A Dissertation

Entitled

Towards a Chemical Taxonomy of Comets: Infrared Spectroscopic Methods for Quantitative Measurements of Cometary Water (With an Independent Chapter on Mars Polar Science)

by

Boncho P. Bonev

Submitted as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Physics

_____
Adviser: Philip B. James

_____
Adviser and external committee member: Michael J. Mumma

_____
Graduate School

The University of Toledo

December 2005

An Abstract of

Towards a Chemical Taxonomy of Comets: Infrared Spectroscopic Methods for Quantitative Measurements of Cometary Water (With an Independent Chapter on Mars Polar Science)

Boncho P. Bonev

Submitted as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Physics

The University of Toledo

December 2005

While comets are often grouped according to their orbital properties, it is increasingly clear that individual groupings likely contain objects formed in diverse regions of the proto-planetary disk. In recent years the emergence of accurate abundance measurements for volatile native constituents has begun to permit a taxonomic classification reflecting the chemical diversity in the composition of the cometary nucleus. The "baseline" of this taxonomy is the measured production of H_2O – the dominant volatile species in comets, which controls the sublimation of other molecules within 2-3 astronomical units from the Sun. Accurate quantitative production rates of

water are critical because they provide the "meter stick" for comparison with organic species; their simultaneous measurement eliminates many sources of systematic error.

This project explores two infrared spectroscopic methods to quantify H_2O production in comets. The first method is based on H_2O non-resonance fluorescence and is well established. The present work improves its application. The initial development of the second method, based on infrared emission from an excited dissociation fragment (OH), is the main original contribution of this work. In addition to obtaining H_2O production rates, the newly developed formalism provides insights about the process of H_2O photo-dissociation, in particular – the relative populations of vibrationally- and highly rotationally-excited states of OH produced by water photolysis.

The five principal contributions of this thesis are: (1) Development of a second method to measure H_2O production at infrared wavelengths, including evaluation of its potential limitations and definition of a future program to further validate it; (2) Observations of the end-product distribution of H_2O photolysis under natural conditions that differ greatly from studies done in terrestrial laboratories; (3) Measurement of H_2O spin temperatures in two comets with realistic treatment of uncertainties; (4) Derivation of reliable water production rates from directly sampled non-resonance fluorescent H_2O emission in two comets. These production rates provide a solid baseline for describing the volatile chemistry of these comets; and (5) Reliable measurements of the H_2O integrated column abundance in comet 9P/Tempel-1, during the observing campaign supporting the NASA's *Deep Impact* mission.

To the last person whom I said “Good Bye” before leaving for Graduate School across the Atlantic; to the one who really loved and understood this song:

Wishes come like a springtime warming.
You try to reach “Columbus’s” way …
Moments are gone like a day belonging
To the season of the snails …

And I, I am running to catch a train;
In my shoe I keep my key,
And running has been a game
But the ocean is for me!
And wonderin’ if I reach
the island of my sea
Will there be white peaks,
or just water rocks beneath?

There is no answer sheet,
There is no gist, no code
There are two stronger feet –
Dedicated to the ROAD!

(Boncho Bonev, 1997, “Roads” ©)

Acknowledgements

Completing this thesis has given me opportunity to meet, collaborate, informally communicate and exchange ideas with a large number of scientists. I would like to give credit to and to thank my mentors Prof. Philip James from the University of Toledo, and Dr. Michael Mumma, from NASA Goddard Space Flight Center; my committee member and informal adviser at a stage of this research Prof. Jon Bjorkman; my thesis Committee, consisting also of Prof. David Ellis, Prof. Larry Curtis, and Prof. Adolf Witt; the Goddard group of infrared spectroscopy of comets and planets lead by Michael Mumma and including people I am privileged to work with: Dr. Neil Dello Russo (who introduced me to spectroscopic data reduction), Dr. Michael DiSanti (among other things, we worked together on improving our algorithms for spectral analysis), Prof. Erika Gibb, Prof. Karen Magee-Sauer, Prof. Robert Novak, and Dr. Geronimo Villanueva; my collaborators in Mars polar science: Dr. Gary Hansen, Dr. David Glenar, and Dr. Mike Wolff; my former mentors: Dr. Boris Komitov in Stara Zagora and Plovdiv, Prof. John Laird and Prof. Dale Smith (who has supported “my cause” for fifteen years) at Bowling Green, and Prof. Armand Delsemme, who served on my Master's thesis committee; all Ritter astronomers; a number of scientists in the Solar System Exploration Division at NASA GSFC, including, but not limited to Dr. Joseph Nuth, Dr. William Maguire, Dr. Gordon Bjoraker, Dr. Michael Smith, Dr. John Pearl, Dr. Tilac Hewagama; the support astronomers at the W. M. Keck Observatory and at the NASA IRTF; my professors in

Toledo in Quantum Mechanics (Dr. Robert Deck) and Electrodynamics (Dr. David Ellis), and Dr. Rick Irving.

I would like to share the delight of being a small part of the community in the Ritter Astrophysical Research Center, along with fellow graduate students and friends like Jennifer Benson, David Horne, John Wisniewski, Uma Vijh. We went together though most of our courses, shared our research excitement, and I hope we will be able to maintain connection.

I thank Prof. Richard Ellis and his graduate student Daniel Stark for conducting the observations of comet C/2004 Q2 (Machholz); their observation produced a very valuable data set used in this thesis.

And most importantly – many thanks to my beloved family and to those dear friends from distant Stara Zagora to near-by Michigan, who have always believed in my dreams, who make me feel stronger during challenges, and because of whom I have never forgotten that there is much more in life than my favorite science! I know that remaining close to them, despite distance and hardship, is the largest accomplishment of my graduate student career.

Table of Contents

List of Tables

List of Figures

Chapter 1

Towards a Chemical Taxonomy of Comets

1.1 "Comets are Fascinating Objects …"

In 1998 I started my first semester in the Master's program of Bowling Green State University with an independent study course titled "Physics of comets". This was my "rookie" experience as a graduate student in the U.S. and I was very enthusiastic with the idea of taking a course suited entirely for my individual scientific interests. I had tried accomplishing this in my (otherwise very good) undergraduate program at Plovdiv University, but such an approach had been unheard of nationwide, so I failed to pass it through the faculty meeting, despite the strong support of the chairman (I passed a proposal for a program in "Astronomy" though, an accomplishment still remembered there). My professor in Bowling Green, Dr. Dale Smith, welcomed my suggestion to include, in addition to the conventional "textbook" topics, a seminar section in our course, discussing recent important contributions in the field. For that purpose Prof. Smith contacted his friend at NASA GSFC Dr. Dave Glenar who passed the request for advice to Dr. Michael Mumma, my thesis co-advisor to be. Mumma sent me several references of indeed interesting papers about X-rays in comets and (more importantly to

me) direct spectroscopic observations of parent volatiles in the near-infrared (NIR). I promptly thanked Mumma over the e-mail and I recall his simple reply:

"*You are welcome. Comets are fascinating objects, and scientifically interesting too.*"

Not completely original, but very well said, so I have remembered it. In fact one might hear this exact phrase referring to many astronomical objects and usually it relates to visual impressions. My fascination in comets, however, was deepened by, but did not originate from, the magnificent views of the elegant Hyakutake (the word "comet" has a female gender in Bulgarian) or the mighty Hale-Bopp, the first two I have seen with a naked eye. It came in the late eighties, after the International Halley Watch, well popularized among the public of my home town, Stara Zagora. I had the chance to join the "school" of Dr. Boris Komitov who along with doing "hard" science and teaching occasionally in college, has spent now 25 years of long hours (many of them underpaid or unpaid) in guiding high school students in their very first steps in research. He opened the world of comets to me, describing the processes of sublimation of ices and the exospheric expansion of the released gases in space. Under his mentorship I wrote my first computer program calculating the thermal balance on the surface of a cometary nucleus. It was based on the "friable sponge model" of Horanyi et al. (1984) which has described successfully the main observed types of asymmetry between the inbound (pre-perihelion) and outbound (post-perihelion) comet light curves. Horanyi et al.'s model explored the idea of Whipple (1950), elaborated by Mendis & Brin (1977; 1978) and Brin & Mendis (1979) about the development of a "mantle": a residual porous dust layer forming on top of a sublimating surface. This was the first time I heard about

sublimation (versus vaporization), about numerical methods and boundary value problems, and also the first time I looked through an original comet paper (Delsemme and Miller 1971, No. 3 from the classical series "Physico-chemical Phenomena in Comets"), although of course I could not fully comprehend it. Still, I was *fascinated* with the opportunity of conducting some research, stretching well beyond the limits of the school textbooks, and with comets along the way. Boris helped me realize a couple of key things beyond comets – the importance of appreciating (and generating) original ideas as the driving force in research, and the equal importance of having the mentality of a marathon runner in pursuing long-term (scientific) goals.

Of course before 1996 (Hyakutake), I envied those who had seen a bright comet with a naked eye. Views of Halley in 1910 were legendary, but also Comet West in the 1970s. So when this happened to me in the back-yard of Plovdiv University, I felt delighted. By that time I and Boris had adapted the aforementioned model to the conditions of Mars (gravity being the main difference) simulating the formation of a residual regolith layer on top a hypothesized surface of "dirty ice". The resulting configuration at mid- and high latitudes was an ice-poor upper surface layer, several dozens of cm thick (the thickness decreasing towards the poles), covering a water ice-rich layer. This was a chapter of my diploma work (senior thesis) in physics and astronomy. In 1997 I got a job as a physicist in the Institute of Astronomy in Sofia, greeting Hale-Bopp in the evenings on my way out of work. I also got some idea about photometric observations of radicals (CN, C_2, CO) and for the first time had somewhat regular access to major scientific journals (my diploma work was done in "blindness" about the developments in the field of planetary research).

In Bowling Green, I took two comet-related courses and worked on a thesis project with Prof. John Laird and Prof. Armand Delsemme, who introduced me into his ideas about the cometary origin of the biosphere. My Master's thesis was on a daughter molecule (CN) and I gave a talk on it at the "Asteroids, Comets, Meteors" meeting in Cornell, after which Dr. Mumma invited me to work with him. At that time, I had already joined those people whose primary interest in comets was pointing to the question: "*What is the chemical composition of the cometary nucleus?*", which leads me to the second part of Michael Mumma's simple reply in 1998:

1.2 "... And Scientifically Interesting Too."

In light of this thesis, there are two primary reasons to have a scientific interest in comets (and there are many more stretching to fields like aeronomy, heliophysics, etc.):

1. Comet nuclei have been considered "chemical fossils" – their primordial composition has been altered significantly less than that of planets since the time of their formation. Therefore the measured chemical composition of comets is a major observational constraint in cosmogony.
2. Comets might have played a (principal) role in the delivery of the building blocks of life, H_2O and pre-biotic organics, on Earth.

The following introductory sections elaborate these points in more detail, keeping the focus on placing the present project in the broad picture of planetary system studies.

1.3 Comet Nuclei. Present Day Reservoirs.

Cometary nuclei are typically km-sized bodies with irregular shapes and overall low albedos. The modern understanding about the nucleus structure is for a weakly bound agglomeration of smaller icy bodies, called *cometesimals* (e.g., Weissman, Asphaug, & Lowry 2004). As first outlined by Whipple (1950), the comet nucleus contains ice (mainly H_2O) and dust (both silicate and carbonaceous). Entering the inner Solar system initiates sublimation processes and the development of a gaseous coma (10^4-10^5 km in visible light), a dust tail, and an ion tail (millions of km long). Spectroscopic observations have revealed that comets contain, in addition to H_2O, biologically important molecules like HCN, H_2CO, C_2H_2, and other organics.

There are two known reservoirs of comet nuclei in the present day Solar system: the *Oort Cloud* and the *Kuiper belt* (referred in various sources as *Kuiper belt*, *Kuiper-Edgeworth belt, or Edgeworth-Kuiper belt*) (cf. Stern 2003; Gladman 2005). The Oort cloud is a milticomponent population of billions of nuclei at heliocentric distances from several thousand to tens of thousands of astronomical units (AU). Beyond about 10,000 AU the Oort cloud becomes roughly spherical. Gravitational effects of various kinds (e.g. galactic tides, random passages of near-by stars, or giant molecular clouds) could perturb individual comets with semimajor axis larger than ~20,000 AU towards the inner solar system. The Oort Cloud comets entering the inner Solar system are characterized by random orbital inclinations with respect to the ecliptic plane with either prograde or retrograde orbits.

The Kuiper belt represents a population in the range of ~30 to hundreds of AU concentrated close to the ecliptic. It is now realized that Pluto has been the first

discovered member of this population. The Kuiper belt is the source of short period comets, directed by collisions or gravitational perturbations towards the inner Solar system with prograde low-inclination orbits.

The understanding about the structure of the Kuiper belt is rapidly advancing. This part of the solar system consists of a main belt, a scattered disk, and an extended scattered disk (Gladman 2005). The main belt is the region just outside the orbits of the four giant planets. It includes comets with semimajor axis between about 35 and 56 AU from the Sun. The classical trans-neptunian objects (nearly circular orbits) and the plutino objects (orbits similar to Pluto's) belong to this group.

The scattered disk consists of comet nuclei of highly eccentric orbits. Their perihelia are near Neptune in the range of 30 to 38 AU, while their aphelia may be hundreds of astronomical units. The scattered disk objects are currently visible for a small fraction of their orbital period while near perihelion.

The extended scattered disk objects have perihelia larger than about 38 AU and (similar to the scattered disk) highly eccentric orbits. Bodies with semimajor axis of about 1000-2000 AU merge into the inner Oort cloud. In difference to the scattered disk population, the orbits in the extended disk are highly stable against close encounters with Neptune over the age of the solar system.

It should be pointed out that among the three populations (main belt, scattered disk and extended scattered disk), the scattered Kuiper disk is the most likely source for most of the active comets injected from the Kuiper belt into the inner solar system (Bernstein et al. 2004).

1.4 Origin of Present Day Reservoirs

It is important to emphasize that the Oort cloud comets formed in a very distant region from their present reservoir. The comet nuclei were among the first to accrete in the early solar nebula – accumulation of silicate, carbonaceous, and icy grains led to the formation of cometesimals, many of which, through collisions, agglomerated in larger bodies (up to several km in size). At heliocentric distance range of 5-35 AU the growing giant planets scattered these "icy planetesimals" to unbound orbits outside the young solar system or to distant orbits in the scattered disk. Galactic tides and/or passing stars (from the denser stellar environment in the cluster where the Sun had formed) have subsequently dispersed a number of objects into the Oort cloud resulting in a spherical distribution outside ~10,000 AU (Gladman 2005).

The Kuiper belt consists of objects formed in situ and of population scattered from the giant planets region prior to and during the Oort cloud build up. The details in creating the multicomponent structure of the Kuiper belt is still not well understood. Current models consider large scale migration of the giant planets (e.g. Levison & Morbidelli 2003, Gomes et al. 2005) and/or the presence of other rogue planetary objects in the early outer solar system (e.g. Petit, Morbidelli, & Valsecchi 1999; Brunini & Melita 2002) as possibly important factors shaping the architecture of the Kuiper belt.

The present day orbital distribution of comet nuclei in their reservoirs provides important clues about giant planet formation and the processes that have "shaped" the outer solar system, including the outward migration of icy planetesimals into the Kuiper belt.

1.5 Cometary Bombardment of the Young Earth

A population of comets formed in the giant planet region has been scattered inward – towards the terrestrial planets. This population might have a significant contribution to the delivery of water and pre-biotic organics to the young Earth. This hypothesis for the *cometary origin of the biosphere* has been discussed by various researchers, but promoted consistently through the years by Delsemme (e.g. 1981; 1989; 1998; 2000). This idea has been a subject of much debate. Delsemme argued that the bulk of the terrestrial water has been delivered by comets formed near Jupiter. Their deuterium enrichment was diminished by steam coming from the inner Solar system and condensing in the Jupiter region into icy grains that formed comet material. Morbidelli et al. (2000), however, concluded that comets from the Jupiter region provided only a very minor contribution at a stage when the growing Earth was approximately half its present size. This work argues that at later stages comets from the trans-Uranian region contributed about 10%, while most of the terrestrial water was delivered by a few planetary embryos formed in the outer asteroid belt.

A comet classification (taxonomy) based on chemical abundances of comet constituents and cosmogonic indicators, like isotopic ratios and (possibly) spin temperatures, has the potential to provide important observational constraints for models describing the conditions in the forming planetary system, as well as different scenarios for the origin of the biosphere.

1.6 Comet Taxonomies

Historically, comets have been grouped according to their orbital properties (e.g. Lardner 1853). Such classification is used to recognize the present day reservoir from which an individual comet came. The modern dynamical taxonomy is based on the Tisserand parameter with respect to Jupiter (T_J, Levison 1996):

$$T_J = \frac{a_J}{a} + 2\left[(1-e^2)\frac{a}{a_J}\right]^{1/2}\cos(i)$$

where a_J is semimajor axis of Jupiter, while a, e, and i are respectively the semimajor axis, eccentricity, and orbital inclination of the comet. T_J is a measure of the influence of the most massive giant planet on the dynamics of the comets. This parameter is approximation of the Jacoby constant, which is an integral of the motion in the circular restricted three-body (Sun – Jupiter – comet) problem. It is also related to the relative velocity between the comet and Jupiter during close encounters:

$$v_{rel} \propto (3 - T_J)^{1/2}$$

Oort cloud comets have values less than two for the dynamical Tisserand parameter with respect to Jupiter. This group includes Halley-type comets, returning long-period comets, and dynamically "new" comets, all belonging to the modern class of *"nearly-isotropic"* comets (Levison 1996; Dones et al. 2004).

On the contrary, the Kuiper belt is the main source for *"ecliptic comets"* characterized by Tisserand parameters exceeding two. It is divided into three sub-classes: *Jupiter-family comets* ($2 < T_J < 3$), *Encke-type comets* ($T_J > 3$, orbits interior to Jupiter), and *Chiron-type comets* ($T_J > 3$, orbits exterior to Jupiter). Hereafter, the term "ecliptic comets" will refer mostly to the first two sub-groups. Note that at present the

term “Jupiter-family” is still sometimes used for both Jupiter-family and Encke-type comets. Also the term “Oort cloud” is often used to indicate the dynamical class of near-isotropic comets.

A more detailed review of the modern dynamical classification is beyond the scope of this work. The background provided so far is adequate to emphasize the gradually growing observational evidence, which is at the heart of the subsequent discussion: comets with similar dynamical properties display diversity with respect to other criteria, in particular the measured abundances of some species (cf. Bockelee-Morvan et al. 2004). This fact justifies the importance of an additional comet taxonomy based on chemistry. Such classifications have initially been conducted at optical wavelengths (cf. A’Hearn et al. 1995; Fink & Hicks 1996). A’Hearn et al.’s analysis indicates that most comets originating in the Kuiper belt appear to be depleted in carbon-chain molecules leading to the observed C_2 and C_3 radicals. The main difficulty in interpreting optical data, however, is that the sampled species are not parent molecules (directly released from nucleus), but daughter fragments (mainly dissociation products). For example, despite the extensive data set of CN, C_2, and C_3 observations, the identity and source (from ice or dust) of their native precursors is still uncertain (cf. Feldman, Cochran, & Combi 2004). Moreover, the principal precursors may in fact differ among comets.

In recent years, the emergence of accurate abundance measurements (from multiple spectral regions) for the dominant volatile (H_2O) and a number of other parent species (HCN, CH_3CN, CH_3OH, CO, CH_4, C_2H_2, C_2H_6, etc.) has initiated a new taxonomic classification based on chemistry. Radio and infrared (IR) observing facilities

play (equally important) major roles in this process since most of the parent molecules emit via rotational and rotational-vibrational transitions, observed respectively at sub-mm/mm and near-infrared (2-5 μm) wavelengths. While this project is focused on IR observations, it should be pointed out that studies in the two spectral regions are complementary and should ultimately lead to a unified cometary classification. Many complex molecules (e.g. ethylene glycol, Crovisier et al. 2004) are more easily detected in the radio. On the other hand, symmetric hydrocarbons do not have rotational transitions, so they are uniquely sampled in the NIR. The sole proven method for directly measuring cometary H_2O from the ground is based on ro-vibrational transitions (c.f. Dello Russo et al. 2005), while pure rotational transitions of water are more commonly sampled by satellite observations (e.g. Lecacheux et al. 2003; Bensch & Bergin 2004). Interpreting the latter is dependent on retrieved H_2O rotational temperatures from NIR data (Frank Bensch, private communication). Another aspect of the complementary nature of radio and NIR observations is that the former are spatially integrated over a large extent of the coma, while the latter are concentrated in the first several hundred to a few thousand km from the nucleus and are spatially resolved.

While the samples of studied comets are still small (especially in the NIR), evidence for chemical diversity is clearly present (Biver et al. 2002, Mumma et al. 2003, Bockelee-Morvan et al. 2004). This has not been a surprise. First, dynamical models suggest that roughly half of the comets remaining in the Oort cloud today originated in the 5-30 AU region, while the remainder formed beyond 30 AU (Dones et al. 2000). As thermo-chemical conditions varied strongly within the protoplanetary disk, chemical differences are to be expected even in a small sample of Oort cloud comets. Second,

comet nuclei might contain cometesimals that migrated from different regions of the forming planetary system, which implies a possible chemical heterogeneity even within a single nucleus.

1.7 The Building Blocks of the New Chemical Taxonomy

What are the principal measurements that will establish the foundation of a chemical taxonomy of comets? Three groups of "observables" can be outlined:

1. Chemical Abundances of different parent molecules.
2. Isotopic enrichments (D/H).
3. Ortho-to-para ratios.

These parameters depend on distance from the young Sun. Comet material that formed beyond 40 AU is expected to display chemistry more characteristic of the natal interstellar cloud core. Conversely, the abundances and isotopic enrichments in comets ejected from the giant planets region would (presumably!) reflect the processing history of ices in the protoplanetary disk. Once established, the listed parameters should be related, via dynamical models, to the present reservoirs of the observed comets and ultimately to their formative regions. Ideally, we would like to provide constraints for modeling the chemical evolution from the natal molecular cloud core to the protoplanetary disk and the young planetary system (e.g. Markwick & Charnley 2003). Such models rely on cometary data.

Deuterium enrichment is an important parameter for testing the origin of terrestrial water. The very few available estimates of D/H in comets exceed the enrichment of terrestrial water by a factor of two. However these results are not

conclusive, because the sampled comets (Halley, Hyakutake, and Hale-Bopp) are not necessarily representative of the population mainly responsible for the presumed cometary bombardment of Earth. Delsemme's argument is that the aforementioned comets formed in the Uranus-Neptune region, and so are expected to have D/H ratios closer to the higher interstellar values. On the other hand, objects formed near Jupiter are expected to have lower D/H values, compatible to that of present day terrestrial water. A similar argument, however, applies to the hydrated planet embryos formed in the outer asteroid belt, considered major water deliverers by Morbidelli et al. (2000). In these scenarios the pending data base of observed isotopic ratios in comets would provide valuable measurements that will help in evaluating the different hypotheses.

1.8 Some Challenges

The ambitious goals stated above immediately raise "a few" critical issues, including (but not limited to):

1. To what extent is the comet material really pristine?

2. Given a set of measured cosmogonic "observables", how precisely can they be related to a given location in the proto-planetary disk? In that regard, to what extent can we account for the expected radial migration of cometesimals and the outward transport of gas and dust from the regions of terrestrial planets? Do we expect sub-populations from the Oort cloud and from the Kuiper belt to have formed in similar regions in the proto-planetary disk, which would imply a similar chemistry?

3. Are the deduced ortho-to-para ratios cosmogonic? How well constrained are the conversion times between different spin states?

4. What effects could produce a difference between the abundances of cometary volatiles observed in the coma and their actual abundances in cometary ices? The major candidates considered have been sublimation fractionation, development of extended sources in the coma, and chemical reactions (Irvine et al. 2000).

5. How important are observational selection effects between Oort cloud and Jupiter family (or Encke-type) comets?

6. To what extent are the chemical abundances from different spectral regions (mainly radio vs. IR) directly comparable?

7. What is the level of synergy between volatile studies and studies of cometary dust? A composition taxonomy should by all means include both volatile chemistry and dust properties.

We will address some of these issues later in the context of particular thesis results. For the moment, we will only comment on the first one, which is very basic and probably the most difficult to resolve.

1.9 How Pristine is the Comet Material?

While arguably considered to be the most primitive bodies known, there is increasing evidence that comet nuclei have been modified by various processes: irradiation by solar and galactic cosmic rays; heating by stars passing through the Oort cloud and by supernovae explosions; accretion of and erosion by interstellar grains; structural changes from amorphous to crystalline water ice as the nucleus is heated above ~120-150K for the first time; sublimation of volatiles in the inner Solar system with

development of a residual dust "mantle"; and collisional evolution within their present reservoirs and before ejection into their present reservoirs (cf. Stern 2003, Weissman et al. 2004).

Except for the last mechanism, these processes are believed to affect at most a thin layer of the nucleus surface. This chapter is being written as the results from our spectroscopic campaign with the Keck II telescope, in support of the Deep Impact mission (A'Hearn et al. 2005) are still "hot from the oven". The major purpose of this Discovery-class space mission was to excavate more pristine material from the interior of a Jupiter family comet. Results from this new work have been included in this thesis.

In recent years, serious arguments have appeared that the initial population of comets in the giant-planet zone experienced severe collisional evolution prior to their ejection to the Oort Cloud. Stern & Weissman (2001) argued that this collisional evolution has been "catastrophic" for much of the initial population. In other dynamical models the collisional effects are not so severe, but there is a consensus that the view of comet nuclei as unprocessed aggregates from the primordial solar nebula needs serious revision or "complete reversal" (Wessman et al. 2004).

In regard to the chemical taxonomy being built, the observed composition of comets should be considered a "window" to the conditions in the early solar system and to the processing history of the cometary ices.

Chapter 2

The "Meter Stick" for Chemical Taxonomy

2.1 The H_2O Paradigm

This project explores two IR spectroscopic methods to quantify H_2O production in comets. The first method is based on H_2O non-resonant fluorescence and is well established. The present work improves its application and reports rotational temperatures, ortho-to-para ratios, and H_2O production rates in two comets. The initial development of the second method, based on IR emission from an excited dissociation fragment (OH), is the main original contribution of this work. In addition to obtaining H_2O production rates, the newly developed formalism provides insights about the process of H_2O photo-dissociation, in particular – the relative populations of vibrationally and highly rotationally excited states of OH produced by water photolysis.

Water in comets presents a paradigm for the release of parent volatiles. Studies of H_2O sublimation from the nucleus and subsequent destruction (primarily via photo-dissociation) are of primary interest in cometary science. The principal reason is that H_2O is (by far) the dominant parent volatile species in comets, and its sublimation controls the release of other molecules within 3-4 AU from the Sun.

H_2O had been considered a potential parent (this term had already been introduced at the time) constituent as early as the 1940s, shortly after the discovery of the OH 3090 Å UV emission (see the historical review of Festou, Keller, & Weaver 2004). In response to Whipple's revolutionary icy-conglomerate model (1950), Delsemme & Swing's (1952) hypothesis for clathrate hydrates of H_2O as the chemical form of nuclear ices implied that water molecules are very abundant in the native cometary material – a concept neither proven, nor discussed in much detail prior to that publication. Following decades of growing indirect observational evidence favoring H_2O as the major volatile, this molecule was detected directly for the first time from the Kuiper Airborne Observatory in abundances exceeding significantly those of all other species sampled (by various ground-based observatories and space missions) in Comet Halley (Mumma et al. 1986).

Studies of cometary water relate directly to each of the three taxonomy elements outlined in the previous chapter.

2.2 Mixing Ratios – the First Building Block of a Chemical Taxonomy

Establishing a chemical taxonomy of comets requires development of reliable methods for quantifying the H_2O production rate. The latter is the "baseline" or the "meter stick" of the new taxonomy. It was pointed out that the chemical abundance of given species is the first taxonomy "building block". This abundance is commonly expressed in terms of the *mixing ratio* (by number) between that species and H_2O:

$$MR(X) = \frac{Q(X)}{Q(H_2O)} \times 100 \quad [\%]$$

where Q(X) and Q(H_2O) are respectively the production rate of the species X and of H_2O, in units of [molecules sec^{-1}].

Behind this choice of a "baseline" is the nature of the cometary nucleus whose dominant volatile constituent is H_2O. This dominance is a common factor among active comets, while the relative abundances of all other parent volatiles can vary (and that's the point for a chemical taxonomy!). Moreover, relating the derived production rates to that of H_2O provides a measure which is independent of the comet's overall activity (to the extent we can detect given species) because the latter is controlled by water (at least, within 3 AU of the sun). The "raw" gas production rates depend on a number of factors, which are <u>not</u> cosmogonic. To list a few: heliocentric distance (the dependencies are often not entirely insolation-driven, i.e. described by the inverse square law); pre- or post-perihelion orbital stage; dynamical class (e.g. dynamically new comets vs. Jupiter family comets); and the fractional active surface of the nucleus. On the other hand, the mixing ratios of different species with H_2O likely encode information about the formation temperatures and processing histories of cometary ices (Mumma, Weissman, & Stern 1993).

Biver el al. (2002) and Mumma et al. (2003) have summarized evidence for chemical diversity among comets sampled respectively at radio and IR wavelengths. Both works used the H_2O production rates as a baseline. Biver et al. also introduced the production rate of HCN as a baseline, because it is the easiest comet molecule to observe in the radio. However, they first firmly established that the ratios between HCN and contemporary measured H_2O production rates varied only weakly in their sample.

2.3 The Importance to Quantify H_2O Production in the IR

Water production rates can be derived by analyzing emissions from dissociation fragments (H, OH, O) sampled in multiple spectral regions (UV, optical, radio, and - with this project - NIR). As already mentioned, H_2O production has also been measured directly in the radio (mainly satellite observations) and in the NIR from the ground. All these approaches are important because they permit comparison of trace molecular detections with contemporary measured H_2O production rates. Indeed, comets are time varying sources, so it is critical to quantify water production at the same time as the production of trace constituents. This approach is desirable, but still risky. The reason is that there are systematic differences between production rates of some parent molecules determined from different spectral regions (and different instruments, telescopes, etc.). These differences impose problems that are not completely understood. Inter-comparisons between production rates obtained from different techniques always require an extra caution. Agreement between absolute production rates may in some cases be more coincidental than fundamental because the various methods are subject to widely different observational approaches, systematic effects (not always understood), modeling assumptions and complexity, and spatial extent of the sampled cometary coma.

The most reliable mixing ratios are obtained when H_2O and other parent volatiles are observed simultaneously within the same instrument setting. Because a number of species are detected in the NIR, it is critical to develop methods for extracting water production rates from that particular spectral region. Therefore it is desirable to "cover"

H_2O and other molecules of interest simultaneously for as many frequency intervals in the IR as possible.

2.4 Isotopic Ratios Based on Direct Observations of H_2O and HDO

The possibility to measure D/H ratios (through HDO/H_2O) adds another dimension to the importance of water studies. As mentioned in Chapter 1, D/H has been measured in only three comets, likely formed in the Uranus-Neptune region. It is desirable not only to increase the overall sample of comets with measured D/H but also to test the deuterium enrichment of *Jovian comets* (not to be confused with Jupiter family comets – *Jovian comets* being the ones formed 5-10 AU from the young Sun). Multiple HDO and H_2O lines are detectable in the NIR. Mumma et al. (2001b) sampled the overall chemistry of one possible candidate of the Jovian group (C/1999 S4 [Linear]), which was however too weak to allow a sensitive search for HDO. Gibb et al. (2002) extended this search to several other comets, but were unable to report a secure detection of HDO. A comet with high enough gas productivity to securely detect HDO with the current high-resolution ground-based spectrometers has yet to be observed. However, the potential is there, and the methodology for H_2O analysis (presented and further developed in this thesis) is also applicable for HDO.

2.5 H_2O Spin Temperature – a Cosmogonic Indicator?

The H_2O spin temperature is another possible cosmogonic indicator (cf. Mumma et al. 1989, 1993). The H_2O molecule is organized into two distinct spin species according to whether the nuclear spin vectors of the hydrogen atoms are parallel (ortho ladder) or anti-parallel (para ladder). The lowest para level lies 23.8 cm-1 (~34 K) below the lowest ortho level, so the ortho-to-para ratio (OPR) is temperature dependent. Therefore para states are increasingly favored at low temperatures of H_2O molecule's formation, while for temperatures above ~50 K the OPR reaches the equilibrium value of 3/1.

A reliable nuclear spin conversion model for H_2O has not been developed yet, and no transition has been observed between the two nuclear spin species. Conversion among different spin species in the gas phase by radiative or collisional processes is forbidden (cf. Mumma et al. 1993, Dello Russo et al. 2005). Recent studies suggest that the nuclear spin conversion times of H_2O are more analogous to the long conversion times of H_2 than to species like CH_4 for which faster conversion can occur (Miani & Tennyson 2004). This theoretical work evaluates the possibility to observe strongly-forbidden rotational-vibrational transitions between ortho- and para levels of H_2O. These authors argue that a successful completion of such experiments would enable the eventual development of a reliable nuclear spin conversion model of H_2O.

Conversions are expected to be very slow in the solid phase too, although various proton-exchange mechanisms should be considered in that case (Bockelee-Morvan et al. 2004).

Based on the best knowledge to date, the H_2O molecules do not undergo spin conversion during their long residence in the comet nucleus or after sublimation in the coma prior dissociation. Under this condition, the nuclear spin temperatures, derived from OPRs in the coma, could be a measure of the chemical formation temperature of the cometary ices.

2.6 Synopsis of the Core Chapters

The following four chapters present the data, the methodology, and the main results of this observational project, the gist of which was given in the beginning of Chapter 2. I will first introduce in Chapter 3 the high-resolution spectroscopic data used. Emphasis is made on the simultaneous detection of H_2O, multiple other parent species, and OH. Chapter 4 is dedicated to the new results from the analysis of H_2O non-resonant fluorescence emission in two Oort-Cloud comets and (as a bonus) one Jupiter-family comet, focusing on the uncertainties in the derived parameters and their physical insight. As already pointed out, this chapter presents an improved implementation of a previously developed method (cf. Dello Russo et al. 2004; 2005). Finally, Chapter 5 and Chapter 6 describe the development of a new method for quantifying H_2O production in comets, the importance and the uncertainties in this approach, and its application to studies of H_2O photo-dissociation, in particular – the production of vibrationally and rotationally-excited OH states as a result of this process.

Chapter 3

High-Resolution Spectroscopy of Comets using NIRSPEC/Keck 2

3.1 Direct Detections of Parent Volatiles in the IR: KAO, CSHELL

The first secure detections of a cometary parent volatile in the IR were achieved in 1985-1986. NASA's Kuiper Airborne Observatory acquired high-resolution spectra of H_2O emission seen mostly in fundamental transitions (Mumma et al. 1986). These unambiguous direct detections of multiple emission lines of the dominant parent volatile were a primary achievement. However, the observational technique (Fourier spectroscopy) did not provide enough sensitivity to detect trace constituents (e.g., Drapatz, Larson, & Davis 1987; Larson et al, 1989).

By the late-1980s observational searches for native constituents had received solid theoretical support. Modeling the radiative pumping (vibrational excitation) by direct solar radiation (cf. Crovisier & Encrenaz 1983; Weaver & Mumma 1984) predicted the wavelength range to sample parent volatiles in the IR: roughly between 2 and 5 μm, where the majority of species have their rotational-vibrational transitions. It was also

realized that the most promising technology to achieve the desired detections would be the echelle grating spectroscopy.

There are three basic features desired for an instrument to conduct gas-phase studies in the IR: high spectral resolving power, sensitivity, and large spectral range that can be sampled within a single instrument setting. These are the directions for improving technology. The trend in modern spectrometers emphasized 2-D detector arrays with increasing size, one dimension corresponding to spectral, the other to spatial information. The latter is obtained via *long-slit spectroscopy*, a classic observing technique in which the intensity of an extended object (like a comet) can be measured along the slit mounted in front of the detector.

The first spectrometer to combine high spectral resolution ($\lambda/\Delta\lambda \sim 25{,}000$ with 1″ wide slit) with a large format 2-D detector array was CSHELL at the NASA Infrared Telescope Facility (IRTF) (Greene, Tokunaga, & Carr 1994). This instrument "revolutionized the field". It lead to the first detection of H_2O from the ground (Mumma et al. 1995), to the discovery of three "radio-invisible" hydrocarbon molecules (CH_4, C_2H_6, Mumma et al. (1996); C_2H_2, Brook et al. [1996]), and to secure detections of CO (previously detected in the UV and tentatively in the IR), HCN (first detected in the radio), and CH_3OH (previously detected in the radio and tentatively in the IR).

CSHELL has an InSb 256x256 detector array allowing extraction of spatial information from extended objects but suffers from a very limited free spectral range, within which just a few (at best) species can be detected simultaneously. The next generation instruments have used larger arrays and utilized *cross-dispersed* optical

designs, thereby increasing tremendously the frequency intervals that can be sampled within a given instrument setting.

3.2 Introduction to NIRSPEC Data

The Near Infrared Spectrometer (NIRSPEC; McLean et al. 1998) is installed at the Keck 2 telescope at the W. M. Keck Observatory atop Mauna Kea, HI, and was commissioned in 1999. NIRSPEC has a 1024x1024 InSb detector array and is cross-dispersed ($\lambda/\Delta\lambda \sim 25{,}000$ with 0.43″ wide entrance slit). All data analyzed in this thesis were acquired with this instrument. Appendix 1 compares between CSHELL/IRTF and NIRSPEC/Keck 2 emphasizing the advantages and disadvantages of using both instruments/facilities. It should be mentioned that although NIRSPEC is the current state-of-the art instrument, the technology is constantly evolving – next generation detectors with larger arrays and/or higher resolving power are being developed. (Needless to say, the analysis methods used in this project are not "tied" to a particular instrument; they are motivated by science, not by technology).

Figure 1 shows a typical NIRSPEC *echellogram*. These are "raw" data from comet C/2004 Q2 (Machholz) taken on Jan 19, 2005 (universal time [UT] will be used in all dates/times reported). Note the six *echelle orders* (21-26) separated by the low-resolution spectrograph (cross-disperser) operating perpendicular to the echelle dispersion. The "horizontal" dimension of each order is the spectral one, the "vertical" dimension is the spatial one and it corresponds to the entrance slit length (24″ here). Note that moving a point source along the slit would cause its spectrum to shift along the spatial dimension of a given order.

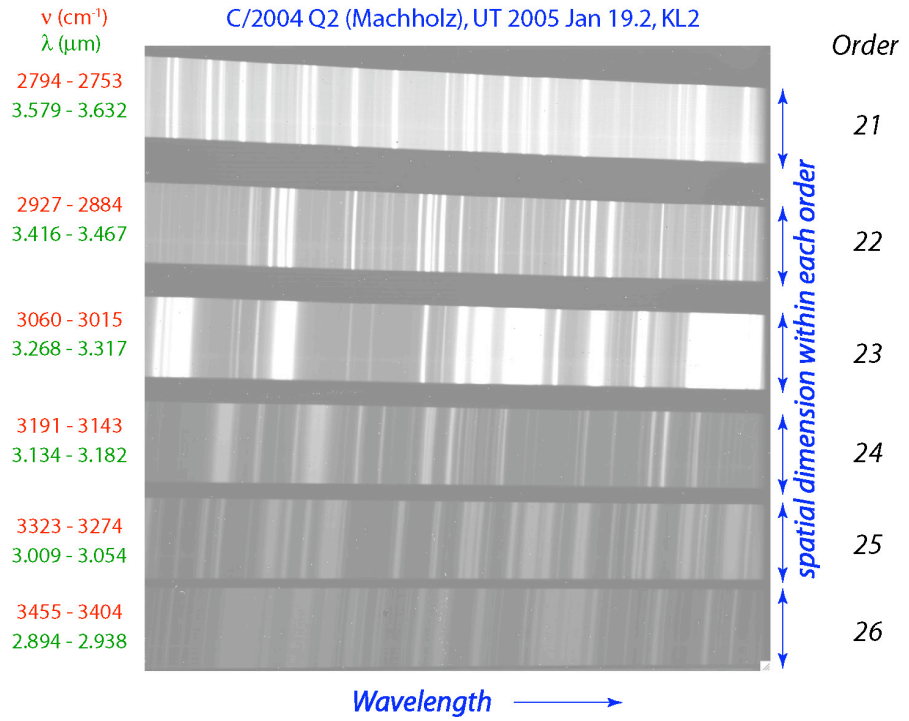

Fig. 1 – "Raw" NIRSPEC echellogram: High-resolution ($\lambda/\delta\lambda \sim 25{,}000$) echelle spectrograph disperses the signal in the horizontal dimension (spectral dimension). A low resolution cross-disperser operating perpendicular to the echelle separates the echelle orders. The frequency ranges for each order are given. The vertical dimension within a given order is the spatial one, corresponding to the length of the entrance slit. The slit used is 24″ long and 0.43″ wide. Note the change of the background levels from order to order.

Where is the comet signature in Figure 1? It is "hidden" by the dominant signal of *background thermal continuum and superimposed sky emission*. Atmospheric sky emission lines of various telluric species (H_2O, CH_4, N_2O, etc.) are clearly seen. This is the first aspect of the data to understand: except for a very bright object (like the mighty Hale-Bopp), the comet signal is weaker than the background. For this reason the dominant source of *photon noise* is the thermal background noise and it is largest at the position of sky emission lines, which constitute a spatially uniform source. It is desirable that the data analyzed be photon noise-limited (versus read noise or other instrumental or modeling effects).

3.3 Revealing the Comet Signal

The comet data are revealed after subtracting two "raw" frames like that of Figure 1. This procedure (ideally) cancels the thermal sky and telescope emission. The residual signal then includes:

1. Cometary continuum dust emission.
2. Cometary molecular emission lines.
3. Telluric absorption lines.

These elements are shown in Figure 2. The comet dust continuum is represented by the pairs of bright and dark strips (A- and B-beam) along the spectral dimension of each order. The B-beam appears negative in the difference frame. Comet emission lines are seen perpendicular to the continuum. Their intensity falls away from the beam centers. The terrestrial atmospheric absorption lines are seen against the comet continuum.

The A- and the B-beams correspond to two nodding positions of the telescope (Figure 3). The object is first centered in the "bottom" (as seen in the figure) half of the slit and the spectrum taken is called an "A-frame". Then the telescope is moved so that the comet is centered in the "top" part of the slit resulting in a "B-frame". To maximize observing efficiency, the nodding sequence is A, B, B, A. Figure 2 shows an A-B difference (more precisely the difference between the means of four A- and of four B-frames). Upon difference, most of the sky emission cancels (residual "sky" is taken out further in the analysis). The comet signal does not cancel because the comet has been moved along the slit (i.e. from the "bottom" to the "top" half of any order shown on Figure 2). Instead, the B-beam appears negative in the A-B frame.

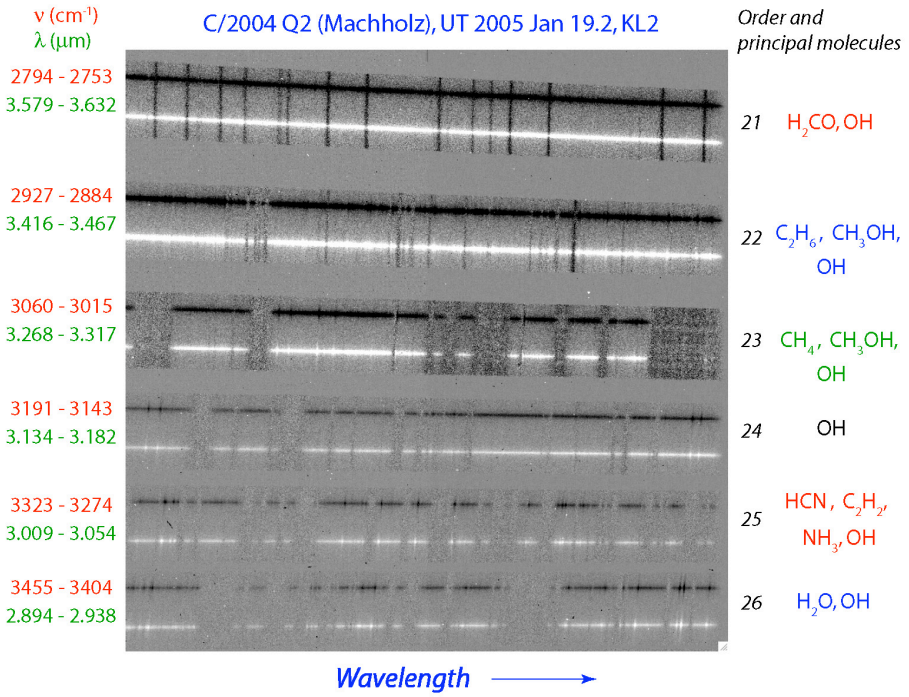

Fig. 2 – A-B echellogram revealing cometary signal: All species indicated are detected simultaneously, which eliminates important systematic effects in the subsequent analysis. Note that the OH lines are spread over multiple frequencies. Cometary emission lines are easier to distinguish against the comet continuum in the lower orders.

This is the basis for nodding "on chip", an approach used when the comet emission line intensities do not extend over one half the slit length. The alternative used for a really extended source (like Hale-Bopp or Mars) is nodding "off chip". In this case for the A-frame the comet is positioned in the slit center while for the B-frame the telescope is moved perpendicular to the slit outside the spatial extension of the source (Figure 3). All NIRSPEC data to date have been nodded "on chip".

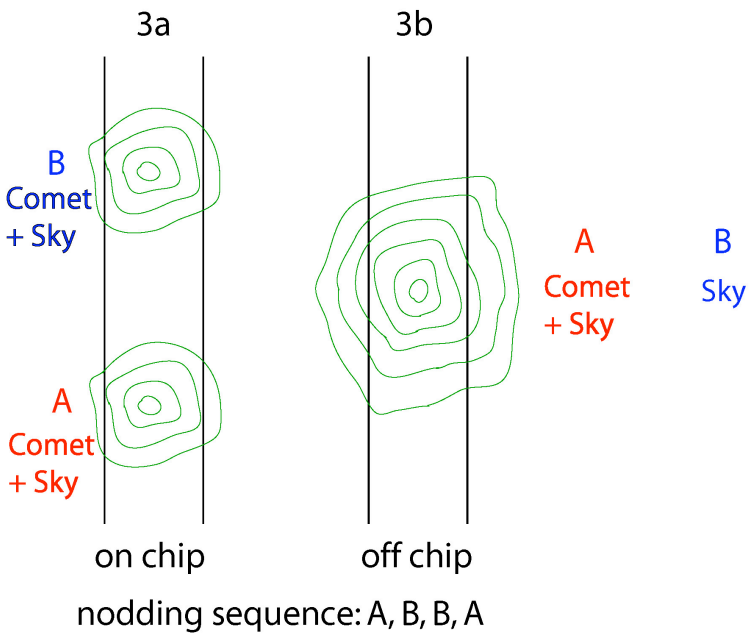

Fig. 3 – Schematic representation of the two commonly used telescope nods.
3a – **nodding "on chip"**. Used for weaker sources. The main advantage is an increased observing efficiency as both the A- and the B-frame contain comet signal. The second advantage is better sky cancellation (see Appendix 2). The disadvantage is that the signal in the middle of the order cancels, which is usually not a serious issue for weaker sources.
3b – **nodding "off chip"**. Used for bright sources that "fill" most of the slit in its longer dimension. The advantage is that maximum spatial information in gained. The disadvantages are lower observing efficiency (the B-frames do not contain comet signal) and harder sky cancellation. The latter is an issue especially in the M-band, where the sky is brighter, therefore fluctuation in the sky intensity between A- and B-frames are more significant.

3.4 Simultaneous Detections of Multiple Parent Species with NIRSPEC

The big advantage of NIRSPEC is the ability to measure many species of interest simultaneously within a single instrument setting, and with high sensitivity. The molecular species commonly observed in the NIR at present can in most cases be sampled within only two or three settings in the L-band (2.9-3.7 μm) and one in the M-band (4.4-5.5 μm). This includes direct measurements of H_2O near 3 and 5 μm and its dissociation product OH at various frequencies in the L-band; hydrocarbons (C_2H_2, CH_4, and C_2H_6); nitrogen bearing molecules (HCN, NH_3); and oxygen bearing molecules (CO, H_2CO, and CH_3OH). At least one representative of each group is sampled within the setting shown on Figure 2. The main results of this project are based on analysis of this setting designated as "KL2".

The simultaneity of detection is a very important aspect of data taken with cross-dispersed IR spectrometers. It eliminates serious sources of systematic error that might affect the determination of individual production rates but cancel out upon determination of mixing ratios. The most important systematic effects are related to seeing, flux calibration, and drift of the comet across or along the slit. They affect the measurements within a given setting the same way, but introduce significant complications otherwise. Our group routinely encounters examples of such inter-setting issues in our on-going projects based on CSHELL data, which (as mentioned above) have a very limited free spectral range. In addition to avoiding such complications, the simultaneous detections with cross-dispersed instruments enable significantly better observing efficiency.

3.5 Data Reduction. On-going Development of the Processing Algorithms for High-Resolution IR Spectroscopic Data.

The reduction of data, also known as the "pre-science part" of the work is often described only in a few words in refereed papers, which is "unfair" because this step is absolutely critical and has required many hours of attention in this project. Having said that, I am leaving the description of the principal data reduction steps for Appendix 2 in order to keep the flow of the science presentation smoother. One issue, however, is imperative to emphasize here: our Goddard group on IR spectroscopy has pioneered the application of CSHELL and NIRSPEC to comets. This requires the development of sophisticated reduction and analysis routines for extraction of quantitative results. All these routines have been created "from scratch" within the group using IDL (Interactive Data language). This development is ongoing! This thesis research has been accomplished in an environment of constantly improving algorithms and the author has contributed significantly to that process. I joined the group during a time of transition, when the focus was shifted from nodded "off chip" CSHELL data (very bright comets) to nodded "on chip" NIRSPEC data (fainter comets). Also, the spectral grasp increased by a factor of 24, requiring major software development. As a result modified approaches, more suitable for the latter data had to be utilized. In addition our understanding of critical reduction steps like flux calibration and dealing with certain types of instrumental noise have become more profound. Last, but not least, our computational capabilities increased dramatically around mid-2004 (switching from an old Solaris workstation to the much more powerful G5 Mac platforms). A reduction procedure that I ran for an hour during my "learning curve" period now runs for several minutes. The increased

computational power has enabled us to apply reduction approaches that could not have been run on our previous platforms. The core results of this thesis have been produced by the current best algorithms.

Chapter 7 mentions some of the author's contributions to the algorithm developments that were required for successful completion of the project and should be considered an intimate part of this thesis work on a par with the science results.

3.6 The Spectral Gallery

The main results of this project are based on reduction and analysis of data in two Oort Cloud comets: C/2000 WM1 (LINEAR) and C/2004 Q2 (Machholtz) (hereafter WM1 and Q2). This section presents spectral extracts from all data analyzed in these two comets – order by order, as each spectral order was reduced separately (see Appendix 2). Complementary results include C/1999 H1 (Lee), C/2001 A2 (LINEAR), and the "Deep Impact" mission target 9P/Tempel-1; spectral extracts from them will be presented in the corresponding sections.

Appendix 3 provides condensed observing logs for the aforementioned comets. The WM1 data were taken on Nov. 25, 2001; the Q2 data – on Jan. 19, 2005. These are both interesting cases illustrating some of the caveats needed when planning comet observations. WM1 was expected to become very bright, but instead showed significantly lower productivity than predicted. Still the Nov. 25 data are high quality because of significant time on source and very favorable atmospheric conditions during observing.

We had not even written a proposal to observe Q2, because this comet was

discovered “too late”: by the time Q2 was known to the community our proposal deadline had passed and we had already submitted for another target. We observed this comet in November 2004, during part of the night when our principal target (also non-sidereal source) was not favorably positioned on the sky. Q2 brightened by early January, 2005, when Prof. Richard Ellis and his graduate student Dan Stark took eight minutes on source for us, for which we are very grateful. Eight minutes on source when the comet is near maximum brightness can make the difference!

Figure 4 shows spectra from orders 26, 24, 23, and 21 from WM1 and Q2 and order 25 from the latter target. All orders belong to the KL2 setting (Figures 1 and 2). The OH lines from order 22 are severely blended with CH_3OH, and therefore have not been analyzed. All spectra have been flux calibrated using the procedure described in Appendix 2.

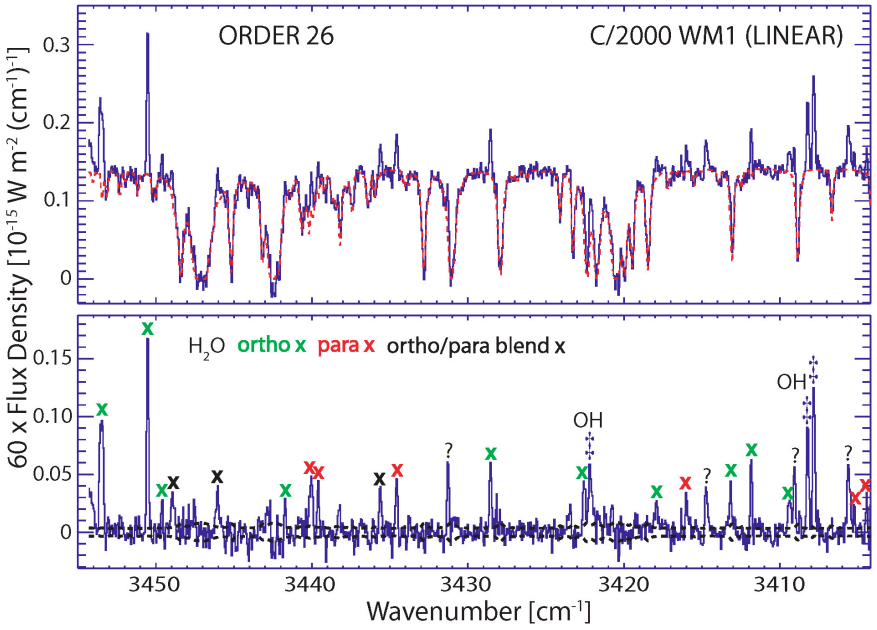

Fig. 4A – Emission from C/2000 WM1 (48 min. on source): spectral extract from order 26 in the KL2 NIRSPEC setting. The spectral order is represented by two panels. Data in both panels are in blue. The red dashed curve in the upper panel represents a best-fit atmospheric transmittance model normalized to the mean continuum intensity of the comet data. The lower panel shows residual spectra after modeling out the telluric absorption. The black dashed lines outline the 1-σ photon noise envelope. The same form of presentation is used in the rest of the spectral gallery.

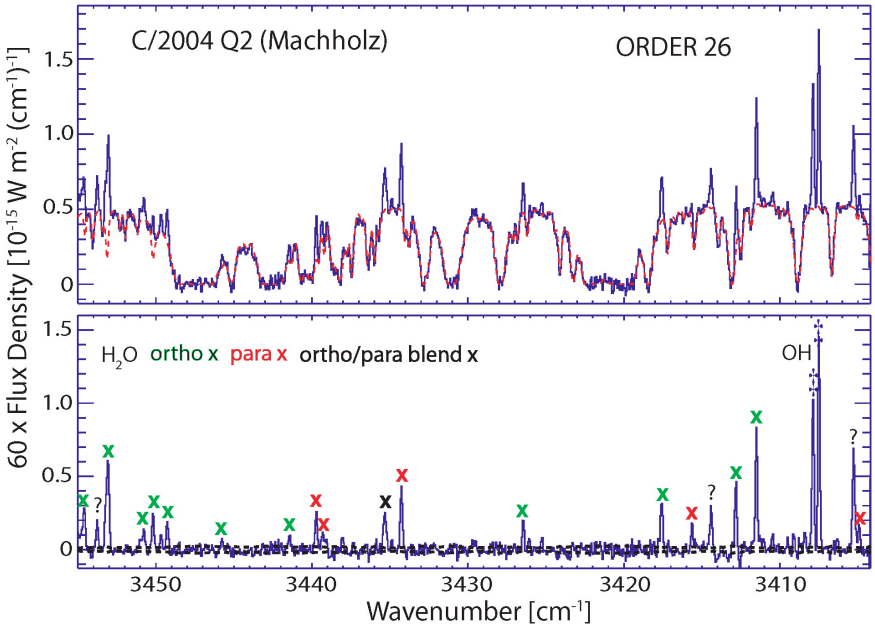

Fig. 4B – Emission from C/2004 Q2 (8 min. on source) – order 26.

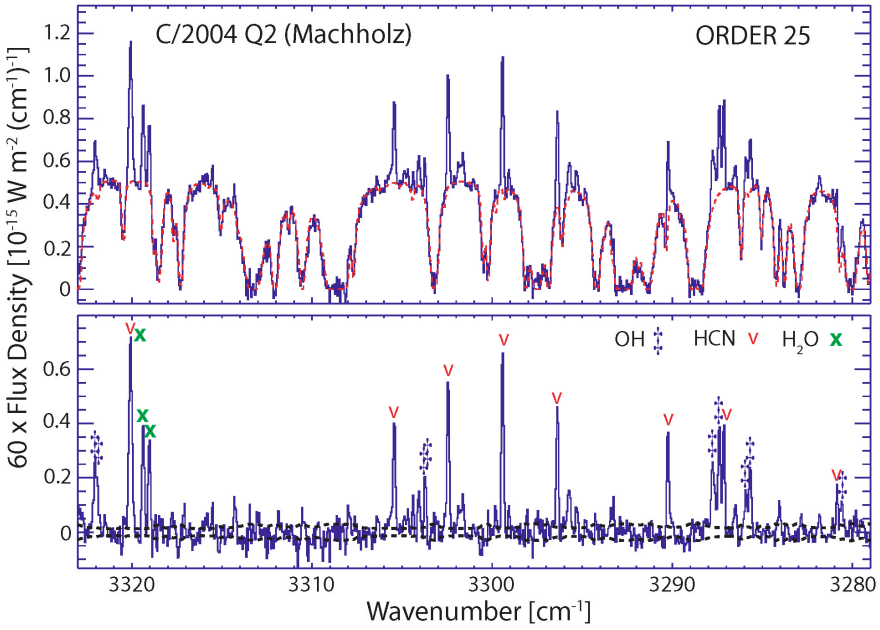

Fig. 4C – Emission from C/2004 Q2 (8 min. on source) – order 25.

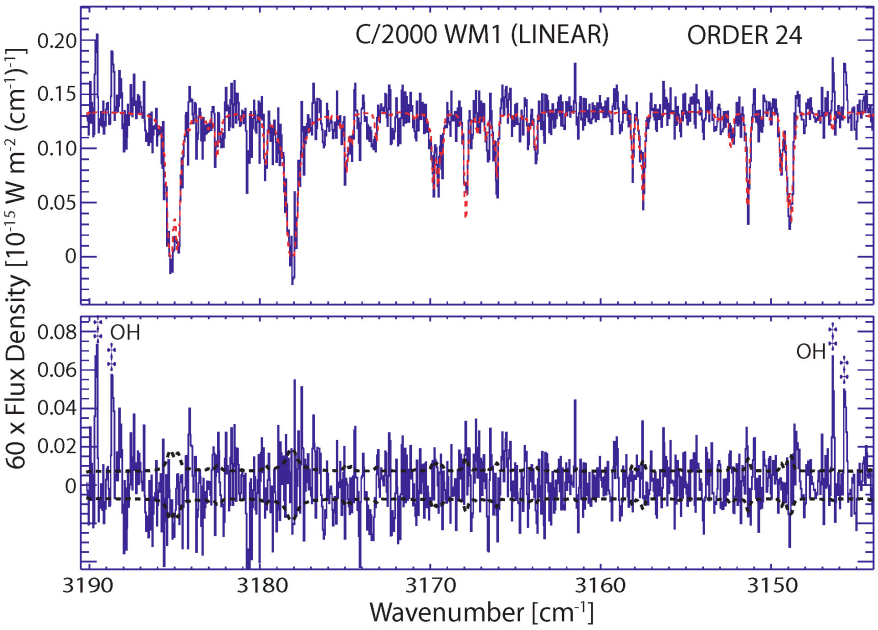

Fig. 4D – Emission from C/2000 WM1 (48 min. on source) – order 24.

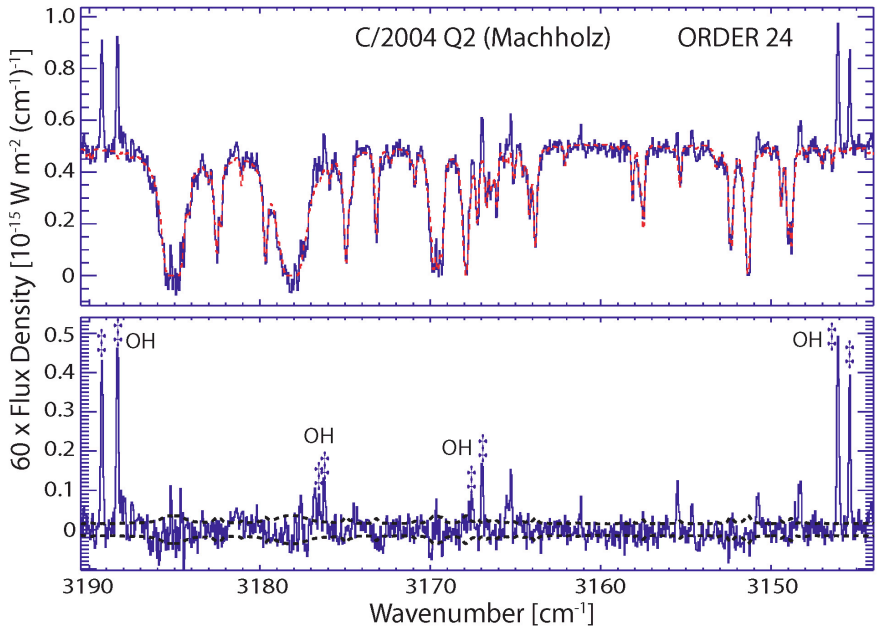

Fig. 4E – Emission from C/2004 Q2 (8 min. on source) – order 24.

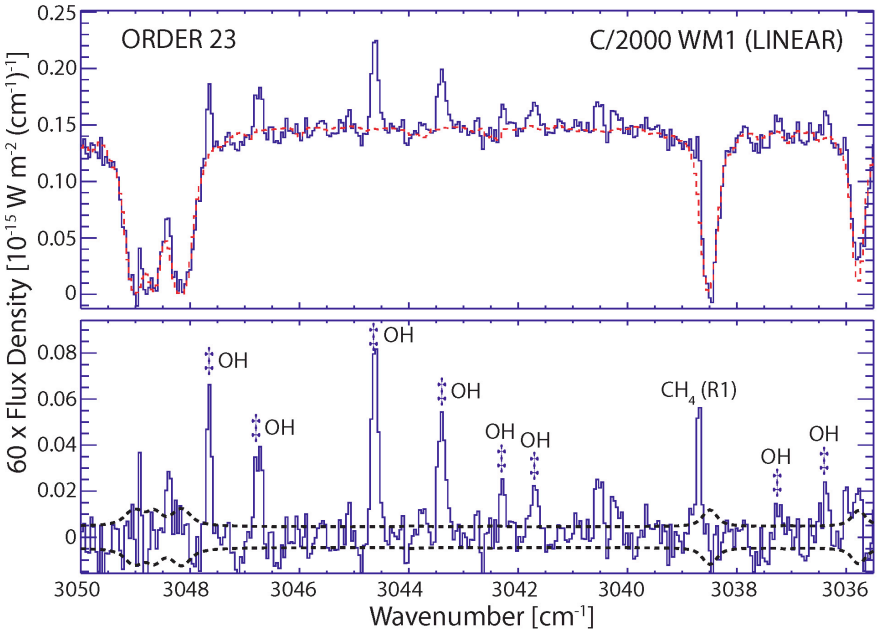

Fig. 4F – Emission from C/2000 WM1 (48 min. on source) – order 23.

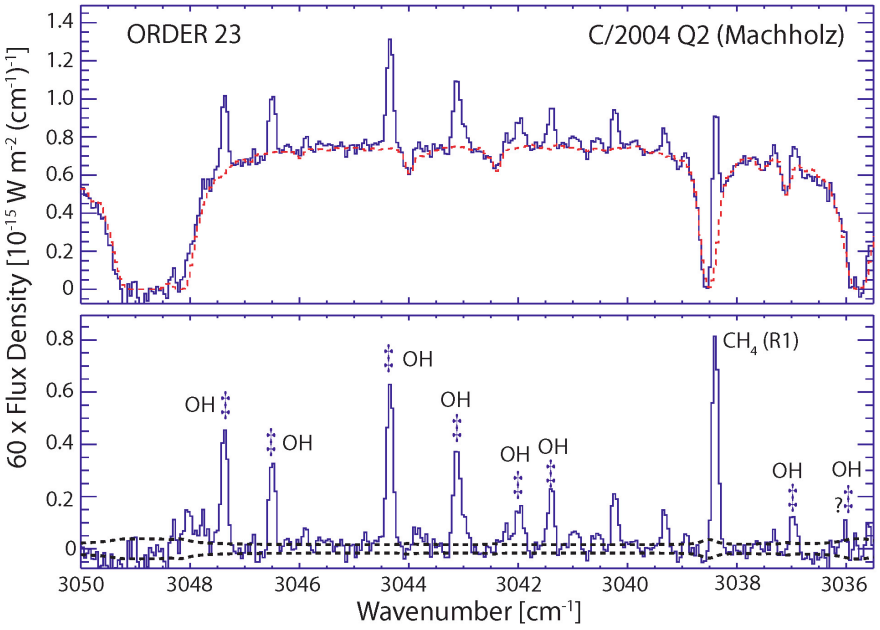

Fig. 4G – Emission from C/2004 Q2 (8 min. on source) – order 23.

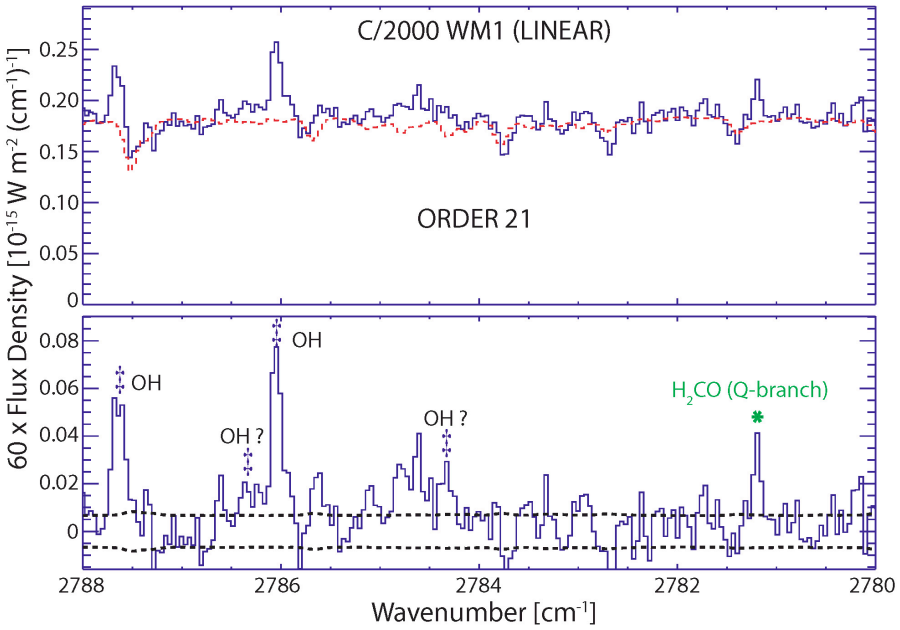

Fig. 4H – Emission from C/2000 WM1 (48 min. on source) – order 21.

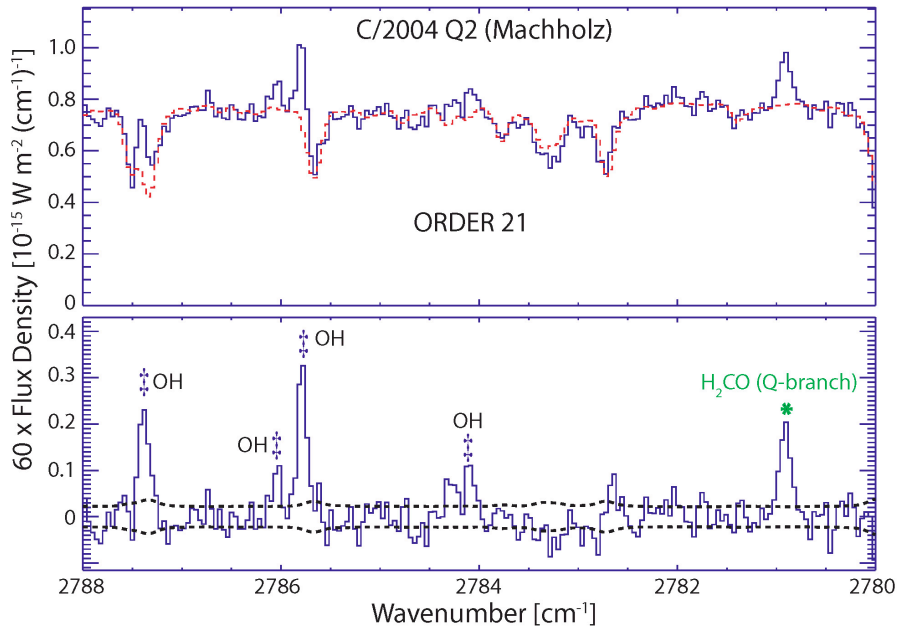

Fig. 4I – Emission from C/2004 Q2 (8 min. on source) – order 21.

