilidade de aprendizagem. As redes neurais têm uma notável habilidade de derivar médias de dados complicados ou imprecisos e podem ser utilizadas para extrair padrões e detectar tendências que são muito complexas para serem percebidas tanto por humanos quanto por outras técnicas computacionais (KIMBALL, 1988a). Uma rede neural treinada pode ser considerada como um *expert* na categoria de informação que foi previamente para a análise. Este expert pode então ser utilizado para fornecer projeções, dadas novas situações de interesse. Existem um grande número de problemas agrícolas reais em que as redes neurais podem ser aplicadas - muitas inclusive já aplicadas com sucesso nas indústrias alimentícias. Como as redes neurais são melhores na identificação de padrões ou tendências em dados, são próprias para a previsão e prognóstico incluindo:

- ✓ previsão de comercialização e armazenagem;
- ✓ controle de processo industrial;
- ✓ pesquisa de consumidores;
- ✓ validação de dados; e

- ✓ gerenciamento de risco;

As redes neurais utilizam um conjunto de elementos de processamento (ou nós). Estes elementos de processamento são interconectados em uma rede que pode identificar padrões nos dados uma vez expostos aos mesmos, ou seja, a rede aprende através da experiência, tais como as pessoas. Esta característica distingue redes neurais de tradicionais programas computacionais, que simplesmente seguem instruções em uma ordem seqüencial fixa.

d.2) Indução

Um banco de dados é um armazém de informações, mas o mais importante é a informação que pode ser inferida deste (HARRISON, 1998). Uma das principais técnicas de inferência disponíveis é a dedução. Dedução é uma técnica de inferência de informação que é uma consequência lógica da informação no banco de dados, tal como o operador *join* aplicado em duas tabelas relacionais onde o primeiro diz respeito aos agricultores e cultivar e o segundo, cultivar e governo federal (liberação de crédito agrícola ou preço mínimo) - infere em um relação entre agricultores e governo federal.

d.3) Árvore de decisão

Árvores de decisões são simples representações de conhecimento e classificam exemplos em um número finito de classes (HARRISON, 1998). Os nós são rotulados com nomes de atributos, os arcos são rotulados com possíveis valores para este atributo e as folhas são rotuladas com diferentes classes. Objetos são classificados através de um caminho percorrendo a árvore - seguindo os arcos que contêm valores que correspondem a atributos no objeto.

Um agente financeiro de crédito agrícola, por exemplo, pode ter registros de agricultores contendo descrições ou atributos. Com histórico de crédito conhecido, os registros pode ser rotulados/classificados como bons, médios ou ruins. Uma técnica de

indução pode produzir um modelo de classificação simbólica que gera uma regra estabelecendo “*se um agricultor tem uma produção com lucro em uma determinada safra \geq R\$ 20.000,00 e tem entre 40-55 anos de idade, e mora em um determinado município da região Sudeste ou Sul do país, então o agricultor tem um bom risco de crédito*”.

Muitas aplicações de mineração de dados utilizam o agrupamento de acordo com a similaridade para segmentar uma base de agricultores/cultivares. O agrupamento de acordo com a otimização de um determinado conjunto de funções é utilizado nas análises de dados, tal como na determinação de taxas de seguros agrícolas, os agricultores podem ser segmentados de acordo com um número de parâmetros e a segmentação de taxa de seguro agrícola ótima pode ser alcançada. O agrupamento/segmentação em bancos de dados são os processos de separar o conjunto de dados em componentes que refletem um padrão consistente de comportamento.

d.4) Visualização de dados

Muitas técnicas e ferramentas existentes têm sido utilizadas como ferramentas de mineração de dados. Estas técnicas e ferramentas fornecem soluções complementares para a mineração de dados. Por exemplo, o elemento inicial da mineração de dados envolve o armazenamento e a acessibilidade de elementos de dados de transações on-line para o uso de ferramentas mais sofisticadas. O *OLAP (on-line Analytic Processing)* envolve sistemas estatísticos tais como *SASTM* e *SPSSTM* para detectar padrões e tendências utilizando modelos estatísticos tais como modelos lineares e não-lineares. Estes sistemas resultam em análises mais diretas baseadas nos resultados da mineração de dados. Sistemas de geração de relatórios utilizam elementos de sistemas de suporte à decisão (*DSS*), sistemas de informações executivas (*executive information systems – EIS*) e fornecem suporte à tomada de decisões. Entretanto, com a mineração de dados a ênfase e o foco serão dados em um uso mais efetivo destes sistemas.

A utilização de planilhas multidimensionais e bancos de dados de visões resumidas dos dados através de múltiplas dimensões, embora populares, necessitam de formulação e revisão do usuário. Análise multidimensional é um método de visualização de dados agregados denominados medidas sobre um conjunto de dimensões tais como cultivar, armazenagem, período de plantio e região de plantio. Tecnologias de mineração de dados realizarão análises automáticas que poderão melhorar o valor da exploração dos dados, suportado por estas ferramentas multidimensionais.

d.4.1) Aprendizagem induzida

Indução é a inferência de informação através de dados e aprendizagem induzida é o processo de construção de modelos onde o ambiente, por exemplo - um banco de dados é analisado em uma visão para a procura de padrões (podendo ser aprendizagem supervisionada ou não supervisionada). Logo, indução pode ser entendido como a extração de padrões. A qualidade do modelo produzido pelos métodos de aprendizagem induzida é tal que o modelo pode ser utilizado para prever o desenvolvimento de situações futuras. O problema é que a maioria dos ambientes têm diferentes estados e conseqüentemente diferentes mudanças entre eles, de modo que não é possível sempre verificar um modelo através de todas as suas situações possíveis.

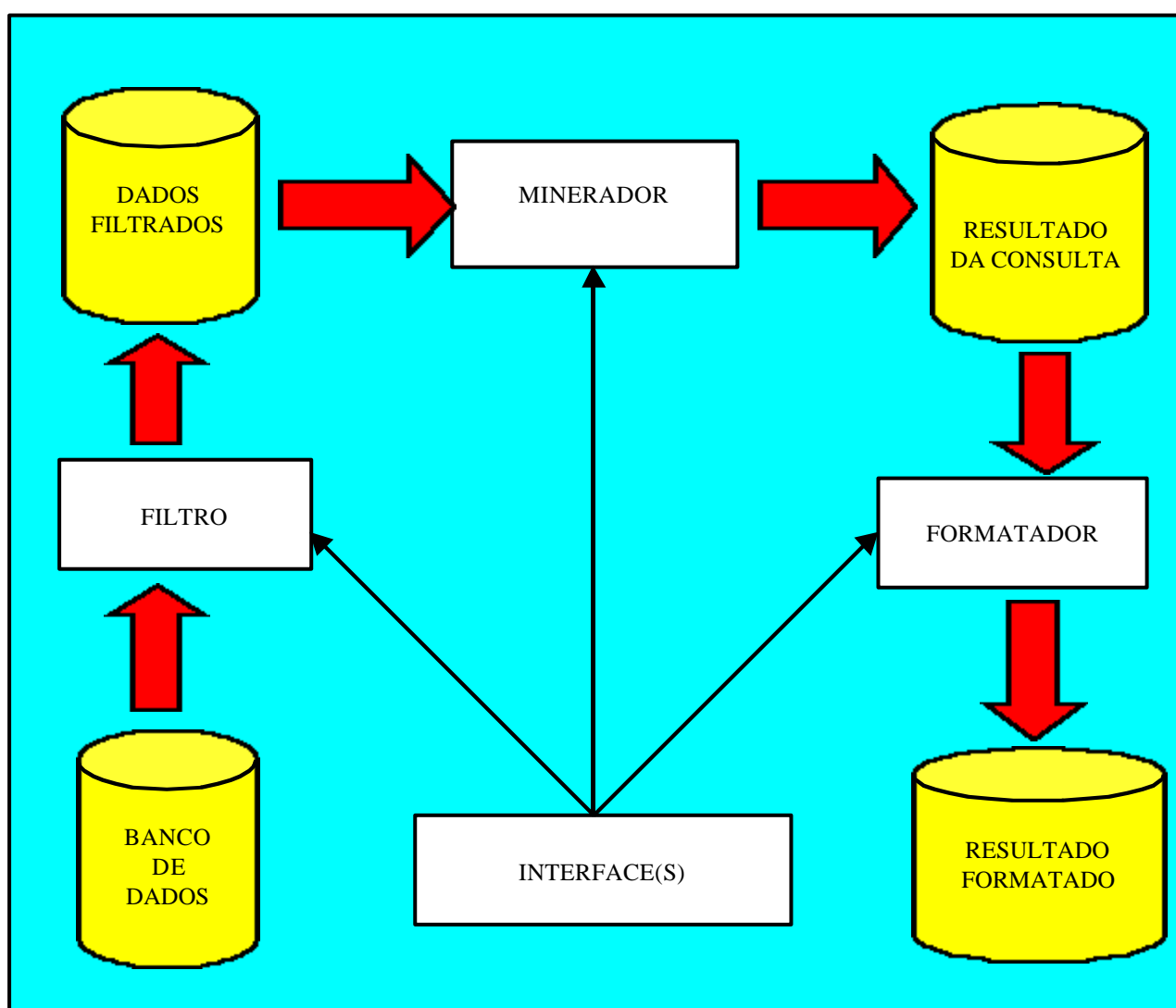
d.4.2) Estatística

A estatística tem uma sólida fundamentação teórica, mas os resultados da estatística podem ser grandes demais e difíceis de interpretar, pois necessitam do usuário para verificar onde e como analisar os dados (KIMBALL, 1988a). A mineração de dados, entretanto, permite que o conhecimento do especialista sobre os dados e técnicas de análise avançadas do computador trabalhem de maneira conjunta. Sistemas de análise estatística tais como SAS e SPSS têm sido utilizados por analistas para a detecção de padrões não comuns e explicações de padrões utilizando modelos estatísticos tais como modelos lineares. A análise estatística

tem um campo enorme de utilização e a mineração de dados não irá substituir tais análises, e sim utilizar análises mais diretas baseadas nos resultados da mineração de dados.

1.14.2 Passos do processo de Data Mining proposto

No Processo de *Data Mining* proposto podemos observar os diversos passos, do banco de dados original à obtenção de resultados.



1.14.2.1 FIGURA 18 – Switch de Software como Processo de Data Mining

Durante este trabalho pretendemos atuar em várias partes deste processo, como descrito a seguir:

- ✓ desenvolvimento de ferramentas de auxílio à extração de dados em banco de dados para servir de entrada a um sistema de *Data Mining*;
- ✓ uma ferramenta de extração de dados junto com uma interface integrada, incorporando a parte de *Data Mining* e que forneça conjuntos de itens que satisfazem o limiar;
- ✓ trabalhar nas heurísticas do algoritmo *A Priori* do *Data Mining* (regras de associação);
- ✓ implementação do algoritmo utilizando paralelização; e
- ✓ uma aplicação específica, utilizando as técnicas acima. A base de dados é o banco de dados do Zoneamento Agrícola.

VALIDAÇÃO

Este trabalho é baseado na quantidade de informações pluviométricas que nos foi fornecida pelo Departamento Nacional de Águas e Energia Elétrica. São 6.000 estações, algumas bem recentes. Precisávamos de séries históricas que permitissem um estudo estatístico mais apurado. Seleccionamos 2.380 estações, com o apoio do DNAEE e do Instituto Nacional de Meteorologia.