Each spectral order is presented in a two-panel plot. The upper panel displays the measured absorption spectrum of the terrestrial atmosphere with a "best fit" atmospheric transmittance model. Cometary emissions are seen against these optimized model spectra. Atmospheric models were synthesized using the Spectral Synthesis Program (SSP; Kunde & Maguire 1974), that accesses the HITRAN molecular data-base (Rothman et al. 1992, 2003). The lower panels show residual spectra after subtraction of the telluric component. These spectra reveal cometary emission from various species. There are uncertainties in the atmospheric model fits, but they have mostly been kept within the limits of the photon noise envelope (also shown in the lower panels), especially in spectral intervals that include cometary emission lines of interest in this study.

A first principle in spectroscopy is that a secure identification of a given species requires detections of its multiple lines. H_2O is seen in non-resonance fluorescence mostly in order 26. The commonly detected lines of HCN are clustered in order 25, while the CH_4 low-excitation R-branch lines are seen in order 23. Order 21 represents part of the spectral range for H_2CO, including the Q-branch.

This spectral gallery reveals OH lines at various frequencies. Such lines are commonly found near emissions of species of primary interest in astronomy and astrobiology like H_2CO, CH_4, and HCN. This proximity in frequency space is an important advantage when OH is used as a proxy for H_2O production, enabling H_2O and the trace constituents to be quantified simultaneously, even when direct water detections are not available.

A note about the atmospheric absorption spectra: any overall trend of the

continuum intensity with frequency (respectively column number) within a given order has been pre-removed via a low order polynomial fit (e.g. Figure 5). For this reason the continua shapes in all orders look mostly "horizontal". Exactly the same fit was removed from the synthetic atmospheric spectrum, so that it is distorted in an identical way to the observed absorption spectrum.

Initially a transmittance function is created with some assumed values of the abundances of terrestrial species. This transmittance function is convolved to the spectral resolving power of NIRSPEC (~25,000 with a 0.43″ wide entrance slit) and is then normalized to match the continuum level of the observed spectrum. Abundances are next adjusted for optimal fits with the observed atmospheric absorption lines. Such normalization and abundance fitting is easier and more reliable if the continua levels in data and transmittance spectrum have been initially subjected to the de-trending procedure. This is a matter of convenience, rather than a strict requirement of the analysis.

It can easily be shown that the described procedure distorts the cometary emission line intensities (Figure 5). Therefore any frequency trend removed before normalizing and fitting the terrestrial component is replaced in the residual spectra. This restores the original ratios between the intensities of the different emission features in the order.

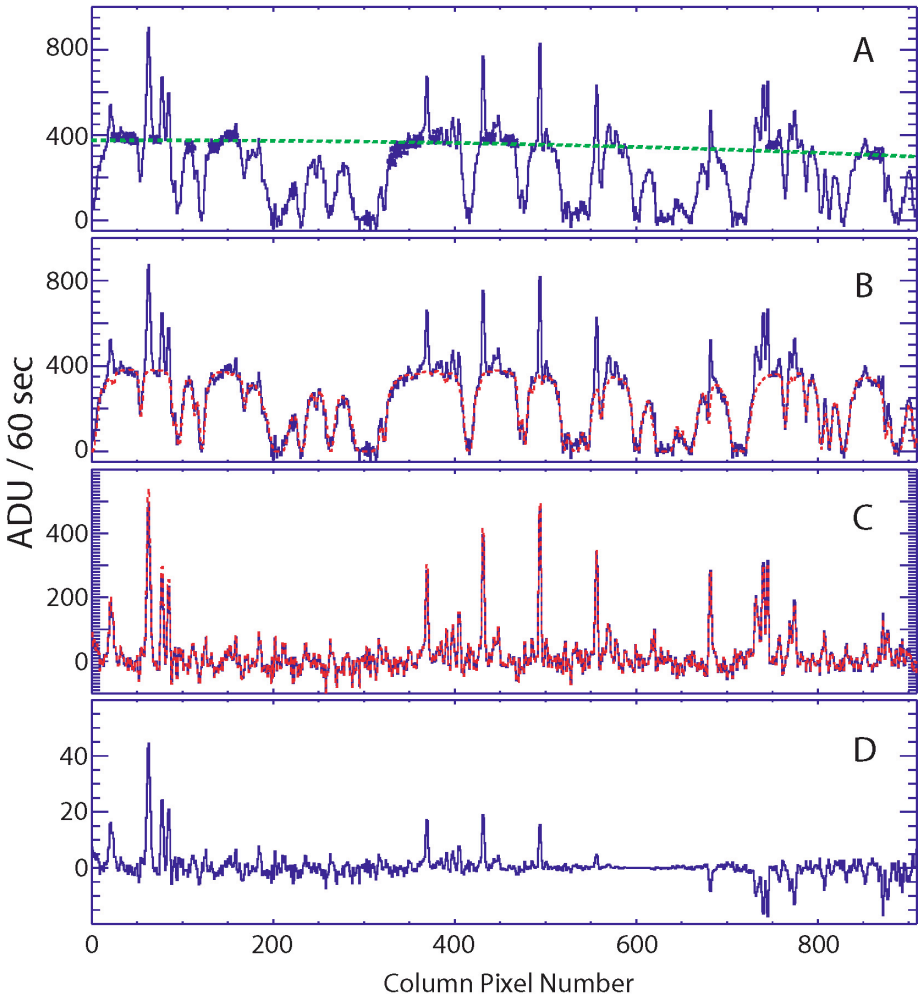

Fig. 5 – Obtaining a residual spectrum.
(A) – Original spectrum and a polynomial fit through the marked points describing the dependence of the overall continuum level on column number. This same fit is divided out from both the original spectrum and the atmospheric transmittance function, before the latter is normalized to the observed continuum levels and fit to the data.
(B) – Resulting spectrum after removing the trend shown on Figure 5A and a best fit atmospheric model. Subtracting this model removes properly the telluric component of the spectrum, but leaves the resulting residuals distorted.
(C) – Solid blue curve: residual resulting from direct subtraction of the terrestrial atmospheric model shown on Figure 5B; dashed red curve: residual corrected for "slope distortion". This correction consists of including back the shape of the green curve shown on Figure 5A.
(D) – Difference between the corrected and the distorted residual (red and blue curves on Figure 5C).

3.7 The Importance of Significant Doppler Shift for Certain Cometary Emission Lines

Cometary emission lines that have terrestrial counterparts (in absorption) can be detected if the comet displays a sufficient Doppler shift of the comet relative to the observing site on Earth. Such is the case for CH_4 (order 23) and CO, whose lines commonly fall in the wings on their corresponding atmospheric absorptions. The synthetic spectrum must be especially accurate in such cases.

On the other hand the observed emissions of H_2O and OH do not have terrestrial counterparts in absorption. These emissions still depend on accidental overlaps with telluric absorptions such as H_2O fundamental bands for example. The dependence is stronger for the H_2O non-resonant fluorescence lines near 2.9 μm (order 26). Water observations in that region definitely benefit from low H_2O vapor content in Earth's atmosphere. A number of OH lines fall within a region of high atmospheric transmittance for a range for comet geocentric velocities. This relative independence of the detectability of OH lines on Doppler shift is a factor favoring their potential usage as a proxy for H_2O production.

Chapter 4

Water Non-Resonant Fluorescent Emission in C/2000 WM1 and C/2004 Q2

The main goal of this chapter is to establish well constrained water production rates from analysis of the H_2O lines seen in non-resonance fluorescence near 2.9 μm (Figure 4, order 26). These directly constrained production rates will be used as a "calibrator" for the OH emission efficiencies derived in Chapter 5, and as a baseline for quantifying the overall volatile chemistry of the comets WM1 and Q2. Along the way, the new results on water will be presented as a working example of the methodology developed to interpret the only type of H_2O emission that can be detected routinely from ground-based observatories.

4.1 Resonant and Non-Resonant Fluorescence

Kuiper Airborne Observatory provided the first secure detection of the dominant parent volatile in comets. The vast majority of the detected water lines were seen in resonant fluorescence (here the term "resonant" in used to indicate that only two vibrational states are involved; the rotational states however, might change). This choice

to target fundamental bands was clear simply because they represent the strongest ro-vibrational transitions. In resonant fluorescence, direct absorption of sunlight excites molecules from the lowest vibrational level to a higher vibrational level. This is followed by a radiative transition back to the ground state. In addition to being strong, emissions from fundamental bands are relatively well understood, thus enabling a reliable interpretation. And yet there is an "issue", an unsolvable obstacle as far as ground-based observers are concerned: severe terrestrial atmospheric extinction. The lowest vibrational state of H_2O is highly populated in the Earth's lower atmosphere, resulting in absorption of the comet photons. Water fundamental bands of course can be detected from airborne and space observatories. But for ground-based observers, the so-called "hot-bands" come to the rescue.

In non-resonant fluorescence, radiative vibrational excitation from the ground state is followed by a cascade into intermediate vibrational levels, which are not significantly populated in the Earth atmosphere (Crovisier 1984; Bockelee-Morvan & Crovisier 1989). As noted in the end of the previous chapter, detections of H_2O in non-resonant fluorescence does not require a specific Doppler shift. Hot-band emission was first detected in the KAO spectra of comet Halley (Larson, Mumma, & Weaver 1987). Observations of H_2O hot-bands have been used routinely to quantify water production directly from the ground for about a decade (e.g. Mumma et al. 1995, 1996; Weaver et al. 1999b; Dello Russo et al. 2000, 2002; Brook et al. 2003). Several hot-bands are now used, the most fruitful regions being near 2.0, 2.9, and 4.7 μm). The region near 2.9 μm has been explored most recently (Dello Russo et al. 2004, 2005). This region contains emissions from six hot-bands such as those shown in our spectral gallery in order 26

(Figure 4). These lines have become a preferable target when using NIRSPEC for two main reasons: first, they and several other important species are detected within the same instrument setting; second, a single echelle order contains a number of emission lines, including several of para-H_2O, allowing us to constrain the rotational temperature and, sometimes, the spin temperature of H_2O. These two parameters are needed in order to fully constrain the water production rate via the hot-band method. They are more difficult to determine from observations in the M-band than from observations near 2.9 μm.

4.2 Principal Elements in the Analysis of H_2O Non-Resonant Fluorescence

Determining H_2O production rates from non-resonant fluorescent emission includes the following principal steps:

1. Development of a H_2O hot-band fluorescence model.
2. Retrieval of the H_2O rotational temperature (T_{rot}).
3. Retrieval of the H_2O ortho-to-para ratio and spin temperature (T_{spin}).
4. Deriving the H_2O production rate (Q[H_2O]) from each individual H_2O emission feature detected.
5. Deriving the overall Q(H_2O) as a weighted mean of the line-by-line measurements.
6. Evaluating the effects of atmospheric seeing on the retrieved Q(H_2O).
7. Evaluating and quantifying as well as possible the different sources of uncertainty in T_{rot}, T_{spin}, and Q(H_2O).

This work has been focused on steps 2-7, adopting recently developed fluorescence models for the H_2O emission near 2.9 μm. In the following sections each part of the analysis sequence will be described using examples from the new work on water lines in order 26 of the KL2 NIRSPEC setting.

4.3 Spatial Distribution of H_2O

Before moving forward with the hot-band method description, it is very important to provide a clear idea about the variation of the H_2O line intensities as a function of projected distance from the cometary nucleus. Though more limited, spatial information is usually available for data nodded "on chip". In this case, sufficient S/N in the spatial profile is achieved by summing the flux of multiple H_2O lines and tracking the variation of this summed flux along the spatial dimension. Figure 6 shows the H_2O spatial profiles observed in comets WM1 and Q2. Both profiles represent the combined flux of many water lines detected in order 26.

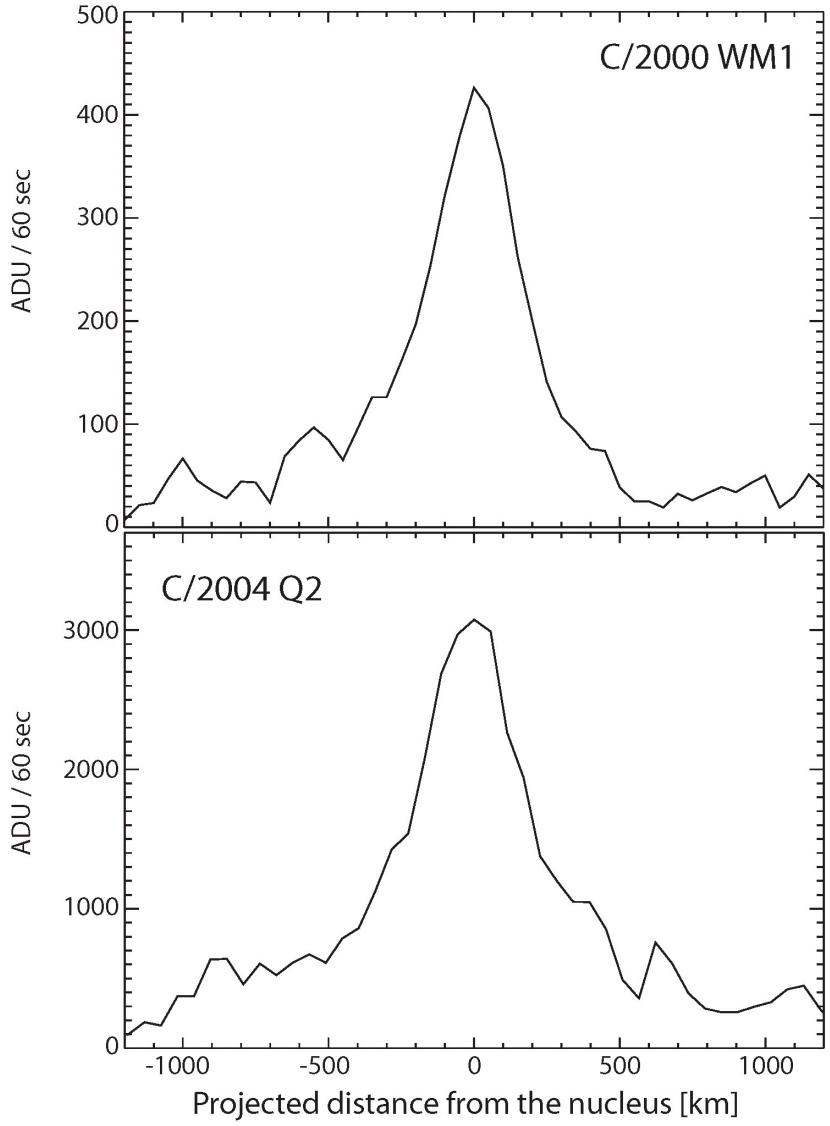

Fig. 6 – Spatial profiles of H_2O in two comets. These profiles represent the summed column-by-column signal from multiple water lines detected in order 26 (Figure 4). It is typical for a distribution of a parent volatile to peak at the nucleus. Note that it is possible that H_2O have a minor extended source of icy grains which cannot be distinguished from these profiles, because their shape is governed by the dominant native component.

The first thing to emphasize is the spatial extent of the sampled cometary coma. NIR observations characteristically cover the innermost several hundred to several thousand km from the nucleus, depending (for a given telescope/instrument combination) on geocentric distance. Recall from the introductory chapter that the visible extent of the cometary atmosphere is about 10^4-10^5 km. Observing very close to the nucleus is a major advantage in interpreting NIR observations of native molecules.

Next, we recognize that the spatial distribution of H_2O is typical for a parent species. The emission intensity peaks at the nucleus a falls off approximately as the inverse of the projected distance from it. By contrast, the distribution of a fluorescent emission from a daughter fragment does not necessarily peak at the nucleus and would look much "flatter" on the small spatial scale of Figure 6. For example Dello Russo et al. (1998) observed NIR transitions of the CN radical in C/1995 O1 (Hale-Bopp) showing that the CN lines fill the slit almost uniformly along the spatial dimension, while the emissions from various parent volatiles exhibited spatial distributions similar to those shown here on Figure 6 (i.e. their intensities falling off much more rapidly).

Note that the quantitative analysis performed in the following sections refers to the innermost coma (~250 km from the nucleus for Q2, and ~230 km for WM1), the region in the spatial profile with highest S/N. All spectral extracts in Figure 4 correspond to this region. This is a characteristic approach for data nodded "on chip". The highest S/N data permit the most reliable line-by-line analysis. Such analysis is always desired. On the other hand, the spatial profiles representing the summed flux of multiple H_2O lines will be used subsequently in evaluating some significant systematic effects on the retrieved production rates.

4.4 H_2O "Hot-Band" Fluorescence Models

The astronomical field of high-resolution molecular spectroscopy is governed by two main drivers. The first one is the evolving technology which enables more sensitive instruments with higher spectral resolving power and larger free spectral ranges. Development of reduction algorithms to handle the increases in both quality and volume of the data is included in that aspect. The second driver is the development of quantum mechanical models that describe emissions from various astronomically observed species.

Predicting the intensities of ro-vibrational lines requires knowledge of the emission rates (also known as fluorescence efficiencies or g-factors) of these lines, expressed in [photons sec^{-1} $molecule^{-1}$] (cf. Bockelee-Morvan et al. 2004). The g-factor for a given transition is defined as the product between the spontaneous emission Einstein coefficient for that transition and the fractional population of the corresponding upper level. Fluorescence models have been developed to determine g-factors as a function of temperature for several H_2O bands seen in non-resonant fluorescence.

A set of recently derived g-factors for H_2O hot-band lines (Dello Russo et al. 2004; 2005) have been adopted in this work and were tested using data from WM1 and Q2. We now discuss the physical assumptions of the hot-band fluorescence models used to derive these H_2O emission rates.

The first assumption is an efficient radiative vibrational relaxation to the ground state of the sublimated H_2O molecules. Sublimated molecules leave the nucleus with vibrational and rotational temperatures expected to be equal to the surface temperature of the nucleus, i.e. about 150-200 K at 1 AU. Except within a few hundred meters from the

nucleus, collisional rates are too small to maintain the initial vibrational population, so this population decays radiatively to the ground vibrational level (Weaver & Mumma 1984). Subsequent vibrational excitation occurs through pumping from the incident solar radiation. As mentioned above, in non-resonant fluorescence the excitation by solar IR photons is followed by decays into intermediate vibrational states giving rise to the observed hot-bands. These transitions are optically thin; however, optical depth effects in the solar pump should be considered in cases of very high gaseous activity and close proximity to the nucleus (a few to hundreds of km from the surface depending on the considered transitions and water production rate) (Dello Russo et al. 2004).

Note that radiative vibrational excitation of H_2O from processes other than direct solar pumping have also been considered, in particular by IR photons scattered or thermally emitted from the nucleus and from the dust. These mechanisms are shown to be negligible except excitation from thermal dust emission, which occurs, however, at longer wavelengths not relevant to this study (Crovisier & Encrenas 1983; Weaver & Mumma 1984).

The second assumption is that the rotational population in the ground vibrational level follows a Boltzmann distribution. The rotational states are thermalized by collisions in the inner coma at the kinetic temperature of the gas. At 1 AU and for moderately high gas productivity ($\sim 10^{29}$ molecules sec^{-1}) the rotational levels are thermalized for distance less than several thousand km, which usually exceeds the spatial extent of the coma sampled by NIR observations (e.g. Figure 6). Moreover, radiative rotational relaxation is not efficient because in the inner coma the rotational transitions of H_2O are optically thick, thereby preventing the de-population of excited rotational levels.

In order to obtain hot-band g-factors one needs to calculate:

1. The initial rotational population of the ground vibrational state. Boltzmann distribution assumed, this means knowledge of the H_2O rotational temperature is required for that state.

2. The pumping and the cascade rates for populating individual rotational levels of the upper vibrational state for a given transition.

3. The rotational and vibrational branching ratios for the observed non-resonant transitions.

4. The ortho-to-para ratio of the H_2O molecule.

We have an expanding library of g-factors calculated for different rotational temperatures assuming OPR = 3. Historically in our group fluorescence models for bands near 2 μm were developed first, followed by models for bands near 4.7 μm, in order to quantify the production of simultaneously measured H_2O and CO. The most recent improvements were accomplished for several hot-bands observed in the 2.9 μm region.

Inasmuch as vibrationally excited states with energies above ~4500 K cannot be populated efficiently in the laboratory, absorption line strengths are generally not available for upward transitions from such states. As a result, calculations of Einstein coefficients for spontaneous emission provide the largest source of uncertainty in the available g-factors of individual lines. Energy levels, wavefunctions and ultimately Einstein A's have been calculated by our collaborators from the theoretical group in the University College London (Tennyson et al. 2004; Barber, Ph.D. thesis, in progress).

It should be emphasized that comets can be used as "ad-hoc" laboratories to test

H_2O fluorescence models. The observations of WM1 and Q2 provide two premium data sets (in some ways the best to date for the 2.9 μm region, as far as NIRSPEC is concerned) for such testing, and this aspect of the work will be emphasized further. Mike DiSanti and I developed a simple but efficient approach to test the reliability of the g-factors available, based on correlation analysis. This approach complements the version of excitation analysis already being used for deriving H_2O rotational temperatures (Dello et al. 2004, 2005).

4.5 De-coupling the Rotational Temperature From the OPR

The hot-band g-factors and respectively the H_2O production rates depend on two parameters that must be retrieved from observations: the rotational temperature and the OPR. These two parameters are coupled through the observed relative intensities among the water lines. To de-couple them, one needs to derive the rotational temperature for each spin species independently. The relative emission intensities among lines of a given spin species of course depend only on T_{rot}, but not on the OPR. The number of the observed ortho lines and their spread in excitation exceed significantly those for the observed para lines in the vast majority of cases. For this practical reason the ortho lines have been chosen for T_{rot} extraction. Blends of orthto and para lines are not included in the analysis.

4.6 Correlation Analysis

The purpose is to conduct an initial test of the fluorescence model at hand against the relative intensities of the ortho lines. Along with this, a first constraint on T_{rot} is also accomplished. The idea is simple:

An OPR of three is assumed at this stage, so the remaining free parameter is the rotational temperature. Then:

1. T_{rot} is varied with a certain step size (1-10 K) within a broad interval of values expected to characterize the inner coma. In this case 50-150 K.

2. For each value of T_{rot} a synthetic spectrum of H_2O emission is built, using the available g-factors. This spectrum is convolved to the resolving power of the data to enable a direct comparison of the modeled and observed relative intensities.

3. The correlation coefficient, R, between the observed and the synthetic spectrum is calculated. The observed spectrum in this case consists of the fluxes measured at the expected position of the ortho lines in our data.

4. The plot of R vs. T_{rot} will be referred as a "temperature correlogram" or "T-R correlogram". The near maximum part in the plot of R vs. T_{rot} corresponds to the most probable rotational temperature range. The sharpness of this maximum is indicative of how well T_{rot} can be constrained by the correlation analysis.

To illustrate this approach Figure 7 presents a comparison of model with data (from order 26, Q2) for three different temperatures, including the optimal one. Figure 8 shows the temperature correlograms for the ortho lines in order 26 for WM1 and Q2.

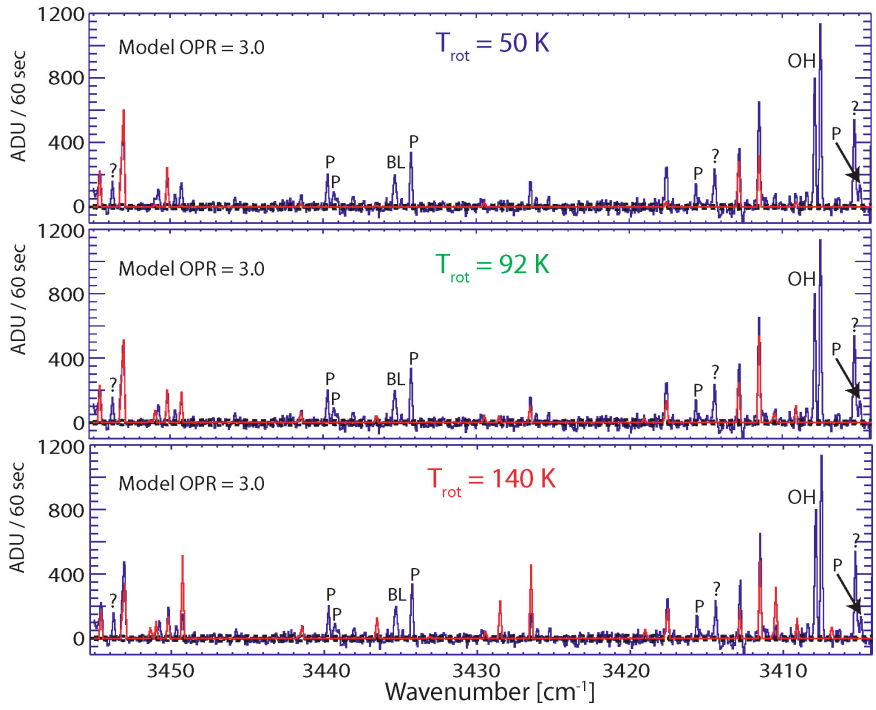

Fig. 7 – Illustration of the correlation analysis. Synthetic spectrum based on fluorescence model for the ortho-H_2O lines (red) is compared to the observed spectrum (blue). The model is scaled to the data for representative purposes only, as the correlation analysis is really a test for the relative line intensities. For the same reason the test may be performed on a spectrum which is not flux calibrated considering that the calibration factor Γ does not vary within order 26 (Appendix 2). The correlation coefficient between model and data is calculated over the frequencies sampling ortho lines. Para lines (P) and ortho/para blends (BL) are indicated for clarity.

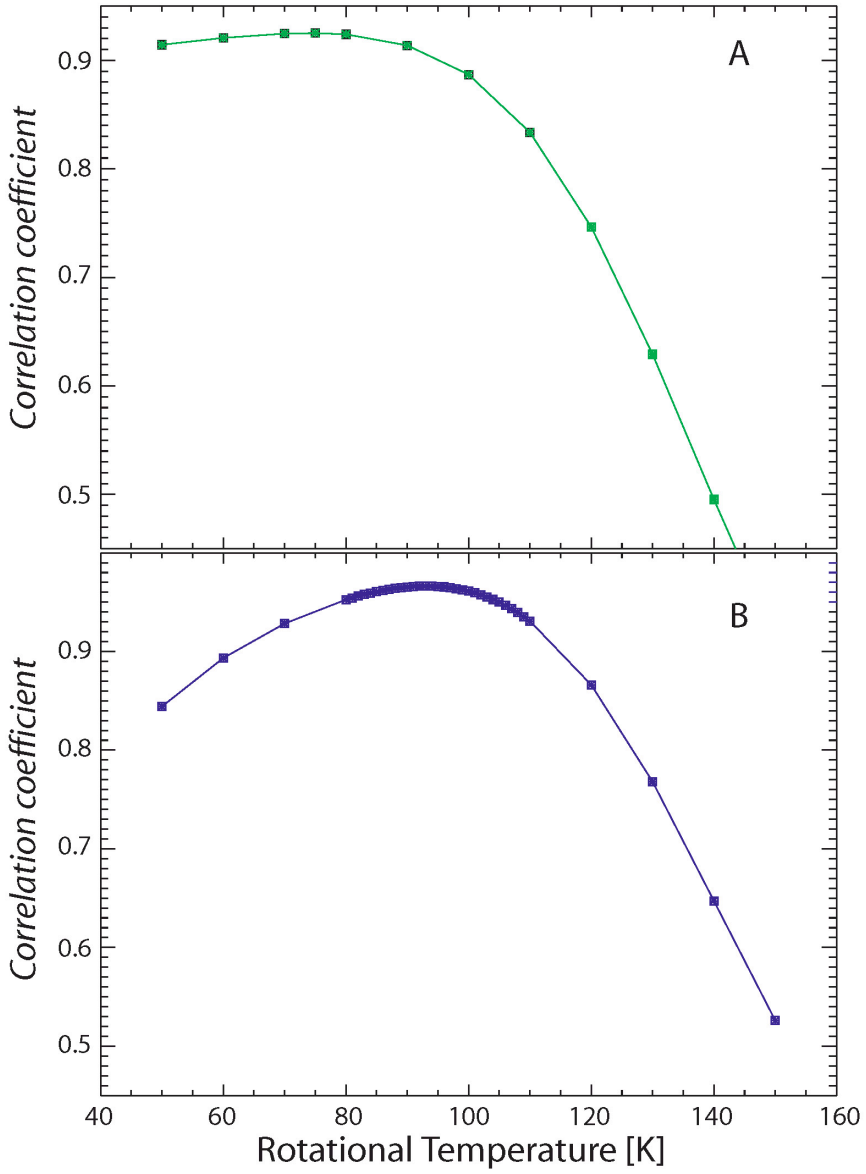

Fig. 8 – Rotational temperature correlograms for the WM1 (A) and Q2 (B) ortho lines in order 26. Note that the high-temperature end of these correlograms is much better constrained.

The WM1 data were analyzed at a time when g-factors were available only at intervals of 10 K, which determined the increment in T_{rot} in the correlation analysis (Figure 8A). I ran an additional fluorescence model for 75 K which was near the maximum of the correlogram. Note that this maximum is not sharp, but the "high end" of the tested temperature interval is pretty well excluded. By constrast, the low temperature end is not so well constrained. The reason is that the sampled frequency range (~3404-3455 cm^{-1}) contains a number of lines excited from relatively high rotational levels in the ground vibrational state. These levels are not expected to be significantly populated below 80-90 K, so emissions from the corresponding excited states are weak or not observed. This is very clear on Figure 7C and shows why the correlation decreases above 80 K. At the same time relative intensities of the strongest ortho lines sampled are less dependent on T_{rot} within the interval 50-80 K, which explains the less steep rise in that part of the correlogram.

By the time the Q2 spectra became available our library of g-factors had expanded to include models at 1 K intervals, which is the increment in T_{rot} in the correlation analysis for that comet (Figure 8B). The most probable temperature range is about 85-100 K, the curve peaking at 93 K.

4.7 Some Comments on the Correlation Analysis

The correlation analysis is really a test of the fluorescence models at hand and a useful one, conducted at every stage of the analysis: here, against only the ortho lines; subsequently, when both T_{rot} and the OPR are constrained - against all H_2O emissions sampled. The correlation analysis presents an efficient way to identify quickly emission

features that are not fit well (or at all) by the synthetic model. The most severe case is for lines that are not fit at any temperature (and OPR). Emissions showing significant deviations are excluded from the subsequent derivation of H_2O production rate, which reduces the uncertainty in the latter. More comments on this issue will be presented further, upon discussing fluorescence models in which both T_{rot} and OPR are optimized.

Note that in difference to other related statistics like χ^2, the correlation coefficient has a fixed range, independent on the type of analyzed data.

Finally, the correlation is really a test for the relative intensities of the H_2O lines, because R is independent of the scaling between model and data. In terms of strictly retrieving T_{rot}, the correlation analysis is useful, but has one particular disadvantage - it does not incorporate explicitly information about the spread in the energy levels involved in the observed transitions. The next section is dedicated to this issue, which ultimately will lead to rotational temperatures with well-constrained confidence intervals.

4.8 Defining Terminology

Complemented by the correlation analysis, the purpose of the next two sections is to establish a reliable value and a confidence interval for the rotational temperature. First, some commonly used terminology should be clearly defined:

In non-resonant fluorescence, three types of ro-vibrational energy levels are involved (Figure 9): (1) the ground ro-vibrational level from which the initial excitation occurs; (2) the upper state of the observed transitions - it is usually populated by excitations from several rotational levels in the ground vibrational state (plus cascade contributions from higher excited vibrational states, which however have much smaller

weight for 2.9 μm bands as shown by Dello Russo et al. 2004); and (3) the lower state of the observed transitions. Note that there are two types of "lower" states: the ground state (the origin of the pump) and the final state of the observed non-resonance transition.

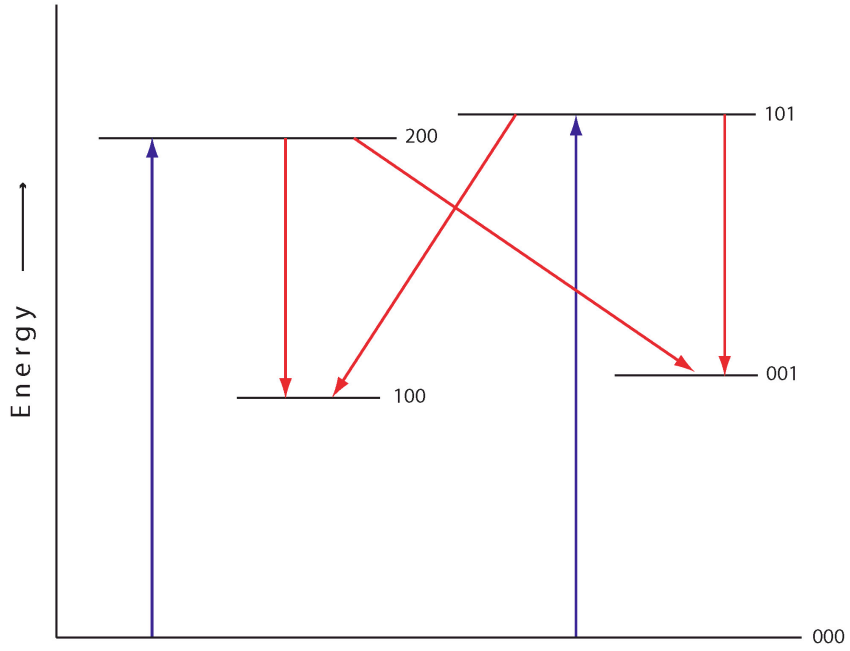

Fig. 9 – Schematic representation (not to scale) of the vibrational energy levels involved in non-resonant fluorescence. The vibrational assignments corresponds to the four major bands sampled near 2.9 μm.

In should be emphasized that the rotational energy within each of these vibrational states is measured against the lowest rotational level in that particular state. Further in the text the term "upper state energy", E_{up}, will refer to the rotational energy of the upper vibrational state for the observed transitions.

The term "lower state energy", E_{low}, here means the average energy of the rotational levels in the ground vibrational state, weighted by their contribution to the population of a given upper ro-vibrational state. E_{low} should not be confused with the rotational energy of the lower vibrational state for the corresponding non-resonance

transition (which is not the ground state).

Note that the upper state energy does not depend on T_{rot}. On the other hand, the lower state energy is a function of the rotational temperature:

$$E_{low}(T_{rot}) = \frac{\sum_i E_i \times pump_i(T_{rot})}{\sum_i pump_i(T_{rot})}$$

Here, E_i is the rotational energy of the i^{th} level in the ground vibrational state; $pump_i(T_{rot})$ is the pump rate (sec^{-1}) from that particular level to the upper ro-vibrational level of interest. The pump rate clearly depends on temperature, because the relative populations within the ground vibrational state are a function of T_{rot}.

4.9 Final Retrieval of the H_2O Rotational Temperature

The modeled g-factors, of the H_2O lines vary with rotational temperature. At the same time the line fluxes atop the terrestrial atmosphere, F (W m^{-2}), reflect the rotational population in the upper vibrational state. This population in turn depends on the rotational population of the ground vibrational state, which is determined by T_{rot} when a Boltzmann distribution holds. At the correct rotational temperature, the quantity $F/g(T_{rot})$ should be independent of the upper state energy (Dello Russo et al. 2004).

Note that here F is the flux of a given H_2O line incident at the top of the atmosphere. It is obtained by dividing the measured flux by the corresponding monochromatic terrestrial atmospheric transmittance modeled at the Doppler-shifted position of a given cometary line.

The fluorescence efficiency, $g(T_{rot})$, is expressed in Watts per molecule and is

calculated for 1 AU:

$$g(R_h = 1AU)\ \ [W \times molecule^{-1}] = g(R_h = 1AU)\ \ [photons \times \sec^{-1} \times molecule^{-1}] \times hc\nu$$

The energy (J) of a photon is designated as "$hc\nu$". The g-factor varies as R_h^{-2} (R_h is the heliocentric distance). Therefore the ratio ($F/[g\ R_h^{-2}]$) (molecules m^{-2}) is proportional to the total H_2O column abundance over the spatial region in the coma corresponding to the flux measurement.

Under this framework the T_{rot} retrieval consists of the following steps:

1. The temperature is varied incrementally at small step size over a broad range, like in the correlation analysis.

2. For each value of T_{rot} the slope in the linear fit "F/g vs. E_{up}" is calculated.

3. The correct rotational temperature is the one for which this slope is zero within error.

This is illustrated in the multi-panel Figure 10. The two top panels (one for WM1 and one for Q2) show the case for T_{rot} lower than its optimal value. This is indicated by a positive slope. For lines with relatively low rotational excitation the g-factors are overestimated, while for high-values of E_{up} the g-factors are underestimated at lower temperatures.

At the optimal rotational temperatures (middle panels) the slope is zero within error. Note that the retrieved values (~75K and ~92K) are consistent with those from the correlation analysis.

When T_{rot} exceeds its optimal value (lower panels) the slope is reversed to negative because the g-factors for the low excitation lines are underestimated, while the modeled emission rates from high-excitation lines are larger than observed.

Figure 11 shows the variation of the slope of "F/g vs. E_{up}" with T_{rot}, illustrating the transition from positive to negative slopes around the optimal temperature.

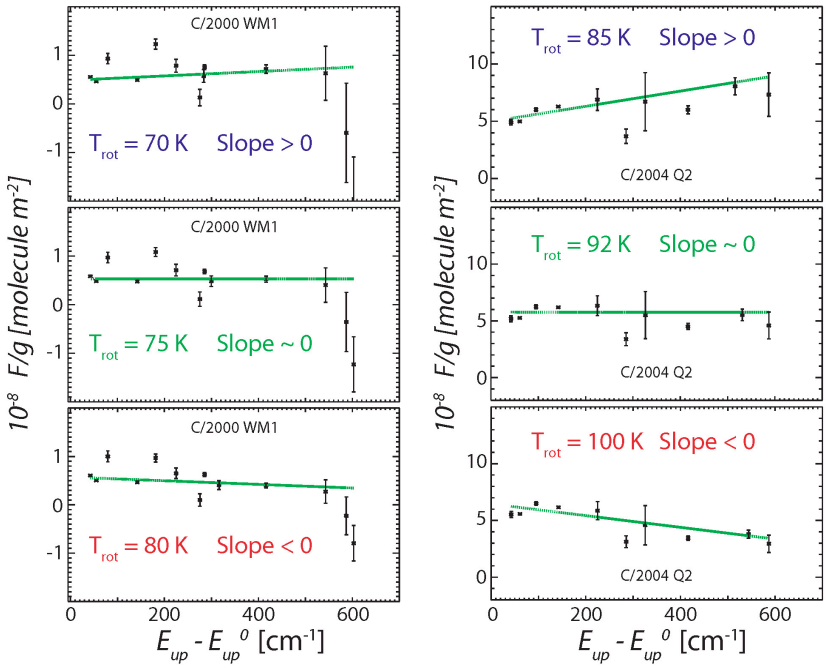

Fig. 10 – Illustration of the excitation analysis used for the final retrieval of the H_2O rotational temperature, based on the intensities of lines of ortho-water. The left panels are for WM1; the right panels are for Q2. The x-axis represents the rotational energy in the upper vibrational state with respect to the energy of its ground rotational level. The quantity "*F*" in the y-axis represents the flux measured at the positions of H_2O lines expected to be detected at some temperature (this flux is also corrected for telluric transmittance). The T_{rot} result for the WM1 is not influenced by the essentially noise measurements for the two lines with highest excitation. These lines could have been seen for higher T_{rot}, so their non-detection is also indicative for the temperature.

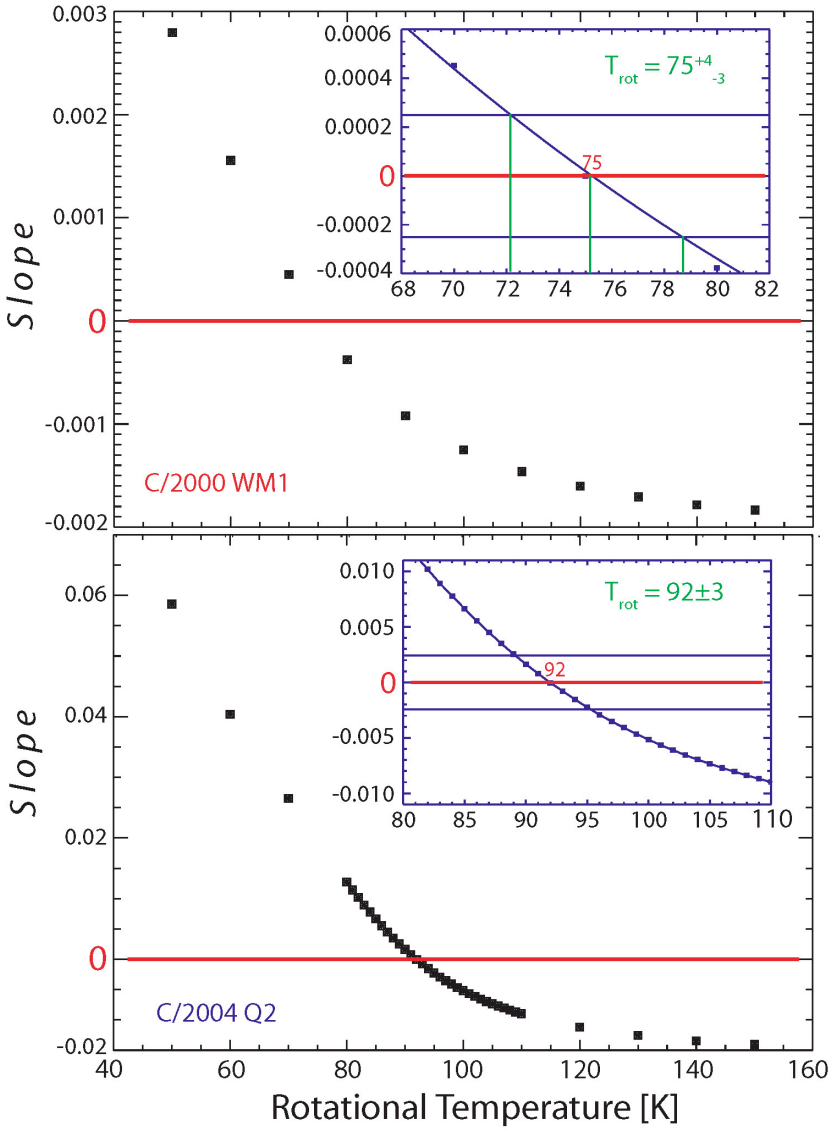

Fig. 11 – Final H_2O rotational temperature retrieval for WM1 (upper panel) and Q2 (lower panel): note the transition of the "F/g(T_{rot}) vs. E_{up}" slope from positive to negative values. The solid curve on the WM1 insert plot is a polynomial fit used to obtain the T_{rot} value and its uncertainties before fluorescence models for every degree became available.

4.10 Uncertainty in T_{rot}

This is the first time uncertainties are discussed – needless to say, this is as important a subject, as retrieving the values of the needed quantities. The error in T_{rot} is calculated in two stages. First, the error in the optimal temperature slope (i.e the zero slope) is evaluated. Then this error is converted into a temperature uncertainty using the "slope vs. T_{rot}" plot like those shown in Figure 11.

Throughout this work two distinct (statistically and physically) types of error are calculated independently whenever a confidence in a slope or a mean is evaluated. These errors are described in Appendix 4 but here we should discuss their physical interpretation:

1. "Stochastic error" - depends on the individual error bars of the quantities involved (here F/g) but not on their spread around the fit line. The term "stochastic" comes from the fact that the uncertainty in the flux is dominated by the stochastic (in nature) photon noise, as discussed earlier. Low stochastic error means high S/N in the data, but not necessarily consistency between line-by-line measurements of the same quantity (e.g. F/g, production rate, etc.).

2. "Standard error" (also referred as "variance error") - depends mainly on the spread of the individual values (here F/g) around the best-fit line (here the linear fit with E_{up}). This error can have various origins but the most important source of uncertainty is likely inaccuracies in the fluorescence models, resulting in over- or underestimated g-factors.

The larger of the two uncertainties is accepted as representative. The standard error usually dominates the stochastic error at the current level of analysis (e.g. Dello

Russo et al. 2005; Mumma et al. 2005; this work). It is important to perceive that the signal-to-noise alone does not contain the whole message as far as uncertainties in T_{rot}, OPR, and production rates are concerned (this will be discussed in more detail later in this thesis).

The errors in slope in particular (both stochastic and standard) also depend on the spread in the x-axis on Figure 10. Physically this translates into the fact that T_{rot} is most accurately determined when a wide range of rotational energy levels are sampled. Our rotational temperatures are constrained quite well because the detected lines of ortho-water represent a wide range of excitation.

The retrieved rotational temperatures are 75 ± 3 K, for WM1 and 92 ± 3 K for Q2 (95% confidence intervals are quoted). These values are consistent with the results from the correlation analysis, confirming the notion that the two approaches complement each other.

4.11 Upper vs. Lower State Energy

The methodology presented in the previous two sections was first introduced by Dello Russo et al. (2004). The error analysis was organized as a part of this thesis and applied initially in Dello Russo et al. (2005). As a new approach there is still some debate about it. In particular, is it more appropriate to use E_{up} or $E_{low}(T_{rot})$ when retrieving T_{rot}?. The main concern is whether or not the two approaches would result in the similar temperature uncertainties. My interpretation of this issue is presented in Appendix 5. The bottom line is that the two alternatives provide very similar results and uncertainties. This is expected because for the optimized temperature the ratio between

observed and predicted flux (the latter depends on g), should be independent of both the upper and the lower state energies. There is only one quantity measured – the column abundance of H_2O (F/g is proportional to it). Therefore the measurements deduced from sampling transitions with different excitation should ideally be the same. Excitation can be represented by the energy in the upper vibrational state, but also $E_{low}(T_{rot})$, because the rotational population of the upper vibrational state depends on the rotational population of the ground state, determined by T_{rot}.

4.12 Ortho-to-Para Ratio of the H_2O Molecule

Chapter 2 introduced the importance of the OPR (alternatively the spin temperature) as a probable cosmogonic parameter. In addition, the determination of a total H_2O production rate requires that we constrain the OPR of cometary water. The algorithm for this retrieval is as follows:

1. The rotational temperature is determined from the relative intensities of the ortho lines, assuming an OPR of three (merely for convenience as our library of g-factors is tabulated for OPR = 3). Ideally one would like to do the same using only the para lines. In practice, only a few para lines are observed in the spectrum (Figure 4, order 26) and they sample a small spread in rotational energies, preventing extraction for T_{rot} from the para ladder.

2. Line-by-line H_2O production rates are derived separately from the ortho- and from the para-lines. Their weighted means lines will be termed Q_{ortho} and Q_{para} respectively. Their uncertainties are also calculated in terms of standard errors and 95% confidence intervals (Appendix 4).

The methodology for extracting production rates will be presented in the following sections. It is enough to point out here that under our model assumptions, the production rate is proportional to the ratio of transmittance-corrected line flux and fluorescence efficiency. Therefore the quantity F/g (where g is in [W molecule^{-1}]) is evaluated from each individual ortho line, and separately from each individual para line. The ratio of the corresponding weighted means, $(F/g)_{ortho}/(F/g)_{para}$ equals the ratio Q_{ortho}/Q_{para}, while the relative errors of these means ($[\sigma_{F/g}/(F/g)]_{ortho}$, and $[\sigma_{F/g}/(F/g)]_{para}$) equal respectively the relative errors in Q_{orhto} and Q_{para}.

For the ortho-to-para ratio:

$$OPR = \frac{Q_{ortho}(T_{rot};OPR = 3) \times 3}{Q_{para}(T_{rot};OPR = 3)} \tag{4.1}$$

$$\sigma_{OPR} = OPR \times \left[\left(\frac{\sigma_{Q_{ortho}}}{Q_{ortho}} \right)^2 + \left(\frac{\sigma_{Q_{para}}}{Q_{para}} \right)^2 \right]^{1/2} \tag{4.2}$$

The multiplicative factor of three comes from the initially assumed value of OPR=3. Figure 12A and 12B show the derived (F/g) values respectively for the individual ortho and para features used in the OPR derivation for WM1. Figure 13 (A and B) presents the same for Q2. Separate lines are designated in these figures with their upper state energies. The idea is to demonstrate the spread in excitation separately for ortho and para features. In case of blends between lines of the same spin species, the summed g-factor has been used to calculate the corresponding F/g. The contributions of individual lines included in the blend have been weighted by the product of their respective g-factors and their monochromatic atmospheric transmittances. The upper state energies in case of a blend have been weighted the same way.

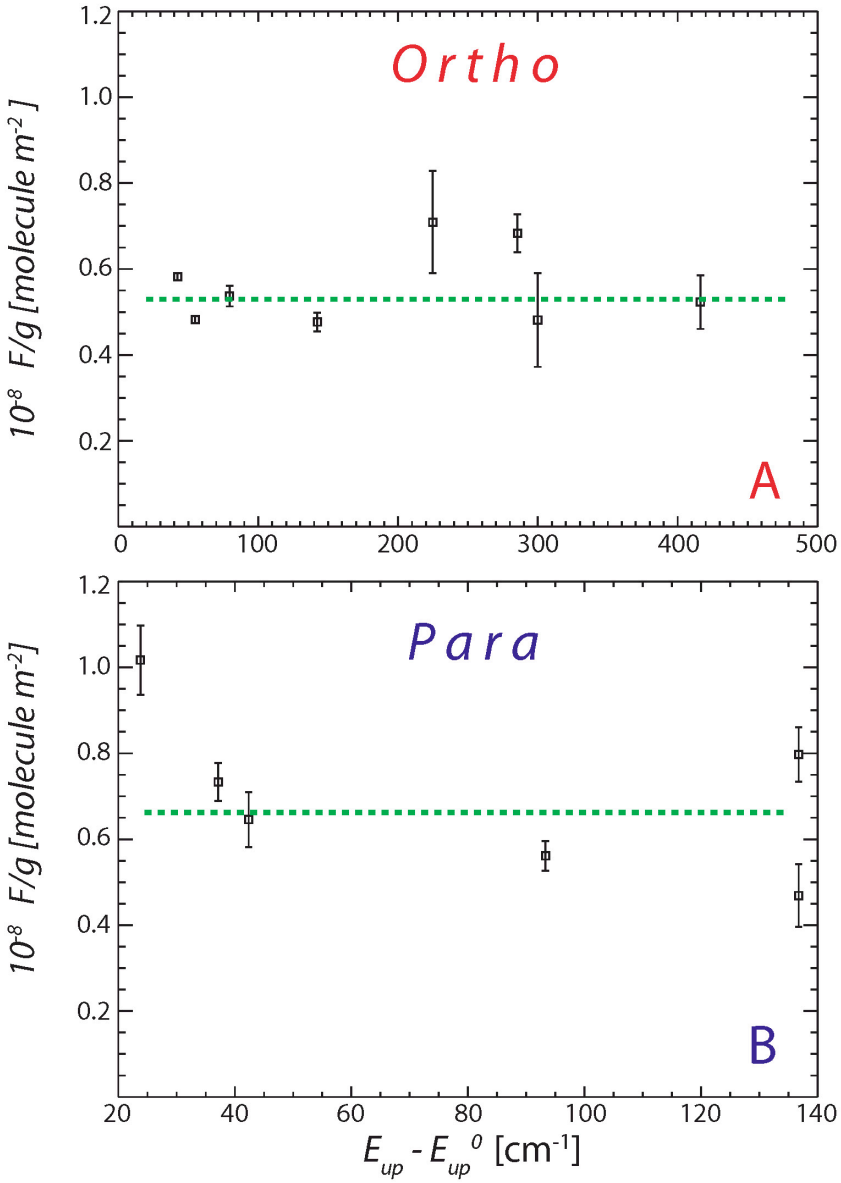

Fig. 12 – Line-by-line values of the line flux – to – g-factor (at 1 AU) ratio (proportional to the H_2O column abundance and production rate) for ortho and para lines used in the final OPR retrieval for WM1. Inclusion or exclusion of the outlier in the para sample (lowest energy point) does not influence the derived OPR.

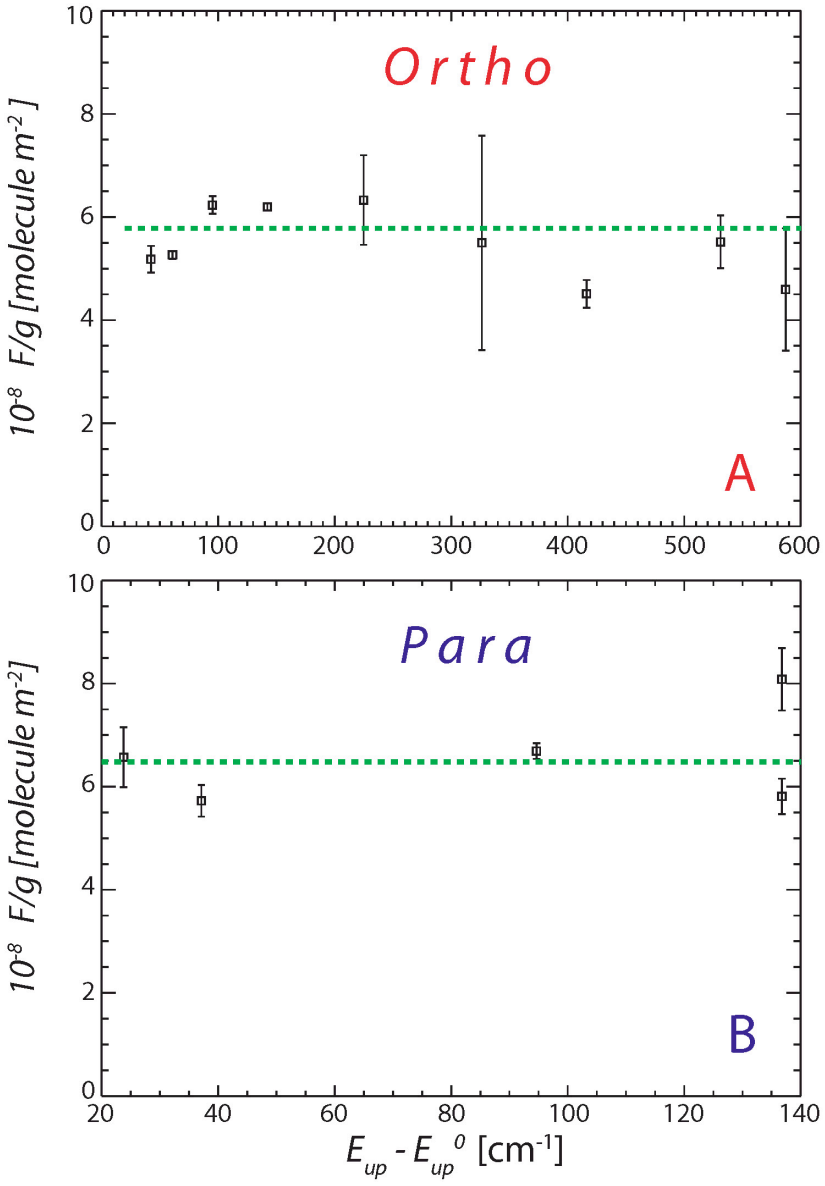

Fig. 13 – Same as Figure 12, but for Q2.

Note that the meaning of the (simple) error expressions given above will be considered later in conjunction with Figure 12 and 13 when discussing uncertainties.

With the above formalism, the OPR is calculated for a range of values of T_{rot}, chosen conservatively to exceed the formal 95% confidence intervals for the obtained rotational temperatures. For WM1 the OPR is calculated for T_{rot} = 70 K; 75K; and 80K. For Q2 the OPR is calculated for T_{rot} = 85K; 92K; and 100K. For 70 K and 80K (WM1) and 85 K and 100 K (Q2), the resulting slopes (F/g vs. E_{up}) are non-zero even above their formal 99% confidence limits. The retrieved OPRs are stable within these ranges, as seen in Table 1 which shows all results.

TABLE 1
H_2O ortho-to-para ratios

T_{rot} [K][a]	OPR
C/2000 WM1	
70	2.3 ± 0.2
75	2.4 ± 0.2
80	2.4 ± 0.3
C/2004 Q2	
85	2.7 ± 0.2
92	2.7 ± 0.1
100	2.6 ± 0.1

[a] The rotational temperature and its error are obtained from the relative fluxes of the ortho lines alone. The sensitivity of the OPR to rotational temperature is then evaluated in this table.

As in the rotational temperature retrieval, correlation analysis was used to eliminate obvious misfits in the model-data comparison before deriving an OPR. However, there is not much "latitude" for eliminating misfits from the para analysis, owing to the small number of lines. This is a limiting factor in the derivation.

The existing g-factors for OPR=3 are converted to g-factors for the optimal OPR:

$$g_{para}(T_{rot};OPR) = g_{para}(T_{rot};OPR=3) \times \frac{4}{OPR+1} \quad (4.3)$$

$$g_{ortho}(T_{rot};OPR) = g_{ortho}(T_{rot};OPR=3) \times \frac{4 \times OPR}{3 \times (OPR+1)} \quad (4.4)$$

Therefore for OPR=2.4 (WM1) the initial g-factors of the ortho and para lines should by multiplied respectively by 0.941 and 1.176. For OPR=2.7 (Q2) the corresponding multipliers are 0.973 and 1.081.

4.13 A Comment on the OPR Derivation

The previous sections describes the practical application for deriving an OPR from a set of g-factors for total production, which are pre-tabulated for a given OPR (in our case OPR = 3). In this section we will show that the described procedure is consistent with the strict definition of ortho-to-para ratio. First we should define some terminology. By definition the ortho-to-para ratio is:

$$OPR = \frac{Q^{ortho-H_2O}(g^{ortho-H_2O}[T_{rot}])}{Q^{para-H_2O}(g^{para-H_2O}[T_{rot}])} = \frac{Q^{total-H_2O} f_{ortho}}{Q^{total-H_2O} f_{pata}} \quad (4.5)$$

$$Q^{ortho-H_2O} + Q^{para-H_2O} = Q(total-H_2O) \equiv Q(H_2O) \quad (4.6)$$

where $Q^{ortho\text{-}H_2O}$ and $Q^{para\text{-}H_2O}$ are the production rates of orhto- and para-water respectively. The g-factors in the definition (4.5) do not relate to the total H_2O production, but to the production of ortho- and para-H_2O respectively (units of W [ortho-H_2O molecule]$^{-1}$ and W [para-H_2O molecule]$^{-1}$). The fluorescent efficiencies depend only on rotational temperature. We do not have their values available in our g-factor library. The total water production rate is designated as $Q^{total\text{-}H_2O}$ or shortly $Q(H_2O)$; f_{ortho}

and f_{para} are respectively the fractions of the total H_2O production rate for the two spin species.

In formula 4.1 O_{ortho} and Q_{para} designated the total (ortho plus para) H_2O production rate derived from the emission lines of ortho-H_2O and para-H_2O respectively. These production rates are determined from the fluorescent efficiencies for total water production (in units of W [H_2O molecule]$^{-1}$) for the ortho and for the para lines. These efficiencies depend on both T_{rot} and OPR and, as mentioned above, we have a library of their values for various rotational temperatures and OPR = 3. Formulae 4.3 and 4.4 show how the g-factors for total water production calculated for a given OPR and for OPR = 3 relate if the rotational temperature has been independently constrained. For two arbitrary model OPRs, x and y (4.3) and (4.4) translate into:

$$g_{para}(T_{rot};y) = g_{para}(T_{rot};x) \times \frac{x+1}{y+1} \tag{4.7}$$

$$g_{ortho}(T_{rot};y) = g_{ortho}(T_{rot};x) \times \frac{x+1}{x} \times \frac{y}{(y+1)} \tag{4.8}$$

In (4.7) the multiplier to the g-factor on the right-hand side represents the ratio between the fractions (of total H_2O) of para-H_2O for OPR = y and OPR = x (this fraction is 1/4 for OPR = 3). In (4.8) the multiplier is the ratio between the fractions of the ortho-H_2O (this fraction is 3/4 for OPR = 3).

From (4.7) and (4.8), and considering that the production rate is inversely proportional to the g-factors, it follows that for two arbitrary OPRs:

$$\frac{Q_{ortho}(T_{rot};x) \times x}{Q_{para}(T_{rot};x)} = \frac{Q_{ortho}(T_{rot};y) \times y}{Q_{para}(T_{rot};y)} = \frac{Q_{ortho}(T_{rot};y) f_{ortho}}{Q_{para}(T_{rot};y) f_{para}} \tag{4.9}$$

where f_{ortho} and f_{para} are the fractional populations for OPR = y. If y is the correct OPR for a given data set, the total water production rate deduced from the ortho-lines should equal to the total water production rate deduced from the para-lines. Therefore:

$$\frac{Q_{ortho}(T_{rot};x) \times x}{Q_{para}(T_{rot};x)} = \frac{Q_{H_2O}(T_{rot};y) \times f_{ortho}}{Q_{H_2O}(T_{rot};y) \times f_{para}} \equiv OPR \qquad (4.10)$$

To summarize starting from the strict definition of the OPR (4.5) we showed its equivalency to the practical application for deriving an OPR, when a set of g-factors for total H_2O production is available.

4.14 Spread of Excitation of the Sampled Para Lines

Aside from illustrating the OPR retrieval, the B-panels of Figures 12 and 13 have an additional implication. Note that the spread in excitation (represented by the upper state energy) is significantly lower for the para lines. This is the main reason for the inability to derive rotational temperatures from the para lines alone, as mentioned eralier. But after constraining T_{rot} from the ortho lines and extracting the OPR, both para and ortho lines can be included in "F/g vs. E_{up}" diagram for consistency check. For the WM1 and Q2 data, such consistency checks result in rotational temperatures that agree with those derived from the ortho lines alone.

4.15 Definition of a "Spherical" Production Rate

With the rotational temperature and OPR determined, the next step is to obtain line-by-line production rates for H_2O. The adopted g-factors have been calculated for the optimal values of these parameters as prescribed in the end of the section on OPRs above.

The production rate of a parent volatile is then given by the formula:

$$Q = \frac{4\pi\Delta^2 F_i}{f(x) g_i \tau}$$

F_i is the flux of the i^{th} H_2O line, at the top of the atmosphere. The g-factor, g_i (W molecule^{-1}) and the photo-dissociation lifetime of H_2O, τ (sec) are both calculated for heliocentric distance of 1 AU. Their product is independent on heliocentric distance ($g_i \propto R_h^{-2}$, $\tau \propto R_h^2$). The geocentric distance is Δ (m).

The quantity *f(x)* is the fraction of molecules expected in the sampled region of the coma. This fraction, as mentioned above, corresponds to the closest ~230 km from the nucleus for WM1 and the closest ~250 km for Q2. Recall these are the central parts of the spatial profiles (Figure 6) with highest S/N. The parameter *x* is the fraction of the H_2O dissociation scale length covered by the spatial extent of the sampled coma (Hoban et al. 1991). The *f(x)* values are available tabulated and are ~10^{-3}, 10^{-4}. For such a small spatial coverage $f(x) \propto (\tau V_{gas})^{-1}$, where V_{gas} is the gas outflow velocity. The formula for production rate (Q) contains the product of τ and *f(x)*. Therefore the derived Q is not very sensitive to the assumed lifetime, but it is sensitive to the outflow velocity (which we cannot measure). We use a value of $\tau = 7.7 \times 10^4$ sec^{-1} for H_2O at 1 AU (Crovisier 1989) and $V_{gas} = 0.8\ R_h^{-0.5}$ km sec^{-1} (the heliocentric distance is in AU). This

approximates the gas outflow velocities measured by radio techniques for comets with similar productivity (e.g. Biver et al. 1999) and has been used for all comets in our database excepting C/1995 O1 (Hale-Bopp), which had unusually high gas production rate (Mumma et al. 2003).

The main assumptions of our methodology are a steady state (the production rate does not vary over the parent lifetime) and a spherically symmetric outflow from the nucleus with a constant velocity. The assumption for a steady state is usually a good approximation, except for special cases like the Deep impact event or natural outbursts (e.g. Hughes 1991), when time variability on a very short time scale characterizes the gas production.

The production rate defined above is called a "spherical production rate", because of the assumption for the outflow of the sublimated molecules. The spherical production rate obtained from the flux measured at the region in the spatial profile closest to the nucleus is referred as to "nucleus-centered". Experience has defined the optimal choice for this nucleus-centered region: it covers ~1.8 arc-seconds along the slit (i.e. ~0.9 arc-seconds on each side of the nucleus). This corresponds to an extract of 9 spatial rows.

4.16 Line-by-Line Nucleus-Centered Production Rates

Whenever possible the production for a given molecule should be quantified independently from each individual detected emission feature. The final production rate is calculated as a weighted mean of the individual measurements – the same way as during the OPR retrieval. The differences between the two procedures are:

1. In the OPR retrieval, weighted mean H_2O production rates are calculated

separately from the ortho and from the para emissions; ortho-para blends are excluded. On the other hand, in the final line-by-line production rate all lines available can be included in the weighted mean.

2. Weighted means for Q_{ortho} and Q_{para} are calculated assuming OPR=3. Only when this assumption holds, $Q_{ortho}=Q_{para}$. Conversely, the final line-by-line production rate is calculated using the optimal values of both T_{rot} and OPR. In the absence of modeling issues (e.g. H_2O fluorescence model; terrestrial atmospheric model fit subtracted to obtain residual spectra) all individual Qs should agree within the uncertainty imposed by the stochastic noise.

After calculating a weighted mean Q, stochastic and standard errors should be evaluated independently (Appendix 4). As already pointed out, the former uncertainty reflects the S/N in the data, while the latter uncertainty reflects the level of disagreement among line-by-line evaluations of the sampled quantity (here the production rate).

Nucleus-centered production rates have been retrieved for each H_2O emission feature detected reliably in the two data sets (WM1 and Q2). The results are displayed on Figure 14 in which a given transition is designated with its corresponding upper state energy.

The weighted mean production rates obtained are:

$Q_{nc} = (1.38 \pm 0.04) \times 10^{28}$ molecules sec^{-1}, for WM1

$Q_{nc} = (1.43 \pm 0.04) \times 10^{29}$ molecules sec^{-1}, for Q2

Both adopted errors are standard, i.e. the spread of the individual Qs around their mean dominates over the S/N noise in the data. The subscript "nc" stamds for "nucleus-

centered". These values for "Q_{nc}" are a subject to a systematic effect that will be evaluated in the next two sections.

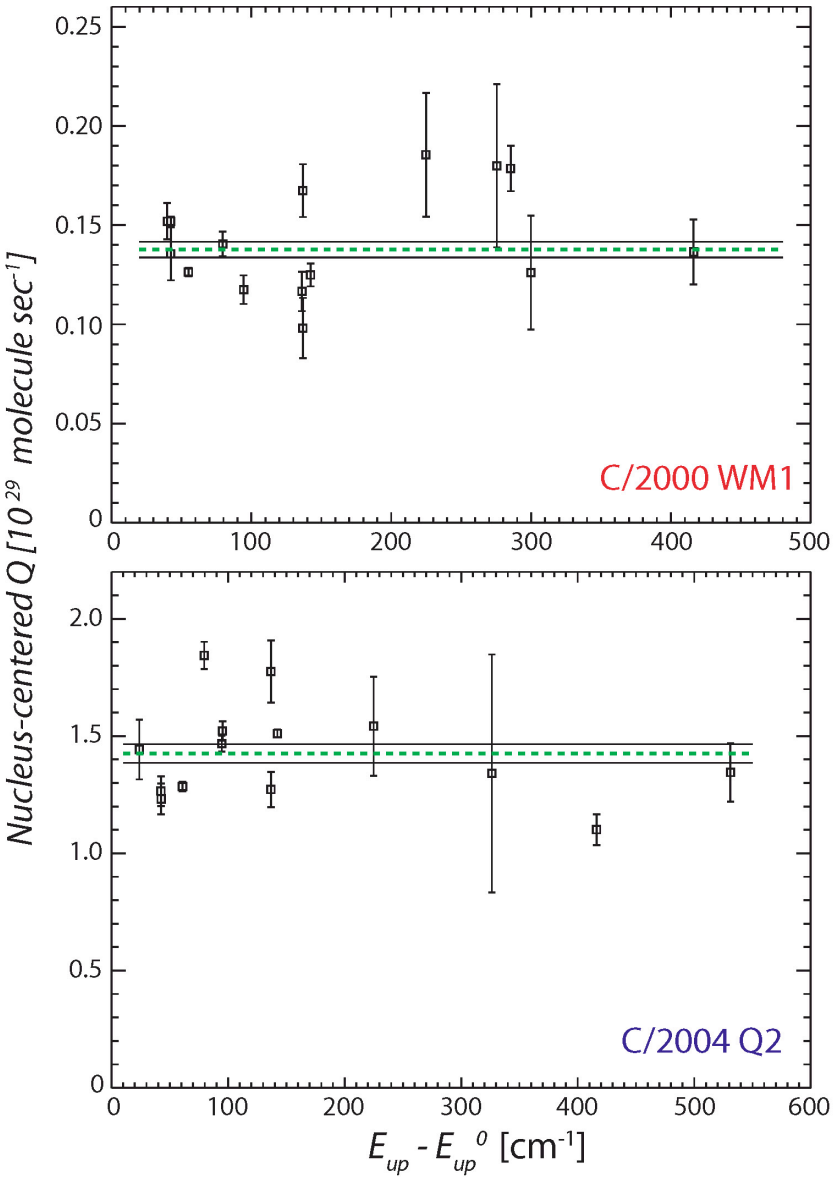

Fig. 14 – Line-by-line nucleus-centered H_2O production rates in two comets. One standard error of the weighted mean is shown.

4.17 "Symmetric" Production Rate, Q-curve, and "Terminal" Production Rate

The spatial distribution of the H_2O emission intensity was introduced in the beginning of this chapter (Figure 6). All subsequent analyses up to this point were done for the central parts of the spatial profiles: the first ~230 km from the nucleus of WM1 and the first ~250 km from the nucleus of Q2. These ranges correspond to spectral extracts including the flux summed over the central (with respect to the peak gas intensity) 9 rows of the fully processed echelle order (see Appendix 2; Figure AP2-1D).

Deriving nucleus-centered production rates has the advantage to sample the region with the highest S/N ratio. This approach is adopted for weak to moderate comets nodded on chip. The nucleus-centered production rates however are systematically underestimated. The reason is that the flux measured in the nucleus-centered region is most affected by seeing (this is found to be the primary factor in most cases) and other effects related to real observing conditions like sub-optimal focus and drift perpendicular to the slit (slit loses; Dello Russo et al. 1998, 2000). The underestimating of nucleus-centered production rate is well understood, based on work with numerous data sets. The required correction is achieved by introducing the concepts of a "symmetric" production rate, a "Q-curve", and a "terminal" production rate, which will be defined in sequence:

Symmetric production rate:

If the spatial profile has adequate S/N, spherical production rates can be calculated from emission fluxes measured away from the nucleus. A "symmetric" production rate is defined as the mean of two spherical production rates taken at the same projected distance from the nucleus on both sides on the spatial profile. The term

"symmetric" is used because the procedure minimizes the effects of asymmetries in the gas outflow. Modeling shows that this approach is a good approximation within first order (Xie & Mumma 1996).

Such asymmetries are sometimes important - mainly for comets with small active surface area and low gas productivity. Such are commonly the Jupiter-family and Encke-type comets. Their gas productivity has been reduced as a result of many passages in the inner solar system resulting in the development of a dust "mantle". For Oort cloud comets with moderate and high productivity (Q exceeding ~10^{28} molecules sec^{-1}) the approach of taking symmetric Qs has proven to be very useful (cf. Mumma et al. 2003).

Q-curve:

The trend of the spherical production rates (Qs) with projected distance from the nucleus defines a Q-curve. Symmetric spherical production rates are usually used in Q-curves. Such Q-curves are shown on Figure 15. These Q-curves are derived from the two spatial profiles displayed on Figure 6, which contain the combined flux of many H_2O lines. Q-curves for individual lines are much less reliable because of lower SN ratio.

"Terminal" or "Global" production rate:

Note that the symmetric spherical production rates reach a terminal value outside ~0.5-1 arc-seconds from the nucleus. These spherical Qs are significantly less sensitive to seeing compared to the nucleus-centered values. The weighted mean of symmetric Qs measured in the spatial region where the production rate reaches its terminal value is called "global" or "terminal" Q. The terminal regions in the Q-curves of WM1 and Q2 are shown with dashed horizontal lines in the two plots on Figure 15.

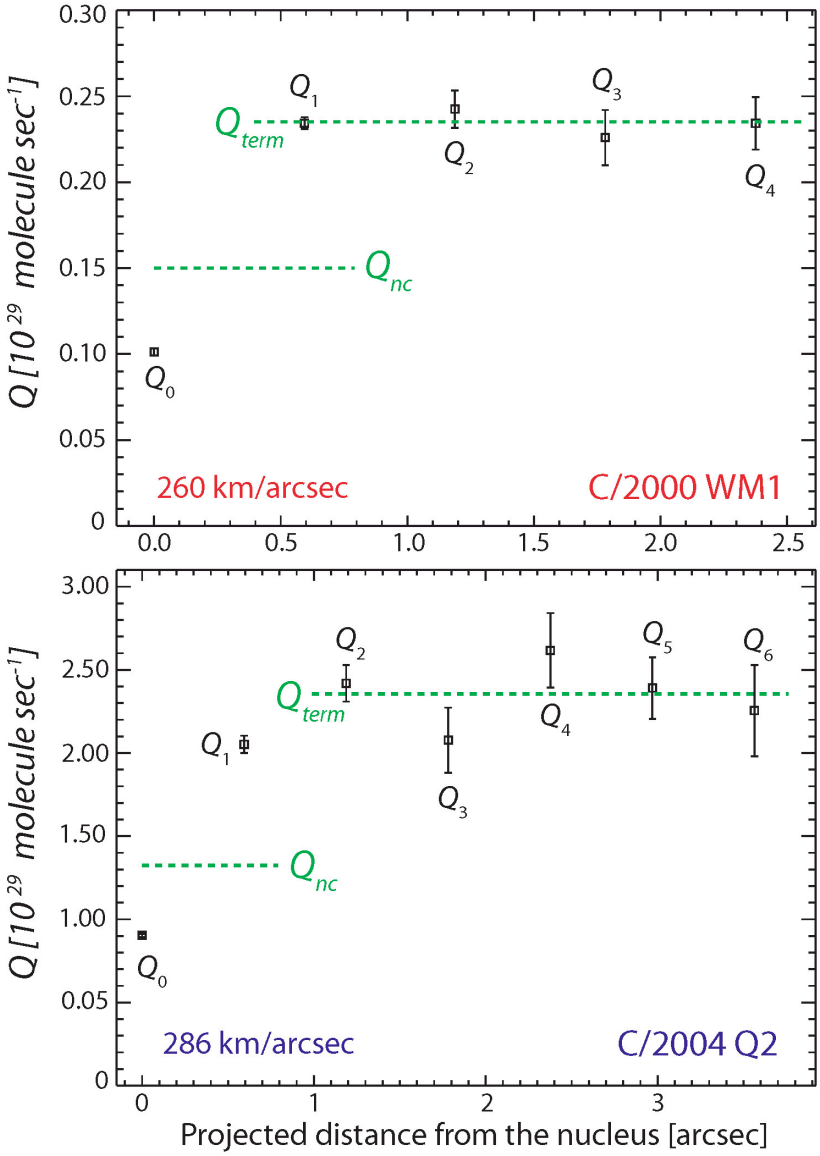

Fig. 15 – Q-curves obtained from the combined H_2O spatial profiles shown on Figure 6. The Q-curve is the trend of the spherical symmetric production rates (Q_i, i=0, 1, 2 …) with projected distance from the nucleus. Each spherical symmetric Q_i corresponds to flux measurement over three spatial pixels (0.198 arcseconds per pixel). For $i > 0$ the measured flux is averaged over two three-pixel extracts equidistant from the nucleus on both sides. For this reason only positive distances are shown in the x-axis. The "terminal" or "global" production rate (Q_{term}) is reached at different projected distances depending on seeing, which was better for the WM1 observation. The extent over the spatial profile for which symmetric Qs are calculated depends on the S/N in the profile. For comet WM1 Q_{term} represents the weighted mean of Q_1, …, Q_4. For comet Q2 – the weighted mean of Q_2, …, Q_6. Note that the nucleus-centered production rate (Q_{nc}) used in quantitative analysis is not Q_o, but the spherical production rate based on the flux measured in the central nine spatial pixels, i.e. about ± 0.9 arcseconds from the nucleus. This measurement is more suitable for line-by-analysis than Q_0. The growth factor is defined as the ratio between the Q_{term} and Q_{nc}.

Why the spherical Qs outside the nucleus-centered region are much less sensitive to seeing and drift effects? Consider the expression for a spherical production rate. In particular - how the line flux (F_i) and the parameter *f(x)* (the fraction of molecules expected in the sampled region of the coma) change with projected distance from the nucleus. In the inner part of the spatial profile they decrease rapidly. The line flux is a measurement influenced by atmospheric seeing or comet drift, while *f(x)* is a modeled parameter always assuming perfectly centered and focused target. If there is a small offset in centering the comet in the slit or the measured flux is smeared by seeing, this flux would not correspond exactly to the value of *f(x)* assumed for ideal conditions. Because the off-centered flux is smaller, their ratio and hence the derived Q would be smaller. On the other hand, away from the nucleus the gradients in line flux and *f(x)* are greatly reduced (see appendix of Hoban et al. 1991). If atmospheric seeing, comet drift or telescope de-focusing offset or smear the measured flux, the resulting error in the ratio $F_i/f(x)$ would be much smaller than in the nucleus-centered case.

An additional factor which "stabilizes" the production rates in the terminal region of the Q-curve is the use of symmetric spherical Qs. A slight image off-centering would produce opposite changes in the $F_i/f(x)$ ratio on both sides on the spatial profile, and therefore the overall effect is reduced by in the symmetric production rate.

4.18 Growth Factor and Correction of the Nucleus-Centered Production Rates

The ratio between the terminal (Q_{term}) and the nucleus-centered (Q_{nc}) production rates, both deduced from the combined spatial profile, is called a "growth" factor (GF_{H2O}):

$$GF_{H2O} = \frac{Q_{term}}{Q_{nc}}$$

For WM1 and Q2 the growth factors are respectively 1.55 ± 0.03 and 1.78 ± 0.06, consistent with the better atmospheric conditions during the WM1 observation, reported in the observing log.

It should be again emphasized that the Q-curves are based on spatial profiles which contain the combined flux from multiple H_2O lines in order 26 (Q-curves based on individual lines are less reliable because of insufficient S/N). In this framework the spherical production rates calculated at different projected distances use the same g-factor. Its value is taken to be the sum of the individual g-factors for the lines included in the Q-curve. At constant g, the derived growth factor is independent of the initial adopted value of fluorescence efficiency which simply cancels out in Q_{term}/Q_{nc}. However a detail, which is well understood but almost never mentioned, is that the g-factors could vary as a function of distance from the nucleus, because the rotational temperature may not be the same (and often cannot be constrained in the wings of the spatial profile as well as in its center). Fortunately in practice the change in spherical Q's along the spatial profile is primarily effected by observing related effects - the ones we would like to quantify. On the other hand, the g-factor, summed over many water lines is much less

sensitive to moderate changes in rotational temperature (as T_{rot} varies some of the individual fluorescence efficiencies increase, while others decrease, depending on lower state energy). Therefore the assumption of a constant g-factor along the Q-curve is quite reasonable for the purpose of deriving a growth factor.

Another aspect to comment on is that the nucleus-centered production rate from the combined flux and g-factor for all lines cannot substitute the weighted mean of the individual nucleus-centered Qs. Line-by-line measurements are always more reliable. One of the reasons is that summing the g-factor hides any model uncertainties.

4.19 H_2O Production Rates in C/2000 WM1 and C/2004 Q2

This is the point to summarize the approach of obtaining production rates from data nodded "on-chip":

1. Nucleus-centered Qs are calculated independently from each individual H_2O emission feature detected. This calculation benefits from using the highest S/N data. At the same time the nucleus-centered values are affected by atmospheric seeing and possibly other observing-related effects. These effects are systematic - they are the same for all lines detected within a single instrument setting.

2. The weighted mean of individual measurements is considered to be the representative nucleus-centered H_2O production rate.

3. A spatial profile containing the summed flux of most (or all) H_2O lines within a certain order is constructed. A Q-curve from this high S/N profile is used to calculate a terminal production rate and a growth factor. The value of the growth factor is practically independent on the assumed effective g-factor for the combined emission

from multiple lines.

4. The weighted-mean nucleus-centered production rates are multiplied by the corresponding growth factors, thereby being corrected for the atmospheric seeing effect.

Figures 14 and 15 (described above), and Figure 16 summarize the analysis of H_2O hot-band emission near 2.9 μm. Note, that most of the methodology presented is not applicable just to the case of H_2O emission but to parent volatiles in general.

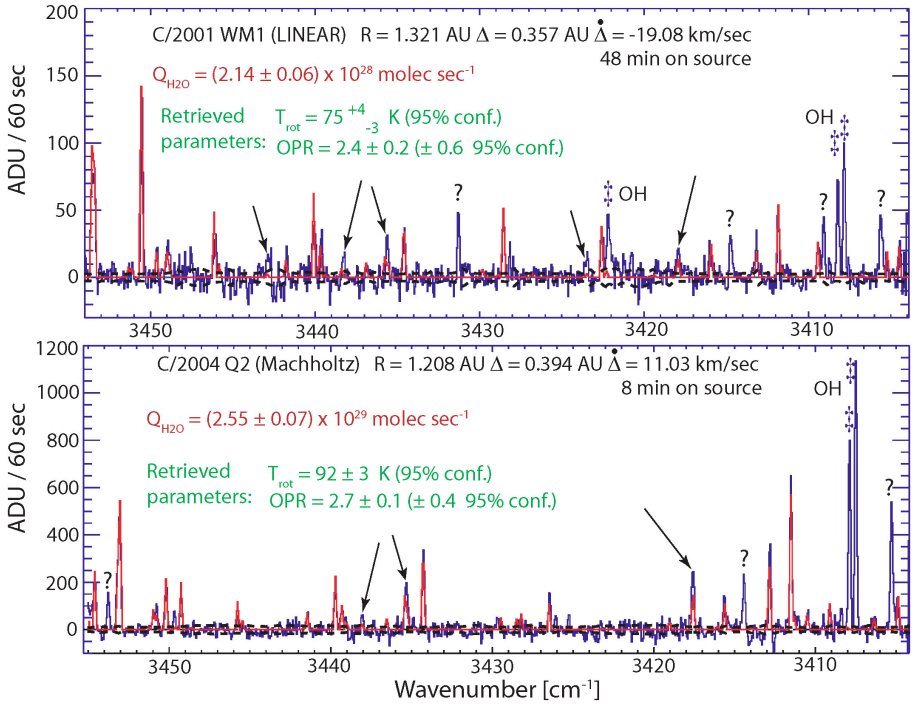

Fig. 16 – H_2O fluorescence models for the retrieved T_{rot} and OPR values (red) tested against the observed spectra (blue). The $\pm 1\sigma$ photon noise enveloped is shown with black dashed curves. The pointers are to features that are not fit well by the model at any rotational temperature. The "standard" error (as defined in the text) is given for H_2O production rates, because this is the error to propagate in determining mixing ratios between other species detected in the KL2 setting and water.

Figure 16 "condenses" the whole chapter into a single two-panel plot. It shows the gist of the observing log (helio- and geo-centric distance, radial velocity with respect to the observing site, and time on source); the retrieved rotational temperatures; ortho-to-para ratios; H_2O production rates; and most importantly, best fit H_2O fluorescence

models compared with the observed spectra. Along with the line-by-line nucleus-centered Qs (Figure 14) and the Q-curves used to correct them for atmospheric seeing (Figure 15), the core analysis for quantifying water production is revealed. The rest of this chapter is left for the discussion of some important implications, including the problem of H_2O spin temperatures outlined in the introduction as possible cosmogonic parameters.

4.20 The Importance to Quantify Model Uncertainties

If possible, I consider presenting line-by-line results for H_2O production rates mandatory. One could certainly use a Q-curve from the combined flux of many H_2O lines, not just to estimate the seeing effect but to derive a H_2O production rate. Its value would be adequate, provided one had a reasonable estimate of the total fluorescence efficiency. However, summing the flux from multiple non-blended lines washes out the critical issue in this analysis: to what degree individual emission features are modeled correctly. We present below some discussion of this point.

Astronomical spectroscopy imposes two very basic constraints on data interpretation. The first is that a secure detection of any species is accomplished through identification of its multiple (at least two in the worse scenario) lines. In the case of H_2O non-resonance fluorescence this requirement is fulfilled as evident from the comparison between model and data on Figure 16.

Second, and equally important, is the condition that the abundances of the species of interest deduced independently from different emission features agree. The extent of disagreement should by all means be incorporated in the uncertainties, as done here with

the “simple” standard error of the mean (Appendix 4). Recall that this error reflects exactly the spread (i.e. the disagreement) of individual measurements. Also note that individual error bars of the production rates on Figure 14 reflect the ±1-σ level of photon noise, indicated with black dashed lines on Figure 16. Some emissions have really high S/N but their analysis gives production rates that are not fully consistent. The biggest outliers are actually not included in the OPR derivation (Figures 12-13) and in the weighted mean *Q* and in Figure 14 they, but are seen in the model-data comparison. Examples of these are the lines marked with arrows on both panels of Figure 16. They seem either blended with unknown features or have underestimated g-factors. The correlation analysis introduced earlier is a very helpful tool to easily identify such obvious misfits. They have been excluded from further quantitative analysis.

In addition to problems related to wrong g-factors or to blends, the standard error also reflects inaccuracies in the terrestrial atmospheric model fit subtracted from the original data in order to obtain a residual spectrum of cometary emission (see Chapter 3).

I will conclude this section by explicitly showing the “simple” and “well-known” formulae evaluating the “stochastic” and the “standard” error of the weighted mean production rate. These errors are defined in Appendix 4, but it is important to underline their different meanings here:

"Stochastic" error of the mean (following Bevington & Robinson 1992):

$$\sigma_{stoch} = \sqrt{\frac{1}{\sum_i \frac{1}{\sigma_{Q_i}^{\ 2}}}}$$

"Standard" error of the mean (following Hoel 1984):

$$\sigma_{std} = \frac{s}{\sqrt{N}}, where$$

$$s^2 = \frac{1}{N-1} \frac{\sum_i \frac{(Q_i - Q_{mean})^2}{\sigma_{Q_i}^2}}{\frac{1}{N}\sum_i \frac{1}{\sigma_{Q_i}^2}}$$

Q_i is the production rate derived from the ith individual line; σ_{Qi} is its uncertainty, derived from photon noise; N is the number of lines sampled; and s^2 is the variance for the sample of production rate measurements.

The "stochastic" (sometimes referred as "reciprocal") error definition has nothing to do with the agreement of line-by-line production rates. Statistically it says only that there is a well-defined mean, which physically translates into high S/N (very important of course), allowing accurate flux measurements. The stochastic error does not provide any information about how accurately these measured fluxes can be interpreted. Obviously strong but not well-modeled lines would still result in a small σ_{stoch}, thereby underestimating the actual uncertainty. This is when σ_{std} and the related to it sample variance (s^2) become useful.

In the current level of fluorescence modeling of the H_2O hot-bands we find that $\sigma_{std} > \sigma_{stoch}$, therefore the "standard" uncertainties are the ones being reported (Dello Russo et al 2004, 2005; this work; Mumma et al. 2005).

4.21 Comets as "Ad-Hoc" Laboratories to Test Fluorescence Models

The preceding section discussed the accuracy of modeling line-by-line

measurements in terms of uncertainty of the retrieved production rate. Earlier in this chapter another important aspect of comparing model to data was mentioned. Comets are used as "ad-hoc" laboratories to test and eventually improve parameters of H_2O fluorescence models, branching ratios from the excited ro-vibrational states in particular. Most of the observed non-resonance transitions cannot be reproduced in the laboratory, so comet data become even more valuable.

The fluorescence models for the fruitful 2.9 μm region are relatively new and have been tried on only a few comet data sets. The first test field were CSHELL spectra from comet 153P/Ikeya-Zhang (Dello Russo et al 2004), followed by NIRSPEC spectra from C/1999 (Lee), C/1999 S4, and C/2001 A2 (Dello Russo et al. 2005). Our project adds two more premium data sets, in some aspects superior to the preceding ones. The Ikeya-Zhang data were taken with CSHELL, which limits severely the number of H_2O lines that can be detected simultaneously. Two out of three comets in the study of Dello Russo et al. (2005) (which yours truly is also part of) were observed at unfavorable atmospheric conditions, namely very high H_2O vapor content in the terrestrial atmosphere. The 2.9 μm region is sensitive to atmospheric extinction, even though the observed transitions do not have direct counterparts in telluric absorption. This resulted in lower S/N spectra and limited the number of lines detected.

Obviously high S/N data is optimal for comparison with fluorescence models. The WM1 data benefit greatly from significant time on source and very low atmospheric H_2O vapor content during the observation. The Q2 data are the best in terms of S/N because of the relatively strong source.

Considering the complexity of ab initio calculations involved in the modeling of

the H_2O energy levels and non-resonance transitions, the overall agreement between model and actual spectra is quite good (Figure 16). However, as discussed above, the accuracy of these new models have not reached the photon noise limit yet. To further improve the precision of the derived g-factors, we identify the lines that are not fit well in each comet and look for systematic deviations between predicted and observed intensities. For example the lines marked with black arrows are misfit in a consistent way in both Q2 and WM1. Our purpose is to reach a point we will be able to estimate "corrected" g-factors for the deviant lines from observed systematic discrepancies seen in multiple data sets. In parallel to our expanding of the test data set by observing more comets, the collaborating UC London group improves the complexity of their theoretical predictions of energy levels and transition probabilities. The effort to improve the model quality comes from both ends in a coordinated fashion.

Interpretation of the unknown features is also a goal. The same unidentified lines have been seen in different comets. They may belong to H_2O or another molecule. The unidentified line near 3431.2 cm^{-1} seen in the WM1 spectrum is not seen in the Q2 spectrum. In principle, this suggests that the line is a not H_2O; however in our case it is simply an issue of increased terrestrial transmittance owing to different Doppler shift of Q2.

Finally it should be mentioned that the correlation analysis as a tool for testing fluorescence models was introduced in this work. While bringing no conceptual insight, it proved to increase the efficiency in evaluating the accuracy of the synthetic spectra. The practical application of the procedure minimizes significantly the time necessary to compare models to data and increases the quality of the comparison.

4.22 Other Types Uncertainties in the Retrieved Parameters

After discussing the stochastic and model uncertainties in some detail, it is appropriate to rank the other sources of error in the H_2O production rate:

Rotational Temperature and OPR:

The rotational temperature is well constrained mainly because the sampled transitions of ortho-water cover a wide range of rotational energy levels (both in the ground and excited vibrational states). The rotational temperature uncertainty is not a significant source of error in the derived ortho-to-para ratio. The OPR was shown to be stable even at a very conservative choice for the T_{rot} uncertainty (Table 1).

The water production rate is more sensitive to variation in the rotational temperature than in the OPR. Fortunately T_{rot} is well constrained and within its imposed limits it is not a considerable factor in the overall uncertainty. This is consistent with Dello Russo et al. (2005). For example varying the rotational temperature between 80 and 90 K does now change the water production rate in WM1 beyond the limits imposed from its standard error.

The effect is similar upon varying the OPR. In some cases there are not enough para lines detected to constrain this parameter at all, but assuming OPR=3 generally provides adequate results (Dello Russo et al. 2004; Mumma et al. 2005). The production rate is inversely proportional to the assumed g-factor. Its dependence on the OPR is weaker than its dependence on T_{rot}. For example, changing the OPR from 3.0 to 2.4 reduces the fluorescence efficiencies of the ortho lines by less than several percent and of the para lines by ~10 %. But the changes are in the opposite directions, so the overall production rate is only weakly affected and definitely within the uncertainties in the line-

by-line values of $Q(H_2O)$. On the contrary, the uncertainty in the OPR has much more serious effect on the precision in the derived spin temperature of cometaty H_2O which is a subject of a follow up section.

Growth factor:

The growth factors represent a minor source of uncertainty. They are well constrained because they are derived from Q-curves of high S/N spatial profiles (Figure 15). The error in the growth factor is usually within ~5% and is dominated by the precision in deriving the symmetric production rate in the terminal region. In constras to the standard error of the weighted mean production rate, the uncertainty from the growth factor is systematic. It affects all individual nucleus-centered production rates within a given instrument setting in the same way and therefore does not affect mixing ratios of trace constituents with H_2O.

Flux calibration:

The dominant source of uncertainty in $Q(H_2O)$ is in most cases the one related to flux calibration (Appendix 2). It is also the one most difficult to quantify. Fortunately the flux calibration error is also a systematic within a given NIRSPEC setting and does not affect mixing ratios.

Summary of the production rate uncertainties for WM1 and Q2 (all results are in units of molecules sec^{-1}):

Standard error of a weighted mean Q:

$$Q_{WM1} = (2.14 \pm 0.06) \times 10^{28}$$
$$Q_{Q2} = (2.55 \pm 0.07) \times 10^{29}$$

The overall spread of line-by-line production rates included in the mean can be estimated from the product of the quoted standard errors and square root of the number of lines in

the sample.

Systematic error from the uncertainty in the growth factor:

$$Q_{WM1} = (2.14 \pm 0.04) \times 10^{28}$$
$$Q_{Q2} = (2.55 \pm 0.09) \times 10^{29}$$

Systematic error from the uncertainty in the flux calibration (this is a reasonable conservative estimate, not a hard number, as discussed in Appendix 2):

$$Q_{WM1} = (2.14 \pm 0.19) \times 10^{28}$$
$$Q_{Q2} = (2.55 \pm 0.26) \times 10^{29}$$

For calculating mixing ratios of other parent volatiles with H_2O the standard error is the one to propagate if working within the same instrument setting. Simultaneous observations of H_2O and other molecules are commonly achieved with NIRSPEC. For this reason, the nature of the standard error has been discussed in detail above.

For calculating mixing ratios of native volatiles observed with instruments/facilities other than NIRSPEC/Keck 2, it should be considered that systematic effects can be produced by various factors related to different instrumentation, modeling assumption used to derive production rates, and spatial extent of the sampled coma. Therefore such comparison is generally less reliable.

4.23 Spin Temperature of H_2O and Interpretation of its Formal Confidence Interval

The conversion between OPR and H_2O spin temperature is presented graphically on Figure 17 for WM1 and Q2 respectively. This plot indicates the values of the found OPRs, their standard errors, and their formal 95% confidence intervals. In the theoretical curve shown, the OPR is defined as the ratio between the ortho and the para partition functions (also known as statistical sums) for a given spin temperature. This plot assumes thermal equilibrium. Under this hypothesis, T_{spin} is interpreted as the formation temperature of the H_2O molecule.

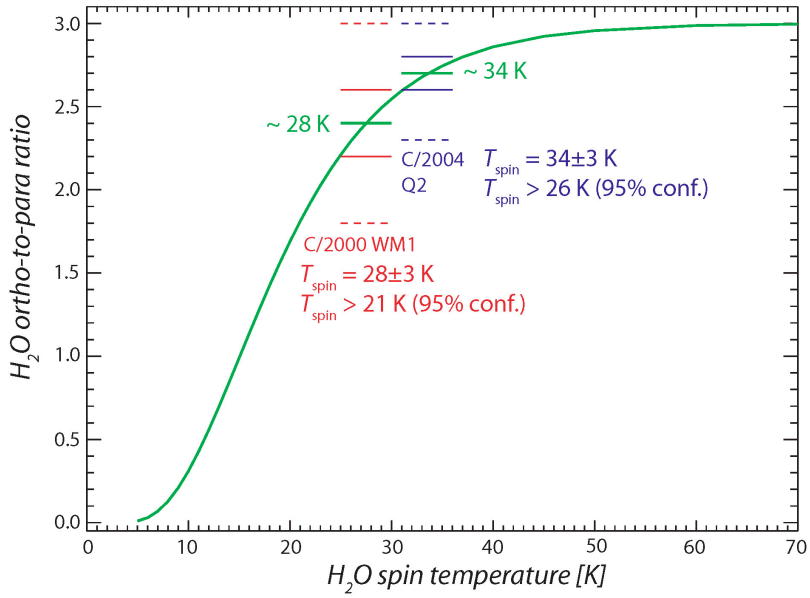

Fig. 17 – Conversion between OPR and spin temperature. The green curve represents the ratio between the total ortho-H_2O and para-H_2O partition functions for a given temperature. The OPR results for WM1 and Q2 are shown with their standard error bars (solid horizontal lines) and 95% formal confidence intervals (dashed lines). Only the former uncertainty is a "hard" number, as discussed in the text.

We should first discuss the importance to include realistic (to our best knowledge) uncertainties in the OPRs and spin temperatures. What do the two types of error intervals shown on Figure 17 mean? The answer is coded into Figures 12 (WM1) and 13 (Q2). As a reminder, these figures display line-by-line measurements of the ratio (F/g), assuming OPR=3. This ratio is essentially a scaled production rate. The way in which the OPR and its uncertainties are calculated have been described in the section "Ortho-to-para ratio of the H_2O molecule." From the definitions in that section it follows that the precision in the OPRs is limited by the scatters around the mean values in Figures 12 and 13. The whole discussion about stochastic vs. standard errors from line-by-line measurements holds here – for a given spin species, it is desirable that the measurements of the same quantity (F/g, respectively Q_{ortho} or Q_{para}) deduced using different lines should be the same. The ideal situation is when all measurement on Figure 12 and on Figure 13 agree within their stochastic errors. But again, at the current level of modeling the g-factors, the photon noise is still not the limiting uncertainty factor.

How do a given error and confidence interval is being assigned represents a critical issue everywhere, but especially for OPRs. One "nuance" to notice is that all uncertainty formulae assume a given distribution of errors. Except for "clear-cut" cases like stochastic noise, this distribution is probably not well (or at all) known. But this is certainly not an excuse to abandon the error calculations - like the often idealized physical models, the error derivations also contain assumptions. The important thing is that when we calculate a standard error, it is unambiguously defined with a meaningful expression. The latter serves its purpose to describe the scatter in multiple measurements of the same quantity via the $(Q_i-Q_{mean})^2$ term.

Unlike the standard error, the so called 95% confidence interval is not a "hard number" given by a well-understood expression. The rule of thumb for this interval is that it corresponds to two (1.96 ~ 2.0) standard errors. However astronomers often deal with small number statistics - the case when the multiplier in front of the standard error exceeds its "conventional" value of 2 (see Appendix 4). We have adopted such formalism for small samples with the clear consciousness that we do not know the distributions of errors leading to the multiplier in front of the standard uncertainty. Therefore the formal 95% confidence interval that accounts for the sizes of samples of ortho and para lines is not considered an absolute quantity (like the standard error) but an indicator for the possibility that the small sample of lines may bias our final result. The key word is "possibility" in contrast to "certainty". The OPR may be on principle be well constrained from a single ortho-para pair of lines, but there is no way to be certain about that, unless this result is confirmed by independent measurements of more lines. One specific example:

The WM1 OPR is 2.4 ± 0.2 based on its standard error. This implies that it is relatively well-constrained within the sample of the lines available for analysis and it implies a spin temperature below the equilibrium value of ~50K. The significantly larger 95% formal confidence interval does not indicate that our measurement is not well-constrained. Instead, it indicates that it is highly desirable to confirm this result based on analysis of a broader frequency range, covering more H_2O emission features. Confirming the same result on any measurement from multiple lines is a key in spectroscopy. In this work, we are mostly limited by the small number of para lines detected.

Therefore assigning an absolute meaning to our OPR formal confidence limits

should be avoided. Quite often I have noticed similar tendency in works of various nature (planetary, solar, geophysical). Putting absolute meaning to statistically derived confidence intervals usually goes into extremes by following the logic that a certain statistical criterion is (is not) fulfilled, therefore the resulting conclusion is (is not) valid, beyond any conditions. Case closed! We would like to avoid such extremes in our OPR (and other) evaluations. These approaches ignore the notion that every statistical criterion should be interpreted only within the framework of its underlying hypothesis. It is then a matter of case-by-case consideration to what extent a certain statistical test is useful and what are its limitations.

To summarize, by considering only "stochastic" errors, our resulting uncertainties in OPR and respectively spin temperature will be seriously underestimated. Additional risk relates to the usage of limited samples of ortho and para lines. Because the T_{spin} is considered an important cosmogonic observable (see Chapter 2), underestimated errors open the door for over-interpretation in terms of possible formation temperatures and mechanisms for the cometary ices. In practice, a theorist who models the formation conditions of cometesimals is not necessarily required to check the meaning of the reported uncertainties of a given OPR. Therefore, it is a responsibility of the observer to declare the sources of error clearly (for example with Figures like 12 and 13), to include uncertainties beyond the photon noise when appropriate, and to keep a high-level of caution when a very limited number of lines is involved into an OPR retrieval.

Finally, note that the same uncertainty in the OPR translates into a different uncertainty in T_{spin} because the relation between these two quantities in not linear (Figure 17).

4.24 H_2O Spin Temperatures in C/2000 WM1 and C/2004 Q2

The above considerations lead to the following interpretation of our spin temperature uncertainties:

1. The most probable spin temperature ranges measured in WM1 and Q2 are respectively 28±3 K and 34±3 K. These ranges are well-constrained within the limits of our sample of ortho and para lines.

2. The OPRs deduced exclude spin temperatures significantly lower than 21 K for WM1 and 26K for Q2.

3. Mainly because of the very limited number of para lines involved in the OPR calculation, spin temperatures above 50 K (consistent with equilibrium) are not completely excluded, but are well outside the aforementioned most probable ranges.

4.25 Comparison with Results from Other Comets

It is most appropriate to discuss these new results with similar measurements in other comets. Table 2 summarizes H_2O OPRs and spin temperatures measured in several comets up to date. There are two groups of measurements: airborne and space observations; and ground-based observations. The former group of measurements interpret H_2O detections from both fundamental and hot-bands; the latter are limited to the weaker hot-band transitions for reasons stated in the beginning of this chapter. It should be remembered here that fundamental bands are in general better understood in terms of fluorescence model parameters but cannot be sampled from the ground. Conversely, ground-based retrievals are now more routine.

TABLE 2
H_2O spin temperatures in comets

Comet	Observatory[a]	OPR[b]	T_{spin} [K]	Reference:
Airborne and Space Observatories				
1P/Halley	KAO	2.5±0.1	29^{+2}_{-1}	Mumma et al. 1987; 1989; 1993
C/1986 P1 (Wilson)	KAO	3.2±0.2	> ~40	Mumma et al. 1989
C/1995 O1 (Hale Bopp)	ISO	2.45±0.1	28^{+3}_{-2}	Crovisier et al. 1997
103P/Hartley	ISO	2.74±0.07[c]	35^{+3}_{-1}	Crovisier et al. 1999
Ground-based Observatories				
C/1999 H1 (Lee)	Keck[d]	2.5±0.4	30^{+15}_{-6}	Dello Russo et al. 2005
C/2000 S4 (LINEAR)	Keck	2.8±0.4	> ~30	Dello Russo et al. 2005
C/2001 A2 (LINEAR)	Keck	1.9±0.4	23^{+4}_{-3}	Dello Russo et al. 2005
C/2000 WM1 (LINEAR)	Keck	2.4±0.2	28 ± 3	This work
C/2004 Q2 (Machholz)	Keck	2.7±0.1	34 ± 3	This work
C/2001 Q4 (NEAT)	Subaru	~2.5	~30	Kawakita et al. 2005b[e] DPS abstract

[a] KAO – Kuiper Airborne Observatory; ISO – Infrared Space Observatory.
[b] 1σ uncertainties are given as these are the commonly reported errors in various works.
[c] Weighted mean of OPRs reported separately for two different dates: 2.76±0.08, and 2.63±0.18.
[d] NIRSPEC data ($\lambda/\delta\lambda$ ~ 25,000) analyzed with "coherent" methodology and uncertainty analysis.
[e] Errors have not been reported in the conference abstract (Division of Planetary Sciences annual meeting), but a follow up paper has been submitted to a referred journal (Neil Dello Russo [abstract's co-author], private communication)

Most OPR retrievals in Table 1 correspond to a spectral resolving power exceeding ~10,000. The exceptions are the ISO results, based on $\lambda/\Delta\lambda \sim 1{,}500$, which are sufficient to resolve the stronger ro-vibrational lines.

The still small sample shown on Table 2 confirms that H_2O in comets can have spin temperatures below the statistical equilibrium value of 50K, an important result at this stage understanding. Low spin temperatures suggest that the H_2O molecules might have formed on cold grains with T_{spin} reflecting the grain temperature, rather than from exothermic gas phase reactions (Tielens & Allamandola 1987). But, as pointed out in the introductory chapter, the interpretation of these results is far from simple. It is quite possible that comets nuclei contain a range of spin temperatures, corresponding to the range of regions in which different cometesimals formed. Therefore, interpreting the observed OPRs as leading to a range of temperatures (vs. a single value) might be physically more justified.

Our spin temperatures are significantly different from the derived rotational temperatures in the two comets, consistent with the notion that T_{spin} is cosmogonic observable reflecting the formation temperature of cometary water and not the physical conditions in the inner coma. This is similar to the results of Dello Russo et al. (2005) and Kawakita et al. (2005a; 2005b), which show no correlation between T_{rot} and T_{spin} in the comets they investigated.

What about parent species other than H_2O? In a series of papers Kawakita et al. derived spin temperatures of NH_3 in three comets included in Table 2: C/1999 S4, C/2001 A2, and C/2001 Q4 (cf. Kawakita et al. 2005a). These authors observed NH_2 in the optical and assumed that this radical comes solely from the photo-dissociation of

NH_3. Kawakita et al. also reported spin temperatures of CH_4 in C/2001 Q4. All their results are around 25-30 K, i.e. in the range of the most probable H_2O spin temperatures for the same comets. This is suggestive that all species formed under a thermal equilibrium at the same temperature. However, their results most likely have underestimated uncertainties – only stochastic errors have been used. For example their NH_2-based OPRs of NH_3 in C/2001 A2 are based on high S/N emissions, but do not agree within error, an effect not accounted for in their final conclusions. In the case of CH_4, their quoted "95 %" confidence intervals are simply not clearly defined. By all means, there is a good reasoning behind Kawakita et al.'s suggestion for a common spin temperature of H_2O, HN_3, and CH_4, but this suggestion's credibility can be solidified only after sufficient uncertainty analysis is presented. To their credit, these authors have already initiated that process with their work of the H_2O spin temperature in C/2001 Q4.

The cliche that more definite conclusions demand more OPR measurements certainly holds here. But this should not underestimate the significance of the fraction of Table 2 that covers the ground-based results. These results have opened the possibility to routinely measure H_2O spin temperatures from the ground. Note that in terms of standard uncertainties, the new results for WM1 and Q2 are coming close to the generally more reliable airborne and space-based retrievals. We are at the point where we can efficiently measure H_2O (and other molecular) OPRs regularly. The direction is to improve the hot-band fluorescence models and to use instruments with larger free spectral ranges to sample more para and more ortho lines.

Ultimately we would like to achieve meaningful comparisons between the deduced spin temperatures with the volatile chemistries (i.e. sampled mixing ratios and

abundances of deuterated species) of the observed comets.

4.26 Pushing the Envelope

All previous OPR retrievals are effectively averages over a single spatial region in the cometary coma. Premium S/N data, similar or better than the one from comet Q2, open the door for spatially-resolved measurements of this important parameter. While this is outside of the scope of the present project, it is worth mentioning that it has reached the point work in that direction. If the OPR is indeed cosmogonic, its values should not vary when retrieved from non-overlapping spatial regions in the coma. On the contrary the rotational temperature could vary as local conditions in the collision-dominated inner coma change with distance from the nucleus. The future standard for OPR retrievals should be from spatially-resolved measurements when possible. If consistent, such results would certainly increase the confidence in the measured spin temperatures.

To demonstrate this approach, the analysis described above has been applied to a spatial region covering ~250-750 km from the nucleus of Q2. It was found that the OPR remained the same within error, but was more loosely defined: 2.8 ± 0.3 vs. 2.7 ± 0.1 for the nucleus-centered extract. Hopefully we will be able to accumulate enough time on source for the next bright comet to allow high-enough S/N spectral extracts for several projected distances away from the nucleus.

Another and very important opportunity which future bright comets will bring is the search for deuterated species. The tools developed for line-by-line analysis of H_2O emissions are applicable for sampling multiple HDO features simultaneously with H_2O.

4.27 Optical Depth Effects

The formalism presented in this Chapter assumes that the medium in the cometary coma is optically thin for the relevant ro-vibrational transitions. In the optically thick case the apparent production rate, derived here will be an underestimate. The observed non-resonance transitions are indeed optically thin, but optical depth effects in the pumping solar radiation might be important close to the nucleus for the Q2 data. The radiative vibrational excitation is primarily in fundamental bands of H_2O, which are much stronger than the hot-bands. Dello Russo et al. (2004) showed that for gas productivity in the order of 10^{29} molecules sec^{-1} even the strongest pumping transitions for the H_2O bands in the 2.9 μm region are optically thin outside only a few km from the nucleus. Therefore the optically thin formalism is valid here even for our nucleus-centered extracts. Optical depth effects in the pump have been shown to be very important as far as several hundred km from the nucleus, but for emissions observed in the M-band, which are not subject of this study.

4.28 Results Summary

H_2O was detected in non-resonance fluorescence in multiple lines near 2.9 μm in two Oort cloud comets - C/2000 WM1 and C/2004 Q2. These two data sets provide spectra with good enough signal-to-noise ratio to allow sensitive tests of the H_2O non-resonance fluorescence model at hand. Well-constrained rotational temperatures and H_2O production rates were retrieved for the dates of observations. These H_2O production

rates establish solid baselines for characterization of the volatile chemistries sampled in WM1 and Q2. In addition they will be used as "calibrators" for the emission efficiencies of OH lines, the subject of the next chapter.

Ortho-to-para ratios and H_2O spin temperatures have also been derived with a precision comparable to that of airborne and space-based observations of fundamental bands. The most probable OPR values suggest formation temperatures representing conditions below statistical equilibrium.

4.29 BONUS: H_2O Non-Resonance Fluorescence in the Jupiter-family comer 9P/Tempel 1 – the Target of the Deep Impact Mission

It is highly appropriate to say a few words about the Keck 2 observations in support of the NASA Deep Impact mission. The primary objective of this mission was to excavate material from below the outermost processed layer of a Jupiter-family comet and to measure the chemical composition of gases released from supposedly more pristine inner layers. This mission was highly successful! In addition to data from instruments available on board, numerous teams joined an organized supporting campaign from both ground-base and space observatories. The ground-based effort, organized by Prof. Karen Meech (Meech et al. 2005) was comparable in scope to the International Halley Watch in the 1980s. The event was covered at all possible wavelength regimes and was scheduled to occur in the beginning of the night for Mauna Kea.

It should be emphasized that ecliptic comets present a very serious observational challenge in the NIR (at the current stage of technology). Their overall gas productivity

is generally weak mainly because they have experienced numerous passages in the inner solar system. This allowed building surface layers of particles too large to be dragged by escaping gases - what is historically known as the dust "mantle", mentioned in the introduction in Chapter 1. This residual surface layer would suppress, and for a large fraction of the comet surface, even "shut-down" sublimation. Therefore ecliptic comets (Jupiter-family and Encke-type) are currently underrepresented in our data base - only in three (including 9P/Tempel-1) some parent molecules have been measured directly. Aside from this observational bias, the main question is: what are the mixing ratios between different parent molecules in ecliptic comets. How do they to compare to the range of abundances seen in Oort cloud comets? The Tempel-1 observations are valuable in that regard, because they sample material from the interior of a Jupiter-family comet that is shielded from possible chemical fractionation of the layers immediately below the residual dust "manltle" (Espinasse et al. 1993).

Our group has acquired a truly unique data set sampling several parent molecules before, during and after the impact night (UT Date July 4, 2005). The contribution of this thesis in support of "Deep Impact" was establishing a reliable baseline of H_2O rotational temperature and column abundance at impact night. A first very good estimate of the H_2O rotational temperature and productivity was provided shortly after impact while still on the observing site in Waimea, HI.

On July 4, the first impression after impact has been an increased dust followed by an increased gas productivity. The dust enhancement was revealed by significant continuum brightening; the gas enhancement - by appearance of several H_2O, C_2H_6, and other lines. Figure 18 shows a composite spectrum from Mumma et al. (2005). The top

spectrum highlights H_2O detections in the "KL1" instrument setting, order 26. Out of the four bright lines, three were used for quantitative rotational temperature retrieval. In spite of having high S/N, line No.4 is exceptional because it lacks a reliably determined g-factor. Correlation analysis outlined rotational temperatures of 35-45 K. This value obtained on the observing site in Waimea, was later confirmed by more careful analysis.

In extracting mixing ratios, spherical production rates were not used because the assumption for steady outflow clearly did not hold during impact. Instead mixing ratios between trace constituents and H_2O were derived from the measured integrated column abundances for these species. The final result of the H_2O column abundance after impact is $(1.0 \pm 0.1) \times 10^{31}$ molecules, implying about a two-fold increase compared to pre-impact (Mumma et al. 2005). The quoted paper and the follow-up work in preparation highlight the results on the volatile chemistry of Tempel-1. It is sufficient to mention here that at a very early stage of analysis, applying the methodology presented in this chapter provided values of H_2O rotational temperature and column abundance necessary for the retrieval of all subsequent results.

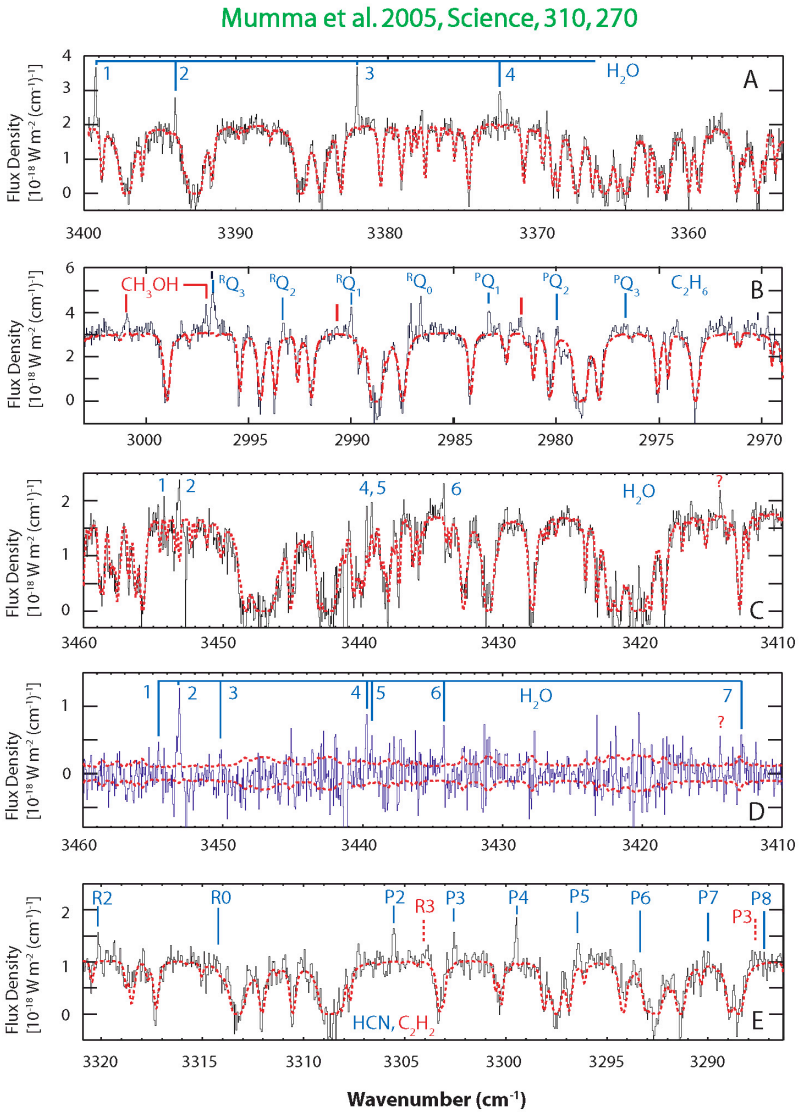

Fig. 18 – Original Figure 2 from Mumma et al. 2005, Science, Special issue on first results following Deep Impact event. Detection of parent volatiles and dust in comet Tempel 1 on UT 4 July, after the impact event. The dashed line in each panel represents the cometary continuum convolved with a synthetic transmittance spectrum of the terrestrial atmosphere. (A and B) Spectra extracted from setting KL1. (A) Four spectral lines of H_2O are shown. Additional lines were detected in order 27, KL1. (B) Six Q-branches of C_2H_6 are detected along with features of CH_3OH. (C to E) Spectra extracted from setting KL2. (C) Five spectral lines of H_2O are marked. (D) Seven H_2O spectral lines are seen in the residuals of (C) after subtraction of the cometary continuum convolved with the atmospheric transmittance. (E) Eight spectral lines of HCN are detected, along with two lines of C_2H_2.

Panel A and the related analysis on H_2O in the KL1 setting is the direct contribution of this thesis to the initial results from the NIRSPEC Deep Impact data. This contribution was completed to a large extent at the observing site in Waimeia, HI. The overall analysis of H_2O pre- and post-impact data was done jointly by Boncho Bonev, Michael DiSanti, and Geronimo Villanueva. Quantitative spectral analysis for all measured species was done with the software algorithms developed by Michael DiSanti and Boncho Bonev initially for the purpose of this thesis and for interpreting H_2CO emission. These algorithms reduced very significantly the time necessary for testing fluorescence models and obtaining rotational temperatures and column abundances.

Chapter 5

Infrared OH Prompt Emission as a Proxy of H_2O Production in Comets

5.1 Motivation

This chapter presents a quantitative analysis of vibrational OH emission in comets. We would like to consider the possibility of an alternative method to measure water production in the NIR on the basis of multiple lines of hydroxyl detected in the L-band. To establish this new approach, this work focuses on the various OH lines from the KL2 NIRSPEC setting presented in our spectral gallery in Chapter 3. These lines belong to the P-branches of the (1-0) and (2-1) vibrational bands of OH (Maillard, Chauville, & Mantz 1976). They belong to transitions which originate from a wide range of rotational levels in the OH upper vibrational state. But before describing the excitation process leading to these levels, it is important to justify the effort:

Why do we strive to quantify H_2O production by alternative (and indirect) means, considering the availability of the "hot-band" method? Wasn't the effort described in the previous chapter convincing enough? Is it really important to complement one IR method based on direct H_2O detection with a second IR method relying on emission from

dissociation product?

Of course, considering every new method for cometary water is justified, because it relates to a critical measurement. But there are much more specific "motivators":

1. The importance of well-defined H_2O production rates, measured simultaneously with other species have already been emphasized – this is the approach leading to the most reliable mixing ratios, which are the first building block of the chemical taxonomy of comets. OH lines are normally detected at various IR wavelengths. Within the L-band their spread is broader than that of the H_2O non-resonance fluorescence emissions. Consequently if we can derive water abundance from OH, this would increase the spectral range over which the production of water and other parent molecules can be measured simultaneously. The most fruitful regions to study H_2O hot-bands are near 2.0, 2.9, and 4.7 μm, but they may not be sampled in targeting other wavelengths, for example the "cometary organics region" (~3.3-3.6 μm). This issue is especially important in work with single order instruments like Phoenix, CRIRES, and the veteran CSHELL.

2. Moreover, even when there is sufficient wavelength coverage to sample simultaneously H_2O and other volatiles, high terrestrial water vapor in the 2.9 μm region may prevent detecting enough hot-band lines to accurately measure T_{rot} and OPR. Recall that while the 2.9 μm region contains the largest number of water emissions and is therefore most suitable for T_{rot} and OPR retrieval, it is also sensitive to atmospheric extinction caused by accidental overlaps with H_2O telluric absorptions in the fundamental bands. In cases of high atmospheric water vapor OH emission could be the rescue since OH lines can be sampled in spectral regions not so affected by atmospheric extinction.

3. An OH-based method is expected to be less sensitive to the H_2O rotational temperature and OPR than the hot-band method (Mumma et al. 2001a). Prior to this project, the new approach had been considered one that might lead to essentially a single-parameter retrieval of $Q(H_2O)$, involving fewer steps than the approach described in the previous chapter. Whether or not our data support this expectation will be revealed in the sections to follow.

4. There is a major motivation to study NIR OH emission for purposes other than quantifying water production in comets. It leads to another example of using comets as natural laboratories in molecular studies. This aspect of the work requires to first review the OH rotational levels and ro-vibrational transitions, and then to discuss the excitation of OH leading to the observed IR emission.

5.2 Nomenclature of OH Ro-Vibrational Levels and Transitions

This chapter will derive new line-by-line parameters describing the NIR OH emission detected in WM1 and Q2 (Figure 4). Therefore, it is appropriate to briefly review the nomenclature of OH ro-vibrational levels and transitions. This review would avoid some ambiguity – the OH energy levels are well understood (e.g. Dieke & Crosswhite 1962; Herzberg 1988), but there is more than one notation for a quantum number and some transitions in the literature.

The NIR samples $^2\Pi$-$^2\Pi$ rotational-vibrational transitions in the ground electronic state of OH (Dieke & Crosswhite 1962). The types of rotational energy levels of the $^2\Pi$ state are presented schematically on Figure 19. The OH lines shown in our spectral gallery belong to the P-branches of the 1-0 (mostly) and 2-1 vibrational bands. For the P-

branch $\Delta J = J'-J'' = -1$, J' and J'' being the rotational quantum numbers of respectively the upper and the lower vibrational levels (for all quantum numbers " ' " and " '' " designate respectively the upper and lower states).

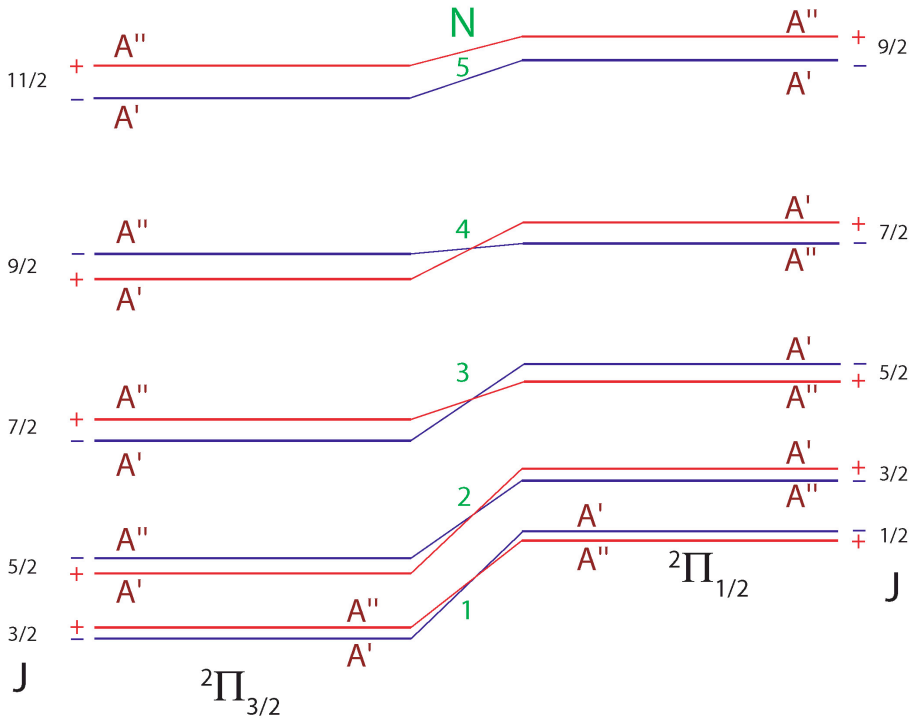

Fig. 19 – Schematic representation (not to scale) of rotational levels in the ground state of OH ($X^2\Pi$).

Two types of energy level splitting are relevant. The first one is caused by the interaction between the total electronic spin and orbital angular momentum (spin-orbit coupling). In the $^2\Pi$ state the resultant of these two momenta in the direction of the internuclear axis (Ω) can be either 1/2 or 3/2 leading to $^2\Pi_{1/2}$ and $^2\Pi_{3/2}$ ladders. The quantum number Ω is well-defined in the absence of rotation or in the case of slow rotation (Hund's case [a]) when the spin is coupled through the orbital angular momentum to the internuclear axis. On the other hand the coupling of the rotation of the two nuclei with the electronic motion is weak.

In the other limiting case, which is fast rotation (Hund's case [b]), the electronic spin is coupled to the axis of rotation of the molecule, rather than the internuclear axis. A given $^2\Pi$ level is split into two states with rotational quantum numbers (J) respectively N-1/2 and N+1/2, where N represents the total angular momentum apart from spin. In the nomenclature of Hund's case [b], the smallest value of this quantum number in unity for Π-states, therefore the smallest J-value is 0.5 ($^2\Pi_{1/2}$).

There are of course intermediate cases between Hund's case [a] and Hund's case [b]. For OH, the transition from the former to the latter as the rotation increases is relatively rapid, with the most rotationally-excited transitions we observe being close or belonging to case [b] (Andresen et al. 1984).

The other type of splitting is caused by the interaction between the rotation of the molecule and the orbital motion of the electrons. This splitting is known as Λ-type doubling (Λ is the projection of the electronic orbital angular momentum on the internulcear axis). The two split levels have opposite parity. They are designated as "+" and "–" depending on the parity of the overall wavefunction. In the high-J limit, the electronic wavefunction of the upper Λ-doublet component for $^2\Pi_{3/2}$ states is antisymmetric with respect to reflection in the plane of molecular rotation, while the lower Λ-doublet component has a symmetric electronic wavefunction. For the $^2\Pi_{1/2}$ ladder, the situation is the same for $J > 3.5$, but reverses for lower J. The Λ-doublet levels with a symmetric and antisymmetric electronic wavefunctions are designated respectively with A' and A". For a given rotational quantum number, the relative population between the A' and A" components is an important characteristic for reactive or photodissociative processes involving open-shell molecules (Alexander et al. 1988).

The most important selection rules to consider are $\Delta J = 0, \pm 1$ (we observed $\Delta J = -1$) and "+" ↔ "–" (Dieke & Crosswhite 1962; Herzberg 1988). There are no strict selection rules for vibrational levels, except that emissions with v'-v'' = 1 (v is the vibrational quantum number) dominate in intensity.

The above background is sufficient to qualitatively interpret the OH line positions shown in our spectral gallery (Figure 4). A full list of detected OH transitions and top-of-the atmosphere line fluxes for WM1 and Q2 is given in Table 3.

The OH lines' rest frequencies are taken from the laboratory measurements of Maillard, Chauville, & Mantz (1976). Another important source is Abrams et al. (1990). These sources agree to a high extent in their reported frequencies but use different notations for the transitions. We have adopted the notation of the former work. This notation refers to the lower rotational state for the corresponding transition and is in the form: P11.5 2^{+}; P12.5 1^{-}. "P" designates the P-branch of the vibrational band; for the first line J''=11.5 (i.e J'=10.5), for the second J''=12.5 (J'=11.5); "2" designates $^{2}\Pi_{1/2}$, "1" designates $^{2}\Pi_{3/2}$; the parity of the lower level is shown in the end with the reminder that the upper states have the opposite parity. If this property is not specified (e.g. P11.5 2) the notation refers to the whole Λ-doublet.

TABLE 3

OH Lines Analyzed in C/2000 WM1 and C/2004 Q2 [a]

OR	Band	Line ID	ν_0 [cm^{-1}]	$F_1(WM1)$	σ_{F1}	$F_1(Q2)$	σ_{F2}
26	1-0	P4.5 1^+	3407.9890	2.63E-19	6.76E-21	3.05E-18	3.04E-20
26	1-0	P4.5 1^-	3407.6069	3.55E-19	7.50E-21	4.16E-18	3.03E-20
25	1-0	P6.5 2^-	3287.8759	—	—	1.18E-18	4.94E-20
25	1-0	P6.5 2^+	3287.4742	—	—	1.33E-18	3.66E-20
25	1-0	P7.5 1^-	3280.7410	—	—	1.11E-18	1.69E-19
25	2-1	P2.5 1^+	3322.1244	—	—	1.52E-18	5.41E-20
25	2-1	P2.5 1^-	3322.0115				
25	2-1	P2.5 2^+	3303.8051	—	—	8.47E-19	5.20E-20
25	2-1	P2.5 2^-	3303.8051				
25	2-1	P3.5 1^-	3286.0109	—	—	4.98E-19	4.14E-20
25	2-1	P3.5 1^+	3285.7591	—	—	7.95E-19	3.81E-20
24	1-0	P9.5 1^-	3189.3929	1.81E-19	1.17E-20	1.19E-18	2.86E-20
24	1-0	P9.5 1^+	3188.4785	1.58E-19	1.18E-20	1.36E-18	3.16E-20
24	1-0	P9.5 2^+	3146.1824	1.86E-19	1.86E-20	1.23E-18	2.82E-20
24	1-0	P9.5 2^-	3145.4909	1.48E-19	1.40E-20	9.41E-19	2.98E-20
24	2-1	P5.5 2^+	3176.5877	—	1.48E-20	2.92E-19	5.17E-20
24	2-1	P5.5 2^-	3176.3171	—	1.44E-20	4.10E-19	4.26E-20
24	2-1	P6.5 1^+	3167.6663	—	9.41E-20	2.56E-19	5.58E-20
24	2-1	P 6.5 1^-	3167.0698	—	1.84E-20	4.92E-19	3.77E-20
23	1-0	P11.5 2^+	3047.4576	1.43E-19	1.16E-20	1.33E-18	3.99E-20
23	1-0	P11.5 2^-	3046.5508	1.09E-19	8.17E-21	8.48E-19	3.06E-20
23	1-0	P12.5 1^+	3044.4280	2.06E-19	7.74E-21	1.58E-18	2.83E-20
23	1-0	P12.5 1^-	3043.2137	1.51E-19	7.94E-21	1.03E-18	2.75E-20
23	2-1	P8.5 2^-	3042.0748	4.96E-20	8.08E-21	4.47E-19	2.73E-20
23	2-1	P8.5 2^+	3041.5069	6.03E-20	7.67E-21	6.16E-19	2.99E-20
23	2-1	P9.5 1^-	3037.0490	3.36E-20	8.54E-21	3.59E-19	4.22E-20
23	2-1	P9.5 1^+	3036.1708	4.54E-20	9.31E-21	4.17E-19	1.44E-19
21	1-0	P16.5 2^-	2787.4646	1.53E-19	1.15E-20	1.05E-18	1.06E-19
21	1-0	P16.5 2^+	2786.1482	5.60E-20	1.04E-20	2.12E-19	3.51E-20
21	1-0	P17.5 1^-	2785.8729	1.71E-19	1.04E-20	7.20E-19	4.89E-20
21	1-0	P17.5 1^+	2784.1880	6.49E-20	1.17E-20	2.39E-19	3.72E-20

[a] OH Lines analyzed quantitatively are given for each NIRSPEC order (OR) of the KL2 setting. Rest frequencies (ν_0) are given. $F_1(WM1)$ and $F_2(Q2)$ are transmittance-corrected line fluxes (W m^{-2}) for the nucleus-centered region (0.43" x 1.78"); their errors (σ_{F1} and σ_{F2}) are estimates of the local photon noise for the corresponding frequencies, which is larger at the presence of sky emission lines in the "raw" data (Chapter 3). Order 25 is analyzed only in C/2004 Q2. The P2.5 1^+ and 1^- lines and the P2.5 2^+ and 2^- lines are blended, so only their summed fluxes are reported. Order 24 lines from the 2-1 band are not detected in WM1, but 3σ upper limits for their equivalent g-factors may be calculated based on the local photon noise.

It should also be noted that we observe the main P-branch and therefore the upper and the lower rotational levels have the same value of Ω (Dieke & Crosswhite 1962). So, in the line notation "1" designates $^2\Pi_{3/2}$ - $^2\Pi_{3/2}$, while "2" designates $^2\Pi_{1/2}$ - $^2\Pi_{1/2}$ transitions.

Most of the detected lines belong to the strongest 1-0 band. The energy separation between the upper and lower ro-vibrational levels is ~2780-3400 cm-1 as indicated by the observed line frequencies.

Two rotational levels with the same value of N (J = N-1/2, and J = N+1/2) lie close together (commonly within a few cm^{-1}). Because of the Λ-doubling of each level, OH lines are commonly detected in "quadruplets". For example, the WM1 spectrum of order 26 shows the quadruplet of N'=3 from the 1-0 band (the unresolved "doublet" for J'=2.5 near 3422 cm^{-1} is not detected in Q2 because of poor atmospheric transmittance); both the Q2 and WM1 spectra show the quadruplets with N'=11 (order 23; 3043-3048 cm^{-1}) and N'=16 (order 21). The brightest OH lines in order 24 belong to two separate quadruplets: N'=8 (doublet in the left part of the order) and N'=9 (doublet in the right part of the order). The "missing" parts of these quadruplets are just outside both order ends. Order 23 also samples the N'=8 quadruplet from the 2-1 band. An important property of out data set is that it samples a wide spread for rotational excitation.

The energy difference between Λ-doublet components is in the order of cm^{-1} or less and increases with rotational quantum numbers. Again it should be stressed that the adopted line notation refers to the lower state, so the P11.5 2^+ line for example originates from a "–" upper state considering the selection rule. Further in the text in will be always stated unambiguously if the reference is to a line transition (e.g. P17.5 1^-) or to its

corresponding upper state (e.g. J'=16.5, $^2\Pi_{3/2}$, "+").

5.3 H_2O Dissociation and OH Prompt Emission

The previous chapter discussed radiative vibrational excitation – the main excitation process leading to the observed molecular emissions in the NIR. We detect multiple species in resonant and/or non-resonant fluorescence. For a given species, the relative intensities of its detected lines depend on the rotational distribution of its ground vibrational state from which the initial ro-vibrational excitation occurs.

This picture is not valid for most of the observed OH IR emission. Its genesis is through a fundamentally different mechanism - the dissociative excitation of H_2O. In particular:

1. H_2O is dissociated into OH and H by the solar UV radiation:

$$h\nu + H_2O \rightarrow OH^*(^2\Pi) + H$$

The "*" in the OH symbol denotes that the hydroxyl is produced in vibrationally- and highly-rotationally excited states. In cometary comae the H_2O molecule is photo-dissociated by UV radiation from two absorption bands (cf. Andresen et al. 1984; Crovisier 1989). The first absorption band of water covers the wavelength range of 136-186 nm. This band dominates the dissociation process because the solar intensity drops rapidly at higher frequencies. Most of the contribution in the second absorption band (λ < 136 nm) comes from the highly-variable solar Ly-α emission.

2. Within milliseconds following dissociation, the excited OH radical transits to a lower ro-vibrational level by emitting an IR quantum. This type of emission is called

"prompt", as the dissociation fragment emits "promptly" after its production.

Most observed OH transitions in the NIR are examples of OH vibrational prompt emission (PE). Note that OH can also be produced in electronically excited states opening the possibility to search for prompt emission in the UV. Prompt emissions from other radicals are also known. The most notable is the $O(^1D)$ emission at 630 nm, used extensively to derive H_2O production rates (cf. Morgenthaler et al. 2001). Another example is the UV prompt emission from CO, following dissociation of CO_2 (Weaver et al. 1994). But our study is solely focused on the IR OH PE in comets. What are the advantages to consider this particular PE?

Two factors potentially make IR PE of OH a compelling proxy for H_2O production. First, the rotationally-excited (also referred as "rotationally-hot") states leading to IR PE can be produced directly by H_2O photolysis, but neither fluorescence of nor collisions with quiescent OH in the cometary coma can efficiently populate them. Second, the newly created OH radical radiates PE only within milliseconds after dissociation of its parent H_2O, thereby minimizing collisional quenching before the ro-vibration transition (Subsequently collisions contribute to cooling the rotational distribution in lower vibrationals state to ambient temperatures [Mumma et al. 2001a]). For these reasons the spatial distribution of OH PE is expected to trace that of its parent H_2O (Crovisier 1989; Bockelee-Morvan & Crovisier 1989). The present work demonstrates that the spatial distribution of H_2O and OH seen in several comets are indeed very similar indicating that the prompt PE is the dominant mechanism for the OH transitions we observe in cometary spectra.

Beyond deriving H_2O production rates, studies of OH PE have the potential to

provide important information about the relative populations of the excited rotational states of OH formed by H_2O photolysis. This is the aforementioned additional (but equally important) aspect motivating this and the following chapter.