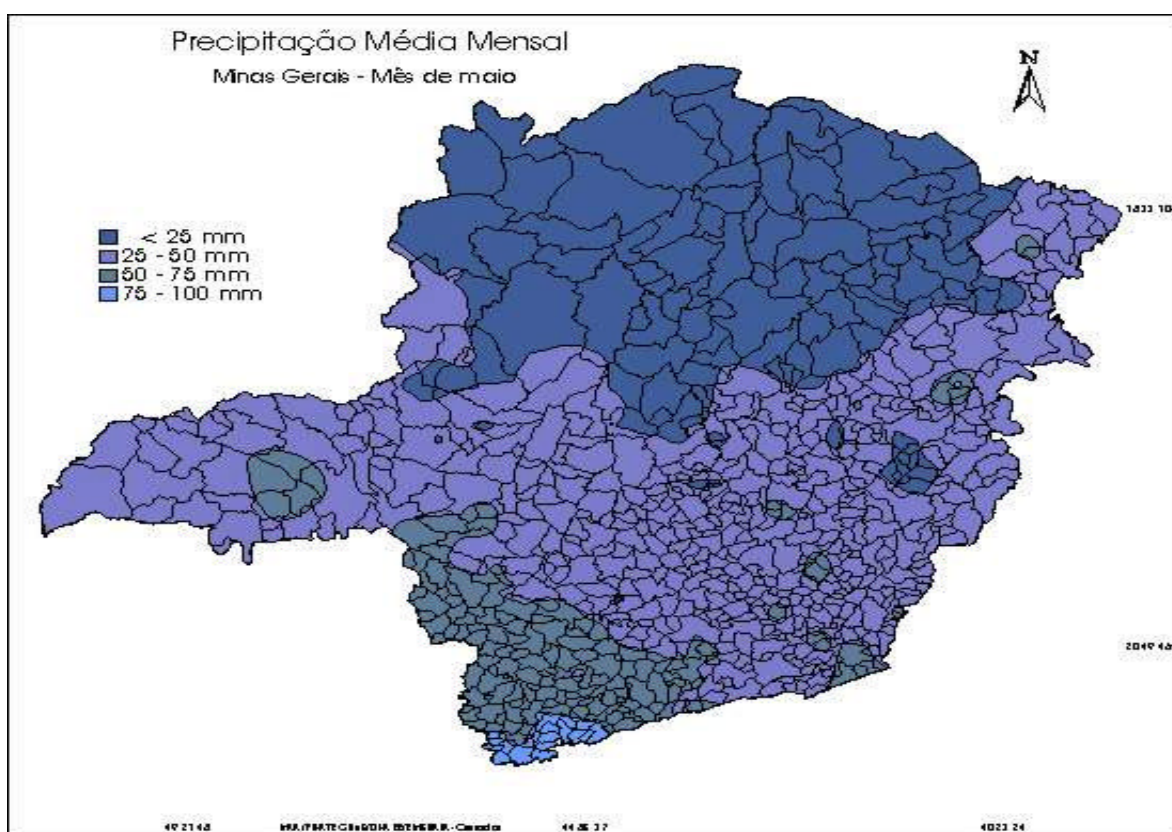
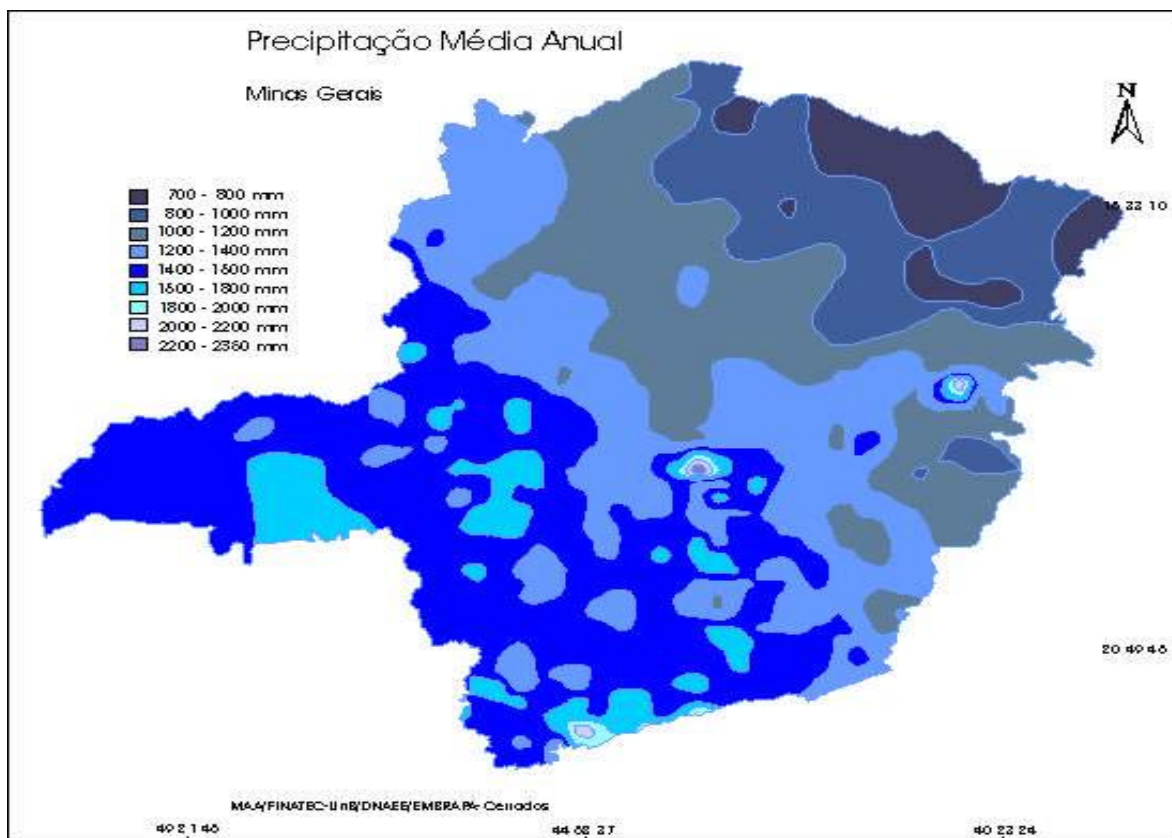
Órgão	Estação	Município	Data	Chuva (mm ³)	Granizo	Vendaval	Geada	Tromba d'água
INMET	83531	JOAQUIM	03/01/1999	48				
NMET	83570	AGUA DE CHAPECO	05/01/1999	42				
INMET	83897	ANGELINA	05/01/1999	68				
INMET	83887	ZORTEIA	06/09/1999	22				
INMET	83883	AGUAS DE CHAPECO	08/09/1999	4	FRACO		FRACA	
INMET	83887	ALTO BELA VISTA	08/09/1999	0,4	FRACO		FRACA	
INMET	83883	ARVOREDO	08/09/1999	4	FRACO		FRACA	
INMET	83883	CAIBI	08/09/1999	4	FRACO		FRACA	
INMET	83887	CAMPOS NOVOS	08/09/1999	0,4	FRACO		FRACA	
INMET	83887	CAPINZAL	08/09/1999	0,4	FRACO		FRACA	
INMET	83883	CAXAMBU DO SUL	08/09/1999	4	FRACO		FRACA	
INMET	83887	CELSO RAMOS	08/09/1999	0,4	FRACO		FRACA	
INMET	83883	CHAPECO	08/09/1999	4	FRACO		FRACA	
INMET	83883	CORDILHEIRA ALTA	08/09/1999	4	FRACO		FRACA	
INMET	83883	CUNHATAI	08/09/1999	4	FRACO		FRACA	
INMET	83883	GUATAMBU	08/09/1999	4	FRACO		FRACA	
INMET	83887	IPIRA	08/09/1999	0,4	FRACO		FRACA	
INMET	83883	ITAPIRANGA	08/09/1999	4	FRACO		FRACA	

A propósito, a grande dificuldade para executar um trabalho dessa amplitude foi os dados aos quais não tínhamos acesso. Nenhum grande projeto de pesquisa tinha esse banco de dados, de magnitude tal, que permitisse analisar o Brasil de uma forma global. Aproveitando a solicitação do Ministério da Agricultura, conseguimos acessar esses dados e iniciar o projeto.

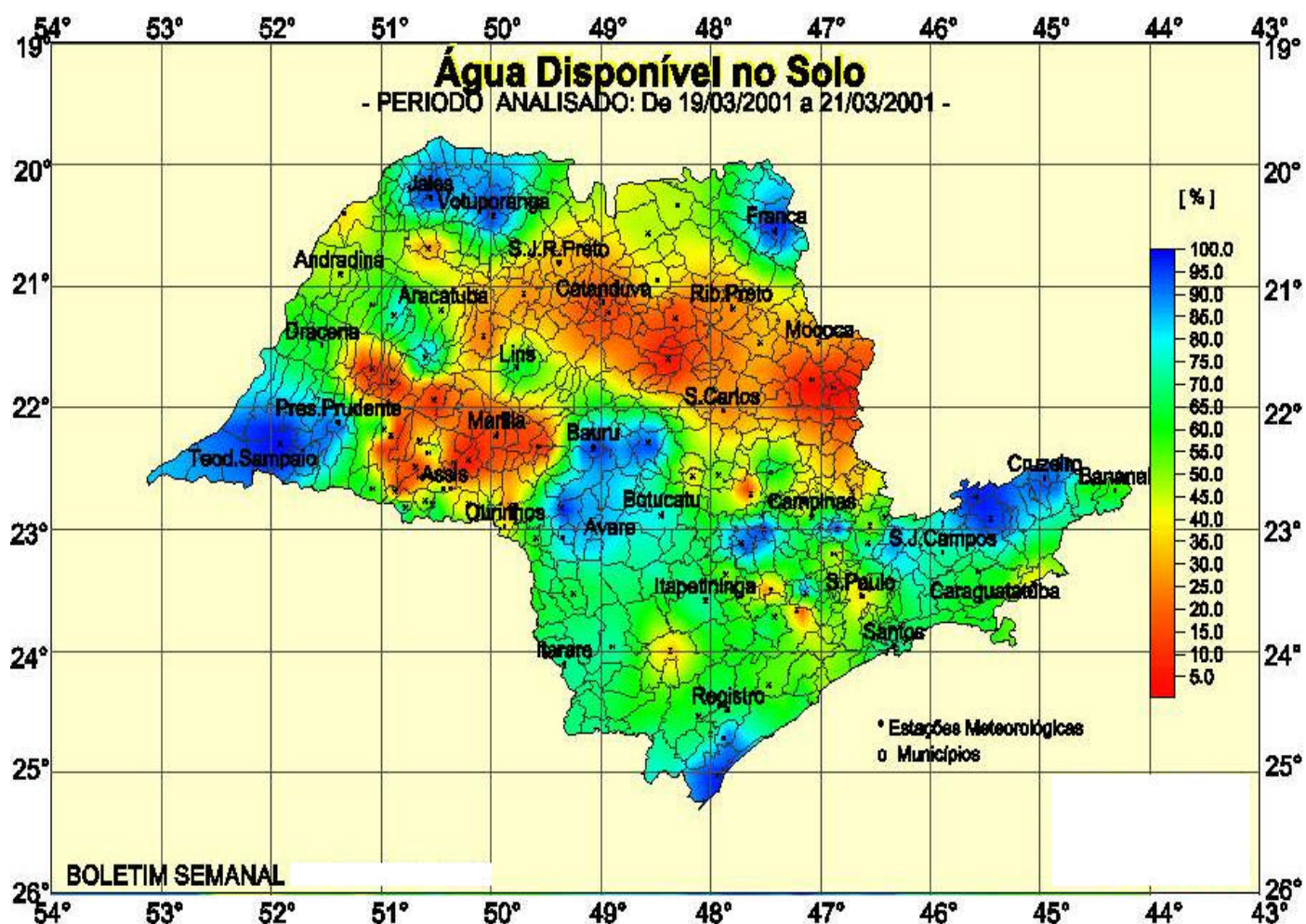
Além dos dados meteorológicos, outros essenciais, para cálculo de um parâmetro bastante utilizado, a evapotranspiração, foram solicitados ao Instituto Nacional de Meteorologia. Uma vez os dados conseguidos, conseguiu-se partir para a caracterização que o Ministério da Agricultura necessitava: o aparecimento dos veranicos, isto é, aqueles períodos secos, durante a estação chuvosa. Começamos a identificar os veranicos com dados das 2.380 estações analisadas, em períodos de 10 dias secos, períodos de 15 dias, períodos de 20 dias, etc., nos meses chuvosos e nos meses secos. Tudo foi quantificado e identificado espacialmente, na forma de mapas, determinando-se as datas de plantio.

[Ver tabela](#)

A idéia era exatamente esta: partindo desses calendários, verificar se era possível deslocar as datas de plantio para se evitar perdas, não caindo nos períodos com alta frequência de veranicos. A esse estudo também era acoplado a análise frequencial de distribuição de chuva (ASSAD, 1999).



De posse desses dados, além da evapotranspiração, ciclo de cultura e todas as demais informações disponíveis, agregamos tudo a um modelo. Foi utilizada a precipitação
 otranspiração, os coeficientes de cultura para cada uma das
 culturas e a disponibilidade de água no solo, da qual temos mais de 5.000 curvas de retenção



problema: o solo brasileiro. A heterogeneidade espacial de solos no Brasil é simplesmente ², os quais devíamos regionalizar. Trabalhávamos com modelos de simulação, considerando classes de retenção de água.

Simplificadamente classificamos os solos como tipo 1, tipo 2 e tipo 3, (baixa, média e alta capacidade de retenção de água, respectivamente) uma vez que nas indicações que o IPEA e o Ministério da Agricultura nos haviam passado, os problemas são seca e chuva forte (ASSAD, 1999).