5.4 OH Prompt Emission in Comets. Equivalent g-Factors.

The production of OH in vibrationally- and rotationally-excited states has been demonstrated in laboratory studies (Carrington 1964; Yamashita 1975; Andresen et al. 1984; Hausler, Andresen, & Schinke 1987; Harich et al. 2000; Nizkorodov et al. 2003). In the context of comets, Mumma (1982) first suggested studying prompt emission from dissociation fragments at IR wavelengths. Subsequent work on OH PE focused on both the UV (observational searches by Bertaux [1986] and Budzien & Feldman [1991]) and the IR spectral regions (modeling of integrated band emission by Crovisier [1989] and of the 1-0 band rotational envelope at 300 K near 2.8 μm by Bockelee-Morvan & Crovisier [1989]). For P-branch lines (such as those we commonly sample) Bockelee-Morvan & Crovisier's study suggested that OH lines at wavelengths longer than ~2.95 μm should be entirely due to prompt emission, while those at shorter wavelengths may be mixed between prompt and fluorescent emission. The study of Crovisier pointed out the prospects to observe IR prompt emission from dissociatively excited OH and predicted that the integrated intensity of the 2-1 band should be about a quarter of that for the 1-0 band for OH production through the first absorption band of water.

The advent of high-resolution NIR spectrometers in the 1990s enabled regular IR detections of OH lines in comets. Such detections were reported near 3.0 and 3.3 μm in

comets C/1996 B2 (Hyakutake; Brooke et al. 1996; Magee-Sauer et al. 2002), C/1995 O1 (Hale-Bopp; Magee-Sauer et al. 1999), C/1999 (Lee; Mumma et al 2001a), and C/2000 WM1 (Gibb et al. 2003). Currently our data base contains spectra with identified OH lines from multiple comets. This commonality of the OH detections further motivates considering this prompt emission as a proxy for H_2O production. The problem is that we lack derived prompt emission efficiencies for the various individual OH lines sampled in NIR.

The OH prompt emission efficiency, also referred as an "equivalent g-factor", is measured in [(OH photons) sec^{-1} (H_2O molecule)$^{-1}$]. Defining this unit is important because our goal is to interpret the measured OH line fluxes in terms of H_2O production rates, not in terms of OH production rates. The physical basis of such interpretation has already been pointed out – the excited ro-vibrational states of OH we sample cannot be produced efficiently by any process in the cometary coma other than H_2O dissociation. Within the formalism for extracting production rates (presented in the previous chapter) the equivalent g-factor of OH plays the same role as the fluorescence efficiency of H_2O in the treatment of non-resonant fluorescence (see the definition of a spherical production rate). The ratio of the observed OH line flux and its corresponding equivalent g-factor at 1AU is intended to be proportional to the H_2O column density and to lead to obtaining $Q(H_2O)$. A spherical water production rate, obtained from OH PE is defined as:

$$Q^*_{spherical} = \frac{4\pi\Delta^2 F_i^*}{f(x)g_{init}(hc\nu)\tau}$$

The resulting water production rate will be denoted as Q^* to distinguish from the quantity based on direct H_2O detections (Q). Note that for simplicity, the specification "H_2O" is dropped from the notation $Q(H_2O)$, as production of no other species but water is

considered in this chapter.

This expression for a spherical production rate is identical to the one introduced in Chapter 4. The difference is only in the interpretation of two quantities. Here F_i* is the flux of the of the ith OH line, incident to the top of the atmosphere, while g_{init} is the initially assumed OH equivalent emission efficiency [(OH photons) sec^{-1} ($\underline{H_2O \text{ molecule}}$)$^{-1}$] for that particular line. The energy of an OH photon (J) is designated as $hc\nu$.

While OH is a well-studied diatomic molecule, there are no firm theoretical models to predict the rotational distribution in its upper vibrational states for the transitions leading to cometary PE. The related modeling challenge is very likely more serious than in the case of non-resonant fluorescence, which itself presents a difficult problem. In the case of OH, the rotational distribution in the upper state is not a result of the well-understood solar IR pumping; also it is not thermal – laboratory studies confirm that much larger temperatures than in the cometary comae are needed to collisionally populate the highly-excited rotational states observed (we confirm that result from comet data in Chapter 6).

In order to understand the OH rotational distribution one needs to quantify the process of H_2O dissociation not by a monochromatic radiation (usually the case in laboratory studies) but by a continuous frequency range of UV radiation, part of which varies on different time scales with the solar activity. We then meet the challenge to quantify the unbound excited electronic states of H_2O from which excited OH is produced (Andresen et al 1984; Hausler et al. 1987) and this modeling involves different degree of complexity in the first and in the second absorption bands of water. The situation might be even more complicated by the possibility that some part of the

observed OH flux in a given band (for example 1-0) does not come from states directly produced by H_2O photolysis but by the quick decay from higher vibrational level (for example v' = 2 through the 2-1 band).

The present work has an observational emphasis. Instead of developing a theoretical model, it adopts the much more direct approach to empirically derive the distribution of OH emission efficiencies. Their values would encompass the integrated effect from all possible processes producing the OH* state initiating the observed PE. These processes include H_2O photolysis (the primary channel) and possible radiative relaxation from already produced OH* in a state of higher vibrational excitation.

Before describing our approach it should be emphasized that its usefulness is twofold:

1. It is leads to line-by-line equivalent g-factors of OH that can be used to derive H_2O production rates in future comets.

2. All modeling efforts require an observational data base for comparison. Our work provides the first cometary data base of this kind. It may be used to constrain theoretical models for the rotational population distribution of the free OH states produced by water photolysis in cometary atmospheres.

5.5 Previous Analysis of NIR OH Prompt Emission

Our primary task is to derive equivalent g-factors for sufficient number of individual OH lines observed in the L-band. It is useful to stress again the importance of line-by-line derived production rates of any cometary species when possible. Previous quantitative work in that direction for OH has been very limited. Mumma et al. (2001a)

sampled the overall chemistry in comet C/1999 H1 (Lee) by covering almost the entire 2.9-3.7 μm range. Notably, this was the first application of NIRSPEC in the L-band (and the first to comets). OH PE lines were observed throughout this spectral region. Mumma et al. introduced the basic idea for the methodology employed in this chapter. They derived (albeit without any uncertainty estimates) the first value of the combined emission efficiency for the "quadruplet" near 3046 cm^{-1} (Figure 4, order 23). Similarly, Mumma et al. (2001b) derived PE efficiencies for parts of this quadruplet and for a doublet near 2996 cm^{-1} for the same comet.

OH emission was not the central subject of either of these two works, so any detailed quantitative analysis was out of their scope. However, the paper on C/2001 H1 (Lee) has a special significance for OH PE studies. First, it reports detections of OH lines, some of which have very likely contributed to previously unidentified features observed at lower spectral resolution. Such a feature had been observed near 2.8 μm, and Mumma et al. confirmed the prediction of Crovisier (1989) that it might be at least partly OH PE. Second, Mumma et al. motivated the necessity for a detailed quantitative line-by-line investigation of the NIR OH PE in comets in order to establish the general usefulness of this emission as a proxy for H_2O production. This thesis initiates the effort.

5.6 Empirical Approach to Derive OH Prompt Emission Efficiencies

The main idea in the retrieval of equivalent g-factors for OH PE is that the already established method for analysis of H_2O non-resonance fluorescence is used initially as a "calibrator" for the PE efficiencies. This is done in several steps:

0. The spatial profiles of simultaneously detected H_2O and OH are directly compared to confirm the notion that the daughter fragment has a parent-like distribution. As pointed out above, this has been expected based on the production mechanism of the excited ro-vibrational states of OH, and on the short time scale for PE following the H_2O dissociation. This step is intentionally indicated as "zero" because it is preliminary. Establishing the parent-like distribution of OH is a very essential prerequisite, rather than a part of the core algorithm to extract g-factors (i.e. this step would be needed even if g-factors are derived by a different approach).

1. The H_2O production rate, Q, is measured directly from non-resonant fluorescence lines detected within the same instrument setting as OH. This part of the analysis was described in the preceding chapter.

For each OH line:

2. The H_2O production rate is also derived from the measured flux of that line and its assumed equivalent g-factor, g_{init}. Otherwise exactly the same methodology as in the H_2O case is applied and the lifetime of H_2O is used in the calculation.

3. The PE g-factor is adjusted until the water production rate deduced from the OH data (Q^*) equals the production rate directly obtained from water lines (Q). In other words we find the g-factor for which:

$$Q^*(g) = Q,$$
$$As \quad Q^* \propto \frac{1}{g},$$
$$g = g_{init} \times \frac{Q^*(g_{init})}{Q}$$

This approach is legitimate – given the simultaneity of OH and H_2O observations, the direct water measurements provides a "calibrator" for the PE efficiency. This work is

the first in which OH g-factors have been derived from truly simultaneous detection of H_2O and OH (Figure 2; Figure 4).

5.7 Empirical Uncertainties. Physical Significance of the Relative g-Factors.

It should be stated explicitly that in the procedure described above Q and Q^* designate terminal production rates. In practice we apply the above formalism by first calculating nucleus-centered values – Q_{nc} and Q^*_{nc}. The former quantity is the weighted mean of individual H_2O line measurements as described in Chapter 4. On the other hand, Q^*_{nc} is a spherical nucleus-centered production rate, calculated from a particular OH line, because the purpose is to derive line-by-line equivalent g-factors.

Recall that obtaining nucleus-centered values has the advantage of sampling the highest S/N parts of the data – an important consideration especially for the case of nodding on-chip. In this framework the H_2O and OH lines are sampled within the same NIRSPEC setting. Therefore Q_{nc} and Q^*_{nc} are equally affected by flux calibration uncertainties if this calibration is performed in a consistent way among the echelle orders (Appendix 2). This effect is well-understood from deriving mixing ratios within the same instrument setting (cf. Mumma et al 2003).

In addition it is a reasonable assumption that the measured nucleus-centered fluxes of H_2O and OH have been affected identically by atmospheric seeing, comet drift or other observing-related effects common to the whole setting. This assumption can be checked by comparing the growth factors for OH and H_2O (GF_{OH} and GF_{H2O} respectively) using appropriate Q-curves (see Chapter 4). If these two growth factors

agree within error, the OH g-factor may be derived from the nucleus-centered spherical production rates:

$$g = g_{init} \times \frac{Q^*_{nc}(g_{init}) \times GF}{Q_{nc} \times GF} = g_{init} \times \frac{Q^*_{nc}(g_{init})}{Q_{nc}}$$

Here *GF* designates the common growth factor for H_2O and OH. This is the most desired and also the most common case. The resulting uncertainty in the g-factor is then given by the expression:

$$\sigma_g = g \times \left[\left(\frac{\sigma_{Q^*_{nc}}}{Q_{Q^*_{nc}}} \right)^2 + \left(\frac{\sigma_{Q_{nc}}}{Q_{nc}} \right)^2 \right]^{1/2}$$

The first term is the relative "stochastic" error propagated from the photon noise uncertainty in the measured flux of the OH line (from the definition of a spherical production rate it follows directly that the relative errors in measured flux and resulting production rate are the same; all quantities in the expression for a spherical Q^* except the line flux are parameters).

The second term is the overall relative error in the directly measured nucleus-centered H_2O production rate. Chapter 4 shows that this error is commonly calculated as a standard error of the weighted mean of individual nucleus-centered production rates. This error is systematic for the OH g-factors. Let us consider the statistical and physical interpretation of this statement:

The nominator in the expression for emission efficiency, "$g_{init}\, Q^*_{nc}(g_{init})$", is calculated separately for each individual line of hydroxyl. On the other hand the denominator, Q_{nc} is the same "calibrator" for all g-factors. Therefore modifying its value would produce identical relative changes in each individual g. This leads to an important conclusion: when we compare the relative g-factor values within a given setting, the

systematic error from the "calibrator" is not included.

The above conclusion seems straightforward in a statistical sense, but it reflects an important physical consideration: the relative intensities of the OH PE lines, which determine the relative g-factors, do not depend on the water production rate. They depend only on the unbound dissociative quantum states of H_2O from which the excited ro-vibrational states of OH originate. The relative g-factors sample the relative populations among the rotationally-excited states of OH which production is initiated by water photolysis. It was pointed out earlier that these populations present a serious modeling challenge, while the comet data would provide valuable constraints on theoretical efforts.

Finally – one minor remark about the error expression which adds in quadrature relative stochastic and systematic errors. We follow the treatment of Taylor (1982) who points out that there is not a rigorous derivation for this formula. However it is appropriate to use it as an estimate of the overall uncertainty and as an indicator which error source dominates this uncertainty. A direct implication of this treatment is that acquiring OH lines with higher S/N would result in more reliable absolute equivalent g-factors only until the stochastic error becomes smaller than the error in the "calibrator" measurement (i.e. the production rate derived directly from H_2O lines).

5.8 The Case of Different Growth Factors of H_2O and OH

Consider two species observed within the same instrument setting (for example OH and H_2O). Is it possible that their corresponding Q-curves display different growth factors even when the fluxes from these two species are affected identically by observing-related effects? The answer is "Yes", and it is valid when there are differences in the observed spatial profiles (respectively Q-curves) which are not caused by observing conditions, but by physical effects in the cometary coma.

The shape of a Q-curve is a function of the spatial profile from which it is constructed. Up to this point the concept of a Q-curve and a growth factor was treated solely as a means to correct for atmospheric seeing, comet drift, or sub-optimal focus. This aspect was covered in the previous chapter dedicated to a single molecule – H_2O. Now when there are two species in consideration, one has to take into account the possibility their spatial distributions to differ because of physical effects. A growth factor difference is a signature for such differences, when observing-related factors are identical (cf. DiSanti et al. 1999; 2001). This might be significant upon particular circumstances even for H_2O fluorescent emission and OH prompt emission, as will be shown with a very important example in this work.

The empirical formalism for deriving OH g-factors is based on terminal Qs, which are equal to the products of the corresponding nucleus-centered values and growth factors. When the OH and H_2O growth factors differ ($GF_{OH} \neq GF_{H_2O}$), the nucleus-centered production rates deduced from hydroxyl and from water lines should be corrected separately before deriving a g-factor:

$$g = g_{init} \times \frac{Q^*_{nc}(g_{init}) \times GF_{OH}}{Q_{nc} \times GF_{H_2O}} = g_{init} \times \frac{Q^*(g_{init})}{Q}$$

Stochastic and systematic errors can be combined in the same fashion as in the previous section, but including the uncertainties in the growth factors. Note that the errors in the growth factors can be summed in quadrature, because GF_{OH} and GF_{H2O} are measured from separate Q-curves constructed from flux coming from different species.

A practical limitation of this application is that the spatial profiles from OH lines have generally lower S/N. The H_2O profiles are better in that regard, because they contain flux summed over larger number of individual lines. Fortunately we have an adequate S/N in the OH profiles for that particular case when it most matters – the comet Q2. The reason for this will be revealed soon.

This concludes the methodology part of this chapter. The results sections open with presenting the spatial profiles and Q-curves of OH PE in several comets. The core results focus on equivalent g-factors for numerous OH lines spanning a wide range of rotational levels in the upper vibrational state for the corresponding transitions.

Like the previous chapter, the main results here focus on the data from comets WM1 and Q2, shown in the spectral gallery (Chapter 3).

5.9 Spatial Distribution of the OH Prompt Emission in C/2000 WM1

The nature of the PE implies that its spatial distribution tracks that of the parent volatile. This is what we generally observe to the extent permitted by the S/N in the H_2O

and OH profiles. For example Figure 20 shows direct comparisons between spatial profiles of H_2O and OH emissions detected in comet WM1. Like the H_2O profiles, the OH profiles are also "combined" – they are constructed from the summed flux of several OH lines within a given order. Each OH profile is scaled to its corresponding H_2O profile by the ratio between the intensities of these two profiles summed over their central five pixels.

It should be emphasized that the spatial profile of the OH PE intensity does not provide information about the distribution of quiescent OH in the coma. This distribution can be tracked by probing OH fluorescent emission. Figure 20 clearly demonstrates that the OH intensity distribution is not characteristic of fluorescence. Unlike PE, the spatial profile of fluorescent OH is much flatter and in most cases does not peak at the nucleus (e.g. Weaver et al. 1999a).

The congruence between the spatial distributions of OH PE and H_2O fluorescent emission results in the agreement between the growth factors derived for OH and H_2O (GF_{OH} and GF_{H2O}). An example is shown on Figure 21 with Q-curves from the WM1 data.

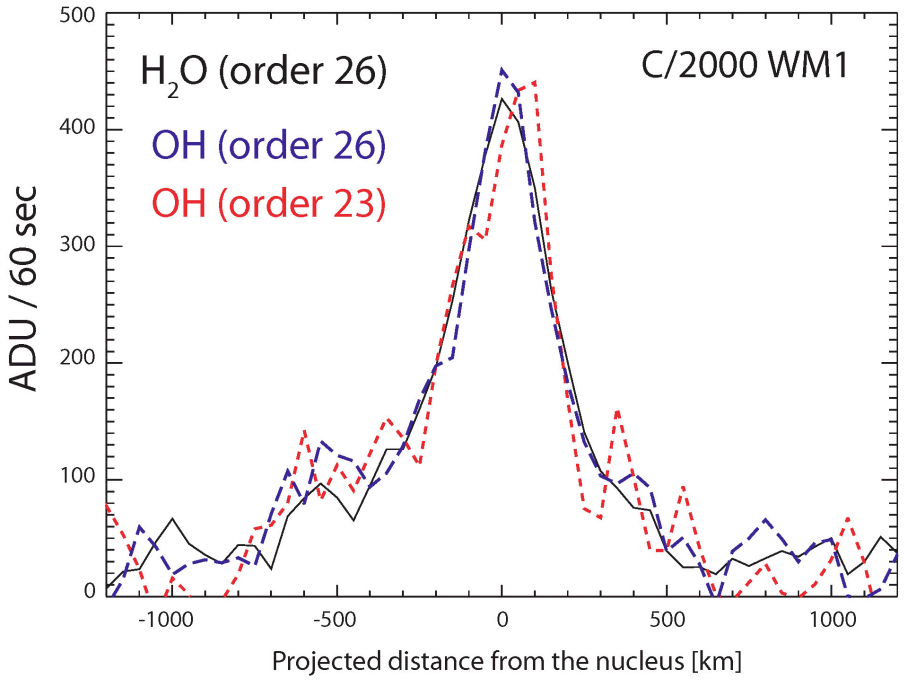

Fig. 20 – Spatial profiles of H_2O and OH emission in C/2000 WM1. The H_2O profile (black) is the same as the one presented on Figure 6. The OH profile from order 26 (blue) contains the combined flux of the two lines near 3407 cm^{-1} (Figure 4A from the Spectral gallery in Chapter 3); the profile from order 23 (red) represents the summed signal of the "quadruplet" spanning between 3043 cm^{-1} and 3048 cm^{-1} (Figure 4F).

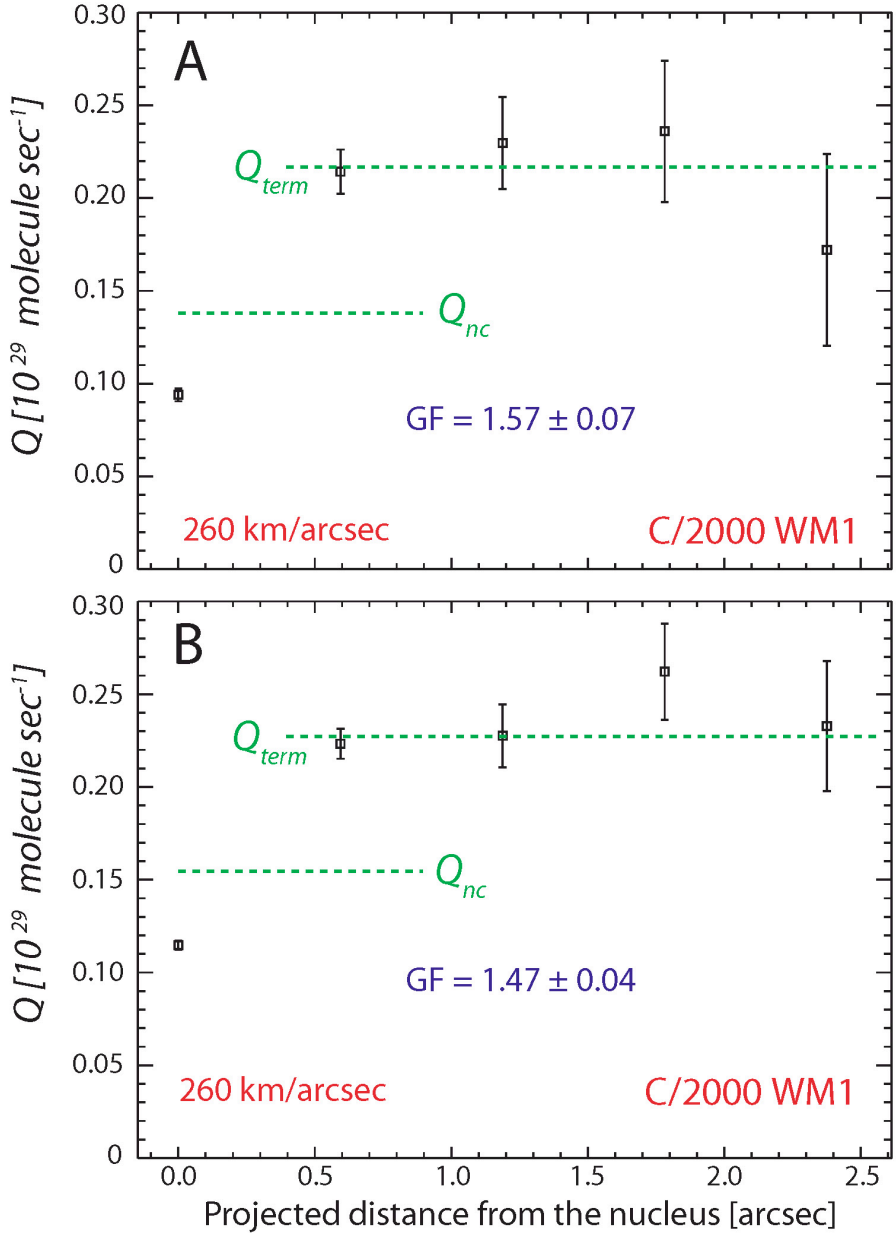

Fig. 21 – Q-curves constructed from the OH spatial profiles shown on Figure 20: (A) order 26; (B) order 23. The derived growth factors are independent on the assumed g-factors and are consistent with the value for H_2O (1.55 ± 0.03).

The agreement between the growth factors derived from OH and H_2O emissions implies that nucleus-centered values of production rates can be used in the derivation of PE g-factors.

Note that similar strong correlation between the distributions of H_2O and OH PE has been observed in two other comets: C/1999 H1 (Lee) and C/2001 A2 (Linear). These distribution were discussed in detail by Bonev et al. (2004), a study which focused on only one OH "quadruplet" and can be treated as the precursor of the main body of work presented here in detail. Because it does not bring an additional conceptual insight, it should be just mentioned that Figure 2 from the cited article displays very similar correlation between the distributions of water and OH PE, as displayed on Figure 20 for WM1.

Finally, OH PE profiles from lines in order 21 in C/2003 T7 have been used in searches for extended component in the co-measured H_2CO (DiSanti et al., in preparation). Such searches include a direct comparison between the spatial distributions of OH and H_2CO assuming the former is parent-like and is tracking the parent H_2O. C/2003 T7 was a particularly bright comet observed by CSHELL with a lot of integrated time on source.

The search for an extended component of formaldehyde is using OH as a reference because H_2O and H_2CO could not be detected simultaneously. The important assumption in this approach is that OH PE has a distribution tracking the distribution of the parent, as in the case of WM1. The next part from our presentation shows that although such an assumption is reasonable in many cases, it cannot be considered valid a-priori in comets with relatively high gas productivity observed close to Earth.

5.10 Spatial Distribution of the OH Prompt Emission in C/2004 Q2

The following sections first demonstrate and then offer an explanation for a difference between the spatial profiles of H_2O and OH identified in comet Q2. It is appropriate to emphasize at the very beginning the two critical implications of this analysis. First, it is significant for deriving PE equivalent g-factors of the OH lines seen in Q2. Second, it has a direct implication in the way the OH-based method for deriving H_2O production rates has to be applied in future comets, especially in the absence of simultaneously quantified H_2O emission.

Comet Q2 presents an interesting case – the spatial profiles of OH and H_2O display a notable difference shown on Figures 22, 23, and 24. Each figure compares the spatial profile of H_2O introduced in Chapter 4 to the spatial profile of OH intensity measured in order 26, 24, and 23 respectively. The OH profile from order 26 contained the combined flux of the two lines near 3407 cm^{-1} (Figure 4B from the Spectral gallery in Chapter 3); the OH profile from order 24 contains the flux summed over the four bright lines, two at each end of the order (Figure 4E); finally the profile from order 23 represents the summed signal of the "quadruplet" spanning between 3043 cm^{-1} and 3048 cm^{-1} (Figure 4G). OH emissions from other orders are not strong enough to construct reliable spatial profiles.

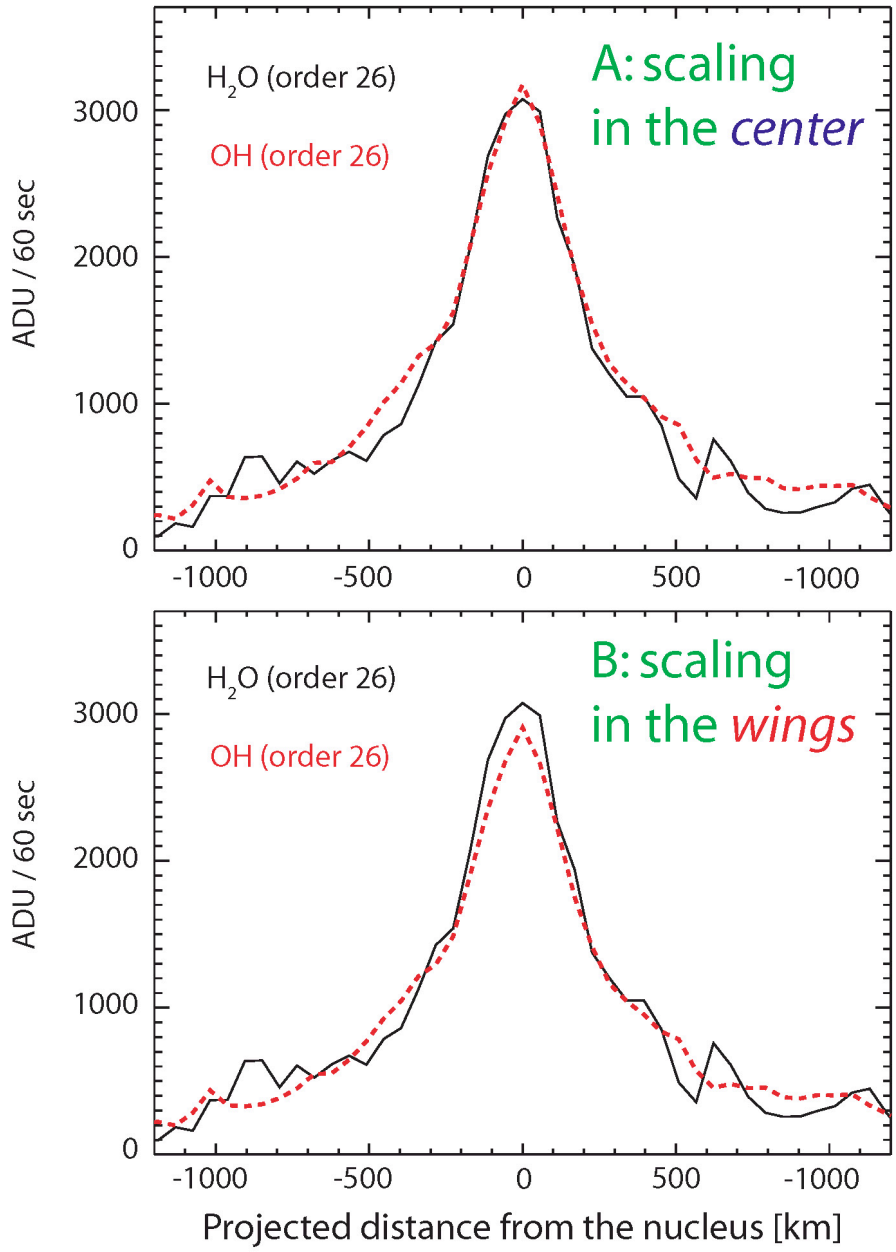

Fig. 22 – Spatial profiles of H_2O and OH emission in C/2004 Q2, **order 26**. The OH profile is scaled to the H_2O profile in two alternative ways described in the text.

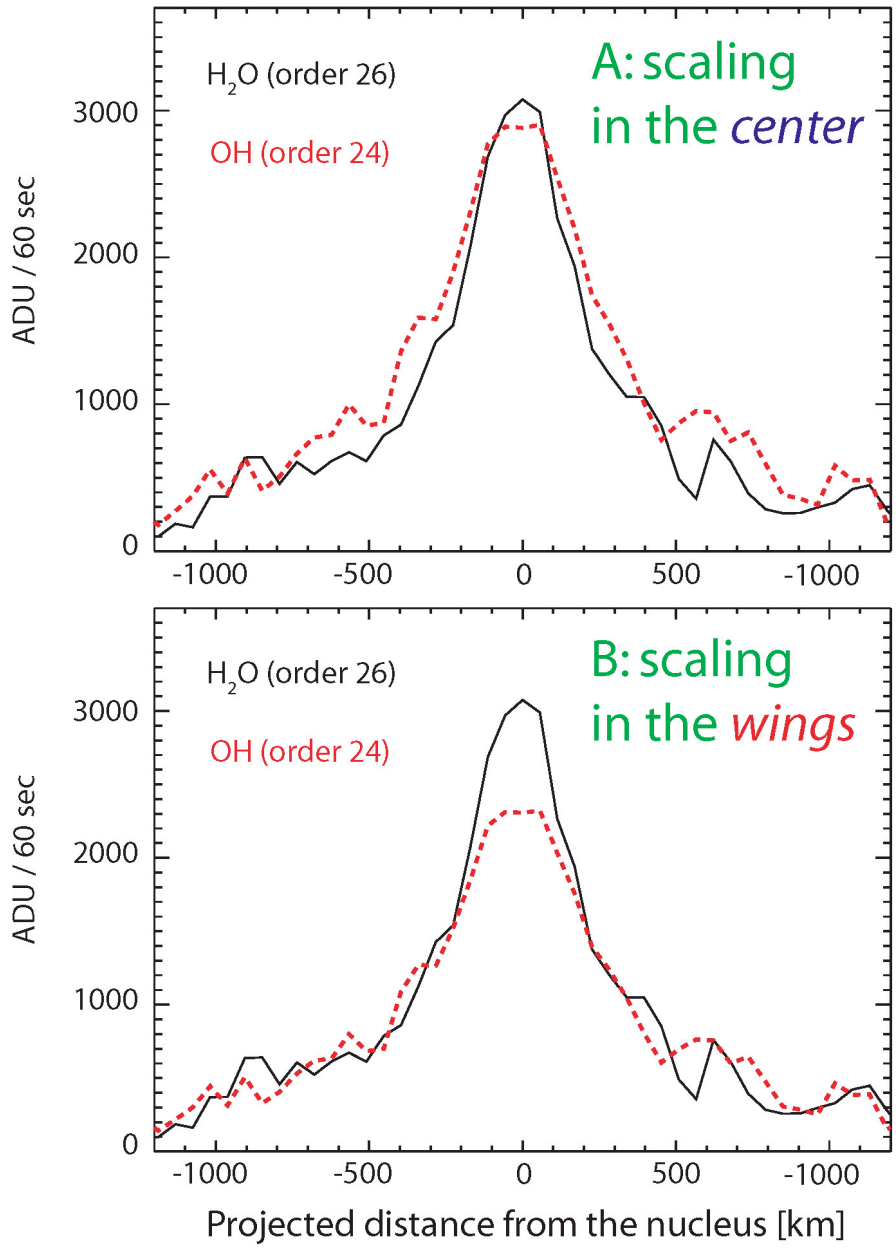

Fig. 23 – Spatial profiles of H_2O and OH emission in C/2004 Q2, **order 24**.

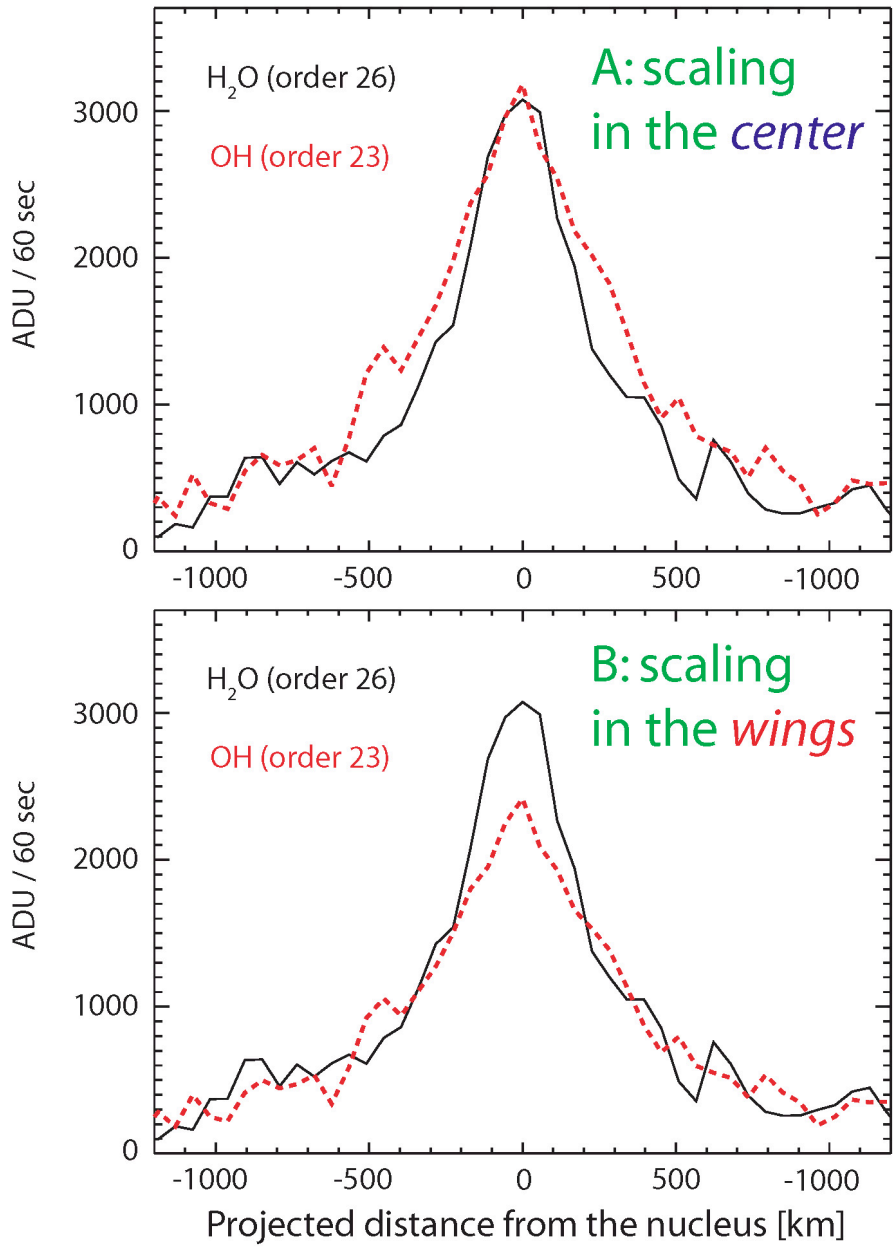

Fig. 24 – Spatial profiles of H_2O and OH emission in C/2004 Q2, **order 23**.

The difference between the H_2O and the OH distribution is demonstrated in two different ways for a better understanding of the effect. The OH profiles are weaker than the H_2O profile, so the former profiles need to be scaled to the latter. Note that the profiles' appearance in comparison plots depends on the way the two distributions are scaled to each other when these distributions differ in shape. Panel A and panel B show two different ways of scaling the OH profile. In panel A the scaling factor is equal to the ratio of intensities measured over the central five spatial pixels in each distribution. In this representation the two profiles are forced to be consistent in their central parts, but the OH profiles look wider otherwise. Alternatively, the B-panels show OH scaling according to intensity ratios from five pixels in the wings on each side of the profile's peak. In this representation there is apparently some OH "missing" in the center. Otherwise the intensity from hydroxyl tracks the water intensity, as expected from PE.

The difference in the spatial profiles propagates to a difference in their corresponding Q-curves. The OH Q-curves are displayed on Figures 25, 26, and 27. These Q-curves look similar to the H_2O Q-curve shown on Figure 15. However, they result in slightly larger growth factors. No such difference is seen in other comets in our data base or reported elsewhere.

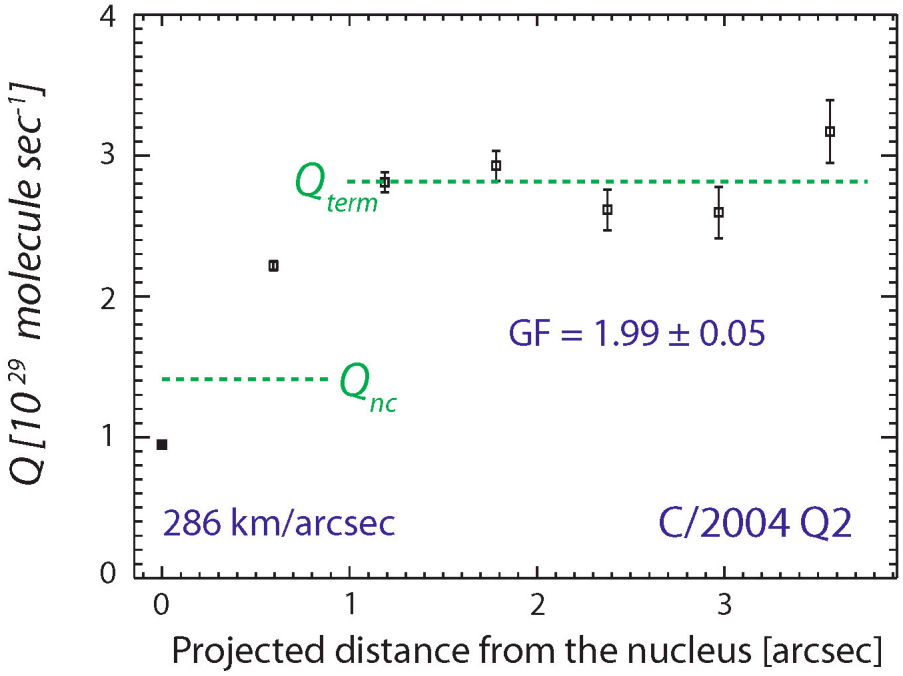

Fig. 25 – Q-curve constructed from the OH spatial profile shown on Figure 22 (**order 26**). An equal arbitrary OH g-factor is assumed in calculating each spherical production rate. The derived growth factor is slightly larger than that for H_2O (1.78 ± 0.06) derived in Chapter 4.

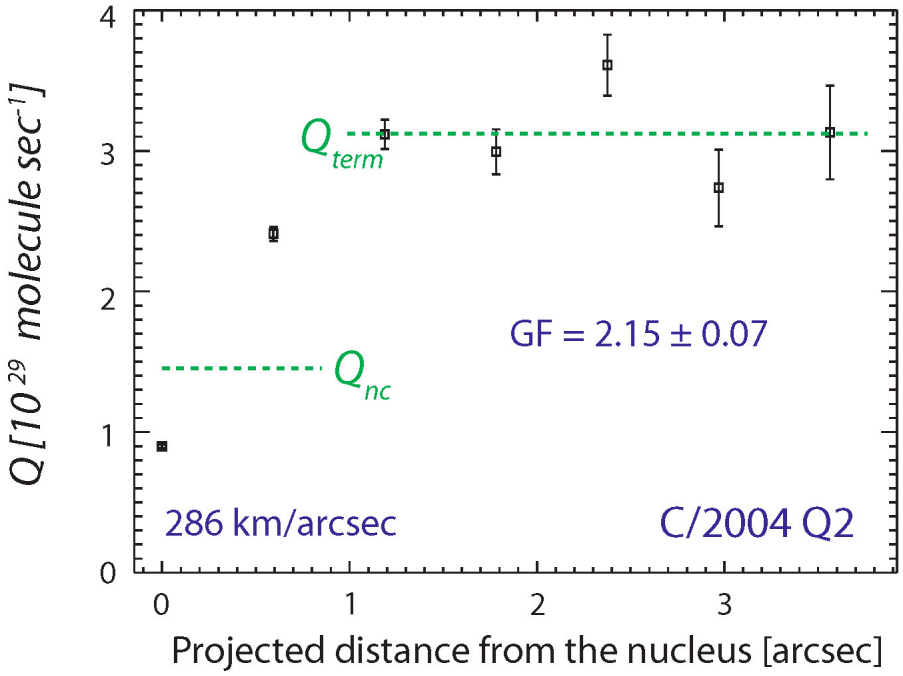

Fig. 26 – Q-curve constructed from the OH spatial profile shown on Figure 23 (**order 24**). An equal arbitrary OH g-factor is assumed in calculating each spherical production rate. The derived growth factor is larger than that for H_2O (1.78 ± 0.06).

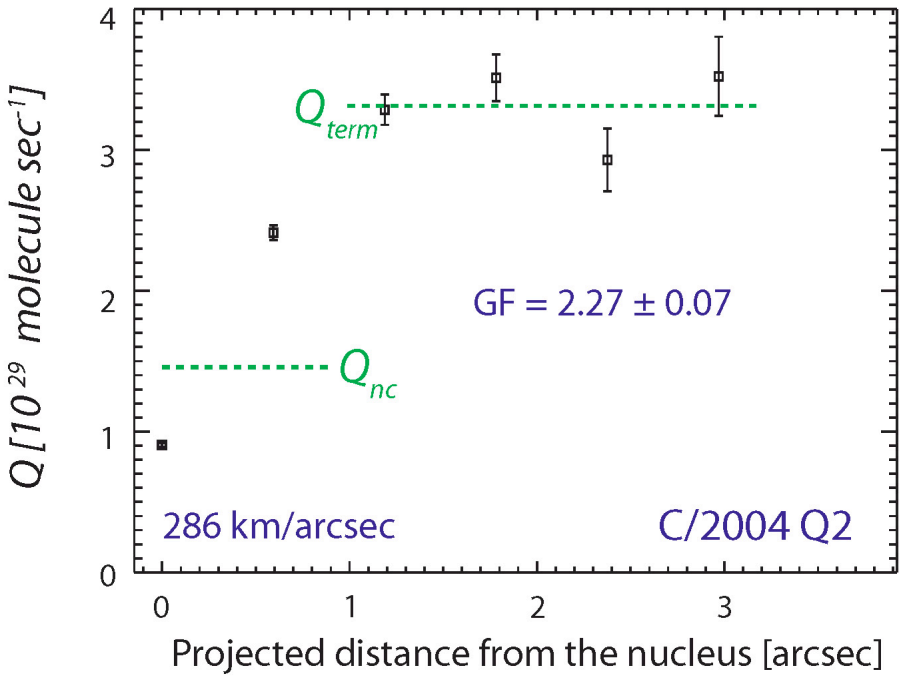

Fig. 27 – Q-curve constructed from the OH spatial profile shown on Figure 24 (**order 23**). An equal arbitrary OH g-factor is assumed in calculating each spherical production rate.. The derived growth factor is larger than that for H_2O (1.78 ± 0.06).

Two issues will be discussed is sequence: first, the larger growth factors for OH compared to H_2O; second, the difference in the OH growth factors derived from order 26 and from orders 24 and 23.

It should be emphasized that the OH and H_2O distributions and Q-curves are still correlated if not in as good agreement as in the case of WM1. The OH profile peaks at the nucleus, consistent with PE, and is definitely much steeper than a distribution dominated by a fluorescent component. But what is the cause the observed difference?

The most plausible interpretation is that the OH emission efficiencies are apparently lower in the region closest to the nucleus than in the wings of the spatial profiles. This would explain the apparent differences in the spatial profiles. The scaling

in Panel A forces the H_2O and OH distributions to be congruent in the region where OH is under-produced, while water follows the normal distribution for a parent volatile. This scaling is therefore artificial. A more realistic is the scaling in Panel B which avoids the inner part of the distribution, where according to our hypothesis the OH g-factors would be lower.

Two explanations for lower PE efficiency in the innermost observed coma have been considered: collisional quenching and optical depth effects in the solar UV. The OH PE efficiency can be lower if the OH is collisionally quenched from the excited vibrational states produced as a result of H_2O photolysis. However, it can be easily shown that the collision time scale may become comparable to the time scale of PE (~ ms) only for number densities in the inner coma much higher than predicted on the basis of our derived H_2O production rates (or even greater values). A detailed treatment of this issue is presented by Mumma (1982) and by Weaver & Mumma (1984) confirming the notion that that radiative vibrational relaxation dominates over collisional vibrational relaxation throughout the observed coma. The other aforementioned reason for under-producing OH in the inner coma is more plausible.

5.11 Optical Depth Effects in the Solar UV

The genesis of the OH PE is photo-dissociation of H_2O by the solar UV radiation. Optical depth effects in the UV would lower the rate of H_2O photolysis and hence – the efficiency for a PE. The spatial distribution of OH would not track the distribution of the parent in that case. The critical question is at what distance from the nucleus, the optical depth in the UV radiation responsible for dissociation of H_2O becomes unity. This

distance will be referred as "critical radius" and is estimated as follows:

For a spherically symmetric outflow and for distances to the nucleus much smaller than the H_2O dissociation scale length (~70,000 km) the number density of H_2O as a function of the distance from the nucleus (r) is given by the expression:

$$n(r) = \frac{Q}{4\pi \cdot V \cdot r^2}, \ \ cm^{-3},$$

where V is the expansion velocity, and Q is the H_2O production rate.

The column density N_{col} from a certain distance from the nucleus R along the line directed to the Sun is:

$$N_{col}(R) = \int_R^\infty n(r)dr = \frac{Q}{4\pi \cdot V \cdot R}, \ \ cm^{-2}.$$

The optical depth at certain UV frequency is:

$$\tau_{UV}(R) = \sigma_{abs} \cdot N_{col}(R),$$

where σ_{abs} is the absorption cross section of H_2O for that particular frequency.

The "critical radius" R_c is defined for $\tau_{UV}(R_c) = 1$:

$$R_c = \frac{\sigma_{abs} \cdot Q}{4\pi \cdot V}$$

For this calculation we conservatively adopted a value of the expansion velocity of 4×10^4 cm/sec (higher value would obviously imply the effect is less important in the observed coma as R_c get close to the nucleus). The absorption cross-section for H_2O has been compiled from Budzien. Festou, & Feldman (1994). We adopt two values: 1.6×10^{-17} cm^2 for Ly-α and 0.5×10^{-17} cm^2, which corresponds to the peak of the first absorption band. As pointed out in earlier in this chapter, most of the water photolysis comes from the first absorption band of H_2O, because the solar flux is larger at its frequencies;

dissociation from Ly-α dominates the second absorption band and also has a significant contribution to the overall photolysis rates.

Table 4 shows estimated critical radii for the four comets in our data base for which the spatial distributions of OH PE and H_2O fluorescent emission have been directly compared. These critical radii are compared to the pixel size (km) along the spatial dimension for NIRSPEC and the corresponding extent of the nucleus-centered region. These sizes depend on the geocentric distance.

TABLE 4
Optical Depth Effects in the UV

Comet	$Q(H_2O)$ [10^{29} molec sec^{-1}][a]	Δ [AU]	R_c (Ly-α) [km]	R_c (FAB) [km]	Spatial pixel size [km][b]	Spatial extent for Q_{nc} [km][c]
C/2000 WM1	0.214	0.357	7	2	51	231
C/2004 Q2	2.550	0.394	81	25	57	255
C/1999 H1	1.431	1.348	46	14	194	871
C/2001 A2	0.430	0.282	14	4	40	182

[a] Calculated from analysis of H_2O non-resonant fluorescence. The production rates for C/2000 WM1 and C/2004 Q2 are presented in Chapter 4; the production rates for C/1999 H1 and C/2001 A2 are taken from Dello Russo et al. (2005).

[b] Pixel size over the spatial dimension for NIRSPEC = 0.198 [arcsec/pixel] x 725.3 [km $arcsec^{-1}$ AU^{-1}] x Δ [AU], where Δ is the geocentric distance.

[c] Q_{nc} is the "nucleus-centered" production rate as defined in Chapter 4 – it corresponds to a flux measurement over the central (with respect to the peak gas emission) nine spatial pixels (i.e. ± 4 pixels with respect to the central one). It should not be confused with the "zeroth" spherical production rate in the Q-curves (Q_0), which corresponds to a flux measurement over the three central spatial pixels, as explained in the caption of Figure 15.

The first thing to emphasize in interpreting Table 4 is that it provides only estimates for the relative importance of possible optical depth effects in the UV for the included comets. The given critical radii are not a subject of a precise modeling. Such modeling would face (among other factors) significant uncertainties in quantifying the

outflow from the nucleus of Q2 (in order to obtain more realistic number densities), and in quantifying the relative contributions of different UV frequencies to the production of each excited state of OH leading to an observed PE.

Table 4 suggests that optical depth effects in the UV are not expected to be important for all comets in the sample except for possibly Q2. Indeed the observed OH and H_2O spatial profiles have been congruent in WM1, C/1999 H1, and C/2001 A2 (e.g. Figure 20). Optical depth effects become important at higher water production rates and small geocentric distances for which the critical radius becomes comparable to the part of the coma that is spatially resolved. WM1 and C/2001 A2 were observed close to Earth, but also exhibited relatively low gas productivity. C/1999 H1 (Lee) had H_2O production rate comparable to the one for Q2, but was observed at significantly larger geocentric distance.

Q2 has the highest H_2O production rate among the comets included in Table 4 and was observed at only 0.394 AU. Although very approximate, the calculation supports the hypothesis that the observed difference between the distributions of OH and H_2O in Q2 is caused by optical depth effects for the wavelengths responsible for water photolysis. Near the centers of the spatial profiles (Figures 22-24) the PE efficiency is lower because the dissociation rates are "damped" by opacity. OH PE is not supposed to track H_2O. In the wings of the profiles, the coma is optically thin at UV wavelengths, H_2O dissociation should be efficient, and OH PE is expected to track the parent. If this is the case, the scaling in the B-panels of Figures 22-24 would be more realistic.

5.12 Interpretation of the Larger Growth Factors in the OH Q-curves

Recall from Chapter 4 that the Q-curve assumes a constant g-factor in the calculation of all spherical production rates. Each spherical production rate is proportional to the ratio between measured flux and assumed g-factor. If the latter does not vary significantly over the spatial extent of the Q-curve, the resulting growth factor would be practically independent of the assumed emission efficiency, as discussed in the section "*Growth factor and correction of the nucleus-centered production rates*". However, if OH is under-produced in the innermost coma, the actual g-factor there would be smaller than in the terminal region of the Q-curve. Then assuming the same g-factor in the nucleus-centered and in the terminal region would result in an underestimated flux – to – g-factor ratio near the nucleus, leading to an underestimated production rate (Q^*_{nc}).

As a result, the ratio Q^*/Q^*_{nc} would be larger than in the case of H_2O. This could explain the "slower" rise to a "terminal" value in the OH Q-curves, which are otherwise quite similar to the H_2O Q-curve (Figure 15).

Because the OH and H_2O were observed within the same NIRSPEC setting, their nucleus-centered fluxes should be identically influenced by atmospheric seeing, drift, etc. The OH growth factors shown on Figures 25-27 are larger than the one for H_2O, because in addition to observing-related effects they correct for the under-production of OH in the nucleus-centered region.

In this situation it is particularly important to derive equivalent g-factors based on "terminal" production rates. In practice g-factors are first determined based on nucleus-centered measurements to take advantage of the best S/N. Then these emission

efficiencies are corrected by the ratio GF_{OH}/GF_{H2O}, as prescribed in the methodology presented for the case of $GF_{OH} \neq GF_{H2O}$.

Despite uncertainty in the interpretation of the difference in the spatial profiles, the ratio in the growth factors provides an adequate correction to the g-factor values deduced from nucleus-centered production rates – within the limits imposed by the uncertainty of this ratio, the resulting H_2O terminal production rates from OH lines equal the directly measured value from the H_2O hot-bands. The described treatment is appropriate for any effect (UV optical depth or other) resulting in underestimated g-factors in the nucleus-centered region of the OH spatial distribution.

The excitation mechanism leading to OH PE in comets implies that this emission should be considered a proxy for H_2O production only when measured over a region in the coma which is optically thin in the UV wavelengths responsible for H_2O dissociation. For comets with $Q \sim 10^{29}$ molecules sec^{-1}, observed close to Earth, optical depth effects might become important in the innermost spatially resolved part of the coma. To avoid the UV opacity factor, H_2O production rates should be derived from the wings of the OH spatial profiles in a distance from the nucleus, where the medium is expected to be optically thin in the UV. If there is not sufficient S/N to determine a reliable Q-curve from a single line, the methodology of this work should be applied: line-by-line values of Q_{nc}* should be subsequently corrected by a growth factor derived from a Q-curve corresponding to a spatial profile containing OH flux summed over several lines.

5.13 Obtaining Terminal Production Rates (Q*) in C/2004 Q2 from OH Lines Corresponding to Different Excitation

It has already been emphasized that the OH equivalent g-factors should be derived from terminal production rates ($Q^* = Q^*_{nc} \times GF_{OH}$)). Directly measured growth factors are available for the P4.5 1 lines (order 26), P9.5 1 and P9.5 2 lines (order 24), and P11.5 2 and P12.5 1 lines (order 23). The individual nucleus-centered production rates from these lines are corrected with their corresponding growth factors. For the P4.5 lines it is 1.99 ± 0.05. Because the values of GF_{OH} in order 24 and 23 do not differ significantly, a weighted mean growth factor (2.21 ± 0.05) is used for both orders. The rest of the lines do not contain high-enough S/N outside the nucleus-centered region to enable an adequate spatial analysis even after individual line fluxes.

There is an apparent correlation between the level of excitation sampled by the OH lines and the necessary correction between nucleus-centered and terminal production rate. The significance of this correlation cannot be validated based on the three Q2 spatial profiles alone. But if UV opacity indeed is the cause for under-producing OH close to the nucleus, the effect should not be the same for all OH lines. It will depend on the optical depth for the frequencies mainly responsible for producing the OH excited state giving rise to a particular transition.

We assume that the GF_{OH} value from orders 24 and 23 is valid also for the 1-0 band lines in order 21 (P16.5 2; P17.5 1), which are of even higher excitation. For the 1-0 band lines from order 25 (P6.5 2; P7.5 1^-) and weighted mean growth factor between the low and high excitation lines is adopted (2.10 ± 0.04). For the 2-1 band lines in order 25 (P2.5 1; P2.5 3; P3.51) we assume the GF_{OH} value from order 26. Finally for the 2-1

band lines from orders 24 (P5.5 2; P6.5 1) and 23 (P8.5 2 P9.5 1) we assume the same growth factor as for the high-excitation 1-0 band lines.

Note that for deriving OH g-factors the important parameter is not GF_{OH} but the ratio GF_{OH}/GF_{H2O} ($GF_{H2O} = 1.78 \pm 0.06$ for Q2).

The limited spatial information in Q2 is not sufficient to neither justify, nor reject the assumptions made for correcting nucleus-centered to terminal values of OH-based H_2O production rates (Q*). Therefore these assumptions are the most serious systematic uncertainty factor in the OH g-factors for this comet. This uncertainty will be addressed long-term in two ways:

1. By comparing the spatial distributions of simultaneously detected OH and H_2O in future comets observed in circumstances similar to Q2, i.e. production rate exceeding 10^{29} molecules sec^{-1} and small geocentric distance allowing to spatially resolve the innermost coma.

2 By testing the OH emission efficiencies from the Q2 data set against observations of OH PE for which OH tracks H_2O throughout the observed coma.

The example with Q2 is very important within the general scope of this chapter, which is to investigate the usefulness of OH PE as a IR proxy for H_2O production in comets. This example shows that very plausible and long-time standing assumption (that OH PE should always track H_2O) needs to be checked when possible in both deriving OH emission efficiencies and in applying a set of these parameters to a new comet, especially when there are not direct simultaneous detections of H_2O. It should be emphasized again that our Q2 g-factors are derived from terminal production rates for which OH PE is tracking water.

5.14 Equivalent OH g-Factors in C/2000 WM1 and C/2004 Q2

Figures 28-31 present emission efficiencies for a number of OH lines detected within the KL2 NIRSPEC setting in two comets. Each figure contains two panels for direct comparison between both comets. Figures 28 ("+" transitions) and 29 ("–" transitions) show g-factors for lines from the 1-0 band; Figures 30 and 31 – g-factors for lines from the 2-1 band. Emission efficiencies for transitions from $^2\Pi_{1/2}$ states are shown as squares; the ones from $^2\Pi_{3/2}$ – as diamonds. All plots have the same vertical scale for an easier comparison of the g-factors: within a comet, and between the two comets. The equivalent g-factors are calculated for heliocentric distance of 1 AU and are assumed to vary as the inverse square of the heliocentric distance.

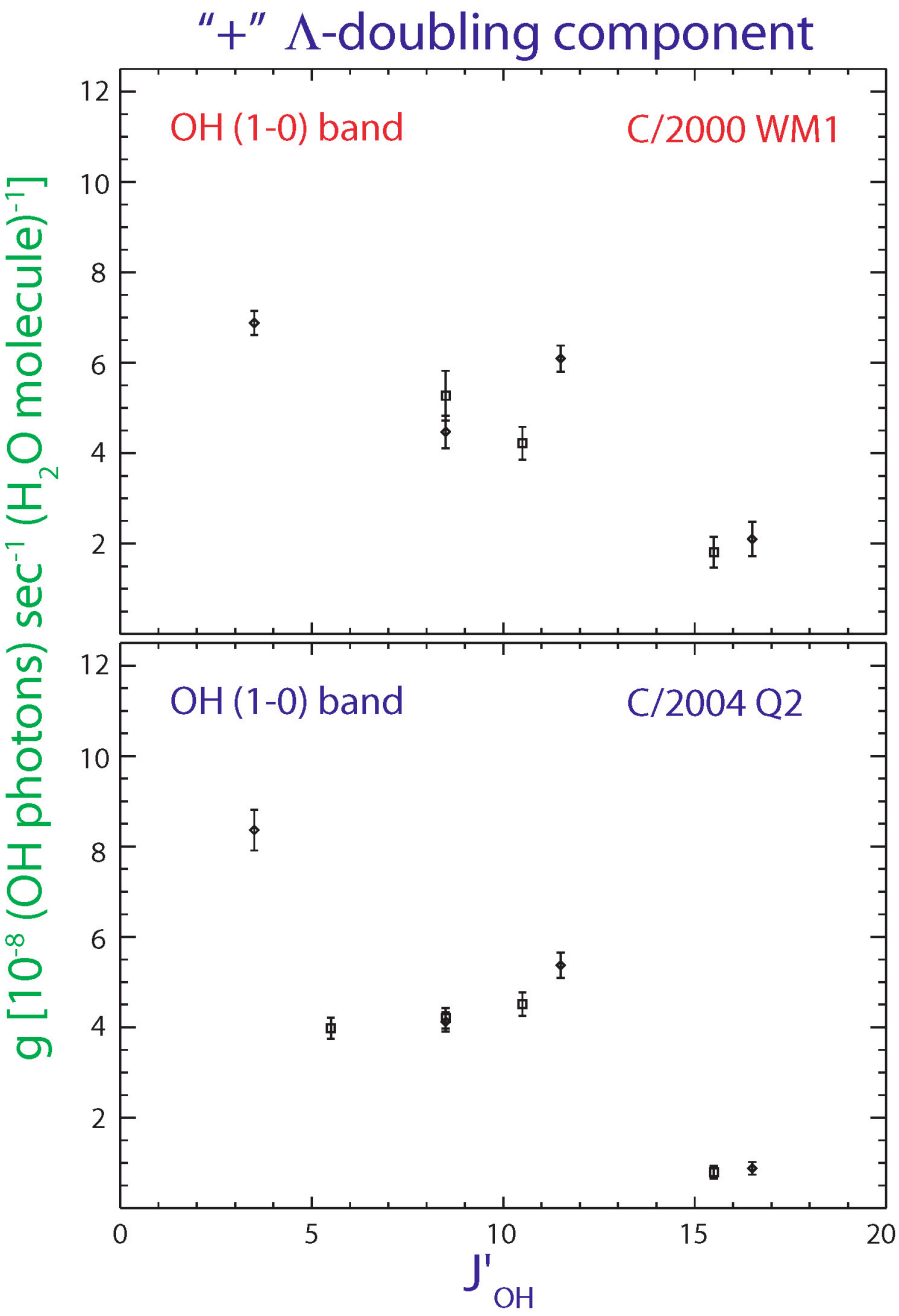

Fig. 28 – Effective g-factors at 1 AU for the OH PE (1-0) band lines detected in C/2000 WM1 and C/2004 Q2 ("+" lines).

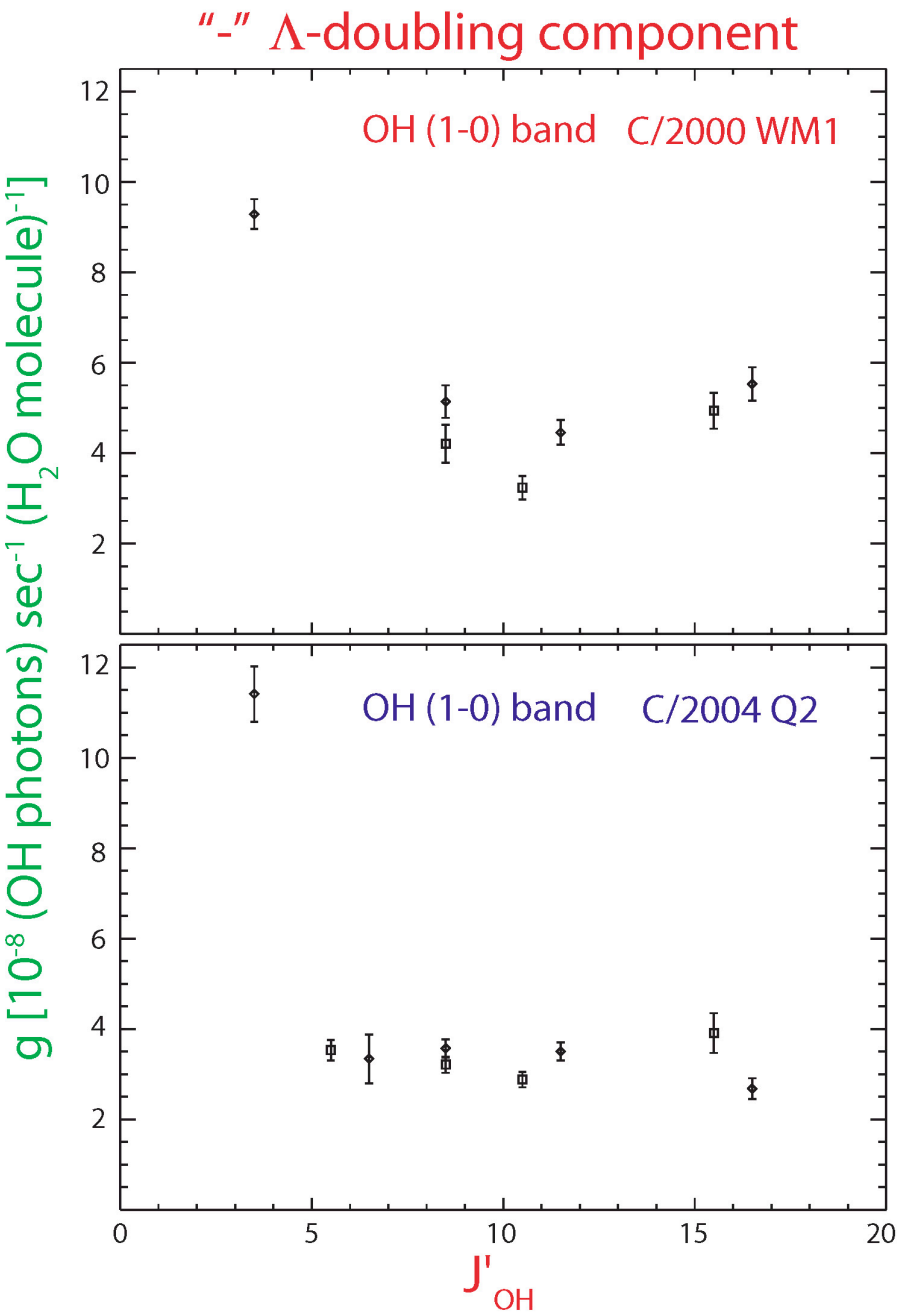

Fig. 29 – Effective g-factors at 1 AU for the OH PE (1-0) band lines detected in C/2000 WM1 and C/2004 Q2 ("–" lines).

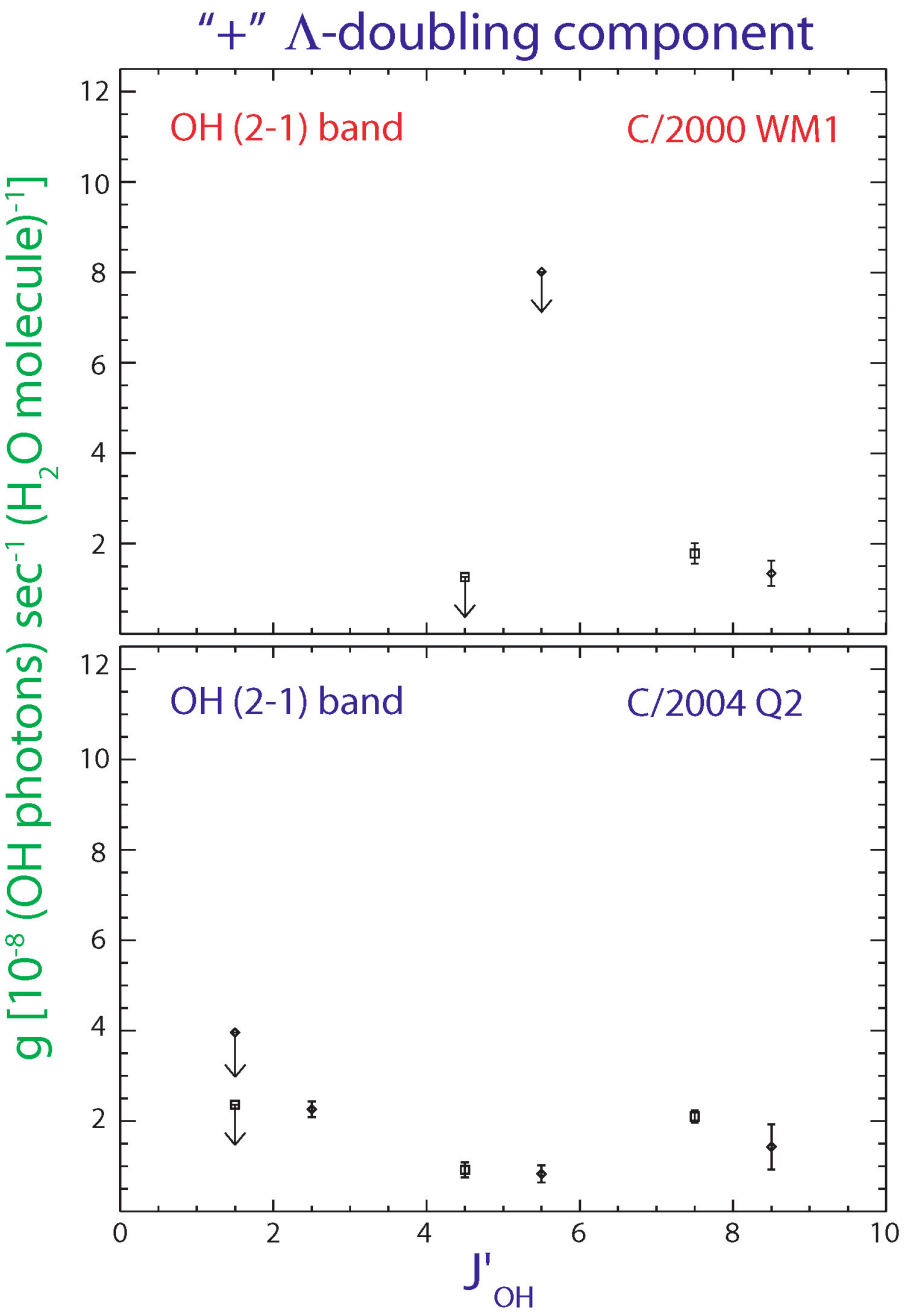

Fig. 30 – Effective g-factors at 1 AU for the OH PE (2-1) band lines detected in C/2000 WM1 and C/2004 Q2 ("+" lines).

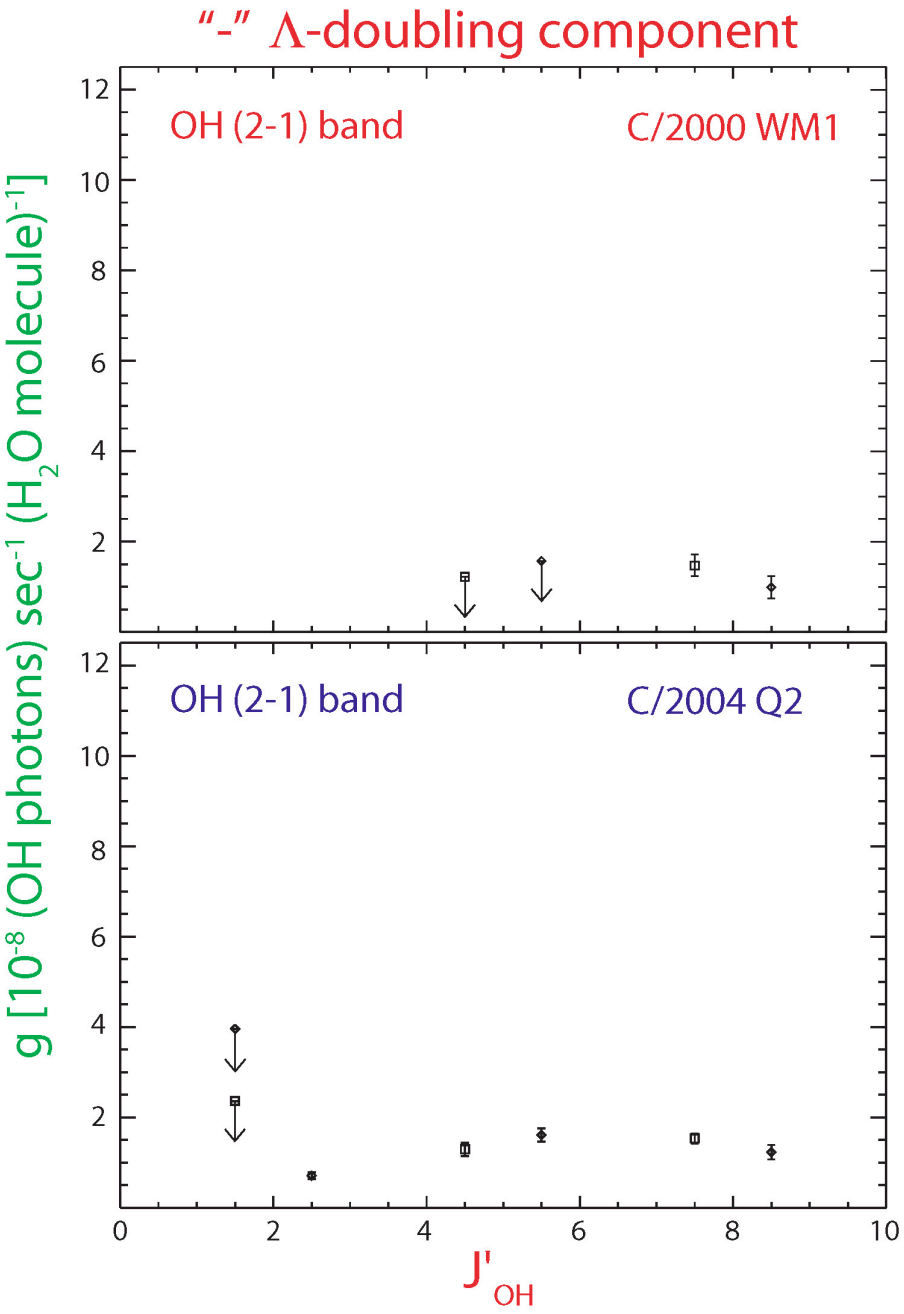

Fig. 31 – Effective g-factors at 1 AU for the OH PE (2-1) band lines detected in C/2000 WM1 and C/2004 Q2 ("–" lines).

Fourteen 1-0 band lines are presented in the WM1 plots; seventeen for Q2. In addition, the blended P3.5 2^+ and 2^- transitions are unambiguously detected in order 26 near 3422 cm^{-1} in WM1. However, their combined g-factors cannot be constrained because the atmospheric transmittance varies significantly in the narrow spectral range containing these lines. But this doublet is definitely strong, because it was detected in WM1 in spite of overall low transmittance (< 40%). In Q2, which has a different Doppler shift, this doublet falls into the core of a strong telluric line.

The weaker 2-1 band lines from the quadruplet of N'=9 (order 23) are detected in both comets. Order 24 samples the N'=5 quadruplet, detected only in Q2. For WM1 3σ limits (based on stochastic noise) are shown on the plots. The lowest J' lines from the 2-1 band are sampled in order 25. They are shown as equal upper limits on the "+" and "–" plots (Q2) because the two Λ-doublet components are blended.

The new OH emission efficiencies are the first calibrated by H_2O production rates derived from truly simultaneous measurements of OH and H_2O. This was enabled by the development of H_2O fluorescence models for the 2.9 μm region allowing to quantify H_2O and OH within a single NIRSPEC setting in the L-band. Another critical implication of the results from the previous chapter should be stressed here. Before quantifying the 2.9 μm bands, H_2O production had been derived from a few lines in the M-band, making simultaneous detections of H_2O and OH impossible. To calibrate two combined g-factors (for one quadruplet and one doublet) in C/1999 H1 (Lee) Mumma et al. (2001a; 2001b) compared H_2O data taken on 1999 August 20 (3 lines in the M-band near 5 μm) to OH data from 1999 August 21. The importance for sampling OH and its calibrator H_2O within the same NIRSPEC setting has already been emphasized – it is the equally

significant as in deriving mixing ratios between trace constituents and H_2O.

5.15 Uncertainties in the g-Factors I: Stochastic and Systematic Errors

These OH g-factors are also the first for which uncertainties have been evaluated – an important part of any empirical analysis. The errors showed in the plot have both stochastic and systematic component. The former reflects photon noise, the latter - the error in the calibrator, i.e. the H_2O production rates discussed in the previous chapter. The latter error is included in the plots because they compare g-factors in two different comets (with water production rates varying by an order of magnitude). For both WM1 and Q2 this error is about 3% of a given emission efficiency value.

It was already pointed out that the major source of uncertainty in the Q2 data set is related to the assumptions made in the choice of growth factors for the emissions for which Q-curves cannot be constructed. This uncertainty cannot be quantified, but a clear strategy to test the aforementioned assumptions was outlined. For the other emissions (i.e. the 1-0 band lines from orders 23, 24, and 26), the errors in the growth factors ratios (GF_{OH}/GF_{H2O}) are 4-5 %.

5.16 Uncertainties in the g-Factors II: Possible Spectral Confusion or Marginal Identification

Spectral confusion between lines of different species is always a reason for caution. Because of the abundance of molecular vibrational transitions in the NIR, this

region can contain overlapping transitions from different species. Considering a blend, one needs to account not only for proximity in line positions, but also for the expected line strengths and the abundances of the involved species. If an emission feature identified as a pure OH line has a contribution from another molecule, the resulting OH g-factor will be an upper limit. This possibility has been carefully evaluated order by order and line by line.

The quadruplet in order 21 (P16.5 2; P17.5 1) is in the spectral region of H_2CO transitions. Their Q-branch near 2781 cm^{-1} is indicated in the spectra of both WM1 and Q2. There is obviously a lot spectral structure to the left of the Q-branch (at higher frequencies). It contains an envelope of multiple overlapping H_2CO transitions and the four OH lines. In parallel to this project, colleague Mike DiSanti is studying the relative intensities of formaldehyde lines as a function of temperature. He is testing the best available H_2CO fluorescence models via correlation analysis against data from several Oort cloud comets. The formaldehyde model does not predict H_2CO contamination for the two middle lines in the OH quadruplet (P16.5 2^+ and P17.5 1^-). On the other hand, the P16.5 2^- and P17.5 1^+ lines could have some H_2CO contribution, depending on the H_2CO abundance in a particular comet. The degree of contamination can be estimated by comparing the relative intensities of unblended H_2CO lines to the expected strengths of the blended H_2CO lines in the model. The contamination is expected to be most serious for the P17.5 1^+ line (the one with the lowest frequency on Figure 4H and 4I), suggesting it may be even weaker than implied by its small g-factor. However, the stronger P16.5 2^- line is always detected at the right position in comparable intensity to P17.5 1^-, regardless

the H_2CO abundance in the particular comet, suggesting that blends with H_2CO in this region may not be severe.

The case of WM1 is less straightforward. Preliminary results from DiSanti indicate that there is a substantial spectral structure that is not accounted for in the H_2CO fluorescence model. Thus it is possible that all four OH lines have blended components and the WM1 g-factors might be overestimated. The effect, if present, could be more severe for the two weaker OH lines (P16.5 2^+ and P17.5 1^+). These lines have intensities comparable to the unexplained features. For this reason they are indicated with "?" in the WM1 spectrum. On the other hand, the two strong lines (P16.5 2^- and P17.5 1^-) have intensities well above the envelope of H_2CO and unknown features. Considering their detection in multiple comets, it is reasonable to assume that they are dominated by an OH emission. Nevertheless, their derived g-factors in WM1 are likely upper limits.

Even when contamination is considered, it is clear that the P16.5 2^- and P17.5 1^- lines are much stronger than the other two quadruplet members: P16.5 2^+ and P17.5 1^+. This preferential population in one of the L-doublet components for high-J OH levels will be emphasized in the discussion of relative g-factors (Chapter 6).

In order 23 most OH lines are unambiguously detected. An occasional exception is the P9.5 1 doublet from the 2-1 band. The primary question for these lines is not a blend with another species but the security of their detection. The P9.5 1+ line is indicated with "?" in the Q2 spectrum (Figure 4G) because it has low atmospheric transmittance (respectively high photon noise, considering that the low transmittance regions contain strong sky emission). Favoring detection is the fact that accounting for

telluric extinction, the g-factors of the two P9.5 1 lines agree between WM1 and Q2 (even after considering the uncertainty in the OH growth factor in Q2).

Order 24 has two strong OH doublets in the 1-0 band: P9.5 1 (near 3189 cm^{-1}) and P9.5 2 (near 3046 cm^{-1}). P9.5 1^+ and P9.5 2^+ have potential contributions from NH_2. However this possibility is unlikely because the corresponding NH_2 bands are weak. Some quantitative analysis has been done colleague Karen Magee-Sauer indicating that NH_2 is not in high abundance in WM1. In addition for both WM1 and Q2 the measured central frequencies of the two emission features correspond to OH, and are not shifted due to a major NH_2 contribution. Therefore the two "+" lines assigned to OH are not expected to have significant contamination from NH_2.

Order 25 contains lines from multiple species and is the most problematic in terms of spectral overlaps. From the 1-0 band, P6.5 2^- and P7.5 1^- likely are blended with P-branch C_2H_2 lines, which are detected in the same order. Therefore, the reported g-factors should be considered upper limits. The P7.5 1^- might have addition contamination by NH_2.

Among the 2-1 band lines, the P2.5 1^+ and 1^- lines are blended together similarly to P2.5 2^+ and 2^-. Because the distance between the Λ-doubling levels decreases as J decreases, the lowest J components cannot be resolved. The P2.5 1 doublet might have minor contamination from NH_2. Quantitative evaluations of the mentioned contributions from C_2H_2 and NH_2 can be done after the abundances of these molecules are evaluated in WM1 and Q2 (Magee-Sauer, in preparation).

Finally order 26, contains multiple H_2O hot-band lines, however, line positions (and in most cases intensities) are well-understood. The H_2O and the few unknown lines

are well separated from the OH 4.5 1^+ and 1^- lines (Figure 4A; 4B and so the OH emission can be presumed to be clear of contamination from other species.

5.17 Uncertainties in the g-Factors III: Possible Contribution from Fluorescent OH

Even though OH shows a parent-like spatial distribution indicative of a prompt emission mechanism, small contributions to the profile from extended fluorescent emission can not be excluded for the detected low-J lines. Determining extended emission from the shapes of spatial profiles is a very coarse technique, so the spatial distribution of OH could still resemble that of H_2O even if a small fluorescent component was present. For the P-branch region of the 1-0 band, Bockelee-Morvan & Crovisier (1989) predicted that all OH lines at wavelengths longer than ~2.95 μm should be entirely due to prompt emission, while lines between 2.88 and 2.95 μm may be mixed.

With very high degree of certainty the (1-0) band OH lines detected in orders 24, 23, and 21 are entirely due to PE. Indeed these are highly rotationally-excited states (J' > 7.5).

The strong doublet in order 26 corresponds to J' = 3.5 and therefore might have a fluorescent component, however, based on spatial data the prompt emission component dominates. Very High S/N spectra nodded off-chip would be more appropriate for this problem as they would provide much more detailed spatial information. It is interesting to note that for all comets in which the doublet near 3407 cm^{-1} is detected, the P4.5 1^- line is stronger than the P4.5 1^+. Their upper states correspond respectively to the "+" and "–" Λ-doubling components of J' = 3.5. These states lie very close together and

therefore it is hard to explain this asymmetry in line strengths solely by fluorescence from the solar continuum. Such asymmetries could be observed if the OH excitation is at least partly related to the dissociation. Either way it is safer to interpret the reported emission efficiencies for low-J lines as equivalent g-factors for *OH emission* versus for *OH prompt emission*.

The low-J lines in the 2-1 band in order 25 are too weak to examine their spatial distribution for signatures of fluorescence. Considering their higher vibrational excitation, they are more likely to be at least partly PE. The 2-1 band lines in order 24 and 23 are with high degree of certainty PE, based on rotational and vibrational excitation.

5.18 Implication for Deriving H_2O Production Rates from OH in Future Comets

Our main purpose was to provide a set of equivalent g-factors measured in at least one comet. This task is accomplished with the derived parameters in WM1 and Q2. It was essential to quantify multiple individual OH lines and to evaluate critically their uncertainties. These have been the two key elements lacking in the very limited previous work on IR OH PE. Deriving equivalent OH g-factors in two comets completes the initial development of the second method to measure H_2O production in the IR. With the two new g-factor data bases (WM1; Q2) the main issue about the IR OH PE shifts from completing the methodology and building an initial data set to assessing the reliability of this method.

The strategy is to use the new parameters to obtain H_2O production rates in future

comets. Our group intends to do this on a regular basis following the methodology presented above. This task can be accomplished now more efficiently by splitting it among group members working on molecules co-measured with OH within a narrow spectral interval. For example DiSanti's work uses the new g-factors of the bright P17.5 1^- and P16.5 2^- lines to estimate H_2O production rates simultaneously with that of H_2CO.

In the cases when OH is co-measured with H_2O, direct comparison between the resulting water production rates would update the existing g-factors, thereby gradually building a multi-comet data base for the emission efficiency of any given line. The clear implication is evaluating and understanding the degree of comet-to-comet variability in these empirical parameters.

We have already advanced in that direction by evaluating OH g-factors for the N' = 11 quadruplet near 3046 cm^{-1} for six observing dates in four comets. They include WM1 and Q2, but also C/1999 H1 (Lee) and C/2001 A2 (Linear). As mentioned above, the study of the latter two comets precedes the results for WM1 and Q2 and is described in detail in a refereed paper (Bonev et al. 2004). We focused first on a single quadruplet while optimizing the methodology, which involved not only the particular OH problem, but the transition from CSHELL data of two very bright comets nodded off-chip (Hyakutake; Hale-Bopp) to NIRSPEC data of weaker comets nodded on-chip, and the transition to higher precision algorithms for data reduction, briefly described in Chapter 3 and in Appendix 2.

TABLE 5
Effective g-factors at 1AU in four comets for
the OH PE (1-0) band quadruplet near 3046 cm^{-1}

Line ID	ν_0 [cm^{-1}]	Comet/UT Date	g-factor at 1 AU
P11.5 2+	3047.4576	C/1999 H1 (1999 Aug 19)	5.53 ± 2.66
		C/1999 H1 (1999 Aug 21)	4.60 ± 0.62
		C/2001 A2 (2001 Jul 9)	3.81 ± 0.64
		C/2001 A2 (2001 Jul 10)	2.68 ± 0.40
		C/2000 WM1 (2001 Nov 25)	4.22 ± 0.36
		C/2004 Q2 (2005 Jan 19)	4.51 ± 0.26
		Weighted Mean	**4.07 ± 0.31**
P11.5 2^-	3046.5508	C/1999 H1 (1999 Aug 19)	4.54 ± 1.29
		C/1999 H1 (1999 Aug 21)	4.42 ± 0.54
		C/2001 A2 (2001 Jul 9)	2.39 ± 0.40
		C/2001 A2 (2001 Jul 10)	2.75 ± 0.39
		C/2000 WM1 (2001 Nov 25)	3.24 ± 0.26
		C/2004 Q2 (2005 Jan 19)	2.88 ± 0.17
		Weighted Mean	**3.00 ± 0.19**

The g-factor is assumed to vary as the inverse square of the heliocentric distance. The values for C/1999 H1 and C/2001 A2 are from Bonev et al. (2004). The uncertainties are 1-σ standard errors; formal 95% confidence intervals correspond to 2.57 σ. The table continues on the next page.

TABLE 5 - continued

Line ID	ν_0 [cm^{-1}]	Comet/UT Date	g-factor at 1 AU
P12.5 1+	3044.4280	C/1999 H1 (1999 Aug 19)	5.76 ± 1.22
		C/1999 H1 (1999 Aug 21)	4.54 ± 0.58
		C/2001 A2 (2001 Jul 9)	4.58 ± 0.70
		C/2001 A2 (2001 Jul 10)	4.22 ± 0.58
		C/2000 WM1 (2001 Nov 25)	6.09 ± 0.29
		C/2004 Q2 (2005 Jan 19)	5.37 ± 0.28
		Weighted Mean	**5.41 ± 0.28**
P12.5 1$^-$	3043.2137	C/1999 H1 (1999 Aug 19)	2.88 ± 0.60
		C/1999 H1 (1999 Aug 21)	3.98 ± 0.46
		C/2001 A2 (2001 Jul 9)	3.99 ± 0.64
		C/2001 A2 (2001 Jul 10)	2.56 ± 0.34
		C/2000 WM1 (2001 Nov 25)	4.46 ± 0.27
		C/2004 Q2 (2005 Jan 19)	3.50 ± 0.20
		Weighted Mean	**3.62 ± 0.28**

Table 5 presents all available g-factors for the N' = 11 quadruplet. The uncertainties in their weighted means account for comet-to-comet variability. Note that the WM1 and Q2 g-factors have fallen mostly within the already established range of the measurements in C/1999 H1 and C/2001 A2, so the overall uncertainty has been reduced. These four g-factors are now the most reliable for independently deriving Q(H_2O) in future comets. This is not only because the larger number of comets in which the four

lines were quantified. In light of the previous two sections, the 1-0 band quadruplet of N' = 11 has no known spectral overlap with other cometary emissions and the corresponding rotational excitation is high enough to ensure that a contribution from OH fluorescent emission would be minimal or (most likely) none.

Analysis of other OH lines has been following. In addition to the aforementioned work of DiSanti on high-J lines, my REU student Yana Radeva (Connecticut College / U. of Maryland) analyses quantitatively the two P4.5 1 lines in order 26 in C/2001 A2. This way both high- and low-rotational excitation will be covered in another comet.

With these studies in progress, the comet-to-comet variation of g-factors for lines outside the N' = 11 quadruplet are still restricted to Q2 and WM1. The non-stochastic uncertainty of a weighted mean of just two measurements cannot be well constrained. But we present the ratio g(WM1)/g(Q2) in Figures 32. Here "g" designates any emission efficiency calibrated in both comets. Figure 32 should be considered in the context of the g-factors being used without having an independent calibrator of $Q(H_2O)$ (a single-order instrument for example).

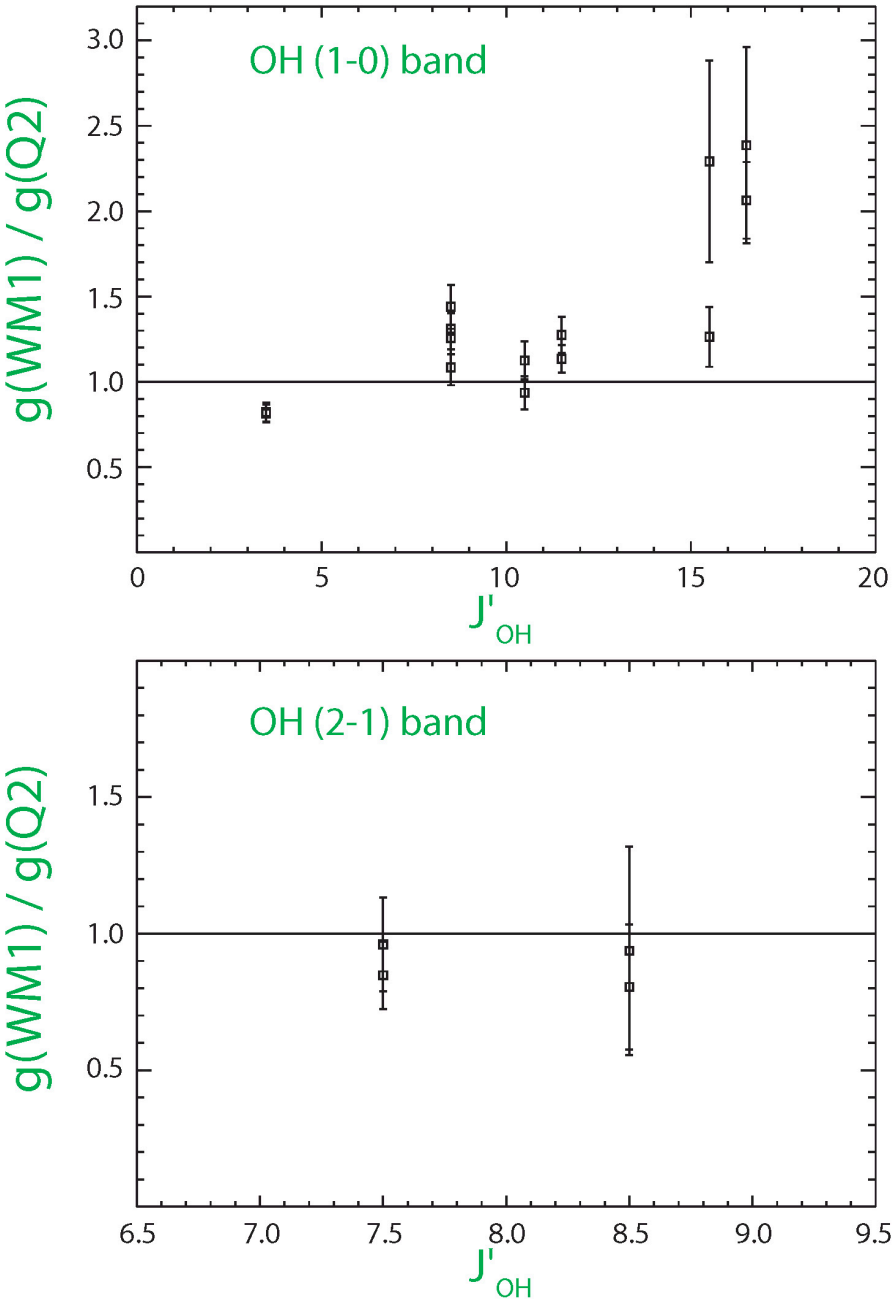

Fig. 32 – Ratio between the equivalent g-factors derived for eighteen OH lines in C/2000 WM1 and C/2004 Q2.

The g-factors for the 2-1 band are in very good agreement in WM1 and Q2, suggesting the N' = 8 quadruplet could be used to derive water production rates. This would be useful because at least some of these lines can be sampled simultanously with CH_4 even within a very narrow spectral grasp. The same is valid of course for the N' = 11 lines from the 1-0 band. The first application of deriving water production rates from the OH lines in order 23 will be presented in the end of this chapter.

Considering uncertainties, the g-factor ratios are close to unity for most 1-0 band lines from order 24 and 23, and for P16.5 2^- line from order 21. The positive ratios for the other strong line in order 21 (P16.5 1^-) raise a caution as we wish to use it for extracting mixing ratios between H_2CO and H_2O. The discrepancy can be resolved assuming that the WM1 values are overestimated because of the spectral overlap with unknown features as discussed above.

Finally the ratio between the two components in the P4.5 1 doublet in order 26 is practically the same for WM1 and Q2, so g(WM1)/g(Q2) appears as a single point in Figure 32. The exact agreement should not be overemphasized. Yana Radeva's REU work has found some variation in the relative intensities of the P4.5 1^- and P4.5 1^+ lines in C/2001 A2. What should be emphasized is that despite some variations in the g-factor ratio for the P4.5 1 doublet, the P4.5 1^+ appears stronger in all comets we have observed with NIRSPEC (A more detailed discussion on relative g-factor within a single comet is presented in Chapter 6).

In cases requiring the usage of g-factors available only on WM1 and Q2, appropriate uncertainties will of course be considered avoiding underestimate.

The main purpose of this chapter has been to launch a multi-line data base of OH emission efficiencies. This goal has been achieved with the results shown on Figures 28-31. Apart from specific evidence of comet-to-comet variability in the new parameters, it is appropriate to ask whether hidden factors might limit the general applicability of deriving H_2O production rates via OH IR PE. To conclude this chapter, the following discussion outlines some important issues in that aspect.

5.19 Discussion

5.19.1 Prompt Emission from OH* and [O(^{1}D)]:

Quantifying water production via OH IR PE has a related technique at optical wavelengths. PE from dissociatively excited [O(^{1}D)] (at 630 nm) has been a very useful proxy for water for many years (cf. Morgenthaler et al. [2001]). The initial steps in developing this method were consistent with water being the sole parent of [O(^{1}D)] (e.g. Spinrad 1982). However, it is now accepted that a significant fraction of [O(^{1}D)] is produced from other precursors (OH, CO, CO_2), especially in the outer coma (e.g., Festou & Feldman 1981; Morgenthaler et al. 2001). In addition, because [O(^{1}D)] is a forbidden transition, it has a comparatively long lifetime (120 s), so collisional quenching may be important in high-density regions near the nucleus.

By contrast, no major production mechanisms for OH PE, other than photolysis of H_2O, are known, and the radiative lifetime of OH* (several milliseconds) is much shorter than that of [O(^{1}D)]. As already pointed out, collisional vibrational relaxation of dissociatively excited OH is expected to be negligible throughout the observed coma.

The [O(^{1}D)] technique is very useful because it has been developed for many

years and its limitations are well understood. Also, optical observations can be conducted in much more facilities than NIR spectroscopy. This technique, however, cannot be a substitute for the development of an indirect method in IR simply because it covers a different frequency range.

5.19.2 Possible Effects of Solar Activity

The effective g-factor might depend on solar activity. The water photolysis rates (respectively OH production) vary in both H_2O bands with solar activity, but the dependence is stronger in the second band, which covers shorter wavelengths. The Ly-α intensity varies significantly during the 11-year solar cycle and also on shorter time scales, but this channel provides less than one-half of the total production of OH (in any state) even under active Sun conditions. Although the dissociation in the second absorption band leads predominantly to OH in its ground electronic state ($X^2\Pi$), there are other channels with substantial contribution (Harich et al 2000). Conversely the first absorption band (less sensitive to solar activity) leads exclusively to OH[$X^2\Pi$], the state within which we sample ro-vibrational transitions.

Comets C/2000 WM1 and C/2001 A2 were observed during the maximum phase of solar cycle 23; C/1999 H1 – in the near maximum phase. On the contrary C/2004 Q2 was observed well after the cycle had entered its "decay" phase approaching minimum. Reduced UV flux under low solar activity qualitatively implies smaller dissociation yields for water, respectively reduced OH PE efficiency. It is however premature to assign the difference between the WM1 and Q2 g-factors for high-J lines unambiguously to solar cycle effects, as we have only one comet sampled near solar minimum. Moreover statistically, the Q2 g-factors for the quadruplet near 3046 cm-1 do not differ

significantly from the values found in C/1999 H1 and C/2001 A2. This is an argument against dominant solar cycle dependence, consistent with the notion that the H_2O dissociation is dominated by the less sensitive to solar activity first absorption band.

Admittedly, the above treatment is simplistic because solar Ly-α varies significantly on short time scales (days), especially around solar maximum. The correct treatment of the problem is to account for such variations for the heliocentric latitude and longitude of the comets, which is out of the scope of this work. Observations in the next 2-3 years will be most diagnostic about the magnitude of the solar activity effects on the g-factors. The Scwabe-Wolff cycle is approaching its minimum. Its conditions are expected to contrast the near maximum activity during 1999-2002 even after considering short term variations in the UV.

5.19.3. Possible Effects of Heliocentric Distance

The g-factors are assumed to vary as the inverse square of the heliocentric distance. This section discusses possible variation other that the assumed R_h^{-2} dependence.

Kawakita et al (2003) suggested that the OH PE mechanism should be re-examined for heliocentric distances around 1.5 AU. They did not detect OH in comet 153P/Ikeya-Zhang when this comet was near R_h =1.5 AU. Their observations were done at a resolving power of 10,000 with IRCS at the Subaru Observatory atop Mauna Kea. These authors derived an upper limit for the H_2O production rate based on the integrated intensity of the quadruplet near 3046 cm^{-1} (P11.5 2; P12.5 1). The only emission efficiency available in the literature for them was the preliminary combined g-factor for the whole quadruplet given by Mumma et al. (2001a) without any uncertainty estimate.

Kawakita et al.'s upper limit was significantly lower than the H_2O production rate deduced from radio observations of OH spanning several dates prior to and following their observations, leading these authors to the suggestion that the emission mechanism of OH has to be reexamined for larger heliocentric distances. It was considered that Mumma et al. observed C/1999 H1 (Lee) near 1 AU, while Kawakita et al. observed 153P/Ikeya-Zhang at 1.5 AU. If a change in the g-factor was the sole cause of Kawakita et al's non-detection, this g-factor would have been significantly lower than the value of Mumma et al. (2001a) of $2 \times 10^{-7} R_h^{-2}$ photons sec^{-1} $(H_2O \text{ molecule})^{-1}$. This value is now an upper end of what we have measured in four comets (Table 5).

In our 2004 paper on OH PE (Bonev et al.) we stated that the suggested alteration of the PE mechanism at large heliocentric cannot be overruled until more comets were observed in such conditions. Our limited NIRSPEC data base at the time for observations at R_h~1.5-1.6 AU did not support but could categorically disprove a drastic change (other that the inverse square dependence) in OH g-factors with respect to $R_h \sim 1$ AU. The clouds started to clear in late 2004. We detected multiple OH PE lines in Q2 on Nov. 28, 2004, an observing run with other primary object, which we used to also accumulate some Q2 data (20 min on source in the KL2 setting when the comet was significantly less bright that during the January observation of Ellis & Stark).

The comet was at 1.493 AU. Despite a problem with abnormally high dark current (the detector temperature was about 10K higher than nominal), the detections of seven high-J OH lines were unambiguous. Only the P9.5 1^- lines was not detected (it has the smallest g-factor among the eight lines sampled in order 23). Figure 33 shows a spectrum from order 23 similar to the one taken ~40 days later, which is displayed in our

spectral gallery. The first thing to notice was that the relative intensities of the OH lines were similar to what we have already observed at smaller R_h, with P12.5 1^+ being the strongest.

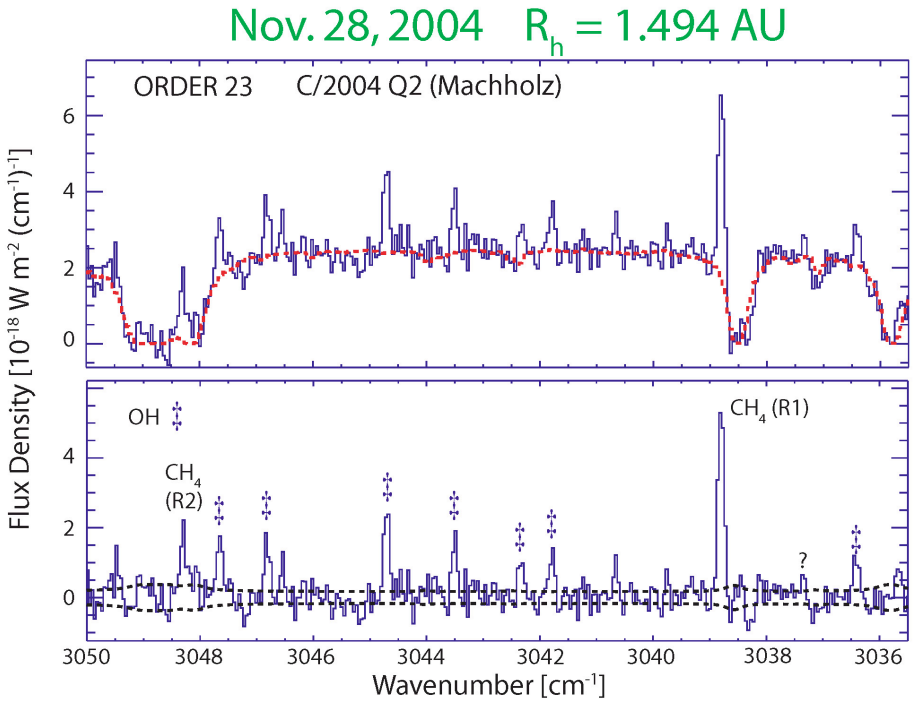

Fig. 33 – Detections of OH prompt emission at heliocentric distance near 1.5 AU; C/2004 Q2, Nov. 28, 2004.

After the observing run we immediately analyzed the Q2 data in an attempt to request more observing time in January as the comet brightens (this effort lead to the Q2 data analyzed here). Subsequently a more careful reduction and analysis allowed to determine H_2O production rates directly from the non-resonance fluorescence lines in order 26, and indirectly from the OH lines detected in order 23. In should be pointed out that the OH-based retrieval is completely independent of the H_2O retrieval. This former retrieval is based on the PE g-factors derived from the January data with only R_h^{-2} correction for the different heliocentric distance in both dates.

The results are displayed on Figure 34. The production rate based on the H_2O

"hot-band" emission in order 26 is indicated with a red dashed line (this value was derived by Mike DiSanti using the methodology from Chapter 4). The production rates from each individual OH line in order 23 are shown as triangles. Their error bars reflect the S/N ratio in these lines. Finally the green dashed line indicated the weighted mean of individual production rates based on OH PE.

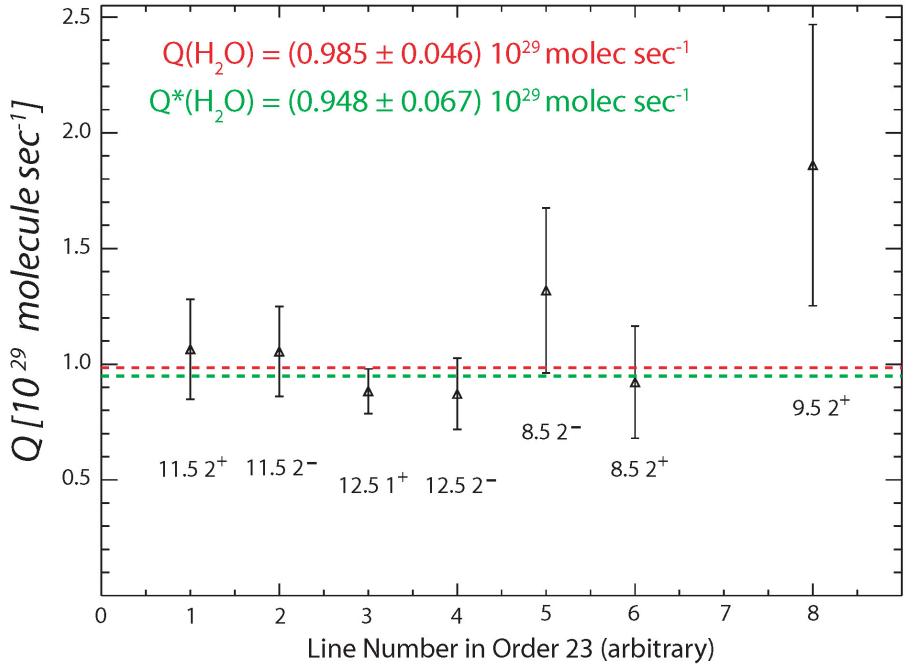

Fig. 34 – H_2O production rates from the OH PE lines shown on Figure 33. The g-factors derived from the January Q2 data have been used.

The agreement between the directly and indirectly measured H_2O production rates (Q and Q*) is very good. These results reject any drastic change (by an order of magnitude as Kawakita et al. imply) in the PE g-factors near 1.5 AU for the first comet in which there are secure detections at this heliocentric distance. Moreover, the agreement with the independently derived value from the water lines in order 26 further validates the emission efficiencies for Q2 derived from the January 19 data.

At present we have secure detections of IR OH PE near 1 AU (C/1999 H1,

C/2001 A2), near 1.2 AU (C/2004 Q2), at 1.35-1.40 AU (C/2001 WM1), and at ~1.5 AU (C/2004 Q2). Aside from R_h^{-2} dependence, no correlation of emission efficiencies with heliocentric distance is implied by these data.

5.19.4. Possible Effects of the Rotational Temperature of the Parent

Is the effective g-factor for OH PE sensitive to the rotational temperature of the parent H_2O? Mumma et al. (2001a) suggested that the OH PE-based method for quantifying water production should be insensitive to T_{rot} because the end product distribution (OH*) tends to average over the initial state distribution of H_2O, causing the rotational distribution of OH PE to have "little memory" of the precursor's rotational distribution prior to photolysis. At first glimpse this contradicts the conclusions by Andresen et al. (1984) and Hausler et al. (1987) that the resultant OH rotational distribution is a strong function of the parent's rotational temperature. However, there is a principal difference between the "set-up" in all laboratory studies of OH production via H_2O photo-dissociation and the comets case. This point will be discussed in the following chapter which shifts the focus from absolute to relative g-factors.

Chapter 6

Comets as "ad-hoc" laboratories: the Rotational Population of Vibrationally-Excited OH produced by H_2O photolysis

The emphasis of the previous chapter was *absolute* g-factors for OH PE: the necessity for these parameters, their derivation, uncertainties, and usefulness in extracting water production rates in comets. The emphasis of this chapter shifts towards *relative* g-factors, which reflect the relative populations of ro-vibrationally excited OH (OH*[$^2\Pi$; v'=1]) produced by water photolysis. In other words, we shift the focus from the problem of H_2O production to H_2O photo-dissociation and the rotational distribution of the product (OH*). Inasmuch as the latter problem has played a principal role in unimolecular dissociation studies, it has received significant experimental and theoretical attention (e.g. Carrington 1964; Yamashita 1975; Andresen et al. 1984; Hausler et al. 1987; Engel, Schinke, & Staemlerr 1988; Harich et al. 2000; Nizkorodov et al. 2003). Our purpose is to augment the existing laboratory data base with a set of measurements of emission from a photo-dissociation product of cometary H_2O. We believe this is a pioneering effort and consequently it is important to distinguish the principal difference between the experimental set-up in a conventional research laboratory and in cometary

the ad-hoc "laboratory".

6.1 Principal Steps of H_2O Photo-dissociation

The process consists of two basic steps (following Hausler et al. 1987):

1. Absorption of a photon leading to an electronically excited state of the parent molecule.

2. Dissociation of that complex to the end products (radicals).

For a triatomic molecule ABC:

$$ABC(i) + h\nu \rightarrow ABC^* \rightarrow AB(f) + C$$

Here "*i*" is the initial quantum state from which dissociation starts. For water it is its ground electronic state; its first excited electronic state is unbound and consequently H_2O spectra do not include electronic transitions. The photon energy $h\nu$ giving rise to photo-dissociation of H_2O corresponds to UV wavelengths (Chapter 5). The type of final products and the quantum state distributions in which they are produced are governed by the unstable excited parent state *ABC**, which in turn depends on the photon energy and initial state population.

In comets the parent H_2O occupies different rotational states of its ground vibrational level, with the distribution governed mostly by collisions in the inner coma and described by the H_2O rotational temperature (Chapter 4). The IR spectra of the final product (OH*) originate from different $^2\Pi$ rotational levels of its ground electronic state. We focus on the 1-0 band transitions because they reflect a broad range of rotational quantum numbers, thereby allowing some investigation of the population distribution in the v'=1 level.

Ro-vibrationally excited OH*($^2\Pi$) can be produced by different dissociation channels depending on the photon energy. Dissociation in the first absoprtion band (FAB) (136-186 nm) leads almost exclusively to this state via the lowest excited electronic state of H_2O. Dissociation in the second absorption band (primarily by the solar Ly-α) (SAB) produces mostly OH*($^2\Pi$). However, other end state products like electronically excited OH*($^2\Sigma^+$) do not have negligible branching ratios.

In the laboratory, the OH*($^2\Pi$) distribution has been measured for dissociative frequencies from both the first and the second absorption bands. In the FAB the rotational distribution of OH*($^2\Pi$) can be approximated by temperatures from a few hundred to about one thousand K as the initial H_2O rotational temperature is varied between ~30 and ~300 K (Andresen et al. 1984; Hausler et al. 1987) . This is a strong dependence! In the SAB the OH*($^2\Pi$) rotational distribution can be extremely "hot", peaking at rotational quantum numbers N' ~ 40-45 equivalent to temperatures of tens of thousands of Kelvin (Harich et al. 2000). This is an extreme case in which about 70% of the initial excitation energy is deposited into OH rotational excitation. The OH distribution for v'=1 reported by Harich et al. has much lower secondary peaks near N'~12 and N'~17-23. Note that for their experiment the H_2O rotational temperature was ~10K. Note also the dramatic difference in the product distribution as compared to the results of Andresen et al. in the FAB.

So what is the principal difference between the setups in physical laboratories and our ad-hoc "laboratories" (the inner comae of WM1 and Q2)? In the former case the energy for dissociation comes from a monochromatic radiation in either the FAB or the SAB. For example Andresen et al. studied photo-dissociation at 157 nm; Hausler et al.

and Nizkorodov et al. – at 193 nm. Work in the SAB, including those involving end products other than OH*($^2\Pi$), has examined mostly dissociation at 121.6 nm (Ly-α).

The dissociation of cometary H_2O originates from a continuum of radiation in both absorption bands. This process therefore involves all possible quantum channels producing OH*($^2\Pi$) from a significant number of H_2O rotational states (unless T_{rot} of H_2O is very low). For only this case, Mumma et al. (2001a) predicted that the OH absolute g-factors would be relatively insensitive to the H_2O rotational temperature, at least for the range in T_{rot} sampled in Oort cloud comets (70-120 K). The overall similarity of the g-factors derived in two comets with rotational temperatures ~75 K and ~92 K weakly supports this conclusion.

From a theoretical standpoint, dissociation from non-monochromatic radiation in both absorption bands is the most challenging case. On the contrary, the theory of dissociation is best understood when H_2O is prepared in a specific initial rotational and vibrational state (Hausler et al. 1987; Nizkorodov et al. 2000) and when photolysis occurs at a monochromatic frequency in the FAB.

Finally the derived OH g-factors include the integrated effect of all processes leading to the production of OH*($^2\Pi$) in the v' = 1 level. The production of these excited states is initiated by photo-dissociation of H_2O. A significant fraction of the population in the first excited vibrational state of OH is expected to be a direct product of the parent's photolysis. For example, for dissociation at 157 nm (FAB) Andresen et al. (1984) found that the population ratios for vibrational levels 0, 1, and 2 are 1:1:0.15, respectively. In addition to direct production from H_2O dissociation, some decay from higher vibrational levels to OH*($^2\Pi$) at v' = 1 is also possible.

6.2 Relative g-Factors for the OH 1-0 Band Lines

Our purpose is to present the new comet data in a form suitable for interpretation by modelers of H_2O dissociation dynamics and end product distributions. We consider three principal questions:

1. What temperature ranges approximate the final OH*($^2\Pi$) rotational distribution in the v'=1 level?

2. How do the relative g-factors vary with OH* rotational excitation (i.e. with J')?

3. What are the ratios between the two Λ-doublet components of the sampled upper ro-vibrational states?

We are searching for trends in the relative g-factors in data from WM1, Q2, and future comets. Systematic effects identified in the rotational distribution of OH* among comets would provide valuable constraints on better understanding H_2O dissociation in the inner cometary coma.

6.3 Rotational Distribution I: Boltzmann Diagrams

Boltzmann analysis of the end OH* rotational distribution has been commonly used in the aformentioned studies in the FAB of H_2O. Figure 35 shows Boltzmann diagrams for WM1 and Q2. These diagrams include all 1-0 band transitions observed in both comets. Note that the zero point energy on the x-axis corresponds to the lowest level sampled (J'=3.5). This was done merely for convenience. The slope of course does not depend on the zero energy choice.

The first thing to emphasize is that the Bolztmann analysis is useful only to the extent that it provides the temperature range which approximates (to a greater or lesser extent) the observed rotational excitation in OH*($^2\Pi$; v'=1). While it is instructive to relate an equivalent temperature to the observed relative g-factors, there is no a-priori assumption that the end product follows a Bolztmann distribution, and that it is described by a single T_{rot} value. This is in contrast to the analysis of the H_2O hot-bands (Chapter 4), in which the assumption of a Boltzmann distribution in the ground vibrational state is physically justified in most cases. H_2O and OH* have completely different mechanisms for rotational excitation: collisions in the inner coma for H_2O versus dissociative excitation of the parent for OH*.

With the above consideration, the equivalent T_{rot} values for OH*($^2\Pi$; v'=1) are ~2214 K in WM1 and ~1980 K for Q2. These temperatures are not very meaningful because the approximation of a single T_{rot} is clearly inadequate. The deviations from the linear fit in both plots are systematic for the middle points and correlate between WM1 and Q2.

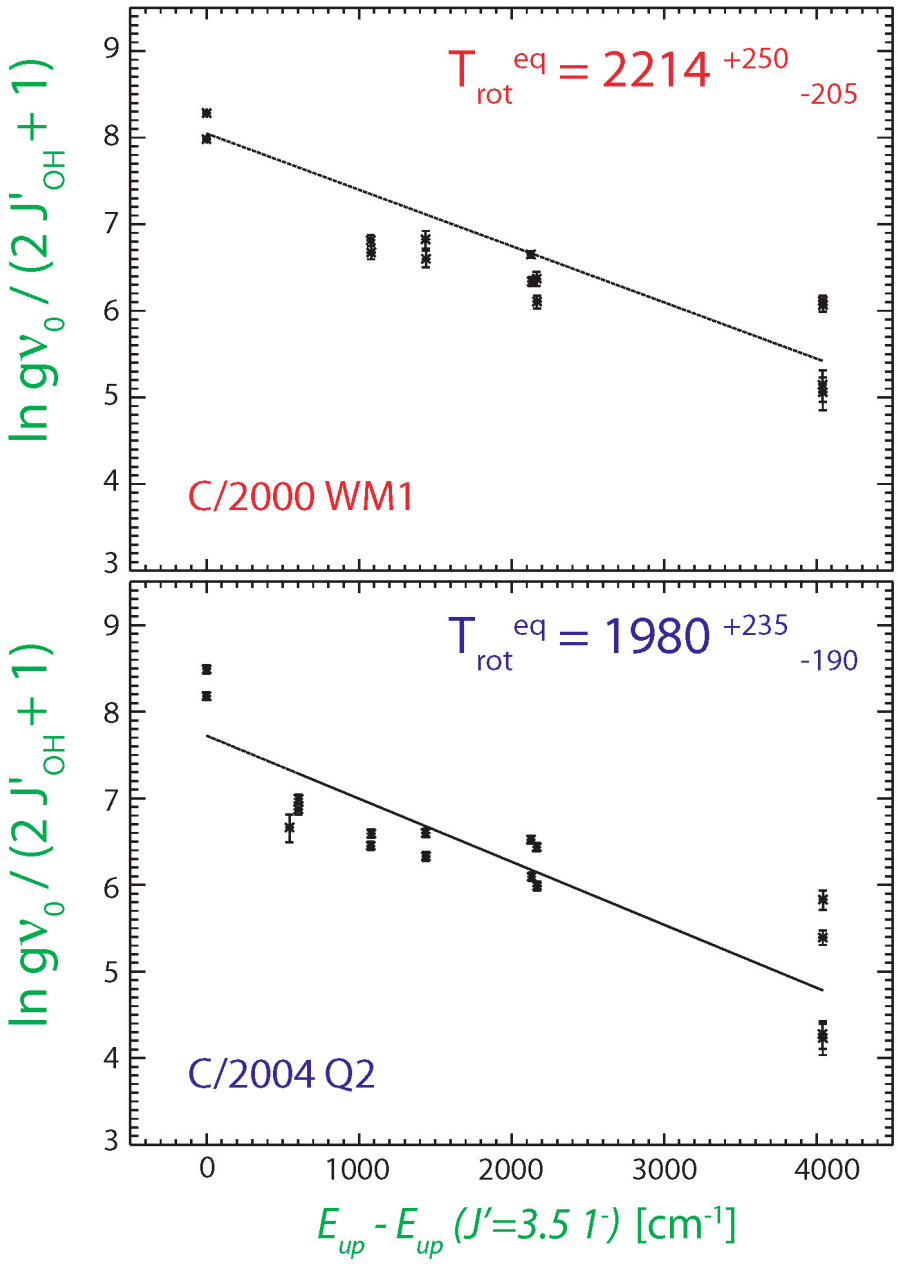

Fig. 35 – Boltzmann diagrams for all analyzed OH 1-0 band lines in the KL2 NIRSPEC setting.

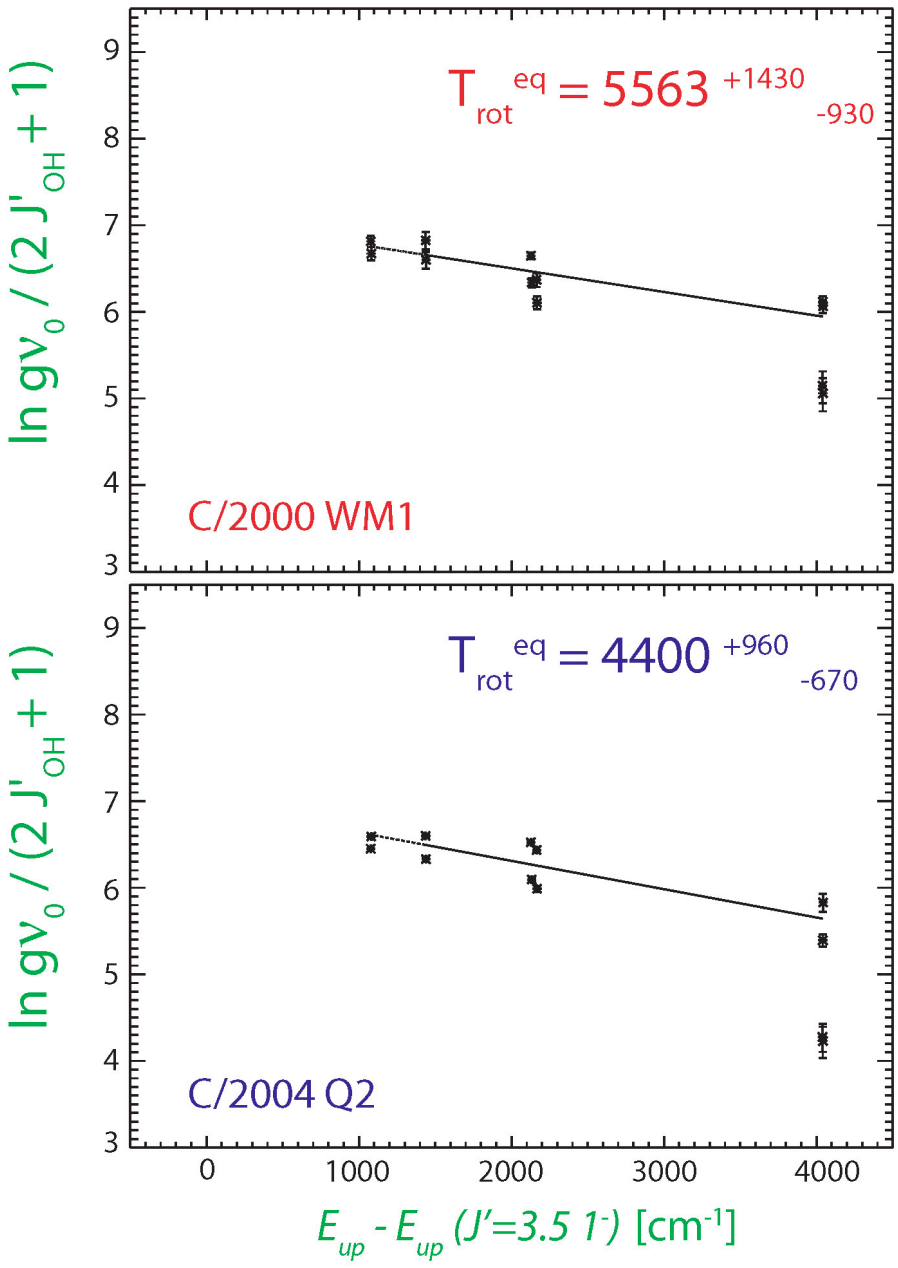

Fig. 36 – Boltzmann diagrams for the OH lines sampling 7.5 < J' < 17.5

We are clearly not observing a single OH* distribution, which is not particularly surprising. This is confirmed by the large change in equivalent T_{rot} after exclusion of the lowest-J lines: the Boltzmam diagrams on Figure 36 include only transitions with J' > 7.5. They result in much larger equivalent rotational temperatures, and these agree between the two comets within error. T_{rot} of several thousand Kelvin is expected for the observed levels of excitation. Such temperatures correspond to the sources used to produce OH IR laboratory spectra: an oxyacetylene flame at ~3000 K (Mailard et al. 1974) and plasma discharge at ~6000 K (Abrams et al. 1990). These temperatures are significantly higher than those retrieved in the laboratory after dissociation at 157 and 193 nm. These laboratory works do not cover excitation higher than J'~10.5.

The congruence of the equivalent temperatures for 7.5 < J' < 17.5 between WM1 and Q2 is an interesting result. But although the higher J' transitions are better approximated by a single T_{rot}, the limitations of the Boltzmann approach (as applied to OH*) are clear from Figure 36. The most striking example is the difference between the line intensities at the high energy (high-J) end, demonstrating that there are substantial effects which cannot be explained by a simple thermal distribution.

In principle, Boltzmann analysis can be done separately on the "1" and "2" transitions to sample the $^2\Pi_{3/2}$ and $^2\Pi_{1/2}$ rotational distributions, respectively. However we do not have enough lines of either type to produce more meaningful conclusions, compared to those based on the analysis already presented.

6.4 Rotational Distribution II: Relative g-Factors and Ratios Between the Λ-doubling Components

A simple plot of relative g-factors does not reveal explicitly an excitation-equivalent temperature but has the advantage of reflecting most directly the observed rotational distribution. The relative g-factors are proportional to the branching ratios for decay of electronically excited H_2O into different OH* states (if OH* is a direct dissociation product). Figures 37 and 38 present the ratios of the line-by-line emission efficiencies to the value for the P12.5 1^+ line. Figure 37 shows relative g-factors for the "+" lines and, therefore, the "–" Λ-doubling component for each given J' state (considering the selection rule of parity change). Figure 38 samples the "–" lines and, therefore, the "+"Λ-doubling component of J'. Note that in all subsequent figures in this chapter transitions from $^2\Pi_{3/2}$ and $^2\Pi_{1/2}$ are indicated with diamonds and squares respectively as on Figures 28 and 29 in Chapter 5.

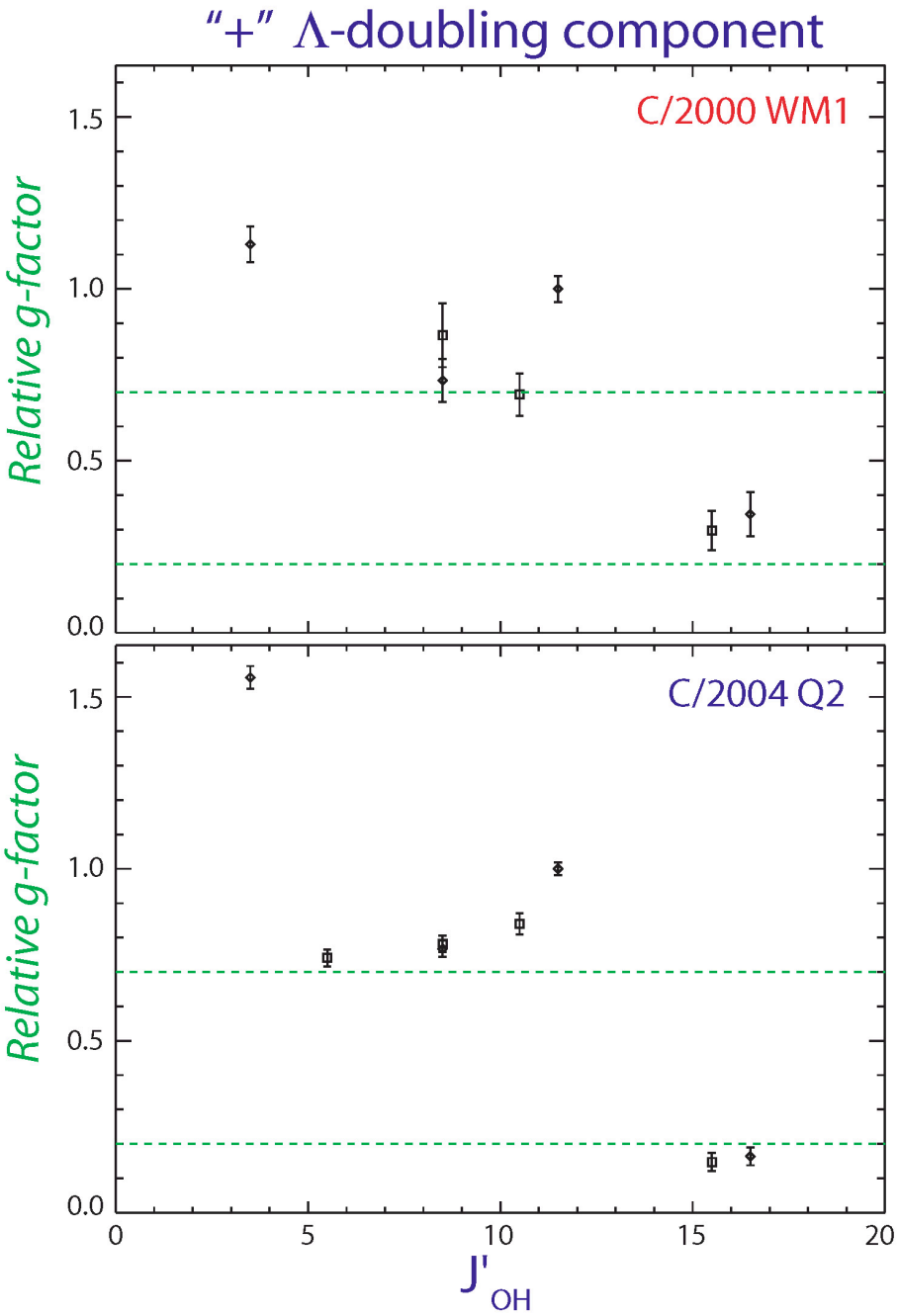

Fig. 37 – OH emission efficiencies of the "+" lines normalized to the g-factors of P12.5 1^+. The dashed green lines do not have any special significance except for guiding.

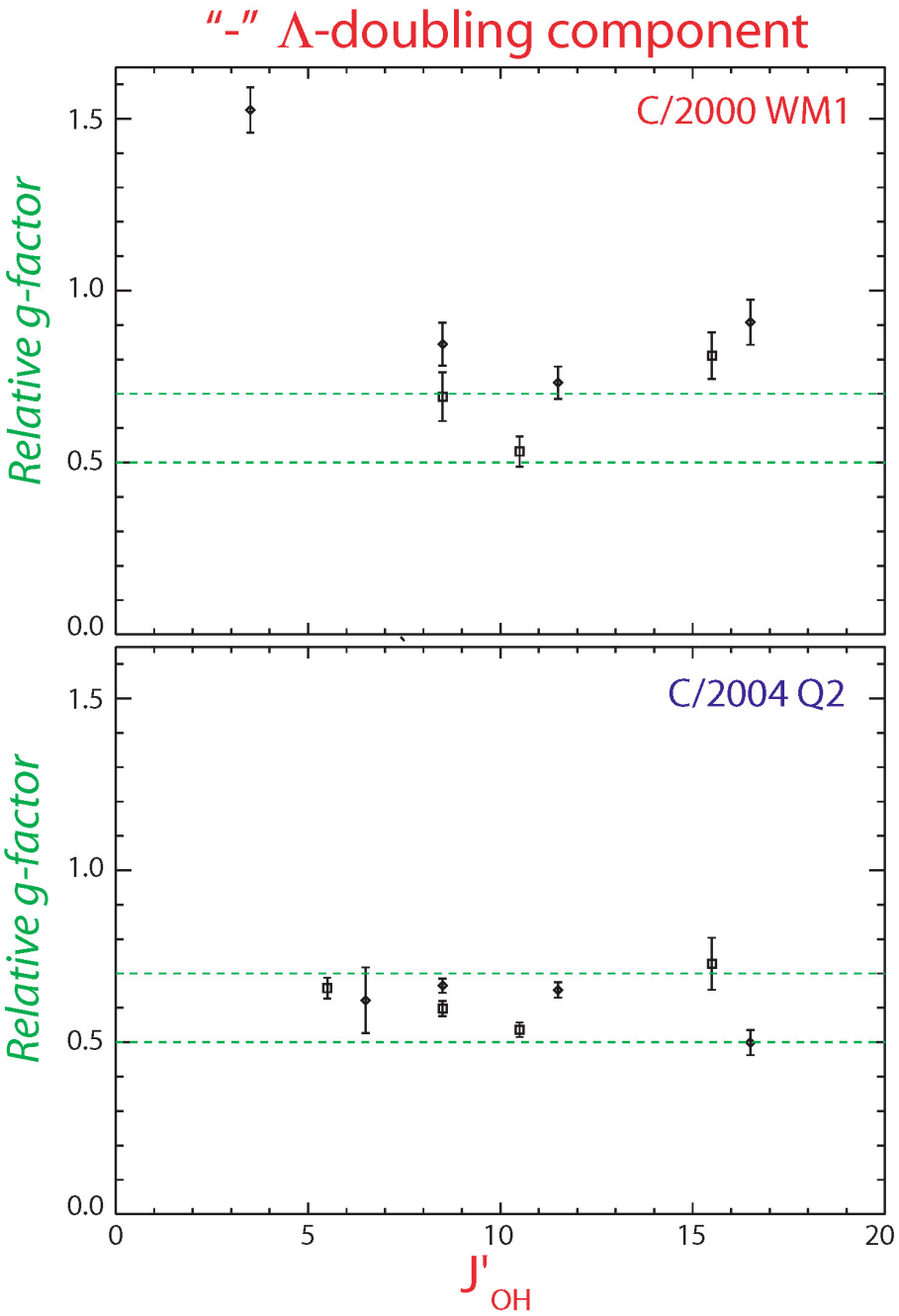

Fig. 38 – OH emission efficiencies of the "–" lines normalized to the g-factors of P12.5 1^+. The dashed green lines do not have any special significance except for guiding.

The ratios between PE efficiencies are very similar for the distributions observed in WM1 and Q2. Most of the relative g-factors agree within the limits of stochastic noise if not within 1σ errors.

Another aspect of the data is revealed by examining the ratios between the g-factors of the "–" and the "+" lines (respectively the "+" and "–" J' states). These ratios are shown in Figures 39 and 40 and are again very similar for the two observed distributions for the majority of lines.

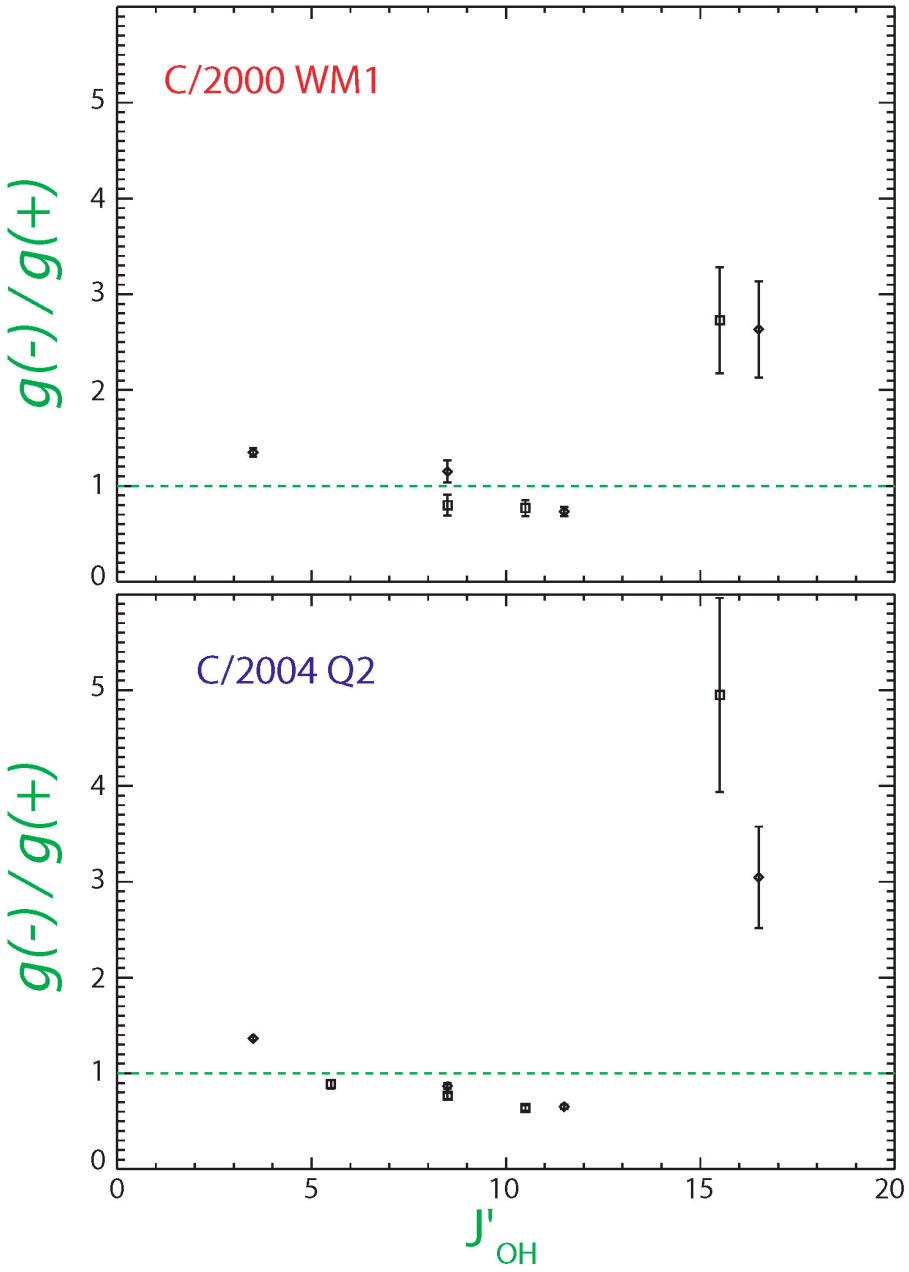

Fig. 39 – Ratio between the g-factors of the "–" and "+" lines from the 1-0 band. The dashed green lines do not have any special significance except for guiding.