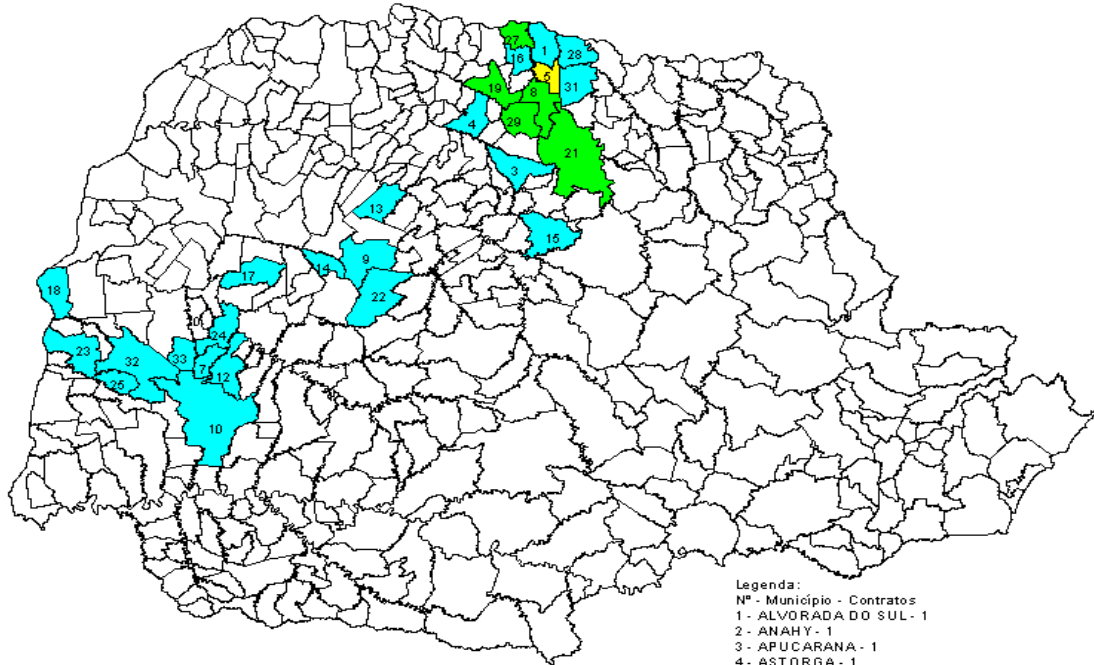
Ao rendimento potencial, foi cruzado o risco de chuva na colheita e, a este, foi cruzado, principalmente, o risco de geada na floração.



relacionados à duração da fase do ciclo das culturas, que estão relacionadas com o consumo de água pelas plantas, durante todo o seu desenvolvimento.

Na primeira fase, há uma parte importante da pesquisa, que é a recuperação de todos os dados, todos os trabalhos que cada Centro de pesquisa da Embrapa já tem, toda experiência de cada um deles na sua área e culturas com que trabalham. Esses dados foram todos compilados e, numa primeira fase, introduzidos em um programa de simulação para cada data de plantio. Simulando como é o desenvolvimento, o consumo de água durante toda fase da cultura e, como resultado, fornece-se o índice, com a sigla ISNA, que significa Índice de Satisfação das Necessidades de Água na fase crítica, a do florescimento dos grãos. Isto é feito por Posto Meteorológico. Trabalhamos com Postos, e estes geralmente são em número menor do que a quantidade de municípios.

Fases de Plantio
Paraná
Trigo - Safra 2000/2000
Situação na Data: 10/05/2000



Fases de Plantio

- Emergência(7 dias)
- Perfilamento(23 dias)
- Espigamento(65 dias)
- Maturação(100 dias)
- Colheita(110 dias)
- Sem informações

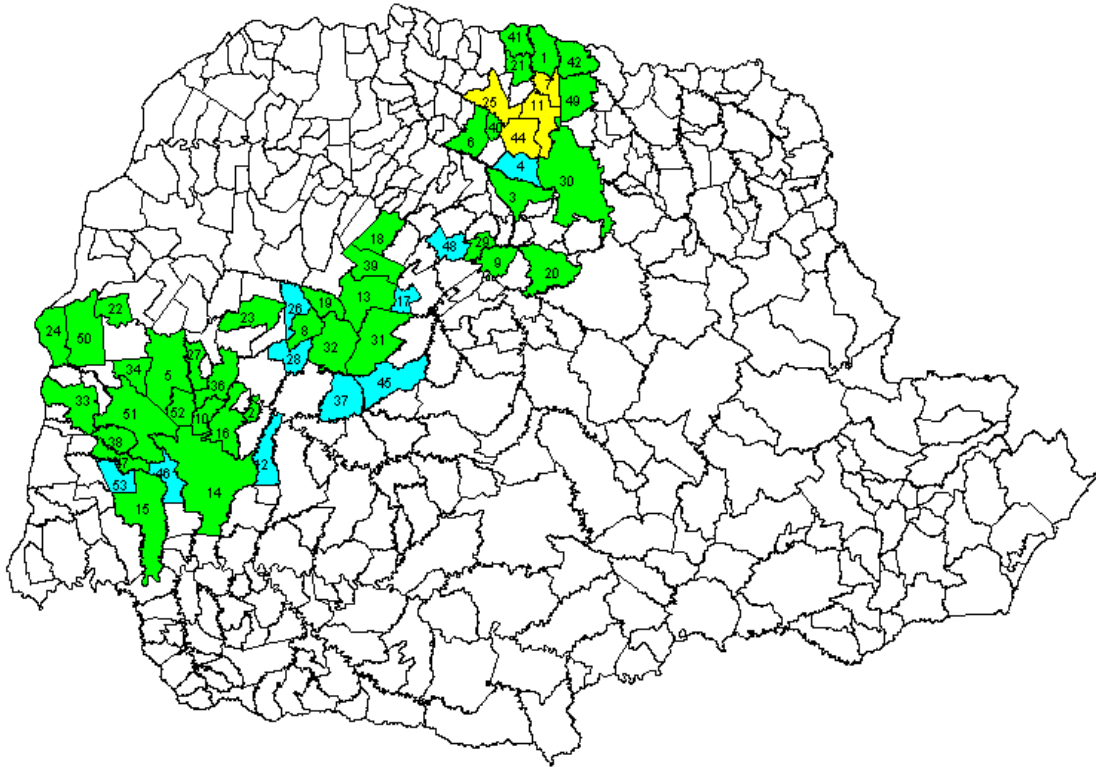
Legenda:

Nº - Município - Contratos

- 1 - ALVORADA DO SUL - 1
- 2 - ANAHY - 1
- 3 - APUCARANA - 1
- 4 - ASTORGA - 1
- 5 - BELA VISTA DO PARAISO - 10
- 6 - BOA ESPERANCA - 3
- 7 - CAFELANDIA - 12
- 8 - CAMBE - 13
- 9 - CAMPO MOURAO - 12
- 10 - CASCAVEL - 9
- 11 - CEU AZUL - 1
- 12 - CORBELIA - 5
- 13 - ENGENHEIRO BELTRAO - 1
- 14 - FAROL - 7
- 15 - FAXINAL - 1
- 16 - FLORESTOPOLIS - 1
- 17 - GOIOERE - 1
- 18 - GUAIRA - 1
- 19 - JAGUAPITA - 7
- 20 - JESUITAS - 9
- 21 - LONDRINA - 3
- 22 - LUIZIANA - 7
- 23 - MARECHAL CANDIDO RONDON - 2
- 24 - NOVA AURORA - 18
- 25 - OURO VERDE DO OESTE - 2
- 26 - PEABIRU - 1
- 27 - PORECATU - 7
- 28 - PRIMEIRO DE MAIO - 1
- 29 - ROLANDIA - 18
- 30 - SAO PEDRO DO IGUACU - 1
- 31 - SERTANOPOLIS - 2
- 32 - TOLEDO - 14
- 33 - TUPASSI - 4

Total:175 Contratos

Fases de Plantio
Paraná
Trigo - Safra 2000/200
Situação na Data: 10/06/2000



Legenda:

Nº. Município - Contratos

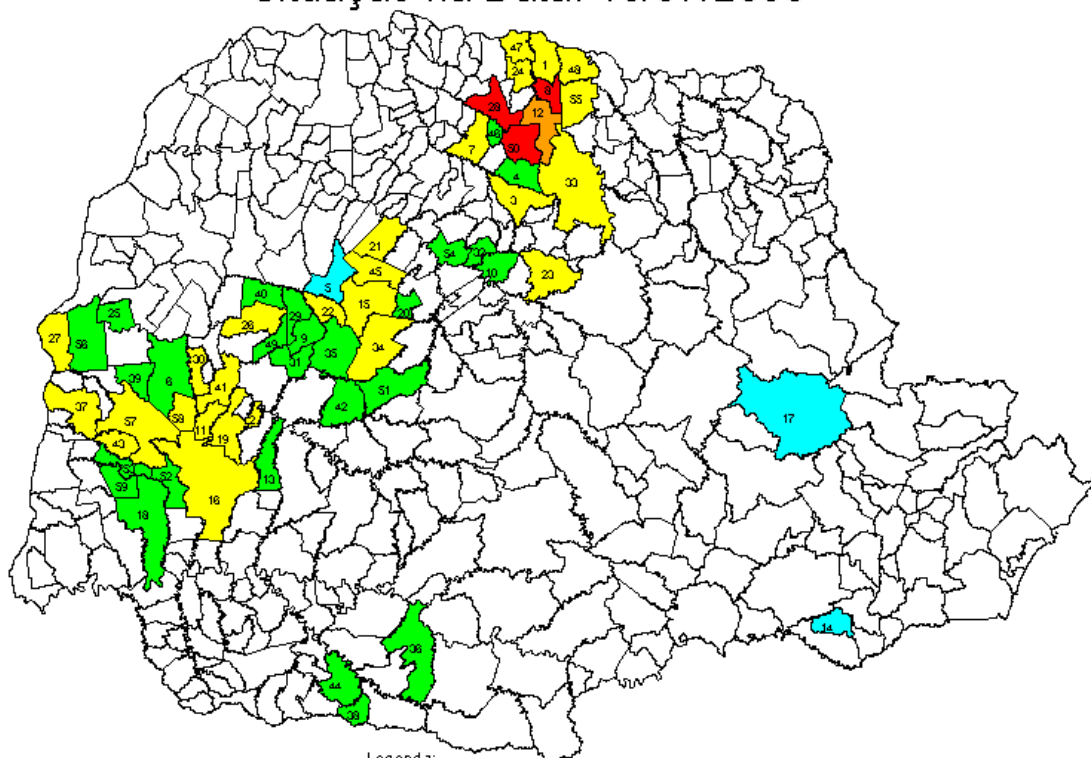
- 1 - ALVORADA DO SUL - 1
- 2 - ANAHY - 1
- 3 - APUCARANA - 1
- 4 - ARAPONGAS - 3
- 5 - ASSIS CHATEAUBRIAND - 7
- 6 - ASTORGA - 3
- 7 - BELA VISTA DO PARAISO - 10
- 8 - BOA ESPERANCA - 14
- 9 - BORRAZOPOLIS - 6
- 10 - CAFELANDIA - 31
- 11 - CAMBE - 16
- 12 - CAMPO BONITO - 1
- 13 - CAMPO MOURAO - 28
- 14 - CASCAVEL - 22
- 15 - CEU AZUL - 1
- 16 - CORBELIA - 6
- 17 - CORUMBATAI DO SUL - 1
- 18 - ENGENHEIRO BELTRAO - 1
- 19 - FAROL - 10
- 20 - FAXINAL - 6
- 21 - FLORESTOPOLIS - 1
- 22 - FRANCISCO ALVES - 2
- 23 - GOIOERE - 2
- 24 - GUAIRA - 6
- 25 - JAGUAPITA - 10
- 26 - JANIOPOLIS - 3
- 27 - JESUITAS - 10

- 28 - JURANDA - 3
 - 29 - KALORE - 1
 - 30 - LONDRINA - 3
 - 31 - LUIZIANA - 20
 - 32 - MAMBORE - 44
 - 33 - MARECHAL CANDIDO RONDON - 2
 - 34 - MARIPA - 1
 - 35 - MOREIRA SALES - 1
 - 36 - NOVA AURORA - 20
 - 37 - NOVA CANTU - 3
 - 38 - OURO VERDE DO OESTE - 9
 - 39 - PEABIRU - 3
 - 40 - PITANGUEIRAS - 3
 - 41 - PORECATU - 10
 - 42 - PRIMEIRO DE MAIO - 1
 - 43 - RANCHO ALEGRE DO OESTE - 1
 - 44 - ROLANDIA - 35
 - 45 - RONCADOR - 6
 - 46 - SANTA TEREZA DO OESTE - 1
 - 47 - SAO PEDRO DO IGUAQU - 1
 - 48 - SAO PEDRO DO IVAI - 1
 - 49 - SERTANOPOLIS - 2
 - 50 - TERRA ROXA - 1
 - 51 - TOLEDO - 73
 - 52 - TUPASSI - 17
 - 53 - VERA CRUZ DO OESTE - 1
- Total: 466 Contratos

Fases de Plantio

- Emergência(7 dias)
- Perfilhamento(23 dias)
- Espigamento(65 dias)
- Maturação(100 dias)
- Colheita(110 dias)
- Sem Informações

Fases de Plantio
Paraná
Trigo - Safra 2000/2000
Situação na Data: 10/07/2000



Legenda:

Nº - Município - Contratos

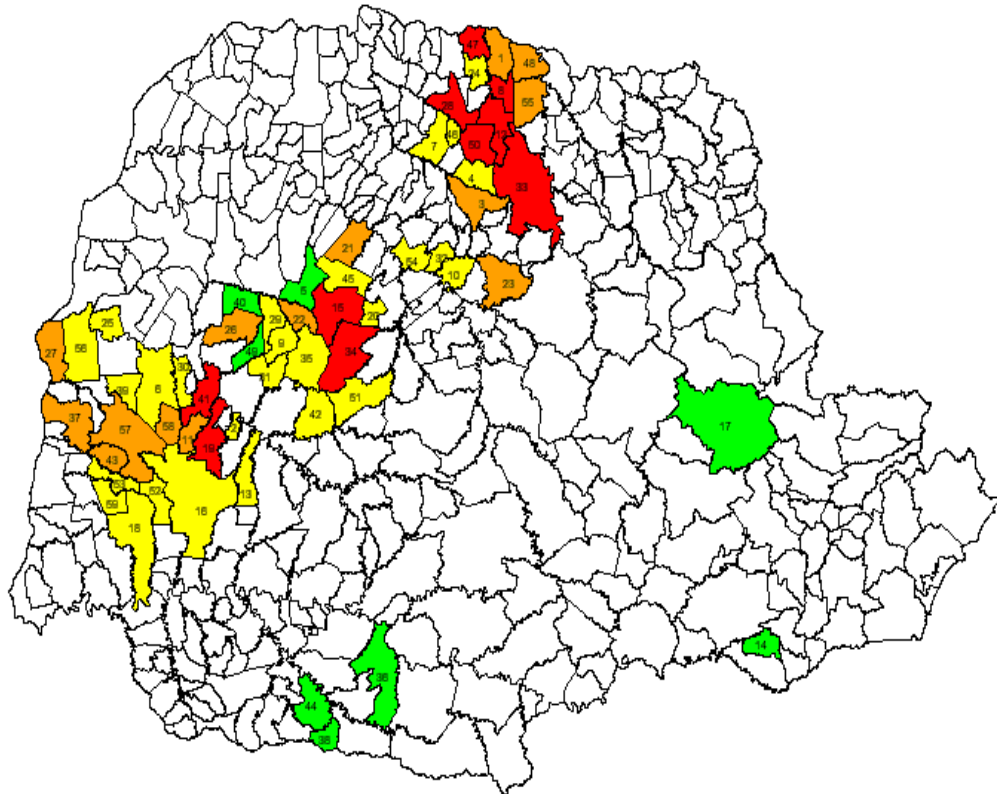
- 1 - ALVORADA DO SUL - 1
- 2 - ANAHY - 1
- 3 - APUCARANA - 1
- 4 - ARAPONGAS - 3
- 5 - ARARUNA - 1
- 6 - ASSIS CHATEAUBRIAND - 24
- 7 - ASTORGA - 3
- 8 - BELA VISTA DO PARAISO - 11
- 9 - BOA ESPERANCA - 16
- 10 - BORRAZOPOLIS - 6
- 11 - CAFELANDIA - 31
- 12 - CAMBE - 17
- 13 - CAMPO BONITO - 1
- 14 - CAMPO DO TENENTE - 1
- 15 - CAMPO MOURAO - 28
- 16 - CASCAVEL - 26
- 17 - CASTRO - 1
- 18 - CEU AZUL - 1
- 19 - CORBELIA - 6
- 20 - CORUMBATAI DO SUL - 1
- 21 - ENGENHEIRO BELTRAO - 1
- 22 - FAROL - 11
- 23 - FAXINAL - 9
- 24 - FLORESTOPOLIS - 1
- 25 - FRANCISCO ALVES - 2
- 26 - GOIOERE - 2
- 27 - GUAIRA - 7
- 28 - JAGUAPITA - 10
- 29 - JANIOPOLIS - 4
- 30 - JESUITAS - 11

- 31 - JURANDA - 3
 - 32 - KALORE - 1
 - 33 - LONDRINA - 3
 - 34 - LUIZIANA - 21
 - 35 - MAMBORE - 48
 - 36 - MANGUEIRINHA - 4
 - 37 - MARECHAL CANDIDO RONDON - 2
 - 38 - MARIOPOLIS - 1
 - 39 - MARIPA - 1
 - 40 - MOREIRA SALES - 1
 - 41 - NOVA AURORA - 20
 - 42 - NOVA CANTU - 3
 - 43 - OURO VERDE DO OESTE - 10
 - 44 - PATO BRANCO - 10
 - 45 - PEABIRU - 3
 - 46 - PITANGUEIRAS - 3
 - 47 - PORCATU - 10
 - 48 - PRIMEIRO DE MAIO - 1
 - 49 - RANCHO ALEGRE DO OESTE - 1
 - 50 - ROLANDIA - 37
 - 51 - RONCADOR - 7
 - 52 - SANTA TEREZA DO OESTE - 1
 - 53 - SAO PEDRO DO IGUAQU - 1
 - 54 - SAO PEDRO DO IVAI - 1
 - 55 - SERTANOPOLIS - 2
 - 56 - TERRA ROXA - 1
 - 57 - TOLEDO - 95
 - 58 - TUPASSI - 21
 - 59 - VERA CRUZ DO OESTE - 1
- Total: 551 Contratos

Fases de Plantio

- Emergência (7 dias)
- Perfilhamento (23 dias)
- Espigamento (65 dias)
- Maturação (100 dias)
- Colheita (110 dias)
- Sem Informações

Fase de Plantio Paraná Trigo - Safra 2000/2000 Situação na data 10/08/2000



Fases de Plantio

- Emergência(7 dias)
- Perfilhamento(23 dias)
- Espigamento(65 dias)
- Maturação(100 dias)
- Colheita(110 dias)
- Sem Informações

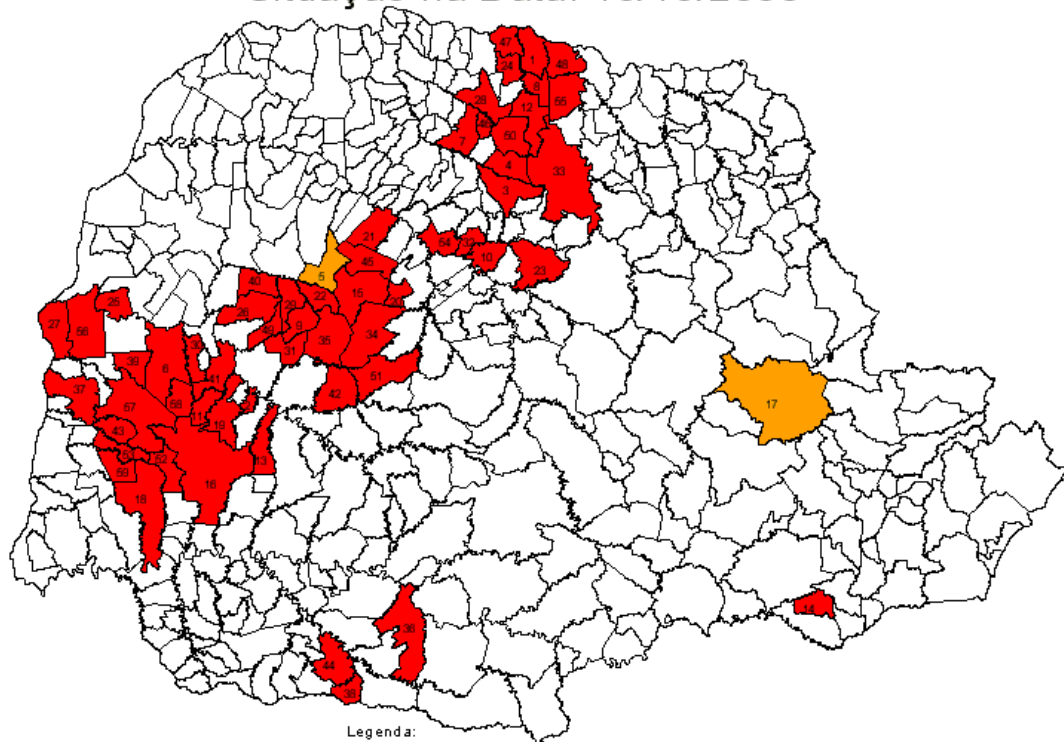
Legenda:

Nº - Município - Contratos

- | | |
|--------------------------------|----------------------------------|
| 1 - ALVORADA DO SUL - 1 | 30 - JESUITAS - 11 |
| 2 - ANAHY - 1 | 31 - JURANDA - 3 |
| 3 - APUCARANA - 1 | 32 - KALORE - 1 |
| 4 - ARAPONGAS - 3 | 33 - LONDRINA - 3 |
| 5 - ARARUNA - 1 | 34 - LUIZIANA - 21 |
| 6 - ASSIS CHATEAUBRIAND - 24 | 35 - MAMBORE - 48 |
| 7 - ASTORGA - 3 | 36 - MANGUEIRINHA - 4 |
| 8 - BELA VISTA DO PARAISO - 11 | 37 - MARECHAL CANDIDO RONDON - 2 |
| 9 - BOA ESPERANCA - 16 | 38 - MARIO POLIS - 1 |
| 10 - BORRAZOPOLIS - 6 | 39 - MARIPA - 1 |
| 11 - CAFELANDIA - 31 | 40 - MOREIRA SALES - 1 |
| 12 - CAMBE - 17 | 41 - NOVA AURORA - 20 |
| 13 - CAMPO BONITO - 1 | 42 - NOVA CANTU - 3 |
| 14 - CAMPO DO TENENTE - 1 | 43 - OURO VERDE DO OESTE - 10 |
| 15 - CAMPO MOURAO - 28 | 44 - PATO BRANCO - 10 |
| 16 - CASCAVEL - 27 | 45 - PEABIRU - 3 |
| 17 - CASTRO - 1 | 46 - PITANGUEIRAS - 3 |
| 18 - CEU AZUL - 1 | 47 - PORCATU - 10 |
| 19 - CORBELIA - 6 | 48 - PRIMEIRO DE MAIO - 1 |
| 20 - CORUMBATAI DO SUL - 1 | 49 - RANCHO ALEGRE DO OESTE - 1 |
| 21 - ENGENHEIRO BELTRAO - 1 | 50 - ROLANDIA - 37 |
| 22 - FAROL - 11 | 51 - RONCADOR - 7 |
| 23 - FAXINAL - 9 | 52 - SANTA TEREZA DO OESTE - 1 |
| 24 - FLORESTOPOLIS - 1 | 53 - SAO PEDRO DO IGUAÇU - 1 |
| 25 - FRANCISCO ALVES - 2 | 54 - SAO PEDRO DO IVAI - 1 |
| 26 - GOIOERE - 2 | 55 - SERTANOPOLIS - 2 |
| 27 - GUAIRA - 7 | 56 - TERRA ROXA - 1 |
| 28 - JAGUAPITA - 10 | 57 - TOLEDO - 95 |
| 29 - JANIOPOLIS - 4 | 58 - TUPASSI - 21 |
| | 59 - VERA CRUZ DO OESTE - 1 |

Total:552 Contratos

Fases de Plantio
Paraná
Trigo - Safra 2000/2000
Situação na Data: 10/10/2000



Legenda:

Nº - Município - Contratos

- 1 - ALVORADA DO SUL - 1
- 2 - ANAHY - 1
- 3 - APUCARANA - 1
- 4 - ARAPONGAS - 3
- 5 - ARARUNA - 1
- 6 - ASSIS CHATEAUBRIAND - 24
- 7 - ASTORGA - 3
- 8 - BELA VISTA DO PARAISO - 11
- 9 - BOA ESPERANCA - 16
- 10 - BORRAZOPOLIS - 6
- 11 - CAFELANDIA - 31
- 12 - CAMBE - 17
- 13 - CAMPO BONITO - 1
- 14 - CAMPO DO TENENTE - 1
- 15 - CAMPO MOURAO - 28
- 16 - CASCAVEL - 29
- 17 - CASTRO - 1
- 18 - CEU AZUL - 1
- 19 - CORBELIA - 6
- 20 - CORUMBATAI DO SUL - 1
- 21 - ENGENHEIRO BELTRAO - 2
- 22 - FAROL - 11
- 23 - FAXINAL - 9
- 24 - FLORESTOPOLIS - 1
- 25 - FRANCISCO ALVES - 2
- 26 - GOIOIERE - 2
- 27 - GUAIRA - 7
- 28 - JAGUAPITA - 10
- 29 - JANIOPOLIS - 4
- 30 - JESUITAS - 11