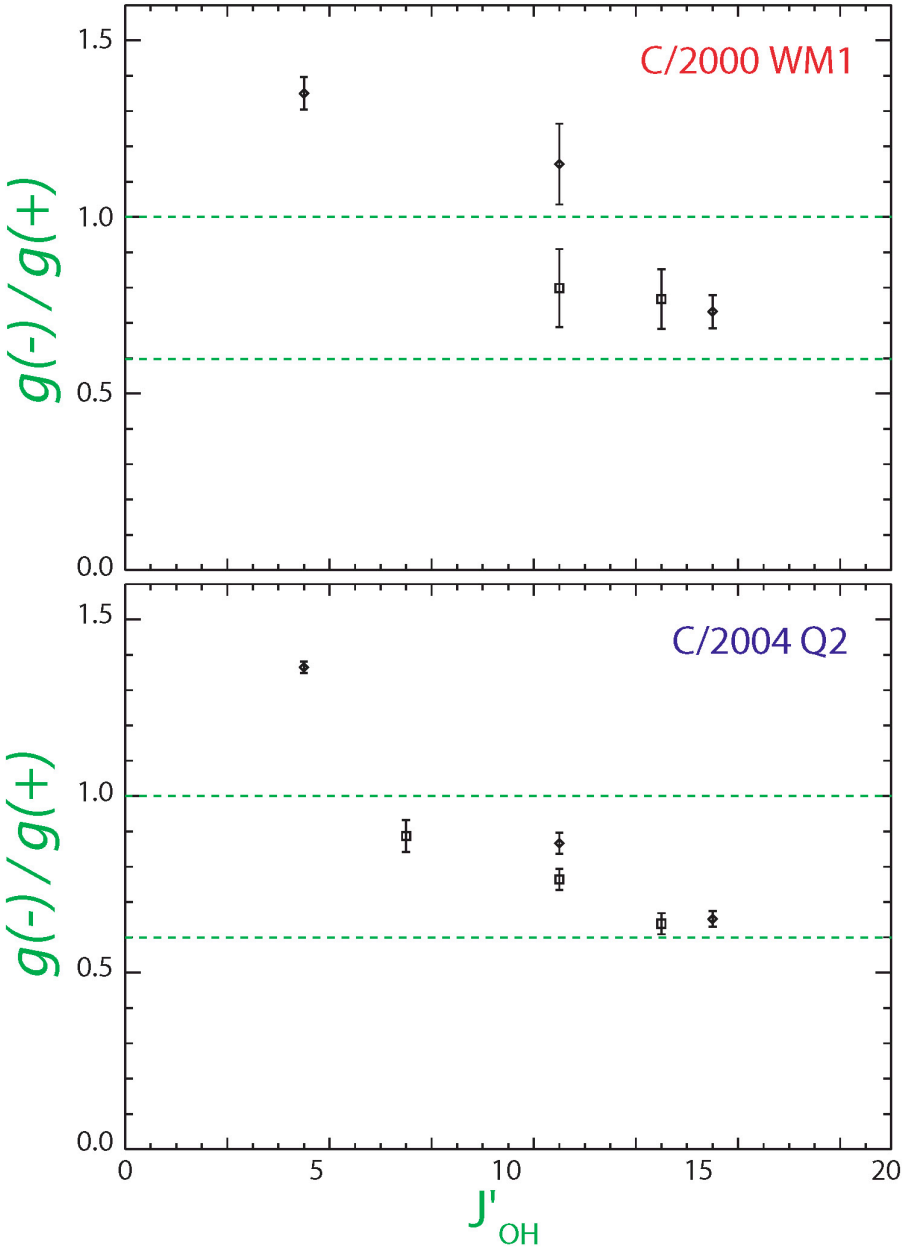

Fig. 40 – Same as Figure 39, but with reduced scale to show the results for J' < 15.5 more clearly. The dashed green lines do not have any special significance except for guiding.

Without over-interpreting results from only two data sets, it is nevertheless important to demonstrate the large extent of agreement in the relative g-factors in WM1 and Q2. Note that this agreements holds for the very few lines sampled in the 2-1 bands, as may be easily inferred from the plots of the absolute g-factors (Figures 29-30).

It is interesting to note that after taking into account all uncertainties discussed in the previous chapter, the "–" lines from the quadruplet in order 21 (N' = 16) are much stronger than the "+" lines. The large error bars in the g-factor rations for P16.5 2^- and 2^+ and P17.5 1^- and 1^+ are because the "+" lines have much lower S/N ratio, while the "–" line intensities are well above the photon noise envelope (Figure 4H; 4I) .

Considering selection rules, the ratio of the "–" to the "+" lines in order 21 implies a strong preferential population of the lower (in energy) Λ-doublet state, which is A' for J' = 15.5 ($^2\Pi_{1/2}$) and J' = 16.5($^2\Pi_{3/2}$). The comet data do not show a preferential population of the upper Λ-doublet state (A") as predicted and observed by Andresen et al. (1984). Their prediction is for the highest-J OH levels for which the Hund's case [b] is a good approximation. Andresen et al. observed up to J'~10.5. We also sample such excitation, but our order 21 lines come from even higher J', thereby presenting a "clean" Hund's case [b]. For such a situation, for a given J' the upper and lower Λ-doublet components A" and A' can be identified by the orientation of the unpaired electron's π-orbital in OH*: these are respectively perpendicular to and in the plane of the molecular rotation.

Strong Λ-doublet inversion was observed at H_2O rotational temperatures of ~30K by Andresen et al. At room temperature, the inversion signature was not observed, with the lower component having higher population, similar to our results. They correspond

to an intermediate T_{rot} for H_2O (70-100 K). So it will be interesting to search for the effect in a comet with low water production rate and likely also low rotational temperature.

Two final remarks: Andresen et al. point out that their treatment applies strictly to the case for a planar dissociation in the FAB. Hausler et al. (1987) mention the possibility that an eventual Λ-doublet inversion distribution may be quenched extremely quickly, but it is not clear how this compares with the (also fast) prompt emission time scale.

6.5 Summary

The most important features of the OH*($^2\Pi$; v'=1) distribution can summarized as follows:

1. The OH PE lines with $7.5 < J' < 17.5$ correspond to excitation temperatures of 4-6 x 10^3 K for both Q2 and WM1.

2. The relative g-factor distributions are very similar for WM1 and Q2, suggesting a common preferential populating of some OH*($^2\Pi$; v'=1) levels.

The purpose of this chapter was not to interpret, but to identify possible systematic effects in the relative PE efficiencies sampled in two comets. It is significant that there are preferential populations of some vibrationally and rotationally-excited states of OH relative to others.

The next important issue is to search for OH PE in a comet with much lower productivity and hence rotational temperature. A good candidate may arise from an

ecliptic comet at sufficiently small geocentric distance to achieve high S/N in a weak source. Chapter 4 had a section on the "Deep Impact" target Tempel-1; that however was not the best candidate because of the non-steady H_2O production combined with its low gas-to-dust ratio. Comet 73P/Schwassman-Wachmann 3, with its close approach to Earth in 2006, is a likely candidate.

Chapter 7

Back to a Broader Perspective

In presenting this observational project, I tried to reflect on the level of detail involved in high-resolution IR spectroscopy of comets: from planning and conducting observing runs, through rigorous data reduction and spectral extraction, to quantitative analysis for a parent volatile and its dissociation product. A good part of the detail included herein has not previously been documented in enough depth to be referenced. Even quality journal articles more often than not do not discuss otherwise critical issues like data reduction caveats and uncertainties of final results. In our work, getting the number is often easier than evaluating the various sources of error that range from photon noise to model uncertainties and spectral confusion. For example, I hope I have managed to convey that reporting spin temperatures (especially) or H_2O production rates incorporating only "stochastic" uncertainties (and therefore without evaluating the consistency between independent line-by-line measurements) could have more serious implications than skipping "minor" details. It can lead to over-interpretation of the particular result (for example, when making cosmogonic interpretations) and therefore is not desirable. I also hope I have managed to express clearly the corner-stones in building a new approach for extracting production rates from spectroscopic data. When initiating a new method, it is necessary to obtain parameters that will allow extraction of

quantitative production rates for the species of interest from its multiple lines. The consistency among these results in an important criterion in establishing the reliability of the method. Another corner-stone is to define a clear strategy for further validation of the method beyond its initial development, as was done for OH prompt emission near the end of Chapter 5.

So, what are the main scientific contributions of this thesis? I would emphasize five main original contributions as follows:

1. *Development of a second method to measure H_2O production at infrared wavelengths, including evaluation of its potential limitations and definition of a future program to further validate it.* While based on a well-known and long-existent idea, the approach of using IR OH prompt emission as a proxy for water production had not been backed up by a detailed quantitative analysis of multiple lines like the one presented here. This has required large simultaneous spectral coverage, as is afforded by NIRSPEC.

2. *Observations of the end-product distribution of H_2O photolysis under natural conditions that differ greatly from studies done in terrestrial laboratories.*

3. *Measurement of H_2O ortho-to-para ratios and spin temperatures in two comets with realistic treatment of uncertainties.* We are now extending the standard measure towards extraction of *spatially-resolved OPRs*, with the clear objective of increasing the level of confidence of these important parameters.

4. *Derivation of reliable water production rates from non-resonance fluorescent H_2O emission sampled directly in two comets.* These production rates provide a solid baseline for describing the volatile chemistry of C/2000 WM1 and C/2004 Q2. Without this baseline, the IR measurements of trace constituents are much more difficult to

interpret.

5. *Reliable measurements of the H_2O integrated column abundance in comet 9P/Tempel-1, during the observing campaign supporting the NASA's Deep Impact mission.*

The previous three chapters focused on scientific methodology and results, but covered to a very limited extent other kinds of contributions of this thesis – especially those related to the ever-evolving development of reduction and analysis algorithms for high-resolution comet spectra. These contributions are numerous and involve each stage of the work. The most important ones are as follows:

1. *Development of an automated algorithm for quantitative analysis of emission from parent volatiles including testing fluorescence models, obtaining rotational temperature, ortho-to-para ratio retrievals, and obtaining line-by-line production rates.* This work was done jointly with Michael DiSanti and has led to a very significant reduction in the analysis time required, plus increased confidence in the results. This effort permitted very rapid (but reliable) analysis of data gathered through our observations supporting Deep Impact, as pointed out near the end of Chapter 4. Another direct application is efficient analysis of the H_2O "hot-bands" in comet C/2004 Q2 (Machholz) observed less than a year ago.

2. *Identifying and correcting critical errors in the existing software.* The most important error corrected was related to systematic inaccuracy in intensity measurements from 2D spectral frames. This inaccuracy depended on the position in a given order and in some cases reached ±15%.

3. *Testing and understanding the limitations of new software*. The most important was the algorithm for wavelength calibration and order straightening described in Appendix 2.

4. *Investigating a serious source of instrumental noise in the NIRSPEC detector, known as "row chop" pattern (jointly with Mike Mumma)*. This effort was required for the WM1 data, in which every eighth row in a "raw" frame exhibited both deviant signal and uncommonly high noise that were not photon-related. (This is only one example of the complex issues which are not addressed by the general software, and which one has to face in a particular data set.)

5. *Conducting a major correction in the algorithm for flat-fielding of NIRSPEC data*. Prior to making this correction not all of the desired part of the cropped spectral order had been properly flat-fielded.

I will conclude by returning to the starting point of this presentation. The *fascination* of comets is subjective and I certainly do not intend the reader to share that perception unless willing. On the other hand, I attempted to convey the *scientific significance* of classifying comets according to the chemistry of their volatile fraction and I tried to present the case for building a "meter stick" (the abundance of H_2O released from the nucleus) in the context of achieving a chemical taxonomy for these objects.

Among the five major contributions listed above, four relate directly to the initial motivation of this thesis – to contribute to the continuous process of establishing the new chemical taxonomy of comets. The result with greatest immediate impact was the accurate measurement of the H_2O column abundance from observations supporting the

"Deep Impact" mission (Chapter 4). This allowed reliable mixing ratios of trace constituents (C_2H_6, CH_3OH, HCN) to be quantified with respect to H_2O, both before and after impact (Mumma et al 2005). 9P/Tempel-1 (the "Deep Impact" mission target) is a Jupiter family comet most likely injected from the scattered Kuiper disk (see Chapter 1). We showed that the chemical composition in the ejecta after impact (expressed in terms of column abundances relative to H_2O) was overall similar to the composition of Oort cloud comets in our data base. This result suggests that although belonging to very different dynamical reservoirs in the present-day solar system, comets from the present scattered disk and from today's Oort cloud could have formed in a similar region in the proro-planetary disk.

The results supporting Deep Impact also suggest that the nucleus of this comet might be heterogeneous, containing cometesimals formed in different regions of the proto-planetary disk. Presenting unambiguous evidence test of compositional heterogeneity within comet nuclei is a significant objective of future observations, because it will test the radial migration of cometesimals in the young solar system, as predicted by dynamical models.

Three main thesis contributions are "long-term investments". Our study of OH vibrational prompt emission (OH*; Chapter 5) shows that this emission can be used as an IR proxy of H_2O production in comets, even when the water production rate cannot be measured directly. For example, as a result of this work, the OH* lines between 3036 and 3048 cm^{-1} are now considered a more reliable tool for determining the relative abundances of H_2O (via OH*) and the co-measured hyper-volatile species CH_4. This mixing ratio varies greatly within a sample of Oort cloud comets (Gibb et al. 2003) and it

is therefore a very important chemical indicator in building a taxonomy. In is interesting to mention, the abundance of another hyper-volatile, CO, also exhibits a great variation with respect to H_2O (cf. DiSanti 1999; 2001; Mumma et al. 2003), but a correlation between the abundances of CH_4 and CO is not evident in our current data set.

Our results for OH prompt emission were used to illustrate the apparent significant change in the methane-to-water mixing ratio for comet C/2001 A2 when comparing observations taken in July and August 2001 (Gibb et al., in preparation). The nucleus of this comet split in two between these dates, thereby exposing fresh (presumably) unprocessed material and creating new sublimation vent(s). The difference in methane-to-water mixing ratio before and after the splitting event supports the notion that the comet nucleus might be internally heterogeneous.

Another important part of the chemical taxonomy is to quantify the native abundances of oxidized species (CO, H_2CO, CH_3OH) that might be linked by hydrogenation on pre-cometary grains. For this purpose, the native source contributions of these molecules must be isolated from their potential extended components (produced by sources in the coma). The best data on H_2CO to date is from a single order instrument that could not simultaneously sample H_2CO and H_2O (or any other strong emission from a parent volatile). But H_2CO was sampled simultaneously with OH prompt emission, whose spatial distribution is characteristic of a native species (over the sampled region in the coma, that is optically thin for the solar UV radiation responsible for H_2O photolysis). Considering the results of this thesis, the search for an extended source of formaldehyde benefits greatly from the fact that it was co-measured with emission from vibrationally and highly-rotationally excited OH.

Our measurements of accurate H_2O spin temperatures in C/2000 WM1 and C/2004 Q2 add to the limited data base of such measurements in comets (Chapter 4). These results are important because they show that the spin temperature of H_2O can be measured both routinely and reliably from the ground, provided high-quality spectra are available. Under this condition, we can consider the gradual growth of a data base of water spin temperatures deduced from measurements of many comets. If the current argument for very long conversion times between ortho-water and para-water is confirmed by future theoretical and experimental work (see discussion in Chapter 2), the distribution of spin temperatures in a large sample of comets could place important constraints for the range of temperatures at which cometary material formed. We emphasize, however, that at the current level of understanding, it would be very premature to relate a given spin temperature unambiguously to a unique time and region of formation of the corresponding comet. Moreover, the possibility of spin conversion in the solid phase (after forming the nucleus) should be investigated. However, this is a very difficult problem experimentally. Observationally, measuring T_{spin} in a given comet over a range of heliocentric distances and positions in the coma, can potentially test these possibilities. This will require a comet with high gas production that is also available for observation with NIRSPEC over an extended period.

Finally, this thesis provides accurate and directly-determined H_2O production rates in C/2000 WM1 and C/2004 Q2, that constitute reliable baselines from which to characterize the overall chemistry (thought mixing ratios) of these two comets (part of this work will be done by the thesis's author). There remain many open questions: Do the mixing ratios of CH_4 fall within the already established large range in eight Oort

cloud comets (Gibb et al. 2003) or they will expand this range even more? Are the C_2H_6 abundances in the two comets consistent with the narrow range observed by us in several Oort cloud comets (cf. Mumma et al. 2003) or will they be more in line with that measured in the "unusual" comet C/1999 S4, likely formed near Jupiter (Mumma et al. 2001b)? Do low C_2H_6 and CO mixing ratios (with respect to water) indicate formation at smaller distances (5-10 AU) from the young Sun, where thermo-chemistry governs the abundances? What is the mixing ratio of native H_2CO and how does it compare to those of chemically related oxidized species (CO, CH_3OH)? Addressing these (and other) issues regarding the overall chemistry depends on the accuracy of H_2O production rates retrieved in this thesis.

To summarize, the present project relates to all principal building blocks of the relatively new chemical classification of comets: mixing ratios $Q(X)/Q(H_2O)$, D/H ratios (quantifying H_2O is half of the problem and the methodology presented applies also to HDO), and spin temperatures of H_2O (the latter are not certain, but are a possible cosmogonic indicator). We would like to understand the cosmogonic implications of the values derived for these observables, and this requires continuing to build on our current data base of comets.

In one sentence, the present project is an IR study of H_2O production and dissociation in comets. It should be pointed out that water is not the "World Cup" of cometary science, but instead is the volatile to which other molecular abundances are compared.

Finally, as in my previous endeavors, this study deepened a long existing interest and also triggered a new one. Before joining the Goddard comet group my questions

revolved around how the observed chemical abundances of cometary volatiles relate to the dynamical classes of these objects. I can state this in a more specific form now:

How did chemical abundances, isotopic ratios, and spin temperatures vary with distance of formation of the constituent cometesimals from the young Sun?

What limits our ability to interpret the values measured for these cometary "observables" , in terms of conditions in the early solar system?

Which species provide the most useful cosmogonic indicators?

Important enough, these questions have been followed by new ones (for me anyway) that opened another realm of interest while working on the cometary H_2O problem. In particular:

What is the level of synergy between studies of cometary composition and studies of molecules observed in proto-stellar environments?

What chemical indicators of potential planet formation can be identified in such environments?

What are their abundance ranges in <u>comets</u>?...

Chapter 8

Extra Section: Effects of Atmospheric Dust on the Recession on the Seasonal and Residual South Polar Cap on Mars - overview

The last chapter of this thesis is an overview of a modeling project, which is not formally connected to comets. Its foundation had been completed before I joined the Goddard group in late 2002. In the summer of 2001 I worked my way through understanding (well beyond the level of controlling input and output) and subsequently modifying a computer code for modeling radiative transfer through atmospheric dust, initially developed by Bjorkman & Wood (2001) for studying circum-stellar disks. The object investigated was the planet Mars, another inspiration for a long-term interest, dating back from one of the "Great oppositions" in the mid-1980s. The studied phenomenon was the spring-time sublimation of the seasonal and residual south polar cap (SPC) of that planet. The problem was to examine *the possible connection (or lack of) between dust storm activity and the rates of recession of the seasonal south polar cap*, as well as *the stability of the residual cap*.

A very important output of this work has been consolidating efforts to constrain

more tightly the frequency-dependent albedos of the SPC ices, based on analysis of multiple spectral regions (visible[VIS], NIR, thermal IR). The *surface albedo spectrum* is the critical parameter in the modeling I am about to describe. Constraining this parameter has been a necessary condition to stretch our modeling results from the level of a "good numerical exercise" to being more definitive.

The driving theme in this chapter will be the "interaction" between modeling the polar cap sublimation rates under dust storm conditions and the studies of the polar surface albedo sprectra, based on both space-craft and ground-based observations in different wavelength regimes. Since 2001 this interaction defined an "upward spiral" of understanding: our initial modeling the sublimation in the polar regions used the best available surface albedo spectra at the time, but also motivated additional studies to better constrain this parameter. This lead to observational papers of polar albedos that in turn enabled more definitive modeling results on the stability of the residual SPC. Table 6 summarizes the overall effort. The subsequent sections describe it in more detail.

TABLE 6

Inter-related papers on modeling the south polar cap sublimation rates under dust storm conditions and constraining surface albedo spectra from observations in multiple spectral regions

Constraining polar surface albedo-spectra	Contribution	Modeling SPC sublimation under dust storm condition	Contribution
Hansen 1997	Accurate determination of optical properties of CO_2 ice		
Hansen 1999	Surface albedo spectra of the SPC based on thermal IR space-craft data (MGS-TES)		
		Bonev et al. 2002	Reporting observations and modeling the spring-time sublimation of the Mountains of Mitchel in two years with very different dust storm histories
		Bonev et al. 2003	Modeling the seasonal SPC sublimation rates under dust storm conditions
James et al. 2005b	Visible albedos of the residual SPC from HST photometric data		
Glenar, Hansen et al. 2005	NIR albedos from ground-based mid-resolution spectra (SPEX at NASA/IRTF)		
Hansen et al. 2005	NIR albedos from space-craft mid-resolution spectra (PFS on MEX)		
Hansen (2005 – in progress); Glenar (2005 – in progress)	NIR albedos from space-craft mid-resolution (PFS on MEX) and from ground-based (SPEX at NASA IRTF) spectra	Bonev (this chapter)	Modeling the stability of the residual SPC against a global dust storm occurring near perihelion

8.1 Effects of Atmospheric Dust on the CO_2 Spring-Time Sublimation Rates on Mars

About 95% of the Martian atmosphere consists of CO_2. Between a quarter and a third from this amount is cycled through the polar caps, condensing out in the fall and winter and sublimating in the spring and summer. This atmosphere-surface interaction defines the *planet's CO_2 cycle*. The polar caps are the sources and sinks of the exchanged CO_2 and therefore play a key role in this cycle. The latent heat released during condensation is a major source for the energy radiated to space by the cap; the solid carbon dioxide sublimes in spring, absorbing the latent heat from the insolation as it transits to a gas phase. Inter-annual variability in the amount of exchanged CO_2 between surface and atmosphere is reflected by apparent variability in the caps' appearances, which have been studied since Hershel by ground-based and in later years by space-craft observations (cf. James, Kieffer, & Paige 1993; James, Hansen, & Titus 2005a).

Our study focuses on south seasonal and residual polar caps. It address two important issues related to the spring-time sublimation of polar ice:

1. What are the causes of the apparent insensitivity of the mean recession rate of the seasonal SPC to atmospheric dust?

2. Can radiative effects from atmospheric dust "break" the stability of the residual SPC after a global dust storm occurring near perihelion orbital position of Mars?

The *seasonal* SPC consists of the carbon dioxide exchanged between surface and atmosphere. Its *recession curve* (or *regression curve*) is defined from the apparent movement of the mean cap radius towards higher latitudes during spring as the solid phase CO_2 sublimes. The *residual* (also known as perennial) SPC is the feature in which

solid CO_2 remains during the summer, even after the complete recession of the seasonal component. The residual SPC is covered mostly by carbon dioxide ice. There is a recent evidence that this layer stays on top of a water ice, which is exposed on the surface in isolated spots.

The reason we are interested in the inter-annual variability of polar caps is their key role in the CO_2 cycle on the planet, which is global phenomenon, having implication on the whole atmosphere. But why we are considering effects of atmospheric dust on SPC stability and respectively on the CO_2 cycle?

The Martian atmosphere contains dust at some background level even at "quiet" conditions. Increased atmospheric dust levels are expected during local, regional, or global dust storms. The latter are planet encircling phenomena (cf. Smith et al. 2002) and have been observed to cover a significant fraction or the whole SPC. Such are the global dust storms in 1971 observed by Mariner-9, in 1977, observed by the Viking orbiters, and in 2001 – observed by the Mars Global Surveyor (MGS). Note that while dust storms on various spatial scales occur throughout the Martian year, the probability for global dust storms is apparently increased during the southern spring and summer season, when Mars is close to perihelion and its summer hemisphere is a subject of an increased insolation because of the orbital eccentricity effect.

Atmospheric dust affects the energy input to the polar cap surface in two ways: dust absorption in the visible reduces the insolation reaching the surface but also increases the energy load in the thermal infrared. Therefore the main effect of atmospheric dust is the visible-to-infrared frequency redistribution of the radiation reaching the polar surface. The surface "response" to this redistribution depends on its

relative absorptive properties between VIS and thermal IR. If the overall surface absorption is dominated by visible photons (i.e. low VIS albedo), the reduction of the direct solar flux caused by absorption from atmospheric dust would lead to a reduced sublimation rates of CO_2. Conversely, if the surface absorbs better in the IR, the atmospheric dust effect would be an accelerated sublimation. Of course a balance between absorptive properties in the IR and in the VIS is also a "scenario". This is the point when knowledge of the surface albedo spectrum becomes critical.

Investigating the aforementioned radiative effects requires modeling the radiative transfer through a dusty planetary atmosphere bounded by a sublimating CO_2 surface. Before describing our modeling approach, it is appropriate to review the key observational evidence for the type of "response" in the south polar recession to global dust storms during the spring and summer seasons.

8.2 Atmospheric Dust and the Seasonal South Polar Cap Recession

Observations of the boundaries of the seasonal CO_2 caps and the pressure curves measured by the Viking landers have been used to search for inter-annual variability in the carbon dioxide cycle. An important aspect of these studies is the lack of evidence for a strong coupling between the global dust storms and the mean recession rate of the seasonal SPC. Ground-based historical data suggests some connection, but disentangling real differences from systematic errors is problematic (James et al. 1987). On the other hand, the Viking pressure curves for years with vastly different dust storm histories have showed only subtle differences (James et al. 1993).

The 1999, 2001, and 2003 spring regressions of the south seasonal polar cap have

been observed in unprecedented detail by MGS in both visible (Mars Orbital Camera, MOC) and IR (Thermal Emission Spectromer, TES) wavelengths. This has provided an excellent opportunity for a precise comparison between the cap regression rates for years with very different atmospheric conditions in early spring. While the early phase of the 2001 regression occurred during a global dust storm (Smith et al. 2002), the 1999 and 2003 seasons were significantly less dusty. It should be emphasized that the MGS observations allow studies not only of the average cap decay but also of localized regions with distinctive albedo properties, such as the Cryptic region (Titus & Kieffer 2002) (low visual albedo) and the Mountains of Mitchel (Titus & Kieffer 2002; Bonev et al. 2002; Benson & James 2005) (high visual albedo). Another advantage of MGS was that the SPC recession in a "dusty" and following "non-dusty" year, could be observed with the same set of space-craft instrumentation, which contributed to the reliability of the direct comparison between the data from 1999, 2001, and 2003.

The main observation results concerning the seasonal cap's response to increased atmospheric dust can be summarized as follows:

1. The average recession curves for 1999, 2001, and 2003 are very similar despite the very different dust storm histories. The seasonal SPC is apparently insensitive to an increased atmospheric dust (Benson & James 2005).

2. Visible bright regions, like the Mountains of Mitchel, showed faster regression rates (by L_s~4°-5°), during the year with a global dust storm (Titus & Kieffer 2002; Bonev et al. 2002; Benson & James 2005). Note that L_s is the areocentric longitude of the Sun in Mars's ecliptic; L_s= 180° for the beginning of southern spring.

3. Low visual albedo regions displayed slower regression (by Ls~4°-5°) in 2001

compared to 1999 (Titus and Kieffer 2002).

Therefore while the average cap recession was not significantly influenced by the global dust storm in 2001, effects were observed on a local scale and were apparently correlated with the visible surface frost albedo. By late 2002 we were able to incorporate the three major types of sublimating surface response to increased atmospheric dust within a single model of radiative transfer through atmospheric dust above a sublimating CO_2 surface. The major parameter was the surface albedo spectrum.

8.3 Atmospheric Dust and the Stability of the Residual South Polar Cap

Is the CO_2 residual cap at the south pole a transitory or a permanent fixture of the current Martian climate? The answer is uncertain. Ground-based observation have been ambiguous. Clearly observable changes less dramatic than complete disappearance of the residual SPC have been documented. The images at similar (~ 200 meter) resolution and near identical L_S acquired by Mars Global Surveyor (MGS) in 2000, by Viking in 1977, and by Mariner 9 in 1972 shown in Figure 41 provide evidence for significant interannual change. On the other hand, there has been little change in the perennial cap at such resolution during the MGS missions.

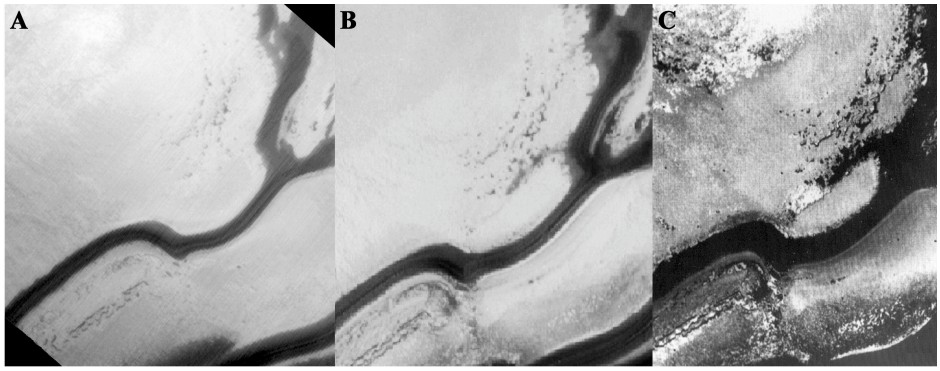

Fig. 41 – Images of the same surface features from the perennial south polar cap taken at near identical L_S and spatial resolution of about 200 meter; (A) Mars Global Surveyor, 2000; (B) Viking, 1977, and (C) Mariner 9, 1972 (the images are credit to NASA).

The integrated visible albedos of the cap inferred from well-calibrated Hubble Space Telescope (HST) photometric data (James et al. 2005b) were high-enough to "provide" a stability of the perennial cap during years with no significant increase in atmospheric dust. This paper ended with a question: whether or not a global dust storm near perihelion, where the amount of solar insolation reaching the pole are largest, can "break" (at least the residual cap's stability). Figure 41 shows that the most significant change in the residual cap's appearance has occurred in 1972 after a planet-encircling dust storm in 1971 which covered the whole cap. On the contrary, observations of the 2001 global storm suggested that not much dust penetrated to the residual cap. So it may be that the residual cap was protected from the dust by circulation resulting from seasonal cap sublimation and due to its high altitude relative to the seasonal cap. In addition, the 2001 dust storm occurred much earlier in the season than the 1971 event. This means relatively smaller importance of radiative effects, considering the lower amount of insolation reaching the pole.

In 2005 a joint effort between several collaborators (coordinated by the thesis's

author) has lead to well-constrained albedo spectra representative for the perennial cap. This in turn allowed to produce more definitive results about the influence of increased atmospheric dust on its stability using the radiative transfer and surface models to be described in the next two sections.

8.4 Modeling the Radiative Transfer Through Atmospheric Dust

We have examined the effects of atmospheric dust on the sublimation rate of a CO_2 surface. For this purpose, we employed the non-iterative Monte Carlo radiative equilibrium code of Bjorkman and Wood (2001), adapted for the case of a plane-parallel dusty planetary atmosphere. Note that although the model atmosphere is one-dimensional, the radiation transfer code is three-dimensional and includes wavelength-dependent dust opacity, anisotropic scattering, and thermal dust emission. The dust temperature is calculated using the condition of radiative equilibrium. Consequently, an input atmospheric temperature profile is not required, even though a version of the code allows this option.

The atmospheric dust particles are assumed to be in local thermodynamic equilibrium. The wavelength-dependent dust absorption and single scattering properties (albedo, absorptive opacity) have been adopted from the study of Wolff and Clancy (2003). These data were constrained using multiple spectral regions, and are result of combining previous work in the visible and IR with subsequent analysis of TES spectra over a wide range of aerosol loading conditions.

For dust scattering, we adopted the phase function of Tomasko et al. (1999) at 965 nm for the visible and a Henyey-Greenstein phase function (with a frequency-dependent

value of the asymmetry parameter) for the IR. Note that at IR wavelengths the absorption starts to dominate over the scattering because the dust albedo drops significantly.

What are the main advantages and limitations of the Monte Carlo approach of Bjorkmam & Wood? Each Monte Carlo algorithm is associated with the generation of a random number from a uniform distribution between zero and unity. The heart of the Monte Carlo approach is the way this uniform distribution is converted to the desired physical distribution.

Before describing this conversion, it is appropriate to point out what are the desired physical distributions – probability for absorption/scattering by atmospheric dust, supplied by the albedo of single scattering; probability for scattering in a certain direction, supplied by the scattering phase function; probability for thermal dust emission at a certain frequency, supplied normally by the product of the dust opacity (κ_ν) and the Plank function ($B_\nu[T]$) for the corresponding dust temperature:

$$\frac{dP}{d\nu} \propto \kappa_\nu B_\nu(T)$$

This approach would require iteration over the atmospheric temperature distribution. However, the approach of Bjorkman & Wood is non-iterative. Their model atmosphere has an initial temperature of zero – an obscure idea at first glance! When I started to describe it in a seminar talk at GSFC two years ago, the question asked immediately was "*Does it work?*". It does for the reason described below:

The radiation field is divided into monochromatic photon packets. Each packet consists of a number of monochromatic photons. All packets have the same total energy, and therefore a different number of physical photons, depending on frequency. When a photon packet is absorbed, the temperature of the absorbing cell in the model grid is

raised corresponding to the packet's energy. Radiative equilibrium requires that an equivalent amount of energy should be reemitted. The frequency of the reemitted photon packet is determined by the probability distribution:

$$\frac{dP}{d\nu} \propto \kappa_\nu \frac{dB_\nu(T)}{dT}.$$

This probability distribution implies that the frequency of each reemitted packet corresponds not to the current cell temperature T, but to the difference spectrum:

$$\kappa_\nu \cdot \left(B_\nu[T] - B_\nu[T - \Delta T]\right),$$

where ΔT is the small increment in temperature as a result of photon packet absorption. After being reemitted the packet continues to be scattered, absorbed, or reemitted again until it escapes from the atmosphere. As a sufficient number of photons packets (~10^5-10^6) is propagated, the initial "zero-temperature" model atmosphere is heated and the final emergent spectral energy distribution relaxes to the correct one, corresponding to the final temperature (T_{final}):

$$\frac{dP}{d\nu}(final) \propto \kappa_\nu B_\nu(T_{final})$$

This approach has been validated by direct comparisons of various modeled spectral energy distributions to the ones expected for the final temperature in a given model grid cell.

The development of this approach is not a part of this thesis, but it was important to outline its main idea, because it leads to a numerical algorithm, which is not-iterative over the atmospheric temperature. As a consequence the algorithm is speed efficient – an important factor in this project, considering the necessity for a large number of model runs to study the parameter space.

In our simulations the photon packets are propagated to a plane-parallel dusty planetary atmosphere bounded by a sublimating CO_2 surface. Some of the packets eventually reach the surface and are reflected or absorbed by it, depending on the surface albedo for the photon's wavelength. In this scheme, the CO_2 sublimation flux is given by the difference between the frequency integrated fluxes absorbed and emitted by the surface. Note, that the spectrally integrated surface emission is constant (independent of optical depth) with a temperature equal to the sublimation temperature of 147 K.

The main limitation of Bjorkman & Wood's approach is that it is designed for temperature-independent opacity source, like atmospheric dust. The same approach cannot be applied in a straightforward manner to a temperature-dependent opacity source. Therefore our modeling does not include the CO_2 15 μm band opacity. Our purpose, however, was not to build a complete atmospheric model and to reproduce precisely its temperature profile, but to simulate the VIS-to-IR redistribution of radiation incident of the surface, caused by absorption and re-emission by atmospheric dust. The non-precise atmospheric temperature profile in our models and the ignored details in the spectrum near 15 μm are second order effects when fluxes integrated over the whole spectral range are modeled. Our modeling results depend mainly on the total dust optical depth in the atmosphere and on the surface albedo spectrum.

We have planned to conduct several model runs using another algorithm which is slower but incorporates explicitly the CO_2 effects in the IR spectrum and on the atmospheric temperature. This would estimate the importance of the ignored effects, but is outside the scope of this chapter.

8.5 The Basis of the Monte Carlo Approach

It was pointed out that (in my understanding) the heart of the Monte Carlo approach is the way an uniform distribution of random numbers is converted to the desired physical distributions, describing the major processes involved (scattering in a certain direction, probability for absorption, thermal emission at a certain frequency). There are various ways to accomplish this conversion. The most straightforward case in for the single scattering albedo, which is essentially a probability for scattering: if the randomly generated number is smaller than the albedo, the photon is scattered, otherwise it is absorbed.

For the other distributions we have adopted the "cumulative probability approach", which is best described by a specific example:

Consider the probability for a certain scattering angle α, given by the phase function of atmospheric dust. We define a cumulative probability as:

$$P_{cum}(\cos[\alpha]) = \int_{-1}^{\alpha} g(\cos[\alpha]) d\cos[\alpha].$$

$P_{cum}(\cos[\alpha])$ is calculated for evenly spaced values of its argument and g(cos[a]) is the phase function. This calculation is displayed graphically on Figure 42 for the phase function of Tomasko et al. (1999). A value of P_{cum} is chosen randomly on the y-axis. A corresponding $\cos(\alpha)$ is determined from the P_{cum} vs. $\cos(\alpha)$ relation. This way each value of P_{cum} in the equal intervals $P_1 - P_2$ and $P_3 - P_4$ (shown on Figure 42) would be picked up with equal chance by the random number generator. This is the uniform distribution. On the other hand a scattering angle in the interval $C_1 - C_2$ would be picked with lower probability than in the much smaller interval $C_3 - C_4$. From the definition of

P_{cum}, it follows directly that the scattering in the angles in the $C_3 - C_4$ interval is much more probable than in the $C_1 - C_2$ interval. Therefore the relation between P_{cum} and $\cos(\alpha)$ is a way to convert an initial uniform distribution to the desired physical non-uniform distribution.

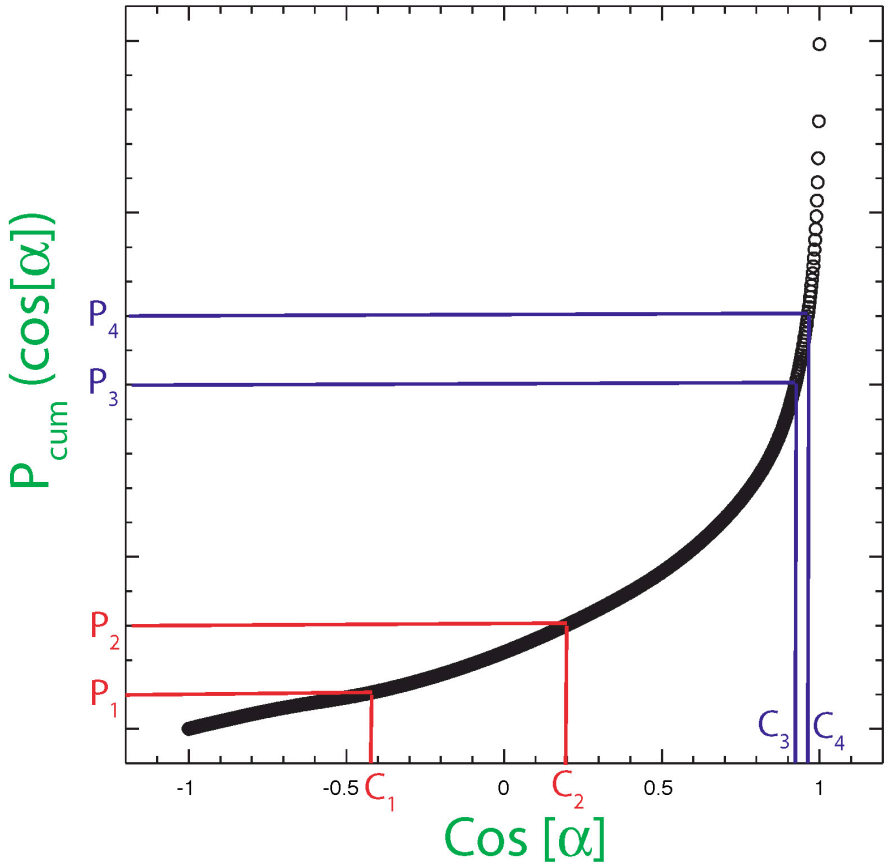

Fig. 42 – Cumulative probability for scattering angle from the phase function of Tomasko et al. (1999).

Figure 43 shows a test of the procedure. The solid line is the phase function that needs to be incorporated into the Monte Carlo algorithm. The individual circles represent the calculated phase function by this algorithm. The circles represent a histogram plot as they are averages over equal bin in $\cos(\alpha)$. For clarity every 100^{th} bin is plotted for $\cos(\alpha) < 0.99$ and every 10^{th} bin is plotted for $\cos(\alpha) > 0.99$. The strongly forward scattering phase function is accurately reproduced using the cumulative probability

approach. Its practical advantage is that it is relatively computationally inexpensive, because the $P_{cum}(\cos[\alpha])$ relation is pre-tabulated before the radiative transfer simulation and intermediate values of $\cos(\alpha)$ are found by linear interpolation.

Similar tests have been performed for other distributions. For example Figure 44 shows a direct comparison between pre-calculated (solid curve) and Monte Carlo simulated (circles) spectral energy distribution for thermal emission of a surface with T_{subl} = 137 K. This comparison also demonstrates the good agreement between desired and simulated distribution.

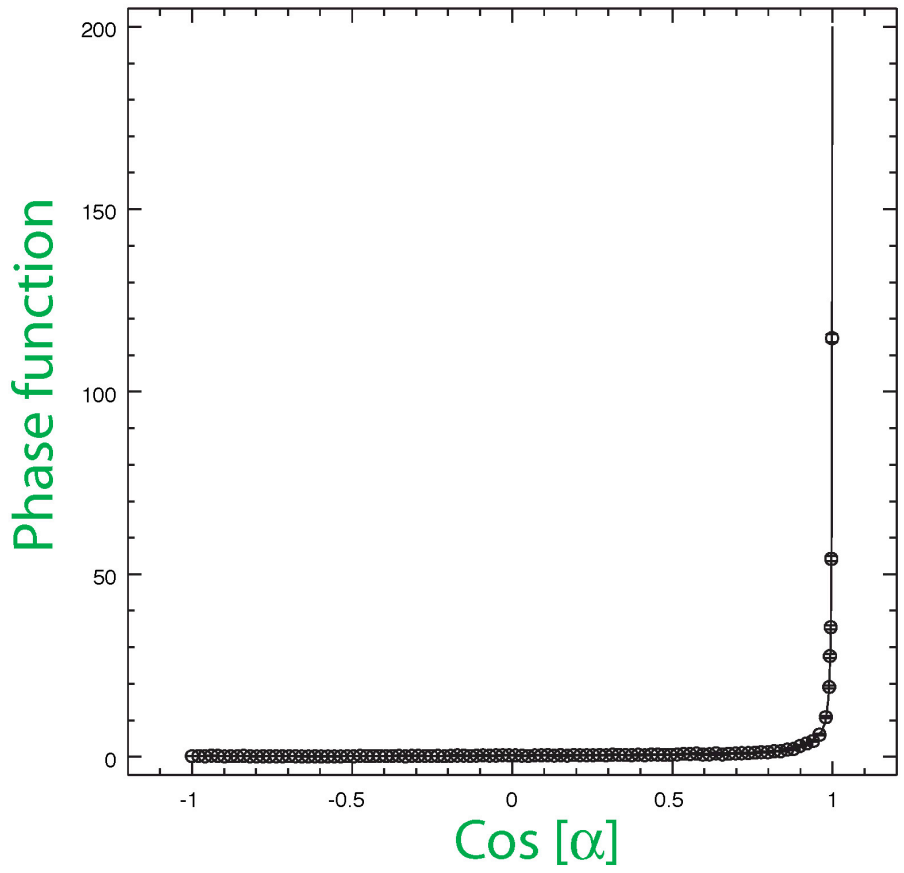

Fig. 43 – Test for implementing the phase function of Tomasko et al. (1999) into the Monte Carlo procedure for scattering by atmospheric dust.

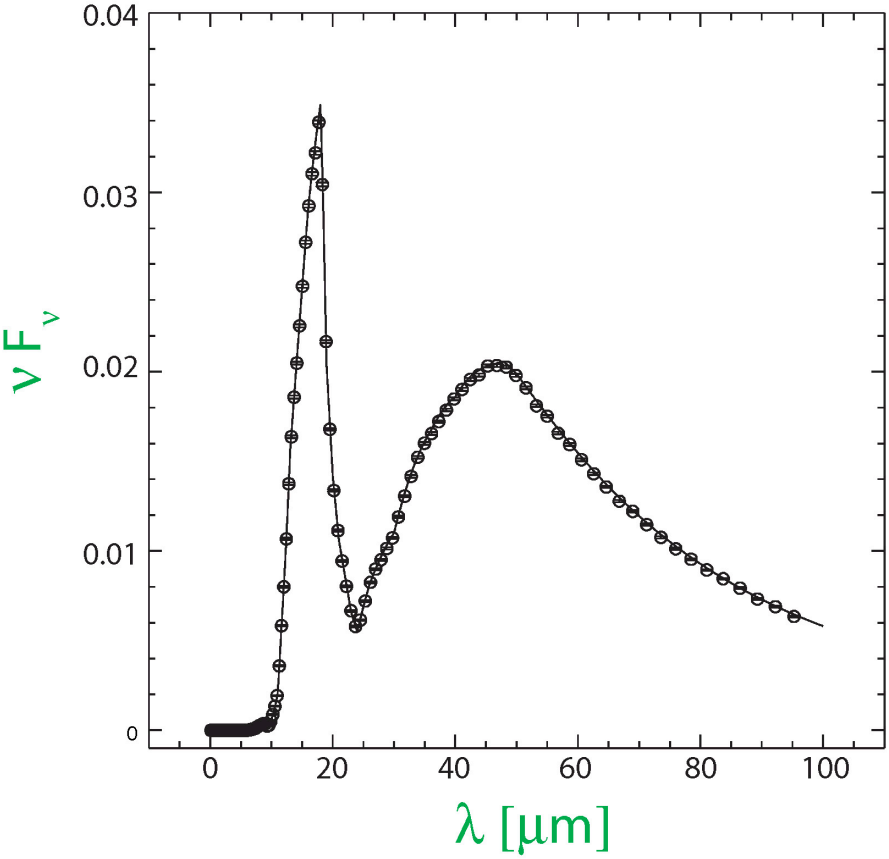

Fig. 44 – Test for the subroutine calculating surface thermal emission within the Monte Carlo algorithm described in the text. The surface flux is normalized to the flux incident to the top of the atmosphere. The surface model is for pure CO_2 ice with very small "grain size".

The examples on Figures 43 and 44 are shown since, they are result of modifying the original code of Bjorkman & Wood as part of this thesis. Figure 43 presents the test for the subroutine treating atmospheric scattering. Figure 44 shows the test of the subroutines related to the surface energy balance. The spectrum on this figure corresponds to a pure CO_2 ice with 100 μm "grain size". The meaning of this term will be clarified in the next section.

8.6 Modeling Surface Albedo Spectra. "Porosity" Versus "Grain Size".

Our dusty model atmosphere is bounded by a sublimating CO_2 surface at 147 K. A very important, but somewhat uncertain, model parameter is the strongly wavelength dependent surface albedo. Constraining this parameter is critical, because it determines how the surface would "respond" to the VIS-to-IR redistribution by atmospheric dust of the flux incident to it.

There are several approaches to model the surface frost albedo. These approaches are similar in terms the radiative transfer physics, but differ in terms of the assumed microphysics of the ice structure. We have adopted the simplest, but the best understood and well-tested approach of Hansen et al. (1999) who approximates the ice medium with a grid of spheres and models the radiated transfer of photons penetrating into this medium. The main advantage of Hansen's approach is that it has successfully reproduced a wide range of polar albedo spectra, measured with different instruments and at different wavelengths. Our work has followed his modeling efforts first in the thermal IR, then in the NIR, and finally in the VIS.

It should be emphasized that our primary purpose is not to interpret the micro-physical structure and properties of the frost but to reproduce, as accurately as possible, it's albedo spectrum – the parameter which enters our sublimation flux calculations. Hansen's model was incorporated in the Bjorkman & Wood's original code during the early stages of this research. This model has only one free parameter for pure CO_2 ice and three free parameters overall, which have to be constrained from observations: CO_2 "grain size", amount of dust intermixed in the frost, and amount of H_2O intermixed in the

frost. The term "grain size" could be misleading. For sizes more than a few mm, it is an approximation of the *photon's mean free path length* (MFPL) in the ice. The term "grain size" has been a subject of some confusion, so hereafter it will be substituted by the symbols MFPL or r[CO_2].

There are other frost albedo models in the literature. Eluszkiewicz et al. (2005) approximate the frost medium as a porous structure with two free parameters for pure CO_2 (versus one in Hansen's model). Eluszkiewicz's model is by all means useful and it indeed also reproduces to a good degree polar albedo spectra. However Eluszkiewicz has emphasized numerous times in conferences that his model's concept of "porosity" is physically more realistic than Hansen's concept of "grain size" in terms of the frost micro-physics. This may be true in some cases like slab ice, but in other cases like freshly deposited fine ice Hansen's model is likely more adequate, because for fine grains the r[CO_2] may be treated as a dimension of a physical grain.

In reality, this whole argument over the micro-structure misses the main point: a real frost structure is much more complicated than these models; it has a rough surface, possibly cracks, and non-spherical voids. If one is focused on a particular surface feature, a more realistic attempt to model the frost structure would be worth it. One the other hand, the purpose of both Hansen's models and our work is to describe the wide variety of polar ice spectra with the simplest and best understood model at hand. Hansen's model, which we use, has continuously been validated against surface spectra, and is able to describe a various polar ices with a single length parameter (MFPL). Therefore its usage in a Monte Carlo routine is justified.

8.7 Constraining the Frost Albedo Model Parameters from Multiple Spectral Regions

A major contribution of this work is that it consolidated efforts to constrain the parameters for the surface frost albedo spectra from multiple spectral regions. We are going to refer to these parameters as "surface frost properties", with the clear perception that these are properties of the modeled frost used to describe the real spectrum. This distinction is not usually made but is implicit in all related work and is important to make. As obvious as it is, the parameters of the CO_2 frost should be interpreted only in the context of the model used to reproduce the real frost spectrum. Our purpose is the most adequate reproduction of the variety of polar ice albedos observed in the south polar cap.

The rest of this chapter will be developed deductively: First, it will outline the three spectral regions diagnostic for the three parameters constraining the surface albedo spectra: the photon's mean free path length, the amount of intermixed water, and the amount of intermixed dust in the CO_2 ice. These spectral regions will be shown in conjunction with our latest models. The subsequent sections will chronologically describe the improvements in the albedo models since 1999 and the related results of modeling the SPC sublimation under dust storm conditions.

Figures 45, 46, and 47 show albedo spectra produced with Hansen's latest models. In each plot two of the three model parameters are fixed, while the other one is varied. Note that the shown ranges do not correspond to the full observed variation of these parameters over the SPC.

Figure 45 shows spectra over the whole spectral range: from the near UV to the

thermal IR. The varied parameter is the amount of intermixed dust. The displayed variation is characteristic for the perennial SPC (which the latest models were used to study). The seasonal cap shows larger range in the amount of intermixed dust.

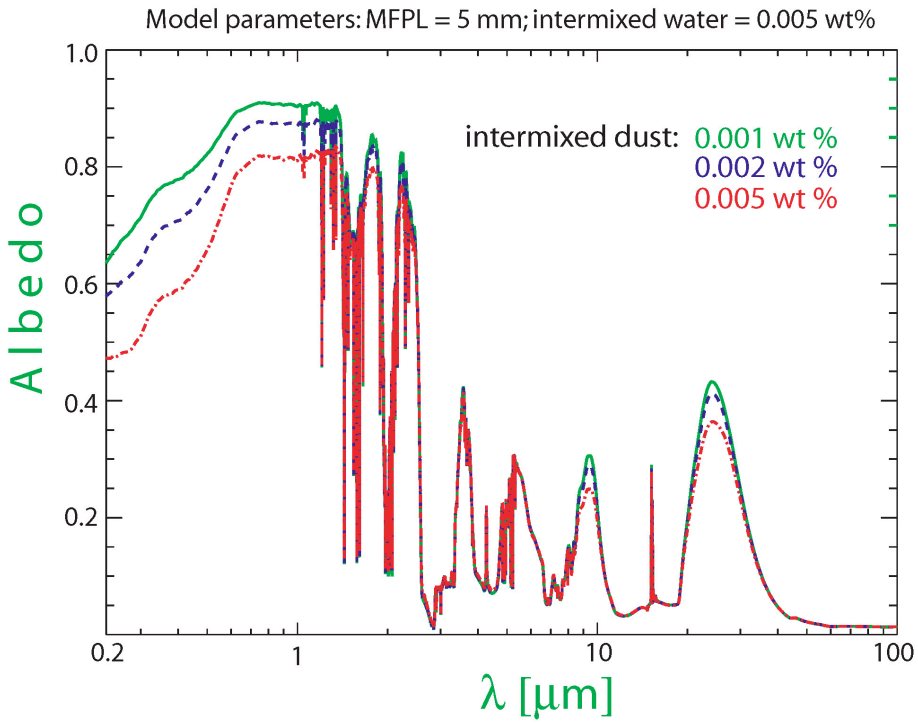

Fig. 45 – Models for CO_2 ice surface albedo spectra for photon incidence angle of 60°, showing the models' sensitivity to the amount of dust intermixed in the frost.

The hight of the albedo maximum between 20 and 30 μm is to some extent diagnostic of the amount of intermixed dust. This maximum is largest for pure CO_2 ice while increased dust tends to "erase" the feature. Hansen (1999) used this feature displayed by TES data to model the great variety of albedo spectra in the seasonal SPC. Although this feature is to a limited extent diagnostic for the amount of intermixed water, Hansen modeled the observed spectra in the thermal IR by essentially two parameter models (MFPL and dust content).

Figure 45 shows that the visible region is more diagnostic for the amount of intermixed dust. Variations of this parameter over the relatively narrow range typical for

the residual cap produce notable differences in the spectral appearance. These differences are more pronounced than in the thermal IR. Therefore the VIS region is diagnostic for intermixed dust, provided the other two parameters have been independently constrained. Recently visible albedos of the south polar caps have been obtained from well-calibrated HST photometric data.

Two frequency ranges in the NIR are particularly diagnostic for the MFPL and the amount of intermixed water respectively. Figure 46 shows modeled albedos in the K-band, with fixed intermixed dust and water contents. The MFPL is constrained from the relative absorption strengths of some spectral features and from the overall shape of the spectrum.

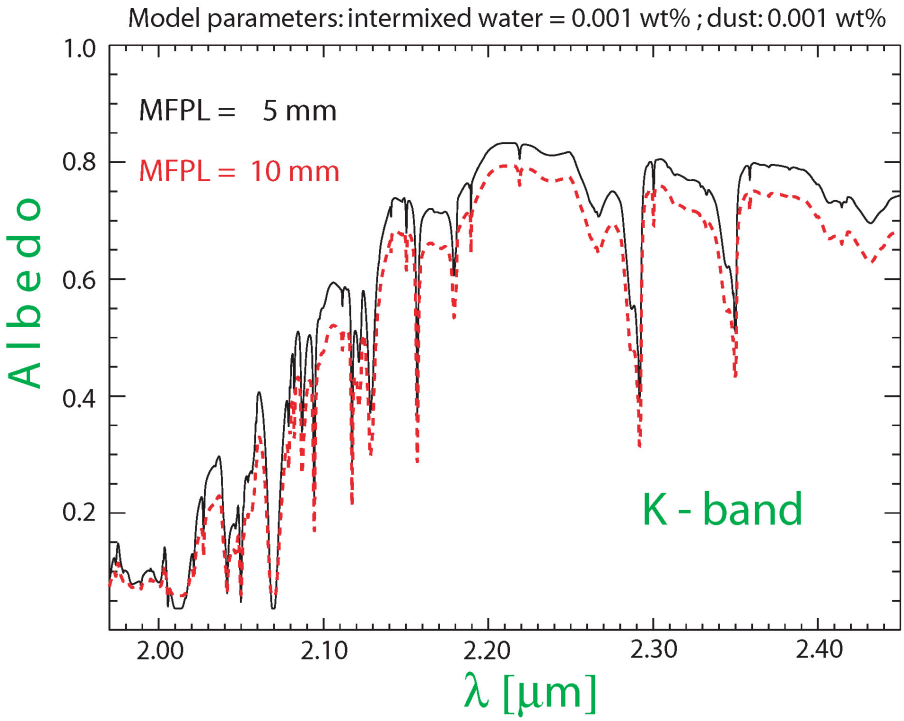

Fig. 46 – Same as Figure 45, but for the K-band region which is most suitable to constrain the mean free photon path in the ice – the only length parameter in Hansen's models.

Figure 47 shows models for the L-band, which is most diagnostic for intermixed

H_2O. The observed spectral features are CO_2, but intermixed H_2O affects dramatically the overall shape of the spectrum and the depth of these features.

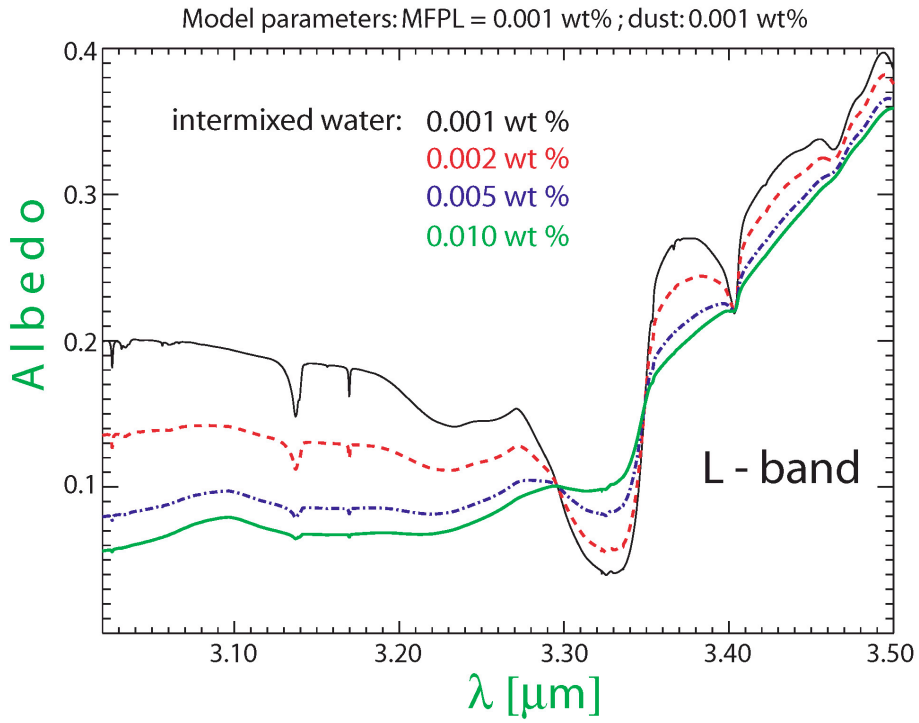

Fig. 47 – Same as Figure 45, but for the L-band region which is most suitable to constrain the amount of water intermixed in the CO_2 ice.

NIR medium resolution data have been acquired from both space-craft and ground-based instruments. Hansen et al. (2005) analyzed perennial SPC spectral from the Planetary Fourier Spectrometer (PFS) on board of Mars Express (MEX). This instrument does not have the spectral resolving power to conduct reliable gas phase studies of trace constituents in the Martian atmosphere, but is well suited to do a solid phase spectroscopy and has provided a valuable data set for the SPC.

Directly motivated by the project described in this chapter, Glenar et al. (2005) and Glenar (in preparation for a collaborative paper) analyzed spectra from the SPEX instrument at NASA IRTF in both the L-band and the K-band. In spite of the very different spatial resolution, Glenar's results (which reflect only the mean residual cap

properties) are very consistent with the space-craft data.

The disadvantage of using data from different instruments and facilities is that these data are "inhomogeneous". They cover vastly different spatial scales and are subject of various interpretation uncertainties. The Earth-based (HST; NASA/IRTF) data have lower spatial resolution, but even space craft spectra cannot resolve all surface feature of varying albedos. Nevertheless our understanding of especially the residual cap frost properties is definitely advance now, in comparison to Hansen's 1999 paper.

In spite of being "inhomogeneous", using data from multiple spectral regions has the serious advantage, because neither of these regions alone can tightly constrain the parameter space of frost properties. On the other hand data from at least the NIR and the visible are much more diagnostic together.

At this point it is time to go back to the earliest stage of this project, when the surface in our model was described by two parameters: a constant visible albedo and a constant IR emissivity. Had we stayed on that stage we would probably repeat with a better atmospheric dust model the main conclusions from previous efforts of modeling the atmospheric dust effects on the CO_2 sublimation. James et al. (1979), Paige & Ingersoll (1985) and a number of other works have already shown by using two parameter (VIS albedo/IR emissivity) that the seasonal SPC can be affected by global dust storms in all possible ways depending on the ratio between these two parameters. In addition of using the best constrained single scattering properties of atmospheric dust by Wolf & Clancy (2003), as well as a state-of-art radiative transfer algorithm, it was important to use surface albedo specta constrained at least partially from real data. The first step was the study of the Mountains of Mitchel.

8.8 Regression of the Mountains of Mitchel Polar Ice After the Onset of a Global Dust Storm on Mars

The Mountains of Mitchel (MM) are located in the south polar region of Mars at about -72°, 330°. They were first reported by Orsmby McKnight Mitchel at the University of Cincinnati (Mitchel 1846). Antoniadi (1930) referred to them as Novus Mons. This high albedo region stands out even when it is still surrounded by darker seasonal CO_2 frost. In early spring, the MM emerge as an extension of the south residual polar cap. They subsequently separate and disappear around solstice. This regressions pattern was observed in most detail in 1999, using the Mars Orbital Camera (MOC) experiment on Mars Global Surveyor (MGS) (James et al. 2001). Earlier work (James et al. 2000) suggested that the MM were susceptible to inter-annual variability, but little difference was observed between the 1997 and 1999 regressions.

In a Geophysical Research Letter (Bonev et al. 2002), we reported the observations for the Martian southern spring in 2001 and compared the regression rates of the MM for 2001 and 1999. Our primary motivation was the very different atmospheric conditions during early spring of these two Martian years. In particular, the early phase of the 2001 regression occurred during a global dust storm, which rivals that observed by Mariner 9 (Smith et al. 2002), while the 1999 spring season was significantly less dusty (very similar to the 1997 season). MOC images (Figure 48) clearly show accelerated sublimation in the MM region in 2001. This accelerated spring-time recession of the was demonstrated independently by Titus & Kieffer (2002) from TES observations based upon surface temperatures. The TES measurements of recession curves are based on the thermal contrast between a sublimating surface and the revealed

"bare ground" which has significantly higher temperature.

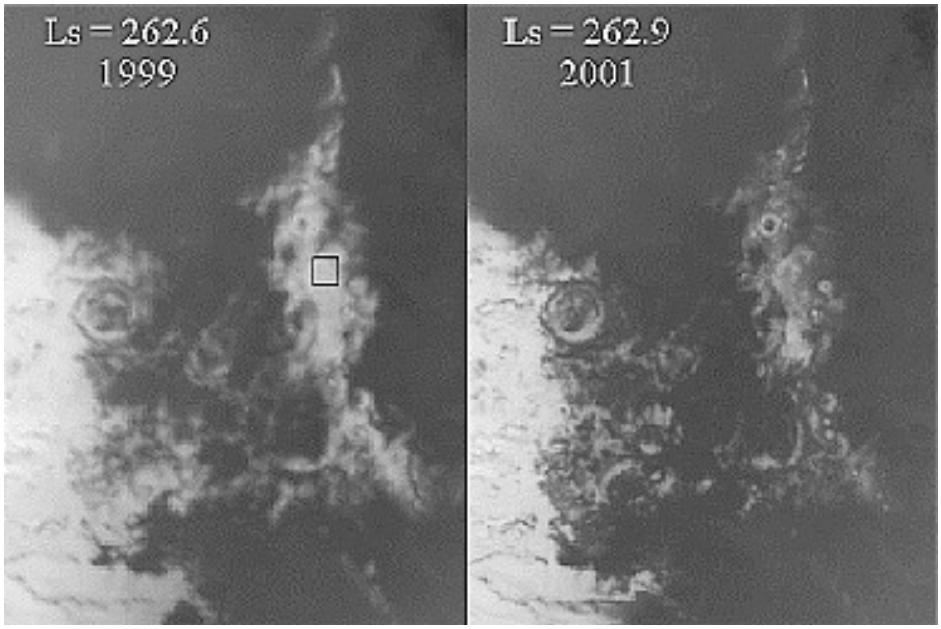

Fig. 48 – MOC red filter images of the Mountains of Mitchel showing faster regression in 2001.

This evidence for interannual variability raised again the question for the effects of atmospheric dust on the CO_2 sublimation rate. As pointed out earlier in this chapter there have been two major effects to consider:

1. Reduction of the direct solar heating due to absorption by dust and,

2. The opposing increase in the absorption of IR thermal dust emission by the surface (cf. James et al. 1979; Paige & Ingersoll 1985).

To estimate the net effect on the CO_2 sublimation rate, we performed our first representative runs with our model for radiative transfer through a dusty atmosphere bounded by a sublimating CO_2 surface. Our purpose was to obtain a calculation of the sublimation rate for a qualitative comparison with the MGS data showing the Mountains of Mitchel regression patterns.

Another related observation which we addressed was from the aforementioned study of Titus & Kieffer. They showed that visibly dark south polar areas, such as the

"cryptic" region (Kieffer et al. 2000), regressed more slowly after the global dust storm as compared to the previous less dusty spring season. It seemed that the cap seasonal cap's "response" to the global dust storms was on a local scale and correlated with the visible albedo.

Hansen's 1999 work provided the only available surface albedo spectra for the south seasonal cap, which were suitable for our problem. This author (now collaborator) had shown that a surface layer of CO_2 particles of MFPL ~ millimeter, mixed with various amounts of dust rather than water ice, is consistent with his analysis of brightness temperatures from TES data (the 20-30 μm region shown on Figure 45).

With parameters constrained only from the thermal IR, Hansen's models implied that, in the visible, the CO_2 albedo is above 0.8 for a wide range of MFPL values, while in the thermal IR, the emissivity shows a distinct minimum around 25 μm whose depth varies with MFPL. Intermixed dust on the surface tends to bring the IR emissivity closer to unity and, most importantly, significantly lowers the visible albedo. Thus the relatively high visual albedo of the Mountain of Mitchel implied a surface with small dust content.

Based on Hansen's surface spectra, we considered two effects of dust in our model: atmospheric dust, which redistributes the incident radiation, and surface dust, which changes the surface albedo. We studied two limiting cases of the latter: a pure CO_2 surface with MFPL = 1 mm (clean ice) and CO_2 frost darkened by one weight percent intermixed dust (dirty ice). This is a very high dust content, on the upper edge of Hansen's; retrievals. The albedo spectra are given in Hansen's 1999 paper, and similarly in the review of James et al (2005a), which really covered the work before 2003.

We modeled the variation of the CO_2 sublimation flux with atmospheric dust optical depth. In all calculations presented in this chapter, the CO_2 sublimation flux is given by the difference between the frequency integrated fluxes absorbed and emitted by the surface.

Figure 49 (Figure 2 in Bonev et al. 2002) shows the variation of the CO_2 sublimation flux with the 550 nm dust optical depth (τ_{550}). This optical depth is about 2–3 times larger than the optical depths at most IR wavelengths. Note that TES team-provided opacities are not available until the cold surface surrounding the Mountain of Mitchel has sublimed because the algorithm they use for retrieval of aerosol opacities is restricted to surface temperatures >220 K (Smith et al. 2001). Therefore, lacking any real estimates, we varied the optical depth up to $\tau_{550} = 2$. The calculations presented are for solar zenith angle near 75°, which corresponds to Ls = 210° and a latitude of -72 (note that the major dust storm has ended by about Ls = 220°). The sublimation flux has been normalized to the total flux incident to the top of the atmosphere. This is a convenient unit for modeling and will be used in presenting all subsequent results.

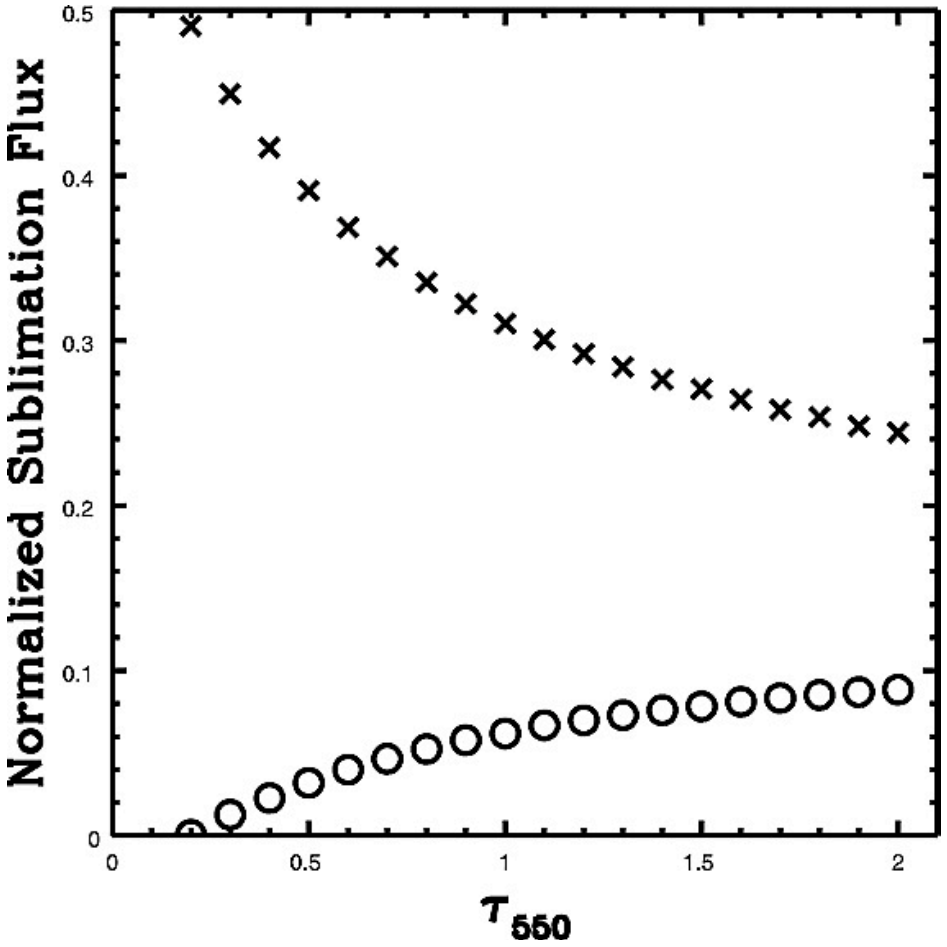

Fig. 49. – Normalized CO_2 Sublimation flux (SF) vs. dust optical depth at 550 nm. Results are shown for two surface albedo spectra: pure CO_2 with a MFPL = 1 mm (circles), and CO_2 frost darkened by one weight percent intermixed dust (crosses). The SF is given by the difference between the frequency integrated fluxes absorbed and emitted by the surface, which are normalized to the frequency integrated flux incident on the top of the atmosphere. Note that a pure CO_2 surface has a very high visual albedo, so the SF is zero for the smallest optical depths.