- 31 - JURANDA - 3
 - 32 - KALORE - 1
 - 33 - LONDRINA - 3
 - 34 - LUIZIANA - 21
 - 35 - MAMBORE - 48
 - 36 - MANGUEIRINHA - 4
 - 37 - MARECHAL CANDIDO RONDON - 2
 - 38 - MARIOPOLIS - 1
 - 39 - MARIPA - 1
 - 40 - MOREIRA SALES - 1
 - 41 - NOVA AURORA - 20
 - 42 - NOVA CANTU - 3
 - 43 - OURO VERDE DO OESTE - 10
 - 44 - PATO BRANCO - 10
 - 45 - PEABIRU - 4
 - 46 - PITANGUEIRAS - 3
 - 47 - PORECATU - 10
 - 48 - PRIMEIRO DE MAIO - 1
 - 49 - RANCHO ALEGRE DO OESTE - 1
 - 50 - ROLANDIA - 37
 - 51 - RONCADOR - 7
 - 52 - SANTA TEREZA DO OESTE - 1
 - 53 - SAO PEDRO DO IGUAQU - 1
 - 54 - SAO PEDRO DO IVAI - 1
 - 55 - SERTANOPOLIS - 2
 - 56 - TERRA ROXA - 1
 - 57 - TOLEDO - 95
 - 58 - TUPASSI - 21
 - 59 - VERA CRUZ DO OESTE - 1
- Total: 556 Contratos

Fases de Plantio

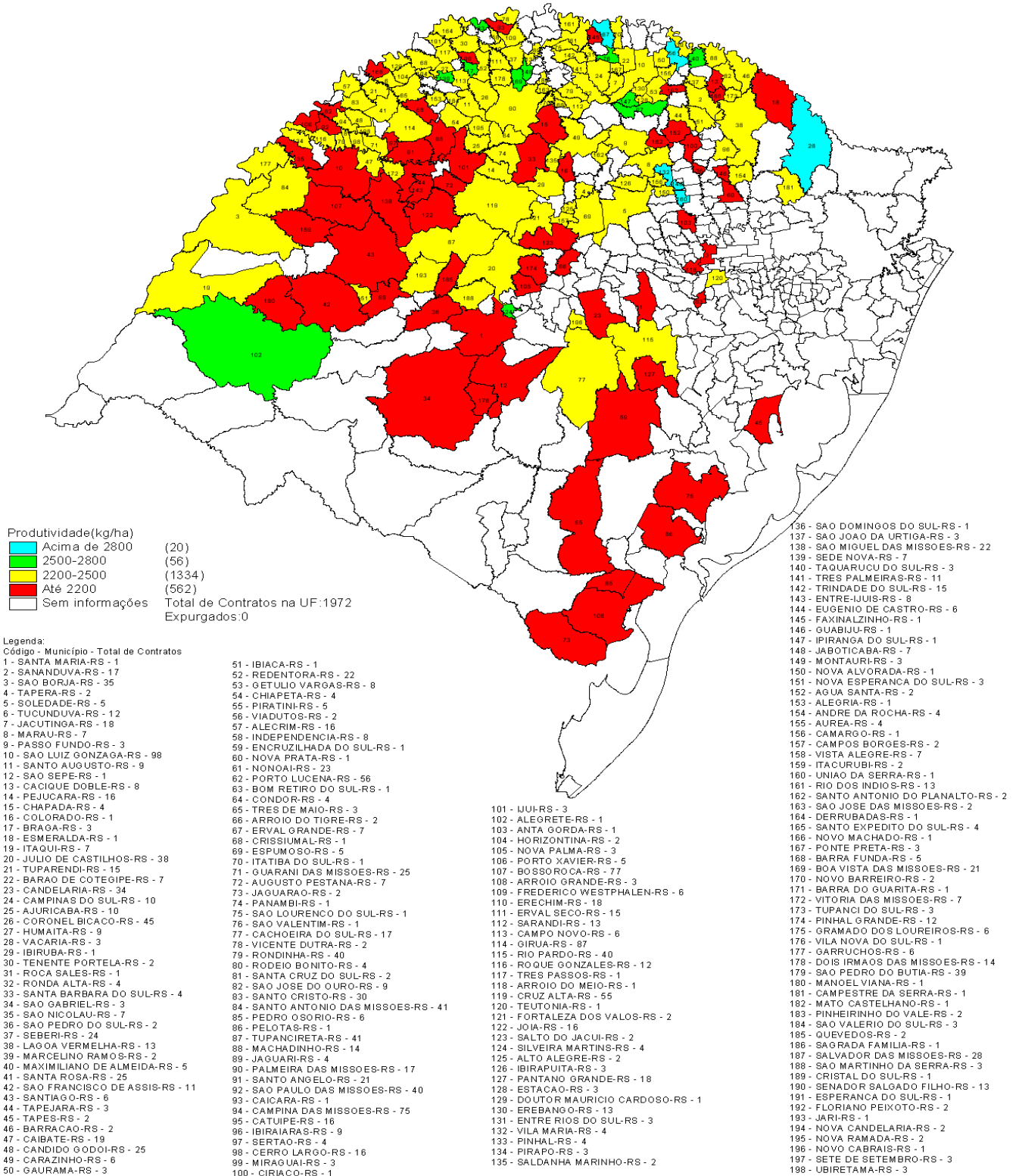
- Emergência(7 dias)
- Perfilhamento(23 dias)
- Espigamento(65 dias)
- Maturação(100 dias)
- Colheita(110 dias)
- Sem Informações

drico

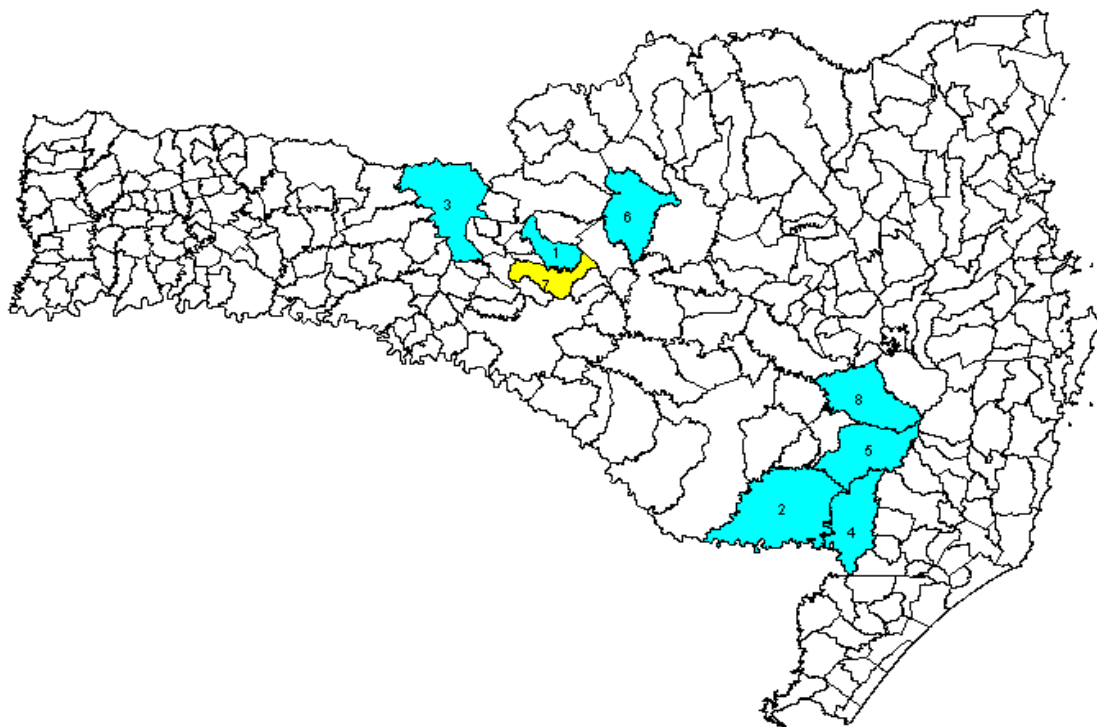
por município, partindo de dados da cultura, dados de chuva, fazendo simulações e definindo o atendimento hídrico por Posto e chegando até o município. A seguir, esses dados, definição do atendimento hídrico nessas culturas, passam por uma filtragem e triagem e, nessa fase, atendimento hídrico, consideramos também o atendimento térmico, porque há culturas com problemas em determinada fase, onde as temperaturas são elevadas ou baixas. Portanto, além do atendimento hídrico, existe uma filtragem, uma triagem no atendimento térmico.

O que interessa nesse primeiro momento são as datas favoráveis de plantio. Essas municípios. São feitos mapas e esses resultados são discutidos com especialistas de cada área e de cada cultura.

Rio Grande do Sul Soja 1999/2000



Santa Catarina
Maçã - Safra 2000/2001
Produtividade (Kg/ha)



Produtividade (Kg/ha)

Acima de 25000
22000-25000
18000-22000
Até 18000
Sem Informações

Legenda:
Código - Município - Total de Contratos
1 - VIDEIRA-SC - 2
2 - SAO JOAQUIM-SC - 3
3 - AGUA DOCE-SC - 4
4 - BOM JARDIM DA SERRA-SC - 4
5 - URUBICI-SC - 8
6 - LEBON REGIS-SC - 1
7 - TANGARA-SC - 3
8 - BOM RETIRO-SC - 5
Total de Contratos na UF:30
Expurgados:0

Custeio Agrícola por Cultura 1999/1999

Cultura	Contratos	Área Total (ha)	Valor Total	Produção Total(Kg)	Área Média	Valor por ha	Produtiv (Kg/ha)
TRIGO	3631	232.062,52	R\$65.104.127,33	539.877.544,00	63,91	R\$280,55	2.326,43
Totalizações	3.631	232.062,52	R\$65.104.127,33	539.877.544,00	63,91	R\$280,55	2.326,43

Fonte: CER-PROAGRO

Custeio Agrícola por Cultura 1999/2000

Cultura	Contratos	Área Total (ha)	Valor Total	Produção Total(Kg)	Área Média	Valor por ha	Produtiv (Kg/ha)
ALGODAO	410	6.994,65	R\$4.806.344,64	13.925.684,00	17,06	R\$687,15	1.990,91
ARROZ	489	43.391,00	R\$18.806.511,85	143.380.800,00	88,73	R\$433,42	3.304,39
FEIJAO	591	9.837,50	R\$5.235.327,19	18.714.170,00	16,65	R\$532,18	1.902,33
MACA	228	1.432,08	R\$7.089.599,31	42.791.000,00	6,28	R\$4.950,56	29.880,31
MILHO	4897	169.574,52	R\$64.855.470,03	880.997.174,00	34,63	R\$382,46	5.195,34
SOJA	3621	224.315,80	R\$68.384.442,04	575.527.024,00	61,95	R\$304,86	2.565,70
Totalizações	10.236	455.545,55	R\$169.177.695,06	1.675.335.852,00	44,50	R\$371,37	3.677,65

Fonte: CER-PROAGRO

Custeio Agrícola por UF 1999/2000

UF	Cultura	Contratos	Área Total (ha)	Valor Total	Produção Total(Kg)	Área Média	Valor por ha	Produtiv (Kg/ha)
BA	ALGODAO	1	42,00	R\$31.185,00	50.400,00	42,00	R\$742,50	1.200,00
BA	FEIJAO	2	80,00	R\$33.622,00	80.000,00	40,00	R\$420,28	1.000,00
BA	MILHO	2	300,00	R\$139.843,64	1.776.000,00	150,00	R\$466,15	5.920,00
CE	ALGODAO	100	1.351,00	R\$469.128,00	1.625.210,00	13,51	R\$347,25	1.202,97
DF	FEIJAO	1	20,00	R\$16.057,66	36.000,00	20,00	R\$802,88	1.800,00
DF	MILHO	2	50,00	R\$29.336,00	300.000,00	25,00	R\$586,72	6.000,00
GO	ALGODAO	22	1.854,00	R\$1.557.150,48	4.288.000,00	84,27	R\$839,89	2.312,84
GO	ARROZ	47	2.700,00	R\$865.025,45	6.569.000,00	57,45	R\$320,38	2.432,96
GO	FEIJAO	2	73,00	R\$46.097,49	155.000,00	36,50	R\$631,47	2.123,29
GO	MILHO	122	9.241,00	R\$4.322.758,92	55.577.910,00	75,75	R\$467,78	6.014,27
GO	SOJA	253	30.577,00	R\$11.735.067,57	86.534.000,00	120,86	R\$383,79	2.830,04
MA	ARROZ	108	9.018,00	R\$3.554.404,12	23.846.800,00	83,50	R\$394,15	2.644,36
MA	MILHO	70	2.735,00	R\$1.298.907,86	10.818.000,00	39,07	R\$474,92	3.955,39
MA	SOJA	52	8.627,00	R\$3.073.392,38	24.259.083,00	165,90	R\$356,25	2.812,00
MG	ALGODAO	20	1.264,00	R\$1.209.052,34	3.044.650,00	63,20	R\$956,53	2.408,74
MG	FEIJAO	2	252,00	R\$182.580,50	605.000,00	126,00	R\$724,53	2.400,79
MG	MILHO	26	2.416,00	R\$1.130.230,05	12.541.000,00	92,92	R\$467,81	5.190,81
MG	SOJA	37	4.662,00	R\$1.757.523,62	11.604.000,00	126,00	R\$376,99	2.489,06
MS	ALGODAO	6	1.071,00	R\$828.332,83	2.512.000,00	178,50	R\$773,42	2.345,47
MS	ARROZ	1	2,00	R\$293,76	2.000,00	2,00	R\$146,88	1.000,00

MS	FEIJAO	1	2,00	R\$563,04	1.000,00	2,00	R\$281,52	500,00
MS	MILHO	38	2.104,00	R\$778.038,77	11.668.000,00	55,37	R\$369,79	5.545,63
MS	SOJA	62	10.937,00	R\$4.132.202,86	29.731.000,00	176,40	R\$377,82	2.718,39
MT	ARROZ	4	930,00	R\$209.035,50	2.196.000,00	232,50	R\$224,77	2.361,29
MT	SOJA	3	778,00	R\$274.995,04	2.108.000,00	259,33	R\$353,46	2.709,51
PB	ALGODAO	77	205,00	R\$97.273,35	285.380,00	2,66	R\$474,50	1.392,10
PI	ARROZ	122	13.560,00	R\$4.236.144,78	29.473.000,00	111,15	R\$312,40	2.173,53
PI	MILHO	24	609,00	R\$172.973,96	1.737.200,00	25,38	R\$284,03	2.852,55
PI	SOJA	24	5.017,00	R\$1.781.355,05	13.313.000,00	209,04	R\$355,06	2.653,58
PR	ALGODAO	32	609,65	R\$412.468,64	1.485.044,00	19,05	R\$676,57	2.435,90
PR	ARROZ	1	14,00	R\$8.767,92	77.000,00	14,00	R\$626,28	5.500,00
PR	FEIJAO	83	1.315,00	R\$678.995,08	2.270.000,00	15,84	R\$516,35	1.726,24
PR	MILHO	411	13.004,02	R\$5.482.729,02	83.916.752,00	31,64	R\$421,62	6.453,14
PR	SOJA	946	44.574,10	R\$13.859.565,28	130.238.441,00	47,12	R\$310,93	2.921,84
RN	ALGODAO	151	526,00	R\$133.386,50	455.000,00	3,48	R\$253,59	865,02
RS	ARROZ	77	5.927,00	R\$3.975.836,91	31.802.000,00	76,97	R\$670,80	5.365,61
RS	FEIJAO	227	1.606,50	R\$544.247,77	2.610.170,00	7,08	R\$338,78	1.624,76
RS	MILHO	3806	117.534,50	R\$41.512.782,26	572.960.312,00	30,88	R\$353,20	4.874,83
RS	SOJA	1975	103.810,70	R\$26.498.028,97	233.311.500,00	52,56	R\$255,25	2.247,47
SC	ARROZ	85	6.214,00	R\$3.924.498,96	35.153.000,00	73,11	R\$631,56	5.657,06
SC	FEIJAO	234	6.376,00	R\$3.678.498,25	12.784.000,00	27,25	R\$576,93	2.005,02
SC	MACA	228	1.432,08	R\$7.089.599,31	42.791.000,00	6,28	R\$4.950,56	29.880,31
SC	MILHO	308	15.795,00	R\$7.530.383,10	100.863.000,00	51,28	R\$476,76	6.385,75
SC	SOJA	216	10.861,00	R\$3.822.522,99	32.931.000,00	50,28	R\$351,95	3.032,04
SP	ALGODAO	1	72,00	R\$68.367,50	180.000,00	72,00	R\$949,55	2.500,00
SP	ARROZ	17	1.789,00	R\$1.031.804,71	7.276.000,00	105,24	R\$576,75	4.067,08
SP	FEIJAO	39	113,00	R\$54.665,40	173.000,00	2,90	R\$483,76	1.530,97
SP	MILHO	80	5.205,00	R\$2.192.612,34	26.264.000,00	65,06	R\$421,25	5.045,92
SP	SOJA	50	4.187,00	R\$1.338.939,78	10.835.000,00	83,74	R\$319,78	2.587,77
TO	ARROZ	27	3.237,00	R\$1.000.699,74	6.986.000,00	119,89	R\$309,14	2.158,17
TO	MILHO	8	581,00	R\$264.874,11	2.575.000,00	72,63	R\$455,89	4.432,01
TO	SOJA	3	285,00	R\$110.848,50	662.000,00	95,00	R\$388,94	2.322,81
Totalizações		10.236	455.545,55	R\$169.177.695,06	1.675.335.852,00	44,50	R\$371,37	3.677,65

Fonte: CER-PROAGRO

Esse é o fluxo. Todas essas fases já foram testadas pelo nosso e por outros grupos.

Atualmente com esta proposta de *Data Warehouse*, inclui-se a qualidade e a quantidade desses dados, com bastante cuidado e critério, para que sejam o mais confiáveis possíveis.

No caso de São Paulo, a rede de Postos utilizada é a do DAEE - Departamento de Estatística. Recebemos desse banco do DAEE dados de um total de 2.100 estações. Dessas estações suprimiram-se algumas, chegando a um número aproximado de 1.300 estações. Destas, retirou-se 390 estações com 30 anos de dados - o mínimo seriam 10 anos - que é um período matematicamente adequado, que comporta todas as variações climáticas normais no Estado. Há regiões que têm mais variações e essa variação também é maior onde há menos Postos. Esses 30 anos de dados dos 390 Postos de chuva, representam o fornecimento de água às plantas, no caso de culturas de sequeiro.

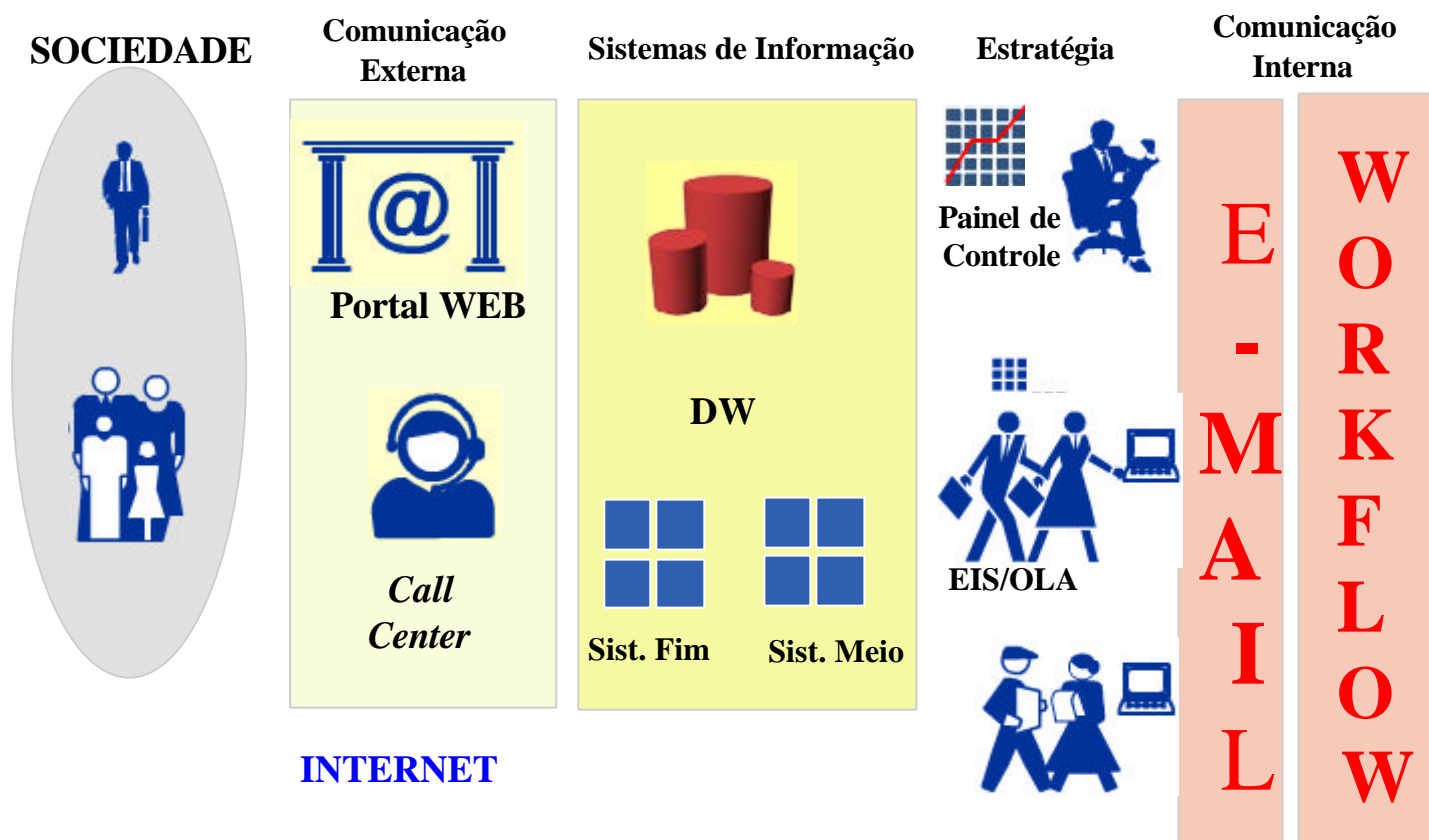
Também utiliza-se um fator econômico que é a duração do ciclo. Ciclos precoces e super precoces, às vezes tem água, às vezes tem a temperatura um pouco acima de 10°C, mas pela duração do ciclo se torna economicamente inviável.

Para ciclos normais utilizamos 165 dias. Multiplicando todos esses parâmetros utilizados num total de 2.500.000 simulações para se chegar às datas favoráveis, considera-se 80% dos casos.

A partir daí, e em tempo real, esses dados pluviométricos estarão disponíveis, transmitidos via satélite. Os dados vão para as centrais estaduais ou ao próprio Instituto de Estatística do Estado, e são encaminhadas ao Ministério da Agricultura, o que virá facilitar enormemente o trabalho de monitoramento e disponibilizados aos bancos para que possam financiar as lavouras.

.Para finalizar, é importante frisar que já conseguimos entregar ao Ministério da Agricultura 1.800.000 informações referentes à data de plantio, tipo de solo, e variedades de cultivares necessárias para nortear o Crédito Agrícola.

1.15 Visão geral do sistema proposto



1.1.12 FIGURA 10 Visão geral do sistema proposto

Benefícios ou resultados alcançados:

Bom para os usuários:

- ◆ Maior participação (qualquer um pode contribuir);
- ◆ Visões personalizadas dos dados e aplicações;
- ◆ Uma única e confiável fonte de informações;

Bom para o IT:

- ◆ fácil de desenvolver e administrar;
- ◆ seguro e gerenciável;
- ◆ padronização do acesso dos usuários finais;

Bom para a corporação:

- ◆ aumenta a produtividade;
- ◆ reduz os custos de infra-estrutura;
- ◆ facilita a tomada de decisões;

Com essa Estrutura Gerencial e de Arquitetura Funcional, pode-se reduzir o consumo de tempo na busca de soluções, e despesas desnecessárias, melhorando com isso a qualidade do produto final.

5 - CONCLUSÃO

Conforme apresentado, o software pode ser encarado como um bem ou serviço, dependendo de sua finalidade específica de uso, e da forma de identificação de sua necessidade e origem. Esses fatores são de fundamental importância na colocação do desenvolvimento do produto. A conceituação de Projeto, permite trabalhar com o produto software da mesma maneira que com outros produtos que venham a ser colocados no mercado. O entendimento da identificação correta dos conceitos de objetivos, metas, especificação de requisitos e origem permitirá uma clareza na fase de estruturação para o desenvolvimento do produto.

Da mesma forma que qualquer tipo de produto, que vise atender a um certo nicho de mercado necessita de uma estrutura de gestão adequada, para que sejam atendidos os objetivos e metas estabelecidos, o software não poderia ser uma exceção à regra. Porém, por se tratar de um produto de uso bastante específico, conforme suas aplicações, o software necessita de metodologias, técnicas e processos especiais para o seu desenvolvimento.

A qualidade desses produtos, alguns de uso bastante generalizado e outros de dedicação exclusiva, está relacionada com a escolha de um melhor caminho para as suas obtenções, e que por meio de uma estrutura de gestão adequada, possibilite o máximo desempenho dos trabalhos e tarefas realizadas pelas equipes envolvidas em seus desenvolvimentos.

O tratamento das informações coletadas junto aos clientes ou usuários deve conduzir a uma clara interpretação de suas reais necessidades, traduzidas em termos adequados ao entendimento do pessoal técnico, que lida diretamente com os processos, fato que agiliza a

A estrutura gerencial é construída sobre um modelo padrão não departamentalizado, adequando-se às características específicas das tarefas de desenvolvimento, assim como a arquitetura funcional, na qual são estabelecidos os procedimentos de controle, recorrência externa, capacitação interna, entre outros.

Portanto, o maior problema em *Data Warehouse* é a qualidade dos dados. Para evitar o princípio de GIGO (*garbage in/garbage out* - literalmente lixo dentro, lixo fora), os dados devem ter valores nulos mínimos, porque isso afeta os resultados da mineração de dados. A chave é, continuamente, monitorar os dados à medida em que vão sendo adicionados ao *Data Warehouse* e fazer um exame formal dos dados através de uma mineração preliminar para assegurar a integridade dos mesmos.