The primary result of our modeling was that, for clean ice, the sublimation rate increases with dust optical depth. Conversely for a dirty ice, the sublimation rate decreases with optical depth. The specific (and uncertain) value of the solar zenith angle we used did not play a major role in this result (models with a solar zenith angle of 60° showed qualitatively the same picture).

For dirty ice, the absorbed flux is dominated by the visible radiation. Consequently the atmospheric dust extinction of the incident sunlight decreases the sublimation flux as the dust optical depth increases. On the other hand, for clean ice, the absorbed flux is dominated by IR radiation, owing to the large visible albedo. As the dust optical depth increases, the atmospheric dust gradually shifts part of the incident solar radiation to IR wavelengths where the overall absorption of the CO_2 surface ice is larger, despite the CO_2 emissivity minimum around 25 μm, thereby increasing the CO_2 sublimation rate. We therefore concluded that the IR thermal dust emission was a plausible explanation for the faster regression of the MM in 2001, assuming the MM ice had a small dust content.

Figure 49 also shows that the sublimation rate of dirty ice is higher than that of clean ice. So an alternative explanation for the faster regression of the MM in 2001 might simply be that newly intermixed dust was deposited in the early stage of the dust storm, rather than the IR redistribution of the incident radiation.

The increased surface dust scenario cannot be ruled out, however, observations provide indirect evidence that the effect of atmospheric dust was dominant. First, MOC data shows that, like in 1999, the Mountains of Mitchel remained one of the brightest polar cap features, which implies a small content of intermixed surface dust. Second, the

aforementioned results of Titus and Kieffer (2002) about the slower recession of the dark south polar areas, such as the "cryptic" region, regressed more slowly after the global dust storm as compared to the previous less dusty spring season. If the dust storm simply increased the surface dust content, one would expect both bright and dark regions to regress faster. For the dark regions to regress more slowly, we require the IR redistribution produced by atmospheric dust. We concluded this opposite regression behavior was in qualitative agreement with our atmospheric dust model.

In 2003 the next recession of the Mountains of Mitchel was monitored by MOC. Benson and James (2005) confirmed the main conclusion from our 2002 paper. There was not a 2003 major dust storm covering the polar region during southern spring. The Mountains of Mitchel recession in that year was very similar to the one in 1999 and slower than the year with a global dust strom (2001) in accordance with our model predictions.

Figure 49 shows the output of the first representative runs (related to Mars) of the Monte Carlo approach described above. The parameter space of surface properties covered only two limiting cases: high visible and dark visible albedo. The next natural step was to fill the gap in the parameter space, because there is a large amount of variation in the visible albedo of surface frost in the SPC. Using albedo spectra representing this variation would allow us to address the question of the apparent "insensitivity" of the mean seasonal cap recession to increased atmospheric dust loads.

8.9 The Average Recession of the Seasonal South Polar Cap

By mid-2002 Gary Hansen had already become an official collaborator and supplied us with a series of models covering the wide range of albedo spectra observed in the seasonal SPC. Most of these spectra represented "typical" frost. Therefore they needed to be used if the average cap's recession had to be considered. The albedo spectra were empirically reproduced by fitting TES brightness temperature data with two parameter models. The resulting spectra covered photon path lengths varying by orders of magnitude and intermixed dust contents of intermediate values compared to our previous parameter study of limiting cases.

The results from our first detailed parameter study are shown on Figures 50, 51, and 52. These results are presented in the same form as in Figure 49. They incorporate all three types of surface responses to increased atmospheric dust load. The lowest curves represent the visibly brightest regions. For them the accelerated sublimation can be explained by the effect of atmospheric dust (shifting the incident energy to frequencies where the surface is more absorptive) or newly deposited surface dust (decreasing the visible albedo). The latter case corresponds to a transition from one curve to another as the atmospheric dust optical depth increases.

The curves corresponding to the fastest sublimation correspond to the visibly darkest regions. Their reduced sublimation fluxes as the dust optical depth increases can be explained only by atmospheric dust extinction in the visible, where these surfaces absorb well in dust-free conditions.

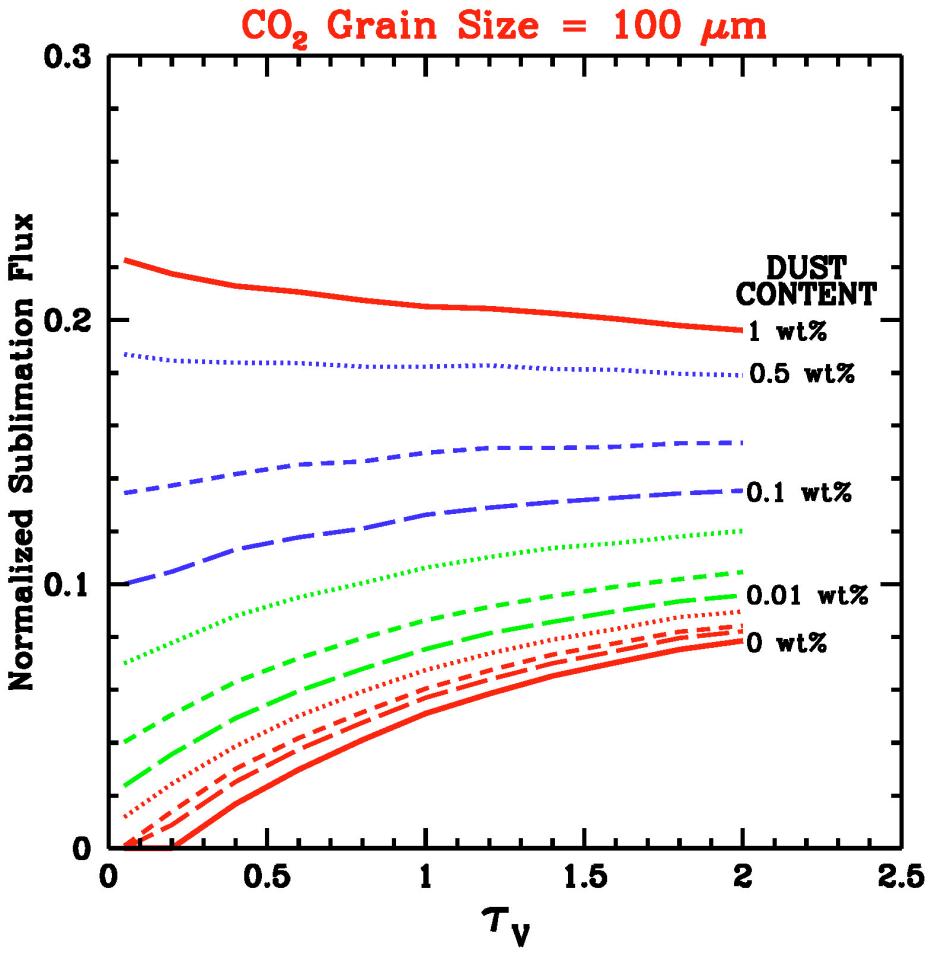

Fig. 50 – CO_2 Sublimation flux vs. total atmospheric dust optical depth at 550 nm for a frost MFPL = 100 μm and various contents of intermixed surface dust. Albedo models from Hansen (1999). The calculations are done for L_s = 210° and solar zenith angle of 70°.

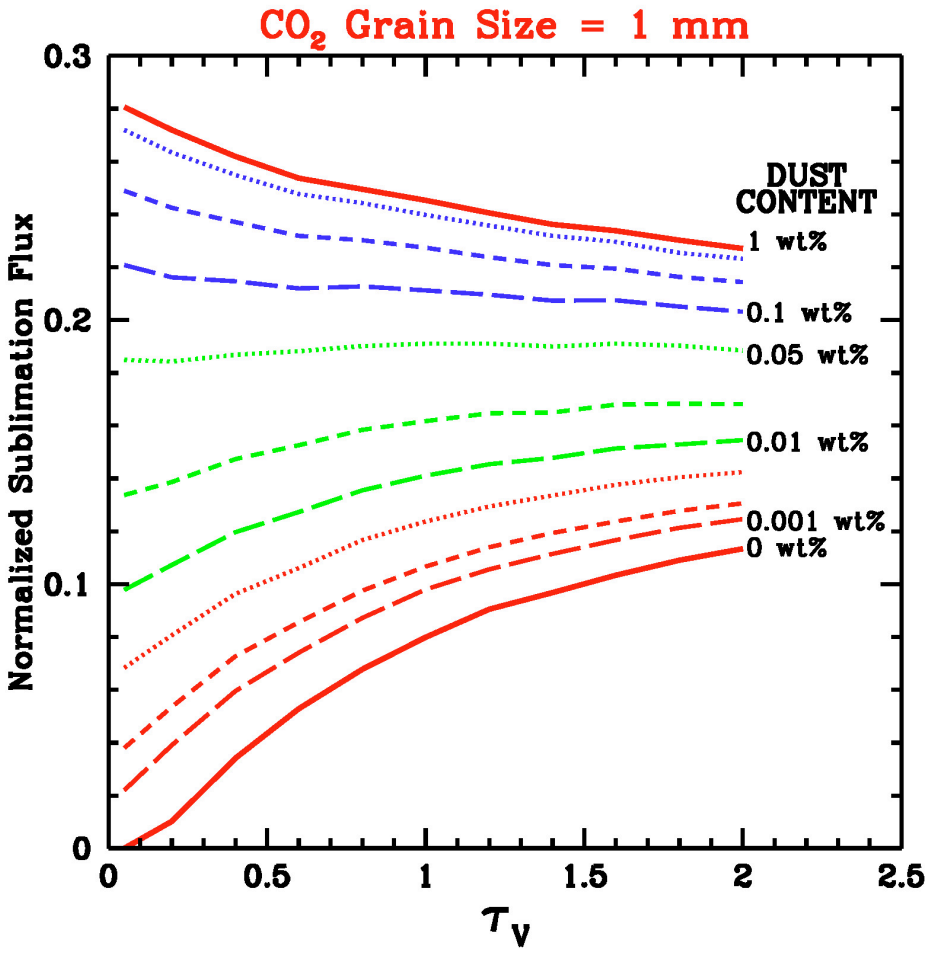

Fig. 51 – Same as Figure 50 but for MFPL = 1 mm.

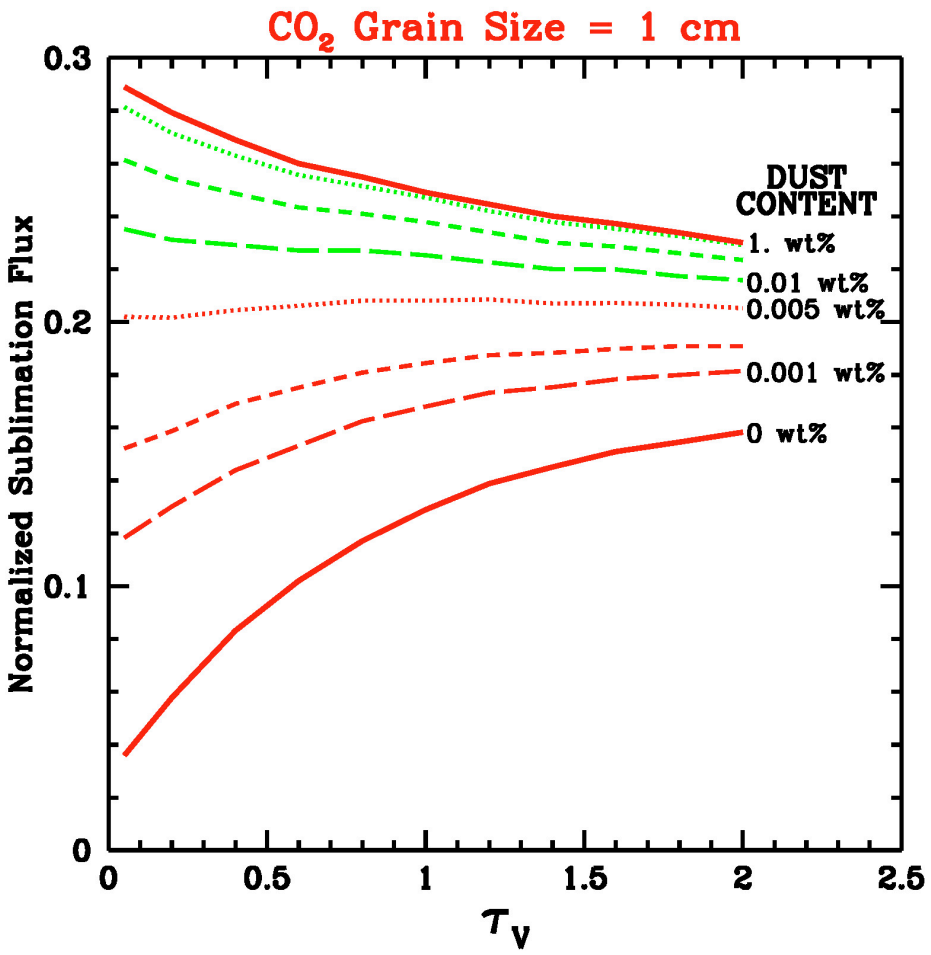

Fig. 52 – Same as Figure 50 but for MFPL = 10 mm.

The new results was the wide range of albedo spectra for which the CO_2 sublimation fluxes showed only subtle variations with the amount of atmospheric dust load. In these cases the surface absorption is distributed equally between visual and IR wavelengths, so the overall atmospheric dust effect is not important. It should be emphasized that the discussed region of the parameter space represents a "typical frost" according to Hansen's 1999 study. Consequently it explains the apparent insensitivity of the average decay rate of the south seasonal cap to dust storm activity. Strong coupling between sublimation and atmospheric dust is expected primarily on local scale for regions with "deviant" surface albedos.

This work was presented at the DPS meeting in 2002. Subsequently Benson & James (2005) measured the seasonal SPC recession curves in 2003. It showed a very little difference compared to the recessions in 1999 and 2001 consistent the our prediction that mean decay rates should be similar from year to year and relatively independent on dust storm histories.

The question remained of whether the behavior of the perennial SPC can be explained within the context of our modeling. The residual cap has an even higher albedo than the Mountains of Mitchel, yet in terms of its large scale appearance it does not seem to have sublimed more in the dust storm year (2001) than during 1999 and 2003. As pointed out earlier in this chapter, no much dust penetrated into the perennial cap during the 2001 storm. The only year when the whole residual cap was certainly covered by dust was 1971. This planet-encircling dust storm was observed by Mariner-9. It occurred near perihelion, when the insolation over the south pole is largest and

therefore radiative effects are expected to be strongest.

The cap's appearance documented by Mariner-9 contrasts its appearance as inferred from MGS observations (Figure 41). It apparently had lost more solid CO_2. Could the 1971 dust storm be the reason? More definite results required better constrained surface albedos.

8.10 Visible and Near-Infrared Observations on the Residual South Polar Cap

The surface models used in the study of the seasonal cap were not "free parameters"; instead they were based on TES brightness temperature data. However, the thermal IR is limited as a diagnostic range, especially in studies of the residual cap whose albedos display a narrower range compared to the seasonal cap and therefore their retrieval requires more sensitive diagnostic frequency intervals. So in 2003 it was clear we should put the modeling efforts on hold until better constrained albedos of the perennial SPC become available (this was also a good point to start focusing on the comet work).

Our results from the previous two sections motivated directly the study of Glenar et al. (2004) who developed their methodology to retrieve average properties of bright regions in the SPC (including the residual cap) from NIR medium resolution spectra acquired with SPEX at NASA/IRTF. Glenar's most recent results are specifically for the perennial cap and are part of the collaborative study intended to unite the efforts of retrieving albedos and modeling the cap's sublimation rates under dust storm conditions into a concise article.

In parallel Gary Hansen and the PFS team conducted an equivalent study but based on much higher spatial resolution space-craft data (Hansen et al. 2005). Their results are also based on NIR and are very sensitive to MFPL (Figure 46) and amount of intermixed water (Figure 47).

In 2003-2004 James lead the effort of measuring visible residual cap albedos. Previous measurement at specific visible wavelengths were conducted from ground-based (Lumme & James 1984) and space-craft (James et al. 2001) observations, but suffered from limited frequency coverage and calibration uncertainties. The new study was conducted with the High Resolution Camera on HST during the very favorable opposition of 2003. The major strengths of this data set are the wavelength coverage, the good photometric accuracy, and the high spatial resolution (for Earth-based telescope). These observations were made using a variety of filters covering a spectral range from 250 to 892 nm; specifically, the filters used were F250W, F330W, F344N, F435W, F475W, F502N, F658N, and F892N. The three digits in each filter name give the central wavelength in nm, and the W and N stand for wide and narrow, respectively.

The main source of uncertainty in the retrieved surface albedos was in correcting for the effects of atmospheric dust, which even at background levels ($\tau_{550} \sim 0.2$) affects the top-of-the atmosphere visible spectra. Background dust optical depth had to be assumed above the south pole, because the TES-retrieval algorithms are still not fully developed for an atmosphere above an icy (i.e. cold) surface. Nevertheless the results of James et al. (2005b) are valuable at least as precise lower limits of the perennial cap albedos for the observing geometry of the HST measurements.

These three data sets are complementary and therefore it was appropriate to used

them in a joint study of the residual cap's properties. I have been "campaigning" for this approach since our first Mountain of Mitchel paper, so it was delightful to see it started.

8.11 Parameters for the New Surface Albedo Models

For modeling the stability of the perennial cap against a global dust storm near perihelium we have constrained the parameter space of surface models, so that the latter reproduce as closely as possible the albedo spectra in the residual cap. Gary Hansen and Dave Glenar had the major involvement in this part of the work. Within the context of Hansen's model we needed to produce a set of model parameters: MFPL, amount of intermixed H_2O, and amount of intermixed dust.

At first stage PFS/MEX and SPEX/NASA IRTF spectra in the K-band were used to constrain the MFPL; independently data in the L-band constrained the amount of intermixed H_2O.

Subsequently, the visible albedos of James et al. (2005b) were fit by models with MFPL and water content ranges independently established from the NIR, while the intermixed dust content was used as a free parameter. Such inter-wavelength approach broke the degeneracy in the parameter space of intermixed dust and H_2O. For a fixed MFPL, the latter parameter raises the visible albedo, while the former lowers it, so a wide range of combinations of the two parameters result in almost the same visible spectra. With range of H_2O amounts constrained from the L-band, the dust was adjusted from the visible.

For modeling visible spectra Hansen first derived a new set of dust optical properties in that spectral range and used then to produce new model albedo spectra (CO_2

ice with intermixed dust and water) for comparison with observations. This comparison also took into account the specific observing geometry of the HST observations in 2003.

Figure 53 shows modeled surface albedo spectra with parameters constrained from the joint analysis in the NIR and in the VIS. The retrieved MFPLs are in the range 5 – 10 mm, with intermixed dust and water ranges respectively 0.001 – 0.005 and 0.001 – 0.01 weight percent.

The visible HST data is also displayed. Albedo lower limits for three different values of L_s near opposition (235°, 251°, 265°) are shown as diamonds; albedos corrected for background atmospheric dust during the 2003 observations are shown as stars.

It should be pointed out that the parameters of the displayed models are retrieved for the specific observing geometries of the corresponding observations by PFS, SPEX, and HST. On the other hand the models on Figure 53 are calculated for a fixed incidence angle of 60° for the photons hitting the surface. The visible data are displayed because the models for the specific observing geometry do no differ substantially from the simpler calculation show here.

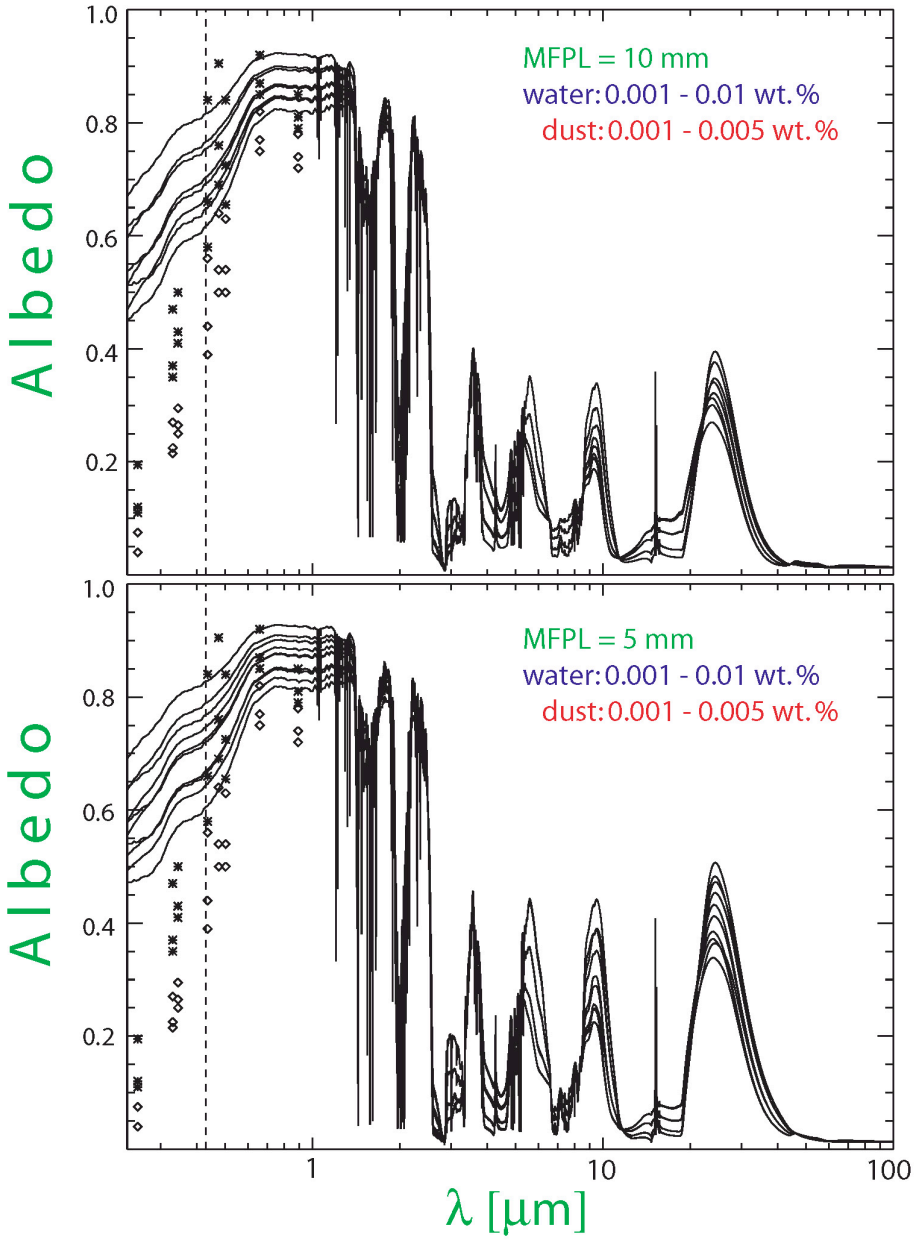

Fig. 53 – The latest models for surface albedo spectra with parameters retrieved from observations of the residual south polar cap.

The produced models match well the NIR spectral features (model – fit comparisons are shown in Hansen et al. 2005) and the VIS data long-ward 400 nm. We still cannot reproduce the albedo decrease in the near UV. This is most probably related to the fact that the properties of the intermixed dust are not well-constrained in the near-UV. Therefore for calculating sublimation rates the following approach was adopted:

1. For $\lambda \geq 420$ nm modeled albedo spectra were used, with parameter ranges on Figure 53.

2. For $\lambda < 420$ nm we used a mean from the observed HST albedos at the corresponding wavelengths and interpolated for intermediate wavelengths in order to reproduce the observed decrease in the albedo towards the near UV. This approach leads to more reliable results of our Monte Carlo modeling that will be described in the next section.

Realizing the uncertainties involved, this is the best approach possible with the current data from the residual SPC. The problem to reproduce the near UV VIS albedos will be addressed with more observation. The Omega spectrometer on board of MEX has for example a VIS channel; there is also a VIS camera on board on a new mission, *Mars Reconnaissance Orbiter*, with primary involvement of Philip James. On the other hand the issue with the UV modeling should not cloud the reliability of the models at hand in most of the VIS and in the NIR – the most diagnostic frequency ranges for retrieving CO_2 frost properties.

8.12 Modeling the Sublimation of the Residual South Polar Cap Under Conditions of a Global Dust Storm Near Perihelion

The last section of this overview shows the results from a simulation of a dust storm similar to the one in 1971. There are no precise optical depth estimates for this event, but it was well documented that the whole residual cap was obscured by dust, and therefore an optical depth significantly larger than unity in the VIS is justified. In order to model the effect of the 1971 global dust storm, we assumed that the optical depth over the perennial SPC could be represented as an abrupt jump from a background opacity level of $\tau_{550} = 0.2$ to $\tau_{550} = 2.0$ at onset. This increase in optical depth is followed by an exponential decay to pre-dust storm levels at $L_S \approx 300^o$, similarly to the observed time scale of the real event in 1971.

For each surface albedo spectrum shown in Figure 53 we calculated the variation in sublimation flux with L_s for two modeling scenarios:

1. Global dust storm scenario in which the change in sublimation flux is governed by the variable atmospheric opacity and by the variable insolation conditions.

2. Background dust scenario, in which the dust optical depth is assumed to be constant ($\tau_{550} = 0.2$) throughout the simulation, so the changes in sublimation flux are solely insolation driven.

The main result of this simulation is that for all surface models shown on Figure 53 the sublimation fluxes have a notable increase in the global dust storm scenario compared to the case of background dust. Because this is an overview chapter, the outputs of only several simulations will be shown. They encompass the magnitude of the

dust storm effect within the range of all input albedos.

Figure 54 shows the variation of sublimation fluxes with L_s for two albedo spectra modeled with MFPL = 5 mm. The results shown on the upper panel corresponds to a brighter VIS albedo between 400 and 800 nm. Figure 55 shows the results for two input albedo spectra modeled with MFPL = 10 mm. In these plots all sublimation fluxes are normalized to the flux incident to the top of the atmosphere for Ls ~ 302° (the point where the two curves converge).

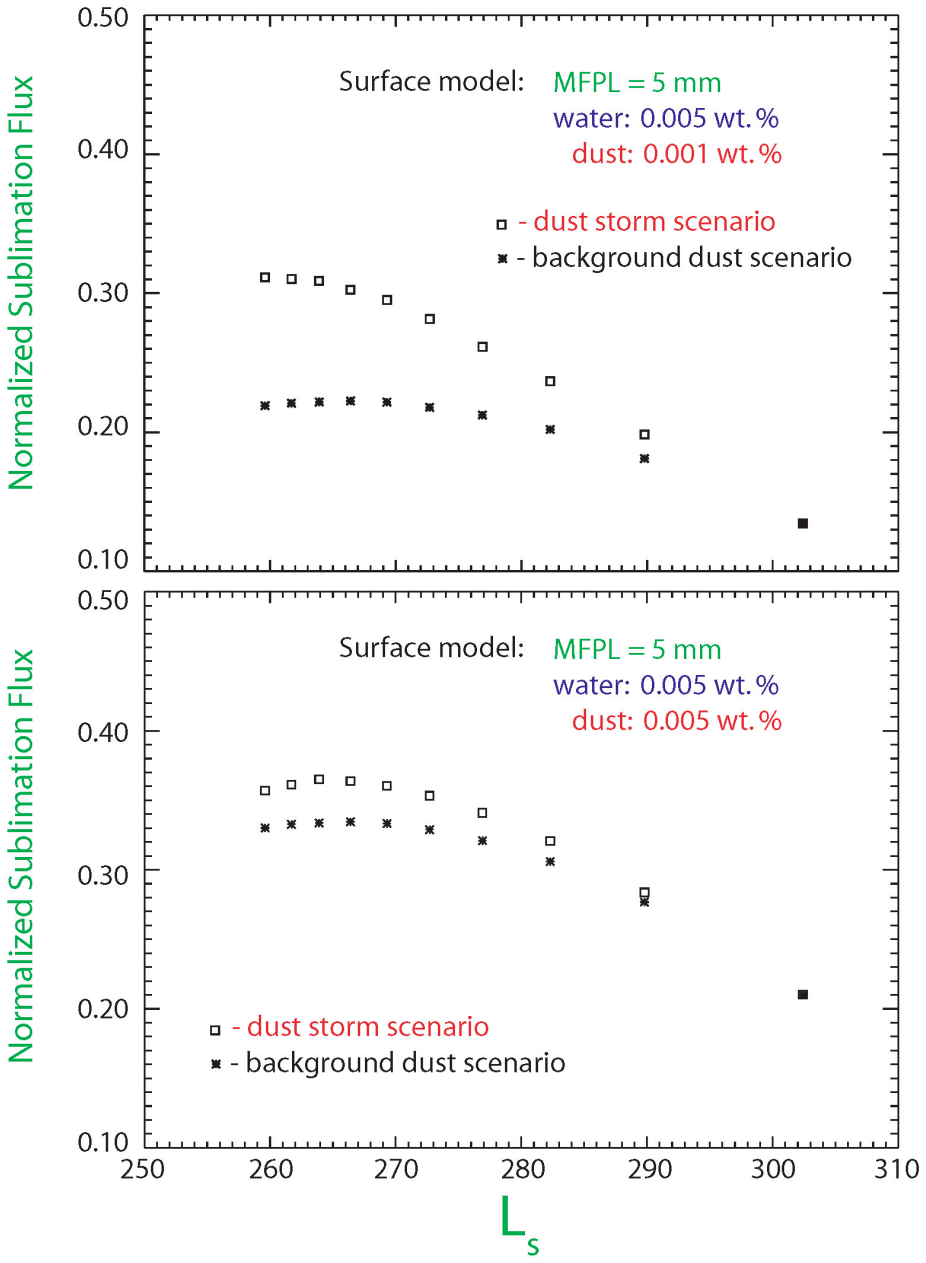

Fig. 54 – Sublimation flux versus L_s for two models with MFPL = 5 mm. For the "dust storm scenario" the optical depth at 550 nm varies from 2.0 ($L_s \sim 259°$) to the background level of 0.2 ($L_s \sim 302°$). For the "background dust scenario" the optical depth at 550 nm is 0.2 for all values of L_s, while the changes in sublimation flux are insolation-driven. The sublimation flux is normalized to the flux incident to the top of the atmosphere for the point near $L_s \sim 302°$ where the two curves converge.

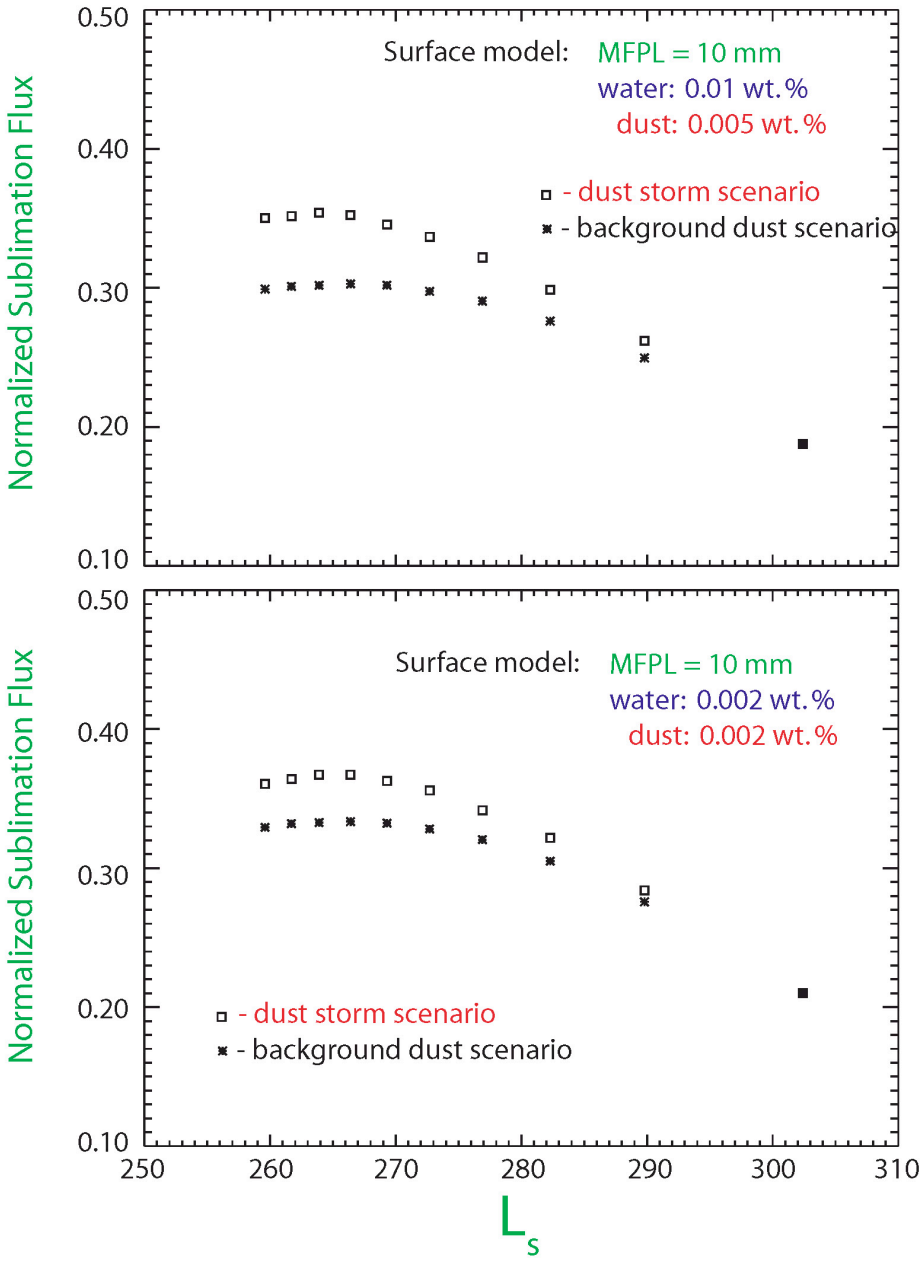

Fig. 55 – Same as Figure 54 but for two models with MFPL = 10 mm.

The main significance of this result is that it is based on parameters consistent with real data from Mars. The most important radiative effects in these simulations are the VIS-to-IR redistribution of energy reaching the surface by atmospheric dust, and the surface absorptive properties over a wide frequency range. The input parameters related to the precise modeling of these two effects are respectively the single scattering properties of Martian aerosols and the surface albedo spectra. Both are based on data from multiple spectral regions as shown by Wolff & Clancy (2003) and Clancy et al. (2003) for the properties of aerosols and in this overview for the surface.

Our new results for the perennial cap are not critically affected by the uncertainty in the near UV albedo. In fact, forcing this albedo to match the HST data is a more conservative choice than using the models shown on Figure 53 for wavelengths shorter than 420 nm. The reason is that these models have higher near UV albedos. Therefore the increase in sublimation flux for them will be even larger because the differential absorption between VIS and IR is larger for high VIS albedos.

Our results for intensified sublimation imply an advanced recession of the residual cap after the 1971 dust storm compared to later "non-dusty" years. This is consistent with the observations as shown on Figure 41.

If the visible albedo of the present day residual cap is representative for the current Martian climate, it is high enough to ensure its stability under conditions of background atmospheric dust (James at al. 2005b). For example Thomas et al. (2005) showed that the surface features known as "Swiss cheese" comprise several different units and that there is evidence for erosion/deposition of some of these units within the last 100 Martian years. One unit has been formed much more recently – after the 1971

dust storm. It is very likely that such an event can partially "break" the perennial cap's stability, leading to a more advanced sublimation in the "dusty" year and to subsequent re-accumulation of surface ice deposits.

Appendix 1: High Resolution ($\lambda/\Delta\lambda$~2-3x10^4) Spectroscopy of Comets. CSHELL/NASA IRTF and NIRSPEC/Keck 2.

Strengths & Weaknesses.

CSHELL (single-order echelle spectrometer at commissioned in 1992) and IRTF:

- 256 x 256-pixel (0.2" pixel size) InSb array detector (sensitivity from ~1.0 - 5.5 μm).
- Circular Variable Filter (CVF).
- **Very limited free spectral range per setting.**
- **The instrument is an "honored veteran" and any serious technical problem has the potential to de-commission it.**
- Slit length = 30" (~150 detector rows on the sky).
- **Active CCD guiding.**
- **Daytime observing capability of IRTF, albeit without guiding.**
- **Planetary Observations are high-priority at IRTF, which enables quality proposals to win significant amount of observing time.**
- **Very good facility for student's initial training because the new user can**

operate the instrument for many hours. In addition daytime observations offer the full range of problems related to tracking non-sidereal objects.

NIRSPEC (cross-dispersed echelle spectrometer commissioned in 1999) and Keck 2:

- 1024x1024-pixel InSb array (0.14”/pixel in spectral dimension, 0.19”/pixel in spatial dimension).
- **Multiple echelle orders observed simultaneously: 6 in L-band (~2.7 - 4.2 μm), 3 at M (~4.4-5.5 μm).**
- **Larger-aperture telescope (10 m vs. 3 m at IRTF) => more light-gathering power & tighter point spread function.**
- Slit length = 24" (~125 detector rows on the sky).
- **Both offset and active guiding for $\lambda < 4.2$ μm only; offset guiding at M-band.**
- **No daytime observing (Keck Observatory policy).**
- **40^{o} elevation limit in west at Keck 2 (physical obstruction).**

Appendix 2: Introduction to NIRSPEC Data Reduction and Flux Calibration

The reduction sequence for NIRSPEC data is applicable (with some adjustments) to all high-resolution echelle spectrometers currently at use or being considered for the next several years. The former group includes CSHELL at NASA IRTF, Phoenix at Gemini-South, and IRCS at Subaru. The latter group includes CRIRES at VLT, the IRCS at Subaru with improved resolving power, and CSHELL-2 at NASA IRTF.

In what follows the raw A- and B-frames will be referred as "sky" frames because their dominant sources of signal are the background thermal continuum and superimposed atmospheric emission lines (see Chapter 3).

In their most general form some of the core procedures are in standard use in astronomy. But all algorithms have been developed at GSFC and tailored specifically to our comet observations. This overview is not designed to be a manual for a new user. Rather, its purpose is to provide a basic idea about the reduction sequence and to point out some caveats that require user consideration.

A2.1 Basic Processing Steps

An example of the "raw" data is shown on Figure 1 (an A-frame) and Figure 2 (A-B difference) in Chapter 3. These data represent an 1024 columns by 1024 rows arrays in which the separate orders are tilted with the tilt varying from order to order. Each spectral order within a given setting is processed separately. The reduction sequence consists of:

1. **Cropping to a 1024 column x 251 row array**, which includes the desired order, from the original 1024 x 1024 array. The new array's size (1024 x 251) is chosen to completely include the most tilted orders (21-23 in the KL2 setting).

2. **Dividing the cropped array by (flat - dark) frame**. The flat and the dark frames should have the same integration time. The flat field is taken using internal continuum lamp. There are two important issues to consider during observing run:

First, the flat field total integration time (the integration time per co-add times the number of co-added frames) should be adjusted so that this flat has a higher signal-to-noise (S/N) than the comet frame – the desire is the data to be limited by the S/N from the source frame, not from the flat.

Second, the flat field and the comet frames must be taken without moving the echelle grating setting. Consider two instrument settings called "KL1" and "KL2". Proper observing sequences can be:

KL1 (comet) – KL1(flat) – Dark – KL2 (flat) – KL2 (comet)

or

Dark - KL1 (flat) – KL1(comet) – KL2(comet) – KL2(flat) - Dark

A non-proper sequence is for example:

KL1 (comet) – KL2(comet) – KL1(flat) – KL2 (flat) – Dark

This is because, in addition to removing pixel-to-pixel sensitivity variations, flat-fielding can be used to remove more slowly varying effects such as a fringe pattern. After changing the instrument setting, the original grating position cannot be reproduced exactly. If moved, the difference between the positions of KL1 (comet) and KL1 (flat) would be very small, but sometimes enough to introduce artifacts like non-properly removed fringe patterns upon flat-field correction (same for KL2 and any other setting).

Note that only the flat field is a subject to a direct dark subtraction. In the comet data, the accumulated dark current cancels out upon taking an A-B difference, while abnormally high dark current pixels ("hot" pixels) from individual A- and B- frames are removed before applying that difference.

3. "Masking": identifying and then removing cosmic ray hits and "hot" pixels from the raw A- and B-frames using standard "sigma-filtering". For an "A1, B1, B2, A2" sequence cosmic ray hits are also identified in (A1-A2) and (B1-B2) frames. Then the resulting mask can be applied to a reduction of the separate AB pairs (A1 and B1, and A2 and B2).

Figure AP2-1A shows an (A-B) frames with procedures 1-3 completed. Figure AP2-2A shows a sky frame (B in this case) with procedures 1-3 completed.

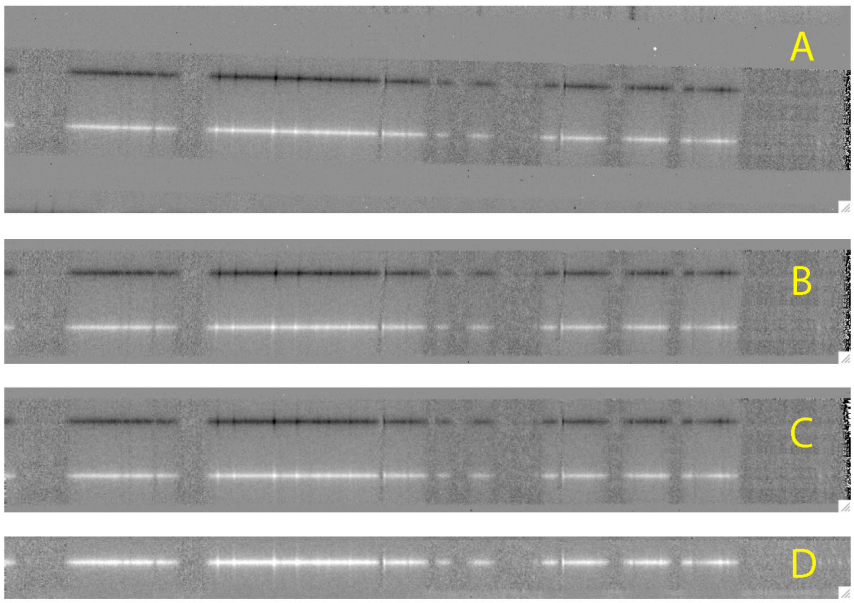

Fig. AP2-1 – Results from a standard processing sequence for an A-B difference. (A) A 1024x251 frame centered on a targeted order, following flat-fielding and bad-pixel correction (i.e., after applying steps 1-3, as described in the text); (B) spectrally-registered frame, using the GPEAK algorithm (step 4); (C) spectrally and spatially registered frame after the DETILT algorithm (step 5-6); (D) "ready for science" spectral frame in which the A- and the B-beams are combined together (step 9).

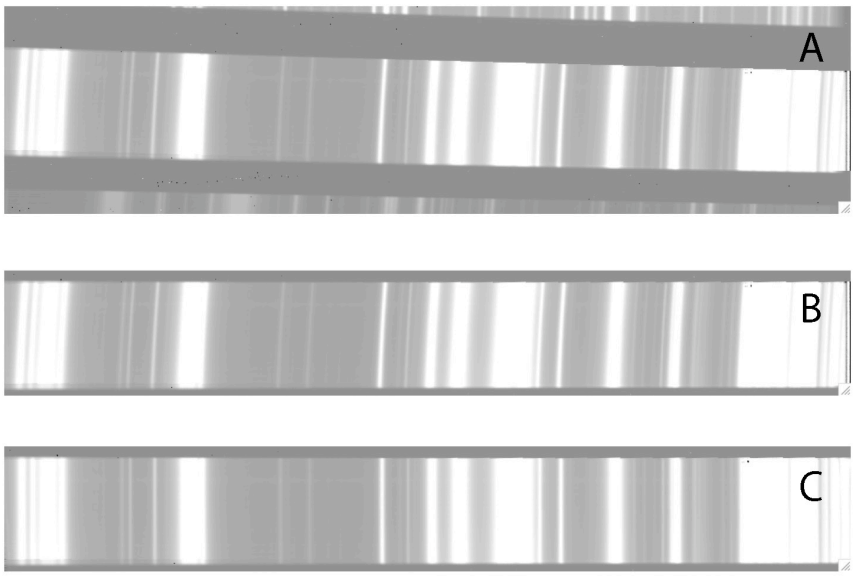

Fig. AP2-2 – Same as figure AP2-1 (A-C) but for a sky frame. Notice the change in the tilt of sky lines as the reduction progresses.

4. Spectral registration ("GPEAK" algorithm). The flat-field corrected and "masked" frames are then re-sampled, so that the spectral dimension falls along the rows of the 2-D array. The result is shown on Figure AP2-1B (A-B) and Figure AP2-2B (sky frame). The re-sampling (to the nearest 0.1 row) is done separately for each half of the order, i.e. separately for the A- and for the B-beam. For each beam a dual gauss fit is applied column-by-column (e.g. Figure AP2-3A). The row number corresponding to the "beam peak" (as determined from the fit) varies with column because the original order is tilted. The resulting peak rows are fit with the corresponding column pixel numbers via a second order polynomial (Figure AP2-3B). This fit approximates the tilt of the analyzed half of the order. The re-sampling procedure that follows is essentially taking this dependence out of the data, so that the spectral dimension falls exactly along the rows of the 2-D array and the beams are "straightened" (compare Figure AP2-1A to AP2-1B).

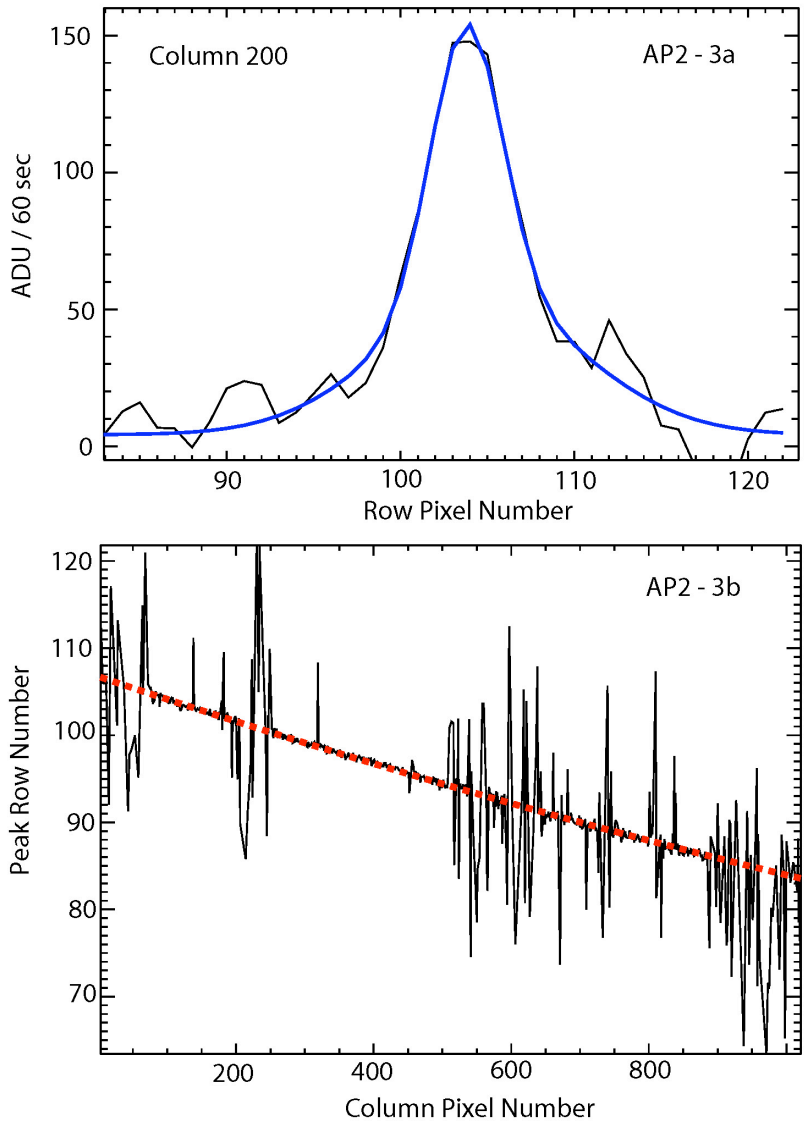

Fig. AP2-3 – The essence of the GPEAK algorithm. (A) a dual gauss fit to the intensity profile along a single column from order 23, KL2. (B) a second order polynomial fit between column numbers and their corresponding peak row numbers, determined from fits like the one shown in (A). The big deviations from the fit line correspond to spectral regions with low atmospheric transmittance.

There is a problem because this procedure applies well only when the comet has a strong continuum. Otherwise the gauss fit of course fails. The remedy in such situation is to apply GPEAK to an equivalent frame of a flux standard star. This frame is from the same setting and order and is cropped exactly the same way as the comet data. Stellar continuum is always strong, except for the columns corresponding to telluric absorption lines. The resulting second order polynomial fits can be used reliably in the spectral re-sampling of the comet frame. Note that after GPEAK the size of the working array is reduced from 1024 x 251 to 1024 x 151.

Figure AP2-2B shows a sky frame after GPEAK. Note that the sky lines are still tilted. The next two procedures are the most critical. They result in a frame that is "registered" in both dimensions: the spectral dimension falls along the rows and the spatial dimension falls along the columns of the resulting array (Figures AP2-1C, and AP2-2C).

5. Frequency calibration. The adopted relation between column pixel number (x) and wavenumber (ν) is:

$$\nu(x) = \nu_{cen} + disp \times |x - x_o| + disp2 \times (x - x_o)^2$$

The central pixel of the order is x_0. The adjustable parameters are the central wavenumber, ν_{cen}, and the first and second order "dispersion" terms, *disp* and *disp2* (not to be confused with $disp^2$). They are calculated separately for the left and right half of the order. It should be pointed out that this is not an exact functional relation, but just an approximation, which however leads to an adequate "straightening" of the sky lines.

The five calibration parameters (central frequency and two dispersion terms on each side) are adjusted by fitting an atmospheric radiance model to a 3-row sky extract

(Figure AP-4). Recall from Chapter 3 that the strongest signal in our comet data is in most cases the sky emission lines. These lines are used as a calibrator, instead of a lamp spectrum. The exact procedure is multi-step and involves deriving an initial solution for the free parameters. This then feeds into an iterative algorithm searching for the best fit in a frequency space. My experience is that the most critical part is providing a good set of initial values for the parameters. This part of the algorithm is in the process of being automated at present (just a reminder that the algorithms are constantly evolving as we gain more experience!).

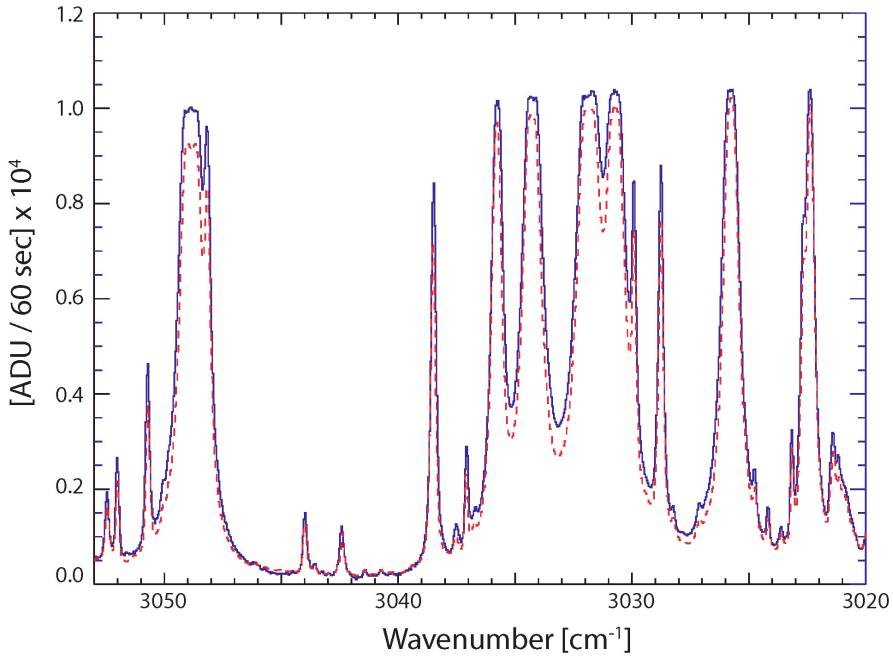

Fig. AP2-4 – A radiance spectrum fit to a 3-row sky spectral extract as a part of the frequency calibration procedure. The fit is only in frequency, not in intensity space. On principle an exact match in intensity is not strictly required for frequency calibration purposes, but would be helpful. This will be accomplished by the algorithm we are currently introducing for common usage (see the "Algorithm development" section for details).

This procedure is done initially for a 3-row sky extract near the bottom of the order (typically near row 20). I found that applying the algorithm for extracts centered at every other row (20, 22, …, ~130) is appropriate. This narrow grid is desirable because

the final iterative solution for a given row (e.g. 30) is used as an initial solution for the next row being fit (32). Of course such an approach was not possible before mid-2004 because we lacked the computational speed required to perform the operation efficiently.

The results of this fitting routine are saved in a "calibration" file (wd_out.dat). Here is an extract from such a file:

...

56 3429.5434 -0.0505018 -0.0503840 1.8266e-06 -1.2654e-06 0.99245

58 3429.5544 -0.0505018 -0.0503840 1.8206e-06 -1.2604e-06 0.99276

60 3429.5639 -0.0505118 -0.0503990 1.8176e-06 -1.2274e-06 0.99259

...

It contains central row number for a 3-row sky extract, ν_{cen} (cm^{-1}), *disp_left*, *disp_right* (cm^{-1} $pixel^{-1}$), *disp2_left*, *disp2_right* (cm^{-1} $pixel^{-2}$), and a correlation coefficient between radiance model and sky spectrum.

6. **Spatial re-sampling ("DETILT" procedure)**. The "calibration file" created in step 5 allows to empirically approximate the dependencies of the five free parameters on row number. Low order (typically 1 or 2) polynomials are used for this purpose:

$$\nu_{cen} = Polyfit1(y)$$
$$disp\_left = Polyfit2(y)$$
$$dist\_right = Polyfit3(y)$$
$$disp2\_left = Polyfit4(y)$$
$$disp2\_right = Polyfit5(y),$$

where y is the row number (ranging roughly between 20 and 130, with a step 2) and the indexing in *Polyfit1*, *Polyfit2*, etc. indicates not the polynomial order but the fact that each parameter is fit with a separate polynomial. Currently all polynomials have to be

from the same order. I have suggested to (may be) improve that by allowing the polynomial order to be adjusted separately for each of the five parameters.

The "DETILT" algorithm is a procedure for spatial re-sampling resulting in a "straightened order" in which the rows fall in the spectral dimension and the columns fall in the spatial dimension (compare the B- and the C- parts of Figures AP2-1 and AP2-2.

7. "DETILT-test". Comets are extended sources. The quality of the desired spatial information depends heavily on the accuracy of the last two steps in the data reduction. We have adopted a simple, but imperative test for the DETILT procedure. Such a test is shown on Figure AP-5. It presents two spectral (row) extracts from the straightened sky-frames (A and B). The one row is near the position of the A-beam (solid blue line), but extracted from the B-frame; the other row is near the position of the B-beam (dashed red line), extracted from the A-frame. The difference (black line) between these two row-extracts from the fully processed sky frames is indicative to the straightening quality.

Figure AP2-5A is an example for unacceptable straightening quality. The sky row difference shows "dispersion" line shapes, typical for residual tilt in the sky lines. In practice, this would propagate in subsequent analyses to less reliable spatial profiles and more problematical identification of nearby weak emission lines. Figure AP2-5B shows an improved and adequate straightening – there are still commonly seen differences in the baselines and in the levels of the two sky extracts (this results from the NIRSPEC optics, which map a sharper image of the slit in the B-beam position than in the A-beam position), but there are no "dispersion" line shapes indicative of inter-row frequency misregistration.

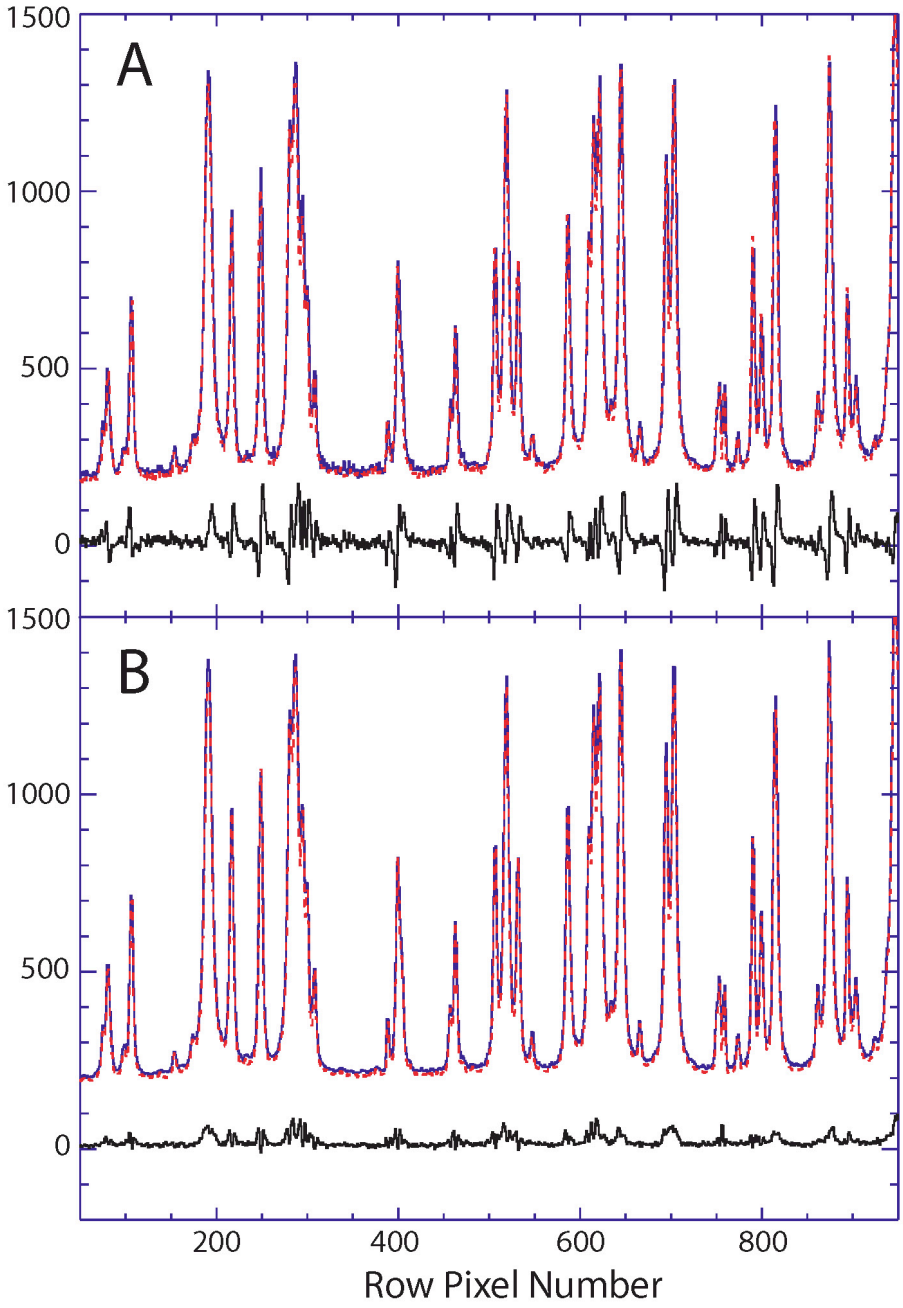

Fig. AP2-5 – The now imperative test for the quality of straightening applied to order 25, KL2 instrument setting (see text for details).

Figure AP2-5B is the best quality straightening we can now achieve. Based on extensive experience from reducing different orders in different data sets, I would say that the work leading to Figure AB2-5B pushes the limits of the current algorithms. DETILT-tests are performed on multiple row extracts and from each order analyzed. Figure AP2-6 shows another example of a DETILT-test, which is for the reduction shown on figures AP2-1 – AP2-4. While this is the best we can do now, our reduction techniques continue to improve.

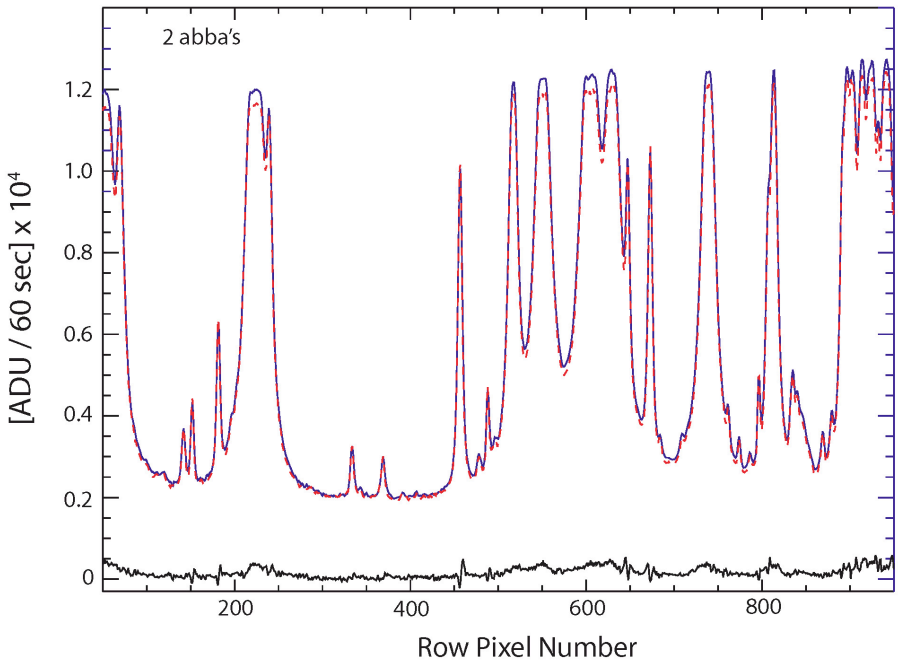

Fig. AP2-6 – Same as Figure AP2-5B, but for order 23, KL2. A few minimal dispersion line shapes are still seen, but the mis-registration is well within a small fraction of a pixel. We hope to overcome even this with the introduction of the GENLIN 2 spectral synthesis program.

8. <u>Algorithm development</u>. The most significant improvement in our reduction algorithms for the last 2.5 years has been achieved in the quality of the order straightening described in steps 4-6. Prior to mid-2004 the sky radiance model was fit to the data for only five row extracts from the sky frame (vs. every other row now), and a fit involving only two parameters (central wave number and first order dispersion) was

applied. Thus the dependence of the frequency calibration parameters was based on fitting just five points, separated by many rows. This could not approximate as well as now the strong dependences of these parameters on row numbers. In addition, the algorithm for fitting a single sky extract and a radiance model worked well for the CSHELL data (much more limited in frequency range) but in many cases could not converge for NIRSPEC data. This situation was changed expecting two bright comets in May 2004. Mike DiSanti put the new algorithms together and passed them to me to test them first on Mars and then on comet data. This is one of the two major improvements, without which the scope of this project would have been much more limited!

The current straightening algorithm has problems. The synthetic spectrum program (SSP, a later development based on Kunde & McGuire 1974) used to calculate the radiance models, introduces uncertainties in the frequency positions of the sky lines. Pressure shifts in line positions are not included and more importantly the calculated rest frequencies have limited precision.

The program version of SSP we use is not flexible toward fitting exactly emission line intensities, because it does not iterate over the atmospheric temperature profile (the line intensities of course depend strongly on temperature, because their source function [proportional to the Planck function] varies strongly with this parameter). This is seen of Figure AP2-4. The fact that we still achieve very good straightening (see the DETILT tests) is mainly because we need to fit in frequency space only, not in line intensity space. Nevertheless, the algorithms are at the process of being updated with a much better program for synthesis of radiance spectra (GENLIN 2, Reference). We expect GENLIN

2 to allow fitting the sky spectra in both frequency and intensity space, with improved precision.

9. <u>Combining A and B beams together</u>. This step is for data nodded "on chip" only. The half of the spectrally and spatially registered (A-B) frame (Figure AP2-1C) corresponding to the B-beam consists of rows 75-150. This half of the array is first multiplied by (-1). The peak rows of the A- and the inverted B-beam are then determined with a precision of about tenth of a pixel. The A- and the B- part of the order are combined into a 1024 x 75 array, in which the peak rows of the two beams coincide. We refer to this procedure as “cropping”. It has two advantages:

First - improvement in the S/N by a factor of up to √2. Second – an (A-B) or (A1-B1-B2+A2) difference does not cancel the sky perfectly. But the residual sky signal is canceled upon cropping the two beams together (at least for data taken in the L-band. In the M-band, where the sky is much brighter, residual sky can be an issue even after cropping).

Figure AP2-1D shows a frame with the two beams cropped together, which is ready (at last!) for spectral analysis.

A2.2 Spatial Registration Between Multiple Spectral Frames

The aforementioned basic processing steps are combined in a “batch” sequence which can be applied for an “A1, B1, B2, A2” set (one complete nodding cycle) or an AB pair. Except when taking “test exposures” of a certain object, the desire is of course to acquire multiple spectral frames. For example, one of the two principal data sets in this project is the KL2 setting of C/2000 WM1. It corresponds to 24 AB pairs or 48 minutes

time on source. After reducing all these AB pairs separately (applying steps 1-8, but not 9, from the previous sub-section), the A- and the B-frames are averaged, resulting in a mean A- and a mean B-frame with improved S/N. Ideally the A- and the B-beams of all reduced frames should fall at the same row position in the order. In other words they have to be spatially registered. In reality this is not always the case.

Comets are non-sidereal objects. Their proper motion is usually very significant and can present a challenge in tracking the object. Non-perfect tracking might be caused by various reasons (e.g. uncertain ephemeris) and it results in a comet drift along and/or across the slit. The latter situation is worse, because signal is lost. On the other hand drifting along the slit can be remedied by a spatial registration of reduced and "rectified" individual A- and B-frames. The following example is for comet C/2000 WM1, KL2 setting taken on Nov. 25, 2001:

Figure AP2-7 shows the measured position (with a precision of about 0.1 spatial pixel) of the A- and the B-beam peaks for each individual frame. The observing log for that night says "*Frame 260: Nodded unexpectedly. Comet jumped several times*". As seen from the figure, the central position of the comet along the slit has changed afterward. Therefore a straight stack of all the data would be spatially miss-registered.

Figure AP2-8 shows the integrated counts from continuum spatial profiles taken at each individual beam. The fact that there is not any trend evident in these plots indicates that the drift has been mostly along and not across the slit.

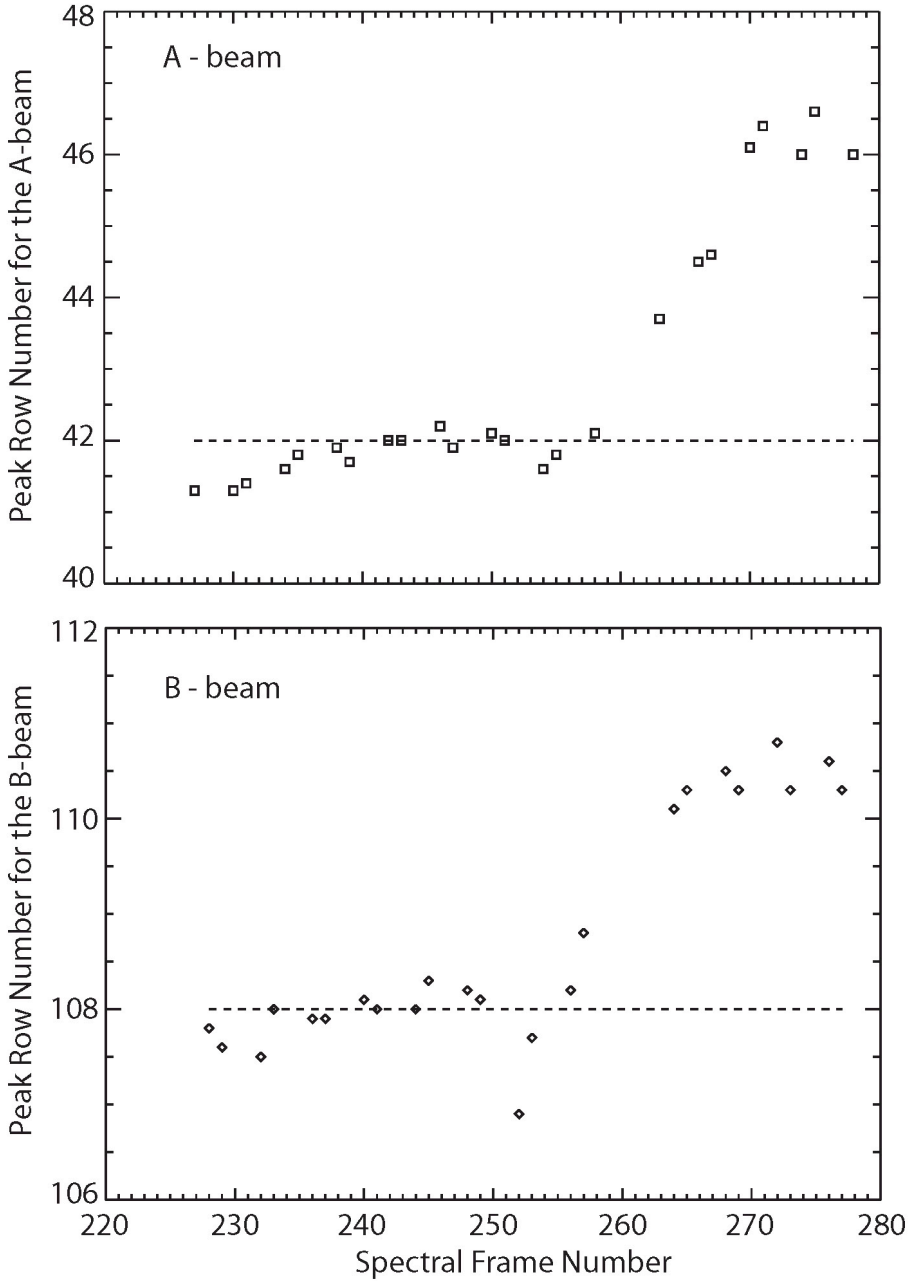

Fig. AP2-7 – Measurement of comet drift along the slit. The intensity of many continuum points has been summed row by row. The resulting spatial profile is fit with a Gaussian, similar to the one shown on Figure AP2-3A. The peak of this Gaussian approximates the "beam-peak", while the area under the Gaussian approximates the integrated intensity of the sampled part of the comet continuum.

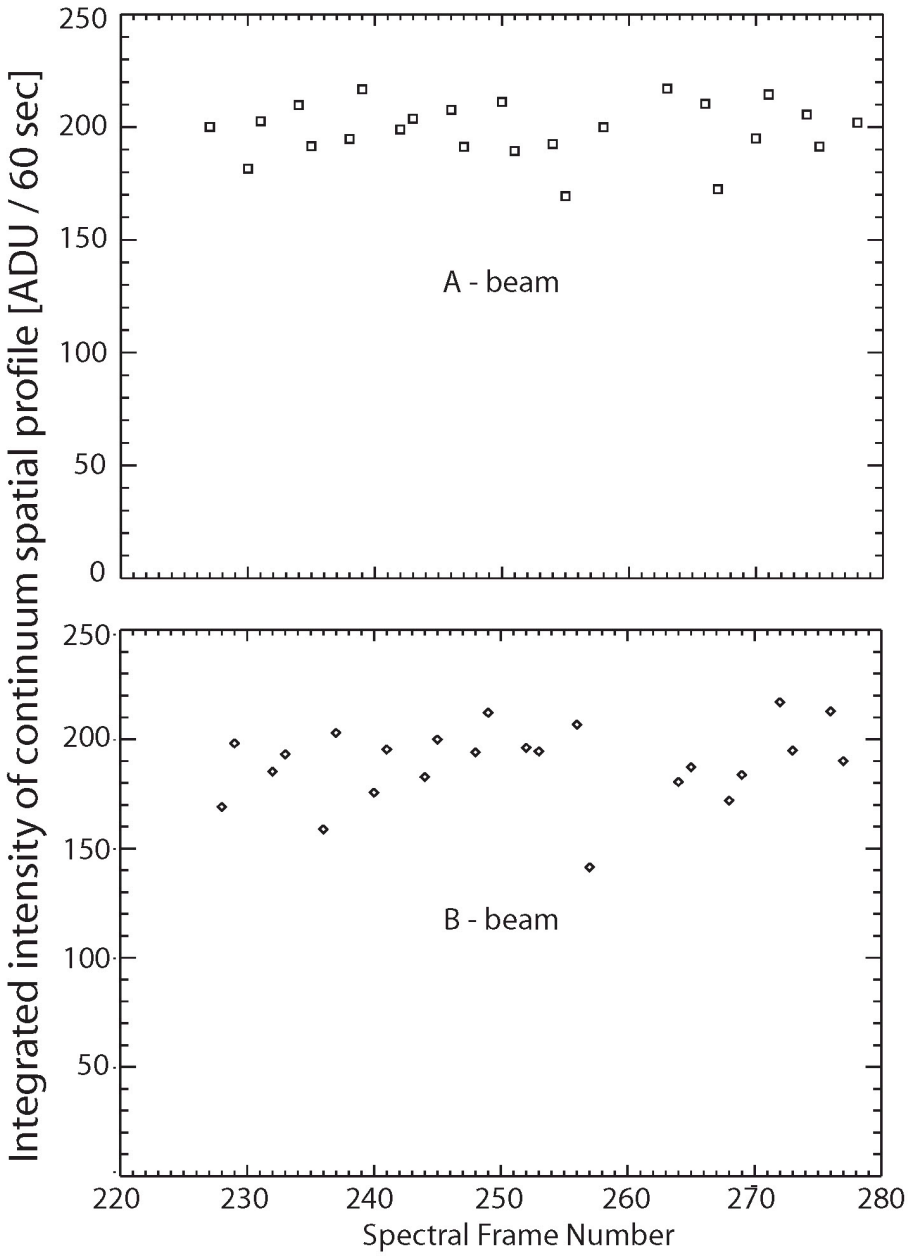

Fig. AP2-8 – The integrated continuum intensity does not show any systematic trend, indicating the comet drift has been mostly along the slit. Such drift is "preferable" versus a drift across the slit, because only in the former case there are no major flux losses from the comet signal.

The multiple reduced frames have been spatially registered (i.e. shifted in the spatial dimension), so that their respective beams peak at the same spatial position in the order. The result is shown on Figure AP2-9. The dashed red curve represents the spatial profile of the continuum intensity (averaged over many columns) for an unregistered stack of the WM1 data. This is done before step 9 of processing sequence, so the two beams are still separated. The solid black curve represents the spatial profile of the registered stack. This profile is much better peaked, justifying the registration procedure.

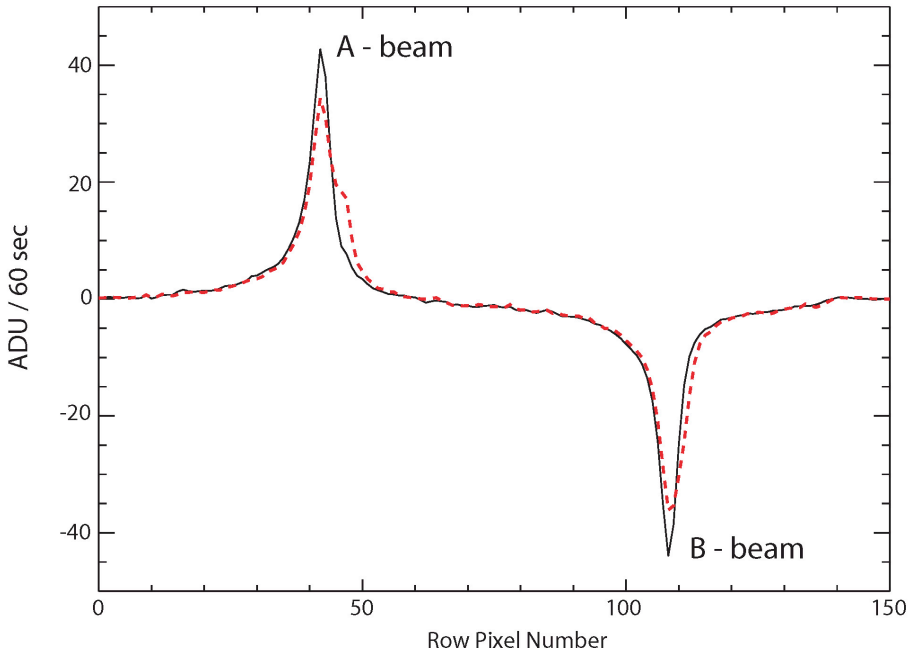

Fig. AP2-9 – Spatial profiles of the comet continuum intensity for a registered (solid) and non-registered stacks. The B-beam appears negative in a (A-B) difference.

WM1 was not a very bright comet. The spatial registration, though laborious at the time, has been critical for producing the science results on H_2O fluorescent emission and OH prompt emission presented in the main text.

A2.3 Flux Calibration

Synopsis:

This procedure derives the conversion factor Γ (also referred as calibration factor) between directly measured instrumental counts per unit time at a certain frequency (ADU sec^{-1}) and their corresponding flux density (W m^{-2} $[cm^{-1}]^{-1}$). There are a number of critical issues, some with a general significance, other – of particular importance for the data sets of interest in this work. Before discussing these issues in some detail, it would probably be beneficial for the reader to see them presented in a condensed form:

1. What is the variation of the calibration factor between different echelle orders within a given NIRSPEC setting?

2. Is there a substantial variation within a separate order? For which orders is such variation observed?

3. NIRSPEC does not have a wide enough slit to collect the whole flux from an IR flux standard star. How do we correct for slit losses?

4. What is more important to quantify precisely for the purposes of the comet work – the absolute flux calibration within a setting or the relative flux calibration between the orders of a given setting? How is the relative uncertainty between orders minimized, so errors in the absolute calibration do not affect the mixing ratio of one comet species to another?

5. The calibration factor is commonly derived from a set of several stellar frames? What criteria are used to choose the optimal value of Γ among separate measurements?

6. How does the relative calibration between orders within a given setting vary from date to date?

The main issue in flux calibration is to perform it in a consistent way between the orders of a NIRSPEC setting. This is the only way to minimize the calibration uncertainty (which is difficult to precisely quantify) when production rates are compared for deriving mixing ratios.

General principles:

The conversion factor (Γ) between directly measured instrumental counts per unit time at a certain frequency and their corresponding flux density is obtained from the following relation:

$$\Gamma = \frac{f_{star}(\nu)}{COUNTS(\nu)/t} \times \frac{\tau_\nu}{F_{cor}}$$

The unit of Γ is [W m^{-2} $(cm^{-1})^{-1}$ $(ADU/sec)^{-1}$]. The quantities on the right hand side are as follows:

$f_{star}(\nu)$ is the predicted flux density of a given IR flux standard star at a frequency ν (cm^{-1}). There are catalogues of IR standards for Mauna Kea (e.g. "*IRTF bright standards*"), which publish the flux standards' magnitudes for a given IR band. These catalogues are commonly posted on the web sites of different observatories. The choice of a standard star depends on the ephemeris of the principal target. We would like to observe the comet in the most favorable circumstances like air mass and maximized time on source for a particular setting. The flux standard is usually chosen to have a similar declination as the target, but to precede or trail it in right ascension by about 2^h, generally depending on whether the target comet is rising or setting. It is also desirable (though not

strictly required) to use an early-type star because it would have a "clean" IR continuum, without absorption features.

The stellar flux at a given frequency is estimated in two steps:

1. The published magnitude (m) for the effective center frequency (ν_0) of a given IR band is converted to flux density units using the standard magnitude-flux relation with Vega (α-Lyrae) as the zero magnitude reference:

$$f_{star}(\nu_0) = f_{m=0}(\nu_0) \times 10^{-0.4m}$$

2. The flux density at the desired frequency ($f_{\text{star}}[\nu]$) is derived from $f_{\text{star}}(\nu_0)$ via the Plank function at a temperature approximating the stellar spectral class (T_{star} , which strictly speaking is a color temperature):

$$f(\nu) = f(\nu_0)\frac{B_\nu(T_{star})}{B_{\nu_0}(T_{star})}$$

The stellar continuum intensity at a given frequency, COUNTS(ν) is measured from a nodded on chip spectrum, which is taken by using the widest slit available in order to collect maximum stellar flux. Unfortunately for NIRSPEC this is only 5-pixels (0.72 arcsec), which is not enough to collect the whole flux of the star. For data taken before November 2003, even a narrower slit (3 pixel; 0.43 arc-sec) was used. Therefore the measured counts have to be corrected for slit losses:

$$F_{cor} = \frac{\text{Predicted counts for an infinitely wide slit}}{\text{Measured counts at the real slit}}$$

The measured stellar counts (reduced to unit integration time) at a certain frequency also need to by divided by the corresponding atmospheric transmittance (τ_ν) to obtain the top-of-the-atmosphere equivalent count rate.

Basic procedure:

The stellar data are reduced using the same algorithms as for the comet. The main difference is that the two beams are not cropped together (step 9). Instead each beam is analyzed separately. There should be at least one (preferably more) A1-B1-B2-A2 sequence for the calibration star, in order to measure the stellar counts multiple times. This measurement is straightforward, so only the gist of it will be given. First the profile of the stellar intensity is measured along the entrance slit. A typical profile is shown on Figure AP2-10A, with schematic representation of the slit displayed on Figure AP2-10B. The integrated Gaussian-equivalent area under this profile is an estimate for the total stellar counts.

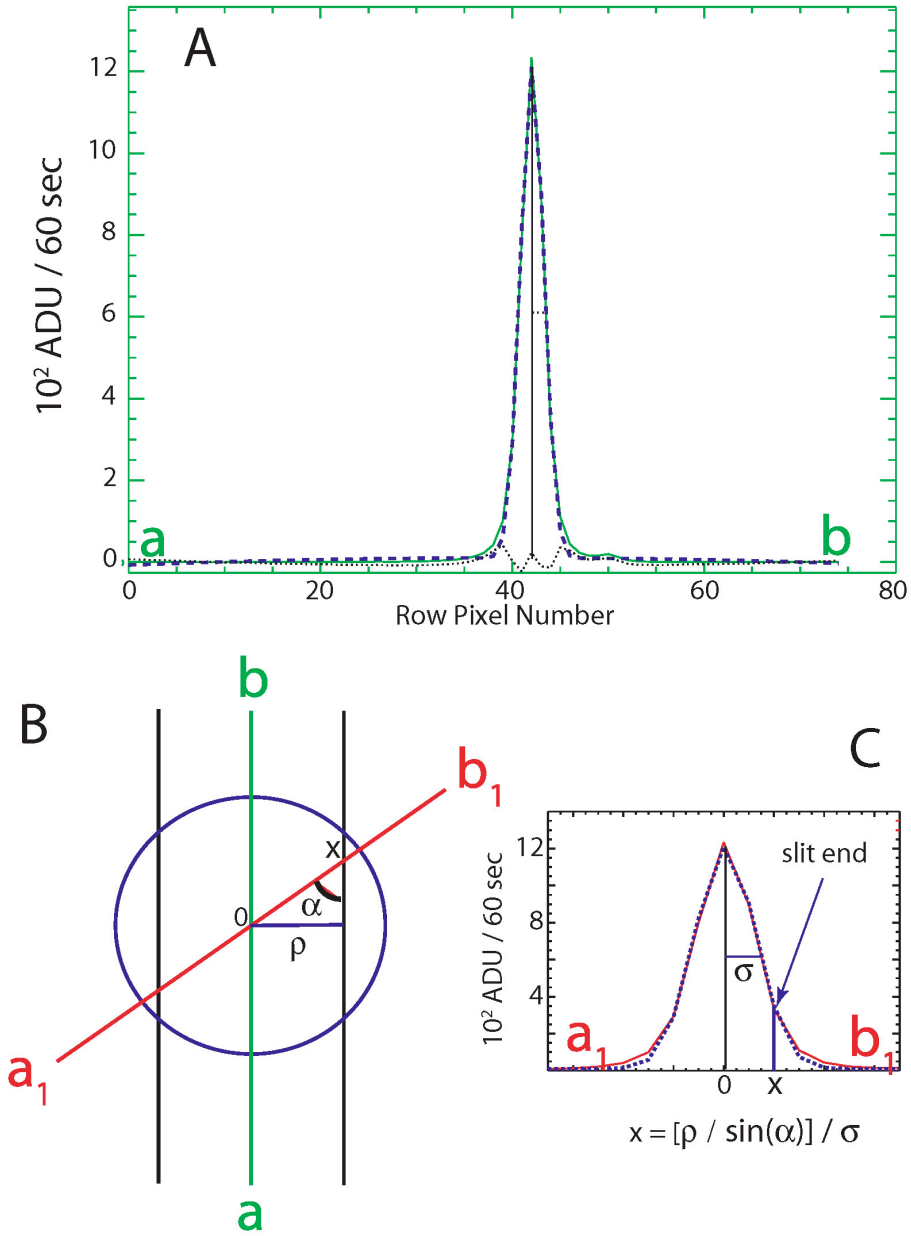

Fig. AP2-10 – (A) stellar continuum profile measured along the slit and a Gaussian fit (dashed line). The area under the fit approximates the total counts from the star; (B) schematic representation of the geometry involved in a slit loss correction – this correction assumes a perfectly centered object and azimuth symmetry in intensity profiles in the plane of the figure. (C) The central part of the intensity profile shown on panel A but this time considered as the profile that would be measured in the absence of a slit in an arbitrary direction a1-b1. The unit on the x-axis is σ – the profile's half width at half maximum.

Approximations:

In a strict sense all quantities in the expression for Γ except the correction factor and the integration time should refer to a particular frequency. Different approximations have been done to that end:

The calibration factor for properly flat-fielded data is normally assumed to be constant within a given echelle order. The stellar counts are measured throughout the spectral dimension of this order for frequencies with approximately equal and high terrestrial atmospheric transmittance (> 85-90%). For example the intensity profile shown on Figure AP2-10A represents the mean signal from multiple frequencies in order 26. The average transmittance over these frequencies is used. The stellar flux density $f_{star}(\nu)$ is calculated for the central frequency of the order.

This approach has been adopted in our group's previous NIRSPEC work. It takes into account the main variation of the calibration factor within a given NIRSPEC setting, namely - the change of Γ from order to order, which is generally larger than possible variations within a given order. Figure AP2-11 shows the variations of the calibrations factors within the KL2 setting for the observations of the two main comets discussed in this thesis (C/2000 WM1 and C/2004 Q2). On the other hand using essentially a mean calibration factor for a whole order ignores the variation of the stellar flux with frequency within that order.

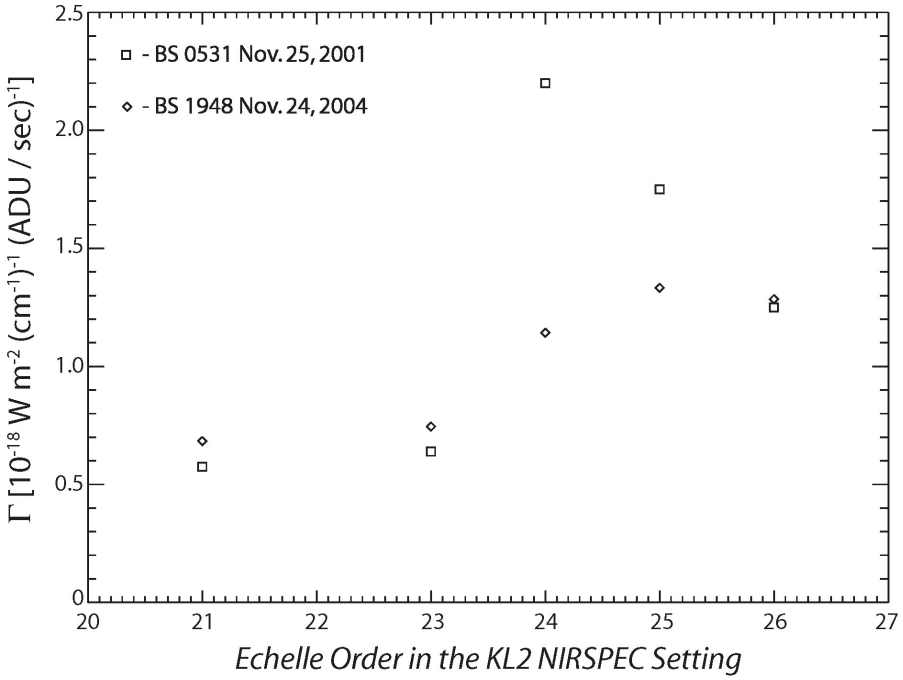

Fig. AP2-11 – Mean flux calibration for all analyzed orders of the KL2 NIRSPEC setting at two dates. Variations within each order are discussed in the text. The BS0531 set is used in the calibration of the C/2000 WM1 data; the BS1948 set - in the calibration of the C/2004 Q2 data. Note the tremendous difference for orders 24 and 25 between the two dates. Absorption from ice accumulated in the window in front of the instrument in November 2001 has caused a significant decrease in sensitivity in these orders. The calibration of order 25 for the 2001 data was done by Karen Magee-Sauer. All other calibration factors are derived by BPB.

In this thesis a higher order of approximation is used by considering not only order to order variation in Γ but also variation within a given order. For this purpose:

1. The stellar counts are measured locally in the order for relatively narrow frequency range. The terrestrial atmospheric transmittance is averaged only over this narrow range, within which τ is usually well-defined.

2. The stellar flux density $f_{star}(\nu)$ is calculated for the central frequency not of the whole order but of that particular narrow range.

In practice three or four values of Γ can be calculated within an order this way. This depends largely on the distribution of atmospheric opacity with frequency. The complication is that there is a limited choice of frequency ranges with high atmospheric transmittance, because of the presence of telluric lines seen in absorption against the stellar continuum (see the spectral gallery in Chapter 3).

No definitive evidence for a variation of the calibration factor in order 26 (which covers the highest frequencies) has been found. Interestingly, some increase in Γ from high to low frequency was noticed for orders 25, 24, 23, and 21. The effect is strongest in order 21 (which covers the lowest frequencies); we find about 15% increase from the high to the low frequency end of this order. This local variation in Γ is a newly discovered effect for us. I have discussed it with colleagues within the group. Our current working hypothesis is that it is an instrumental effect, related to flat-fielding and there is a consensus that it would affect the same way both comet and calibration star data. Therefore in this work "local" calibration factors have been used within a given order when appropriate. This is in principle more adequate in terms of the strict definition of Γ, which is a function on frequency.

Correction for slit losses:

The main problem in the flux calibration of NIRSPEC data is that even the widest entrance slit blocks part of the stellar signal. The correction factor for slit losses (F_{cor}) is calculated under the following assumptions:

1. The star has been perfectly centered in the slit.

2. The 2D distribution of the stellar continuum intensity is azimuthally symmetric.

These assumptions are illustrated on the already introduced Figure AP2-10. The B-part of this figure shows a schematic representation of the entrance slit. The circle symbolizes the stellar image, partly blocked by the slit. The stellar continuum profile is shown on panel A. It is measured along the slit
in the a-b direction as indicated. It is assumed that in the absence of a slit, a stellar profile taken in an arbitrary direction (a_1-b_1) would be identical to the measured profile, as shown on panel C. The slit loss, predicted in the a_1-b_1 direction is calculated as the ratio:

$$2 \times \frac{\int_{0}^{\infty} Gaussian}{\int_{0}^{x} Gaussian},$$

where Gaussian refers to the fit approximating the shape of the continuum profile and the coordinate x is in units of σ – the half width at half maximum of this Gaussian, as indicated in the figure. The integrated correction factor sums the slit losses over all directions.

Minimizing the overall uncertainty in flux calibration within a setting:

The two assumptions inherent in the derivation of F_{cor} are idealized, because atmospheric seeing causes both off-centering and possibly anisotropic intensity distribution in the plane of the drawing in Figure AP10-B. This is the principal reason limiting our ability to quantify the uncertainty in the flux calibration factor. It has been conservatively assumed that the flux calibration factor might be wrong by about 10%, however the underlying considerations for this are qualitative.

However the main issue with NIRSPEC is not the absolute calibration, but instead the relative calibration factors between the orders of a given instrument setting. We are primarily interested in mixing ratios between abundances from different species. There is a well-understood way to minimize the effect of slit losses in the relative (order-to-order) flux calibration uncertainty:

Recall that in practice the stellar counts are measured from a series of spectral frames. Each of these frames would be affected to a different extent by the atmospheric effects and would exhibit a different slit loss. On the other hand, the signal in the orders within a given NIRSPEC setting would be subject to the same slit loss for a single frame, aside from low order effects like varying optical distortion at different frequencies. Indeed, the correction factors we measure within a setting for the same stellar frame agree to a very high degree between different orders.

Table AP2-1 shows a representative example of flux calibration of two orders within the KL2 setting. It shows all relevant quantities included in the expression for Γ. Each individual stellar frame results in different correction factor, but the correction factors for a given frame agree among orders. Both orders use the calibration factors

derived from frame 90. Therefore these calibration factors would be almost equally affected by slit loss, so this factor would not affect the relative calibration. The critical implication of this is that mixing ratios between abundances of species detected in two different orders would be more reliable and relatively independent of the loosely bound uncertainty in the absolute flux calibration.

TABLE AP2-1
Flux calibration for two NIRSPEC orders

Frame Number[a]	Profile Height[b] [ADU / 5 sec]	Integrated Counts[c] [ADU / 5 sec]	Correction factor for slit losses	Γ [d,e] [Wm^{-2} $(cm^{-1})^{-1}$ $(ADU/sec)^{-1}$]
ORDER 23: f = 1.182E-14 W m^{-2} $(cm^{-1})^{-1}$; τ = 0.912				
89	14916	48299	1.06	1.05E-18
90	**16794**	**65174**	**1.11**	**7.45E-19**
91	16576	63570	1.11	7.66E-19
92	14165	50630	1.09	9.80E-19
93	14386	46720	1.06	1.09E-18
94	14871	62853	1.14	7.53E-19
95	12204	36407	1.05	1.42E-18
96	16350	38621	1.02	1.37E-18
ORDER 26: f = 1.495E-14 W m^{-2} $(cm^{-1})^{-1}$; τ = 0.914				
89	12030	36536	1.05	1.78E-18
90	**13553**	**48851**	**1.09**	**1.28E-18**
91	12980	47762	1.09	1.31E-18
92	10200	39048	1.11	1.58E-18
93	11456	35977	1.06	1.80E-18
94	12198	47610	1.11	1.29E-18
95	9956	28082	1.04	2.35E-18
95	12528	30497	1.02	2.20E-18

[a] – Spectral frames for IR flux standard BS 1948 taken on Nov. 28, 2004; exposure time = 5 sec.

[b] – Height of a Gaussian fit approximating the stellar intensity profile like the one shown on Figure AP2-10A.

[c] – The area under the Gaussian profile = Height $\sigma (2\pi)^{1/2}$, where σ is the profile's half width at half maximum.

[d] – Average calibration for the order as explained in the text.

[e] – The minimum calibration factor corresponds to the highest profile with largest integrated counts. This factor corresponds to the same frame (90) in both orders.

Finally, what governs the choice of Γ among results from individual frames? It depends on a case by case basis, but in the vast majority of situations the lowest value is preferred – the one corresponding to the highest counts collected from the star if this is also the tallest (best peaked) stellar intensity profile. In theory this would imply the smallest slit loss. Indeed, for data taken with the narrower (3-pixel) slit we always observed an inverse correlation between collected counts and correction factors: the highest intensity profiles like the one shown on Figure AP2-10A are also the narrowest and provide the minimum Γ. However for 5-pixel slit data taken under favorable atmospheric conditions, the spread in correction factors could be small. The lowest calibration factor still should correspond to the highest number of collected counts and the best peaked intensity profile. However it does not necessarily correspond to the smallest F_{cor}. One has of course always to be reminded that the assumptions in calculating the correction factor do not exactly hold. Therefore the value of Γ based on the largest collected stellar flux and tallest profile is still preferable.

If the spread among individual calibration factors is very small and there is no best peaked outlier, a mean instead of minimum Γ can be chosen, considering it would average over the relatively small slit losses of individual frames. The important issue is again to use the same choice (mean vs. minimum Γ) for all orders.

For the flux calibration of the C/2000 WM1 spectra a mean calibration factor is used within each order; the calibration of C/2004 Q2 is based on minimum Γ because the spread of individual calibration factors within an order was much larger, but there was a definite outlier with best peaked profile as seen in Table AP2-1. The calibration in the C/2004 Q2 data set imposed a serious additional problem, which cannot not overlooked.

Flux calibration for the C/2004 Q2 data taken on January 19, 2005:

The data on C/2004 Q2 taken on UT January 19, 2004 is a valuable set to quantify the overall chemistry of this comet. After initial data reduction and detection of multiple species in non-calibrated spectra we found that the stellar frames for that night were saturated, so we could not use them for flux calibration. This situation might seem hopeless, but instead it has a reasonable solution which has not compromised the scientific merit of this data set. Stellar flux standard spectra, taken about a month and a half earlier under similar atmospheric conditions, were found to be suitable for calibrating the January data. It should be pointed out that using calibration data from a different night has not been an automatic decision, but a subject of the following consideration:

All major results in our work, including the ones in this thesis depend critically not on the absolute but on the relative calibration between the orders within a setting, as discussed above. Of course the relative calibration within a NIRSPEC setting might change between observing runs for various reasons, a major one being ice accumulating at times on the window in front of the instrument, as mentioned in the caption of Figure AP2-11. We examined this possibility by a direct comparison between the intensity levels over different NIRSPEC orders in flat field frames taken on Nov. 28, 2004 and Jan. 19, 2005. This comparison is shown on Figure AP2-12.

A date-to-date change in the relative levels among the orders in a dark-corrected flat field would reflect a change of the relative calibration between these orders (in a given NIRSPEC setting). However, the comparison between dark-corrected flats from both days indicates that the relative calibration has not changed between these days to an

extent that would affect any important result based on the ratio between production rates (i.e. mixing ratios of different species, and OH equivalent g-factors, which are the subject of Chapters 5 and 6 in the main text). Therefore using stellar calibration data from Nov. 28, 2004 in the analysis of the comet data from Jan. 19, 2005 is justified.

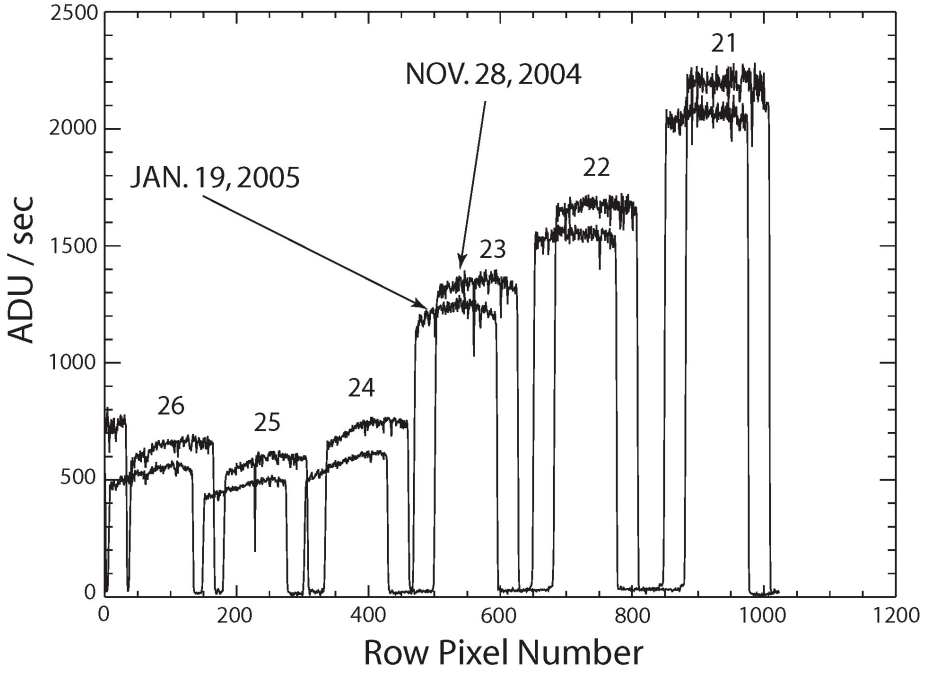

Fig. AP2-12 – Column profiles from flat-field frames taken on Nov. 28, 2004 and Jan. 19, 2005. Order numbers are indicated in the plot. The two frames are corrected for dark current and their signal is scaled to an integration time of one second. The two column profiles are shifted because the cross-disperser position for the KL2 setting was slightly changed on the second date. Note that the relative mean intensities between the separate NIRSPEC orders for a given date is very similar in both dates.

This case is not an unique example for "stability" in the relative calibration from date to date. Similar behavior has been verified in our data sets for cases in which there are adequate flat field and star frames for both compared dates (so that the flat field test for a change in relative calibration could be verified). However it should be stressed that such stability cannot be assumed a-priori, but must always be verified by a direct comparison such as the one presented on Figure AP2-12.

Moreover, our absolute calibration is probably also adequate at worst. The H_2O production rate derived in Chapter 4 of the main text falls within the range of predicted comet productivity for early January based on previous observations. While this is a semi-quantitative test, it indicates the lack of a drastic change in Γ between the dates in question. This is consistent with the experience of senior group members, that unless there is an accumulated ice on the window in front of the instrument set-up, the calibration factors for flat fielded data derived on different dates are similar.

Summary on flux calibrations issues:

1. The principal effect to quantify is the variation of the calibration factor between different echelle orders of a given NIRSPEC setting.

2. Additional variation (~10%) is found and accounted for within orders, the effect being more profound at increasing wavelength (i.e. with decreasing order number).

3. A principal uncertainty in the absolute flux calibration is the correction for slit losses. This uncertainty is minimized for the purpose of taking mixing ratios by determining the calibration factor from a set of measurements in consistent way for all orders.

4. The choice of optimal calibration factor is decided on a case by case basis, but usually the chosen value corresponds to the stellar intensity profile with maximum height and integrated stellar counts, which results in a minimum value of Γ. If all values of Γ (obtained from multiple stellar frames) are close and there is not an outlier as described in the previous sentence, an average value is also appropriate.

5. The relative mean intensities between the orders in the KL2 setting of dark-corrected flat field frames were found to be very similar for two different observing

dates. This similarity indicates that the relative calibration between the setting's orders has not changed significantly between these two dates. Therefore using calibration factors from the first date in the analysis of the comet data from the second date would still provide reliable mixing ratios

Discussing these calibration problems, I hope I have successfully conveyed an additional (an substantial) advantage of cross-dispersed versus single-order instruments. The former not only cover broader frequency range, but also allow more consistent reduction of important yet difficultly quantified uncertainties, which become systematic for the whole setting if an appropriate reduction or flux calibration approach is employed.

Appendix 3: Observing Logs

TABLE AP3-1

Comet	UT date	R_h [AU]	Δ [AU]	Δ–dot [km/sec]	T_{int} [min]	NIRSPEC Setting	Orders analyzed
C/1999 Lee	19 Aug 1999	1.049	1.381	-28.4	10	KL2	23
C/1999 Lee	21 Aug 1999	1.076	1.348	-29.0	24	KL2	23
C/2001 A2	9 Jul 2001	1.160	0.285	11.4	28	KL2	23
C/2001 A2	10 Jul 2001	1.173	0.282	12.4	40	KL2	23
C/2000 WM1	25 Nov. 2001	1.321	0.357	-19.1	48	KL2	21,23,24,26
C/2004 Q2	28 Nov. 2004	1.493	0.655	-21.8	20	KL2	23
C/2004 Q2	19 Jan. 2005	1.208	0.394	11.0	8	KL2	21,23,24,25,26
9P/Tempel 1	04 Jul 2005	1.506	0.894	9.2	28	KL1	26,27

Notes: R_h, Δ, and Δ-dot are respectively heliocentric and geocentric distance, and radial velocity with respect to the observing site; T_{int} is total integration time on source. Only the NIRSPEC settings analyzed as a part of this thesis are listed in the table.

The major results in this work are based on C/2000 WM1 and C/2004 Q2 (Jan. 19, 2005) and are presented in detail in Chapters 4, 5, and 6. The last section in Chapter 4 is about the observation of 9P/Tempel-1 which was a part of the ground-based observing

campaign in support of the NASA "Deep Impact" mission. The contribution of this thesis to the 9P/Tempel-1 observations has been presented in Mumma et al. (2005). Work on C/1999 (Lee) and C/2001 A2, relevant to this project, is described by Mumma et al. (2001a) (only on the first comet), Bonev et al. (2004), and Dello Russo et al. (2005). The main results from Bonev et al. are shown in Chapter 5. The results from Dello Russo et al. are referred to in Chapter 4, and those from Mumma et al. – in Chapter 5.

Appendix 4: Basic Statistics

Two simple statistical procedures are involved in the analysis presented in Chapters 4, 5, and 6: deriving a weighted mean and a weighted linear slope. Their related physical implications are discussed in detail in the main text. This appendix summarizes the used apparatus from elementary statistics and basic error analysis. The gist is to distinguish between the two errors used: the "stochastic" and "standard" error. The term "stochastic" implies that this error describes the uncertainty in derived production rates, rotational temperatures, and ortho-to-para ratios induced by photon noise. The "standard" or "variance" error reflects the spread of the derived quantities around their mean or linear slope line and relates to uncertainties beyond the photon noise limits.

The used expressions have been compiled from three literature sources:

1. Bevington & Robinson (1992), for the definition of stochastic errors and of weighted sample variance of a mean and of a linear fit.

2. Hoel (1984), for the treatment of the "standard" errors.

3. Arkin & Colton (1970) which together with Hoel's text is used for the formal treatment of small number statistics.

Definition of weighted mean of individual measurements y_i ($i = 1, .., N$) with uncertainties σ_i (in reference to production rates, σ_i is the propagated flux measurement error):

$$y_{mean} = \frac{\sum_i y_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2}$$

Stochastic error of a weighted mean:

$$\sigma_{stoch} = \sqrt{\frac{1}{\sum_i \frac{1}{\sigma_i^2}}}$$

"Standard" or "variance" error of a weighted mean:

$$\sigma_{std} = \frac{s}{\sqrt{N}}, \quad where$$

$$s^2 = \frac{1}{N-1} \frac{\sum_i \frac{(y_i - y_{mean})^2}{\sigma_i^2}}{\frac{1}{N} \sum_i \frac{1}{\sigma_i^2}}$$

The sample variance is indicated as σ^2 and relates to the spread of the individual measurements around their mean value.

Linear fit of quantity y (for e.g. "F/g" in Chapter 4) vs. x (e.g. upper state energy):

$$y_{fit} = a + b(x)$$

"Stochastic" error of the slope in a linear fit:

$$\sigma_b^{stoch} = \sqrt{\frac{1}{\Delta}\sum_i \frac{1}{\sigma_i^2}}, \quad where$$

$$\Delta = \sum_i \frac{1}{\sigma_i^2}\sum_i \frac{x_i^2}{\sigma_i^2} - \left(\sum_i \frac{x_i}{\sigma_i^2}\right)^2$$

"Standard" or "variance"error in a slope of a linear fit:

$$\sigma_b^{std} = \frac{s_y}{s_x\sqrt{N-1}}, \quad where$$

$$s_y^2 = \frac{1}{N-2}\frac{\sum_i \frac{(y_i - y_{fit}[x_i])^2}{\sigma_i^2}}{\frac{1}{N}\sum_i \frac{1}{\sigma_i^2}}$$

$$s_x^2 = \frac{\sum_i (x_i - x_{mean})^2}{N-1}$$

Note that both errors depend on the spread in the x-value which in the rotational temperature analysis reflects the fact the this quantity is better constrained when a wide range of rotational levels is sampled.

Formal 95% confidence intervals

The formal 95% percent confidence interval of a weighted mean or slope for a large sample of measurements is ~1.96 standard errors if the latter are normally distributed. In the limit of a small sample the multiplier in front of the standard error becomes larger than 1.96 and is tabulated for various sample sizes. This multiplier is derived from a variable with the Sturdent's t-distribution. For example for a sample of 6 points the formal 95% confidence interval corresponds to 2.57 standard errors of a mean, and to 2.78 standard errors of a linear slope.

The most important principle is not to apply these errors and confidence intervals "blindly" but first to understand the assumptions behind their formal derivations, and second to consider their meaning in light of the particulat physical problem at hand. The interpretation of 95% confidence intervals is most important in the discussion of ortho-to-para ratios deduced when only a few orhto and (especially) para lines are available.

Appendix 5: Lower vs. Upper state Energy in the H_2O Rotational Temperature Retrieval

Chapter 4 describes the derivation of the H_2O rotational temperature (T_{rot}). The main idea is that for the right value of this parameter, the quantity F/g should be independent on the upper state energy (E_{up}), where F designates the flux measured at the expected position of a particular line and corrected for atmospheric transmittance. The line g-factor is temperature dependent. The line fluxes reflect the rotational population in the upper vibrational state, which in turn depends on the ground state's rotational population. The latter is described by T_{rot}, assuming Boltzmann distribution is a good approximation (which is normally the case as discussed in the main text). Therefore the measured line flux depends on E_{up}, but also on $E_{low}(T_{rot})$ through E_{up}.

The question raised in Chapter 4 is whether or not using E_{low} instead of E_{up} in the T_{rot} retrieval would result in different temperature value and uncertainty. It has been expected that both approaches would provide very similar temperatures, but it has not been clear if the uncertainty would be the same when using lower versus upper state energy.

Previously the same problem was examined by Dello Russo et al. (2004) based on analysis of CSHELL data with the conclusion that using either parameter provides

consistent results. However their temperature uncertainties were significantly larger (and therefore easier to overlap) than the ones reported here, so the problem has to be re-examined.

This issue is investigated, by a direct comparison of the two alternative T_{rot} retrievals in C/2000 WM1. The optimal value derived in Chapter 4 from the $F/g(T_{rot})$ vs. E_{up} relation is 75 K. First, this temperature was fixed and the slope of "$F/g(T_{rot}$ =75 K) vs. E_{low} (T_{rot} =75 K)" was evaluated. Figures AP5-1A and AP5-1B show respectively "F/g vs. E_{up}" and "F/g vs. E_{low}" diagrams. The slope in the latter representation is zero within error: $(1.2 \pm 3.3)x10^4$ (95% conf.). Therefore the rotational temperature value would be very close to 75 K if inferred independently from the "F/g vs. E_{low}" relation. Indeed, an independent retrieval from the "F/g vs. E_{low}" slope variation with T_{rot} results in 76 K.

What about the uncertainty? Statistically it depends on the error in the zero slope, which in turn depends on the spread in the x-axis. In the case of "F/g vs. E_{up}" the x-axis (E_{up}) is fixed upon varying T_{rot} because the upper state energy does not depend on T_{rot}. In the case of "F/g vs. $E_{low}(T_{rot})$", the x-axis is not fixed because the weighted lower state energy varies with temperature. The other factors influencing the uncertainty are discussed in the main text and remain unchanged.

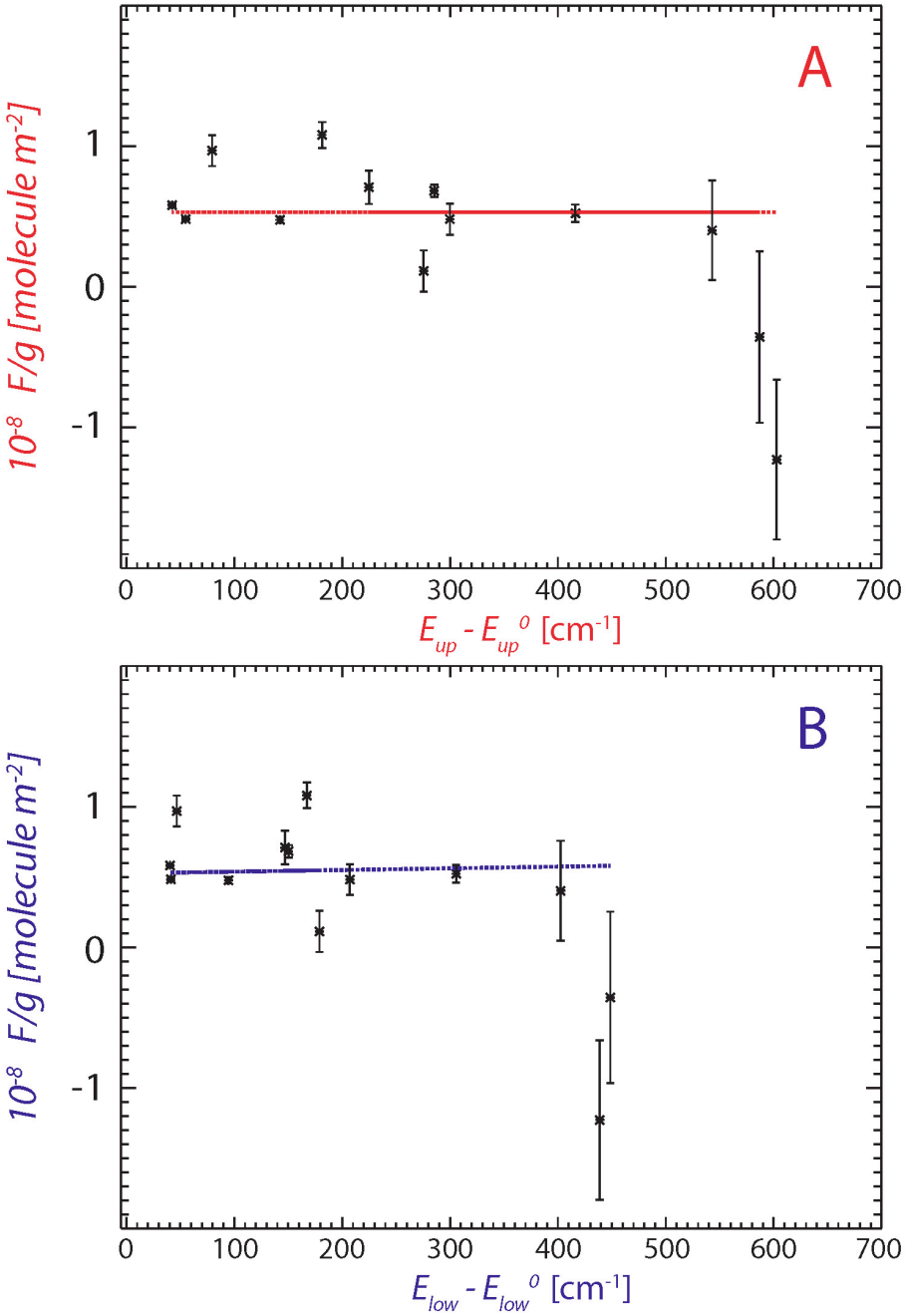

Fig. AP5-1 – "F/g vs. E_{up}" and "F/g vs. E_{low}" diagrams for the optimal temperature reported in Chapter 4 (75 K). This temperature was found using the upper state energy in the retrieval. For the same value of T_{rot} the slope in panel B (when the lower state energy is used instead) is zero within error ([1.2 ± 3.3] 10^{-4}; 95% conf.), indicating that both approaches are consistent.

For a given T_{rot}, the spread in energies is smaller than in the E_{low} case, as seen on Figure AP5-1B. So for the optimal rotational temperature, the zero-slope's uncertainty is larger in the "F/g vs. $E_{low}(T_{rot})$" representation. But how this propagates to uncertainty in the rotational temperature?

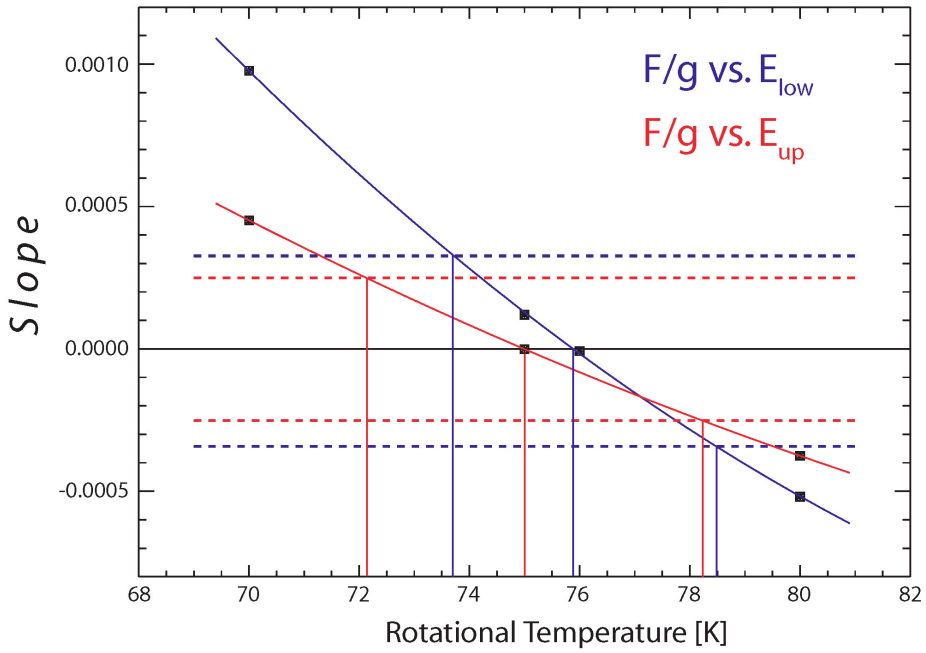

Fig. AP5-2 – A direct comparison between retrieving the uncertainties in rotational temperatures using "F/g vs. E_{up}" (red) and "F/g vs. E_{low}" (blue). The solid curves are polynomial fits, used to the obtain the T_{rot} value and its uncertainties. Formal 95% confidence intervals in the zero-slope are indicated with the dashed lines. Note the two factors determining the rotational temperature uncertainty: the uncertainty in the zero slope of "F/g vs. E_{up}" or "F/g vs. E_{low}" diagrams (e.g. Figure AP5-1) and the rate of slope change with rotational temperature.

Figures AP5-2 shows how the slope varies with T_{rot} in the "F/g vs. E_{up}" and "F/g vs. $E_{low}(T_{rot})$" representations. To emphasize the difference, only the near optimal temperature range in covered. While the zero-slope error is larger in the $E_{low}(T_{rot})$ case, the slope transits faster from positive to negative values as T_{rot} increases. As a result the uncertainty in T_{rot} remains very similar to the one retrieved in the "F/g vs. E_{up}" situation. Using E_{low} results in $T_{rot} = 76^{+3}_{-2}$ K, against $T_{rot} = 75 \pm 3$ K (95 % conf.) when using E_{up} (values and uncertainties are rounded to the nearest integer). Therefore both approaches

provide practically the same most probable rotational temperature ranges.

Statistically the faster change in the slope with T_{rot} is caused by the same reason as the larger zero-slope error – the smaller spread in energies in the E_{low} representation. Physically the two representations should have very similar outputs (as shown here) for both rotational temperature and its uncertainty because populating the E_{up}-levels, depends on the $E_{low}(T_{rot})$ distribution.

References for Chapters 1-7

A'Hearn, M. F., Millis, R. L., Schleicher, D. G., Osip, D. J., & Birch, P. V. 1995, Icarus, 118, 223

A'Hearn, A. F., et. al. 2005, Science, 310, 258

Abrams, Mark C., Davis, S. P., Rao, M. L. P., & Engleman Jr., R., 1990, ApJ, 363, 326

Alexander, M. H. et al. 1988, J. Chem. Phys., 89, 1749

Andresen, P., Ondrey, G. S., Titze, B., & Rothe, E. W. 1984, J. Chem. Phys., 80, 2548

Arkin, H., & Colton, R. R. 1970, Statistical Methods (New York: Barnes & Noble)

Bensch, F., & Bergin, E. A. 2004, ApJ, 615, 531

Bernstein, G. M., Trilling, D. E., Allen, R. L., Brown, M. E., Holman, M., Malhotra, R., AJ., 128, 1364.

Bertaux, J. L. 1986, A&A, 160, L7

Bevington, P. R., & Robinson, D. K. 1992, Data Reduction and Error Analysis for the Physical Sciences (New York: McGraw-Hill)

Biver, N., et al. 1999, AJ, 118, 1850

Biver, N., et al. 2002, Earth Moon Planets, 90, 323

Bockelee-Morvan, D., & Crovisier, J. 1989, A&A, 216, 278

Bockelee-Morvan, D., Crovisier, J., Mumma, M. J., & Weaver, H. A. 2004, In Comets II, ed. M. C. Festou, H. U. Keller, & H. A. Weaver (Tucson: The University of Arizona Press)

Bonev, B. P., Mumma, M. J., Dello Russo, N., Gibb, E. L., DiSanti, M. A., & Magee-Sauer 2004, ApJ, 615, 1048

Brin, C. D., & Mendis, D. A. 1979, ApJ, 229, 402

Brooke, T. Y., Tokunaga, A. T., Weaver, H. A., Crovisier, J., Bockelee-Morvan, D., & Crisp, D. 1996, Nature, 383, 606

Brooke, T. Y., Weaver, H. A., Chin, G., Bockelee-Morvan, D., Kim, S. J., & Xu, L.-H. 2003, Icarus, 166, 167

Brunini, A., & Melita, M. 2002, Icarus, 160, 32

Budzien, S. A., & Feldman, P. D. 1991, Icarus, 90, 308

Budzien, S. A., Festou, M. C., & Feldman, P. D. 1994, Icarus, 107, 164

Carrington, T. 1964, J. Chem. Phys., 41, 2012

Crovisier, J., & Encrenaz, T. 1983, A&A, 126, 170

Crovisier, J., 1984, A&A, 130, 361

Crovisier, J. 1989, A&A, 213, 459

Crovisier, J., Leech, K., Bockelee-Morvan, D., Brooke, T. Y., Hanner, M. S., Altieri, B., Keller, H. U., & Lellouch, E. 1997, Science, 275, 1904

Crovisier, J., et al. 1999, in The Universe as Seen by ISO, ed. P. Cox & M. F. Kessler (ESA SP-427; Noordwijk: ESA), 137

Crovisier, J., Bockelee-Morvan, D., Biver, N., Colom, P., Despois, D., & and Lis, D. C. 2004, A&A, 428, L35

Dello Russo, N., DiSanti, M. A.,Mumma, M. J.,Magee-Sauer, K., & Rettig, T. W. 1998, Icarus, 135, 377

Dello Russo, N., Mumma, M. J., DiSanti, M. A., Magee-Sauer, K., Novak, R., & Rettig, T. W. 2000, Icarus, 143, 324

Dello Russo, N., Mumma, M. J., DiSanti, M. A., & Magee-Sauer, K. 2002, J. Geophys. Res., 107(E11), 5095

Dello Russo, N., DiSanti, M. A., Magee-Sauer, K., Gibb, E. L., Mumma, M. J., Barber, R. J., &Tennyson, J. 2004, Icarus, 168, 186

Dello Russo, N., Bonev, B. P., DiSanti, M. A., Mumma, M. J., Gibb, E. L., Magee-Sauer, K., Barber, R. J., &Tennyson, J. 2005, ApJ, 621, 537.

Delsemme, A. H. 1981, L'Astronomie, 95, 293

Delsemme, A. H. 1989, Adv. Space Res., 9, 25

Delsemme, A. H. 1998, Plan. Space Sci., 47, 125

Delsemme, A. H. 2000, Icarus, 146, 313

Delsemme, A. H., & Miller, D. C. 1971, Plan. Space Sci., 19, 1229

Delsemme, A. H., & Swings, P. 1952, Annales d'Astrophysique, 15, 1

Dieke, G. H., & Crosswhite, H. M. 1962, J. Quant. Spectrosc. Radiat. Transfer, 2, 97

DiSanti, M. A., M. J. Mumma, N. Dello Russo, Magee-Sauer, K., Novak, R., Rettig, T. W. 1999, Nature, 399, 662

DiSanti, M. A., Mumma, M. J., Dello Russo, N., & Magee-Sauer, K. 2001, Icarus, 153, 361

Dones, L., Levison, H., Duncan, M., Weissman, P. 2000, BAAS, 32, 1060

Dones, L., Weissman, P. R., Levison, H. F., & Duncan, M. J. 2004, In Comets II, ed. M. C. Festou, H. U. Keller, & H. A. Weaver (Tucson: The University of Arizona Press)

Drapatz, S., Larson, H. P., & Davis, D. S. 1987, A&A, 187, 497

Engel, V., Schinke R., & Staemmler V. 1988, *J. Chem. Phys.* 88, 129

Espinasse, S., Coradini, A., Capria, M. T., Capaccioni, F., Orosei, R., Salomone, M., Federico, C. 1993, Plan. Space. Sci., 41, 409

Feldman, P. D., Cochran, A. L., & Combi, M. R. 2004, In Comets II, ed. M. C. Festou, H. U. Keller, & H. A. Weaver (Tucson: The University of Arizona Press)

Festou, M., & Feldman, P. D. 1981, A&A, 103, 154

Festou, M. C., Keller, H. U., & Weaver, H. A. 2004, In Comets II, ed. M. C. Festou, H. U. Keller, & H. A. Weaver (Tucson: The University of Arizona Press)

Fink, U., & Hicks, M.D. 1996, ApJ, 459, 729

Gibb, E. L., Mumma, M. J., Disanti, M. A., Dello Russo, N., & Magee-Sauer, K. 2002, In: Proceedings of Asteroids, Comets, Meteors - ACM 2002, Ed. Barbara Warmbein. ESA SP-500, ESTEC (Noordwijk), 705

Gibb, E. L.,Mumma,M. J., Dello Russo, N., DiSanti, M. A., & Magee-Sauer, K.2003, Icarus, 165, 391

Gladman, B. 2005, Science, 307, 71

Gomes, R., Levison, H. F., Tsiganis, K., & Morbidelli, A. 2005, Nature, 435, 466

Greene, T. P., Tokunaga, A. T., & Carr, J. S., 1994, in Infrared Astronomy with Arrays: The Next Generation , ed. I. McLean, (Kluwer), 511

Harich, S. A., Hwang, D. W. H., Yang, Xuef., Lin, J. J., Yang, Xuem., & Dixon, R. N. 2000, J. Chem. Phys., 113, 10073

Hausler, D., Andresen, P., & Schinke R. 1987, J. Chem. Phys., 87, 3949

Herzberg, G. 1988, The Spectra and Structure of Simple Free Radicals (New York: Dover)

Hoban, S., Mumma, M. J., Reuter, D. C., DiSanti, M., Joyce, R. R. 1991, Icarus, 93, 122

Hoel, P. G. 1984, Introduction to Mathematical Statistics (New York: Wiley)

Horanyi, M., Gombosi, T. E., Gravens, T. E., Korosmezey, A., Kecskemety, K., Nagy, A. F., & Szego, K. 1984, ApJ, 278, 449

Hughes, D. W. 1991, In Comets in the Post-Halley Era, ed. R. L Newburn, Jr., M. Neugebauer, & J. Rahe (Kluwer), 825

Irvine, W. M., Schloerb, F. P., Crovisier, J., Fegley, Br., & Mumma, M. J. 2000, In Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russel (Tuscon: The University of Arizona Press)

Kawakita, H., Watanabe, J., Kinoshita, D., Ishiguro, M., & Nakamura, R. 2003, ApJ, 590, 573

Kawakita, H., Watanabe, J., Furusho, R., Fuse, T., & Boice, D. C. 2005a, ApJ. 623, L49

Kawakita, H., et. al. 2005b, BAAS, 37, 647

Kunde, V. G., & Maguire, W. C. 1974, J. Quant. Spectrosc. Radiat. Transf., 14, 803

Lardner, D. 1853, MNRAS, 13, 188

Larson, H. P., Mumma, M. J., & Weaver, H. A. 1987, A&A, 187, 391

Larson, H. P., Weaver, H. A., Mumma, M. J., & Drapatz, S. 1989, ApJ, 338, 1106

Lecacheux, A., et al. 2003, A&A, 402, L55

Levison, H. F. 1996, in Completing the inventory of the Solar system (ASP Conference Series), 107, 173

Levison, H. F., Morbidelli, A. 2003, Nature, 426, 419

Magee-Sauer, K., Mumma, M. J., DiSanti, M. A., Dello Russo, N., & Rettig, T. W. 1999, Icarus, 142, 498

Magee-Sauer, K., Mumma, M. J., DiSanti, M. A., & Dello Russo, N. 2002, J. Geophys. Res., 107, E11, 6

Maillard, J. P., Chauville, J., & Mantz, A. W. 1976, J. Mol. Spectrosc., 63, 120

Markwick, A. J., & Charnley, S. B. 2003, BAAS, 968

McLean, I. S., et al. 1998, Proc. SPIE, 3354, 566

Meech, K. J., et. al. 2005, Science, 310, 265

Mendis, D. A., & Brin, C. D. 1977, Moon, 17, 77

_____ 1978, Moon and Planets, 18, 77

Miani, A., & Tennyson, J. 2004, J. Chem. Phys., 120, 2732

Morbidelli, A., Chambers, J., Lunine, J. I., Petit, J. M., Robert, F., Valsecchi, G. B., & Cyr, K. E. 2000, Meteoritics Plan. Sci., 35, 1309

Morgenthaler, J. P. 2001, ApJ, 563, 451

Mumma, M. J. 1982, in Vibrational-Rotational Spectroscopy for Planetary Atmospheres, ed. M. J. Mumma, K. Fox, & J. Hornstein (NASA CP-2223; Washington: NASA), 717

Mumma, M. J., Weaver, H. A., Larson, H. P., Williams, M., & Davis, D. S. 1986, Science, 232, 1523

Mumma, M. J., Weaver, H. A., & Larson, H. P. 1987, A&A, 187, 419

Mumma, M. J., Blass, W. E., Weaver, H. A., & Larson, H. P. 1989 in Formation and Evolution of Planetary Systems, ed. H. A. Weaver & L. Danly (Cambridge: Cambridge Univ. Press), 157

Mumma, M. J., Weissman, P. R., & S. A. Stern 1993, in Protostars and Planets III, ed. E. H. Levy and J. I. Lunine (Univ. Ariz. Press, Tucson), 1177

Mumma, M. J., DiSanti, M. A., Tokunaga, A., & Roettger, E. E. 1995, BAAS, 27, 1144

Mumma, M. J., DiSanti, M. A., Dello Russo, N., Fomenkova, M., Magee-Sauer, K., Kaminski, C. D., & Xie, D. X. 1996, Science, 272, 1310

Mumma, M. J., et al. 2001a, ApJ, 546, 1183

Mumma, M. J., et al. 2001b, Science, 292, 1334

Mumma, M. J., DiSanti, M. A., Dello Russo, N., Magee-Sauer, K., Gibb, E. L., & Novak, R. 2003, Adv. Space Res., 31, 2563

Mumma, M. J., et al. 2005, Science, 310, 270

Nizkorodov, S. A., Ziemkiewicz, M., Myers, T. L., & Nesbitt, D. J. 2003, J. Chem. Phys., 119, 10158

Petit, J. M., Morbidelli, A., & Valsecchi, G. B. 1999, Icarus, 141, 367

Rothman, L. S., et al. 1992, J. Quant. Spectrosc. Radiat. Transf., 48, 469

Rothman, L. S., et al. 2003, J. Quant. Spectrosc. Radiat. Transf., 82, 5

Spinrad, H. 1982, PASP, 94, 1008

Stern, S. A. 2003, Nature, 424, 639

Stern, S. A., & Weissman, P. R. 2001, Nature, 409, 590

Taylor, J. R. 1982, An Introduction to Error Analysis (Mill Valley: University Science Books)

Tennyson, J., Kostin, M. A., Barletta, P., Harris, G. J., Polyansky, O. L., Ramanlal, J., & Zobov, N. F. 2004, Comput. Phys. Commun., 163, 85

Tielens, A. G. G. M., & Allamandola, L. J. 1987, in Interstellar Processes, ed. D. J. Hollenbach & H. A. Thronson, Jr. (Dordrecht: Reidel), 397

Weaver, H. A., & Mumma, M. J. 1984, ApJ, 276, 782

Weaver, H. A., Feldman, P. D., McPhate, J. B., A'Hearn, M. F., Arpigny, C., & Smith, T. E. 1994, ApJ, 422, 374

Weaver, H. A., Feldman, P. D., A'Hearn, M. F., Arpigny, C., Brandt, J. C., & Stern, S. A. 1999a, Icarus, 141, 1

Weaver, H. A., Chin, G., Bockelee-Morvan, D., Crovisier, J., Brooke, T. Y., Cruikshank, D. P., & Geballe, T. R. 1999b, Icarus, 142, 482

Weissman, P. R., Asphaug, E, & Lowry, S. C. 2004, In Comets II, ed. M. C. Festou, H. U. Keller, & H. A. Weaver (Tucson: The University of Arizona Press)

Whipple, F. L. 1950, ApJ, 111. 375

Xie, X., & Mumma, M. J. 1996, ApJ, 464, 457

Yamashita, I. 1975, J. Phys. Soc. Japan, 39, 205

References for Chapter 8

Antoniadi, E. M. 1930, La Planete Mars, Herman, Paris

Benson, J. L., & James, P. B. 2005, Icarus, 174, 513

Bjorkman, J. E, & Wood, K. W. 2001, ApJ, 554, 615

Bonev, B.P., James, P.B., Bjorkman, J.E., & Wolff, M.J. 2002, Geophys. Res. Lett., 29, 2017

Bonev, B. P., James, P. B., Wolff, M. J., Bjorkman, J. E., Hansen, G. B., Benson, J. L. 2003, Sixth International Conference on Mars, July 20-25 2003, Pasadena, California, abstract no. 3111

Clancy, R.T., Wolff, M.J., Christensen, & P.R., 2003. J. Geophys. Res. 108 (E9), 5098.

Eluszkiewicz, J., Moncet, J. L., Titus, T. N., & Hansen, G. B. 2005, 174, 524

Glenar, D. A., Hansen, G., Bjoraker, G., Blaney, D., Smith, M., & Pearl, J. C., 2005, Icarus 174, 600

Hansen, G. B. 1997, J. Geophys. Res., 102, 21,569

Hansen, G. B. 1999, J. Geophys. Res., 104, 16,471

Hansen, G. B., et al. 2005, Planet. Space. Sci., 53, 1089

James, P. B., Briggs, G. Barnes, J. & Spruck, A. 1979, J. Geophys. Res., 84

James, P. B., Malolepszy, K. M., Martin, L. J. 1987, Icarus, 71, 298

James, P.B., Kieffer, H.H., & Paige, D.A. 1993, In: Mars, eds. Kieffer, H.H., et al. (Tucson: University of Arizona Press, Tucson), 934.

James, P. B., et al. 2000, Icarus, 144, 410

James, P. B., Cantor, B. A. & Davis S. 2001, J. Geophys. Res., 106, 23,635

James, P. B., Hansen, G. B., & Titus, T. N. 2005a, Adv. Space Res., 35, 14

James, P. B., Bonev, B. P., & Wolff, M. J. 2005b, Icarus, 174, 596

Kieffer, H. H., Titus, T. N., Mullins, K. F., &Christensen, P. R. 2000, J. Geophys. Res., 105, 9653

Lumme, K., & James, P.B. 1984, Icarus, 58, 363

Mitchel, O. M. 1846, Sidereal Messenger, 2, 101

Paige, D.A., Ingersoll, A.P., 1985, Science, 228, 1160

Smith, M. D., Pearl, J. C., Conrath, B. J., & Christensen, P. R. 2001, J. Geophys. Res., 106, 23,929

Smith, M. D., Conrath, B. J., Pearl, J. C., & Christensen P. R. 2002, Icarus, 157, 259

Thomas, P. C., Malin, M. C., James, P. B., Cantor, B. A., Williams, R. M. E., & Gierasch, P. 2005, Icarus, 174, 535

Titus, T. N., & Kieffer, H. H. 2002, Lunar and Planetary Science XXXIII, Houston, Texas, abstract no.2071

Tomasko, M. G., Doose, L. R., Lemmon, M., Smith, P. H., & Wegryn E. 1999, J. Geophys. Res., 104, 8987

Wolff, M.J., & Clancy, R.T., 2003. J. Geophys. Res. 108 (E9), 5097.