Os *Data warehouses* requerem o carregamento incremental de novos dados em bases periódicas entre janelas de tempo estreitas; o desempenho do processamento não deve restringir artificialmente o volume de dados requerido para o negócio. Muitas etapas devem ser tomadas para o carregamento de dados novos ou modificados dentro do *Data Warehouse*, incluindo conversões de dados, filtragem, reformatação, checagem de integridade, armazenamento físico, indexação e modificação dos metadados conforme visto na Figura 18 (Swicth de Softawe como processo de *Data Mining* - página 56).

A vantagem da aplicação dessa metodologia reside no fato de que logo no início do planejamento, evidenciam-se os pontos de conflitos, gargalos tecnológicos, o que evita desperdício de energia em caminhos que não conduzirão ao alcance das metas e objetivos, apurados diretamente junto aos usuários. Outra vantagem é a fácil visualização, por se tratar de um conjunto de matrizes constituídas por símbolos gráficos. Essa facilidade de visualização aumenta também a velocidade no remanejamento de atividades, e correções de planjenamento.

Para competir efetivamente, empresas ou órgãos públicos devem ser capazes de compreender seus dados disponíveis. Identificar padrões e tomar decisões num prazo que permita se manter a competitividade. A mineração de dados tem sido aplicada para achar respostas no processo de minimização de custos, gerenciamento de armazenagem e geração

Algumas das áreas não comerciais em que a mineração de dados pode atuar, incluem as indústrias agrícolas, ciências, prevenção de incêndios florestais, identificação de estruturas químicas, detecção de pragas vegetais ou animais, diagnóstico climático e planejamento de política agrícola. Muito da ênfase agora está sendo dada no estabelecimento de *Data Warehouse*, pois um dos pré-requisitos para a mineração de dados é que tais dados estejam livres de anomalias. Frequentemente, sínteses sobre os dados derivados da *Data Warehouse* levam a futuras consultas e mais análises subseqüentes.

Conforme apresentado no decorrer desta dissertação, cada um dos exemplos citados tem um contexto comum, eles utilizam o conhecimento sobre agricultores ou produtores implícitos no *Data Warehouse* para reduzir custos e melhorar o valor da relação com o consumidor. Estas organizações ou órgãos públicos podem então, concentrar seus esforços em cultivares mais importantes (lucrativos) e ainda, desenvolver estratégias para políticas agrícolas direcionadas no Brasil.

BIBLIOGRAFIA

- ADLER, P. & WINOGRAD, T. (eds.). **Usability: Turning Technologies Into Tools**. New York (NY) : Oxford University Press, 1993.
- ALLEN Cypher. Watch What I do. **Journal of Visual Languages and Computing**, 1994a.
- ANDERSEN, P.B.; HOLMQVIST, T.B.; JENSEN, F.F. (eds.). **The computer as Medium**. Cambridge : University Press, 1993.
- BARNETT, V. **Comparative Statistical Inference**. New York : John Wiley & Sons, 1982.
- BRAND, S. **The Media Lab: Inventing the future at MIT**. New York : Viking, 1987.
- BRASIL. MINISTÉRIO DA AGRICULTURA E DO ABASTECIMENTO.
<http://www.agricultura.gov.br/>
- BUCHANAN, B.; BARSTOW, D; BECHTEL, R. et al. Constructing Expert Systems. In HAYESROTH F., WATERMAN, D. and LENAT, D. (eds.). **Building Expert Systems**, pp. 127-168, 1983.
- CAMPOS, Vicente Falconi. **TQC: Controle da qualidade total**, 1994.
- CARD, S.; MORAN, T.; NEWELL A. **The Psychology of Human-Computer Interaction**. Hillsdale (NJ) : Lawrence Erlbaum Associates.
- CLEAL, D.; HEATON, N. **Knowledge-Based Systems: Implications for Human-Computer Interfaces**. Chichester, England : Ellis Horwood Books, 1988.
- COCHRAN, W.G. **Sampling Techniques**. New York : John Wiley & Sons Ltd., 1963.
- CONAB – Companhia Nacional de Abastecimento. **Sensibilização para a Qualidade**. Brasília, 2000.
- EAGLEMANN, J.R. An Experimentally derived model for actual evapotranspiration. **Agricultural Meteorology**. Amsterdam, v. 8, p. 385-394, 1971.
- EUREKA, William E.; RYAN, Nancy E. **Perspectivas gerenciais do desdobramento da função qualidade**. Qualitymark Editora, 1994.
- EVERITT, B. **Cluster Analysis**. London : Heinemann Educational Books Ltd., 1974.

- GUINTA, Lawrence R.; PRAIZLER, Nancy C. **Manual do QFD**. LTC, 1994.
- HAN , G.J. & SHAPIRO , S.S. **Statistical Models in Engineering**. New York : John : Wiley & Sons, Inc., 1968.
- HARMAN, P.; KING, D. **Expert Systems Artificial Intelligence in Business**. Wile (NY), 1985.
- HARRISON, Thomas H. **Intranet Data Warehouse. Ferramentas e Técnicas para a Utilização do Data Warehouse na Intranet**. São Paulo : Berkeley Brasil ,1998.
- HENDRIX, G.; SACREDOTTI, F. Natural Language Processing: The Field in Perspective. In ANDRIOLE, S. (ed.). **Application of Artificial Intelligence**. Princeton (NJ) : Petrocelli Books, 1985.
- HJELMSLEV, Louis. **Prolegomena to a Theory of Language**. Univ. of Wisconsin Press, 1969.
- INMON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro : Editora Campus, 1997.
- JOHNSON, N.L. **Systems of Frequency Curves Generated by the Methods of Translation**, Biometrika, n. 36, p.149, 1949.
- JUNGDUCK, K.; COURTNEY, J. Applications of Knowledge Acquisition Techniques to Decision Support Systems. **Working paper, Business Analysis and Research, College of Business Administration** Texas : A&M University, 1987.
- KAY, A. **Computer Software**. Scientific American, 251, pp. 52-59, 1984.
- KEEN, P. **Competing in Time: Using Telecommunications to competitive advantage**. Ballinger Cambridge, Mass, 1988.
- KERSHBERG, I. Expert Database Systems. **Proceedings of the First International Conference on Expert Database Systems**. Kiawah Island, South Carolina, 1984.
- KIMBALL, Ralph. **The Data Warehouse Toolkit. Técnicas para Construção de Data Warehouses Dimensionais**. São Paulo : Makron Books, 1998a.
- KIMBALL, Ralph; REEVES, Laura; ROSS, Margy; THORNTHWAITE, Warren. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**. New York : John Wiley & Sons, 1998b.
- KING, D. Knowledge Acquisition in Intelligent Decision Support Systems : A Case for Natural Language. **Transactions of Seventh International Conference on Decision Support Systems**. San Francisco, pp.115-117, 1987b.

- _____. ERGO: An Explanation Facility for Decision Support Systems. **Proceedings of the 6th International Workshop on Expert Systems and Their Applications**. France, v. II, pp. 991-1011, 1986a.
- LEINWEBER, D. Finance. In BARTEE, T. (ed.). **Expert Systems and Artificial Intelligence**. Howard Sams and Company, Indiana-polis, Ind., pp. 33-60,1988.
- LEITE, J.C. **Modelos e Formalismos para a Engenharia Semiótica de Usuário**, 1998.
- LINDGARD, G. **Usability Testing and System Evaluation**. London (UK) : Chapman & Hall, 1994.
- MARTINS, I. H. **Um instrumento de Análise Semiótica para Linguagens Visuais de Interfaces**. PUC Rio, 1998.
- MICROSOFT CORPORATION. **The Windows Interface: Guidelines for Software Design**. Redmond : Microsoft Press, 1995.
- MORAN, T. The Command Language Grammars: a representation for the user interface of interactive computer systems. **International Journal of Man-machine Studies**, 1988.
- NORMAN, D. Cognitive Artifacts. In Carrol (ed.) **Designing Interaction : Psychology at the Human-Computer Interface**, 1986b.
- NORMAN, D. & Draper. **User Centered System Design**, Hillsdale (NJ) : Lawrence Erlbaum, 1986a.
- PEIRCE, Charles Sanders. **Semiótica**. The Collected Papers, nº 46, 2 ed. Editora Perspectiva, 1995.
- PREECE. **Human Computer Interaction**, 1994.
- PROAGRO. Ministério da Agricultura e do Abastecimento. 2000/2001, nº 05. Brasília, 2000.
- RIOS NETO, Atair. **A sobrevivência das empresas e a gestão por qualidade total** José dos Campos, 1993.
- ROCHA, Ana Regina Cavalcanti da. **Análise e Projeto Estruturado de Sistemas**. Rio de Janeiro : Editora Campus, 1987.
- ROCKHART, J.; D. DeLong. **Executive Support Systems**. Dow Jones-Irvin, Homewood, Ille,1988.
- ROTH, E., BENNETT, K., WOODS, D. **Human Interaction with Intelligent Machines**, R&D paper 87-1C6-HUSCI-P5. Pittsburg (PA) : Westinghouse R&D Center, 1987.

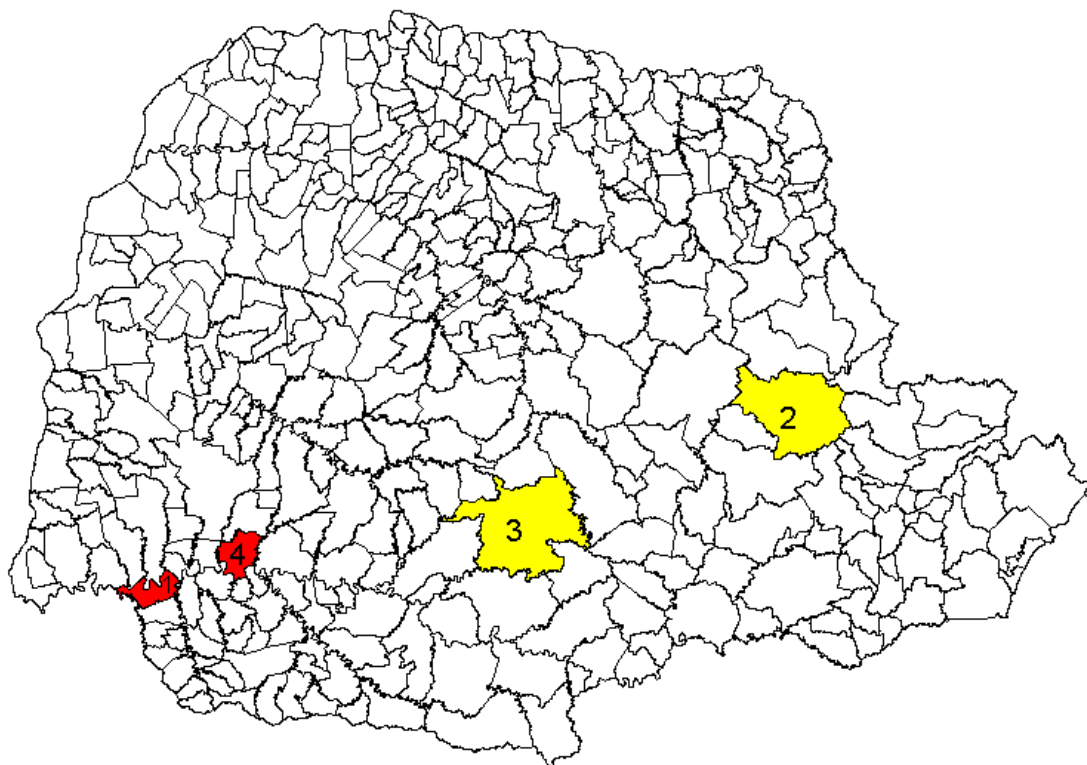
- SANTAELLA, Lúcia. O que é Semiótica. **Coleção Primeiros Passos**. Editora Brasiliense, 1998.
- SCANK, R.; REISBECK, C. **Inside Computer Understanding**, Hillsdale (NJ) ;: Lawrence Erlbaum 1981.
- SCHWARTZ, S. **Applied Natural Language Processing**. Princeton (NJ) : Petrocelli Books, 1988.
- SOUZA, C. S. de; LEITE, J. C.; PRATES, R.O.; BARBOSA, S.D.J. **Interação Humano-Computador: Perspectivas Cognitiva e Semiótica**. Rio de Janeiro, jul. 1999.
- SOWA, J.. **Conceptual Structures**. Addison – Wesley, Reading, Mass, 1984.
- STRAFACCI JR, Valter. **Qualidade Total aplicada ao gerenciamento de projeto**. Instituto Tecnológico de Aeronáutica (ITA- Brasil), 1996.
- UNIVERSIDADE DE MINAS GERAIS. <http://www.dcc.ufmg.br>
- UNIVERSIDADE DE SANTA CATARINA. <http://www.ufsc.br>
- WALTZ, D. State of the Art in Natural-Language Understanding. In LEHNERT, W. and RINGLE, M. (eds.) **Strategies for Natural Language Processing**. Hillsdale (NJ) : Lawrence Erlbaum Associates, pp.3-36, 1982.
- WARD , J.H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, 58 , 236-244, 1963.
- WINOGRAD, T. **Bringing Design to Software**, 1996.
- WOODS, W. Knowledge Representation. In BARTEE, T. (ed.). **Expert Systems and Artificial Intelligence**. Indianapolis, pp. 147-176, 1988b.
- _____. Knowledge Representation: What's Important About It? In CERCONE, N. and MCCALLA, G. (eds.). **The Knowledge Frontier**. New York, pp. 44-79, 1987a.

, aluno

de matrícula n ° 99902377 da Universidade Federal de Santa Catarina – UFSC.

SÓSTHENES RAMOS PRADO FILHO
Coordenador de Informática – COINF/MA
Ministério da Agricultura e do Abastecimento

Fases de Plantio
Paraná
Feijão 2000/2001
Situação na data: 20/01/2001

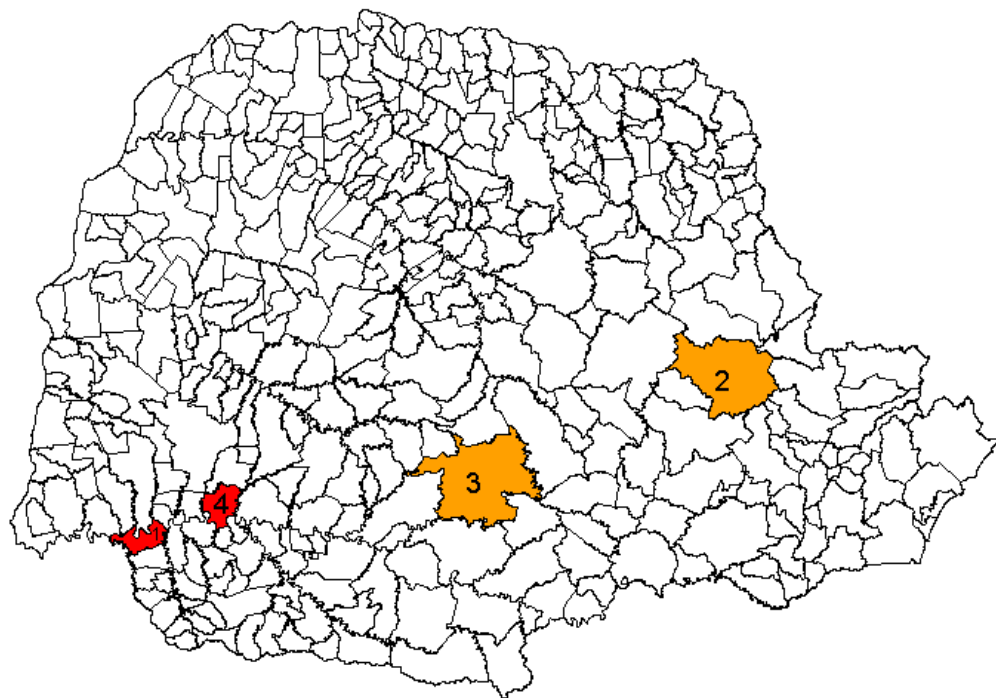


Fases
Emergência(07 dias)
Início Floração(40 dias)
Formação de Vagens(60 dias)
Final Maturação(85 dias)
Colheita(100 dias)
Sem Informações

Ciclo: Único

Legenda:
Nº - Município - Contratos
1 - CAPANEMA - 1
2 - CASTRO - 10
3 - GUARAPUAVA - 2
4 - TRES BARRAS DO PARANA - 12
Total:25 Contratos

Fases de Plantio
Paraná
Feijão 2000/2001
Situação na data: 20/02/2001



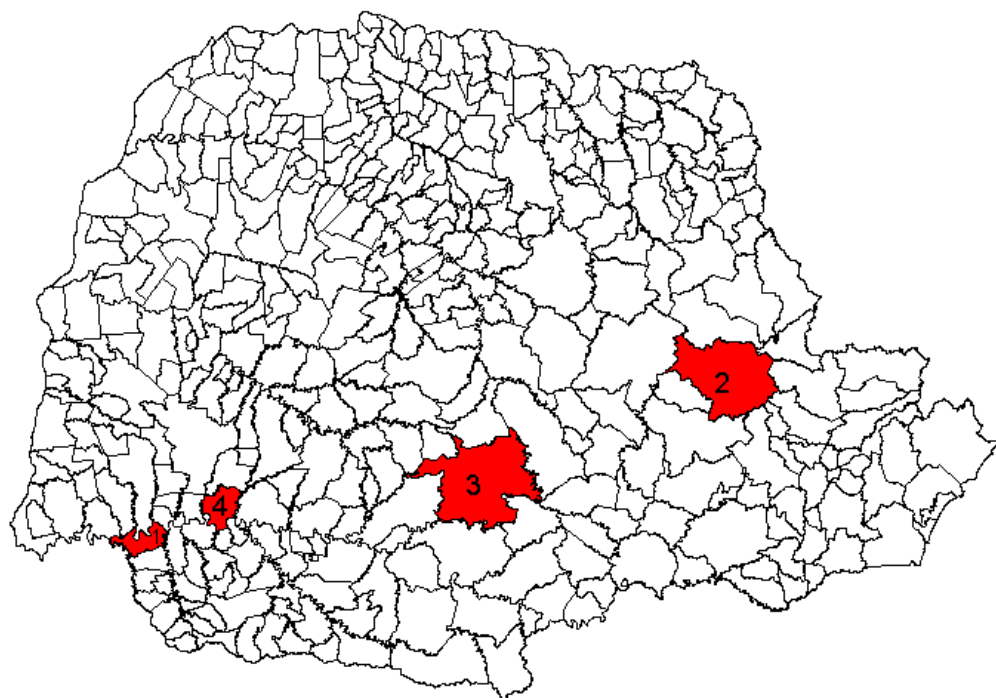
Fases

- Emergência(07 dias)
 - Início Floração(40 dias)
 - Formação de Vagens(60 dias)
 - Final Maturação(85 dias)
 - Colheita(100 dias)
 - Sem Informações
- Ciclo: Único







Legenda:

- Nº - Município - Contratos
- 1 - CAPANEMA - 1
 - 2 - CASTRO - 10
 - 3 - GUARAPUAVA - 2
 - 4 - TRES BARRAS DO PARANA - 12
- Total:25 Contratos

Fases de Plantio
Paraná
Feijão 2000/2001
Situação na data: 20/03/2001



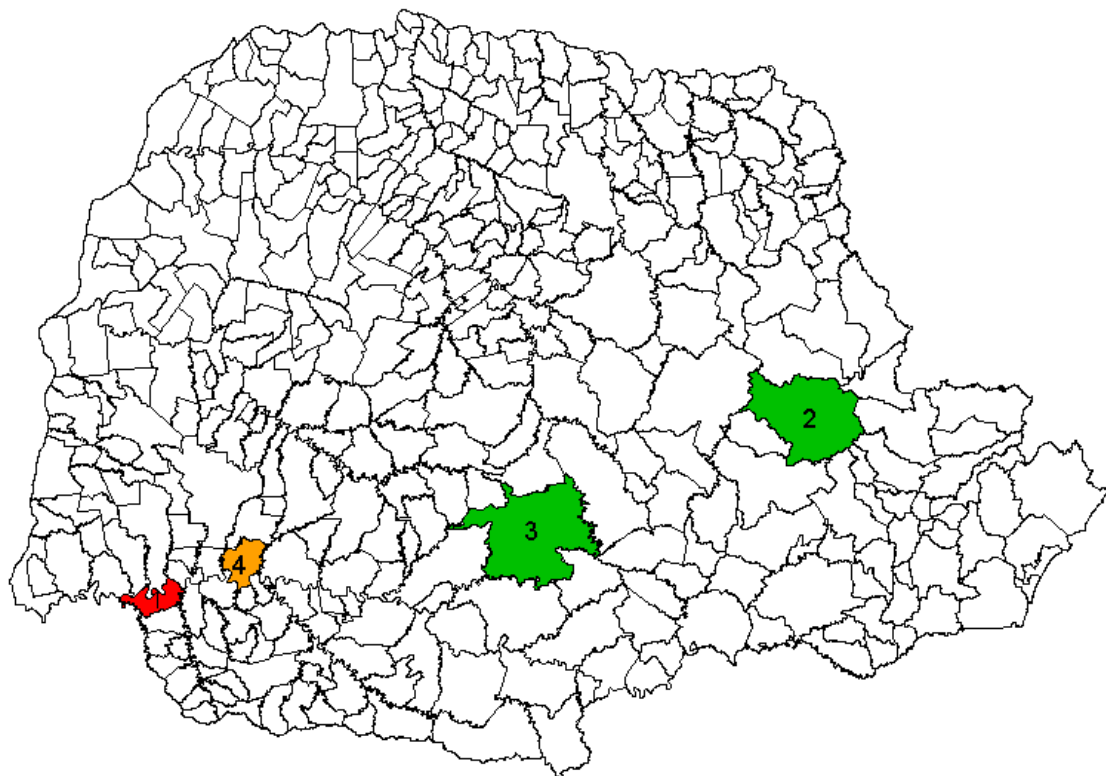
Fases

-  Emergência(07 dias)
-  Início Floração(40 dias)
-  Formação de Vagens(60 dias)
-  Final Maturação(85 dias)
-  Colheita(100 dias)
-  Sem Informações

Legenda:

- Nº - Município - Contratos
- 1 - CAPANEMA - 1
 - 2 - CASTRO - 10
 - 3 - GUARAPUAVA - 2
 - 4 - TRES BARRAS DO PARANA - 12
- Total:25 Contratos

Fases de Plantio
Paraná
Feijão 2000/2001
Situação na data: 20/12/2000



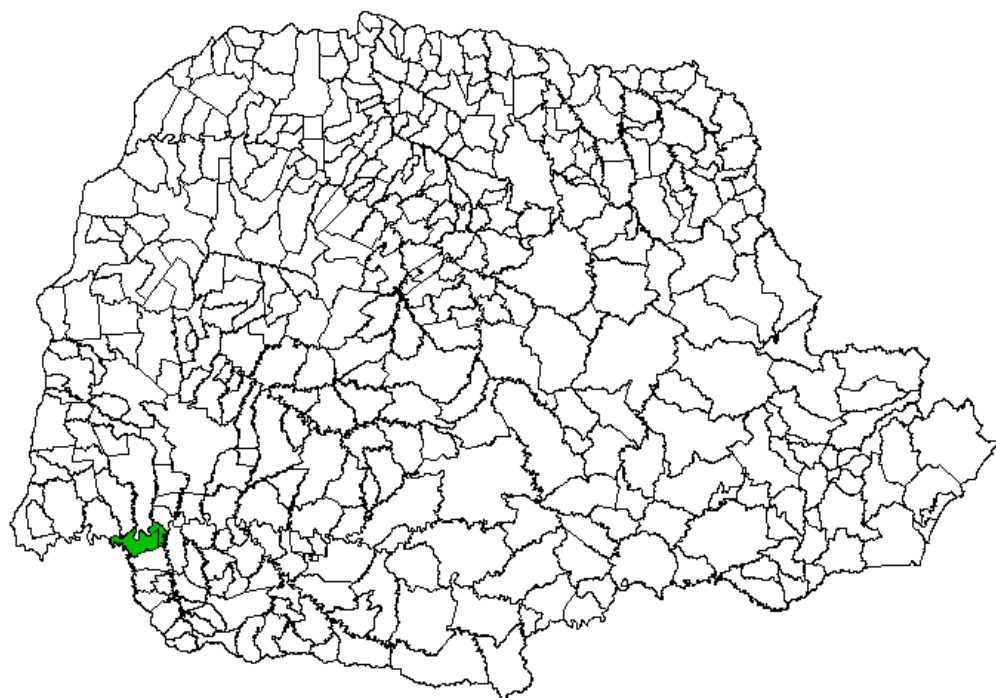
Fases

Emergência(07 dias)
Início Floração(40 dias)
Formação de Vagens(60 dias)
Final Maturação(85 dias)
Colheita(100 dias)
Sem Informações

Legenda:

Nº - Município - Contratos
1 - CAPANEMA - 1
2 - CASTRO - 10
3 - GUARAPUAVA - 2
4 - TRES BARRAS DO PARANA - 12
Total:25 Contratos

Fases de Plantio
Paraná
Feijão 2000/2001
Situação na data: 20/09/2000



Fases

- Emergência(07 dias)
- Início Floração(40 dias)
- Formação de Vagens(60 dias)
- Final Maturação(85 dias)
- Colheita(100 dias)
- Sem Informações

Legenda:

Nº - Município - Contratos
1 - CAPANEMA - 1
Total:1 Contratos